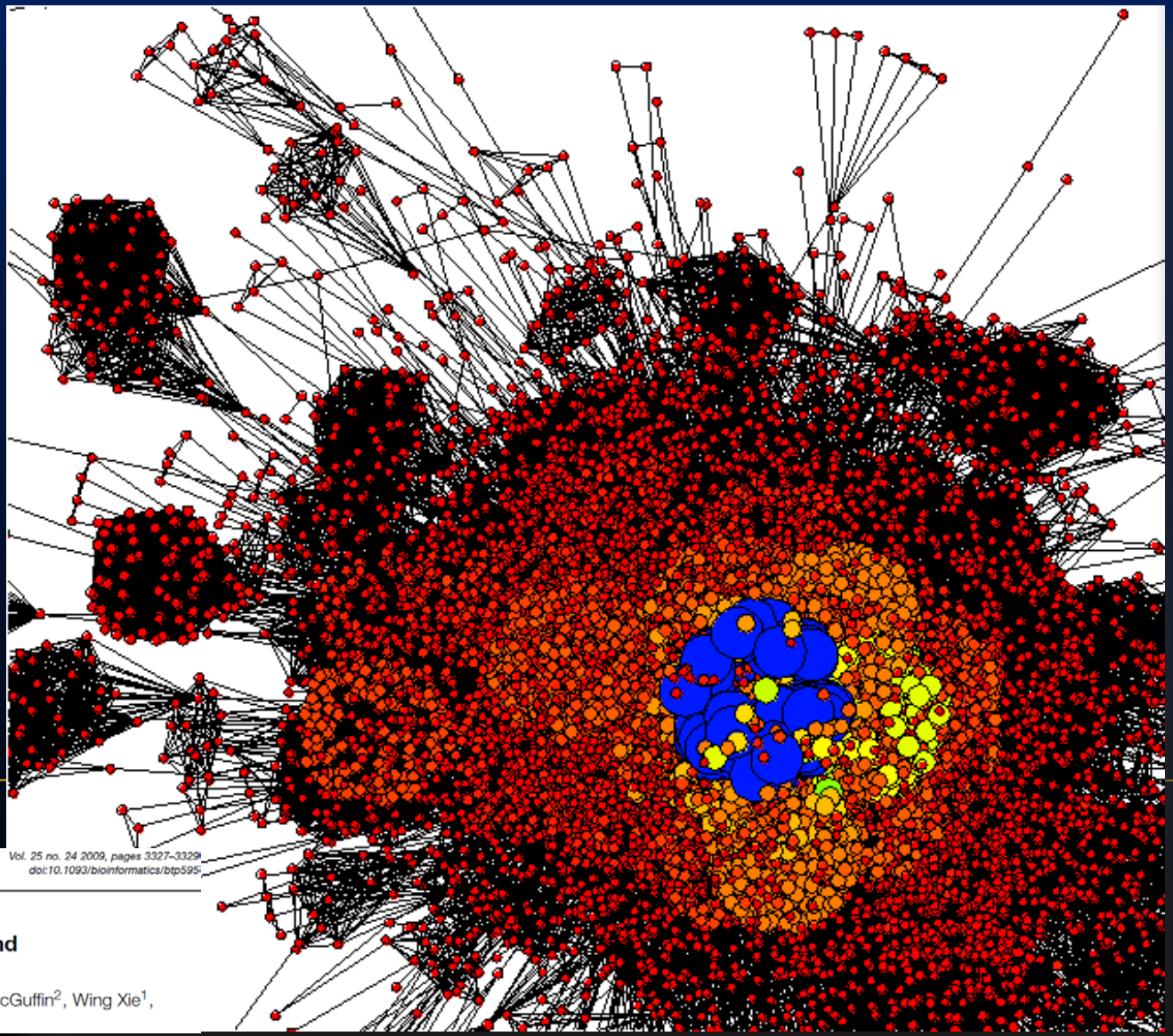
SYSTEMATIC IDENTIFICATION AND CHARACTERIZATION OF CANCER BIOMARKERS

Igor Jurisica

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

University Health Network
Ontario Cancer Institute

175 University of Toronto
VELUT ARBOR ÆVO



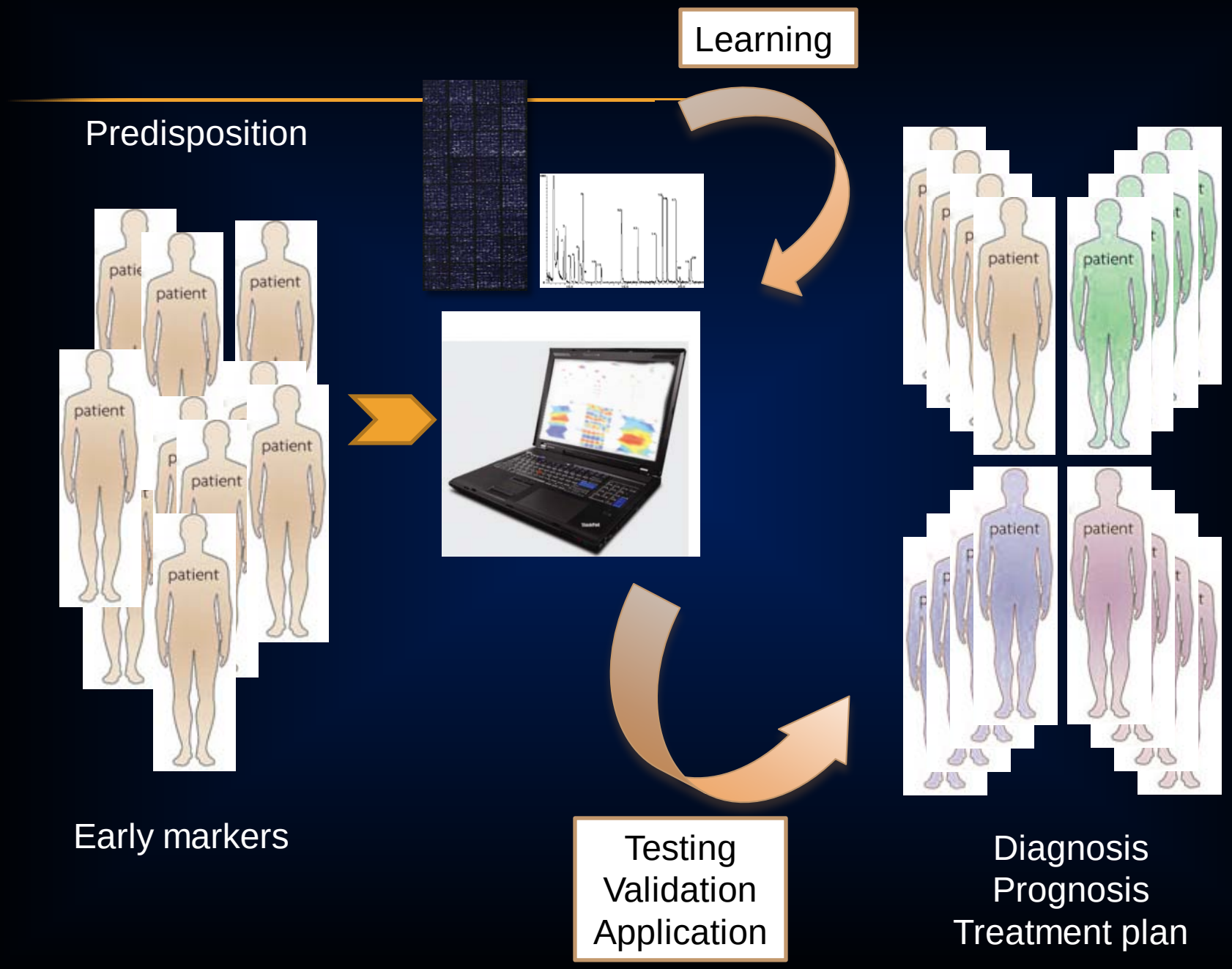
BIOINFORMATICS APPLICATIONS NOTE

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

Systems biology

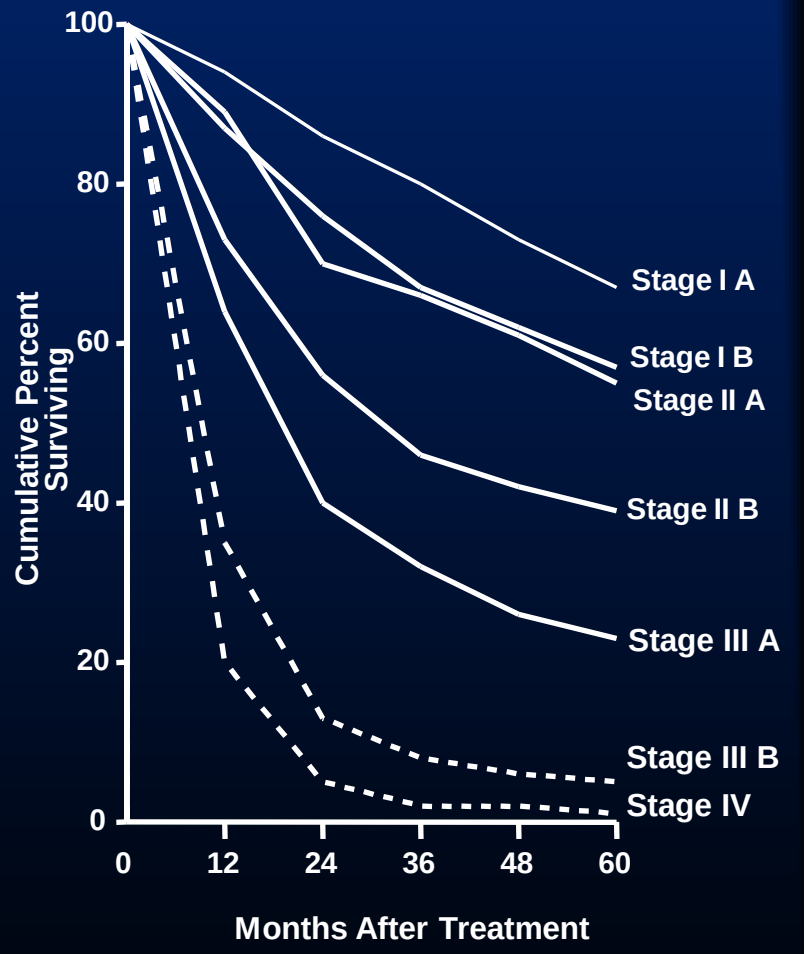
NAViGaTOR: Network Analysis, Visualization and Graphing Toronto

Kevin R. Brown[1], David Otasek[1], Muhammad Ali[1], Michael J. McGuffin[2], Wing Xie[1], Baiju Devani[1], Ian Lawson van Toch[1] and Igor Jurisica[1,3,*]


Learning
Predisposition
patient
Early markers
Testing
Validation
Application
Diagnosis
Prognosis
Treatment plan

STRATEGY TO REDUCE MORTALITY

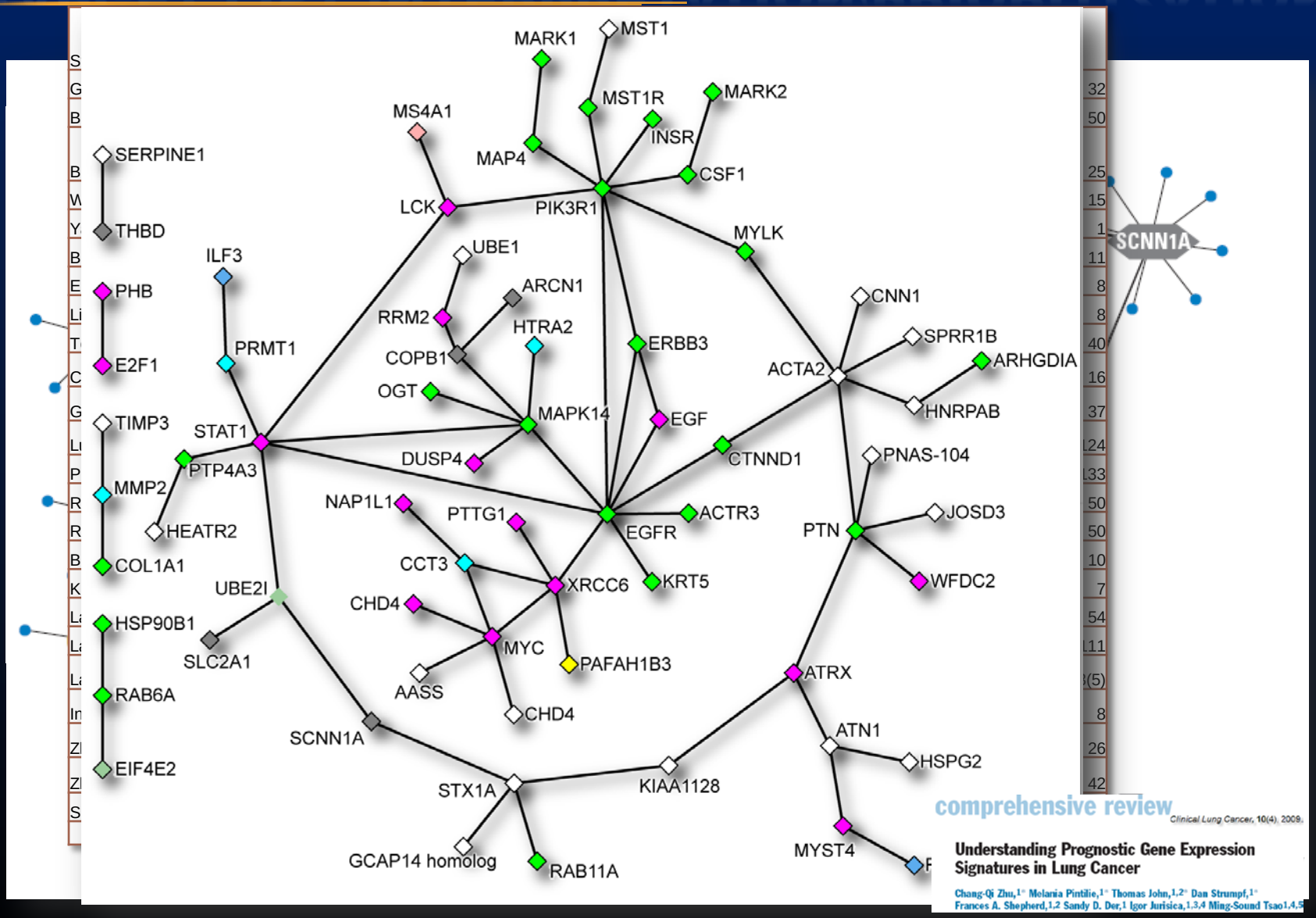
- Increase number of cases diagnosed at earlier stage
  - Biomarkers for early detection
- Individualized treatment
  - Biomarkers for treatment selection & monitoring
- Improved treatment
  - Understanding disease mechanism
  - New medicines



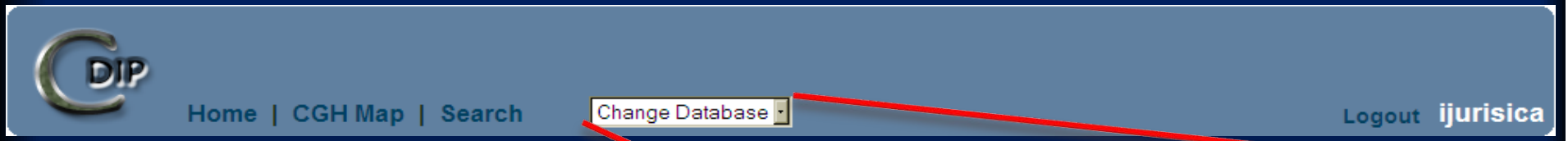
CHALLENGES

- 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 20*08
  - 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

TARGET LIST INTERPRETATION/PRIORITIZATION



ADDRESSING THE GAP



	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

Annotate

Cancer signatures

Annotate

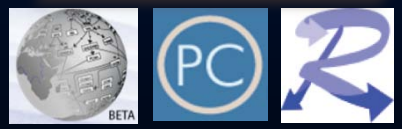
Protein Interactions & pathways

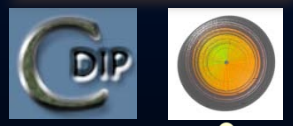
Validate

Focus

Prioritize

Curation & Annotation | Tools | Prioritization | **Validation** | Applications

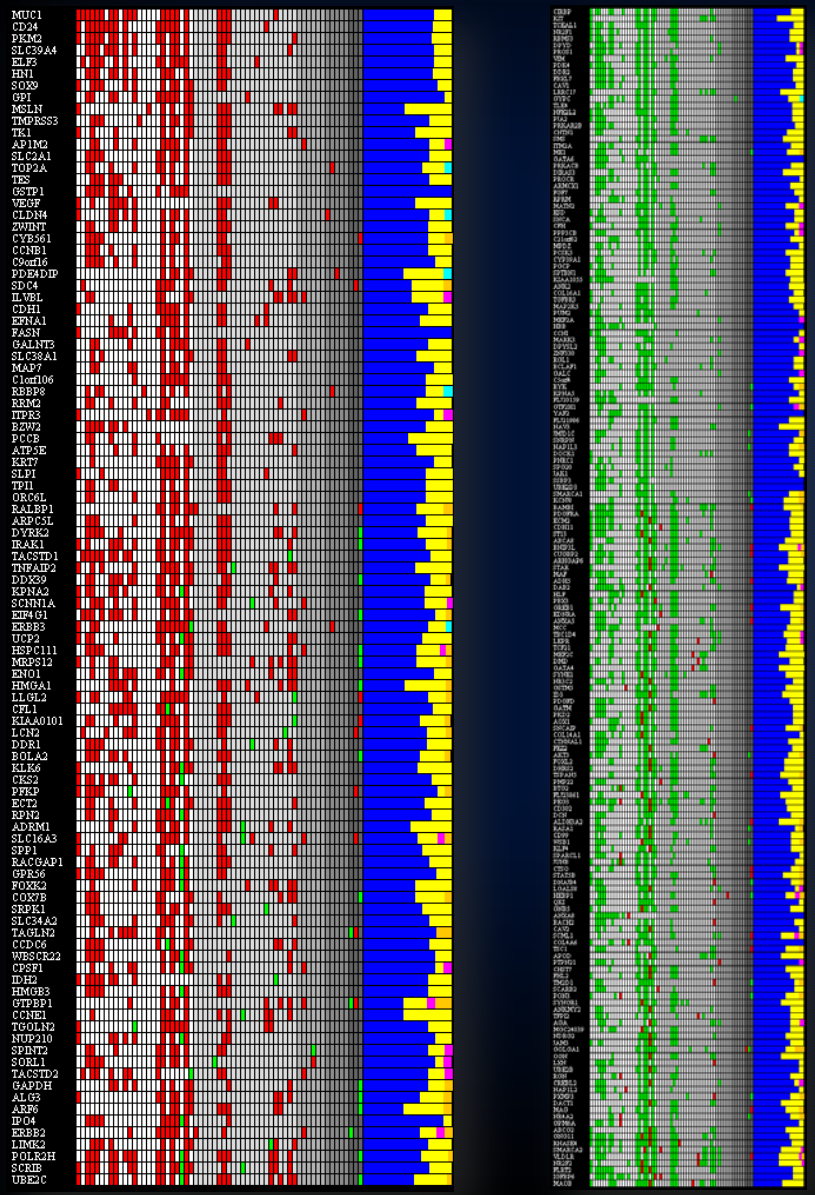




Lung, ovarian, breast
prostate, head&neck

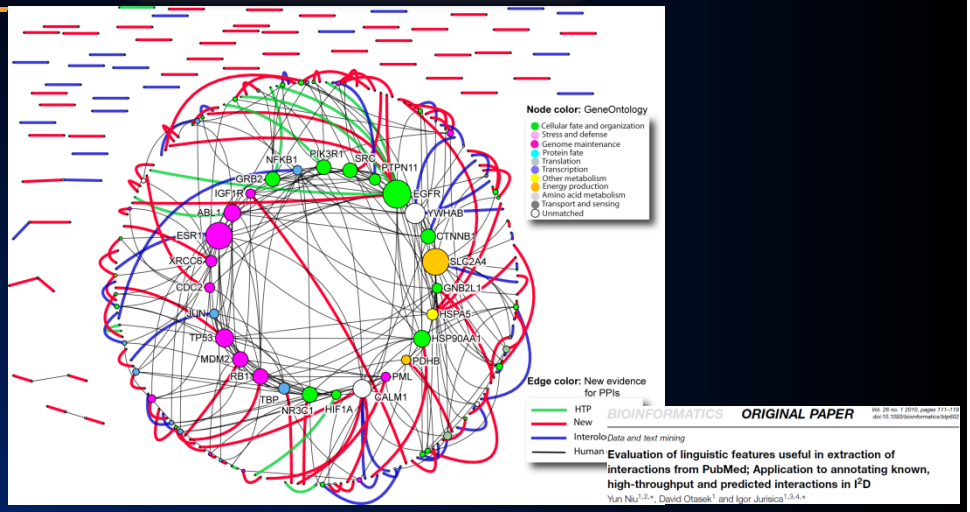
Signatures

ophid.utoronto.ca/cdip



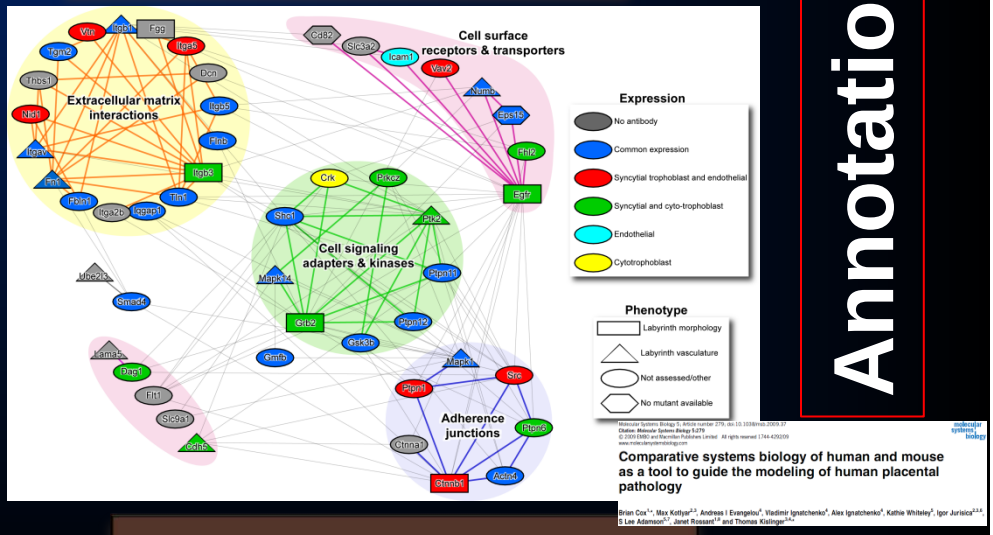
Interactions

ophid.utoronto.ca/i2d



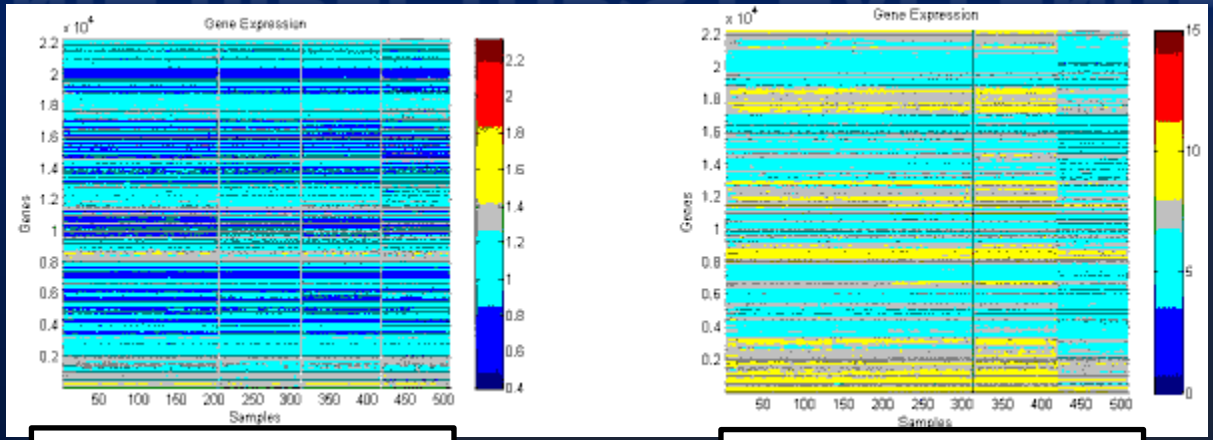
ophid.utoronto.ca/navigator

Annotation



ophid.utoronto.ca/genecards

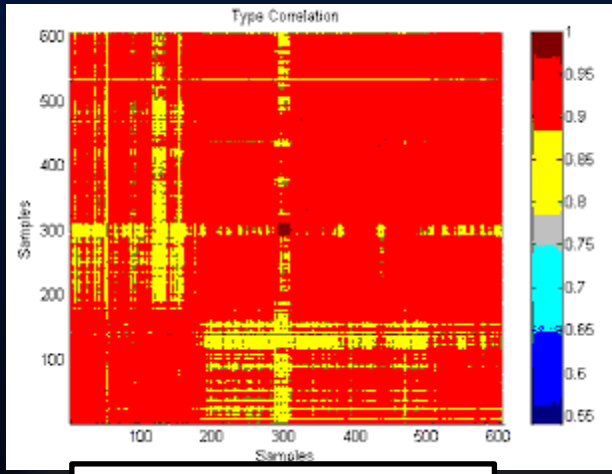
NIH DIRECTOR'S CHALLENGE



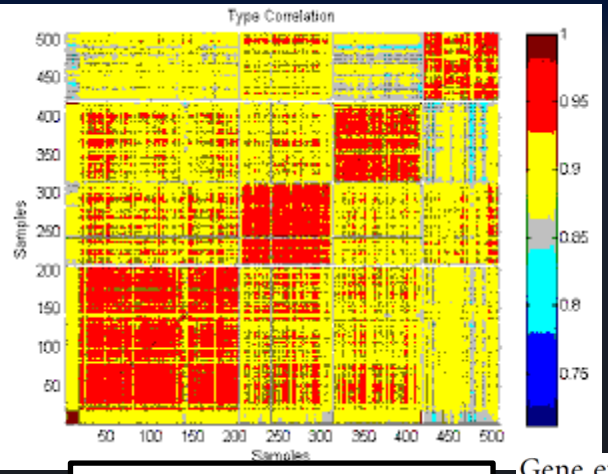
MICH MSKC MOFF HARV

MICH MSKC MOFF HARV

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



HARV MOFF MICH MSKC



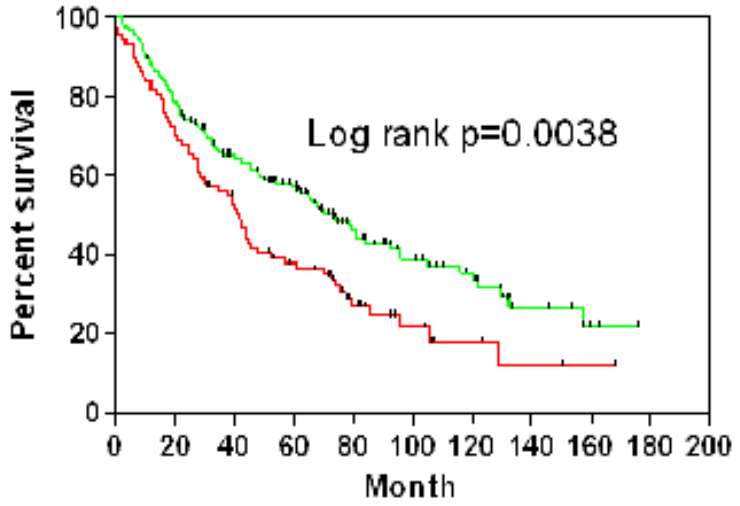
MICH MSKC MOFF HARV

nature medicine

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 Giordano[8], David E Misek[3,9], Andrew C Chang[3,9], Chang Qi Zhu[6], Daniel Strumpf[6], Samir Hanash[3], Frances A Shepherd[6], Keyue Ding[10], Lesley Seymour[10], Katsuhiko Naoki[11], Nathan Pennell[11], Barbara Weir[11], Roel Verhaak[11], Christine Ladd-Acosta[12], Todd Golub[12], Michael Gruidl[5], Anupama Sharma[5], Janos Szoke[7], Maureen Zakowski[7], Valerie Rusch[7], Mark Kris[7], Agnes Viale[7], Noriko Motoi[7], William Travis[7], Barbara Conley[13], 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[3,9,17]

BTSVQ – ALL PATIENTS



Factor	p	HR(95%CI)
Site	0.058	1.46(0.99-2.16)
Stage	<0.001	2.07(1.7-2.52)
Histology	0.21	1.08(0.96-1.21)
Age	0.007	1.02(1.006-1.04)

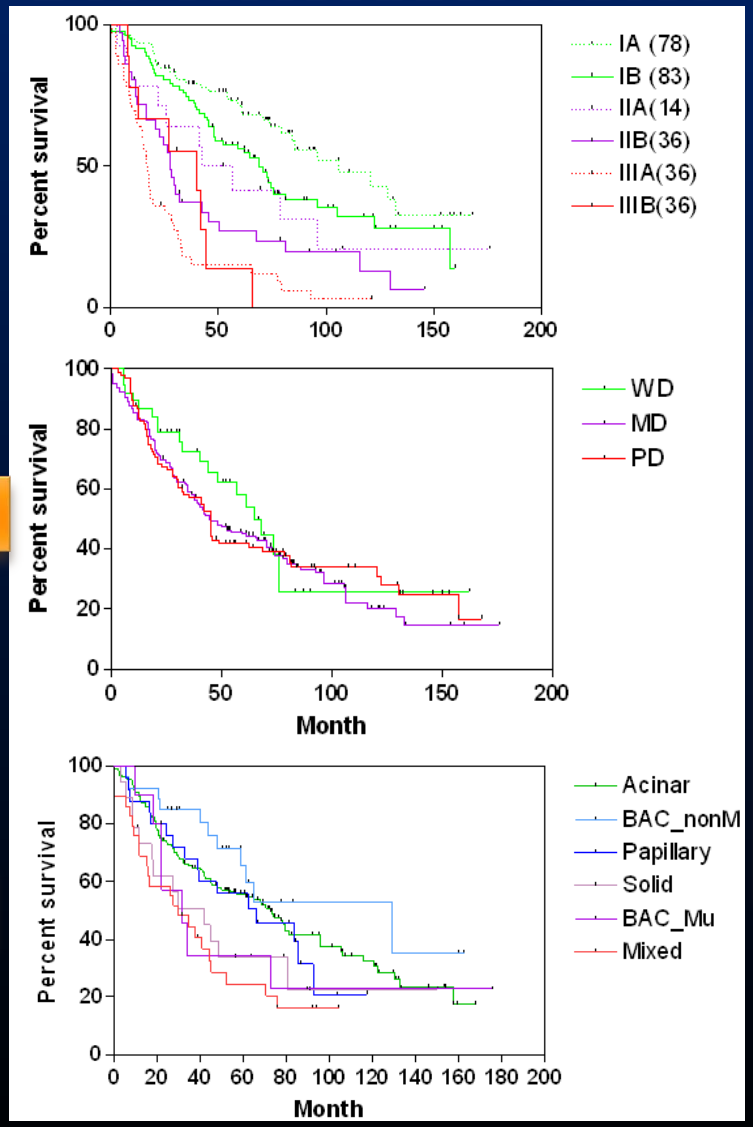
CQ Zhu

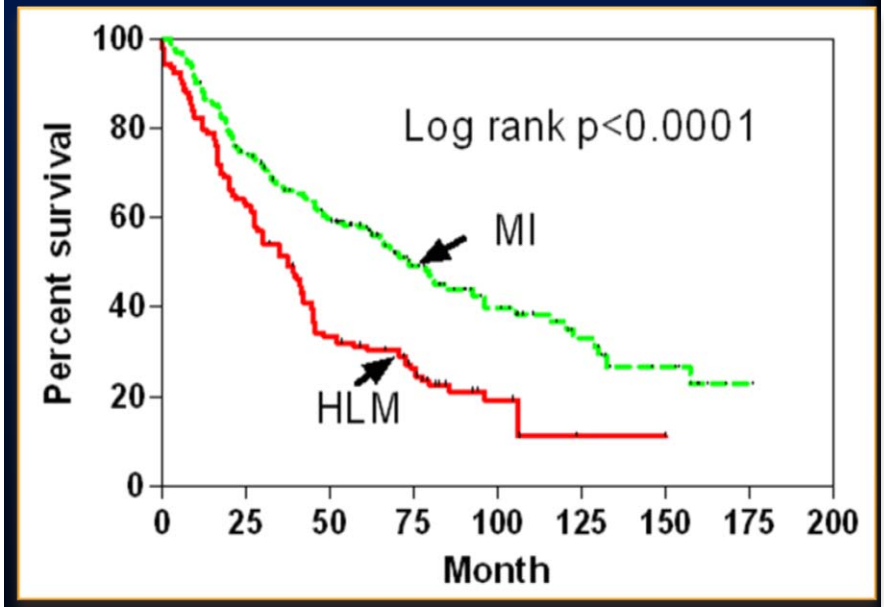
NOT AS PROMISING AS IT MAY SEEM

Stage

Differentiation

Histology





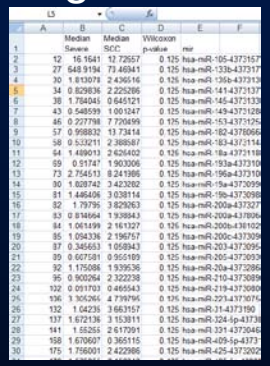
Site

CQ Zhu

SYSTEMATIC SIGNATURE SELECTION

Start with N=1 “signature”

Rank N-gene signatures

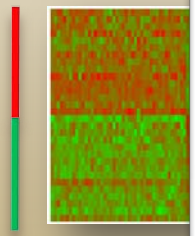


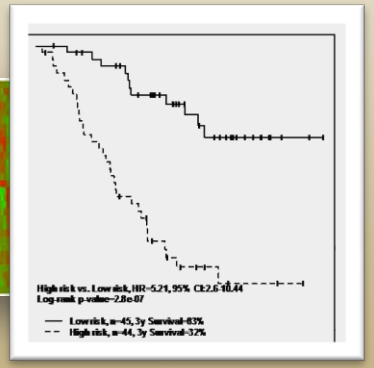
Select the best

Evaluate the signature

K-medians clustering

Survival diff test





N+1 gene signatures

Improvement?

Yes

No

Final signature

PNAS

Prognostic gene signatures for non-small-cell lung cancer

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

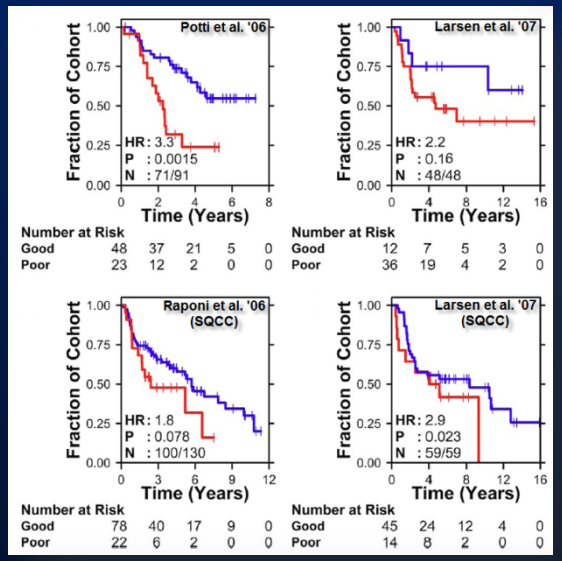
Departments of [a]Medical Biophysics, [d]Medicine, [e]Laboratory Medicine and Pathology, and [f]Computer Science, University of Toronto, Toronto, ON, Canada M5S 1A1; [b]Ontario Cancer Institute, University Health Network, Toronto, ON, Canada M5G 2M9; and [c]Division of Medical Oncology, Princess Margaret Hospital, Toronto, ON, Canada M5G 2M9

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

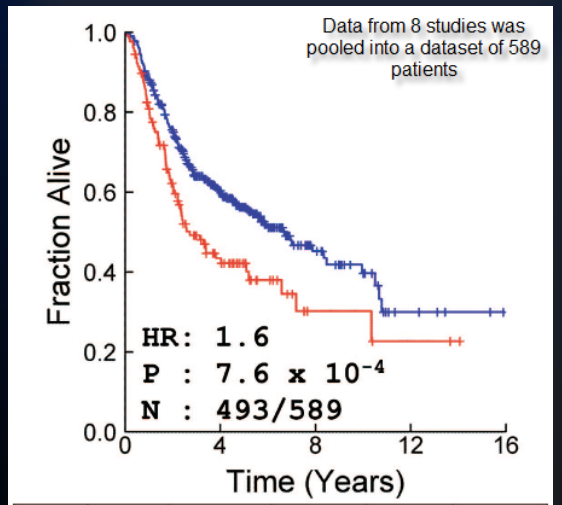
THERE IS MORE THAN ONE SIGNATURE

- 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



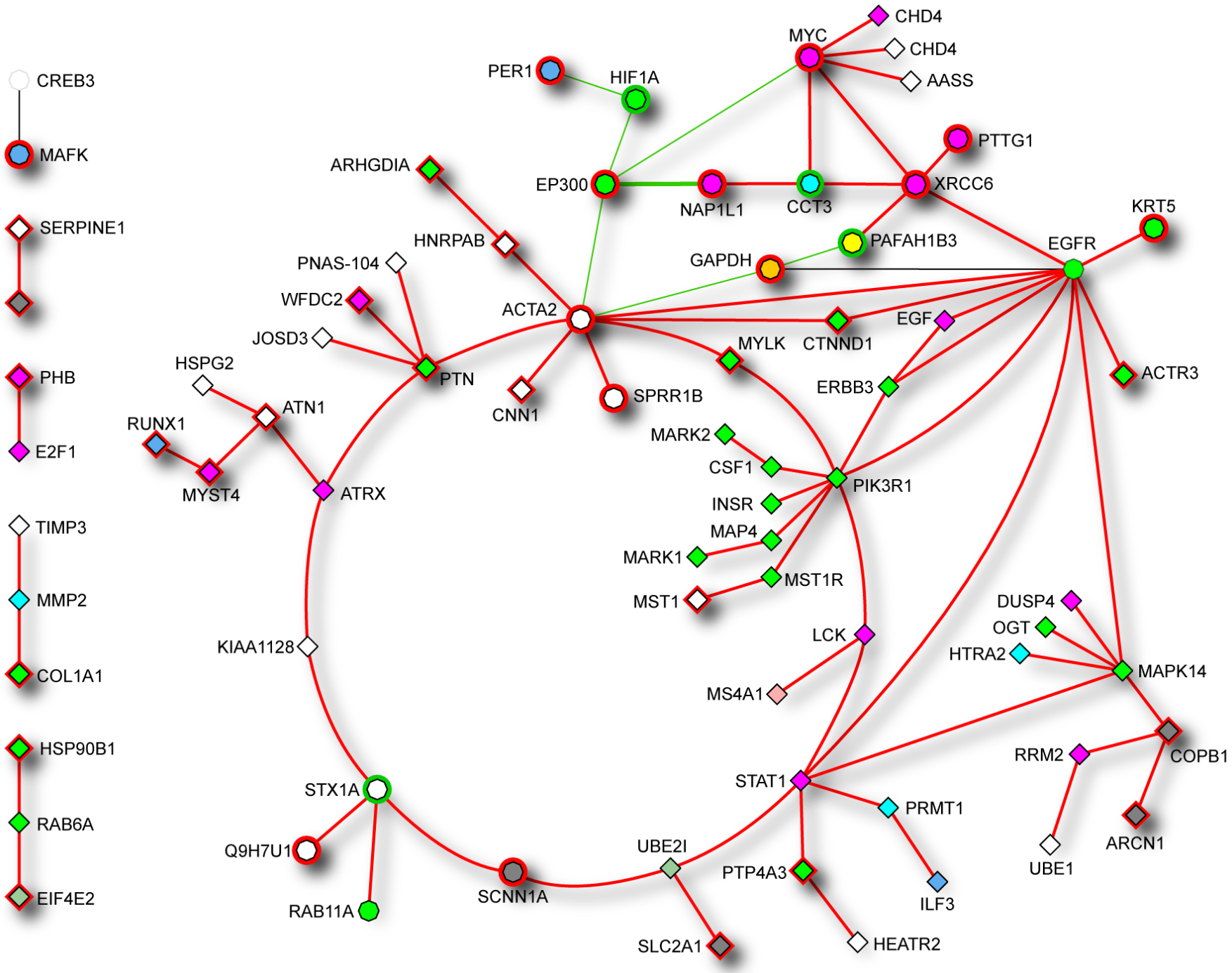
Good	365	159	30	5	0
Poor	128	35	4	2	0

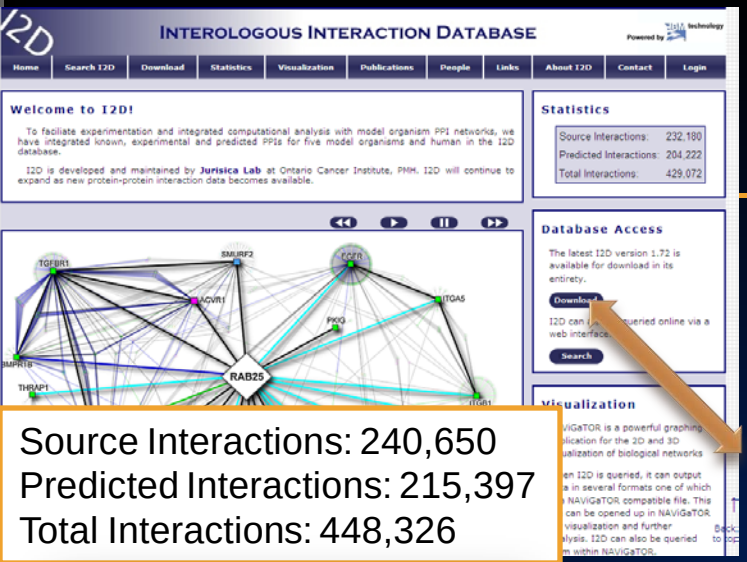
Prognostic gene signatures for non-small-cell lung cancer

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

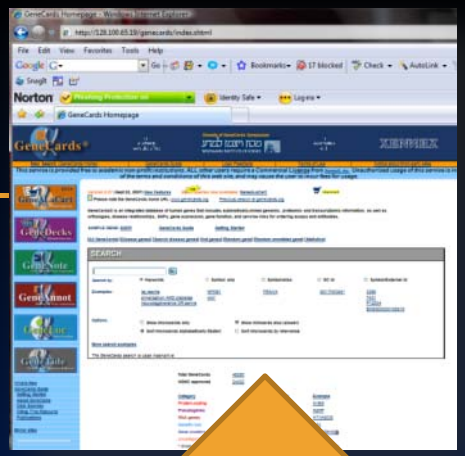
Departments of [a]Medical Biophysics, [d]Medicine, [e]Laboratory Medicine and Pathology, and [f]Computer Science, University of Toronto, Toronto, ON, Canada M5S 1A1; [b]Ontario Cancer Institute, University Health Network, Toronto, ON, Canada M5G 2M9; and [c]Division of Medical Oncology, Princess Margaret Hospital, Toronto, ON, Canada M5G 2M9

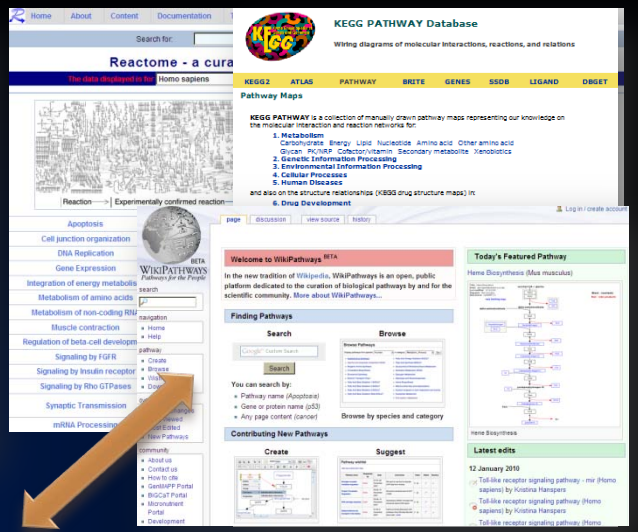
Edited by Tak Wah Mak, University of Toronto, Toronto, ON, Canada, and approved December 23, 2008 (received for review September 21, 2008)

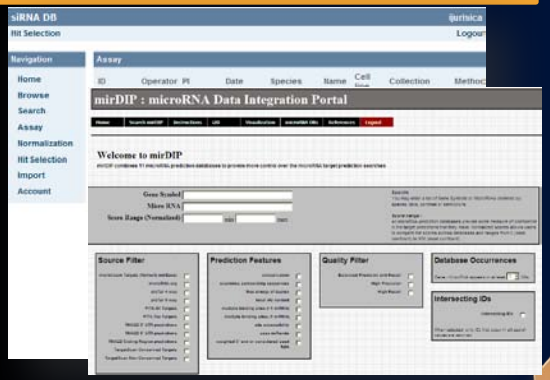

CREB3
MAFK
SERPINE1
PHB
E2F1
TIMP3
MMP2
COL1A1
HSP90B1
RAB6A
EIF4E2
CHD4
MYC
CHD4
AASS
PER1
HIF1A
PTTG1
ARHGDIA
EP300
NAP1L1
CCT3
XRCC6
KRT5
HNRPAB
PAFAH1B3
EGFR
PNAS-104
GAPDH
WFDC2
ACTA2
EGF
JOSD3
MYLK
CTNND1
HSPG2
PTN
ACTR3
ATN1
CNN1
SPRR1B
ERBB3
RUNX1
MARK2
CSF1
MYST4
ATRX
PIK3R1
INSR
MAP4
MARK1
MST1R
MST1
DUSP4
OGT
LCK
KIAA1128
HTRA2
MAPK14
MS4A1
RRM2
COPB1
STX1A
STAT1
PRMT1
Q9H7U1
UBE2I
ARCN1
UBE1
PTP4A3
SCNN1A
ILF3
RAB11A
SLC2A1
HEATR2

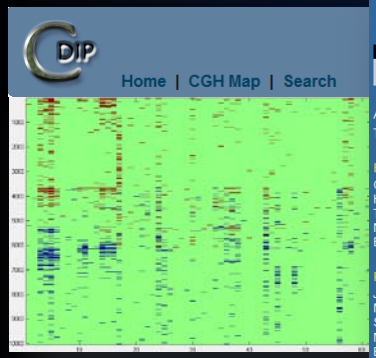


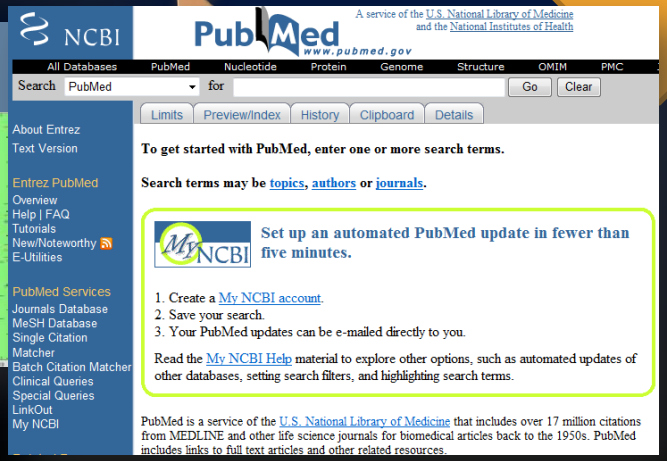
Source Interactions: 240,650
Predicted Interactions: 215,397
Total Interactions: 448,326

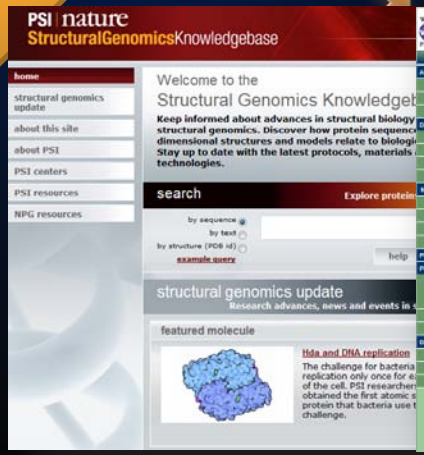




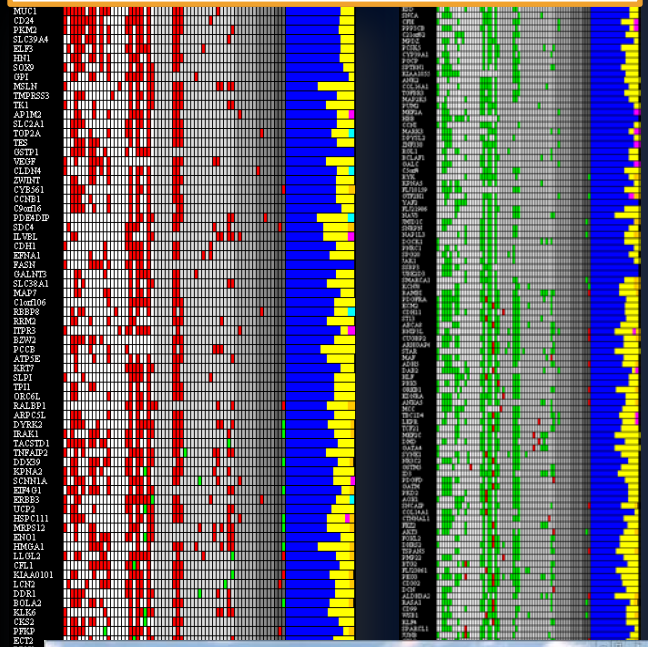


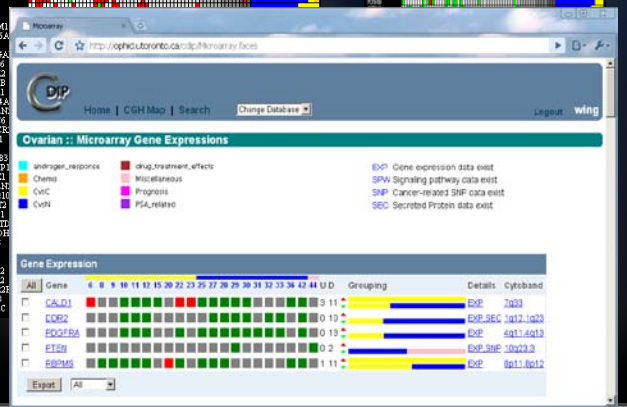




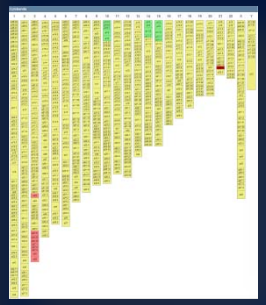


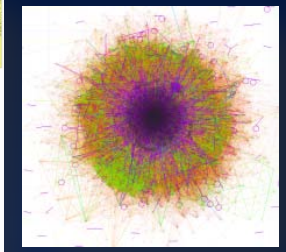
Blue	Cancer *vs.* Normal
Yellow	Cancer *vs.* cancer
Magenta	Prognosis
Cyan	Androgen response
Orange	PSA related
Pink	Drug/treatment effect

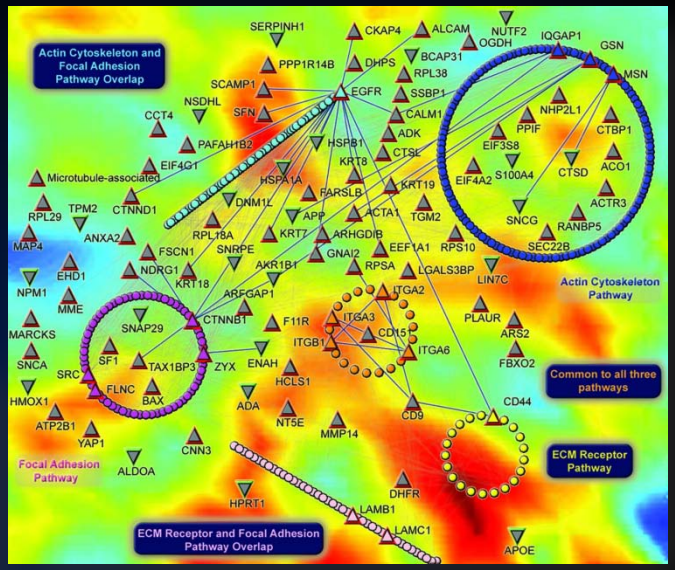




- **12,975 unique genes**
- 60% or more in one direction & at least ≥6 studies in one direction
  - **UP** **198**
  - **DOWN** **408**







Gene from Max number of studies : NBL1 (8↑ & 11↓)

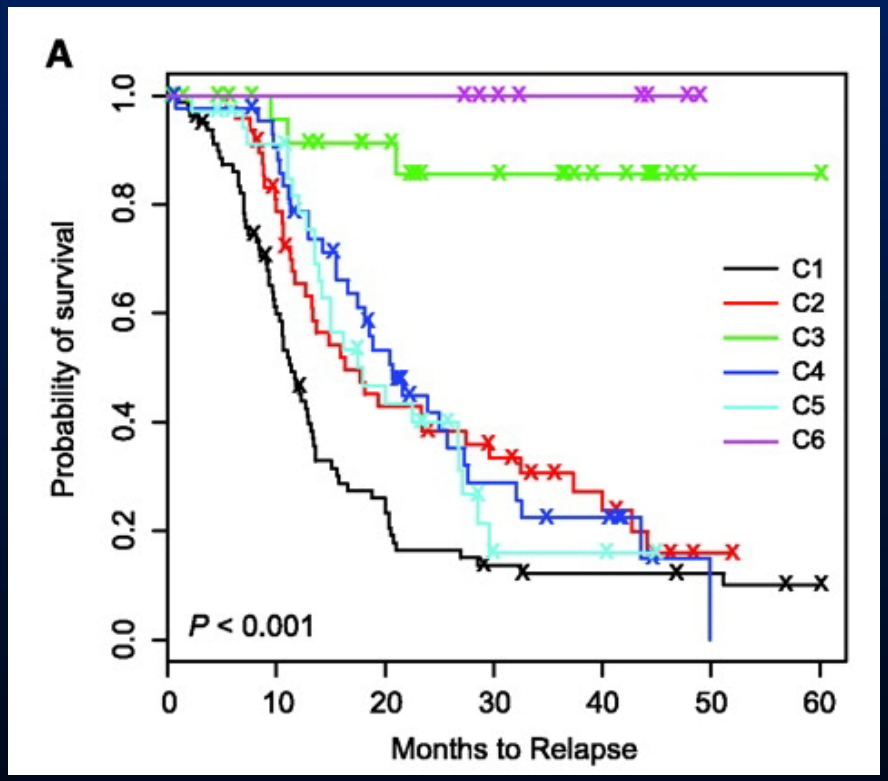
Gene with Max Ups : Myc (14↑ & 2↓)
Gene with 100% Up (max #studies) : NEURL (9↑)

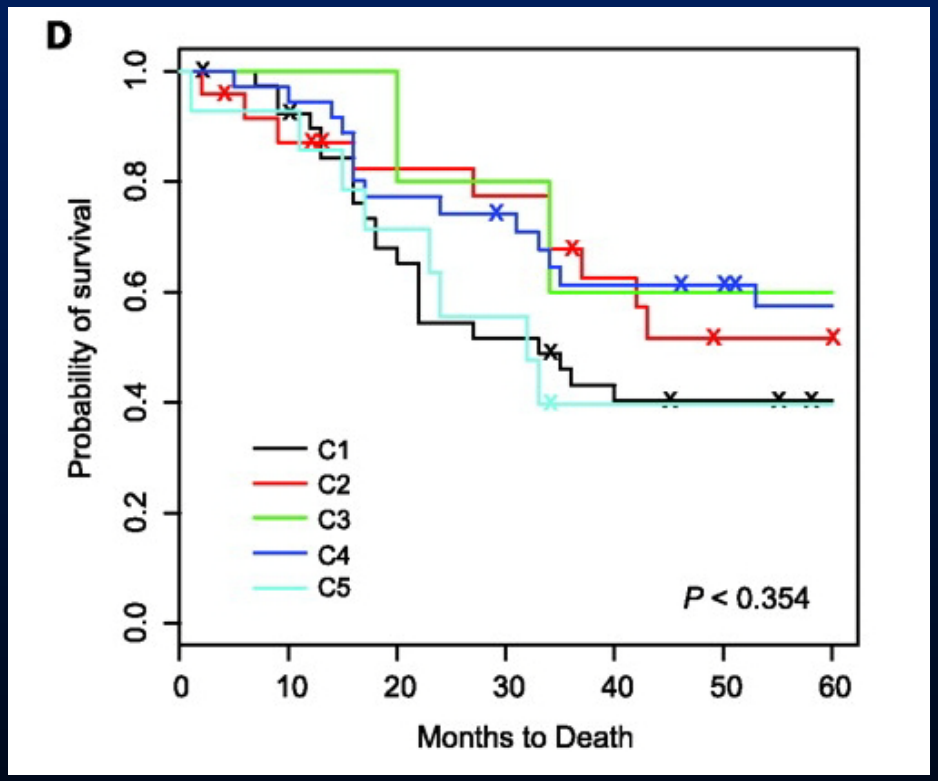
Gene with Max Downs : MYLK (14↓ & 4↑)
Gene with 100% Down (max #studies) : SERPINB1 (11↓)

CGH data:

Amplification : 3q (3q27.3, 3q27.2, 3q27.1, 3q26.33, 3q26.32, 3q26.31, 3q29, 3q28, 3q23)
Deletion : 10p, 14p, 15q/p (15q11.1, 14p13, 14q12, 14q11.2, 14q11.1, 10p15.3, 10p15.2, 10p15.1, 15q13, 15q12, 15p11.2,15p11.1, 10p14, 10p13

OVARIAN CANCER





Training

Independent Validation

Clin Cancer Res. 2008 Aug 15;14(16):5198-208.

Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome.

Tothill RW, Tinker AV, George J, Brown R, Fox SB, Lade S, Johnson DS, Trivett MK, Etemadmoghadam D, Locandro B, Traficante N, Fereday S, Hung JA, Chiew YE, Haviv I; Australian Ovarian Cancer Study Group, Gertig D, DeFazio A, Bowtell DD.

J Clin Oncol. 2007 Feb 10;25(5):517-25.

An integrated genomic-based approach to individualized treatment of patients with advanced-stage ovarian cancer.

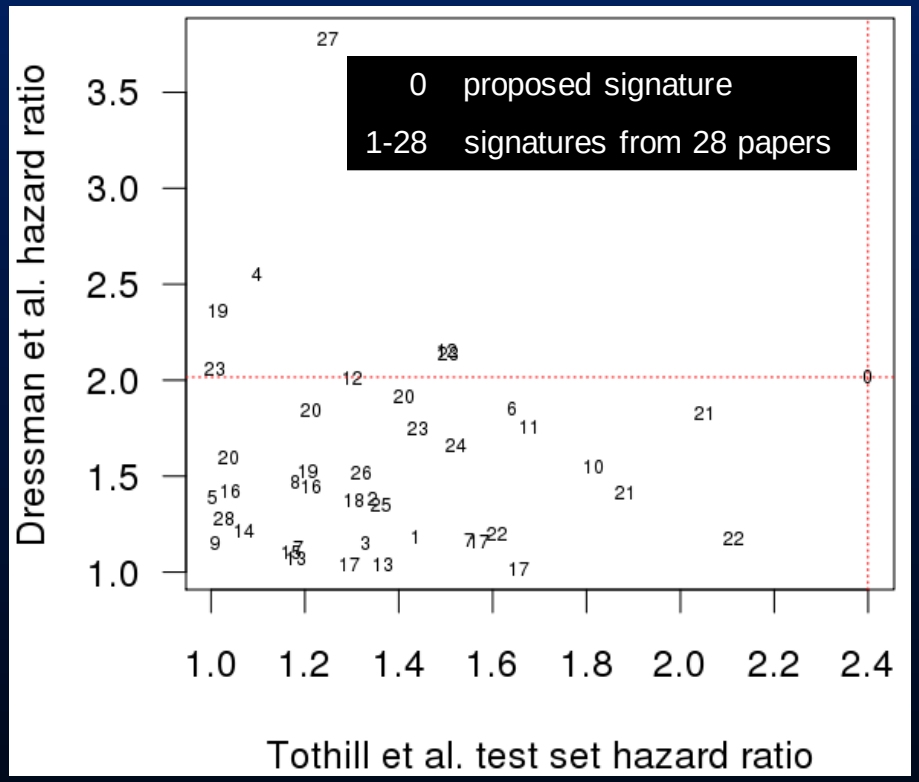
Dressman HK, Berchuck A, Chan G, Zhai J, Bild A, Sayer R, Cragun J, Clarke J, Whitaker RS, Li L, Gray J, Marks J, Ginsburg GS, Potti A, West M, Nevins JR, Lancaster JM.

Nucleic Acids Res. 2010 Jan;38(Database issue):D716-25. Epub 2009 Nov 24.

GeneSigDB--a curated database of gene expression signatures.

Culhane AC, Schwarzl T, Sultana R, Picard KC, Picard SC, Lu TH, Franklin KR, French SJ, Papenhausen G, Correll M

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

....

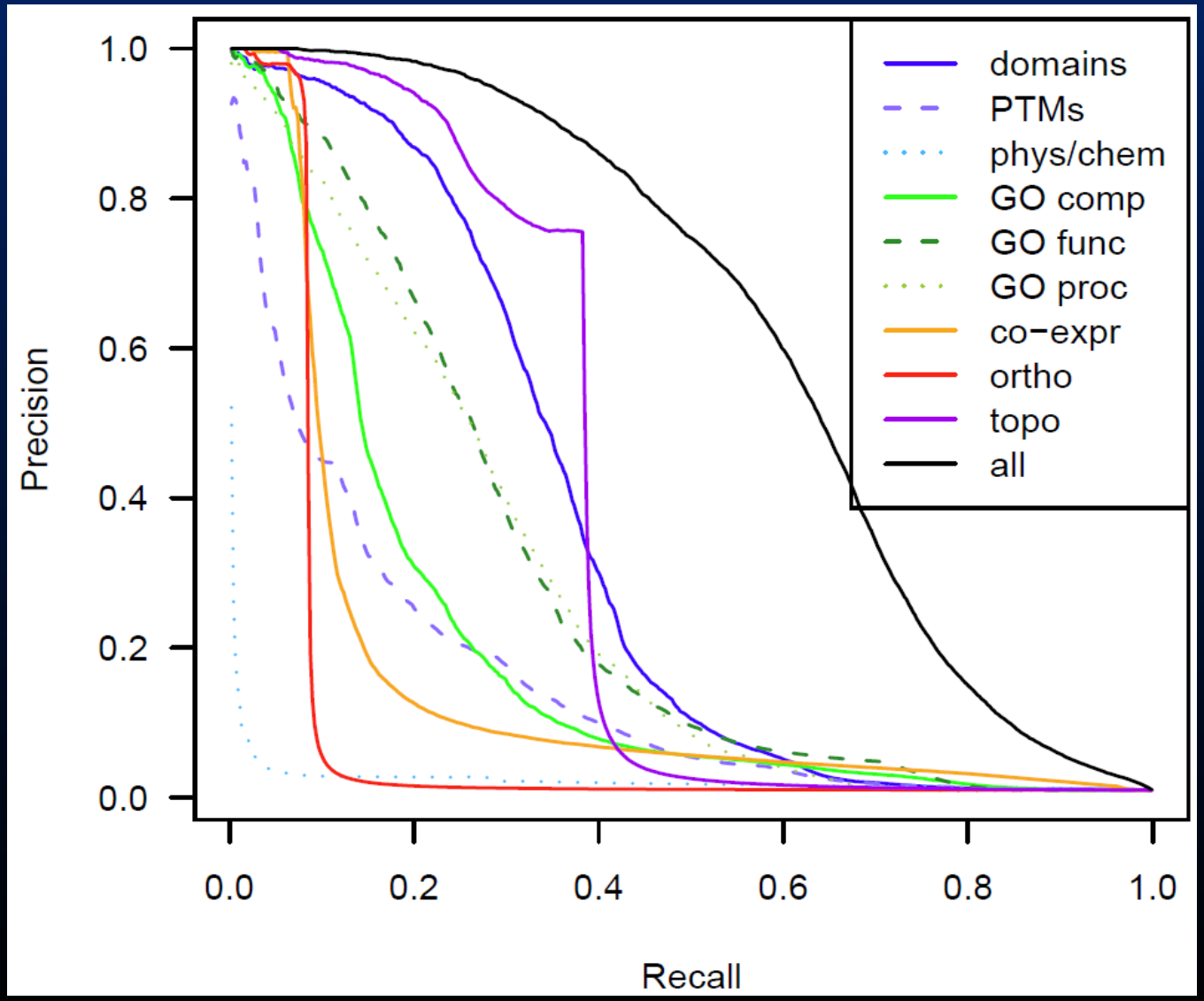
30,363 human interactions

303,630 random human protein pairs

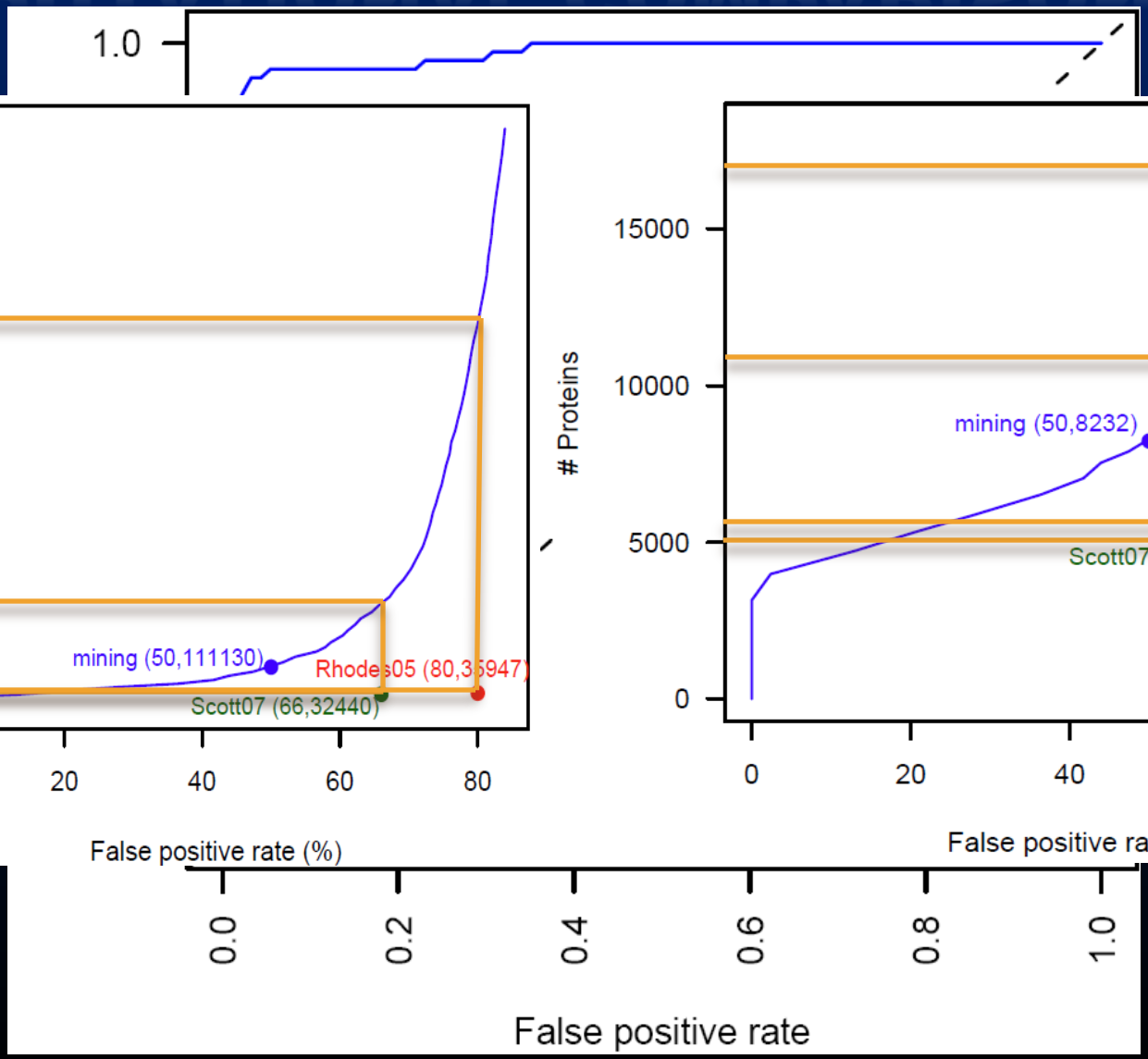
Protein Pair	Gene Co-expression	Ortholog Co-occurrence	Domains (log odds)	GO (Semantic Similarity)			Interacts
				Local	Function	Process	
P08133 , P55327	-0.041	-0.285	3.906	0.166	0.041	1.000	1
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
P17612 , P98175							0
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:

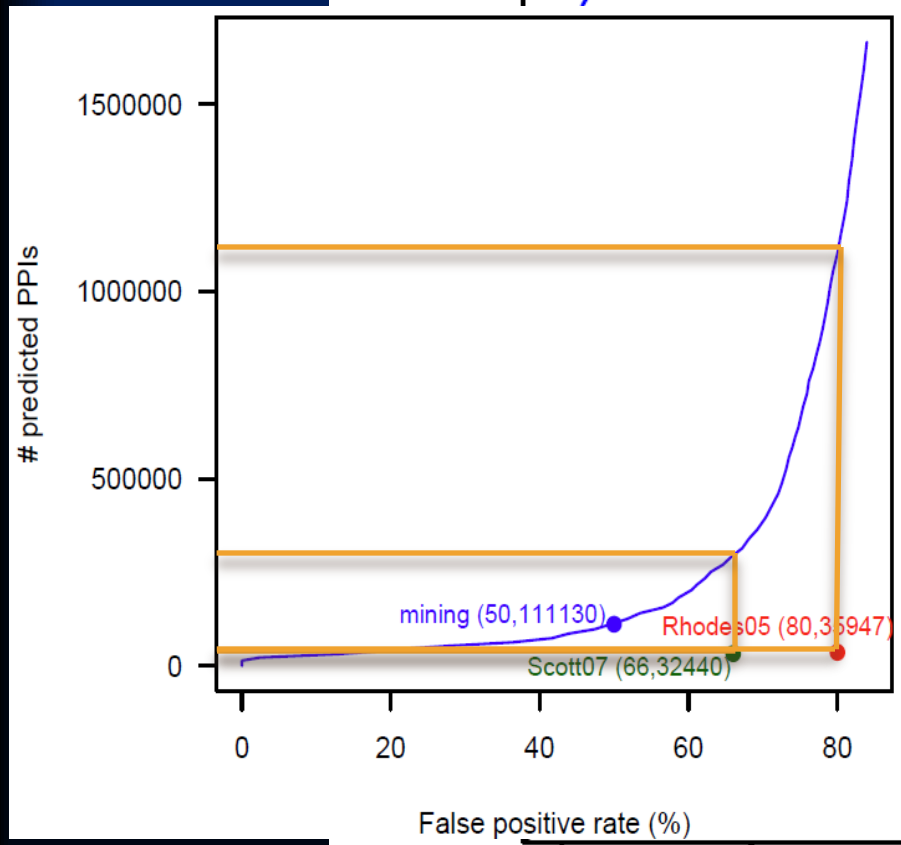
- Association mining algorithm identifies sets of features that occur predominantly in PPIs.
- Decision tree algorithm is then used to predict PPIs.

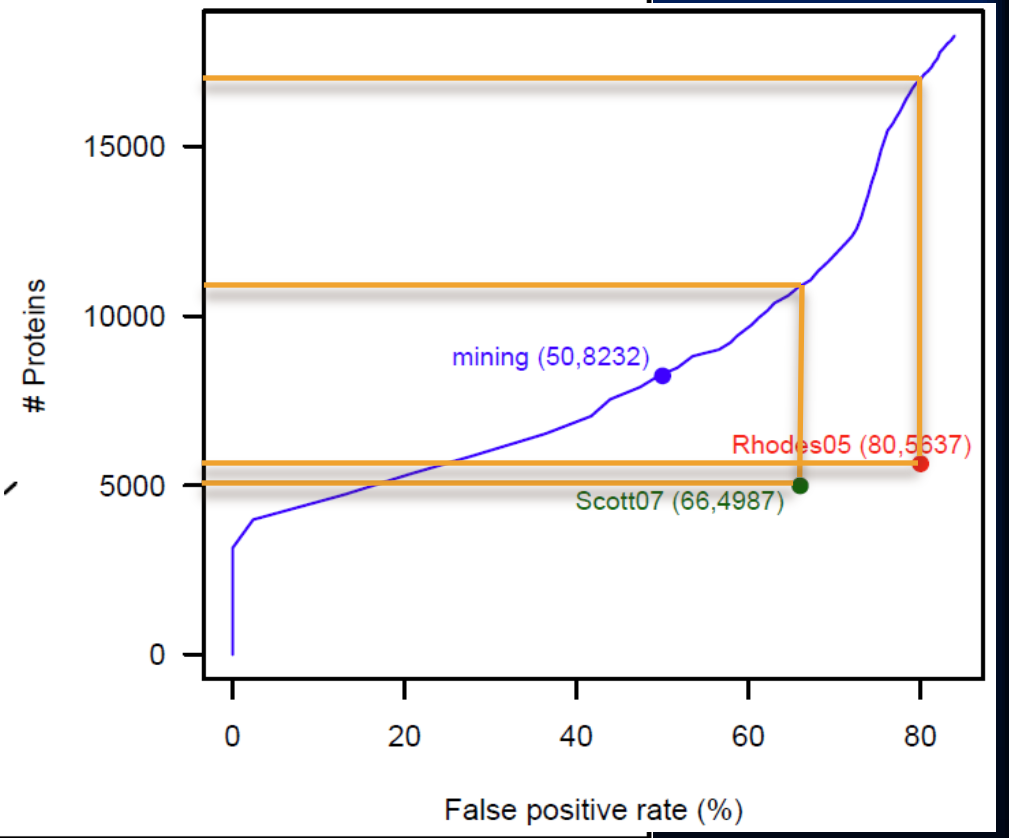
FPCLASS


1.0
0.8
0.6
0.4
0.2
0.0
Precision
0.0
0.2
0.4
0.6
0.8
1.0
Recall
domains
PTMs
phys/chem
GO comp
GO func
GO proc
co−expr
ortho
topo
all

COMPUTATIONAL COMPARISON

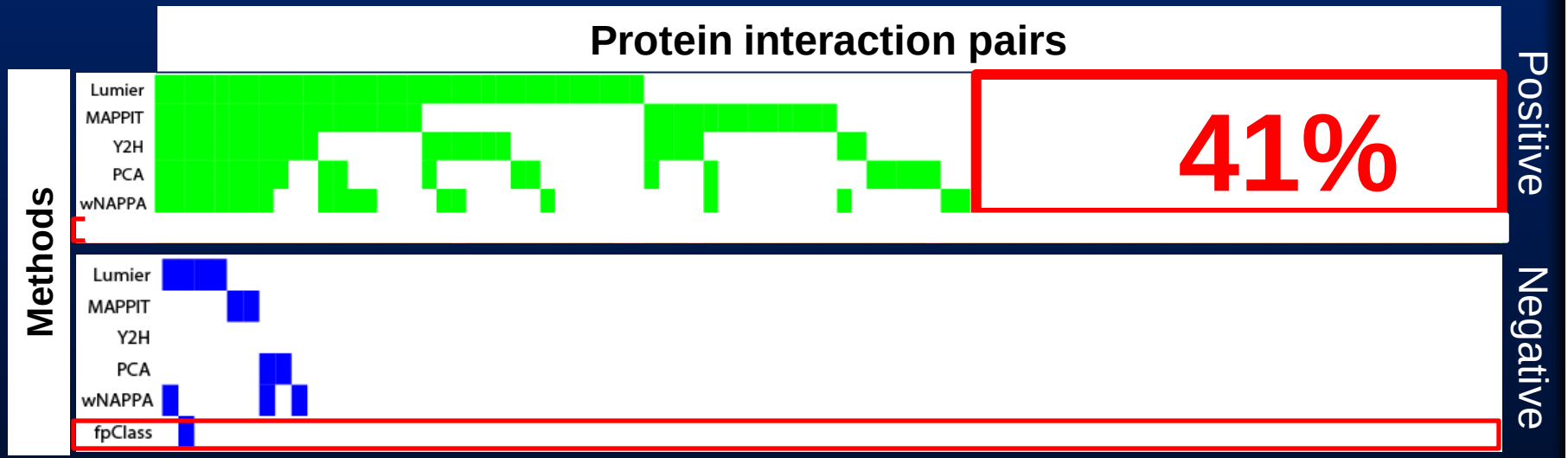






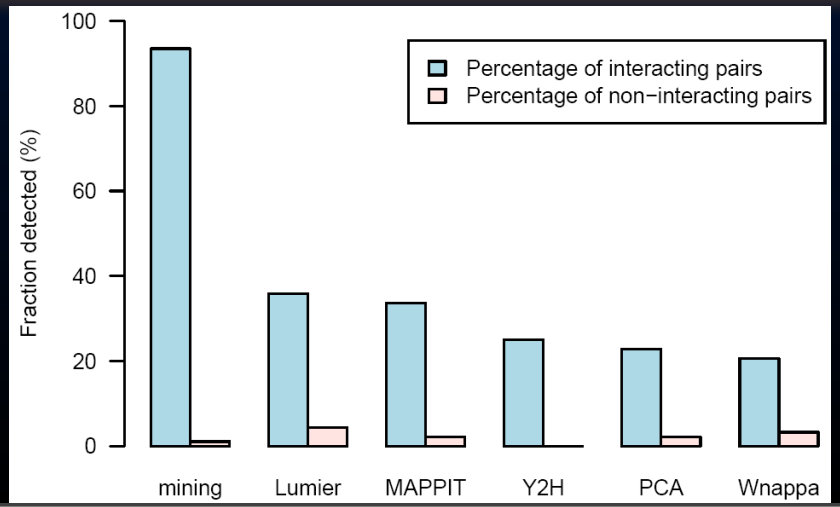
DATA MINING & HTP VALIDATION

Braun et al., *Nat Methods*, 2009

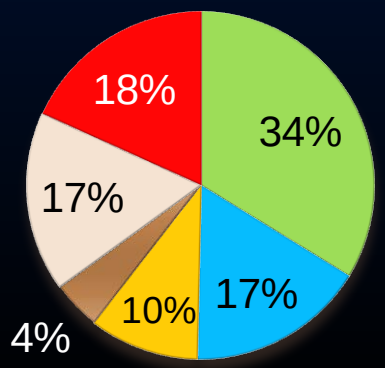


Individual HTP: 25-**35%** interactions

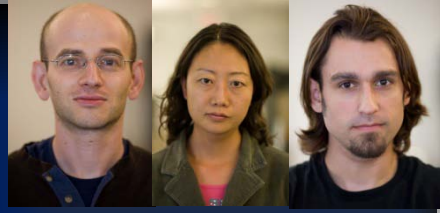
All HTP combined: **59%** interactions



Data Mining

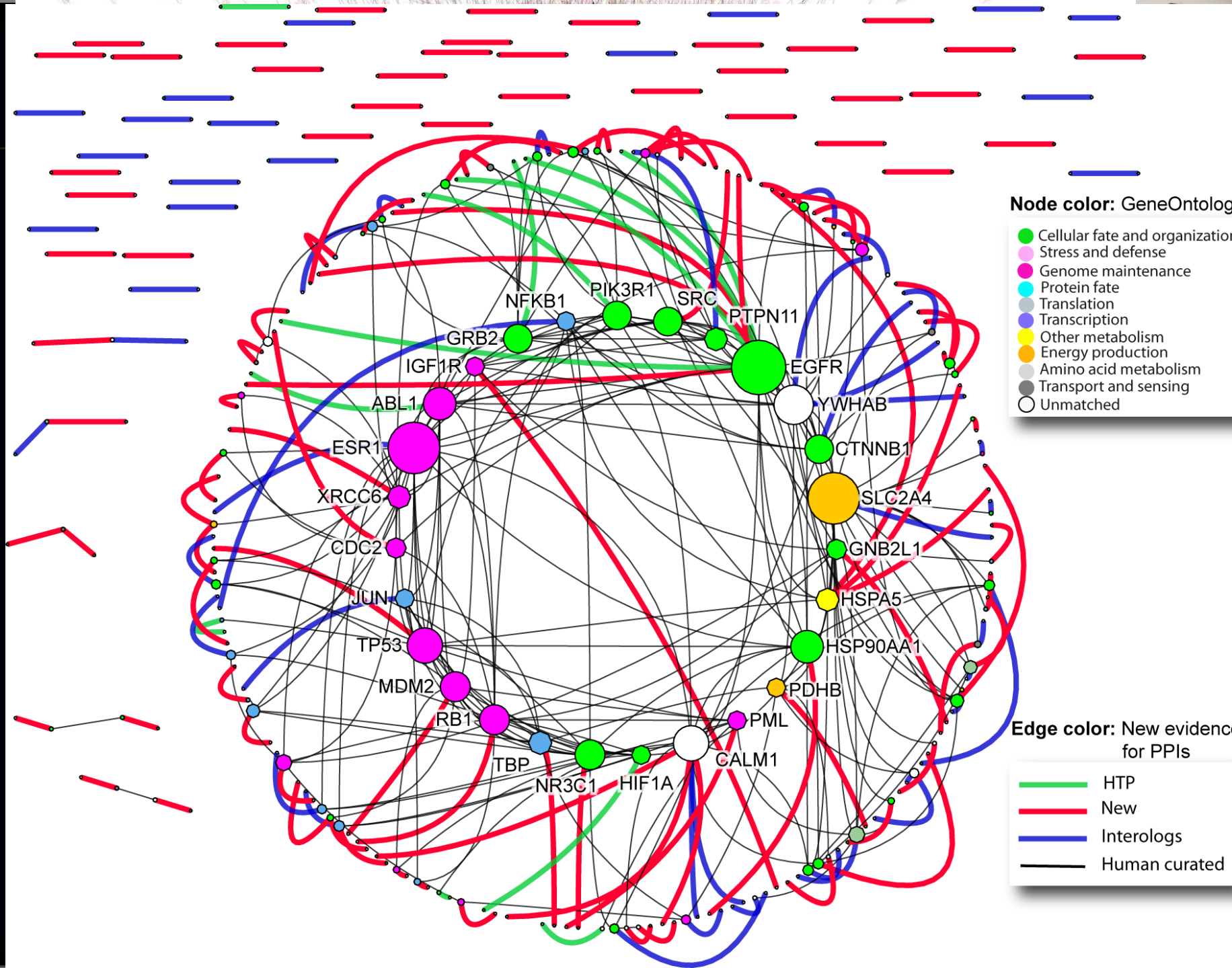


- Lumier
- MS
- PA
- Y2H
- coIP
- TM

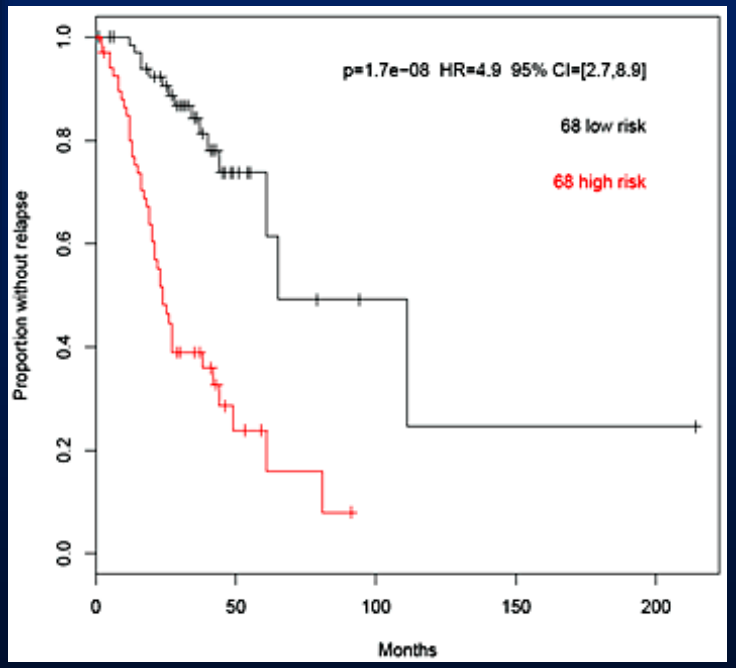


TEXT MINING & CURATION

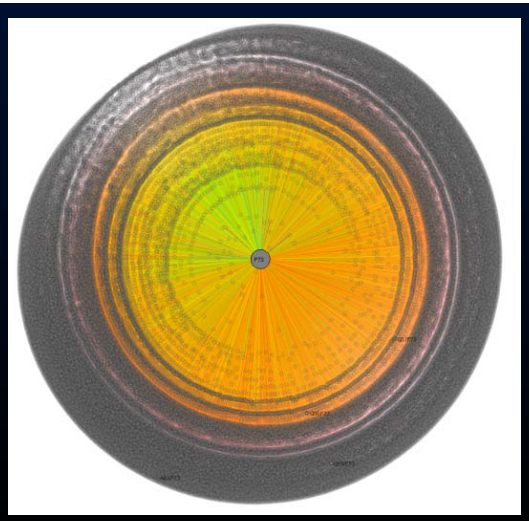
- 176,075 known and predicted PPIs in I2D
  - 113,230 had no evidence and 62,845 had some evidence from text mining
  - 9,398 abstracts had SVM score >0.5, of which 6,290 manuscripts published prior to 2003 and 8,847 prior to 2006
  - these 9,398 abstracts cover 13,800 unique PPIs
- Distribution across 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.

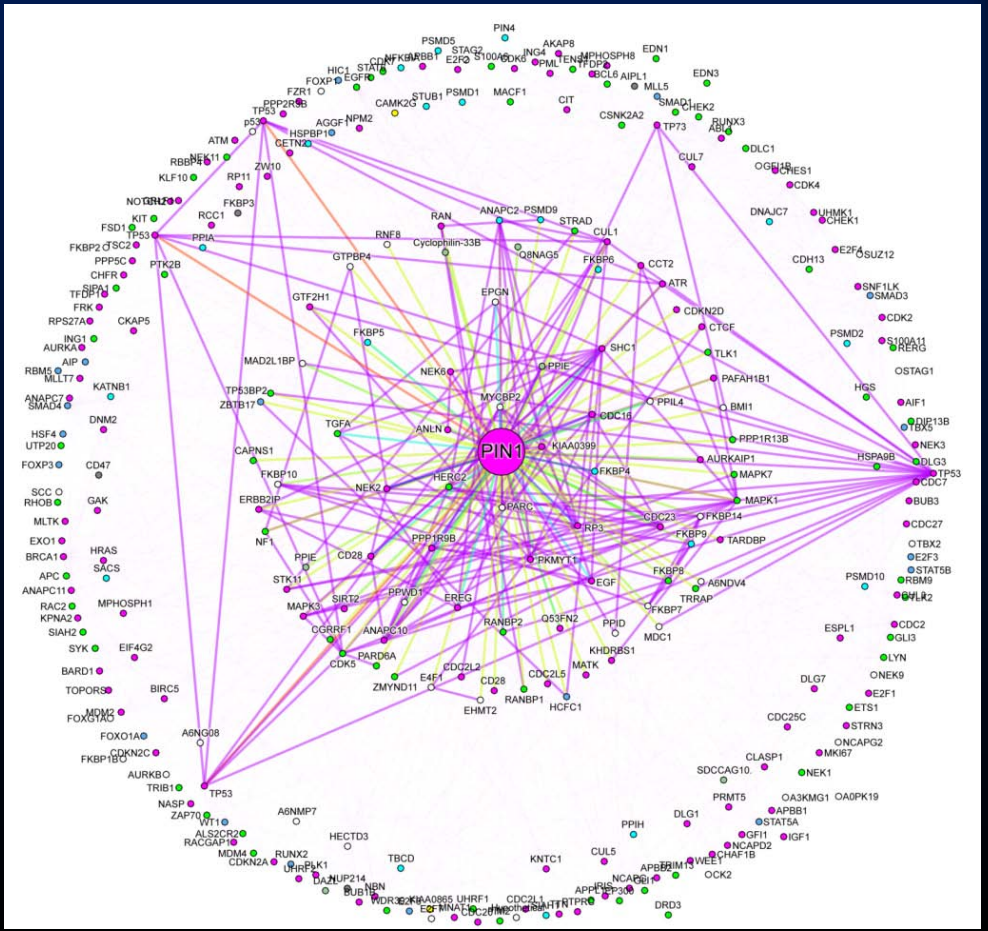

Node color: GeneOntolog
Cellular fate and organizatio
Stress and defense
Genome maintenance
Protein fate
Translation
Transcription
Other metabolism
Energy production
Amino acid metabolism
Transport and sensing
Unmatched
Edge color: New evidenc
for PPIs
HTP
New
Interologs
Human curated
NFKB1
PIK3R1
SRC
PTPN11
GRB2
IGF1R
EGFR
ABL1
YWHAB
ESR1
CTNNB1
XRCC6
SLC2A4
CDC2
GNB2L1
JUN
HSPA5
TP53
HSP90AA1
MDM2
PDHB
RB1
PML
TBP
CALM1
NR3C1
HIF1A

VALIDATING 6,631 PREDICTED PPIS

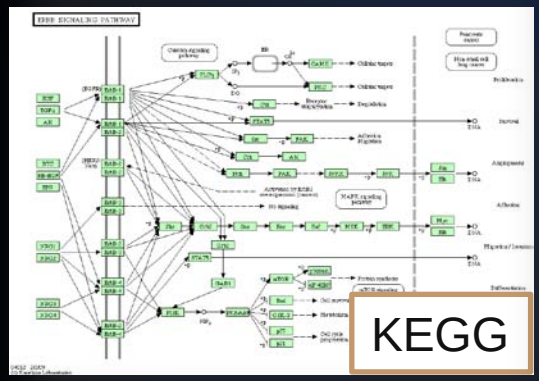


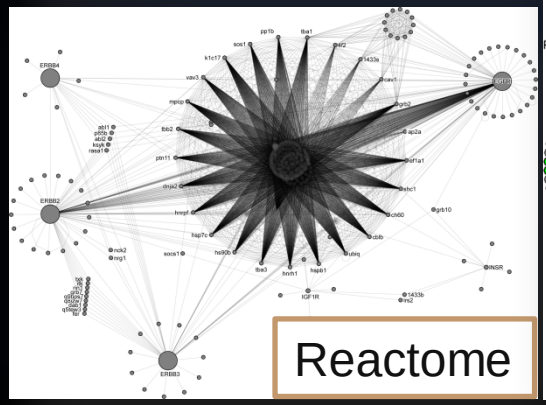
	SVM	Co-occurrence
OVARIAN	56	631
PROSTATE	33	799
LUNG	3	37

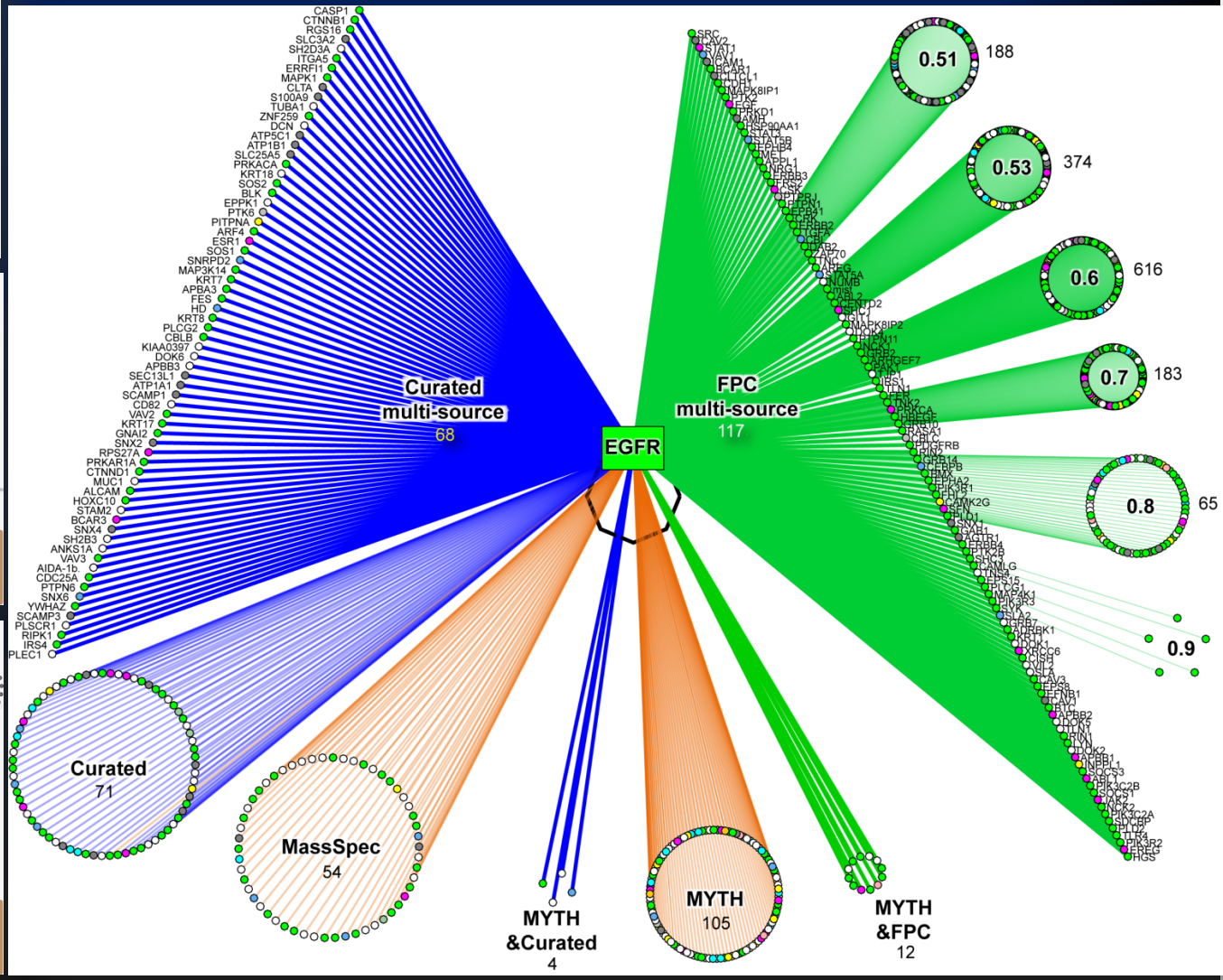




EGFR







RESEARCH ARTICLE

CELL BIOLOGY

Science Signaling

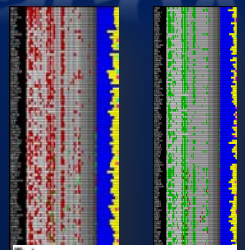
Regulation of Epidermal Growth Factor Receptor Trafficking by Lysine Deacetylase HDAC6

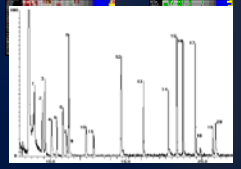
Yonathan Lissanu Deribe,[1] Philipp Wild,[1] Akhila Chandrashaker,[2] Jasna Curak,[3] Mirko H. H. Schmidt,[1,4] Yannis Kalaidzidis,[2,5] Natasa Milutinovic,[3] Irina Kratchmarova,[6] Lukas Buerkle,[3,7] Michael J. Fetchko,[3] Philipp Schmidt,[1] Saranya Kittanakom,[3] Kevin R. Brown,[8] Igor Jurisica,[8,9,10]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

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

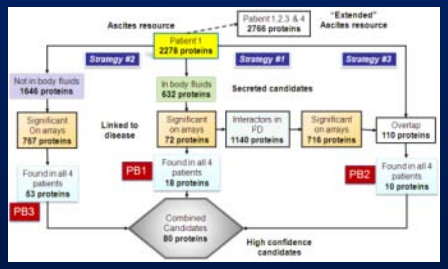
DISCOVERY PIPELINE

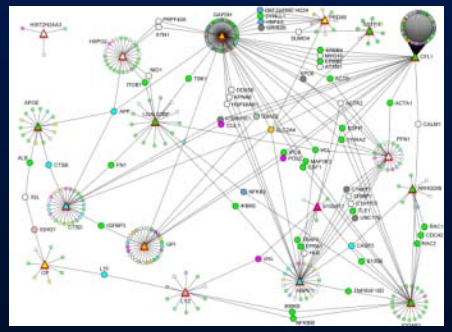
Data





+

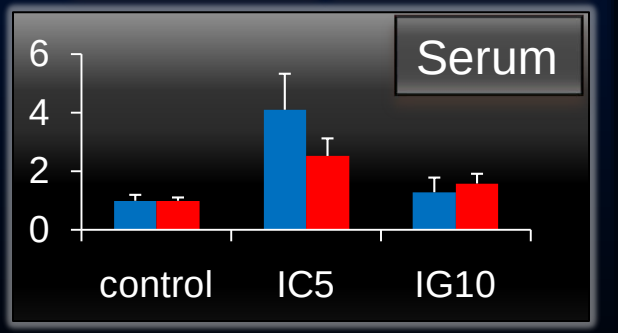




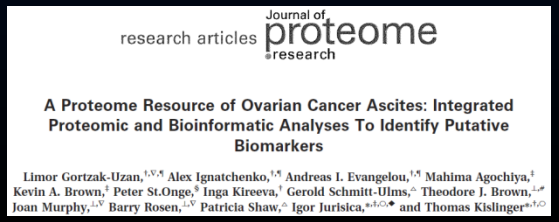
Analysis

=

Prediction & Validation & Modeling



Tools & Resources

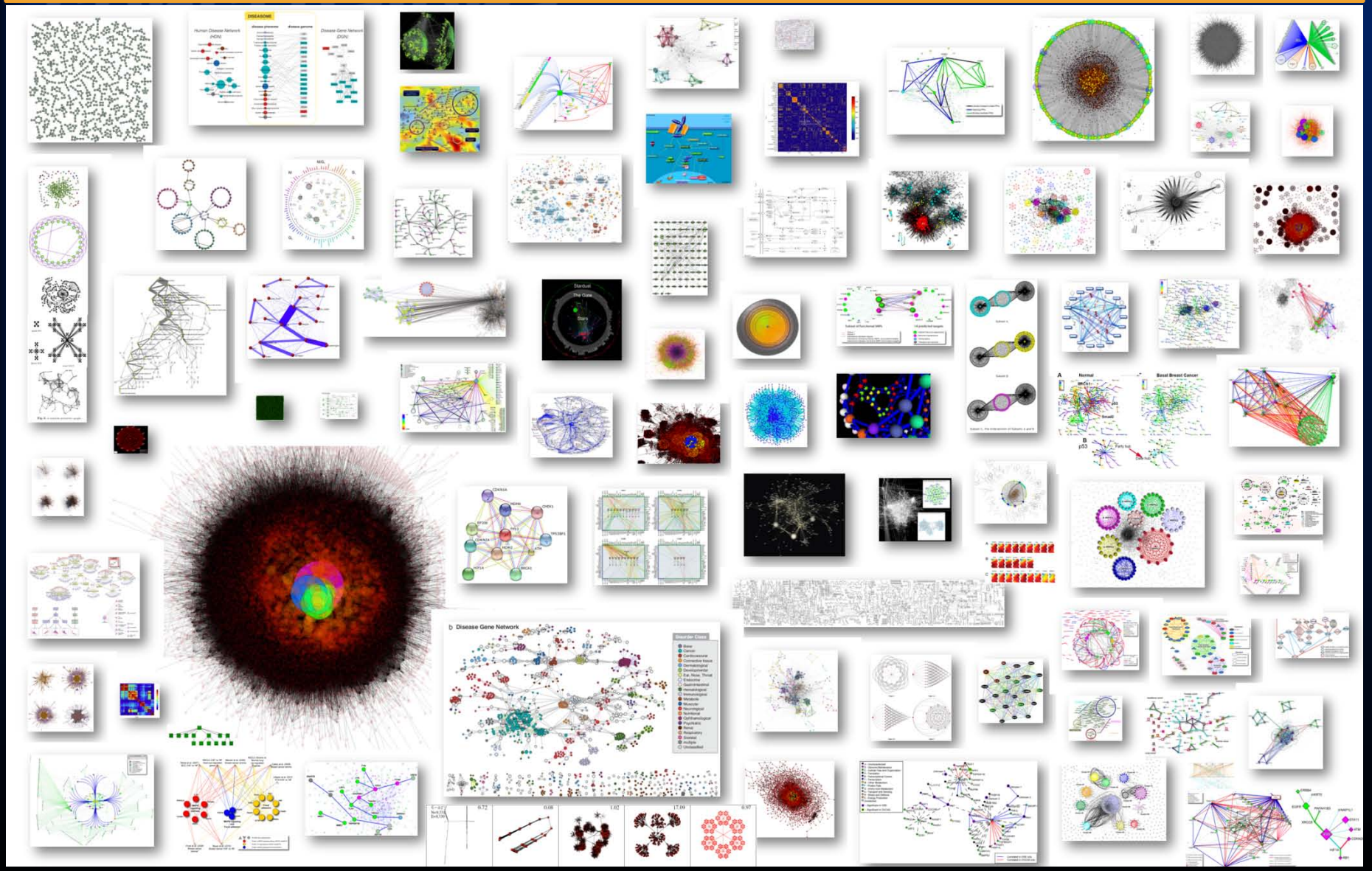


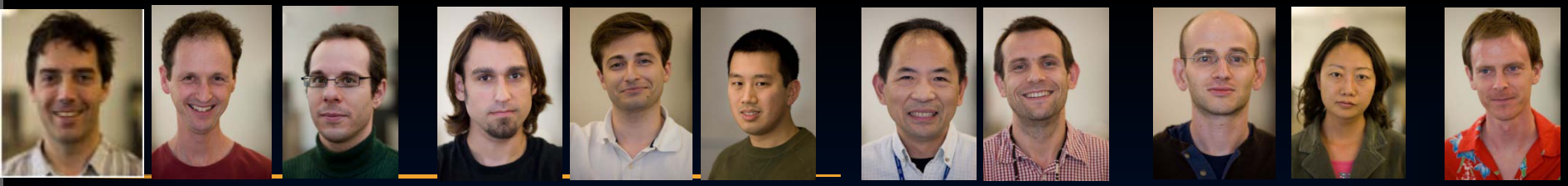
research articles Journal of proteome research

A Proteome Resource of Ovarian Cancer Ascites: Integrated Proteomic and Bioinformatic Analyses To Identify Putative Biomarkers

Limor Gortzak-Uzan,[†,∇,¶] Alex Ignatchenko,[†,¶] Andreas I. Evangelou,[†,¶] Mahima Agochiya,[‡] Kevin A. Brown,[‡] Peter St.Onge,[§] Inga Kireeva,[†] Gerold Schmitt-Ulms,[△] Theodore J. Brown,[△,#] Joan Murphy,[⊥,∇] Barry Rosen,[⊥,∇] Patricia Shaw,[△] Igor Jurisica,[*,‡,○,♦] and Thomas Kislinger[*,†,○]

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

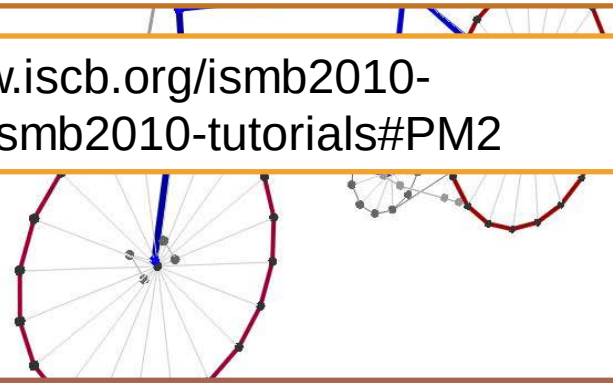
M. S. Tsao, F. Shepherd, L. Penn , M. Pintilie,
L. Penn, R. Bristow, N. Fleshner
I. Stagljar, I. Dikic
A. Oza, A. Jurisicova, T. Kislinger, T. Colgan
G. Mills, …

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

Canada Research Chairs
NSERC CRSNG
Powered by IBM technology
GenomeCanada
CIHR IRSC
Ontario

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