20.109
LABORATORY FUNDAMENTALS IN BIOLOGICAL ENGINEERING

MODULE 2
EXPRESSION ENGINEERING

Lecture # 4
Leona Samson
April 2nd 2009
Snapshot of the next four weeks

We will eliminate the expression of various genes using

(i) RNA interference technology
(ii) Cultured mouse ES cells
(iii) Chemiluminescent proteins
(iv) DNA microarrays
siRNA knockdown of expression of Renilla Luciferase plus various mouse gene

* Prepare cell extracts and measure Luciferase activities
siRNA knockdown of expression of Renilla Luciferase plus various mouse gene

Isolate total RNA in order to measure relative levels of all mRNAs – with special attention to YGI
Monitor mRNA expression level for every mouse gene in one single experiment.

How can we measure the level of thousands of mRNA species present in a particular cell type?

Now that we know the DNA sequence for every gene, this is possible!
How did we measure mRNA levels one at a time? This depends on Nucleic Acid Hybridization.

The specificity of G pairing with C and A pairing with T (or U) drives hybridization and provides a mechanism for quantitatively assessing the amount of a specific mRNA species in cells.

[Diagram of Nucleic Acid Hybridization]

http://www.accessexcellence.org/AB/GG/nucleic.html

Nucleic Acid Hybridization

32P- label
Let's first back-up. How did we measure mRNA levels one or two at a time? **Northern Blots**

All mRNAs separated by size

**Specific mRNAs Lit-up by radioactive cDNA probes**

**Northern Blots**

- **Transfer RNA by membrane blotting**
- **Cross-link RNA to the membrane using UV light**
- **Hybridize membrane with denatured $^{32}$P-cDNA probe**

**Formaldehyde-agarose gel**

- RNA Mkr 1
- RNA Mkr 2
- RNA Mkr 3

- 28S RNA
- 18S RNA

**AMG probe**

**GAPDH probe**
How to monitor mRNA expression level for every gene: Global transcriptional profiling

- Carry out thousands Northern Blots?
- Instead – DNA microarrays were developed
- DNA microarrays for global transcriptional profiling were not feasible before the sequencing of whole genomes.
The immobilized mRNA population is probed (hybridized) with $^{32}$P-labeled DNA sequences specific for one or two genes.
Northern Blots

Immobilized mRNA population hybridized with labeled DNA probe representing one or two genes

DNA Microarrays

Immobilized DNA probes representing all possible genes hybridized with labeled mRNA population
Need to achieve two things:

(i) Immobilize (array) thousands of DNA probes specific for each individual mRNA gene product

(ii) Label mRNA populations
Up to 20,000 probes per slide
The probes can be cDNAs (~ 1Kb) or oligonucleotides (20-70 mers)
Robots designed to spot up to 20,000 DNA probes per slide

The probes can be cDNAs (~ 1Kb) or oligonucleotides (20-70 mers)
The arrays we’ll be using……..

Agilent’s non-contact industrial inkjet printing process uniformly deposits oligo monomers onto specially-prepared glass slides. Both the catalog and custom microarrays are manufactured using Agilent’s non-contact in situ synthesis process of printing 60-mer length oligonucleotide probes, base-by-base, from digital sequence files. This is achieved with an inkjet process which delivers extremely small, accurate volumes (picoliters) of the chemicals to be spotted. Standard phosphoramide chemistry used in the reactions allows for very high coupling efficiencies to be maintained at each step in the synthesis of the full-length oligonucleotide. Precise quantities are reproducibly deposited “on the fly.” This engineering feat is achieved without stopping to make contact with the slide surface and without introducing surface-contact feature anomalies, resulting in consistent spot uniformity and traceability.

Ink-jet Technology

4x44K spots “features”

Need to achieve two things:

(i) Immobilize (array) thousands of probes specific for each individual gene

(ii) Label mRNA populations

Copy the population of purified mRNA species such that they are fluorescently labeled – hybridize to the array
RNA Sample 1

5’ ——— AAAAAAAA 3’

Anneal primer containing capture sequence I

5’ ——— AAAAAAAA 3’

TTTTTTT 5’

cDNA synthesis

5’ ——— AAAAAAAA 3’

TTTTTTT 5’

3’ ———

degrade RNA, hybridize array then bind Cy3

3’ ——— TTTTTT 5’

RNA Sample 2

5’ ——— AAAAAAAA 3’

Anneal primer containing capture sequence II

5’ ——— AAAAAAAA 3’

TTTTTTT 5’

cDNA synthesis

5’ ——— AAAAAAAA 3’

TTTTTTT 5’

3’ ———

degrade RNA, hybridize array then bind Cy5

3’ ——— TTTTTT 5’
Cells in state A

Cells in state B

Isolate mRNA populations

Label copies of mRNA species with RED or GREEN

Hybridize to the microarray

MIX
Cells in state B

Cells in state A

**DYE SWAP**

Isolate mRNA populations

Label copies of mRNA species with RED or GREEN

MIX

Hybridize to the microarray
siRNA knockdown of expression of Renilla Luciferase plus various mouse gene

Isolate total RNA in order to measure relative levels of all mRNAs – with special attention to YGI
What's happening at each spot?

Hybridization
mRNA present much higher in State A than State B
mRNA present much higher in State B than State A
mRNA present at equal levels in States A and B
Northern Blot vs. Microarray

**A**

- **DMC1**
- **SPS1**
- **DIT1**
- **SPS100**

**B**

- **DMC1**
- **SPS1**
- **DIT1**
- **SPS100**

**Hours:** 0 2 5 6 7 9 11

**Fold repression:**
- >20
- 10x
- 3x
- 1:1
- 3x
- 10x
- >20

**Fold induction:**
- >20
Hierarchical clustering to group together similarly regulated genes

Each colored vertical line in the horizontal lane displays the relative expression level of a single mRNA

13 time points, and several thousand genes

Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 14863-14868, December 1998
Cluster analysis and display of genome-wide expression patterns
This is the “two color” technology (you will use something similar), but there is another common technology that uses one color....
Affymetrix Technology

RNA: 2.5μg poly A*  
20 μg total

- **cDNA Synthesis**
- **IVT and Biotin Labeling**
- **Hybridization to Oligonucleotide Microarray**
- **Wash and Stain with Streptavidin-Phycoerythrin in Fluidics Module**
- **Scan with Confocal Laser Scanner and Analyze Data**
Affymetrix has focused on light-directed synthesis for the construction of high-density DNA probe arrays using two techniques: photolithography and solid-phase DNA synthesis. Synthetic linkers are attached modified with photochemically removable protecting groups to a glass substrate and direct light through a photolithographic mask to specific areas on the surface to produce localized photodeprotection.

Figure 1. Synthesis Confirmation.

<table>
<thead>
<tr>
<th>Mask Cycles</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>C</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>C</td>
<td>G</td>
<td>T</td>
</tr>
<tr>
<td>Probe 1</td>
<td>A</td>
<td>C</td>
<td>-</td>
<td>T</td>
<td>-</td>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probe 2</td>
<td></td>
<td>G</td>
<td>-</td>
<td>A</td>
<td>-</td>
<td>G</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>Probe 3</td>
<td>A</td>
<td>-</td>
<td>G</td>
<td>-</td>
<td>A</td>
<td>-</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>Probe 4</td>
<td>C</td>
<td>-</td>
<td>T</td>
<td></td>
<td>C</td>
<td>-</td>
<td>T</td>
<td></td>
</tr>
</tbody>
</table>

Probes Synthesized

- Probe 1: ACGT
- Probe 2: GATG
- Probe 3: AGAG
- Probe 4: CTCT
Affymetrix Technology
RNA fragments with fluorescent tags from sample to be tested

RNA fragment hybridizes with DNA on GeneChip® array
Shining a laser light at GeneChip® array causes tagged DNA fragments that hybridized to glow

Non-hybridized DNA

Hybridized DNA
GeneChip Probe Array

Hybridized Probe Cell

- Single stranded, labeled RNA target
- Oligonucleotide probe
- Millions of copies of a specific oligonucleotide probe
- >200,000 different complementary probes

Image of Hybridized Probe Array

24μm

1.28cm
Oligonucleotide “One Color Chips”

Target preparation

Array preparation

“One Gene”
Untreated
MMS Treated
Yeast cells treated with DNA damaging agents like those in tobacco smoke.
Affymetrix Oligonucleotide Based Expression Arrays

Jelinsky and Samson, 1999
## Northern Blot vs. Affymetrix Chip

<table>
<thead>
<tr>
<th>Northern</th>
<th>GeneChip Array</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5 hr</td>
<td>Untreated</td>
</tr>
<tr>
<td>1 hr</td>
<td></td>
</tr>
<tr>
<td>2 hr</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GeneChip Array</th>
</tr>
</thead>
<tbody>
<tr>
<td>YFL061W</td>
</tr>
<tr>
<td>SNZ1</td>
</tr>
<tr>
<td>DBP2</td>
</tr>
<tr>
<td>PHO3</td>
</tr>
<tr>
<td>ACT1</td>
</tr>
<tr>
<td>CDC19</td>
</tr>
</tbody>
</table>
# 'Next Generation' Sequencing Technologies

<table>
<thead>
<tr>
<th>Platform</th>
<th>Amplification</th>
<th>Sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harvard/Danaher/Agencourt/ABI</td>
<td>Emulsion PCR</td>
<td>Ligase</td>
</tr>
<tr>
<td>454 Life Sciences / Roche</td>
<td>Emulsion PCR</td>
<td>Polymerase - pyrosequencing</td>
</tr>
<tr>
<td>Solexa/Illumina</td>
<td>Bridge PCR</td>
<td>Polymerase - reversible terminator</td>
</tr>
<tr>
<td>Helicos</td>
<td>None</td>
<td>Polymerase - single base extension</td>
</tr>
<tr>
<td>Pacific Biosciences</td>
<td>None</td>
<td>Polymerase - active site</td>
</tr>
<tr>
<td>occupancy</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Solexa/Illumina DNA Sequencing Platform

1. PREPARE GENOMIC DNA SAMPLE
   - Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

2. ATTACH DNA TO SURFACE
   - Bind single-stranded fragments randomly to the inside surface of the flow cell channels.
   - Adapter
   - DNA fragment
   - Dense lawn of primers

3. BRIDGE AMPLIFICATION
   - Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.
Solexa/Illumina DNA Sequencing Platform

4. Fragments become double-stranded

5. Denature the double-stranded molecules

6. Complete amplification

The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.

Denaturation leaves single-stranded templates anchored to the substrate.

Several million dense clusters of double-stranded DNA are generated in each channel of the flow cell.
Solexa/Illumina DNA Sequencing Platform

7. Determine first base
   First chemistry cycle: to initiate the first sequencing cycle, add all four labeled reversible terminators, primers and DNA polymerase enzyme to the flow cell.

8. Image first base
   After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.

9. Determine second base
   Second chemistry cycle: to initiate the next sequencing cycle, add all four labeled reversible terminators and enzyme to the flow cell.
Reversible Terminator Chemistry

Solexa/Illumina DNA Sequencing Platform

7. DETERMINE FIRST BASE
First chemistry cycle: to initiate the first sequencing cycle, add all four labeled reversible terminators, primers and DNA polymerase enzyme to the flow cell.

8. IMAGE FIRST BASE
After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.

9. DETERMINE SECOND BASE
Second chemistry cycle: to initiate the next sequencing cycle, add all four labeled reversible terminators and enzyme to the flow cell.
Solexa/Illumina DNA Sequencing Platform

10. IMAGE SECOND CHEMISTRY CYCLE

After laser excitation, collect the image data as before. Record the identity of the second base for each cluster.

11. SEQUENCE READS OVER MULTIPLE CHEMISTRY CYCLES

Repeat cycles of sequencing to determine the sequence of bases in a given fragment a single base at a time.

12. ALIGN DATA

Align data, compare to a reference, and identify sequence differences.
In the MIT BioMicro Center

~70bp “reads”

7 million reads per channel (flow cell)!!!

From 7 million sequences can count/calculate the relative abundance of each original mRNA species - i.e. the transcriptional profile
### 'Next Generation' Sequencing Technologies

<table>
<thead>
<tr>
<th>Platform</th>
<th>Amplification</th>
<th>Sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harvard/Danaher/Agencourt/ABI</td>
<td>Emulsion PCR</td>
<td>Ligase</td>
</tr>
<tr>
<td>454 Life Sciences / Roche</td>
<td>Emulsion PCR</td>
<td>Polymerase - pyrosequencing</td>
</tr>
<tr>
<td>Solexa/Illumina</td>
<td>Bridge PCR</td>
<td>Polymerase - reversible terminator</td>
</tr>
<tr>
<td>Helicos</td>
<td>None</td>
<td>Polymerase - single base extension</td>
</tr>
<tr>
<td>Pacific Biosciences</td>
<td>None</td>
<td>Polymerase - active site</td>
</tr>
<tr>
<td>occupancy</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>