

Pathway analysis to connect information across cancer genomes

ICGC BAWG Telecon, July 12, 2013

**Josh Stuart, Professor
UC Santa Cruz**



With Mutations, Context Matters

- There are many recurrent, but low frequency mutations that are not well characterized.
- Some patients with “right” mutation don’t respond. Why?
- Many cancers have one of several “novel” mutations. Can these be targeted with current approaches?
- The mode of action, loss-of-function or gain-of-function (LOF/GOF), of mutations can improve our understanding of disease mechanisms and treatment.
- Pathway-based methods can complement prediction of LOF or GOF.
 - Complimentary to LoH, methylation, amplification, ...



Two Themes



1) Predict
Drivers



2) Predict
Essential



Two Themes



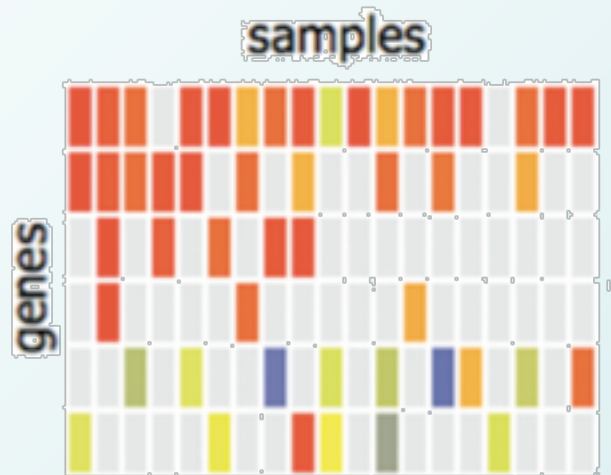
1) Predict
Drivers



2) Predict
Essential



Predicting Drivers w/ Frequency Analysis



- Recurrently mutated genes are more likely to be impactful
- Correct background mutation rate?
- Biased for “early” rather than “late” drivers?



Predicting Drivers w/ Frequency Analysis

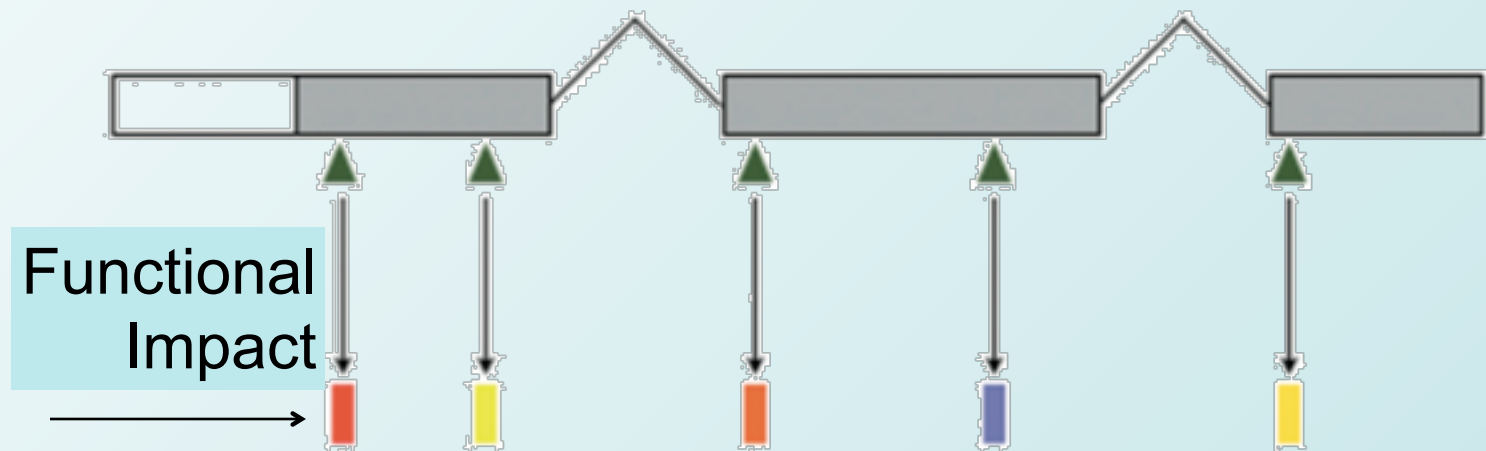


- Recurrently mutated genes are more likely to be impactful
- Correct background mutation rate?
- Biased for “early” rather than “late” drivers?

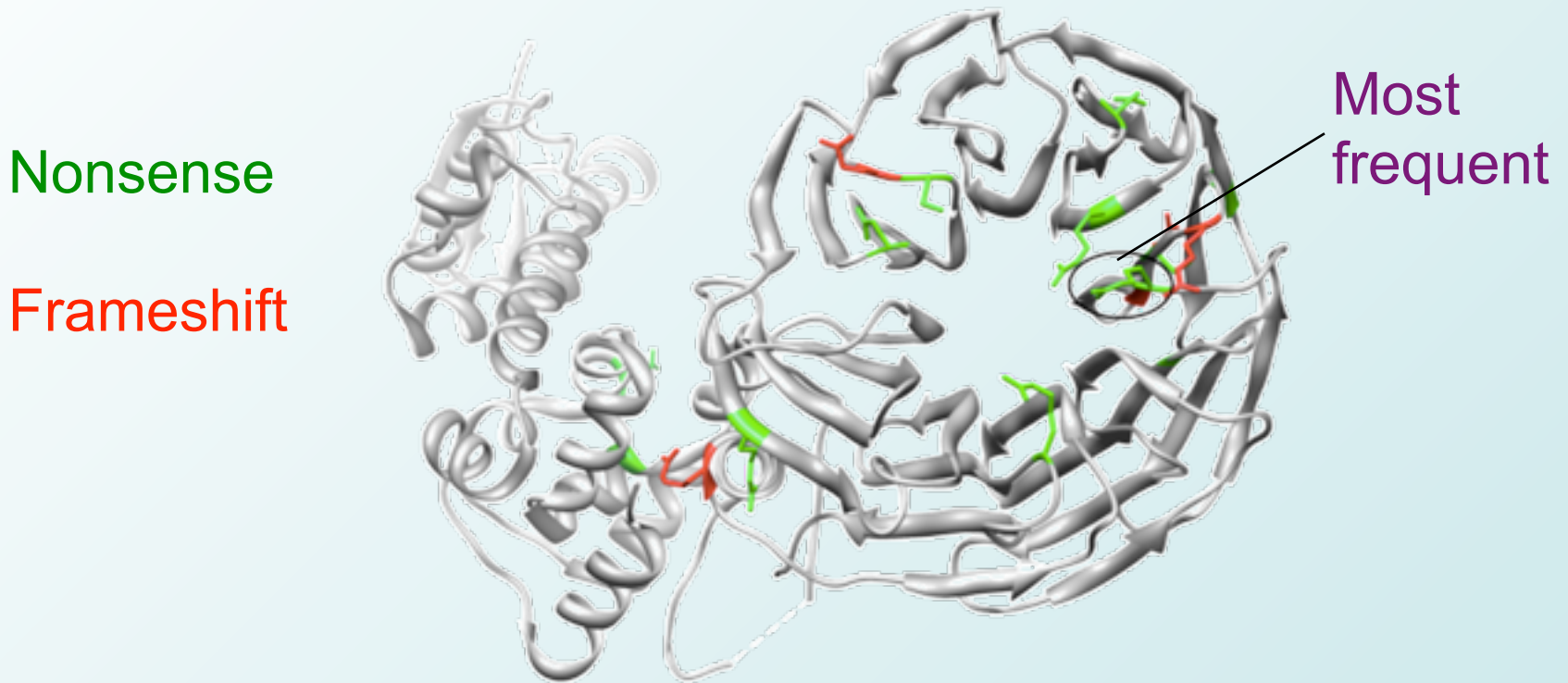


Predicting Drivers with Sequence Analysis

- Protein domains
 - Disruptions in key domains (e.g DNA binding, kinase-interacting)
- Conservation
 - Mutations in conserved residues or positions more important
 - Synonymous / non-synonymous ratio suggests selection
- Non-random patterning across protein sequence/structure.



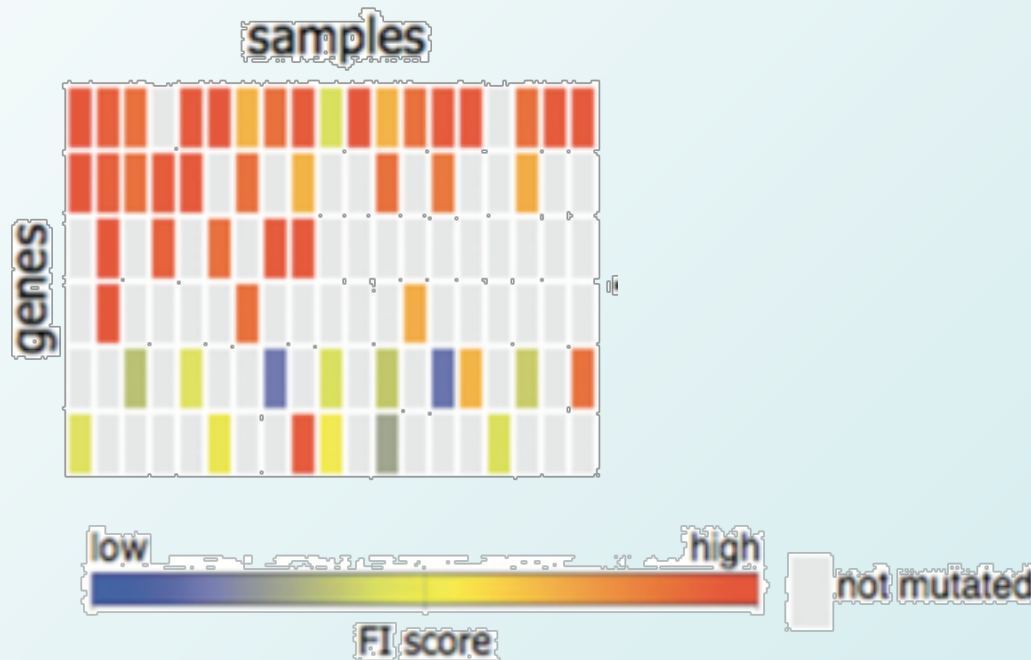
Methods to Gauge Mutation Impact



- FBXW7 mutations in colorectal cancer cluster in 3D space in beta propeller affecting its ability to ubiquitylate



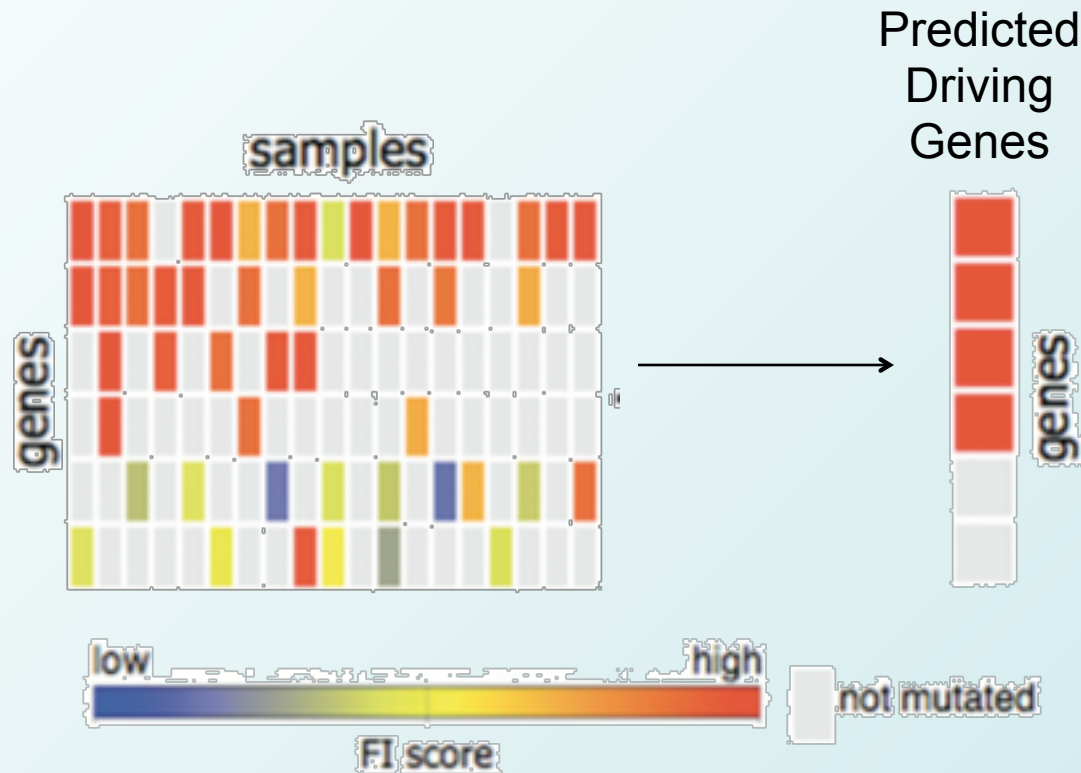
Predicting Drivers with a Combination of Frequency and Sequence Analysis



- Predicted functional impact can be accumulated across samples
- Sensitive: Find more rare mutations w/ higher impact
- Specific: Weed out frequent, spurious mutations (e.g. TITAN)



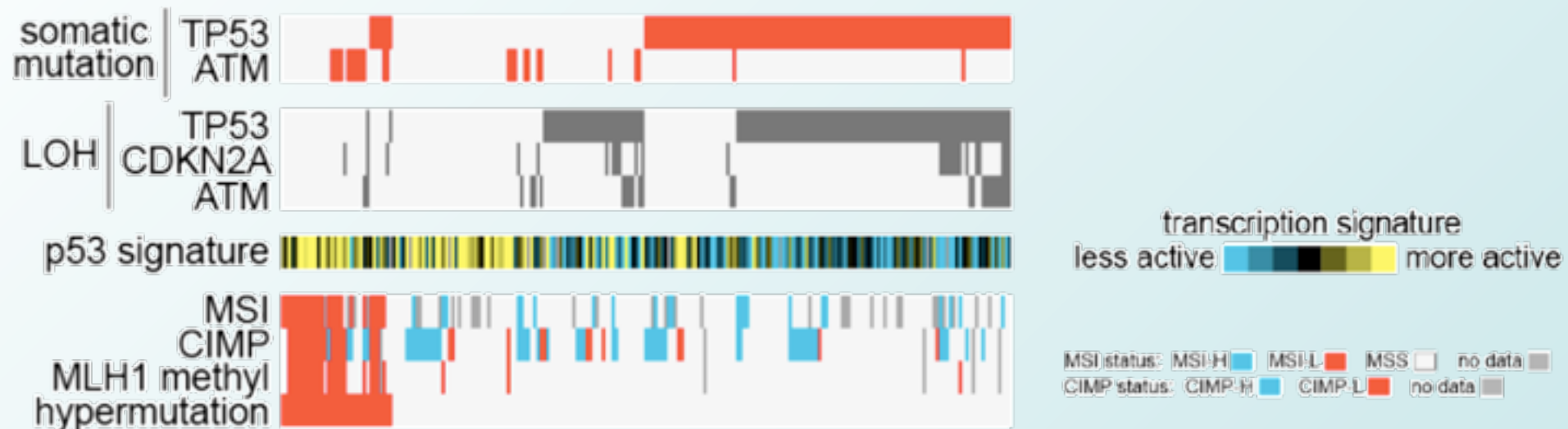
Predicting Drivers with a Combination of Frequency and Sequence Analysis



- Predicted functional impact can be accumulated across samples
- Sensitive: Find more rare mutations w/ higher impact
- Specific: Weed out frequent, spurious mutations (e.g. TITAN)

Gene Signatures of Mutations Shed Light on Impact

TP53 pathway activity in CRC



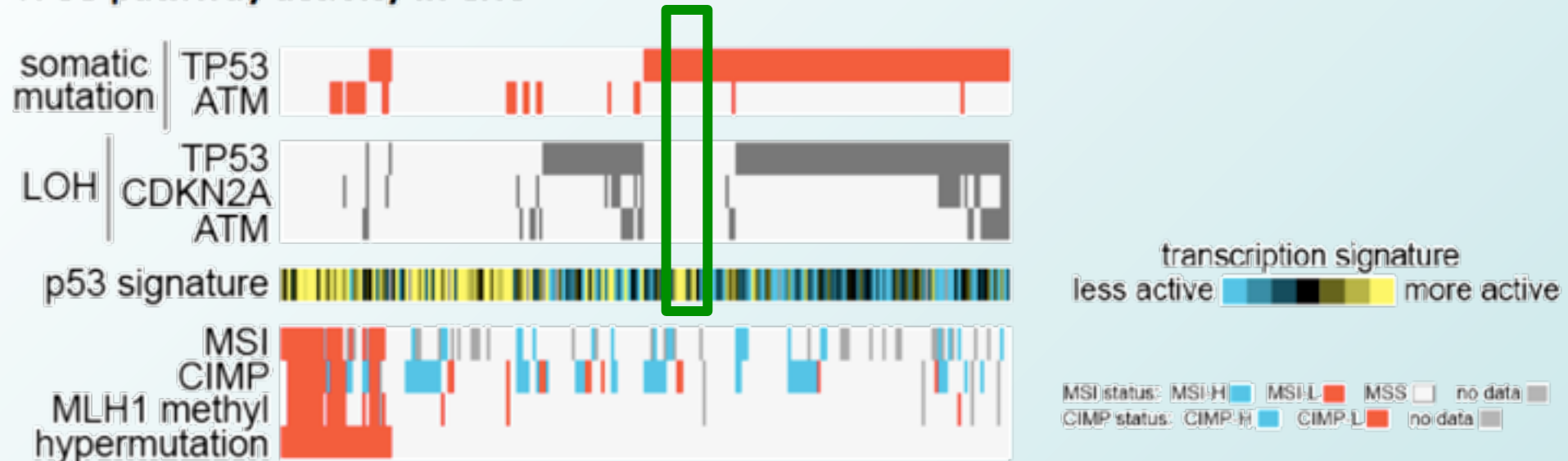
- Gene expression signatures can predict presence/absence of mutations
- Training predictors difficult if gene can have both GOF and LOF events



Gene Signatures of Mutations Shed Light on Impact

Gain of function events in TP53?

TP53 pathway activity in CRC

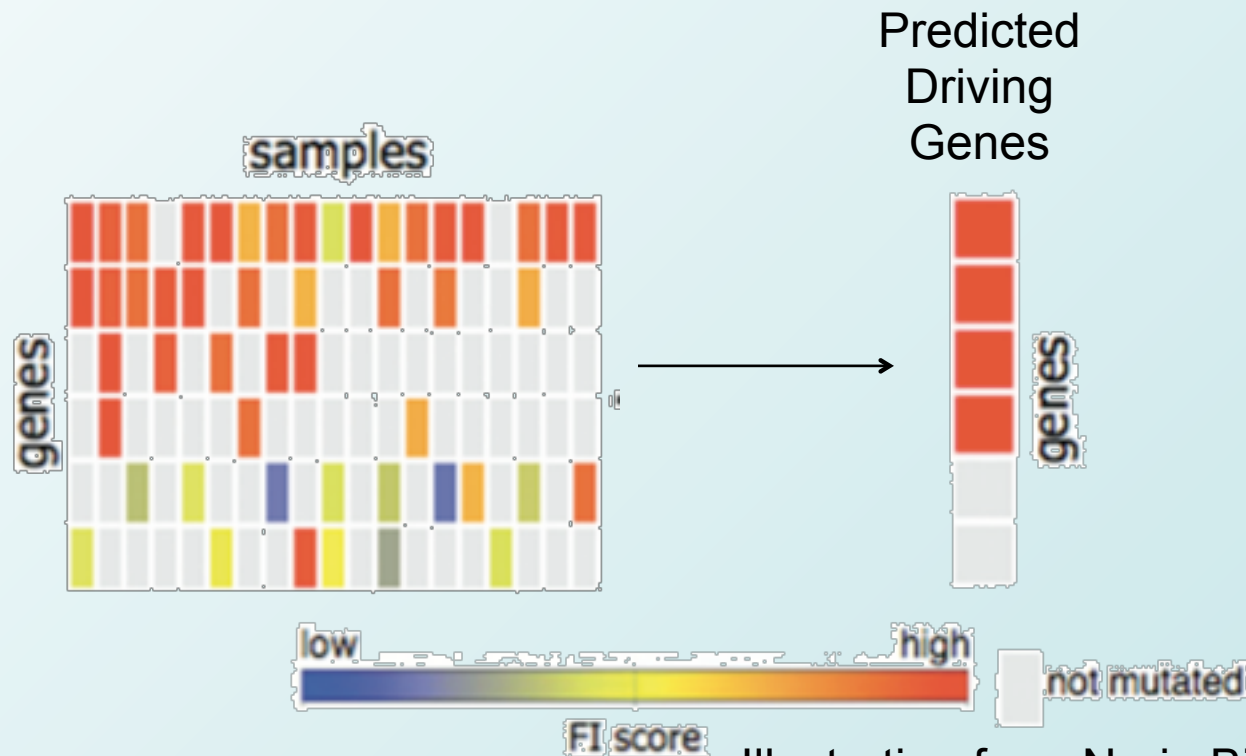


- Gene expression signatures can predict presence/absence of mutations
- Training predictors difficult if gene can have both GOF and LOF events



Are Pathway the Mutable Unit?

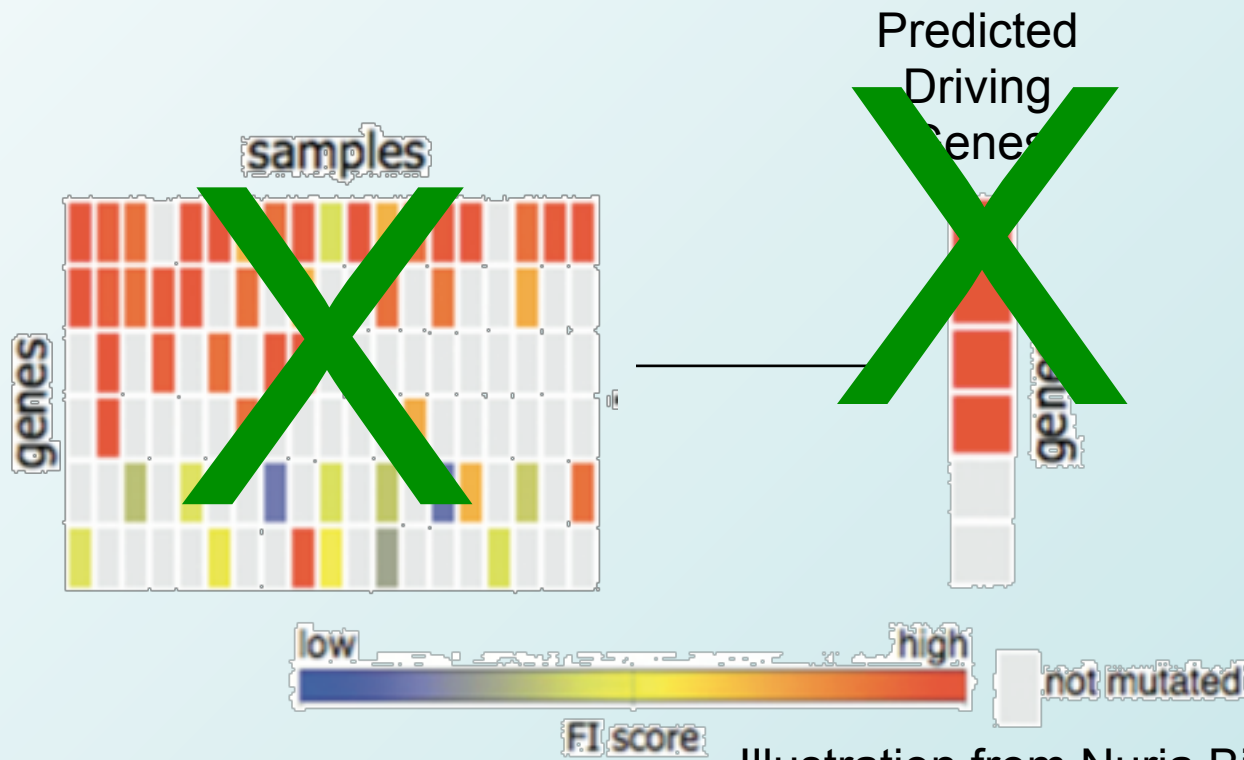
- Genes in common pathway show Mutually Exclusive patterns
 - MEMo (Ciriello et al, MSKCC, 2011)
- Mutations are clustered in protein-protein networks
 - HotNet (Vandin and Raphael, Brown, 2011)



Are Pathway the Mutable Unit?

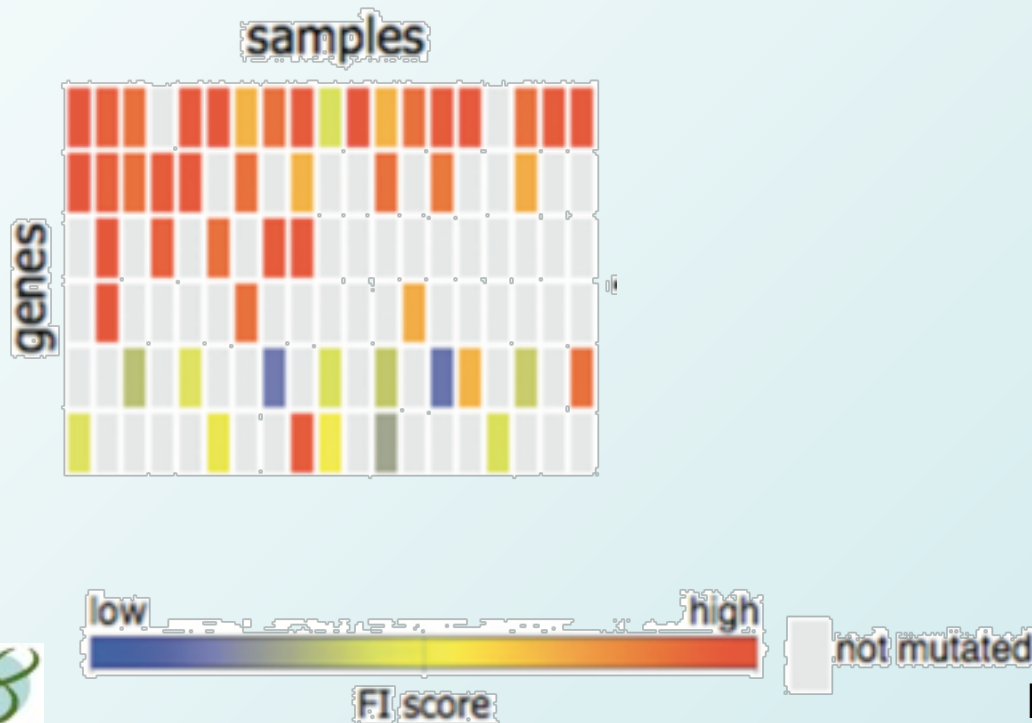
- Genes in common pathway show Mutually Exclusive patterns
 - MEMo (Ciriello et al, MSKCC, 2011)
- Mutations are clustered in protein-protein networks
 - HotNet (Vandin and Raphael, Brown, 2011)

Don't think in terms of single genes in isolation...



Pathway as the Mutable Unit

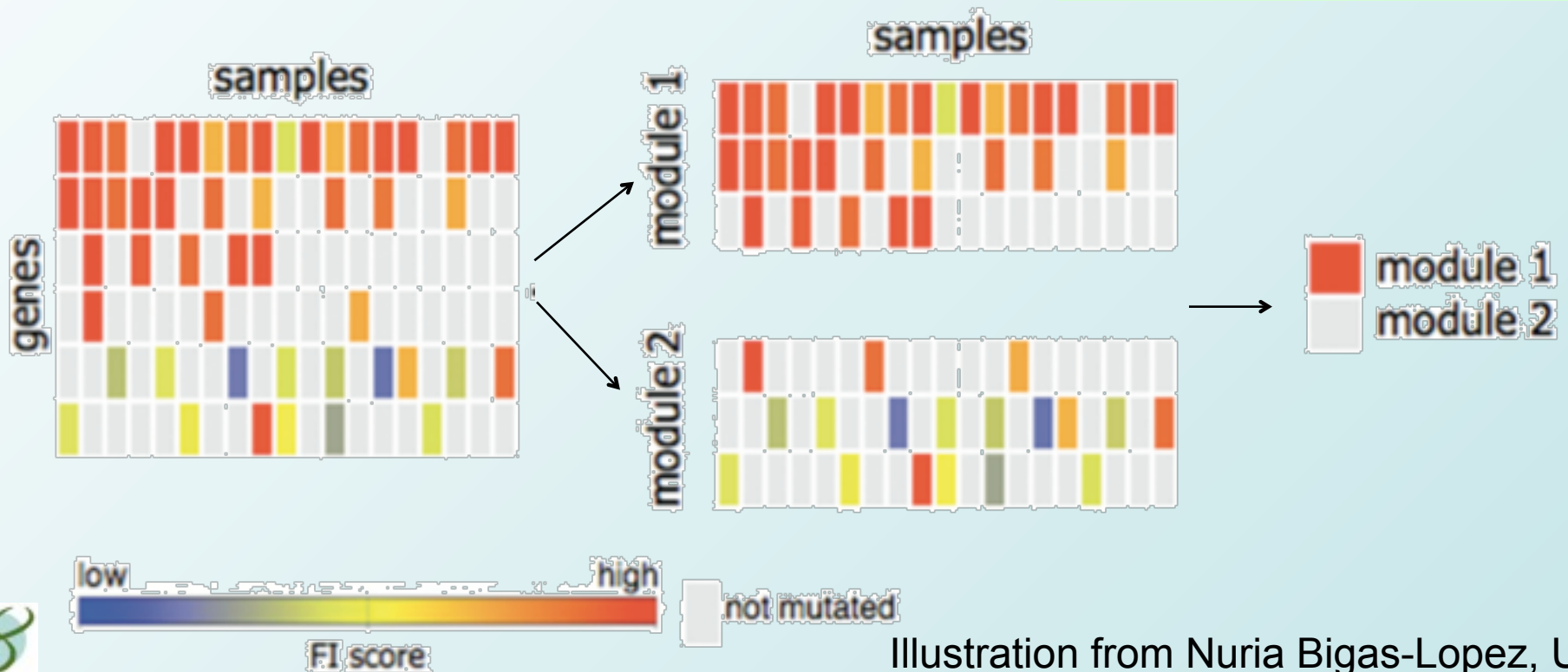
- Genes in common pathway show Mut Ex patterns
 - MEMo (Ciriello et al, MSKCC, 2011)
- Mutations are clustered in protein-protein networks
 - HotNet (Vandin and Raphael, Brown, 2011)



Pathway as the Mutable Unit

- Genes in common pathway show Mut Ex patterns
 - MEMo (Ciriello et al, MSKCC, 2011)
- Mutations are clustered in protein-protein networks
 - HotNet (Vandin and Raphael, Brown, 2011)

Pathways may reveal significant patterns of disruption



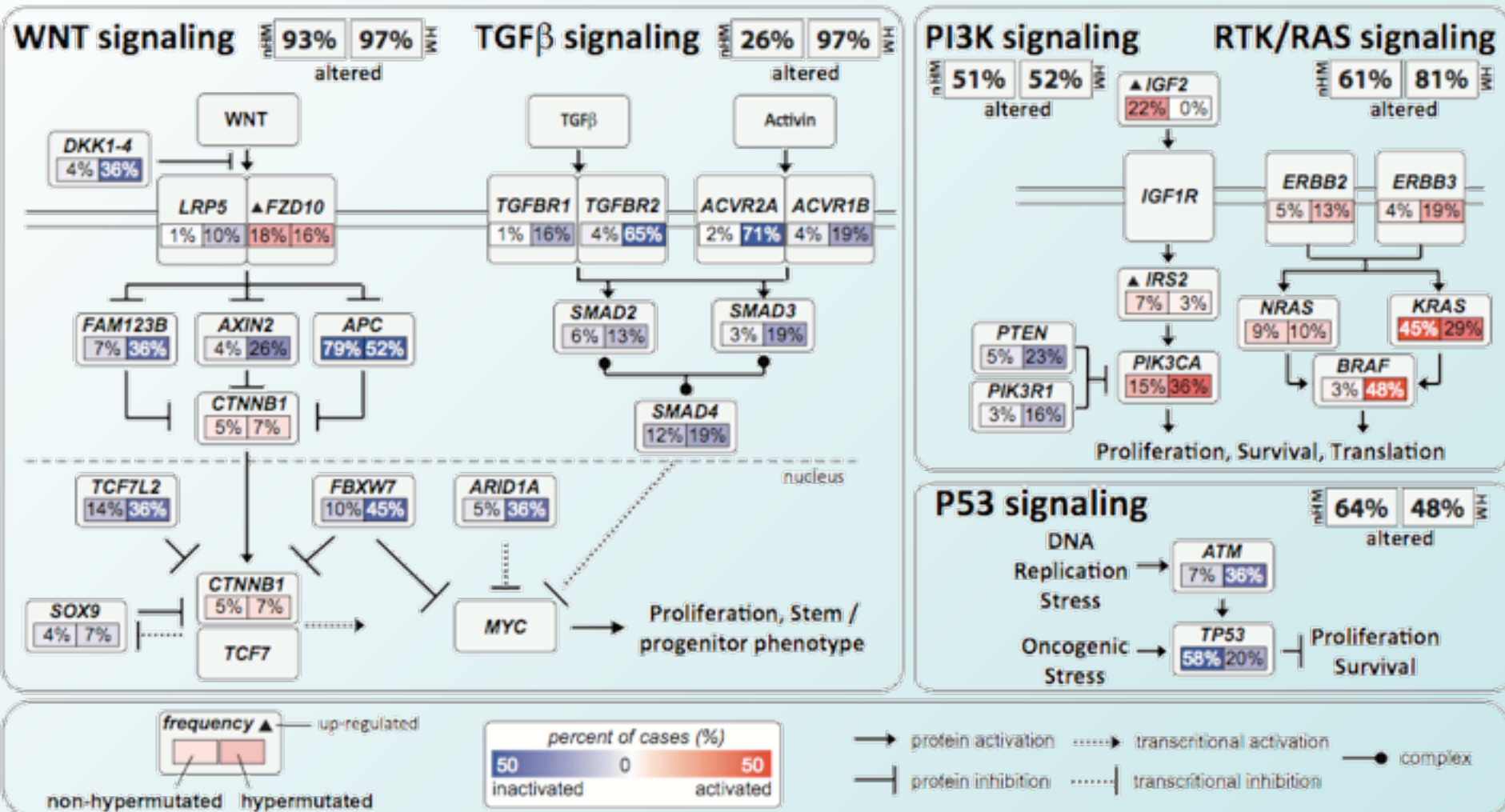
Mutual Exclusion among genes in key pathways

TCGA Colorectal MEMo result:



Mutual Exclusion among genes in key pathways

TCGA Colorectal Set:

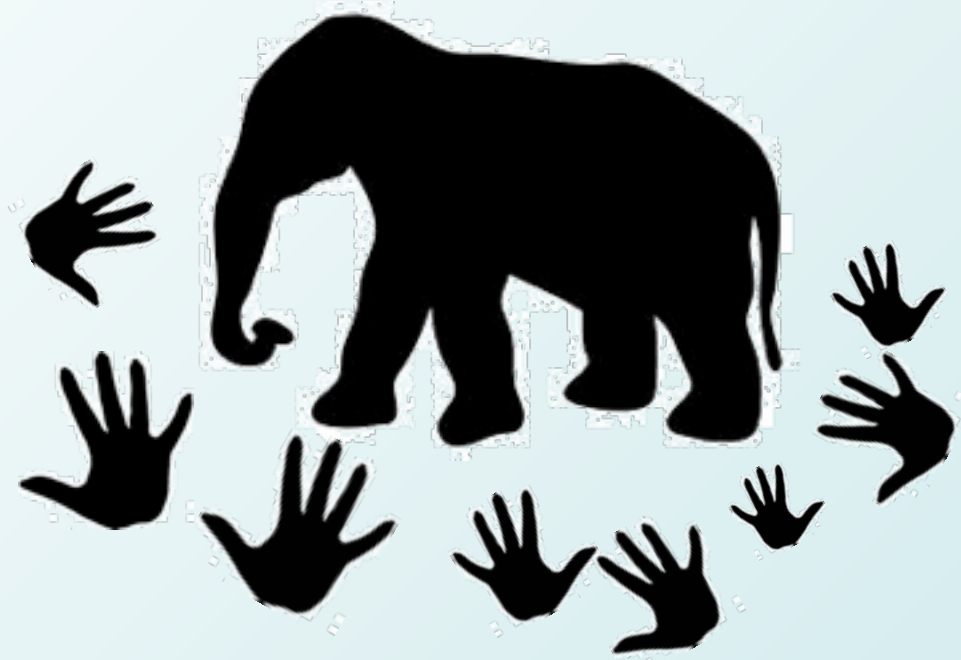


Overview of pathway-guided approach

- Integrate many data sources to gain accurate view of how genes are functioning in pathways
- Predict the functional consequences of mutations by quantifying the effect on the surrounding pathway



TCGA: Multiple views of a beast



Flood of Data Analysis Challenges

Genomics, Functional Genomics, Metabolomics, Epigenomics =



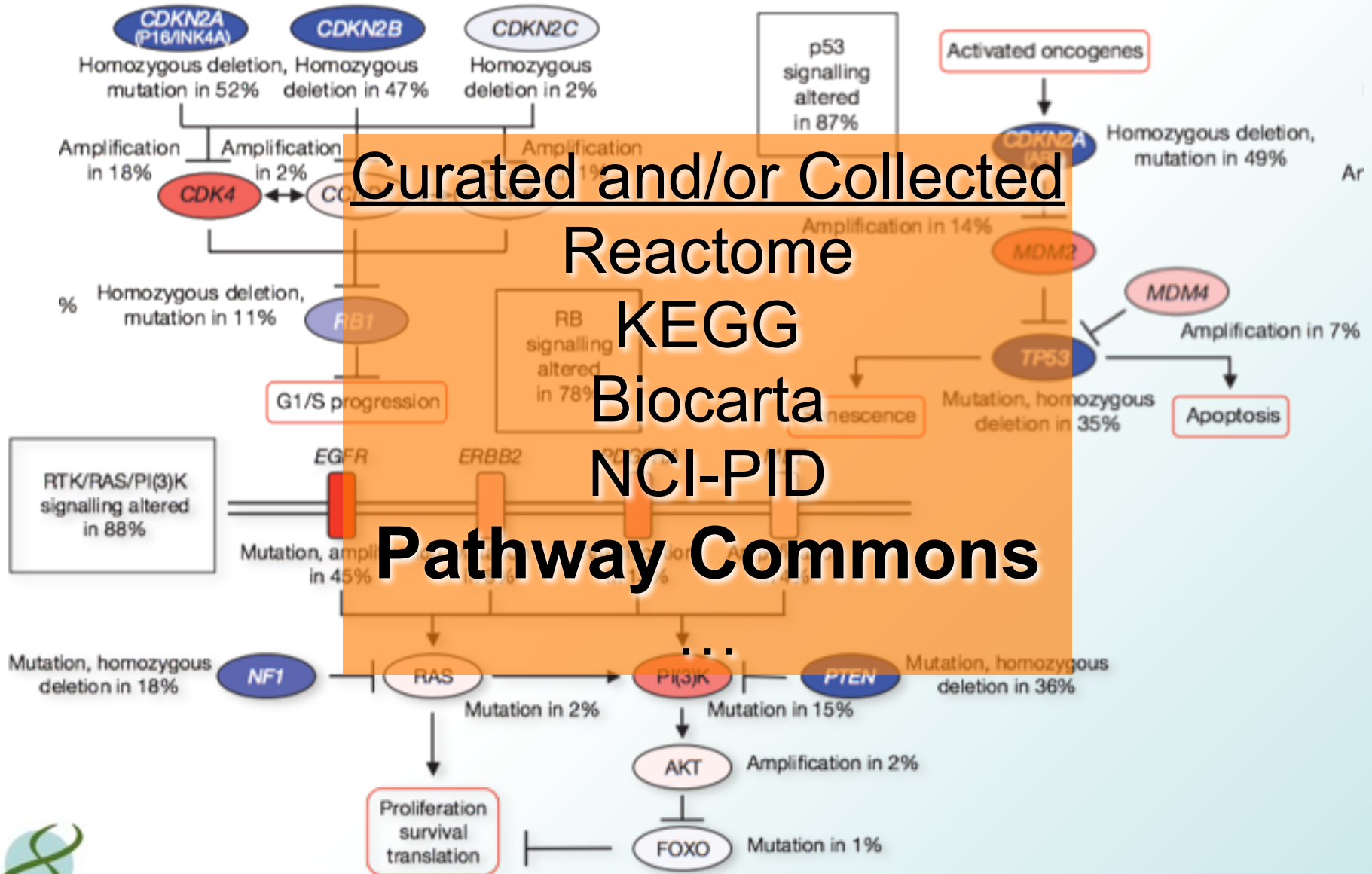
Multiple, Possibly
Conflicting Signals

Expression

Copy Number
Alterations

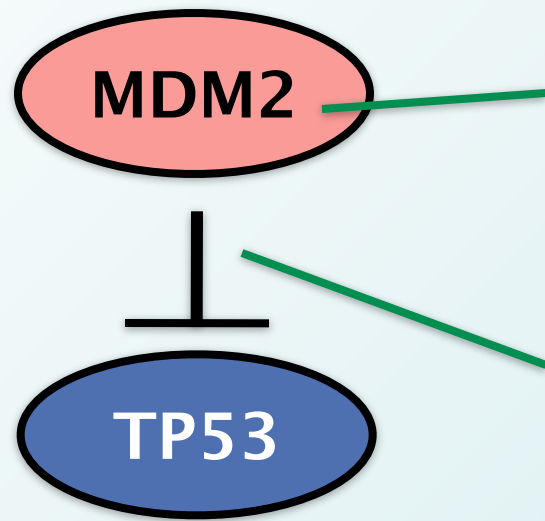
DNA Methylation

Much Cell Machinery Known: Gene circuitry now available.



Integration Approach: Detailed models of expression and interaction

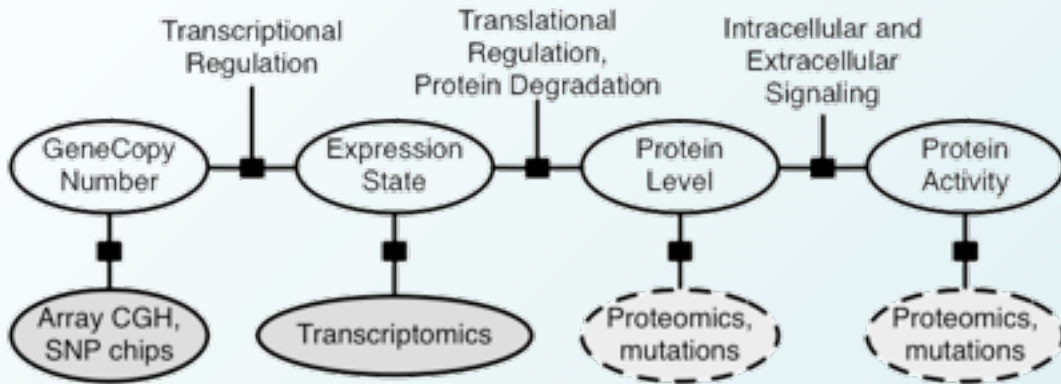
Two Parts:



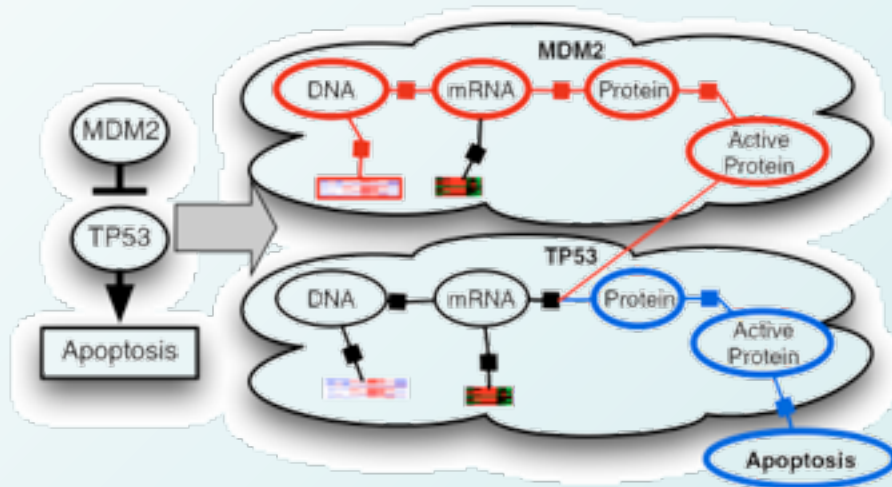
1. Gene Level Model (central dogma)
2. Interaction Model (regulation)



PARDIGM Gene Model to Integrate Data



1. Central Dogma-Like Gene Model of Activity



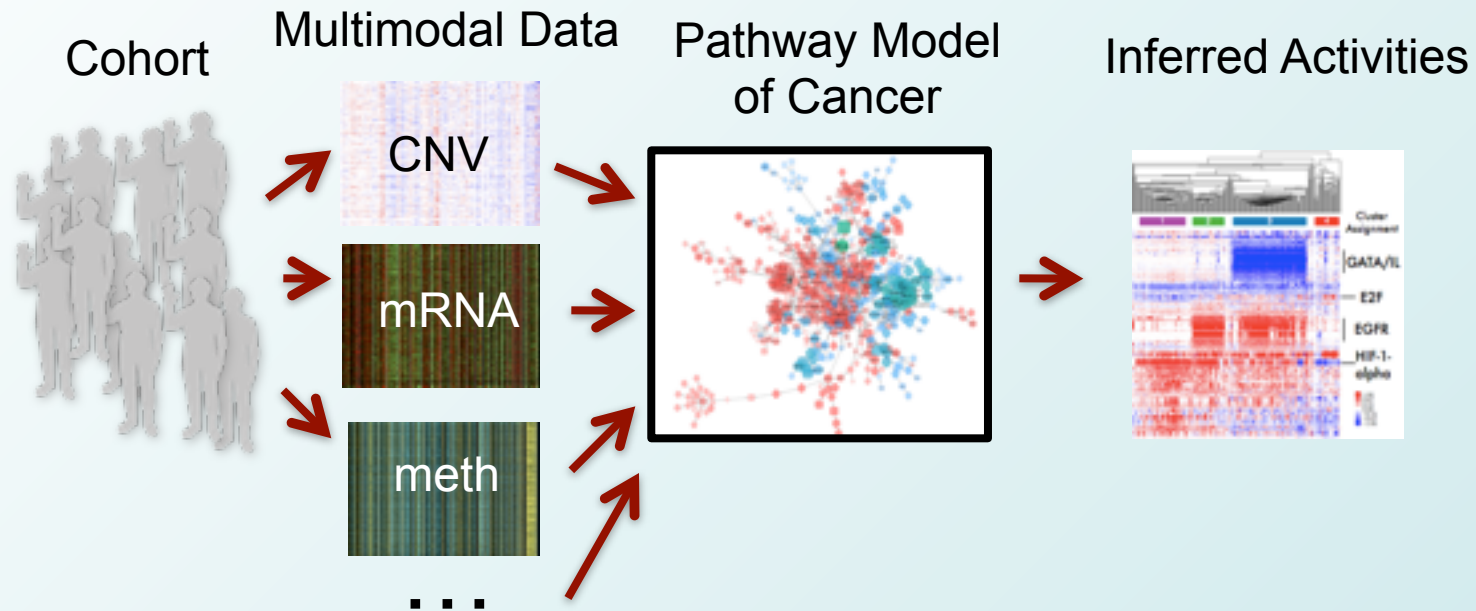
2. Interactions that connect to specific points in gene regulation map



Charlie Vaske
Steve Benz



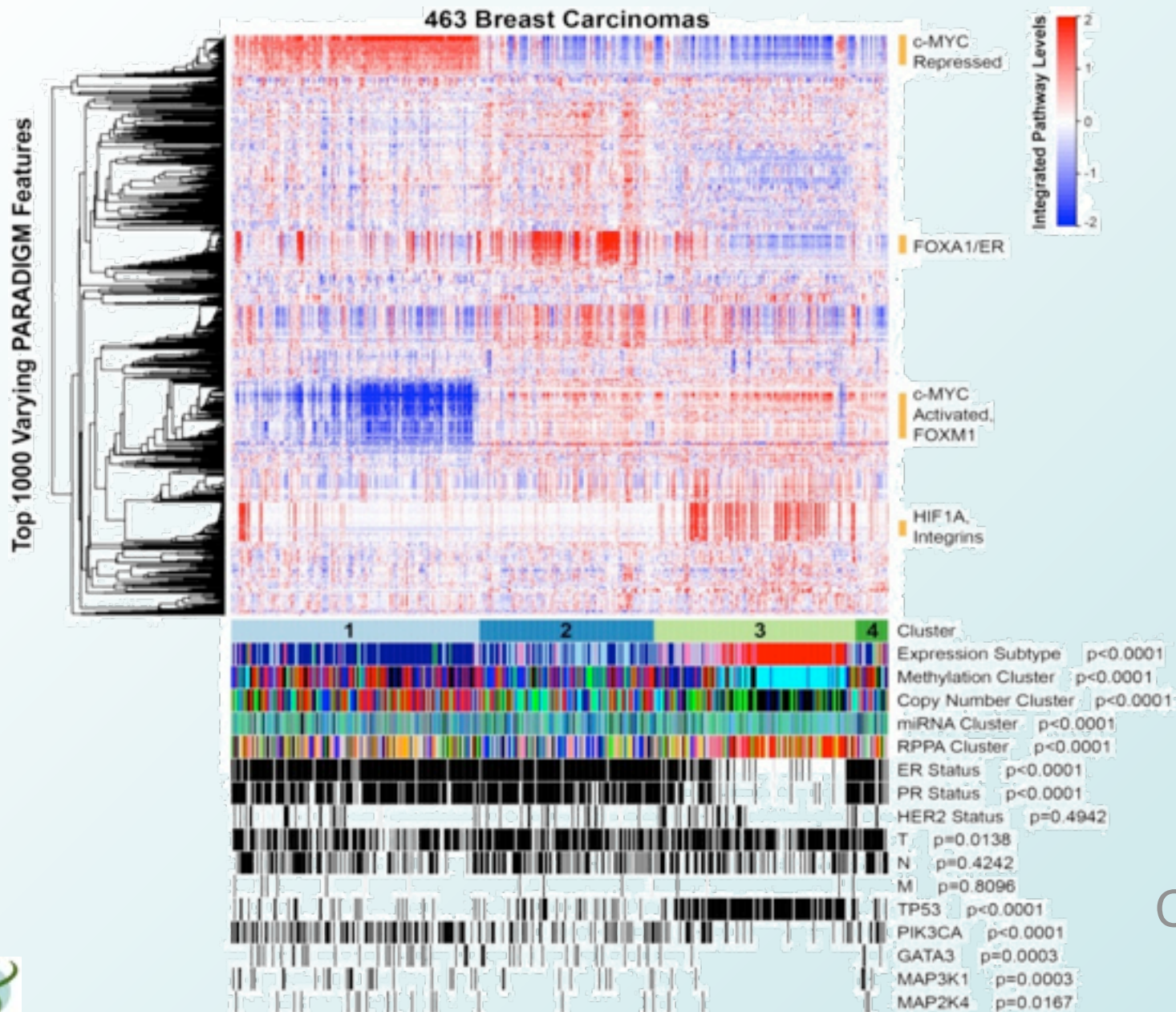
Integrated Pathway Analysis for Cancer



- Integrated dataset for downstream analysis
- Inferred activities reflect neighborhood of influence around a gene.
- Can boost signal for survival analysis and mutation impact



PARADIGM in TCGA patient BRCA tumors



Christina Yau,
Buck Inst



How do you distinguish drivers from passengers?

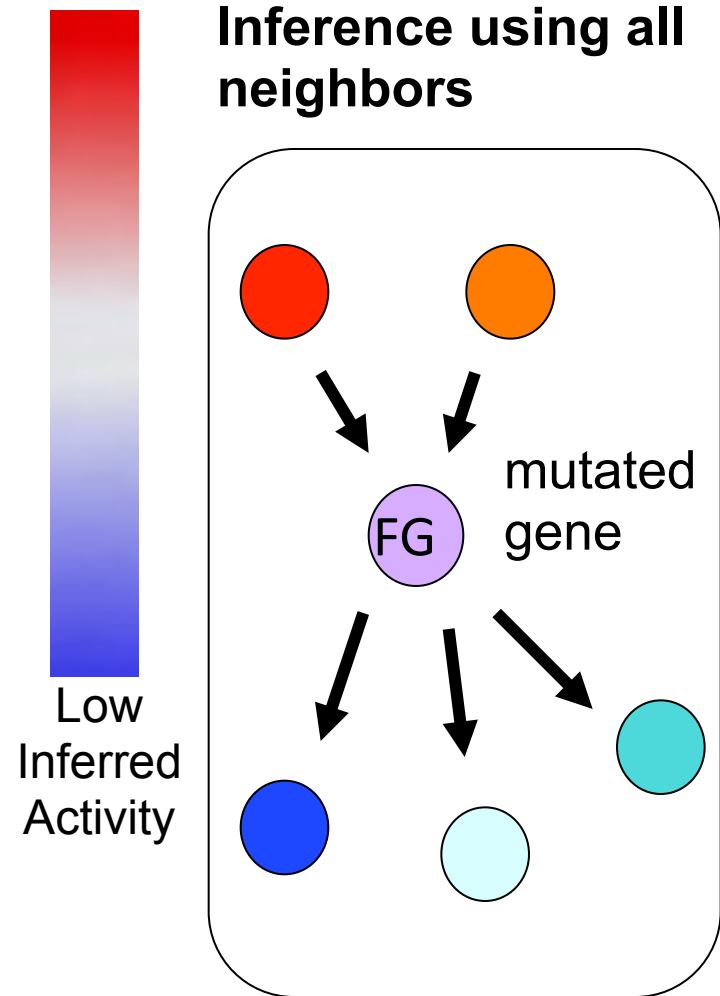
- GOF: increase subsystem function: as in *Pedal stuck down*.
- LOF: decrease subsystem function: as in *Needle broken*.



PARADIGM-Shift Predicting the Impact of

gh Inferred
Activity

**Inference using all
neighbors**

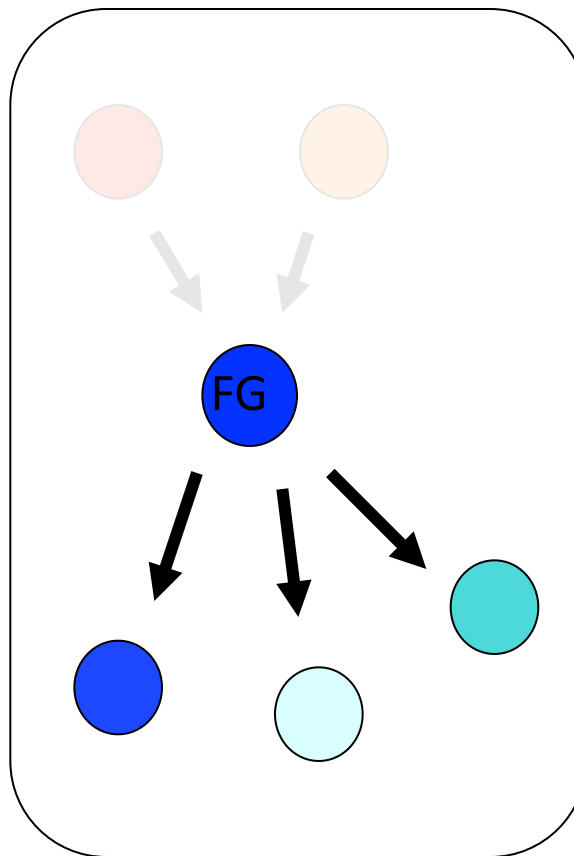
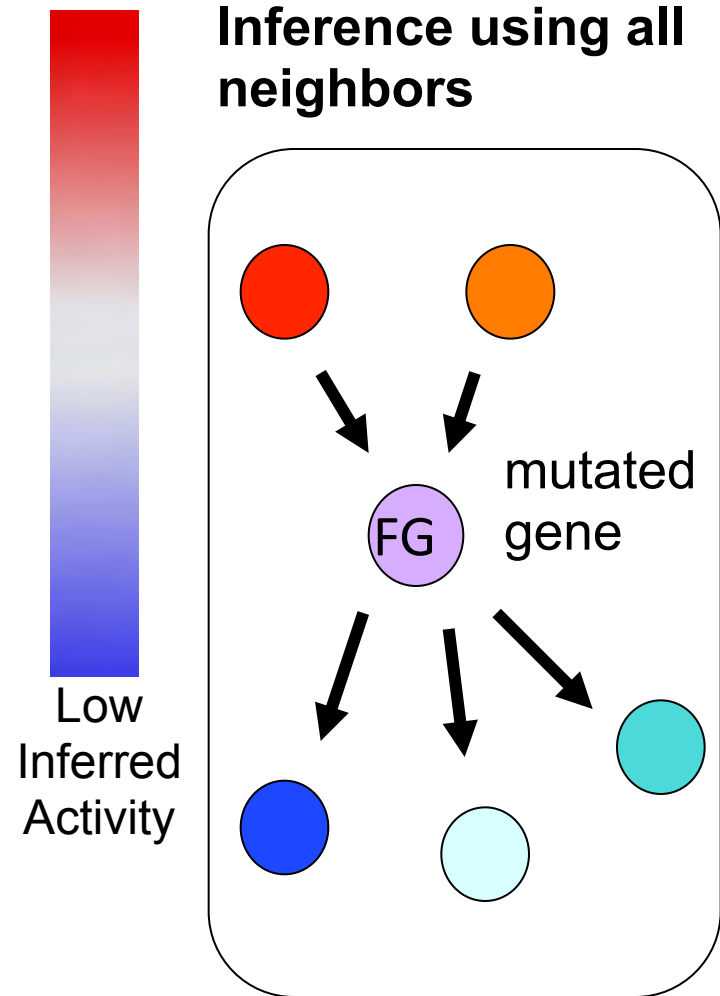


PARADIGM-Shift Predicting the Impact of

gh Inferred Activity

Inference using all neighbors

Inference using downstream neighbors

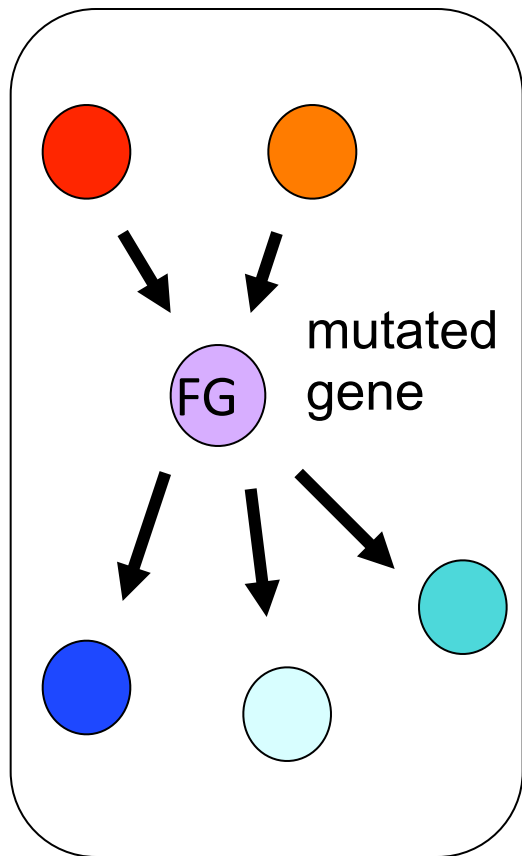


PARADIGM-Shift Predicting the Impact of

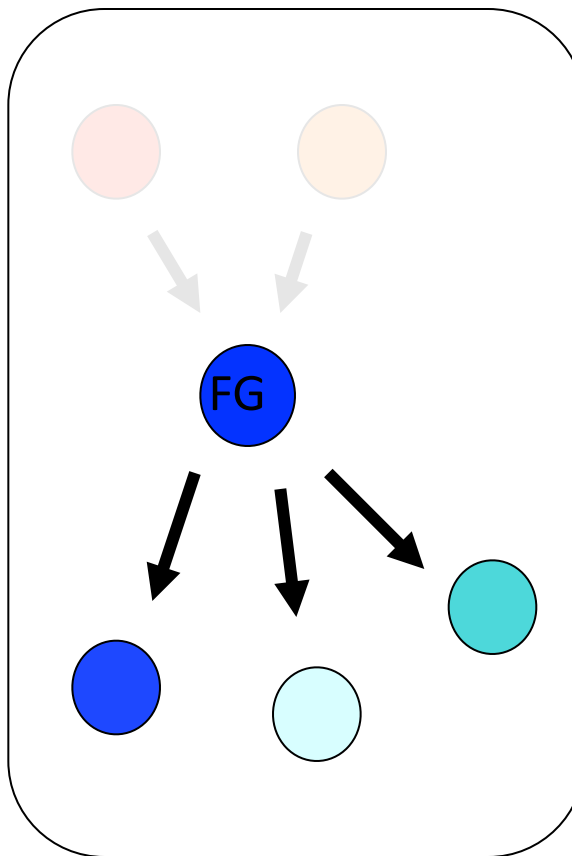
gh Inferred Activity



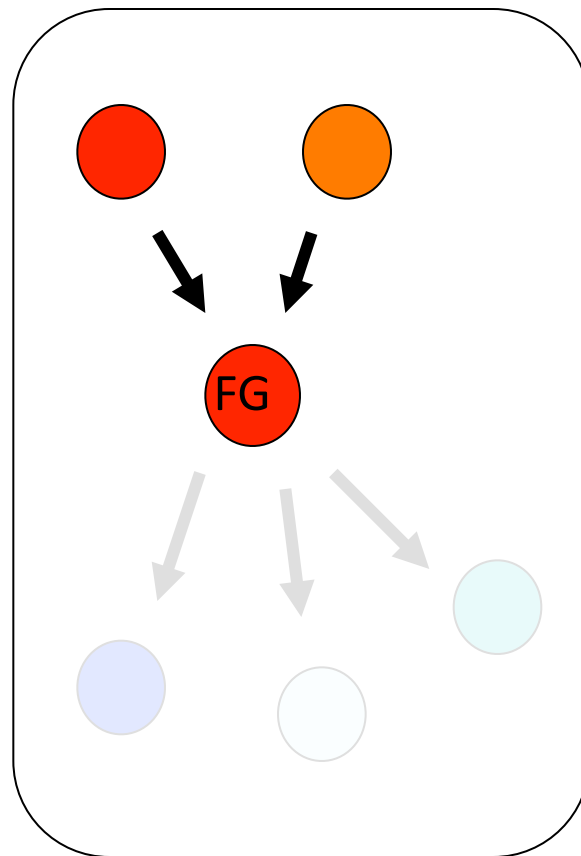
Inference using all neighbors



Inference using downstream neighbors

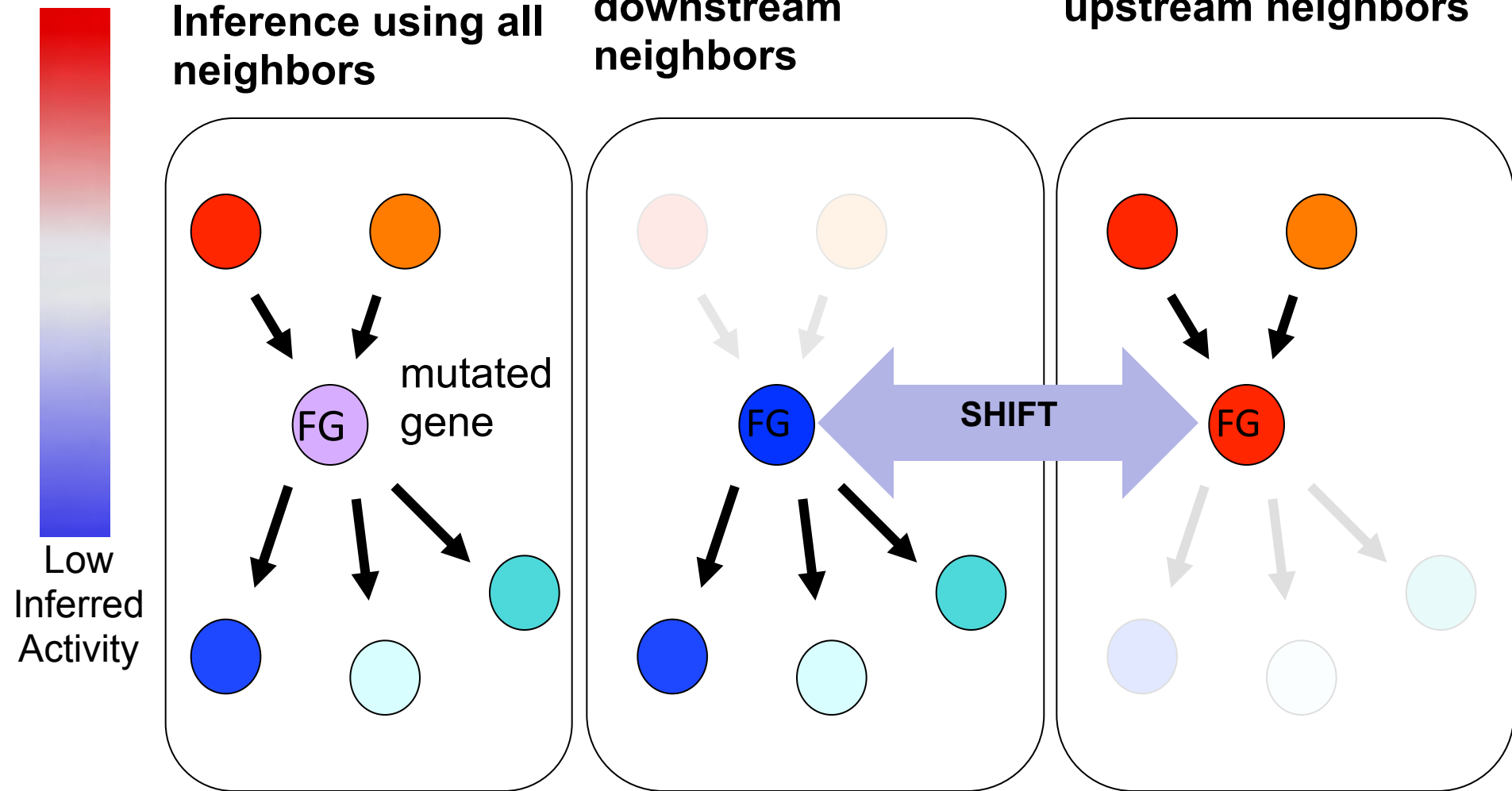


Inference using upstream neighbors

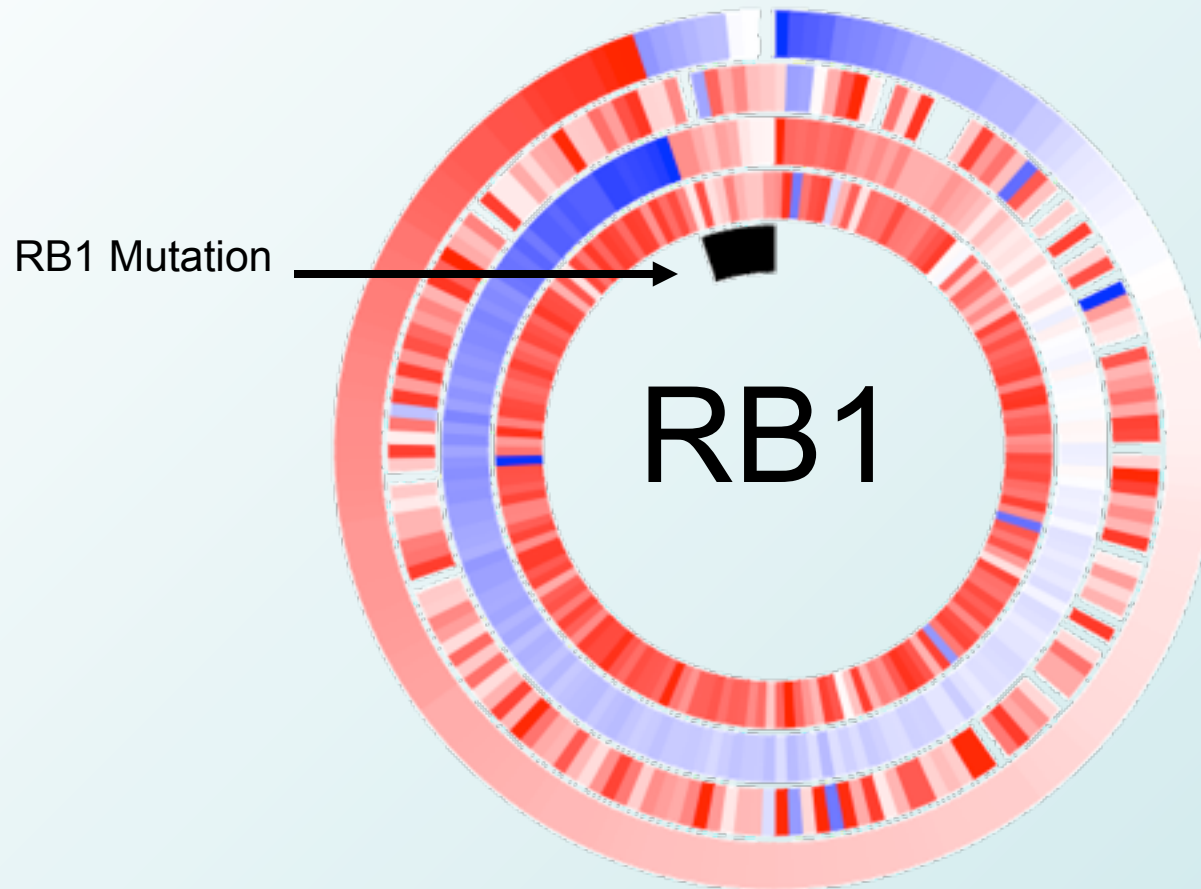


PARADIGM-Shift Predicting the Impact of

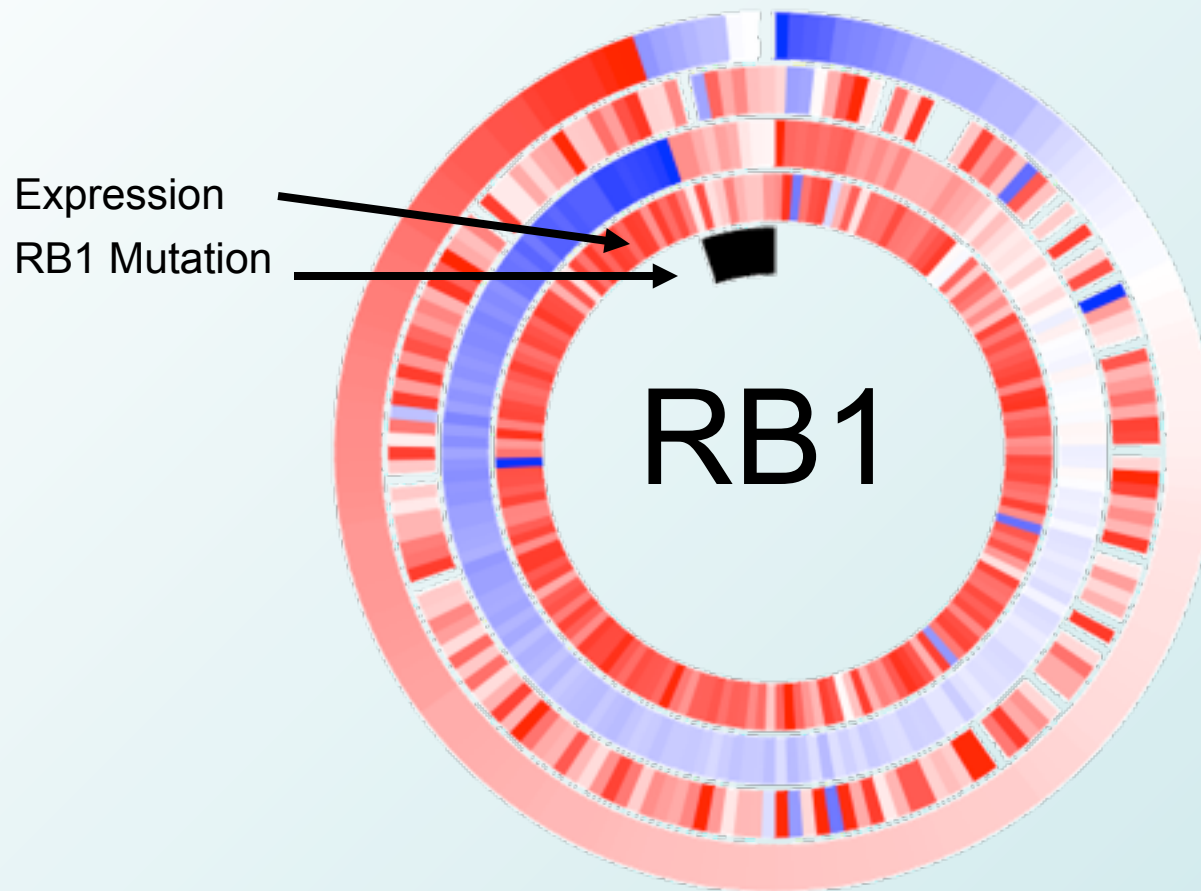
gh Inferred Activity



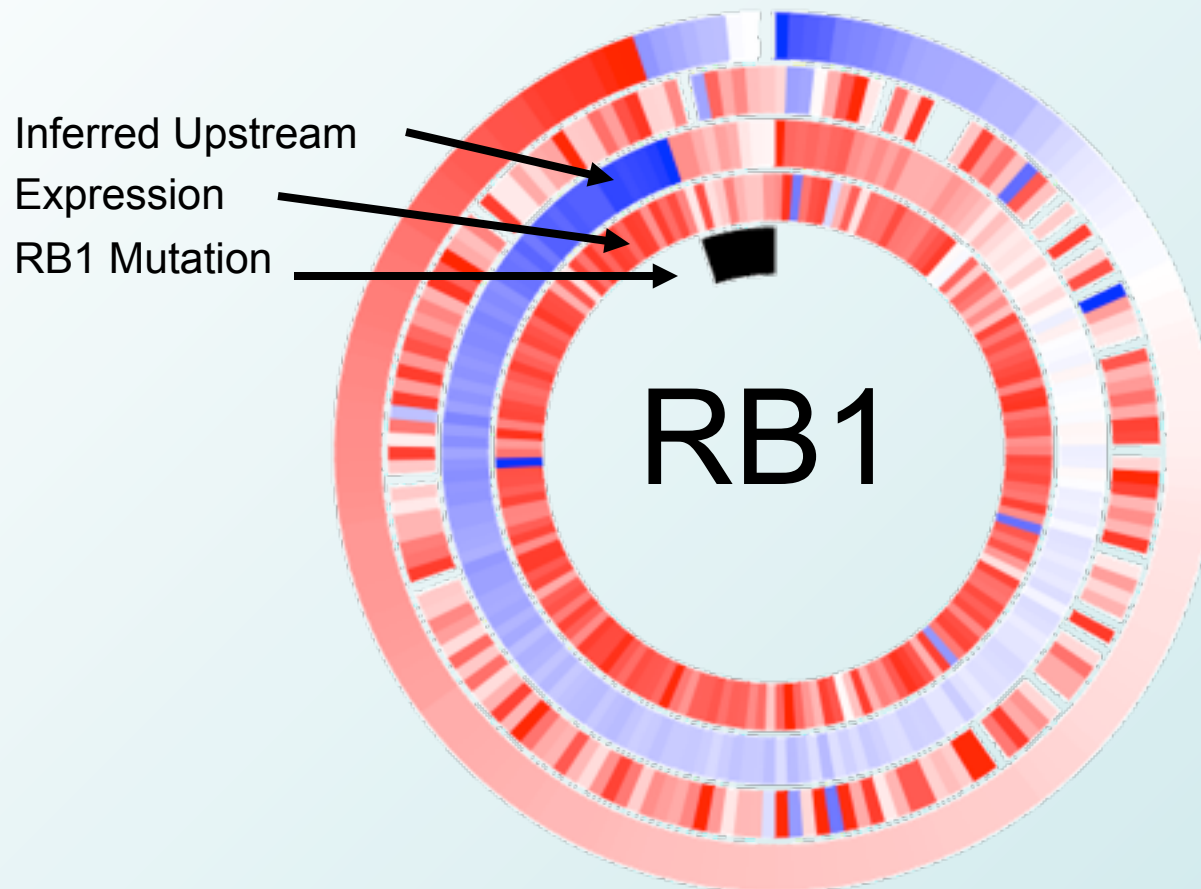
RB1 Loss-of-Function (GBM)



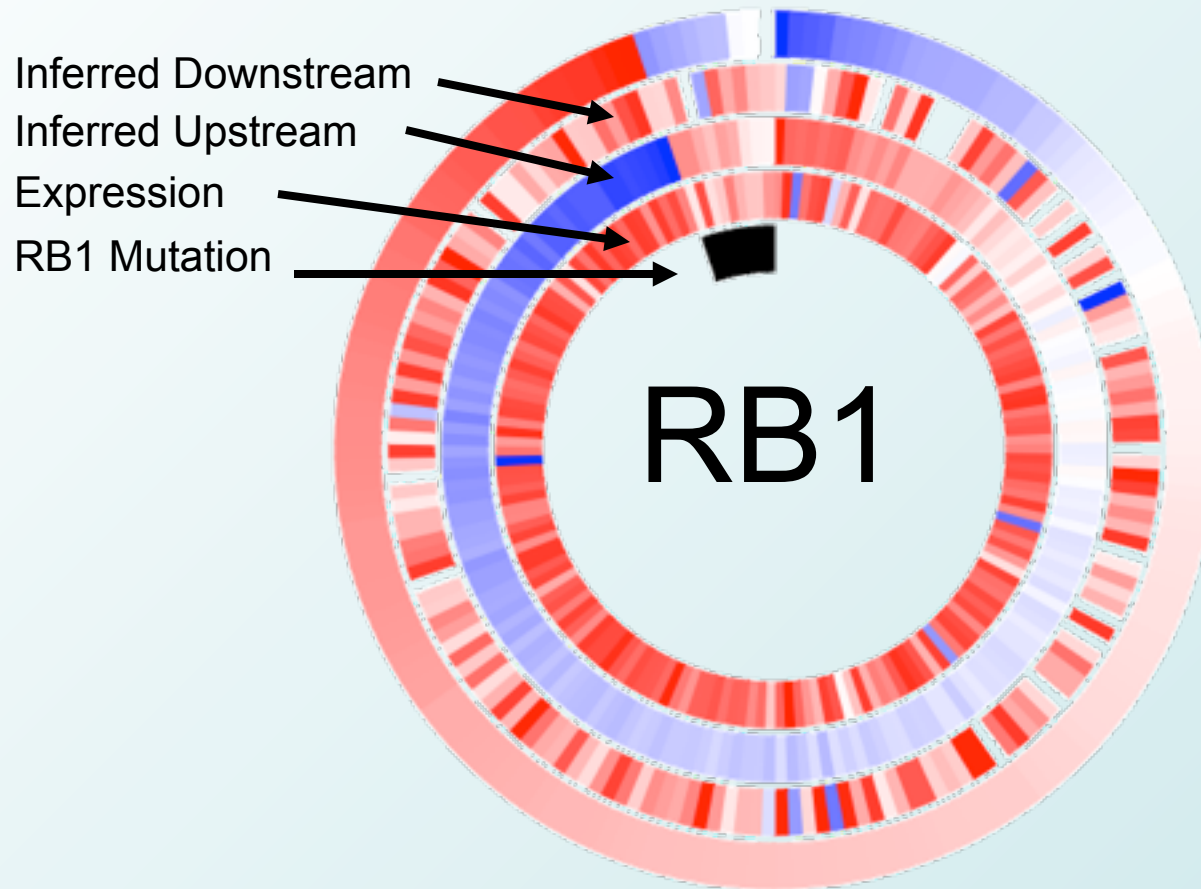
RB1 Loss-of-Function (GBM)



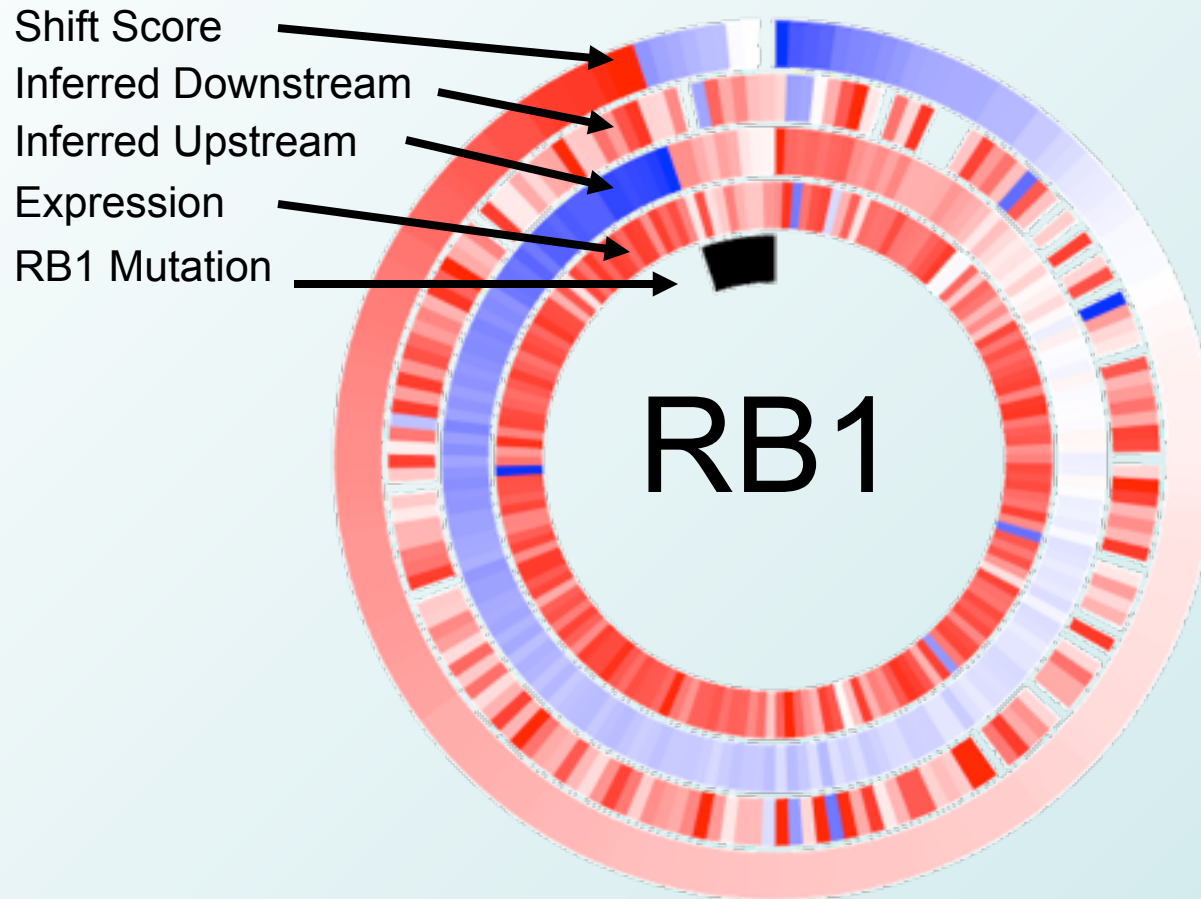
RB1 Loss-of-Function (GBM)



RB1 Loss-of-Function (GBM)

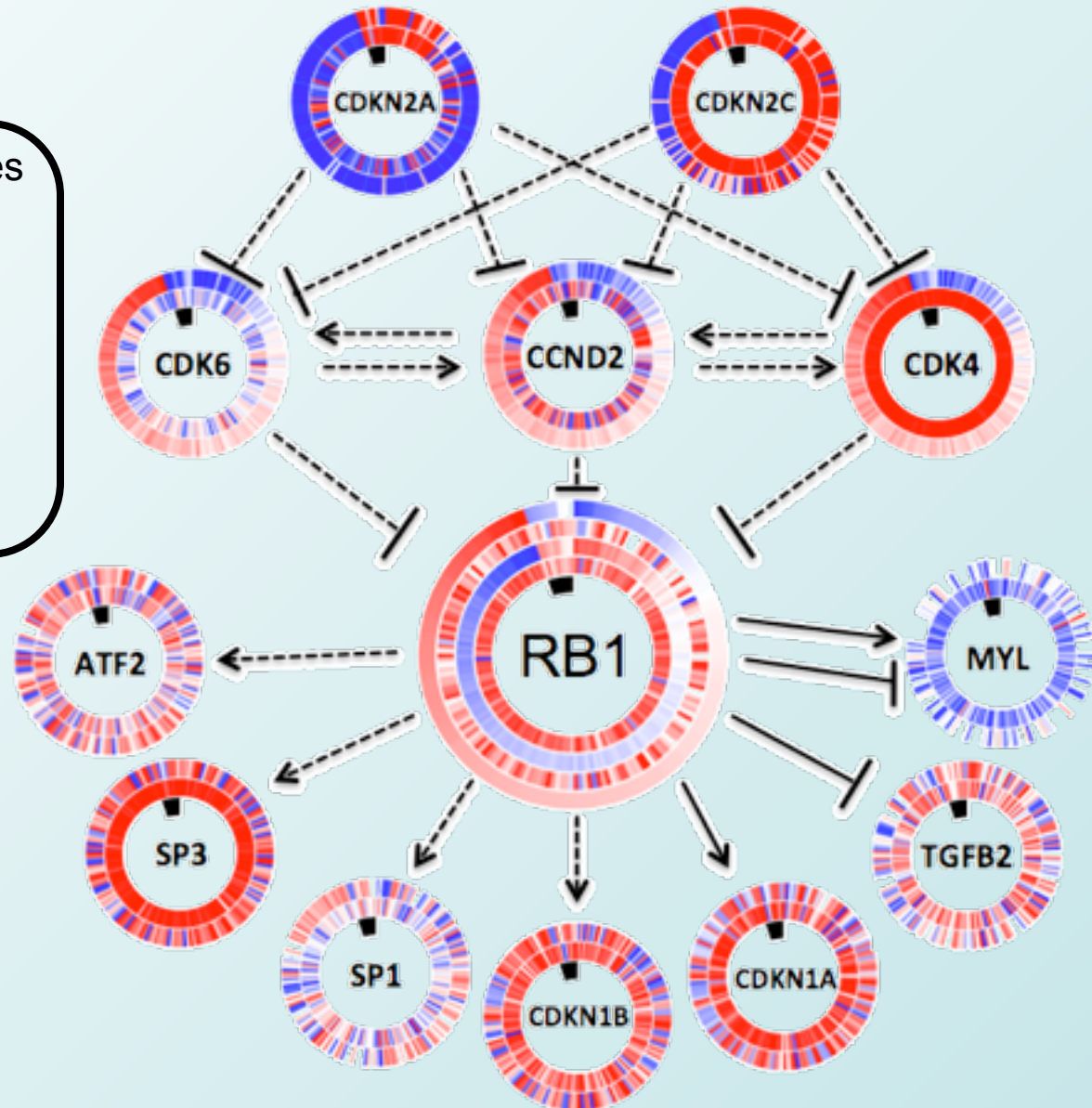
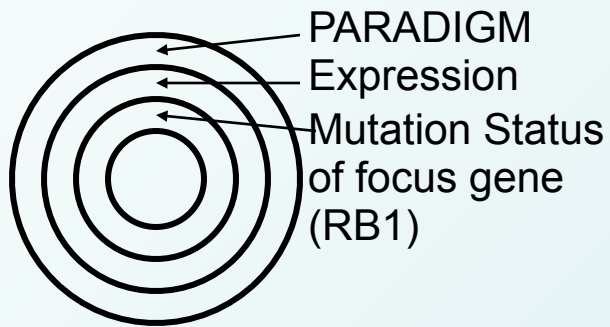


RB1 Loss-of-Function (GBM)



RB1 Loss-of-Function (GBM)

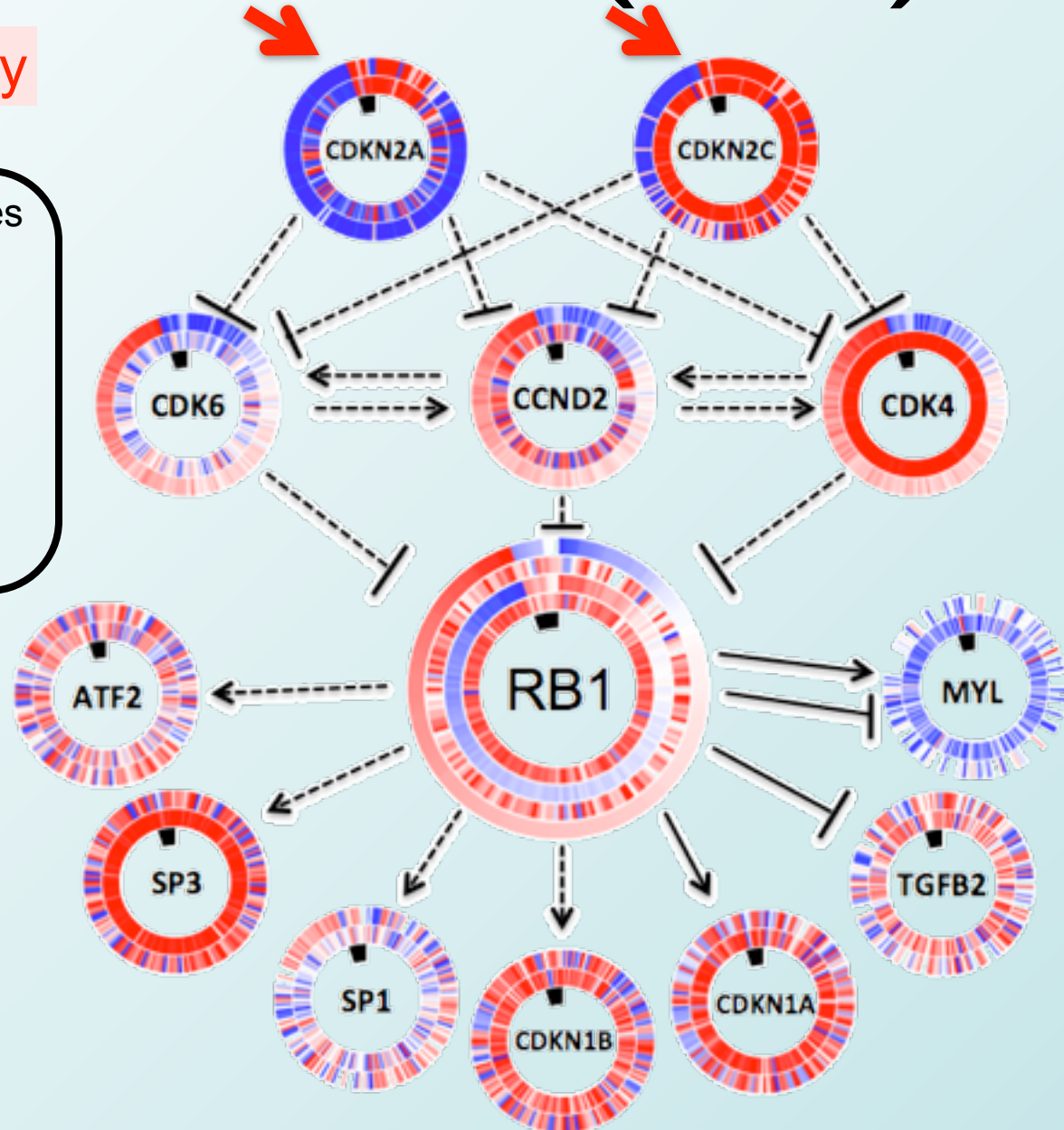
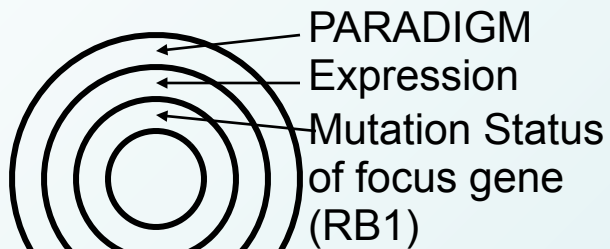
Upstream and Downstream Genes



RB1 Loss-of-Function (GBM)

High Activator Activity

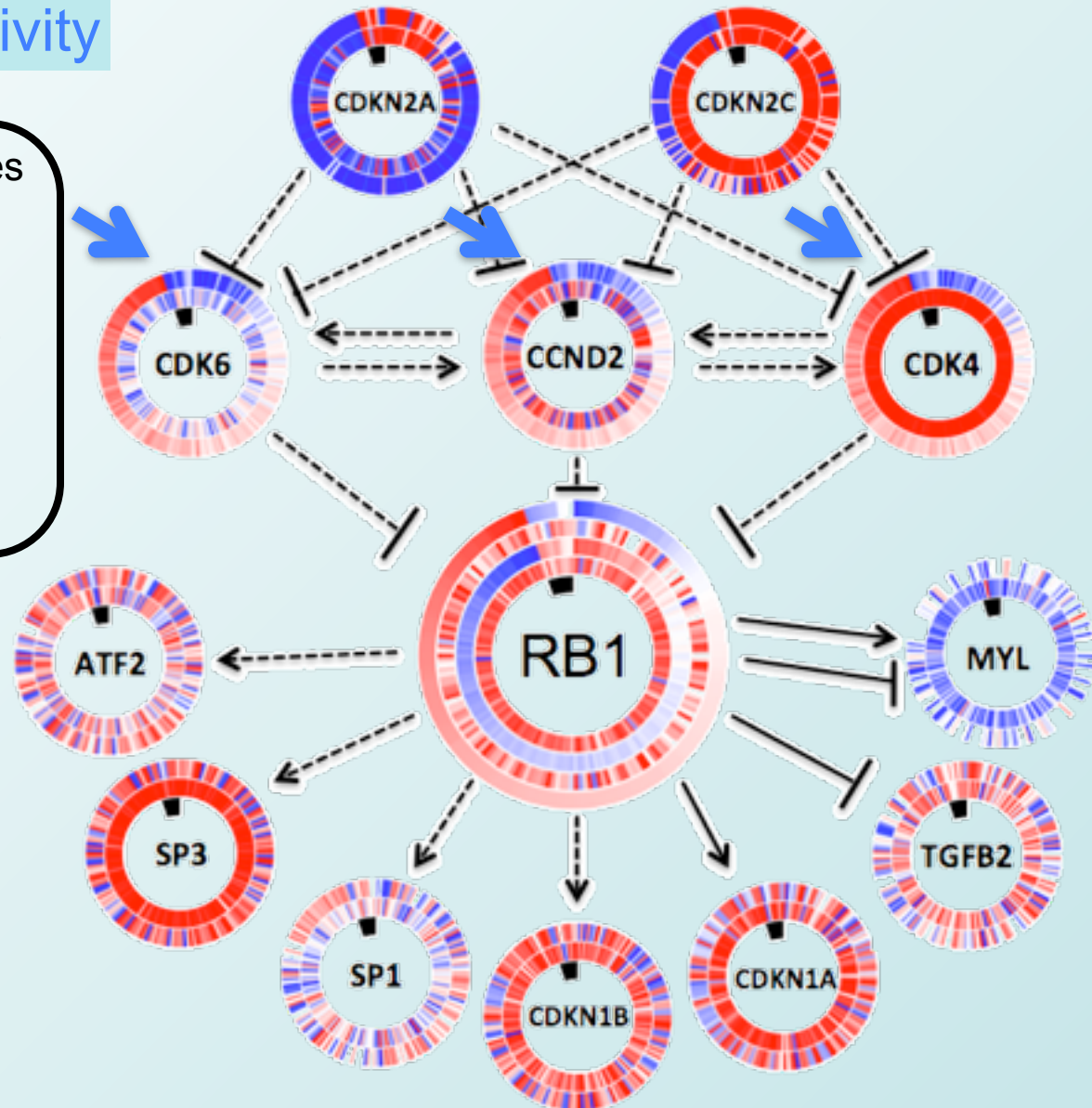
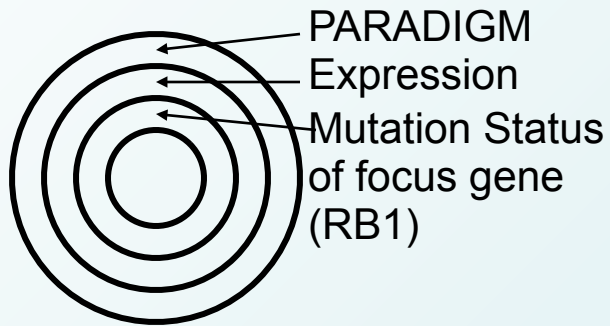
Upstream and Downstream Genes



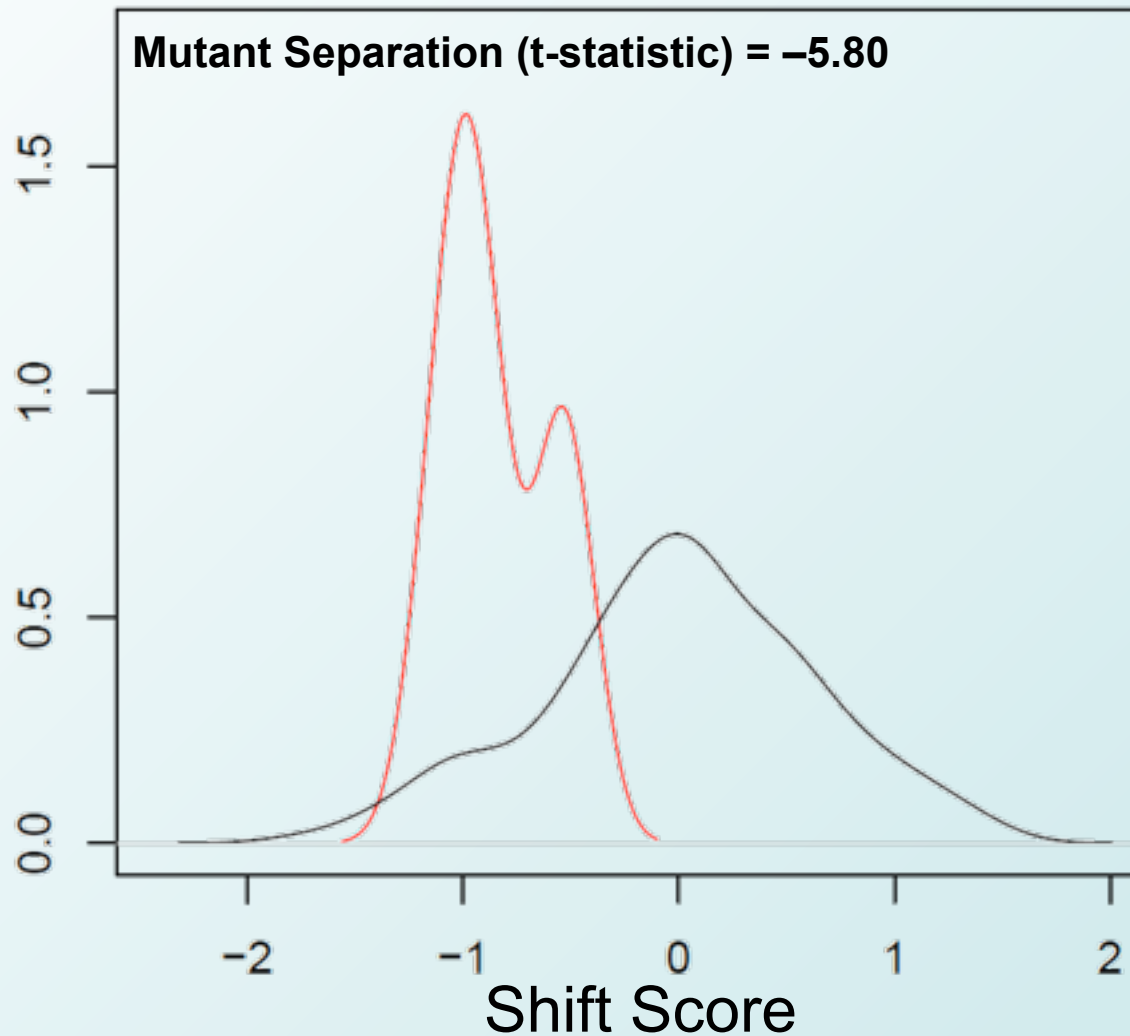
RB1 Loss-of-Function (GBM)

Low Inhibitor Activity

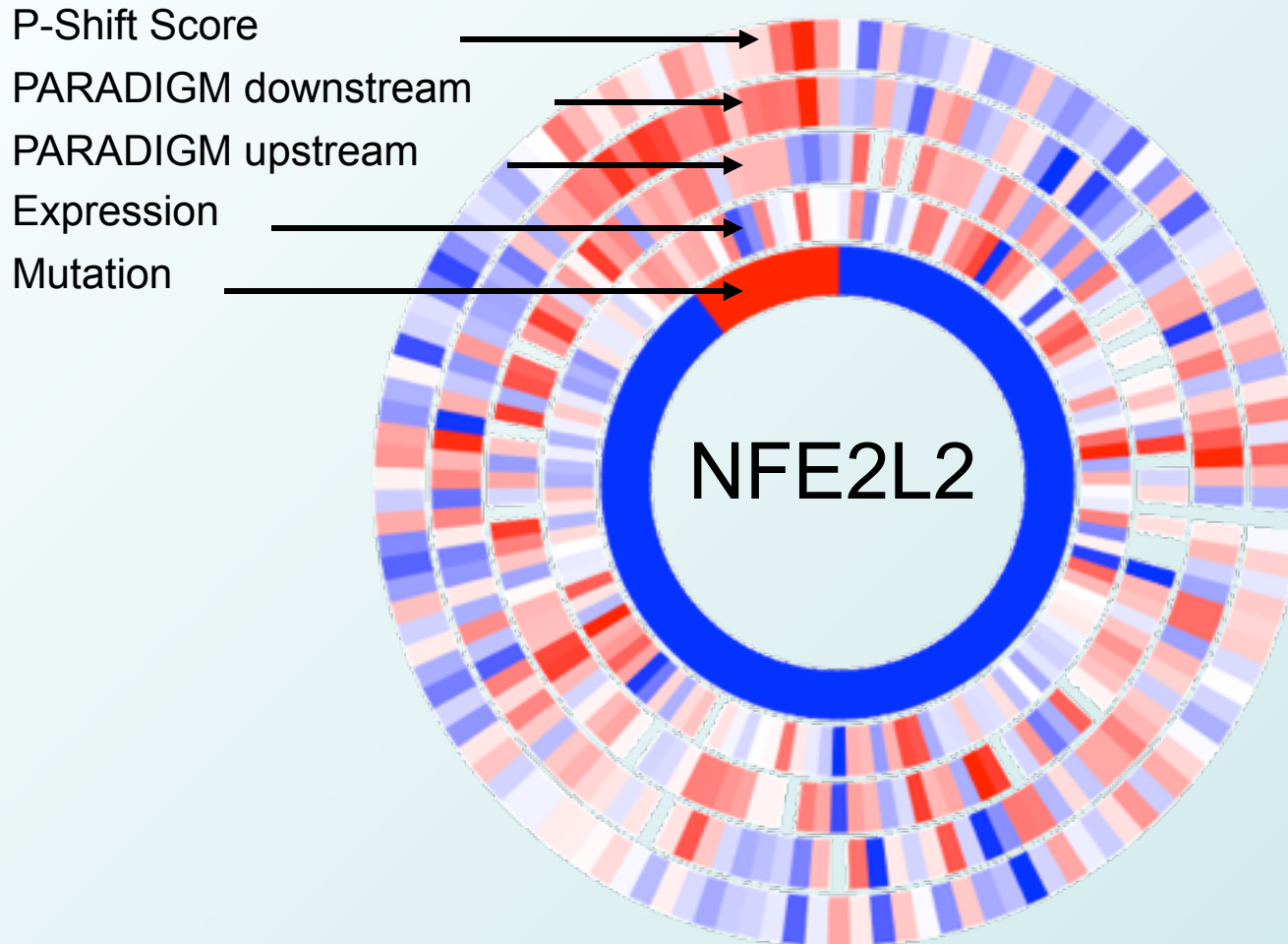
Upstream and Downstream Genes



Shift Scores differ in mutated versus non-mutated samples

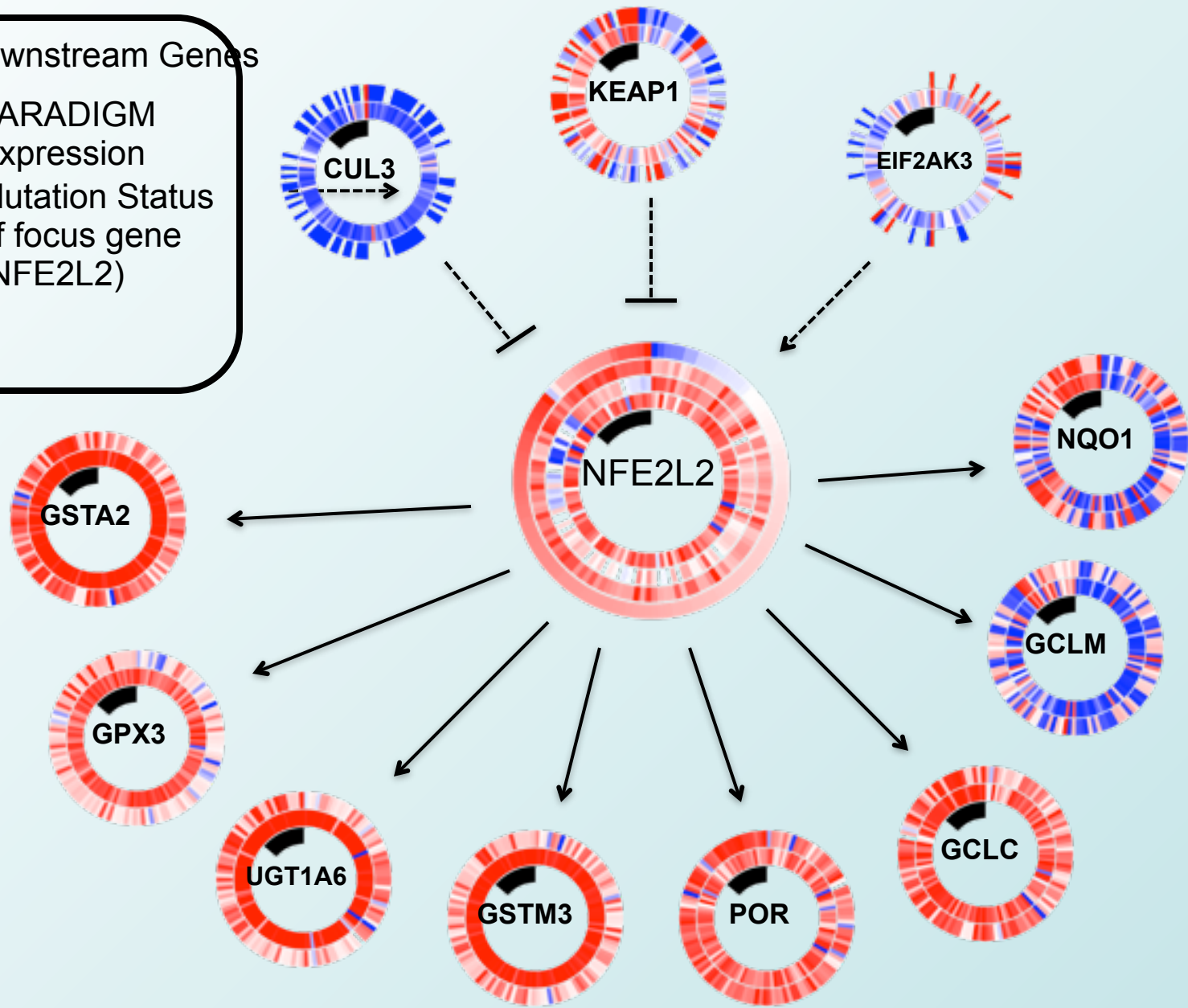
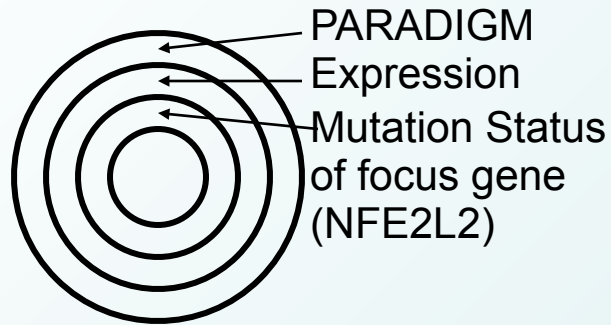


Gain-of-Function (LUSC)



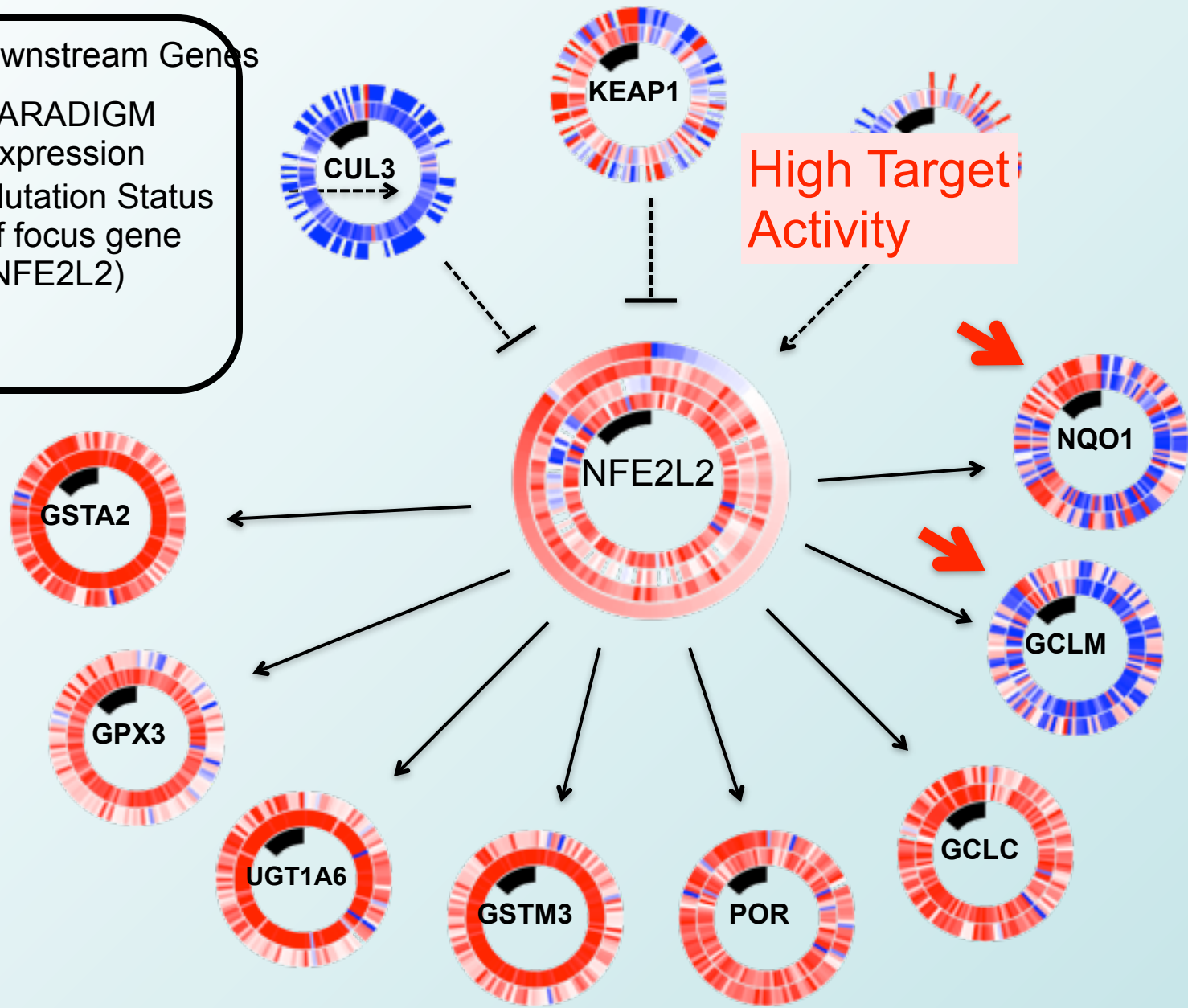
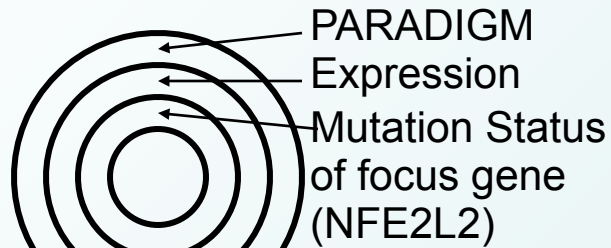
NFE2L2 Gain-of-Function (LUSC)

Upstream and Downstream Genes

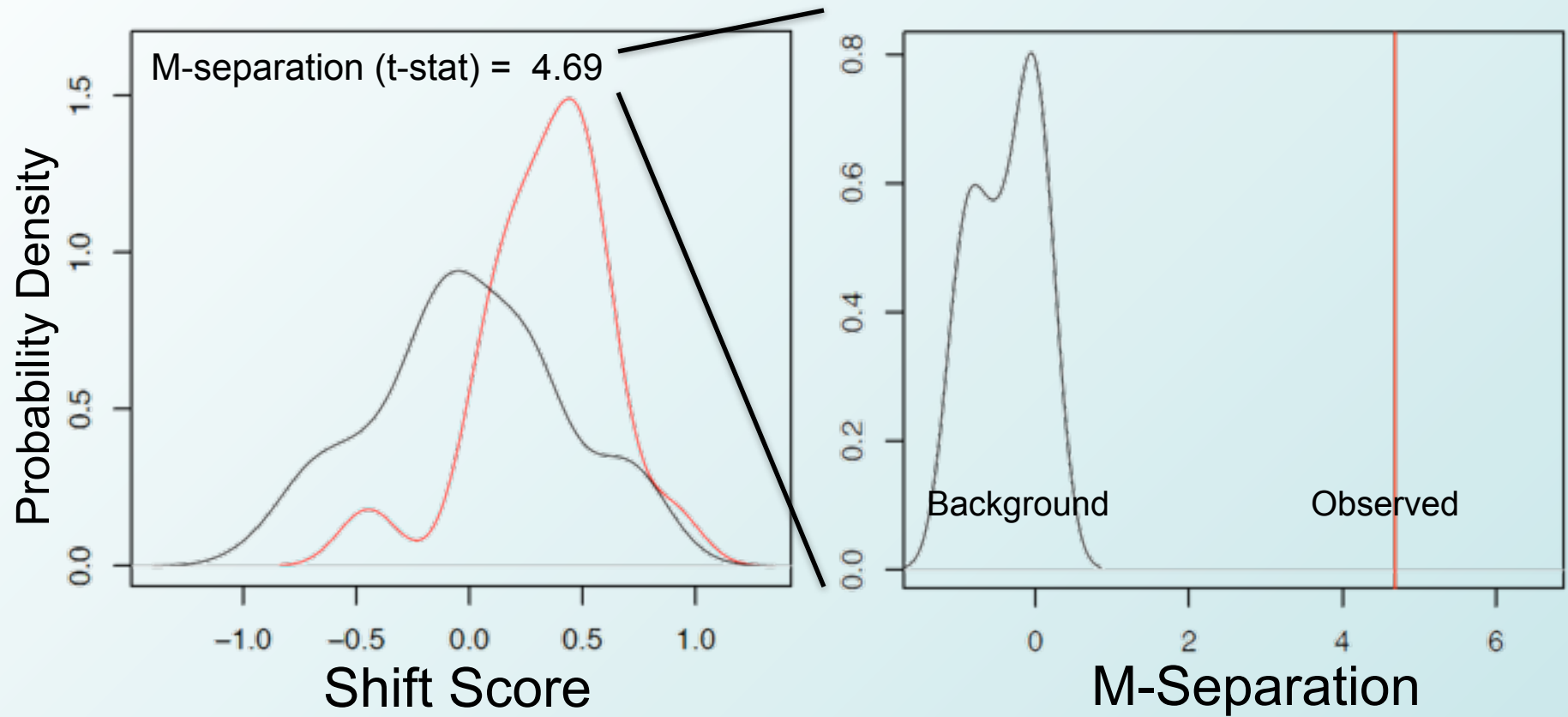


NFE2L2 Gain-of-Function (LUSC)

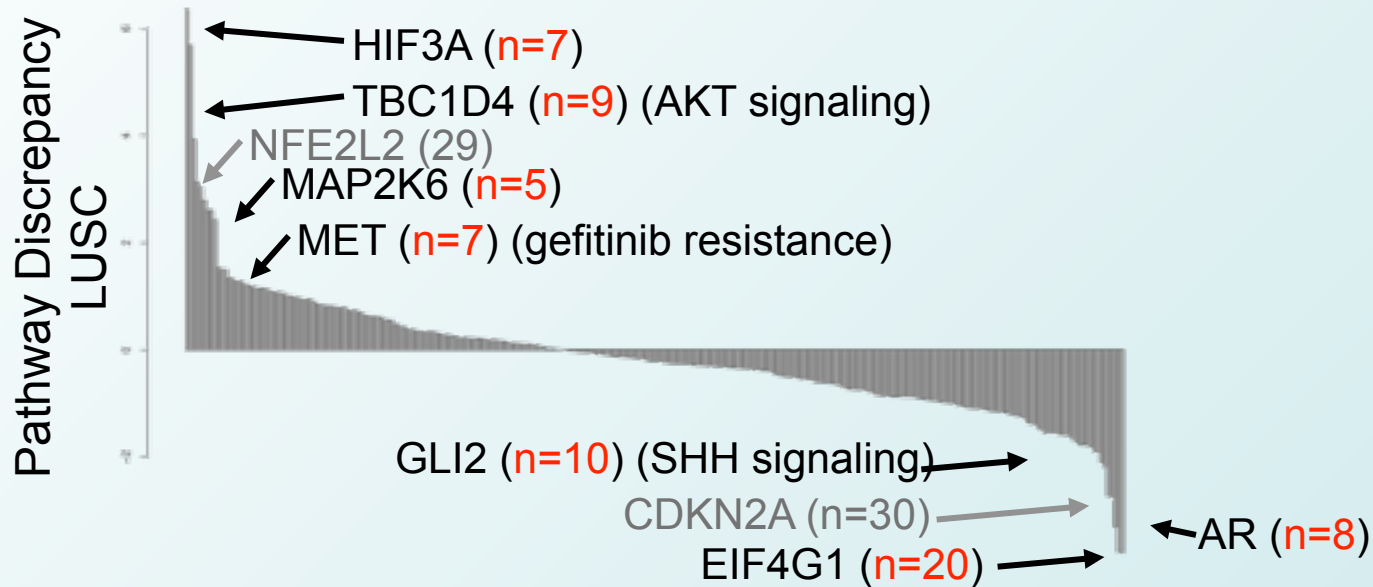
Upstream and Downstream Genes



NFE2L2 Mutant Separation and Significance



PARADIGM-Shift gives orthogonal view of the importance of mutations



- Enables probing into infrequent events
- Can detect non-coding mutation impact (pseudo FPs)
- Can detect presence of pathway compensation for those seemingly functional mutations (pseudo FPs)
- Extend beyond mutations
- Limited to genes w/ pathway representation

Two Themes



1) Predict
Drivers



2) Predict
Essential



PATHMARK: Identify Pathway-based “markers” that underlie sub-types

- Identify sub-pathways that distinguish patients sub-types (e.g. mutant vs. non-mutant, response to drug, etc) Insight from contrast
- Predict mutation impact on pathway “neighborhood”
- Identify master control points for drug targeting.
- Predict outcomes with quantitative simulations.



Sam Ng



Ted Goldstein



Defining Pathway Signatures for Mutations and Sub-Types

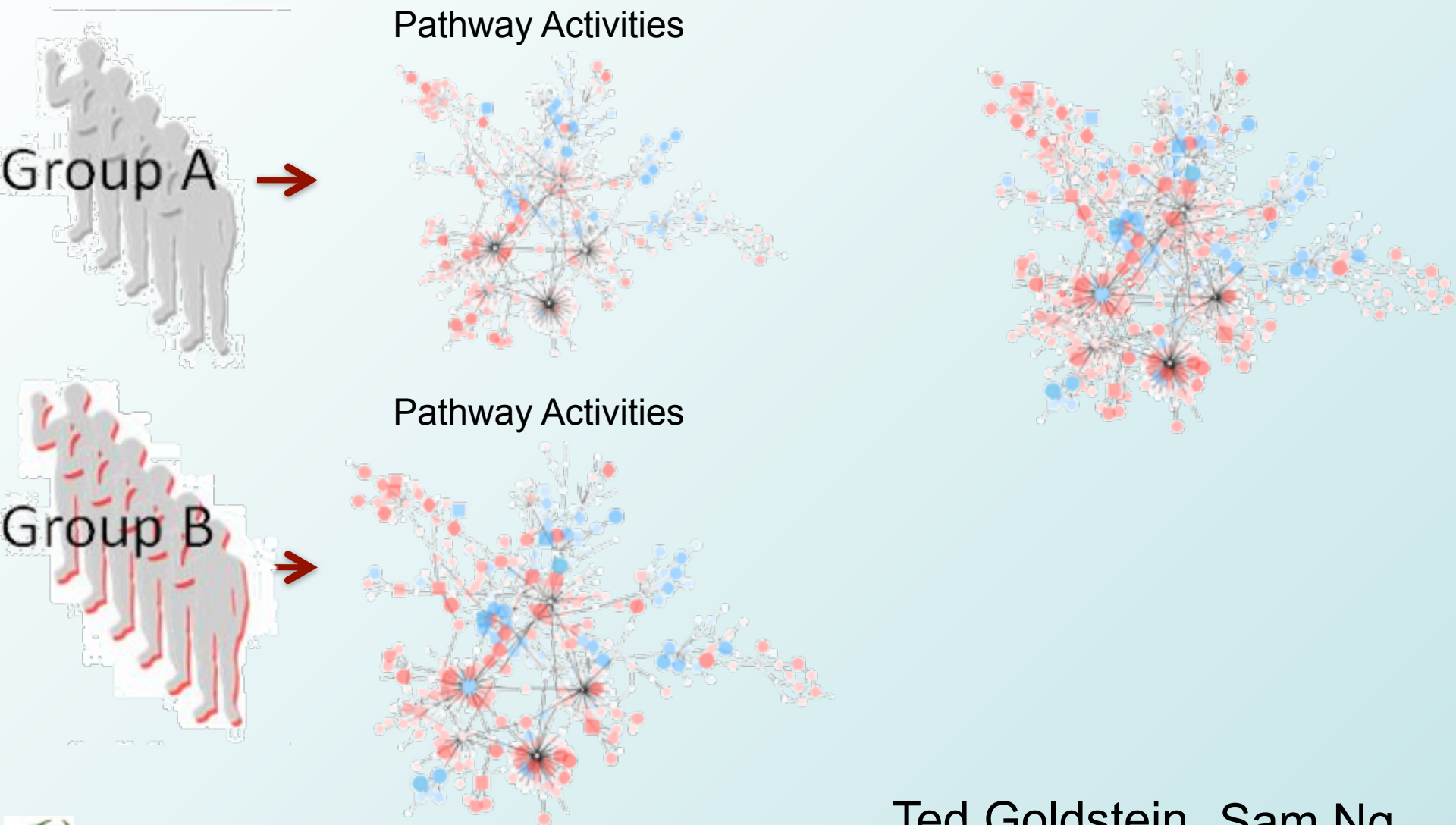
- Build a signature for every mutation and tumor/clinical event.
- Correlate every signature to each other.
- Reveals common molecular similarities between different divisions of patient subgroups
- Mutations in novel genes may “phenocopy” mutations in known genes



Ted Goldstein



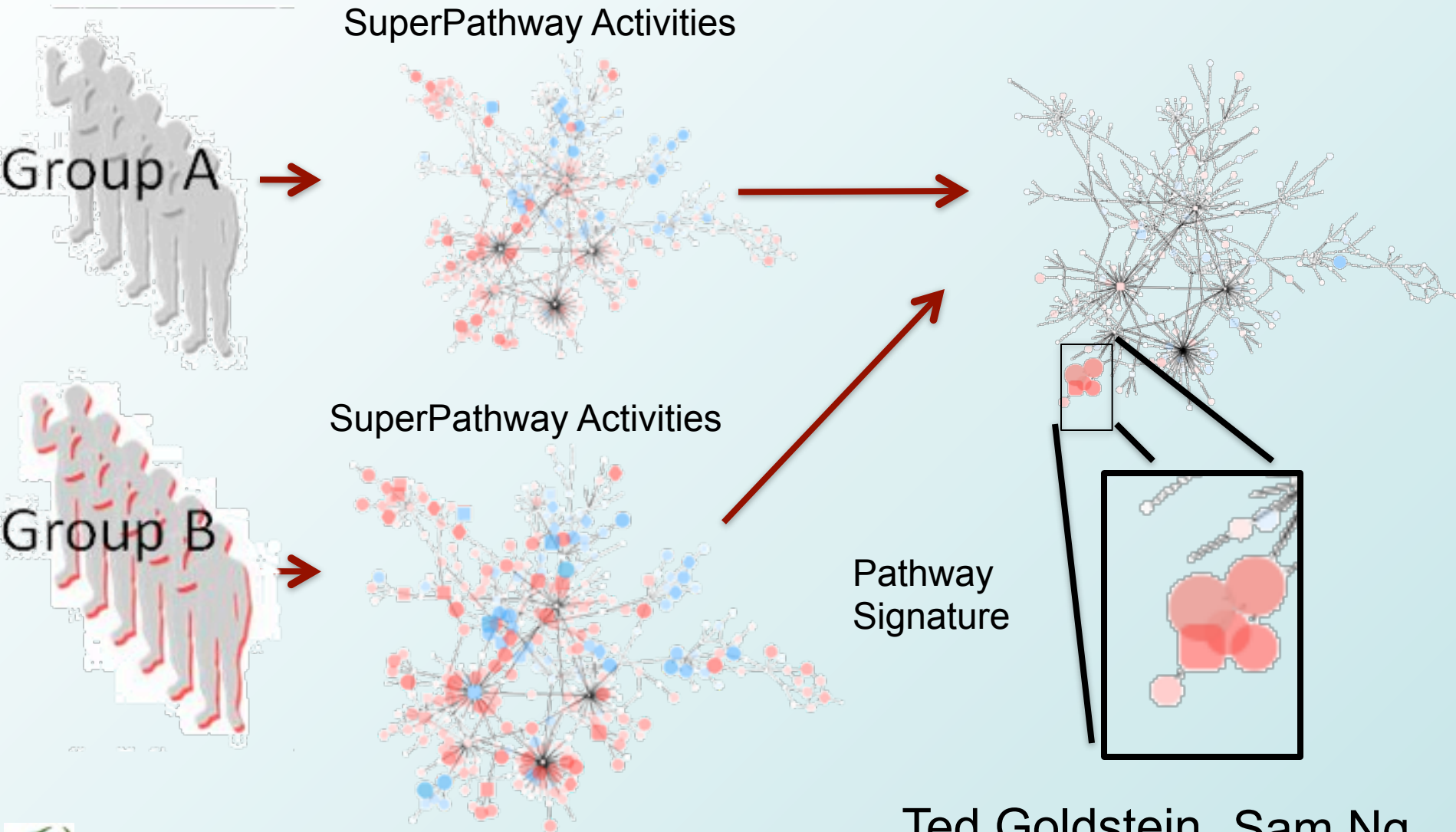
PathMark: Differential Subnetworks from a “SuperPathway”



Ted Goldstein Sam Ng



PathMark: Differential Subnetworks from a “SuperPathway”



Ted Goldstein Sam Ng

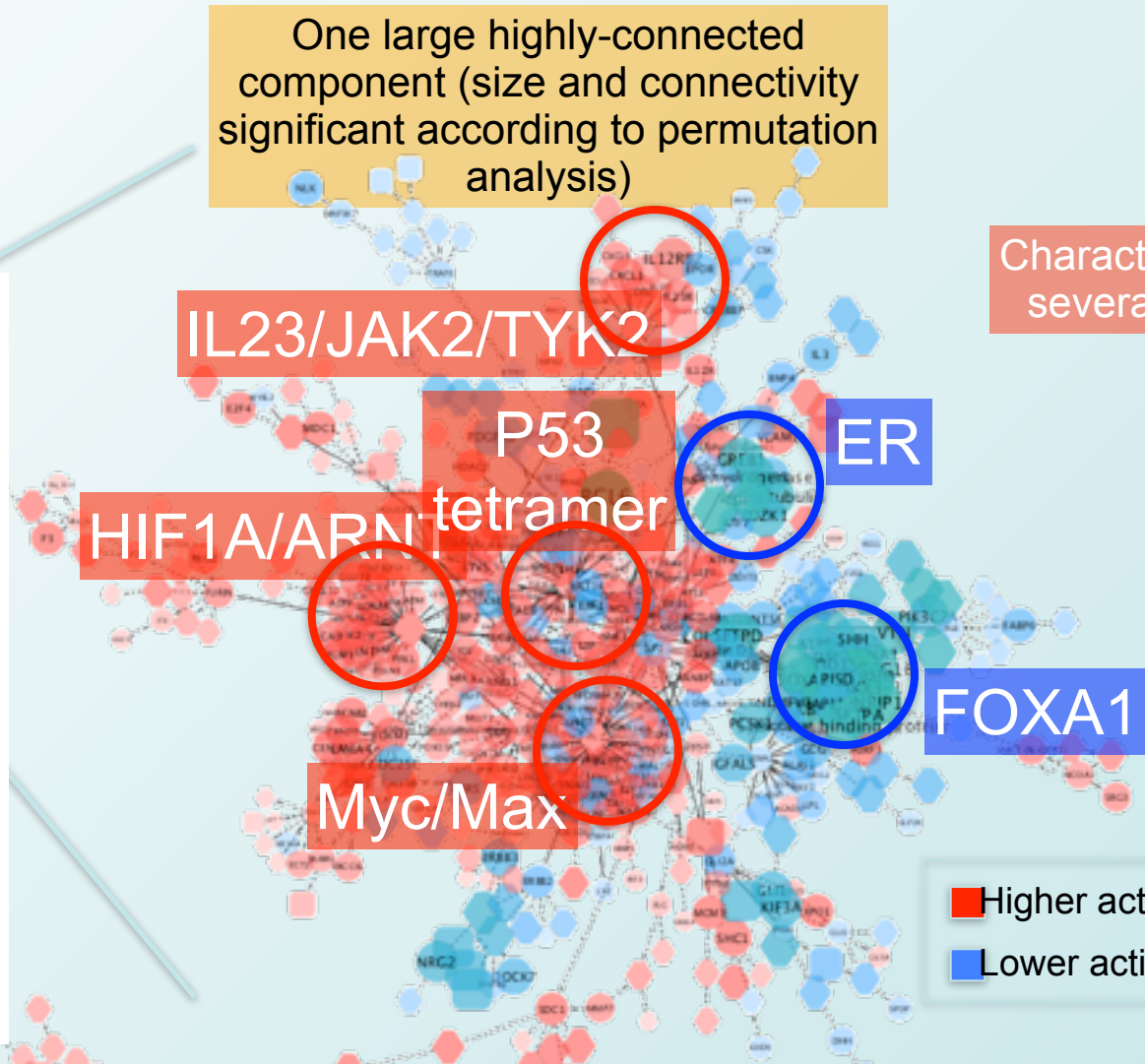


Basal Breast Pathway Markers

980 pathway concepts
1048 interactions

One large highly-connected component (size and connectivity significant according to permutation analysis)

Characterized by several “hubs”



Sam Ng, Ted Goldstein

Identify master controllers using SPIA (signaling pathway impact analysis)

- Google PageRank for Networks
- Determines affect of a given pathway on each node
- Calculates perturbation factor for each node in the network
- Takes into account regulatory logic of interactions.

Impact factor:

$$IF(g_i) = s(g_i) + \beta_{ij} \times \frac{IF(g_j)}{N_{up}(g_j)}$$

Systems biology

A novel signaling pathway impact analysis

Adi Laurentiu Tarca^{1,2}, Sorin Draghici^{1,*}, Purvesh Khatri¹, Sonia S. Hassan², Pooja Mittal², Jung-sun Kim², Chong Jai Kim², Juan Pedro Kusanovic² and Roberto Romero²

¹Department of Computer Science, Wayne State University, 431 State Hall, Detroit, MI 48202 and

²Perinatology Research Branch-NIH/NICHD, 4 Brush, 3990 John R, Detroit, MI 48201, USA

Received on July 28, 2008; revised on October 29, 2008; accepted on November 4, 2008



Identify master controllers using SPIA (signaling pathway impact analysis)

- Google PageRank for Networks
- Determines affect of a given pathway on each node
- Calculates perturbation factor for each node in the network
- Takes into account regulatory logic of interactions.

Impact
factor:

$$IF(g_i) = s(g_i) + \beta_{ij} \times \frac{IF(g_j)}{N_{up}(g_j)}$$

Systems biology

Google's PageRank-Like

A novel signaling pathway impact analysis

Adi Laurentiu Tarca^{1,2}, Sorin Draghici^{1,*}, Purvesh Khatri¹, Sonia S. Hassan², Pooja Mittal², Jung-sun Kim², Chong Jai Kim², Juan Pedro Kusanovic² and Roberto Romero²

¹Department of Computer Science, Wayne State University, 431 State Hall, Detroit, MI 48202 and

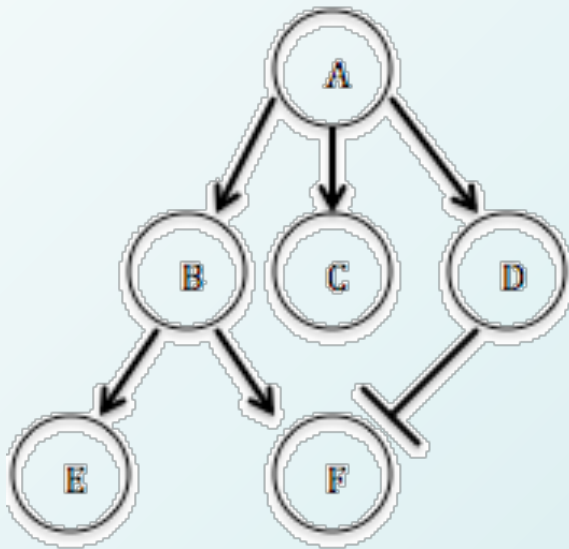
²Perinatology Research Branch-NIH/NICHD, 4 Brush, 3990 John R, Detroit, MI 48201, USA

Received on July 28, 2008; revised on October 29, 2008; accepted on November 4, 2008



Slight Trick: Run SPIA in reverse

- Reverse edges in Super Pathway
- High scoring genes now those at the “top” of the pathway

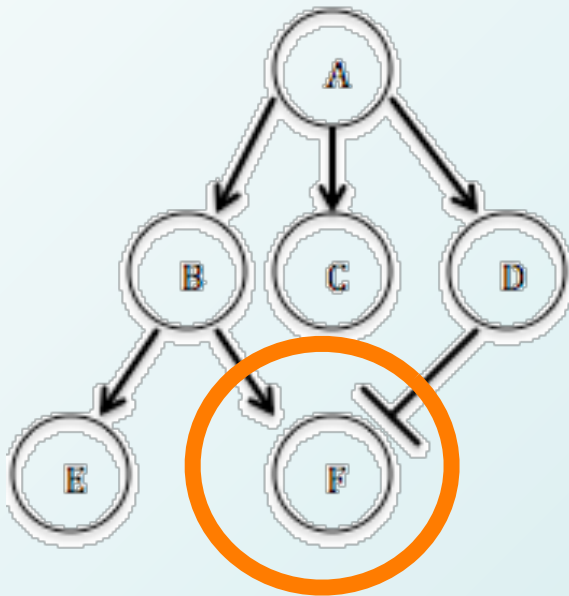


PageRank finds
highly referenced



Slight Trick: Run SPIA in reverse

- Reverse edges in Super Pathway
- High scoring genes now those at the “top” of the pathway

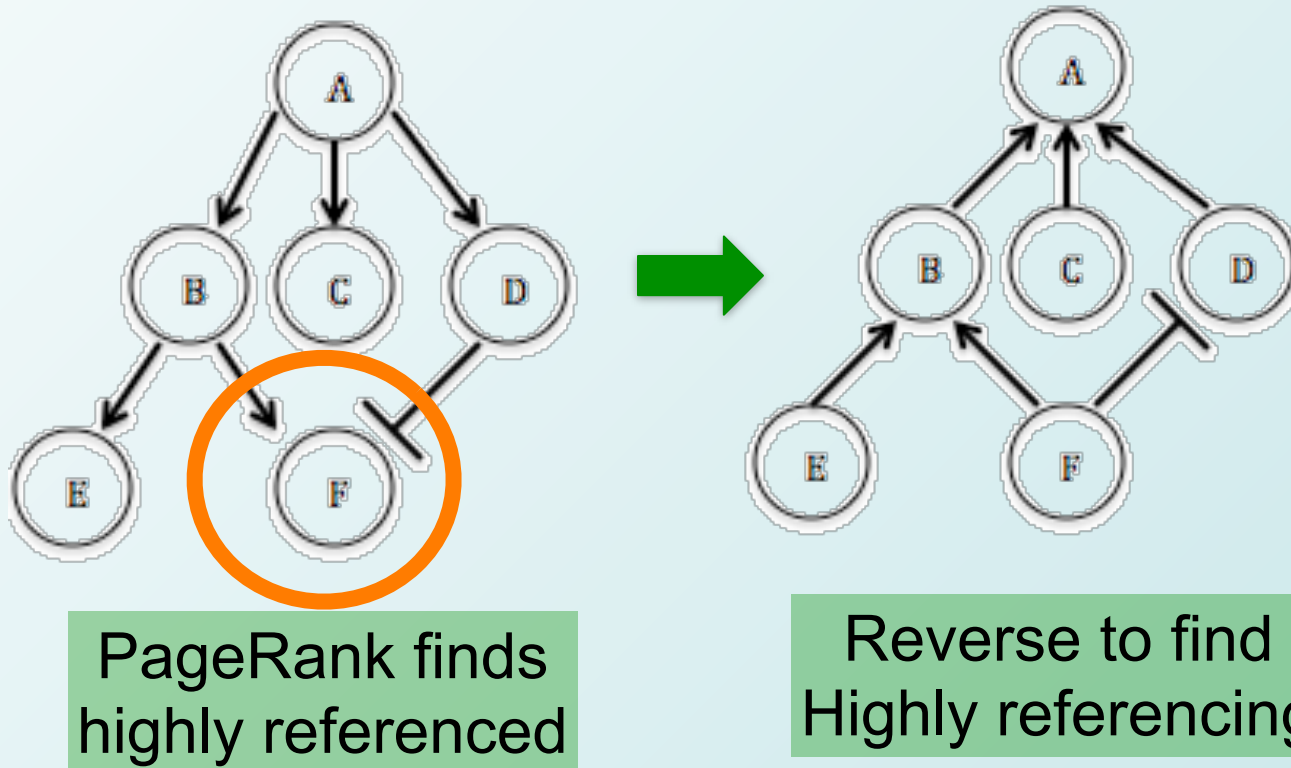


PageRank finds
highly referenced



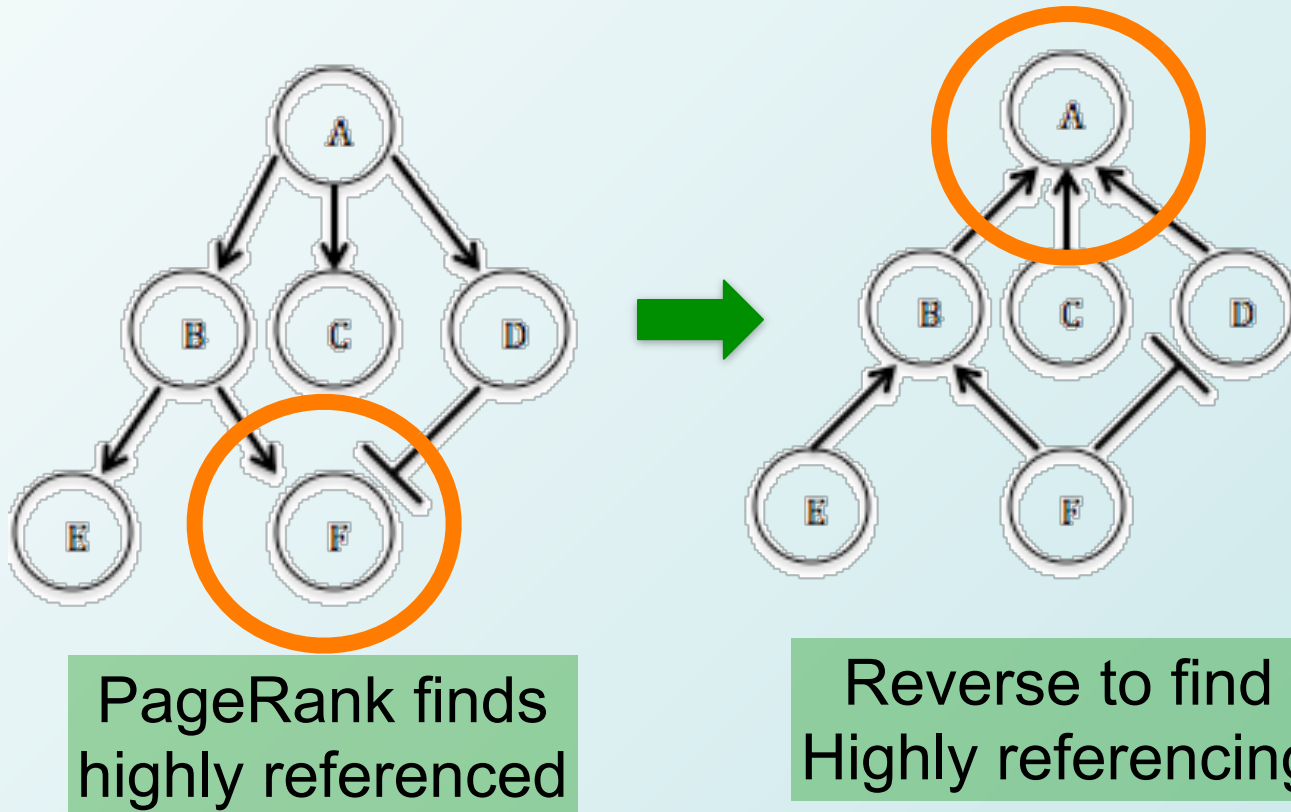
Slight Trick: Run SPIA in reverse

- Reverse edges in Super Pathway
- High scoring genes now those at the “top” of the pathway

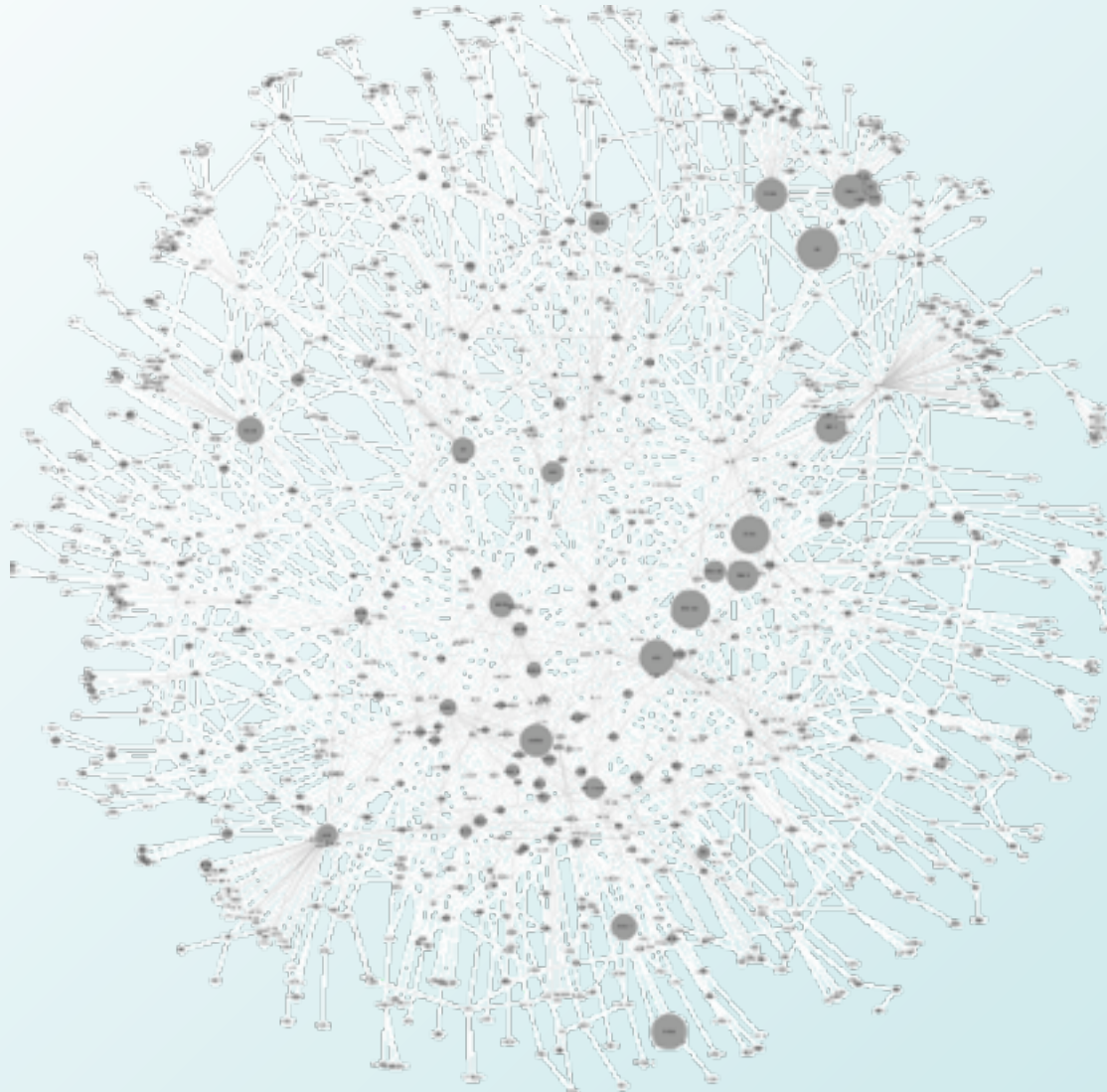


Slight Trick: Run SPIA in reverse

- Reverse edges in Super Pathway
- High scoring genes now those at the “top” of the pathway



Master Controller Analysis on Breast Cell Lines



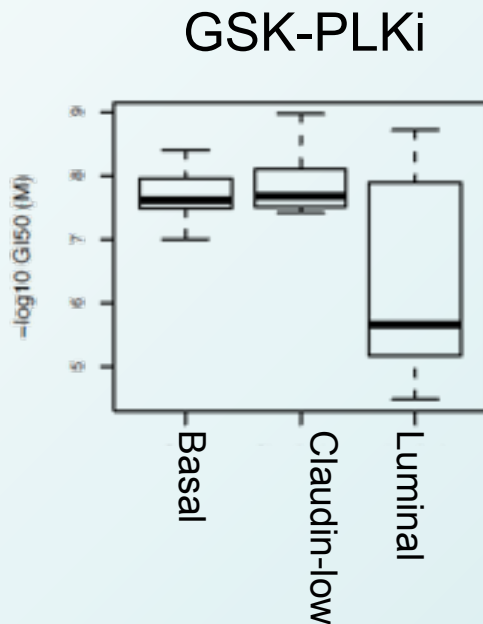
Basal
Luminal

Yulia Newton



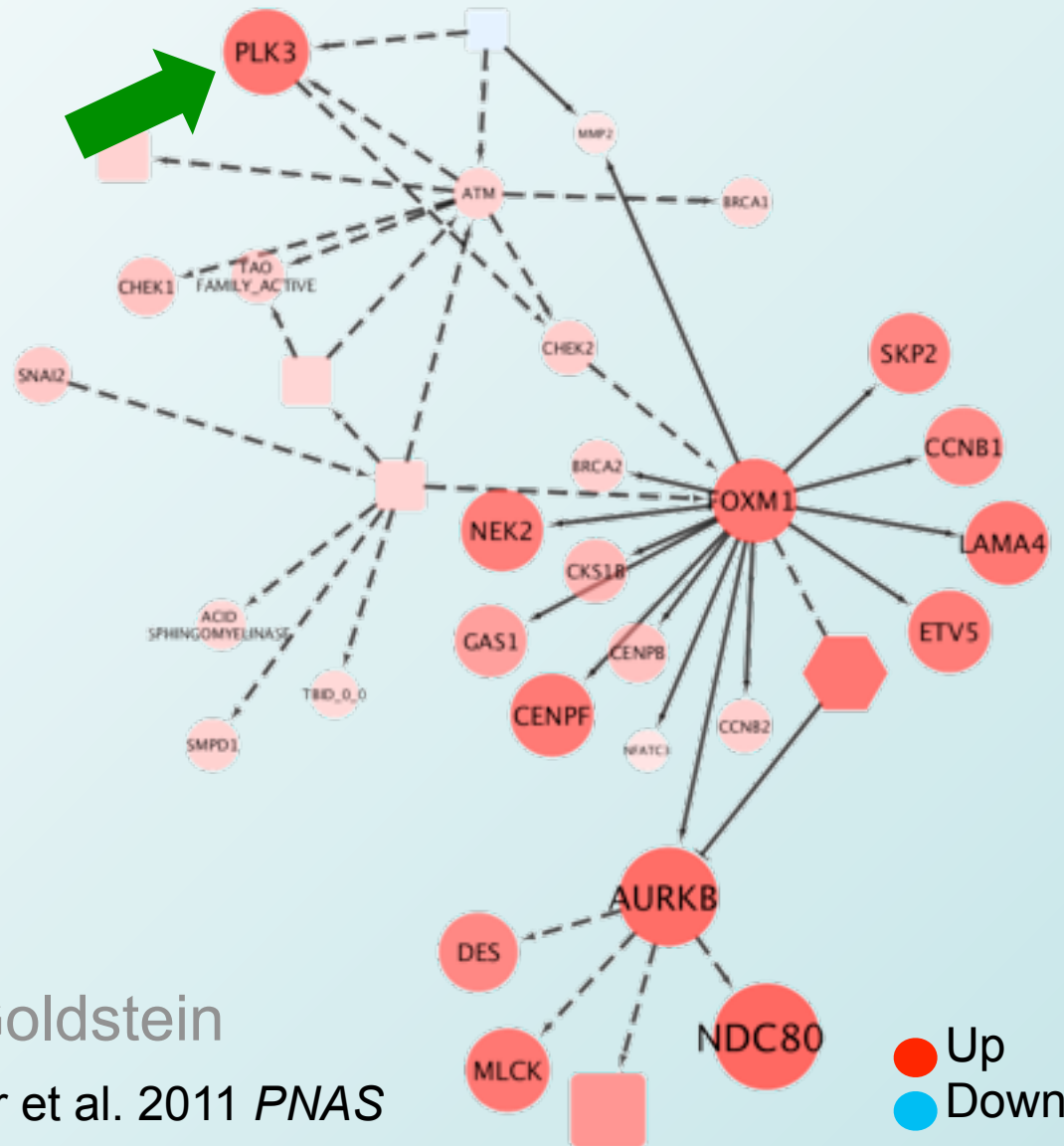
Master regulators predict response to drugs:

- DNA damage network is upregulated in basal breast cancers
- Basal breast cancers are sensitive to PLK inhibitors



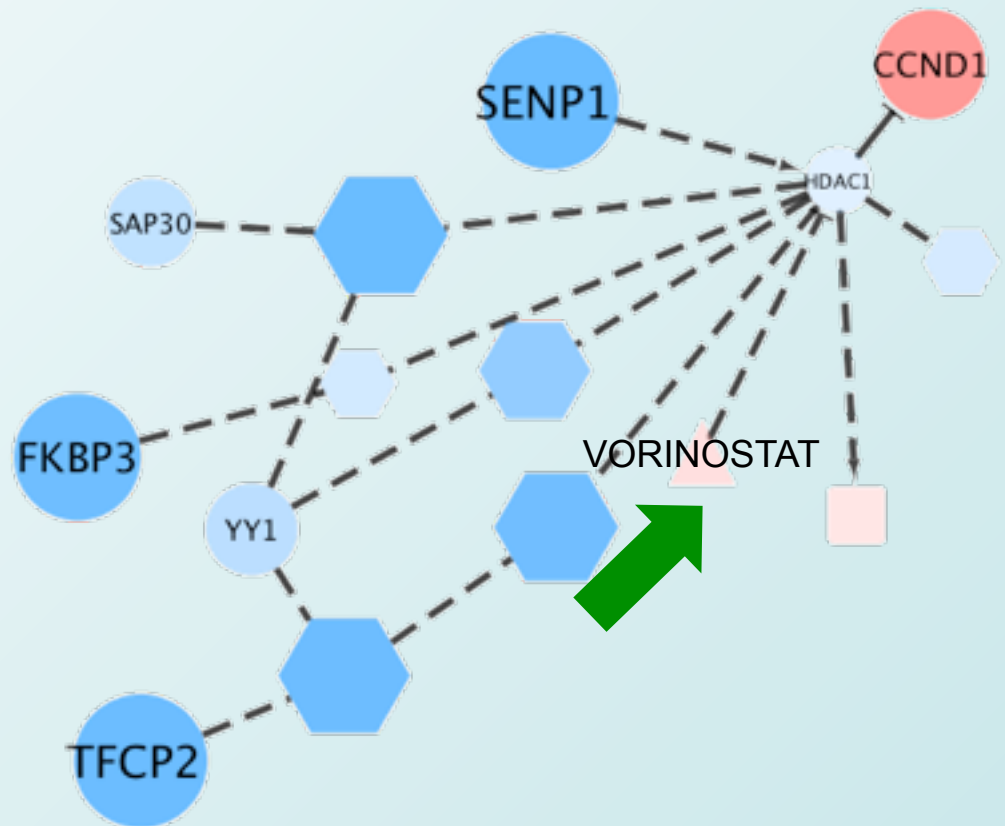
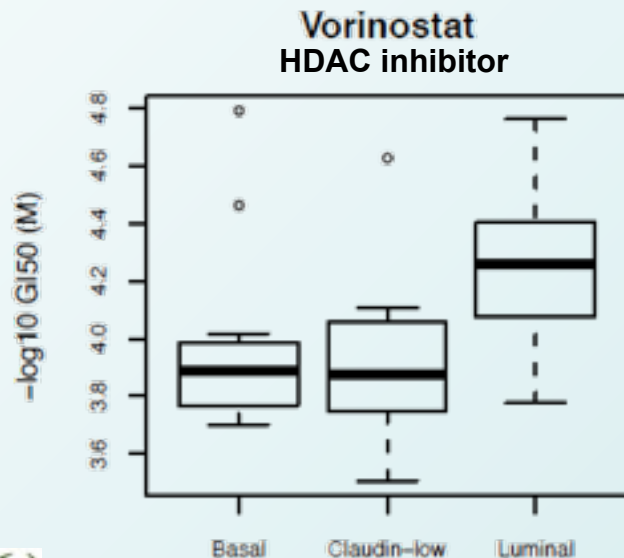
Ng, Goldstein

Heiser et al. 2011 *PNAS*

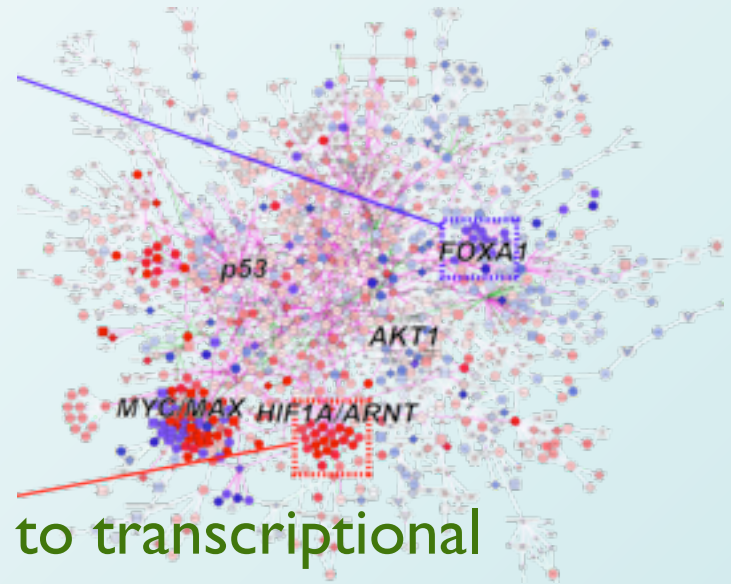
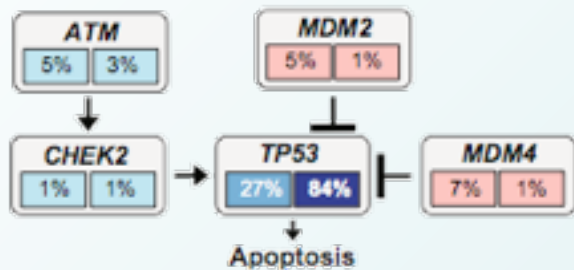


HDAC inhibitors predicted for luminal breast

- HDAC Network is down-regulated in basal breast cancer cell lines
- Basal/CL breast cancers are resistant to HDAC inhibitors



Essential genes mediate signaling in cancer



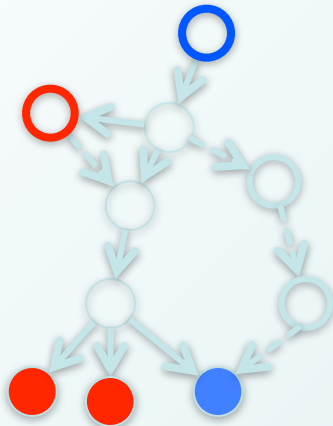
- What circuitry connects mutations to transcriptional changes?
 - Mutations → general (epi-) genomic perturbation
 - Expression → activity
- Mutation/perturbation and expression/activity treated as heat diffusing on a network
 - **HotNet**, Vandin F, Upfal E, B.J. Raphael, 2008.
 - HotNet used in ovarian to implicate Notch pathway
- Find subnetworks that link genetic to mRNA and protein-level changes.



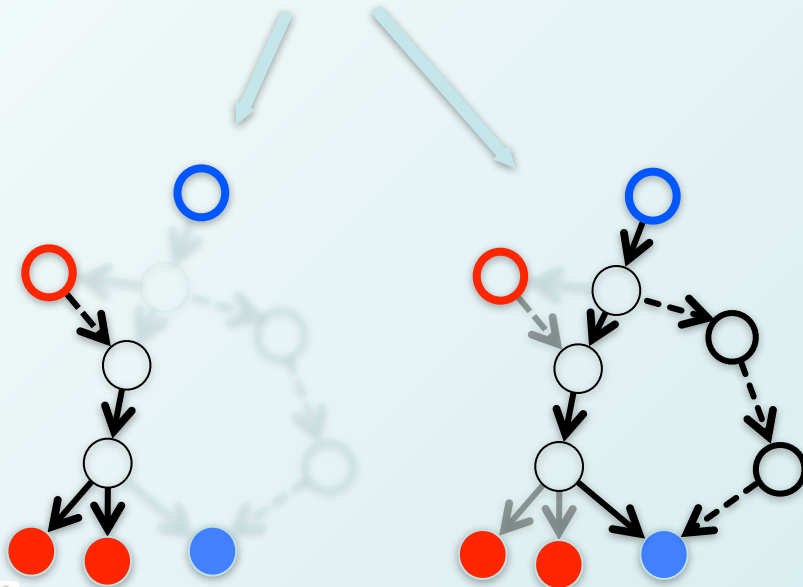
Evan Paull
ISMB
Oral Poster



TieDIE: Problem Definition



- Find logically consistent paths connecting source to target genes (“causal” paths).
- Generate many reasonable, testable hypotheses that explain the tumor state.



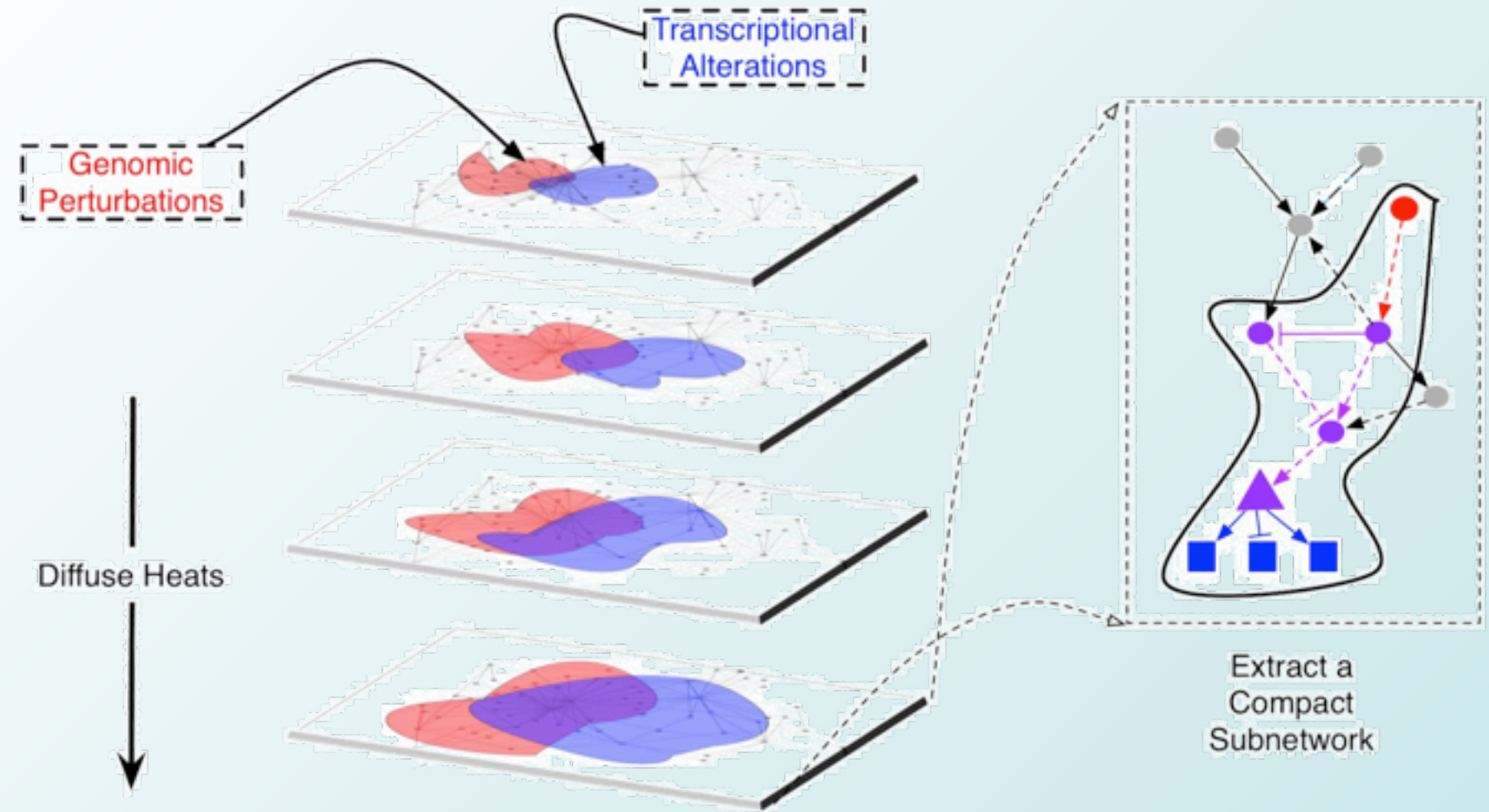
Mutation/perturbation and expression/activity treated as heat diffusing on a network

-**HotNet**, Vandin F, Upfal E, B.J. Raphael, 2008.

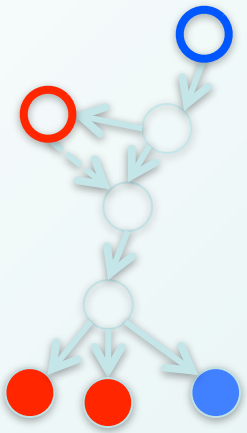
-**HotNet** used in ovarian to implicate Notch pathway



TieDIE “Double Diffusion” Overview



Evaluation

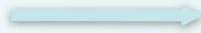
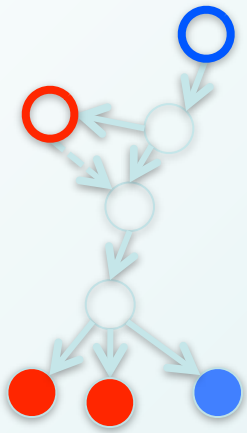


Evaluation

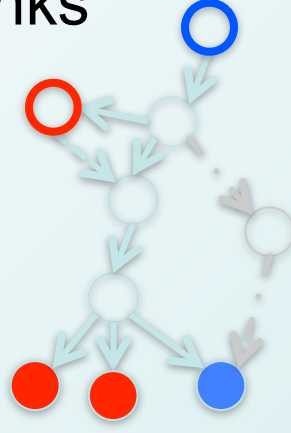
Add “Dummy”
Links



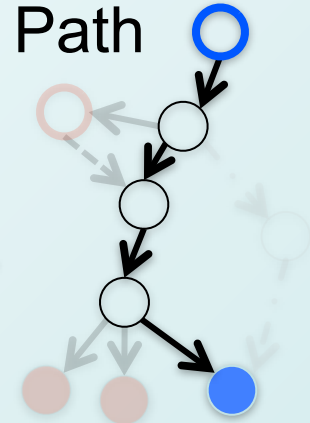
Evaluation



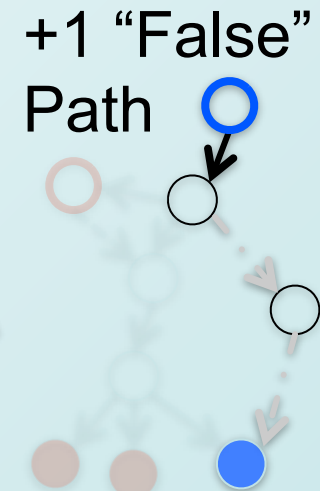
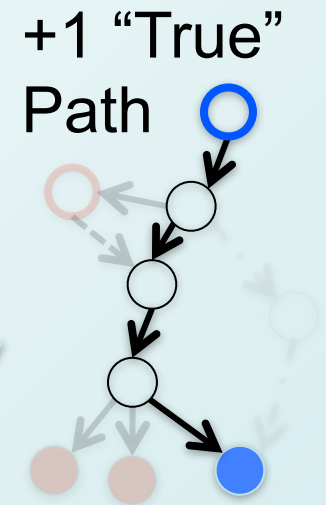
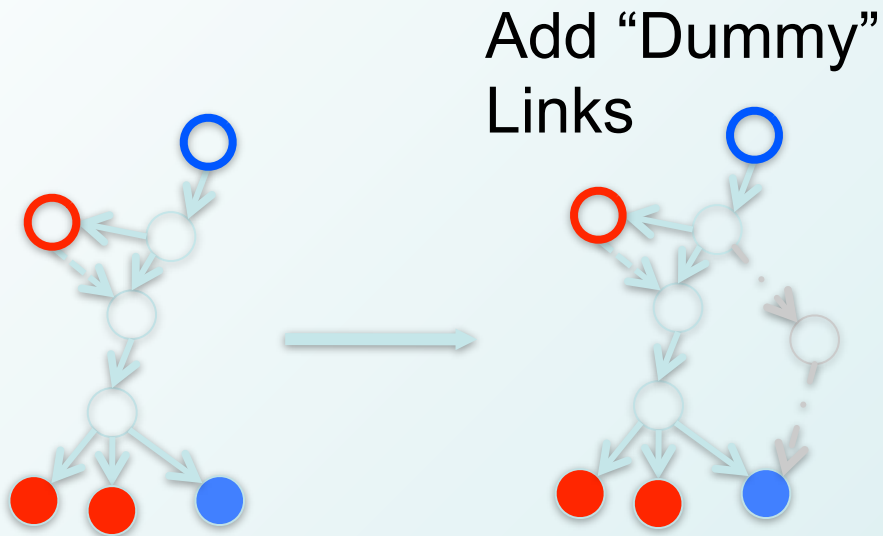
Add “Dummy”
Links



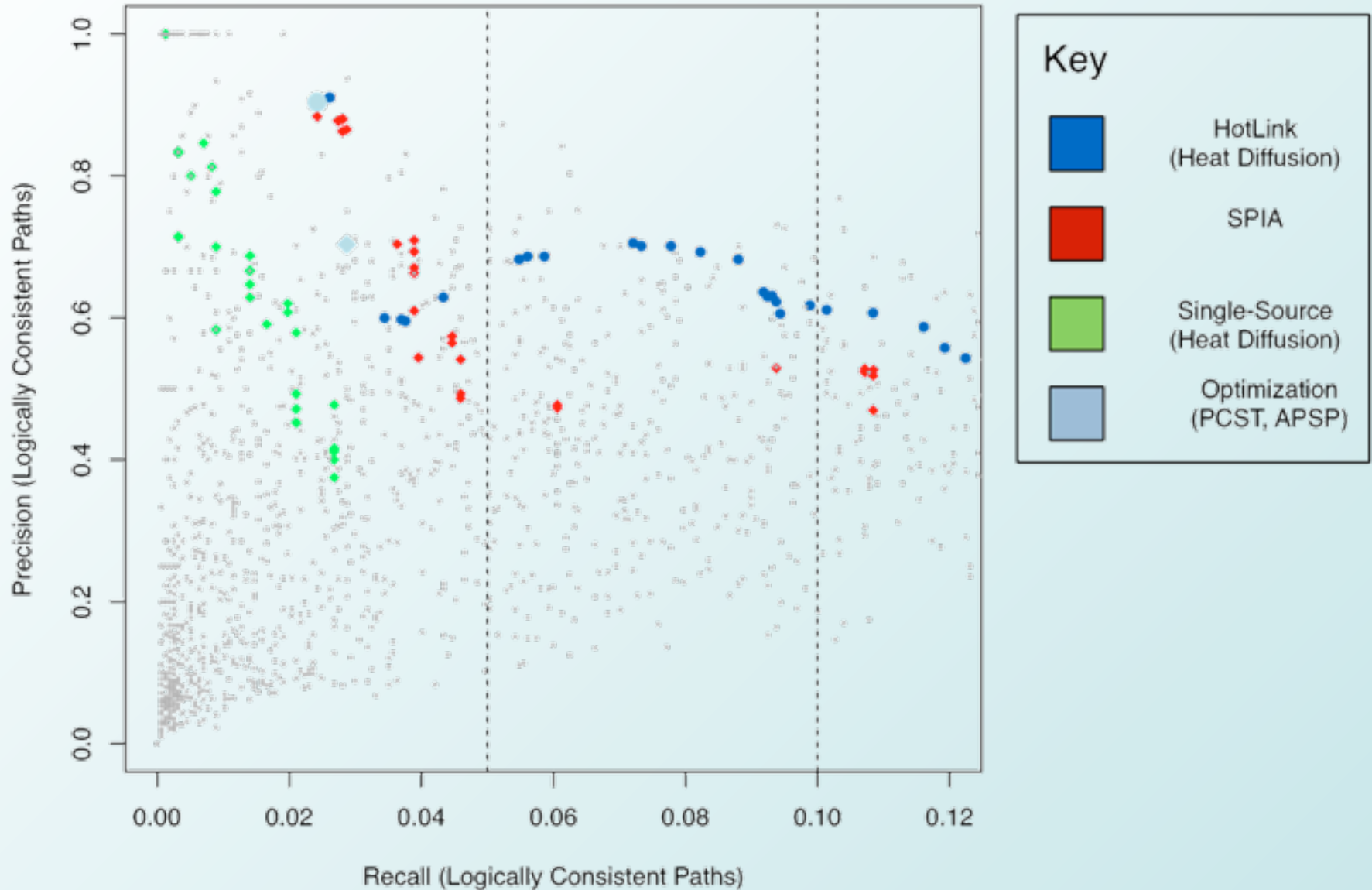
+1 “True”
Path



Evaluation



Double diffusion finds more logically coherent



TieDIE,TCGA Breast Cancer (BRCA) Analysis

- patient tumor/match-normal samples for 533 patients, each with genomic sequencing data and microarray expression
- Differential analysis: Basal vs Luminal A subtypes
- “Source” set: 12 genomic perturbations
- “Target” set: 370 differentially expressed transcription factors



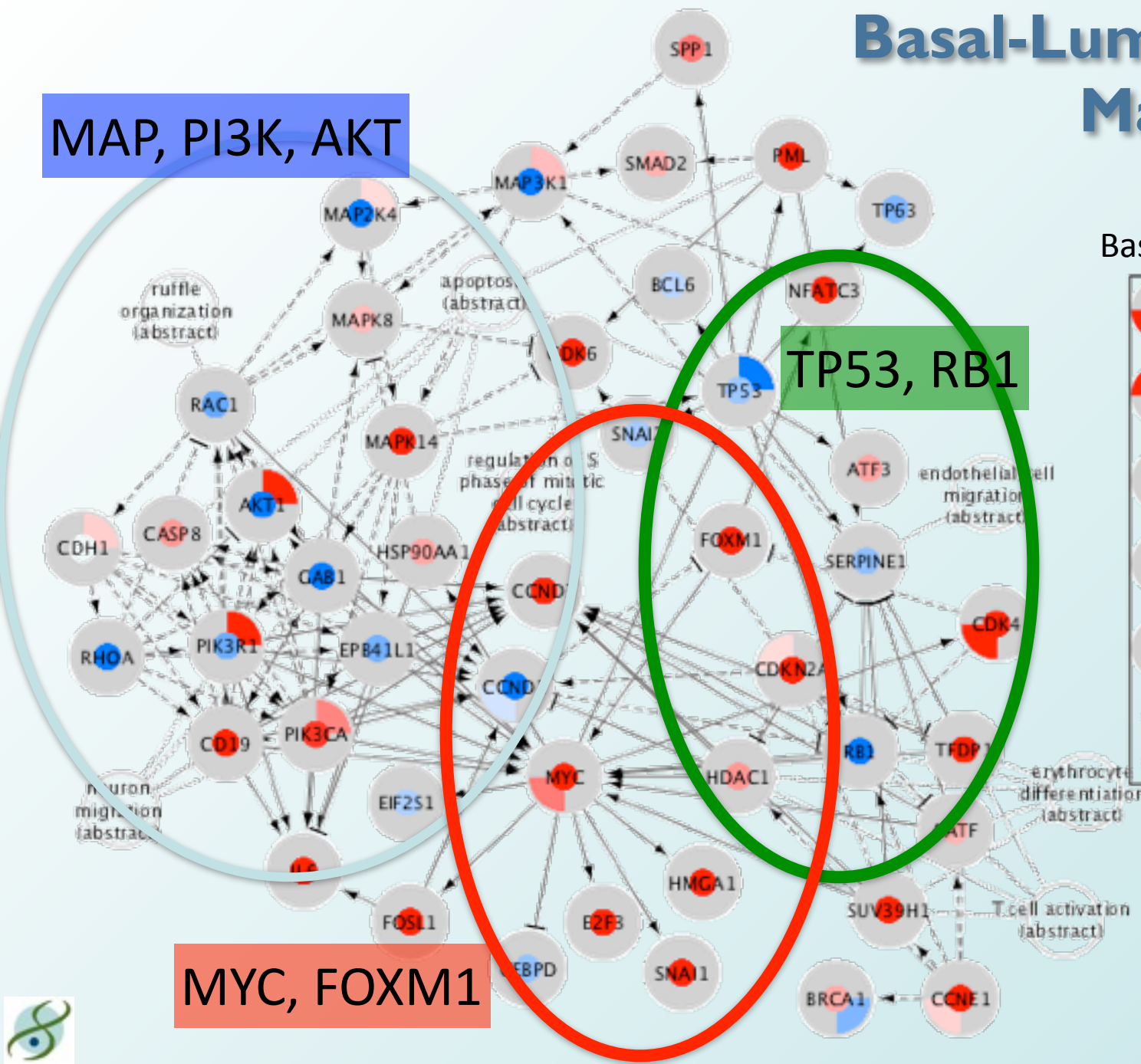
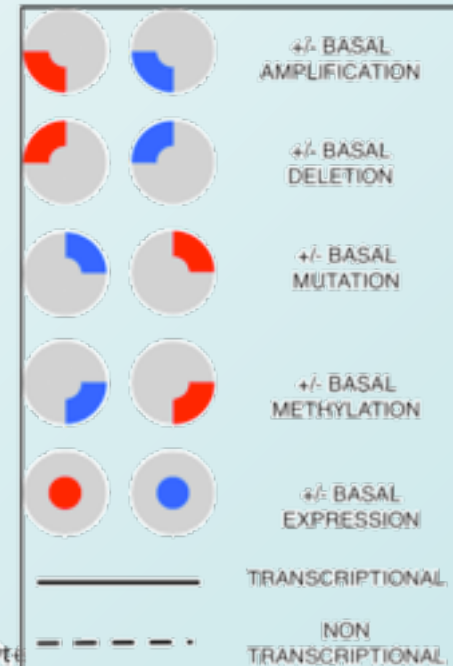
Basal-LumA TieDIE Map

MAP, PI3K, AKT

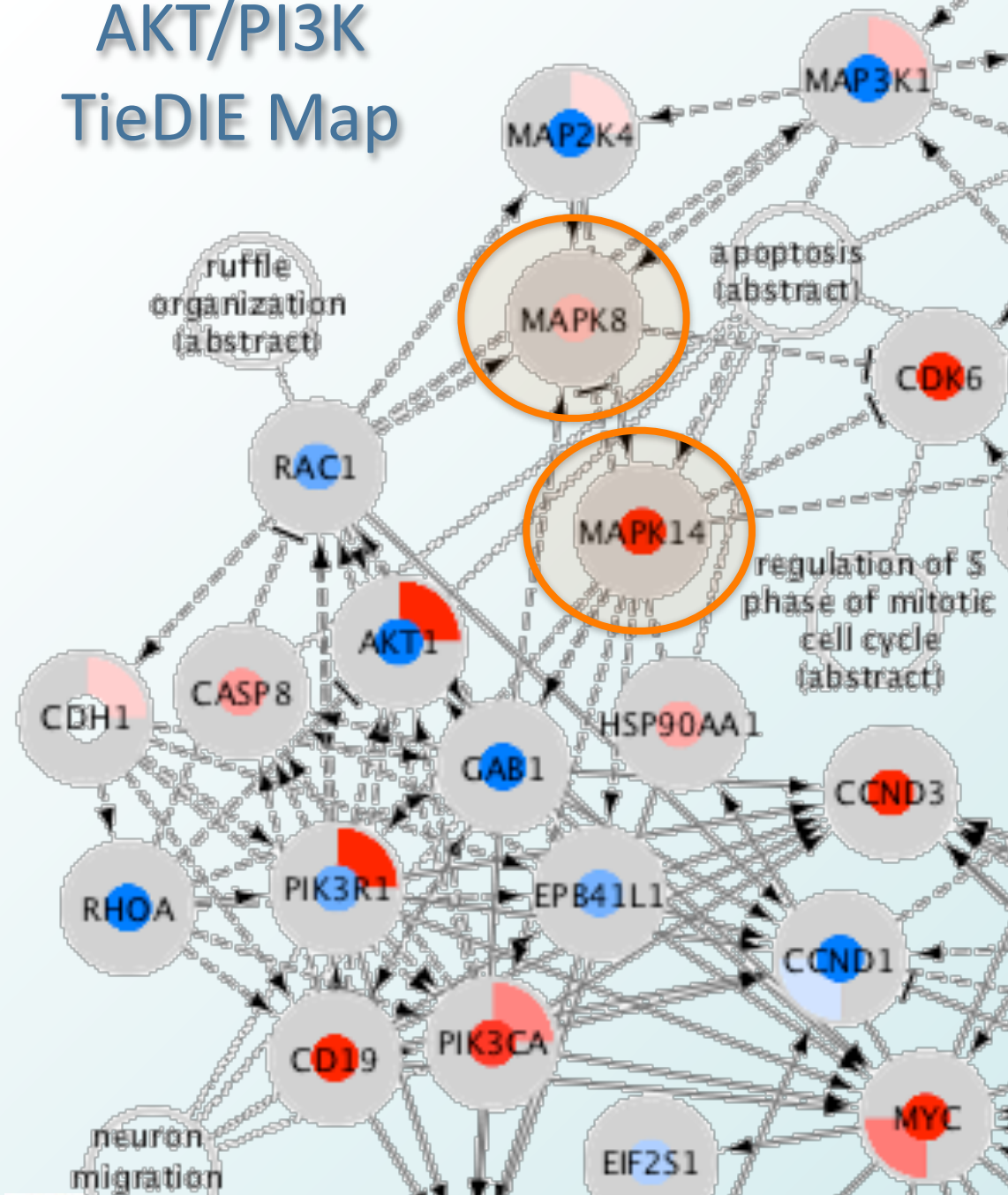
Basal LumA

TP53, RB1

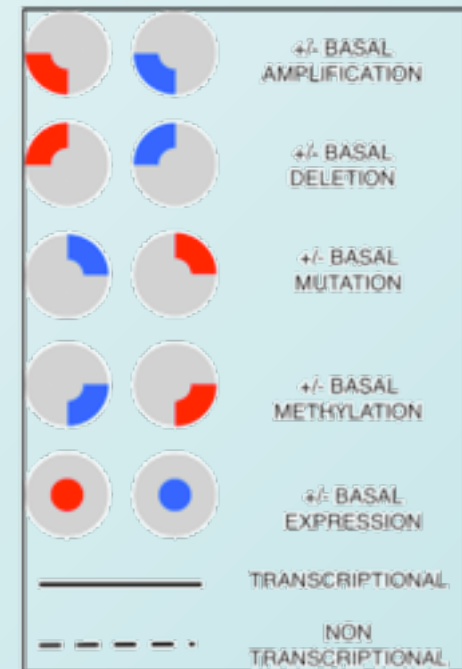
MYC, FOXM1



AKT/PI3K TieDIE Map

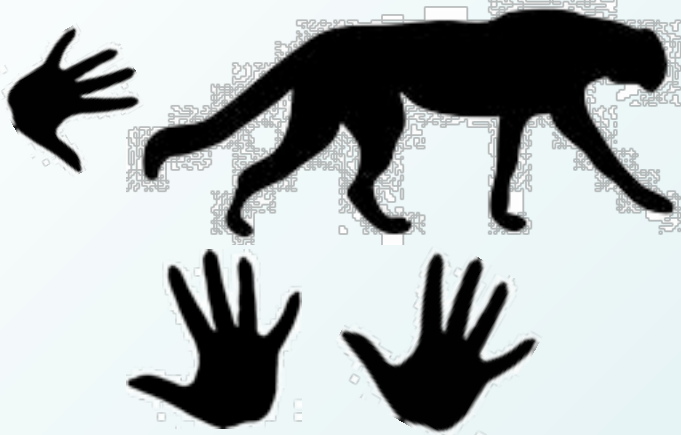


MAPK8, MAPK14 (p38-alpha)
identified as mediators.

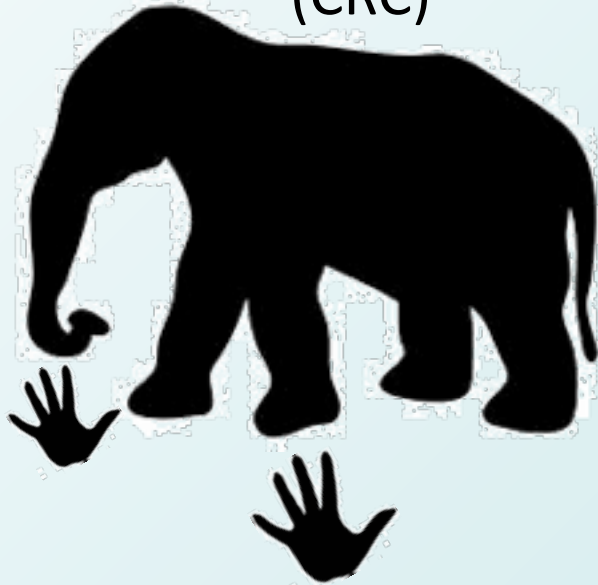


TCGA: Multiple views of many beasts

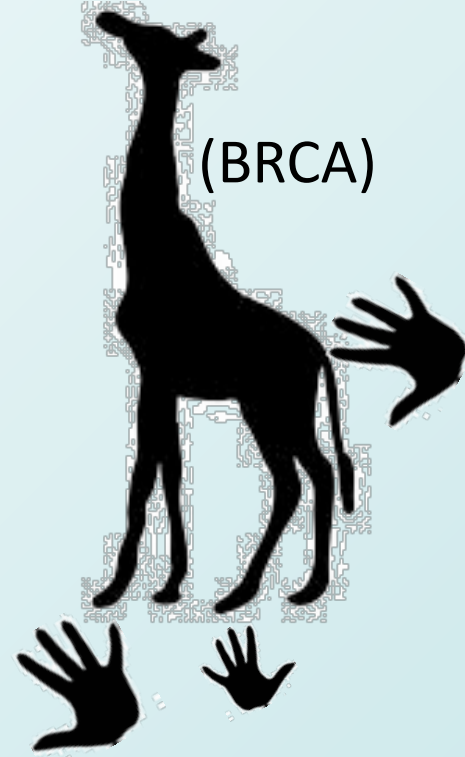
(GBM)



(CRC)



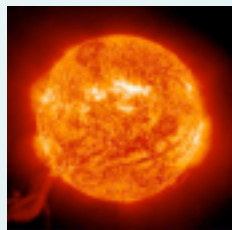
(BRCA)



Motivations for Comparative “Pan-Cancer”

- Statistical power increase
- More than additive gained by comparative analysis?
- Is TP53 mutation associated with the same transcriptional outcomes?
- Are there examples of core cancer signals that transcend cell of origin? Core subtypes?
- Can these subtypes be treated therapeutically similar?

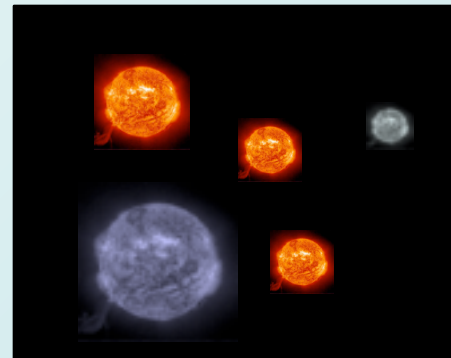
Analyze **One**
Star



- composition
- motion
- mass

vs.

Analyze **Many**
Stars



- types
- ages
- interactions
- evolution



Discover Pan-Cancer Sub-Types

PI3K



WNT / MYC



Reveal common
Pathway alterations

unknown



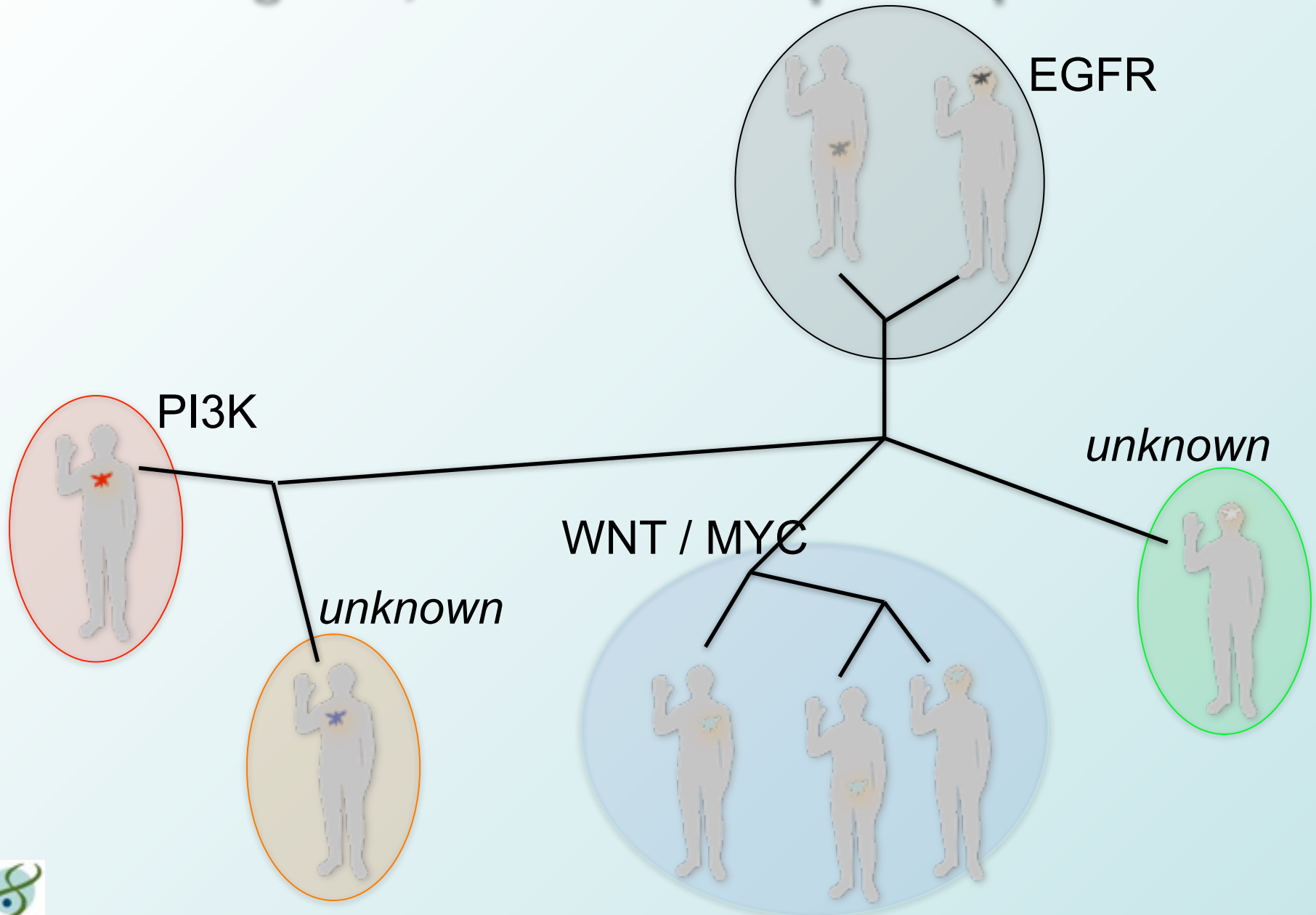
unknown



EGFR



Define a global, data-driven map of all pan-cancer



What is the TCGA Pan-Cancer Project?

➤ Multi-institution collaborative analysis of currently 12 tumor types

➤ Integrated data for ~5000 samples *

- 3808 w/ mRNA-seq
- 5191 w/ copy number
- 5074 w/ methylation
- 3082 w/ mutation data
- 2785 w/ proteomics data (RPPA)
- 4306 w/ miRNA-seq

➤ Thematic Analyses

- Mutation trends, impact
- Structural landscapes
- Non-coding, splicing
- Pathogens, pseudogenes
- Intrinsic subtypes
- Pathways, networks

(credit: K. Hoadley)



The hang-loose Santa Cruz vibe lives on



* Based on 3/26/13 Firehose Report, Mike Noble, Broad



The PanCan12 DataSet (Level 3 data)

Tissue in PanCan12

2012_12_21 stddata Run

PanCan12

Tumor	BCR	Clinical	CN	Low Methylation	miRNA	miRNAseq	miR	miRseq	RPPA	MAF	
BLCA	153	108	137	100	138	0	96	0	124	54	28
BRCA	929	866	874	0	888	529	841	0	868	408	507
CESC	134	32	102	0	122	0	97	0	122	0	36
COAD	423	423	413	69	420	155	192	0	407	269	155
COADREAD	592	591	575	104	582	224	264	0	550	399	224
DLBC	28	0	18	0	17	0	0	0	16	0	0
ESCA	20	0	0	0	0	0	0	0	0	0	0
GBM	598	565	563	0	405	542	161	491	0	214	291
HNSC	343	315	306	96	310	0	303	0	309	212	306
KICH	66	0	65	0	65	0	65	0	0	0	0
KIRC	502	502	493	0	500	72	469	0	480	454	293
KIRP	149	103	103	0	103	16	63	0	103	0	100
LAML	202	200	0	0	194	0	179	0	187	0	199
LGG	222	208	180	0	176	27	110	0	180	0	34
LIHC	99	62	97	0	98	0	17	0	96	0	0
LUAD	508	333	356	0	430	32	353	0	365	237	229
LUSC	389	327	343	0	359	154	223	0	332	195	178
OV	592	580	566	0	584	574	297	570	454	412	316
PAAD	57	0	48	0	40	0	0	0	34	0	34
PANCAN12	5320	4838	4706	406	4890	2181	3303	1061	4154	2785	2819
PANCAN18	6851	5815	6071	601	6151	2224	4242	1061	5330	3173	3728
PRAD	180	148	171	0	172	0	140	0	170	0	83
READ	169	168	162	35	162	69	72	0	143	130	69
SARC	52	0	29	0	29	0	0	0	29	0	0
SKCM	288	138	273	101	253	0	247	0	240	164	253
STAD	257	162	237	0	145	0	43	0	134	0	116
THCA	435	218	401	94	412	0	379	0	349	224	323
UCEC	512	451	493	106	500	54	333	0	485	200	248
Totals	7307	5909	6430	601	6522	2224	4680	1061	5627	3173	3798

The PanCan12 DataSet (Level 3 data)

Tissue in PanCan12

2012_12_21 stddata Run

PanCan12

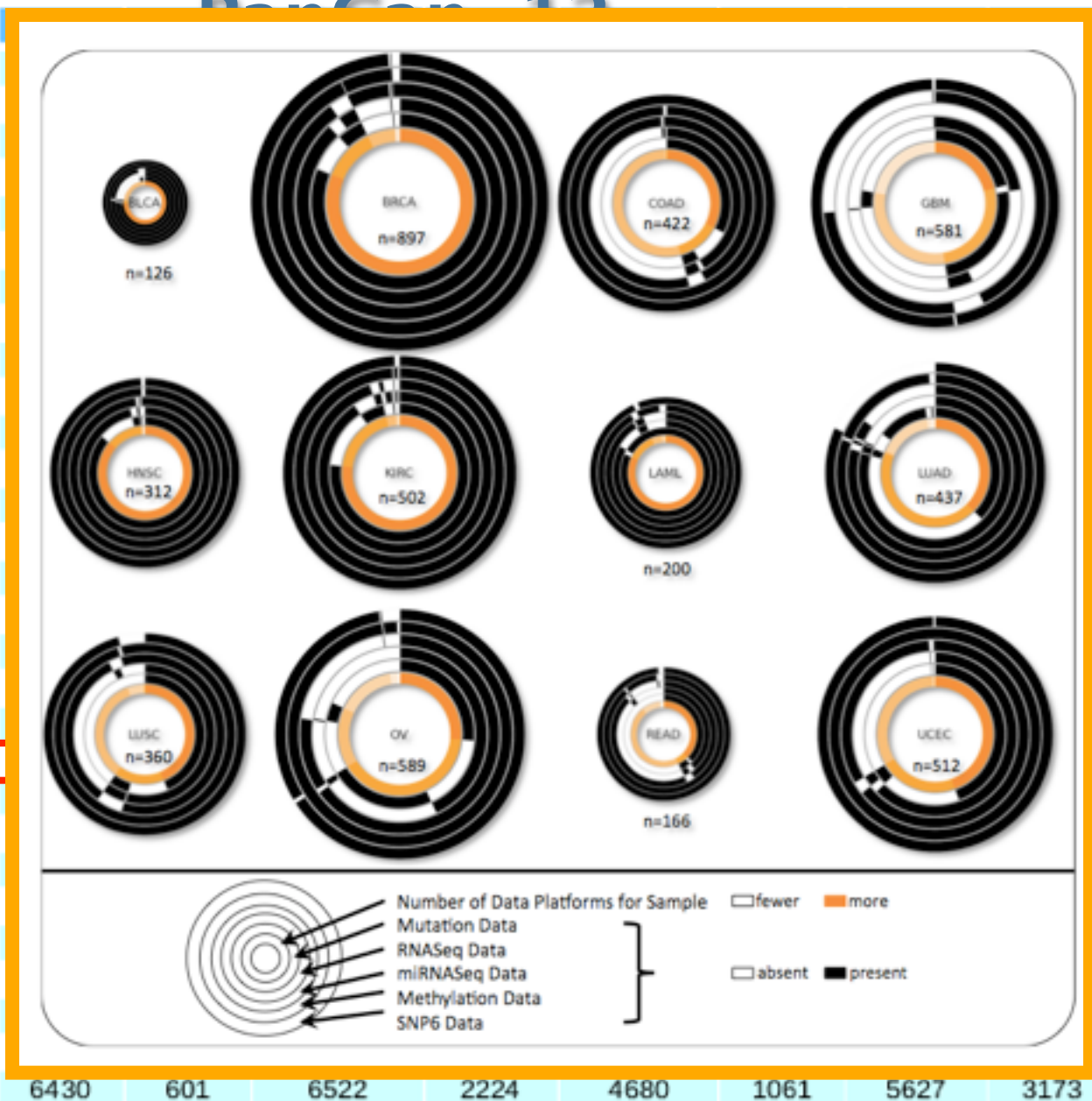
Tumor	BCR	Clinical	CN	Low Methylation	miRNA	miRNAseq	miR	miRseq	RPPA	MAF	
BLCA	153	108	137	100	138	0	96	0	124	54	28
BRCA	929	866	874	0	888	529	841	0	868	408	507
CESC	134	32	102	0	122	0	97	0	122	0	36
COAD	423	423	413	69	420	155	192	0	407	269	155
COADREAD	592	591	575	104	582	224	264	0	550	399	224
DLBC	28	0	18	0	17	0	0	0	16	0	0
ESCA	20	0	0	0	0	0	0	0	0	0	0
GBM	598	565	563	0	405	542	161	491	0	214	291
HNSC	343	315	306	96	310	0	303	0	309	212	306
KICH	66	0	65	0	65	0	65	0	0	0	0
KIRC	502	502	493	0	500	72	469	0	480	454	293
KIRP	149	103	103	0	103	16	63	0	103	0	100
LAML	202	200	0	0	194	0	179	0	187	0	199
LGG	222	208	180	0	176	27	110	0	180	0	34
LIHC	99	62	97	0	98	0	17	0	96	0	0
LUAD	508	333	356	0	430	32	353	0	365	237	229
LUSC	389	327	343	0	359	154	223	0	332	195	178
OV	592	580	566	0	584	574	297	570	454	412	316
PAAD	57	0	48	0	40	0	0	0	34	0	34
PANCAN12	5320	4838	4706	406	4890	2181	3303	1061	4154	2785	2819
PANCAN18	6851	5815	6071	601	6151	2224	4242	1061	5330	3173	3728
PRAD	180	148	171	0	172	0	140	0	170	0	83
READ	169	168	162	35	162	69	72	0	143	130	69
SARC	52	0	29	0	29	0	0	0	29	0	0
SKCM	288	138	273	101	253	0	247	0	240	164	253
STAD	257	162	237	0	145	0	43	0	134	0	116
THCA	435	218	401	94	412	0	379	0	349	224	323
UCEC	512	451	493	106	500	54	333	0	485	200	248
Totals	7307	5909	6430	601	6522	2224	4680	1061	5627	3173	3798

The PanCan12 DataSet (Level 3 data)

Tissue in PanCan12

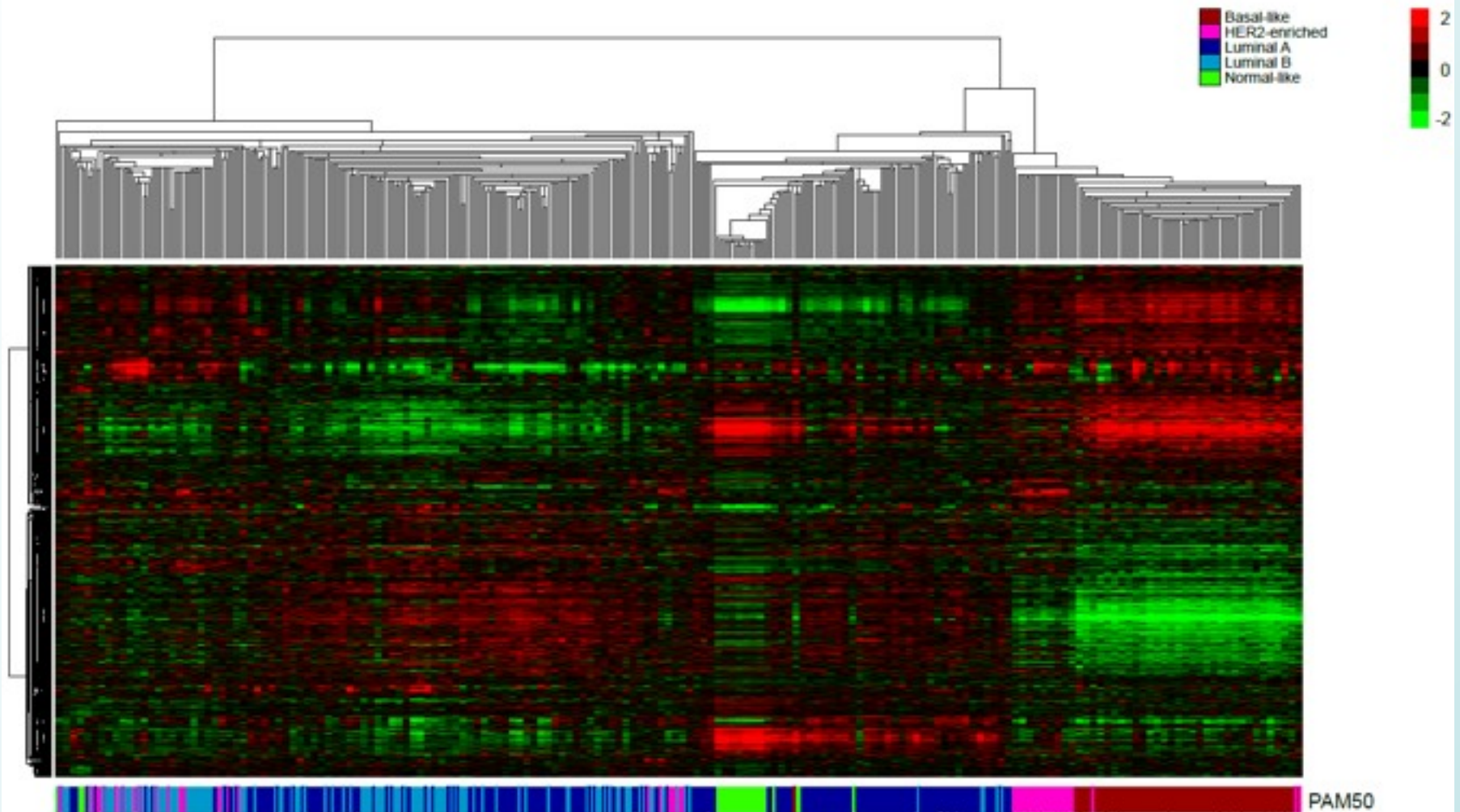
2012_12_21 stddata Run

Tumor	BCR	Clinical
BLCA	153	108
BRCA	929	866
CESC	134	32
COAD	423	423
COADREAD	592	591
DLBC	28	0
ESCA	20	0
GBM	598	565
HNSC	343	315
KICH	66	0
KIRC	502	502
KIRP	149	103
LAML	202	200
LGG	222	208
LIHC	99	62
LUAD	508	333
LUSC	389	327
OV	592	580
PAAD	57	0
PANCAN12	5320	4838
PANCAN18	6851	5815
PRAD	180	148
READ	169	168
SARC	52	0
SKCM	288	138
STAD	257	162
THCA	435	218
UCEC	512	451
Totals	7307	5909



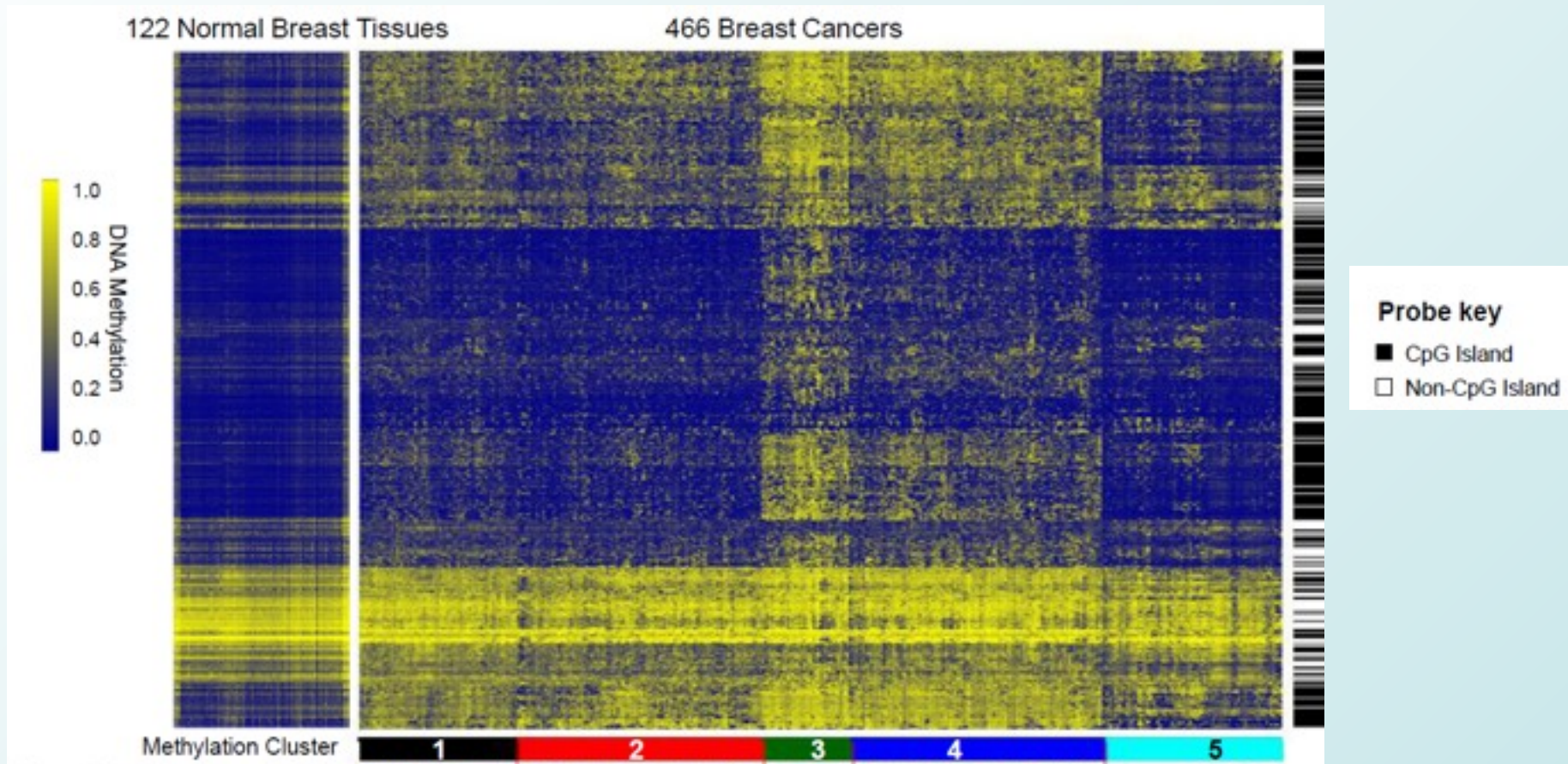
MAF
28
507
36
155
224
0
0
291
306
0
293
100
199
34
0
229
178
316
34
2819
3728
83
69
0
253
116
323
248
3798

BRCA mRNA Expression



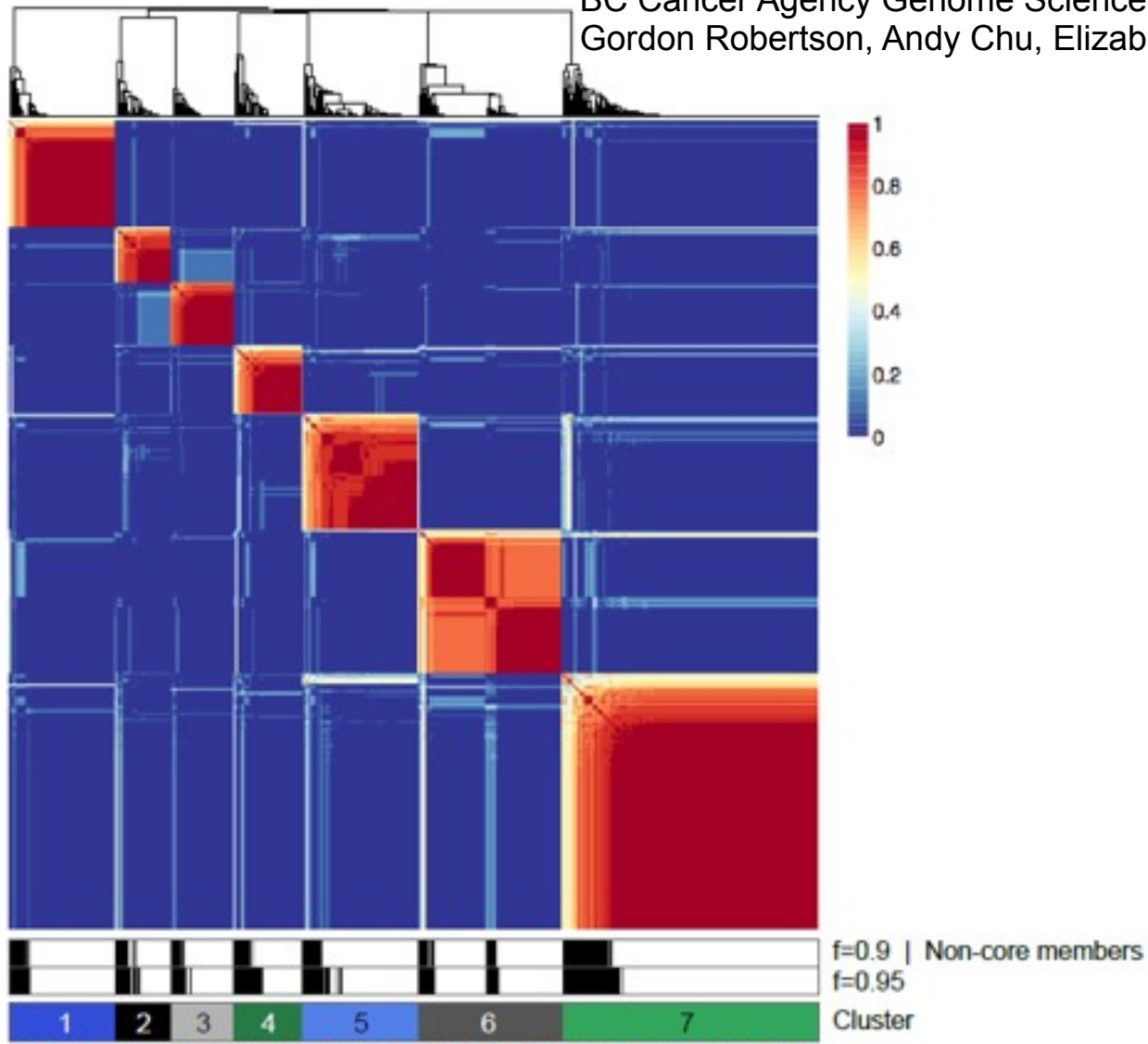
BRCA Methylation

USC: Swapna Mahurkar, Simeen Malik, Hui Shen, Dan Weisenberger, Peter Laird

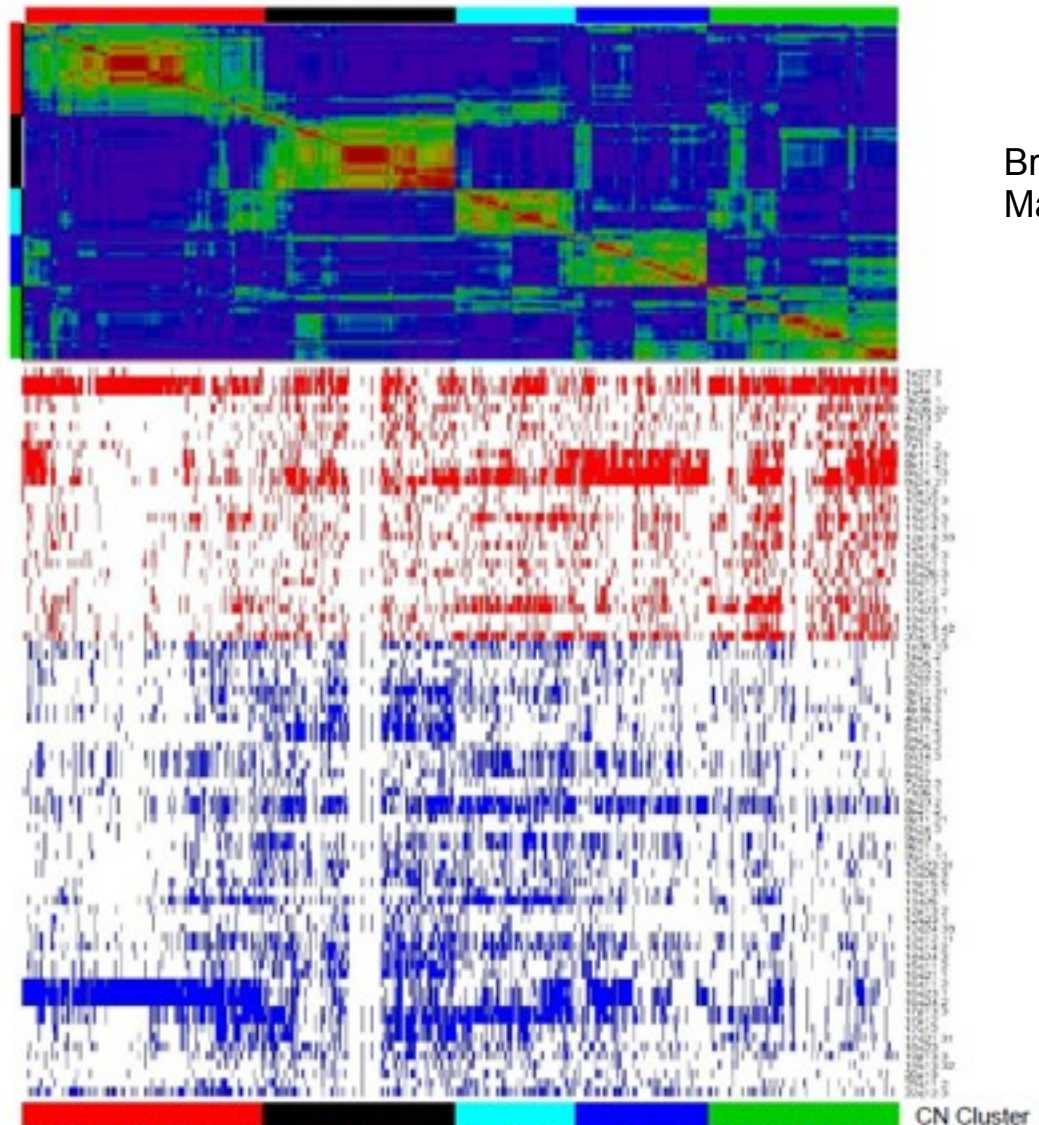


BRCA miRNA Expression

BC Cancer Agency Genome Sciences Centre:
Gordon Robertson, Andy Chu, Elizabeth Chun



BRCA Copy Number

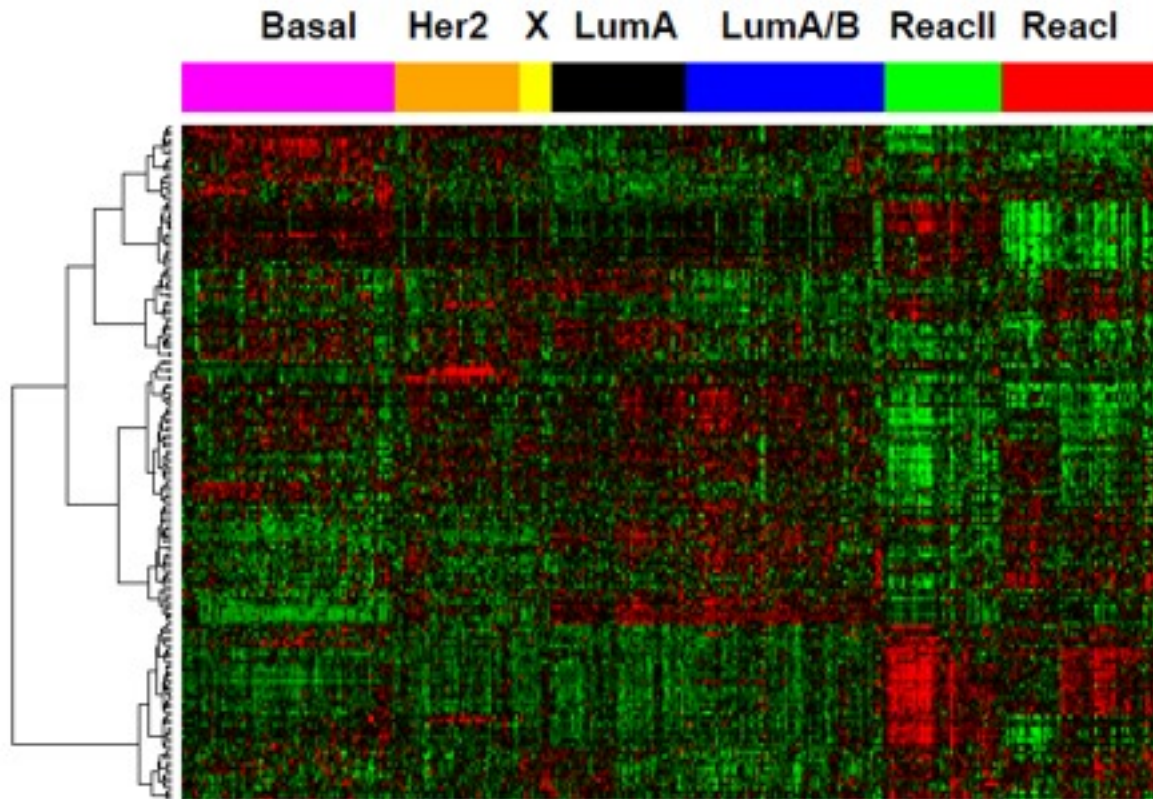


Broad Institute: Andrew Cherniack,
Matthew Meyerson



BRCA Protein - RPPA

MDACC: Nancy Shih, Roel Verhaak, Gordon Mills



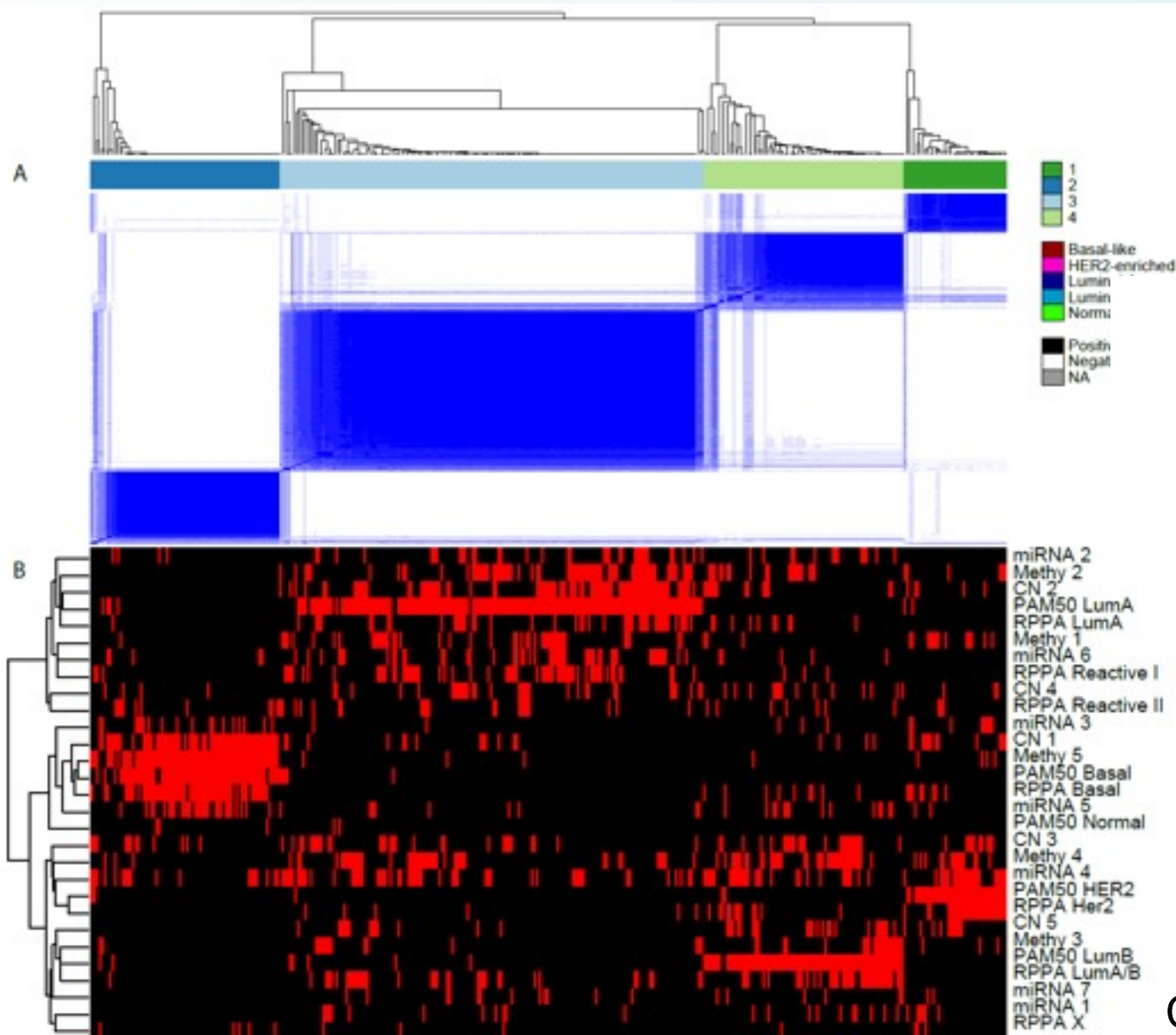
BRCA “Cluster of Clusters”

- Classifications from 5 different platforms
 - mRNA expression, miRNA expression, copy number, protein, and DNA methylation
- How to compare the different classifications?
- Turned each classification methods (5-7 subtypes each) into dummy variables and ran through consensus cluster

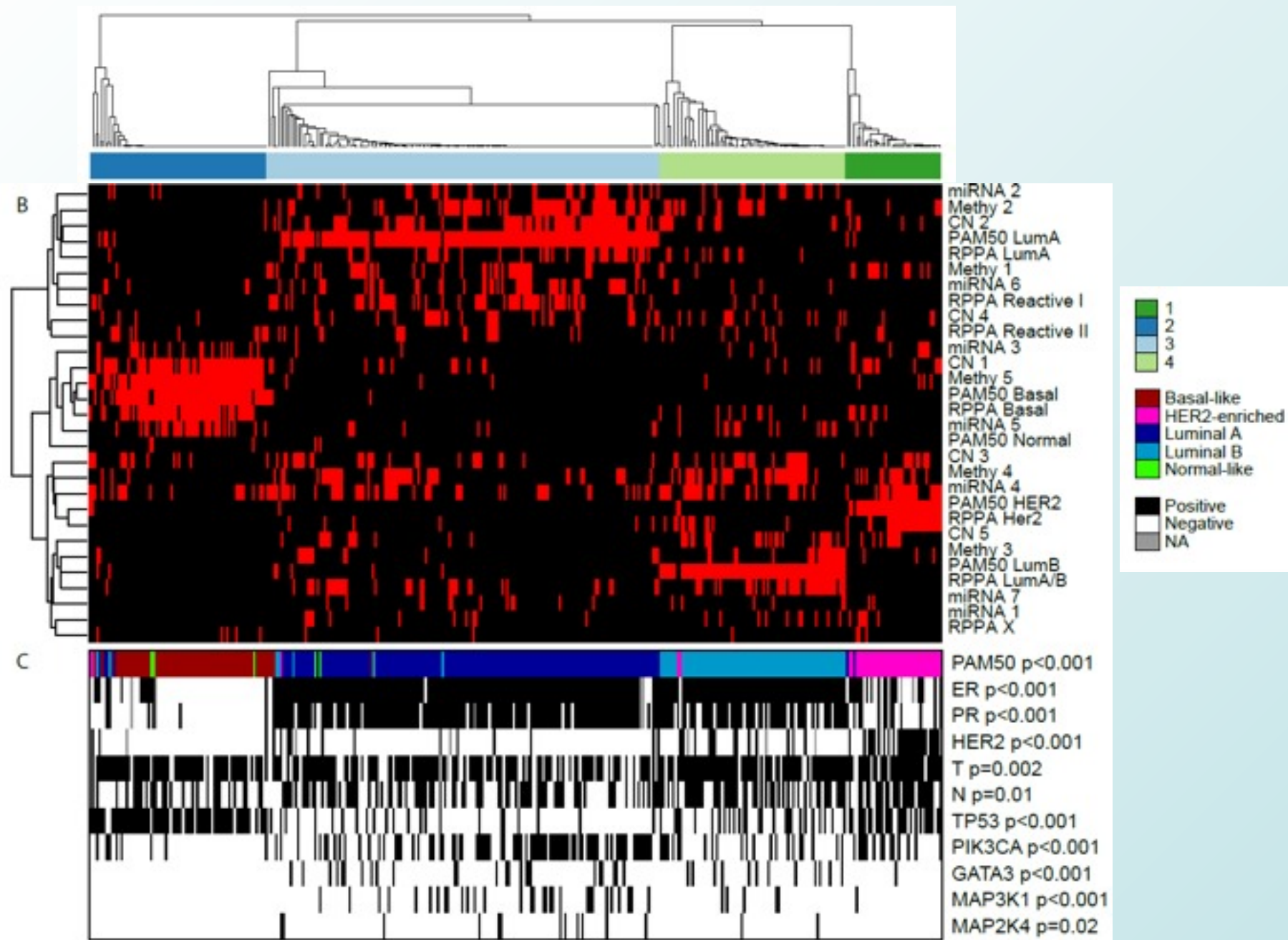
Katherine Hoadley, Chuck Perou (UNC)



BRCA “Cluster of Clusters”

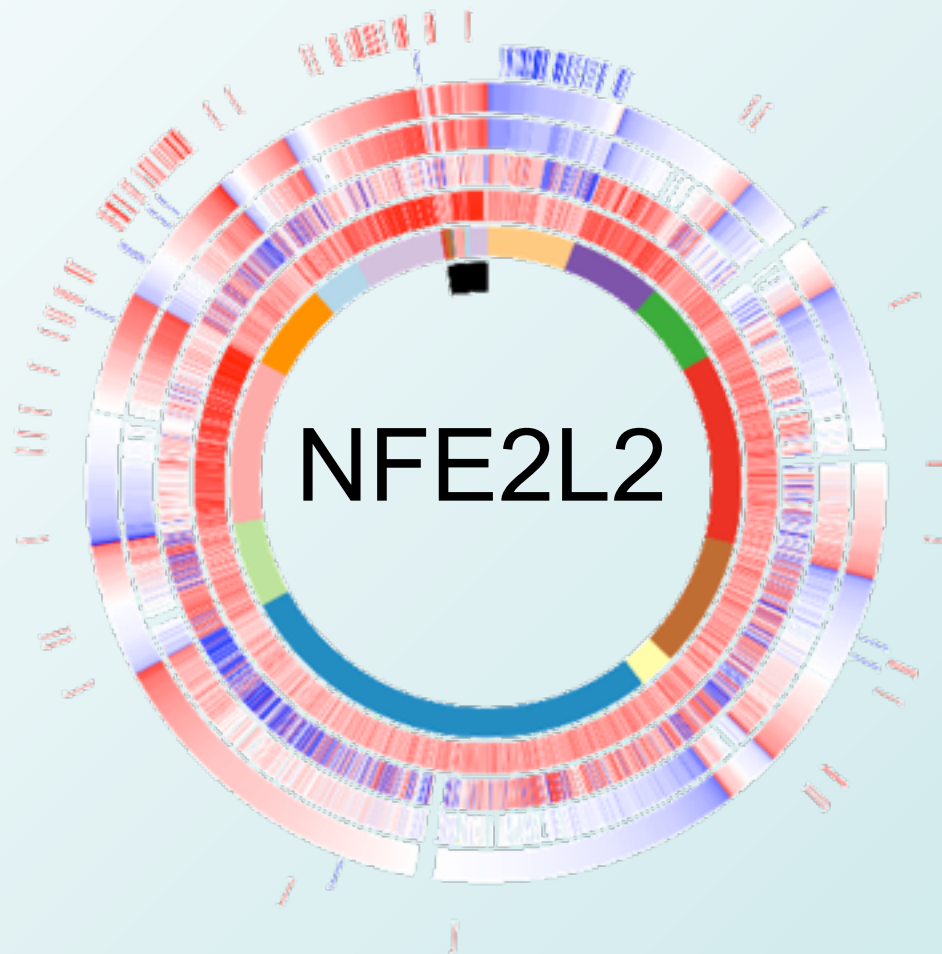


Katherine Hoadley,
Chuck Perou (UNC)



NFE2L2 GOF (Pan-Can)

Tissue	GSEA Score
LUSC	0.65582
BLCA	0.500254
LUAD	0.435044
HNSC	0.27281
COAD	0.215413
BRCA	0.213536
READ	0.164221
UCEC	-0.162436
KIRC	-0.218411
GBM	-0.36728
OV	-0.396528
LAML	-0.79203



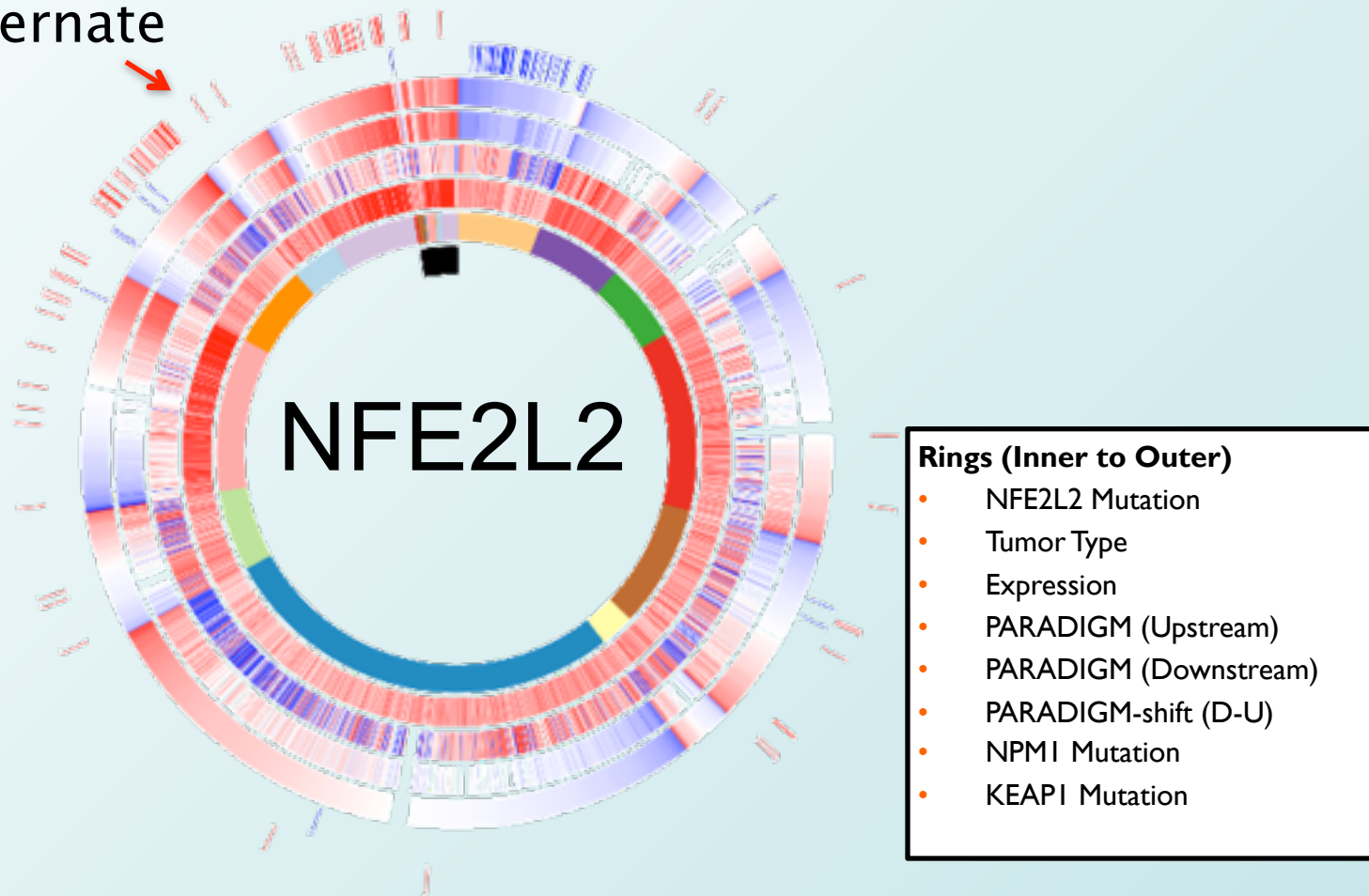
Rings (Inner to Outer)

- NFE2L2 Mutation
- Tumor Type
- Expression
- PARADIGM (Upstream)
- PARADIGM (Downstream)
- PARADIGM-shift (D-U)
- NPM1 Mutation
- KEAP1 Mutation

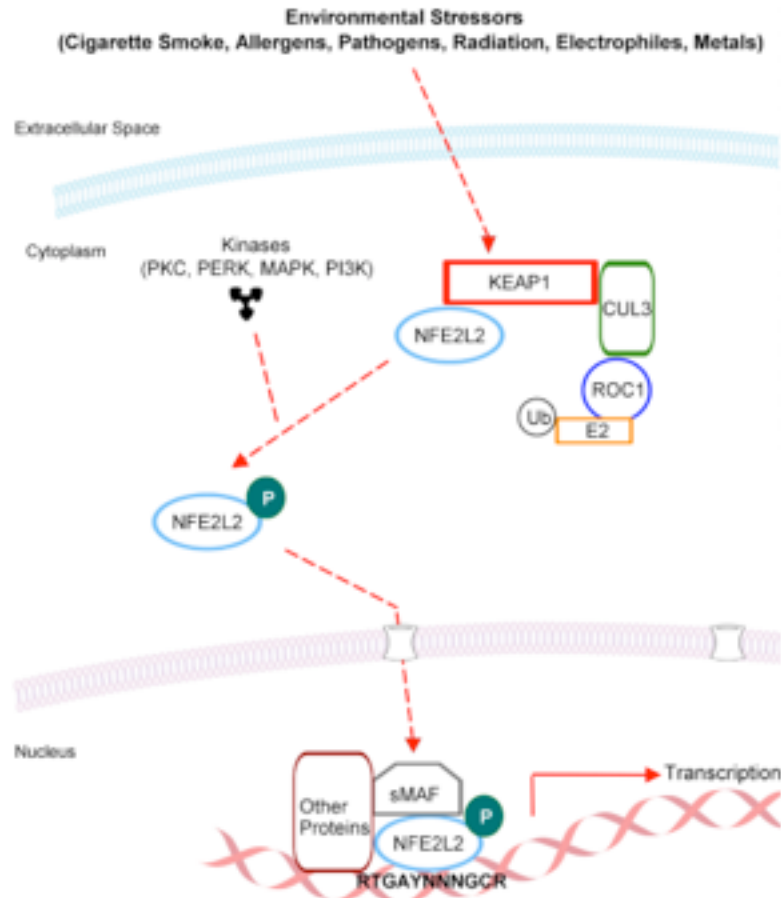


NFE2L2 GOF (Pan-Can)

KEAP1 Mutation is highly correlated with high PARADIGM-shift score and occurs across many Tumor Types and indicates an alternate mechanism of activating NFE2L2



NFE2L2 GOF (Pan-Can)



Gene	GSEA	Adjusted P-Value
NPM1	- 0.721524	< 0.02
KEAP1	0.70183	< 0.02
CUL3	0.615936	< 0.02
OR2C3	0.600244	< 0.02
GP2	0.586752	< 0.02
POTEE	0.57908	< 0.02
NLRP14	0.572907	< 0.02
OR2T4	0.569149	< 0.02
CCDC40	0.569149	< 0.02
C10orf71	0.564996	0.02



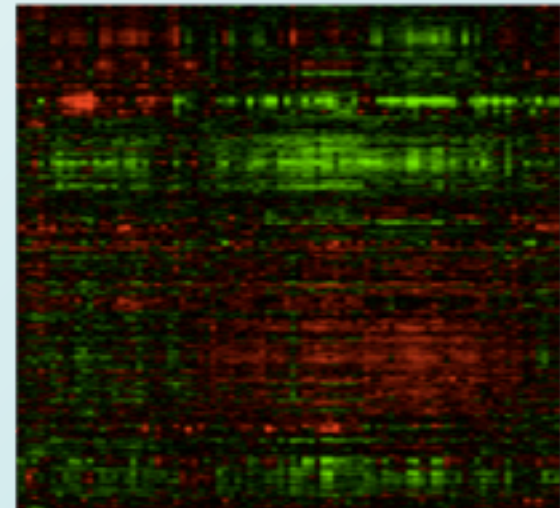
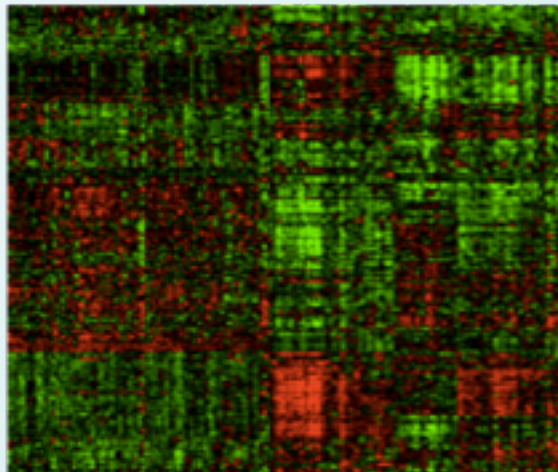
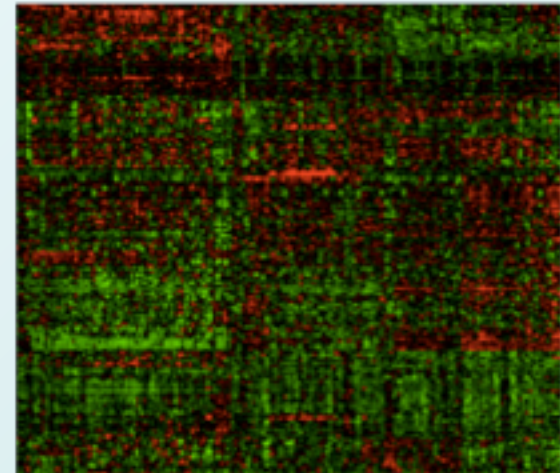
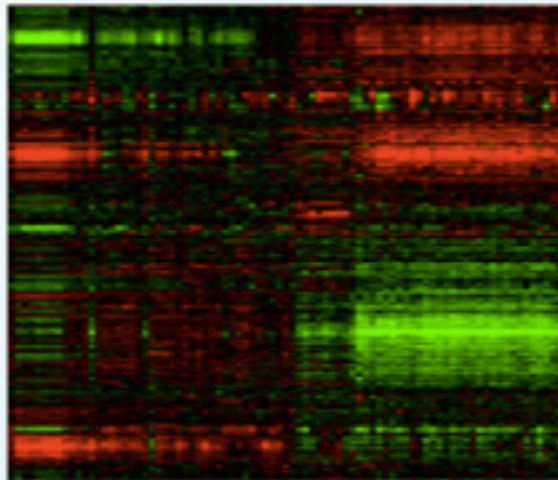
Heatmap-Matrices for HT data

- Most popular viewing modality in biology
- Matrix of genes by samples.
- Color by gene activity i in sample j .

 Cluster patterns

Heatmap-Matrices for HT Data

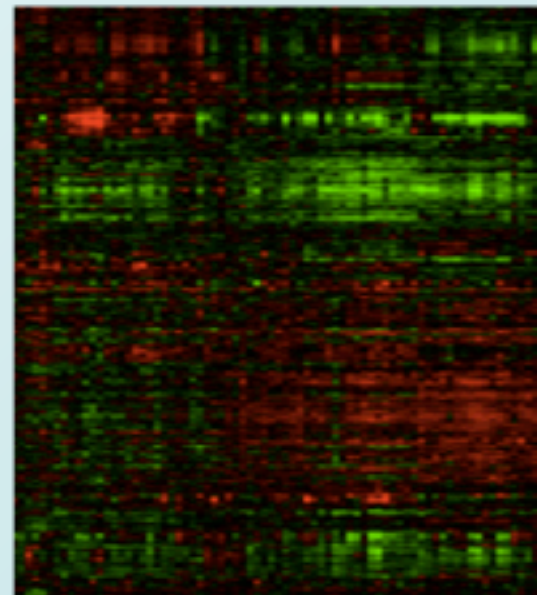
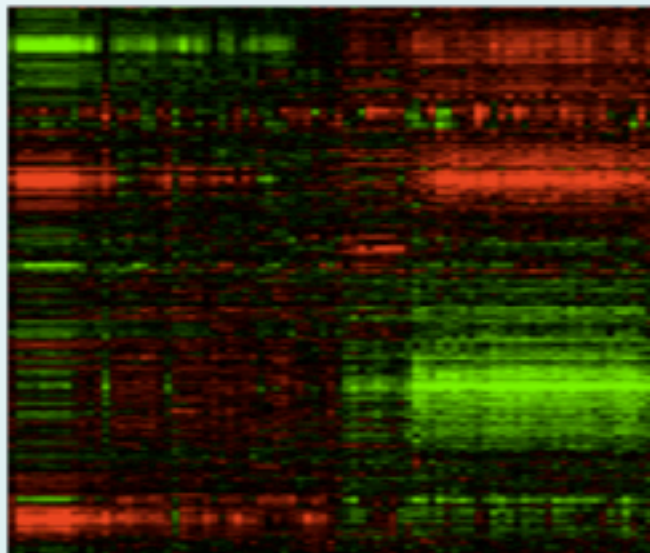
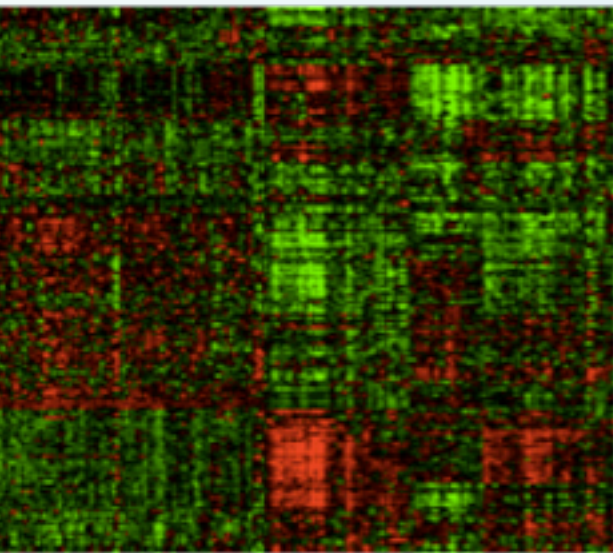
- Here are 4 more:



Quiz

Which Matrix is missing?

■ 1? 3? 4?



HeatMap-Matrices

- Great for making colorful t-shirts...



- eh, maybe not



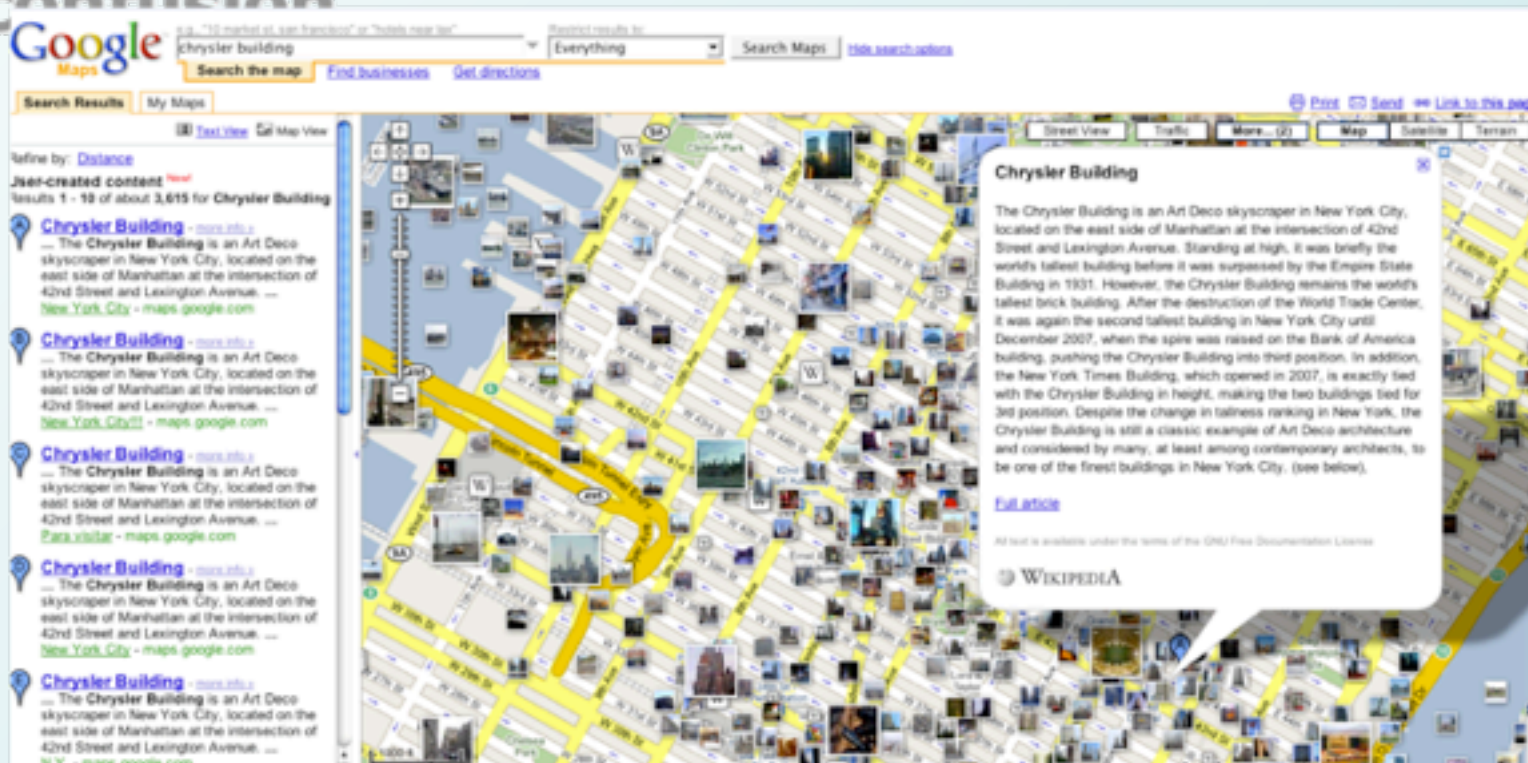
Matrices to 2D-Maps

- **Heatmaps Good for viewing activity patterns**
 - But relationships among samples obscured
 - Samples arranged in 1D
 - Space has no biological meaning
 - Equals 12-point font
- **Maps give global coordinate system to overlay information**



Google Map for Cell Signatures

- Define set coordinates
- Natural fit for humans
- Can overlay lots of information w/o confusion

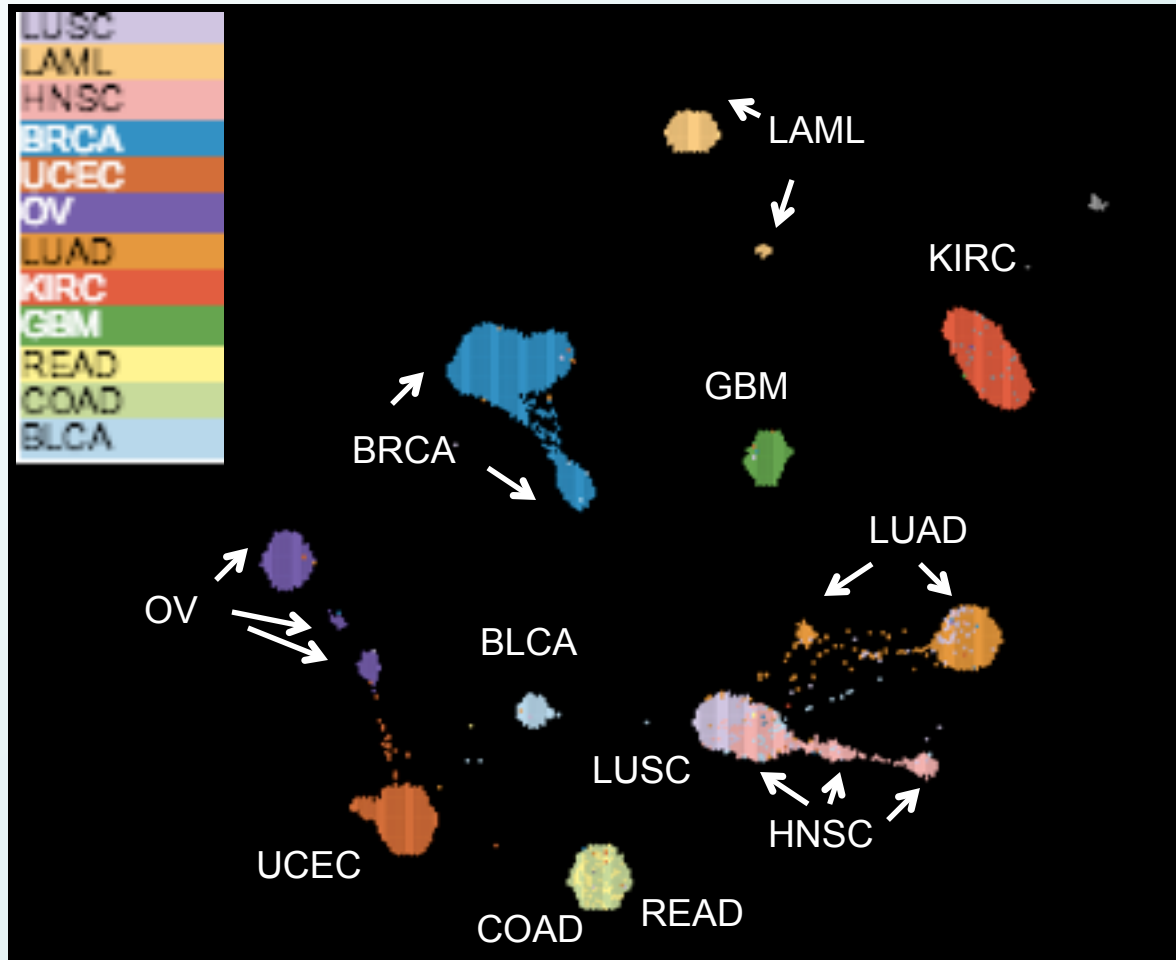


2D-Map for Tumor Similarities

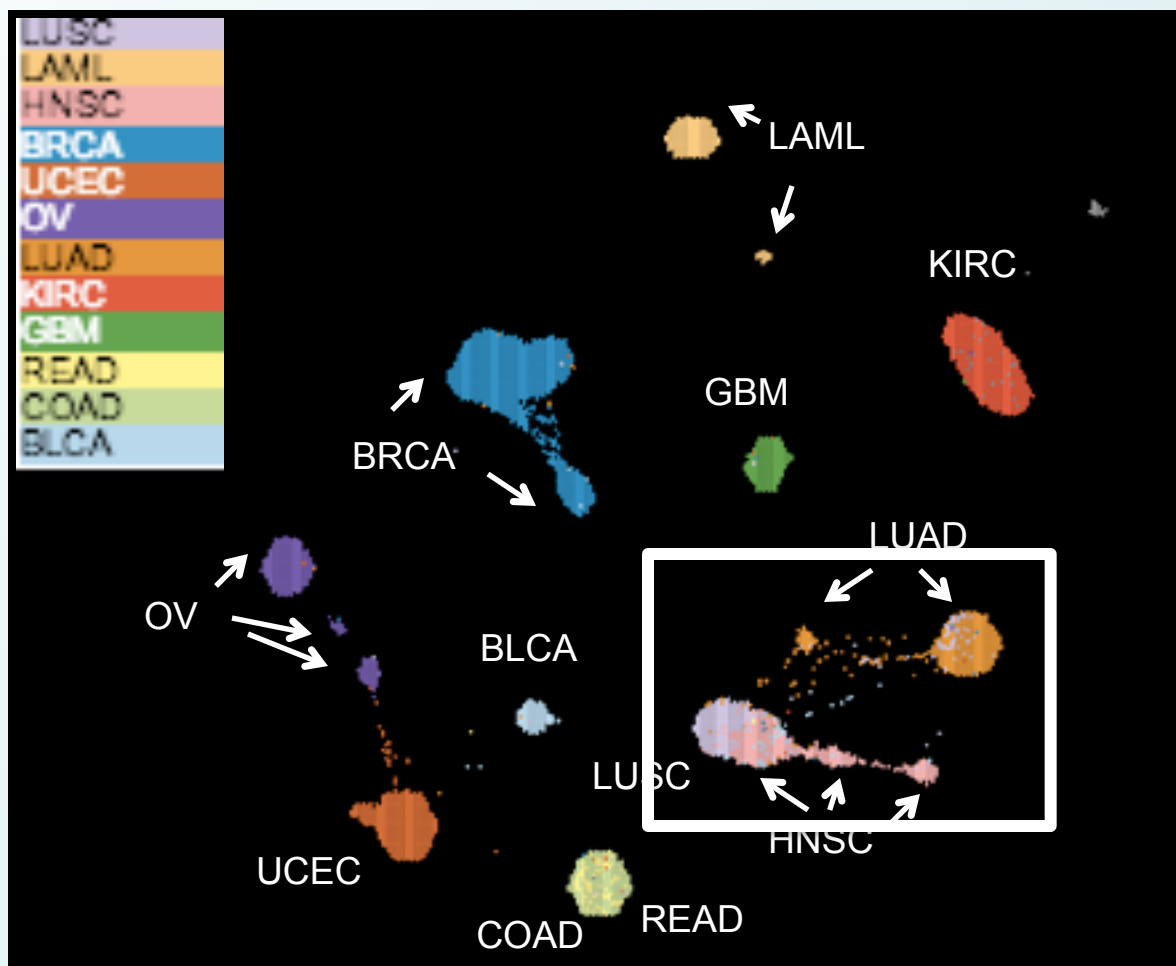
- **Point = Sample**
- **Samples connected by springs**
- **Springs strong for samples w/ molecular similarity**
- **Integrated Similarity Based on mRNA, miRNA, proteomics**
 - **Use Context Likelihood of Relationship Trick**



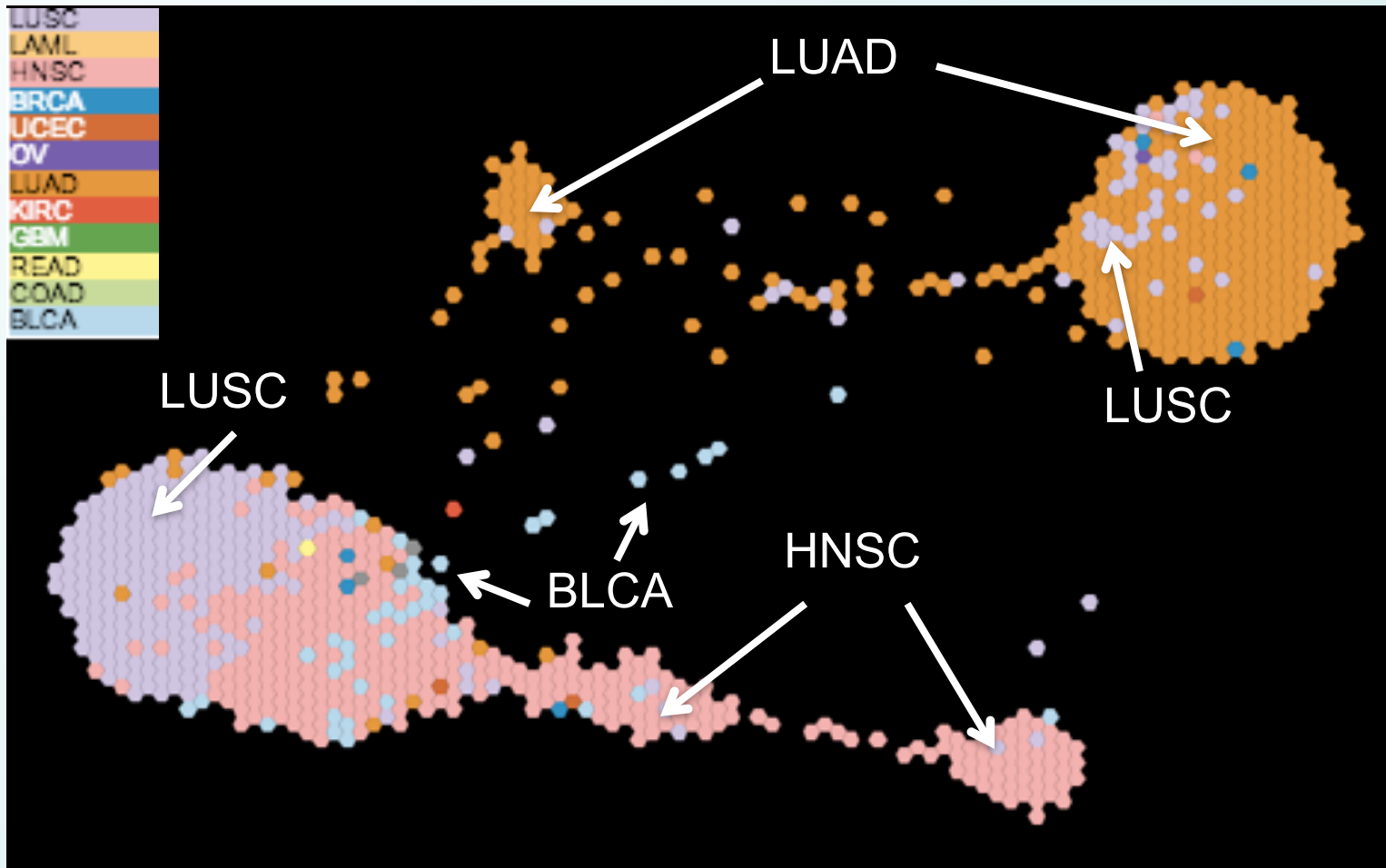
Integrated expression map of pancan12



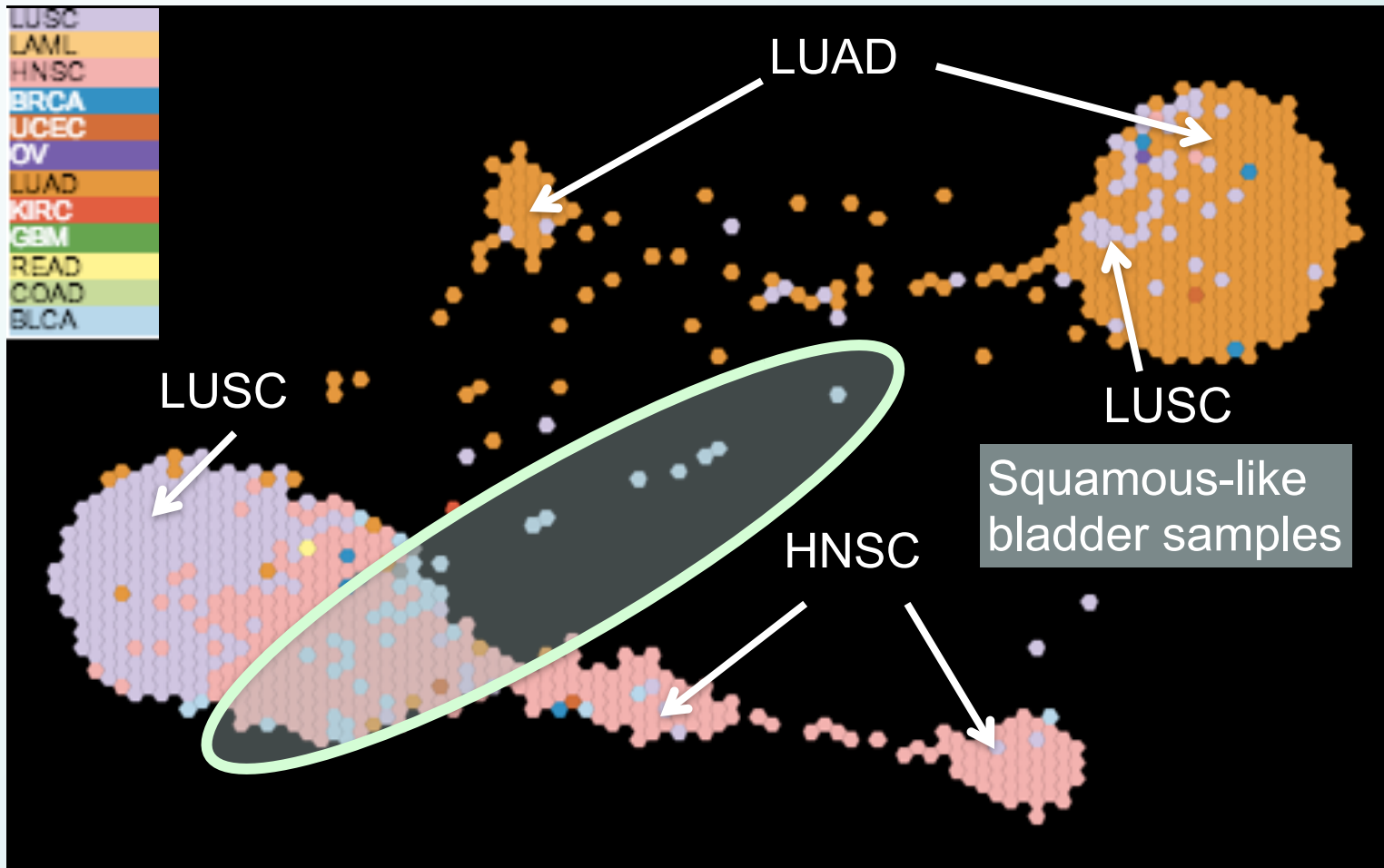
Integrated expression map of pancan12



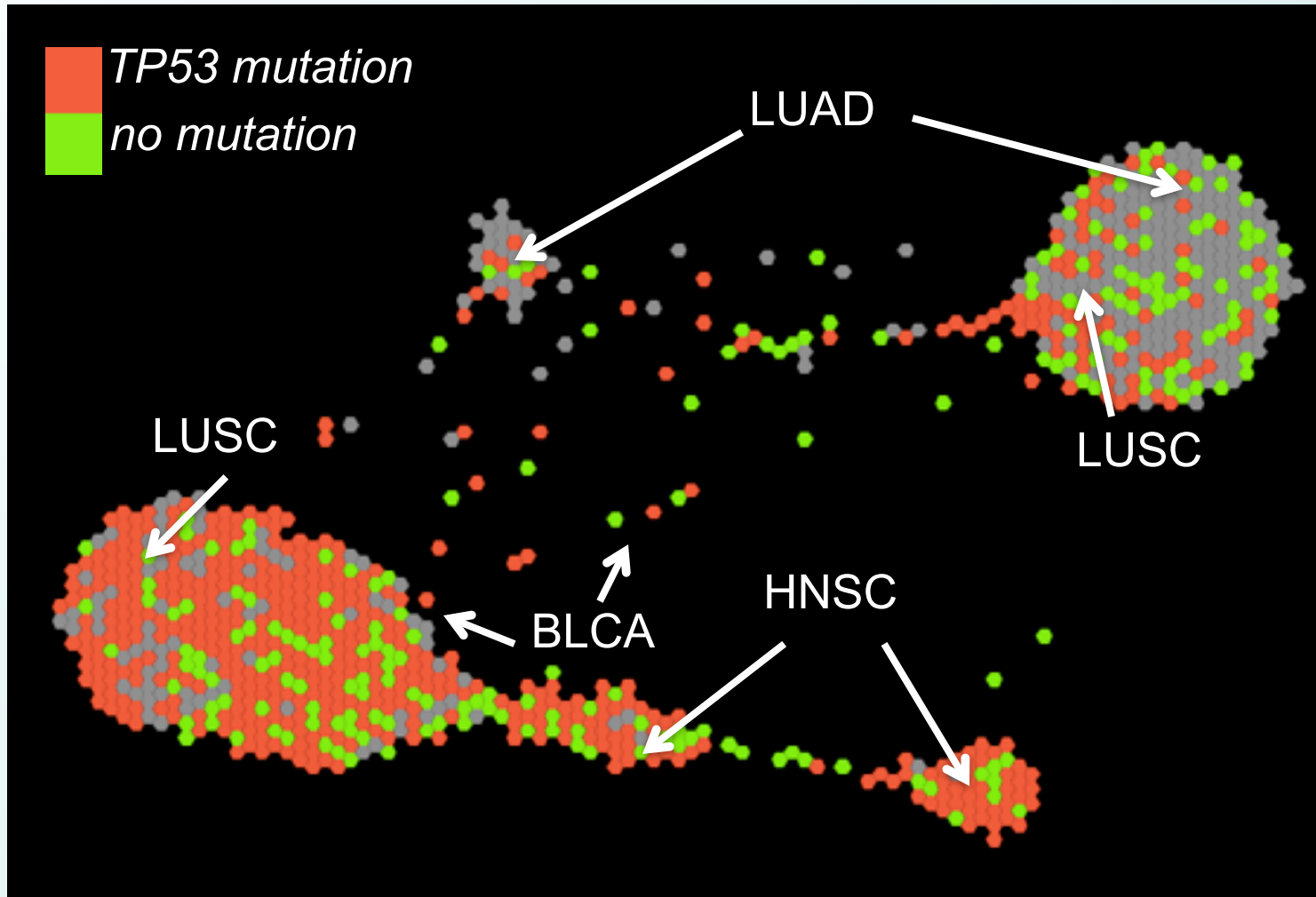
Squamous “island”



Squamous “island”

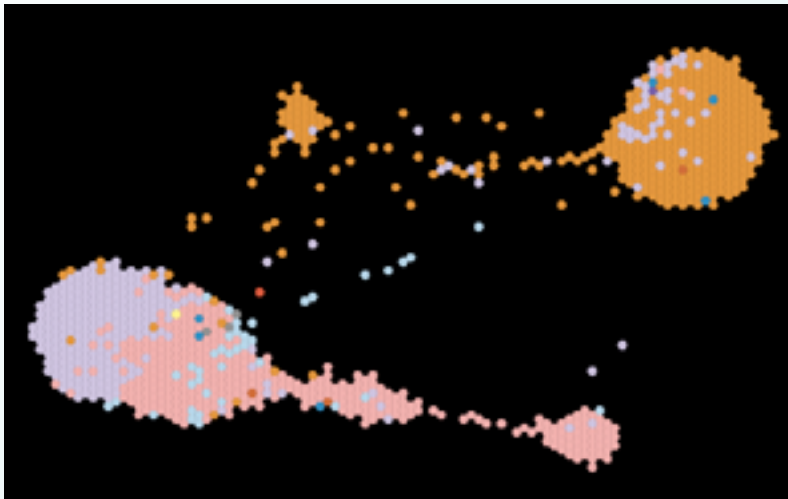


TP53 mutations on squamous “island”

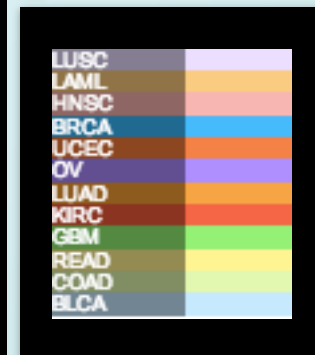
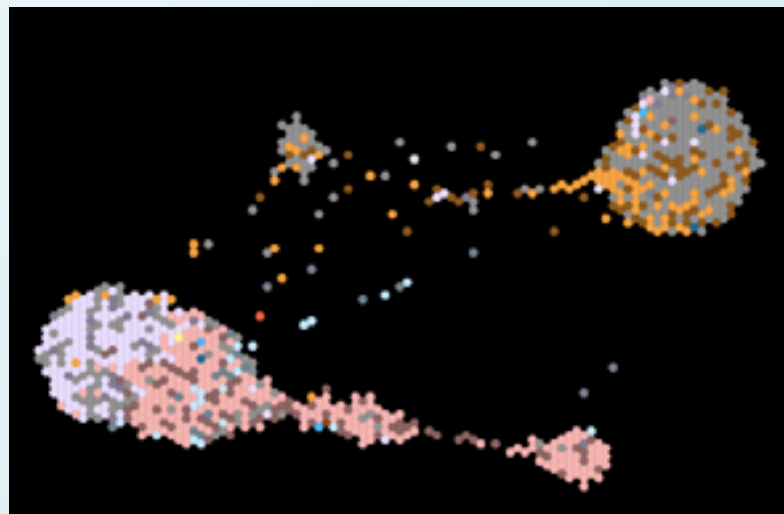
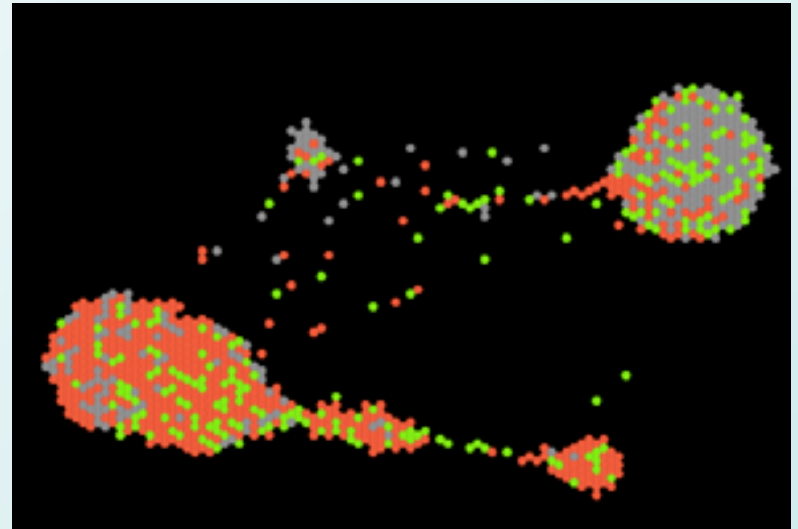


Overlay Mutations on top of tissue

tissue map

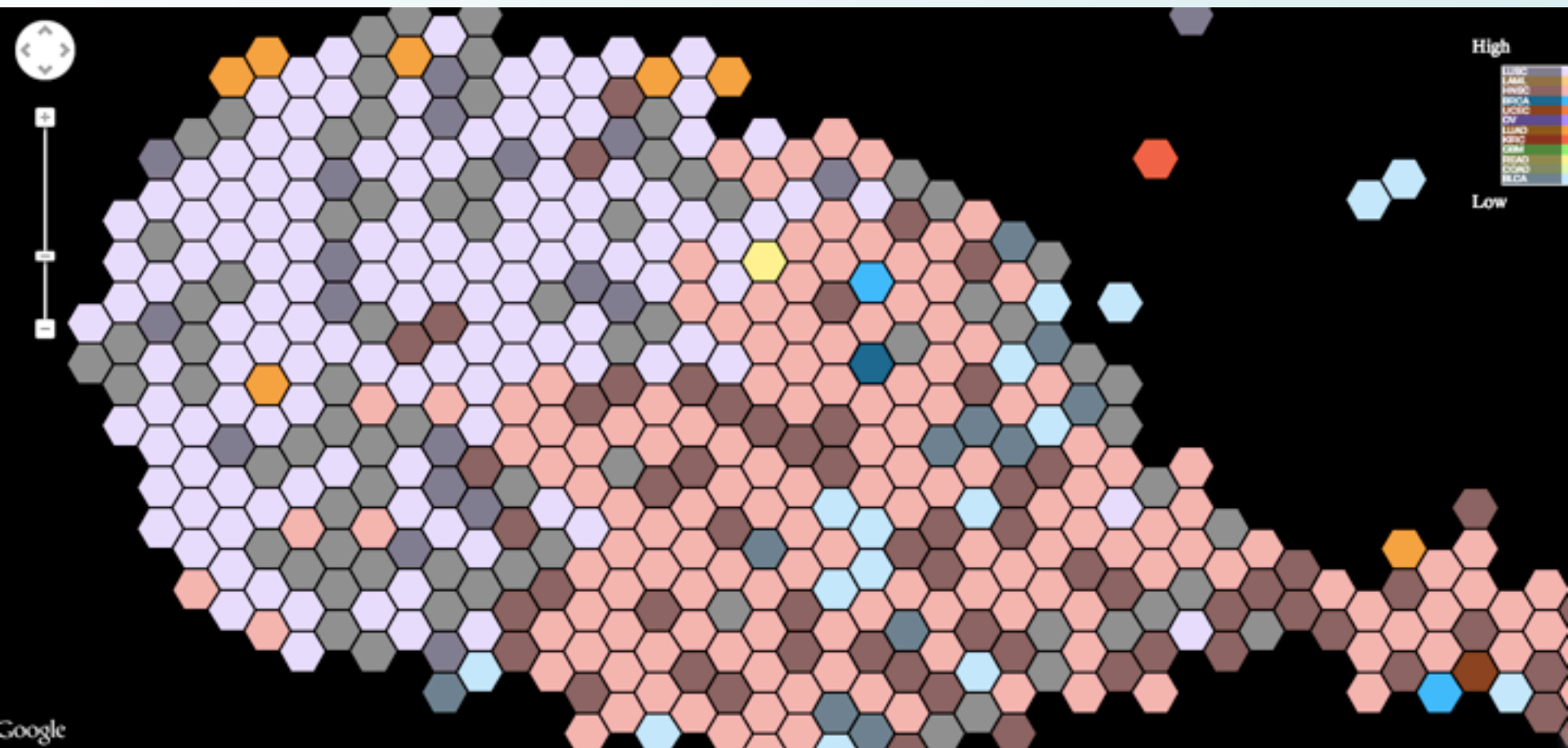


TP53 map



Overlay mutation on tissue

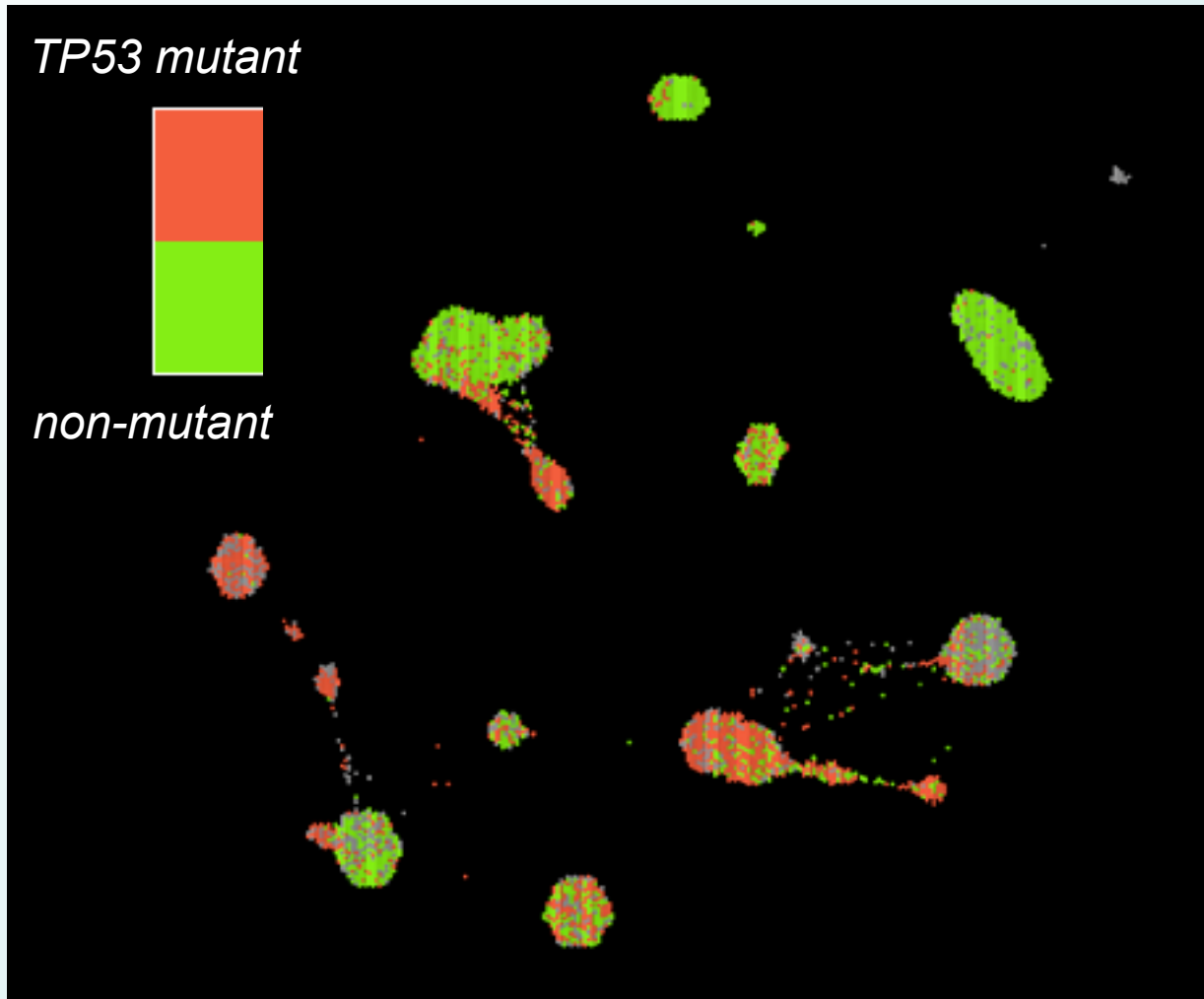
Tissue and TP53 Status



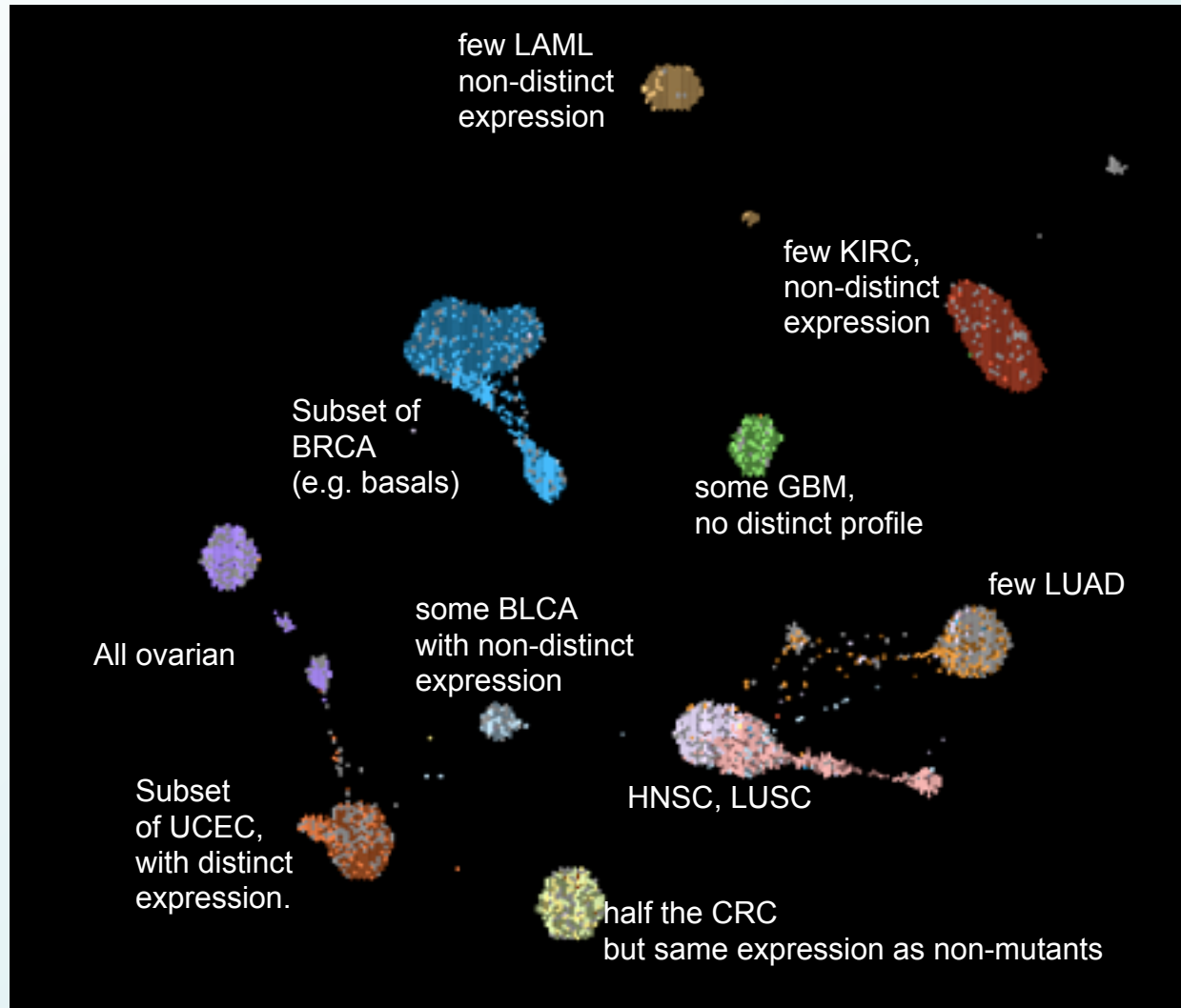
Google



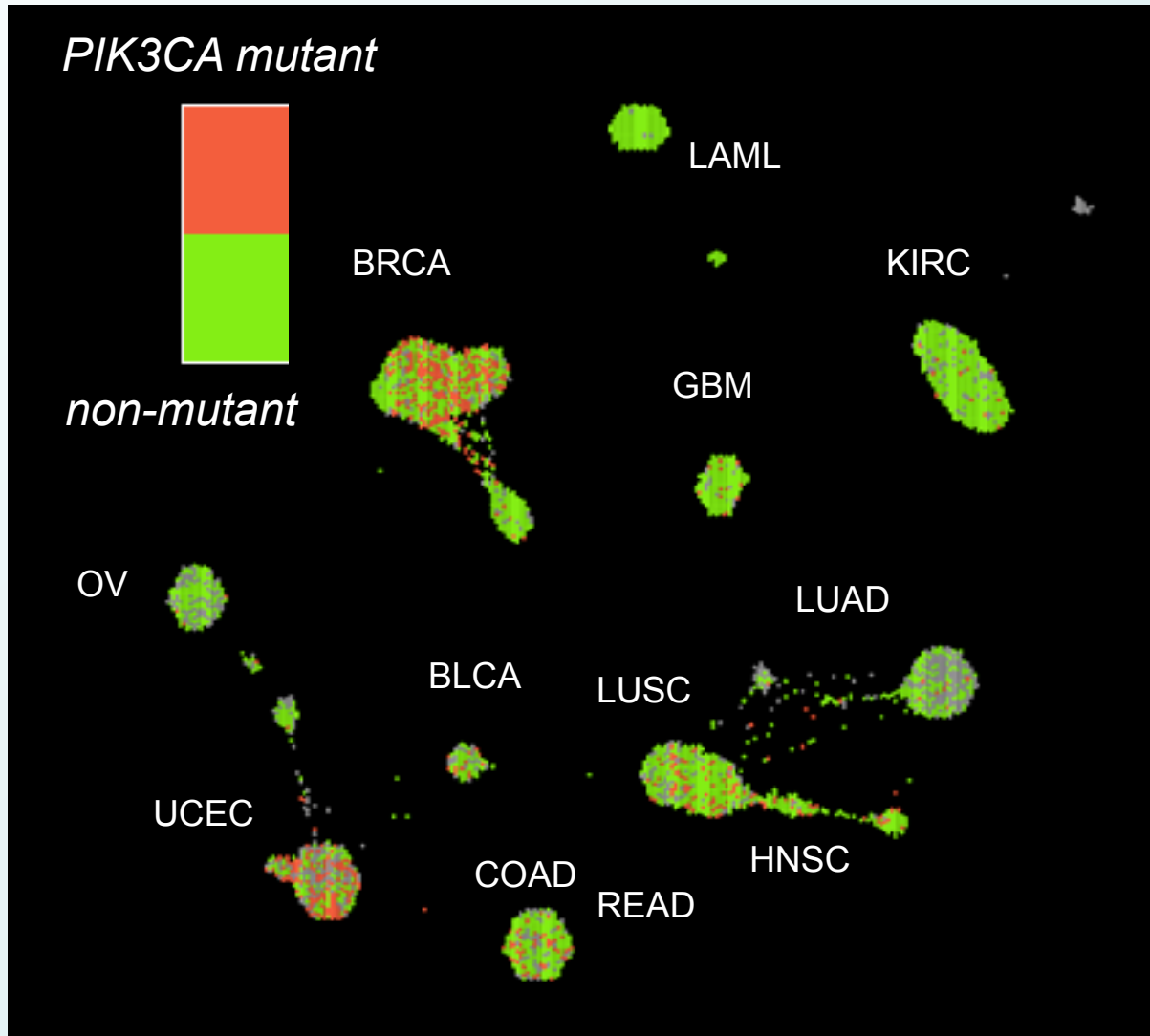
Full view of TP53 mutations



Overlay view of TP53 mutations



View of PIK3CA mutations



Summary



Summary

- Modeling information flow on known pathways gives view of gene activity.
- Patient stratification into pathway-based subtypes
- Sub-networks provide pathway-based signatures of sub-types and mutations.
- Loss- and gain-of-function predicted from pathway neighbors for even rare mutations.
- Identify interlinking genes associated with mutations to implicate additional targets even in LOF cases
 - E.g. Target MYC-related pathways in certain TP53-deficient cells?
- Current: Pan-cancer analysis to identify cell-of-origin or tissue independent drivers of cancer.
- Current: New browsers beyond heatmaps for cell signatures



UCSC Integrative Genomics Group

Marcos Woehrmann



Sam Ng



Dan Carlin



Evan Paull



James Durbin



Chris Szeto



Artem Sokolov



Chris Wong



CENTER FOR BIOMOLECULAR SCIENCE & ENGINEERING
promoting discovery and invention for human health and well-being





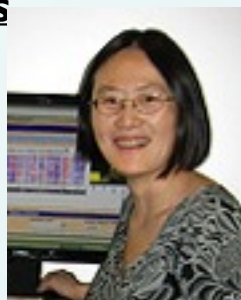
David Haussler

Acknowledgments

UCSC Cancer Genomics

- Kyle Ellrott
- Brian Craft
- Chris Wilks
- Amie Radenbaugh
- Mia Grifford
- Sofie Salama
- Steve Benz

Jing Zhu



UCSC Genome Browser Staff

- Mark Diekins
- Melissa Cline
- Jorge Garcia
- Erich Weiler

Buck Institute for Aging Chris Benz,

- Christina Yau
- Sean Mooney
- Janita Thusberg



Collaborators

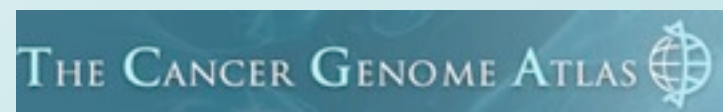
- Joe Gray, LBL
- Laura Heiser, LBL
- Eric Collisson, UCSF
- Nuria Lopez-Bigas, UPF
- Abel Gonzalez, UPF

Funding Agencies

- NCI/NIH
- SU2C
- NHGRI
- AACR
- UCSF Comprehensive Cancer Center
- QB3

Broad Institute

- Gaddy Getz
- Mike Noble
- Daniel DeCara



TCGA Pan-Can AWG



MSKCC

Chris Sander
Nikolaus Schultz
Giovanni Ciriello
Ethan Cerami
Marc Ladanyi
Anders Jacobsen



UNC

Katherine Hoadley
Chuck Perou



JSC

Peter Laird
Hui Shen

UCSC



David Haussler
Jing Zhu
Ted Goldstein
Sam Ng
Evan Paull
Chris Szeto

Broad



Matt Meyerson
Gaddy Getz
Rameen Beroukhim
Andy Cherniak
Scott Carter

MDACC

Roel Verhaak
Gordon Mills
John Weinstein
Rehan Akbani
Nancy Shih
Diana Hubbard
Hoon Kim
Matti Nykter



NCI



Kenna Shaw
Greg Eley
Heidi Sofia



UBC

Gordon Robertson
Andy Chu
Anders Jacobsen
Preethi Gunaratne,
Weimin Xiao



ISB

Ilya Shmulevich
Sheila Reynolds

TCGA Data on UCSC Cancer Browser

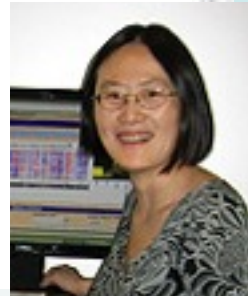
The screenshot shows the UCSC Cancer Browser interface. On the left, a sidebar lists various cancer types and data sources. A red circle highlights the TCGA data source. The main panel displays three tracks for Breast Cell Line data:

- Breast Cell Line Exon Expression (Gray Lab) • N=55**: A heatmap showing gene expression levels across various cell lines.
- Breast Cell Line Copy Number Segmented (Gray Lab) • N=46**: A heatmap showing copy number variations across various cell lines.
- Breast Cell Line DNA Methylation (Gray Lab) • N=50**: A heatmap showing DNA methylation levels across various cell lines.

The tracks are color-coded: red for high expression, blue for low expression, and yellow for copy number variations. The DNA methylation track shows black and white bars representing methylation levels.

TCGA data

- Regular automatic uploads
 - Data from DCC
 - Results from Firehose



Jing Zhu

