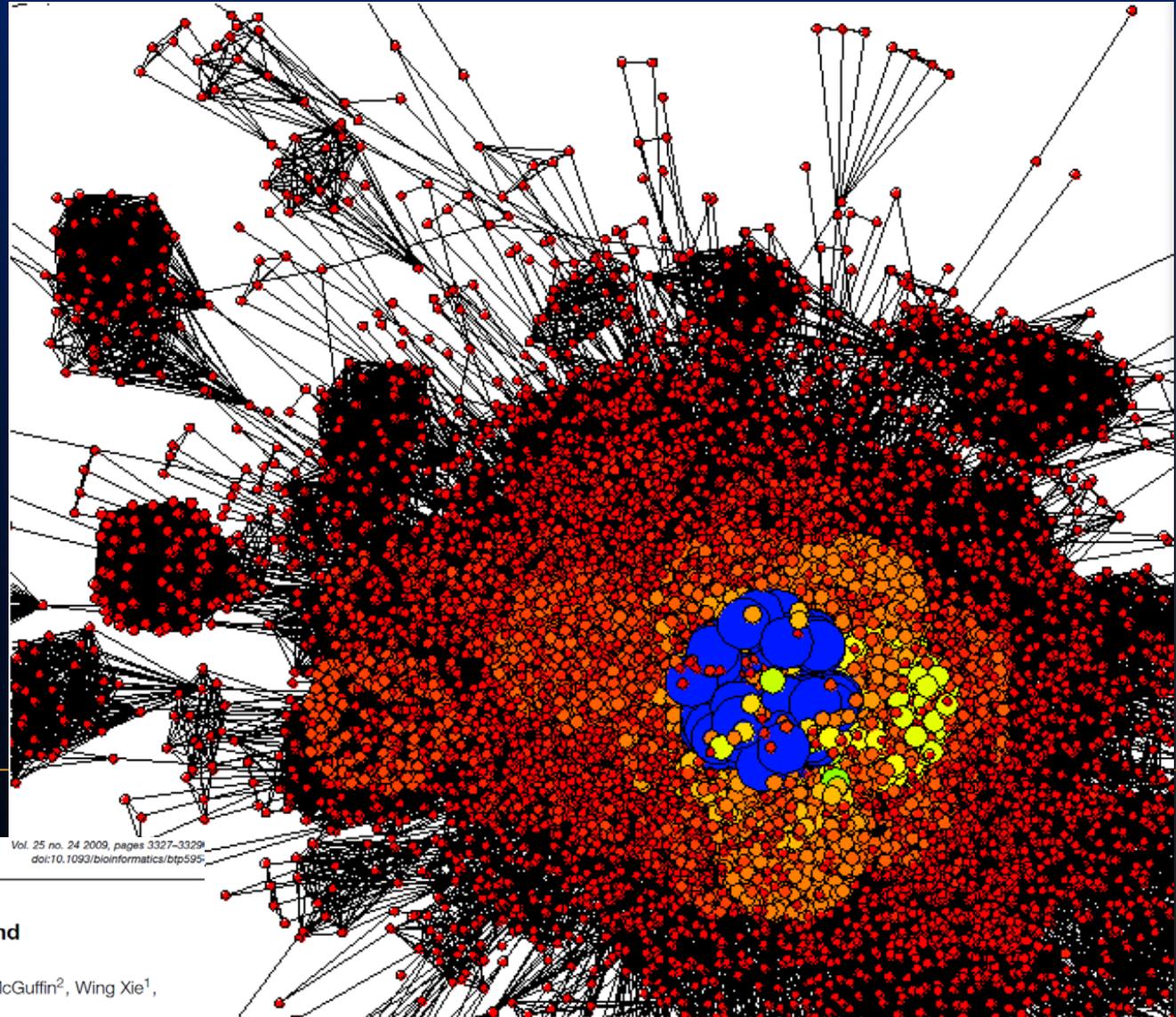


SYSTEMATIC IDENTIFICATION AND CHARACTERIZATION OF CANCER BIOMARKERS

Igor
Jurisica

<http://www.cs.utoronto.ca/~juris>



Vol. 25 no. 24 2009, pages 3327–3329
doi:10.1093/bioinformatics/btp595

BIOINFORMATICS APPLICATIONS NOTE

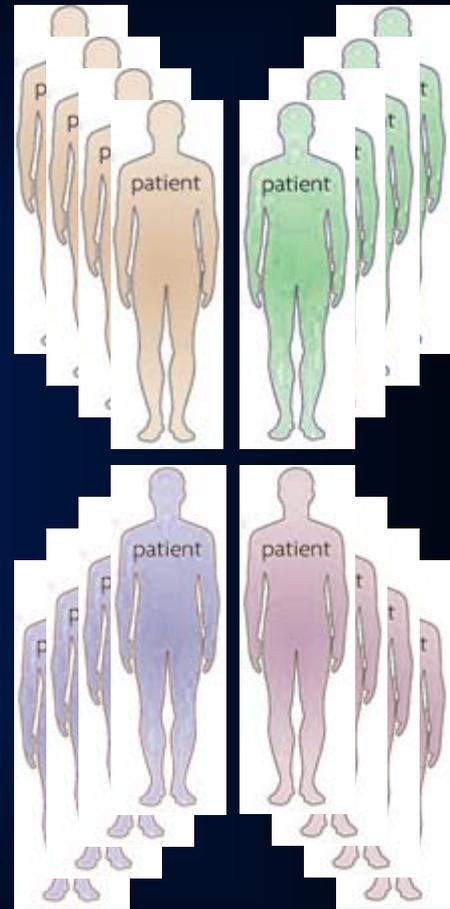
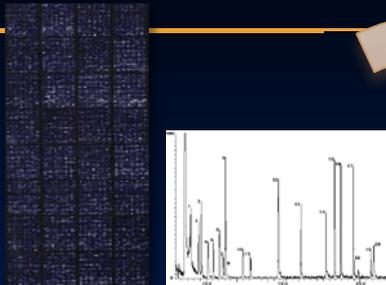
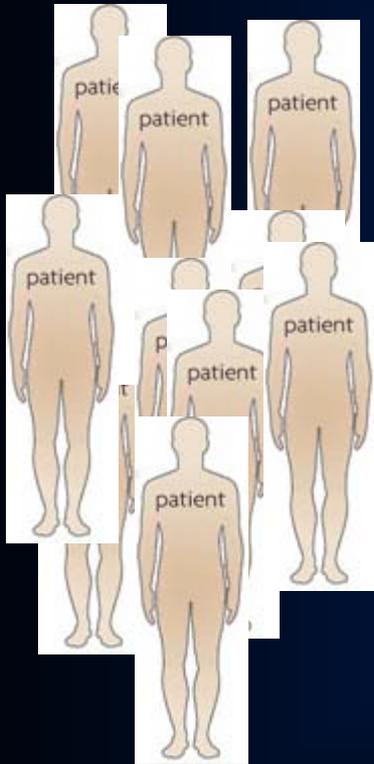
Systems biology

NAVIGATOR: Network Analysis, Visualization and Graphing Toronto

Kevin R. Brown¹, David Otasek¹, Muhammad Ali¹, Michael J. McGuffin², Wing Xie¹,
Baiju Devani¹, Ian Lawson van Toch¹ and Igor Jurisica^{1,3,*}

Learning

Predisposition



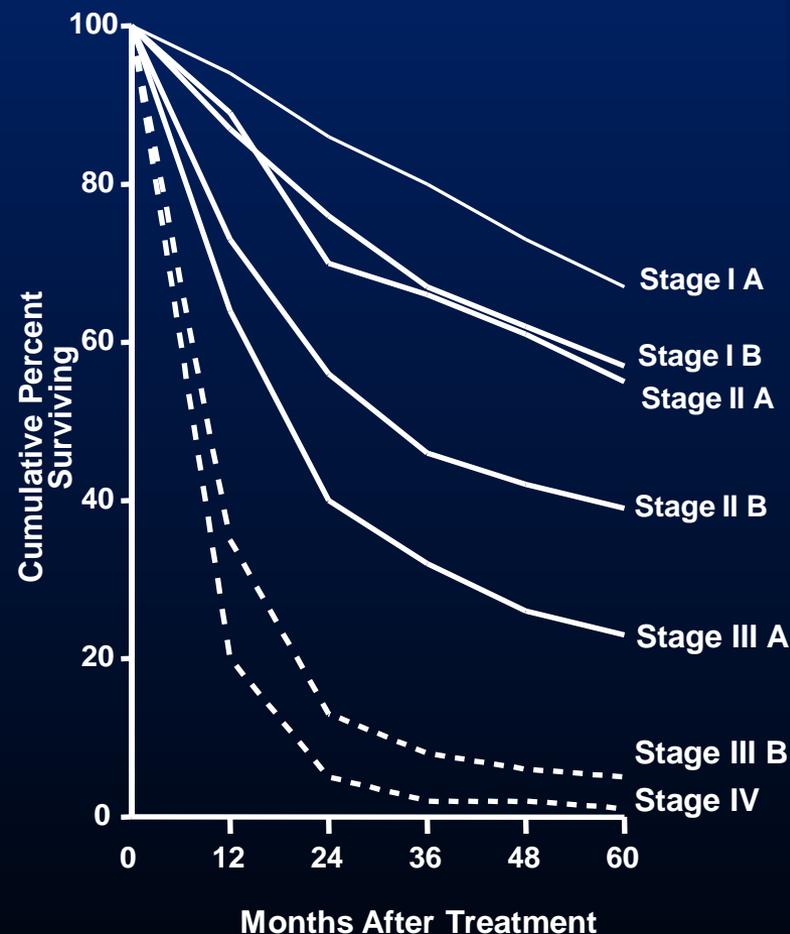
Early markers

Testing
Validation
Application

Diagnosis
Prognosis
Treatment plan

STRATEGY TO REDUCE MORTALITY

- ✗ Increase number of cases diagnosed at earlier stage
 - + Biomarkers for early detection
- ✗ Individualized treatment
 - + Biomarkers for treatment selection & monitoring
- ✗ Improved treatment
 - + Understanding disease mechanism
 - + New medicines



CHALLENGES

- ✘ Signatures do not overlap
 - + most do not validate on external datasets
 - Subramanian & Simon *JNCI*, 2010
 - Ein-Dor et al. *PNAS* 2006
 - Shedden et al. *Nat Med* 2008
 - Lau et al., *JCO* 2007, Boutros et al. *PNAS* 2009.
- ✘ Interactomes comprise false positives and FALSE NEGATIVES
 - + ~40% of signature genes do not have known PPIs
- ✘ PPI dynamics is mostly missing
- ✘ PDB has only ~60K structures

ADDRESSING THE GAP



Home | CGH Map | Search

Change Database ▾

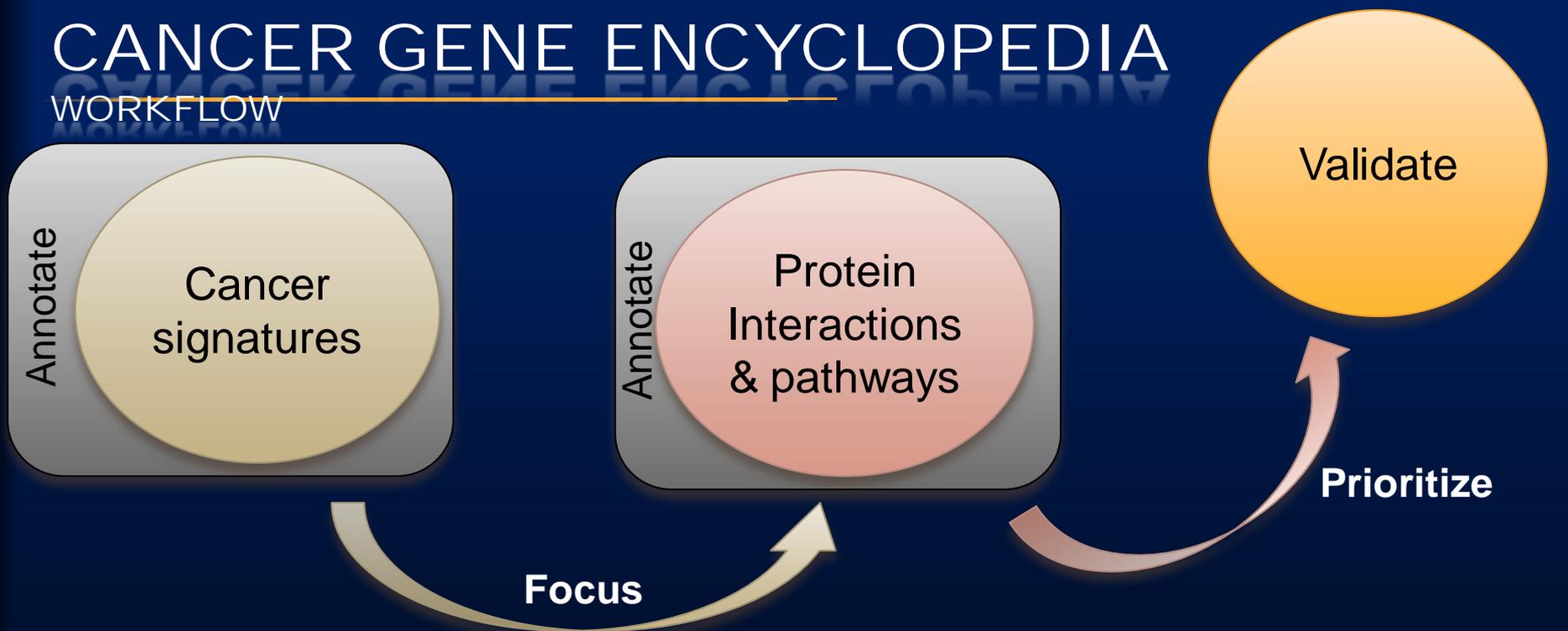
Logout **ijuristica**

	Ovarian	Prostate	Lung	H&N
Number of individual analyses	61	119	26	108
Differential genes in CDIP	9,818	12,985	1,586	2,774

	Ovarian	Prostate	Lung	H&N
Cancer targets in CDIP	9,818	12,985	1,586	2,774
Cancer targets with no PPIs in I2D	3,962 (40.35%)	5,235 (40.32%)	611 (38.52%)	1,329 (47.91%)
Maximum predicted PPI probability from FPClass	0.946	0.946	0.8887	0.930
Average predicted PPI probability from FPClass	0.357	0.338	0.341	0.364
Number of predicted PPIs supported by text mining with SVM score >0.5	1,183	728	65	NA

CANCER GENE ENCYCLOPEDIA

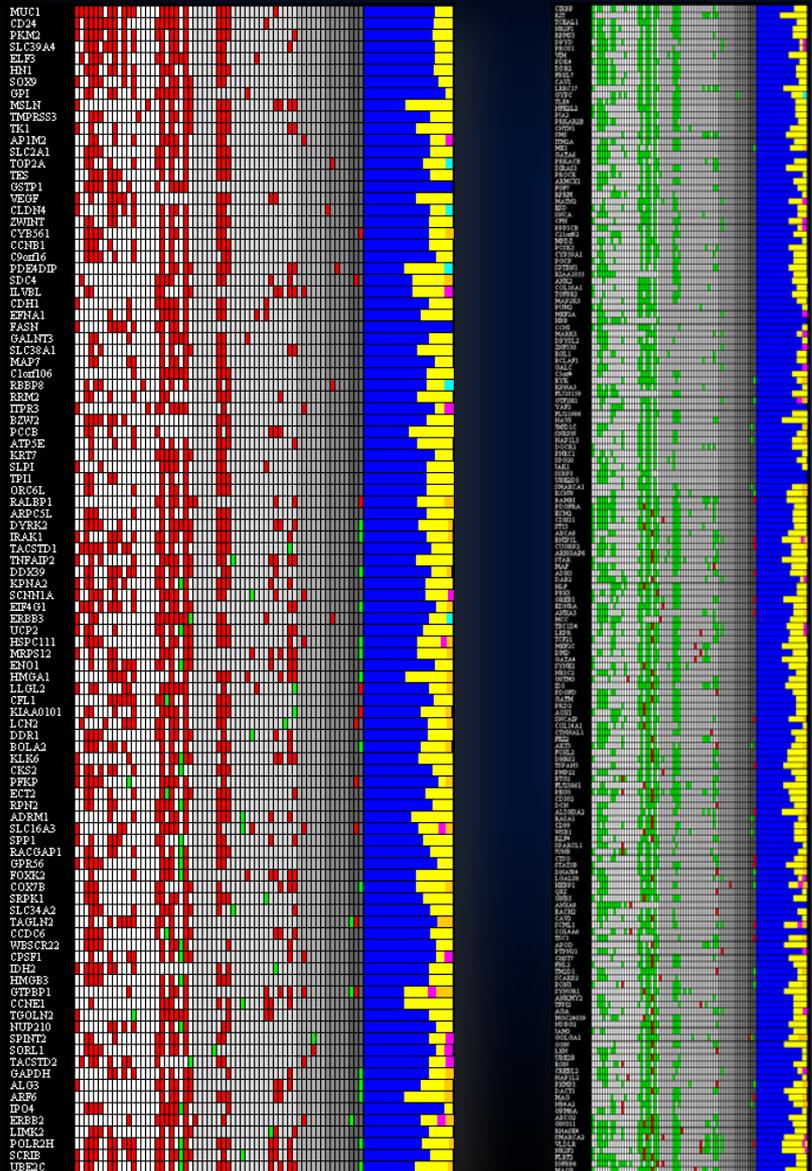
WORKFLOW



Lung, ovarian, breast prostate, head&neck

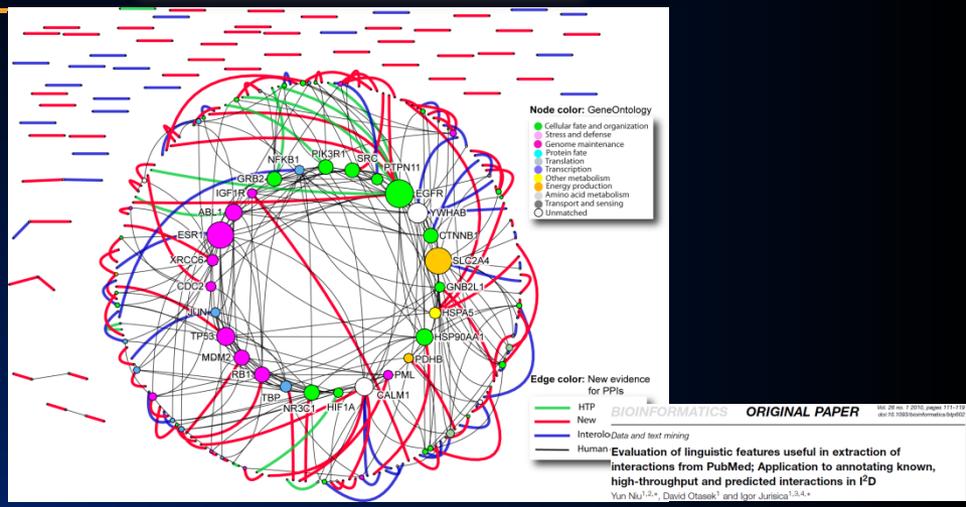
Signatures

ophid.utoronto.ca/cdip

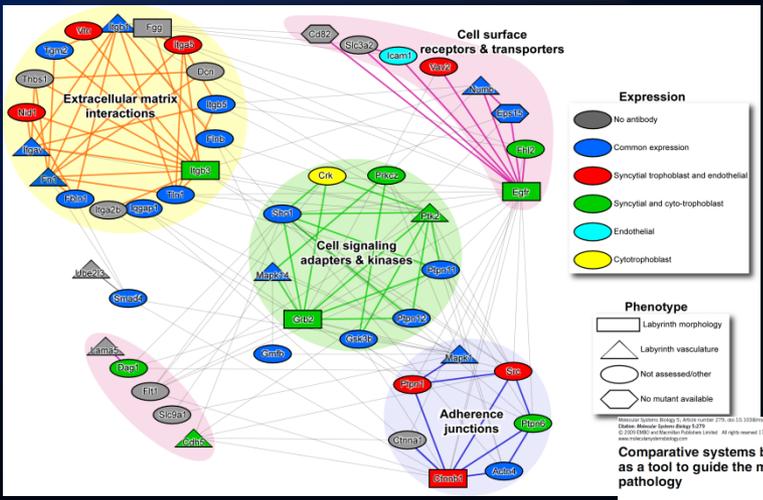


Interactions

ophid.utoronto.ca/i2d



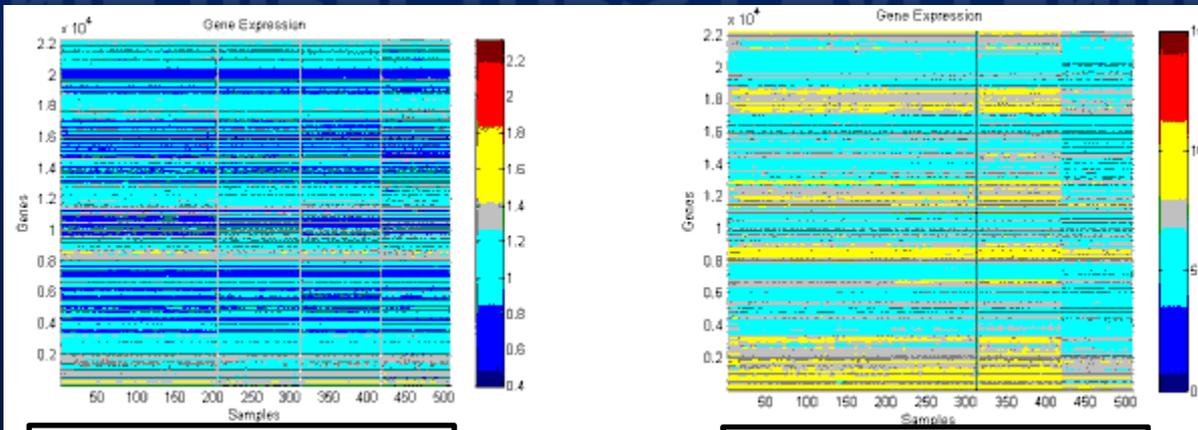
ophid.utoronto.ca/navigator



ophid.utoronto.ca/genecards

Annotation

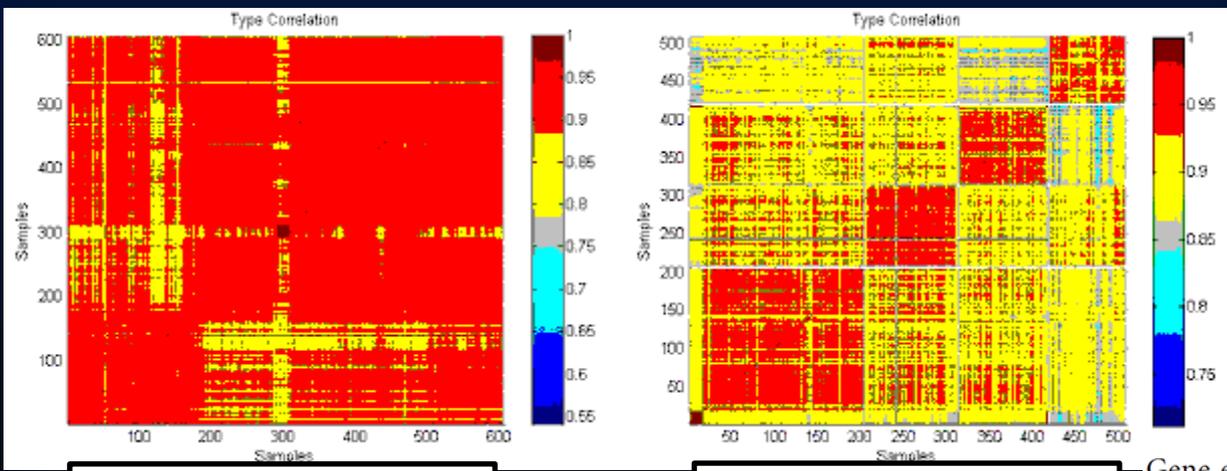
NIH DIRECTOR'S CHALLENGE



MICH MSK MOFF HARV

MICH MSK MOFF HARV

- ~500 lung adenocarc.
- Affy U133
- same preprocessing
- different analyses
- blinded validation



HARV MOFF MICH MSK

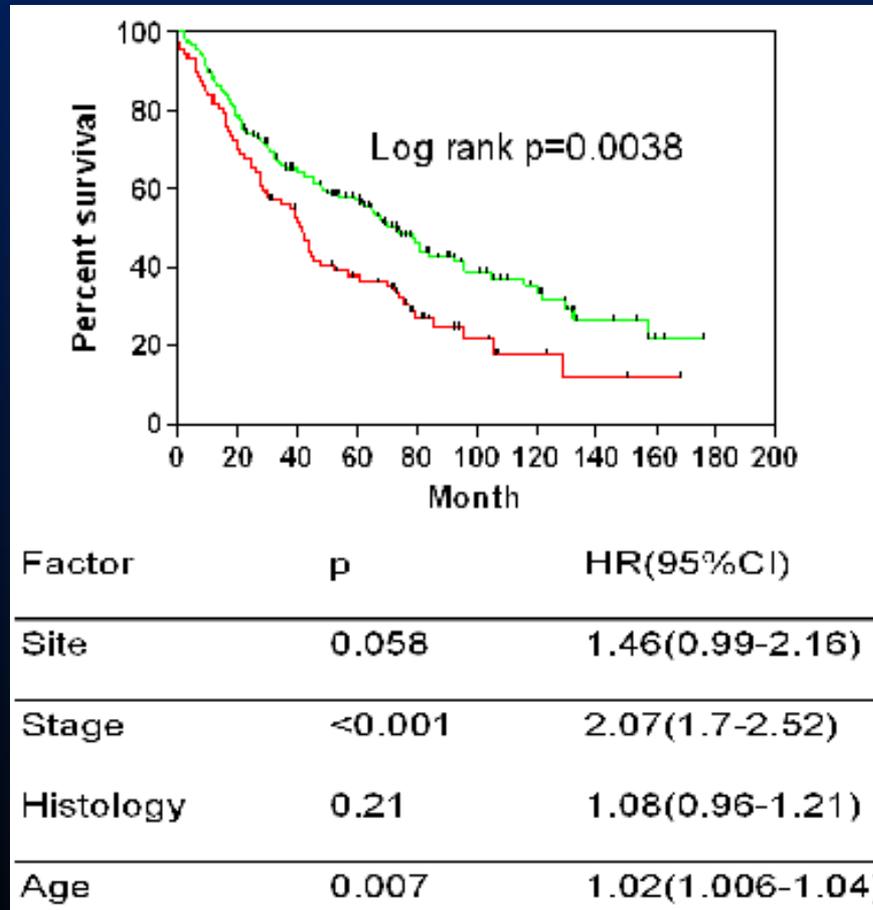
MICH MSK MOFF HARV



Gene expression–based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study

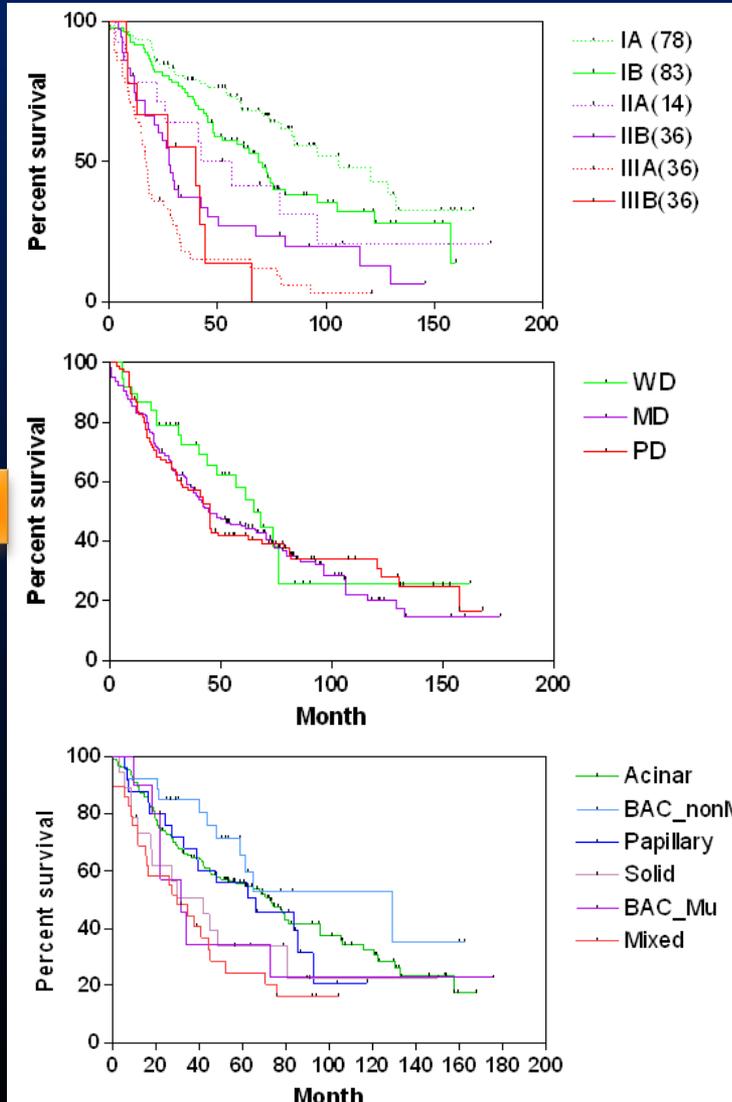
Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma:¹
 Kerby Shedden^{2,3,17}, Jeremy M G Taylor^{3,4,17}, Steven A Enkemann^{5,17}, Ming-Sound Tsao^{6,17},
 Timothy J Yeatman^{5,17}, William L Gerald^{7,17}, Steven Eschrich^{5,17}, Igor Jurisica^{6,17}, Thomas J Giordano⁸,
 David E Misek^{3,5}, Andrew C Chang^{3,9}, Chang Qi Zhu⁶, Daniel Strumpf⁶, Samir Hanash³, Frances A Shepherd⁶,
 Keyue Ding¹⁰, Lesley Seymour¹⁰, Katsuhiko Naoki¹¹, Nathan Pennell¹¹, Barbara Weir¹¹, Roel Verhaak¹¹,
 Christine Ladd-Acosta¹², Todd Golub¹², Michael Gruidl⁵, Anupama Sharma⁵, Janos Szoke⁷, Maureen Zakowski⁷,
 Valerie Rusch⁷, Mark Kris⁷, Agnes Viale⁷, Noriko Motoi⁷, William Travis⁷, Barbara Conley¹³,
 Venkatraman E Seshan^{14,17}, Matthew Meyerson^{11,12,17}, Rork Kuick^{3,17}, Kevin K Dobbin^{15,17}, Tracy Lively^{16,17},
 James W Jacobson^{16,17} & David G Beer^{3,5,17}

BTSVQ – ALL PATIENTS

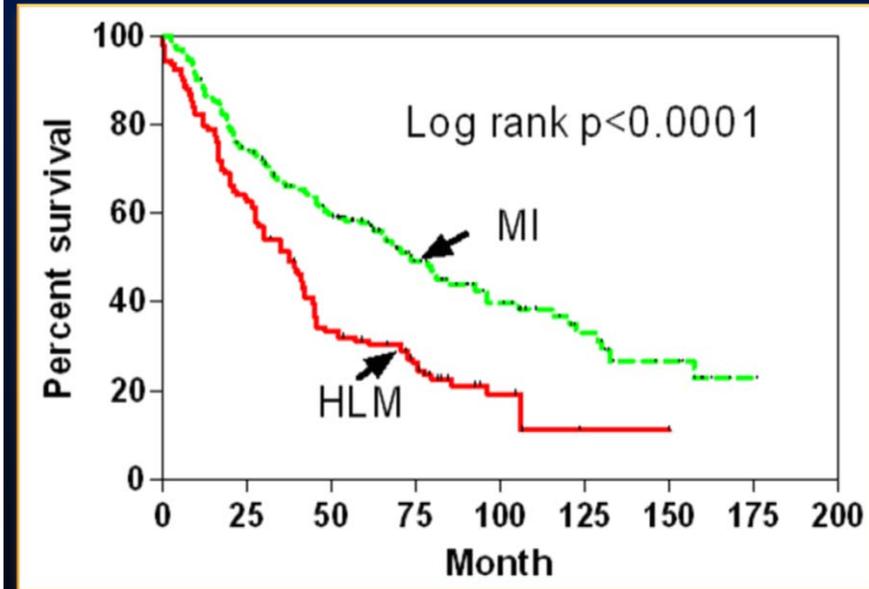


NOT AS PROMISING AS IT MAY SEEM

Stage



Differentiation



Site

Histology

SYSTEMATIC SIGNATURE SELECTION

Start with N=1 "signature"



Rank N-gene signatures

LS	A	B	C	D	E	F
	Median	Median	Median	Wilcoxon		mir
1	Severe	SCC				
2	12	18.1641	12.7257	0.125	hsa-miR-105-372157	
3	27	648.9136	73.46941	0.125	hsa-miR-133b-323117	
4	30	1.813078	2.436516	0.125	hsa-miR-136b-437313	
5	34	0.829636	2.225296	0.125	hsa-miR-141437137	
6	38	1.764045	0.645121	0.125	hsa-miR-165-3721326	
7	43	0.548559	1.091347	0.125	hsa-miR-149-437128	
8	46	0.207788	7.726999	0.125	hsa-miR-151-3721726	
9	57	0.698832	11.78414	0.125	hsa-miR-162-3726666	
10	58	0.533211	2.388887	0.125	hsa-miR-183-4373114	
11	64	1.409013	2.626402	0.125	hsa-miR-18a-4273188	
12	69	0.91747	1.903006	0.125	hsa-miR-193a-437310	
13	73	2.754513	8.241986	0.125	hsa-miR-196a-437319	
14	90	1.029742	2.423202	0.125	hsa-miR-19a-3720999	
15	81	1.485406	3.038114	0.125	hsa-miR-19b-4372988	
16	82	1.79739	3.829263	0.125	hsa-miR-200b-4373227	
17	83	0.918664	1.938643	0.125	hsa-miR-200a-437806	
18	84	1.061499	2.181327	0.125	hsa-miR-200b-438102	
19	85	1.094136	2.196757	0.125	hsa-miR-200c-4373209	
20	87	0.345613	1.058943	0.125	hsa-miR-203-372056	
21	30	0.607581	0.955189	0.125	hsa-miR-205-4372933	
22	32	1.175088	1.939536	0.125	hsa-miR-219-4372096	
23	35	0.902064	2.322238	0.125	hsa-miR-218-4372689	
24	132	0.091703	0.465643	0.125	hsa-miR-219-4372689	
25	936	3.305266	4.739796	0.125	hsa-miR-221a-372076	
26	132	1.64236	1.661957	0.125	hsa-miR-21-4373186	
27	137	1.672136	3.153811	0.125	hsa-miR-324-37313	
28	941	1.88253	2.617291	0.125	hsa-miR-3291-4372644	
29	158	1.670907	0.365115	0.125	hsa-miR-409-437317	
30	175	1.756801	2.422986	0.125	hsa-miR-425-4372022	

Select the best

Evaluate the signature

K-medians clustering

Survival diff test

N+1 gene signatures

Improvement?

Final signature

Yes

No

Prognostic gene signatures for non-small-cell lung cancer

Paul C. Boutros^{1,2}, Suzanne K. Lau^{3,4}, Melania Pintilie¹, Ni Liu⁵, Frances A. Shepherd^{1,6}, Sandy D. Der^{1,6}, Ming-Sound Tsao^{1,6}, Linda Z. Penn^{1,6}, and Igor Jurisica^{1,2,7}

Departments of ¹Medical Biophysics, ²Medicine, ³Laboratory Medicine and Pathology, and ⁴Computer Science, University of Toronto, Toronto, ON, Canada; ⁵MSG 1A1, ⁶Ontario Cancer Institute, University Health Network, Toronto, ON, Canada; ⁷MSG 2M9, and ⁸Division of Medical Oncology, Princess Margaret Hospital, Toronto, ON, Canada; ⁹MSG 2M9

Edited by Tak Wah Mak, University of Toronto, Toronto, ON, Canada, and approved December 23, 2008 (received for review September 21, 2008)

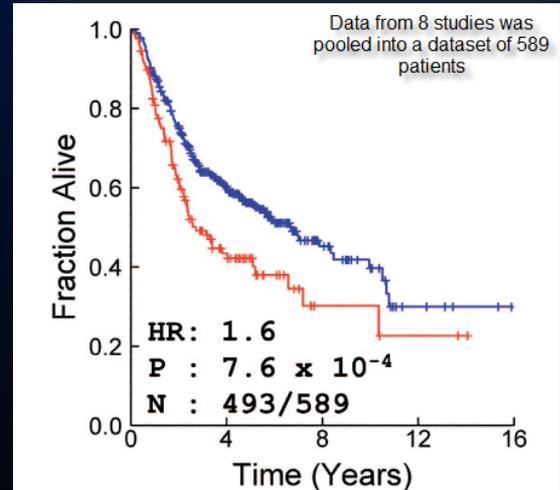
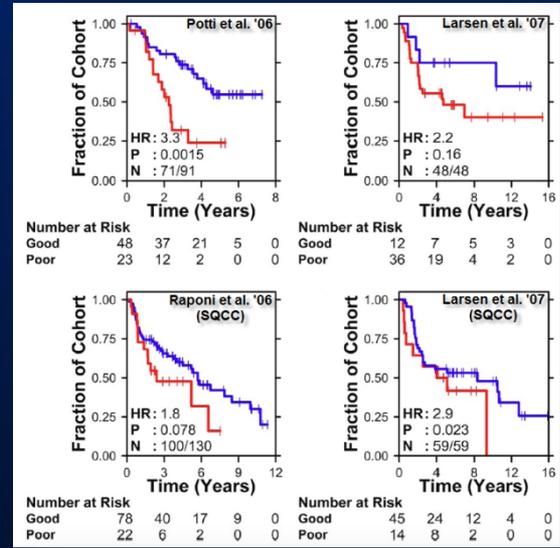


THERE IS MORE THAN ONE SIGNATURE

- ✘ Re-analysing Lau et al., *JCO* 2007 data with mSD
- + Validated 6 gene signature

STX1A HIF1A CCT3 MAFK RNF5 HLA_DPB1

- ✘ 113 genes in 4 test datasets



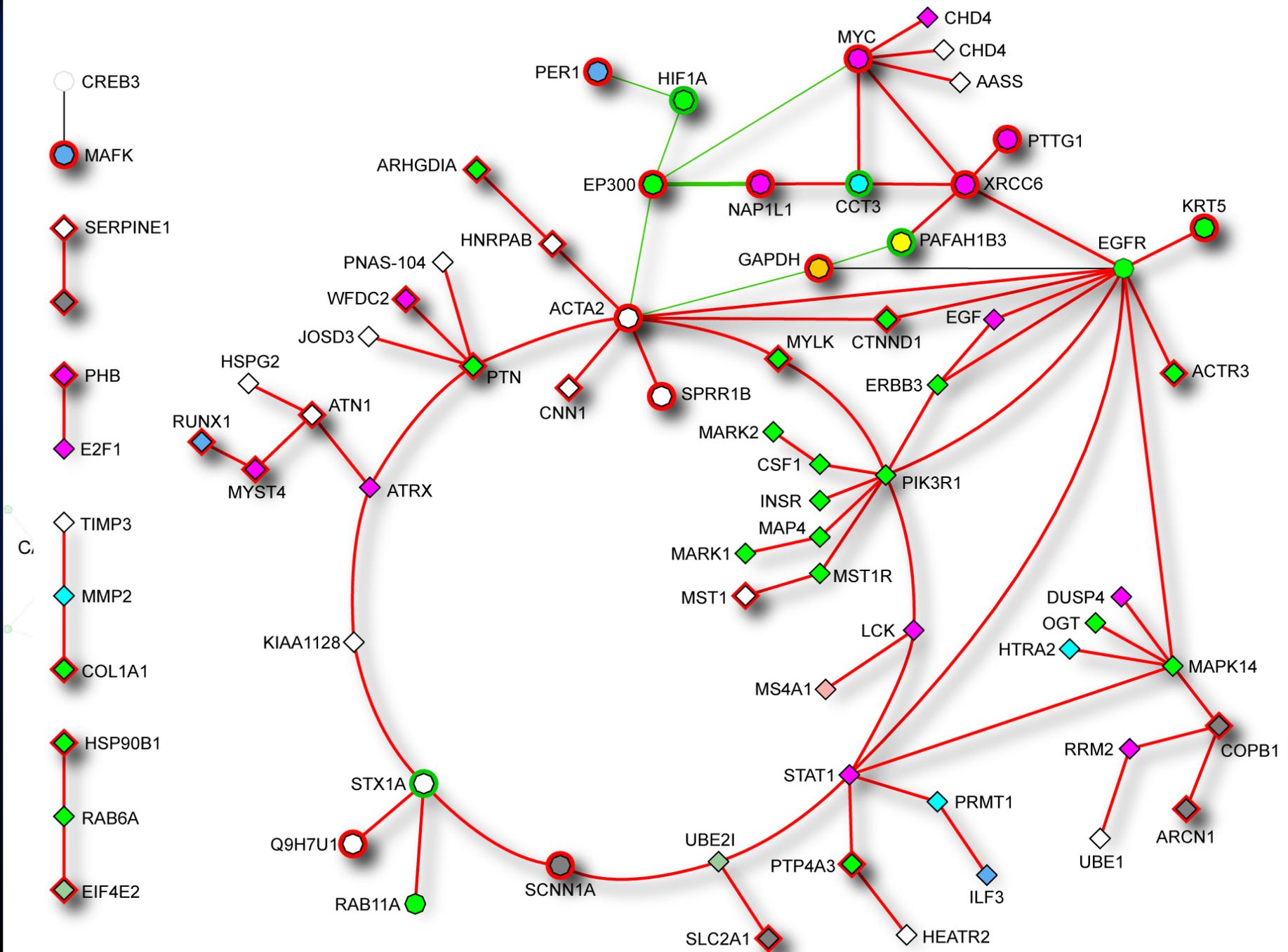
Time (Years)	0	4	8	12	16
Good	365	159	30	5	0
Poor	128	35	4	2	0

Prognostic gene signatures for non-small-cell lung cancer

Paul C. Boutros^{a,b,1}, Suzanne K. Lau^{a,b}, Melania Pintilie^a, Ni Liu^b, Frances A. Shepherd^{c,d}, Sandy D. Der^{b,e}, Ming-Sound Tsao^{b,h}, Linda Z. Penn^{a,b}, and Igor Jurisica^{a,b,h,1,2}

Departments of ^aMedical Biophysics, ^bMedicine, ^cLaboratory Medicine and Pathology, and ^dComputer Science, University of Toronto, Toronto, ON, Canada; ^eMSG 1A1; ^fOntario Cancer Institute; University Health Network, Toronto, ON, Canada; ^gMSG 2M9; and ^hDivision of Medical Oncology, Princess Margaret Hospital, Toronto, ON, Canada; ⁱMSG 2M9

Edited by Tak Wah Mak, University of Toronto, Toronto, ON, Canada, and approved December 23, 2008 (received for review September 21, 2008)



C.

INTEROLOGOUS INTERACTION DATABASE

Powered by Technology

Home Search I2D Download Statistics Visualization Publications People Links About I2D Contact Login

Welcome to I2D!

To facilitate experimentation and integrated computational analysis with model organism PPI networks, we have integrated known, experimental and predicted PPIs for five model organisms and human in the I2D database.

I2D is developed and maintained by **Juriska Lab** at Ontario Cancer Institute, PHN. I2D will continue to expand as new protein-protein interaction data becomes available.

Statistics

Source Interactions: 232,180
 Predicted Interactions: 204,222
 Total Interactions: 429,072

Database Access

The latest I2D version 1.72 is available for download in its entirety.

Download

I2D can also be queried online via a web interface.

Search

Visualization

NAVIGATOR is a powerful graphing application for the 2D and 3D visualization of biological networks.

When I2D is queried, it can output in several formats one of which is NAVIGATOR compatible file. This can be opened up in NAVIGATOR visualization and further analysis. I2D can also be queried within NAVIGATOR.

Source Interactions: 240,650
Predicted Interactions: 215,397
Total Interactions: 448,326

GenCards Homepage

Fix Edit View Favorites Tools Help

Google

Norton

GenCards

Search

GenCards is a comprehensive database of gene and protein information. It provides a central location for researchers to find and access a wide range of biological data, including gene names, protein structures, and functional annotations.

KEGG PATHWAY Database

Wiring diagrams of molecular interactions, reactions, and relations

Home About Content Documentation

Search for

Reactome - a cura

The data (reactome) for Homo sapiens

KEGG2 ATLAS PATHWAY BRITE GENES SSB LGAND DBET

Pathway Maps

KEGG PATHWAY is a collection of manually drawn pathway maps representing our knowledge on the molecular interaction and reaction networks for:

- Metabolism
- Carbohydrate Energy Lipid Nucleotide Amino acid Other amino acid
- Cycle PCNA Cell cycle Ubiquitin Secondary metabolite Neurotransmitters
- Genetic Information Processing
- Environmental Information Processing
- Cellular Processes
- Human Diseases

and also on the structure relationships (KEGG drug structure maps) in:

- Drug Development

page discussion view source history

Welcome to WikiPathways

Today's Featured Pathway

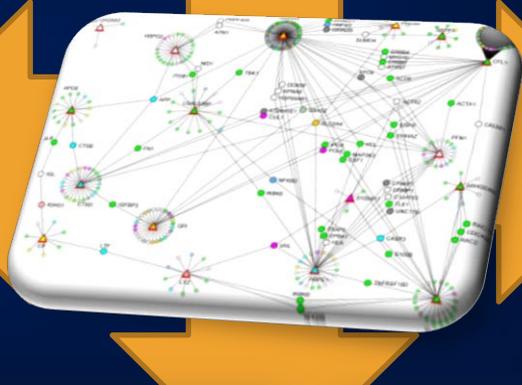
Here Diacylglycerol (Mus musculus)

Search

Browse

Contributing New Pathways

Create Suggest



miRNA DB

Hit Selection

Navigation

Assay

mirDIP: microRNA Data Integration Portal

Welcome to mirDIP

Source Filter

Prediction Features

Quality Filter

Database Occurrences

Interacting IDs

Toronto Centre for Pharmacogenomics

Centre for Modeling Human Disease

About CMHD

CMHD Services

Gene Sequencing

ENU Mutagenesis

CGI Map

Search

Home | CGI Map | Search

About Entrez

Text Version

Entrez PubMed

Overview

Help | FAQ

Tutorials

New/Noteworthy

E-Utilities

PubMed Services

Journals Database

MeSH Database

Single Citation

Matcher

Batch Citation Matcher

Clinical Queries

Special Queries

LinkOut

My NCBI

NCBI

PubMed

A service of the U.S. National Library of Medicine and the National Institutes of Health

www.pubmed.gov

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC

Search PubMed for

Limits Preview/Index History Clipboard Details

To get started with PubMed, enter one or more search terms.

Search terms may be [topics](#), [authors](#) or [journals](#).

Set up an automated PubMed update in fewer than five minutes.

1. Create a [My NCBI account](#).
2. Save your search.
3. Your PubMed updates can be e-mailed directly to you.

Read the [My NCBI Help](#) material to explore other options, such as automated updates of other databases, setting search filters, and highlighting search terms.

PubMed is a service of the U.S. National Library of Medicine that includes over 17 million citations from MEDLINE and other life science journals for biomedical articles back to the 1950s. PubMed includes links to full text articles and other related resources.

PSI nature

StructuralGenomicsKnowledgebase

Welcome to the Structural Genomics Knowledgebase

Keep informed about advances in structural biology, structural genomics, and models related to biological technologies.

search

Explore proteins

structural genomics update

Research advances, news and events in structural genomics

featured molecule

IsdA and DNA replication

The challenge for bacteria replication only once for each of the cell. PSI researchers obtained the first atomic structure of the IsdA protein that bacteria used to challenge.

WORLDWIDE PDB

PROTEIN DATA BANK

Access the PDB FTP

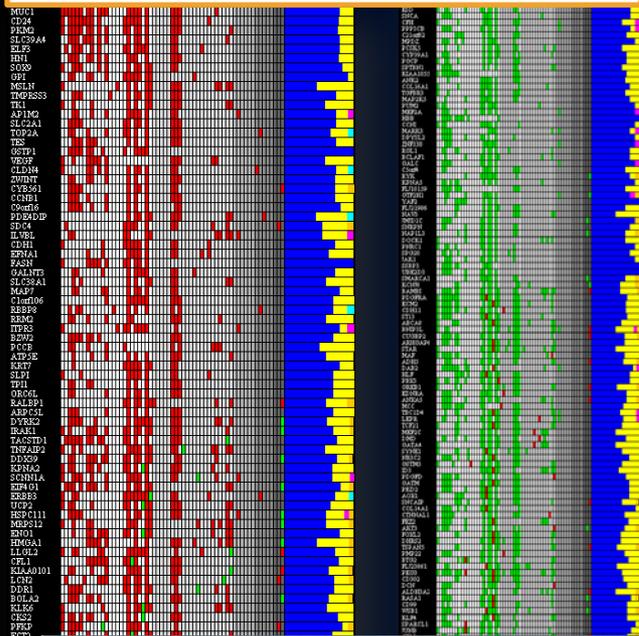
Time-stamped Copies of PDB Archive Available via FTP

Announcement: Data Processing Versioning Procedures

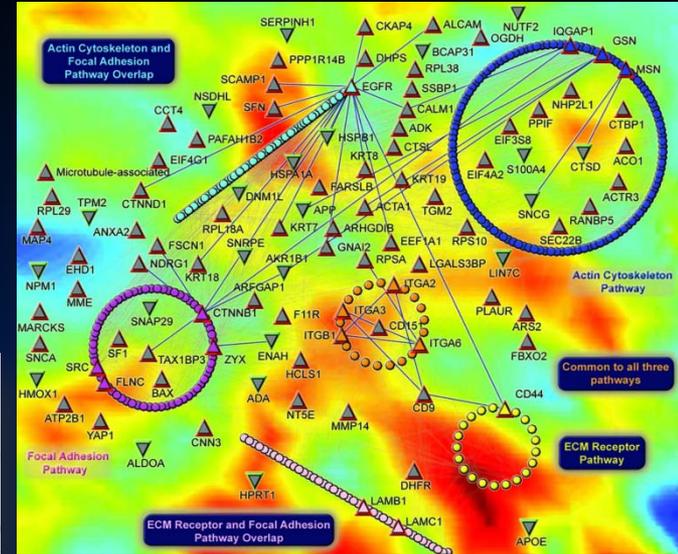
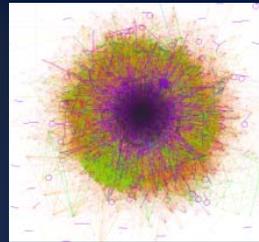
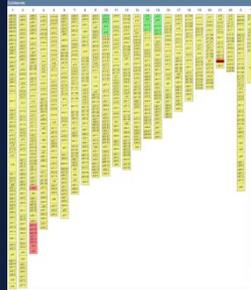
Date in the PDB archive currently follows either PDB File Format Version 3.0 or 3.1. This is indicated in the file name. Version 3.0 is the format used for files released as a result of the Nomenclature Project.

Begining March 4, 2008, 3.0.0 will be indicated in the file name with the name "V3.0" when a version 3.0 file is released as Version 3.1 - HEAD BORG.

Blue Cancer vs. Normal
 Yellow Cancer vs. cancer
 Magenta Prognosis
 Cyan Androgen response
 Orange PSA related
 Pink Drug/treatment effect



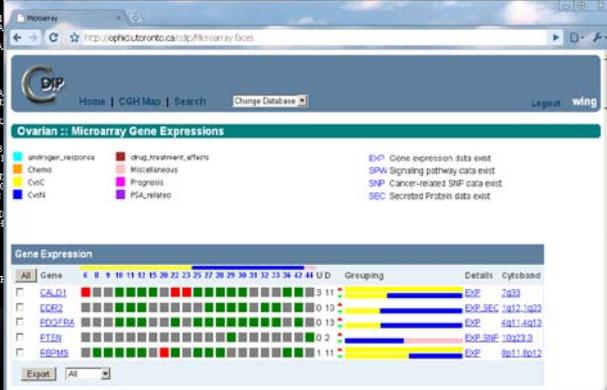
- **12,975 unique genes**
- 60% or more in one direction & at least ≥ 6 studies in one direction
- **UP** **198**
- **DOWN** **408**



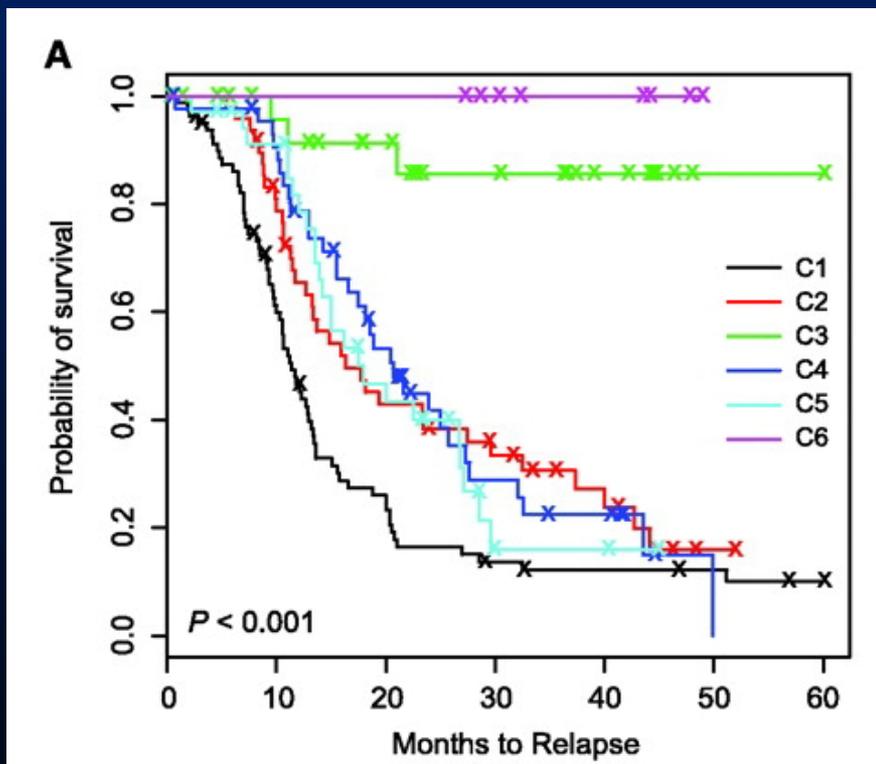
- Gene from Max number of studies : NBL1 (8 \uparrow & 11 \downarrow)
- Gene with Max Ups : Myc (14 \uparrow & 2 \downarrow)
- Gene with 100% Up (max #studies) : NEURL (9 \uparrow)
- Gene with Max Downs : MYLK (14 \downarrow & 4 \uparrow)
- Gene with 100% Down (max #studies) : SERPINB1 (11 \downarrow)

CGH data:

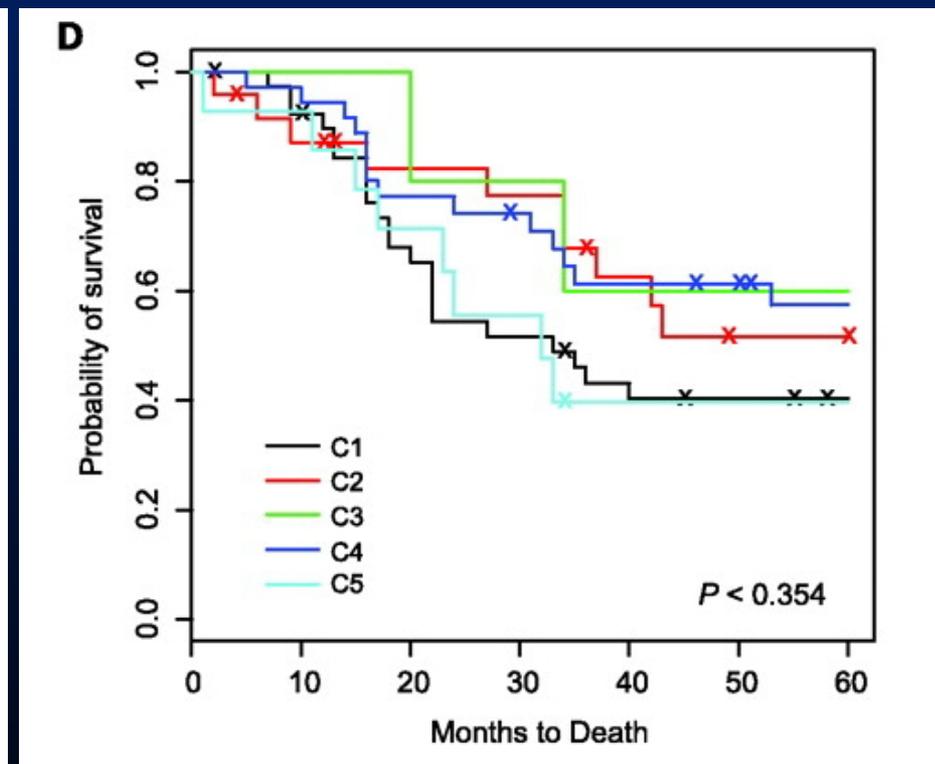
- Amplification : 3q (3q27.3, 3q27.2, 3q27.1, 3q26.33, 3q26.32, 3q26.31, 3q29, 3q28, 3q23)
- Deletion : 10p, 14p, 15q/p



OVARIAN CANCER



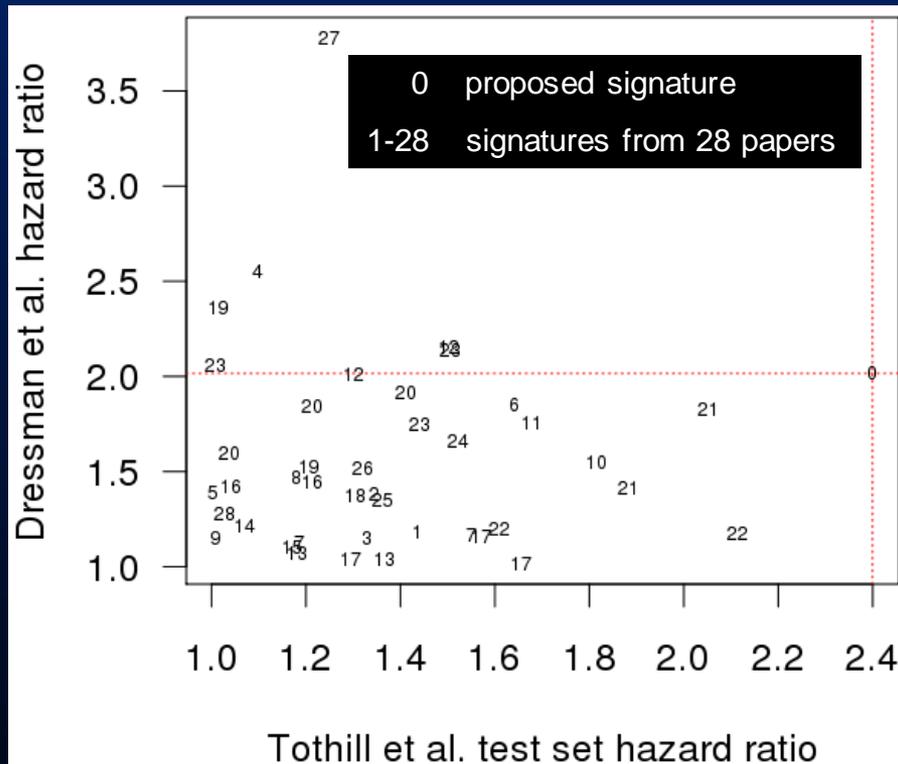
Training



Independent Validation



EOC PROGNOSTIC SIGNATURES



plot ID	PMID
1	11779836
2	12067990
3	12629520
4	14522886
5	14556934
6	14871800
7	14996724
8	15505275
9	15661559
10	15788660
11	16204010
12	16287073
13	16404635
14	16421595
15	16557592
16	16964388
17	16969345
18	17072343
19	17177833
20	17254359
21	17346539
22	18398031
23	18593951
24	18687136
25	19112514
26	19192944
27	19318476
28	19440550

k-means clustering (k=2) to compare the hazard ratio of the proposed risk score to signatures from 28 papers.

Publication #27 used some of the same samples as the Dressman et al. paper, so its performance is not an independent validation.

CANCER INTERACTOME

- ✘ Thousands of prognostic/predictive signatures need to be integrated, prioritized & understood (Lau et al., *JCO*, 2007; Boutros et al., *PNAS*, 2009)
- ✘ ~40% of cancer-related proteins lack interactions (CDIP)
- ✘ HTP methods detect individually 25-35% of interactions and combined 59% of interactions (Braun et al., *Nat Methods*, 2009)
- ✘ Data mining can essentially eliminate this gap



SPARSE INTERACTOME

- ✗ I2D
 - + 5,665 SwissProt proteins with 0 PPIs
 - + 3,563 proteins with only 1 known PPI
 - + 1,141 proteins only predicted PPIs
- ✗ UniProt (SwissProt/Trembl)
 - + 74,946 proteins with primary IDs (Dec, 2008)
 - + 36,899 proteins with non-redundant sequences
 - + 23,013 protein fragments
- ✗ Determine interaction scores for all pairs of proteins in UniProt: ~2.6 billion protein pairs



PREDICTING PPIS

correlated gene expression across human tissues

orthologs in the same set of species

domains that rarely occur in non-interacting protein pairs

similar GO annotation

Protein Pair	Gene Co-expression	Ortholog Co-occurrence	Domains (log odds)	GO (Semantic Similarity)			Interacts
				Local	Function	Process	
P08133 , P55327	-0.041	-0.285	3.906	0.166	0.041	1.000	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
P17612 , P98175							0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

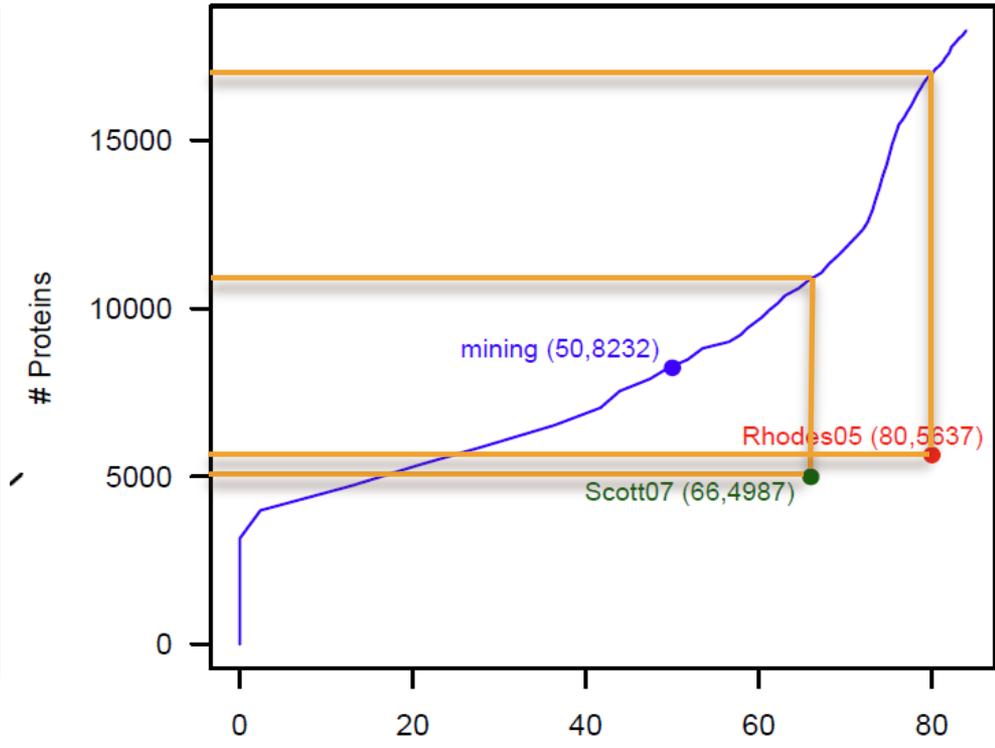
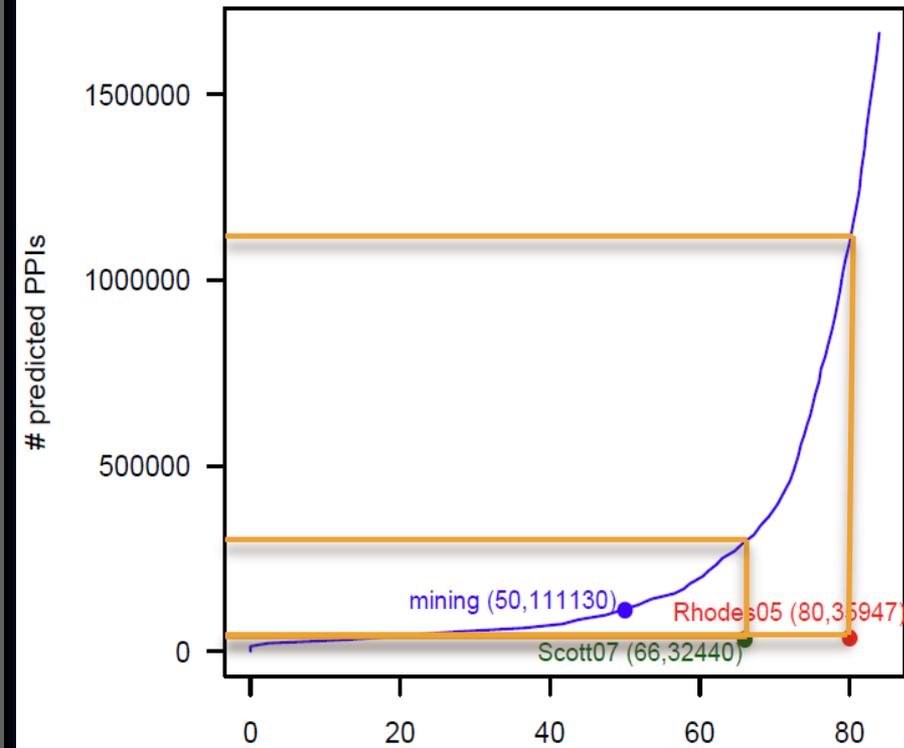
30,363 human interactions

303,630 random human protein pairs

- Association mining algorithm identifies sets of features that occur predominantly in PPIs.
- Decision tree algorithm is then used to predict PPIs.



COMPUTATIONAL COMPARISON



False positive rate (%)

False positive rate (%)

0.0

0.2

0.4

0.6

0.8

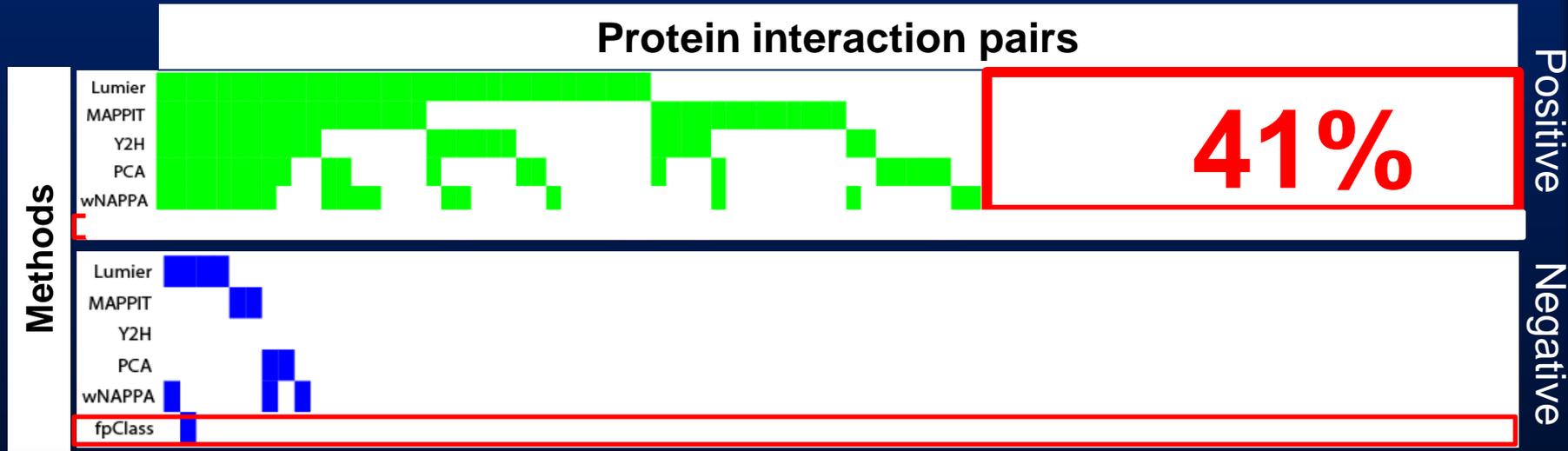
1.0

False positive rate



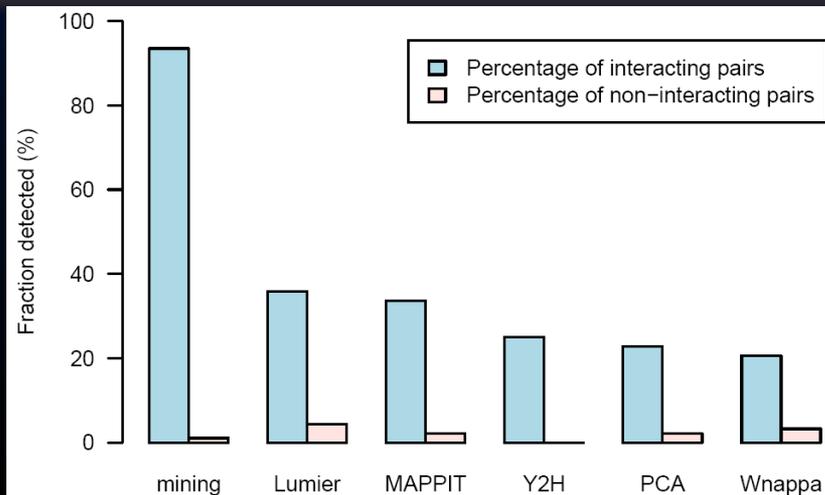
DATA MINING & HTP VALIDATION

Braun et al., Nat Methods, 2009

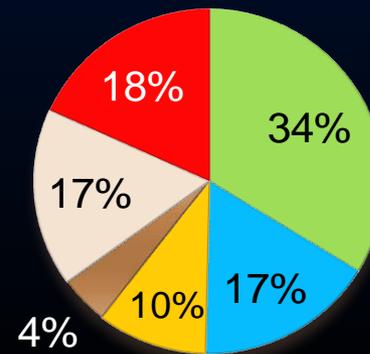


Individual HTP: 25-35% interactions

All HTP combined: 59% interactions

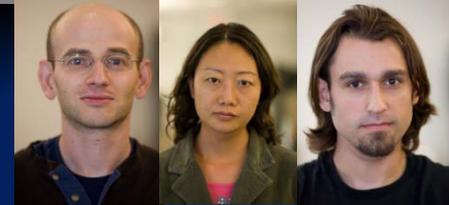


Data Mining

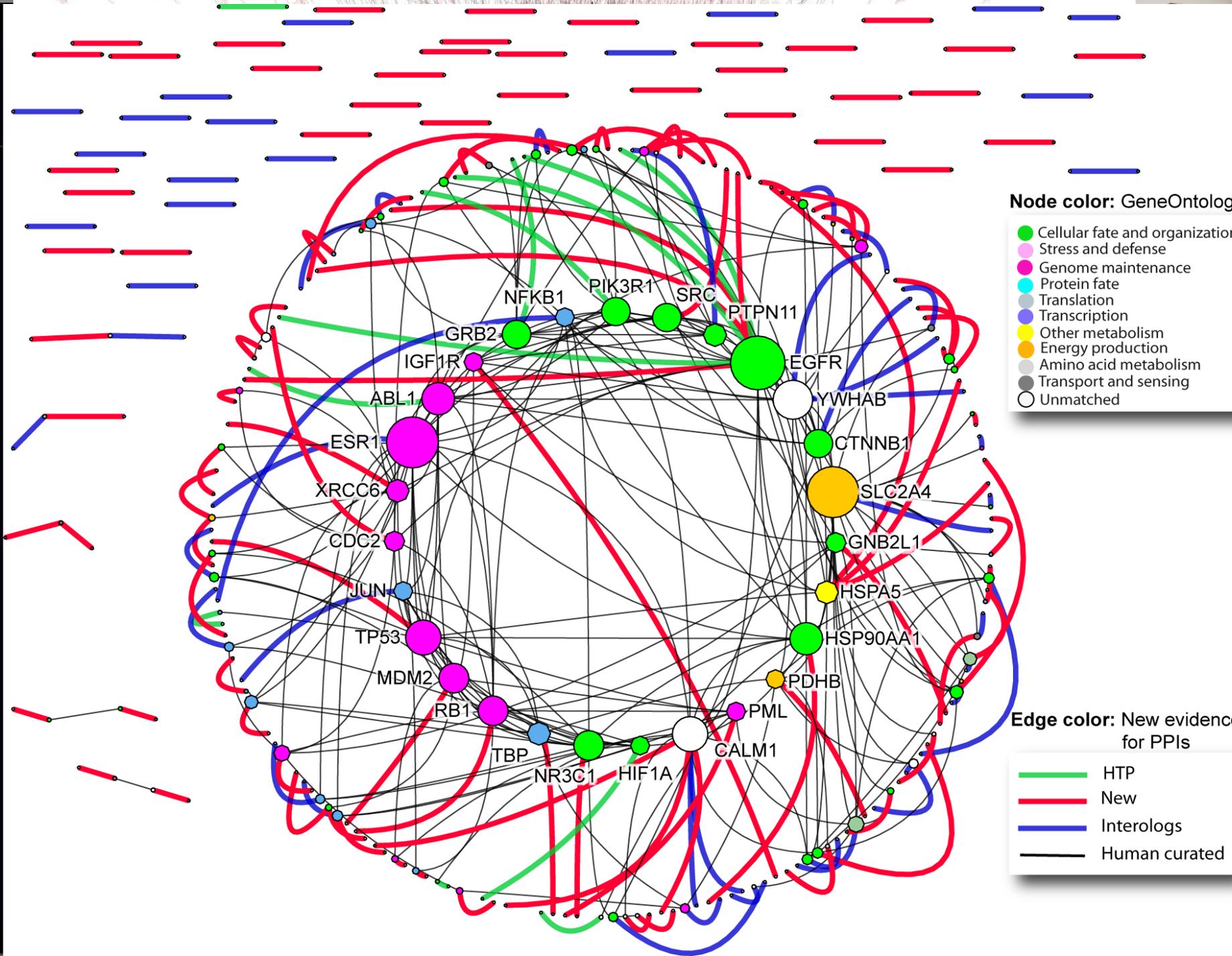


- Lumier
- MS
- PA
- Y2H
- coIP
- TM

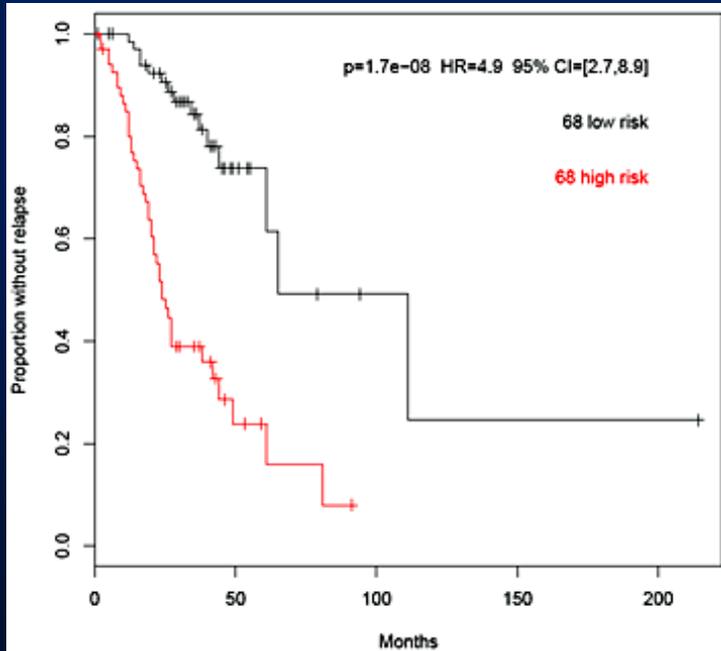
TEXT MINING & CURATION



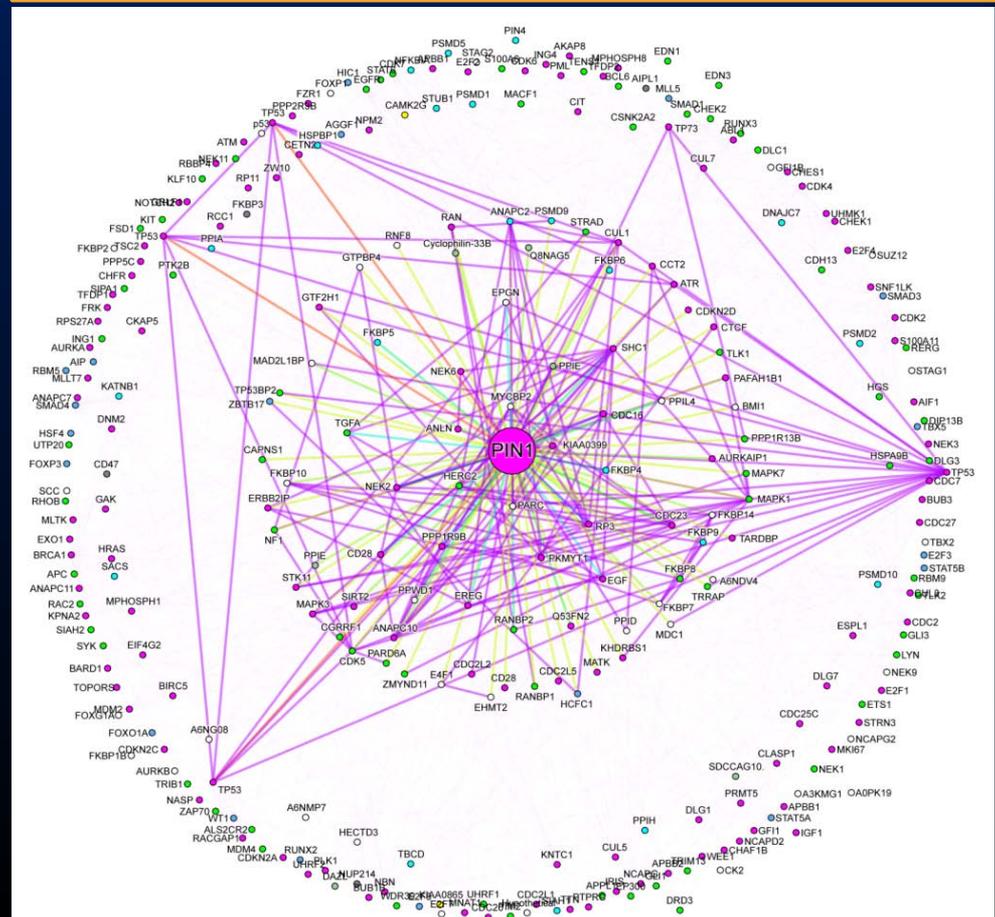
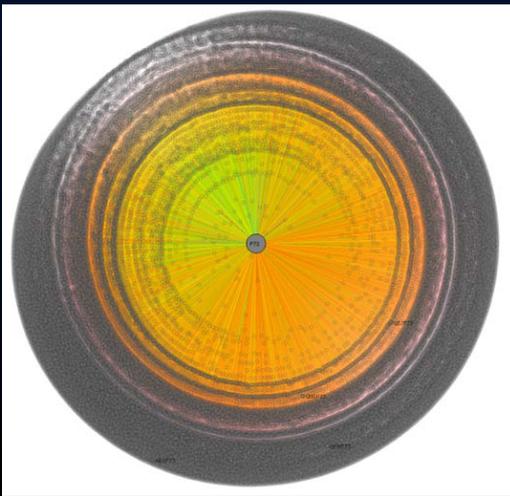
- ✗ 176,075 known and predicted PPIs in I2D
 - + 113,230 had no evidence and 62,845 had some evidence from text mining
 - + 9,398 abstracts had SVM score >0.5 , of which 6,290 manuscripts published prior to 2003 and 8,847 prior to 2006
 - + these 9,398 abstracts cover 13,800 unique PPIs
- ✗ Distribution across journals covered by IMEx:
 - + Cancer Cell: 1; Cell: 81; EMBO J: 213; EMBO Rep: 21; FEBS Let: 157; Nat Struct Mol Biol: 5; PNAS: 349; Biochem Biophys Res Commun: 308.
- ✗ Journal “PPI value”:
 - + JBC: 2,093; Mol Cell Bio: 480; Oncogene: 302; J Immunol: 239; Blood: 152; Nature: 116; Cancer Res: 98.



VALIDATING 6,631 PREDICTED PPIS



	SVM	Co-occurrence
OVARIAN	56	631
PROSTATE	33	799
LUNG	3	37

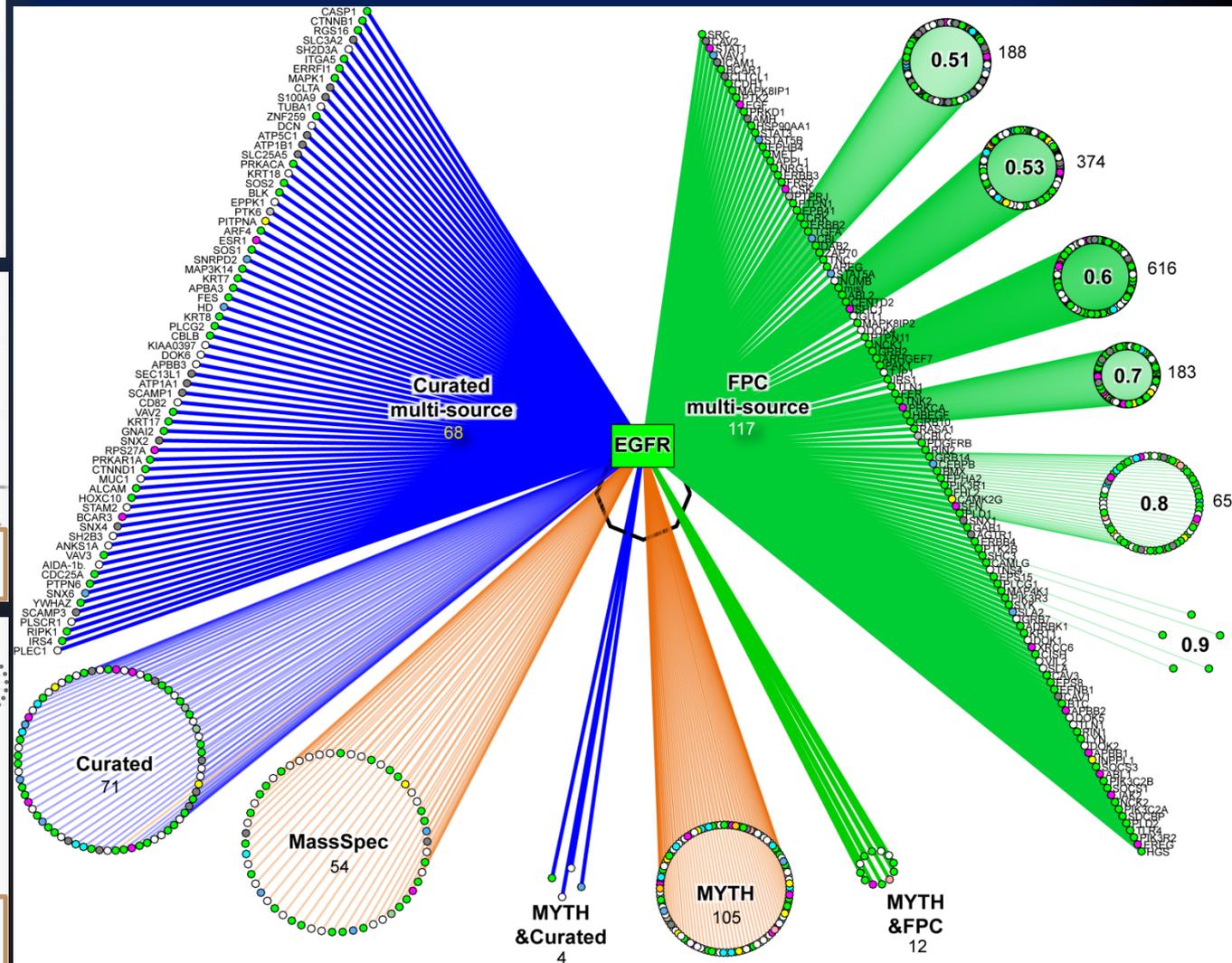
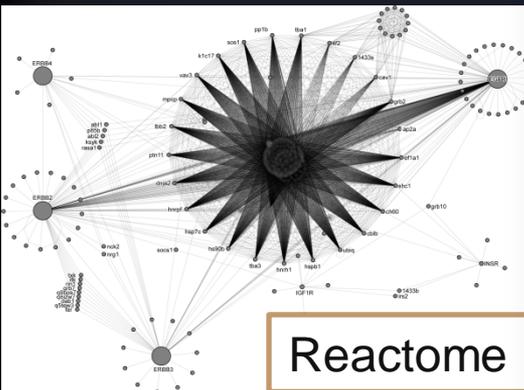
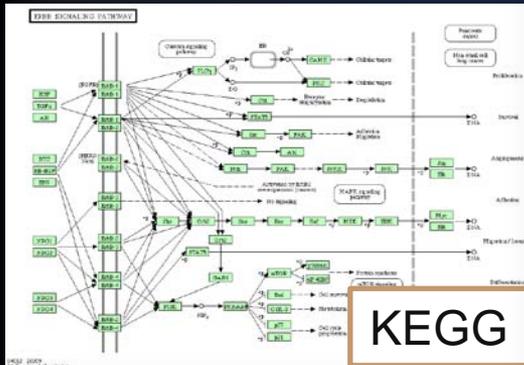
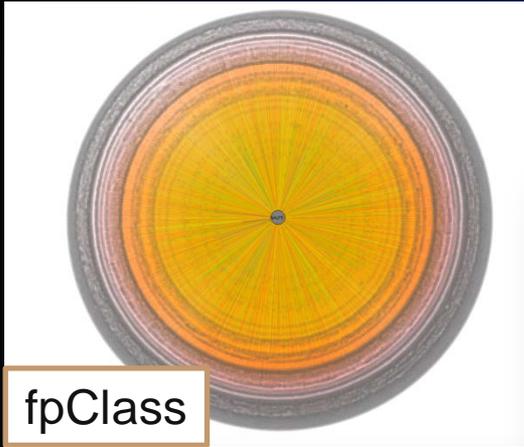


Regulation of Epidermal Growth Factor Receptor Trafficking by Lysine Deacetylase HDAC6

Yonathan Lissanu Deribe,¹ Philipp Wild,¹ Akhila Chandrashaker,² Jasna Curak,³ Mirko H. H. Schmidt,^{1,4} Yannis Kalaidzidis,^{2,5} Natasa Milutinovic,³ Irina Kratchmarova,⁶ Lukas Buerkle,^{3,7} Michael J. Fetchko,³ Philipp Schmidt,¹ Saranya Kittanakom,³ Kevin R. Brown,⁸ Igor Jurisica,^{8,9,10} Blagoy Blagoev,⁶ Marino Zerial,² Igor Stagljar,^{3*} Ivan Dikic^{1,11,12*}

(Published 22 December 2009; Volume 2 Issue 102 ra84)

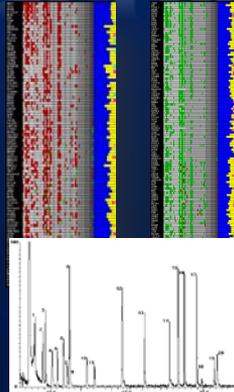
EGFR



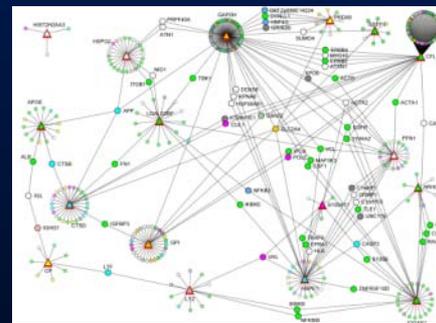
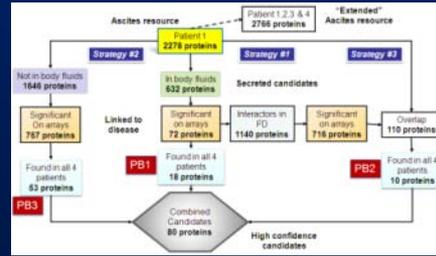


DISCOVERY PIPELINE

Data
+
Tools & Resources



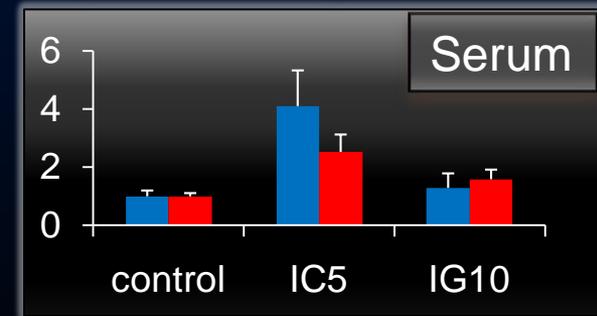
+



Analysis

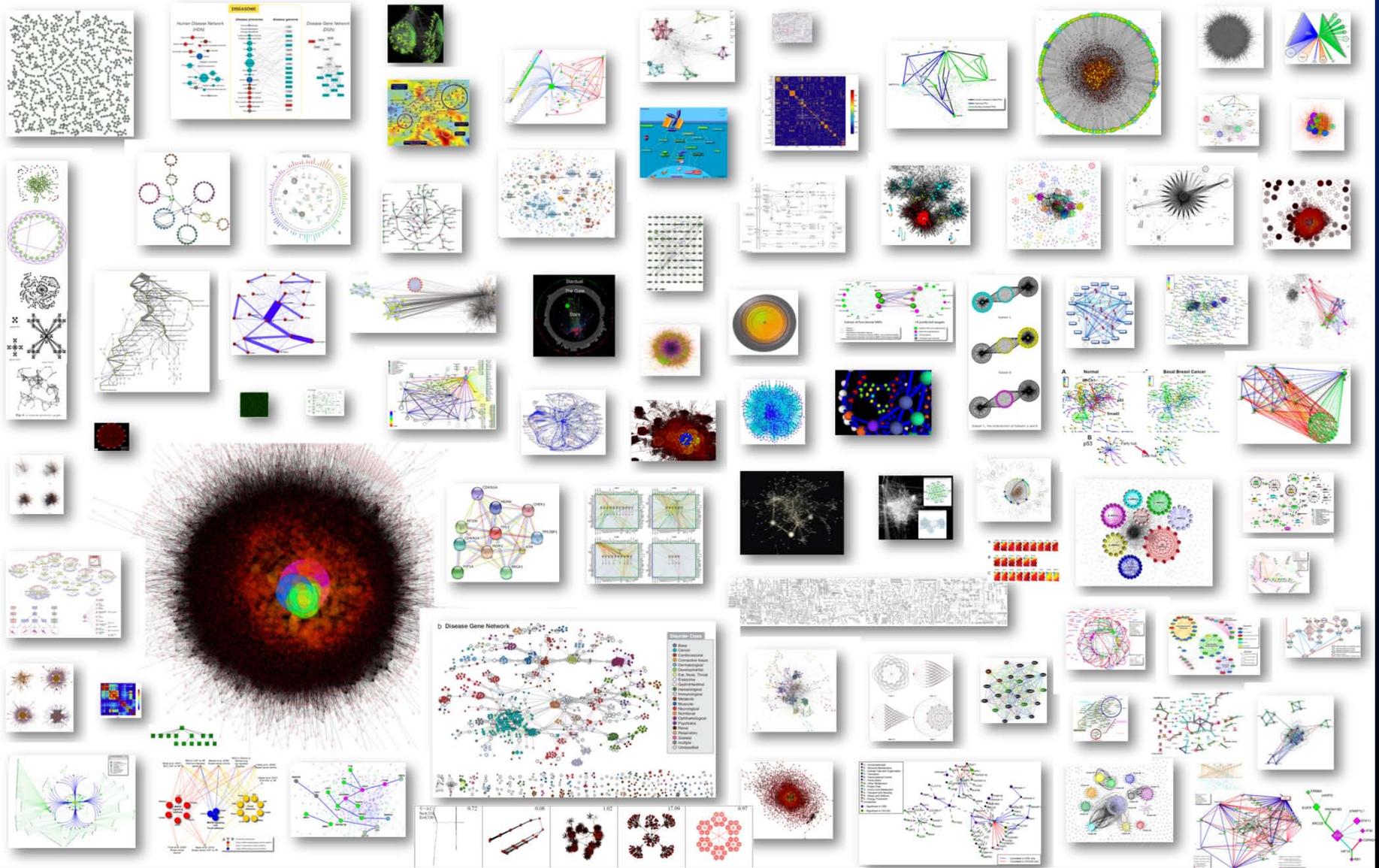
=

Prediction & Validation & Modeling



Journal of proteome research
research articles
A Proteome Resource of Ovarian Cancer Ascites: Integrated Proteomic and Bioinformatic Analyses To Identify Putative Biomarkers
Limor Gortzak-Uzan,^{1,2,3,4} Alex Ignatenko,^{1,5} Andreas I. Evangelou,^{1,5} Mahima Agochiya,¹ Kevin A. Brown,¹ Peter St-Onge,³ Inga Kireeva,¹ Gerold Schmitt-Ulms,⁶ Theodore J. Brown,^{1,6} Joan Murphy,^{1,5} Barry Rosen,^{1,5} Patricia Shaw,⁶ Igor Jurisica,^{6,7,8,9} and Thomas Kislinger^{1,10}

Size Does Matter





D. Strumpf, W. Xie, D. Otasek, A. Muhammad
 A. Teisanu, **R. Lu**
L. Waldron, Y. Niu, M. Kotlyar, C. A. Cumbaa
 K. Fortney, R. Yan, S. Rahmati, E. Shirdel, D. Rosu, S. Wong
 R. Ramnarine, T. R. Ramnarine, C. To,
 H. Li, L. Qi

M. S. Tsao, F. Shepherd, L. Penn, M. Pintilie,
L. Penn, R. Bristow, N. Fleshner
I. Stagljär, I. Dikic
A. Oza, A. Jurisicova, T. Kislinger, T. Colgan
G. Mills, ...

M. McGuffin, B. Devani, I. Van Toch, C. Frantz, A. Barta, F. Breard, K. Brown
 M. Soloviev, S. Grant, J. Jiang, P. Boutros

<http://www.cs.utoronto.ca/~juris>

<http://www.iscb.org/ismb2010-program/595-ismb2010-special-session-details#session5>

<http://www.iscb.org/ismb2010-program/ismb2010-tutorials#PM2>

<http://iscbnews.blogspot.com/2010/05/iscb-member-feedback-sought-on-draft.html>



ophid.utoronto.ca/i2d

ophid.utoronto.ca/genecards

ophid.utoronto.ca/navigator

ophid.utoronto.ca/cdip

ophid.utoronto.ca/mirdip

<http://www.conquercancer.ca>

<http://www.worldcommunitygrid.org>

