CHAPTER ONE

# Microarrays: Making Them and Using Them

## SECTION 1.1 INTRODUCTION

A DNA microarray consists of a solid surface, usually a microscope slide,[1] onto which DNA molecules have been chemically bonded. The purpose of a microarray is to detect the presence and abundance of labelled nucleic acids in a biological sample, which will hybridise to the DNA on the array via Watson–Crick duplex formation, and which can be detected via the label. In the majority of microarray experiments, the labelled nucleic acids are derived from the mRNA of a sample or tissue, and so the microarray measures gene expression. The power of a microarray is that there may be many thousands of different DNA molecules bonded to an array, and so it is possible to measure the expression of many thousands of genes simultaneously.

This book is about the bioinformatics of DNA microarrays: the mathematics, statistics and computing you will need to design microarray experiments; to acquire, analyse and store your data; and to share your results with other scientists. One of the features of microarray technology is the level of bioinformatics required: it is not possible to perform a meaningful microarray experiment without bioinformatics involvement at every stage.

However, this chapter is different from the remainder of the book. While the other chapters discuss bioinformatics, the aim of this chapter is to set out the basics of the chemistry and biology of microarray technology. It is hoped that someone new to the technology will be able to read this chapter and gain an understanding of the laboratory process and how it impacts the quality of the data. The chapter is arranged into two further sections:

Section 1.2: Making Microarrays, describes the main technologies by which microarrays are manufactured.
Section 1.3: Using Microarrays, describes what happens in a microarray laboratory when a microarray experiment is performed.

## SECTION 1.2 MAKING MICROARRAYS

There are two main technologies for making microarrays: **robotic spotting** and **in-situ synthesis**.

[1] Historically, microarrays have also been produced using nylon filters and larger glass slides.

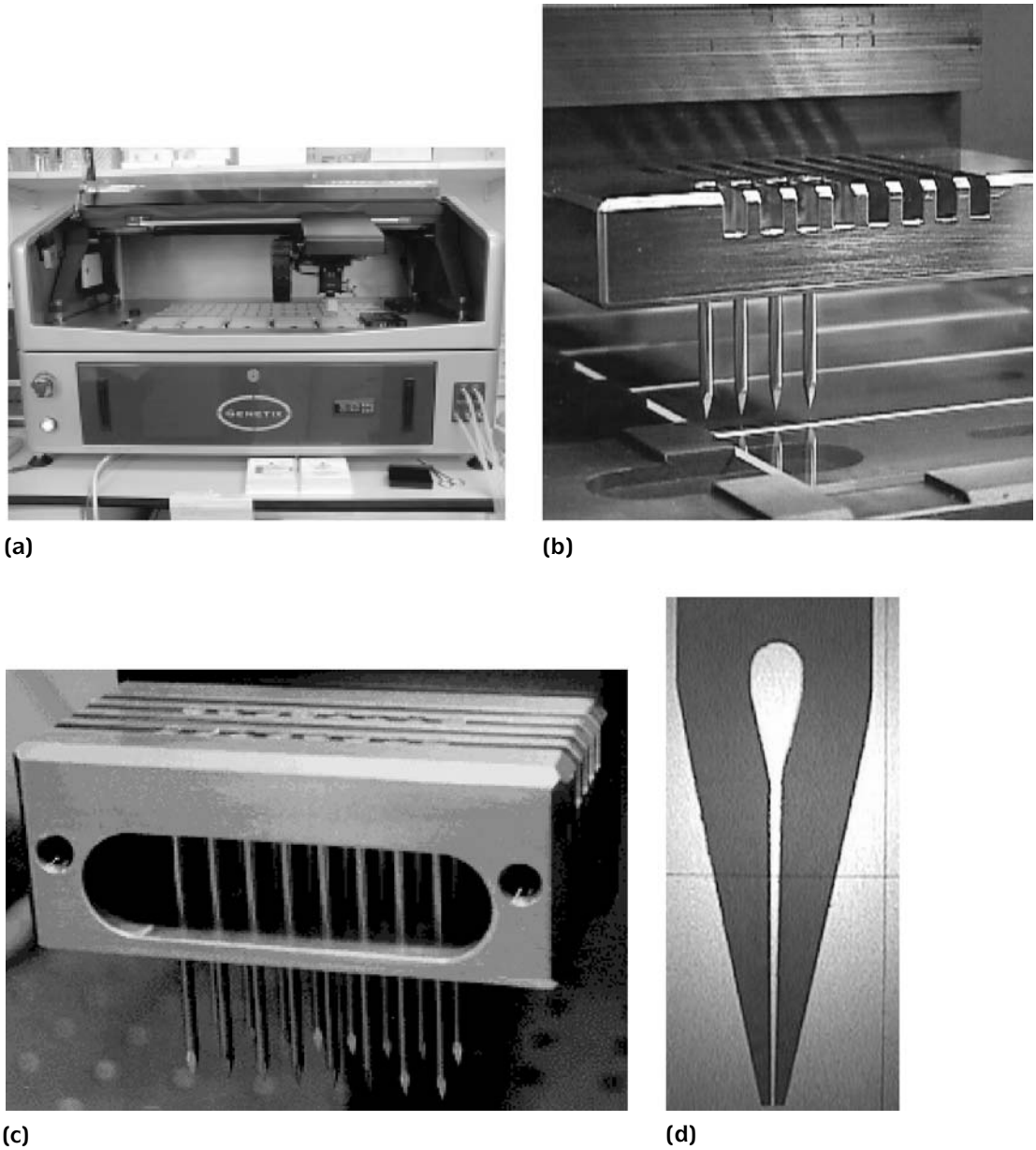


**Figure 1.1: Spotting robot. (a)** An example of a spotting robot. There are many different robots on the market; this is a Genetix spotting robot located at the Mouse Genetics Unit in Harwell, Oxfordshire. **(b)** The pins are held in a cassette in a rectangular grid, which in turn is held on a robot arm that can be moved between the microtiter well plates and the glass arrays to deposit liquid. **(c)** The number of pins in the cassette can vary. The more pins, the greater the throughput of the robot, but the greater the propensity for pin-to-pin variability. Each pin will spot a different grid on the array (Chapter 4). **(d)** Most pins in modern use have a reservoir that holds sample and so can print multiple features – usually on different arrays – from a single visit to the well containing probe. Earlier robots use solid pins, which can only print one feature before needing to collect more DNA from the well.

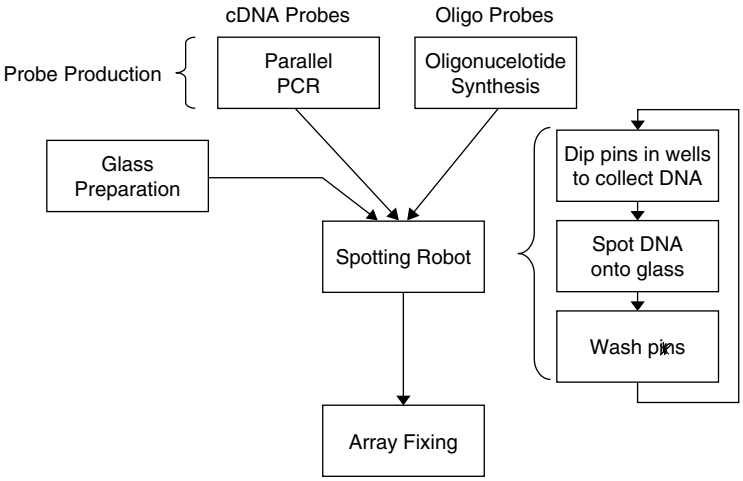


**Figure 1.2: Spotted array synthesis.** There are several steps involved in the synthesis of spotted arrays. First is the production of the probes. cDNA probes are made via highly parallel PCR; oligonucleotide probes have to be presynthesised. The next step is the spotting step. The robot arm moves the cassette containing the pins over one of the microtiter plates containing probe and dips the pins into the wells to collect DNA. The arm is then moved over the first array and the cassette is moved down so that the pins touch the glass and DNA is deposited on the surface. If more than one array is being synthesised, the cassette is moved to the subsequent arrays. Before collecting the next DNA to be spotted, the pins are washed to ensure no contamination. The final step of array production is fixing, in which the surface of the glass is modified so that no additional DNA can stick to it.

## Spotted Microarrays

This is the technology by which the first microarrays were manufactured. The array is made using a spotting robot (Figure 1.1a) via three main steps (Figure 1.2):

1. Making the DNA probes[2] to put on the array
2. Spotting the DNA onto the glass surface of the array with the spotting robot
3. Postspotting processing of the glass slide

There are three main types of spotted array (Table 1.1), which can be subdivided in two ways: by the type of DNA probe, or by the attachment chemistry of the probe to the glass.

The DNA probes used on a spotted array can either be polymerase chain reaction (PCR) products or oligonucleotides. In the first case, highly parallel PCR is used to amplify DNA from a clone library, and the amplified DNA is purified. In the second case, DNA oligonucleotides are presynthesised for use on the array.

[2] There are now three camps in the microarray community as to what to call the DNA on the array and the DNA in solution. Throughout this book, we will use the "Southern" terminology and refer to the DNA on the array as *probes* and the labeled DNA in solution as *target*. Other researchers refer to the DNA on the array as *target* and the labeled DNA in solution as *probe*. More recently, MIAME (Minimal Information About a Microarray Experiment) introduced a new convention of referring to the DNA on the array as *reporters* and the DNA in solution as the *hybridisation extract*. MIAME conventions are described in full in Chapter 11, and MIAME terms are detailed in the Appendix.

**TABLE 1.1**

| | Surface Chemistry | |
|---|---|---|
| **DNA Probes** | **Covalent** | **Non-covalent** |
| Oligonucleotides | ✓ | |
| cDNAs | ✓ | ✓ |

*Note:* There are three types of spotted microarrays, which can be thought about in two different ways. The DNA probes can be oligonucleotides or cDNAs; the surface chemistry can be covalent or non-covalent. Oligonucleotide probes can only be attached covalently; cDNA probes can be attached either covalently or non-covalently. Covalent attachment is via an aliphatic amine ($NH_2$) group added to the 5′ end of the DNA probe, and consequently the probes are tethered to the glass from the 5′ end. Non-covalent attachment is via electrostatic attraction between amine groups on the glass slide and the phosphate groups on the DNA probe backbone; thus the DNA probe is attached to the glass by its backbone.

The attachment chemistry can either be covalent or non-covalent. With covalent attachment, a primary aliphatic amine ($NH_2$) group is added to the DNA probe and the probe is attached to the glass by making a covalent bond between this group and chemical linkers on the glass. With oligonucleotide probes, the amine group can be added to either end of the oligonucleotide during synthesis, although it is more usual to add it to the 5′ end of the oligonucleotide. With cDNA probes, the amine group is added to the 5′ end of the PCR primer from which the probes are made. Thus the cDNA probes are always attached from the 5′ end.

With non-covalent attachment, the bonding of the probe to the array is via electrostatic attraction between the phosphate backbone of the DNA probe and $NH_2$ groups attached to the surface of the glass. The interaction takes place at several locations along the DNA backbone, so that the probe is tethered to the glass at many points. Because most oligonucleotide probes are shorter than cDNAs, these interactions are not strong enough to anchor oligonucleotide probes to glass. Therefore, non-covalent attachment is usually only used for cDNA microarrays.

The DNA probes are organised in microtiter well plates, typically 384 well plates. Most modern spotting robots will use a number of plates to print arrays, so the plates are arranged in a "hotel," whereby the robot is able to gain successive access to each of the plates. The spotting robot itself consists of a series of pins arranged as a grid and held in a cassette (Figures 1.1b and 1.1c). The pins are used to transfer liquid from the microtiter plates to the glass array.[3]

There are a number of different designs of pins. The first spotting robots used solid pins (Figure 1.1b); these can only hold enough liquid for one spot on the array, thus requiring the pin cassette to return to the plate containing probe before printing the next spot. Most array-making robots today have pins with a reservoir that holds

[3] Not every spotting robot is a pin-based system: Perkin Elmer sell some robots which use a piezoelectric system to fire tiny drops of liquid onto the arrays. These are in the minority in the microarray field.

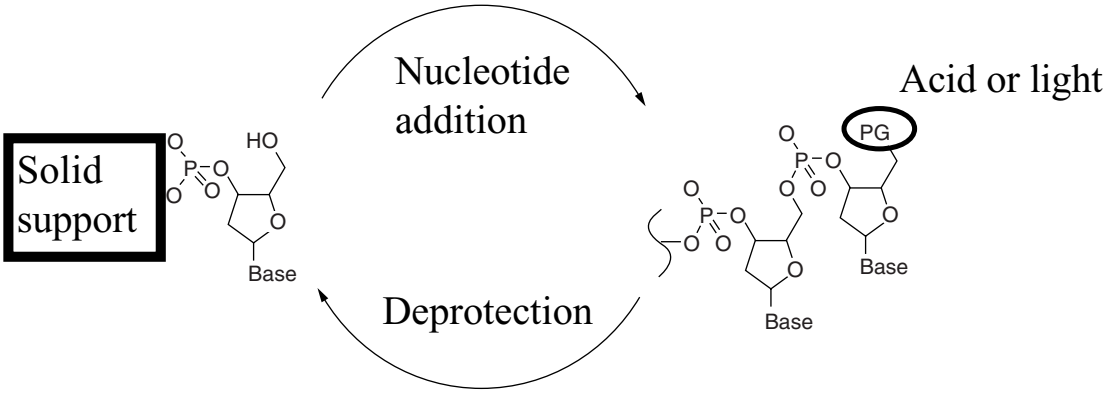


**Figure 1.3: In-situ synthesis of oligonucleotides.** The oligonucleotides are built on the glass array one base at a time. At each step, the base is added via the reaction between the hydroxyl group 5′ of the terminal base and the phosphate group of the next base. There is a protective group on the 5′ of the base being added, which prevents the addition of more than one base at each step. Following addition, there is a deprotection step at which the protective group is converted to a hydroxyl group to allow addition of the next base.

liquid (Figure 1.1d). This enables higher throughput production of arrays because each probe can be spotted on several arrays without the need to return the pins to the sample plates.

The typical printing process follows five steps (Figure 1.2):

1. The pins are dipped into the wells to collect the first batch of DNA.
2. This DNA is spotted onto a number of different arrays, depending on the number of arrays being made and the amount of liquid the pins can hold.
3. The pins are washed to remove any residual solution and ensure no contamination of the next sample.
4. The pins are dipped into the next set of wells.
5. Return to step 2 and repeat until the array is complete.

In the final phase of array production, the surface of the array can be *fixed* so that no further DNA can attach to it. There are many fixing processes that depend on the precise chemistry on the surface of the glass. The desired outcome is always the same: we do not want DNA target from the sample to stick to the glass of the array during hybridisation, so the surface must be modified so this does not happen. It is also common to modify the surface so that the glass becomes more hydrophilic because this aids mixing of the target solution during the hybridisation stage. Some microarray production facilities do not fix their arrays.

### In-Situ Synthesised Oligonucleotide Arrays

These arrays are fundamentally different from spotted arrays: instead of presynthesising oligonucleotides, oligos are built up base-by-base on the surface of the array (Figure 1.3). This takes place by covalent reaction between the 5′ hydroxyl group of

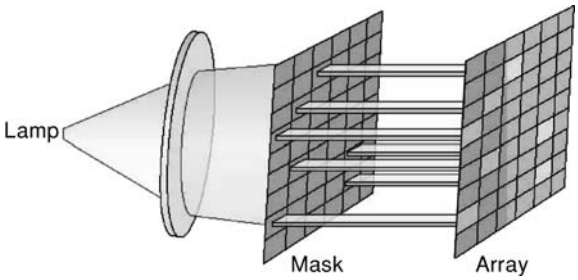


**Figure 1.4: Affymetrix technology.** Affymetrix arrays are manufactured using in-situ synthesis with a light-mediated deprotection step. During each round of synthesis, a single base is added to appropriate parts of the array. A mask is used to direct light to the appropriate regions of the array so that the base is added to the correct features. Each step requires a different mask. The masks are expensive to produce, but once made, it is straightforward to use them to manufacture a large number of identical arrays. (Reproduced with Permission from Affymetrix Inc.)

the sugar of the last nucleotide to be attached and the phosphate group of the next nucleotide. Each nucleotide added to the oligonucleotide on the glass has a protective group on its 5′ position to prevent the addition of more than one base during each round of synthesis. The protective group is then converted to a hydroxyl group either with acid or with light before the next round of synthesis. The different methods for deprotection lead to the three main technologies for making in-situ synthesised arrays:

1. Photodeprotection using masks: this is the basis of the Affymetrix® technology.
2. Photodeprotection without masks: this is the method used by Nimblegen and Febit.
3. Chemical deprotection with synthesis via inkjet technology: this is the method used by Rosetta, Agilent and Oxford Gene Technology.

### Affymetrix Technology

Affymetrix arrays use light to convert the protective group on the terminal nucleotide into a hydroxyl group to which further bases can be added. The light is directed to appropriate features using *masks* that allow light to pass to some areas of the array but not to others (Figure 1.4). This technique is known as photolithography and was first applied to the manufacture of silicon chips. Each step of synthesis requires a different mask, and each mask is expensive to produce. However, once a mask set has been designed and made, it is straightforward to produce a large number of identical arrays. Thus Affymetrix technology is well suited for making large numbers of "standard" arrays that can be widely used throughout the community.

### Maskless Photodeprotection Technology

This technology is similar to Affymetrix technology in that light is used to convert the protective group at each step of synthesis. However, instead of using masks, the light is directed via micromirror arrays, such as those made by Texas Instruments. These are solid-state silicon devices that are at the core of some data projectors: an array of mirrors is computer controlled and can be used to direct light to appropriate parts

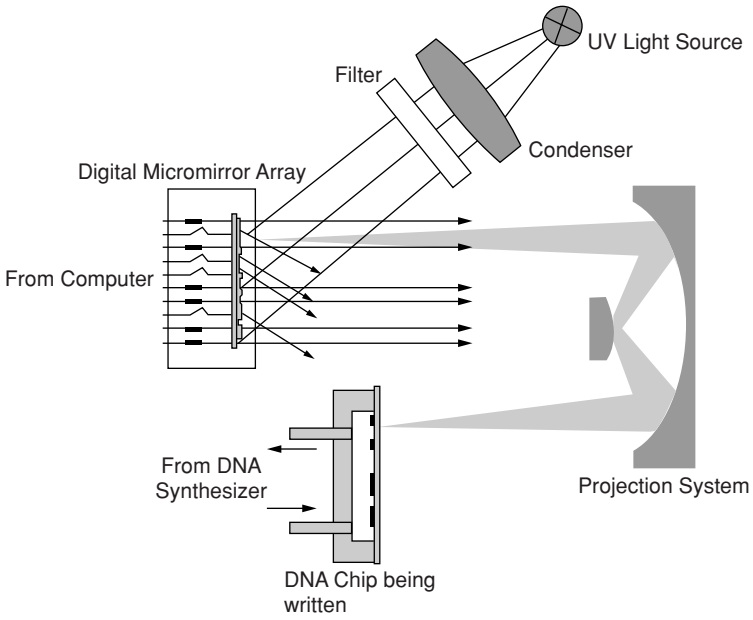


**Figure 1.5: Maskless photodeprotection.** This system also uses light-mediated deprotection. However, instead of using a physical mask, the array is synthesised using a computer-controlled micromirror array. This consists of a large number of mirrors embedded on a silicon chip, each of which can move between two positions: one position to reflect light, and the other to block light. At each step, the mirrors direct light to the appropriate parts of the array. This technology is used by Nimblegen and Febit.

of the glass slide at each step of oligonucelotide synthesis (Figure 1.5). This is the technology used by Nimblegen and Febit.

### Inkjet Array Synthesis

Instead of using light to convert the protective group, deprotection takes place chemically, using the same chemistry as a standard DNA synthesiser. At each step of synthesis, droplets of the appropriate base are fired onto the desired spot on the glass slide via the same nozzles that are used for inkjet printers; but instead of firing cyan, magenta, yellow and black ink, the nozzles fire A, C, G and T nucleotides (Figure 1.6).

One of the main advantages of micromirror and inkjet technologies over both Affymetrix technology and spotted arrays is that the oligonucleotide being synthesised on each feature is entirely controlled by the computer input given to the array-maker at the time of array production. Therefore, these technologies are highly flexible, with each array able to contain any oligonucleotide the operator wishes. However, these technologies are also less efficient for making large numbers of identical arrays.

### Synthesis Yields

The different methods of oligonucleotide synthesis have different coupling efficiencies: this is the proportion of nucleotides that are successfully added at each step of synthesis. Photodeprotection has a coupling efficiency of approximately 95%, whereas

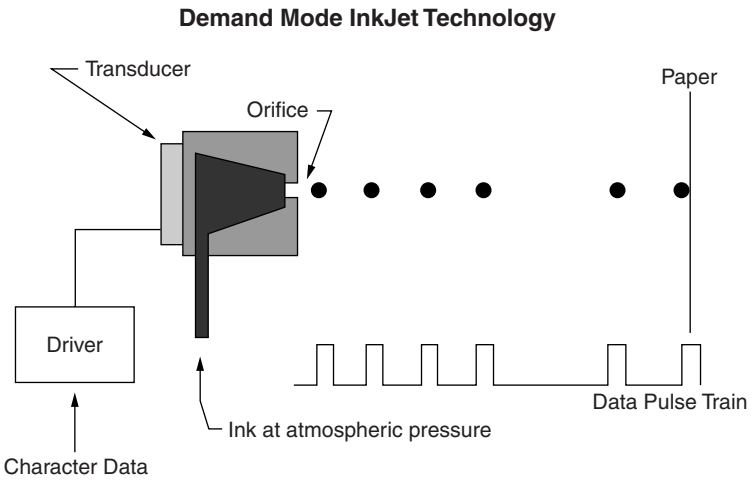


**Figure 1.6: Inkjet array synthesis.** This technology uses chemical deprotection to synthesise the oligonucleotides. The bases are fired onto the array using modified inkjet nozzles, which, instead of firing different coloured ink, fire different nucleotides. At each step, the appropriate nucleotide is fired onto each spot on the array. The nozzles are computer controlled, so any oligonucleotides can be synthesised on the array simply by specifying the sequences in a computer file. This is the technology used by Rosetta, Agilent and Oxford Gene Technology.

acid-mediated deprotection of dimethoxytrityl protecting groups has a coupling efficiency of approximately 98%. The effect on the yield of full-length oligos is dependent on the length of the oligonucleotide being synthesised: the longer the oligonucleotide, the worse the yield. This dependence is multiplicative, so that even a small difference in coupling efficiency can make a large difference in the yield of long oligonucleotides (Table 1.2).

The composition of the final population of oligonucleotides produced depends on whether or not a capping reaction is included during synthesis. Capping is used by Affymetrix and prevents further synthesis on a failed oligonucleotide. As a result,

**TABLE 1.2**

| Oligonucleotide Length ($s$) | Coupling Efficiency ($p$) | Oligonucleotide Yield ($p^s$) |
|---|---|---|
| 25 | 95% | 28% |
| 25 | 98% | 60% |
| 60 | 95% | 5% |
| 60 | 98% | 30% |

*Note:* The yield of in-situ synthesised oligonucleotides of desired length $s$ depends on the coupling efficiency $p$ according to the formula yield $= p^s$. So the longer the oligonucleotide, the worse the yield. Photodeprotection has a coupling efficiency of approximately 95%, while chemical deprotection has a coupling efficiency of approximately 98%. For a 25-base oligonucleotide, the yields are 28 and 60%, respectively. For a 60-base oligonucleotide, the yields are 5 and 30%, respectively. This is why Affymetrix is restricted to making 25-base oligonucleotides: the coupling efficiency is too low to produce longer oligos. Companies using chemical deprotection are able to synthesise 60-base oligos with similar yield to Affymetrix's 25-base oligos.

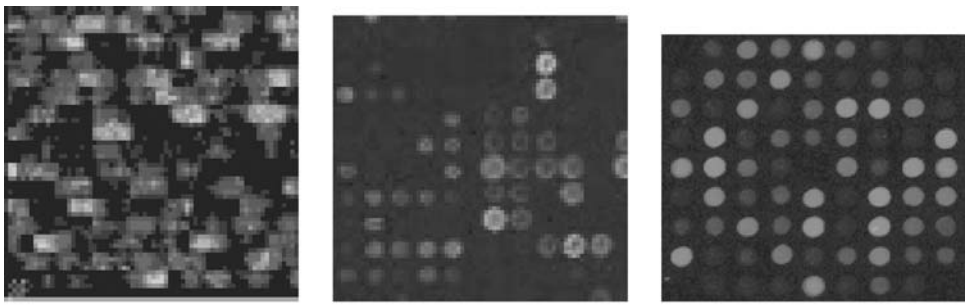


**Figure 1.7: Array quality. (a)** On Affymetrix arrays the features are rectangular regions. The masks refract light, so there is leakage of signal from one feature to the next. The Affymetrix image-processing software compensates for this by using only the interior portions of the features. **(b)** Spotted arrays produce spots of variable size and quality. This image shows some of this variation; we cover image processing of spotted arrays in detail in Chapter 4. **(c)** Inkjet arrays tend to be of the highest quality, with regular, even spots. (*Please see also the color section at the middle of the book.*)

all oligonucleotides on a feature will have the same start, but will be of different lengths (e.g., with a coupling efficiency of 95%, each feature will be 4.8% monomers, 4.5% dimers, 4.3% trimers, etc.). In contrast, uncapped oligonucleotides allow further synthesis to take place. Therefore, all the oligonucleotides on a feature will be of similar length but may contain random deletions (e.g., with a coupling efficiency of 95% and synthesis of 20 mers, the average probe length would be 19 bases, with such probes containing one deletion).

### Spot Quality

The quality of the features depends on the method of array production (Figure 1.7). Spotted array images can be of variable quality, and Chapter 4 is dedicated to the bioinformatics of image processing associated with these arrays. Affymetrix arrays have the problem that the masks refract light, so light leaks into overlapping features; Affymetrix compensates for this with their image-processing software, so the user need not worry about this problem. Inkjet arrays tend to produce the highest quality features.

## SECTION 1.3 USING MICROARRAYS

There are four laboratory steps in using a microarray to measure gene expression in a sample (Figure 1.8):

1. Sample preparation and labelling
2. Hybridisation
3. Washing
4. Image acquisition

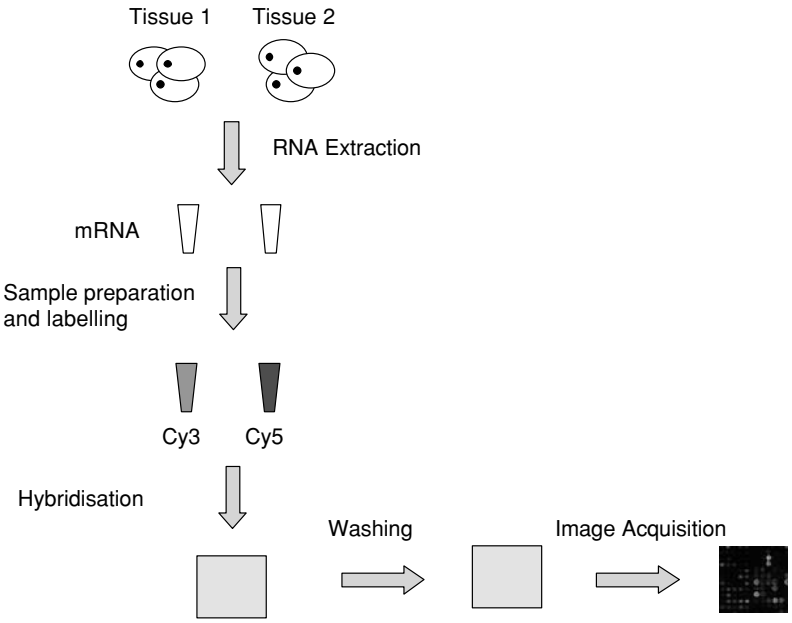


**Figure 1.8: Steps in using a microarray.** The first step is to extract the RNA from the tissue(s) of interest. With most technologies, it is common to prepare two samples and label them with two different dyes, usually Cy3 (green) and Cy5 (red). The samples are hybridised to the array simultaneously and incubated for between 12 and 24 hours at between 45 and 65°C. The array is then washed to remove sample that is not hybridised to the features.

## Sample Preparation and Labelling

There are a number of different ways in which a sample can be prepared and labelled for microarray experiments. In all cases, the first step is to extract the RNA from the tissue of interest. This procedure can be difficult to reproduce, and there is much variability among the individual scientists performing the extraction.

The labelling step depends on the technology used. For the Affymetrix platform, one constructs a biotin-labelled complementary RNA for hybridising to the GeneChip®. The protocols are very carefully defined by Affymetrix,[4] so every Affymetrix laboratory should be performing identical steps. This has the advantage that it is easier to compare the results of experiments performed in different Affymetrix laboratories, because the procedures they will have followed should be the same.

Although it is possible to hybridise complementary RNA to other types of microarrays, it is much more common to hybridise a complementary DNA to these arrays. In the past, the DNA has been radioactively labelled, but now most laboratories use fluorescent labelling, with the two dyes Cy3 (excited by a green laser) and Cy5 (excited by a red laser). In the most common experiments, two samples are hybridised to the arrays, one labelled with each dye; this allows the simultaneous measurement of both

[4] See the reference to the Affymetrix manual at the end of the chapter.