Weighted Gene Co-Expression Network Analysis (WGCNA) as a Bridge for Extrapolation Between Species.

Improving interpretation of nonclinical results using modularity to reduce complexity without loss of biological information.

Jim Stevens  
*Distinguished Research Fellow*  
*Lilly Research Laboratory*  
stevens_james_l@lilly.com
Multi-Scale Modeling of Pathophysiology

Liver Response
Liver Function Gate
Lobule Response
Multicellular Signals
Cell Fate Gate
Cellular Response
Signal Integration
Signal Generation

Organ/Tissue
Organism

Pathophysiological Complexity

Dynamic Complexity (dose/time)

Compound class specific
Tissue stereotypic
←adaptive/progressive→

Multiscale/hybrid model

Cell fate gate
HPC-hepatocellular
NPC-nonparenchymal

Chemical
Biochemical
Organelle
Cell (HPC)
Multi-cell (HPC+NPC)

protein damage
autophagy
cell death
NPC activation
cholestasis; fibrosis

Signal Generation
Signal Transduction
Signal Integration

Δ initial conditions
-hypertrophic-
normal
-damaged-

min-> hrs
hrs-> weeks
days-> months

Biological Complexity

←adaptive/progressive→

Liver Function Gate
Liver Response
Lobule Response
Cell Fate Gate
Cellular Response
Signal Integration
Signal Generation

Liver
HPC
HPC/NPC

Chemical
Biochemical
Organelle
Cell (HPC)
Multi-cell (HPC+NPC)

protein damage
autophagy
cell death
NPC activation
cholestasis; fibrosis

Δ initial conditions
-hypertrophic-
normal
-damaged-

min-> hrs
hrs-> weeks
days-> months

Biological Complexity

←adaptive/progressive→

Liver Function Gate
Liver Response
Lobule Response
Cell Fate Gate
Cellular Response
Signal Integration
Signal Generation

Liver
HPC
HPC/NPC

Chemical
Biochemical
Organelle
Cell (HPC)
Multi-cell (HPC+NPC)

protein damage
autophagy
cell death
NPC activation
cholestasis; fibrosis

Δ initial conditions
-hypertrophic-
normal
-damaged-

min-> hrs
hrs-> weeks
days-> months

Biological Complexity

←adaptive/progressive→

Liver Function Gate
Liver Response
Lobule Response
Cell Fate Gate
Cellular Response
Signal Integration
Signal Generation

Liver
HPC
HPC/NPC

Chemical
Biochemical
Organelle
Cell (HPC)
Multi-cell (HPC+NPC)

protein damage
autophagy
cell death
NPC activation
cholestasis; fibrosis

Δ initial conditions
-hypertrophic-
normal
-damaged-

min-> hrs
hrs-> weeks
days-> months

Biological Complexity

←adaptive/progressive→
Overview

♦ **Modeling biological complexity**
  • The modular nature of complex systems.
  • Leveraging modular systems models using gene expression data.
  • Translating gene expression data into biological understanding.
  • Knowing what we don’t know.

♦ **Molecular Pathogenesis**
  • Correlating expression modules with pathology.
  • Closing the loop from transcription factor to pathogenesis.
  • Predicting adaptive vs progressive responses.

♦ **Applications**
  • Separating injury signals from tissue stereotypic response.
  • Understanding *in vitro* models.
  • Translation to human.
Biological Systems are Modular Across Scales of Complexity

♦ **Modularity** refers to “…pattern[s] of connectedness in which elements of a system (e.g. mRNAs) are grouped into highly connected subsets.”
  - Modules can be arranged in hierarchies using looser connections between modules.

♦ Modular behavior can be captured in unsupervised network models using **coalescent properties** of the system.
  - **Physical** interactions – protein interaction networks
  - **Dynamic** interactions – gene regulatory networks
  - **Statistical** interactions – individual elements connected to phenotype

♦ **Co-regulation** in transcriptional networks is a **coalescent property** of biological systems – networks self-assemble.
  - Connected at level of transcriptional control, e.g. Hox gene networks, Nrf2, etc.
  - Defined/modelled statistically to yield co-expression modules.

♦ Modeling complex systems as networks/modules has advantages:
  - Avoids the ‘curse of dimensionality.’
  - If $2 \times 10^4$ genes form $2 \times 10^2$ modules complexity is reduced by 99%.
  - Biological content is retained.
  - Network visualization applied to modular systems improves data interpretation.
Co-expression modules (genes that respond similarly to drugs): 1 readout per module; the Eigengene (EiG).

9074 liver genes
Form Follows Function

GO-CC (closed) and GO-BP (open)

Color by term
- actin cytoskeleton
- actin cytoskeleton organization
- cell cycle
- condensed chromosome
- endoplasmic reticulum
- extracellular matrix
- extracellular matrix organization
- glutathione biosynthetic process
- mitotic spindle
- proteasomal protein catabolic process
- proteasome complex
- response to endoplasmic reticulum stress
- ribosome
- ribosome biogenesis

Shape by term
- actin cytoskeleton
- actin cytoskeleton organization
- cell cycle
- condensed chromosome
- endoplasmic reticulum
- extracellular matrix
- extracellular matrix organization
- glutathione biosynthetic process
- mitotic spindle
- proteasomal protein catabolic process
- proteasome complex
- response to endoplasmic reticulum stress
- ribosome
- ribosome biogenesis

Histopathology @ 4h

Vacuolation
Infiltration
Hypertrophy-TG
Hematopoiesis
Fibrosis
Degeneration
Apoptosis

Apoptosis
Degeneration
Fibrosis
Hematopoiesis
Hypertrophy-TG
Infiltration
Vacuolation

0 0.5 1 1.5 2 2.5 3
Overview

♦ Modeling biological complexity
  • The modular nature of complex systems.
  • Leveraging modular systems models using gene expression data.
  • Translating gene expression data into biological understanding.
  • Knowing what we don’t know.

♦ Molecular Pathogenesis
  • Correlating expression modules with pathology.
  • Closing the loop from transcription factor to pathogenesis.
  • Predicting adaptive vs progressive responses.

♦ Applications
  • Separating injury signals from tissue stereotypic response.
  • Understanding *in vitro* models.
  • Translation to human.
Increased cJun Binding After Bile Duct Ligation

(a) Percent change for select clinical chemistry endpoints.

(b) Severity scores for histology findings

(c) Scores for selected modules in bile duct ligated rats.

(d) TXG_MAP modules EG scores @1 day...

(e) …and 14 days

(f) ChIP-seq @ 1 and 14 day after ligation. Modules (>20 genes) are ordered by branch.

Sutherland et al. The Pharmacogenomics Journal advance 2017
Location on the TXG_MAP for Modules that are Significant for Adverse Concurrent and Adverse@29d.

**PANELS A and B:** No adjustment for absAveEG to control for the high effect size for adverse concurrent and %DEG.

a) Top-ranked modules selected for effect size >1.0 and p-value <10^{-17}

b) Top-ranked modules selected for effect size >1.0 and p-value <10^{-5}

c) This is outline in detail in Fig 2 of Sutherland et al.

**PANELS C and D:** Modules with p-adjust values were selected to illustrate the effect of controlling for absAveEG as a covariate.

a) Top-ranked modules selected for effect size >1.0 and p-adj <10^{-3}.

b) Top-ranked modules selected for effect sizes >1.0 and p-adj <10^{-3}.

c) This is outline in detail in Fig 2 of Sutherland et al.
Overview

♦ **Modeling biological complexity**
  - The modular nature of complex systems.
  - Leveraging modular systems models using gene expression data.
  - Translating gene expression data into biological understanding.
  - Knowing what we don’t know.

♦ **Molecular Pathogenesis**
  - Correlating expression modules with pathology.
  - Closing the loop from transcription factor to pathogenesis.
  - Predicting adaptive vs progressive responses.

♦ **Applications**
  - Separating injury signals from tissue stereotypic response.
  - Understanding *in vitro* models.
  - Translation to human.
Time Dependence of I/R Liver Injury

88: cell-cell junction; intermediate filament
70: cell cycle arrest
122: Erk1/Erk2 MAPK Pathway
332: protein refolding
336: establishment of epithelial polarity

Best Comparitors:
- Bortezomib r = 0.7
- Cycloheximide r = 0.69
- N,N diethyl nitrosamine r = 0.68
- Phorone r = 0.68

Lateral: ischemic @ 4 hr
Caudal: perfused @ 4 hr

r = 0.51

Best Comparitors:
Partial hepatectomy 6 hr r = 0.92
Partial hepatectomy 3 hr r = 0.87

ISCHM: nonISCHM @ 4 hr r = 0.51
- Lateral difference drivers relate to heat shock, cell cycle arrest, cell-cell junction changes
- Caudal difference drives relate to ribosomal RNA processing
- Evidence of distinct responses

ISCHM: nonISCHM @ 1d r = 0.84
- Similarity driven by DNA replication, ribosomal biogenesis and tubulin formation
- Evidence of convergent pathobiology

MODULES > 20 genes
- 3m: ribosome
- 14m: tubulin folding CCT/Tric
- 37m: DNA replication

117m: RNA processing
125: nucleospeck
1m: KEGG: Spliceosome
198: micro-ribonucleoprotein complex
Effect of placing hepatocytes in culture in the context of ~4000 rat liver experiments

- A: mouse liver vs. TG rat liver
- B: MPH vs mouse liver
- C: 100 mg/kg methapyrilene @ 29 days
- D: 30 mg/kg N-nitrosodiethylamine @ 15 days
- E: HPH vs. human liver
- F: 10 mg/kg cycloheximide @ 9 hrs
- G: HepG2 vs. human liver
- H: TG RPH vs. TG rat liver
- I: DM RPH vs. DM rat liver
- J: 2337 mg/kg aminosalicylic acid @ 1 day
- K: 1 mg/kg bortezomib @ 9 hrs

Average module score
(degree of transcriptional perturbation)

HYPOTHESIS: Probability of Human Liver Toxicity Given Nonclinical Toxicity is a Function of Network EiG, Preservation and Effect Size.

\[ p(hLT|rLT) = f \text{(networks)} \]
\[ p(hLT|rLT) = f \text{(EiG, eff size, preservation)} \]

Where:
- EiG – eigengene score (How much did it change?)
- Eff Size – Is the network associated with adverse outcomes?
- Preservation - Z-score (Is the network preserved in human?)
Comparison of BDH molecular subtypes in rats versus human liver disease.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Rat Scores</th>
<th>Human Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>sub-acute BDH (3-6 hours BDL)</td>
<td>0.0 -0.1</td>
<td>0.4 0.8 0.3 -0.1 0.0 0.0 0.3</td>
</tr>
<tr>
<td>acute BDH (0.5-3 days BDL)</td>
<td>0.5 0.0</td>
<td>0.9 3.4 2.6 0.7 0.9 -1.1 -0.8</td>
</tr>
<tr>
<td>chronic BDH (5-14 days BDL)</td>
<td>0.1 -0.8</td>
<td>1.0 3.2 3.5 3.3 2.2 -1.3 -1.4</td>
</tr>
<tr>
<td>BDH-carcinogens</td>
<td>1.3 1.5</td>
<td>1.5 2.9 3.1 1.5 1.2 -2.7 -2.6</td>
</tr>
<tr>
<td>type 2 diabetes (GSE23343)</td>
<td>0.1 -0.2</td>
<td>0.1 0.7 0.6 0.2 0.4 -0.5 -0.9</td>
</tr>
<tr>
<td>normal liver from obese patients (GSE48452)</td>
<td>-1.4 -1.8</td>
<td>-0.7 -0.4 -0.2 0.1 2.0 0.2 -0.1</td>
</tr>
<tr>
<td>NAFL (GSE48452)</td>
<td>-1.8 -2.1</td>
<td>-1.1 -1.1 -0.4 -0.2 2.1 0.5 0.1</td>
</tr>
<tr>
<td>NASH (GSE48452)</td>
<td>-0.8 -1.7</td>
<td>1.0 1.1 1.1 1.9 2.2 -0.9 -0.4</td>
</tr>
<tr>
<td>biliary atresia (GSE46960)</td>
<td>-0.9 -1.1</td>
<td>0.4 1.5 1.2 3.0 1.0 -1.0 -1.7</td>
</tr>
<tr>
<td>acute liver failure, etiology HBV (GSE38941)</td>
<td>0.1 -0.8</td>
<td>1.4 1.0 3.0 5.3 4.5 -6.7 -10.2</td>
</tr>
<tr>
<td>alcoholic hepatitis (GSE28619)</td>
<td>1.4 1.1</td>
<td>1.5 3.1 2.4 3.1 0.9 -1.4 -0.6</td>
</tr>
<tr>
<td>cirrhosis adjacent to tumor, HCC, etiology HCV (GSE17856)</td>
<td>0.4 -0.4</td>
<td>1.4 0.7 1.7 1.1 1.8 -2.5 -2.1</td>
</tr>
<tr>
<td>HCC, etiology HCV (GSE17856)</td>
<td>2.9 1.5</td>
<td>4.9 -0.7 1.7 -2.8 -1.0 -6.3 -5.4</td>
</tr>
<tr>
<td>HCC, various etiologies (GSE62232)</td>
<td>0.8 0.3</td>
<td>2.3 1.1 1.0 0.3 -0.3 -2.8 -3.2</td>
</tr>
<tr>
<td>Hepatoblastoma (GSE75271)</td>
<td>0.3 0.1</td>
<td>3.9 -1.1 -0.1 1.7 -0.7 -4.8 -9.7</td>
</tr>
</tbody>
</table>

Module scores are averaged across treatments in each BDH subtype (rats) from Figure 5 (top heatmap) and human samples (bottom heatmap) available in each Gene Expression Omnibus (GEO) series, identified via their accession number.
Making the TXG_MAP Public:
Developing an open-source platform for toxicogenomics research

Web-based WGCNA:

- Cloud-hosted platform to access data and computational methods to increase reproducibility and ease of use.
- Collaboration between Indiana Biosciences Research Institute, Dow Agrosciences and Eli Lilly; additional participants welcome
- Current status: proof-of-concept website allowing access to DM, TG data and various analysis methods

Open Source Visualization Tools:

- Shiny R-based visualization approach to WGCNA analysis.
- Collaboration between Lilly and Leiden University with support from EU-ToxRISK and the TransQST projects - European-based.
- Current status: proof-of-concept visualization allowing dendogram visualization and access to enrichment data.
- Includes additional build for human and rat hepatocytes data from TG-Gates.
Making the TXG_MAP Public:
Developing an open-source platform for toxicogenomics research
Multi-Scale Modeling of Pathophysiology

Liver Response
Liver Function Gate
Lobule Response
Multicellular Signals
Cell Fate Gate
Cellular Response
Signal Integration
Signal Generation
Chemical
Biochemical
Organelle
Cell (HPC)
Multi-cell (HPC+NPC)
Organ/Tissue
Organism

Dynamic Complexity (dose/time)

Compound class specific

Tissue stereotypic

← adaptive / progressive →

Multiscale/hybrid model

protein damage

autophagy

cell death

HPC+NPC

NPC activation

cholestasis; fibrosis

HPC-hepatocellular
NPC-nonparenchymal

Liver Response

Min-> hrs

Hrs-> weeks

Days-> months

Biological Complexity
QUESTIONS?