

Cell signaling networks

Part 1: Predicting signaling networks from the genome

Part 2: Network classification of disease

Cary MP, Bader GD, Sander C

Pathway information for systems biology

FEBS Lett. 2005 Mar 21;579(8):1815-20

Pawson T, Nash P

Assembly of cell regulatory systems through protein interaction domains

Science. 2003 Apr 18;300(5618):445-52

Gary Bader

Donnelly Centre

University of Toronto

Mar.18 2011 - BCB420



Donnelly Centre
for Cellular + Biomolecular Research



UNIVERSITY OF
TORONTO



<http://baderlab.org>

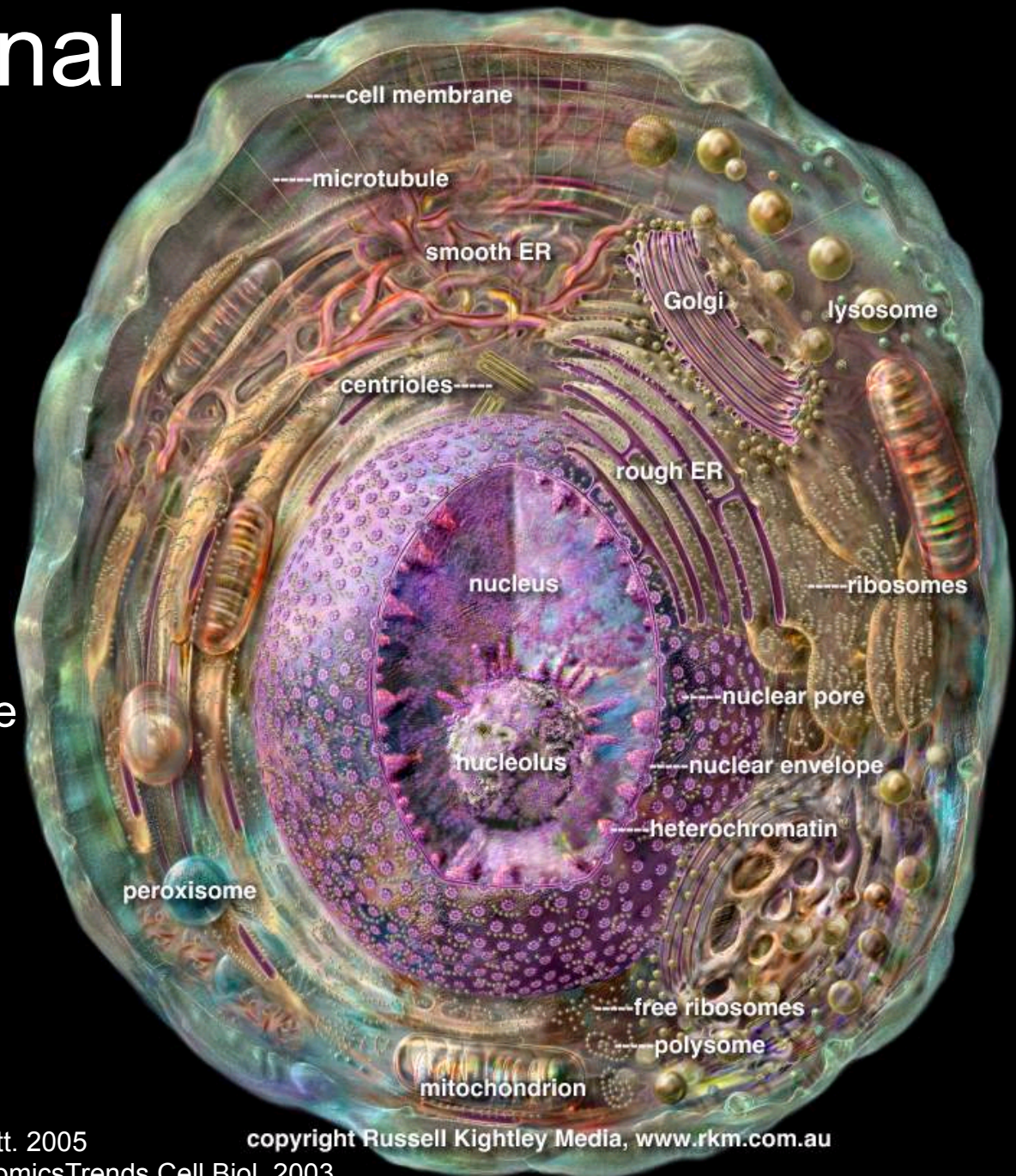
Computational Cell Map

Read map to understand

- Cell processes
- Gene function
- Disease effects
- Map evolution

Map the cell

- Predict map from genome
- Multiple perturbation mapping
- Active cell map
- Map visualization and analysis software



How are biological networks in the cell
encoded in the genome?

Can we accurately predict biologically relevant interactions from
a genome?

How do genome sequence changes underlying disease affect
the molecular network in the cell?

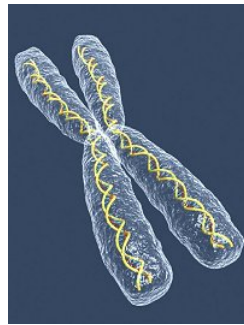
Can we predict how well model pathways or phenotypes will
translate to human?

Can we design new networks *de novo*?

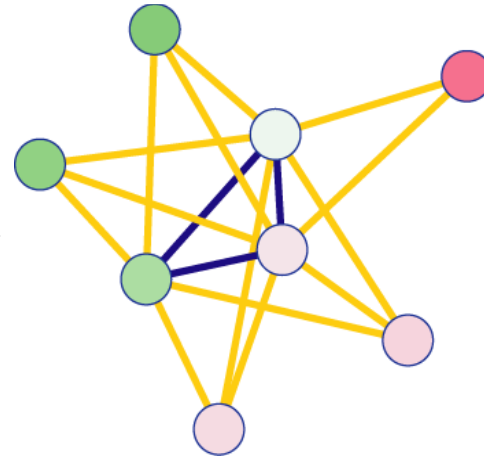


Predicting Protein Interaction Networks From the Genome

- Ideally:



Accurately
Predict



- Reality:

- Not currently possible
- Protein interaction prediction likely as hard as protein folding, in general e.g. induced fit

Predicting Networks

- Map via orthology relationships

- Metabolic pathways

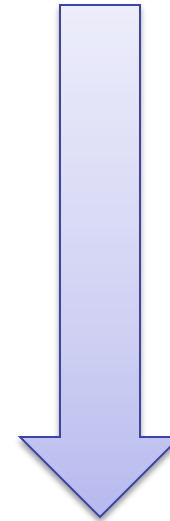
- E.g. KEGG, BioCyc, metaSHARK

- Protein-protein interactions

- E.g. I2D(OPHID), HomoMINT

- Signaling pathways

- E.g. Reactome



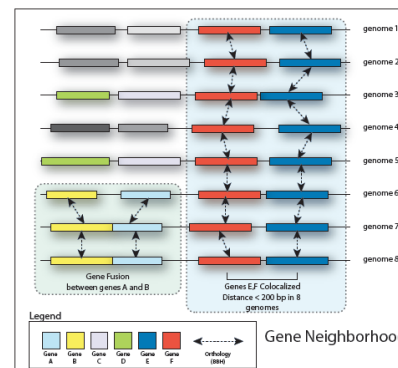
Higher accuracy
(more conserved)

Lower accuracy
(less conserved)

- Infer using functional associations

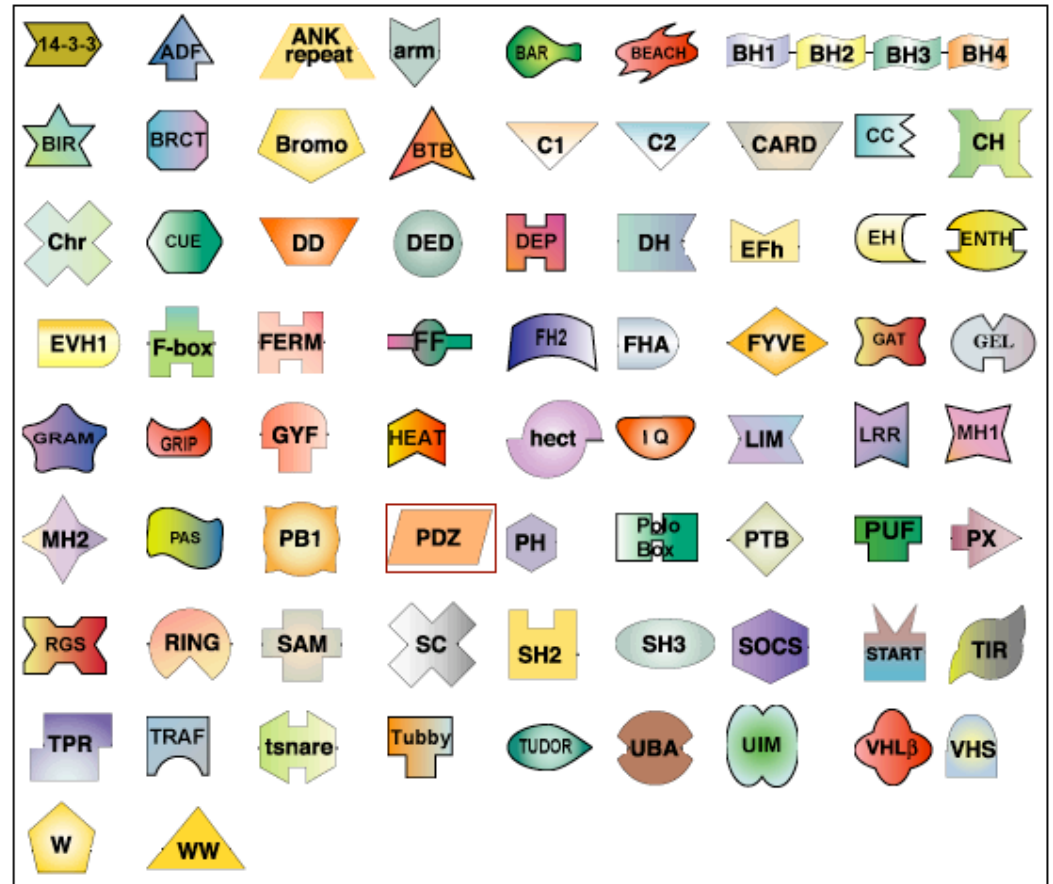
- Phylogenetic profile, Rosetta Stone

Gene A	0	0	0	0	0	1	1	1
Gene B	0	0	0	0	0	1	1	1
Gene C	0	0	1	0	1	0	0	0
Gene D	0	0	1	0	1	0	0	0
Gene E	1	1	1	1	1	1	1	1
Gene F	1	1	1	1	1	1	1	1
	genome 1	genome 2	genome 3	genome 4	genome 5	genome 6	genome 7	genome 8



Peptide Recognition Domains

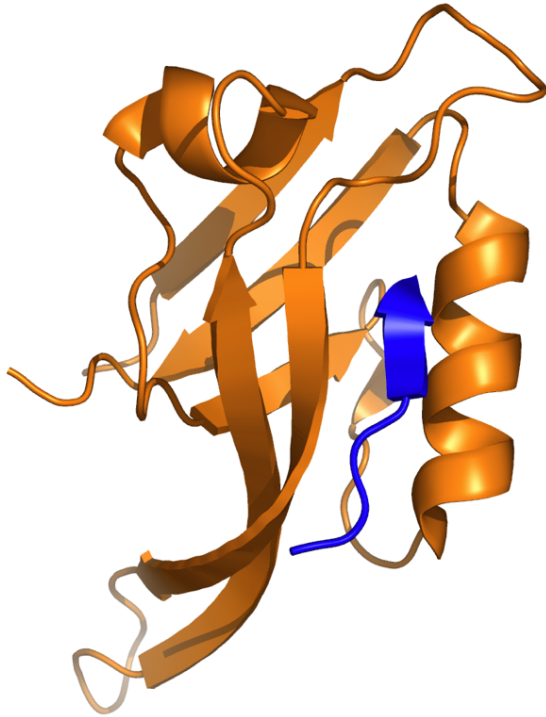
- Simple binding sites
- Well studied
- Numerous
- Biologically important
 - Eukaryotic signaling systems often involve modular protein-protein interaction domains



<http://pawsonlab.mshri.on.ca/>

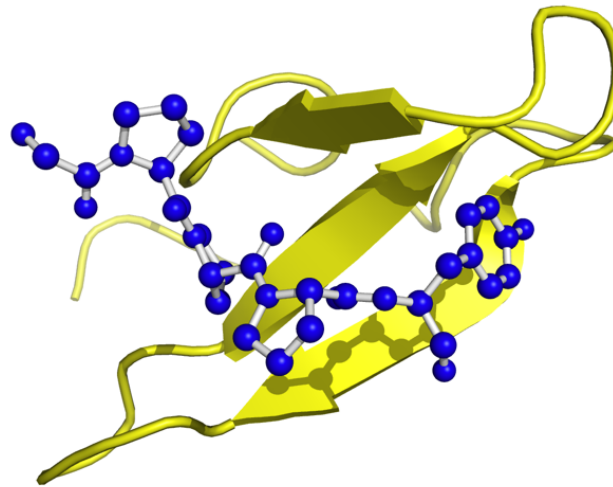
<http://nashlab.uchicago.edu/domains/>

Focus: PDZ, WW, SH3



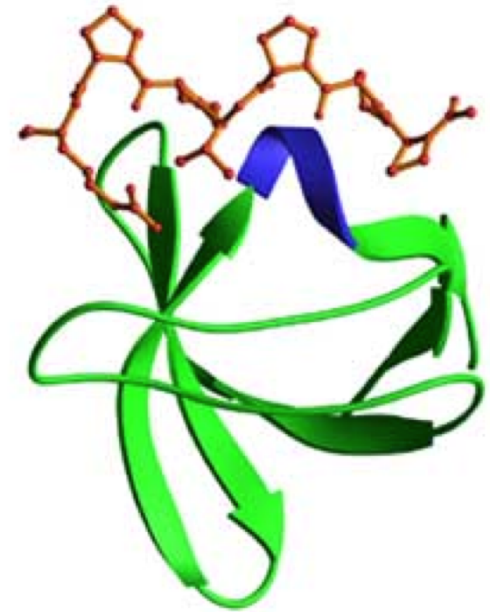
PDZ

Human: ~250



WW

~90



SH3

~300

Protein Domain Interaction Network Prediction

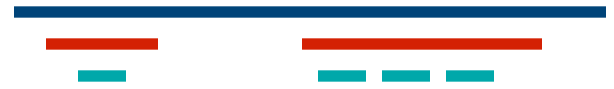
Genome



Gene and protein prediction



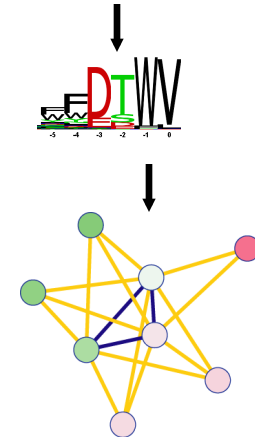
Domain prediction



Specificity prediction



Protein-protein interaction prediction



Protein Domain Interaction Network Prediction

Genome



Gene and protein prediction



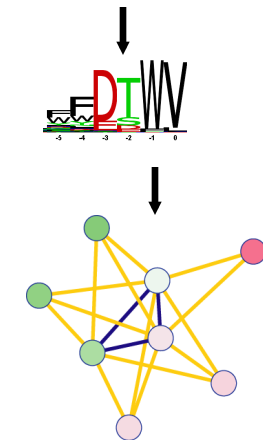
Domain prediction



Specificity prediction



Protein-protein interaction prediction



Protein Domain Interaction Network Prediction

Genome



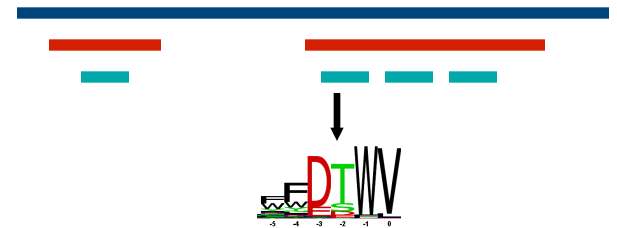
Gene and protein prediction



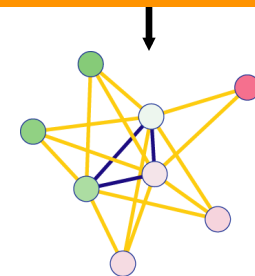
Domain prediction



Specificity prediction



Protein-protein interaction prediction

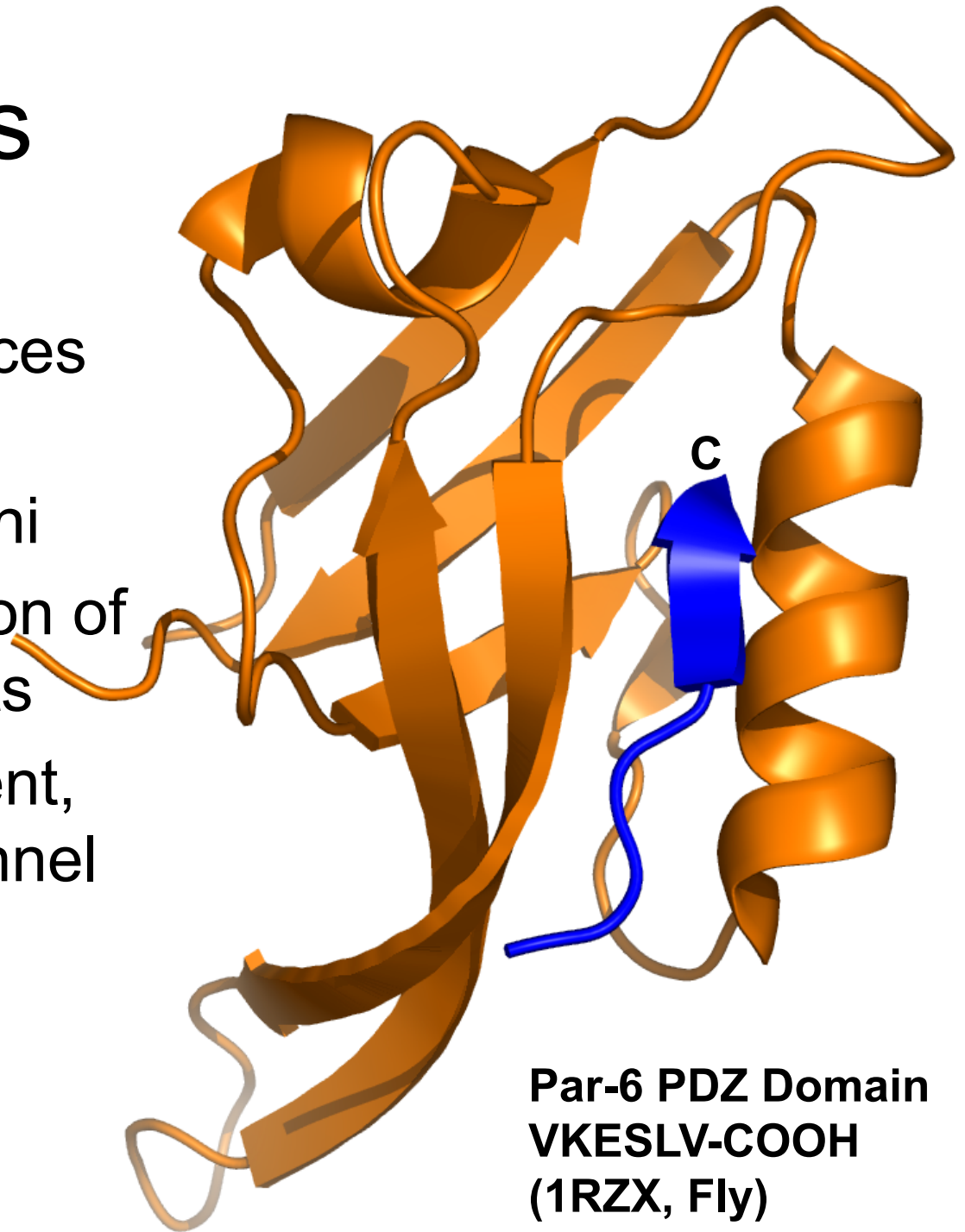


PDZ Domains

- 80-90 aa's, 5-6 beta strands, 2 alpha helices
- Recognize hydrophobic C-termini
- Membrane localization of signaling components
- Neuronal development, cell polarity, ion channel regulation

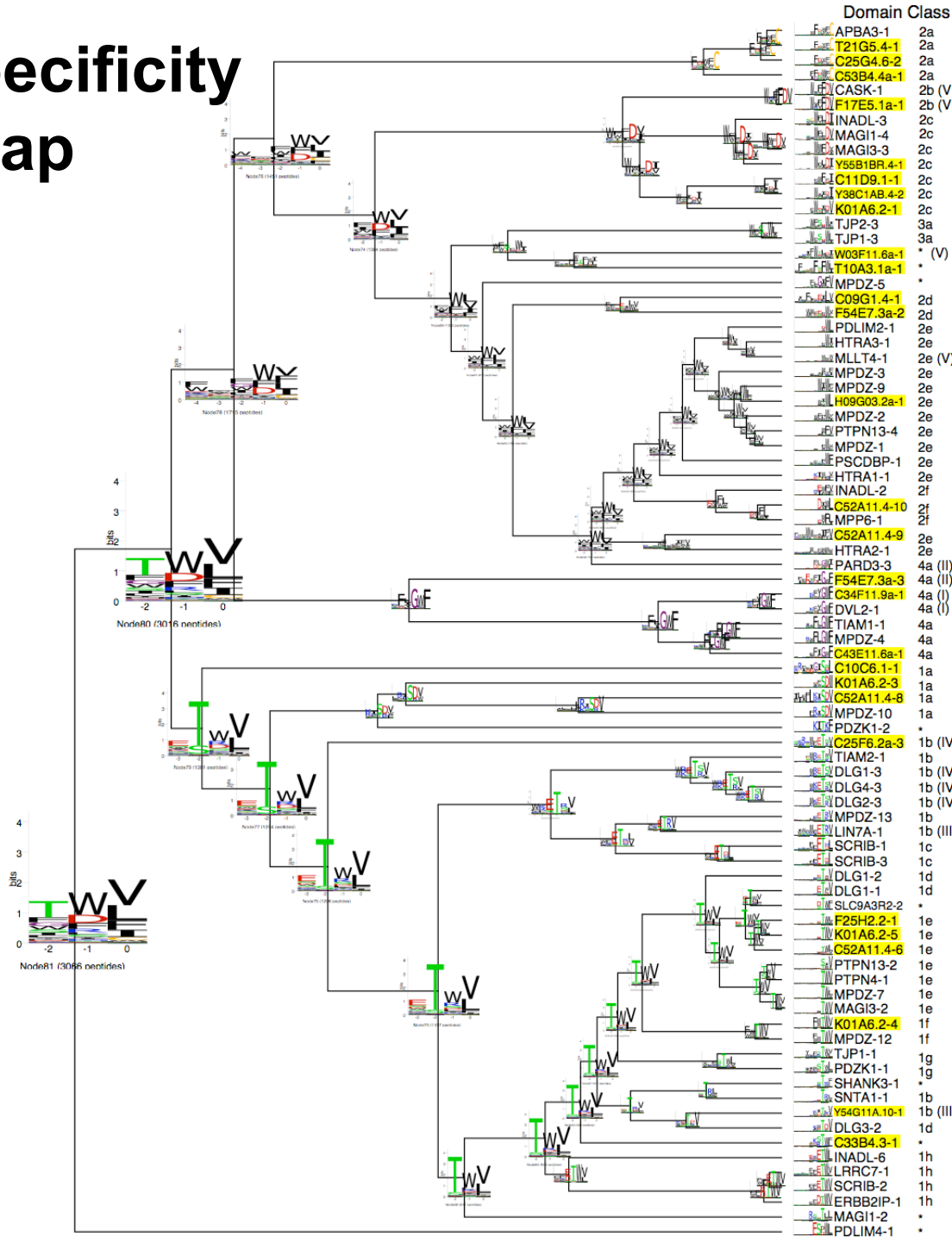
Dev Sidhu

Tonikian et al. PLoS Biology
Sep.2008



Par-6 PDZ Domain
VKESLV-COOH
(1RZX, Fly)

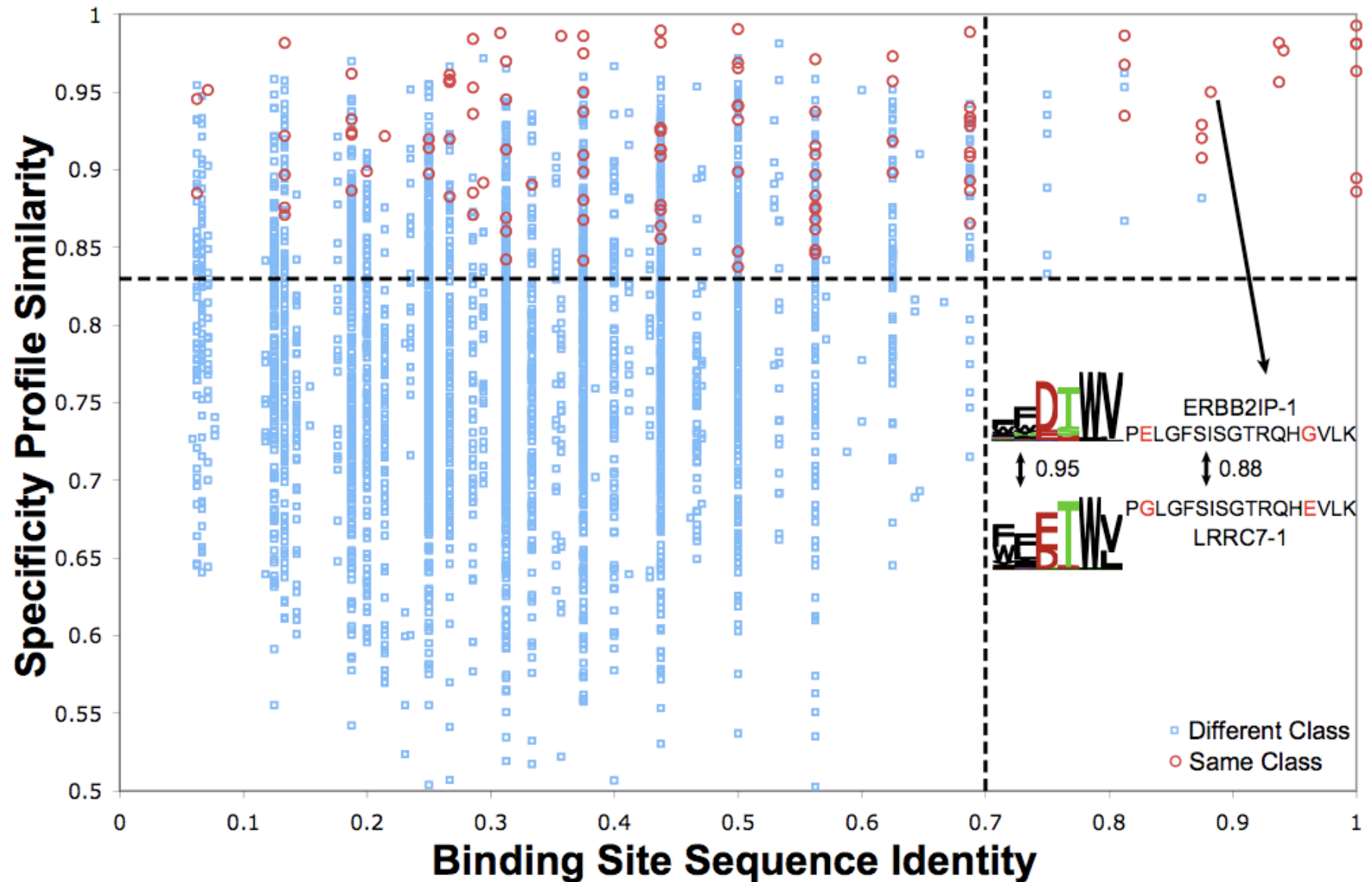
PDZ Specificity Map



Class 2: XΦXΦ

Class 1: X[T/S]XΦ

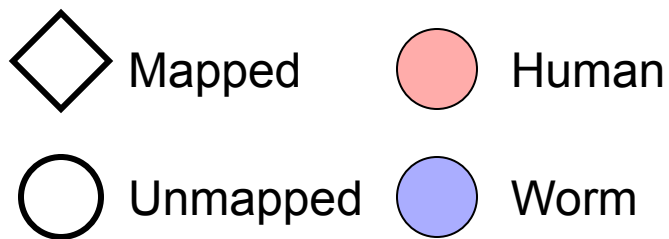
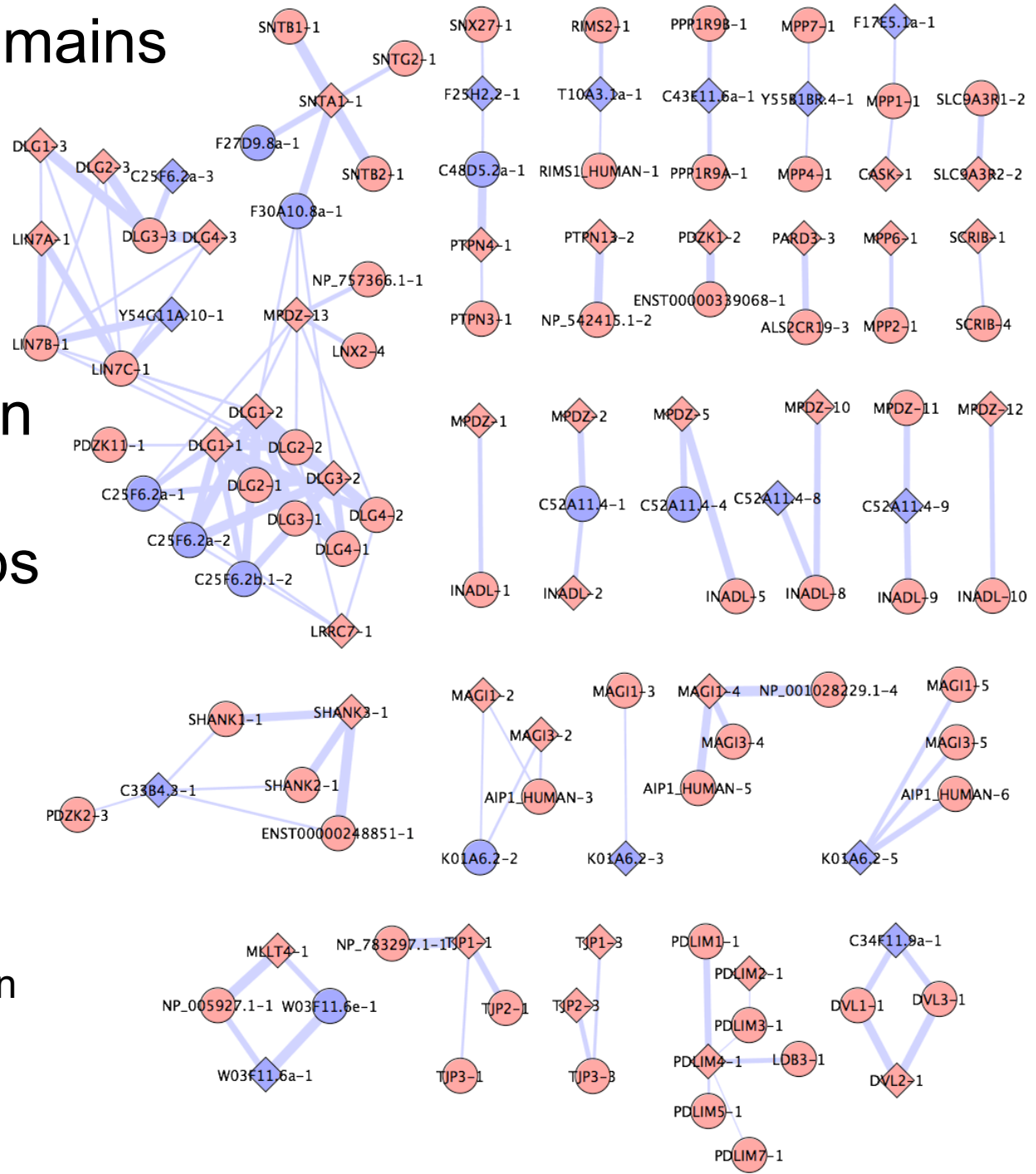
Sequence Predicts Specificity



50 mapped PDZ domains
 >70% similar to 69
 unmapped PDZ

Double coverage to
 45% of worm/human

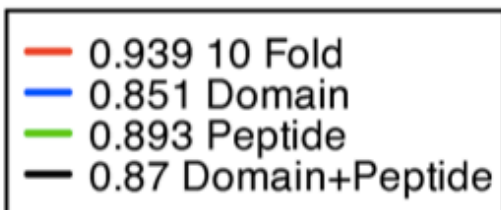
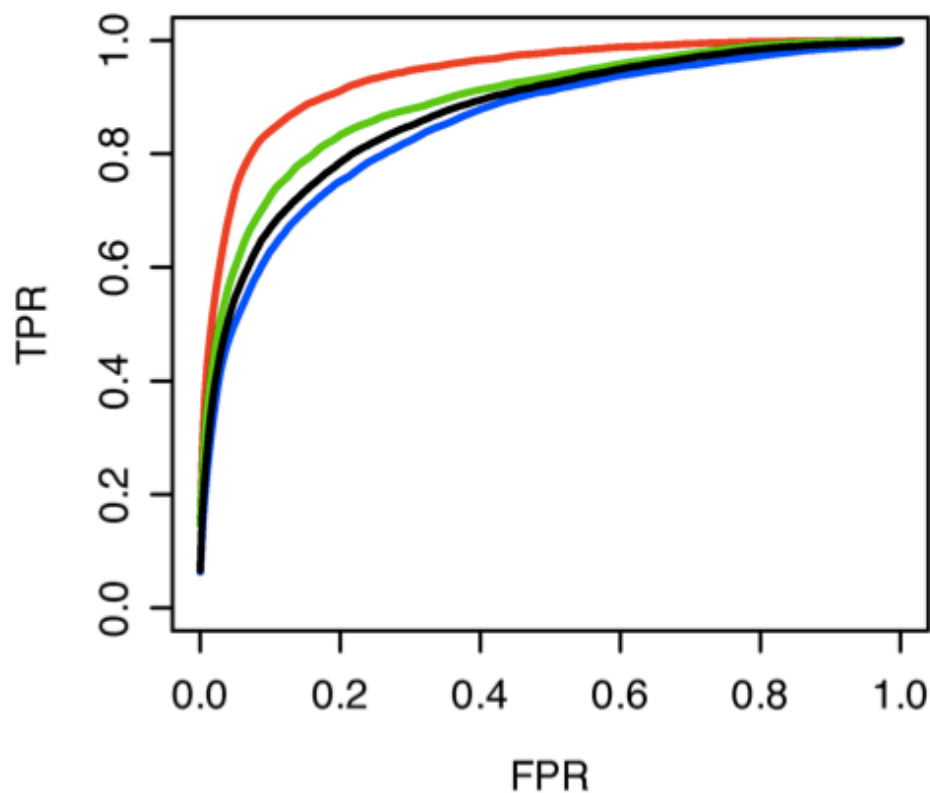
33 more PDZ groups
 110 singletons



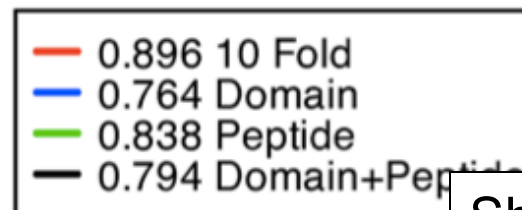
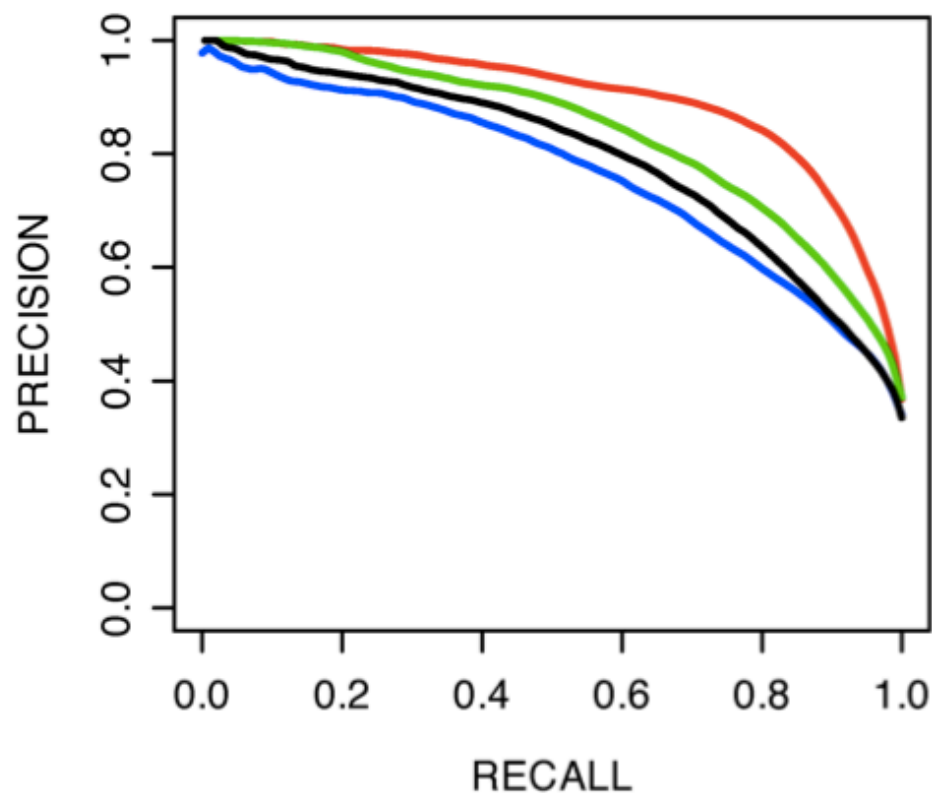
PDZ-Peptide Interaction Prediction

SVM using binding site contact map feature encoding
Cross validation, independent comparison with known PPIs

ROC



Precision Recall



Shirley Hui

Proteome scanning to predict PDZ domain interactions using SVMs

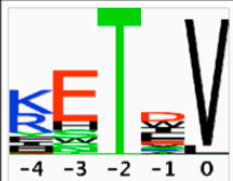
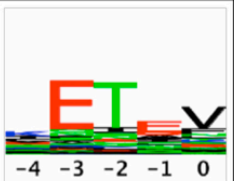
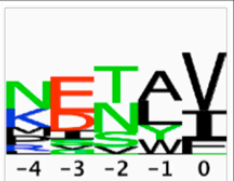
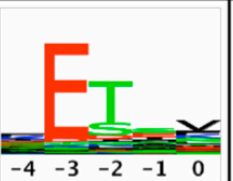
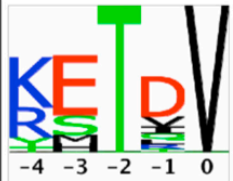
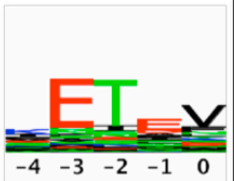
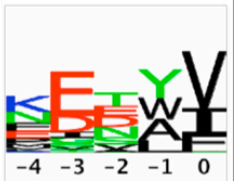
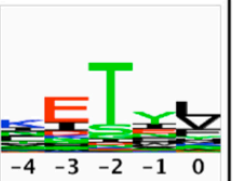



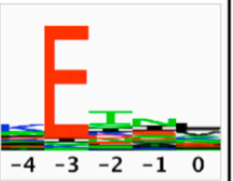
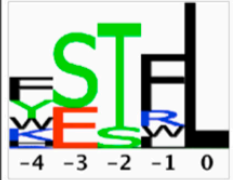
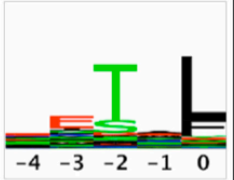
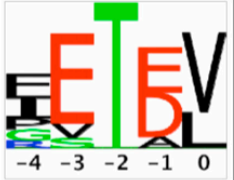
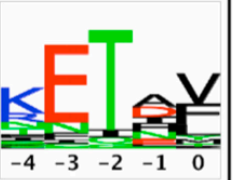
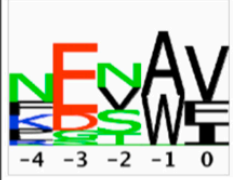
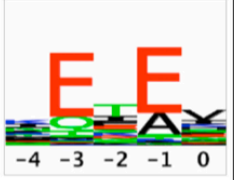

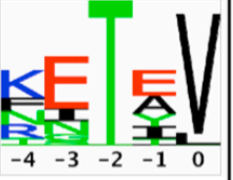
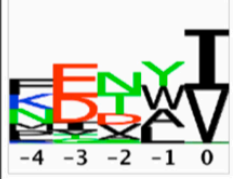
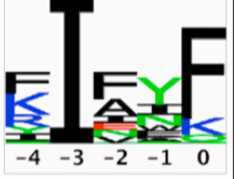

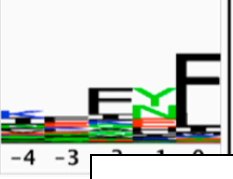
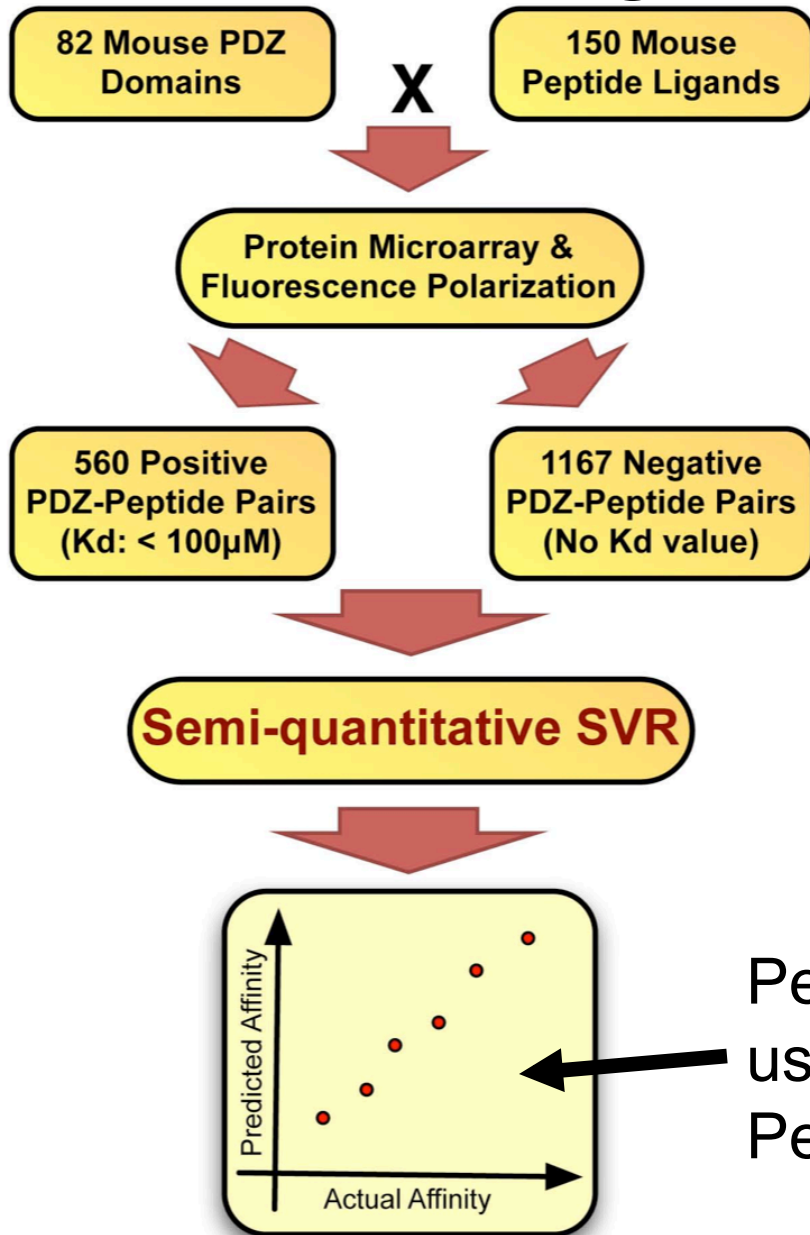
Domain Name	NN Sim	Experiment	SVM Predicted	Profile Sim	Domain Name	NN Sim	Experiment	SVM Predicted	Profile Sim
DLG1-2 Human	1			0.751	LIN7-1 Worm	1			0.688
DLG3-2 Human	1			0.682	MPZ1-6 Worm	0.69			0.729
MLLT4-1 Human	0.69			0.62	STN2-1 Worm	0.81			0.688
PDZK1-1 Human	0.81			0.691	LAP4-2 Fly	0.88			0.725
DLG1-3 Worm	0.94			0.671	LAP4-3 Fly	0.75			0.735
DSH-1 Worm	0.81			0.507	PATJ-2 Fly	0.81			0.565

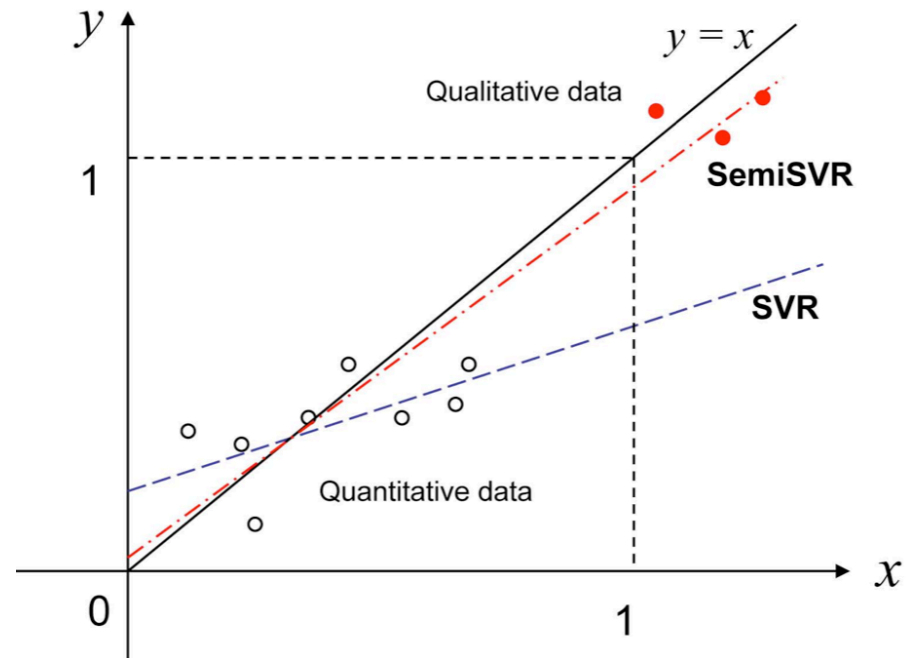
Figure 4

Shirley Hui

Predicting PDZ-Peptide Affinity



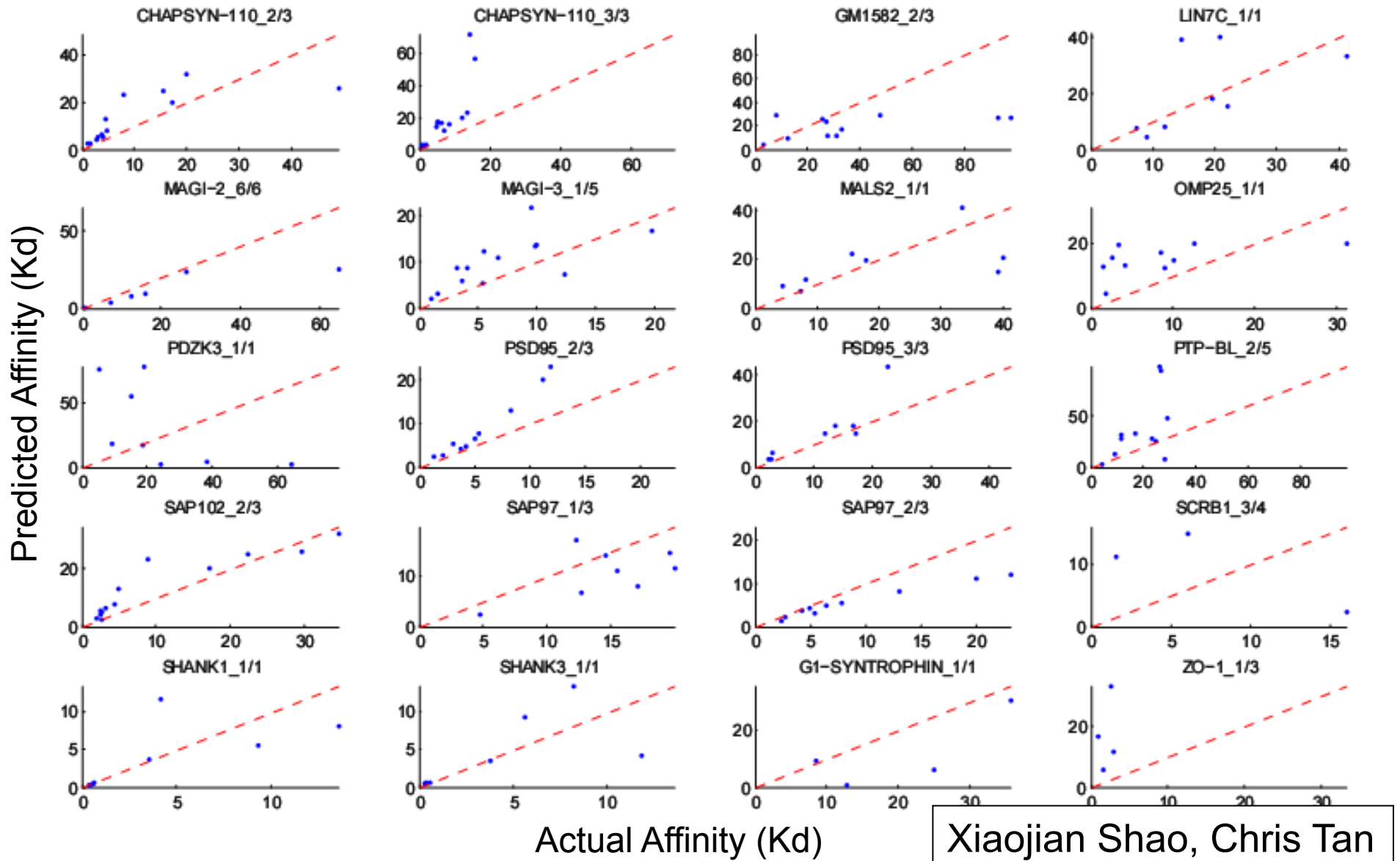
Data: Stiffler, MacBeath (Science 2007)



Performance assessed using Spearman or Pearson correlation

Xiaojian Shao, Chris Tan

Affinity Predicted from Sequence



Xiaojian Shao, Chris Tan

Incorporating Negative Data Improves Quantitative Performance

Measure	Feature encoding	SemiSVR	SVR
Spearman	WS_118AAs	0.605	0.501
	BS_16AAs	0.594	0.425
	CoreBS_10AAs	0.594	0.487
Pearson	WS_118AAs	0.653	0.574
	BS_16AAs	0.636	0.556
	CoreBS_10AAs	0.649	0.585

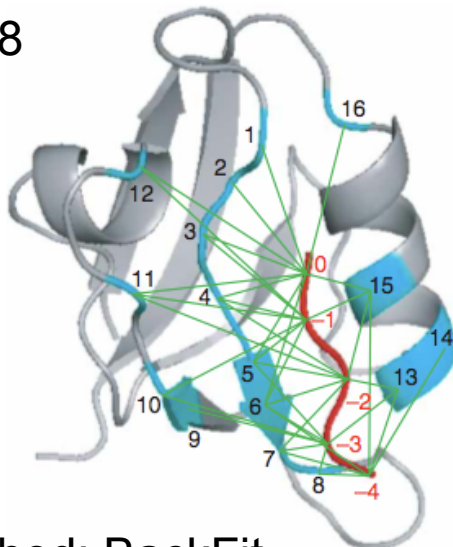
Average Performance Across all PDZ Domains

For WS_118AAs, BS_16AAs and CoreBS_10AAs feature encoding, polynomial kernel with $p=2$ was used for both PDZ domain and peptide sequence. The Pearson correlations were calculated after taking \log_{10} on the actual K_D .

Xiaojian Shao, Chris Tan

Comparison with existing methods

Chen et al. Nat Biotech 2008



Method: BackFit

Features: 38 Pairs

SemiSVR does better in many cases

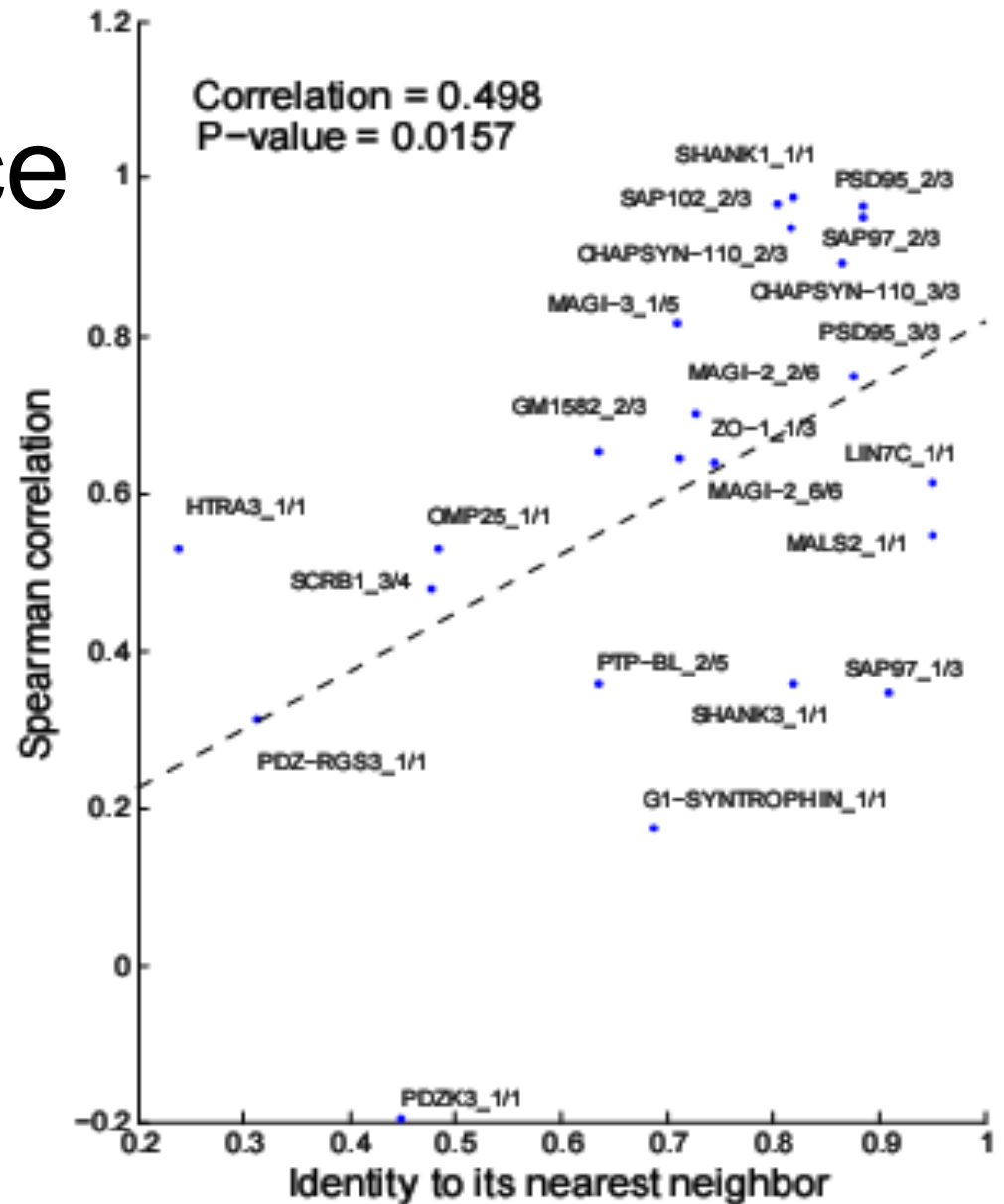
Performance	Spearman correlation			Pearson correlation		
	SemiSVR 118AAs	SemiSVR 38pairs	BackFit	SemiSVR 118AAs	SemiSVR 38pairs	BackFit
CHAPSYN-110_2/3	0.936	0.921	0.802	0.936	0.942	0.793
CHAPSYN-110_3/3	0.891	0.831	0.593	0.874	0.834	0.501
GM1582_2/3	0.653	0.344	0.356	0.580	0.586	0.185
HTRA3_1/1	0.527	0.006	0.2	0.646	0.254	0.130
LIN7C_1/1	0.612	0.394	-0.37	0.682	0.588	-0.168
MAGI-2_2/6	0.700	0.724	0.113	0.769	0.845	0.206
MAGI-2_6/6	0.637	-0.135	0.278	0.689	-0.464	0.169
MAGI-3_1/5	0.816	0.699	0.544	0.876	0.797	0.516
MALS2_1/1	0.545	0.592	0.172	0.612	0.736	0.147
OMP25_1/1	0.528	0.519	0.318	0.504	0.458	0.366
PDZK3_1/1	-0.197	-0.621	-0.224	0.039	-0.544	0.018
PDZ-RGS3_1/1	0.310	0.125	-0.081	0.027	0.117	0.072
PSD95_2/3	0.965	0.905	0.526	0.917	0.897	0.663
PSD95_3/3	0.747	0.817	0.223	0.880	0.899	0.165
PTP-BL_2/5	0.356	0.608	0.184	0.401	0.709	0.159
SAP102_2/3	0.968	0.94	0.907	0.938	0.938	0.936
SAP97_1/3	0.345	0.37	-0.164	0.755	0.563	0.138
SAP97_2/3	0.952	0.942	0.766	0.949	0.927	0.852
SCRB1_3/4	0.479	0.515	0.697	0.694	0.625	0.776
SHANK1_1/1	0.976	0.942	0.954	0.980	0.968	0.956
SHANK3_1/1	0.358	0.517	0.687	0.509	0.639	0.700
G1-SYNTROPHIN_1/1	0.172	0.268	0.524	0.129	0.247	0.483
ZO-1_1/3	0.643	0.186	0.263	0.646	0.217	0.164
Average Performance	0.605	0.496	0.359	0.653	0.556	0.388

Xiaojian Shao, Chris Tan

Performance Increases if Similar Sequence in Training Set

Identity between each tested PDZ to its nearest neighbor is calculated to the other 81 PDZ domains from the 118AAs of the PDZ domain sequence.

Xiaojian Shao, Chris Tan



Protein Domain Interaction Network Prediction

Genome



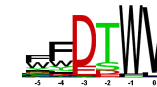
Gene and protein prediction



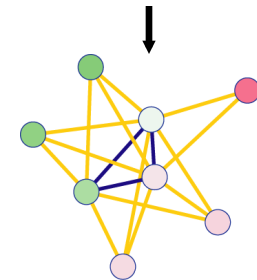
Domain prediction



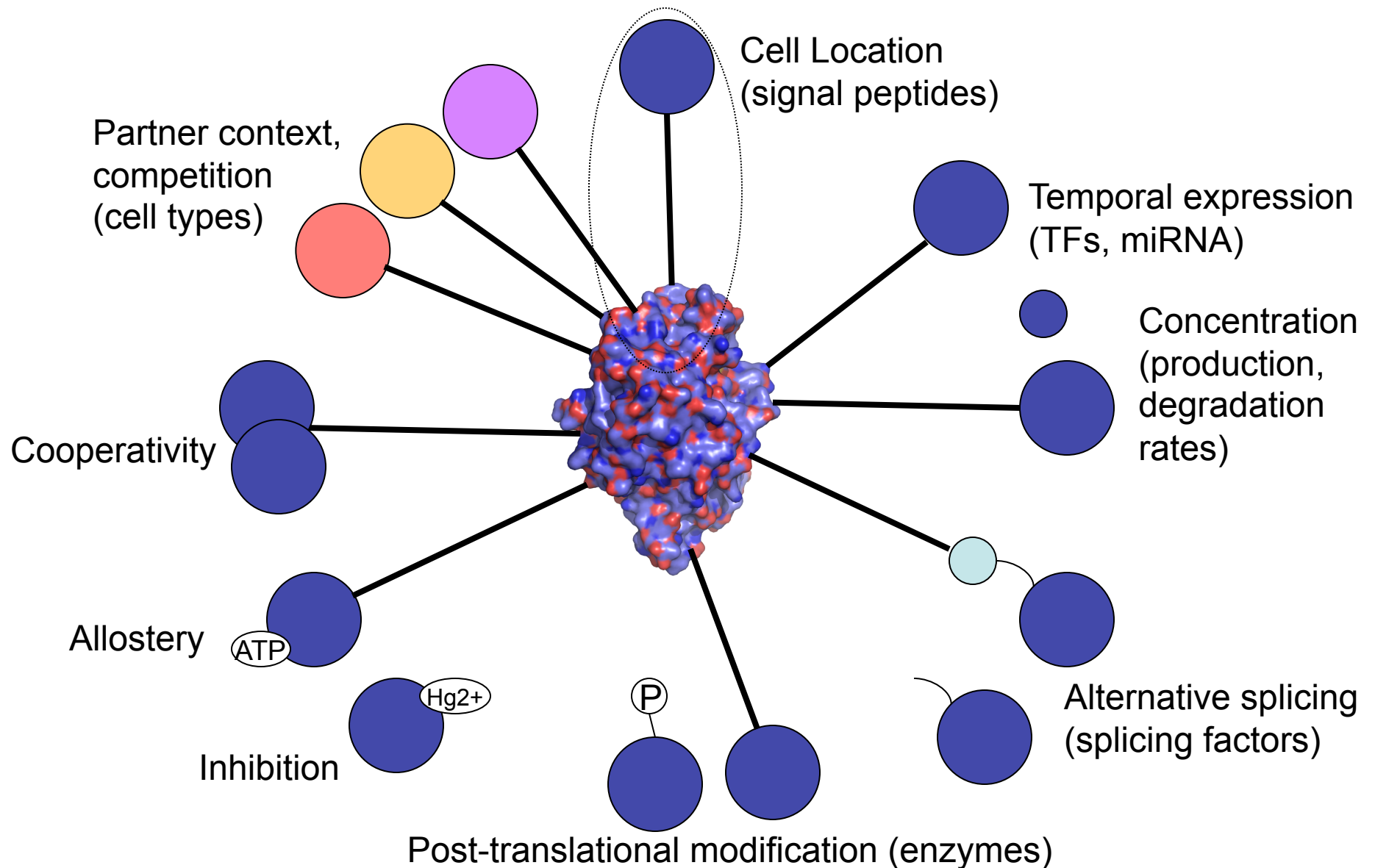
Specificity prediction



Protein-protein interaction prediction

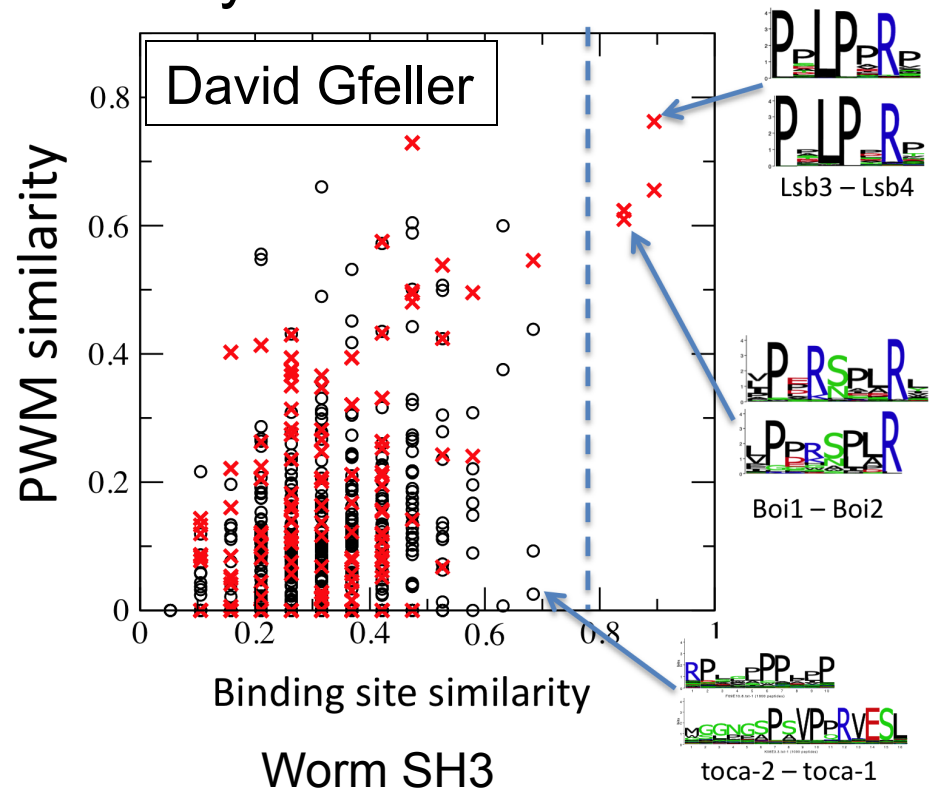


In vivo Interaction Regulation



Conclusions – Part 1

- Specificity and affinity predicted from sequence for PDZ domains
 - Also for WW, but likely not as easy for SH3 and SH2
- Which PDZ to study next? (target selection)
- Negative binding information is extremely useful
- Need more data across species and from different experimental methods



Mapping SH3 Networks Across Species

- Yeast*
 - Phage
 - Peptide display
 - Y2H
- 27 SH3 domains cloned
 - Phage
 - Peptide display
 - Y2H
- High confidence binding site resolution interaction network
 - Bayesian (benchmark)
- Worm
 - Phage, Y2H
- 80/84 SH3 domains
 - Phage, Y2H
- (AD)-ORFeome and AD-cDNA Y2H prey libraries
- 1889 PPIs, incl 13 hubs and 893 full length hits
- High confidence binding site resolution interaction network
 - Phage & Y2H

Xiaofeng Xin
Boone, Sidhu

David Gfeller

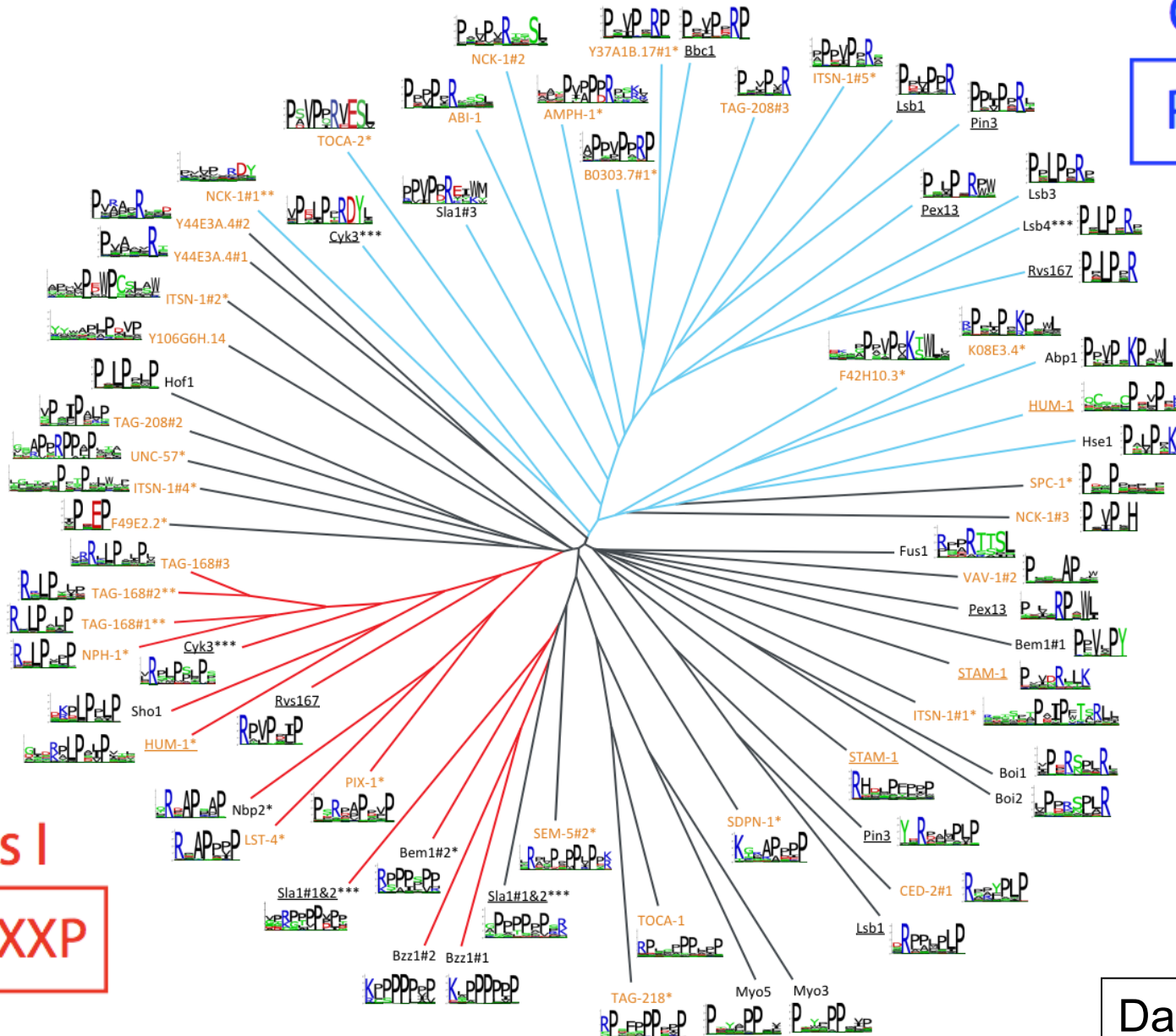
Philip Kim

* Raffi Tonikian, Xiaofeng Xin, Christopher P. Toret, David Gfeller, Christiane Landgraf, Simona Panni, Serena Paoluzi, Luisa Castagnoli, Bridget Currell, Somasekar Seshagiri, Haiyuan Yu, Barbara Winsor, Marc Vidal, Mark B. Gerstein, Gary D. Bader, Rudolf Volkmer, Gianni Cesareni, David G. Drubin, Philip M. Kim, Sachdev Sidhu, Charles Boone – PLoS Biology, Oct.2009

SH3 Specificity Conserved from Yeast To Worm

Class II

PXXPX+

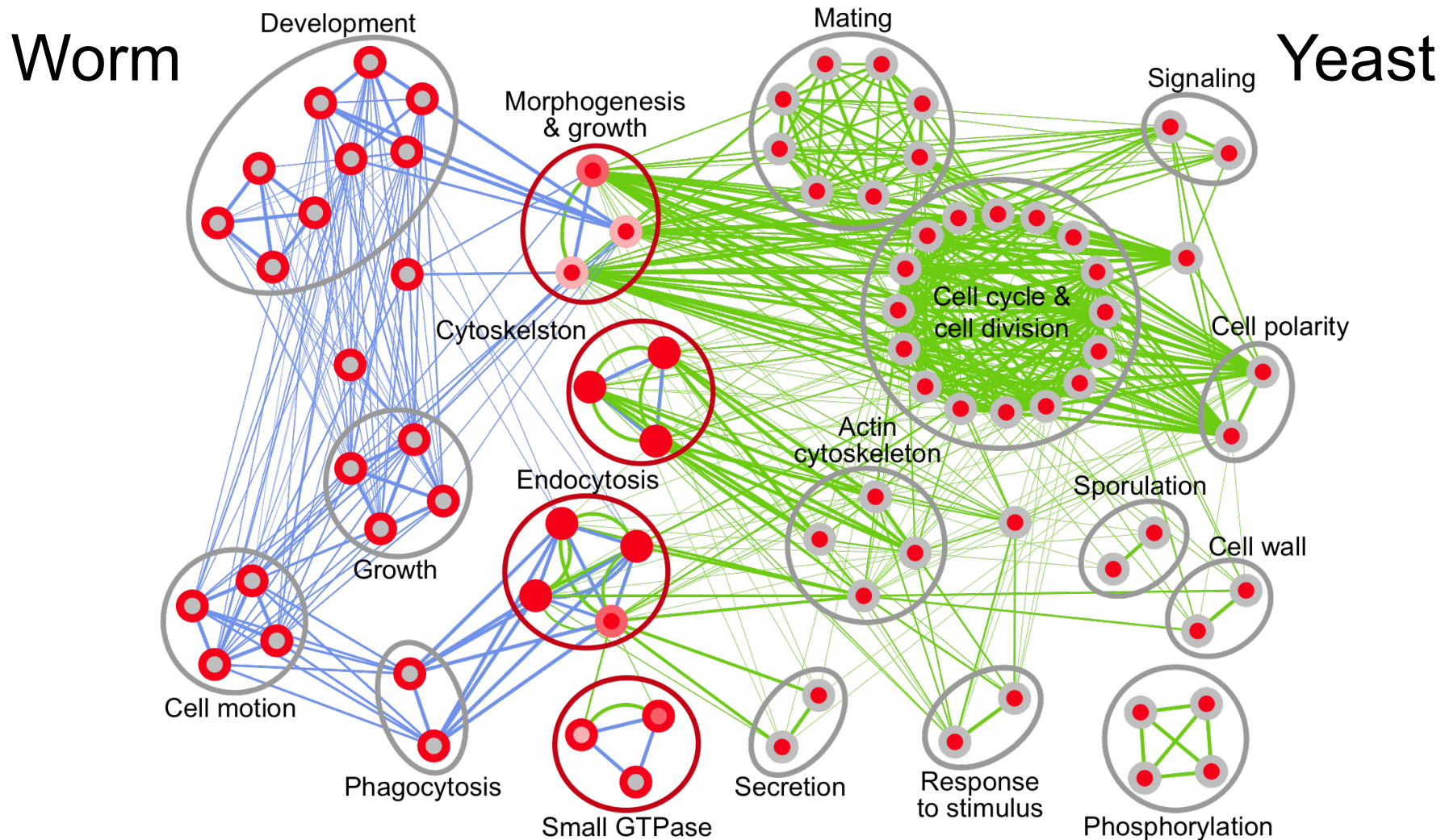


Class I

+XXPXXP

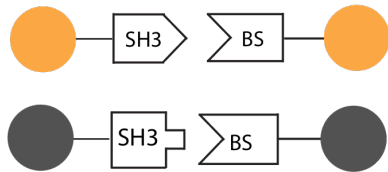
David Gfeller

Yeast & Worm Networks Enriched in Endocytosis



Network is Heavily Rewired

Case 1: Domain Changes



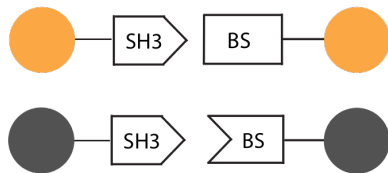
Example

	STAM-1	ACT-2
Worm:		ALPHAILRLDLAGR
Yeast:		SLPHAILRI DLAGR
	Hse1p	Act1p

Homolog SH3 domains involved in this case

STAM-1	HUM-1	HUM-1
Hse1p	Myo3p	Myo5p

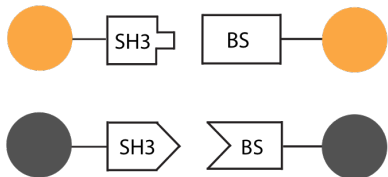
Case 2: Binding Site Changes



	SDPN-1	YMEL-1
Worm:		SRNGINRKPLDIEAT
Yeast:		SRNIPPPPPPPPKP
	Bzz1p#1	Yta12p

SDPN-1	SDPN-1	K08E3.4
Bzz1p#1	Bzz1p#2	Abp1p

Case 3: Both change



	HUM-1	BUB-1
Worm:		SDTKSASEKDVSDSD
Yeast:		QDLSSQPPVVPKST
	Myo3p	Bub1p

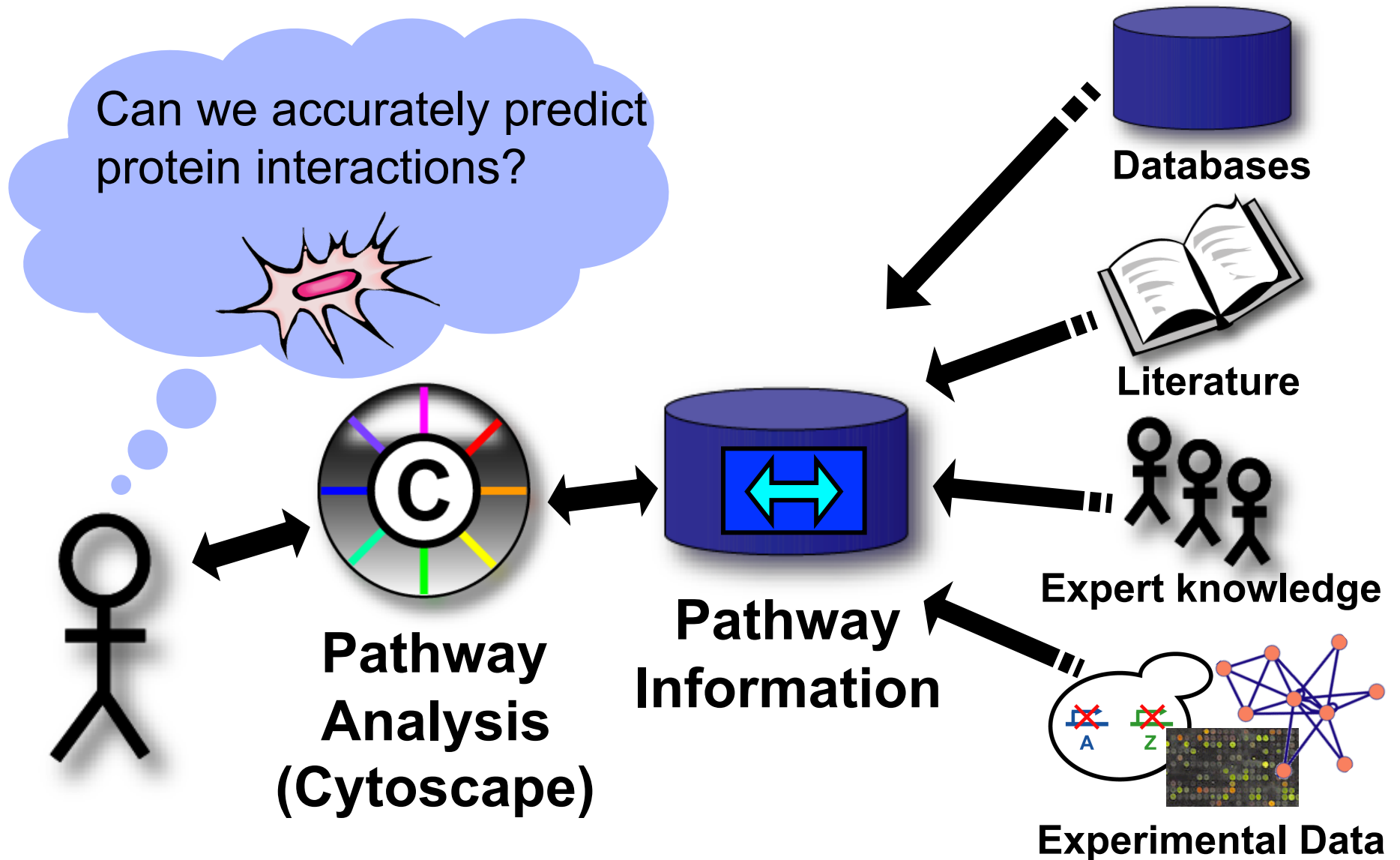
STAM-1	HUM-1	HUM-1
Hse1p	Myo3p	Myo5p

7 ortholog pairs, 5 with yeast & worm profiles
No conserved interactions!

Conclusions – Part 2

- Endocytosis role of SH3 domains is conserved functionally but not structurally
- Low number of yeast/worm SH3 orthologs
 - Higher between worm and human
- Increased coverage may change this conclusion
- Interolog methods will fail in this case
- Functional orthologs?
- Need to experimentally map SH3 networks in additional species

Cell map exploration and analysis



http://pathguide.org

Pathguide» the pathway resource list

Navigation

- Protein-Protein Interactions
- Metabolic Pathways
- Signaling Pathways
- Pathway Diagrams
- Transcription Factors / Gene Regulatory Networks
- Protein-Compound Interactions
- Genetic Interaction Networks
- Protein Sequence Focused
- Other

Search

Organisms
All

Availability
All

Standards
All

Reset Search

Statistics

Analyze Pathguide

Contact

Comments, Questions, Suggestions are Always Welcome!

Complete Listing of All Pathguide Resources

Pathguide contains information about **222** biological pathway resources. Click on a link to go to the resource home page or 'Details' for a description page. Databases that are free and those supporting BioPAX, CellML, PSI-MI or SBML standards are respectively indicated.

If you know of a pathway resource that is not listed here, or have other questions or comments, please [send us an e-mail](#).

>300 Pathway Databases!

Many new search options are available

Get the Stats

Detailed Pathguide resource statistics now available

Pathguide Published

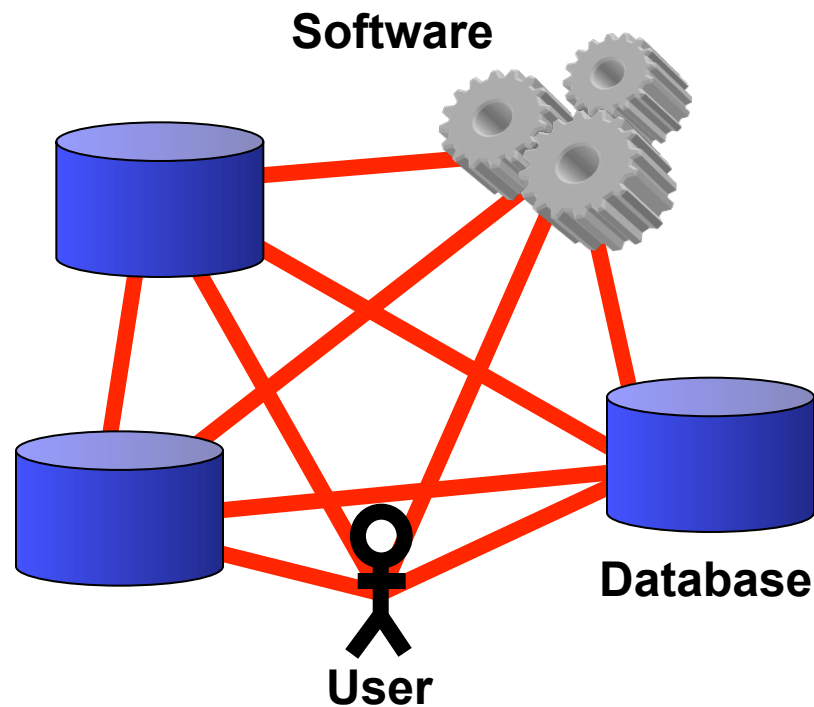
Please cite the [Pathguide](#).

Protein-Protein Interactions

Database Name (Order: alphabetically by web popularity)	Full Record	Availability	Standards
3DID - 3D interacting domains	Details	Free	
ABCdb - Archaea and Bacteria ABC transporter database	Details	Free	
AfCS - Alliance for Cellular Signaling Molecule Pages Database	Details	Free	
AllFuse - Functional Associations of Proteins in Complete Genomes	Details	Free	
ASEdb - Alanine Scanning Energetics Database	Details	Free	
ASPD - Artificial Selected Proteins/Peptides Database	Details	?	
BID - Binding Interface Database	Details	Free	
BIND - Biomolecular Interaction Network Database	Details	Free	PSI-MI
BindingDB - The Binding Database	Details	Free	
BioGRID - General Repository for Interaction Datasets	Details		PSI-MI
BRITE - Biomolecular Relations in Information Transmission and Expression	Details	Free	
CA1Neuron - Pathways of the hippocampal CA1 neuron	Details	Free	
Cancer Cell Map - The Cancer Cell Map	Details	Free	BioPAX
CSP - Cytokine Signaling Pathway Database	Details	Free	
CTDB - Calmodulin Target Database	Details	Free	
DDIB - Database of Domain Interactions and Bindings	Details	Free	
DIP - Database of Interacting Proteins	Details		PSI-MI
Doodle - Database of oligomerization domains from lambda experiments	Details	Free	
DopaNet - DopaNet	Details	Free	
DRC - Database of Ribosomal Crosslinks	Details	Free	
DSM - Dynamic Signaling Maps	Details		
FIMM - Functional Molecular Immunology	Details	Free	
FusionDB - Prokaryote Gene Fusion Events	Details	Free	

Vuk Pavlovic

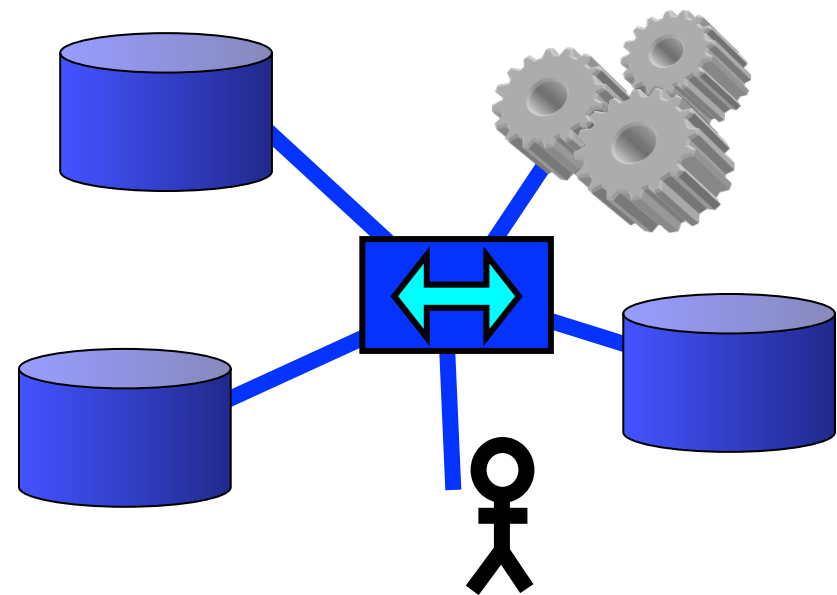
Biological Pathway Exchange (BioPAX)



Before BioPAX

>100 DBs and tools

Tower of Babel



After BioPAX

Unifying language

Reduces work, promotes collaboration, increases accessibility

BioPAX Pathway Language

- Represent:
 - Metabolic pathways
 - Signaling pathways
 - Protein-protein, molecular interactions
 - Gene regulatory pathways
 - Genetic interactions
- Community effort: pathway databases distribute pathway information in standard format

BioPAX Supporting Groups

Current Participants

- Memorial Sloan-Kettering Cancer Center: E.Demir, M. Cary, C. Sander
- University of Toronto: G. Bader
- SRI Bioinformatics Research Group: P. Karp, S. Paley, J. Pick
- Bilkent University: U. Dogrusoz
- Université Libre de Bruxelles: C. Lemer
- CBRC Japan: K. Fukuda
- Dana Farber Cancer Institute: J. Zucker
- Millennium: J. Rees, A. Ruttenberg
- Cold Spring Harbor/EBI: G. Wu, M. Gillespie, P. D'Eustachio, I. Vastrik, L. Stein
- BioPathways Consortium: J. Luciano, E. Neumann, A. Regev, V. Schachter
- Argonne National Laboratory: N. Maltsev, E. Marland, M.Syed
- Harvard: F. Gibbons
- AstraZeneca: E. Pichler
- BIOBASE: E. Wingender, F. Schacherer
- NCI: M. Aladjem, C. Schaefer
- Università di Milano Bicocca, Pasteur, Rennes: A. Splendiani
- Vassar College: K. Dahlquist
- Columbia: A. Rzhetsky

Collaborating Organizations

- Proteomics Standards Initiative (PSI)
- Systems Biology Markup Language (SBML)
- CellML
- Chemical Markup Language (CML)

Databases

- BioCyc, WIT, KEGG, PharmGKB, aMAZE, INOH, Transpath, Reactome, PATIKA, eMIM, NCI PID, CellMap, NetPath

Wouldn't be possible without

Gene Ontology

Protégé, U.Manchester, Stanford

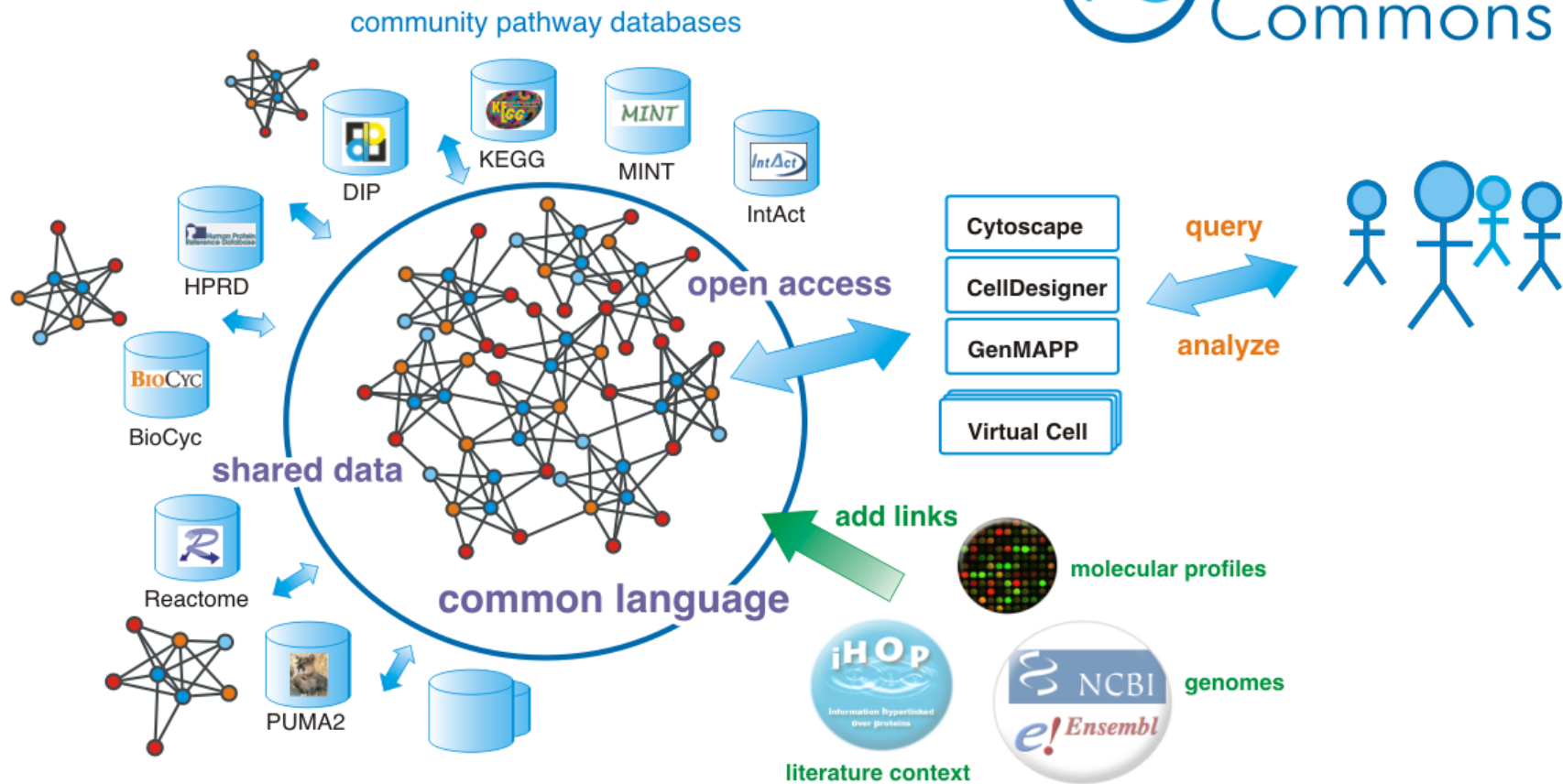
Grants/Support

- Department of Energy (Workshop)
- caBIG



Pathway Commons: A Public Library

<http://pathwaycommons.org>



- Books: Pathways
- Lingua Franca: BioPAX

- Index: cPath pathway database software
- Translators: translators to BioPAX

<http://cytoscape.org>

Network visualization and analysis

Pathway comparison
Literature mining
Gene Ontology analysis
Active modules
Complex detection
Network motif search

UCSD, ISB, Agilent,
MSKCC, Pasteur,
UCSF

Cytoscape Desktop (New Session)

File Edit View Select Layout Plugins Help

Sample2 Search: yal003w

CytoPanel 1

Network Editor

Network	Nodes	Edges
galFiltered.sif	331(19)	362(35)

CytoPanel 2

Node Attribute Browser (galFiltered.sif)

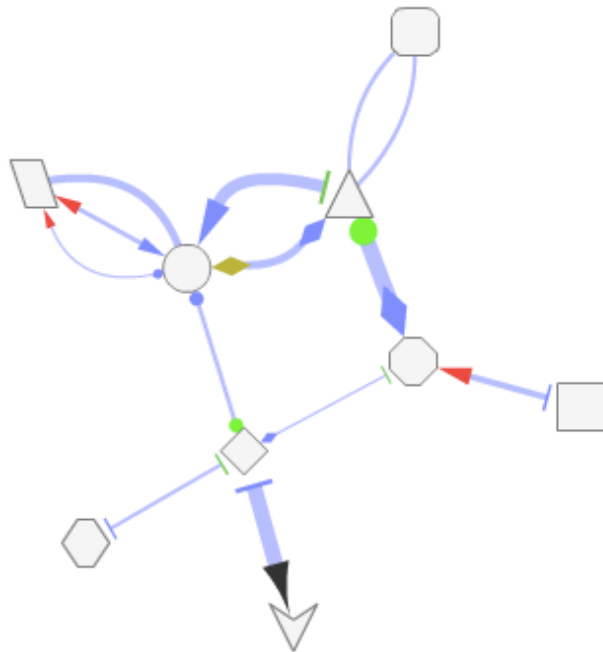
ID	gal1RGexp	gal1RGsig	gal4RGexp	gal4RGsig	gal80Rexp	gal80Rsig
YGL008C	-0.352	1.0007E-5	-0.282	7.1366E-4	-0.573	1.2622E-5
YCL067C	0.169	0.0012873	-0.085	0.11481	0.301	0.0027E-5
YNL145W	-0.764	3.148E-11	-0.098	0.05338	-1.237	1.1916E-5
YMP043W	-0.183	0.0035372	-0.654	4.2514E-6	0.457	2.4112E-5

Welcome to Cytoscape 2.4.0-b1

Right-click + drag to ZOOM Middle-click + drag to PAN



Save file Open file Style ▾ Layout ▾



Examples Visual style Filter Properties

Shapes example



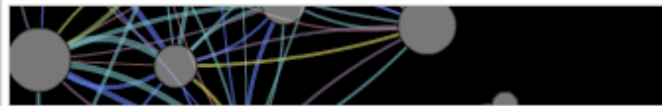
A graph that contains all possible shapes for nodes and arrows

Petersen example



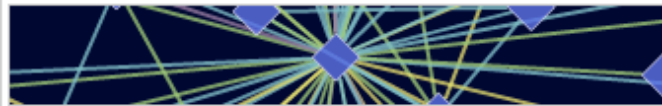
The Petersen graph

Disconnected example



A graph that contains several, disconnected components

Genetics example



A modified graph from GeneMANIA with different visual styles

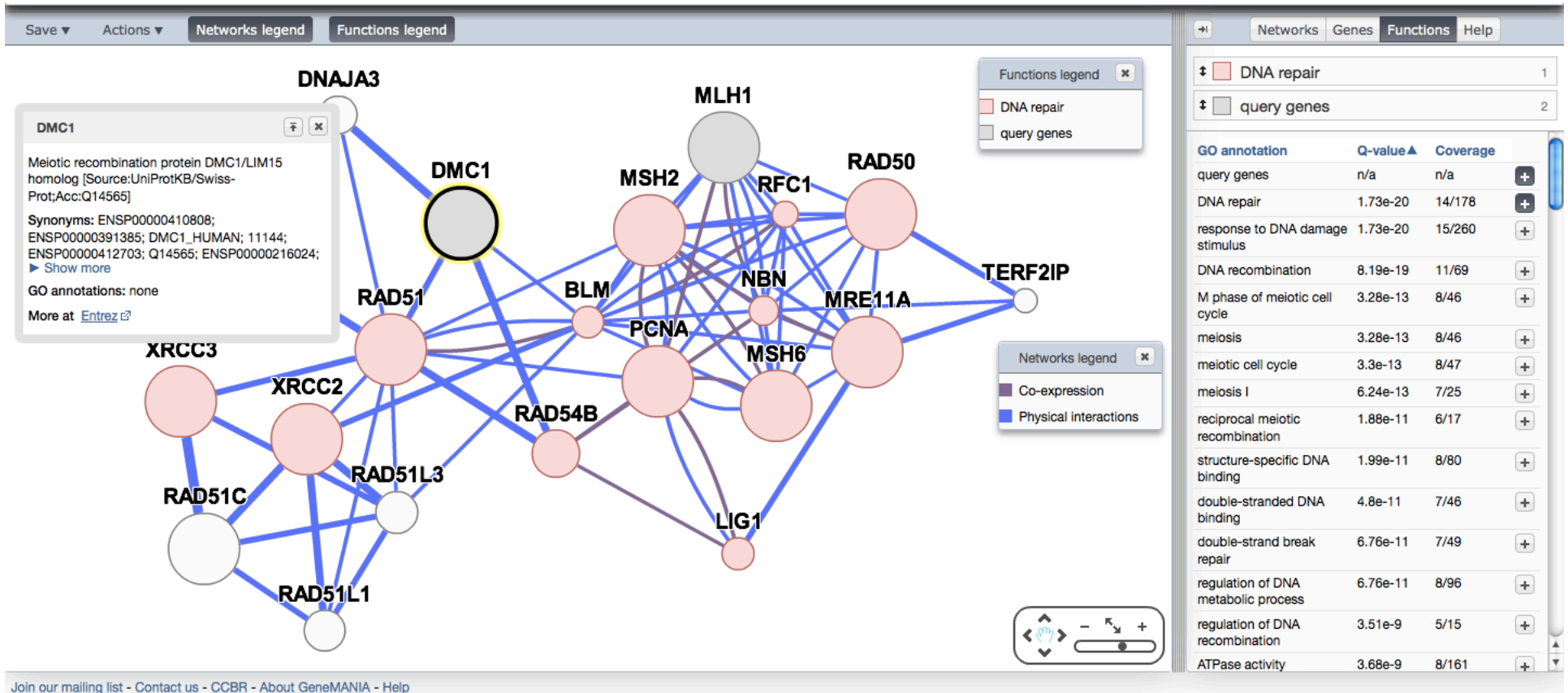
http://www.genemania.org

GENEMANIA
fast gene function predictions

Find genes in

related to

[Show advanced options](#)



[Join our mailing list](#) - [Contact us](#) - [CCBR](#) - [About GeneMANIA](#) - [Help](#)

- Guilt-by-association principle
- Biological networks are combined intelligently to optimize prediction accuracy
- Algorithm is more fast and accurate than its peers

Gene Function Prediction

Quaid Morris (CCBR)

Rashad Badrawi, Ovi Comes, Sylva Donaldson, Christian Lopes, Farzana Kazi, Jason Montojo, Harold Rodriguez, Khalid Zuberi

Acknowledgements

PDZ & SH3 Data

Dev Sidhu (Phage)

Raffi Tonikian, Yingnan Zhang, Andreas Ernst, Heike Held, Stephen Szinsky, Yan Wu

Charlie Boone, Marc Vidal (Y2H)

Xiaofeng Xin

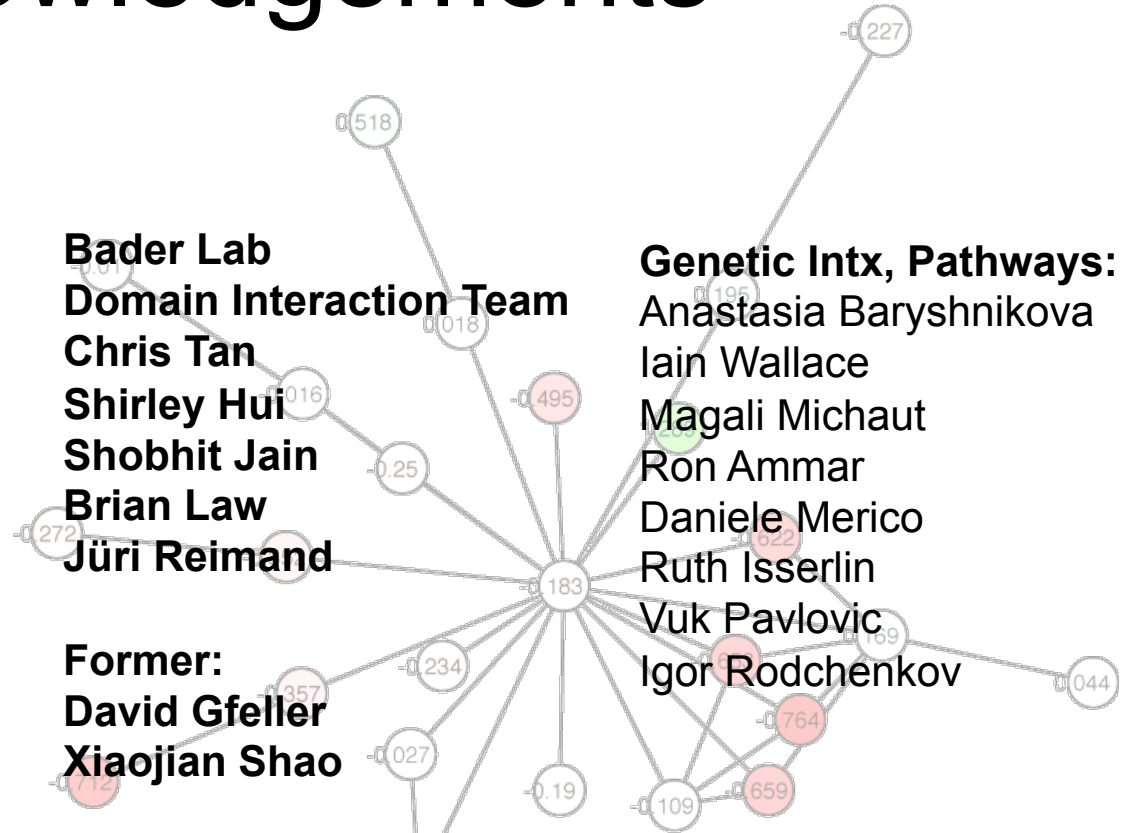
Bader Lab Domain Interaction Team

Chris Tan
Shirley Hui
Shobhit Jain
Brian Law
Jüri Reimand

Former:
David Gfeller
Xiaojian Shao

Genetic Intx, Pathways:

Anastasia Baryshnikova
Iain Wallace
Magali Michaut
Ron Ammar
Daniele Merico
Ruth Isserlin
Vuk Pavlovic
Igor Rodchenkov



Funding



CIHR IRSC



GenomeCanada



Ontario Genomics Institute



NSERC
CRSNG



Canada Foundation for Innovation
Fondation canadienne pour l'innovation



Ontario



www.GeneMANIA.org
Quaid Morris (Donnelly)
Rashad Badrawi, Ovi Comes, Sylva Donaldson,
Christian Lopes,
Farzana Kazi, Jason Montojo,
Harold Rodriguez,
Khalid Zuberi

<http://baderlab.org>

Acknowledgements

Cytoscape

Trey Ideker (UCSD)

Mike Smoot, Kei Ono, Peng Liang Wang (Ryan Kelley, Nerius Landys, Chris Workman, Mark Anderson, Nada Amin, Owen Ozier, Jonathan Wang)

Lee Hood, Ilya Schmulevich (ISB)

Sarah Killcoyne (Iliana Avila-Campillo, Rowan Christmas, Andrew Markiel, Larissa Kamenkovich, Paul Shannon)

Benno Schwikowski (Pasteur)

Mathieu Michaud (Melissa Cline, Tero Aittokallio, Andrea Splendiani)

Chris Sander (MSKCC)

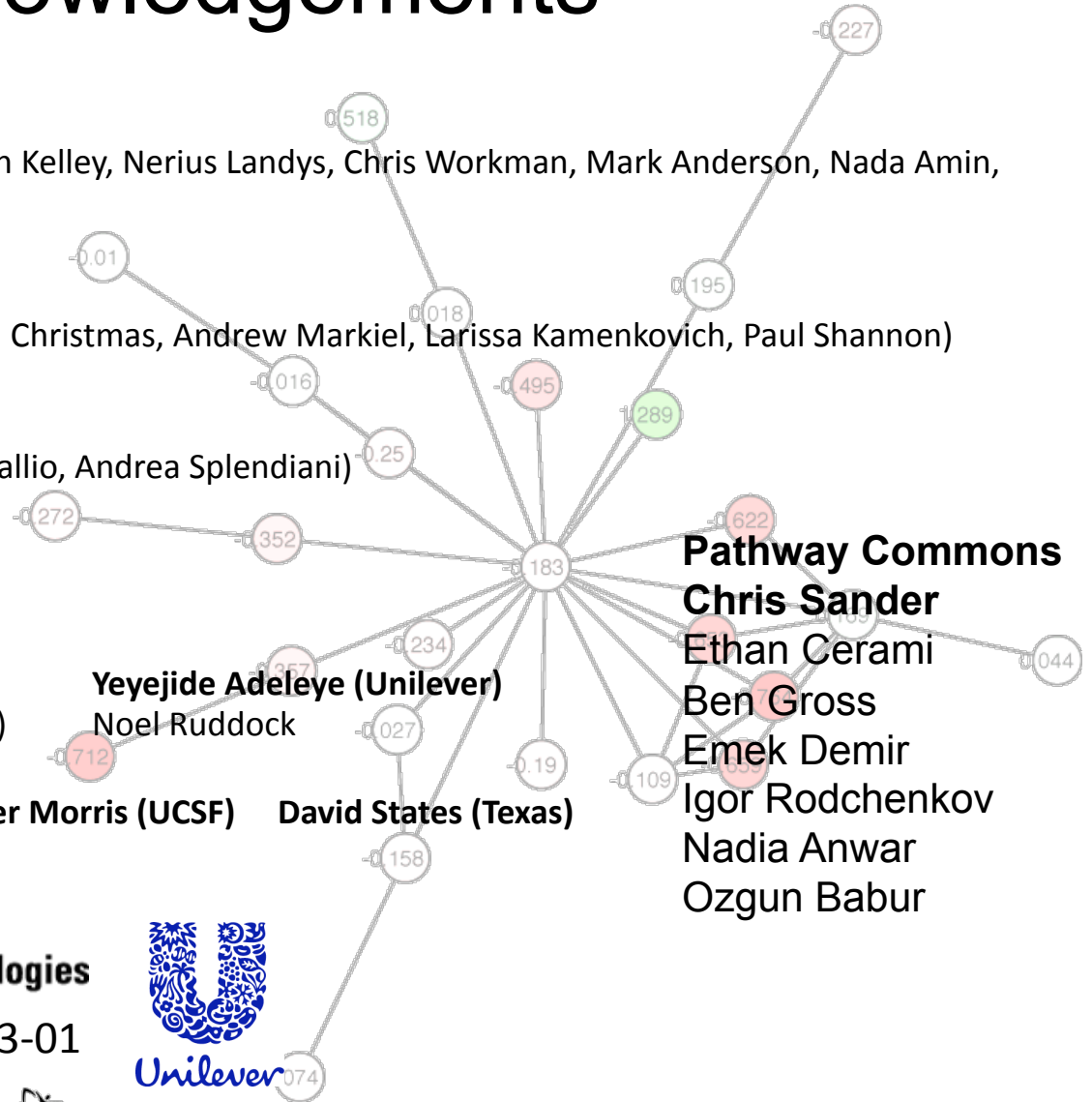
Ethan Cerami, Ben Gross (Robert Sheridan)

Annette Adler (Agilent)

Allan Kuchinsky, Mike Creech (Aditya Vailaya)

Bruce Conklin (UCSF)

Alex Pico, Kristina Hanspers



Pathway Commons

Chris Sander

Ethan Cerami

Ben Gross

Emek Demir

Igor Rodchenkov

Nadia Anwar

Ozgun Babur

Yeyejide Adeleye (Unilever)

Noel Ruddock

Scooter Morris (UCSF)

David States (Texas)



NIGMS GM070743-01



GenomeCanada



OntarioGenomicsInstitute



Ontario

<http://baderlab.org>

Network classification of disease

- Traditional: Gene association
- Limitations: Too many genes reduces statistical power
- New: Active cell map based approaches combining network and molecular profiles

Liu M, Liberzon A, Kong SW, Lai WR, Park PJ, Kohane IS, Kasif S
Network-based analysis of affected biological processes in type 2 diabetes models
PLoS Genet. 2007 Jun;3(6):e96

Chuang HY, Lee E, Liu YT, Lee D, Ideker T
Network-based classification of breast cancer metastasis
Mol Syst Biol. 2007;3:140. Epub 2007 Oct 16

Efroni S, Schaefer CF, Buetow KH
Identification of key processes underlying cancer phenotypes using biologic pathway analysis
PLoS ONE. 2007 May 9;2(5):e425

Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, Bull S, Pawson T, Morris Q, Wrana JL.
Dynamic modularity in protein interaction networks predicts breast cancer outcome.
Nat Biotechnol. 2009 Feb 1