Integrated Analysis of Cancer Pathways with the UCSC Cancer Genomics Browser

Josh Stuart, UCSC

Advances in Bioinformatics and Genomics Talk, Feb 19, 2010

UCSC Cancer Pathways Group

Collaborating Labs

Haussler Lab, UCSC

Gray Lab, LBL

Benz Lab, Buck Inst

Funding

NIH TCGA

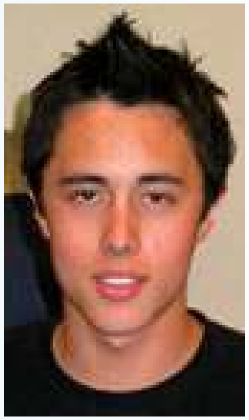
Stand Up To Cancer


James Durbin


Steve Benz


Charlie Vaske


Zack Sanborn


Chris Szeto


Jing Zhu

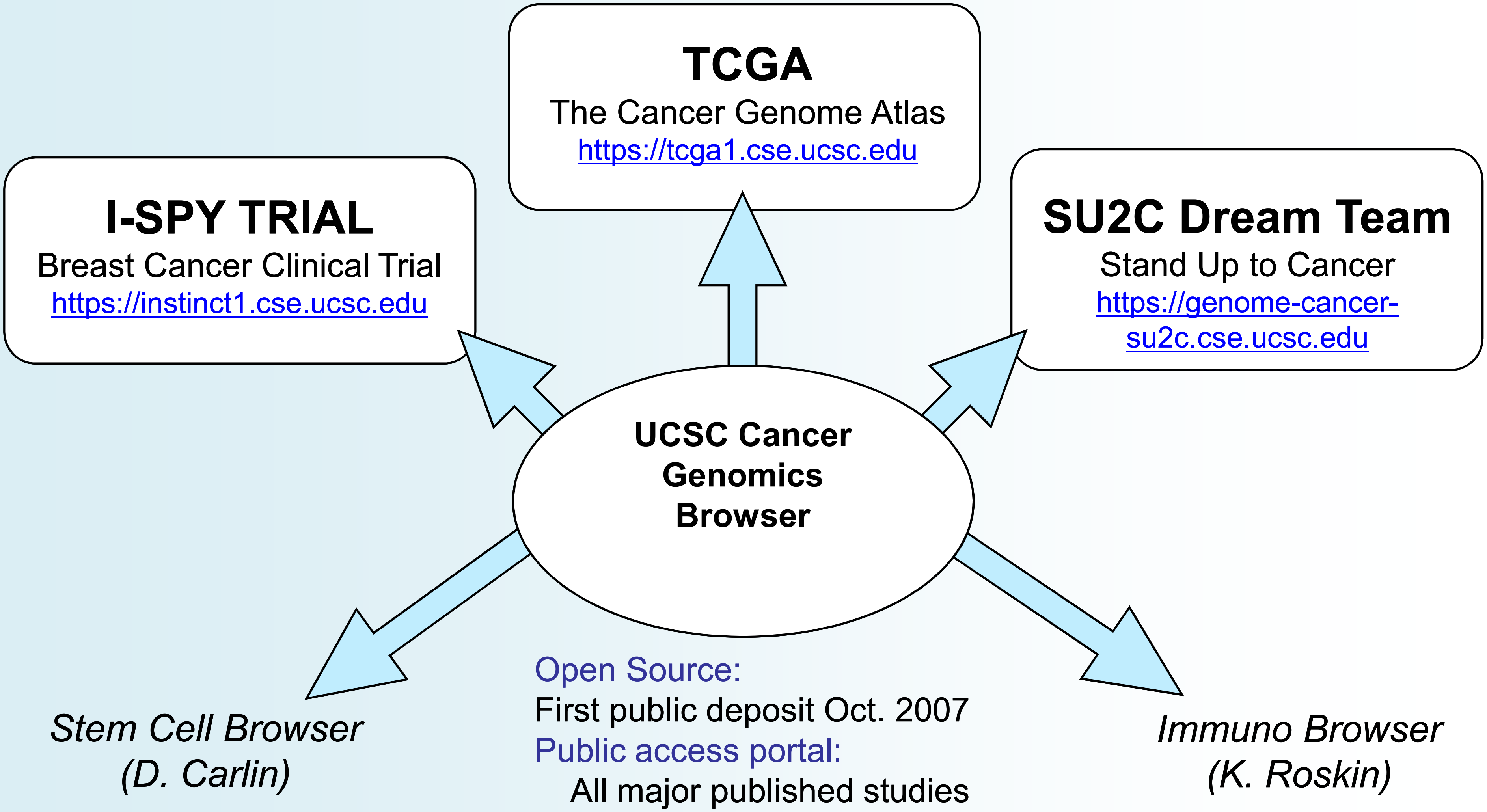
Outline

- High-throughput visualization w/ the UCSC Cancer Genomics Browser
- Identifying patient-specific pathway alterations
- Identifying novel pathway interactions w/ genetic knock-outs

Outline

- High-throughput visualization w/ the UCSC Cancer Genomics Browser
- Identifying patient-specific pathway alterations
- Identifying novel pathway interactions w/ genetic knock-outs

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

Stem Cell Browser
(D. Carlin)

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

Immuno Browser
(K. Roskin)

UCSC Cancer Informatics Group Aims

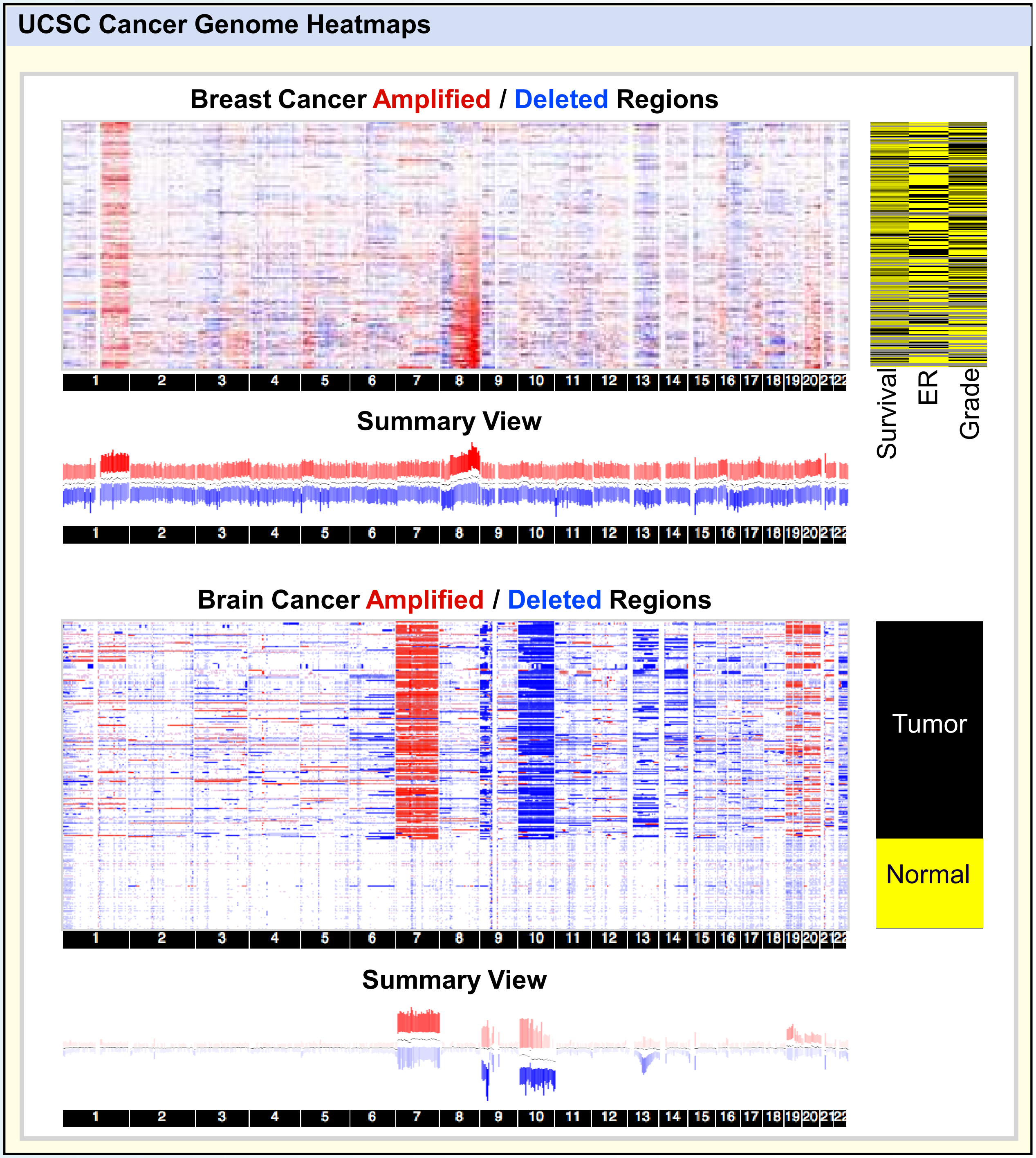
- Secure database to manage and communicate data for national cancer projects (TCGA, SU2C, I-SPY)
- Web interface to help browse and download the data (and results and models…).
- Integrated tools to detect and display altered pathways.
- Predictors of specific patient subsets based on omics and pathway-level data.
- Recommend drugs based on patient-specific pathway alterations

UCSC Cancer Genomics Browser

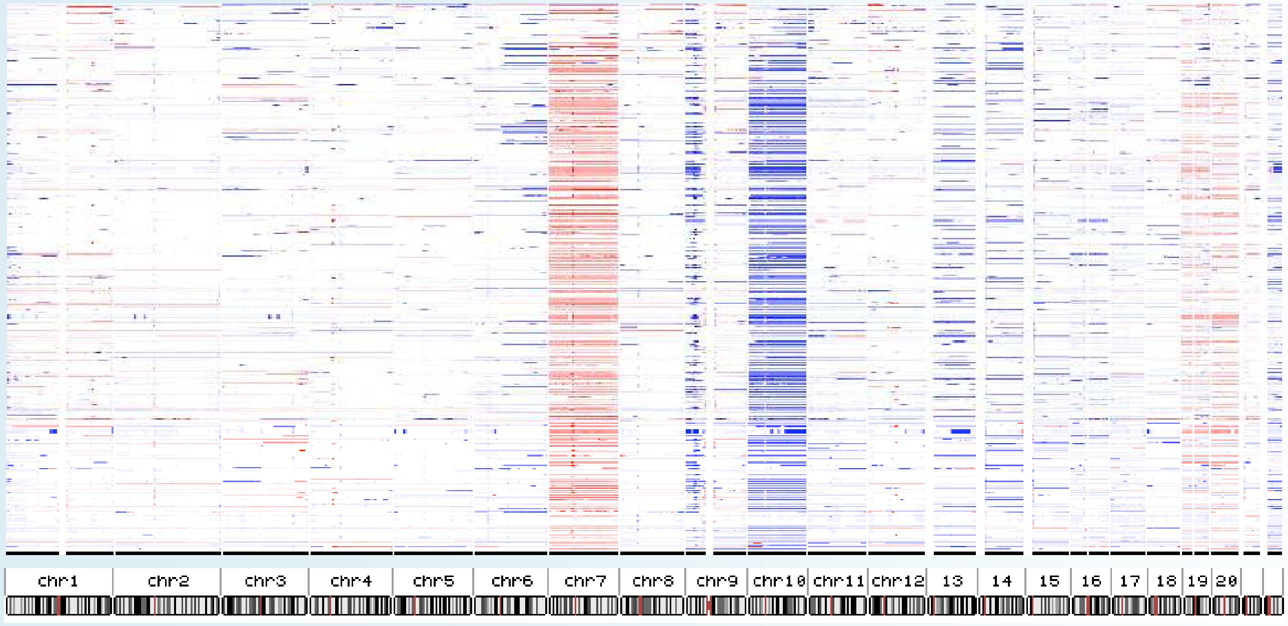
- Simultaneously display patient genomic *and* clinical data from multiple studies
- Highly configurable display
- Growing list of available genome-wide statistical tests
- Integrated with popular UCSC Genome Browser and its vast store of genomic information

---

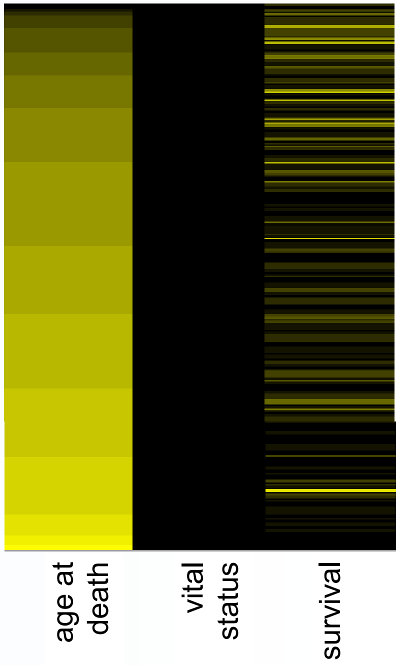
- Compatible with all modern web browsers
- jQuery Javascript Frontend
- Robust C / MySQL Server Backend
- Distributed Databases
- Multi-level Data Security Model



Genome/clinical heatmap view



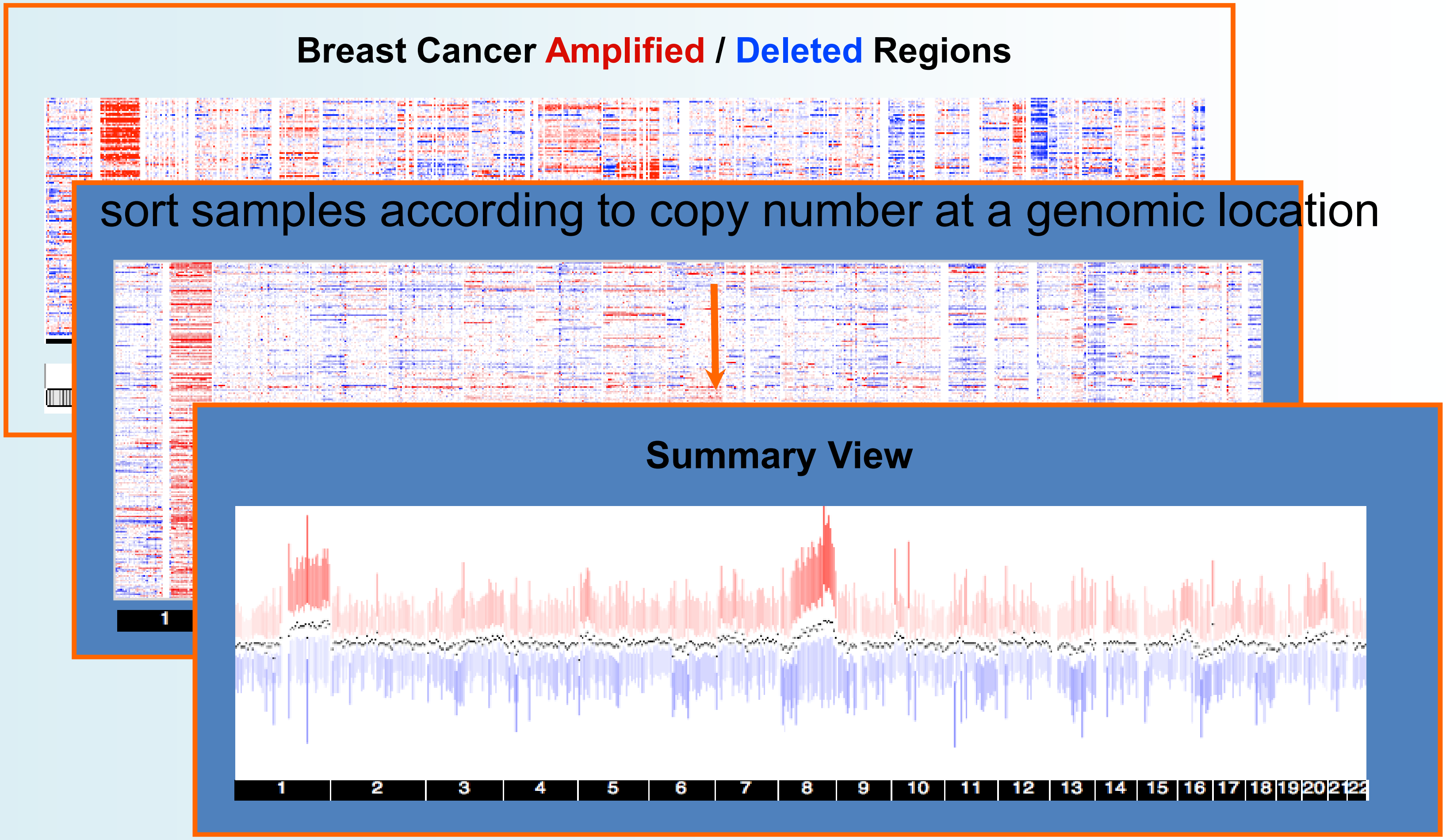
Copy number heatmap



clinical heatmap

TCGA GBM data
478 GBM tumor samples
Segmented CGH

Genomic View of Copy Number Data



Breast Cancer Amplified / Deleted Regions

sort samples according to copy number at a genomic location

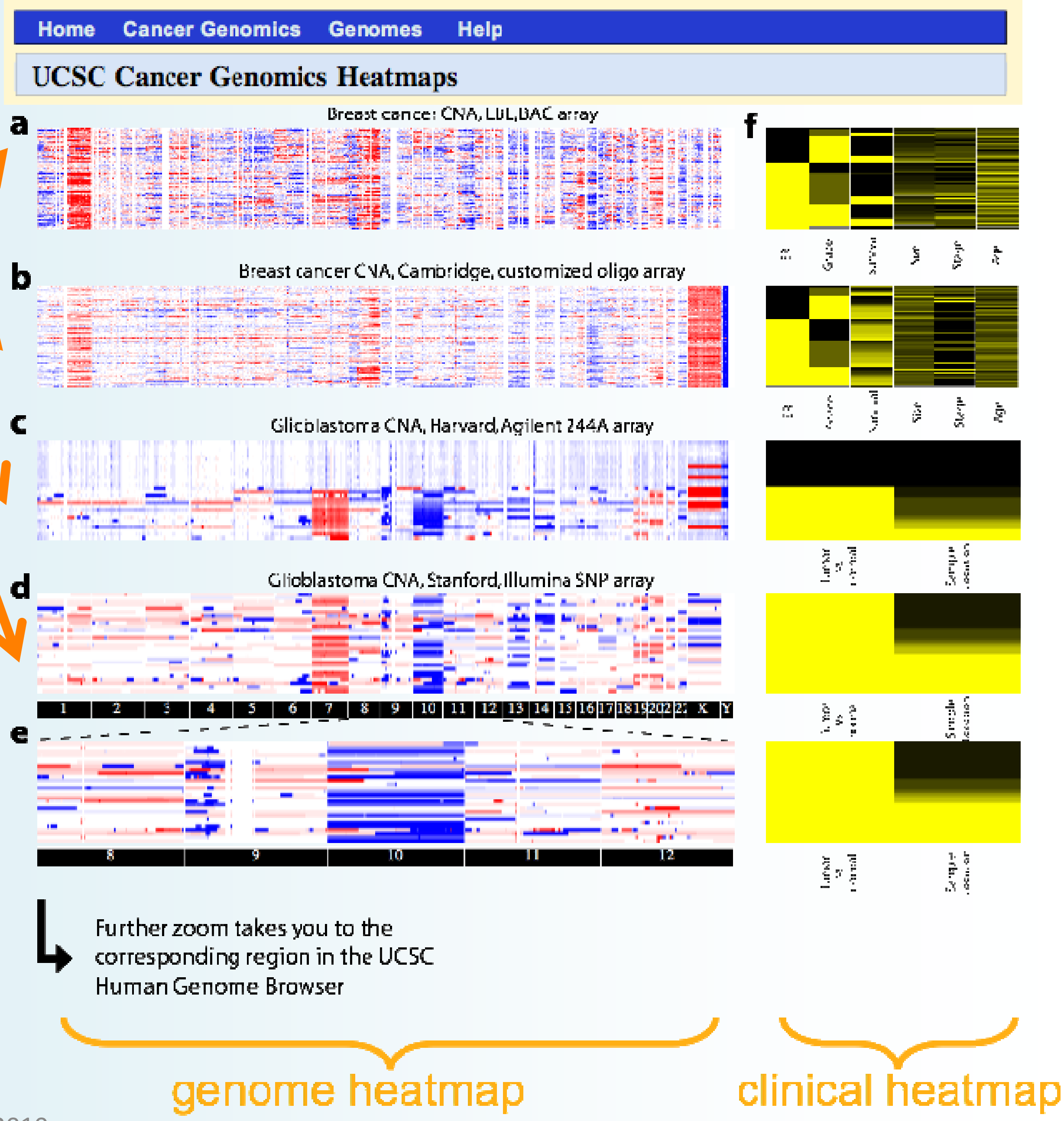
Summary View

UCSC Cancer Genomics Browser

- Simultaneously display patient genomic and clinical data from multiple studies
- Highly configurable *interactive* display
- Online genome-wide statistical tests

- **Public Browser**: genome-cancer.ucsc.edu
- **TCGA Browser** with secured access control: tcga.cse.ucsc.edu

tracks



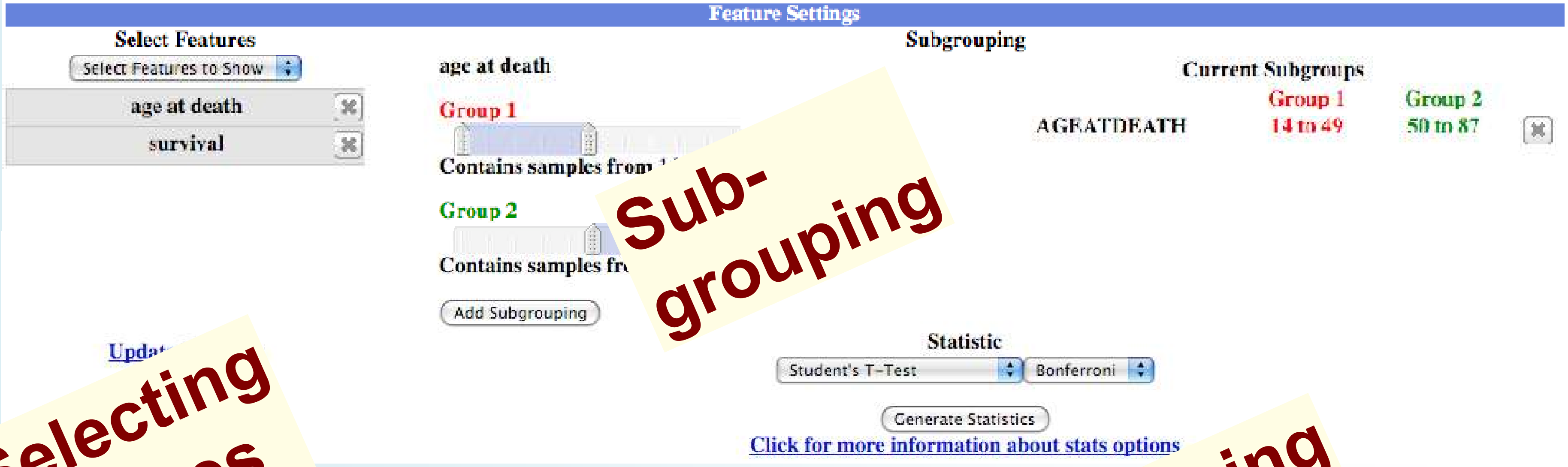
Further zoom takes you to the corresponding region in the UCSC Human Genome Browser

genome heatmap

clinical heatmap

Subgroup Samples using Clinical Features

Feature Setting



	Group 1	Group 2
AGEATDEATH	14 to 49	50 to 87

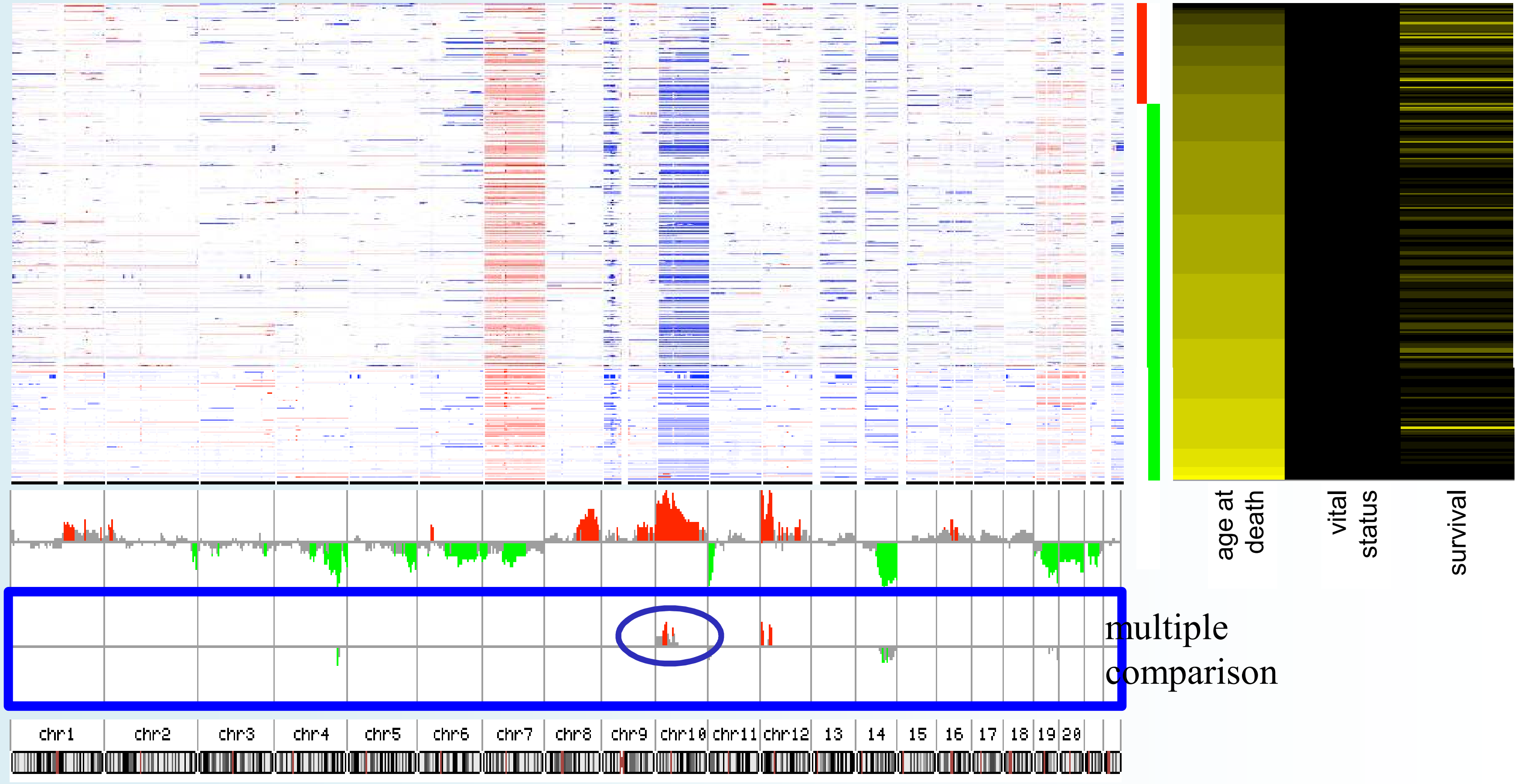
Statistic: Student's T-Test, Bonferroni

Selecting features

Sub-grouping

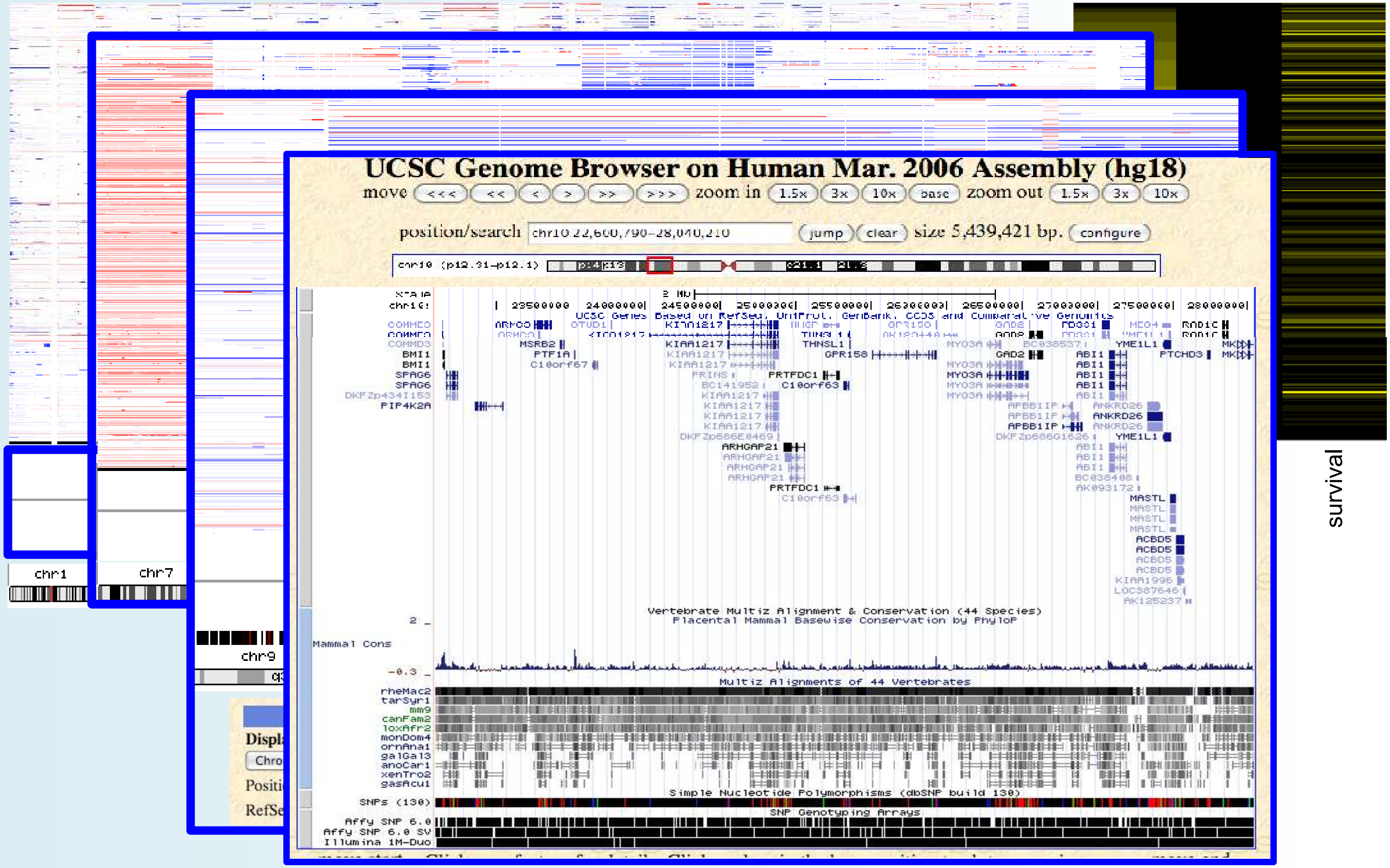
Choosing tests

Genome/clinical heatmap view



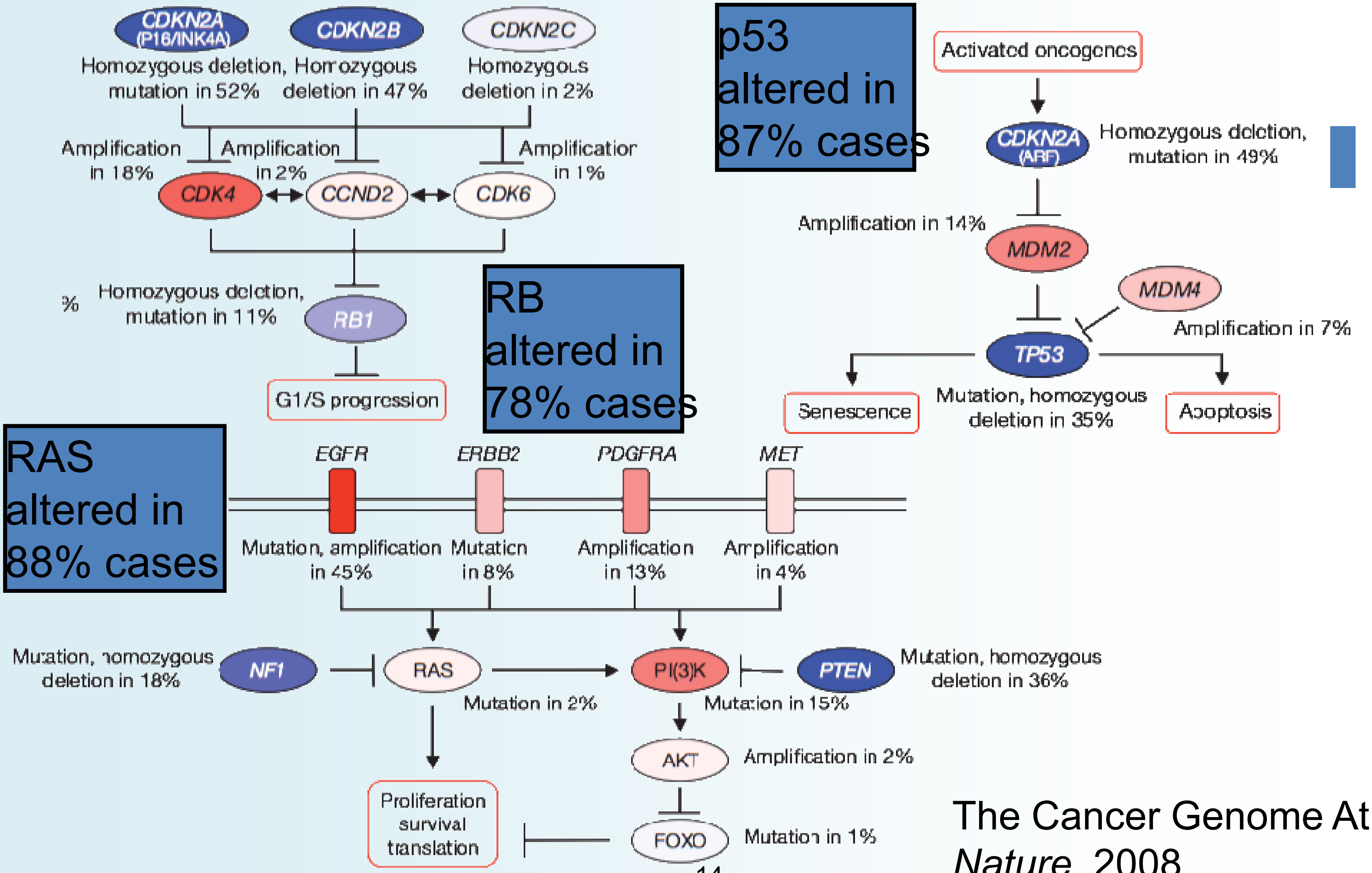
multiple comparison

Zooming in on the heatmap



position/search chr10:22,600,790-28,040,210 size 5,439,421 bp.

Pathways as genetic unit



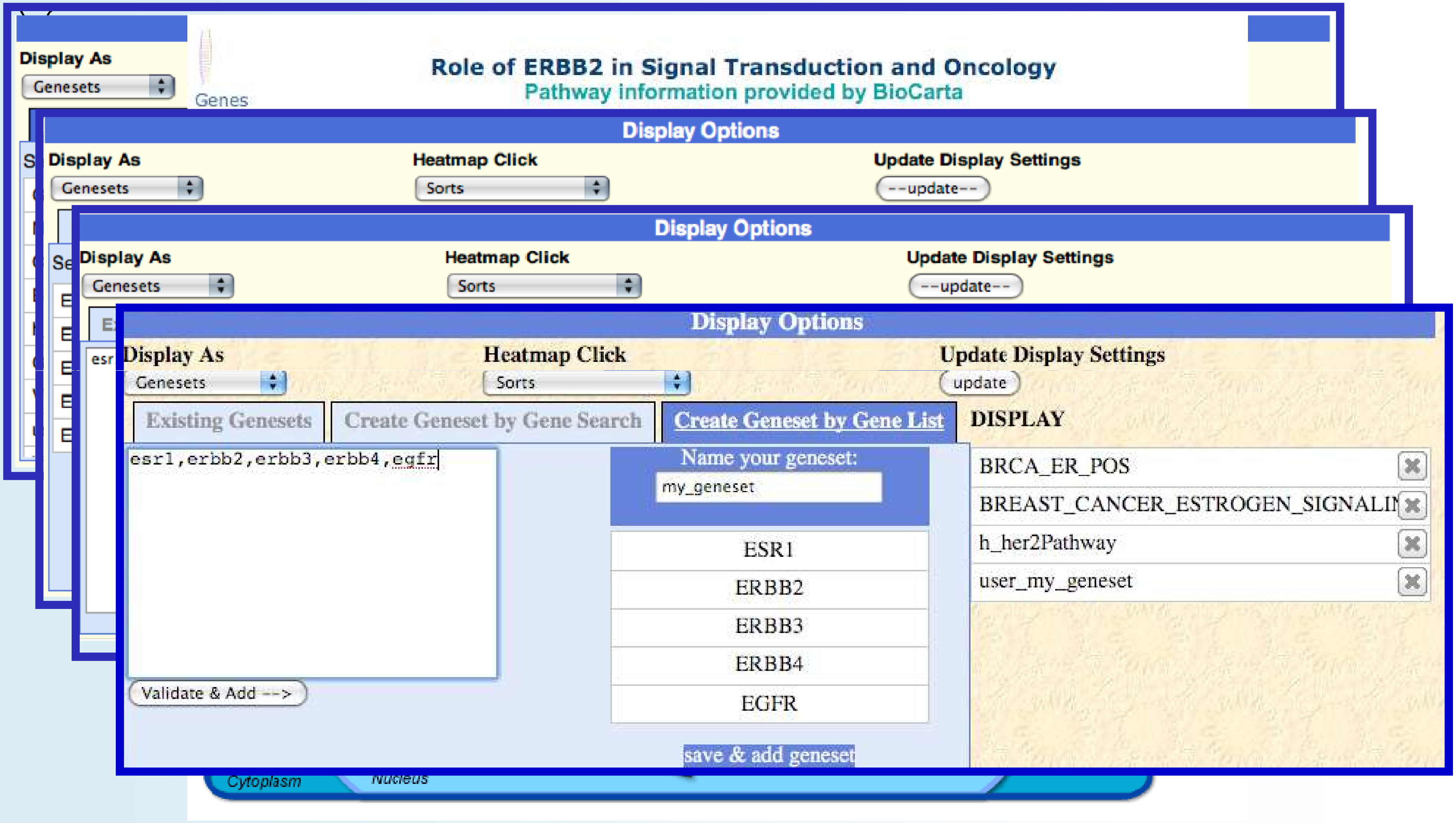
p53 altered in 87% cases

RB altered in 78% cases

RAS altered in 88% cases

The Cancer Genome Atlas
Nature, 2008

Selecting Pathway or Geneset



Role of ERBB2 in Signal Transduction and Oncology

Pathway information provided by BioCarta

Display Options

Display As: Genesets

Heatmap Click: Sorts

Update Display Settings

esr1,erbb2,erbb3,erbb4,egfr

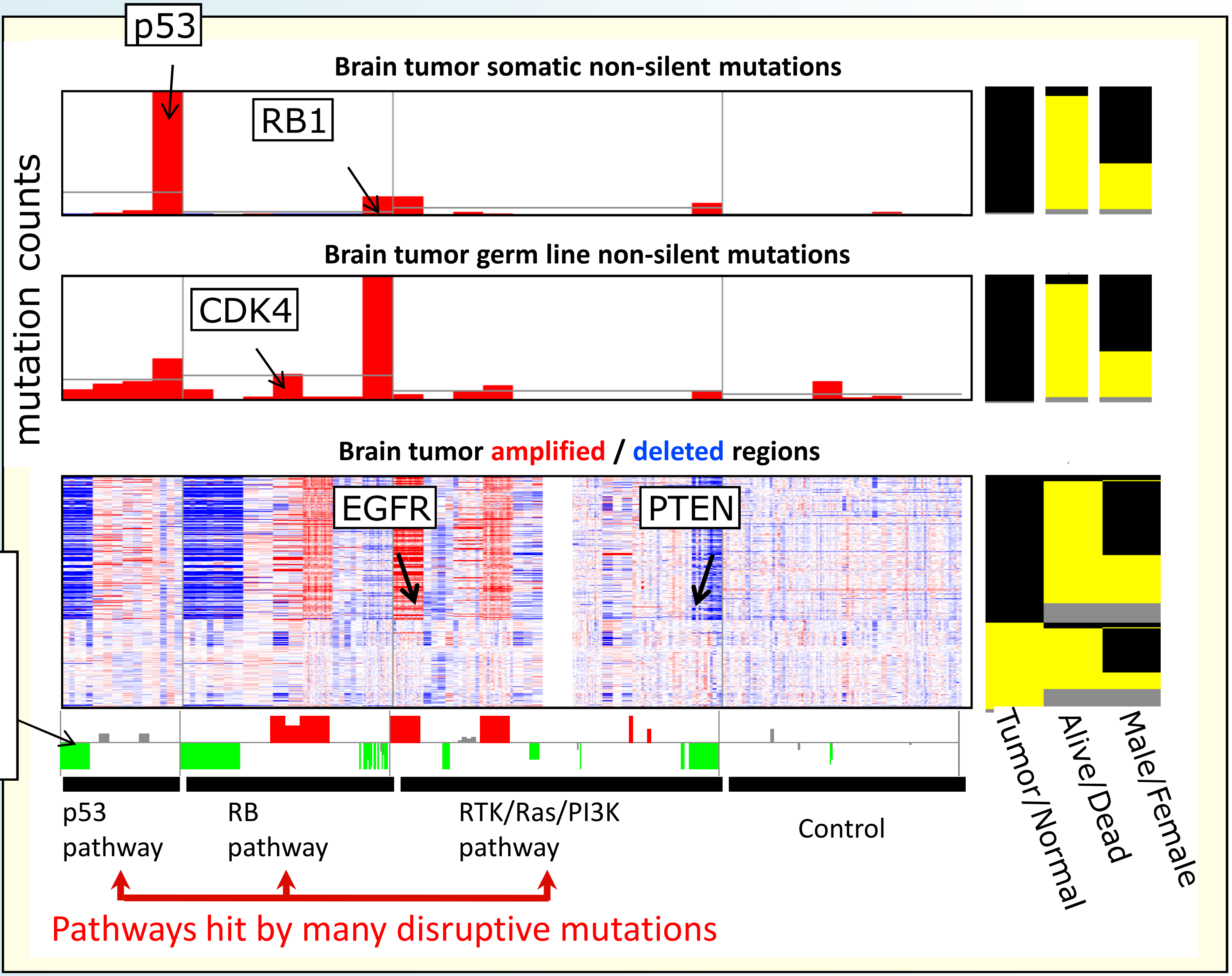
Name your geneset: my_geneset

ESR1
ERBB2
ERBB3
ERBB4
EGFR

DISPLAY

BRCA_ER_POS
BREAST_CANCER_ESTROGEN_SIGNALIN
h_her2Pathway
user_my_geneset

Geneset View: Deletions, duplications and mutations disrupt interacting pathways of genes



t-test:
red = signif. up in tumor,
green = signif. down in tumor

Pathways hit by many disruptive mutations

TCGA Research Network, *Nature* 2008

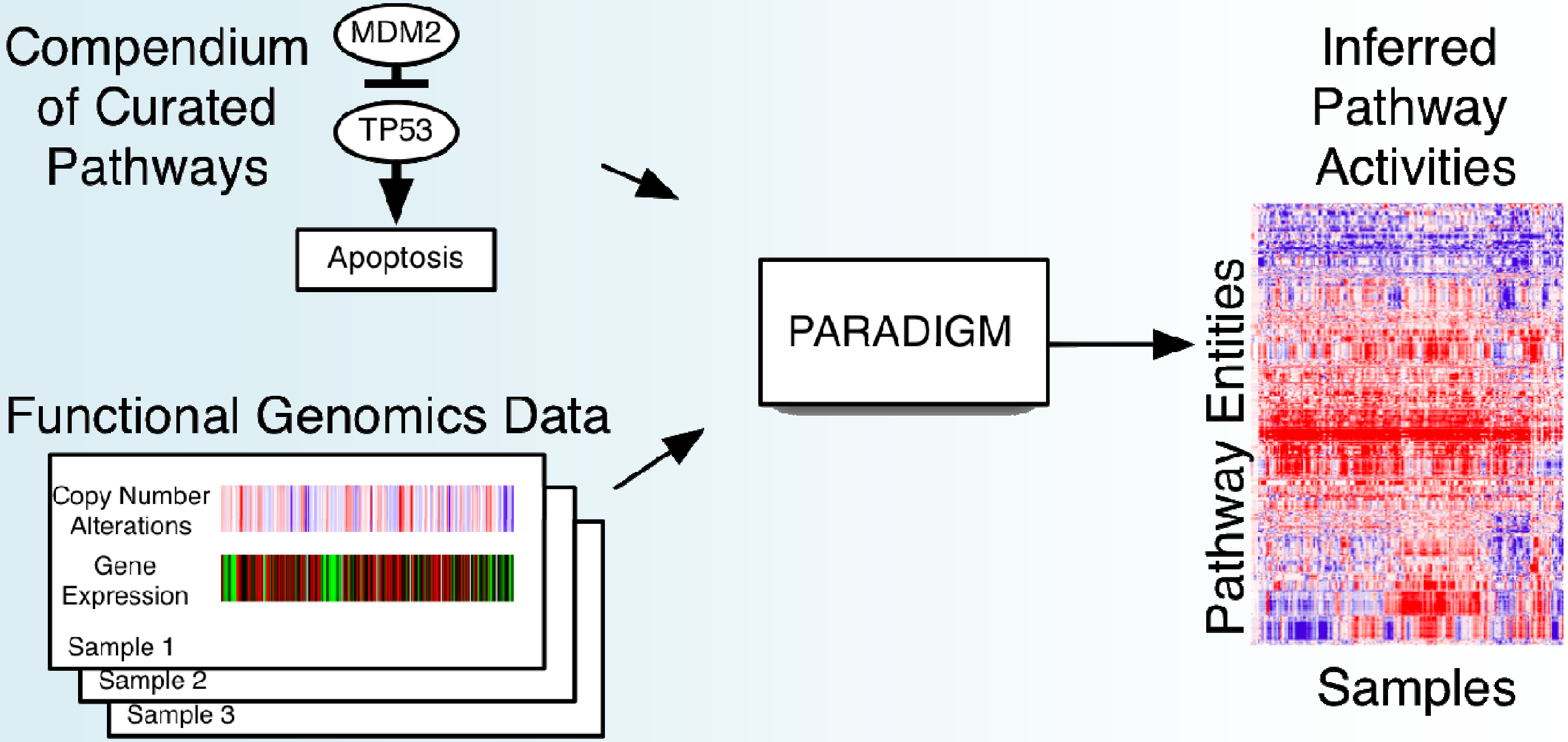
Outline

- High-throughput visualization w/ the UCSC Cancer Genomics Browser
- Identifying patient-specific pathway alterations
- Identifying novel pathway interactions w/ genetic knock-outs

Patient-specific alterations for *known* pathways

- Collect publicly available pathways (NCI, WikiPathways, KEGG, …)
- Convert to graphical model
- Infer pathway "entities" for each patient sample
  - e.g. gene activities, apoptosis, DNA repair, small molecules, complexes, etc.
- Train classifiers w/ pathway activities
  - E.g. 3-year disease-free survival

PARADIGM Overview



Compendium of Curated Pathways

MDM2

TP53

Apoptosis

Functional Genomics Data

Copy Number Alterations

Gene Expression

Sample 1

Sample 2

Sample 3

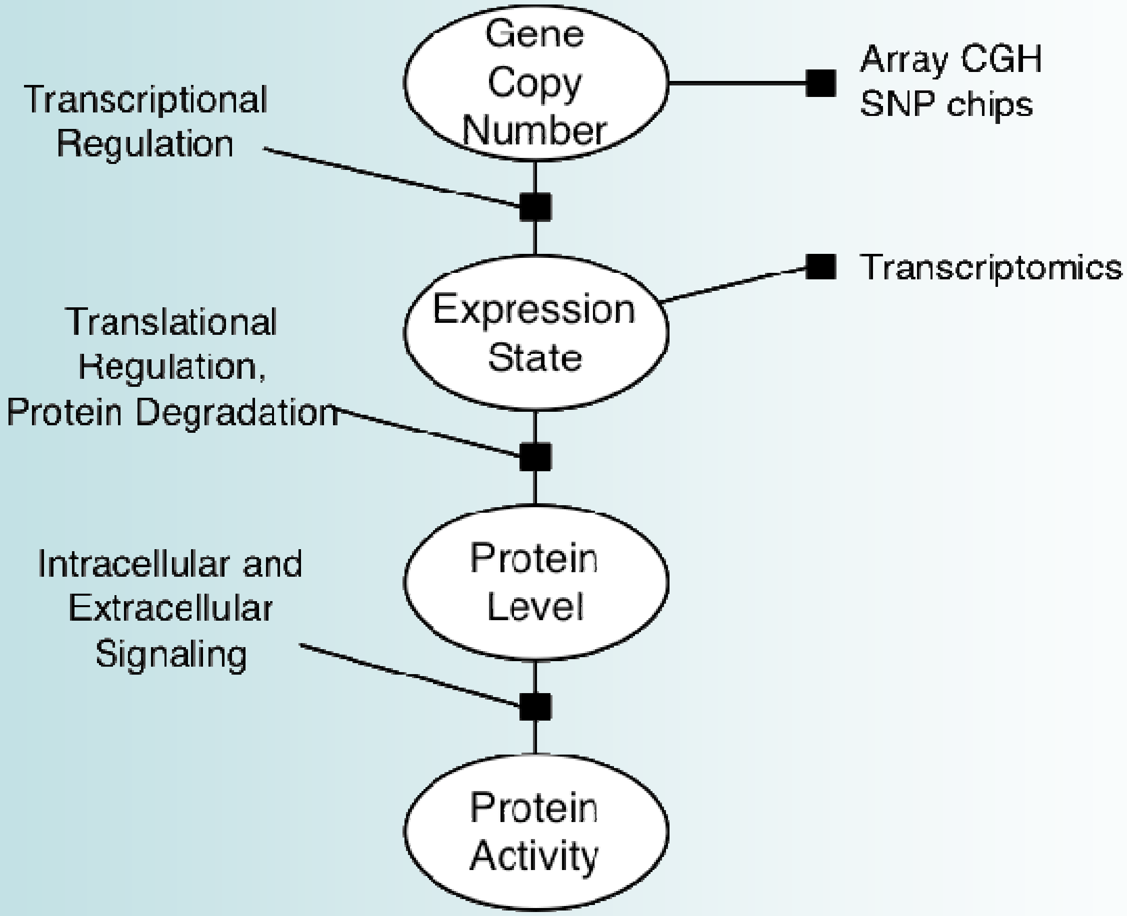
PARADIGM

Inferred Pathway Activities

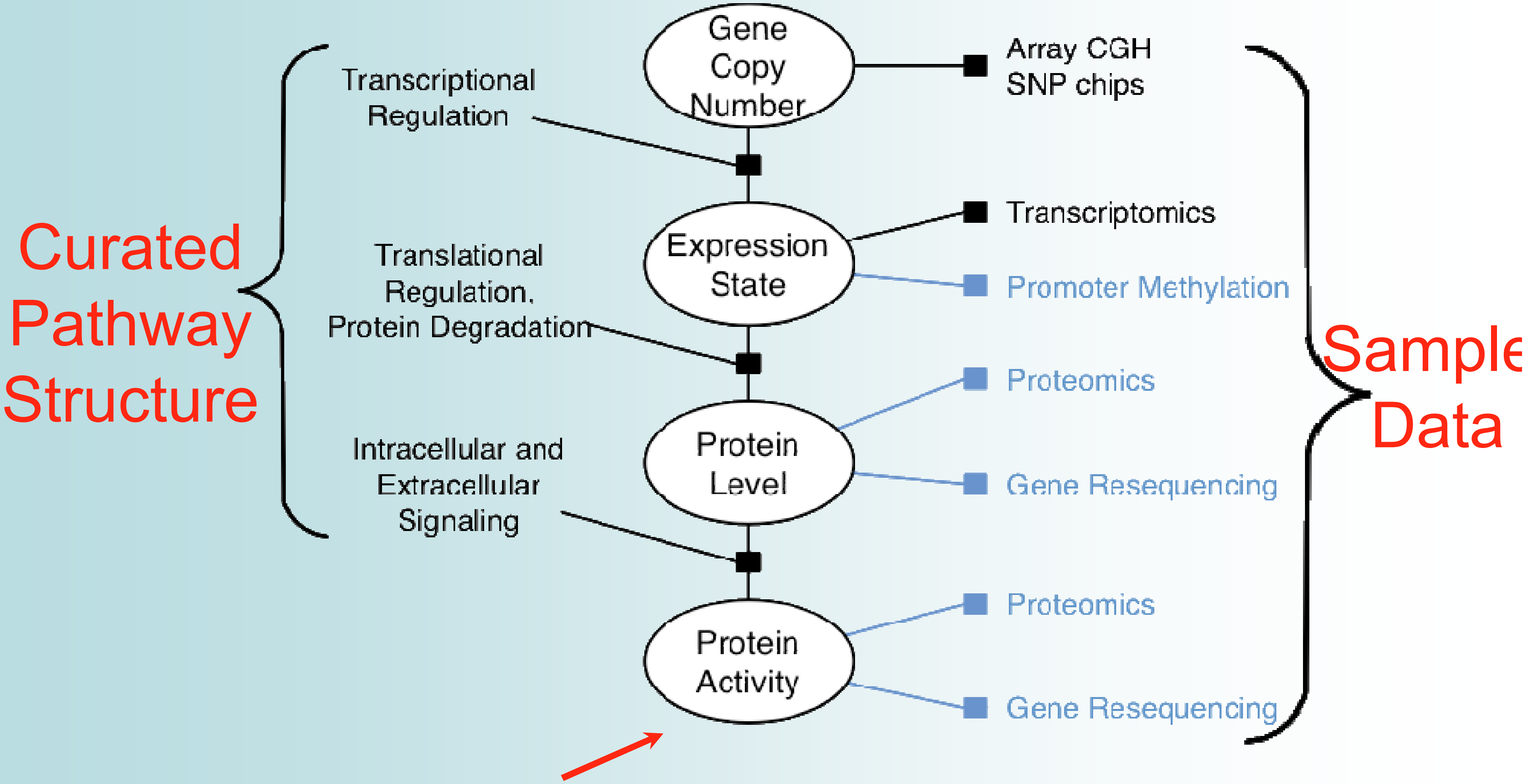
Pathway Entities

Samples

Integration Model

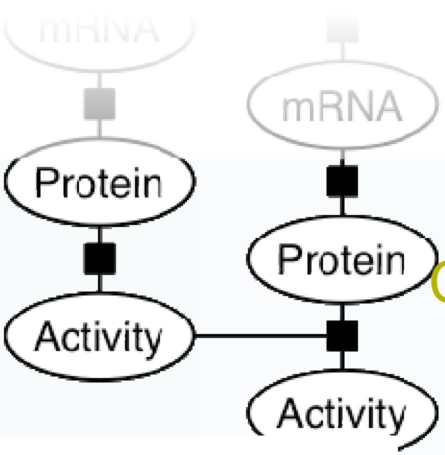


Pathway Activities

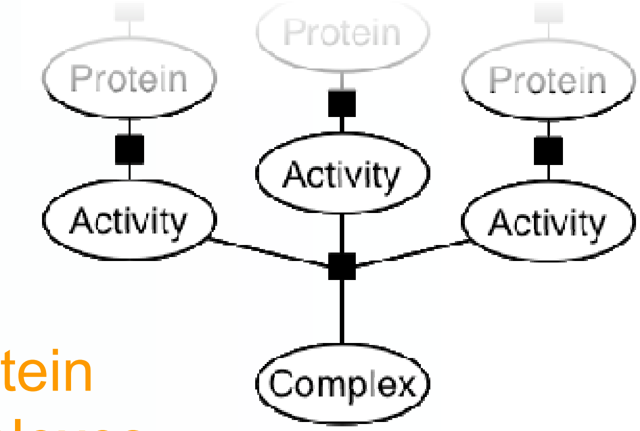


Activity Estimates

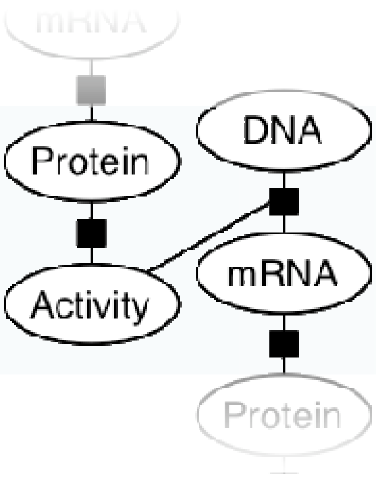
Pathway Components



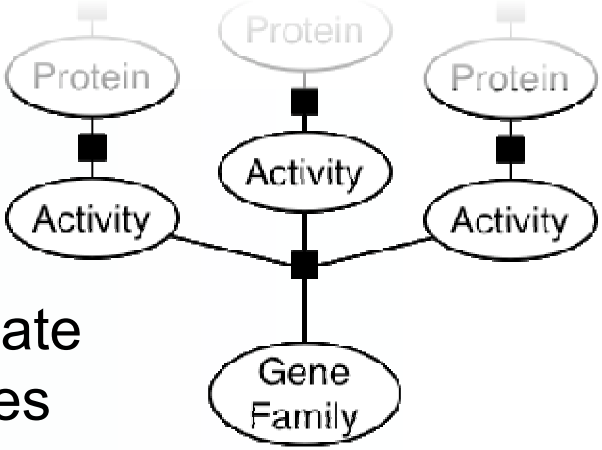
Kinase Cascade



Protein Complexes

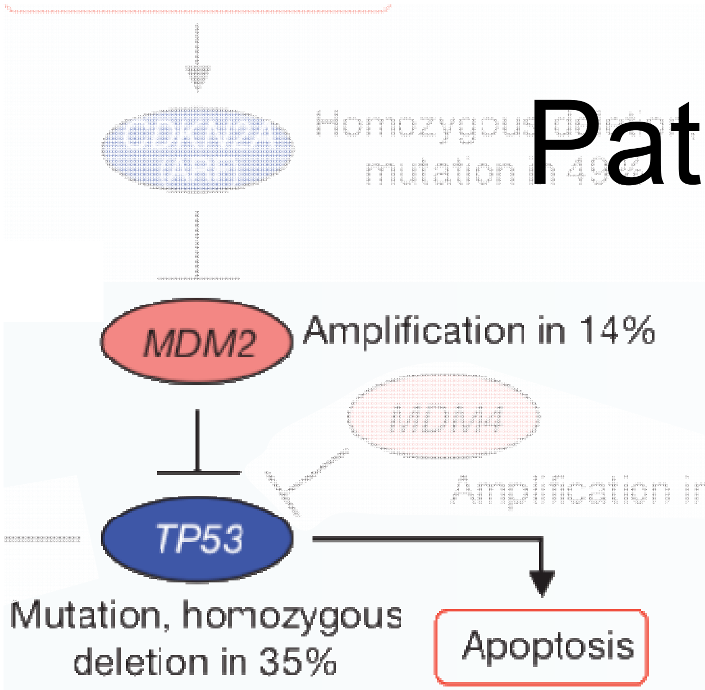


Transcriptional Regulation

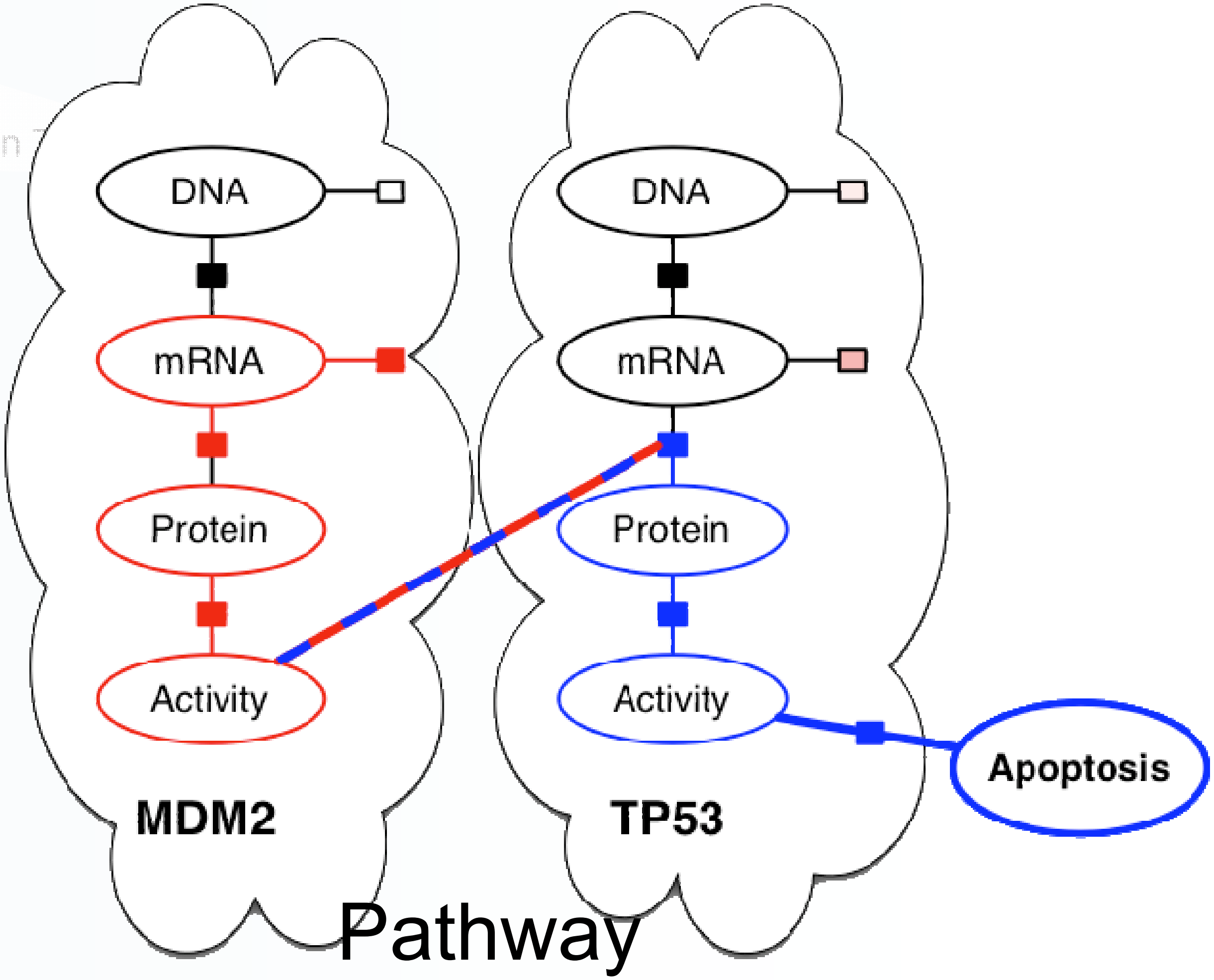


Alternate Genes

Pathway Context



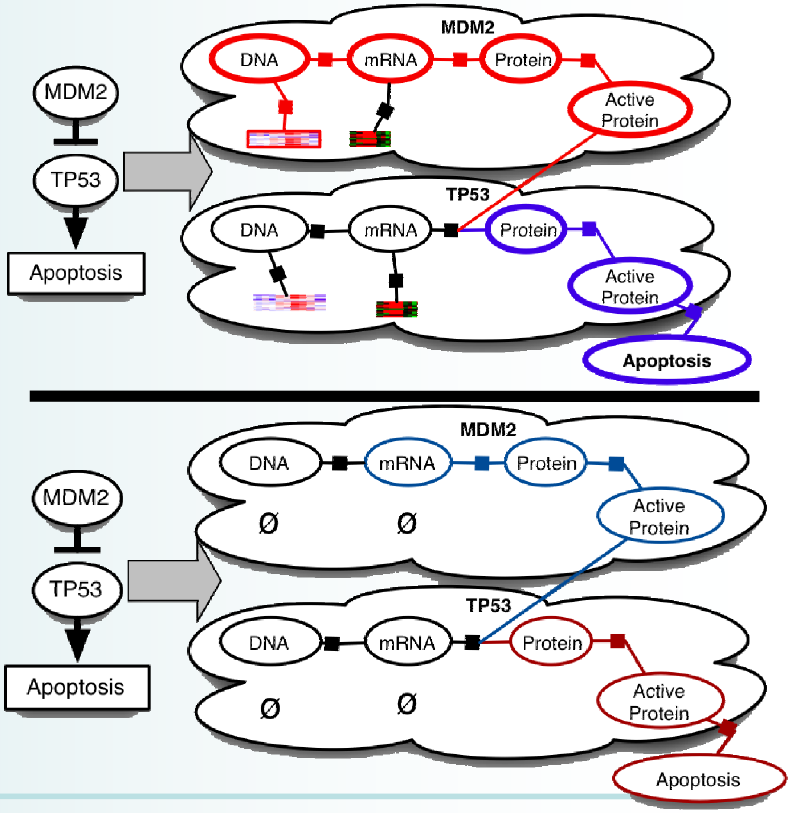
Pathway Schematic



Pathway Schematic

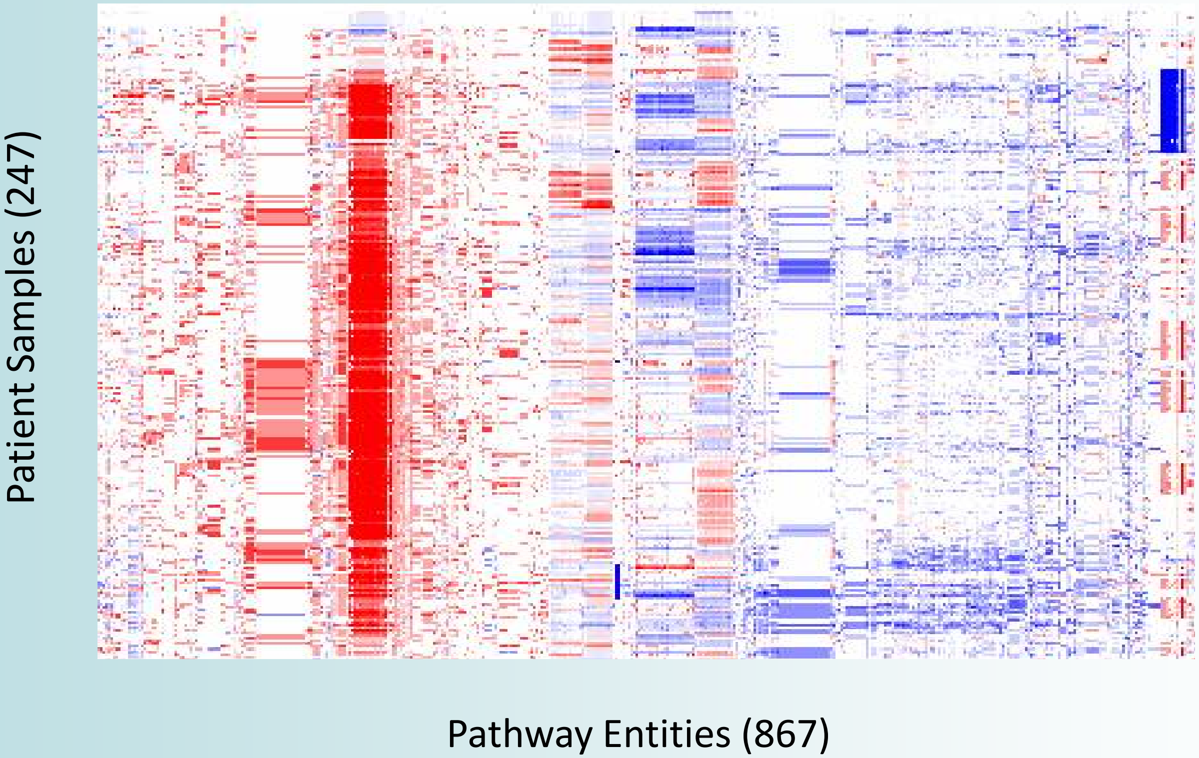
Inferred Pathway Activities

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

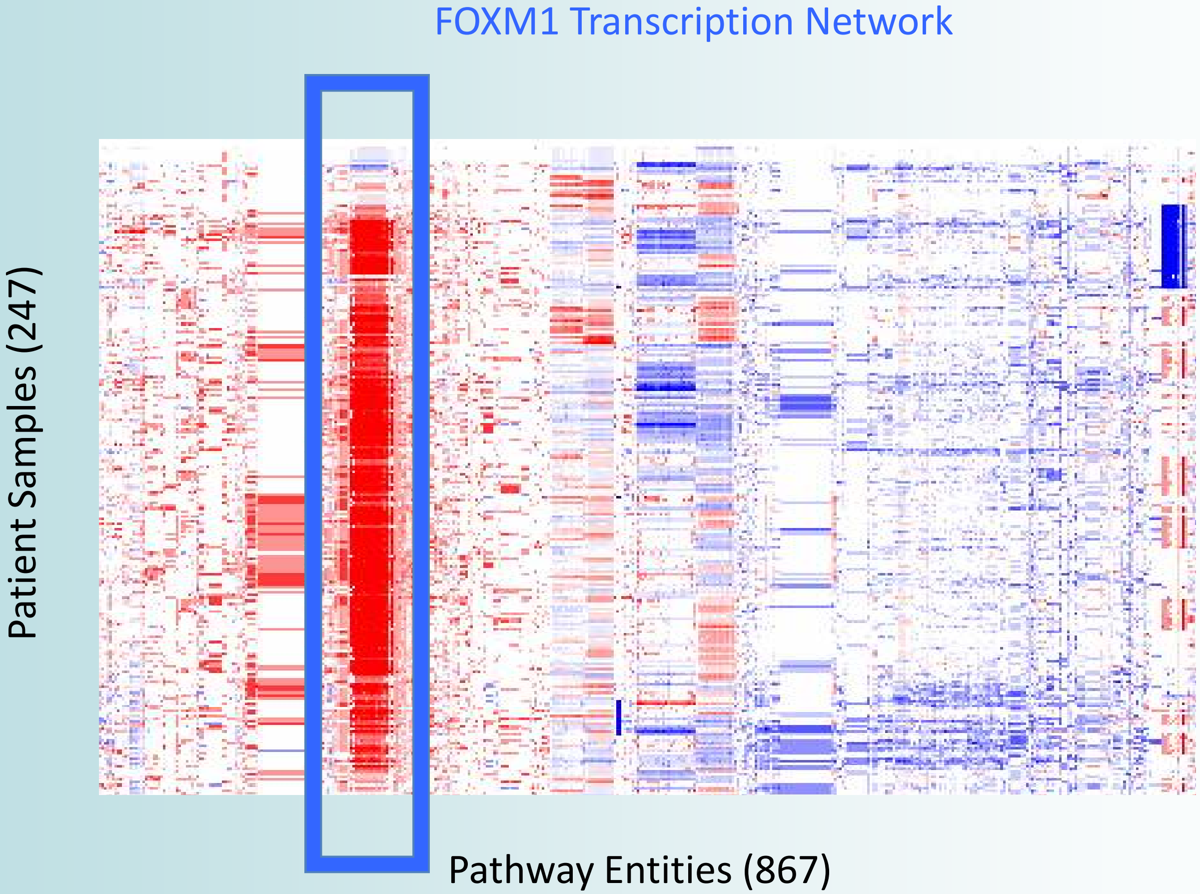


$$IPA(i) = \begin{cases} L(i,1) & \text{L(i,1)>L(i,-1) and L(i,1)>L(i,0)} \\ -L(i,-1) & \text{L(i,-1)>L(i,1) and L(i,-1)>L(i,0)} \\ 0 & \text{otherwise.} \end{cases}$$

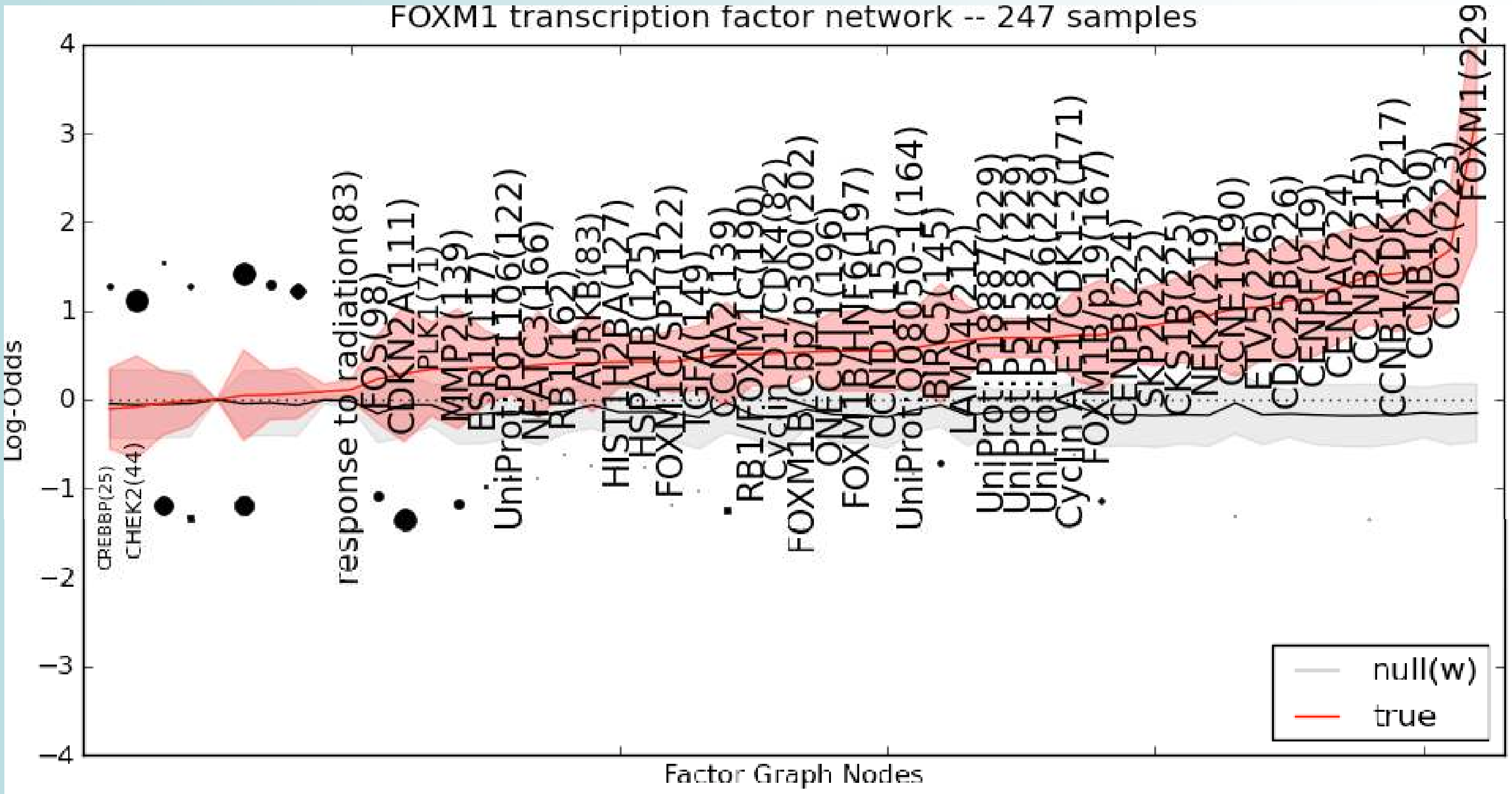
Ovarian Cancer
Inferred Pathway Activities



FOXM1 pathway altered in majority of OV patients

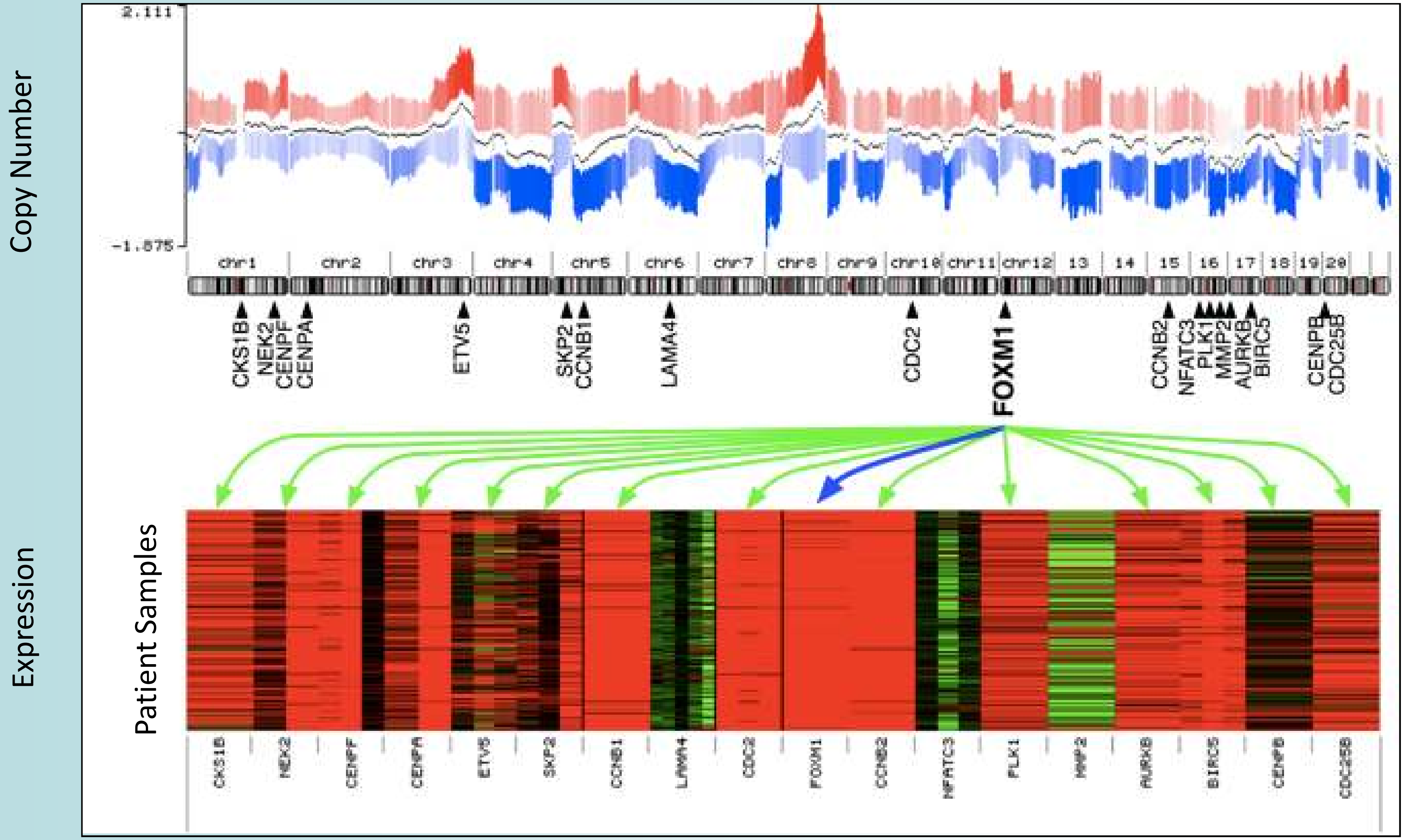


FOXM1 Pathway Significance

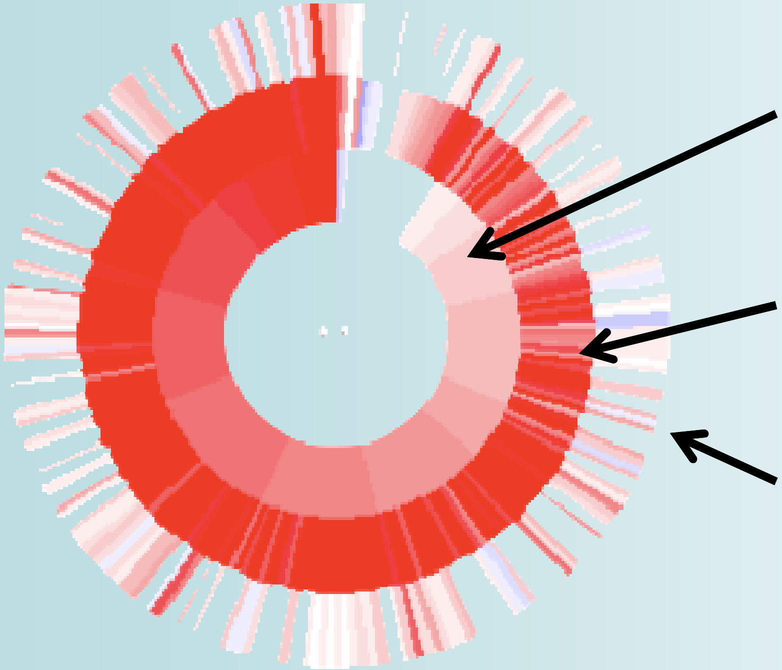


- Significance assessed w/ *simulated networks*
- TF known to play a role in many cancers including breast and cervical cancers.

FOXM1 Amplification Leads to Secondary Responses Predicted by its Pathway Interactions



CircleMap of FOXM1 Transcription Factor and Targets



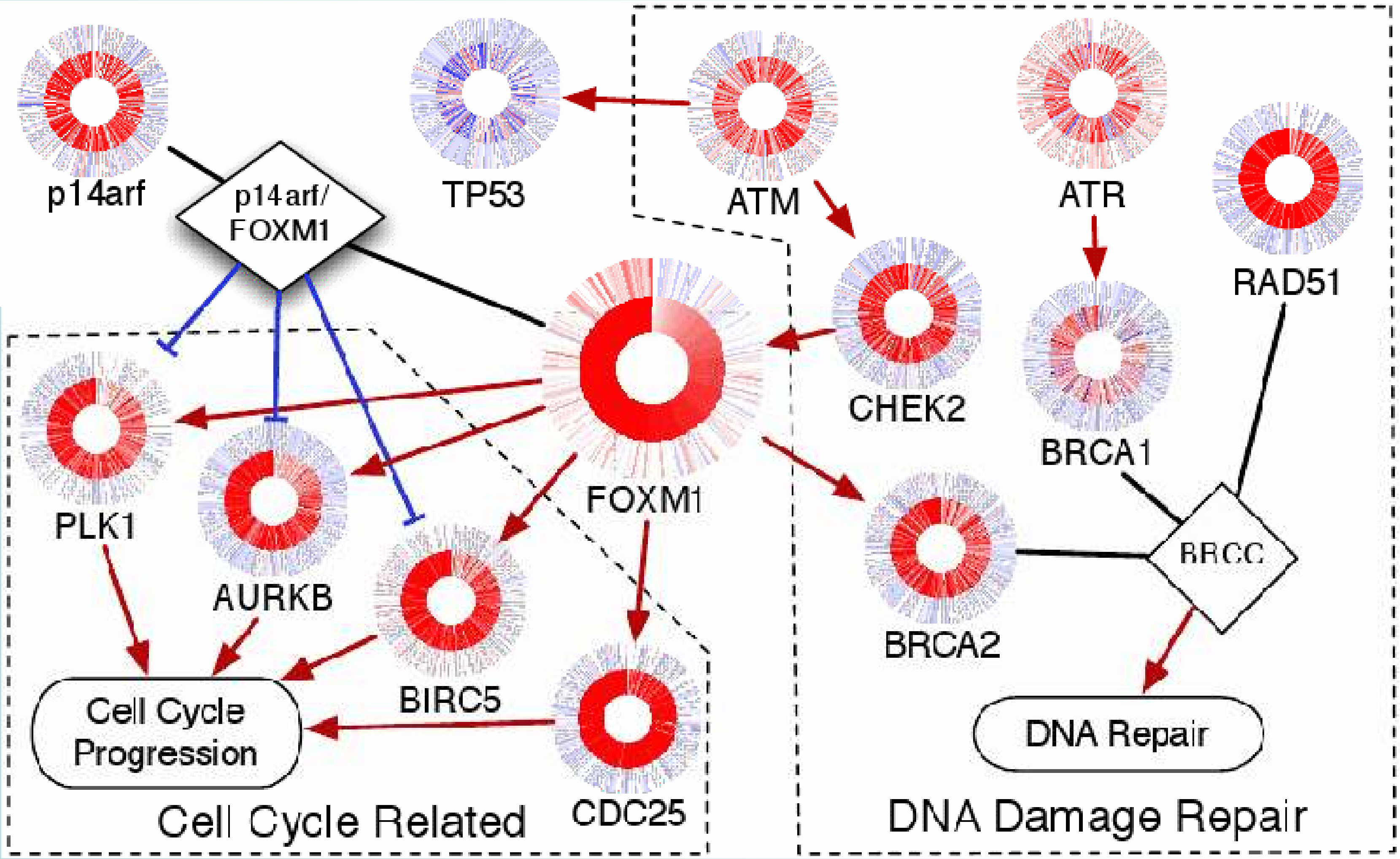
IPAs From PARADIGM

Expression vs. Fallopian Normal

Copy Number Variation

Each tickmark around the circle represents a single sample

FOXM1 central to DNA repair and cell proliferation



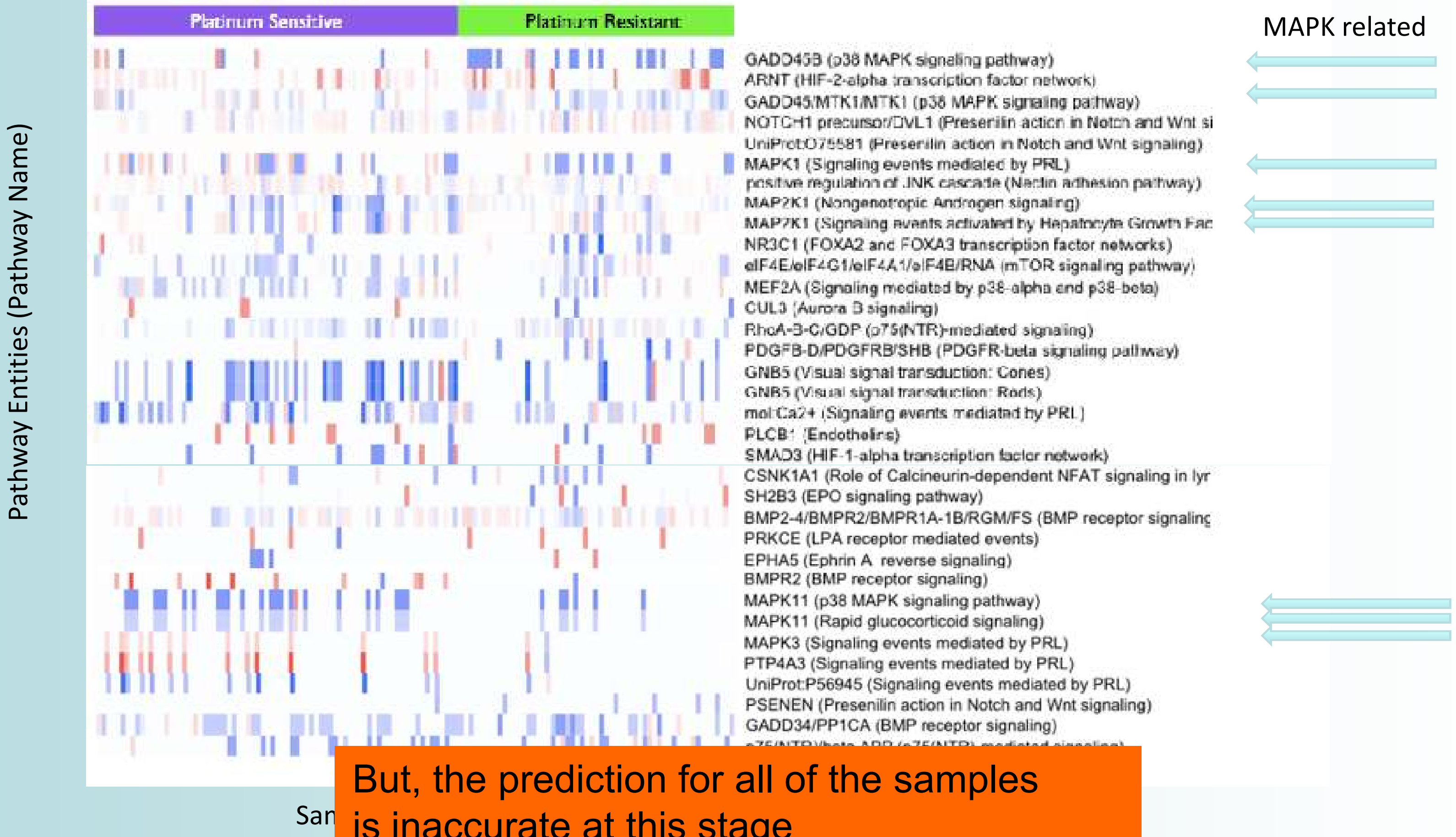
Predict Platinum Sensitivity with Machine-Learning Classifiers

- Are any pathway entities correlated with Platinum response?

- Can we train a classifier to use a combination of activities to improve prediction of platinum sensitivity?

Pathway Entities Improve Platinum Sensitivity Predictions

Top Predictive Pathway Entities for Platinum Sensitivity



But, the prediction for all of the samples is inaccurate at this stage

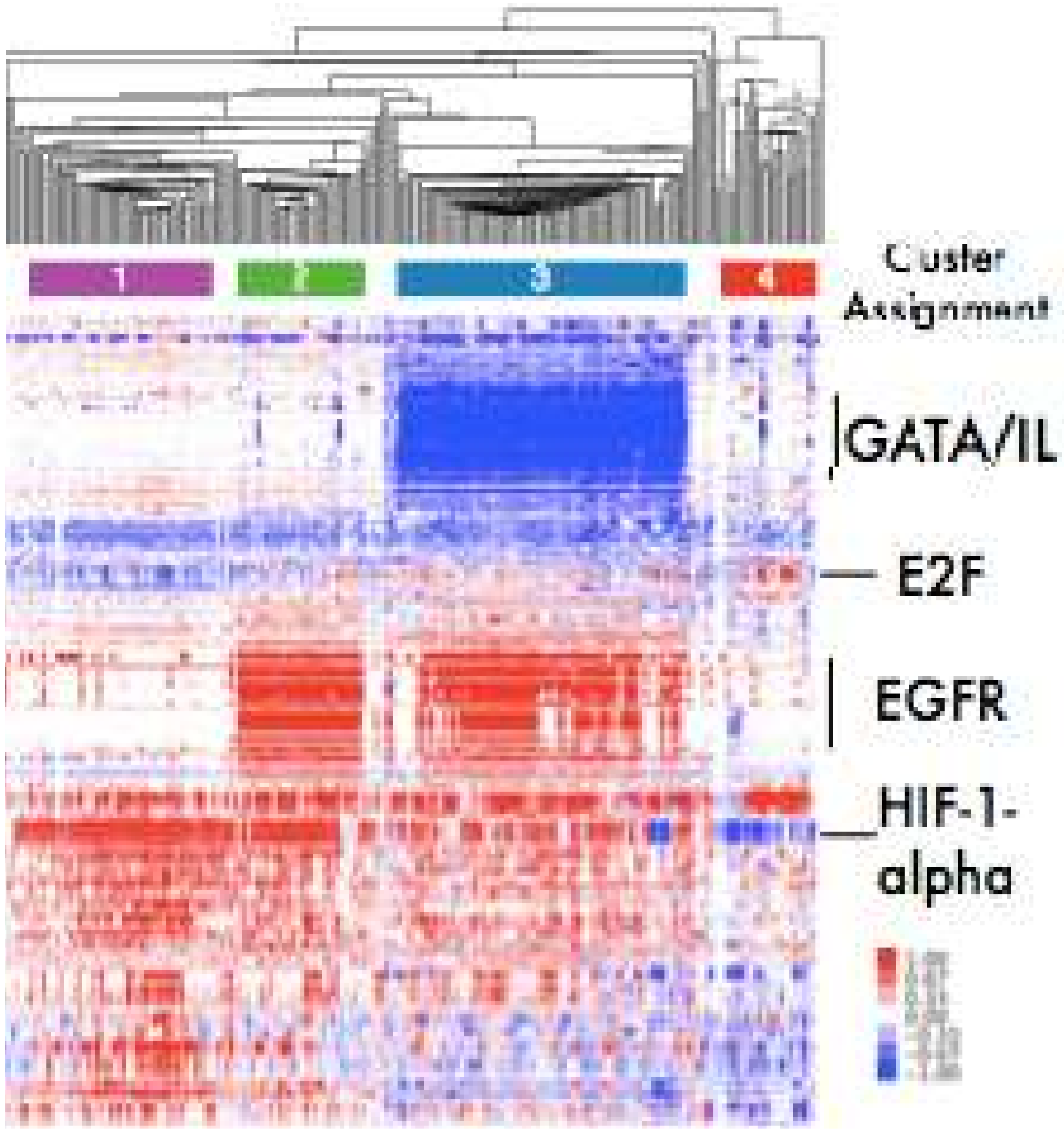
Prediction Accuracy

Expression Data	CGH Data	Expression & CGH Data	Pathway Inference Data
< 50%	< 50%	53.15%	**75%**

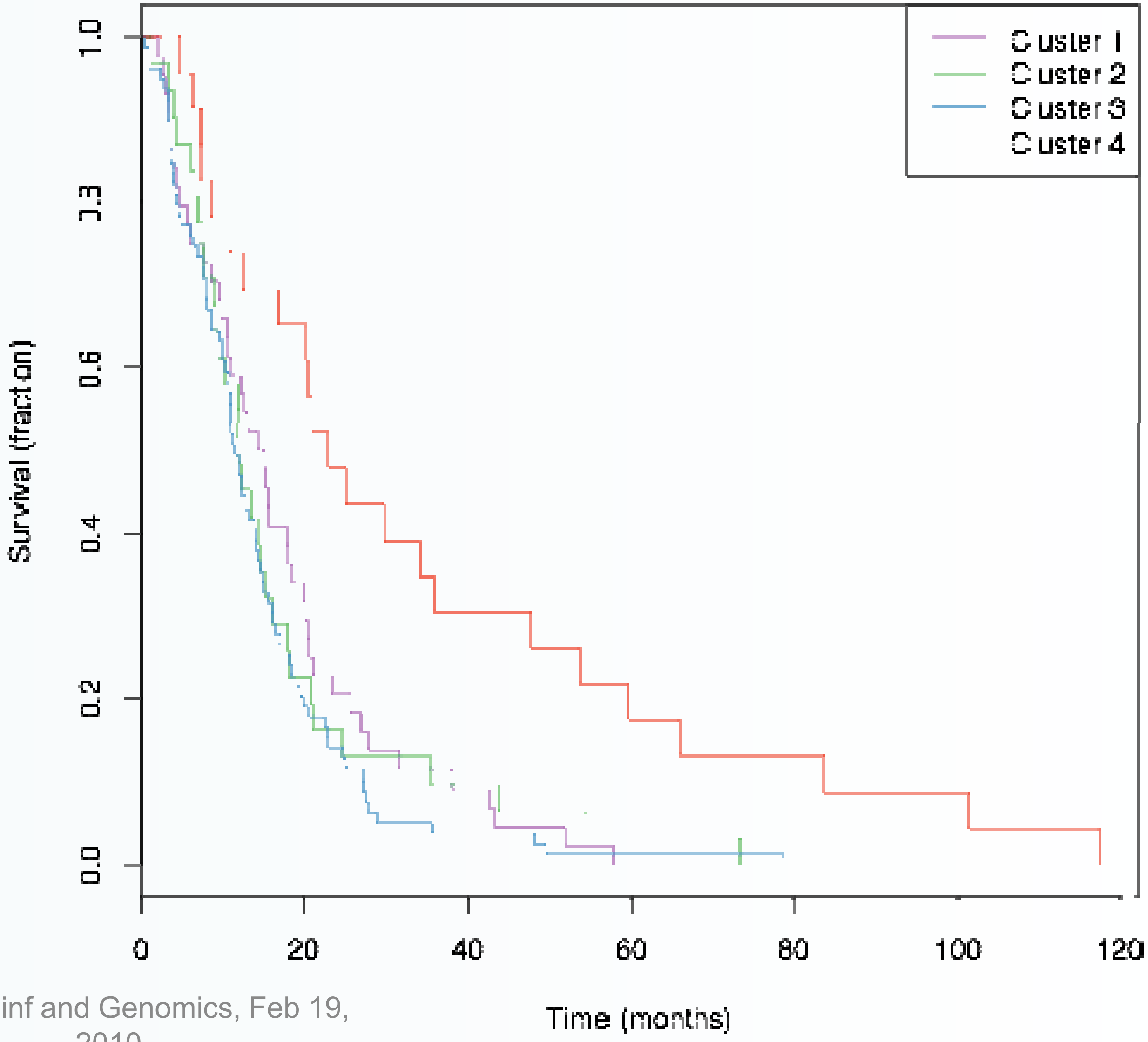
Predict on Patient Sub-groups

- Classifiers forced to make a prediction for all patients are not highly accurate (75%).
- Can we accurately classify a subset of patients?
- Idea: allow classifiers to return “I don’t know” for a sample.
  - A two-step classifer: 1) Decide if sample is classifiable; 2) If so, classify as sensitive/insensitive
- We can predict, at ~90% accuracy for 23% of the samples (35 out of 153)
  - 22 predicted as low (1 incorrect)
  - 12 predicted as high (4 incorrect)
- (Still working on this; results are preliminary)

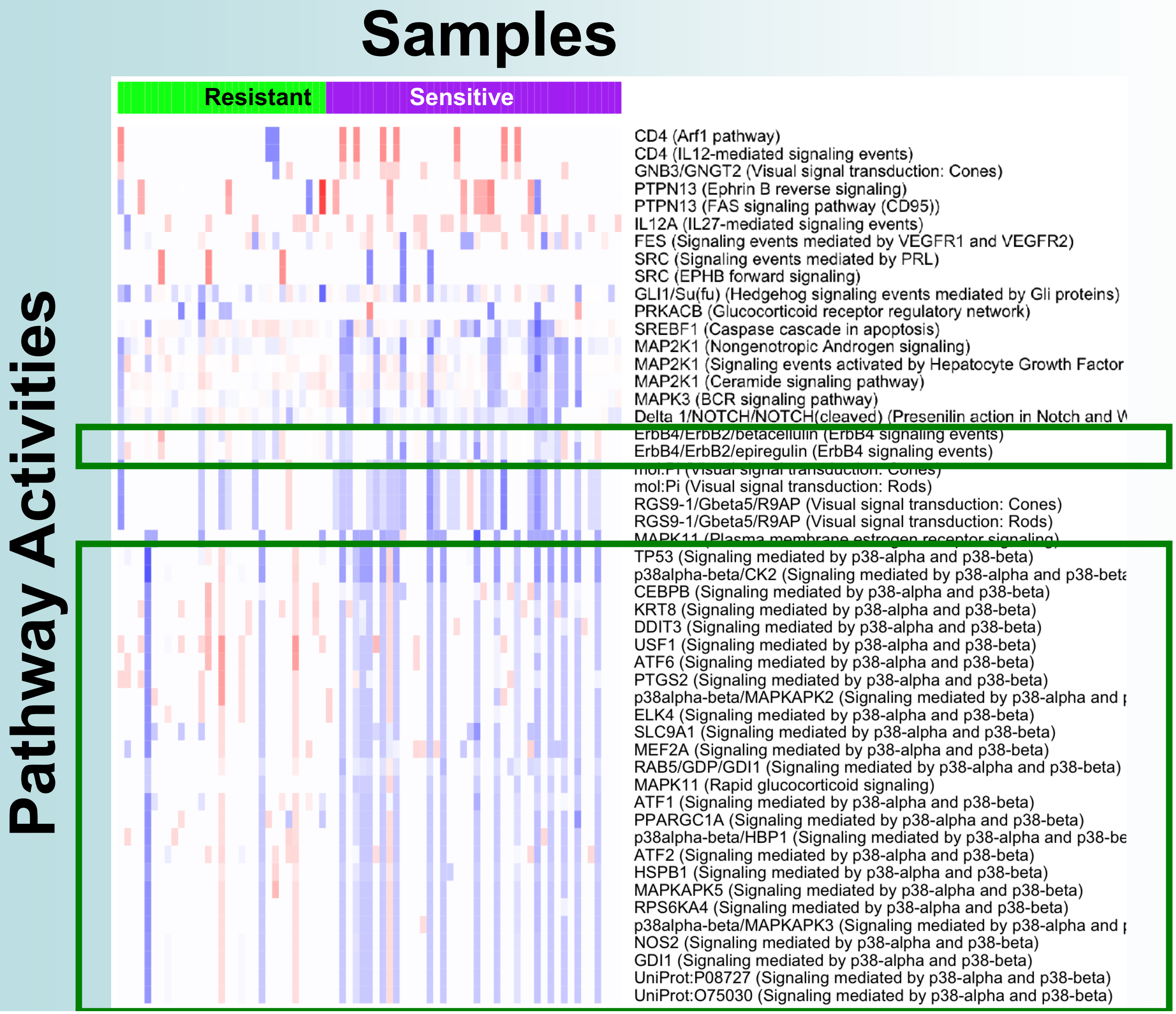
PARADIGM Outputs (IPAs) Create Obvious Clusters in GBM



GBM Clusters Significantly Separate Patients on Survival



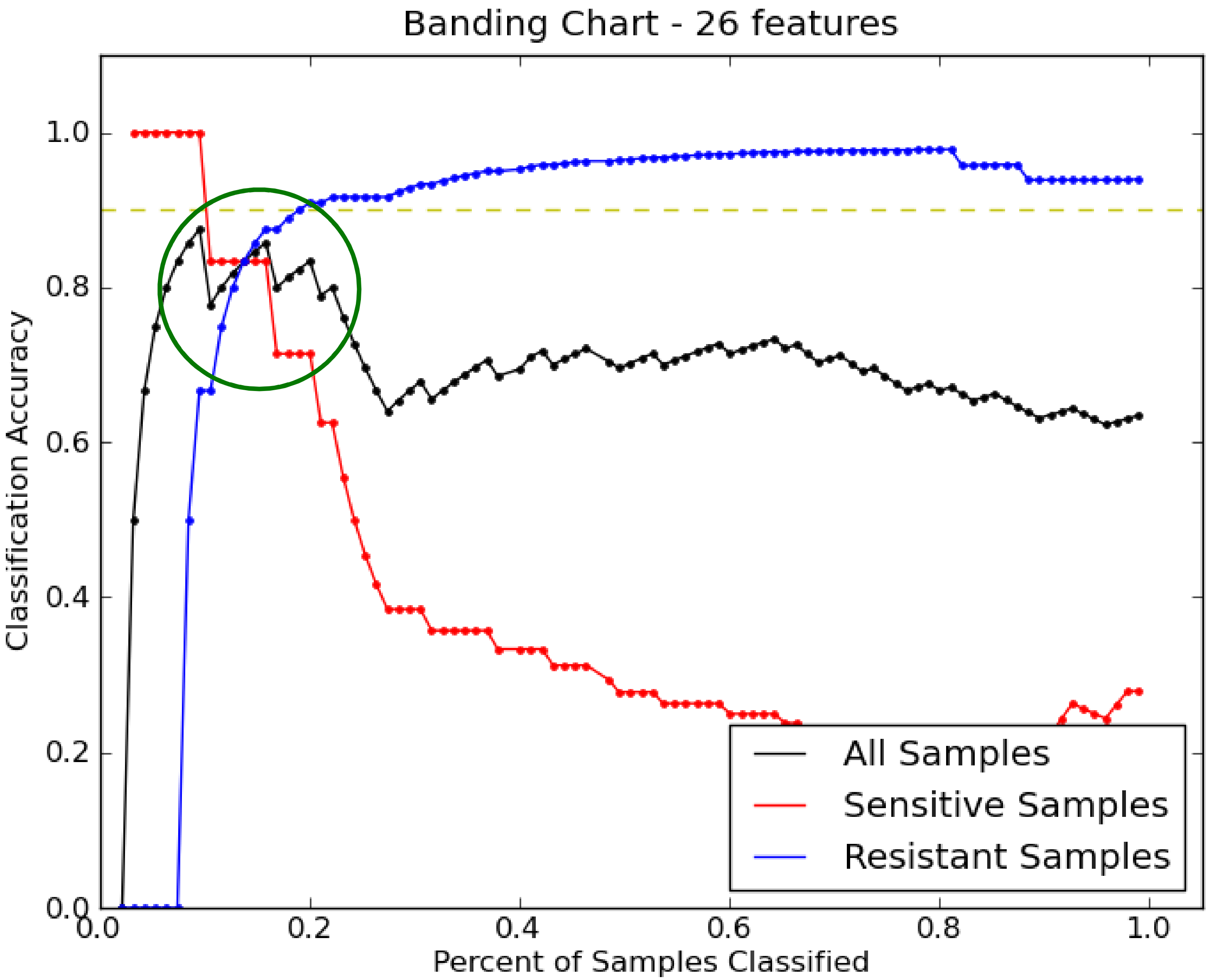
Top Ovarian Cancer Features Using Pathway Scores



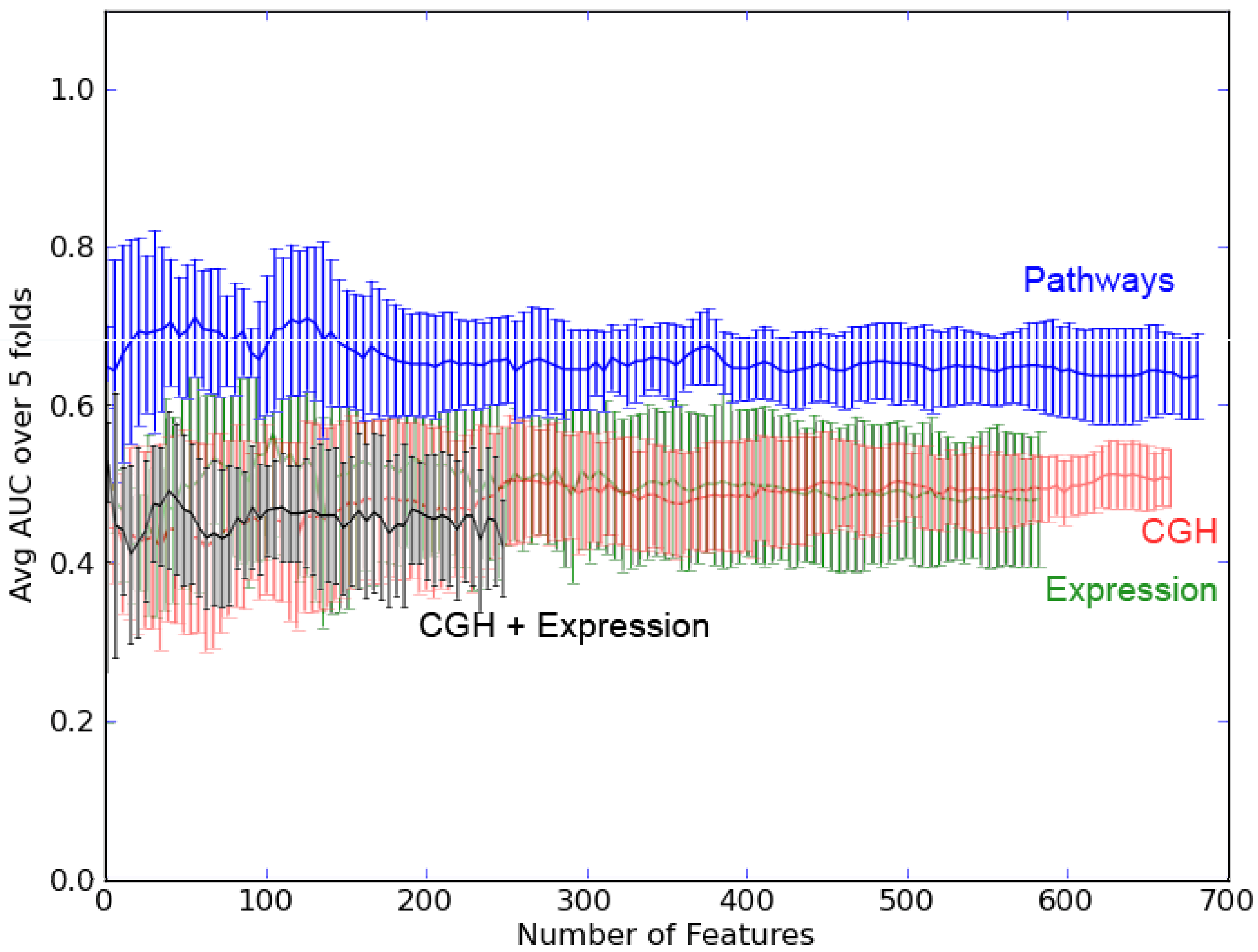
ErbB4 pathway

p38 pathway

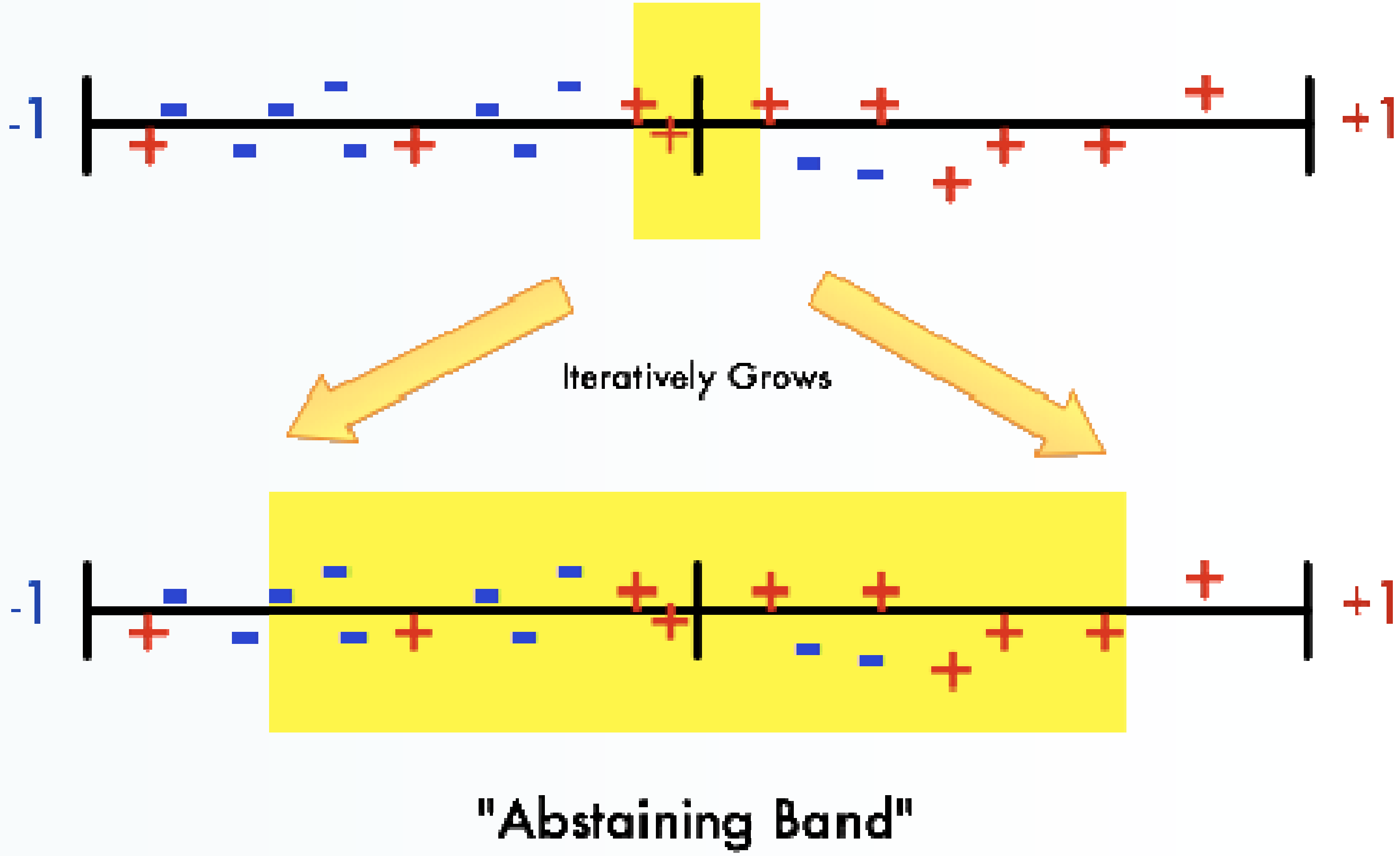
Classifying Platinum Sensitivity With Pathway Scores



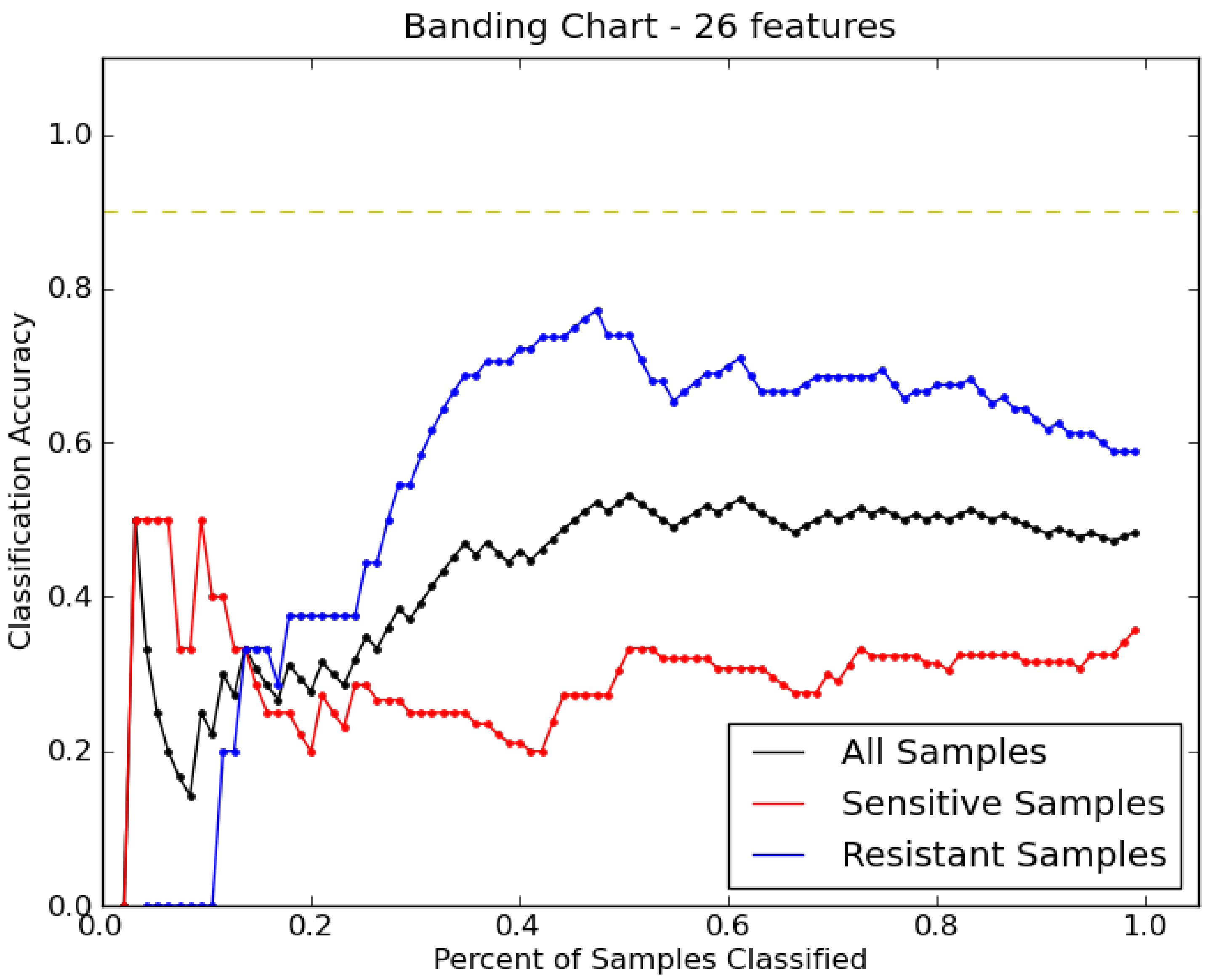
Pathway Activities More Informative for Predicting Platinum Sensitivity than Input Data



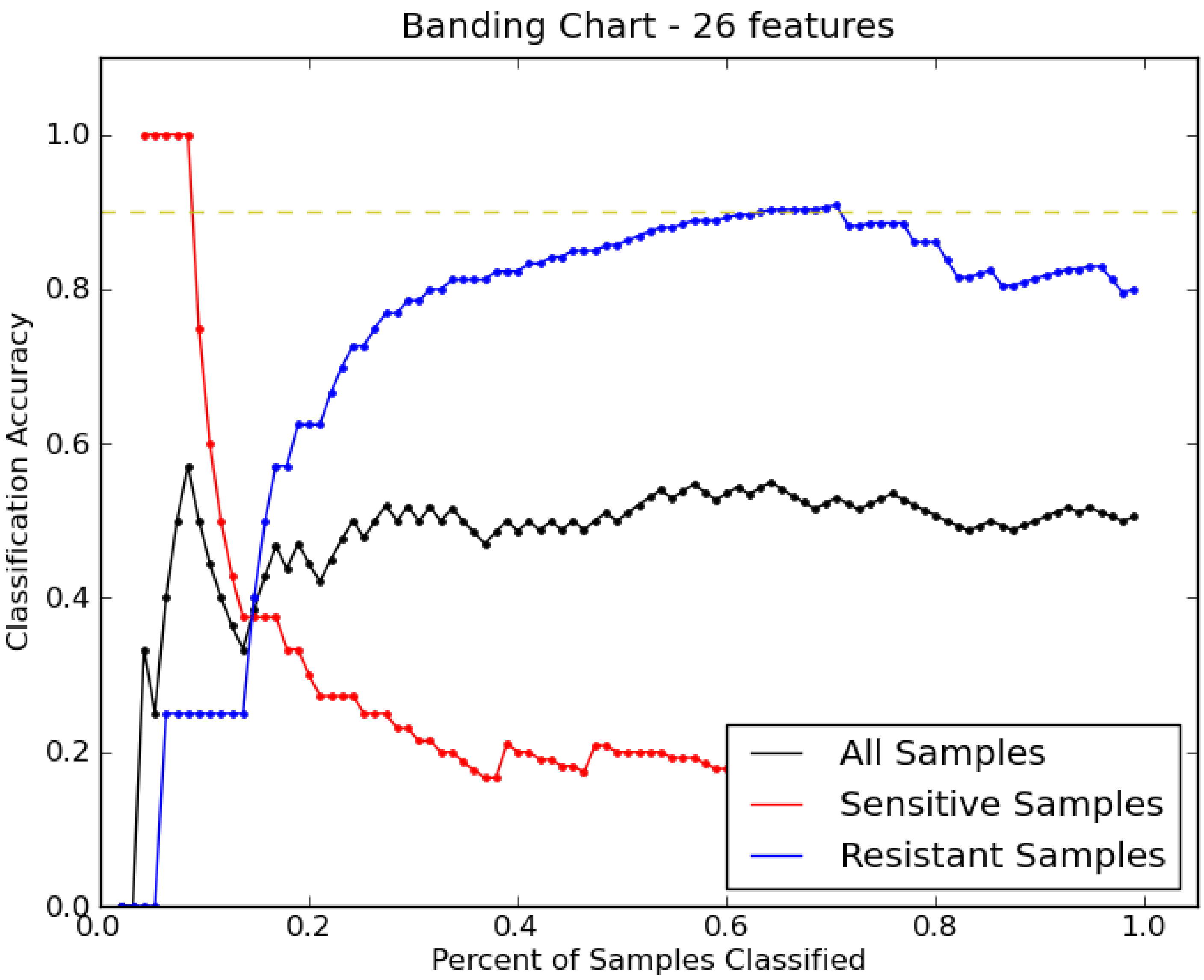
Using SVM-Light to build a simple abstaining classifier



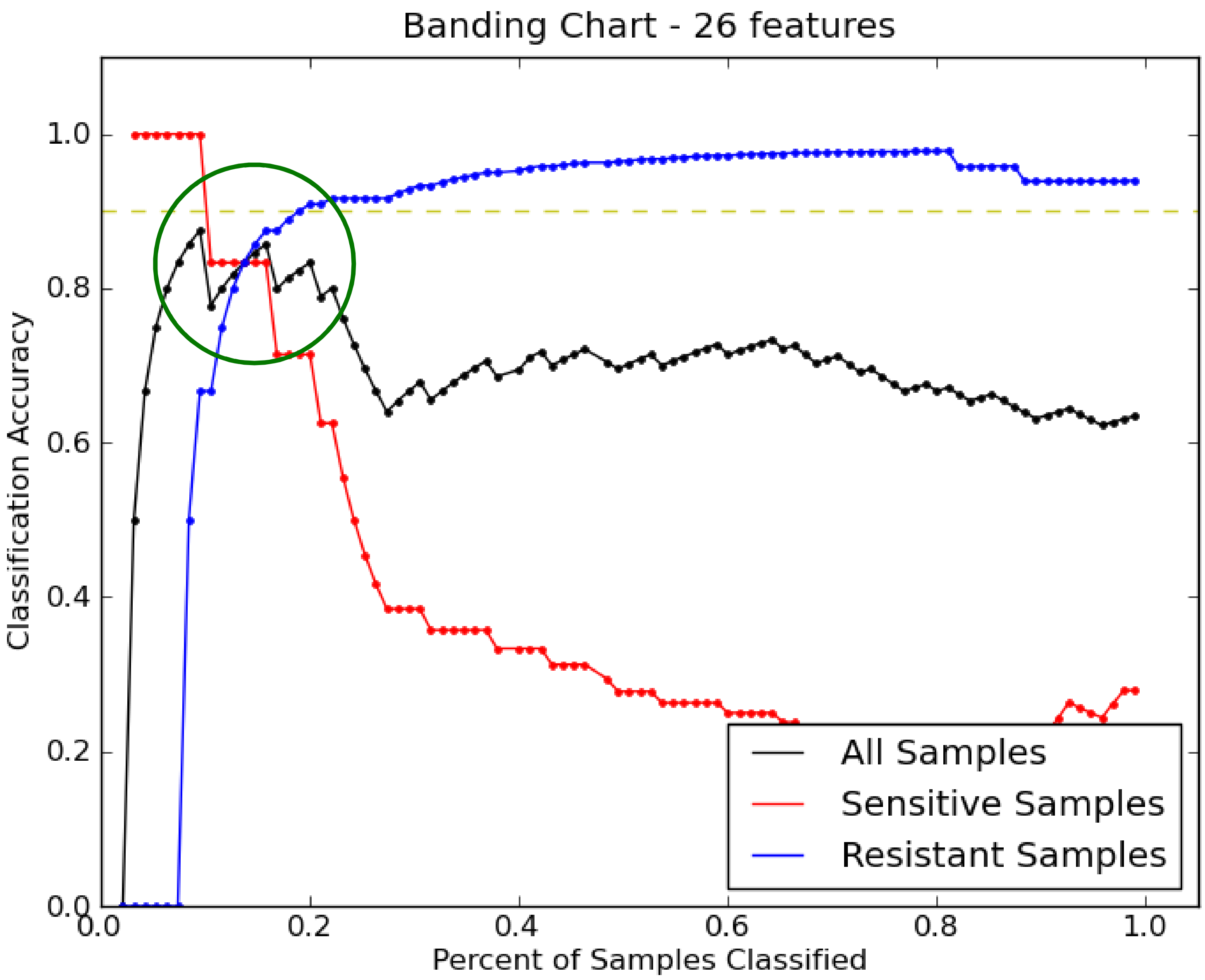
Classifying Platinum Sensitivity With Expression Scores



Classifying Platinum Sensitivity With Copy Number Scores



Classifying Platinum Sensitivity With Pathway Scores



Automated Pathway Scoring

- Better models agree with most of the data
- Allow integration of multiple kinds of data
- Probabilistic graphical models allow automatic detection of relevant pathways
- On some datasets, better clinical outcome predictors
- Limitation: can’t discover anything new

Outline

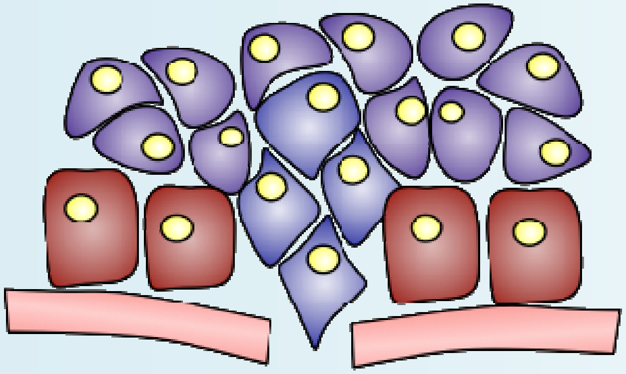
- High-throughput visualization w/ the UCSC Cancer Genomics Browser
- Identifying patient-specific pathway alterations
- Identifying novel pathway interactions w/ genetic knock-outs

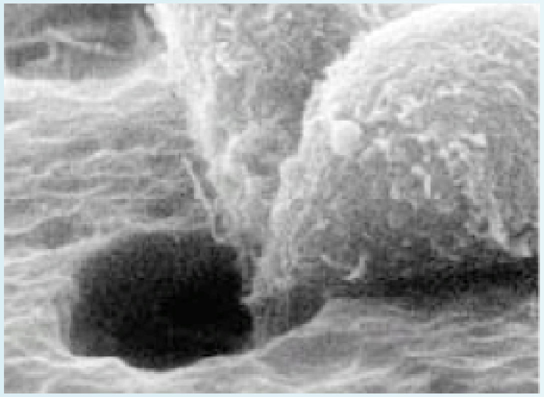
Beyond the "usual suspects": Novel Pathways

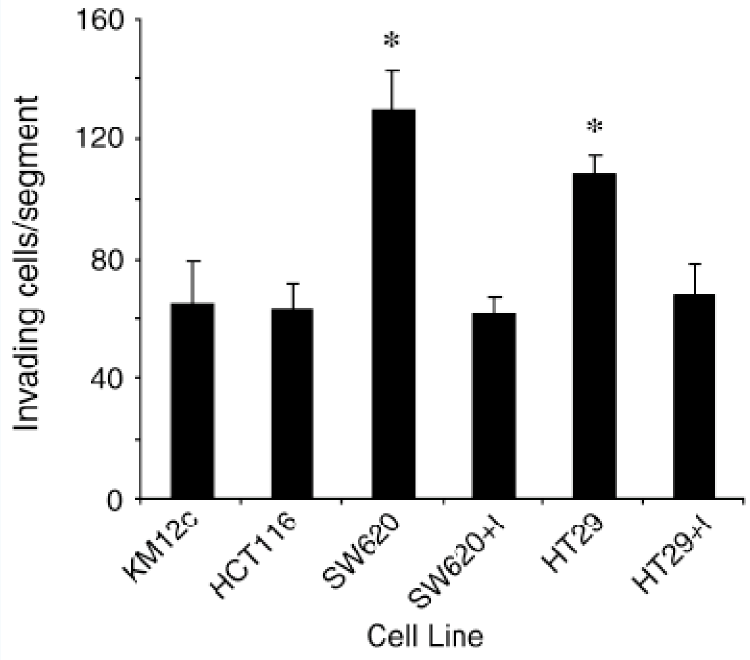
- Many pathways not fully characterized

1. Find interactions (structure) between known genes using fast, available experimental techniques
2. Predict new members of the pathway
3. Validate, then iterate

Application to Cancer Invasion



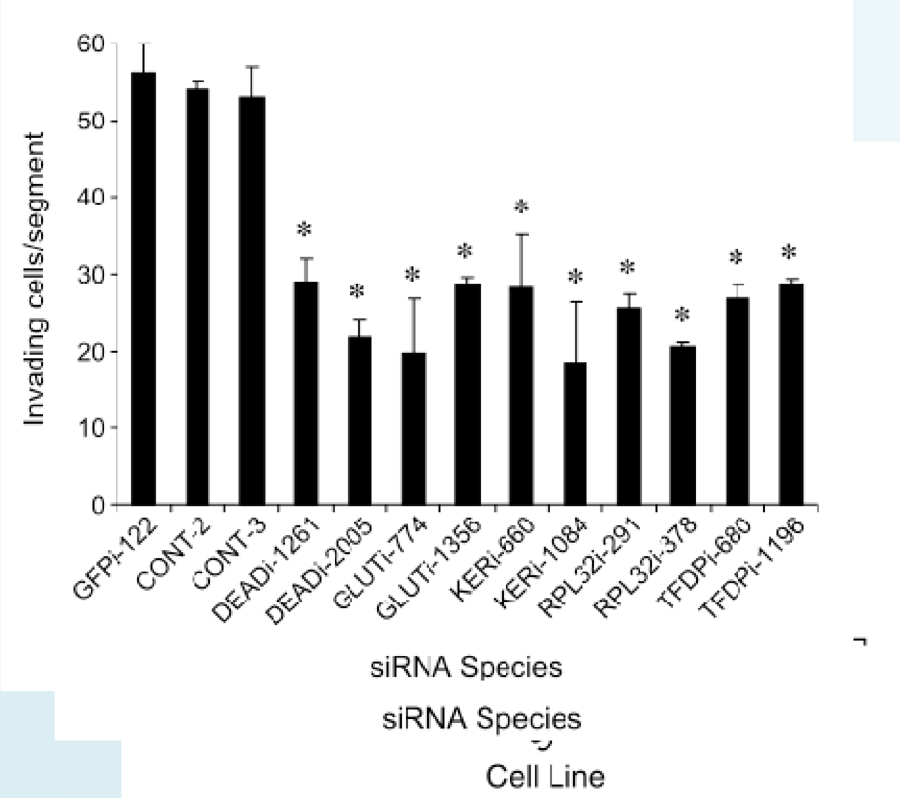


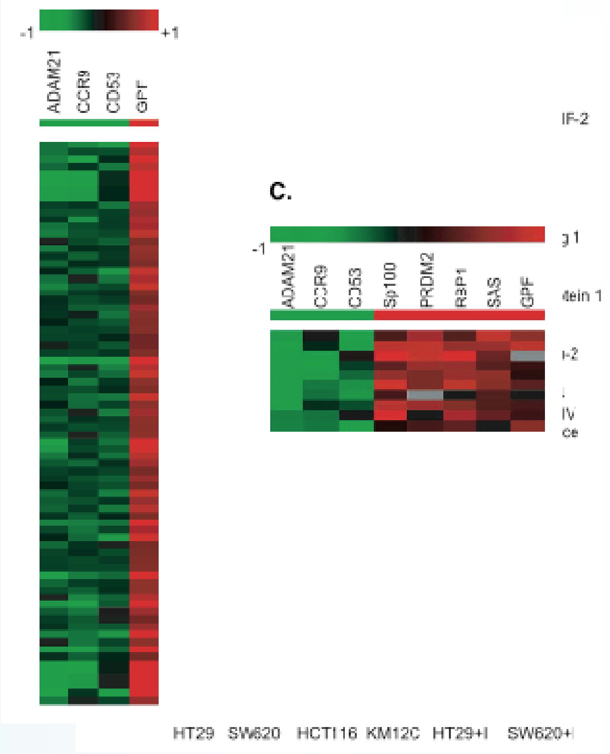


Phenotype assay
(Invasion in Matrigel)

Rosalyn Irby, Norman Lee, *et al., Cancer Res.*, 2005

Colon Cancer Invasion





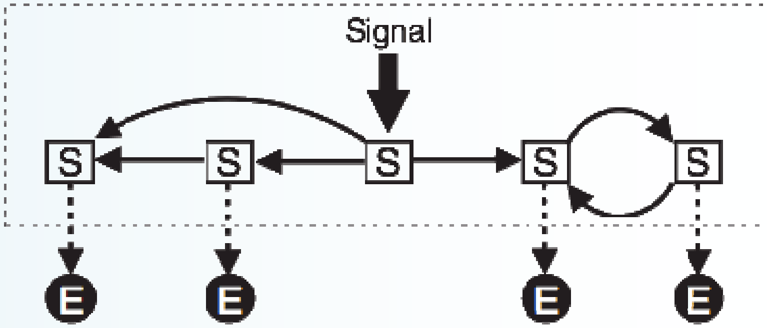
Phenotype assay
(Invasion in Matrigel)

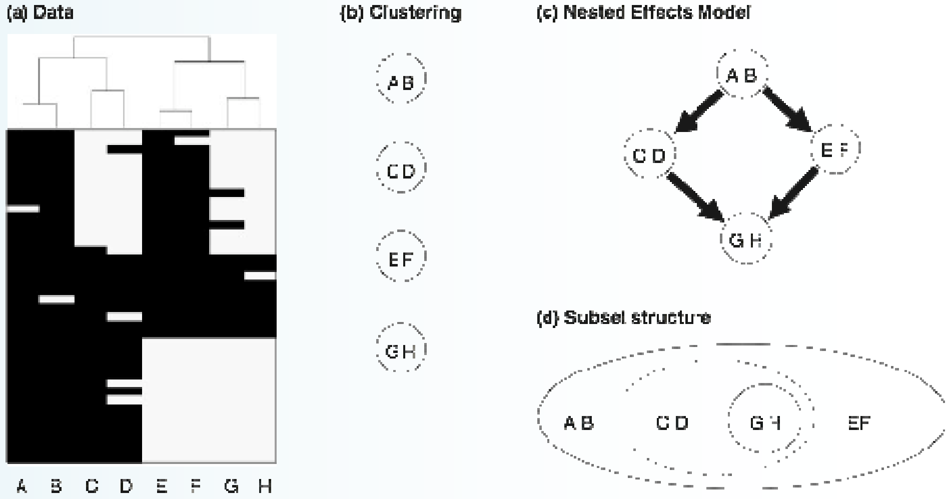
Expression Profile
Find gene candidates

Rosalyn Irby, Norman Lee, *et al.*, *Cancer Res.*, 2005

Nested Effects Inference Method

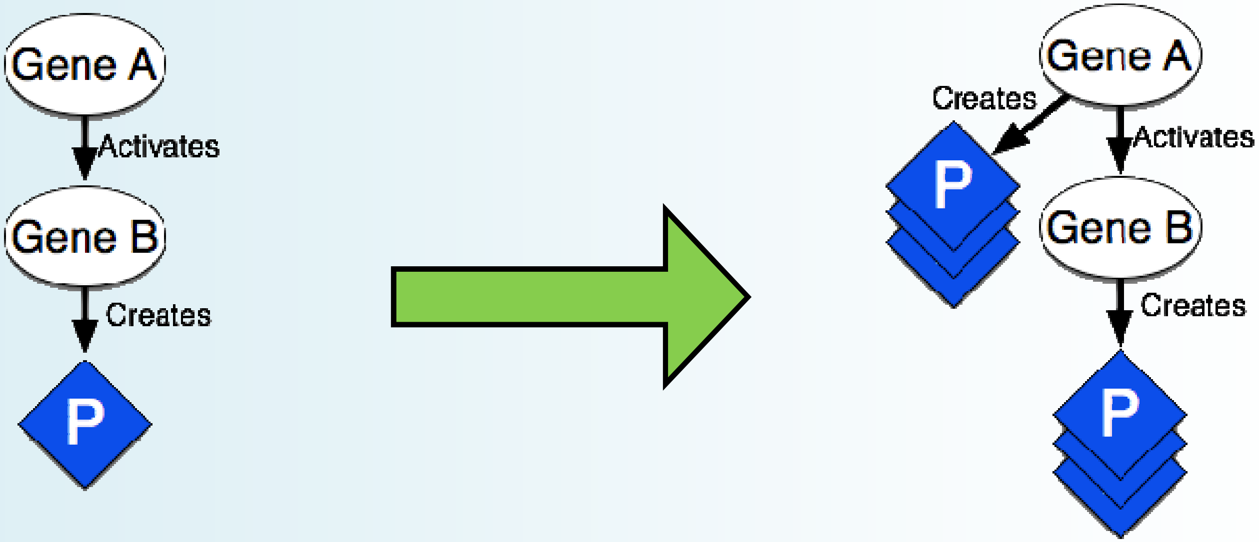
- Nested Effects Model
  - Indirect observations
  - Transcriptional & NON-transcriptional regulation
- Challenges
  - Throwing out sign of data
  - Computationally difficult hypothesis space





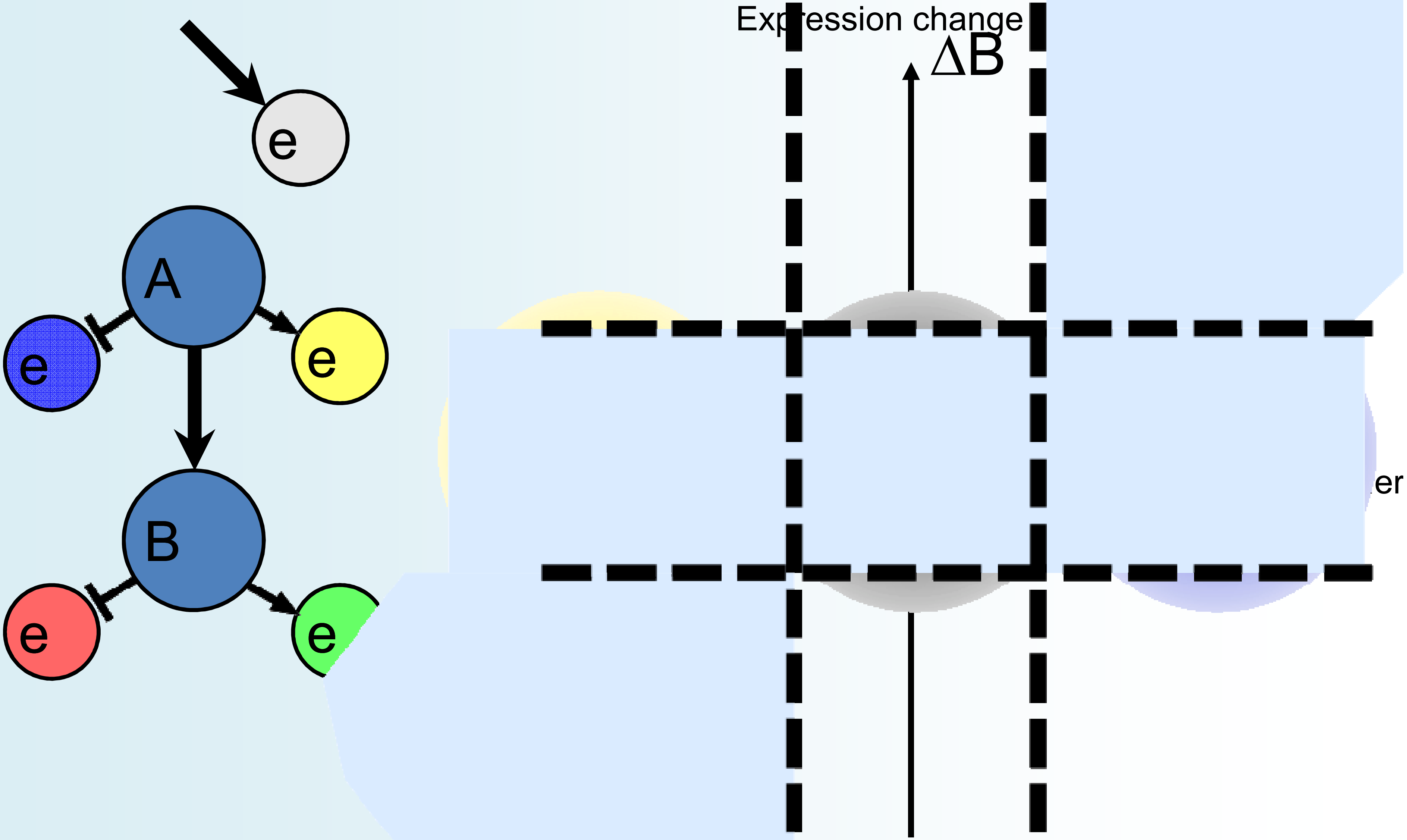
Markowetz *et al.*, 2005 & 200

Microarrays Measure Many Phenotypes

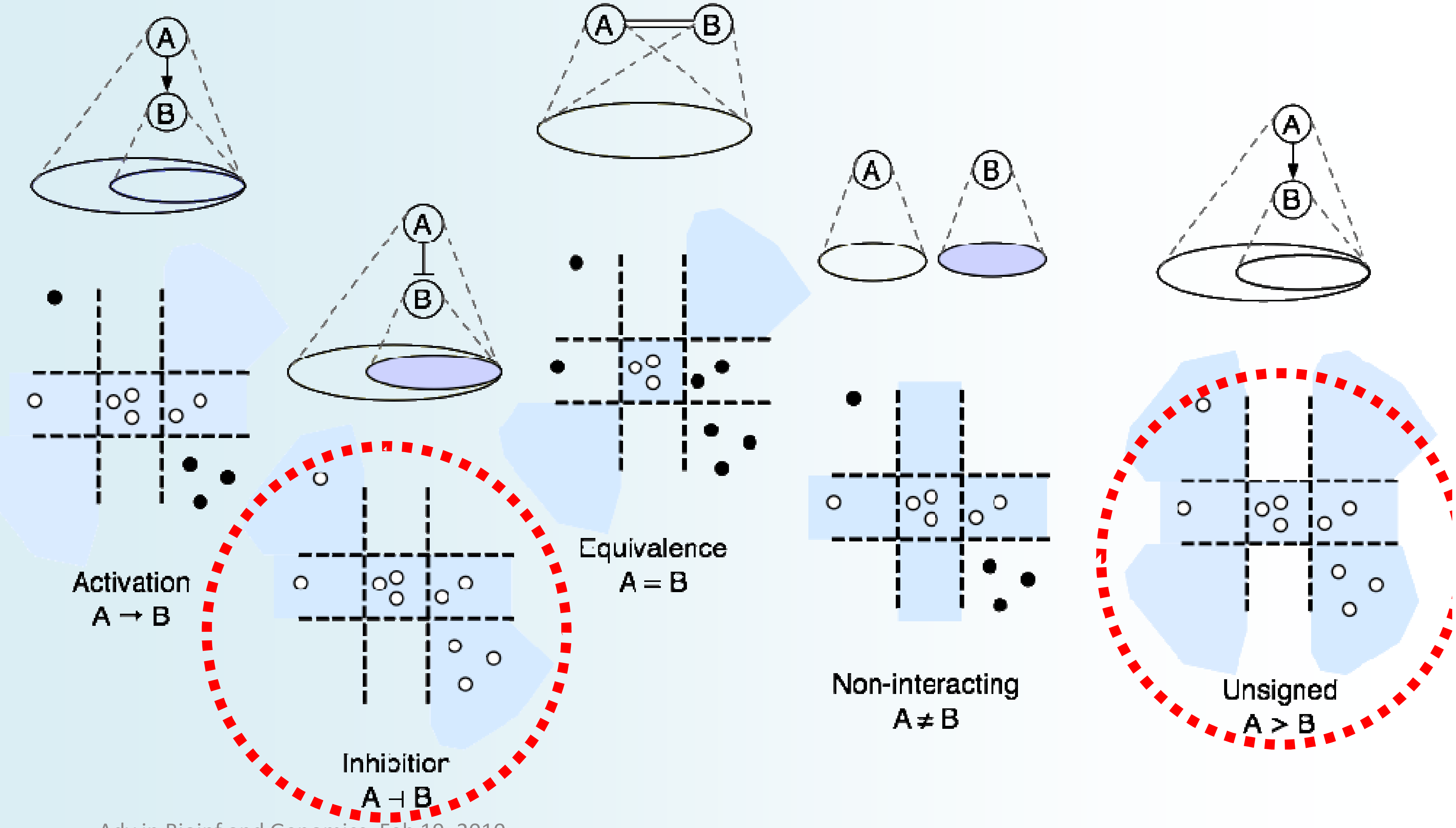


Switch
Regulatory
Epistasis

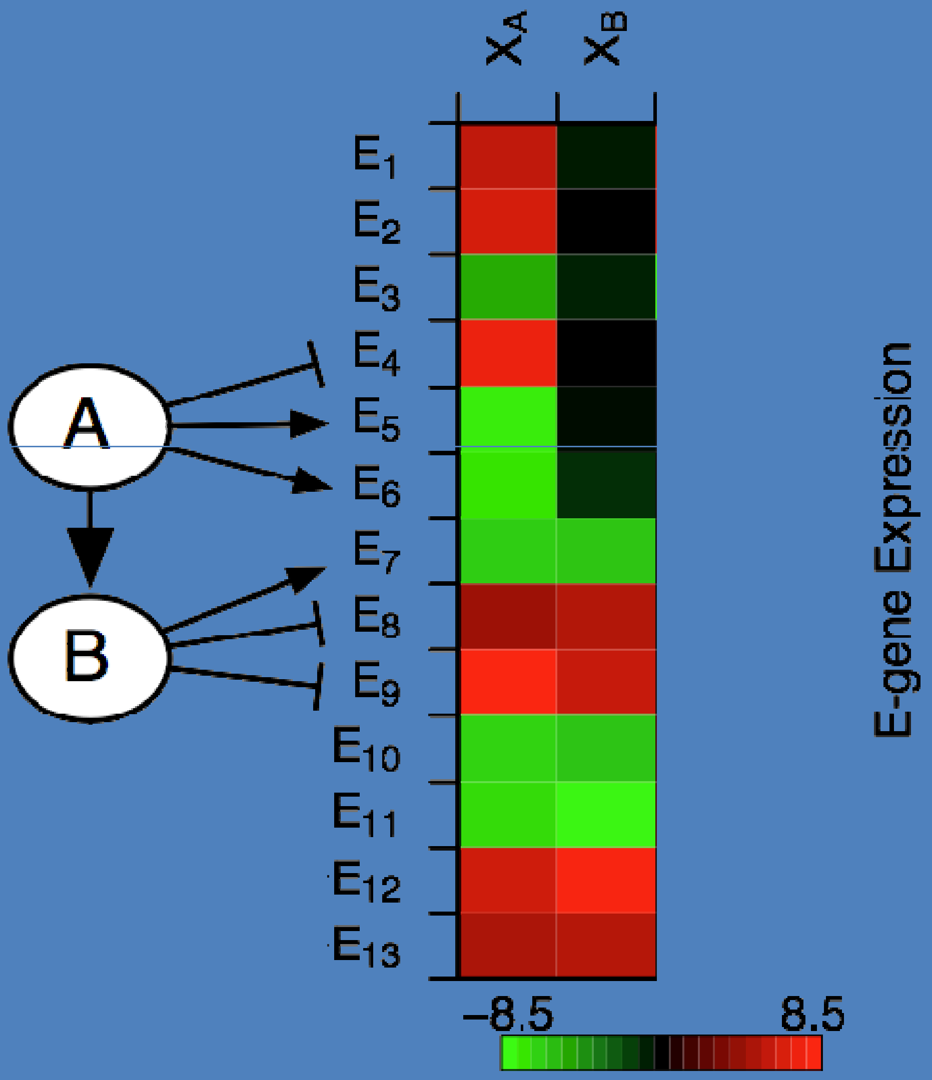
Nested
Effects
Model


e
A
e
e
B
e
e
Expression change
ΔB
er

Start from pairs

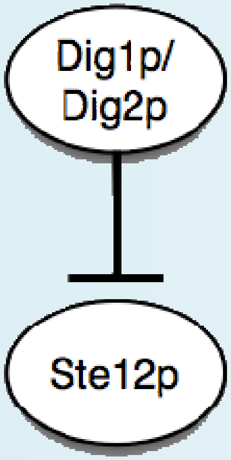


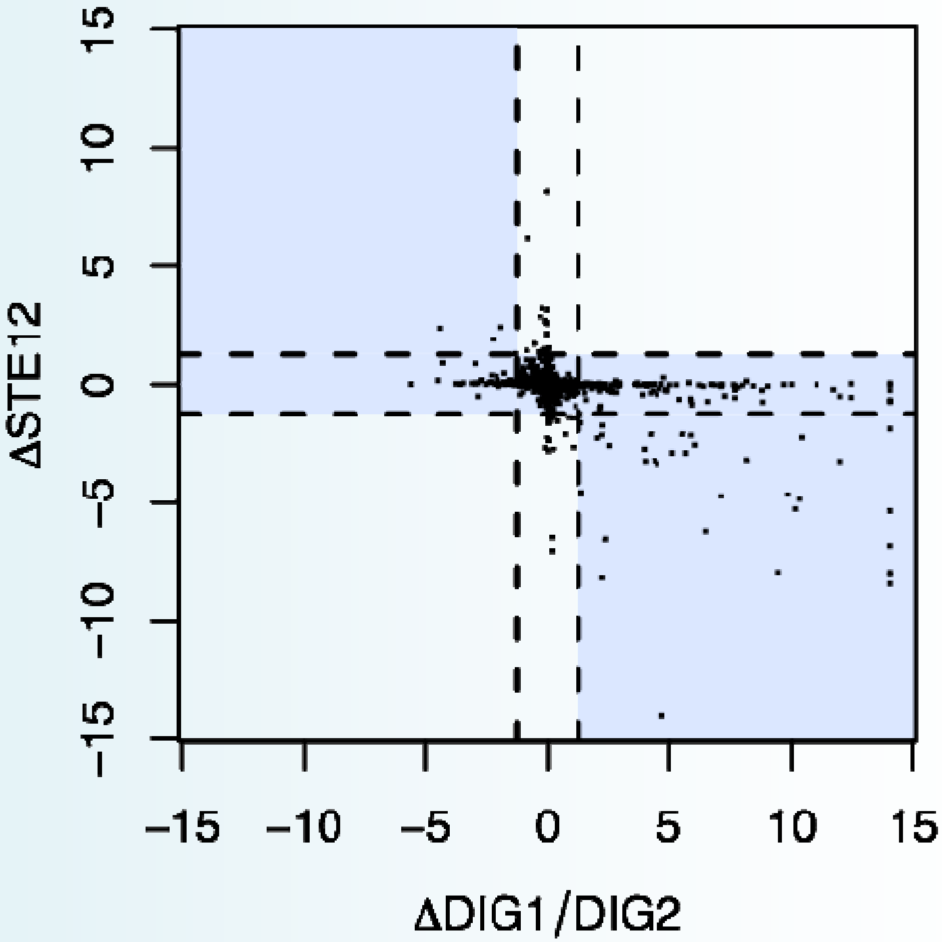
Toy Example



X_A
X_B
A
B
E_1
E_2
E_3
E_4
E_5
E_6
E_7
E_8
E_9
E_{10}
E_{11}
E_{12}
E_{13}
–8.5
8.5
E-gene Expression

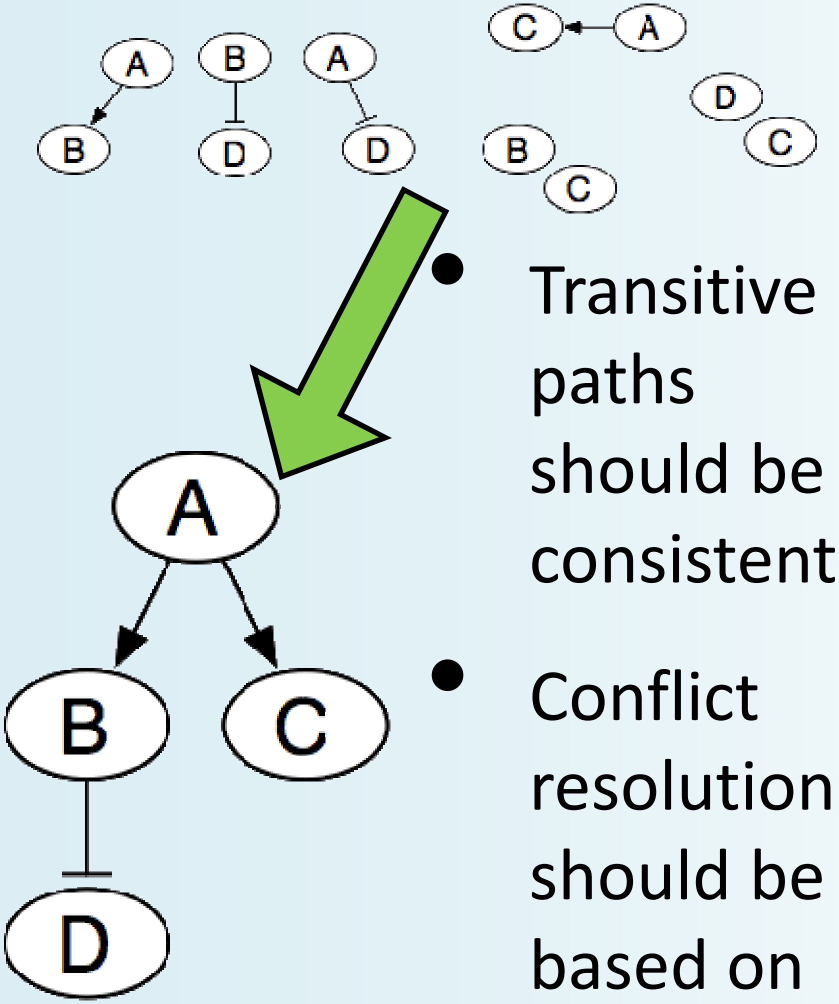
Observations from Yeast



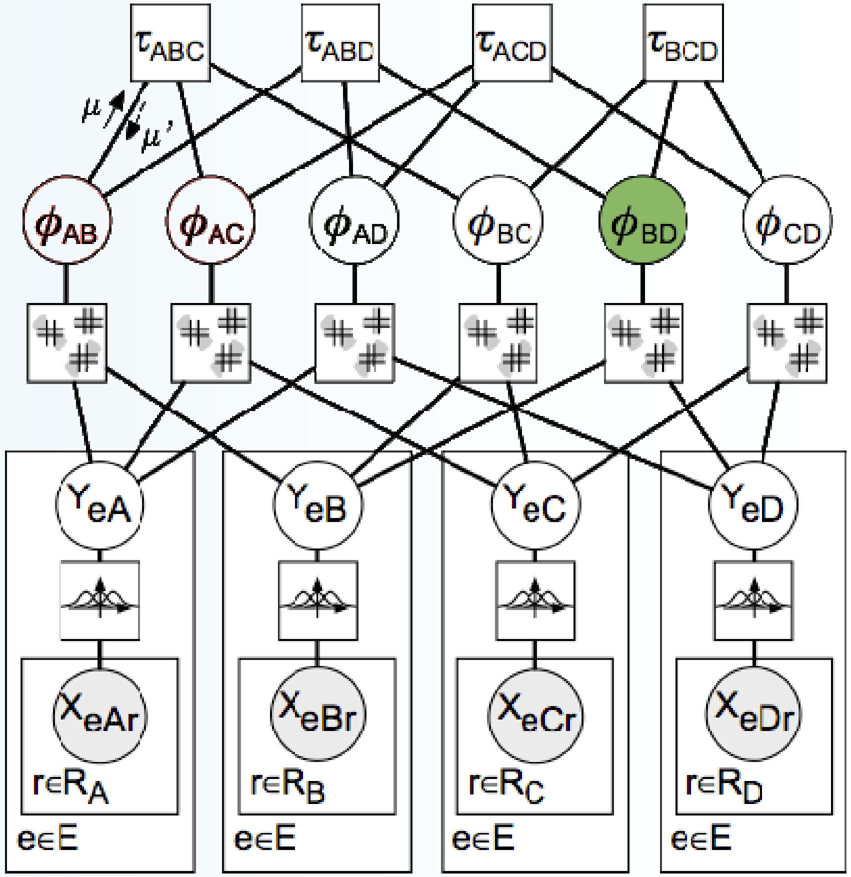


Data from Hughes *et al.*, 20

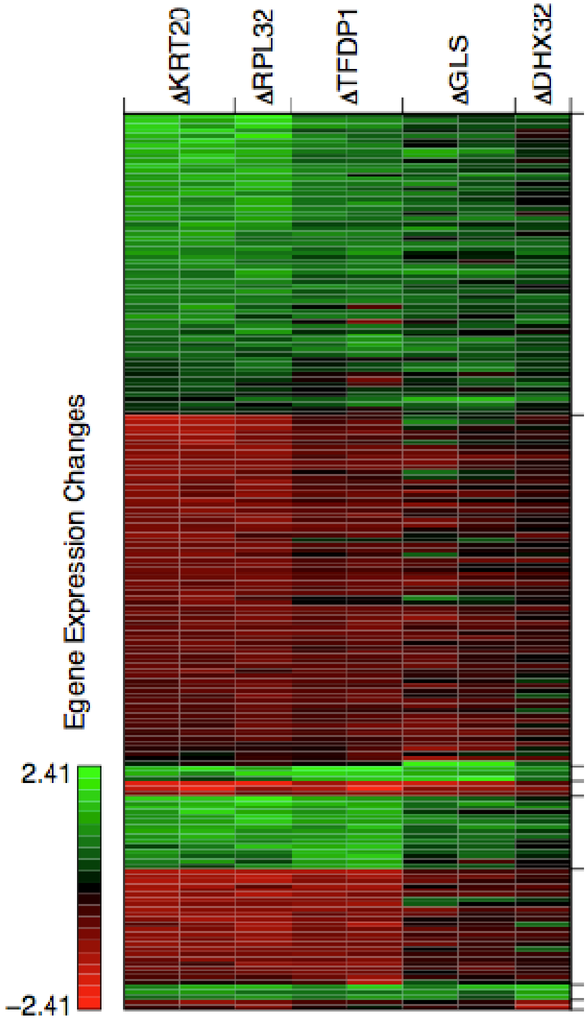
Scoring All Possible Nets with One Dual Net



- Transitive paths should be consistent
- Conflict resolution should be based on evidence

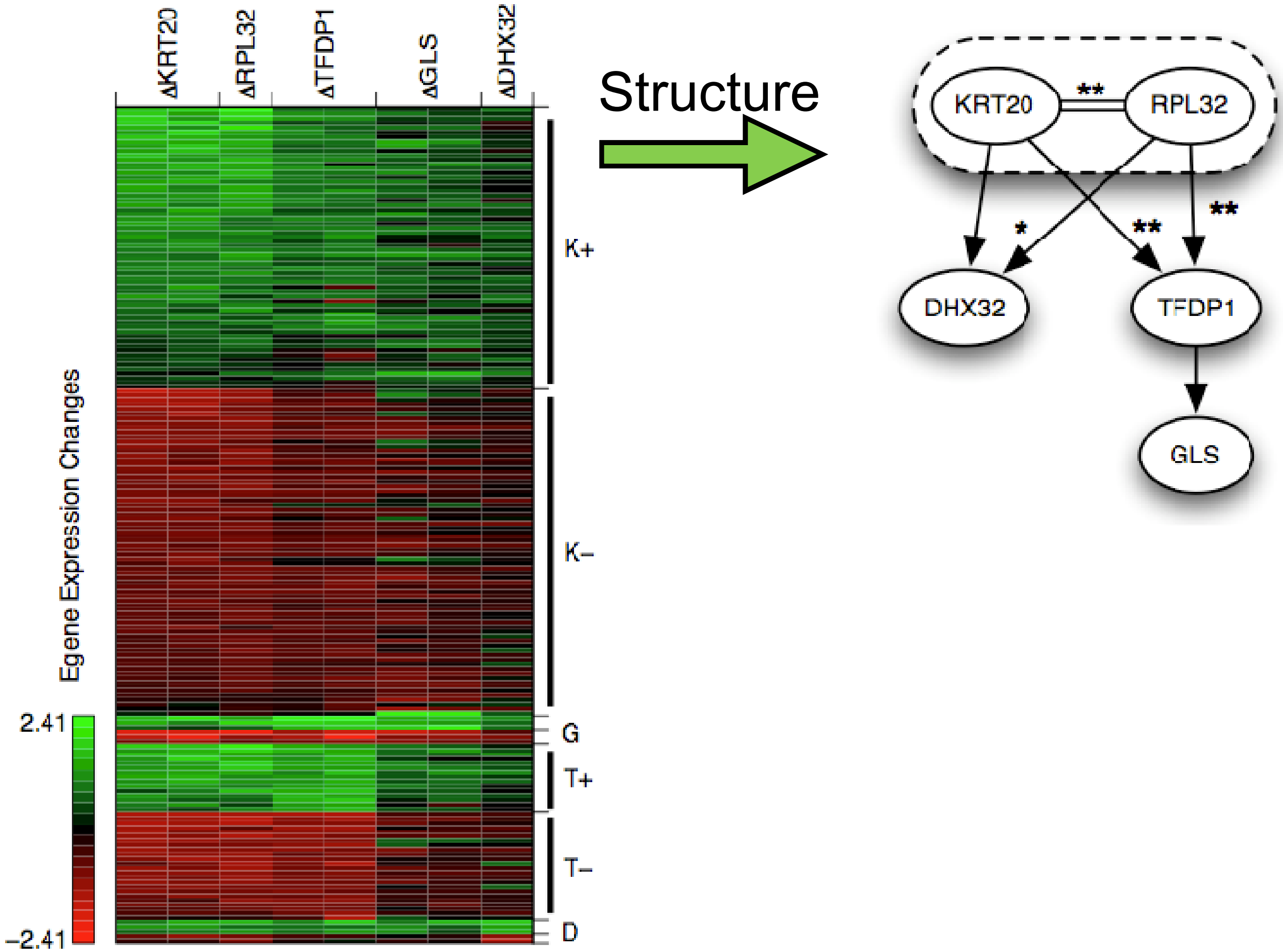


Colon Cancer Invasion



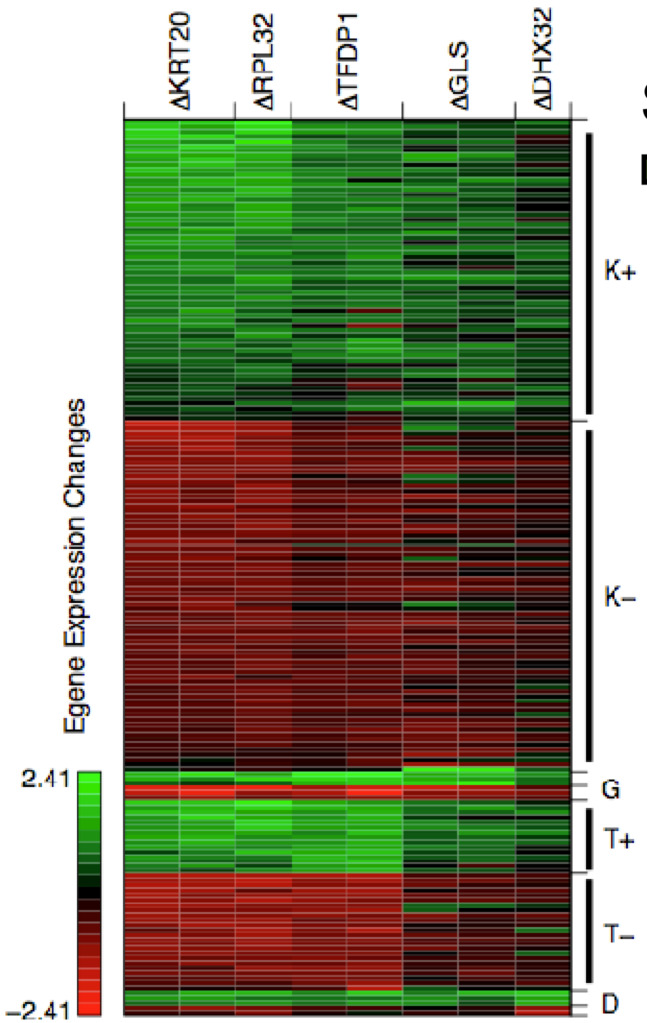
Vaske, *et al.*, 2009, *PloS Comp. Bio*

Colon Cancer Invasion

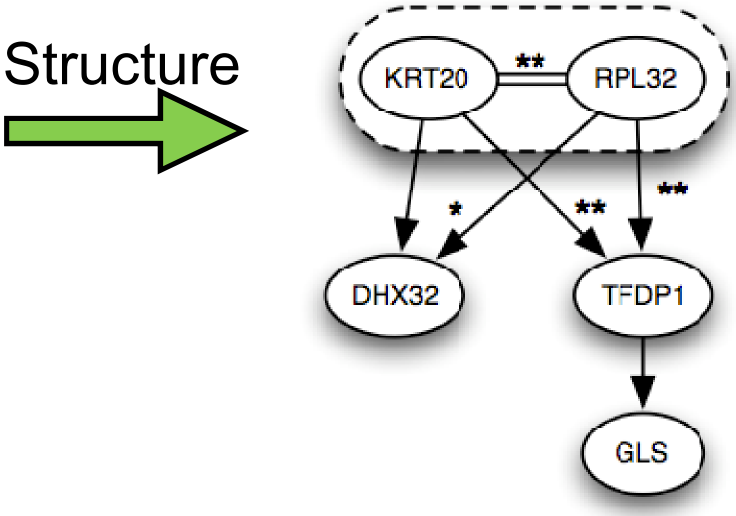


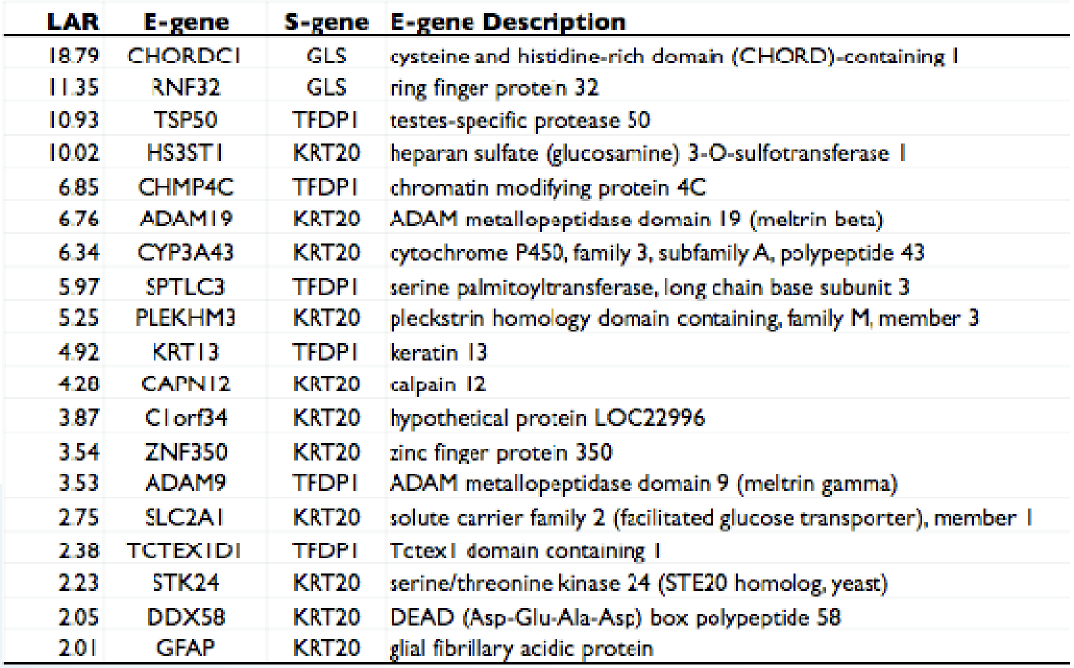
Vaske, *et al.*, 2009, *PloS Comp. Bio*

Colon Cancer Invasion



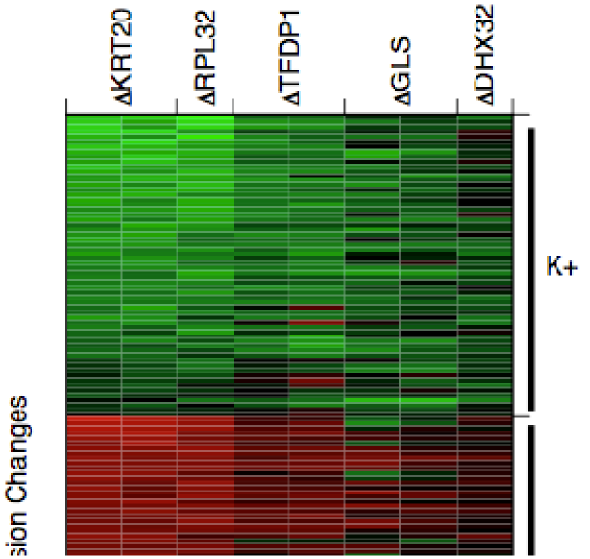
Vaske, *et al.*, 2009, *PloS Comp. Bio*

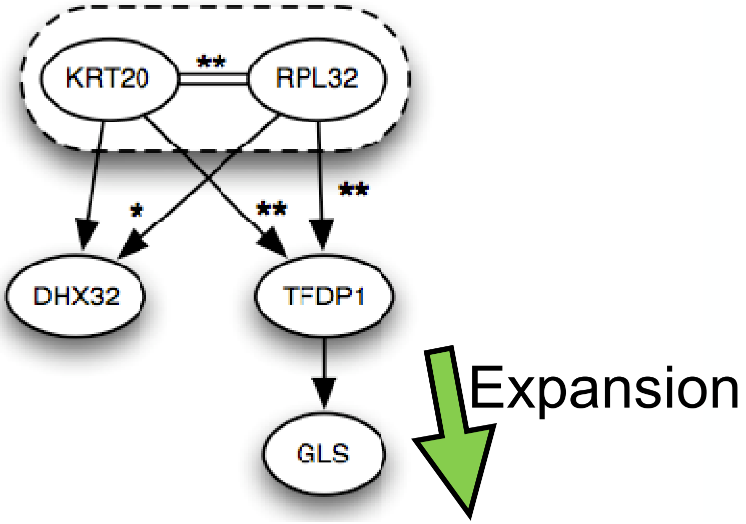


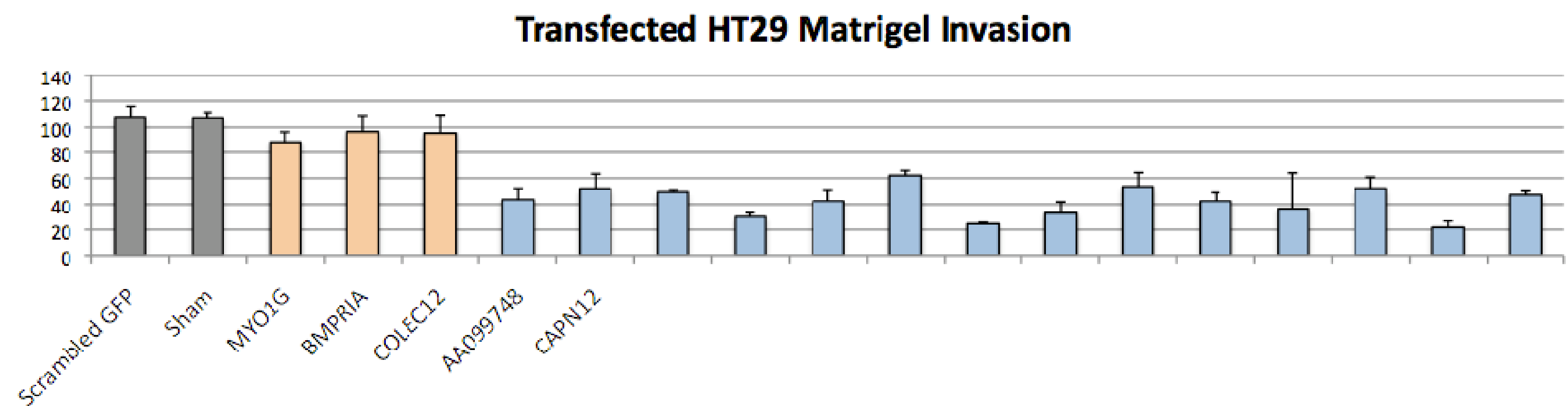


LAR	E-gene	S-gene	E-gene Description
18.79	CHORDC1	GLS	cysteine and histidine-rich domain (CHORD)-containing 1
11.35	RNF32	GLS	ring finger protein 32
10.93	TSP50	TFDP1	testes-specific protease 50
10.02	HS3ST1	KRT20	heparan sulfate (glucosamine) 3-O-sulfotransferase 1
6.85	CHMP4C	TFDP1	chromatin modifying protein 4C
6.76	ADAM19	KRT20	ADAM metallopeptidase domain 19 (meltrin beta)
6.34	CYP3A43	KRT20	cytochrome P450, family 3, subfamily A, polypeptide 43
5.97	SPTLC3	TFDP1	serine palmitoyltransferase, long chain base subunit 3
5.25	PLEKHM3	KRT20	pleckstrin homology domain containing, family M, member 3
4.92	KRT13	TFDP1	keratin 13
4.28	CAPN12	KRT20	calpain 12
3.87	C1orf34	KRT20	hypothetical protein LOC22996
3.54	ZNF350	KRT20	zinc finger protein 350
3.53	ADAM9	TFDP1	ADAM metallopeptidase domain 9 (meltrin gamma)
2.75	SLC2A1	KRT20	solute carrier family 2 (facilitated glucose transporter), member 1
2.38	TCTEX1D1	TFDP1	Tctex1 domain containing 1
2.23	STK24	KRT20	serine/threonine kinase 24 (STE20 homolog, yeast)
2.05	DDX58	KRT20	DEAD (Asp-Glu-Ala-Asp) box polypeptide 58
2.01	GFAP	KRT20	glial fibrillary acidic protein

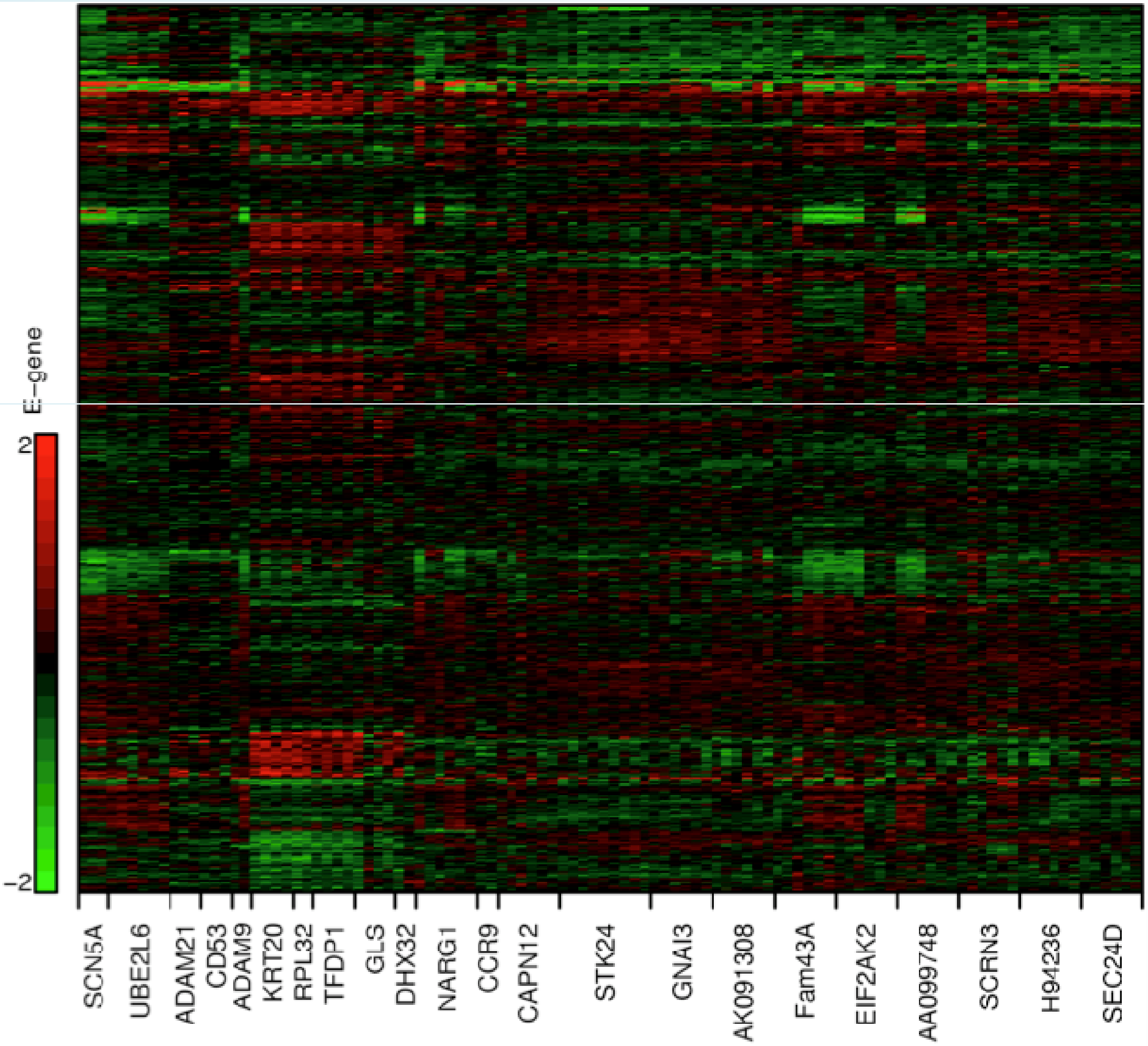
Colon Cancer Invasion



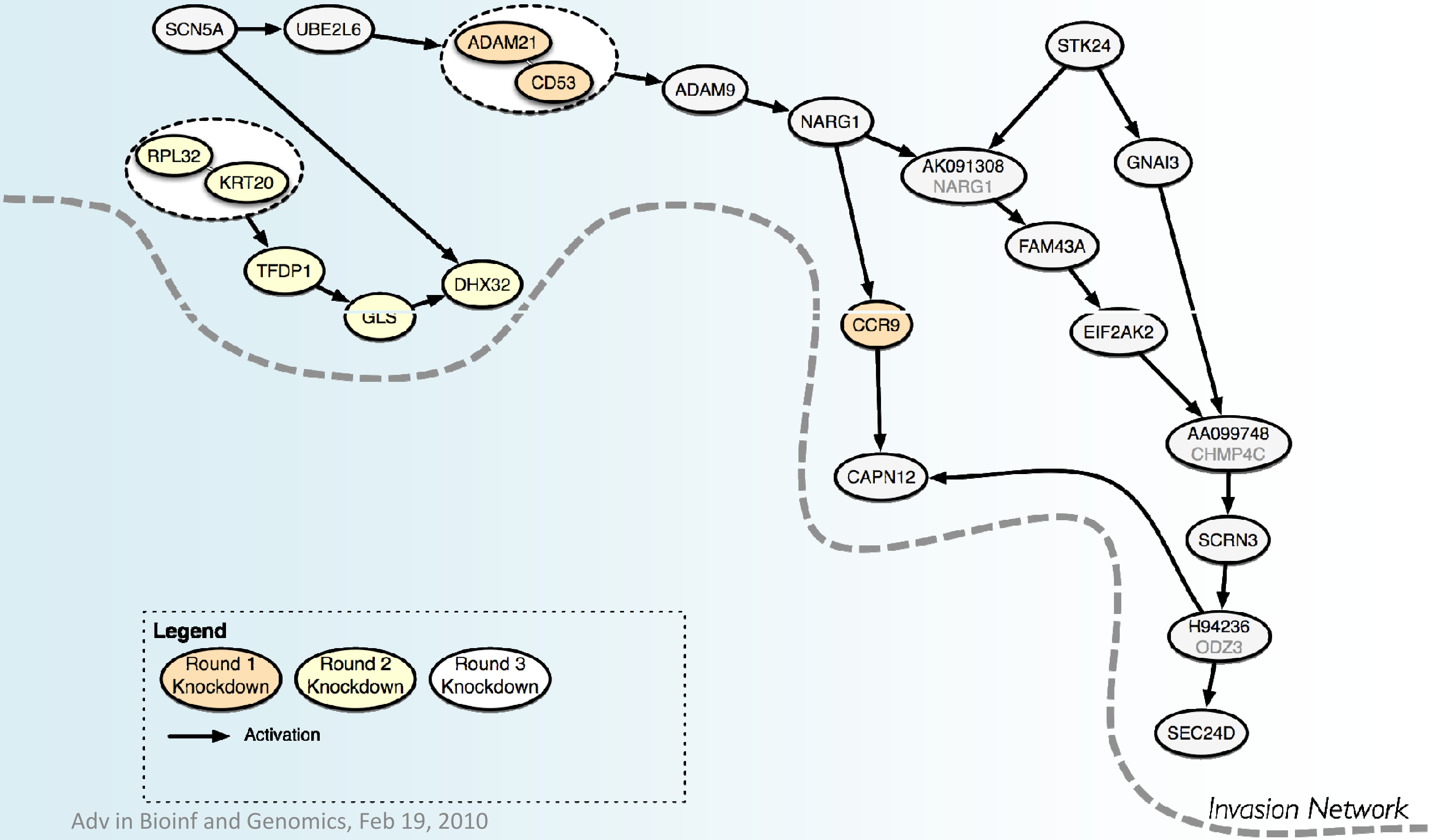




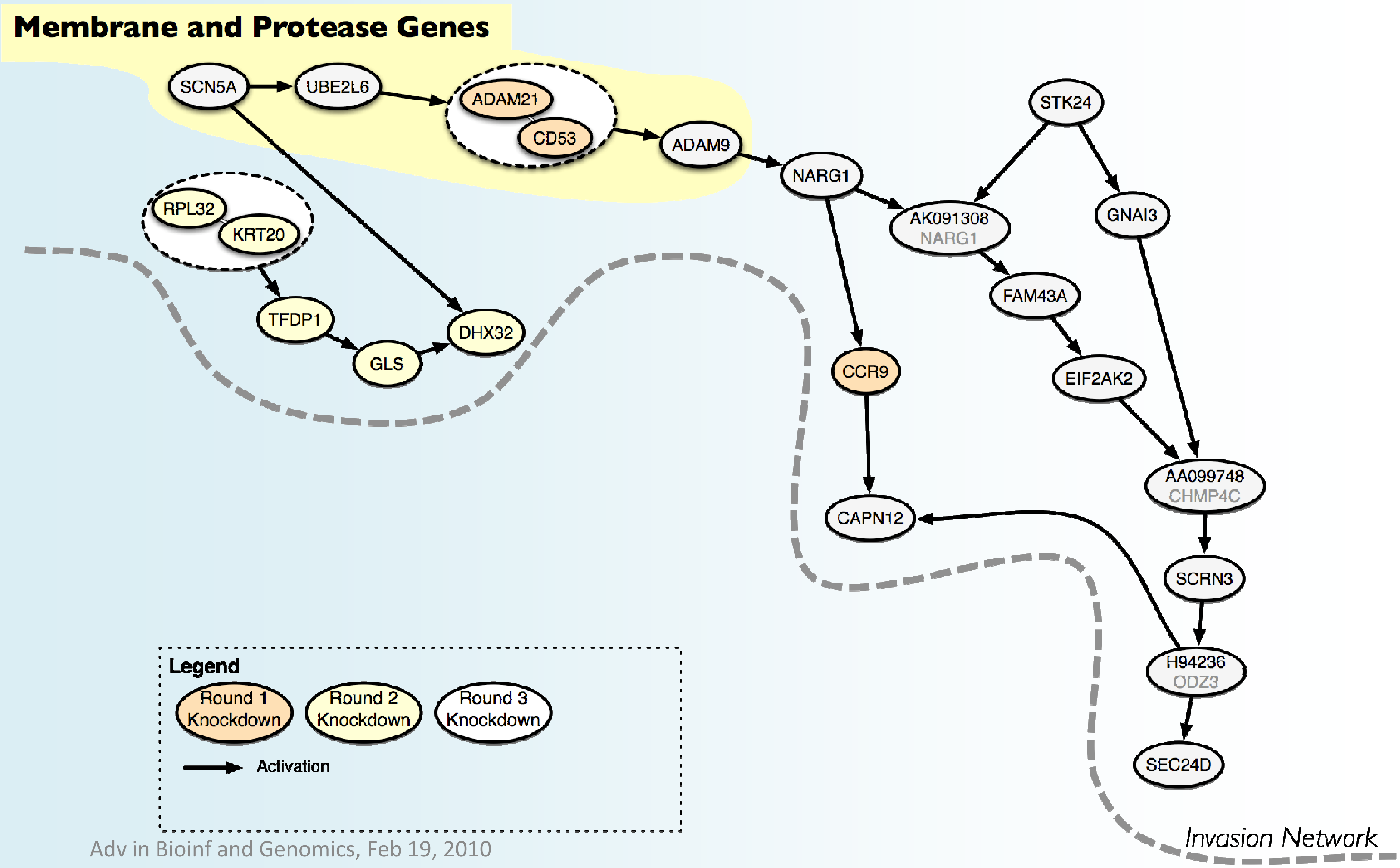
Iterative Pathway Expansion



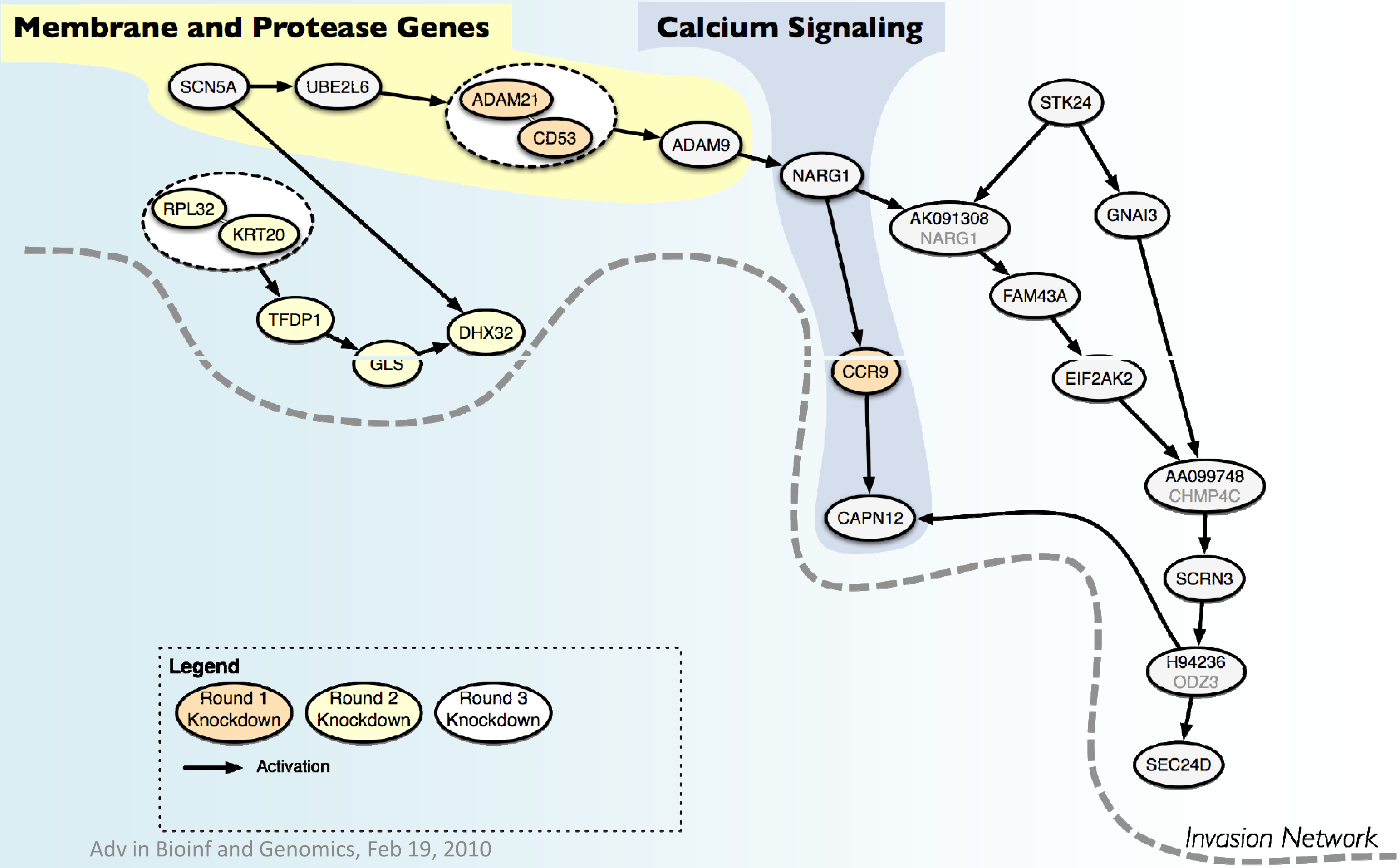
Invasion Network



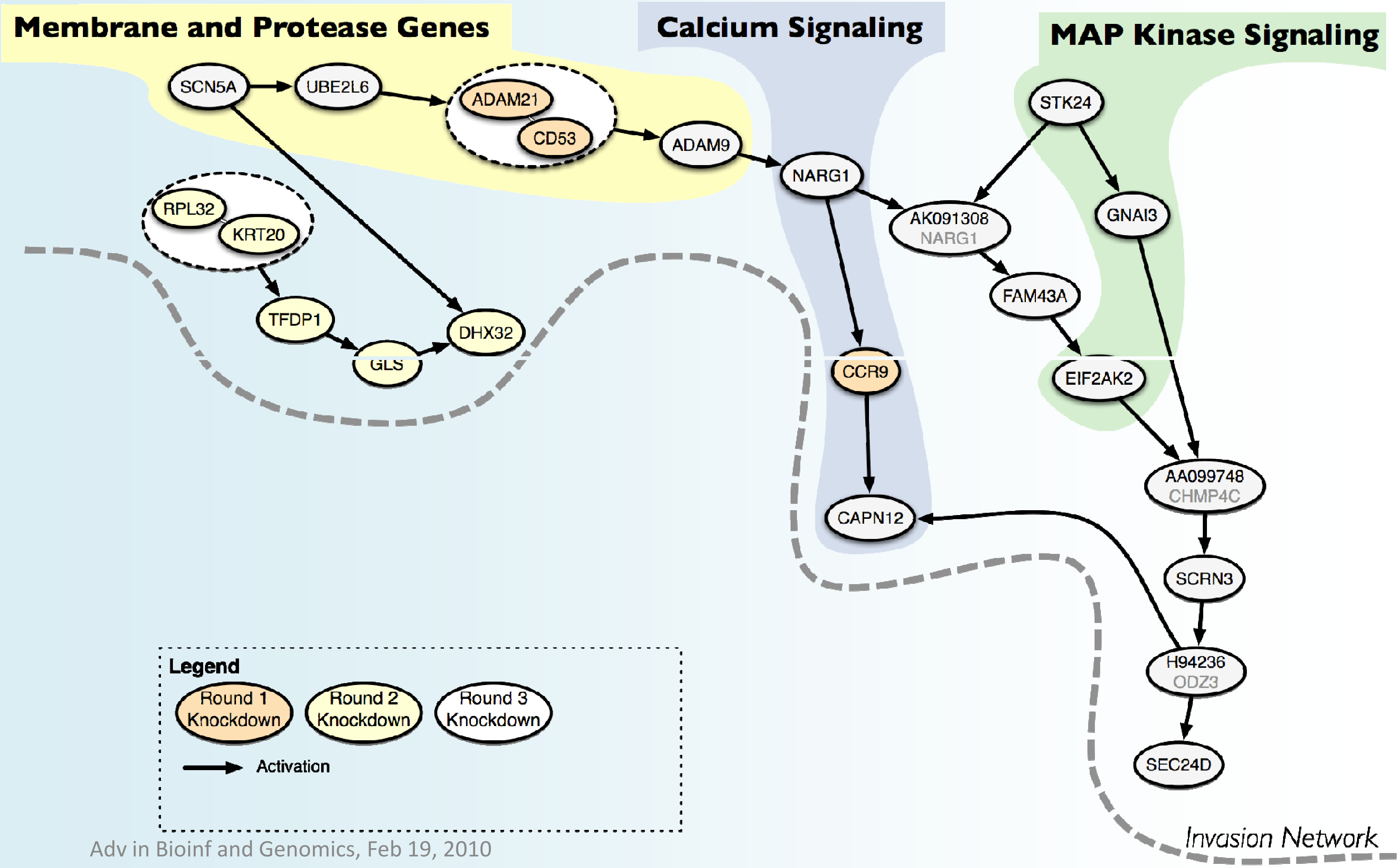
Invasion Network



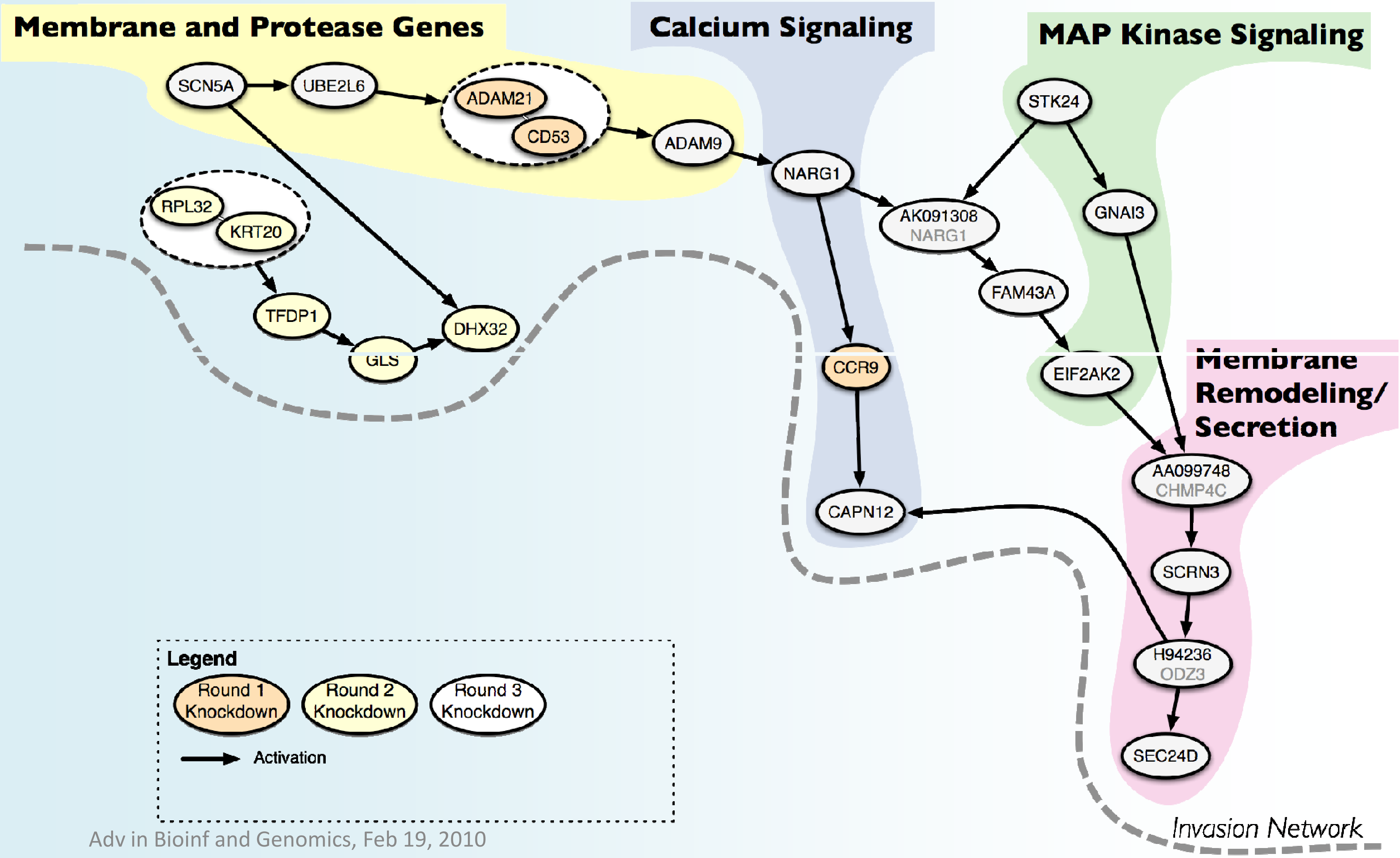
Invasion Network



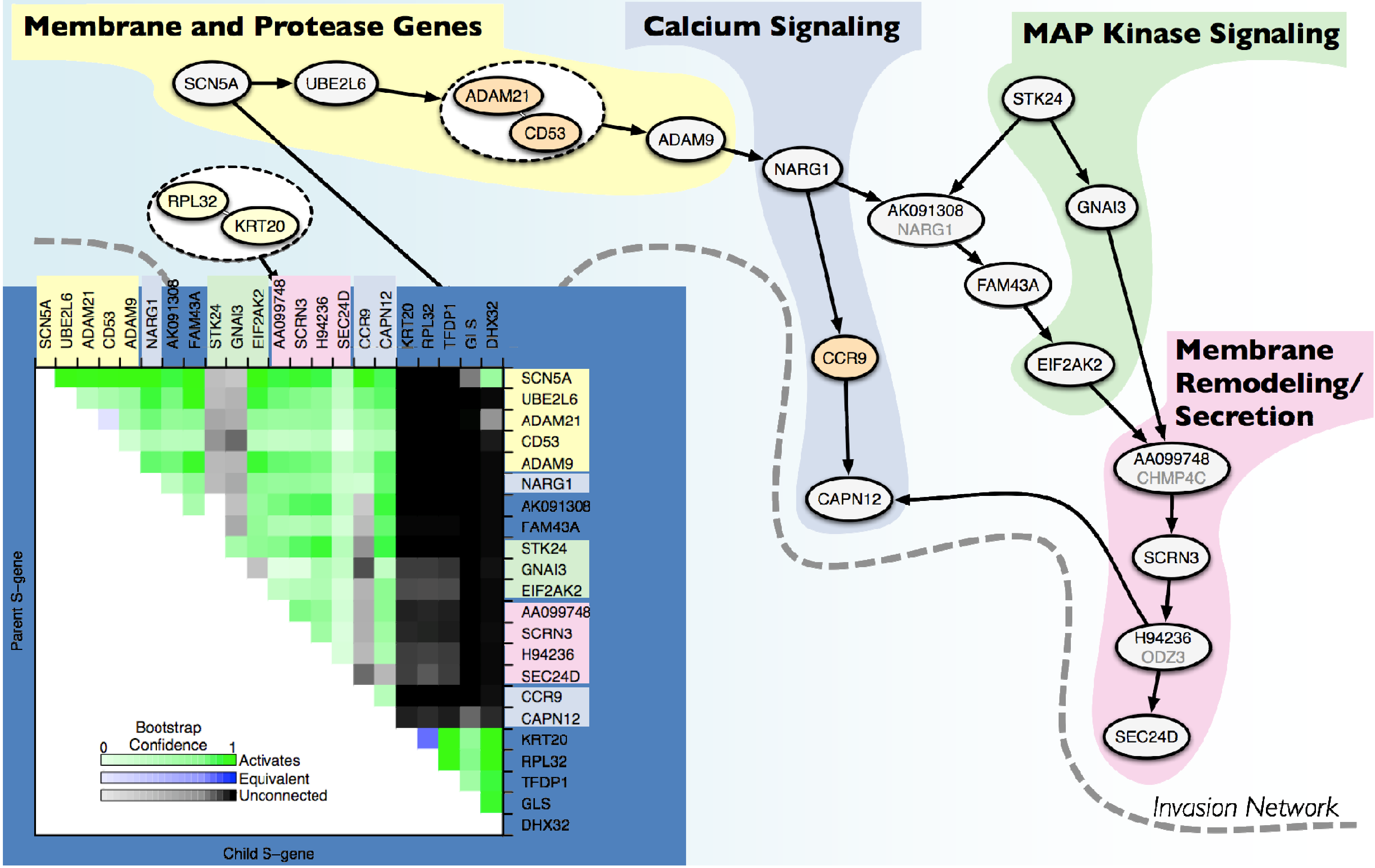
Invasion Network



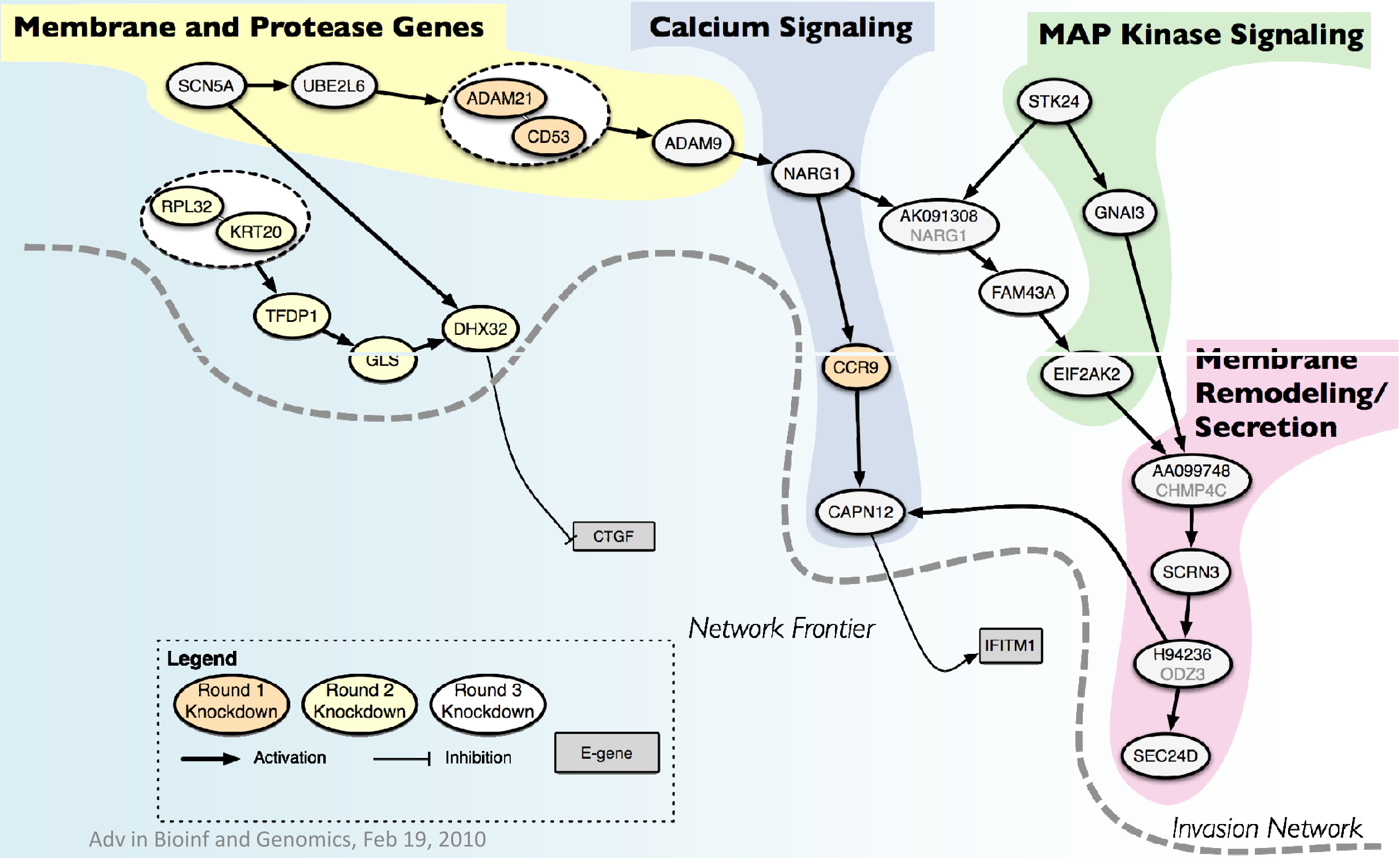
Invasion Network



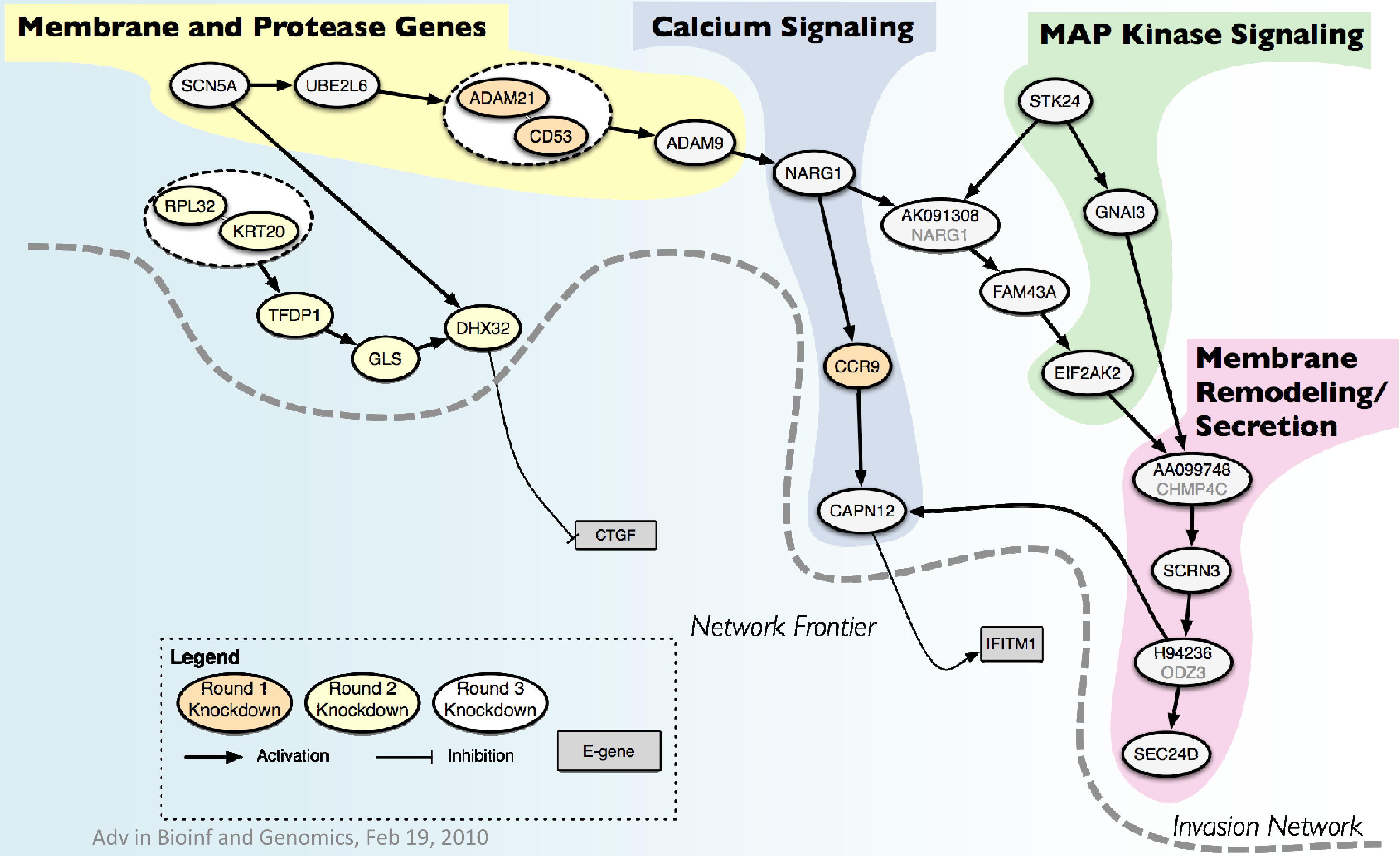
Invasion Network



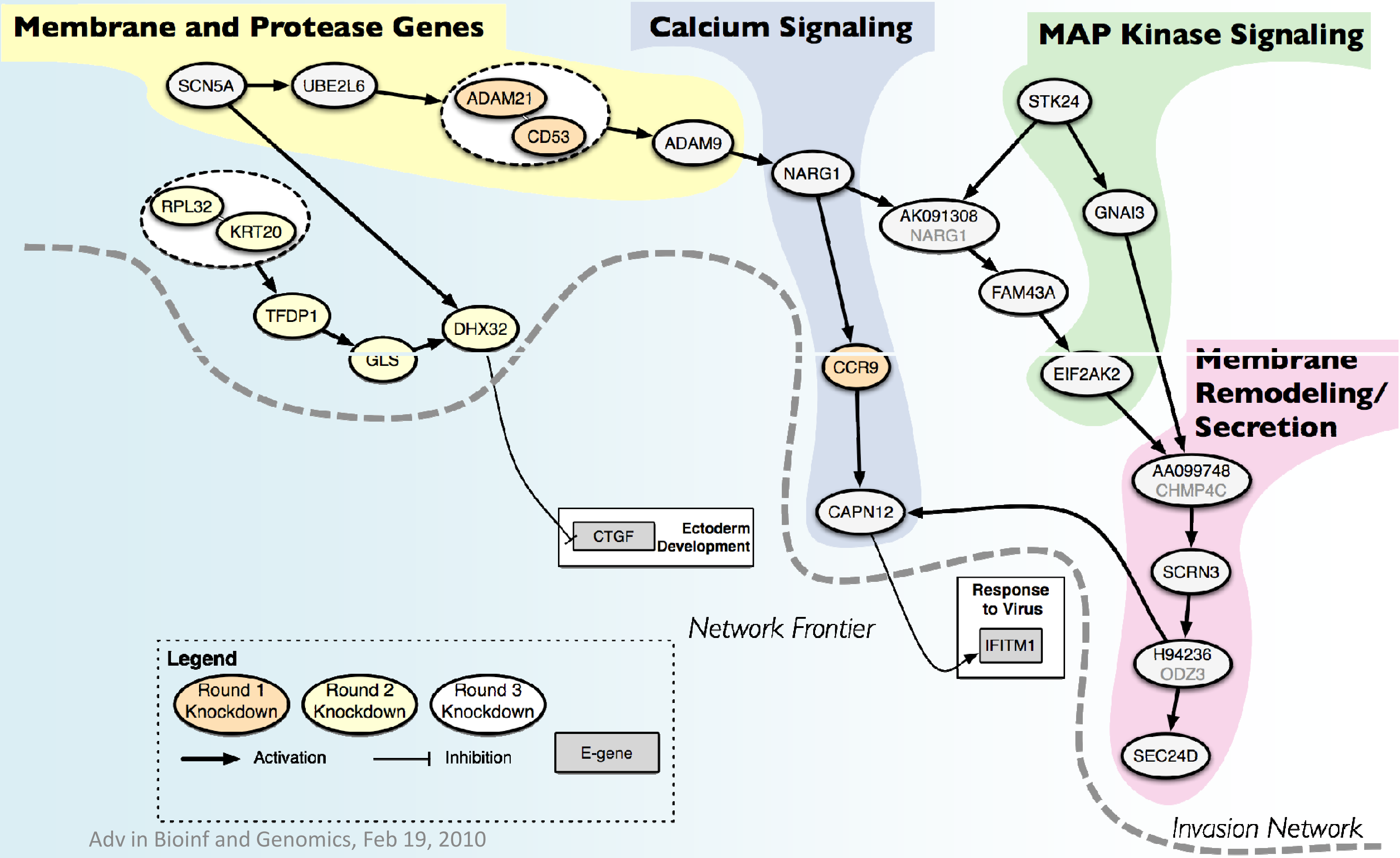
Invasion Network



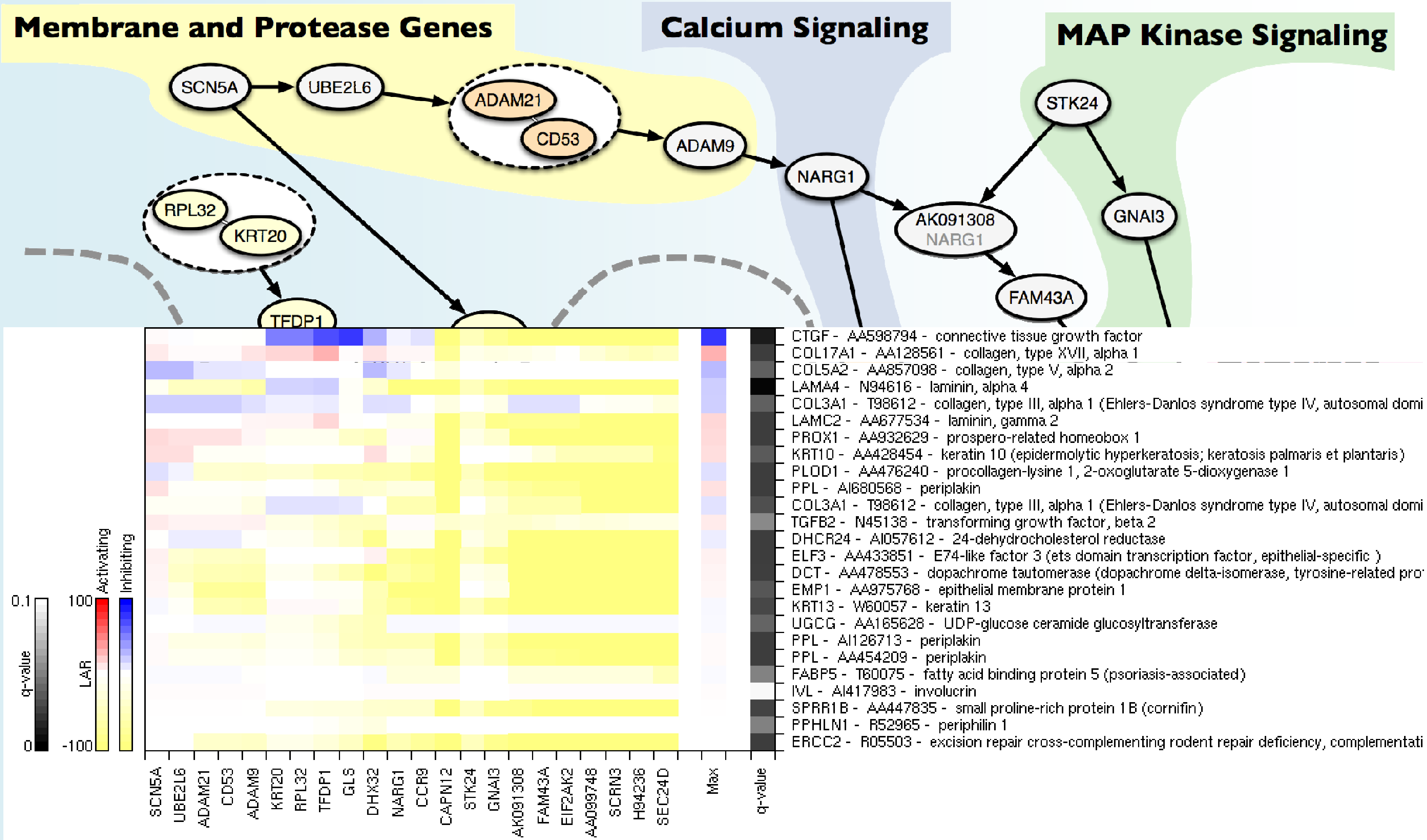
Invasion Network



Invasion Network

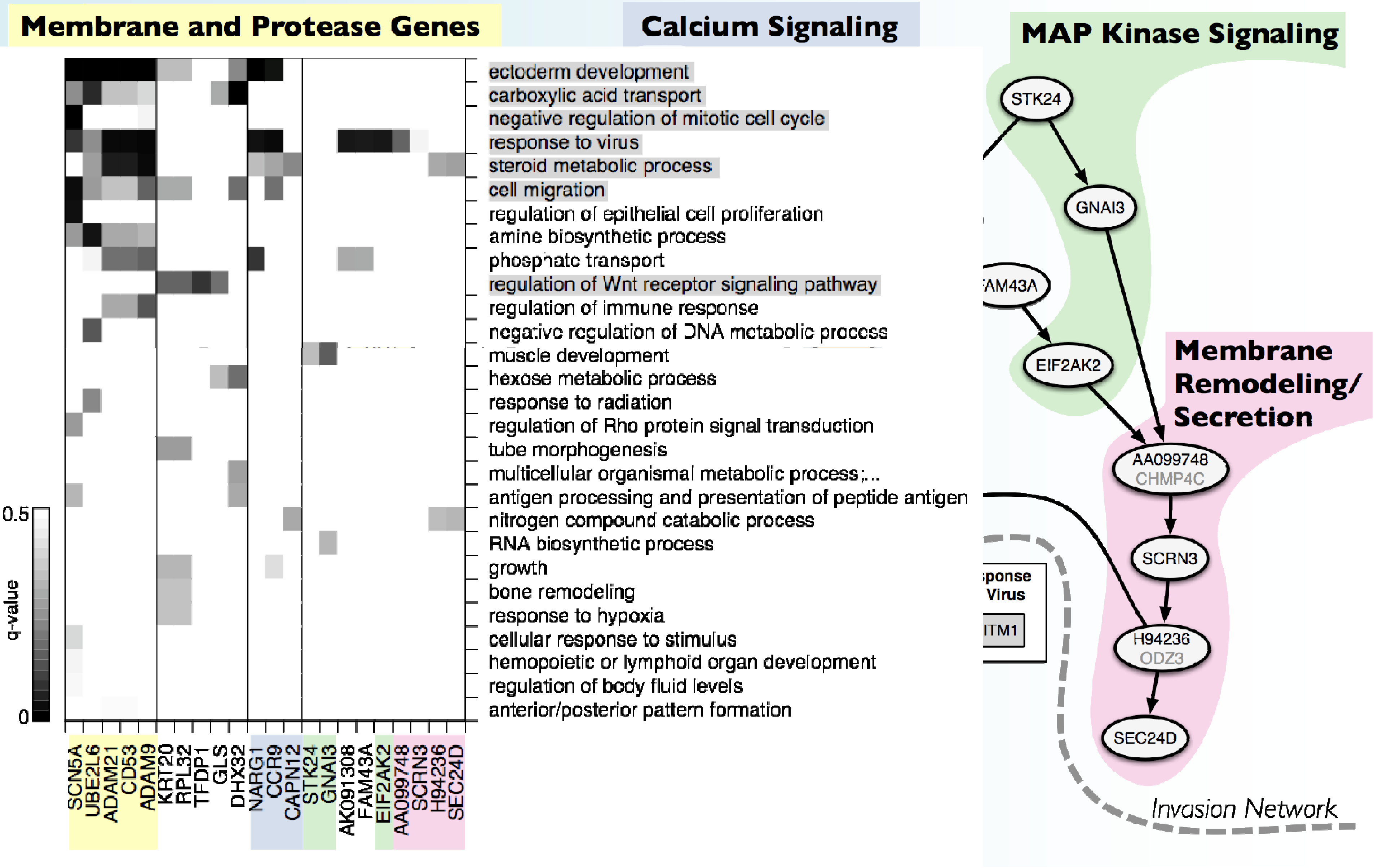


Invasion Network

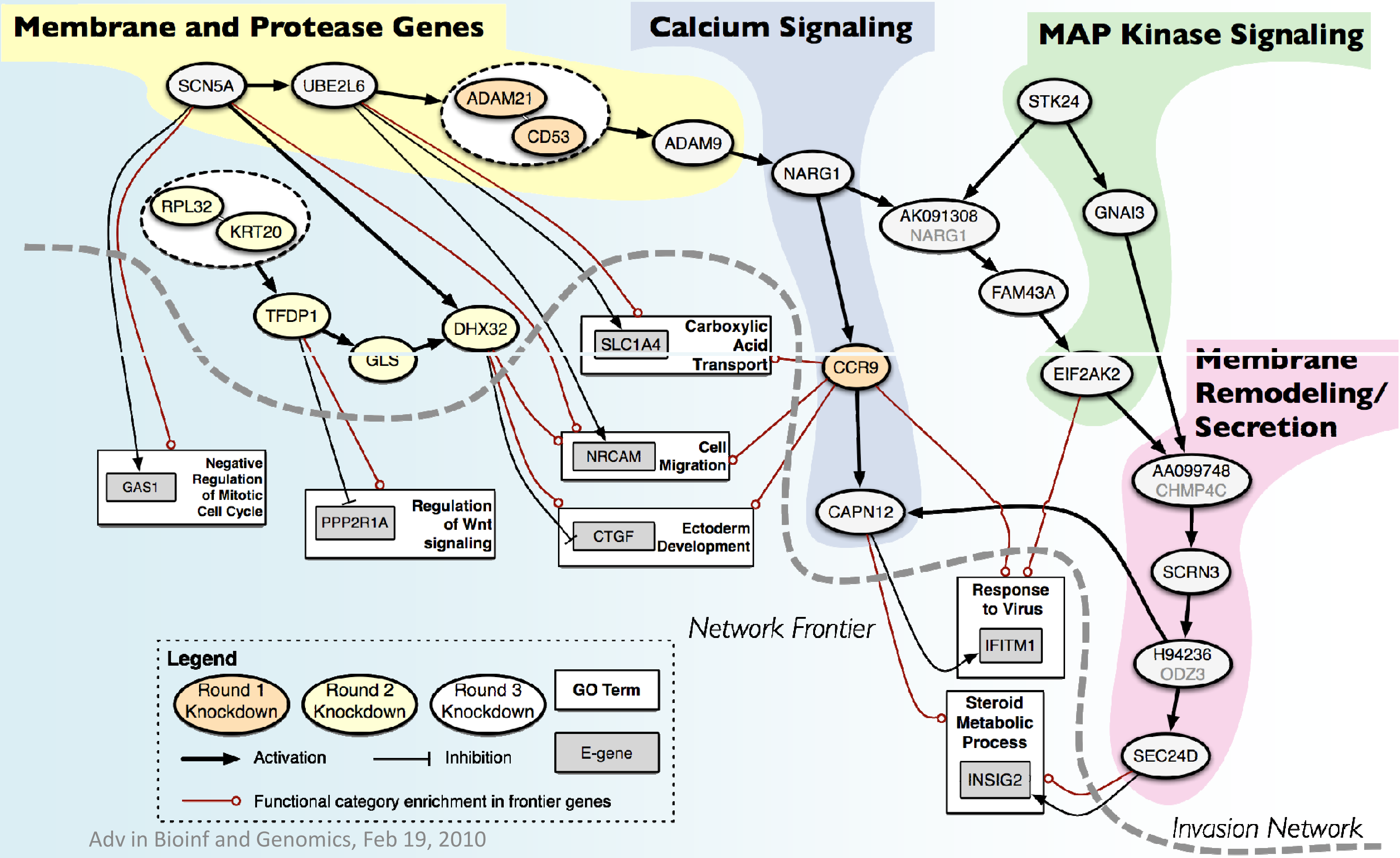


CTGF - AA598794 - connective tissue growth factor
COL17A1 - AA128561 - collagen, type XVII, alpha 1
COL5A2 - AA857098 - collagen, type V, alpha 2
LAMA4 - N94616 - laminin, alpha 4
COL3A1 - T98612 - collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal domi
LAMC2 - AA677534 - laminin, gamma 2
PROX1 - AA932629 - prospero-related homeobox 1
KRT10 - AA428454 - keratin 10 (epidermolytic hyperkeratosis; keratosis palmaris et plantaris)
PLOD1 - AA476240 - procollagen-lysine 1, 2-oxoglutarate 5-dioxygenase 1
PPL - AI680568 - periplakin
COL3A1 - T98612 - collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal domi
TGFB2 - N45138 - transforming growth factor, beta 2
DHCR24 - AI057612 - 24-dehydrocholesterol reductase
ELF3 - AA433851 - E74-like factor 3 (ets domain transcription factor, epithelial-specific)
DCT - AA478553 - dopachrome tautomerase (dopachrome delta-isomerase, tyrosine-related pro
EMP1 - AA975768 - epithelial membrane protein 1
KRT13 - W60057 - keratin 13
UGCG - AA165628 - UDP-glucose ceramide glucosyltransferase
PPL - AI126713 - periplakin
PPL - AA454209 - periplakin
FABP5 - T60075 - fatty acid binding protein 5 (psoriasis-associated)
IVL - AI417983 - involucrin
SPRR1B - AA447835 - small proline-rich protein 1B (cornifin)
PPHLN1 - R52965 - periphilin 1
ERCC2 - R05503 - excision repair cross-complementing rodent repair deficiency, complementati

Invasion Network



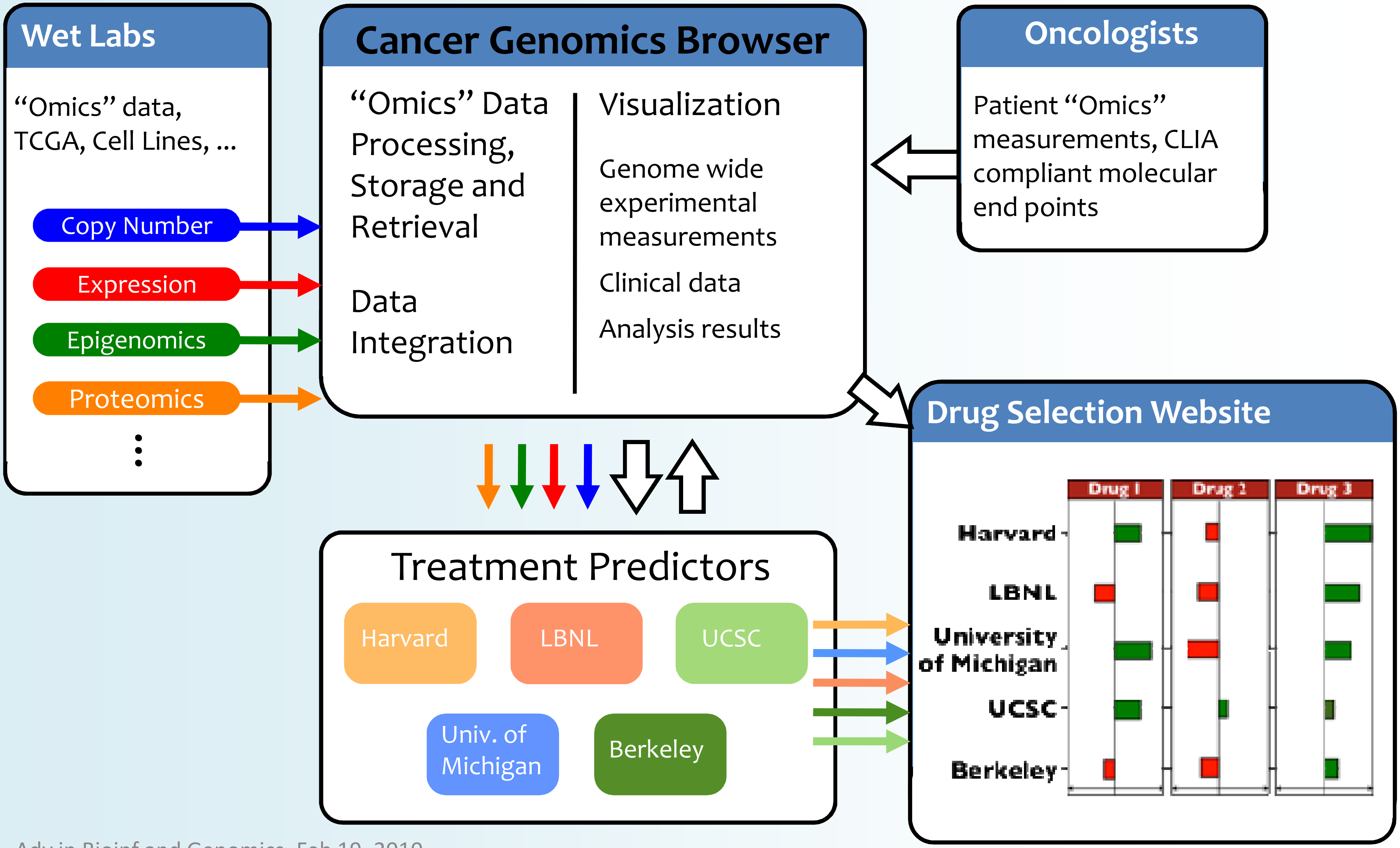
Invasion Network



Summary

- KO microarrays provide causal information
- Inferred pathways provide better framework for identifying new members compared to correlation
- Could use to identify new interactions
- Add *de novo* interactions to known
  - Measure increase in likelihood across patients
- Next steps will attempt to simulate KOs on pathways to predict good drug targets

UCSC Cancer Genomics Browser
Public Research Portal



Wet Labs

"Omics" data, TCGA, Cell Lines, …

- Copy Number
- Expression
- Epigenomics
- Proteomics
- ⋮

Cancer Genomics Browser

"Omics" Data Processing, Storage and Retrieval

Data Integration

Visualization

Genome wide experimental measurements

Clinical data

Analysis results

Oncologists

Patient "Omics" measurements, CLIA compliant molecular end points

Treatment Predictors

- Harvard
- LBNL
- UCSC
- Univ. of Michigan
- Berkeley

Drug Selection Website

THE CANCER GENOME ATLAS

STAND UP TO CANCER

NATIONAL INSTITUTES OF HEALTH

qb3 ucb-ucsc-ucsf

- **UCSC Cancer Genomics**
- **David Haussler**
- **Jim Kent**
- Josh Stuart
- Jingchun Zhu
- Ting Wang
- Fan Hsu
- Zack Sanborn
- Steve Benz
- Christopher Szeto
- Larry Meyer
- James Durbin
- Tracy Ballinger

- UCSC Genome Browser Staff

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

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