



BROWN



Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes

Nature Genetics 47, 106-114 (2015) doi:10.1038/ng.3168

Max Leiserson*
ISMB
July 13, 2015

Joint work with: F. Vandin*, H-T. Wu, J.R. Dobson, J.V. Eldridge, J.L. Thomas, A. Papoutsaki, Y. Kim, B. Niu, M. McLellan, M.S. Lawrence, A. Gonzalez-Perez, D. Tamborero, Y. Cheng, G.A. Ryslik, N. Lopez-Bigas, G. Getz, L. Ding & B.J. Raphael

* equal contribution

Identifying cancer driver genes

Science

AAAS

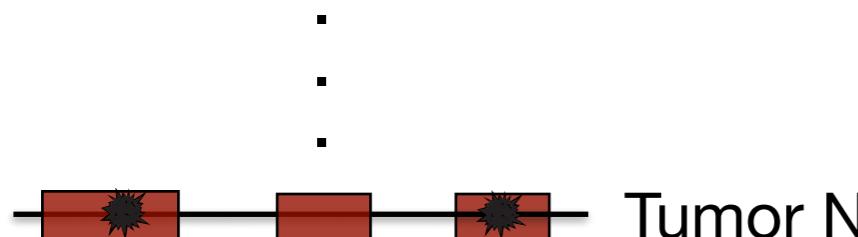
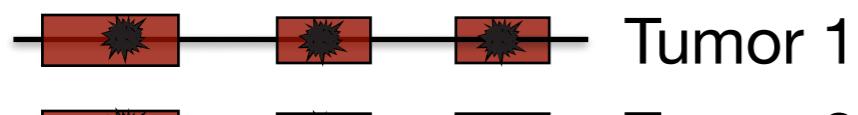
Cancer Genome Landscapes “>99.9% of mutations are passengers”
Vogelstein *et al.* (2013) “3-8 drivers per tumor”

Identifying cancer driver genes



Cancer Genome Landscapes “>99.9% of mutations are passengers”
Vogelstein *et al.* (2013) “3-8 drivers per tumor”

Compare variation across tumors



■ = gene ⚡ = mutation

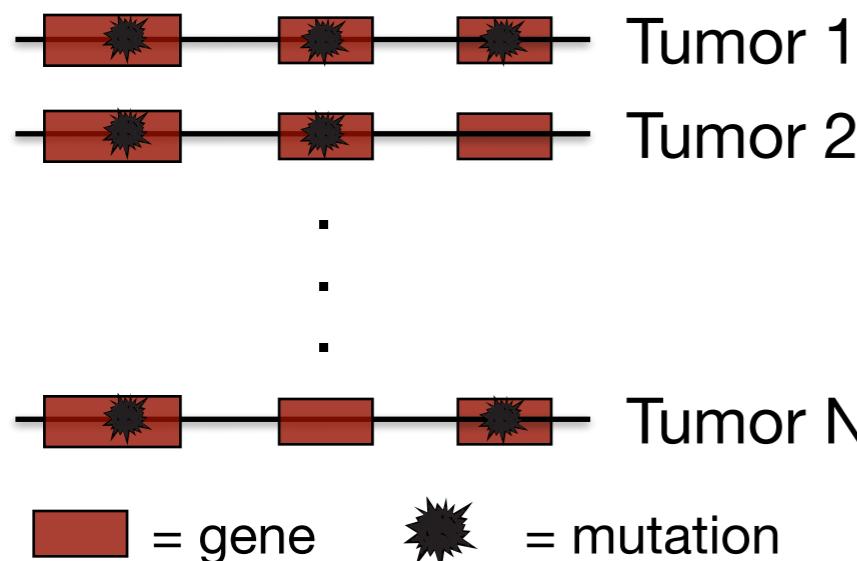
- Single nucleotide variants
- Copy number aberrations
- Gene expression
-

Identifying cancer driver genes

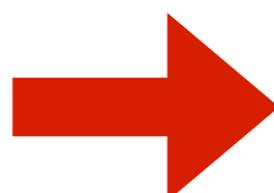


Cancer Genome Landscapes “>99.9% of mutations are passengers”
Vogelstein *et al.* (2013) “3-8 drivers per tumor”

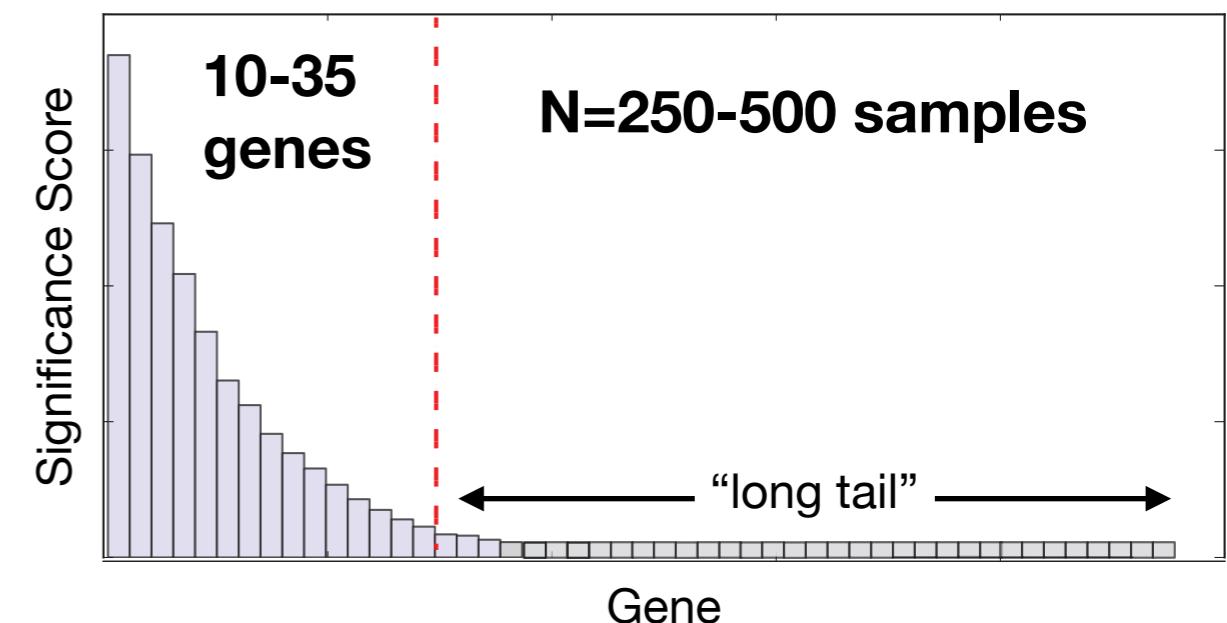
Compare variation across tumors



- Single nucleotide variants
- Copy number aberrations
- Gene expression
- ...



Identify cancer driver genes

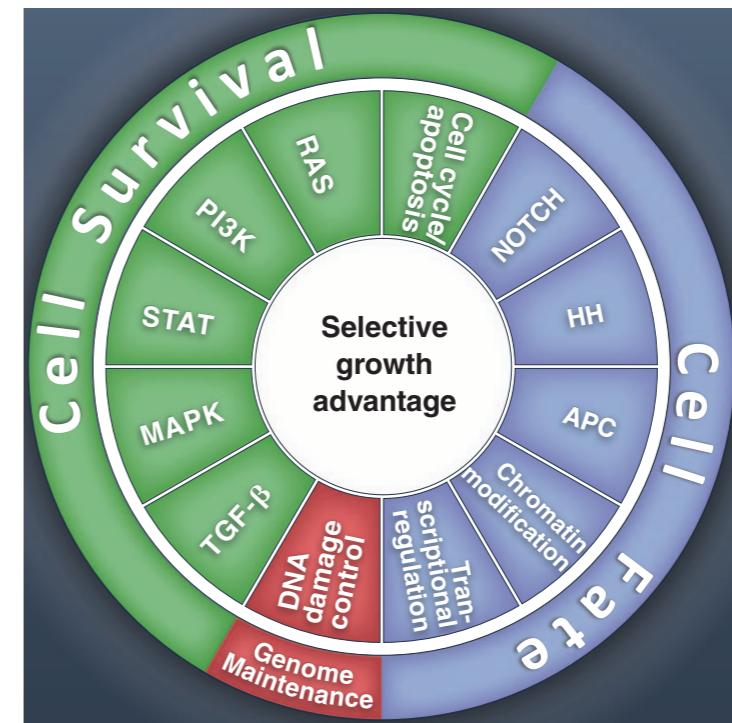


Mutations weighted by:

- Recurrence
- Gene length
- Mutation context
- Expression level
- Replication timing
- ...

Cancer driver mutations target pathways

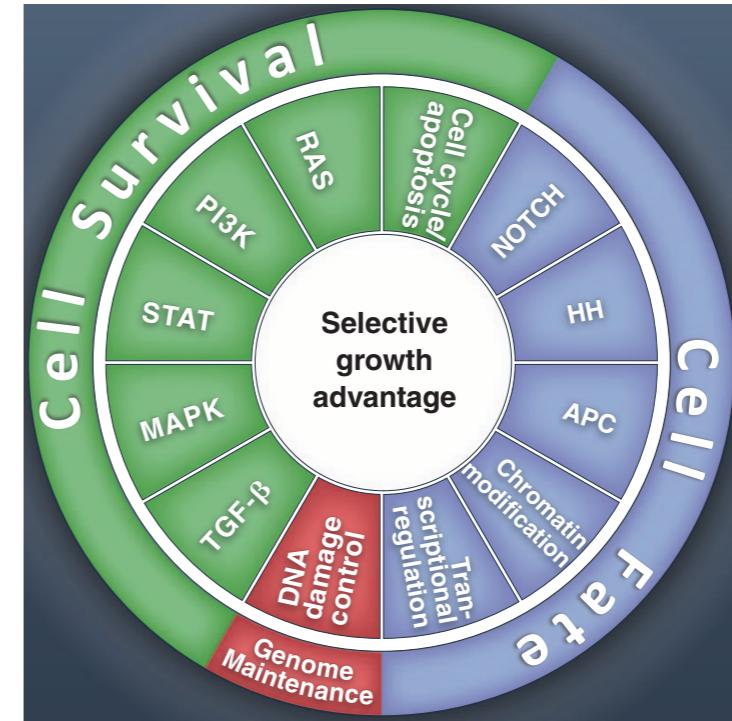
**Driver mutations confer a growth advantage to the tumor
→ driver genes are members of cancer signaling pathways**



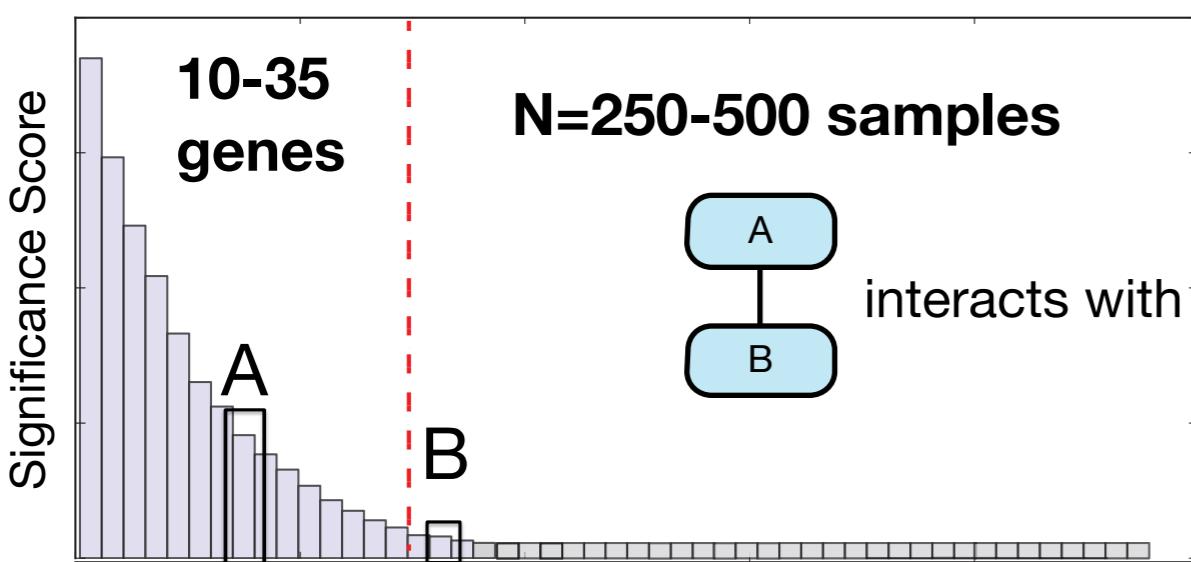
Vogelstein et al.
(Science 2013)

Cancer driver mutations target pathways

**Driver mutations confer a growth advantage to the tumor
→ driver genes are members of cancer signaling pathways**

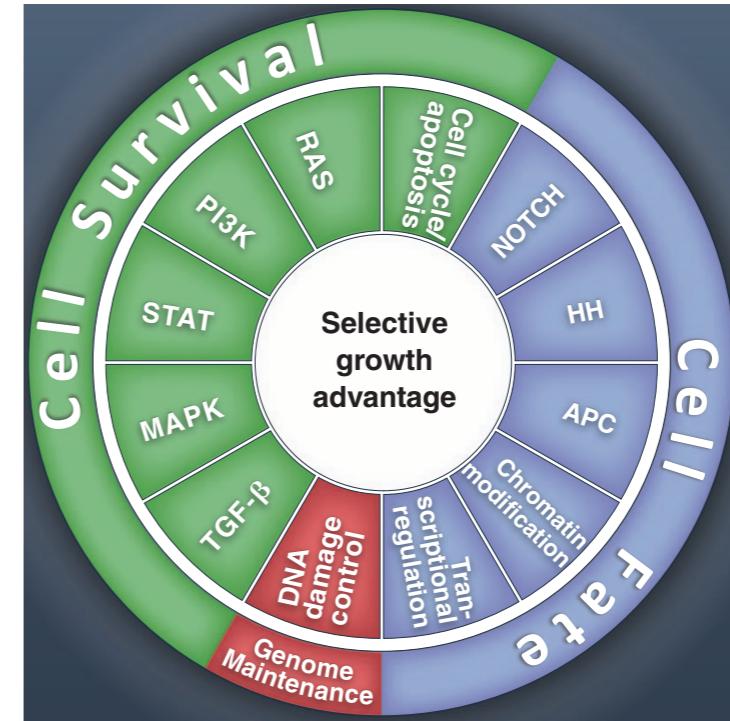


Vogelstein et al.
(Science 2013)

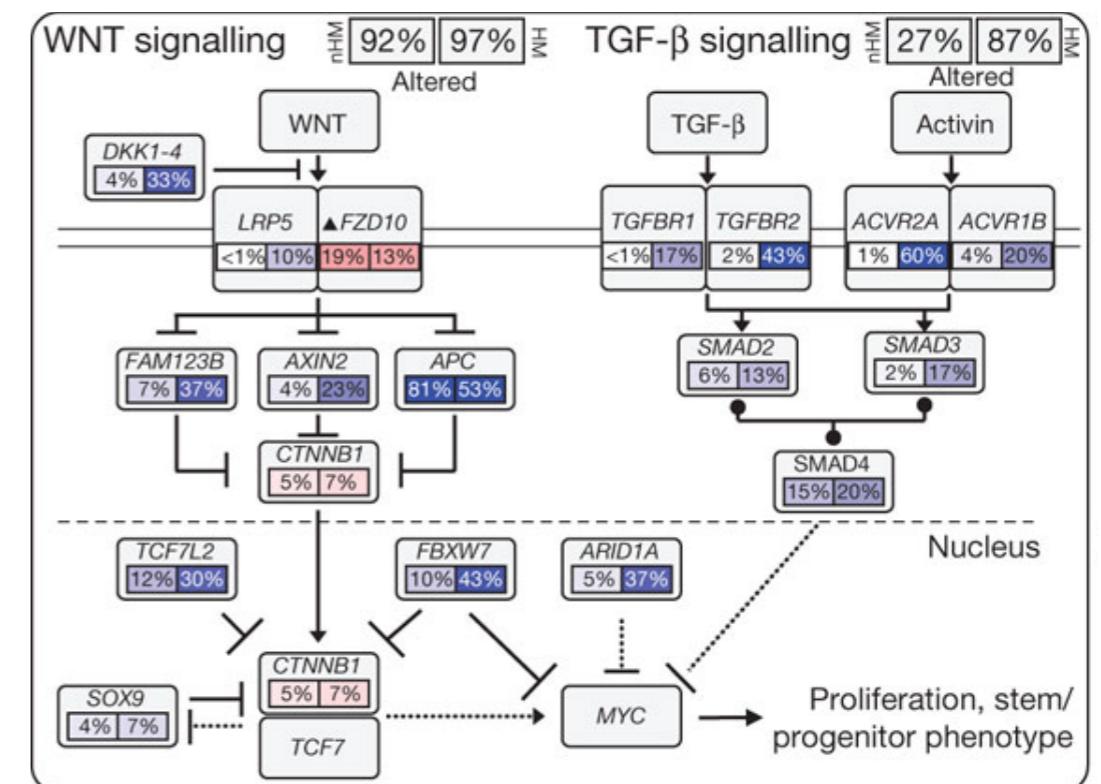
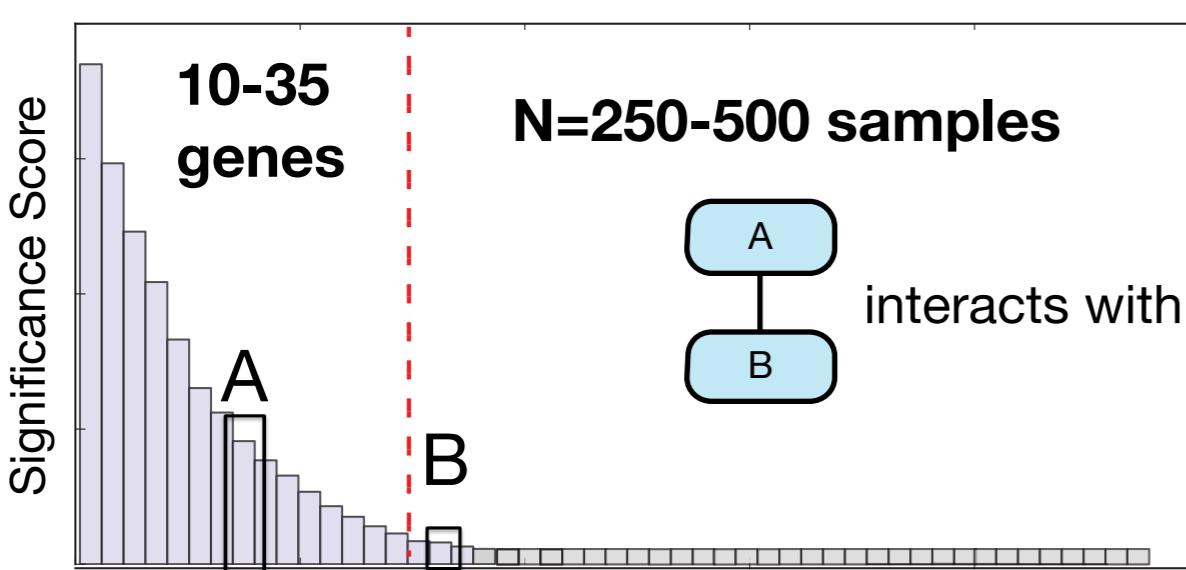


Cancer driver mutations target pathways

Driver mutations confer a growth advantage to the tumor
→ driver genes are members of cancer signaling pathways



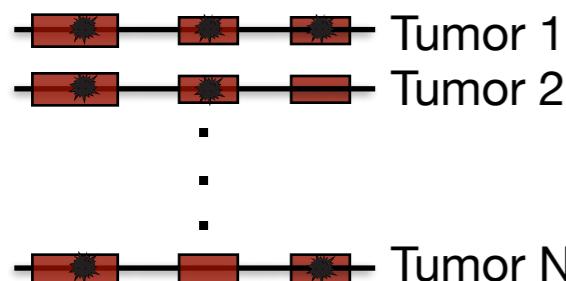
Vogelstein et al.
(Science 2013)



Testing known gene sets and pathways

Input data

Mutation data

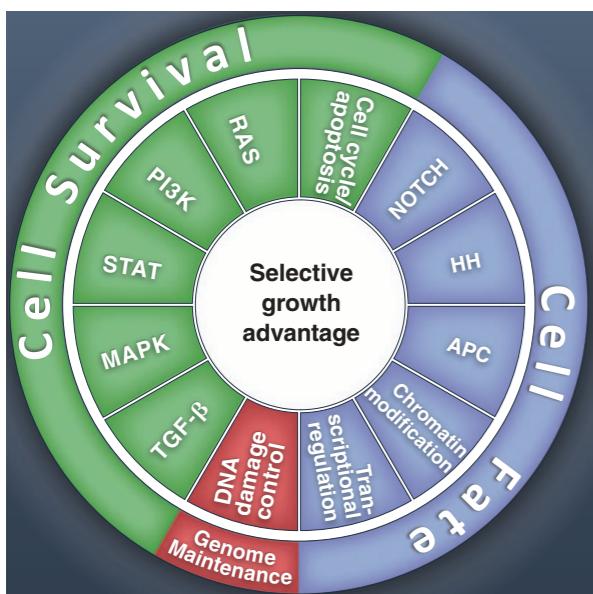


Enrichment tests

GSEA [1,2]
DAVID [3,4]

(e.g. most mutated genes:
EGFR, KRAS, BRAF)

Gene set database



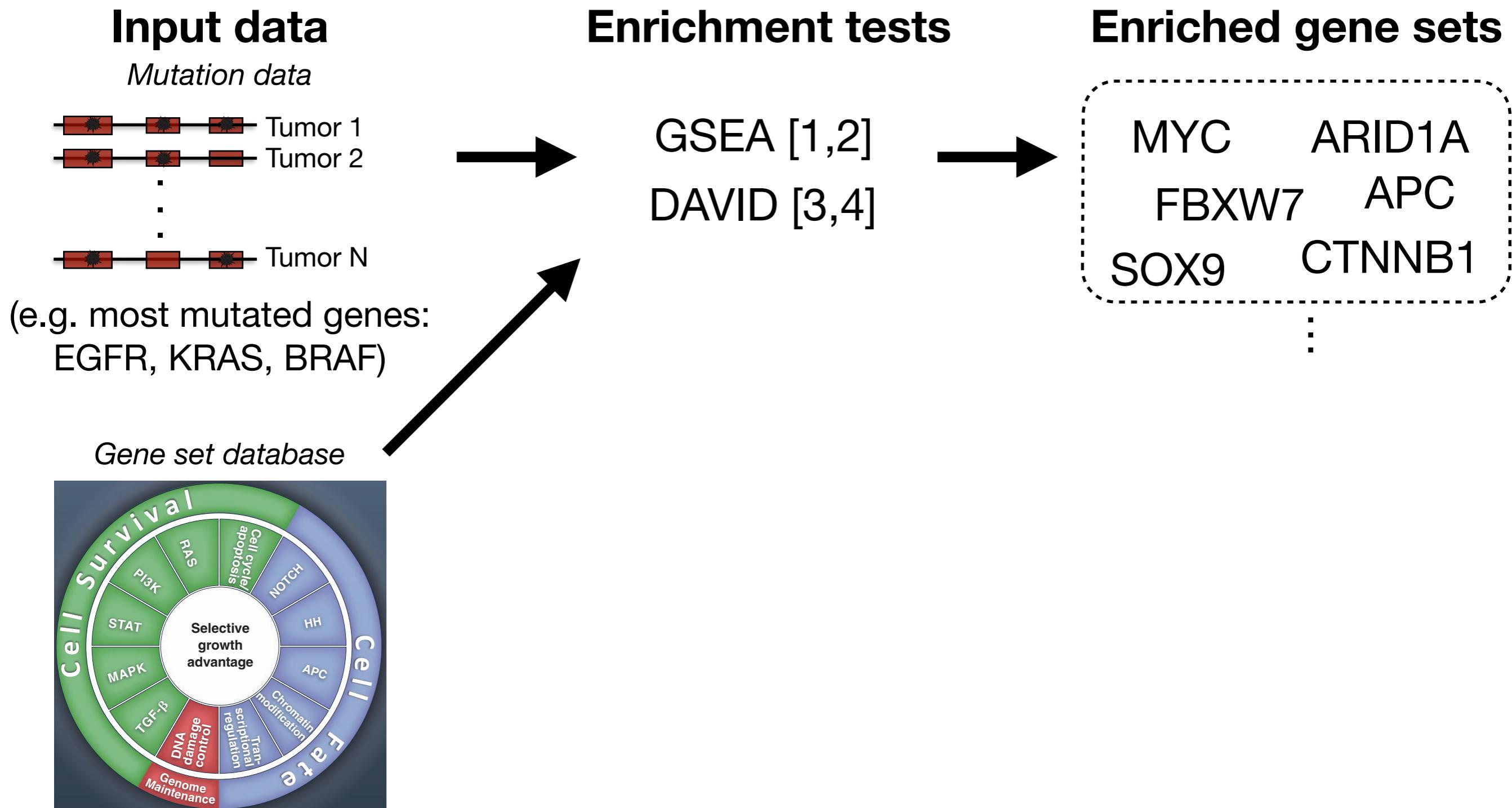
[1] Mootha *et al.* *Nat. Genet.* (2003).

[2] Subramanian *et al.* *PNAS* (2005).

[3] Huang *et al.* *Nat. Protoc.* (2009).

[4] Huang *et al.* *Nucleic Acids Res.* (2009)

Testing known gene sets and pathways



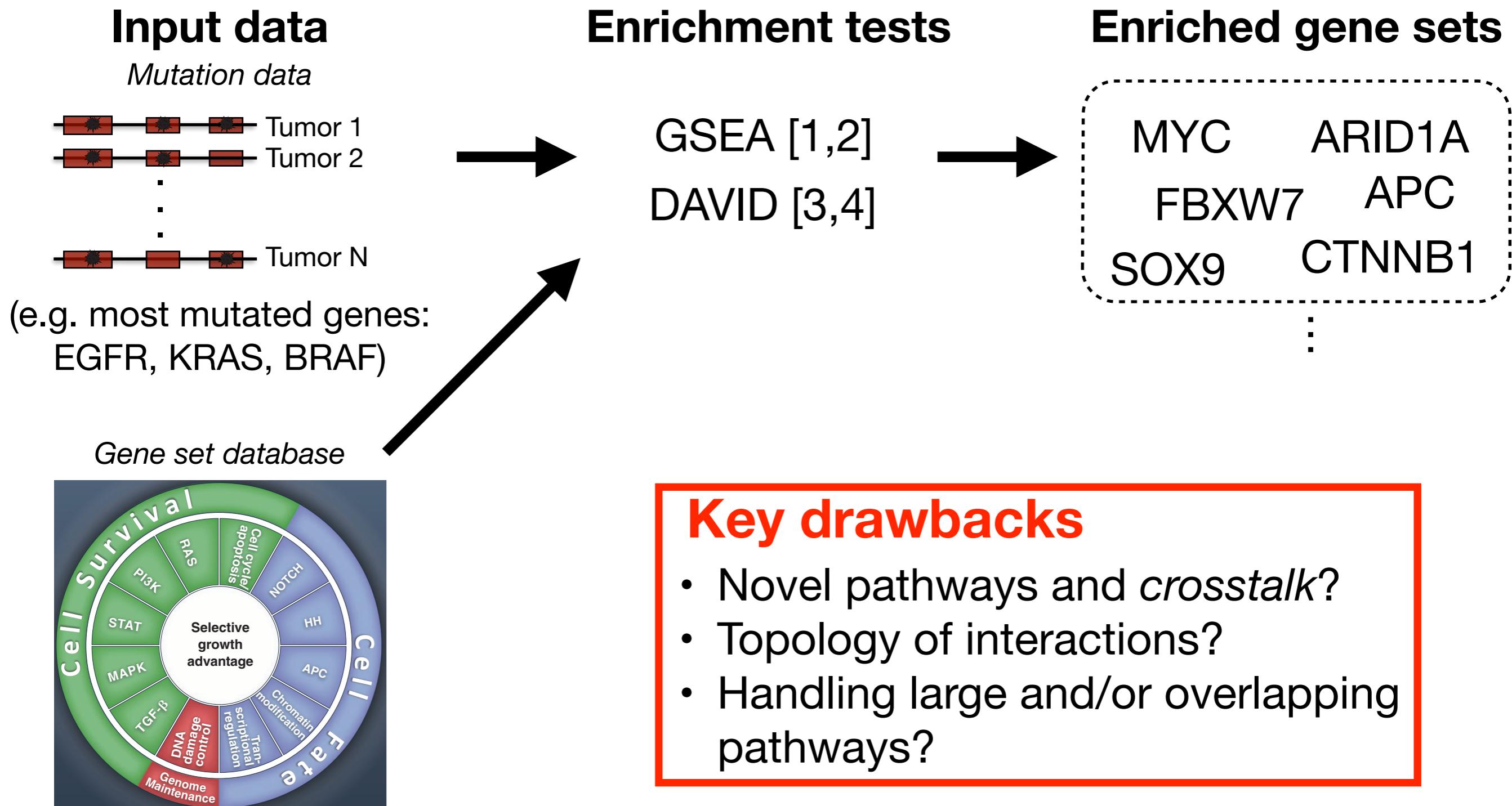
[1] Mootha *et al.* *Nat. Genet.* (2003).

[2] Subramanian *et al.* *PNAS* (2005).

[3] Huang *et al.* *Nat. Protoc.* (2009).

[4] Huang *et al.* *Nucleic Acids Res.* (2009)

Testing known gene sets and pathways



[1] Mootha *et al.* *Nat. Genet.* (2003).

[2] Subramanian *et al.* *PNAS* (2005).

[3] Huang *et al.* *Nat. Protoc.* (2009).

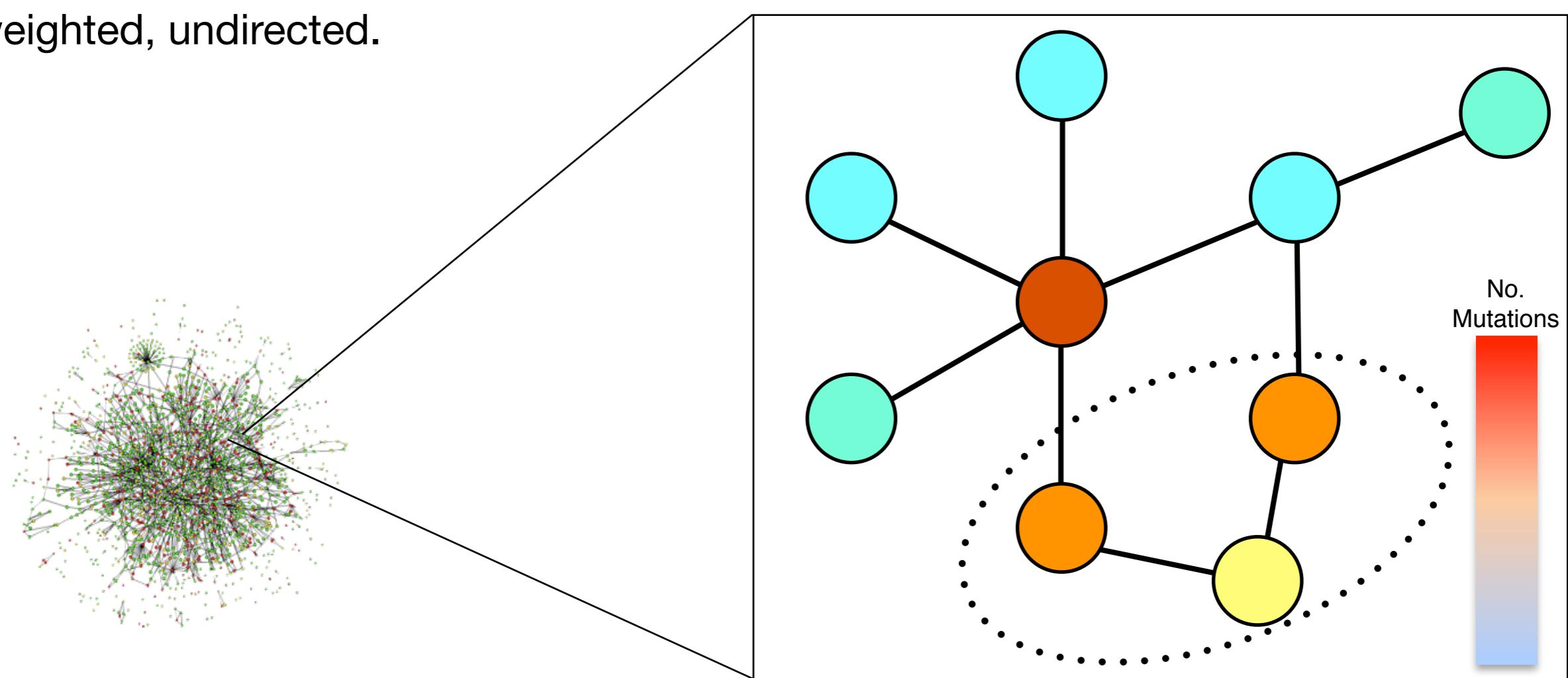
[4] Huang *et al.* *Nucleic Acids Res.* (2009)

Significantly mutated subnetworks of a protein-protein interaction network

Protein-protein interaction networks

- Nodes: genes/protein
- Edges: connect genes if the proteins they encode physically interact.
- Unweighted, undirected.

Goal: identify connected subnetworks with more mutations than expected by chance.

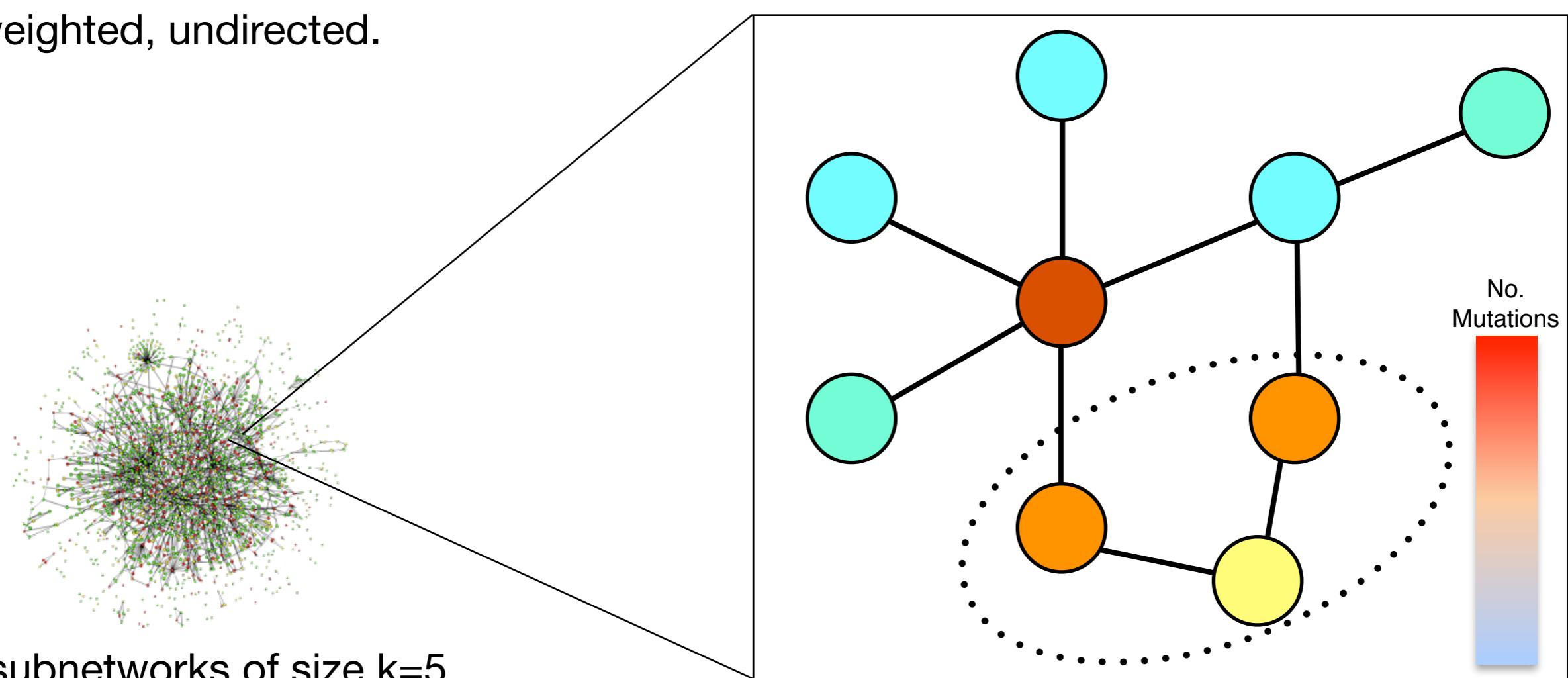


Significantly mutated subnetworks of a protein-protein interaction network

Protein-protein interaction networks

- Nodes: genes/protein
- Edges: connect genes if the proteins they encode physically interact.
- Unweighted, undirected.

Goal: identify connected subnetworks with more mutations than expected by chance.

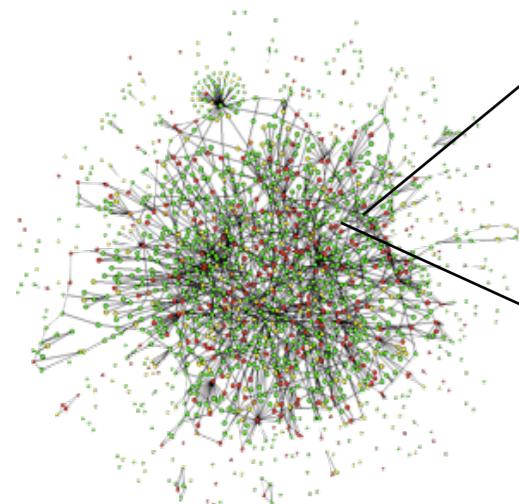


Significantly mutated subnetworks of a protein-protein interaction network

Protein-protein interaction networks

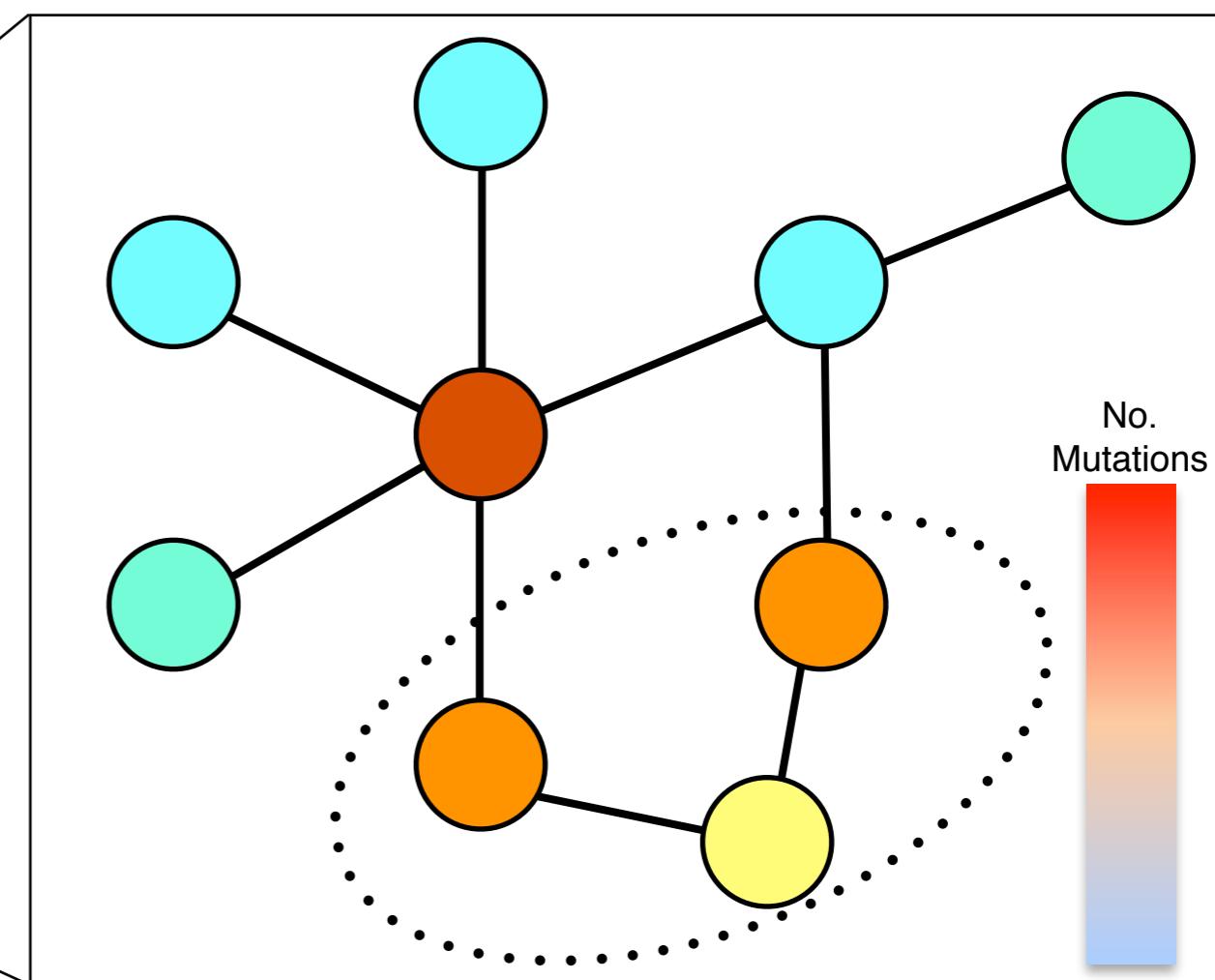
Network	Nodes	Edges	Diameter	ASP
HPRD	9,205	36,720	14	4.22
HINT+HI2012	9,859	40,705	14	4.08
iRefIndex	12,129	91,809	12	3.64
MultiNet	14,399	109,570	9	3.39

Low diameter → Most genes have a high-scoring neighbor



~ 10^{18} subnetworks of size k=5

Goal: identify connected subnetworks with more mutations than expected by chance.



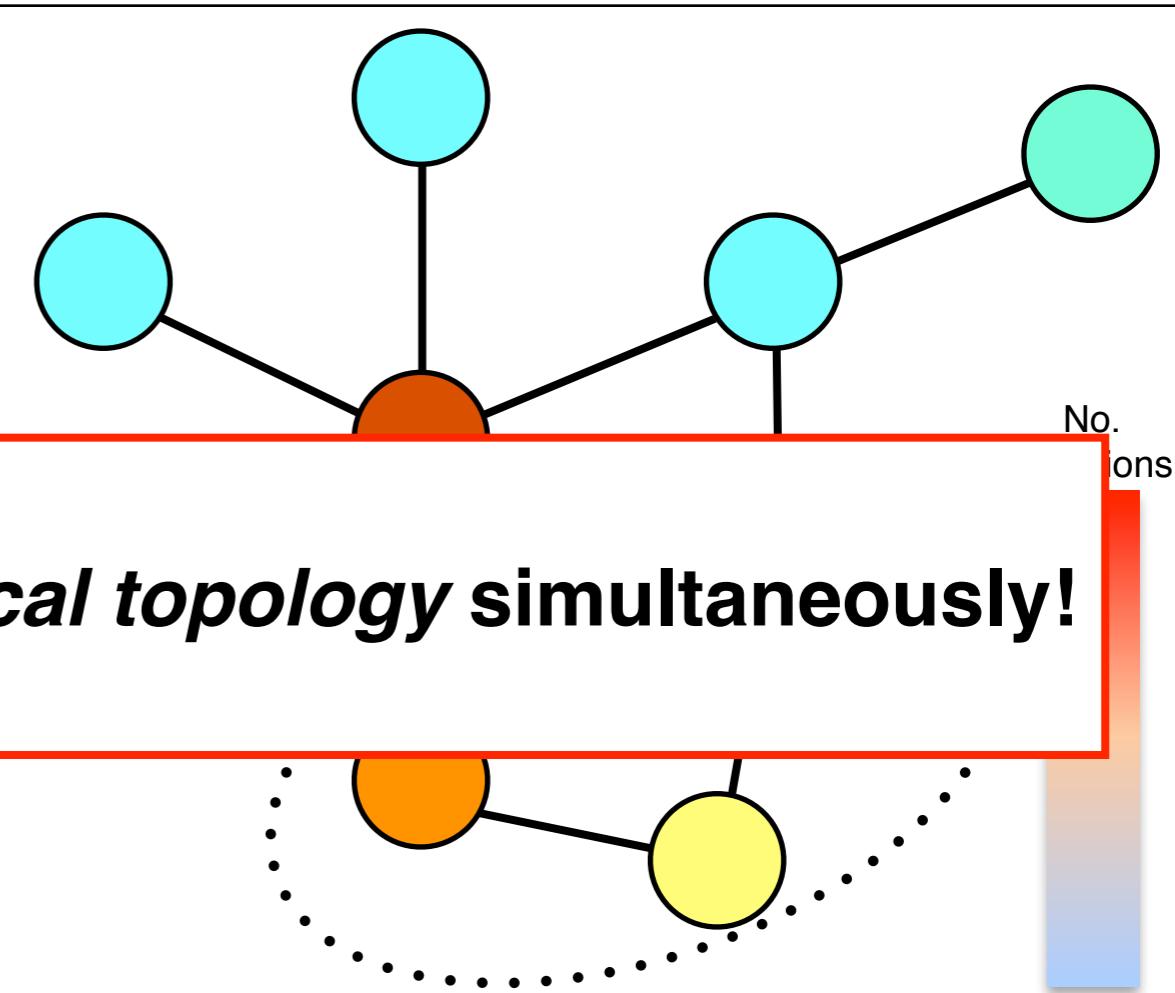
Significantly mutated subnetworks of a protein-protein interaction network

Protein-protein interaction networks

Network	Nodes	Edges	Diameter	ASP
HPRD	9,205	36,720	14	4.22
HINT+HI2012	9,859	40,705	14	4.08
iRefIndex	12,129	91,809	12	3.64
MultiNet	14,399	109,570	9	3.39

Low diameter → Most genes have a high-scoring neighbor

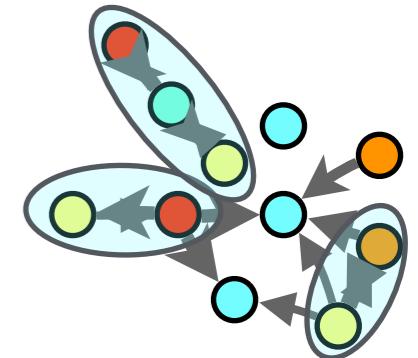
Goal: identify connected subnetworks with more mutations than expected by chance.



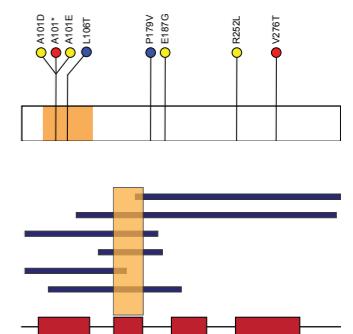
~ 10^{18} subnetworks of size $k=5$

Outline

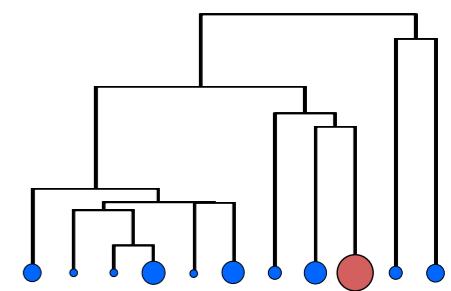
1. A new algorithm, **HotNet2**.



2. Comparison of HotNet2 to similar methods.

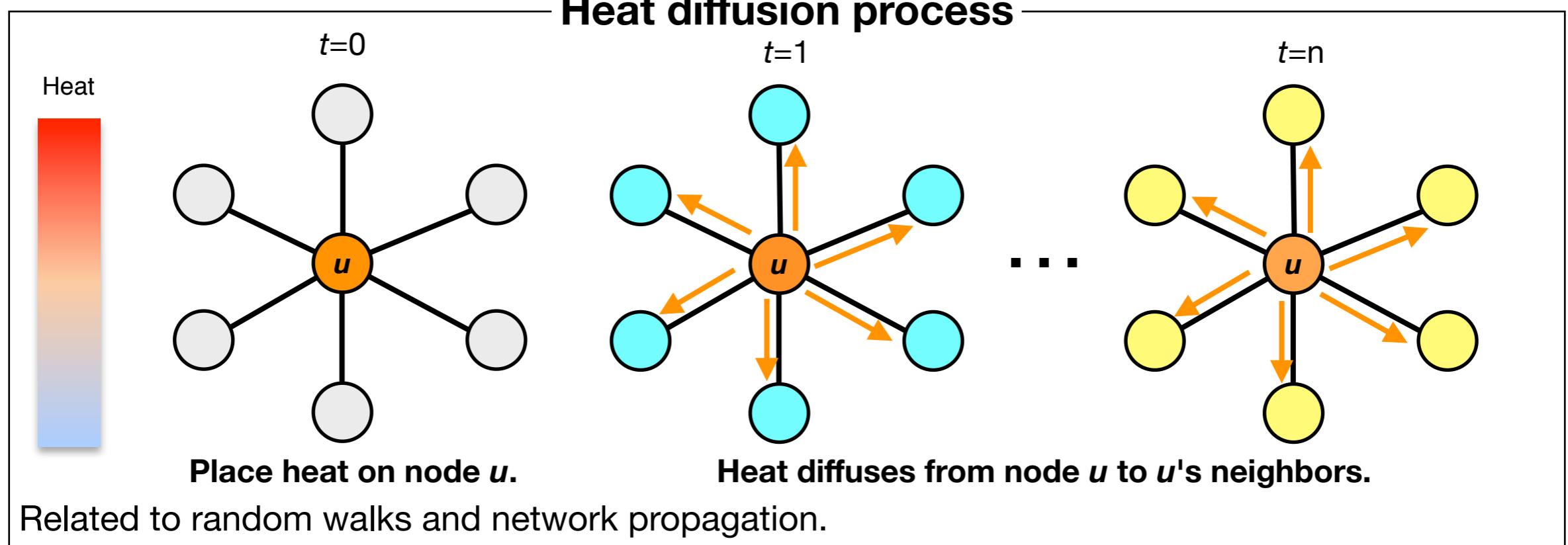


3. Application to TCGA Pan-Cancer data.

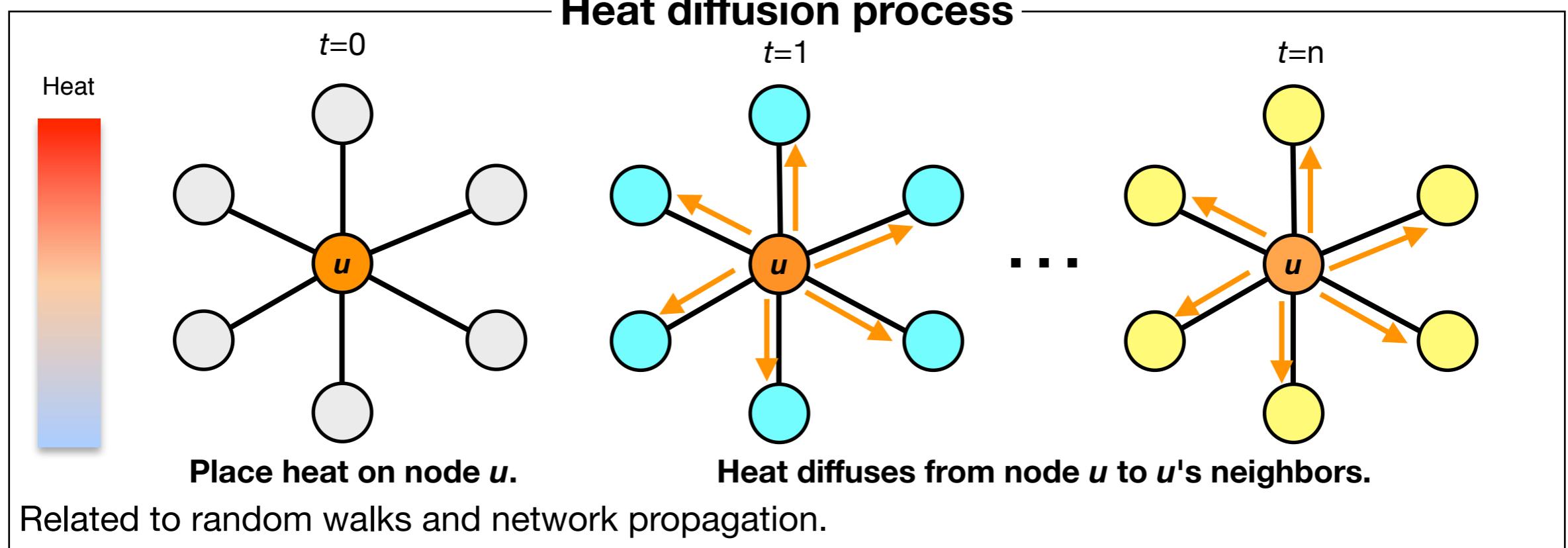


4. Latest work.

Encoding mutations and graph topology with heat diffusion

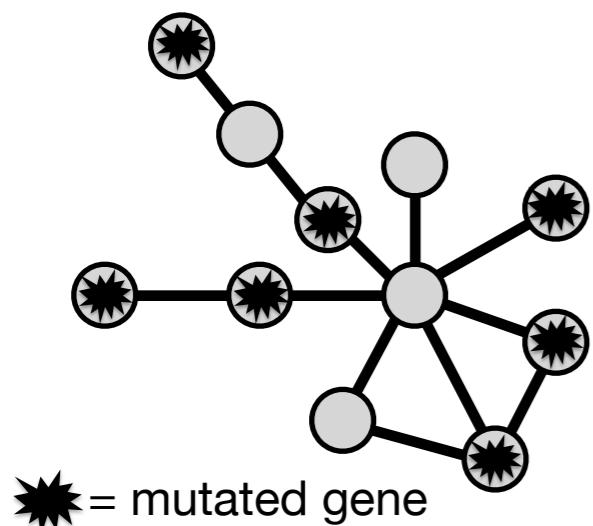


Encoding mutations and graph topology with heat diffusion

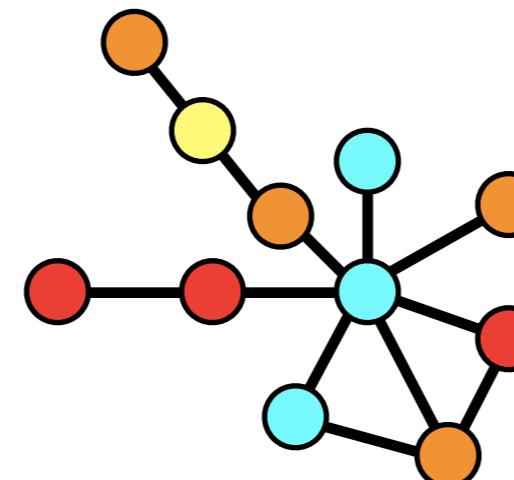


HotNet (Vandin et al. JCB & RECOMB 2010)

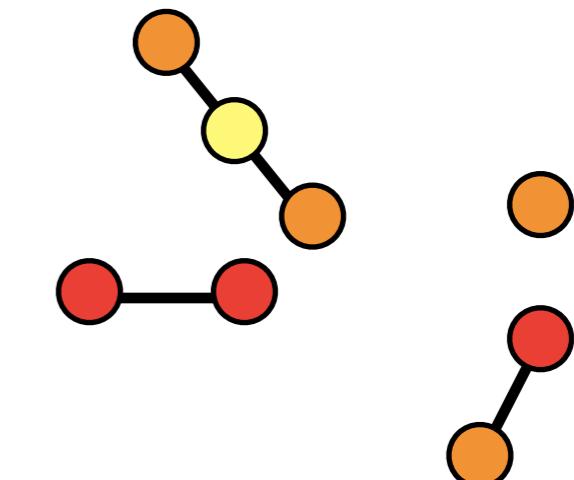
Mutations = heat sources



Heat diffusion



Extract “hot” subnetworks



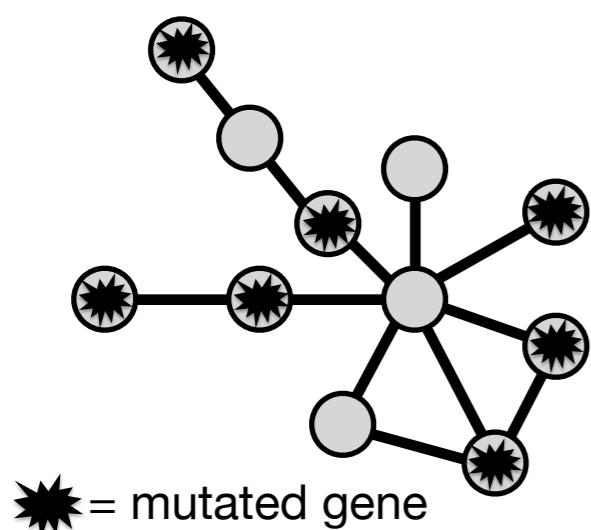
HotNet applied to TCGA data

TCGA Papers (~300 samples)

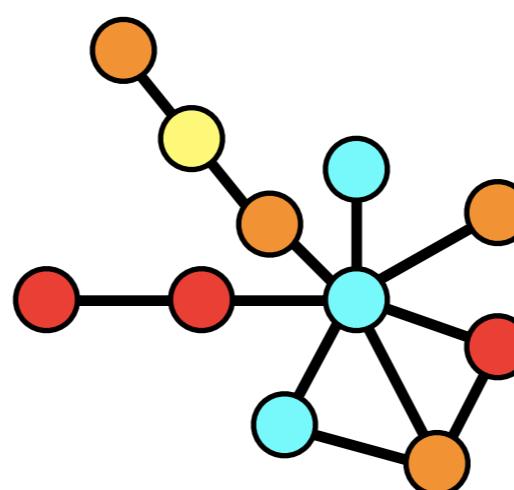
- Leukemia (*NEJM* 2013)
- Kidney (*Nature* 2011)
- Ovarian (*Nature* 2011)

HotNet (Vandin et al. *JCB & RECOMB* 2010)

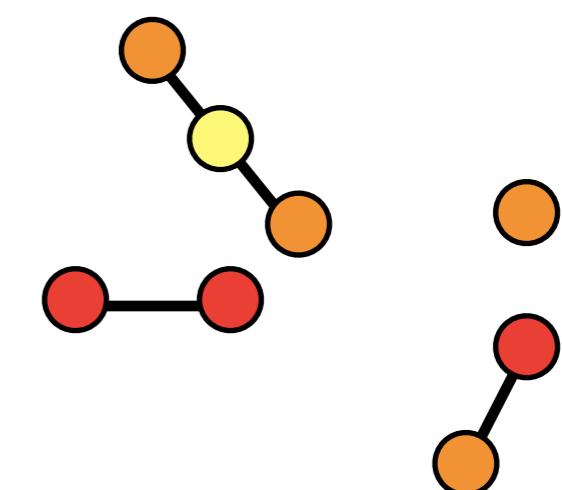
Mutations = heat sources



Heat diffusion



Extract “hot” subnetworks



HotNet applied to TCGA data

TCGA Papers (~300 samples)

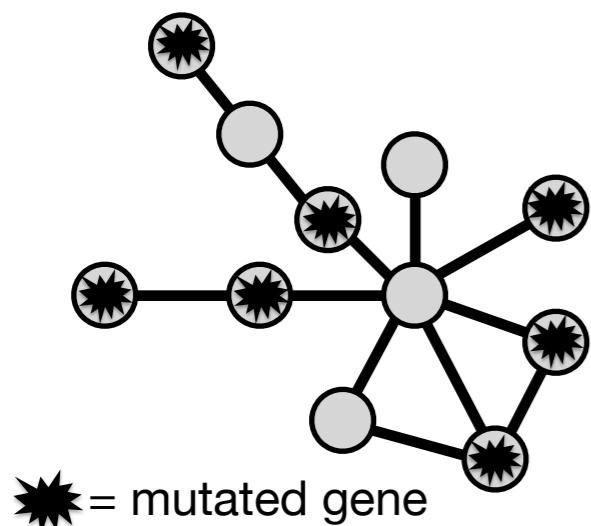
- Leukemia (*NEJM* 2013)
- Kidney (*Nature* 2011)
- Ovarian (*Nature* 2011)



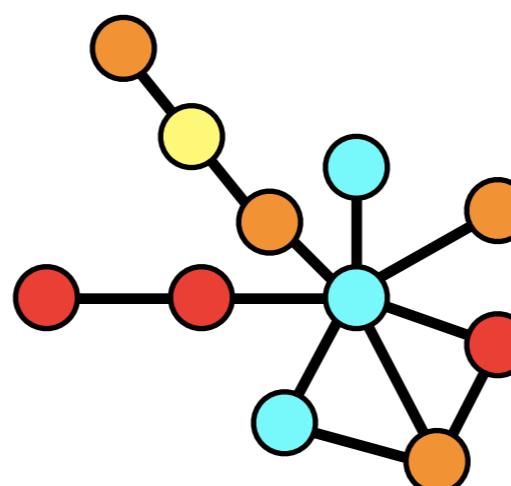
TCGA Pan-Cancer (>3000 samples)

HotNet (Vandin et al. *JCB & RECOMB* 2010)

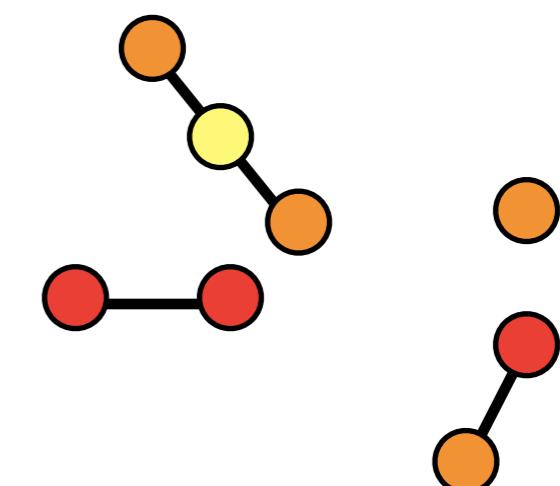
Mutations = heat sources



Heat diffusion



Extract “hot” subnetworks



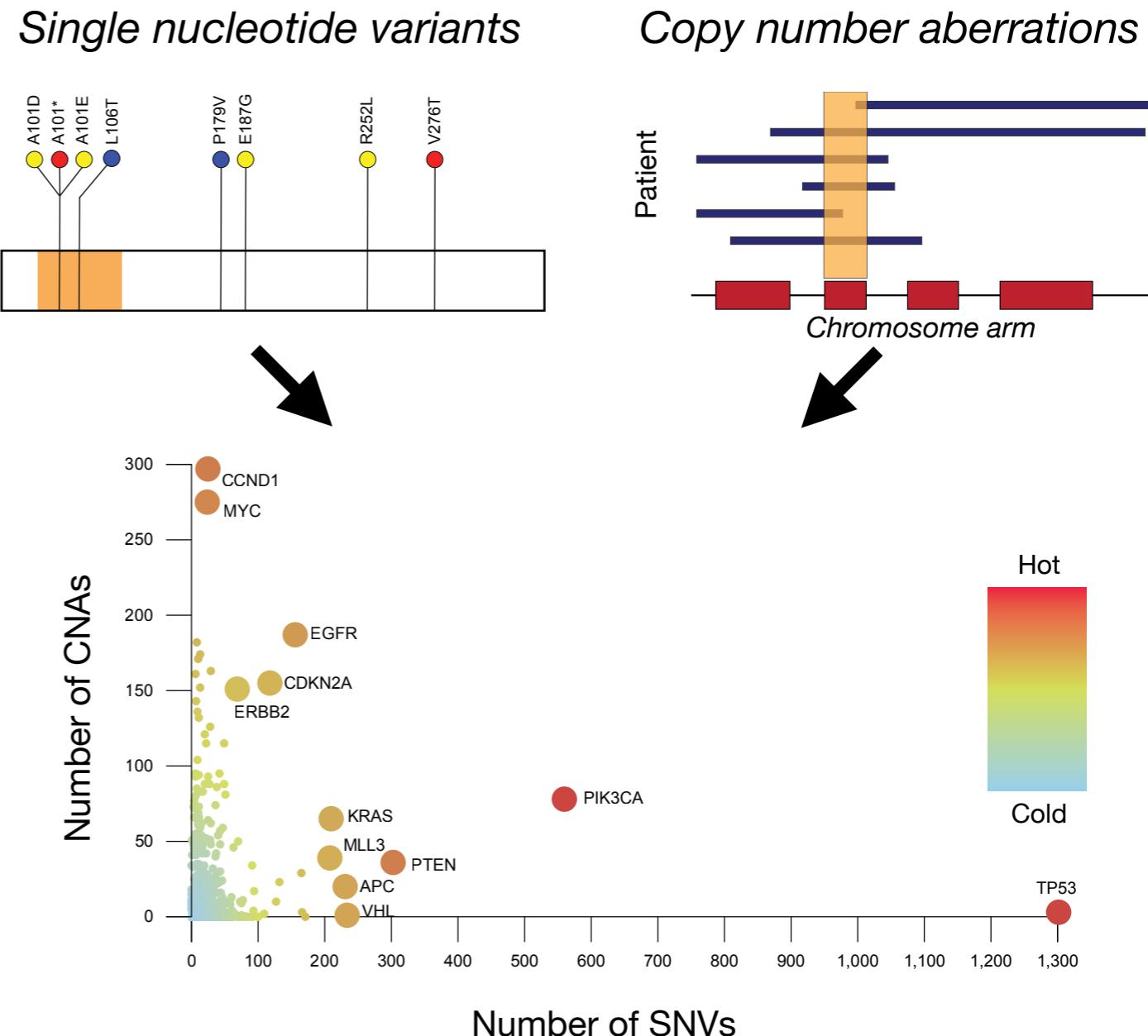
TCGA Pan-Cancer

Tumor samples

Cancer	Samples	Color
BLCA	99	
BRCA	772	
COAD/READ	224	
GBM	291	
HNSC	306	
KIRC	417	
LAML	196	
LUAD	230	
LUSC	178	
OV	316	
UCEC	248	

3,110 tumors of 12 cancer types

Mutations

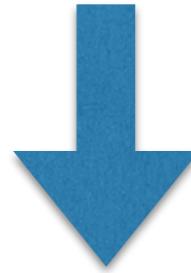


SNVs and CNAs in 3,110 samples
among 11,565 expressed genes

TCGA Pan-Cancer

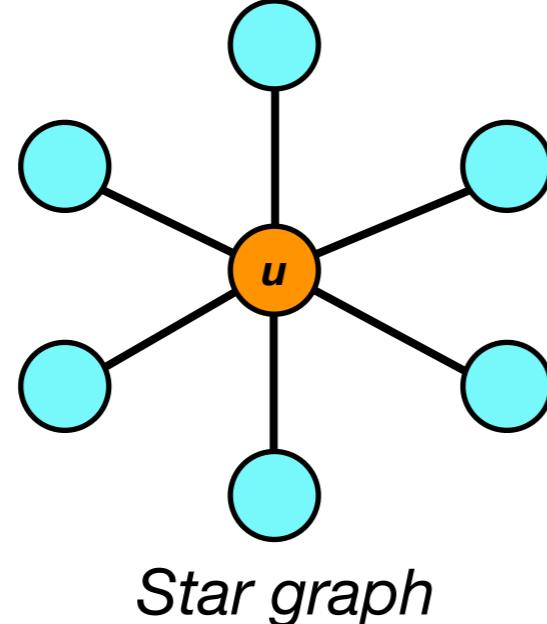
Mutations

HotNet

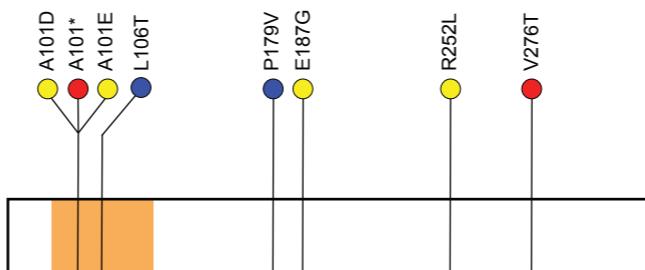


Potential artifacts

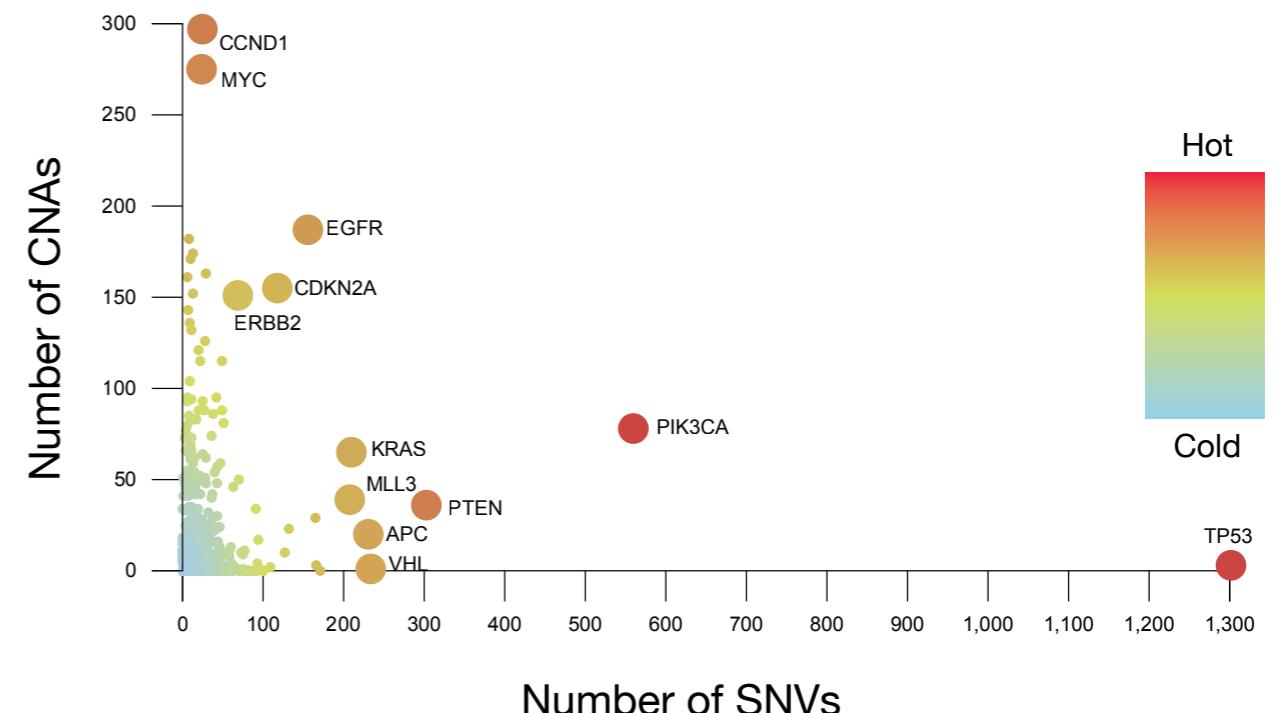
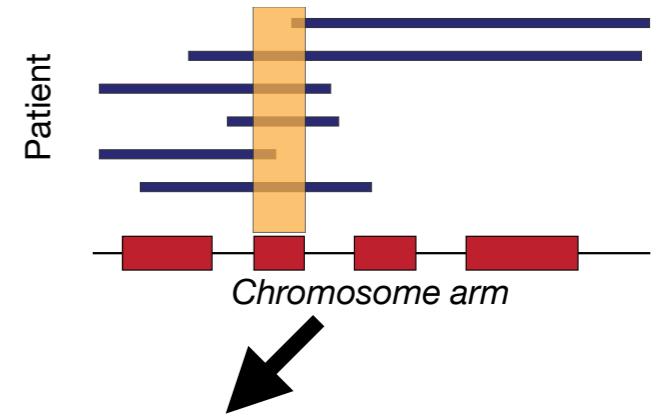
Hot nodes with high degree often form large “star” subnetworks with many cold nodes



Single nucleotide variants



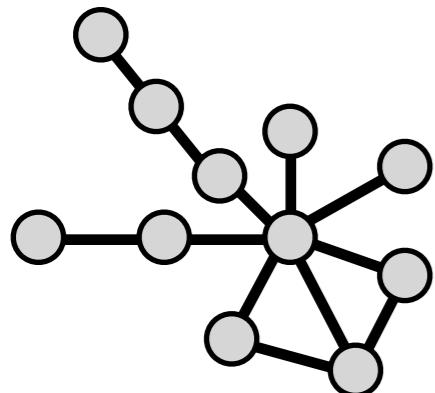
Copy number aberrations



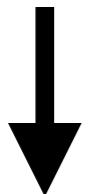
SNVs and CNAs in 3,110 samples among 11,565 expressed genes

HotNet Algorithm

Input

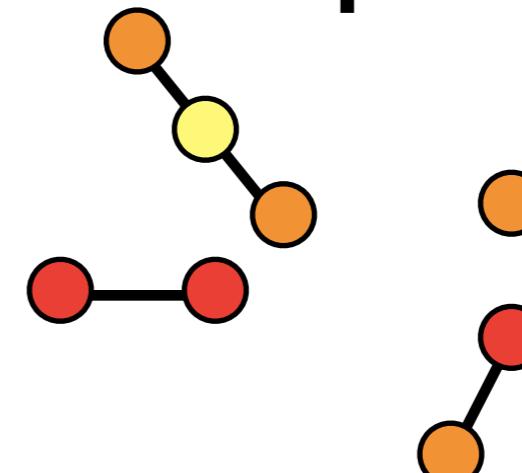


\mathbf{A} = adjacency matrix \mathbf{h} = gene scores



$$\begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}$$

Output



Connected components



Threshold at δ

$$\left[\begin{array}{c} \text{Heat kernel} \\ f(\mathbf{A}, t) \end{array} \right]$$

*Diffusion matrix
(symmetric)*

Time parameter t

$$\begin{pmatrix} h_1 & & 0 \\ \vdots & \ddots & \vdots \\ 0 & & h_n \end{pmatrix} =$$

$$\begin{pmatrix} s_{11} & \dots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{n1} & \dots & s_{nn} \end{pmatrix}$$

symmetrize



$$\begin{pmatrix} r_{11} & \dots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{n1} & \dots & r_{nn} \end{pmatrix}$$

Similarity matrix

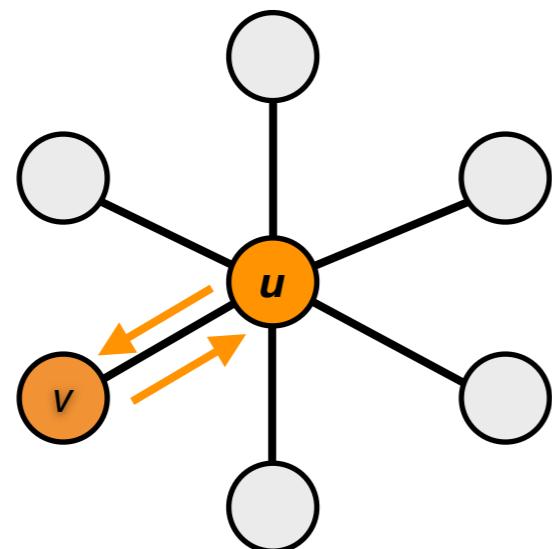
s_{ij} = heat on vertex i at time t given initial heat h_j on vertex j at time 0.

Direction of heat is important

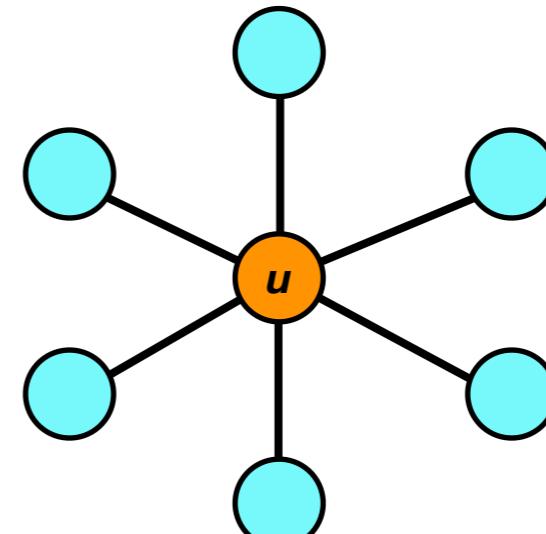
→ HotNet can fail

**HotNet's similarity matrix
is symmetric**

u sends the same
heat to v even
though u has much
higher degree



Potential artifacts



Star graph

Hot nodes with high
degree often form
large “star”
subnetworks with
many cold nodes

$$\left[\begin{array}{c} \text{Heat kernel} \\ f(A, t) \end{array} \right]$$

$$\begin{pmatrix} h_1 & & 0 \\ & \ddots & \\ 0 & & h_n \end{pmatrix}$$

$$\begin{pmatrix} s_{11} & \dots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{n1} & \dots & s_{nn} \end{pmatrix}$$

symmetrize

$$\begin{pmatrix} r_{11} & \dots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{n1} & \dots & r_{nn} \end{pmatrix}$$

*Diffusion matrix
(symmetric)*

Time parameter t

Similarity matrix

s_{ij} = heat on vertex i at time t given initial heat h_j on vertex j at time 0.

HotNet2 algorithm

(HotNet diffusion oriented subnetworks)

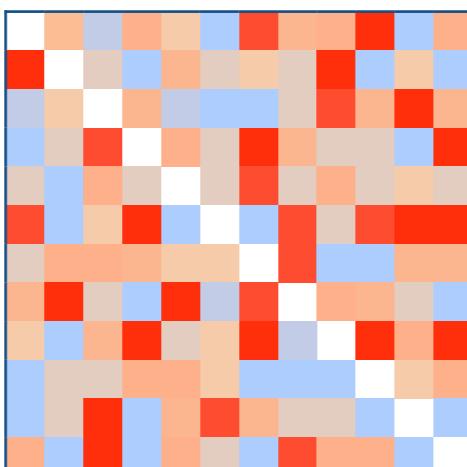
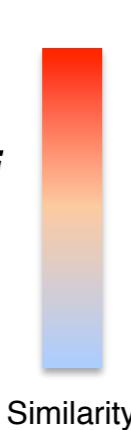
Need to consider the source of heat



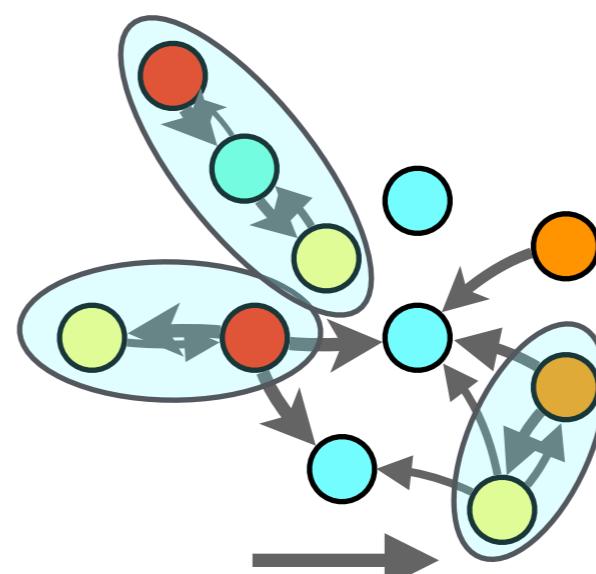
Encode directionality with asymmetric heat diffusion.

- Hot genes do not necessarily implicate their neighbors.
- Hot subnetworks have a directed path between each pair of nodes.

s_{ij} = heat on vertex i at *equilibrium* given initial heat h_j on vertex j at time 0.



Similarity matrix S
Asymmetric!



Identify *strongly* connected components



HotNet2 algorithm

(HotNet diffusion oriented subnetworks)

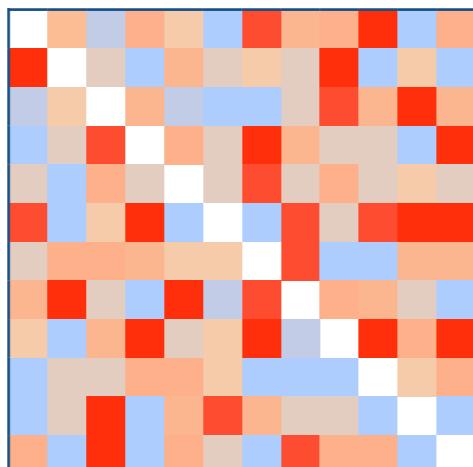
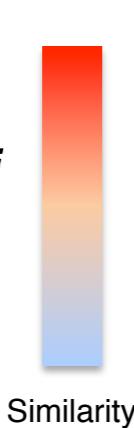
Need to consider the source of heat



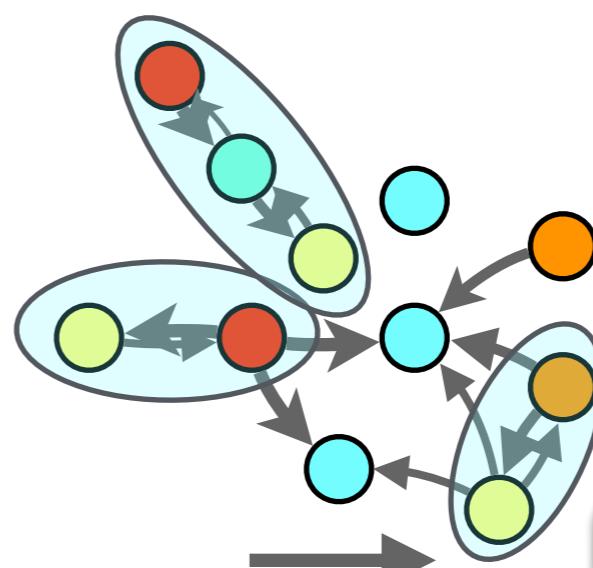
Encode directionality with asymmetric heat diffusion.

- Hot genes do not necessarily implicate their neighbors.
- Hot subnetworks have a directed path between each pair of nodes.

s_{ij} = heat on vertex i at *equilibrium* given initial heat h_j on vertex j at time 0.



Similarity matrix S
Asymmetric!



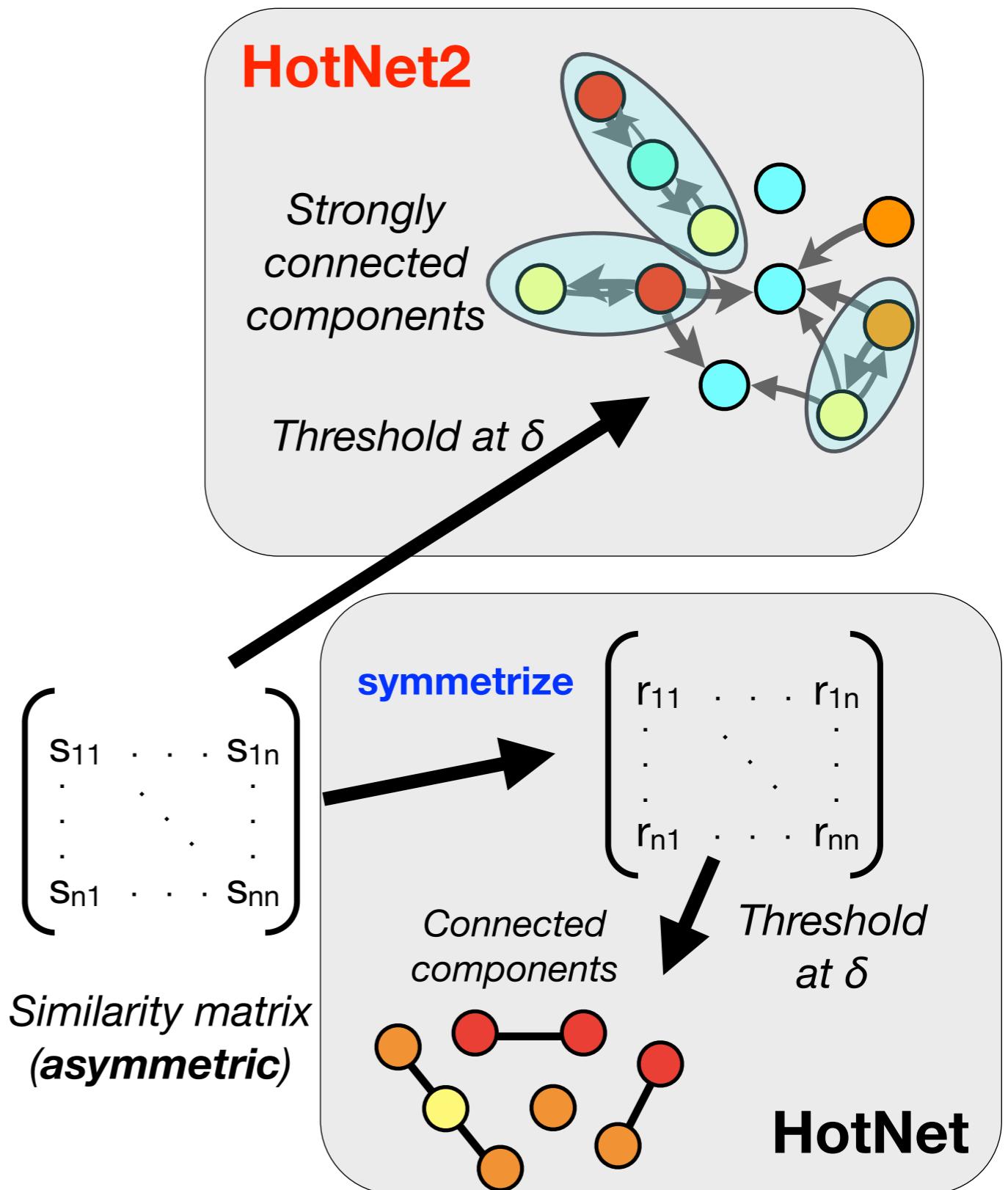
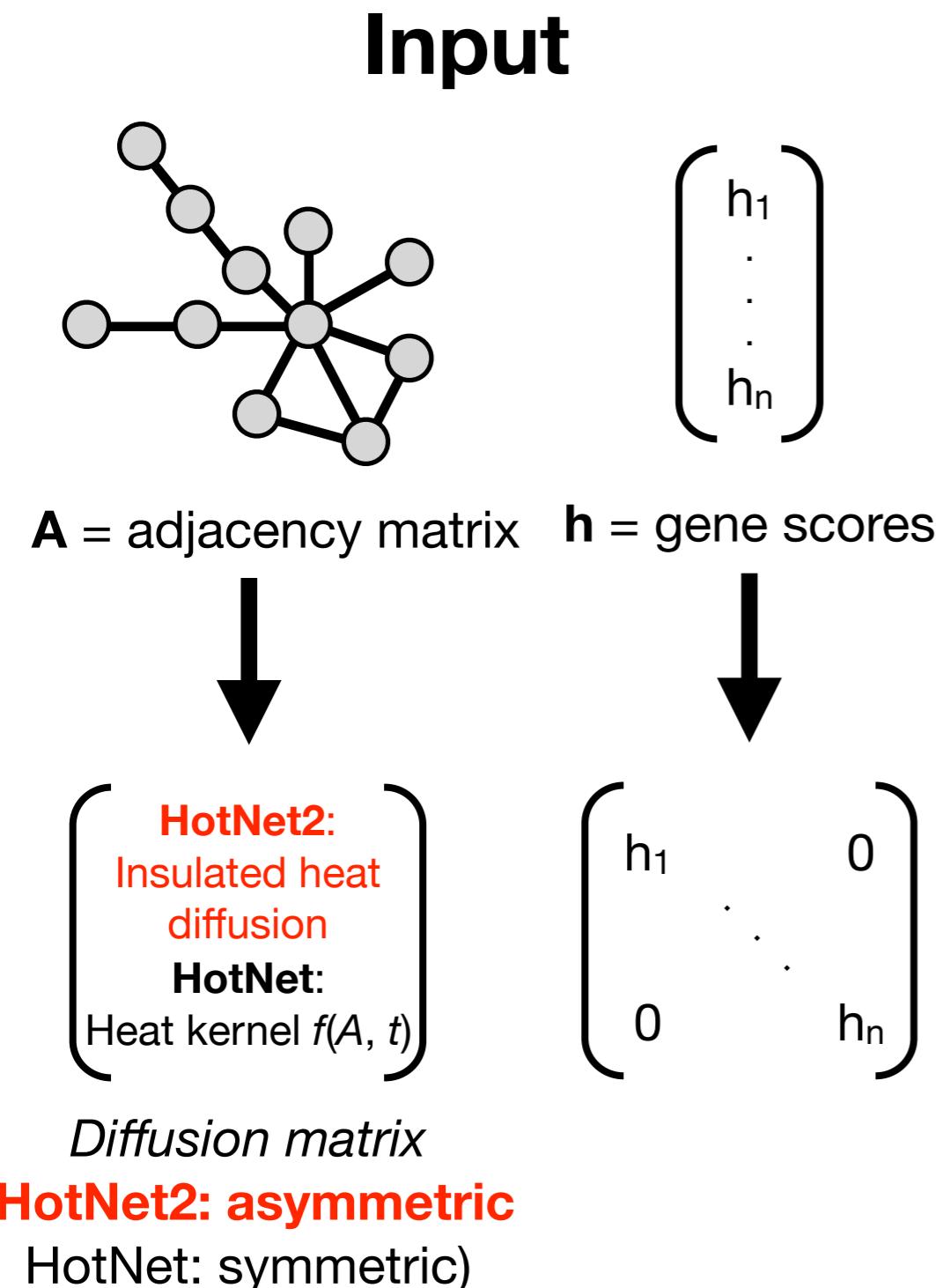
similarity of u and v

Identify *strongly* connected components

TCGA Papers
Gastric, *Nature* (2014)
Thyroid, *Cell* (2014)
...

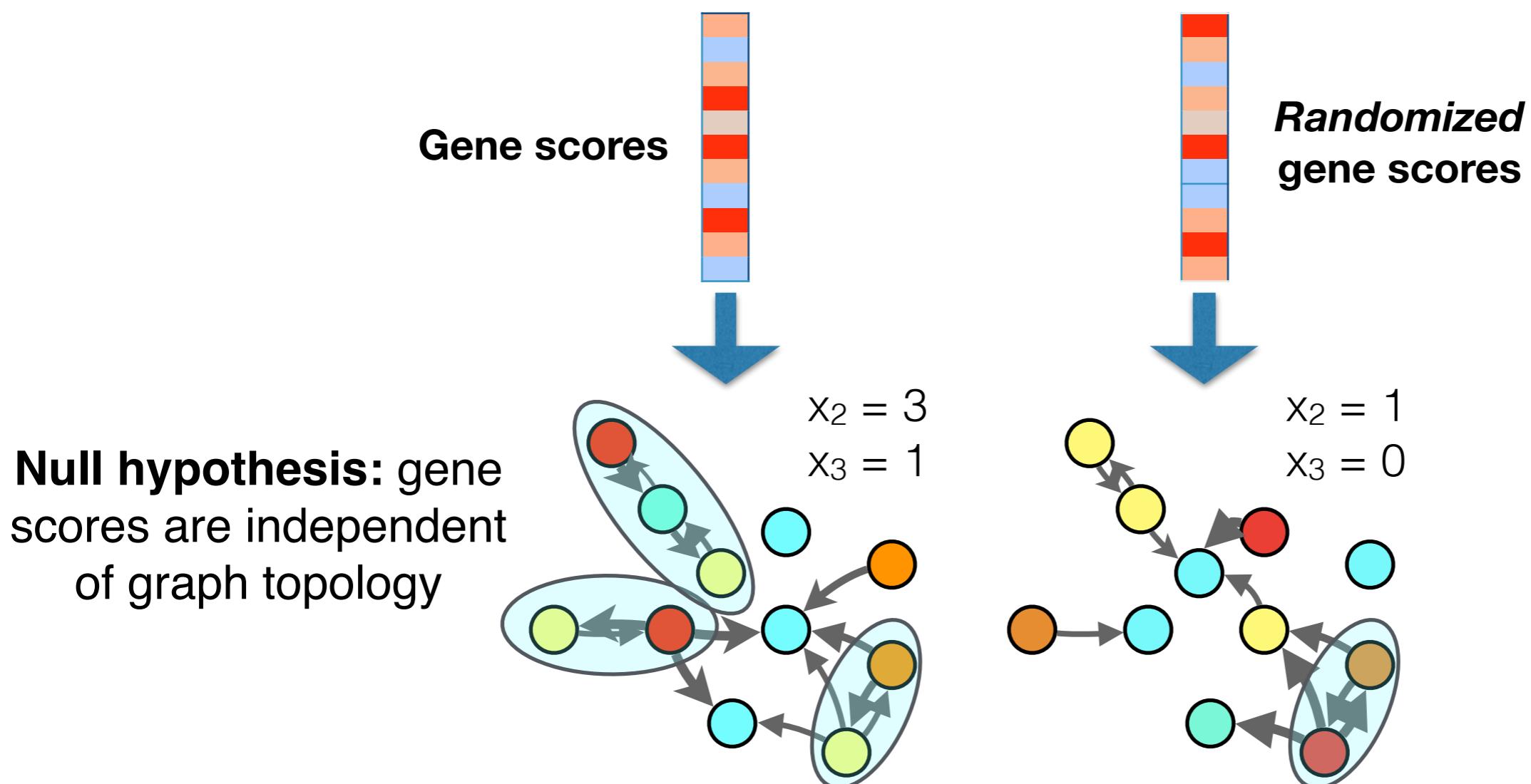


HotNet2 vs. HotNet



Statistical test

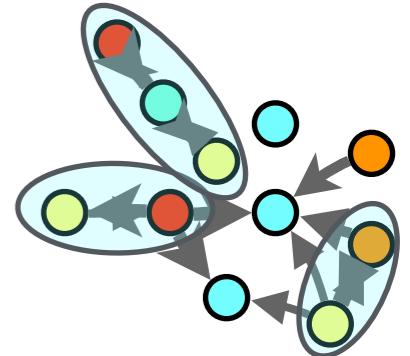
Evaluate output with rigorously bounded False Discovery Rate (FDR)



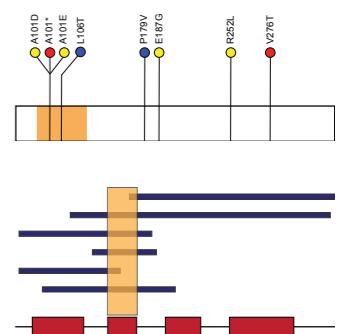
X_k : number of subnetworks of size $\geq k$
 $\Pr(X_k \geq x_k \mid h, \delta)$

Outline

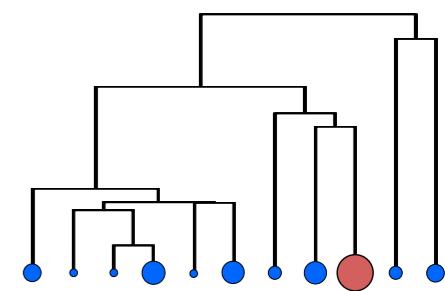
1. A new algorithm, HotNet2.



2. Comparison of HotNet2 to similar methods.



3. Application to TCGA Pan-Cancer data.



4. Latest work.

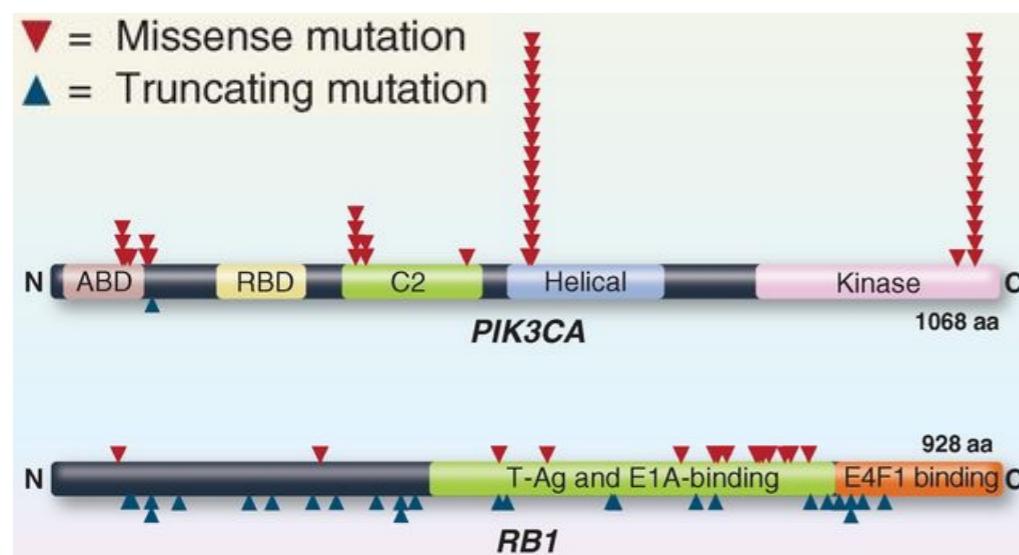
HotNet2 outperforms other methods on real data

No gold standard dataset → compare methods at identifying putative cancer genes

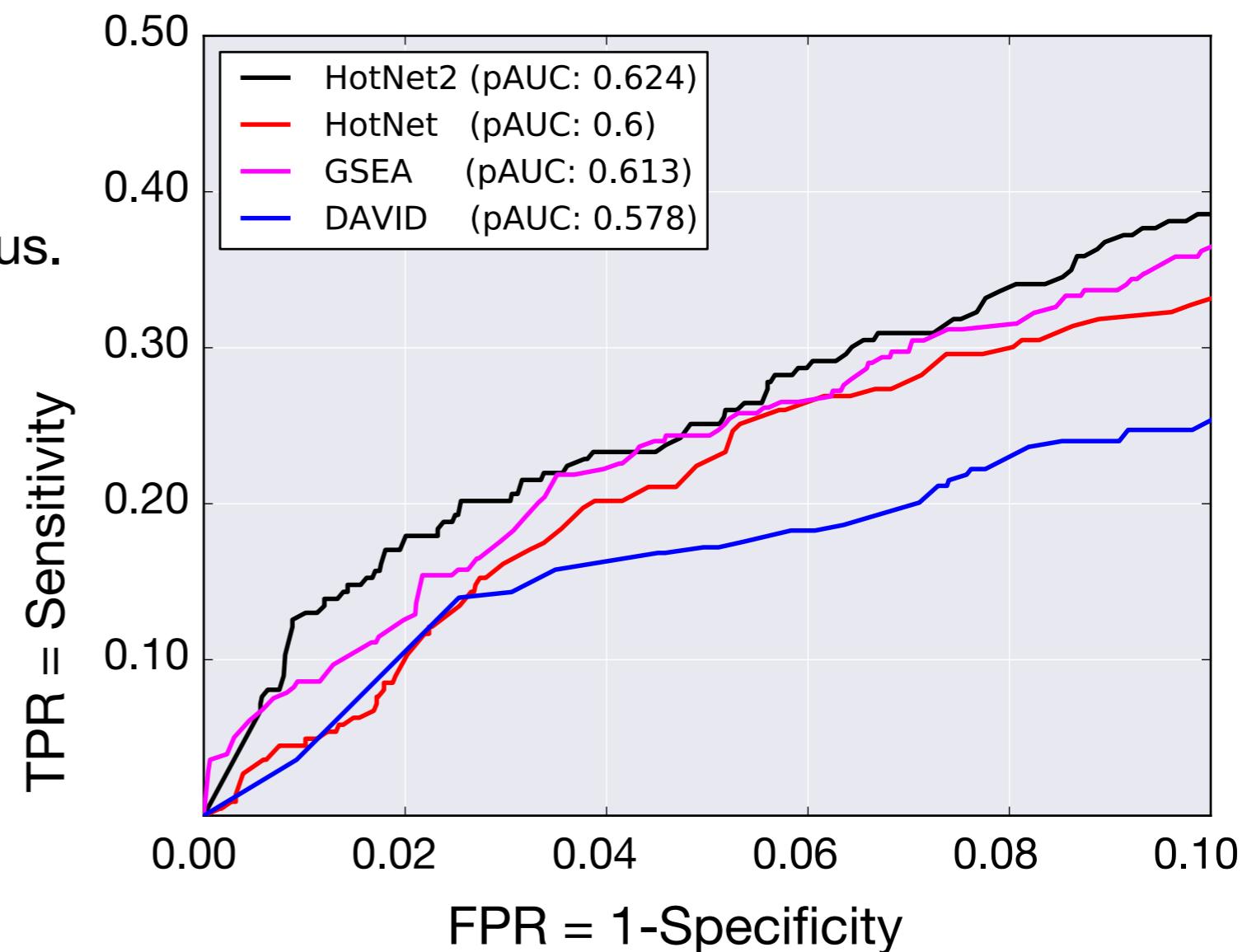
Dataset of *putative* cancer genes

Cancer genes have:

1. ≥ 20% truncating mutations; or,
2. ≥ 20% mutations clustered at a locus.

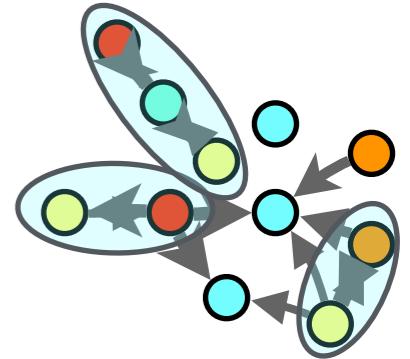


Vogelstein et al. (Science, 2013)

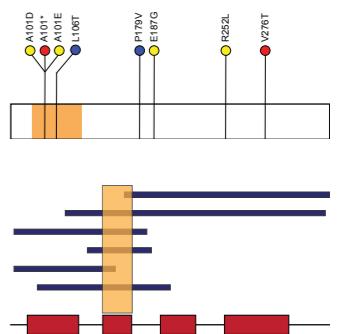


Outline

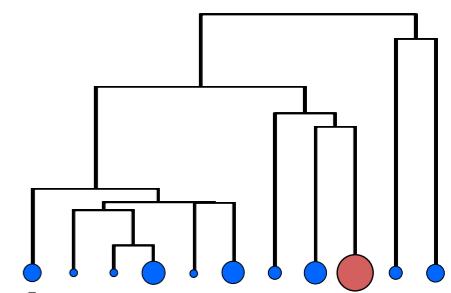
1. A new algorithm, HotNet2.



2. Comparison of HotNet2 to similar methods.



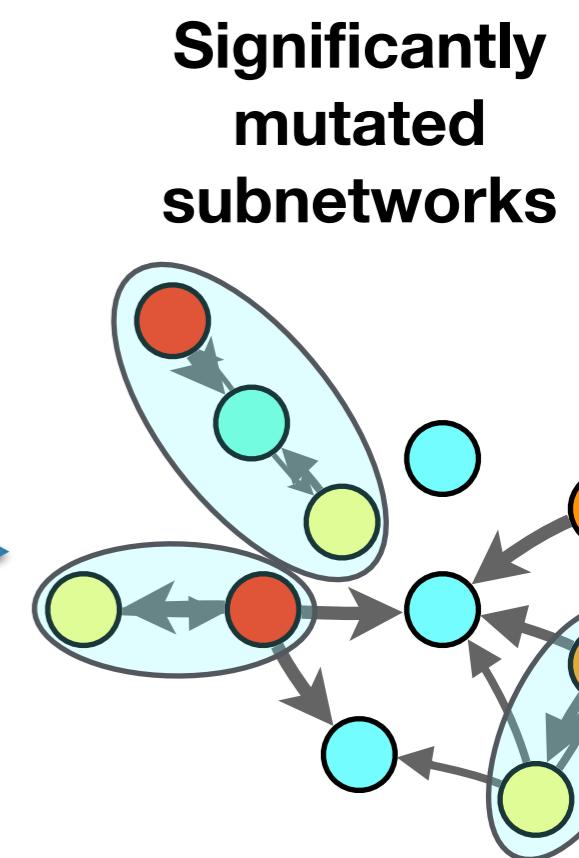
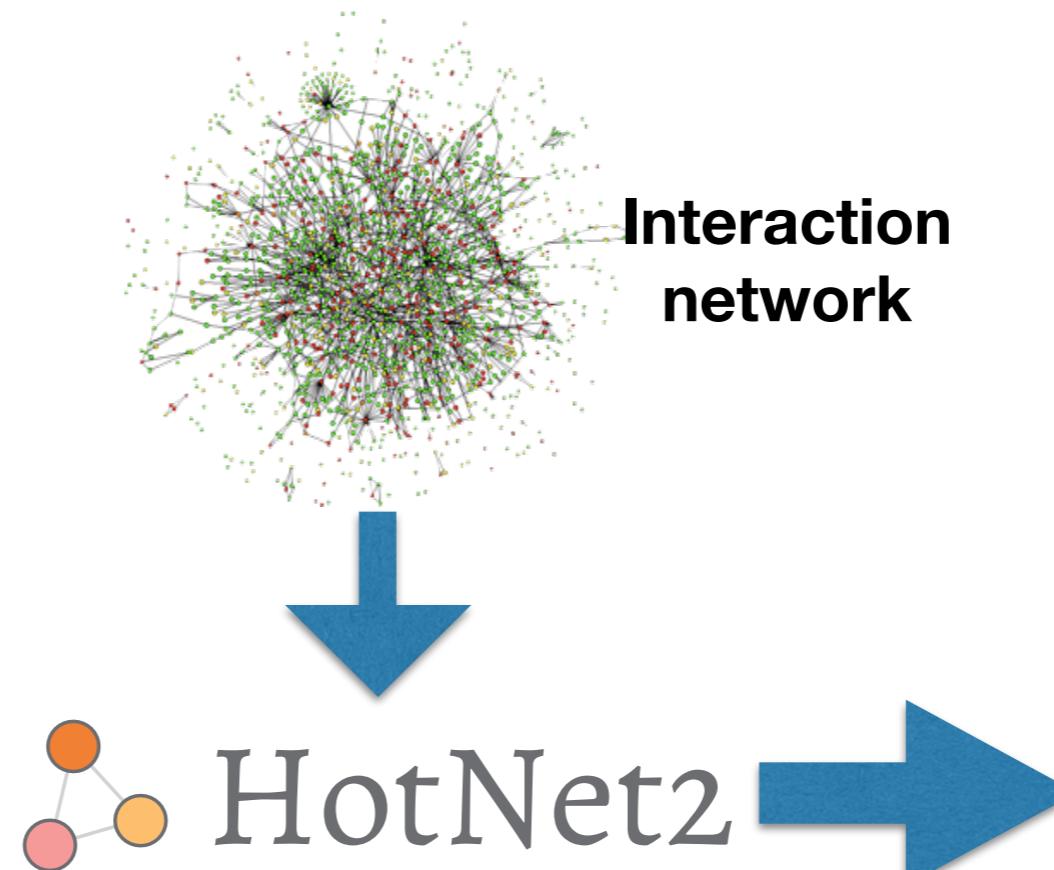
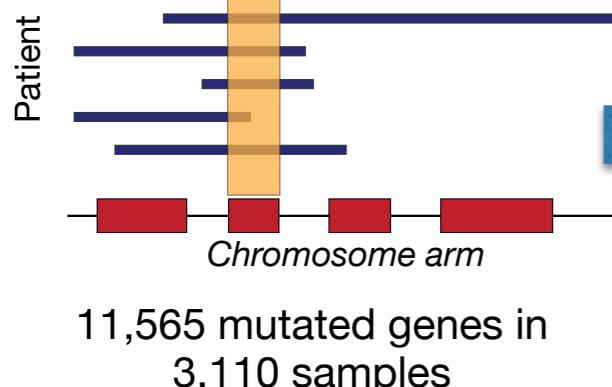
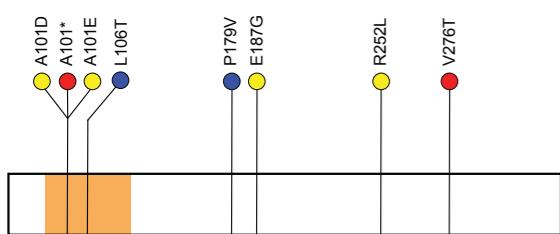
3. Application to TCGA Pan-Cancer data.



4. Latest work.

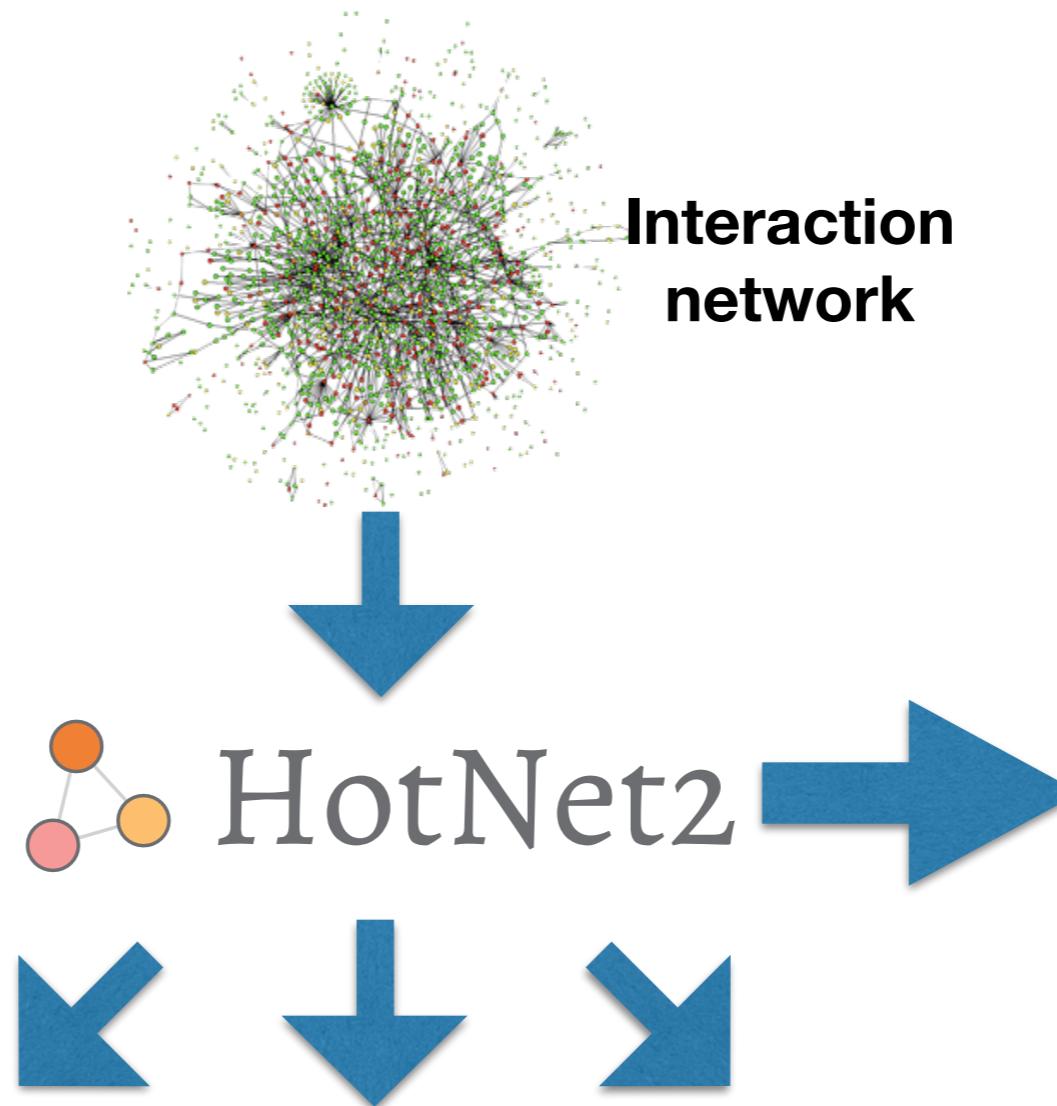
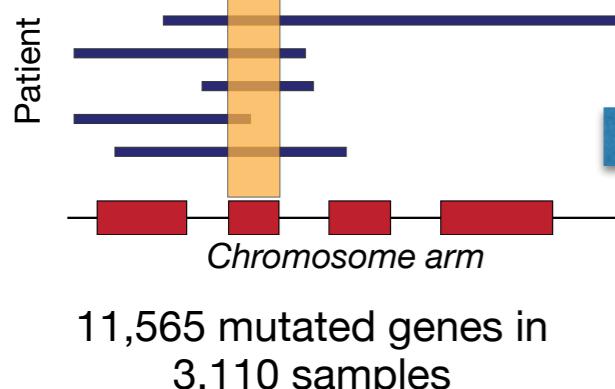
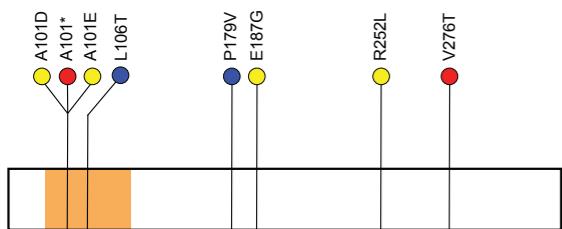
HotNet2 runs on TCGA Pan-Cancer dataset

Mutation & copy number data



HotNet2 runs on TCGA Pan-Cancer dataset

Mutation & copy number data



HINT+HI2012 ($P < 0.01$)

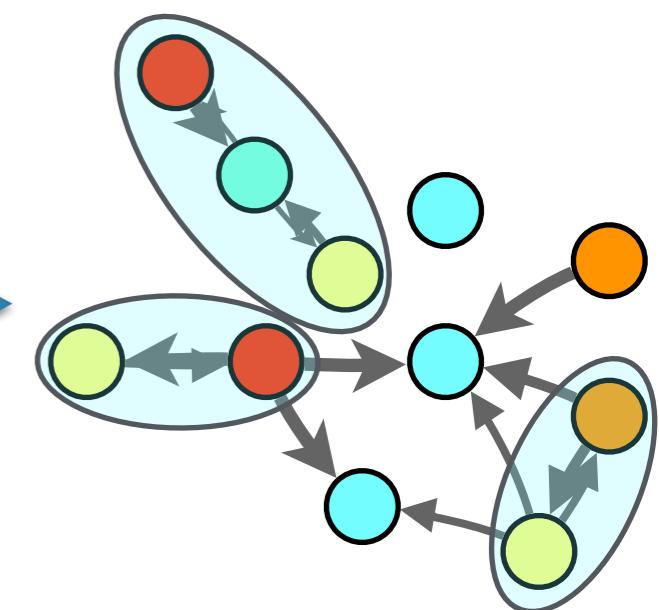
40,704 interactions
9,858 proteins

iRefIndex 9.0 ($P < 0.01$)

91,808 interactions
12,128 proteins

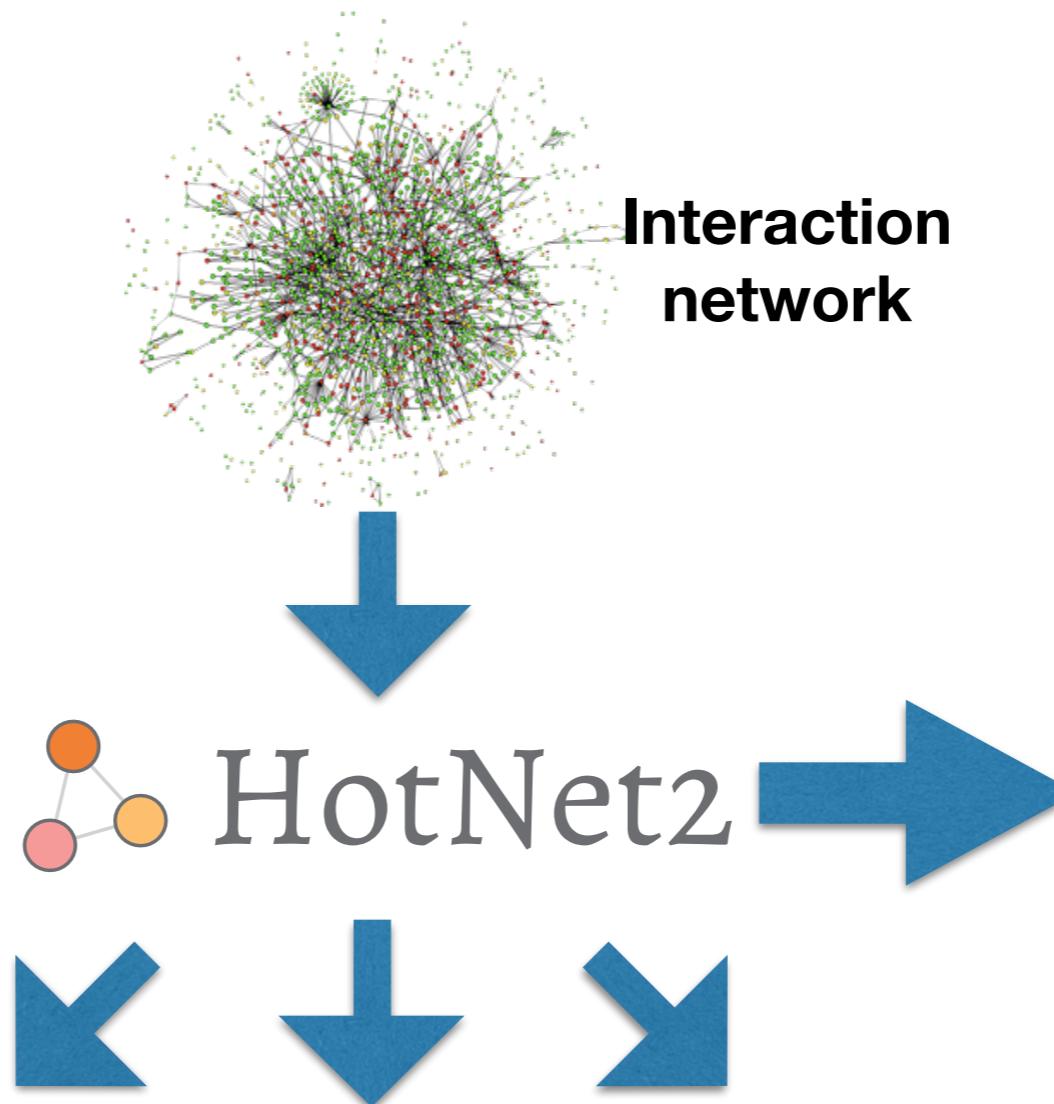
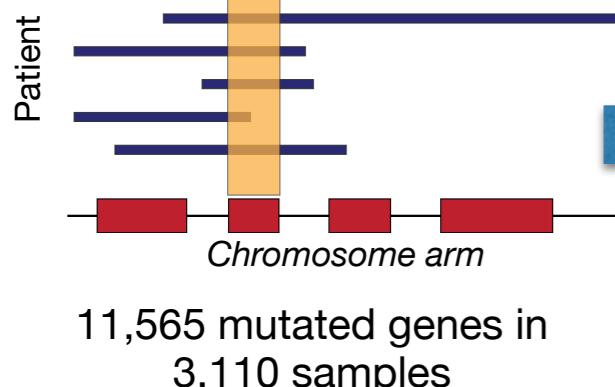
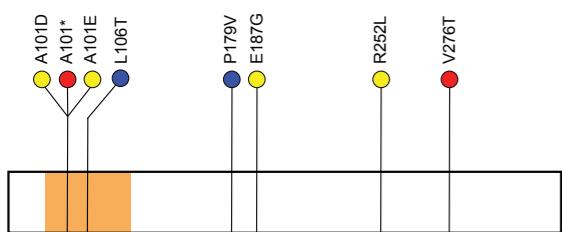
Multinet ($P < 0.01$)

109,569 interactions
14,398 proteins



HotNet2 runs on TCGA Pan-Cancer dataset

Mutation & copy number data



HINT+HI2012 ($P < 0.01$)

40,704 interactions
9,858 proteins

iRefIndex 9.0 ($P < 0.01$)

91,808 interactions
12,128 proteins

Multinet ($P < 0.01$)

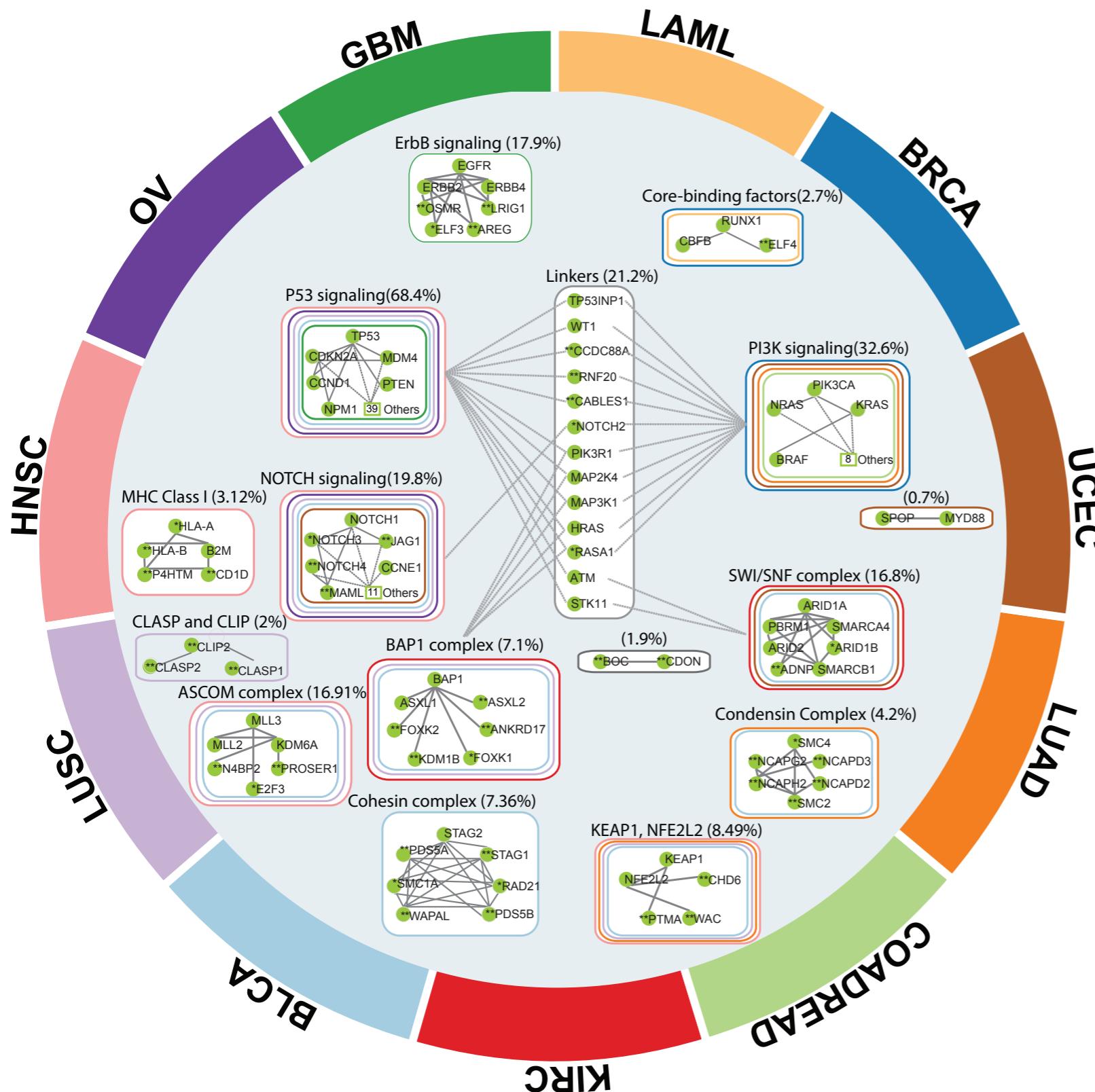
109,569 interactions
14,398 proteins

Consensus subnetworks

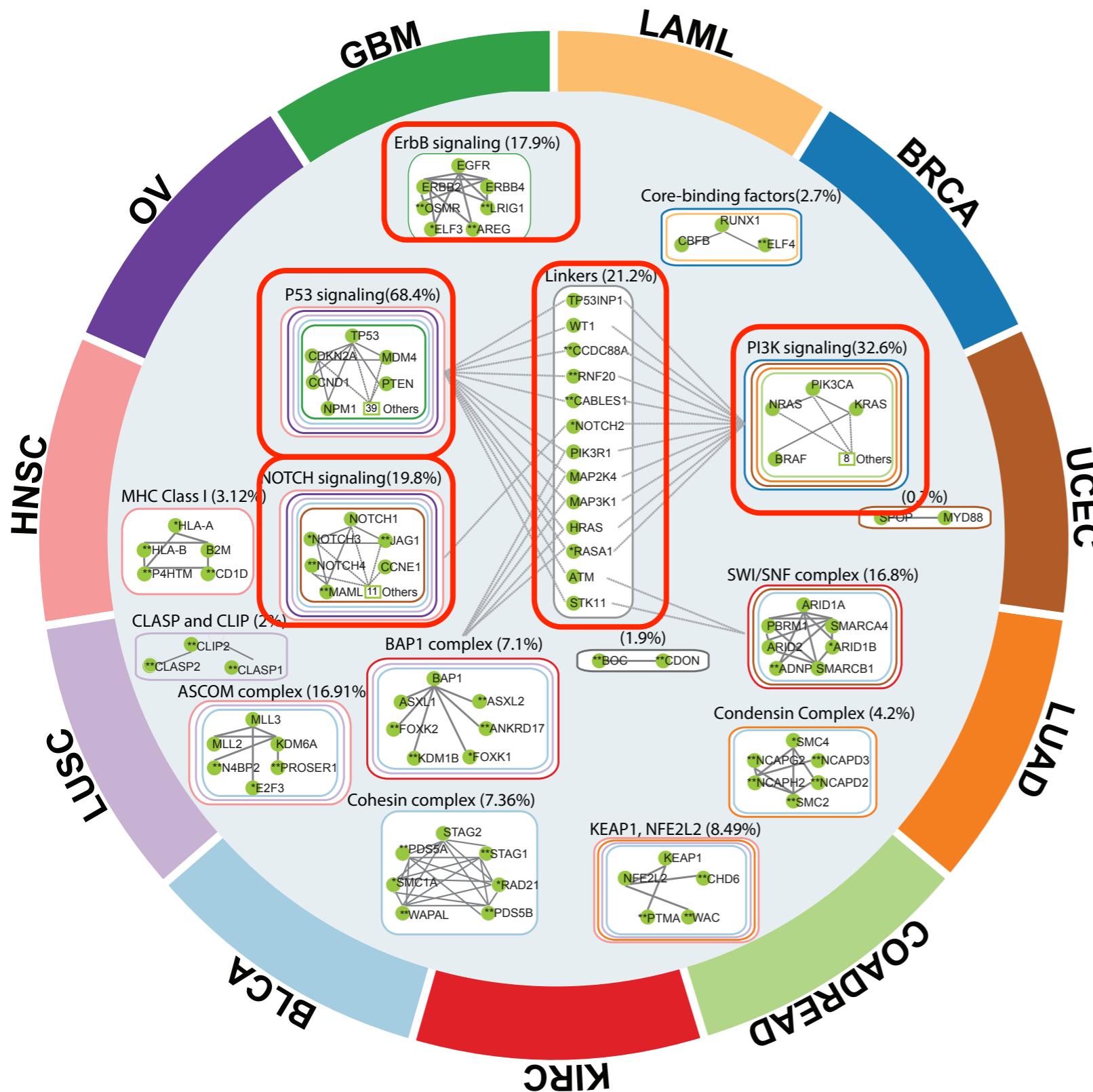
16 consensus subnetworks with ≥ 4 genes ($P=0.004$)
13 “linkers” between consensus subnetworks

HotNet2 Consensus Subnetworks

Frequently and rarely mutated cancer genes



HotNet2 Consensus Subnetworks

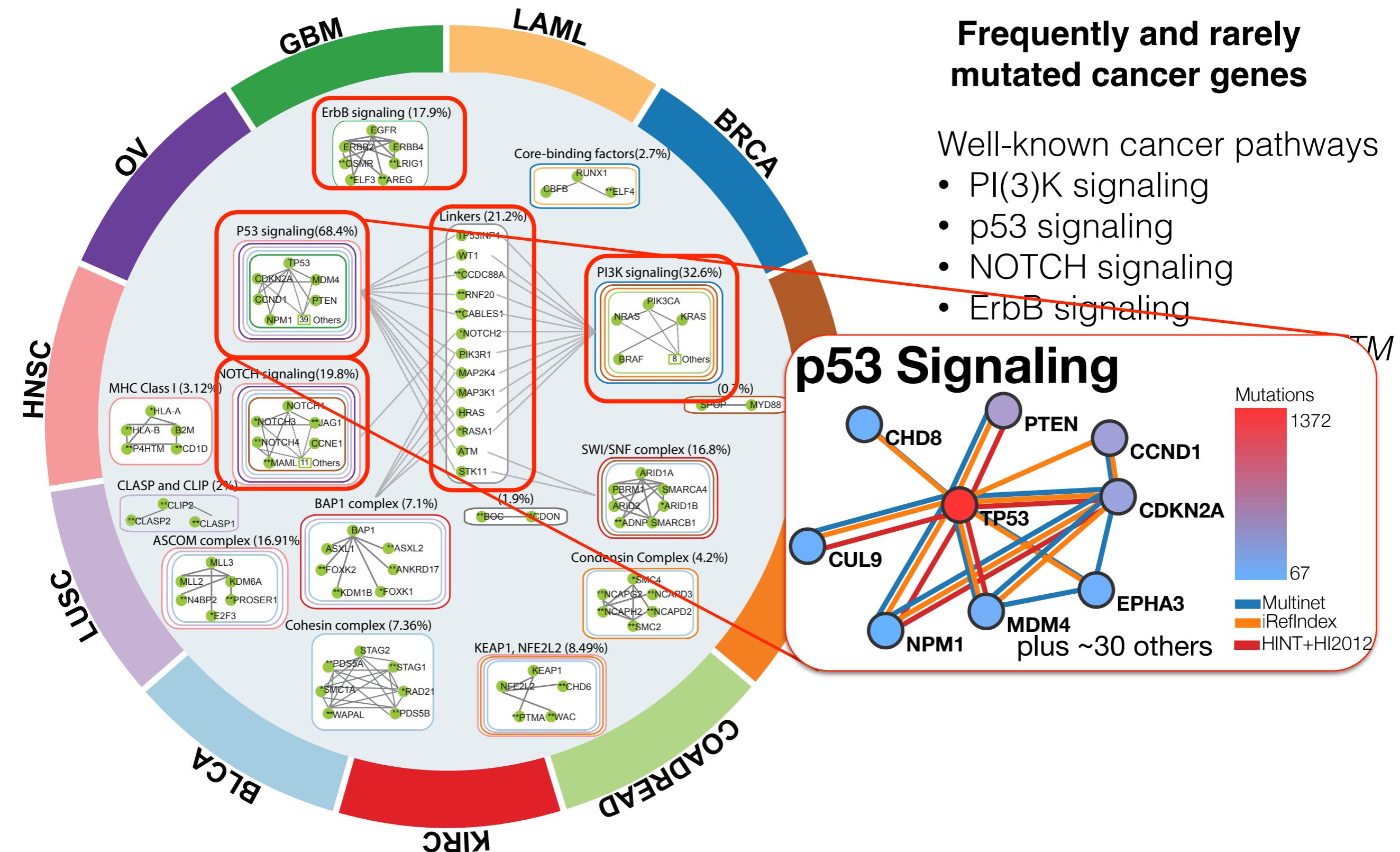


Frequently and rarely mutated cancer genes

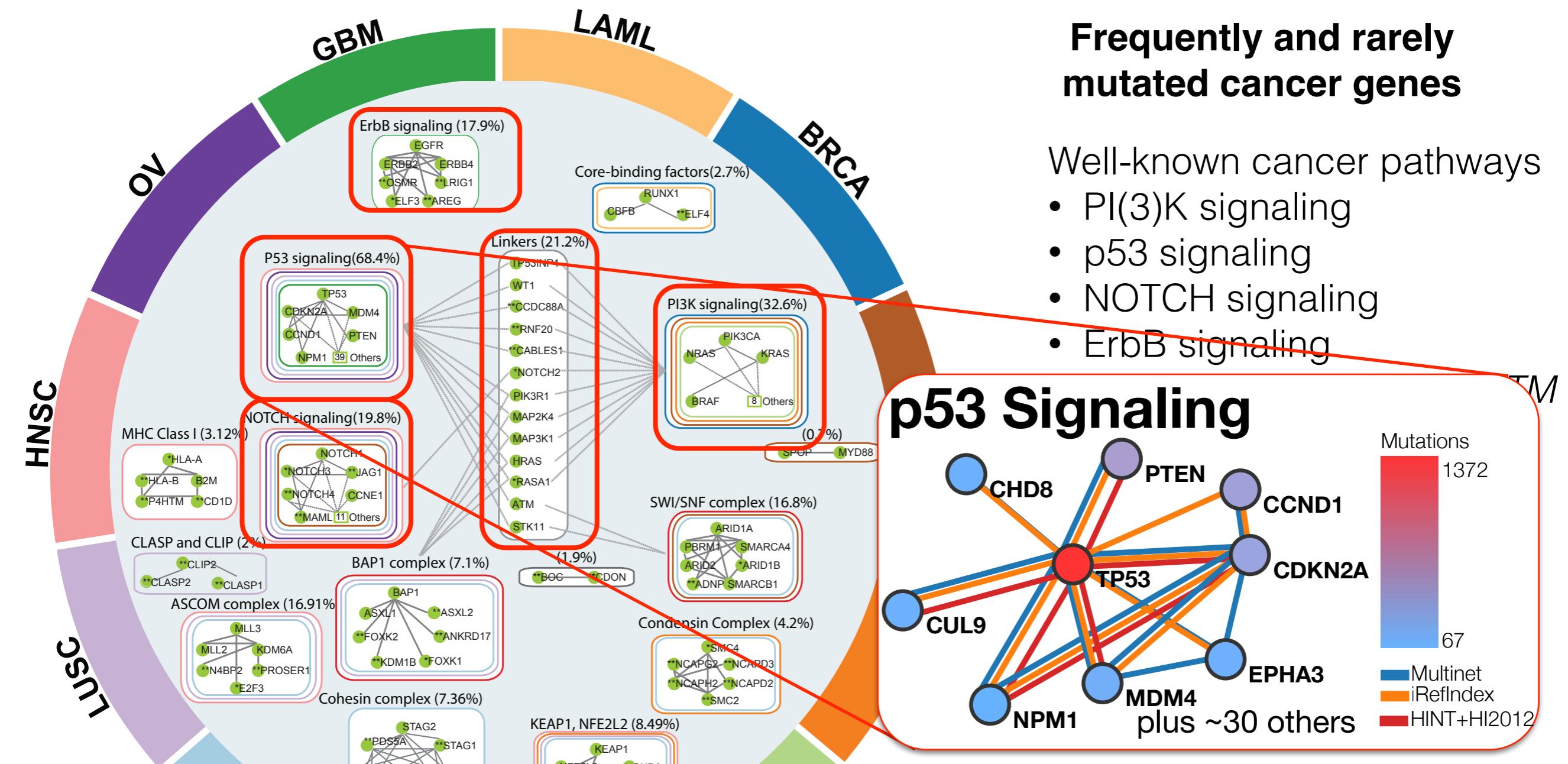
Well-known cancer pathways

- PI(3)K signaling
- p53 signaling
- NOTCH signaling
- ErbB signaling
- Linkers: *HRAS*, *STK11*, *ATM*

HotNet2 Consensus Subnetworks



HotNet2 Consensus Subnetworks



CHD8 (mutated in 54 of 3110 samples)

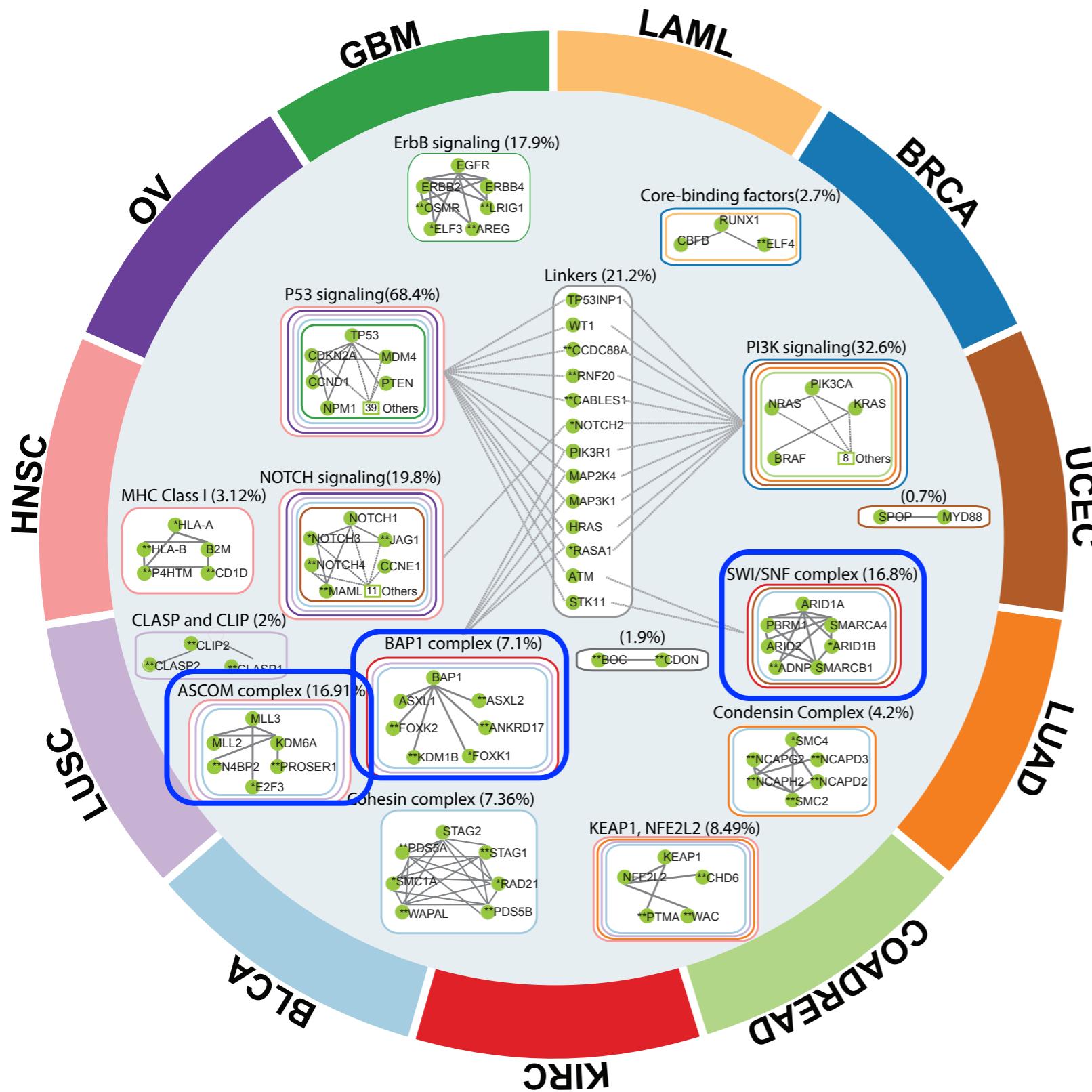
Legend
Cancer types

BLCA	BRCA	COADREAD	GBM	HNSC
KIRC	LAML	LUAD	LUSC	OV

Mutation types

Inactivating SNV	SNV
Amplification	Deletion

HotNet2 Consensus Subnetworks



Frequently and rarely mutated cancer genes

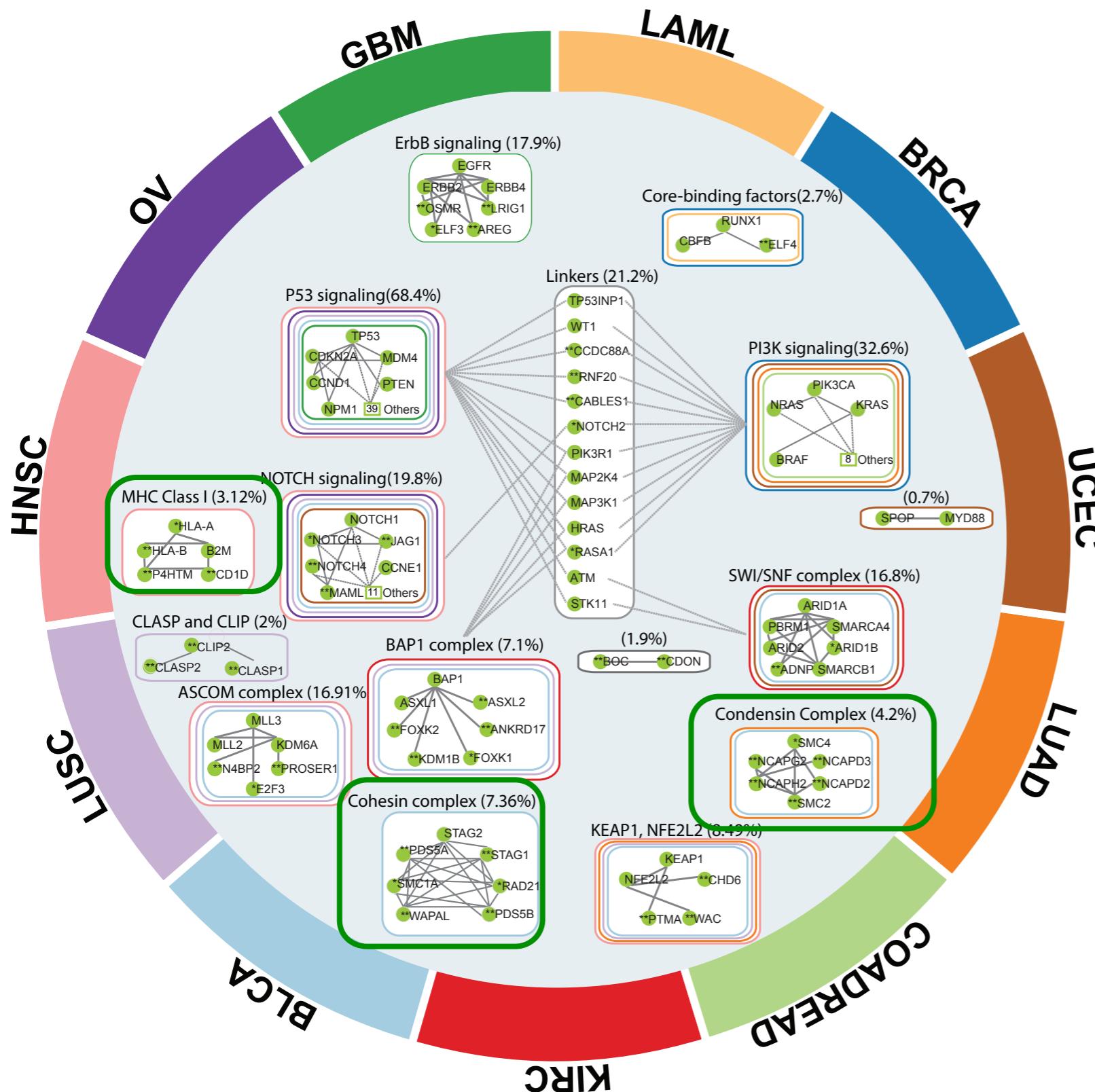
Well-known cancer pathways

- PI(3)K signaling
- p53 signaling
- NOTCH signaling
- ErbB signaling
- Linkers: *HRAS*, *STK11*, *ATM*

Recently characterized complexes:

- SWI/SNF complex
- ASCOM complex
- BAP1 complex

HotNet2 Consensus Subnetworks



Frequently and rarely mutated cancer genes

Well-known cancer pathways

- PI(3)K signaling
- p53 signaling
- NOTCH signaling
- ErbB signaling
- Linkers: *HRAS*, *STK11*, *ATM*

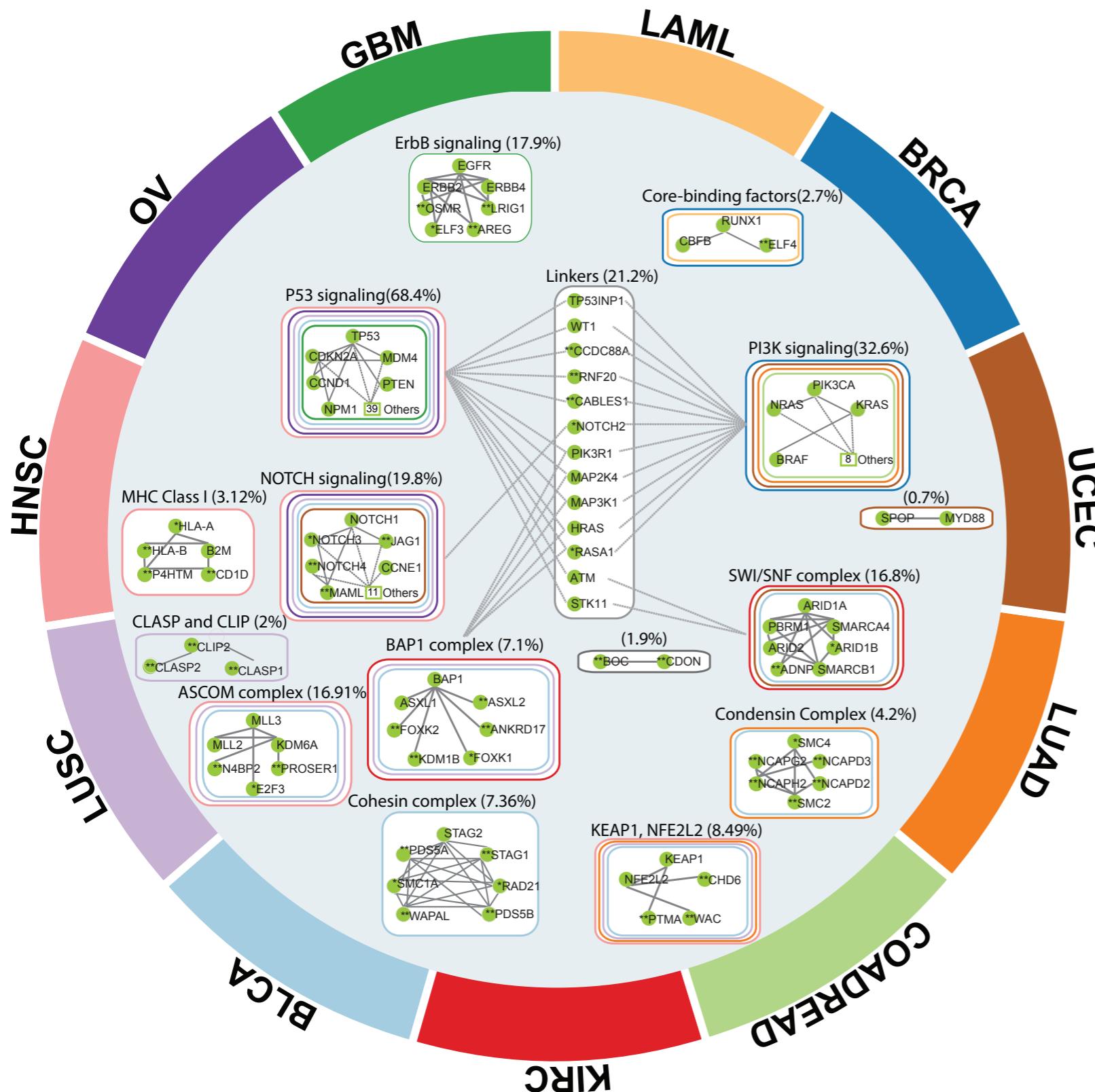
Recently characterized complexes:

- SWI/SNF complex
- ASCOM complex
- BAP1 complex

Potentially novel complexes:

- Cohesin complex
- Condensin complex
- MHC Class I proteins

HotNet2 Consensus Subnetworks



Frequently and rarely mutated cancer genes

Well-known cancer pathways

- PI(3)K signaling
- p53 signaling
- NOTCH signaling
- ErbB signaling
- Linkers: *HRAS*, *STK11*, *ATM*

Recently characterized complexes:

- SWI/SNF complex
- ASCOM complex
- BAP1 complex

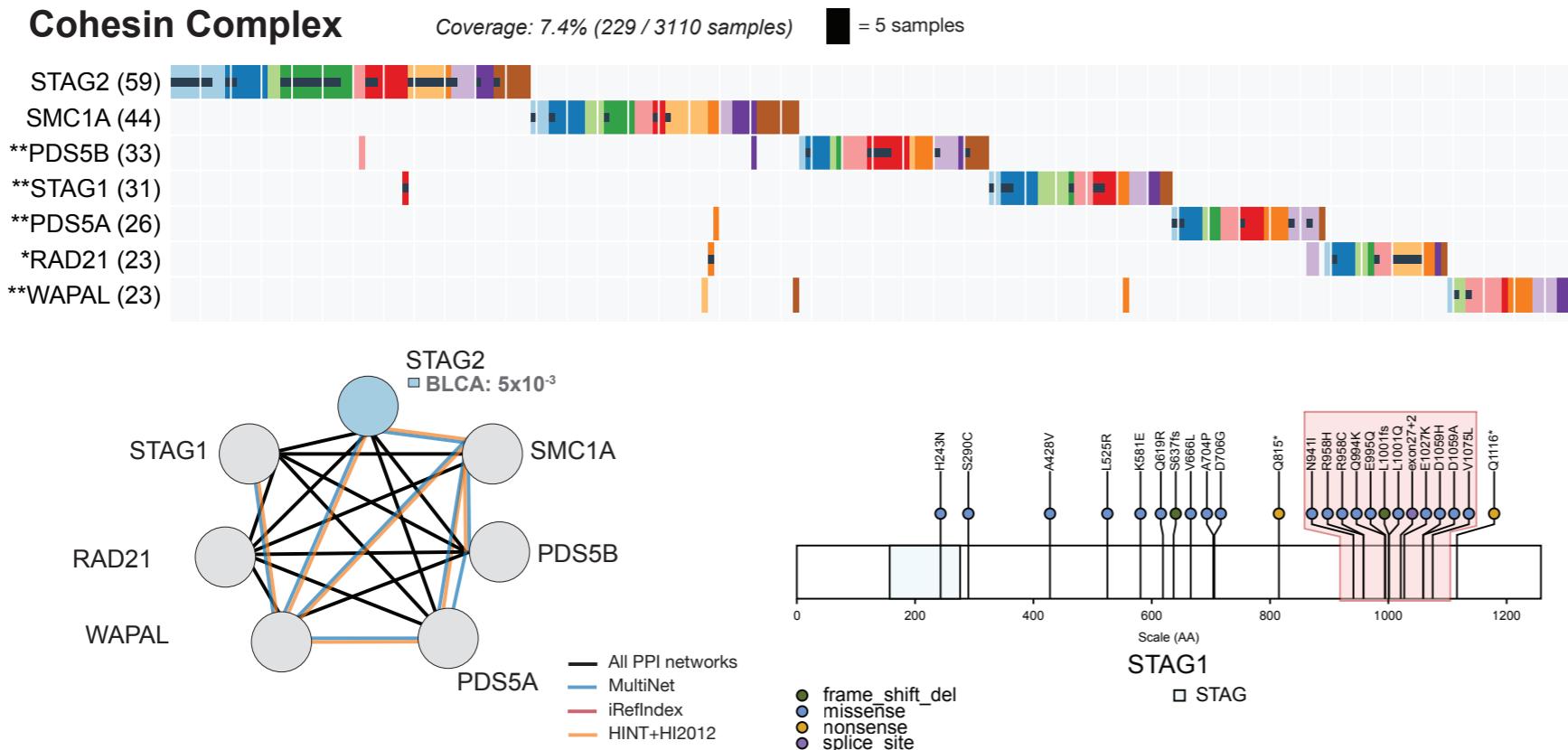
Potentially novel complexes:

- Cohesin complex
- Condensin complex
- MHC Class I proteins



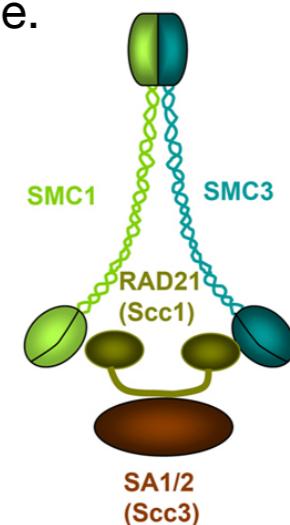
Cohesin and condensin complexes

Cohesin Complex



Cohesin complex

- 4/5 members of complex
- Involved in sister chromatid cohesion and gene regulation.
- Mutated in >4% of samples in each cancer type.

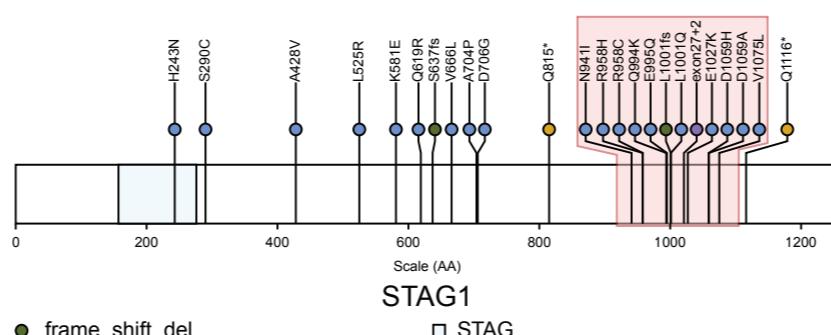
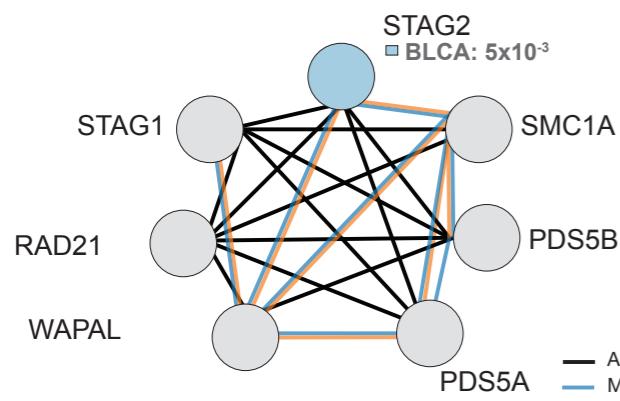
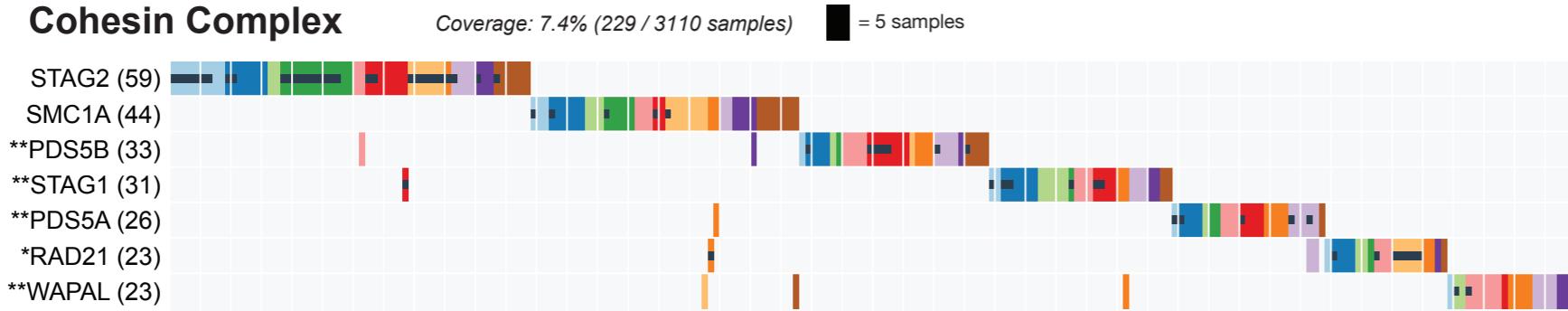


Legend

Cancer types	BLCA	BRCA	COADREAD	GBM	HNSC	Inactivating SNV	SNV
KIRC	LAML	LUAD	LUSC	OV	UCEC	Amplification	Deletion

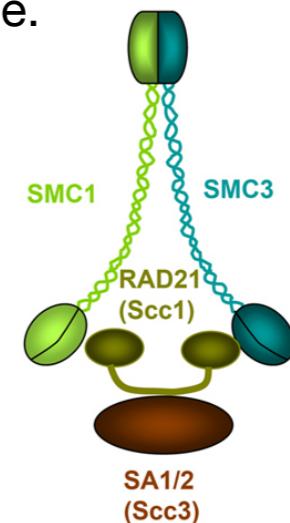
Cohesin and condensin complexes

Cohesin Complex

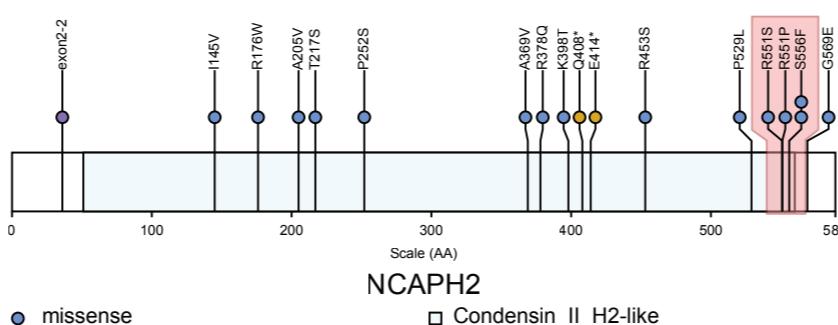
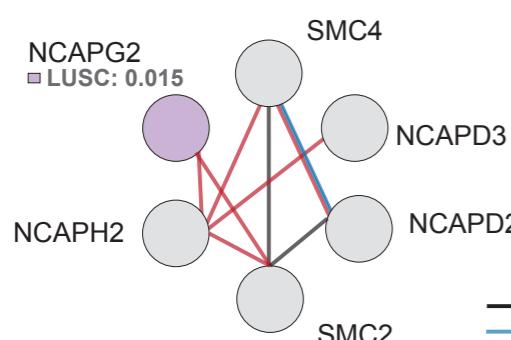
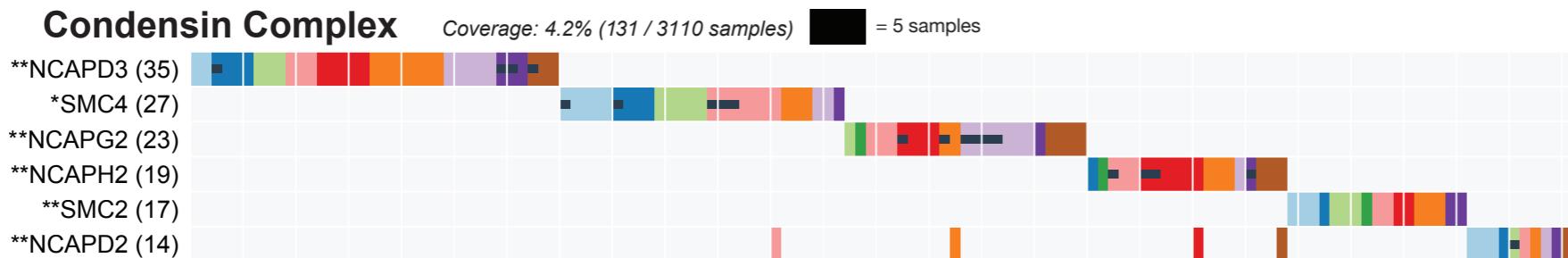


Cohesin complex

- 4/5 members of complex
- Involved in sister chromatid cohesion and gene regulation.
- Mutated in >4% of samples in each cancer type.



Condensin Complex



Condensin complex

- 6/8 members of complex
- Involved in sister chromatid condensation and gene regulation.
- Somatic mutations and expression validated using whole-genome sequencing and RNA-Seq

Legend

Cancer types	BLCA	BRCA	COADREAD	GBM	HNSC	Inactivating SNV	SNV
	KIRC	LAML	LUAD	LUSC	OV	Amplification	Deletion

Cohesin and condensin complexes

Cohesin Complex Coverage: 7.4% (229 / 3110 samples) █ = 5 samples

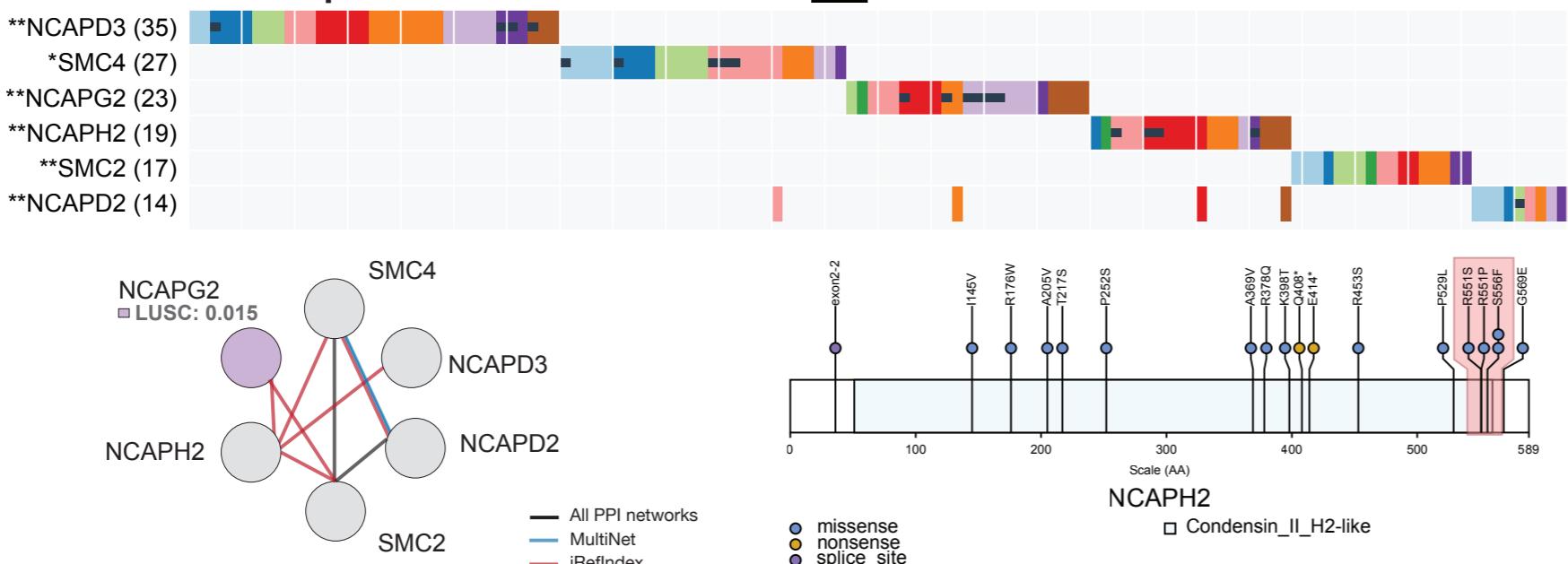
Combinations of Mutually Exclusive Alterations (CoMEt)

Leiserson, Wu, et al. RECOMB & Genome Biology (2015)

- Statistical score for mutually exclusive mutations
- MCMC algorithm for identifying combinations
- Software:
 - Python (<https://github.com/raphael-group/comet>)
 - R (<http://cran.r-project.org/package=cometExactTest>)



Condensin Complex Coverage: 4.2% (131 / 3110 samples) █ = 5 samples



Cohesin Complex Coverage: 7.4% (229 / 3110 samples) █ = 5 samples
4/5 members of complex involved in sister chromatid cohesion and gene regulation.
Mutated in >4% of samples across cancer type.



Hsin-Ta Wu

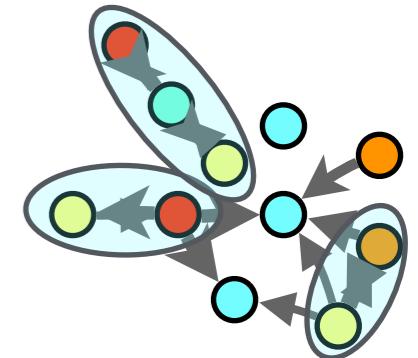
Condensin complex

- 6/8 members of complex
- Involved in sister chromatid condensation and gene regulation.
- Somatic mutations and expression validated using whole-genome sequencing and RNA-Seq

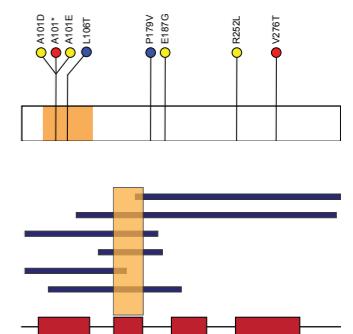
Legend	Cancer types								Mutation types				
	BLCA	BRCA	COADREAD	GBM	HNSC	KIRC	LAML	LUAD	LUSC	OV	UCEC	Inactivating SNV	SNV
												Amplification	Deletion

Outline

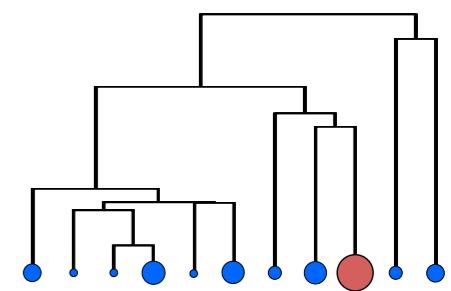
1. A new algorithm, HotNet2.



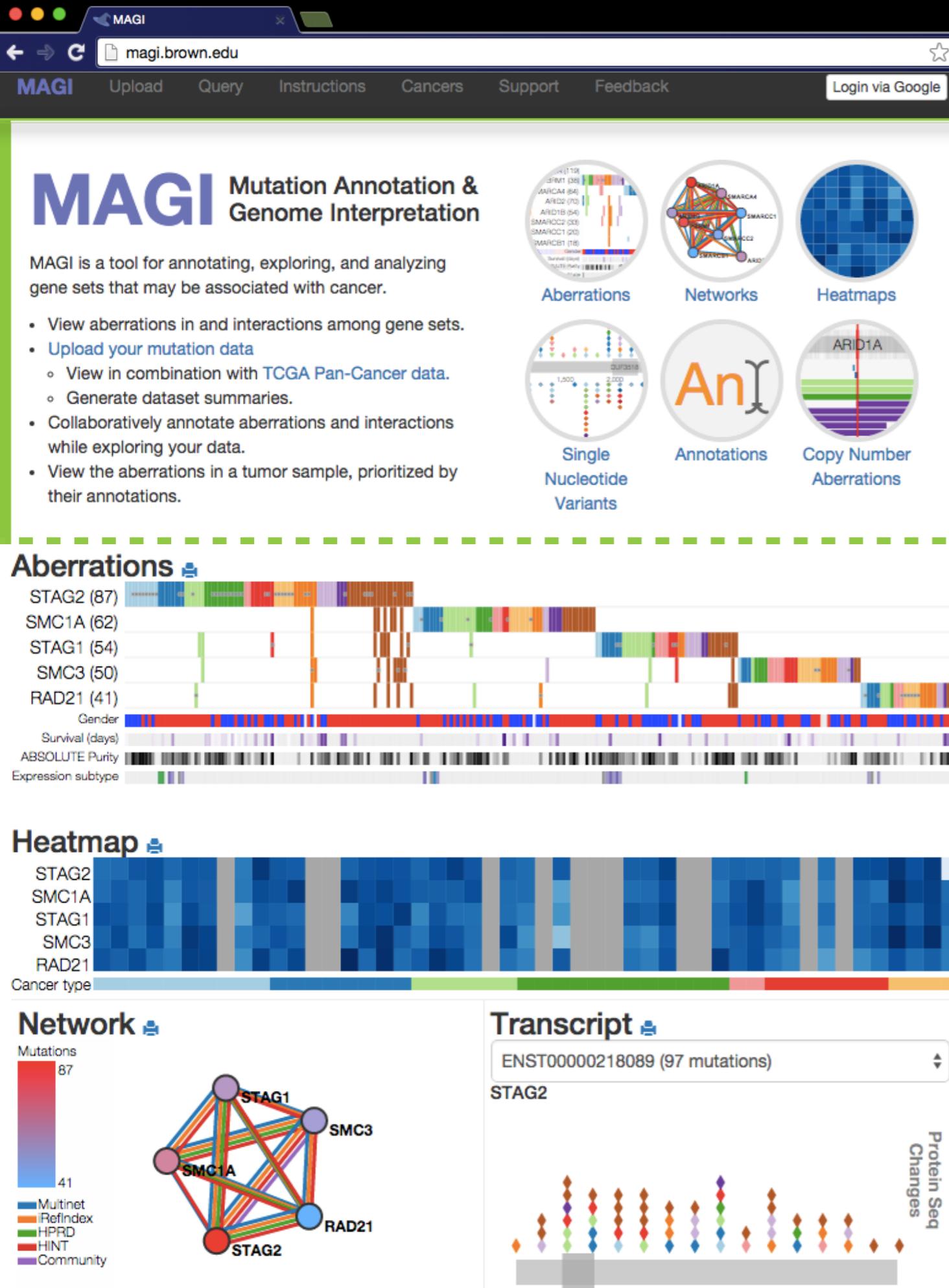
2. Comparison of HotNet2 to similar methods.



3. Application to TCGA Pan-Cancer data.



4. Latest work.



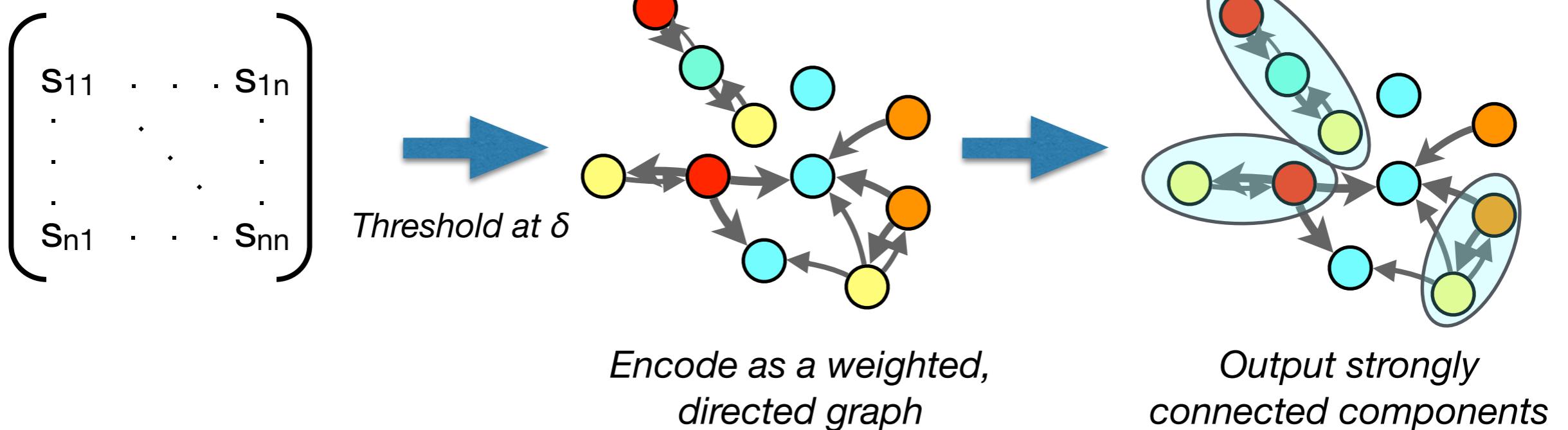
MAGI: Web application for exploring and annotating cancer genomics data.



- Mutations are annotated with literature citations.
- Upload your mutations in a simple web form.
- Combine your data with TCGA data.

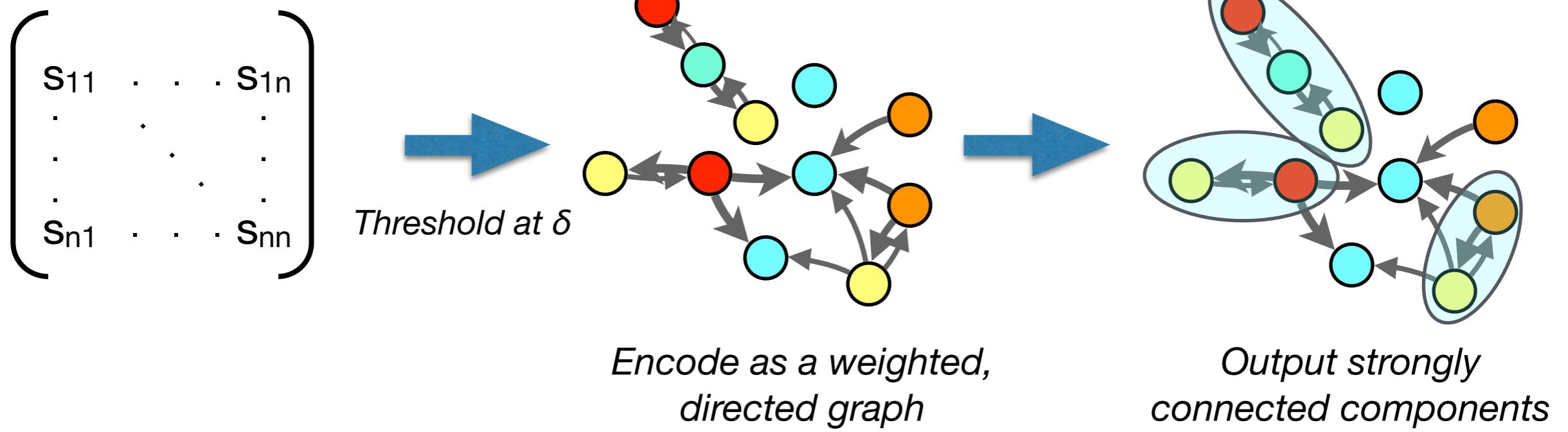
Strongly connected components form a hierarchy

HotNet2: identifying “hot spots” in the network

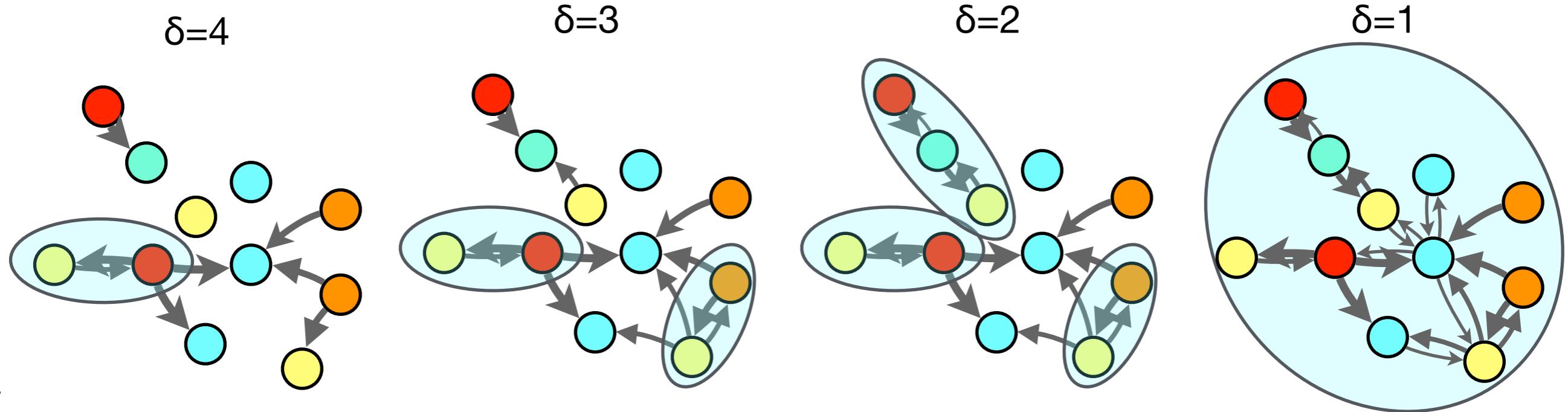


Strongly connected components form a hierarchy

HotNet2: identifying “hot spots” in the network

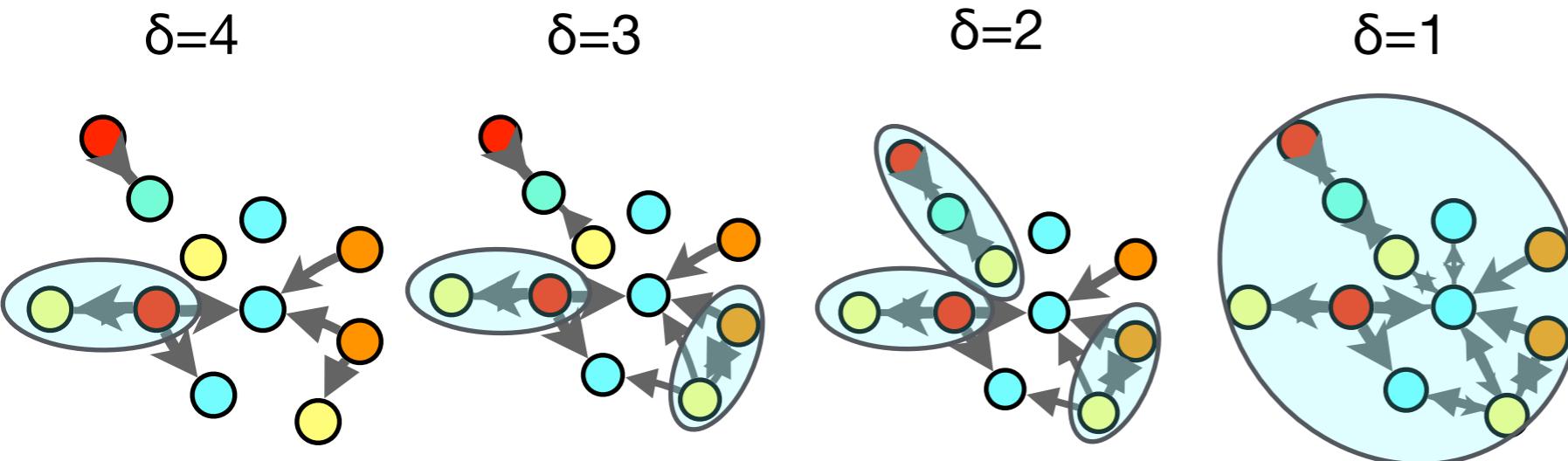


Components merge as δ decreases



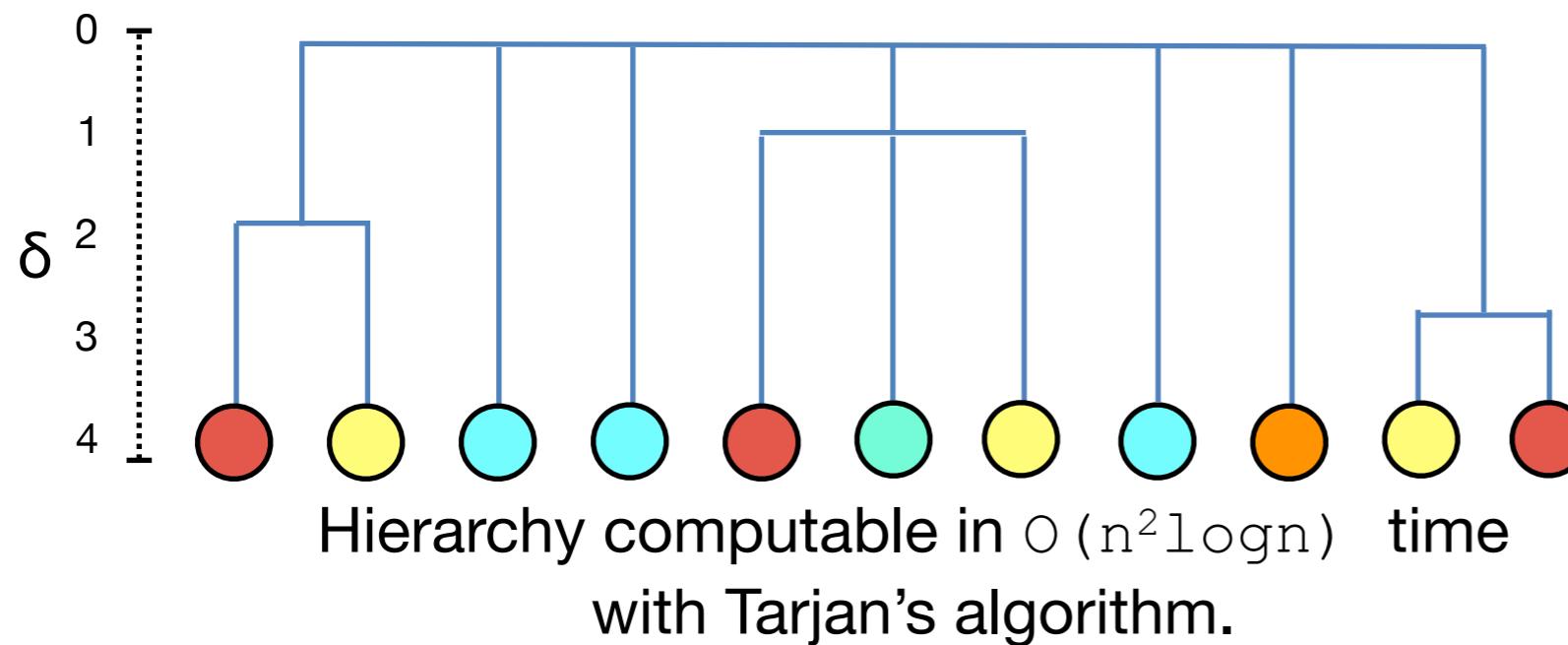
Strongly connected components form a hierarchy

Components merge as δ decreases



Every value of δ gives a partition of the vertices into strongly connected components.

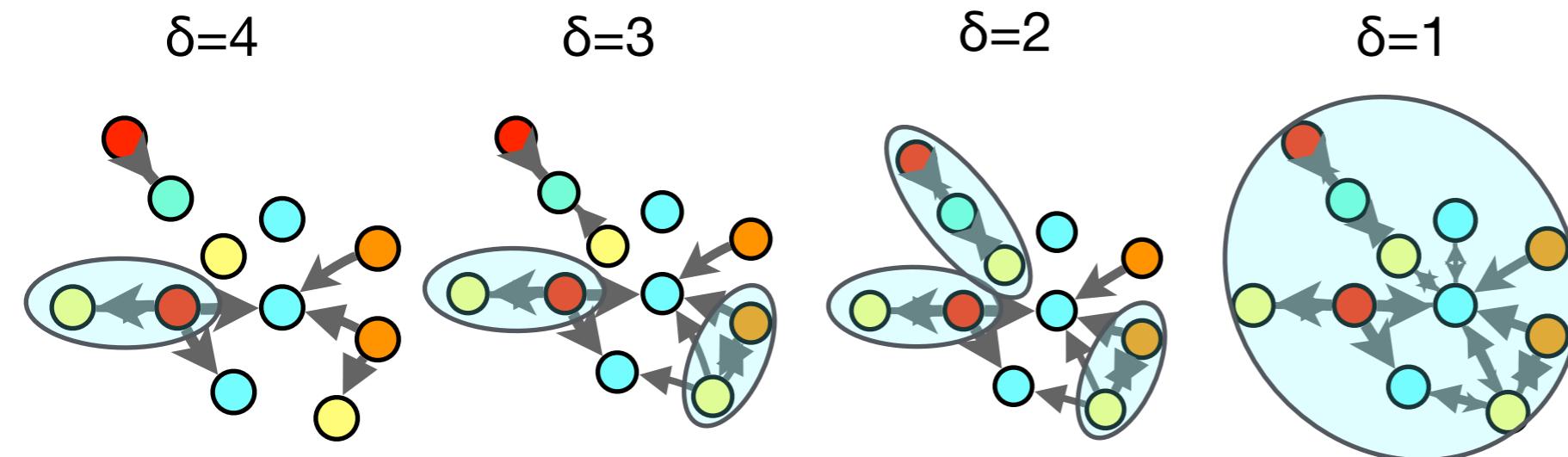
Hierarchy encodes mutations and topology



Matt Reyna

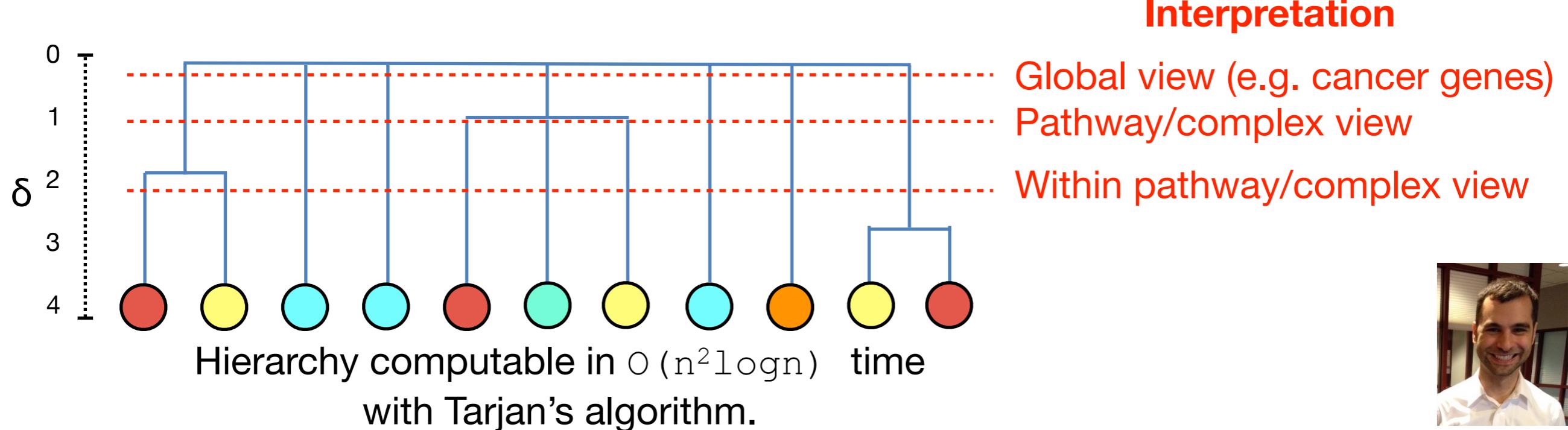
Strongly connected components form a hierarchy

Components merge as δ decreases



Every value of δ gives a partition of the vertices into strongly connected components.

Hierarchy encodes mutations and topology

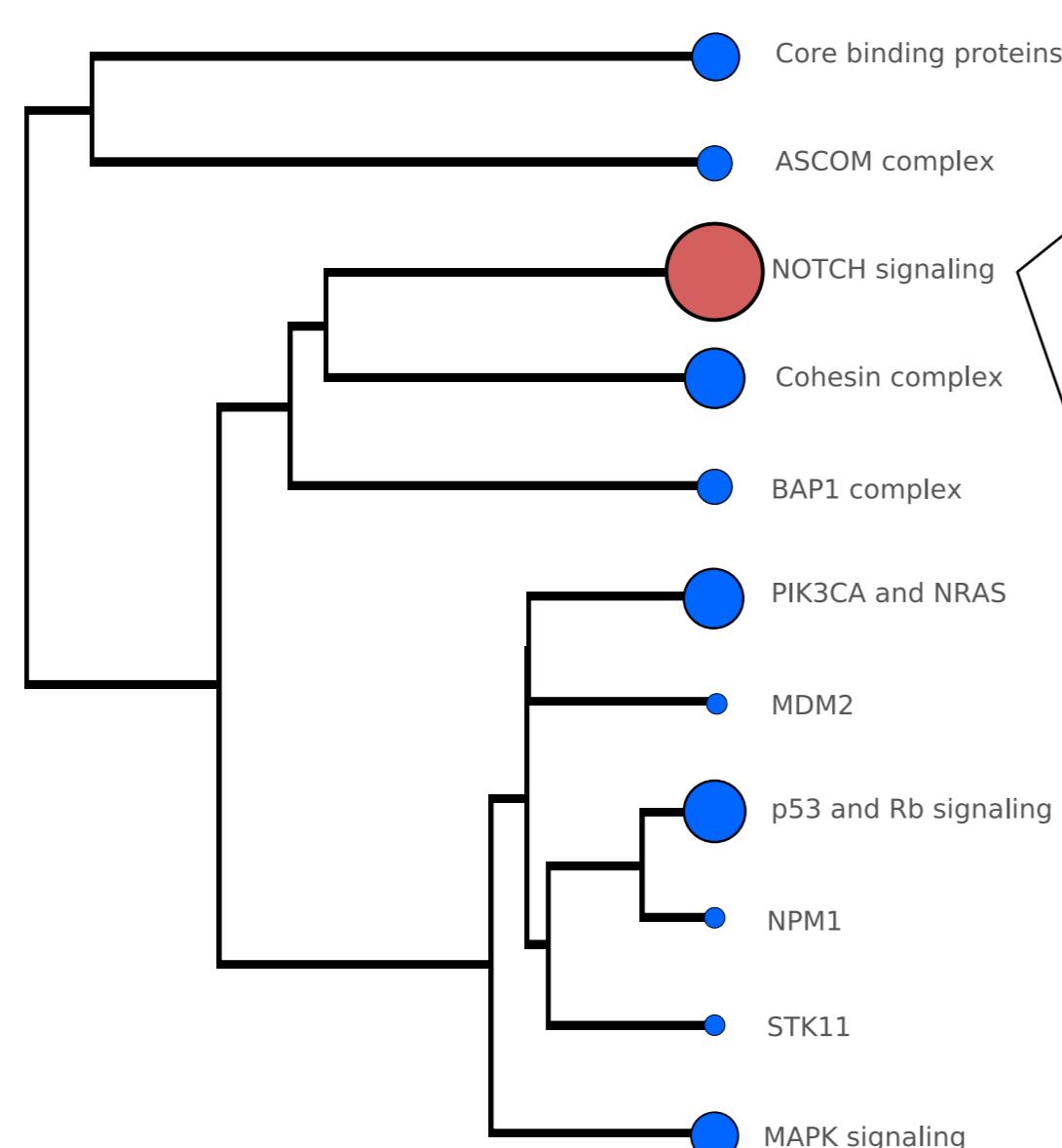


Matt Reyna

Hierarchies of HotNet2 subnetworks

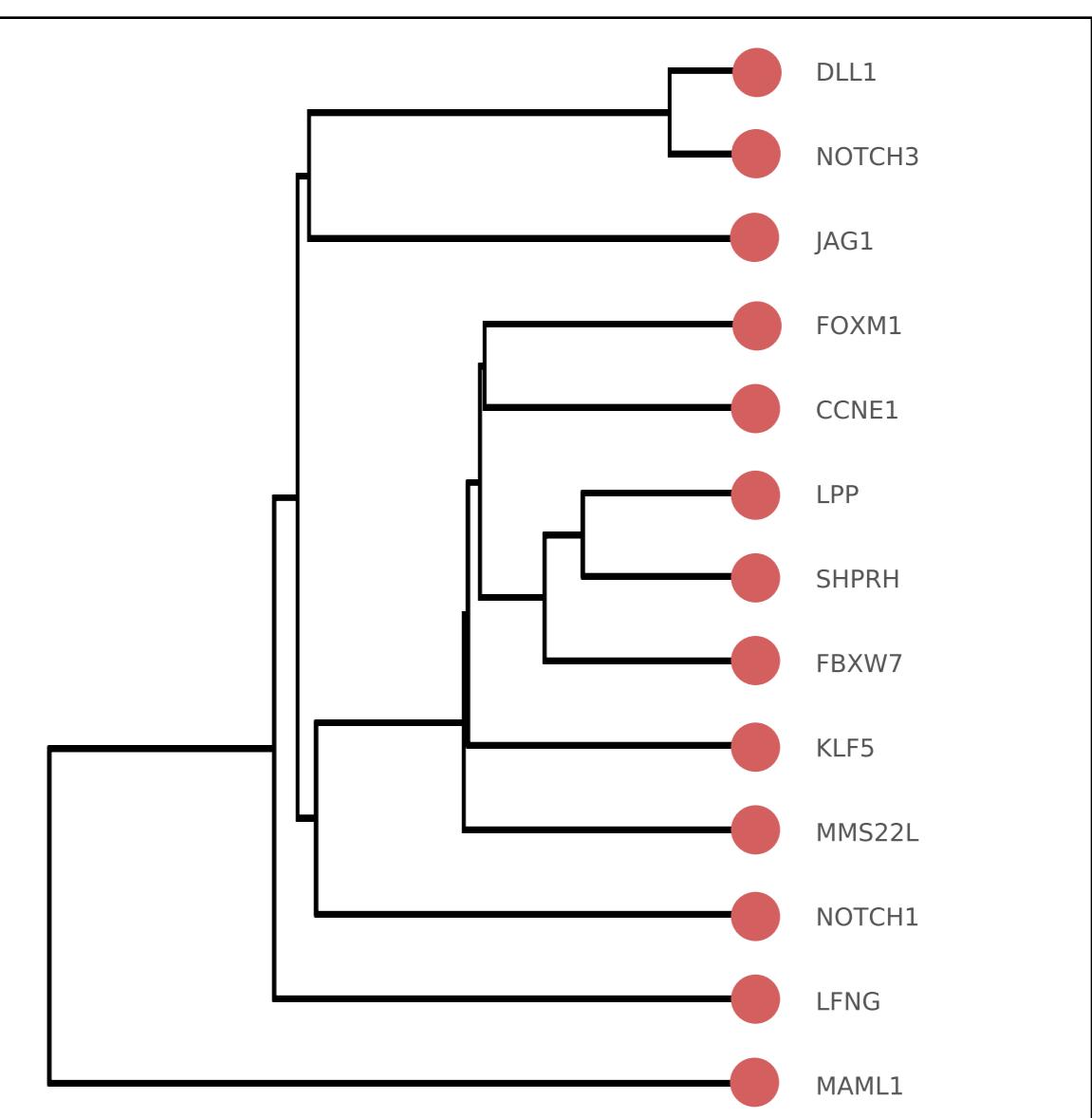
Across subnetworks

*HotNet2 results on
TCGA Pan-Cancer data*

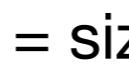
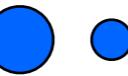


Within subnetwork

*Subnetwork overlapping the
NOTCH signaling pathway*



δ



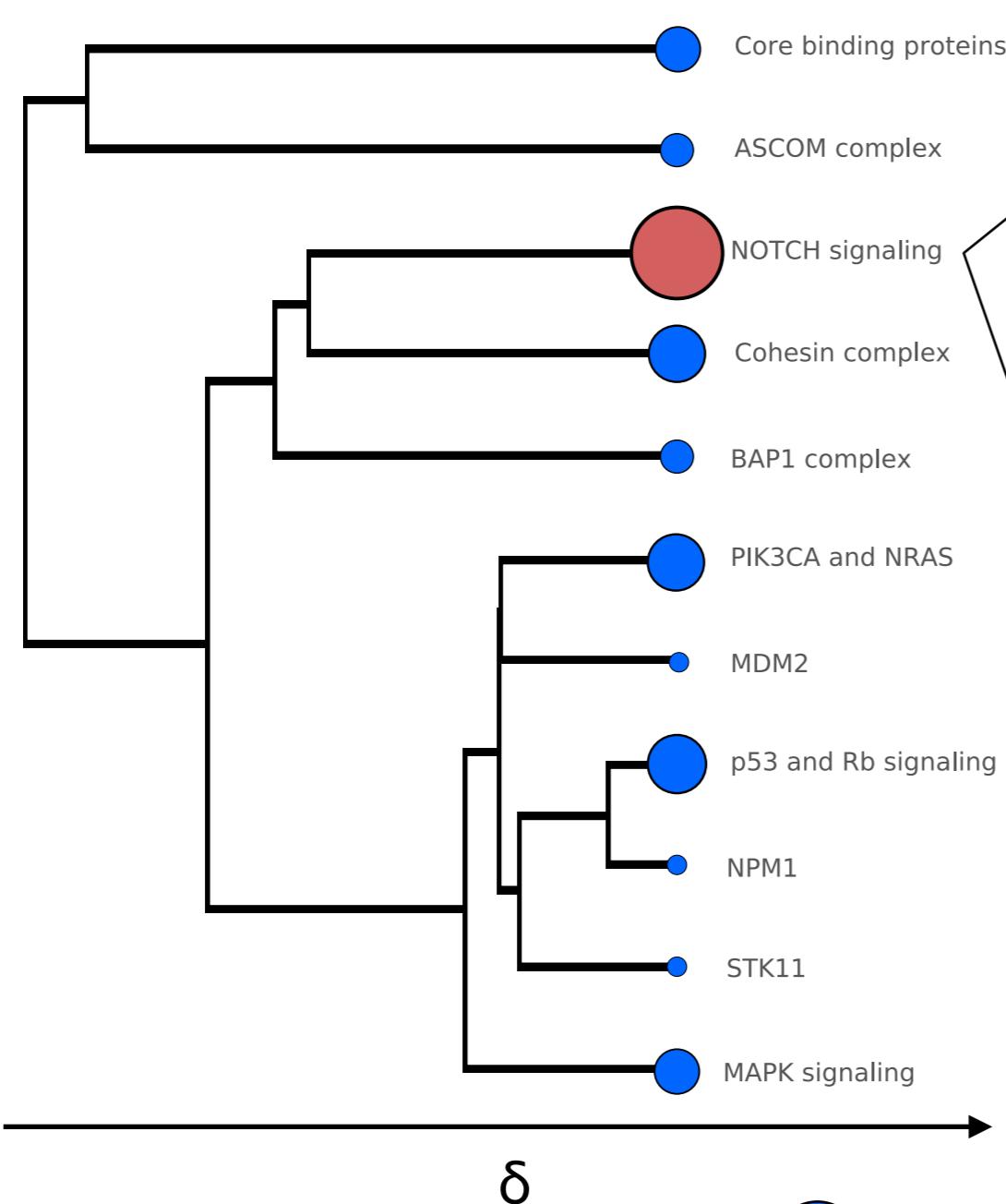
= size of subnetwork

δ

Hierarchies of HotNet2 subnetworks

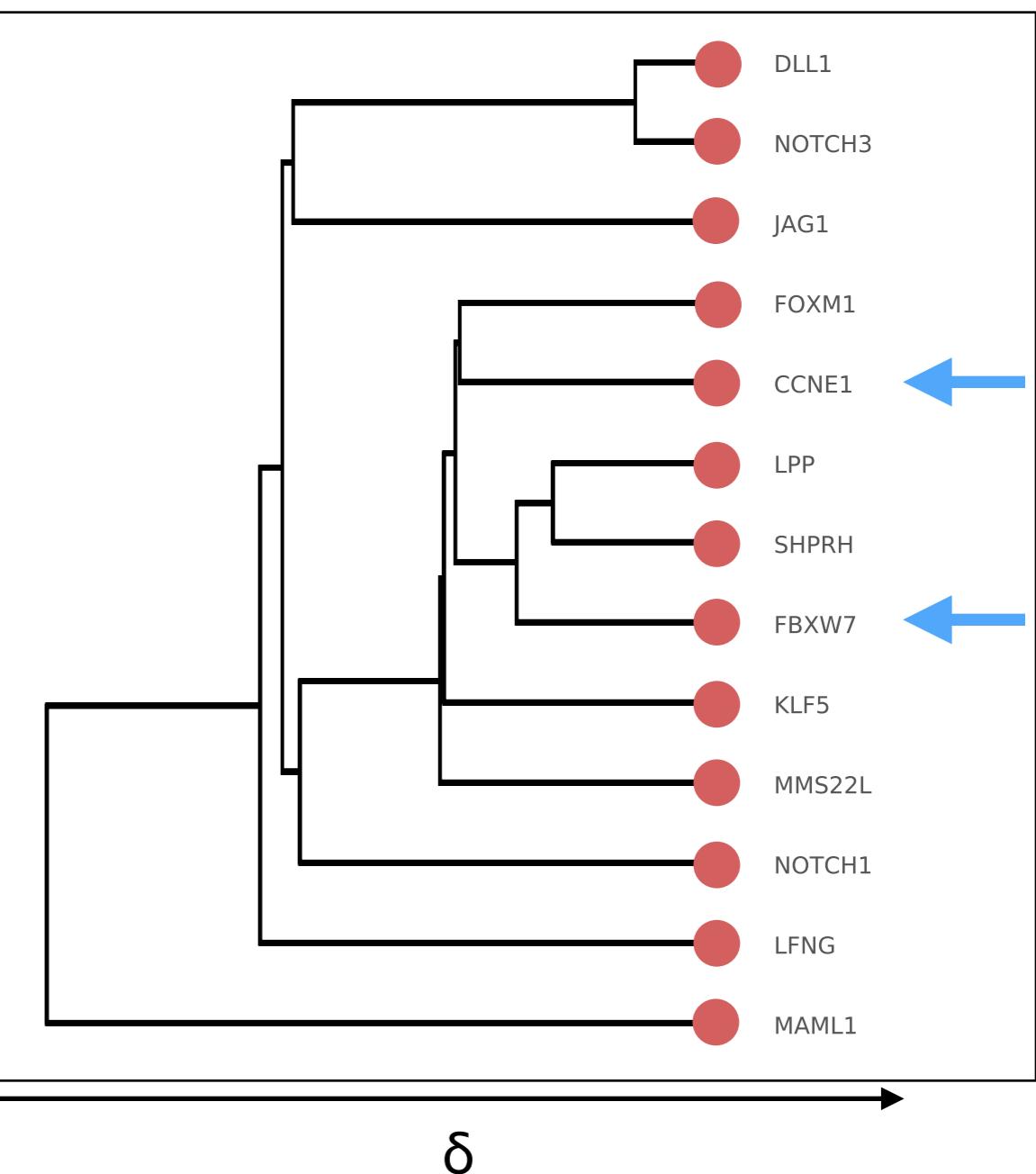
Across subnetworks

*HotNet2 results on
TCGA Pan-Cancer data*



Within subnetwork

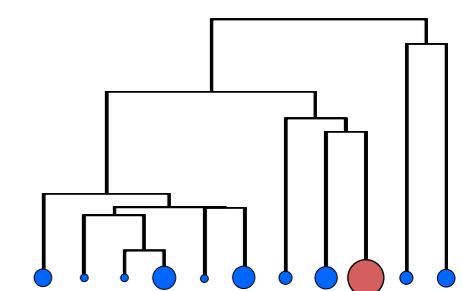
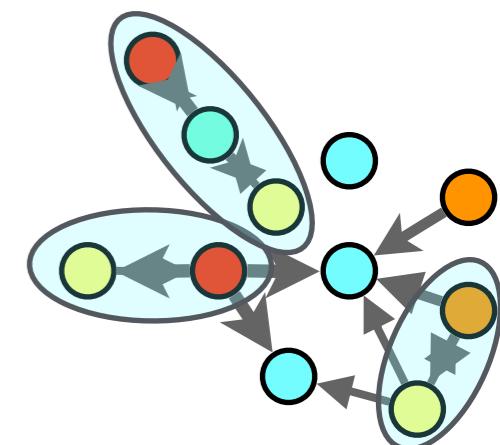
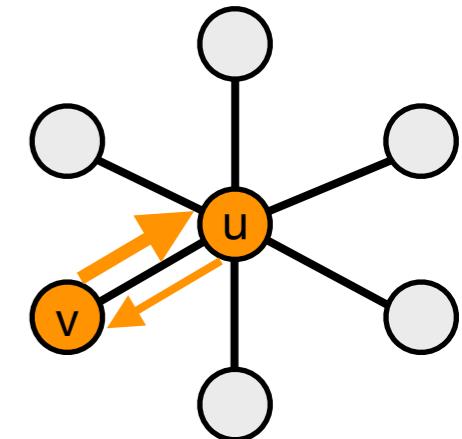
*Subnetwork overlapping the
NOTCH signaling pathway*



● ● ● = size of subnetwork

Summary

- HotNet2: Novel algorithm that analyzes topology and mutations simultaneously with asymmetric heat diffusion.
- Identifies known and novel pathways and complexes with frequently and rarely mutated genes on TCGA Pan-Cancer data.
- Recent work: CoMEt algorithm and MAGI web server
- Ongoing work: Hierarchies of strong components?



Acknowledgements

Research Group

Ben Raphael

Fabio Vandin

Hsin-Ta Wu

Jason R. Dobson

Matt Reyna

Jonathan V. Eldridge

Alexandra Papoutsaki

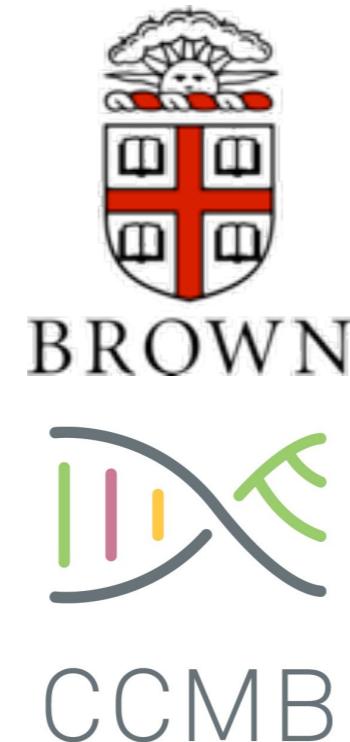
Jacob Thomas

Younhun Kim

Connor Gramazio

David Laidlaw

Jason Hu



<http://compbio.cs.brown.edu/software/>

Funding & Data

*Travel funding to ISMB/ECCB 2015
was generously provided by ISCB*



National Science Foundation
WHERE DISCOVERIES BEGIN
The Cancer Genome Atlas

Collaborators

Beifang Niu

Michael McLellan

Li Ding

Michael Lawrence

Gad Getz

Nuria Lopez-Bigas

Abel Gonzalez-Perez

David Tamborero

Yuwei Chang

Greg Ryslik



Yale University

