Using Microarrays to Study Cell Cycle Regulation

This is a light version with some pictures missing (3.2 Mb)
• Introduction to cell division and the cell cycle

• Cell culture synchronization and arrest methods

• Microarrays and periodically expressed genes

• Our own research
The Origin of Life

How does one cell become many?
The cell cycle
The cell cycle – a hot scientific topic

Papers about microarray analysis of the cell cycle have been extensively cited

- Spellman et al., 1998 (yeast)  697 citations
- Cho et al., 1998 (yeast)  385 citations
- Cho et al., 2002 (human)  61 citations

Cell cycle related research topics:
- Cancer, apoptosis, development, metabolic engineering, biomass increase, fertility, protein degradation, phosphorylation, signalling, etc.
A complex biological system
Evolutionary Conservation & Check-points

• The **core machinery** of the cell cycle is largely conserved throughout eukaryotes (from yeast to humans)

• Especially the multi-cellular organisms have extensive and sophisticated control mechanisms (**check-points**) that check if one step (e.g. DNA replication) is completed and error-free, before proceeding to the next steps (e.g. segregation of the chromosomes)
Different levels of control in the cell

• The activity of a protein depends on
  – regulation of its transcription (by activator/inhibitors)
  – its mRNA stability
  – the efficiency of translation (from mRNA to protein)
  – stability of the mature protein
  – the subcellular localization of the protein
  – regulation of the protein function by binding of activator and inhibitors
  – regulation of the protein function by post-translational modification (e.g. phosphorylation)
  – targeted degradation of the protein (e.g. by ubiquitination)
The Just-in-time principle

- Some products you use all the time, while others are only needed once in a while

- People tend to buy or order things right before they need them...
Just-in-time in the cell cycle
Periodically expressed genes
Outline

• Introduction to cell division and the cell cycle

• Cell culture synchronization and arrest methods

• Microarrays and periodically expressed genes

• Our own research
In a normal culture cells grow out of synchrony
**Arrest-and-release synchronization**

- **Temperature sensitive mutants** – unable to pass a certain point in the cell cycle at high temperatures

- **Alpha factor** – all cells arrest in the same stage as long as a factor is present in the culture
• Introduction to cell division and the cell cycle

• Cell culture synchronization and arrest methods

• Microarrays and periodically expressed genes

• Our own research
Using microarrays with synchronized cells

- Take samples from a synchronized cell culture
- Hybridize each sample to a microarray chip
- Normalize to make the chips comparable
- String the data-points together to a profile
- Look for genes whose expression is periodic
Time-series versus comparative studies

• Most microarray experiments compare measurements from condition A with condition B to find those genes that are differentially expressed
  – the typical analysis is a statistical test, e.g. t-test

• Time-series experiments are usually aimed at studying how gene expression changes over time (i.e. identifying genes that show a specific pattern of expression)
  – you can’t make a t-test, instead you have to look for some pattern!
Look for sine/cosine-like waves
Measuring Periodicity

- **Fourier-scoring**: extracting the "signal" at interdivision frequency.

\[
D_i = \sqrt{< A_i >^2 + < B_i >^2}
\]

\[
A = \sum_{all \ t} \sin(\omega t + \Phi)x_i(t)
\]

\[
B = \sum_{all \ t} \cos(\omega t + \Phi)x_i(t)
\]
Fourier score distribution
Comparison of Three Microarray Studies

- Visual inspection of expression profiles (Cho et al., 1998)
- Fourier analysis and correlation with profiles of known genes (Spellman et al., 1998)
- Statistical modelling, single pulse model (Zhao et al., 2001)
The thresholding problem

Yeast

Human
**Sensitivity** versus **Specificity**

**Sensitivity** = how many of the edible mushrooms do you accept

**Specificity** = how many of the mushrooms you accept are edible
The Thresholding problem

- Every threshold corresponds to a sensitivity and a specificity
- There (almost always) a trade-off between the two
• Introduction to cell division and the cell cycle
• Cell culture synchronization and arrest methods
• Microarrays and periodically expressed genes
• Our own research
Comparison of Replicates in Human Cells

Overlap between the 875 most periodic genes in three replicates, analyzed individually.

(our periodicity analysis of Whitfield et al.’s time-series data on double-Thymidine-block-arrested human HeLa cells)
Potential reasons for the lack of overlap

The discrepancies originate from

• Slightly different experimental conditions

• Handling of samples, amplification, hybridization, etc.

• Different analysis methods identify different sets of genes

• Thresholding difficult, because the grey zone area constitutes a large fraction of the score distribution

• Signal-to-noise ratio low for weakly expressed – and regulated genes
Prediction of Cell Cycle Genes

List of putative cell cycle regulated genes
Experimental Validation of Predictions

- Fermentation of synchronized yeast culture
- Samples taken at 20 min intervals
- Experiment covers two whole cell cycles

- Samples analyzed on the Febit Geniom microarray platform
- Probe design optimized with OligoWiz
- Non-linear normalization with Qspline
Experimental setup - sampling

• **Sampling process:**
  – Samples are shaken with ice (2°C I a few sec)
  – Centrifuged 1-1/2 min.
  – Supernatant discarded
  – Frozen in liquid N2 (-196°C)
  – Stored at -80°C until RNA extraction
Normalizing cell cycle expression data

Un-normalized Febit data

Normalized Febit data (invariant rank)
Probe variation within and between processors

Variation in probe intensity
Probe specific background subtraction

Variation in Probe Intensity

![Graph showing variation in probe intensity with raw and background subtracted data.](image-url)
Variation in Gene Expression Index

- 70% quantile
- Li & Wong

Variation (sd/mean)
Normalization

• Qspline (Workman et al., 2002)

• Li & Wong (dChip) gene expression index

• Global background subtraction

• Data log transformed, mean subtracted
Fourier analysis of time-series data

The Fourier score depends on

- relative magnitude of regulation
- number of time points
  - coverage (number of cycles)
  - resolution (samples per cycle)

\[
A = \sum_{all \ t} \sin(\omega t + \Phi)x_i(t)
\]

\[
B = \sum_{all \ t} \cos(\omega t + \Phi)x_i(t)
\]

\[
D_i = \sqrt{<A_i>^2 + <B_i>^2}
\]
### Permutation of data

<table>
<thead>
<tr>
<th></th>
<th>60</th>
<th>80</th>
<th>100</th>
<th>120</th>
<th>140</th>
<th>160</th>
<th>180</th>
<th>200</th>
<th>220</th>
<th>240</th>
<th>260</th>
<th>280</th>
</tr>
</thead>
<tbody>
<tr>
<td>YAL001C</td>
<td>-0.50</td>
<td>-0.08</td>
<td>-0.28</td>
<td>-0.62</td>
<td>-0.42</td>
<td>-1.14</td>
<td>-0.30</td>
<td>0.01</td>
<td>-0.06</td>
<td>0.34</td>
<td>1.12</td>
<td>0.54</td>
</tr>
<tr>
<td>Permutation 1</td>
<td>-0.62</td>
<td>-0.06</td>
<td>-0.28</td>
<td>0.54</td>
<td>-0.30</td>
<td>0.01</td>
<td>-0.50</td>
<td>-1.14</td>
<td>-0.06</td>
<td>1.12</td>
<td>-0.08</td>
<td>0.34</td>
</tr>
<tr>
<td>Permutation 2</td>
<td>0.01</td>
<td>-1.14</td>
<td>0.34</td>
<td>-0.30</td>
<td>-0.08</td>
<td>-0.62</td>
<td>-0.42</td>
<td>0.54</td>
<td>-0.28</td>
<td>1.12</td>
<td>-0.06</td>
<td>-0.50</td>
</tr>
<tr>
<td>Permutation 10000</td>
<td>-0.08</td>
<td>-0.30</td>
<td>-0.06</td>
<td>-0.50</td>
<td>0.54</td>
<td>-0.28</td>
<td>-0.34</td>
<td>-0.42</td>
<td>1.12</td>
<td>-1.14</td>
<td>-0.62</td>
<td>0.01</td>
</tr>
<tr>
<td>Fourier score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>1.98</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Permutation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

P-value = Fraction of scores for permuted data above observed score
### High confidence (FDR 0 - 10 %)

<table>
<thead>
<tr>
<th>All genes in set</th>
<th>Known periodic genes included</th>
<th>Bound by Cell cycle Transcription factors</th>
<th>Molecular Function Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>204</td>
<td>38 19%</td>
<td>78 38%</td>
<td>44 22%</td>
</tr>
<tr>
<td>71</td>
<td>20 28%</td>
<td>38 54%</td>
<td>16 23%</td>
</tr>
<tr>
<td>57</td>
<td>16 28%</td>
<td>29 51%</td>
<td>5 9%</td>
</tr>
<tr>
<td>13</td>
<td>1 8%</td>
<td>3 23%</td>
<td>2 15%</td>
</tr>
<tr>
<td>63</td>
<td>1 2%</td>
<td>8 13%</td>
<td>21 33%</td>
</tr>
</tbody>
</table>

### Intermediary confidence (FDR 10 - 20 %)

<table>
<thead>
<tr>
<th>All genes in set</th>
<th>Known periodic genes included</th>
<th>Bound by Cell cycle Transcription factors</th>
<th>Molecular Function Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>168</td>
<td>18 11%</td>
<td>49 29%</td>
<td>39 23%</td>
</tr>
<tr>
<td>39</td>
<td>10 26%</td>
<td>19 49%</td>
<td>7 18%</td>
</tr>
<tr>
<td>44</td>
<td>5 11%</td>
<td>18 41%</td>
<td>11 25%</td>
</tr>
<tr>
<td>9</td>
<td>2 22%</td>
<td>2 22%</td>
<td>4 44%</td>
</tr>
<tr>
<td>76</td>
<td>1 1%</td>
<td>9 12%</td>
<td>17 22%</td>
</tr>
</tbody>
</table>
Preliminary Validation Results

- More than 100 new periodic genes identified/validated
- For many of them, a role in the cell cycle is supported by other sources of evidence

<table>
<thead>
<tr>
<th>Gene</th>
<th>p-value</th>
<th>Neural Network score</th>
<th>GO Biological Process &amp; Gene Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene A</td>
<td>0.0009</td>
<td>0.76</td>
<td>Regulates the cell size requirement for passage through Start and commitment to cell division</td>
</tr>
<tr>
<td>Gene B</td>
<td>0.0026</td>
<td>0.70</td>
<td>cyclin involved in G1/S transition of mitotic cell cycle</td>
</tr>
<tr>
<td>Gene C</td>
<td>0.0081</td>
<td>0.59</td>
<td>Involved in cell cycle dependent gene expression</td>
</tr>
<tr>
<td>Gene D</td>
<td>0.0111</td>
<td>0.76</td>
<td>cell wall organization and biogenesis*</td>
</tr>
<tr>
<td>Gene E</td>
<td>0.0142</td>
<td>0.90</td>
<td>Required for spindle pole body duplication and a mitotic checkpoint function.</td>
</tr>
<tr>
<td>Gene F</td>
<td>0.0169</td>
<td>0.85</td>
<td>DNA repair*</td>
</tr>
<tr>
<td>Gene G</td>
<td>0.0192</td>
<td>0.74</td>
<td>G1/S transition of mitotic cell cycle*</td>
</tr>
<tr>
<td>Gene H</td>
<td>0.0222</td>
<td>0.76</td>
<td>DNA repair*</td>
</tr>
<tr>
<td>Gene I</td>
<td>0.0247</td>
<td>0.75</td>
<td>cellular morphogenesis*</td>
</tr>
<tr>
<td>Gene J</td>
<td>0.0255</td>
<td>0.81</td>
<td>regulation of exit from mitosis</td>
</tr>
<tr>
<td>Gene K</td>
<td>0.0353</td>
<td>0.46</td>
<td>Protein with similarity to putative glycosidase of the cell wall</td>
</tr>
<tr>
<td>Gene L</td>
<td>0.0482</td>
<td>0.74</td>
<td>G2/M transition of mitotic cell cycle*</td>
</tr>
<tr>
<td>Gene M</td>
<td>0.0520</td>
<td>0.81</td>
<td>chromatin assembly/disassembly*</td>
</tr>
<tr>
<td>Gene N</td>
<td>0.0630</td>
<td>0.92</td>
<td>actin cytoskeleton organization and biogenesis*</td>
</tr>
</tbody>
</table>
Proteins of Unknown Function

- About 30% of the newly discovered periodic genes have no known functional role

<table>
<thead>
<tr>
<th>Gene</th>
<th>p-value</th>
<th>Intensity</th>
<th>Neural Network Score</th>
<th>Binding of known cell cycle regulators</th>
<th>ProtFun Function prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene A</td>
<td>0.0004</td>
<td>0.20</td>
<td>0.79</td>
<td>FKH2 &amp; NDD1</td>
<td>cell envelope, enzyme ligase</td>
</tr>
<tr>
<td>Gene B</td>
<td>0.0014</td>
<td>0.41</td>
<td>0.56</td>
<td>SVM4 &amp; SVM6</td>
<td>Transcription regulation</td>
</tr>
<tr>
<td>Gene C</td>
<td>0.0016</td>
<td>0.15</td>
<td>0.80</td>
<td>FKH1 &amp; FKH2</td>
<td>cell envelope</td>
</tr>
<tr>
<td>Gene D</td>
<td>0.0019</td>
<td>0.54</td>
<td>0.76</td>
<td></td>
<td>Regulatory functions</td>
</tr>
<tr>
<td>Gene E</td>
<td>0.0063</td>
<td>0.63</td>
<td>0.84</td>
<td></td>
<td>Regulatoty functions, Transcription regulation</td>
</tr>
<tr>
<td>Gene F</td>
<td>0.0079</td>
<td>0.04</td>
<td>0.89</td>
<td>ACE2, SVM5 &amp; MCM1</td>
<td>cell envelope</td>
</tr>
<tr>
<td>Gene G</td>
<td>0.0093</td>
<td>0.12</td>
<td>0.76</td>
<td>SVM4</td>
<td>Transport and binding, Transporter</td>
</tr>
</tbody>
</table>
• Then you analyze your results and write an article about them!
No man is an island…

**Project coordinator**  
- Ulrik de Lichtenberg

**Synchronization and sampling**  
- Rasmus Wernersson & Flemming Bryde Hansen

**Probe design**  
- H. Bjørn Nielsen & Rasmus Wernersson

**Data analysis**  
- Thomas Skøt Jensen, Ulrik de Lichtenberg & H. Bjørn Nielsen

**Being the Boss**  
- Steen Knudsen & Søren Brunak