Biochemistry Review II

BBSI @ Pitt 2004

Rajan Munshi

Center for Computational Biology and Bioinformatics
University of Pittsburgh School of Medicine

May 26, 2004
Outline

- Cell Cycle
- Signal Transduction
- DNA Microarrays
Cell Cycle

- Four phases of the cell cycle:
  - Mitosis (M phase)
  - Gap 1 (G1 phase)
  - DNA Synthesis (S phase)
  - Gap 2 (G2 phase)
  - A fifth “phase”: G0 (quiescence)
Cell Cycle Phases

- M phase: cell division; each cell gets 1 copy of the genome
- G1 phase: cell growth; preparation for DNA replication
- S phase: DNA synthesis (replication)
- G2 phase: preparation for M phase
Chromatin Packaging

- Why does DNA in interphase “look” different from DNA in mitosis?
- Higher order of packaging
- Mitotic phase: DNA packaged into chromosomes
- Interphase: DNA present as chromatin
- “beads-on-a-string”
- beads = nucleosomes
- nucleosomes = DNA wrapped around histones
- Mitotic chromosomes = transcriptionally inactive (heterochromatin)
- Interphase chromatin = transcriptionally active (euchromatin)
Cell Cycle Control

- Web animation
- Checkpoints controlled by proteins
- Important group of checkpoint proteins are the cyclins
- So termed because their levels “cycle” during different phases
- Cyclins, by themselves, are inactive
- Have to associate with cyclin-dependent kinases (cdk)
- Cdk levels are invariant throughout the cell cycle
- G1 cyclin — cyclin D (cdk4)
- S-phase cyclins — cyclins A and E (cdk2)
- G2 cyclins — cyclin B (cdc2 (cdk1))
Cyclins and cdks

- **Cyclins**
  - G1 cyclin (cyclin D)
  - S-phase cyclins (cyclins E and A)
  - mitotic cyclins (cyclins B and A)

- **Cdks**
  - G1 Cdk (cdk4)
  - S-phase cdk (cdk2)
  - M-phase cdk (cdc2 (Cdk1))
Cell Cycle Checkpoints

- **G2**
  - Cyclin B-cdc2
  - Nuclear envelope breakdown; assembly of mitotic spindle; activation of APC

- **M**
  - Cyclin A-cdc2
  - Dissociation of APC; exit from M

- **G1**
  - Cyclin A-cdk2
  - Restriction point, R

- **S**
  - Cyclin E-cdk2
  - Cyclin D-cdk4

In molecular terms, passing R is the phosphorylation of the retinoblastoma protein, Rb. Unphosphorylated Rb binds the transcription factor, E2F; the phosphorylated form cannot bind E2F, thereby allowing E2F to modulate gene expression.
Regulation via Phosphorylation

- Phosphorylation and dephosphorylation regulate many key events
- Cell cycle control
- Signal transduction
- Transcription
Signal Transduction

- Ensures that a signal is converted from one form to another
- From the exterior of the cell to the interior
- Retain original signal content
Steps in Signal Transduction

- Signal is sent. e.g. hormone, non-steroid ligand (epinephrine)
- Recognition of the signal by the cell via a receptor.
- Receptors can be present on the cell membrane or in the cytosol
- *Internal signaling molecules* transduce and amplify the signal
- Carried out via a *signaling cascade*, with multiple regulatory steps
- E.g. Glycogen breakdown in response to epinephrine
Cell Receptors

- Ion-channel linked: involved in rapid synaptic signaling between excitable cells; mediated by neurotransmitters
- Enzyme-linked receptors: when activated, either function directly as enzymes or are associated with enzymes.
- G-protein coupled receptors (GPCR)
GPCRs

- Largest family of cell-surface receptors
- Biological functions include smell, taste, vision, neurotransmission, blood pressure, embryogenesis, cell growth, development
- Rhodopsin is the only GPCR with a known 3D structure
- Contains 7 membrane traversing α helices (7TM)
- N terminal – outside cell, C terminal – inside cell
- Ligand binding outside cell induces conformational change detected inside cell
- Mediating molecule is a G protein (hence the name GPCR)
- Heterotrimeric GTP-binding regulatory protein (α, β, γ)
- Activated G protein transmits signal by binding to other proteins (e.g. adenylate cyclase: converts ATP to cAMP)
GPCR Structure
GPCR Structure (contd.)
Signal reception

Signal mediation and amplification

Regulation by reversible Phosphorylation and dephosphorylation

Signal effects
DNA Microarrays: An Overview
Also called

- DNA chips
- biochips
- gene chips / gene arrays
- genome chips / genome arrays
What is a microarray?

- An arrangement of DNA sequences on a solid support
- Each microarray contains thousands of genes
- Able to simultaneously monitor gene expression levels in all these genes
- Used for:
  - gene expression studies
  - disease diagnosis
  - pharmacogenetics (drug discovery)
  - toxicogenomics
Types

- Two basic microarray technologies
- cDNA arrays (Stanford)
- High-density oligonucleotide arrays (Affymetrix)
- Each technology has its merits and demerits
Definition
High-density oligonucleotide arrays

- Pioneered by Affymetrix (GeneChip®)
- DNA probe sequences are 25-mer fragments
- Built *in situ* (“on-chip”) by photolithography
- Uses 1 fluorescent dye
High-density oligonucleotide arrays

- Each sequence is represented by a probe set
- 1 probe set = 16 probe pairs
- Each probe pair = 1 Perfect Match (PM) probe cell and 1 MisMatch (MM) probe cell
- PM = perfectly complementary to target
- MM = central base is mismatched to target
Affymetrix Probe Sets

5' 3'

Perfect Match (PM)

MisMatch (MM)

Probe set (102353_at)

Probe pair

GTACTTCCATGCCTAGCTAGCTAGT

GTACTTCCATGCATACTAGCTAGCTAGT

Perfect Match (PM)

MisMatch (MM)
Affymetrix chip
A Single Probe set
cDNA arrays

- Also known as spotted arrays
- Support can be glass or membrane
- DNA sequences are robotically “imprinted”
- Sequences can range from 30 bp to 2 kb
- Sequences are cDNA clones
- Uses 2 fluorescent dyes (cy3, cy5)
cDNA arrays overview
cDNA arrays

Animation
(Courtesy: Dr. A. Malcolm Campbell, Davidson College, NC)
(www.bio.davidson.edu/courses/genomics/chip/chip.html)
Genome-on-a-chip (yeast)
## General Steps

<table>
<thead>
<tr>
<th>Probe</th>
<th>Chip Fabrication</th>
<th>Target</th>
<th>Assay</th>
<th>Readout</th>
<th>Informatics</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA or cDNA with known identity</td>
<td>Putting probes on chip (robotic imprinting, photolithography)</td>
<td>Fluorescently labeled cDNA (single channel, dual channel)</td>
<td>Hybridization (Southern Blot)</td>
<td>Fluorescence intensities, fold-change ratios (up- or down-regulated)</td>
<td>Visualization, data mining</td>
</tr>
</tbody>
</table>
Analysis

- Low-level analysis
  - Extraction of signal intensities
  - Normalization of samples

- High-level analysis
  - Unsupervised learning (clustering)
    - Aggregation of a collection of data into clusters based on different features in a data set (e.g. hierarchical clustering, SOM)
  - Supervised learning (class discovery)
    - Incorporates knowledge of class label information to make distinctions of interest by using a training set.
Low-level analysis

Gene Expression Intensity (Signal)

In other words, a numerical value is obtained

Now, these values can be compared because fluorescence intensity is directly proportional to gene expression.
High-level analysis

<table>
<thead>
<tr>
<th>Probe Set</th>
<th>Accession</th>
<th>M1-0</th>
<th>M1-1</th>
<th>M1-2</th>
<th>M1-3</th>
<th>M1-4</th>
<th>M1-6</th>
<th>M1-8</th>
<th>M1-10</th>
<th>M1-12</th>
<th>M1-13</th>
<th>M1-14</th>
<th>M1-15</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFFEN1</td>
<td>M10832</td>
<td>80.43</td>
<td>90.26</td>
<td>94.25</td>
<td>88.87</td>
<td>92.23</td>
<td>96.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN2</td>
<td>M7987</td>
<td>95.09</td>
<td>95.09</td>
<td>120.52</td>
<td>120.52</td>
<td>86.20</td>
<td>101.70</td>
<td>98.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN3</td>
<td>M25935</td>
<td>50.00</td>
<td>54.00</td>
<td>48.00</td>
<td>48.00</td>
<td>48.00</td>
<td>48.00</td>
<td>48.00</td>
<td>48.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN4</td>
<td>M18834</td>
<td>519.38</td>
<td>549.27</td>
<td>712.04</td>
<td>884.61</td>
<td>867.34</td>
<td>814.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN5</td>
<td>J04423</td>
<td>456.23</td>
<td>768.35</td>
<td>347.41</td>
<td>353.26</td>
<td>256.20</td>
<td>230.09</td>
<td>230.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN6</td>
<td>J04423</td>
<td>888.22</td>
<td>928.23</td>
<td>498.79</td>
<td>410.29</td>
<td>587.63</td>
<td>394.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN7</td>
<td>J04423</td>
<td>592.85</td>
<td>777.07</td>
<td>492.35</td>
<td>495.84</td>
<td>382.93</td>
<td>394.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN8</td>
<td>J04423</td>
<td>1262.96</td>
<td>1800.34</td>
<td>1030.91</td>
<td>957.47</td>
<td>191.67</td>
<td>297.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN9</td>
<td>J04423</td>
<td>1056.25</td>
<td>1518.82</td>
<td>787.47</td>
<td>765.66</td>
<td>715.04</td>
<td>718.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN10</td>
<td>J04423</td>
<td>1545.23</td>
<td>2043.69</td>
<td>1042.41</td>
<td>1325.91</td>
<td>1105.97</td>
<td>1000.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN11</td>
<td>J04423</td>
<td>4759.14</td>
<td>6168.65</td>
<td>4139.77</td>
<td>3732.00</td>
<td>3932.16</td>
<td>3061.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN12</td>
<td>J04423</td>
<td>7838.64</td>
<td>9652.75</td>
<td>6168.41</td>
<td>6566.99</td>
<td>5619.94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN13</td>
<td>J04423</td>
<td>9515.45</td>
<td>1141.46</td>
<td>792.21</td>
<td>792.21</td>
<td>792.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN14</td>
<td>J04423</td>
<td>9274.23</td>
<td>120.71</td>
<td>105.18</td>
<td>79.66</td>
<td>79.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN15</td>
<td>J04423</td>
<td>106.29</td>
<td>141.72</td>
<td>103.99</td>
<td>92.20</td>
<td>87.04</td>
<td>94.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN16</td>
<td>J04423</td>
<td>104.29</td>
<td>234.80</td>
<td>177.92</td>
<td>164.80</td>
<td>203.52</td>
<td>162.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN17</td>
<td>J04423</td>
<td>152.24</td>
<td>134.12</td>
<td>136.57</td>
<td>157.06</td>
<td>140.92</td>
<td>149.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN18</td>
<td>J04423</td>
<td>116.43</td>
<td>139.42</td>
<td>108.76</td>
<td>108.76</td>
<td>108.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN19</td>
<td>J04423</td>
<td>126.39</td>
<td>173.24</td>
<td>134.53</td>
<td>123.5</td>
<td>120.56</td>
<td>140.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN20</td>
<td>J04423</td>
<td>915.45</td>
<td>296.23</td>
<td>296.23</td>
<td>296.23</td>
<td>296.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN21</td>
<td>J04423</td>
<td>214.96</td>
<td>302.45</td>
<td>194.21</td>
<td>175.31</td>
<td>156.52</td>
<td>174.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN22</td>
<td>J04423</td>
<td>240.87</td>
<td>319.84</td>
<td>226.72</td>
<td>182.48</td>
<td>196.41</td>
<td>210.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN23</td>
<td>J04423</td>
<td>46.89</td>
<td>49.41</td>
<td>49.41</td>
<td>49.41</td>
<td>49.41</td>
<td>49.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN24</td>
<td>J04423</td>
<td>37.94</td>
<td>46.30</td>
<td>42.80</td>
<td>33.77</td>
<td>36.3</td>
<td>35.97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN25</td>
<td>J04423</td>
<td>56.25</td>
<td>55.25</td>
<td>53.24</td>
<td>53.24</td>
<td>53.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN26</td>
<td>J04423</td>
<td>50.69</td>
<td>51.69</td>
<td>54.01</td>
<td>54.01</td>
<td>54.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN27</td>
<td>J10713</td>
<td>88.19</td>
<td>71.25</td>
<td>88.80</td>
<td>88.80</td>
<td>88.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN28</td>
<td>J10713</td>
<td>59.51</td>
<td>51.51</td>
<td>54.01</td>
<td>54.01</td>
<td>54.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN29</td>
<td>J10713</td>
<td>41.38</td>
<td>40.19</td>
<td>37.51</td>
<td>37.51</td>
<td>37.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFFEN30</td>
<td>J10713</td>
<td>1071.2</td>
<td>114.61</td>
<td>103.73</td>
<td>103.73</td>
<td>103.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Now what??
High-level analysis (Hierarchical Clustering)

- Algorithm that “pairs” similarly expressed genes
- Uses Pearson’s correlation coefficient ($r$)
- Useful to gain a general understanding of genes involved in pathways
Time course of serum stimulation of human fibroblasts

- Identify clusters of genes that are co-regulated
- Identification of novel genes
- Very widespread method for microarray analysis
High-level analysis (self-organizing maps)

- Algorithm that clusters genes based on similar expression values
- Useful for finding patterns in biological data
- Cocaine study
- 5 regions of the rat brain under treated and untreated conditions
- e.g. cluster 3
Overall Goal

>10,000 genes

Identify potential therapeutic targets

<50 genes

Experimental confirmation
Potential Problems

- Local contamination
Array Contamination
Potential Problems

- Local contamination
- Normalization
- Statistical significance of difference in expression
- cDNA arrays
  - must have the genes cloned
  - need relatively pure product
- Affymetrix arrays
  - need sequence information
Additional Reading

- Affymetrix website: www.affymetrix.com
- Stanford University: genome-www.stanford.edu
- Nature Genetics, vol. 21 supplement, “The Chipping Forecast”
- www.microarray.org
- www.gene-chips.com/
- ihome.cuhk.edu.hk/~b400559/array.html
- www.stat.wisc.edu/~yandell/statgen/reference/array.html