Sample-specific cancer pathway analysis using PARADIGM

Stephen Benz

Thesis Advancement
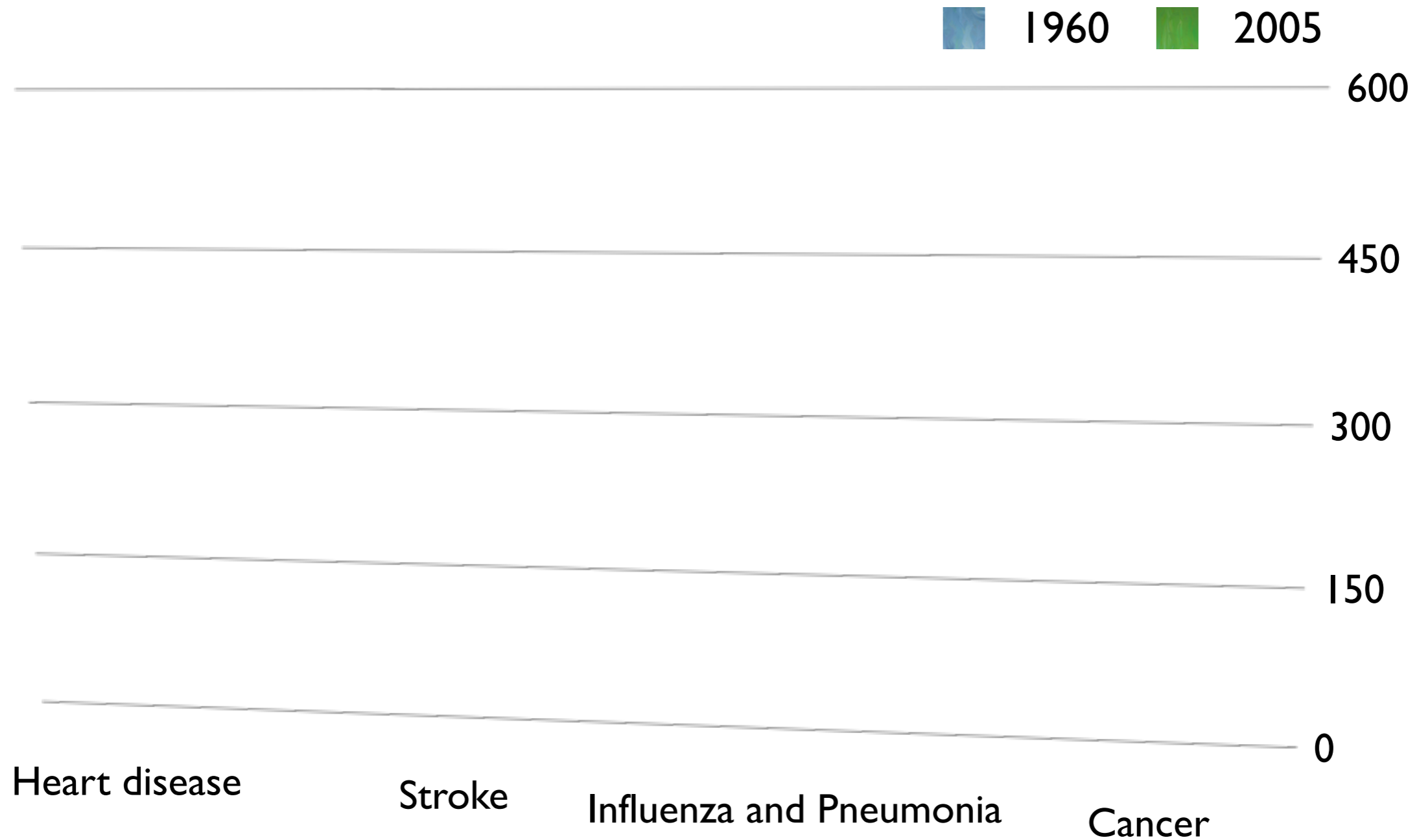June 15, 2010

Outline

I. Introduction

II. Previous Work

III. Aim 1: PARADIGM - Integrated Pathway Analysis

IV. Aim 2: Modeling pathways using sequence data

V. Aim 3: Directing combination therapies

Outline

I. Introduction

II. Previous Work

III. Aim 1: PARADIGM - Integrated Pathway Analysis

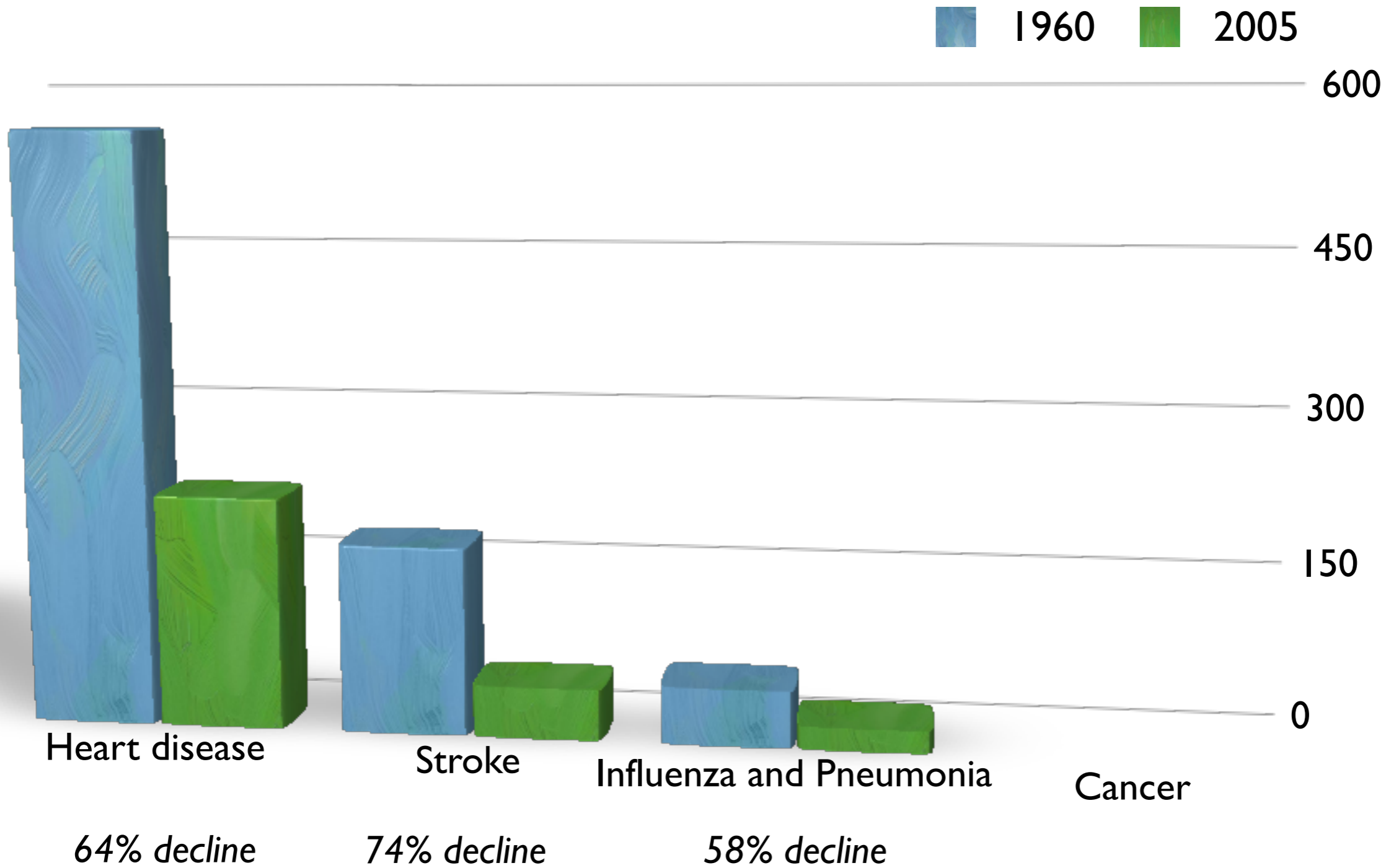IV. Aim 2: Modeling pathways using sequence data

V. Aim 3: Directing combination therapies
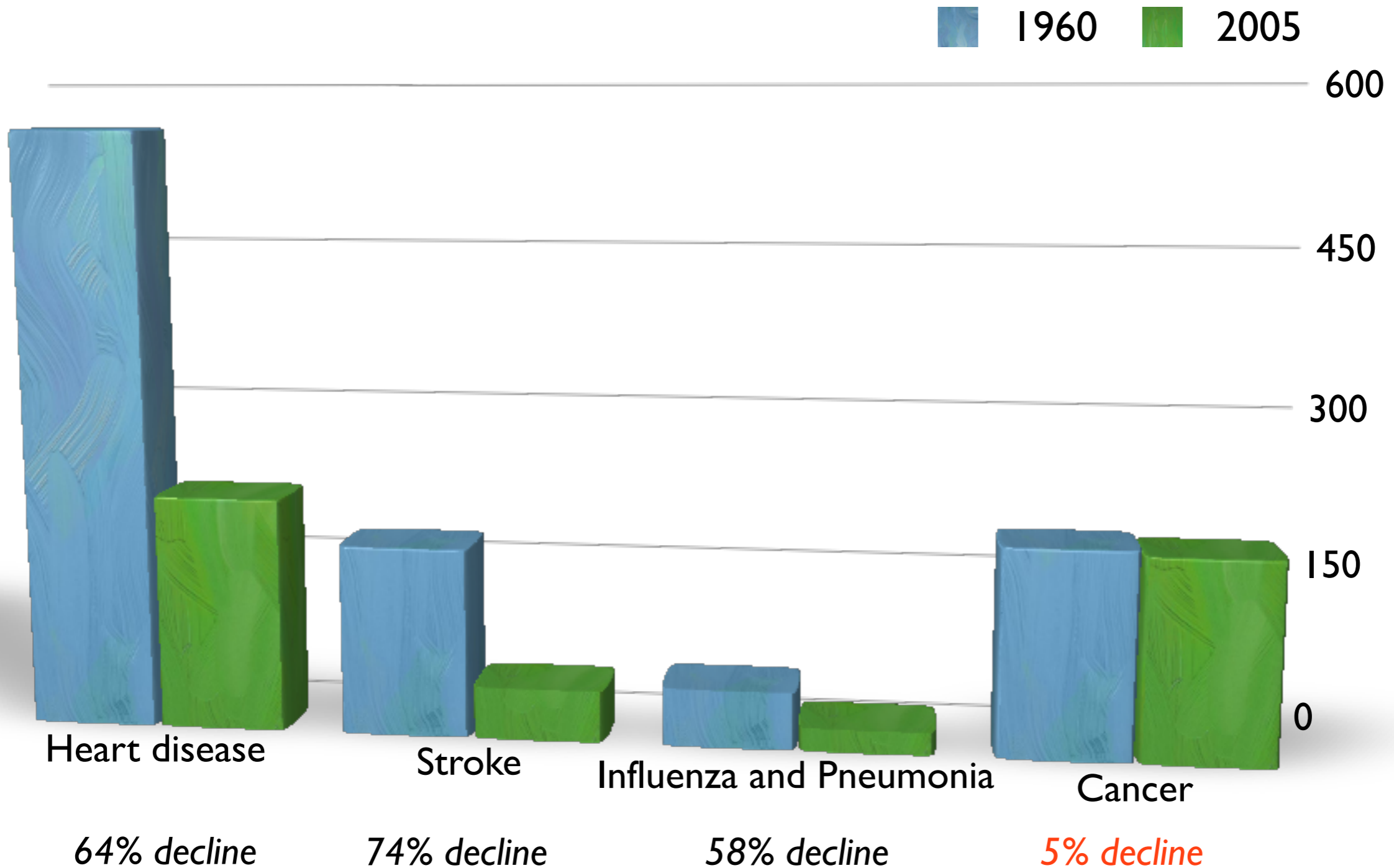
Deaths in the US

1960   2005

600

450

300

150

0

Heart disease   Stroke   Influenza and Pneumonia   Cancer

Deaths in the US

Source: U.S. National Center for Health Statistics, National Vital Statistics Reports, Vol. 54, No. 19, June 28, 2006. Obtained at http://www.infoplease.com/ipa/A0005124.html
Inspired by the New York Times article "As Other Death Rates Fall, Cancer's Scarcely Moves." Gina Kolata. April 24, 2009.

Tuesday, June 15, 2010

Deaths in the US

Source: U.S. National Center for Health Statistics, National Vital Statistics Reports, Vol. 54, No. 19, June 28, 2006. Obtained at http://www.infoplease.com/ipa/A0005124.html
Inspired by the New York Times article "As Other Death Rates Fall, Cancer's Scarcely Moves." Gina Kolata. April 24, 2009.

Tuesday, June 15, 2010
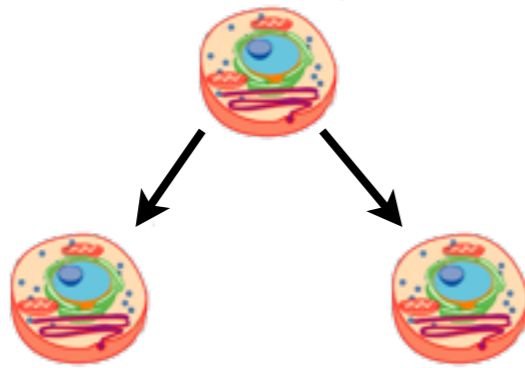
Mechanics of Cancer

Normal Cell Division

Mechanics of Cancer
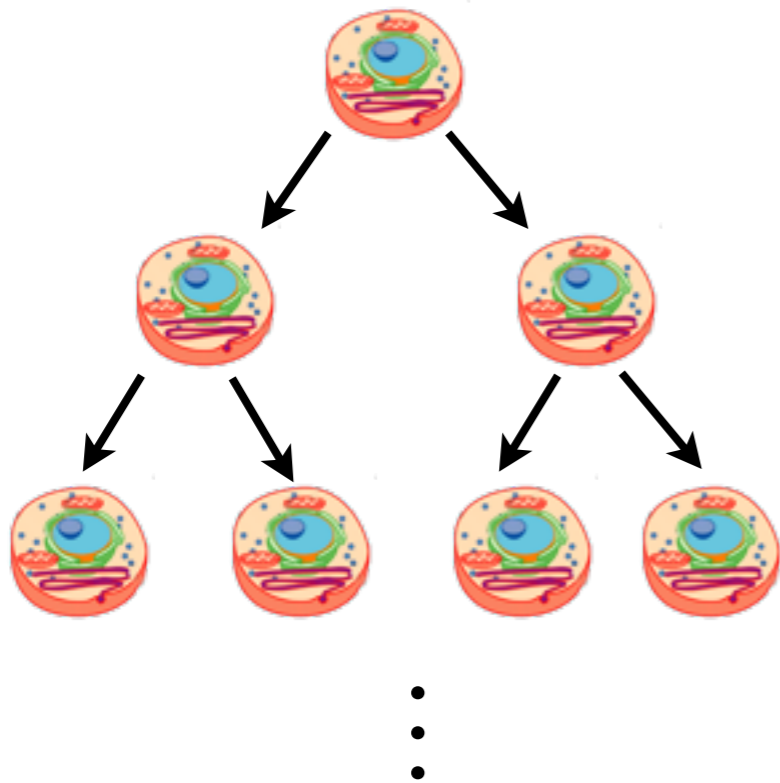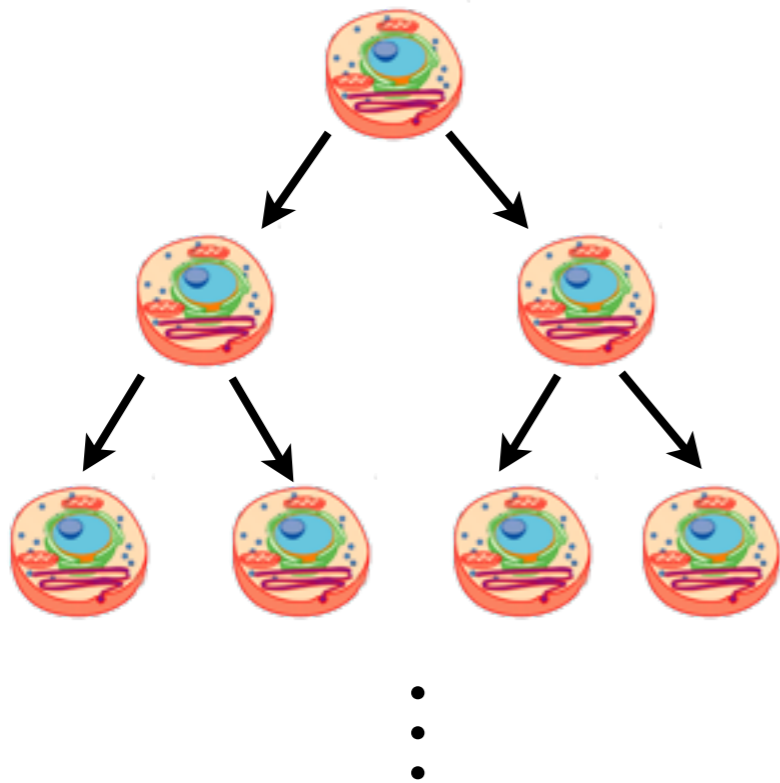
Normal Cell Division

Mechanics of Cancer

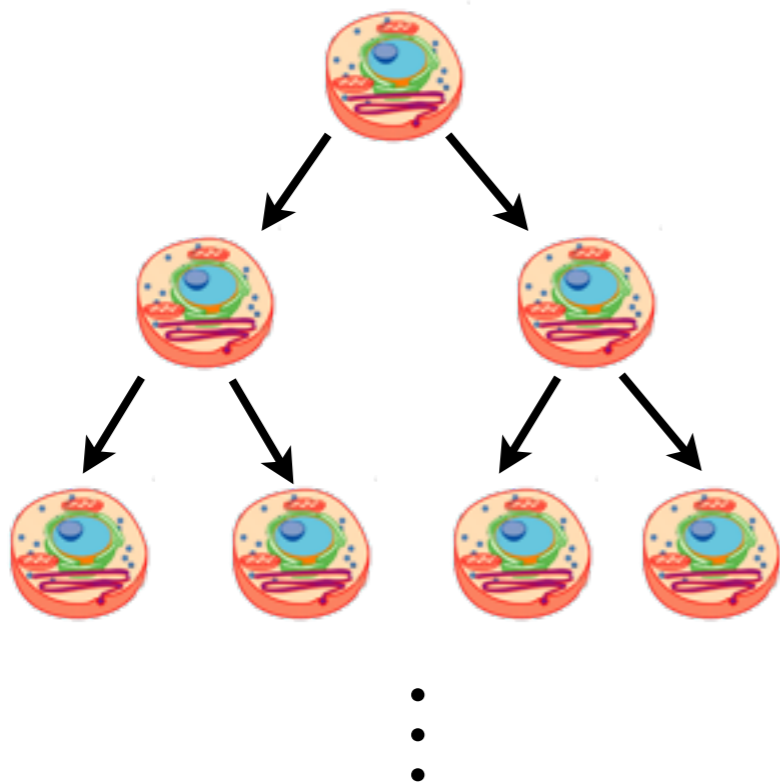Normal Cell Division

Mechanics of Cancer

Normal Cell Division
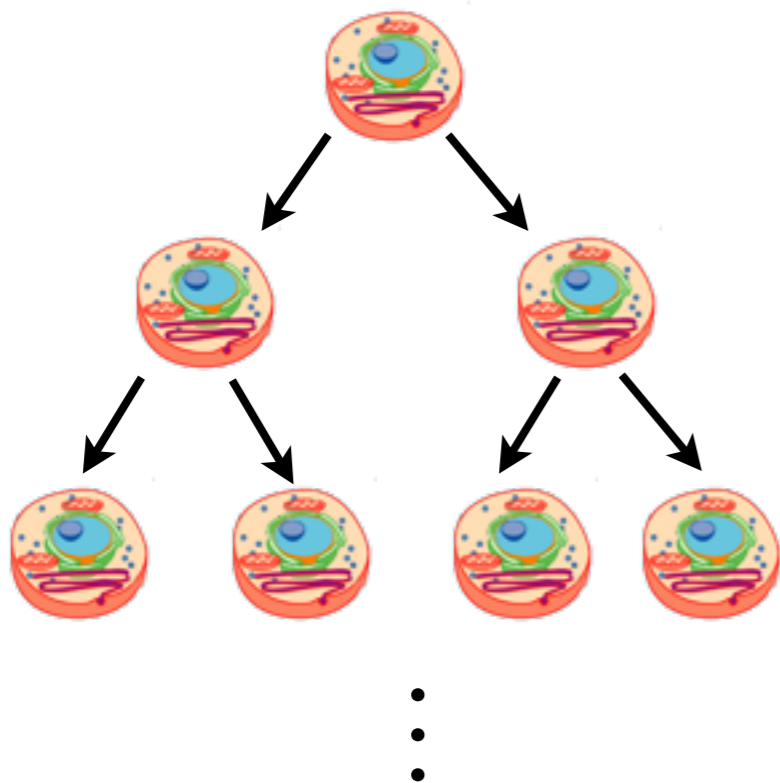
Cancer

Mechanics of Cancer

Normal Cell Division

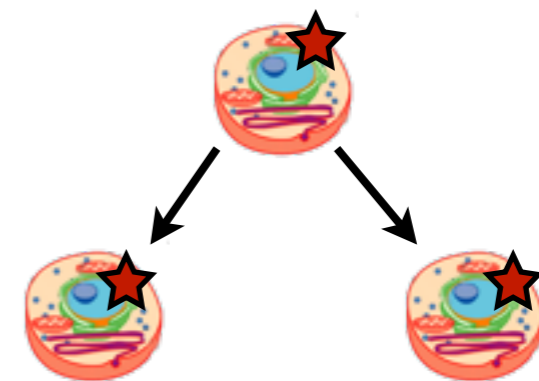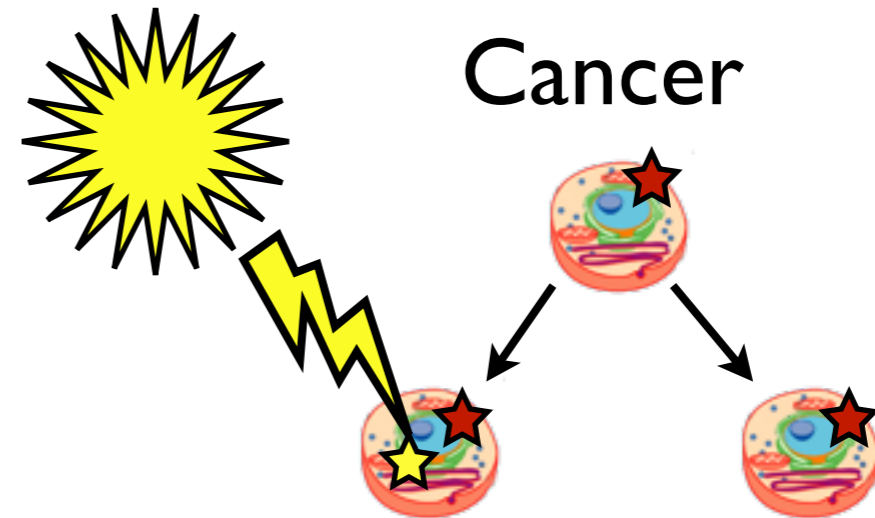Cancer

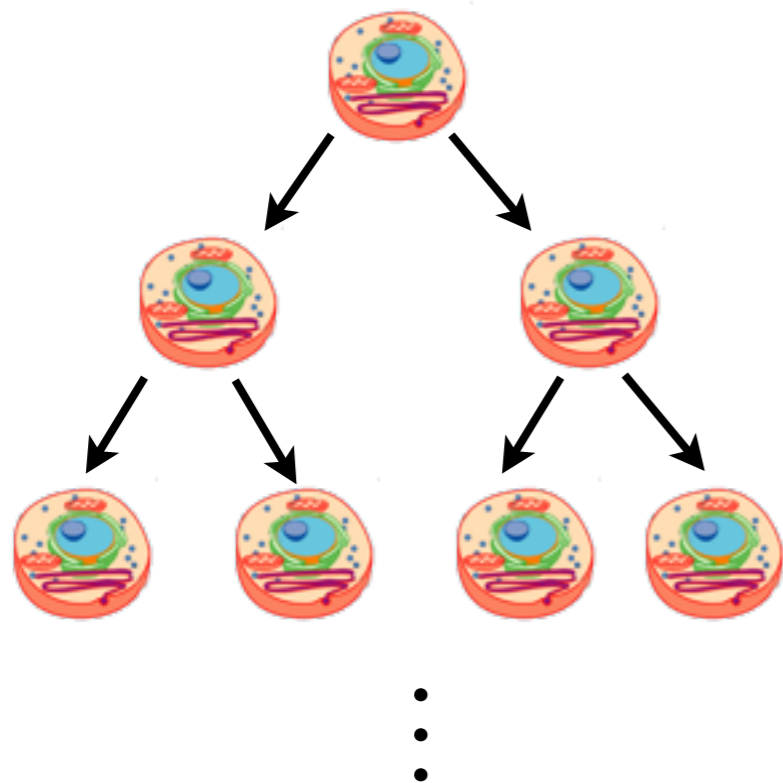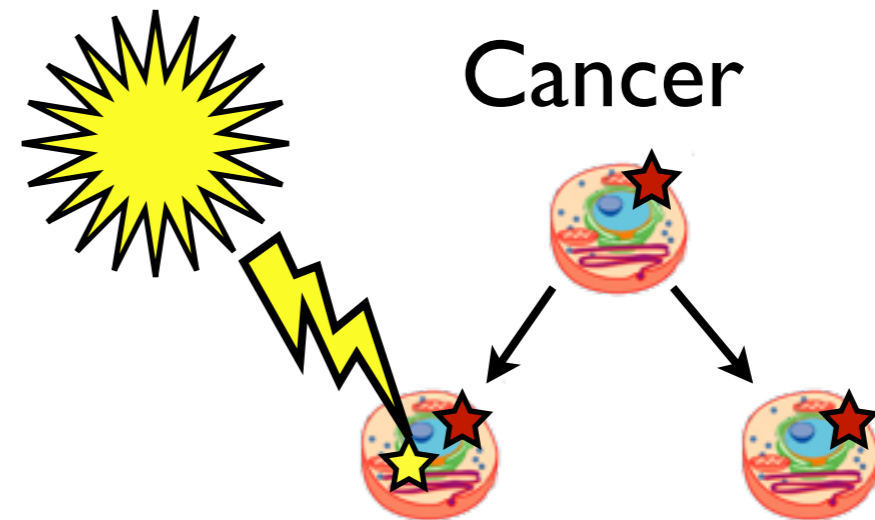⭐ Inherited Mutation

Mechanics of Cancer

Normal Cell Division

Cancer
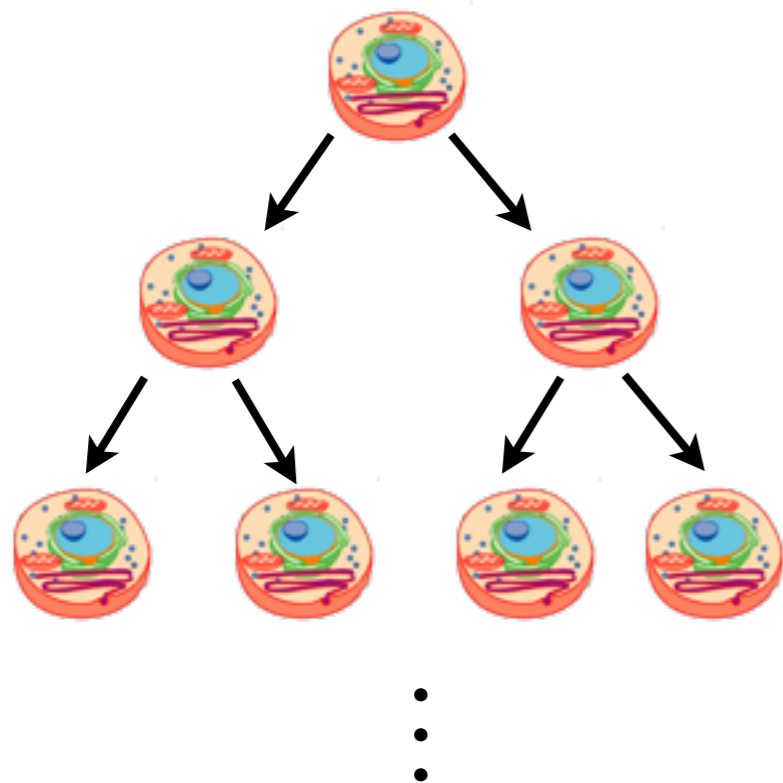

⭐ Inherited Mutation

Mechanics of Cancer

Mechanics of Cancer

Normal Cell Division

Cancer

⭐ Inherited Mutation

⭐ Somatic Mutation

Mechanics of Cancer

Normal Cell Division

Cancer

Inherited Mutation

Somatic Mutation
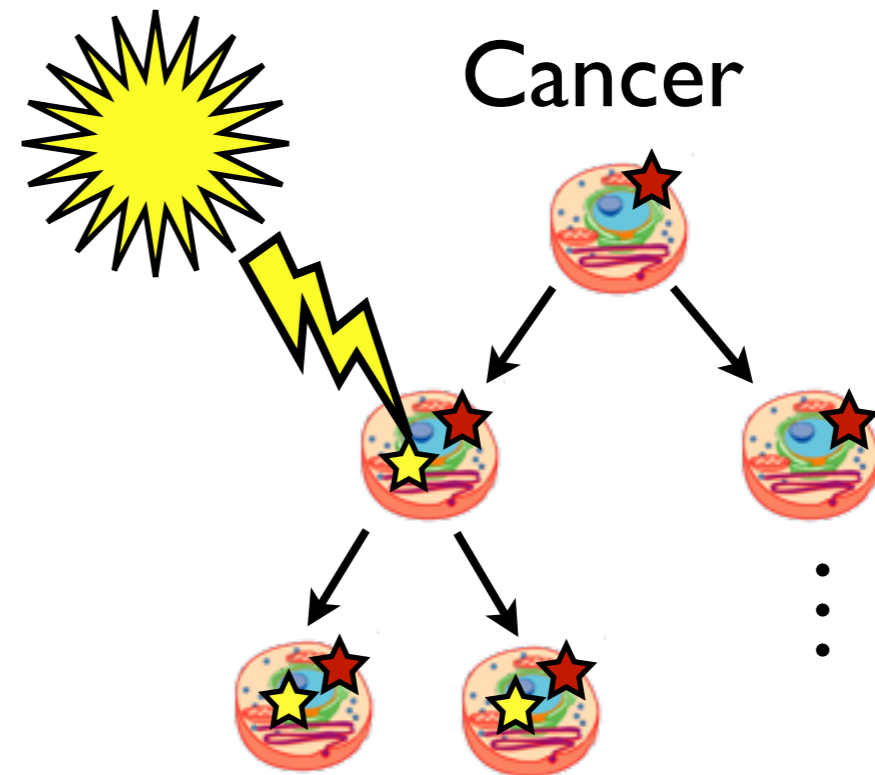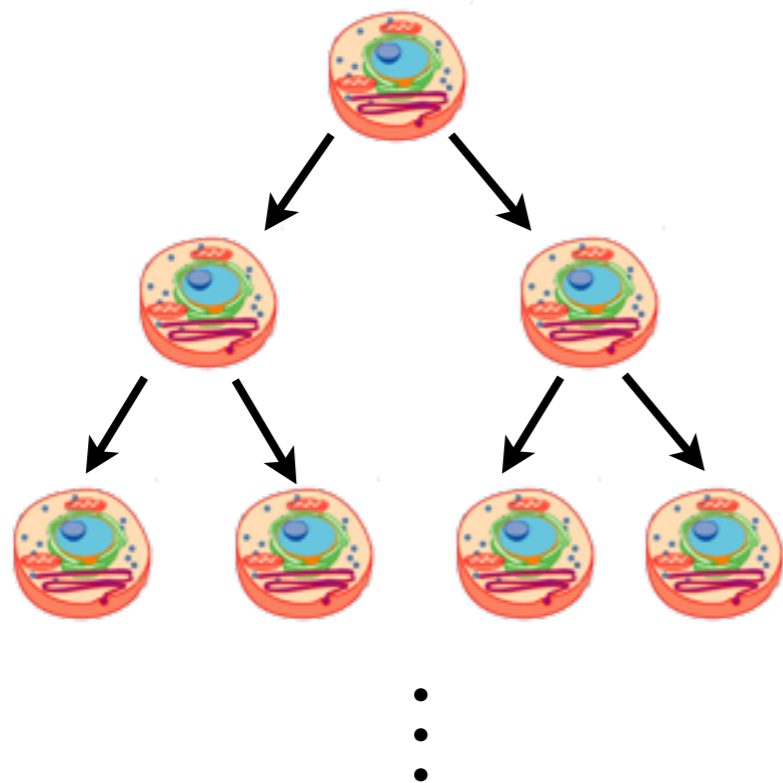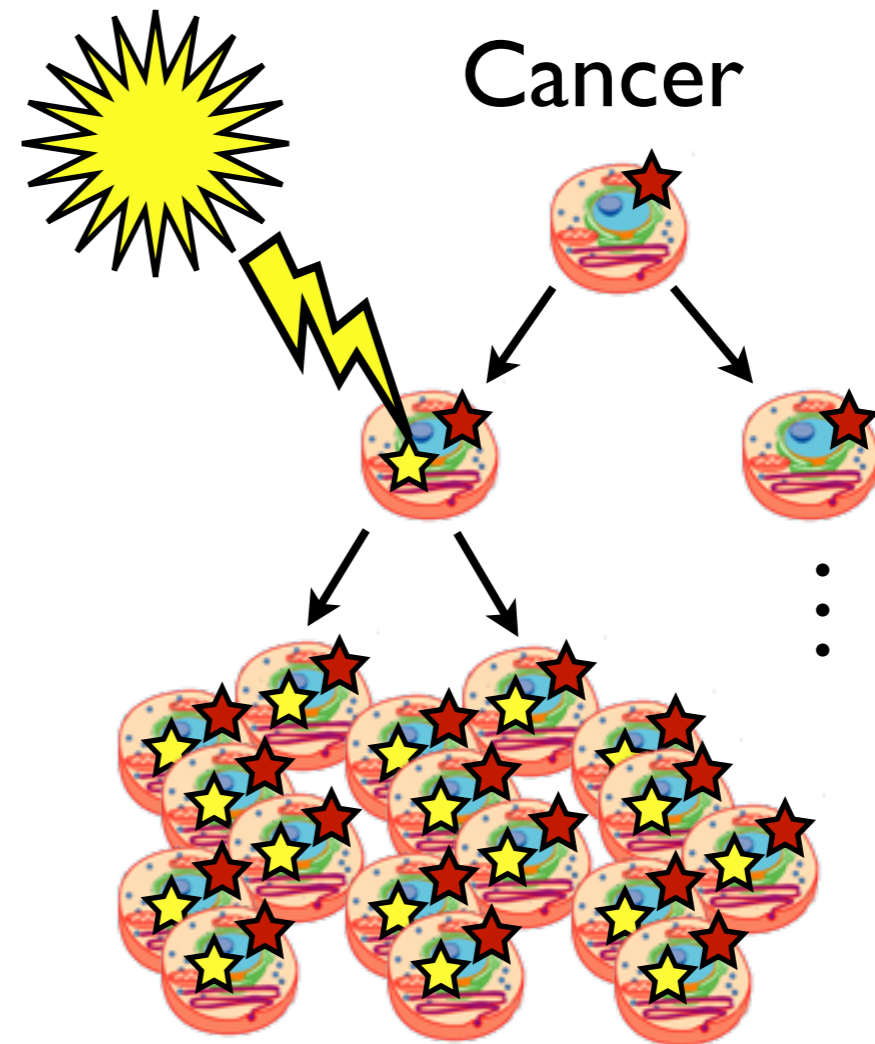
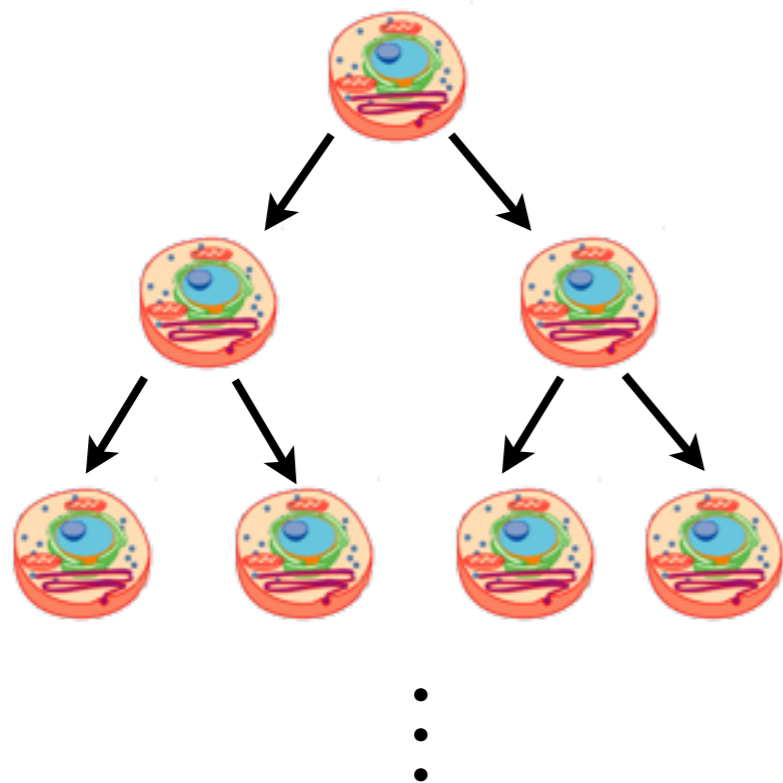# Mechanics of Cancer

Normal Cell Division

Cancer

★ Inherited Mutation

★ Somatic Mutation

Mechanics of Cancer

Normal Cell Division

Cancer
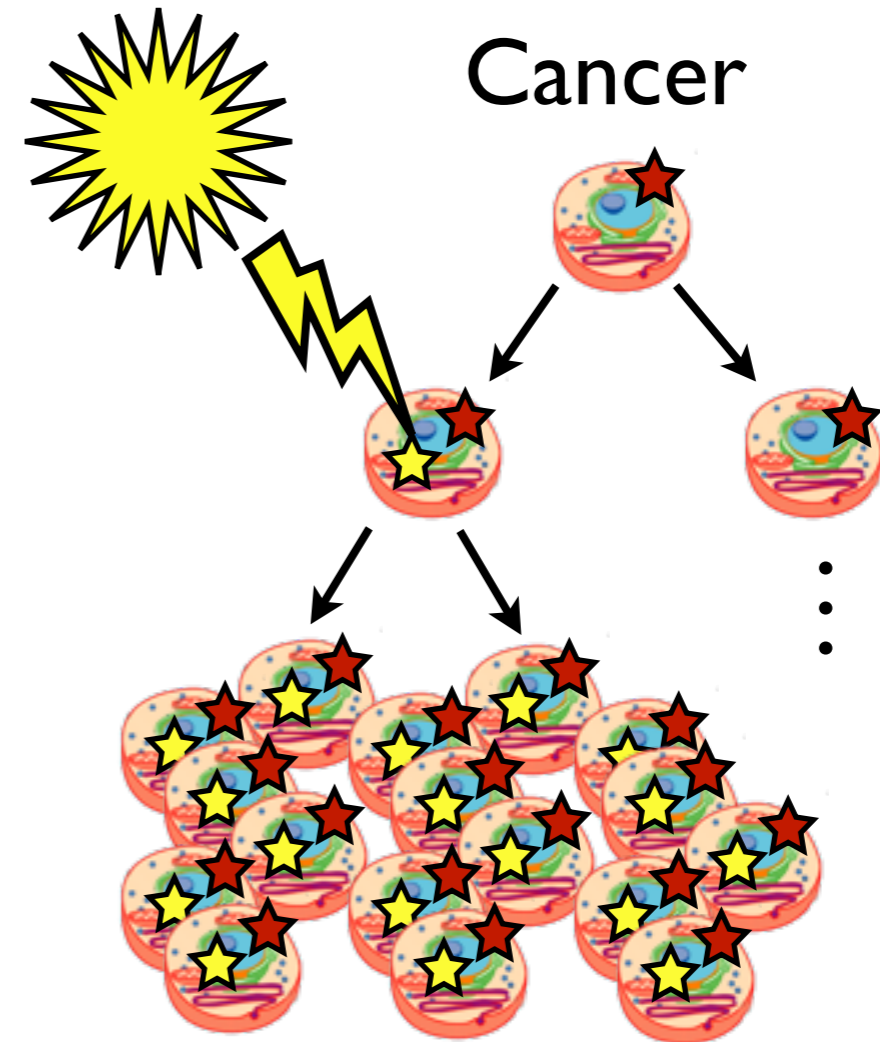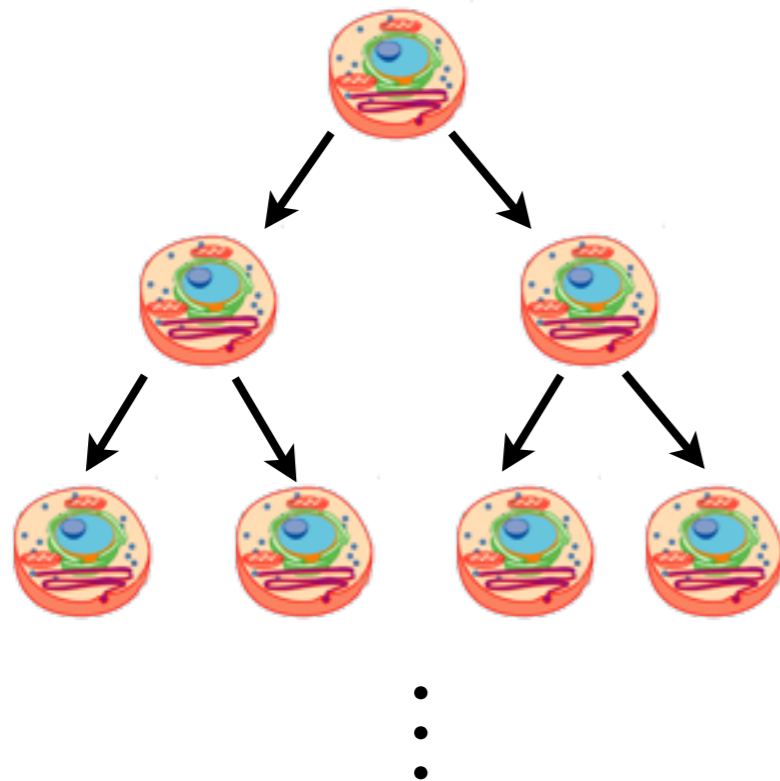
★ Inherited Mutation

★ Somatic Mutation

Another Possibility: Over-activation of an oncogene

Treatment Options

Treatment Options

- Classical Treatments

  - Surgery
  - Chemotherapy

Treatment Options

- Classical Treatments

  - Surgery
  - Chemotherapy

- New Kids on the Block: Targeted Therapies

  - Tamoxifen (1977)
  - Trastuzumab (Herceptin) (1998)
  - 30+ targeted therapies now FDA approved

Data Processing
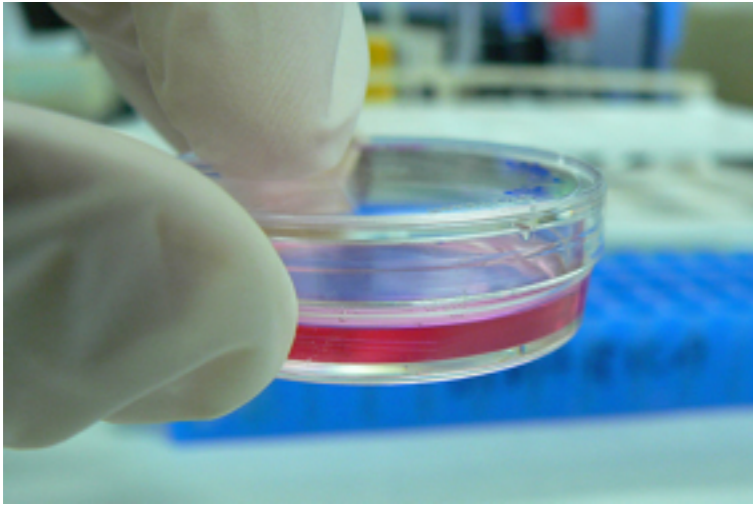
Single Gene Assays

~1 datapoint

Data Processing
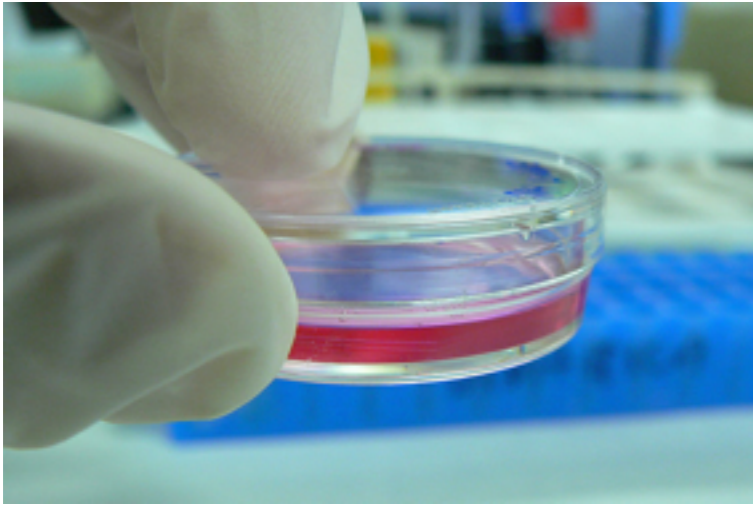
Single Gene Assays
~1 datapoint
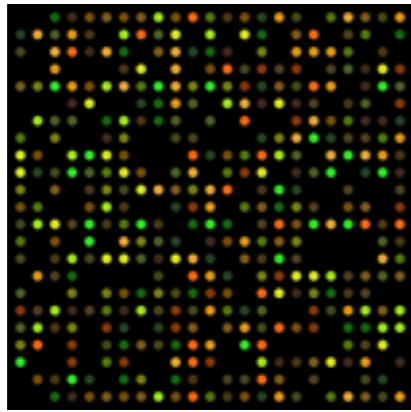
Microarray Revolution (1990)
~1 million datapoints

Data Processing
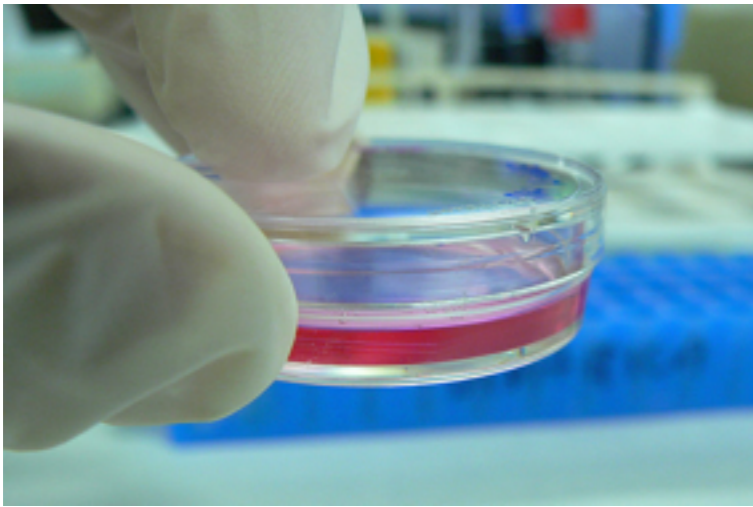
Single Gene Assays
~1 datapoint
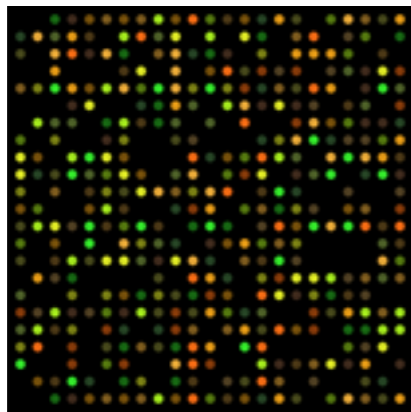
Microarray Revolution (1990)
~1 million datapoints

Sequencing Revolution (2010)
>200 billion datapoints

Image Credits: Wikimedia Foundation

Data Processing

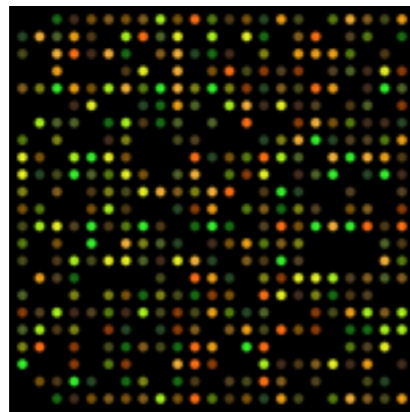Single Gene Assays

~1 datapoint

Microarray Revolution (1990)

~1 million datapoints

Sequencing Revolution (2010)

>200 billion datapoints

How do we make this information clinically actionable?

Approach the Problem with Pathways
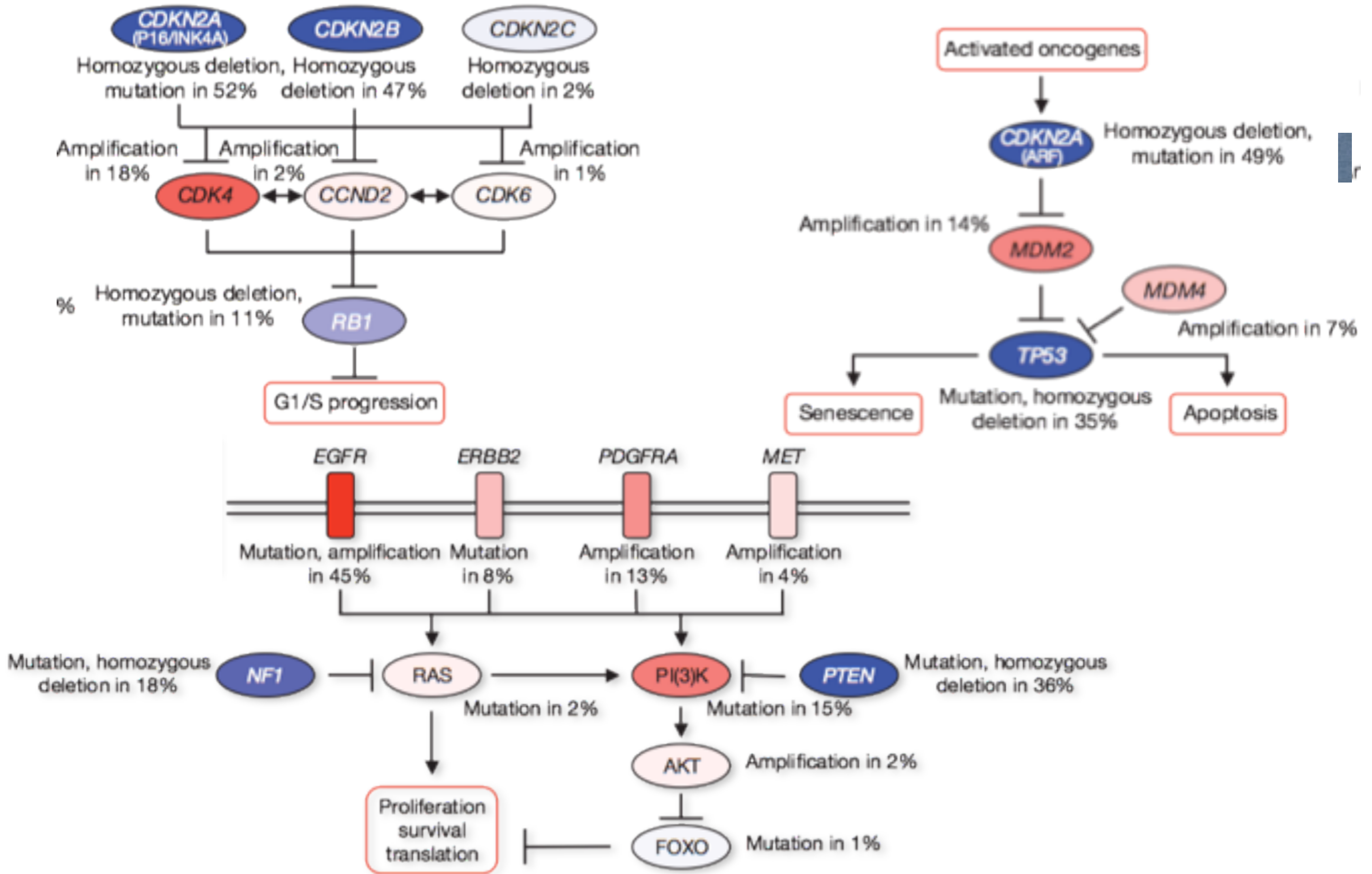
The Cancer Genome Atlas, *Nature*, 2008

Approach the Problem with Pathways

The Cancer Genome Atlas, *Nature*, 2008
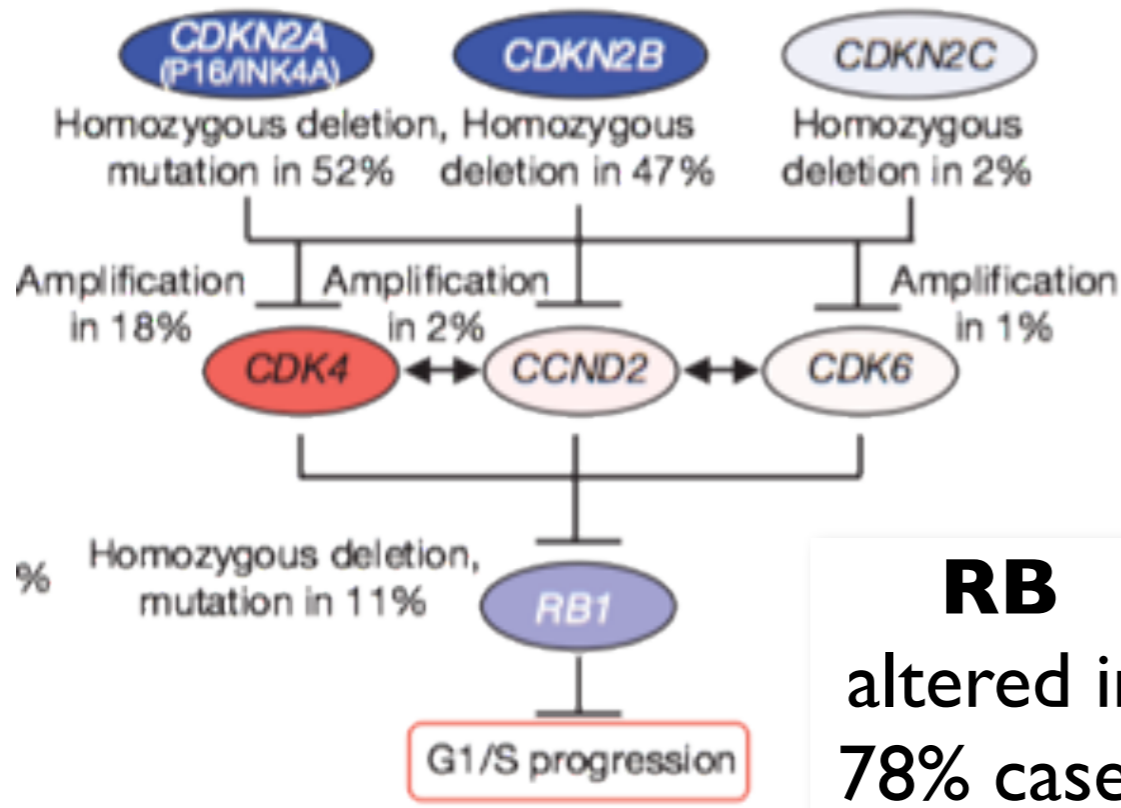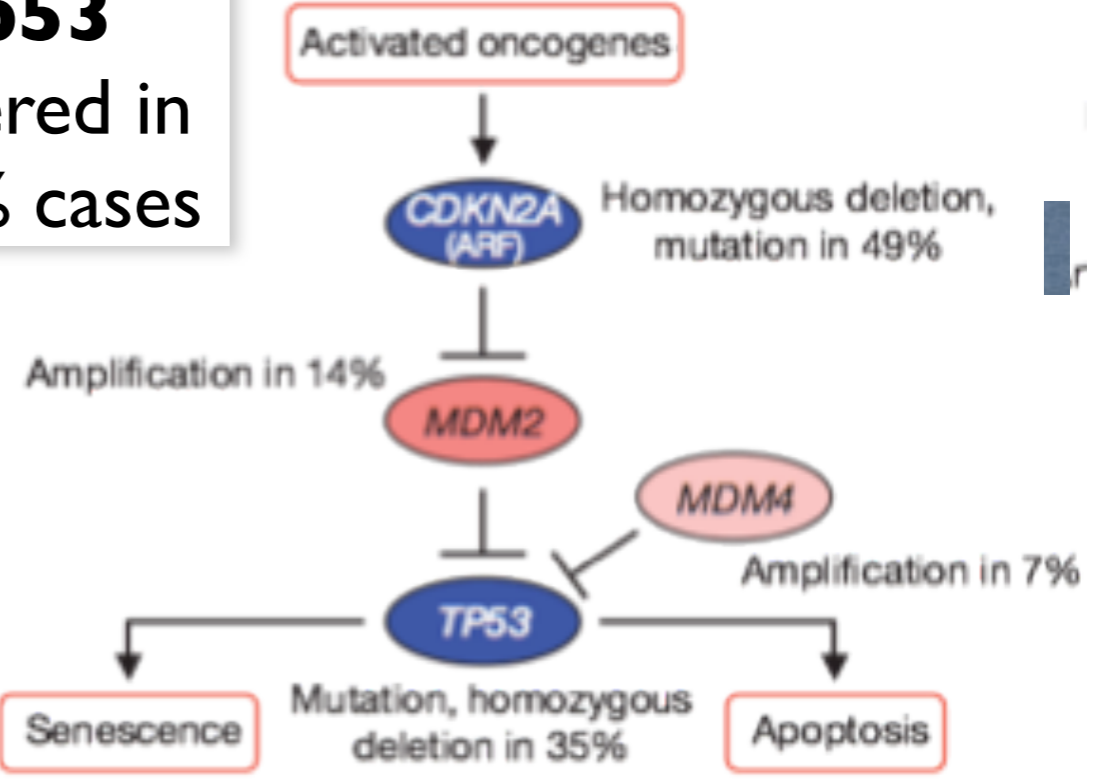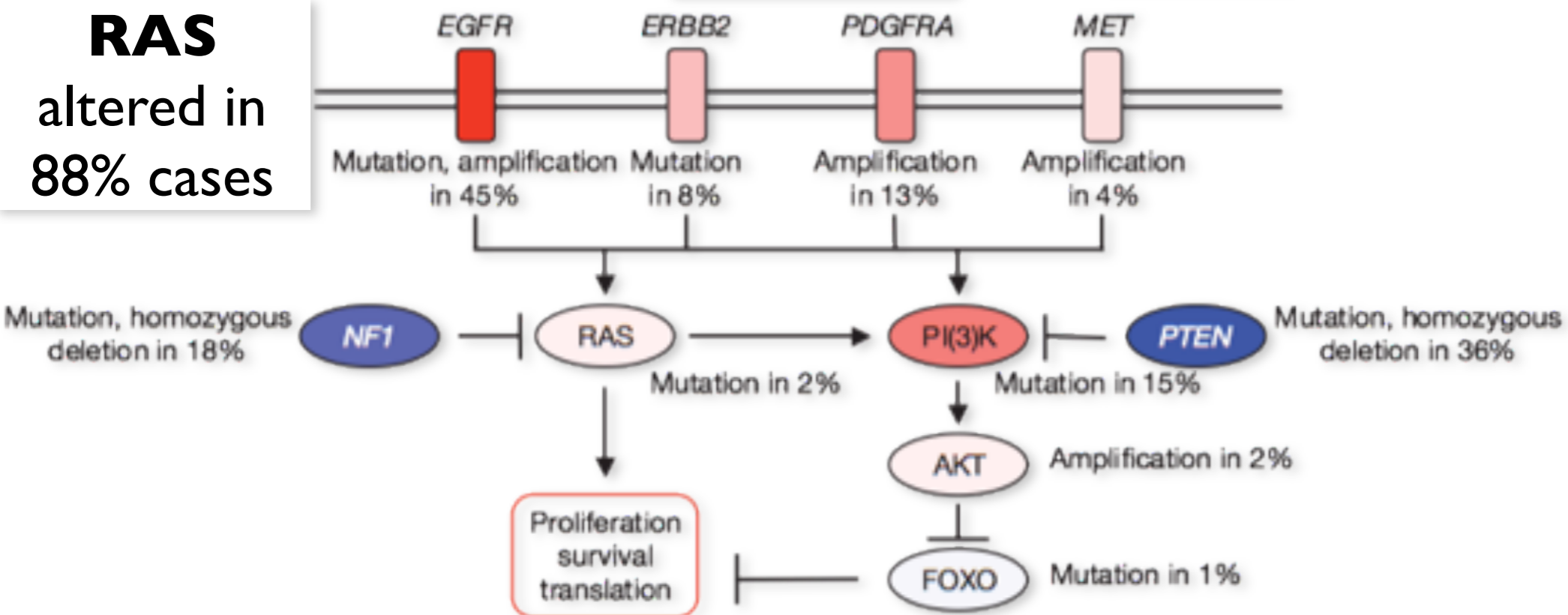
Analyzing Pathways

- Overlay data on curated pathway structure

  - (TCGA Research Network, 2008)

- Treat pathways as genesets (bags of genes)

  - GSEA (Subramanian et al., 2005)

- Bayesian Networks

  - (Segal et al., 2001)

Factor Graphs

Bipartite graph
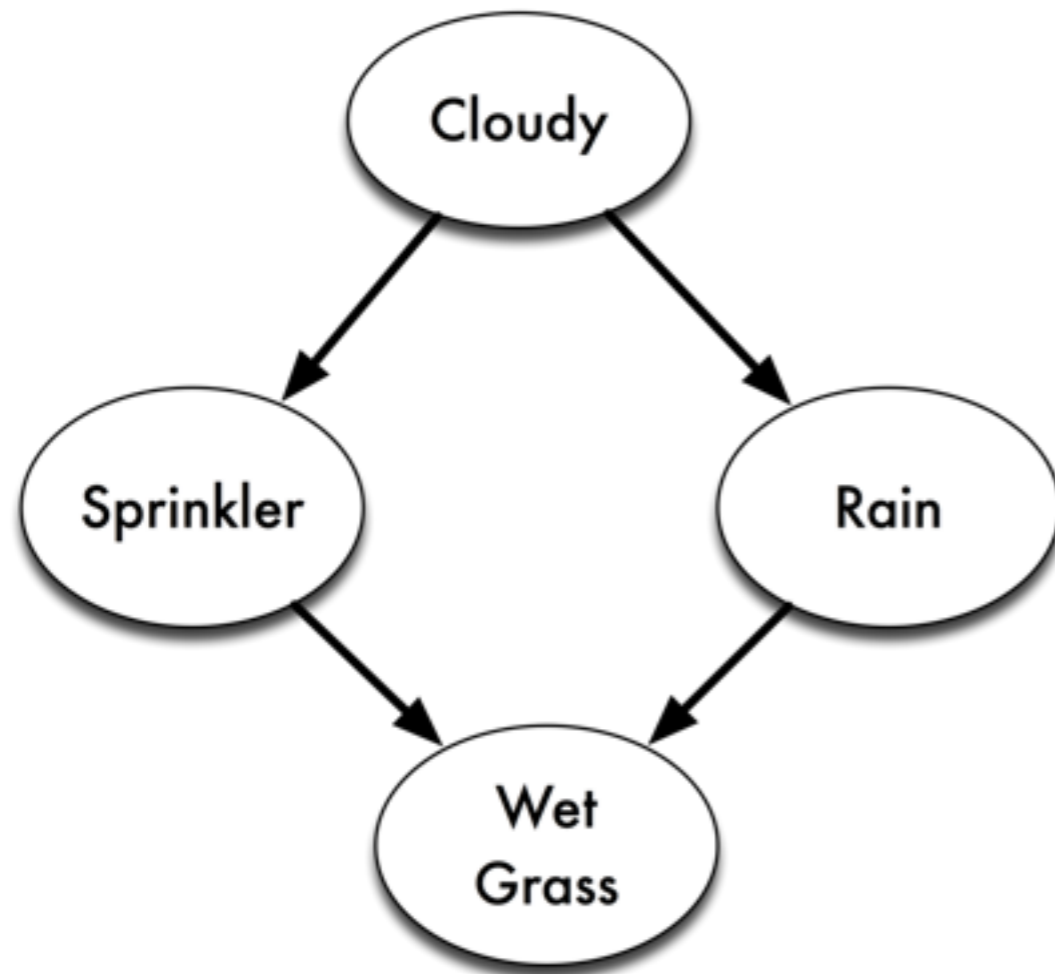
Generalization of Bayes and Markov Nets

Factor Graphs

Bipartite graph

Generalization of Bayes and Markov Nets
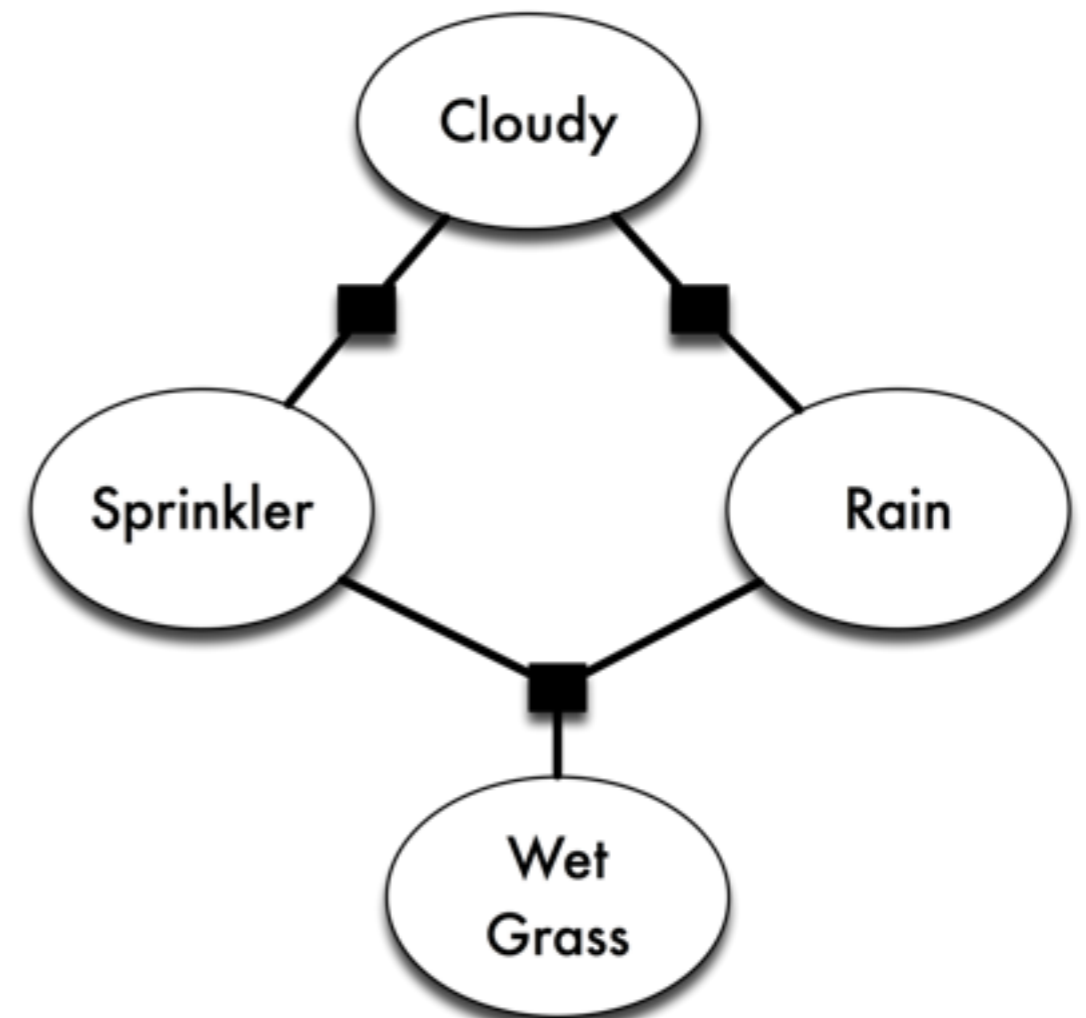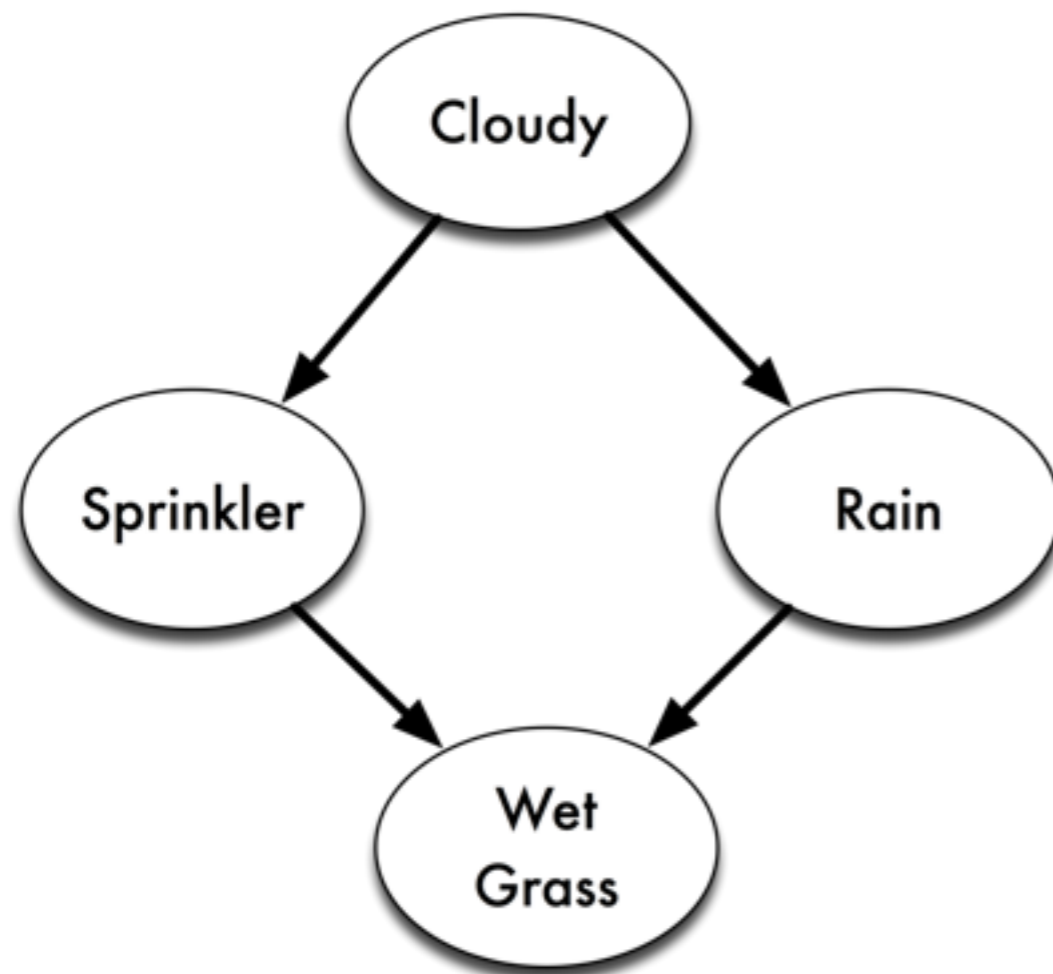
Factor Graphs

Bipartite graph

Generalization of Bayes and Markov Nets
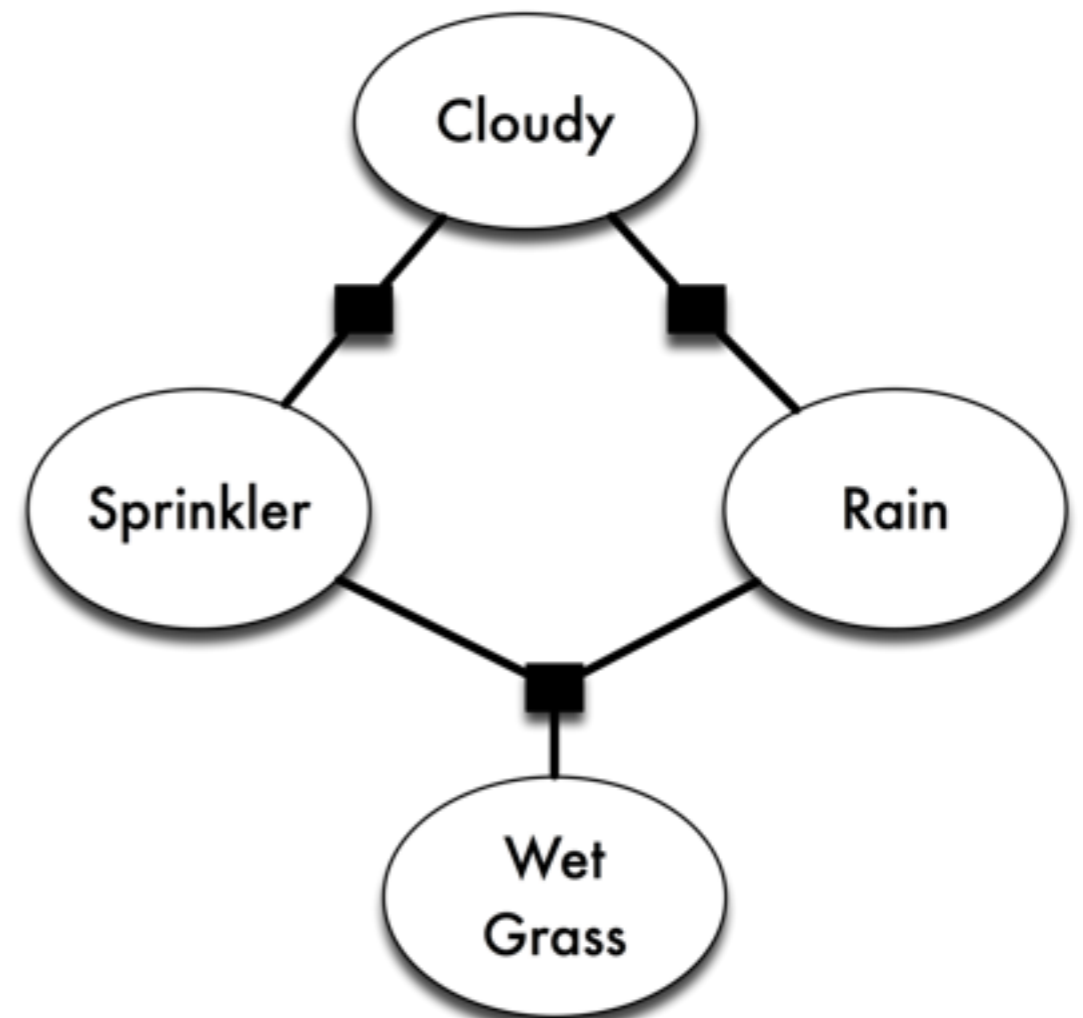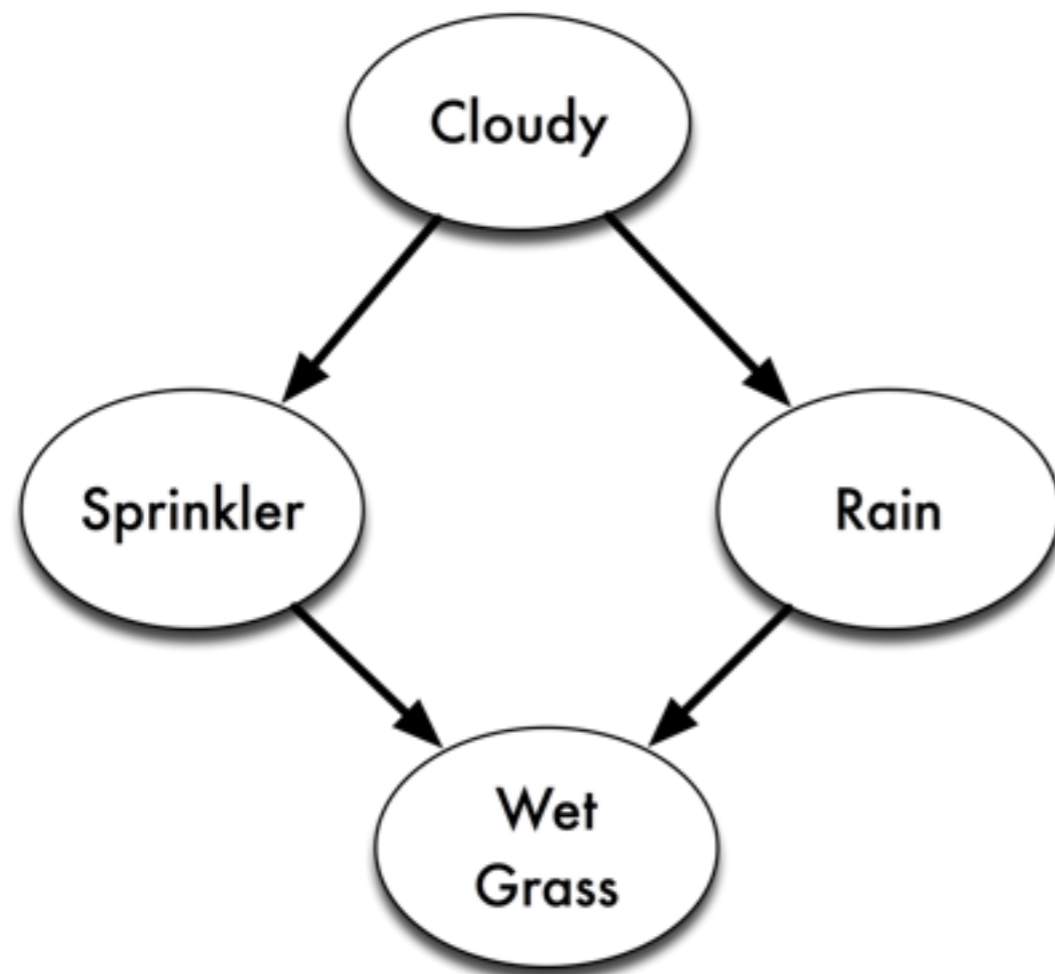
Factor Graphs have been used successfully to model pathways, including thesis work of Charles Vaske

Outline

I. Introduction

II. **Previous Work**

III. Aim 1: PARADIGM - Integrated Pathway Analysis

IV. Aim 2: Modeling pathways using sequence data

V. Aim 3: Directing combination therapies

UCSC Cancer Genomics Browser
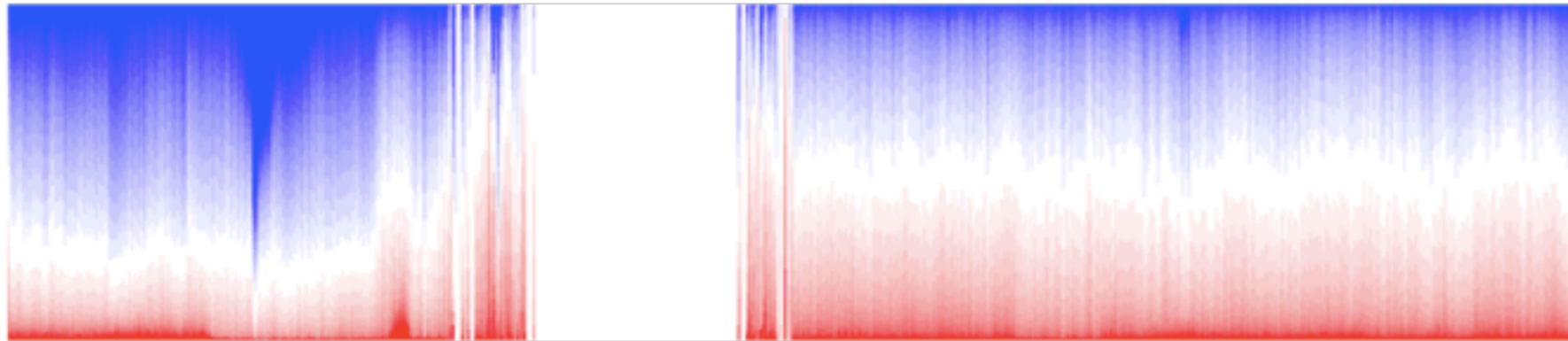
Pancreatic Cancer (Jones et al. 2008)

Glioblastoma multiforme (TCGA 2008)

Melanoma (Lin et al. 2008)

chr9   25000000| 50000000| 75000000| 100000000| 125000000|

p23   21.3   q12   21.13   31.1   32 33.1

Chromosome 9

http://genome-cancer.ucsc.edu
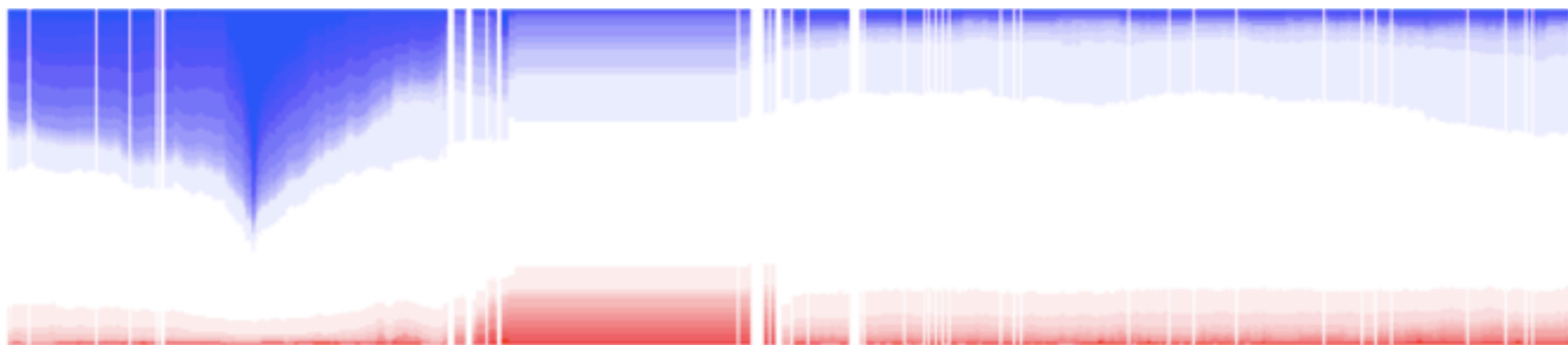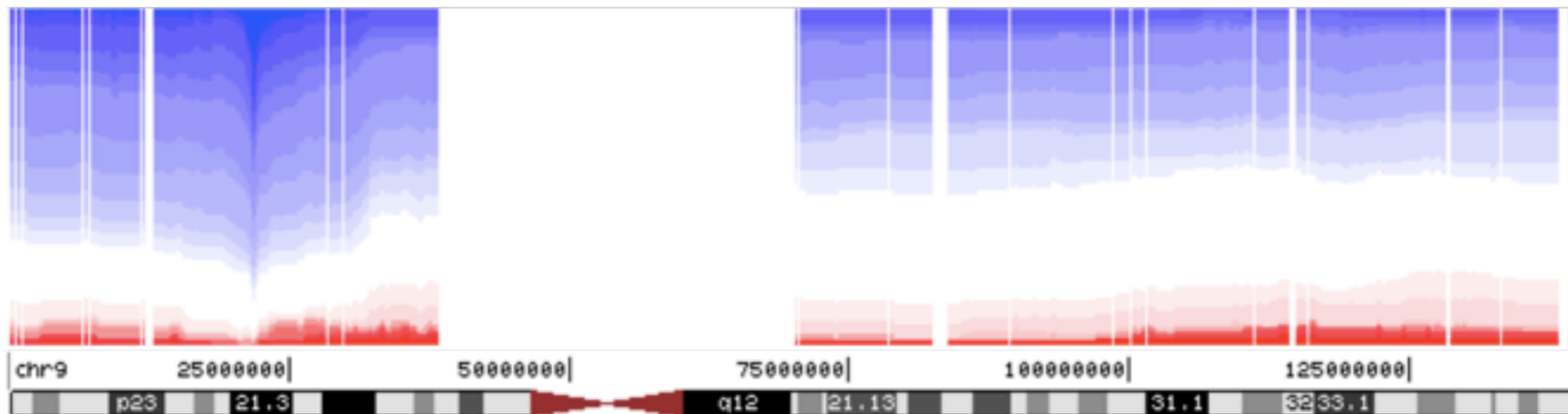
UCSC Cancer Genomics Browser

Pancreatic Cancer (Jones et al. 2008)
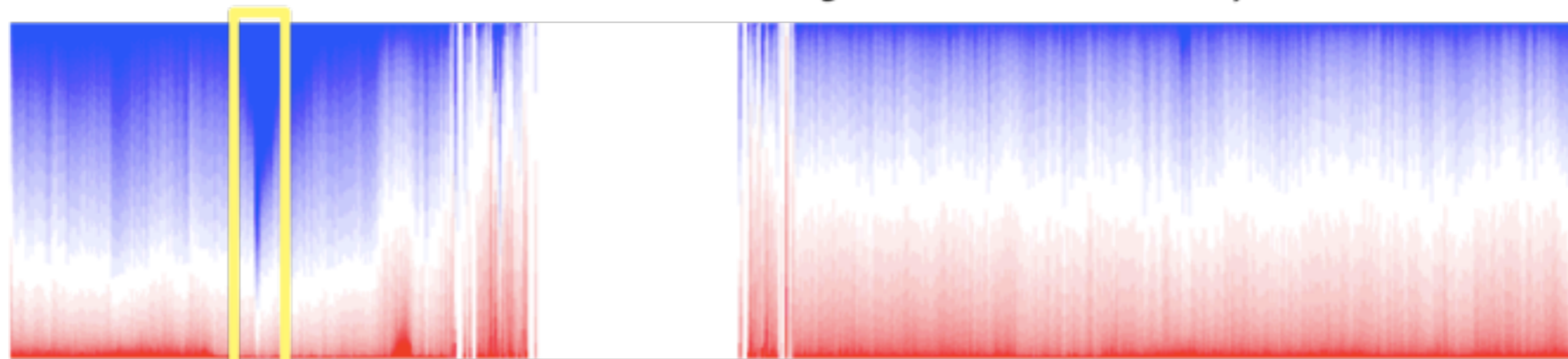
Glioblastoma multiforme (TCGA 2008)
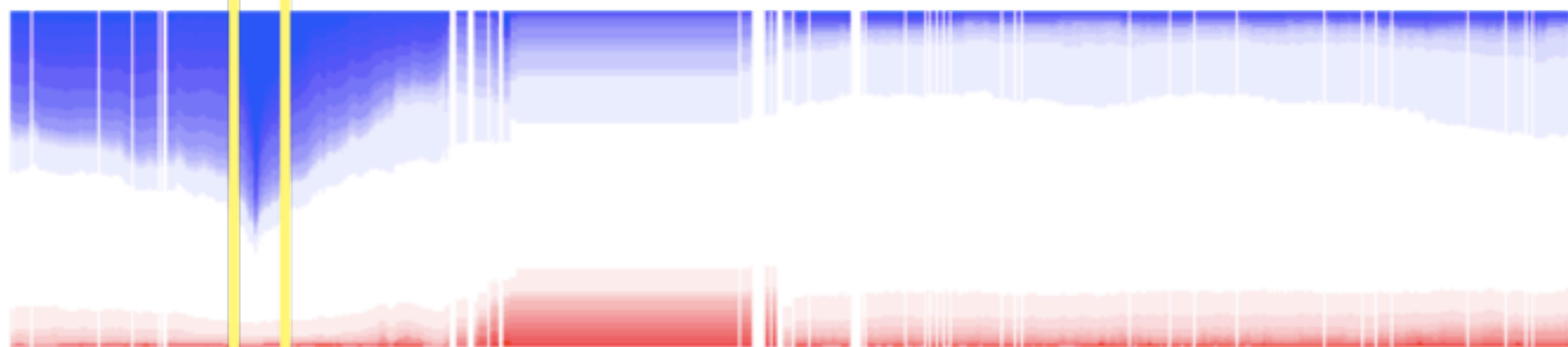
Melanoma (Lin et al. 2008)

CDKN2A/B

Chromosome 9

http://genome-cancer.ucsc.edu

Visualizing Cancer Genesets

Brain tumor somatic non-silent mutations

Brain tumor germline non-silent mutations

Brain tumor amplified / deleted regions

a

b

c

Normal

Tumor

d

p53 pathway

RB pathway

RTK/Ras/PI3K pathway

Control

Normal/Tumor

Dead/Alive

Male/Female
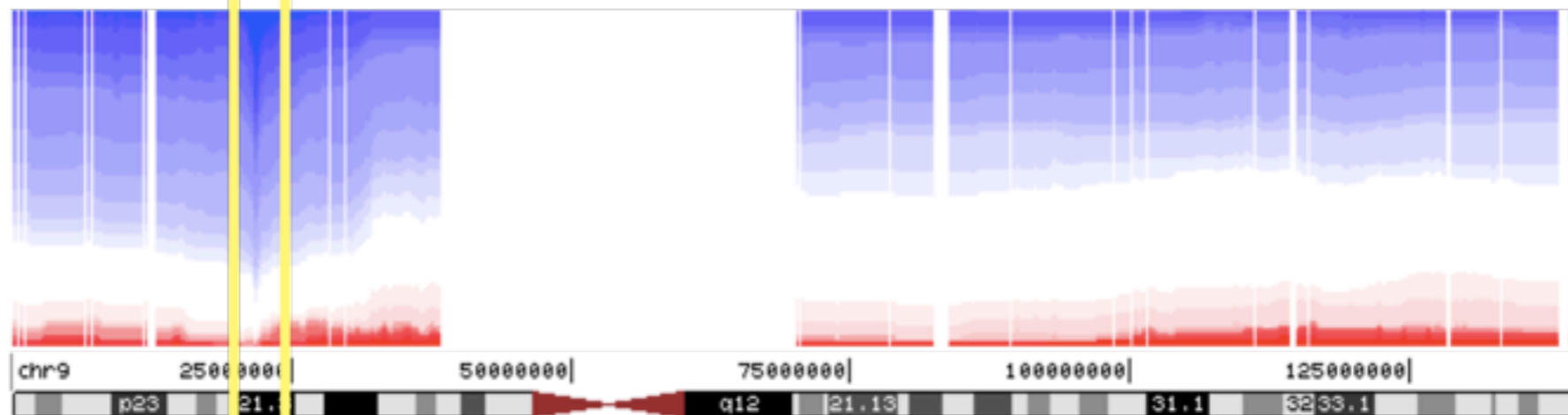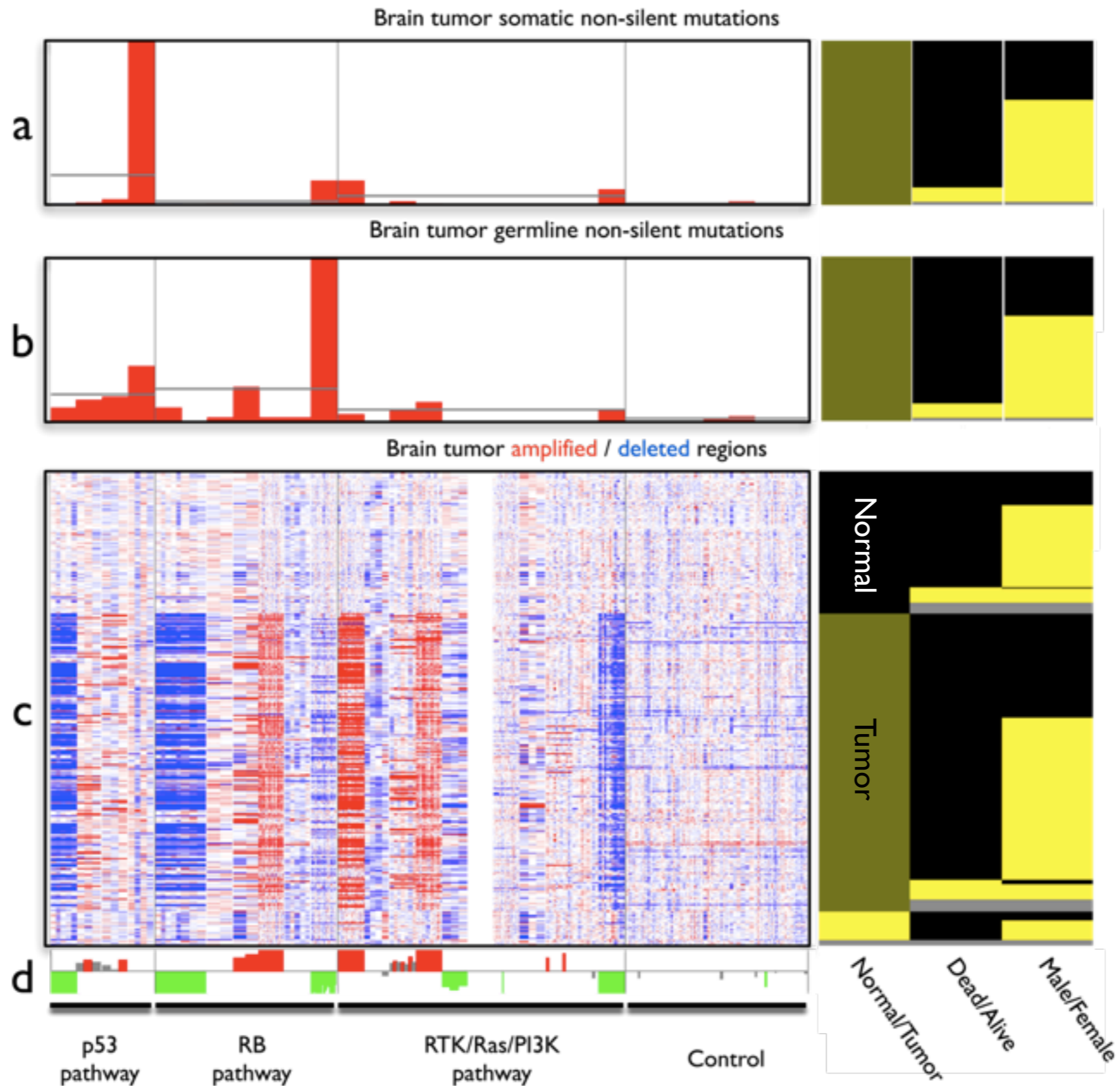
Visualizing Cancer Genesets
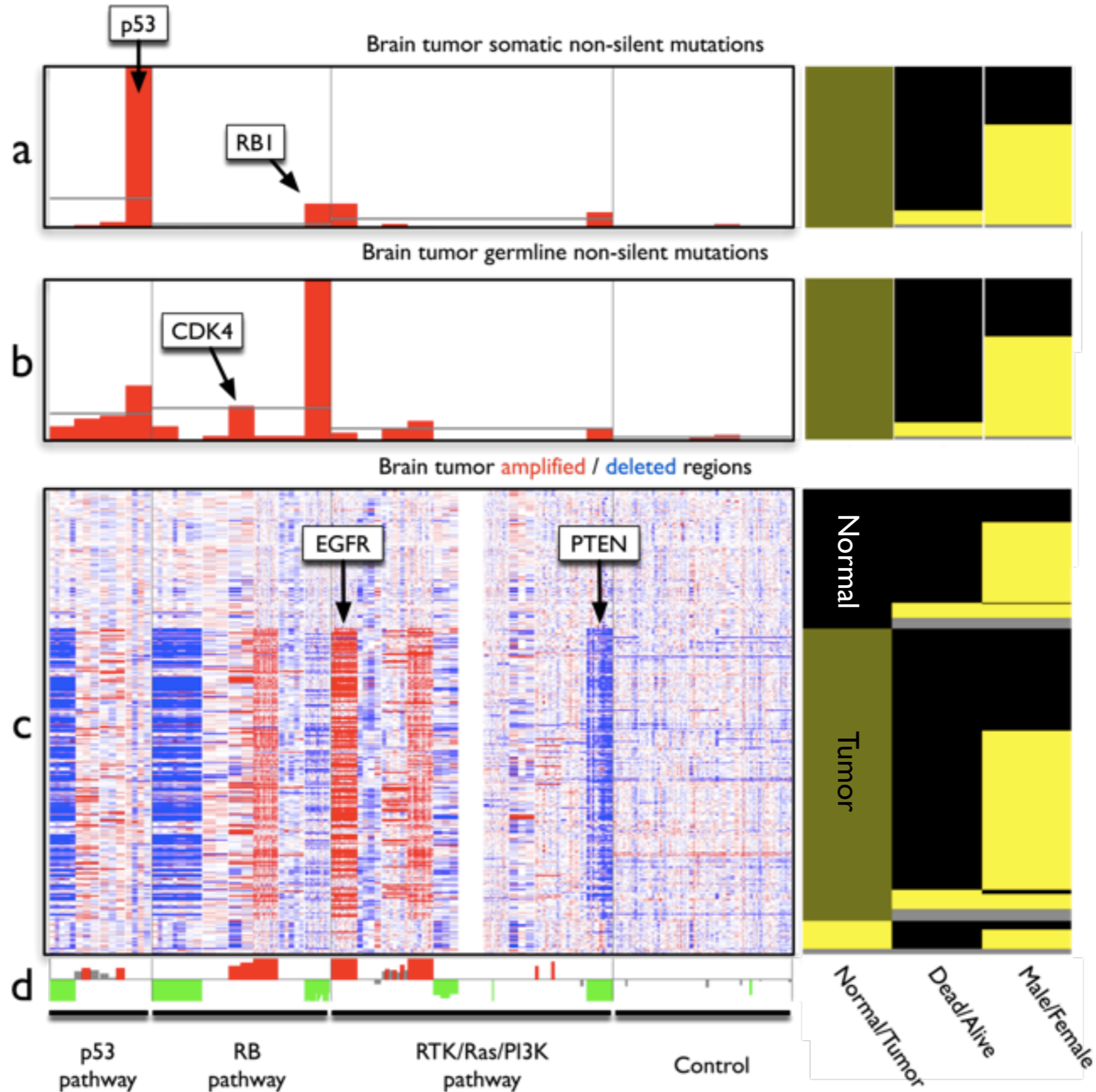
UCSC Cancer Genomics Browser
genome-cancer.ucsc.edu

TCGA
The Cancer Genome Atlas
https://tcga1.cse.ucsc.edu

I-SPY TRIAL
Breast Cancer Clinical Trial
https://instinct1.cse.ucsc.edu

SU2C Dream Team
Stand Up to Cancer
https://genome-cancer-su2c.cse.ucsc.edu

UCSC Cancer Genomics Browser

ENCODE

Open Source:
First public deposit Oct. 2007
Public access portal:
All major published studies

Spin-off Browsers
Immuno Browser
Stem Cell Browser
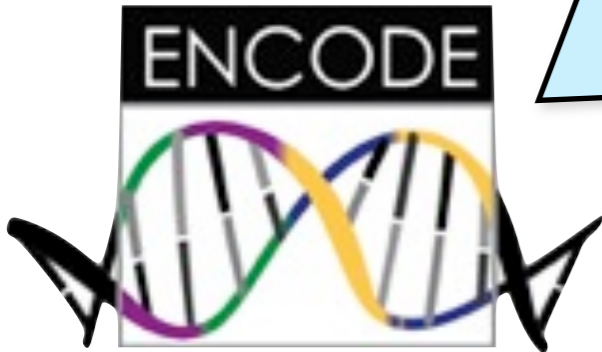
Integrated Data

I-SPY TRIAL
Breast Cancer Clinical Trial
https://instinct1.cse.ucsc.edu
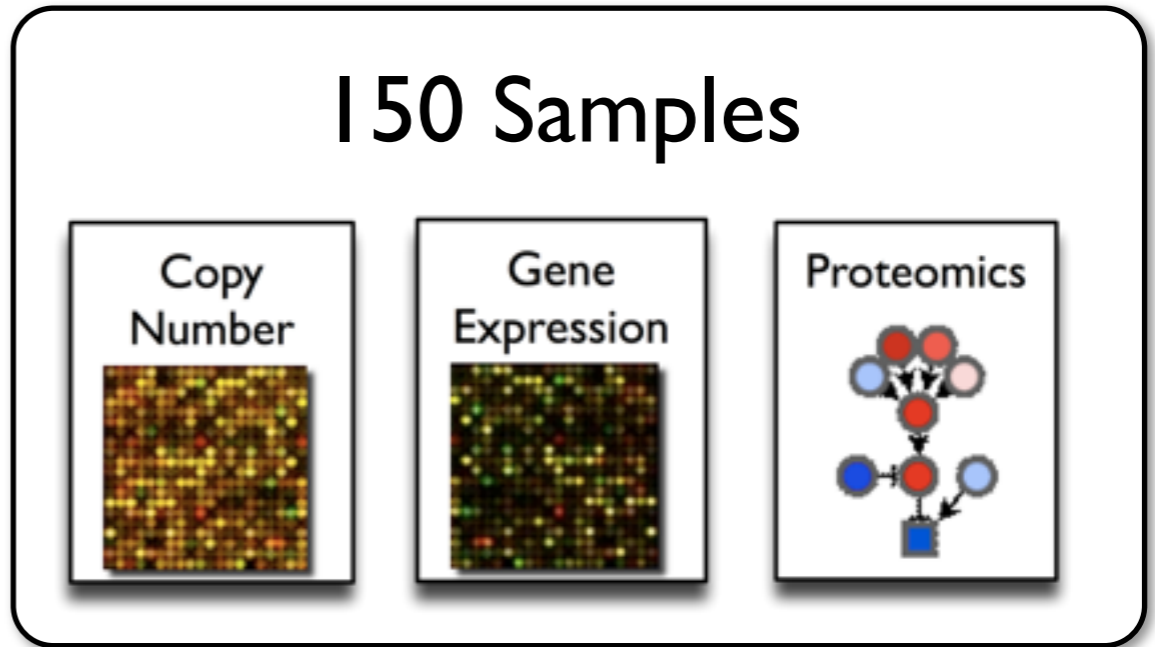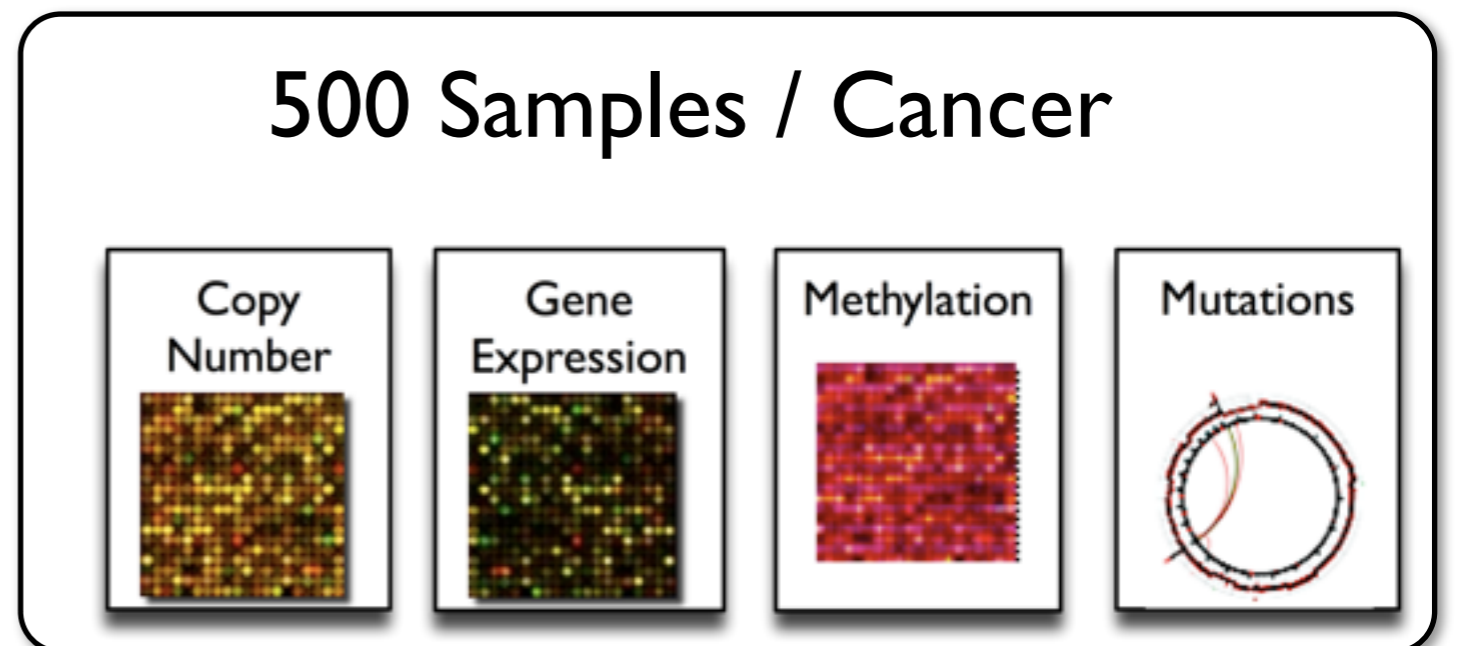
150 Samples

Copy Number

Gene Expression

Proteomics

Integrated Data

BioIntegrator: Pipeline for Automated Cancer Sample Analysis

Database backend for storage efficiency and normalization

BioIntegrator: Pipeline for Automated Cancer Sample Analysis

Database backend for storage efficiency and normalization

Web Frontend for Data Visualization:
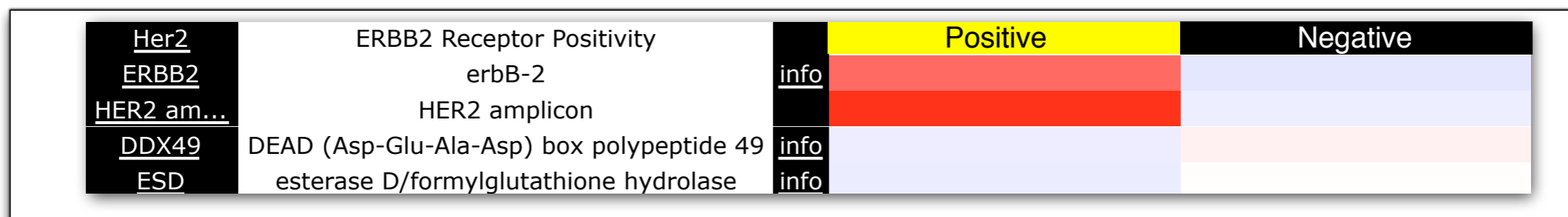
A

B

BioIntegrator: Meta-Analysis Module
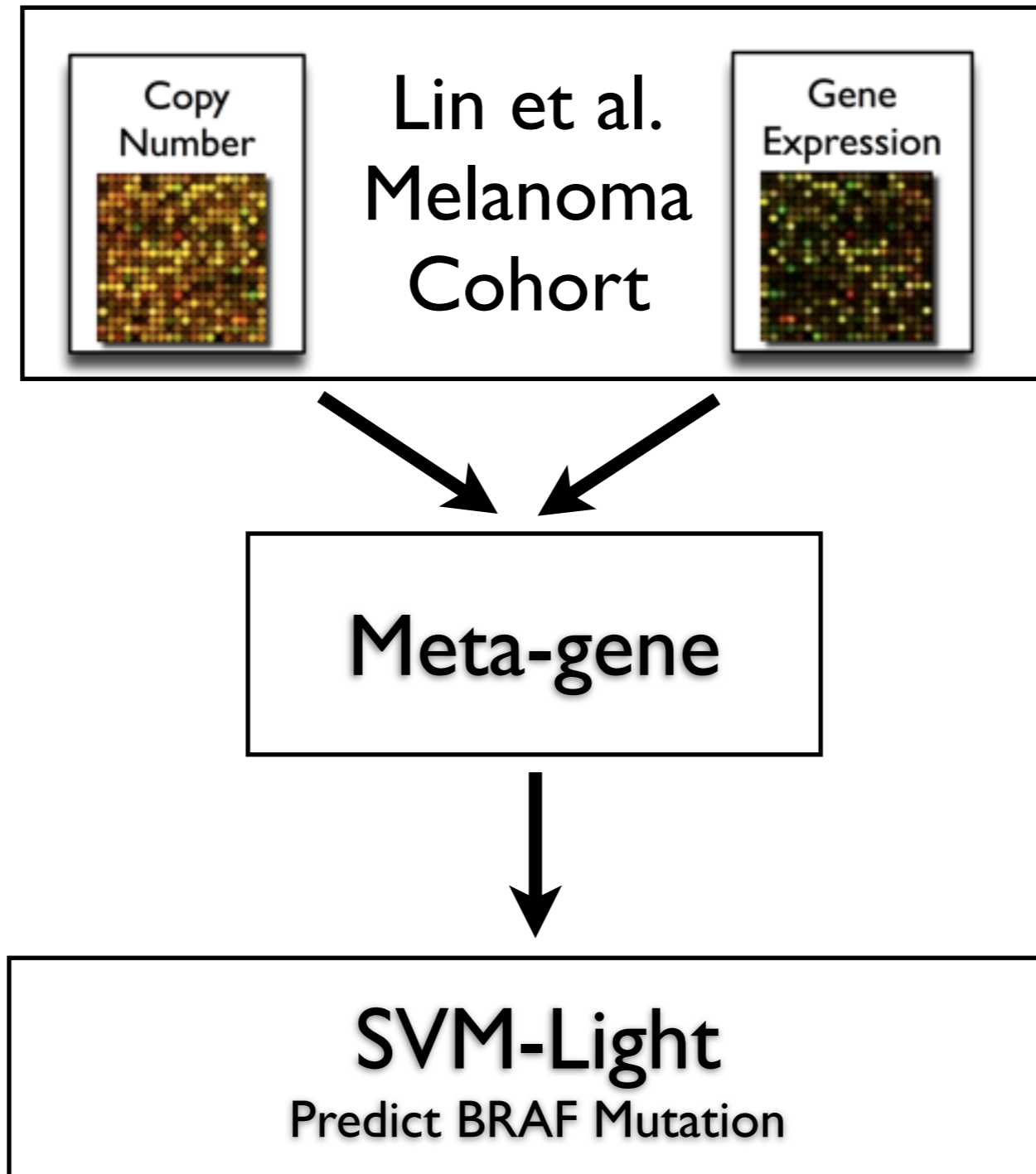
BioIntegrator: Meta-Analysis Module
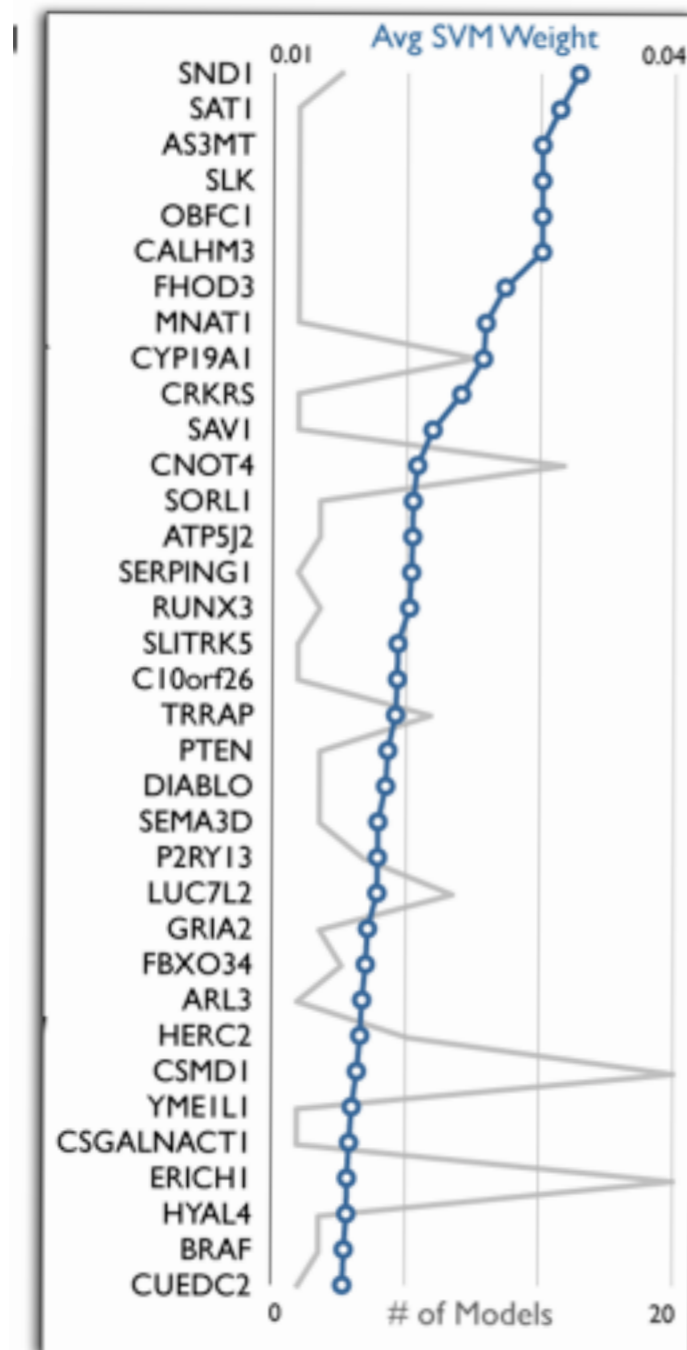
BioIntegrator: Meta-Analysis Module
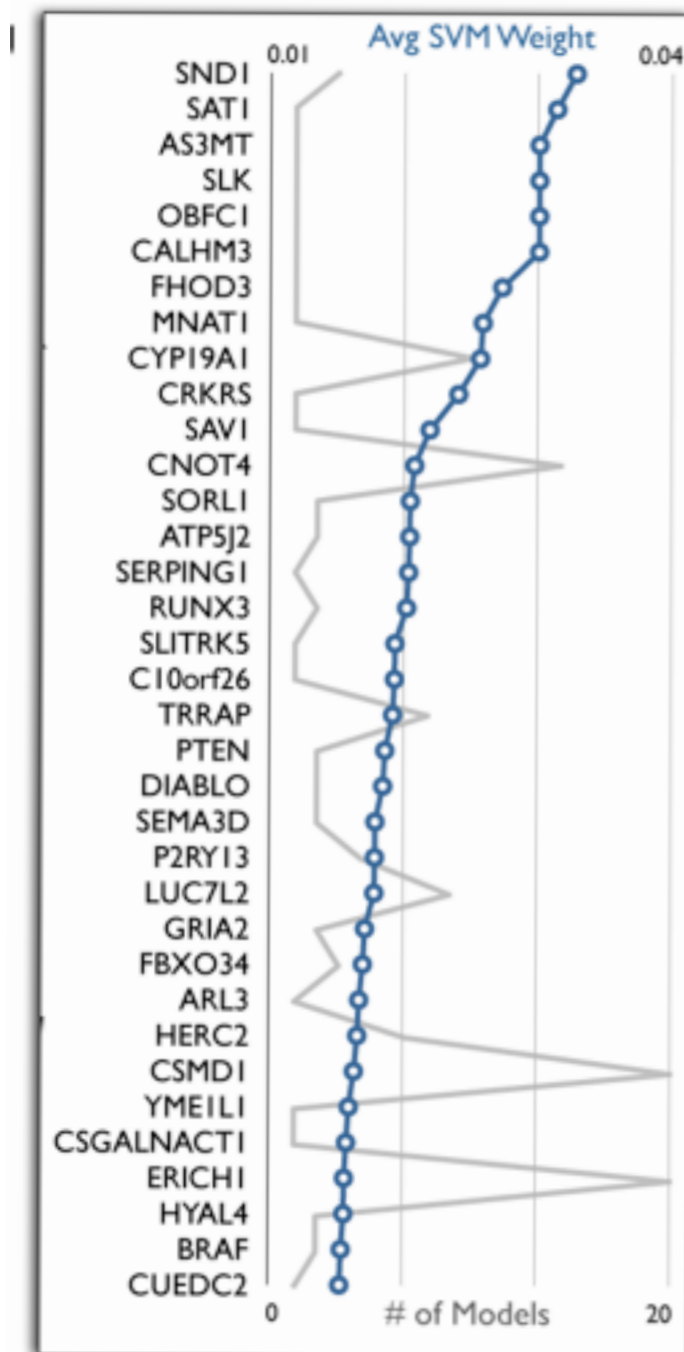
Models > 91% accuracy
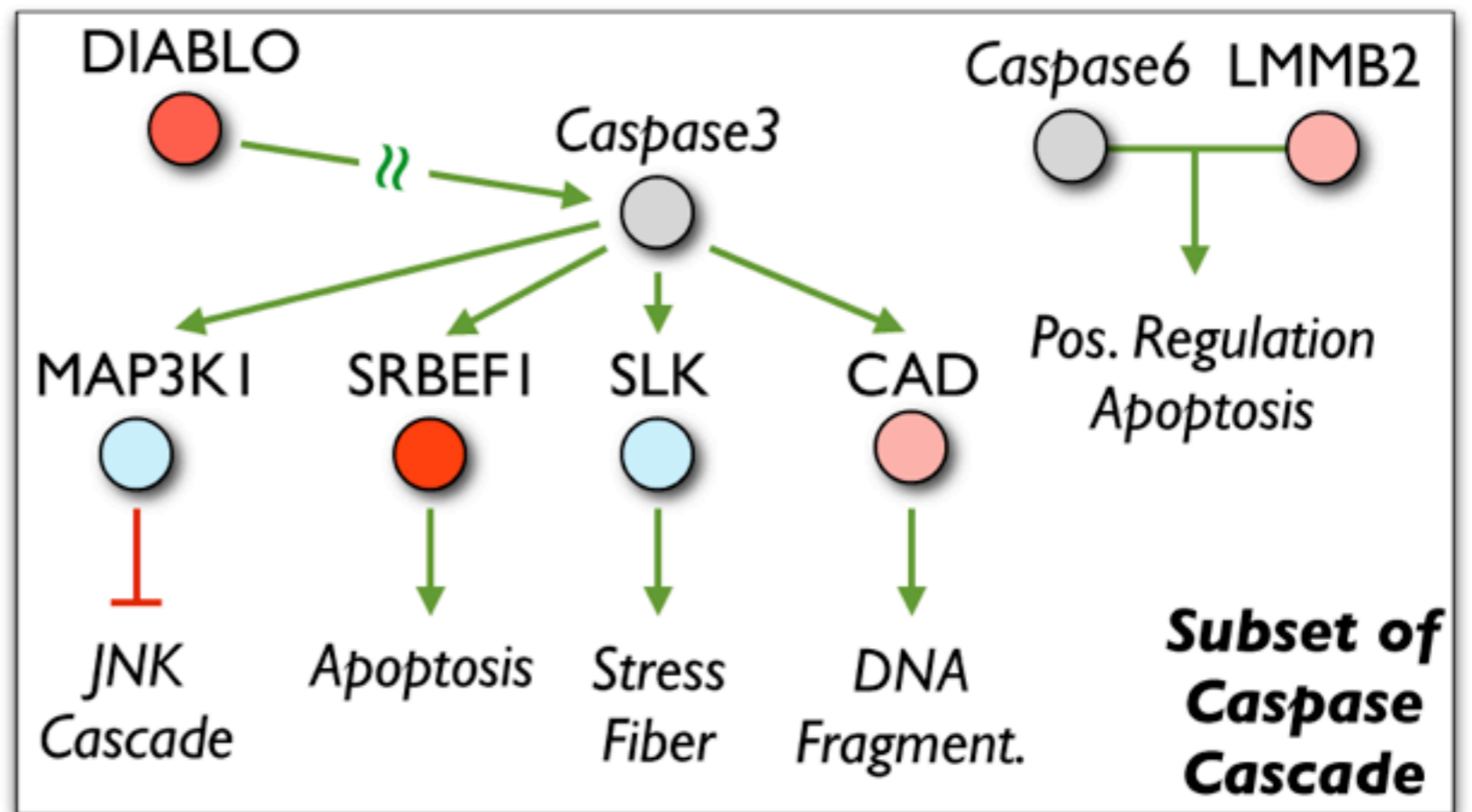
Enrichment: P < 4e-6

Outline

I. Introduction

II. Previous Work

III. Aim 1: PARADIGM - Integrated Pathway Analysis

IV. Aim 2: Modeling pathways using sequence data

V. Aim 3: Directing combination therapies

PAthway Recognition Algorithm using Data Integration on Genomic Models (PARADIGM)
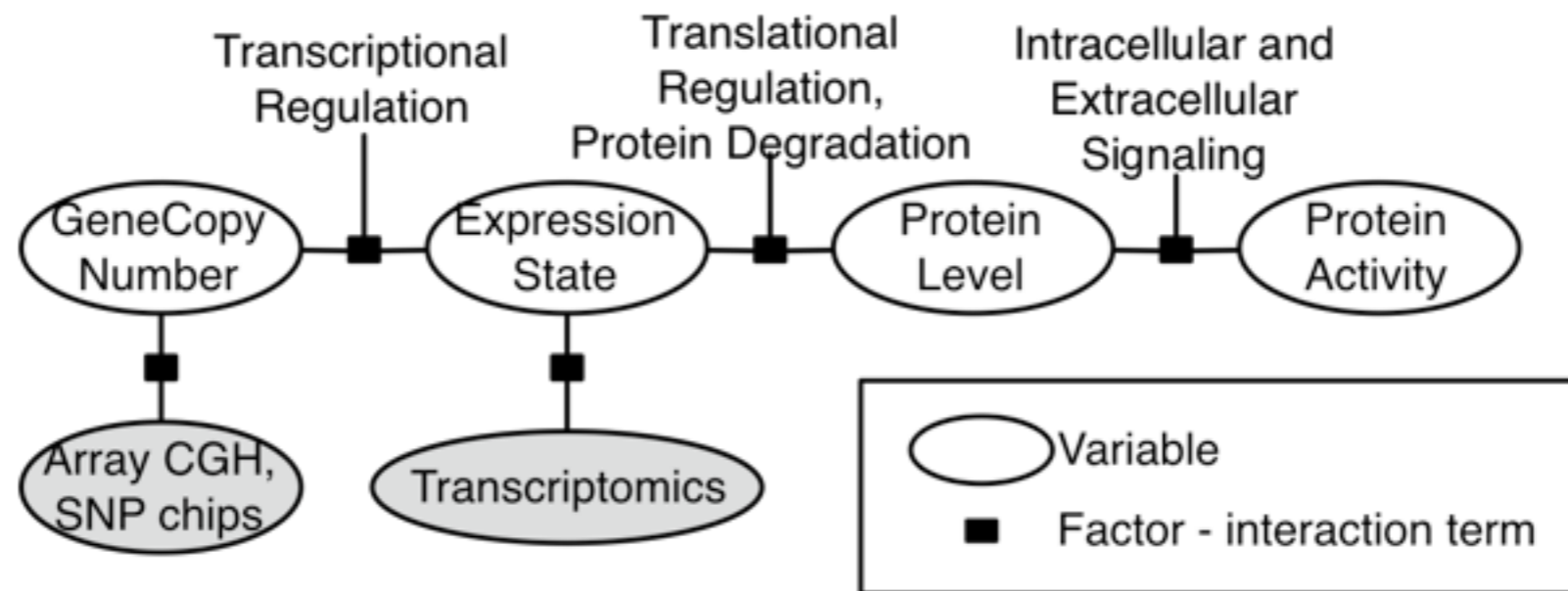
Vaske and Benz et al., Bioinformatics (2010)

PAthway Recognition Algorithm using Data Integration on Genomic Models (PARADIGM)

Vaske and Benz et al., Bioinformatics (2010)
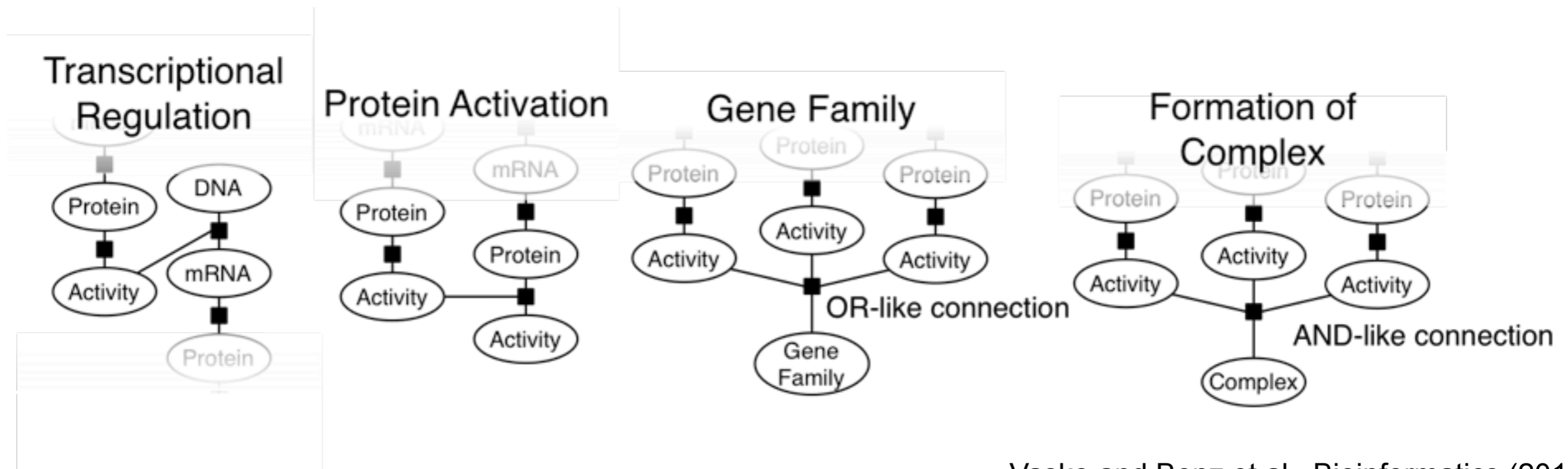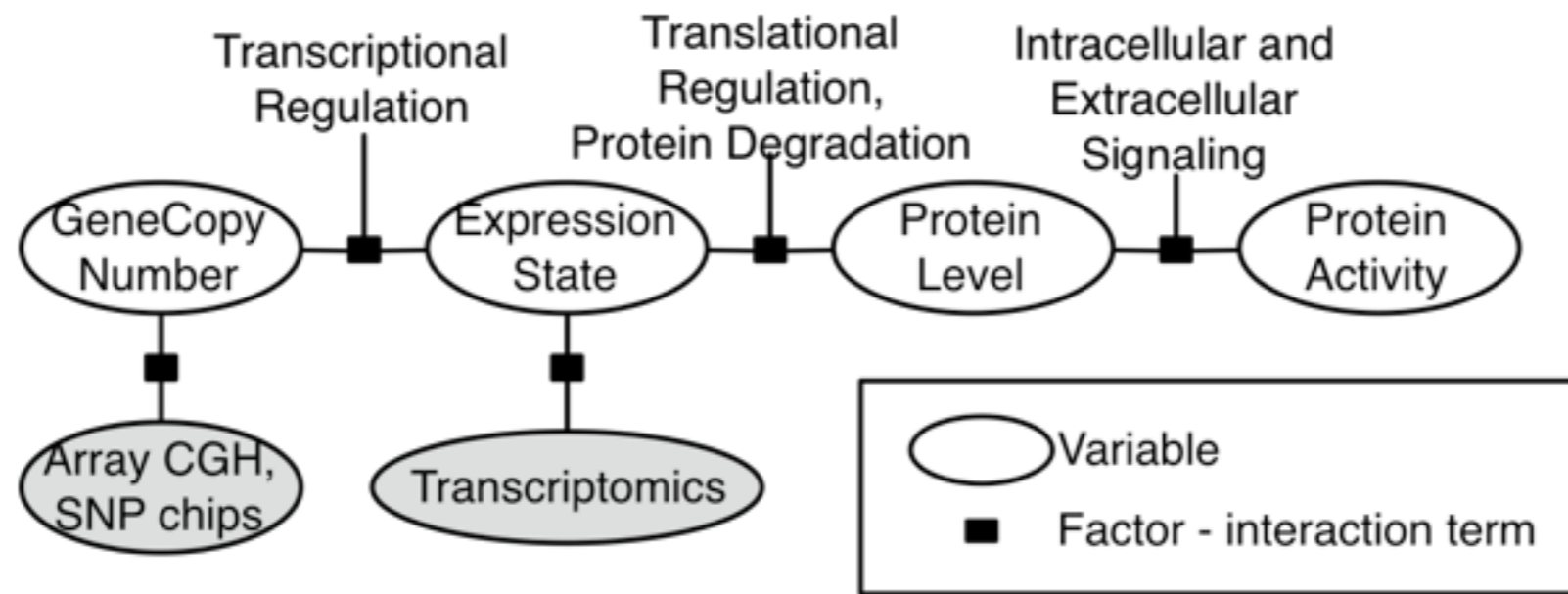
Small Pathway Example
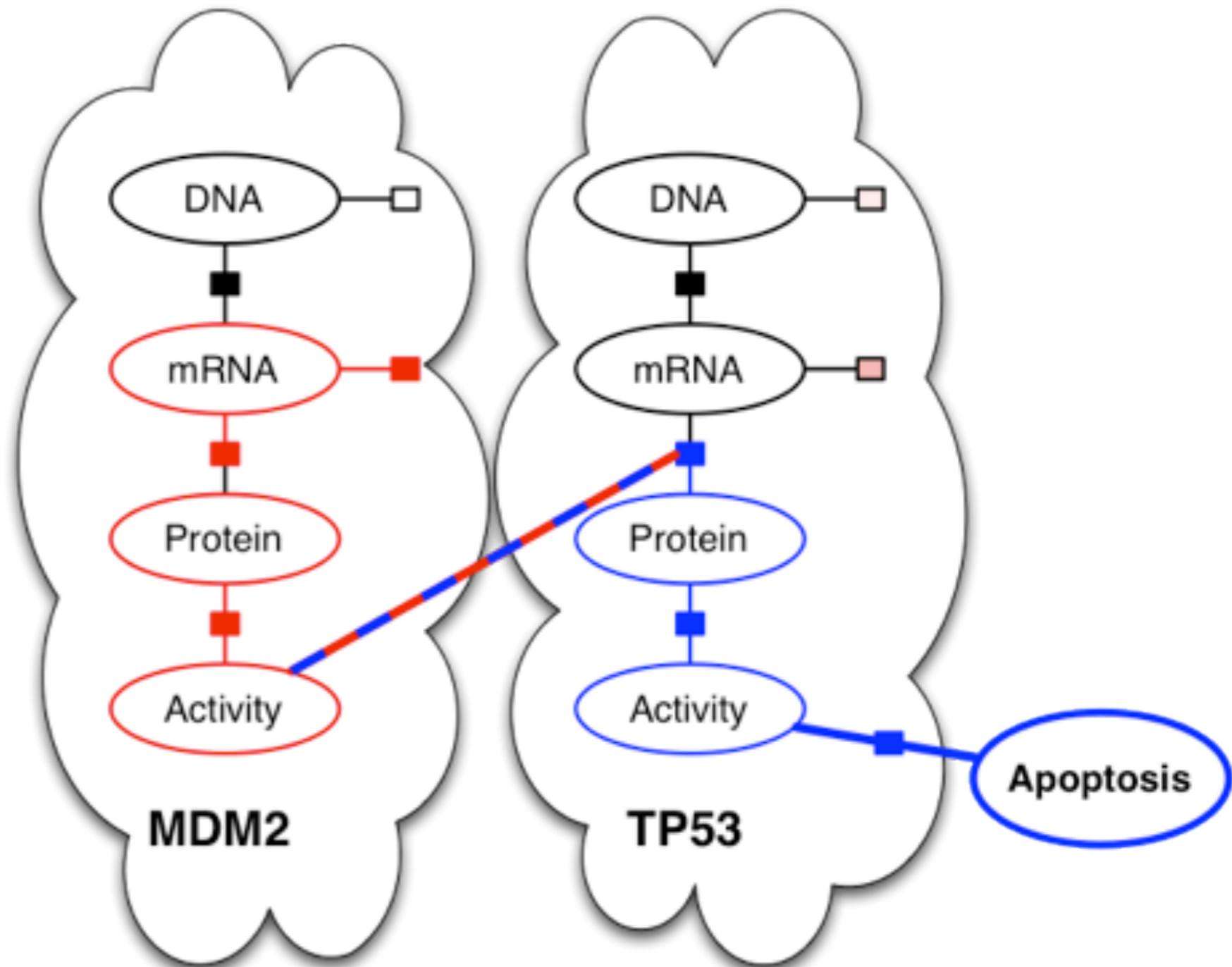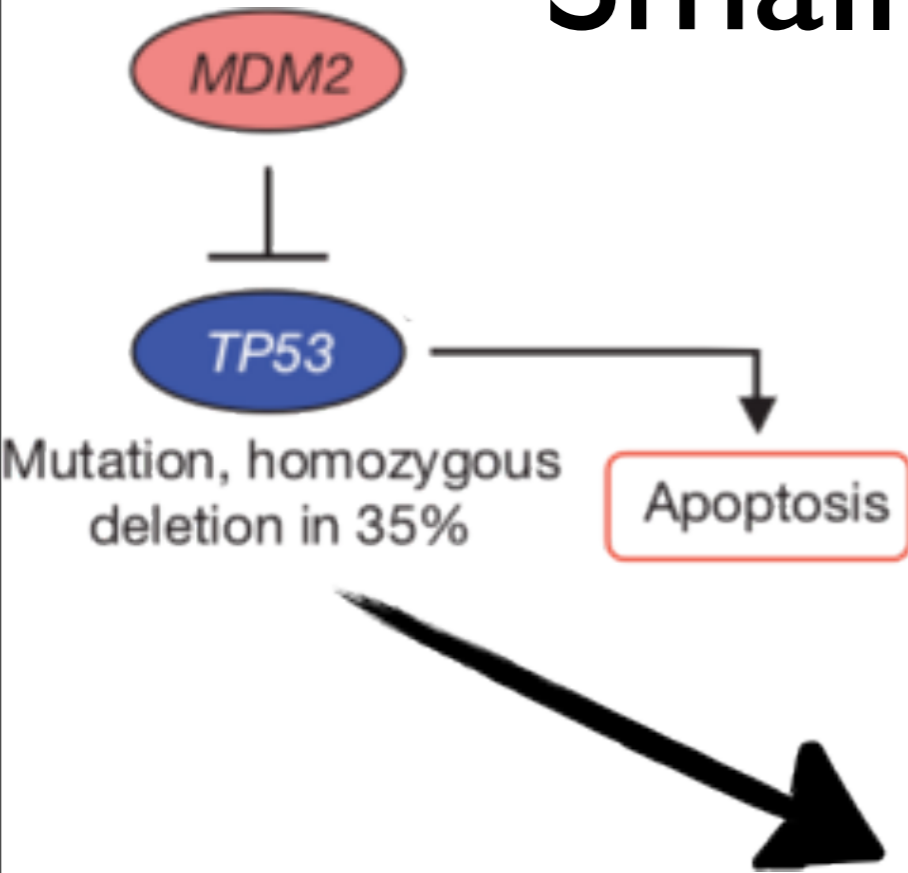
Pathway Schematic

Integrated Pathway Activities (IPAs)

$$L(i,s) = log\left(\frac{P(D, x_i = s | \Phi)}{P(D, x_i \neq s | \Phi)}\right) =$$
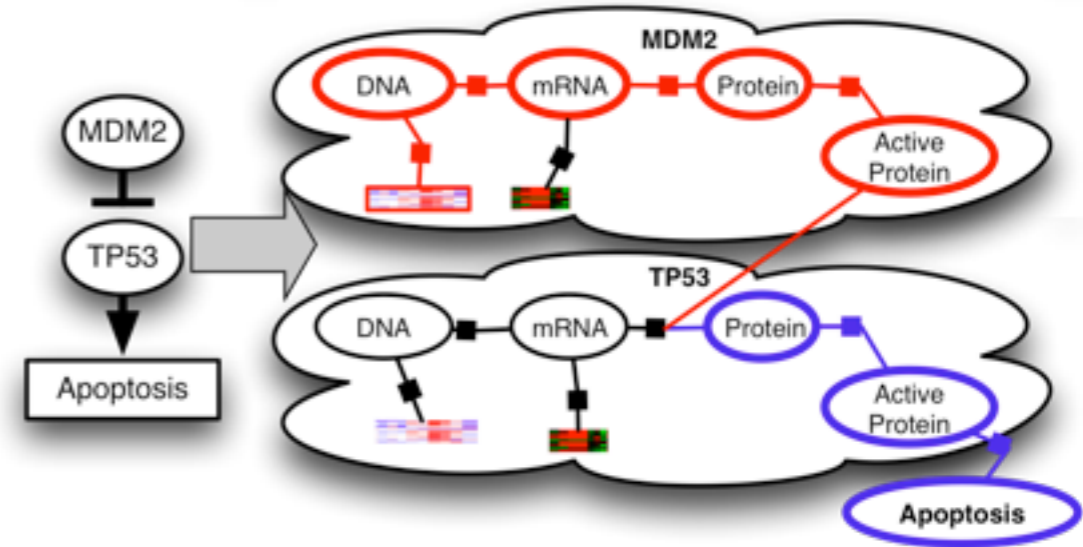
Integrated Pathway Activities (IPAs)

$$L(i, s) = \frac{log(\frac{P(D, x_i = s | \Phi)}{P(D, x_i \neq s | \Phi)})}{- \; log(\frac{P(x_i = s | \Phi)}{P(x_i \neq s | \Phi)})} =$$
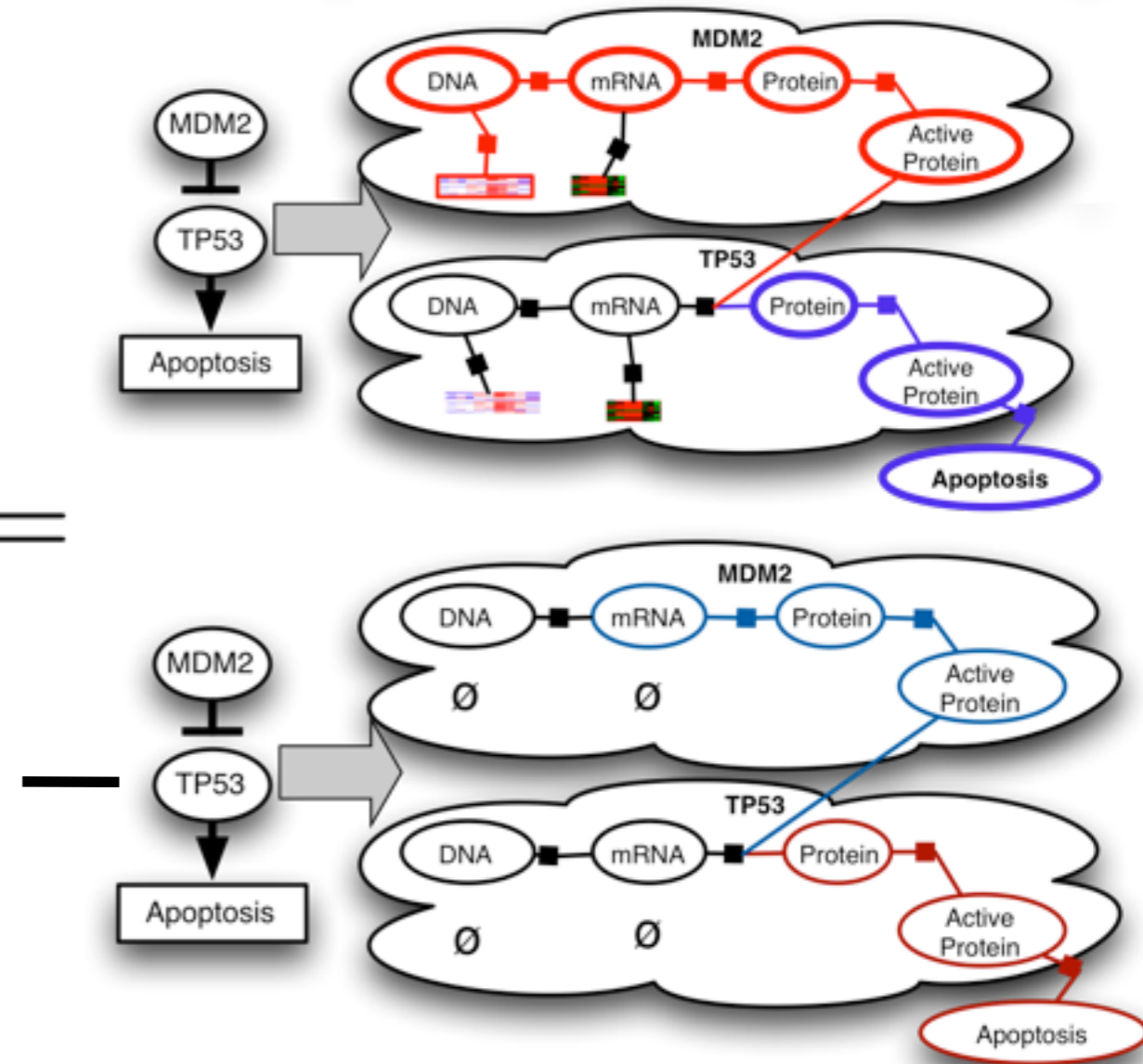
Integrated Pathway Activities (IPAs)

$$L(i, s) = \frac{log\left(\frac{P(D, x_i = s | \Phi)}{P(D, x_i \neq s | \Phi)}\right)}{- \; log\left(\frac{P(x_i = s | \Phi)}{P(x_i \neq s | \Phi)}\right)}$$
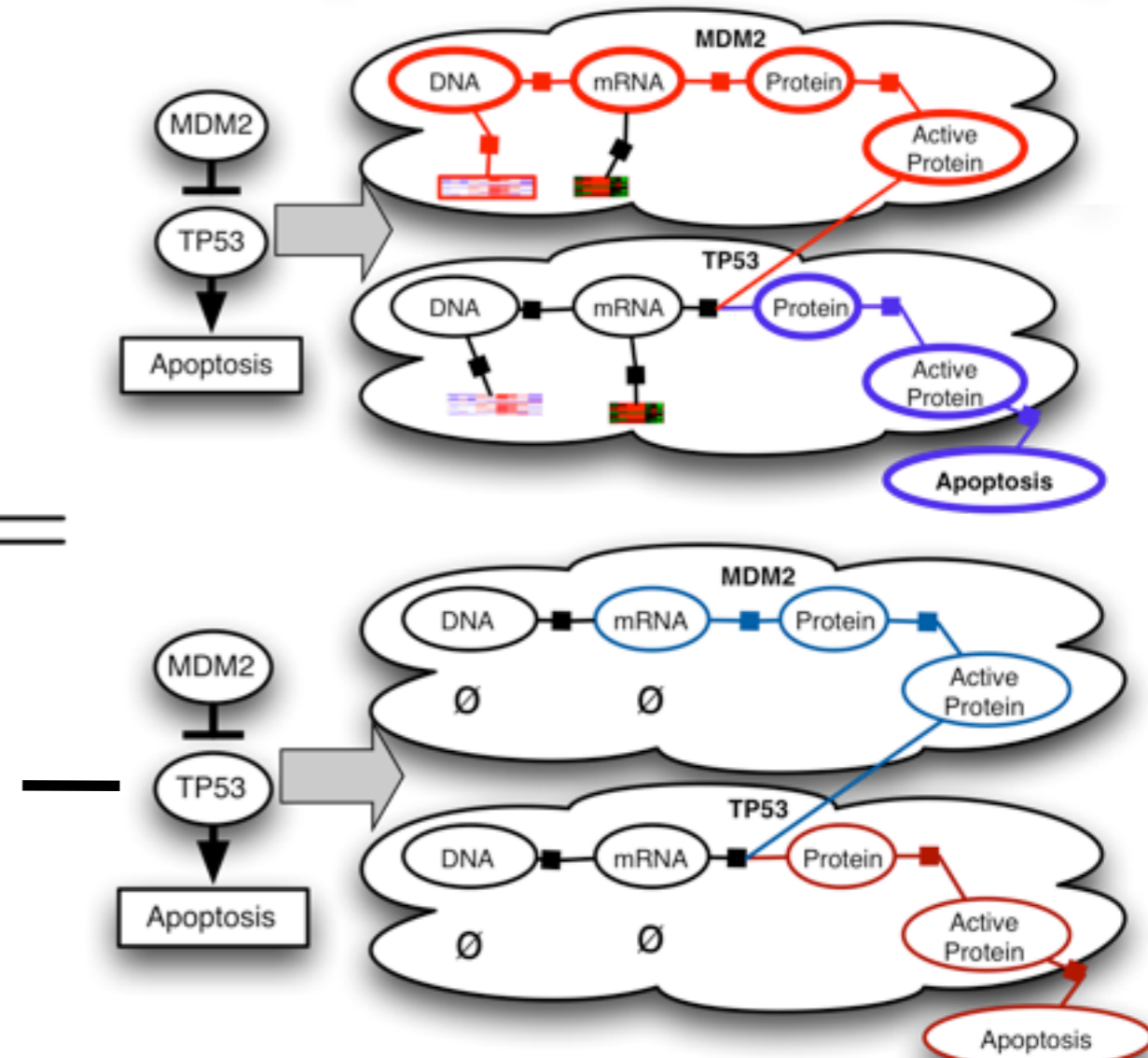
$$IPA(i) = \begin{cases} L(i, 1) & L(i,1) > L(i,-1) \text{ and } L(i,1) > L(i,0) \\ -L(i, -1) & L(i,-1) > L(i,1) \text{ and } L(i,-1) > L(i,0) \\ 0 & \text{otherwise.} \end{cases}$$

Detecting Top Pathways

Permute data to create "fake" samples across the dataset

Detecting Top Pathways

Permute data to create "fake" samples across the dataset

HIF-1-alpha transcription factor network -- 113 samples
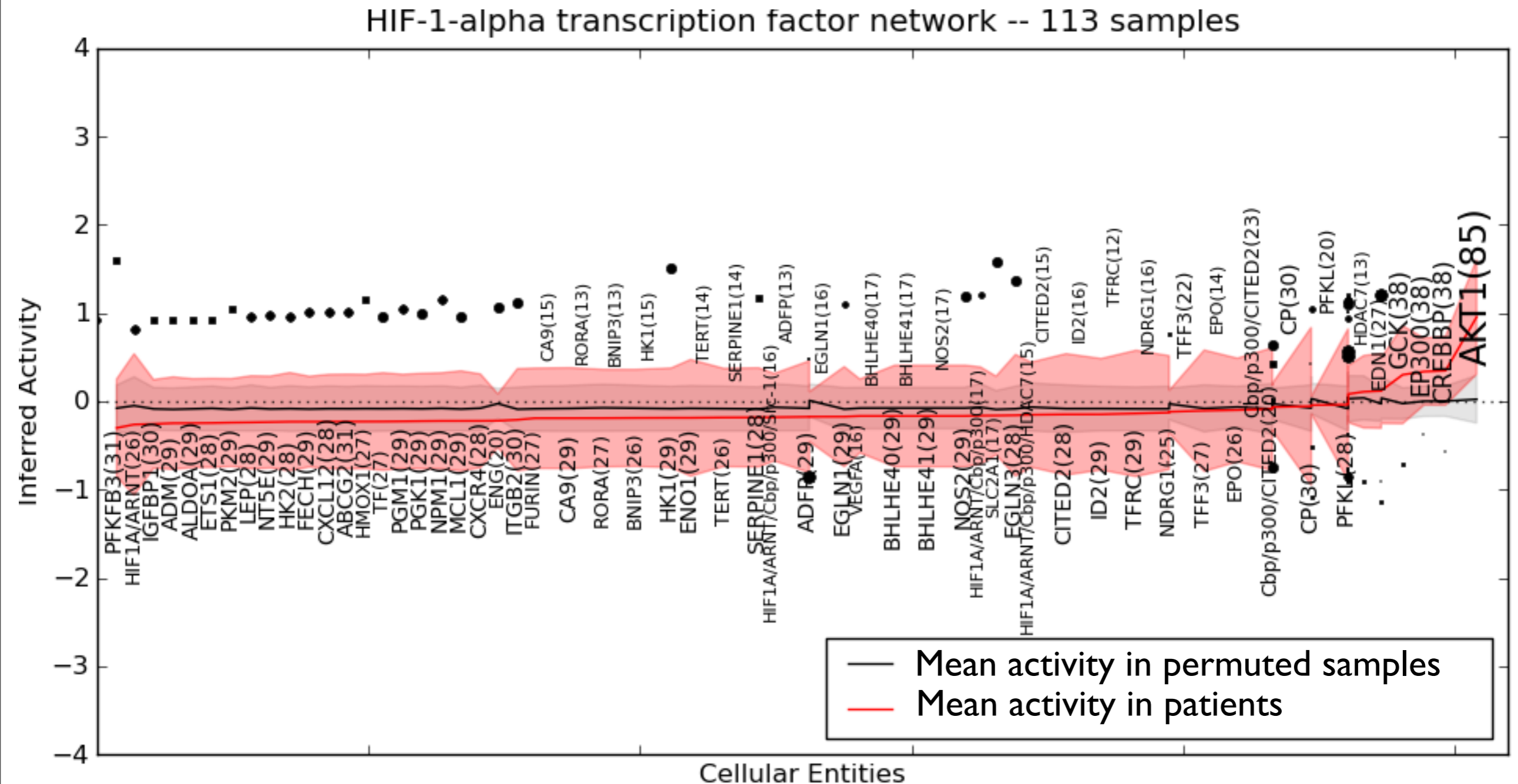
Breast Cancer Cohort: Chin et al. 2007 and Naderi et al. 2007

Signaling Pathway Impact Analysis (SPIA)

- Score pathways using expression data

- Weight genes by degree of connectivity

- Opposite of Google PageRank

- Compare by creating "decoy" pathways

Tarca et al. Bioinformatics 2008

PARADIGM Outperforms SPIA

PARADIGM Outperforms SPIA

AUCs	PARADIGM	SPIA
Breast	0.669	0.602
GBM	0.642	0.604

SPIA Implementation by Dent Earl

Clustering IPAs Stratifies GBM Patients
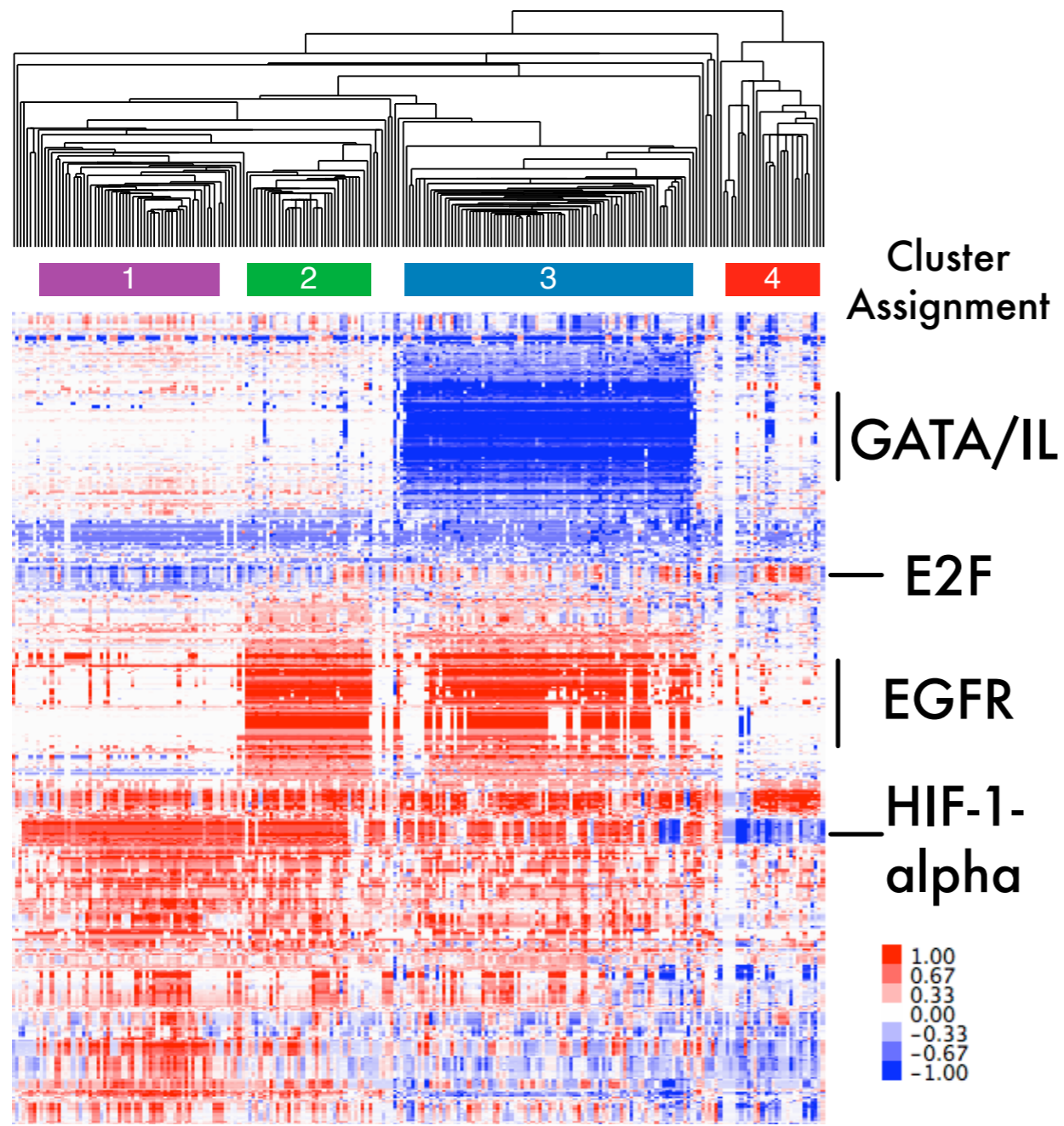
Clustering IPAs Stratifies GBM Patients
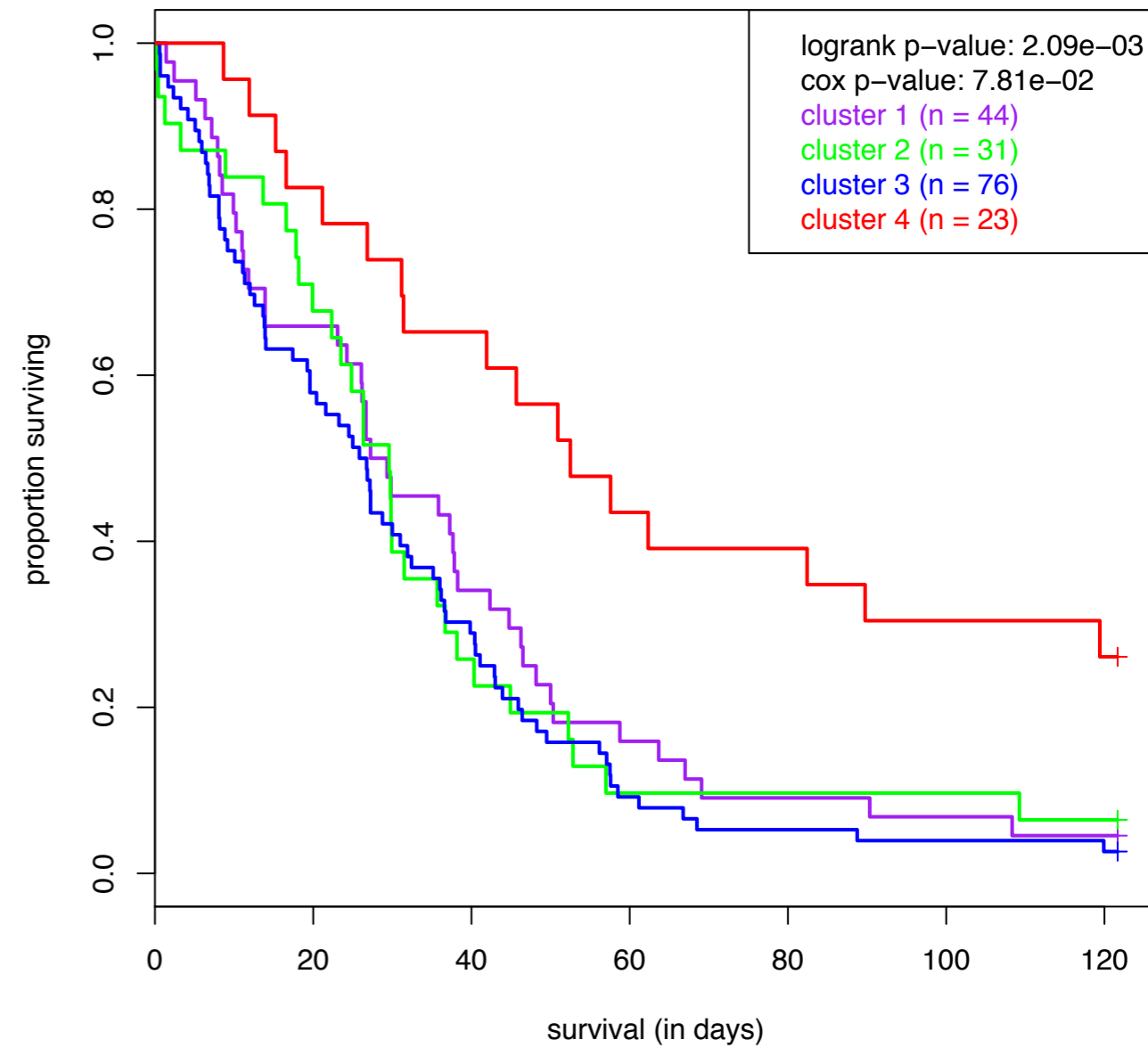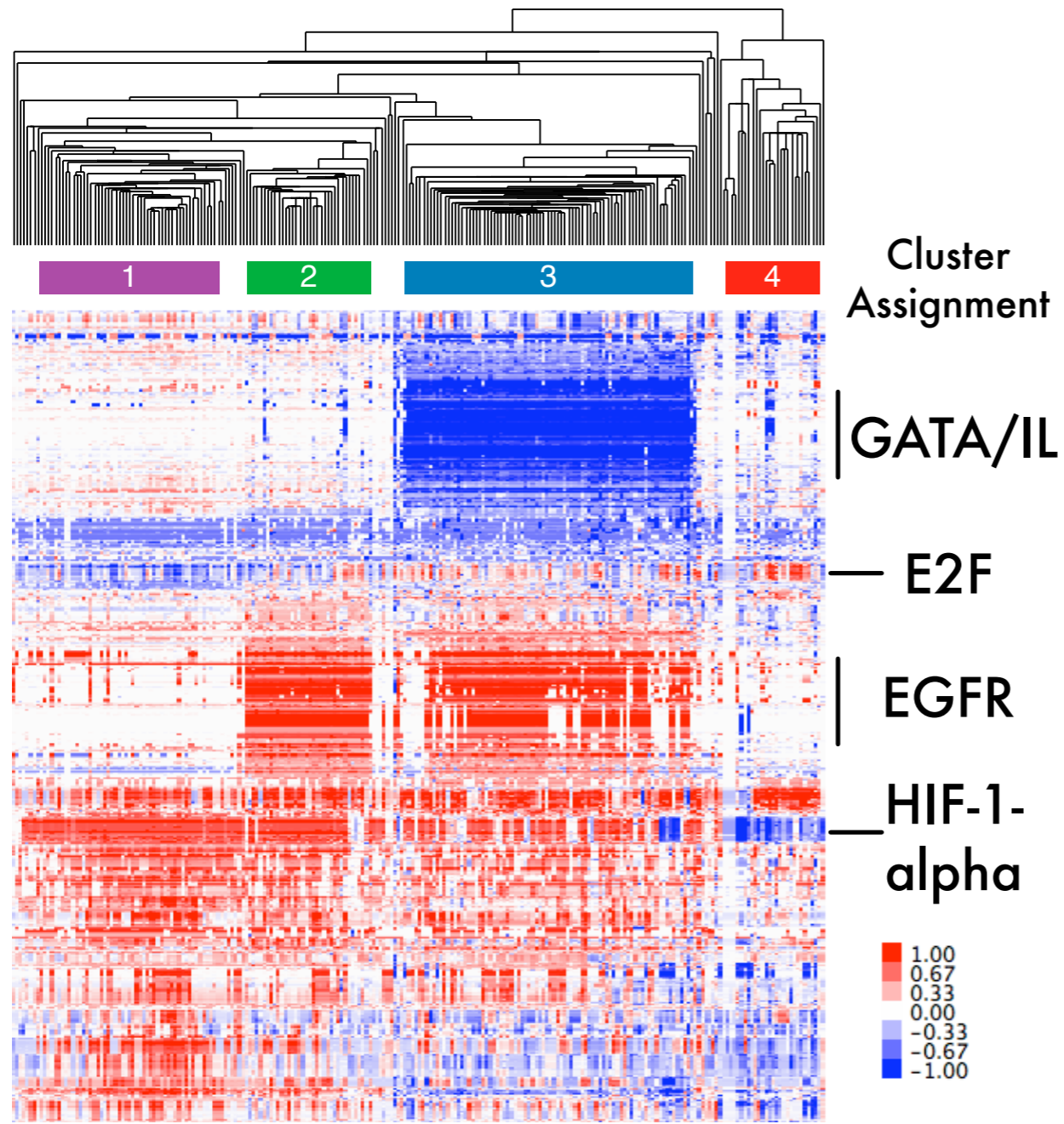
Clustering IPAs Stratifies OV Patients
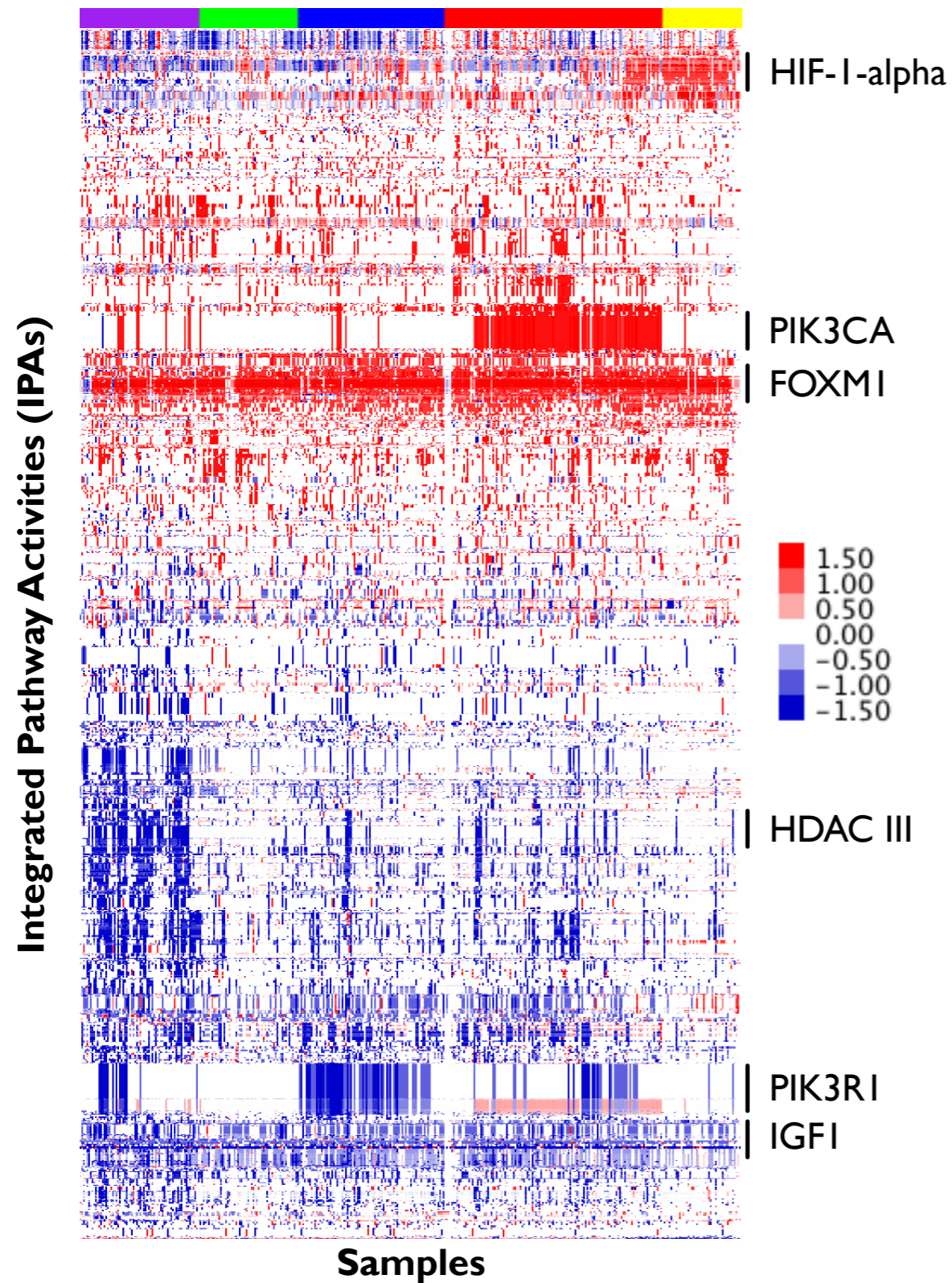
Clustering IPAs Stratifies OV Patients
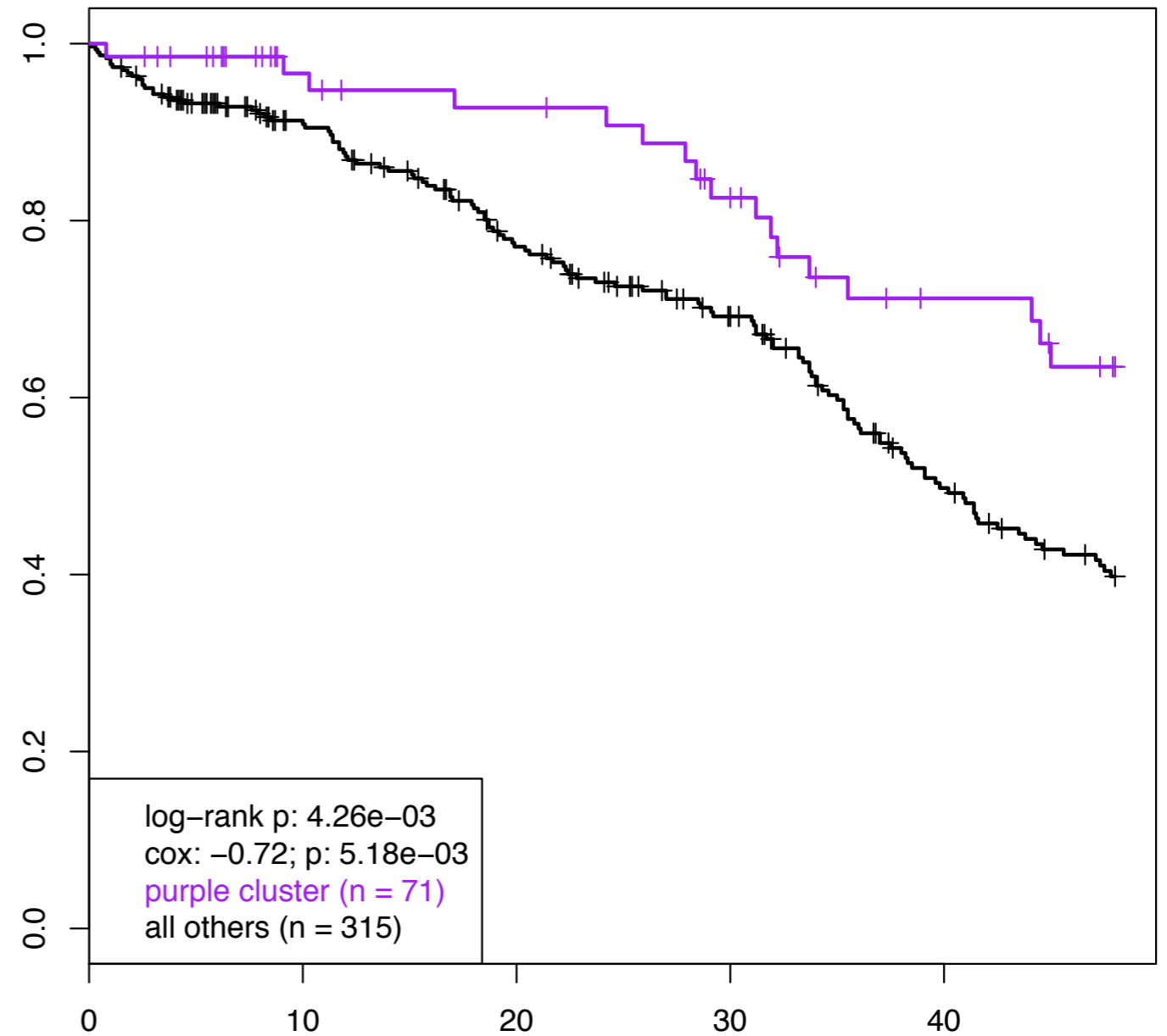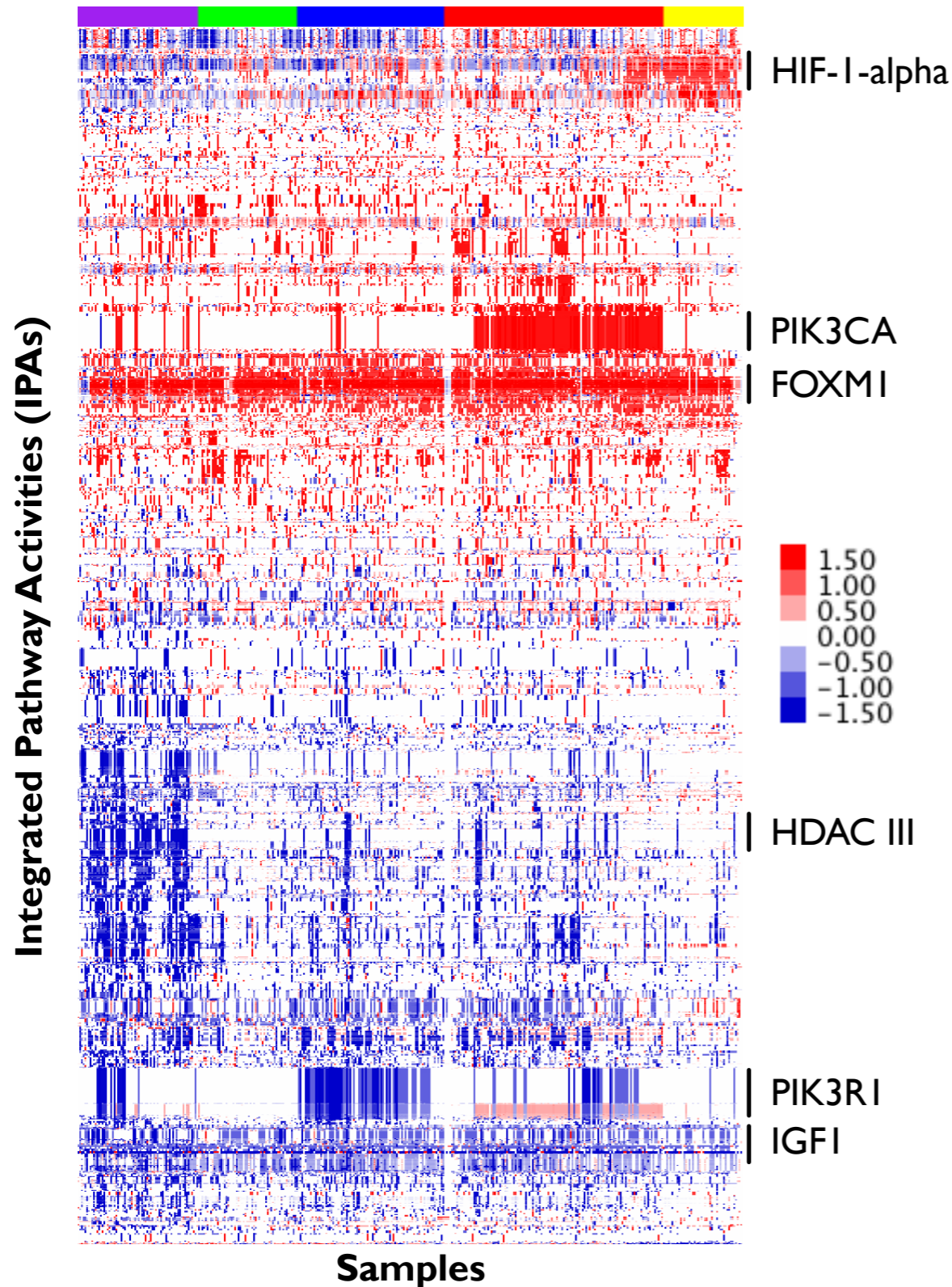
Future Work

- Incorporate Mutations, Methylation, miRNA

- More pathway sources - especially metabolic

- Try a "unified" pathway of cell

Outline

I. Introduction

II. Previous Work

III. Aim 1: PARADIGM - Integrated Pathway Analysis

IV. Aim 2: Modeling pathways using sequence data

V. Aim 3: Directing combination therapies

Mainstream Sequencing Has Arrived

- 100 full genome sequences available

- TCGA: 50 per cancer type (20+ types)

- PARADIGM - underlying model still works!

Ovarian IPA Heatmap
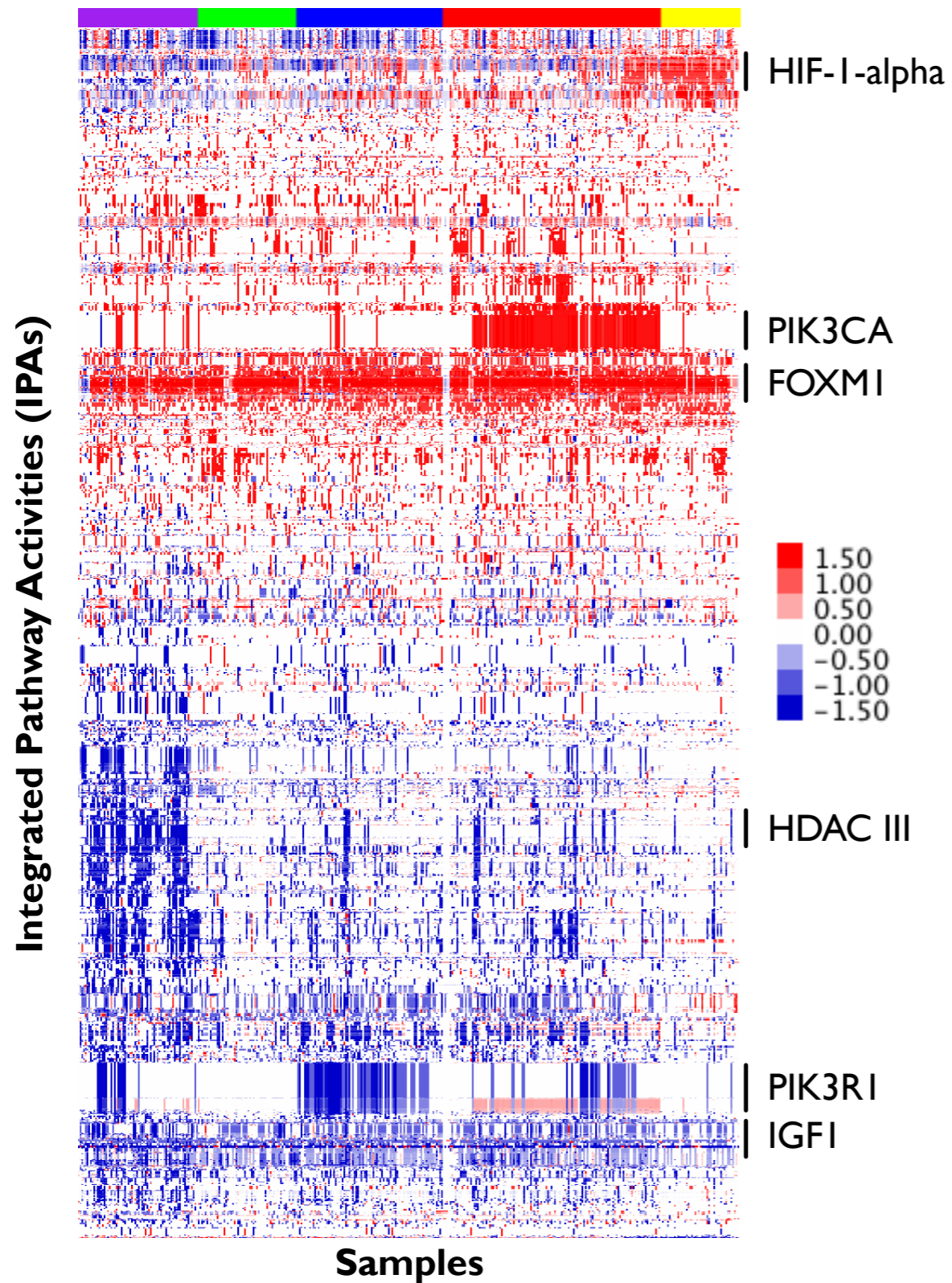
Ovarian IPA Heatmap
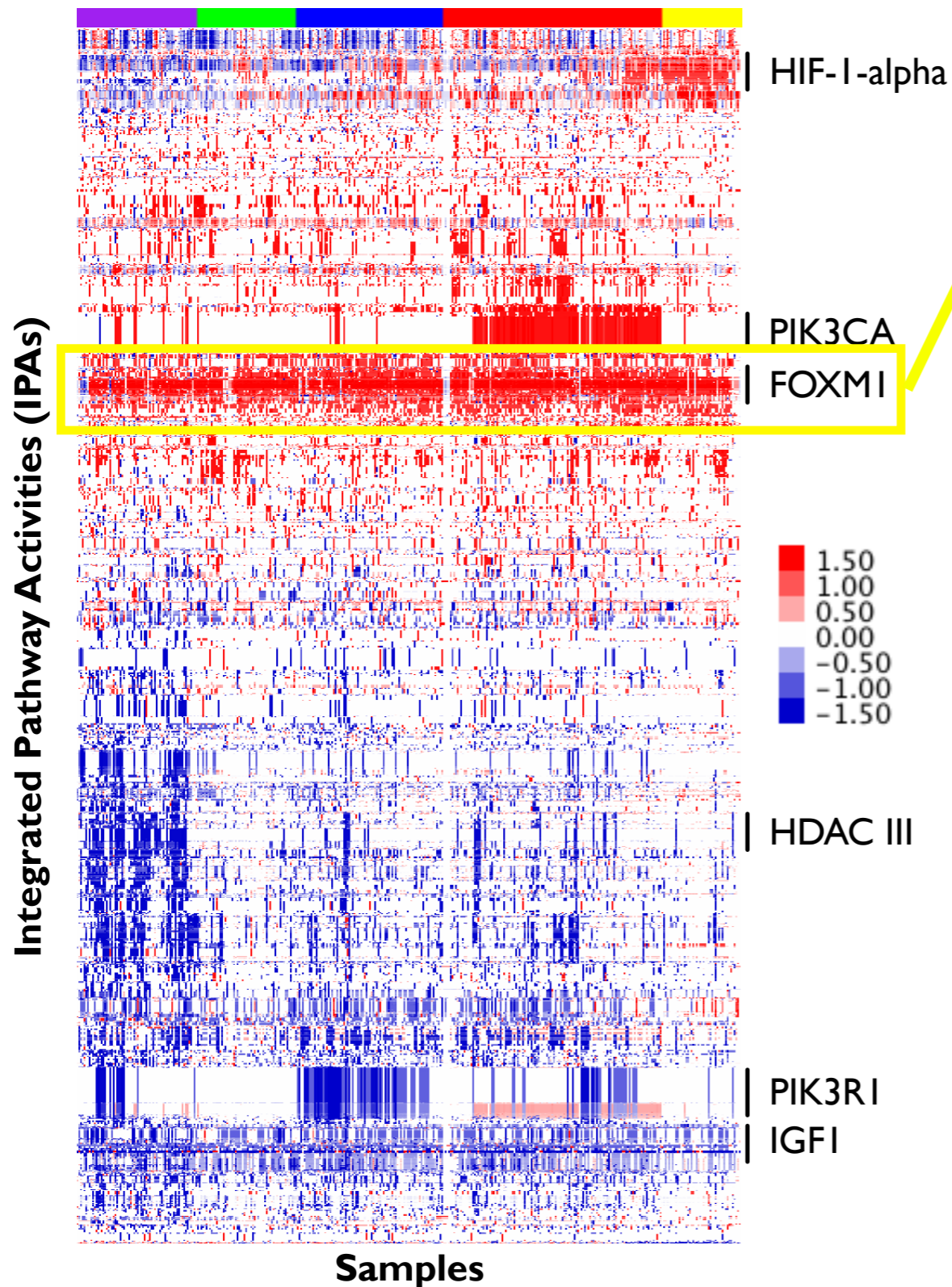
Pathway Name	Percentage
FOXM1 transcription factor network	49.79%
FOXA2 and FOXA3 transcription factor	14.06%
TCGA08 Retinoblastoma	11.78%
Endothelins	11.65%
Calcium signaling in the CD4+ TCR pathway	11.43%
Glucocorticoid receptor regulatory network	11.29%
TCR signaling in naïve CD4+ T cells	11.24%
TRAIL signaling pathway	10.95%
IL23-mediated signaling events	10.89%
ErbB4 signaling events	10.71%
IL12-mediated signaling events	10.53%
Canonical NF-kappaB pathway	10.49%
Regulation of cytoplasmic and nuclear	10.21%
Wnt signaling	10.17%
HIF-1-alpha transcription factor network	10.13%
PLK1 signaling events	10.04%

Ovarian Samples Have Significantly Higher Activity

FOXM1 transcription factor network -- 330 samples
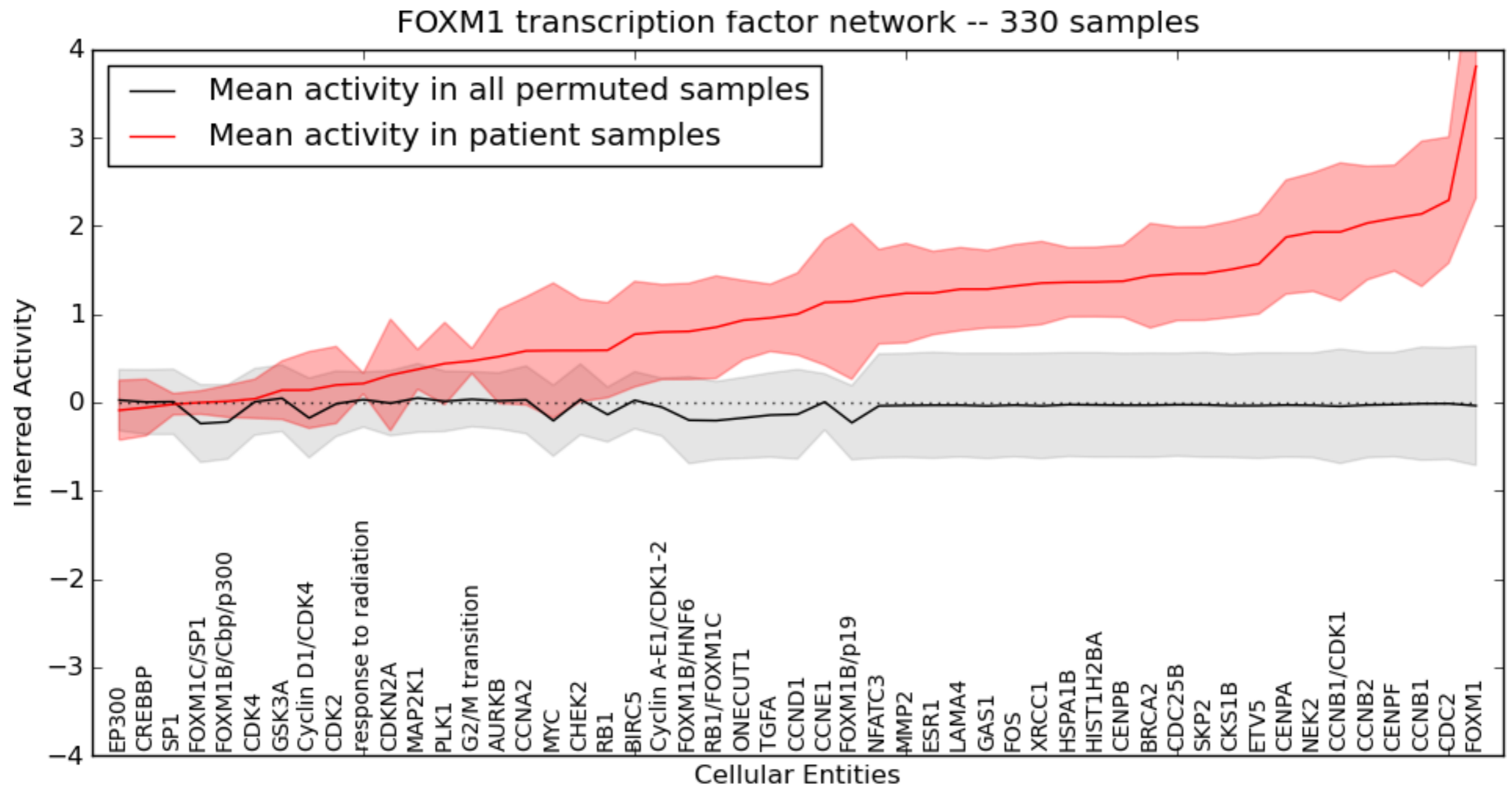
Forkhead Box M1 (FOXM1)
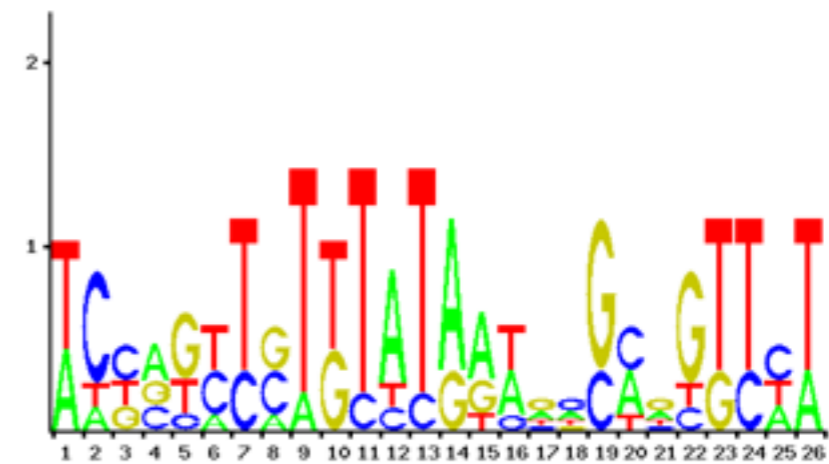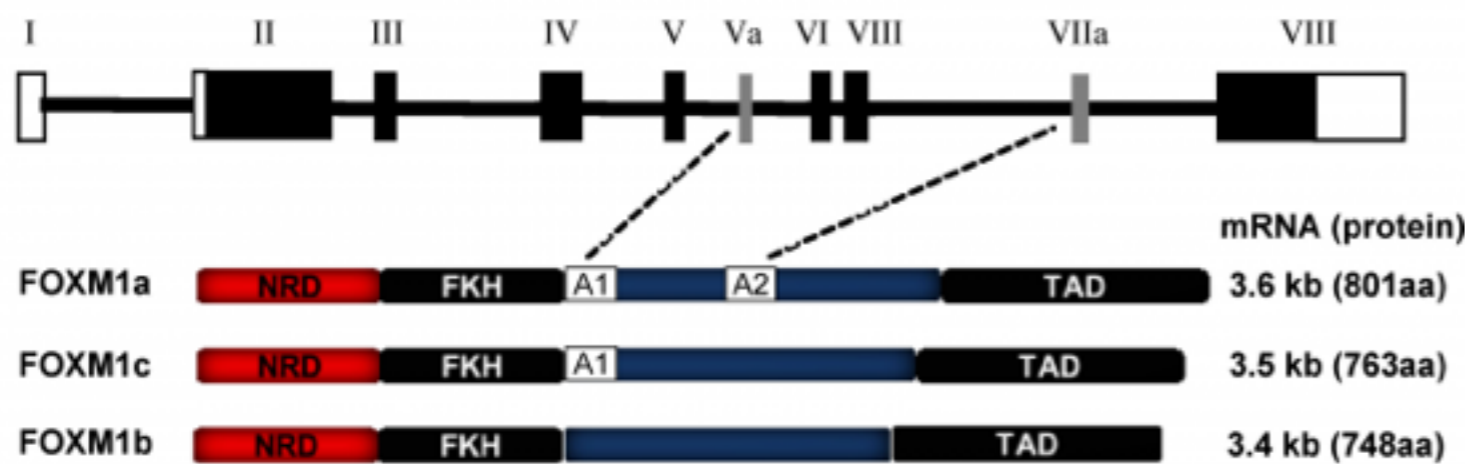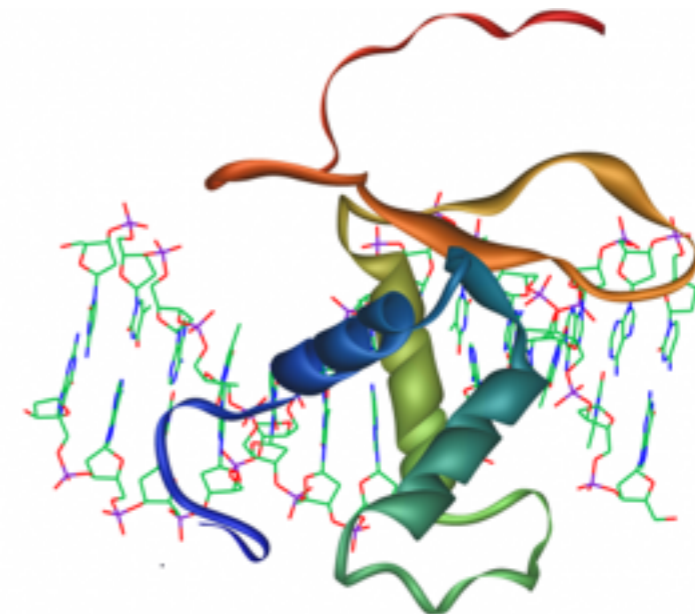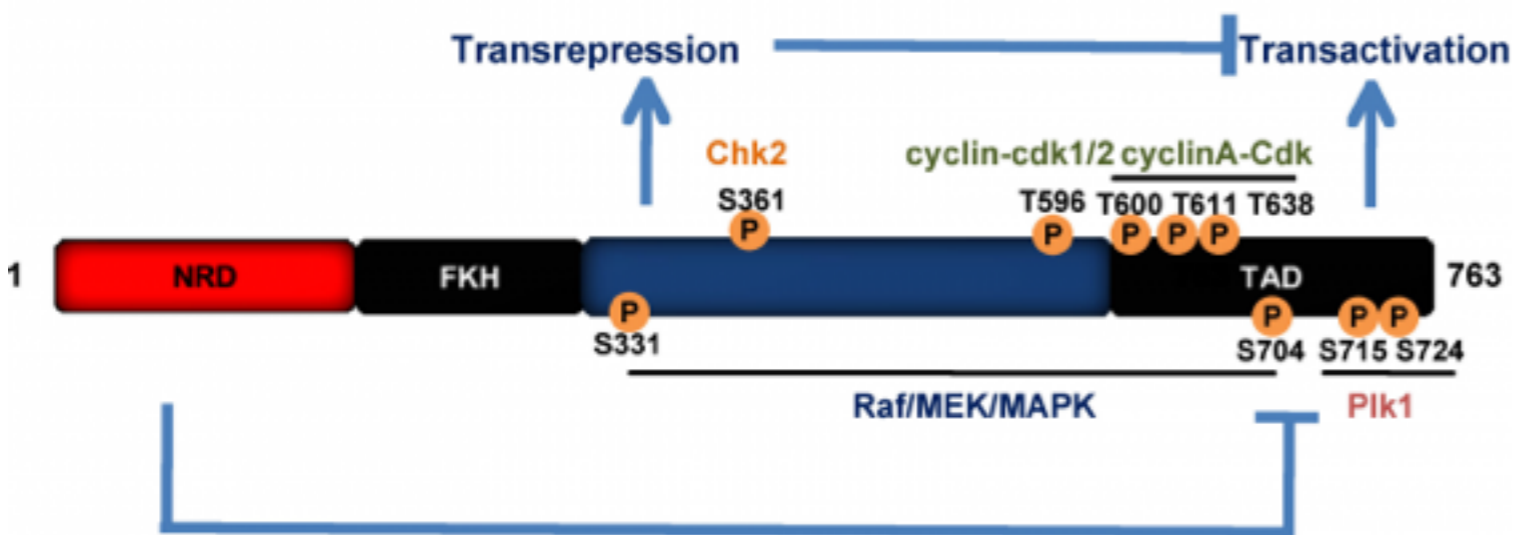
Transcription Factor Encyclopedia (http://www.cisreg.ca/cgi-bin/tfe/articles.pl?tfid=437)

Transcription Factor at the Crossroads of Repair and Cell Cycle
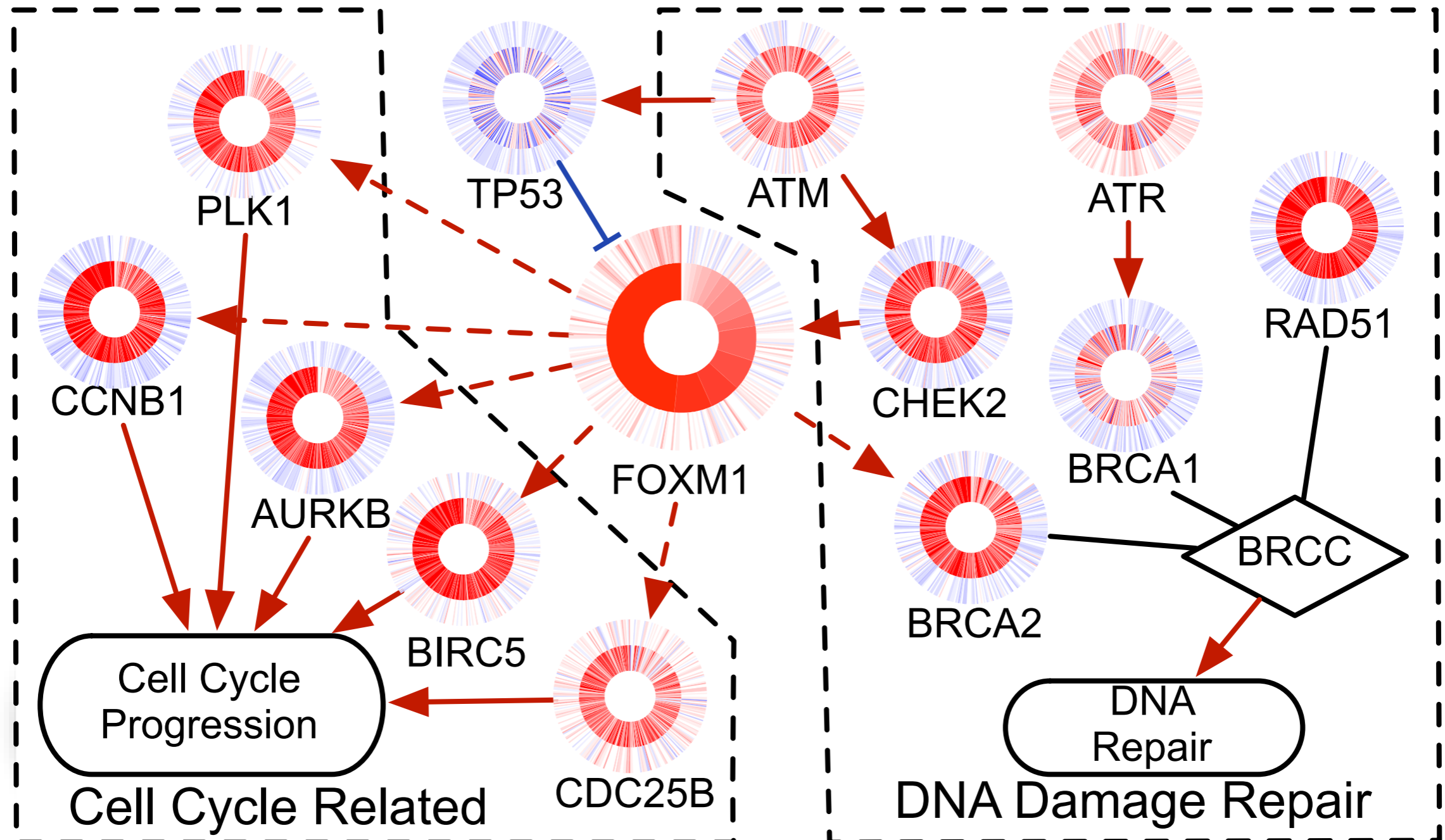
Sequencing can help us!

- Need to be able to detect isoform-specific transcripts

- RNA-Seq using Velvet/Columbus/Oasis pipeline and SOLAS*

- Eventually move to allele-specific signals

Velvet/Columbus/Oasis by D. Zerbino   *or other isoform-detecting algorithms

Isoform- and Allele-specific gene model
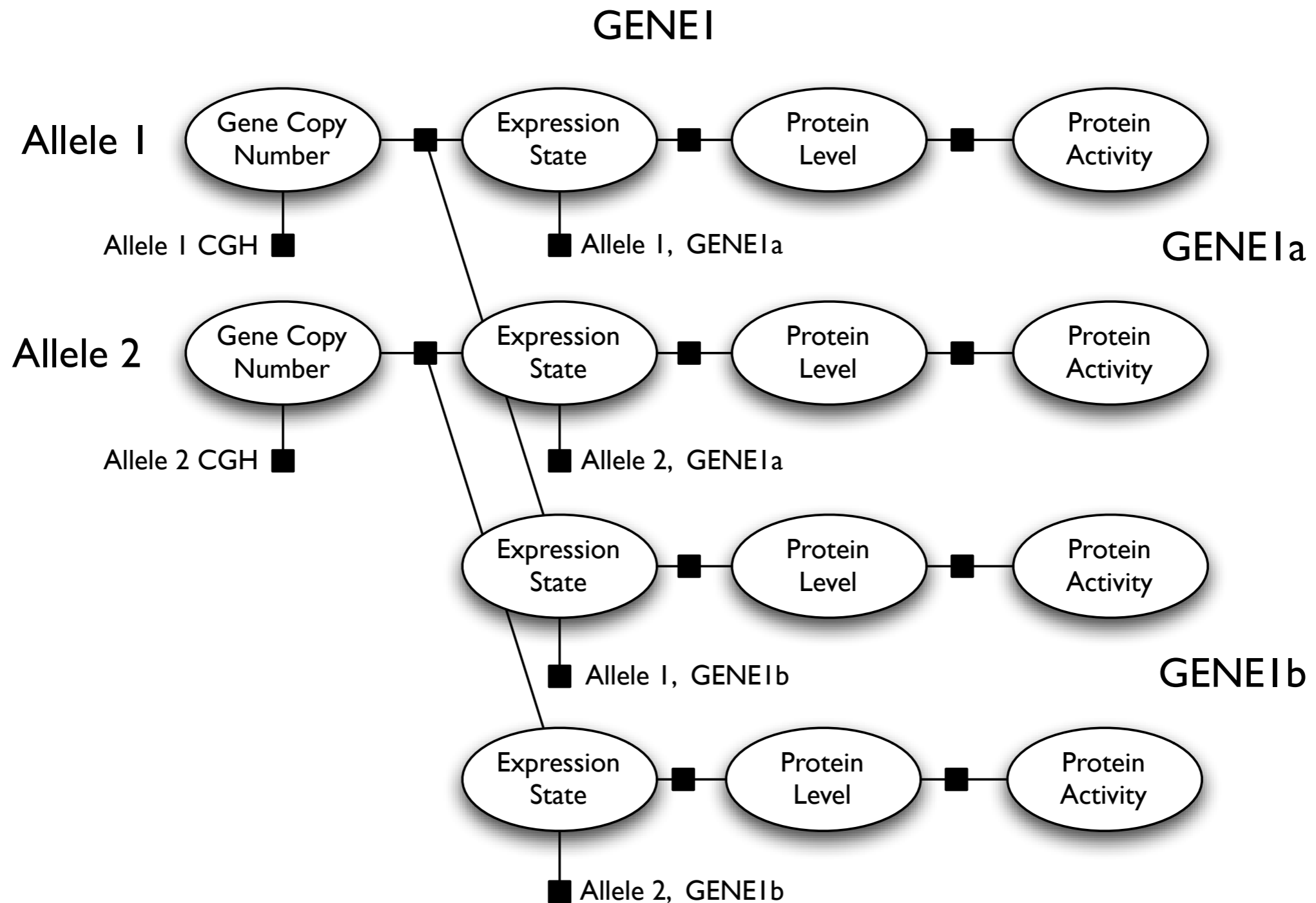
Gene Fusions

- Detectible by RNA-Seq

- Uncommon, but critical to model

- Dynamically add new pathway representing fusion
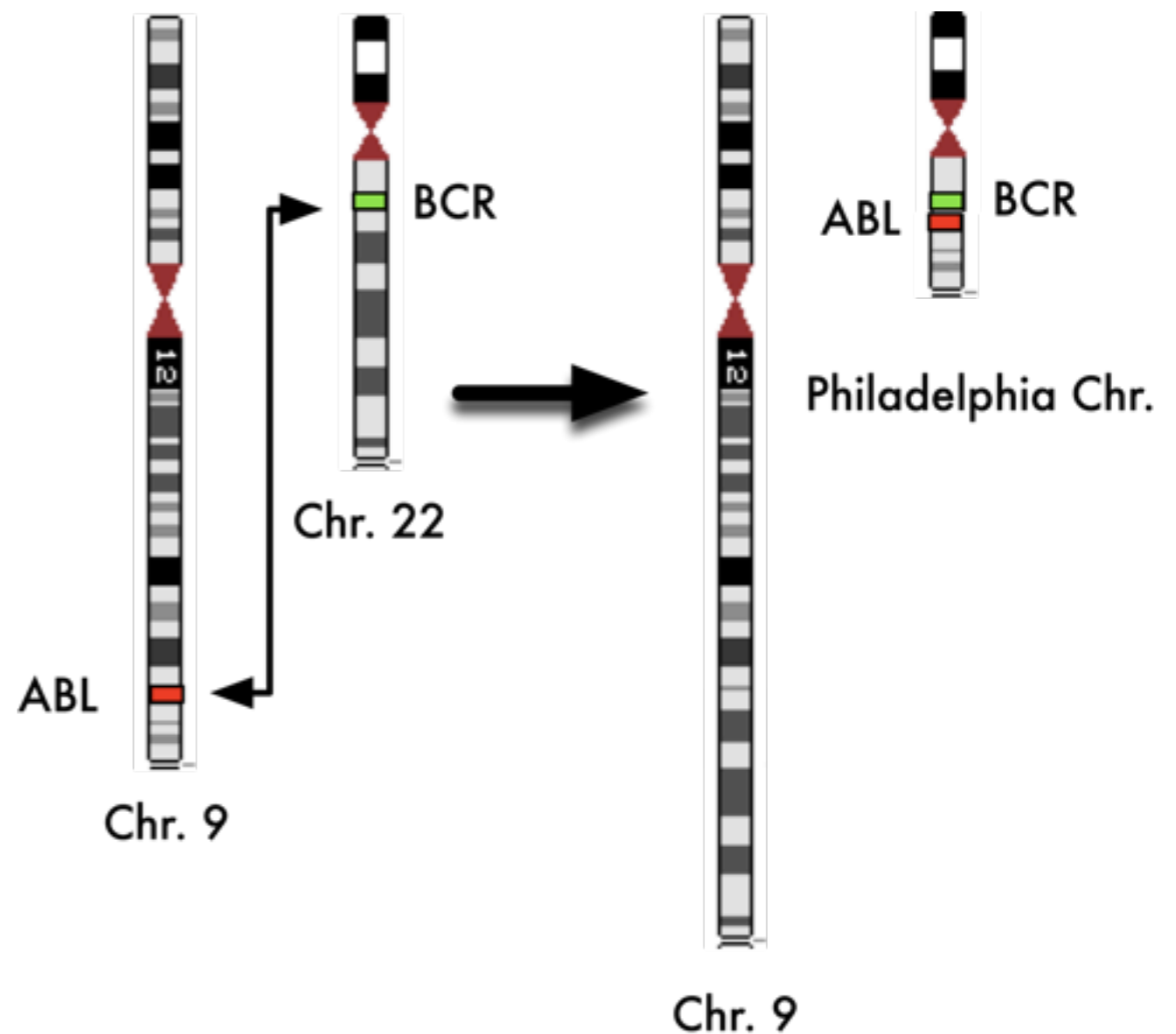
Future Work

- Work on regularizing sequencing data for attaching to the model

- Add support for allele- and isoform-specific central dogma

- Model Gene Fusions

Outline

I. Introduction

II. Previous Work

III. Aim 1: PARADIGM - Integrated Pathway Analysis

IV. Aim 2: Modeling pathways using sequence data

V. Aim 3: Directing combination therapies

Overview

Goal: Identify shRNAs that can be used to convert drug resistant samples into sensitive

- Identify critical points in pathway

- Simulate shRNA combinations *in silico*

- Use cell lines to test predictions

IPAs of Breast Cancer Cell Lines
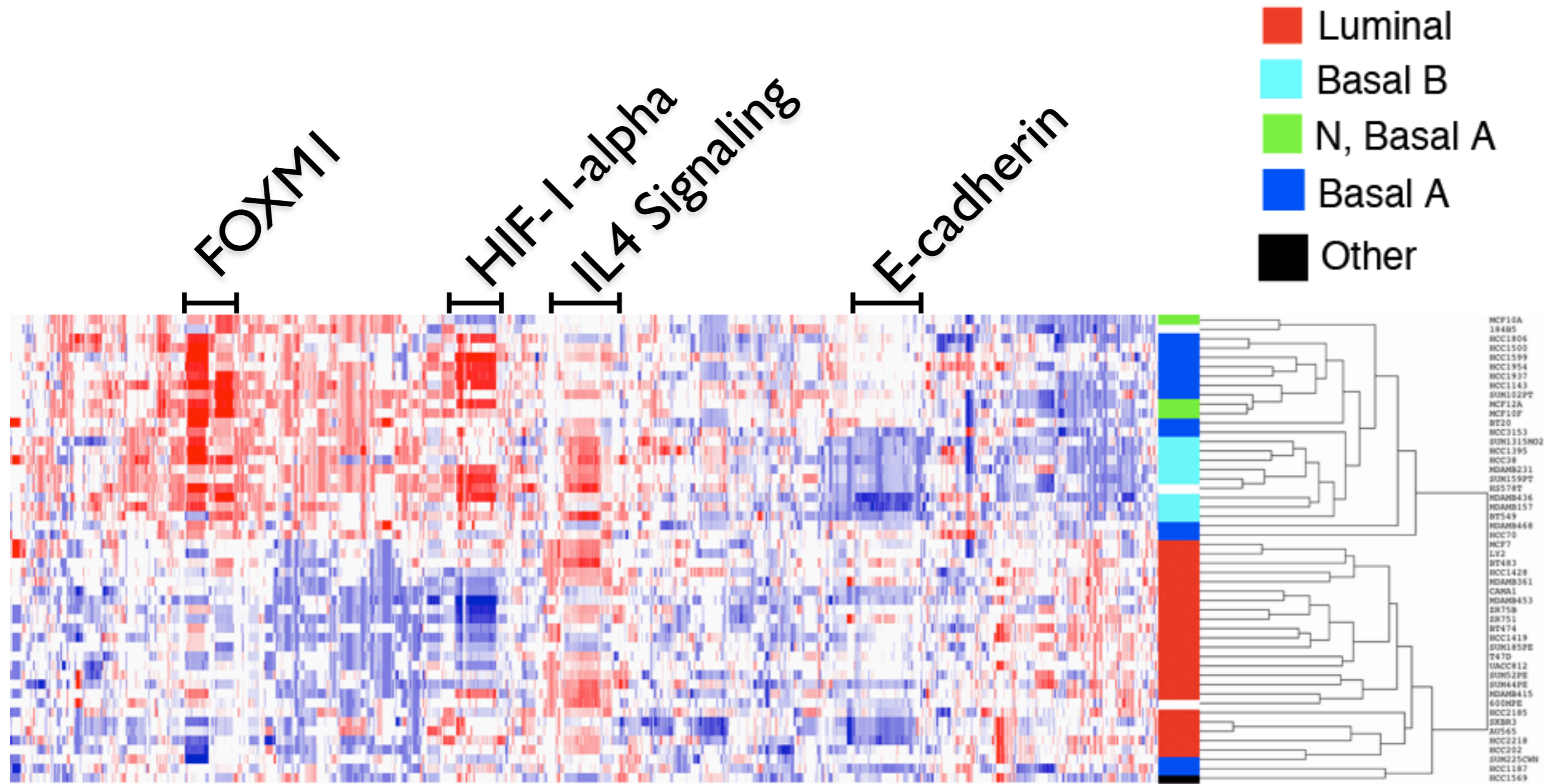
Cell Line Drug Responses & Known Targets

50 breast cancer cell lines, 70 drugs

Erlotinib
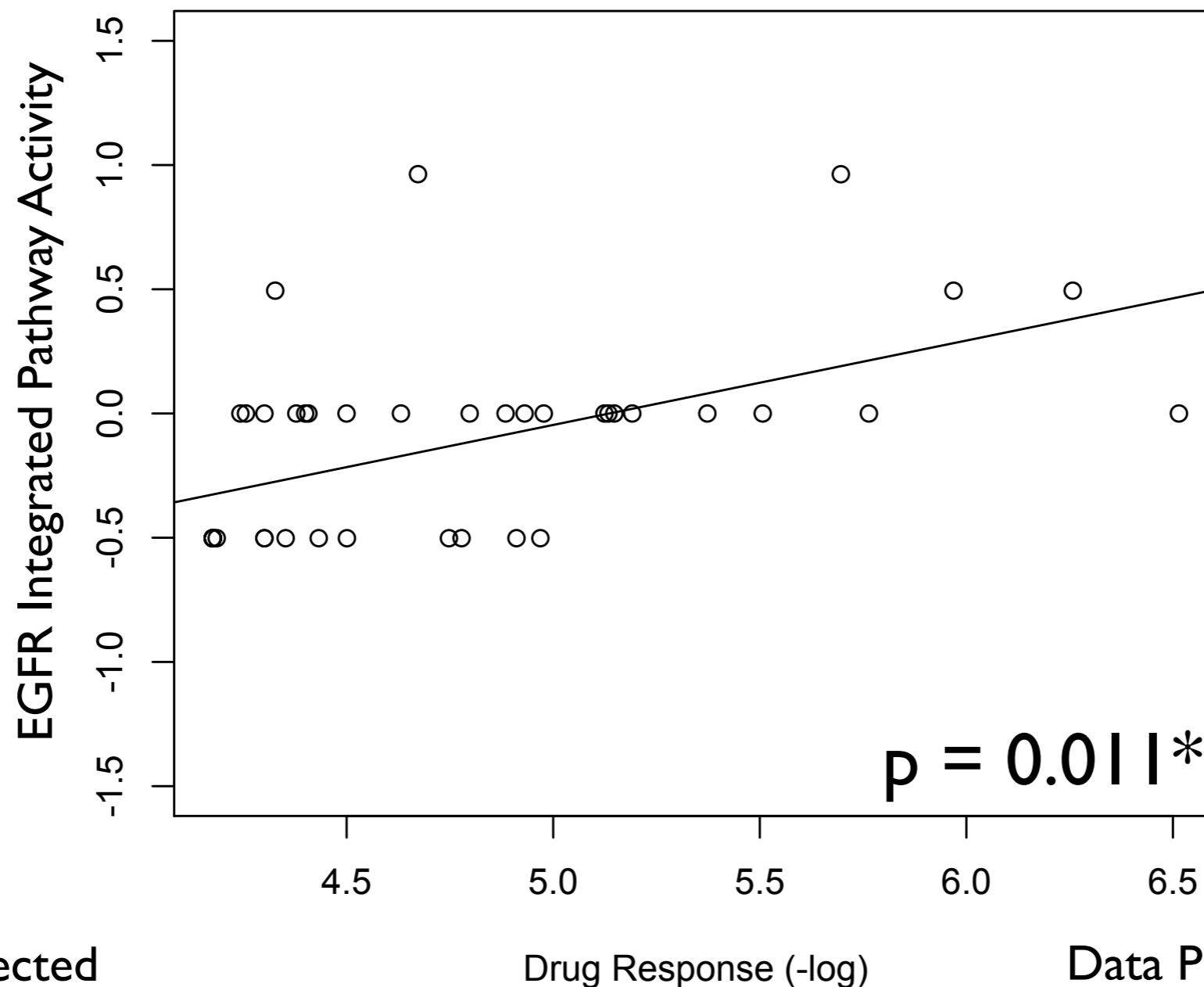
*Bonferroni Corrected

Data Provided by Gray Lab, LBNL

Identify Key Pathway Entities

Feature Selection & Support Vector Machines to find all possible contributing entities

Identify Key Pathway Entities

Feature Selection & Support Vector Machines to find all possible contributing entities

	E1	E2	E3	E4	E5	E6	...
S	5	4	5	-3	-2	-3	...
S	6	5	5	-4	-5	-4	...

Identify Key Pathway Entities

Feature Selection & Support Vector Machines to find all possible contributing entities

	E1	E2	E3	E4	E5	E6	...
S	5	4	5	-3	-2	-3	...
S	6	5	5	-4	-5	-4	...
R	-2	-3	-3	4	5	5	...
R	-3	-4	-3	3	4	4	...
R	-2	-4	-6	5	5	3	...

Find Targets to Test with shRNA

Use medioid of sensitive vectors as target

	E1	E2	E3	E4	E5	E6	...
M-S	5.5	4.5	5	-3.5	-3.5	-3.5	...

Find Targets to Test with shRNA

Use medioid of sensitive vectors as target

	E1	E2	E3	E4	E5	E6	...
M-S	5.5	4.5	5	-3.5	-3.5	-3.5	...

Disconnect parents of each entity to simulate shRNA
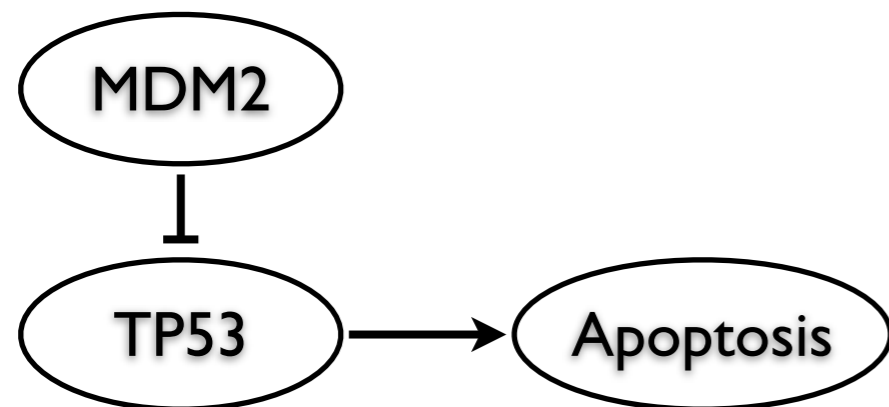
Find Targets to Test with shRNA

Use medioid of sensitive vectors as target

	E1	E2	E3	E4	E5	E6	...
M-S	5.5	4.5	5	-3.5	-3.5	-3.5	...

Disconnect parents of each entity to simulate shRNA
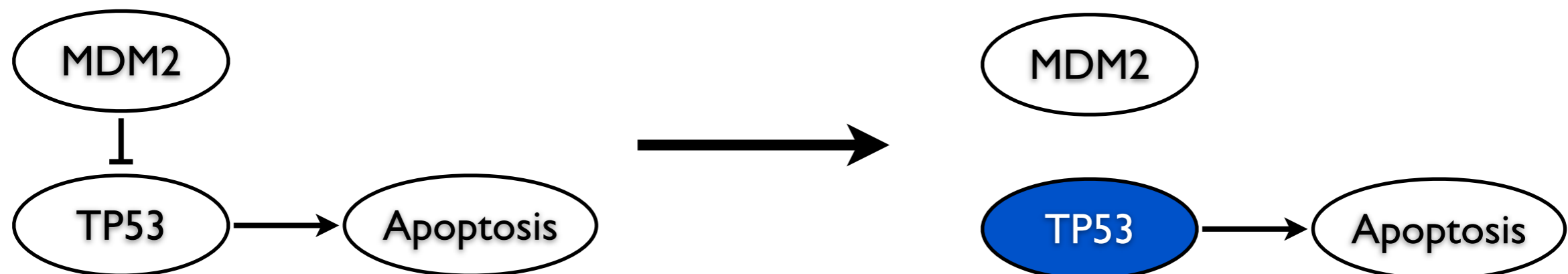
Find Targets to Test with shRNA

Use medioid of sensitive vectors as target

	E1	E2	E3	E4	E5	E6	...
M-S	5.5	4.5	5	-3.5	-3.5	-3.5	...

Disconnect parents of each entity to simulate shRNA

Rank all entities and take the top targets to the wet lab

Future Work

- Identify key pathways associated with treatment resistance

- Rank genes by disconnecting links

- Test targets using shRNA screens

Bringing it all together...

Goal: Provide clinically relevant information

- Assist treatment decisions

- Find new targets

- Focus on the individual tumors

Acknowledgments

- **Advancement Committee**
  - **David Haussler**
  - **Josh Stuart**
  - **Nader Pourmand**
  - **Scott Lokey**

- **UCSC Cancer Genomics**
  - Zack Sanborn
  - Jingchun Zhu
  - Christopher Szeto
  - Larry Meyer
  - Tracy Ballinger
  - Daniel Zerbino
  - Sam Ng
  - Mia Grifford
  - James Durbin
  - Fan Hsu
  - Sofie Salama

  - UCSC Genome Browser Staff
  - Cluster Admins

- **Collaborators**
  - The Cancer Genome Atlas
  - Stand Up To Cancer
  - Christopher Benz, Buck Institute
  - Laura Esserman, UCSF
  - Joe Gray, LBL
  - Eric Collisson, UCSF
  - Ting Wang, Washington University
  - Charles Vaske, Princeton

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