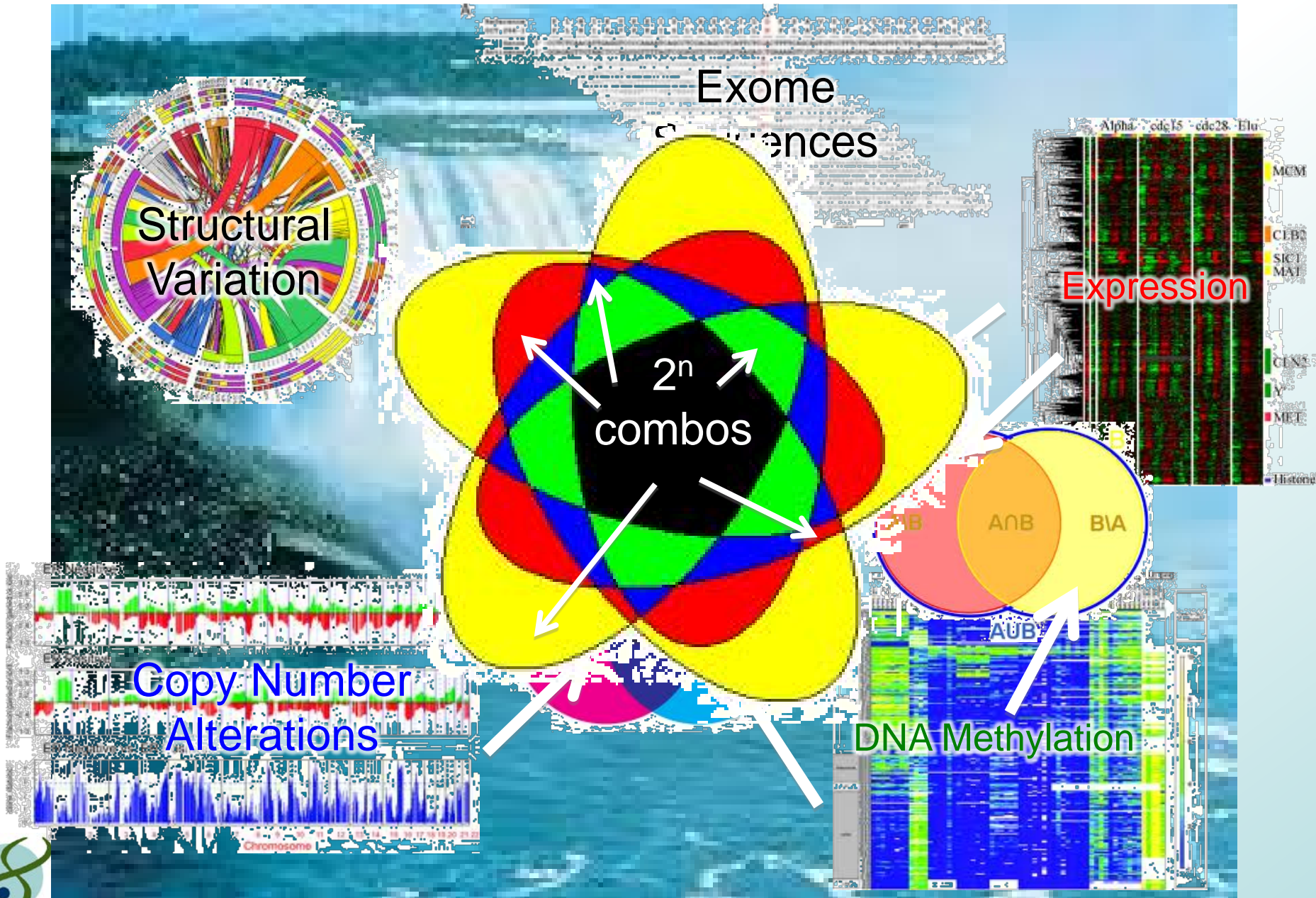


Patient-specific pathway analysis using PARADIGM identifies key activities in multiple cancers

**Josh Stuart, UC Santa Cruz
TCGA Symposium
National Harbor, Nov 18, 2011**



Flood of Data Analysis Challenges



Flood of Data Analysis Challenges

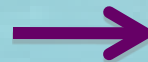


**Multiple, Possibly
Conflicting Signals**

Expression

CLB
SIC
MAT
CLN
Y
MET
Histones

**This is What it
Does to You**

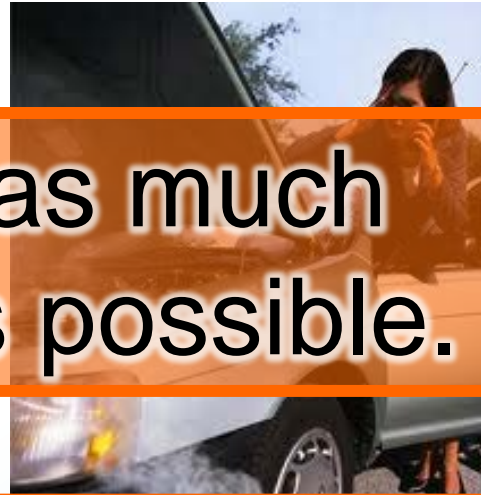


Analysis of disease samples like automotive repair (or detective work or other sleuthing)

Patient Sample 1



Patient Sample 2



Sleuths use as much knowledge as possible.

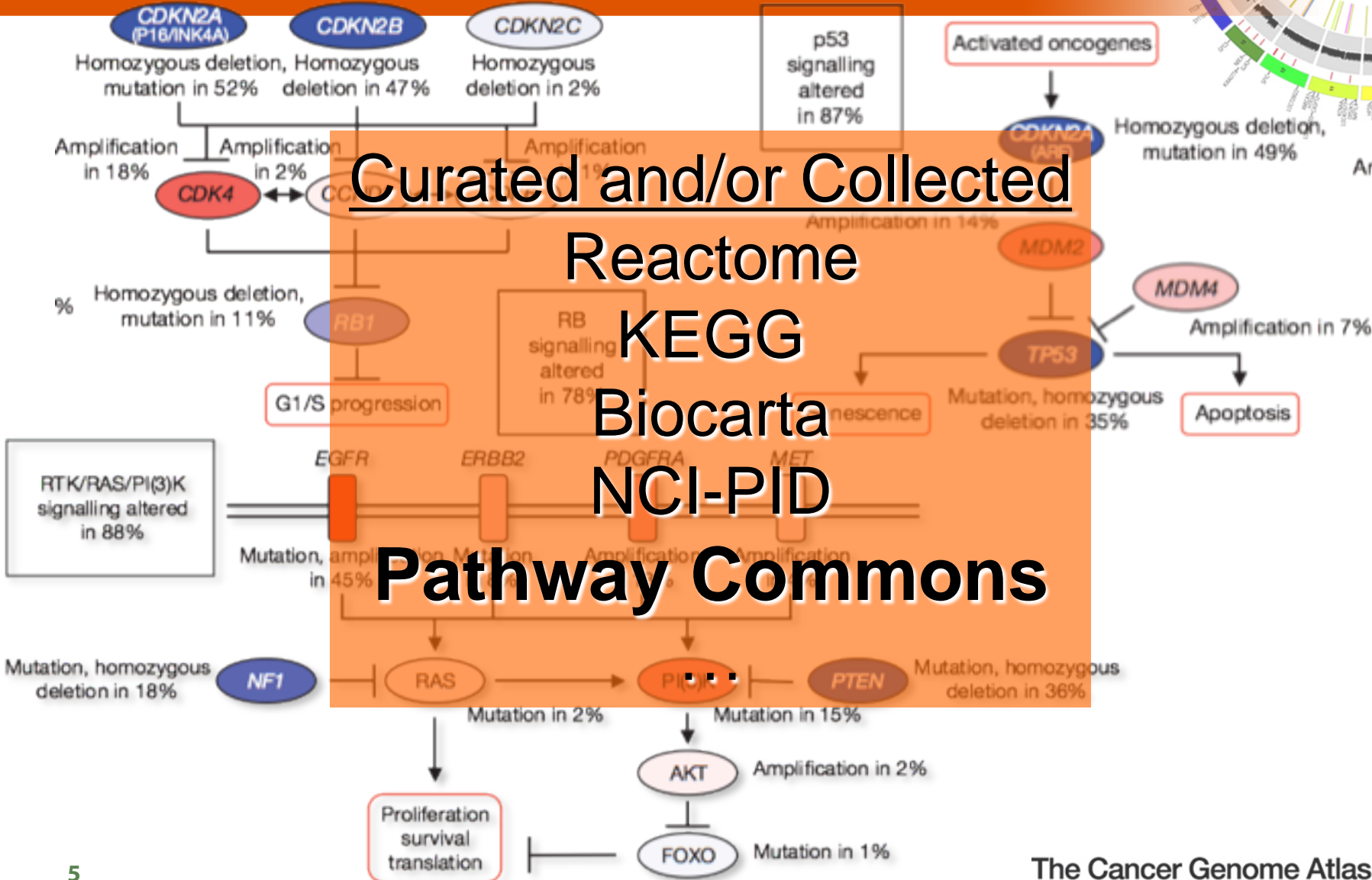
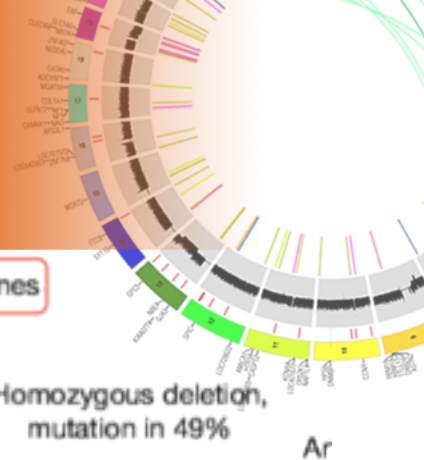
Patient Sample 3



Patient Sample N



Much Cell Machinery Known: Gene circuitry now available.

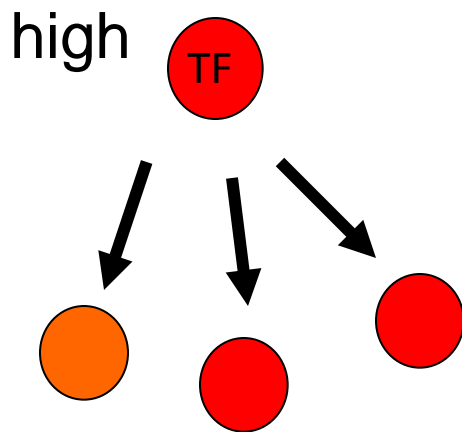


Integration key to correct interpretation of gene function

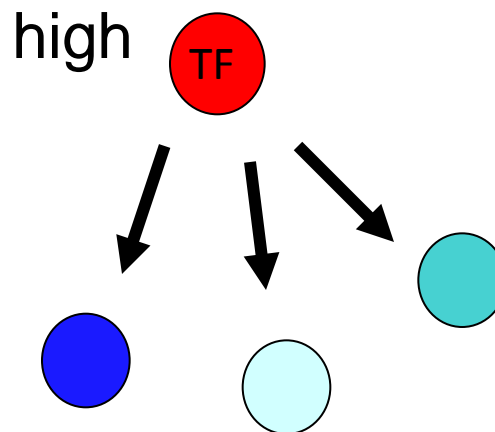
- Expression not always an indicator of activity
- Downstream effects often provide clues



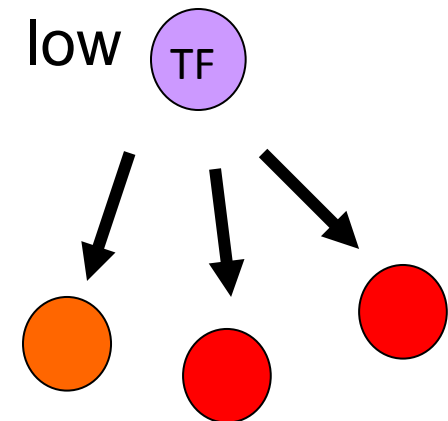
Expression of 3 transcription factors:



Inference:
TF is **ON**
(expression reflects activity)



Inference:
TF is **OFF**
(high expression but inactive)



Inference:
TF is **ON**
(low-expression but active)

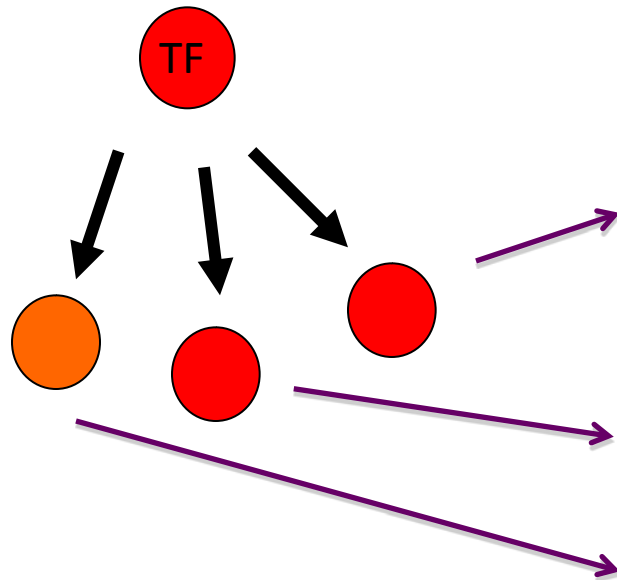
Integration key to correct interpretation of gene function



- Need multiple data modalities to get it right.

BUT, targets are amplified

Expression -> TF **ON**



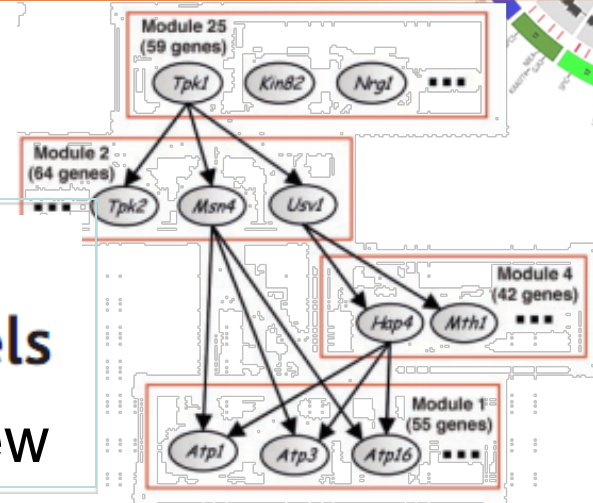
Copy Number -> TF **OFF**

Lowers our belief in active TF because **explained away** by *cis* evidence.

Probabilistic Graphical Models: A Language for Integrative Genomics

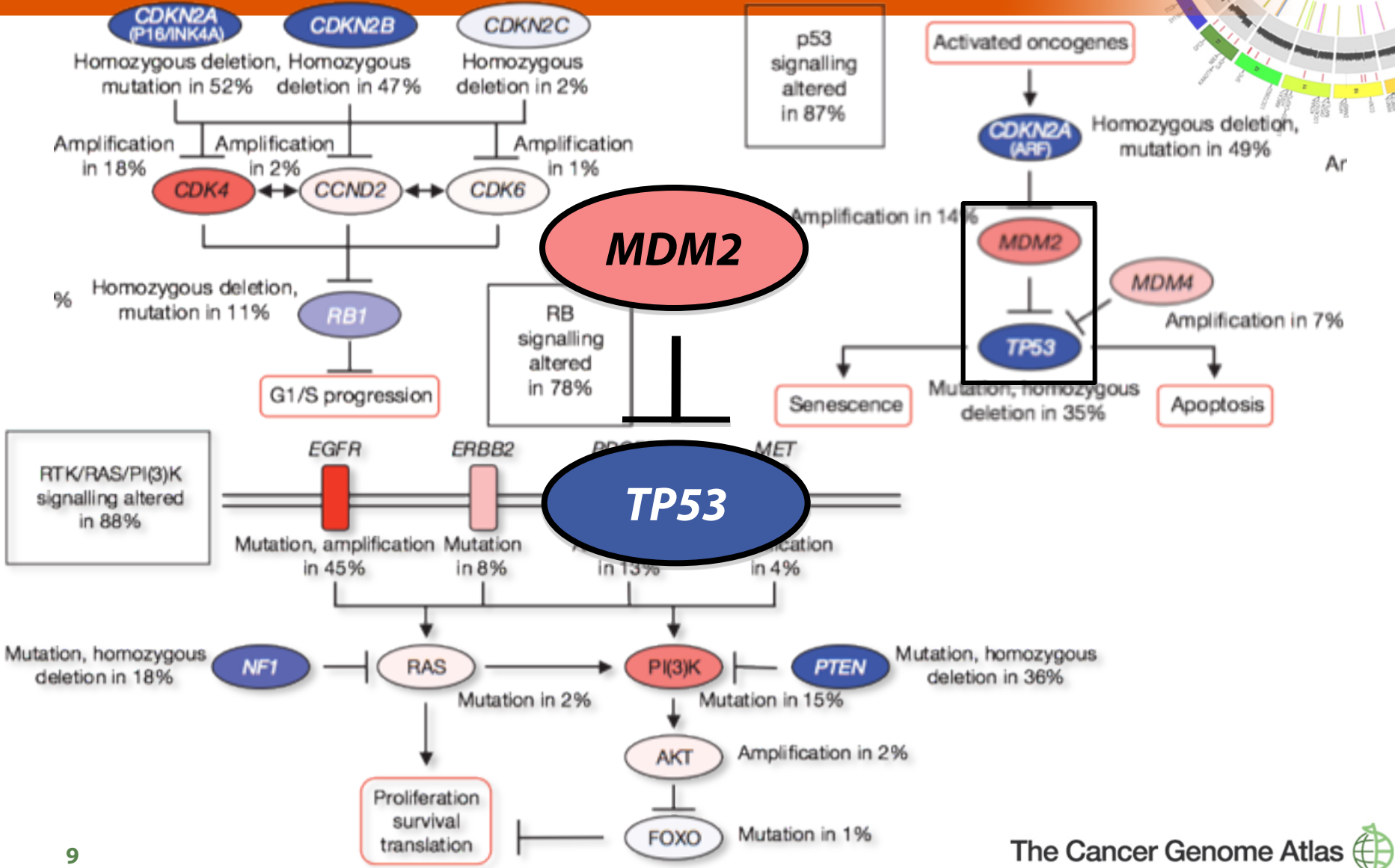
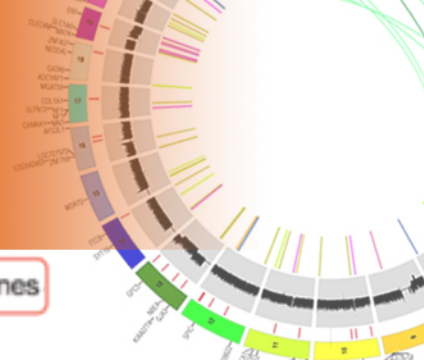
Inferring Cellular Networks Using Probabilistic Graphical Models

Nir Friedman, *Science* (2004) - Review



- Generalize HMMs, Kalman Filters, Regression, Boolean Nets, etc.
- Language of probability ties together multiple aspects of gene function & regulation
- Enable data-driven discovery of biological mechanisms
- Seminal work: J. Pearl, D. Heckerman, E. Horvitz, G. Cooper, R. Schacter, D. Koller, N. Friedman, M. Jordan, ...
- Recent work: E. Segal, E Schadt, A. Hartemink, D. Pe'er, ...

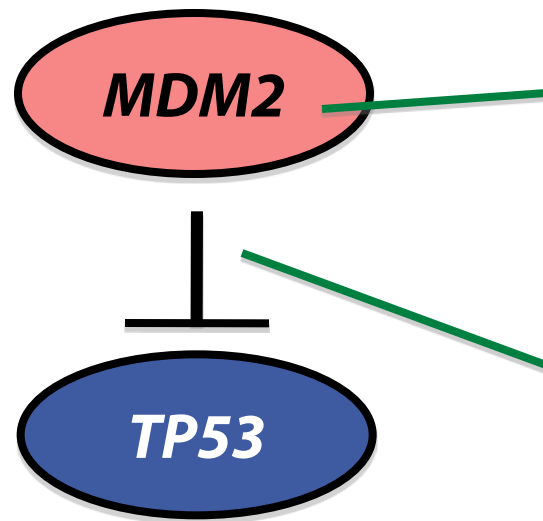
Integration Approach: Detailed models of gene expression and interaction



Integration Approach: Detailed models of expression and interaction

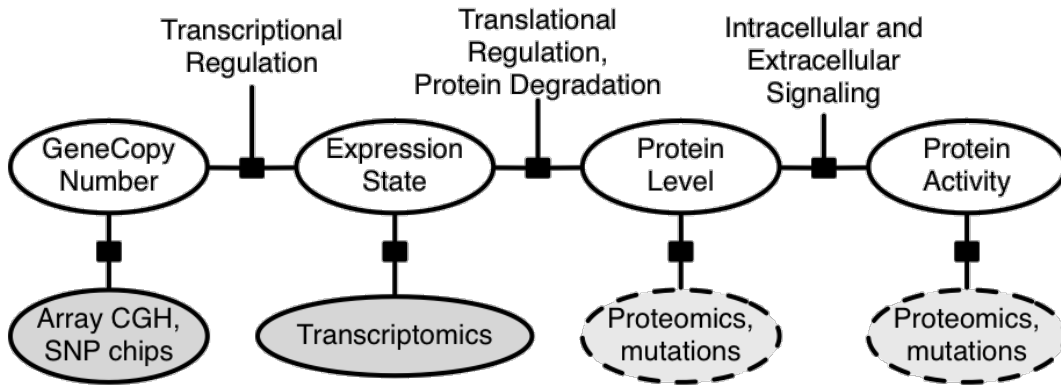
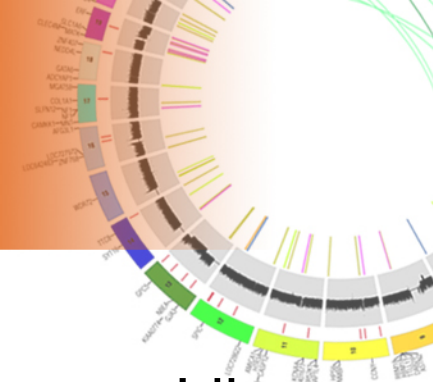


Two Parts:

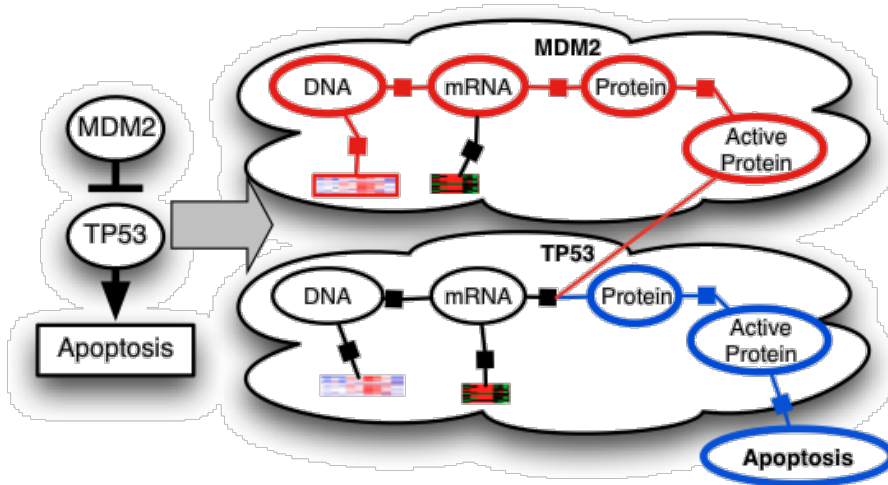
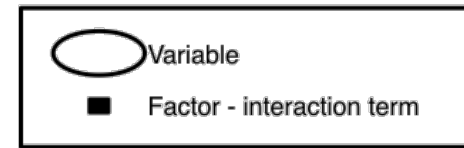


1. Gene Level Model (central dogma)
2. Interaction Model (regulation)

PARDIGM Gene Model to Integrate Data

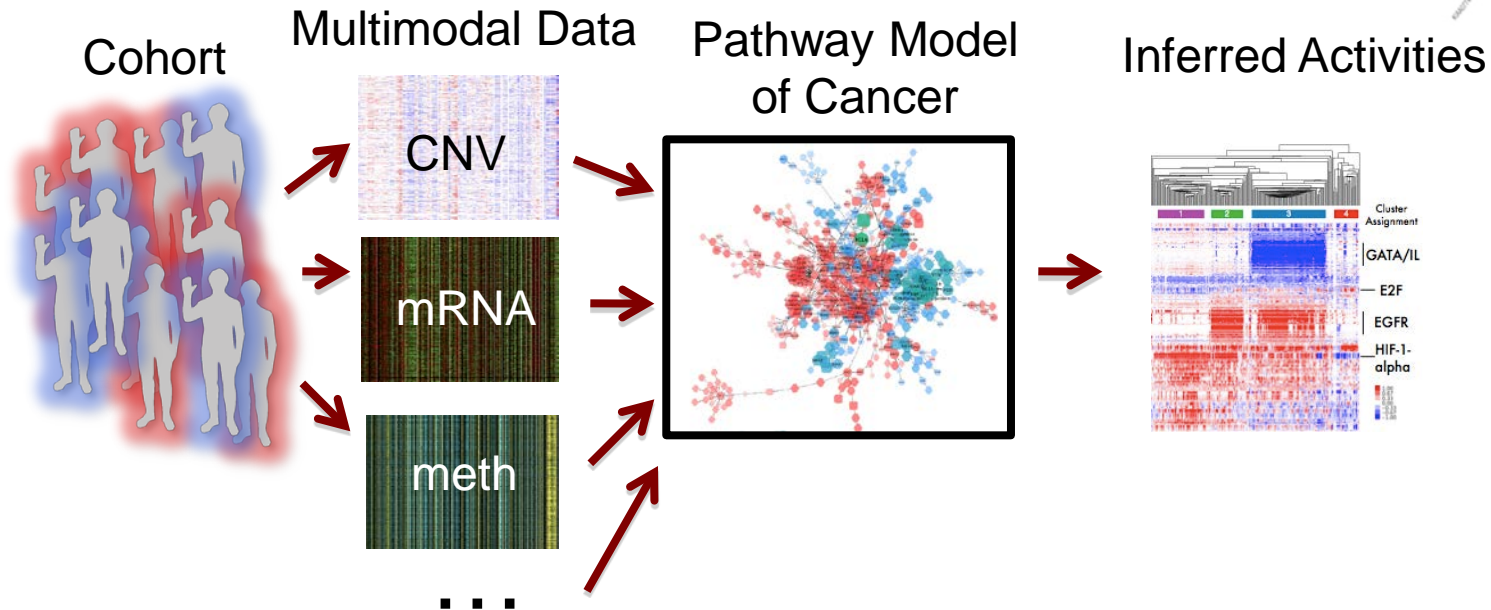


1. Central Dogma-Like Gene Model of Activity



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

Integrated Pathway Analysis for Cancer



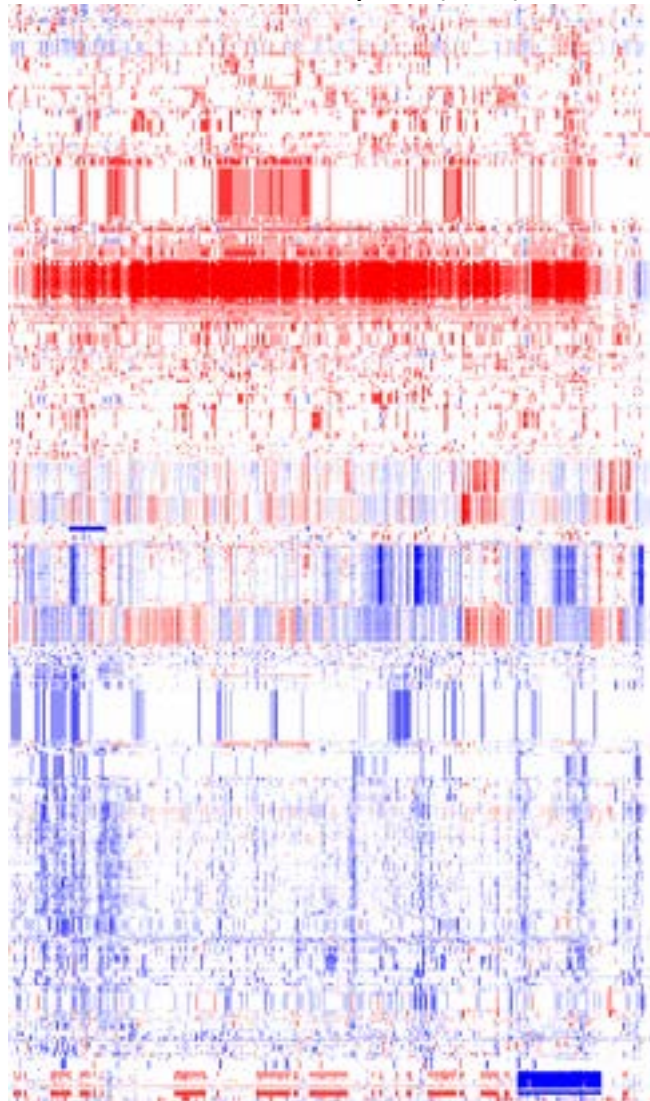
- Integrated dataset for downstream analysis
- Inferred activities reflect neighborhood of influence around a gene.
- Can boost signal for survival analysis and mutation impact

TCGA Ovarian Cancer Inferred Pathway Activities



Patient Samples (247)

Pathway Concepts (867)



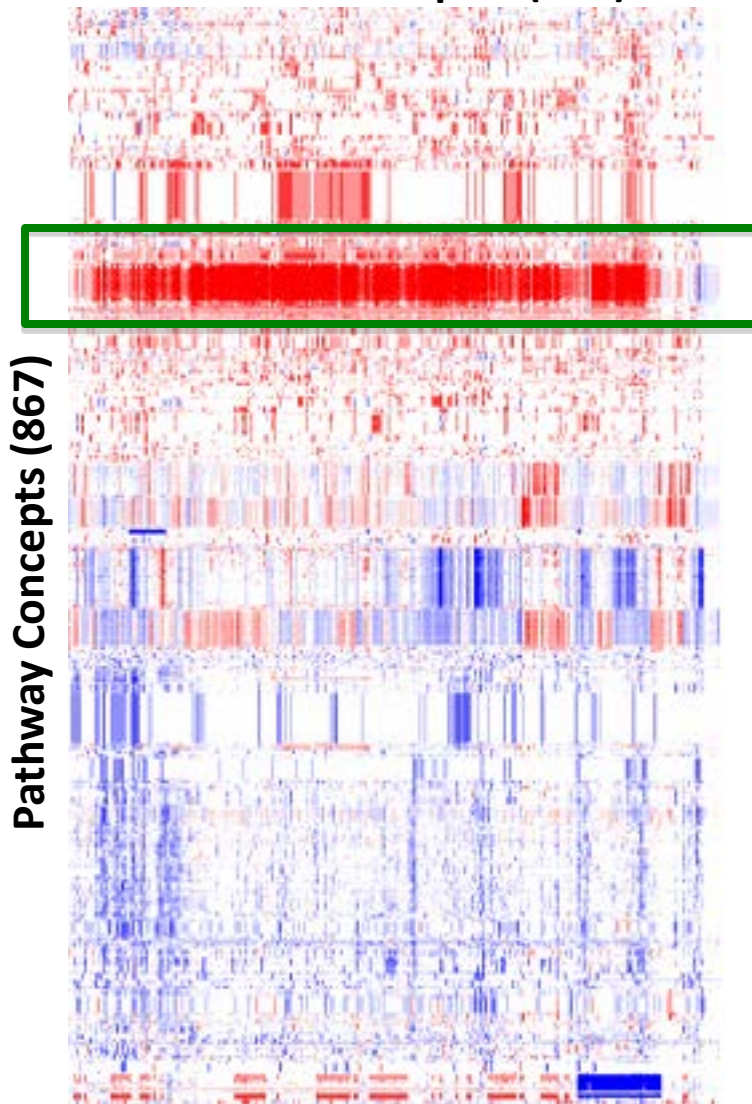
TCGA Network. 2011. *Nature*

The Cancer Genome Atlas 

Ovarian: FOXM1 pathway altered in majority of serous ovarian tumors



Patient Samples (247)



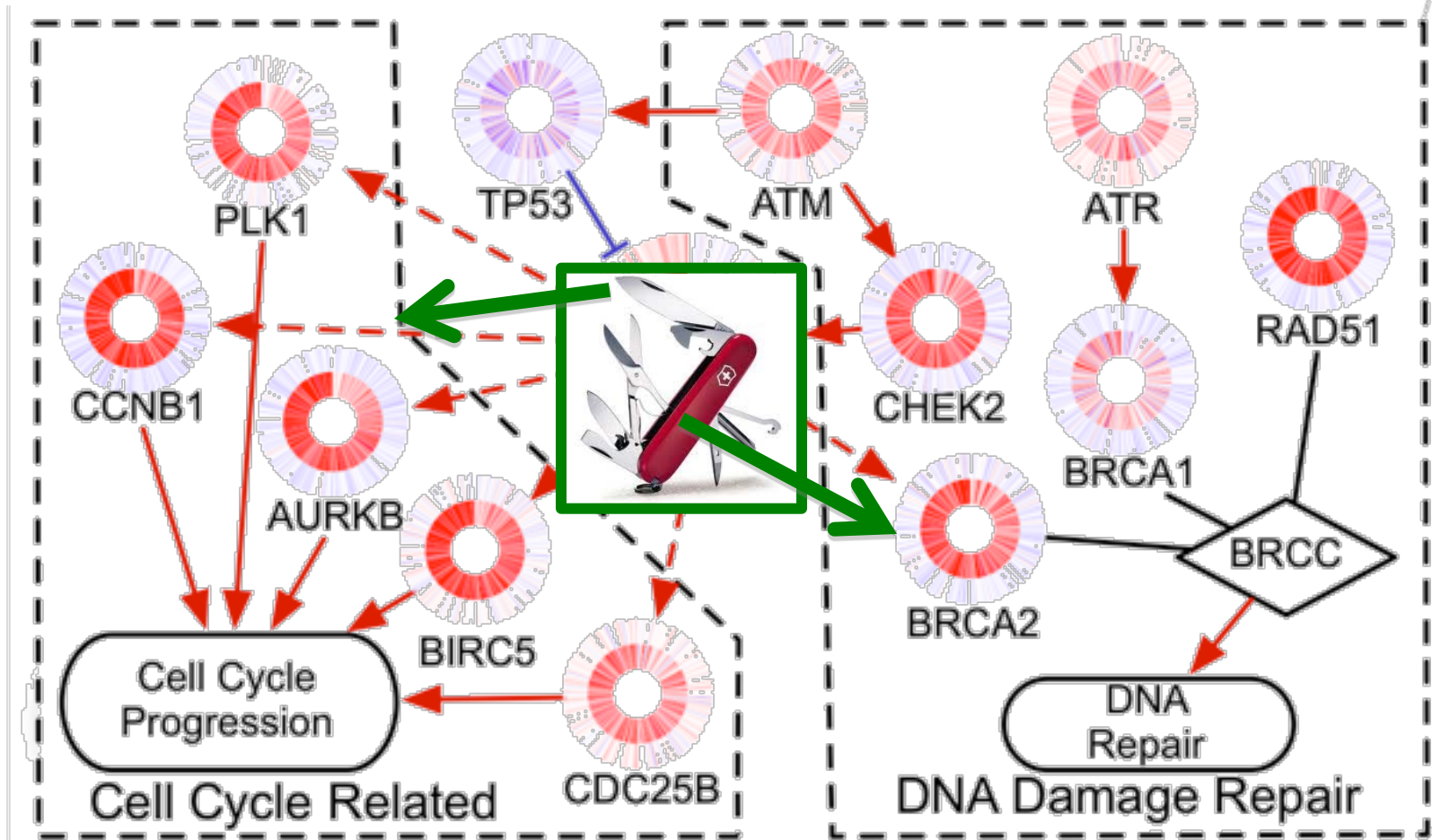
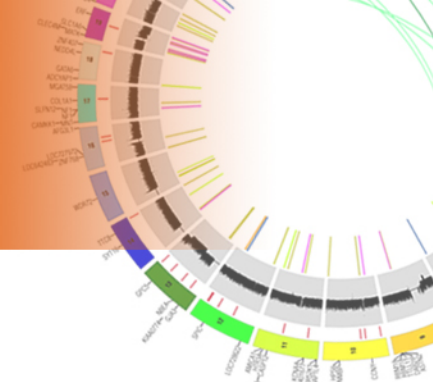
Pathway Concepts (867)

FOXM1 Transcription Network

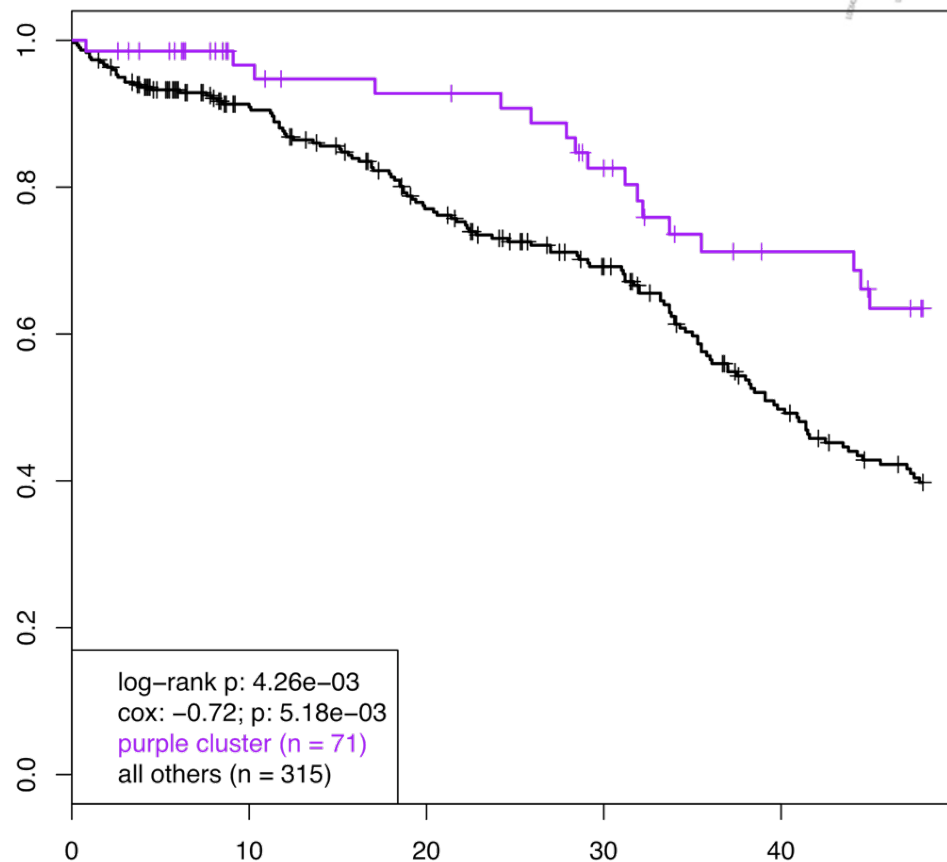
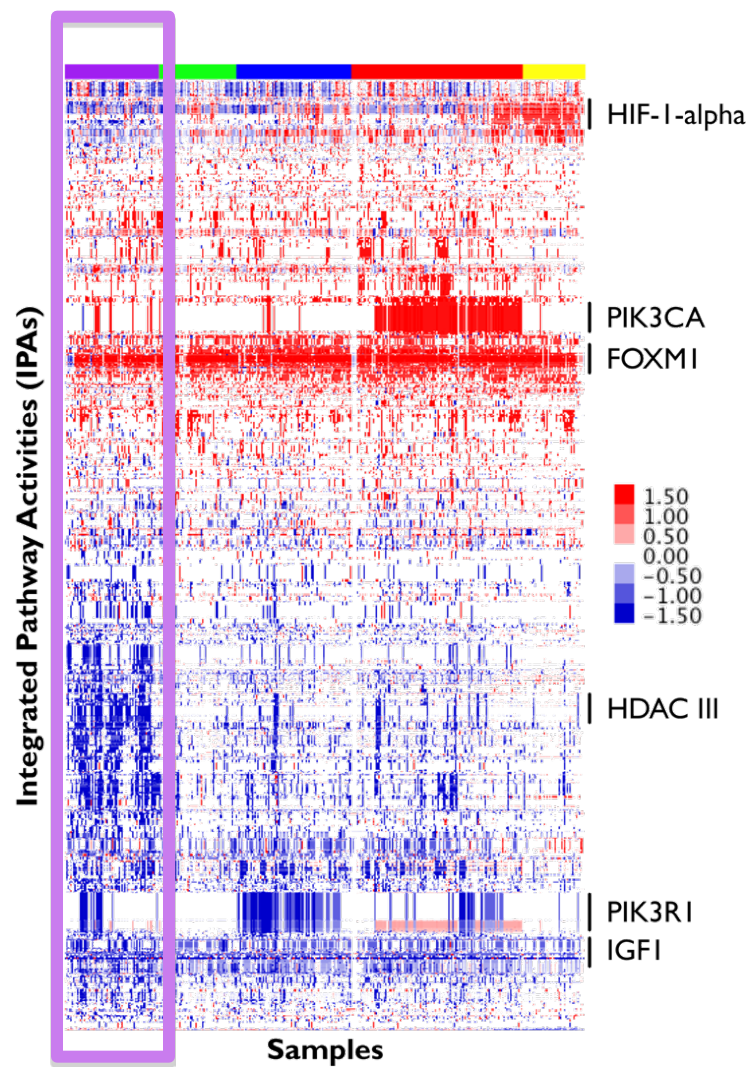
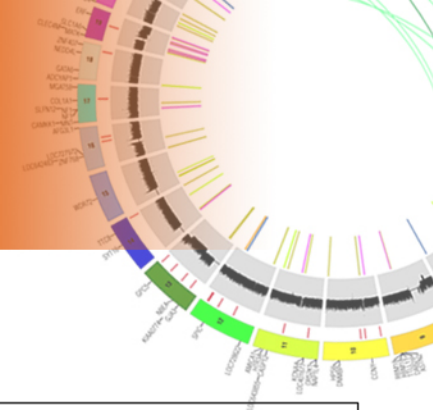
TCGA Network. 2011. *Nature*

The Cancer Genome Atlas 

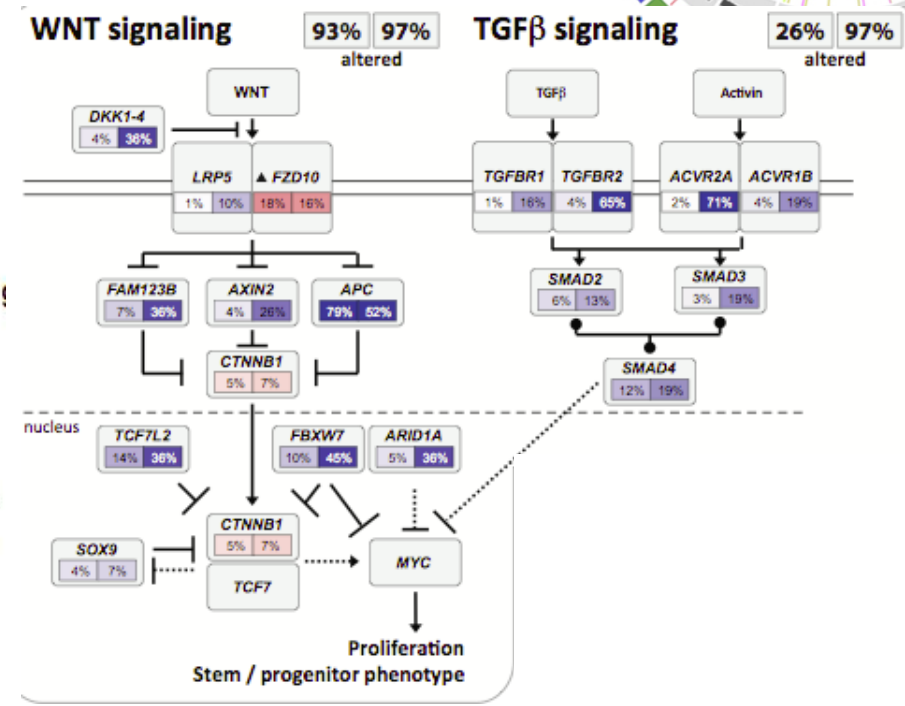
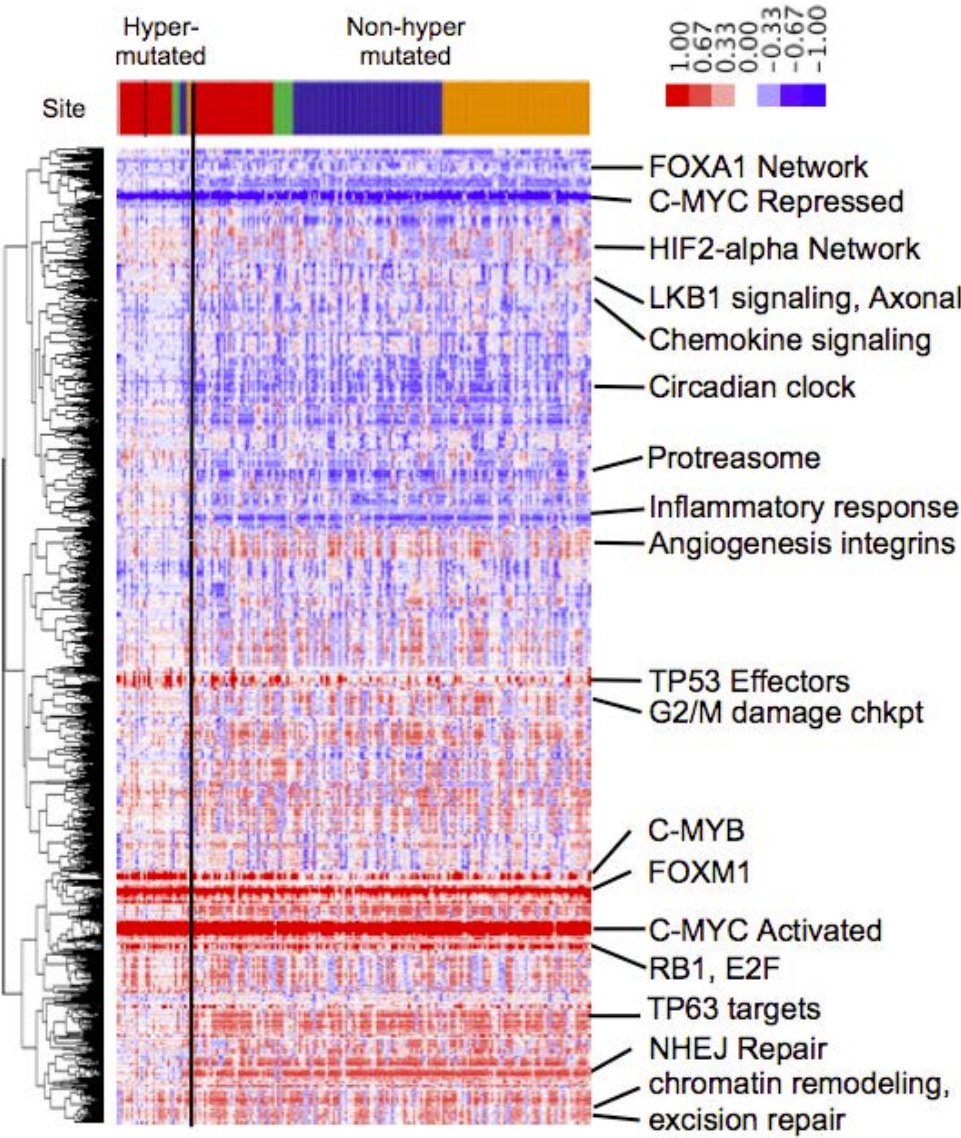
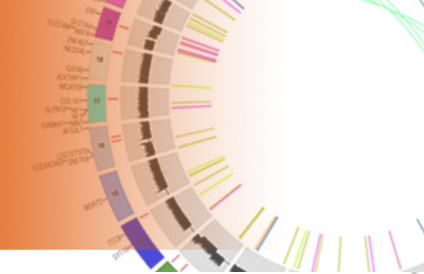
FOXM1 central to cross-talk between DNA repair and cell proliferation in Ovarian Cancer



Ovarian: IPLs stratify by survival time



MYC is characteristically altered in CRC

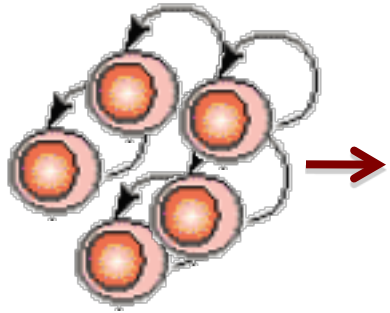


- Cohort-wide disruption of C-MYC
- Common downstream consequence of WNT and TGFβ pathway alterations.

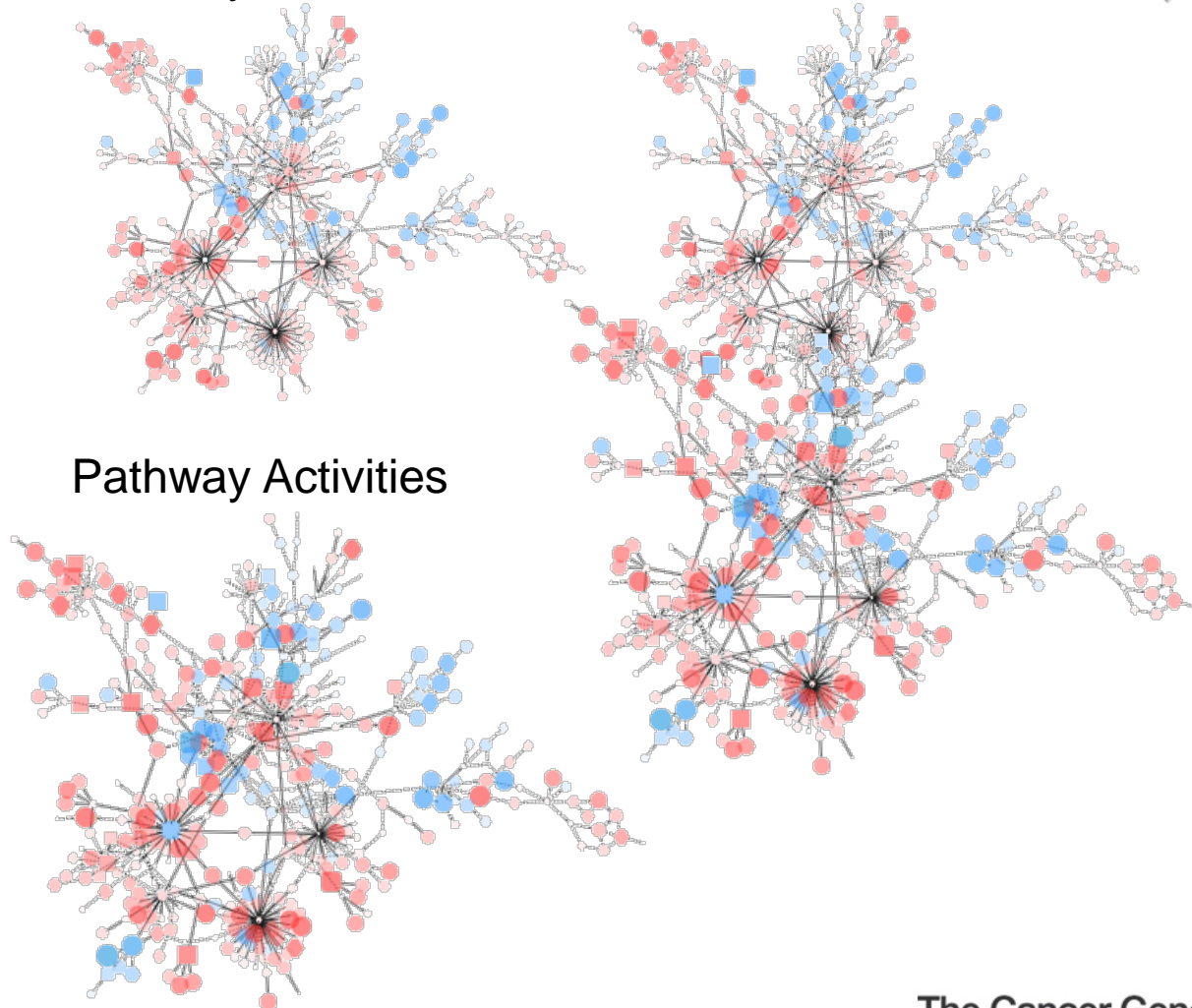
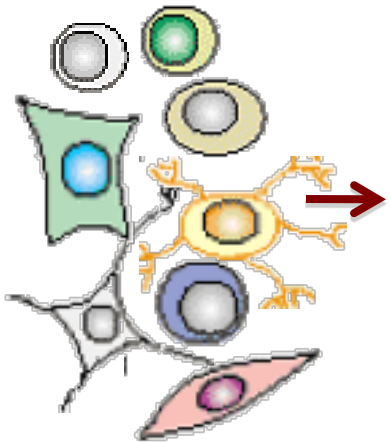
Pathway Signatures: Differential Subnetworks from a "SuperPathway"



Pathway Activities

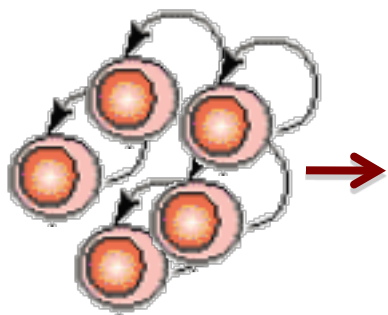


Pathway Activities

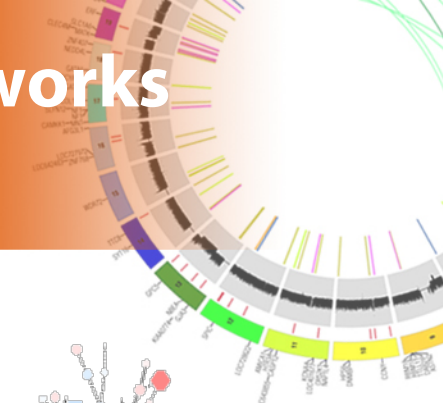
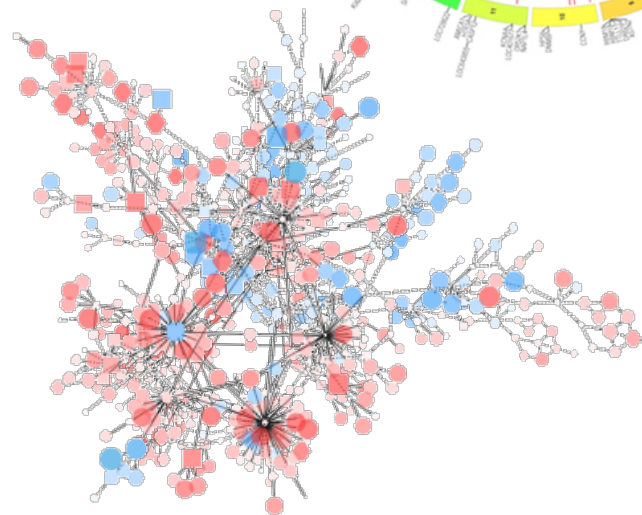
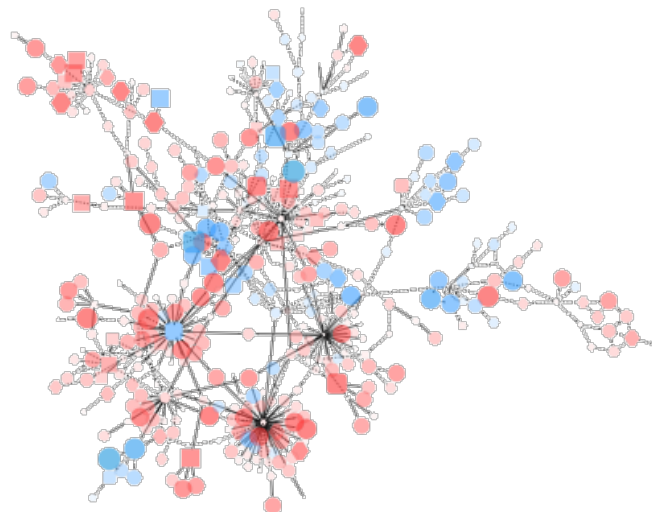
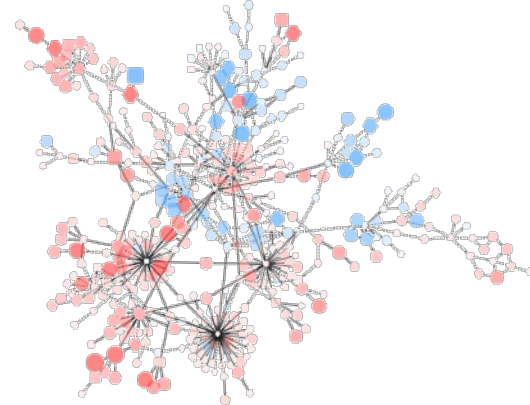
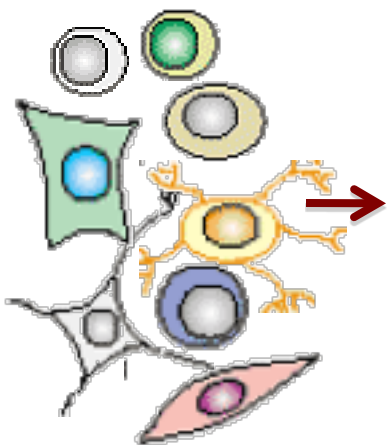


Pathway Signatures: Differential Subnetworks from a "SuperPathway"

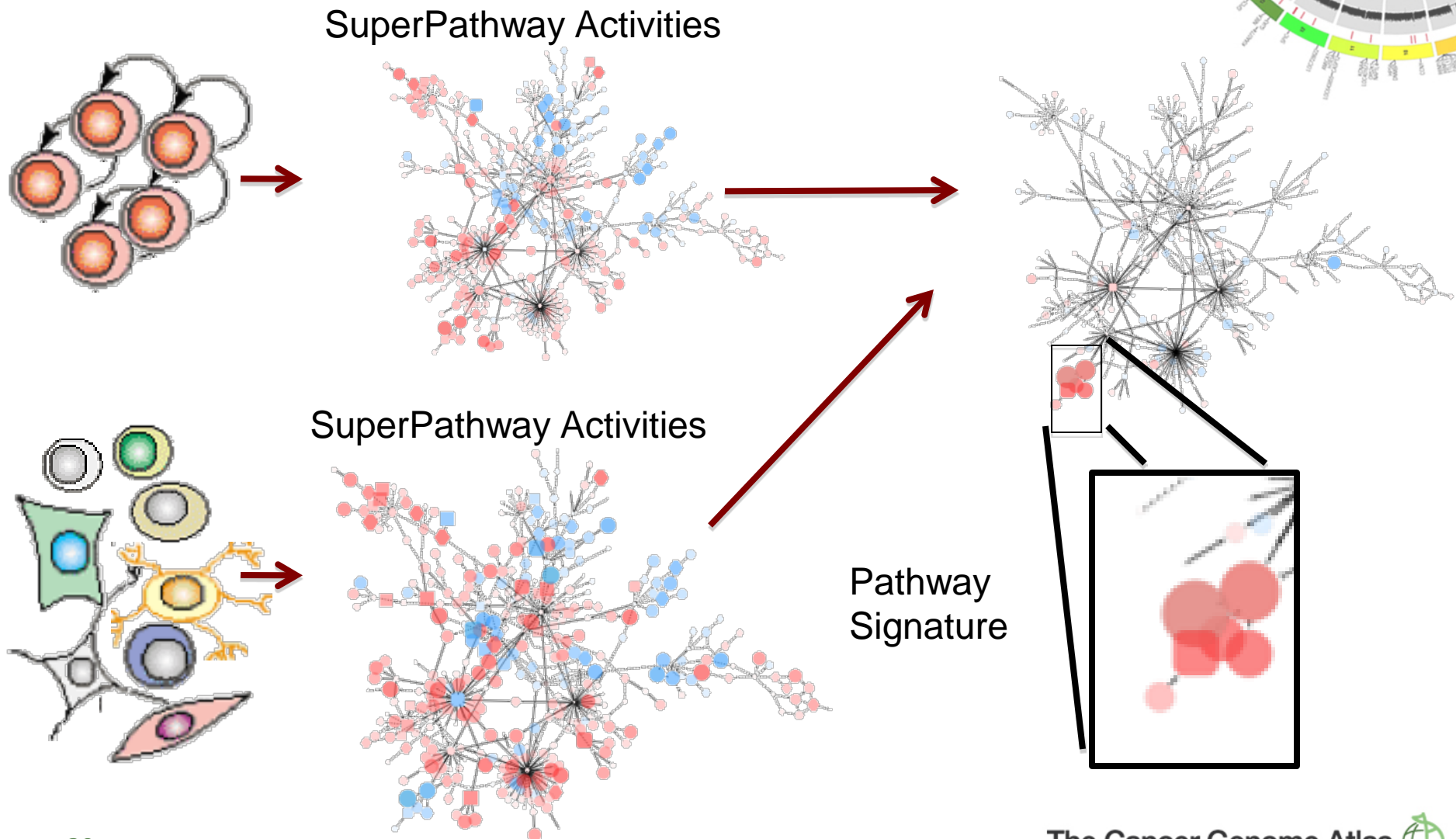
Pathway Activities



Pathway Activities



Pathway Signatures: Differential Subnetworks from a "SuperPathway"



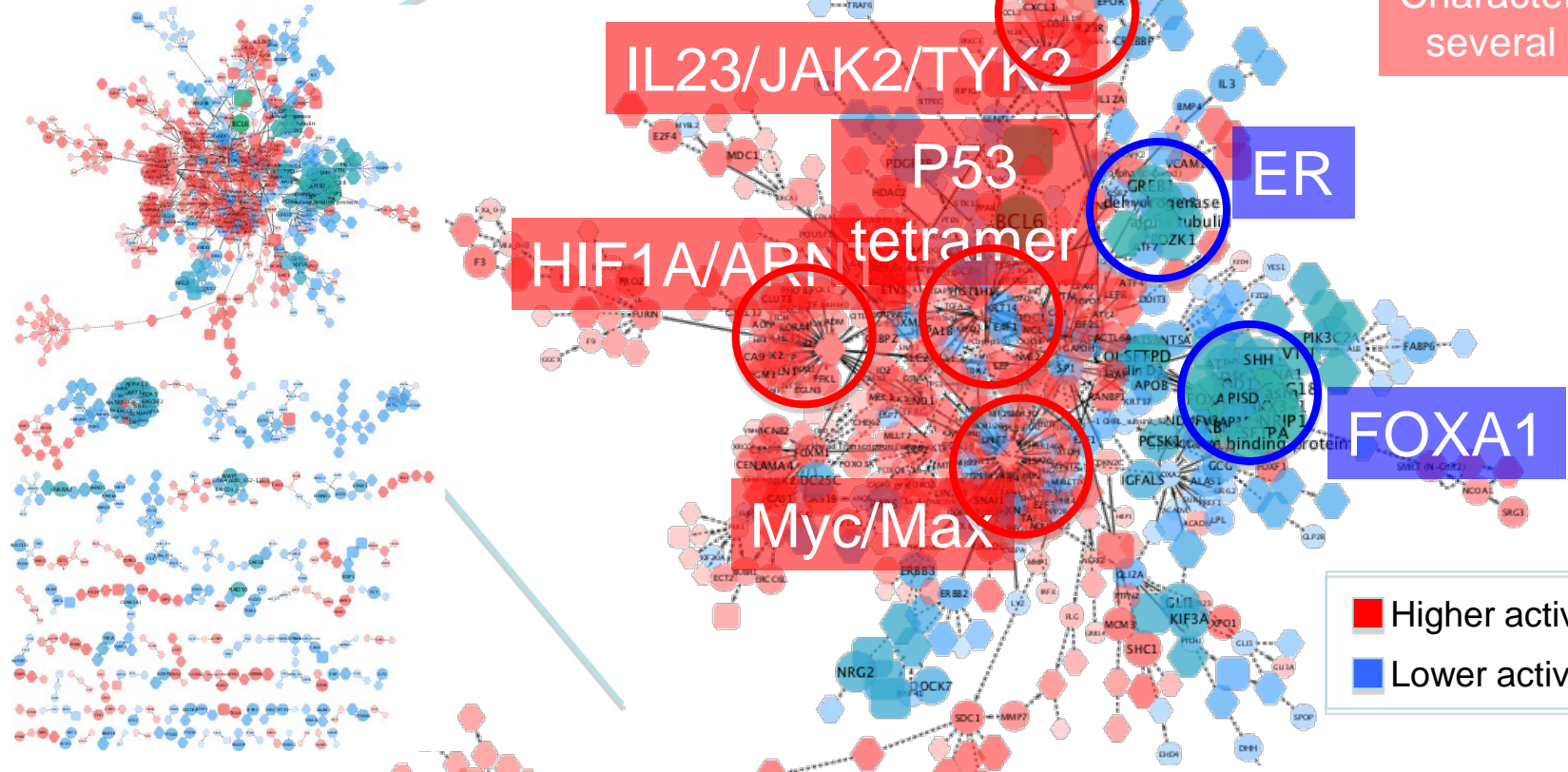
Triple Negative Breast Pathway Markers Identified from 50 Cell Lines



980 pathway concepts
1048 interactions

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

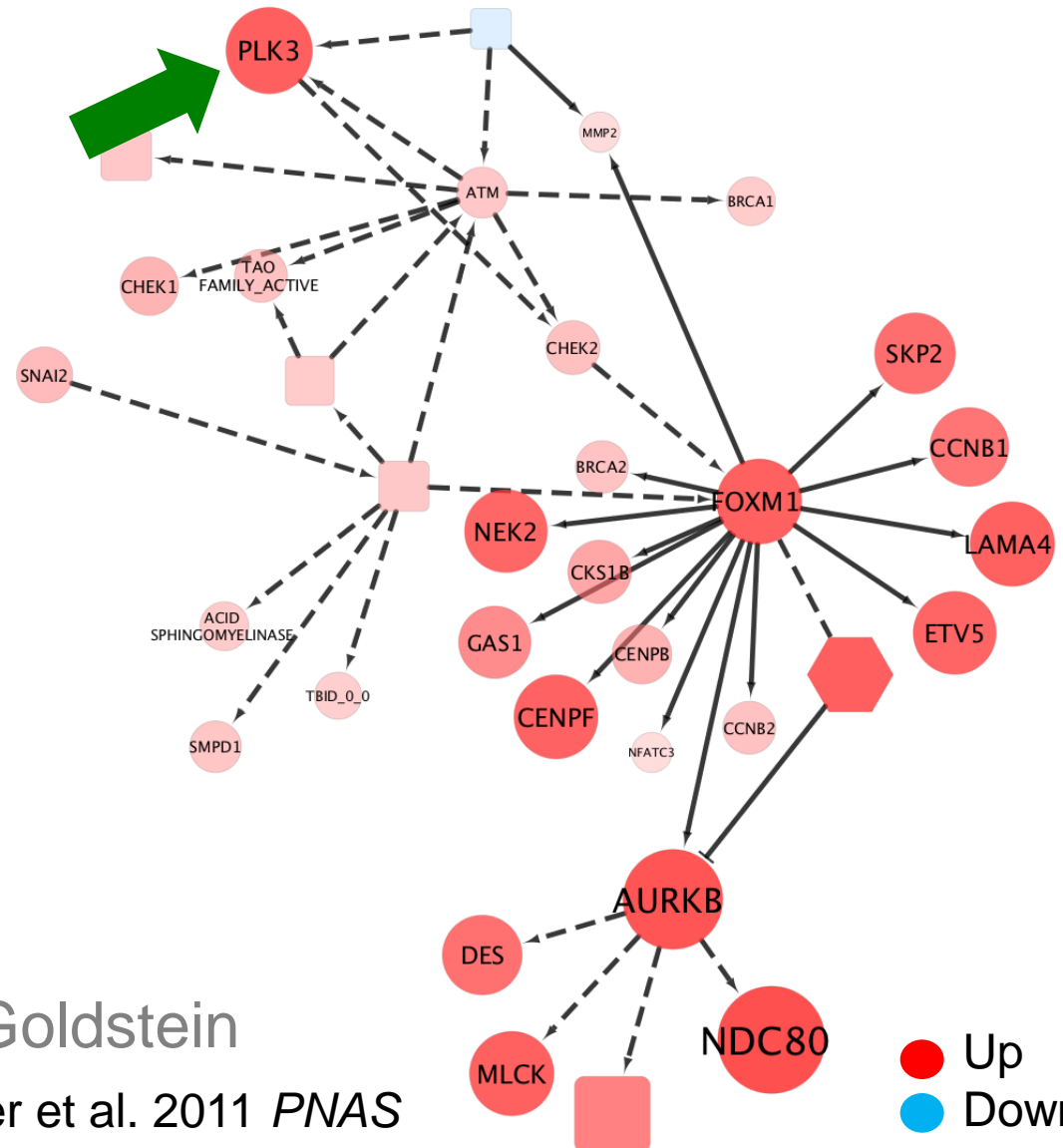
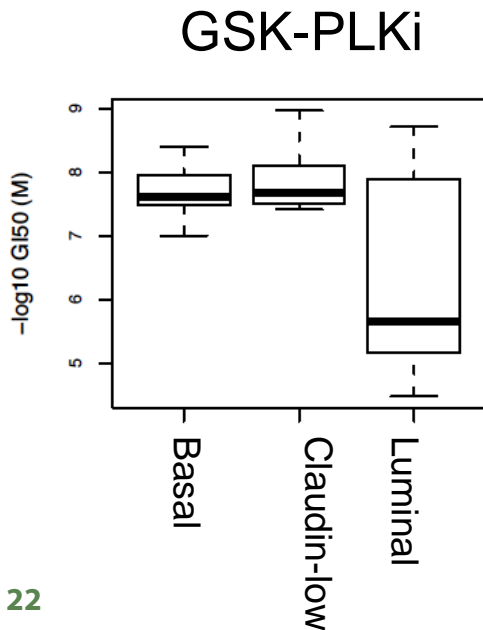
Characterized by several "hubs"



■ Higher activity in ER-
■ Lower activity in ER-

Master regulators predict response to drugs: PLK1 predicted as a target for basal breast

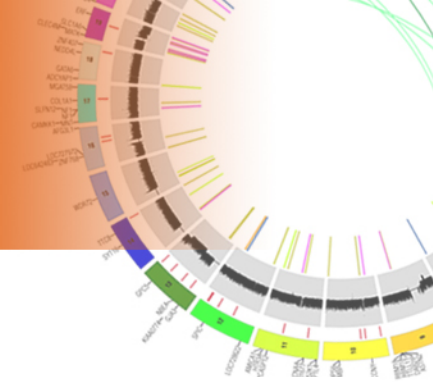
- DNA damage network is upregulated in basal breast cancers
- Basal breast cancers are sensitive to PLK inhibitors



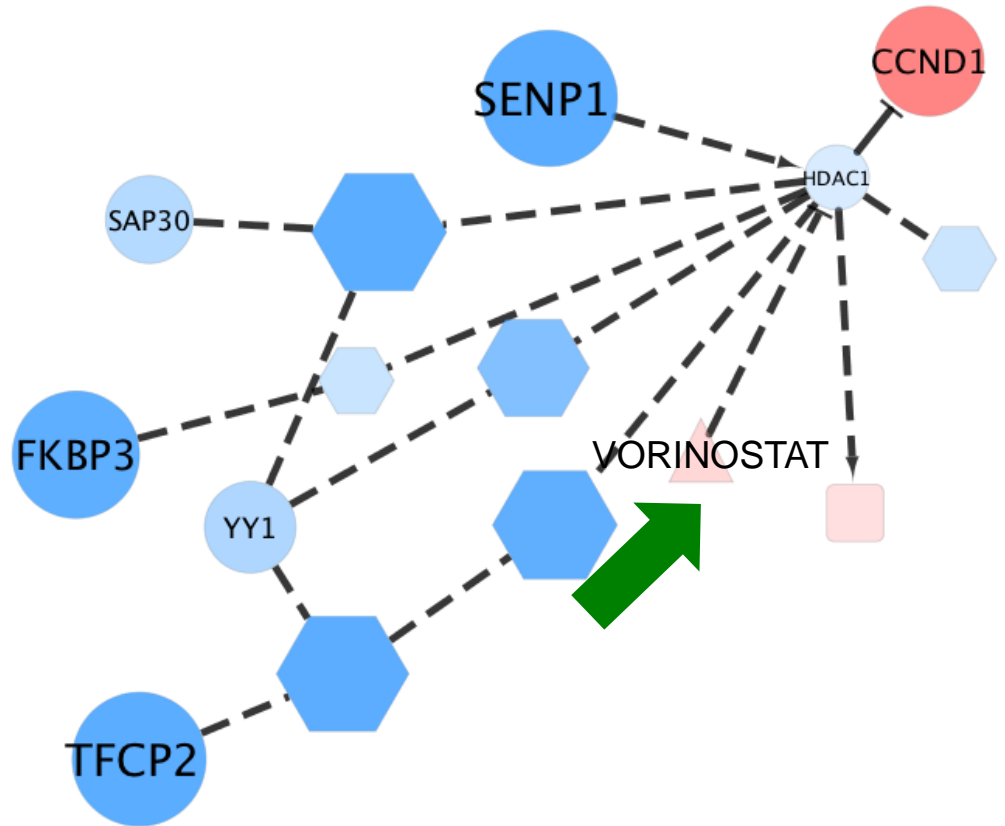
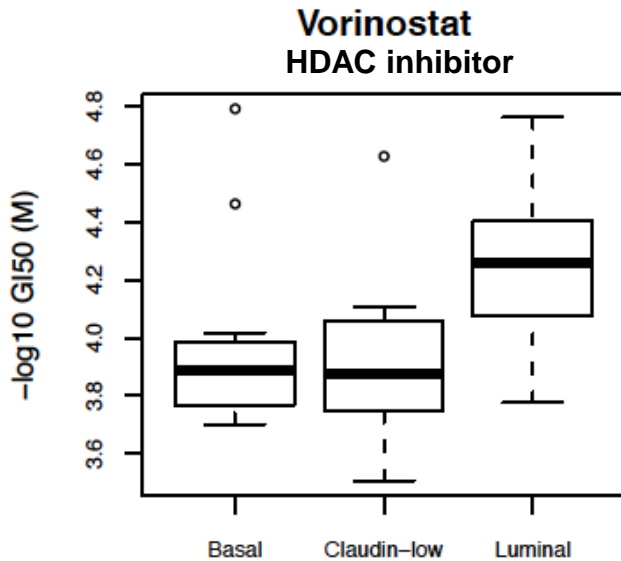
Ng, Goldstein

Heiser et al. 2011 *PNAS*

HDAC inhibitors predicted for luminal breast

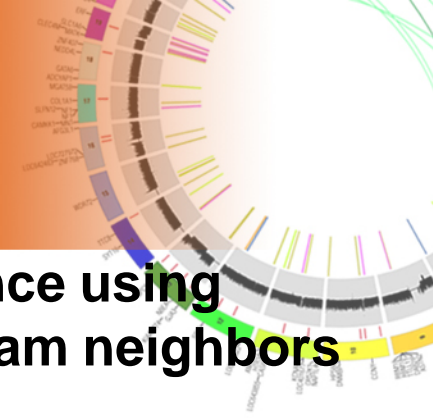


- HDAC Network is down-regulated in basal breast cancer cell lines
- Basal/CL breast cancers are resistant to HDAC inhibitors

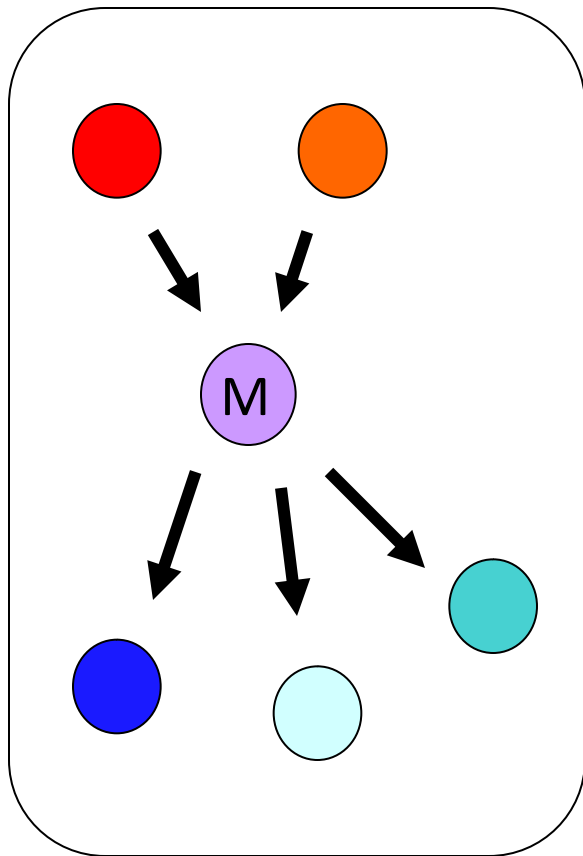


Ng, Goldstein

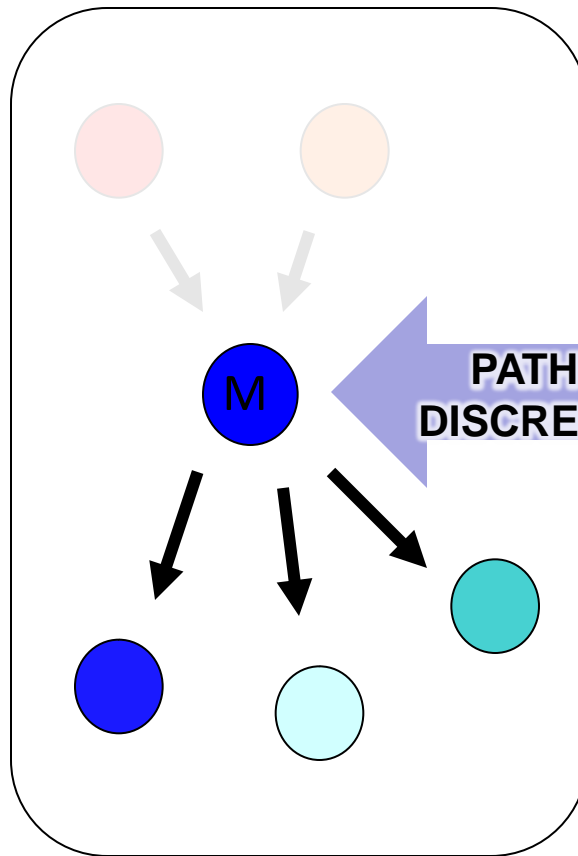
Predicting the Impact of Mutations On Genetic Pathways



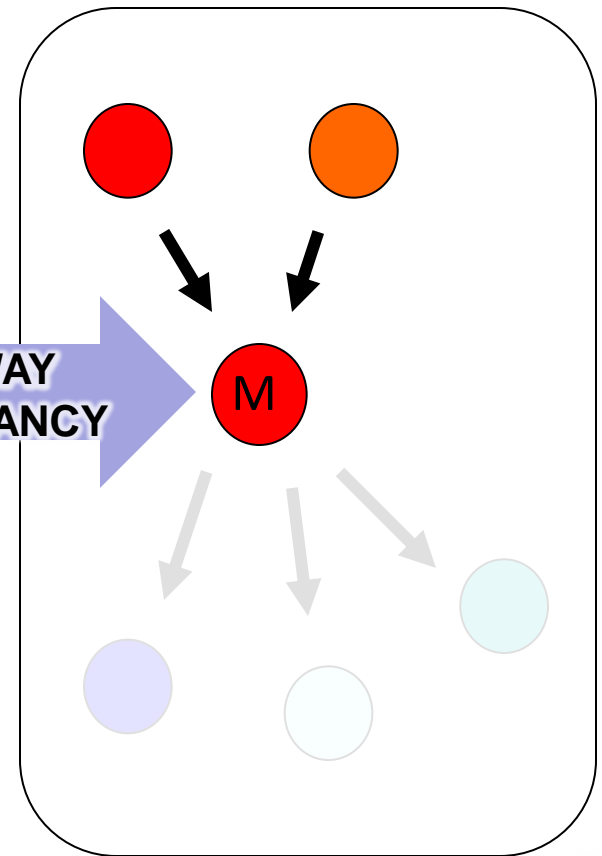
Inference using all neighbors



Inference using downstream neighbors



Inference using upstream neighbors



← PATHWAY DISCREPANCY →

RB1 Loss-of-Function (GBM)



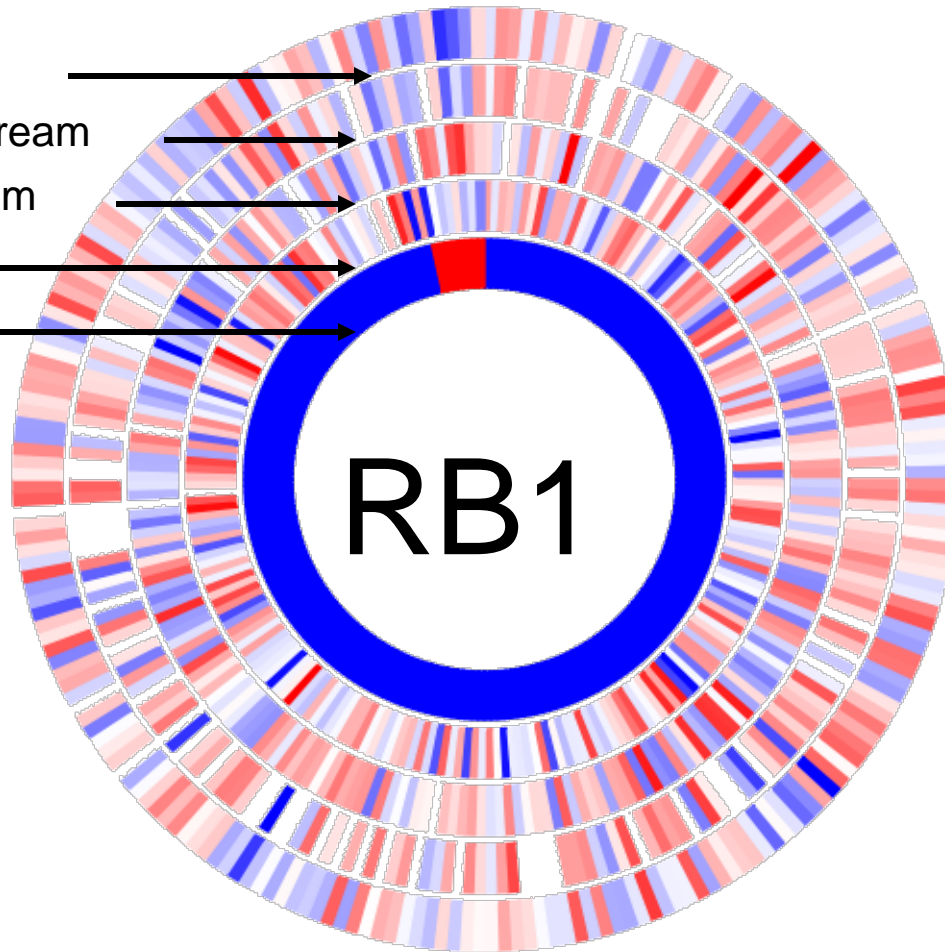
Discrepancy Score

PARADIGM downstream

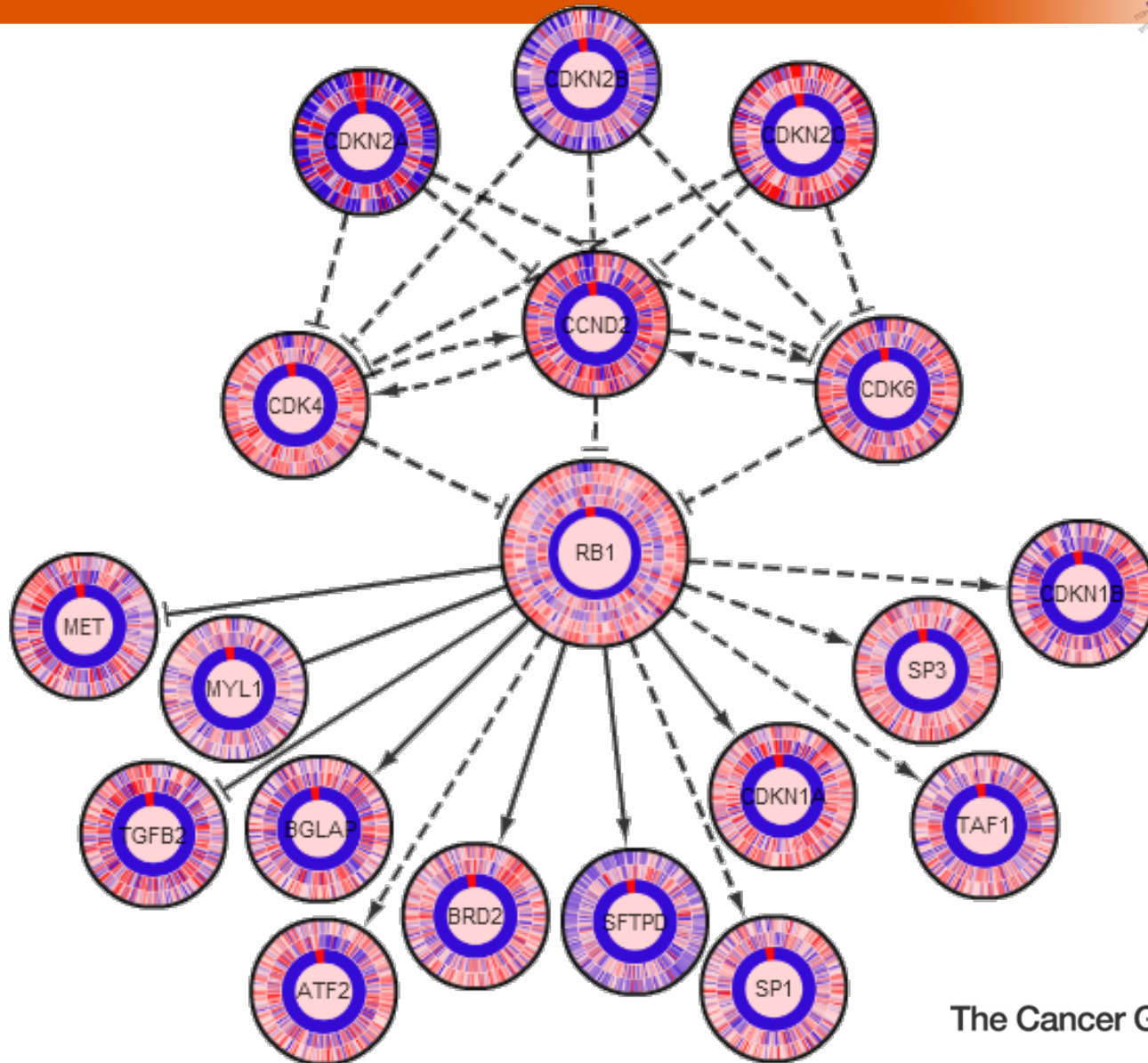
PARADIGM upstream

Expression

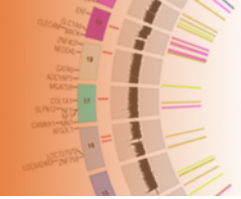
Mutation



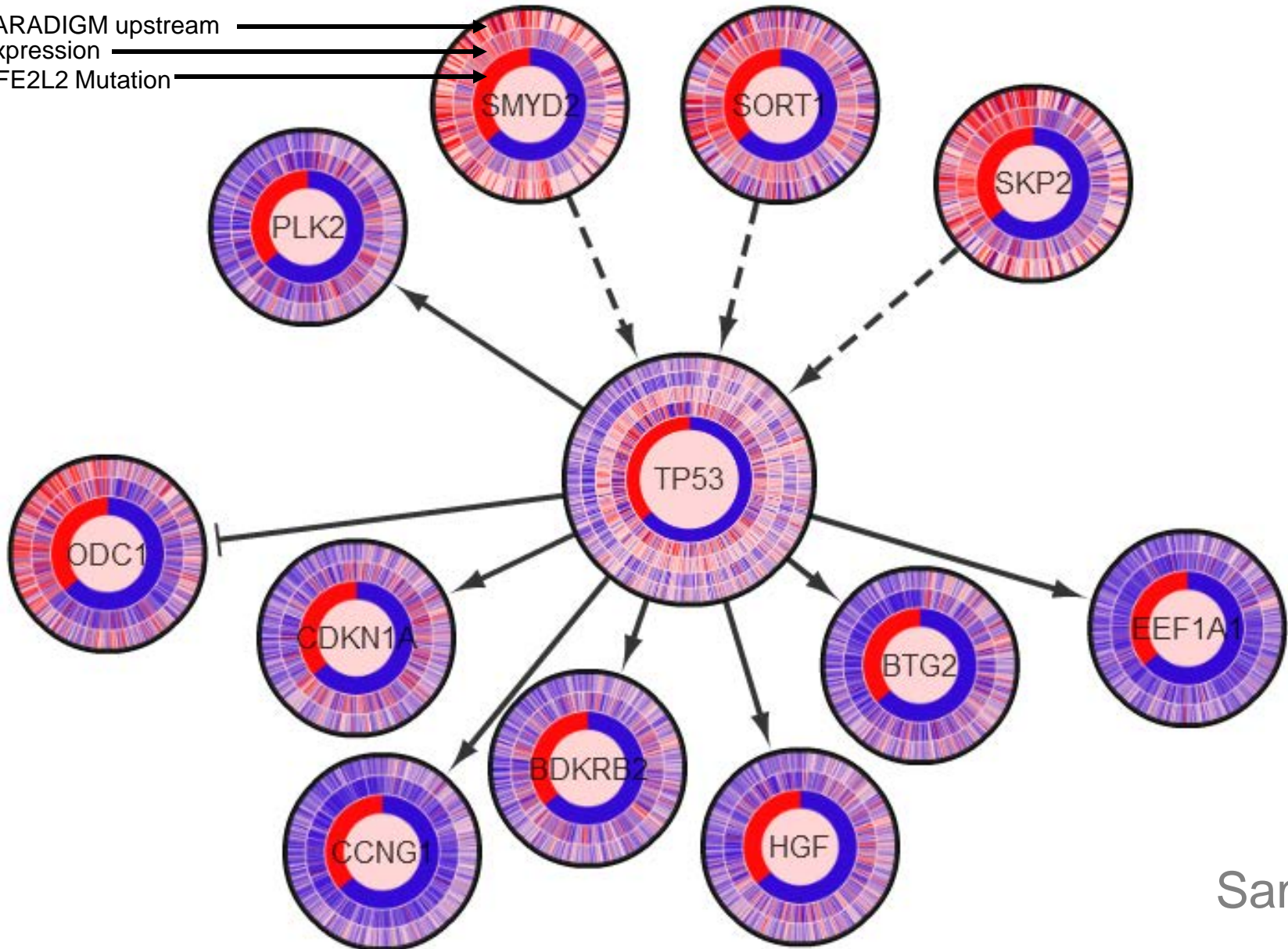
RB1 Network (GBM)



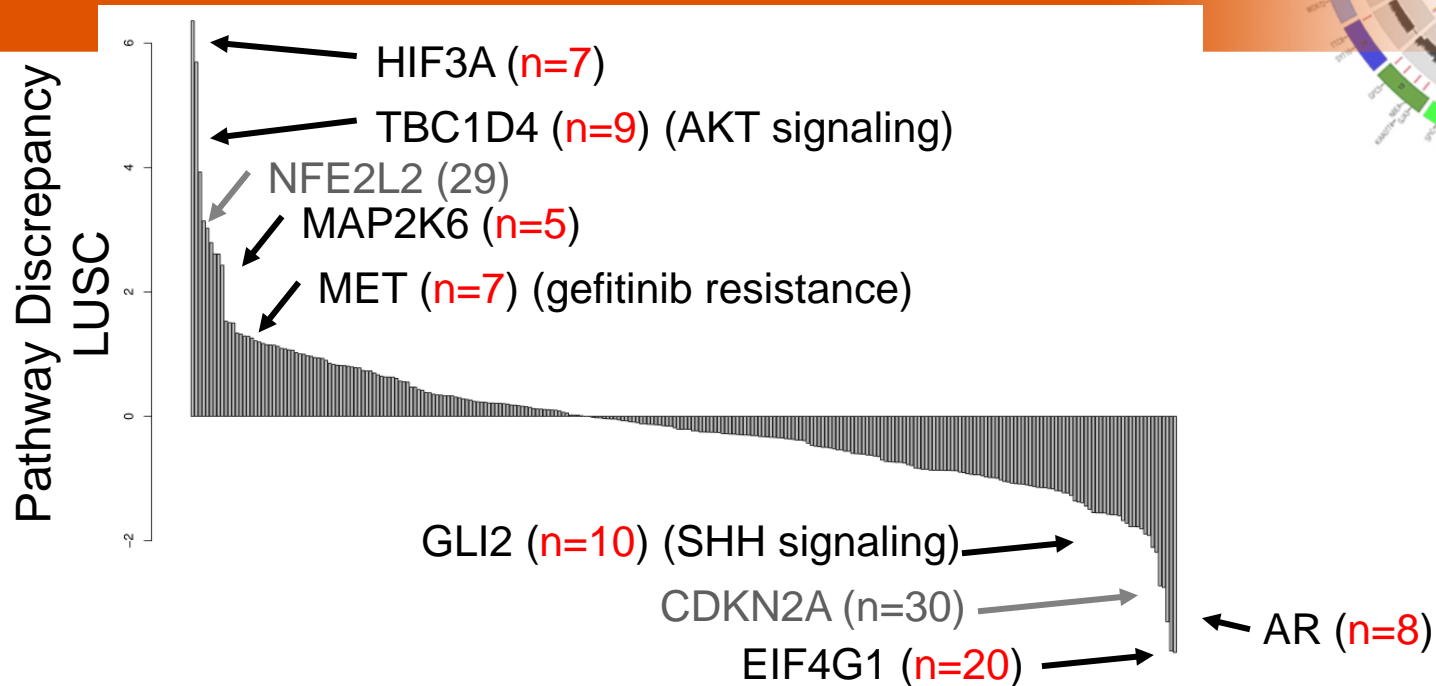
TP53 Network



PARADIGM upstream
Expression
NFE2L2 Mutation



Pathway discrepancy 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

Correlates to mutations?



GC039 C-MYB transcription factor network
braf_mut
histological_type=Colon_Mucinous_Adenocarcinoma
hypermethylation
methclust=CIMP.H
methclust=CIMP.L
methclust=Cluster3
mlh1_hypermethylation
mlh1_silenced
mrnaclust=CIN
mrnaclust=MSI/CIMP
msi_mda=MSI-H
msi_nch=MSI-H
mutfreq
pik3ca_mut
tp53_mut
vascular_invasion_present=YES
GC003 Validated targets of C-MYC transcriptional repression
msi_mda=MSI-L

GC001 FOXA1 transcription factor network
GC002 Validated targets of C-MYC transcriptional repression
GC003 Validated targets of C-MYC transcriptional repression

GC004

GC005

GC006 Chemokine receptors bind chemokines
GC007 HIF-2-alpha transcription factor network

GC008 LKB1 signaling events

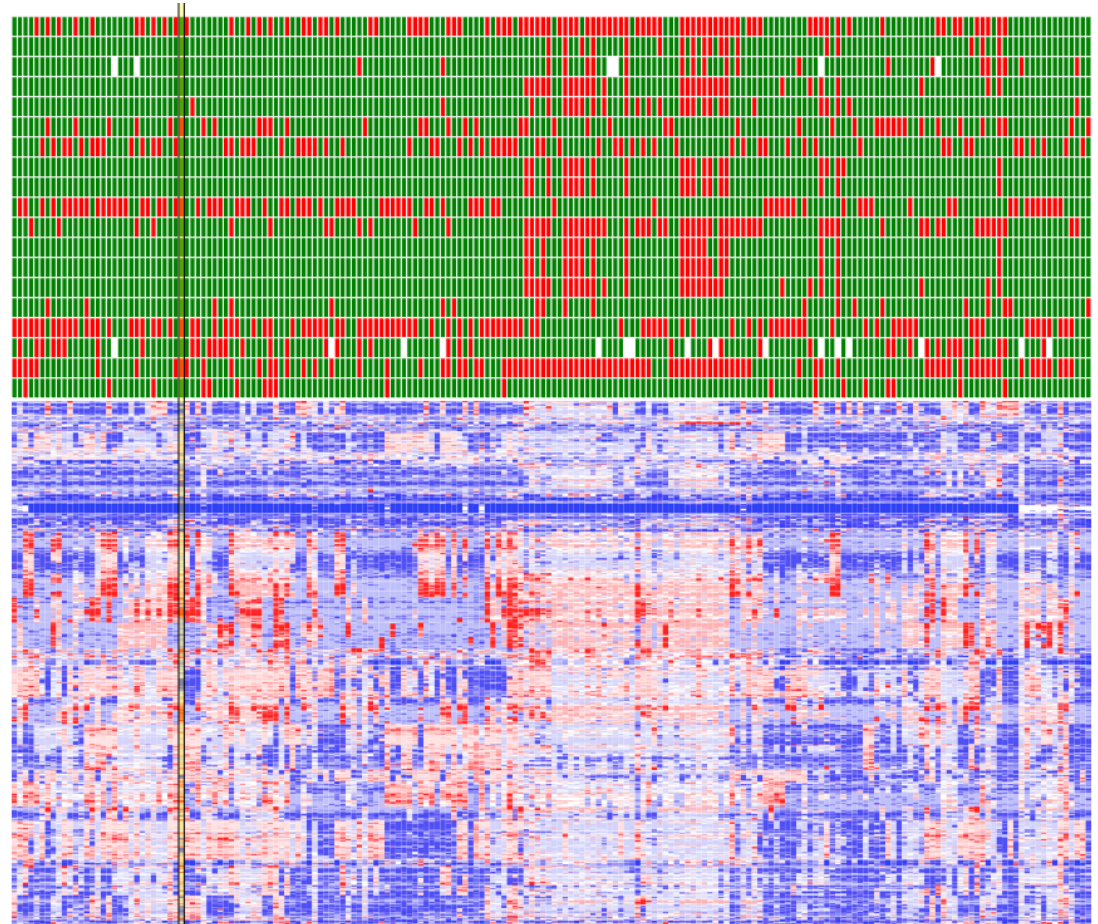
GC009

GC010 P2Y receptors

GC011 Olfactory Signaling Pathway

GC012 Ion transport by P-type ATPases

GC013 Circadian Clock

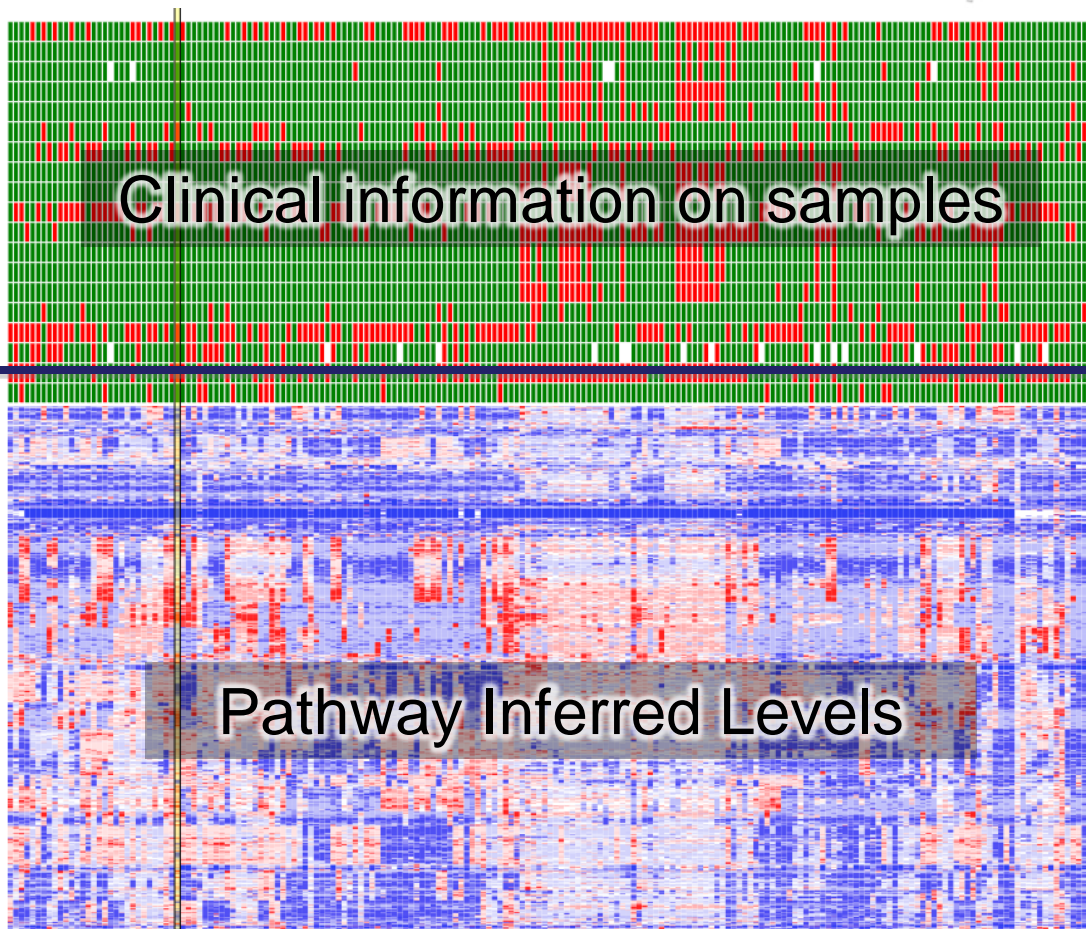


What about when we don't have pathway information for a gene?



GC039 C-MYB transcription factor network
braf_mut
histological_type=Colon_Mucinous_Adenocarcinoma
hypermethylation
methclust=CIMP.H
methclust=CIMP.L
methclust=Cluster3
mlh1_hypermethylation
mlh1_silenced
mrnaclust=CIN
mrnaclust=MSI/CIMP
msi_mda=MSI-H
msi_nch=MSI-H
mutfreq
pik3ca_mut
tp53_mut
vascular_invasion_present=YES
GC002 Validated targets of C-MYC transcriptional repression
msi_mda=MSI-L

GC001 FOXA1 transcription factor network
GC002 Validated targets of C-MYC transcriptional repression
GC003 Validated targets of C-MYC transcriptional repression
GC004
GC005
GC006 Chemokine receptors bind chemokines
GC007 HIF-2-alpha transcription factor network
GC008 LKB1 signaling events
GC009
GC010 P2Y receptors
GC011 Olfactory Signaling Pathway
GC012 Ion transport by P-type ATPases
GC013 Circadian Clock



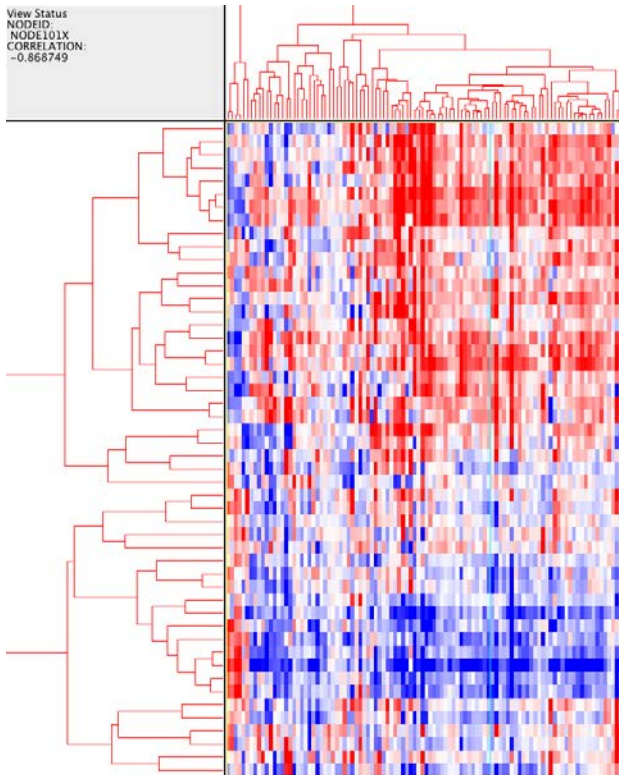
Mutation Association to Pathways

- What pathway activities is a mutation's presence associated?
- Can we classify mutations based on these associations?



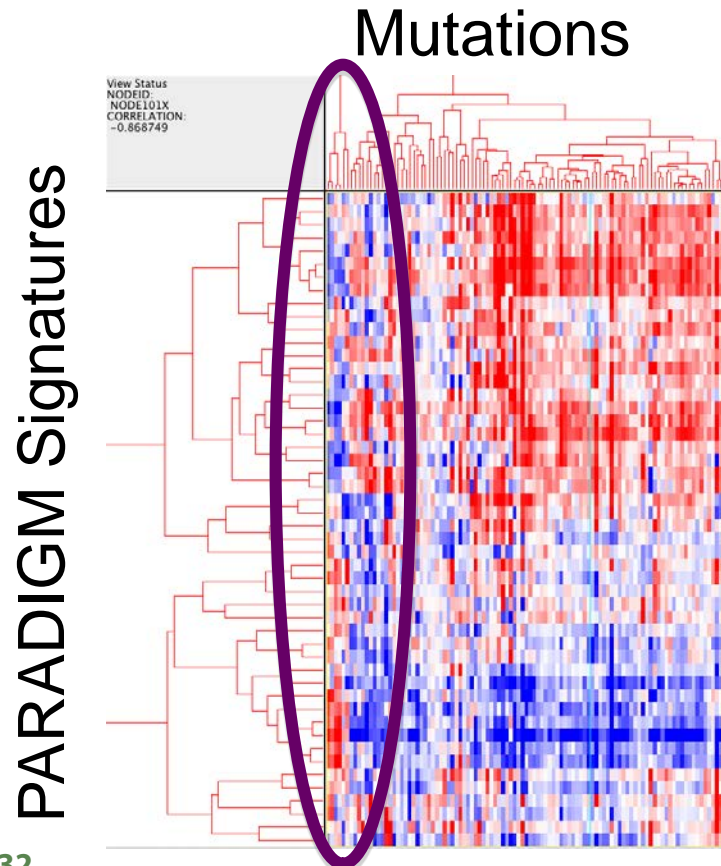
PARADIGM Signatures

Mutations



Mutation Association to Pathways

- What pathway activities is a mutation's presence associated?
- Can we classify mutations based on these associations?



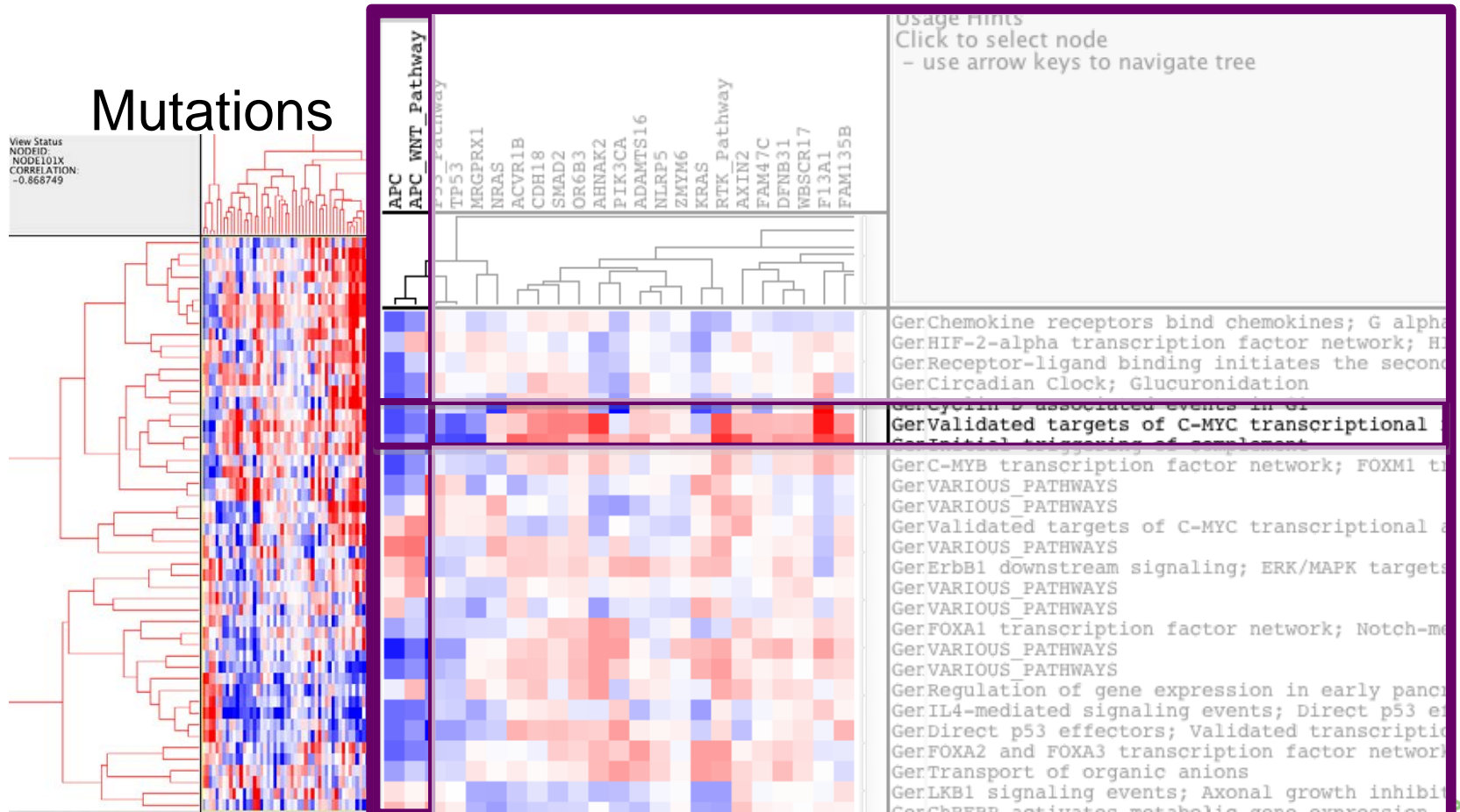
APC and TP53

Mutation Association to Pathways



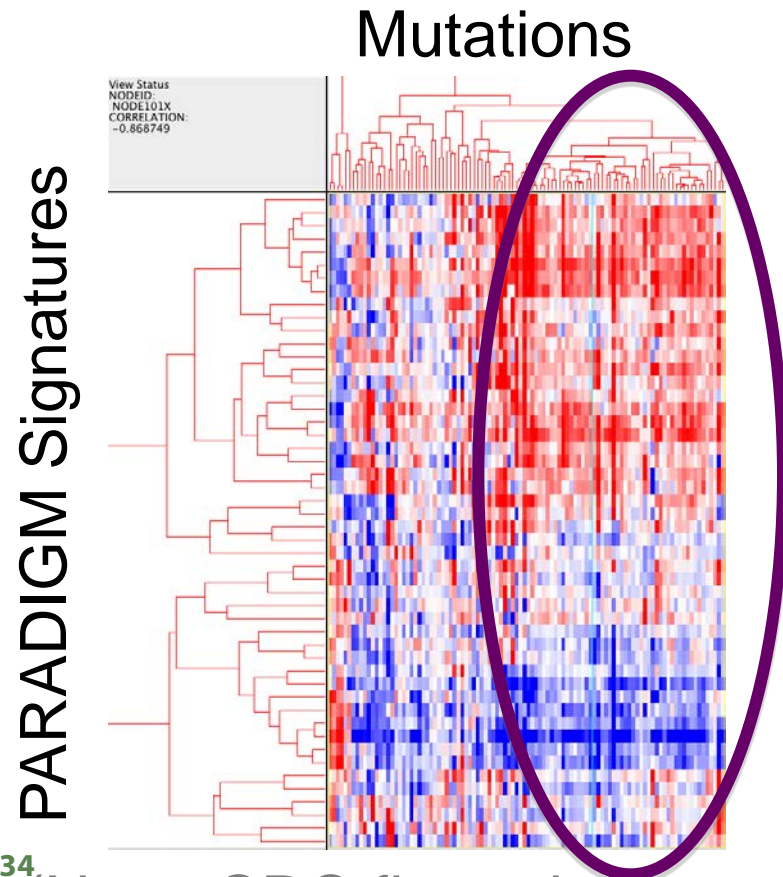
- What pathway activities is a mutation's presence associated?
- Can we classify mutations based on these associations?

PARADIGM Signatures



Mutation Association to Pathways

- What pathway activities is a mutation's presence associated?
- Can we classify mutations based on these associations?



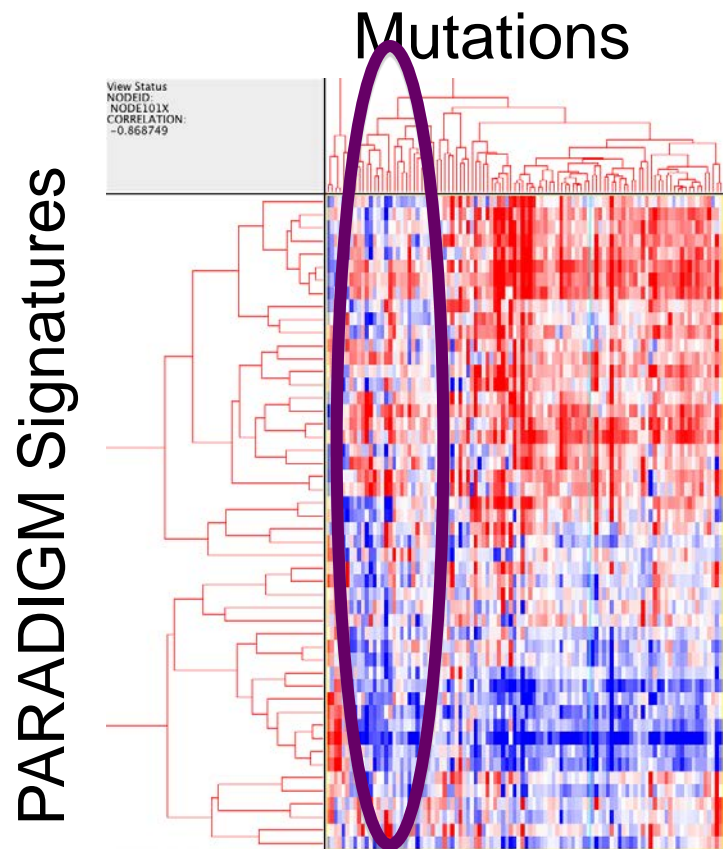
TGFB Pathway mutations

Ted Goldstein
The Cancer Genome Atlas 

34 (Note: CRC figure below; soon for BRCA)

Mutation Association to Pathways

- What pathway activities is a mutation's presence associated?
- Can we classify mutations based on these associations?



PIK3CA, RTK pathway, KRAS

Ted Goldstein

The Cancer Genome Atlas 

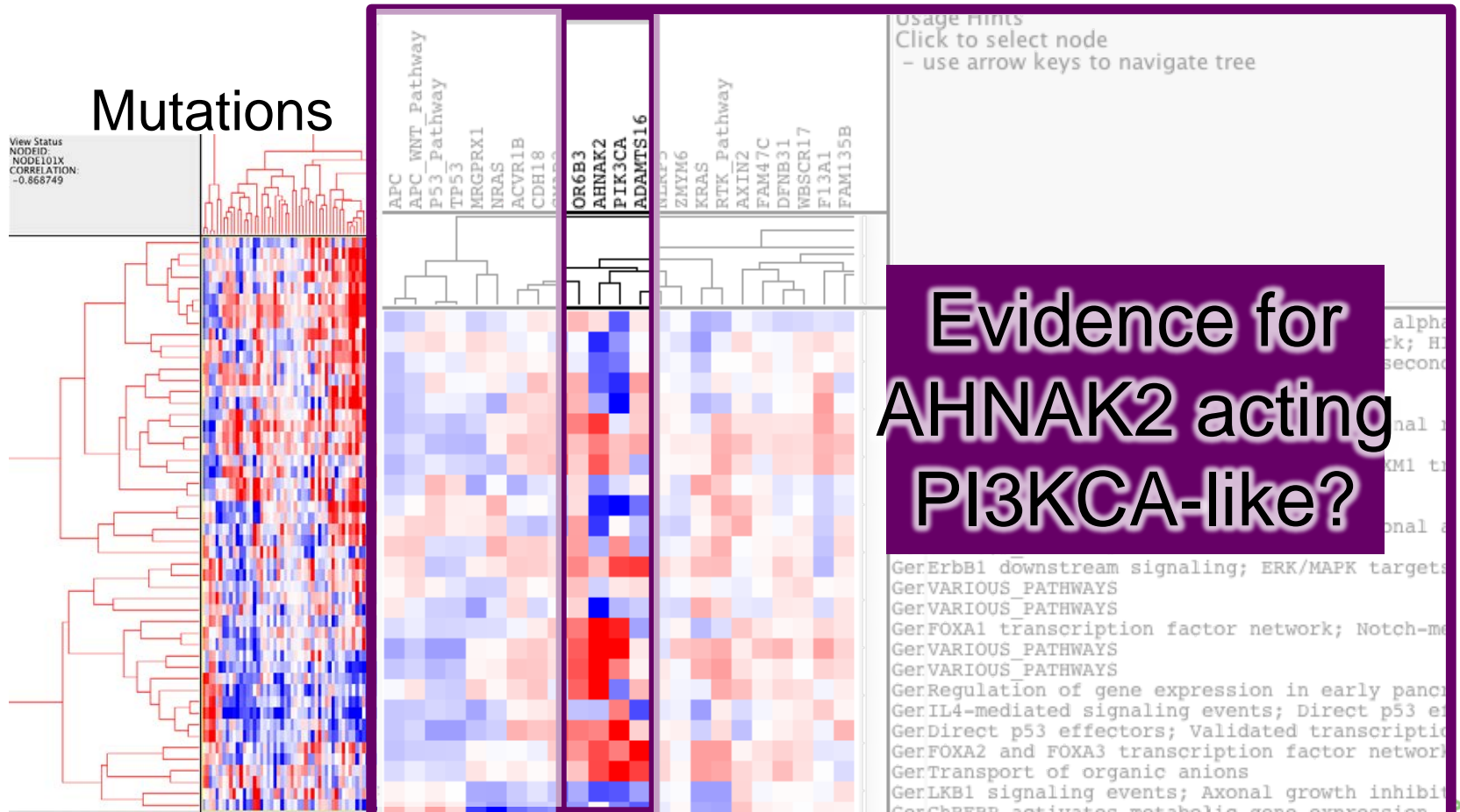
35 (Note: CRC figure below; soon for BRCA)

Mutation Association to Pathways



- What pathway activities is a mutation's presence associated?
- Can we classify mutations based on these associations?

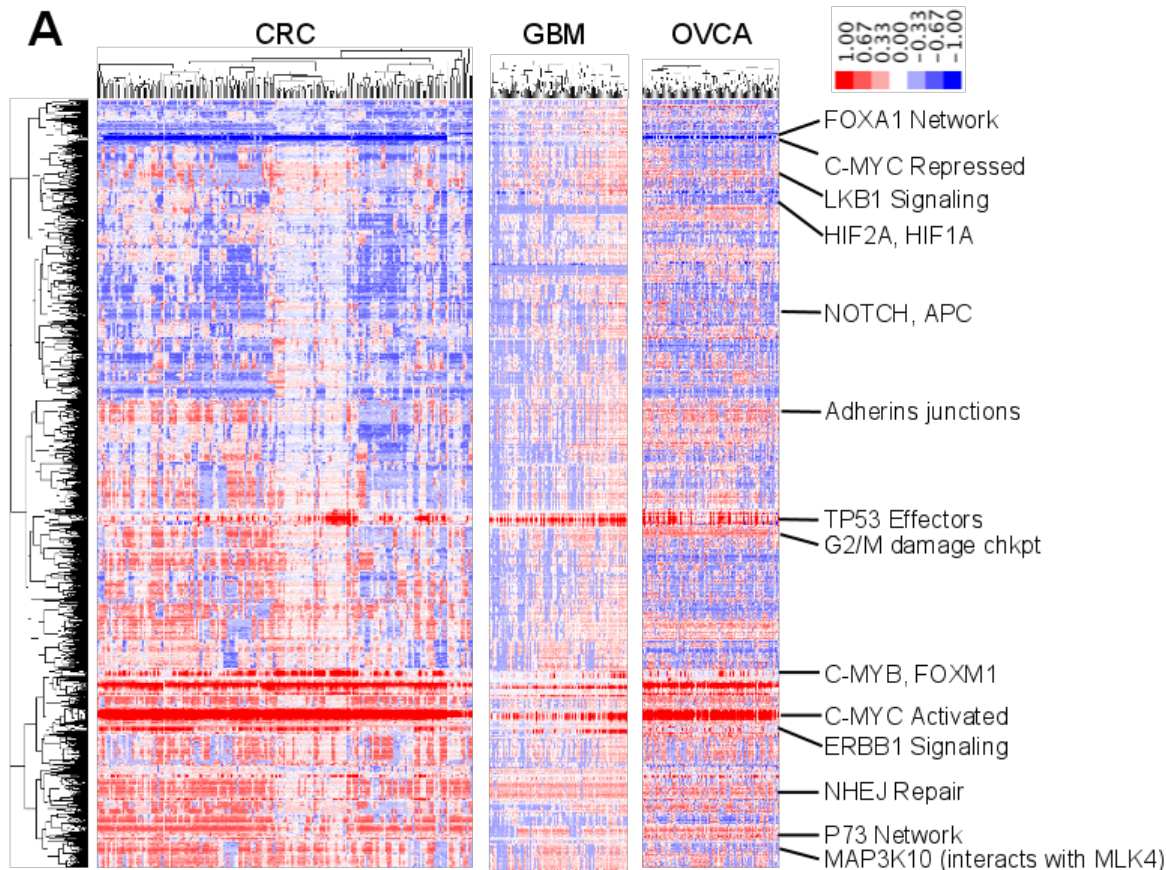
PARADIGM Signatures



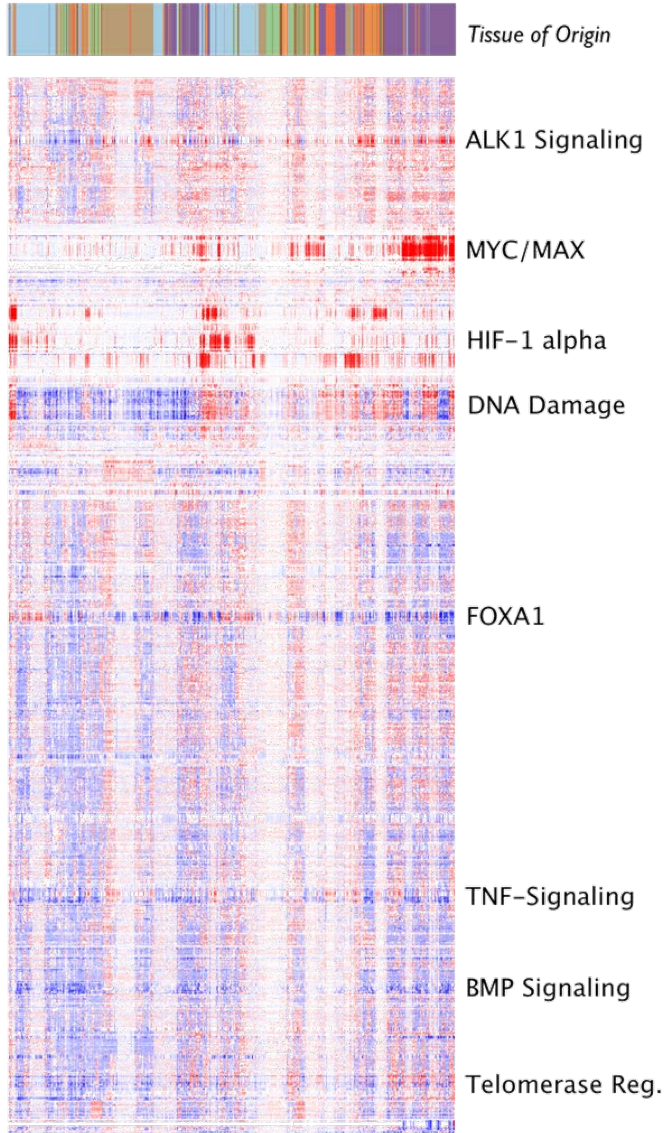
Pan-Cancer: Pathway signatures will connect molecular subtypes across tissues



- Projection of CRC modulated pathways onto GBM and OVCA



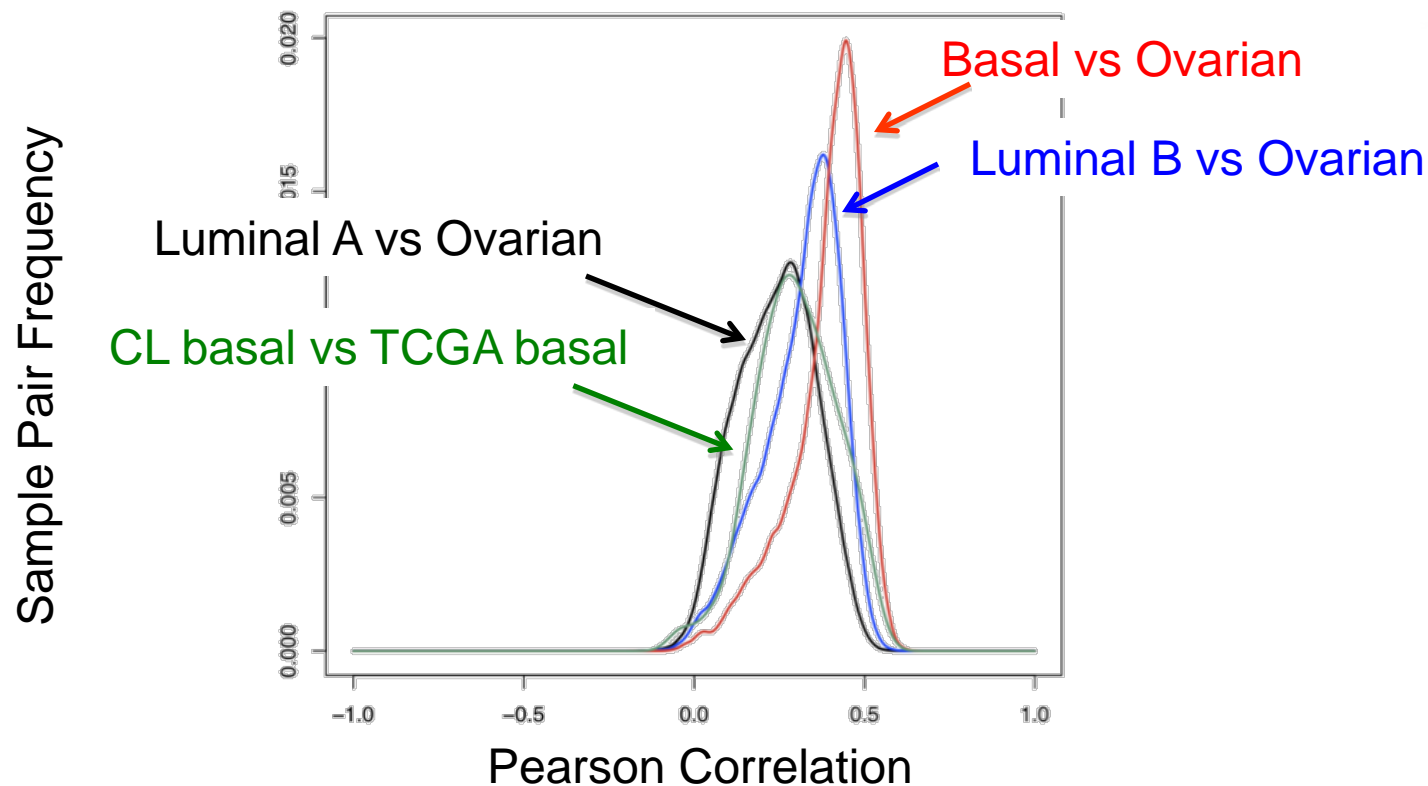
Global Pan-Cancer Map



1382 tumor samples:

- 377 OV
- 69 KIRC
- 251 GBM
- 339 BRCA
- 117 LUSC
- 21 LUAD
- 67 READ
- 141 COAD

Is there a basal disease? – BRCA vs OVCA



- TCGA ovarian more like basal than luminal breast

Summary



- Model information flow to accurately model gene activity using multi-modal data.
- Focus first on known biology
 - Now going after novel biology (new genes and interactions)
- Patient stratification into pathway-based subtypes
- Sub-networks are predictive markers and can be used to simulate scenarios (like drug inhibition)
- Even rare mutations can be assessed for biological significance.
- Enables multi- and pan-cancer analyses

Connecting the dots: A drug for “rare toe carcinoma” (RTCA)



- TCGA cataloging many signatures of tumors: mutation spectrum, altered genes, and pathway activities
 - E.g. patient presents w/ RTCA and has HER2 amplification
- Subtypes, and ultimately single samples can be connected by these signatures
 - RTCA signature checks out w/ PAM50
- We should also engage signatures from external datasets to inform TCGA data (e.g. Connectivity Map)
 - Signature matches lapatinib sensitivity signature
- Provide a basis to bootstrap clinical findings
 - Prescribe lapatinib to RTCA patient

Shout out to the Broad Team



- PARADIGM now included in Firehose
 - Public now can access CPU-intensive results
- Special THANKS to Daniel DeCara.

UCSC Integrative Genomics Group

Marcos Woehrmann



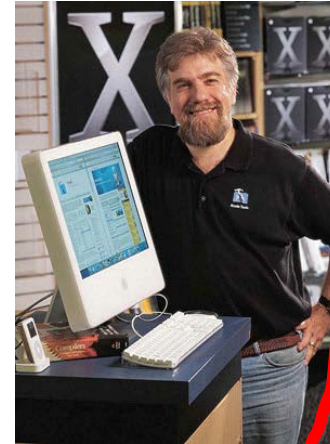
Sam Ng



Dan Carlin



Ted Golstein



Evan Paull



James Durbin



Chris Szeto



Artem Sokolov



Daniel Sam

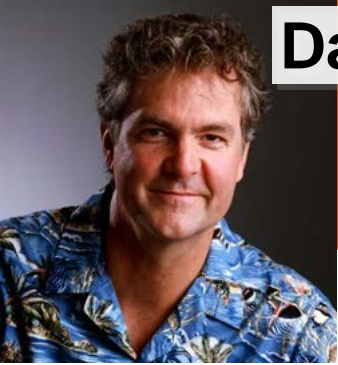


Chris Wong



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





Acknowledgments

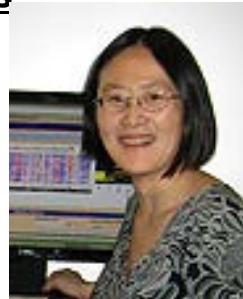
Chris Benz,



UCSC Cancer Genomics

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

Jing Zhu



UCSC Genome Browser Staff

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

Buck Institute for Aging

- Christina Yau
- Sean Mooney
- Janita Thusberg

Collaborators

- Laura Esserman, UCSF
- Joe Gray, LBL
- Laura Heiser, LBL
- Eric Collisson, UCSF

Funding Agencies

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

Broad Institute

- Gaddy Getz
- Mike Noble
- Daniel DeCara

