ARBS
Annual Review of Biomedical Sciences
ISSN 1806-8774 - http://arbs.biblioteca.unesp.br

unesp

Functional Genomics: Identifying Critical Determinants of Term and Preterm Labor

Michael Sean Esplin
Department of Obstetrics and Gynecology, University of Utah
Rm. 2B200, 50 N. Medical Dr., Salt Lake City, UT 84132, USA.
E-mail: mesplin@hsc.utah.edu

Abstract

Esplin MS. Functional Genomics: Identifying Critical Determinants of Term and Preterm Labor. ARBS Ann Rev Biomed Sci 2005;7:1-6. The completion of the human genome project provides an opportunity to understand molecular control at the cellular level that is unprecedented. The most important next step is the move from sequence information to functional analysis of the genes that have been identified, a process that may be referred to as functional genomics. To date, the genes that initiate and maintain labor or those that are responsible for many physiologic processes have been identified only on a case-by-case basis. However, relatively new molecular techniques, such as cDNA microarray provide the means to identify such genes in a systematic and relatively rapid fashion, allowing the development of new or improved therapeutic regimens to augment term labor or inhibit preterm labor. In the present study the history and uses of microarray technology is reviewed, emphasizing how the advent of cDNA microarray techniques has changed the approach of functional genomics and efficient method for gene expression monitoring.
Key words: Genomics, Proterm Labor, Microarray Technology, DNA.

Mini-review Received: 02/08/2004 Accepted: 23/12/2004

Contents

Introduction
History of microarray
Performing the microarray
Single gene approach
Cluster analysis
Uses of microarray technology
Conclusion
References

Introduction

The completion of the human genome project provides an opportunity to understand molecular control at the cellular level that is unprecedented. However, obtaining the complete sequence of the genome is only the first step in the process. New DNA sequences may represent previously unknown genes. The most important next step is the move from sequence information to functional analysis of the genes that have been identified, a process that may be referred to as functional genomics. Evaluation of this new database may yield new insights into the expression patterns and potential function of each newly identified gene in the moment-to-moment control of the cell or the underlying etiologies of abnormal physiologic conditions.

Preterm labor complicates as many as 10% of all pregnancies. However, preterm infants account for more than 70% of the neonatal morbidity and mortality (Rush *et al.* 1976). Preterm infants are 40 times more likely to die in the neonatal period than term counterparts (Lipper *et al.,* 1990). The risk of serious complications including cerebral palsy, respiratory illness, blindness and deafness is increased in premature neonates as compared with term counterparts (European Community Study, 1990). Long-term neurologic and developmental deficits have been identified in as many as 70% of children with very low (< 800 grams) birth weights (Lipper *et al.,* 1990). Inevitably, these complications are associated with substantial costs. An estimated $5 billion was spent in 1985 in the United States alone for the neonatal care of preterm infants.

Despite the magnitude of this clinically challenging problem, our understanding of the basic underlying etiology and our ability to affect the outcome remain limited. Current tocolytic therapy (e.g. â-mimetics, magnesium sulfate or indomethacin) carries significant risks to both mother and fetus. Moreover, there exists no incontrovertible evidence as to the effectiveness of the aforementioned therapeutic agents in preventing or arresting preterm labor. Clearly, development of effective tocolytic therapy requires a greater understanding of the cellular and molecular mechanisms involved with the onset of labor.

Labor may be viewed as a cascade of events designed to convert a quiescent, unresponsive, non-excitable and poorly coordinated uterus to an active, highly responsive, excitable and coordinated organ capable of sustaining the synchronous contractions of labor. According to current concepts, the initiation of labor and the subsequent maintenance of regular, organized uterine contractions leading to the delivery of a fetus rely upon an orderly sequence of up- and down-regulation of a series of key genes in gestational tissues including the myometrium.

The expression patterns of genes and thus the relative levels of RNA and gene products and their interaction with the environment likely control the performances of tissue-specific functions, such as labor. These processes likely involve multiple gene pathways involving hundreds or even thousands of genes. Understanding the timing and level of expression within a given cell or tissue type may elucidate the functional role of a given gene. Alteration in expression pattern between tissue types, during disease states or following specific stimulation or perturbations may also shed light on the function of the gene of interest. It is also likely that genes with similar expression patterns (i.e. similar temporal expressions or response to stimuli) may have similar functions or roles; therefore the grouping of gene expression may be a valuable tool for functional analysis.

The shear quantity of information found in the human genome may prove to be a difficulty in screening for functionality. Previous attempts to identify and characterize novel genes involved in development or other cellular process have been limited by a gene-by-gene approach. New high throughput techniques may allow for the assessment of multiple DNA sequences in a rapid and efficient manner. The advent of cDNA microarray techniques has changed the approach of functional genomics and these techniques have become a popular method for gene expression monitoring.

History Of Microarray

Complementary DNA (cDNA) microarray evolved from and improved upon previous forms of differential screening, which have employed gel or filter-based techniques. The use of high-speed robotic printing of samples on a glass slide allows for miniaturization of the array as well as an increase in the numbers of expressed sequence tags (EST) or unique DNA sequences that may be evaluated. The use of fluorescence detection systems with two colors facilitates comparison of independent samples.

Since it was first described in 1995, microarray technology has been used to evaluate gene expression in cell cycle regulation in multiple organisms. In the human it has been employed to study a variety of disease processes including different types of cancer and inflammation (rheumatoid arthritis and bowel disease) (Schena *et al.*, 1995). For profiling mRNA expression cDNA microarray is now one of the most widely used methods.

Performing The Microarray

The exact techniques used to perform the microarray evaluation have been fully described (Schena *et al.*, 1996; Hedge *et al.*, 2000) yet a brief discussion of the process involved in performing the microarray will allow for a discussion of the benefits as well as some of the limitation of this technique. The first step involves fabrication of the array. DNA segments representing the collection of genes to be evaluated are amplified by PCR and mechanically spotted at high density on a glass microscope slide. The use of the high throughput robotic spotter increases the density of the spots and thus the number of sequences that may be evaluated in a single hybridization. The cDNA to be used for the microarray may include fully sequenced clones, ESTs or randomly chosen clones from any library of interest. Once the spotting has been completed, the cDNAs are chemically linked to the glass surface and denatured by heat treatment. If the cDNAs are derived from a library, low abundance cDNAs are unlikely to be spotted and the library must be normalized to reduce the redundant spotting of highly expressed genes.

During the hybridization step, the microarray slide is queried by co-hybridization assay using two or more fluorescently labeled probes prepared from mRNA from the cellular phenotypes of interest. The cDNA probes are prepared by reverse transcription in the presence of Cy3 (green) or Cy5 (red) labeled dUTP. A confocal laser scanner is then used to measure fluorescence intensities. A separate scan is conducted for each fluorophore and then a combination of the two scans allowing a comparison of relative expression levels of all the genes represented on the array. The use of a single hybridization for both mRNA populations of interest avoids the complications that arise when one is forced to compare results from independent hybridizations.

The final step in the microarray experiment is the analysis and interpretation of data. This is the portion of the experiment that varies the most and is possibly the most important to the outcome of the research.

Single Gene Approach

We have used cDNA micro-array technology to compare gene expression in the quiescent and laboring lower uterine segment and to construct a labor-selective myometrial cDNA library. From this library a selected group of genes was characterized using real time RT-PCR, Northern blot analysis, ELISA, in-situ hybridization, and immunohistochemistry in samples of myometrium, membranes, placenta and amniotic fluid from quiescent and laboring patients to verify differential expression and identify the cell types responsible for gene expression. Greater than 15,000 cDNA sequences were screened using micro-array and 30 were found to be consistently up-regulated in the lower uterine segment during labor.

For this type of study, the ordering of genes is based on the extremes of expression or repression. It is generally accepted that a gene with an expression level more than two to three standard deviation above the background expression may be considered “up-regulated”. Similarly, a gene with an expression level that decreases more than two to

three standard deviations as compared to the background expression is considered "down-regulated". Expression levels in this type of analysis may then be considered as fold increases or decreases as compared with controls.

The use of statistical methods, such as analysis of variance (ANOVA), to evaluate microarray results provides a more robust method of analysis. ANOVA provides an estimate of the level of confidence in differential expression in the form of a p value and may result in a more standard and easily reproducible method to analyze microarray data.

There are several advantages to the single gene method. It is a reliable way to screen large numbers of genes (potentially the entire genome) in a relatively short time. The single gene approach may also be used to identify the temporal changes in gene expression during a specific cellular process (i.e. labor or ovulation) or following a response to a stimulus or perturbation. It is also not necessary to have a pre-conceived notion about the exact genes or mechanisms that will be important, thus a broader and less biased view may be obtained. Because arrays often contain DNA sequences for genes of unknown function, relatively novel genes may be found to play a critical role in the cellular process being evaluated.

One major disadvantage of the single gene approach is that it represents an over simplification of the data. Of the thousands of genes that are evaluated, few may be identified as "up-regulated". For example, Heller *et al.* (1997) reported their results of microarray analysis of inflammatory disease related genes and found only 3 genes up-regulated of the 1,000 genes evaluated from a cDNA library of peripheral blood. Thus, large amounts of data (i.e. the expression profile of the remaining 977 genes) are overlooked in this type of analysis.

A second disadvantage of the single gene approach is the lack of evaluation of relationships between genes. No information is obtained with respect to the inter-relationship between expression patterns among groups of genes. Thus, the analysis fails to account for the complex interactions that are taking place in vivo.

Finally, this approach is subject to experimental error since the focus of analysis is on the extremes of expression (i.e. genes that are highly expressed or repressed) and therefore experimental artifact may mask potential findings. The reproducibility of expression patterns is dependent on the precision of the spotting of the array, efficiency of conversion of the mRNA to cDNA probe, the consistency of the hybridization conditions and the accuracy of the measurement of fluorescent intensities. Lee *et al.* (2000) found that any single microarray output is subject to substantial variability. The reproducibility of microarray was assessed by printing triplicates of the same 288-cDNA sequences on one slide and hybridizing with probes from a single source. They found that false positives were common. A single hybridization may have a misclassification rate (the inaccurate labeling of a gene as expressed or repressed) as high as 9%. The reliability and accuracy of the results increase with repetition of the analysis. In this study, the maximum attainable precision was attained after three replicates with a final error rate of 0.7%. Therefore, the analysis must be repeated at least three times to verify the results.

Cluster Analysis

Cluster analysis refers to grouping of genes with similar patterns of expression. This method is used to evaluate data sets comprised of a time-course of expression levels (serial evaluation of expression levels over time) such as changes in a specific tissue or cell type during a developmental process or following a stimulus or perturbation. This type of analysis enables the investigator to evaluate the expression levels of thousands of genes simultaneously within a cell to provide a broad picture of the state of the cell.

Clustering methods may be divided into two general classes referred to as unsupervised or supervised clustering. The unsupervised method requires no predefined references therefore no prior knowledge of the expected gene expression is needed to conduct the analysis. In supervised clustering, novel genes are analyzed in reference to previously characterized genes and thus requires some knowledge of the gene expression and function.

In cases where little is known about gene expression, both methods may be used together (unsupervised followed by supervised) to improve the analysis.

Pairwise average-linkage cluster analysis is one example of an unsupervised clustering method that has been utilized to describe large libraries of gene expression data (Eisen *et al.*, 1998). This form of hierarchical clustering creates phylogenetic trees to reflect higher-order relationships between genes with similar expression patterns. A dendogram is constructed in which the branch lengths among genes also reflect the degree of similarity of expression. All genes evaluated in an experiment are eventually included in a single tree.

The first step in the unsupervised method is the determination of similarity between genes using a mathematical description. The data resulting from this analysis may be combined with a graphical representation, which assigns a color to quantitatively and qualitatively expression data at each point. The color of each cell is based on the fluorescence ratio and thus expression level of the particular gene. Highly expressed genes at a certain time point appear as green on the graph, while those with extremely low expression appear red. This graphic data presentation allows for the assimilation and evaluation of large amount of data.

Eisen *et al.* (1998) have used the pairwise average-linkage cluster method to analyze gene expression in the budding yeast *Saccharomyces cerevisiae* during cell division, sporulation and temperature shocks. They noted that the method correctly identified groups of genes with similar function. Genes of unrelated sequence but similar function such as stress response, mitochondrial protein synthesis and histone genes, were found to cluster tightly together with respect to expression levels confirming the utility of this method of analysis.

Supervised clustering methods take advantage of the knowledge that already exists about the function of groups of genes within a specific process. Several different supervised techniques have been described including Parzen windows, Fischer's linear discriminant, decision tree classifiers and support vector machines (SVM). SVMs use a training set to specify, in advance, which data should cluster together. The SVM is given a set of genes with a common function, such as genes coding for ribosomal proteins, and a set of genes that are known not to be members of the functional class, in this case to be used as a training sample. Using this information, the SVM would learn to discriminate between members and non-members of this functional class using expression data. Training samples may be constructed from the literature and other databases. Information as simple as common structure or previous evidence of co-expression may be utilized to form training samples. Brown *et al.* (2000) compared the ability of these methods to analyze the expression patterns of the *Saccharomyces cerevisiae* under the same conditions as used by Eisen *et al.* (1998). The SVMs were trained to recognize six functional classes of genes including genes involved in the tricarboxylic acid (TCA) cycle, respiration, cytoplasmic ribosomes, proteasomes and helix-turn-helix proteins as a control. Brown *et al.* (2000) found that SVMs best identified sets of genes with common function using expression data.

Uses Of Microarray Technology

The potential uses of the microarray technology include: 1) the identification of single genes that are consistently expressed during a given cellular process, 2) the classification of genes into functional groups based on patterns or clustering of gene expression over time or throughout a cellular process or response to stimulus and 3) the classification of tissue types based on gene expression. The experimental design that is required to answer each of these questions differs only in the tissue types that are chosen for evaluation and the method of data analysis and interpretation scheme that is employed.

Conclusion

To date, the genes that initiate and maintain labor or those that are responsible for many physiologic processes have been identified only on a case-by-case basis. However, relatively new molecular techniques, such as cDNA microarray provide the means to

identify such genes in a systematic and relatively rapid fashion. The identification of differentially expressed genes in the laboring term human myometrium may improve the understanding of myometrial function thereby allowing for the development of new or improved therapeutic regimens to augment term labor or inhibit preterm labor.

References

Brown MP, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, Ares M Jr, Haussler D. Knowledge-based analysis of microarray gene expression data by using support vector machines. Proc Natl Acad Sci USA 2000;97:262-7.

Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genomewide expression patterns. Proc Natl Acad Sci USA 1998;95:14863-8.

Hegde P, Qi R, Abernathy K, Gay C, Dharap S, Gaspard R, Hughes JE, Snesrud E, Lee N, Quackenbush J. A concise guide to cDNA microarray analysis. Biotechniques 2000;29:548-54.

Heller RA, Schena M, Chai A, Shalon D, Bedilion T, Gilmore J, Woolley DE, Davis RW. Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. Proc Natl Acad Sci USA 1997;94:2150-5.

Lee ML, Kuo FC, Whitmore GA, Sklar J. Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. Proc Natl Acad Sci USA 2000;97:9834-9.

Lipper EG, Ross GS, Auld PA, Glassman MB. Survival and outcome of infants weighing less than 800 grams at birth. Am J Obstet Gynecol 1990;163:146-50.

Rush RW, Keirse MJ, Howat P, Baum JD, Anderson AB, Turnbull AC. Contribution of preterm delivery to perinatal mortality. Br Med J 1976;2:965-8.

Schena M, Shalon M, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 1995;270:467-70.

Schena M. Genome analysis with gene expression microarrays. Bioessays 1996;18:427-31.

Working group on the very low birthweight infant. European Community collaborative study of outcome of pregnancy between 22 and 28 weeks. gestation. Lancet 1990; 336:782-4.