An integrative network analysis for the new target identification of Crohn’s Disease

Jing Wang, Ph. D.
May 14, 2018
Background

Crohn’s disease (CD) is caused by a complex interaction among host genetic background, microbial shifts, and environmental cues.

Low remission rate because of heterogeneity of pathogenic mechanisms

<table>
<thead>
<tr>
<th>Drug Type</th>
<th>Overall rate: N (%)</th>
<th>Range (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systemic steroids in population-based cohorts</td>
<td>95/183 (52)</td>
<td>48-58</td>
</tr>
<tr>
<td>Systemic steroids in randomised, controlled trials</td>
<td>79/132 (60)</td>
<td>47-83</td>
</tr>
<tr>
<td>Methotrexate</td>
<td>52/135 (39)</td>
<td>19.3-66.7</td>
</tr>
<tr>
<td>Infliximab</td>
<td>27/83 (32)</td>
<td>25-48</td>
</tr>
<tr>
<td>Adalimumab</td>
<td>58/225 (26)</td>
<td>18-36</td>
</tr>
<tr>
<td>Certolizumab pegol</td>
<td>109/550 (20)</td>
<td>19-23</td>
</tr>
</tbody>
</table>

Peyrin-Biroulet et al. 2011
Identification of causal variations based on the genomics data

**Genome-wide Association Study (GWAS)**

Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk. Genome-wide meta-analysis increases to 71 the number of confirmed loci. A Genome-Wide Scan of Ashkenazi Jewish Crohn’s Disease Suggests Novel Susceptibility Loci. Host–microbe interactions have shaped the genetic estimation and partitioning of (co)heritability of inflammatory bowel disease from GWAS and immunochip data. Over 200 IBD-associated loci have been identified.

Only 26% CDs are contributed by these known genetic variants.
Identification of regulators based on the transcriptomics data

Genetics
- tag SNP
- e.g. rs8099917
- tag SNP

Causal variant(s)
- e.g. rs12379860

Gene (e.g. IL-28B)
- e.g. Chromosome 11

Environment
- e.g. Smoking
- e.g. Dietary intake
- e.g. Antibiotics

Microbiome

eQTL

Regulators

Causal Reasoning

Downstream gene expression

Phenotype
### Data sets with inconsistent conditions

<table>
<thead>
<tr>
<th>Data</th>
<th>Patient</th>
<th>Tissue</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE57945</td>
<td>Treat-naïve pediatric patients</td>
<td>Ileum biopsies</td>
</tr>
<tr>
<td>GSE83687</td>
<td>Patients with advanced disease</td>
<td>Ileum, colon Surgical specimens</td>
</tr>
<tr>
<td>GSE10083</td>
<td>Patients refractory to anti-TNF treatment</td>
<td>Ileum, colon, rectum</td>
</tr>
</tbody>
</table>

Different types of data sets

Conserved regulators
Module analysis and module-specific regulators

Regulator1  Regulator2  Regulator3

Module-specific regulators (Segal et al. 2003)
Analysis Workflow

Diverse data → Module Identification → Conserved modules

Conserved Molecular Regulators

MetaBase network
<table>
<thead>
<tr>
<th>Studies</th>
<th>Data Type</th>
<th>#Sample</th>
<th>Disease Type</th>
<th>Tissues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal A</td>
<td>RNAseq</td>
<td></td>
<td>CD, UC</td>
<td>colon</td>
</tr>
<tr>
<td>Internal B</td>
<td>RNAseq</td>
<td></td>
<td>CD, UC</td>
<td>colon</td>
</tr>
<tr>
<td>GSE16879</td>
<td>Microarray</td>
<td>25</td>
<td>CD, TNFR/NR</td>
<td>colon</td>
</tr>
<tr>
<td>GSE57945</td>
<td>RNAseq</td>
<td>319</td>
<td>Treat-naïve Pediatric CD/UC</td>
<td>Ileum</td>
</tr>
<tr>
<td>GSE83687</td>
<td>RNAseq</td>
<td>134</td>
<td>Advanced IBD CD, UC</td>
<td>Ileum, colon</td>
</tr>
<tr>
<td>GSE10083</td>
<td>Microarray</td>
<td>118</td>
<td>TNF IR CD</td>
<td>Ileum, colon, rectum</td>
</tr>
</tbody>
</table>
WGCNA Analysis Workflow

**Construct network**
Tools: Pearson correlation, Soft thresholding
Rationale: make use of interaction patterns between genes

**Identify modules**
Tools: TOM, Hierarchical clustering
Rationale: module- (pathway-) based analysis

**Find conserved modules**
Tools: Hypergeometric test for each pair of modules
Rationale: Conserved modules should have significant partners in at least two other datasets

Adapted from WGCNA tutorial
Steve Horvath, UCLA
WGCNA results

<table>
<thead>
<tr>
<th>Data</th>
<th>#Modules</th>
<th>#Conserved Modules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal A</td>
<td>28</td>
<td>8</td>
</tr>
<tr>
<td>Internal B</td>
<td>24</td>
<td>10</td>
</tr>
<tr>
<td>GSE16879</td>
<td>73</td>
<td>14</td>
</tr>
<tr>
<td>GSE57945</td>
<td>17</td>
<td>3</td>
</tr>
<tr>
<td>GSE83687</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>GSE10083</td>
<td>43</td>
<td>6</td>
</tr>
</tbody>
</table>
### Functional annotations of WGCNA modules

<table>
<thead>
<tr>
<th>Biological Function</th>
<th>Internal A</th>
<th>Internal B</th>
<th>GSE16879</th>
<th>GSE83687</th>
<th>GSE10083</th>
<th>GSE57945</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inflammation</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Neuronal Signaling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Metabolism</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Fibrosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proliferation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aldosterone Signaling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tissue Factor/Angiogenesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tissue Junction</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unassigned</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

**Metabolic Alterations to the Mucosal Microbiota in Inflammatory Bowel Disease**

**The Influence of the Gut Microbiome on Obesity, Metabolic Syndrome and Gastrointestinal Disease**
1. Change the network to a computational causal graph

2. Predict the causal relationship

\[ S_h^+ := \{v \in T(G_C) | d(h,v) \leq \Delta, d(h,v) < d(h,-v) \} \]
\[ S_h^- := \{v \in T(G_C) | d(h,-v) \leq \Delta, d(h,-v) < d(h,v) \} \]
\[ S_h^0 := \{v \in T(G_C) | d(h,v) > \Delta \lor d(h,v) = d(h,-v) \} \]
\[ s(h,G^\pm) = (|S_h^+ \cap G^+| + |S_h^- \cap G^-|) - (|S_h^+ \cap G^-| + |S_h^- \cap G^+|) \]

Chindelevitch et al. 2012
Regulator identification

Conserved Modules
- Module1
- Module2
- Module3
- ModuleN

Regulatory network

Causal Reasoning
- Regulator A, B
- Regulator A, C
- Regulator D, E
- Regulator G, H

510 Regulators
Regulator ranking

**Ranking criteria**

- Conserved regulators
- Crohns SNP loci gene
  
  Host-microbe interactions have shaped the genetic architecture

510 regulators

**Key genes in the context specific networks**

**Differentially expressed genes/proteins**

**TNF (Tumor Necrosis Factor)**

**OSM (Oncostatin M)**

*Oncostatin M drives intestinal inflammation and predicts response to tumor necrosis factor–neutralizing therapy in patients with inflammatory bowel disease*
Conclusion

- Integrative network analysis identified common key modules across multiple independent diseases datasets.
- Combination of a number of network and module analysis algorithms (WGCNA, Causal reasoning, ARACHNE) identified 510 regulators based on the common modules and regulatory network.
- Additional ranking criteria were used to help Target Validation group to prioritize the list of target candidates.
Acknowledgement

Sherry Cao
Roger Liu
Mike Macoritto
Li Li

Thank you for your attention!