

Impacts of DNA Microarray Technology in Gene Therapy

Yadollah Omidi, Amir Ata Saei and Jaleh Barar
*Research Center for Pharmaceutical Nanotechnology,
Faculty of Pharmacy, Tabriz University of Medical Sciences,
Tabriz,
Iran*

1. Introduction

Invention of microarray technology in the early 1990s revolutionized the genomic research, where it was exploited for global gene expression profiling through screening the expression of thousands of genes simultaneously (Watson et al., 1998). Northern blotting and reverse transcription polymerase chain reaction (RT-PCR) which were the traditional techniques for monitoring changes in mRNA levels were replaced with high density arrays, which proved to impart comprehensive data and to be better in the course of time.

Nowadays, microarrays are used for genotyping, large scale single nucleotide polymorphism analysis, comparative genomic hybridizations, identification of new genes, establishing genetic networks and as a more routine function, gene expression profiling (Bednar, 2000). Providing a unique tool for the determination of gene expression at transcriptomic level, it confers simultaneous measurement of large fractions of the genome which can facilitates the recognition of specific gene expression patterns associated with a certain disease or a specific pharmaceutical. Detection of the inimitable genomic signature of any active compound is deemed to be another important application of microarray technology, upon which "intrinsic genomic impacts" of any pharmacologically active agent can be clarified. And presumably, such toxicity predication may promise notable information about each individual resulting in unique patterns of gene expression that, in turn, exhibit individual specific responses to a particular toxic substance.

Basically, the discovery of diagnostic biomarkers has been the most promising feature of microarrays up to now and microarray technology has shown a great potential in predicting gene function and tracking the interactions among genetic networks too (Xiang & Chen, 2000). Microarray methodology has also been applied for analysis of proteins and metabolites, which are the principle controllers of gene expression, to verify the results at the molecular level. This in turn can extend our understanding of gene expression patterns and molecular pathways even though other techniques such as NMR, mass spectroscopy, gas and liquid chromatography can be employed for metabolite profiling. Having exploited such techniques, for example, the identity and quantity of different molecules can be determined in the CSF, urine or any other biological sample. Thus, by merging the classical techniques with new high-density microarray, the "omics" technology has been devised and

implemented for investigation on genomics, proteomics, cytomics and metabolomics (Jares, 2006; Rosa et al., 2006). The microarray is by far one of the best tools for pursuing such impacts. For implementation of this technology, however, one needs to be familiar with the different methods used in microarray data analyses and the ways for more efficient applications of such methods to enhance the output of a microarray screening process. Fig.1 represents schematic steps of microarray (Saei & Omid, 2011).

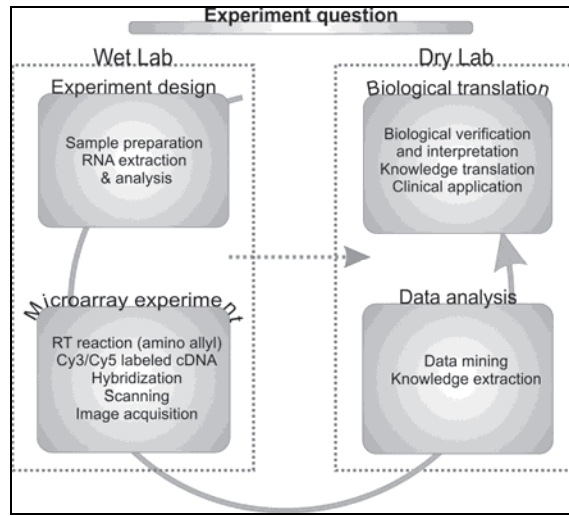


Fig. 1. Schematic steps of microarray, adapted with permission from (Saei & Omid, 2011).

The main focus of this chapter is to provide the importance of DNA microarray technology in gene therapy, focusing on the application of DNA microarray technology (wet lab approach) and also giving simple descriptions on microarray data analysis as well as knowledge extraction (dry lab approach). In the next sections, we discuss some important information on the DNA microarray gene expression profiling.

2. DNA microarray technology

DNA microarrays consist of microscopic arrays of large sets of DNA sequences that are immobilized on solid substrates such as glass (slide array which are identical in dimensions to microscope slides) used as powerful tools for identification and/or quantification of many specific DNA sequences. Fig. 2 shows schematic illustration of spotted genes on a glass slide array.

For printing of arrays of the desired genes, glass slide arrays are commonly used. In practice, prior to spotting the genes, the surface of the microarray slides are uniformly coated with a chemical compound which is able to interact with and immobilize nucleic acids irreversibly and attach them onto the surface. Nucleic acids (e.g., clone library, cDNAs, PCR products, long oligonucleotides, microRNA) are then deposited on the coated surface using contact printing, after which the slides are further treated (using UV light or baking at 80°C) to crosslink the genes spots onto the slide surface. The printed slide arrays are normally stored (at room temperature and dark) until required for experimentation (Hobman et al., 2007).

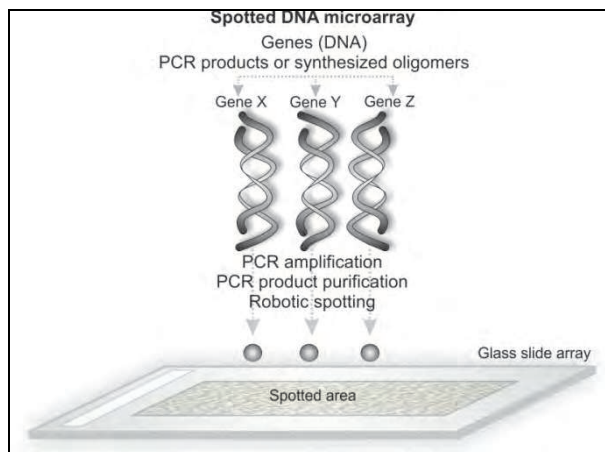


Fig. 2. Schematic illustration of spotted genes on a glass slide array. Glass slide arrays are produced by the robotic spotter that spots genes (e.g., PCR products, cDNAs, clone libraries or long oligonucleotides) onto coated glass slides. Each spot on the array represents a particular contiguous gene fragment, i.e. 40–70 nucleotides for oligonucleotide arrays, or several hundred nucleotides for PCR products, adapted with permission from (Saei & Omidi, 2011).

In DNA arrays, each spot on the array corresponds to a particular single gene. To spot greater numbers of genes on the slide arrays, the genes must be miniaturized and printed using tiny amount of biological material. The high throughput methods of sample preparation tend to be restricted in the amount of biological materials and concomitant with miniaturization of the array there has also been a trend towards decreasing the amount of biological material (e.g. RNA) used in the array experiments. Such approach can make the entire process much easier and more cost-effective. It should be also noticed that the fabrication of microarray screening systems are swiftly moving towards nanoarrays (Chen & Li, 2007), upon which some limitations of microarrays such as requirement for relatively large sample volumes and elongated incubation time can be resolved. We use the glass slide arrays commercially available by the Ocimum Biosolutions (www.ocimumbio.com/) mainly for gene expression profiling in cancer nanomedicine studies. The spotted glass slide arrays are now widely being used for research purposes as their fabrication is simple and cost-effective even though the spotting of genes tend not to be a simple approach by means of a printing machine (spotter). For printing the glass slide microarrays, the noncontact deposition and contact deposition methods have been widely used. The first uses ink-jet printing technology to deposit small droplets of liquid of pre-synthesized oligonucleotides or other biological materials onto a solid surface. The second uses physical deposition of small volumes of liquid from a metal pin onto the slide surface. Both of these technologies exploit microfluidics/nanofluidics approaches.

Basically, nucleic acid hybridization is the keystone of the DNA microarray technology. When denatured, single stranded nucleic acids are incubated together under certain conditions, hybridization can occur, upon which the formation of base paired duplex molecules can be prompted through G:C and/or A:T hydrogen bonds base pairing. As a matter of fact the nucleic acids hybridization process may be influenced by the

concentration and complexity of the sample, which also can be improved through manipulation of time, temperature and ionic strength of the hybridization buffer (Hobman et al., 2007). Maladjustment of these parameters may lead to some inadvertent mismatches in the nucleic acid duplex, resulting in undesired outcomes. Removal of mismatched hybrids is generally conducted under increasing stringency (i.e., by decreasing salt concentration and/or increasing temperature). Given that the removal of mismatched hybridization is difficult and may affect the end point results of the experiment, thus the more suitable experimental design with best variable adjustment should be considered (Hobman et al., 2007).

In general, some major applications and/or technologies of DNA microarrays are: 1) gene expression profiling, 2) pathogen detection and characterization, 3) comparative genome hybridization (CGH), 4) mutation and single nucleotide polymorphism (SNP) detection, 5) genotyping, 6) whole genome resequencing, 7) determining protein-DNA interactions using ChIP-chip (also known as ChIP-on-chip), 8) regulatory RNA studies, 9) alternative splicing detection (exon junction array), 10) RNA binding protein analysis, 11) microRNA studies, and 12) methylome analysis (DNA methylation profiling).

Recently, the US Food and Drug Administration (FDA) has conducted a project named "MicroArray Quality Control (MAQC)" upon which FDA aims to develop standards and quality control metrics which will eventually allow the use of microarray data in drug discovery, clinical practice and regulatory decision-making procedures. MAQC has gone to phase III now (i.e., MAQC-III, also called Sequencing Quality Control (SEQC)). The MAQC-III aims at: 1) assessing the technical performance of next-generation sequencing platforms, through generating benchmark datasets with reference samples, and 2) evaluating the advantages and limitations of various bioinformatics strategies in RNA and DNA analyses (Shi, 2008).

In practice, DNA microarray can be divided into two main dimensions, i.e., wet lab (Fig. 3A) and dry lab dimensions (Fig. 3B). The main steps of experimental approach for transcriptomics DNA microarray are shown in Fig. 3. Based on experiment question, the arrays are printed (Fig. 2) or the desired format of commercially available slide arrays are purchased. The wet lab experiments consists of: 1) sample preparation and total RNA extraction, 2) RT reaction and cDNA labeling (preparation for direct or indirect labeling, e.g., using aminoallyl tagged dUTP), 3) cyanine dyes (Cy3/Cy5) post labeling, 4) co hybridization of labeled cDNAs (Cy3-cDNA and Cy5-cDNA), 5) washing and scanning, 6) image acquisition, and 7) data analysis.

3. Wet lab experiments

3.1 Experimental design and sample preparation

A key to successful implementation of the DNA microarray technology is experimental design and quality of samples, i.e. the more accurate design, the more reliable microarray data. The user must make precise questions/hypotheses to be addressed through recruiting such technology, so are the statistical implications (e.g., replications, power analysis, clustering, principal component analysis (PCA), etc.). It has to be clarified that whether comparisons made on the microarray are to be direct or indirect (by making comparisons within or between slides, the so called type I/II experiments), and also whether experimental design needs to be single or multi-factorial (Shi, 2008).

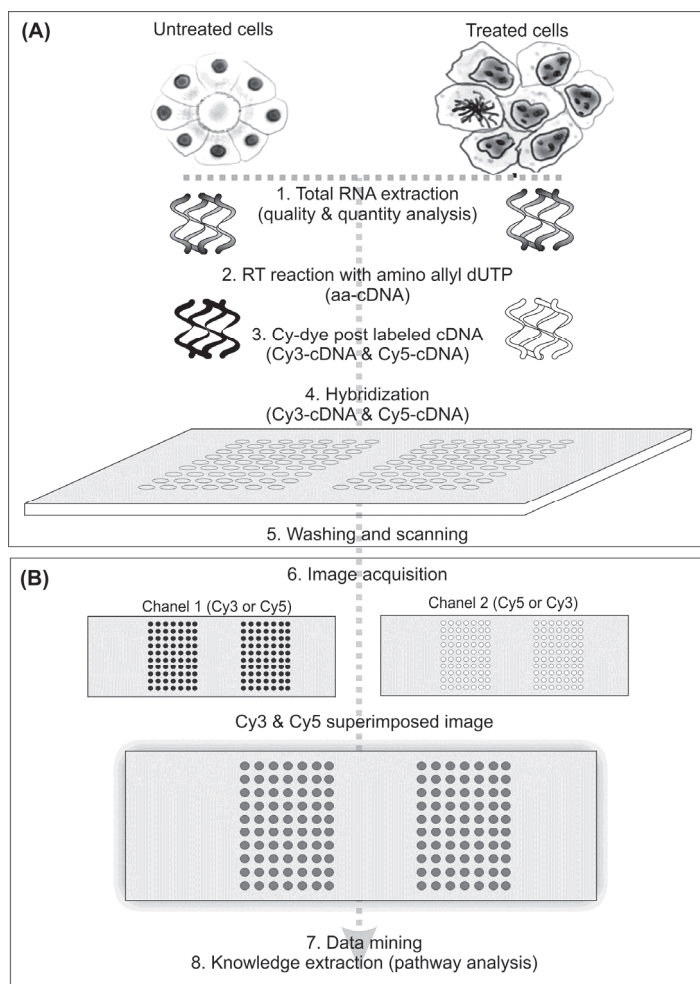


Fig. 3. Main steps of the experimental approach of transcriptomics DNA microarray for wet (A) and dry (B) lab experiments; adapted with permission from (Saei & Omidi, 2011).

In sample preparation, usually total RNA is isolated from the designated sample (e.g., animal/plant cells). Total RNA (10-20 μg), in the case of indirect labeling, is then converted to cDNA using tagged dNTPs (e.g., aminoallyl dUTP) and labeled with cyanine dyes (Cy3/Cy5) before hybridization to the array. Fundamentally, of critical importance to any successful transcriptomics experiment is believed to be the quality, quantity, and integrity of the isolated RNA (total RNA or mRNA) used in microarray experiments. Among various commercially available RNA isolation kits (e.g., from Sigma-Aldrich, Ambion, Qiagen, and Promega), we have successfully used TriReagent™ from Sigma-Aldrich Co. (Omidi et al., 2003). It should be evoked that most of these isolation kits are designed for quick, easy and reliable RNA extraction. Besides, in general, it is crucial to choose the best working protocol for RNA extraction considering vulnerability of RNA for degradation since the RNA even in

the cell appears to show instability and short half-life to some extent. Successful use of commercially available kits such as RNAprotect™ (Ambion) and RNAlater™ (Qiagen) has been also reported. In the case of total RNA use, it should be noticed that only a small fraction of total RNA is mRNA, thus 10-20 µg of total RNA is commonly used to obtain sufficient mRNA for a transcriptomics microarray experiment. However, prior to conducting microarray experiment, the quality and quantity of the isolated total RNA has to be checked - the electrophoresed total RNA must be intact, showing 28s and 18s bands respectively at ~5 kbp and ~2 kbp using agarose gel electrophoresis, and the UV absorbance ratio (both for 260/280 and 260/230) must be about 1.8-2 using UV microspectrophotometers (e.g., Nanodrop™).

3.2 Labeling and hybridization

For labeling of samples, direct and indirect labeling of cDNA are widely used methods. In both approaches, the total RNA or mRNA is converted into cDNA by means of RT reaction (using either oligo dT priming to the 3' polyadenylation site, or random hexamer/nonamer). The main difference between these two approaches is direct or indirect incorporation of fluorescent dyes (e.g., cyanine dyes, CyDyes™ (Cy3 or Cy5), Alexa™ dyes) into the cDNA, i.e. direct incorporation of dye into the cDNA by RT reaction (direct labeling method), and indirect incorporation of dye through amino-allyl dUTP into the cDNA (indirect labeling method). Although the direct incorporation method is simple and rapid, it seems there is a bias of incorporation in this method so that the cDNA is labeled at higher efficiency by Cy3 than by Cy5. To resolve this issue, dye swap experiments in treated and untreated samples can be performed, in which in the replicate experiment the control sample is labeled with Cy3 instead of Cy5 and vice-versa for test sample; and comparing the results and analysis of combined outcome. The indirect labeling method display less biased dye incorporation, however its main disadvantage is that the protocols are more intricate and time consuming. In our research center (<http://nano.tbzmed.ac.ir/>), we follow the protocol of the supplier "Ocimum Biosolutions". Briefly, after sticking the gas-tight sealing system (Gene Frame™), the hybridization buffer is preheated at 42°C for 10 min. The labeled cDNA is dissolved in an appropriate amount of hybridization buffer. The hybridization mixture is heated at 95°C for 3 min and subsequently cooled on ice for maximum 3 min. The appropriate volume of probe/hybridization solution is pipetted at one end of the frame and the polyester cover slip is carefully placed over the frame. The arrays are incubated at 42°C on a shaker in dark for 16-24 h. For washing, the following washing buffers (WB) are used: WB1 (2x saline-sodium citrate (SSC), 0.1 % sodium dodecyl sulfate (SDS)), WB2 (1 x SSC), WB3 (0.5x SSC). All washing steps are performed at room temperature for 3-5 min each step. All buffers should be filter sterilized. To dry the slide arrays, prior to scanning with a scanner, they are briefly gently centrifuged.

4. Data mining

A key point for translation of DNA microarray data into clinical application is data mining which is deemed to be the most confounding part. Data Mining is all about automating the process of searching for patterns in the data - it is an iterative process of discovery. As shown in Fig. 4, implementation of image processing and bioinformatics methodologies appears to be crucial for obtaining a sensible outcome from microarray data.

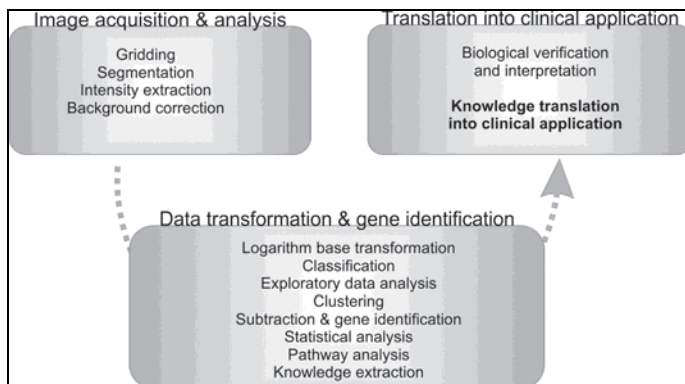


Fig. 4. Translation of DNA microarray data into clinical applications; adapted with permission from (Saei & Omidi, 2011).

Exploratory data analysis on microarray can help discover new patterns and networks in gene expression. The patterns that data mining discovers can have various forms including: trends in data over time, clusters of data defined by important combinations of variables, and finally evolution of these clusters over time. Patterns could be derived from analyzing the change in expression of the genes, and new insights could be gained into the underlying biology. Using the data retrieved from the microarray we can determine gene function in the cell, identify targets for new drug designation and reveal molecular networks and pathways

4.1 Image acquisition and primary analysis

Although the correct analysis and interpretation of microarray data is highly dependent upon image acquisition and data production, surprisingly less attention is usually given to such crucial steps which include: 1) scanning of the image, 2) spot recognition, 3) gridding 4) segmentation, and 5) intensity extraction.

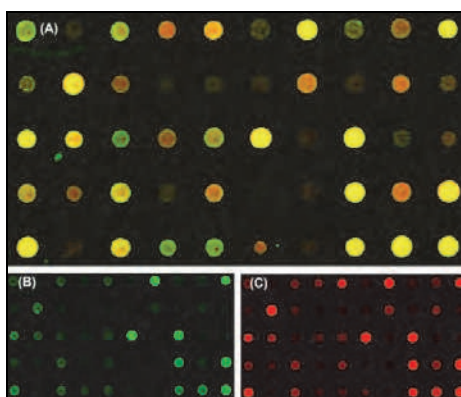


Fig. 5. Typical fluorescent images of hybridized cDNA microarray. A) Superimposed fluorescent image of Cy3-cDNA and Cy5-cDNA hybridization. B) Fluorescent images of Cy3 channel. C) Fluorescent images of Cy5 channel. Data are adapted with permission from (Barar et al., 2009).

Fig. 5 represents typical fluorescent images of hybridized cDNA microarray for superimposed image of Cy3- and Cy5-labeled cDNA as well as Cy3-labeled hybridized cDNA of untreated control cells and Cy5-labeled hybridized cDNA of treated cells.

For identification of each spot or gene on the slide, a grid has to be applied to the spots on the slide using appropriate softwares (e.g., we generally use TECAN scanner with ArrayPro™ software). However, a manual proofreading needs to be performed to ensure that all the spots on the array fall on the grid correctly (Lonardi & Luo, 2004). Once the gridding fulfilled, segmentation can be performed to separate the spots from the background and define the shape of each spot – there are always different shapes that do not have a constant diameter. Inside the circle is the signal we seek and outside is the background. Many methods are included in softwares for segmentation such as fixed circle segmentation (FCS), adaptive circle segmentation (ACS), adaptive shape segmentation (ASS) and histogram segmentation (HS). The selection of the best method largely depends on the quality of the produced images and personal interest and experience (Ahmed et al., 2004). As the name indicates, the FCS treats all spots equally. When almost all of the spots on the microarray are circular, the ACS has been proved to be the best approach, whereas ASS finds the shape which best fits the spot when the spots are not circular. The HS method considers pixel intensities in and around each spot to decide whether that specific pixel belongs to that spot (Ahmed et al., 2004; Bengtsson & Bengtsson, 2006; Lehmußola et al., 2006; Qin et al., 2005). After separation of the signal from the background, it comes to intensity extraction, after which the intensity ratios are calculated. To minimize the systematic errors, poor quality spots should be removed – that is any spot with intensity lower than the background plus two standard deviations should be excluded.

4.2 Normalization and transformation

Normalization and transformation of data is the first step after primary analysis and includes background subtraction, normalization, ratio calculation and log transformation (Geller et al., 2003; Quackenbush, 2002). Spotted arrays are routinely used for simultaneous analysis of gene expression in untreated control and treated cells. The mRNA extracted from untreated control is labeled with cy3, whereas mRNA obtained from treated cells is labeled with cy5. To minimize the impact of these two dyes and for verification of results, dye flipping appears to be a helpful strategy.

Cy3 and Cy5 data will be normalized individually and then the expression ratio (as logarithmic scale) will be calculated. The advantage of reporting \log_2 instead of the intact expression ratio is that in this way all the ratios fall between -1 and +1. Genes with a \log_2 of +1, have an upregulation factor of 2 and genes with a \log_2 of -1 have a downregulation factor of 2. Genes which possess constant expressions in the two samples will have a \log_2 of 0. Normalization is the most important process in transformation and compensates for the experimental variability of the data and assumes that all of the genes on an array or a subset of them have an average expression ratio of 1. In the scatter plot of cy5 versus cy3 intensities, most of the genes usually form a cluster along a straight line. Cy5 intensity is usually less than that of the Cy3. This is normally seen in the scatter plots with deviations from the one to one ratio. The slope of the line (Cy5 vs. Cy3) in the scatter plot should be one in the ideal case where there are no inconsistencies and/or all the variables are same for the two samples, but this is not usually the case. Normalization is, in fact, the calculation of the slope which best fits the line. This is usually performed using regression techniques which provide the slope value and helps removal of the unwanted effects, at the cost of losing

some information such as the absolute expression values. The most often applied methods for the normalization of the dye bias include: 1) local mean, 2) global mean, 3) LOcally WEighed Scatter-plot Smoothing (LOWESS), and 4) bayesian analyses. We have effectively used the LOWESS normalization method. Fig. 6 represents a typical small array showing LOWESS based normalization of data as scatter plot.

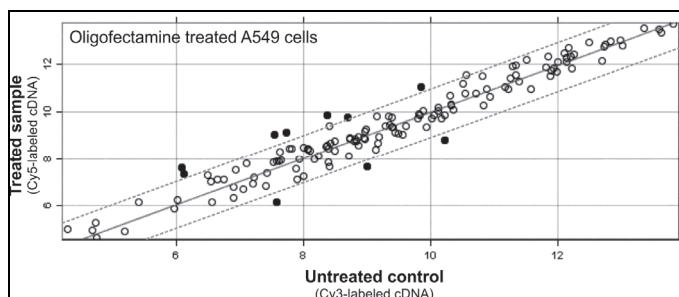


Fig. 6. A typical scatter plot for small array representing normalized data using LOWESS method. Data represent gene expression changes induced by OligofectamineTM nanoliposome used for gene (antisense) delivery in A549 cells. White and Black circles respectively show unchanged and up-/ down- regulated genes (with at least 2-fold changes). Data are adapted with permission from (Barar et al., 2009).

Now the question is how can we select a specific normalization method which best fits the structure of our data? In fact, the most simple normalization method is using housekeeping genes as reference points in the two chips and multiplying all the intensities by a constant until the expression of the controls get equal in the arrays under comparison (Quackenbush, 2001). The global normalization (total intensity normalization) assumes that the total amount of mRNA in a cell is fixed and the intensities of one chip are multiplied by a constant until the sum of the intensities of the genes in one chip equals the sum of the intensities in the other. This method is still a simple approach and does not remove the signal-dependent bias (Quackenbush, 2001). Regression techniques can be used instead of aforementioned methods for calculating the best-fit slope of the line (Kepler et al., 2002). If regression techniques for normalization are applied locally, it will be known as LOWESS which can be used to heal the nonlinearity in the data. LOWESS appears to be a powerful method in the normalization of the intensity-dependent dye bias and has shown to be good enough in many studies (Berger et al., 2004; Quackenbush, 2001). Once data were normalized, the expression ratio is calculated and then the \log_2 transformation is applied. To ensure about the quality of data produced, the expression of repeated genes on array as internal controls and the coefficient variation (CV) of normalized untreated controls of various experiments as external control can be checked.

4.3 Selection of differentially expressed genes and significance examination

The prime aim of a microarray analysis is to clarify the genes which have been differentially up or down regulated. The traditional method which was used for identifying the differentially expressed genes was setting a fixed cutoff threshold (usually in 2-fold) to infer significance. Genes with intensity ratio above or below the fixed ratio were said to be down or up regulated (Baggerly et al., 2001). This method is called fold change method (Draghici,

2002), but it is now believed to be inefficient due to the lack of a statistical basis. For example, if the condition under study changes the gene expression profile slightly, then this method will give no differentially regulated genes and as a result the sensitivity will be almost zero. In contrast to the former example, if the affecting parameter level is too high to produce the optimal detectable effect and the threshold is set at two, many genes will be selected as differentially regulated, the method will have low specificity and many genes will pass our filter as false positives (Cui & Churchill, 2003; Gusnanto et al., 2007). Another disadvantage is that, in microarray technology genes with little expression levels have a low signal to noise ratio, which can be demonstrated by the funnel shape at the bottom of the expression distribution in the scatter plot. It should be evoked that losing some data will be inevitable.

Another method for selecting differentially expressed genes is the unusual ratio which involves the selection of the genes for which the ratio of the experiment and control value is at certain distance that is usually taken ± 2 standard deviation. In fact, the selected genes should have an "experiment: control" ratio of at least two standard deviations away from the related mean value. Basically, this procedure is accomplished by applying a z transformation to the log ratio values (Draghici, 2002).

4.4 Dimension reduction

A major problem in microarray analysis is the large number of dimensions. In gene expression experiments each gene and each experiment may represent one dimension. Visual analysis of the microarray data in the original format does not yield much because if a slide array (spotted with 1000 genes) is used to examine a particular disease in 10 patients, the resultant data will be of high dimensionality (which would be a matrix of 10 by 1000). To make the most of the data, feature reduction or dimension reduction is used which renders this possible by finding the minimum number of genes that can best describe data.

Dimensionality reduction algorithms can be classified into "feature selection" and "feature extraction". Feature selection is to select k dimensions, out of the original d dimensions, that can best represent the original data set. Feature extraction is to find a new set of k dimensions that are some combination of the original d dimensions. The most popular feature extraction algorithms may be the linear projection method such as PCA for unsupervised learning and linear discriminant analysis (LDA) for supervised learning. The number of dimensions after the methods are applied will be two as most analyses are performed in two dimensions x and y . There is a possibility of losing some weak but important data in this way as they cannot compete with the impact of more important features. The methods used in this section of analysis are PCA, correspondence analysis (CA), multi-dimensional scaling (MDS), and cluster analysis (Dugas et al., 2004; Nguyen, 2005; Shannon et al., 2003).

Of the dimension reduction methods, PCA is now mostly used method as a tool in exploratory data analysis and for making predictive models. PCA involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. PCA ignores the dimensions in which data do not vary much. PCA is closely related to factor analysis; indeed, some statistical packages deliberately conflate the two techniques.

4.5 Grouping methods

After performing multiple experiments under different conditions (e.g., number of patients or various time points), genes that reveal similar expression intensities can be grouped under a subset of conditions. Likewise based on the pattern of the distinguishing genes, boundaries between different subtypes (e.g., for cancer) can be set (Belacel et al., 2006). Grouping methods as “unsupervised method” or “supervised method” (the so called clustering and classification methods, respectively) can help identify co-expressed genes (Juan & Huang, 2007).

4.5.1 Unsupervised grouping: clustering

Unlike supervised methods, no data is provided in unsupervised methods that include hierarchical clustering, k means clustering and self-organizing maps (Au et al., 2004; Steinley, 2006). The efficiency of all these methods relies on whether high quality data has been produced.

Most clustering algorithms utilize a sort of distance metric (the distance between two expression vectors), based on which they group similar profiles together. This means that genes closer to each other in space are grouped together. They use either statistical correlation coefficient (from -1 to +1) or the Euclidean distance (the square root of the sum of the squared differences between corresponding feature values). The common measures of dissimilarity are “Euclidean distance” and “Manhattan distance” (Leung & Cavalieri, 2003; Quackenbush, 2001). Fig. 7 represents a schematic illustration for Euclidean distance clustering.

The Euclidean distance (Ed) and Manhattan distance (Md) are given by equations (1) and (2), respectively:

$$d_{Ed} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

$$d_{Md} = \sum_{i=1}^n |x_i - y_i| \quad (2)$$

Where x_i and y_i are the measured expression values, respectively, for genes x and y in the experiment i , and n is the number of the experiments under the analysis.

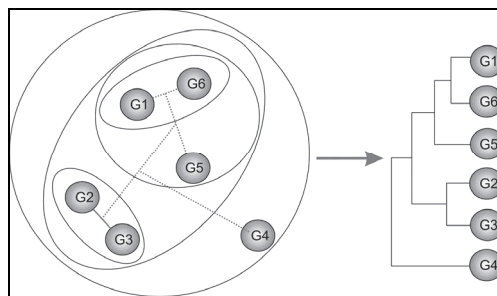


Fig. 7. Schematic illustration of Euclidean distance clustering of expressed genes (G); adapted from (Saei & Omid, 2011).

The best way is to calculate the distance between a gene and the centroid of the cluster, however sometimes the distance between the gene and the nearest neighbor is calculated. Hierarchical clustering is a tree-like representation of data, which can be applied in two manners, "bottom-up" or "top-down" methods (Chipman & Tibshirani, 2006). The bottom-up method starts with each object representing one cluster of size 1 and at each following step, the two closest clusters are joined until all objects are in a single cluster of size n . The top-down method is the reverse phase of what was discussed and is usually of more interest when we aim to identify a few clusters (Chipman & Tibshirani, 2006). One major problem using the hierarchical clustering is that the number of the genes which undergo clustering should not exceed several thousands, because the process may fail or become time consuming since the algorithm calculates the distance between all genes. To resolve such problem, one of the solutions is to exclude the unchanged genes, or alternatively applying other methods such as K-means clustering which is faster than hierarchical clustering (Quackenbush, 2001). In fact, the K-means clustering uses the prior information about the expected number of the clusters (k) which can be obtained from principle component analysis. Each gene is randomly assigned to one of the clusters and then the distance between each gene with the centroid (average expression vector) of the clusters is calculated. If one gene is closer to the centroid of another cluster, it is moved to that cluster. The centroids are recalculated after the assignment of all the genes to their nearest centroids in an iterative process. This process continues until there is no change in the members of the clusters and the algorithm stops here (Steinley, 2006; Wilkin & Huang, 2008). Basically, the hierarchal clustering represents the relation of the genes to each other even though this is not the case with K-means clustering.

Self-organizing maps (SOMs) have been embedded with artificial neural networks (ANNs). And, based on the amount of the similarity of the expression vectors to reference, it is possible to clarify which gene belongs to which cluster. In contrast to other clustering algorithms, SOMs provide maps with the most similar clusters along each other's sides which make it a little bit different from K-means clustering (Covell et al., 2003).

One deficiency of the discussed algorithms is that a gene can only belong to just one cluster in the dataset, while a gene may be member of more than one cluster due to its different functions in the cell. An example to this can be those proteins that are the major players of complicated networks in the cell, e.g. second messengers of the cell transduction pathways or epidermal growth factor receptor (EGFR) which is known to play various roles in different cells. A fuzzy logic can be alternatively exploited to solve such problem, because it pursues clustering with overlapping. In fuzzy clustering, a gene may belong to different clusters with specific membership degrees between zero and one specified to each cluster. The sum of the membership degrees of each gene in fuzzy clustering is one (Dembele & Kastner, 2003).

As discussed above different clustering methods have different efficiencies and different algorithms and thus are likely to produce different results. Even two hierarchical clustering algorithms in two softwares are slightly different and do not completely match, thus the results vary from algorithm to algorithm and from software to software. This means one gene may be placed in different groups or clusters when applying different algorithms. Therefore, upon our experiences, it is recommended to apply various methods and different parameters to the data analysis and then, based on the result decide what to do. Function prediction is another technique which can be applied after cluster analysis (Joshi et al., 2004). Genes that behave similarly in response to different conditions are likely to have

similar functions. Thus, the function of orphan genes (genes with unknown function) can be inferred from the known function of other genes placed in the same cluster.

4.5.2 Supervised grouping: classification, class prediction using the input data

In supervised methods the network is fed with some data for which the gene families are known. Thus based on the previous data provided for the network, it predicts that a new feature -that has not been used in the training step- belongs to a specific family (Peng, 2006; Wang et al., 2005). This is called "machine learning". The classifier is trained using the expression of all genes (which have passed the t-test or ANOVA for determination of the significance) as the input and then selects the features (genes) that contribute the most to your condition, and then classifier is validated on a set of sample that was not used in feature selection in the later step. For example, if you have two cancer subtypes you can easily classify the two subtypes and select the genes that are differentially expressed in one subtype and not in the other or using known genes you can classify the cancer subtypes. For comparison of more samples you need to find genes that are expressed differentially just in one subtype. To make a general model, one needs to apply cross-validation method (i.e., to split data into different training and prediction sets and to try the classifier) or leave-one-out method (i.e., crossing out one of the features and screening the effect of deletion on the dataset) to the data.

Supervised methods are almost good at pattern recognition and dividing the samples into groups, but so far have not shown to be effective enough in the new samples, i.e. when it comes to prediction they have lower efficiency. The supervised methods include neural networks, logistic regression and linear discriminant analysis (LDA) (Liao & Chin, 2007; Linder et al., 2007; Shen & Tan, 2005; Tai & Pan, 2007). Nearest neighbor classifier is one of the simplest classifiers (Shen & Hasegawa, 2008). If one has already performed a principal component analysis or clustering, then it will be possible to determine the most probable number of classes (as in K-means) in the data and the algorithm will work well.

For higher number of example it is better to use more advanced methods such as ANN which use the brain neurons' logic. In use of the classification methods, however, overfitting or overtraining of the classifier algorithm is a major problem, which means if the classifier is provided with too many parameters a false model may perfectly fit the data (Babyak, 2004; Hawkins, 2004). When the degrees of freedom in parameter selection are more than the information content, the ability of the model is diminished or destroyed. In machine learning, the algorithm is trained using some part of the data for which we know what the output will be. After training, the model should be able to predict the output for the other part of data for which the output is unknown. When overfitting happens, the performance of the model on training examples will increase, but the prediction power will be lower. Contrary to this, if the network is not complex enough it will fail to detect the signal. This is called underfitting. But a too complex network will fit even the noise, leading to overfitting which can produce excessive variance in the outputs and instead of capturing the desired pattern or trend in the data, it will memorize the training data (Babyak, 2004; Hawkins, 2004). The best way to chuck out overfitting is to use lots of training data. We should have at least 10 observations per variable. If one is looking at 30 variables, the model should be at least provided with 300 events. To avoid overfitting, it is also possible to use additional techniques such as cross-validation, weight decay, bayesian learning, early stopping and also model comparison (Babyak, 2004; Braga-Neto & Dougherty, 2004; Hawkins, 2004).

4.6 Pathway analysis

Pathway analysis should be towards functional enrichment to establish networks between genes and proteins. This may provide better understanding about the dynamics of gene expression that may grant a platform for reverse engineering of regulatory networks. Understanding the expression dynamics of gene networks helps us infer innate complexities and phenomenological networks among genes. Defining the true place of the genes in cell networks is indeed the foundation of 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 (Guo et al., 2006; Li et al., 2007; Mircean et al., 2004; Park et al., 2007).

When the picture of gene interactions are established, the nodes with many connections in the network are thought to be of crucial role in the cell function and can be considered as probable specific targets for drug delivery or even design of new drugs (e.g., in the case with EGFR or other Tyrosine Kinase genes). These pathways can tell us where the drugs act and also where our carriers induce toxicity (Dewey, 2002). They also help us predict the function of unknown genes and investigate their true place in the pathways, where gene expression or metabolic changes have been produced due to mutations or changed environmental conditions. Using these techniques, the mechanisms underlying diseases can be discovered.

One of the methods used in pathway analysis is comparing the gene list to a pathway which gives a p value as a result. The scoring enrichment methods compare a list of the genes to that of a pathway and count the hits; the greater the number of the hits, the greater the score and the enrichment (Curtis et al., 2005). Other hit-counting methods used by MAPP Finder and Odds ratio are hit-counting methods which indicate the degree of enrichment (Doniger et al., 2003). GenMAPP, an open source package that allows users to visualize microarray and proteomics data in the context of biological pathways (freely available at <http://www.genmapp.org/>), represents biological pathways in a special file format called 'MAPPs' which are independent of the gene expression data. It is used to group genes by any organizing principle (e.g., apoptosis pathways). The gene set enrichment analysis (GSEA) is a novel method, which uses all the data on the microarray in the order of expression, determining whether an a priori defined set of genes shows statistically significant, concordant differences between two biological states (e.g. phenotypes) (Subramanian et al., 2005). This algorithm compares each gene, in turn, to the pathway or class, and enrichment score is specified (Subramanian et al., 2007). The GSEA and Z score analysis provide p values which are easy to interpret, but it is not clear whether these values are biologically relevant. As for other methods described above, using different criteria Z scores can yield different results, because they are sensitive to gene list and pathway length (Curtis et al., 2005).

In the gene ontology project (<http://www.geneontology.org>), the genes have been classified into a hierarchy, based on the similarities in functions. The Kyoto Encyclopedia of Genes and Genomes (KEGG) (<http://www.genome.kp/kegg/>) provides searchable pathways (Aoki & Kanehisa, 2005; Kanehisa, 2002). However, the major problem encountered is that there are many genes with unknown functions or pathways, thus attempt should be done in discovering more pathways and connecting them with existing ones. With the future data incorporated into the pathway analysis, there will sure be improvements in the information produced by the microarrays. One best approach for establishing relationships in the data is to focus mainly on the genes on which a large amount of data is available.

Reverse engineering of regulatory networks can be fulfilled using two methods, one of which is “time-series data” and second one is “steady-state data” of gene knockouts. In the first approach, the magnitude of expression of a certain gene at a certain time point is a function of expression amount of the other genes at all previous time points. In the second 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 is assessed. For example, if by deleting gene A, gene B is upregulated, then one can relatively think that gene A, either directly or indirectly, suppresses the expression of gene B. However, these methods and in particular reverse engineering still lack full applicability, because there is a need for more discovered sophisticated networks in the cells in order to identify the hidden role of different molecules in the circuitry of gene regulation (Curtis et al., 2005; Martin et al., 2007).

5. Microarray gene expression profiling blunders

Exploratory data analysis on microarray can help discover new patterns and networks in gene expression. Using the data retrieved from the microarray one can determine gene function in the cell, identify targets for new drug designation and reveal molecular networks and pathways (Lee et al., 2008).

In recent years, many methods have been devised which have now found critical applications in the analysis of the microarray data. The methods used to explore deep into the microarray data include the PCA, the singular value decomposition (SVD), the SOMs and the support vector machine (Nahar et al., 2007; Pandey et al., 2007; Wall et al., 2001; Xiao et al., 2003; Zhu et al., 2009). Some of these methods such as PCA and clustering methods are used for data reduction as the vast nature of the microarray data does not allow the analysis of the genes one by one and holism is also not possible (Leung & Cavalieri, 2003; Quackenbush, 2001). Most applicable methods are unsupervised, and do not require additional prior information to feed the algorithm. As our understanding and knowledge of the genes and their expression in different cells and tissues is increased, incorporation of the known data will empower the microarray data analysis which can help show some real potential of data. As the scientific demands grow, we will need more sophisticated bioinformatics and biostatistics tools to fulfill the different needs in the analysis of the microarray data. Distinguishing the truly differentially regulated genes from noises or other genes is a real challenge in the first steps, because the filtered data is going to be used as a basis for the next steps in data mining. Unfortunately there are more genes than conditions and this makes the analysis of the data difficult. To render the microarray data more reliable, one needs to use more stringent data cutoffs such as the Bonferonni correction (not the two-fold cutoff) as the selection measure of the differentially regulated genes and reducing the number of the false positives in the data set (Bland & Altman, 1995). It should be highlighted that obtaining microarray data appears to be easy, however the analysis of the data is usually tedious and time consuming (Loring, 2006), thus there should be special focus on the analysis of the microarray data. However, because of our poor understanding of cell pathways and molecular biology and also not incorporating the gene information when analyzing the data, we are not able to dig deep into the jam of data produced by microarrays. Another problem is that some sections of the analysis deal with genes one by one considering them as individuals, whereas others look at the data in a holistic way. After finding a responsible gene for a specific trait or function in the cell using microarray

technology, other methods (e.g., RT-PCR, gene silencing, northern blotting) can be applied on the specific gene to assure that all the steps in the microarray analysis have been performed correctly. The reverse is also true; i.e. microarray can be used to verify the other techniques (Hembruff et al., 2005).

If microarray technology is to be employed in genomic impacts investigations of the gene delivery systems, then it should be noticed that microarrays are prone to many errors both in the experimental handling and data analysis steps. In fact, it should be perceived that all we have is the mere fluorescence intensity emitted from the two dyes and the ratio of the two intensities, steps which are vulnerable to many blunders. In different steps in a microarray analysis, multiple errors can be produced. These include the efficiency of RNA isolation, amount of mRNA used, cDNA generation, amplification, selective incorporation of cyanine dyes (i.e., Cy₃ or Cy₅), the quality of the array, hybridization and washing condition, spatial bias on the array during hybridization, scanning procedures, random errors and other variations in the conditions (Pounds, 2006). Because of these variations and errors, there exists a need for replicate experiments for normalizing the data, i.e., the more replicates in experiment, the less variance in the data. Taking the log of ratio of the intensities can eliminate the possible non-linearity in the data. Instead of using the 2-fold increase/decrease in the ratio intensities for identifying differentially regulated genes, some statistically based techniques (e.g., ANOVA and maximum likelihood analysis) can be applied (Bakewell & Wit, 2005; Churchill, 2004). It should be evoked that the result of a microarray experiment is largely dependent on statistical methods used and the results from two separate analyses do not usually concord because data analysis is a multiple step process, with each step being vulnerable to personal misunderstanding and bad interpretations and also because many combinations of the analytic methods are possible (Loring, 2006). One factor that contributes to the deterioration of the matter is that different laboratories often use different defined approaches to their analyses. Analyzing the data using different procedures makes the final results different. Each of these processes can be carried out using various methods, for example one can use hierarchical clustering, K-means clustering or self-organizing maps in the clustering step. As another example different linkage methods are available for hierarchical clustering that the analyst is free to choose from, each of them producing slightly different outcomes.

Recently some attempts have been performed to bring these divergent processes into one homogenous procedure. For overcoming the problem, the Microarray Gene Expression Data (MGED) society (<http://www.mged.org>) proposed standards termed as "Minimum Information About a Microarray Experiment (MIAME) standards" for homogenization of the microarray experiments and data handling (Whetzel et al., 2006). Many journals today require submissions to be in compliance with MIAME, upon which many experiments are now in concordance with MIAME in experimental design, implementation and data analysis. This may provide researchers a platform to share their data, to cooperate for better data analysis and comparison of experiments. The six most critical elements in MIAME are: 1) the raw data for each hybridization, 2) the final normalized data for the set of hybridizations, 3) the essential sample annotation including experimental factors and their values (e.g., compound and dose in a dose response experiment), 4) experimental design, 5) sufficient annotation of the array (e.g., gene identifiers, genomic coordinates, probe sequences or reference commercial array catalog number), and 6) the essential laboratory and data processing protocols; reader is directed to see (<http://www.mged.org/Workgroups/MIAME/miame.html>) for a review of these standards. Unfortunately, the

variations in two separate experiments can also be due to the true variation in the samples. For optimal minimization of the errors and avoidance of irreproducible measurements in the microarray analyses, it is often recommended to conduct the experiment in three separate replicates. By further increasing the number of the repeats, the confidence level increases even though it is not usually practical due to economic issues. Basically, the order in which the data analysis methods are applied appears to be: 1) image analysis, 2) normalization, 3) clustering, 4) pathway analysis, 5) promoter analysis, 6) function prediction, and 7) classification (Allison et al., 2006; De, V et al., 2007; Olson, 2006).

6. Gene therapy trials and microarray technology

Gene therapy appears to be one of the most challenging fields even though gene therapy researches have been tarnished with the death of Jesse Gelsinger in 1999 (the first publicly identified person to die in a gene therapy trial). However, many positive phase II/III trials seem to accelerate these researches. Table 1 represents the latest state of the gene therapy clinical trials worldwide.

Phase	Number and percentage of clinical trials (1989-2010)
Phase I	995 (60.5%)
Phase I/II	308 (18.7%)
Phase II	267 (16.2%)
Phase II/III	13 (0.8%)
Phase III	57 (3.5%)
Phase IV	2 (0.1%)
Single subject	2 (0.1%)
Total	1644

Table 1. Gene therapy clinical trials worldwide, data are from (Edelstein, 2011).

For example, Contusugene ladenovec (Advexin™, INGN-201; Introgen Therapeutics Inc.) has been announced as a first-generation gene therapy for cancer, targeting the cancer suppressor gene p53. Based upon the phase III trial results showing fairly good responses of certain patients to Advexin™, it has been accepted to be reviewed by the European Medicines Agency. Having observed additive/synergistic effects in a variety of tumor types (non-small-cell lung carcinomas, squamous cell carcinoma of the head and neck, hepatocellular carcinoma, glioma, breast, prostate and colorectal cancers), researchers in the field of gene therapy hope that these advancements will eventually help restore gene-based treatment modalities alone and in combination with chemotherapy and radiation (Senzer & Nemunaitis, 2009). This is just a beginning of a “molecular big-bang”, a domain of science which requires mechanistic investigations on early and late influences of gene-based modalities for their successful translation into clinical practice.

Of these gene therapy clinical trials, as shown in Table 1, about 4% have reached the final stage and it is expected to see several gene based medicines in clinic soon. About 64.5% and 8.7% of them respectively belong to cancer and cardiovascular gene therapies, which highlight importance of these diseases (Edelstein, 2011). Table 2 represents the latest state of the gene delivery vectors used for gene therapy clinical trials worldwide.

Gene delivery vector	Number and percentage of clinical trials (1989-2010)
Adenovirus	400 (23.8%)
Retrovirus	344 (20.5%)
Naked/Plasmid DNA	304 (17.7%)
Vaccinia virus	133 (7.9%)
Lipofection	109 (6.5%)
Poxvirus	93 (5.5%)
Adeno-associated virus	75 (4.5%)
Herpes simplex virus	56 (3.3%)
Lentivirus	29 (1.7%)
Other categories	82 (4.9%)
Unknown	55 (3.3%)

Table 2. Gene delivery vectors used for gene therapy clinical trials worldwide, data are from (Edelstein, 2011).

So far, DNA microarray technology has been employed for detection of unknown gene-based biomarkers in various diseases, and also for clinical diagnosis. For example, in 2010 Qin and Tian performed a microarray gene analysis of peripheral whole blood in normal adult male rats after long-term growth hormone (GH) gene therapy. They reported that 61 genes were found to be differentially ($p < 0.05$) expressed 24 weeks after receiving GH gene therapy. These genes were mainly associated with angiogenesis, oncogenesis, apoptosis, immune networks, signaling pathways, general metabolism, type I diabetes mellitus, carbon fixation, cell adhesion molecules, and cytokine-cytokine receptor interaction. The results imply that exogenous GH gene expression in normal subjects is likely to induce cellular changes in the metabolism, signal pathways and immunity. Based on such screening, these researchers claimed eight genes as candidate biomarkers (Qin & Tian, 2010).

Since genomedicines (e.g., antisense oligo deoxy nucleotides (As-ODN), RNA interference) can selectively block disease-causing genes, the responsible genes of diseases (e.g., cancer genes) have been chosen as potential targets even though undesired nonspecific side effects may tarnish the actual mechanism of gene therapy, blemishing clinical development of gene based therapeutics. Using DNA microarray technology, Cho et al. have conducted a systematic characterization of gene expression in cells exposed to antisense and showed that in a sequence-specific manner, antisense targeted to protein kinase A R1alpha alters expression of the clusters of coordinately expressed genes at a specific stage of cell growth, differentiation, and activation (Cho et al., 2001). These researchers showed that the genes defined the proliferation-transformation signature were down-regulated, however those genes defined the differentiation-reverse transformation signature were up-regulated in antisense-treated cancer cells and tumors. And based on such findings, they concluded that the defining As-ODNs on the basis of their effects on global gene expression may lead to identification of clinically relevant antisense therapeutics and can identify which molecular and cellular events might be important in complex biological processes.

A brief search of the relevant database of scientific literatures such as Medline, Toxnet and BIDS clearly reveals existence of a fairly large number of investigations upon gene delivery

and/or targeting, while little works have been undertaken on cytogenomic impacts of such gene delivery vectors. It is still not totally clear what genomic impacts can be elicited by a particular genomedicine or even by its viral/non-viral carrier(s). Although, the gene/drug delivery systems are generally announced as inert materials, it seems that in many occasions they are not compatible in a genomic level. Thus, we may even face with new terminologies such as “functional excipients” which can inevitably impose non-specific intrinsic cytogenomic changes in target cells/tissues (Omidi et al., 2005a).

7. Genomic impacts of gene delivery systems

To date, the traditional chemotherapies have become an indispensable part of the cancer treatment procedures which almost always requires that the patient take high doses of multiple drugs. Thus, to avoid the inevitable occurrence of side effects, there is an emergent need for smart targeted delivery systems that provide lower toxicity and higher specificity - that will also be able to reduce the required dose of the drugs. Since gene therapy (using As-ODN, siRNA, aptamer, ribozyme, etc.) can target disease at genomic level, this approach has attained great attention. To deliver nucleic acids to target cells, gene carriers such as viral and nonviral vectors (Table 2) have been widely used to overcome systemic and subcellular barriers. In fact, such carriers like drug delivery systems can largely influence the pharmacokinetic properties (i.e., absorption, distribution, metabolism, and elimination) of drugs which are incorporated with designated carriers. Thus, there has been an increasing interest in design of delivery vehicles which are capable of delivering drugs (of any origin) inside cancer cells that is now an issue in design of molecular Trojan horses (Kim et al., 2008; Portney & Ozkan, 2006; Sawant et al., 2006).

It has been revealed that most gene delivery systems can nonspecifically induce alterations in the gene expression profile of the target cells (Omidi et al., 2005a); an impact which is either parallel with the therapeutic purpose or contrary to it or even null. From another viewpoint the individual effect of the carriers can be cut off into two categories (i.e., early and late impacts). For example, we have observed that most of the cationic gene delivery systems (e.g., linear and branched polyethylenimine (PEI), cationic lipids and dendrimers) are able to induce gene expression changes intrinsically (Hollins et al., 2007; Omidi et al., 2005b; Omidi, 2008). The determination of the early effects of a gene delivery system appears to be quite easy; however investigation on the late effects in a course of time is a cost demanding and time consuming trial. If a genomedicine is to enter clinic, there has to be some investigations to prove its safety. Time-series analysis can be a predictable section of these investigations. Up to now, many attempts have been made to design some gene delivery systems (viral and non-viral vectors). Some of these vectors include monoclonal antibodies, liposomes, metabolites, peptide hormones, cytokines, growth factors, viral and bacteriophage particles, nanoparticles and dendrimers. The choice of the carrier depends on the drug and also the cell to which the drug is going to be delivered. Almost all these carriers have suffered from either poor uptake and little transgene expression (non-viral vectors), or immunogenicity (viral vectors). With little efficacy of the carrier systems, some of them are still far from clinic and the gene therapy field needs a lot of time to reach the magic of genomedicines for treatment of diseases.

Viral vectors have a higher efficiency but their successful applications need more investigations. Expression of the recombinant viruses in the target cells is transient and can produce an inflammatory response in the patient. Another disadvantage of these vectors is

that it is not known whether the re-administration is safe due to the immunological responses. Unfortunately the first volunteer, Jesse Gelsinger who was undergone a gene therapy with adenoviral vectors died due to the activated viruses in the phase I gene therapy trial in 1999 (Ferber, 2001). By far, as shown in Table 2, adenoviruses and retroviruses (viral vectors) and liposomes (nonviral vectors) have been the most frequently used viral vectors for gene therapy clinical trials. Molecular Trojans were introduced into the field of drug targeting, when the need for more specific drug targeting was felt. For brain drug delivery using molecular Trojan horses (Pardridge, 2008; Patel et al., 2009), different technologies are used including fusion protein technology, avidin biotin technology and Trojan horse liposome technology. Recently, tat protein transduction domain (PTD) as cell penetrating peptides have been used as molecular Trojans to cross biological barriers (Dietz & Bahr, 2004).

In non-viral vectors, DNA condensation and packaging (using cationic lipids or polymers) can protect gene based therapies against degradation by nucleases. These systems can also be equipped with specific ligands which help the carrier find its way in the hostile environment of the body through specific receptor-ligand interactions (Duzgunes et al., 2003; Nie et al., 2006; Putnam et al., 2001).

Ideally a targeted therapy should be specific, but assuming the vast variety of the proteins and receptors on the target cell surface, a substance may inadvertently bind to the cell surface receptors nonspecifically. Thus, we may witness many other interactions other than the desired one upon implementation of a drug or a target therapy system. A glucocorticoid, for example, can bind to myriad of cellular receptors and accordingly may have many nonspecific interactions. Hence it is not far from logic if we think that many of its interactions are still unknown. This may guide us towards a new field of science "nonspecific genomic/proteomic signature of chemicals" (the so called genomic impacts) (Hollins et al., 2007). Accordingly, screening the pseudo-targets of a substance or carrier can be of great help in choosing the best carrier for a molecule, as nonspecific signaling pathways can be induced by carriers alone. In this regard, implementation of microarray for development and approval of gene based medicines can assist us in establishing the genetic fingerprints of such modalities. In fact, now there is a growing hope that global gene expression profiling will replace histopathology and other traditional techniques used in toxicology and will become the primary tool for safety evaluation of drugs and their carriers. Comparison of the gene expression profiles between a healthy and mutant cell can be used to reveal the site of the interaction of a desired drug. Therefore, the detailed characterization of the genomic signature for each delivery system -both the complex and the vector- has to be performed in different cells, since different cells respond to stimuli in different ways (Omidi et al., 2005b). This clearly means that our future challenge will be finding and establishing the genomic signature for each carrier in specific cell lines -the way genes feel the toxicants and react against them and finally define the safety margins for them. This challenge requires implementation of high throughput microarray based screening methodology from bench works to data mining.

8. Final remarks

With many different attempts to translate gene therapy investigation into clinic, researchers in this field are still feeling the optimism that they felt years ago. In fact, although the initial burst of excitement over gene therapy hit inevitable hurdles, it appears that still we must

work hard to learn more and more to provide improved treatments with minimal undesired side effects/impacts. For example, in cancer gene therapy, we have started learning which genes to choose (for suppression/stimulation), how to improve expression, and amplify their cancer-killing abilities by adding other approaches, such as cancer vaccines and chemotherapy. The entire music of life played by genes are yet to be fully understood, for which we need to get the practical and theoretical skills to comprehend the orchestral meaning of clusters of expressed genes in relation to a particular disease.

Now, we are just targeting a single or (at maximum) couple of biomarkers to fight a disease. This perception has to be improved through our understanding on entire picture of genomic expression of a biological function/disease in relation with so many others (genes/clusters of networked genes) that are directly/indirectly affecting the biological end point phenomena. Besides, a rapid, accurate, and reliable diagnostic method is necessary for identification of a disease for development of a suitable therapy, which consequently can reduce the mortality rate and also the cost of treatment. The recently developed molecular diagnostic assays (based on DNA-DNA or protein-DNA hybridization of clinical samples) appear to hopefully provide a robust platform, allowing effective diagnosis of different diseases with high speed, sensitivity, and specificity.

In the "functional genomics era", challenges on determination of proper analytical methods seem to be smoothly being shifted towards the post-analytical challenges, for which the microarray technology can provide a revolutionary analytical platform for concurrent analysis of thousands of genes in a single experiment. Such an approach confers enormous potential in the study of biological processes in health and disease. Based on recent FDA project on microarray (MAQC), this technology is going to become a potentially important tool in diagnostic applications and drug discovery too. In fact, the microarray based investigations have provided the vital impulse for biomedical experiments including: 1) identification of disease-causing genes in malignancies, 2) regulatory genes in the cell cycle mechanism, 3) investigation on genomic impacts of pharmacotherapies/gene therapies. The final point of these studies will guide us to translate our basic genomic knowledge into more sensible clinical practice, even for development of individual therapies, perhaps by identifying genes for new and unique potential drug targets and predicting drug responsiveness for individual patients. Given altogether, it seems we can use such technology for prevention strategies either to improve the health and life styles.

In the future, it is deemed that most of the investments on functional genomics will largely dependent upon research movement beyond the microarray based exploratory stages, so that new functional genomic pipeline demand sensible translation of a list of genes resulting from a microarray analysis. The real value and meaningful sense of gene expression changes need distinctive interpretation of a change by means of biological validation. Thus, the numerical verification of expression levels of thousands of genes need to be discovered, perhaps by finding biological relationships between the genes which require appropriate online systems to reveal new biological pathways in relation with clusters of the regulatory biomolecules. Unfortunately such an online tool has yet to be invented and till that comes into reality; the analysis of microarray data sets will be less rewarding. Other important issue seems to be the growing large size of microarray data and untapped information available in different data bases – the question is how we can extract the best results from existing data sets in relation with similar sets of data? Can we build an online network for such aim? It should be clarified that any set of microarray data has potential to be re-analyzed based upon many different integrative biological/clinical concepts.

9. References

- Ahmed, A.A.; Vias, M.; Iyer, N.G.; Caldas, C. & Brenton, J.D. (2004). Microarray segmentation methods significantly influence data precision. *Nucleic Acids Res.*, Vol.32, No.5, (March 2004), pp. e50, ISSN 0305-1048
- Allison, D.B.; Cui, X.; Page, G.P. & Sabripour, M. (2006). Microarray data analysis: from disarray to consolidation and consensus. *Nat.Rev.Genet.*, Vol.7, No.1, (January 2006), pp. 55-65, ISSN 1471-0056
- Aoki, K.F. & Kanehisa, M. (2005). Using the KEGG database resource. *Curr.Protoc.Bioinformatics.*, Vol.Chapter 1 (October 2005), pp. Unit, ISSN 1934-3396
- Au, N.H.; Cheang, M.; Huntsman, D.G.; Yorida, E.; Coldman, A.; Elliott, W.M.; Bebb, G.; Flint, J.; English, J.; Gilks, C.B. & Grimes, H.L. (2004). Evaluation of immunohistochemical markers in non-small cell lung cancer by unsupervised hierarchical clustering analysis: a tissue microarray study of 284 cases and 18 markers. *J.Pathol.*, Vol.204, No.1, (September 2004), pp. 101-109, ISSN 1096-9896
- Babiyak, M.A. (2004). What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom.Med.*, Vol.66, No.3, (May 2004), pp. 411-421, ISSN 1534-7796
- Baggerly, K.A.; Coombes, K.R.; Hess, K.R.; Stivers, D.N.; Abruzzo, L.V. & Zhang, W. (2001). Identifying differentially expressed genes in cDNA microarray experiments. *J.Comput.Biol.*, Vol.8, No.6, (2001), pp. 639-659, ISSN 1066-5277
- Bakewell, D.J. & Wit, E. (2005). Weighted analysis of microarray gene expression using maximum likelihood. *Bioinformatics.*, Vol.21, No.6, (March 2005), pp. 723-729, ISSN 1367-4803
- Barar, J.; Hamzeiy, H.; Mortazavi-Tabatabaei, S.A.; Hashemi-Aghdam, S.E. & Omid, Y. (2009). Genomic signature and toxicogenomics comparison of polycationic gene delivery nanosystems in human alveolar epithelial A549 cells. *Daru*, Vol.17, No.3, (2009), pp. 139-147, ISSN 1560-8115
- Bednar, M. (2000). DNA microarray technology and application. *Med.Sci.Monit.*, Vol.6, No.4, (July 2000), pp. 796-800, ISSN 1234-1010
- Belacel, N.; Wang, Q. & Cuperlovic-Culf, M. (2006). Clustering methods for microarray gene expression data. *OMICS.*, Vol.10, No.4, (2006), pp. 507-531, ISSN 1536-2310
- Bengtsson, A. & Bengtsson, H. (2006). Microarray image analysis: background estimation using quantile and morphological filters. *BMC.Bioinformatics.*, Vol.7 (2006), pp. 96, ISSN 1471-2105
- Berger, J.A.; Hautaniemi, S.; Jarvinen, A.K.; Edgren, H.; Mitra, S.K. & Astola, J. (2004). Optimized LOWESS normalization parameter selection for DNA microarray data. *BMC.Bioinformatics.*, Vol.5, No.1, (December 2004), pp. 194, ISSN 1471-2105
- Bland, J.M. & Altman, D.G. (1995). Multiple significance tests: the Bonferroni method. *BMJ*, Vol.310, No.6973, (January 1995), pp. 170, ISSN 0959-8138
- Braga-Neto, U.M. & Dougherty, E.R. (2004). Is cross-validation valid for small-sample microarray classification? *Bioinformatics.*, Vol.20, No.3, (February 2004), pp. 374-380, ISSN 1367-4803
- Cangelosi, R. & Goriely, A. (2007). Component retention in principal component analysis with application to cDNA microarray data. *Biol.Direct.*, Vol.2 (2007), pp. 2, ISSN 1745-6150

- Chen, H. & Li, J. (2007). Nanotechnology: moving from microarrays toward nanoarrays. *Methods Mol.Biol.*, Vol.381 (2007), pp. 411-436,
- Chipman, H. & Tibshirani, R. (2006). Hybrid hierarchical clustering with applications to microarray data. *Biostatistics.*, Vol.7, No.2, (April 2006), pp. 286-301, ISSN 1465-4644
- Cho, Y.S.; Kim, M.K.; Cheadle, C.; Neary, C.; Becker, K.G. & Cho-Chung, Y.S. (2001). Antisense DNAs as multisite genomic modulators identified by DNA microarray. *Proc.Natl.Acad.Sci.U.S.A.*, Vol.98, No.17, (August 2001), pp. 9819-9823, ISSN 0027-8424
- Churchill, G.A. (2004). Using ANOVA to analyze microarray data. *Biotechniques*, Vol.37, No.2, (August 2004), pp. 173-5, 177, ISSN 0736-6205
- Covell, D.G.; Wallqvist, A.; Rabow, A.A. & Thanki, N. (2003). Molecular classification of cancer: unsupervised self-organizing map analysis of gene expression microarray data. *Mol.Cancer Ther.*, Vol.2, No.3, (March 2003), pp. 317-332, ISSN 1535-7163
- Cui, X. & Churchill, G.A. (2003). Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol.*, Vol.4, No.4, (2003), pp. 210, ISSN 1465-6914
- Curtis, R.K.; Oresic, M. & Vidal-Puig, A. (2005). Pathways to the analysis of microarray data. *Trends Biotechnol.*, Vol.23, No.8, (August 2005), pp. 429-435, ISSN 0167-7799
- De, B., V; Al-Mulla, F. & Pot, B. (2007). Methods for microarray data analysis. *Methods Mol.Biol.*, Vol.382 (2007), pp. 373-391, ISSN 1064-3745
- Dembele, D. & Kastner, P. (2003). Fuzzy C-means method for clustering microarray data. *Bioinformatics.*, Vol.19, No.8, (May 2003), pp. 973-980, ISSN 1367-4803
- Dewey, T.G. (2002). From microarrays to networks: mining expression time series. *Drug Discovery Today*, Vol.7, No.20, (October 2002), pp. s170-s175, ISSN 1359-6446
- Dietz, G.P. & Bahr, M. (2004). Delivery of bioactive molecules into the cell: the Trojan horse approach. *Mol.Cell Neurosci.*, Vol.27, No.2, (October 2004), pp. 85-131, ISSN 1044-7431
- Doniger, S.W.; Salomonis, N.; Dahlquist, K.D.; Vranizan, K.; Lawlor, S.C. & Conklin, B.R. (2003). MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. *Genome Biol.*, Vol.4, No.1, (2003), pp. R7, ISSN 1465-6914
- Draghici, S. (2002). Statistical intelligence: effective analysis of high-density microarray data. *Drug Discov.Today*, Vol.7, No.11 Suppl, (June 2002), pp. S55-S63, ISSN 1359-6446
- Dugas, M.; Merk, S.; Breit, S. & Dirschedl, P. (2004). mdclust--exploratory microarray analysis by multidimensional clustering. *Bioinformatics.*, Vol.20, No.6, (April 2004), pp. 931-936, ISSN 1367-4803
- Duzgunes, N.; De Ilarduya, C.T.; Simoes, S.; Zhdanov, R.I.; Konopka, K. & Pedroso de Lima, M.C. (2003). Cationic liposomes for gene delivery: novel cationic lipids and enhancement by proteins and peptides. *Curr Med Chem*, Vol.10, No.14, (2003), pp. 1213-20, ISSN 1568-0169
- Edelstein, M. (2011). Gene therapy clinical trials worldwide In: *Wiley InterSciences*, (March 2011), Available from: <http://www.wiley.com/legacy/wileychi/genmed/clinical/>
- Ferber, D. (2001). Gene therapy. Safer and virus-free? *Science*, Vol.294, No.5547, (2001), pp. 1638-42, ISSN 0036-8075
- Geller, S.C.; Gregg, J.P.; Hagerman, P. & Rocke, D.M. (2003). Transformation and normalization of oligonucleotide microarray data. *Bioinformatics.*, Vol.19, No.14, (September 2003), pp. 1817-1823, ISSN 1367-4803

- Guo, S.; Lu, J.; Subramanian, A. & Sonenshein, G.E. (2006). Microarray-assisted pathway analysis identifies mitogen-activated protein kinase signaling as a mediator of resistance to the green tea polyphenol epigallocatechin 3-gallate in her-2/neu-overexpressing breast cancer cells. *Cancer Res.*, Vol.66, No.10, (May 2006), pp. 5322-5329, ISSN 0008-5472
- Gusnanto, A.; Calza, S. & Pawitan, Y. (2007). Identification of differentially expressed genes and false discovery rate in microarray studies. *Curr.Opin.Lipidol.*, Vol.18, No.2, (April 2007), pp. 187-193, ISSN 0957-9672
- Hawkins, D.M. (2004). The problem of overfitting. *J.Chem.Inf.Comput.Sci.*, Vol.44, No.1, (January 2004), pp. 1-12, ISSN 0095-2338
- Hembruff, S.L.; Villeneuve, D.J. & Parissenti, A.M. (2005). The optimization of quantitative reverse transcription PCR for verification of cDNA microarray data. *Anal.Biochem.*, Vol.345, No.2, (October 2005), pp. 237-249, ISSN 0003-2697
- Hobman, J. L., Jones, A., and Constantinidou, C. (2007). Introduction to microarray technology, In: *Microarray technology through applications*, Falciani, F., pp. 1-52, Taylor & Francis Group, ISBN 9780415378536, New York
- Hollins, A.J.; Omid, Y.; Benter, I.F. & Akhtar, S. (2007). Toxicogenomics of drug delivery systems: Exploiting delivery system-induced changes in target gene expression to enhance siRNA activity. *J.Drug Target*, Vol.15, No.1, (January 2007), pp. 83-88, ISSN 1061-186X
- Jares, P. (2006). DNA microarray applications in functional genomics. *Ultrastruct.Pathol.*, Vol.30, No.3, (May 2006), pp. 209-219, ISSN 0191-3123
- Joshi, T.; Chen, Y.; Alexandrov, N. & Xu, D. (2004). Cellular function prediction and biological pathway discovery in Arabidopsis thaliana using microarray data. *Conf.Proc.IEEE Eng Med.Biol.Soc.*, Vol.4 (2004), pp. 2881-2884, ISSN 1557-170X
- Juan, H.F. & Huang, H.C. (2007). Bioinformatics: microarray data clustering and functional classification. *Methods Mol.Biol.*, Vol.382 (2007), pp. 405-416, ISSN 1064-3745
- Kanehisa, M. (2002). The KEGG database. *Novartis.Found.Symp.*, Vol.247 (2002), pp. 91-101, ISSN 1528-2511
- Kepler, T.B.; Crosby, L. & Morgan, K.T. (2002). Normalization and analysis of DNA microarray data by self-consistency and local regression. *Genome Biol.*, Vol.3, No.7, (June 2002), pp. 1-12, ISSN 1465-6906
- Kim, S.; Kim, J.H.; Jeon, O.; Kwon, I.C. & Park, K. (2008). Engineered polymers for advanced drug delivery. *Eur.J.Pharm.Biopharm.*, Vol.71, No.3, (October 2008), pp. 420-430, ISSN 0939-6411
- Lee, J.K.; Williams, P.D. & Cheon, S. (2008). Data mining in genomics. *Clin.Lab Med.*, Vol.28, No.1, (March 2008), pp. 145-166, ISSN 0272-2712
- Lehmussola, A.; Ruusuvaori, P. & Yli-Harja, O. (2006). Evaluating the performance of microarray segmentation algorithms. *Bioinformatics.*, Vol.22, No.23, (December 2006), pp. 2910-2917, ISSN 1367-4803
- Leung, Y.F. & Cavalieri, D. (2003). Fundamentals of cDNA microarray data analysis. *Trends Genet.*, Vol.19, No.11, (November 2003), pp. 649-659, ISSN 0168-9525
- Li, X.; Wang, H.; Touma, E.; Rousseau, E.; Quigg, R.J. & Ryaby, J.T. (2007). Genetic network and pathway analysis of differentially expressed proteins during critical cellular events in fracture repair. *J.Cell Biochem.*, Vol.100, No.2, (February 2007), pp. 527-543, ISSN 1097-4644

- Liao, J.G. & Chin, K.V. (2007). Logistic regression for disease classification using microarray data: model selection in a large p and small n case. *Bioinformatics.*, Vol.23, No.15, (August 2007), pp. 1945-1951, ISSN 1367-4803
- Linder, R.; Richards, T. & Wagner, M. (2007). Microarray data classified by artificial neural networks. *Methods Mol.Biol.*, Vol.382 (2007), pp. 345-372, ISSN 1064-3745
- Lonardi, S. & Luo, Y. (2004). Gridding and compression of microarray images. *Proc.IEEE Comput.Syst.Bioinform.Conf.* (2004), pp. 122-130, ISSN 1471-2105
- Loring, J.F. (2006). Evolution of microarray analysis. *Neurobiol.Aging*, Vol.27, No.8, (August 2006), pp. 1084-1086, ISSN 0197-4580
- Martin, S.; Zhang, Z.; Martino, A. & Faulon, J.L. (2007). Boolean dynamics of genetic regulatory networks inferred from microarray time series data. *Bioinformatics.*, Vol.23, No.7, (April 2007), pp. 866-874, ISSN 1367-4803
- Mircean, C.; Tabus, I.; Kobayashi, T.; Yamaguchi, M.; Shiku, H.; Shmulevich, I. & Zhang, W. (2004). Pathway analysis of informative genes from microarray data reveals that metabolism and signal transduction genes distinguish different subtypes of lymphomas. *Int.J.Oncol.*, Vol.24, No.3, (March 2004), pp. 497-504, ISSN 1019-6439
- Nahar, J.; Ali, S. & Chen, Y.P. (2007). Microarray data classification using automatic SVM kernel selection. *DNA Cell Biol.*, Vol.26, No.10, (October 2007), pp. 707-712, ISSN 1044-5498
- Nguyen, D.V. (2005). Partial least squares dimension reduction for microarray gene expression data with a censored response. *Math.Biosci.*, Vol.193, No.1, (January 2005), pp. 119-137, ISSN 0025-5564
- Nie, Y.; Zhang, Z.R. & Duan, Y.R. (2006). Combined use of polycationic peptide and biodegradable macromolecular polymer as a novel gene delivery system: a preliminary study. *Drug Deliv.*, Vol.13, No.6, (November 2006), pp. 441-446, ISSN 1071-7544
- Olson, N.E. (2006). The microarray data analysis process: from raw data to biological significance. *NeuroRx.*, Vol.3, No.3, (July 2006), pp. 373-383, ISSN 1545-5343
- Omidi, Y. (2008). Microarray Analysis of the Toxicogenomics and the Genotoxic Potential of a Cationic Lipid-Based Gene Delivery Nanosystem in Human Alveolar Epithelial A549 Cells. *Toxicology Mechanisms and Methods*, Vol.18, No.4, (2008), pp. 369-378,
- Omidi, Y.; Barar, J. & Akhtar, S. (2005a). Toxicogenomics of cationic lipid-based vectors for gene therapy: impact of microarray technology. *Curr.Drug Deliv.*, Vol.2, No.4, (October 2005), pp. 429-441, ISSN 1567-2018
- Omidi, Y.; Hollins, A.J.; Benboubetra, M.; Drayton, R.; Benter, I.F. & Akhtar, S. (2003). Toxicogenomics of non-viral vectors for gene therapy: a microarray study of lipofectin- and oligofectamine-induced gene expression changes in human epithelial cells. *J.Drug Target*, Vol.11, No.6, (July 2003), pp. 311-323, ISSN 1567-2018
- Omidi, Y.; Hollins, A.J.; Drayton, R.M. & Akhtar, S. (2005b). Polypropylenimine dendrimer-induced gene expression changes: the effect of complexation with DNA, dendrimer generation and cell type. *J.Drug Target*, Vol.13, No.7, (August 2005), pp. 431-443, ISSN 1061-186X
- Pandey, G.; Yoshikawa, K.; Hirasawa, T.; Nagahisa, K.; Katakura, Y.; Furusawa, C.; Shimizu, H. & Shioya, S. (2007). Extracting the hidden features in saline osmotic tolerance in *Saccharomyces cerevisiae* from DNA microarray data using the self-organizing

- map: biosynthesis of amino acids. *Appl.Microbiol.Biotechnol.*, Vol.75, No.2, (May 2007), pp. 415-426, ISSN 0175-7598
- Pardridge, W.M. (2008). Re-engineering biopharmaceuticals for delivery to brain with molecular Trojan horses. *Bioconjug.Chem.*, Vol.19, No.7, (July 2008), pp. 1327-1338, ISSN 1043-1802
- Park, S.E.; Lee, M.J.; Yang, M.H.; Ahn, K.Y.; Jang, S.I.; Suh, Y.J.; Myung, H.; You, J.C. & Park, J.H. (2007). Expression profiles and pathway analysis in HEK 293 T cells overexpressing HIV-1 Tat and nucleocapsid using cDNA microarray. *J.Microbiol.Biotechnol.*, Vol.17, No.1, (January 2007), pp. 154-161, ISSN 1017-7825
- Patel, M.M.; Goyal, B.R.; Bhadada, S.V.; Bhatt, J.S. & Amin, A.F. (2009). Getting into the brain : approaches to enhance brain drug delivery. *CNS.Drugs*, Vol.23, No.1, (2009), pp. 35-58, ISSN 1172-7047
- Peng, Y. (June 2006). A novel ensemble machine learning for robust microarray data classification. *Comput.Biol.Med.*, Vol.36, No.6, (June 2006), pp. 553-573, ISSN 0010-4825
- Portney, N.G. & Ozkan, M. (2006). Nano-oncology: drug delivery, imaging, and sensing. *Anal.Bioanal.Chem.*, Vol.384, No.3, (February 2006), pp. 620-630, ISSN 1618-2642
- Pounds, S.B. (2006). Estimation and control of multiple testing error rates for microarray studies. *Brief.Bioinform.*, Vol.7, No.1, (March 2006), pp. 25-36, ISSN 1477-4054
- Putnam, D.; Gentry, C.A.; Pack, D.W. & Langer, R. (2001). Polymer-based gene delivery with low cytotoxicity by a unique balance of side-chain termini. *Proc Natl Acad Sci U S A*, Vol.98, No.3, (2001), pp. 1200-5, ISSN 0027-8424
- Qin, L.; Rueda, L.; Ali, A. & Ngom, A. (2005). Spot detection and image segmentation in DNA microarray data. *Appl.Bioinformatics.*, Vol.4, No.1, (2005), pp. 1-11, ISSN 1175-5636
- Qin, Y. & Tian, Y.P. (2010). A microarray gene analysis of peripheral whole blood in normal adult male rats after long-term GH gene therapy. *Cell Mol.Biol.Lett.*, Vol.15, No.2, (June 2010), pp. 177-195, ISSN 1425-8153
- Quackenbush, J. (2001). Computational analysis of microarray data. *Nat.Rev.Genet.*, Vol.2, No.6, (June 2001), pp. 418-427,
- Quackenbush, J. (2002). Microarray data normalization and transformation. *Nat.Genet.*, Vol.32 Suppl (December 2002), pp. 496-501, ISSN 1061-4036
- Rosa, G.J.; de, L.N. & Rosa, A.J. (2006). Review of microarray experimental design strategies for genetical genomics studies. *Physiol Genomics*, Vol.28, No.1, (December 2006), pp. 15-23, ISSN 1094-8341
- Saei, A.A. & Omid, Y. (2011). A glance at DNA microarray technology and applications. *BioImpacts*, Vol.1, No.2, (2011), pp. in press, ISSN 2228-5652
- Sawant, R.M.; Hurley, J.P.; Salmaso, S.; Kale, A.; Tolcheva, E.; Levchenko, T.S. & Torchilin, V.P. (2006). "SMART" drug delivery systems: double-targeted pH-responsive pharmaceutical nanocarriers. *Bioconjug.Chem.*, Vol.17, No.4, (July 2006), pp. 943-949, ISSN 1043-1802
- Senzer, N. & Nemunaitis, J. (2009). A review of contusogene ladenovex (Advexin) p53 therapy. *Curr.Opin.Mol.Ther.*, Vol.11, No.1, (February 2009), pp. 54-61, ISSN 1464-8431
- Shannon, W.; Culverhouse, R. & Duncan, J. (2003). Analyzing microarray data using cluster analysis. *Pharmacogenomics.*, Vol.4, No.1, (January 2003), pp. 41-52, ISSN 1462-2416

- Shen, F. & Hasegawa, O. (2008). A fast nearest neighbor classifier based on self-organizing incremental neural network. *Neural Netw.*, Vol.21, No.10, (December 2008), pp. 1537-1547, ISSN 0893-6080
- Shen, L. & Tan, E.C. (2005). Dimension reduction-based penalized logistic regression for cancer classification using microarray data. *IEEE/ACM.Trans.Comput.Biol.Bioinform.*, Vol.2, No.2, (April 2005), pp. 166-175, ISSN 1545-5963
- Shi, L. (2011). The MicroArray Quality Control (MAQC) Project: An FDA-Led Effort Toward Personalized Medicine In: *U.S.Food and Drug Administration*, (March 2011), Available from:
<http://www.fda.gov/downloads/ScienceResearch/BioinformaticsTools/MicroarrayQualityControlProject/ucm111050.pdf>
- Steinley, D. (2006). K-means clustering: a half-century synthesis. *Br.J.Math.Stat.Psychol.*, Vol.59, No.Pt 1, (May 2006), pp. 1-34, ISSN 0007-1102
- Subramanian, A.; Kuehn, H.; Gould, J.; Tamayo, P. & Mesirov, J.P. (2007). GSEA-P: a desktop application for Gene Set Enrichment Analysis. *Bioinformatics.*, Vol.23, No.23, (December 2007), pp. 3251-3253, ISSN 1367-4803
- Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S. & Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc.Natl.Acad.Sci.U.S.A*, Vol.102, No.43, (October 2005), pp. 15545-15550, ISSN 0027-8424
- Tai, F. & Pan, W. (2007). Incorporating prior knowledge of gene functional groups into regularized discriminant analysis of microarray data. *Bioinformatics.*, Vol.23, No.23, (December 2007), pp. 3170-3177, ISSN 1367-4803
- Wall, M.E.; Dyck, P.A. & Brettin, T.S. (2001). SVDMAN--singular value decomposition analysis of microarray data. *Bioinformatics.*, Vol.17, No.6, (June 2001), pp. 566-568, ISSN 1367-4803
- Wang, A. & Gehan, E.A. (2005). Gene selection for microarray data analysis using principal component analysis. *Stat.Med.*, Vol.24, No.13, (July 2005), pp. 2069-2087, ISSN 0277-6715
- Wang, Y.; Tetko, I.V.; Hall, M.A.; Frank, E.; Facius, A.; Mayer, K.F. & Mewes, H.W. (2005). Gene selection from microarray data for cancer classification--a machine learning approach. *Comput.Biol.Chem.*, Vol.29, No.1, (February 2005), pp. 37-46, ISSN 1476-9271
- Watson, A.; Mazumder, A.; Stewart, M. & Balasubramanian, S. (1998). Technology for microarray analysis of gene expression. *Curr.Opin.Biotechnol.*, Vol.9, No.6, (December 1998), pp. 609-614, ISSN 0958-1669
- Whetzel, P.L.; Parkinson, H.; Causton, H.C.; Fan, L.; Fostel, J.; Fragoso, G.; Game, L.; Heiskanen, M.; Morrison, N.; Rocca-Serra, P.; Sansone, S.A.; Taylor, C.; White, J. & Stoeckert, C.J., Jr. (2006). The MGED Ontology: a resource for semantics-based description of microarray experiments. *Bioinformatics.*, Vol.22, No.7, (April 2006), pp. 866-873, ISSN 1367-4803
- Wilkin, G.A. & Huang, X. (2008). A practical comparison of two K-Means clustering algorithms. *BMC.Bioinformatics.*, Vol.9 Suppl 6 (2008), pp. S19, ISSN 1471-2105
- Xiang, C.C. & Chen, Y. (2000). cDNA microarray technology and its applications. *Biotechnol.Adv.*, Vol.18, No.1, (March 2000), pp. 35-46, ISSN 0734-9750

- Xiao, L.; Wang, K.; Teng, Y. & Zhang, J. (2003). Component plane presentation integrated self-organizing map for microarray data analysis. *FEBS Lett.*, Vol.538, No.1-3, (March 2003), pp. 117-124, ISSN 0014-5793
- Zhu, Y.; Shen, X. & Pan, W. (2009). Network-based support vector machine for classification of microarray samples. *BMC.Bioinformatics.*, Vol.10 Suppl 1 (2009), pp. S21, ISSN 1471-2105



Gene Therapy - Developments and Future Perspectives

Edited by Prof. Chunsheng Kang

ISBN 978-953-307-617-1

Hard cover, 356 pages

Publisher InTech

Published online 22, June, 2011

Published in print edition June, 2011

The aim of this book is to cover key aspects of existing problems in the field of development and future perspectives in gene therapy. Contributions consist of basic and translational research, as well as clinical experiences, and they outline functional mechanisms, predictive approaches, patient-related studies and upcoming challenges in this stimulating but also controversial field of gene therapy research. This source will make our doctors become comfortable with the common problems of gene therapy and inspire others to delve a bit more deeply into a topic of interest.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Jaleh Barar, Amir Ata Saei and Yadollah Omid (2011). Impacts of DNA Microarray Technology in Gene Therapy, *Gene Therapy - Developments and Future Perspectives*, Prof. Chunsheng Kang (Ed.), ISBN: 978-953-307-617-1, InTech, Available from: <http://www.intechopen.com/books/gene-therapy-developments-and-future-perspectives/impacts-of-dna-microarray-technology-in-gene-therapy>

INTECH

open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.