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

Josh Stuart, UC Santa Cruz

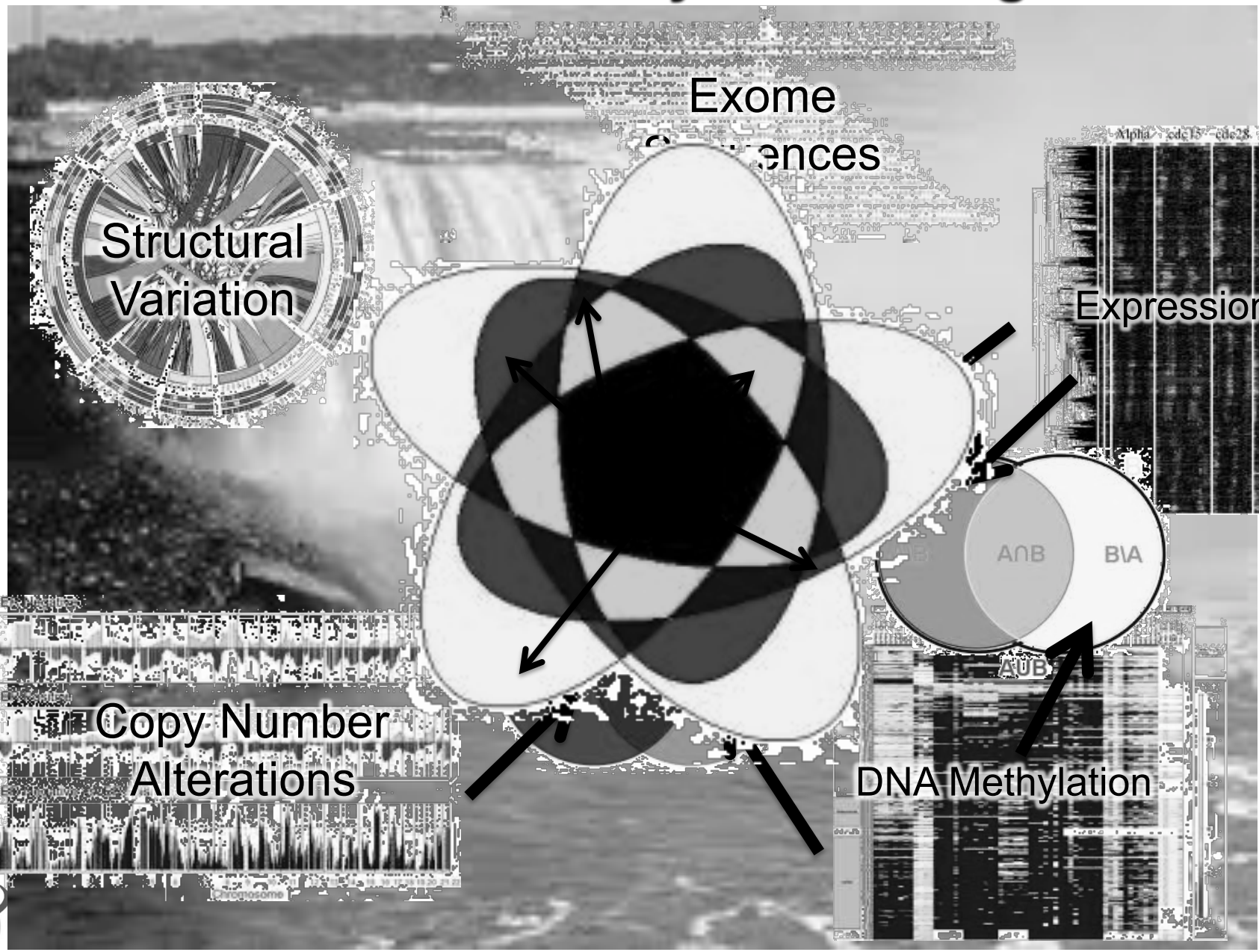
AACR-NCI-EORTC International Conference

San Francisco, CA, Nov 15, 2011

Disclosures

- SAB and stock owner of Five^3 Genomics.
- All work presented part of TCGA consortium.

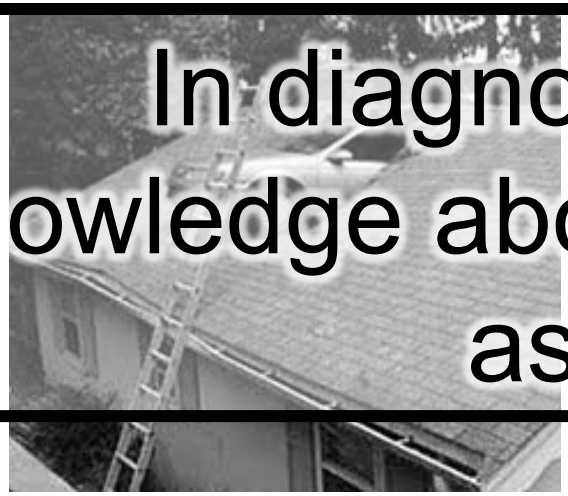
Flood of Data Analysis Challenges

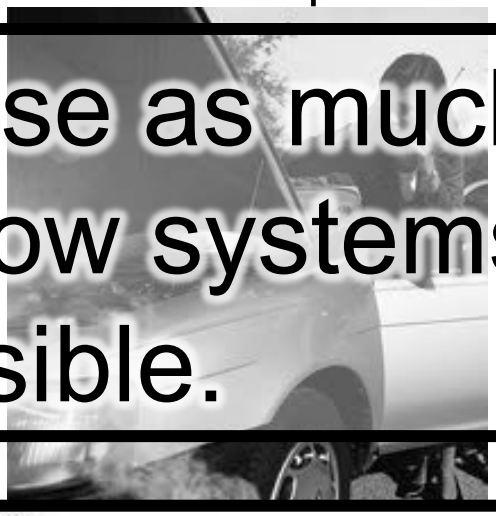


Analysis of disease samples like Investigating a series of different car breakdowns

Patient Sample 1

Patient Sample 2

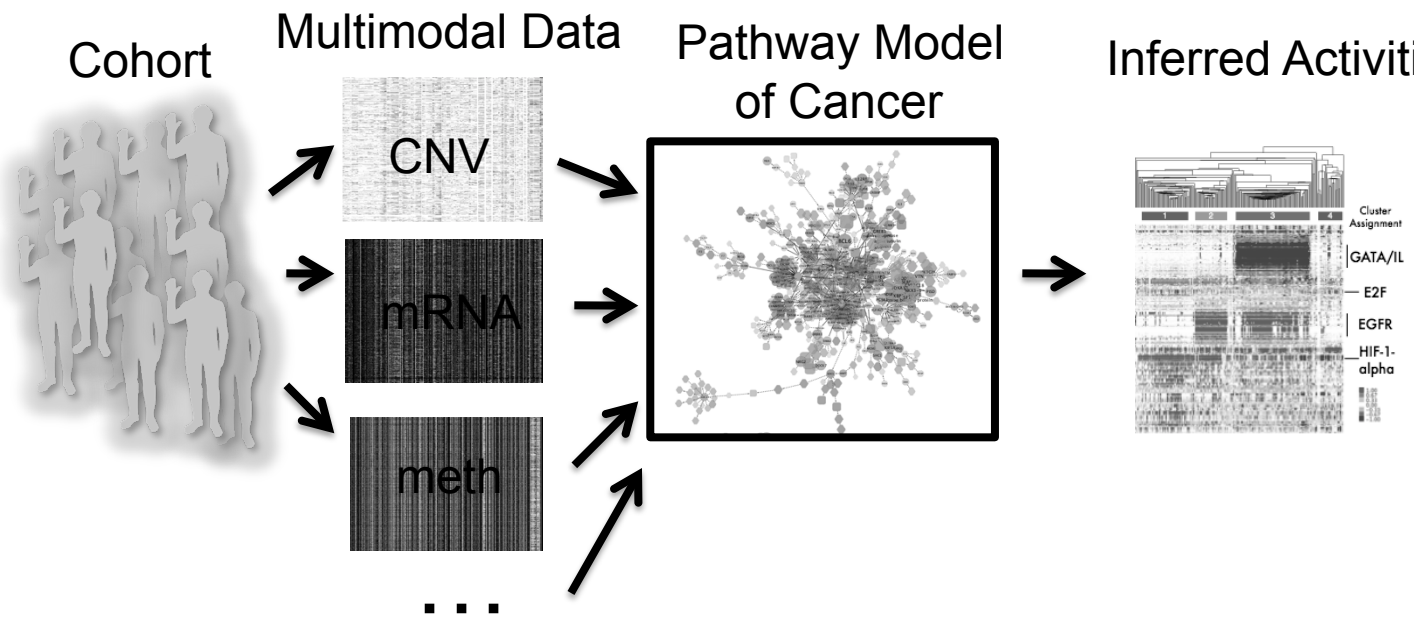




In diagnosis use as much knowledge about how systems work as possible.

ple *N*

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

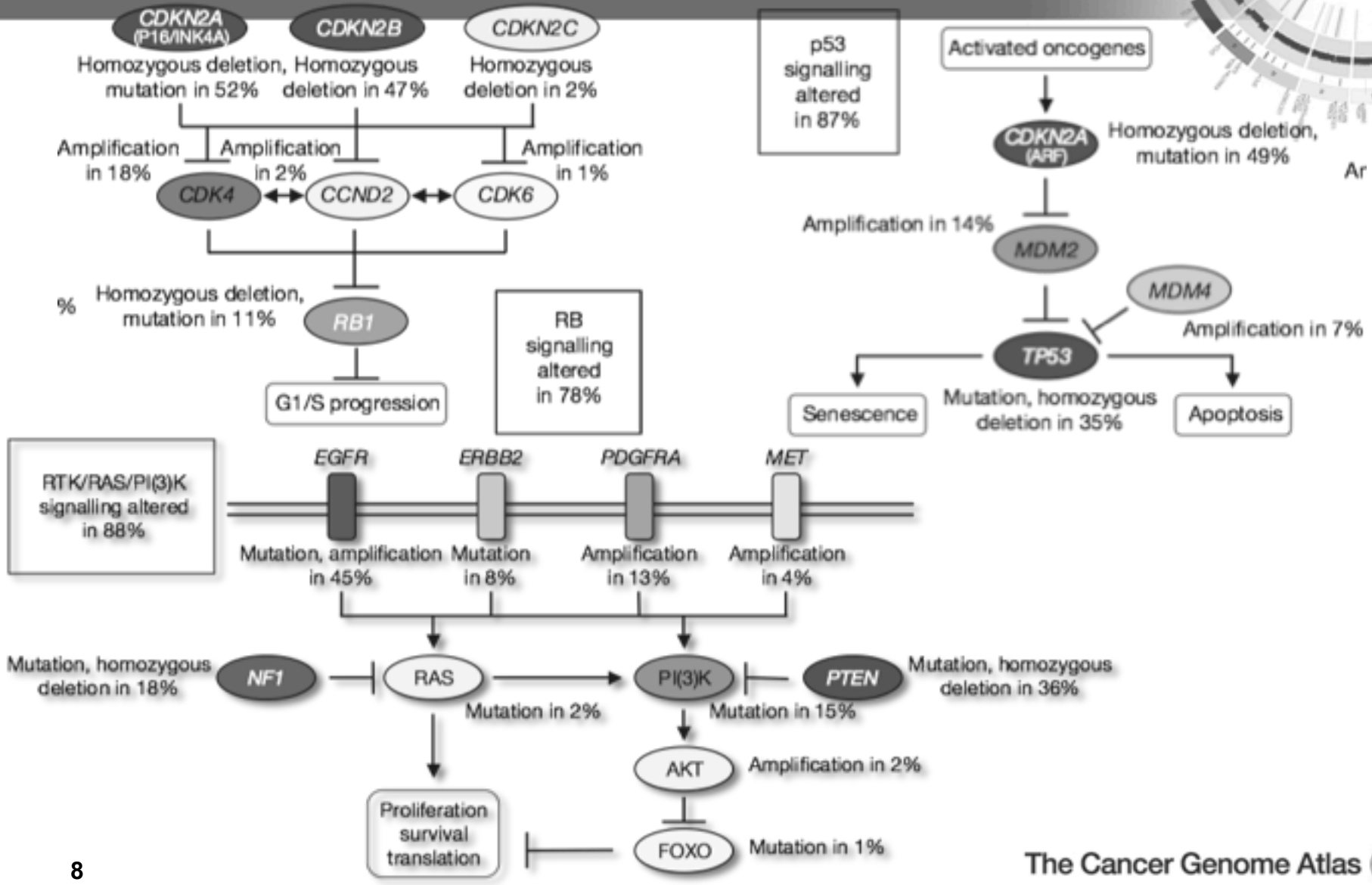
Outline

- Integrated Pathway Approach
- Application to find Pathway Biomarkers of Cancer
- Application to predict impact of mutations
- Pan-Cancer initial look

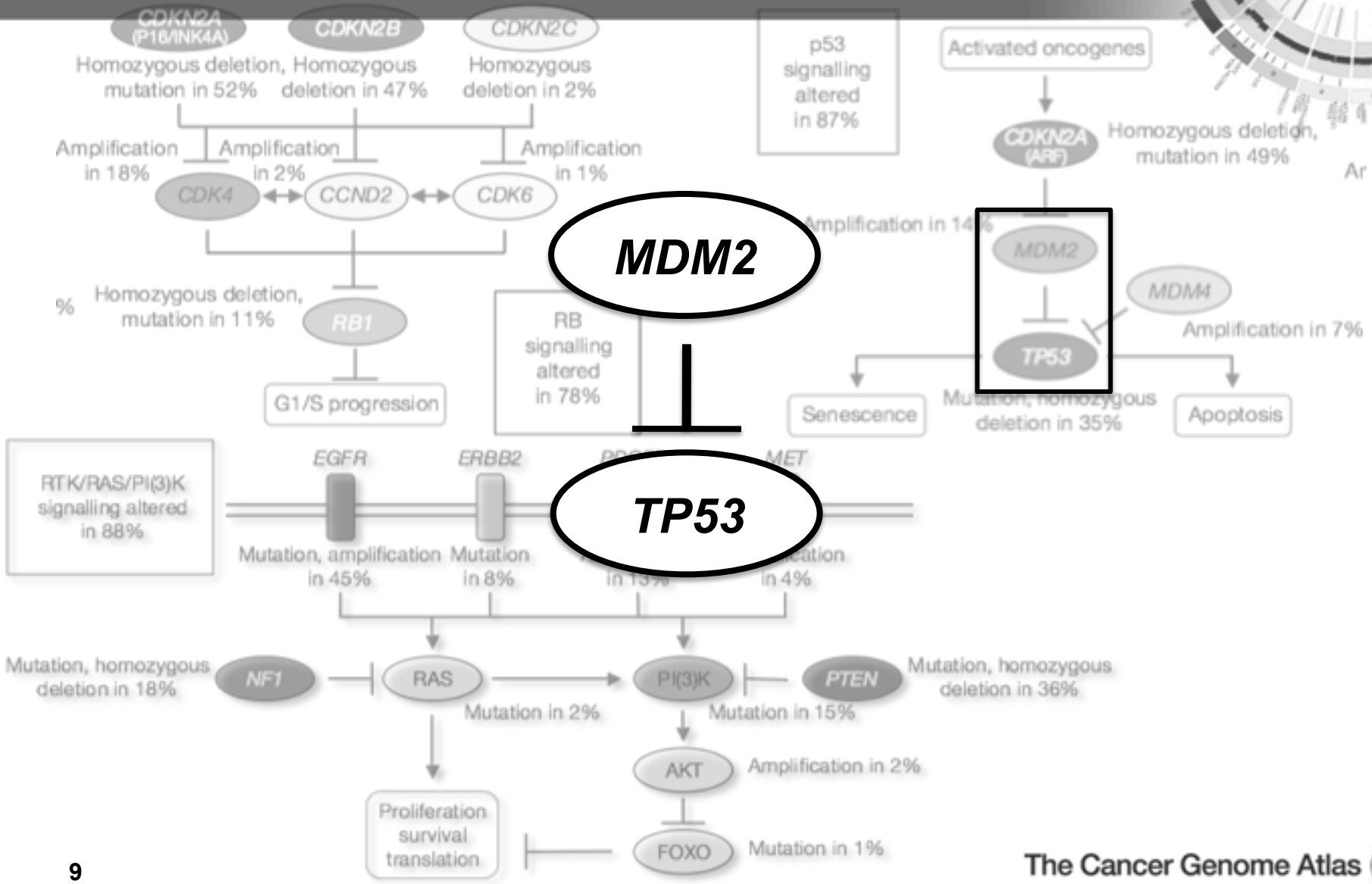
Outline

- Integrated Pathway Approach
- Application to find Pathway Biomarkers of Cancer
- Application to predict impact of mutations
- Pan-Cancer initial look

Main Approach: Detailed models of gene expression and interaction

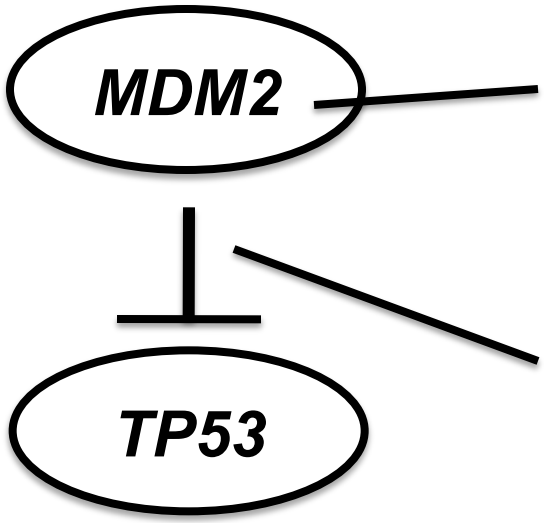


Main Approach: Detailed models of gene expression and interaction



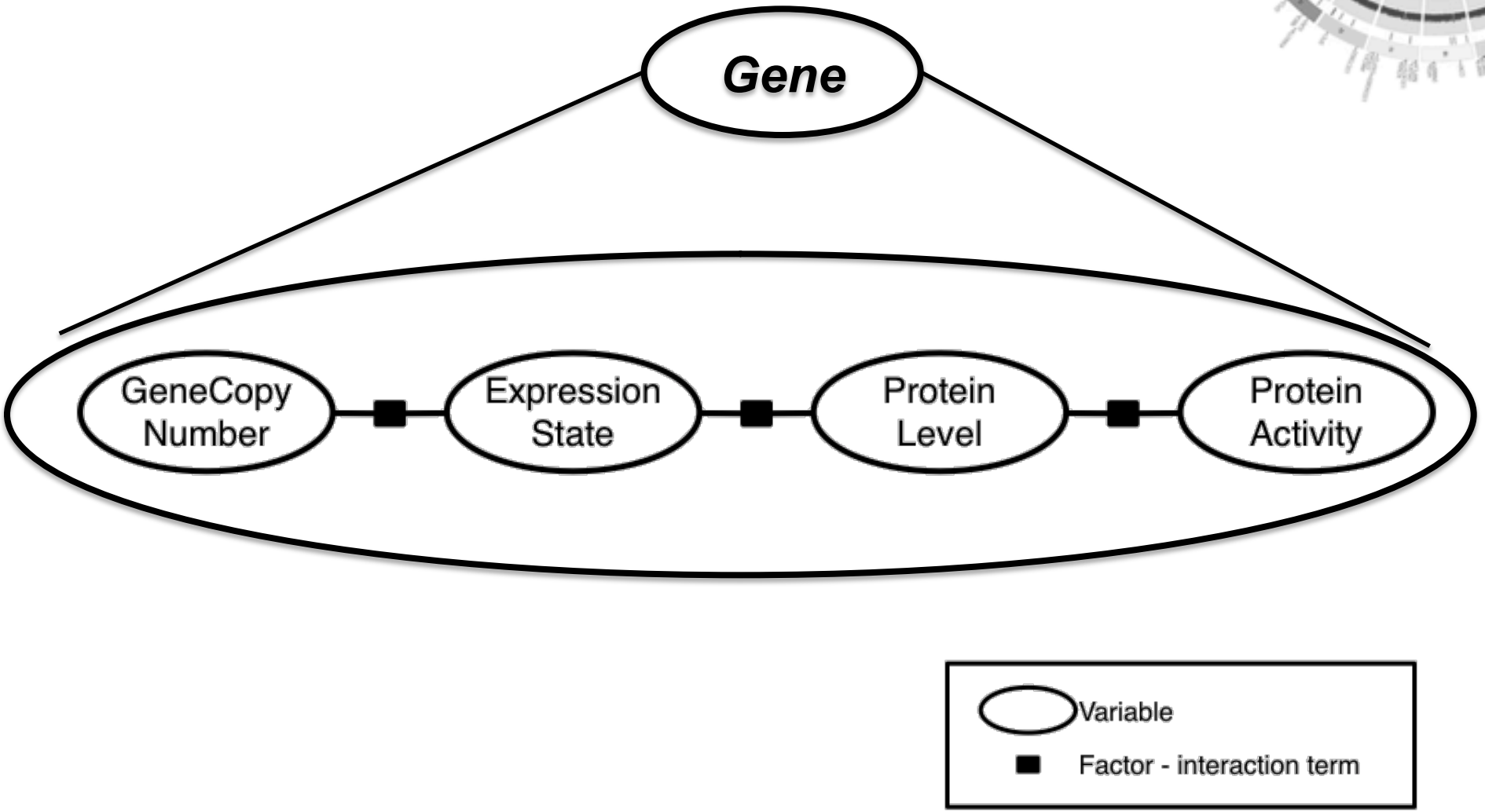
Main Approach: Detailed models of expression and interaction

Two Parts:



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

Pathway Recognition Algorithm Using Data Integration on Genomic Models (PARADIGM)

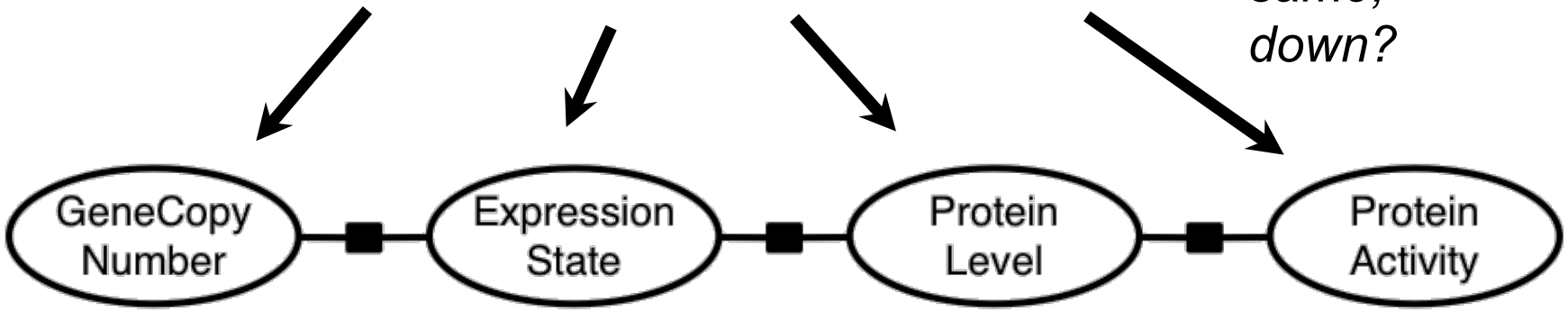


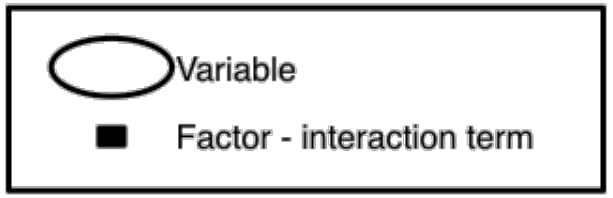
Vaske et al. 2010. *Bioinformatics*

PARADIGM
Gene Model to Integrate Data

relative to non-cancer,
is this sample:
up,
same,
down?

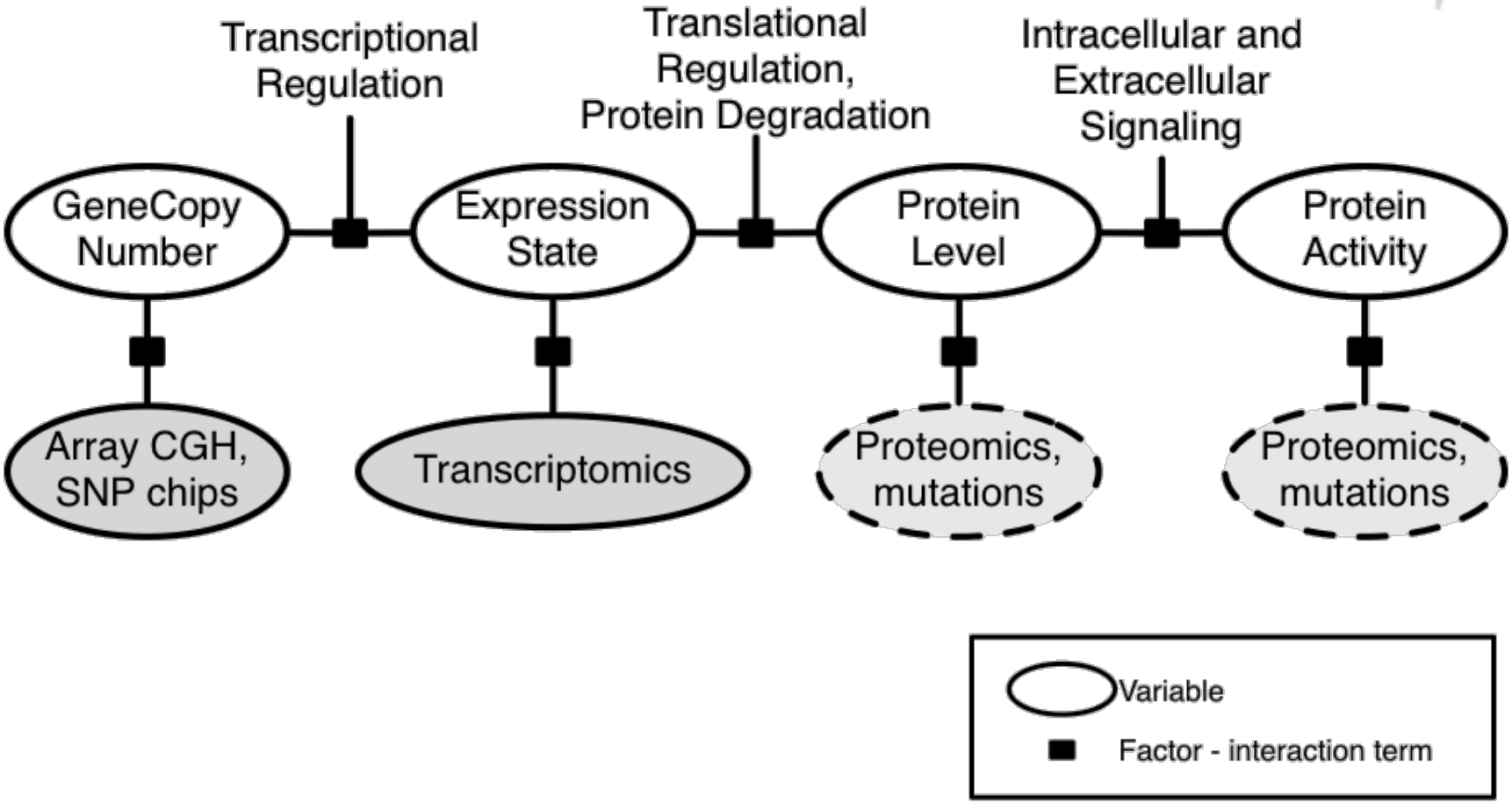
3-state discrete variables





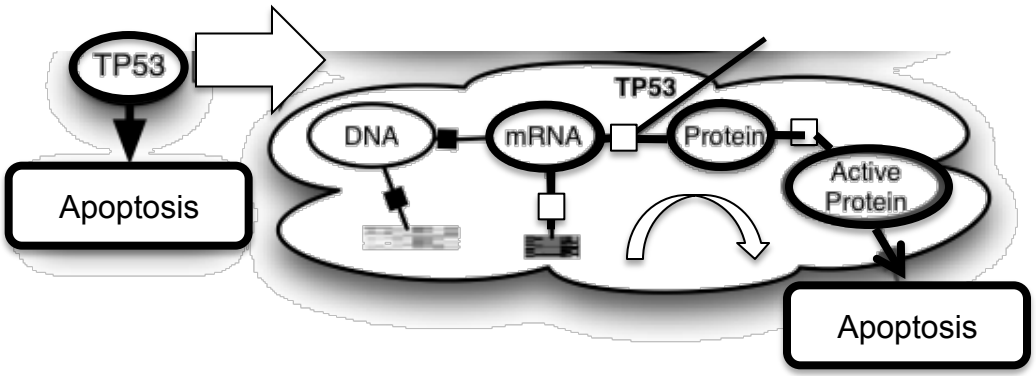
Vaske et al. 2010. *Bioinformatics*

PARDIGM Gene Model to Integrate Data



Vaske et al. 2010. *Bioinformatics*

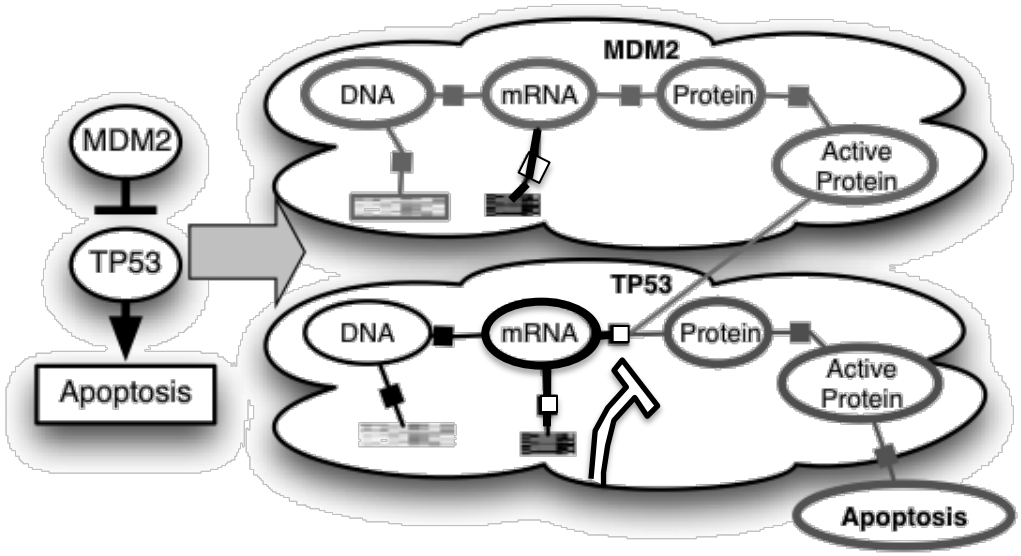
Interactions Matter



- Given information about the expression of TP53 alone
- Reasoning predicts apoptosis is in tact in these cells.

Vaske et al. 2010. *Bioinformatics*

Interactions Matter



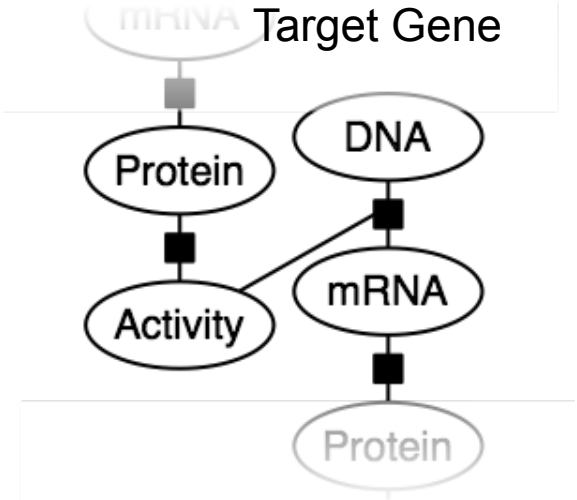
- Given the interaction and data about MDM2.
- apoptosis inference reversed

Vaske et al. 2010. *Bioinformatics*

PARADIGM Interaction Components

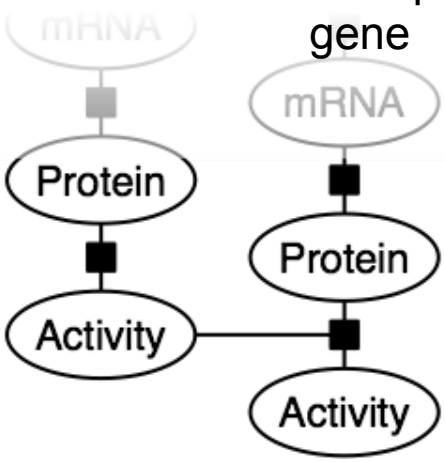
Transcription Factor

Target Gene



Kinase

Phosphorylated gene

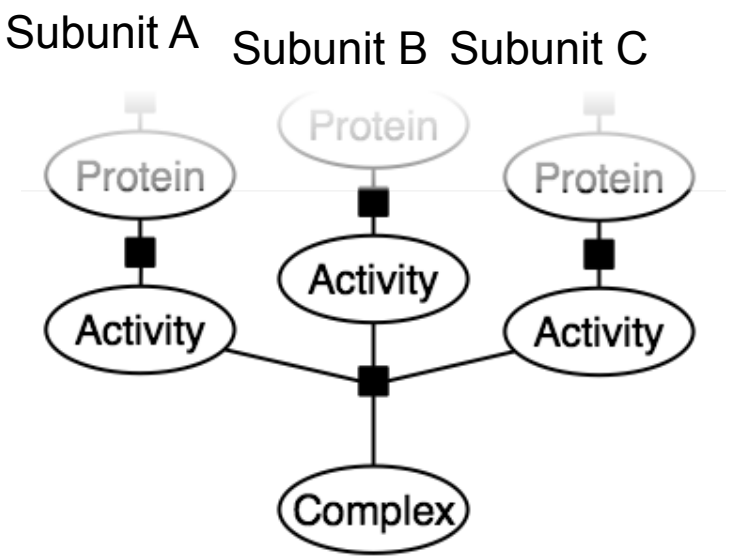


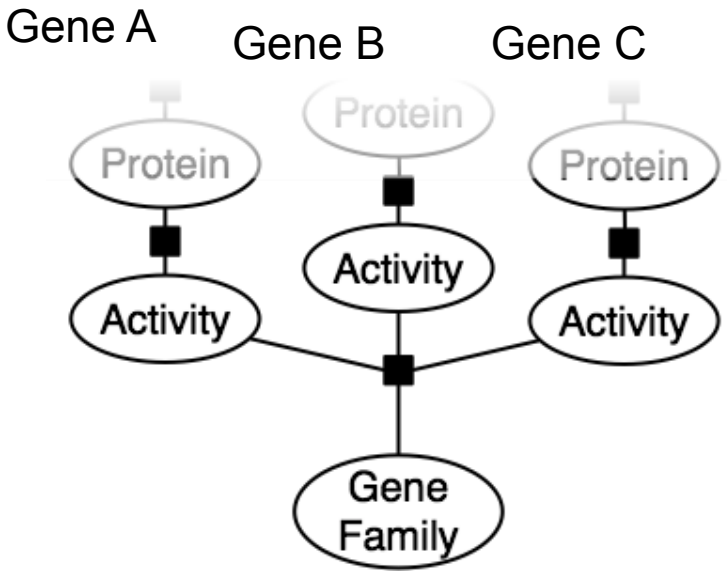
Transcriptional Regulation

Post-translational Modification

Vaske et al. 2010. *Bioinformatics*

PARADIGM Interaction Components



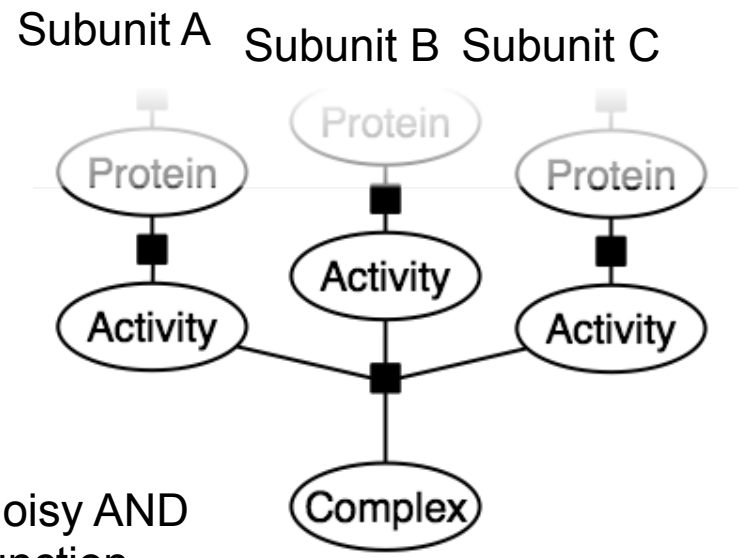


Protein Complex

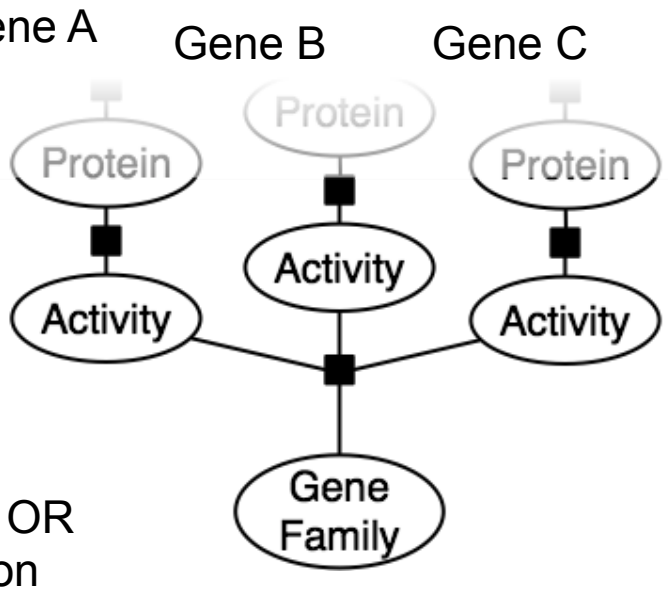
Gene family, proteins with interchangeable function

Vaske et al. 2010. *Bioinformatics*

PARADIGM Interaction Components



Noisy AND function



Noisy OR function

Protein Complex

Gene family, proteins with interchangeable function

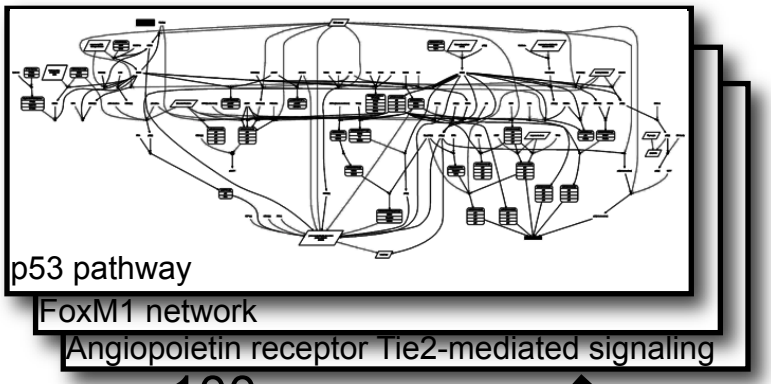
Vaske et al. 2010. *Bioinformatics*

Outline

- Integrated Pathway Approach
- Application to find Pathway Biomarkers of Cancer
- Application to predict impact of mutations
- Pan-Cancer initial look

Pathway Interpretation of Omics Data

PARADIGM Pathway Analysis



~100 Pathways

NATIONAL CANCER INSTITUTE nature PathwayInteractionDatabase

316 TCGA Ovarian Samples

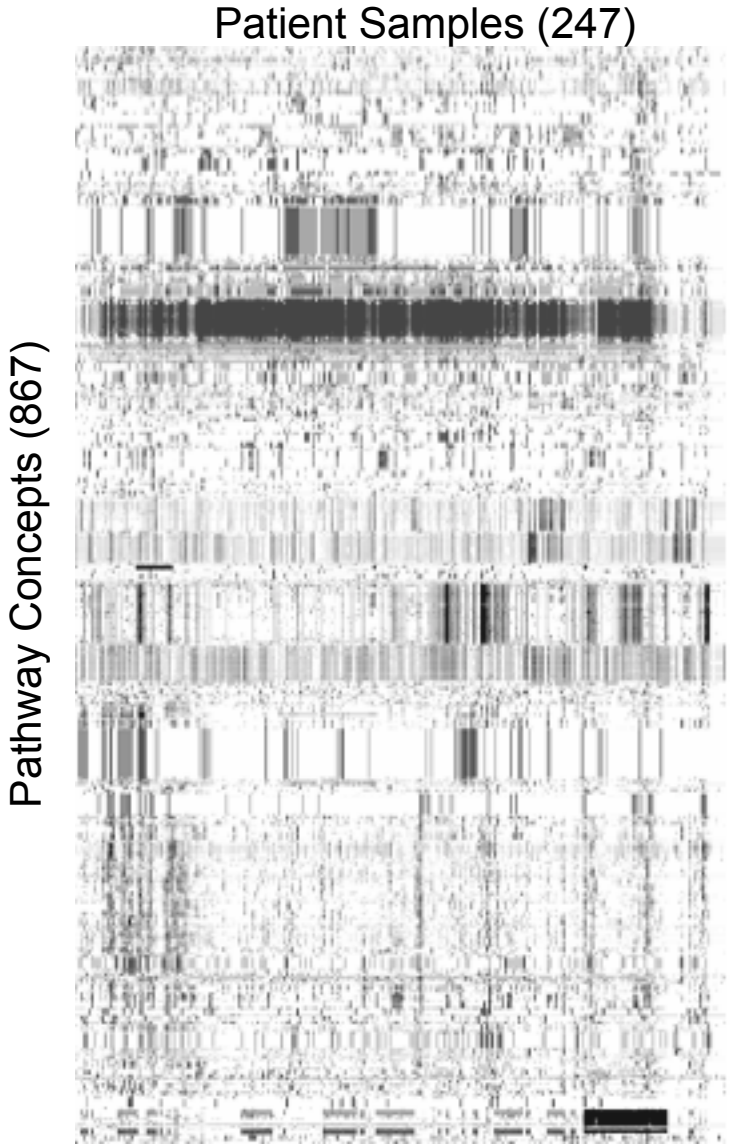
Per-sample integrated pathway levels

Tumor Samples

Concepts

TCGA Network. 2011. *Nature*

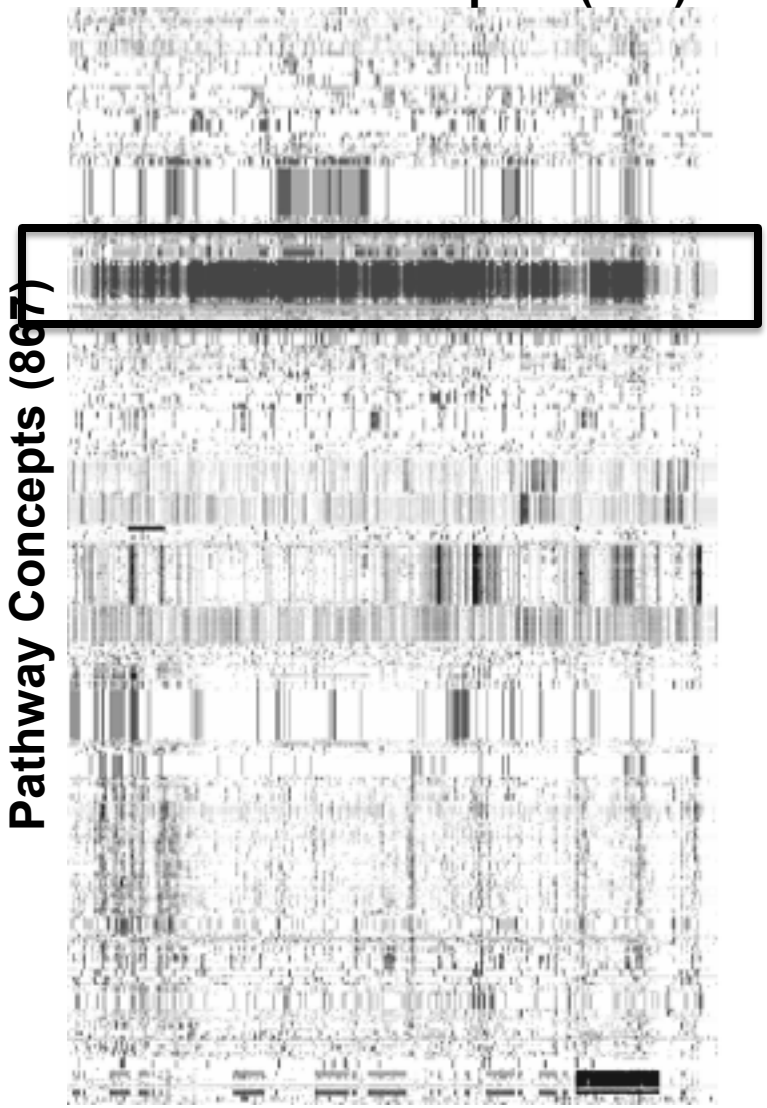
TCGA Ovarian Cancer Inferred Pathway Activities



TCGA Network. 2011. *Nature*

The Cancer Genome Atlas

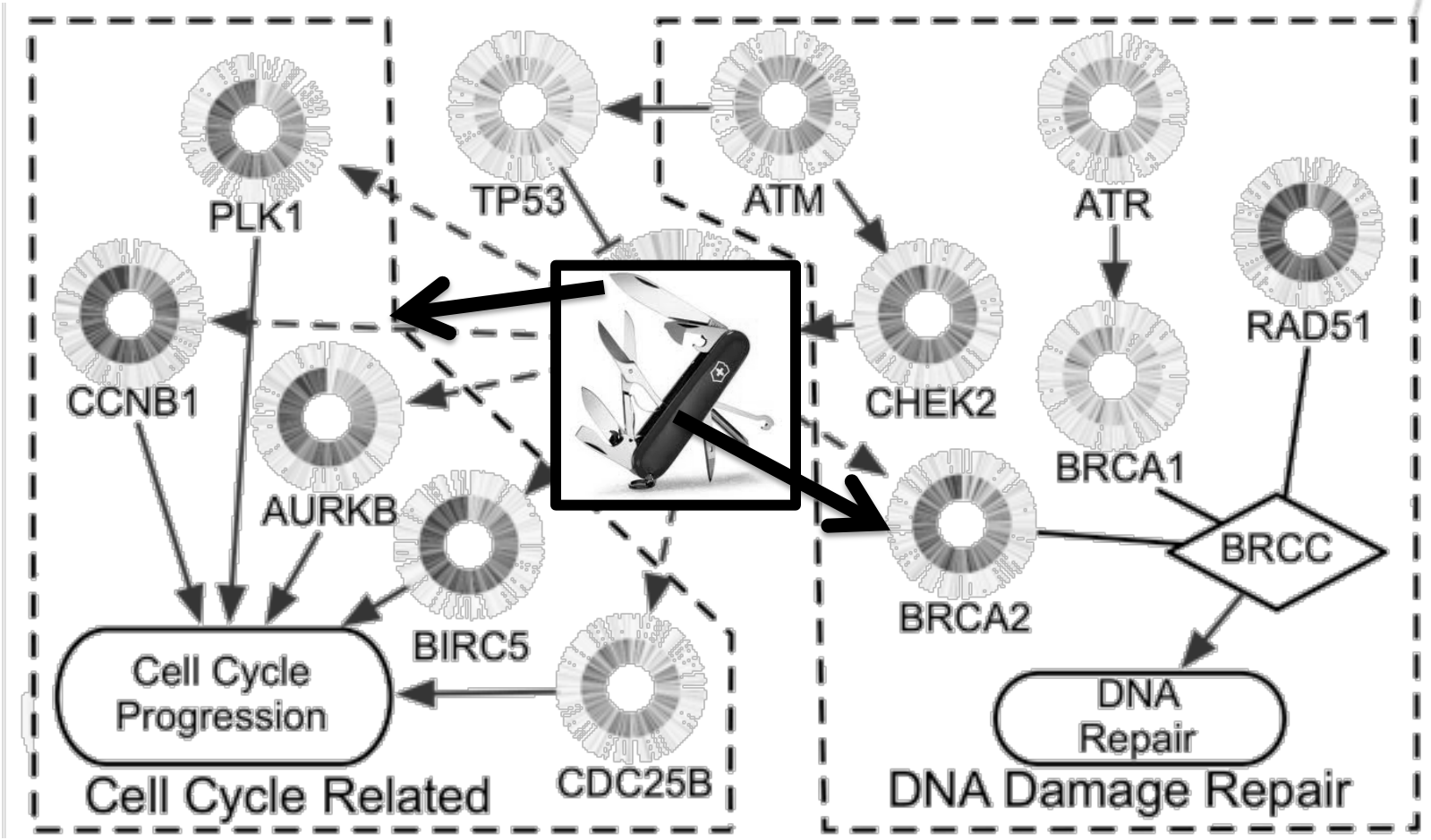
Ovarian: FOXM1 pathway altered in majority of serous ovarian tumors



FOXM1 Transcription Network

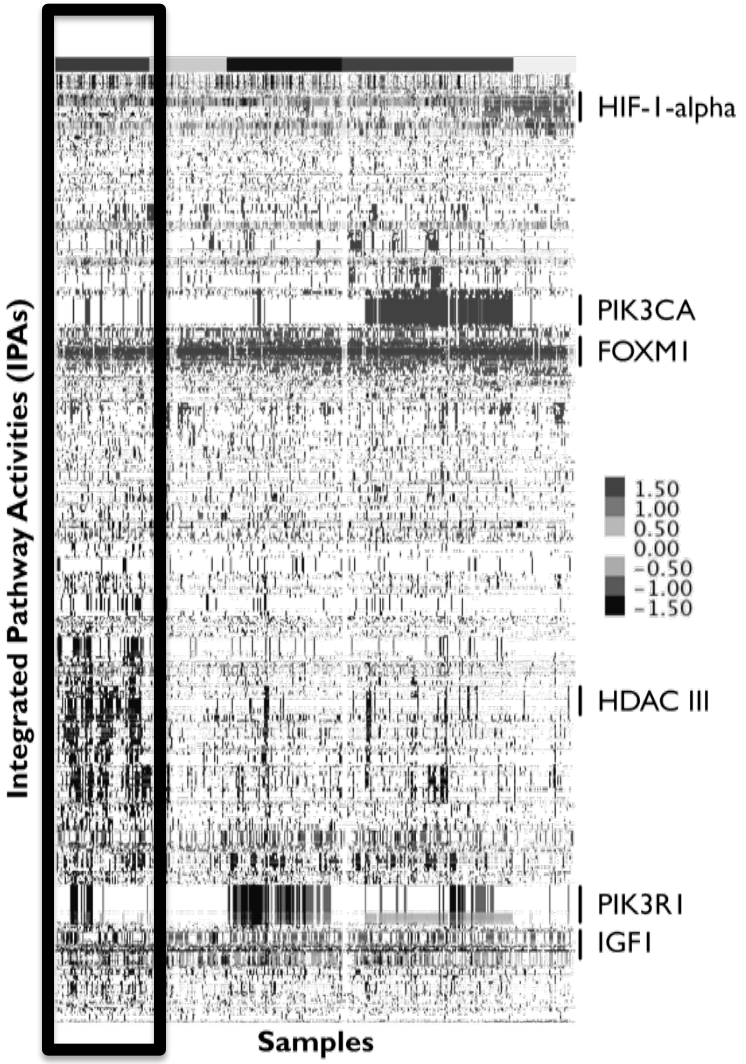
TCGA Network. 2011. *Nature*

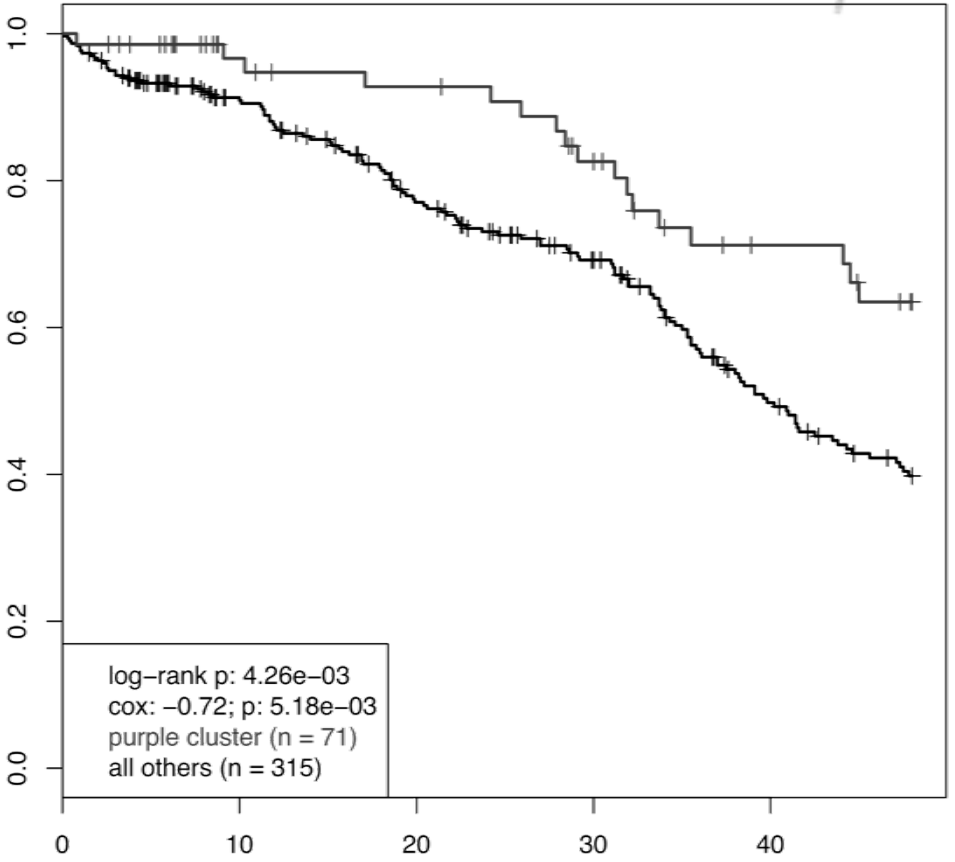
FOXM1 central to cross-talk between DNA repair and cell proliferation in Ovarian Cancer



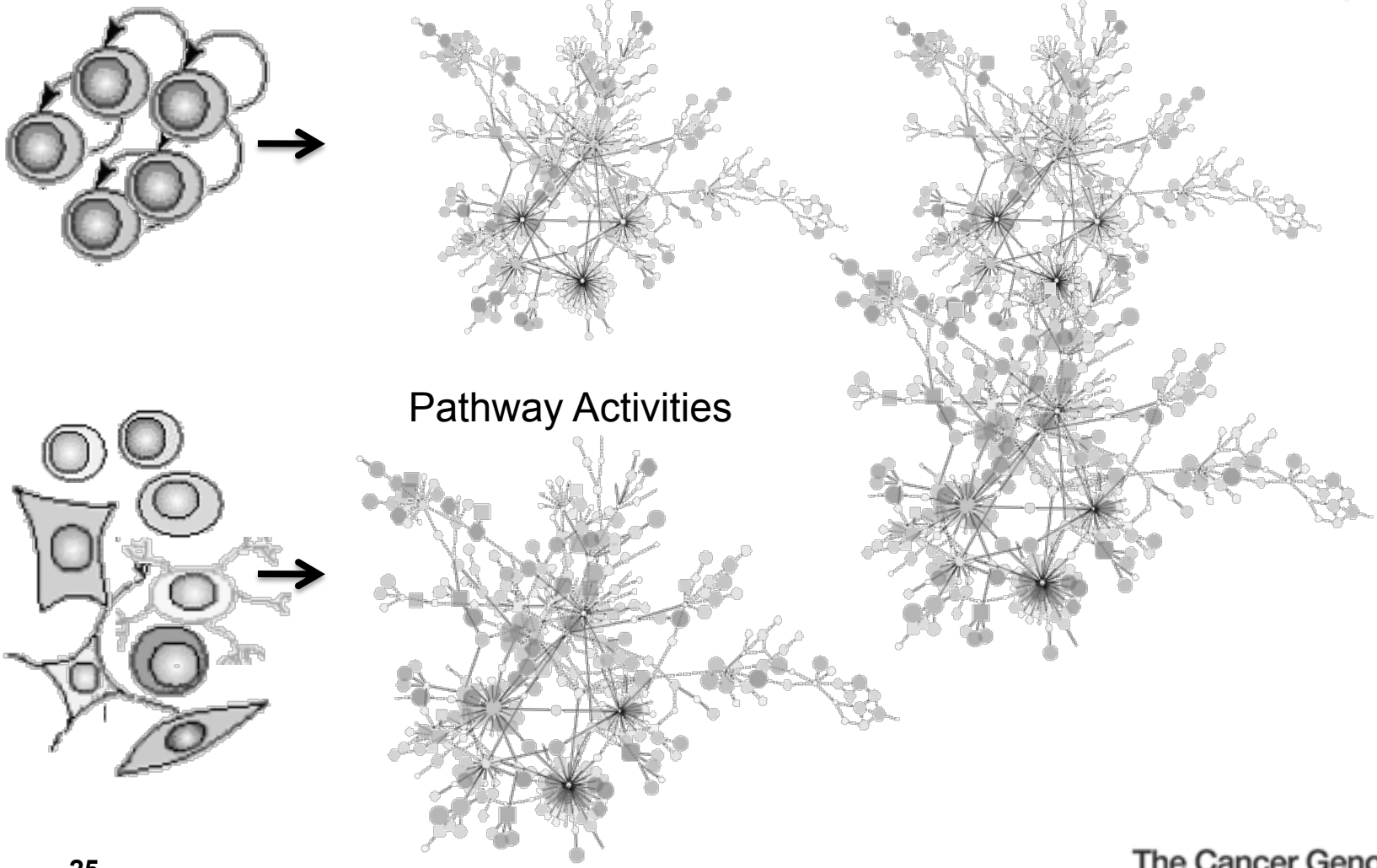
TCGA Network. *Nature* 2011

Ovarian: IPLs statify by survival time

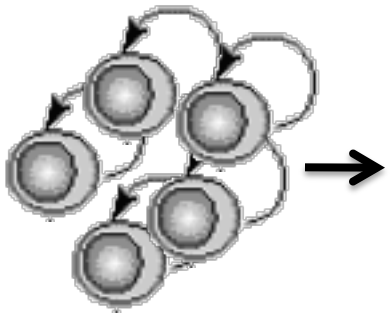




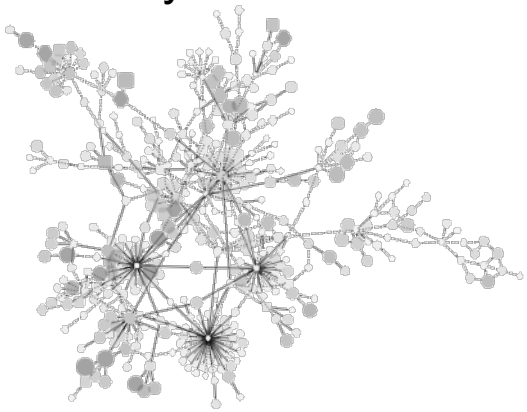
Pathway Signatures: Differential Subnetworks from a "SuperPathway"

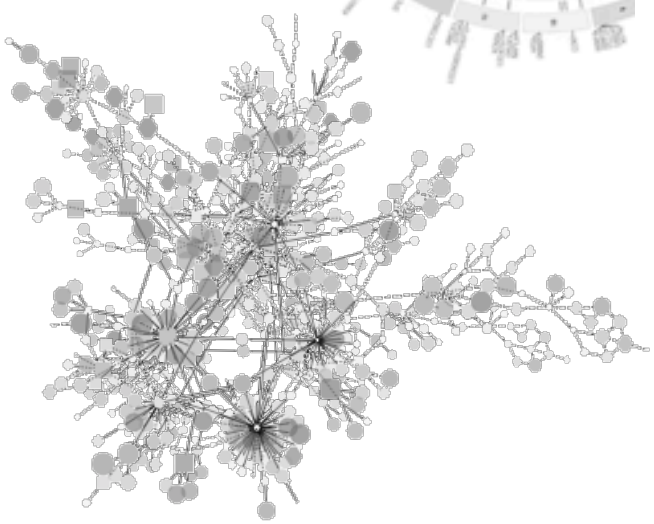


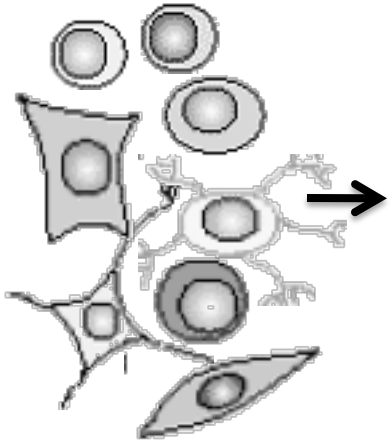
Pathway Signatures: Differential Subnetworks from a "SuperPathway"



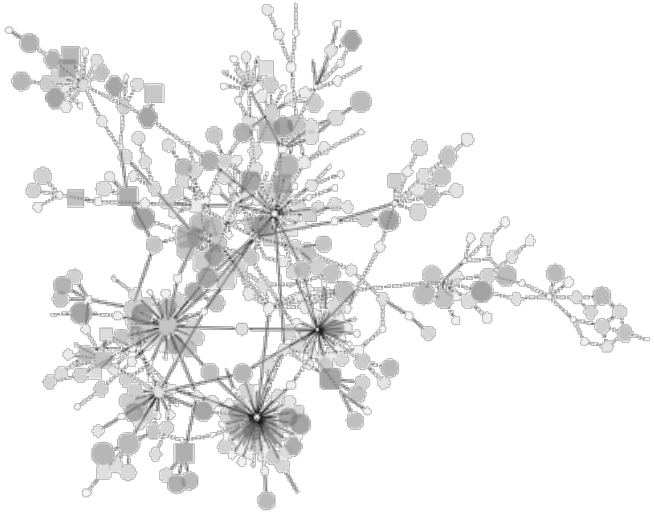
Pathway Activities



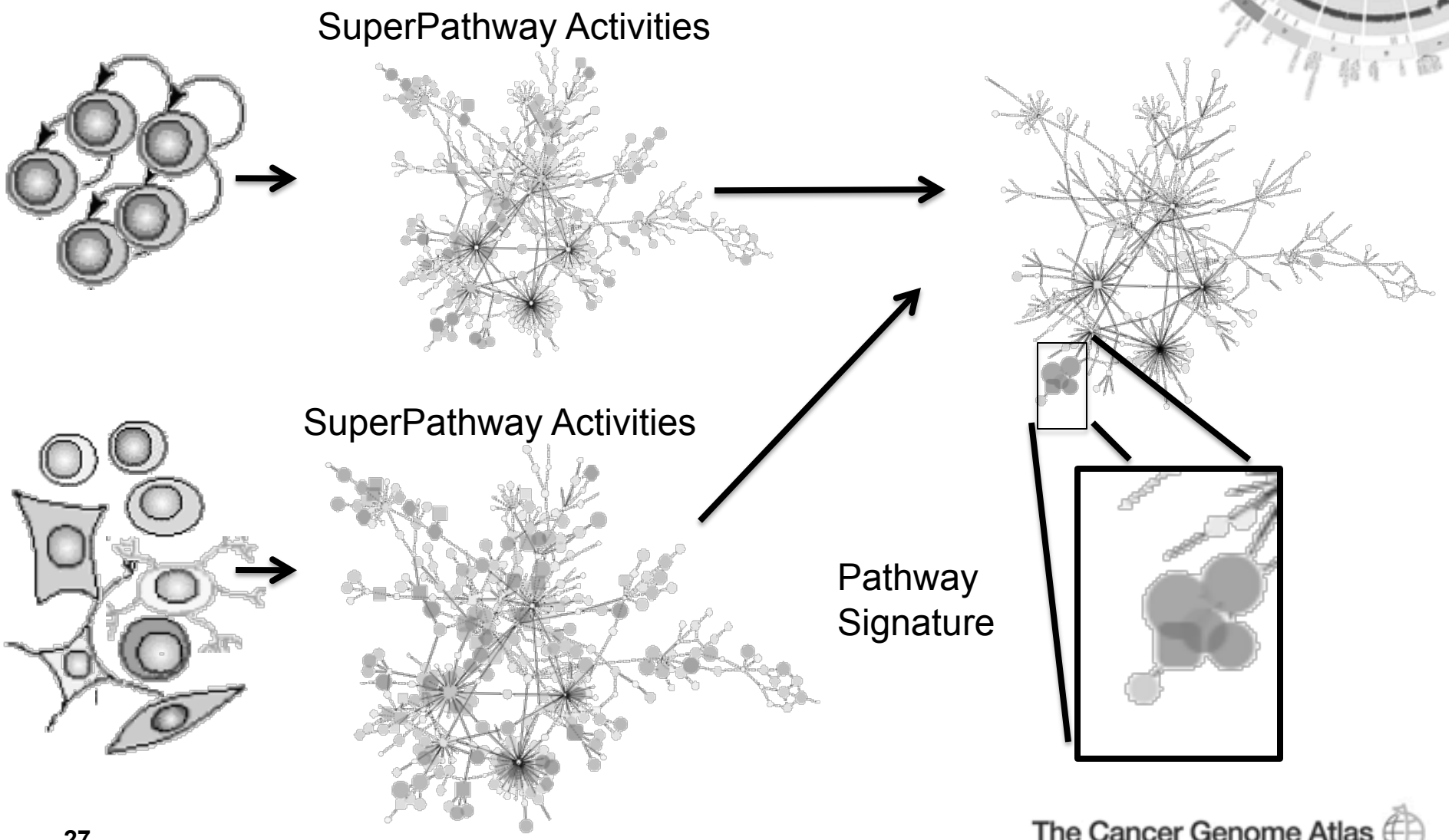




Pathway Activities



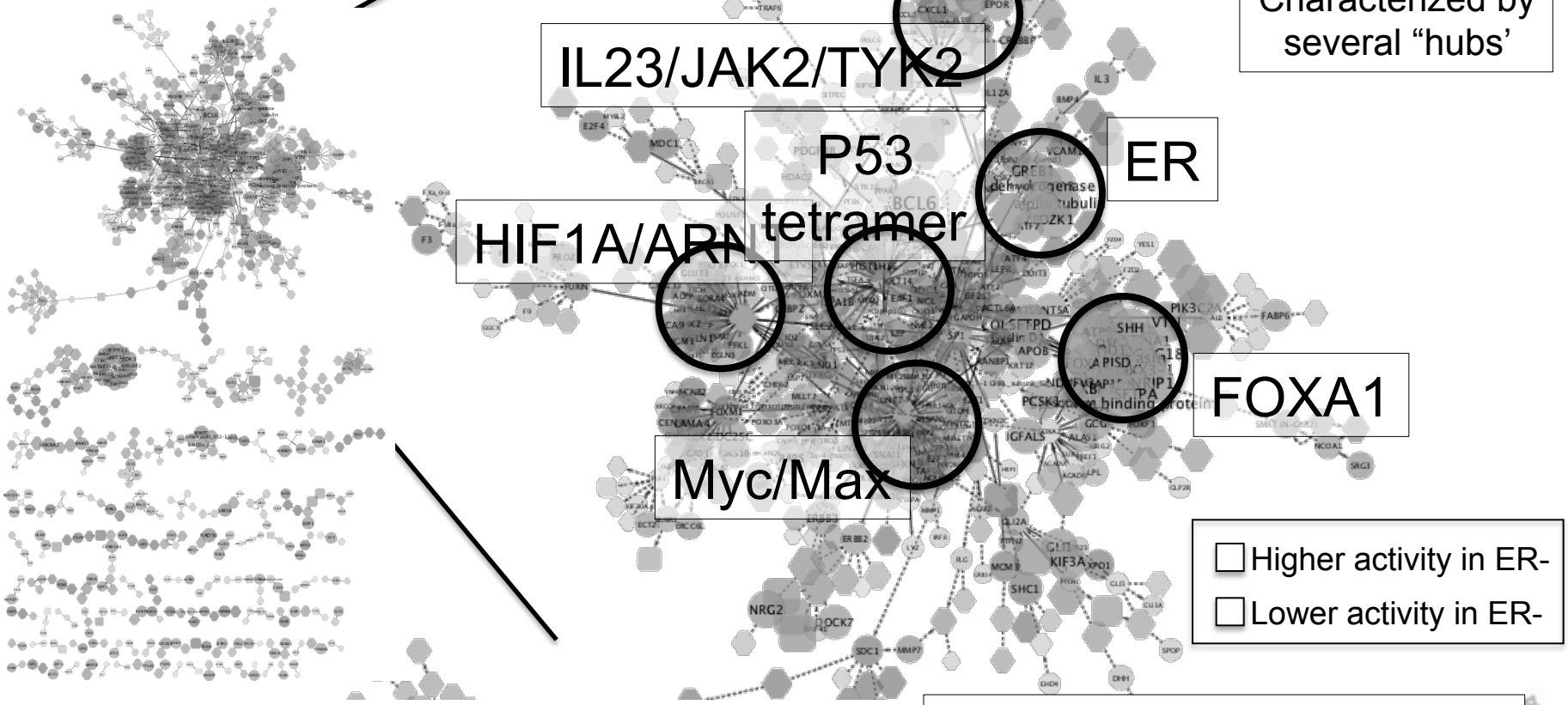
Pathway Signatures: Differential Subnetworks from a "SuperPathway"



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)



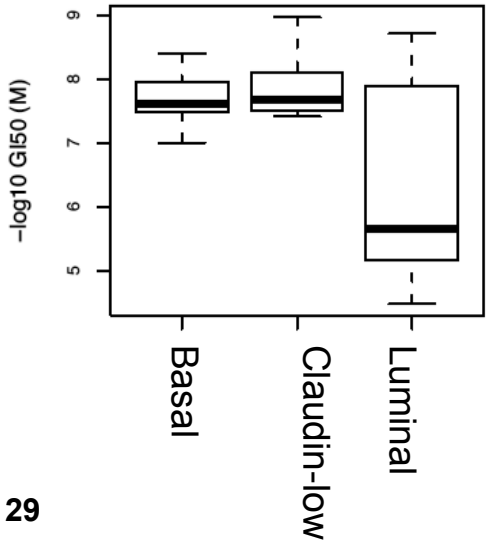
Characterized by several “hubs’

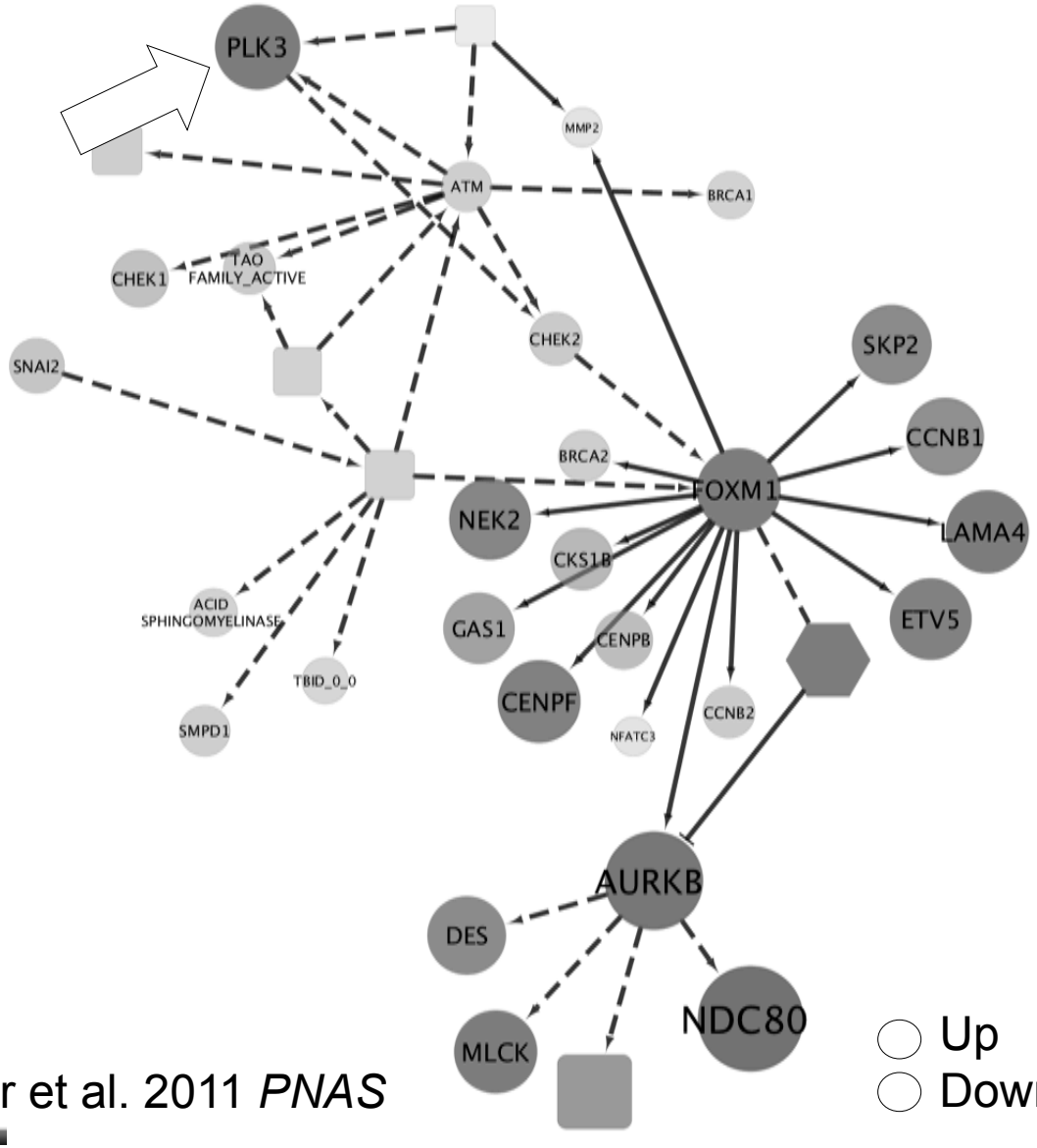
Sam Ng, Ted Goldstein

Networks can predict response to treatment: FOXM1/PLK/DNA Damage Network

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

GSK-PLKi

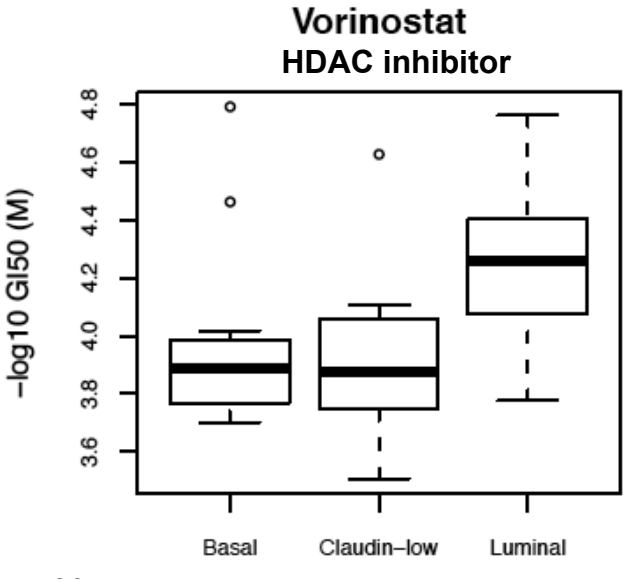


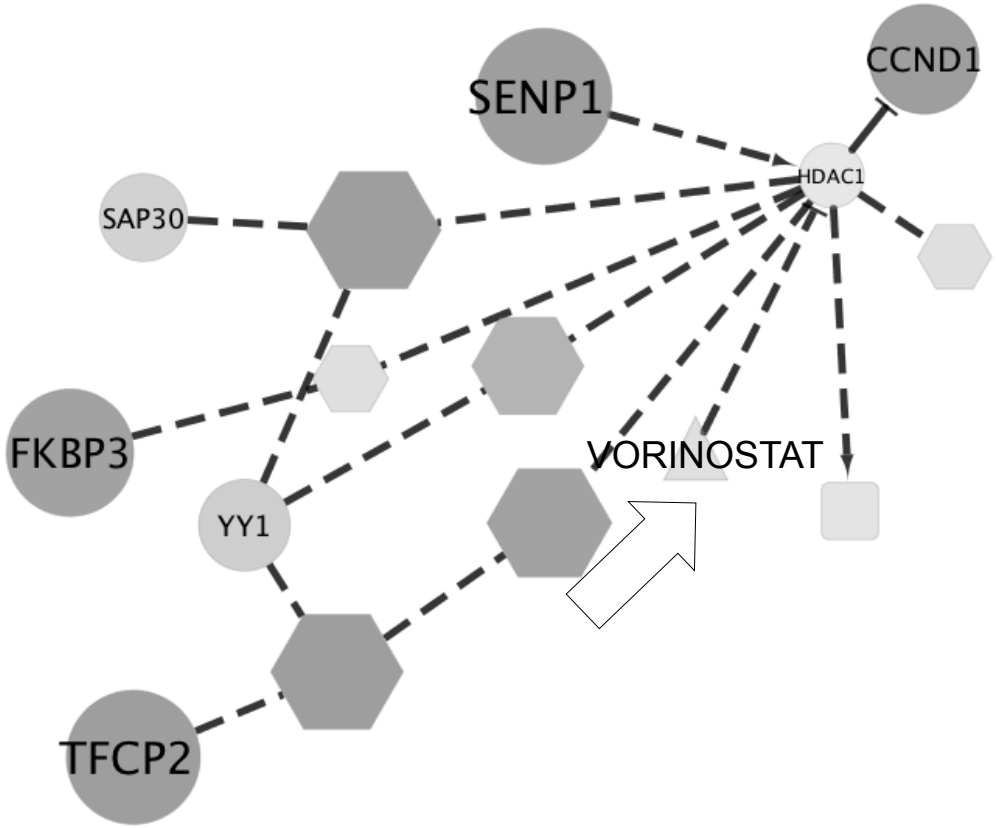


Heiser et al. 2011 *PNAS*

Networks can predict response to treatment: HDAC Network

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





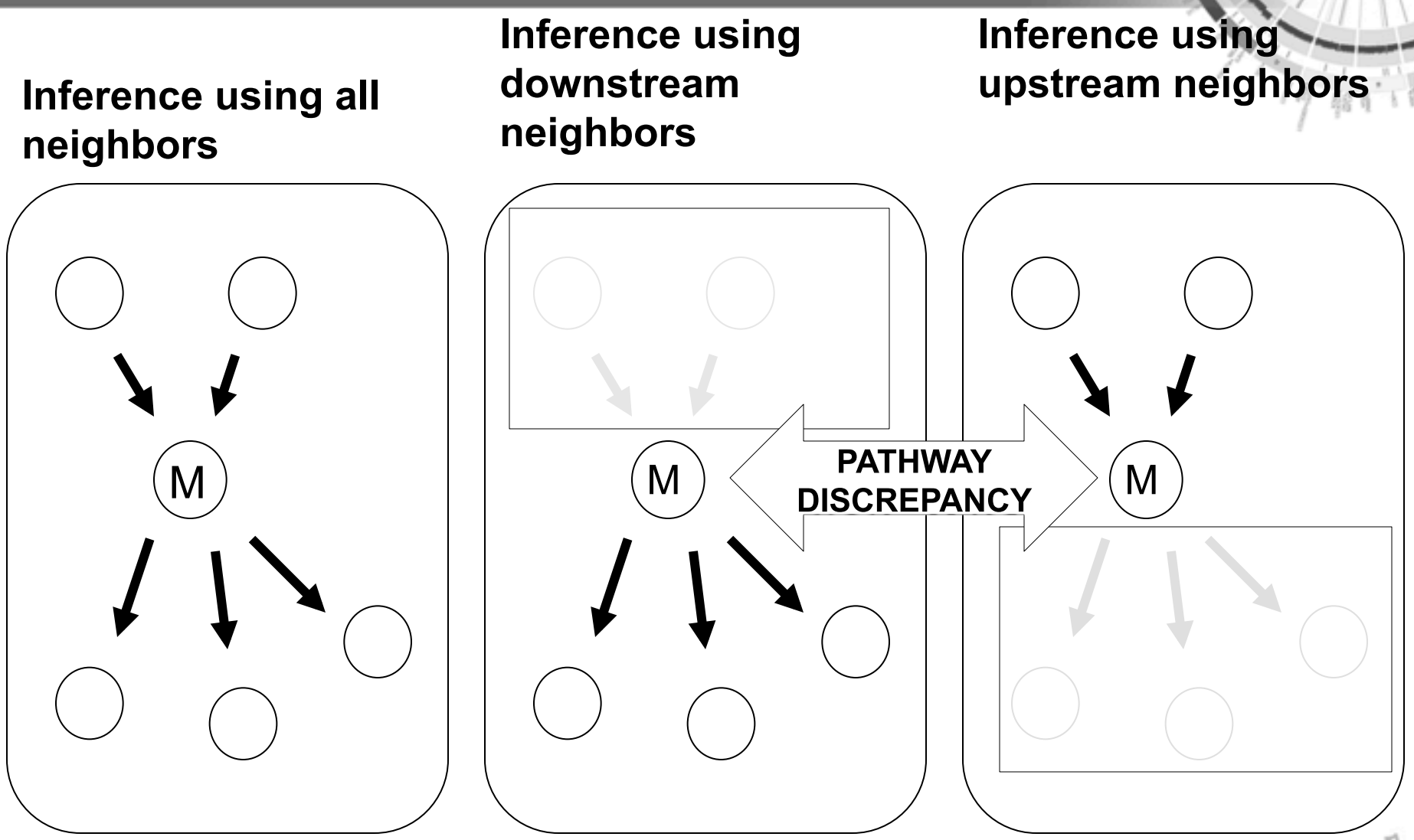
Heiser et al. 2011 *PNAS*

The Cancer Genome Atlas

Outline

- Integrated Pathway Approach
- Application to find Pathway Biomarkers of Cancer
- Application to predict impact of mutations
- Pan-Cancer initial look

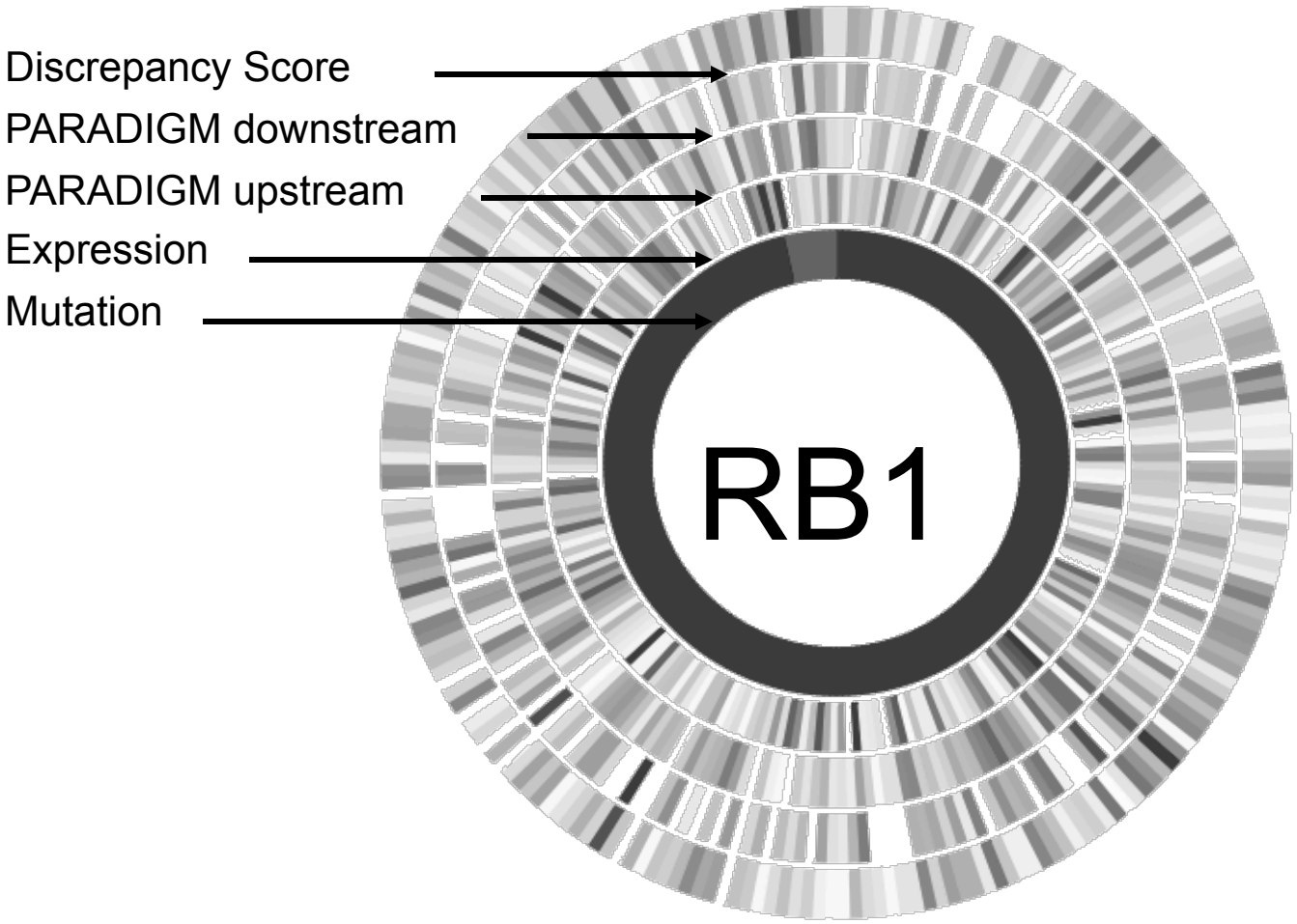
Predicting the Impact of Mutations On Genetic Pathways



Sam Ng

The Cancer Genome Atlas

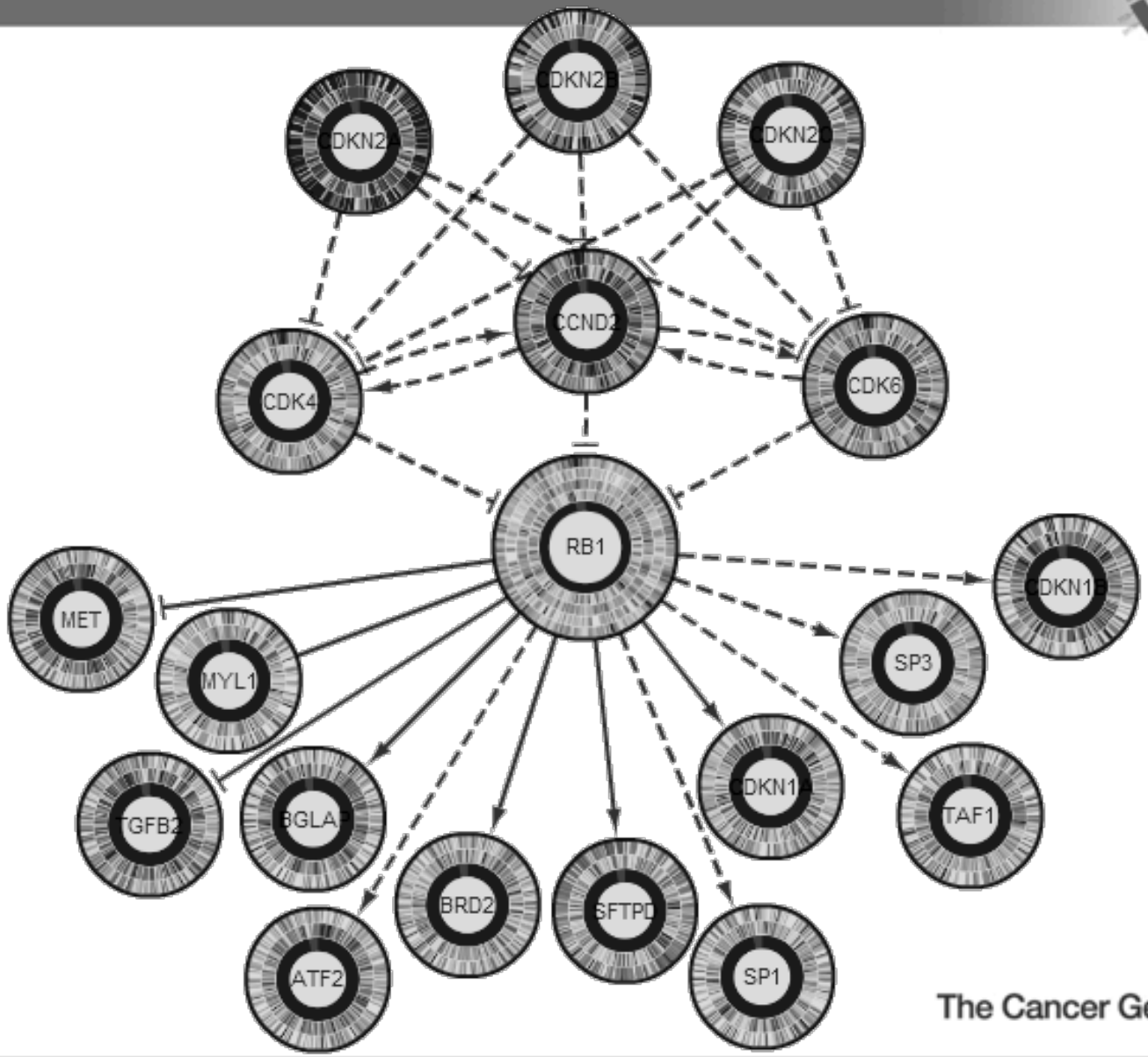
RB1 Loss-of-Function (GBM)



Sam Ng

The Cancer Genome Atlas

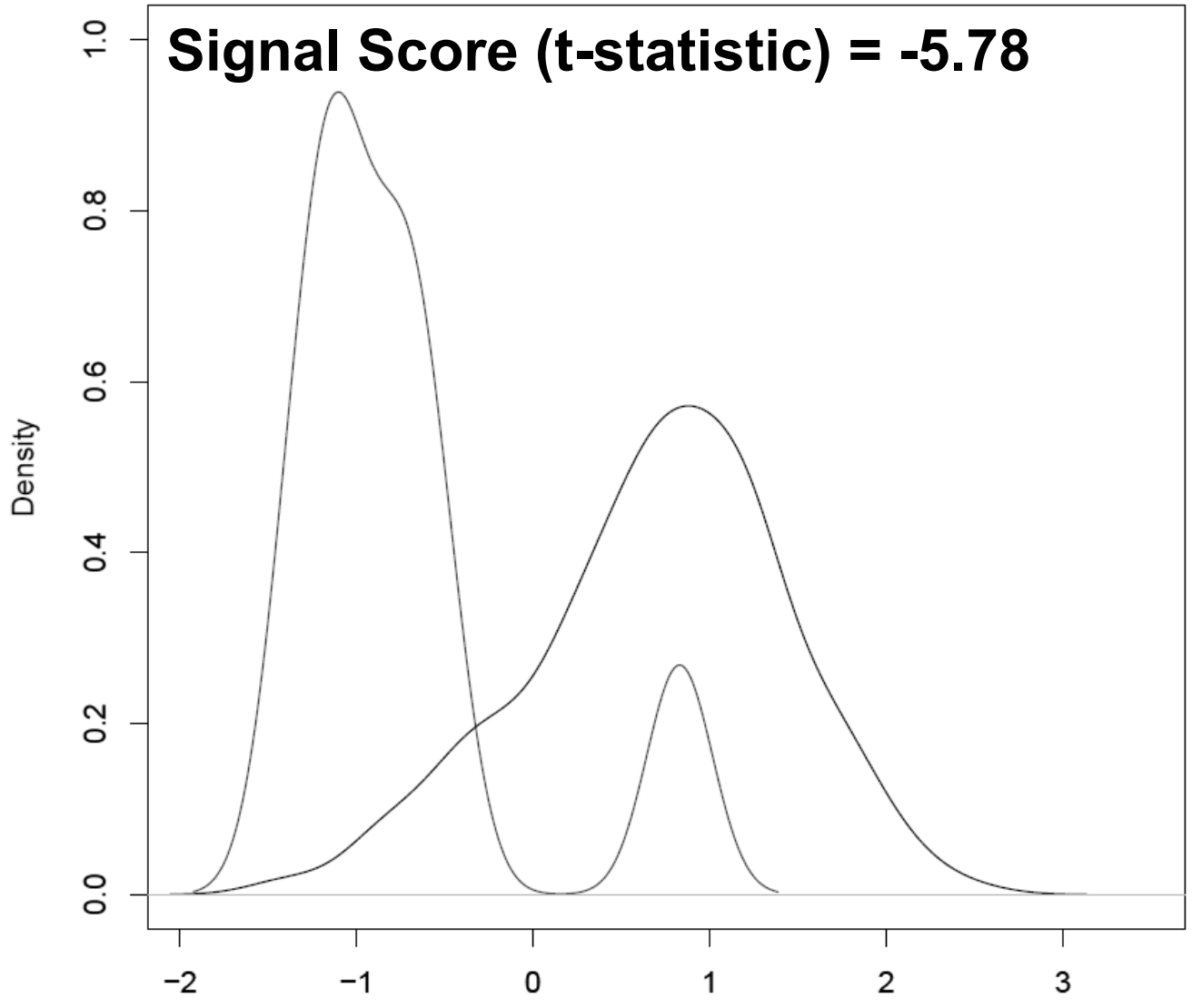
RB1 Network (GBM)



Sam Ng

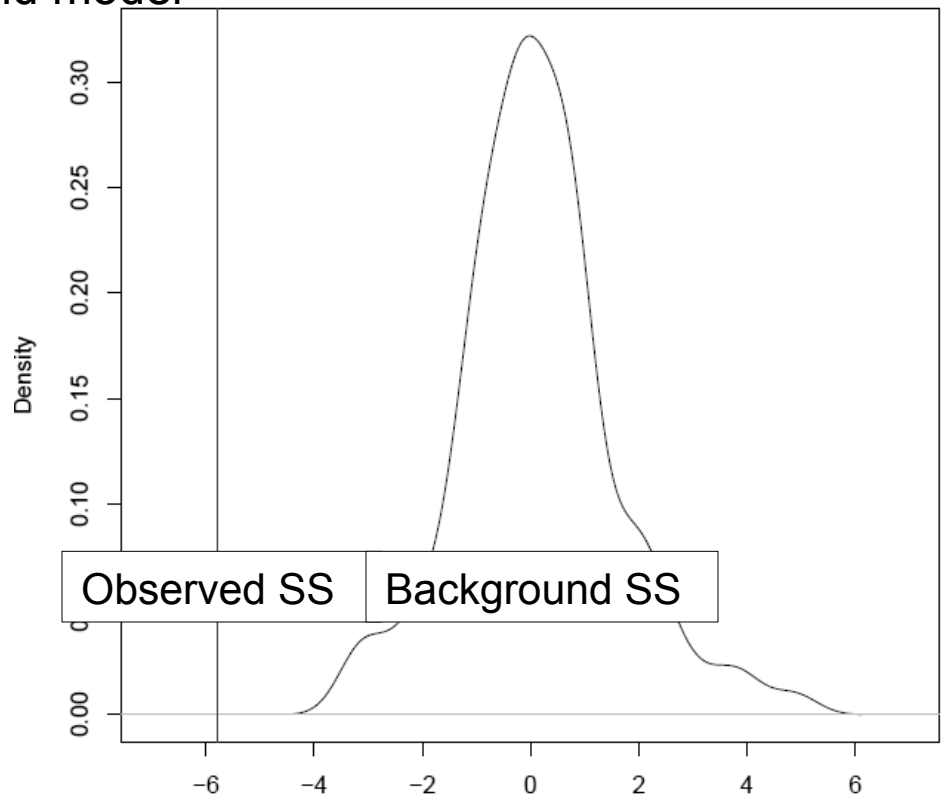
The Cancer Genome Atlas

RB1 Discrepancy Scores distinguish mutated vs non-mutated samples

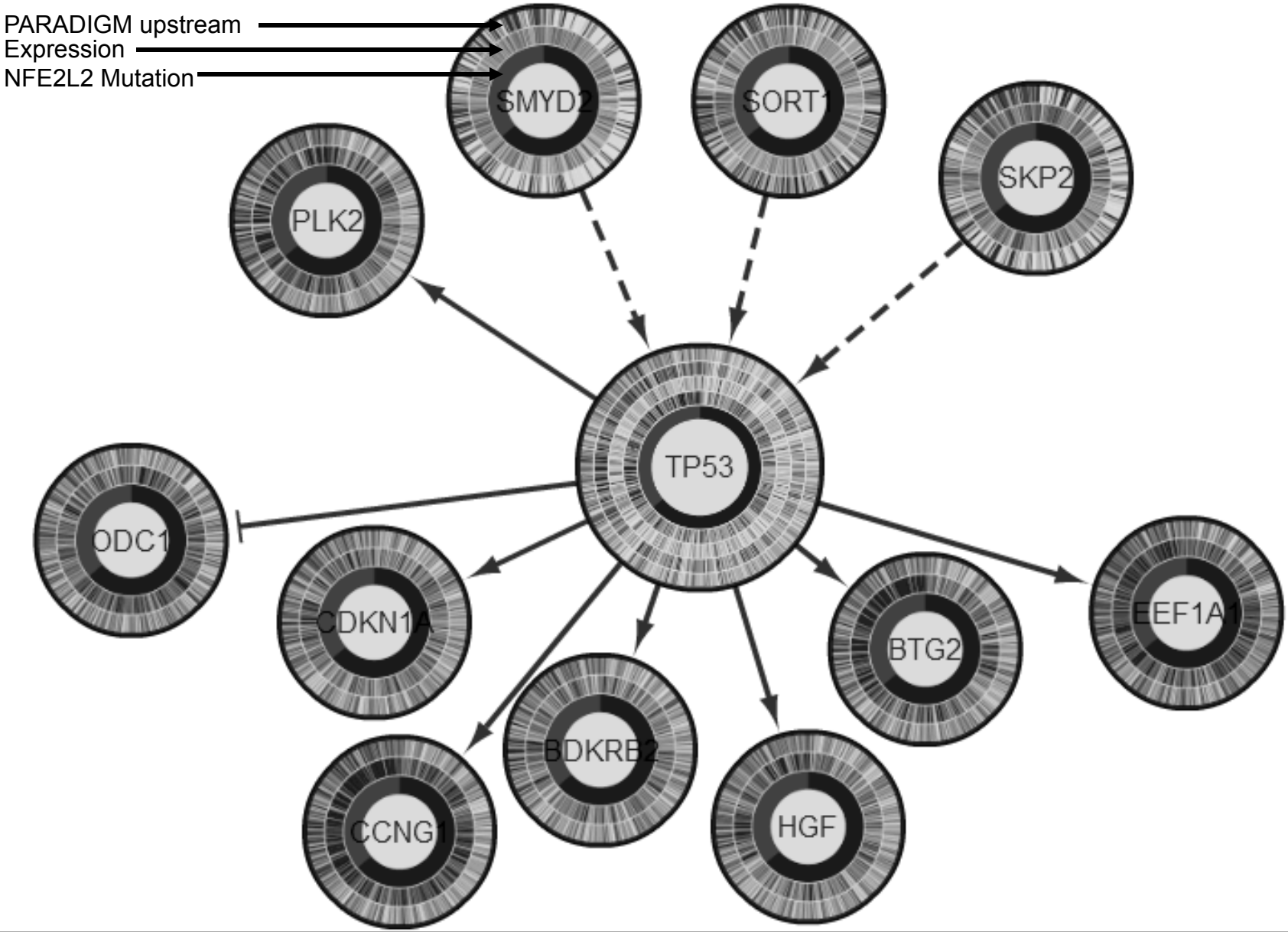


RB1 discrepancy distinction is significant

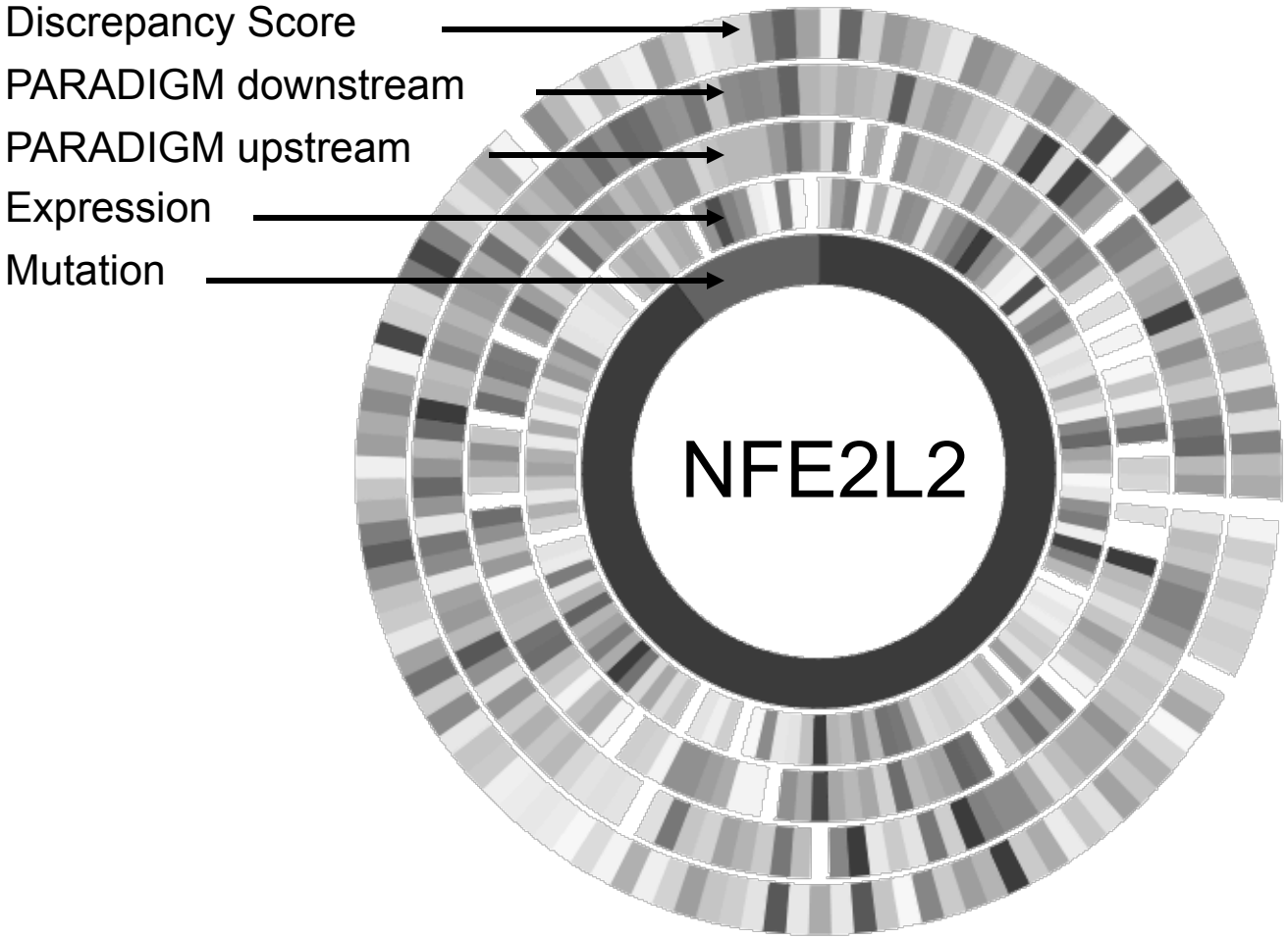
- Given the same network topology, how likely would we call a gain/loss of function
  - Background model: permute gene labels in our dataset
  - Compare observed signal score to signal scores (SS) obtained from background model



TP53 Network



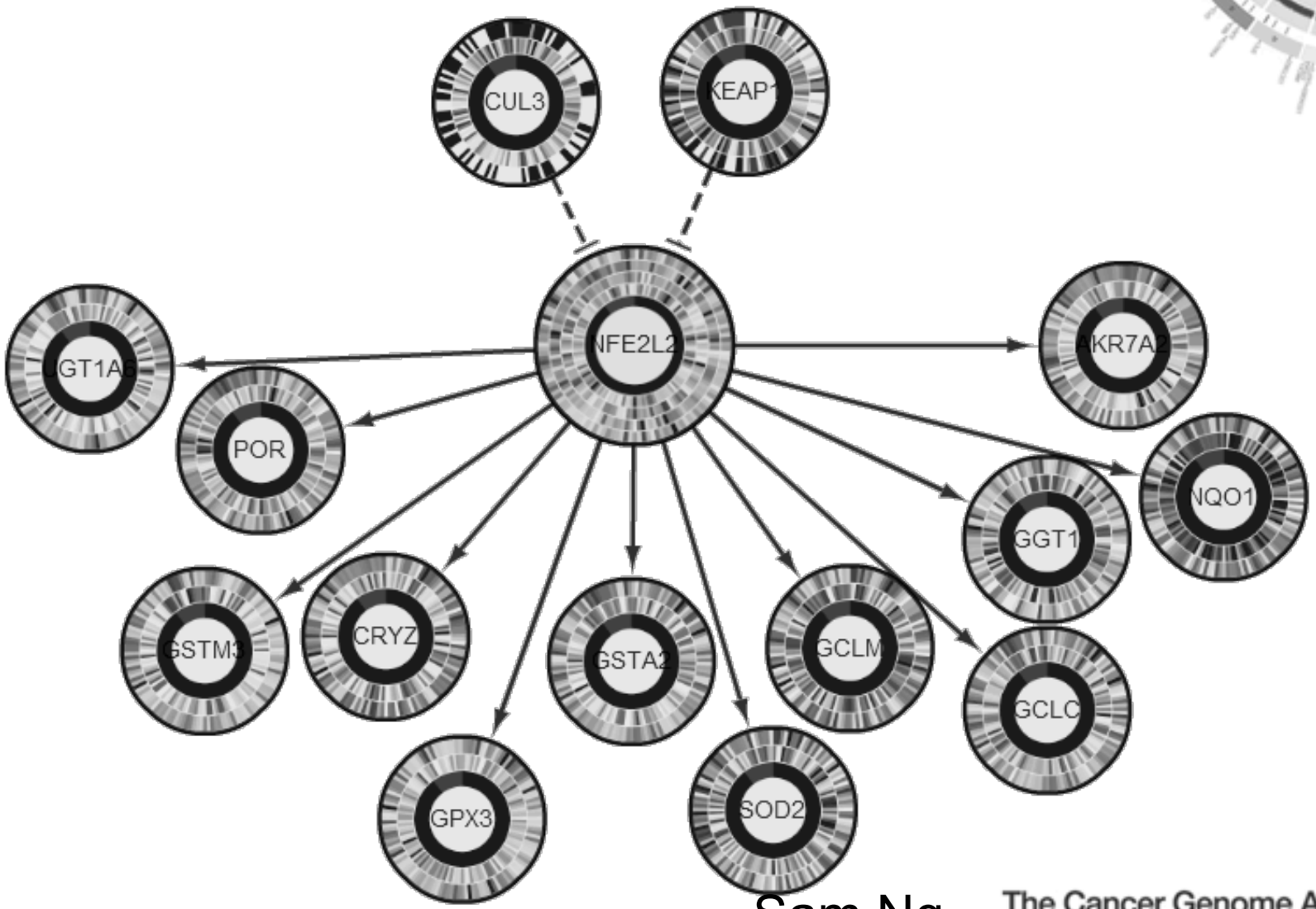
Gain-of-Function (LUSC)



Sam Ng

The Cancer Genome Atlas

NRF2 Network (LUSC)

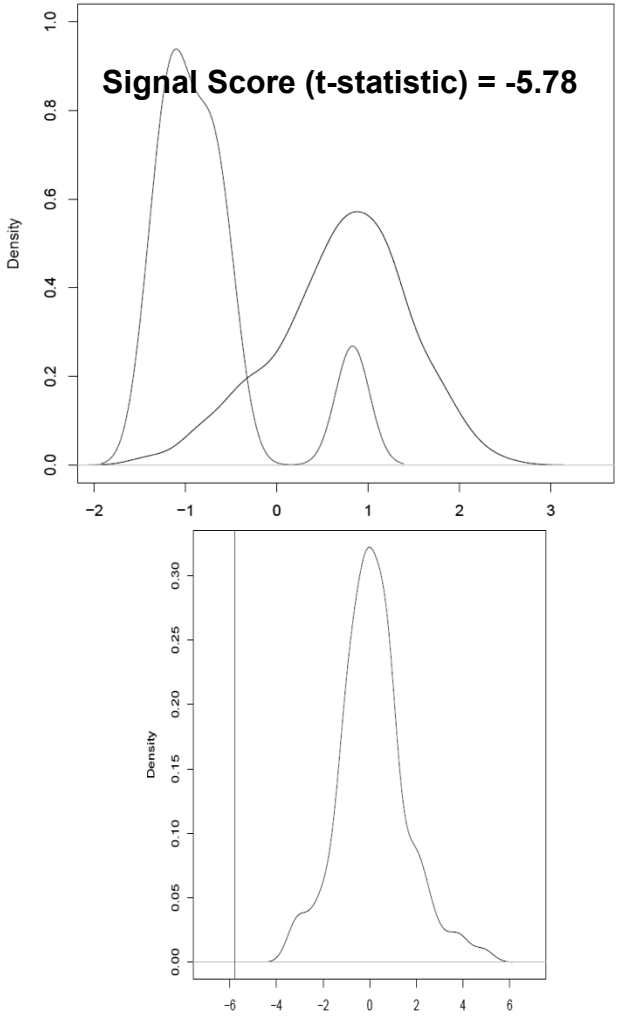


Sam Ng

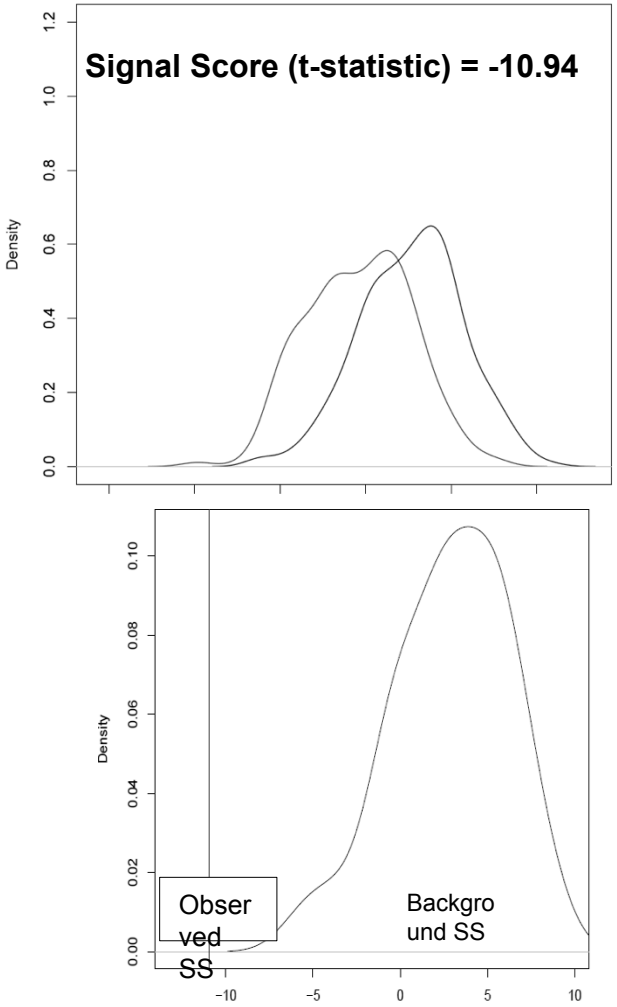
The Cancer Genome Atlas

Discrepancy scores are sensitive

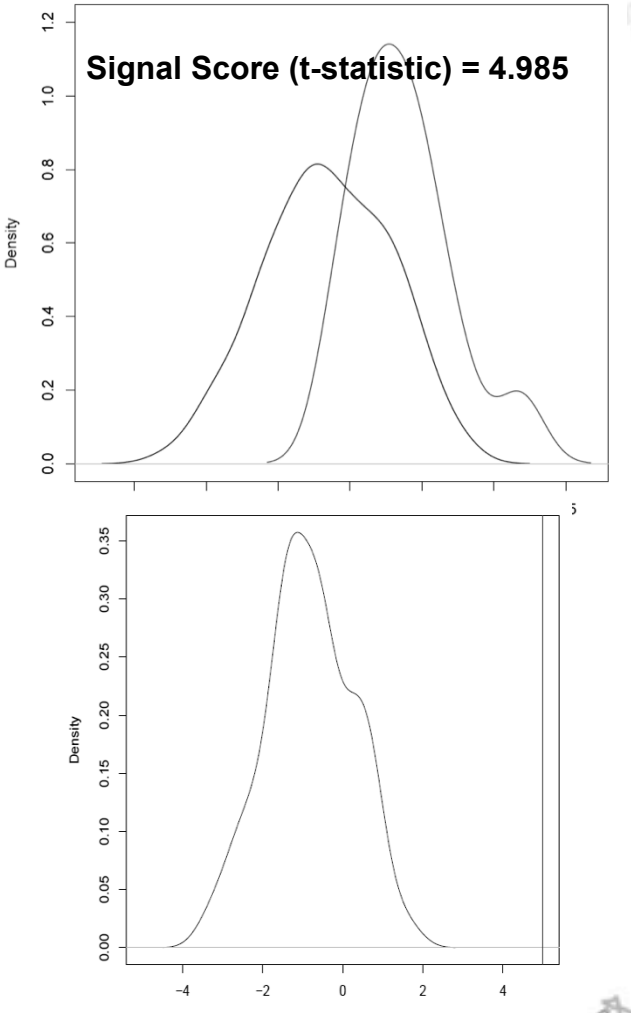
RB1



TP53



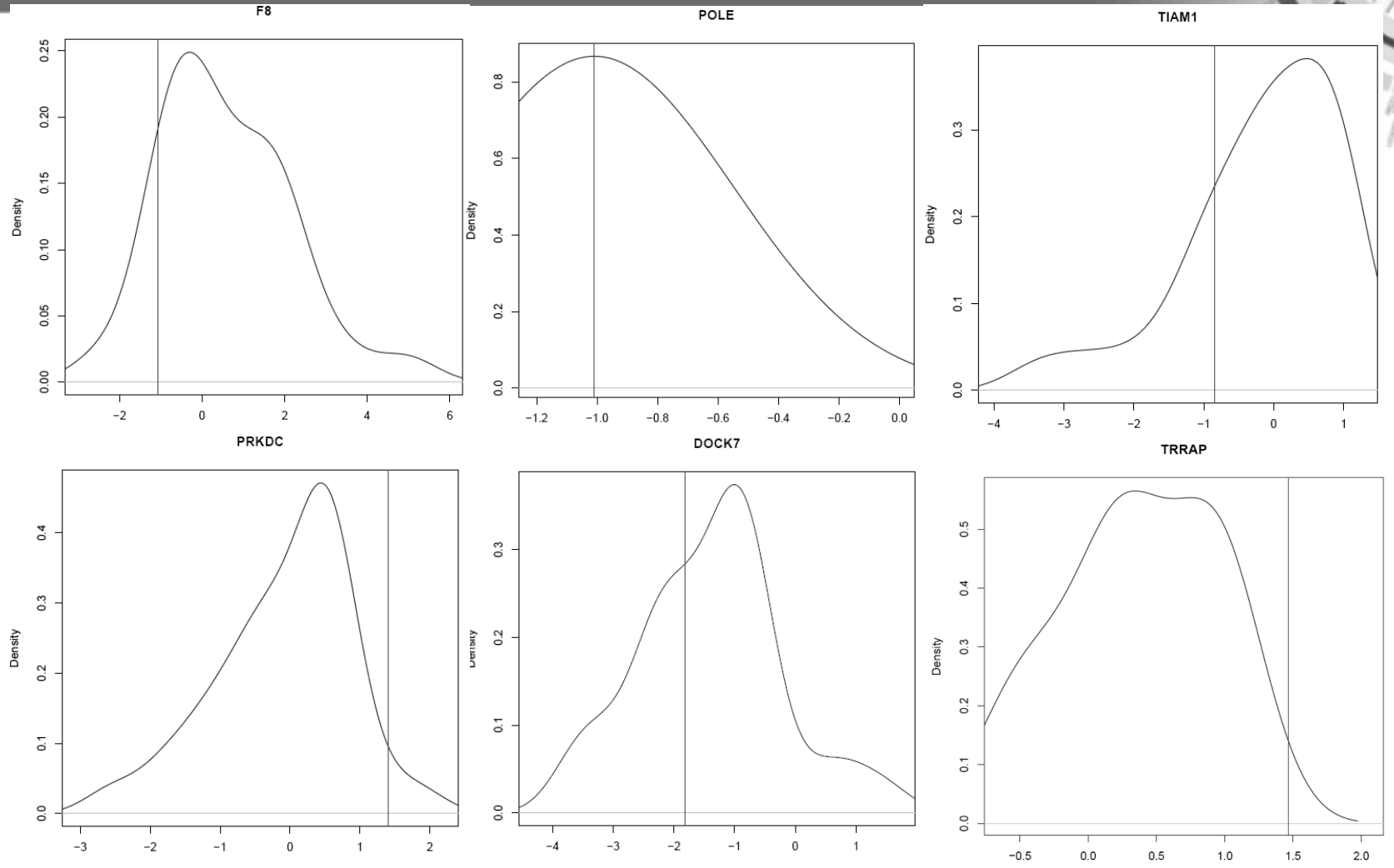
NFE2L2



Passenger Mutations should not have distinctive discrepancies

- Is the discrepancy specific?
- Negative control: calculate scores for "passenger" mutations
- Passengers:
  - insignificant by MutSig ($p > 0.10$)
  - well-represented in our pathways
- Discrepancy of these "neutral" mutations should be close to what's expected by chance (from permuted)

Discrepancies of Passenger Mutations are NOT distinctive



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
hypermut
methclust=CIMP.H
methclust=CIMP.L
methclust=Cluster3
mlh1_hypermet
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

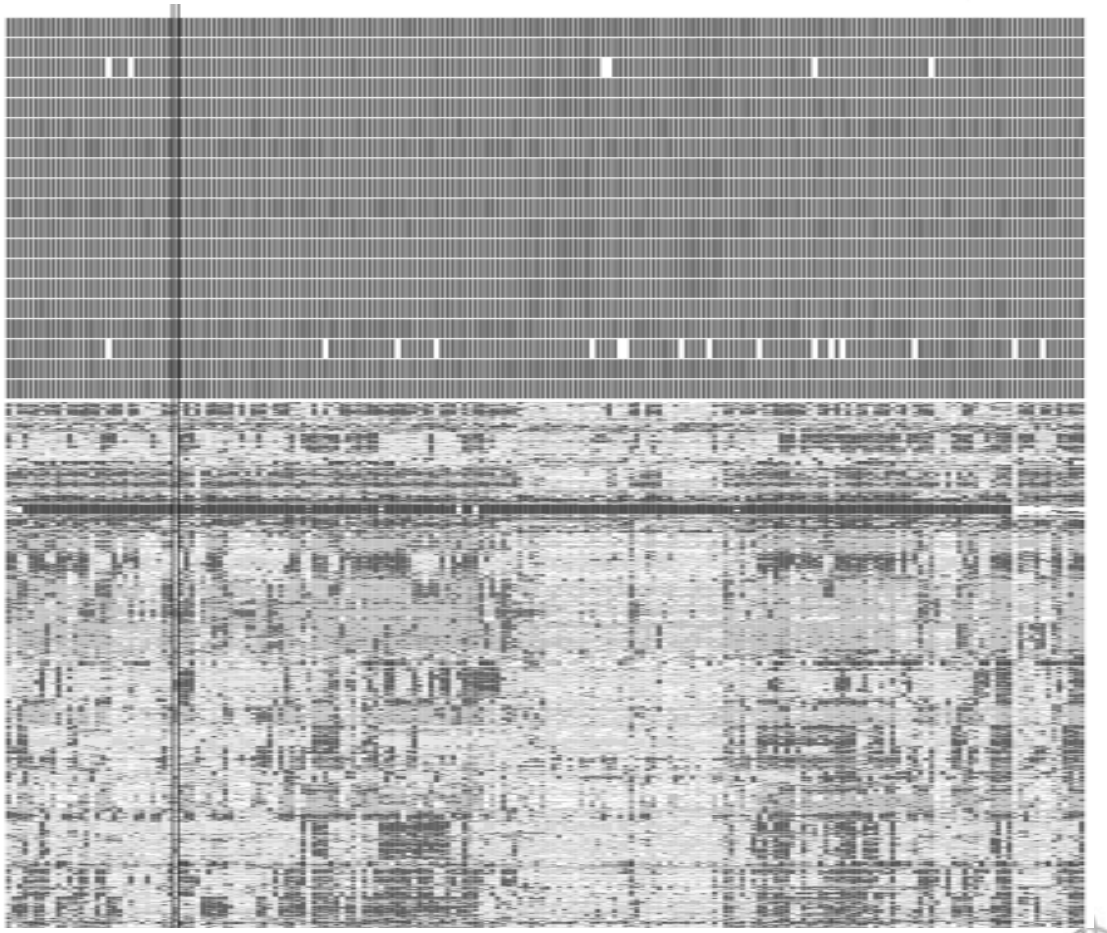
GC009

GC010 P2Y receptors

GC011 Olfactory Signaling Pathway

GC012 Ion transport by P-type ATPases

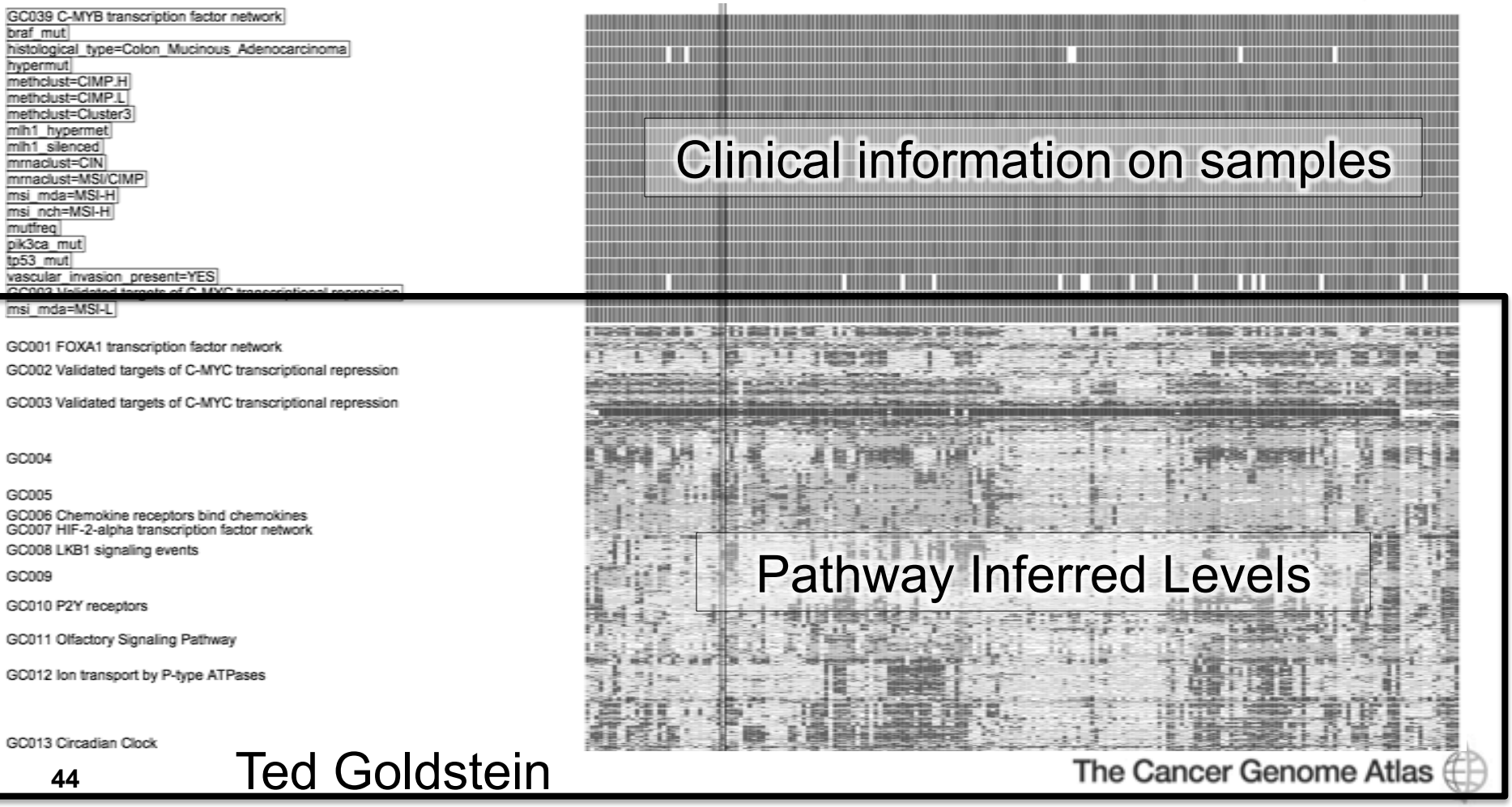
GC013 Circadian Clock



Ted Goldstein

The Cancer Genome Atlas

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

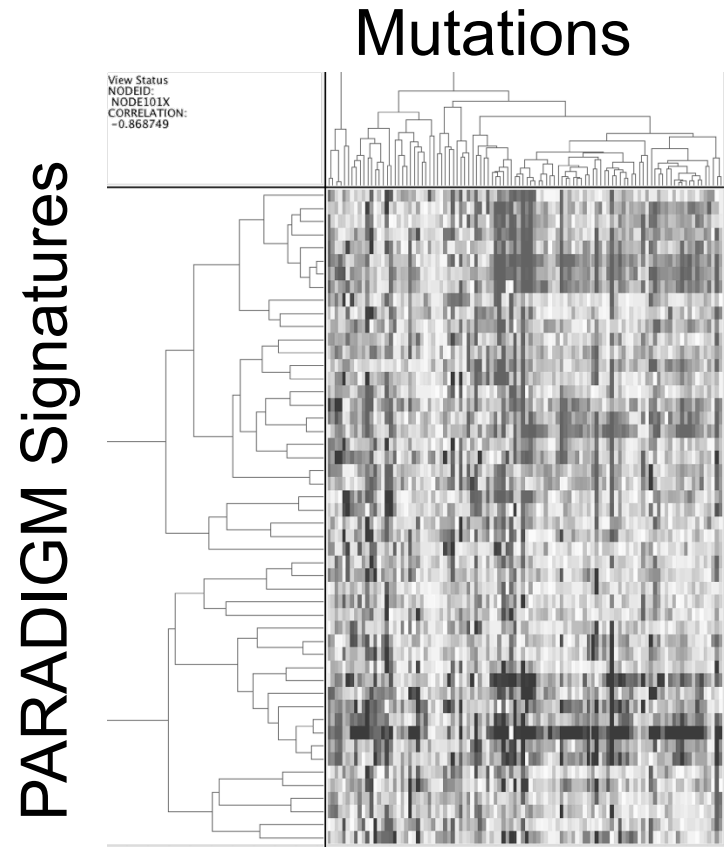


Ted Goldstein

The Cancer Genome Atlas

Mutation Association to Pathways

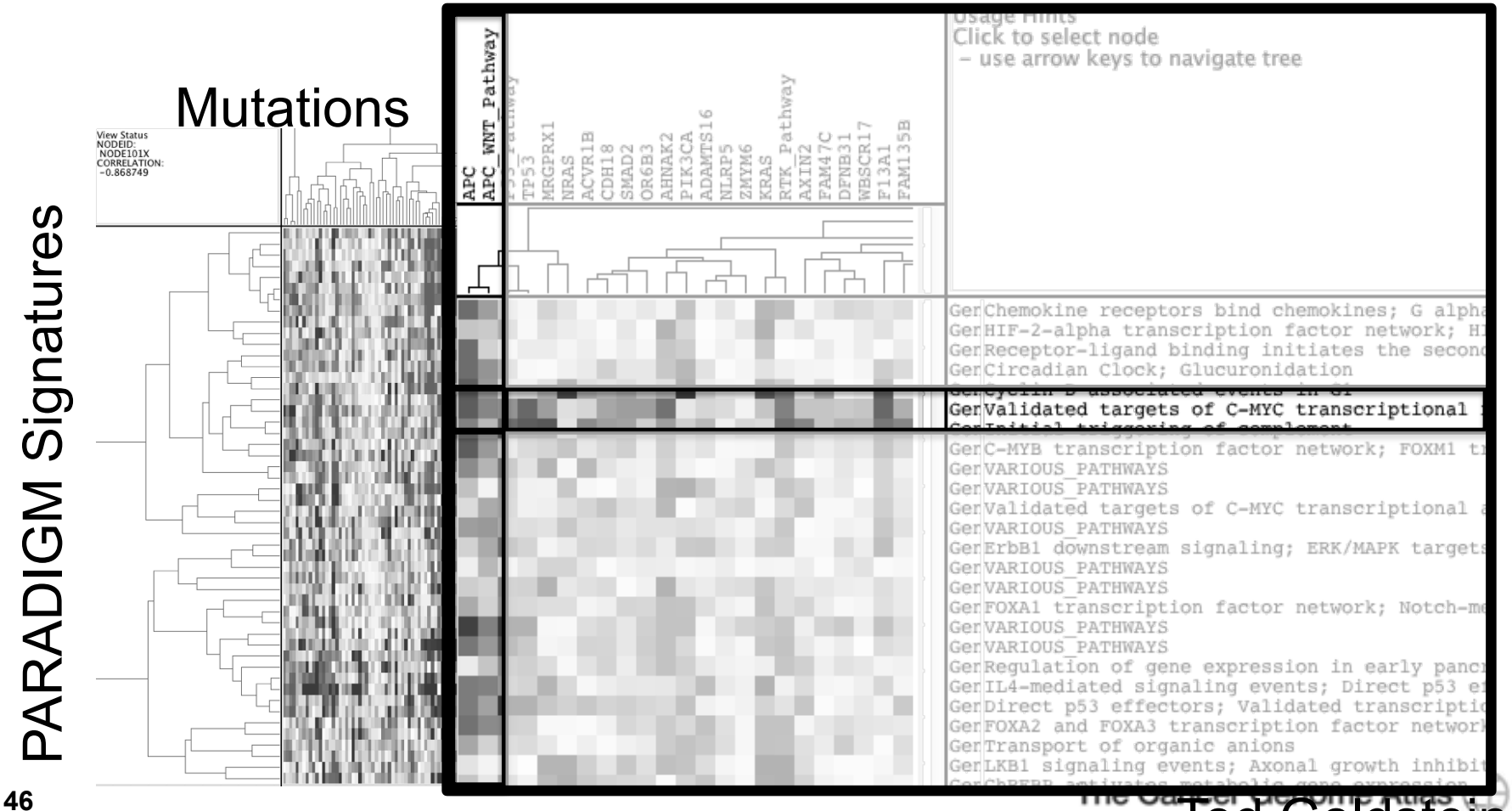
- What pathway activities is a mutation's presence associated?
- Can we classify mutations based on these associations?



Ted Goldstein
The Cancer Genome Atlas

Mutation Association to Pathways

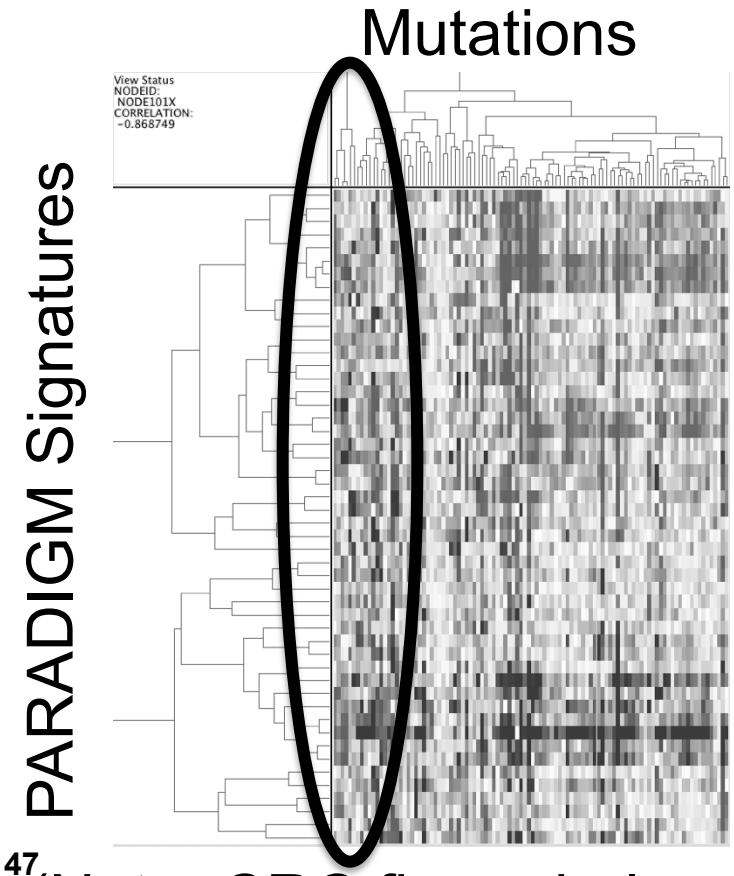
- What pathway activities is a mutation's presence associated?
- Can we classify mutations based on these associations?



Ted Goldstein

Mutation Association to Pathways

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

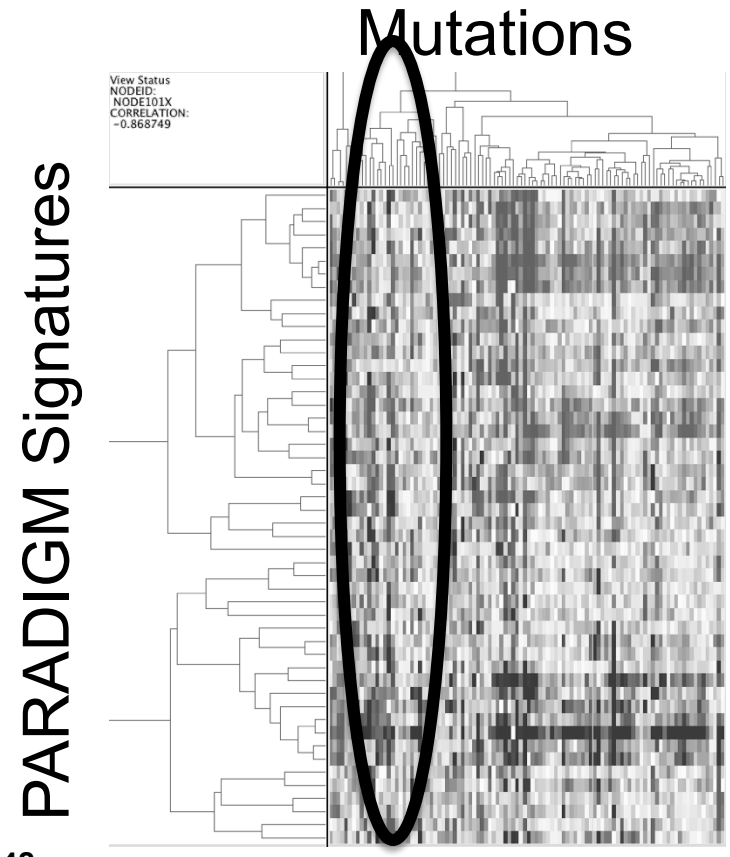


APC and TP53

(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?

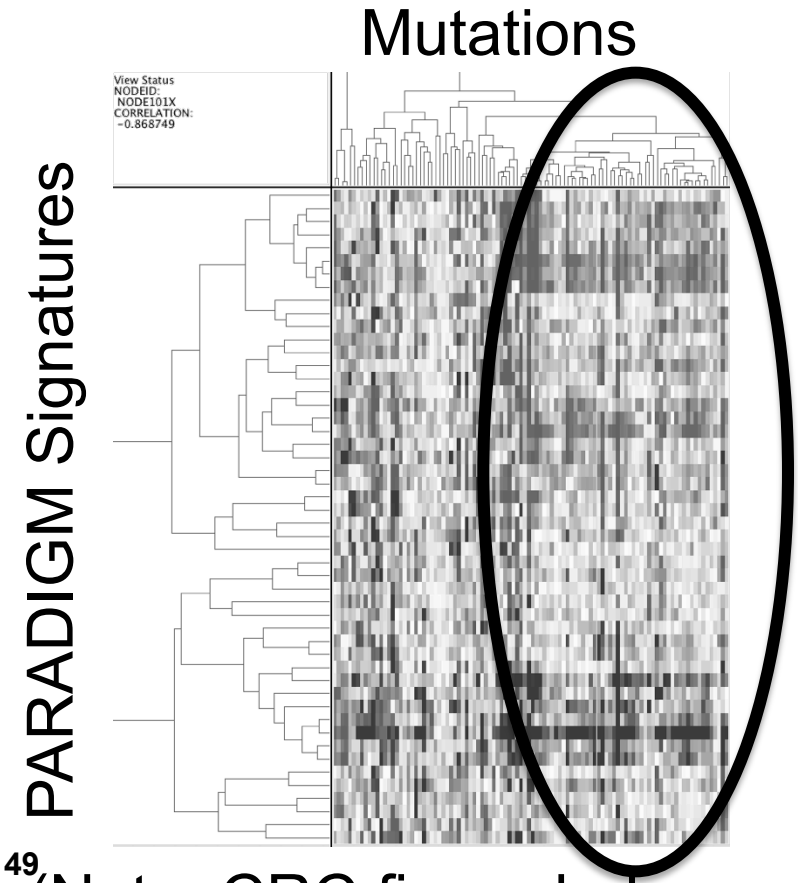


PIK3CA, RTK pathway, KRAS

(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?

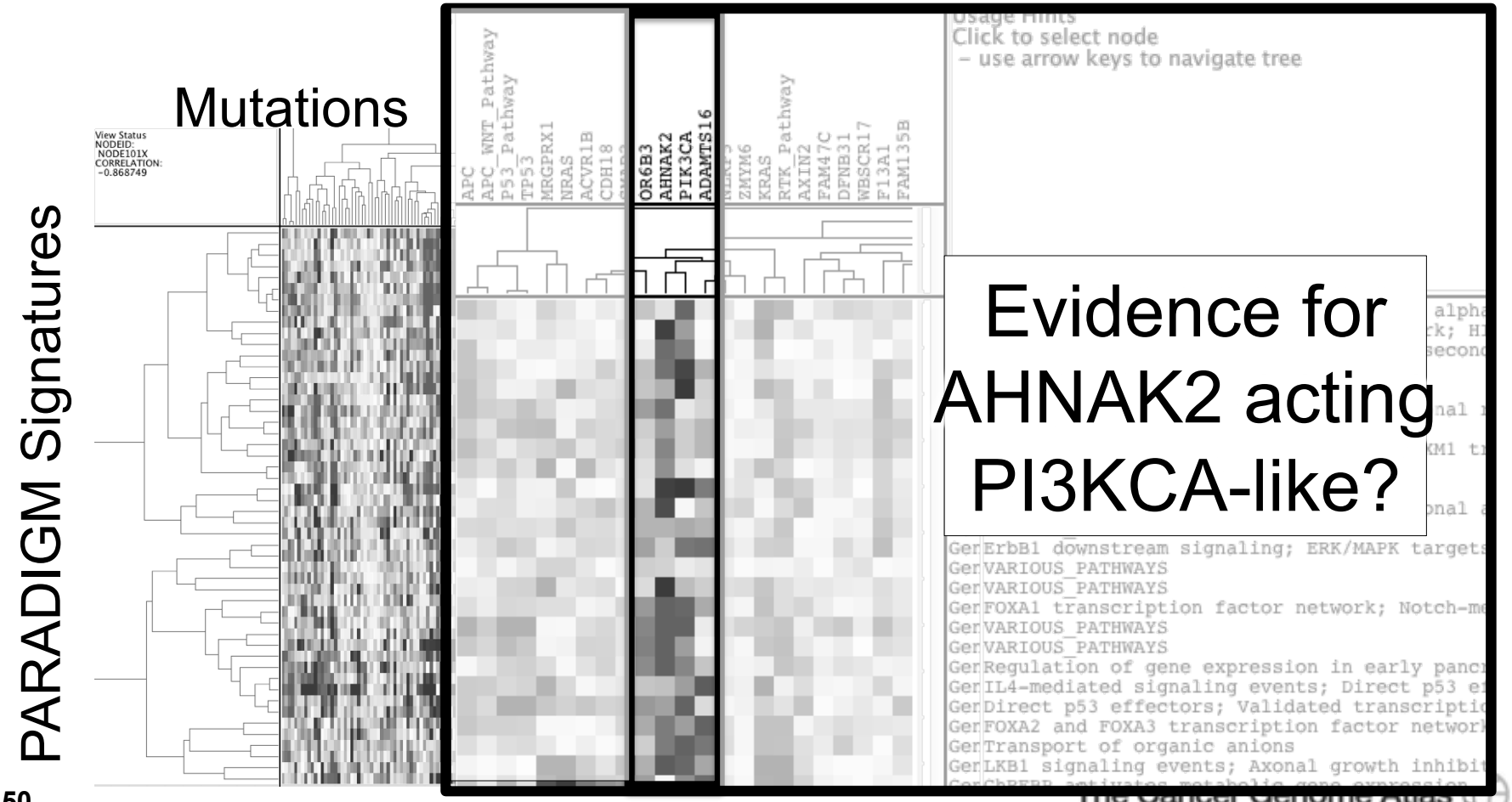


TGFB Pathway mutations

(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?



(Note: CRC figure below; soon for BRCA)

Ted Goldstein

Outline

- Integrated Pathway Approach
- Application to find Pathway Biomarkers of Cancer
- Application to predict impact of mutations
- Pan-Cancer initial look

Sub-type connections across cancers can be explored with pathway activities.

- Do samples of one subtype share pathway activities with another subtype?
- May provide therapeutic option
  - E.g. “rare toe carcinoma” has HER2-amplified signature; try herceptin on “rare toe carcinoma” (E. Collisson)
- Unsupervised analysis: compare direct signatures
- Supervised analysis
  - Train computer to recognize subtype X. Does it recognize subtype Y?
  - Perform *reciprocal prediction*: Also train on Y to predict X.

Is there a basal disease?

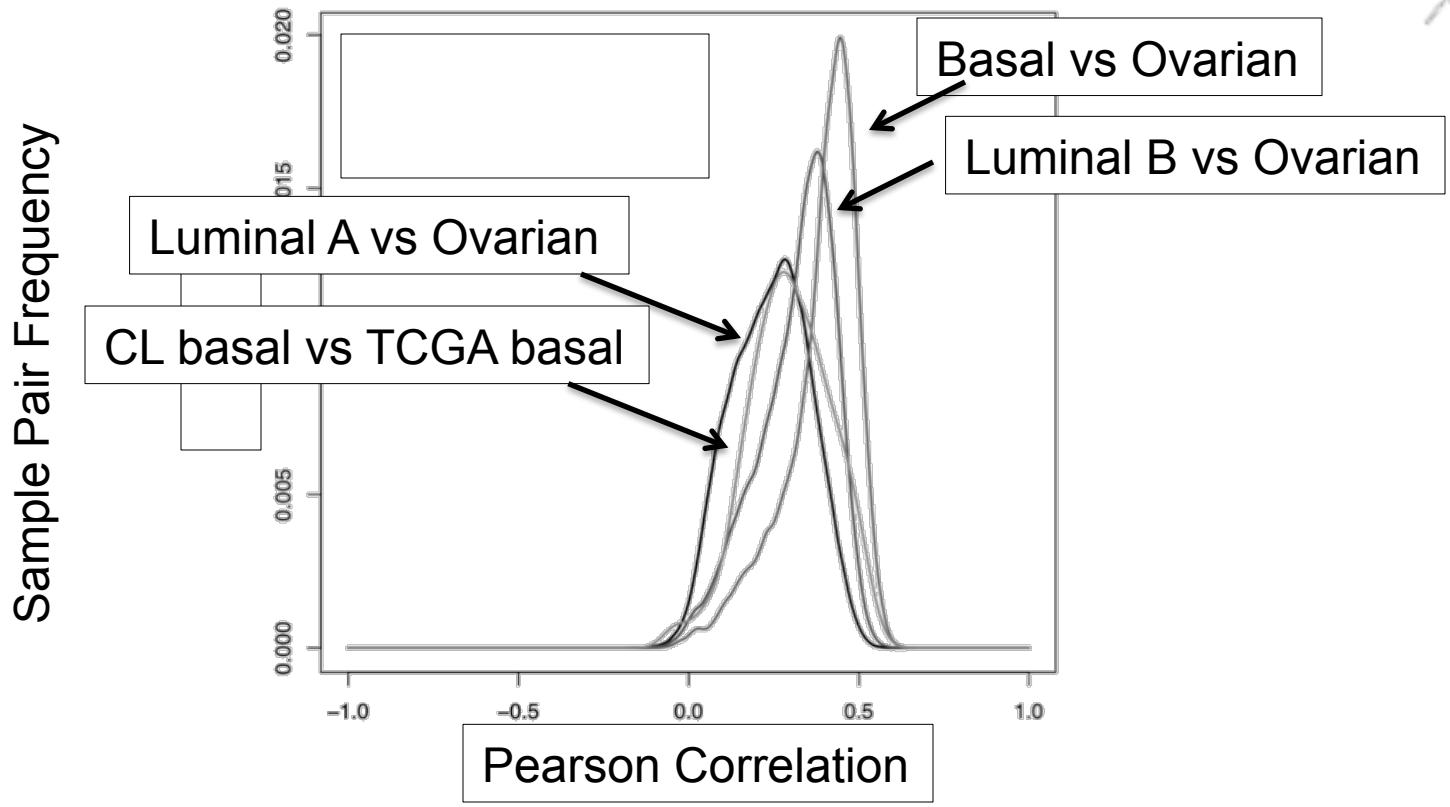
TCGA Breast

Different Breast Platform

Cell line models

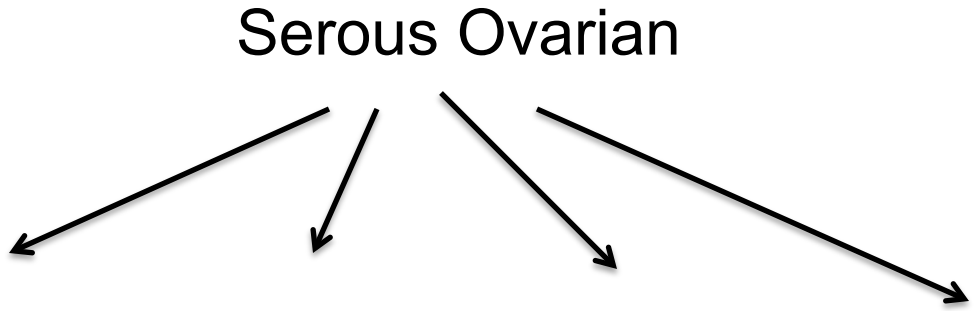
TCGA Ovarian

Is there a basal disease? – BRCA vs OVCA



- TCGA ovarian more like basal than luminal breast

Supervised: OVCA score as basal on On basal vs. luminal predictors



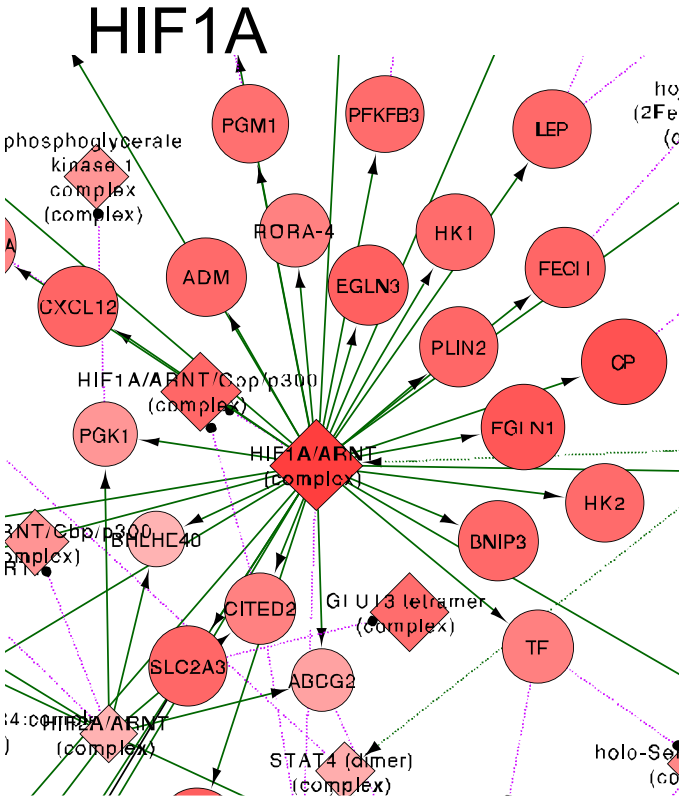
Basal predictors separate OVCA subtypes.

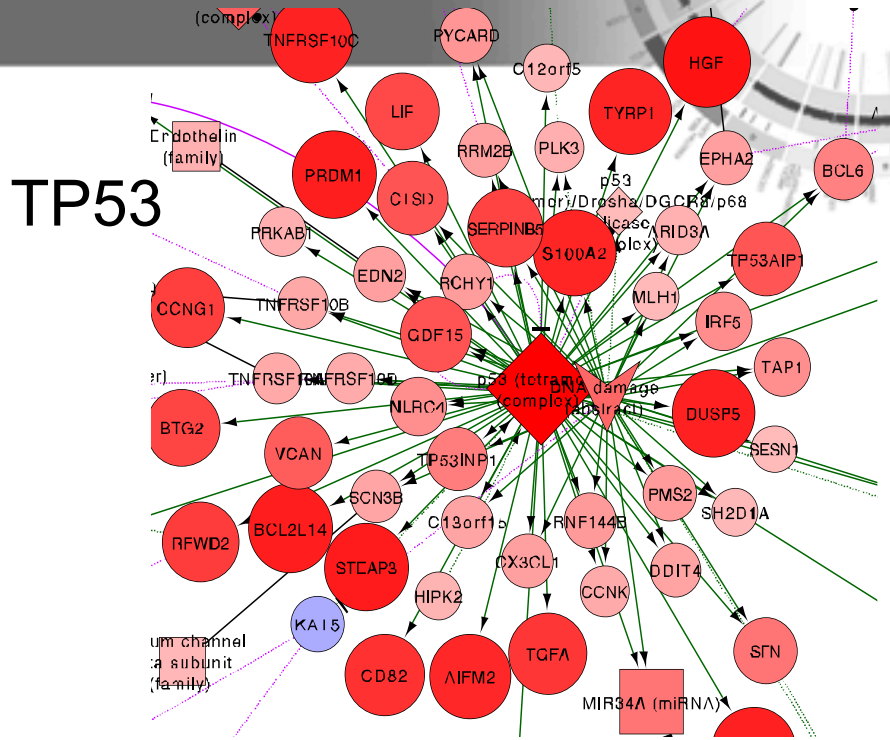
Lung to breast comparison

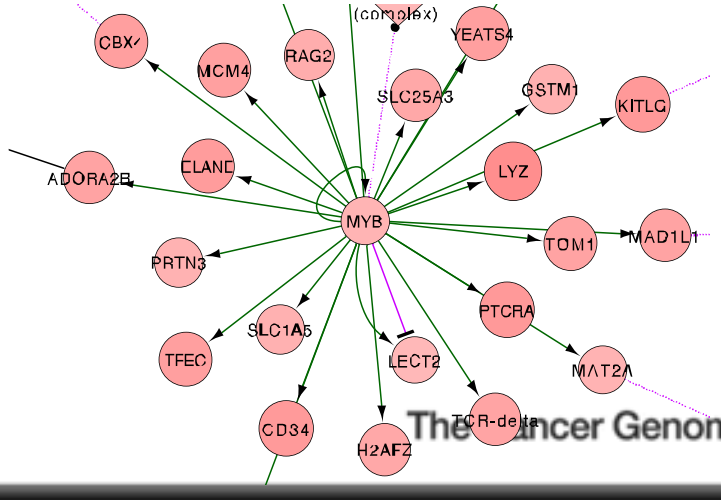
- Training LUSC vs. LUAD -> basal BRCA
  - 70% accuracy (173/250)

- Training basal vs. luminal -> LUSC
  - 94% accuracy (130/138)

Hubs predictive of basal



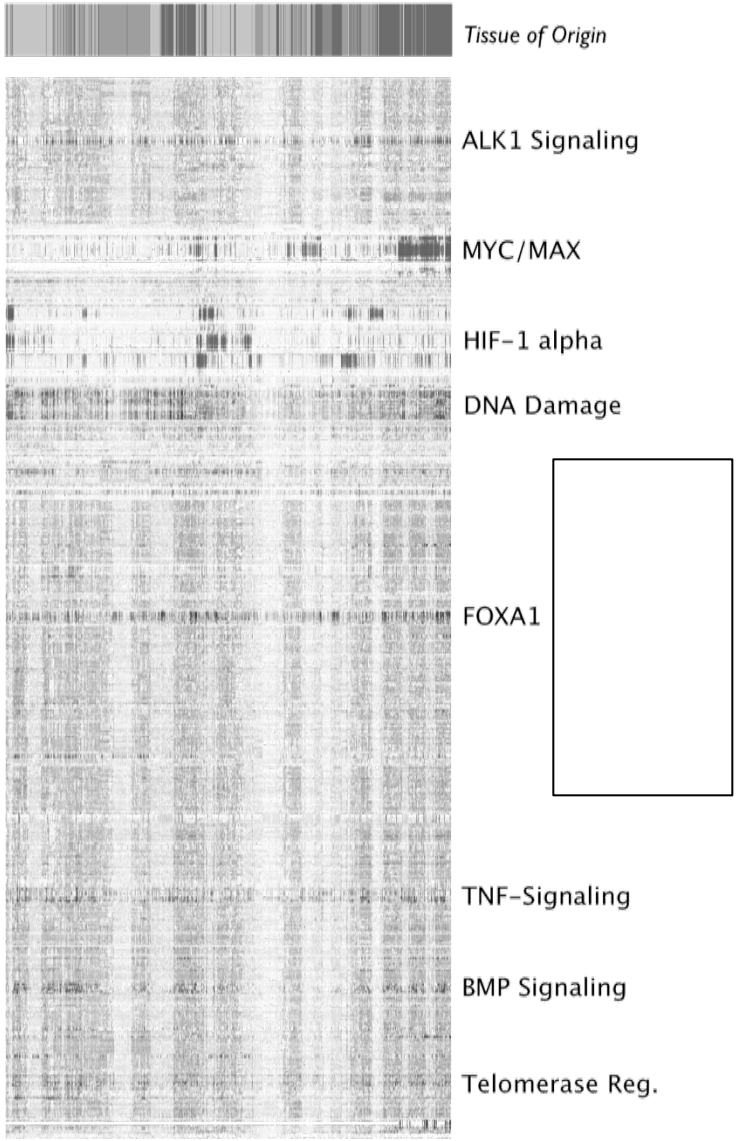




MYB

The Cancer Genome Atlas

Global Pan-Cancer Map



1382 tumor samples:

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

unpublished

The Cancer Genome Atlas

Navigating the landscape with pathways

- Provides a powerful integration framework for a large number of data types
- Focuses results on known biology
- Provides a method to stratify patients more accurately than using the original data
- Sub-networks are predictive markers and can be used to simulate scenarios (like drug inhibition)
- Allows integration of data across cohorts in different cancers

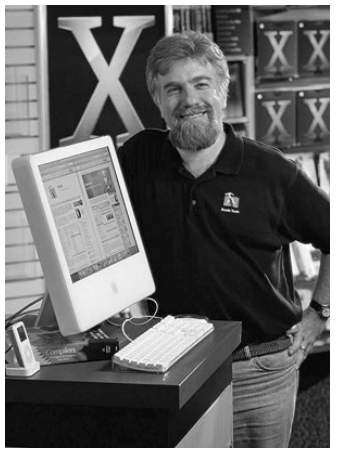
UCSC Integrative Genomics Group

Marcos Woehrmann

Sam Ng

Dan Carlin

Ted Golstein



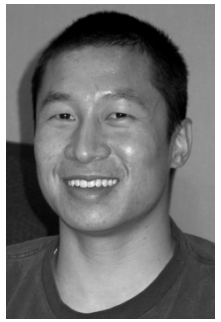
Evan Paull

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

UC SANTA CRUZ

Acknowledgments

UCSC Cancer Genomics

- Kyle Ellrott
- Brian Craft
- Erich Weiler
- Chris Wilks
- Sam Ng
- Amie Radenbaugh
- Mia Grifford
- Sofie Salama
- Steve Benz
- Charles Vaske
- Zack Sanborn
- Larry Meyer
- Tracy Ballinger
- Daniel Zerbino

- UCSC Genome Browser Staff
- Cluster Admins

- Collaborators
- The Cancer Genome Atlas
- Stand Up To Cancer
- Christopher Benz, Buck Institut
- Laura Esserman, UCSF
- Joe Gray, LBL
- Laura Heiser, LBL
- Eric Collisson, UCSF
- Ting Wang, Washington University

David Haussler

Jing Zhu

- Funding Agencies
- CIRM
- NCI/NIH
- NHGRI
- American Association for Cancer Research (AACR)
- UCSF Comprehensive Cancer Center
- California Institute for Quantitative Biosciences (QB3)