Differential Network Analysis in Mouse Expression Data

Tova Fuller
Steve Horvath
Department of Human Genetics
University of California, Los Angeles
BIOCOMP'07 Conference, 6/26/07
Outline

• Introduction:
  – Single versus differential network analysis
• Differential Network construction
• Results
• Functional Analysis
• Conclusion
Goals of Single Network Analysis

• Identifying genetic pathways (modules)
• Finding key drivers (hub genes)
• Modeling the relationships between:
  – Transcriptome
  – Clinical traits / Phenotypes
  – Genetic marker data
Validation set 1  
Validation set 2  

Single Network WGCNA  

1 gene co-expression network  
Multiple data sets may be used for validation
Goals of Differential Network Analysis

• Uncover differences in modules and connectivity in different data sets
  – Ex: Human versus chimpanzee brains (Oldham et al. 2006)

• Differing topology in multiple networks reveals genes/pathways that are wired differently in different sample populations

Differential Network WGCNA

2+ gene co-expression networks
Identify genes and pathways that are:
1. Differentially expressed
2. Differentially wired
BxH Mouse Data

• Single network analysis female BxH mice revealed a weight-related module (Ghazalpour et al. 2006)

• **Samples:** Constructed networks from mice from extrema of weight spectrum:
  – Network 1: 30 leanest mice
  – Network 2: 30 heaviest mice

• **Transcripts:** Used 3421 most connected and varying transcripts

Methods

Compute Comparison Metrics
- Difference in expression: t-test statistic
- Compare difference in connectivity: DiffK

Identify significantly different genes/pathways
Permutation test

Functional analysis of significant genes/pathways
DAVID database
Primary literature
Computing Comparison Metrics

**Differential Expression**

$t$-test statistic computed for each gene, $t(i)$

**Differential Connectivity**

\[
K_1(i) = \frac{k_1(i)}{\max(k_1)} \quad K_2(i) = \frac{k_2(i)}{\max(k_2)}
\]

$DiffK(i)$: difference in normalized connectivities for each gene:

\[
DiffK(i) = K_1(i) - K_2(i)
\]
We visualize the comparison metrics via a sector plot:

- x-axis: $\text{DiffK}$
- y-axis: $t$ statistics

We establish sector boundaries to identify regions of differentially expressed and/or connected regions

- $|t| = 1.96$ corresponding to $p = 0.05$
- $|\text{DiffK}| = 0.4$
Permutation test: Identifying significant sectors

\[
p_j = \frac{\text{#}(\text{obs}_j \text{perm}_j) + 1}{\text{no.perms} + 1}
\]
Sector Plot Results

Colored by Network 1 Modules, cor= 0.176

Permutated, cor= -0.0764

0.001 0.001

0.001 0.01
Functional Analysis

**SECTOR 3**
- High t statistic
- High DiffK
- Yellow module in lean
- Grey in obese
- (63 genes)

**SECTOR 5**
- Low t statistic
- High Diff K
- (28 genes)

Genes in these sectors have higher connectivity in lean than obese mice:

\[\sim \text{pathways potentially disregulated in obesity}\ \sim\]
Sector 3: Functional Analysis Results

DAVID Database

• “Extracellular”:
  – extracellular region (38% of genes $p = 1.8 \times 10^{-4}$)
  – extracellular space (34% of genes $p = 5.7 \times 10^{-4}$)
• signaling (36% of genes $p = 5.4 \times 10^{-4}$)
• cell adhesion (16% of genes $p = 7.7 \times 10^{-4}$)
• glycoproteins (34% of genes $p = 1.6 \times 10^{-3}$)
• 12 terms for epidermal growth factor or its related proteins
  – EGF-like 1 (8.2% of genes $p = 8.7 \times 10^{-4}$),
  – EGF-like 3 (6.6% of genes $p = 1.6 \times 10^{-3}$),
  – EGF-like 2 (6.6% of genes $p = 6.0 \times 10^{-3}$),
  – EGF (8.2% of genes $p = 0.013$)
  – EGF_CA (6.6% of genes $p = 0.015$)
Sector 3: Functional Analysis Results
Primary Literature

- Results supported by a study on EGF levels in mice (Kurachi et al. 1993)
  - EGF found to be increased in obese mice
  - Obesity was reversed in these mice by:
    - Administration of anti-EGF
    - Sialoadenectomy

Sector 5:  
Functional Analysis Results  
DAVID Database

- Enzyme inhibitor activity ($p = 2.9 \times 10^{-3}$)*
- Protease inhibitor activity ($p = 6.0 \times 10^{-3}$)
- Endopeptidase inhibitor activity ($p = 6.0 \times 10^{-3}$)
- Dephosphorylation ($p = 0.012$)
- Protein amino acid dephosphorylation ($p = 0.012$)
- Serine-type endopeptidase inhibitor activity ($p = 0.042$)

* $p$ values shown are corrected using Bonferroni correction
Sector 5: Functional Analysis Results
Primary Literature

*Itih1* and *Itih3*
- Enriched for all categories shown previously
- Located near a QTL for hyperinsulinemia (Almind and Kahn 2004)
- *Itih3* identified as a gene candidate for obesity-related traits based on differential expression in murine hypothalamus (Bischof and Wevrick 2005)

*Serpina3n* and *Serpina10*
- Enriched for enzyme inhibitor, protease inhibitor, and endopeptidase inhibitor
- *Serpina10*, or Protein Z-dependent protease inhibitor (ZPI) has been found to be associated with venous thrombosis (Van de Water et al. 2004)

Conclusions

• Differential Network Analysis reveals pathways that are both differentially regulated and connected in mouse obesity
  – Genes that are differentially connected may/may not be differentially expressed

• Primary literature supports biological plausibility of these pathways in weight related disorders
  – Sector 3 & EGF pathways: potential EGF causality in obesity
  – Sector 5 & serine protease pathways: potential link between obesity and venous thrombosis

• These results help identify targets for validation with biological experiments
Acknowledgements

**Guidance**

**HORVATH LAB**
Steve Horvath  
Jason Aten  
Jun Dong  
Peter Langfelder  
Ai Li  
Wen Lin  
Anja Presson  
Lin Wang  
Wei Zhao

**Collaboration**

**LUSIS LAB**
Jake Lusis  
Anatole Ghazalpour  
Thomas Drake

**Funding**

UCLA Medical Scientist Training Program (MD/PhD)

An R tutorial may be found at:
http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/DifferentialNetworkAnalysis