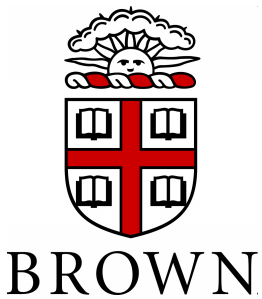
Network Analysis of Mutations Across Cancer Types

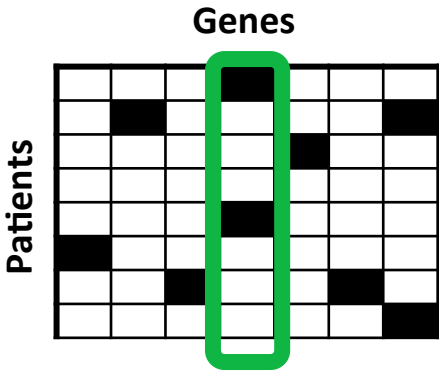
Ben Raphael
Fabio Vandin, Max Leiserson, Hsin-Ta Wu



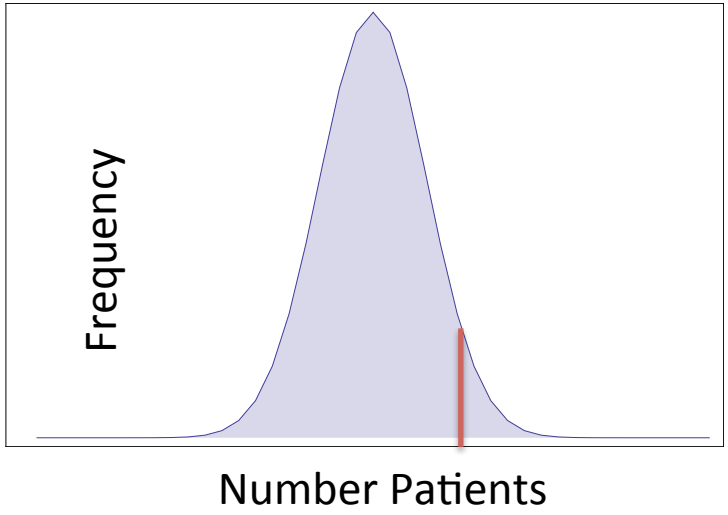
Department of Computer Science
Center for Computational Molecular Biology

Significantly Mutated Genes

Mutation Matrix

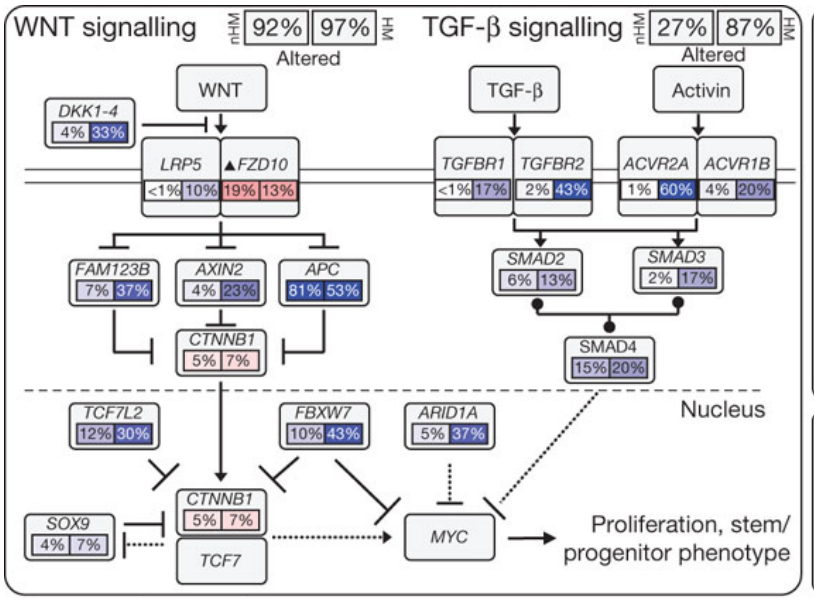


Statistical test



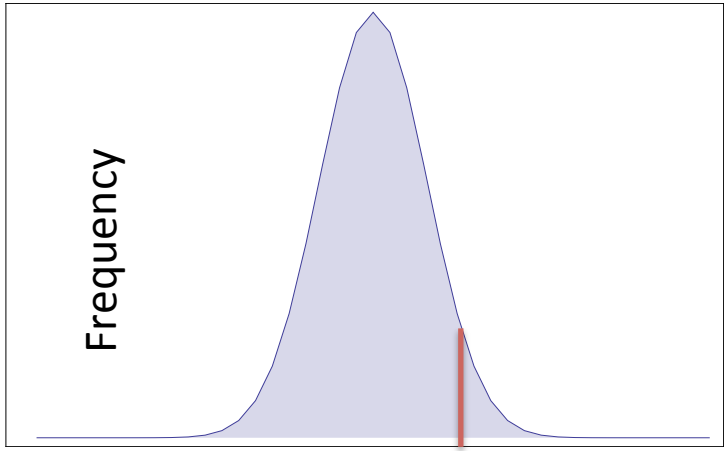
Study	Num. Samples	Num. SMG
TCGA Ovarian (2011)	316	10
TCGA Breast (2012)	510	35
TCGA Colorectal (2012)	276	32

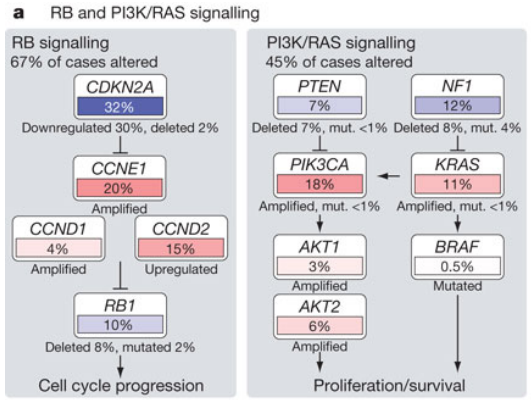
Significantly Mutated Genes → Pathways



TCGA Colorectal (*Nature* 2012)

Statistical test

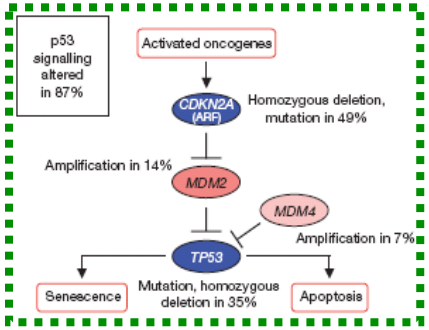


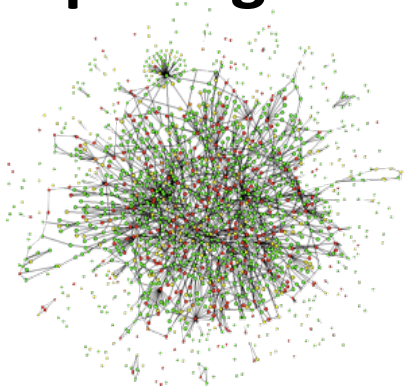


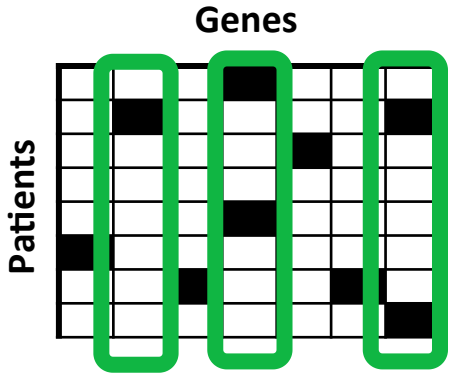
TCGA Ovarian (*Nature* 2011)

Advantages of Large Datasets

Prior knowledge of groups of genes







Known pathways

Interaction Network

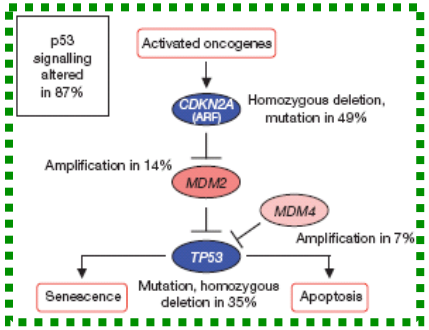
None

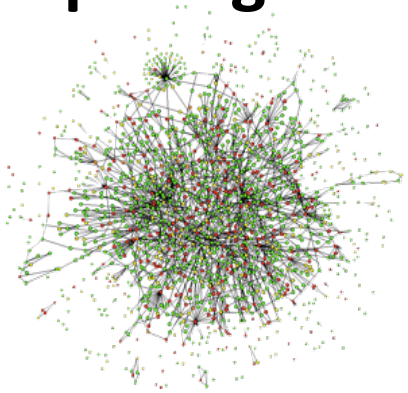
Prior knowledge

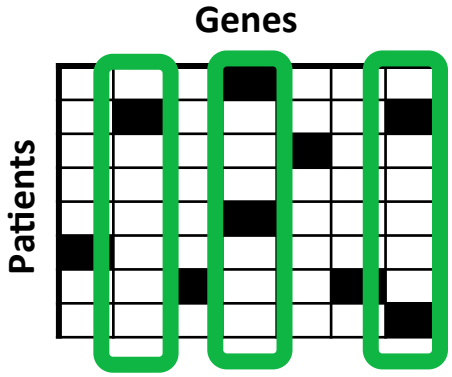
- Novel pathways or interactions between pathways (*crosstalk*)
- Topology of interactions

Two Algorithms

Prior knowledge of groups of genes



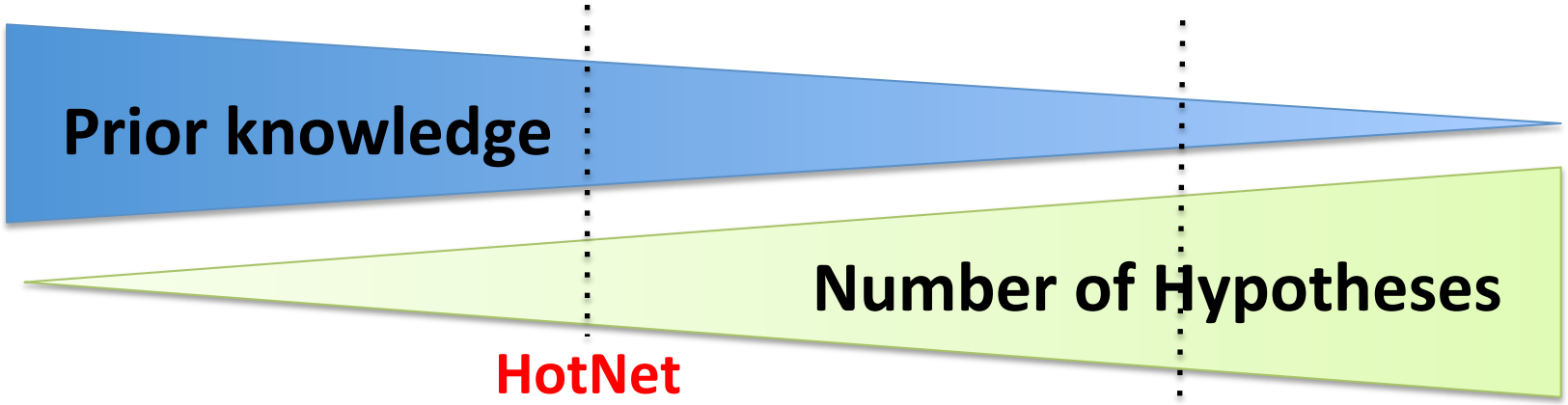




Known pathways

Interaction Network

None



HotNet

subnetworks of interaction network

Dendrix

Exclusive gene sets

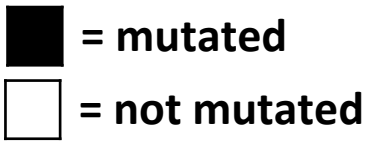
HotNet: Problem Definition

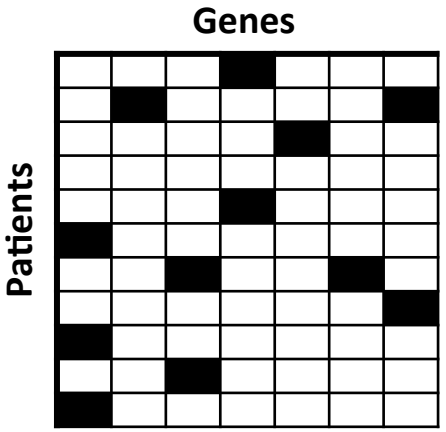
Given:

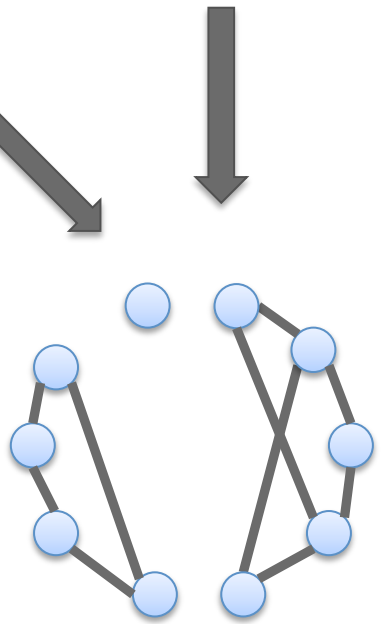
1. Network $G = (V, E)$

   V = genes. E = interactions b/w genes

2. Binary mutation matrix



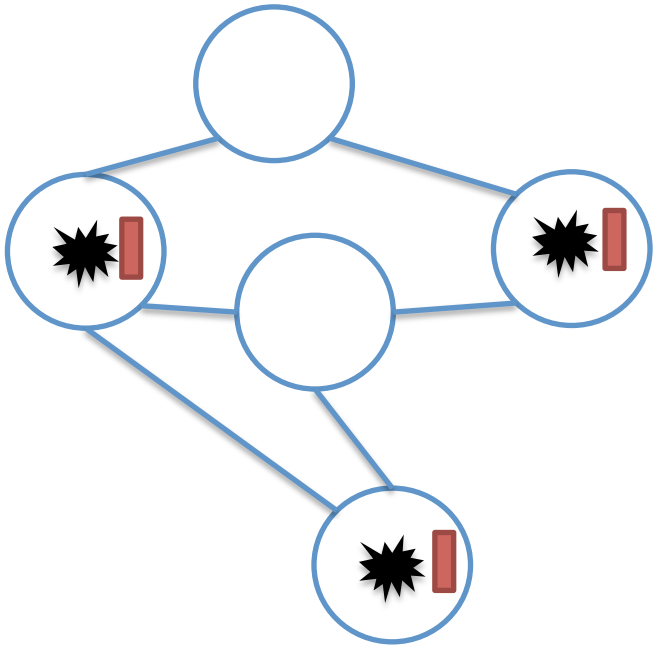




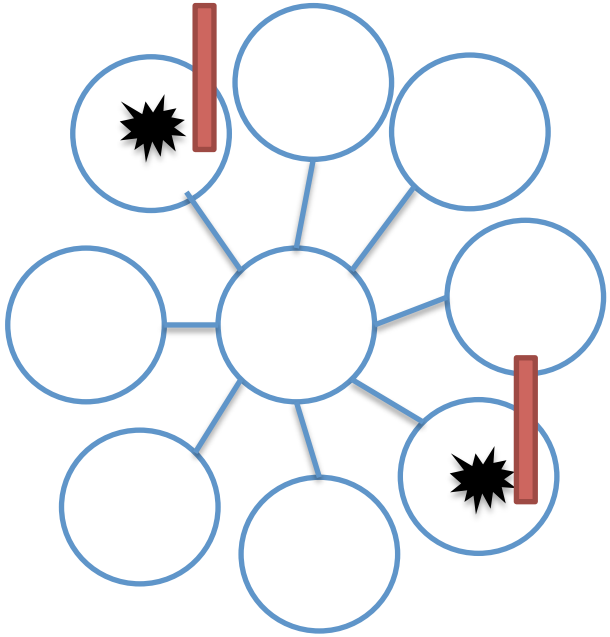
Find: *Connected subnetworks* mutated in a **significant** number of patients.

Subnetwork Properties

Mutation frequency/score AND network topology

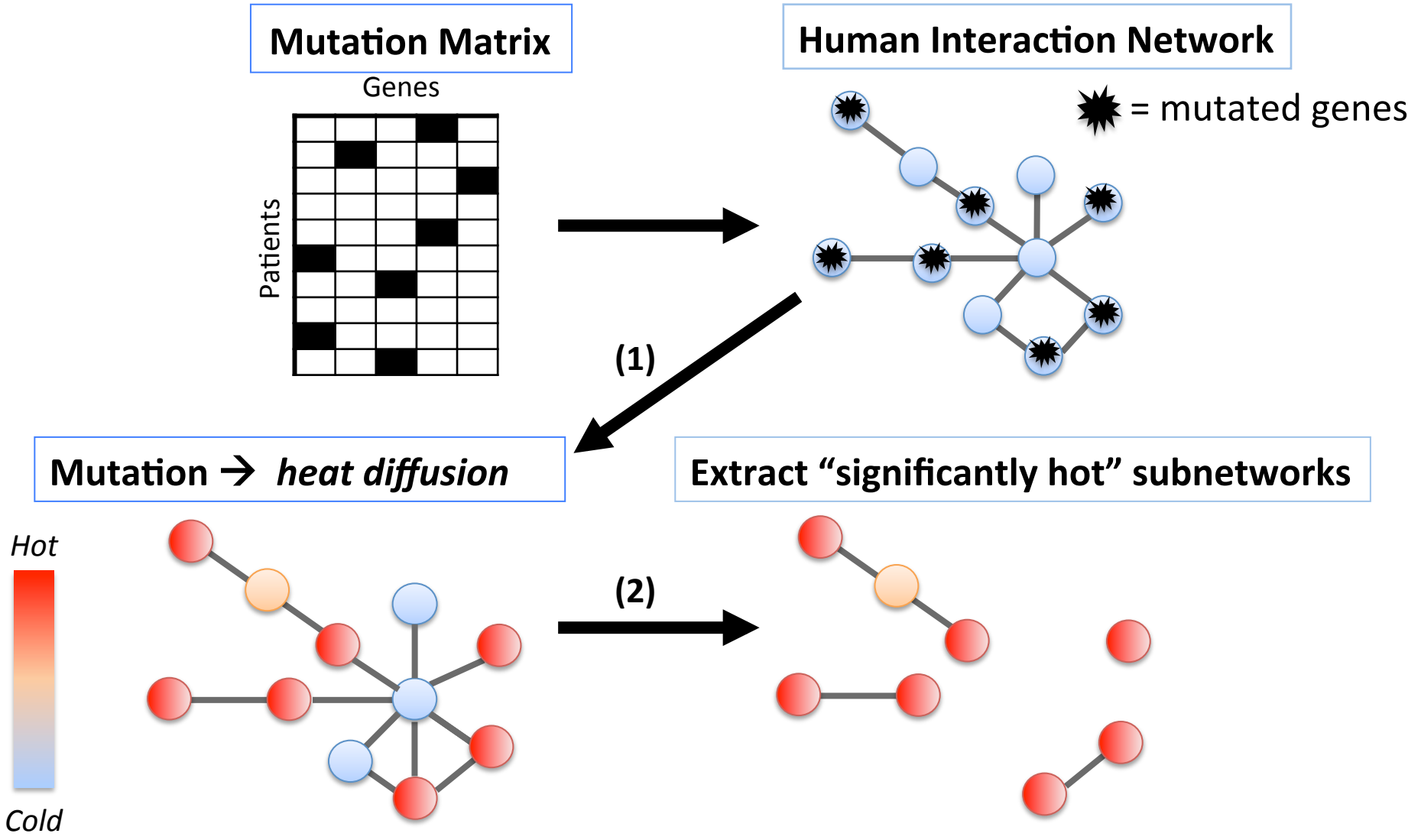


- Moderate frequency/score
- Highly connected



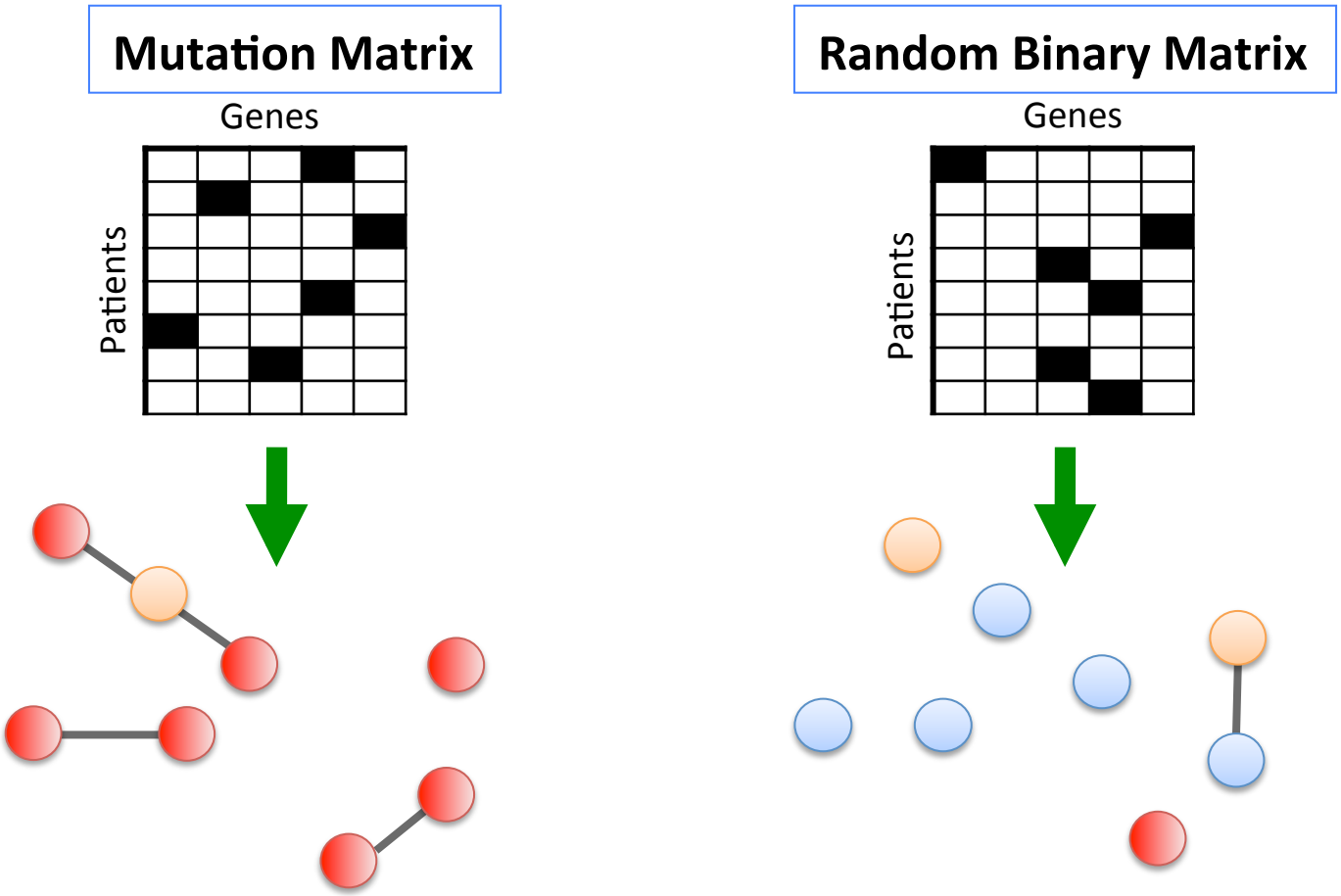
- High frequency/score
- Connected through high-degree node.

Mutated subnetworks: *HotNet**



*F. Vandin, E. Upfal, and B. J. Raphael. *J. Comp.Biol*. (2011). Also *RECOMB* (2010).

Statistical Test



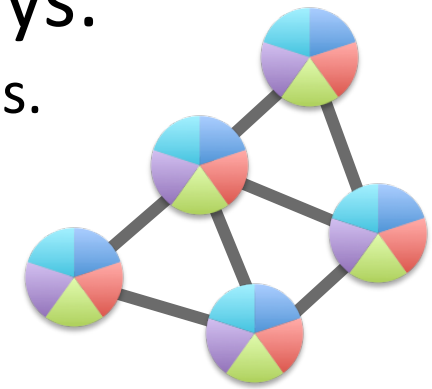
X_s = ***number*** of subnetworks $\geq s$ genes

Two-stage multi-hypothesis test: Rigorously bound FDR.

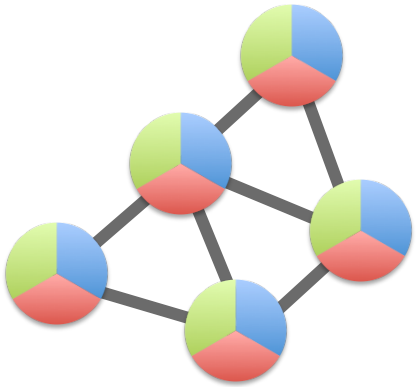
HotNet PanCancer: Goals

Find subnetworks/pathways:

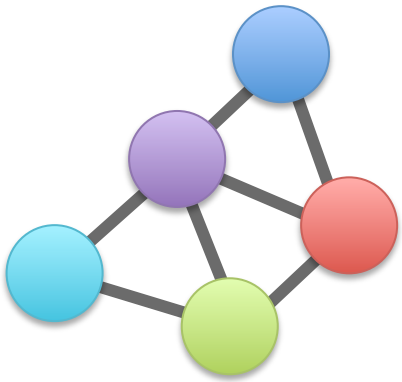
1. Mutated *across* all cancer types.



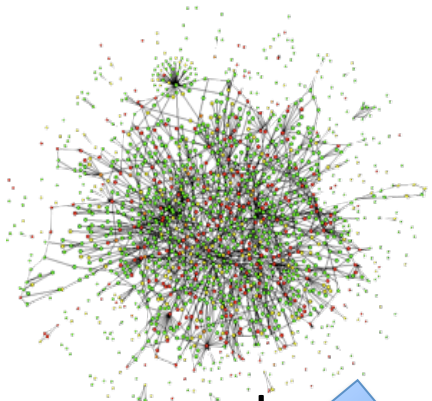
2. Mutated in a subset of cancer types.



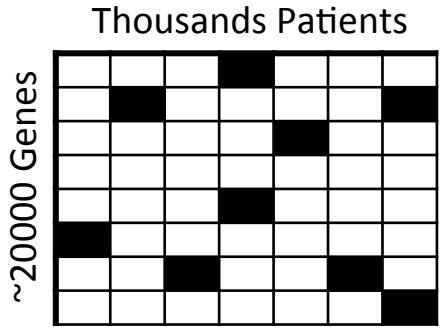
3. Mutated *across* all/many cancer types, but with type-specific mutations.



HotNet PanCancer

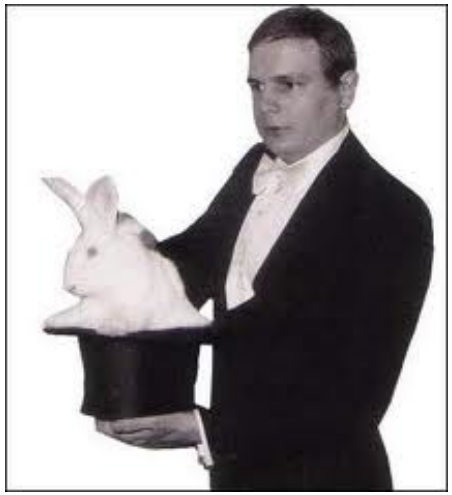


Interaction network
> 10K proteins
> 200K interactions



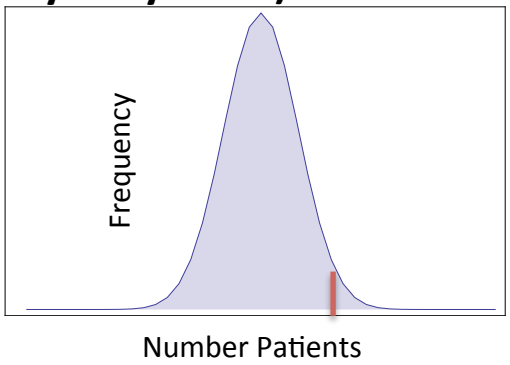
Mutation &
copy number

HotNet



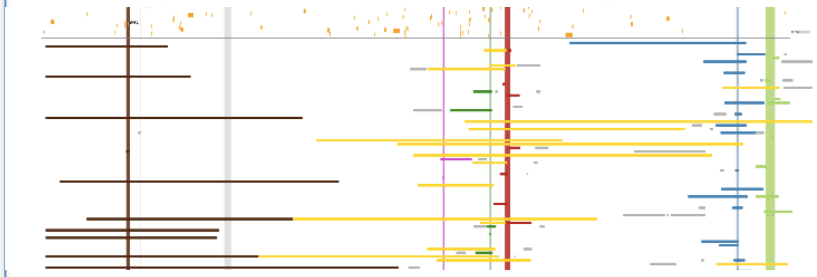
HotNet PanCancer

SNVs (nonsynonymous) and small indels

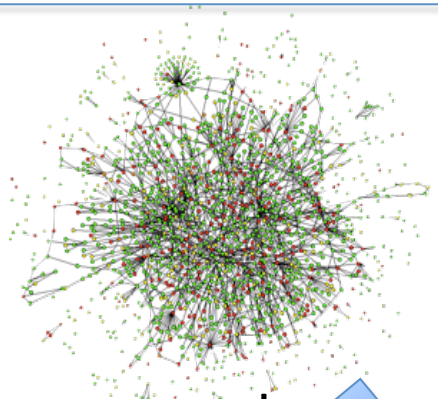


≥ 0.8% frequency
or significant in ≥ 1 cancer type

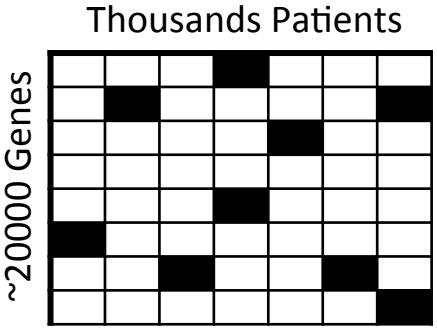
Copy number aberrations



Target gene selection: GISTIC max peaks



Interaction network
> 10K proteins
> 200K interactions

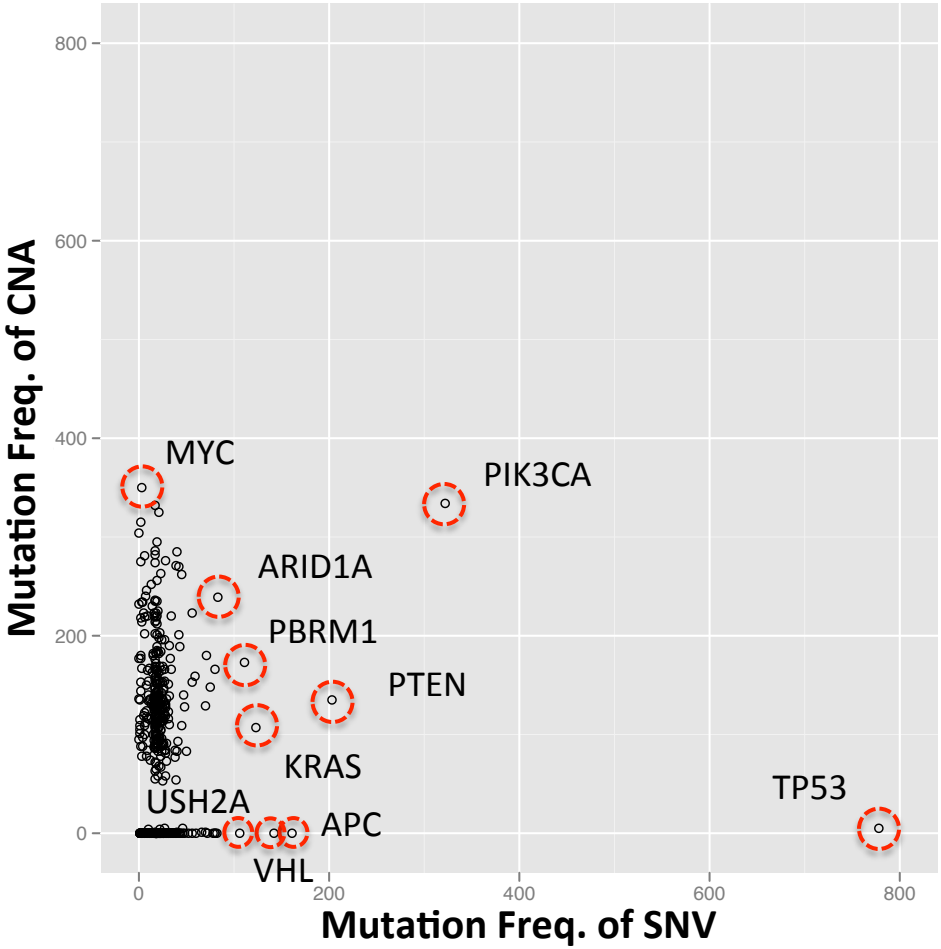


Mutation &
copy number

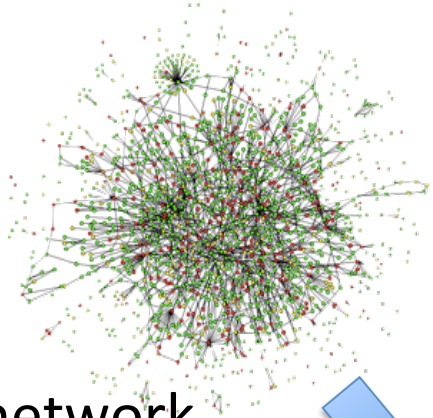
Mutation Data Summary

Total: 765 genes, 1984 samples

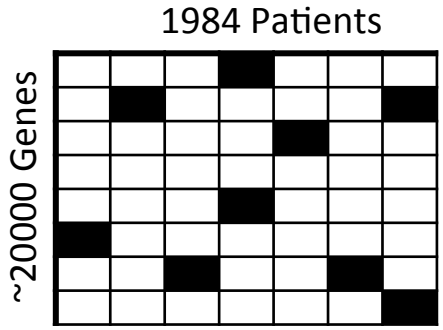
Cancer (sub)type		Num samples
AML		200
COADREAD		185
GBM		261
Basal-like	BRCA	87
HER2	BRCA	61
LuminalA	BRCA	215
LuminalB	BRCA	119
KIRC		292
LUSC		86
OV		311
UCEC		166



HotNet: iRef Network

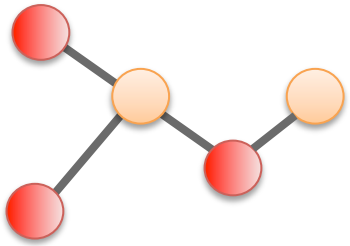


iRef 9.0 network
212,746 interaction
14,384 proteins



Mutation &
copy number

HotNet



11 subnetworks with ≥ 3 genes ($P < 0.01$)

iRef Summary

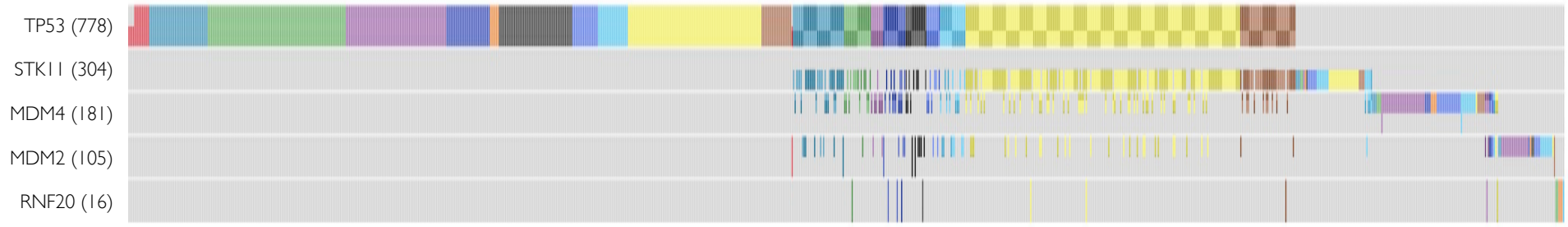
Subnetwork	Coverage
TP53, MDM4, MDM2, STK11, RNF20	48.6% (964 / 1984 samples)
PIK3CA, KRAS, NRAS, IRS2	45.6% (904 / 1984 samples)
EGFR, ERBB2, ERBB3, INSR, KIT, NTRK1, PDGFRB, CBLB	44.3% (878 / 1984 samples)
RB1, CDKN2A, CDK4, CCND1, CEBPA,, DNMT1	40.1% (796 / 1984 samples)
BRIP1, ATR, TP53BP1, BLM, BRCA1, FANCA, FANCM	32.3% (641 / 1984 samples)
PTEN, PIK3R1, PDGFRA, SOS1	32.1% (636 / 1984 samples)
BAP1, ASXL1, ASXL2, ANKHD1-EIF4EBP3, HCFC1, ANKRD17,	30.8% (611 / 1984 samples)
CTNNB1, MYO7A, KDR, CDH1	29.4% (584 / 1984 samples)
Cohesin	28.8% (571 / 1984 samples)
MYC, SETDB1, DNMT3A	28.7% (569 / 1984 samples)
ARID1A, ARID2, BAZ1B	13.2% (261 / 1984 samples)

TP53, STK11, MDM4, *et al.*

Mutations Types

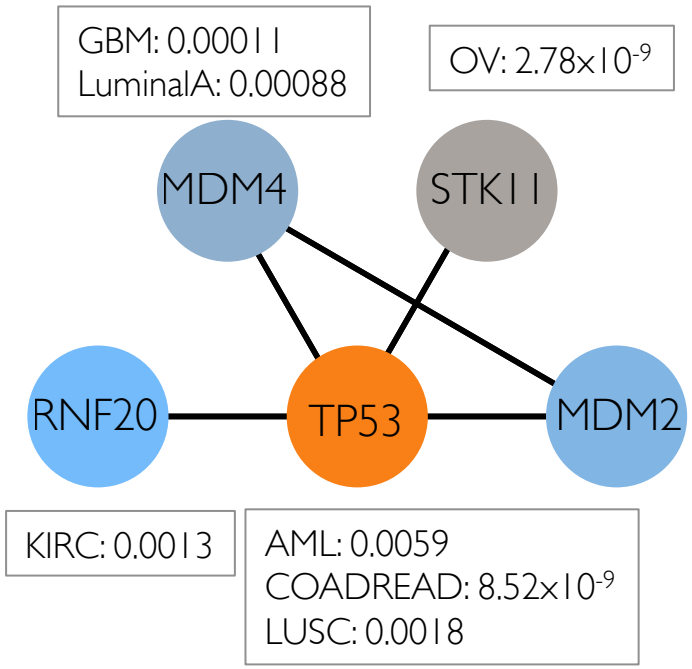
SNV AMP DEL

Coverage: 48.2% (957/1984 samples)



Mutations / Gene

778
741
704
667
630
593
556
519
482
445
408
370
333
296
259
222
185
148
111
74
37



Mutations by Cancer

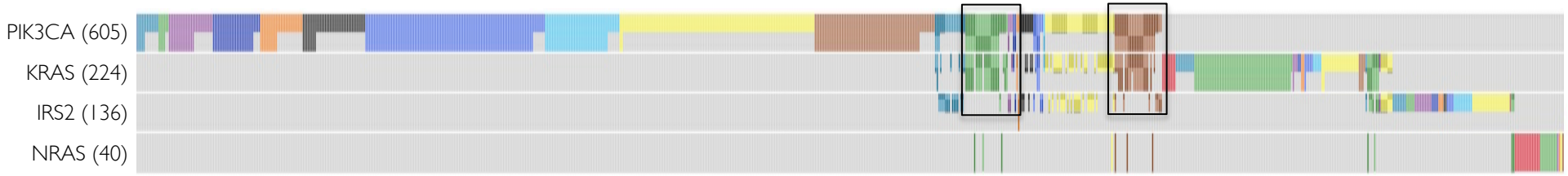
- AML (7.5%)
- Luminal A (25.1%)
- Luminal B (58.8%)
- **Basal-like (931.%) (p = 5.73x10^{-19})**
- HER2-enriched (82%)
- **COADREAD (63.2%) (p = 0.0001)**
- GBM (48.3%)
- KIRC (5.1%)
- **OV (95.2%) (p = 1.18x10^{-83})**
- UCEC (40.4%)
- **LUSC (76.7%) (p = 3.88x10^{-7})**

PIK3CA and RAS interactions

Mutations Types

SNV AMP DEL

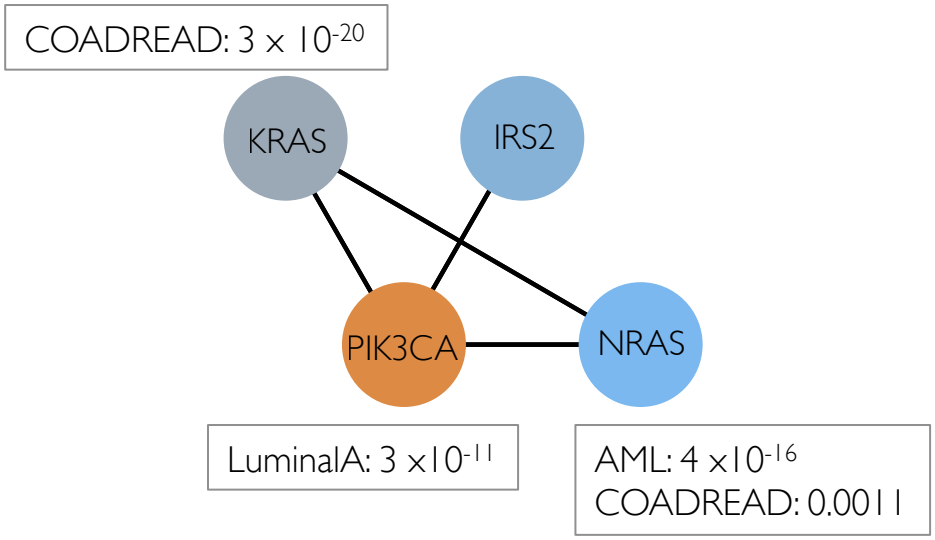
Coverage: 42.4% (842/1984 samples)



D.J. Burgess.
Nature Reviews Cancer **11**, 389 (June 2011)

Mutations / Gene

778
741
704
667
630
593
556
519
482
445
408
370
333
296
259
222
185
148
11
74
37

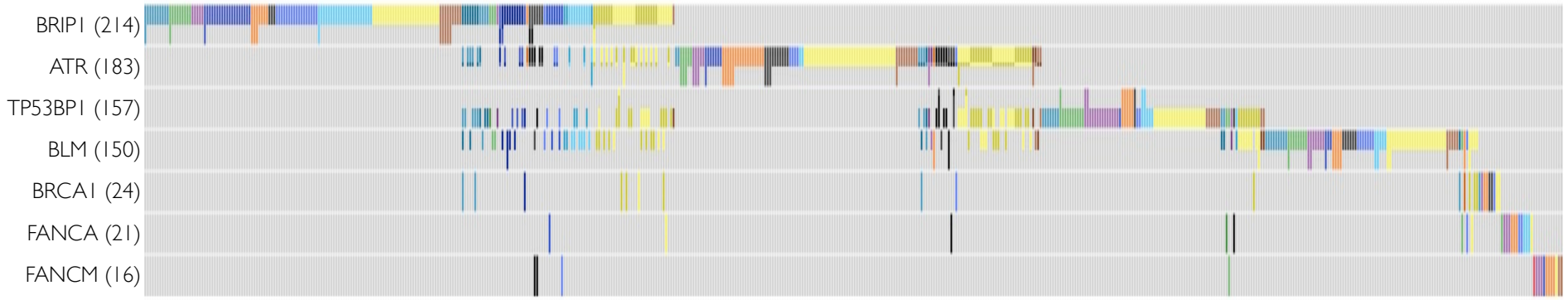


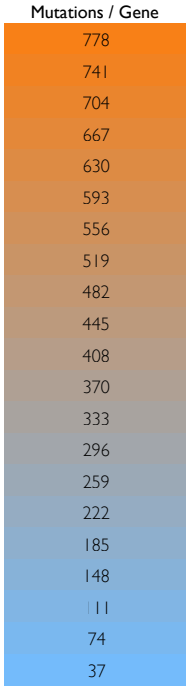
Mutations by Cancer

- AML (11.5%)
- **Luminal A (56.3%) (p = 0.0001)**
- Luminal B (51.3%)
- Basal-like (57.5%)
- HER2-enriched (59%)
- **COADREAD (61.1%) (p=7.91x10^{-7})**
- GBM (16.9%)
- KIRC (11.3%)
- **OV (66.9%) (p = 3.12x10^{-20})**
- **UCEC (63.3%) (p = 1.68x10^{-7})**
- LUSC (54.7%)

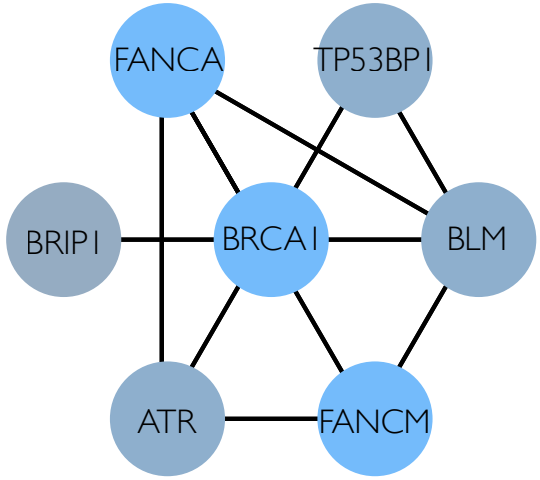
DNA Repair

Coverage: 28.8% (572/1984 samples)





HER2-enriched: 0.00078
LuminalB: 0.00084



Mutations by Cancer

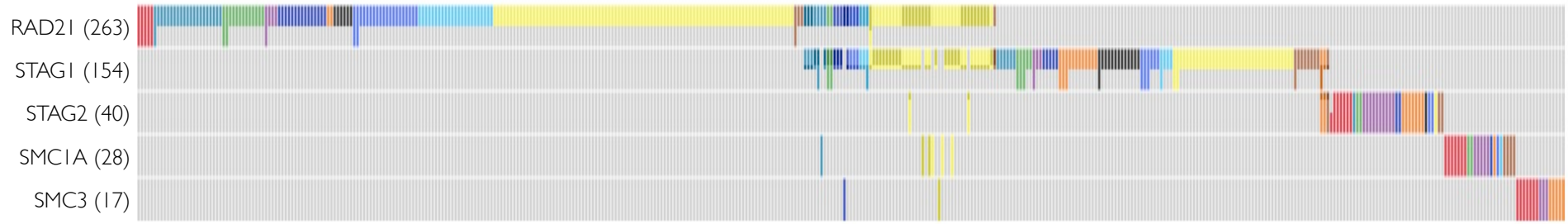
- AML (0.5%)
- Luminal A (20%)
- Luminal B (42%)
- **Basal-like (54.%) (p = 4.91x10^{-6})**
- **HER2-enriched (68.9%) (p = 6.60x10^{-10})**
- COADREAD (21.1%)
- GBM (16.1%)
- KIRC (15.4%)
- **OV (60.5%) (p = 2.75x10^{-36})**
- UCEC (22.9%)
- LUSC (43%)

Cohesin Complex

Mutations Types

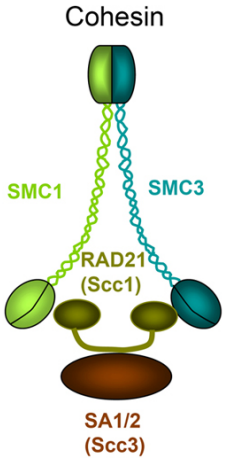
SNV AMP DEL

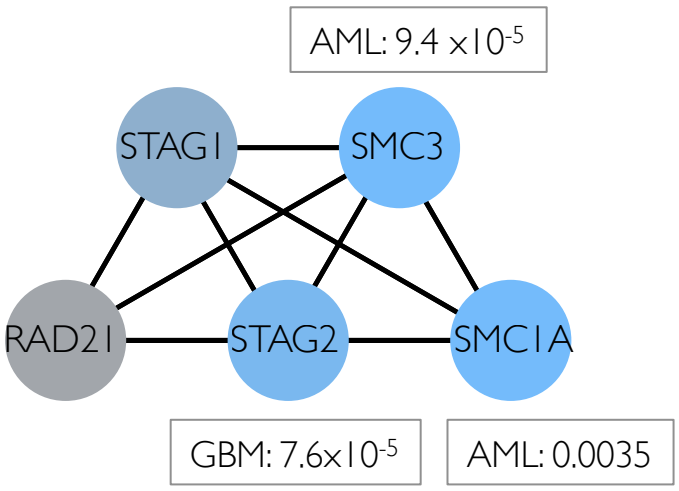
Coverage: 22% (437/1984 samples)



Mutations / Gene

778
741
704
667
630
593
556
519
482
445
408
370
333
296
259
222
185
148
111
74
37

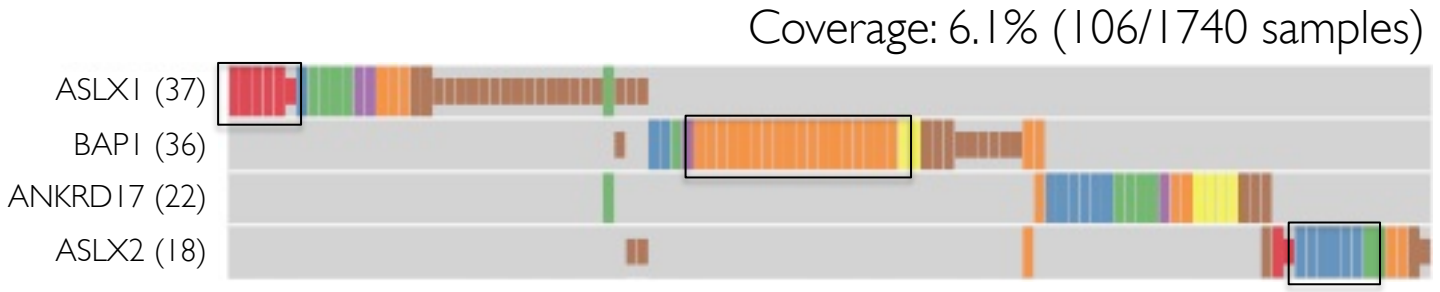




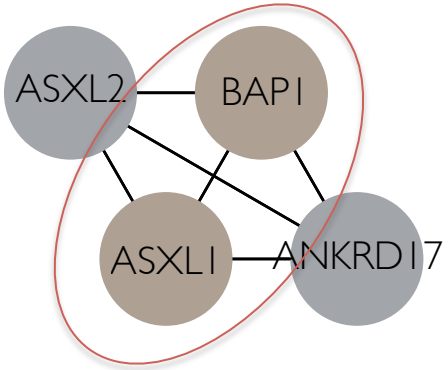
Mutations by Cancer

- AML (13%)
- Luminal A (14.9%)
- Luminal B (26.1%)
- **Basal-like (40.2%) (p = 0.0008)**
- **HER2-enriched (45.9%) (p = 0.0002)**
- COADREAD (13%)
- GBM (9.6%)
- KIRC (9.9%)
- OV (54%)
- UCEC (11.4%)
- LUSC (23.3%)

Polycomb group proteins



PR-DUB complex



Cancer types

- AML (8)
- BRCA (15)
- COADREAD (12)
- GBM (4)
- KIRC (27)
- OV (6)
- UCEC (34)

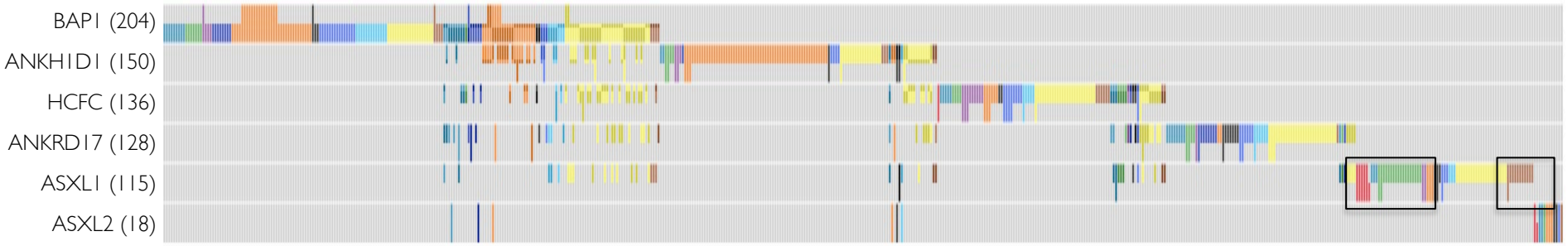
Mutations / Gene

1 2 3 4 5 7 9 12 17 23 32 44 60 82 112 153 210 287 393 538 737

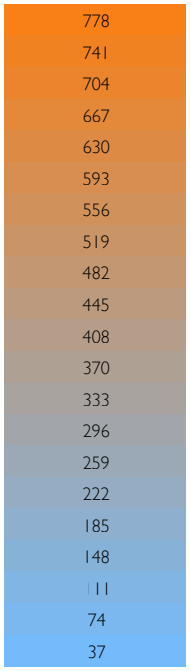
Mutation type

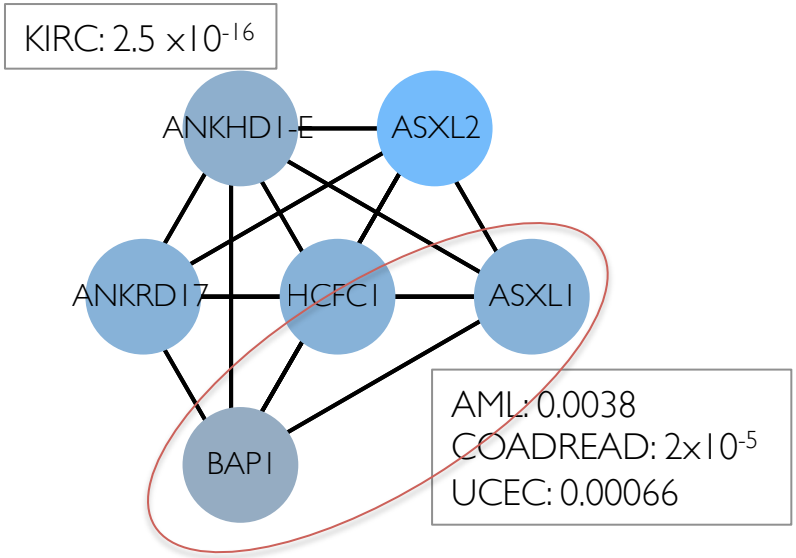
SNV CNA

Polycomb Group Proteins



Coverage: 29% (575/1984 samples)



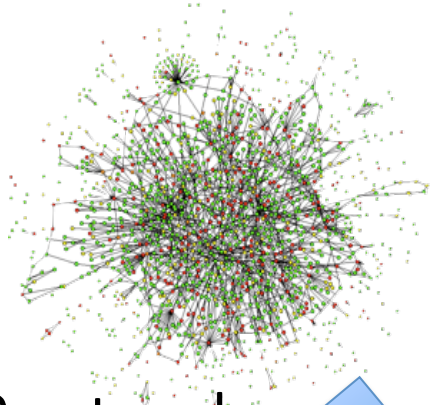


PR-DUB complex

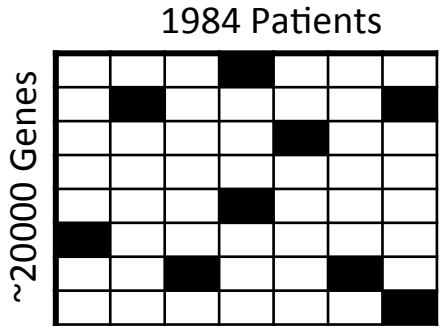
Mutations by Cancer

- AML (4.5%)
- Luminal A (19.1%)
- Luminal B (29.4%)
- **Basal-like (49.4%) (p = 0.0003)**
- HER2-enriched (41%)
- COADREAD (23.8%)
- GBM (7.7%)
- **KIRC (44.5%) ($p = 7.49 \times 10^{-9}$)**
- **OV (54.7%) ($p = 2.50 \times 10^{-24}$)**
- UCEC (20.5%)
- LUSC (27.9%)

HotNet: HPRD Network

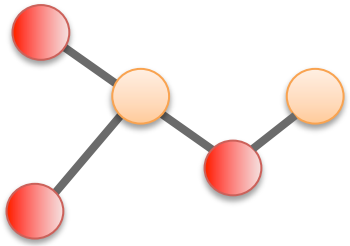


HPRD 9.0 network
39,173 interaction
9,617 proteins



Mutation &
copy number

HotNet



20 subnetworks with ≥ 3 genes ($P < 0.01$)

HPRD Summary

Subnetwork	Coverage
PIK3CA and neighbors	64.8% (1286 / 1984 samples)
ARID1A, PBRM1, and neighbors	64.6% (1281 / 1984 samples)
Cohesin and neighbors	47.6% (945 / 1984 samples)
C3, THBS1, CFH, COL3A1, ITGAX, LAMA1	30.7% (609 / 1984 samples)
TG, LPA, ANKS1B, LRP2	28.2% (559 / 1984 samples)
AXIN2, RP1, APC, PTPN13	27.0% (535 / 1984 samples)
BRIP1, BAP1, BRCA1	24.3% (483 / 1984 samples)
PTPRS, CDH1, PTPRK, PPFIA2, PTPRM	24.3% (482 / 1984 samples)
MYC, DNMT3A, DNMT1	22.8% (453 / 1984 samples)
ANK1, L1CAM, NCAN, TNC	21.9% (435 / 1984 samples)
CCNE1, NOTCH3, NOTCH4, FBXW7	20.5% (406 / 1984 samples)
SLIT3, ROBO2, ROBO1, SLIT2, SRGAP1	17.9% (355 / 1984 samples)
MCL1, BCL2L1, BCLAF1	16.8% (333 / 1984 samples)
NTRK1, NTRK3, RASGRF1	16.3% (324 / 1984 samples)
MYO7A, NFE2L2, KEAP1	16.1% (319 / 1984 samples)
MLLT4, EPHA7, NRXN3	15.2% (302 / 1984 samples)
COL4A1, COL4A5, COL4A6, FBLN2	11.8% (234 / 1984 samples)
OPRM1, GRM1, GPRASP1	10.8% (215 / 1984 samples)
CDH18, CDH12, CDH10, CDH9	10.6% (211 / 1984 samples)
FLNA, CALCR, GRM8, CASR	10.1% (201 / 1984 samples)

HPRD Summary

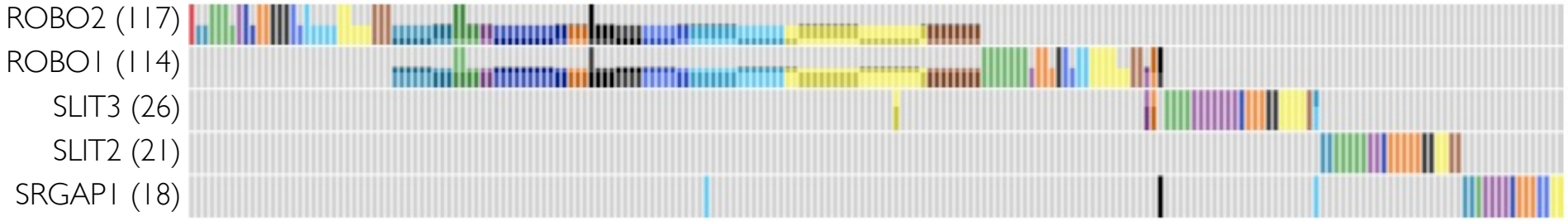
Subnetwork	Coverage	
PIK3CA and neighbors	64.8% (1286 / 1984 samples)	
AR1D1A, PBRM1, and neighbors	64.6% (1281 / 1984 samples)	KIRC
Cohesin and neighbors	47.6% (945 / 1984 samples)	
C3, THBS1, CFH, COL3A1, ITGAX, LAMA1	30.7% (609 / 1984 samples)	
TG, LPA, ANKS1B, LRP2	28.2% (559 / 1984 samples)	
AXIN2, RP1, APC, PTPN13	27.0% (535 / 1984 samples)	
BRIP1, BAP1, BRCA1	24.3% (483 / 1984 samples)	
PTPRS, CDH1, PTPRK, PPFIA2, PTPRM	24.3% (482 / 1984 samples)	
MYC, DNMT3A, DNMT1	22.8% (453 / 1984 samples)	
ANK1, L1CAM, NCAN, TNC	21.9% (435 / 1984 samples)	
CCNE1, NOTCH3, NOTCH4, FBXW7	20.5% (406 / 1984 samples)	OV
SLIT3, ROBO2, ROBO1, SLIT2, SRGAP1	17.9% (355 / 1984 samples)	
MCL1, BCL2L1, BCLAF1	16.8% (333 / 1984 samples)	
NTRK1, NTRK3, RASGRF1	16.3% (324 / 1984 samples)	
MYO7A, NFE2L2, KEAP1	16.1% (319 / 1984 samples)	
MLLT4, EPHA7, NRXN3	15.2% (302 / 1984 samples)	
COL4A1, COL4A5, COL4A6, FBLN2	11.8% (234 / 1984 samples)	
OPRM1, GRM1, GPRASP1	10.8% (215 / 1984 samples)	
CDH18, CDH12, CDH10, CDH9	10.6% (211 / 1984 samples)	
FLNA, CALCR, GRM8, CASR	10.1% (201 / 1984 samples)	

SLIT/ROBO signaling

Mutations Types

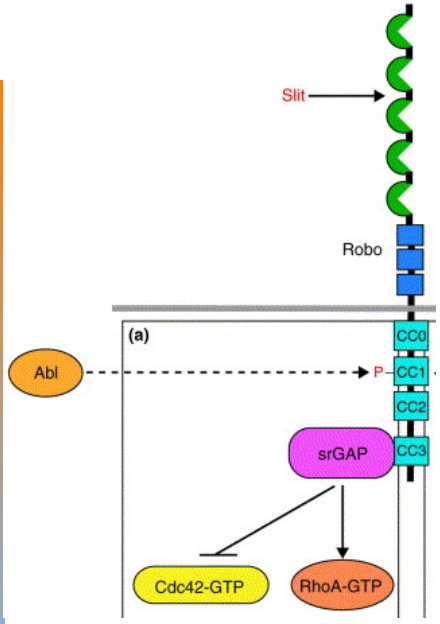
SNV AMP DEL

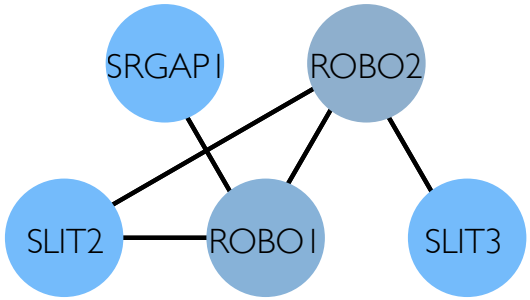
Coverage: 10.3% (203/1984 samples)



Mutations / Gene

605
576
547
519
490
461
432
403
375
346
317
288
259
230
202
173
144
115
86
58
29



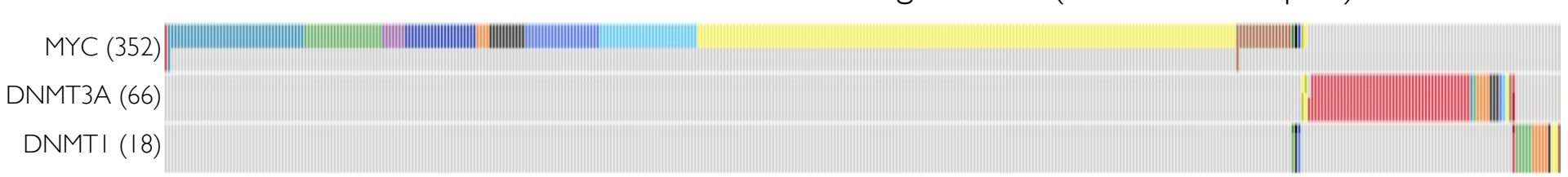


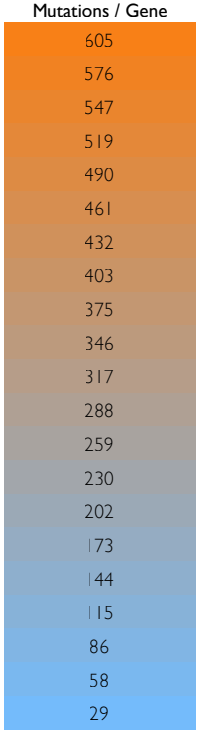
GBM: 0.002

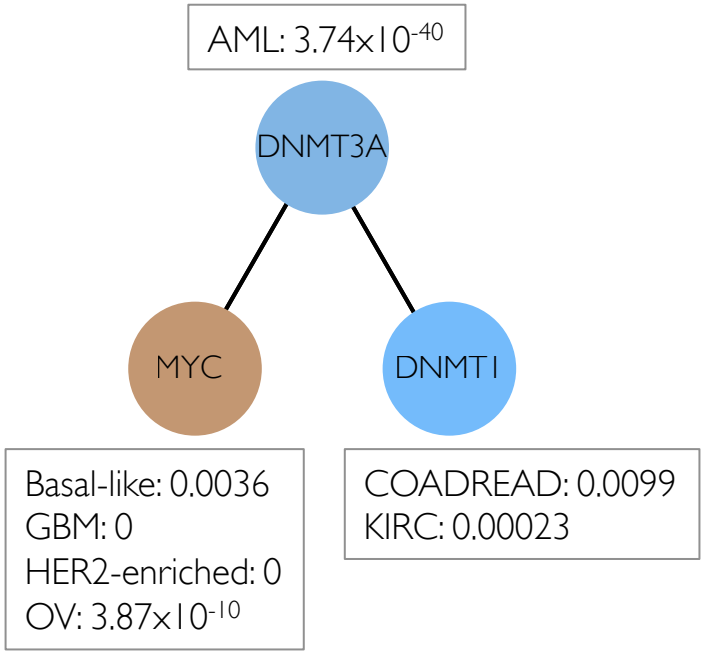
Mutations by Cancer

- AML (0.5%)
- Luminal A (6%)
- Luminal B (18.5%)
- Basal-like (17.2%)
- **HER2-enriched (26.2%) (p = 0.002)**
- COADREAD (13.5%)
- GBM (6.9%)
- KIRC (6.8%)
- OV (12.9%)
- UCEC (9.6%)
- LUSC (19.8%)

MYC, DNMT3A, DNMT1



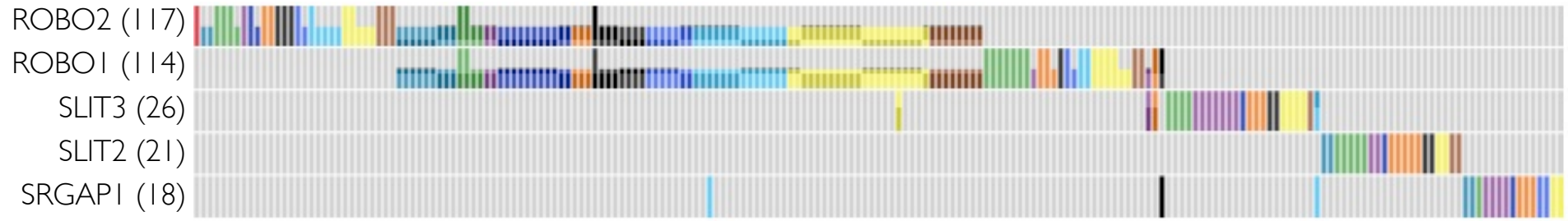




Mutations by Cancer

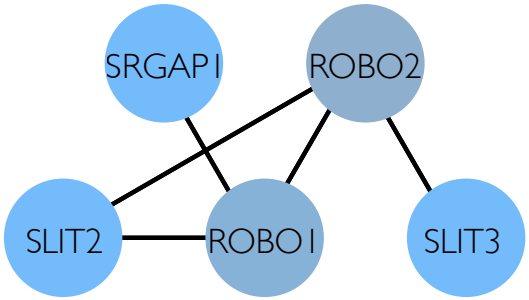
- AML (26%)
- Luminal A (11.6%)
- Luminal B (26.1%)
- Basal-like (49.4%)
- HER2-enriched (36.1%)
- COADREAD (16.8%)
- GBM (2.7%)
- KIRC (4.5%)
- **OV (55%) (p = 3.63x10^{-45})**
- UCEC (11.4%)
- LUSC (18.6%)

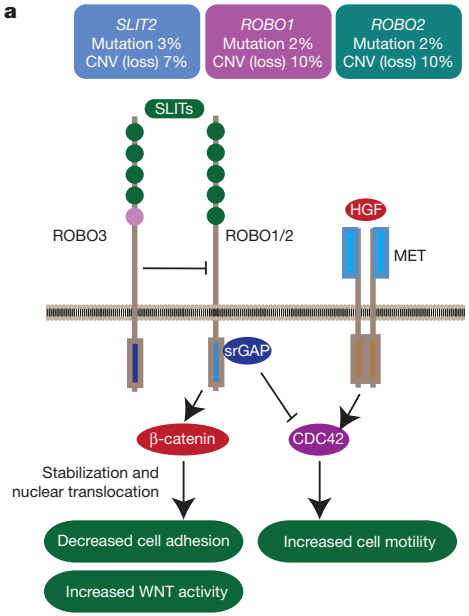
SLIT/ROBO signaling



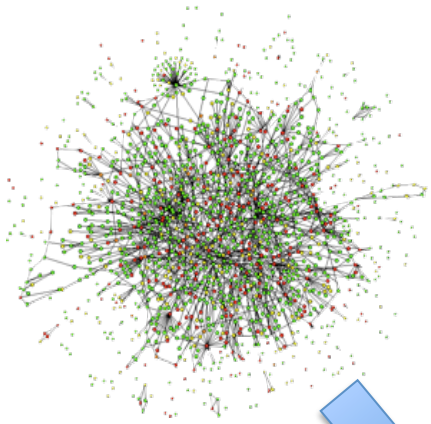
Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes

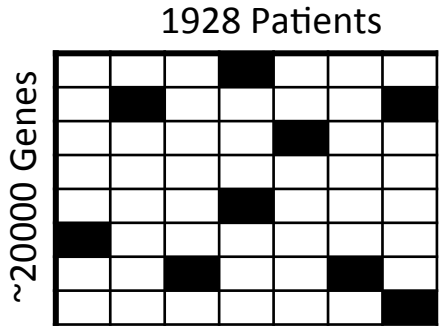
Nature (Nov. 15, 2012) [a]





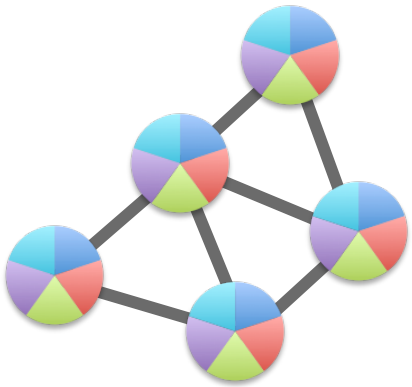
Summary

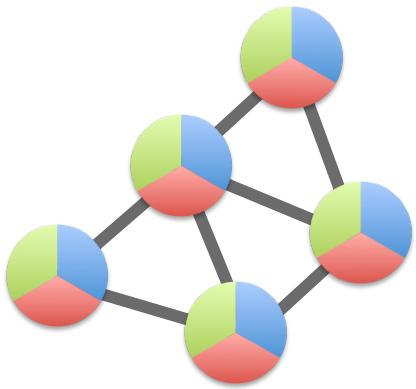


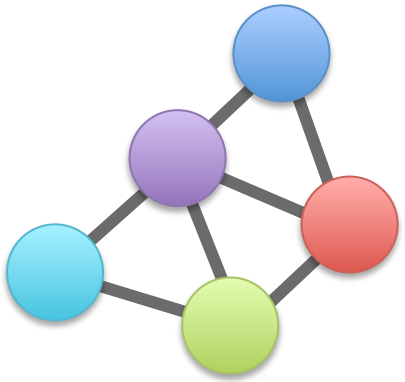


Mutation &
copy number

HotNet







Next Steps

- Additional QC on mutation data
  - Target gene selection on copy number aberrations
- Incorporate background mutation model
- Incorporate other data types: e.g. expression and/or methylation
- Dendrix: exclusivity

Acknowledgements

Fabio Vandin
Eli Upfal
Max Leiserson
Hsin-Ta Wu
Adrien Deschamps
Jason Shum

Tim Ley
Li Ding
Chris Miller
Elaine Mardis
Rick Wilson
and others…

HotNet and Dendrix Available

Web: http://compbio.cs.brown.edu/software