

Methods for quantitation of gene expression

Vigdis Nygaard, Eivind Hovig

Department of Tumor Biology, Institute for Cancer Research, Norwegian Radium Hospital, 0310 Oslo, Norway

TABLE OF CONTENTS

1. Abstract
2. Measuring gene products at the transcriptional level
3. A reductionist approach versus a high throughput approach
4. Common methods for measuring mRNA levels
 - 4.1. Subtractive hybridization
 - 4.2. Differential display
 - 4.3. RNase protection assay
 - 4.4. Quantitative real time RT-PCR
 - 4.5. Real competitive PCR
 - 4.6. Serial analysis of gene expression (SAGE)
 - 4.7. Massive parallel signature sequencing (MPSS)
 - 4.8. Microarray technology
 - 4.8.1. Principle of microarray technology and array platforms
 - 4.8.2. Target preparation and hybridization
 - 4.9. General technology comparison
5. Increasing technology applicability by resolving restrictions and limitations
 - 5.1. Global mRNA amplification
6. Concluding remarks and future perspective
7. References

1. ABSTRACT

Gene expression of protein encoding genes can be quantitatively measured at the transcriptional level by a number of low- to high-throughput methods. The sensitivity of each method is dependent on both the intrinsic properties of the respective technology and the absolute number of each mRNA molecule to be measured. For these reasons, the correlation of measurements between technological platforms may be variable. Due to the complexity of the transcriptome, the purpose of a gene expression study dictates the choice of method as each is connected to a set of advantages and disadvantages. Strategies such as global mRNA amplification of small samples, have been implemented to overcome previous limitations. However, stochastic events will limit quantitative measurements of any tool when in-put levels are extremely low. Due to the versatile nature of microarray technology, this method will likely persist as a highly applied tool to query the levels of non-coding transcripts, a new expansion in the field of gene expression analysis although possible advances of the technology may occur.

2. MEASURING GENE PRODUCTS AT THE TRANSCRIPTIONAL LEVEL

Nucleic acids in the form of DNA and RNA control heredity and cellular function. DNA and RNA can be studied in a qualitative manner, such as by investigating sequences, mutations and splicing, or quantitatively by interrogating how much of each sequence is present, such as measuring DNA amplifications/deletions or messenger RNA (mRNA) molecules expressed by gene activation. The focus of this presentation is the quantitative methodology and aspects of RNA-based gene expression analysis, with an emphasis on microarray technology.

The traditional view of the term gene expression comprises the two-step process of converting information from DNA in the nucleus into protein. The first step is transcription which takes place in the nucleus of a cell. During the transcription of a gene, DNA-dependent RNA polymerase II directs the synthesis of an mRNA molecule that is a complementary copy of the DNA template. This primary transcript, which has the same organization as the

genomic sequence (DNA), is generally modified by additions to either ends of the molecule (5' and 3' end), and spliced to remove sequences interrupting (introns) the protein coding sequences (exons). The processed mRNA is then transported into the cytoplasm. In the second step, the nucleotide sequence of the processed mRNA molecule is translated into amino acids comprising a protein. Gene expression can thus be measured both at an RNA level and at the protein level.

Gene expression analysis at the transcriptional level (step 1) implies measurements of the abundance or transcript copy numbers of specific RNA sequences. The correlation between mRNA levels and the presence of the respective proteins in the cell is not straightforward. Still, in many cases some estimates about the protein levels can be made from mRNA transcript abundances. RNA analysis is a widely used angle to study gene expression, as direct quantification of proteins is technically difficult at present (1,2), due mostly to the chemical complexity of proteins relative to RNA. The experimental conditions to measure mRNA abundance are nearly the same for all genes compared to the large range of optimal conditions that are specific for each protein molecule, rendering the establishment of high throughput protein assays difficult (1).

The central event in transcription is the RNA polymerase-catalyzed copying of the sequence of the DNA template strand into complementary RNA transcripts. The initiation of gene transcription from a gene template is the result of activation by multiple factors that increase the relative ability of its promoter to be recognized and bind polymerase. The transcription cascade that follows includes promoter clearance, mRNA transcript synthesis, elongation, processing and maturation. The events in the cascade are regulated by the interplay of a multitude of factors turning the transcription machinery on or off (3). Hence, the steady-state level of specific mRNAs in a cell that we measure with analysis tools, depends on the balance between the efficiency of these events, stochastic processes exerted on the involved components, in addition to the stability of the synthesized mRNA transcript. The mRNA measure does not however, directly reveal the intensity of these processes.

The estimated number of protein encoding genes in the human genome has in recent years been reduced from 100 000 to between 20 000 and 25 000 unique genes (4,5). However, the number of transcribed unique mRNA sequences exceeds the total gene number as demonstrated by the vast numbers of expressed sequence tags (ESTs) that have been reported and submitted to databases. The discrepancy is most likely due to the existence of numerous splice alternatives of protein encoding transcripts, alternative sites for transcription start, alternative polyadenylation sites and the presence of transcripts from uncharacterized, regulatory genes not encoding proteins (e.g. Xist), in addition to contamination of unspliced, immature mRNA (6). According to the European Bioinformatics Institute (EBI) website, of the 16293 genes with confirmed intron/exon features, 13572 genes had at

least more than one transcript variant (7). The use of alternative transcription start sites and/or alternative polyadenylation sites often accompanies alternative splicing (8). Results from two studies showed that proportion of genes with alternative polyadenylation sites was estimated to 29% and 54 %, respectively (9,10). The accumulation of noncoding transcript in EST libraries, often singleton clones, have been erroneously interpreted as contamination or noise (11). However, more recent comprehensive studies have shown that transcription is not limited to protein encoding genes but also includes an abundant group of non coding transcripts, many of which also harbour polyadenylation sites (12-14).

At a given point in time, a human cell expresses a fraction of the 20,000 to 25,000 protein-encoding genes carried by the DNA in the nucleus. This fraction, commonly referred to as the transcriptome, is highly dynamic and changes rapidly in response to cellular events such as cell cycle state, or by perturbations through inducible stimuli (15,16). The concept transcriptome can be expanded to include all elements transcribed at a certain time point, not only protein encoding genes. The scale of non-coding transcripts has just recently begun to be realized, adding new dimensions to the transcriptome. RNA-based gene expression analyses are used to study questions such as which genes (or elements) are transcribed, when and where (which tissues) are genes expressed, how many are expressed ubiquitously and identification of differences in gene expression. This knowledge is important for elucidating the complex relation between activation and de-activation of genes during physiological (e.g. cell cycle, differentiation, development, stimuli and response) as well as pathological processes (e.g. cause of disease, onset, consequences, and progression). In the context of human health management, gene expression analysis may aid in the discovery of which genes may be potential targets for intervention in a therapeutic setting, and to elucidate how drugs and drug candidates work.

The textbook description of total RNA composition in a eukaryotic cell lists ribosomal RNA (rRNA), transfer RNA (tRNA) and mRNA. A more up to date description of total RNA would divide RNA into two classes, coding RNA (mRNA) molecules which are translated into proteins, and non-coding RNAs, which are functional as RNA molecules rather than proteins. The non-coding RNAs (ncRNAs), include rRNA and tRNA, but also a number of smaller ncRNAs such as microRNA, small nuclear RNA (snRNA), small nucleolar RNA (snoRNA) and small interfering RNA (siRNA). Current studies are revealing increasingly important functions, including key regulators of transcription regulation, clearly warranting further analysis of these RNA families (17). However, going back to the traditional view, RNA-based gene expression analysis would be focused on protein encoding mRNAs, although the methodology applied could in fact be screening non-coding transcripts as well, though undetected or not interpreted by the investigator. For simplicity, this traditional train of thought throughout the text will be maintained for the sake of simplicity, and the implications

of previously unrecognized transcriptional activity with respect to gene expression analysis will be commented towards the end.

The fraction of mRNA in total RNA is in the range of 1-3%. The mRNA of a typical somatic cell is traditionally divided into three frequency classes, high, medium and low expressing genes (18). The experimental evidence for this view is, however, limited. Although the true frequency is unknown, results from expression assay studies indicate that the distribution is likely to differ between cell types (19). Hybridization kinetic studies have indicated that the distinction of three abundance classes is one of several possible categorizations (20). It has also been shown that the number of abundance classes within a type of cell may vary between stimulated and unstimulated groups, transcription being heavily induced in the stimulated cells (20). Regardless of the discrete or continuous copy number distribution, it is apparent that transcript abundances do not follow a normal distribution, but is skewed, having a heavy tail of low expressing genes. There are about 12 000 different transcripts per cell, and over 90% of these are represented by the low abundance class, having 1-15 copies of the transcript per cell or even significantly less than 1 copy per cell (21). According to high throughput gene expression analysis, more than 83% of the collective transcripts were present as low as one copy per cell (22). A transcriptome analysis of colon cancer cells showed that the most highly expressed genes ($n=623$) accounted for almost half of the mRNA content (22).

An important aspect of the gene expression measurements obtained by analytical tools is that the values represent the average transcript levels per gene from all the cells in the sample queried. Whether the samples are relatively homogeneous or heterogeneous is therefore a relevant consideration when addressing the results. The more pure the cell population is, the easier it is to assign distinctive gene expression patterns to specific cells, and not have results confounded with measurements from contaminating cells (23).

Considering the high levels of biological complexity of eukaryotic transcriptomes, an ideal raw data set from a gene expression tool would capture the frequencies of all transcribed protein coding- and non-coding RNAs, as well as transcript modifications such as splicing, and sequence. However, even with today's technology, fulfilling this list of demands entirely is a challenging quest. The performance of each of the current tools is reliant on the intrinsic technological properties. Common for all, is that assay sensitivity is a major determining factor with respect to quantitative reliability and is dependent on both the inherent technology and the absolute number of each specific mRNA molecule to be measured. Collectively, typical factors affecting quantitative measurements are small samples, unreliable detection of low expressing genes, experimental errors, ambiguities in the identification of many transcripts and cross-hybridization.

3. A REDUCTIONIST APPROACH VERSUS A HIGH THROUGHPUT APPROACH

Molecular biology has traditionally taken a reductionist approach to biological questions, focusing on a single gene, protein or a single reaction at a time, often using simple *in vitro* systems. This approach is hypothesis driven and assumes that the effects at the organism level can be reduced to causes at the gene level. Prior knowledge is required to some extent. For RNA expression studies, the classical reductionist method is the nucleic acid hybridization procedure known as Northern blotting (24). In Northern blotting, a labeled template of the gene of interest is hybridized to a number of RNA samples that are first separated according to molecular weight by electrophoresis and then transferred onto a nylon membrane. Quantification is typically done by normalizing the radioactive signals from the hybridized probe within a sample against the baseline of a monitored housekeeping gene detected across all the samples on the same blot, but in a separate hybridization event. This implies the assumption that housekeeping genes are transcribed at relatively equal steady-state levels in different cell populations. In this way, Northern blotting allows one gene to be inspected at a time in a limited number of samples. The technique is time and sample material consuming, and accurate quantification of mRNA levels is limited.

The reductionist approach has more or less been replaced by the popular integrative approach. The transition from a reductionist view to integrative biology has been possible due to available sequence data obtained by successful genome sequencing efforts, medium/high throughput gene expression technologies and computer technology. The aim of integrative biology in RNA analysis is to provide insight into global gene expression in order to understand complex biological systems. Genes do not operate alone, but in complex patterns, therefore characterization of more than one gene at a time increases the chance of observing the interplay between genes. As demonstrated numerous times in the literature, these integrative approaches do not exclude a reductionist angle when mining the results. In other words, although high throughput data are generated, attention is often focused on a few selected genes while discarding the remaining data.

A wide range of technologies exist for the analysis of gene expression at the level of mRNA. The following presentation of mainstream low and high throughput techniques includes subtractive hybridization (25), differential display (26), RNase protection assay (27), quantitative real time RT-PCR (qRT-PCR) (28), real competitive PCR (29), serial analysis of gene expression (SAGE) (30), massive parallel signature sequencing (MPSS) (31) and one of the most popular techniques, microarray hybridization (32). Figure 1 depicts the chronological introduction of these methods along the timeline. Subtractive hybridization and differential display represent solely analysis of differential gene expression (DGE) yielding only rough quantitative estimates, thus appearing relatively archaic compared to alternative techniques presented on the timeline (Figure 1). They are,

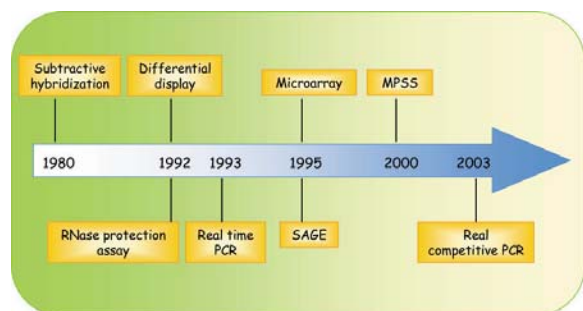


Figure 1. The chronological order of publications introducing novel gene expression technologies.

however, the progenitors of a number of more refined DGE strategies of which a few are mentioned in a following section. RNase protection assay, qRT-PCR and real competitive PCR generally query a single selected gene of interest per experiment. The sensitivity that these techniques offer have encouraged attempts to alter the throughput level resulting in multiprobe RNase protection assays (*e.g.* BD Riboquant™ kits) and at least one low density assay based on qRT-PCR (33). The key advantage of this latter assay is the ability to assess gene expression abundance in multiple genes (up to 384 genes) in a single sample. The remaining three technologies query a vast number of genes in a single experiment set-up. The results from gene transcript measurements are commonly reported within the categories of identification, classification and particularly, differential expression through side-by-side comparison of mRNA from different samples.

4. COMMON METHODS FOR MEASURING MRNA LEVELS

Due to the nature of RNA transcripts (unstable and easily forms secondary structures) and the state of today's technology, mRNA quantification methods involve multistep procedures. For quantification of mRNA levels it is important to point out that each step in the respective procedures can contribute to the introduction of variability and thus potentially be bottlenecks for the throughput of reproducible and reliable mRNA measurements. Quality control checkpoints throughout the procedures are advisable.

The first two steps are common for all the methods to be discussed. The third step is also common with the exception of the RNase protection assay. The first step is the sampling of cells/tissue to be investigated. Ideally this sample should be as homogenous as possible. In the case of a mixed population, the transcript amount from the cells of interest is difficult to decipher as the measurement is based on an average value of all the cells queried (23). The availability of material to be analyzed may be a limiting factor restricting the use of certain technologies. One option is to pool RNA from individual samples. Although it may be practical and reduce costs, discussions of validity are prevalent typically within the microarray community. The gene expression averaging across pooled samples reduces the chance to observe

sample variability and weak gene expression patterns may be washed out, but substantive differences may be easier to detect (34). Effects and guidelines for RNA pooling strategies have been presented in light of microarray experimental designs (35,36). Further in the sampling process, time of sample material harvest may be an important consideration in order to detect relevant information about response or progression of originating events. The handling of the biological material should strive towards minimal disturbance that could induce changes in gene expression. During the second step, RNA purification, it is highly recommended to work in an RNase-free environment as to avoid RNA degradation and reduction of sample quality, as RNA is extremely delicate once removed from the cell. For gene expression analysis, it is generally assumed that the isolated RNA represents all species of mRNAs in the same proportions as in the original cells, although this is not necessarily true and difficult to verify. Hence, uncertainties are linked to the current analysis tools claiming absolute quantification of purified mRNA transcripts, as the assumption may not be entirely fulfilled. The third step is also common for the most popular techniques, and constitutes the conversion of mRNA into cDNA by reverse transcription (Figure 2). There are several types of reverse transcriptase commercially available for this purpose, and enzymatic efficiency is an asset. Efficiency is not only a function of the characteristics of the reverse enzyme applied, but also template abundance. Efficiency is significantly lower for rare mRNA templates (37). It is common practice to initiate the reaction with oligo dT primers, random hexamers or gene specific primers that anneal to the templates. The latter type of primers reduce background priming due to their transcript specificity, while oligo dT and hexamers maximize the number of mRNA molecules converted into cDNA. After reverse transcription, the procedures linked to the different technologies deviate in strategies.

4.1. Subtractive hybridization

Subtractive hybridization represents an early technique with the aim of identifying differentially expressed genes (25). The method enriches the cDNA population of interest (tester) with differentially expressed gene transcripts, by removing commonly expressed genes through cDNA hybridization to mRNA of the sample being compared (driver). The transcripts that are unique or overexpressed in the sample of interest remain as single stranded cDNAs and can be separated from cDNA/mRNA duplexes by one of several methods, such as gel display. The subtracted cDNA molecules are then further characterized by retrieval, sequencing and identification. As an mRNA quantification method, it is relatively coarse, in addition to being inefficient for obtaining low abundance genes. However, the method is relatively simple, inexpensive and can be performed in most averagely equipped molecular biology laboratories. Preliminary knowledge of sequence is not necessary with this technique. To reduce the amount of RNA needed and the number of false positives, while augmenting the inclusion of low expressing genes, PCR-based subtractive hybridization strategies have been developed, including representational difference analysis (RDA) (38), and

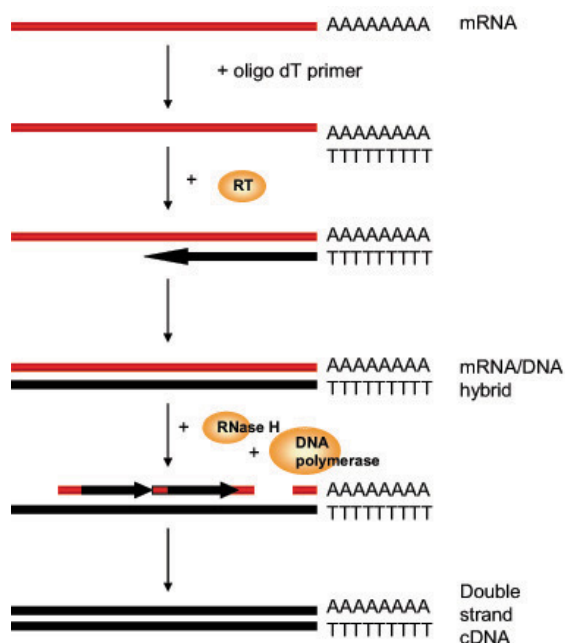


Figure 2. Scheme of the conversion of mRNA transcripts into double stranded cDNA templates. The first strand is primed by oligo dT primers and synthesized by a reverse transcriptase enzyme. The second strand is generated by partial digestion of the mRNA molecule with RNase H, leaving small fragments used to prime the synthesis of second strand by a DNA polymerase.

subtractive suppression hybridization (SSH) (39). Further modifications to RDA aimed to increase the representation of rare transcripts differentially expressed have also been described (40), and adds to the list of subtractive hybridization protocol variants.

4.2. Differential display

The general scheme of differential display involves conversion of mRNA into cDNA using anchored oligo dT primers followed by a combination of short random primers so that batches of the transcriptome are amplified in different subsets (26). These subsets of cDNA are further amplified and labeled with either isotopes or fluorescent dyes by PCR using a set of second primers that are short and random in sequence. The second primers are designed so that each will recognize 50-100 mRNAs under a given PCR condition. The PCR products are then displayed on denaturing polyacrylamide gels. Side-by-side comparisons of such complementary patterns from two or several cell samples may reveal differences in gene expression through band intensities. Differentially expressed cDNA can be retrieved and sequenced for further molecular characterization. The method does not require any specialized equipment other than that found in a reasonably well-equipped molecular laboratory. Further, the method does not rely on prior gene annotations. The major difference between this technique and subtractive hybridization is that there is no enrichment and selection step of differentially expressed transcripts. In subtractive hybridization, two samples (or two groups) are

compared, while in differential display a moderate number of samples can be compared in parallel. The low ability to identify rare transcripts, the laborious downstream work once differential bands are detected and specifically the high rate of false positives, presumably due to the suboptimal primer designs, are clear disadvantages of the original technique (41). Attempts to replace selection of mRNA by random primers with restriction enzyme digestions in order to reduce the number of false positives, have led to the development of close to a dozen restriction enzyme-based procedures such as total gene expression analysis (TOGA) (42). Alternative refinements of the technique, *e.g.* long-distance differential display-PCR, have been developed with the aim of increasing the sequence information obtained with the original protocol (43).

4.3. RNase protection assay

The RNase protection assay does not require the formation of cDNA (27). However, it does require prior knowledge of the gene of interest in order to synthesize a labeled antisense RNA probe complementary to the target RNA. This labeled probe is hybridized to the RNA sample to be examined. The sample is then treated with single-strand-specific RNase that degrades unhybridized probe and non-target sample RNA. The resistant probe:target hybrids are separated on a denaturing polyacrylamide gel for detection and quantification. The amount of probe is directly proportional to the amount of target mRNA in the sample. If an internal control probe or synthetic sense strand is included in the experiment, the assay can be used for relative and absolute quantification, respectively. Although sensitivity with respect to number of mRNA copies detectable is greater than for the classical Northern blotting, it is limited compared to quantitative RT-PCR (44). There is no need to invest in expensive equipment to follow this protocol. Several RNA targets may be assayed simultaneously as long as the protected fragments are of significantly different molecular weights, so that they can clearly be separated on the gel. A number of commercially available multiprobe RNase protection assays exist.

4.4. Quantitative real time RT-PCR

Real time RT-PCR is a kinetic-based quantification technique (28). The conversion from mRNA into cDNA is followed by PCR amplification. The products generated by gene specific PCR amplification emit fluorescent signals, which are measured during each PCR cycle and are directly proportional to the amount of template in the initial sample. It is only during the exponential phase of the PCR reaction that it is possible to use the fluorescent signal values to calculate and determine the initial amount of template. This method is a close-tube system and requires a method of detecting the accumulating PCR products, and recording the results during each PCR cycle in real time. Several fluorescent based methods exist to generate a signal in real time PCR experiments. The most commonly used are SYBR Green intercalating dyes, Taqman hydrolysis probes, molecular beacons and dual hybridization probes (45). SYBR green offers the least expensive method, as there is no need for target-specific fluorescent probes. During the polymerization step, more and more SYBR green dye molecules bind to the newly

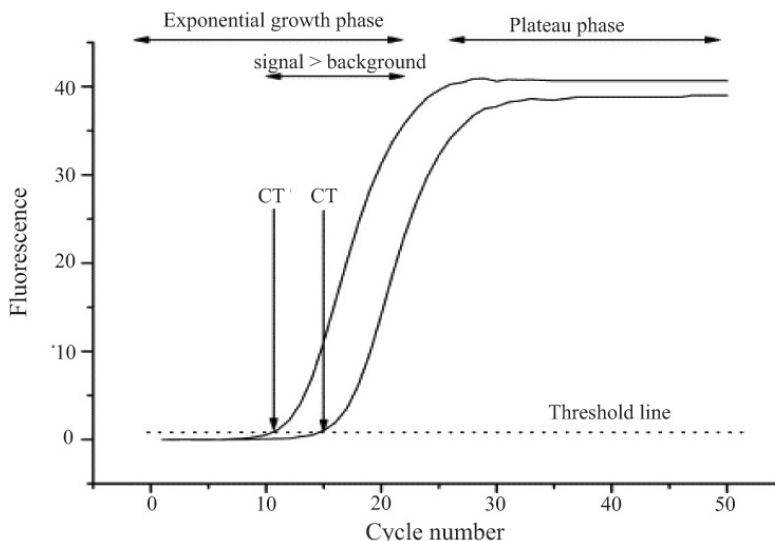


Figure 3. Real time detection of PCR products from two samples, that vary in initial number of template copies. A threshold is set sufficiently above background and the number of cycles necessary to reach threshold (CT), is registered. The figure was adapted from Kubista *et al.*, (56).

synthesized DNA, and the increase in fluorescence can be monitored in real time. Taqman, molecular beacon and dual hybridization probes are sequence specific and designed to hybridize to the PCR template between the first set of primers, conveying further specificity to the assay. When hybridized, the quenched signal of these sequence specific primers are released and detected by the apparatus in real time in the case of Taqman and molecular beacon probes. Dual hybridization probes rely on the head-to-tail annealing of one donor fluorescein probe and one acceptor fluorescein probe, bringing the two dyes in close proximity to one another, resulting in an energy transfer and emission of light.

In the real time synthesis of PCR products, fluorescent signal accumulates as the PCR cycle number increase, as a function of product increase. A threshold is calculated as a function of the background, typically as ten times the standard deviation of the background/baseline fluorescent signal (46). The number of PCR cycles necessary to generate sufficient signal to reach this threshold, is defined as the cycle threshold or Ct for a sample (Figure 3). The Ct values of different samples are used to calculate the abundance of transcript template for each sample. (46).

There are principally two types of quantitative RT-PCR set ups: relative quantification and standard curve absolute quantification. Relative quantification is a commonly used method whereby the expression level of the gene of interest is compared with a control gene (*e.g.* a housekeeping gene). The relative difference in initial template is calculated from the difference in cycle number needed to reach the Ct of the gene of interest compared to the control gene and is expressed as a ratio value. The application of relative quantification infers the assumption that the control gene does not vary in expression level amongst the samples interrogated. Standard curve

quantification relates the PCR signal to input copy number using a calibration curve. An *in vitro* transcribed external RNA or a DNA template can be used in constructing a standard curve to calculate copy numbers for absolute quantification. Known amounts of external molecules, ranging from high to low copy numbers, are assayed and the Ct versus copy number of the standard is plotted. The accuracy of determining absolute transcript numbers is directly dependent on the precision of the quantification of the RNA (DNA) standard. A disadvantage with both relative and absolute quantification is that this strategy assumes that the target transcript and control/standard transcript amplify with similar efficiency. This is not necessarily true and may potentially confound data results (47).

In an earlier paragraph the existence of a higher throughput level of qRT-PCR exemplified by low density arrays was mentioned. It should be clarified that these microfluidic chips split the sample into as many as 384 wells, each well interrogating one gene, respectively. The main problem amplifying multiple specific targets in a single reaction tube/well is the limited availability of multiple fluorescent reporters that can be distinguished from each other in a multiplex assay (45).

4.5. Real competitive PCR

Real competitive PCR combines competitive PCR, single base extension and matrix assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS) (29). The novelty of this procedure was the introduction of MALDI-TOF MS since the conventional competitive PCR, described by Becker-Andre and Hahlbrock (48), had disadvantages with respect to resolving the PCR products and thus quantification. MALDI-TOF MS produces mass spectra of DNA products at high resolutions, allowing accurate determination of the

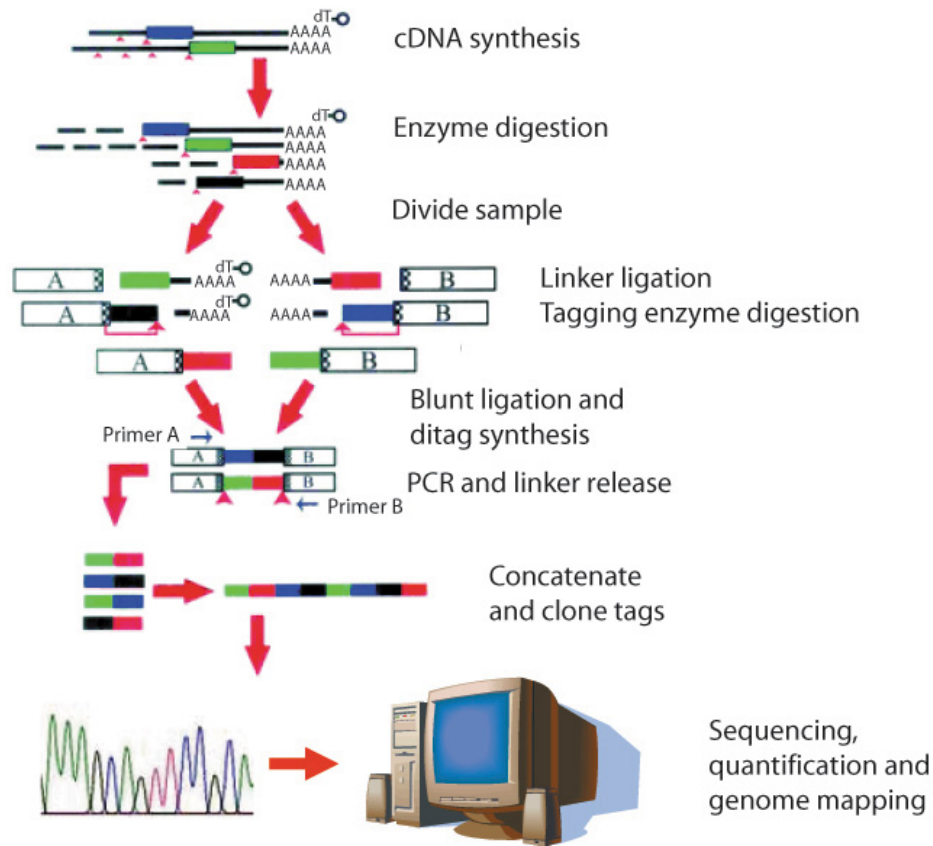


Figure 4. Scheme of the conversion of mRNA transcripts into double stranded cDNA templates. First strand is primed by oligo dT primers and synthesized by a reverse transcriptase enzyme. The second strand is generated by partial digestion of the mRNA molecule with RNase H, leaving small fragments used to prime the synthesis of second strand by a DNA polymerase.

bases in addition to quantitative information (49,50). In the improved real competitive PCR method, the cDNA synthesized from mRNA during reverse transcription is spiked with a synthetic oligonucleotide, termed the competitor as in the original protocol. This competitor template is identical to the transcript of interest with the exception of a single base in the middle of the template (mutation site). The competitor and transcript of interest are co-amplified in a PCR reaction. Following PCR, a base extension reaction is performed with a base extension primer. In this reaction, the base extension primer is designed as to hybridize right next to the mutation site and either one of two bases are added for the competitor and the cDNA of interest, yielding two products with different molecular weights. These two products are separated and the ratio of their concentrations is quantified by MALDI-TOF MS, whereby the amount of competitor spiked in, is a known factor. Thus it is possible to do absolute gene transcript quantification. The use of an internal competitor reduces the tube-to-tube variability in PCR amplification and the use of the specific base extension primers reduces the contamination of non-specific PCR products. The procedure requires specific instruments for mass spectrometry but has the potentials for increased throughput by multiplexing and/or automation (29).

4.6. Serial analysis of gene expression (SAGE)

In 1995, Velculescu *et al.* (30), presented a sequenced based technique for quantifying gene expression in a direct and high throughput manner called serial analysis of gene expression (SAGE). SAGE is mainly based on two principles, creation of short sequence tags from each transcript in the sample, and concatenation of these tags to allow efficient sequencing. The sequence information can then be mapped to the genome sequence and the quantification of specific transcripts is based on the frequency of occurrence. The short sequence stretch can be as minimal as 9 bp, assuming random nucleotide distribution throughout the genome. A 9–10 bp nucleotide tag can distinguish 4^9 (262 144) and 4^{10} (1 048 576) transcripts, and hence the transcript population of a cell population can theoretically be represented, assuming 300 000 transcripts per cell (21). The first step in the SAGE procedure is to synthesize double stranded

cDNA from mRNA using oligo dT primers bound to magnetic beads. The cDNA is then cleaved by a 4-base restriction enzyme (such as *NlaIII*), which cuts on average every 256 bp (Figure 4). Only the 3' ends are recovered by positive selection of cDNA bound to the magnetic beads. The bound cDNA is then split into two

pools and two different linkers (A and B) are ligated to the (*NlaIII*) cohesive termini so that each pool is defined by one linker, respectively. The linkers contain a recognition site for a tagging enzyme that cleaves 13-14 bp downstream of the recognition site. Cleavage by this enzyme creates a unique oligonucleotide known as the SAGE tag. The released SAGE tags from the oligo dT beads are recovered, blunted and ligated to each other to give rise to ditags with a tail-to-tail orientation. The ditags are then PCR amplified, released from the linkers by *NlaIII* digestion, gel purified, serially ligated to obtain concatemers, cloned and sequenced using an automated sequencer.

The coverage and sensitivity of SAGE is dependent upon the number of tags sequenced (51,52). A SAGE library typically holds 50 000 tags sampled from the entire transcriptome of the sample assayed. The sampling is unbiased, but the reflection of transcripts in the original sample may be distorted as only a small fraction of the total transcriptome is assessed. Evidently, the method is relatively laborious and requires a substantial sequencing effort, limiting the sample throughput number. Due to the limited sequence information contained in the tags, adequate gene transcript assignments require extensive bioinformatics support and sequence information. SAGE represents an open system with full transcriptome coverage and (53). many unique tags have been identified by SAGE (54). However, one problem with SAGE is that several tags have no match to known sequences in the databases (51). In this case, one must also not exclude the possibility of nucleotide misincorporation during the amplification or sequencing errors. Another problem is that tags find several matches in the databases, thus making correct gene assignment ambiguous. This has been addressed in attempts to increase the specificity by prolongation of the tags by various methods such as LongSage (55). Some gene transcripts are lost from the library to due the lack of recognition site for the restriction enzyme used to generate the tags.

Although SAGE does have some limitations, the method provides absolute quantification of gene expression levels in the sampled material, and thus allows direct comparison between genes within the sample and between different samples. The NIH Cancer Genome Project (CGAP) maintains a SAGE database for various normal and cancerous tissue and cell lines, which can be accessed through a web interface known as SAGE Genie.

4.7. Massive parallel signature sequencing (MPSS)

Massive parallel signature sequencing (MPSS) is a sequenced based approach that can be used to identify and quantify mRNA transcripts present in a sample similar to SAGE but the biochemical manipulation and sequencing approach differ substantially (31). MPSS allows mRNA transcripts to be identified through the generation of a 17-20 bp signature sequence adjacent to the 3'-end of the 3'-most site of the designated restriction enzyme (commonly *Sau3A* or *Dpn II*). Each signature sequence is cloned onto one of a million microbeads. The technique ensures that only one type of DNA sequence is on a microbead. So if

there are 50 copies of a specific transcript in the biological sample, these transcripts will be captured onto 50 different microbeads, each bead holding roughly 100 000 amplified copies of the specific signature. The microbeads are then arrayed in a flow cell for sequencing and quantification. The sequence signatures are deciphered by the parallel identification of four bases by hybridization to fluorescently labeled encoders (Figure 5). Each of the encoders has a unique label which is detected after hybridization by taking an image of the microbead array. The next step is to cleave and remove that set of four bases and reveal the next four bases for a new round of hybridization to encoders and image acquisition. The raw output is a list of 17-20 bp signature sequences, that can be annotated to the human genome for gene identification. The longer tag sequence confers a higher specificity than the classical SAGE tag of 9-10 bp. The level of unique gene expression is represented by the count of transcripts present per million molecules, similar to SAGE output. A significant advantage is the larger library size compared with SAGE. An MPSS library typically holds 1 million signature tags, which is roughly 20 times the size of a SAGE library. Some of the disadvantages related to SAGE apply to MPSS as well, such as loss of certain transcripts due to lack of restriction enzyme recognition site and ambiguity in tag annotation. The high sensitivity and absolute gene expression certainly favors MPSS. However, the technology is only available through Lynxgen Therapeutics, Inc. (now Solexa Inc).

4.8. Microarray technology

As with SAGE, we reached the decade mark in 2005 for the microarray technology. The advent of this technology received tremendous interest and was rapidly accepted as one of the most promising approaches that molecular biology had to offer. A microarray is by definition a large collection of gene specific DNA fragments, of known sequences aligned in an orderly fashion on a solid surface to which labeled samples are hybridized. Expression levels are indirectly measured when using arrays, by means of quantitative detection of fluorescent dye or other signaling molecules recorded by a scanning device. With microarrays containing sequences representative of all human genes, the highly appealing ultimate goal was to determine the expression level of every RNA species transcribed in a cell or tissue in a single experiment.

4.8.1. Principle of microarray technology and array platforms

The principles behind cDNA microarray technology presented by Brown and co-workers in 1995 (32), were in fact extensions of the standard nucleic acid hybridization methods used on solid surfaces such as Northern blots. In the preceding moderate-throughput techniques, dot blots and macroarrays, nucleic acids deposited in an orderly fashion on porous nylon membranes, allowed the monitoring of gene expression of multiple genes (56-58). Detection was performed by radioactive labeling of the hybridized material. Development of high throughput microarray technology took advantage of achievements within recent assembly of

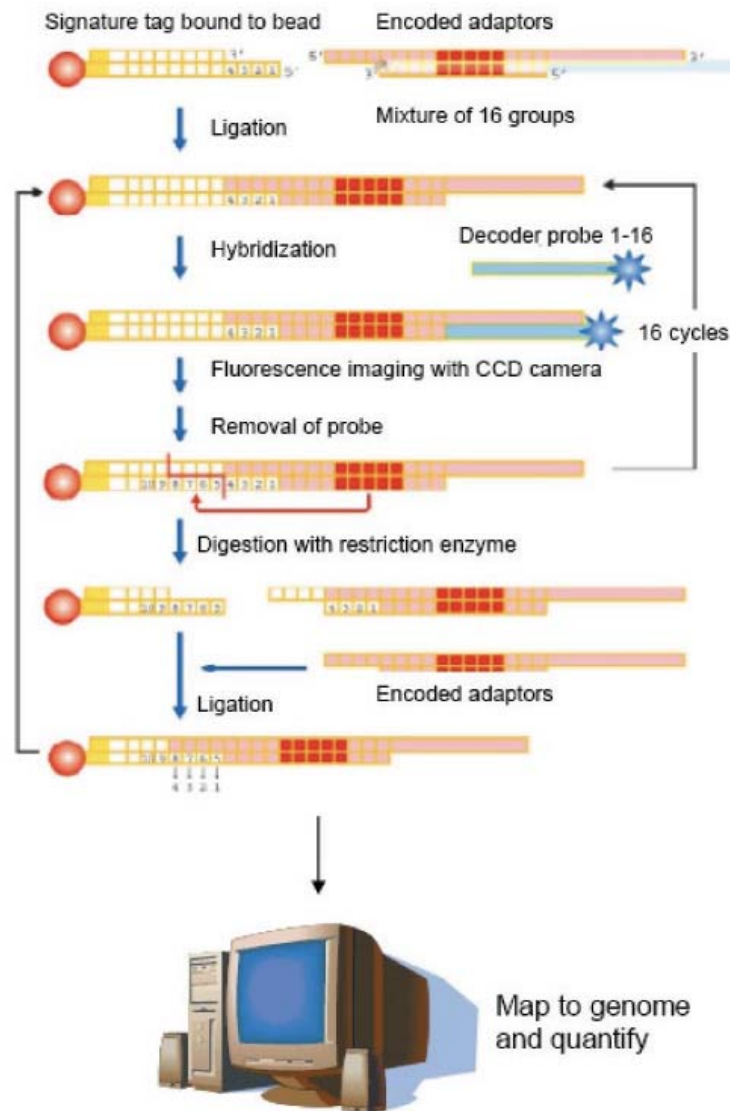


Figure 5. The principle of MPSS sequencing. Sequencing of the signature tag is done in a repeated series of reactions: restriction enzyme cutting, ligation of encoded adaptor and hybridization of decoder probe followed by signal detection and removal of probe before the next cycle is initiated. As with SAGE, the sequenced tags are mapped to the genome and quantified. (Figure adapted from <http://www.takarabioeurope.com/pdf/hd/BV11>).

cDNA collections, novel ideas for biochemistry of high density deposition of nucleic acids on small solid surface areas, in addition to fluorescent labeling of nucleic acids and subsequent detection technology. The steps in a microarray procedure involve converting all mRNA expressed in a cell sample into labeled cDNA using reverse transcriptase. The resulting complex cDNA mixture is then hybridized, in a two channel (co-hybridization of two samples) procedure, to the cDNA fragments or oligonucleotides spotted on a microscope slide to determine the gene expression levels of thousands of genes simultaneously. The steps are described in more detail below, with a bias towards the use of fluorescent detection of hybridized transcripts.

Regardless of the kind of arrays, it has been shown that different platforms perform differently (59-62), although it was established that the gene expression profiles are more influenced by biology rather than technological differences between the array formats (63). For cross-platform comparisons, there are generally several possible approaches to match corresponding measurements of the respective probes. One simple method would be to find a common gene identifier such as Unigene IDs for the different probe sets and compare at gene level. It is common to purchase the probe libraries from a commercial vendor and thus leaving it up the manufacturer to select the characteristics of the probes such as length, position of the probe within the mRNA sequence and homology to

sequences found in other genes. These parameters are important to ensure reproducibility and reliability of the gene expression levels. A thorough cross-platform comparison is therefore challenging due to the variable position of the probes on the reference mRNA sequence. Improper probe set matching may be one of the contributing factors to observed discrepancies between platforms. Carter *et al.* (64), showed how array platform inconsistencies were reduced when only comparing probes sets that had overlapping gene sequences. The largest cross-platform study to date that relied on sequence based mapping of all common probes, reported significant correlation for genes being detected as highly expressed in most platforms (65). However, larger discrepancies were observed for transcripts expressed at low levels.

4.8.2. Target preparation and hybridization

Standard microarray labeling protocols are based on reverse transcription of mRNA, either from purified total RNA or poly(A) mRNA extracts, to generate labeled cDNA targets, typically using fluorescent dyes such as cyanine-3 (Cy3) and cyanine-5 (Cy5) (66). Separate labeling reactions are performed for each sample. The initial protocol is based on direct labeling, whereby reverse transcription is primed with an oligo dT primer and the strand synthesis occurs in the presence of fluorescently labeled nucleotides incorporating the dyes into the newly synthesized strands (32). Cy3 and Cy5 can be well separated in terms of their excitation and emission spectra. However the Cy-dye nucleotides are rather bulky, and alternatively, aminoallyl modified nucleotides can be used in the reverse transcription step to achieve indirect labeling. After cDNA synthesis, a separate coupling reaction of either Cy3 or Cy5 dye molecules to the free amine group on the aminoallyl nucleotide is performed. There are a number of alternative labeling strategies that have been proposed *e.g.* the use of dendrimer technology (67). To generalize, the main goal of novel target labeling techniques (those without amplification steps) is to increase the sensitivity of microarray experiments in attempts to reduce the amount of RNA required (68-70). So far, the most commonly used methods have reduced requirements from hundreds of microgram to tens of microgram and even down to 1 microgram. To be able to profile even smaller samples, several RNA and cDNA amplification strategies have been developed and will be introduced in more detail in a latter section. However, it is worth mentioning due to the widespread use of their arrays, that Affymetrix has set the synthesis of labeled, complementary RNA (cRNA) generated through an *in vitro* transcription step as a standard for target preparation with the use of their oligo arrays (71). Originally, the cRNA samples were labeled with biotin, but as described by 't Hoen *et al.*, (72) indirect labeling of Cy3 and Cy5 via the incorporation of aminoallyl-nucleotide works well. This method can also be used on in-house printed spotted oligo arrays and has been evaluated by Park *et al.*, (73).

The next step in the microarray experimental pipeline is to hybridize the target mixture onto the array, either underneath a coverslip sitting directly on top of the array in the case of manual hybridization, or into a hybridization chamber holding the array in an automatic

hybridization station. The temperature of the hybridization system is determined by the type of array, hybridization buffer composition and is finally optimized by the user. Typically the hybridization step is allowed to proceed for 16-20 hours.

4.9. General technology comparison

Table 1 (page 45) summarizes the features, advantages and disadvantages of the common mRNA transcript quantification techniques. These are the primary features for consideration when selecting a technology to measure mRNA transcripts. It is likely that cost, sensitivity and gene coverage are among the first aspects to be evaluated during selection of technology, having in mind the number of genes to be observed, the number of samples to inspect and evidently, the purpose of the study. Generally, qRT-PCR and real competitive PCR are considered as more sensitive and accurate than the global gene expression techniques (SAGE, MPSS and microarray). However, they are not high throughput assays, although advances are in progress to increase throughput. Further, the sensitivity of SAGE, MPSS and microarray technology in terms of RNA quantity required is under constant improvements, especially with respect to inclusion of a sample amplification procedure. Sensitivity in terms of resolution, on the other hand, differs between these high throughput technologies. SAGE and MPSS measure transcripts in increments of 1 count (= 1 transcript). The increments in signal units in the microarray technology are difficult to extrapolate in terms transcript copies, as it is dependent not only on the transcript quantities but also a number of experimental variables, including the dynamic range of the scanner. It is clear that each technology described has its advantages and disadvantages, and no platform offers complete superiority with respect to an ideal data capture setting (mRNA frequency, splicing, sequence, *etc.*).

One question is how can techniques such as differential display and subtractive hybridization survive in the digital era of high-throughput methods? The gap in difference in sample material requirement, *e.g.* microarray experiments requiring more than suppression subtractive hybridization (SSH), is closing in and thus not a major consideration anymore. In a review article, it was proposed that future prospects of differential display was likely to be directed by continuous improvements of the method and potentials to combine such differential expression techniques with *e.g.* microarrays, allowing complementation and synergism (74). An example of a possible strategy would be to generate labeled targets from the products of a representational differential display (RDA) experiment as an approach to reduce the number of genes to be analyzed in a specific manner. However, this does not satisfy the whole transcriptome analysis approach. Hence, the purpose of the gene expression study certainly influences the choice of technology, and there is an arsenal of techniques to choose from, including a number of methods that have not been described here.

Another reasonable question to raise, considering the plethora of gene expression analysis methods, is how

Table 1. A comparative overview of essential properties of the gene expression methods surveyed

| | Subtractive hybridization | Differential Display | RNase protection assay | q RT-PCR | Real competitive PCR | SAGE | MPSS | Microarray |
|---|---|---|------------------------------------|--|------------------------------------|--|--|---|
| Equipment requirements | Common molecular biology equipment | Common molecular biology equipment | Common molecular biology equipment | Thermocycler with fluorescent detection system | MALDI-TOF MS apparatus | High-throughput sequencer | Proprietary equipment | (Arrayer), arrays, scanner |
| Cost | Low | Low | Low | Moderate | Moderate | High | High | Moderate-high |
| Prior knowledge of sequence | No | No | Yes | Yes | Yes | Needed when mapping tags to genome | Needed when mapping tags to genome | Yes |
| Quantitative | Coarse and relative | Coarse and relative | Coarse and relative | Relative and absolute | Absolute | Absolute | Absolute | Relative and absolute ¹ |
| Throughput 1. gene nr. 2. sample nr. | 1. Medium 2. Low | 1. Medium 2. Low | 1. Low 2. Low | 1. Low (medium) 2. Low (medium) | 1. Low (medium) 2. Low (medium) | 1. High 2. Low | 1. High 2. Medium-high | 1. High 2. Medium-high |
| Sensitivity 1. RNA quantity 2. Resolution | 1. Varies accord. to protocol, potentially high 2. ND ² | 1. Varies accord. to protocol, potentially high 2. ND | 1. ND 2. ND | 1. High 2. High | 1. High 2. High | 1. Relatively high if sample amplification is included 2. High | 1. Relatively high if sample amplification is included 2. High | 1. Relatively high if sample amplification is included 2. Moderate |
| Specificity | Needs further characterization (sequencing), but can yield high specificity | Needs further characterization (sequencing), but can yield high specificity | High | High | High | Moderate-high Depends on tag length, library size, sequencing errors and ambiguous tag annotation | Moderate-high Depends on errors introduced in library and sequencing steps and ambiguous annotation | Moderate-high Depends on probe design and cross hybridization activity |
| Reproducibility | ND | ND | ND | High | High | Reproducible but dependent on tag numbers | ND | Inter-lab reproducibility high |
| Full transcriptome coverage | Yes | Yes | No | No | No | Yes | Yes | No |
| Laborious | Yes | Yes | Yes | No (automation) | No | Yes | No (automation) | No /moderate |

¹Absolute quantification of microarray measurements requires the use of specific parameters to convert the signal into equivalent transcript quantities. An example of a proposed strategy was published by Frigessi *et al.*, (98). ²ND = no data acquired to specify feature.

well do the data compare between technologies and does it matter how we measure mRNA abundance. Investigations of overlap and consistency, is of great value for downstream attempts to integrate data from different technologies in order to construct complete transcript profiles of various tissues. This is a challenging issue, due to the considerable presence of ambiguity regarding data analysis, interpretation and strategies for correlation calculations between different technologies. A direct comparison of data generated by two or more technologies requires identical samples, matched probes sets (preferably sequence matched probe sets), and comparable measurement units. Due to the measurement scale (units), systems providing absolute quantification are easier to compare than relative quantification measurements.

Microarray, SAGE and MPSS represent technologies for which inter-platform agreement is

currently of high interest due to the rapid accumulation of enormous amounts of global gene expression data. To date, these techniques represent the main contributors to complete transcriptome profiling. However, there are relatively few comparison studies in the literature. One reason for this may be the restricted access to more than one technology and the expense involved in a purely comparative study. Further, data comparison between microarray, SAGE and MPSS is not straightforward due to the inherent technological differences. One main challenge is the choice of comparable measurement units. The frequent strategies applied are correlation between signal intensity (microarray) versus tag counts (SAGE, MPSS), or using two biological samples to compare ratios obtained with either technique. It could be argued that comparison of hybridization-based signals and sequence-based tag counts are indirect and can only demonstrate a trend. A few

studies have compared microarray data against SAGE counts and found that correlation was good for high expressing genes and for large ratio changes, but modest for overall measurements (75-78). One study examined array data versus MPSS data and found quite a diverging gene expression profile, as many genes were measured by one platform, but not detected by the other (79).

The general conclusions, drawn from inter-platform comparative studies, range from overall- to modest- to limited agreement. These discrepancies call for data validation regardless of which technique is applied. Quantitative real time RT-PCR stands out as a commonly used validation method to confirm measurements obtained with other techniques, particularly microarray. In other words, qRT-PCR data is generally considered as the true standard, although this may not necessarily be correct. However, the low throughput restricts the number of genes selected for further validation. A common observation from these validation analyses is that microarray data underestimates the magnitude of transcript quantity differences between samples.

So, in principle, and with respect to the actual measurement observed, it does matter which tool is applied to assess mRNA abundances levels. This is due to differences that arise from the intrinsic properties of the technologies themselves, and also from the various processing and analytical steps involved. A simple, specific example of divergence in measurements due to intrinsic technological properties, is querying the abundance level of a gene with several alternative splicing products. An oligoarray with splice specific probes would distinguish the splice products, while SAGE and MPSS would not, unless the splice information by chance was captured in the short sequence tags. A step towards greater concordance between studies, particularly biological studies, may require complete technology insight as to how the respective technologies handle the different complexities of the transcriptome. In addition, the observed discrepancies between platforms underline the need for standardizing procedures and also publicly available data repositories, allowing data integration and resources for construction of tissue specific transcriptomes.

5. INCREASING TECHNOLOGY APPLICABILITY BY RESOLVING RESTRICTIONS AND LIMITATIONS

From the literature, it is evident that all of the gene expression analysis technologies have undergone development to alleviate some of the drawbacks and to further refine the methods. This observation also applies for microarray technology, the most common gene expression tool for high throughput analysis in modern laboratories today. Developments to further increase the applicability and sensitivity of the microarray procedure have rapidly been implemented, and this rapid pace of development is expected to continue.

5.1. Global mRNA amplification

Initially, one of the main hurdles for application of microarray expression analysis was the large amount of

mRNA that each experiment required. The introductory microarray technology paper stated that the amount of mRNA used for target preparation was 5 µg (32). In terms of total RNA, this figure converts to 165-500 µg assuming 1-3% mRNA content. This material quantity requirement restricted the use of microarrays, as not all investigators could provide the required amount of RNA from their cells of interest. However, for the investigators who could provide bulk tissue, the question was whether it was possible to decipher the complex expression patterns and extract the gene expression profiles from the cells of interest, considering the heterogeneity of the cell types present in the tissue. Standard protocols applied today require much less material, although they still do not encompass a number of samples, such as clinically important biopsies or fine needle aspirates. Efforts to substantially reduce the amount of required RNA have focused on two main strategies, signal amplification and sample amplification. The aim of signal amplification is to increase the fluorescent signal emitted per mRNA molecule. The purpose of sample amplification, or global mRNA amplification, is to increase the number of transcripts to sufficient quantities for labeling and hybridization. The latter approach has an increased overall applicability with respect to sample amount input range, compared to the current signal amplification such as the dendrimer-based technology (67). Although laborious, mRNA amplification can be monitored during several steps of the procedure and the amount of material generated is generally sufficient for multiple hybridizations. Amplification can be performed either linearly, using T7-based *in vitro* transcription (the classical Eberwine method (80), see Figure 6), or exponentially by PCR-based strategies (81) or a combination of both (82).

The most important aspect for any amplification protocol to be used in combination with quantitative analysis of gene expression is that the relative transcript abundance present in the initial mRNA sample is maintained throughout the procedure. The presence of transcript abundance bias during the up scaling procedure renders the output data quantitatively unreliable. Systematic assessments of potential distortions to relative transcript abundance or other limitations of global mRNA amplification are therefore important steps before presenting data generated from amplified samples and drawing biological conclusions.

Despite the lack of such systematic assessments, sample amplification became a method of choice for profiling small samples. The Affymetrix platform (high density oligonucleotide arrays) integrated at an early stage the use of linear mRNA amplification as a standard step. The first two papers to quantitatively inspect the differences in gene expression measurements before and after linear amplification, concluded both that the concordance was high, although there were some discrepancies that increased as the input of mRNA into the reaction decreased (83,84). These observations have been confirmed in a number of published studies (85-89). Many of these studies are informative, but rather limited, due to choice of statistical analysis approaches and to reporting

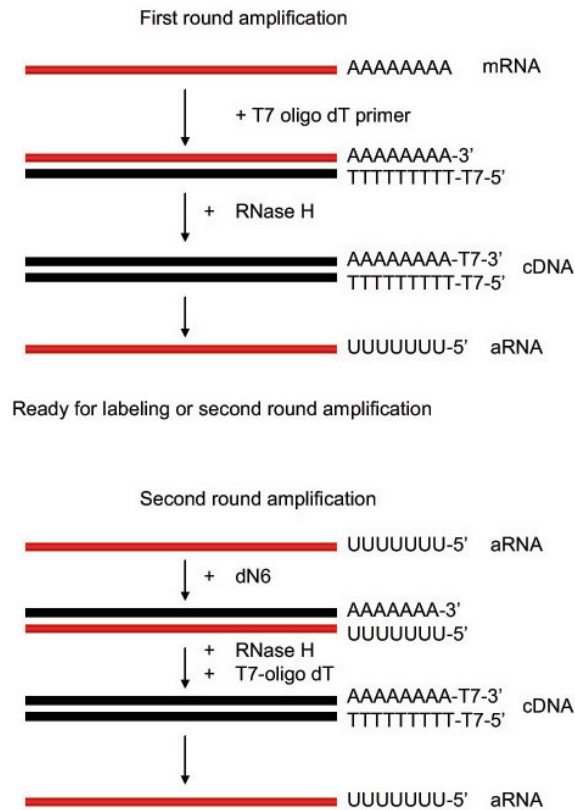


Figure 6. Flowchart of a global and linear mRNA amplification procedure that generates antisense RNA (aRNA). This figure is based on the classical Eberwine method presented by Van Gelder *et al.* (1). An oligo dT primer containing a T7 polymerase binding site is used to prime the first strand cDNA synthesis. Digestion of the mRNA strand in the mRNA-cDNA hybrid by RNase H leaves small fragments of RNA, which are used to prime second strand cDNA synthesis. Antisense RNA is then transcribed by T7 RNA polymerase. Second and subsequent rounds of amplification are initiated by random priming.

results from only subsets of genes. These subsets were often chosen according to selection criteria, such as a two-fold ratio difference, which more or less are replaced by more statistically sound methods. In fact, to accommodate new microarray data analysis approaches, it is more informative to examine the degree of fidelity in a global manner and hence across all genes (90).

Technically, it has been demonstrated that the efficiency of certain amplification protocols for mRNA even allow single cell profiling. However, many studies have observed a markedly reduced correlation with extremely small samples, especially for mRNA transcripts in the low abundance range (83,84,87). A common feature for studies employing minute samples is the reduced number of gene specific transcripts detected on the arrays (91-93). Few investigators have established the lower boundaries with respect to amplification fidelity, but report that variability is increased in experiments with low RNA

input values. In fact, questionable reliability of quantitative measures from low abundant mRNA transcripts is an issue that affects all of the gene expression technologies mentioned when starting with a highly diluted and complex mRNA template mix. It has been reported that in a microarray setting, quantitative accuracy of expression measurements was greatly affected by stochastic effects exerted on low abundant mRNA molecules. The lower abundance of any template, the smaller the probability its true abundance will be maintained in the amplified product (94). This implies that the yield of quantitative data from scarce material is restricted to a few highly expressed genes. Validation of microarray data is generally performed using qRT-PCR. However, as mentioned above, this technique also suffers from the same inconsistencies at low copy numbers and may similarly not represent true measurements. These features require the investigator to fully understand the risk and take the consequence by filtering out unreliable data as not to confer biological significance to invalid quantitative data.

As mentioned above, the technical possibilities to globally amplify mRNA from a single cell have been presented in the published literature. If we return to the theme of the transcriptome of a sample, however, there is an additional aspect of stochastic perturbation to have in mind when looking into the transcriptome of single cells. In addition to stochastic effects exerted on the templates during global sample amplification, investigators are also confronted by stochastic aspects of gene expression. Stochastic fluctuations are considered to be significant in small systems, where components are present at very low concentrations (95). These processes introduce random variation in gene expression among supposedly identical cells, which is referred to as gene expression noise (96). The detection of illegitimate transcripts (low level transcription of tissue-specific gene in nonspecific cells (97)) exemplifies the presence of gene noise. Hence, the extraction of biological significance of gene expression data obtained from amplified single cell material is burdened by stochastic effects both on cellular gene expression, and on the global transcript amplification process. The difficulties lie in distinguishing genuine, reliable transcription levels from stochastically generated noise.

6. CONCLUDING REMARKS AND FUTURE PERSPECTIVE

To date, a substantial number of biological investigations have included mRNA amplification prior to expression analysis and the popular Affymetrix array platform has incorporated mRNA amplification as a standard step in their experimental procedure. Although some ratios are distorted, there are several advantages to include an amplification process. However, defining a limit with respect to minimum amount of total RNA input is rather overlooked. It is accepted that the relative maintenance diminishes but that does not appear to pose a restriction towards conducting studies with extremely small samples. The impression from current gene expression profiling-literature, is that the prevailing, unspoken

consensus says that potentially noisy microarray data is better than none. Qualitative information can surely be obtained, but for quantitative information, we claim otherwise. There is a sample input quantity limit due to stochastic fluctuations and strategies exist to assess this threshold to avoid presenting misleading data. The microarray community is large, and it is likely that the technology users demand more than one published study emphasizing on thresholds excluding extremely small samples before re-evaluating the basis from which to draw biological conclusion in limited sample situations. Until such time, we urge critical evaluation of the data published from mRNA amplified from scarce material.

To understand genes important to a particular phenotype, research has been focused on protein encoding genes. The classical view has been that DNA stores the genetic information, proteins are the effector molecules and mRNAs are the intermediate products. Analysis of gene expression products at the intermediate RNA transcriptional level has been regarded as a decent strategy even though RNA levels do not necessarily correspond to protein levels. Thus, this has provided the basis for the vast amount of gene expression studies particularly by application of microarray technology. The availability, cost, relatively feasible protocols, probe design and flexibility are features that favour the usage of microarray platforms. It may be observed that microarray experiments detect more genes in spite of being a closed system compared to the open systems defined by SAGE and MPSS. The possibility of converting raw signal intensities into universal absolute units provides additional advantages to the use of microarray technology as it opens up for facilitated transcriptome assembly and data comparison.

Recently, we are experiencing a new boost for gene expression analysis at the transcriptional level due to the discovery of abundant non-protein encoding RNA transcripts. Microarray technology has already established itself as a versatile tool to analyze non-coding RNAs through custom tailored array formats (99,100). As most microarrays have been designed to address targets from coding regions in the genome, these new formats have to be designed to include probes representing non-coding genomic regions in order to assess ncRNA expression activity. Applications of RNA amplification strategies are also highly relevant in this newly expanded field of gene expression analysis. It is probable that amplification procedures will provide a key step to facilitate microarray experiments on small samples of ncRNA. The first commercially available kits for amplification of miRNA have already seen the light, one of which (SenseAmp Plus for miRNA, Genisphere Inc) was demonstrated in a published study (101). Clearly, we are entering a new and exciting era of RNA-based gene expression analysis where knowledge from the past decade with high throughput technologies can be applied and further developed to unravel the content and function of the diverse set of RNA molecules transcribed from the genome in particular cells, tissues and physiological or pathological conditions. The future looks bright for microarray technology. However, one should be aware that technical progression may rapidly

change the microarray technique as we know it today. Possible advances include merging microarrays with emerging techniques such as bead array technology and/or microfluidic devices.

7. REFERENCES

1. Maercker, C.: Protein arrays in functional genome research. *Biosci Rep*, 25, 57-70 (2005)
2. Bertone, P. and Snyder, M.: Advances in functional protein microarray technology. *Febs J*, 272, 5400-5411 (2005)
3. Naar, A.M., Lemon, B.D. and Tjian, R.: Transcriptional coactivator complexes. *Annu Rev Biochem*, 70, 475-501 (2001)
4. Human Genome Project Information http://www.ornl.gov/sci/techresources/Human_Genome/faq/genenumber.shtml
5. Ewing, B. and Green, P.: Analysis of expressed sequence tags indicates 35,000 human genes. *Nat Genet*, 25, 232-234 (2000)
6. Claverie, J.M.: Gene number. What if there are only 30,000 human genes? *Science*, 291, 1255-1257 (2001)
7. AltSplice database of alternative spliced events <http://www.ebi.ac.uk/asd/altsplice/humrel2.html>
8. Zavolan, M., Kondo, S., Schonbach, C., Adachi, J., Hume, D.A., Hayashizaki, Y. and Gaasterland, T.: Impact of alternative initiation, splicing, and termination on the diversity of the mRNA transcripts encoded by the mouse transcriptome. *Genome Res*, 13, 1290-1300 (2003)
9. Beaulieu, E., Freier, S., Wyatt, J.R., Claverie, J.M. and Gautheret, D.: Patterns of variant polyadenylation signal usage in human genes. *Genome Res*, 10, 1001-1010 (2000)
10. Tian, B., Hu, J., Zhang, H. and Lutz, C.S.: A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res*, 33, 201-212 (2005)
11. Claverie, J.M.: Fewer genes, more noncoding RNA. *Science*, 309, 1529-1530 (2005)
12. Schadt, E.E., Edwards, S.W., GuhaThakurta, D., Holder, D., Ying, L., Svetnik, V., Leonardson, A., Hart, K.W., Russell, A., Li, G. et al.: A comprehensive transcript index of the human genome generated using microarrays and computational approaches. *Genome Biol*, 5, R73 (2004)
13. Ravasi, T., Suzuki, H., Pang, K.C., Katayama, S., Furuno, M., Okunishi, R., Fukuda, S., Ru, K., Frith, M.C., Gongora, M.M. et al.: Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res*, 16, 11-19 (2006)

14. Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G. et al.: Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science*, 308, 1149-1154 (2005)
15. Ueda, H.R., Hayashi, S., Matsuyama, S., Yomo, T., Hashimoto, S., Kay, S.A., Hogenesch, J.B. and Iino, M.: Universality and flexibility in gene expression from bacteria to human. *Proc Natl Acad Sci U S A*, 101, 3765-3769 (2004)
16. Darzacq, X., Singer, R.H. and Shav-Tal, Y.: Dynamics of transcription and mRNA export. *Curr Opin Cell Biol*, 17, 332-339 (2005)
17. Mendes Soares, L.M. and Valcarcel, J.: The expanding transcriptome: the genome as the 'Book of Sand'. *Embo J*, 25, 923-931 (2006)
18. Bishop, J.O., Morton, J.G., Rosbash, M. and Richardson, M.: Three abundance classes in HeLa cell messenger RNA. *Nature*, 250, 199-204 (1974)
19. Reverter, A., McWilliam, S.M., Barris, W. and Dalrymple, B.P.: A rapid method for computationally inferring transcriptome coverage and microarray sensitivity. *Bioinformatics*, 21, 80-89 (2005)
20. Quinlan, T.J., Beeler, G.W., Jr., Cox, R.F., Elder, P.K., Moses, H.L. and Getz, M.J.: The concept of mRNA abundance classes: a critical reevaluation. *Nucleic Acids Res*, 5, 1611-1625 (1978)
21. Hastie, N.D. and Bishop, J.O.: The expression of three abundance classes of messenger RNA in mouse tissues. *Cell*, 9, 761-774 (1976)
22. Velculescu, V.E., Madden, S.L., Zhang, L., Lash, A.E., Yu, J., Rago, C., Lal, A., Wang, C.J., Beaudry, G.A., Ciriello, K.M. et al.: Analysis of human transcriptomes. *Nat Genet*, 23, 387-388 (1999)
23. Szanislo, P., Wang, N., Sinha, M., Reece, L.M., Van Hook, J.W., Luxon, B.A. and Leary, J.F.: Getting the right cells to the array: Gene expression microarray analysis of cell mixtures and sorted cells. *Cytometry A*, 59, 191-202 (2004)
24. Alwine, J.C., Kemp, D.J. and Stark, G.R.: Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc Natl Acad Sci U S A*, 74, 5350-5354 (1977)
25. Zimmermann, C.R., Orr, W.C., Leclerc, R.F., Barnard, E.C. and Timberlake, W.E.: Molecular cloning and selection of genes regulated in *Aspergillus* development. *Cell*, 21, 709-715 (1980)
26. Liang, P. and Pardee, A.B.: Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*, 257, 967-971 (1992)
27. Hod, Y.: A simplified ribonuclease protection assay. *Biotechniques*, 13, 852-854 (1992)
28. Higuchi, R., Fockler, C., Dollinger, G. and Watson, R.: Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology (N Y)*, 11, 1026-1030 (1993)
29. Ding, C. and Cantor, C.R.: A high-throughput gene expression analysis technique using competitive PCR and matrix-assisted laser desorption ionization time-of-flight MS. *Proc Natl Acad Sci U S A*, 100, 3059-3064 (2003)
30. Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W.: Serial analysis of gene expression. *Science*, 270, 484-487 (1995)
31. Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M. et al.: Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol*, 18, 630-634 (2000)
32. Schena, M., Shalon, D., Davis, R.W. and Brown, P.O.: Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270, 467-470 (1995)
33. Abruzzo, L.V., Lee, K.Y., Fuller, A., Silverman, A., Keating, M.J., Medeiros, L.J. and Coombes, K.R.: Validation of oligonucleotide microarray data using microfluidic low-density arrays: a new statistical method to normalize real-time RT-PCR data. *Biotechniques*, 38, 785-792 (2005)
34. Dobbin, K., Shih, J.H. and Simon, R.: Questions and answers on design of dual-label microarrays for identifying differentially expressed genes. *J Natl Cancer Inst*, 95, 1362-1369 (2003)
35. Kendzierski, C., Irizarry, R.A., Chen, K.S., Haag, J.D. and Gould, M.N.: On the utility of pooling biological samples in microarray experiments. *Proc Natl Acad Sci U S A*, 102, 4252-4257 (2005)
36. Zhang, S.D. and Gant, T.W.: Effect of pooling samples on the efficiency of comparative studies using microarrays. *Bioinformatics*, 21, 4378-4383 (2005)
37. Karrer, E.E., Lincoln, J.E., Hogenhout, S., Bennett, A.B., Bostock, R.M., Martineau, B., Lucas, W.J., Gilchrist, D.G. and Alexander, D.: In situ isolation of mRNA from individual plant cells: creation of cell-specific cDNA libraries. *Proc Natl Acad Sci U S A*, 92, 3814-3818 (1995)
38. Hubank, M. and Schatz, D.G.: Identifying differences in mRNA expression by representational difference analysis of cDNA. *Nucleic Acids Res*, 22, 5640-5648 (1994)
39. Diatchenko, L., Lau, Y.F., Campbell, A.P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K.,

- Gurskaya, N., Sverdlov, E.D. et al.: Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci U S A*, 93, 6025-6030 (1996)
40. O'Hara, E.F., Williams, M.B., Rott, L., Abola, P., Hansen, N., Jones, T., Gurjal, M.R., Federspiel, N. and Butcher, E.C.: Modified representational difference analysis: isolation of differentially expressed mRNAs from rare cell populations. *Anal Biochem*, 336, 221-230 (2005)
41. Liang, P.: A decade of differential display. *Biotechniques*, 33, 338-344, 346 (2002)
42. Sutcliffe, J.G., Foye, P.E., Erlander, M.G., Hilbush, B.S., Bodzin, L.J., Durham, J.T. and Hasel, K.W.: TOGA: an automated parsing technology for analyzing expression of nearly all genes. *Proc Natl Acad Sci U S A*, 97, 1976-1981 (2000)
43. Jurecic, R., Nachtman, R.G., Colicos, S.M. and Belmont, J.W.: Identification and cloning of differentially expressed genes by long-distance differential display. *Anal Biochem*, 259, 235-244 (1998)
44. Rottman, J.B.: The ribonuclease protection assay: a powerful tool for the veterinary pathologist. *Vet Pathol*, 39, 2-9 (2002)
45. Bustin, S.A.: Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol*, 25, 169-193 (2000)
46. Kubista, M., Andrade, J.M., Bengtsson, M., Forootan, A., Jonak, J., Lind, K., Sindelka, R., Sjoback, R., Sjogreen, B., Strombom, L. et al.: The real-time polymerase chain reaction. *Mol Aspects Med*, 27, 95-125 (2006)
47. Ginzinger, D.G.: Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. *Exp Hematol*, 30, 503-512 (2002)
48. Becker-Andre, M. and Hahlbrock, K.: 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 (1989)
49. Edwards, J.R., Ruparel, H. and Ju, J.: Mass-spectrometry DNA sequencing. *Mutat Res*, 573, 3-12 (2005)
50. McCullough, R.M., Cantor, C.R. and Ding, C.: High-throughput alternative splicing quantification by primer extension and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Nucleic Acids Res*, 33, e99 (2005)
51. Yamamoto, M., Wakatsuki, T., Hada, A. and Ryo, A.: Use of serial analysis of gene expression (SAGE) technology. *J Immunol Methods*, 250, 45-66 (2001)
52. Ruijter, J.M., Van Kampen, A.H. and Baas, F.: Statistical evaluation of SAGE libraries: consequences for experimental design. *Physiol Genomics*, 11, 37-44 (2002)
53. Patino, W.D., Mian, O.Y. and Hwang, P.M.: Serial analysis of gene expression: technical considerations and applications to cardiovascular biology. *Circ Res*, 91, 565-569 (2002)
54. Chen, J., Sun, M., Lee, S., Zhou, G., Rowley, J.D. and Wang, S.M.: Identifying novel transcripts and novel genes in the human genome by using novel SAGE tags. *Proc Natl Acad Sci U S A*, 99, 12257-12262 (2002)
55. Saha, S., Sparks, A.B., Rago, C., Akmaev, V., Wang, C.J., Vogelstein, B., Kinzler, K.W. and Velculescu, V.E.: Using the transcriptome to annotate the genome. *Nat Biotechnol*, 20, 508-512 (2002)
56. Chuang, S.E., Daniels, D.L. and Blattner, F.R.: Global regulation of gene expression in *Escherichia coli*. *J Bacteriol*, 175, 2026-2036 (1993)
57. Chalifour, L.E., Fahmy, R., Holder, E.L., Hutchinson, E.W., Osterland, C.K., Schipper, H.M. and Wang, E.: A method for analysis of gene expression patterns. *Anal Biochem*, 216, 299-304 (1994)
58. Drmanac, S. and Drmanac, R.: Processing of cDNA and genomic kilobase-size clones for massive screening, mapping and sequencing by hybridization. *Biotechniques*, 17, 328-329, 332-326 (1994)
59. Kuo, W.P., Jenssen, T.K., Butte, A.J., Ohno-Machado, L. and Kohane, I.S.: Analysis of matched mRNA measurements from two different microarray technologies. *Bioinformatics*, 18, 405-412 (2002)
60. Li, J., Pankratz, M. and Johnson, J.A.: Differential gene expression patterns revealed by oligonucleotide versus long cDNA arrays. *Toxicol Sci*, 69, 383-390 (2002)
61. Tan, P.K., Downey, T.J., Spitznagel, E.L., Jr., Xu, P., Fu, D., Dimitrov, D.S., Lempicki, R.A., Raaka, B.M. and Cam, M.C.: Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res*, 31, 5676-5684 (2003)
62. Irizarry, R.A., Warren, D., Spencer, F., Kim, I.F., Biswal, S., Frank, B.C., Gabrielson, E., Garcia, J.G., Geoghegan, J., Germino, G. et al.: Multiple-laboratory comparison of microarray platforms. *Nat Methods*, 2, 345-350 (2005)
63. Yauk, C.L., Berndt, M.L., Williams, A. and Douglas, G.R.: Comprehensive comparison of six microarray technologies. *Nucleic Acids Res*, 32, e124 (2004)
64. Carter, S.L., Eklund, A.C., Mecham, B.H., Kohane, I.S. and Szallasi, Z.: Redefinition of Affymetrix probe sets by sequence overlap with cDNA microarray probes

reduces cross-platform inconsistencies in cancer-associated gene expression measurements. *BMC Bioinformatics*, 6, 107 (2005)

65. Kuo, W.P., Liu, F., Trimarchi, J., Punzo, C., Lombardi, M., Sarang, J., Whipple, M.E., Maysuria, M., Serikawa, K., Lee, S.Y. et al.: A sequence-oriented comparison of gene expression measurements across different hybridization-based technologies. *Nat Biotechnol* (2006)

66. Holloway, A.J., van Laar, R.K., Tothill, R.W. and Bowtell, D.D.: Options available--from start to finish--for obtaining data from DNA microarrays II. *Nat Genet*, 32 Suppl, 481-489 (2002)

67. Stears, R.L., Getts, R.C. and Gullans, S.R.: A novel, sensitive detection system for high-density microarrays using dendrimer technology. *Physiol Genomics*, 3, 93-99 (2000)

68. Karsten, S.L., Van Deerlin, V.M., Sabatti, C., Gill, L.H. and Geschwind, D.H.: An evaluation of tyramide signal amplification and archived fixed and frozen tissue in microarray gene expression analysis. *Nucleic Acids Res*, 30, E4 (2002)

69. Gupta, V., Cherkassky, A., Chatis, P., Joseph, R., Johnson, A.L., Broadbent, J., Erickson, T. and DiMeo, J.: Directly labeled mRNA produces highly precise and unbiased differential gene expression data. *Nucleic Acids Res*, 31, e13 (2003)

70. Huber, M., Wei, T.F., Muller, U.R., Lefebvre, P.A., Marla, S.S. and Bao, Y.P.: Gold nanoparticle probe-based gene expression analysis with unamplified total human RNA. *Nucleic Acids Res*, 32, e137 (2004)

71. Mahadevappa, M. and Warrington, J.A.: A high-density probe array sample preparation method using 10- to 100-fold fewer cells. *Nat Biotechnol*, 17, 1134-1136 (1999)

72. t Hoen, P.A., de Kort, F., van Ommen, G.J. and den Dunnen, J.T.: Fluorescent labelling of cRNA for microarray applications. *Nucleic Acids Res*, 31, e20 (2003)

73. Park, J.Y., Kim, S.Y., Lee, J.H., Song, J., Noh, J.H., Lee, S.H., Park, W.S., Yoo, N.J., Lee, J.Y. and Nam, S.W.: Application of amplified RNA and evaluation of cRNA targets for spotted-oligonucleotide microarray. *Biochem Biophys Res Commun*, 325, 1346-1352 (2004)

74. Broude, N.E.: Differential display in the time of microarrays. *Expert Rev Mol Diagn*, 2, 209-216 (2002)

75. Ishii, M., Hashimoto, S., Tsutsumi, S., Wada, Y., Matsushima, K., Kodama, T. and Aburatani, H.: Direct comparison of GeneChip and SAGE on the quantitative accuracy in transcript profiling analysis. *Genomics*, 68, 136-143 DOI: 10.1006/geno.2000.6284. (2000)

76. Evans, S.J., Datson, N.A., Kabbaj, M., Thompson, R.C., Vreugdenhil, E., De Kloet, E.R., Watson, S.J. and

Akil, H.: Evaluation of Affymetrix Gene Chip sensitivity in rat hippocampal tissue using SAGE analysis. *Serial Analysis of Gene Expression. Eur J Neurosci*, 16, 409-413 (2002)

77. van Ruissen, F., Ruijter, J.M., Schaaf, G.J., Asgharnegad, L., Zwijnenburg, D.A., Kool, M. and Baas, F.: Evaluation of the similarity of gene expression data estimated with SAGE and Affymetrix GeneChips. *BMC Genomics*, 6, 91 (2005)

78. Kim, H.L.: Comparison of oligonucleotide-microarray and serial analysis of gene expression (SAGE) in transcript profiling analysis of megakaryocytes derived from CD34+ cells. *Exp Mol Med*, 35, 460-466 (2003)

79. Oudes, A.J., Roach, J.C., Walashek, L.S., Eichner, L.J., True, L.D., Vessella, R.L. and Liu, A.Y.: Application of Affymetrix array and Massively Parallel Signature Sequencing for identification of genes involved in prostate cancer progression. *BMC Cancer*, 5, 86 (2005)

80. Van Gelder, R.N., von Zastrow, M.E., Yool, A., Dement, W.C., Barchas, J.D. and Eberwine, J.H.: Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc. Natl. Acad. Sci. U S A*, 87, 1663-1667 (1990)

81. Iscove, N.N., Barbara, M., Gu, M., Gibson, M., Modi, C. and Winegarden, N.: Representation is faithfully preserved in global cDNA amplified exponentially from sub-picogram quantities of mRNA. *Nat. Biotechnol.* (2002)

82. Aoyagi, K., Tatsuta, T., Nishigaki, M., Akimoto, S., Tanabe, C., Omoto, Y., Hayashi, S., Sakamoto, H., Sakamoto, M., Yoshida, T. et al.: A faithful method for PCR-mediated global mRNA amplification and its integration into microarray analysis on laser-captured cells. *Biochem Biophys Res Commun*, 300, 915-920 (2003)

83. Wang, E., Miller, L.D., Ohnmacht, G.A., Liu, E.T. and Marincola, F.M.: High-fidelity mRNA amplification for gene profiling. *Nat. Biotechnol.*, 18, 457-459 (2000)

84. Baugh, L.R., Hill, A.A., Brown, E.L. and Hunter, C.P.: Quantitative analysis of mRNA amplification by in vitro transcription. *Nucleic Acids Res.*, 29, E29 (2001)

85. Zhao, H., Hastie, T., Whitfield, M.L., Borresen-Dale, A.L. and Jeffrey, S.S.: Optimization and evaluation of T7 based RNA linear amplification protocols for cDNA microarray analysis. *BMC Genomics*, 3, 31 (2002)

86. Puskas, L.G., Zvara, A., Hackler, L., Jr. and Van Hummelen, P.: RNA amplification results in reproducible microarray data with slight ratio bias. *Biotechniques*, 32, 1330-1334, 1336, 1338, 1340 (2002)

87. Scheidl, S.J., Nilsson, S., Kalen, M., Hellstrom, M., Takemoto, M., Hakansson, J. and Lindahl, P.: mRNA expression profiling of laser microbeam microdissected

cells from slender embryonic structures. *Am. J. Pathol.*, 160, 801-813 (2002)

88. Hu, L., Wang, J., Baggerly, K., Wang, H., Fuller, G.N., Hamilton, S.R., Coombes, K.R. and Zhang, W.: Obtaining reliable information from minute amounts of RNA using cDNA microarrays. *BMC Genomics*, 3, 16 (2002)

89. Feldman, A.L., Costouros, N.G., Wang, E., Qian, M., Marincola, F.M., Alexander, H.R. and Libutti, S.K.: Advantages of mRNA amplification for microarray analysis. *Biotechniques*, 33, 906-912, 914 (2002)

90. Segal, E., Friedman, N., Kaminski, N., Regev, A. and Koller, D.: From signatures to models: understanding cancer using microarrays. *Nat Genet*, 37 Suppl, S38-45 (2005)

91. Nygaard, V., Holden, M., Loland, A., Langaas, M., Myklebost, O. and Hovig, E.: Limitations of mRNA amplification from small-size cell samples. *BMC Genomics*, 6, 147 DOI: 10.1186/1471-2164-6-147. (2005)

92. Nakagawa, T. and Schwartz, J.P.: Gene expression profiles of reactive astrocytes in dopamine-depleted striatum. *Brain Pathol*, 14, 275-280 (2004)

93. Mohr, S., Bottin, M.C., Lannes, B., Neuville, A., Bellocq, J.P., Keith, G. and Rihn, B.H.: Microdissection, mRNA amplification and microarray: a study of pleural mesothelial and malignant mesothelioma cells. *Biochimie*, 86, 13-19 (2004)

94. Stenman, J., Lintula, S., Rissanen, O., Finne, P., Hedstrom, J., Palotie, A. and Orpana, A.: Quantitative detection of low-copy-number mRNAs differing at single nucleotide positions. *Biotechniques*, 34, 172-177 (2003)

95. Swain, P.S., Elowitz, M.B. and Siggia, E.D.: Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci U S A*, 99, 12795-12800 (2002)

96. Raser, J.M. and O'Shea, E.K.: Noise in gene expression: origins, consequences, and control. *Science*, 309, 2010-2013 (2005)

97. Chelly, J., Concordet, J.P., Kaplan, J.C. and Kahn, A.: Illegitimate transcription: transcription of any gene in any cell type. *Proc Natl Acad Sci U S A*, 86, 2617-2621 (1989)

98. Frigessi, A., van de Wiel, M.A., Holden, M., Svendsrud, D.H., Glad, I.K. and Lyng, H.: Genome-wide estimation of transcript concentrations from spotted cDNA microarray data. *Nucleic Acids Res*, 33, e143 (2005)

99. Aravin, A. and Tuschl, T.: Identification and characterization of small RNAs involved in RNA silencing. *FEBS Lett*, 579, 5830-5840 (2005)

100. Huttenhofer, A. and Vogel, J.: Experimental approaches to identify non-coding RNAs. *Nucleic Acids Res*, 34, 635-646 (2006)

101. Mattie, M.D., Benz, C.C., Bowers, J., Sensinger, K., Wong, L., Scott, G.K., Fedele, V., Ginzinger, D.G., Getts, R.C. and Haqq, C.M.: Optimized high-throughput microRNA expression profiling provides novel biomarker assessment of clinical prostate and breast cancer biopsies. *Mol Cancer*, 5, 24 (2006)

Key Words: mRNA amplification, mRNA Measurements, Microarrays, High Throughput Expression Analysis, Review

Send correspondence to: Eivind Hovig, Department of Tumor Biology, Institute for Cancer Research, Norwegian Radium Hospital, 0310 Oslo, Norway, Tel: 47-22935416, Fax: 47-22522421, E-mail: ehovig@ifi.uio.no

<http://www.bioscience.org/current/vol14.htm>