# *Development of predictive models to identify pathways relevant to compound sensitivity in cancer.*

**James R. Ortega**

**Florida International University**

**Mentor: Amrita Basu, PhD**
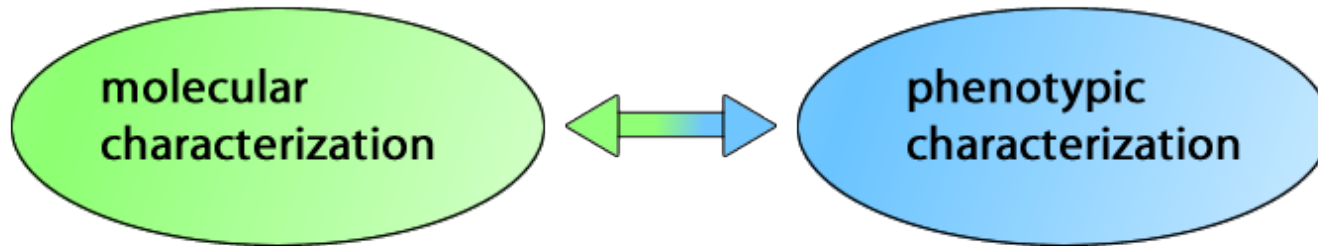
**Clemons Group**

**Broad Institute SRPG**

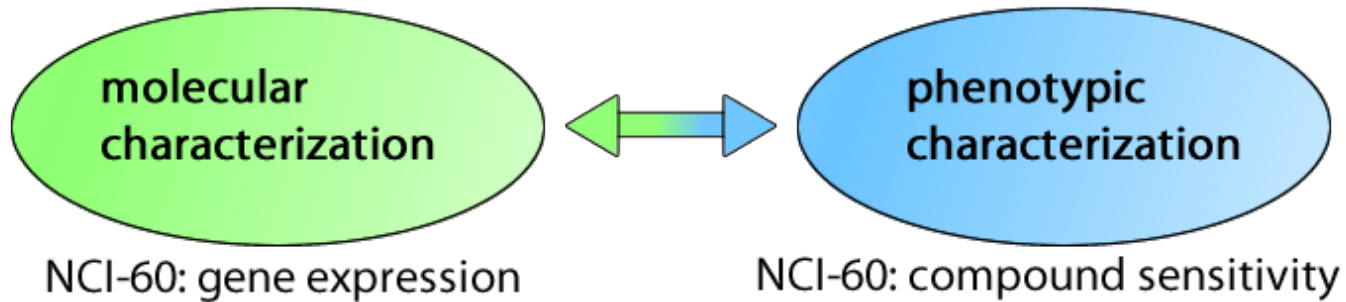# Importance

- Aim: To characterize the biological determinants of compound sensitivity in cancer.

  - Personalization and improved therapeutics.

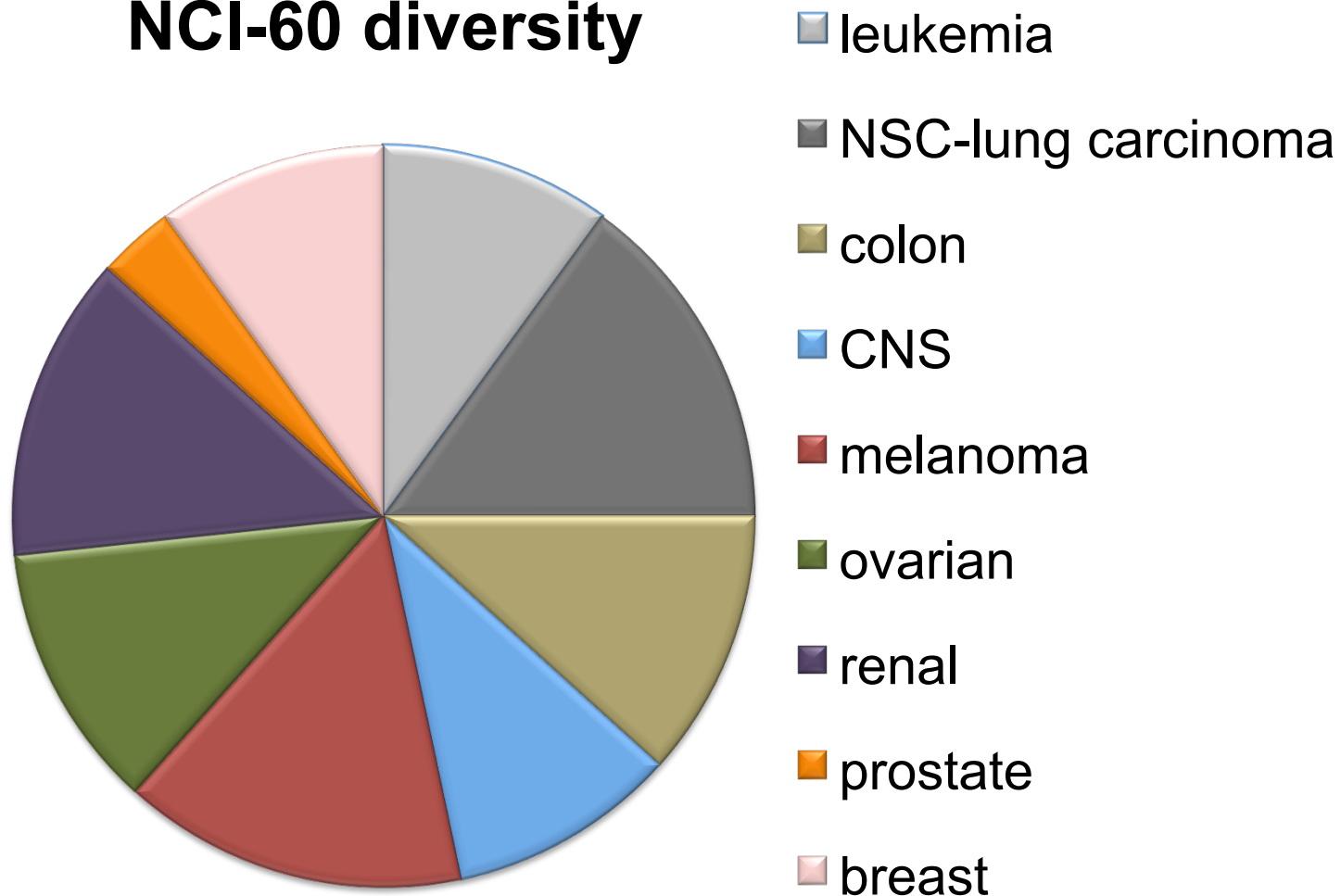**BROAD** INSTITUTE

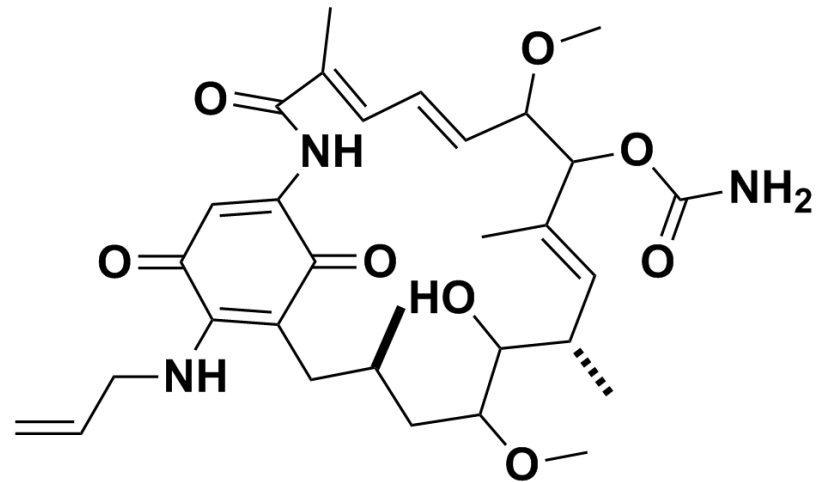# Methodology

# Methodology

# The NCI-60 cancer cell lines

## NCI-60 diversity



- leukemia
- NSC-lung carcinoma
- colon
- CNS
- melanoma
- ovarian
- renal
- prostate
- breast

Monks et al. (1991) . J Natl Cancer Inst. 83(11):757-66
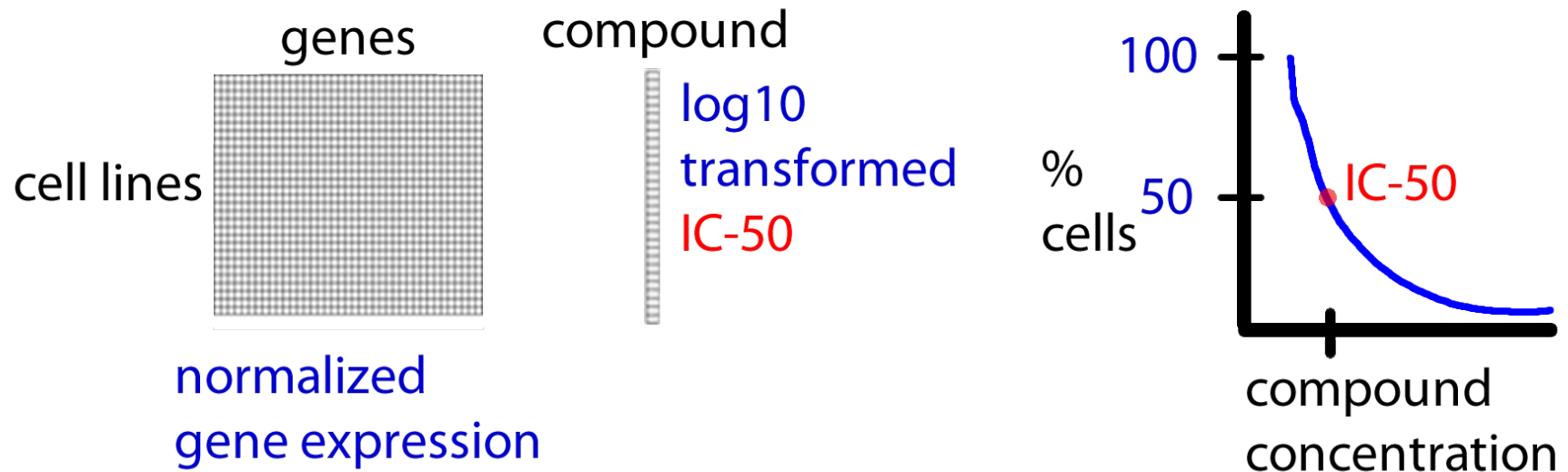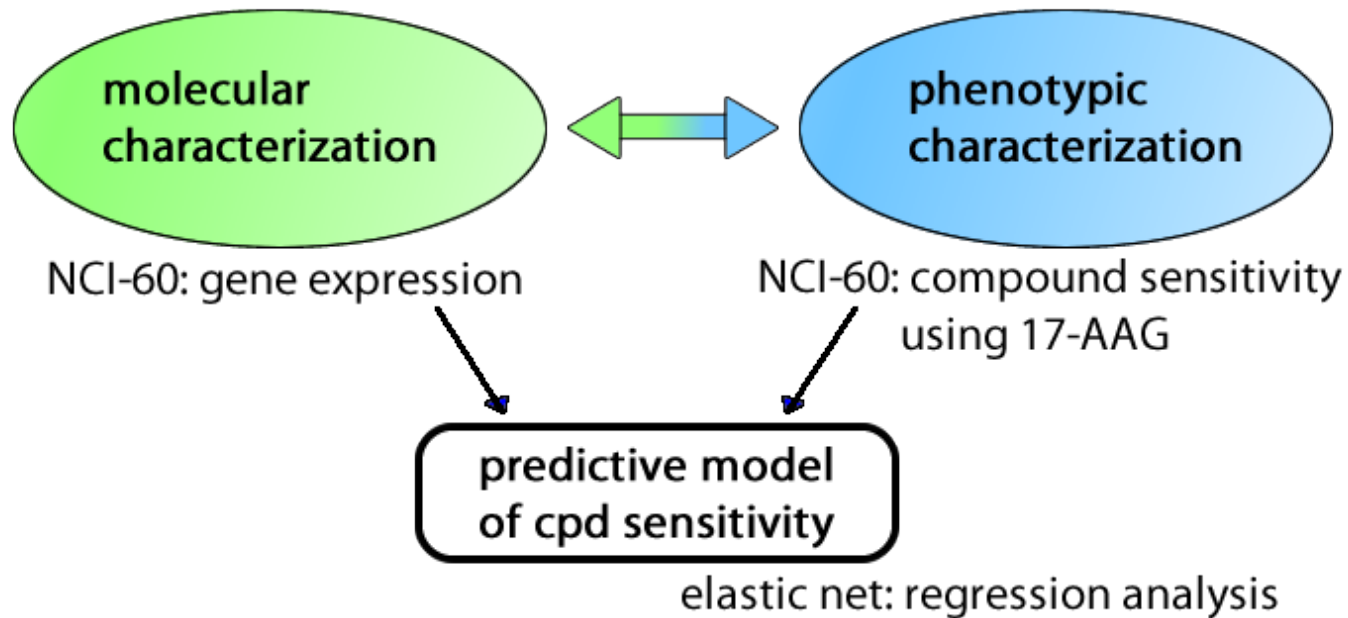
BROAD
INSTITUTE

# 17-allylamino-geldanamycin(17-AAG)

- Being investigated in many cancers, including phase II clinical trials for breast cancer
- Inhibitor of Hsp90

Modi et al. (2011) Clin Cancer Res [Epub ahead of print]
Schulte & Neckers (1998) Cancer Chemother Pharmacol 42(4):273-9

BROAD
INSTITUTE

# Data pipeline



genes

cell lines

normalized
gene expression

compound

log10
transformed
IC-50

100

%
cells 50

IC-50

compound
concentration

BROAD
INSTITUTE

# Methodology



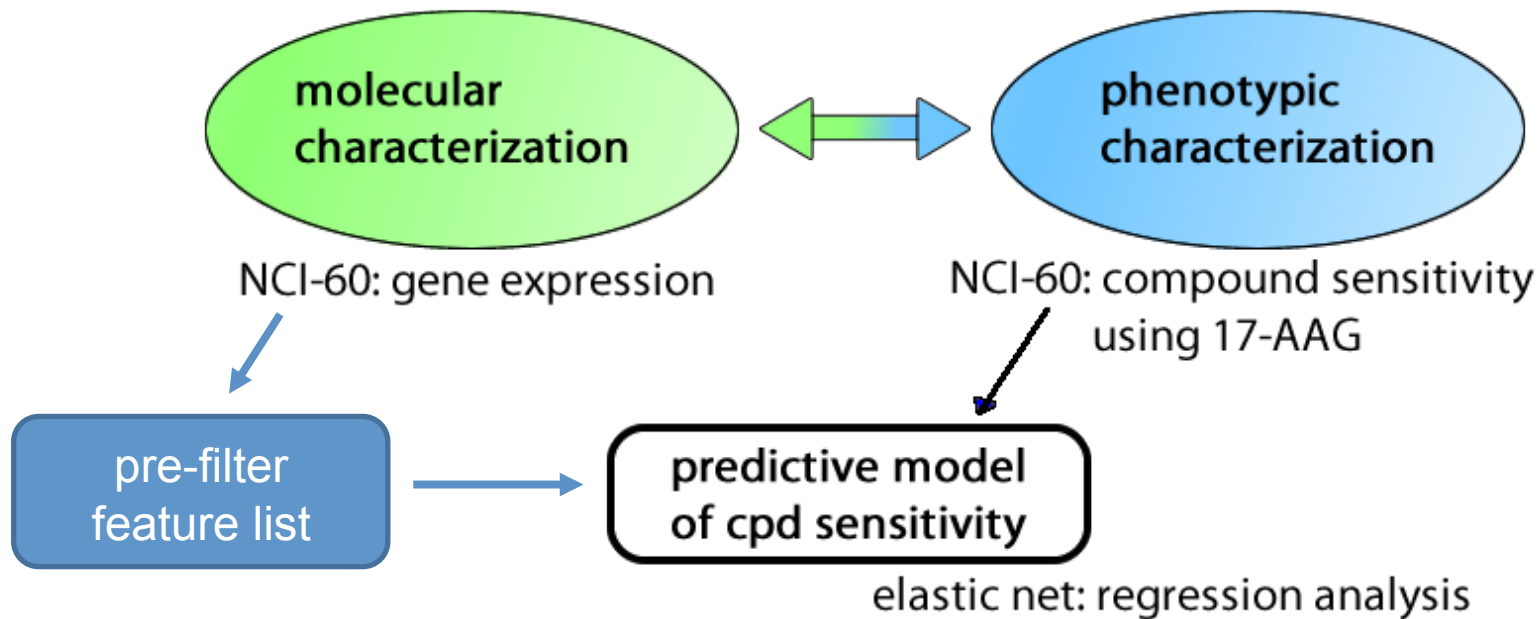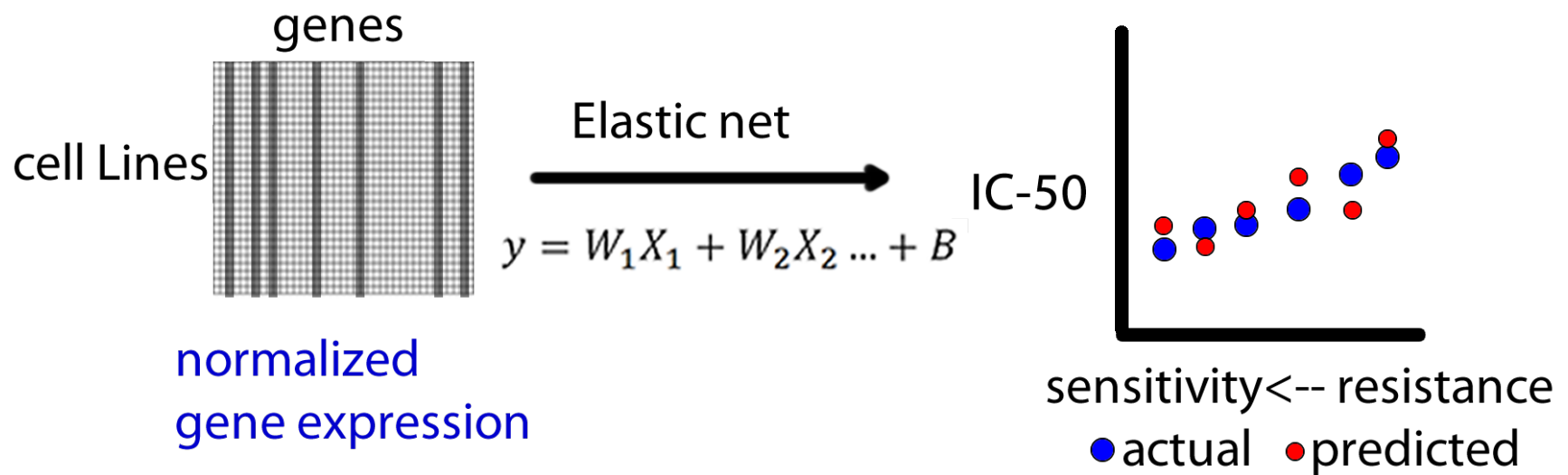Zou & Hastie (2005) *Appl Stat J Roy St-C* 67(2):301-320 BROAD INSTITUTE

# Elastic net regression

- Allows for
  - The analysis of high-dimension data sets
  - Quantitative predictions
  - Simple model interpretation
- Limitation
  - Sensitive to the number of input features
- Implementation
  - Ranking and scheduling algorithms for feature selection

BROAD INSTITUTE

# Methodology



molecular characterization

NCI-60: gene expression

phenotypic characterization

NCI-60: compound sensitivity using 17-AAG

pre-filter feature list

predictive model of cpd sensitivity

elastic net: regression analysis

Zou & Hastie (2005) *Appl Stat J Roy St-C* 67(2):301-320

BROAD INSTITUTE

# Data pipeline



genes

cell Lines

normalized
gene expression

Elastic net

$$y = W_1 X_1 + W_2 X_2 \ldots + B$$

IC-50

sensitivity<-- resistance

● actual   ● predicted

BROAD
INSTITUTE

# Scheduling algorithm results



17-allylamino-geldanamycin:RMSE vs number of top features selected

204 input gene expression features
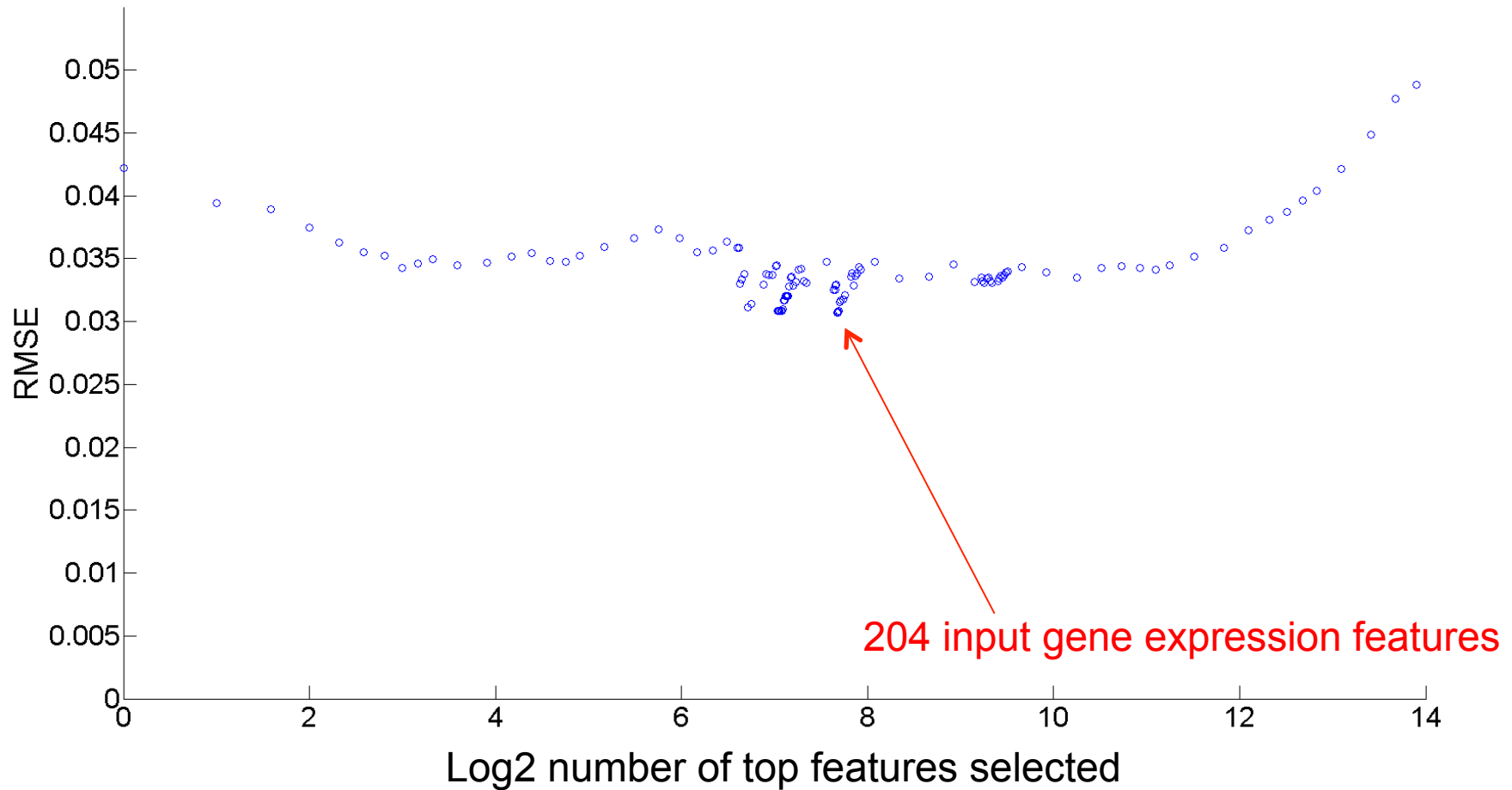
# Elastic net predictions



NCI-60 compound sensitivity:17-allylamino-geldanamycin

sensitivity <- resistance
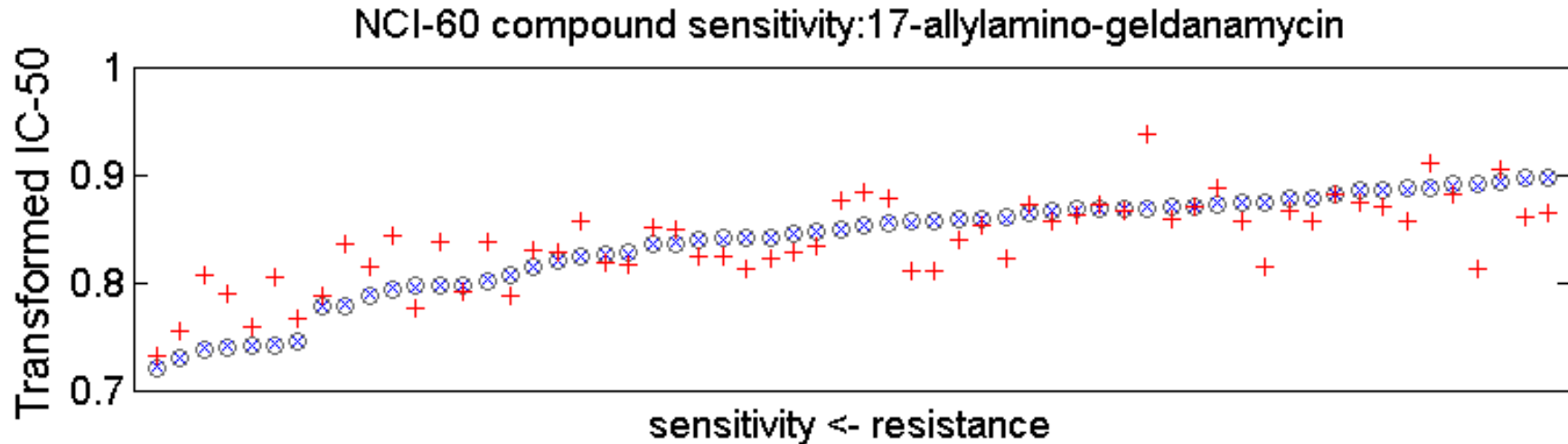
Feature: expression
BROAD CPD ID: BRD-K81473043
Compound name: 17-allylamino-geldanamycin
Num cell lines: 60
alpha: 0.3
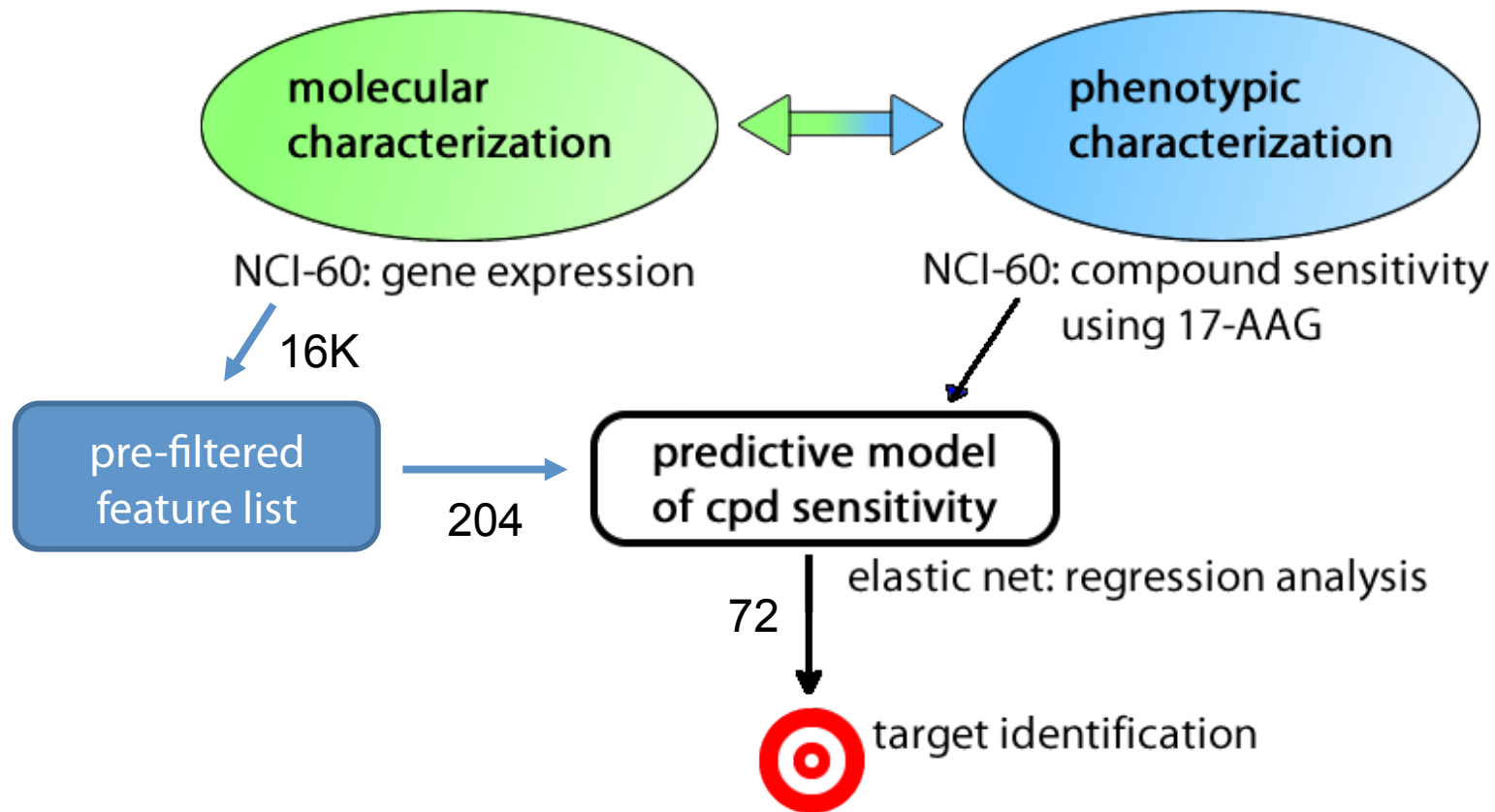lambda: 0.00024592
Nonzero features: 72
Intercept: 0.83656
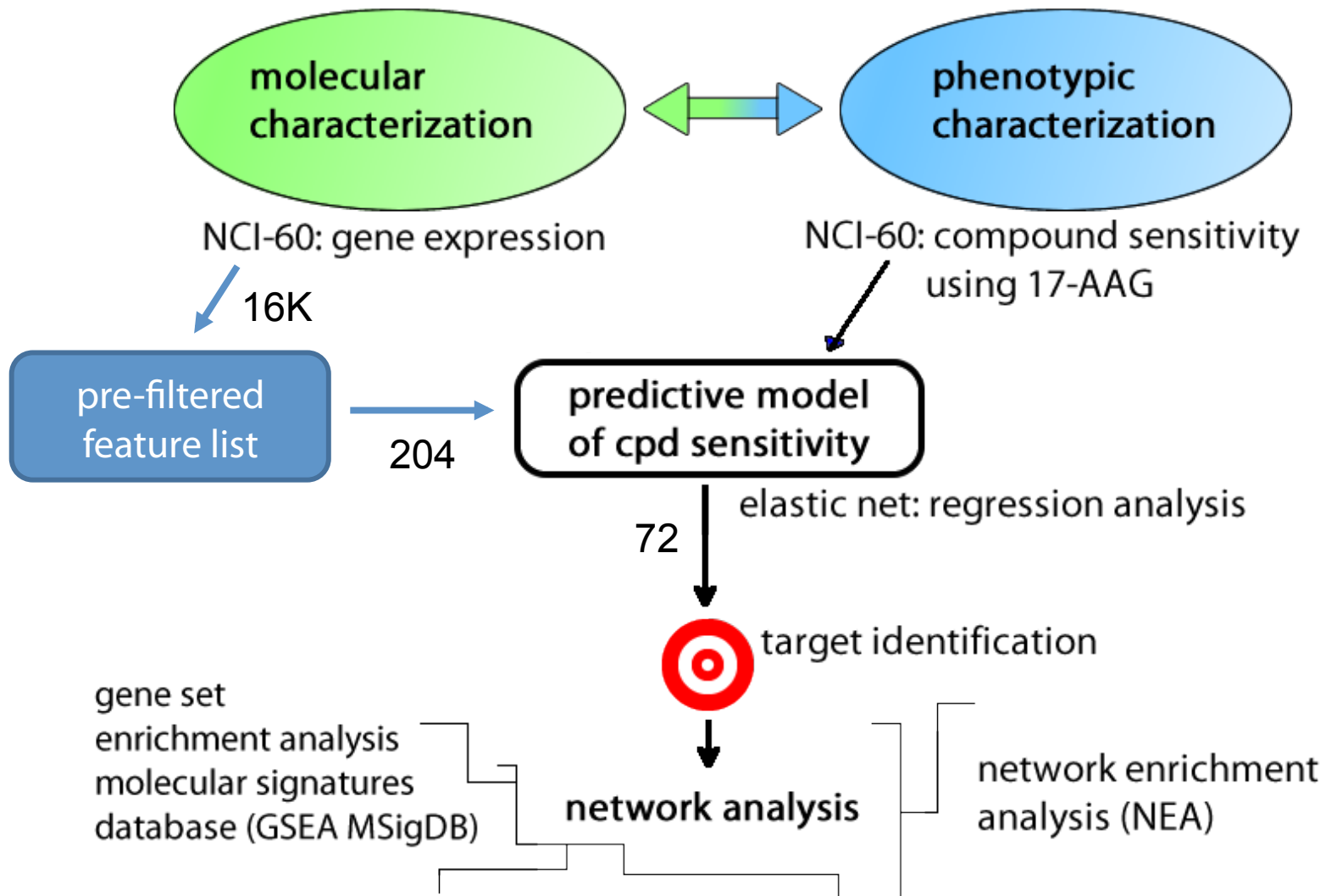Model RMS error: 0.03073
Weighted RMS error: 0.03073
run_id: 201107212030-87051-A204.BRD-K81473043

- o→ Actual sensitivity
- x →Prediction using all 60 cell lines
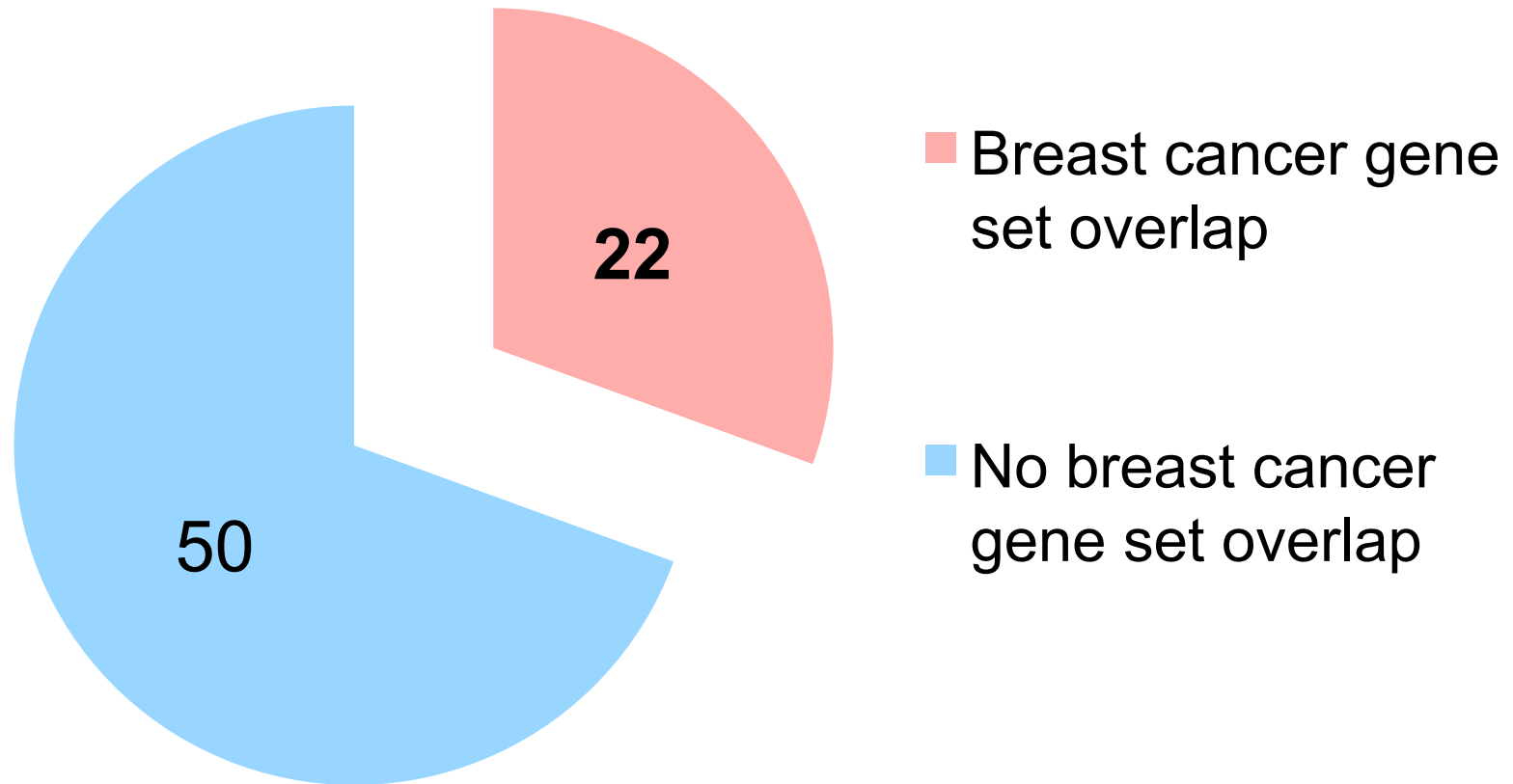- + →Prediction using leave-one-out cross-validation

BROAD
INSTITUTE

# Methodology



Zou & Hastie (2005) *Appl Stat J Roy St-C* 67(2):301-320

BROAD INSTITUTE

# Methodology



Subramanian et al. (2005) *PNAS* 102(43):15545-15550
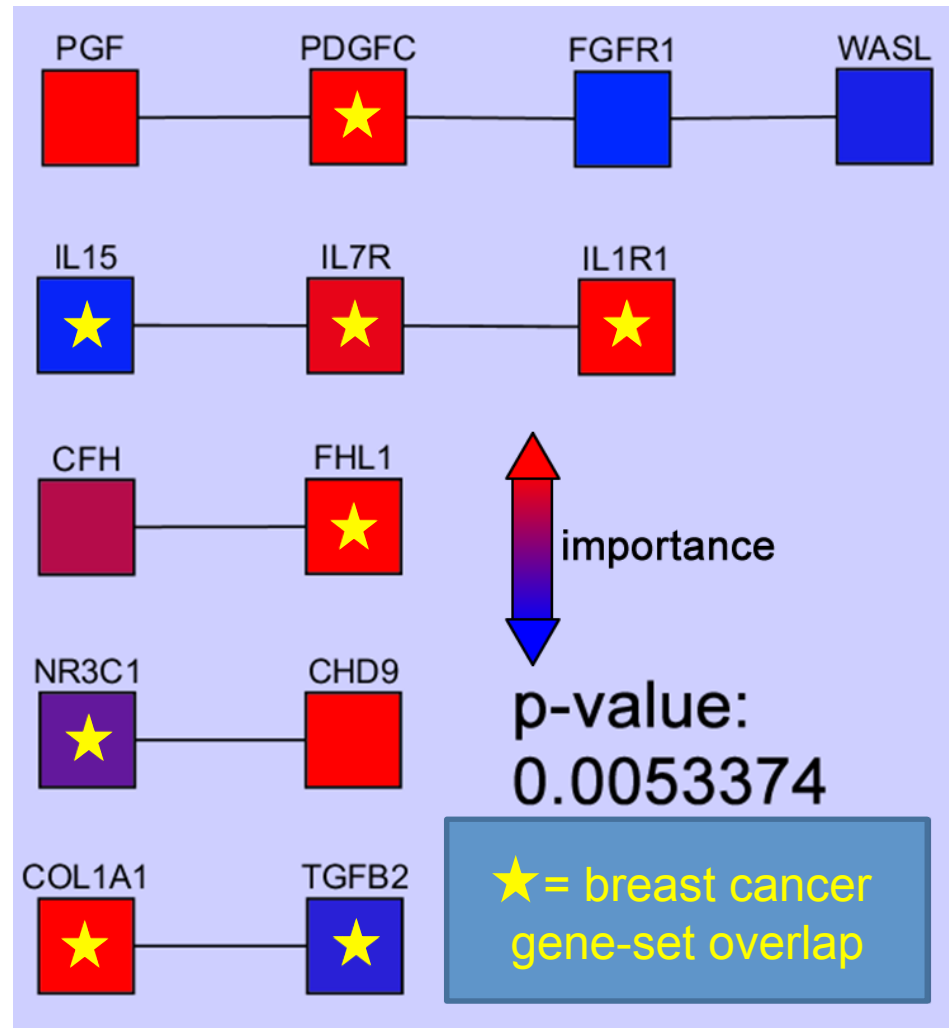
**BROAD** INSTITUTE

# GSEA MSigDB network analysis results

**GSEA MSigDB breast cancer overlaps**

# Network Enrichment Analysis results

# Conclusion

- Our method has provided a framework for future analyses of the genetic basis for compound sensitivity in cancer.

- In addition, our method of analysis could be useful for the study of other diseases and biological contexts.

**BROAD** INSTITUTE

# Acknowledgements

- ## Computational Chemical Biology
  - Amrita Basu
  - Nicole Bodycombe
  - Hyman Carrinski
  - Paul Clemons
  - Vladimir Dancik
  - Joshua Gilbert
  - Taner Kaya
  - Sandrine Muller

- ## Diversity Initiative
  - Bruce Birren
  - Nicole Edmonds
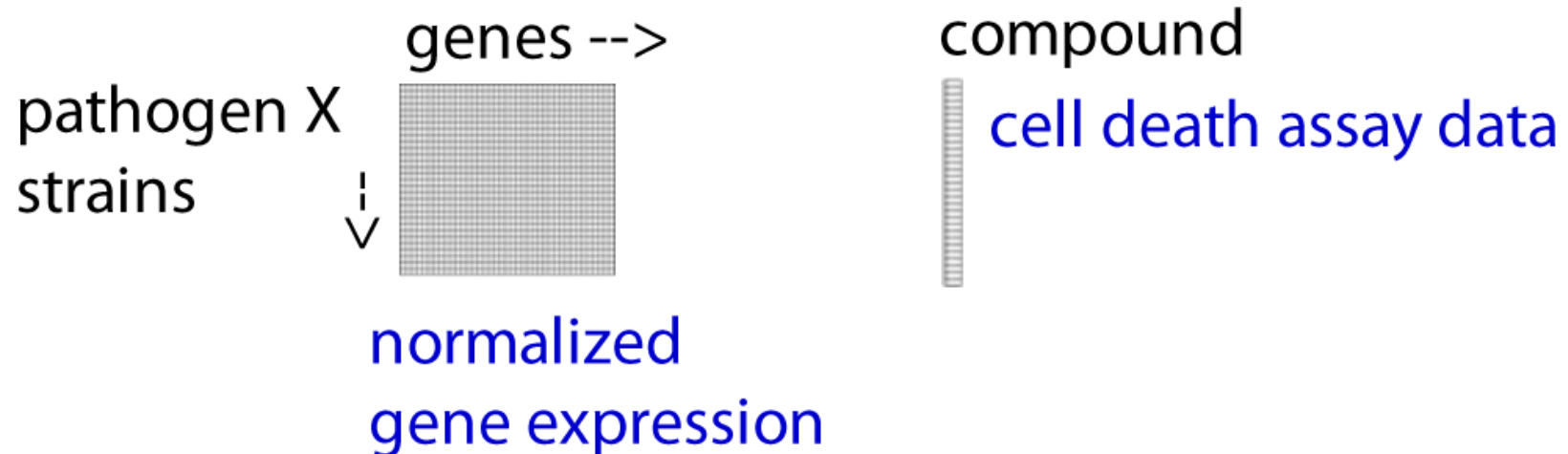  - Eboney Smith

**BROAD** INSTITUTE

# Potential applications

- Identifying pathways relevant to stem cell differentiation.
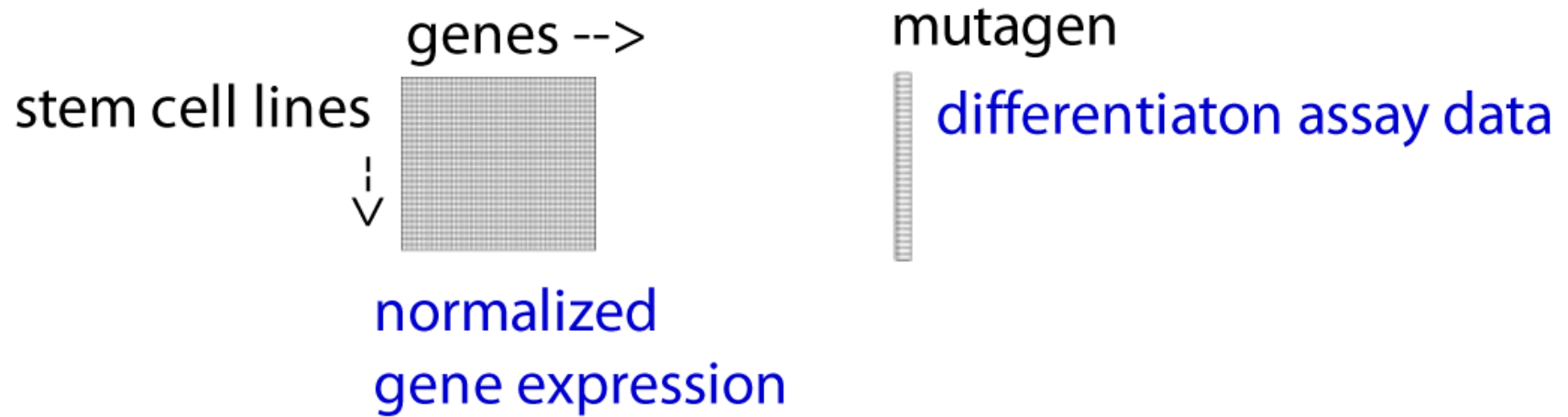
# Potential applications

- Identifying pathways relevant to compound sensitivity in infectious diseases

genes -->

compound

pathogen X strains

cell death assay data

normalized gene expression

BROAD INSTITUTE

# Potential applications

- Identifying small molecule probes relevant to the genetics of a given cellular context.

BROAD INSTITUTE