Guidelines to perform a successful microarray experiment

This document tries to answer some of the basic questions to be considered when planning a microarray experiment.

Table of contents

Guidelines to perform a successful microarray experiment..........................................................1
1. Are microarrays appropriate to achieve our goals?.................................................................1
2. Carefully plan your experiment..............................................................................................2
3. What kind of microarrays should I use?..................................................................................2
5. What should I provide UCTS and UEB with?........................................................................2
6. Data Analysis..........................................................................................................................3
   6.1. Do I need statistical advice for my study design?..............................................................3
   6.2. Who should analyze the data............................................................................................3
7. What kind of results do I get after study?................................................................................4
Appendices.....................................................................................................................................5
   Appendix I: Templates to provide information.......................................................................5
      1. Information about the experiment...................................................................................5
      2. Allocation of each sample to each experimental condition..............................................6
   Appendix II: Considerations for a good experimental design...............................................7
References........................................................................................................................................9

0. Talk to UCTS and UEB

Figure 1. Working flow chart to design an appropriate microarray experiment.

1. Are microarrays appropriate to achieve our goals?

   If …
   • you are interested in few candidate genes ⇒ RT-QPCR or LDA may be better options
   • you wish to do gene fishing ⇒ microarrays may be appropriate.
      Small number of replicates (e.g 4/condition) may be enough
   • you wish to build predictors ⇒ microarrays may be appropriate.
      High number of replicates (e.g. 30/condition) may be necessary.
2. Carefully plan your experiment

- Keep things simple (pair-wise comparisons, simple time-courses).
- DO NOT perform studies in different batches
- Use Balanced design: An experimental design where all cells (i.e. treatment combinations or effects to study) have the same number of observations.
- Try to avoid confounding design: An experiment where the value of a main effect estimate comes from both the main effect itself and also contamination or bias from some interactions (i.e: treatment; sex...)

3. What kind of microarrays should I use?

- If you are interested in gene expression UCTS recommends to use Affymetrix platform.
- You may choose between
  
  - **3’ Gene Expression**: Affymetrix 3’ expression arrays interrogate each gene with multiple independent probes located at 3’-End. This array allows you to investigate gene level expression (see Figure 2). This array is available in many species [1].
  - **Exon arrays and Gene arrays**: These arrays measure gene expression at the transcript and exon level. Combining the unbiased whole-transcript assay with probes interrogating the entire length of the gene, GeneChip® Exon [2] and Gene Arrays [3] will allow you to detect transcriptional events not possible with 3’-biased arrays (see Figure 2). Only available in human, mouse and rat species.

![Figure 2](image)

**Figure 2.** Probe distribution through the genes in the three types of arrays. Panel a) 3’Gene Arrays, panel b) Gene Arrays, panel c) Exon Arrays.

5. What should I provide UCTS and UEB with?

1. Clearly state your primary and secondary objectives. You may use the template in Appendix I.1.
2. Clearly state your experimental design. Which factors are considered, what type of experiment is it, ...You may use the template in Appendix I.1.
3. Provide an appropriate quantity (usually at least 50 picograms are needed) of good quality RNA (RIN of at least 7 is recommended).
4. Provide the sample allocation to experimental conditions. Use the template in Appendix I.2 to indicate the name of the sample, the conditions (factors) to which it belongs, and other covariates you consider important.

There are many issues to account for in the design of your microarray experiment. Appendix II lists some aspects that you must consider.
6. Data Analysis

6.1. Do I need statistical advice for my study design?

Microarray experiments are biological experiments, so the most important considerations will be biological. It will, however, be very useful to involve UEB in the design, analysis and interpretation process to prevent some common pitfalls, such as underestimating the “multiple testing problem” involved in examining thousands of genes at once. And of course you should always be aware of the basic statistical and philosophical issues involved in any successful experimental design.

6.2. How are the data analyzed?

- It is recommended that data analysis is performed by experienced statisticians such as the UEB team.
- However, if you prefer to do it by yourself, commercial software (Partek and Ingenuity), and public software (UEB on-line resources) are available.

![Figure 3. Outline of the different steps for the analysis of microarrays](image-url)
7. What kind of results do I get after study?

- **UCTS** delivers a DVD with raw data (binary files).
  
  **WARNING**: This is the data source that belongs to you, but it cannot be read with standard office-like software.

- **UEB** delivers a results report with tables of list(s) of genes, plots, biological significance,...
Appendices

Appendix I: Templates to provide information

1. Information about the experiment

<table>
<thead>
<tr>
<th>Study Proposal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Contact Information</strong></td>
</tr>
<tr>
<td>Name</td>
</tr>
<tr>
<td>Lab/Department</td>
</tr>
<tr>
<td>Institution</td>
</tr>
<tr>
<td>Telf</td>
</tr>
<tr>
<td>e-mail</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>General Information of the Study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General goals</strong></td>
</tr>
<tr>
<td>(Write in short an overview of your research)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Main objectives</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>(Describe simply and clearly your/s objectives)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Specific objectives</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>(Based on your 'Main objectives', write down accurately what are the comparisons that you are interested in)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Material</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sample Information</strong></td>
</tr>
<tr>
<td><strong>Organism</strong></td>
</tr>
<tr>
<td><strong>Sample types</strong></td>
</tr>
<tr>
<td>(patients, animals, tissues, cell lines,...)</td>
</tr>
<tr>
<td><strong>Sample size</strong></td>
</tr>
<tr>
<td>(number of chips, sequences,...)</td>
</tr>
</tbody>
</table>
2. Allocation of each sample to each experimental condition

<table>
<thead>
<tr>
<th>Sample Number</th>
<th>Sample Name</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>…</th>
<th>Factor k</th>
</tr>
</thead>
</table>
Appendix II: Considerations for a good experimental design

How many replicates will I need?

As many as possible! For an exploratory analysis 3 or 4 replicates are usually sufficient, unless the data are particularly noisy (e.g. samples from very small numbers of cells) or the expected effect is particularly small (e.g. changes occur only in very few, specialized cells in the sample).

Using less than 4 replicates is NOT A GOOD IDEA.

Most important are: (1) the use of biological replicates and (2) to do equal numbers of replicates for each condition/comparison to keep the later analysis simple.

What is a biological replicate?

Biological replicates reflect true variation among experimental units (i.e. the use of several cell lines, mice, rats, tissue samples, patients,...).

A well designed experiment is fundamental to obtain reliable data, which allow to assess differential expression of a set of genes under investigation in an adequate manner. The use of biological replicates plays a crucial role in any statistical analysis to obtain inference that could be generalized to a biological population.

![Figure 4. Obtaining good replicates to be analyzed.](image)

In the absence of biological replicates, the experiment and the interpretation of its results are confined to the single subjects under investigation.

With regard to cell lines, they are equivalents treatments, generally speaking. Each cell line should be replicated.

Should I do technical replicates?

NO, unless you are planning a technical instead of a biological study. Repeated hybridization of the same biological sample is a waste of resources.

Technical replicates DO NOT PROVIDE ANY BIOLOGICALLY USEFUL INFORMATION.
Should I pool my samples?

It is tempting to pool samples to save hybridization costs.

Try to avoid pooling samples a priori. Unless you expect to find interesting inter-individual variations there is little to argue against pooling.

Pooling samples could be appropriate in comparative studies, but is highly recommended to ask UEB about it.

Pools should be well-performed. It is IMPORTANT to pool the biological material (tissue, cells), NOT the purified RNA or labeled cDNA!

How should I provide RNA samples?

Preparation of samples, labeling, hybridization, and other steps of microarray experiment can contribute to technical variation, which can significantly impact the quality of array data.

To ensure highly reproducible microarray data, technical variation should be minimized by controlling the quality of the RNA samples and by efficient labeling.
References