WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Review on: Gene Expression Analysis Techniques and its Application

Kavya S

PG Student, Vivekanandha College of Engineering For Women

DOI: https://doi.org/10.55248/gengpi.5.0424.1145

ABSTRACT

Beadle and Tatum presented research in 1941 that would provide an explanation for the fundamental tenet of molecular biology: proteins carry out the functions of cells by converting DNA via an intermediary molecule known as RNA. As biomedical research currently aims to elucidate the mechanisms by which a given disease arises, gene expression studies have shown to be an invaluable resource. The phrase "gene expression" refers only to the process from gene activation to mature protein localization in the appropriate compartment, where it functions and aids in the expression of the cell's phenotype. The goal of the expression research is to find and measure a gene's messenger RNA (mRNA) levels. The goal of the expression research is to find and measure a gene's messenger RNA (mRNA) levels. Alwine et al.'s Northern Blot, published in 1977, marked the beginning of the development of RNA-based gene expression research. Although in situ hybridization was independently invented in 1969 by Gall and Pardue and John et al., Coghlan did not use this approach to detect mRNA until 1986. Many of the methods used to quantify RNA are no longer in use since more information may be obtained using other, more recent methods. For transcriptome analysis, qPCR, expression microarrays, and RNA-seq are now the most used methods. We shall review these strategies in this chapter.

Keywords: Gene expression; Microarrays; RNA sequencing; Transcriptome ; qPCR ; hybridization

INTRODUCTION

The analysis of gene expression has become increasingly significant in biological and medical research. One gene at a time was the focus of traditional gene function and regulation research techniques. As a result, a great deal of significant biological changes were either overlooked or overlooked unexpectedly. Many techniques have been developed in the last ten years to enable researchers to look at the levels of mRNA expression of thousands of genes in a single experiment. Thanks to cutting-edge technologies like Microarray, there is an abundance of biological data available. Thousands of genes are expressed simultaneously, and a microarray can identify which genes are expressed and to what degree. Efficient computational approaches are needed to analyze and extract informative knowledge from these data. Preparing the data for gene expression could be required for a number of reasons. Sometimes it is impossible to quantify the expression values for some genes correctly due to technical issues or improper management of the microarrays [1]. Depending on the distribution of the repeated values, one can either average them or pick the median to collapse them into a single value. The literature on the analysis of gene expression data also suggests that the classification algorithms may be able to more easily and accurately model the underlying structure in the training data by normalizing and transforming the data using variance stabilizing transforms like logarithm and cubic-root [2-6]. Transcription Factors (TFs) are specialized proteins that attach to DNA promoter regions to interfere with the rate of protein synthesis. This causes an abrupt change in the expression levels of genes. There is a chance that this intervention will go wrong. An increase in the rate of protein synthesis is referred to as activation, or up positive regulation, while a reduction is referred to as inhibition, or down negative regulation. The nature of microarrays and other genetic data differs from that of classical data. It has frequently been required to modify current methods or create new ones in order to match the circumstances faced. Since changes in an organism's or a cell's physiology will inevitably result in changes in the pattern of gene expression, the analysis of gene expression is crucial in many biological study domains. Clinical samples, both healthy and sick, may be used in expression studies to find new biomarkers. Likewise, one can get understanding of the physiological effects of genetic alteration in plants by the examination of gene expression. Finding hidden patterns in gene expression data is a huge opportunity to improve our knowledge of functional genomics. for concurrently tracking thousands of genes' expression levels[7-11]

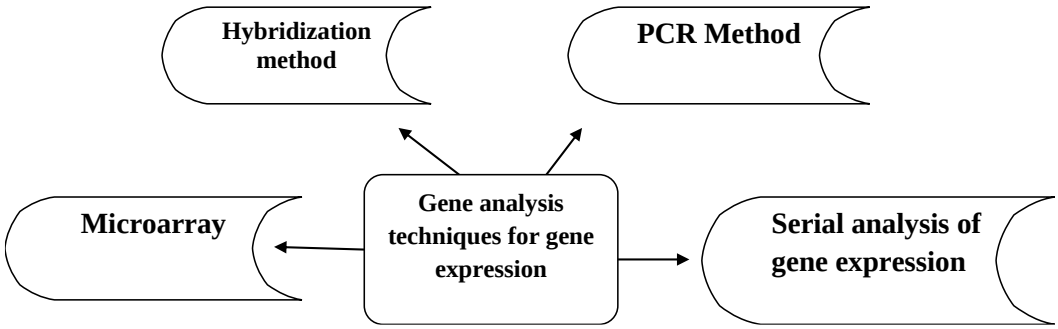


Fig No :1- Analysis Techniques for Gene Expression

Overview of Gene Expression

The process by which a gene's information is translated into a function is known as gene expression. The transcription of RNA molecules that code for proteins or non-coding RNA molecules with different roles is primarily responsible for this. One way to conceptualize gene expression is as a "volume control" that regulates the amount and timing of the synthesis of proteins and RNA molecules, as well as their location. Gene expression is a highly regulated process that varies significantly between cell types and environmental factors. Many genes produce RNA and proteins that control the expression of other genes. Observing a phenotype linked to a gene or evaluating the functional activity of a gene product can also be used to determine where, when, and how much a gene is expressed. The intricate process by which a cell uses its genetic code to produce useful products is known as gene expression. Multiple phases of regulation govern this process, and any disruption could result in illnesses like cancer. In addition to highlighting significant historical findings about gene expression, this documentary explains how different DNA base combinations encode the amino acids that make up proteins. A review of various methods for measuring gene expression and examining its regulation follows an exploration of important questions in the field of gene expression research. Lastly, we examine how these methods are now being applied by scientists to the study of gene expression.

Gene Expression Analysis Techniques

Hybridization method

Hybridization approaches are considered conventional procedures. They are typically employed to find specific sequences (targets) in a complicated mixture of RNA or DNA molecules. Typically, DNA or RNA is transferred to nitrocellulose or, more frequently, nylon in order to become immobilized. Linings Single-stranded complementary probes can have radioactive or non-radioactive labels. Probes form hydrogen bonds with their corresponding target sequence when they are hybridized to the filter. After that, the hybridized probe is removed, and the specifically-bound probe is found via color reaction or autoradiography.

Double-stranded DNA associates itself in a stable but reversible way when hydrogen bonds are formed. It is crucial that the two DNA strands can split and reassemble under physiological settings without breaking covalent connections in order for transcription and replication to occur in vivo. The in vitro methods discussed are based on this natural function.Under this heading. The process of pairing complementary single-stranded DNA or RNA molecules to create hybrid double-stranded molecules is known as hybridization [12]

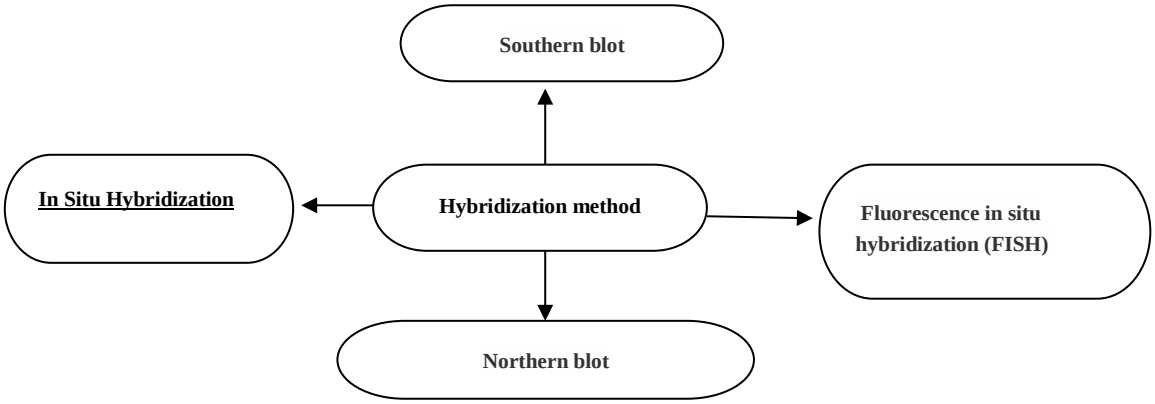
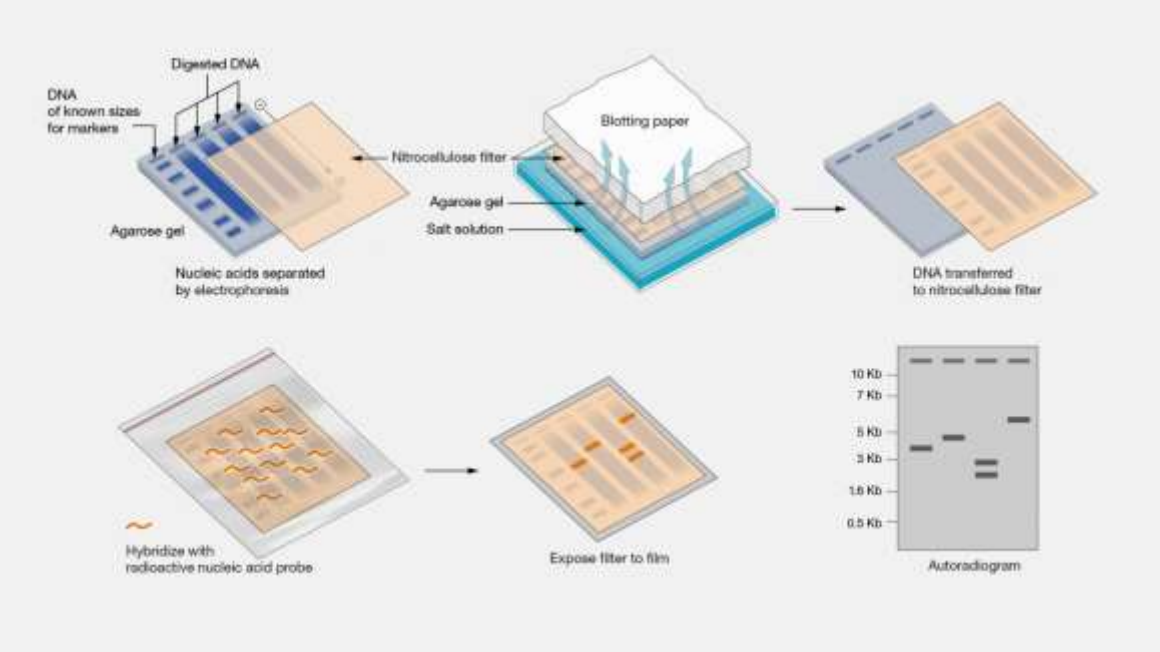


Fig No: 2 – Hybridization Methods

Southern blot:

The first of these techniques developed was the Southern blot, named after Dr. Edwin Southern who developed it to identify specific DNA sequences. In the realm of molecular biology, southern blotting is a detection technique used to identify the target DNA sequences in the DNA sample. First, DNA molecules are electrophoresed and then hybridized in a blotting membrane. Next, DNA from the gel is transferred to the blotting membrane in a transfer step.

Blotting techniques are a supplement to gel electrophoresis, which is a highly effective technique for isolating proteins, RNA, and DNA. They make it possible to identify particular molecules within the mixture that the gel has separated. Every technique has a common phase that involves moving molecules from the gel across a porous membrane; this is often accomplished by soaking a solution through both the membrane and the gel with the use of absorbent paper. Many vendors offer more sophisticated "electroblotting" equipment, which is especially helpful for transfers from polyacrylamide gels, which are less porous than agarose. Specific sequences in the membrane are identified for DNA and RNA5 using molecular hybridization with labeled nucleic acid probes; proteins are identified through the use of labeled antibodies[13]



PCR METHODS

Polymerase chain reaction (PCR) is a technology for exponential amplification of a fragment of DNA.(The PCR is covered by patents owned by HoffmanLa Roche. A license is required to use the PCR process.) Because PCR can identify only one molecule at its sensitivity limit, it is an excellent qualitative method for the targeted identification of uncommon DNA sequences. It can also be used as a quantitative analytical tool because, under the right circumstances, the yield of amplified DNA is proportionate to the original number of target molecules. Since its initial description in 1985, PCR has developed into a collection of diverse methodologies that are nearly always employed in forensics, food technology, environmental testing, biotechnology, clinical research, basic biological research, biotechnology, and other fields. While alternative strategies for amplification of nucleic acids have been described, PCR remains by far the most widely used.

Reverse Transcription PCR

Reverse Transcription PCR: If RNA is first reverse-transcribed to complementary DNA, PCR can be employed to detect and quantify RNA.(cDNA). Reverse transcriptases derived from the avian myeloblastoma virus (AMV-RT) or the Maloney murine leukemia virus (MMLV-RT) are frequently employed. Reverse transcription is usually carried out at 37–421C for 15–60 min, using either the PCR buffer or the buffer ideal for reverse transcription (RT).These enzymes are inactivated during the initial denaturation segment of PCR due to their heat-labile nature. Reverse transcription can denature RNA secondary structure at greater temperatures thanks to heat-stable reverse transcriptases. When Mn2 þ is present, some RNA templates are active for DNA-dependent DNA polymerases, such as Tth DNA polymerase, an enzyme from Thermus thermophilus. The same enzyme that is used for PCR can be utilized to synthesise cDNA when reverse transcription is performed using MnCl2. Mg2Í must be added before PCR and Mn2Í must be eliminated or chelated if a precise balance between MnCl2 and MgCl2 cannot be determined. In situ PCR, or RT-PCR, is a technique that can be used to detect cells expressing specific genes or determine whether disease-related genes are present in intact cells. It is possible to automate the process on microscope slides with thermal cyclers [14]

Real Time PCR

"Real-time PCR" or "fluorescence based PCR" enables the measurement of nucleic acids extracted from tissues or cells, the comparison of different infection stages, the identification of chromosomal translocations, the genotyping of single nucleotide polymorphisms, and ascertain the samples' degree of gene expression, and so forth. The most sensitive and precise technique for nucleic acid detection and quantification is now real-time PCR.

Understanding fundamental cellular mechanisms and identifying changes in gene expression levels in response to particular biological stimuli (e.g., growth factor or pharmaceutical agent) require quantitative measurement of specific gene expression using quantitative PCR (qPCR) [16]. The introduction of the Real-time PCR technology has greatly simplified the process of quantifying nucleic acids [17]. It is mostly used for two reasons: either as a primary investigative tool to determine gene expression or as a secondary tool to validate the results of DNA microarrays [18]

MICROARRAY

Gene expression analysis using DNA microarrays is a potent tool in our fields of expertise, which center on food and plant studies. In both cases domains, knowledge of the processes and interactions that take place in organisms and cells is crucial. It is anticipated that DNA microarray technology will be helpful in the following areas, among others: (i) risk assessment in transgenic agricultural products through the analysis of altered gene expression; (ii) food component functional and toxicological effects investigation; and (iii) using mutant analysis to unravel gene function and metabolic pathways in plants. While complicated sample identification and detection are other applications of DNA microarray technology, the focus of this review is on gene expression analysis. In addition to other technical details, the potential improvements of this technology will be discuss and illustrated by some of the above-mentioned applications[19]

SAGE

Serial analysis of gene expression (SAGE) uses mRNA from a particular sample to create complementary DNA (cDNA) fragments which are then amplified and sequenced using high-throughput sequencing technology. SAGE operates on the basis of tags that are able to recognize the original transcript and quick sequencing of chains of tags that are connected. Through the process of joining the cDNA segments into a lengthy chain, sequencing is effectively made simpler.

Method

Two basic principles form the foundation of the technique.

(1) Enough information can be found in a short oligonucleotide sequence tag (10 or 11 base pairs) to uniquely identify a transcript. These tags are used to identify genes within mRNA, as well as the relative abundance of their transcripts.

(2) Because SAGE uses serial processing, which analyzes 25–50 SAGE tags on each lane of a DNA sequencer, concatenation of short sequence tags enables the effective study of transcripts in a serial way.

The foundation of SAGE is the creation of brief nucleotide sequences (tags) consisting of nine or ten base pairs from a specific location within each species of mRNA. A sequence-based method known as serial analysis of gene expression (SAGE), which is simply an expedited form of EST sequencing, enables a thousands of transcripts quickly analyzed . This method involves first creating a unique sequence tag (10–14 bp) in the cell or tissue of interest that contains enough information for each transcript. Next, sequence tags are ligated to produce cloned and sequenced concatemers, which are then compared with databases to identify variations in the expression of the genes that the tags have identified.After mRNA is reverse transcribed using a biotinylated oligo(dT) primer and digested with a frequent-cutting restriction enzyme, sequence tags are produced.

Enzyme known as an anchoring enzyme, which, for example, limits 4-bp recognition sites. After being further purified using streptavidin beads, the 3' region of the cDNA is divided into two fractions, A and B, which are ligated using primers A' and B'. These primers have a type IIs restriction enzyme recognition site, which cleaves DNA at a specific distance from the recognition site. BsmI is one instance of this kind of restriction enzyme. Following extraction of the digested and unbound DNA segment and digestion with the appropriate type-IIs restriction enzyme (referred to as the "tagging enzyme"), the eluted DNA fragments are ligated and amplified using the primers. B' and A'. The anchoring enzyme is used to remove primer sites A and B during PCR. The DNA pieces can form concatemers, which can be cloned into a vector, thanks to the sticky ends that are therefore formed. Every clone has brief, distinct tags for at least 20 genes. All SAGE libraries are sequenced in order to establish transcription profiles. Given that every sequencing process provides data for at least 20 genes, it is tens of thousands of transcripts' worth of data points can be produced with a small sequencing effort. Sequence tag counts or clusters to yield the relative abundance of each gene. This brief sequence tag is enough for the majority of genes to give a unique identification through common database searches. By using PCR or hybridization-based techniques, the SAGE tag can be used to create a cDNA clone for a previously unidentified gene.[20]

APPLICATIONS OF GENE EXPRESSION ANALYSIS TECHNIQUES;

- **Hybridization method**

✓ When analyzing the expression and regulation of genes in both normal and pathologic tissues, hybridization techniques are an effective tool .One of hybridization main benefits is its localization capability.

- ✓ mRNA at the cellular level in many tissues, hence enhancing the findings of additional molecular methods like Northern blot hybridization for the investigation of particular genes. The study of mRNA encoding oncogenes, growth factors and their receptors, hormones and hormone receptors, cytokines, structural proteins, and hormones has benefited greatly from the application of hybridization approaches.
- ✓ Collagenase, enzymes, and other things. The hybridization techniques have numerous real-world uses in the field of tumor pathology. Researchers are looking at the relationship between oncogene expression and prognosis in neuroblastomas and epithelial neoplasms like breast, lung, colon, and prostate cancers.
- ✓ carcinomas. Potential uses of hybridization techniques in pathologic diagnosis include the identification of genes encoding cell structural proteins, such as tumor-associated indicators. Diagnostic pathology laboratories employ various techniques, such as nonisotopic hybridization methods to localize immunoglobulin light chain in mRNAs in hyperplastic and neoplastic lymphoproliferative disorders , albumin mRNA to differentiate between hepatocellular and metastatic carcinomas in the liver, and chromogranin/secretogranin mRNAs to classify neuroendocrine tumors [15]

- **PCR**

✓ Using PCR, a DNA sequence can be amplified millions or billions of times, producing enough DNA copies to be analyzed using other techniques. For instance, the DNA may be visualized by gel electrophoresis, sent for sequencing, or digested with restriction enzymes and cloned into a plasmid.

✓ PCR is used in many research labs, and it also has practical applications in forensics, genetic testing, and diagnostics. For instance, PCR is used to amplify genes associated with genetic disorders from the DNA of patients (or from fetal DNA, in the case of prenatal testing).

✓ PCR can also be used to test for a bacterium or DNA virus in a patient's body: if the pathogen is present, it may be possible to amplify regions of its DNA from a blood or tissue sample.

- **Microarray**
  - ✓ There are several uses for DNA microarray technology. Its application in the safety evaluation of food plant genetic alteration is the first.
  - ✓ This should consider potential unwanted side effects from the genetic alteration in addition to the assessment of the recently added trait.
  - ✓ Currently, the latter is accomplished by comparing the converted line to its traditional counterpart(s) based on a small number of distinct macro- and micronutrients as well as anti-nutritional factors, such as naturally occurring toxins.
  - ✓ This method has limits, too, as it's unclear if the elements under investigation are the most crucial or the only ones to look for when it comes to food safety. As a result, this kind of investigation offers little insight into the possible consequences on human health.
  - ✓ For assessing safety, DNA microarray technology might be a suitable substitute[19]
- **SAGE**
  - ✓ How SAGE can be applied to biological research is demonstrated by a study of novel cancer indicators.
  - ✓ In an effort to find markers that could aid in the early diagnosis of pancreatic cancer, researchers analyzed the levels of gene expression in malignant and non-cancerous tissues.
  - ✓ The researchers searched the database for genes selectively expressed in pancreatic cancer as the outcomes of a SAGE analysis of numerous representative tissues had already been made available online.
  - ✓ This allowed them to pinpoint a gene known as prostate stem cell antigen (PCSA) that had not before been linked to pancreatic cancer.
  - ✓ The advantage of SAGE over many other methods is that SAGE can provide both quantitative and qualitative data about gene expression, and the ability to accumulate and compare SAGE tag data from a variety of samples.
  - ✓ The disadvantages are related to the technical difficulties in generating good SAGE libraries and in analyzing the data. So far, only a limited number of research groups have effectively used SAGE.

CONCLUSION

One highly effective method for discovering novel and significant details in cutting-edge biological applications is gene expression analysis. These approaches will help scientists uncover new information, comprehend intricate gene regulatory networks, and comprehend the entirety of biological processes while analyzing any given gene. In the future, this will aid in the treatment of numerous hereditary illnesses. Programmers will always have difficulty classifying microarrays. Appropriate programming tools and higher-level concepts merely help the process along; they don't make it easy.

REFERENCE:

1. Herrero J, Diaz-Uriarte R, Dopazo J. Gene expression data preprocessing. Bioinformatics 2003; 19: 655-6

2. Durbin BP, Hardin JS, Hawkins DM, Rocke DM. A variancestabilizing transformation for gene-expression microarray data. Bioinformatics 2002; 18: S105-10

3. Ekstrom CT, Bak S, Kristensen C, Rudemo M. Spot shape modelling and data transformations for microarrays. Bioinformatics 2004; 20: 2270-78

4. Huber W, von Heydebreck A, Sultmann H, Poustka A, Vingron M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 2002; 18: S96-104

5. Geller SC, Gregg JP, Hagerman P, Rocke DM. Transformation and normalization of oligo-nucleotide microarray data. Bioinformatics 2003; 19: 1817-23.

6. Thygesen HH, Zwinderman AH. Comparing transformation methods for DNA microarray data. BMC Bioinformatics 2004; 5:77

7. Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, Hwang SY, Brown PO, Davis RW: Yeast microarrays for genome wide parallel genetic and gene expression analysis. Proc Natl Acad Sci USA 1997;94:13057– 13062

8. Hsiao L, Stears R, Hong R, Gullans S: Prospective use of DNA microarrays for evaluating renal function and disease. Curr Opin Nephrol Hypertens 2000;9:253–258

9. Kurella M, Hsiao L, Yoshida T, Randall J, Chow G, Sarang S, Jensen R, Gullans S: DNA microarray analysis of complex biologic processes. J Am Soc Nephrol 2001;12:1072–1078

10. Nagasawa Y, Takenaka M, Kaimori J, Matsuoka Y, Akagi Y, Tsujie M, Imai E, Hori M: Rapid and diverse changes in gene expression in the kidneys of protein-overloaded proteinuria mice detected by microarray analysis. Nephrol Dial Transplant 2001;16:923–931

11. Moch H, Schraml P, Bubendorf L, Mirlacher M, Kononen J, Gasser T, Mihatsch M, Kallioniemi O, Sauter G: High-throughput tissue microarray analysis to evaluate genes uncovered by cDNA microarray screening in renal cell carcinoma. Am J Pathol 1999;154:981–986

12. KAI-OLAF NETZER : Hybridization Methods (Southern and Northern Blotting), F. Hildebrandt et al. (eds.), Techniques in Molecular Medicine.1999

13. Southern blotting Ed Southern, Department of Biochemistry, University of Oxford; 2006; doi:10.1038/nprot.2006.73

14. J M Wages Jr., Palmetto Consulting & Research, Tupelo, MS, USA & 2005, Elsevier Ltd

15. In Situ Hybridization: Methods and Applications Long Jin1 and Ricardo V. Lloyd1 * 1Department of Laboratory Medicine and Pathology. L Jin, RV Lloyd - Journal of clinical laboratory analysis, 1997 - Wiley Online Library

16. Bustin, S. A. (2000). Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J. Mol. Endocrinol., 25, 2, 169-193.

17. Bustin, S. A. (2002). Quantification of mRNA using real-time reverse transcription PCR (RTPCR): trends and problems. J. Mol. Endocrinol., 29, 1, 23-39.

18. Valasek, M. A., Repa, J. J. (2005) The power of real-time PCR. Adv Physiol Educ., 29, 3, 151- 159.

19. Nicole L.W. van Hal a, Oscar Vorst ,et.all; The application of DNA microarrays in gene expression analysis; Journal of Biotechnology 78 (2000) 271–280

20. Kin-Ying To, Identification of Differential Gene Expression by High Throughput Analysis, Institute of BioAgricultural Sciences, Academia Sinica, Combinatorial Chemistry & High Throughput Screening, 2000, 3, 235-241