Correlating Compound Classes and Gene Expression Data

Paul Blower
LeadScope, Inc.
April 4, 2001
Objectives of Study

- Identify compound classes for which the activity is well correlated with expression pattern of specific genes

- Increase knowledge of molecular mechanisms without direct testing of compound vs gene expression
Overview

- Summary of NCI data
- Methods for:
  - Selecting genes
  - Finding compound classes
- Examples
  - Correlation of melanoma and leukemia genes with quinones
Datasets: Conceptual Framework

- **S** (structure) 4463 compounds
- **A** (activity) 4463 compounds
- **T** (target) 3748 genes

* J. Weinstein, *Science* 1997, **275**, 343-49
NCI Cancer Screening: A Matrix

- Tests compounds against 60 tumor cell lines
  - Breast (8)  Leukemia (6)  Ovarian (6)
  - CNS (6)    Lung (9)     Prostate (2)
  - Colon (7)  Melanoma (7) Kidney (8)

- Each compound tested at 5 concentrations, usually $10^{-4}M - 10^{-8}M$

- Determine: $GI_{50}$, TGI and $LC_{50}$
NCI Gene Expression: T Matrix*

- Compared cDNA from individual cell line (resting) with cDNA from a pool of 12 cell lines
- Microarray contained 9,703 DNA elements
  - 3,700 named genes
  - 1,900 human homologs
  - 4,100 ESTs
- Selected 3748 genes passing QC and sequence verified

AT Matrix

4463 compounds  x  3748 genes = AT

60 cell lines

3748 genes
Example of A- and T-Matrices

- Breast
- CNS
- Colon
- Leukemia
- Lung
- Melanoma
- Ovarian
- Renal

Gene 486233

Compound 656238
S(AT) Matrix

\[
\begin{array}{c}
\text{S} \times \text{AT} = \text{SAT}
\end{array}
\]
Selection of Genes from T Matrix

- Genes with high variance over 60 cell lines
- Genes correlated with compound activity from AT matrix
- Genes correlated with cell origins (melanoma, leukemia, etc.)
  - individual cell clusters
  - multiple cell clusters
Selection of Genes from T Matrix

- Studentized range test with 7 cell clusters*
  - 476 genes with significant difference between 2 clusters
  - 391 genes (82%) involved melanoma or leukemia

- Calculate sub-[AT] matrix

- Select genes with high compound correlations

Distribution of Correlation Coefficients

![Bar chart showing the distribution of correlation coefficients for the Rab7 gene.](chart.png)
Finding Compound Classes

- Load gene correlations into LeadScope™ as compound properties

- Structure-based datamining
  - sorted feature list
  - recursive partitioning*
  - simulated annealing*

- Construct user-defined features to refine compound classes

* Research prototype only
Compound Classes

Compound class correlated with melanoma gene *Rab7*

![benzothiophenedione](image1)

Compound class correlated with leukemia gene *LCP1*

![indolonaphthoquinone](image2)
Quinone Cytotoxicity

Redox cycling

Alkylation

\[
\text{NADH} \rightarrow \text{NAD}^+ \rightarrow \text{O}^\cdot \rightarrow \text{O}_2 \rightarrow \text{HO}_2^\cdot \rightarrow \text{OH} \rightarrow \text{SR}
\]
Quinone Anticancer Agents

- Mitomycin C
- Adriamycin $R = \text{CH}_2\text{OH}$
- Daunorubicin $R = \text{CH}_3$
- EO9
- Actinomycin D (portion)
## Quinone/Gene Correlations

<table>
<thead>
<tr>
<th>Class</th>
<th>Count</th>
<th>NFATx</th>
<th>Rab7</th>
</tr>
</thead>
<tbody>
<tr>
<td>actinomycin</td>
<td>12</td>
<td>-1.36</td>
<td>1.44</td>
</tr>
<tr>
<td>anthraquinone</td>
<td>65</td>
<td>2.86</td>
<td>-6.52</td>
</tr>
<tr>
<td>aziridinylquinone</td>
<td>11</td>
<td>-1.84</td>
<td>0.93</td>
</tr>
<tr>
<td>benzothiophenedione</td>
<td>23</td>
<td>-7.40</td>
<td>10.96</td>
</tr>
<tr>
<td>indoledione</td>
<td>7</td>
<td>-0.84</td>
<td>-0.07</td>
</tr>
<tr>
<td>indolonaphthoquinone</td>
<td>20</td>
<td>6.36</td>
<td>-2.62</td>
</tr>
<tr>
<td>Gene ID</td>
<td>656238</td>
<td>661223</td>
<td>Gene Description</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>--------</td>
<td>-----------------------------------------------------------</td>
</tr>
<tr>
<td>486233</td>
<td>0.67</td>
<td>-0.20</td>
<td>Human small GTP binding protein Rab7</td>
</tr>
<tr>
<td>270505</td>
<td>0.65</td>
<td>-0.68</td>
<td>H.sapiens mRNA matrix metalloproteinase 1</td>
</tr>
<tr>
<td>472138</td>
<td>0.62</td>
<td>-0.57</td>
<td>Homo sapiens sgk gene Chr.6</td>
</tr>
<tr>
<td>248589</td>
<td>0.60</td>
<td>-0.51</td>
<td>Homo sapiens sgk gene Chr.</td>
</tr>
<tr>
<td>416332</td>
<td>0.59</td>
<td>-0.35</td>
<td>Human frezzled (fre) mRNA</td>
</tr>
<tr>
<td>486086</td>
<td>0.59</td>
<td>-0.39</td>
<td>PIGF Phosphatidylinositol glycan</td>
</tr>
<tr>
<td>366379</td>
<td>0.58</td>
<td>-0.59</td>
<td>RNASE1 Ribonuclease (pancreatic)</td>
</tr>
<tr>
<td>364929</td>
<td>0.58</td>
<td>-0.54</td>
<td>Serine/Threonine-Protein Kinase R1</td>
</tr>
<tr>
<td>486471</td>
<td>0.58</td>
<td>-0.62</td>
<td>CALD1 Caldesmon</td>
</tr>
</tbody>
</table>
Quinone/Gene Correlations

Correlation coefficient depends on substituent
Quinone/Gene Correlations
Possible GeneChip Experiment

Whole space:
~ $10^6$ gene chips
5000 compounds
60 cell lines
@ 3 times

In Silico
3 compounds
12 cell lines

Partial space:
~ $10^2$ gene chips
3 compounds
12 cell lines
@ 3 times
Summary

- Genes correlated with tissue type and compound activity

- Compound classes correlated with specific genes:
  - benzothiophenedione
  - indolonaphthoquinone

- In-silico technique aids experimental design
Acknowledgement

John Weinstein  NCI
Samuel Richman  NCI
Michael Fligner  Ohio State Univ.
Joseph Verducci  Ohio State Univ.
Chihae Yang  LeadScope, Inc.
Limin Yu  LeadScope, Inc.