Application of Genomic Technology to Toxicology
Identifying Predictive Biomarkers and Assessing the Impact of Chemicals on Cell Signaling Networks

Rusty Thomas
CIIT Centers for Health Research
SOT Colgate Palmolive Lecture
San Diego, 2006
Is There a Role for Genomics in the 3Rs...
Is There a Role for Genomics in the 3Rs...

Hazard Identification

Application of Genomic and Metabonomics Technology to Identify Biomarkers Predictive of Rodent Cancer Bioassay

Dose Response Assessment

Application of High-Coverage Functional Genomic Screens to Dissect Cell Signaling Networks
Application of Genomics to Hazard ID

Background and Significance

• Two-year rodent bioassays play a central role in evaluating both the carcinogenic potential of a chemical and generating quantitative information on the dose-response behavior for chemical risk assessments.

• Due to the resource-intensive nature of these studies, each bioassay costs $2 to $4 million and takes over three years to complete.

• Since its inception in 1978, the number of chemicals currently tested by the National Toxicology Program (NTP) stands at 530 in long-term studies and 70 in short-term tests.

• Currently, there are approximately 80,000 chemicals registered for commercial use in the United States with 2,000 more added each year.
Experimental Flow Chart

Chemicals Previously Tested in Rodent Two-Year Bioassay

Positive for Tumors

Subchronic Exposure

Identify Biomarkers

Transcriptomics

Metabonomics

Proteomics

Construct Statistical Models

Validate Biomarkers

Negative for Tumors
# Experimental Design

<table>
<thead>
<tr>
<th>Chemical</th>
<th>Genetic Tox</th>
<th>Exposure Route</th>
<th>MTD Dose</th>
<th>NTP Designation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,5-Naphthalenediamine (NAPD)</td>
<td>+</td>
<td>Feed</td>
<td>2,000 ppm</td>
<td>Positive</td>
</tr>
<tr>
<td>Benzofuran (BFUR)</td>
<td>-</td>
<td>Gavage</td>
<td>240 mg/kg</td>
<td>Positive</td>
</tr>
<tr>
<td>N-(1-naphthyl)ethylenediamine dihydrochloride (NEDD)</td>
<td>+</td>
<td>Feed</td>
<td>2,000 ppm</td>
<td>Negative</td>
</tr>
<tr>
<td>Pentachloronitrobenzene (PCNB)</td>
<td>-</td>
<td>Feed</td>
<td>8,187 ppm</td>
<td>Negative</td>
</tr>
<tr>
<td>Corn oil (CCON)</td>
<td></td>
<td>Gavage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feed (FCON)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Experimental Design
Experimental Results

Lung Gene Expression

Liver Gene Expression

Urinary Metabolites
Experimental Results

**Ces1**

- Lung

**E130013N09Rik**

- Liver

Normalized Expression

- CCON, FCON, NEDD, PCNB, NAPD, BFUR

- NONCARC → CARC → NONCARC
Experimental Results

qRT-PCR Confirmation

A. Ces1

Fold Change in Gene Expression

Carc NonCarc

B. E130013N09Rik

Fold Change in Gene Expression

Carc NonCarc
Application of Genomics to Hazard ID

Results and Conclusions

• Statistical classification analysis of gene expression data in both the lung and liver samples exhibited 100% predictive accuracy based on three-fold cross-validation.

• Statistical classification analysis of urinary metabolites exhibited 88% predictive accuracy based on three-fold cross-validation.

• The two carcinogenic chemicals used in the study showed common changes despite different chemical structures, genotoxicity categories, and potential modes-of-action.
Is There a Role for Genomics in the 3Rs...

Application of Genomic and Metabonomics Technology to Identify Biomarkers Predictive of Rodent Cancer Bioassay

Application of High-Coverage Functional Genomic Screens to Dissect Cell Signaling Networks
Application of Genomics to Dose Response Assessment

Lessons from the Industrial Revolution...
Application of Genomics to Dose Response Assessment

A fundamental understanding of the underlying biology will provide…

– potential molecular targets involved
– logic of the signaling network
– shape of the dose response curve
– how well this response is conserved across species
Examples With Drosophila and MAPK Signaling

Boolean Model of Drosophila Segment Polarity Gene Expression


ODE Model of PDGF Stimulated MAPK

Conolly and Zhang, unpublished data.
Application of Genomics to Dose Response Assessment
How Do We Implement a Functional Genomics Approach to Dissect These Networks?

Assemble the “Parts List”

Identify How the Pieces Fit Together in Each Subsystem

Understand How the Subsystems Interconnect
How Do We Implement a Functional Genomics Approach to Dissect These Networks?

Assemble the “Parts List”

Identify How the Pieces Fit Together in Each Subsystem

Understand How the Subsystems Interconnect
Assembling the “Parts List”

Anatomy of a Screen: Constructing The Assay

Stimulus + Full-length Genes or siRNAs
Loss of function

Two “Functional” Approaches

Full-length Genes
Gain of function

Cellular Assay (Promoter/RE Reporter)
Assembling the “Parts List”

Anatomy of a Screen

Arrayed, full-length gene or siRNA set in 384-well plates

Transfect genes into reporter cells

Identify hits

Identify components of the signaling pathway

Gene1 Gene2 Gene3
Gene4 Gene5 Gene6

Identify hits
Assembling the “Parts List”

Anatomy of a Screen

Movie Showing Robotic Preparation of Full-length cDNA library
Assembling the “Parts List”

Anatomy of a Screen

Movie Showing Robotic Screening of Full-length cDNA library
Preliminary Results

NFκB Gain-of-Function and Loss-of-Function Screens

Screen Type: Full-length Gene
Gain-of-Function
Genes Screened: ~14,000

Screen Type: Full-length Gene
Loss-of-Function
Genes Screened: ~14,000

Graph showing frequency and cumulative percentage of fold induction.

Venn diagram indicating 75 genes induced by IL-1, 110 by TNFα, and 61 by both.

IL-1
TNFα
75
110
61
How Do We Implement a Functional Genomics Approach to Dissect These Networks?

Assemble the “Parts List”

Identify How the Pieces Fit Together in Each Subsystem

Understand How the Subsystems Interconnect
Identify How the Pieces Fit Together

Anatomy of a Screen: Organizing the Pathway

RNAi Knockdown or Dominant Negative

Reduced or No Reporter Activity

cDNA Expression

RNAi Knockdown or Dominant Negative

cDNA Expression

Reporter Activity
Identify How the Pieces Fit Together

Anatomy of a Screen: Organizing the Pathway

<table>
<thead>
<tr>
<th>Individual Screen Matrices</th>
<th>Combined Matrix</th>
<th>Functional Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRAF2 IKK 1</td>
<td>TRAF2 IKK 1</td>
<td>TNF</td>
</tr>
<tr>
<td>NFKB IKK -1</td>
<td>TRAF2 NFKB 1</td>
<td>TRAF2</td>
</tr>
<tr>
<td>TNF IKK 1</td>
<td>NFKB IKK -1</td>
<td>IKK</td>
</tr>
<tr>
<td>IKK TRAF2 -1</td>
<td>TNF TRAF2 1</td>
<td>NFKB</td>
</tr>
<tr>
<td>NFKB TRAF2 -1</td>
<td>TNF IKK 1</td>
<td></td>
</tr>
<tr>
<td>TNF TRAF2 1</td>
<td>NFKB TNF -1</td>
<td></td>
</tr>
</tbody>
</table>
Identify How the Pieces Fit Together

Anatomy of a Screen: Organizing the Pathway
Preliminary Results

Functional Network Map
**Preliminary Results**

**Tissue Specific Networks**

*Based on expression data in www.symatlas.org*
Preliminary Results

Tissue Specific Networks

*Expression values based on GNF Symatlas (www.symatlas.org)
How Do We Implement a Functional Genomics Approach to Dissect These Networks?

Assemble the “Parts List”

Identify How the Pieces Fit Together in Each Subsystem

Understand How the Subsystems Interconnect
Understanding How the Subsystems Interconnect

Upstream Cellular Signaling Pathway

Primary Expression Changes

Secondary Expression Changes

Tertiary Expression Changes

Transcriptional Alterations

Centers For Health Research
Understanding How the Subsystems Interconnect

Example with the Stress Response Network

- **STRESS**
- Denatured Proteins
- HSPs
- HSF1
- Free HSF1
- Homotrimers
- Cytoplasm
- Nucleus
- Heat Shock Induced Transcription
- HSE
Understanding How the Subsystems Interconnect

HSF1 mRNA

siHSF1 #1

US     HS     US      HS       US      HS
siHSF1#1 siHSF1#2 siLuc

HSF1 Protein

CIIT

Centers For Health Research
Understanding How the Subsystems Interconnect

![Gene Expression Heatmap]

Time (hrs) 0.5 2 4

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>200664_s_at</td>
<td>DNAJB1</td>
</tr>
<tr>
<td>200666_s_at</td>
<td>DNAJB1</td>
</tr>
<tr>
<td>200800_s_at</td>
<td>HSPA1A</td>
</tr>
<tr>
<td>202581_at</td>
<td>HSPA1B</td>
</tr>
<tr>
<td>117_at</td>
<td>HSPA6</td>
</tr>
<tr>
<td>213418_at</td>
<td>HSPA6</td>
</tr>
<tr>
<td>204420_at</td>
<td>FOSL1</td>
</tr>
<tr>
<td>214315_x_at</td>
<td>CALR</td>
</tr>
<tr>
<td>203239_s_at</td>
<td>CNOT3</td>
</tr>
<tr>
<td>210317_s_at</td>
<td>YWAHE</td>
</tr>
<tr>
<td>212553_at</td>
<td>KIAA0460</td>
</tr>
<tr>
<td>218251_at</td>
<td>MID1IP1</td>
</tr>
<tr>
<td>218607_s_at</td>
<td>SDAD1</td>
</tr>
<tr>
<td>204494_s_at</td>
<td>DFKZP434H132</td>
</tr>
<tr>
<td>204693_at</td>
<td>CDC42EP1</td>
</tr>
<tr>
<td>206724_at</td>
<td>CBX4</td>
</tr>
<tr>
<td>213038_at</td>
<td>IRBDC3</td>
</tr>
<tr>
<td>36564_at</td>
<td>IRBDC3</td>
</tr>
<tr>
<td>203254_s_at</td>
<td>TLN1</td>
</tr>
<tr>
<td>209344_at</td>
<td>TPM4</td>
</tr>
<tr>
<td>211965_at</td>
<td>ZFP36L1</td>
</tr>
<tr>
<td>37462_i_at</td>
<td>SF3A2</td>
</tr>
<tr>
<td>210996_s_at</td>
<td>YWHAE</td>
</tr>
<tr>
<td>209381_x_at</td>
<td>SF3A2</td>
</tr>
<tr>
<td>220748_s_at</td>
<td>ZNF580</td>
</tr>
<tr>
<td>218887_at</td>
<td>MRPL2</td>
</tr>
<tr>
<td>218010_x_at</td>
<td>C20orf149</td>
</tr>
</tbody>
</table>

Fold Change
- >8: Positive HSF1 regulation
- 4: Positive HSF1 regulation
- 2: Positive HSF1 regulation
- 1: Positive HSF1 regulation
- -2: Negative HSF1 regulation
- -4: Negative HSF1 regulation
- <8: Negative HSF1 regulation
- No Regulation
## Understanding How the Subsystems Interconnect

Gene ontology (GO) analysis of genes with significant regulation by HSF1 and significant alterations in expression following proteotoxic stress.

<table>
<thead>
<tr>
<th>Biological Process GO Category</th>
<th>p-value</th>
<th>Mol. Function GO Category</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response to unfolded protein</td>
<td>0.000</td>
<td>DNA binding</td>
<td>0.000</td>
</tr>
<tr>
<td>Regulation of transcription, DNA-dependent</td>
<td>0.000</td>
<td>Metal ion binding</td>
<td>0.000</td>
</tr>
<tr>
<td>Protein folding</td>
<td>0.000</td>
<td>Unfolded protein binding</td>
<td>0.000</td>
</tr>
<tr>
<td>Transcription</td>
<td>0.000</td>
<td>Zinc ion binding</td>
<td>0.000</td>
</tr>
<tr>
<td>Chromosome organization and biogenesis</td>
<td>0.000</td>
<td>Heat shock protein binding</td>
<td>0.000</td>
</tr>
<tr>
<td>Nuclear mRNA splicing, via spliceosome</td>
<td>0.001</td>
<td>Nucleic acid binding</td>
<td>0.000</td>
</tr>
<tr>
<td>Anti-apoptosis</td>
<td>0.009</td>
<td>Small GTPase regulator activity</td>
<td>0.003</td>
</tr>
<tr>
<td>DNA repair</td>
<td>0.010</td>
<td>Ubiquitin thiolesterase activity</td>
<td>0.017</td>
</tr>
<tr>
<td>Protein ubiquitination</td>
<td>0.012</td>
<td>Cysteine-type endopeptidase activity</td>
<td>0.021</td>
</tr>
<tr>
<td>Response to virus</td>
<td>0.020</td>
<td>Transcription factor activity</td>
<td>0.023</td>
</tr>
<tr>
<td>Negative regulation of transcription</td>
<td>0.022</td>
<td>Chromatin binding</td>
<td>0.024</td>
</tr>
<tr>
<td>Response to stress</td>
<td>0.029</td>
<td>Binding</td>
<td>0.027</td>
</tr>
<tr>
<td>Regulation of cell cycle</td>
<td>0.029</td>
<td>Helicase activity</td>
<td>0.032</td>
</tr>
<tr>
<td>RNA splicing</td>
<td>0.031</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Understanding How the Subsystems Interconnect

**Protein Folding**
GO:0006457

**DNA Repair**
GO:0006281

**Anti-apoptosis**
GO:0006916

**Nuclear mRNA Splicing**
GO:0000398
The Ultimate Goal

Apply Genomic Tools in an Integrated Approach
The Ultimate Goal

Apply Genomic Tools in an Integrated Approach
Acknowledgements

- Research Associates
  Linda Pluta
  Dana Stanley

- Postdocs
  Tom Halsey
  Todd Page

- Bioinformatics
  Longlong Yang

- Senior Management
  William Greenlee
  Dave Dorman
  Mel Andersen

- Funding
  American Chemistry Council
  NIEHS

- Collaborators
  John Hogenesch (Scripps)
  John Walker (GNF)
  Russ Wolfinger (SAS)
  Jeff MacDonald (UNC)
  Tom Kodadek (UTSMS)
  Devanjan Sikder (UTSMS)