# Construction of microarrays and their application to virus analysis

**Michael J. Allen, Bela Tiwari, Matthias E. Futschik, and Debbie Lindell**

## Abstract

In this virus-focused chapter, we deal primarily with the use of microarrays for expression analysis (the most popular usage) of host and virus systems during infection. We examine aspects related to array platform choice (spotted and oligonucleotide arrays), probe and array design considerations, experimental procedures and data analysis, normalization, processing, and curation. We also provide in-depth examples for the study of viral transcriptome analysis for both spotted long oligonucleotide (coccolithoviruses) and Affymetrix GeneChip (cyanophage) arrays.

## Guidelines

### Microarray platforms

At every stage of microarray design and construction, there are various options that can be taken depending on the available budget, the number of probes on the array, and the local infrastructure and facilities available (Fig. 1). It is crucial that before microarray design begins, the scientific questions of interest are considered and defined, so that a microarray fit for the purpose is constructed. For example: Can the virus system be accurately studied on a microarray independent of the host system? Are there any genes common to both virus and host genomes? How much virus message will be present in relation to host message at various stages of infection? Will any amplification of message be needed? The answers to these sorts of questions can have a profound impact on the nature of the microarray developed. Other questions that may affect your microarray design include the following: How many microarray experiments will be run? Will the microarray be used by just one research group, or will it be made available to other interested parties? What local microarray infrastructure is available?

Here we focus primarily on two array platforms: spotted arrays and Affymetrix GeneChip arrays. We also occasionally comment on Agilent arrays when relevant. Other array types exist, but these three are common platforms that provide a good overview of the different approaches currently employed. Each array platform offers its own particular advantages and disadvantages, which we discuss. We start off describing high-density custom oligonucleotide systems and then move on to the custom spotted microarrays. One issue we will not touch on in this review is specific costs, which are highly subjective and prone to changes in the future; here, we discuss only the current costs of each system relative to other options presently available. With respect to cost calculations, the stripping and reuse of previously used microarrays is technically feasible for many types of microarray system, but we strongly advise against this practice as it can introduce uncontrollable variability.

### High-density oligonucleotide microarrays

High-density oligo - nucleotide microarray systems (such as the Affymetrix and Agilent systems) usually offer the best reliability, reproducibility, and coverage. Many high-density microarrays are constructed using a process known as photolithography (Affymetrix and Nimblegen), whereby light is used to stimulate in situ DNA synthesis in defined positions; others use inkjet technology (such as Agilent) (Table 1). In both these cases, single-stranded oligonucleotides are sequentially synthesized base by base, directly on the solid surface of the array. Whereas Affymetrix technology uses a number of short 25-mer probes per gene (generally 8–11), Agilent arrays consist of a single 60-mer probe for each gene. The manufacturing process is an extremely robust procedure with no noticeable differences between arrays. A powerful application of high-density microarrays is to produce what is referred to as a tiling array. Tiling arrays are designed independently of annotation data, cover entire stretches of genomic sequence (usually total genomes) in an unbiased manner (e.g., 25-mer probes designed with a space of approximately 50 bases in between, along the length of the whole genome irrespective of annotation), and allow the identification of novel transcribed sequences (often unannotated) as well as regulatory elements. By determining which probes generate positive signals and their intensity relative to neighboring probes, regions of the genome that are transcribed (i.e., the genes) or are regulatory can be easily identified. Tiling arrays, developed by companies including Affymetrix, Nimblegen, and Agilent, are incredibly powerful but are commonly restricted to model organisms for which there is a large commercial market.

Affymetrix expression arrays can produce highly reproducible data. These arrays are designed and manufactured by Affymetrix based on empirical but proprietary information and therefore are the easiest for the researcher to "design," although also the most costly. After RNA is extracted, it can be labeled by the researcher or, for a cost, at an Affymetrix array facility. Hybridization of the arrays is carried out at a specialized array facility, generally by facility personnel. This makes this procedure relatively simple, especially for researchers not so familiar with RNA work, but requires that the researcher find a reliable facility. The greatest disadvantage of these arrays is the high cost incurred for the custom design necessary for nonmodel systems. Cost per array is also quite high (with a minimum order of 90 arrays).

Agilent arrays—both probes and array layout—can be designed for free (using their Web program at earray.chem.agilent.com/earray) or can be designed by Agilent for a reasonable fee. A major advantage of the Agilent platform is the flexibility for probe and array design it provides, being feasible to order any number of arrays at a time (even a single array) and redesign probes for subsequent arrays. RNA labeling can be carried out by the researcher or at an array facility; however, hybridization and scanning are generally carried out at an array facility, again making the process quite simple for the researcher. The biggest disadvantage of the Agilent arrays is the cost per array, which is considerably higher than for the other platforms. It is possible, however, to hybridize multiple samples in different compartments on one slide if the number of probes is small enough (for example, if investigating the expression dynamics of the viral genome alone), making the price more reasonable per sample.

Thus, for high-density oligonucleotide arrays, Affymetrix is the platform of choice for systems when a large number of experiments will be carried out, although the cost for array design is quite high. Agilent is the platform of choice when high design flexibility is desired and few samples will be investigated, although the cost per array is quite high.

### Custom spotted microarrays

The costs associated with developing high-density microarrays can make them financially prohibitive. The development of a custom spotted array is an economical alternative, although there is a price to pay in reduced array reproducibility. A spotted microarray is also the platform of choice for large-volume experiments when high design flexibility is required. The researcher has the added advantage of having complete control over array design and total flexibility in its use. For spotted microarrays, the most common method is to use a glass slide as the basis for printing. Glass slide arrays offer researchers great control: they can generate their own labeled samples, perform their own hybridizations, and scan using their own equipment. Depending on the infrastructure, once microarrays have been fabricated, all that is needed is a microarray scanner and some basic laboratory equipment. Probes can be immobilized onto glass slides by a variety of techniques. Initially, this was done by physical contact between robotically controlled pins and the slide surface. It is now more common to use a technique known as piezoelectric printing, which is akin to ink-jet printing and provides greater control over spotting quality and quantity. As the print head moves across the array, electrical stimulation causes the DNA to be delivered onto the surface via tiny jets in a noncontact process. The process of "slide printing" is time consuming, technically demanding, and requires expensive robotic machinery. Often this is beyond the budget of most research laboratories; however, microarray printing facilities are now commonplace and offer a cheap and reliable method of fabricating microarrays.

Perhaps the greatest advantage of glass slide microarrays over their high-density counterparts is the flexibility they offer in the nature of the material printed on the slide. Many different types of material, including PCR products, plasmid libraries (cDNA/Expressed Sequence Tag [EST], plasmid, shotgun [randomly fragmented DNA]), or presynthesized oligonucleotides have been printed on glass slide microarrays depending on what was available and most cost-effective for a given project. We strongly recommend the use of long, presynthesized oligonucleotides when sequence information is available. Presynthesized oligonucleotides provide high specificity and allow design of probes with similar hybridization characteristics. A length between 50 and 70 bases generally provides a good balance between sensitivity, specificity, cost, and consideration of the decreased coupling efficiency during synthesis with increasing probe length. As the price of generating longer probes has decreased over recent years, we recommend probes of approximately 70 base pairs (although in theory they can be synthesized to any size required). Long oligonucleotide probes have the added advantages of not needing certain quality checks required for other material types, for example, amplifying and verifying the quality of PCR fragments, as well as avoiding the problems associated with hybridizing to probes of different length and different annealing temperatures, which arise when using PCR and plasmid probes.

For those who do not have access to preexisting sequence data, probes can be generated from plasmid libraries. Here, researchers can hone in on a small number of unknown probes using a microarray, and sequence only relevant plasmids of direct interest to them toward the end of their experimental work. This used to be a popular use of microarrays; however, the price of sequencing and of generating synthetic oligonucleotides have dropped considerably in recent years. Thus, for most purposes, it is preferable to generate, sequence, and design long oligonucleotides, rather than generate physical materials such as plasmid libraries or PCR products for spotting onto arrays. High-throughput sequencing methods have already been used to provide sequence for use in designing transcriptomic arrays (Vera et al. 2008), and this is likely to become a popular avenue in the future.

For long oligonucleotide arrays, researchers must choose how much of the process they want to undertake locally versus the cost in time and money. Designing long oligonucleotide probes is within the abilities of most researchers, thanks to the ready availability of probe design software, both

commercial and free. However, one should not underestimate the time required to design an array layout and to quality check this array before any experimentation can begin. For many researchers, the services offered by a company, along with a quality guarantee for what they provide and the time scale within which they will provide it, may make the initially larger financial outlay worth it in the end.

When engaging a company, the normal rules stand: Ensure that you lay out clearly what you need and expect, and that you read the terms and conditions of their service in detail. Many companies and facilities have much experience working with model organisms and the types of chip designs one might desire for studying such organisms. The needs of the viral community can be somewhat different, however; for example, the requirement to spot multiple, unrelated genomes (e.g., virus and host) on a single chip, where the characteristics of these genomes can be quite different. Will the company print probes of different lengths? If not, how do they plan to deal with the likelihood of different melting temperature profiles among genes in each organism? If you are doing a diversity study, or screening for new genotypes, will they help you design appropriate probes? How many probes do you anticipate spotting? What is the minimum number of arrays you can have made? (For example, study the implications on choice if you are only going to run a few experiments/hybridizations with this design.) Could/should the company do the hybridizations for you?

If you are designing your own long oligonucleotide spotted array, then you must take into account the physical characteristics of the probes to include on the array, what types of sequences to represent, how many probes per gene (or region) to include, whether replicate spots will be printed, and what controls to print on the array. Some of these issues are common to both high-density and spotted arrays, although many are issues specific to just spotted arrays. Each of these topics is covered briefly below, as well as a brief comment on available microarray probe design software. In general, unless you already have the facilities locally, we recommend that you find a microarray facility to work with and undertake a detailed conversation with them about the needs of your experimental system.

### Array design considerations

Regardless of which array platform will be used, it is imperative that the researcher decides how the RNA will be labeled, as this determines whether the array should contain probes that are identical to or complementary to mRNA—termed a sense or antisense array, respectively, by Affymetrix. For example, a protocol that carries out reverse transcription to produce cDNA requires an antisense array, whereas some protocols that include an amplification step will produce DNA that is identical (sense) to the original RNA.

Probes bound to their targets should have approximately the same melting temperature (typically 50–60°C) across the array. Some oligonucleotide design software (see section below) will allow you to design probes with a range of lengths. This can be useful if more than one organism (e.g., host and virus) with different nucleotide characteristics (e.g., GC content) will be represented on the array. If you decide to design probes of different lengths, make sure that the company you are ordering the oligonucleotides from will manufacture them.

Probes need to be specific to the target of interest and sensitive enough to detect low levels of that target. Potential for secondary structure in the probe or the target can affect sensitivity; this is particularly relevant to the longer probes on spotted arrays. Some oligonucleotide design software include calculations for this potential. The specificity of a probe for its target will depend on how many

mismatches there are between them, and also the location and arrangement of those mismatches along the probe sequence (Letowski et al.2004). Gene-specific probes should have little or no sequence similarity to nontarget genes that may be present in the sample. Comprehensive studies on the effect of probe–target characteristics have provided tables of empirical results for essential design criteria for gene-specific and group-specific probes (He et al. 2005, Karaman et al. 2005). Note that some software use free energy in place of sequence identity as a measure of oligonucleotide specificity.

For bacterial arrays, where total RNA (i.e., mRNA, rRNA, and tRNA) is labeled, it is important to ensure that none of the probes are similar to ribosomal and transfer RNA sequences, as even very low labeling efficiency of these abundant RNAs is likely to mask mRNA levels. In addition, probes for bacteria and bacterial viruses should not be 3′ biased, as random hexamers, rather than polyT priming, will be used in the majority of protocols for making cDNA. Therefore, if a single probe is designed per gene it should be positioned toward the 5′ end of genes. If multiple probes are designed, as in Affymetrix arrays, we suggest these be spread across the gene, although they could also be designed toward the 5′ end of the gene.

If there is sufficient room on the array, designing probes in intergenic regions will enable identification of small unannotated genes that may have escaped annotation, as well as small noncoding RNAs that are often found in these regions (Steglich et al. 2008). The researcher should also consider whether probes for the detection of antisense RNAs will enhance the utility of the array. Perhaps rather than including probes that are antisense to all mRNAs, preliminary experiments for the detection of mRNAs, noncoding RNAs, and antisense RNAs could be carried out by Solexa- or 454-like sequencing, which would greatly inform on probe design (O. Wurtzel and R. Sorek, pers. comm.). Alternatively, as mentioned above, a tiling array could be designed across the genome on both strands.

High-density Affymetrix arrays can contain many probes on each array, making it worth considering including multiple genomes per array. However, if the genomes of two of these organisms will be investigated at the same time, as for host and viral genomes, probes with little cross-hybridization between the genomes must be designed (see "Probe design" below) and can be empirically tested with DNA from each organism. This is particularly important, as today we know that viral genomes often include host-like genes (Hughes and Friedman 2005, Moreira and Brochier-Armanet 2008, Monier et al. 2009). If this design is not feasible, and indeed whenever potential cross-hybridizing probes are being used, it is best to confirm the microarray results or carry out single-gene analysis from the outset, with a method capable of differentiating between host and viral copies of the genes—for example, using quantitative RT-PCR. Conversely, if the multiple genomes on the array will be investigated independently, i.e., separate arrays for each sample type, then there is no need to ensure that the probes do not cross-hybridize.

The types of probe sequences required on a microarray designed to study biodiversity will be different from those used on an array to study differential expression. Combinations of probes might be used to detect particular species or the presence of novel genotypes in a sample (Wang et al. 2002, 2003a; Rich et al. 2008). A recent review outlines microarray studies in microbial ecology (Gentry et al. 2006), including an overview of the types of probes included on arrays used in different types of studies. From this point in this review, we assume that the question of interest requires detection of only unique genes.

### Probe design

Many computer programs are available to aid in the design of long oligonucleotide probes for microarrays. These programs usually determine specificity and the potential for cross-hybridization potential of all probes designed. Some of these programs are commercial, but many are open source and freely available for academic use (Li et al. 2002, Herold and Rasooly 2003, Nielsen et al. 2003, Rouillard et al. 2003, Wang and Seed 2003, Chou et al. 2004, Reymond et al. 2004, Chung et al. 2005, Li et al. 2005, Nordberg 2005, Stenberg et al. 2005, Schretter and Milinkovitch 2006). When looking for a software package, important considerations include on what basis it chooses probes, what type of input data it requires, what format the output data will be in, how easy it is to install, and how easy it is to use. Ideally, there will be empirical evidence available on how well the software has worked in designing probes for microarrays already in use. Another key consideration is whether the software is installed and runs on a local machine or on a remote machine (e.g., entering your data via a Web site). Running programs locally has the benefit that data are secure and private, whereas running programs remotely depends on the good will of someone else, and in essence involves passing your sequence set onto someone else's machine for processing. On the upside, the machine at the other end may be more powerful and therefore quicker (or not!) than what is available locally, and the maintenance and installation of the software is someone else's responsibility. You also need to ensure that you enter your sequence data in the appropriate direction (see probe direction section above). More detailed outlines of software considerations, as well as desirable probe characteristics, can be found in specific reviews of the topic (Millard and Tiwari 2009). As a good starting point, the authors have found that the Picky and Yoda software packages are both very user friendly. They are available fromcomplex.gdcb.iastate.edu/download/Picky/index.html and pathport.vbi.vt.edu/YODA, respectively. In addition, many oligonucleotide software design packages are listed, and some reviewed, at nebc.nox.ac.uk/tools/bioinformatics-docs/otherbioinf/oligo-nucleotide-design.

## Controls

Control spots are vital in the assessment of quality, sensitivity, and reliability of microarray experiments. Different types of controls can be used to assess various quality aspects. We strongly recommend the inclusion of spike-in labeling controls—probes for RNA that will be spiked into the sample at defined concentrations, before the labeling procedure. They can be used to estimate the minimum amount of transcript as well as the minimum change in transcript abundance detectable by the array technology being used. It is important to include a sufficient number of spikein controls so that they can be added at varying concentrations to cover the signal intensity range of the experimental transcripts. Control probes should be placed on the array such that they appear across the spatial dimensions of the array. If sufficient controls are included, they can be used for data normalization (covered below). Affymetrix arrays include probes for detecting spike-in hybridization controls as a default, but extra probes for labeling need to be requested. Commercial control probes and their partner targets are available from companies such as Stratagene. Obviously, all spike-in controls should be for organisms other than those being investigated and should show no crosshybridization to the experimental genomes. When working with unusual organisms, especially those for which no sequence is available in the public databases, it is worth checking with the company involved to ensure that their controls do not contain sequences similar to anything in your samples. If you don't have sequences for your own samples, or you cannot get information on control sequences from the company, you may need to run test hybridizations without the labeled control targets added to ensure that there is no cross-hybridization. The inclusion of appropriate controls at the fabrication stage is particularly pertinent for virus-focused microarrays, where virus message may or may not be present at all in the early stages of the experiment (i.e., the uninfected state of a transcriptional profiling experiment). This is discussed further in the data normalization section below. It is also useful to include empty spots, or probes for which no spike-ins will be added, which can

provide an indication of nonspecific background signal. For spotted arrays, print-buffer spots are useful to check that no carryover effects occur during the printing of spotted arrays, as such artifacts may especially compromise the measurement of weakly expressed genes.

Genome annotation is often a process in flux, changing as better bioinformatics tools are developed and more experimental information becomes available. It is therefore useful to be able to change the annotation of genes on the array. It is possible for bioinformaticians to re-annotate arrays themselves for all platforms, but it is not always trivial. We have found that Affymetrix will support the need to change the appropriate files free of charge for a while, but will eventually start charging for this service; we suggest that the number of times Affymetrix will carry out this process be negotiated with them as part of the design contract. Importantly, once files with the updated annotation are included, old array data can be reanalyzed in light of the newly annotated protein coding genes, ncRNAs, or regulatory elements.

### *Microarray experimental design*

As in any scientific endeavor, appropriate experimental design is crucial. Many aspects of planning microarray experiments are the same as for other types of experiment requiring statistical analysis of the resulting data. For those who are not comfortable with statistics, there is a solution: find a collaborator who is. This is not a glib statement, it is a serious recommendation. Planning your experiment with someone who is familiar with microarray statistics and experimental design can mean the difference between generating data that allows you to address your questions of interest or generating data providing little or no scope for meaningful analysis.

The design will include defining the type and number of samples needed, certain aspects of their preparation and replication, the number of slides to be hybridized, and what samples will be hybridized to the same slides in the case of twocolor experiments. Too often, researchers perform an entire experiment and then provide a block of data to a statistician to be analyzed, having unknowingly introduced bias or without including appropriate replicates. When this occurs, it can make analysis difficult or impossible, meaning that all your time, samples, and money have just been wasted. Below we briefly outline some basic considerations for experimental design of microarray experiments.

Replicates: Different samples of the same type (e.g., samples of the organism exposed to the same treatment) are referred to as biological replicates. If you measure the same sample twice (e.g., take the same extract and put it on two microarray slides), this is a technical replicate. Common questions that arise when planning a microarray experiment are how many biological and technical replicates are needed? If your aim is to compare gene expression levels between treatments or conditions, then measurements from biological replicates are essential. We recommend that the minimum number of biological replicates considered for most situations is three, with four to six a more desirable number for spotted arrays. Note that if variability is high, even this number of biological replicates will be inadequate for certain types of analyses. Be wary of limiting the number of samples or experimental replicates too much based purely on cost or ease of obtaining samples, as this may lead to the inability to derive useful information from the experiment (and the money would have been better saved than spent on the microarrays).

Technical slide replicates inform on the signal variation due to "uninteresting" differences such as different slides, different hybridization chambers, different researchers carrying out protocols, etc. Technical replicates allow for certain types of quality control checks as well as providing greater

precision for a given measurement. To be able to glean meaningful data from such replicates, it is necessary to use sophisticated statistical software to set up statistical models that take technical replication into account.

A different sort of technical replication common in microarray systems is the inclusion of replicate probes on the arrays. This is especially common with arrays where there are only a small number of unique probes, as is the case in many viral–host arrays. Replicate probe spots provide information about signal variation related to position on the array. Options for the analysis of replicate spots include taking a simple arithmetic average of the measure for these replicate spots or using an error term in the statistical model for each gene to account for the variation between spots (Smyth et al. 2005). We recommend the latter. It is important to choose software capable of taking replicate spots into consideration and knowing how such replicate spots are treated. For example, some software assumes that spots with the same name are the same probe, whereas other software assumes that spot replicates are equidistant across the array.

In broad terms, we recommend as many true biological replicates as possible, with technical replicate spots within arrays providing useful information about technical variation if the data are handled appropriately. Although technical slide replicates are important to run during the initial setting up of your array-related protocols, they can be of limited use later on. We generally do not use technical replicate slides once the system has been set up, as biological variability is usually much greater than the technical variation between measurements, providing little additional knowledge of the biological system. Therefore we suggest adding more biological replication rather than including technical replicates.

Culture infection concerns: For those studying viral systems grown in lab culture, the division between biological and technical replicates, and single and pooled samples, is often not clear. Every flask of a host–virus system is a pool of organisms. Cultures in two flasks from the same exact source sample are similar to a technical replicate: measuring gene expression in these two samples is likely to give you an idea of how technical differences (slight temperature or light variation, flask conditions) affect gene expression in each pool of virus–host in each flask. We therefore suggest that cultures be grown separately over many generations to get an indication of the biological variation possible in this host–virus system. Where possible, we also suggest carrying out biological replication at different times (for example 1 month apart) to further control for differences that may relate to the particular run of an experiment. Another consideration for virus infection experiments when host gene expression is being investigated is carrying out paired experiments where each replicate culture is divided into two subcultures. One of the subcultures is infected with the virus and the second serves as the uninfected paired control. This can help reduce potentially irrelevant variability related to culture physiology that is not related to the infection process. It is important to note that standard statistical tests commonly used in microarray analysis often include the assumption of independent, identically distributed measurement errors. This means that each measurement from a sample you consider a biological replicate should be an independent measurement from truly different cultures and not from quasitechnical replicates, such as flasks grown from the same original culture.

Pooling: Pooling biological samples is often considered as a way to keep costs down (as a given number of samples can be measured using fewer arrays) or to reduce "noisy" variation. However, pooling biological replicates allows you to measure only a mean value for those samples. This leads to the lack of a measure of the biological variability in the system—a measure that is essential to determine the significance of changes in expression of one condition relative to another. If you do not know the inherent variability in your expression levels for a given treatment, you cannot determine

when expression is significantly different or within the level of normal variability. In general, we advise against pooling, as it is likely to mask results of interest, while usually not providing any real benefit. If you decide to pool samples anyway, it is vital to take the pooling into account when interpreting your experimental results. As a rule, pool samples only if the experiment is purely exploratory, for example, providing preliminary data, and if you intend to screen all candidate genes in each biological replicate by other techniques such as quantitative RT-PCR.

Variability and confounding: Experiments should be designed in a manner that ensures that sources of technical variability are not aligned, or *confounded*, with treatment types. Examples of confounding include the use of arrays from different batches for different conditions and the use of spotted arrays printed early in a print run for one condition and those from late in the run for another condition. To circumvent these types of problems, array use should be randomized. The order of slide use can be randomized by generating a set of random numbers. An excellent place to get random numbers for this purpose is www.random.org/sequences.

Wherever possible, one researcher should carry out a particular task for all samples. If this is not feasible, it is important to build the design so that each researcher looks after equal numbers of samples from each condition—preferably for paired infected and uninfected samples where appropriate. In a similar manner, it is important to ensure that hybridization of microarrays from one condition are not hybridized on one day and those from the other on a different day. This way you will not confound the effect of the researcher or the day on which they were handled with the condition itself.

### *Microarray hybridization designs*

A variety of experimental hybridization designs are commonly referred to in the microarray literature (Kerr and Churchill 2001). The key distinction between the design types is whether they allow direct or indirect comparison of samples. Direct comparisons refer to the hybridization of two samples to a single slide, providing a ratio indicating the relative expression levels of the two samples. Indirect comparisons refer to taking measurements from different slides and comparing them. Single-color designs, as used with the Affymetrix platform, are a type of indirect design. One sample is applied to each slide, and comparisons take place by considering biological replicate measurements of treated versus untreated samples. As such, one-color designs are relatively straightforward to devise and analyze. Two-color designs are more complicated. Thus, the rest of this section provides an overview of two-color microarray designs specific to spotted microarrays.

A common indirect design used with two-color microarray experiments is the reference design. In a reference design experiment, each sample is hybridized onto an array along with a *reference sample* (Fig. 2). This reference sample should ideally have a hybridization signal for all genes of interest during the course of your experiment, as you will be working with expression ratios. A common and generally effective choice for a reference sample is a pooled sample made from aliquots from each of the samples in the experiment (Kerr et al. 2007). Genomic DNA can also be used as a reference, although a different labeling strategy (i.e., labeling DNA, not RNA) must be used to generate such a reference sample. Despite this, it provides advantages in that every gene in your organism is represented at the same level above background, and it enables comparisons across experiments. In reference design experiments, the signal ratio is made up of your sample signal compared to a reference signal, and the ratios from different slides are then compared to one another. The indirect nature of reference design experiments makes them somewhat less efficient than direct designs, and they require more arrays (see below). However, they are relatively more straightforward and flexible

than other designs.

An alternative approach for two-color arrays is direct designs (Fig. 2). These can provide a more accurate estimate of differences between samples, as each sample is hybridized to the same slide as the sample it is being compared with. Much of the technical variation is cancelled out when these ratios are taken. In this model, uninfected samples could be compared directly with infected samples taken from the same time point.

A common extension to direct designs is loop designs (sometimes extended to interwoven loop designs) (Kerr and Churchill 2001, Kerr 2003) (Fig. 3). Here again, two samples of interest are hybridized to each slide; however, these designs involve a combination of direct and indirect comparisons. By comparing two conditions through a chain of other conditions, samples can be compared directly with other samples with a multiple-pairwise methodology (Pirooznia et al. 2008). These designs have the potential to be more efficient than a standard reference design and have stronger statistical power, but are considerably more complex (see Fig. 3). We recommend ILOOP, a freely available Web-based program, as a useful tool in finding optimal loop designs for two-color microarray experiments (Pirooznia et al. 2008). Be aware, though, that not all analysis software is capable of handling data from loop designs.

For virus–host studies using two-color arrays where either virus or both host and virus gene expression is to be investigated, we generally prefer a reference design. This is because uninfected samples have no real signal from the virus probes, which makes taking a direct ratio between uninfected and infected samples very problematic, as would be done in a direct or loop design. If, however, only host gene expression is of interest, direct or loop designs can work well. In a two-color experiment, each slide will have two samples hybridized to it. One will be labeled with Cy3 dye, and one with Cy5 dye. A dye swap involves labeling samples from a particular condition with the Cy3 on one slide, and with Cy5 on another slide. The aim of dye swaps are to avoid identifying genes as potentially interesting when in fact a relatively strong or weak signal is due to a bias associated with the dye a sample has been labeled with. A dye swap can be carried out with technical or biological replicates; however, for reasons discussed above, we recommend biological dye swaps if direct or loop designs will be used. For example, a sample pair from the treated and untreated conditions are labeled with Cy3 and Cy5, respectively, whereas samples from another biological replicate pair (from the same treated and untreated conditions) are labeled instead with Cy5 and Cy3, respectively. For reference designs, spot ratios biased by the dye used can be avoided by labeling the reference sample with a particular dye and the experimental samples with the other dye for the whole experiment.

### Sample labeling and microarray hybridization

Once you have designed your microarray, determined the experimental plan, and have extracted nucleic acid samples in hand (see "Case studies" in "Assessment" for in-depth examples), you are ready to proceed with the sample labeling and array hybridization. The method used for labeling your samples depends on the chemical nature of your sample (i.e., DNA or RNA) and the amount of sample available (low amounts of starting material may require an amplification of message step). Labeling of mRNA requires the use of reverse transcriptase, whereas labeling DNA requires the use of the Klenow fragment of DNA polymerase (i.e., minus the proofreading activity). Nucleotide mixes for these labeling reactions may require some optimization depending on the GC content of the systems under study. One needs to decide whether to use a direct or indirect labeling method (for spotted arrays); random, specific, or oligo dT primers; an amplification of message step. Some commercial kits using

derivatives of the Eberwine (1996) method claim to quantitatively amplify message by up to a millionfold, but are very expensive. Direct labeling is the cheapest methodfor spotted arrays and involves directly incorporating a florescent label conjugated to a nucleotide during polymerization of the complementary strand. The increased size of these synthetic nucleotides causes an unavoidable decrease in efficiency in the labeling reaction. Alternatively, indirect labeling involves the incorporation of an aminoallyl-modified nucleotide in the initial step followed by a second step involving the chemical addition of a fluorescent dye ester to the aminoallyl-modified nucleotide. The structural similarity of aminoallyl nucleotides to normal nucleotides circumvents the lower labeling efficiency related to direct labeling and generally gives stronger signal strength. We therefore recommend indirect labeling despite the fact that it is a more expensive method.

The precise hybridization conditions will also need to be optimized for each microarray. The optimum temperature for hybridization should be empirically tested with the temperature calculated by probe design software serving as an initial guide. Decreasing or increasing hybridization temperature (leading to a respective increase and decrease in crosshybridization potential) can have a profound impact on the quality of the data. Volumes and sample buffer (typically 3× SSC, 0.1% SDS) can also be manipulated to optimize hybridization conditions. Once conditions have been optimized, they can be kept constant for a particular array and sample type. Indeed, it is essential that hybridization conditions are kept identical within a particular experiment. If different sample types are to be used with the same array, however, it may be necessary to change hybridization conditions between experiments. For example, for environmental samples you may wish to decrease hybridization temperature to maximize signal, or alternatively increase temperature to increase specificity of signal. It is essential for researchers to realize, however, that such differences in conditions will prevent a direct comparison between experiments.

### Image acquisition and quantification

Once the sample has been hybridized, microarray signal intensities are collected via image acquisition with a microarray scanner. Software is then used to visualize the image, find features (i.e., the spots), and quantify the signal in each feature. The scanner to be used for microarray image capture depends on your platform and the available infrastructure. Affymetrix microarrays require a specialized Affymetrix scanner, and scanning is generally carried out at an array facility. For other platforms, scanning may be done with a variety of scanners in an automated, highthroughput fashion using standard settings for laser power, image position, and pixel size or on a slide-by-slide basis with parameters optimized for each microarray. If you are purchasing a scanner, your choice will depend on factors such as the resolution required (between 5 and 10 µm is usual for standard spotted microarrays), the number of microarrays you intend to analyze, and the likelihood of requiring more than the two lasers. Excellent scanners commonly used include the Axon GenePix series, PerkinElmer ScanArray series, and Agilent scanners. Image analysis software is generally supplied with the scanner, which ensures that the image is in the correct format for processing. For example, the GenePix series use GenePix Pro software, whereas PerkinElmer suggests ScanArray Express for their ScanArray series. In principle, however, any scanner can generate images which can be assessed using any microarray image analysis software. Other suitable microarray software packages include BlueGnome, ArrayVision (GE Healthcare), and ImaGene (BioDiscovery), all of which perform well. We therefore recommend that you try out a number of different software packages (demo formats can generally be downloaded from the Web) and choose according to your technical requirements and ease of use.

### Data processing

Microarray data are inherently noisy and require careful processing before statistical analysis. Pre-analysis steps include quality control, background correction, and normalization. If you are working on a system with more than one probe per gene (e.g., Affymetrix), there will also be a summarization step, where a summary measure for a gene is generated from the multiple probes representing that gene. A good overview of the steps involved in preprocessing microarray data are given in chapter 1 of the 2005 Bioconductor book (Huber et al. 2005). Here we outline basic considerations, but direct the reader to other chapters of the same book that cover these topics in greater detail.

Quality assessment of the microarray data can include a variety of methods, for example, looking at regenerated images of background signal to assess spatial irregularities, plots of log signal value compared to intensity pre- and postnormalization (so-called MA plots), and box plots of slide data pre- and postnormalization and assessing spot quality flag information where this has been supplied by the image capture software. Hierarchical clustering (see "Data Analysis" below) of data pre- and postnormalization is very useful to look for whether the data are grouping according to non-interesting factors such as which day an analysis was carried out. For quality assessment methods for Affymetrix arrays, we also direct the reader to a white paper on the Affymetrix Web site (www.affymetrix.com/support/technical/whitepapers/exon_ arrays_qa_whitepaper.pdf). Image capture software and analysis software manuals usually include some information on quality assessment and control. Of note is that different software programs often tag spots with indicators of measurement quality, and different analysis software will provide different ways of dealing with these tags. We tend to work only with high-quality spots, as microarray data are noisy enough without including lower-quality spots, even if downweighted, for analysis purposes.

Background correction methods aim to remove background signal from spot signal measurements. Background correction is applied to data before normalization, although it is not always obvious, as it may be included in a software choice that encompasses a number of steps, such as background correction, normalization, and summarization, under a single command.

Normalization describes a variety of methods to correct microarray data for variation introduced by experimental procedures rather than biological differences between samples. For example, differences in detection efficiency, dye labeling, fluorescence yields, and total amount of cRNA/cDNA loaded onto the array can affect the measured signal intensities. If neglected, these factors can be erroneously interpreted as changes in expression. It is therefore important to correct for such variability before further data analyses. Choosing an appropriate normalization method is a crucial step, since it has considerable influence on the results (Hoffman et al. 2002). Normalizations can be applied to data within each array, to account for intra-array issues such as dye bias, location-dependent bias, and intensity-dependent bias. They can also be applied across arrays, usually referred to as between-array normalization, to address scale differences between arrays. The aim here is to make data comparable across arrays. For example, the popular quantile normalization (Bolstad et al. 2003) shifts the distributions of signals on arrays, resulting in the same empirical distribution across arrays and across channels.

Depending on the type of microarray platform used, different normalization schemes can be applied, with within-array normalizations being applied before between-array normalizations. For one-color microarrays, the first data processing step is the calculation of the summary indices for each gene based on the corresponding probes. Two methods are widely used for this task: Microarray Suite (MAS)/GeneChip Operating Software (GCOS) by Affymetrix and Robust Multi-array Average (RMA) introduced by Irizarry and coworkers (Irizarry et al. 2003). MAS/GCOS calculates the summary indices by averaging probe signals for each array individually. In contrast, RMA simultaneously calculates

summary indices for all arrays included in the experiment. Because it incorporates both probe and array effects in the calculation, it can correct for systematic difference in signal intensities between arrays and thus provides a first level of normalization. Note that this implicitly assumes that total (logged) expression intensity should be equal for the different arrays. The obtained distributions of summary indices reflecting the expression levels can be further adjusted subsequently. For MAS, scaling of the median expression to a chosen level is usually performed, if we can assume that the majority of measured genes are not differentially expressed. A common additional adjustment for RMA-processed data are quantile normalization transforming expression levels to have the same distribution for different arrays (Bolstad et al. 2003).

Several comparisons have been conducted so far between MAS and RMA processed data, but the results remain inconclusive. For example, RMA performed favorably for a benchmark dataset with a small number of spike-in controls, for which the concentrations were known (Cope et al. 2004). In another study, where a considerably larger number of spike controls was used, MAS outperformed RMA (Choe et al. 2005). The latter study, however, has been criticized for the use of a flawed design (Dabney and Storey 2006, Irizarry et al. 2006).

These contrasting results for different datasets indicate that the performance of normalization procedures can strongly depend on the dataset being processed. This is not surprising, since all of the normalization procedures are based on specific assumptions which may not hold for different datasets. When assumptions are violated, normalization might fail and can lead to erroneous results. Thus, it is of vital importance that researchers carefully check the suitability of assumptions.

To normalize two-color arrays, the signal intensities of the cohybridized samples are used. The most basic method is global normalization, which is the linear scaling of the total intensities in each channel to the same value. Nonlinear effects in the signal intensities are frequently observed, however—a phenomenon referred to as dye bias (see earlier discussion of dye swaps). This kind of artifact causes signals to be systematically larger in one channel for the low-intensity range even after balancing the average signal intensities of both channels. To cope with such bias, several intensity-dependent normalization procedures have been introduced. Some of these elaborate methods can cope with potential spatial artifacts (Yang et al. 2002, Yang and Speed 2002, Futschik and Crompton 2004). It is important to emphasize, however, that such nonlinear methods require that most of the genes assayed are nondifferentially expressed or that the differential expression is symmetrical, i.e., the overall up- and downregulation is balanced. This is not always the case for virus-derived samples—for example, if overall host genome expression declines during viral infection or if the array is dominated by virus probes only.

A solution is to scale the data to signal from reference probes that are kept constant across the experiment. The employed reference can be of either endogenous (e.g., so-called housekeeping genes) or exogenous (i.e., spike-in controls) origin. Normalization based on housekeeping genes predates most other normalization procedures for two-color arrays (DeRisi et al. 1996) and assumes that the genes chosen really do not change under the experimental conditions. If you cannot guarantee a set of expressed but nonchanging genes, we recommend you use spike-in controls. Validation, for example using quantitative RT-PCR, can be used to select an optimal normalization procedure (Lindell et al. 2007). Reference-based normalization methods seem intuitively very attractive, but careful checks are necessary to ensure they are working as desired. We have performed comparisons of several normalization schemes for microarray experiments for virus-infected *Prochlorococcus* cells, for which we expected that the majority of assayed genes underwent differential expression (Lindell et al. 2007). The microarrays were customized Affymetrix GeneChips

including spike-in hybridization controls. Neither the normalization by spike-in control nor by rRNA selected as housekeeping genes yielded superior results compared with other normalization strategies when using qRT-PCR data as the gold standard in an independent comparison (Lindell et al. unpubl. results). Possible reasons for the inferior performance of spike-in controls could be their limited number and their restricted intensity range on Affymetrix GeneChips. In the case of normalization based on rRNA, their high expression levels and their possible saturation on the array might cause difficulties in their use as reference. Having said this, we have had satisfactory results using reference-based normalizations on two-color arrays. As with all aspects of microarray experimental planning, nothing should be taken for granted. The use of spike-in controls combined with external checks of representative genes (e.g., quantitative RT-PCR), and plotting of data and controls pre- and postnormalization, are necessary to have confidence in your microarray results.

### Data analysis

Clustering is a popular approach to explore large microarray data sets. The aim of clustering is to assign genes or arrays to groups based on their similarity: genes or arrays displaying similar expression profiles, where you define what *similar* means by choosing an appropriate dissimilarity measure, should be assigned to the same clusters, whereas genes or arrays displaying distinct expression profiles should be placed in different clusters. Using cluster analysis, we can detect prominent expression patterns, coexpressed genes, and similar conditions, which can be further examined for their biological meaning. Many different clustering methods have been applied to microarray data. Generally, two types of clustering exists: hierarchical and partitional (Jain and Dubes 1988). Hierarchical clustering creates a set of nested clusters, so that clusters on a higher level are composed of smaller clusters on lower levels. The resulting hierarchy of clusters are conventionally presented as a treelike structure, the so-called dendrogram. To perform hierarchical clustering, we proceed in a sequential manner. In each step, we calculate the pairwise distances between all clusters and merge the ones with the smallest distance. In contrast, in partitional clustering, all objects are simultaneously assigned to clusters. This type of clustering typically aims to optimize an objective function for a given number of clusters. A prime example of this clustering approach is k-means clustering, which seeks to minimize the between-cluster variation in an iterative manner. Hierarchical and partitional clustering both have their advantages and disadvantages. One strength of hierarchical clustering is that it defines relations between and within clusters. However, the sequential procedure used can be sensitive to the high noise level that is frequently contained in microarray data. Partitional clustering tends to be more robust to noise, but commonly fails to present within-cluster structures. Notably, some partitional clustering methods can reveal internal cluster structures and are still highly noise-robust. For instance, fuzzy clustering assigns genes graded membership to clusters, i.e., it can indicate how strongly a gene is associated with a cluster. Thus, genes that are tightly clustered obtain a large membership (with values close to 1), whereas genes with noisy expression patterns receive low membership values. In contrast to conventional clustering methods, fuzzy clustering even allows genes to be placed in several clusters (Futschik and Crompton 2004). A variety of software packages have been developed for clustering analysis of microarray data.
Popular standalone software packages for performing and visualizing hierarchical clustering are Cluster 3.0 and Java TreeView, respectively (bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster; jtreeview.sourceforge.net). Alternatively, several Web servers enable online cluster analyses (e.g., EBI expression profiler, www.ebi.ac.uk/expressionprofiler).

A difficult question is how many clusters can be reliably retrieved from the observed expression data. The difficulty arises from the complexity of microarray data and a high noise component. Frequently, different cluster structures are apparent depending on the resolution. For example, several main

clusters may exist, but each might display subclusters. Furthermore, the noise component can lead to overlapping clusters, for which a separation might be not justified. Despite these difficulties, some tools have been developed to help researchers choose the accurate number of clusters and judge the reliability of the results. Classic approaches are based on the so-called figures of merit. These measures capture a desired feature that we seek to optimize. To obtain an accurate clustering, we seek to optimize the figure of merit. An example is the Dunn index, defined as the ratio between the minimal intracluster and the maximal intercluster distance (Dunn 1974). Used as a figure of merit, we aim to minimize the Dunn index to obtain tight clusters that are well separated. A common drawback, however, is that these measures assume non-overlapping clusters, which are typically not the case for microarray data. An alternative approach is based on measuring the stability of clusters with respect to data perturbations, e.g., through resampling or addition of noise. According to this concept, reliable clusters are those that are maintained in spite of perturbation (Bittner et al. 2000, Levine and Domany 2001). For instance, we could cluster genes using only a subset of measurements and examine the resulting clusters. As reliable clusters should not depend on single measurements, they should still be detectable using partial data. Finally, the inspection of the functional composition of clusters can give us clues about reliability. This strategy assumes that genes sharing the same function tend to be coexpressed and thus should be placed into the same cluster. By optimizing the enrichment of functional gene categories, the number of clusters can be chosen (Gibbons and Roth 2002).

Despite these tools, assessing the quality of clusters remains challenging. Researchers are advised to apply several clustering approaches to their datasets, as a single method often works well with some data sets, but may perform poorly in others. In practice, we propose that clustering should be seen primarily as exploratory analysis that can then be followed up with more stringent computational and experimental examination. A good introduction to clustering as applied to microarray analysis is given in chapter 7 of Wit and McClure (2004).

### Statistical significance of differential expression

A typical aim of microarray experiments is to identify genes that are differentially expressed under different conditions. Because of the large number of genes measured by microarray technologies and random fluctuations in gene expression, we expect a number of genes to show differences in expression simply by chance. Therefore, we use stringent statistical approaches to analyze microarray data for differential expression. In classic statistical testing, we compare a test statistic calculated from our data to a distribution of that test statistic expected under the *null hypothesis* that the gene is not being differentially expressed. The comparison of our test statistic to this distribution results in a *P* value. *P* values indicate how often you would expect to see data as extreme as that you just observed if the gene is not being differentially expressed. P values are not direct indicators of the probability of the gene being or not being differentially expressed. For differential gene expression studies, a small *P* value (e.g., <0.01 for a single test, see below) indicates that we would rarely see a test statistic that extreme if the gene measurements in one condition really were from the same distribution of expressions as the condition we are comparing them to. For example, with $P = 0.01$, we would expect to see values this extreme in about 1 of every 100 samplings if the data we are working with were sampled from the same distribution as the one we are comparing it to. Another school of statistics, the Bayesian school, provides more readily interpretable probabilities, but can be harder to apply in practice. *P* values are closely related to a more readily interpretable value, the false discovery rate (Benjamini and Hochberg 1995), which is commonly used to evaluate microarray results and is discussed further below. For a good introduction to all the aspects of statistical testing for differential expression of microarray data mentioned here, we recommend chapter 8 of Wit and

McClure (2004).

There are two general categories of statistical tests:
1. Parametric tests. These assume that the populations being compared can be described by particular distributions. For instance, certain tests assume that the underlying distribution is normal. For microarray analyses, the distribution being referred to is the distribution of expression values for a given gene; it is not about the distribution of expression values across genes. Typically, having sufficient biological replicates for parametric statistical analysis is a challenge. Some available statistical methods, so-called *local pooled error* methods, aim to decrease the number of replicates required to carry out reliable statistical tests by pooling the sample variation of genes with similar expression intensity; i.e., they fit the error with respect to the signal intensity based on the observed data (Baldi and Long 2001, Tusher et al. 2001, Jain et al. 2003, Smyth 2004).
2. Nonparametric tests. These do not make assumptions about the underlying distribution. Such tests commonly rank the data in value order and then carry out tests based on the order. This is very useful when we do not know the underlying distribution of what we are measuring. However, nonparametric tests have less power to detect differences and thus require more replicates to give similar confidence when interpreting your data.

Parametric tests involve comparing the test statistics calculated using your data to a standard distribution with particular parameters. If you have sufficient biological replication in your experimental design, you can instead compare your test statistic to the expected null distribution of that statistic (i.e., when a gene is not differentially expressed), which you generate through bootstrapping, or through permutation, of your own data set. For example, instead of comparing a *t*-test statistic to a standard Student *t* distribution to determine a *P* value, you would generate a distribution of the test statistic from resampling your own data set in defined ways and compare your test statistic to that distribution. Issues relating to bootstrapping and permutation are discussed in Wit and McClure (2004).

Until this point, we have really been discussing testing for differential expression of a single gene between two conditions. For a comparison yielding a P value of 0.01, we would expect to see data this extreme in about 1 of every 100 tests if the gene were not differentially expressed. So, if we test 10,000 genes, and 10 genes are really different between conditions, we could still expect around 110 genes with P values less than or equal to 0.01. Of these, 100 had "extreme" mean expression values due to chance, with 10 of them truly differentially expressed between conditions. We need to try and increase our ability to discern genes that are truly differentially expressed. In other words, we need to adjust our results to take into account that we are carrying out multiple testing. Different types of multiple testing corrections are used in microarray studies (Dudoit et al. 2003), but arguably the most popular is the false discovery rate (FDR) (Benjamini and Hochberg 1995). The FDR method allows you to define the proportion of false positives you would find tolerable in your results. It then returns the largest list of genes classified as differentially expressed that includes this specified, expected percentage of nondifferentially expressed genes. Numerous software tools can help researchers to assess the significance of differential expression. Notably, a highly powerful and flexible platform, also for other aspects of microarray data analysis, is the Bioconductor project (www.bioconductor.org). Alternatively, more specialized software solutions such as SAM (www-stat.stanford.edu/~tibs/*SAM*) or BRB Array Tools (linus.nci.nih.gov/BRB-ArrayTools.html) can be applied to calculate false discovery rates for gene expression data.

***Experiment annotation and data submission***

Many journals require microarray experimental data to be submitted to a public repository as a condition of publication. Indeed, some funding agencies require researchers to agree to make their data publicly available at the end of the project; for many researchers, the most sensible way to do this will be by submitting to a known public repository such as the EBI's ArrayExpress (www.ebi.ac.uk/arrayexpress) or the NCBI's GEO (www.ncbi.nlm.nih.gov/geo). Both these databases require adequate annotation of the experiment, including at least the information required by the Minimum Information about a Microarray Experiment (MIAME) standard (Brazma et al. 2001). For researchers engaging in environmental experiments, it is worth also referring to the MIAME Env extension (Morrison et al. 2006).

The importance of annotating data properly and making it publicly available has led to many new "minimum information" checklists for different domains. The Minimum Information for Biological and Biomedical Investigations (MIBBI) portal (http://mibbi.sourceforge.net) is a good place to look for standards lists. Data standards list *what* the minimum requirements are for describing a data set. How the data should be described is usually defined by a set of terms or an ontology. For key microarray experimental concepts, the Microarray Gene Expression Data Society (MGED) make the MGED Ontology available. Use of this ontology, sometimes in combination with other domain-specific ontologies, is highly recommended when annotating your experiments. The Open Biomedical Ontologies (OBO) consortium (Smith et al. 2007) Web site (http://obofoundry.org) is probably the best place to see if ontologies relevant to your area already exist or are under development.

Public data repositories offer tools to facilitate submission of data, and some external tools support export in a format acceptable to the public repositories. A "non-exhaustive list of possible MIAME compliant software" is held on the MGED site at http://www.mged.org/Workgroups/MIAME/miame_software.html. For researchers in the environmental sciences, the software maxdLoad2 supports annotation using the MIAME Env extension of the MIAME standard (Hancock et al. 2005).

### Summary

Microarrays offer great potential, but the statistical analysis needs careful consideration from the outset. Our general recommendations are as follows:

1. If the samples required to address a particular question are too difficult to generate or collect, you should adjust the question you are asking. Your data will not miraculously provide answers to the original question if you do not have sufficient or appropriate samples.
2. If you are not experienced with statistics, then find a collaborator who is. Access to shiny software with easy-to-use menus is not the same thing as statistical knowledge.
3. Know what software you (or your collaborator) are going to use for the analysis and be sure that it is capable of analyzing data generated under a particular design. It is also a good idea to define how technical replicate spots and technical replicate slides will be handled if these are part of your design.
4. Technical replicates provide a different type of information than biological replicates and can be difficult to handle appropriately using some software. For many purposes, more biological replicates is a better option than carrying out technical replicate hybridizations.
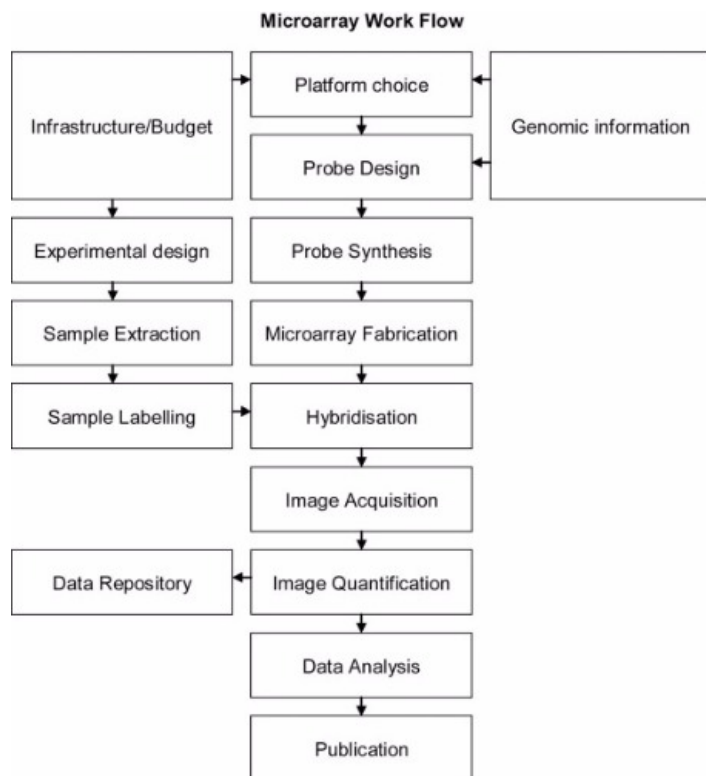
### Figures and Tables

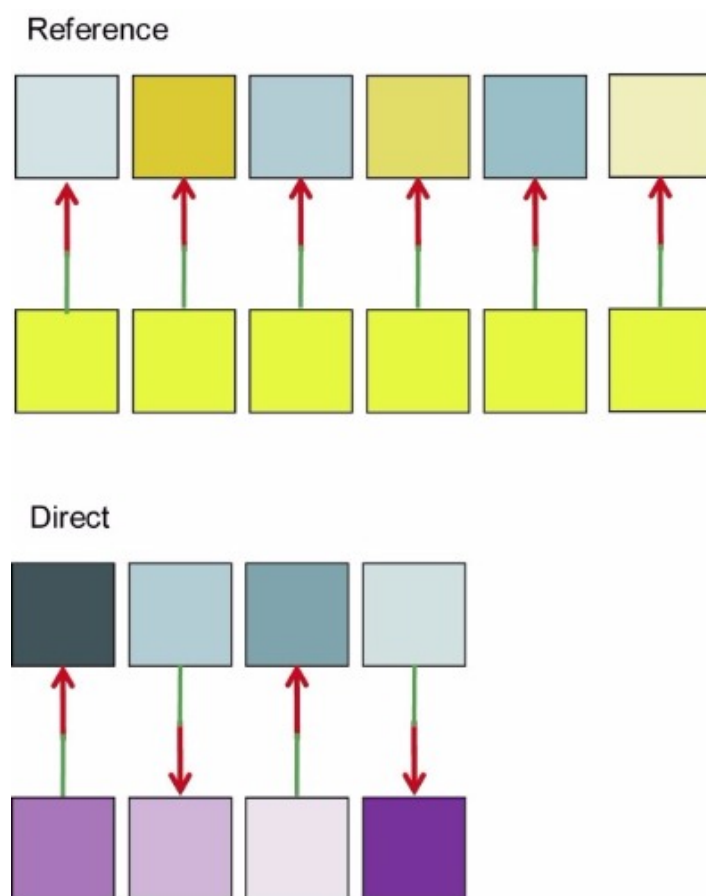**Fig. 1.** Microarray work flow.



**Fig 2.** Examples of reference and direct designs. Samples connected by an arrow are hybridized to the same array. Dye color can be inferred by the direction of the arrow (arrow head, Cy5 dye; tail, Cy3

dye). Here, we have also represented this by coloring the head of the arrow in red and the tail of the array in green. The reference design is shown with two experimental sample types (orange and blue), one reference sample (yellow), three biological replicates (shades of orange and blue), and no dye swaps (arrow directions) with a total of six slides. The direct design is shown with two experimental sample types (blue and purple), four biological replicates (shades of blue and purple), two pairs of dye swaps (arrow directions) with a total of four slides.
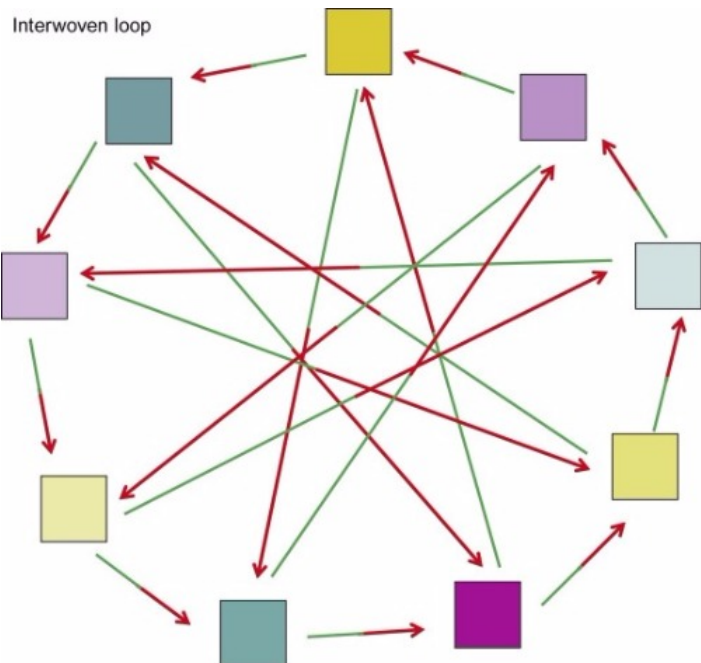


**Fig 3.** Example of an interwoven loop design shown with three experimental sample types (blue, orange, and purple), three biological replicates (shades of blue, orange, and purple), each sample hybridized the same number of times and with both Cy3 and Cy5 (also the same number of times), with a total of 27 slides.

**Table 1.** Commercial suppliers of high-density microarrays.

| Company | Web site | Description |
| --- | --- | --- |
| Affymetrix | http://www.affymetrix.com/estore/ | High-density chips can be designed with up to 1.3 million features, with 25mer probes. Features are 11 µm in size. |
| Agilent | www.chem.agilent.com | High-density arrays with up to 243,504 features per array printed on standard glass slides. Features are 65 µm in size. Standard probe length is a 60mer, but any length between 25 and 60 bases possible. |
| CombiMatrix | www.combimatrix.com | Array chips featuring 12,000 35-40mer probes, feature size 44 µm. |
| Nimblegen | www.nimblegen.com | High-density arrays with up to 2.1 million features, 50–75mer probes printed on standard glass slides. Features are 13 µm in size. |

*Full paper available at [http://www.aslo.org/books/mave/MAVE_034.pdf](http://www.aslo.org/books/mave/MAVE_034.pdf)*

## Protocol