LECTURE 3
MICROARRAY
HYBRIDIZATION
David Galbraith
Graduate level course Methods in Cell Biology and Genomics (PLS539)

Much information concerning microarrays!

http://cals.arizona.edu/classes/pls539/
Microarray Target Production

- Using small fluorochromes: Cy3, Cy5, others.
  - Direct Incorporation.
  - Indirect Incorporation.
- Signal Amplification without RNA Amplification.
  - Dendrimer technologies (Genisphere).
  - RLS particles (Invitrogen).
- Amplification of RNA prior to Target Production.
Direct Labeling

Advantages.
- Simple procedure.
- Works well.

Disadvantages.
- Cost.
- Limited to Cy3 and Cy5.
- Dye-specific effects during incorporation.
- Inhibition of RT processivity.
Absence of dye effects during direct incorporation: mutant vs. wild-type

Major factor seems to be the source of the RT
Indirect Labeling

1st strand cDNA synthesis → Aminoallyl dXTPs → Dye incorporation → Fluorescent nucleophilic reagents

Advantages.
• Cost.
• Dye flexibility.

Disadvantages.
• Clean-up.
RNA input levels required without amplification.

- We require about 1-2 $\mu$g of polyA+ RNA for direct or indirect incorporation methods.

- Using long oligo microarrays, we need to employ random priming, since probes are not necessarily at the 3’-ends of the genes.

- This requires about 100 $\mu$g of total RNA.

- If this amount is unavailable, amplification techniques must be employed.
The Need for Amplification

RNA amounts relative to cell numbers and amplification levels.

1. Arabidopsis RNA content: \(~1\) pg/cell
2. 1 \(\mu\)g total RNA \(~1 \times 10^6\) cells
   \(~1 \times 10^7\) nuclei
3. 2-rounds of RNA amplification: \(~1 \times 10^5\) fold
Target Amplification Techniques

- Eberwine-based.
- PCR-based.
- Strand Displacement-based.
- Other.
Amplification Technique Evaluation
Criteria

• Sensitivity.
• Strand orientation.
• Coverage (genome-wise and gene-wise).
• Linearity.
• Reproducibility.
• Limiting factors.
• cRNA or cDNA targets?
RNA Amplification - Eberwine Method

First Cycle

5’---------------------AAAAA 3’
3’ TTTTT-- T7 5’

1st strand cDNA synthesis

Second Cycle

5’NNN 3’ 5’NNN 3’
3’---------------------UUUUU 5’

2nd strand cDNA synthesis

5’NNN----------NNN------AAAAA 3’
3’---------------------UUUUU 5’
3’T TTTT-- T7 5’

in vitro transcription

Problem of 5’- truncation
RNA Amplification – Eberwine-based Strand Reversal

First Cycle

1st strand cDNA synthesis

Second Cycle

2nd strand cDNA synthesis

in vitro transcription

Devised by F-C Gong; avoids truncation
Indirect Labeling with One-Round of Amplification

1\textsuperscript{st} and 2\textsuperscript{nd} strand cDNA synthesis

\textit{In vitro} transcription

Aminoallyl UTP

Dye incorporation

Fluorescent nucleophilic reagents

RNA-DNA hybridization
Target Amplification Techniques

ExpressArt® -- micro-, nano-, and pico-kits.

- Conversion of mRNA to cDNA using an anchored oligo-dT primer during first RT reaction. Removal of all RNAs using heat-labile RNAases.

- Conversion of ss cDNAs to ds DNAs using a special 30-mer primer construct. This contains a unique 21-mer sequence, followed by six random nucleotides and a fixed trinucleotide sequence at the 3’ end. A mix of several primers is used with different 3’-terminal trinucleotides. These determine the potential primer elongation sites, discontinuously distributed over the annealed templates.
Target Amplification Techniques

ExpressArt®

- Conversion of the ds cDNA to ds cDNA containing a T7 promoter site.
- Amplification by *in vitro* transcription.
- Process can be repeated once more (nano) or twice more (pico).
- Advantages are (a) discrimination against rRNA (based on primer selection), (b) limited loss of size of products), and (c) high fidelity and amplification.
Expected yields and length distribution of amplified RNAs

Please note that mRNA content varies with sample type, ranging between 1 and 5% of total RNA. Data presented were obtained with total RNA isolated from cultured cells (about 3% mRNA).

<table>
<thead>
<tr>
<th>Pico Version</th>
<th>1&lt;sup&gt;st&lt;/sup&gt; aRNA</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt; aRNA</th>
<th>3&lt;sup&gt;rd&lt;/sup&gt; aRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input total RNA</td>
<td>100 pg</td>
<td>not applicable</td>
<td>5 ± 2 µg</td>
</tr>
<tr>
<td>1 pg</td>
<td>not applicable</td>
<td>5 ± 2 µg</td>
<td>&gt; 50 µg</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nano Version</th>
<th>1 ng</th>
<th>5 ± 2 µg</th>
<th>&gt; 50 µg</th>
</tr>
</thead>
</table>

| Nano version or Nano plus Version | 1 ng | 5 ± 2 µg | > 50 µg |

<table>
<thead>
<tr>
<th>Micro Version</th>
<th>1&lt;sup&gt;st&lt;/sup&gt; round amplified RNA</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt; round amplified RNA</th>
<th>3&lt;sup&gt;rd&lt;/sup&gt; round amplified RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input total RNA</td>
<td>Yield with 4 h in vitro transcription</td>
<td>Yield with 16 h in vitro transcription</td>
<td>not recommended</td>
</tr>
<tr>
<td>500 ng</td>
<td>5 ± 2 µg</td>
<td>10 ± 4 µg</td>
<td></td>
</tr>
<tr>
<td>3 µg</td>
<td>20 ± 8 µg</td>
<td>50 ± 16 µg</td>
<td></td>
</tr>
<tr>
<td>5 µg</td>
<td>45 ± 10 µg</td>
<td>not recommended</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example Electropherograms of ExpressArt® amplified mRNAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA profiles were obtained with the Agilent Bioanalyzer</td>
</tr>
</tbody>
</table>

Contact:
Bob Boehmer (Qon-teq, LLC) -- bobboehmer@qon-teq.com
Conventional Eberwine ExpressART amplification
2-rounds (100ng) vs. 2-rounds (10ng)  
2-rounds (100ng) vs 3-rounds (5ng)

For more details of this kit, talk to Roger Barthelson
Other advantages of RNA amplification

• Amplification produces large amounts of RNA, which means that hybridizations can be repeated several times (RNA is no longer limiting).

• RNA-DNA duplexes form with greater affinity than DNA-DNA duplexes.

• The amplification process serves to purify the RNA from contaminants, giving a cleaner hybridization.

• Troubleshooting amplification is generally cheaper than troubleshooting labeling.
Other examples of situations requiring amplification

Linking RNA transcript accumulation and protein translation

• Cell type-specific analysis of polyribosomal RNA.
We are developing microarray-based methods that selectively examine global transcript abundances within polyribosomes in a cell type-specific manner.

To what extent does cytoplasmic transcript abundance relate to protein production?
Analysis of Global Gene Expression at the Level of Translation

- To what extent does transcript abundance (nuclear or cytoplasmic) relate to protein production?
- Can methods be devised to analyze global translational activity in a cell type-specific manner?
Global Cell Type-specific Translational Regulation

• Produce transgenic plants expressing ribosomal proteins that are epitope-tagged, under the control of constitutive or cell type-specific promoters.

• Isolate polyribosomes from whole tissue homogenates.

• Immunoprecipitate polysomes using anti-Tag antibodies.

• Isolate mRNA from polysomes and use for microarray hybridization.
Status

- Suitable ribosomal proteins and epitope tags have been identified (His$_6$-FLAG, and RPL18).
- Transgenic plants have been produced.
- Microarrays have been hybridized using targets amplified from immunoprecipitated polysomal RNA.
- Data analysis is underway.
(A) HF-RPL18 accumulates in ribosomes (P-170 fraction) of leaf extracts. Control - Arabidopsis transformed with the empty His6-FLAG (HF) vector. Immunodetection with FLAG-HRP conjugated-antibody (α-FLAG, 1:1000) and antiserum prepared against r-protein S6 (α-RPS6, 1:5000) or cytosolic GAPDH (α-GAPDH, 1:5000). (B) HF-RPL18 60S subunits and polyribosomes of sucrose density gradient fractionated cell extracts. (C) HF-RPL18 and RPL23a-FH complexes immunoaffinity purified with anti-FLAG M2 agarose resin (Eluate). (D) Coomassie blue stained proteins isolated by affinity purification or conventional centrifugation (P170). (E) Absorbance profile of complexes isolated by conventional centrifugation (P-170) and affinity purification of HF-RPL18 (Eluate). (F) RNA blot hybridization analysis of HF-RPL18 complexes. (G) RT-PCR analyses of high (RPS9) and low (PABP2, WRKY75) mRNAs from HF-RPL18 complexes. Comparison of total RNA (T) and affinity-purified RNA (E).
(A) HF-RPL18 accumulates in ribosomes (P-170 fraction) isolated from leaves.

Control - Arabidopsis transformed with the empty His6-FLAG (HF) vector. Immunodetection with FLAG-HRP conjugated-antibody (α-FLAG, 1:1000) and antiserum prepared against r-protein S6 (α-RPS6, 1:5000) or cytosolic GAPDH (α-GAPDH, 1:5000).
HF-RPL18 is found in subunits and polyribosomes of sucrose density gradient fractionated cell extracts.
<table>
<thead>
<tr>
<th>kDa</th>
<th>Control</th>
<th>HF-RPL18</th>
<th>RPL23a-FH</th>
<th>Control</th>
<th>HF-RPL18</th>
<th>RPL23a-FH</th>
<th>Control</th>
<th>HF-RPL18</th>
<th>RPL23a-FH</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(C) HF-RPL18 and RPL23a-FH complexes immunoaffinity purified with anti-FLAG M2 agarose resin (Eluate).
(E) Polysomes can be isolated by affinity chromatography.

Absorbance profile of complexes isolated by conventional centrifugation (P-170) and affinity purification of HF-RPL18 (Eluate).
RNA extracted from immunoprecipitated polysomes can be detected using the Bioanalyzer (C), and can be amplified and employed as microarray targets (D). Amplification is both linear and reproducible (E).
Availability of RNA relative to Experimental Design

How best to compare samples within the context of an experiment?

Microarrays are costly. What is the optimal design for microarray experiments?

- We want analysis of the data and interpretation of the results to be as simple and powerful as possible.
- Consider two-color experiments.
- Each microarray gives us the relative abundances of two sets of mRNAs.
- Between slide variation is greater than within slide variation.
- Biological variation is greater than technical variation.
Microarrays are costly. What is the optimal design for microarray experiments?

- The most important design issue is which mRNAs are to be labeled with which fluor, and which are to be hybridized together on the same slide.
- This includes determining the number of replications that will be done.
Graphical representation of microarray experiments

Convention: Nodes (square boxes) represent mRNA samples. Arrow head implies mRNA sample is labeled with Cy5; arrow tail implies mRNA samples is labeled with Cy3; number on arrow indicates number of replicates.

(a) 5 replicates of mRNA from state A (Cy3-labeled) and from state B (Cy5-labeled)

(b) Simplest loop design for three states.
Graphical representation of microarray experiments

Consider the accuracy of measurements depicted by Panel b.

- Vertices A and B are joined by two different paths. A path of length one directly links A and B; a path of length two links A with B via C.
- The direct ratio measurement $\log(A/B)$ is likely to be more accurate than the indirect $\log(A/C) - \log(C/B)$. 
Figure 1 | **Direct versus indirect designs.** Two possible designs that compare gene expression in two cell-population samples T and C.  

**a** | In a direct comparison, the differential expression of the genes in samples T and C is measured directly on the same slide (in a single experiment).  

**b** | In an indirect comparison, expression levels of samples T and C are measured separately on two different slides. The log ratio \( \log_2(T/C) \) is estimated by the difference \( \log_2(T/R) - \log_2(C/R) \).  

R, reference.
Direct versus Indirect Experiments

- Assuming mRNA is not limiting, the variance associated with measuring log(T/C) is $\sigma^2$, and if we were to average duplicate experiments, the variance would be $\sigma^2/2$.

- In contrast, if we use an indirect measurement with respect to a common reference (R), the variance of each measurement will be $\sigma^2$, and the variance of the difference $(\log(T/R) - \log(C/R))$ will be $2\sigma^2$.

- As a general rule, direct comparisons are more accurate than indirect comparisons.
Dye Swap Experiments

- For some experimenters, not all mRNAs are equally labeled by Cy3 and Cy5, with a small minority of mRNAs showing preference for one or the other dye.
- Dye swap experiments account for this bias. They can be done as direct comparisons, or in loop designs.
Box 3 | Issues that affect the design of array experiments

Scientific
• Aim of the experiment.
• Specific questions to be answered and how they are prioritized.
• How will the experiments answer the posed questions?

Practical (logistic)
• Types of mRNA samples: reference, control, treatment 1 (T1), and so on.
• Amount of material available: count the amount of mRNA involved in one channel of one hybridization as one unit.
• Number of slides available for the experiment.

Other factors
• The experimental process before hybridization: sample isolation, mRNA extraction, amplification and labelling.
• Controls planned: positive, negative, ratio, and so on.
• Verification method: northern blot, reverse transcriptase (RT)-PCR, *in situ* hybridization, and so on.
Experimental Constraints on Design

• The experiment must be feasible, given the number of microarrays and amounts of mRNA available.

• When the experiment is done, the more important scientific questions should be answered more accurately than the less important ones.

• Experiments can be done as direct comparisons, or in loop designs.

• Examples of single-factor experiments.
Single-Factor Experiments

• Example in which 3 mRNAs from different sources are being compared.

• All pairwise comparisons are of equal interest, such as comparison between three different tissues.

• Scientific aim: identify genes that are differentially regulated in the pairwise comparisons (i.e. in the different tissues).

• Experimental possibilities are shown on the next slide.
Single-Factor Experiments

- Assume variance for log ratios within a slide for a given gene is $\sigma^2$.
- The third column entry provides the average variance for the pairwise comparisons.
- Physical constraints govern the design choices – if unlimited amounts of reference RNA are available, but only one sample from each of the test tissues, design I is the only one that can be used.
- If two RNA samples are available from A, B, & C, designs II and III are possible. Design II uses twice as many slides as Design III. Design III provides more direct comparisons between the tissues, and is more accurate in this respect.
- As the number of samples increases, the reference designs (designs I and II) scale upwards in an analogous fashion. The “all-pairs” design (design III) becomes unfeasible, since the amounts of mRNA required become too great.

### Table 1 | Single-factor experiments

<table>
<thead>
<tr>
<th>Design choices</th>
<th>Number of slides</th>
<th>Units of material (number of samples)</th>
<th>Average variance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Indirect designs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Design I</td>
<td>3</td>
<td>$A = B = C = 1$</td>
<td>2.00</td>
</tr>
<tr>
<td>Design II</td>
<td>6</td>
<td>$A = B = C = 2$</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Direct design</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Design III</td>
<td>3</td>
<td>$A = B = C = 2$</td>
<td>0.67</td>
</tr>
</tbody>
</table>

Variance of estimated effects for three different designs of single-factor experiments. $\sigma^2$ was set to 1 throughout.
Single-Factor Experiments

- An alternative is the Loop Design (Kerr and Churchill).
- Note: the direction of the arrows is arbitrary, and they can be bidirectional.
- Drawbacks of this design are that samples that are not being directly compared are estimated less accurately.
- Another problem can be that of “broken” links.
Experimental Design

Multiple sample comparison in a complex loop
Further Design Issues

• See Speed and Churchill web pages:
  • http://stat-www.berkeley.edu/users/terry/zarray/Html/
  • http://aretha.jax.org/jax-cgi/churchill/index.cgi
Types of design employed in the Galbraith lab
Example 1

Treatment 1  Control Plants 1

Treatment 2  Control Plants 2
Example 2

• Three-way (up to n-way) designs are also possible. Here a three-way hybridization design, with four biological replications, requires a total of 12 microarrays (still affordable).
ANOVA Description

I. Normalization Model (mixed model ANOVA)

\[
\text{Log}_2 (\text{spot intensity}) = \text{treatment} + \text{slide} + \text{dye} + \text{metarow (slide)} + \text{metacolumn (slide)}
\]

Residuals from this model represent spot intensities that have been adjusted for systematic (i.e., across all genes) linear effects of treatment, individual slide, dye, and metarows and metacolumns within each slide.

The residuals are used as raw data to determine differences among treatments within each gene.

II. Gene Model (mixed model ANOVA)

For Each Element:

Residuals from ANOVA model = treatment + slide + dye

“Treatment” effect yields a significance value based on the differences between treatment levels (adjusted for slide and dye, etc.) and variability within treatment levels.

Note: only array elements detecting a large difference between treatments relative to variation within treatments will be classified as significant.

Long Oligonucleotide-based microarrays

- The ANOVA model allows us to predict, on a gene-by-gene basis, the significance with which genes are differentially expressed between the different conditions.
- Examining the treatment/control comparisons, we can generate a statistic indicating the probability that treatments are different for each gene represented on the array.
Genes showing significant differences at $P = 0.05$ are located below the red line.

Even very small fold changes can have low $P$-values if variability is small within treatments. The biological significance of such small changes merits further study.
Speculations as to where this is going

• We print around 8,000 microarrays per year.

• We have equipment to analyze 30 microarrays per run (Axon GenePix 4000AL)
Speculations as to where this is going

- This means we can contemplate microarray experiments involving very large numbers of microarrays.

- Example: Expression profiling of 800 different T-DNA insertion mutants (with three biological replications) would require 2,400 microarrays.
Questions

• Experimental design?

• Computational requirements for data analysis.
Experimental Design

- Requires 28 microarrays for all possible direct comparisons.
Experimental Design
Experimental Design

- Note: concept is scalable.
- It simply requires that you figure out the degree of acceptable separation of comparisons in the experimental design.
It is unclear whether such an experiment is computationally feasible.
Technical Issues in Microarray Hybridization

- Hybridization can be done manually, or using a variety of automatic and semi-automatic machines.

- Manual hybridization:
  - Follow our instructions! This includes all steps, such as the cross-linking, the requirement for blocking reagents, the temperatures of hybridization and washing, the volumes of the targets, and so-on.
  - Make sure the hybridization oven accurately maintains the set temperature, and that moisture levels are maintained.
  - Make sure the platform is properly leveled: an improperly leveled platform can result in the accumulation of hybridization solution on one side of the microarray, giving rise to partial hybridization.
  - Make sure you have denatured the target before applying it onto the microarray.
  - Ensure dust and smudge-free conditions when handling slides.
  - Avoid photobleaching by scanning the slides immediately and keeping them under light-tight conditions.
Technical Issues in Microarray Hybridization

- Automated hybridization:
  - Make sure you set the temperature correctly.
  - Make sure that appropriate solution are loaded within reservoirs.
  - Make sure you have denatured the targets before applying them onto the microarrays.
  - Use manual hybridization!
Manual Hybridization

A typical manual hybridization takes place under controlled temperature conditions within a humid chamber. A small volume (250μl) of target is used, without mixing with a raised coverslip configuration.