Combating Viruses by Targeting Host Proteins

T. M. Murali

Department of Computer Science
Virginia Tech
murali@cs.vt.edu
http://bioinformatics.cs.vt.edu/~murali

Introduction to Computational Biology and Bioinformatics (CS 3824)
Oct 18, 20, 2011
Infectious Diseases

- Tenacious and major public-health problem all over the world.
- Second leading cause of death after cardiovascular disease.
- New infectious diseases are emerging and old ones are re-emerging.
- Rate of development of new medicines is flat.
Pathogens are Becoming Drug-Resistant

Pssst! Hey kid! Wanna be a Superbug...?
Stick some of this into your genome...
Even penicillin won't be able to harm you...

It was on a short-cut through the hospital kitchens that Albert was first approached by a member of the Antibiotic Resistance.
Motivation HDFs Algorithms Cross Validation Predictions

Pathogens are Becoming Drug-Resistant

Evidence of Artemisinin-Resistant Malaria in Western Cambodia

TO THE EDITOR: Although artemisinins are potent and rapidly acting antimalarial drugs, their widespread use for treating patients with Plasmodium falciparum malaria raises the question of emerging drug resistance. Artemisinin monotherapy should not be used in areas where malaria is endemic; it requires an extended admission (30 mg per kilogram per day) or quinine (25 mg per kilogram per day) in a split dose every 8 hours for 7 days (54 patients). The study was approved by ethics review committees in Cambodia and the United States and was conducted from October 2006 through March 2007. Written informed consent was obtained.

Our Pigs, Our Food, Our Health

One of the many industrial hog farms outside Camden, Ind.

By NICHOLAS D. KRISTOF
Published: March 11, 2009

Once-a-day HIV drug cocktail — in one pill — wins FDA approval

WASHINGTON — The federal government on Wednesday approved the first HIV treatment that packs a triple-drug cocktail into a once-a-day pill.

Doctors say the salmon-colored pill will vastly simplify AIDS care and turn what a few years ago was a bothersome regimen of 20 or 30 tablets to one pill taken before bed.

To be sold as Atripla, the pill includes doses of three drugs now sold in the US by two companies. The drugs are Bristol-Myers Squibb’s Sustiva and GlaxoMajordamo’s Truvada, a combo of Viread and Emtriva.

Taking the trio as a single pill makes it less likely that patients will miss doses, which would allow the virus to rebound and become resistant to treatment, doctors say. Keeping the virus in check also helps lower the risk that a patient will infect someone else.

New Wave of Antibiotic-Resistant Bacteria

"New strains of 'Gram-negative' bacteria have become resistant to all safe antibiotics. Though methicillin-resistant Staphylococcus aureus (MRSA) is the best-known antibiotic-resistant germ, the new class of resistant bacteria could be more dangerous still. The bacteria, classified as Gram-negative because of their reaction to the so-called Gram stain test, can cause severe pneumonia and infections of the urinary tract, bloodstream, and other parts of the body. Their cell structure makes them more difficult to attack with antibiotics than Gram-positive organisms like MRSA. The only antibiotics - colistin and polymyxin B - that still have efficacy against Gram-negative bacteria produce dangerous side effects: kidney damage and nerve damage. Patients who are infected with Gram-negative bacteria must make the unsavory choice between life with kidney damage or death with intact kidneys. Recently, some new strains of Gram-negative bacteria have shown resistance against even colistin and polymyxin B. Infection with these new strains typically means death for the patient."
One-Bug-One-Drug

- Drugs against infectious diseases target pathogen proteins.
- Targeted proteins can mutate and make pathogen drug resistant.
- Problem is especially acute for RNA viruses such as HIV-1.
One-Bug-One-Drug ⇒ Many-Bugs-One-Drug

- Drugs against infectious diseases target pathogen proteins.
- Targeted proteins can mutate and make pathogen drug resistant.
- Problem is especially acute for RNA viruses such as HIV-1.
- Develop drugs that target human proteins.
- Prioritize human proteins interacting with multiple pathogens.
Viral Dependency Factors

- RNA viruses like HIV have very few genes.
- Viral dependency factor (VDF): *human* protein that virus needs to replicate and propagate.
RNA viruses like HIV have very few genes.

Viral dependency factor (VDF): *human* protein that virus needs to replicate and propagate.

Recent genome-wide experiments have discovered dozens of VDFs for HIV, flu virus, West Nile virus, Hepatitis C virus, and the bacterium that causes Tuberculosis.

HDF: HIV dependency factor.
Three Screens for HDFs

- Use RNAi to silence each human gene in turn and measure the extent to which HIV propagates.
- Brass et al., Science, February 2008: 275 HDFs
- König et al., Cell, October 2008: 296 HDFs
- Zhou et al., Cell Host and Microbe, November 2008: 375 HDFs
(Brass et al., Science., 2008)
Three Screens for HDFs

- Use RNAi to silence each human gene in turn and measure the extent to which HIV propagates.
- Brass et al., Science, February 2008: 275 HDFs
- König et al., Cell, October 2008: 296 HDFs
- Zhou et al., Cell Host and Microbe, November 2008: 375 HDFs
# Overlap Between the Three Screens

<table>
<thead>
<tr>
<th>Table Name (Size)</th>
<th>siRNA HIV König (293)</th>
<th>siRNA HIV Brass (283)</th>
<th>siRNA HIV Zhou (303)</th>
<th>siRNA HIV Fellay (63)</th>
<th>Particle Associated HIV (248)</th>
<th>HARC Nef (6)</th>
<th>HARC Tat (69)</th>
<th>HARC Rev (56)</th>
<th>BIND INT HIV (23)</th>
<th>NCBI Interactions (1,434)</th>
<th>siRNA Flu Fly (98)</th>
</tr>
</thead>
<tbody>
<tr>
<td>siRNA HIV König (293)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>siRNA HIV Brass (283)</td>
<td>$&lt;0.001$ (13)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>siRNA HIV Zhou (303)</td>
<td>0.024 (9)</td>
<td>$&lt;0.001$ (18)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>siRNA HIV Fellay (63)</td>
<td>1 (0)</td>
<td>0.511 (1)</td>
<td>0.541 (1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Particle Associated HIV (248)</td>
<td>0.154 (5)</td>
<td>0.035 (6)</td>
<td>0.07 (6)</td>
<td>0.108 (2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HARC Nef (6)</td>
<td>1 (0)</td>
<td>1 (0)</td>
<td>1 (0)</td>
<td>1 (0)</td>
<td>1 (0)</td>
<td>$&lt;0.001$ (2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HARC Tat (69)</td>
<td>1 (0)</td>
<td>0.004 (4)</td>
<td>0.052 (3)</td>
<td>1 (0)</td>
<td>0.027 (3)</td>
<td>1 (0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HARC Rev (56)</td>
<td>0.125 (2)</td>
<td>0.44 (1)</td>
<td>0.469 (1)</td>
<td>1 (0)</td>
<td>$&lt;0.001$ (10)</td>
<td>1 (0)</td>
<td>1 (0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIND INT HIV (23)</td>
<td>$&lt;0.001$ (3)</td>
<td>1 (0)</td>
<td>0.232 (1)</td>
<td>1 (0)</td>
<td>0.191 (1)</td>
<td>1 (0)</td>
<td>1 (0)</td>
<td>0.07 (1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCBI Interactions (1,434)</td>
<td>$&lt;0.001$ (53)</td>
<td>$&lt;0.001$ (39)</td>
<td>$&lt;0.001$ (40)</td>
<td>0.234 (5)</td>
<td>$&lt;0.001$ (94)</td>
<td>1 (0)</td>
<td>$&lt;0.001$ (21)</td>
<td>$&lt;0.001$ (21)</td>
<td>0.009 (5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>siRNA Flu Fly (98)</td>
<td>$&lt;0.001$ (13)</td>
<td>0.125 (3)</td>
<td>0.738 (1)</td>
<td>1 (0)</td>
<td>$&lt;0.001$ (9)</td>
<td>1 (0)</td>
<td>1 (0)</td>
<td>0.002 (3)</td>
<td>1 (0)</td>
<td></td>
<td>$&lt;0.001$ (20)</td>
</tr>
<tr>
<td>siRNA WNV (305)</td>
<td>0.02 (8)</td>
<td>0.004 (9)</td>
<td>0.693 (3)</td>
<td>0.14 (2)</td>
<td>0.013 (8)</td>
<td>1 (0)</td>
<td>0.061 (3)</td>
<td>0.481 (1)</td>
<td>1 (0)</td>
<td>0.006 (29)</td>
<td>0.337 (2)</td>
</tr>
</tbody>
</table>

Each entry in the table shows the p-values (determined by comparison to random simulation) and the number of overlapping genes in parenthesis. Set names are as in Table 1. doi:10.1371/journal.ppat.1000437.t002

(Bushman et al., *PLoS Path.*, 2009)
## Overlap Between the Three Screens

<table>
<thead>
<tr>
<th>Table Name (Size)</th>
<th>siRNA HIV König (293)</th>
<th>siRNA HIV Brass (283)</th>
<th>siRNA HIV Zhou (303)</th>
</tr>
</thead>
<tbody>
<tr>
<td>siRNA HIV König (293)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>siRNA HIV Brass (283)</td>
<td>&lt;0.001 (13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>siRNA HIV Zhou (303)</td>
<td>0.024 (9)</td>
<td>&lt;0.001 (18)</td>
<td></td>
</tr>
<tr>
<td>siRNA HIV Fellay (63)</td>
<td>1 (0)</td>
<td>0.511 (1)</td>
<td>0.541 (1)</td>
</tr>
<tr>
<td>Particle Associated HIV (248)</td>
<td>0.154 (5)</td>
<td>0.035 (6)</td>
<td>0.07 (6)</td>
</tr>
<tr>
<td>HARC Nef (6)</td>
<td>1 (0)</td>
<td>1 (0)</td>
<td>1 (0)</td>
</tr>
<tr>
<td>HARC Tat (69)</td>
<td>1 (0)</td>
<td>0.004 (4)</td>
<td>0.052 (3)</td>
</tr>
<tr>
<td>HARC Rev (56)</td>
<td>0.125 (2)</td>
<td>0.44 (1)</td>
<td>0.469 (1)</td>
</tr>
<tr>
<td>BIND INT HIV (22)</td>
<td>≤0.001 (2)</td>
<td>1 (0)</td>
<td>0.222 (1)</td>
</tr>
</tbody>
</table>
### Reasons for Low Overlap

**Motivation** HDFs Algorithms Cross Validation Predictions

---

#### (Goff, *Cell*, 2008)

<table>
<thead>
<tr>
<th>Host Cell Line</th>
<th>Time of siRNA Treatment</th>
<th>Virus Challenge</th>
<th>Time of Scoring Post-infection</th>
<th>Readout</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brass et al. HeLa (CD4+, (\beta)-gal reporter)</td>
<td>72 hr</td>
<td>Live HIV-1 (III B)</td>
<td>48 hr; 48 hr in new cells</td>
<td>p24 (CA); reporter activation</td>
</tr>
<tr>
<td>König et al. 293T</td>
<td>48 hr</td>
<td>HIV-1 luc vector, VSV-G pseudotyped</td>
<td>24 hr</td>
<td>Luc reporter</td>
</tr>
<tr>
<td>Zhou et al. HeLa (CD4+, (\beta)-gal reporter)</td>
<td>24 hr</td>
<td>Live HIV-1 (HXB2 isolate)</td>
<td>48 hr; 96 hr activation</td>
<td>(\beta)-gal reporter</td>
</tr>
</tbody>
</table>
HDFs Screens Uncover Common Pathways

(Bushman et al., PLoS Path., 2009)
Intriguing statement in König et al.:

“...an additional 64 genes reported by Brass et al. directly interact with a confirmed gene in our study.”
Goal of this Work

▶ Intriguing statement in König et al.:

“... an additional 64 genes reported by Brass et al. directly interact with a confirmed gene in our study.”

▶ Hypotheses:

1. We can exploit the social network of genes to predict new HDFs.
2. Combining the HDFs from three screens will improve prediction capability.

▶ Collaboration with Michael Katze (Dept of Microbiology, Univ. of Washington) and Brett Tyler (VBI).
Datasets

- PPI network: 71,461 interactions involving 9,595 proteins (BIND, DIP, HPRD, IntAct, MINT, MIPS, and Reactome).
- Weighting schemes: Unweighted and hypergeometric-based (Goldberg and Roth, *PNAS*, 2003).

<table>
<thead>
<tr>
<th>Study name</th>
<th>#genes</th>
<th>#genes in PPI network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brass (B)</td>
<td>275</td>
<td>157</td>
</tr>
<tr>
<td>Konig (K)</td>
<td>296</td>
<td>199</td>
</tr>
<tr>
<td>Zhou (Z)</td>
<td>375</td>
<td>215</td>
</tr>
<tr>
<td>Brass, Konig, or Zhou (BKZ)</td>
<td>908</td>
<td>545</td>
</tr>
<tr>
<td>Essential genes</td>
<td>483</td>
<td>373</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study name</th>
<th>#genes</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brass</td>
<td>275</td>
<td>0.807</td>
</tr>
<tr>
<td>Konig</td>
<td>296</td>
<td>0.013</td>
</tr>
<tr>
<td>Zhou</td>
<td>375</td>
<td>0.2</td>
</tr>
<tr>
<td>BKZ</td>
<td>908</td>
<td>0.112</td>
</tr>
</tbody>
</table>

Human orthologs of essential genes in mouse.
Datasets

- **PPI network**: 71,461 interactions involving 9,595 proteins (BIND, DIP, HPRD, IntAct, MINT, MIPS, and Reactome).

- **Weighting schemes**: Unweighted and hypergeometric-based (Goldberg and Roth, *PNAS*, 2003).

- **Positive examples**:

<table>
<thead>
<tr>
<th>Study name</th>
<th>#genes</th>
<th>#genes in PPI network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brass (B)</td>
<td>275</td>
<td>157</td>
</tr>
<tr>
<td>Konig (K)</td>
<td>296</td>
<td>199</td>
</tr>
<tr>
<td>Zhou (Z)</td>
<td>375</td>
<td>215</td>
</tr>
<tr>
<td>Brass, Konig, or Zhou (BKZ)</td>
<td>908</td>
<td>545</td>
</tr>
</tbody>
</table>

- **Negative examples**: Human orthologs of essential genes in mouse.
Datasets

- **PPI network**: 71,461 interactions involving 9,595 proteins (BIND, DIP, HPRD, IntAct, MINT, MIPS, and Reactome).

- **Weighting schemes**: Unweighted and hypergeometric-based (Goldberg and Roth, *PNAS*, 2003).

- **Positive examples**:

<table>
<thead>
<tr>
<th>Study name</th>
<th>#genes</th>
<th>#genes in PPI network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brass (B)</td>
<td>275</td>
<td>157</td>
</tr>
<tr>
<td>Konig (K)</td>
<td>296</td>
<td>199</td>
</tr>
<tr>
<td>Zhou (Z)</td>
<td>375</td>
<td>215</td>
</tr>
<tr>
<td>Brass, Konig, or Zhou (BKZ)</td>
<td>908</td>
<td>545</td>
</tr>
<tr>
<td>Essential genes</td>
<td>483</td>
<td>373</td>
</tr>
</tbody>
</table>

- **Negative examples**: Human orthologs of essential genes in mouse.

<table>
<thead>
<tr>
<th>Study name</th>
<th>#genes</th>
<th>#genes that are essential</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brass</td>
<td>275</td>
<td>5</td>
</tr>
<tr>
<td>Konig</td>
<td>296</td>
<td>14</td>
</tr>
<tr>
<td>Zhou</td>
<td>375</td>
<td>12</td>
</tr>
<tr>
<td>BKZ</td>
<td>908</td>
<td>28</td>
</tr>
</tbody>
</table>
Exploit Social Network of Genes

- Exploit network structure to determine whether unknown examples have the same function as the known HDFs or the essential genes.
# Algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Uses negative examples</th>
<th>Parameters</th>
<th>Values tested</th>
</tr>
</thead>
<tbody>
<tr>
<td>SinkSource</td>
<td>Yes</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Local</td>
<td>Yes</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Hopfield</td>
<td>Yes</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Local+</td>
<td>No</td>
<td>No</td>
<td>Identical to Functional Flow with one phase</td>
</tr>
<tr>
<td>SinkSource+</td>
<td>No</td>
<td>Weight of artificial edges</td>
<td>0.01, 0.1, 0.5, 1, 2, 10, and 100</td>
</tr>
<tr>
<td>FunctionalFlow</td>
<td>No</td>
<td>#phases</td>
<td>3, 5, 7</td>
</tr>
</tbody>
</table>
Evaluating Algorithms

- Four datasets (B, K, Z, BKZ) on unweighted and weighted PPI network.
- Two-fold cross validation.
- Averaged over 10 independent runs.
Evaluating Algorithms

- Four datasets (B, K, Z, BKZ) on unweighted and weighted PPI network.
- Two-fold cross validation.
- Averaged over 10 independent runs.
- $k$-fold cross validation:
  1. Partition union of positive and negative examples into $k$ groups, uniformly at random.
  2. For each group, use algorithm to predict the state of each positive/negative example in that group using all other examples.
  3. Sort all positive and negative examples in decreasing order of $s(u)$.
  4. For each threshold on prediction confidence, compute the number of true positives ($tp$), false positives ($fp$), true negatives ($tn$), and false negatives ($fn$).
  5. For each threshold on prediction confidence, compute precision ($tp/(tp + fp)$), recall ($tp/(tp + fn)$), and false positive rate ($fp/(fp + tn)$).
Cross-Validation Results, Unweighted PPI Network

SinkSource, unweighted network

Precision vs. Recall graph for different methods:
- BKZ
- B
- K
- Z

T. M. Murali
CS 3824
Combating Viruses by Targeting Host Proteins
Cross-Validation Results, Unweighted PPI Network

SinkSource+, unweighted network

- BKZ
- B
- K
- Z

Precision vs. Recall graph showing the performance of different algorithms on the PPI network.
Cross-Validation Results, Unweighted PPI Network

Local+, unweighted network

- BKZ
- B
- K
- Z

T. M. Murali
CS 3824
Combating Viruses by Targeting Host Proteins
Cross-Validation Results, Unweighted PPI Network

Area under Precision-Recall Curve, unweighted network

H, L, SS, FF 1, FF 7, L+, SS+

Area under Precision-Recall Curve

B, K, Z, BKZ

T. M. Murali CS 3824 Combating Viruses by Targeting Host Proteins
Cross-Validation Results, Unweighted PPI Network

Brass-Konig-Zhou HDFs, unweighted network

Precision vs Recall plot for various methods:
- SS
- H
- L
- SS+
- FF 7
- L+

T. M. Murali
CS 3824
Combating Viruses by Targeting Host Proteins
Cross-Validation Results, Unweighted PPI Network

Brass-Konig-Zhou HDFs, unweighted network

Precision
Recall
SS
SS+

T. M. Murali
CS 3824
Combating Viruses by Targeting Host Proteins
Cross-Validation Results, Unweighted PPI Network

Functional Flow 1, 3, 5 and 7 phases, unweighted network

Precision vs Recall for different functional flows (FF 1, 3, 5, 7) in a unweighted PPI network.
Cross-Validation Results, Unweighted PPI Network

SinkSource+, unweighted network

Precision vs. Recall graph with different colors and markers for various parameters (0.01, 0.1, 0.5, 1, 2, 10, 100). The graph shows the performance of the algorithm under cross-validation.
Cross-Validation Results, Weighted PPI Network

Area under Precision-Recall Curve, weighted network

T. M. Murali
CS 3824
Combating Viruses by Targeting Host Proteins
Cross-Validation Results, Weighted PPI Network

Area under Precision-Recall Curve, unweighted network

- B
- K
- Z
- BKZ

T. M. Murali
CS 3824
Combating Viruses by Targeting Host Proteins
Summary of Cross-Validation Results

- Performance for BKZ dataset dominates the other three curves at most values of recall.
- SinkSource+, Local+, and Functional Flow with 7 phases achieve highest precision values when recall is less than 20%.
- SinkSource has the best performance for values of recall greater than 20%.
- The two top-performing algorithms (SinkSource, Hopfield) show very little variation in AUPRC values (0.744 to 0.745).
- The results for weighted versions of the network did not substantially differ from those for the unweighted network.
- Further analysis with predictions made by SinkSource+ and by SinkSource.
SS+ and SS Make Some Common Predictions

Jacquard coefficient of predictions made by SinkSource+ and by SinkSource

#predictions
Independent Evaluation of Predictions

- HDFs were discovered in cell lines.
- Protein-protein interactions discovered in a wide variety of experiments.
- These data are from cells that are not the natural environment for HIV.
- Need to evaluate predictions using independent HIV-relevant data.

1. Which biological processes are enriched in BKZ or in predicted HDFs?
2. Do BKZ or predicted HDFs interact with HIV proteins?
3. How do BKZ and predicted HDFs cluster within the human PPI network?
4. What are the expression programs of BKZ or predicted HDFs in infected individuals?
Independent Evaluation of Predictions

- HDFs were discovered in cell lines.
- Protein-protein interactions discovered in a wide variety of experiments.
- These data are from cells that are not the natural environment for HIV.
- Need to evaluate predictions using independent HIV-relevant data.

1. Which biological processes are enriched in BKZ or in predicted HDFs?
2. Do BKZ or predicted HDFs interact with HIV proteins?
3. How do BKZ and predicted HDFs cluster within the human PPI network?
4. What are the expression programs of BKZ or predicted HDFs in infected individuals?
1. Functions Enriched in BKZ and in Predicted HDFs

- Used FuncAssociate software (Berriz et al., Bioinformatics, 2009).
  - BKZ HDFs are unordered.
  - Ordered predicted HDFs by decreasing value of $s(u)$.
- Computed Gene Ontology (GO) terms enriched in these lists.
- Used a $p$-value cutoff of 0.05.
Predicted HDFs are enriched in HIV-related GO terms

- 53 GO terms enriched in BKZ HDFs.
- 1229 GO terms enriched in HDFs predicted by Local+.
- 199/1229 GO terms were enriched among the predicted HDFs with the 1000 highest ranks.
Predicted HDFs are enriched in HIV-related GO terms

- 53 GO terms enriched in BKZ HDFs.
- 1229 GO terms enriched in HDFs predicted by Local+.
- 199/1229 GO terms were enriched among the predicted HDFs with the 1000 highest ranks.
- 218 GO terms split into three classes.
  1. 33 enriched in both BKZ and predicted HDFs: also identified by (Bushman et al., *PLoS Path.*, 2009), proteasome, RNA polymerase, mediator complex, transcriptional elongation, and RNA binding/splicing.
  2. 19 terms enriched in BKZ HDFs but not in predicted HDFs: non-specific GO terms such as protein complex, primary metabolic process, and organelle part.
  3. 166 terms enriched only in predicted HDFs: specializations of non-specific terms, e.g., the Ndc80 complex (GO:0031262) and MIS12/MIND type complex (GO:0000444).
2. Overlap with HIV Interactors

- Do predicted HDFs interact with HIV proteins?
- NCBI HIV database lists 1433 unique human proteins that interact with HIV proteins (HIV interactors).
- 120 BKZ HDFs interact with HIV proteins ($p$-value $1.02 \times 10^{-10}$).
2. Overlap with HIV Interactors

- Do predicted HDFs interact with HIV proteins?
- NCBI HIV database lists 1433 unique human proteins that interact with HIV proteins (HIV interactors).
- 120 BKZ HDFs interact with HIV proteins \((p\text{-value } 1.02 \times 10^{-10})\).

1. Sorted predicted HDFs in decreasing order of \(s(u)\).
2. Computed intersection of HIV interactors with top \(k\) predicted HDFs, for \(k = 100, 200, \ldots\).
3. Computed the \(p\)-value of observed intersection sizes using Fisher’s exact test.
HDFs Predicted by SinkSource+ Interact with HIV

Fraction of BKZ or predicted HDFs that interact with HIV

(a) SinkSource+

(b) SinkSource

T. M. Murali  CS 3824  Combating Viruses by Targeting Host Proteins
HDFs Predicted by SinkSource+ Interact with HIV

p-values of BKZ or predicted HDFs that interact with HIV

SinkSource+

SinkSource
3. Clustering BKZ and Predicted HDFs

- Computed sub-network induced by BKZ and HDFs predicted by SinkSource+: 9,452 nodes and 71,367 PPIs.
- Modification multiplied internal node weight computed by MCODE with weight computed by SinkSource+. 
MCODE Clusters

- MCODE computed 157 clusters, spanning 2,476 nodes (70%) and 22,212 PPIs (44%).
MCODE Clusters

- MCODE computed 157 clusters, spanning 2,476 nodes (70%) and 22,212 PPIs (44%).
- Only 9 MCODE clusters have a statistically significant overlap with BKZ HDFs.

<table>
<thead>
<tr>
<th>Cluster id</th>
<th>#proteins</th>
<th>#BKZ HDFs</th>
<th>Fraction of BKZ HDFs</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>87</td>
<td>17</td>
<td>0.2</td>
<td>$3.5 \times 10^{-4}$</td>
</tr>
<tr>
<td>2</td>
<td>139</td>
<td>24</td>
<td>0.17</td>
<td>$7.2 \times 10^{-5}$</td>
</tr>
<tr>
<td>6</td>
<td>19</td>
<td>9</td>
<td>0.47</td>
<td>$4.9 \times 10^{-6}$</td>
</tr>
<tr>
<td>9</td>
<td>14</td>
<td>7</td>
<td>0.5</td>
<td>$2.5 \times 10^{-5}$</td>
</tr>
<tr>
<td>11</td>
<td>12</td>
<td>6</td>
<td>0.5</td>
<td>$2.9 \times 10^{-32}$</td>
</tr>
<tr>
<td>14</td>
<td>10</td>
<td>5</td>
<td>0.5</td>
<td>$2.1 \times 10^{-3}$</td>
</tr>
<tr>
<td>29</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>$3 \times 10^{-6}$</td>
</tr>
<tr>
<td>87</td>
<td>4</td>
<td>3</td>
<td>0.75</td>
<td>$6.2 \times 10^{-3}$</td>
</tr>
<tr>
<td>90</td>
<td>6</td>
<td>2</td>
<td>0.33</td>
<td>$2.7 \times 10^{-2}$</td>
</tr>
</tbody>
</table>
## GO Terms Enriched in MCODE Clusters

<table>
<thead>
<tr>
<th>Cluster index</th>
<th>#proteins</th>
<th>#PPIs</th>
<th>Enriched GO functions</th>
<th>p-value</th>
<th>#proteins GO term in cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>87</td>
<td>3738</td>
<td>RNA metabolic process</td>
<td>$2.1 \times 10^{-58}$</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Spliceosome</td>
<td>$2.4 \times 10^{-41}$</td>
<td>49</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
<td>1225</td>
<td>Kinetochore</td>
<td>$5.2 \times 10^{-38}$</td>
<td>29</td>
</tr>
<tr>
<td>2</td>
<td>139</td>
<td>2647</td>
<td>Respiratory chain</td>
<td>$4.4 \times 10^{-72}$</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NADH dehydrogenase</td>
<td>$1.9 \times 10^{-52}$</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Proteasome</td>
<td>$8.3 \times 10^{-44}$</td>
<td>32</td>
</tr>
<tr>
<td>3</td>
<td>49</td>
<td>650</td>
<td>Ribosome</td>
<td>$5 \times 10^{-42}$</td>
<td>33</td>
</tr>
<tr>
<td>4</td>
<td>42</td>
<td>467</td>
<td>Ribosome</td>
<td>$1.2 \times 10^{-24}$</td>
<td>24</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>190</td>
<td>small GTPase mediated signal transduction</td>
<td>$3.1 \times 10^{-9}$</td>
<td>19</td>
</tr>
<tr>
<td>6</td>
<td>19</td>
<td>171</td>
<td>RNA elongation</td>
<td>$4.5 \times 10^{-12}$</td>
<td>14</td>
</tr>
<tr>
<td>7</td>
<td>15</td>
<td>105</td>
<td>MHC protein complex</td>
<td>$7.6 \times 10^{-15}$</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>15</td>
<td>102</td>
<td>Pore complex</td>
<td>$10^{-15}$</td>
<td>11</td>
</tr>
</tbody>
</table>
4. Relating HDFs to HIV Progression

- Used gene expression data from SIV-infected non-human primates.
  - African Green Monkeys (AGMs) are natural hosts for SIV.
  - Pig-tailed Macaques (PMs) are susceptible to SIV.
- Lederer et al., *PLoS Path.*, 2009 conducted a longitudinal study of SIV infection in AGMs and PTs:
  - Day 10 and day 45 after SIV infection.
  - Three tissues: blood, colon, and lymph nodes.
- Used ANOVA to compute “diagnostic genes,” the genes that were differentially expressed between AGMs and PMs.
- Computed the statistical significance of the overlap between diagnostic genes and known/predicted HDFs.
Predicted HDFs Overlap with Diagnostic Genes

Fraction of BKZ or predicted HDFs that are differentially expressed
SinkSource+
0.09
0.095
0.1
0.105
0.11

Fraction of HDFs
Rank

Blood (Day 10)

Colon (Day 10)

Lymph node (Day 10)

Blood (Day 45)

Colon (Day 45)

Lymph node (Day 45)
Predicted HDFs Overlap with Diagnostic Genes

Fraction of BKZ or predicted HDFs that are differentially expressed

SinkSource+

Fraction of HDFs

Lymph node (Day 10)
Predicted HDFs Overlap with Diagnostic Genes

Fraction of BKZ or predicted HDFs that are differentially expressed

SinkSource+
 0.09
 0.095
 0.1
 0.105
 0.11

Blood (Day 10)

Colon (Day 10)

Lymph node (Day 10)

Blood (Day 45)

Colon (Day 45)

Lymph node (Day 45)
**Predicted HDFs Overlap with Diagnostic Genes**

*p*-values of BKZ or predicted HDFs that are differentially expressed

\[ \text{SinkSource}^+ \]

- **Blood (Day 10)**
  - \( \text{-log(p-value)} \) vs. Rank
  - No significant \( p \)-values

- **Colon (Day 10)**
  - \( \text{-log(p-value)} \) vs. Rank
  - Significant \( p \)-values

- **Lymph node (Day 10)**
  - \( \text{-log(p-value)} \) vs. Rank
  - \( p \)-values above 1.5 indicate significant expression

- **Blood (Day 45)**
  - \( \text{-log(p-value)} \) vs. Rank
  - No significant \( p \)-values

- **Colon (Day 45)**
  - \( \text{-log(p-value)} \) vs. Rank
  - \( p \)-values above 0.002 indicate significant expression

- **Lymph node (Day 45)**
  - \( \text{-log(p-value)} \) vs. Rank
  - \( p \)-values above 0.02 indicate significant expression

---

**T. M. Murali**

**CS 3824**

**Combating Viruses by Targeting Host Proteins**
Predicted HDFs Overlap with Diagnostic Genes

Fraction of BKZ or predicted HDFs that are differentially expressed

SinkSource

Blood (Day 10)

Colon (Day 10)

Lymph node (Day 10)

Blood (Day 45)

Colon (Day 45)

Lymph node (Day 45)
Conclusions

- HDFs neighbour each other in the human PPI network.
- Clustering can be exploited to predict HDFs with high confidence.
- Predicted HDFs interact with HIV to a much more significant extent than known HDFs.
- Predicted and BKZ HDFs participate in host cellular processes that are subverted by HIV during its life cycle.
- The gene expression programme of predicted HDFs in acute pathogenic SIV infection is significantly different from that during non-pathogenic infection, especially in lymph nodes.
Acknowledgments

- Michael G. Katze
- Brett M. Tyler
- Matthew D. Dyer
- David Badger

- NIH grants P30DA015625, P51RR000166, and R24RR016354 to MGK.
- VBI Fellows programme to TMM and BMT.
- ASPIRES program at VT to TMM
Further Research

► Understand why Local+ has such good performance.
► Can we synthesise ideas like FunctionalFlow for low recall with ideas like SinkSource for high recall?
► Understand off-target effects in siRNA screens. Are there hidden HDFs that are common to all three screens?
► Analyse data for Influenza virus.
► Combine dependency factor datasets to predict broad spectrum drug targets for RNA viruses.