Chapter 1

INTRODUCTION TO MICROARRAY ANALYSIS

1.0 Introduction

Microarray analysis has emerged in the last few years as a flexible method for analyzing large numbers of nucleic acid fragments in parallel. Its origins can be traced to several different disciplines and techniques. Microarrays can be seen as a continued development of molecular biology hybridization methods, as an extension of the use of fluorescence microscopy in cell biology, as well as a diagnostic assay using capture to solid surface as a way to reduce the amount of analytes needed. The convergence of ideas and principles utilized in these fields, together with technological advancements in preparing miniaturized collections of nucleic acids on solid supports, have all contributed to the emergence of microarray and microchip technologies.

In molecular biology, analysis of nucleic acids by hybridization is a universally adopted key method for analysis. Filter-based dot blot analysis has been used for a long time as a convenient method for analyzing multiple samples by hybridization. Classical gene expression analysis methods such as Northern blotting, reverse transcriptase polymerase chain reaction (RT-PCR) and nuclease protection assays, are best suited for analyzing a limited number of genes and samples at a time. By reversing the Northern blotting principle so that the labelled moiety is derived from the mRNA sample and the immobilized fractions are the known sequences traditionally used as probes, filter-based gene expression analysis has enabled simultaneous determination of expression levels of thousands of genes in one experiment. Because of the ease of use of these filter-based methods and their compatibility with general lab equipment, these macroarrays have been widely adopted for gene expression studies (1). One disadvantage to using this method has been the relatively large size and the autofluorescence of the membrane, which prevents efficient use of multiplexed fluorescent probes and subsequently limits the number of samples that can be analyzed in each experiment (Fig 1).

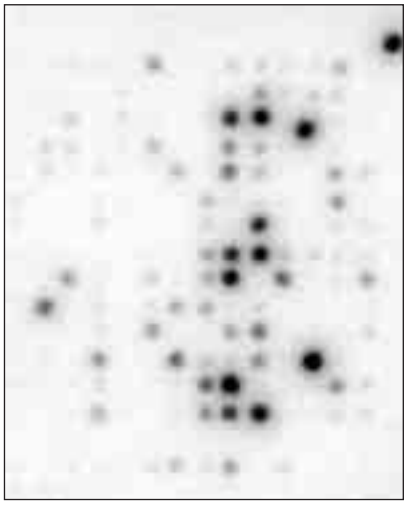


Fig 1. Comparison of a macroarray and microarray. A close-up of a filter macroarray (a) hybridized with ^{32}P-labelled cDNA probe and a microarray (b) hybridized with two different cDNA probes, one labelled with Cy™3 and the other with Cy5. The array images are shown approximately to the same scale.

Utilizing microscope slides to immobilize cells and chromosomes precedes filter-based gene expression analysis, as well as proven methods such as immunohistochemistry, immunocytochemistry and *in situ* hybridization. By combining fluorescence analysis of multiplexed probes with microscopy, fluorescent *in situ* hybridization (FISH) has enabled detection of nucleic acids within cells and chromosomes, and has been found useful in gene expression and genomic analysis.

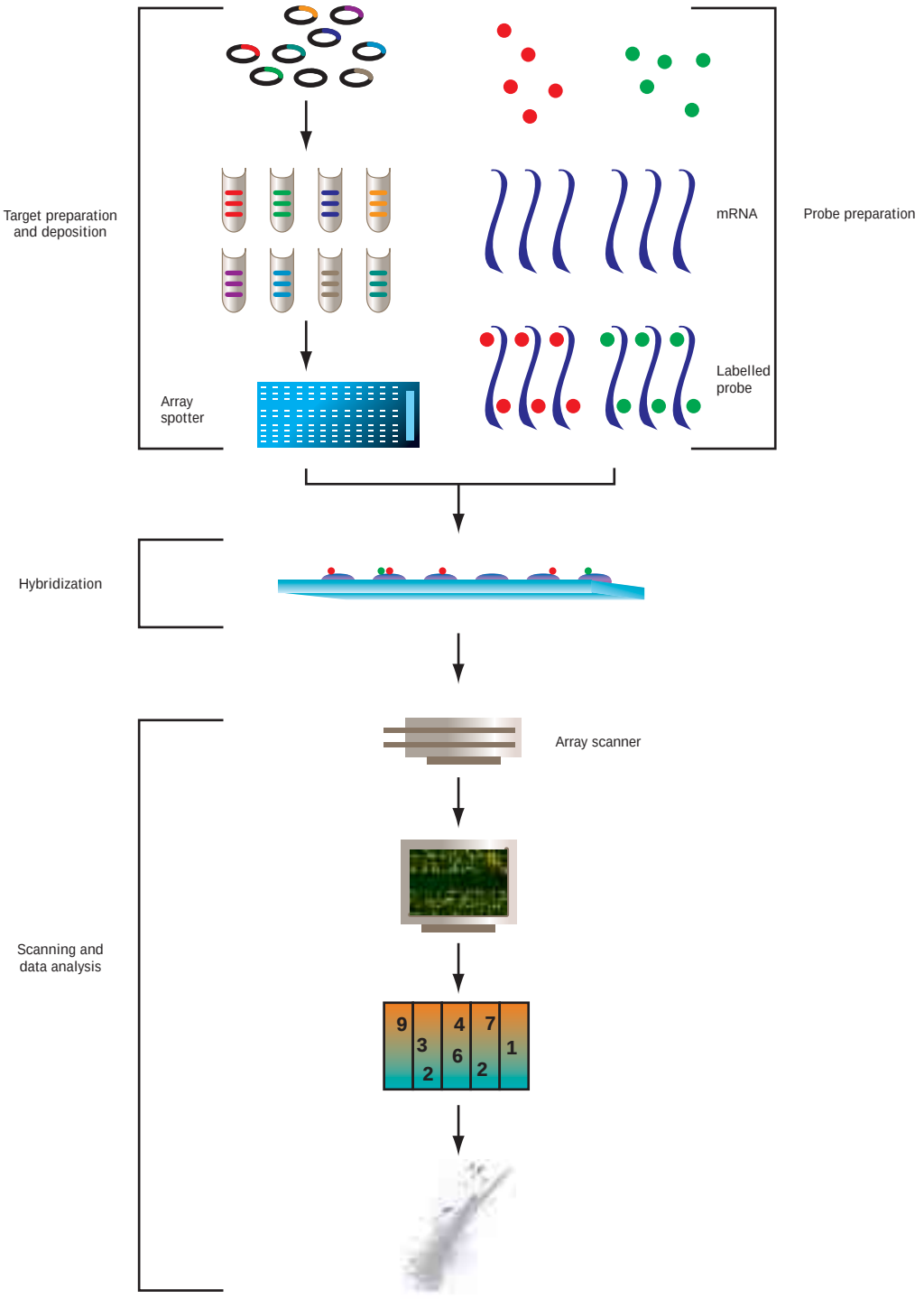
These two methods of analysis were brought together by advancements made in attaching nucleic acid sequences to a glass support. Borrowing a technique from semiconductor manufacturing, photolithography was used to synthesize oligonucleotides directly onto a glass support. Separately, a procedure called contact printing was used to deposit purified nucleic acid onto a slide surface (2). These methods have since made it possible to miniaturize the macroarray experiment, so to speak, by using microscope slides instead of membrane filters. In 1995 and 1996, the first papers in which the term 'microarray' was used in its current meaning were published by the laboratory of Pat Brown at Stanford University (3). The rapid adoption of this technique is illustrated by the publication in 2001 of over 900 papers on the use of microarray technique.

1.1 Principles of microarray analysis

Despite the variety of technical solutions that have been developed for performing microarray analysis, all are miniaturized hybridization assays for studying thousands of nucleic acid fragments simultaneously. All microarray systems (Fig 2) share the following key components:

- the array, which contains immobilized nucleic acid sequences, or 'targets'
- one or more labelled samples or 'probes', that are hybridized with the microarray
- a detection system that quantitates the hybridization signal

Fig 2. Principles of microarrays.



1.1.1 Nomenclature for microarrays

The terms ‘probe’ and ‘target’ are sometimes used interchangeably to describe either the labelled sample or the immobilized nucleic acids. In this handbook the immobilized nucleic acid is referred to as the target and the labelled sample as the probe (Fig 3).

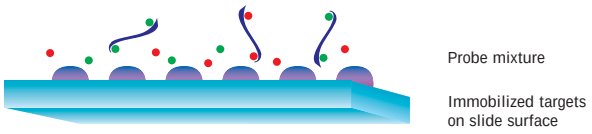


Fig 3. Target and probe. Targets are the immobilized nucleic acids on the slide surface. A probe consisting of two identical populations of nucleic acids labelled with different fluorescent dyes is shown.

1.1.2 Microarrays

Microarrays consist of a collection of nucleic acid sequences immobilized onto a solid support so that each unique sequence forms a tiny feature, called a ‘spot’ or ‘target’. These nucleic acids are obtained in numerous ways, and there are different methods for depositing them onto microarray slides (refer to chapter 3). The size of these spots varies from one system to another, but it is usually less than two hundred micrometers in diameter. A glass slide or glass wafer acts as the solid support onto which up to tens of thousands of spots can be arrayed in a total area of a few square centimeters (Fig 4).

Fig 4. A glass microarray. A standard and mirrored glass microscope slide that contains several thousands of immobilized cDNA fragments. Because of the small amounts of nucleic acid present and their tiny size, the target spots are not visible with the naked eye.

1.1.3 Probe labelling

The microarray sample that is being analyzed, whether it is mRNA for a gene expression study or DNA derived from genomic analysis, is converted to a labelled population of nucleic acids, the probe. These probes frequently consist of several thousands of different labelled nucleic acid fragments. The complexity of microarray hybridization—over 10 000 different labelled fragments interrogating up to 100 000 different immobilized sequences—is greater than that encountered in other routine molecular biology experiments. Therefore, this hybridization should be carried out under conditions that do not promote annealing of non-complementary fragments.

Fluorescent dyes, and especially the cyanine dyes Cy3 and Cy5, have been adopted as the predominant label in microarray analysis. Fluorescence has the advantage of permitting the detection of two or more different signals in one experiment. This has allowed investigators to perform comparative analysis of two or more samples on one microarray. It has also increased the accuracy and throughput of microarray analysis over filter-based macroarrays, in which only one radioactively labelled sample can be conveniently analyzed at a time.

1.1.4 Microarray hybridization

In a microarray hybridization, the labelled fragments in the probe are expected to form duplexes with their immobilized complementary targets. This requires that the nucleic acids are single-stranded and accessible to each other. The number of duplexes formed reflects the relative number of each specific fragment in the probe, as long as the amount of immobilized target nucleic acid is in excess and not limiting the kinetics of hybridization. Two or more samples labelled with different fluorescent dyes can be hybridized simultaneously, resulting in simultaneous hybridization taking place at each target spot. By measuring the different fluorescent signals associated with each spot, the relative abundance of specific sequences in each of the samples can be determined.

1.1.5 Scanning and data analysis

Microarray scanners typically contain two different lasers that emit light at wavelengths that are suitable for exciting the fluorescent dyes used as labels. A confocal microscope attached to a detector system records the emitted light from each of the microarray spots, allowing high-resolution detection of the hybridization signals.

Despite their small size, microarrays generate large quantities of data even from a single experiment. As a typical experiment will involve the use of several analyzed samples on replicate arrays, the use of computerized data processing is necessary in order to handle the amount of data generated and to gain maximum information from the experiment. This can be achieved by specialized software that extracts primary data from scanned microarray slide images, normalizes this data to remove the influence of experimental variation, and finally manipulates the data so that biologically meaningful conclusions can be made.

1.2 Applications of microarray analysis

The versatility of microarray analysis is confirmed by its rapid emergence as a general molecular biology analytical technique. Increasing numbers of researchers within academic institutions and industrial laboratories are now exploiting this technology in diverse biomedical disciplines. Microarrays have not become a replacement to established techniques, but more a novel, high-power approach to perform analyses that were previously time consuming.

By using information derived from the several complete or near complete genome sequences, including the human genome, it is now possible to perform genome-wide experiments using microarray technology. This has already been demonstrated for *S. cerevisiae* where all the expressed genes are known. As microarrays can contain thousands of targets, both characterized and uncharacterized, experiments can be conducted without prior hypotheses. This combined with the millions of data points that are possible to analyze in one experiment, microarray analysis has enabled global analysis of biological processes. Gene expression analysis, genome analysis, and drug discovery have been three of the main areas in which microarray analysis has been applied so far.

1.3 Gene expression analysis

Gene expression analysis examines the composition of cellular messenger RNA populations. The identity of transcripts that make up these populations and their expression levels are informative of cell state and activity of genes and, as the precursors of translated proteins, changes in mRNA levels are related to changes in the proteome.

1.3.1 Traditional techniques

Traditional gene expression analysis has used techniques such as Northern blotting, RT-PCR and nuclease protection assays. More advanced methods—some of these include differential display, subtractive hybridization, representational difference analysis, expressed sequence tags, cDNA fragment fingerprinting, and serial analysis of gene expression—have enabled the discovery of novel differentially expressed genes (4). However, the technical challenges of these methods still limit their use to the analysis of just a few samples at a time. Microarray analysis, in contrast, allows the analysis of thousands of genes in multiple samples with relative ease.

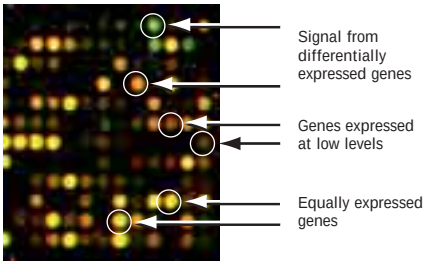


Fig 5. Dual color differential microarray analysis. Dual color microarray hybridization signals are typically represented as false color images in which signals from one dye are presented in red (Cy5 in this case) and signals from the other dye in green (Cy3). If equal signal is obtained from a spot, it will appear yellow. Shades of green and red denote differences in relative abundance in favor of one or the other sample. As the screen appearance of microarray images can be easily manipulated, information gained from such images can be misleading.

1.3.2 Gene expression analysis with microarrays

A typical microarray gene expression analysis experiment compares the relative expression levels of specific transcripts in two samples. One of these samples is a control and the other is derived from cells whose response or status is being investigated. Each of these samples is labelled with a different fluorescent dye, and equal amounts of the labelled samples are combined and hybridized with the microarray. The fluorescent signals corresponding to the two dyes are measured independently from each spot after hybridization. After normalization, the intensity of the two hybridization signals can be compared. Equal signal from both samples suggests equal expression in both samples (Fig 5).

Microarray analysis does not give information about absolute gene expression levels in the samples. This is because the intensity of the fluorescent signals is not only proportional to the number of hybridized fragments but also to the length of these fragments and the number of fluorescent labels each fragment carries, i.e. labelling density. As these are determined by the unique nucleotide sequence of each gene and transcript, they will vary from gene to gene. If two samples have been labelled under similar conditions, the length and labelling density of specific transcripts will be similar in the two samples, making it possible to compare the relative abundance of the transcripts in the two samples. A strong hybridization signal from microarray analysis does not necessarily correspond to a highly expressed gene; it could be derived, for example, from a gene that is expressed at a relatively low level but yields long, highly-labelled probe fragments.

Gene expression analysis with microarrays has been applied to numerous mammalian tissues, plants, yeast, and bacteria alike (1, 5, 6, 7, 8). These studies have examined the effects of treating cells with chemicals, the consequences of over-expression of regulatory factors in transfected cells, and compared mutant strains with parental strains to delineate functional pathways. In cancer research microarrays have been used to find gene expression changes in transformed cells and metastases, to identify diagnostic markers, and to classify tumors based on their gene expression profiles (9, 10, 11).

1.4 Genomic analysis

Microarrays are proving to be useful tools for genomic analysis. Identification of new genes by examining nucleic acid sequences derived from open reading frames has proved to be an efficient way of annotating the human genome and facilitating the use of genomic information for experimental purposes (12). Understanding of gene regulation is advanced by elucidation of transcription factor gene interactions. For example, by combining immunoprecipitation of transcription factor-DNA complexes to microarray identification of DNA fragments on a genomic microarray, it was possible to identify functional regulatory elements in the yeast genome (13). Furthermore, microarrays can be used for predicting splice variants of transcripts and analyzing genomic fragments derived from genetic analysis methods, such as genomic mismatch scanning and representational difference analysis (14, 15). Oligonucleotide microarrays have been applied to analysis of known single nucleotide polymorphisms (SNPs) and mutations (16, 17). Samples can be sequenced using microarray hybridization (18), thus providing convenient means for identifying new genetic variants.

1.5 Drug discovery

As a typical drug discovery process takes several years and incurs high costs, and only a few drug candidates result in approved drugs, methods that increase the efficiency of the process and improve the chances of developing effective drugs have been welcomed.

Microarrays have been found to provide useful information in the different stages of the drug discovery process (8, 15, 19). Identification of potential drug targets can be aided by elucidating metabolic pathways by looking for co-expressed genes. The protein targets of drug treatments can be identified by finding a protein that causes the same changes as a drug when removed from cells. Once drug candidates have been identified and selected, microarrays can be used to define their toxic properties by examining expression profiles induced by drug treatments (20). On the other hand, different function modes of drugs were identified based on the gene expression changes they elicited (21).

Chapter 2

GENETIC CONTENT OF MICROARRAYS

2.0 Introduction

The genetic content of microarrays resides in the immobilized nucleic acid sequences on the microarray. The identity of these sequences determines what information can be obtained from array experiments and how reliable this information is. As microarrays enable simultaneous interrogation of up to tens or hundreds of thousands of targets with one or more labelled probes, generation of accurate data demands that only specific interactions result in detectable signals. Several strategies for preparing the immobilized target nucleic acids for microarrays exist (Fig 6). These nucleic acids can be synthesized directly on the microarray or they can be purified cDNA clones, other DNA fragments or oligonucleotides, which are deposited onto the array by a printing process. This flexibility of using either partially characterized sequences or defined oligonucleotides as targets has improved the application of microarray analysis to different biological problems in a number of species.

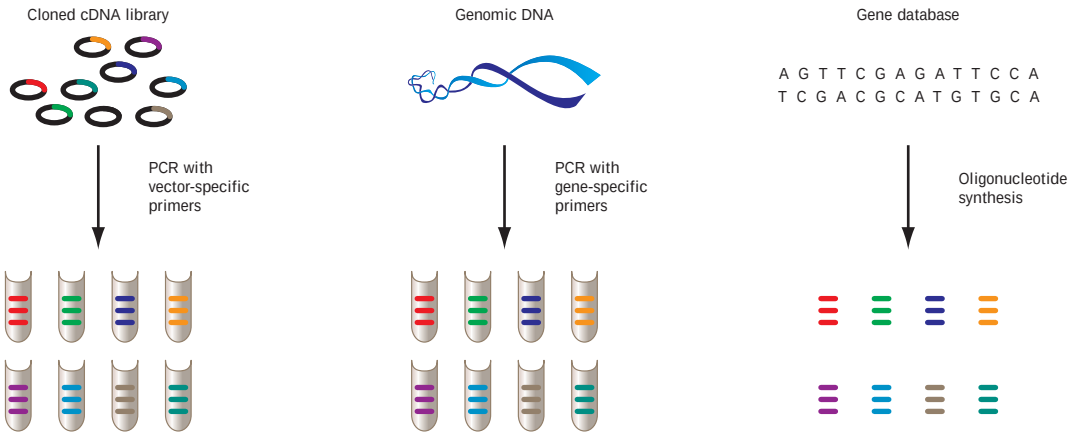


Fig 6. Sources of microarray target sequences. Some of the common strategies for obtaining targets for microarray analysis are illustrated.

2.1 Oligonucleotides as genetic content

2.1.1 Printed oligonucleotide arrays

Oligonucletides can be attached onto microarrays by depositing modified oligos onto a specially treated glass surface. The deposition can be achieved with a variety of contact and non-contact printing methods. See chapter 3, section 3.1.2 on common deposition methods for a detailed overview of this process.

Depending on the source of oligonucleotide content, the length of these oligonucleotides typically varies between 50–70 nucleotides (22, 23). Different microarrays can be easily prepared with the deposition method by choosing different sets of oligonucleotides for array printing. This method is increasingly advantageous because customized microarrays can be prepared in a researcher's own laboratory using microarray spotter equipment.

Regardless of the array fabrication method, the use of oligonucleotides requires that the nucleotide sequences of the intended targets are known. The publication of the human genome sequence as well as partial or complete sequences of several other organisms has facilitated this task. However, the accuracy of the information in the databases has a critical impact on the quality of the arrays. Errors in sequence entries can result in oligonucleotides that do not function in hybridization, because nucleotide mismatches can prevent efficient hybridization from taking place or non-complementary target strands are used by mistake. As more and more genes are identified and their sequences elucidated, the power of oligonucleotide arrays will increase.

2.1.2 Benefits of oligonucleotide arrays

Oligonucleotide targets have several benefits over cDNA targets.

- Different parts of the same gene can be represented on the array. This enables a more robust design of microarray experiments as the same gene can be probed independently for the same information in the same experiment.
- Oligonucleotides can be designed to distinguish between alternative splicing variants as well as different alleles. Oligonucleotides offer precise control over the genetic composition on the arrays. With a judicious choice of oligos, it is possible to discriminate between related gene sequences and study different members of gene families simultaneously.

- Oligonucleotide targets are readily available from commercial manufacturers or synthesized by researchers.
- The time and effort required to prepare oligonucleotides for array printing is less than that required for preparing cloned targets by molecular biology methods.

2.1.3 Design of oligonucleotides

The design of oligonucleotide targets should take into account factors that influence the specificity and strength of hybridization with labelled probes. The specificity can be estimated by comparing the oligonucleotide sequence with known gene sequences. Predicting the strength of hybridization is more difficult, however. Computer algorithms have been developed for selecting target oligonucleotides. Some general rules for oligonucleotide selection have been established:

- Repeat sequences should be avoided, including polynucleotide stretches, repetitive genomic elements, and palindromic sequences.
- The chosen sequences should not be homologous to other genes, but one short homologous stretch may still produce enough specificity in hybridization (22, 24, 25).
- The length of the oligonucleotide, its nucleotide sequence, as well as the positions of mismatches in the oligo, all influence the behavior of the oligo in hybridization.
- It is important to choose a fairly even distribution of all four nucleotides in the sequence.
- Testing of oligonucleotide targets before including them on arrays can help to eliminate sequences that will not perform well.
- The use of computer algorithms may also facilitate the selection of target oligonucleotides (23).

Target sequences may not be accessible to probe molecules near the attachment site on the solid support. mSpacer sequences can be used to increase hybridization efficiency. These are additional sequences added to the oligo sequence to move it further away from the solid support (2). 40-atom long spacers were found to result in up to 150-fold increase in hybridization signal on oligonucleotide arrays (26).

2.2 DNA fragments as genetic content

Development of mechanical microspotting methods and instruments, which can be used to deposit nucleic acid solutions onto glass surface, has enabled the use of cDNA clones and other DNA fragments as microarray targets (27). These methods allow quick and adaptable construction of microarrays that can be customized according to different experimental needs.

2.2.1 Sources of DNA targets

The nucleic acid fragments used for microarray construction can be derived from a number of sources. For gene expression microarrays the fragments are typically derived from either cDNA clones or amplified from exon sequences. Libraries of cDNA clones, expressed sequence tags, clones isolated from subtraction libraries in which the number of highly expressed sequences has been minimized, or PCR-amplified fragments corresponding to open reading frames in genomic DNA have been used as targets (28, 29). It is not always necessary to fully sequence the cDNA clones before using them on microarrays, nor have prior information about their expression in tissues. On the other hand, if the cDNA sequence is known, it is possible to select areas of cDNAs that hybridize with higher specificity to sequences derived from one gene only and which do not hybridize with other related sequences. Many 3' untranslated sequences can also contain repetitive genomic elements that will compromise hybridization specificity and should not be present in microarray targets.

As microarray analysis will usually involve the examination of thousands of fragments in one experiment, acquiring and maintaining large collections of nucleic acid fragments is labor-intensive and expensive. While access to genetic content has previously limited, to some extent, the adoption of microarray technology, the availability of ready-printed microarray slides from both commercial companies and academic consortiums has helped alleviate this problem.

2.2.2 Preparation of DNA targets

Before using for microarray spotting, DNA targets need to be amplified and purified. Typically, PCR amplification is used, and universal primers complementary for vector sequences simplify the process (30). It is possible to amplify the target sequences starting from bacterial cultures, purified plasmids, or RNA, if reverse transcription is performed before amplification. With PCR, it is possible to amplify only part of the DNA target or clone. This allows for the removal of sequences that might compromise hybridization specificity. The amplified DNA needs to be purified to remove enzymes, nucleotides, and buffer components, all

a)

a)

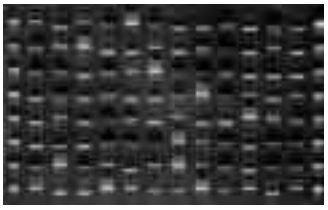


Fig 7. Separation of PCR* fragments with Ready-to-Run Electrophoresis System.

of which can interfere with the microarray analysis if present in target solution. Column purification methods, such as GFX™ PCR DNA and Gel Band Purification Kit, can be used for this purpose. Whatever methods are used for amplification and purification, it is most important to verify that the amplified fragments are the right size, do not contain other contaminating sequences and that they are present in known quantities. Agarose gel electrophoresis is a convenient way of performing this analysis. The Ready-to-Run Electrophoresis System, which is capable of separating up to 96 samples in 5 min, is well suited for this task (Fig 7).

Special care is needed when large collections of nucleic acid fragments are handled simultaneously. It has been estimated that as much as 5–30% of clones in some collections are wrongly labelled or contaminated with other sequences (31, 32). It is important that the genes identified with microarray analysis are verified with other techniques.

2.2.3 Desired properties of DNA targets

An optimal length for DNA targets is between 300–800 nucleotides. Fragments of this length can be efficiently attached to the microarray slide surface, where they form specific and stable hybrids. Figure 8 shows that the retention of UV-immobilized double-stranded DNA targets on aminosilane-treated microarray slides increases slightly with increasing length of the molecules. With increasing length, however, the concentration of DNA required to guarantee the deposition of a sufficiently high number of target molecules within a spot increases. This creates practical problems for using long DNA sequences as targets, as it can become difficult to ensure that the targets are not limiting the hybridization reaction. In order to obtain accurate results from competitive microarray hybridization, the target molecules must be in excess of the corresponding labelled probe molecules. Otherwise, hybridization signals will be saturated.

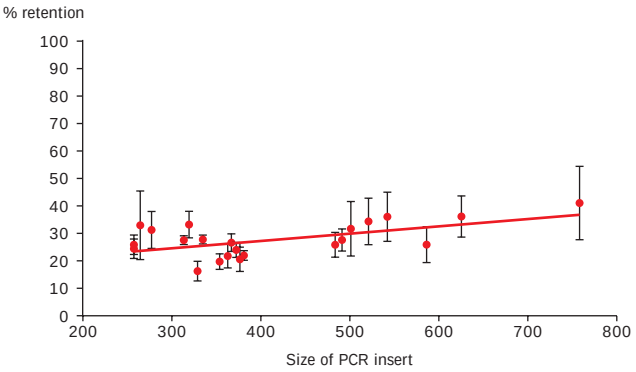


Fig 8. Retention of microarray targets of different lengths on aminosilane-treated microarray slide.

As was a requirement for oligonucleotide targets, DNA targets should not contain repetitive sequences, and they should contain sequences that are unique to one particular gene. Examination of potential cross hybridization between related sequences, such as those derived from a gene family, has revealed that more than 80% homology between targets results in hybridization signals that are not specific for one gene. However, even a lower degree of similarity was found to result in cross hybridization, suggesting that interpretation of microarray data must take the nature of the target sequences into account.

DNA targets do not need to be single-stranded. Spotting from denaturing solutions is enough to render even double-stranded targets available for hybridization. However, single-stranded DNA, which can be generated by asymmetric PCR or by exonuclease digestion of partially protected fragments (2), can also be used as targets in microarray analysis.

2.3 Control targets

Because microarray analysis is a complex process, there is a need for the use of effective controls during the whole process. For gene expression microarrays, the hybridization signal is influenced by a number of variable factors, including the number of specific transcripts in the labelled samples, the labelling method, the properties of microarray printer pens, hybridization conditions, and slide surface chemistry. Furthermore, variation in microarray signal is observed not only between different slides but also between different replica targets spotted onto different locations of the microarray slide. In this context, it is important to be able to draw conclusions from the validity of microarray results and to identify experiments that did not proceed optimally. Control sequences included on microarrays are the key factor to aid in these functions.

Different strategies have been devised for microarray control purposes. These include the use of spiked exogenous sequences of known quantities (25, 33) and housekeeping genes (34), i.e. genes whose expression is not expected to change under experimental conditions.

A control system that consists of both control targets and RNA spikes can monitor most aspects of the microarray process. A control strategy adopted in the Lucidea™ Universal ScoreCard™ combines the use of different types of control targets for spotting onto microarray slides and spikes added to samples before labelling. Together these elements cover aspects of slide printing, sample labelling, slide pretreatment, and hybridization. In a typical experiment, up to 24 replicas of the ScoreCard

sequences would be included on a slide printed with 12 pens. This number of replicas allows calculation of quality indicators that report on variation between different pens and spot sets as well as the overall dynamic range and precision of signals.

In order to gain maximum information from the quality of microarrays, positive, negative, ratio, dynamic range controls, and normalization controls are typically used. Table 1 lists the properties and main utilities of these control types. As a microarray hybridization involves thousands of different nucleic acid fragments, it is important that sequences used as controls are selected and functionally tested to avoid unspecific or cross species hybridization. Negative controls, on the other hand, need to represent different nucleic acid sequences to be able to capture the occurrence of random hybridization events. Blank spots that contain no DNA are useful as negative controls too, but are not sufficient on their own.

The use of oligonucleotides as targets allows the use of mismatched sequences as controls. By comparing the signal from the correct sequence to that from the mismatched sequence, the reliability of each signal can be assessed individually (24).

Table 1. Control target types.

Control type	Composition	Purpose
Positive control	■ Pooled genomic DNA	■ Control for labelling and hybridization success
Negative control	■ DNA fragments derived from unrelated species	■ Specificity of hybridization ■ Detection limit
Ratio control	■ Two different sequences spiked into each sample before labelling at different amounts	■ Success of labelling and hybridization ■ Color discrimination
Dynamic range control	■ Different sequences spiked into samples before labelling at different molar amounts	■ Success of labelling and hybridization ■ Color balance ■ Dynamic range of detection ■ Detection limit and saturation of signal

Chapter 3

MANUFACTURING OF MICROARRAY SLIDES

3.0 Introduction

Microarray analysis is invariably performed on a glass slide, which enables the performance of high-throughput miniaturized hybridization assays with fluorescently labelled samples—a significant improvement over the use of membrane support.

Microarray manufacture requires three distinct components:

- production method
- microarray slide
- target genetic content

In this chapter the properties of deposition methods, instruments, and microarray slides are discussed.

3.1 Production methods

3.1.1 Oligo synthesis

Two parallel approaches have been developed for the production of microarray slides. Nucleic acid targets can either be synthesized directly onto the microarray slide, or purified targets can be deposited onto a solid surface that is capable of binding nucleic acids.

Adopting a photolithographic masking method used in the semiconductor industry, oligo synthesis is begun by attaching chemically modified linker groups, which contain photochemically removable protective groups, onto the glass surface (39). By masking different predefined positions of the glass at different steps, it is possible to synthesize different oligonucleotides at different locations. Target synthesis proceeds in a step-wise fashion using a different light-impermeable mask for each round. In each step, the unprotected areas are first activated with light which removes the light sensitive protective groups. Exposure of the activated areas to a nucleoside solution results in chemical attachment of the nucleoside to the activated positions. This process is then repeated by using a different mask and a new nucleotide until all nucleotides have been added to the oligo (35).

This method (Fig 9) produces arrays of small features that are anchored at their 3' ends to the array surface. Each feature is made up of oligonucleotides that all have the same nucleotide sequence. These arrays have a high density: an area of 1.6 cm can contain up to 400 000 features. Additionally, multiple arrays can be synthesized simultaneously onto a large glass wafer, further automating the manufacturing process. The wafers are then cut into individual arrays in preparation for use.

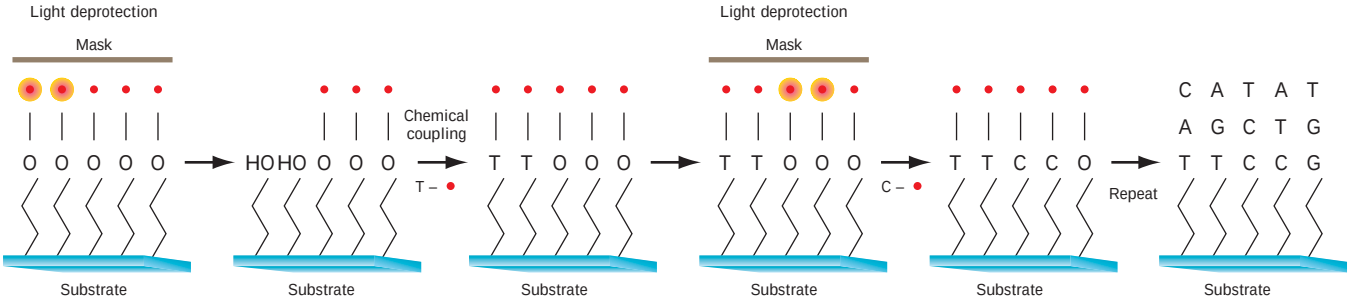


Fig 9. Microarray manufacturing using photolithography.

3.1.2 Deposition

Using common deposition methods, purified nucleic acids are attached to a modified glass slide. Typically, small volumes of nucleic acid solution—nanoliters or picoliters—are transferred onto the glass slide. Deposition methods are equally suitable for preparing microarrays containing oligonucleotides, cDNA sequences, as well as genomic DNA. Deposition methods are commonly used for preparing customized microarray slides.

The deposition chemistry involves a chemical reation between molecular groups on the glass surface and the oligo, resulting in the formation of covalent bonds that bind the ologonucleotide onto the array. There are many different suitable attachment chemistries. One is the coupling of amine-modified oligonucleotides to aldehyde slides. Another is the derivatization of 5' phosphate groups with imidazole, followed by reaction with the amine slide surface. A third is the use of bifunctional cross-linkers to couple aminated oligos to aminated glass (2).

Many different techniques have been developed for the deposition process, some of which are reviewed in this chapter. Regardless of the technique used, however, the manufacturing process should meet several criteria. Variation in the quantity of targets deposited, the shape of spots, the regularity of the array pattern, and the carryover of targets could all detrimentally affect the accuracy of microarray data.

3.1.3 Requirements of microarray spotting methods and instruments

Spot size and density

The microarray spots should be small and discernible from each other. The spots should be deposited in grid-like fashion, at equal distances from each other. It is important to immobilize the slides during printing, as even the slightest movement can distort the microarray pattern.

Spot reproducibility

The spots should be of uniform size and shape, and they should contain equal amounts of the target nucleic acids. This requires careful calibration and matching of individual printing pens.

Environmental control

Environmental conditions can have a significant effect on the quality of the spotted slides. Clean environment is important because dust particles can become fixed onto slides, causing background signals in microarray hybridization and spot finding problems during data analysis. Controlling the humidity helps to avoid changes in sample concentrations due to evaporation during printing and when spots are drying. High humidity levels may cause spots to smear whereas low humidity levels may cause evaporation from the sample plate. Under high temperature conditions targets will dry rapidly at the outer edges of the spot, thereby causing poor spot uniformity. This effect can cause a donut-shaped spot morphology. A humidity between 10–70% has been found to be most suitable for a microarray application.

Sample carryover

There must not be any carryover of previous target during the printing process.

Throughput

The printing process should be fast to allow timely printing of slides. The total time need for sample retrieval, printing, and washing of the printer pens needs to be considered.

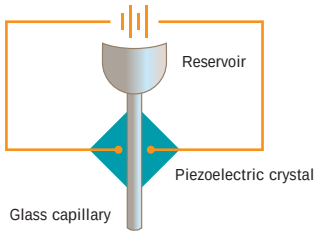


Fig 10. Diagram of a piezoelectric microarray printing system.

3.1.4 Non-contact deposition

Non-contact deposition has been adapted for microarray manufacture from the modern ink jet printing industry. As the name implies, the printing heads do not touch the surface of the microarray. Piezoelectric printing and syringe-solenoid methods are the two common variations of this method.

- In piezoelectric printing (Fig 10) the target solution is drawn into a capillary that is in contact with a piezoelectric crystal. Application of voltage to the crystal results in a slight conformational change, squeezing the capillary. A small volume of sample is deposited onto the glass surface. This method allows for very rapid spotting times. Very small volumes can be delivered, as the distortion of the crystal shape can be accurately controlled. However, this deposition method is prone to problems caused by air bubbles, which can cause poor spot morphology.
- Syringe-solenoid deposition (Fig 11) uses a syringe pump positive displacement method to deposit nanoliter volumes onto a slide. A syringe that provides the pressure source is connected to a micro-solenoid valve. The sample is drawn up the dispensing tip via the syringe. The system is pressurized and the opening of the micro-solenoid valve allows small volumes of sample to be deposited onto the surface. This system is not as rapid as that of piezoelectric printing, and it is not able to deposit sub-nanoliter volumes; however, deposition volumes are very precise and reproducible.

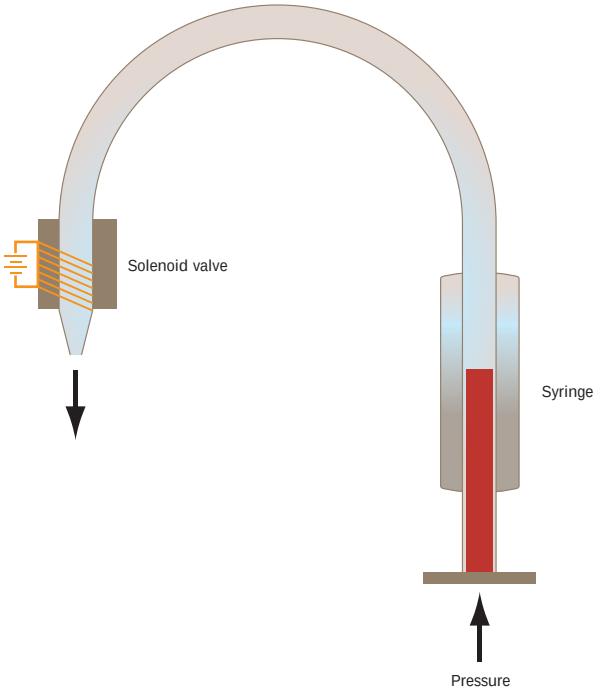


Fig 11. Diagram of a syringe-solenoid microarray printing system.

3.1.5 Contact deposition

In contact deposition, solid, hollow, or split-open pen designs are used to transfer target nucleic acid onto the slide surface. These pens are dipped into the target solution, a small volume of which adheres to the pen. When the pen comes into contact with the slide surface, a fraction of the nucleic acid solution on the pen is deposited onto the glass surface. One sample uptake of the pen allows for several spots to be printed. For achieving high throughput, several pens are used simultaneously, each of which typically deposits a different nucleic acid solution. Successful spot deposition is achieved by using pens that are quantitatively tested to ensure performance, such as those made by Amersham Biosciences.

Contact deposition requires less target nucleic acid solution than the non-contact methods and also results in smaller spots that can be packed more densely on the microarray surface.

Table 2. Comparison of characteristics of different microarray printing methods.

	Contact pen printing	Piezoelectric printing	Syringe-Solenoid printing
Microtiter plate well volume (microliters)	10–30	20–50	20–50
Uptake volume (microliters)	0.2–1.0	5–10	5–10
Spot volume (nanoliters)	0.5–2.5	5–100	0.1–10
Spot size (nanometers)	75–250	250–500	125–175

3.2 Lucidea Array Spotter

Lucidea Array Spotter (Fig 12) is a new contact deposition microarray spotter from Amersham Biosciences. The Lucidea Array Spotter is part of the Lucidea platform of products offered by Amersham Biosciences for microarray analysis. These products include Lucidea Array Spotter, Lucidea SlidePro Hybridizer (see chapter 9), and Lucidea Universal ScoreCard (see chapter 11). The key features of Lucidea Array Spotter are:

- Patent pending, stainless steel capillary pens that conserve sample and uniformly deposit picoliter volumes of target (Fig 13). From a single sample uptake of less than 200 nl, up to 150 spots can be spotted in duplicate, across each of 75 slides. The design of the pens (Fig 14) minimizes clogging with target solution and simplifies washing after each sample, resulting in no detectable carryover or mixing of samples during printing. To achieve good spot uniformity, the pens in each pen set are quantitatively tested during manufacturing to ensure performance.
- A newly designed five-step wash system eliminates the possibility of sample carryover as shown with dye-labelled DNA testing (Fig 15).
- Lucidea Array Spotter allows for monitoring and control of humidity and temperature monitoring during spotting (Fig 16).
- The target plates are kept in an area with minimized airflow to reduce evaporation while the printing is in progress.
- The spotting chamber is encased inside the enclosure of the instrument, thus limiting the access of particulates during slide printing.
- Several user-defined spotting modes are available to print arrays in up to four replicates per slide. Lucidea Spotting Pens can handle spotting fluids with significantly different viscosity.
- Control software integrates all aspects of spotter operation.

Fig 12. Lucidea Array Spotter.

Fig 13. Lucidea Spotting Pen Set.

Fig 14. Side view of the tip of Lucidea Spotting Pen. Target solution is drawn by capillary action to the narrow opening in the tip of the pen.

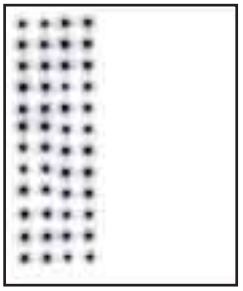
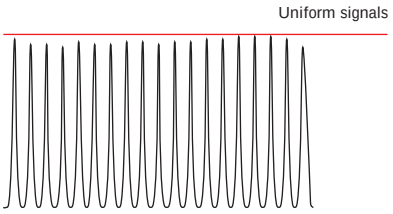


Fig 15. Wash system in Lucidea Array Spotter removes previous target solution efficiently, resulting in virtually no detectable carryover of sample during printing. The above experiment was performed with radiolabelled ^{32}P-cDNA spotted on a nylon membrane. The image on the right shows results after the pens have been washed, dipped into a sample blank, and then spotted.

Spots



Scanned

Fig 16. Controlled temperature and humidity result in relatively even spot intensities and morphology.

3.3 Microarray slides

At present the most commonly used support for microarrays are standard glass microscope slides that offer flat and rigid support with low intrinsic background fluorescence. However, there are quality differences between different manufacturer's glass slides. Careful analysis of slides before they are used for microarray printing is recommended. Furthermore, it is very important to ensure that microarray slides are absolutely clean.

3.3.1 Slide surface chemistries

Nucleic acids will not attach efficiently to an untreated glass slide. Therefore, different surface chemistries have been developed to facilitate the attachment of targets to the slide. These treatments not only enable the binding of targets, but also determine the density of molecules that can be attached per surface unit.

The uniformity and thickness of the surface coating on the slide is critical for good quality microarray results, as this will influence spot uniformity and morphology, DNA binding, as well as background signals from microarray hybridization. Variation in slide coating can contribute to the variation in microarray signals and decrease the resolution of a microarray experiment. Uneven slide coating can also lead to poor attachment of deposited nucleic acid, which may come loose during microarray hybridization.

Commonly used slide surface modifications include the introduction of aldehyde, amino, or poly-lysine groups onto the slide surface. Aminosilane slides give highly consistent and reproducible data with high signal to noise values, and they are most favorable for use in microarray experiments.

3.3.2 Common slide types

Aldehyde slides

Amino-modified DNA can be attached to microarray slides that have been modified with aldehyde groups (Fig 17). The amino group can be introduced into DNA in a PCR amplification reaction using amino-modified oligonucleotides. The aliphatic amine on the amino-modified DNA acts as a nucleophile, attacking the carbon atom on reactive aldehydes covalently attached to the surface of the slide. An unstable intermediate is converted to a Schiff base through a dehydration reaction ($-H_2O$), and the DNA is bound to the surface. To minimize fluorescent background, the unreacted aldehyde groups are reduced to non-reactive primary alcohols by treatment with sodium borohydride ($NaBH_4$). Aromatic amines on the G, C, and A bases of naturally occurring DNA can also react with aldehyde groups. The efficiency of this side reaction is ~0.01% for short oligonucleotides and ~10% for double-stranded PCR products (36).

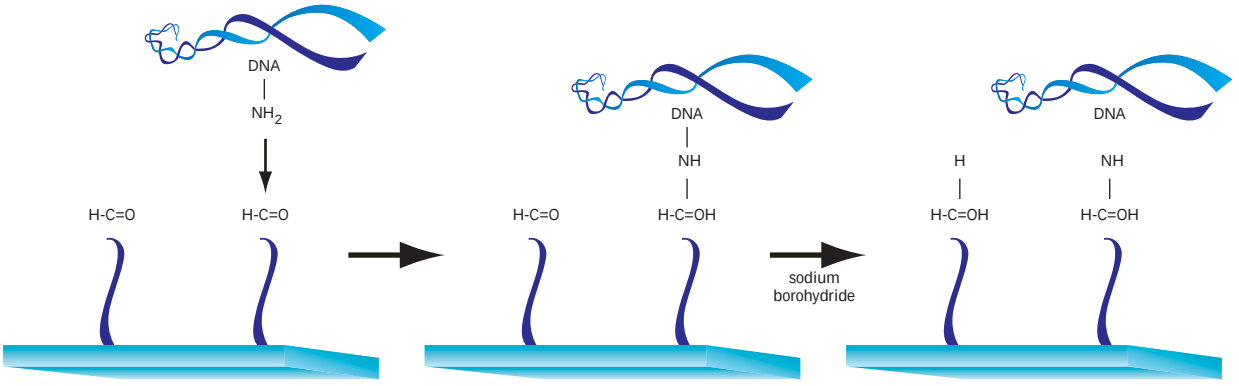


Fig 17. Attachment of amine-modified DNA to aldehyde slide.

Amine slides

Amine groups can be introduced onto microarray slides by treating cleaned glass with aminosilane, such as 3-aminopropyltrimethoxysilane (Fig 18). Vapor treatment of slides gives generally better results than deposition by a dipping method (37). Unmodified DNA can be attached to amine-modified slides, via interactions between negatively charged phosphate groups on the DNA and the positively charged slide surface. This interaction helps ensure denaturation of the DNA as well as increase its binding affinity to the slide surface. UV treatment can be used to further immobilize the DNA onto the slide surface. Attachment via electrostatic interactions is suitable for binding DNA fragments that are longer than 60–70 nucleotides. For attaching oligonucleotides to amine-modified glass, chemical coupling methods must be used (2).

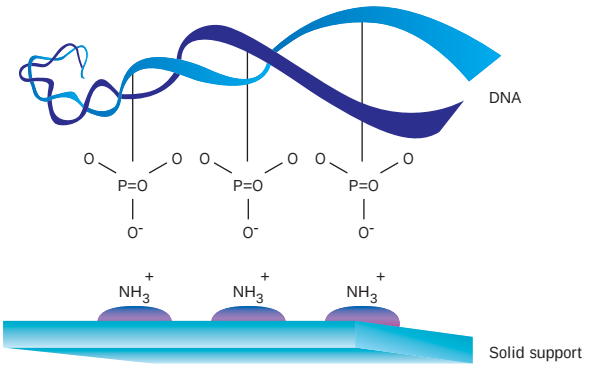


Fig 18. Attachment of unmodified DNA to an amine-modified slide surface. The DNA binds to the surface of the slide via an electrostatic interaction. The positive amines in the silane coating will attract the negative phosphate backbone of the DNA.

Poly-lysine slides

Treatment of the slide with poly-lysine creates a positively charged surface to which unmodified DNA can bind via ionic interactions (38).

3.3.3 Reflective slides

A large proportion of the fluorescent light emitted from the hybridized probe is scattered in all directions when using regular glass arrays. The introduction of a reflective surface below the spotting surface enables a significant amount of this scattered output to be directed towards the detector, hence increasing the amount of signal detected by the system. These reflective slides are constructed by adding a layer of aluminium above the glass surface. Figure 19 shows a diagram of a reflective slide.

Fig 19. Diagram of the structure of a reflective microarray slide.

Signal enhancement is further achieved if an optimal thickness of silicon dioxide is used as a spacer on top of the reflective layer (Fig 20). It has been found that a thickness corresponding to $^1/_4$ of the excitation wavelength results in optimal signal enhancement for a particular dye. In a typical microarray experiment two different dyes, such as Cy3 and Cy5, are used. As these dyes have different excitation maxima, it is not possible to enhance the excitation of both dyes simultaneously. Since the fluorescence from Cy3 is already enhanced when the dye is bound to molecules, it is more critical to increase the fluorescent signal from Cy5-labelled molecules. Hence, Lucidea Reflective Slides have been designed to specifically enhance Cy5 signals from microarray experiments, resulting in better balanced signals from both dyes.

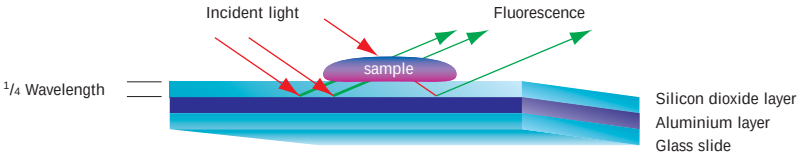


Fig 20. The enhancement of signal by a reflective slide.

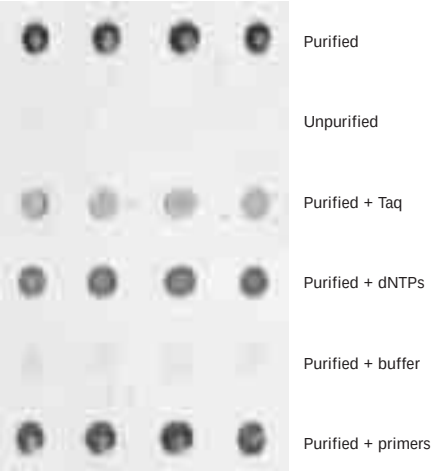


Fig 21. Purification of PCR-amplified targets. cDNA targets were amplified with PCR, then purified with column chromatography. The indicated reagents were added to the purified target DNA before spotting. Significantly decreased hybridization signals were observed from all targets containing impurities as compared with purified targets.

3.4 Target nucleic acids

The third critical component in microarray manufacturing is the target nucleic acid. Factors influencing the choice of target sequences are described in Chapter 2.

Microarray targets must be available in high enough concentration to allow a sufficient number of molecules to be deposited onto the slide. The purity of target solutions is important for both the efficient attachment of nucleic acids to the slide surface and the availability of the immobilized targets for hybridization. PCR-amplified targets must be purified to remove dNTPs, primers, DNA polymerase, buffer salts, and detergents. Column chromatography methods are suitable for this purpose. As shown in Figure 21, the presence of these compounds is detrimental to the success of microarray hybridization.

The targets, once attached to the microarray surface, are only available for hybridization when they are present in a denatured, single-stranded form. This can be achieved by spotting the targets under denaturing conditions. Typically, targets are dissolved in high salt solutions such as 3 × SSC, or in denaturing solvents such as DMSO.

3.5 Critical success factors for microarray preparation

- Always handle microarray slides in a clean environment.
- Never use gloves that contain powder as the powder will invariably get onto the slides and cause background signals.
- Never touch the array surface, only handle slides from sides.
- Only use low fluorescence microscope slides for microarray manufacturing.
- Clean microarray slides efficiently before applying slide surface treatment to them.
- Verify the purity and concentration of targets before using them for slide printing.
- Handle target plates with care to avoid drying out or mixing of targets.
- Follow the instructions provided with the microarray spotter carefully.
- Make sure that spotting is carried out under known and controlled temperature and humidity.
- Microarray slides have a limited shelf life, so prepare and use microarray slides in a timely fashion.
- Always store slides dessicated and protected from light.

Chapter 4

FLUORESCENT LABELS IN MICROARRAY ANALYSIS

4.0 Introduction

Most researchers performing microarray analysis prefer to use fluorescent dyes as labels in these experiments as their use offers high sensitivity of detection and enables detection of different dyes simultaneously. Furthermore, fluorescent dyes do not carry the hazards associated with radioactive markers. In this chapter, the general properties of fluorescence and CyDye™ fluorophores are discussed.

4.1 Definition of fluorescence

Fluorescence can be defined as the molecular absorption of light energy (photon) at one wavelength and its re-emission at another wavelength. Molecules that absorb light are known as chromophores. Molecules that both absorb and emit light are known as fluorochromes, or fluorophores.

Light is a high frequency electromagnetic wave, and the energy of the photon is inversely proportional to its wavelength (λ). Thus photons towards the blue end of the spectrum, i.e. light photons with shorter wavelengths, have a higher energy than those towards the red end of light spectrum (Fig 22).

The process of fluorescence is a three-phase one, consisting of excitation, the excited state, and emission.

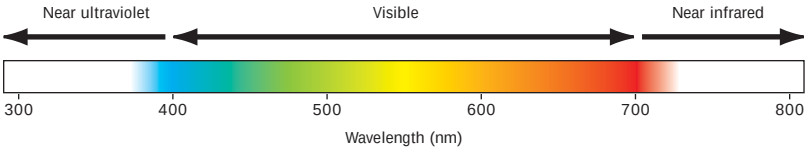


Fig 22. Light spectrum.

4.1.1 Excitation

Extinction coefficient (ε) is a measure for a fluorophore's ability to absorb light energy. When a photon of light energy (hv_{EX}) of the appropriate wavelength is absorbed by a fluorophore, an electron is boosted to a higher, unstable excited energy state. The difference between the ground state (S_0) and the higher energy state (S_n) is a property of the fluorophore; it is equivalent to the energy of the photon absorbed. Because photons with shorter wavelengths have higher energy, the shorter the wavelength of the absorbed photon, the higher the excited energy state reached by the fluorophore. The wavelength at which the fluorophore has maximum excitation is determined by the structural properties of that fluorophore.

4.1.2 The excited state

The excited state typically lasts a fraction of a second. During this state some of the energy absorbed may be dissipated in the form of vibrational and rotational energy, often resulting in localized heating. The fluorophore thus loses some of the energy it has absorbed from excitation, prior to any fluorescent emission taking place. It is for this reason that the quantum yield (ϕ) of a particular fluorophore (ratio of the number of photons emitted to the number of photons absorbed) is between 0 and 1.0. The quantum yield of a fluorophore can be greatly influenced by the medium in which it resides. For example, unconjugated Cy3 in phosphate buffered saline solution (PBS) has a quantum yield of 0.04; thus a large proportion of the energy absorbed by each dye molecule is lost to its surrounding solution. However, in glycerol the quantum yield increases more than ten-fold to 0.52.

4.1.3 Emission

Once the fluorophore has reached the lowest vibrational energy level within the electronic excited state (S_1), the electron falls from the excited state to the ground state (S_0). It is at this point in the decay process that light is emitted at a specific wavelength (hv_{EM}). Because some energy between excitation and emission has already been lost, the emitted photon has less energy than the original photon absorbed by the fluorophore (Fig 23). Therefore, the emitted light has a longer wavelength. The difference between the maximum excitation wavelength and the maximum emission wavelength is known as the Stokes shift ($hv_{EX} - hv_{EM}$).

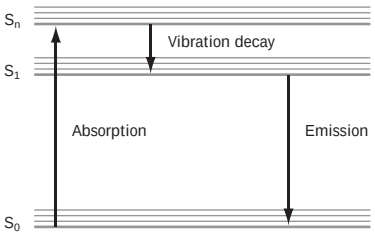


Fig 23. Diagram illustrating the energy levels of the fluorescence process.

4.1.4 Photobleaching

The fluorescent process is rapid (10^{-8} seconds) and cyclical, enabling the fluorophore molecule to be excited repeatedly. It must be considered, however, that the excited state of a fluorophore is generally more chemically reactive than the ground state. In conditions of intense light, the fluor may gradually lose its fluorescent properties, a phenomenon known as "photobleaching". This results in lower fluorescent output from the fluorophore after prolonged exposure to light or repeated excitation. For more photosensitive dyes, such as fluorescein, photobleaching may be a significant factor when using instrumentation with high laser power. In contrast, as seen in Figure 24, CyDye fluors are more resistant to photobleaching, which makes them more suitable for multiple applications.

An excellent approach to reduce photobleaching is to maximize detection sensitivity so that the excitation intensity can be reduced.

The brightness (intensity of output) of a fluorophore is proportional to both the extinction coefficient (ε, the molecule's ability to absorb light energy) and the quantum yield (ϕ, the molecule's efficiency to re-emit light). Both of these are constants under specific static environmental conditions. Consequently, fluorophores with very different characteristics may give a comparable signal brightness. Fluorescein, for example, has a molar extinction coefficient of ~70 000 and a quantum yield of ~0.9, whereas Cy5 has values of ~200 000 and 0.3 respectively. However both are of similar brightness.

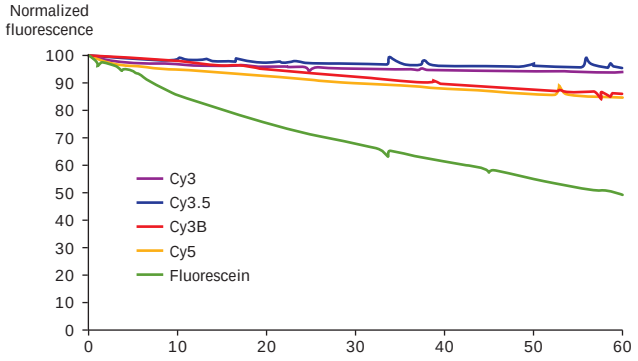


Fig 24. Photostability of fluorophores under "natural" light conditions.

4.2 Applications of fluorescence

The fluorescent process, combined with appropriate imaging instrumentation, enables a sensitive and quantitative detection method for microarray expression analysis.

4.2.1 Benefits of fluorescent labelling

The particular Stokes shift properties of individual fluorophores make it possible to separate excitation light from emission light with the use of optical filters. Good spectral separation enables high sensitivity of detection and yields low background.

By choosing fluorophores that have different pairs of excitation and detection wavelengths, it is possible to excite and detect multiple dyes in the same sample. This method enables multiplexing with dyes—labelling two or more samples with different dyes that have different absorption and emission spectra—and makes it possible to analyze several samples simultaneously. Figure 25 shows the absorbance and emission spectra of Cy3 and Cy5, the most widely used pair of fluorescent dyes in microarray analysis. The minimal overlap between the Cy3 and Cy5 spectra demonstrates how it is possible to detect both dyes with minimal cross talk (overlap) between their respective signals.

Another advantage of using fluorescent dyes as labels is that they are less hazardous than radioactive compounds and offer significantly increased stability, or longer shelf life. The availability of fluorophores conjugated to many different chemical groups enables the labelling of nucleic acids, proteins, lipids, and carbohydrate molecules. Furthermore, fluorescent dyes can also be used as reporters to detect changes in pH, ion concentrations, or dye environment.

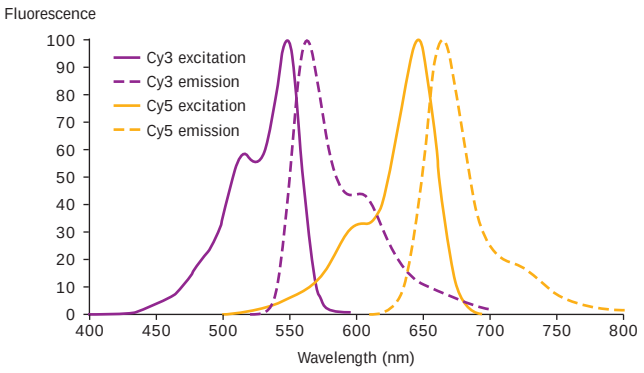


Fig 25. Excitation and emission spectra of Cy3 and Cy5 dyes.

4.2.2 Fluorescent quenching

Fluorescent quenching (Fig 26) causes decreased fluorescent signals. It occurs when two or more fluorophores are in close proximity to each other and the excitation energy is dissipated in interactions between the adjacent dye molecules, rather than emitted as fluorescent light. Chemical structures as well as photochemical properties of dyes determine the distance at which two fluors will quench.

Quenching can occur when samples have been labelled too densely or when too much labelled sample is used in hybridization. Over-labelling not only results in the loss of linearity between fluorescent signal emitted and the number of fluors present, but will, in extreme cases, reduce the signal to levels that cannot be observed. In microarray analysis, quenching may also occur when two probe strands come into close proximity of each other. This is likely to be most apparent in the presence of highly expressed transcripts, where a very large number of labelled molecules are bound densely at a precise location. See chapter 6, section 6.2.4 for further overview of this process.

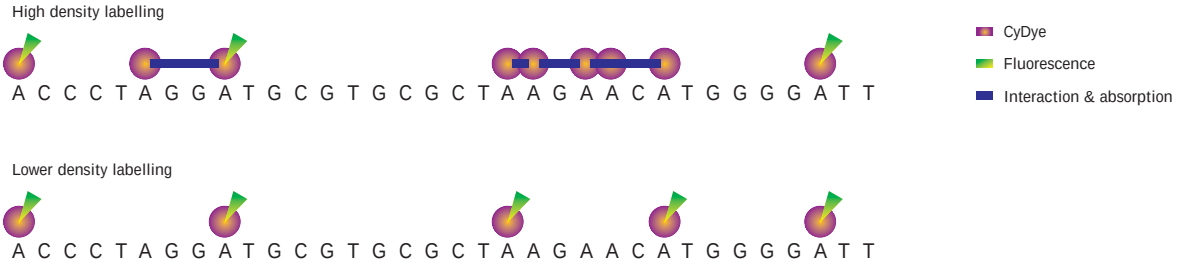


Fig 26. A schematic representation of the effect of quenching under varying labelling conditions. Fluorescent signal from the same nucleic acid fragment labelled with high and low densities is depicted. Green shows fluorescent emission, whereas purple denotes dissipation of energy between closely spaced fluors (pink circles).

4.3 CyDye fluorophores

4.3.1 Chemical structure of cyanine dyes

The cyanine, or CyDye, family of fluorescent dyes were first used in the photographic industry as film sensitizers. They were subsequently discovered for use in molecular biology applications when a CyDye was coupled to a N-succinimide ester, to form the first dUTP (40, 41).

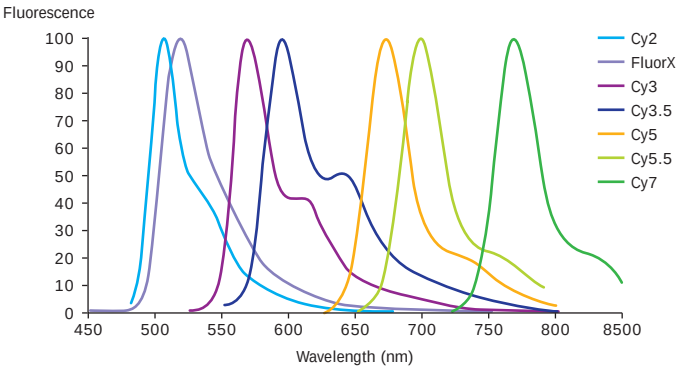
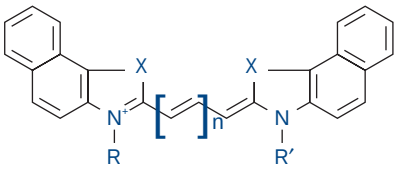


Fig 27. The emission spectra of some CyDye fluors.



X = O, NR, $C(CH_3)_2$, S, Se
R and R′ = alkyl or aryl
n = 0, 1, 2, 3, 4 or 5

Fig 28. Core structure of the cyanine dyes.

This family of fluors consists of a chemically related group of dyes whose emission spectra span the spectrum of visible light (Fig 27). CyDye fluors share a core structure consisting of two heterocyclic indocyanine ring structures joined by a polymethine bridge (Fig 28). Each dye differs in the structure of this bridge. Adding pairs of conjugated C atoms to the polymethine chain results in a wavelength shift of ~100 nm, for example Cy3 (550 nm) and Cy5 (650 nm). An important modification of Cyanine dyes is sulfonation of the indocyanine rings, as shown in Figure 29. The sulfonate acid groups increase the solubility of the dyes. In addition they reduce aggregation of dye molecules, as the introduction of negative charge makes the dye molecules repel each other. This results in a decrease of fluorescence quenching.

The multiplexing properties of CyDye fluors were further increased by synthesizing Cy3.5 and Cy5.5 (Fig 30). The addition of benzene rings shifts their absorbance and emission spectra up by approximately 30 nm to the red. Two additional sulfonate groups are needed to increase solubility in order to overcome aggregation due to benzene rings.

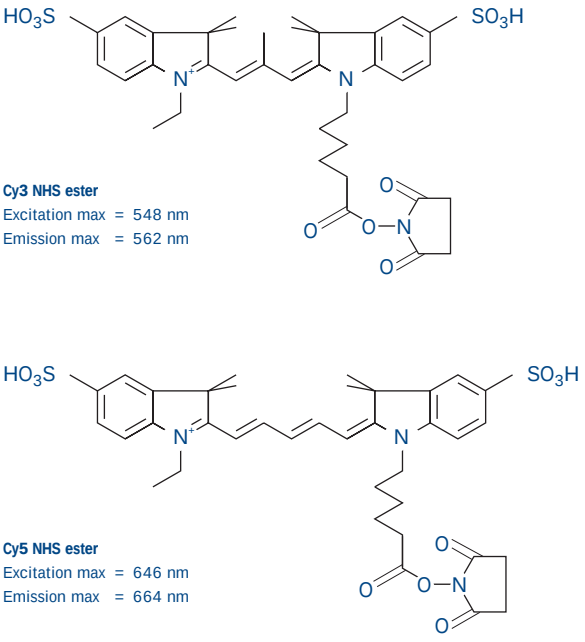


Fig 29. Structure of Cy3 NHS ester (3 carbon bridge) and Cy5 NHS ester (5 carbon bridge).

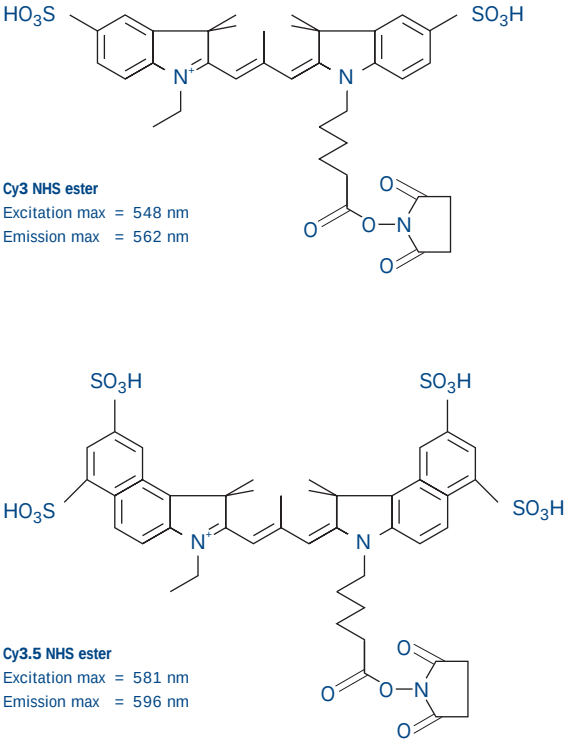


Fig 30. Structures of Cy3 and Cy3.5 NHS esters.

4.3.2 Fluorescent and chemical properties of CyDyes

Cy3 and Cy5 have become the most commonly used pair of CyDyes. This can be attributed to the following factors:

- The photostability of Cy3 and Cy5 is higher than that of other widely used dyes (Fig 24), making the use of CyDye fluors more practical.
- Cy3 and Cy5 are bright dyes that give strong fluorescent signals.
- The good spectral separation of Cy3 and Cy5 means that each can be excited at a different wavelength and their emissions can be detected separately. This requires the use of two different lasers, typically a 532 nm and 633 nm laser for Cy3 and Cy5, respectively. By choosing optical filters that only collect emitted light from part of the spectra, Cy3 and Cy5 signals can be measured with minimal overlap.
- The fluorescence of Cy3 and Cy5 is minimally affected by factors such as pH and the presence of DMSO. Additionally, these dyes can withstand temperatures and conditions normally encountered in molecular biology applications.

4.3.3 Handling of CyDye fluors

Correct handling of CyDye and CyDye compounds helps to conserve their fluorescent properties. When handling these substances, observe the following precautions:

- Minimize the exposure of CyDye compounds to all light sources.
- Store CyDye in amber tubes, in light-safe containers, or wrapped in aluminium foil to protect them from light during storage.
- Take CyDye out of their protective container only when ready to use and return to dark immediately after use.
- If using CyDye nucleotides, prepare single use aliquots to avoid freeze thaw.
- Protect CyDye NHS esters and other CyDye labelling conjugates such as CyDye Direct™ from moisture by storing dessicated.
- Do not store CyDye in solutions containing concentrated amines. Phosphate buffer or water is preferred over Tris-buffers.

Table 3. Photochemical properties of selected CyDye fluors.

CyDye	Abs. Max. (nm)	Em. Max. (nm)	Molar extinction coefficient ($M^{-1}cm^{-1}$)
Cy2	489	506	~150 000
Cy3	550	570	150 000
Cy3.5	581	596	150 000
Cy5	649	670	250 000
Cy5.5	675	694	250 000
Cy7	743	767	~250 000

4.4 Use of fluorescent dyes in microarray analysis

Multiple labelling strategies have been developed for incorporating fluorescent labels into microarray probes. These techniques are discussed in chapter 6. The photochemical properties of fluorescent dyes, especially the positions of their excitation and emission peaks, determine the specification of scanning instruments, the laser type, and the choice of emission filters that are required to separate and detect fluorescent signals from particular dyes. The popularity of pairing Cy3 and Cy5 as labels has led to the development of microarray scanning instruments that are optimally suited for detection of these dyes.

General requirements for detecting fluorescent signals in microarray analysis are:

- High enough resolution to image a large area (2 × 6 cm) in a short period of time.
- At least two fluorescent spectra must be distinguished to accommodate differential gene expression experiments using two fluorescent dyes.
- The wide range of message abundance levels requires an instrument with a low fluorescent detection threshold to allow detection of rare messages and wide linear dynamic range to measure the more abundant messages.
- The entire area of the microarray must be scanned uniformly to ensure reproducibility.

4.4.1 Characteristics of fluorescent detection

The energy of the emitted fluorescent light is a statistical function of the available energy levels in the fluorochrome, but it is independent of the intensity of the light used to excite the fluorophore. In contrast, the intensity of the emitted fluorescent light varies with the intensity and wavelength of incident light and the brightness and concentration of the fluorochrome. When more intense light is used to illuminate a sample, more fluorochrome molecules are excited, and the number of photons emitted, i.e. the number of electrons falling to the ground state, increases. If the illumination is very intense, all the fluorochrome molecules are in the excited state most of the time—this is called saturation.

When the illumination wavelength and intensity are held constant, as with the use of a controlled laser light source, the number of photons emitted is a linear function of the number of fluorochrome molecules present. At very high fluorochrome concentrations, the signal becomes non-linear because the fluorochrome molecules are so dense that excitation occurs only at or near the surface of the sample. Additionally, some of the emitted light is reabsorbed by other fluorochrome molecules (self-absorption).

The amount of light emitted by a given number of fluorochrome molecules can be increased by repeated cycles of excitation. In practice, however, if the excitation light intensity and fluorochrome concentration are held constant, the total emitted light becomes a function of how long the excitation beam continues to illuminate those fluorochrome molecules (dwell time). If the dwell time is long relative to the lifetime of the excited state, each fluorochrome molecule can undergo many excitation and emission cycles.

Measuring fluorescent light intensity (emitted photons) can be accomplished with any photosensitive device. For example, for detection of low-intensity light, a photomultiplier tube (PMT) can be used. This is simply a photoelectric cell with a built-in amplifier. When light of sufficient energy hits the photocathode in the PMT, electrons are emitted, and the resulting current is amplified. The strength of the current is proportional to the intensity of the light detected. The light intensity is usually reported in arbitrary units, such as relative fluorescence units (RFU).

If fluorescent samples are detected with a system that uses an intense light source to excite the dyes, photobleaching can occur. This irreversible destruction of an excited fluorophore will result in a loss of brightness, or emission light intensity. As photobleaching is a consequence of excitation, altering the characteristics of detection, such as increasing the voltage of PMT to allow more sensitive detection, will not cause it. Microarray slides hybridized with Cy3- or Cy5-labelled probes can be scanned several times with commercial microarray scanners without considerable loss of fluorescent emission.

Chapter 5

PREPARATION OF RNA SAMPLES FOR MICROARRAY ANALYSIS

5.0 Introduction

There are multiple steps involved in isolating and preparing RNA samples to be used for microarray analysis. Discussed in this chapter are factors affecting the quality of analyzed RNA, protecting RNA samples from contamination and degradation, purifying RNA, and finally, characterizing the purified RNA.

5.1 Factors affecting RNA quality

The quality of information obtained from microarray experiments is primarily dependent upon the quality of RNA analyzed. Ideally, the RNA should be devoid of DNA, protein, carbohydrates, lipids, and other compounds. The presence of these substances will not only make it difficult to correctly estimate the amount of RNA present in the sample, but can contribute to fluorescent background signals in the array hybridization. Degradation of RNA, whether by enzymatic or chemical means, results in the loss of gene expression information from the labelled samples. Furthermore, if the quantity or quality of the two samples being compared differ, misleading conclusions can be made.

Compared with DNA, RNA is relatively unstable and can be degraded either enzymatically, chemically or physically. Ribonuclease enzymes (RNAses) degrade RNA into short oligonucleotides in a rapid reaction. They are present in all cells and can be derived from a variety of environmental sources, such as the hands, skin, and hair; bacterial or viral contamination of solutions; or remnants of previous reagents in lab glassware. Inactivation of these enzymes is difficult; therefore it is essential that precautions are taken to ensure that RNA degradation is minimized during isolation, purification, and storage.

5.2 Protecting RNA from degradation

RNA can be degraded if it comes into contact with any source of RNAse. Discussed in this section are the many ways in which RNA samples can become contaminated and how to protect them from subsequent degradation if contamination occurs.

5.2.1 Protecting RNA from contamination by environment

An easy way of accidentally contaminating RNA preparations is by transferring nucleases from the investigator's hands, skin, or hair to the sample. The transfer of RNAses can also take place via equipment, bench surfaces, and door handles. Common molecular biology protocols, such as plasmid preparation, involve the use of high amounts of RNAses and can lead to RNAse contamination if handled in the same location. In order to protect RNA samples from contamination, the following precautions are recommended:

- Wear disposable gloves while handling RNA samples and while preparing other solutions for RNA work. Use a clean pair of gloves if potential contamination of any kind occurs.
- Perform RNA work in a separate area of the laboratory where no RNAses are allowed. Before RNA handling, clean the bench surface with a detergent, such as RNAse ZAP™ (Ambion).
- Lab equipment, such as tissue homogenizers, non-disposable centrifuge tubes, gel tanks, and trays that can come into contact with RNA, should be reserved for RNA work. If this is not possible, large equipment or those items made from materials that do not withstand autoclaving temperatures should be cleaned with RNAse inactivating detergents.

5.2.2 Protecting RNA during preparation of reagents

In order to obtain good quality RNA and to maintain its integrity during subsequent analysis, all reagents should be prepared so that they do not contain any traces of RNAses.

- Use only disposable plastic tubes and pipette tips for RNA work. Clean plastic-ware should be baked at 120 °C for at least 20 min to reduce RNAse (and other nuclease) contamination. Before baking, pack centrifuge tubes into glass beakers and cover them with aluminium foil. Additionally, wrap tip racks in foil to keep the baked items free of contamination after treatment.
- Set aside reagents for use in RNA work only. Always use baked or disposable spatulas and weighing trays for measuring out reagents.

- Treatment with diethyl pyrocarbonate (DEPC) renders RNAses inactive and can be used to clean solutions and labware before use in protocols involving RNA. As DEPC is toxic it should be handled with appropriate care. Prepare 0.2% (v/v) solution of DEPC in water and soak clean equipment and glassware in the solution for at least 1 h. Rinse the treated equipment with sterile water and let dry in a clean place where no further contamination is encountered. Finally, autoclave the equipment/labware in sealed autoclave bags. This is necessary as autoclaves become contaminated by RNAses from other autoclaved materials or from the water used in the autoclaving process.
- Reagent solutions can be treated by adding 0.2% (v/v) DEPC into the solution. After the solution has been left to stand for a couple of hours, it can be autoclaved to remove DEPC by heat degradation. Solutions containing Tris-buffers cannot be prepared in this way. However, water that has been treated with DEPC and then autoclaved, can be used to make up any Tris-buffers.
- Use DEPC-treated water in all RNA protocols.
- Store sterilized equipment and solutions unopened in a clean environment, away from any potential sources of contamination.

5.2.3 Protecting RNA during isolation process

Isolation of RNA requires disruption of cellular structures, which leads to the release of RNAses. Rapid degradation of RNA follows if these RNAses come in contact with RNA during cell homogenization. The following protocols are recommended:

- Use strong denaturing agents such as guanidium isothiocyanate (GITC) or guanidium hydrochloride in RNA extraction protocols as they will denature RNAses (and other enzymes) efficiently and quickly.
- RNAse inhibitor, which is a placental protein that inactivates RNAses by binding to them, can also be used during RNA isolation. However, the effectiveness of the RNAse inhibitor can be affected by the composition of extraction solutions, and denaturation of the inhibitor can release active RNAses.
- The longer the time elapsed between collecting cells or tissue and preparing a denatured homogenate, the more chance there is for RNA degradation to take place. Furthermore, during this time, changes in gene expression can also take place as cells react to the change in their environment. Therefore it is good practice to work as quickly as possible when preparing biological samples for RNA extraction.

- Harvest cultured cells directly into a solution that contains GITC to ensure minimal degradation. Pipette cell lysis solution directly onto a washed cell monolayer. This will lead to immediate cell lysis. Scrape the cells off and transfer into a centrifuge tube. The lysate can be homogenized by drawing it several times through a needle with a syringe or by vigorous mixing until the lysate becomes clear and homogeneous. Cell lysates prepared in this way can be stored frozen, preferably at -70 °C, until RNA isolation protocol is completed.
- Freeze tissue samples rapidly in liquid nitrogen, as this helps to minimize RNA degradation. It is advisable to cut large samples into small pieces before freezing them as this will greatly facilitate their subsequent use for RNA extraction. These samples must be stored frozen, preferably in liquid nitrogen. Thawing of the samples, at any stage, will result in RNA degradation as the freezing process causes some cellular damage. For best results, mechanically pulverize frozen tissue samples while they are kept cold with liquid nitrogen. This can be done with a pestle and mortar. The cold powder should then be dissolved in lysis buffer containing GITC to prepare a clear homogenate.

5.2.4 Protecting RNA during storage and handling

RNA should be protected during storage and handling. The following protocols are recommended:

- RNA is not stable at alkaline pH. This property can be exploited to degrade RNA selectively from DNA. However, if the aim is to maintain RNA intact, all solutions that come into contact with RNA should be neutral or mildly acidic. As the pH of laboratory water can vary, using dilute buffers, such as TE pH 7.6, is recommended.
- RNA degradation is more rapid at high temperatures. For long term storage, storing RNA solutions at -70 °C is recommended. All RNA solutions should be stored on ice while working with them and kept thawed for the minimum time needed. As with other nucleic acids, avoiding freeze-thaw cycles is important. If large amounts of RNA are prepared at one time, it is recommended that the purified nucleic acid is aliquoted for storage and only the required number of aliquots are thawed at any time.

5.3 Choosing an RNA isolation method

5.3.1 Methods for purifying total RNA

Numerous RNA isolation methods have been published and a variety of RNA isolation kits are available (42, 43). The key criteria in choosing a method should be to achieve a high yield of intact and pure RNA. Obtaining long RNA molecules can be problematic, and the choice of purification method will influence results (44).

5.3.2 Critical factors in RNA isolation

The main factors in isolating good RNA are the composition of the cell lysis buffer, the method of cell disruption, and the method used for separating RNA from protein, DNA, and other compounds. The nature of the biological sample is also relevant as some RNA isolation methods may be more suitable for certain tissues. Whereas soft tissues or cultured cells disrupt quickly and efficiently, and methods using mild cell lysis buffers can give good results, harder tissues containing large amounts of connective tissue, such as muscle, will require the use of strong chaotropic agents such as GITC. Amersham Biosciences RNA extraction kits, such as QuickPrep™ Total RNA Extraction Kit, RNA Extraction Kit, and QuickPrep *Micro* mRNA Purification Kit, all contain either guanidium hydrochloride or guanidium isothiocyanate in the lysis buffer, and are suitable for use with a wide variety of cells and tissues. In general, RNA preparations that use chaotropic agents in the lysis buffer tend to give the best results. However, the protocols are more laborious, involve the use of toxic chemicals, and take longer.

Choosing an efficient cell/tissue disruption method for RNA extraction is important. If cell lysis is not complete, the yield of RNA will be compromised. Mechanical disruption using tissue homogenizers, vigorous vortexing, needle and syringe, sonication, pestle and mortar, and bead milling are among commonly used methods. The main considerations in choosing a disruption method are the amount of each sample and the time that is needed for preparation of a cell lysate.

A variety of techniques can be used to differentially separate RNA from other cellular compounds. Precipitation with lithium chloride, acid extraction phenol/chloroform, binding to an absorbent matrix or cesium chloride gradient centrifugation can be used successfully to purify RNA. However, the quality and quantity of RNA obtained with these methods can vary.

The presence of contaminating DNA in total RNA samples can cause problems in microarray analysis. Most labelling methods will label both RNA and DNA with equal efficiency. Labelled DNA can hybridize with microarray targets and can lead to high level hybridization signals that are not derived from transcripts. RNA cannot be quantitated separately from DNA, so an accurate estimation of the amount of RNA in contaminated RNA preparations is impossible. Therefore, it is advisable to treat total RNA preparations to remove DNA contaminants before using the RNA for labelling. This can be achieved with DNase I treatment or by using CsCl gradient centrifugation to separate RNA from DNA.

5.3.3 Isolating RNA from difficult samples

The nature of some biological samples may necessitate the use of modified RNA extraction strategies to avoid contamination of RNA samples with other compounds. For example, plant tissues can contain polyphenols and polysaccharides. Precipitation with polyvinyl pyrrolidone can be used to remove these substances from RNA preparations. The hardness of cell walls and outer protective structures can also pose a problem. Freezing the samples followed by mechanical grinding may be necessary to efficiently disrupt cell walls and to release cellular RNA. In some cases, as with yeast that has cell walls that can form capsules, disruption of cellular structures increases access to RNA. Digestion with enzymes, such as zymolase, can be used to weaken the cell walls before mechanical disruption with bead milling to lyse the cells. Similarly, isolation of bacterial RNA benefits from the use of enzymes that digest and weaken outer supportive structures. Lysozyme treatment followed by mechanical bead milling is a suitable approach for disrupting bacterial cells for RNA extraction.

5.3.4 Purification of eukaryotic mRNA

Most eukaryotic transcripts contain a poly-A tail, and this property can be exploited to separate transcripts from other RNA molecules. Incubation of total RNA with oligonucleotides containing a poly-T sequence, otherwise called oligo(dT), will result in the hybridization between the poly-A tail of transcripts and the oligonucleotides. By attaching the oligo(dT) to a solid support, it is possible to specifically separate transcripts away from other RNA molecules. QuickPrep *Micro* mRNA Purification Kit uses oligo(dT) cellulose for extraction of mRNA.

Although purification of mRNA lengthens the sample preparation protocols, it provides several benefits for microarray analysis:

- Probes prepared from mRNA usually give higher signal to noise values on arrays than probes prepared from similar amounts of total RNA.
- Total RNA preparations are more likely to contain compounds other than RNA, which can interfere with the labelling or hybridization steps.
- The yield of labelled cDNA is higher from mRNA than from total RNA, because alternative priming strategies that use oligo(dT) can be used.
- It is easier to prepare labelled probes corresponding to the 5' ends of transcripts from mRNA populations than from total RNA.

5.3.5 Purification of prokaryotic mRNA

Purification of mRNA from prokaryotes is difficult as most transcripts lack poly-A tails. However, strategies have been developed to polyadenylate 3' ends of bacterial transcripts in crude extracts.

Enrichment for bacterial mRNA can also be achieved by selective degradation of ribosomal RNA. By synthesizing first-strand cDNA selectively from ribosomal RNA with the use of specific primers, RNAse H can be used to degrade the RNA strand in the resulting double-stranded hybrid. DNAse I digestion can then remove the DNA strand, resulting in the enrichment of transcripts. Up to 80% enrichment can be achieved with this method (45).

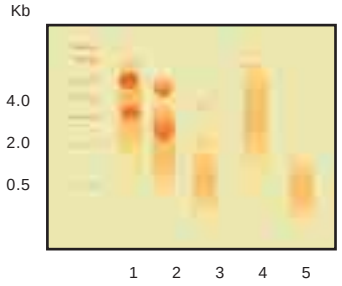


Fig 31. Schematic illustration of typical results obtained from analyzing the quality of RNA with a denaturing agarose gel. Denatured RNA samples were electrophoretically separated in an agarose gel and visualized with UV light after ethidium bromide staining. Lane 1. Intact total RNA with both ribosomal RNA bands present as sharp and bright bands. High abundance transcripts can be discerned as distinct bands. Lane 2. Partially degraded total RNA. Although ribosomal bands are still visible, the average size of RNA is smaller, and no distinct transcript bands are visible. Lane 3. Badly degraded total RNA. Most of the RNA is shorter than 1 kb, and ribosomal RNA bands appear as smears. Lane 4. Intact mRNA. Fragments longer than 4 kb should be visible and abundant transcripts should appear as distinct bands. Lane 5. Degraded mRNA. The transcripts appear as a fast migrating smear.

5.4 Characterization of purified RNA

Microarray gene expression data is derived from a comparison of hybridization signals obtained from two samples. Accurate results are only obtained when the two samples are of the same quality. Comparison of a partially degraded RNA sample to an intact sample can artificially show some genes as being more highly expressed in the better quality sample. Likewise, differences in the amount of RNA in different samples can also give biased data. Therefore, it is highly advisable to verify both the quantity and quality of the RNA or mRNA sample before their use in microarray analysis.

5.4.1 Measuring the amount of RNA

The amount of RNA can be quantified by measuring the absorbance of RNA solution at 260 nm. Pure RNA solution that contains 40 µg of RNA per mL will give absorbance of 1 AU. This method works well with clean RNA samples that are devoid of other contaminating substances. Unfortunately, this is rarely the case with total RNA samples purified with simple methods. Proteins and DNA or other compounds, such as those released from some affinity chromatography columns, will absorb at 260 nm. This absorbance is indistinguishable from that of RNA and will give an artificially high estimate for the amount of RNA in the sample. By measuring absorption spectra from 200 to 350 nm, some conclusions on the purity of RNA can be made. However, the presence of DNA in the sample can still go undetected. Therefore, it is important to use an RNA isolation method that specifically removes DNA from the purified sample.

5.4.2 Verifying the quality of RNA

Agarose gel electrophoresis performed under denaturing conditions can be used to analyze the quality of RNA. Gels containing formaldehyde have been traditionally used for this purpose, but denaturation of RNA hairpins by glyoxal has gained popularity, as this method does not involve the use of large quantities of harmful chemicals (42, 43, 46). Both methods, followed by staining of the gels with nucleic acid binding stains, such as ethidium bromide or Vistra Green™, are useful for observing overall differences in RNA quality. It is relatively easy to see if a sample is badly degraded as the ribosomal bands appear as smears. However, it may be difficult to detect more subtle degradation of transcripts unless large amounts of sample are used. Figure 31 shows a schematic illustration of typical results from RNA gel electrophoresis. Transferring the gel onto a membrane for Northern blotting analysis can give more precise information, as the status of specific transcripts in different samples can be analyzed.

Reverse transcriptase PCR (RT-PCR) offers a convenient, fast, and versatile method for obtaining information regarding the quality of RNA preparations from small sample amounts. First-strand cDNA synthesized with oligo(dT) priming can be used as PCR template.

By choosing specific primer pairs, it is possible to determine whether different transcripts are intact in the samples. Performing PCR amplification for a limited number of cycles, or with real time detection, makes it possible to estimate the relative amount of specific transcripts in different samples. By amplifying transcripts derived from genes whose expression is not expected to vary under experimental conditions, it is also possible to compare the amount of mRNA in the different samples. By choosing several pairs of primers from one preferably long transcript, and targeting its 5' central and 3' regions, it may be possible to observe partial degradation of samples. Performing PCR with primers that are derived from gene introns can reveal the presence of genomic DNA.

5.5 General recommendations for preparing RNA samples for microarray analysis

In conclusion, the following general recommendations for preparing RNA for microarray analysis are given:

- Minimize the degradation of RNA at all handling stages.
- Choose an RNA purification method that gives good yields of pure and intact RNA from your samples, even if this means using a complicated protocol.
- Measure the amount of RNA before using it for microarray labelling.
- Verify the quality of the RNA before using it for microarray labelling.
- If possible, purify mRNA for use in microarray analysis.
- Prepare all the samples for microarray analysis with the same protocol.

Chapter 6

SAMPLE LABELLING FOR GENE EXPRESSION ANALYSIS

6.0 Introduction

In differential gene expression analysis two or more RNA samples are compared to identify differences in the abundance and identity of the transcripts they contain. In order to convert the information contained in the transcript populations into a form that can be hybridized with microarrays and subsequently detected, the transcript populations need to be labelled. This can be achieved using different methods; an ideal method retains both the information carried by the identity of the transcripts as well as their relative abundance in the sample.

6.1 The diversity of transcript populations

Messenger RNA molecules, otherwise called transcripts, carry the genetic information encoded in genes. In most cells these transcripts constitute only a small proportion of the total RNA, whereas ribosomal and transfer RNA account for more than 98%. In any cell type, the transcript population typically consists of thousands of distinct transcripts, most of which are transcribed from different genes (although splice variants of genes exist too). These transcripts can be present in widely varying amounts ranging from just a few copies per cell to thousands of copies. Furthermore, the relative levels of transcripts are constantly changing as the cell responds to different environmental signals. The amount of transcripts is estimated to follow a normal distribution in which a small number of genes are expressed at high or very low levels. The majority of the genes are expressed at intermediate levels.

Genes come in different sizes, with different numbers and sizes of exons. The size of transcripts reflects this by varying from a few hundred nucleotides to up to about 20 000 nucleotides. Average length of transcripts is estimated to be between 1.5–1.7 kb.

6.2 Requirements of labelling methods

6.2.1 Retaining gene expression information

The labelling methods used in microarray analysis must cope with the inherent diversity of transcript sequences and create representations that contain all the information present in the original transcript population. Thus an ideal labelling system is neither biased towards any nucleotide sequences, nor does it label differently transcripts of different sizes or sequences that are expressed at different levels.

In reality, existing labelling methods do not convert all information into labelled form. Enzymatic methods are limited to copying certain nucleic acid sequences, whereas the instability of some transcripts is a general problem for all methods.

6.2.2 Length of labelled fragments

Accurate information about gene expression can only be deduced from microarray experiments if the labelled nucleic acids can hybridize efficiently and with specificity to their complementary targets. The length of the labelled fragment is an important factor in determining these parameters. Fragments longer than 100 nucleotides can hybridize strongly enough with their target sequences to withstand stringent hybridization and wash conditions. The hybridization kinetics of shorter fragments is faster. For optimal hybridization, probes consisting of fragments of 200–500 nucleotides long are recommended. Longer fragments may not find their targets as efficiently as shorter fragments, but will produce a higher signal when hybridized as they carry more labelled molecules.

6.2.3 Yield of labelled probe

The amount of labelled probe prepared by the labelling method is important to the sensitivity of microarray experiments. This is because the efficiency of the labelling process is critical in determining the lowest amount of mRNA that can be used to generate detectable signals from microarrays. With higher amounts of mRNA available, differences in probe yield from different labelling methods can make the difference between being able to hybridize one or several slides with one probe. Ideally, the labelling method should transform each transcript into a labelled fragment, without any bias towards more highly expressed sequences. If the labelling method results in net amplification of nucleic acid in the labelling process, the amplification process should be linear, i.e. the original ratios of expression levels within the sample should not be changed in the amplification process.

6.2.4 Optimum labelling density

The fluorescent labels used in microarray analysis bring their own restrictions to labelling protocols. If two or more fluorescent molecules are in close proximity of each other, a significant portion of the absorbed light energy can be spent on interactions between different molecules and dissipated as heat. This will result in less than the expected amount of fluorescence being emitted from the sample, and moreover, the amount of fluorescence is no longer directly proportional to the number of fluors in the sample. This phenomenon is called “quenching”, and it is an inherent property of fluorescent molecules. Each fluorophore has slightly different quenching properties that are determined by its chemical structure. In practical terms this means that for each fluor there is an optimal labelling density, or distance between attached labels, which will produce maximum fluorescence from a labelled nucleic acid fragment. Exceeding this optimum labelling density results in decreased fluorescent signal (Fig 32). Therefore, in order to achieve highest sensitivity of detection, the labelling method used in microarray experiments should be optimized to yield fragments that are labelled at the maximum density as determined by the labelling fluors. See chapter 4, section 4.2.2 for additional information.

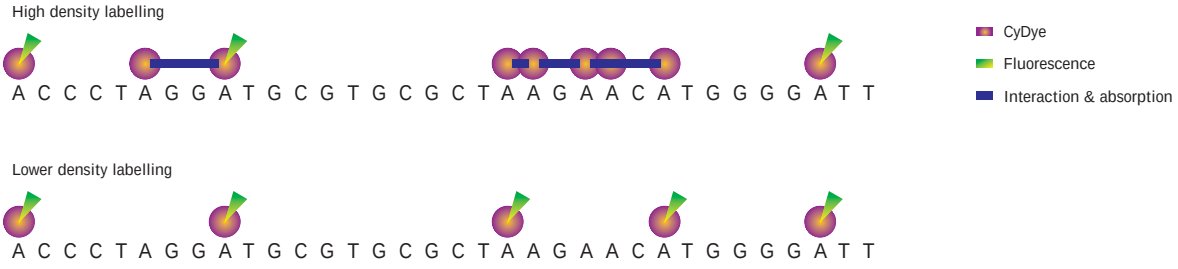


Fig 32. Quenching and labelling density. Fluorescent output from two identical nucleic acid strands labelled to different labelling densities is depicted. Green denotes fluorescent emission whereas purple shows energy being lost in intermolecular interactions between adjacent fluorophores. Lower labelling density results in higher fluorescent signal.

6.2.5 Equal labelling with different fluors

The purpose of differential gene expression analysis is to detect relative differences in the number of specific transcripts between two or more samples. This requires that the two samples hybridize competitively with the immobilized targets and differences in relative signals primarily reflect changes in the number of the transcripts. From a technical point of view, this is best achieved when equal numbers of equally labelled nucleic acid fragments are compared. As two different fluors are used in two-color analysis, no imbalance due to the properties of the fluors should be present in the labelled populations. In extreme cases such imbalances can lead to false positive signals from gene expression microarrays. The labelling method should produce the same labelling density and size distribution of labelled fragments, regardless of the fluorescent dye label used.

6.2.6 Nucleotide sequence preferences

As all labelling methods attach the fluorescent label in a specific manner, usually via certain nucleotides, the nucleotide sequence of the molecules being labelled can have a significant effect on labelling density and also on the length of labelled fragments (Fig 33). For example, incorporation of CyDye dUTP instead of CyDye dCTP into cDNA that is C-rich is less likely to result in quenching because the likelihood of incorporating two CyDye dCTP into close proximity will be lower. Conversely, highly A-rich sequences are best labelled using CyDye dCTP as label.

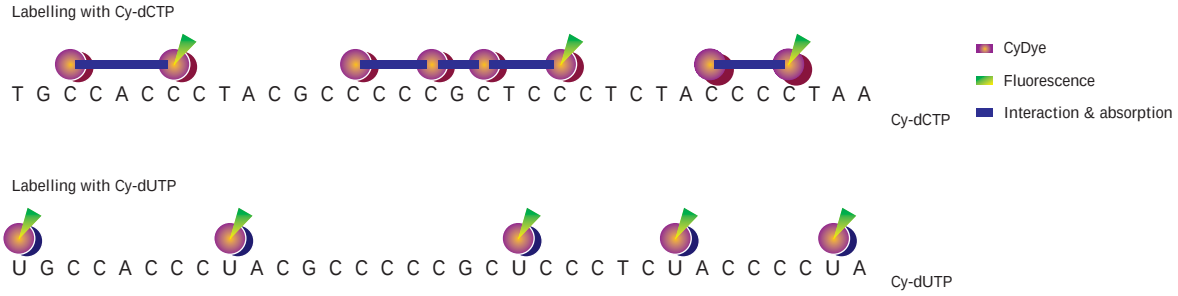


Fig 33. The effect of sequence on labelling density and fluorescent signal. The same nucleotide fragment has been labelled with Cy-dCTP and Cy-dUTP, using a similar ratio of label to cold nucleotide. As the fragment is GC-rich, the use of Cy-dCTP results in much higher labelling density on this fragment than the use of Cy-dUTP, which can only be incorporated in a few positions. However, the fluorescent signal from the Cy-dUTP labelled fragment is greater than from the more densely labelled fragment, because quenching becomes a significant factor in the situation of high labelling density.

As probes used in microarray analysis are complex mixtures of nucleic acid sequences, careful optimization of labelling methods is needed to ensure that quenching is not a practical problem for the majority of transcripts. However, it may not be possible to create labelling conditions that would be perfect for all sequences. Performing a 'yellow experiment', in which the same sample is labelled with both fluors and then used in microarray hybridization, can help to identify the targets that do not give balanced signals, as well as to optimize the microarray system in general. By choosing target sequences that do not show high nucleotide sequence bias, it is also possible to control this aspect of labelling.

Increasing labelling density is not the best solution for increasing signal from microarrays. Using a lower labelling density is more likely to result in higher signal. As fluorescent molecules tend to be fairly large, high labelling densities can lead to changes in the hybridization kinetics of labelled molecules (47). The melting temperature (Tm) of highly substituted nucleic acids are lower than those of unlabelled nucleic acids and can result in lower hybridization signal.

6.3 Labelling strategies

6.3.1 mRNA vs total RNA

Only a small proportion, about 1.5–2.5%, of cellular RNA is mRNA. Prior to hybridization, mRNA must be purified and labelled. Since most of the cellular RNA is ribosomal RNA, specific protocols are used to separate mRNA from ribosomal RNA.

Eukaryotic transcripts usually have a poly-dA tail at their 3' ends. This property can be exploited by using a complementary poly-dT sequence to capture polyadenylated transcripts away from other RNA species, as well as from other molecules and impurities. Because of the additional purification steps involved in the preparation of mRNA, mRNA samples tend to give higher signal to noise values on microarrays than total RNA samples.

6.3.2 Priming with oligo(dT)

In the labelling reaction, mRNA can be selected from total RNA for use as the labelling template by using oligo(dT) primers that will hybridize with the poly-A tail in transcripts. Addition of an anchoring base to the 3' end of these primers directs cDNA synthesis to the beginning of the poly-A stretch. This has the advantage of producing labelled fragments that are devoid of most of the repetitive sequence. This priming method will result in only one copy of cDNA that contains primarily 3' sequences synthesized from each transcript. If the targets on the microarray are derived from 5' ends of long cDNAs, probes labelled directly in cDNA synthesis using oligo(dT) priming may not produce complementary fragments to these targets, resulting in absence of signal in hybridization.

6.3.3 Random priming

Random priming, in which a mixture of oligonucleotides comprised of all sequence variants of a short sequence of defined length are used as primers, can be used to produce probes that contain sequences derived from all parts of transcripts. Typically, each transcript is copied into several non-overlapping probe fragments. Because of their longer length and ability to form more stable duplexes, nonamers are preferred over hexamers and give higher yields of cDNA. Random priming is only compatible with mRNA templates, as random primers can anneal to all RNA molecules. cDNA synthesis from total RNA with random priming will produce a large quantity of short fragments that lack specificity in hybridization and usually give rise to high background signals. As the proportion of label incorporated into cDNA derived from mRNA is going to be very small under these conditions, the specific signals from microarray spots will be low. Conversely, as most of the fluorescent label is incorporated into sequences derived from ribosomal RNA, unspecific hybridization can become a problem.

6.3.4 Other priming strategies

Highest yield of cDNA from mRNA (without probe amplification) is achieved with the use of both oligo(dT) and random primers together (Fig 34). This strategy has the highest likelihood of copying all parts of transcripts into probe, and therefore, is suitable for use with target sequences that are derived from varying parts of genes.

It is also possible to use specific primers to copy transcripts into probe. As each sequence requires the synthesis of a specific primer, this approach can be costly and require a new set of primers to be prepared for each different microarray. The advantage of this approach is that only those sequences that are analyzed on the microarray are labelled. It is also possible to use total RNA as a sample.

6.3.5 Amount of primer

Regardless of the type of primer used, its concentration should be in excess of the number of possible binding sites on transcripts so that its availability is not limiting cDNA synthesis. cDNA synthesis with general priming strategies (oligo[dT] and random nonamers) may be biased towards highly expressed transcripts if too much mRNA is used. The specific priming strategies may be biased against high-expressing transcripts if the amount of each primer is not sufficient to cover the whole expression range.

6.3.6 Labelling bacterial RNA

Bacterial mRNA lacks poly-A tails and selecting transcripts for labelling from total RNA is not as easy as with eukaryotic RNA. It is possible to remove ribosomal RNA sequences by converting them into cDNA: RNA hybrids, followed by digestion with RNAse H and DNAse I to selectively remove the double-stranded sequences. Random priming strategies can be used successfully with bacterial RNA, if the stringency of hybridization is controlled carefully to counteract the high proportion of label associated with ribosomal RNA sequences. Alternatively, priming strategies that utilize gene specific primers or short primers that are able to prime from several genes have been used (48, 49). As gene specific primers prime cDNA synthesis only from those genes that are being studied on the array, they can help to increase specificity of hybridization and signal to noise values.

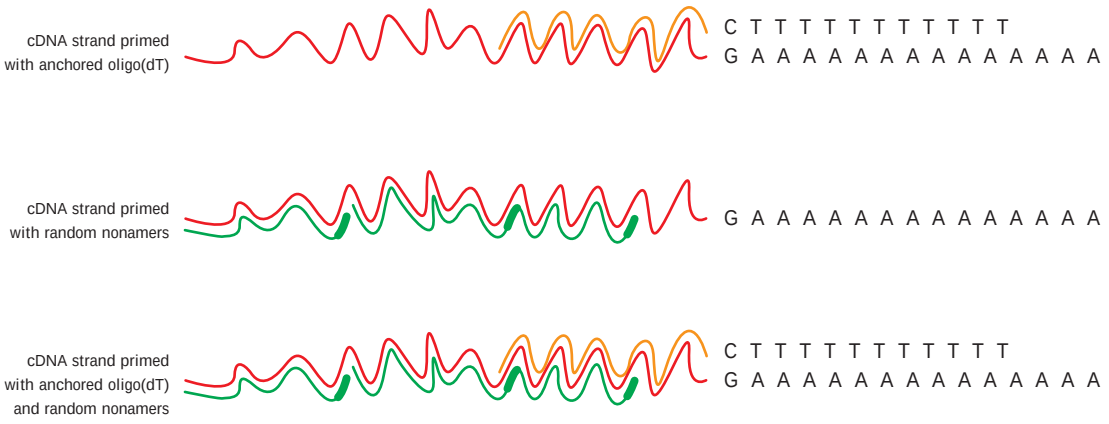


Fig 34. The use of different primers in cDNA synthesis. cDNA strand synthesized with anchored oligo(dT) priming is shown in yellow and those primed with nonamers in green. The combined use of both primers results in the probe being prepared from all parts of the transcript.

6.4 Enzymatic labelling methods

6.4.1 Labelling strategies

Several strategies based on molecular biology or chemical reactions have been developed for labelling samples for gene expression microarray analysis. The availability of fluorescent labels in different reactive forms has contributed to the diversity of labelling methods. All these strategies have in common that they start with an RNA population (Fig 35). Molecular biology strategies rely on the use of enzymes to convert mRNA into new populations of nucleic acids, either DNA or RNA. Combining two or more enzymatic reactions into one protocol widens the choice further. Using more than one enzyme for labelling, however, has the disadvantage that the information carried by the original population is likely to change more than by using a single enzyme. This is because some information is lost in each enzymatic conversion step, and as the lost information is dependent on the sequence of the transcripts and the properties of the enzyme, the representations synthesized by each enzyme will be different. Chemical methods have the advantage that no copying of nucleic acid to another form takes place; instead the labelling moiety reacts with the nucleic acids to form covalently modified, labelled probe population.

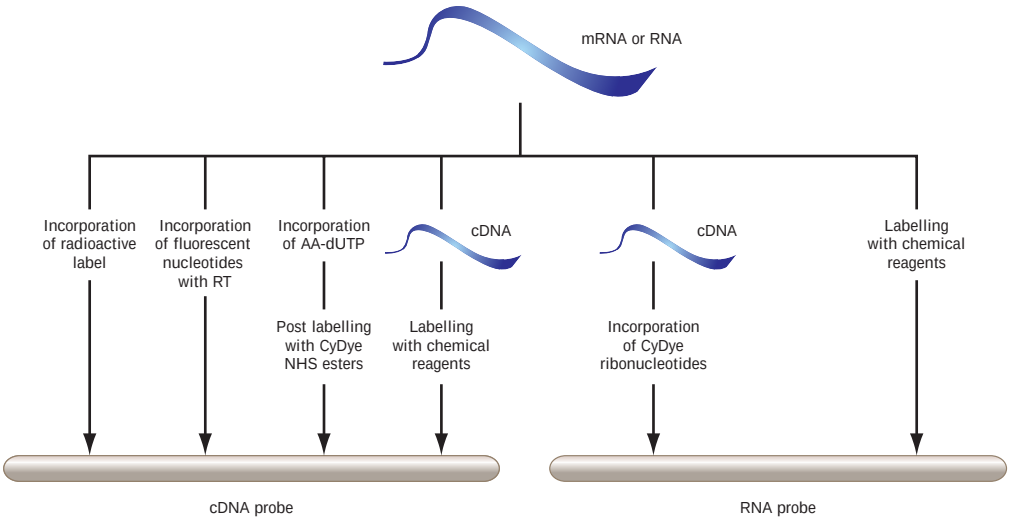


Fig 35. Labelling strategies for gene expression microarrays.

6.4.2 Fluorescent labels

Fluorescent dyes, especially the cyanine dyes Cy3 and Cy5, are the most popular choice for dual color microarray analysis. The main benefit of using CyDye fluors in particular is that two dyes can be excited and detected from the same slide. CyDye fluors also produce bright signals and have a wide dynamic range of detection, so both weak and strong signals can be detected in the same experiment. Fluorescent dyes can be directly incorporated into nucleic acid by either enzymatic or chemical methods.

6.5 Labelling in first-strand synthesis

6.5.1 Principle

One of the simplest and most popular labelling strategies is to convert mRNA population into a labelled first-strand cDNA population. This is achieved by copying the transcripts into cDNA molecules with a reverse transcriptase while incorporating a modified CyDye nucleotide. The cDNA synthesis can be primed with a choice of primers including random primers, anchored oligo(dT) as well as gene specific primers. This allows the use of both mRNA and total RNA as sample.

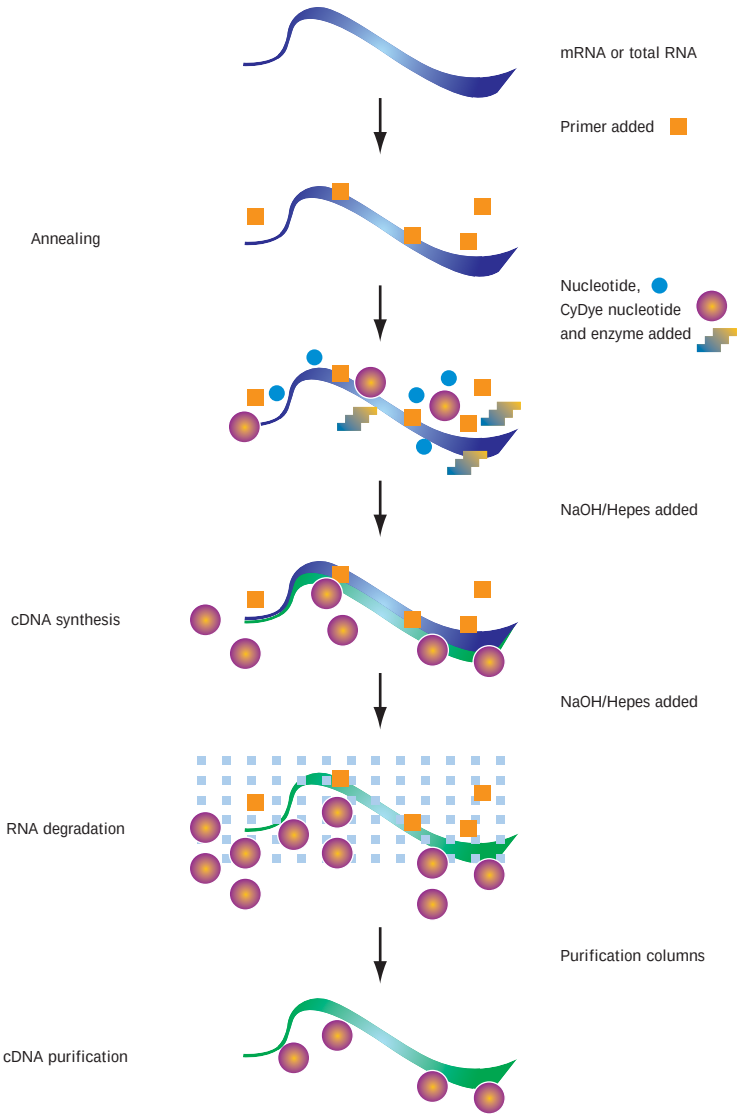


Fig 36. Principle of labelling in first-strand cDNA synthesis. Fluorescent nucleotide (pink circles) is incorporated into first-strand cDNA by a reverse transcriptase. After degradation of the mRNA template strand, labelled single-stranded cDNA probe can be purified.

6.5.2 Optimization

The incorporation of fluorescently labelled nucleotides is the rate-limiting step of this labelling method. This is because all polymerase enzymes, both DNA and RNA dependent polymerases, incorporate unlabelled nucleotides more efficiently than larger fluorescent dye nucleotides. The incorporation kinetics are very much dependent on the identity of the enzyme, and even homologous enzymes can have very different properties. The ratio between the fluorescently labelled nucleotide and the corresponding unlabelled nucleotide in the labelling reaction determines the incorporation of fluors into cDNA. As different fluorescent dyes and nucleotides have different structures, it is necessary to optimize this ratio separately for each combination of dye and nucleotide. Furthermore, optimized ratios determined for one enzyme will not necessarily give optimal results with other enzymes.

Labelling in first-strand synthesis does not produce long cDNA molecules. This is because elongation of the nucleotide chain is dependent on the previous nucleotides having been incorporated. The incorporation of more than one fluorescent label consecutively into cDNA is not favored by polymerases. If the ratio of fluorescent nucleotide to unlabelled nucleotide is high, a highly labelled cDNA can be transcribed, but cDNA synthesis will stall if the mRNA sequence requires several labelled nucleotides to be incorporated in a row. Only short fragments will be made, resulting in low yield of cDNA. Lowering the nucleotide ratio can increase the yield of cDNA, but this will compromise labelling density and the brightness of the probe. A balance between these two factors and the consequences of quenching at high labelling densities must be attained for optimal results. In practice this requires evaluation of fluorescent signals on microarrays from probes labelled at different nucleotide ratios, and this demands considerable effort.

6.6 cDNA Post Labelling

6.6.1 Principle

The shortcomings of the first-strand cDNA labelling method and the availability of CyDye as reactive N-hydroxyl succinimidal dyes (NHS-dyes) have led to the development of cDNA post-labelling method.

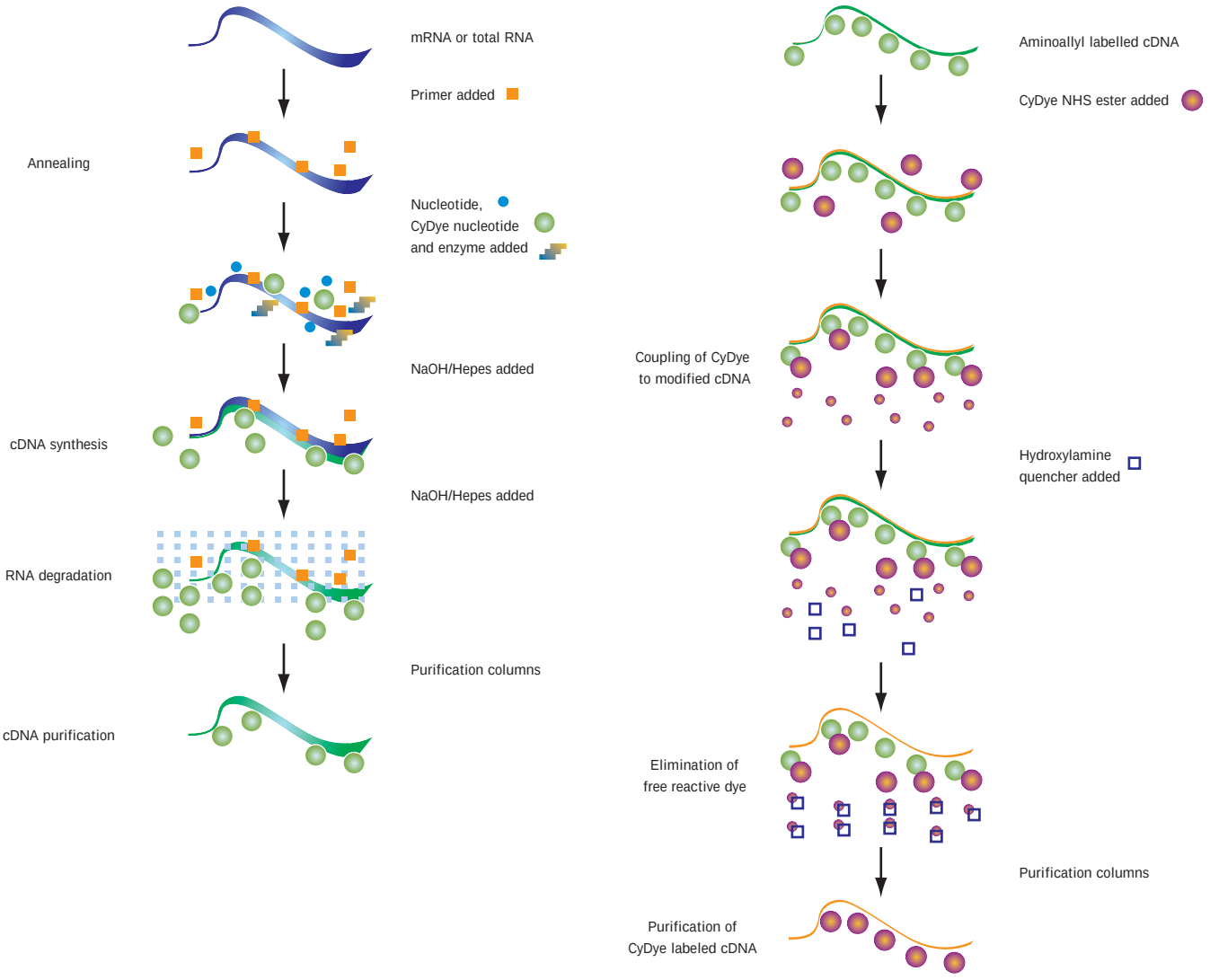


Fig 37. The principle of post-labelling is illustrated. mRNA is converted into first-strand cDNA that contains aminoallyl-dUTP. After elimination of mRNA template, the amine groups on cDNA are reacted with CyDye-NHS ester, resulting in the generation of fluorescently labelled cDNA. Excess of NHS-ester can be neutralized with hydroxylamine, and labelled cDNA is purified for use.

Amersham Biosciences has developed CyScribe™ Post-Labelling Kit (see section 6.14) for this method. In this method, amine-modified cDNA is first synthesized by incorporating aminoallyl-modified nucleotide in first-strand cDNA by a reverse transcriptase. After removal of RNA template and purification of the amine-modified cDNA, chemical labelling with N-hydroxyl succinimidyl-ester derivative of CyDye is performed. NHS-esters react with amine groups to form a covalent bond between CyDye and the amine group while the NHS-ester group is released (Fig 38). A high excess of CyDye NHS-ester is required for efficient reaction, and any non-reacted label is neutralized with an excess of small amine such as hydroxylamine. The cDNA needs to be re-purified after labelling to remove CyDye that is not incorporated into labelled cDNA.

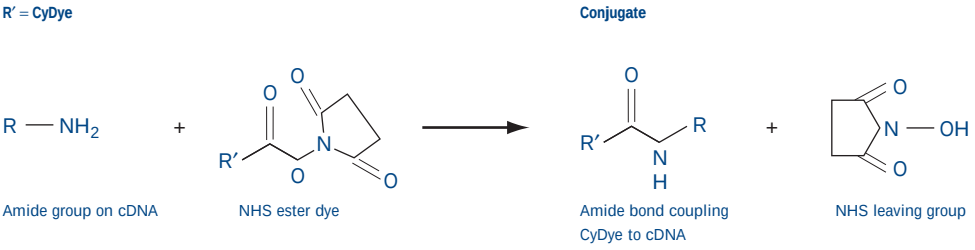


Fig 38. The use of CyDye NHS-ester in labelling amine-modified cDNA

6.6.2 Benefits

The main benefits of this method over the first-strand cDNA labelling method are derived from the more efficient incorporation of the smaller aminoallyl nucleotide compared with the bulkier CyDye nucleotides. As a consequence, the yield of cDNA is considerably higher than with the first-strand labelling method. The cDNA fragments synthesized in the post-labelling method are also longer. Further benefits of the post-labelling method stem from the chemical labelling step itself. The number of amine groups on cDNA is the main factor influencing labelling density. The sequences of the cDNAs being labelled do not have a major impact on labelling outcome. Because of this, more random attachment of labels is achieved than with the first-strand cDNA labelling method. Furthermore, as the labelling process is not dependent on the structure of different fluorescent dyes, it is easier to achieve equal labelling with both Cy3 and Cy5, and the labelling method introduces less variation into microarray analysis. This is illustrated in tighter scatter plots derived from microarray hybridizations in which the mRNA samples were labelled with the post-labelling method (Fig 39). As the extent of experimental variation is reduced, the detection of smaller changes in gene expression between two samples is improved.

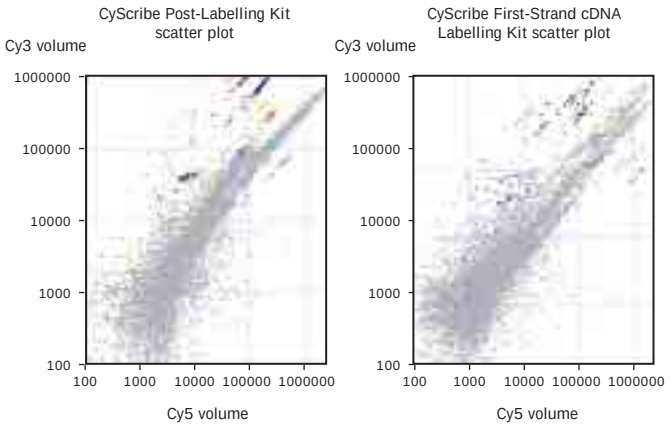


Fig 39. Scatterplots of gene expression microarray analyses comparing skeletal muscle and placental mRNA samples. Identical mRNA samples were labelled with the post-labelling method or with first-strand cDNA labelling. Identical slides were hybridized with equal amounts of all probes.

6.6.3 Chemical considerations

Coupling of CyDye NHS-ester to amine-modified cDNA requires mildly alkaline pH, which is provided by the coupling buffer. However, if they are not carefully removed, buffer components or acidic residues from preceding steps can alter the pH of this buffer. Furthermore, any amine groups present on other compounds will compete with the amine-modified cDNA for CyDye incorporation. Hence the free aminoallyl-dUTP, Tris-buffer, and reverse transcriptase enzyme must be removed from the cDNA preparation before labelling. This can be best achieved with affinity column chromatography methods, such as GFX columns, in which amine-modified cDNA is bound to the matrix, other compounds are washed away, and cDNA is finally eluted with water. Alternatively, standard ethanol precipitation also works well and provides the added benefit of concentrating the cDNA ready for CyDye coupling.

NHS-esters are readily hydrolyzed with water, even with the small amount of moisture present in laboratory air. Because of this, aliquots of CyDye NHS-esters must always be stored desiccated and protected from light. Storing CyDye NHS-esters in solution can lead to rapid loss of reactivity. As these reactive dyes were originally developed for protein labelling, the quantities provided commercially were adjusted for this application. This necessitated aliquoting of reactive dyes before their use for microarray sample labelling and frequently resulted in decreased activity as a consequence of handling and storage. The availability of individually foil-packed, pre-dispensed, and freeze-dried aliquots of Cy3 and Cy5 NHS-esters for microarray analysis has removed this problem. Furthermore, the quality of the CyDye NHS-ester in CyDye Post-Labelling Reactive Dye Packs is higher than available otherwise.

6.7 RNA amplification and labelling in RNA synthesis

6.7.1 RNA amplification

The amount of RNA sample can be a limiting factor for microarray analysis, and it may be necessary to amplify RNA before analysis. In the most commonly used protocol, the mRNA population is first converted into a double-stranded cDNA that contains a promoter sequence for viral RNA polymerase, such as T7, T3, or SP6 polymerase (Fig 40).
This can be achieved by using a modified oligo(dT) primer containing a 5' extension coding for the viral promoter. Each resulting cDNA molecule will contain one RNA polymerase promoter sequence. By including the corresponding RNA polymerase and ribonucleotides in the reaction, several RNA copies can be synthesized from each template.

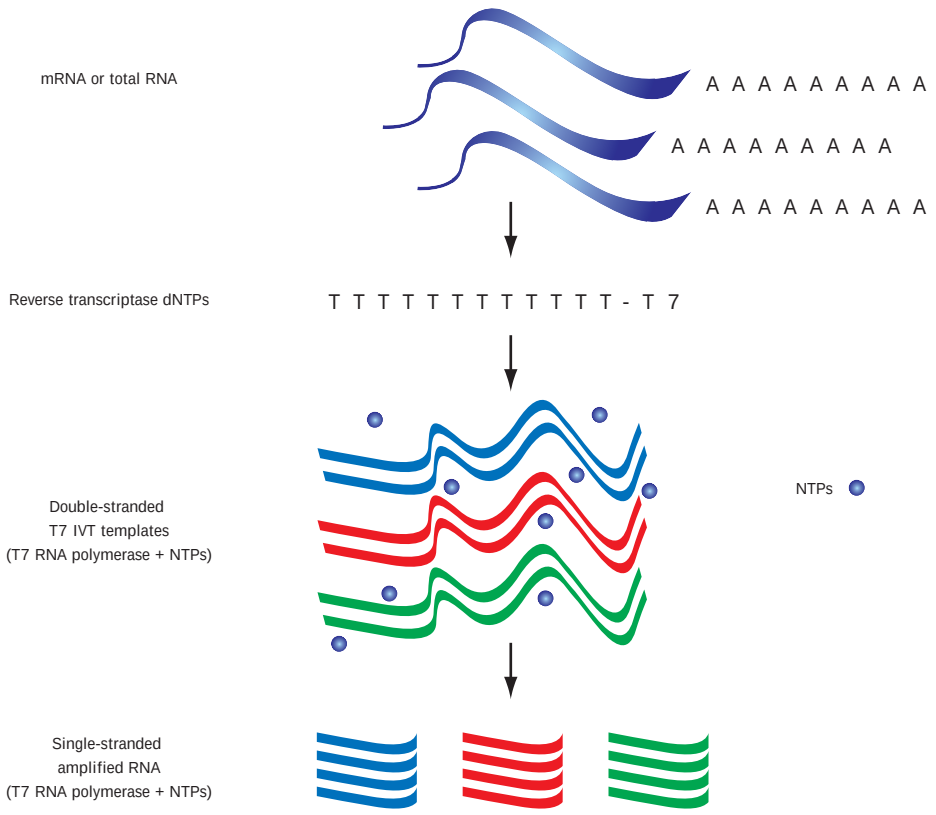


Fig 40. Principle of RNA amplification with RNA polymerases.

This results in a linear amplification of the cDNA pool into RNA without altering the relative abundance of sequences in the mixture significantly. However, the length of synthesized fragments will be shorter than the original templates. 2000-fold amplification of starting RNA can be achieved with this method, in one round of amplification (50, 51). This RNA amplification strategy can be used on its own, and the amplified RNA population can be labelled separately using other methods. Alternatively, labelling and amplification can be performed together by including a CyDye ribonucleotide in the reaction (Fig 41). In some cases when only a few cells are available for the preparation of RNA sample, such as when laser capture micro (LCM) dissected samples are analyzed, several rounds of RNA amplification can be performed to acquire enough RNA for labelling (52).

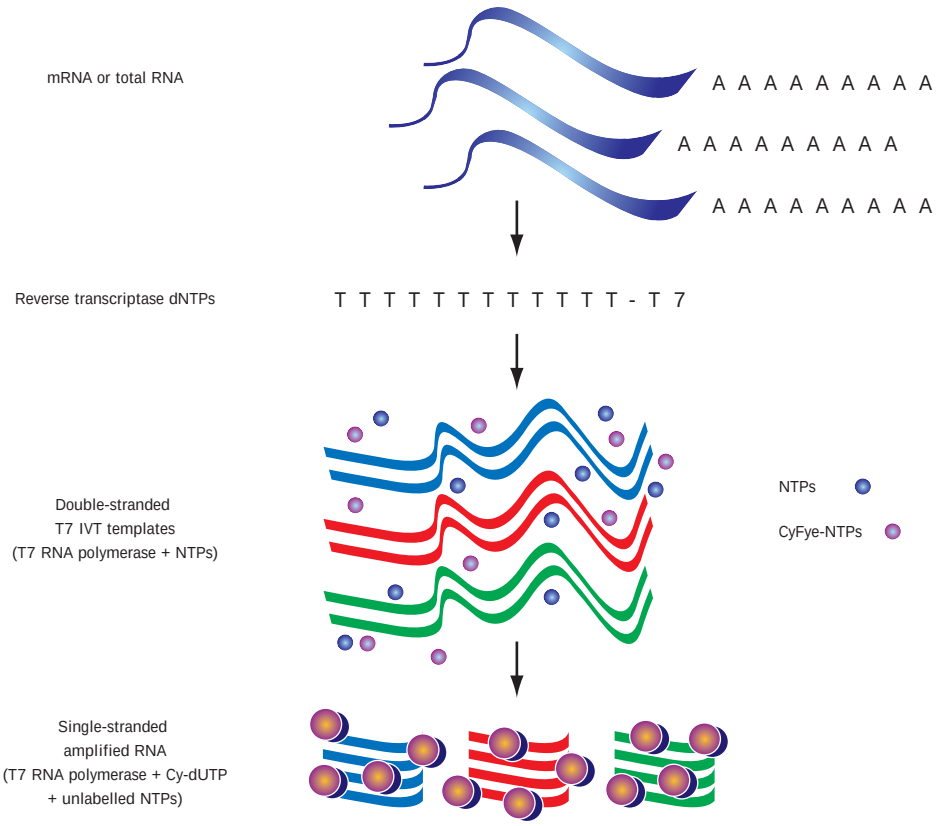


Fig 41. Principle of preparation of fluorescently labelled RNA probe with RNA polymerase amplification.

6.7.2 Other amplification methods

Amplification of RNA can also be achieved by other means, including using limited numbers of PCR cycles to amplify double-stranded cDNA (53). However, PCR-based methods have not gained wide popularity in microarray analysis, because of the problems associated in logarithmic amplification of complex nucleic acid mixtures. Sequence and size differences in nucleic acid fragments can influence their amplification rate and result in selection of sub-populations of sequences (50). Regardless of the RNA amplification strategy, the method should neither alter the relative abundance of different transcripts in the sample nor result in the creation of sequence chimeras in which sequences from two or more transcripts are joined together. Linear amplification strategies avoid these pitfalls and are favored in microarray analysis.

6.7.3 Labelling with RNA amplification

Synthetic RNA can be labelled in synthesis reactions by incorporating a CyDye ribonucleotide. The main factor determining both labelling density and yield of labelled RNA in this method is the ratio of CyDye ribonucleotide to the corresponding unlabelled ribonucleotide. As is the case for DNA polymerases, RNA polymerases incorporate the unlabelled nucleotide more efficiently, and a compromise between labelling density and yield of RNA is necessary. This requires careful optimization of the labelling conditions, combined with the analysis of fluorescent signal derived from the probes.

RNA labelling can also be performed by using hapten-labelled ribonucleotide, such as biotinylated ribonucleotides. This strategy has the advantage of giving higher yields of cDNA, as the incorporation of large dye nucleotide is not a rate-limiting factor. However, the need to perform additional detection steps adds to the length of the protocol and makes the use of more than one color at a time more difficult.

6.7.4 Fragmentation of RNA probes

The secondary structure of RNA can interfere with hybridization to targets. Fragmentation of RNA probe into smaller fragments of 50–200 nucleotides can be performed to overcome this. This can be achieved in controlled fashion by exposure of RNA to potassium and magnesium ions (24). Careful handling of RNA is necessary at all stages to minimize uncontrolled degradation of RNA to nucleotides and short fragments. If RNA probes are used in hybridization, precautions must be taken throughout the whole microarray procedure.

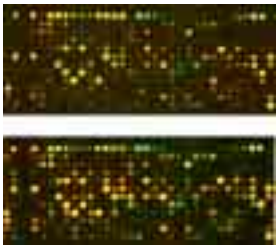


Fig 42. Expression patterns obtained from gene expression microarrays comparing skeletal muscle and placental mRNA samples. Identical mRNA samples were labelled with the post-labelling method and first-strand cDNA labelling methods using CyScribe Labelling Kits. Equal amounts of all probes were hybridized with identical microarrays. Although the overall hybridization patterns are very similar, some significant differences in signal intensities of individual spots can be seen. These reflect the size differences of labelled fragments obtained with the two labelling methods. Despite the different appearance of the two microarrays, both gave similar information on differential gene expression.

6.8 Expression patterns

The intensity of hybridization signals from microarrays is determined not only by the number of hybridized probe fragments but also by their length and the number of fluorescent labels that are associated with them, or the labelling density. Similar labelling densities can be achieved with different labelling methods. However, with different methods, the average length of the labelled fragments varies. In practice, the intensity of signals derived from two identical samples labelled with different methods will be different. Hence, gene expression microarrays do not give quantitative information about gene expression. However, if two samples are labelled by the same method using two different fluorescent labels, the relative intensity of the signals will reflect the relative abundance of the specific transcripts in those two samples. Information about differential gene expression can therefore be gained (Fig 42). Labelling the two samples to be compared under identical conditions is therefore extremely important for guaranteeing that experimental artifacts do no lead to wrong conclusions from microarray results.

6.9 Random prime labelling of DNA

A modified random prime labelling method can be used to label DNA with CyDye. In this method, Klenow polymerase incorporates fluorescently labelled nucleotides in a DNA synthesis reaction, which is primed with random nonamer or hexamer primers. This method is a practical solution for genomic microarrays, although direct chemical labelling methods can also be used. Random prime labelling methods are not recommended because two different enzymes are needed to convert mRNA into labelled form.

6.10 Direct chemical labelling of mRNA

When an enzyme is used to convert a mRNA population into another nucleic acid population, some information is lost because the ability of different enzymes to copy through different nucleic acid sequences varies. By using a chemical labelling method, it is possible to label mRNA directly by using a chemical reaction, coupling modified CyDye reagent to RNA molecules. These methods are simple to perform, as no modification of RNA is required before labelling. Furthermore, these methods are less prone to discrimination against certain nucleotide sequences that are difficult templates for polymerase-based labelling systems. It is important to note that any chemical that avidly reacts with nucleic acid molecules is potentially toxic and will require careful handling and adequate safety measures to be taken.

6.11 Purification of labelled probes

Regardless of the labelling strategy, it is necessary to purify the labelled nucleic acid after labelling, as the amount of incorporated fluorescent dye is typically only a small fraction of all the dye present in the sample. The recovery of labelled nucleic acid from purification is a major limiting factor for most labelling methods. Products such as AutoSeq™ G-50, GFX, and QIAquick™ spin columns have been developed for the purification of double-stranded cDNA. They can also be used to remove unincorporated CyDye nucleotides away from fluorescently labelled single-stranded cDNA, but their use does not result in optimal recovery of labelled material. Typically, less than 40% of the labelled probe is recovered, and the recovery can vary considerably between different samples. This variation not only reduces the amount of data that can be generated with microarray analysis, but also significantly contributes to poor quality of results if the amount of probes are not adjusted before hybridization. When small amounts of template are labelled, the loss of labelled cDNA tends to be higher, and as little as 5% of probe may be recovered.

All of these methods, however, are relatively successful in removing the free dye nucleotide. Ethanol precipitation is not an option for use with most labelling methods, as it can result in the formation of dye aggregates that will produce intense speckled background in array hybridization.

Appreciation of these problems has lead to the development of a novel GFX purification system, CyScribe GFX Purification Kit, which has been specifically tailored for purification of single-stranded CyDye labelled cDNA. As this purification system is based on binding of the labelled nucleic acid to a customized GFX matrix, it is also suitable for use with the post-labelling method. These CyScribe GFX columns give excellent yield of purified cDNA from different synthesis scales, including small-scale samples. Therefore, the use of this purification system downstream of optimized labelling kits can improve the sensitivity of microarray analysis, enable the use of smaller amounts of RNA samples, or increase the number of replica slides that can be hybridized with one sample.

6.12 The CyScribe family of labelling kits

The CyScribe family of labelling kits from Amersham Biosciences has been developed to offer a range of optimized labelling products for producing CyDye labelled microarray probes. These kits enable flexible choice of different labelling methods to suit the different needs of researchers.

6.13 CyScribe First-Strand cDNA Labelling Kit

6.13.1 Features of the kit

CyScribe First-Strand cDNA Labelling Kit has been developed for preparing highly fluorescent CyDye labelled probes for microarray analysis using first-strand cDNA labelling. The kit has been optimized for the use of either CyDye dCTP or CyDye dUTP nucleotides as labels. Two nucleotide mixes are provided in the kit to provide optimal nucleotide ratios for each type of nucleotide when 1 nmol of CyDye nucleotide is used per reaction. The compositions of these solutions have been optimized so that labelled cDNA will contain an attached CyDye fluor to every 12–25 nucleotides synthesized, regardless of the nucleotide used.

CyScribe First-Strand cDNA Labelling Kit contains both anchored oligo(dT) primers and random nonamer primers, allowing flexible choice of templates. As little as 50 ng of mRNA and 2.5 µg of total RNA can be labelled per reaction with the kit and used successfully in microarray hybridization on one slide (Fig 43). Recommended highest amounts of

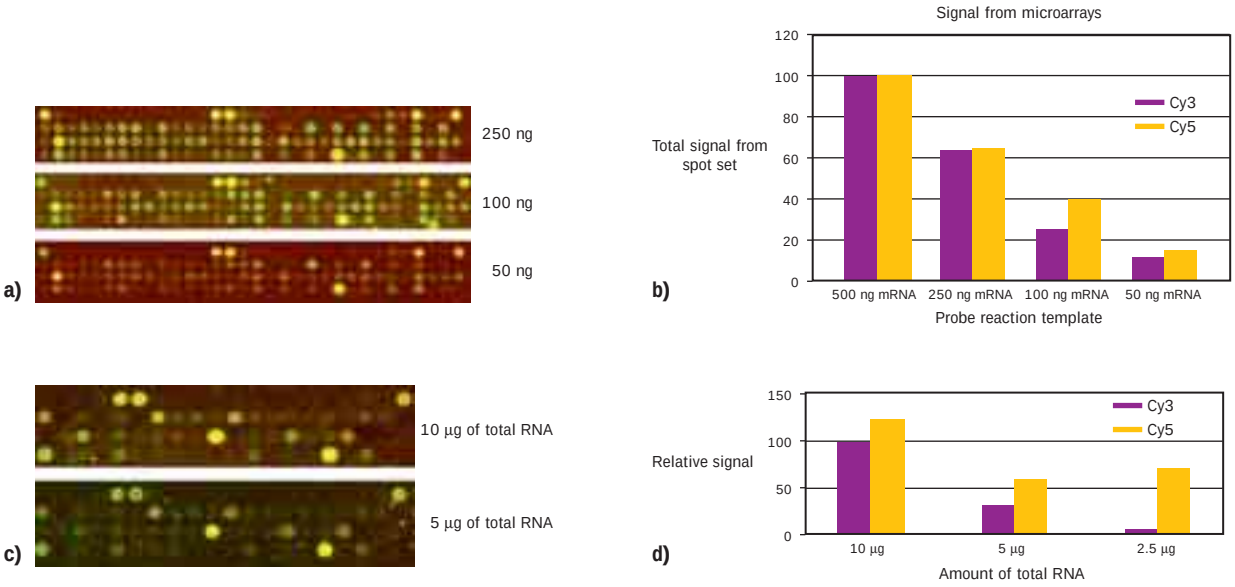


Fig 43. The use of CyScribe First-Strand cDNA Labelling Kit with small template amounts. The indicated amounts of mRNA and total RNA templates were labelled with Cy3 and Cy5 in duplicate reactions, purified, and all of the recovered probe was used in a microarray hybridization with duplicate slides. Panels A and C show part of microarray images obtained with mRNA probes and total RNA probes, respectively. Panels B and D show quantified Cy3 and Cy5 signals from these arrays. As the amounts of probes were not adjusted before hybridization, a high amount of variation in signal intensities was observed in this experiment.

template are 500 ng of mRNA and 25 µg of total RNA per reaction. These amounts of template will generate enough probe for several microarray hybridizations. Highest yield of labelled probe is obtained with dual priming, when both oligo(dT) and random primers are used together. This priming strategy is only compatible with the use of mRNA templates. The choice or amount of RNA template does not affect the labelling density achieved with the kit.

6.13.2 Degradation of RNA template

Degradation of the RNA template after cDNA synthesis is necessary to prevent the labelled probe from hybridizing with the original template in solution instead of the microarray targets during microarray hybridization. The removal of RNA can be performed enzymatically, by using RNAse H enzyme to digest the RNA component of RNA DNA heterohybrid molecules. A simpler option is to degrade the RNA strands by raising the pH of the probe solution. The CyScribe Kits contain an efficient RNA degradation protocol which has been developed to minimize pH fluctuation observed in earlier protocols in which small volumes of concentrated alkali and acids were used. In this improved protocol, RNA is degraded by the addition of 2 µl of 2.5 M sodium hydroxide, and after incubation the solution is neutralized with 10 µl of 2 M Hepes free acid.

The CyScribe First-Strand cDNA Labelling Kit also contains a mixture of synthetic mRNA molecules, as control RNA template, that can be used to gain familiarity of the labelling technique, or to troubleshoot problems. Because of its synthetic nature this control RNA is not suitable for microarray hybridization.

6.13.3 Critical success factors for CyScribe First-Strand cDNA Labelling Kit

- Only label RNA that is intact, clean, and in known quantity.
- If possible, purify mRNA for best labelling results and highest yield of labelled probe.
- Do not exceed the recommendations for template amount.
- Do not alter the amount of CyDye in reaction.
- Pipette all volumes exactly.
- Protect CyDye from light during all handling and storage.
- Do not alter the RNA degradation protocol.
- Monitor the success of purification.
- Measure the amount of purified probe before performing microarray hybridization.

6.14 CyScribe Post-Labelling Kit

6.14.1 Features of the kit

CyScribe Post-Labelling Kit has been developed to offer an optimal and convenient solution for using post-labelling methods in microarray analysis. Each kit provides reagents for performing 12 cDNA synthesis and labelling reactions with both Cy3 and Cy5 fluors. CyScript reverse transcriptase, which is a highly efficient enzyme and gives high yields of cDNA, is used to synthesize amine-modified cDNA by incorporation of aminoallyl-dUTP into first-strand cDNA. The amount of this modified nucleotide in the cDNA synthesis reaction has been adjusted to give an optimal labelling density with CyDye fluors.

The kit includes a protocol for an improved RNA degradation method and two alternative methods for removing amine-containing impurities from amine-modified cDNA. Purified cDNA is reacted with an amount of reactive CyDye NHS-ester that has been chosen to give high and reproducible labelling density, similar to that achieved with the CyScribe First-Strand cDNA Labelling Kit. Protocols for reacting excess NHS-esters with hydroxylamine and subsequent purification of labelled cDNA with column chromatography are also provided.

A practical difficulty in performing CyDye post-labelling of microarray samples has been the necessity for aliquoting and storing CyDye NHS-esters when traditional reagents developed originally for protein labelling have been used. Because of the instability of NHS-esters in moist conditions, their storage in standard laboratory conditions can result in significant loss of reactivity in just a few weeks. CyScribe Post-Labelling Kit solves this problem by providing ready-to-use CyDye NHS-esters in individually dispensed aliquots. These dyes have been sealed in foil to protect them from light and contain desiccant for extra protection. The reactive dyes in these aliquots are also guaranteed to contain over 75% reactive dye content, thus providing the highest quality reagents for microarray labelling.

CyScribe Post-Labelling Kit includes both oligo(dT) and random nonamer primers, offering flexible use of both total and messenger RNA as template. Because this kit yields high amounts of cDNA, it is recommended that 500 ng or less of mRNA and 25 µg or less of total RNA are used per reaction. Highest yield of cDNA is obtained from mRNA using both types of primers together.

6.14.2 Comparison of performance with CyScribe First-Strand cDNA Labelling Kit

Properties of labelled probe:

- The CyScribe Post-Labelling Kit synthesizes about three times as much cDNA than CyScribe First-Strand cDNA Labelling Kit from an equal amount of template. However, because this kit involves two purification steps in which some of the cDNA is lost, approximately two-fold increase in final probe yield is achieved.
- Both kits have been optimized to give similar labelling densities: on average a CyDye fluor is attached at every 12–25 nucleotides.
- As the average length of cDNAs synthesized with the post-labelling kit is longer, higher signals can be obtained from some targets (ones with long transcripts) with this kit.
- The post-labelling method is not influenced by individual nucleotide sequences to the same extent as the first-strand cDNA labelling method is. Therefore, this method can cope better with sequences that contain repeated nucleotide stretches, which can lead to chain termination in first-strand labelling.
- Because of the chemical nature of the labelling process, more random distribution of CyDye fluors over labelled cDNAs is obtained than with the first-strand labelling method in which labelling is modified by sequence-specific events.

6.14.3 Identification of differential gene expression with CyScribe kits

The performance of the two CyScribe labelling kits in identifying differential gene expression was investigated in a model experiment in which human skeletal muscle and placental mRNA populations were compared. Replicate labelling reactions were performed with both systems, purified probes were pooled, and replica slides were hybridized with 25 pmol of each probe. Examples of the expression patterns produced by the two CyScribe Labelling Kits from these

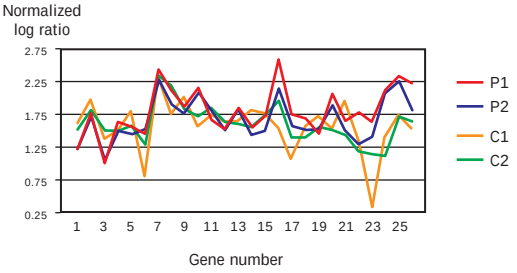


Fig 44. Identification of muscle-specific gene expression with CyScribe Labelling Kits. Data is shown from two replica slides hybridized with skeletal muscle and placental cDNA probes labelled with CyScribe First-Strand cDNA Labelling Kit and CyScribe-Post Labelling Kit.

experiments, as shown in Figure 44, highlight the different intensity of several gene-specific signals. However, when muscle-specific gene expression was examined, essentially the same genes were identified with both methods as being more highly expressed in muscle tissue.

In order to compare the performance of the CyScribe Labelling Kits in analyzing gene expression over a wide range of expression values, 31 random gene sequences were selected for analysis. In skeletal muscle mRNA, the expression levels of these genes cover the whole dynamic range of values (data not shown).

Figure 45 depicts normalized gene expression log ratios for the 31 selected gene sequences. Despite the differences in the intensity of the observed Cy3 and Cy5 signals generated by the two labelling systems (data not shown), the extent of differential gene expression revealed by the two methods is concordant for most of the genes. On the whole, variation between replica slides is greater than variation between the two labelling methods. Some genes show slightly different (within 0.5 log units) values for differential gene expression and may indicate the presence of gene sequence-specific labelling events. However, this data does not support the conclusion that either of the methods is labelling with a systematic bias towards one of the fluors. Rather, a particular labelling method may be more suitable for extracting information from certain gene sequences, and combining data from experiments using different labelling principles could enhance chances of identifying significant difference in gene expression between two samples.

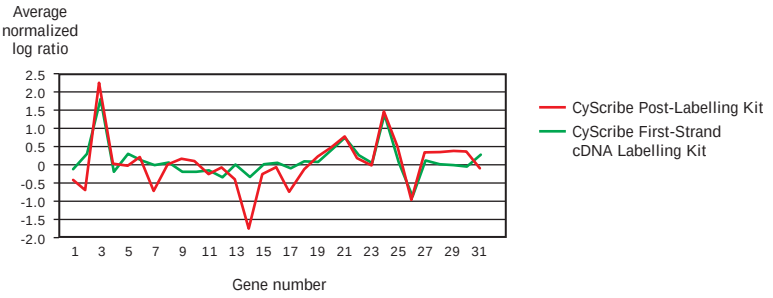


Fig 45. Differential gene expression determined with CyScribe First-Strand cDNA Labelling Kit and CyScribe Post-Labelling Kit for 31 randomly chosen genes in skeletal muscle and placental mRNA samples. Data from two replica slides, each of which contained two identical spot sets for each of the labelling methods, is shown.

6.14.4 Critical success factors for CyScribe Post-Labelling Kit

- Only label RNA that is intact, clean, and in known quantity.
- If possible, purify mRNA for best labelling results and highest yield of labelled probe.
- Do not exceed the recommendations for template amount.
- Do not use random nonamers with total RNA.
- Purification of amine-modified cDNA is critical for labelling success: other amines and acidic ions should be removed.
- Do not alter the RNA degradation protocol.
- Only dissolve CyDye-NHS esters immediately before use.
- Do not reuse CyDye-NHS ester solutions.
- Pipette all volumes exactly.
- Protect CyDye from light during all handling and storage.
- Monitor the success of purification.

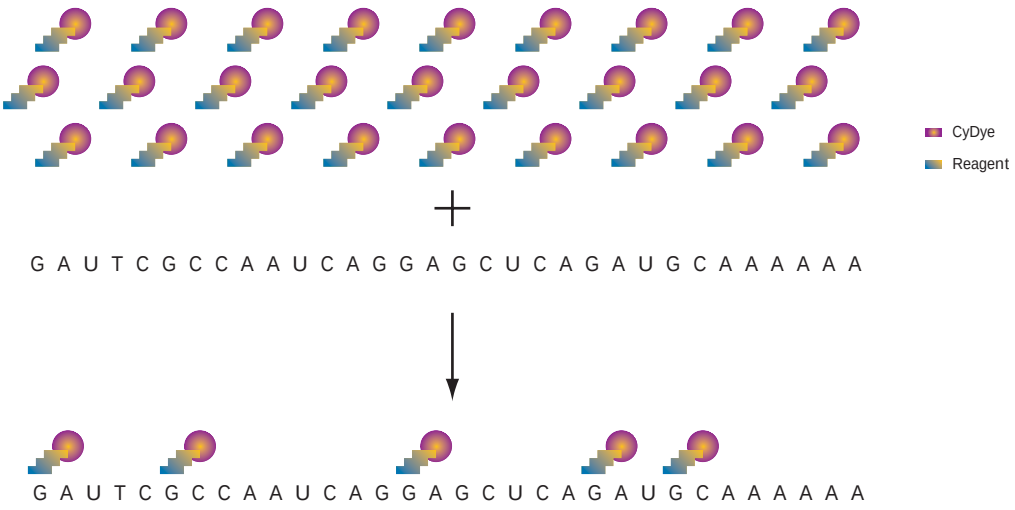


Fig 46. Principle of mRNA labelling with CyScribe Direct mRNA Labelling Kit.

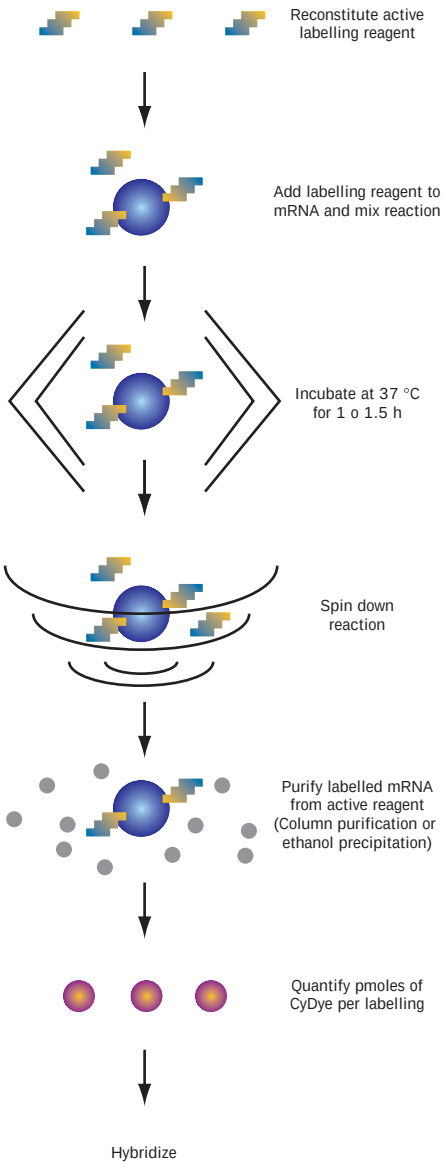


Fig 47. The method for labelling mRNA with CyScribe Direct mRNA Labelling Kit.

6.15 CyScribe Direct mRNA Labelling Kit

6.15.1 Principle

CyScribe Direct™ mRNA Labelling Kit utilizes Cy3 Direct and Cy5 Direct labelling reagents to generate highly labelled mRNA probes for use in microarray applications. These labelling reagents contain a CyDye fluor attached to a chemical group which can react efficiently with guanine residues, resulting in covalent attachment of CyDye to mRNA (Fig 46). Because of this high reactivity with nucleic acids, Cy Direct™ reagent is hazardous and requires careful handling. When in contact with water, the reagent hydrolyzes and loses its reactivity. As seen in Figure 47, the whole labelling process, including removal of excess dyes, can be performed in less than 2 h.

Because of the chemical nature of the labelling, attachment of both Cy3 and Cy5 is equally efficient and even. Furthermore, the reaction does not require any enzymatic modification of the mRNA prior to labelling. The attachment of CyDye to mRNA does not interfere with subsequent hybridization with DNA targets; thus mRNA labelled with CyScribe Direct mRNA Labelling Kit is suitable for use as microarray probe. Because mRNA can be used directly as the probe, no loss of information because of incomplete or biased transcription process occurs.

6.15.2 Application

Labelling reactions performed with CyScribe Direct mRNA Labelling Kit can be scaled up or down to accommodate different amounts of template: as little as 250 ng or as much as 5 µg of mRNA can be labelled in a single reaction. The ratio of the CyDye Direct Labelling Reagent to mRNA determines the labelling density achieved with the kit. Extending the labelling time beyond 1 h can result in higher labelling efficiency than shorter times. The length of transcripts is an important factor in determining the intensity of hybridization signals from mRNA probes labelled with the CyScribe Direct mRNA Labelling Kit. As seen in Figure 48, the expression patterns obtained with the direct labelling method differ considerably from those generated with the first-strand labelling method. This evidence suggests that the direct labelling method may be advantageous for analyzing gene expression from transcripts that are difficult templates for reverse transcription.

The CyScribe Direct mRNA Labelling Kit has been developed for use with purified mRNA that is free from contaminating DNA, proteins, or nucleotides. The kit requires the use of purified mRNA as template because the chemical labelling reaction cannot discriminate between transcripts and other species of RNA. Labelling of total RNA will result in low signal to noise values from microarray experiment.

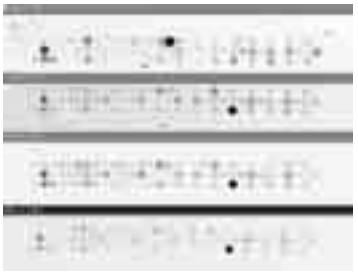


Fig 48. Hybridization of microarrays with varying amounts of mRNA labelled with Cy3 using CyScribe Direct mRNA Labelling Kit. Results from a DNA probe generated with CyScribe First-Strand cDNA Labelling Kit are shown for comparison.

6.16 Critical success factors for sample labelling

- Observe safety precautions while performing the labelling reaction.
- It is imperative that the mRNA is free from contaminating ribosomal RNA, DNA, and proteins.
- Handle RNA so that degradation is avoided. See Chapter 5 for advice. Protect the CyDye Direct labelling reagent from water or any moisture, as it can be inactivated on contact.
- Store the reagent vial and reaction tubes sealed from the environment.
- Protect all reactions and labelling reagents from light when storing and handling.
- Minimize RNAse contamination of mRNA probes during microarray hybridization.

Table 4. Choosing the right CyScribe Kit.

Feature	CyScribe Direct mRNA Labelling Kit	CyScribe First Strand cDNA Labelling Kit	CyScribe Post Labelling Kit
Brightness of signals	●	●	●
Even incorporation of Cy3 and Cy5	●	○	●
Starting material	mRNA	mRNA or total RNA	mRNA or total RNA
Quantity of starting material using total RNA	—	2.5 – 25 µg	2.5 – 25 µg
Quantity of starting material using mRNA	250 ng – 1 µg	50 ng – 1 µg	100 – 500 ng
Possible to prepare a batch of unlabelled cDNA and store	no	no	yes
Simplicity of protocol	●	●	○
Time from RNA to probe	2 h	3 h	5.5 h
Suitable for less experienced users	●	●	○
Labelling density	20 – 35 nuc	12 – 25 nuc	12 – 25 nuc

● = Highly recommended ● = Recommended ○ = Suitable — = Not suitable

Chapter 7

CHARACTERIZATION OF LABELLING MICROARRAY PROBES

7.0 Introduction

When performing gene expression microarrays, it is important to characterize the labelled probes in order to avoid experimental error derived from variation between the probes. Even under carefully controlled conditions, some differences in the amounts of nucleic acid samples, labelling success, and recovery of material from the purification system will occur. In order to account for these artifacts and to ensure that these variations are not carried through to hybridization, it is highly recommended that the properties of the labelled probes are determined before microarray hybridization.

Several methods can be used for characterizing the properties of the labelled probes. As a minimum we recommend routinely measuring the amount of CyDye in the labelled and purified sample. If problems do occur in microarray hybridizations, the other methods described below can be used for troubleshooting purposes.

- The quantity of CyDye and nucleic acid in the labelled probe can be determined using spectrophotometry.
- Radioactive spiking can be used to derive information about the amount of nucleic acid in the sample and the success of purification.
- Gel electrophoresis and fluorescence scanning can be used to analyze the molecular distribution of the probe, its purity, and relative fluorescence.

7.1 Determination of the amount of CyDye in a labelled sample with spectrophotometry

Spectrophotometry can be used to determine the amount of CyDye incorporated into labelled nucleic acid. This can be achieved by measuring the absorbance of the solution containing the nucleic acid at the absorption maximum for Cy3 and Cy5. These wavelengths are 550 nm for Cy3 and 650 nm for Cy5. From the known extinction coefficients corresponding to these wavelengths, the concentration and amount of CyDye in the sample can be calculated. The amount of CyDye in the purified sample can be used as a guide to optimize the amount of probe in the hybridization. Best results are achieved when the amounts of Cy3 and Cy5 in dual color hybridization are equal.

It is necessary to purify the labelled nucleic acid before performing the spectrophotometry analysis, as any residual, unincorporated CyDye labelled nucleotides will interfere with the detection of CyDye labelled cDNA.

7.1.1 Measuring the amount of CyDye in the probe with spectrophotometry

Follow the steps below to measure the amount of incorporated CyDye:

❶

Sample preparation

- Dilute an aliquot of purified probe with water. The required volume depends on the size of available measuring cells. It is recommended to use the smallest cells possible. Small volume disposable measuring cells are available, and they offer the added advantage that a separate cell can be used for each sample, thus minimizing cross contamination from one sample to another. However, it is also possible to use regular 100 µl glass cells, but these need to be cleaned thoroughly by rinsing with sterile water between samples.

❷

Measurement

- Measure the absorbance spectrum of the sample from 200 to 700 nm against a blank. Although only the absorbance values at the absorbance peaks are needed to calculate the amount of CyDye in the sample, the shape of the absorption spectra can give additional information about the quality of the labelled sample. The absorption spectrum of CyDye contains two peaks, of which the second should be of higher intensity (Fig 49). The first peak indicates the presence of intermolecular interactions between different dye molecules. If both peaks are of nearly similar intensity, this is a sign of quenching, i.e. loss of fluorescent signal because absorbed light energy is spent on intermolecular interactions. This is usually associated with over-labelling of the sample.

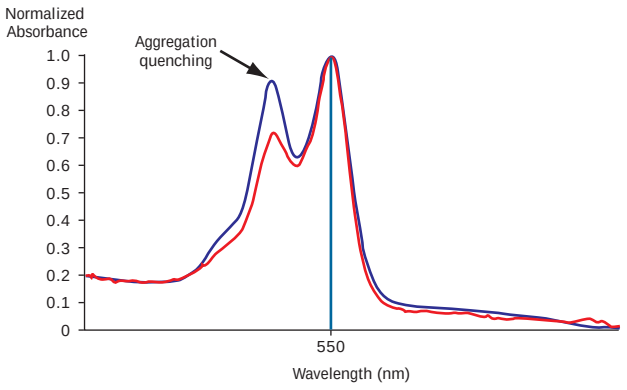


Fig 49. Quantification of the amount of CyDye in a labelled probe with spectrophotometry: examples of absorption spectra from Cy3-labelled samples. The presence of two absorption peaks of near identical intensity (gray line) is a sign of intermolecular interactions and aggregation, which can reduce fluorescent signal from the probe. The red line shows a typical profile of Cy3 absorption spectrum.

- Measure absorbance at 550 nm for Cy3 and at 650 nm for Cy5 using cuvettes with 1 cm path length to determine the amount of CyDye in the sample. The observed absorbance values depend on how much the sample was diluted before measurement, how much RNA was used in the labelling, the labelling method used, and the efficiency of labelling. Typically, values around 0.050 would be expected from first-strand cDNA labelling reactions in which 1 µg of mRNA is used as a template, when the sample is diluted to 100 µl before measurement. Choose a dilution that will give a measurement that is clearly discernible from background values. Sometimes it is necessary to use all of the labelled sample for analysis.
- Recover the measured sample for future use. It may be necessary to concentrate the sample before using for microarray hybridization. This can be performed by drying down the sample in a vacuum concentrator or by letting sample that has been heated to 60 °C evaporate to dryness. It is important to protect the sample from light during all handling.

❸

Calculation

- The amounts of Cy3 and Cy5 incorporated into probes can be calculated from their respective extinction coefficients and using the following equations:

Extinction coefficients

150 000 M^{-1} cm^{-1} at 550 nm for Cy3 and
250 000 M^{-1} cm^{-1} at 650 nm for Cy5

Calculation equations

pmol Cy3 in purified sample =
$(A_{550} / 150\ 000) \times$ dilution factor $\times$ (z µl) $\times$ (w cm) $\times$ 1012
where
A_{550} = absorbance at 550 nm
z µl = the volume of sample after purification
w cm = optical path in cuvette

pmol Cy5 in measured sample =
$(A_{650} / 250\ 000) \times$ dilution factor $\times$ (z µl) $\times$ (w cm) $\times 10^{12}$
where
A_{650} = absorbance at 650 nm
z µl = the volume of sample after purification
w cm = optical path in cuvette

7.2 Determination of the amount of labelled nucleic acid in the sample

7.2.1 UV spectrophotometry

Two methods can be used to determine the amount of labelled nucleic acid in the sample: spectrophotometry and radioactive spiking. Nucleic acids absorb at 260 nm, and absorption measurement at this wavelength can be used for quantitative purposes. However, several purification systems release other materials absorbing near or at this wavelength and, depending on the amounts of these compounds, most of the absorbance at 260 nm will be derived from something other than nucleic acid. Therefore absorption spectra from 200 to 400 nm should be measured, and only if the peak at 260 nm is clearly distinguishable from absorption at near wavelengths, should estimation of nucleic acid amount be made (Fig 50).

The following approximations can be used to calculate the amount of nucleic acid in the probe:

1 A_{260} unit of double-stranded DNA corresponds to 50 µg/ml

1 A_{260} unit of single-stranded DNA corresponds to 37 µg/ml

1 A_{260} unit of single-stranded RNA corresponds to 40 µg/ml

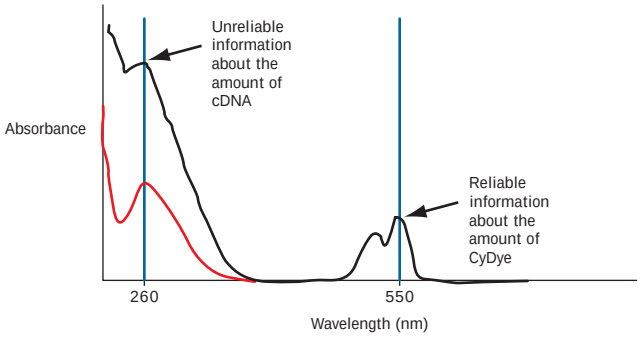


Fig 50. Quantification of the amount of nucleic acid in probe with UV spectrophotometry. The sample, from which the black spectrum was generated, contains impurities absorbing near 260 nm and does not give reliable information about the amount of nucleic acid in the sample. The sample from which the red spectrum was measured is cleaner and the A_{260} can be used for quantification. In both cases, the absorbance spectra for Cy3 at 550 nm gives usable information.

7.2.2 Spiking with radioactive nucleotide

If more accurate quantification of the labelled nucleic acid is needed, the labelling reactions can be spiked with a small amount of a radioactive nucleotide. This enables the determination of the amount of nucleic acid synthesized as well as the amount of sample recovered from purification. This approach can be used with all labelling methods in which new nucleic acid is synthesized. For labelling in cDNA synthesis, a

deoxynucleotide must be used, and for labelling of synthetic RNA, a ribonucleotide is needed. In order to minimize interference with the labelling process, the nucleotides used as spike and label should be different. For example, dATP is suitable for use as a spike with both CyDye-dCTPs and CyDye-dUTPs. Only small amounts of the radioactive nucleotide are needed—2 µCi per sample or less is adequate—and isotopes of lower energy, such as ^{33}P, can be used. Note that the absolute amount of the radioactive spike is not critical, as long as the radioactivity can be measured accurately. The low concentration of radioactive nucleotide solutions means that the spiked nucleotide will not contribute significantly to the total amount of that nucleotide in the reaction. The amount of the radioactive isotope incorporated into labelled sample is relatively small and does not interfere with the detection of fluorescence.

The information derived from the radioactive spike can be used to estimate the amount of nucleic acid synthesized. It is reasonable to assume that the radioactive nucleotide and unlabelled nucleotide are incorporated by enzymes at a similar rate. Therefore, if the percentage of the radioactive nucleotide incorporated into the synthesized product is known, it can be concluded that the same percentage of the unlabelled nucleotide is incorporated as well. If the total amount of this nucleotide in the labelling mix is known, and it is assumed that all four nucleotides are incorporated with equal efficiency (this is probably not absolutely true), it is possible to calculate the amount of nucleic acid synthesized. This method only estimates the amount of nucleic acid synthesized, but in most cases gives more accurate data than UV spectrophotometry.

In order to quantify the amount of nucleic acid synthesized by spiking, a small amount of radioactive nucleotide is added to the labelling reaction. In order to keep the total reaction volume unchanged, the amount of water pipetted into the reaction must be adjusted. The radioactive nucleotide should not contain any colored marker or stabilization agents, as these compounds are fluorescent in the same regions of visible spectrum as CyDye and can interfere with determination of CyDye amount.

Note: Because of the radioactivity present in the spike, necessary precautions for working with radioactive materials must be followed.

No other special modifications are needed for the labelling reactions. After the labelling reaction has been completed and before the sample is purified, the incorporation of the radioactive nucleotide into the synthesized nucleic acid needs to be determined. The incorporated radioactive nucleotide can be easily separated from free nucleotides by thin layer chromatography on PEI cellulose chromatography plates (Fig 51).

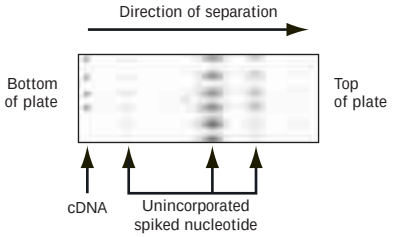


Fig 51. Thin layer chromatography analysis of the incorporation of radioactive spike into labelled cDNA sample. The labelled nucleic acid does not migrate far from the sample application line, whereas the unincorporated radioactive nucleotide moves up toward the top of the plate.

7.2.3 Thin layer chromatography analysis of spike incorporation

To perform thin layer chromatography, follow the steps outlined below:

❶

Preparation

- Prepare a 20 × 20 cm glass-backed PEI cellulose chromatography plate (available from Merck) for use by cutting a thin linear groove into the PEI cellulose layer at 1 cm distance from the top edge of the plate. Mark sample positions along a line that is 3 cm from the bottom edge of the plate.

❷

Titration

- Pipet 0.5 µl samples of labelling reactions in duplicate along the marked line. The sample spots should be about 1 cm apart. Do not damage the PEI cellulose layer with the tip.

❸

Separation

- Place the plate in a rectangular chromatography tank so that the bottom of the plate is immersed 2 cm deep in 1 M K_2HPO_4. Make sure that the level of the buffer is below the level of the marked sample line. Cover the tank and let sample separation take place. Newly synthesized nucleic acid will not move far from the sample line whereas free nucleotides will move progressively towards the top of the plate.
- When the buffer has reached the top groove, i.e. when the samples have migrated the full available length of the 20 cm plate, remove the plate from the tank and let it air dry.

❹

Imaging

- Wrap the plate in cling-film and expose it to a phosphor screen for 1–6 h. Take care not to overexpose the phosphor screen as it will saturate the signal. Some trial and error may be needed to determine correct exposure time.
- Scan the phosphor screen on a Typhoon™ Variable Mode Imager, using the recommended settings. As an alternative to using phosphor screens, any instrument that can accommodate chromatography plates and measure radioactivity quantitatively can be used.

❺

Analysis

- Calculate the proportion of the radioactive nucleotide that is associated with the synthesized nucleic acid. This is the incorporation percentage. For example, the ImageQuant™ software can be used for this purpose.
- Calculate the yield of nucleic acid as follows:

  Yield of nucleic acid =

  (mol of cold nucleotide in labelling mix) × (incorporation percentage) × 4 × 330 g/mol

  This formula assumes that all four nucleotides are incorporated in equal proportions (hence, times 4) and that the molecular weight of average nucleotide is 330 g/mol.

  For example first-strand cDNA labelling was performed with 2 nmol of dATP in the reaction. 20% of [α-^{33}P]dATP was incorporated into cDNA.

  Yield of cDNA =
  2×10^{-9} mol × 20% × 4 × 330 g/mol = 528 ng

7.3 Calculation of labelling density

Labelling density can be defined as the amount of CyDye incorporated into a known amount of nucleic acid. It can be expressed as pmol of CyDye incorporated into µg of nucleic acid, or can be converted to the number of CyDye molecules per 100 nucleotides. Labelling density is a measure of the average distance between CyDye fluors on the labelled nucleic acid. Samples labelled successfully, with an optimized protocol, will usually have similar labelling densities, but problems with the labelling reagents or in performing the protocol can result in variation in the incorporation of CyDye into cDNA. For example, in the cDNA post-labelling method, exposure of the CyDye-NHS esters to moisture during storage would result in low labelling density without affecting the amount of cDNA synthesized.

Once the amount of CyDye incorporated into nucleic acid and the amount of the nucleic acid are known, labelling density can be calculated.

$$\text{Labelling density} = \frac{\text{pmol CyDye in labelled sample}}{\text{µg nucleic acid in labelled sample}}$$

1 µg of cDNA contains approximately $1 \times 10^{-6} / 330 = 3030$ pmol of nucleotides.

Hence labelling density of 100 pmol/µg equals 100 pmol/3030 pmol of nucleotides = 3.3 CyDye nucleotides per 100 nucleotides = 3030 pmol of nucleotides.

7.4 Recovering labelled nucleic acid after purification

The recovery of CyDye labelled nucleic acids from purification systems can be variable. In order to draw conclusions about the performance of the purification process, two aspects need to be considered. First, the recovery of the labelled material needs to be determined. Poor recovery can limit the amount of slides that can be hybridized with the sample, and the true benefits of labelling methods may not be realized. Second, the presence of free CyDye needs to be assessed. Free CyDye in the purified probe will affect the determination of CyDye amount in the labelled nucleic acid and can result in too little probe being used in hybridization. Free CyDye, especially free CyDye-

$$\text{recovery \%} = \frac{(\text{cpm after}) \times (\text{volume recovered}) \times 100\%}{(\text{cpm before}) \times (\text{incorporation \%}/100) \times (\text{volume purified})}$$

nucleotide, can also contribute to hybridization background and give rise to speckled images that are difficult to quantify. Gel electrophoresis is a convenient method for determining whether the purified probes are free of unincorporated dye.

Radioactive spiking, as explained above, can be used to obtain information about the purification process. Scintillation counting of a small aliquot (0.5–1 µl) of labelling reaction sampled before and after purification can be used to calculate the proportion of cDNA (or other nucleic acid) recovered. For this calculation it is necessary to know what proportion of the radioactive nucleotide was incorporated into the nucleic acid. This can be determined from the unpurified labelling reaction with thin layer chromatography analysis as explained above.

7.5 Analysis of the composition and fluorescence of labelled sample

Gel electrophoresis can be used to investigate the quality of the labelled sample. Separation of unincorporated CyDye from labelled nucleic acids can be achieved with denaturing polyacrylamide gel electrophoresis (PAGE) or agarose gel electrophoresis. After electrophoretic separation, the gel can be scanned to detect Cy3 and/or Cy5 fluorescence using a multipurpose scanner, such as Typhoon 9410 Variable Mode Imager. From the scanned image it is possible to detect how much free CyDye there is present in the sample. If PAGE gels are used, the size range of the labelled nucleic acid population can be examined. This can give an indication of the quality of the starting population of RNA and success of the labelling reactions. If unpurified sample is separated alongside purified samples, it is possible to estimate the recovery of probe from purification. By comparing the relative fluorescence of different labelled samples, the relative amount of CyDye in each sample can be estimated (Fig 52).

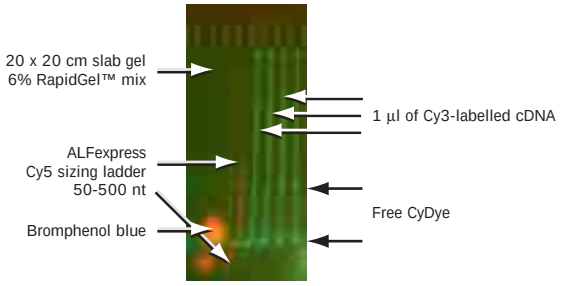


Fig 52. PAGE analysis of Cy3-labelled cDNA probes. 1 µl samples of Cy3-labelled and purified cDNA probes prepared with cDNA post-labelling method were separated in 6% sequencing gel. The gel was scanned for Cy3 and Cy5 fluorescence with Typhoon scanner to detect ALFexpress™ Sizer and the labelled probes The cDNA probe consists of fragments longer than 150 nucleotides and some unincorporated Cy3-reactive dye is present, as indicated. All four samples show similar relative fluorescence and quality.

7.5.1 Analyzing CyDye labelled probes with PAGE

In order to analyze CyDye labelled probes using PAGE, follow the procedures detailed below:

❶

Preparation

- PAGE gels suitable for analysis of CyDye labelled samples can be prepared from standard 6% or 8% (w/v) sequencing gel mix such as RapidGel-XL-6%. The gel can be cast using equipment designed for sequencing, but since single base pair resolution is not required, slab gel instruments can be used. These provide the added benefit of thicker and deeper wells that simplify sample loading.
- 0.1–1 µl of purified labelling reaction is enough for detection of CyDye labelled nucleic acid by fluorescence scanning. Add 2 µl of formamide and 8 µl of water to each sample. Do not use normal loading/denaturation buffers which contain dyes such as bromophenol blue or xylene cyanol, as these will interfere with the detection of CyDye fluorescence.
- Dilute 4 µl of fluorescent markers such as ALFexpress Sizer 50-500 with 4 µl of water and 2 µl of formamide. This sizer contains a Cy5-labelled DNA marker ladder.

❷

Denature

- Denature all samples by boiling for 2 min at 95 °C. Snap cool on ice before loading on to the gel.

❸

Electrophoresis

- Load samples to a gel that has been pre-electrophoresed for 15–30 min. Perform electrophoresis according to the instructions provided with your equipment. Use 1× TBE as buffer. Protect the samples from light during the electrophoresis.
- In order to help monitor the progress of electrophoresis, you can load a small aliquot of loading buffer (1 µl) containing bromphenol blue into a side well of the gel that is well separated from the wells containing labelled cDNAs. Stop the electrophoresis when the bromphenol dye is still well within the gel. Unincorporated CyDye will migrate faster than bromphenol blue.
- Remove one of the gel plates before scanning and make sure that the back of the remaining plate is clean. Do not let the gel dry before scanning.

❹

Imaging

- Scan the gel on a Typhoon Variable Mode Imager. Detect Cy3 by excitation with 532 nm laser and using emission filter 555 BP 20. Detect Cy5 by excitation with 633 nm laser and using emission filter 670 BP 30. Set PMT to 800 V, focal plane to +3 mm, and use normal sensitivity. The PMT values may need to be adjusted to account for different amounts of sample in the gel.

Agarose gel electrophoresis can be used instead of PAGE. Single-stranded CyDye labelled nucleic acid fragments will not migrate true to their size in standard agarose gels, but valid information about the purity and fluorescence of labelled sample can be obtained. 0.1–1 µl of labelling reaction can be diluted by adding 2 µl of 50% (v/v) glycerol and water to 10 µl. No loading dyes should be used. Denaturation of samples is not necessary before gel loading. Run the gel in 1× TBE or 1× TAE.

Chapter 8

MICROARRAY HYBRIDIZATION

8.0 Introduction

The process of hybridization is typically performed in order to identify and quantitate nucleic acids within a larger sample. Generally, it involves annealing a single-stranded nucleic acid to a target complementary strand. Southern blotting is one well-established hybridization method. In this technique, samples of genomic DNA—the target—are attached to a membrane and incubated within a solution of fluorescently labelled DNA—the probe (54). Binding of probe molecules to the sample on the membrane highlights complementary sequences, and the intensity of signal is proportional to the amount of immobilized sample.

The microarray hybridization technique works in a very similar way to that of Southern blotting, except that it is carried out in reverse, and the target is first attached to a slide instead of a membrane.

There are several critical factors to performing a successful microarray hybridization. The following are discussed in detail in this chapter:

- Pre-hybridization
- Hybridization conditions
- Hybridization buffer
- Stringency washes

8.1 Overview of the microarray hybridization process

❶

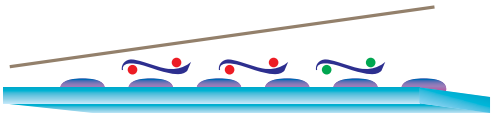
Attachment

Genes of interest are spotted onto a solid surface by the array spotter. These are known as the targets. Attachment chemistry will often be required to ensure that the DNA remains attached to the slide surface throughout the hybridization process.

❷

Hybridization

Hybridization buffer containing a known amount of labelled sample DNA—often referred to as probe— is then placed on the slide surface. A coverslip can then be carefully placed on top of the slide.



The slide is then incubated in a humid environment for up to 16 h. During this time the labelled probe is in contact with the targets on the slide. If the sequence homology is good then the probe will adhere to the target.

❸

Washing

Once the hybridization is complete, the slides are washed, and buffer and probe of little or no homology to the target will be washed away, leaving the labelled probe of high homology attached to the target and available for detection.

8.2 Pre-hybridization

Pre-hybridization consists of incubating the spotted slide in a buffer in the absence of probe. Different slide chemistries require slightly different pre-hybridization protocols that vary in the type of buffer used. Consult manufacturers' recommendations to find out what is the best procedure for each slide. Pre-hybridization prepares the microarray for hybridization in the following two ways:

- Badly adhered target is washed away during pre-treatment. If this step is omitted, this target will often wash off the slide surface during the hybridization and will hybridize to the probe in solution, thus competing with the immobilized targets. This can decrease hybridization signal.
- Pre-treatment ensures that the target is available for hybridization. The target is normally double-stranded DNA and, although targets are frequently spotted in a denaturing solution such as DMSO, most microarray protocols do not contain a specific denaturization step. Pre-hybridization may also act to block any sites on the slide surface that are capable of binding the probe nonspecifically, thus improving the backgrounds.

8.3 Hybridization

There are several widely used methods for carrying out the hybridization, either using automated instruments or performing the procedure manually. In this chapter we describe the manual hybridization method, which is the most widely used. General properties of manual hybridization will be discussed, followed by advice for choosing a suitable hybridization method for different microarray slide types.

8.3.1 Coverslip method

In the coverslip method, hybridization buffer containing the probe is incubated on the microarray under a coverslip. This way only a small volume, typically about 30 µl, of buffer is needed. The coverslip method is a very convenient one in that it requires no special equipment. However, a microarray slide under coverslip is prone to drying out, especially around the edges, causing most of the array to become unusable. Originally, coverslips were sealed on the slide. This prevented drying but made the coverslip difficult to remove prior to detection. A more practical approach is to carry out the hybridization in a humid environment, thus preventing evaporation of the hybridization buffer from beneath the coverslip. This can be achieved with anything from a

a)

b)

Fig 53. Practical solutions for ensuring that a humid environment is maintained during manual microarray hybridization. A plastic box (a) containing a platform raised above moistened tissues is sufficient for hybridizing a few slides. Commercially available humid chamber (b) holds up to 40 slides on removable trays and fits into most lab ovens.

humid chamber with slide trays and a reservoir, to a plastic box lined with wet tissues at the base (Fig 53). In either case, the atmosphere within the chamber must remain at >95% humidity throughout the 16-h hybridization. Equally important is not to allow the slide to come in contact with the water, which may dilute the probe or cause water marks on the slide surface.

8.3.2 Hybridization buffer

The hybridization buffer and conditions used are vital for successful results. Hybridization buffers vary considerably but will normally contain the following components:

- a buffering component that acts to stabilize variations in pH
- a detergent that acts to lower the surface tension and allow the buffer to flow easily under a coverslip
- compounds that act as rate enhancers, volume excluders, or to speed up the hybridization and lower the T_m

Melting temperature (T_m) is the temperature at which 50% of the probe is denatured. This temperature will be affected by both the size and G-C content of the probe fragments, but the effect is minimized by optimizing the salt content and formulation of commercial hybridization buffers, thus making them suitable for use with most probes without optimization. Formamide is a denaturing reagent that is often used to lower the T_m of the probe and hence the temperature of hybridization. The optimum hybridization temperature for microarrays, in aqueous buffers, will be high (65–75 °C). At these high temperatures drying out of the slide becomes more of a problem; the probe is also more likely to degrade. The addition of formamide to a buffer decreases the T_m by 0.65 °C for every 1% concentration; therefore, the addition of 50% formamide to the hybridization lowers the optimum temperature to a more reasonable 42 °C (55). However, hybridizations carried out in formamide should be left for 16 h, unless the probe concentration is increased.

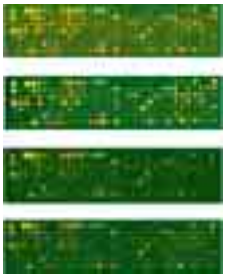


Fig 54. Cy3- and Cy5-labelled probes are hybridized on Amersham Biosciences reflective slides at 75, 25, 8 and 3 pmol of each dye per slide. Although hybridization signals can be detected using as little as 3 pmol of each probe, the intensity of signals is greatly increased by using 25 pmol of each probe, thus allowing the rarer messages to be visualized.

8.3.3 Probe blocking

Most manufacturers recommend some type of probe blocking either prior to or during the hybridization to prevent nonspecific hybridization of probe to common genetic elements. One common blocking agent is poly-dA oligo, which hybridizes to poly-dT tails (formed during the cDNA probe synthesis by the poly-dA oligo, and prevents probe from hybridizing with poly-A sequences often present in targets. Other types provide more general forms of blocking, such as salmon sperm and yeast tRNA to block non-specific binding and the inclusion of Cot1 DNA™ mop up repetitive sequences. The blocking agents are normally added to the labelled probe/hybridization buffer solution prior to applying to the slide surface. The solution can then be heated to denature any double-stranded DNA and to allow the blocking to take place, before setting up the hybridization reaction.

8.3.4 Probe concentration

The amount of probe to add to a hybridization will vary, depending on the samples used, the slide type, and what information is expected to be gained. Slide manufacturers will recommend optimum probe concentrations to use in hybridizations with their slides. If two or more colors are being used, it is important that exactly the same amount of probe labelled with each dye is added so as not to skew the results in favor of one of the probes. For glass slides 30 pmol of each labelled dye is sufficient for most systems, but this should be reduced by as much as half when using mirrored slides, which contain a reflective layer capable of enhancing signal intensities. Increasing the amount of probe used will increase the result obtained but only up to a certain point (Fig 54), beyond which the increase in background levels will actually decrease the amount of background-corrected signal (Fig 55).

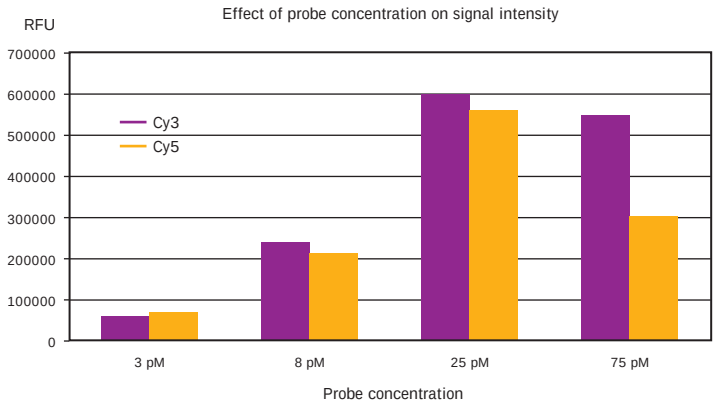


Fig 55. When increased amounts of labelled probe are used, there is an increase in background levels. This decreases the amount of background-corrected signal detected from the microarray slide.

8.3.5 Probe depletion and target saturation

As in most manual hybridizations the reaction is carried out under a coverslip, which means that there is sufficient solution under the coverslip for complete coverage of the array but no room for any movement or flow of the buffer under the coverslip during the hybridization. Whether a labelled probe fragment finds its complementary target on the microarray is therefore simply relying on diffusion of the probe. Research has shown that a 20 bp oligo diffuses a distance of 3.6 mm in an 18 h period, therefore a labelled cDNA sample is hardly going to move during an overnight hybridization. This means that if there were several replicate target spots concentrated in a small area on an array, these spots would be competing for a limited amount of complementary sequences within diffusion distance. This is called probe depletion, and it can limit the signal obtained from microarray. This will be most relevant for those transcripts that are present in low numbers in the labelled samples, as the signal from these spots may fall below detection sensitivity of the microarray system.

The sensitivity of the microarray system is determined by several factors including the amount of label attached to the probe molecules, the level of background signal, and the sensitivity of the scanner. Furthermore, the rate at which the probe molecules find their targets is a more critical determinant of sensitivity than the amount of spotted target. For most low- to medium-abundance genes, the amount of spotted target is in a huge excess over the probe molecules. For a high abundance gene, the amount of probe in solution starts to approach the amount of target present, which can lead to target saturation. Target saturation will be determined by factors such as the amount of target initially spotted on the slide and the amount retained on that slide after pre-treatment, as well as the percentage of that target available for hybridization and the efficiency of hybridization. Together, the sensitivity of detection and target saturation determine the dynamic range of the microarray experiment.

8.3.6 Use of hybridization chambers

Hybridization chambers are used in order to overcome any problem arising from the amount of buffer used under coverslips and possible probe depletion. These are plastic chambers that hold an individual slide and a larger volume of hybridization buffer. The amount of labelled probe added remains the same. These are then incubated overnight in a water bath or oven. In order to introduce a significant amount of mixing, using an automated hybridizer is recommended.

8.3.7 Practical tips for setting up coverslip hybridization

This protocol gives instructions how to set up a microarray hybridization using coverslips. It should be performed with clean slides and coverslips, preferably in a clean room or under a hood. It is recommended to use plastic coverslips that have been packed in plastic film. Do not use gloves that contain powder, as this can easily get onto microarray slides and cause background problems.

❶

Pre-hybridization

- Store spotted slides in a desiccator until use.
- Read through the pre-treatment and hybridization protocols thoroughly before use, as buffers often need preparing and preheating before use.
- Pretreat the number of slides required for the experiment. Some manufacturers say that pre-treated slides can be stored before use, but not all, so it is worth checking before doing a large batch.
- During the pre-treatment stage, prepare the probes. This will often involve drying down equivalent amounts of the two probes of interest together and reconstituting them in the manufacturer's recommended buffer. Some protocols require heating the probe before use; the probe prepared by reverse transcription will be single-stranded and therefore should not require denaturing before use.

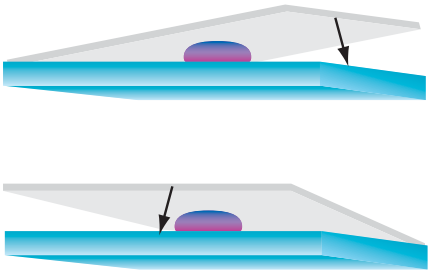


Fig 56. Lowering the coverslip.

❷

Hybridization

- Once the probe is prepared, lay the spotted slide, DNA side up, on a clean surface. Absorbent tissue that does not release any fibers is a good choice of surface. This is important, as most slides are glass, and dirt on the rear of the slide will affect the result on the front of the slide (this is obviously not an issue with opaque or mirrored slides).
- Using a pipette, transfer the required amount of hybridization buffer/probe mixture onto the slide. Avoid touching the slide surface with the pipette tip. Try to deposit the mixture along the short side of the slide, away from the spotted area.
- Take a clean coverslip and place it on the slide near the probe mixture, and allow surface tension to speed the buffer along the coverslip. Then gently lower the coverslip, avoiding trapping air bubbles underneath. There are several ways of lowering the coverslip, two of which are illustrated in Figure 56.
- If air bubbles become trapped beneath a coverslip (Fig 57), do not move the coverslip to try and remove them. Movement of the coverslip will result in damage to the targets themselves. Most small air bubbles will disperse once the slide is transferred to hybridization temperature. Larger bubbles can be 'encouraged' to move by gently pressing on the surface of the coverslip with a pipette tip.
- The probe mixture is light sensitive, so once the coverslip is on, place the slide in the humid chamber and incubate overnight in the dark.

Note: If using RNA probes, it is important to take appropriate precautions to protect all reagents from nucleases (see Chapter 5 on RNA handling).

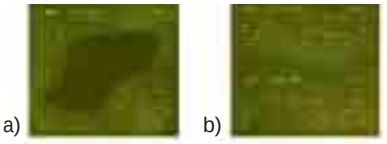


Fig 57. Typical problems encountered in microarray hybridization. Trapping of air bubbles (a) beneath the coverslip will lead to areas on the array that fail to hybridize at all. Allowing the slide to dry out during the hybridization will lead to high patchy backgrounds (b) that may cause difficulty at the analysis stage.

8.4 Stringency washes

The purpose of the post-hybridization washes is to remove all unattached and loosely bound probe molecules. This prevents false positive signals and removes all components of the hybridization buffer, preventing background noise in the form of smearing and speckles. Again, as the slides are light sensitive at this stage, the washing steps should be carried out in the dark so as to minimize signal loss due to bleaching of the fluorescent dyes. Once the slides have been washed, they should immediately be dried by centrifugation or nitrogen steam to prevent smearing while drying. The slides should then be stored in the dark in a desiccator and scanned as soon as possible. If, once scanned, it is found that the slides have high background or low stringency, it is worth re-washing the slide and re-scanning.

The stringency washes will affect the amount of labelled probe retained on the slide for the final analysis. While it is obviously important to remove all the loosely bound probe, it is important to not strip the bound probe. Generally, stringency washes are carried out in a SSC/SDS solutions of different concentrations, with the primary washes often being carried out at the same temperature as the hybridization. Primary wash solutions have a high salt content (typically 1–2× SSC/0.1% SDS buffer), and they remove most of the hybridization buffer components. The secondary washes are performed with low salt buffer (typically 0.1× SSC/0.1% SDS), and they will remove the loosely bound probe from the blot. It also serves to remove any remaining salt from the primary washes. Failing to warm solutions thoroughly before use will lower their effectiveness and may lead to increases in background noise. Conversely, warming solutions too much (as often happens if a microwave oven is used) or using too low a salt concentration in the buffers, will strip precious signal from the blot. Check the manufacturer's protocols for exact wash dilution volumes. Manufacturers often suggest a water dip prior to drying the slides to prevent smearing. Check the protocols provided with the buffer components for instructions on performing the water dip.

8.5 Microarray slides

8.5.1 Choosing the right hybridization protocol for different slide types

Most manufacturers of microarray slides will provide a hybridization protocol that they have optimized for their system. The following table lists some of the more commonly used slides and a brief summary of tested hybridization protocols for them. This in not an exhaustive list and the protocols are only for reference. Please refer to the manufacturer's own protocols prior to use.

Table 5. Microarray slide types and their characteristics.

Slide type	Manufacturer	Spotting chemistry	Hybridization buffer	Pretreatment protocol	Hybridization
Lucidea Reflective Slides	Amersham Biosciences	50% DMSO	Version 2 (4×)-formamide based	None required	42 °C overnight
CMT-GAPS	Corning	50% DMSO or 3× SSC	25% formamide, 5× SSC, 0.1% SDS	25% formamide, 5× SSC, 0.1% SDS for 45 min at 42 °C	42 °C overnight
SigmaScreen™	Sigma-Aldrich	3× SSC	ArrayHyb	1% SDS for 2 min, water rinse. Boiling water for 2 min, ethanol dip	50 °C 6 h overnight
Type I	Clontech	150 mM Na phosphate	GlassHyb	Optional: 70 mM succinic anhydride in 315 ml 1-methyl-2-pyrrolidinone and 35 ml Na borate pH 8–15 min at RT. Boiling water for 2 min. Ethanol dip.	50 °C overnight
Type II	Clontech	150 mM Na phosphate	GlassHyb	70 mM succinic anhydride in 315 ml 1-methyl-2-pyrrolidinone and 35 ml Na borate pH 8–15 min at RT. Boiling water for for 2 min. Ethanol dip.	50 °C overnight
Super Amine	Telechem Int.	5× SSC	UniHyb (1.25×)	0.1% SDS twice for 2 min at RT followed by a water rinse. Incubate in bolling water for 3 min before drying.	42–65 °C overnight

Chapter 9

LUCIDEA SLIDEPRO HYBRIDIZER

Fig 58. Lucidea SlidePro Hybridizer.

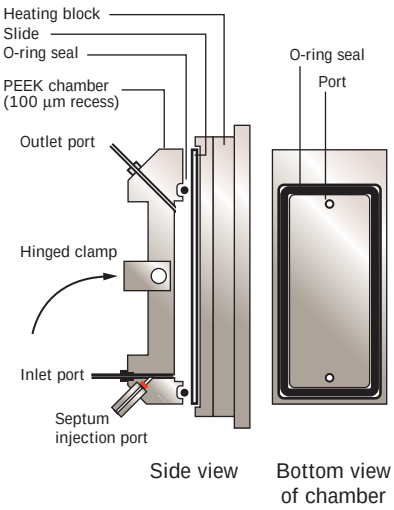


Fig 59. Schematic of individual slide chamber.

9.0 Introduction

A complex system, microarray analysis is affected by a number of experimental factors, including target preparation, physical deposition of the targets, slide chemistry, probe chemistry, hybridization, and detection of the fluorescent signal (37). In order to improve the efficiency of microarray analysis, each source of variation must be eliminated or minimized. Control of environmental conditions during hybridization in particular, is critical in producing and maintaining a consistent fluorescent signal. Lucidea SlidePro was designed to overcome the problems associated with hybridization variability.

9.1 Features of Lucidea SlidePro Hybridizer

Lucidea SlidePro Hybridizer (Fig 58) has a modular format that consists of a base control unit, up to four additional modules, and control software on a laptop computer. Each unit contains six individually temperature-controlled chambers, each of which holds a standard microscope slide. With all five modules a total of 30 slides can be processed. The multi-module protocol software allows each module to be started at different times and different experiments can be conducted on each module, thereby increasing user flexibility. Also, each module has its own pump, which allows a faster processing time.

Lucidea SlidePro is capable of automating a variety of chemical and biochemical techniques in which incubation and wash steps are performed at varying temperatures. It is primarily used in microarray analysis to automate the pre-hybridization, hybridization, and washing of microarray slides. It has been designed to be used in conjunction with the Lucidea range of microarray instruments and reagents.

In Lucidea SlidePro, each slide is held in a chamber sealed with a patented O-ring. Pre-treatment and wash solutions are drawn into the chamber, from up to five reservoirs, and deposited into a waste bottle. Each module can run off one set of wash botttles, or multiple modules can run off the same set of wash bottles. Hybridization samples are injected through a septum port at the lower end of the chamber (Fig 59).

9.2 Benefits of Lucidea SlidePro Hybridizer

The key benefits of Lucidea SlidePro include:

- Improved uniformity of signals and Cy3/Cy5 ratios within and between slides.
- Temperature and mixing controls for each chamber. Small volumes are drawn in and out of the chamber to provide continuous mixing. The volume, speed, and length of mixing can also be individually controlled, making optimization easier.
- Temperature and wash solution controls for each chamber. This improves the reproducibility of signals both within and between slides and from user to user. Greater reproducibility can increase the accuracy of results with fewer numbers of replicates.
- Enhanced detection of rare messages. The mixing of probe during hybridization results in enhanced signals, without the need to use increased amounts of probe as compared with manual hybridization. Increased signal strength compared to background enables detection of low signals from rare messages.
- Rapid start-up time and ease of use. Experiments to determine optimal hybridization parameters, such as temperature and washing, are performed with ease. Standard protocols are provided to decrease time to optimize experimental procedures. Different conditions can be tested within a single run, using up to 30 slides with five modules. The software has a help feature for fast start-up and troubleshooting.
- Facilitates reuse of probe. Probe can be removed from the hybridization chamber via the injection port, allowing samples to be reused multiple times. The instrument is paused after hybridization and the probe removed with a syringe. The instrument continues with washing and drying of slides. While overall signals decrease with probe reuse, the Cy3/Cy5 ratios are not significantly altered.
- Reduced demands on user time. The user needs only to load slides and inject the probe. Following automated hybridization, washing, and drying, the slides are removed from the instrument ready to scan.

9.3 Validation in microarray hybridization

We describe here how a number of hybridization factors—including time, probe reuse and probe concentration—affect the signal-to-noise values detected from microarray experiments. Despite the effects of these factors on overall signal strength, the Cy3/Cy5 ratio remains constant, suggesting that differential expression may be determined under conditions which do not provide an optimal signal. Three experimental applications were used to demonstrate the utility of Lucidea SlidePro.

9.3.1 Comparison of automated and manual hybridization

Twenty-four standard silanized glass slides were spotted with p53 cDNA. Twelve slides were hybridized manually (approximately 30 µl of hybridization sample was placed on a microarray slide, under a coverslip and incubated in a humidified container in a hybridization oven) and twelve were hybridized in Lucidea SlidePro. All were hybridized using Cy3-labelled human skeletal muscle and Cy5-labelled human skeletal muscle.

Hybridization in Lucidea SlidePro produced increased signal intensity and more consistent Cy3/Cy5 ratios with very low variability compared to the manual method. Analysis of variance (ANOVA) comparison of the Cy3/Cy5 ratios showed significantly less variation in the Lucidea SlidePro processed slides compared to manually processed slides (Fig 60).

Coefficient of variation		
Type 5	Lucidea SlidePro	Manual
Within slide	8.6%	35.9%
Between slides	4.1%	10.7%
Total	12.7%	46.6%

Fig 60. Lucidea SlidePro vs manual ANOVA coefficient of variation values for glass slides hybridized in Lucidea SlidePro or manually.

Results suggest that hybridization efficiency and data reproducibility could be improved using Lucidea SlidePro as compared to the manual methodology (Fig 61-62).

9.3.2 The effect of probe mixing

Since diffusion rates on solid surfaces are much lower than those in solution, compensation for localized depletion of probe may not occur within the time frame of a hybridization (37, 56). Lucidea SlidePro provides mixing during hybridization, which ensures a constant probe concentration and thereby eliminates depletion effects.

Initial experiments were designed to determine whether mixing could enhance hybridization efficiency. The relative signals of a serial dilution of target hybridized under static or mixing conditions were assessed (Fig 63). Despite greater than 300-fold dilution of probe, mixing during hybridization enhanced the signal detected around 5-fold at the highest concentration of target. This result suggests that localized probe depletion effects could be reduced by mixing the sample.

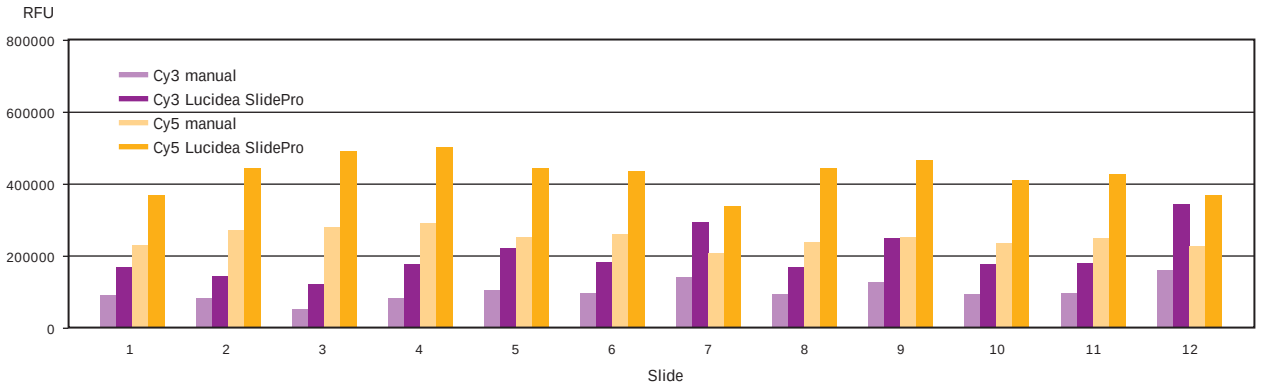


Fig 61. Lucidea SlidePro vs manual mean signal intensities.

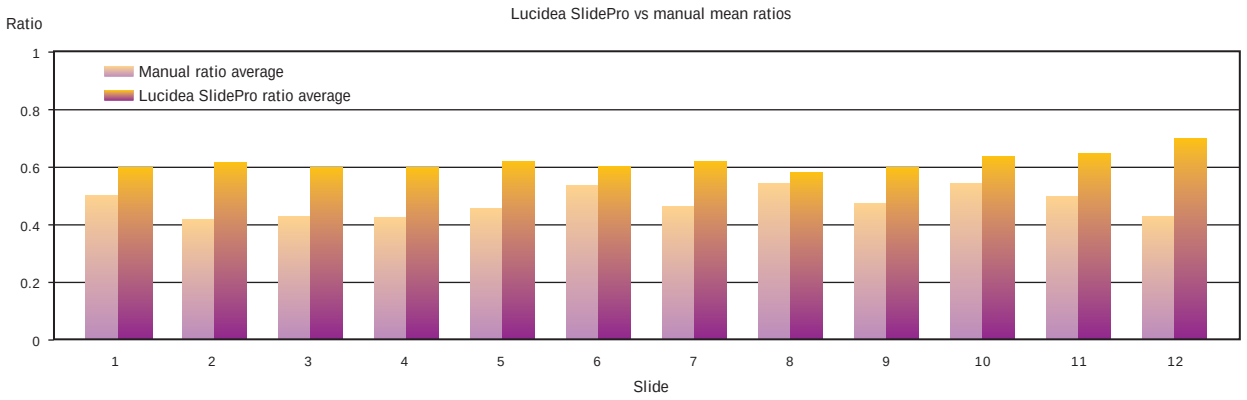


Fig 62. Lucidea SlidePro vs manual mean Cy3/Cy5 ratios.

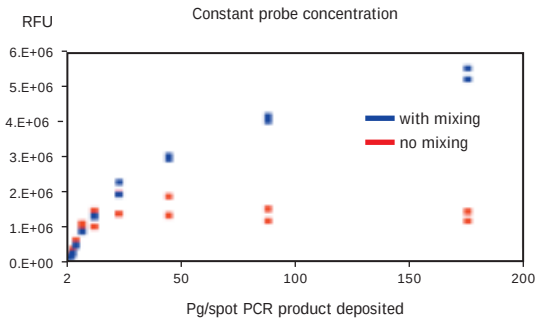


Fig 63. The effect of mixing on hybridization signal.

9.3.3 Reuse of probe

Since probe is often generated from mRNA samples that are in short supply, it may be desirable to use a labelled probe for multiple rounds of hybridization. This is possible using Lucidea SlidePro, since probe can be removed directly through the injection port following hybridization. The effect of probe reuse on overall signal strength and Cy3/Cy5 ratios was assessed. Duplicate reflective slides were hybridized in Lucidea SlidePro with 200 µl of Cy3-labelled skeletal muscle and Cy5-labelled placenta cDNA (40 pmol total) per chamber. Following hybridization and immediately prior to washing the slides, probe was recovered from all chambers using a syringe/needle through the injection port. Approximately 60% of the injection volume (120 µl) was recovered from each chamber. Each probe was then reconstituted to the original volume of 200 µl in 1× version 2 hybridization buffer and reinjected into chambers containing fresh slides. This constituted the first reuse of probe. The procedure was repeated for the second reuse.

Although the signal strength decreased over multiple hybridizations (Fig 64), the Cy3/Cy5 ratios were relatively unaffected (Fig 65). This suggests that the individual Cy3/Cy5 ratios for each gene should remain constant, provided the signal is strong enough to be detected above background. Furthermore, overall background was reduced with multiple rounds of hybridization, providing an additional benefit.

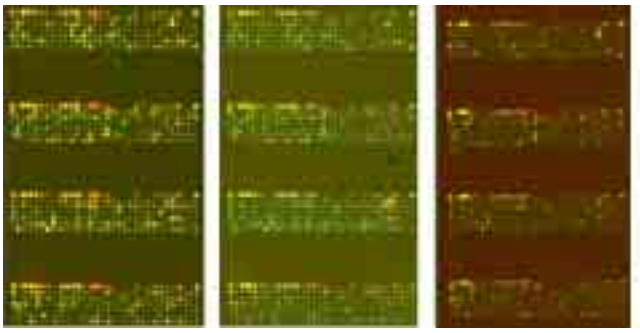


Fig 64. Reuse of probe. Scanned images showing arrays following overnight hybridization in Lucidea SlidePro with fresh probe, first r euse, and second reuse.

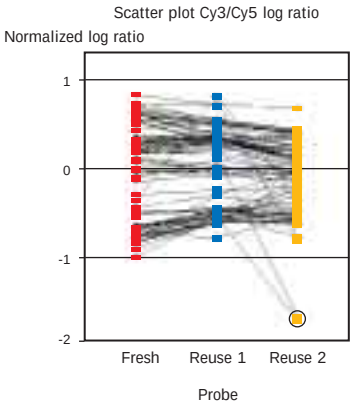


Fig 65. Reuse of probe. Cy3/Cy5 ratios of selected genes following hybridization with fresh probe, reuse 1, and reuse 2.

9.3.4 Length of hybridization

Standard microarray protocols call for overnight hybridization (12–18 h). Time course studies suggest that effective hybridization can take much longer than this (data not shown). In order to assess whether Lucidea SlidePro enhances the kinetics of hybridization, the relative signal strength following different hybridization times was investigated. Arrayed reflective slides were hybridized with 20 pmol each of Cy3 skeletal muscle and Cy5 placenta probe per slide. Total mean signal intensities (Fig 66) increased with time up to 16 h hybridization, but Cy3/Cy5 ratios (Fig 67) remained consistent between the 3-h, 6-h, and 15-h timepoints.

Another set of microarray slides were spotted with Lucidea Universal ScoreCard dynamic range controls (control elements that are used to evaluate the dynamic range and sensitivity of the system), and the amount of probe was increased to 60 pmol labelled probe per slide. Slides were hybridized for 3, 6, 9, 12, and 15 h in Lucidea SlidePro. Mean signal intensities were similar (Fig 68) for all hybridization times with dynamic ranges representing 2000 copies (DR 2), 200 copies (DR 3), and 20 copies (DR 5). These results suggest that shorter hybridization times may be used for higher throughput, provided that sufficient signal is obtained to detect low expressing genes of interest. Furthermore, increased probe concentrations may be used to decrease the hybridization time required for signal detection provided that the mean signal above background is not compromised.

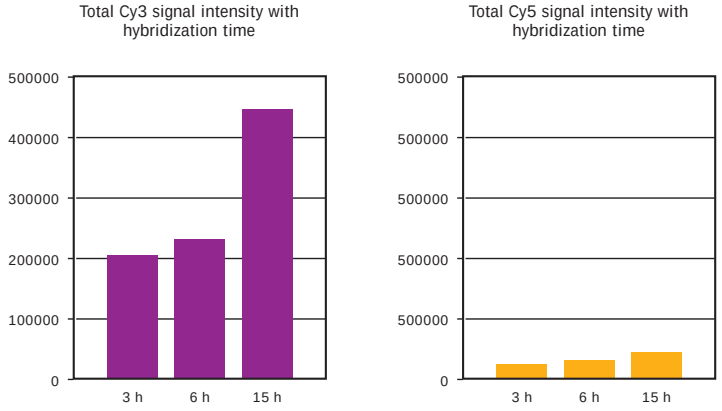


Fig 66. Effect of hybridization time. Total mean Cy3 and Cy5 signals following 3-h, 6-h, and 15-h hybridization in Lucidea SlidePro.

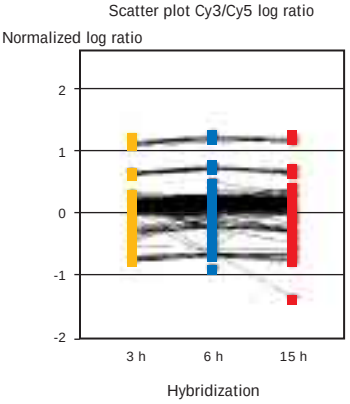


Fig 67. Effect of hybridization time. Cy3/Cy5 ratios of selected genes following 3-h, 6-h, and 15-h hybridization in Lucidea SlidePro.

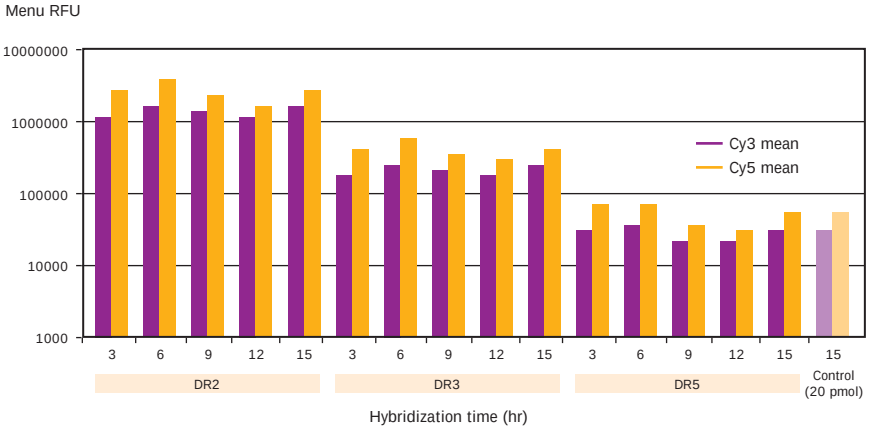


Fig 68. Effect of hybridization time. Total mean Cy3 and Cy5 signals of dynamic range controls.

9.4 Using different types of microarray slides

The dimensions of the slide chamber are specific and only standard microscope slides—25–25.5 mm × 75.5–76.0 mm, and 0.95–1.15 mm thick—may be used with Lucidea SlidePro. The hybridization chamber will fit arrays of up to 20 × 59 mm, allowing for a barcode on one end of the slide. The array should be spotted at least 2.5 mm from the edge. It is important that the bar code does not exceed 10 mm in width, or it will lie under the O-ring seal and cause leakage.

Lucidea SlidePro has been optimized for use with Amersham Biosciences version 2 hybridization buffer and microarray slides. Lucidea SlidePro can also be used to hybridize other manufacturer's slides (Fig 69). These protocols can be found on the Amersham Biosciences web site (www.amershambiosciences.com).

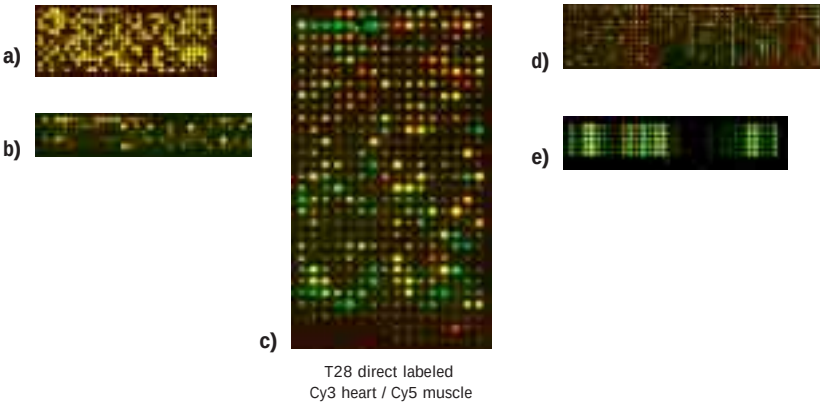


Fig 69. Lucidea SlidePro hybridization with various commercially available slides.

Chapter 10

FLUORESCENCE IMAGING SYSTEMS IN MICROARRAY ANALYSIS

10.0 Introduction

All fluorescence imaging systems require the following key elements (Fig 70):

- Excitation source
- Light delivery optics
- Light collection optics
- Filtration of the emitted light
- Detection, amplification and digitization of the emitted light

In this chapter, various types of scanner systems are discussed. Their light delivery and light collection mechanisms, signal detection and amplification, and overall performance are detailed, as well as criteria for selecting appropriate fluorochromes and filters for use with the scanner.

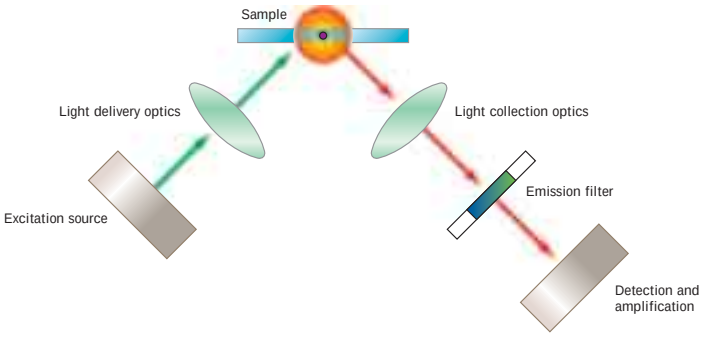


Fig 70. Components of a general fluorescence imaging system.

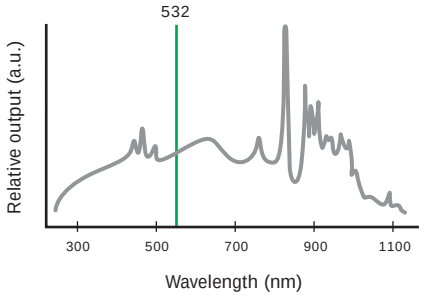


Fig 71. Spectral output of light from a xenon lamp and Nd:YAG laser. The "relative output" axis is scaled arbitrarily for the two light sources. The 532-nm line of the Nd:YAG laser is shown in green.

10.1 Requirements of a fluorescence imaging system

10.1.1 Excitation sources and light delivery optics

Light energy is essential to fluorescence. Light sources fall into two broad categories—wide-area, broad-wavelength sources, such as UV and xenon arc lamps, and line sources with discrete wavelengths, such as lasers (Fig 71). Broad-wavelength excitation sources are used in fluorescence spectrometers and camera imaging systems. Although the spectral output of a lamp is broad, it can be tuned to a narrow band of excitation light with the use of gratings or filters. In contrast, lasers deliver a narrow beam of collimated light that is predominantly monochromatic. In most camera systems, excitation light is delivered to the sample by direct illumination of the imaging field, with the excitation source positioned either above, below, or to the side of the sample.

Laser-based imaging systems, on the other hand, use more sophisticated optical paths, comprising mirrors and lenses, to direct the excitation beam to the sample. Some filtering of the laser light may also be required before the excitation beam is directed to the sample. For microarray applications, laser-based instruments are substantially favored, therefore CCD scanners will not be discussed.

10.1.2 Light collection optics

High-quality optical elements, such as lenses, mirrors, and filters, are integral components of any efficient imaging system. Optical filters are typically made from laminates of multiple glass elements. Filters can be coated to selectively absorb or reflect different wavelengths of light, thus creating the best combination of wavelength selection, linearity, and transmission properties.

10.1.3 Filtration of the emitted light

Although emitted fluorescent light radiates from a fluorochrome in all directions, it is typically collected from only a relatively small cone angle on one side of the sample. For this reason, light collection optics must be as efficient as possible. Any laser light that is reflected or scattered by the sample must be rejected from the collection pathway by a series of optical filters. Emitted light can also be filtered to select only the range or band of wavelengths that is of interest to the user. Systems that employ more than one detector require additional beam splitter filters to separate and direct the emitted light along separate paths to the individual detectors.

10.1.4 Detection, amplification, and digitization

For detection and quantification of emitted light, either a photomultiplier tube (PMT) or a charge-coupled device (CCD) can be used. In both cases, photon energy from emitted fluorescent light is converted into electrical energy, thereby producing a measurable signal that is proportional to the number of photons detected. After the emitted light is detected and amplified, the analog signal from a PMT or CCD detector is converted to a digital signal. The process of digitization turns a measured continuous analog signal into discrete numbers representing intensity levels. The number of intensity levels available is based on the digital resolution of the instrument, which is usually given as a number of bits, which increases exponentially by two. 8-bit, 12-bit, and 16-bit digital files correspond to the number of intensity levels allocated within that image file (256, 4096 and 65 536, respectively).

Digital resolution defines the ability to resolve two signals with similar intensities. Since only a limited number of intensity levels are available, it is unavoidable that this conversion process introduces a certain amount of error. To allow ample discrimination between similar signals and to keep the error as low as possible, the distribution of the available intensity levels should correspond well to the linear dynamic range of a detector. There are two methods of distributing intensity levels. A linear (even) distribution has the same spacing for all the intensity levels, allowing measurement across the dynamic range with the same absolute accuracy. However, relative digitization error increases as signals become smaller. A non-linear distribution (e.g. logarithmic or square root functions) divides the lower end of the signal range into more levels while combining the high end signals into fewer intensity levels. Thus, the absolute accuracy decreases with higher signals, but the relative digitization error remains more constant across the dynamic range.

10.2 Scanner systems

10.2.1 Excitation sources

Most fluorescence scanner devices used in life science research employ laser light for excitation. A laser source produces a narrow beam of highly monochromatic, coherent, and collimated light. The combination of focused energy and narrow beam-width contributes to the excellent sensitivity and resolution possible with a laser scanner. The active medium of a laser—the material that is made to emit light—is commonly a solid state (glass, crystal), liquid, or gas (57). Gas lasers and solid-state lasers both provide a wide range of specific wavelength choices for different imaging needs. Other light sources used in imaging systems include light emitting diodes (LEDs), which are more compact and less expensive than lasers, but produce a wide-band, low-power output.

Lasers

There are several commonly used types of lasers.

- Argon ion lasers produce a variety of wavelengths including 457 nm, 488 nm and 514 nm that are useful for excitation of many common fluorochromes, such as fluorescein and Cy2.
- Helium neon or HeNe lasers, which generate a single wavelength of light (633 nm), are popular in many laser scanners, and can be used to excite Cy5.
- Neodymium:Yttrium Aluminum Garnet (Nd:YAG) solid-state lasers, when frequency-doubled, generate a strong line at 532 nm which can be used to excite Cy3.
- Diode lasers (or semiconductor diode lasers) are compact lasers. Because of their small size and light weight, these light sources can be integrated directly into the scanning mechanism of a fluorescence imager. Diode lasers are inexpensive and are generally limited to wavelengths above 635 nm.

Light Emitting Diodes (LEDs)

As an alternative to lasers, the LED produces an output with a much wider bandwidth (over 60 nm) and a wide range of power from low to moderate output. Because LED light emissions are doughnut-shaped, and not collimated, the source must be mounted very close to the sample using lenses to tightly focus the light. LEDs are considerably smaller, lighter, and less expensive than lasers. They are available in the visible wavelength range above 430 nm.

10.3 Excitation light delivery

Because light from a laser is well-collimated and of sufficient power, delivery of excitation light to the sample is relatively straightforward, with only negligible losses incurred during the process. For lasers that produce multiple wavelengths of light, the desired line(s) can be selected by using filters that exclude unwanted wavelengths, while allowing the selected line to pass at a very high transmission percentage. Excitation filters are also necessary with single-line lasers, as their output is not 100% pure. Optical lenses are used to align the laser beam, and mirrors can be used to redirect the beam within the instrument. One of the main considerations in delivering light using a laser scanning system is that the light source is a point, while the sample typically occupies a relatively large two-dimensional space. Effective sample coverage can be achieved by rapidly moving the excitation beam across the sample in two dimensions. There are two ways to move and spread the point source across the sample, which are discussed below.

10.3.1 Galvanometer-based systems

Galvanometer-based systems use a small, rapidly oscillating mirror to deflect the laser beam, effectively creating a line source (Fig 72). By using relatively simple optics, the beam can be deflected very quickly, resulting in a short scan time. Compared to confocal systems, galvanometer-based scanners are useful for imaging thick samples due to the ability to collect more fluorescent signal in the vertical dimension. However, since the

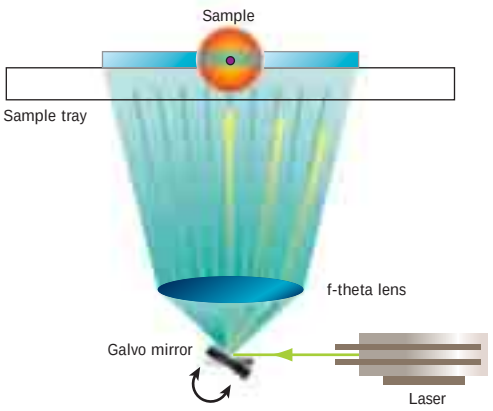


Fig 72. Galvanometer-controlled scanning mechanism. Light is emitted from the laser in a single, straight line. The galvanometer mirror moves rapidly back and forth redirecting the laser beam and illuminating the sample across its entire width (X-axis). The f-theta lens reduces the angle of the excitation beam delivered to the sample. The entire sample is illuminated either by the galvanometer mechanism moving along the length of the sample (Y-axis) or the sample moving relative to the scanning mechanism.

excitation beam does not illuminate the sample from the same angle in every position, a parallax effect can result. The term parallax here refers to the shift in apparent position of targets, predominately at the outer boundaries of the scan area. Additionally, the arc of excitation light created by the galvanometer mirror produces some variations in the effective excitation energy reaching the sample at different points across the arc. These effects can be minimized with an f-theta lens, but when the angle of incident excitation light varies over the imaging field, some spatial distortion can still occur in the resulting image.

10.3.2 Moving-head scanners

Moving-head scanners use an optical mechanism that is equidistant from the sample. This means that the angle and path length of the excitation beam is identical at any point on the sample (Fig 73). This eliminates variations in power density and spatial distortion common with galvanometer-based systems. Although scan times are longer with a moving-head design, the benefits of uniformity in both light delivery and collection of fluorescence are indispensable for accurate signal quantification. For microarray scanners an alternative method is to move the stage that contains the microarray slide. In some scanners the stage is moved in one direction while the galvanometer moves the laser beam across in the second dimension.

10.4 Light collection

The light collection optics in a scanner system must be designed to efficiently collect as much of the emitted fluorescent light as possible. Laser light that is reflected or scattered by the sample is generally

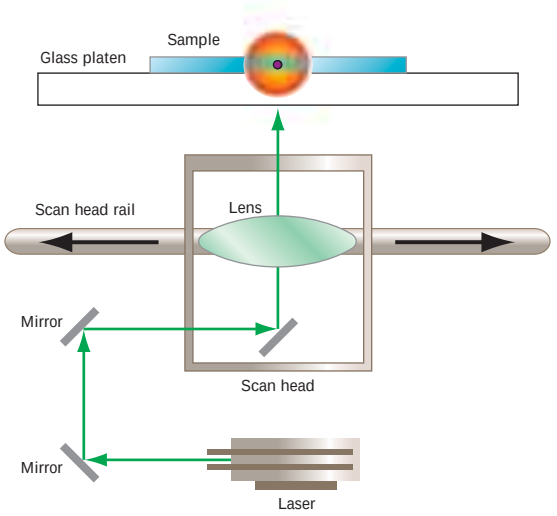


Fig 73. Moving-head scanning mechanism. The light beam from the laser is folded by a series of mirrors and ultimately reflected onto the sample. The sample is illuminated across its width as the scan head moves along the scan head rail (X-axis). The entire sample is illuminated by the scan head, laser, and mirrors tracking along the length of the sample (Y-axis).

rejected from the collection pathway by a laser-blocking filter, the design of which is to exclude the light produced by the laser source, while passing all other emitted light. Light collection schemes vary depending on the nature of the excitation system. With galvanometer systems, the emitted fluorescence must be gathered in a wide line across the sample. This is usually achieved with a linear lens (fiber bundle or light bar), positioned beneath the sample, that tracks with the excitation line, collecting fluorescence independently at each pixel. Although this system is effective, it can produce image artifacts. At the edges of the scan area where the angle of the excitation beam, relative to the sample, is farthest from perpendicular, some spatial distortion may occur. Where very high signal levels are present, stimulation of fluorescence from sample areas that are adjacent to the pixel under investigation can result in an inaccurate signal measurement from that pixel, an artifact known as flaring or blooming.

With moving-head systems, emitted light is collected directly below the point of sample excitation. Again, it is important to collect as much of the emitted light as possible to maintain high sensitivity. This can be achieved by using large collection lenses, or lenses with large numerical apertures (NA). Since the NA is directly related to the full angle of the cone of light rays that a lens can collect, the higher the NA, the greater the signal resolution and brightness (58). Moving-head designs can also include confocal optical elements that detect light from only a narrow vertical plane in the sample. This improves sensitivity by focusing and collecting emission light from the point of interest while reducing the background signal and noise from out-of-focus regions in the sample (Fig 74). Additionally, the parallel motion of moving head designs removes other artifacts associated with galvanometer-based systems, such as spatial distortion and the flaring or blooming associated with high activity samples.

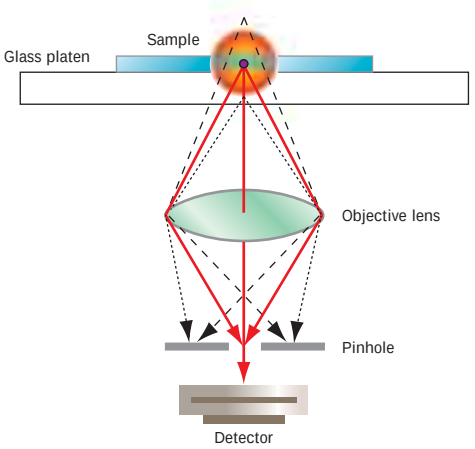


Fig 74. Illustration of confocal optics. Fluorescence from the sample is collected by an objective lens and directed toward a pinhole aperture. The pinhole allows the emitted light from a narrow focal plane (red solid lines) to pass to the detector, while blocking most of the out-of-focus light (black dashed lines).

10.5 Signal detection and amplification

The first stage in fluorescent signal detection is selection of only the desired emission wavelengths from the label or dye. In single-channel or single-label experiments, emission filters are designed to allow only a well-defined spectrum of emitted light to reach the detector. Any remaining stray excitation or scattered light is rejected. Because the intensity of the laser light is many orders of magnitude greater than the emitted light, even a small fraction of laser light reaching the detector will significantly increase background. Filtration is also used to reduce background fluorescence or inherent autofluorescence originating from either the sample itself or the sample matrix gel, membrane, or microplate. In multichannel or multi-label experiments using instrumentation with dual detectors, additional filtering is required upstream of the previously described emission filter. During the initial stage of collection in these experiments, fluorescence from two different labels within the same sample is collected simultaneously as a mixed signal. A dichroic beam splitter must be included to spectrally resolve the contribution from each label and then direct the light to appropriate emission filters (Fig 75). At a specified wavelength, the beam splitter partitions the incident fluorescent light beam into two beams, passing one and reflecting the other. The reflected light creates a second channel that is filtered independently and detected by a separate detector. In this way, the fluorescent signal from each label is determined accurately in both spatial and quantitative terms.

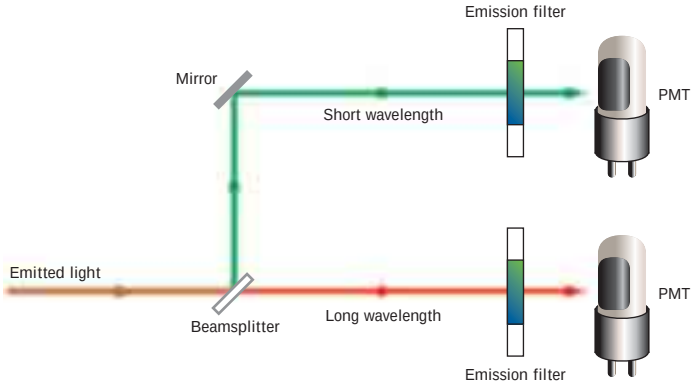


Fig 75. Use of a beam splitter or dichroic filter with two separate PMTs. Light from a dual color sample enters the emission optics as a combination of wavelengths. A dichroic beam splitter distinguishes light on the basis of wavelength. Wavelengths above the beam splitter range pass through, those below are reflected. In this way two channels are created. These two channels can then be filtered and detected independently.

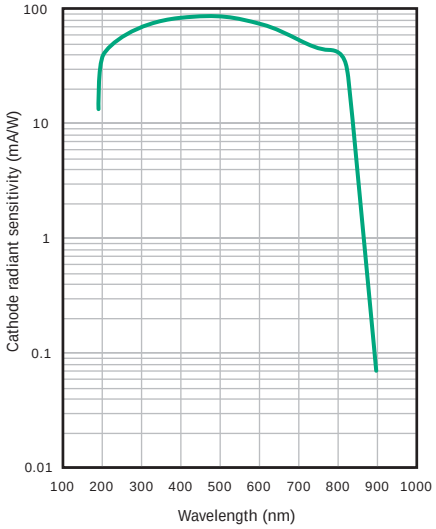


Fig 76. An example of the response of a PMT versus wavelength.

After the fluorescent emission has been filtered and only the desired wavelengths remain, the light is detected and quantified. Because the intensity of light at this stage is very small, a PMT must be used to detect it. In the PMT, photons of light hit a photocathode and are converted into electrons which are then accelerated in a voltage gradient and multiplied from 10^6 to 10^7 times. This produces a measurable electrical signal that is proportional to the number of photons detected. The response of a PMT is typically useful over a wavelength range of 300–800 nm (Fig 76). High-performance PMTs extend this range to 200–900 nm.

10.6 System performance

The performance of a laser scanner system is described in terms of system resolution, linearity, uniformity, and sensitivity.

Resolution can be defined in terms of both spatial and amplitude resolution. Spatial resolution of an instrument refers to its ability to distinguish between two very closely positioned objects. It is a function of the diameter of the light beam when it reaches the sample and the distance between adjacent measurements. Spatial resolution is dependent on, but not equivalent to, the pixel size of the image. Spatial resolution improves as pixel size is reduced. Systems with higher spatial resolution can not only detect smaller objects, but can also discriminate more accurately between closely spaced targets. However, an image with a 100-µm pixel size will not have a spatial resolution of 100 µm. The pixel size refers to the collection sampling interval of the image. According to a fundamental sampling principle, the Nyquist Criterion, the smallest resolvable object in an image is no better than twice the sampling interval (59). Thus, to resolve a 100-µm sample, the sampling interval must be at most 50 µm. Amplitude resolution, or gray-level quantification, describes the minimum difference that is distinguishable between levels of light intensity (or fluorescence) detected from the sample (60). For example, an imaging system with 16-bit digitization can resolve and accurately quantify 65 536 different values of light intensity from a fluorescent sample.

Linearity of a laser scanner is the signal range over which the instrument yields a linear response to fluorochrome concentration and is therefore useful for accurate quantification. The linear dynamic range can be defined in at least 3 ways:

1) the electronic dynamic range of the scanner

2) the chemical dynamic range of the fluorescent dyes used

3) the biological dynamic range of the system under study

A scanning system with a wide dynamic range can detect and accurately quantify signals from both very low- and very high-intensity targets in the same scan. The linear dynamic range of most laser scanner instruments is between 10^4 and 10^5.

Uniformity across the entire scan area is critical for reliable quantitation. A given fluorescent signal should yield the same measurement at any position within the imaging field. Moving-head scanners, in particular, deliver flat-field illumination and uniform collection of fluorescent emissions across the entire scan area.

Detection limit is the minimum amount of sample that can be detected by an instrument at a known confidence level. From an economical standpoint, instruments with better detection limits are more cost-effective because they require less fluorescent sample for analysis.

10.7 Fluorochrome and Filter Selection

To generate fluorescence, excitation light delivered to the sample must be within the absorption spectrum of the fluorochrome. Generally, the closer the excitation wavelength is to the peak absorption wavelength of the fluorochrome, the greater the excitation efficiency. Appropriate filters are usually built into scanner instruments for laser line selection and elimination of unwanted background light. Fixed or interchangeable optical filters that are suitable for the emission profile of the fluorochromes are then used to refine the emitted fluorescence, such that only the desired wavelengths are passed to the detector. Matching a fluorochrome label with a suitable excitation source and emission filter is the key to optimal detection efficiency.

10.7.1 Types of emission filters

The composition of emission filters used in fluorescence scanners and cameras ranges from simple colored glass to glass laminates coated with thin interference films. Coated interference filters generally deliver excellent performance through their selective reflection and transmission effects. Three types of optical emission filters are commonly used.

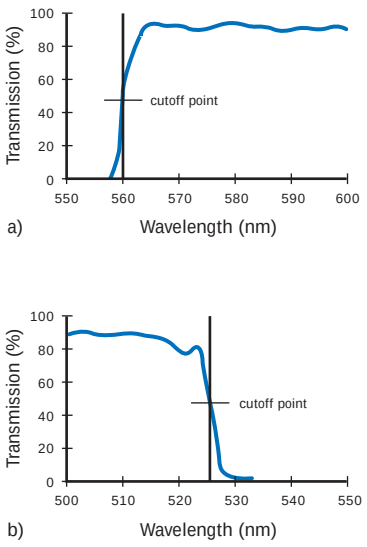


Fig 77. Transmission profiles for a (a) 560-nm long-pass and a (b) 526-nm short-pass filter. The cutoff points are noted.

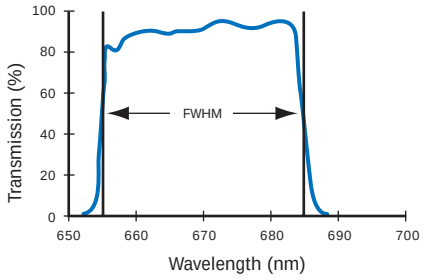


Fig 78. Transmission profile for a band-pass (670 BP 30) filter. The full-width at half-maximum (FWHM) transmission of 30 nm is indicated by the arrows.

Long-pass (LP) filters pass light that is longer than a specified wavelength and reject all shorter wavelengths. A good quality long-pass filter is characterized by a steep transition between rejected and transmitted wavelengths (Fig 77a). Long-pass filters are named for the wavelength at the midpoint of the transition between the rejected and transmitted light (cutoff point). For example, the cutoff point in the transmission spectrum of a 560 LP filter is 560 nm, where 50% of the maximum transmittance is rejected. The name of a long-pass filter may also include other designations, such as OG (orange glass), RG (red glass), E (emission), LP (long-pass), or EFLP (edge filter long-pass). OG and RG are colored glass absorption filters, whereas E, LP, and EFLP filters are coated interference filters. Colored glass filters are less expensive and have more gradual transition slopes than coated interference filters.

Short-pass (SP) filters reject wavelengths that are longer than a specified value and pass shorter wavelengths. Like long-pass filters, short-pass filters are named according to their cutoff point. For example, a 526 SP filter rejects 50% of the maximum transmittance at 526 nm (Fig 77b).

Band-pass (BP) filters allow a band of selected wavelengths to pass through, while rejecting all shorter and longer wavelengths. Band-pass filters provides very sharp cutoffs with very little transmission of the rejected wavelengths. High-performance band-pass filters are also referred to as Discriminating Filters (DF). The name of a band-pass filter is typically made up of two parts:

- the wavelength of the band center (the 670 BP 30 filter passes a band of light centered at 670 nm [Fig 78]);
- the full-width at half-maximum transmission (FWHM) (the 670 BP 30 filter passes light over a wavelength range of 30 nm [655–685 nm] with an efficiency equal to or greater than half the maximum transmittance of the filter).

Band-pass filters with an FWHM of 20–30 nm are optimal for most fluorescence applications, including multi-label experiments. Filters with FWHMs greater than 30 nm allow collection of light at more wavelengths and give a higher total signal; however, they are less able to discriminate between closely spaced, overlapping emission spectra in multichannel experiments. Filters with FWHMs narrower than 20 nm transmit less signal and are most useful with fluorochromes with very narrow emission spectra.

10.8 Using emission filters to improve sensitivity and linearity range

When selectable emission filters are available in an imaging system, filter choice will influence the sensitivity and dynamic range of an assay. In general, if image background signal is high, adding an interchangeable filter may improve the sensitivity and dynamic range of the assay. The background signal from some matrices (gels and membranes) has a broad, relatively flat spectrum. In such cases, a band-pass filter can remove the portion of the background signal comprising wavelengths that are longer or shorter than the fluorochrome emissions. By selecting a filter that transmits a band at or near the emission peak of the fluorochrome of interest, the background signal is typically reduced with only slight attenuation of the signal from the fluorochrome. Therefore, the use of an appropriate band-pass filter should improve the overall signal-to-noise ratio (S/N).

To determine if a filter is needed, scans should be performed with and without the filter while other conditions remain constant. The resulting S/N values should then be compared to determine the more efficient configuration. Interchangeable filters can also be used in fluorescence scanners to attenuate the sample signal itself so that it falls within the linear range of the system. Although scanning the sample at a reduced PMT voltage can attenuate the signal, the response of the PMT may not be linear if the voltage is set below the instrument manufacturer's recommendation. If further attenuation is necessary to prevent saturation of the PMT, the addition of an appropriate emission filter can decrease the signal reaching the detector.

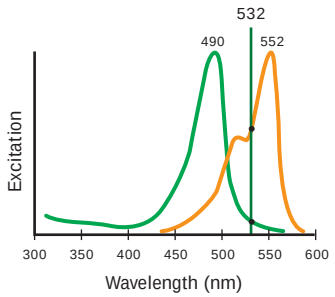


Fig 79. Excitation of fluorescein (green) and Cy3 (orange) using 532-nm laser light. The absorption spectra of Cy3 and fluorescein are overlaid with the 532-nm wavelength line of the Nd:YAG laser.

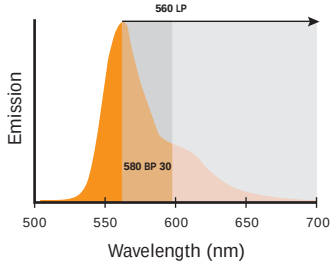


Fig 80. Filtering of Cy3 fluorescence using either a 580 BP 30 (dark gray area) or a 560 LP filter (light and dark gray areas).

10.9 General guidelines for selecting fluorochromes and filters

10.9.1 Single-color imaging

Excitation efficiency is usually highest when the fluorochrome's absorption maximum correlates closely with the excitation wavelength of the imaging system. However, the absorption profiles of most fluorochromes are rather broad, and some fluorochromes have a second (or additional) absorption peak or a long "tail" in their spectra. It is not mandatory that the fluorochrome's major absorption peak matches exactly the available excitation wavelength for efficient excitation. For example, the absorption maxima of the fluorescein and Cy3 fluorochromes are 490 nm and 552 nm respectively (Fig 79). Excitation of either dye using the 532-nm wavelength line of the Nd:YAG laser may seem to be inefficient, since the laser produces light that is 40 nm above the absorption peak of fluorescein and 20 nm below that of Cy3. In practice, however, delivery of a high level of excitation energy at 532 nm does efficiently excite both fluorochromes. For emission, selecting a filter that transmits a band at or near the emission peak of the fluorochrome generally improves the sensitivity and linear range of the measurement. Figure 80 shows collection of Cy3 fluorescence using either a 580 BP 30 or a 560 LP emission filter.

10.9.2 Multicolor imaging

Multicolor imaging allows detection and resolution of multiple targets using fluorescent labels with different spectral properties. The ability to multiplex or detect multiple labels in the same experiment is both time and cost-effective and improves accuracy for some assays. Analysis using a single label can require a set of experiments or many repetitions of the same experiment to generate one set of data. For example, single-label analysis of gene expression from two different tissues requires two separate hybridization to different gene arrays, or consecutive hybridization to the same array with stripping and re-probing. With a dual-label approach, however, the DNA probes from the two tissue types are labelled with different fluorochromes and used simultaneously with the same gene array. In this way, experimental error is reduced because only one array is used, and hybridization conditions for the two probes are identical. Additionally, by using a two-channel scan, expression data is rapidly collected from both tissues, thus streamlining analysis.

The process for multicolor image acquisition varies depending on the imaging system. An imager with a single detector acquires consecutive images using different emission filters and, in some cases, different excitation light. When two detectors are available, the combined or mixed fluorescence from two different labels is collected at the same time and then resolved by filtering before the signal reaches the detectors. Implementation of dual detection requires a beam splitter filter to spectrally split the mixed fluorescent signal, directing the resulting two emission beams to separate emission filters (optimal for each fluorochrome), and finally to the detectors. A beam splitter, or dichroic reflector, is specified to function as either a short-pass or long-pass filter relative to the desired transition wavelength. For example, a beam splitter that reflects light shorter than the transition wavelength and passes longer wavelengths is effectively acting as a long-pass filter (Fig 75).

10.9.3 Fluorochrome selection in multicolor experiments

When designing multicolor experiments, two key elements must be considered: the fluorochromes used and the emission filters available. As with any fluorescence experiment, the excitation wavelength of the scanner must fall within the absorption spectrum of the fluorochromes used. Additionally, the emission spectra of different fluorochromes selected for an experiment should be relatively well resolved from each other. However, some spectral overlap between emission profiles is almost unavoidable. To minimize cross-contamination, fluorochromes with well-separated emission peaks should be chosen along with emission filters that allow reasonable spectral discrimination between the fluorochrome emission profiles. Figure 86 shows the emission overlap between two common fluorochromes and the use of band-pass filters to discriminate the spectra. For best results, fluorochromes with emission peaks at least 30 nm apart should be chosen. A fluorescence scanner is most useful for multicolor experiments when it provides selectable emission filters suitable for a variety of labels. A range of narrow band-pass filters that match the peak emission wavelengths of commonly used fluorochrome labels will address most multicolor imaging needs.

Fig 81. Typhoon Variable Mode Imager.

10.10 Amersham Biosciences imaging systems

Amersham Biosciences offers a variety of imaging instruments that are well suited for use in microarray analysis. For more information, please consult Amersham Biosciences web site at www.amershambiosciences.com.

10.10.1 Typhoon 9210: High performance laser scanning system

Excitation sources: 532-nm Nd:YAG and 633-nm HeNe lasers

Filters: 6 emission filters and 3 beamsplitters (up to 13 emission filter positions)

Detection: 2 high sensitivity PMTs

Imaging modes: 4 modes. 2 modes for fluorescence detection, chemiluminescence, storage phosphor

Scanning area: 35 × 43 cm

Maximum resolution: 10 µm

Sample types: microarrays, gel sandwiches, agarose and polyacrylamide gels, blots, microplates, TLC plates, and macroarrays

10.10.2 Typhoon 9410: High performance laser scanning system

Excitation sources: 532-nm Nd:YAG, 633-nm HeNe, and 457-nm and 488-nm Argon lasers

Filters: 7 emission filters and 3 beamsplitters (up to 13 emission filter positions)

Detection: 2 high sensitivity PMTs

Imaging modes: 5 modes. 3 modes for fluorescence detection, chemiluminescence, storage phosphor, chemifluorescence

Scanning area: 35 × 43 cm

Maximum resolution: 10 µm

Sample types: microarrays, gel sandwiches, agarose and polyacrylamide gels, blots, microplates, TLC plates, and macroarrays

Notes: Versatile fluorescence and radioactive imager that can scan microarrays but also contains an extra blue laser

Chapter 11

DESIGN, CONTROLS AND DATA ANALYSIS OF MICROARRAY EXPERIMENTS

11.0 Introduction

Methods for the analysis of microarray data are still evolving, and there is no standard experimental design or method of data analysis for microarray experiments at the present time. However, some efforts are being made to set a common annotation and standards for microarray data in order to create public databases for microarray results (61, 62, 63). Meanwhile, in this chapter some important considerations for analyzing microarray data are discussed.

11.1 Experimental design

Data analysis begins with experimental design. When planning a microarray experiment it is important to consider sources of variation within the experiment. These can arise from the samples reflecting differences in gene expression between individual animals or different tissue culture plates. Furthermore, time-dependent variation in gene expression levels resulting from circadian rhythms can also be a factor. Experimental variations may also occur due to variation within the experiment itself. In order to ensure that these experimental errors can be identified, slides should contain replicate spots of each target and replicate slides should be analyzed with pooled or multiple mRNA samples. This replication enables the use of statistical tools such as averages and standard deviations to monitor the extent of experimental variation (64).

Lucidea Universal ScoreCard has been developed by Amersham Biosciences to address the need for controls. It is a set of controls used to validate and normalize microarray experimental data. It is further described in section 6 of this chapter.

Another prudent measure is to perform reverse color or ‘flip-flop’ experiments. In these experiments the two mRNA samples being compared in a microarray experiment are labelled separately with both Cy3 and Cy5. Replica slides are hybridized with both combinations of probes. By comparing the signal ratios from the reversed slides, it is possible to identify data that is affected more by the labelling process or quality of mRNA than by changes in gene expression levels.

Once the mRNA extraction, labelling, hybridization, and scanning are complete, the final stage in the microarray experiment is data analysis. This is a complex multi-step process and is illustrated in Figure 82. The steps of data analysis are described in further detail in this chapter.

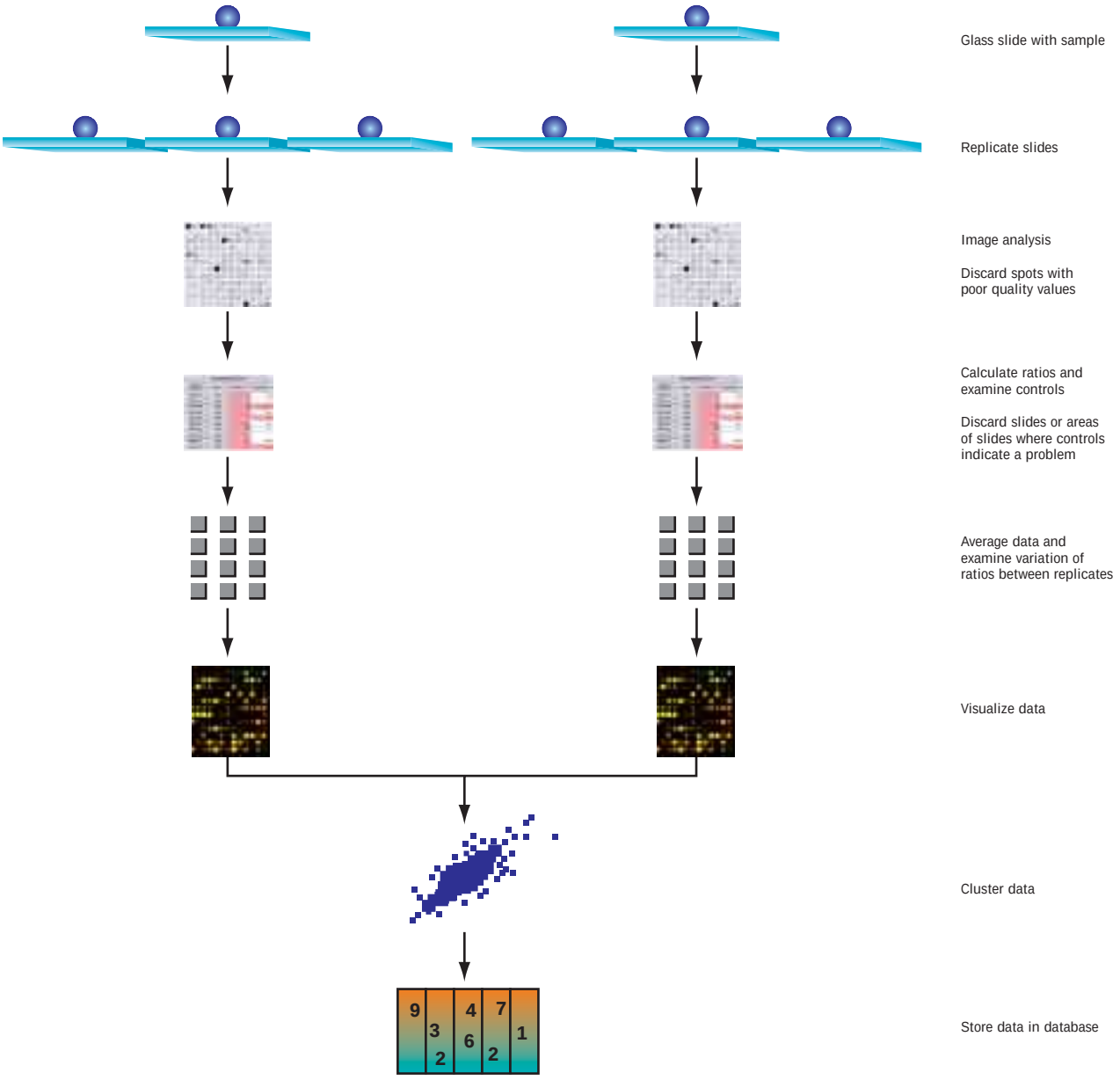


Fig 82. Stages of microarray data analysis.

11.2 Overview of microarray data analysis

Microarray data analysis consists of four main steps:

- Image analysis
- Examination of controls
- Data normalization
- Visualization and clustering

Image analysis uses a dedicated software to quantitate the fluorescent intensity at each spot. Normally, this involves a process called spotfinding. The second step is to examine the controls on the arrays. Normalization is performed next, followed by calculation of mean ratios. The data can then be visualized in a graphical form, and clustered, such that meaningful trends can be found among data from multiple slides and experiments.

The amount of data obtained from microarray experiments is vast and can be generated very rapidly. However, it is important to know the quality of the data. There are three types of quality values that can be used. It is recommended that all three are used within a microarray experiment. The types of values are:

- A series of metrics reported by the image analysis software to ensure that the spots that have been quantitated appear to be good spots, for example, regularly shaped.
- A series of controls on your microarray to ensure the hybridization has occurred correctly. These will indicate how specific and efficient the hybridization has been.
- Analysis of the data from replicate targets. Replicates are critical for indicating how good the overall data is and whether the results obtained are statistically meaningful (64).

11.3 Image analysis

A scanned microarray image records the fluorescent intensities of all pixels in the image area, including pixels from within (signal) and outside (background) the DNA spots. The first step in the microarray workflow is to locate the spots. Consider the following objectives:

- accurately define the positions of every DNA spot in the image
- provide appropriate measurement of fluorescence intensity for each spot by quantifying the intensities of pixels within and outside the DNA spots
- provide quality metrics that give estimation of the accuracy of the intensity measurement

The first step, alternately called “gridding”, can usually be performed using dedicated software, such as ArrayVision or Lucidea Automated Spotfinder.

The process of spotfinding begins by defining a grid, or an array of circles, that indicates the expected size of each spot, how far away they are spaced, and how they are arranged in an array, all regardless of the intensities of individual spots. This information can be measured in pixel or micron units. Once a grid is defined, it is overlaid onto the scanned image such that the circles are nearly exactly aligned with the spots on the microarray image. This spotfinding process can be automated using spotfinding software, which serves to eliminate the tedious task of manual alignment. In addition to the following descriptions, see the spotfinding software help guide for instructions how to most effectively use this tool.

- **Manual**: This method involves first dividing the grid into several subgrids, and then visually aligning each smaller subgrid with the corresponding area of the image by adjusting the position of circles.
- **Semi-automated**: This is where the software algorithm finds the spots, but some user intervention is required.
- **Automated**: This is where minimal user intervention is required. These software packages can automatically analyze multiple images while eliminating the need for supervision.

The next step in image analysis is to determine the signal present inside and outside the spot. The background signal is then subtracted from the spot signal to give background-corrected signal. Whereas spot signal is calculated from within the positioned grid, the background signal is determined by calculating the average pixel intensity in a user-defined region. Although mean or median background signal can be used, median values are more resistant to variation in background caused, for example, by fluorescent speckles. Some of the various regions in which the background can be calculated are illustrated in Figure 83. It is important to use a background correction method such that any pixels from the spot do not get included in the background. This can easily occur if the spots are close together.

Fig 83. Some of the background region options available to the user in the ArrayVision™ Image Analysis Software. The green represents the spots enclosed by the grid while the blue encloses the background region. Image analysis quality metrics calculated by analysis software are increasingly used to highlight data that may be unreliable and should be omitted from further analysis. Typical causes of suspect data include dust speckles over spots, poor spot morphology, very low or very high signal.

Fig 84. Lucidea Automated Spotfinder has a simple, intuitive user interface to initiate the automatic processing of microarray images. In this example, four images are loaded for analysis.

Fig 85. ArrayVision software provides automated analysis of radioisotopic or fluorescent macro- and microarrays.

11.4 Spotfinding software offered by Amersham Biosciences

11.4.1 Lucidea Automated Spotfinder

Lucidea Automated Spotfinder processes microarray images by performing spotfinding and data extraction in an automated fashion with virtually no manual intervention (Fig 84). The output from Lucidea Automated Spotfinder includes the signal intensity for each spot, plus quality metrics to assess individual spots as well as the overall image. The software is compatible with images produced by commercially available scanners. Several images can be analyzed at once in a batch mode, without manual inspection or image manipulation. Lucidea Automated Spotfinder features include:

- fully automated spot finding and data extraction
- multiple reporting options and data export
- user-defined templates for analyzing single or multiple images
- metrics for assessing data quality
- background subtraction

11.4.2 ArrayVision

ArrayVision software is a semi-automated software used for performing image analysis (Fig 85). Some of its features include:

- automated alignment of quantification grid over array
- choice of methods for background signal removal
- quality metrics
- reporting tools and data export
- visualization tools for viewing array images

11.5 Use of controls in microarray experiments

As with all experiments, microarrays should contain a series of controls to ensure that the data obtained from the arrays is accurate. Therefore, included on microarrays should be some cDNA or oligos which are expected to give a negative result, and some which should give a positive result. The types of controls that should be included are discussed below.

11.5.1 Negative controls

Negative controls are spotted DNA sequences that should not hybridize with any labelled probe. The negative controls used should ideally come from organisms that are only distantly related to the organism being studied in the experiment. For example with human microarrays, sequences from bacterial genes, intergenic regions, plant genes, or double-stranded poly-dA are often used for this purpose (25, 65, 66). Under optimal analysis conditions negative control spots should not give any signal at all. However, if the stringency of the hybridization is not high enough, non-specific hybridization between labelled cDNA molecules in the probe and unrelated target sequences on the array may take place, resulting in detectable signals from negative control spots. The higher the signal from negative controls, the less reliable is the data from the whole slide. These negative controls are particularly important to include if the spotted array consists of oligonucleotides because in these types of arrays, a lower hybridization stringency may be used. Negative controls can also be used to detect contamination between targets during spotting. Placing negative control targets after positive control targets that are always expected to give signal can do this.

11.5.2 Poly-adenylated DNA and CotI DNA

When using oligo(dT) to prime first-strand cDNA synthesis, it is possible that the oligo(dT) will prime within the poly-A tail of the mRNA. If this occurs there will be a string of dT bases within the probe. It is possible that the targets spotted may also contain a similar string of poly-A sequences, particularly if the targets were derived originally from an EST library which had been made by the use of oligo(dT) primer. In order to prevent cross reactivity of the poly-dT sequences within the probe with potential poly-dA sequences in the targets, a poly-dA oligo of 80 bases can be included in the hybridization to block the poly-dT (65). In order to ensure that this process has occurred correctly, it is good to include as a negative control, a poly-dA sequence spotted on the microarray. If these spots are negative this suggests that the blocking has occurred effectively.

Another sequence that may cause problems within the probe is derived from repetitive sequences such as Alu-repeat sequences. These sequences can be blocked by the inclusion of Cot-1 DNA in the hybridization. To ensure that this blocking has occurred correctly, the spotting of Cot-1 DNA as a negative control is recommended.

11.5.3 Positive controls

Labelled DNA

DNA can be labelled with CyDye fluors using polymerase chain reaction (PCR), or in the case of oligos, during the synthesis of the oligos (many oligo manufacturers offer this service). When the labelled DNA is spotted onto the array, the DNA will be fluorescent and serve as a useful positive control for verifying that the target DNA is binding effectively to the slide surface during the hybridization and washes. Total genomic DNA can also be used as a positive target. Positive controls placed on different locations of the slide can help in the spotfinding process by providing clearly detectable signals in known positions, regardless of the type of probe used.

Fig 86. To validate, normalize, and filter microarray data, four different types of controls are supplied:

1. Calibration controls: the signal intensities of ten individual controls span 4.5 orders of magnitude for both Cy3 and Cy5 channels. These controls can be used to generate a calibration curve.
2. Ratio controls: eight ratio controls are provided at both low and high expression levels and are used to evaluate precision of ratios.
3. Negative controls: two controls are used to estimate non-specific hybridization and potential carryover with the microarray system.
4. Utility controls: three individual controls can be used to troubleshoot and examine sample preparations, or they can be used as additional ratio or calibration controls.

Spikes for determining sensitivity and dynamic range

Using some of the negative control genes discussed above can be used to make a different set of positive controls. RNA can be synthesized by *in vitro* transcription from plasmids carrying these negative control sequences. This synthetic mRNA can then be included in the Cy3 and Cy5 labelling reactions at known concentrations. This process is known as spiking, and the synthetic mRNA are the spikes. Several different mRNA spikes can be used and spiked in at different concentrations, resulting in a set of controls that can give the researcher a value for the linear dynamic range and sensitivity of the assay. These controls are known as dynamic range controls. In addition, different spikes can be spiked into the Cy3 and Cy5 labelling reactions at different concentrations. These types of controls are known as ratio controls. After hybridization of the probe to the slide, washing, and scanning, the Cy3 and Cy5 signals obtained from the ratio controls can be compared with the known amount of mRNA spiked into the labelling reactions and the theoretical known ratios. Therefore a series of spikes can determine how sensitive the hybridization has been and how accurate the data obtained is.

11.5.4 Housekeeping gene controls

Some genes are expressed relatively consistently within many different cell types. These are called housekeeping genes. Examples of such genes are actin, glyceraldehyde 3-phosphate dehydrogenase (GAPDH), and tubulin. These housekeeping genes can be included in microarray experiments as controls to ensure that the hybridization has occurred, and they can also act as a normalization factor. However, the expression of these genes can vary under experimental conditions, and relying on the use of one or few housekeeping genes can result in skewed data.

11.5.5 Controls for measuring pen-to-pen variation

Housekeeping gene, spikes, or positive controls can be spotted in replicate across the slide. If a different pen on the microarray spotter spots each of these replicates, this may provide some information on the pen-to-pen variation of the spotter. It should be considered that the variation of each of these controls will be dependent not only on pen-to-pen variation but also on the variation in the slide surface and any variation in the hybridization. Therefore, if a high pen-to-pen variation is seen, a pen test on the spotter should be performed.

11.6 Control products offered by Amersham Biosciences

11.6.1 Lucidea Universal Scorecard

Lucidea Universal ScoreCard contains a set of 23 artificial genes that serve as analytical controls to validate and normalize microarray data. The controls are composed of DNA sequences from yeast intergenic regions, and their performance has been shown to be independent of a wide variety of species. This system can be used as a universal reference for validating and normalizing microarray data as well as for creating a calibration curve for determining limit of detection, linear range, and saturation of microarray experiment (Fig 86).

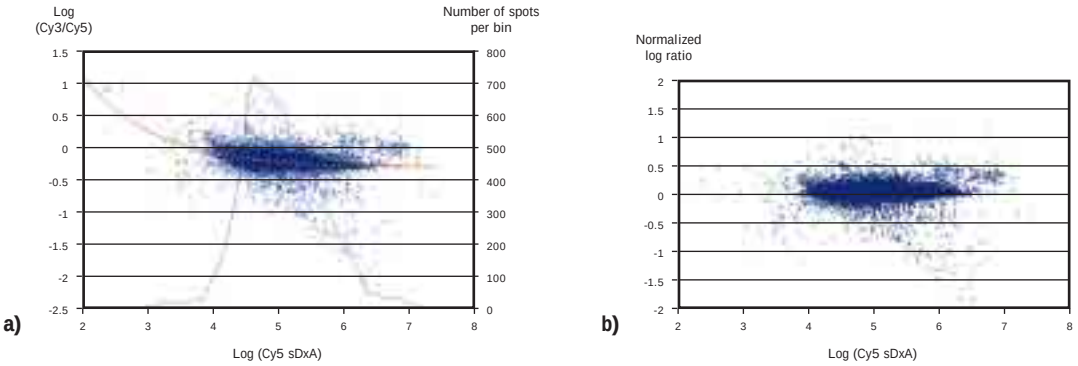


Fig 87. Applying non-linear normalization to microarray data. Panel A shows distribution of log(Cy3/Cy5) values plotted against Cy5 signal values (blue crosses) from a typical microarray experiment. The orange line denotes the average relationship between these log ratios as a function of Cy5 signals. As can be seen from the shape of the curve, this relationship is non-linear over the distribution of Cy5 signals. The blue line shows distribution of Cy5 signal intensities. Panel B shows the same data after it has been normalized using the non-linear normalization algorithm.

11.7 Normalization

In order to compare ratio data from one microarray slide to another microarray slide, the ratio data needs to be normalized to correct for experimental variation. The reason for this is that from one slide to another there will be differences between the relative Cy3 and Cy5 signals due to one or more of the following:

- the amounts of mRNA used in the Cy3 and Cy5 labelling reaction
- efficiency of detection of the Cy3 and Cy5 by the detection system within the scanner
- relative incorporation differences of the Cy3 and Cy5 reverse transcriptases

11.7.1 Linear normalization

Linear normalization assumes that there is a single normalization factor required over the whole signal range. There are three methods to obtain this factor:

- use signal ratios from housekeeping genes
- use signal ratios from spikes which have been added to the labelling reaction in equal amounts
- use total signal, which is the summation of all the Cy3 signals and all the Cy5 signals

In the housekeeping gene or spike gene methods, it is assumed that the spots corresponding to these targets give a ratio of 1. The total signal method assumes that addition of all the signals in the Cy3 and the Cy5 channel should give a Cy3/Cy5 ratio of 1. The normalization factor can be calculated from the observed ratios for the housekeeping genes, spikes, or total signals to give a conversion factor that results in the expected ratio for the control spots.

For example, a housekeeping gene in experiment A gives a ratio of 2, while the gene of interest has a Cy3/Cy5 ratio of 5. Therefore the normalized ratio for the gene of interest is 5/2 = 2.5. In experiment C, the housekeeping gene has a ratio of 3, while the gene of interest has a Cy3/Cy5 ratio of 7.5. The normalized ratio is 7.5/3 = 2.5.

The three methods discussed above assume that the normalization factor is constant over the whole signal range, which in most cases is not.

11.7.2 Non-linear normalization

It has been found that linear normalization is not necessarily accurate (67). Examining the data shown in Figure 87a can show this. In this experiment, Cy3-labelled muscle and Cy5-labelled muscle probes (identical mRNA labelled with two different fluors) were hybridized to a single slide. The results of this experiment are plotted below as a plot of log Cy3/Cy5 ratio against the log Cy5 signal (Fig 87b). As the same mRNA was used, it would be expected that the log Cy3/Cy5 ratio should be constant over all Cy5 signals. However, as can be seen from the graph, the ratio is higher at low Cy5 signal levels compared to the figure at high Cy5. Therefore, the normalization factor used for spots with a low Cy5 signal should be different from the normalization factor that is used at high Cy5 signals.

There are two principal non-linear methods of calculating the normalization factor. One method is to rank all the data points according to their Cy3+Cy5 signal. Then for each 50 genes calculate the normalization factor for those 50, in the same way as total normalization is carried out. A more precise way is to fit a curve to the data so that the normalization factor for each point can be calculated. Software packages are commercially available that can perform this kind of normalization. A non-linear normalization generally results in a more accurate normalization than linear normalization.

11.7.3 Post normalization

Once the normalization procedure has been carried out for all the data points, the behavior of controls is examined next. The negative controls should have a signal-to-noise ratio of less than 3 [(SNR = [average signal – average background]/standard deviation of background). Any higher SNR than 3 suggests that the data obtained from this experiment may not be accurate.

If a series of dynamic range controls have been included, this is one way to estimate sensitivity. A control which has a SNR above 3 would be regarded as having been detected. If spikes have been included, such that the spikes have been placed in the Cy3 reaction at a different amount compared to the Cy5 reaction, then the theoretical ratio can be compared to the actual ratio. Finally, controls such as pen-to-pen variation controls may suggest other potential problems within the experiment.

If the data from the slide meets the criteria set by the researcher then the next step is to look at the variation between replicates. It is recommended that each experiment be repeated several times, and there are statistical

criteria as to how many replicates should be carried out to give a certain degree of confidence in your results (64). Replicates could be within multiple spots on the same slide, but can also be carried out using several different slides. Typically log ratios are calculated so that ratios less than 1 appear as a negative value. The coefficient of variation (CV = [Standard deviation of the ratio] / [mean ratio]) can also be calculated and provides a simple measure of the value of a particular data point. Data points with high CVs can be highlighted or discarded. Often ratios with high CVs are due to low signal in one or both of the channels. A standard practice by some researchers is therefore to discard spots which have a signal in both channels with a SNR below 3. If there is a low signal (below a SNR of 3) in one channel, then the ratio could be considered to be an arbitrary fixed value to avoid very large ratios (to prevent, for example if Cy5 signal is zero, the Cy3/Cy5 would be infinite). In addition, it should be remembered that there may be significant biological variation that must be taken into account when designing experiments.

11.8 Visualization and clustering

After microarray data is normalized to account for differences in Cy3 and Cy5 signals, it can subsequently be exported to any number of data visualization software for further analysis. These software products can be used to mine the data for significant changes in gene expression. The process of visualization can significantly enhance data analysis. It can provide helpful features, such as data integration, customized query devices, and pattern recognition. Clustering data points, or genes, that show similar responses on microarray analysis can be used to identify genes that have similar gene expression patterns and which possibly belong to the same pathway.

Fig 88. Exploring gene expression data using hierarchical clustering, and principal component analysis with Spotfire DecisionSite for Functional Genomics.

11.9 Visualization software products offered by Amersham Biosciences

11.9.1 Spotfire DecisionSite for Functional Genomics

This software combines the core capabilities of Spotfire™ DecisionSite™ with specialized tools for interrogating and extracting information from microarray data. In addition to simplified access to data and information, Spotfire DecisionSite for Functional Genomics provides researchers with leading analytical methods used in gene expression analysis. Dynamic visualizations and interaction with computational results help researchers in validating and prioritizing target genes (Fig 88).

Some of the features of this software include:

- easy access to information, in any format, wherever it resides
- visually interactive representations of enriched data sets for enhanced analysis
- publication and sharing of results for collaborative decision-making
- idenification of key patterns with distinction calculation, hierarchical, bi-directional hierarchical, and K-means cluster analysis
- preparation of data for analysis with standard array and gene based normalization
- identification of entities exhibiting characteristic or signature profiles with ad hoc profile search and analysis

11.10 Scierra Microarray Laboratory Workflow Systems: Information Management Systems for Microarray Laboratories

Microarray technology is rapidly becoming the mainstream platform for high-throughput gene expression analysis. As microarray experiments generate vast amounts of experimental and biological data, an urgent need is created for informatics tools that can manage the microarray workflow process more effectively and efficiently. Scierra™ Microarray Laboratory Workflow System, as a part of a larger integrated laboratory information managment system, is designed to address this need.

11.10.1 Scierra Laboratory Workflow Systems

Scierra Laboratory Workflow Systems (LWS) are a series of bioinformatics solutions to aid in the collection, annotation, collation, curation, and analysis of biological data. The platform includes four products:

- Scierra Sequencing LWS System
- Scierra Genotyping LWS System
- Scierra Microarray LWS System
- Scierra Proteomics LWS System

These Scierra LWS products are built on a common software framework that manages and tracks all aspects of an experiment. Each system mirrors the natural workflow found within the laboratory, and succeeds in linking manual processes, instrumentation, software, and reagent use into one system. This integration enables the collection and comparison of each type of biological information. The open framework design allows the introduction of new instruments and reagents, thereby providing a scalable system that will grow as needs increase.

11.10.2 Scierra LWS architecture

Scierra LWS is a three-tiered application comprised of an Oracle™ database, a middleware application server, and multiple clients. Scierra LWS can accept data from most available network client sources, including computers running Windows™ 2000 or Windows NT™. Work is easily requested and results readily accessed through a standard browser-based user interface that communicates to the middleware application server.

11.10.3 Scierra Microarray Laboratory Workflow System (MA-LWS)

A typical microarray workflow involves array content preparation, array production, sample preparation and labelling, array hybridization, scanning, and image analysis.

Scierra MA-LWS has been designed to mirror the workflow of the typical microarray lab (Fig 89). It allows users to perform many tasks. First, users can organize large numbers of samples and experiments based on projects. Users can also manage and track a large variety of samples and reagents, including:

- crude biological samples, including blood, tissue, and cells
- total RNA, mRNA, and labelled cDNA
- spotting plates and spotting sample types, including cDNA and oligo
- spotting substrates, chemistry, and protocol
- arrayed slides

Scierra MA-LWS allows users to manage and track every activity in the microarray workflow, including:

- array preparation—spotting custom arrays and pre-arrayed slides
- sample preparation and labelling
- hybridization, scanning, and image analysis

Scierra MA-LWS makes it possible to integrate components of different microarray systems, including spotters and scanners, and image analysis software. Users can store and effectively retrieve large amounts of information. Flexible reporting tools provide for standard and user-defined queries across different activities. With Scierra MA-LWS, the precious microarray data is completely captured and securely stored in the database for further analysis.

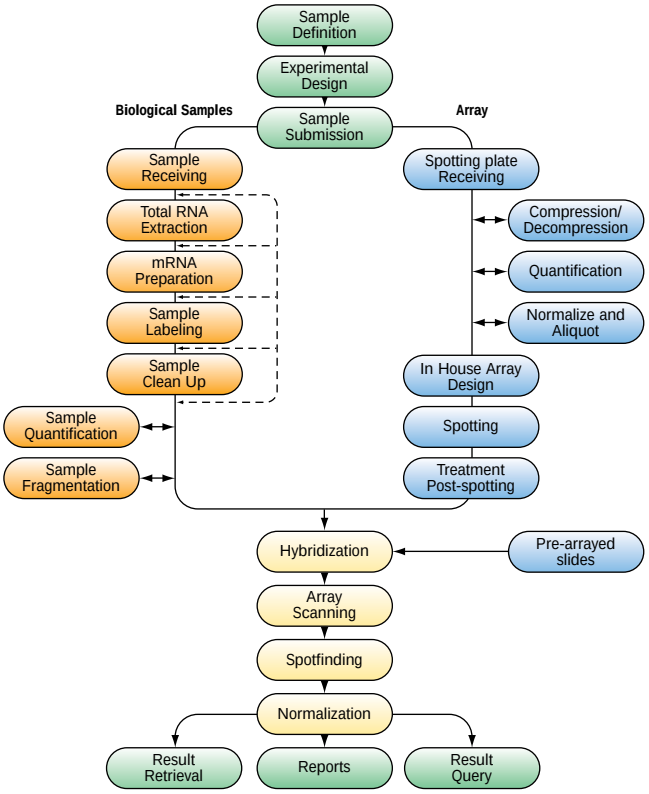


Fig 89. Scierra MA-LWS mirrors the microarray workflow.

Chapter 12

TROUBLESHOOTING MICROARRAY EXPERIMENTS

12.0 Introduction

Microarray analysis is a complicated multistep process consisting of discrete steps as shown in Figure 90. Each of these steps is critical in determining whether the experiment is successful, and problems encountered at any stage will be detrimental to the quality of data obtained. Troubleshooting microarray experiments needs to consider all these steps.

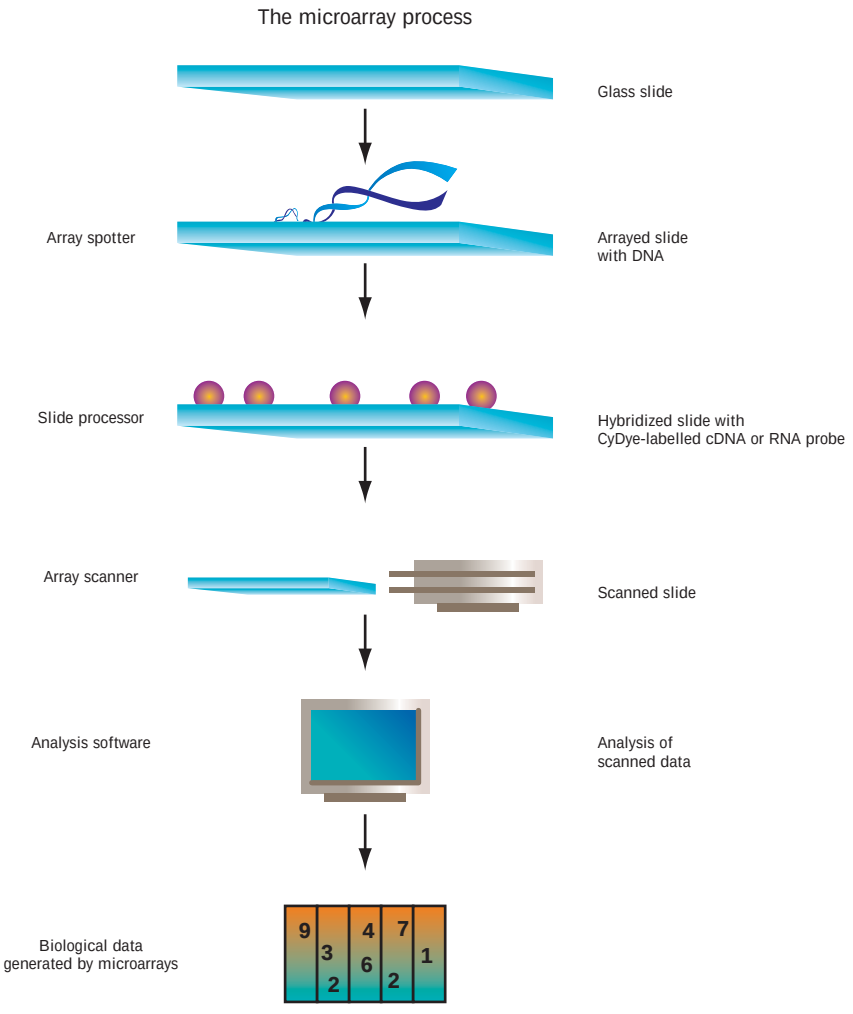


Fig 90. Flowchart of microarray experiment.

The output from a microarray experiment is the intensity of hybridization signals, which reflect the expression levels of the corresponding genes in the analyzed samples. However, other experimental factors also have significant influence on the magnitude of these signals. Some of these factors are listed in Figure 91.

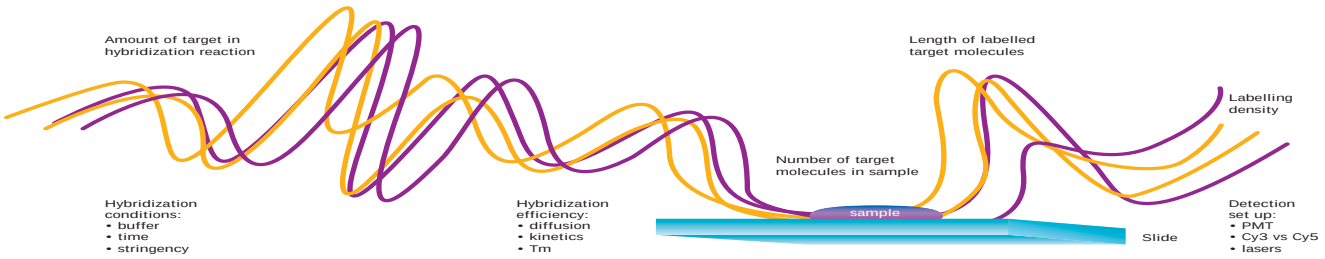


Fig 91. Factors influencing the intensity of observed microarray signals.

12.1 Optimization of the microarray system

As microarray analysis is not trivial to perform, careful optimization of the microarray system is recommended. Due to the wide choice of reagents, consumables and instruments from various manufacturers, it is important to optimize the selected combination of materials and protocols before starting real experiments. Some reagents may not work well with other reagents; for example, spotting buffer may not be compatible with all different slide surfaces, and different hybridization buffers may give very different results on identical slides.

System optimization should test all microarray components together using target sequences and probes that are as close to real samples as possible. Such experiments should incorporate controls, as discussed in Chapter 11. Combining the use of realistic sets of targets and control reagents, as illustrated in Figure 92, gives the best results. The aim of

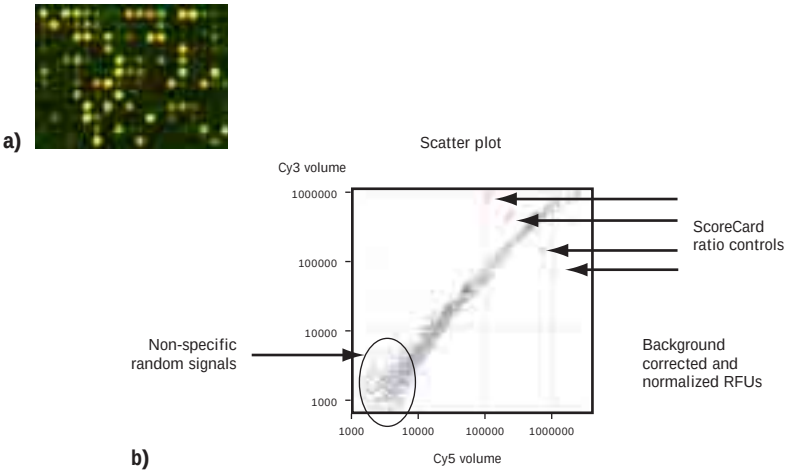


Fig 92. Panel A shows a scanned microarray image from a yellow experiment in which skeletal muscle mRNA was labelled with Cy3 and Cy5. Only a proportion of the microarray slide is shown. Panel B shows a scatterplot derived from a yellow experiment. The microarray used in this experiment contained ratio control targets in addition to cDNA clones and mRNA corresponding to these control sequences was spiked into the mRNA before labelling. These ratio control spots appear as green and red dots on the array image and are also shown on the scatterplot, where they fall above and below the scatter line.

system optimization is to find the best overall protocol for the system and to determine what is the standard performance of the system.

12.1.1 The yellow experiment

The yellow experiment is an efficient tool for microarray system optimization. In this experiment, the same RNA sample is labelled with both fluorescent dyes, typically with Cy3 and Cy5 fluors. Hybridization of equal amounts of both probes onto a microarray should produce equal hybridization signals from both colors. In a false color array image, all target spots should appear as yellow dots. As computer screen images of microarray data can be misleading, analyzing numerical data from a yellow experiment as a scatter plot is more informative. As no differential gene expression is expected, normalized Cy3 signals plotted against normalized Cy5 signals should appear as a straight line. Figure 92 shows an example of microarray data generated from a typical yellow experiment. Because RNA isolated from different cell cultures or individuals is likely to contain slightly different levels of some transcripts, it is recommended that pooled RNA obtained from several RNA isolations is used for system optimization. Alternatively, purified RNA is also available commercially.

Yellow experiments only require one type of RNA sample, but provide information on all aspects of the microarray process. In contrast, experiments in which two different RNA species are used have the added complication that differential gene expression will be present in unknown quantities. In reverse-color experiments, in which both samples are labelled separately with two colors and hybridizations are performed with both possible combinations, any inherent variation between the quality of the two arrays can complicate the optimization process and result in wrong conclusions. For example, high and uneven Cy3 background on one of the slides in a reverse-color experiment can give the appearance of unbalanced labelling with one fluorescent dye. For ‘real’ experiments in which information about gene expression is being sought, reverse color experiments are useful.

12.2 Experimental design and execution

12.2.1 Experimental variation

Experimental variation must be taken into account when designing and carrying out microarray experiments (Fig 93). No two microarray experiments, even if replicas of each other, will give exactly the same results. Each step of the process contributes to this variation, which may mask the presence of differential gene expression, leading to false negative results. If the amount of variation is not known, false positive results can also be obtained if randomly varying results are taken at face value.

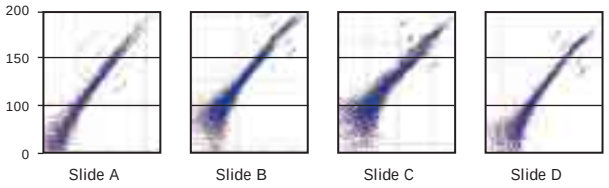


Fig 93. Experimental variation. Four identical microarray slides were hybridized simultaneously with equal aliquots of the same probe. Scatterplots of normalized gene expression data are shown. Variation in background fluorescence arising from uneven slide surfaces was the major contributing factor to experimental variation in this case.

12.2.2 Replication in microarray analysis

Replication is the key to identifying and quantifying variation in microarray experiments. It has been found that performing three replica microarray hybridizations with different slides reduced the misclassification of gene expression compared with performing single hybridizations (64). Data calculated and pooled from all these replicas enables statistical determination of experimental variation in terms of standard deviation and coefficient of variance (CV).

Microarray experiments should contain the following:

- Each target should be present in at least two, preferably more, copies on the microarray.
- Multiple slides should be hybridized with each probe pair.
- Multiple RNA samples should be obtained for each experimental condition.

12.2.3 Step controls

Separately monitoring the various steps in microarray analysis, while lengthening the protocols, provides quality control information. Step controls also offer the following benefits:

- Researchers' resources are used most efficiently, with the least amount of waste.
- Problems can be identified on a step-by-step basis, and most likely causes obtained.
- Information from intermediate steps allows conclusions about the overall validity of microarray results to be drawn.

Ideal step controls provide numerical or visual information that unequivocally characterizes the success of that step. Recommended control procedures are listed in Table 6. **The control measures indicated in bold should be included in every microarray experiment**. Additional controls are recommended to be performed when new protocols or reagents are being tested. It is advisable to prepare some extra reagents if control procedures are performed. Greatest benefits from step controls are derived when the results are evaluated before continuing with the experiment.

12.3 Performing microarray analysis

There are several critical points to performing a successful microarray hybridization experiment. Please take note of the following:

- Follow all protocols precisely.
- Be consistent across experiments when several separate hybridizations are involved.
- Maintain precise technique when performing the microarray experiment and analyzing the results.
- Always use appropriate reagents for each protocol; alterations can be a source of error in final data analysis.
- Keep record of individual reagents used in each experiment as it can be useful in identifying causes of problems.

Table 7 lists some typical problems in microarray analysis and their likely causes. Most of these problems can be identified and avoided by following the recommended quality control measures listed in Table 7, which can also aid in identifying the cause of the problem. Often there can be several contributing causes to any problem.

Table 6. Quality control measures for microarray analysis.

Step	Control mearsures
Preparation of cDNA targets by PCR	■ **Verify presence of only one band in agarose electrophoresis.** ■ **Determine quantity of target DNA.** ■ Sequence target to verify identity.
Microarray slide coating	■ Scan to detect presence of background fluorescence or dirt particles. ■ Perform additional surface tests.
Microarray printing	■ Use fluorescent DNA as control for printing quality. ■ Use DNA-binding dyes to detect printing DNA on slide. ■ **Perform a test yellow experiment with well-characterized RNA preparation for each printing batch and especially when new targets are introduced.** ■ **Incude plenty of control targets on slide.** ■ **Note down temperature and humidity of printing chamber.** ■ **Keep track of how many times targets have been used.** ■ **Keep track of when slides were printed.**
RNA isolation	■ **Check purity and integrity of RNA by gel electrophoresis or RT-PCR.** ■ **Determine quantity of RNA.**
Sample labelling and purification	■ **Spike in synthetic control mRNA.** ■ Perform control labelling reaction with control RNA. ■ **Determine incorporation of CyDye into labelled sample by spectrophotometry.** ■ **Determine whether purified sample contains free CyDye by gel electrophoresis.** ■ Determine size of labelled nucleic acid fragments with gel electrophoresis. ■ Determine the amount of labelled nucleic acid. ■ Determine the amount of CyDye per microgram of labelled nucleic acid (labelling density). ■ Determine the recovery of labelled cDNA from purification step by radioactive spiking or gel analysis.
Hybridization and stringency washes	■ **Use equal and optimal amounts of Cy3- and Cy5-labelled probes in the hybridization.** ■ **Include positive and negative hybridization controls.** ■ Perform hybridization without probe.
Scanning	■ Test scanner performance in order to verify correct functioning. ■ **Perform scans at different settings to ensure optimal data collection.** ■ **Visually inspect scanned images to detect any obvious blemishes or areas of poor data.**
Data analysis	■ **Background correct and normalize data before drawing conclusions.** ■ **Examine how well normalization worked.** ■ **Determine the amount of variation in experiment.** ■ **Never trust data from one slide.**
Verification of microarray results	■ **Use independent analytical techniques to verify whether the results obtained from microarray analysis are reproducible and biologically significant.**

Table 7. Troubleshooting microarray experiments,

Symptom	Possible cause	Remedy
No hybridization signal.	■ Target concentration too low.	■ Determine target concentration before slide spotting.
	■ Targets not clean enough.	■ Remove PCR components from targets before slide spotting.
	■ Poor retention of targets on slide.	■ Prepare new microarray slides. Check that spotting buffer and protocol are compatible with slide type.
	■ No transcripts in RNA sample.	■ Obtain new RNA/mRNA sample and test it before labelling.
	■ Failed labelling reaction.	■ Always check the success of labelling reaction before using it in hybridization.
	■ Faulty component in labelling reaction.	■ Test components of labelling reaction against new reagents. Use control RNA.
	■ Loss of probe in purification step.	■ Check success of probe purification before use.
	■ Poor hybridization.	■ Check that hybridization buffer and protocol are compatible with slide type.
	■ Failure of scanning instrument.	■ Test performance of scanner with known amounts of fluors.
	■ Detection sensitivity too low.	■ Adjust detection sensitivity.
	■ CyDye have been exposed to light during handling.	■ Protect CyDye from light.
	■ Target genes not expressed in examined tissue.	■ Use housekeeping genes and positive controls to ascertain proper functioning of the system.
	■ Human error at some stage.	■ Repeat experiment and use step controls to monitor progress.
Low or undetectable Cy3 and/or Cy5 signal.	■ Poor retention of targets on slide. Identical probes were hybridized with two different types of slide that contained the same targets (Fig 94n, 94o).	■ Check purity and concentration of targets. Use slides of different batch. Use different slide type.
	■ Targets are old.	■ Prepare new targets for spotting.
	■ Poor labelling reaction with one dye. One or more components faulty in labelling reaction.	■ Check success of labelling reaction. Check performance of all components of labelling reaction such as nucleotides, enzyme and fluorescent nucleotides or reactive dyes.
	■ Loss of probe in purification.	■ Check performance of purification. Do not purify Cy3 and Cy5 probes together.
	■ Unequal amount of Cy3 and Cy5.	■ Use equal amounts of probes in hybridization.
	■ Too little probe in hybridization.	■ Measure the amount of probe before hybridization. Use more RNA to prepare probe.
	■ Free CyDye in probe.	■ Optimize probe purification.
	■ Poor quality RNA sample or samples.	■ Test RNA before labelling.
	■ Too much or too little quantity of RNA.	■ Measure amount of RNA before labelling.
	■ RNA contaminated by DNA.	■ Use DNAse I to remove DNA.

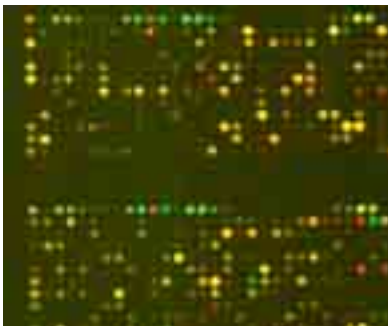
Table 7 cont'd. Troubleshooting microarray experiments.

Symptom	Possible cause	Remedy
	■ Hybridization conditions not optimal.	■ Check compatibility of hybridization buffer and slide surface. Use lower hybridization stringency.
	■ High background in hybridization.	■ Use slides of a different batch. Optimize hybridization and wash protocol.
	■ Overexposure of Cy3 and Cy5 to light during storage and handling.	■ Protect CyDye from light always.
	■ Laser source not working optimally.	■ Check laser performance.
	■ Detection sensitivity not optimal.	■ Optimize laser power and detection sensitivity settings.
	■ Detection sensitivity not optimal.	■ Optimize laser power and detection sensitivity settings.
	■ Human error.	■ Repeat experiment with step controls.
Unbalanced Cy3 and Cy5 signals.	■ Too high labelling density leading to quenching of one fluorescent dye.	■ Label to a lower density.
	■ Too much CyDye nucleotide in labelling reaction.	■ Use less CyDye nucleotide.
	■ Too much of one probe in hybridization leading to quenching.	■ Optimize the amount of probe in hybridization. Measure the amount of probe in hybridization.
	■ Some nucleotide sequences label poorly.	■ Use a different labelled nucleotide. Use a different labelling method.
	■ Poor or variable quality of RNA sample/samples. Biological variation in RNA samples.	■ Use good quality RNA for labelling. Use pooled RNA samples.
	■ High fluorescent background in . one color.	■ See separate entry.
	■ Normalization method is not adequate.	■ Use a different normalization method.
	■ High amount of variation in experiment.	■ Optimize and standardize experimental conditions to reduce amount of variation.
High background, weak specific signals.	■ Poor labelling reaction.	■ Check success of labelling reaction before hybridization.
	■ Random nonamers used for labelling total RNA.	■ Prepare probe with oligo(dT) primer.
	■ Probe fragments very short.	■ Use good quality RNA. Check purity of RNA. Re-purify RNA.
Uneven fluorescent background on slide.	■ Poor slide quality with an uneven coating. Often background is higher on one side of slide (Fig 94a).	■ Use slides of a different batch. Optimize slide surface treatment protocol. Use different source of microscope slides.
	■ Fluorescent background from pre-hybridization or hybridization solution (Fig 94f).	■ Optimize wash protocol. Include water dip at the end. Dry slides quickly.

Table 7 cont'd. Troubleshooting microarray experiments.

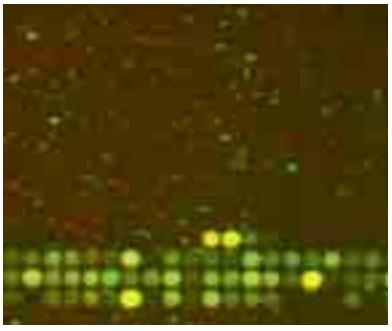
Symptom	Possible cause	Remedy
	■ Salts in wash buffer dried onto dried onto slide (Fig 94b).	■ Dip slide into water before drying.
	■ Powder from lab gloves adheres to slide.	■ Handle slides using powder-free gloves.
	■ Edge of coverslip has dried during manual hybridization (Fig 94m).	■ Perform hybridization under humid conditions.
Most spots give high uniform signal.	■ High amount of unspecific signals.	■ Increase stringency of hybridization and washes.
Even fluorescent background.	■ Poor slide quality (Fig 94g, 94i).	■ Use slides from different batch. Use different source of microscope slides.
	■ Too much probe used in hybridization.	■ Quantify probe before use.
	■ Over labelling of sample.	■ Optimize amount of probe to use. Optimize labelling density.
Speckled background on slide.	■ CyDye nucleotides remain in probe (Fig 94d).	■ Optimize purification of probe. Check probe for presence of free CyDye.
Particles seen on slide.	■ Dust particles have been fixed onto slide.	■ Always handle slides in clean environment. Use air stream to remove any dust particles from dry slides before spotting and use.
	■ Slide surface is scratched (Fig 94p).	■ Handle slides with care using forceps.
	■ Finger prints seen on slide.	■ Never touch slides with bare hands.
Bubble effect on slide.	■ Air has been trapped under coverslip (Fig 94c).	■ Remove air bubbles from hybridization.
Spots appear as comets with tails.	■ Hybridized probe is coming loose during low stringency wash/water dip (Fig 94j).	■ Optimize wash conditions. Dry slides quickly after washes and water dip.
Deformed spots.	■ Doughnut spots (Fig 94l).	■ Control humidity of spotting process.
	■ Tiny spots.(Fig 94e).	■ Test printing pen performance.
	■ Variably sized spots (Fig 94k).	■ Wrong spotting buffer for slide chemistry.
	■ Negative spots caused by slide background which is higher than hybridization signals (Fig 94h).	■ Use slides of a different batch.

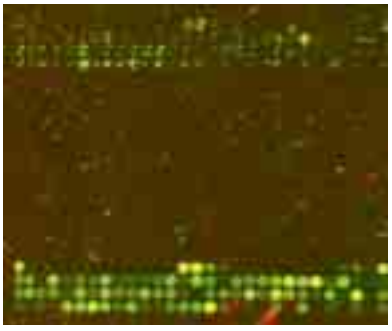
Fig 94. Troubleshooting microarray experiments.

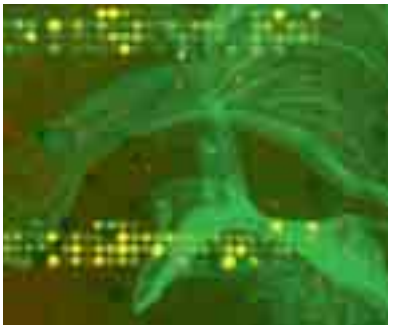
a) 

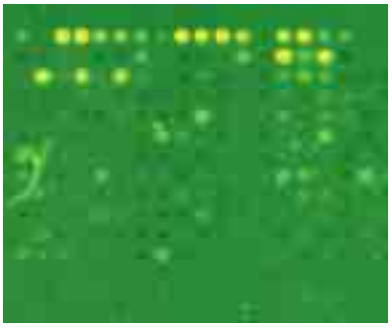
b) 

c) 

d) 

e) 

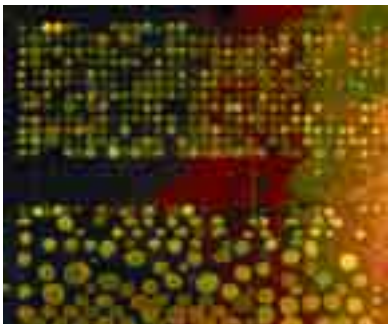
f) 

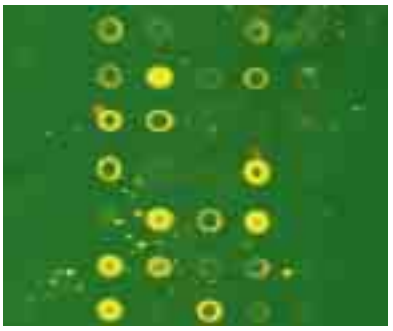
g) 

h) 

i) 

j) 

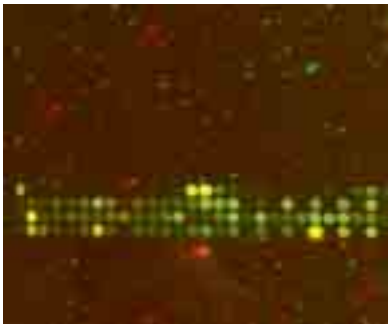
k) 

l) 

m) 

n) 

o) 

p) 

References

References cited in text

1. Baldwin, D. *et al*. A comparison of gel-based, nylon filter and microarray techniques to detect differential RNA expression in plants. *Curr Opinions in Biol.* **2**, 96–103 (1999).

2. Watson, A. *et al*. Technology for microarray analysis of gene expression. *Current Opinions in Biotech*. **9** 609–614 (1998).

3. Schena, M. *et al*. Microarrays: Biotechnology's discovery platform for functional genomics. *Trends in Biotech*. **16**, 301–306 (1998).

4. Kozian, D. H. and Kirschbaum, B. J. Comparative gene expression analysis. *Trends in Biotech*. **17**, 73–78 (1999).

5. Braxton, S. and Bedilion, T. The integration of microarray information in the drug development process. *Current Opinions in Biotech*. **9**, 643–649 (1998).

6. Mirnics, K. *et al*. Analysis of complex brain disorders with gene expression microarrays: Schizophrenia as a disease of the synapse. *Trends in Neuroscience* **24**, 479–486 (2001).

7. Schulze, A. and Downward, J. Navigating gene expression using microarrays — a technology review. *Nature Cell Biology* **3**, e190–e195 (2001).

8. Van Berkum, N. L. and Holstege, F. C. P. DNA microarrays: raising the profile. *Current Opinions in Biotech*. **12**, 48–52 (2001).

9. Alizadeh, A. A. *et al*. Towards a novel classification of human malignancies based on gene expression patterns. *J Pathol*. **195(1)**, 41–52 (2001).

10. DeRisi, J. *et al*. Use of cDNA microarray to analyze gene expression patterns in human cancer. *Nature Genetics* **14**, 457–460 (1996).

11. Rew, D. A. DNA microarray technology in cancer research. *European Journal of Surgical Oncology* **27**, 504–508 (2001).

12. Shoemaker, D. D. *et al*. Experimental annotation of the human genome using microarray technology. *Nature* **409**, 922–927 (2001).

13. Lieb, J. D. *et al*. Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nature Genetics* **28**, 327–334 (2001).

14. Hu, G. K. *et al*. Predicting splice variant from DNA chip expression data. *Genome Research* **11**, 1237–1245 (2001).

15. Meltzer, P. S. Spotting the target: microarrays for disease gene discovery. *Current Opin in Genetics and Dev*. **11**, 258–263 (2001).

16. Sapolsky, R. J. *et al*. High-throughput polymorphism screening and genotyping with high-density oligonucleotide arrays. *Genet Anal*. **14(5-6)**, 187–92 (1999).

17. Larsen, L. A. *et al*. Recent developments in high-throughput mutation screening. *Pharmacogenomics* **2(4)**, 38799 (2001).

18. Drobyshev, A. *et al*. Sequence analysis by hybridization with oligonucleotide microchip: identification of beta-thalassemia mutations. *Gene* **188(1)**, 45–52 (1997).

19. Lockhardt, D. J. and Winzeler, E. A. Genomics, gene expression and DNA arrays. *Nature* **405**, 827–836 (2000).

20. Jain, K. K. Applications of biochip and microarray systems in pharmacogenomics. *Pharmacogenomics* **1**, 289–307 (2000).

21. Gray, N. S. *et al*. Exploiting chemical libraries, structure and genomics in the search for kinase inhibitors. *Science* **281**, 533–538 (1998).

22. Kane, M. D. *et al*. Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res*. **28**, 4552–4557 (2000).

23. Li, F. and Stormo, G. D. Selection of optimal DNA oligos for gene expression arrays. *Bioinformatics* **17**, 1067–1076 (2001).

24. Lockhart, D. J. *et al*. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotech*. **14(13)**, 1675–1680 (1996).

25. Schena, M. *et al*. Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes. *Proc. Natl. Acad. Sci.* **93**, 10614–10619 (1996).

26. Schepinov, M. S. *et al*. Steric factors influencing hybridization of nucleic acid to oligonucleotide arrays. *Nucl Acids Res*. **25**, 1155–1161 (1997).

27. Schena, M. *et al*. Microarrays: biotechnology's discovery platform for functional genomics. *Trends Biotechnol*. **17**, 217–218 (1997).

28. Welford, S. M. *et al*. Detection of differentially expressed genes in primary tumor tissues using representational differences analysis coupled to microarray hybridization. *Nucleic Acids Res*. **26**, 3059–65 (1998).

29. Penn, S. G. *et al*. Mining the human genome using microarrays of open reading frames. *Nat Genet 2000* **26**, 315–318 (2000).

30. Hedge, P. *et al*. A concise guide to cDNA microarray analysis. *BioTechniques* **29**, 548–562 (2000).

31. Knight, J. When the chips are down. *Nature* **410**, 860–861 (2001).

32. Taylor, E. *et al*. Sequence verification as quality control. *Biotechniques* **31**, 62–65 (2001).

33. Schuchhardt, J. *et al*. Normalization strategies for cDNA microarrays. *Nucl. Acids Res*. **28**, e47 (2000).

34. Yamanaka, Y. *et al*. Gene expression profiles of human small airway epithelial cells treated with low doses of 14- and 16-membered macrolides. *Biochem. Biophys. Res. Comm*. **287**, 198–203 (2001).

35. Lipschutz, R. J. *et al*. High density synthetic oligonucleotide arrays. *Nature genetics* **21(supplement)**, 20–24 (1999).

36. Zammatteo, N. *et al*. Comparison between different strategies of covalent attachment of DNA to glass surfaces to build DNA microarrays. *Anal Biochem*. **280(1)**, 143–50 (2000).

37. Worley, J. *et al*. *Microarray Biochip Technology* (Schena, M., ed.), Eaton Publishing/BioTechniques Books, Natick, MA, pp. 65–85 (2000).

38. DeRisi, J. L. *et al*. Exploring the metabolic and genetic control of gene expression on genomic scale. *Science* **278**, 680–686 (1997).

39. Fodor, S. P. *et al*. Light-directed, spatially addressable parallel chemical synthesis. *Science* **251**, 767–73 (1991).

40. Mujumdar, R. B. *et al*. Cyanine dye labeling reagents: sulfoindocyanine succinimidyl esters. *Bioconjug Chem*. **4(2)**, 105–11 (1993).

41. Yu, H. *et al*. Cyanine dye dUTP analogs for enzymatic labeling of DNA probes. *Nucleic Acids Res*. **22**, 3226–32 (1994).

42. Sambrook, J. *et al*. *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989).

43. Ausubel, F. M., *Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley-Interscience, New York (2000).

44. Gruffat, D. *et al*. Comparison of four methods for isolating large mRNA: Apolipoprotein B mRNA in bovine and rat livers. *Anal Biochem*. **249**, 77–83 (1996).

45. Rosenow, C. *et al*. Prokaryotic RNA preparation methods useful for high density array analysis: comparison of two approaches. *Nucl. Acid Res*. **29**, e112 (2001).

46. (Farrell, R. E., ed.), *RNA Methodologies, A laboratory guide for Isolation and Characterization*, Academic Press, Inc., New York, pp. 125–157 (1997).

47. Randolp, J. B. and Waggoner, A. S. Stability, specificity and fluorescence brightness of multiply-labeled fluorescent DNA probes. *Nucl Acid Res*. **25**, 2923–2929 (1997).

48. Smoot L. M. *et al*. Global differential gene expression in response to growth temperature alteration in group A Streptococcus. *Proc. Natl. Acad. Sci*. **98**, 10416–421 (2001).

49. Talaat A. M. *et al*. Genome-directed primers for selective labeling of bacterial transcripts for DNA microarray analysis. *Nature Biotech*. **18**, 679–682 (2000).

50. Van Gelder, R. N. *et al*. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci*. **87**, 1663–1667 (1990).

51. Eberwine, J. H. *et al*. Analysis of gene expression in single live neurons. *Proc Natl Acad Sci*. **89**, 3010–3014 (1992).

52. Luo, L. *et al*. Gene expression profiles of laser-captured adjacent neuronal subtypes. *Nature Medicine* **5**, 117–122 (1999).

53. Herrler, M. Use of SMART-generated cDNA for differential gene expression. *J Molecular Medicine* **78**, B23 (2000).

54. Southern, E. M. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol*. **98**, 503 (1975).

55. Casey, J. and Davidson, N. Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide. *Nucleic Acids Res*. **4**, 1539–1551 (1977).

56. Southern, E. *et al*. Molecular interactions on microarrays. *Nat Genet*. **21(1 Suppl)**, 5–9 (1999).

57. O'Shea, D., in *Introduction to Lasers and their applications*, Addison-Wesley, Reading, MA, pp. 51–78 (1978).

58. Smith, W. J., in *Modern Optical Engineering*, McGraw Hill, Boston, MA, pp. 142–145 (1990).

59. Skoog, D. A. *et al*., in *Principles of Instrumental analysis*, Harcourt Brace, Philadelphia, p. 108 (1998).

60. Gonzalez, R. C. and Woods, R. E., in *Digital Image Processing*, Addison-Wesley, Reading, MA, pp. 31–37 (1978).

61. Brazma A. *et al*. One-stop shop for microarray data. *Nature*. **403**, 699–700 (2000).

62. Brazma, A. and Vilo, J. Gene expression data analysis. *FEBS Letters* **480**, 17–24 (2000).

63. Brazma, A. *et al*. Minimum information about a microarray experiment (MIAME) — toward standards for microarray data. *Nature genetics*. **29**, 365–371 (2001).

64. Lee, M. *et al*. Importance of replication in microarray gene expression studies: Statistical methods and evidence from repetitive cDNA hybridizations. *Proc Natl Acad Sci*. **97**, 9834–9839 (2000).

65. Bernard, K. *et al*. Multiplex messenger assays: simultaneous, quantitative measurement of expression of many genes in the context of T cell activation. *Nucl Acids Res*. **24**, 1435–1442 (1996).

66. Yue, H. *et al*. An evaluation of the performance of cDNA microarrays for detecting changes in global mRNA expression. *Nucleic Acids Res*. **29**, e41 (2001).

67. Tseng, G. C. *et al*. Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucl. Acids Res*. **29**, 2549–2557 (2001).

68. Hemmilä, I.A., *Applications of Fluorescence in Immunoassays*, John Wiely and Sons, Inc. New York (1991).

Product	Quantity	Code Number
Microarray probe preparation		
CyScribe First-Strand cDNA Labelling Kit	25 reactions	RPN6200
CyScribe First-Strand cDNA Labelling Kit with CyScribe GFX Purification Kit	25 reactions	RPN6200X
CyScribe First-Strand cDNA Labelling System-dUTP	50 reactions	RPN6201
CyScribe First-Strand cDNA Labelling System-dUTP with CyScribe GFX Purification Kit	50 reactions	RPN6201X
CyScribe First-Strand cDNA Labelling System-dCTP	50 reactions	RPN6202
CyScribe First-Strand cDNA Labelling System-dCTP with CyScribe GFX Purification Kit	50 reactions	RPN6202X
CyScribe Post-Labelling Kit	24 reactions	RPN5660
CyScribe Post-Labelling Kit with CyScribe GFX Purification Kit	24 reactions	RPN5660X
CyScribe Direct mRNA Labelling Kit	24 reactions	RPN5665
Cy3-dCTP	25 nmol	PA53021
Cy5-dCTP	25 nmol	PA55021
Cy3-dUTP	25 nmol	PA53022
Cy5-dUTP	25 nmol	PA55022
Cy3-UTP	100 nmol	PA53026
Cy5-UTP	100 nmol	PA55026
CyDye Post-Labelling Reactive Dye Pack	24 reactions	RPN5661
Lucidea brand products		
Lucidea Array Spotter	1	63-0040-09
Lucidea SlidePro Module 1	1	18-1162-01
Lucidea SlidePro Module 2	1	18-1162-02
Lucidea SlidePro Module 3	1	18-1162-03
Lucidea SlidePro Module 4	1	18-1162-04
Lucidea SlidePro Module 5	1	18-1162-05
Lucidea Automated Spotfinder	1	63-0038-18
Lucidea Universal ScoreCard	200 hybridizations	63-0042-85
Lucidea Reflective Slides		Available soon
Microarray scanning systems		
Typhoon 9610 Variable Mode Imager	1	63-0038-55

Product	Quantity	Code Number
Image analysis software		
ArrayVision for Scanners	1	ARV-100
Spotfire DecisionSite for Functional Genomics	1	63-0036-56
ImageQuant Solutions for Windows 2000 for Scanners	1	63-0035-16
Bioinformatics products		
Scierra Microarray Laboratory Workflow System		
Electrophoresis systems		
Ready-to-Run Separations Unit	1	80-6460-95
RNA purification products		
RNase-free water	500 ml	US70783
QuickPrep Total RNA Extraction Kit	1	27-9271-01
RNA Extraction Kit	1	27-9270-01
QuickPrep Micro mRNA Purification Kit	1	27-9255-01
Spectrophotometry products		
Ultraspec 3300 pro UV/Visible Spectrophotometer	1	80-2112-33
Hybridization equipment and reagents		
Microarray Hybridization Solutions version 2	1	RPK0325
Humid Hybridization Cabinet for microarrays	1	RPK0176
SSC 20×	100 ml	US19629
SDS 20%	500 ml	US75832
Other		
Hybond-N+ Membrane	50	RPN82B
Vistra Green Nucleic Acid Stain	500 ml	RPN5786
RapidGel-XL - 6%	100 ml	US75861
RapidGel-XL - 8%	100 ml	US75862
ALFexpress Sizer 50-500	50	27-4539-01
TBE Buffer Pre-mixed Powder 10×	6 bottles	US70454

Index