Combating Viruses by Targeting Host Proteins

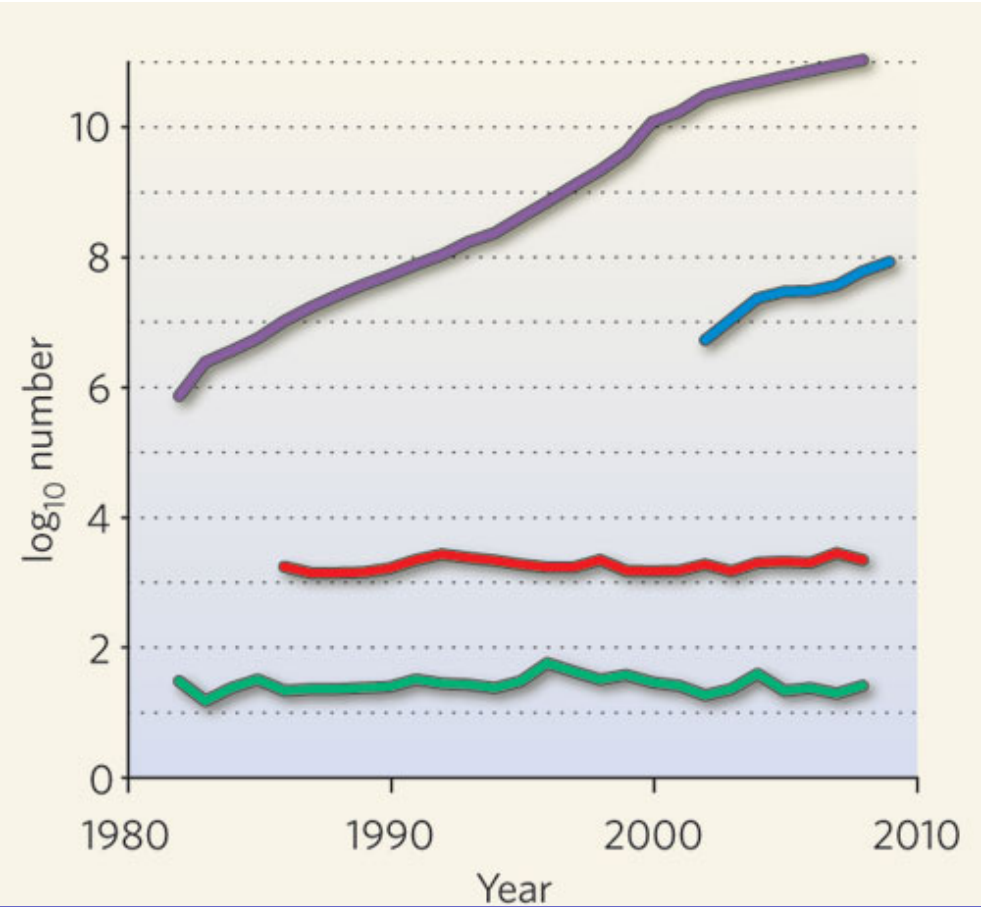
T. M. Murali

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

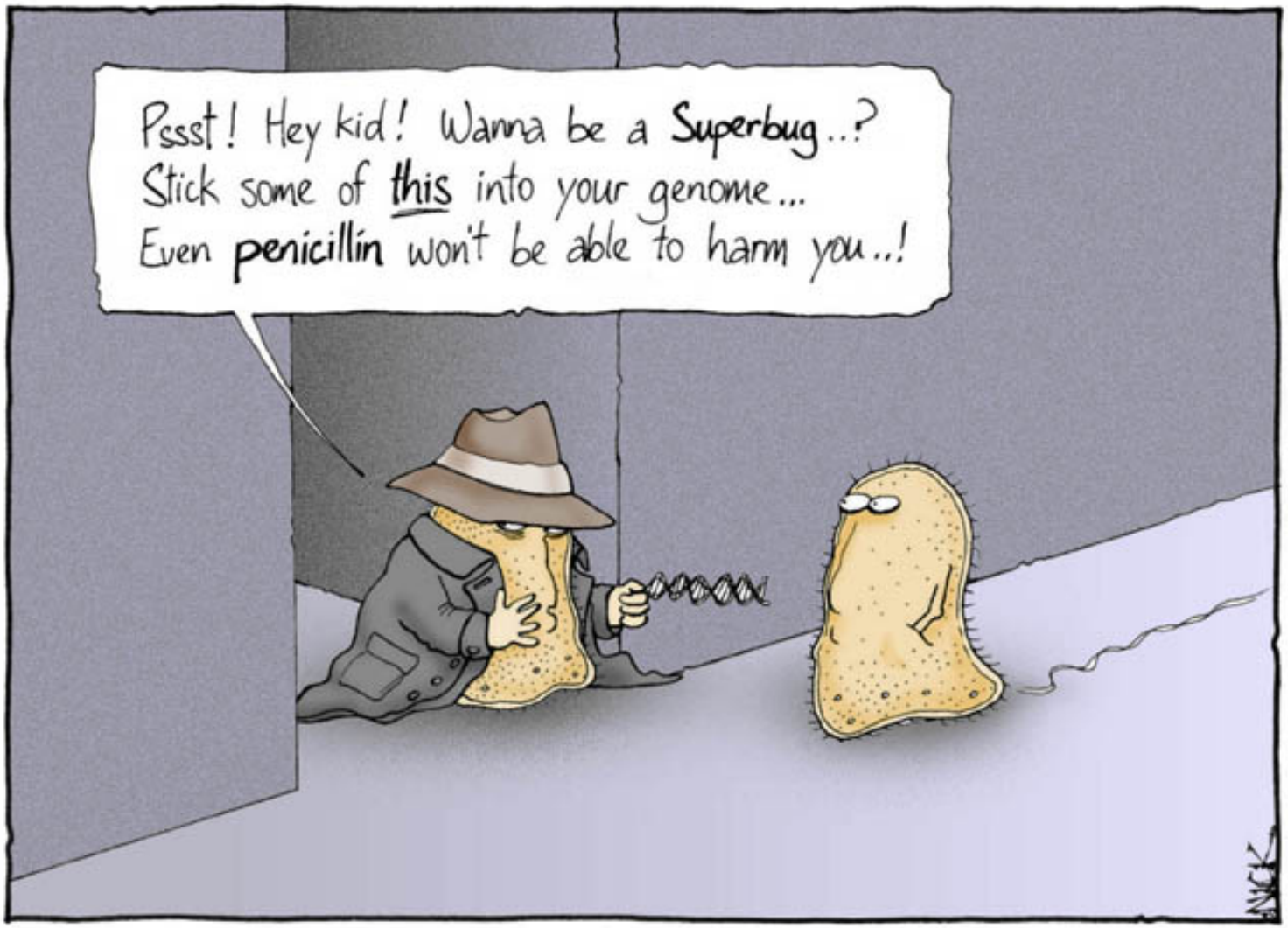
Introduction to Computational Biology and Bioinformatics (CS 3824)
Oct 18, 20, 2011

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



It was on a short-cut through the hospital kitchens that Albert was first approached by a member of the Antibiotic Resistance.

Pathogens are Becoming Drug-Resistant

CORRESPONDENCE

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.[1,2] Artemisinin monotherapy should not be used in areas where malaria is endemic; it requires an extended administration (7 days) or quinine (30 mg per kilogram per day) plus tetracycline (25 mg per kilogram per day) in a split dose every 8 hours for 7 days (34 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

OP-ED COLUMNIST

Our Pigs, Our Food, Our Health

Nicholas D. Kristof/The New York Times

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

By NICHOLAS D. KRISTOF
Published: March 11, 2009

SIGN IN TO

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

Updated 7/16/2006 7:02 AM ET

E-mail | Save | Print | Reprints & Permissions | RSS

Enlarge By Haraz N. Ghanbari, AP

Andrew von Eschenbach, acting director of the Food and Drug Administration, announces a new fixed-dose, once-a-day pill, called Atripla, for the treatment of HIV-positive patients.

By Steve Sternberg, USA TODAY

WASHINGTON — The federal government on Wednesday approved the first HIV treatment that packs a triple-drug cocktail into a one-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 USA by two companies. The drugs are Bristol-Myers Squib's Sustiva and Gilead Pharmaceutical'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.

Slashdot NEWS FOR NERDS. STUFF THAT MATTERS.

Stories Recent Popular Search

Science: New Wave of Antibiotic-Resistant Bacteria

Posted by kdawson on Sunday February 28, @05:57PM
from the gram-of-prevention dept.

reporter writes

> "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.

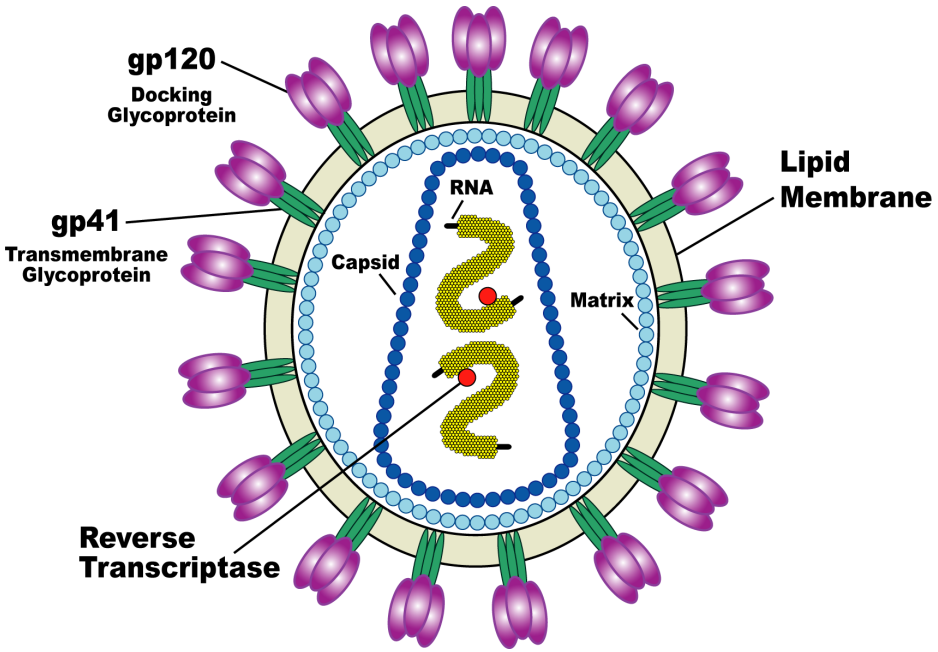
One-Bug-One-Drug $\Rightarrow$ Many-Bugs-One-Drug

- 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.

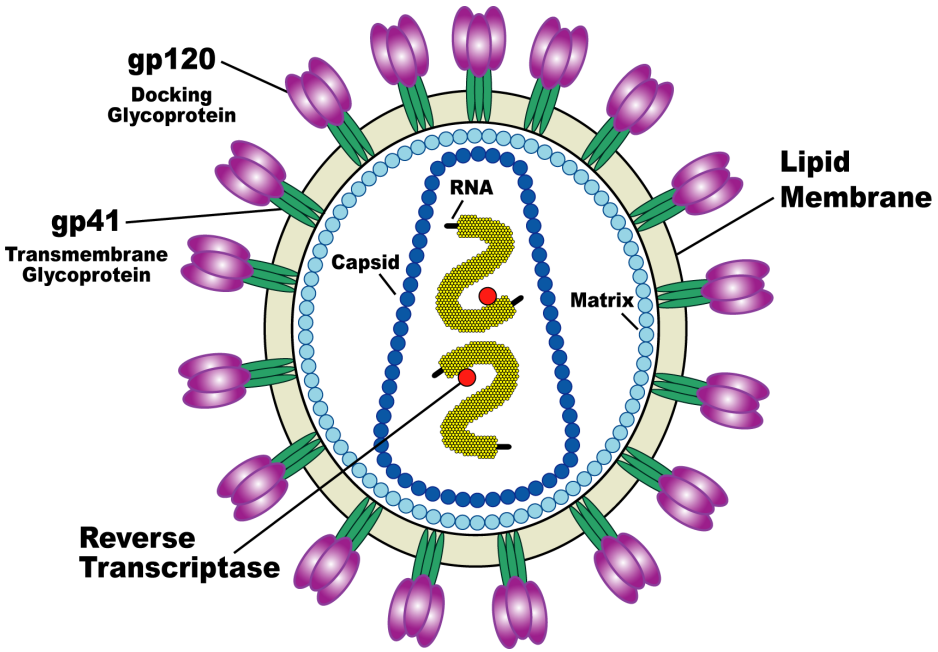
- Pathogens interact with central human proteins in the human protein interaction network (Dyer, Murali, and Sobral, *PLoS Pathog*, 2008).

Viral Dependency Factors



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

Viral Dependency Factors



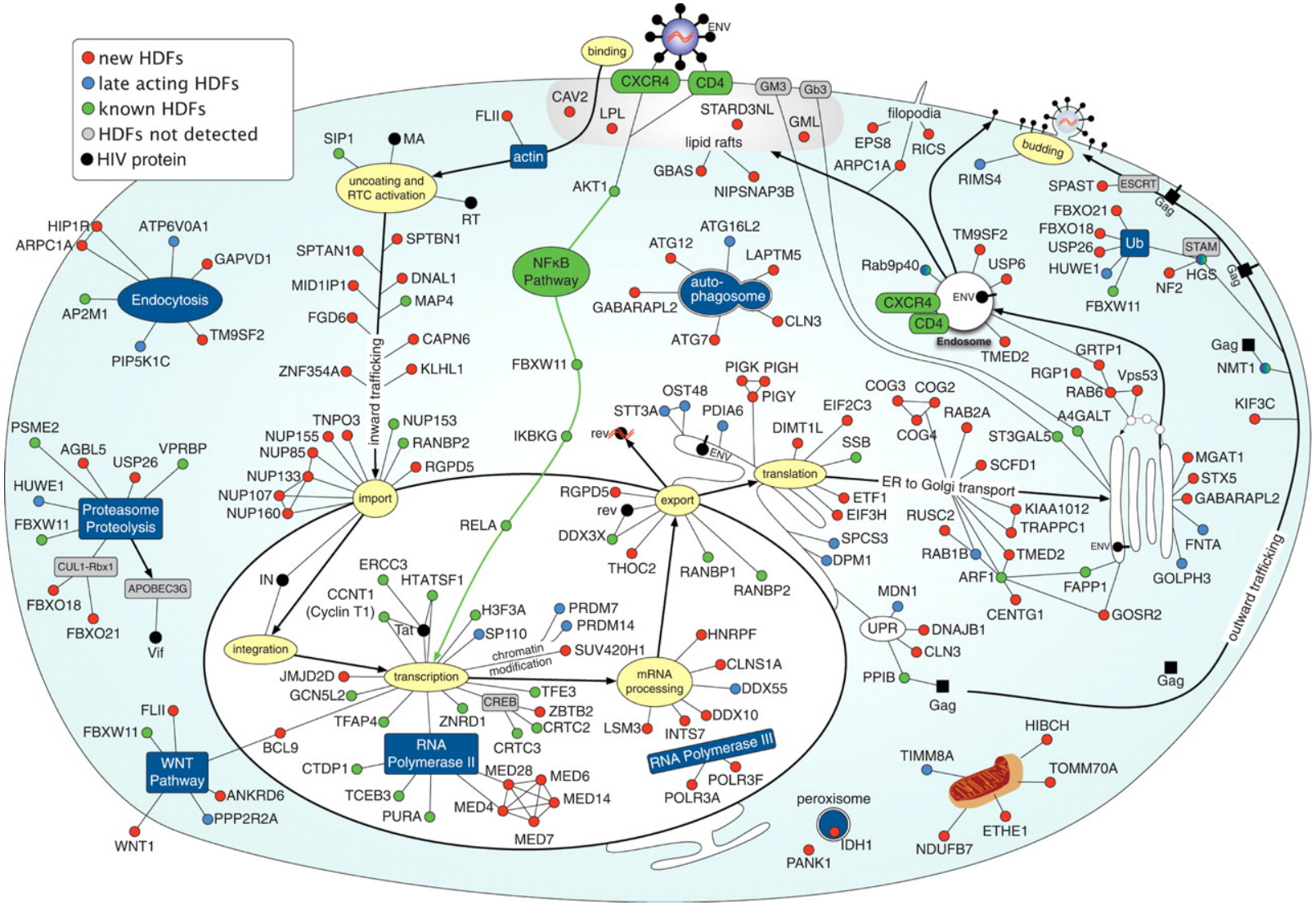
- RNA viruses like HIV have very few genes.
- Viral dependency factor (VDF): *human* protein that virus needs to replicate and propagate.

- 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

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

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

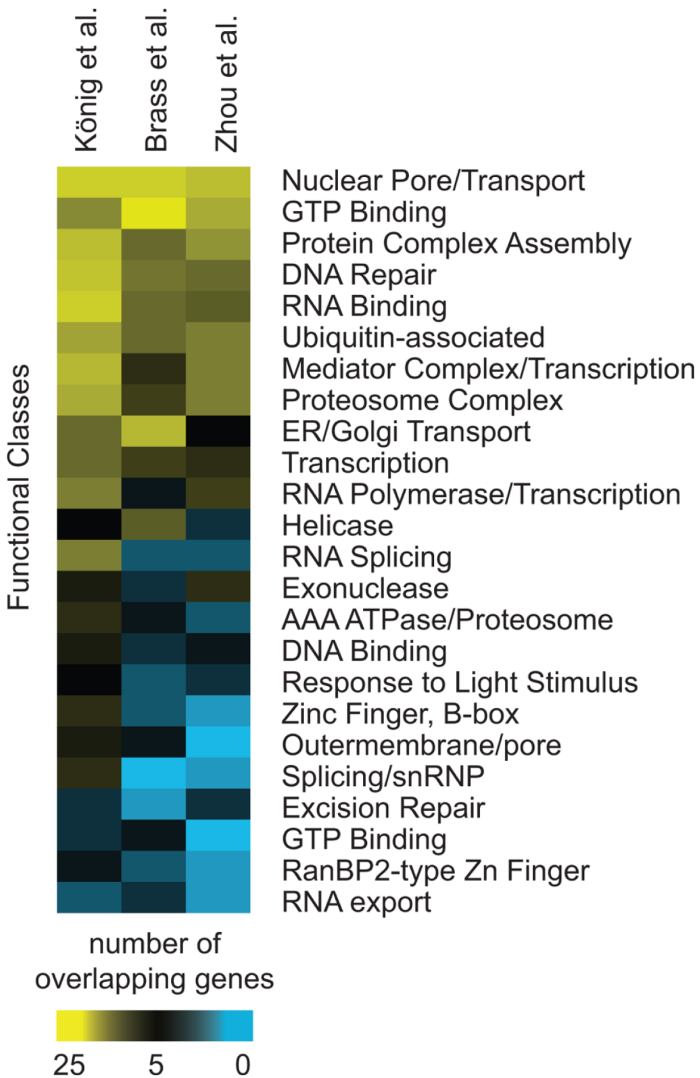
Table Name (Size)	siRNA HIV König (293)	siRNA HIV Brass (283)	siRNA HIV Zhou (303)
siRNA HIV König (293)			
siRNA HIV Brass (283)	$<$0.001 (13)		
siRNA HIV Zhou (303)	0.024 (9)	$<$0.001 (18)	
siRNA HIV Fellay (63)	1 (0)	0.511 (1)	0.541 (1)
Particle Associated HIV (248)	0.154 (5)	0.035 (6)	0.07 (6)
HARC Nef (6)	1 (0)	1 (0)	1 (0)
HARC Tat (69)	1 (0)	0.004 (4)	0.052 (3)
HARC Rev (56)	0.125 (2)	0.44 (1)	0.469 (1)
RIND-INT HIV (33)	$<$0.001 (3)	1 (0)	0.232 (1)

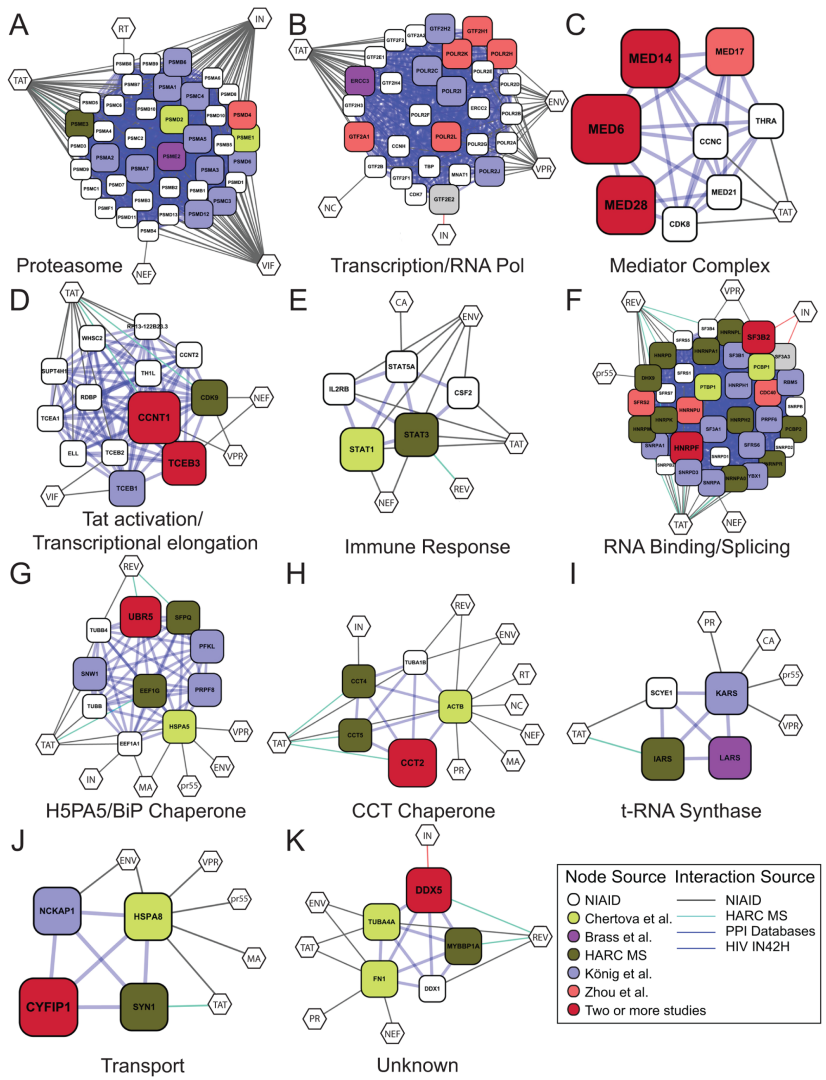
Reasons for Low Overlap

(Goff, *Cell*, 2008)

	Host Cell Line	Time of siRNA Treatment	Virus Challenge	Time of Scoring Post-infection	Readout
Brass et al.	HeLa (CD4+, β-gal reporter)	72 hr	Live HIV-1 (III B)	48 hr; 48 hr in new cells	p24 (CA); reporter activation
König et al.	293T	48 hr	HIV-1 luc vector, VSV-G pseudotyped	24 hr	Luc reporter
Zhou et al.	HeLa (CD4+, β-gal reporter)	24 hr	Live HIV-1 (HXB2 isolate)	48 hr; 96 hr activation	β-gal reporter

HDFs Screens Uncover Common Pathways





(Bushman et al., *PLoS Path.*, 2009)

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."

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 capabililty.
- 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).

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:

Study name	#genes	#genes in PPI network
Brass (B)	275	157
Konig (K)	296	199
Zhou (Z)	375	215
Brass, Konig, or Zhou (BKZ)	908	545

Datasets

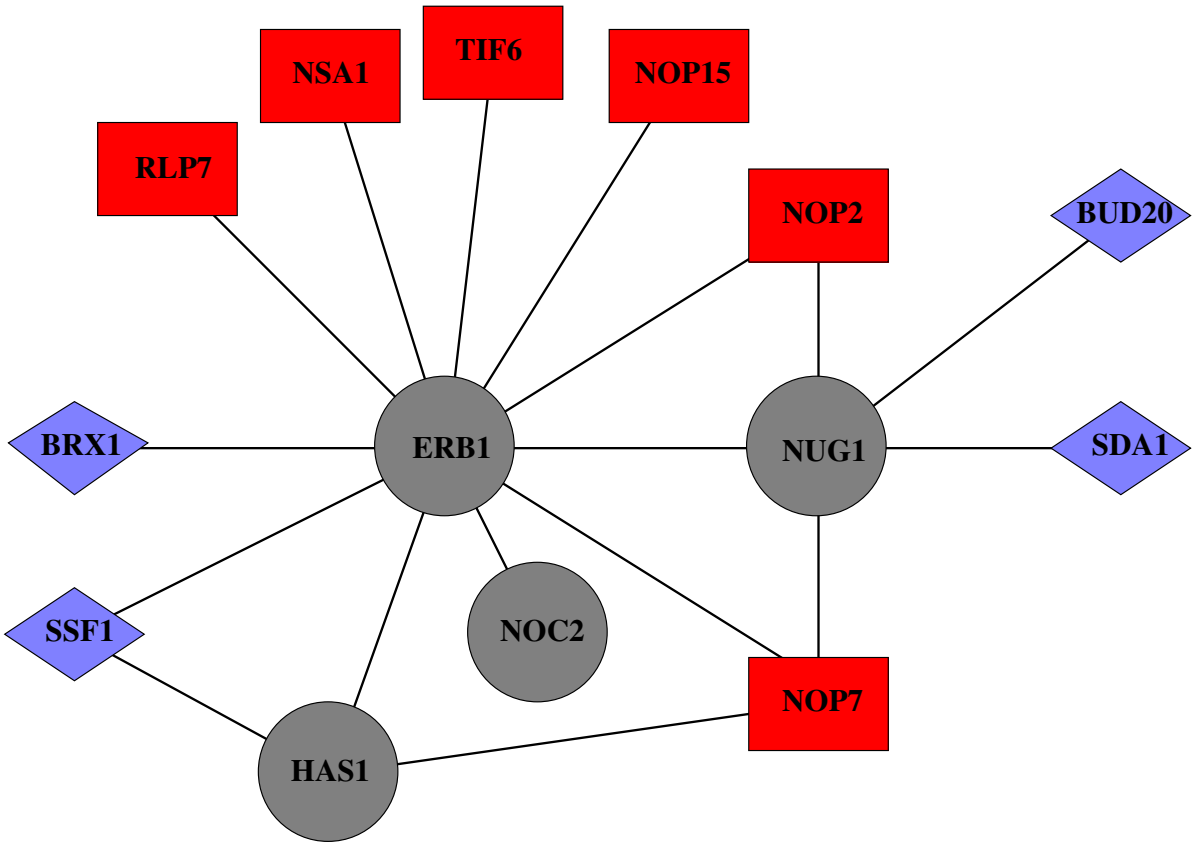
- PPI network: 71,461 interactions involving 9,595 proteins (BIND, DIP, HPRD, IntAct, MINT, MIPS, and Reactome).
- Weighting schemes: Unweighted and hypergeometric-based (Goldberg and Roth, *PNAS*, 2003).
- Positive examples:

Study name	#genes	#genes in PPI network
Brass (B)	275	157
Konig (K)	296	199
Zhou (Z)	375	215
Brass, Konig, or Zhou (BKZ)	908	545
Essential genes	483	373

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

Study name	#genes	#genes that are essential	p-value
Brass	275	5	0.807
Konig	296	14	0.013
Zhou	375	12	0.2
BKZ	908	28	0.112

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

Algorithm	Uses negative examples	Parameters	Values tested
SinkSource	Yes	None	
Local	Yes	None	
Hopfield	Yes	None	
Local+	No	No	Identical to Functional Flow with one phase
SinkSource+	No	Weight of artificial edges	0.01. 0.1, 0.5, 1, 2, 10, and 100
FunctionalFlow	No	#phases	3, 5, 7

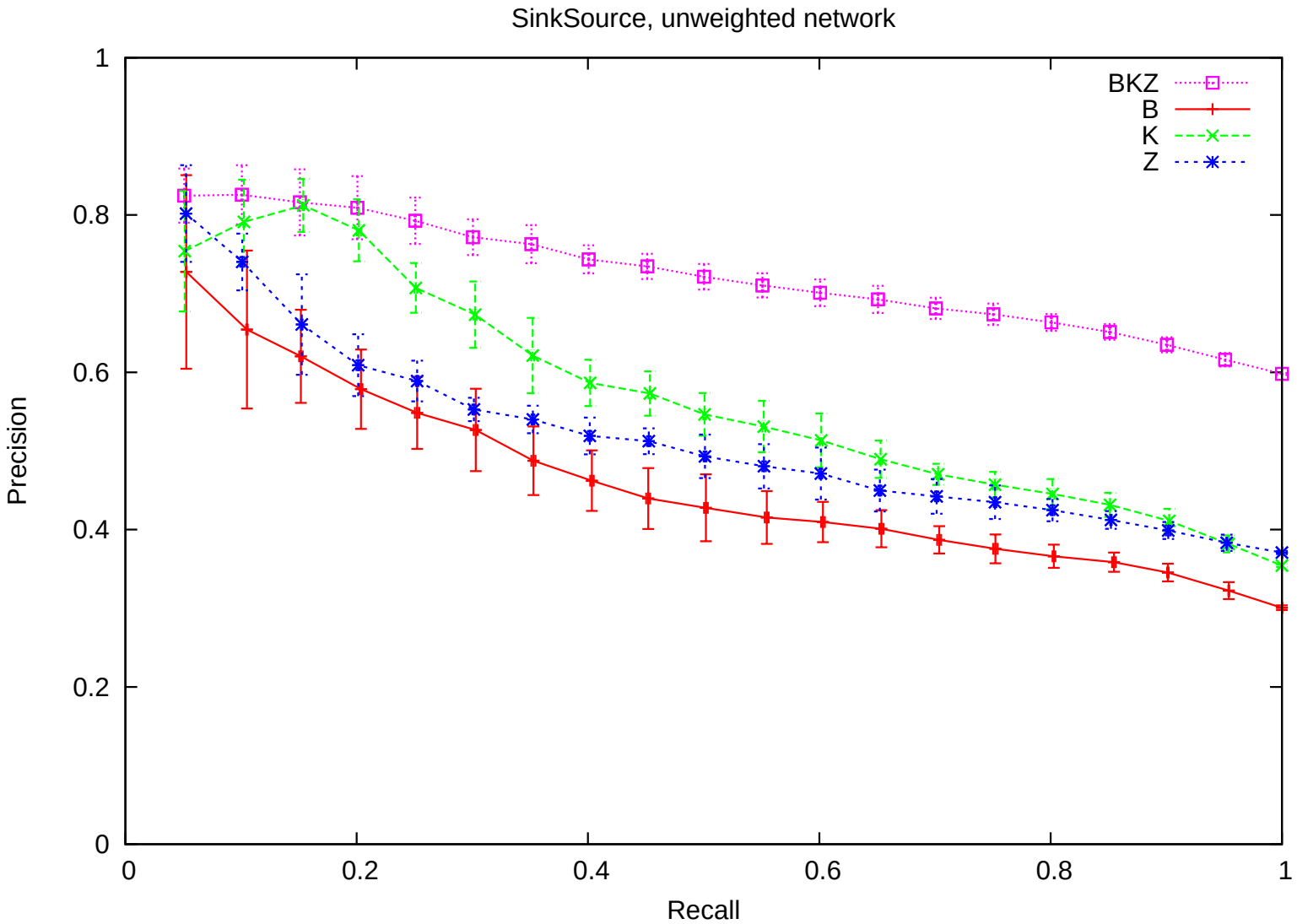
Evaluating Algorithms

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

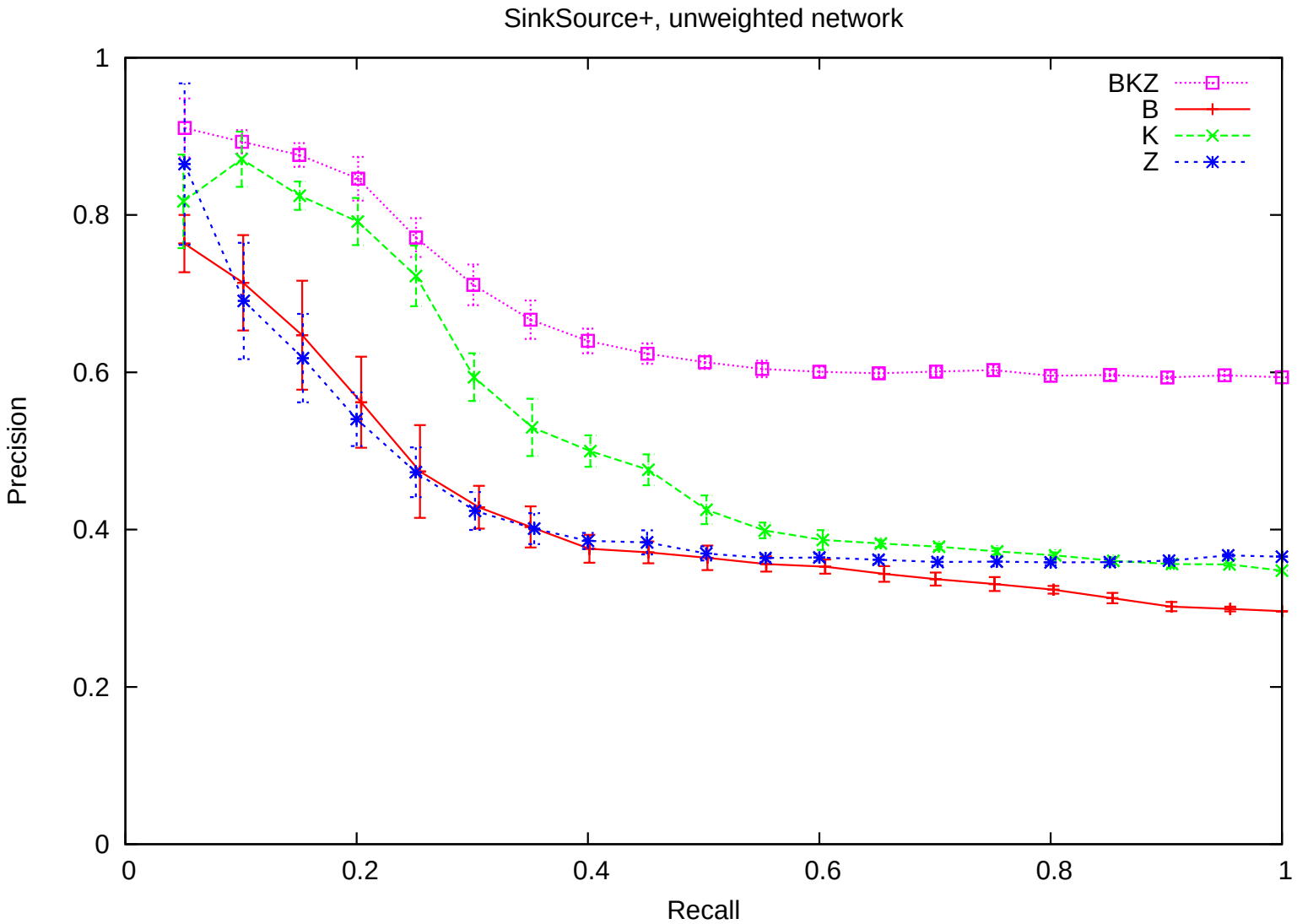
Evaluating Algorithms

- 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 ($fp/(fp + tn)$).

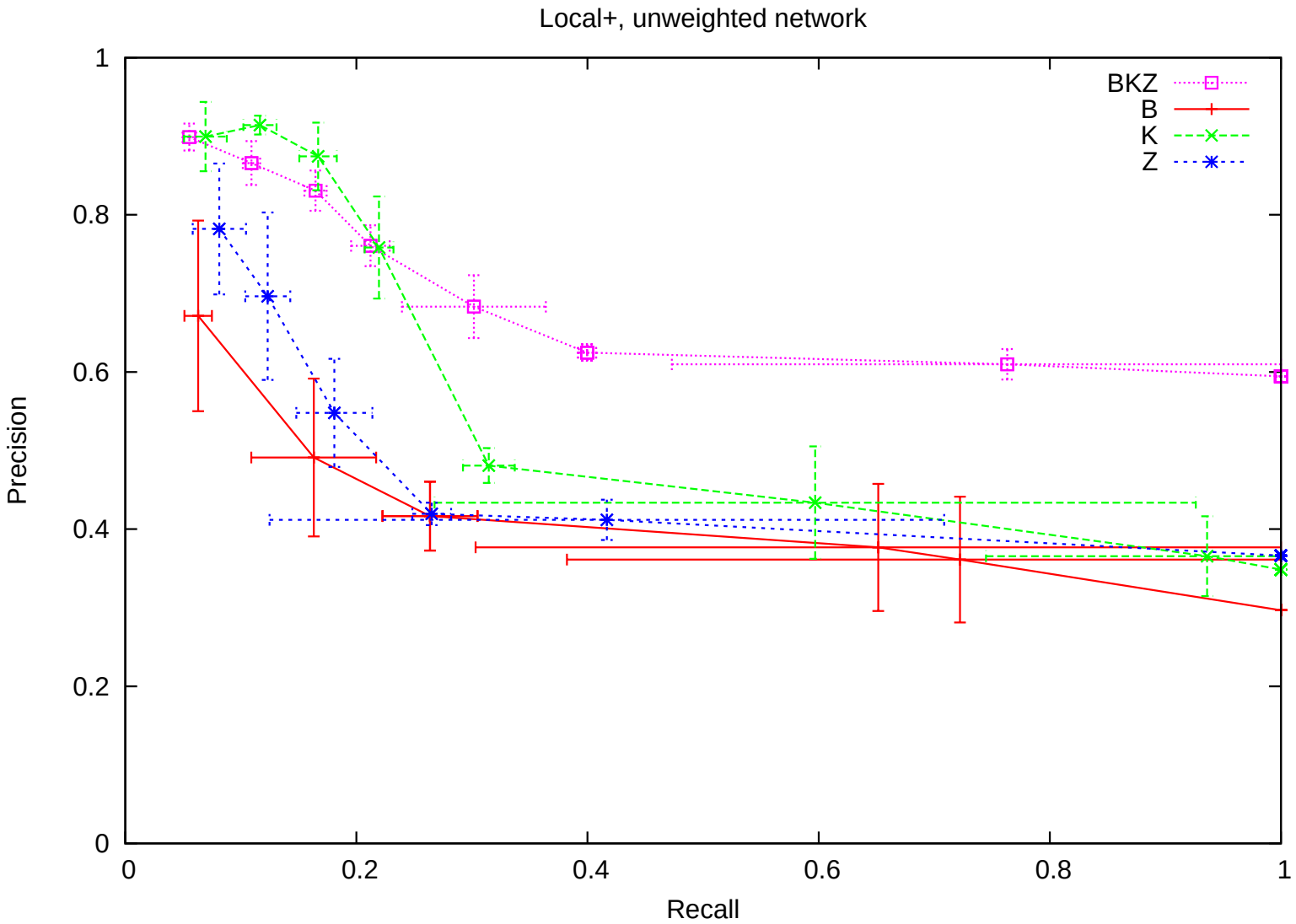
Cross-Validation Results, Unweighted PPI Network



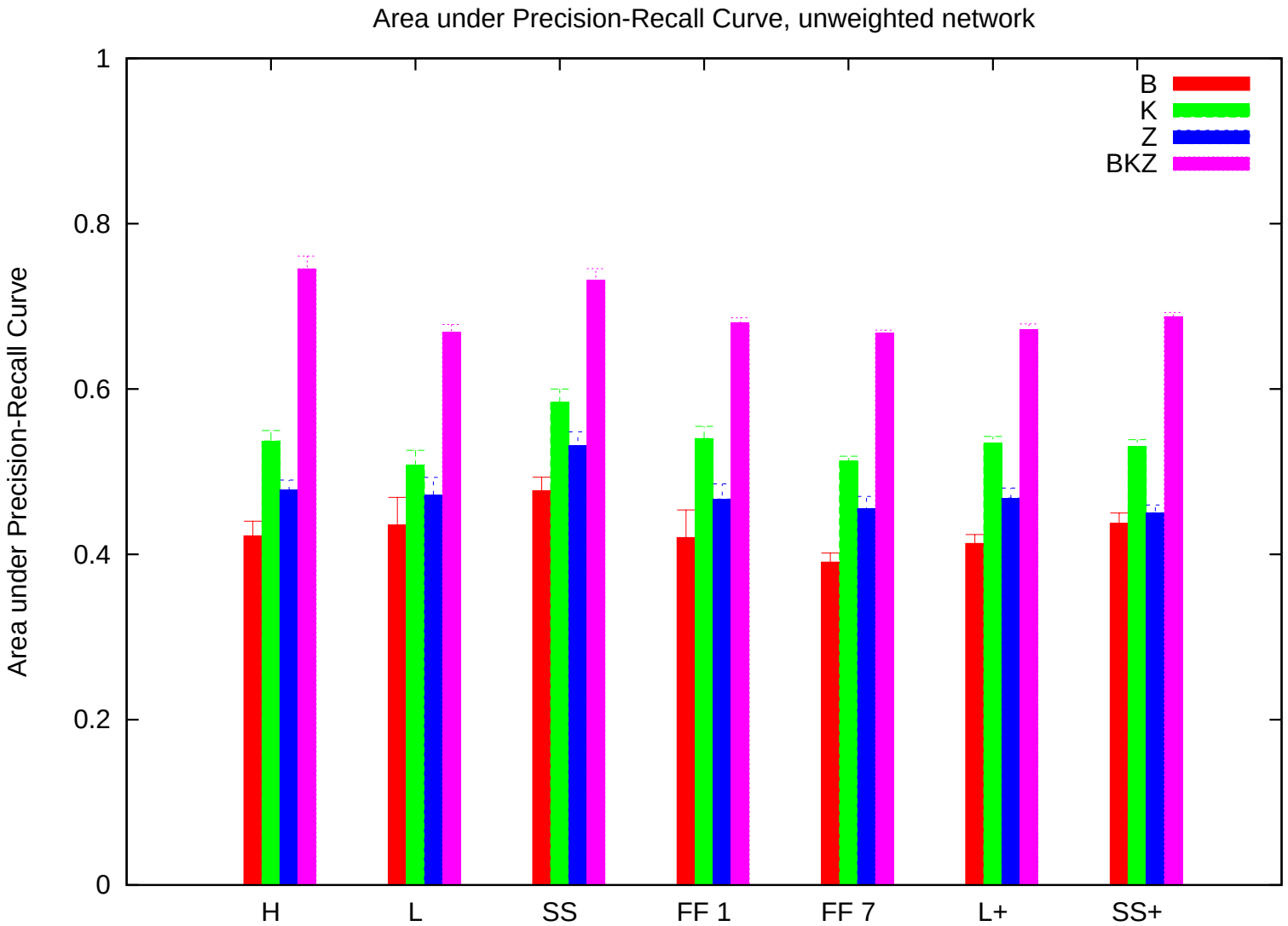
Cross-Validation Results, Unweighted PPI Network



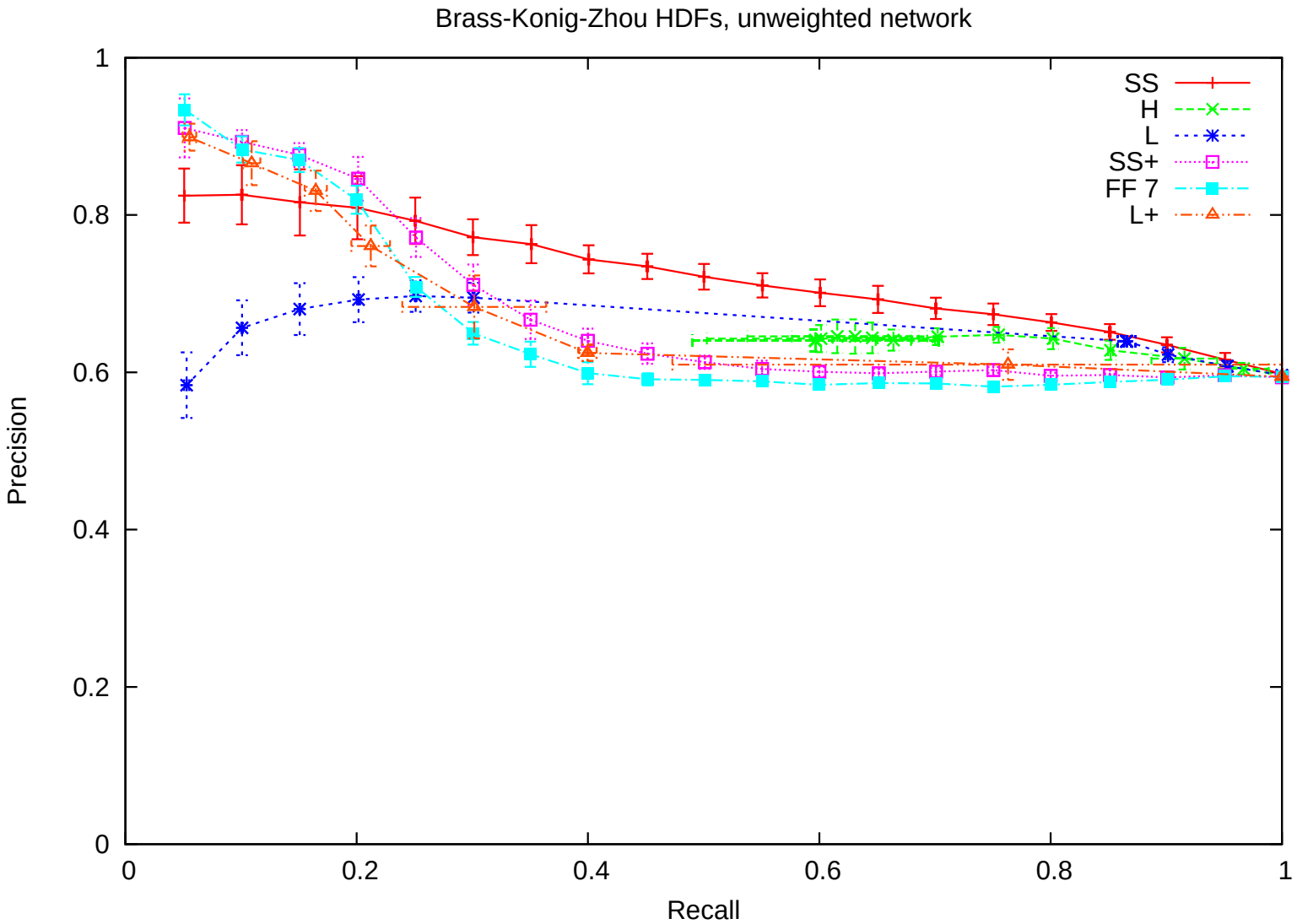
Cross-Validation Results, Unweighted PPI Network



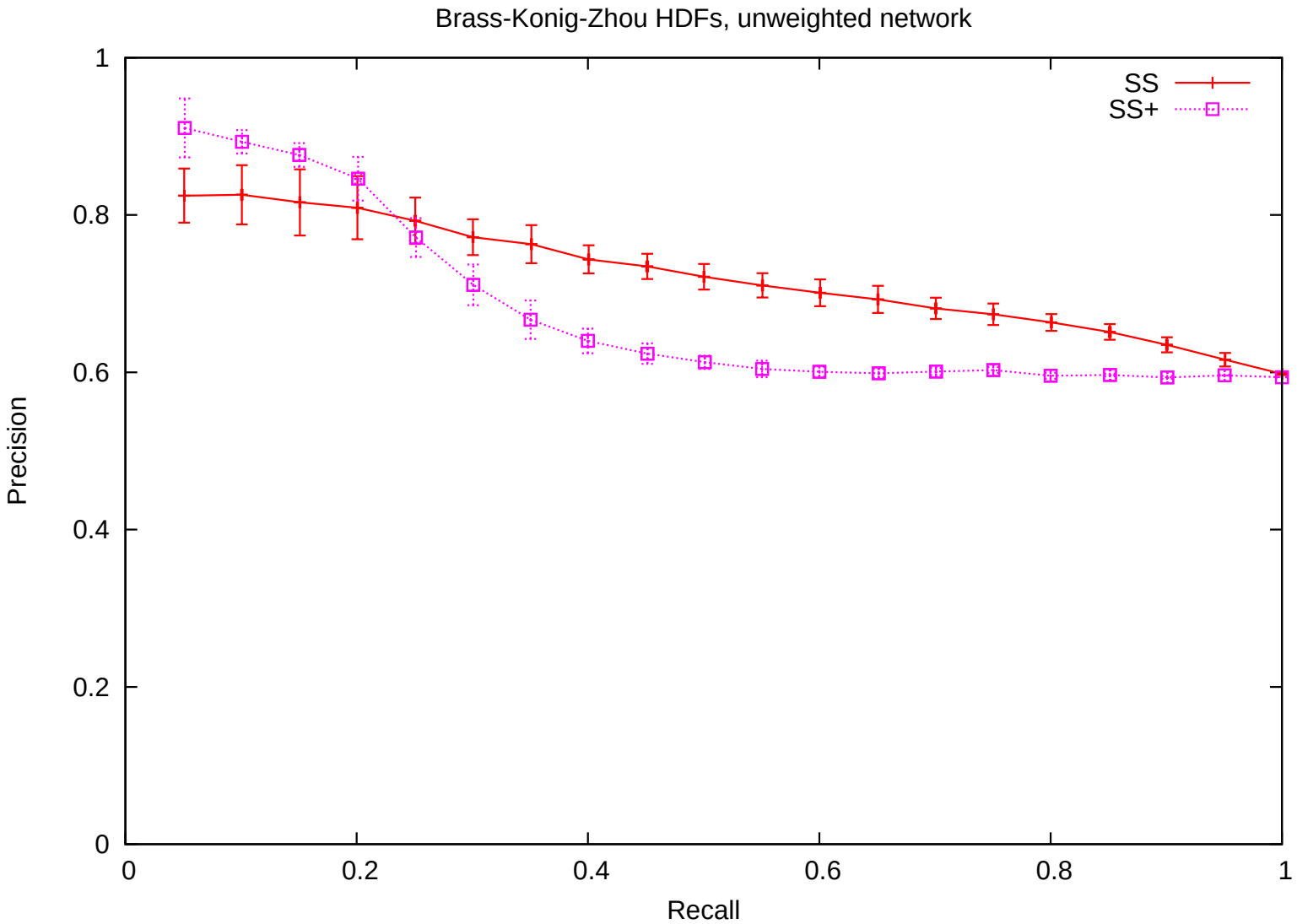
Cross-Validation Results, Unweighted PPI Network



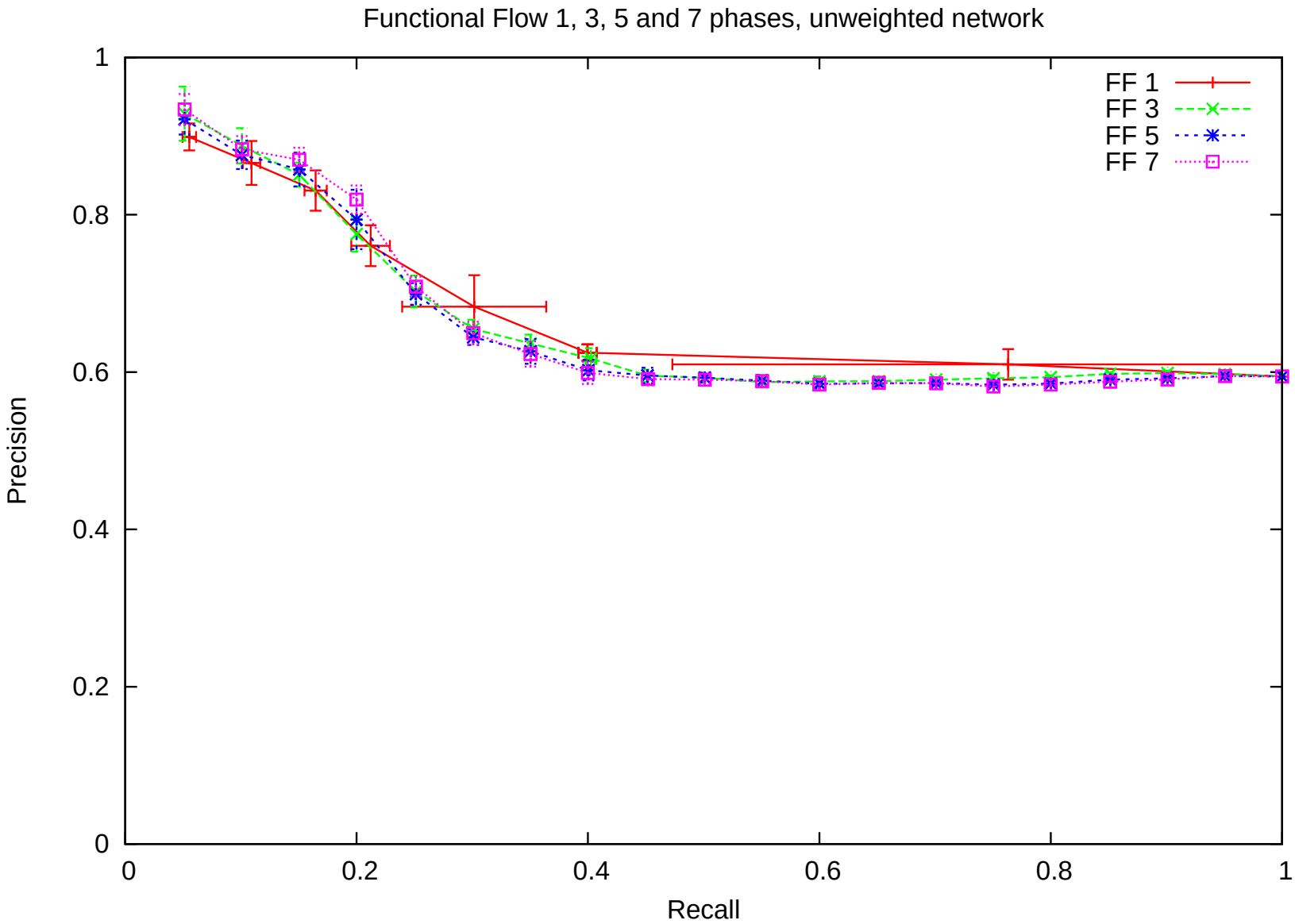
Cross-Validation Results, Unweighted PPI Network



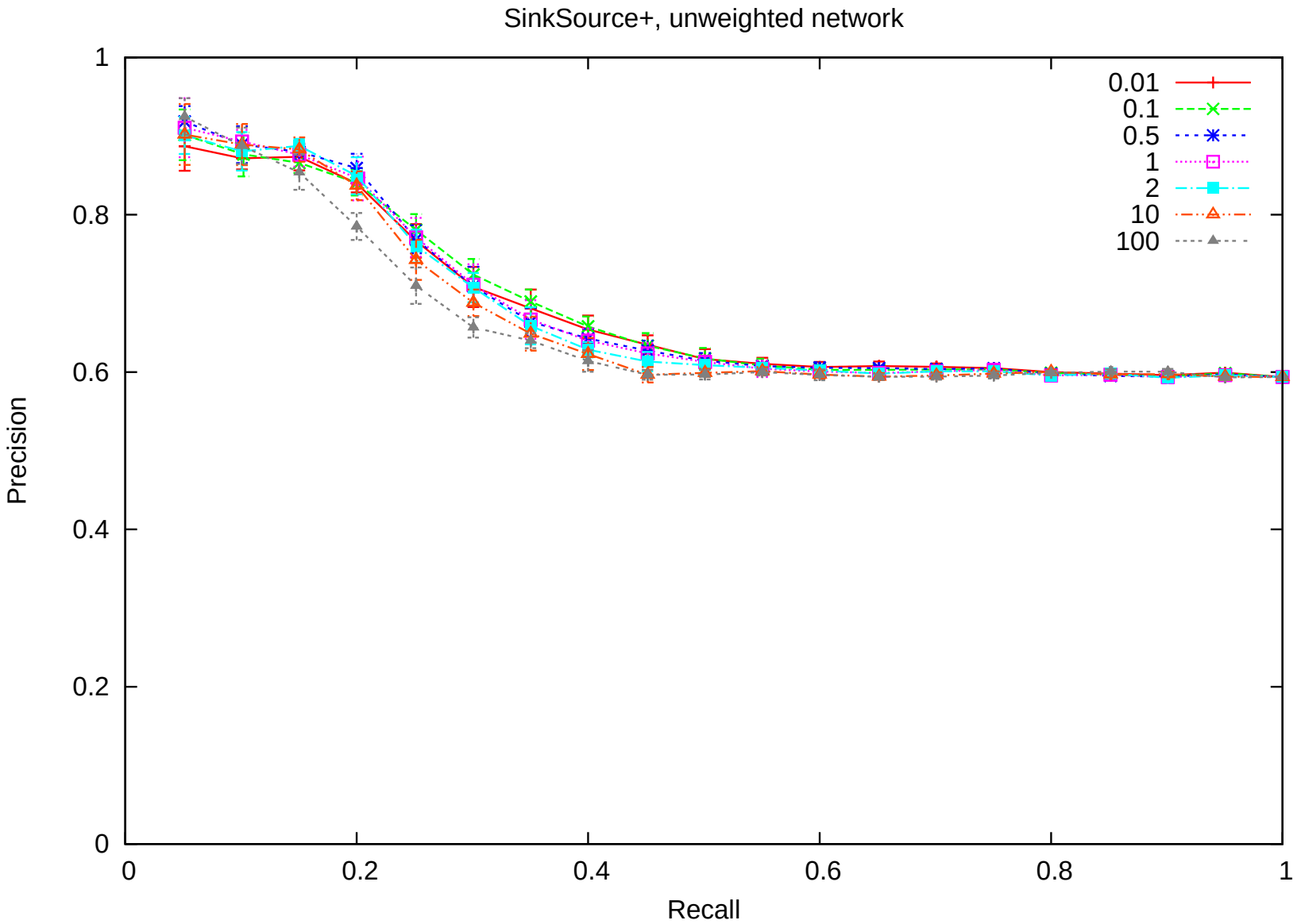
Cross-Validation Results, Unweighted PPI Network



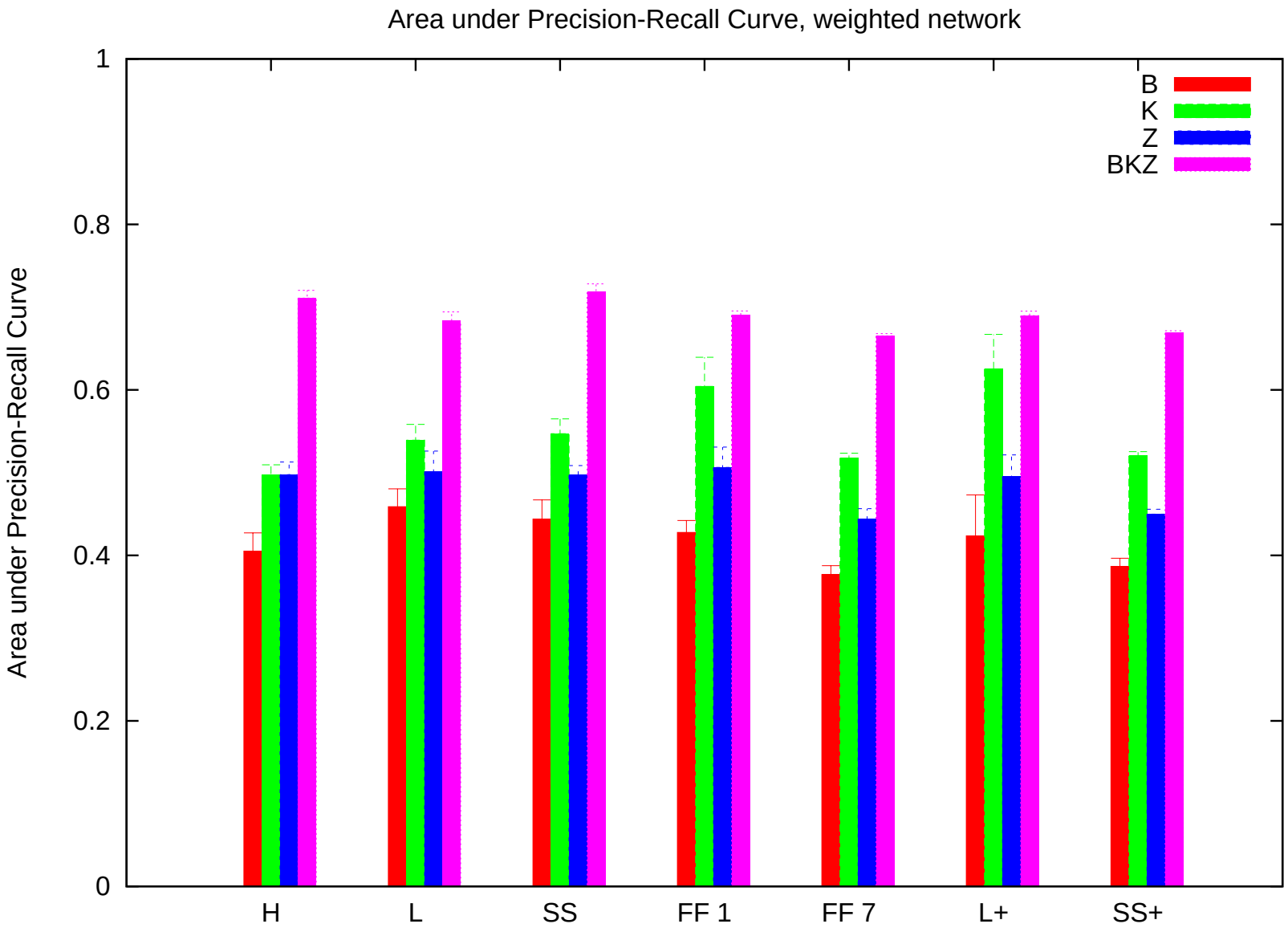
Cross-Validation Results, Unweighted PPI Network



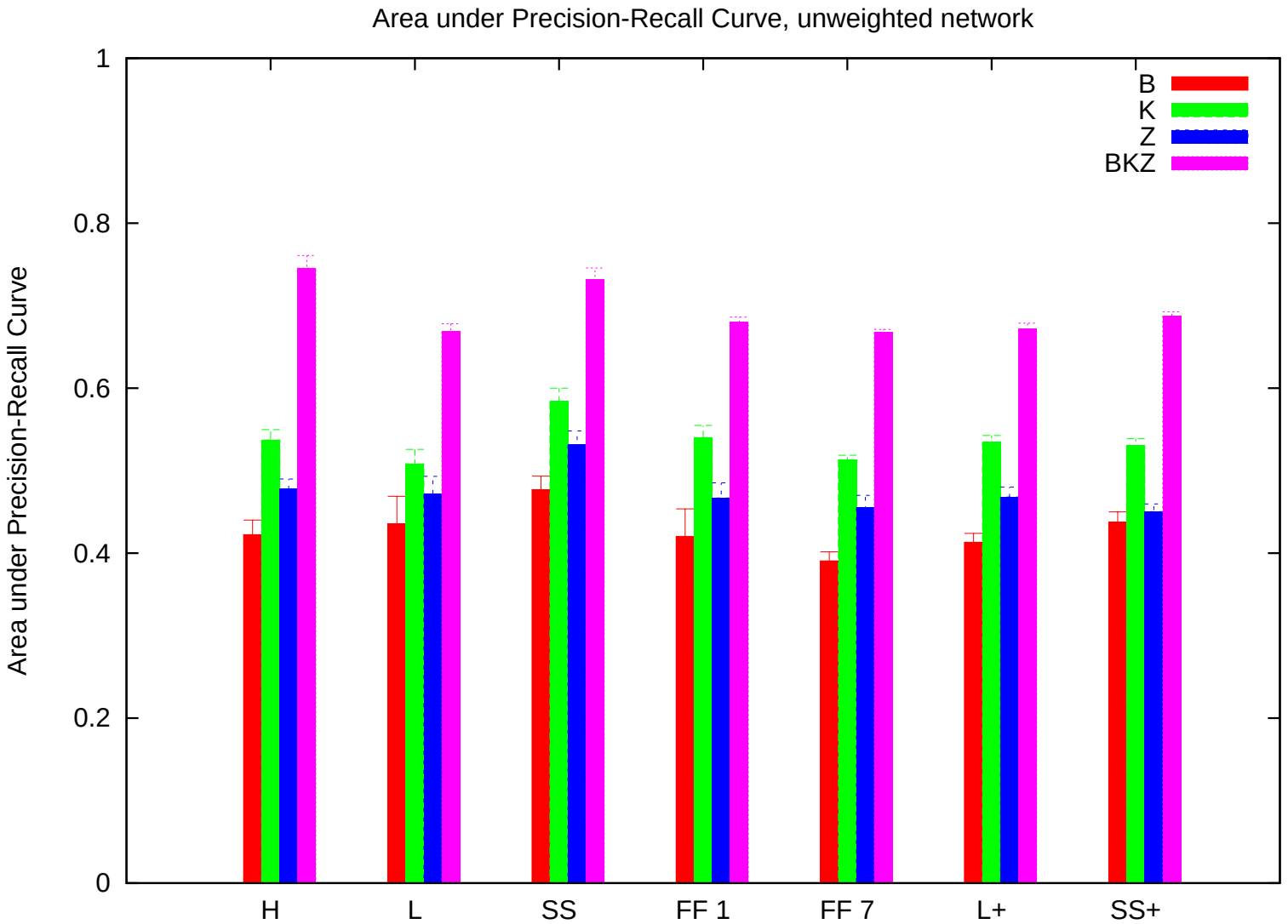
Cross-Validation Results, Unweighted PPI Network



Cross-Validation Results, Weighted PPI Network



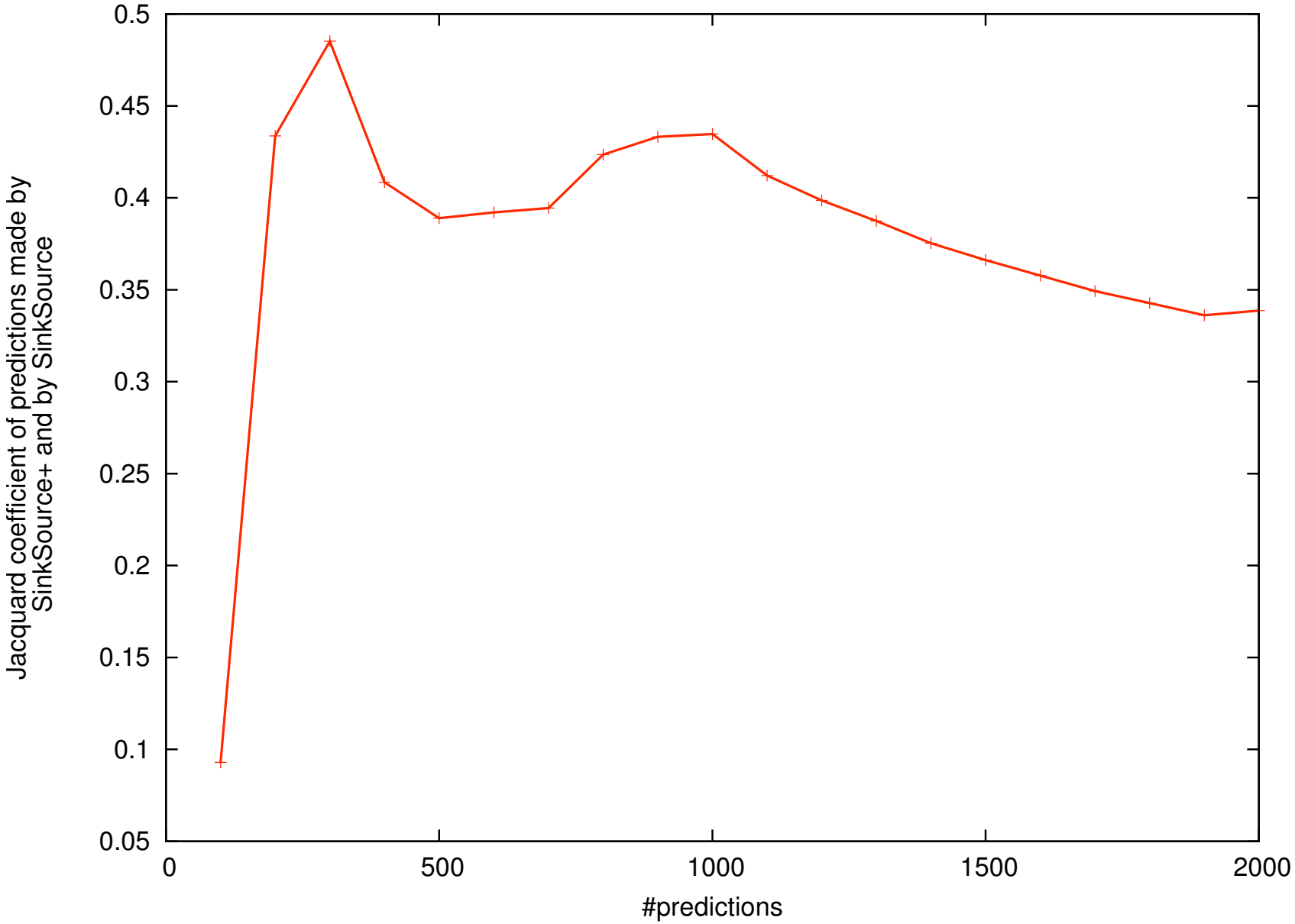
Cross-Validation Results, Weighted PPI Network



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



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.

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 1.02×10^{-10}).

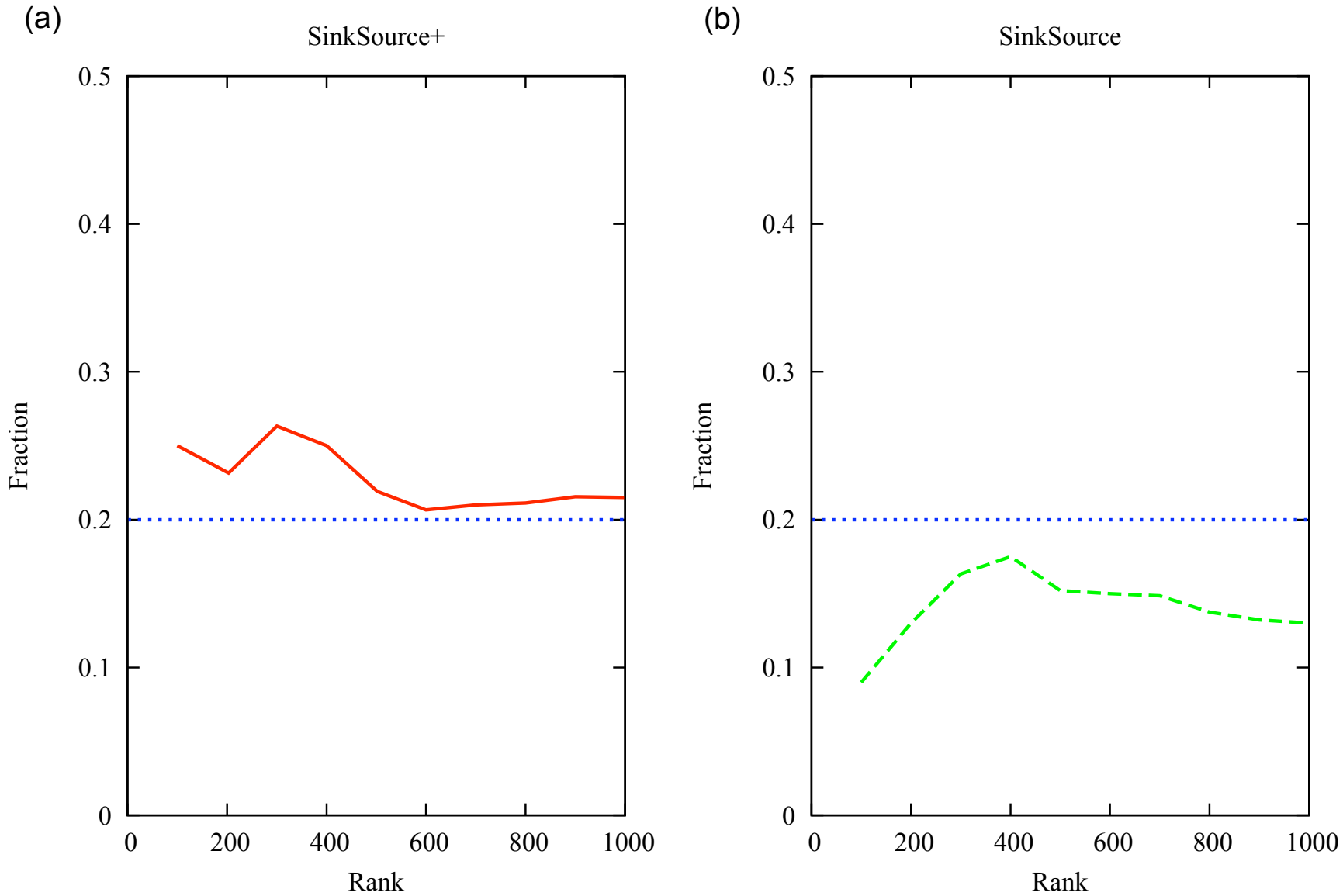
2. Overlap with HIV Interactors

- 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-value 1.02×10^{-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.

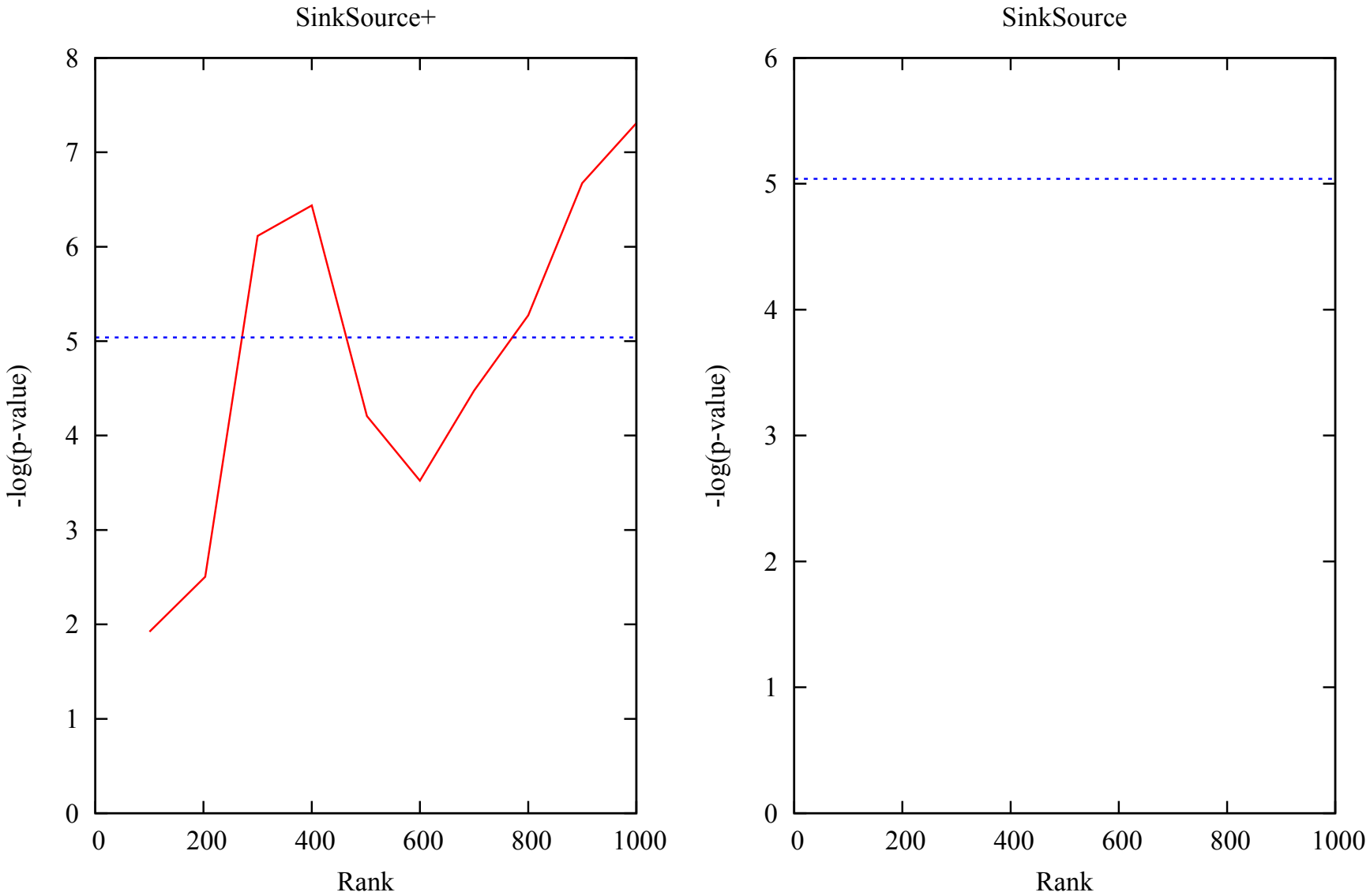
HDFs Predicted by SinkSource+ Interact with HIV

Fraction of BKZ or predicted HDFs that interact with HIV



HDFs Predicted by SinkSource+ Interact with HIV

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



3. Clustering BKZ and Predicted HDFs

- Computed sub-network induced by BKZ and HDFs predicted by SinkSource+: 9,452 nodes and 71,367 PPIs.
- Applied a modified version of the MCODE algorithm (Bader and Hogue, *BMC Bioinfo*, 2003).
- 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.

Cluster id	#proteins	#BKZ HDFs	Fraction of BKZ HDFs	p-value
0	87	17	0.2	3.5×10^{-4}
2	139	24	0.17	7.2×10^{-5}
6	19	9	0.47	4.9×10^{-6}
9	14	7	0.5	2.5×10^{-5}
11	12	6	0.5	2.9×10^{-32}
14	10	5	0.5	2.1×10^{-3}
29	5	5	1	3×10^{-6}
87	4	3	0.75	6.2×10^{-3}
90	6	2	0.33	2.7×10^{-2}

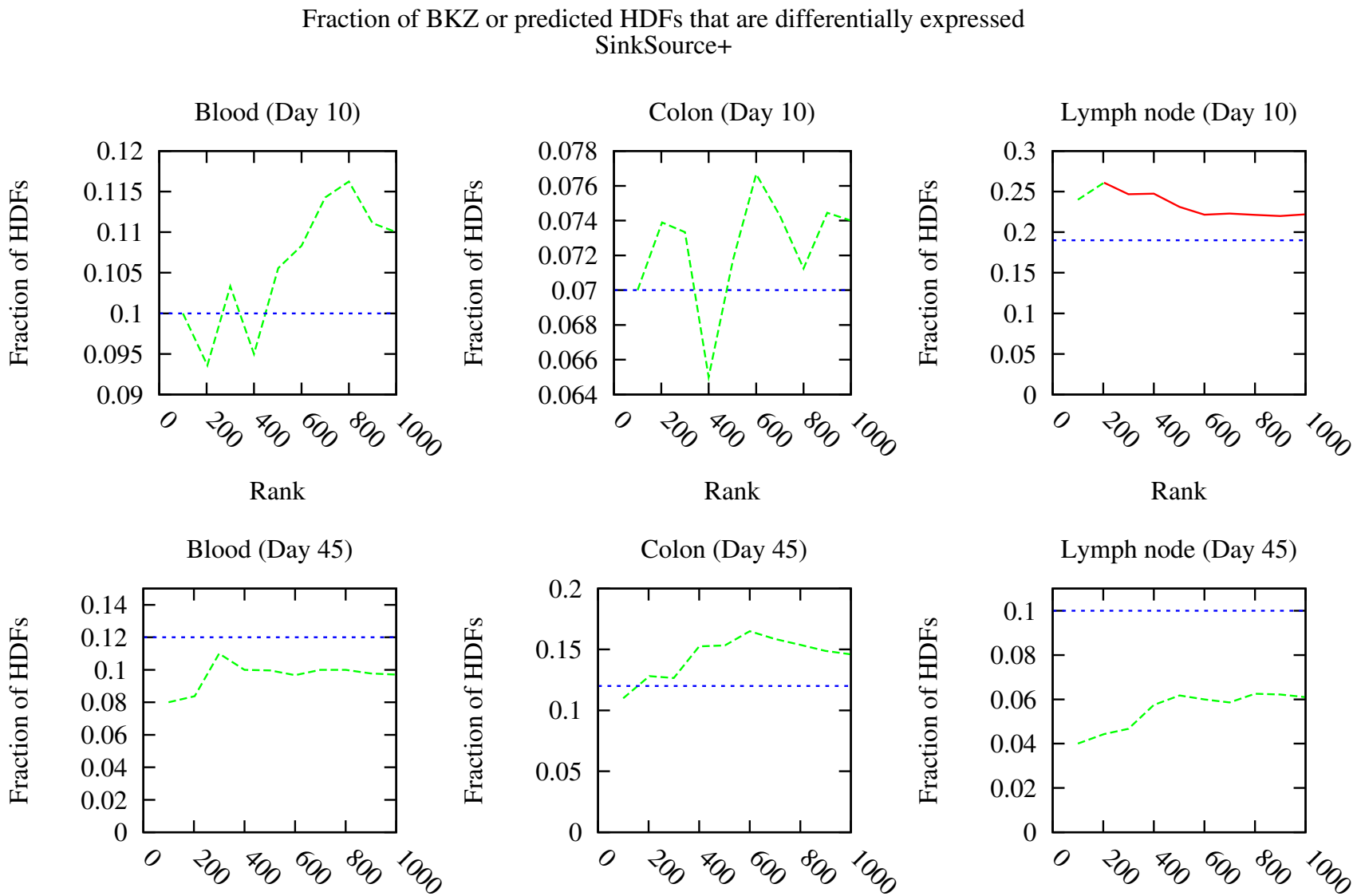
GO Terms Enriched in MCODE Clusters

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

4. Relating HDFs to HIV Progression

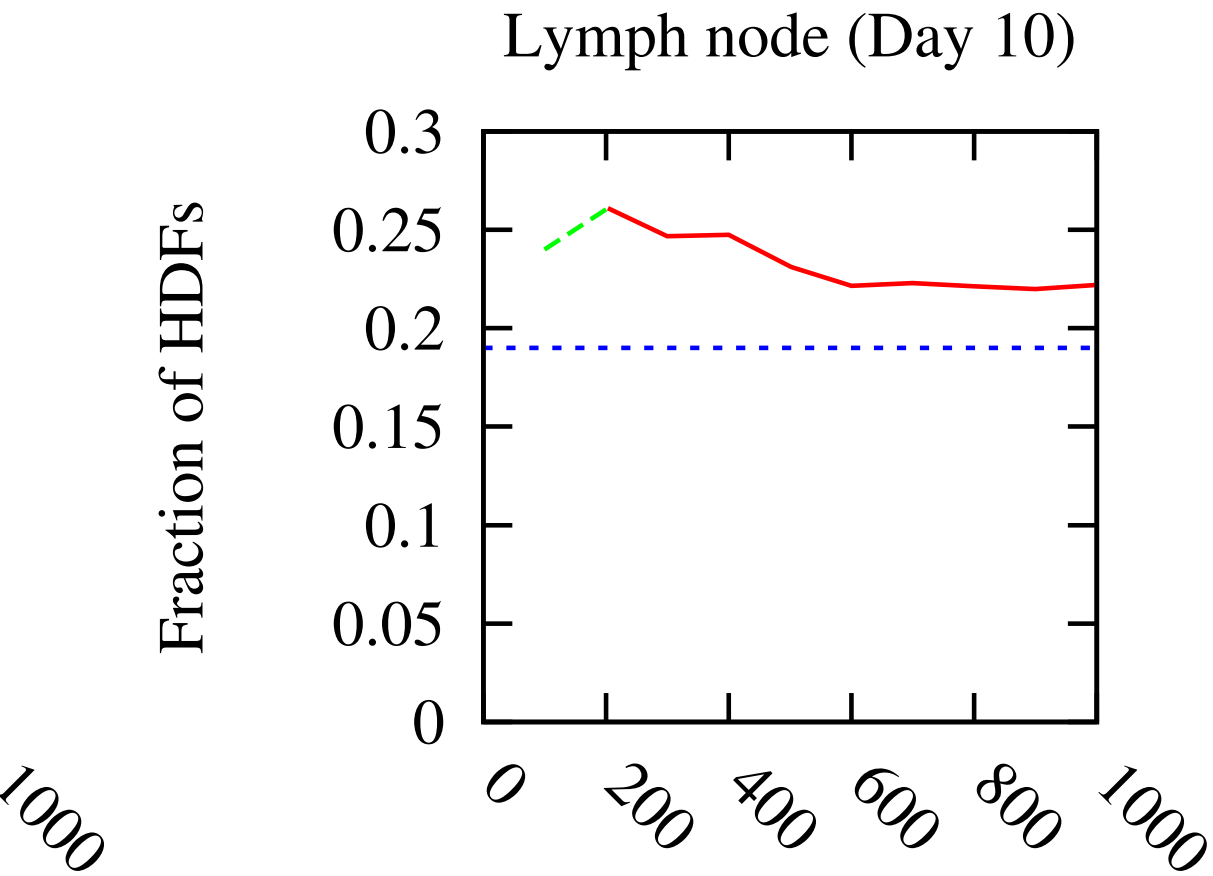
- Used gene expression data from SIV-infected non-human primates.
  - African Green Monkeys (AGMs) are natural hosts for SIV.
  - Pig-tailed Macaques (PMs) are susceptible to SIV.
- Lederer et al., *PLoS Path.*, 2009 conducted a longitudinal study of SIV infection in AGMs and PTs:
  - Day 10 and day 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

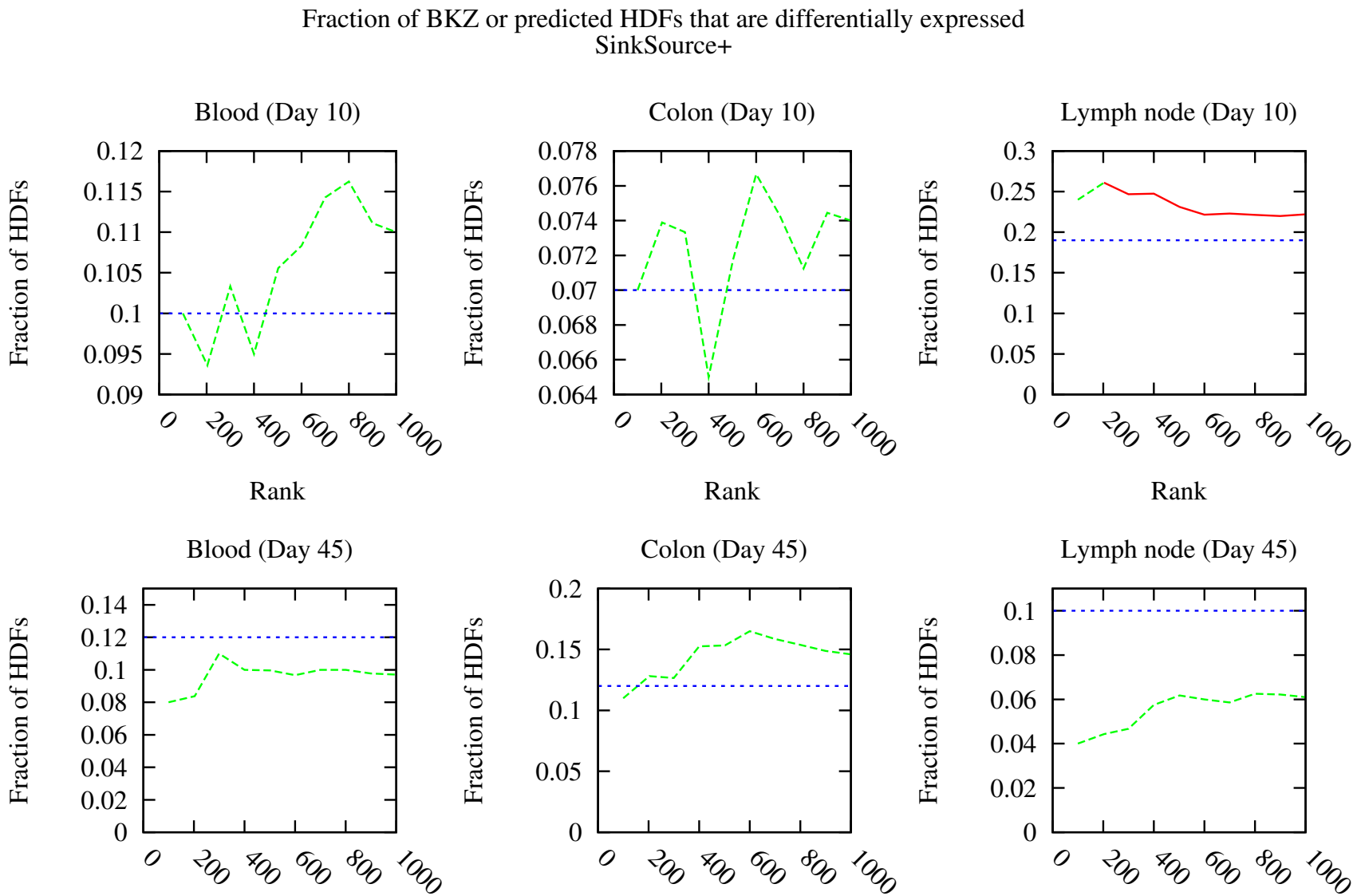


Predicted HDFs Overlap with Diagnostic Genes

ntially expressed

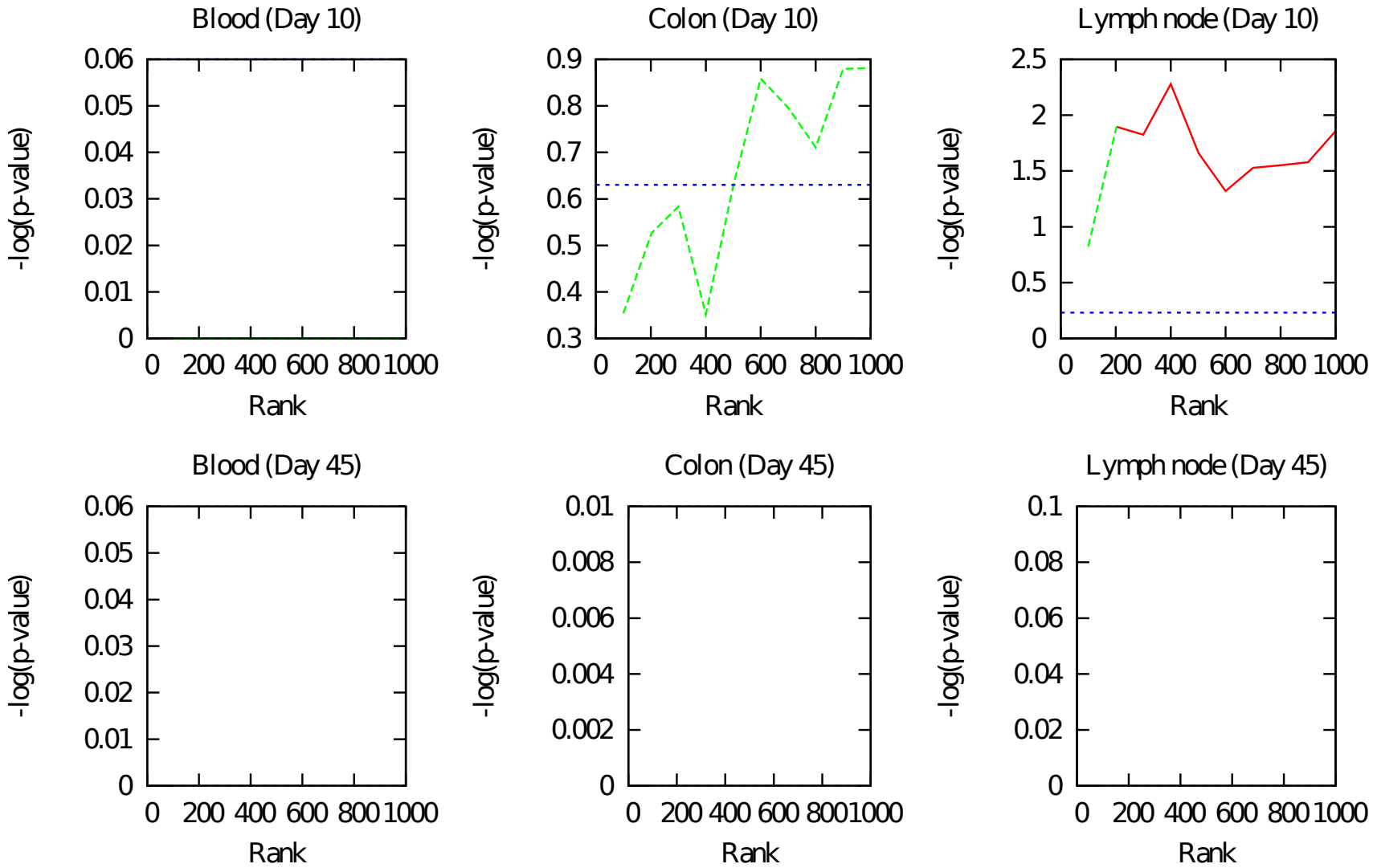


Predicted HDFs Overlap with Diagnostic Genes

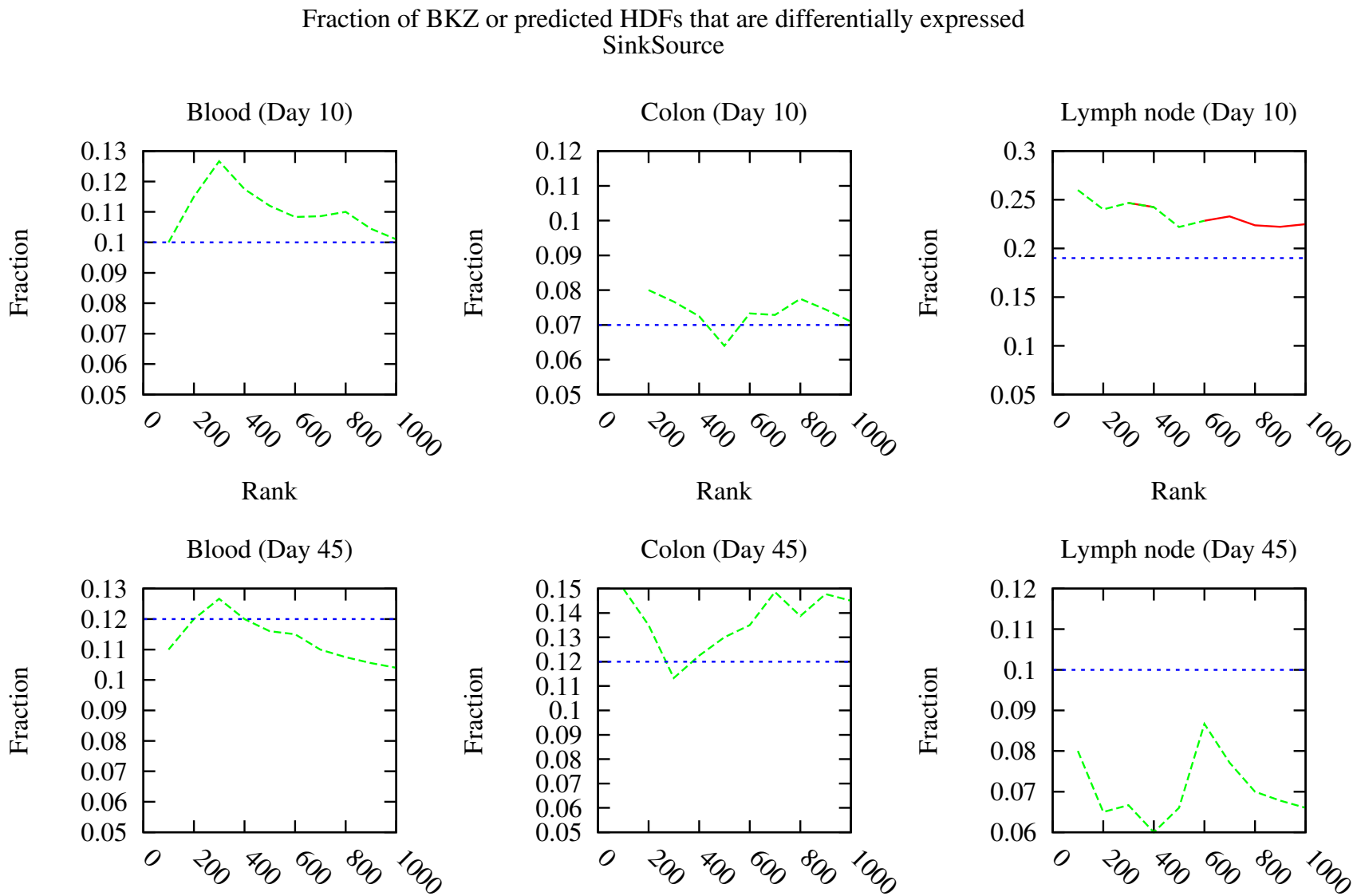


Predicted HDFs Overlap with Diagnostic Genes

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



Predicted HDFs Overlap with Diagnostic Genes



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.