

Review

## Quantitative Analysis of Nucleic Acids - the Last Few Years of Progress

Chunming Ding\* and Charles R. Cantor\*<sup>†</sup>

\*Bioinformatics Program and Center for Advanced Biotechnology, Boston University, Boston, Massachusetts 02215, USA

<sup>†</sup>SEQUENOM Inc., San Diego, California 92121, USA

Received 23 December 2003

DNA and RNA quantifications are widely used in biological and biomedical research. In the last ten years, many technologies have been developed to enable automated and high-throughput analyses. In this review, we first give a brief overview of how DNA and RNA quantifications are carried out. Then, five technologies (microarrays, SAGE, differential display, real time PCR and real competitive PCR) are introduced, with an emphasis on how these technologies can be applied and what their limitations are. The technologies are also evaluated in terms of a few key aspects of nucleic acids quantification such as accuracy, sensitivity, specificity, cost and throughput.

**Keywords:** Gene expression, Quantitative PCR, Microarray, SAGE, Differential display

### Introduction

Nucleic acids (DNA and RNA) are among the most fundamental molecules of all life forms. DNA and RNA can be studied in two different manners. One can study the qualitative aspects such as genome sequences (Lander *et al.*, 2001; Venter *et al.*, 2001) and DNA mutations (Sachidanandam *et al.*, 2001). Alternatively, in quantitative studies, one asks how much of each sequence is present, such as gene expression analysis (Lockhart and Winzeler, 2000), DNA duplications/deletions (Ginzinger *et al.*, 2000; Nigro *et al.*, 2001) and viral titer determinations (Nyberg-Hoffman *et al.*, 1997). This review focuses on the quantitative aspects of nucleic acid research. Technological advances in the last few years will be reviewed. For each technology, we will attempt

to discuss how the analysis works, the advantages and disadvantages, as well as its specific applications. In the end, we hope the readers will be able to choose the most appropriate technology for their own research, and to be able to critically assess results obtained from different technology platforms.

Nucleic acids quantification includes RNA and DNA analyses. Quantitative RNA analysis, or gene expression analysis (or transcriptional profiling for high-throughput analysis), aims to measure the concentrations (or copy numbers) of specific RNA sequences. In most cases, the average steady state mRNA concentrations from many cells (hundreds to millions) are quantified. Quantitative DNA analysis can be used to detect genomic DNA duplications/deletions such as chromosome aneuploidy and loss of heterozygosity (LOH). We will use RNA quantification in most of the text, but the same principles also apply to DNA quantification.

An overview of the critical steps in quantifying mRNA is illustrated in Fig. 1. It is important to point out that each step can contribute to the bottleneck for high-throughput, accurate and reproducible mRNA quantification. An ideal technology should have a robust scheme to deal with the potential pitfalls in these steps.

For gene expression analysis, it is important to obtain a homogeneous sample containing only the desired cell type for RNA preparation. Two techniques, laser capture microdissection (Emmert-Buck *et al.*, 1996; Mikulowska-Mennis *et al.*, 2002) and flow cytometry (for review, see Herzenberg *et al.*, 2002) can be used to obtain homogenous cells from complex tissue samples. Since cells can respond to environmental changes by changing their expression profiles, it is important that minimum disturbance is inflicted during cell purification and RNA preparation. RNase inhibition reagents (such as *RNAlater* from Ambion, Austin, USA) are often added prior to cell lysis. Many commercial kits are now available for total RNA and mRNA preparations (such as Ambion and Qiagen kits, Qiagen, Hilden, Germany). For reverse transcription, various engineered AMV reverse transcriptase (*Sensiscript* from Qiagen; *ThermoScript* from

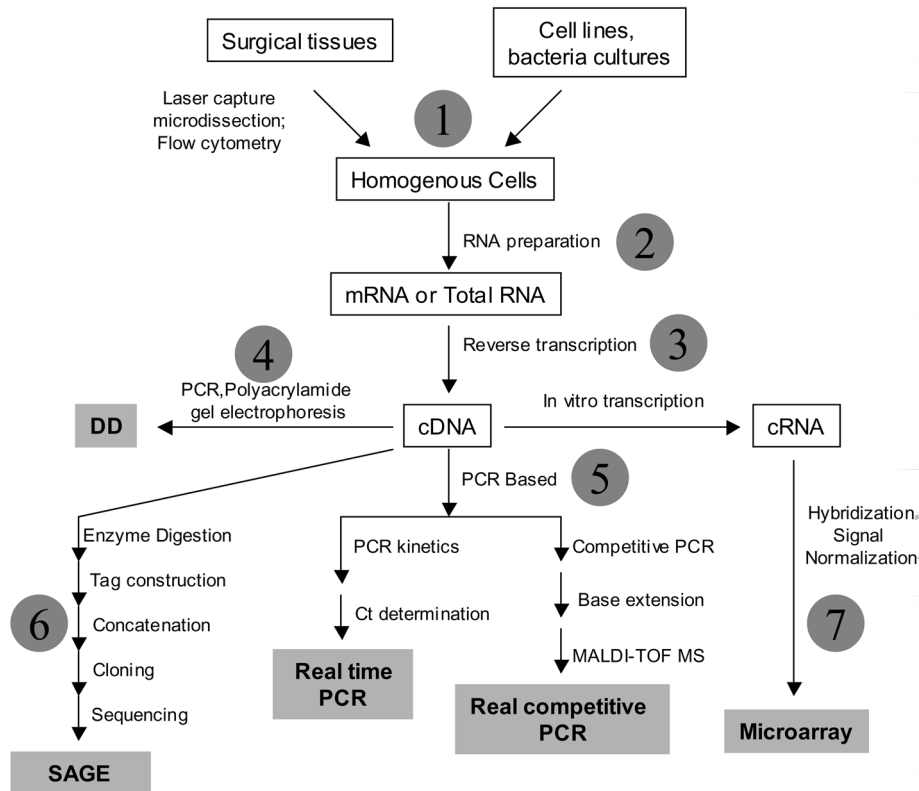
\*To whom correspondence should be addressed.

Tel: 1-617-353-8496; Fax: 1-617-353-8501

E-mail: cmding@bu.edu

Tel: 1-858-202-9012; Fax: 1-858-202-9020

E-mail: ccantor@sequenom.com



**Fig. 1.** Overview of mRNA quantifications. The entire process has many potential pitfalls and some of them are highlighted. Preparation of homogeneous cells. This step can be particularly difficult when a specific human tissue is needed. Laser capture microdissection (LCM) and flow cytometry can be used to obtain homogeneous cells. 2. RNA preparation. It is important to prevent RNAs from degradation in the preparation step by adding RNase inhibitors. In addition, shorter RNA or RNA without a poly(A) tail might be lost in the purification steps. 3. Reverse transcription. It is well known that RNAs can form quite complex secondary structures which can cause problems for efficient and consistent reverse transcription. 4-7 are covered in the text for each particular technology.

Invitrogen, San Diego, USA; Improm-II from Promega, Madison, USA), engineered M-MLV reverse transcriptase (SuperScript from Life Technologies, Grand Island, USA) or rTth reverse transcriptase/DNA polymerase (AppliedBiosystems, Foster, USA) can be used with either random hexamer, Oligo(dT) or gene specific primers. Due to the sometimes very significant secondary structures of RNA molecules, it is important that the reverse transcription efficiency is monitored and can be normalized. It is a common practice that one or more housekeeping genes are used for normalization. Choosing the gene(s) for normalization can have a significant impact on the results and there is no consensus on which genes to use (Goidin *et al.*, 2001; Solanas *et al.*, 2001; Vandesompele *et al.*, 2002).

## Technologies for Nucleic Acid Quantification

There are numerous technologies available for DNA and RNA quantification. It is clearly unrealistic to cover all of them in this brief review. We attempt to cover five medium/high-throughput technologies that we think have or will have the

biggest impact on DNA and RNA quantification.

It is important to point out that the reviewed techniques are developed for different purposes. Thus, it is neither fair nor appropriate to directly compare the techniques and conclude which technique is the best. Rather, we intend to illustrate how the techniques have successfully been applied to investigate biological and medical problems. Only after that do we provide a cross-platform comparison so that readers will be able to choose the technique(s) that are most suitable for their specific needs.

**Microarrays** DNA microarrays use the same principle of Watson-Crick base pairing as other traditional techniques such as Northern and Southern blotting (Southern, 1975; Alwine *et al.*, 1977). In Northern and Southern blotting, target DNA/RNA sequences are immobilized and labeled probe sequences in excess are added for hybridization. In DNA microarrays, probe sequences (either short oligonucleotides or cDNA sequences) are immobilized at pre-determined positions at high density, and labeled (usually fluorescent) target sequences are added and allowed to hybridize (Schena *et al.*, 1995; Lockhart *et al.*, 1996). After washing away

nonspecifically bound targets, the signal intensity of the remaining bound target sequences is proportional to the amount of the target sequence; this is the basis for nucleic acids quantification.

There are two widely adopted methods for microarray production: robotic deposition of nucleic acids (PCR products or oligonucleotides) onto a glass slide and *in situ* synthesis of oligonucleotides (for review, see (Cheung *et al.*, 1999)). For short oligonucleotide (20 to 25 nucleotides in length) microarrays, it is common to have multiple (10 to 20) probes for each mRNA transcript, and signals are statistically estimated from all probes. Target sequences can be labeled in several different ways (for review, see (Lockhart and Winzeler, 2000)). When RNA quantity is sufficient (>20 µg total RNA), the RNA can be labeled directly with biotin (Lockhart *et al.*, 1996). When RNA quantity is limited, RNA is first converted to double-stranded cDNA, and subsequently transcribed into labeled anti-sense RNA by *in vitro* transcription (Wodicka *et al.*, 1997). This approach provides 20 to 200 fold amplification of the target sequences. For comparing the expression between two different samples, Cy3/Cy5 labeling is widely adopted. One sample is labeled with Cy3 and the other sample is labeled with Cy5. The two labeled samples are co-hybridized to an array. The array is scanned at wavelengths specific for Cy3 and Cy5, and the relative abundance of each gene is derived from the signal intensity ratio of the two scans.

A very clever adaptation of the microarray technology for genomic DNA analysis is array Comparative Genomic Hybridization (array CGH), which has very promising clinical applications in cancer (for reviews, see Kashiwagi and Uchida, 2000; Albertson, 2003), preimplantation genetic diagnosis (PGD) (for reviews, see Wells and Levy, 2003) and potentially many other applications.

Microarrays are probably the most popular technique for high-throughput transcriptional profiling. In the early days, when there were inadequate and often erroneous genomic data and gene annotations, microarrays were criticized for creating a lot of dirty data (Mir, 2000; Wooster, 2000; Knight, 2001a and 2001b). Many of the criticisms were valid. However, with the completion of the human genome and many other genomes, it is only a matter of time before we will have highly accurate annotation for most (if not all) of the genes. Thus, it is possible to eliminate the previous errors in probe designs. The problem of cross-hybridization, although not completely resolvable, can be minimized by careful probe design.

Early success stories with microarrays include classification of cancers (Golub *et al.*, 1999), determination of therapeutic responses (Alizadeh *et al.*, 2000), exploring signaling pathways (Fambrough *et al.*, 1999; Harkin *et al.*, 1999; Lee *et al.*, 1999), gene function prediction based on various clustering algorithms (Cho *et al.*, 1998; Arbeitman *et al.*, 2002) and many others. The power of microarrays lies mostly in the simultaneous quantification of thousands to tens of thousands of genes or DNAs. Thus, it is possible to carry out

extensive comparisons, either for gene expression changes between a control group and a disease group, or among genes within the same sample. In a typical experiment where the transcriptional profiles of two sample groups (a control and a disease group) are obtained, the bad news and the good news are exactly the same (too) many gene hits are identified. It is important to design the microarray experiments carefully and with a sufficient sample size (Lee and Whitmore, 2002; Cui and Churchill, 2003). In addition and probably more importantly, it is also important to validate the identified gene targets in a larger (ideally independent) sample group with a more accurate technique such as real time PCR or real competitive PCR. Microarrays serve best as a screening platform to narrow down the potential gene targets (to this end it resembles the yeast two hybrid technique). However, the false positive rate for a quick and dirty microarray experiment is not well understood. Therefore, further validation is essential.

**Real-Time PCR** Polymerase chain reaction (PCR) is by far the most sensitive technique available for nucleic acid detection when appropriate amplification primers can be designed. The sensitivity of PCR is due to the exponential amplification nature of the procedure (Equation 1).

$$N_n = N_0 \times (1 + E)^n \quad (1)$$

where  $N_n$  is the number of DNA molecules after  $n$  cycles of PCR,  $N_0$  is the number of molecules before PCR, and  $E$  is the efficiency of amplification ( $0 < E < 1$ ; in an optimized assay,  $E$  is close to 1 before PCR reaches the plateau phase).

From equation 1, it is clear that  $E$  plays a dominant role in the amplification process. Even a minor change in  $E$  can result in dramatically different amplification kinetics. For example, with  $E$  of 1 versus 0.95 (a mere 5% difference in amplification efficiency), the difference in  $N_n$  will be 275% after 40 cycles.  $E$  is often dependent on the length and the sequence of the DNA to be amplified, how the primers are designed, the reaction buffer ( $Mg^{2+}$ , dNTP, etc) concentrations, and temperatures and times for each step of a PCR cycle. To make the situation even worse, even for the same reaction,  $E$  can change over the course of amplification due to loss of enzyme activity, accumulation of pyrophosphate molecules and consumption of reagents.  $E$  will approach zero as PCR reaches the plateau phase. For these reasons, it is very difficult, if not impossible, to directly quantify  $N_0$  by measuring  $N_n$  after some pre-determined number of PCR cycles. Competitive PCR and real time PCR are two techniques that can overcome some of these limitations.

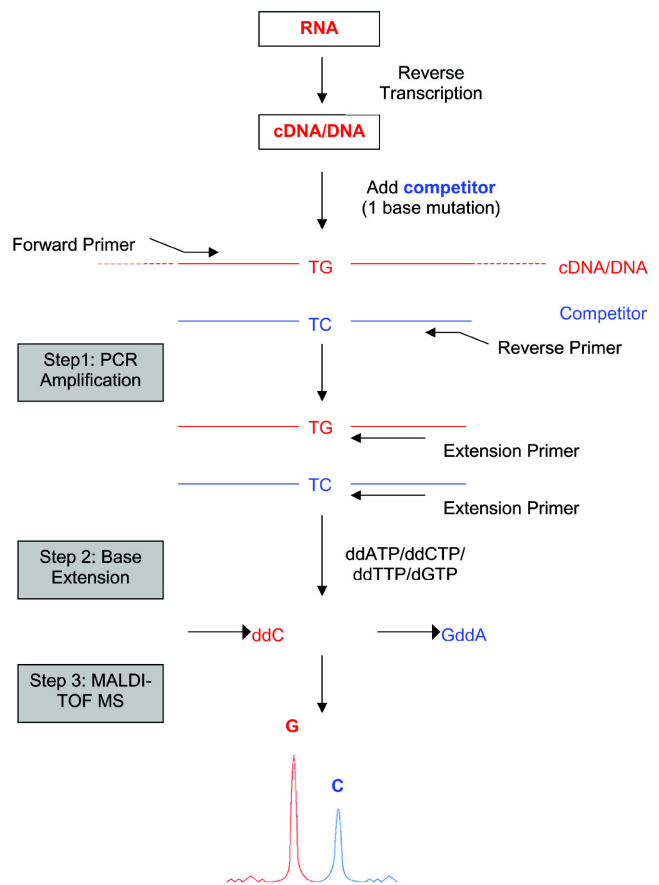
Real time PCR is a kinetics-based quantification technique. The amount of PCR product is measured at each PCR cycle for the entire PCR process. Except for the first few PCR cycles where the PCR product is too little to be accurately quantified, the  $E$  of each PCR reaction can be quantified. The correlation between the starting template concentration and the amplification kinetics is exploited for the quantification of

nucleic acids. We will use the Taqman™ system (AppliedBiosystems, Inc.) to illustrate how real time PCR works. Three oligonucleotides are used for quantifying each sequence. Two of them are the primers for the PCR amplification. The third oligonucleotide (the probe) is designed to hybridize specifically to the amplified sequence. It has a reporter fluorescent dye (R) at the 5'-end and a quencher dye (Q) at the 3'-end. When the probe is intact, it is non-fluorescent due to fluorescence quenching by Q when it is close to R. In the process of primer extension of PCR, the DNA polymerase displaces and cleaves the probe using its 5' to 3' exonuclease activity. The cleavage separates the dye and the quencher so that the dye now is fluorescent. The increase in fluorescence is thus directly correlated to the accumulation of the PCR product and can be used to monitor the PCR product in real time. Other popular real time PCR monitoring methods include molecular beacons (Tyagi and Kramer, 1996) and SYBR Green (Morrison *et al.*, 1998).

For nucleic acids quantification, a certain fluorescence level, typically 10 times the standard deviation of the average signal of the baseline fluorescence signal, is used to define the threshold cycle (Ct) (Gibson *et al.*, 1996). The Ct value, under ideal conditions, is inversely proportional to the logarithm of the template concentration (copy number). The PCR at Ct is always in the exponential phase; thus efficiently eliminating the PCR plateau phase problem.

Real time PCR is widely considered as the gold standard for nucleic acid quantification. It is commonly used for validation of microarray results. It is also used when the starting materials (cells, tissues, RNAs) are limited and high sensitivity is needed (Fink *et al.*, 1998), or when high accuracy is needed (Ginzinger *et al.*, 2000). However, the Taqman-based probes can be prohibitively expensive for a typical academic research laboratory, and PCR assay optimization can take some significant amount of time. Cheaper detection alternatives such as SYBR green may suffer in specificity and accuracy. Thus, in most cases, only a few genes are analyzed by real time PCR.

**Real competitive PCR** As discussed in the real time PCR section, PCR efficiency (E) can differ for the same set of PCR reactions due to the differences in buffer and enzyme concentrations and PCR cycling. Thus, in real time PCR, it is often necessary to make a supermix of PCR solutions and aliquot them into each individual reaction. Competitive PCR is another method that has a long history of use in nucleic acid quantification (Becker-Andre and Hahlbrock, 1989). A DNA standard with a synthetic point mutation to differentiate it from the gene of interest is spiked into the cDNA sample and co-amplified with the gene of interest. Since the gene and the DNA standard are virtually identical in sequence, they are amplified with the same kinetics throughout the PCR. The introduced mutation either creates or eliminates a restriction enzyme site, providing a way of distinguishing the standard and the gene. However, heteroduplex DNA formed by the



**Fig. 2.** The real competitive PCR approach for gene expression analysis. Total RNA is reverse transcribed with random hexamers. Then, a competing DNA oligonucleotide (typically 80 bases long) with one base difference from the gene of interest is added prior to PCR. A base extension reaction at the mutation site is carried out by adding a ThermoSequenase and 3 ddNTPs and 1 dNTP to produce two oligonucleotide products with different molecular weight. These two products are subsequently detected and quantified by MALDI-TOF MS with Allelotyping software (SEQUENOM, Inc.).

standard and the gene is resistant to enzymatic digestion, which complicates the analysis (Becker-Andre and Hahlbrock, 1989). Furthermore, the gel electrophoresis technique used to separate and quantify the two PCR products is labor-intensive and has a low dynamic range. Ion pair-reversed phase-high performance liquid chromatography (IP-RP-HPLC) was introduced for PCR product separation and quantification to improve the throughput (Hayward-Lester *et al.*, 1996).

Recently, matrix assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS) was adapted for analyzing gene expression (Ding and Cantor, 2003b). As illustrated in Fig. 2, this technique, dubbed as real competitive PCR (rc PCR), combines competitive PCR, single base extension and MALDI-TOF MS. After isolation of RNA and reverse transcription, cDNA is spiked with a synthetic

oligonucleotide (the competitor) with an identical sequence except one single base roughly in the middle of the sequence to the cDNA of interest. The competitor and the cDNA of interest are co-amplified with PCR. Excess dNTPs are removed by shrimp alkaline phosphatase treatment after PCR. Then, a base extension reaction is carried out with a base extension primer, a combination of ddNTPs and dNTPs and a ThermoSequenase. The base extension primer hybridizes right next to the mutation site and either one or two bases are added for the competitor and the cDNA, yielding two oligonucleotide products with different molecular weights. These two extension products are separated, and the ratio of their concentrations is quantified by MALDI-TOF MS.

Since the amount of competitor spiked in is known, the absolute concentration of the cDNA can be easily calculated. Thus, it is possible to do absolute gene expression analysis. Like other point mutation based competitor PCR techniques, the competitor and the cDNA are virtually identical in sequence and are amplified with the same kinetics. Using an internal competitor also makes the quantification not susceptible to well-to-well (or tube-to-tube) variations in PCR amplification, and it is potentially more accurate than kinetics based PCR quantification techniques. The quantification is not PCR cycle dependent (Ding and Cantor, 2003b). Furthermore, the real competitive PCR technique can use any desired single point mutation and the mutation does not need to introduce or eliminate a restriction enzyme site. The rcPCR technique also naturally solves the problem of heteroduplex DNA formation by the virtue of the base extension reaction. The linear amplification from the base extension reaction also increases the sensitivity, and single DNA molecule sensitivity can be readily achieved (Ding and Cantor, 2003a). In addition, the specificity is also improved since nonspecific PCR products are not likely to be a suitable template for the base extension reaction.

In terms of the number of different transcripts that can be analyzed, real time and real competitive PCR do not come close to the throughput of microarrays, SAGE or differential display. These two techniques are not applicable to fishing expeditions. However, these two techniques have substantially higher accuracy, sensitivity and reproducibility. In addition, they are more flexible about the number of samples that can be analyzed. Thus, they are most suitable when the number of genes is limited, and the sample size is large. The small set of selected genes can come from an initial high-throughput screening experiment by microarray, a previously established biological system where genes of interest are known, or a high-throughput mutation-based genetics study (Kruglyak, 1999).

**Differential display** Differential display (DD) is probably the only technique covered in this review that does not necessarily require some special, often expensive equipments (Liang and Pardee, 1992). It also does not rely on previous knowledge of the gene sequences. First, mRNAs from

different cells are reverse transcribed into cDNAs using three anchored oligo(dT) primers that differ from each other only at the base right next to the poly(dT) sequence. The three resulting cDNA preparations are further amplified by adding a set of second primers that are short (~13 bp) and arbitrary in sequence. Each combination of anchored oligo(dT) primers and short second primers will amplify about 50 to 100 different mRNAs. The PCR products are labeled with isotopes or fluorescent dyes and separated by denaturing polyacrylamide gel electrophoresis. The cDNA patterns are thus displayed on the gel. When two cDNA samples are displayed side by side, expression changes can be detected. Partial gene sequences can subsequently be cloned and sequenced. To cover the entire transcriptome, a few hundred different reactions are needed. To achieve this, fluorescent differential display (Ito *et al.*, 1994) and automation (Irie *et al.*, 2000) have been developed.

Since differential display does not require any special instrument, it is accessible to any reasonably equipped molecular biology laboratory. In a relatively high throughput and somewhat labor-intensive fashion, one can start a fishing expedition to identify genes that are differentially expressed among different samples/tissues/cell types. This is a quite popular technique in academic research laboratories, and it has been applied to many types of organisms including bacteria, yeast, flies, plants, and higher mammals (for reviews, see Ivanova and Ivanov, 2002; Liang, 2002; Liao and Freedman, 2002; Simon, 2002; Stein and Liang, 2002; Yamazaki and Saito, 2002). The components of the differential display technology (RT-PCR, DNA sequencing gel electrophoresis and cDNA cloning) are well established, and their advantages and disadvantages are well documented. Differential display protocols are also well documented (Liang and Pardee, 1997). Another advantage of differential display is that it does not rely on prior gene annotations. Thus, it is possible to identify novel genes. However, differential display probably is not the ideal platform for whole-genome transcriptional profiling, even with the higher throughput and more automated fluorescence-based platform.

**SAGE** Serial analysis of gene expression (SAGE) is a sequencing based technique for gene expression analysis (Velculescu *et al.*, 1995). Double-stranded cDNAs, synthesized using biotinylated oligo(dT) primers from an mRNA sample, are digested with a 4-base restriction enzyme (such as *Nla* III, which cuts on average every 256 bp). Only the 3'-ends of the cDNA sequences are recovered with streptavidin beads via biotin-streptavidin binding. These cDNA sequences are then divided into two pools, ligated to two different linkers (A and B) containing the recognition site for *Bsm* FI, a type IIS restriction enzyme that cleaves 13-14 bp downstream of the recognition site. A digestion by *Bsm* FI releases 14 bp fragments, each with a 10 bp gene-specific tag that are concatenated into long sequences and subsequently cloned into plasmids. Taking advantage of high-throughput

sequencers, 20,000 to 100,000 tags are sequenced. The frequencies of gene-specific tags are used to estimate the expression levels of the genes.

One advantage of SAGE is that it can provide an absolute quantification of gene expression. Thus, it is possible to combine results from different laboratories to construct a SAGE library (<http://www.ncbi.nlm.nih.gov/SAGE>). Another advantage is that novel transcripts can be discovered (Chen *et al.*, 2002). However, a gene must have a recognition site for the chosen enzyme (such as *Nla* III) to be included in the analysis. In addition, it is well documented that sequence features such as high GC content can affect SAGE analysis (Margulies and Innis, 2000). In addition, the 10-base tag sequences used in SAGE are not always unique. Using longer SAGE tags can provide more sequence information for more accurate gene identification (Chen *et al.*, 2003), but the analysis throughput is compromised. SAGE also has limited sensitivity related to the number of sequenced tags. Even with sequencing of 100,000 tags, a gene that is expressed on average less than one copy/cell cannot be accurately quantified (or even be robustly detected).

## A Comprehensive Overview of the Five Technologies

In this section, we will give a comprehensive overview of the various techniques. Again it is important to point out that the techniques are designed for different purposes, and it is neither fair nor appropriate to conclude which technique is the best. The comparison is intended to give the readers a sense of which technique(s) might be most appropriate for a specific experiment/project. We will analyze the accuracy, sensitivity, specificity, throughput and cost, as well as a few other issues. The main points are summarized in Table 1.

**Accuracy and reproducibility** Real time PCR and real competitive PCR are the two most accurate and reproducible techniques for DNA and RNA quantifications. Real time PCR has been widely accepted as the gold standard for accurate gene expression analysis. Although competitive PCR has been developed for a long time, real competitive PCR is a new comer in the field (the first paper describing the technique was published only 7 months ago at the time of the writing of this review). The use of an internal standard by real competitive PCR for each gene of interest can potentially offer higher accuracy by eliminating problems that arise from differential amplification kinetics in different reactions. A coefficient of variation (CV) of less than 10% (Ding and Cantor, 2003b) or even 4% (Sequenom Inc., unpublished data) can routinely be achieved. For real time PCR, a CV of 1-3% for the Ct value can also be achieved. However, since the mRNA concentration is roughly proportional to  $(1/2)^{Ct}$ , the CV for mRNA concentration can be substantially higher (for example, if Ct is 25 and the CV for Ct is 2%, the CV for gene

expression can be as high as 44%). Extremely high accuracy is important to detect small, yet biologically significant expression changes (Yan *et al.*, 2002).

For SAGE, the accuracy and reproducibility depend on several factors. As mentioned earlier, there are intrinsic problems associated with cloning and sequencing. In addition, it is increasingly harder to analyze genes with low expression levels. Statistical models (such as  $\chi^2$ -test, see Man *et al.*, 2000) are used for evaluating the statistical significance of the expression changes.

For microarrays, there has been much improved reproducibility due to better array production and data analysis. However, the intrinsic problems associated with cross-hybridization can yield dubious results for genes with low expression levels or very similar paralogs. A two-fold expression change is often set as the guideline for significant quantification. It is possible that smaller changes can be robustly detected for highly expression genes, but it is also possible that even larger than two-fold changes cannot be detected for low expression genes. Sophisticated algorithms have been developed for assessing the statistical significance of expression changes (Koopberg *et al.*, 2002; Cheadle *et al.*, 2003; Hyduke *et al.*, 2003; Leung and Cavalieri, 2003; Singhal *et al.*, 2003).

**Sensitivity and specificity** Sensitivity is an issue often somewhat overlooked in high-throughput analyses. The argument is that for many high-throughput analyses, it is often sufficient to obtain enough gene targets. Thus, missing a few thousand genes is not a problem. This was probably a valid (and practical) approach in early high-throughput studies. However, it is often true that, in a clinical setting, the amount of sample is often limited. Additionally, gene-gene interactions are showing increasing importance in studying gene function, but many regulatory genes are expressed at low levels.

Both real time PCR and competitive PCR are extremely sensitive. A few molecules of cDNA sequences are sufficient for detection and quantification (Ding and Cantor, 2003a; Palmer *et al.*, 2003). These techniques are the only options when analyzing low expression level genes from limited samples (such as cells selected by laser capture microdissection, Fink *et al.*, 1998).

Specificity is also very important since it provides a confidence level for the measurements. For short oligonucleotide-based microarrays, single base mismatch oligonucleotides are spotted next to the perfect match oligonucleotides to qualitatively (and somewhat quantitatively) estimate the effects of cross-hybridization. The specificity of low expression genes is likely to be lower since cross-hybridizations by homologous sequences can account for a larger percentage of the signal intensity. For SAGE, the specificity can be problematic when there are sequencing errors and an inability to map the tags to the transcriptome uniquely. Some of the real time PCR techniques (such as

**Table 1.** Overview of different quantification platforms

	Microarrays	Differential Display	SAGE	Real Time PCR	Real Competitive PCR
Requirement	Arrayer, array scanner	Common molecular biology equipment	High-throughput Sequencer	Thermocycler with automatic fluorescence detection	Nanodispenser, MALDI-TOF MS
Sensitivity	1:300,000 mRNA 0.2 µg mRNA (or 20 µg total RNA)	Varies, generally high	1: Total Tag Sequenced (Typically 50,000)	Highest, up to single-digit copy numbers (ng to pg total RNA)	Highest, up to single-digit copy numbers (ng to pg total RNA)
Specificity	Possible cross-hybridization	Further sequencing can give high specificity	Possible mis-assignment or no assignment of tags	Varies, can be very high with gene specific probes	Highest, single nucleotide specificity for known mutations
Throughput 1. Gene Number 2. Sample Number 3. Cost	1. Highest, can cover the whole human transcriptome 2. ~ ten samples/day 3. \$500-\$1000/chip	1. Medium 2. Low-medium 3. Low	1. High, thousands 2. Low-medium 3. medium-high	Balanced between 1 and 2. 100 genes from 10 samples or 10 genes from 100 samples 3. a few dollars/gene	Balanced between 1 and 2. 100 genes from 10 samples or 10 genes from 100 samples 3. \$2/gene
Accuracy/ Reproducibility	Accuracy dependent on expression levels. (2 fold changes); Reproducible in the same lab, less reproducible in different labs, different arrayers	Generally, larger expression changes are detected	Accuracy dependent on the number of total tags, Taq sequenced; Reproducible, but tag number dependent	High, 20-50% expression changes detectable; Highly reproducible, but may need assay optimization	High, 10-20% expression changes detectable; Highly reproducible
Gene sequences need to be known	Yes	No	No, but will have problem assigning the tags	Yes	Yes
Alternative Splicing Detection	Possible, with specially designed chips	Possible	Unlikely	Possible	Possible
Allelic Expression	Possible, with specially designed chips	Unlikely	Unlikely	Possible, need extensive assay optimization	Yes
Absolute quantification	Unlikely	Unlikely	Yes	Yes, with standard curve construction	Yes
Dynamic range	10 <sup>5</sup>	10-100	Hard to assess, but 10-100 fold possible	10 <sup>5</sup> -10 <sup>6</sup>	10 <sup>6</sup> , but 3 different competitor concentrations needed

Taqman) use a third gene-specific primer that is very likely to increase the specificity. The real competitive PCR technique also uses a third gene-specific primer for the base extension reaction. In addition, the molecular weights of the extension products (generally oligonucleotides of 18-25 bases long) are accurately ( $\pm$  a few Da) measured by MALDI-TOF MS. When the extension primer is designed next to a potential polymorphic site, real competitive PCR has the ability to

simultaneously quantify the expressions of the two different alleles of the same gene even though they differ by as few as one single base (Ding and Cantor, unpublished data). This can be especially advantageous for quantifying different subtypes of viruses.

**Throughput and cost** Microarray technology has the highest throughput in terms of gene numbers per experiment.

It can analyze the whole transcriptome from one sample on a single chip (such as Affymetrix Human Genome U133 Plus 2.0 Array). SAGE also has high throughput due to the high capacity of capillary sequencers. Neither real time PCR nor real competitive PCR has the same level of throughput in terms of gene numbers. However, these two techniques are more flexible in terms of sample numbers. For example, 20 genes from a few hundred different samples can be analyzed in one single day using real competitive PCR. All these four technologies require special and quite expensive instruments and can be expensive to run for an academic laboratory. They are more likely to be hosted by a central support laboratory so that the cost can be shared by many laboratories in a research institute. Differential display is the only technique in this review that does not necessarily require any expensive equipment and can be carried out in any reasonably equipped molecular biology laboratory.

**Future directions** We think differential display will continue to be widely used in academic laboratories for relatively small-scale identification of genes differentially expressed among different samples. SAGE will continue to be important in building a reference database for gene expression analysis. There will be more improvements in microarray probe designs, experiment designs and data analyses. Microarray will continue to serve as the main platform for high-throughput transcriptional profiling. We also expect real time PCR and real competitive PCR to play a significant role in fine-scale, more accurate and specific DNA and RNA quantification.

## References

- Albertson, D. G. (2003) Profiling breast cancer by array CGH. *Breast Cancer Res. Treat.* **78**, 289-298.
- Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., *et al.* (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503-511.
- Alwine, J. C., Kemp, D. J. and Stark, G. R. (1977) Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc. Natl. Acad. Sci. USA* **74**, 5350-5354.
- Arbeitman, M. N., Furlong, E. E., Imam, F., Johnson, E., Null, B. H., Baker, B. S., Krasnow, M. A., Scott, M. P., Davis, R. W. and White, K. P. (2002) Gene expression during the life cycle of *Drosophila melanogaster*. *Science* **297**, 2270-2275.
- Becker-Andre, M. and Hahlbrock, K. (1989) Absolute mRNA quantification using the polymerase chain reaction (PCR). A novel approach by a PCR aided transcript titration assay (PATTY). *Nucleic Acids Res.* **17**, 9437-9446.
- Cheadle, C., Vawter, M. P., Freed, W. J. and Becker, K. G. (2003) Analysis of microarray data using z score transformation. *J. Mol. Diagn.* **5**, 73-81.
- Chen, J., Sun, M., Lee, S., Zhou, G., Rowley, J. D. and Wang, S. M. (2002) Identifying novel transcripts and novel genes in the human genome by using novel SAGE tags. *Proc. Natl. Acad. Sci. USA* **99**, 12257-12262.
- Chen, J. J., Lee, S., Zhou, G., Rowley, J. D. and Wang, S. M. (2003) Generation of longer cDNA fragments from SAGE tags for gene identification. *Methods Mol. Biol.* **221**, 207-222.
- Cheung, V. G., Morley, M., Aguilar, F., Massimi, A., Kucherlapati, R. and Childs, G. (1999) Making and reading microarrays. *Nat. Genet.* **21**, 15-19.
- Cho, R. J., Campbell, M. J., Winzeler, E. A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T. G., Gabrielian, A. E., Landsman, D., Lockhart, D. J. and Davis, R. W. (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell.* **2**, 65-73.
- Cui, X. and Churchill, G. A. (2003) Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol.* **4**, 210.
- Ding, C. and Cantor, C. R. (2003a) Direct molecular haplotyping of long-range genomic DNA with MI-PCR. *Proc. Natl. Acad. Sci. USA* **100**, 7449-7453.
- Ding, C. and Cantor, C. R. (2003b) A high-throughput gene expression analysis technique using competitive PCR and matrix-assisted laser desorption ionization time-of-flight MS. *Proc. Natl. Acad. Sci. USA* **100**, 3059-3064.
- Emmert-Buck, M. R., Bonner, R. F., Smith, P. D., Chuaqui, R. F., Zhuang, Z., Goldstein, S. R., Weiss, R. A. and Liotta, L. A. (1996) Laser capture microdissection. *Science* **274**, 998-1001.
- Fambrough, D., McClure, K., Kazlauskas, A. and Lander, E. S. (1999) Diverse signaling pathways activated by growth factor receptors induce broadly overlapping, rather than independent, sets of genes. *Cell* **97**, 727-741.
- Fink, L., Seeger, W., Ermert, L., Hanze, J., Stahl, U., Grimminger, F., Kummer, W. and Bohle, R. M. (1998) Real-time quantitative RT-PCR after laser-assisted cell picking. *Nat. Med.* **4**, 1329-1333.
- Gibson, U. E., Heid, C. A. and Williams, P. M. (1996) A novel method for real time quantitative RT-PCR. *Genome Res.* **6**, 995-1001.
- Ginzinger, D. G., Godfrey, T. E., Nigro, J., Moore, D. H. 2nd., Suzuki, S., Pallavicini, M. G., Gray, J. W. and Jensen, R. H. (2000) Measurement of DNA copy number at microsatellite loci using quantitative PCR analysis. *Cancer Res.* **60**, 5405-5409.
- Goidin, D., Mamessier, A., Staquet, M. J., Schmitt, D. and Berthier-Vergnes, O. (2001) Ribosomal 18S RNA prevails over glyceraldehyde-3-phosphate dehydrogenase and beta-actin genes as internal standard for quantitative comparison of mRNA levels in invasive and noninvasive human melanoma cell subpopulations. *Anal. Biochem.* **295**, 17-21.
- Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D. and Lander, E. S. (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531-537.
- Harkin, D. P., Bean, J. M., Miklos, D., Song, Y. H., Truong, V. B., Englert, C., Christians, F. C., Ellisen, L. W., Maheswaran, S., Oliner, J. D. and Haber, D. A. (1999) Induction of GADD45 and JNK/SAPK-dependent apoptosis following inducible expression of BRCA1. *Cell* **97**, 575-586.
- Hayward-Lester, A., Oefner P. J. and Doris P. A. (1996) Rapid



- quantification of gene expression by competitive RT-PCR and ion-pair reversed-phase HPLC. *Biotechniques* **20**, 250-257.
- Herzenberg, L. A., Parks, D., Sahaf, B., Perez, O., Roederer, M. and Herzenberg, L. A. (2002) The history and future of the fluorescence activated cell sorter and flow cytometry: a view from Stanford. *Clin. Chem.* **48**, 1819-1827.
- Hyduke, D. R., Rohlin, L., Kao, K. C. and Liao, J. C. (2003) A software package for cDNA microarray data normalization and assessing confidence intervals. *Omics* **7**, 227-234.
- Irie, T., Oshida, T., Hasegawa, H., Matsuoka, Y., Li, T., Oya, Y., Tanaka, T., Tsujimoto, G. and Kambara, H. (2000) Automated DNA fragment collection by capillary array gel electrophoresis in search of differentially expressed genes. *Electrophoresis* **21**, 367-374.
- Ito, T., Kito, K., Adati, N., Mitsui, Y., Hagiwara, H. and Sakaki, Y. (1994) Fluorescent differential display: arbitrarily primed RT-PCR fingerprinting on an automated DNA sequencer. *FEBS Lett* **351**, 231-236.
- Ivanova, A. V. and Ivanov, S. V. (2002) Differential display analysis of gene expression in yeast. *Cell. Mol. Life Sci.* **59**, 1241-1245.
- Kashiwagi, H. and Uchida, K. (2000) Genome-wide profiling of gene amplification and deletion in cancer. *Hum. Cell* **13**, 135-141.
- Knight, J. (2001a) Geneticists' work in disarray as DNA-chip producer pulls the plug. *Nature* **414**, 135-136.
- Knight, J. (2001b) When the chips are down. *Nature* **410**, 860-861.
- Kooperberg, C., Sipione, S., LeBlanc, M., Strand, A. D., Cattaneo, E. and Olson, J. M. (2002) Evaluating test statistics to select interesting genes in microarray experiments. *Hum. Mol. Genet.* **11**, 2223-2232.
- Kruglyak, L. (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat. Genet.* **22**, 139-144.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.
- Lee, M. L. and Whitmore, G. A. (2002) Power and sample size for DNA microarray studies. *Stat. Med.* **21**, 3543-3570.
- Lee, S. B., Huang, K., Palmer, R., Truong, V. B., Herzlinger, D., Kolquist, K. A., Wong, J., Paulding, C., Yoon, S. K., Gerald, W., Oliner, J. D. and Haber, D. A. (1999) The Wilms tumor suppressor WT1 encodes a transcriptional activator of amphiregulin. *Cell* **98**, 663-673.
- Leung, Y. F. and Cavalieri, D. (2003) Fundamentals of cDNA microarray data analysis. *Trends Genet.* **19**, 649-659.
- Liang, P. (2002) A decade of differential display. *Biotechniques* **33**, 338-44, 346.
- Liang, P. and Pardee, A. B. (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257**, 967-971.
- Liang, P. and Pardee, A. B. (1997) *Differential display methods and protocols*. Humana Press, Totowa, NJ, USA.
- Liao, V. H. and Freedman, J. H. (2002) Differential display analysis of gene expression in invertebrates. *Cell. Mol. Life Sci.* **59**, 1256-1263.
- Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. and Brown, E. L. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* **14**, 1675-1680.
- Lockhart, D. J. and Winzler, E. A. (2000) Genomics, gene expression and DNA arrays. *Nature* **405**, 827-836.
- Man, M. Z., Wang, X. and Wang, Y. (2000) POWER\_SAGE: comparing statistical tests for SAGE experiments. *Bioinformatics* **16**, 953-959.
- Margulies, E. H. and Innis, J. W. (2000) eSAGE: managing and analysing data generated with serial analysis of gene expression (SAGE). *Bioinformatics* **16**, 650-651.
- Mikulowska-Mennis, A., Taylor, T. B., Vishnu, P., Michie, S. A., Raja, R., Horner, N. and Kunitake, S. T. (2002) High-quality RNA from cells isolated by laser capture microdissection. *Biotechniques* **33**, 176-179.
- Mir, K. U. (2000) The hypothesis is there is no hypothesis. The Microarray Meeting, Scottsdale, Arizona, USA, 22-25 September 1999. *Trends Genet.* **16**, 63-64.
- Morrison, T. B., Weis, J. J. and Wittwer, C. T. (1998) Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. *Biotechniques* **24**, 954-8, 960, 962.
- Nigro, J. M., Takahashi, M. A., Ginzinger, D. G., Law, M., Passe, S., Jenkins, R. B. and Aldape, K. (2001) Detection of 1p and 19q loss in oligodendroglioma by quantitative microsatellite analysis, a real-time quantitative polymerase chain reaction assay. *Am. J. Pathol.* **158**, 1253-1262.
- Nyberg-Hoffman, C., Shabram, P., Li, W., Giroux, D. and Aguilar-Cordova, E. (1997) Sensitivity and reproducibility in adenoviral infectious titer determination. *Nat. Med.* **3**, 808-811.
- Palmer, S., Wiegand, A. P., Maldarelli, F., Bazmi, H., Mican, J. M., Polis, M., Dewar, R. L., Planta, A., Liu, S., Metcalf, J. A., Mellors, J. W. and Coffin, J. M. (2003) New real-time reverse transcriptase-initiated PCR assay with single-copy sensitivity for human immunodeficiency virus type 1 RNA in plasma. *J. Clin. Microbiol.* **41**, 4531-4536.
- Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., Sherry, S., Mullikin, J. C., Mortimore, B. J., Willey, D. L., *et al.* (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928-933.
- Schena, M., Shalon, D., Davis, R. W. and Brown, P. O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467-470.
- Simon, H. G. (2002) Messenger RNA differential display strategies in birds and amphibians. *Cell. Mol. Life Sci.* **59**, 1264-1273.
- Singhal, S., Kyvernitis, C. G., Johnson, S. W., Kaiser, L. R., Liebman, M. N. and Albelda, S. M. (2003) Microarray data simulator for improved selection of differentially expressed genes. *Cancer Biol. Ther.* **2**, 383-391.
- Solanas, M., Moral, R. and Eschrich, E. (2001). Unsuitability of using ribosomal RNA as loading control for Northern blot analyses related to the imbalance between messenger and ribosomal RNA content in rat mammary tumors. *Anal. Biochem.* **288**, 99-102.
- Southern, E. M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503-517.
- Stein, S. and Liang, P. (2002) Differential display analysis of gene

- expression in mammals: a p53 story. *Cell. Mol. Life Sci.* **59**, 1274-1279.
- Tyagi, S. and Kramer, F. R. (1996) Molecular beacons: probes that fluoresce upon hybridization. *Nat. Biotechnol.* **14**, 303-308.
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. and Speleman, F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, RESEARCH0034.
- Velculescu, V. E., Zhang, L., Vogelstein, B. and Kinzler, K. W. (1995) Serial analysis of gene expression. *Science* **270**, 484-487.
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., *et al.* (2001) The sequence of the human genome. *Science* **291**, 1304-1351.
- Wells, D. and Levy, B. (2003) Cytogenetics in reproductive medicine: the contribution of comparative genomic hybridization (CGH). *Bioessays* **25**, 289-300.
- Wodicka, L., Dong, H., Mittmann, M., Ho, M. H. and Lockhart, D. J. (1997) Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat. Biotechnol.* **15**, 1359-1367.
- Wooster, R. (2000) Cancer classification with DNA microarrays is less more? *Trends Genet.* **16**, 327-329.
- Yamazaki, M. and Saito, K. (2002) Differential display analysis of gene expression in plants. *Cell. Mol. Life Sci.* **59**, 1246-1255.
- Yan, H., Dobbie, Z., Gruber, S. B., Markowitz, S., Romans, K., Giardiello, F. M., Kinzler, K. W. and Vogelstein, B. (2002) Small changes in expression affect predisposition to tumorigenesis. *Nat. Genet.* **30**, 25-26.