Application of Weighted Gene Co-Expression Network Analysis (WGCNA) to Dose Response Analysis.

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

Chemical
Biochemical
Organelle
Cell (HPC)
Multi-cell (HPC+NPC)
Organ/Tissue
Organism

Dynamic Complexity (dose/time)

Compound class specific

Tissue stereotypic

←adaptive/progressive→

HPC-hepatocellular
NPC-nonparenchymal

Liver -decision gate

Multi-scale/hybrid model

protein damage
autophagy
cell death

HPC+NPC

NPC activation
cholestasis; fibrosis

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.
  - Reducing redundancy in MSigDB.
  - Knowing what we don’t know.

♦ **Understanding Molecular Pathogenesis**
  - Correlating expression modules with pathology.
  - Closing the loop from transcription factor to pathogenesis.
  - Predicting adaptive vs progressive responses.
  - Closing the loop on transcriptional control (addendum slides).

♦ **Applications of WGCNA to Dose Response Analysis**
  - Separating injury signals from tissue stereotypic response.
  - Perturbing network in culture.
  - 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.” (modified from Wagner et al.¹)
  - 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-modeled statistically to yield co-expression modules.

♦ Modeling complex systems as networks/modules has advantages:
  - Avoids the ‘curse of dimensionality.’
  - If 2X10⁴ genes form 2X10² modules complexity is reduced by 99%.
  - Biological content is retained.
  - Network visualization applied to modular systems improves data interpretation.
WGCNA² - Form Follows Function

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

Co-expression modules (genes that respond similarly to drugs): 1 readout per module; the Eigengene (EiG).

WGCNA:85 - Atf4 nutrient depletion

Histopathology @ 4h
Reducing Redundancy of MSigDB Information.

Screen shot from Spotfire TXG_MAP tool for WGCNA:Liver_8, a module highly enriched in genes associated with the proteasome.
Sutherland JJ et al. The Pharmacogenomics Journal advance 2017
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.
  • Reducing redundancy in MSigDB.
  • Knowing what we don’t know.

♦ Understanding Molecular Pathogenesis
  • Correlating expression modules with pathology.
  • Closing the loop from transcription factor to pathogenesis.
  • Predicting adaptive vs progressive responses.
  • Closing the loop on transcriptional control (addendum slides).

♦ Applications of WGCNA to Dose Response Analysis
  • Separating injury signals from tissue stereotypic response.
  • Perturbing network in culture.
  • Translation to human.
Predicting Adverse Responses for Adaptive vs. Progressive Tox-Phenotypes

Panels A and B: Average absolute eigengene score (Abs(aveEig)) for all module changes for three dose levels of acetaminophen (A) and methylenedianiline (B) from TG-GATES 28d repeat dose rat liver studies. Note the scales for A and B differ. The TXG_MAP outputs shown correspond to the highest dose in the upper panel at different time points corresponding to the red circled time points.

PANELS C and D: Top-ranked modules selected for effect size >1.0 and p-adjust <10^-3. are shown for any tox-phenotypes considered adverse and either concurrent (C – present at the same time as gene expression) or predictive of adverse tox-phenotypes occurring at any time later than 1 day (D). Discussed in detail in Fig 2 of Sutherland et al. (2017).
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.
  - Reducing redundancy in MSigDB.
  - Knowing what we don’t know.

♦ **Understanding Molecular Pathogenesis**
  - Correlating expression modules with pathology.
  - Closing the loop from transcription factor to pathogenesis.
  - Predicting adaptive vs progressive responses.
  - Closing the loop on transcriptional control (addendum slides).

♦ **Applications of WGCNA to Dose Response Analysis**
  - Separating injury signals from tissue stereotypic response.
  - Perturbing network in culture.
  - Translation to human.
The Importance of Time Series

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

**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
Perturbations Caused by Placing Hepatocytes in Culture.

Average module score
(degree of transcriptional perturbation)

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

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

\[ p(hLT|rLT) = f(\text{networks}) \]

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

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?)
QUESTIONS?
REFERENCES


2. Horvath, S. et al. Weighted Gene Co-expression Network Analysis (WGCNA) 
http://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/