Basic Network Concepts

• Node

• Link
  – Directed or undirected
  – Weighted or unweighted

• Connectivity
  – degree, outdegree, indegree

  – degree, betweenness, closeness, eigenvector (PageRank)

• Module
  – a group of highly interacting nodes
What can we know from networks?

Courtesy Palla et al., Nature 2005
Multiple Scales of Gene Networks

- Global Networks
- Modules/Communities
- Local Subnetworks
- Key Drivers

Association Networks
Regulatory Networks
Multiscale Gene Network Analysis

Association Network

Large-scale Gene Expression Data

Causal Network

Key Driver Analysis

Zhang B et al., Cell (2013)
Zhang B and Jun Zhu, WCE (2013)
Yang X*, Zhang B* et al., Genome Res. (2010)
Key Driver Analysis

1. **Gene Set (G)**
   - **Network (N)**
   - **Define a subnetwork ($N_G$)**
     - Are the nodes of $N_G$ in the set G?
       - **Yes**
         - **Search each node’s neighborhoods for the one most enriched for the list**
         - **Identify candidate drivers by FET**
         - **Identify global & local drivers**
         - **Promote MEN outliers as global**
         - **Output key drivers & subnetwork**
         - **Visualize key drivers & subnetwork (via Cytoscape plugin)**
       - **No**
         - **Find each node’s neighborhood**
         - **Identify candidate drivers by NBH**
         - **Identify global & local drivers**
         - **Promote hub nodes as global**
         - **Output key drivers & subnetwork**
         - **Visualize key drivers & subnetwork (via Cytoscape plugin)**

   Dynamic NGH Search

   Static NGH Search
Dynamic Neighborhood Search (DNS)

- **Input**
  - a set of nodes \( (G) \)
  - a directed/undirected network \( (N) \)

- **Procedure**
  - Generate a sub-network \( N_G \), defined as the set of nodes in \( N \) that are no more than \( h \)-layers away from the nodes in \( G \).
  - Search the \( h \)-layer neighborhood \( (h=1,..,H) \) for each gene in \( N_G \) \( (HLN_{g,h}) \) for the optimal \( h^* \), such that
    \[
    ES_{h^*} = \max(ES_{h,g}) \forall g \in N_G, h \in \{1..H\}
    \]
    where \( ES_{h,g} \) is the computed enrichment statistic for \( HLN_{g,h} \)
  - A node becomes a candidate driver if its HLN is significantly enriched for the nodes in \( G \)
  - Candidate drivers without any parent node (i.e., root nodes in directed networks) are designated as global drivers and the rest are local drivers.
Static Neighborhood Search (SNS)

• Input
  – a set of nodes \(G\)
  – a directed/undirected network \(N\)

• Procedure
  – Compute the size of the h-layer neighborhood (HLN) for each node.
  – Let \(\mu\) be the sizes of HLNs and \(d\) be the out-degrees for all the nodes.
  – Nodes with HLN sizes greater than \(\bar{\mu} + \sigma(\mu)\) are considered as candidate drivers.
  – Candidate drivers without any parent node (i.e., root nodes in directed networks) are designated as global drivers and the rest are local drivers.
  – Promote hub nodes nodes with out-degrees above \(\bar{d} + 2\sigma(d)\) as global drivers.
Application I: Key Drivers of eQTL Hotspots

- A genotypic and expression data from a yeast cross of 112 segregants constructed from the BY and RM strains of *S. cerevisiae*
- Expression quantitative trait loci (eQTL) analysis identified 13 chromosomal regions harboring a large number of eQTL, i.e., eQTL hot spots.
- A Bayesian network reconstructed by integrating genotypic, gene expression, protein-protein interaction and transcription factor binding site (TFBS) data remains the most predictive (Zhu J, Zhang B et al., Nat Genet 2008)
- We apply DNS to identify key drivers of the eQTL hotspots, as oppose to the original static search.
### Key Drivers of eQTL Hotspots

#### eQTL Hotspots

<table>
<thead>
<tr>
<th>eQTL hotspot</th>
<th>Hotspot chr.</th>
<th>Hotspot base-pair position</th>
<th>the original KDA (Zhu, Zhang et al. 2008)</th>
<th>KDA L1</th>
<th>KDA L2</th>
<th>KDA L3</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>560000</td>
<td>TBS1, TOS1, ARA1, CSH1, SUP45, CNS1, AMN1</td>
<td>TBS1, ARA1, CSH1, SUP45, CNS1, PWP2</td>
<td>TBS1, TOS1, ARA1, CSH1, SUP45, CNS1, ENP2, NOP7</td>
<td>TBS1, TOS1, ARA1, CSH1, SUP45, CNS1, NMD3, RPF1</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>1.00E+05</td>
<td>LEU2, ILV6, NFS1, CIT2, MATAHPA1</td>
<td>LEU2, BAP2, OAC1</td>
<td>BAP2, LEU2, OAC1, RTG3</td>
<td>LEU2, BAP2, OAC1, RTG3</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>230000</td>
<td>MATAHPA1</td>
<td>MATAHPA1</td>
<td>MATAHPA1</td>
<td>MATAHPA1</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>130000</td>
<td>URA3</td>
<td>URA3</td>
<td>URA3</td>
<td>URA3</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>130000</td>
<td>GPA1</td>
<td>GPA1</td>
<td>GPA1</td>
<td>GPA1</td>
</tr>
<tr>
<td>8</td>
<td>12</td>
<td>680000</td>
<td>HAP1</td>
<td>HAP1</td>
<td>HAP1</td>
<td>HAP1</td>
</tr>
<tr>
<td>9</td>
<td>12</td>
<td>107000</td>
<td>YRF1-4, YRF1-5, YLR464W</td>
<td>YRF1-4</td>
<td>YRF1-4</td>
<td>YRF1-4</td>
</tr>
<tr>
<td>11</td>
<td>14</td>
<td>503000</td>
<td>SAL1, TOP2</td>
<td>SAL1, RSM24, RSM25, MRPL3</td>
<td>SAL1, RSM24, RSM25, MRPL3</td>
<td>SAL1, RSM24, RSM25, MRPL3</td>
</tr>
<tr>
<td>12</td>
<td>15</td>
<td>180000</td>
<td>PHM7</td>
<td>PHM7, TFS1, YGR043C, HXT7, TKL2, GDB1, YGR052W, GDB1</td>
<td>PHM7, TFS1, YGR043C, HXT7, TKL2, GDB1, YGR052W, GDB1</td>
<td>PHM7, TFS1, YGR043C, HXT7, TKL2, GDB1, YGR052W, GDB1</td>
</tr>
<tr>
<td>10</td>
<td>13</td>
<td>70000</td>
<td>GCV1</td>
<td>GCV1</td>
<td>GCV1</td>
<td>GCV1</td>
</tr>
<tr>
<td>13</td>
<td>15</td>
<td>590000</td>
<td>ATP5</td>
<td>ATP5, ATP20</td>
<td>ATP5, ATP20</td>
<td>ATP5, ATP20</td>
</tr>
</tbody>
</table>
New Key Drivers for eQTL Hotspot 11

- SAL1 is the only one predicted by the original approach.
- KDA identified three new regulators, RSM24, RSM25 and MRPL3.
- 98 genes on the Hotspot 11 are the downstream of SAL1 (p<7e-95).
- 142 genes on the Hotspot 11 are the downstream of RSM25 (p<1.42e-114).
Key Drivers of eQTL Hotspot 12

- PHM7 is the only regulator identified by the original approach.
- KDA uncovered 6 new regulators.
- The neighborhoods of TFS1, YGR043C, TKL2 and YGR052W are more significantly enriched for the genes links to the hot spot than that of PHM7.
## Application II: Inflammatome Driver Genes

<table>
<thead>
<tr>
<th>Disease</th>
<th>Model</th>
<th>Species</th>
<th>Tissue profiled</th>
<th># of Cases</th>
<th># of Controls</th>
<th># of Total Arrays</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asthma</td>
<td>OVA</td>
<td>Mouse</td>
<td>Lung</td>
<td>5</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>COPD</td>
<td>IL-1b Tg</td>
<td>Mouse</td>
<td>Lung</td>
<td>5</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Fibrosis</td>
<td>TGFb Tg</td>
<td>Mouse</td>
<td>Lung</td>
<td>4</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Atherosclerosis</td>
<td>ApoE KO HFD</td>
<td>Mouse</td>
<td>Aorta</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Diabetes</td>
<td>db/db</td>
<td>Mouse</td>
<td>Adipose</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Diabetes</td>
<td>db/db</td>
<td>Mouse</td>
<td>Islet</td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Obesity</td>
<td>ob/ob</td>
<td>Mouse</td>
<td>Adipose</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Multiple</td>
<td>LPS</td>
<td>Rat</td>
<td>Liver</td>
<td>4</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Stroke</td>
<td>MCAO</td>
<td>Rat</td>
<td>Brain</td>
<td>4</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Neuropathic pain</td>
<td>Chung</td>
<td>Rat</td>
<td>DRG</td>
<td>4</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Inflammation pain</td>
<td>CGN</td>
<td>Rat</td>
<td>Skin</td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Sarcopenia</td>
<td>Aged vs. Young</td>
<td>Rat</td>
<td>Muscle</td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
</tbody>
</table>

### Consistency of up- and down-regulated genes in 12 disease models.

<table>
<thead>
<tr>
<th>Gene Set</th>
<th># of Gene</th>
<th>Overlap</th>
<th>Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up-regulated</td>
<td>83</td>
<td>208</td>
<td>704</td>
</tr>
<tr>
<td>Down-regulated</td>
<td>36</td>
<td>78</td>
<td>639</td>
</tr>
</tbody>
</table>

### Gene Ontology Categories enriched in the Inflammatome Signature.

<table>
<thead>
<tr>
<th>Category</th>
<th>Similar Set: Up-regulated</th>
<th>Similar Set: Down-regulated</th>
</tr>
</thead>
<tbody>
<tr>
<td>inflammatory response</td>
<td>4.76E-61</td>
<td>3.32E-11</td>
</tr>
<tr>
<td>leukocyte activation</td>
<td>2.13E-32</td>
<td>1.34E-08</td>
</tr>
<tr>
<td>regulation of immune response</td>
<td>1.44E-25</td>
<td>4.03E-06</td>
</tr>
<tr>
<td>cytokine production</td>
<td>6.10E-18</td>
<td>1.30E-05</td>
</tr>
<tr>
<td>Chemotaxis</td>
<td>4.97E-16</td>
<td>9.18E-05</td>
</tr>
<tr>
<td>humoral immune response</td>
<td>3.25E-14</td>
<td>0.000122</td>
</tr>
</tbody>
</table>

Inflammatome Networks in Human Liver and Adipose

## Inflammatome Signature and Drivers versus MGI Phenotype Database

![Bar chart showing percentage of genes with MGI phenotype(s) for different groups.

<table>
<thead>
<tr>
<th>Group</th>
<th>No. of genes</th>
<th>No. of gene tested in the MGI phenotype database</th>
<th>No. of genes with MGI phenotype(s)</th>
<th>% tested genes with phenotype(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>top 55 key drivers</td>
<td>55</td>
<td>19</td>
<td>14</td>
<td>73.7</td>
</tr>
<tr>
<td>key drivers</td>
<td>151</td>
<td>44</td>
<td>28</td>
<td>63.6</td>
</tr>
<tr>
<td>local drivers</td>
<td>212</td>
<td>57</td>
<td>33</td>
<td>57.9</td>
</tr>
<tr>
<td>non-drivers</td>
<td>2098</td>
<td>609</td>
<td>239</td>
<td>39.2</td>
</tr>
</tbody>
</table>


Application III: Drivers of Cancer Genomic Alterations

- Tumor outcome:
  - Good: free metastasis for \( \geq 5 \) years
  - Bad: metastasis for \( < 5 \) years
- 109 recurrent regions of inferred CNV identified from 4 data sets

Causal Network of Breast Cancer Metastasis

- Global Driver
- Global Driver (validated)
- Local Driver
- Local Driver (validated)
Validation of Causal Drivers

A

Fold Enrichment

- nonRR: 0.9
- RR Gain: 2.1 *
- non Drivers: 2 *
- Drivers: 6.8 *
- Local Drivers: 4.8 *
- Global Drivers: 7.8 *

B

Fold Enrichment

- nonRR: 0.9
- RR Gain: 1.8 *
- non Drivers: 1.7 *
- Drivers: 5.4 *
- Local Drivers: 4.4 *
- Global Drivers: 5.9 *
Application IV: Key Driver Genes of Alzheimer’s Disease

Relevance of the Predicted Drivers to AD

• The 13 well known AD susceptibility genes (Bertram, McQueen et al. 2007)
• Seven (7) of the 13 known AD susceptibility genes were included in the multi-tissue network
• Three (3) of the 7 genes, CST3, PSEN1 and TF, are the predicted drivers, representing a 5.5-fold enrichment (P=1.06e-3) while the rest four are not drivers, i.e., they are underrepresented in the non-driver genes (0.62 fold-enrichment, P=0.99).
• The predicted drivers are 9 times more likely to be the known AD susceptibility genes than the non-driver genes.
Validation of TYROBP Networks

- full-length TYROBP specific target
- truncated TYROBP specific target
- common target
Summary

• An algorithm (KDA) was developed to identify key drivers of biological networks based on various centrality measurements.

• Key regulators predicted by KDA appear to be biologically more important than non-drivers.

• Many key regulators predicted by KDA have been validated at various stages in complex human diseases.

• More comprehensive network analysis methods need be explored to further understand the complexity of biological networks and their underlying biology.
Acknowledgements

Sage Bionetworks

- Justin Guinney
- Dave A Henderson
Openings for Postdoctoral Fellow & Senior Scientist Positions in Computational Neuroscience & Cancer Biology

http://research.mssm.edu/multiscalenetwork/Opportunities.html
Stay Connected with Us!

@IcahnInstitute