Protein interaction networks and signaling pathways

Anthony Gitter
Biostatistics and Medical Informatics, UW-Madison
Morgridge Institute for Research

April 11, 2019

These slides, excluding third-party material, are licensed under CC BY-NC 4.0 by Anthony Gitter
Networks from spreadsheet-scale data

RNA-seq

Proteomics

Phosphoproteomics

Goal: Prioritize testable hypotheses

RNA-seq

Proteomics

Phosphoproteomics
Case study: systems biology of latent KSHV infection
Kaposi’s Sarcoma-associated Herpesvirus (KSHV)

- KSHV typically in dormant (latent) state
  - Apoptosis ↓
  - Immune response ↓
  - Cell growth ↑
  - Angiogenesis ↑
Compare host cell responses in KSHV and mock infections

RNA-seq
Proteomics
Phosphoproteomics

Sychev et al. *PLoS Pathogens* 2017
Compare host cell responses in KSHV and mock infections

RNA-seq

Proteomics

Phosphoproteomics

How do we interpret the systems-scale “omic” data?
Interpretation idea 1

Find our favorite gene on the list

<table>
<thead>
<tr>
<th>GeneID</th>
<th>Gene</th>
<th>logFC</th>
<th>PValue</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1059</td>
<td>KLHL6</td>
<td>0.8737</td>
<td>0.0002</td>
<td>0.01</td>
</tr>
<tr>
<td>1060</td>
<td>SSX2IP</td>
<td>0.8672</td>
<td>0.01</td>
<td>0.0543</td>
</tr>
<tr>
<td>1061</td>
<td>STAT3</td>
<td>0.8484</td>
<td>7E-11</td>
<td>4E-09</td>
</tr>
<tr>
<td>1062</td>
<td>HPSE</td>
<td>0.8497</td>
<td>0.0419</td>
<td>0.1584</td>
</tr>
<tr>
<td>1063</td>
<td>LINC00665</td>
<td>0.83</td>
<td>0.0986</td>
<td>0.2877</td>
</tr>
<tr>
<td>1064</td>
<td>PCDHB15</td>
<td>0.8801</td>
<td>0.0028</td>
<td>0.0197</td>
</tr>
<tr>
<td>1065</td>
<td>MTUS1</td>
<td>0.841</td>
<td>1E-07</td>
<td>3E-06</td>
</tr>
<tr>
<td>1066</td>
<td>PIK3C2B</td>
<td>0.8708</td>
<td>0.0003</td>
<td>0.0034</td>
</tr>
<tr>
<td>1067</td>
<td>FZD5</td>
<td>0.878</td>
<td>0.0039</td>
<td>0.0256</td>
</tr>
<tr>
<td>1068</td>
<td>BTN3A1</td>
<td>0.8755</td>
<td>7E-07</td>
<td>2E-05</td>
</tr>
<tr>
<td>1069</td>
<td>MDK</td>
<td>0.887</td>
<td>3E-06</td>
<td>6E-05</td>
</tr>
</tbody>
</table>
Interpretation idea 2

Find our favorite enriched pathway or Gene Ontology term

<table>
<thead>
<tr>
<th>GeneID</th>
<th>Gene</th>
<th>GO:Biological Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>443</td>
<td>PRUNE2</td>
<td>apoptotic process</td>
</tr>
<tr>
<td>1264</td>
<td>TRIM69</td>
<td>apoptotic process</td>
</tr>
<tr>
<td>1483</td>
<td>MAGEH1</td>
<td>apoptotic process</td>
</tr>
<tr>
<td>2058</td>
<td>GRAMD4</td>
<td>apoptotic process</td>
</tr>
<tr>
<td>2548</td>
<td>CLPTM1L</td>
<td>apoptotic process</td>
</tr>
<tr>
<td>4984</td>
<td>TMEM219</td>
<td>apoptotic process</td>
</tr>
<tr>
<td>5098</td>
<td>PPP3R1</td>
<td>apoptotic process</td>
</tr>
<tr>
<td>5207</td>
<td>BCL7C</td>
<td>apoptotic process</td>
</tr>
<tr>
<td>5754</td>
<td>EMC4</td>
<td>apoptotic process</td>
</tr>
<tr>
<td>5892</td>
<td>ZNF346</td>
<td>apoptotic process</td>
</tr>
<tr>
<td>6165</td>
<td>RNF34</td>
<td>apoptotic process</td>
</tr>
</tbody>
</table>
Interpretation idea 3

Infer Bayesian network or dependency network

3 mock replicates, 3 KSHV replicates

RNA-seq  Proteomics  Phosphoproteomics
Connecting the dots with protein interactions

Protein complex

Two or more proteins physically bind

Post-translational modification

One protein chemically modifies another
Connecting the dots with protein interactions

Post-translational modification

Complex

Construct network

Protein-protein interaction (PPI) network
Interpretation idea 4

Use prior knowledge of PPI to find relationships among KSHV-related proteins

Condition-specific protein information

Generic PPI network
https://github.com/fraenkel-lab/OmicsIntegrator

From networks to pathways

1. Assign costs to edges based on reliability
2. Find the minimum cost subgraph that connects all nodes

Protein-protein interactions
Steiner tree problem

1. Designate special nodes based on KSHV scores
2. Assign costs to edges based on reliability
3. Collect special nodes while minimizing edge costs
Prize-collecting Steiner forest

1. Assign prizes to proteins based on KSHV scores
2. Assign costs to edges based on reliability
3. Collect prizes while minimizing edge costs

What are the prizes and costs?
KSHV score for proteins

• Proteomics: small p-value → high score

• RNA-seq: infer transcription factor (TF) activities

Mock  KSHV  Mock  KSHV

High TF score  Low TF score
PPI network costs for edges

### SLN1-YPD1:

Cost = 1 – confidence score

#### Evidence

<table>
<thead>
<tr>
<th>Source DB</th>
<th>Source ID</th>
<th>Interaction Type</th>
<th>PSI MI Code</th>
<th>PubMed ID</th>
<th>Detection Type</th>
<th>PSI MI Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>intact</td>
<td>EBI-1037552</td>
<td>direct interaction</td>
<td>MI:0407</td>
<td>14656441</td>
<td>phage display</td>
<td>MI:0084</td>
</tr>
<tr>
<td>intact</td>
<td>EBI-1037552</td>
<td>direct interaction</td>
<td>MI:0407</td>
<td>14656441</td>
<td>demethylase assay</td>
<td>MI:0870</td>
</tr>
<tr>
<td>biogrid</td>
<td>147693</td>
<td>direct interaction</td>
<td>MI:0407</td>
<td>15628880</td>
<td>t7 phage display</td>
<td>MI:0108</td>
</tr>
<tr>
<td>bind</td>
<td>169826</td>
<td>molecular interaction</td>
<td>MI:0000</td>
<td>8808622</td>
<td>unknown</td>
<td>unknown</td>
</tr>
<tr>
<td>biogrid</td>
<td>143168</td>
<td>direct interaction</td>
<td>MI:0407</td>
<td>8808622</td>
<td>fluorescence technology</td>
<td>MI:0051</td>
</tr>
<tr>
<td>biogrid</td>
<td>147695</td>
<td>direct interaction</td>
<td>MI:0407</td>
<td>9882653</td>
<td>t7 phage display</td>
<td>MI:0108</td>
</tr>
<tr>
<td>biogrid</td>
<td>264432</td>
<td>direct interaction</td>
<td>MI:0407</td>
<td>17559414</td>
<td>t7 phage display</td>
<td>MI:0108</td>
</tr>
<tr>
<td>biogrid</td>
<td>147692</td>
<td>direct interaction</td>
<td>MI:0407</td>
<td>15628880</td>
<td>t7 phage display</td>
<td>MI:0108</td>
</tr>
<tr>
<td>bind_translation</td>
<td>169826</td>
<td>molecular interaction</td>
<td>MI:0000</td>
<td>8808622</td>
<td>electron tomography</td>
<td>MI:0410</td>
</tr>
</tbody>
</table>
Prize-collecting Steiner forest

1. Assign prizes to proteins based on KSHV scores
2. Assign costs to edges based on reliability
3. Collect prizes while minimizing edge costs

How do we find an approximate solution?
Prize-collecting Steiner forest objective

3. Collect prizes while minimizing edge costs

Find forest (network)

Minimize

\[ o(F) = \lambda \sum_{v \notin V_F} p(v) + \sum_{e \in E_F} c(e) \]

Controls tradeoff between prizes and costs

Prizes of excluded vertices (nodes)

Costs of included edges
Constructing a probabilistic approximation algorithm

How many possible subgraphs?
How can we perform inference in that state space?

msgsteiner algorithm Bailly-Bechet et al. *PNAS* 2011
Transform global tree constraint into local constraints

Add Root node (R) and Null node (*)
Transform global tree constraint into local constraints

Create two variables for every node \( i \):

- \( p_{a_i} \) – parent node
- \( d_i \) – depth from root
Translate omitted prize cost into edge cost

\[ p_{a_i} = * \]
\[ d_i = \infty \]
\[ c_{i,*} = \lambda \, p(v_i) \]

Node excluded from tree incurs cost equal to its prize
Computing objective function with new variables

\[ H(d, pa) = \sum_{v_i \in V} c_{i, pa_i} \]

Minimize subject to tree constraint

\( p_{ai} = \text{parent} \)
\( d_i = \text{depth} \)
\( c_{i,j} = \text{edge cost} \)
Probabilistic optimization

Boltzmann-Gibbs distribution

\[ P(d, pa) \propto e^{-\beta H(d, pa)} \]

Statistical mechanics: cavity method
Probabilistic graphical models: belief propagation
Iterative optimization with belief propagation

Node $i$ receives messages from neighbors

Node $i$ updates state of $pa_i$ and $d_i$
Combining forests into pathways
Returning to the latent KSHV infection study
Latent KSHV “pathway”
Zooming in on regions of interest

Peroxisome subnetwork

KSHV latency gene
Proteomic
p-Proteomic
TF
Steiner node

Up
Down

Outlined KEGG pathway

April 11, 2019
STAT/BMI 877: Statistical Methods for Molecular Biology
Peroxisome validation

- **Peroxisome**: organelle involved in metabolism, very long chain fatty acids

Peroxisome counts increase upon infection

Knocking down peroxisomal enzymes increases cell death only in KSHV-infected cells
Related resources
Extensions to the Prize Collecting Steiner Forest approach

- Multi-task network modeling
- Temporal information on nodes
- New implementations