Strategies for Mining the NCI's Screening Databases:

Data (NCI60, Xenograft)

Informatics (Bio and Chemo)

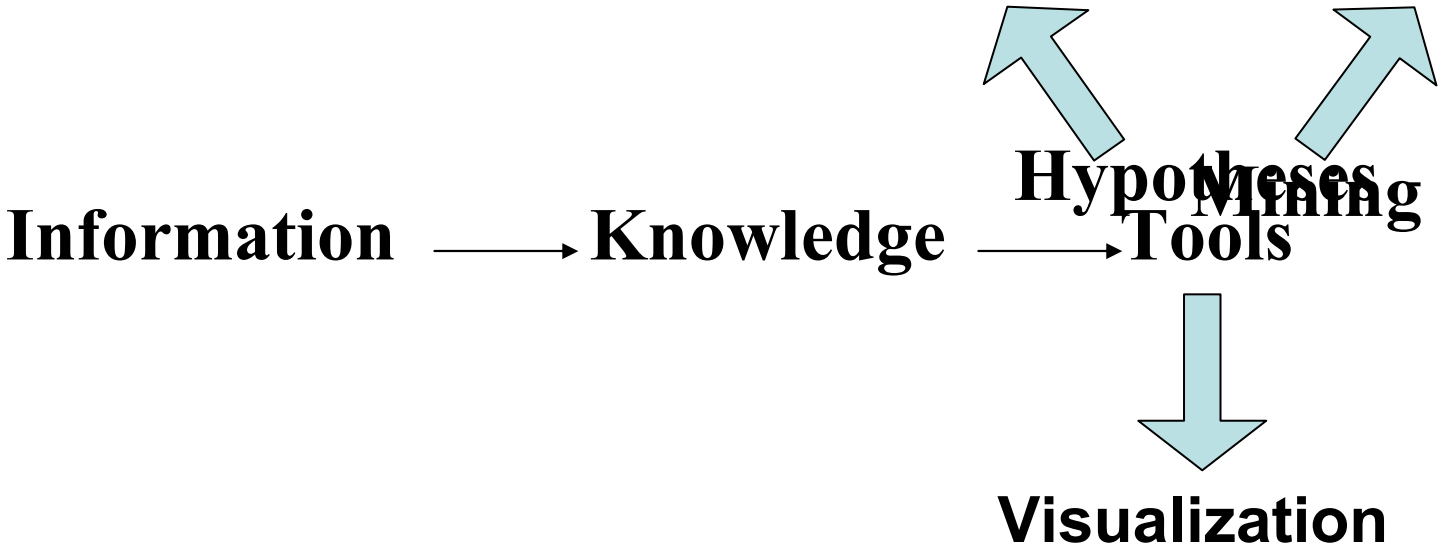
Laboratory of Computational Technologies

Anders Wallqvist, Ruili Huang, Narmada Thanki, Xiang-Jun Lu, Alfred Rabow

NIH/NCI/DCTD/DTP/STB

Drs. Doroshow, Collins, Shoemaker

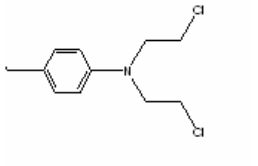
spheroid.ncifcrf.gov

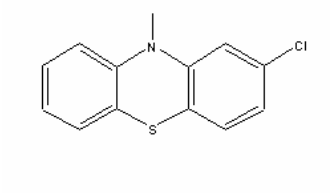


- Successes
- Pitfalls
- Strategies
- Recommendations

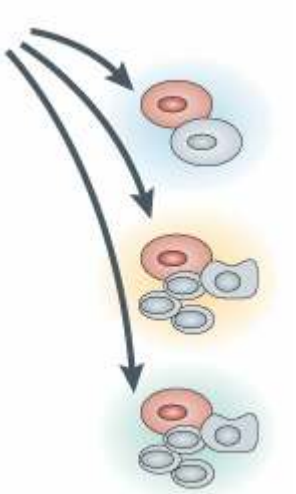
Data Generation / Data Analysis

Compounds

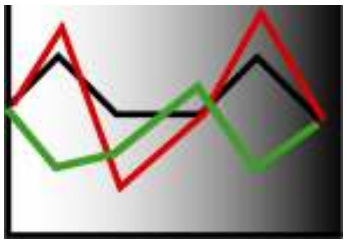




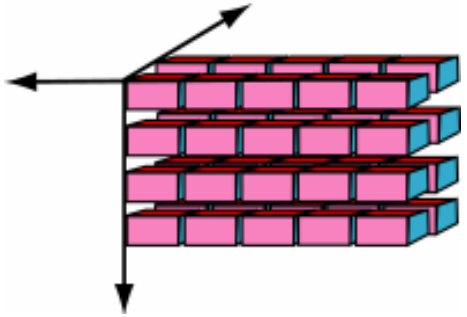
Functional screen

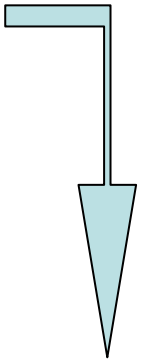


Phenotypic readout

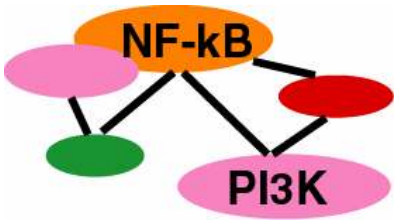


Database



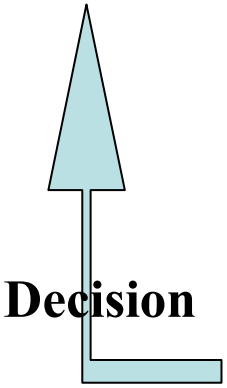


Gene Function

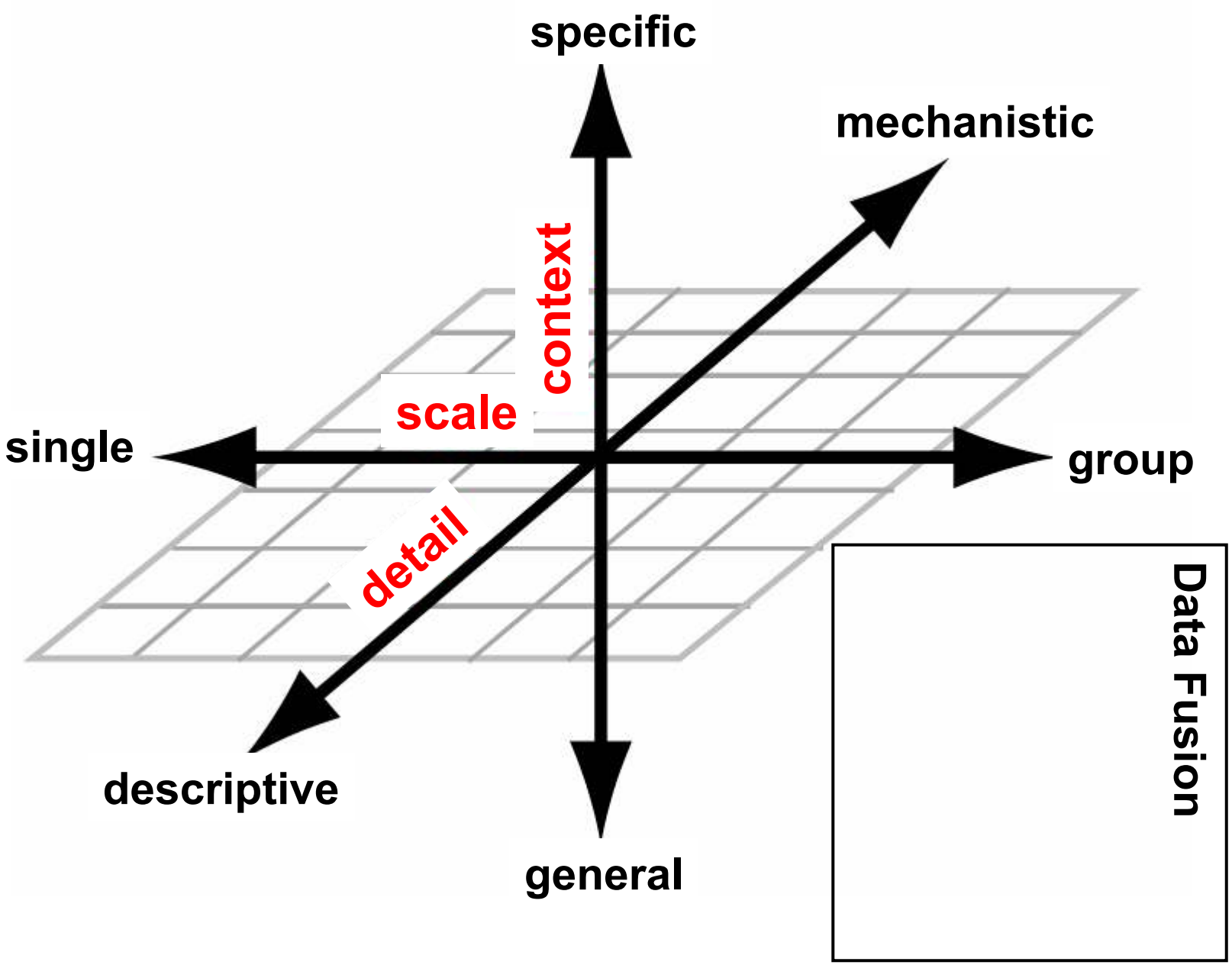


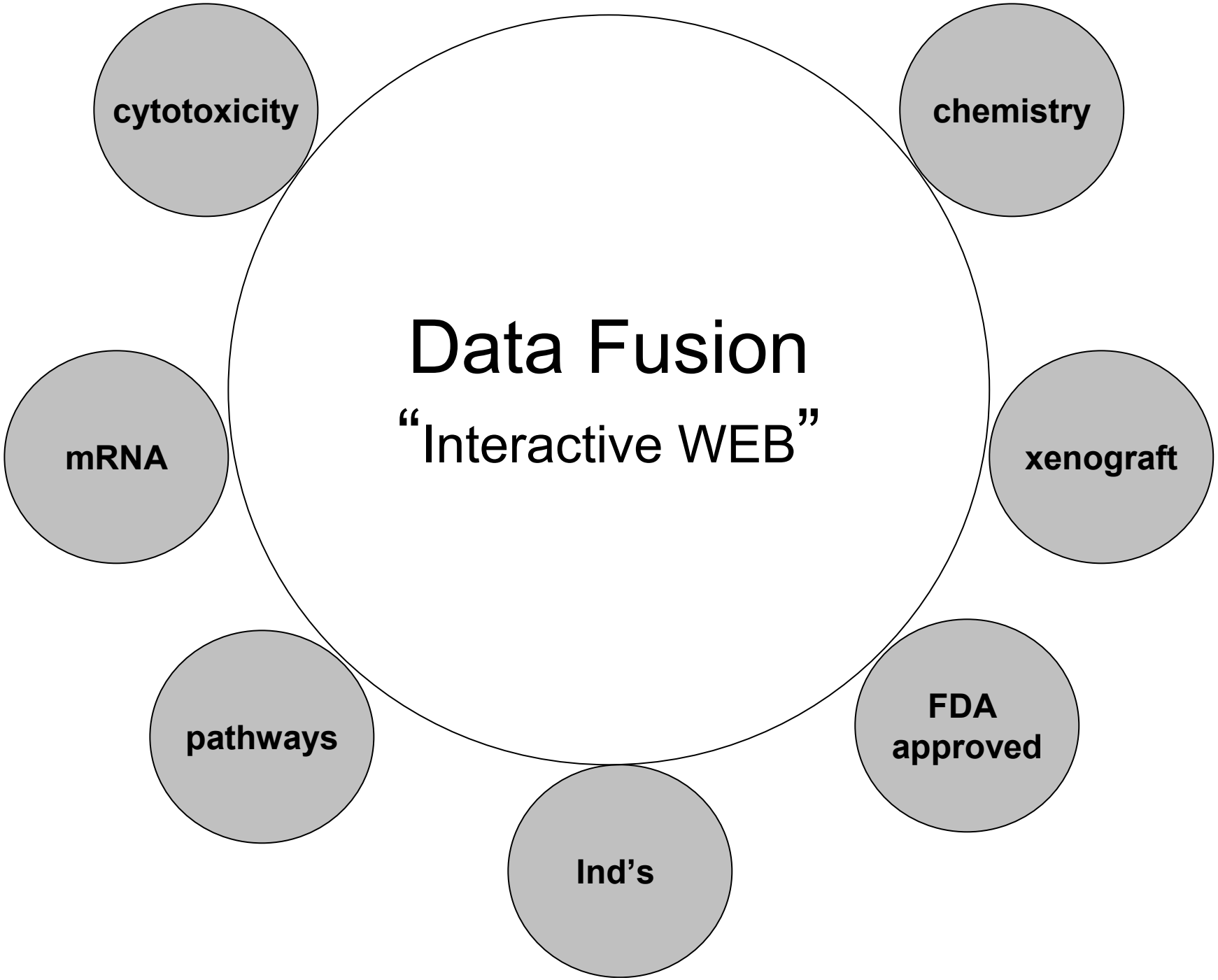
Statistics Mathematics

Decision

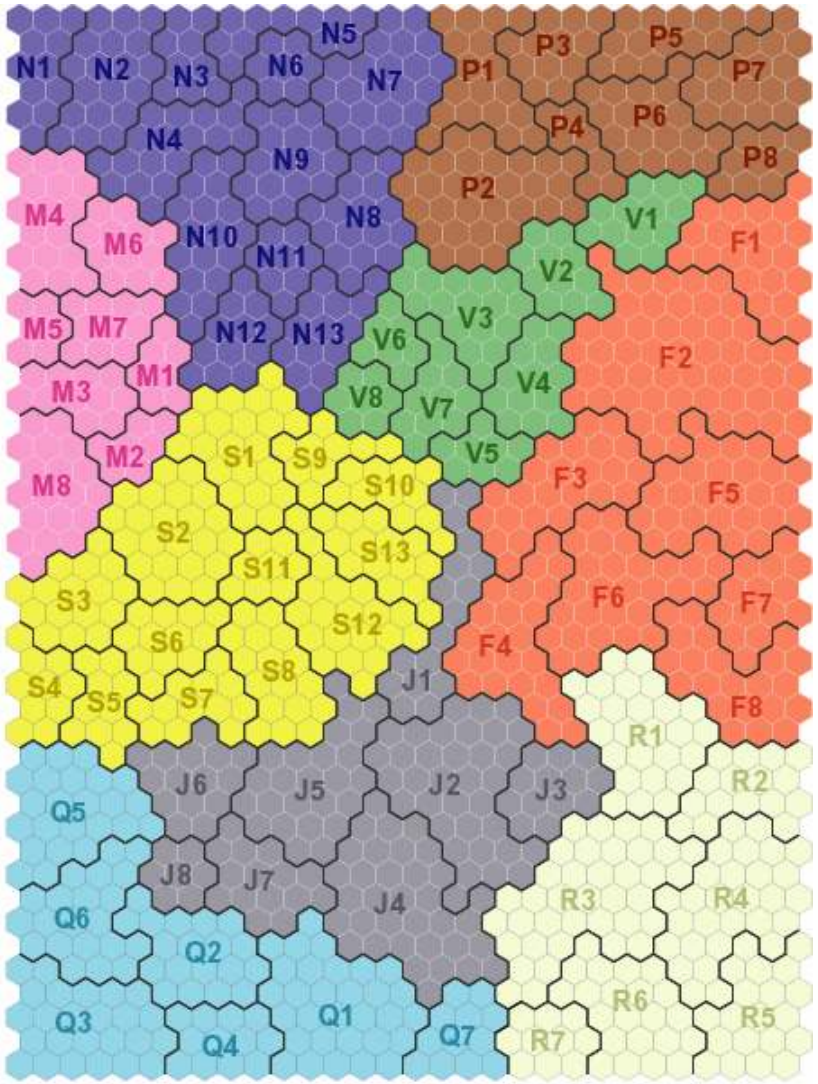


Drug Function


specific
mechanistic
context
scale
single
group
detail
descriptive
general
Data Fusion


cytotoxicity
chemistry
Data Fusion
"Interactive WEB"
mRNA
xenograft
pathways
FDA approved
Ind's

GI_{50} SOM



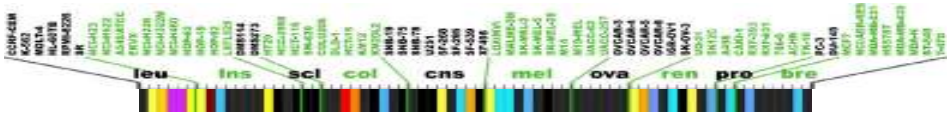
Rabow et al. J Med Chem (2002)
Wallqvist et al. JCIM (2006)

cytotoxicity **Successes**

NCI_{60}: lung, renal, colorectal, ovary, breast, prostate, central nervous system, melanoma, leukemia

~100,000 compounds

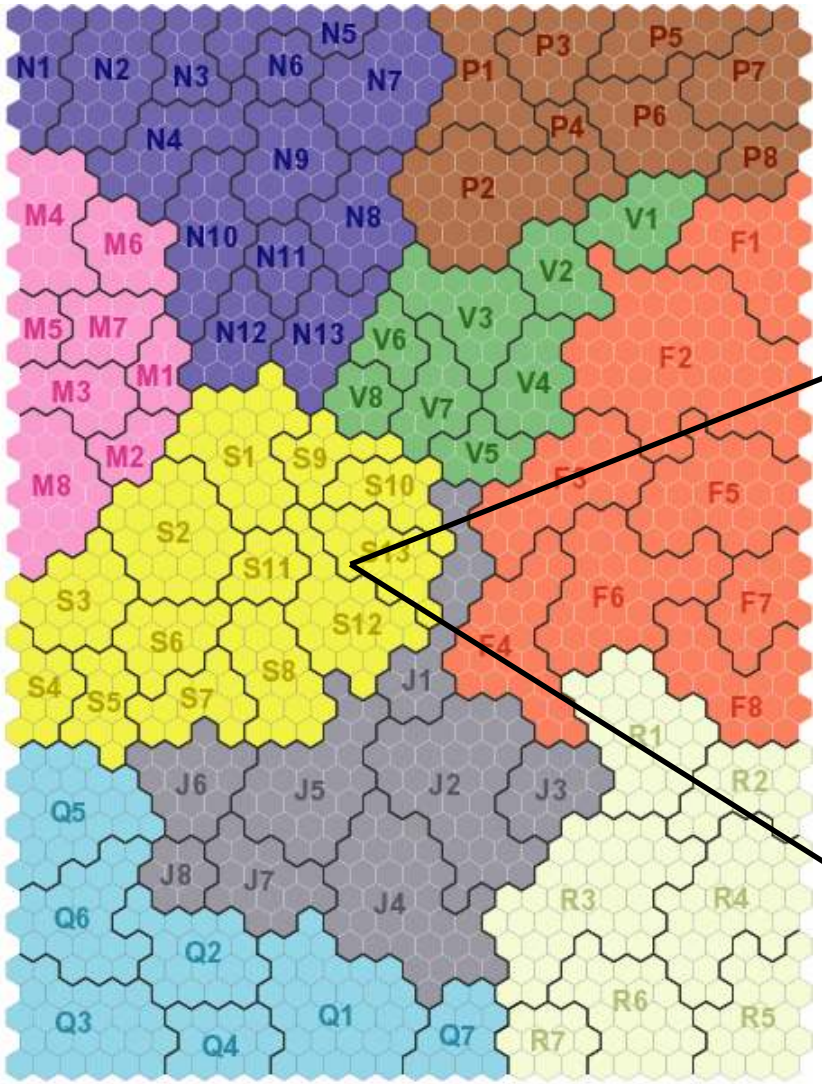
GI_{50}: 50% cancer cell growth inhibition concentration

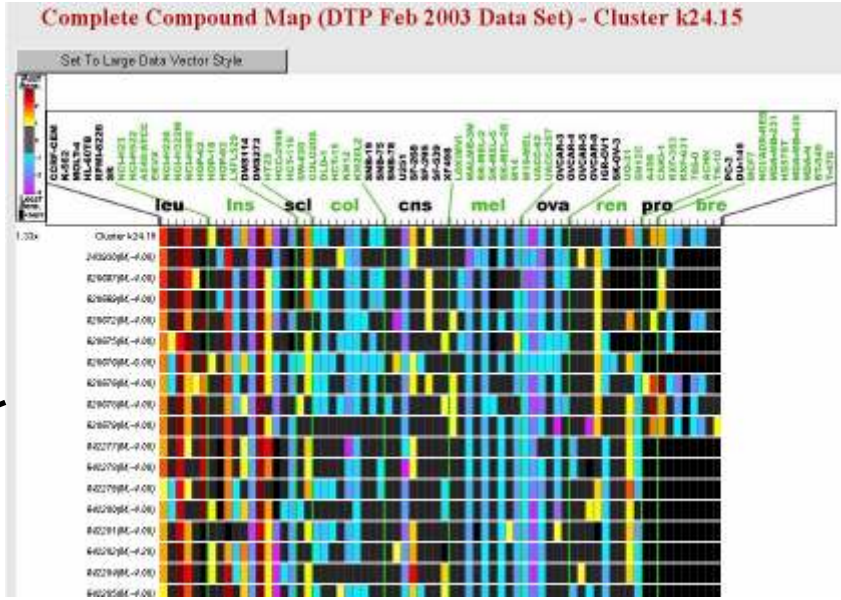


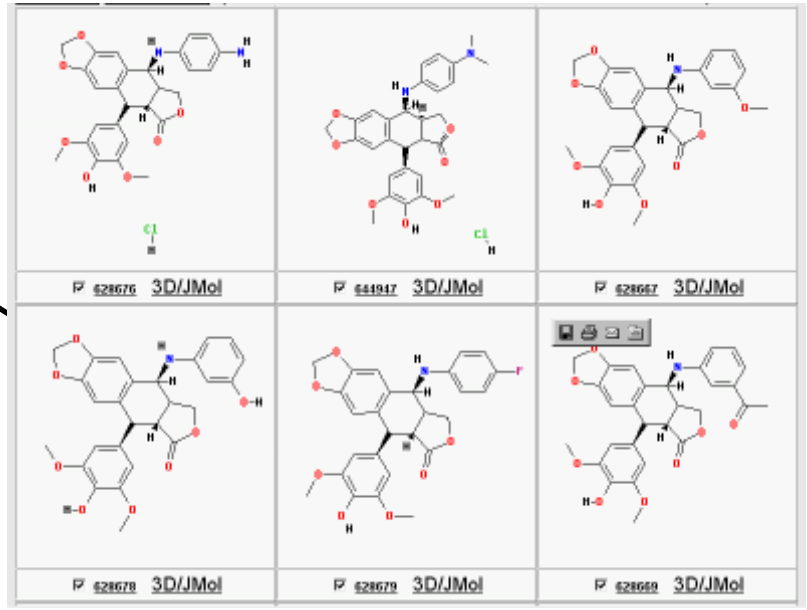
Legend to Cellular Growth Inhibition SOM

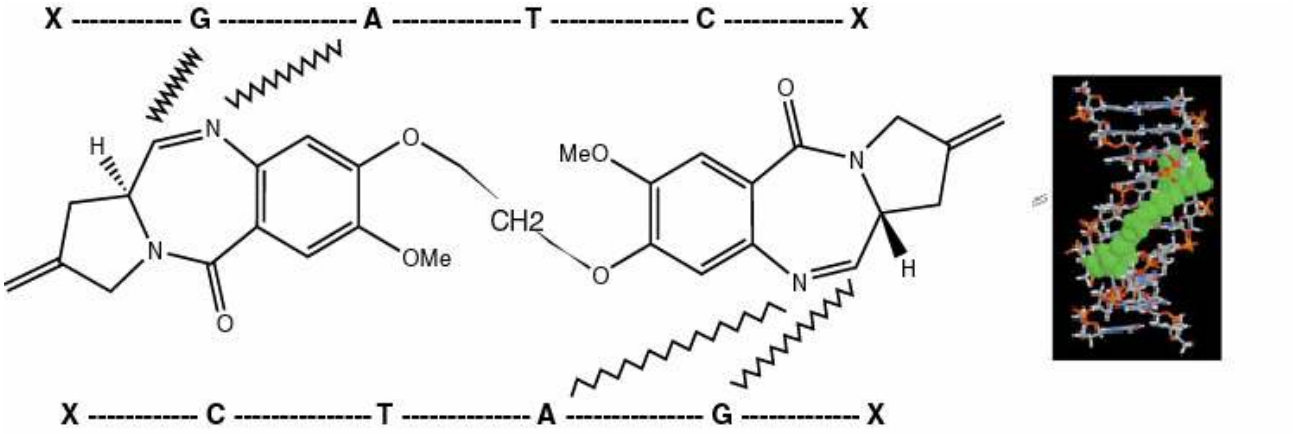
- Nucleic Acid Biosynthesis
- Xenobiotic Metabolism
- Membrane Transport/Integrity
- Mitosis
- Oxidative Metabolism
- Kinase/Phosphatase
- Work in Progress

GI_{50} SOM


N1
N2
N3
N5
N6
N7
N4
N9
N10
N11
N8
N12
N13
M4
M6
M5
M7
M3
M1
M2
M8
S1
S9
S10
S2
S11
S3
S6
S12
S4
S5
S7
S8
J1
J6
J5
J2
J3
J8
J7
J4
Q5
Q6
Q2
Q3
Q4
Q1
Q7
P1
P3
P5
P7
P4
P6
P8
P2
V1
V2
V3
V6
V4
V8
V7
V5
F1
F2
F5
F6
F7
F4
F8
R1
R2
R3
R4
R6
R5
R7

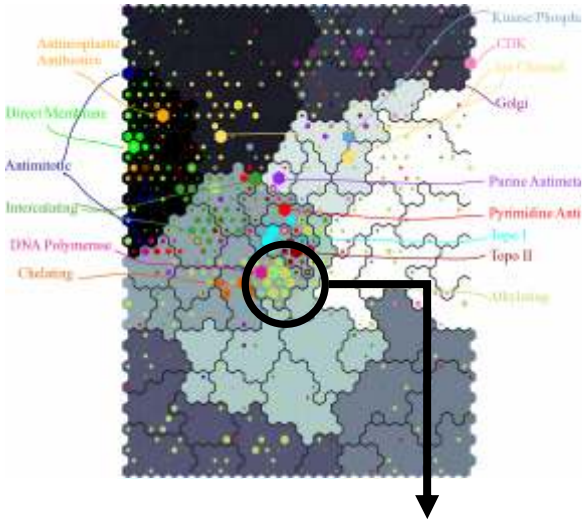

Complete Compound Map (DTP Feb 2003 Data Set) - Cluster k24.15
Set To Large Data Vector Style
leu
lns
scl
col
cns
mel
ova
ren
pro
bre


628676 3D/JMol
644947 3D/JMol
628667 3D/JMol
628678 3D/JMol
628679 3D/JMol
628669 3D/JMol



SJG-136 NSC694501

Neidle & Thurston Nat.Rev. 2005



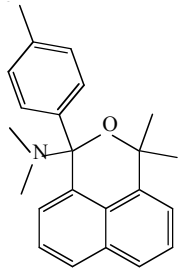
Alkylating

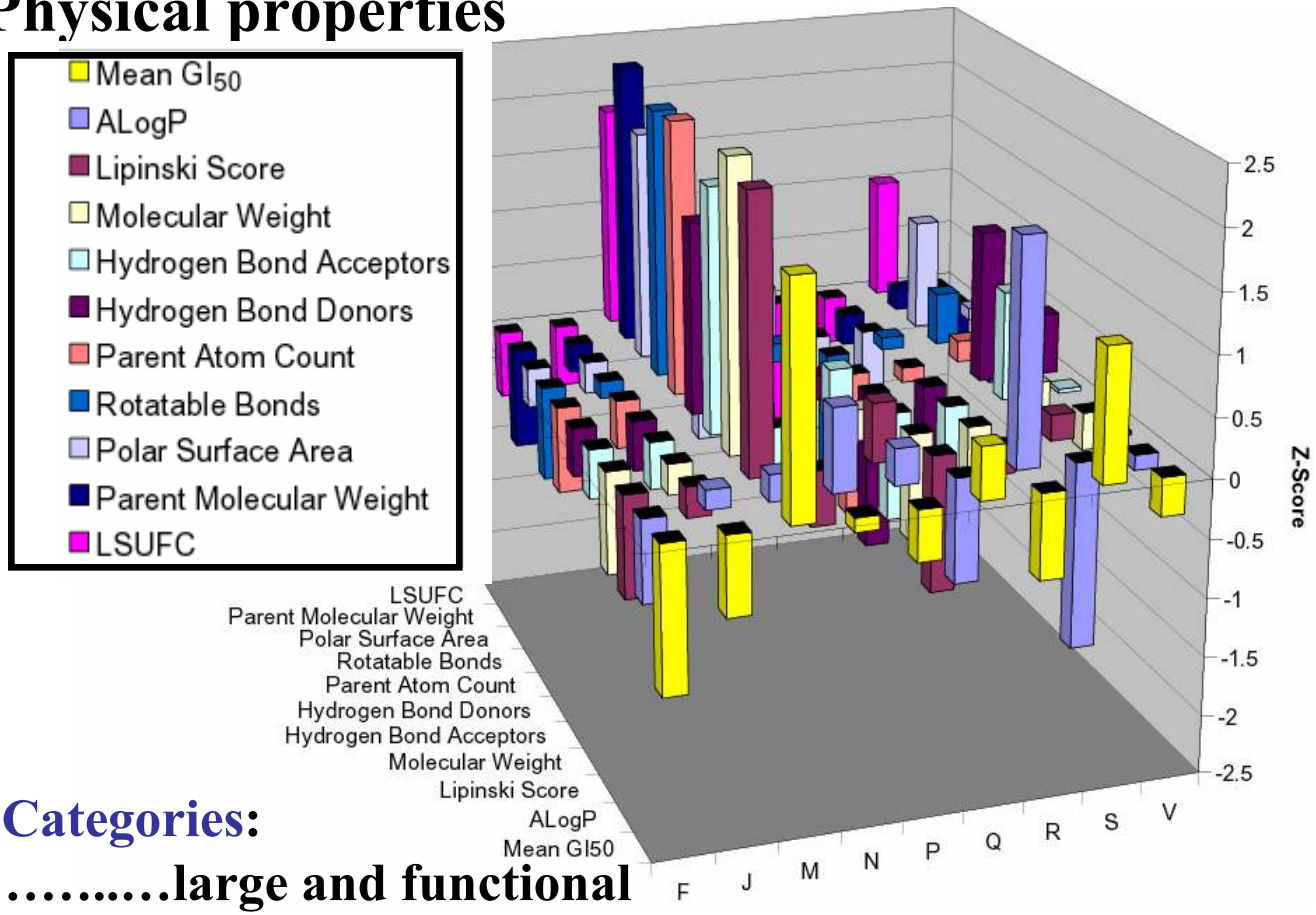
chemistry

success

Huang et al. J Med Chem 49:1964-1979 (2006)

Atoms and bonds **Physical properties**





Mechanism of Action Categories:

[M] Anti-mitotic……..…large and functional

[S] DNA synthesis……..low lipohilicity

[P] Phosphatase/kinase..…....most diverse signal

[R] Membrane active……high lipophilicity

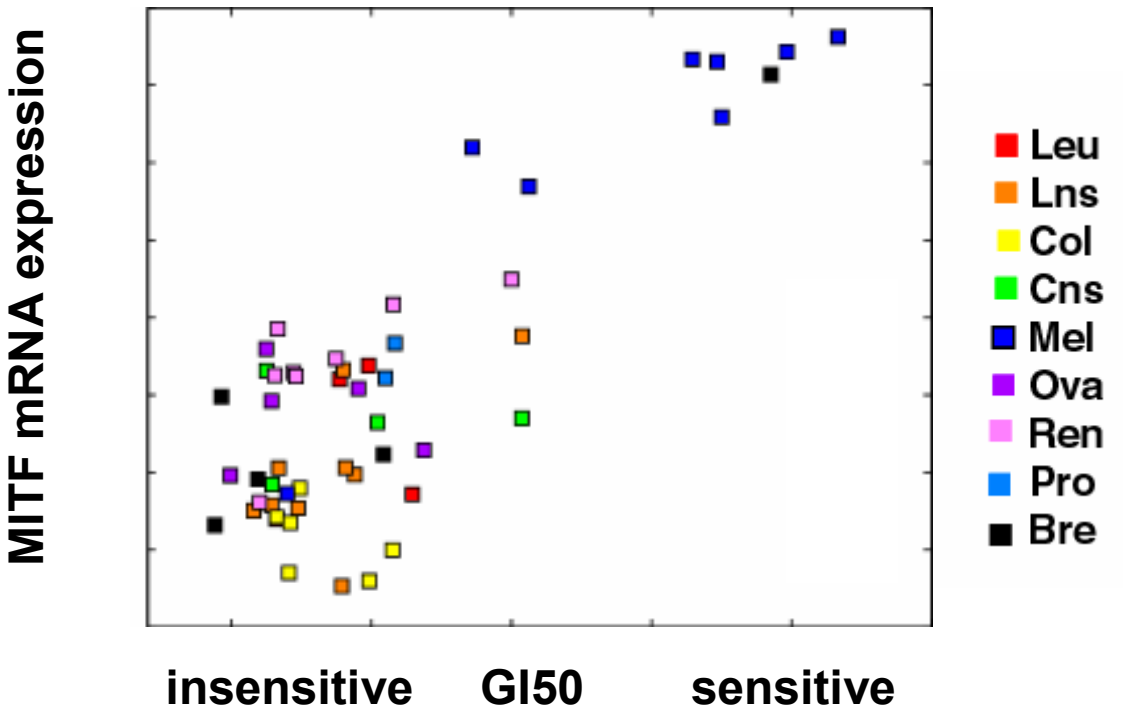
[Q] Xenobiotic metabolism…reactive groups

Chemistries: Modeling GI50

Selecting potent compounds

- **large (many previous studies)**
- **complex: many features**
  (Oprea, Blake, Veber, Veith)
- **local effects (SOM regionalization)**
  - **potency scales with selectivity**
    (Huang et al.)

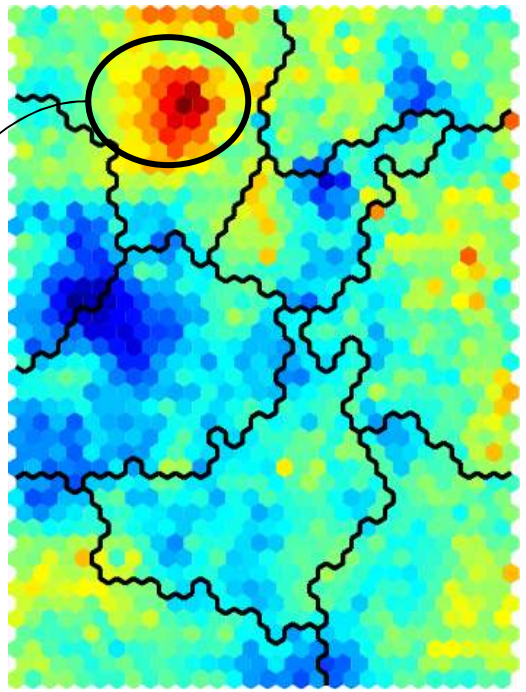
Gene Expression vs GI50



mRNA

"Success"

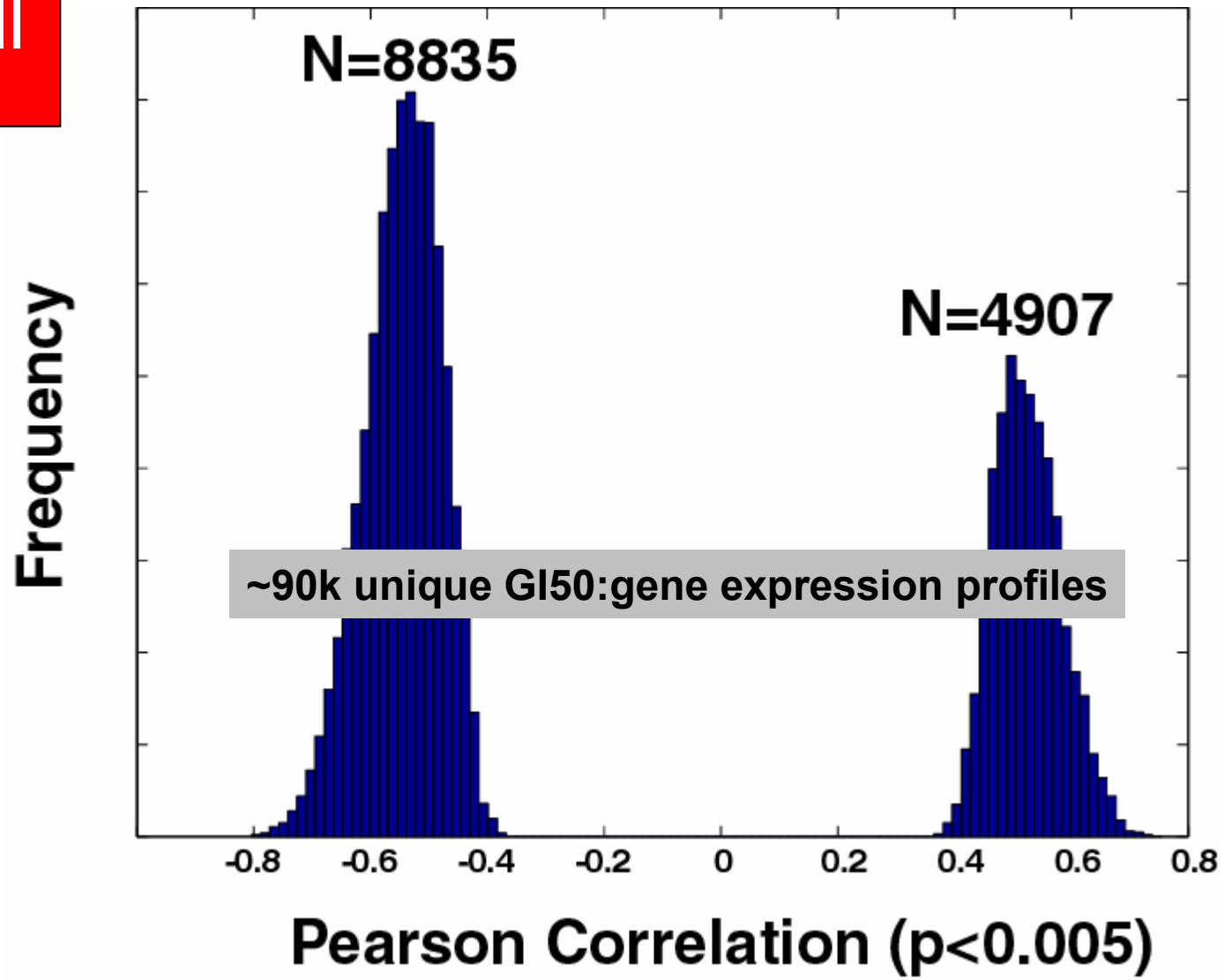
GI50 correlation



6 NSCs selected from highest + correlations
Hypothemycin, LF, PD98059
Rosen et al.
Sellers et al.

Pitfall

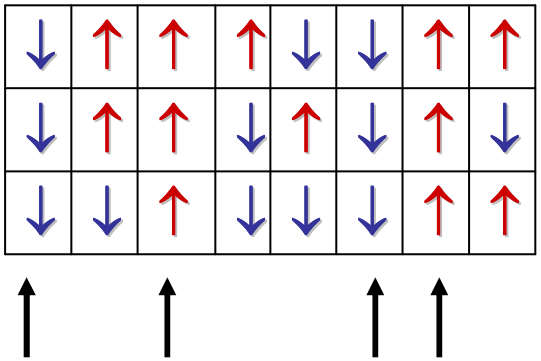
GI50:Gene Expression Correlations



Linking Pathway Gene Expression to GI_{50}

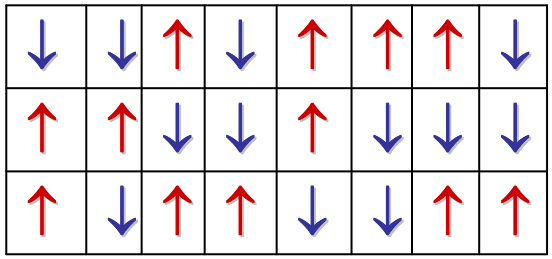
cells

genes



coherent pathway

cells



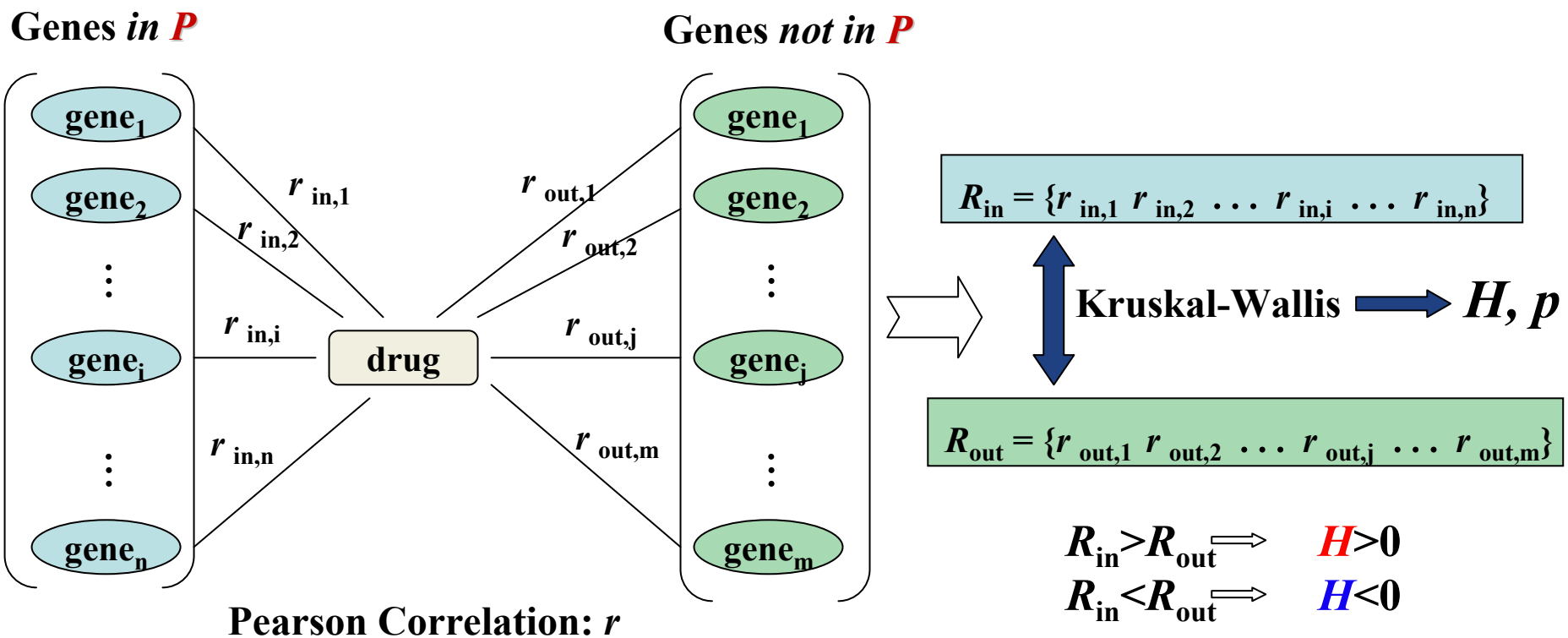
non-coherent pathway

Huang et al. Genomics 87:315-328 (2006)

Linking Pathway Gene Expression to GI_{50}

Huang et al. Genomics 87:315-328 (2006)

For pathway *P*:



Drug is significantly associated with *P* if: $H>0$ and $p<0.05$

***H* defines a Fitness Score for pathways against GI_{50}**

Relating Fitness Scores to Drug Response

Nucleotide sugar metabolism

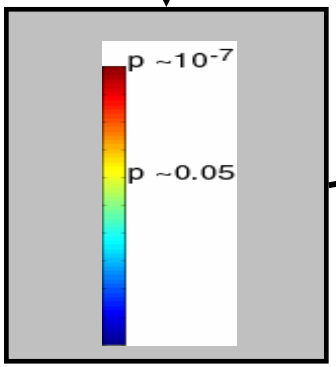
Kegg (hsa00520)

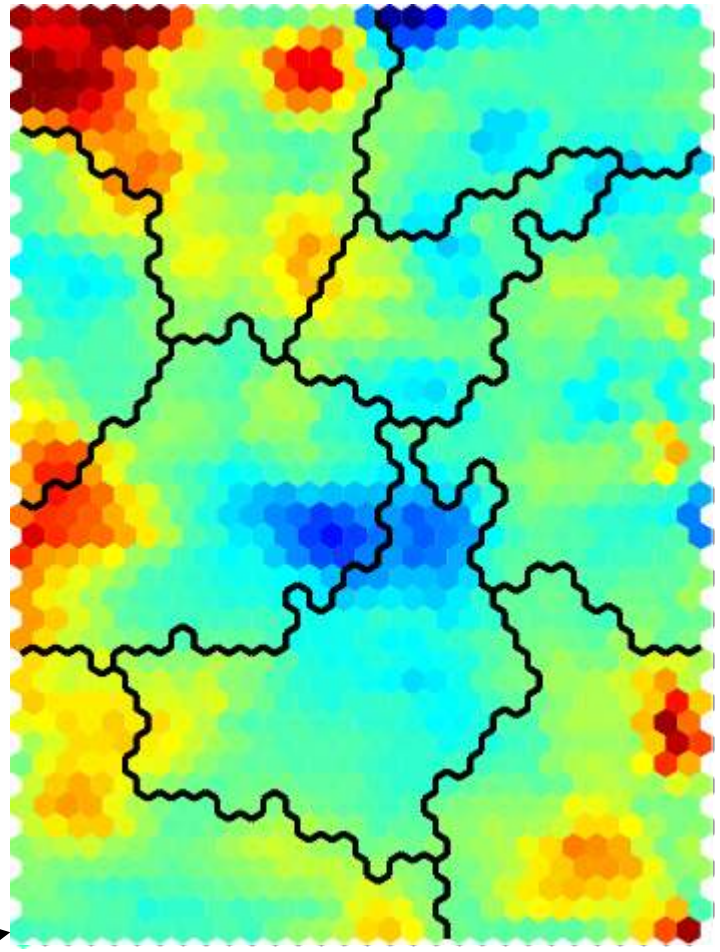
24 member pathway

Pearson Correlation

$mRNA_{pathway}$: $GI50_{node}$

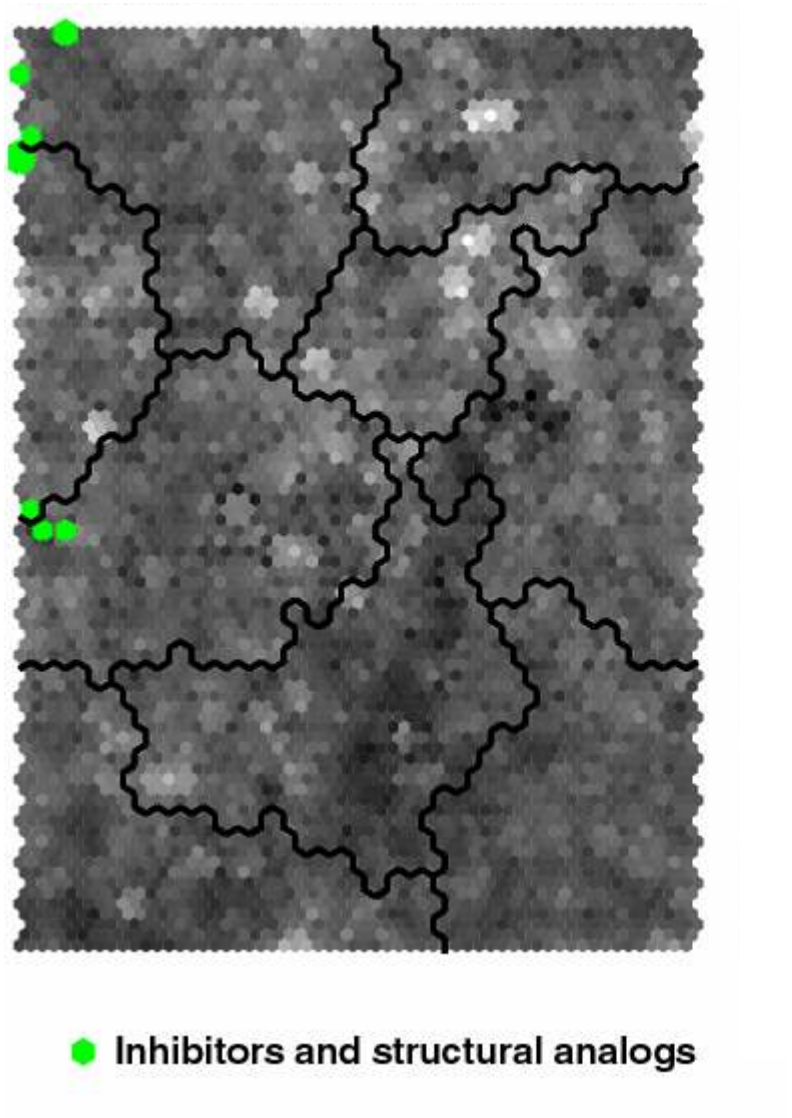
Kruskall Wallis Statistic

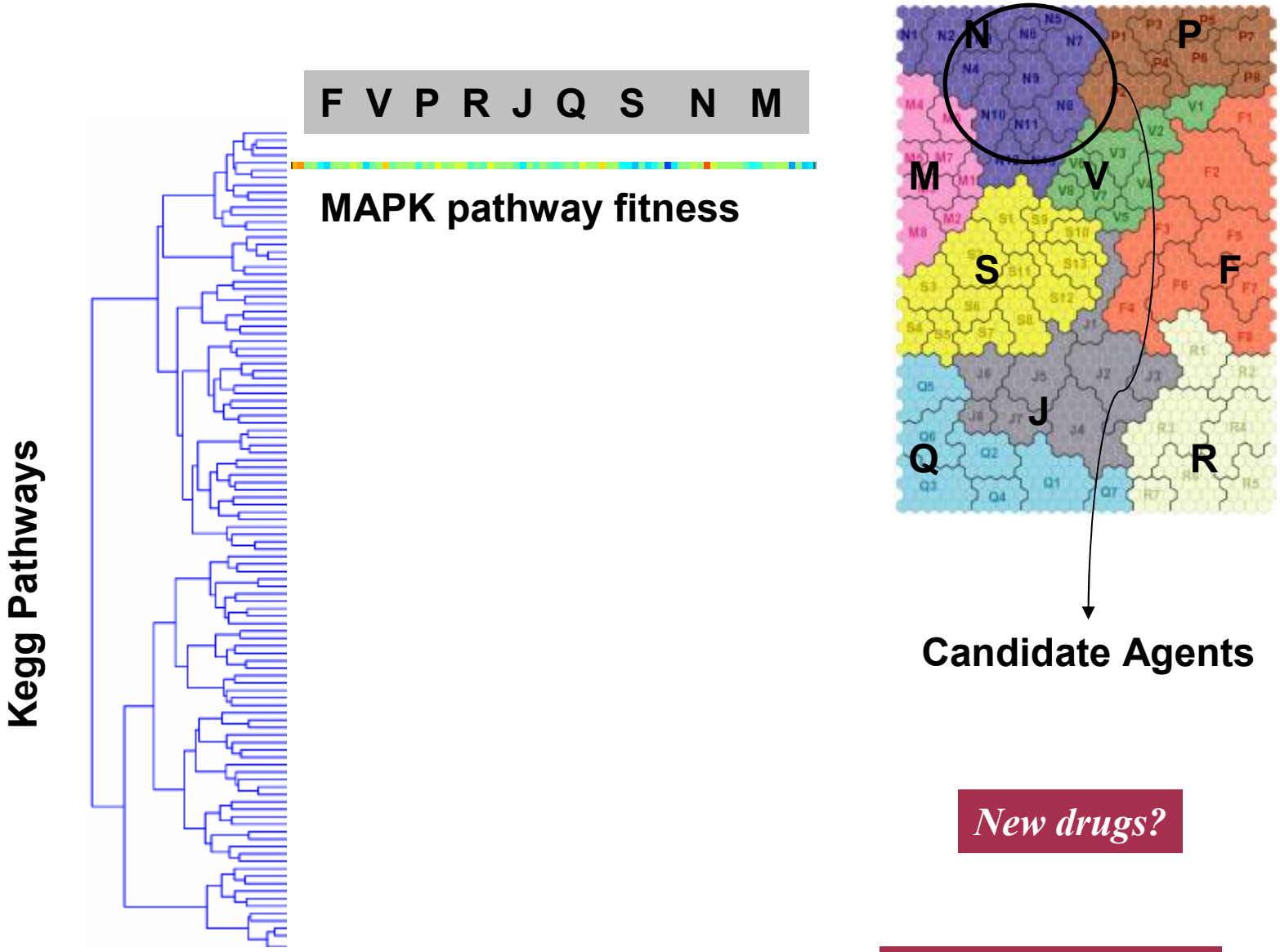


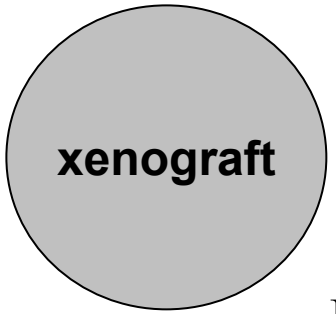


Pathway Fitness (coherence)

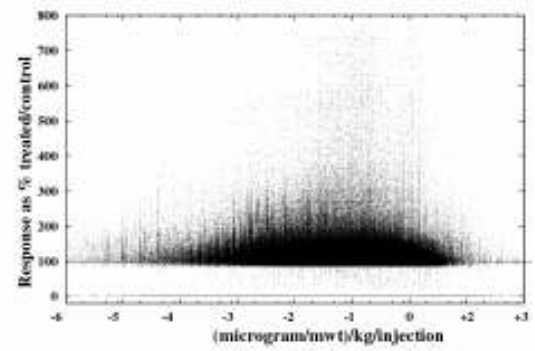
Potential inhibitors of L-asparaginase biosynthesis: Mokotoff JMC, 1981, Richards, Ann Rev, 2006




F V P R J Q S N M
MAPK pathway fitness
Kegg Pathways
N
P
M
V
S
F
J
Q
R
Candidate Agents
New drugs?
New targets/MOA?



Xenograft data

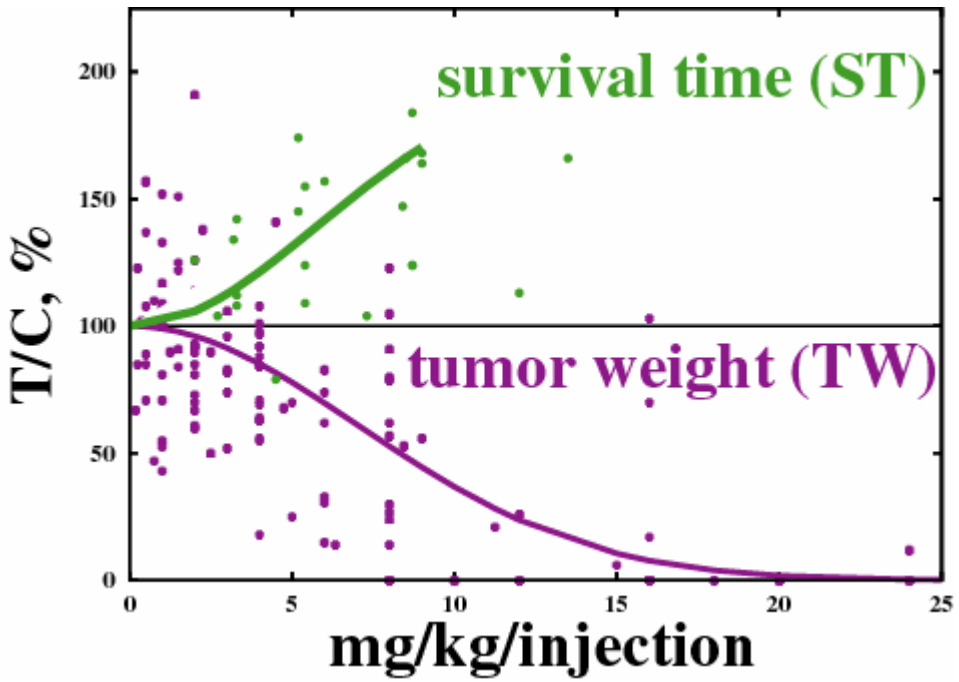


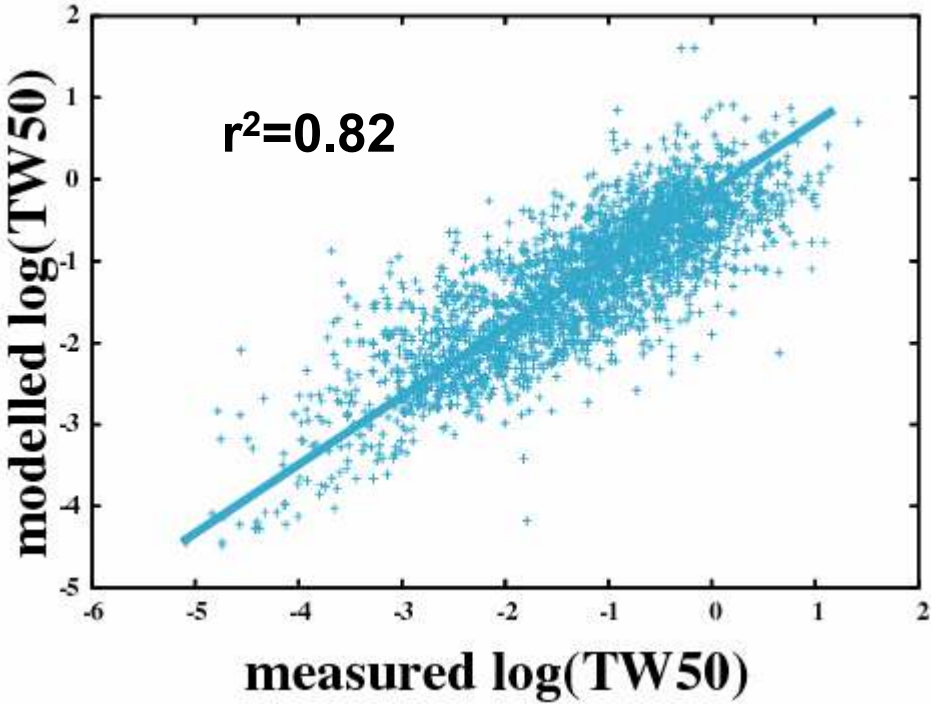
Experimental Design

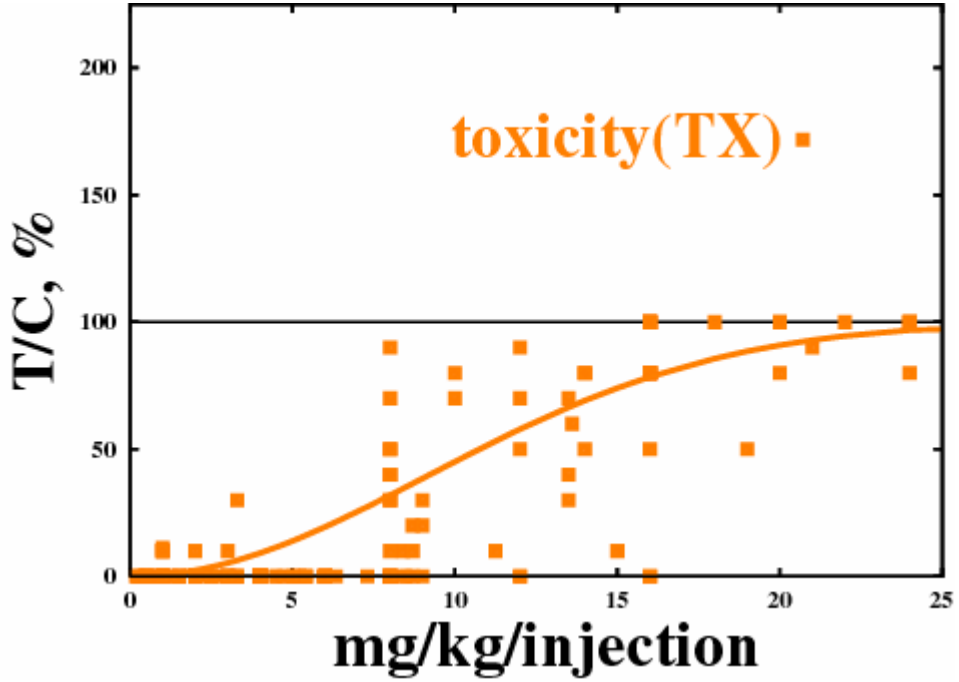
- **1363 NSC tested**
- **31 formulations**
- **187 treatment schedules**
- **50 tumor models**
- **6 implantation sites**
- **15 mice strains**
- **>5,000 combinations of experiments**

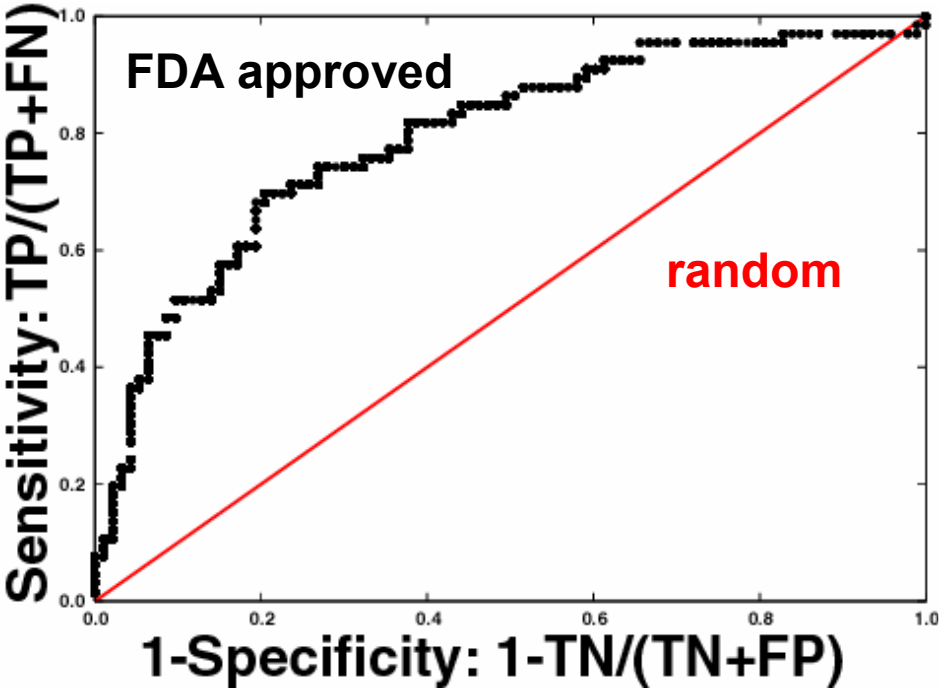
Measurements

- **Tumor weight reduction (TW50)**
- **Survival time (ST150)**
- **Toxicity (survival $_{\text{control vs treament}}$)**
- **Therapeutic index (TW50,ST150/Tox)**


survival time (ST)
tumor weight (TW)
T/C, %
mg/kg/injection

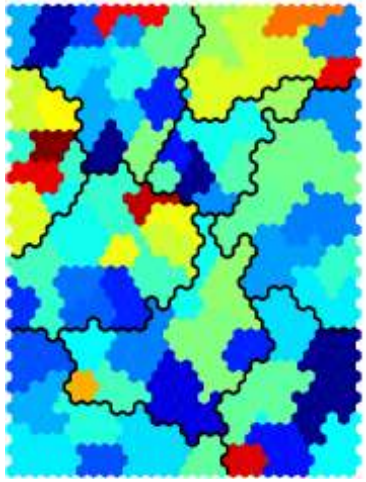

r²=0.82
modelled log(TW50)
measured log(TW50)

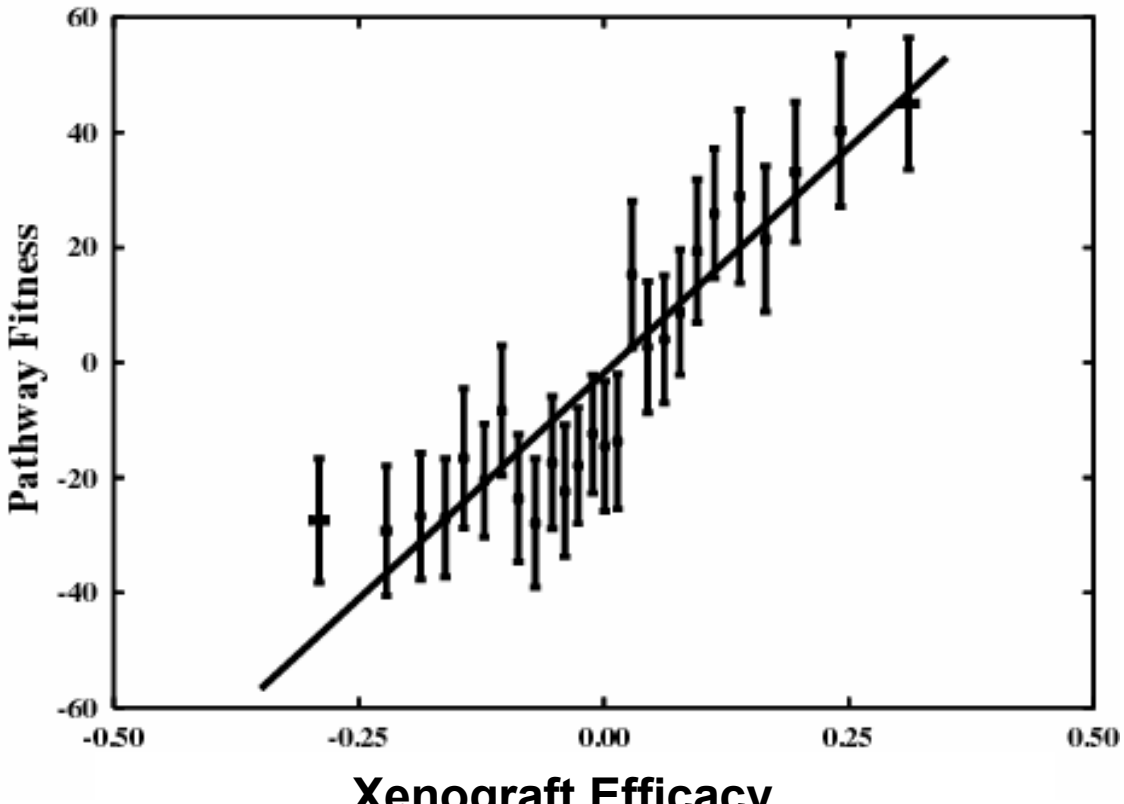

toxicity(TX)
T/C, %
mg/kg/injection


FDA approved
random
Sensitivity: TP/(TP+FN)
1-Specificity: 1-TN/(TN+FP)

Activity Class

Fitness Scores

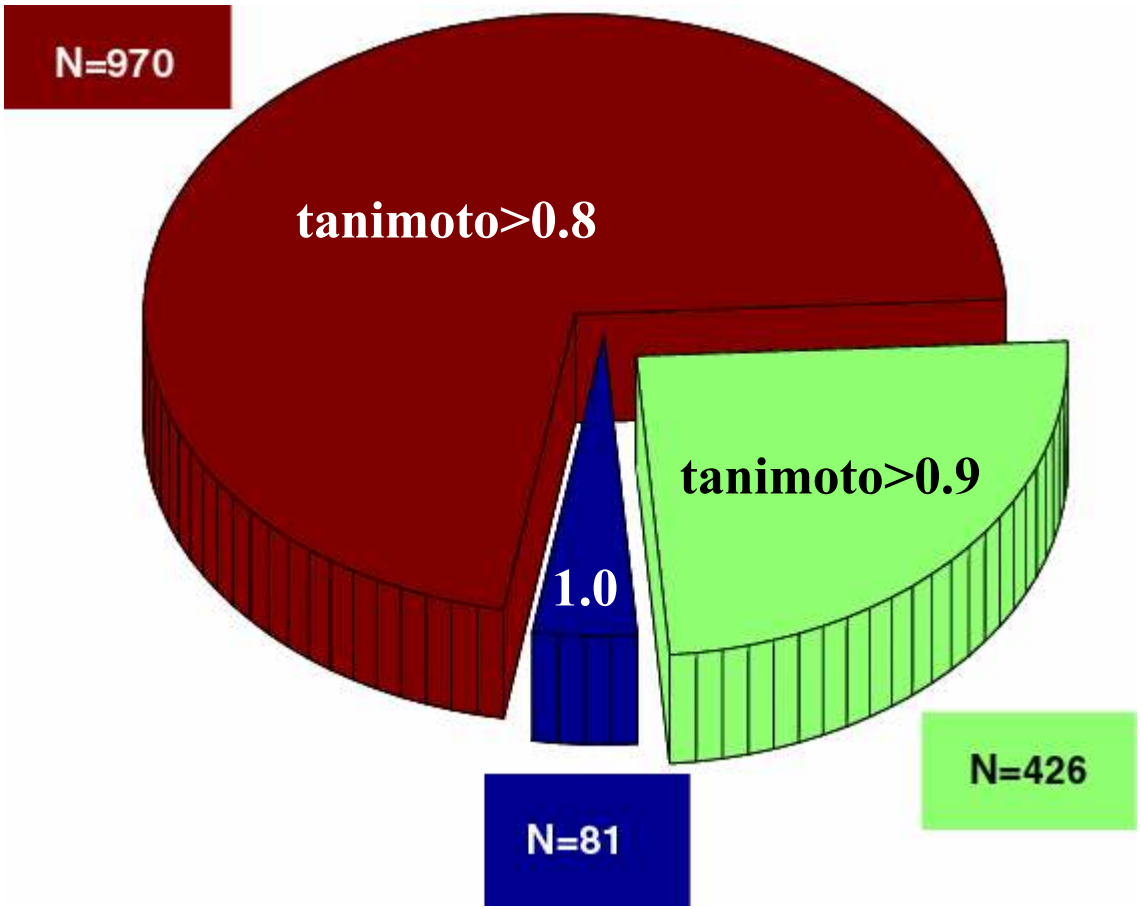




Ind's

399 Anticancer Medicines in Development (283 nonbiologicals)123 (45%) have structural analogs in NCI screening set www.phrma.org

success



Recommendations

Statistics:

- **beyond sorting**
- **clustering**
  - **SOM**
- **decision trees**
  - **random forests**
- **curse of dimensionality**
  - **false positives**
  - **positive predicitive value**

Data Sharing

- **chemistry**
- **gene expression**
- **mutation**
- **SNP**
- **‘cancer genes’**
- **negative results**

Reverse mining

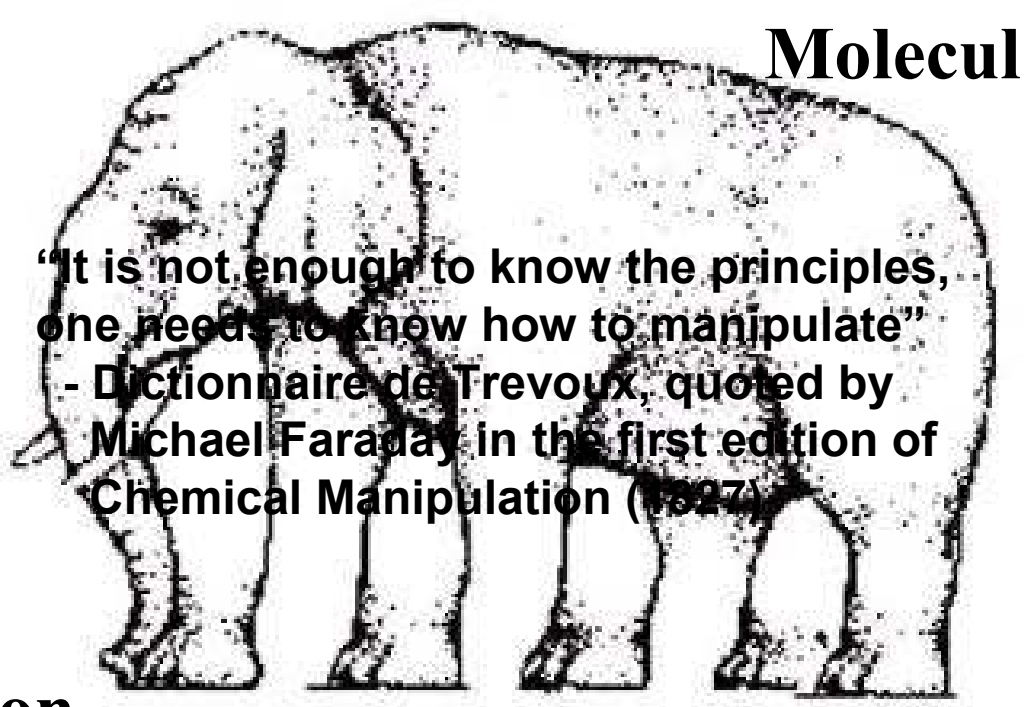
- **retrospective testing**
  - **clinical trials**
  - **preclinincal data**

Cellular growth inhibition

microRNA

Molecular properties

Toxicology



"It is not enough to know the principles, one needs to know how to manipulate"
- Dictionnaire de Trevoux, quoted by Michael Faraday in the first edition of Chemical Manipulation (1827)

Proteomics

Clinical trials

Xenografts

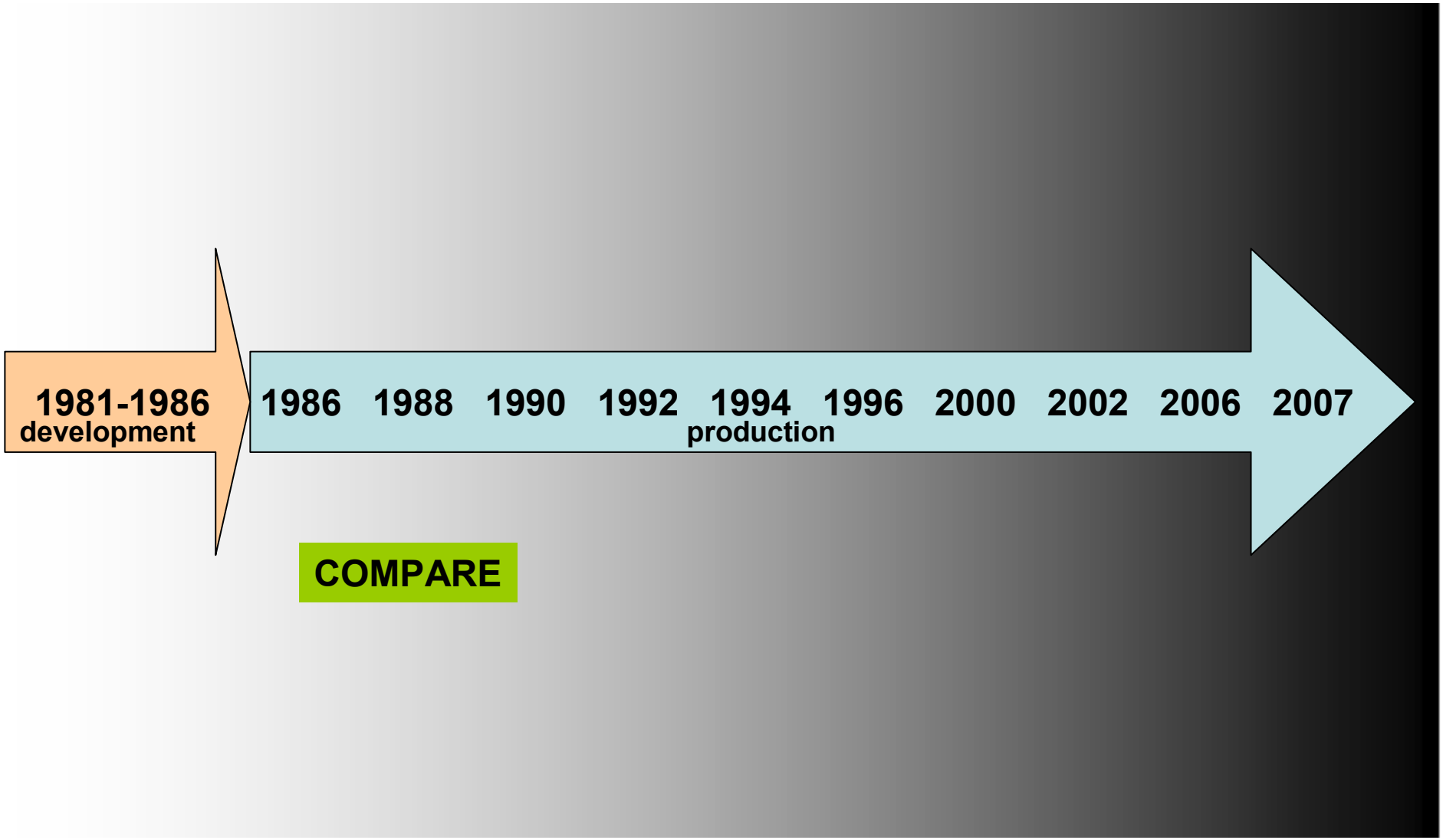
Gene expression

Karyotype

SNP copy number

Methylation status

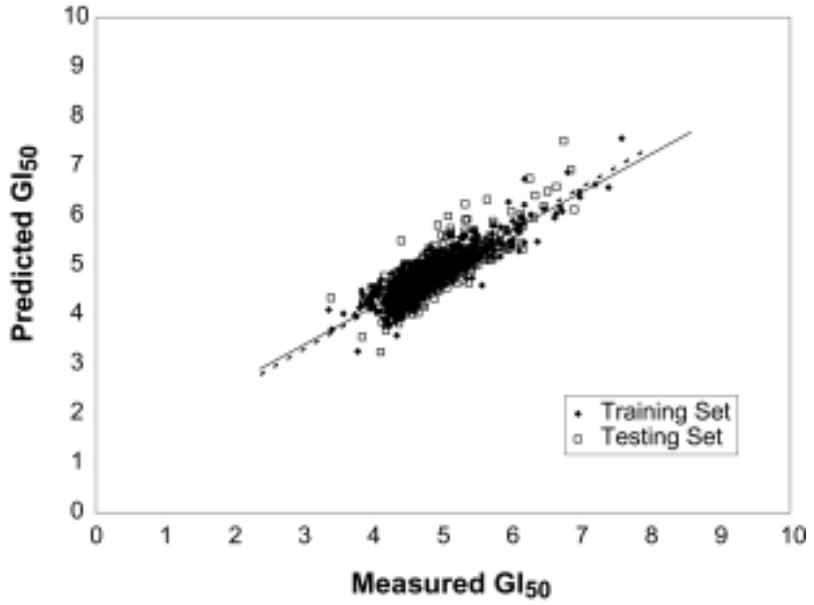
NCI-60 Timeline
Shoemaker, Nat. Rev. Cancer, 2006



Chemistries: Modeling GI50

$$GI_{50} = F_1(\text{properties}) = c_1x_1 + c_2x_2 + \ldots + c_Nx_N$$

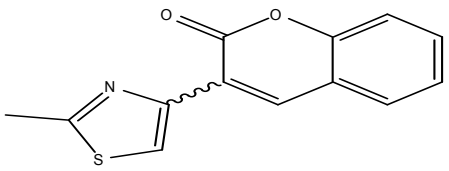
AlogP Mwt LSUFC

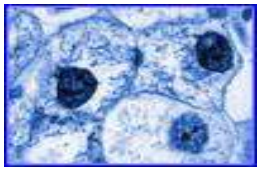


Training: $r^2 = 0.77$
Testing: $r^2 = 0.67$

Xenograft data

Rx



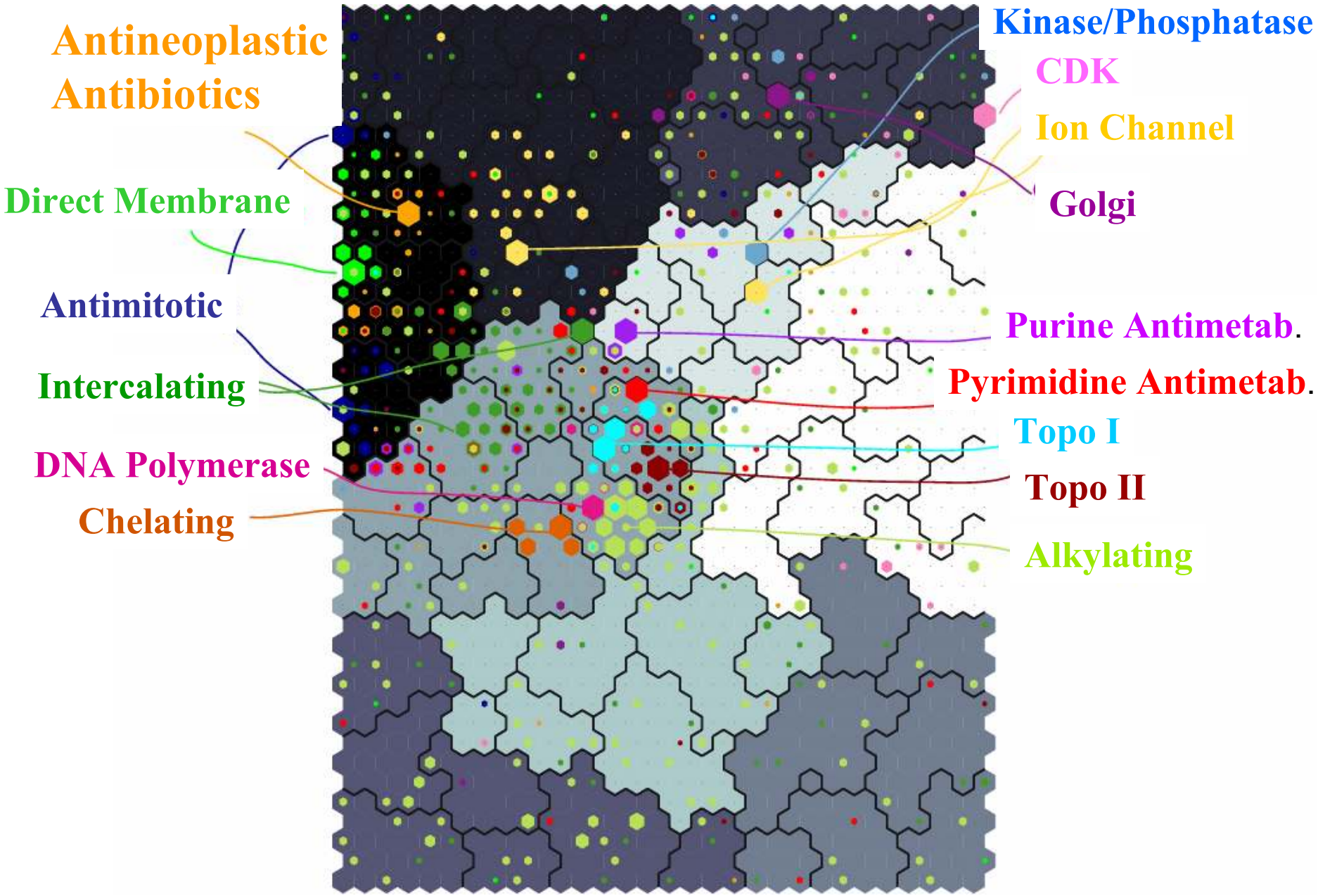


outcome = **B** × [(**treatment**) (**chemistry**)(**cellular growth inhibition**)]

exptl design **properties** **GI50**

Treatment variations alone account for a log order of difference in efficacy outcome

Molecular Classes



	Act 1	Act 2	Act 3	Interm.	Act 4
GI50	✓	✓			
Chemistry		✓			
mRNA					
Pathways					
Xenograft					

	Act 1	Act 2	Act 3	Interm.	Act 4
GI50	✓	✓	✓	✓	✓
Chemistry		✓	✓		✓
mRNA			✓	✓	✓
Pathways				✓	✓
Xenograft					✓

	Act 1	Act 2	Act 3	Interm.	Act 4
GI50	✓	✓	✓		
Chemistry		✓	✓		
mRNA			✓		
Pathways					
Xenograft					

Chemistry Meets Biology

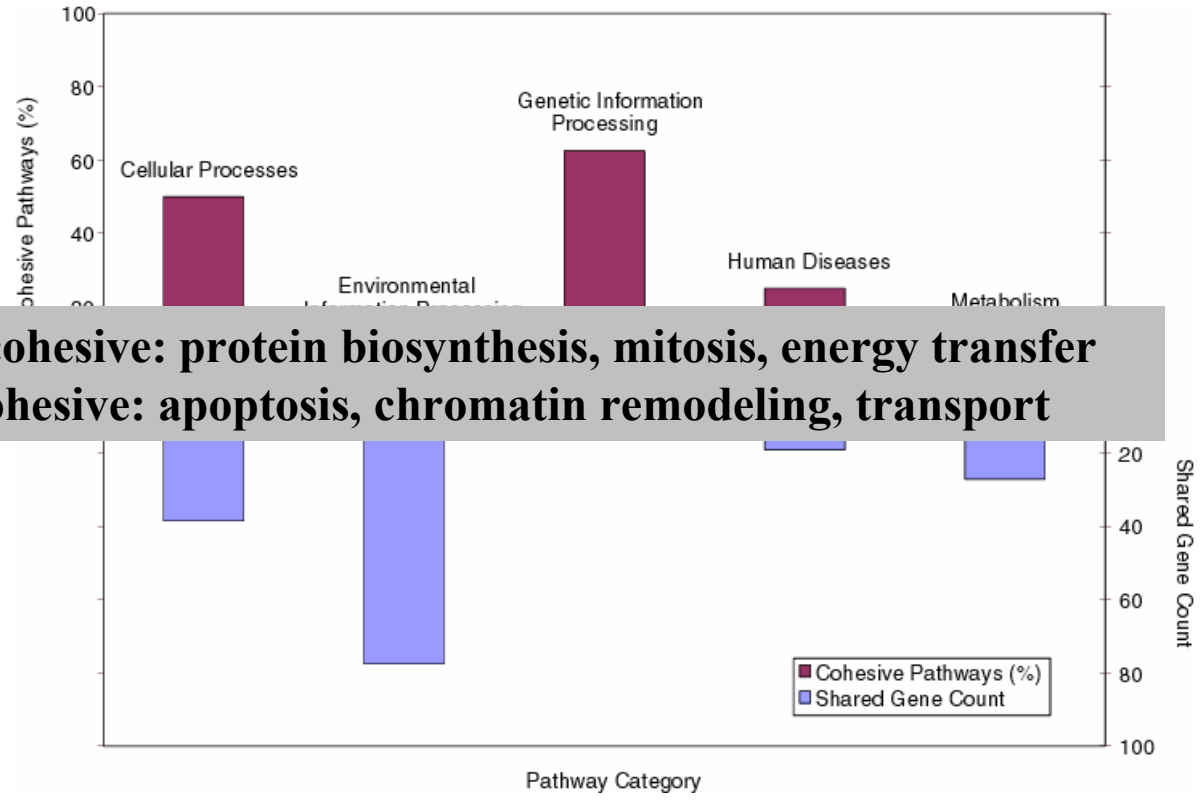
	Act 1	Act 2	Act 3	Interm.	Act 4
GI50	✓	✓	✓	✓	
Chemistry		✓	✓		
mRNA			✓	✓	
Pathways				✓	
Xenograft					

Pathway Fitness - Cohesiveness

- **Relationship between the number of genes in a pathway that are shared with other pathways and the cohesiveness of the pathway**

– **Genetic Information Processing**
  - highest percentage of cohesive pathways
  - least number of shared genes

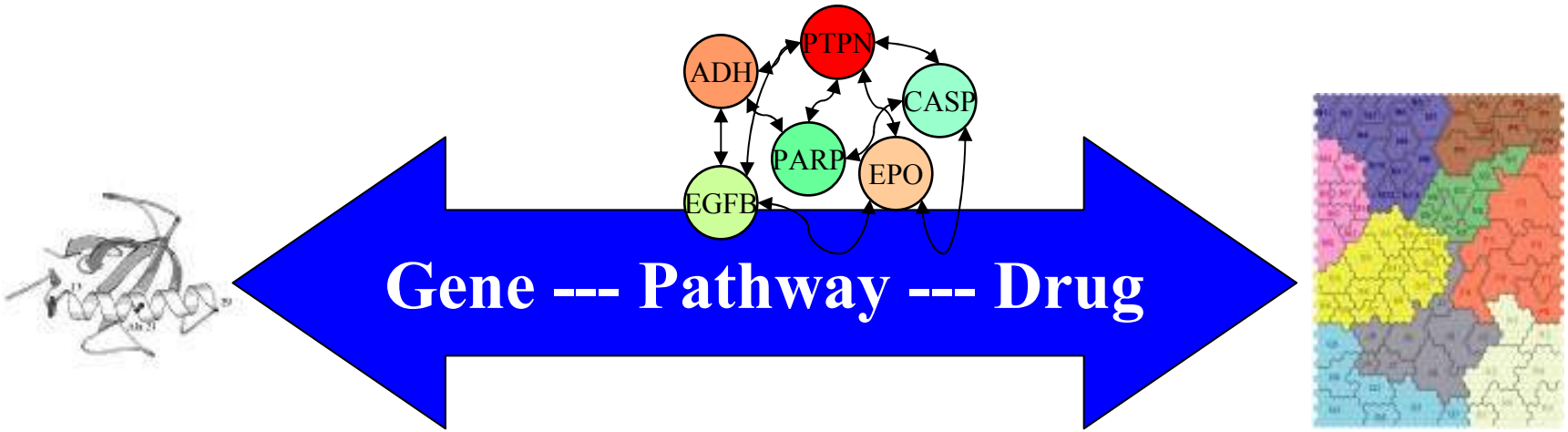
– **Environmental Information Processing**
  - lowest percentage of cohesive pathways
  - largest number of shared genes

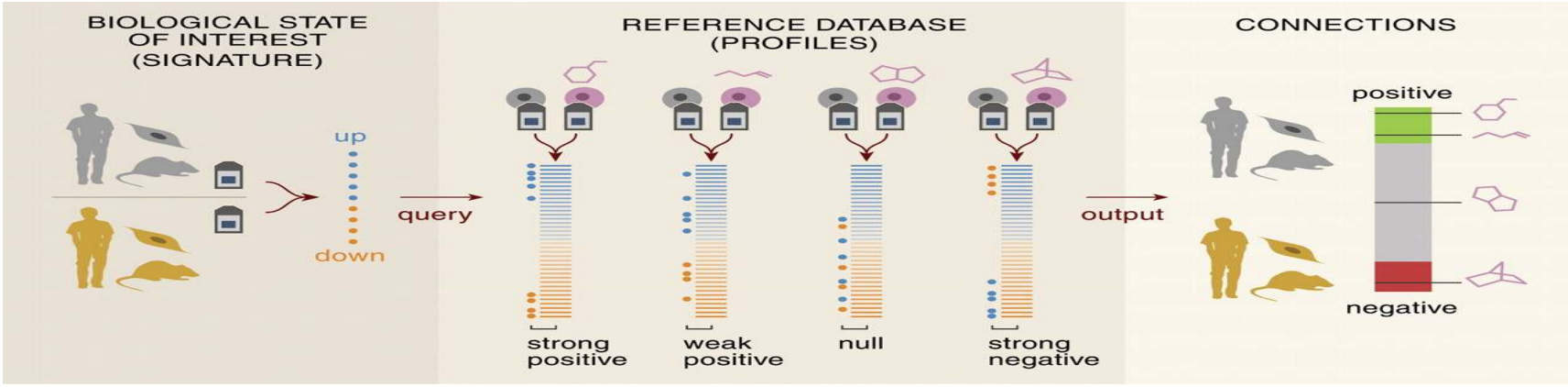


More cohesive: protein biosynthesis, mitosis, energy transfer
Less cohesive: apoptosis, chromatin remodeling, transport

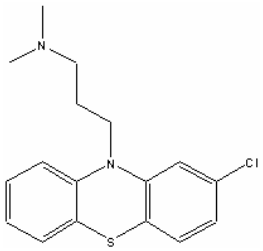
Huang et al. Genomics (2006)

Huang et al. Mol. Cancer Therapeutics (2006)

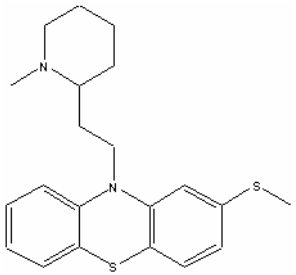




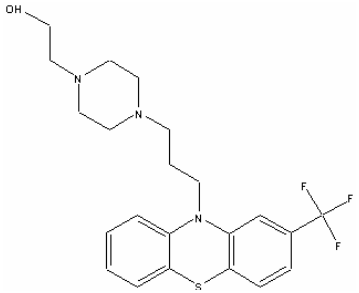
Connectivity Maps
Lamb et al., 2006



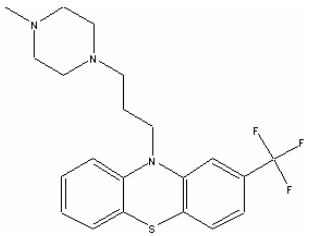
chlorpromazine



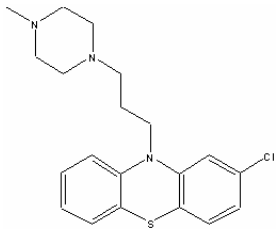
thioridazine



fluphenazine



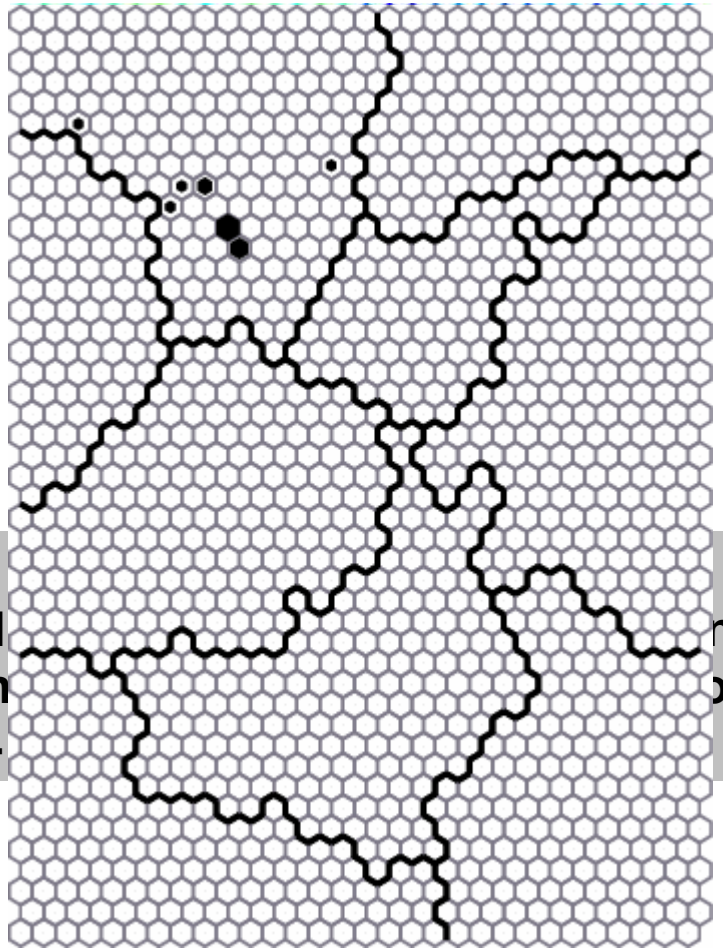
trifluoperazine

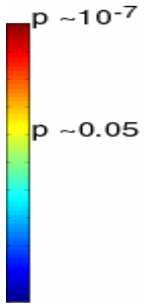


prochlorperazine

GO:3707 Steroid hormone receptor activity (PPARG, RXR, ESRR)

GO:199992 Diacylglycerol Binding (DAK, PKC)

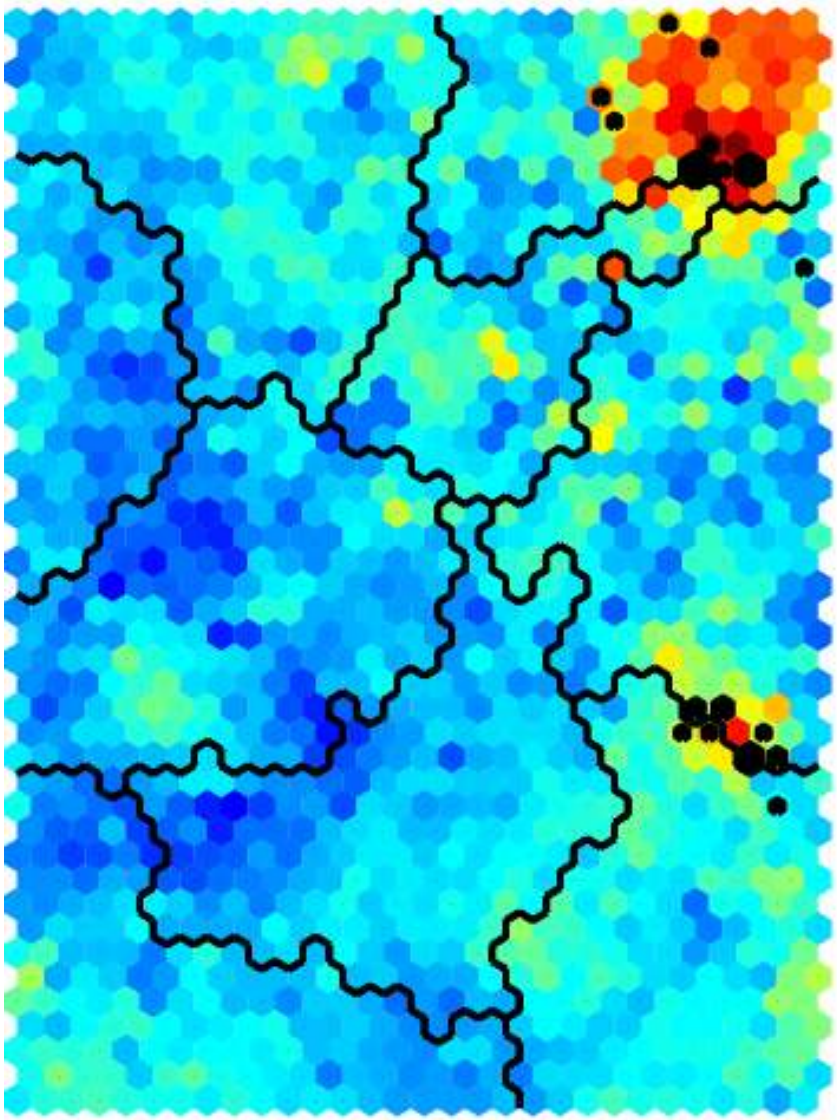




Pathway Fitness (coherence)

PPARgamma agonists amel n via inhibition of diacylglycerol-protein kin of diacylglycerol kinase. Verrier et al. Circ. Res, 2004

Rapamycin Family



h_vegf
h_ires
h_ran
GO:74 cell cycle
GO:1525 angiogenesis
GO:6099 tca cycle

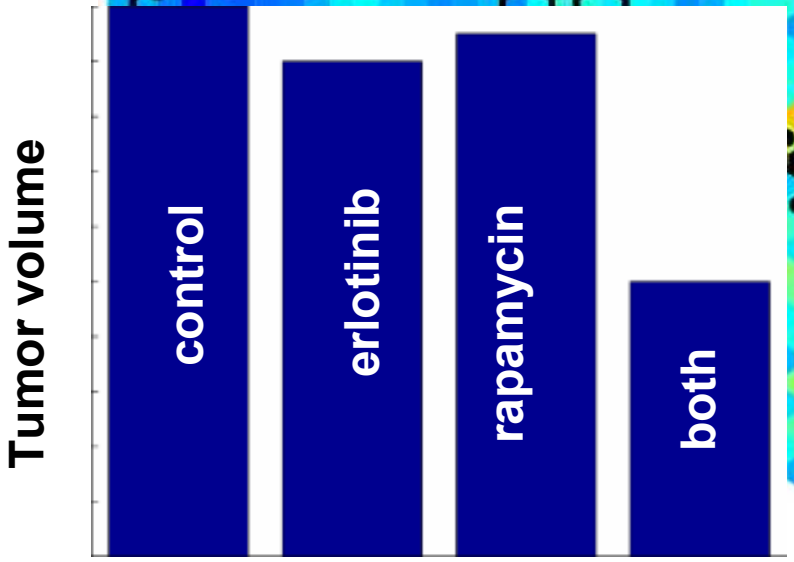
h_ctc
GO:3724 RNA helicase
GO:6631 fatty acid metabolism

- Structurally similar to temsirolimus (N=24)
  179 gene espressions are correlated with these 24 NSCs

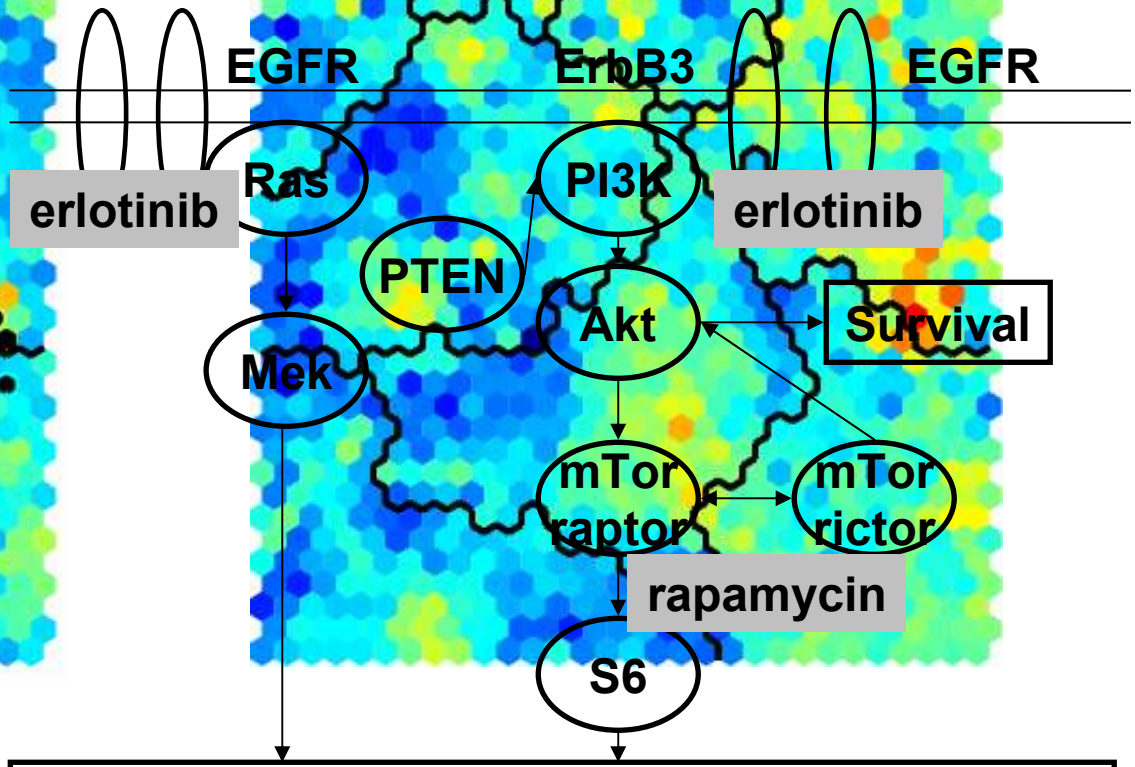
Rapamycin

Erlotinib

Rapamycin synergizes with epidermal growth factor receptor inhibitor Erlotinib in non-small-cell lung, pancreatic, colon and breast tumors. Buck et al. Mol Cancer Therapeutics, Nov. 2006



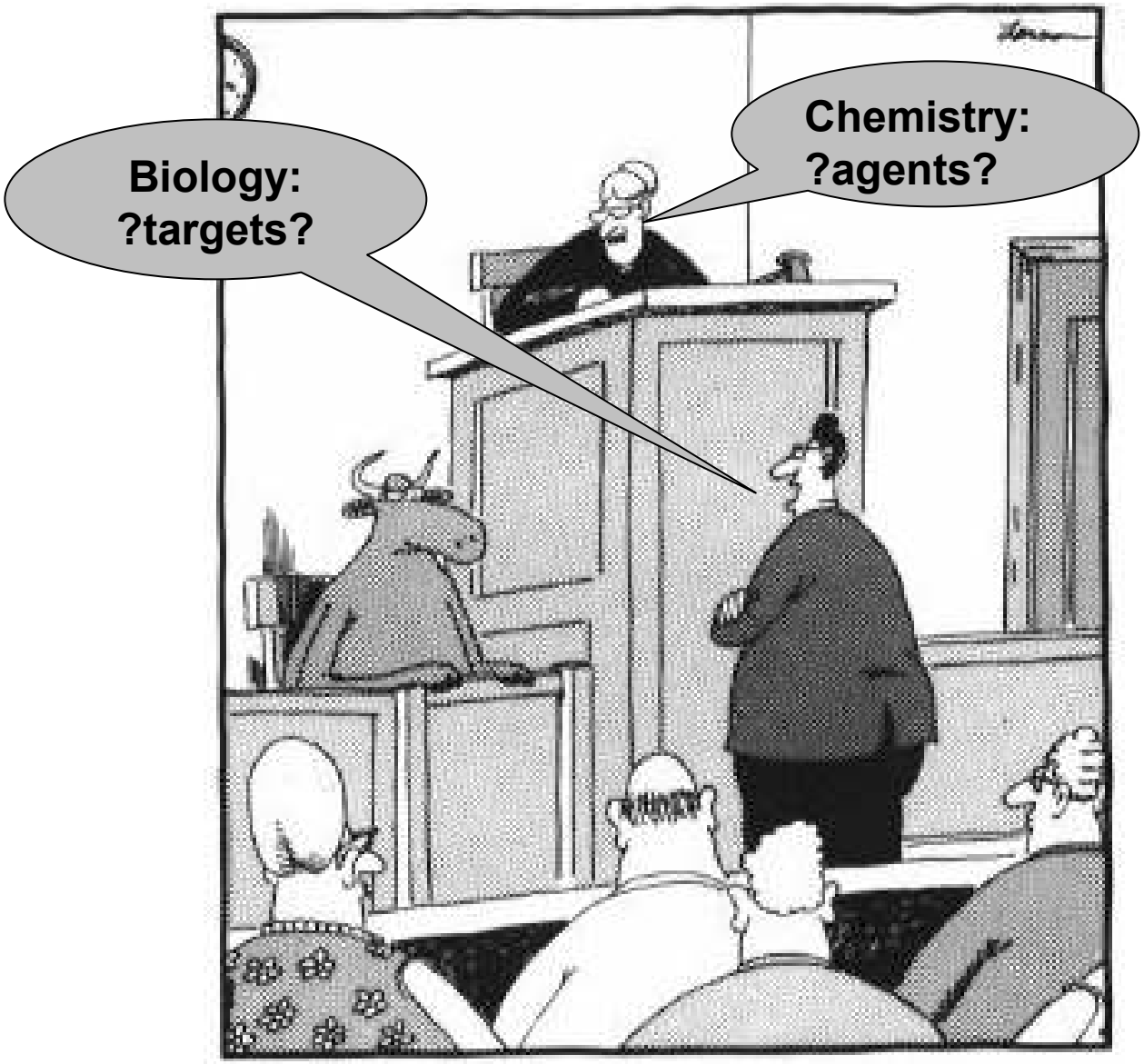
Combination effects



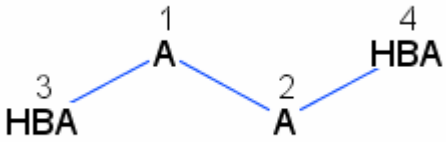
Proliferation/Cell Cycle Progression

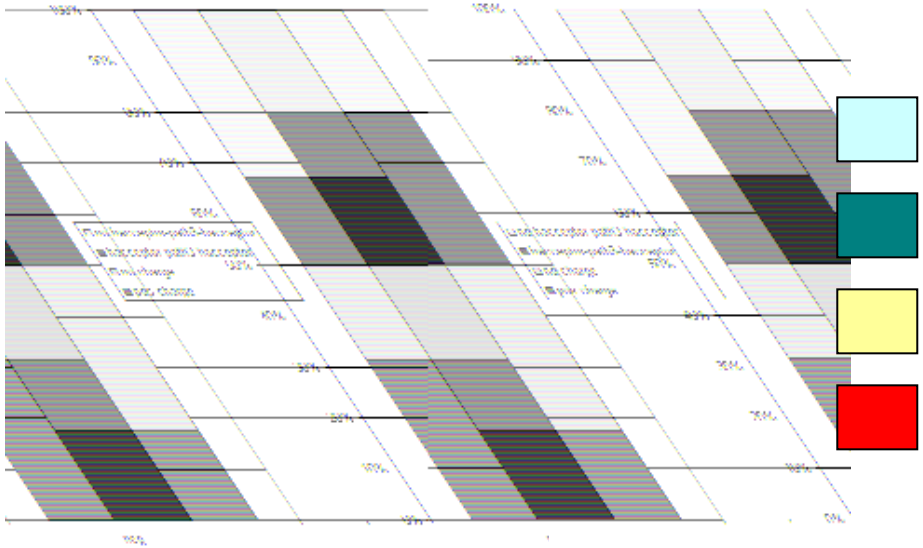
Chemistry Meets Biology

	Act 1	Act 2	Act 3	Interm.	Act 4
GI50	✓	✓	✓	✓	✓
Chemistry		✓	✓		✓
mRNA			✓	✓	
Pathways				✓	✓
Xenograft					✓



Look. We know *<u>that</u>* it works ---- *<u>that</u>* is no longer the question. What we now want to know is *<u>how</u>*… <u>How</u> now brown cow?"

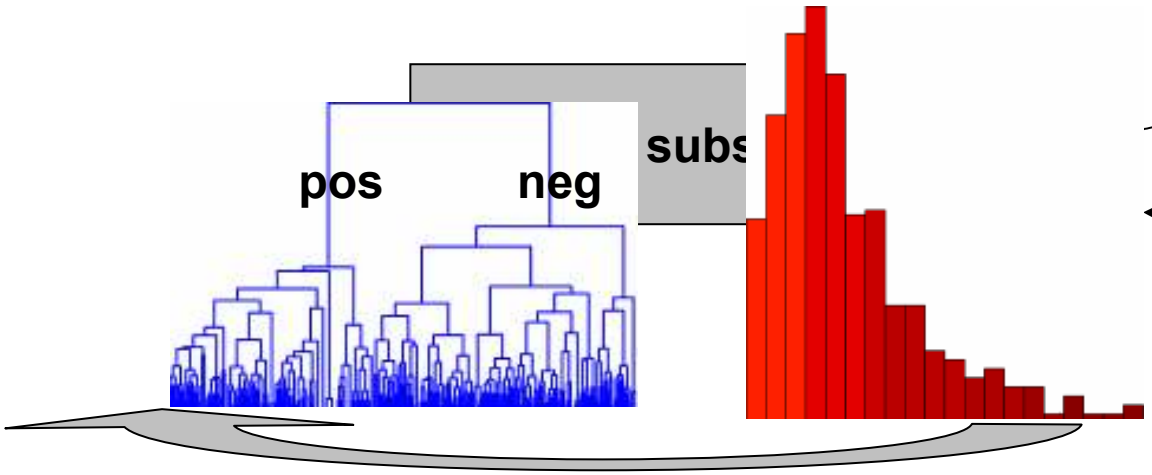


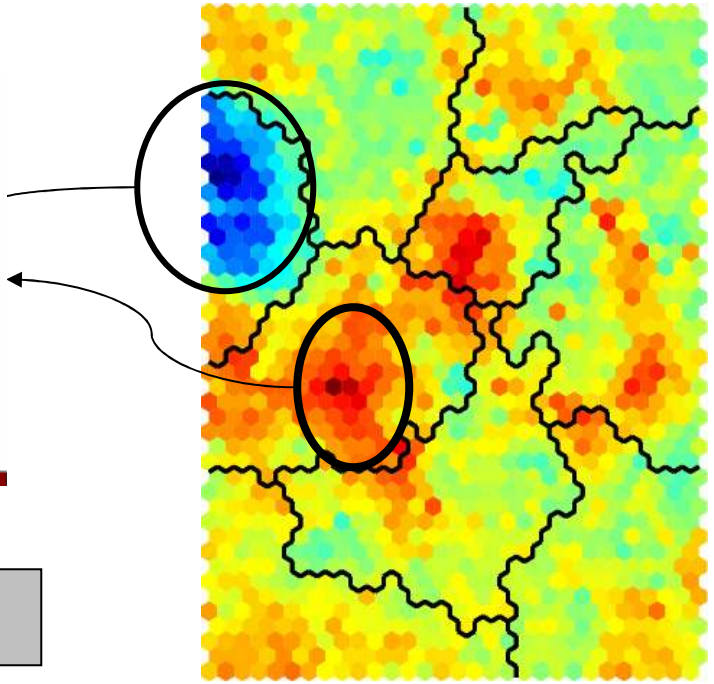


H-acceptor path3

+ charge

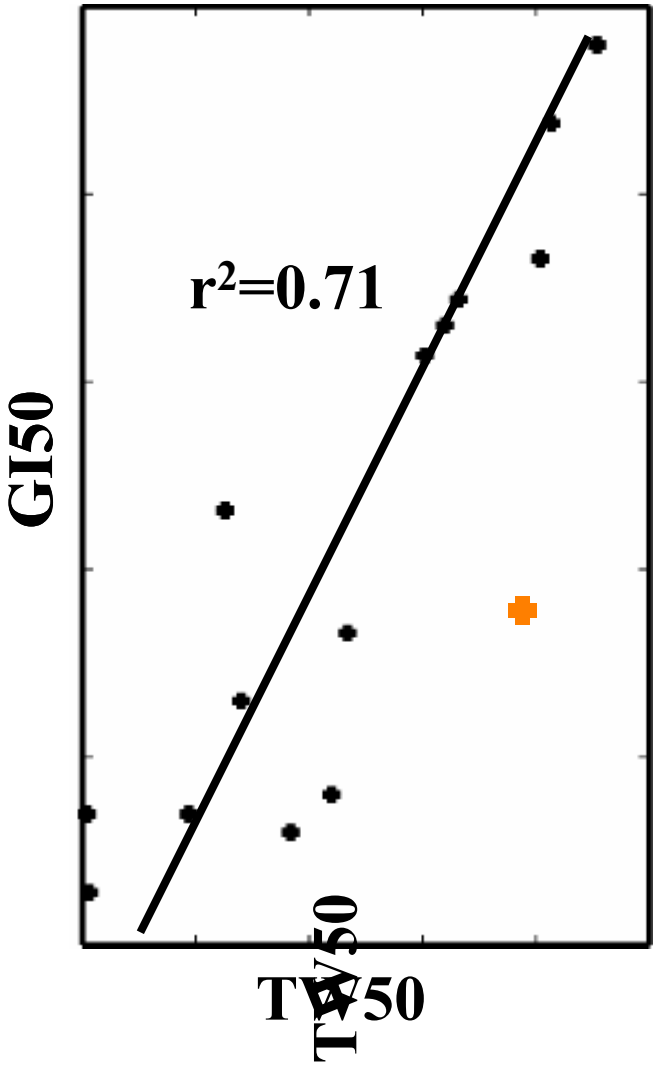
POS NEG

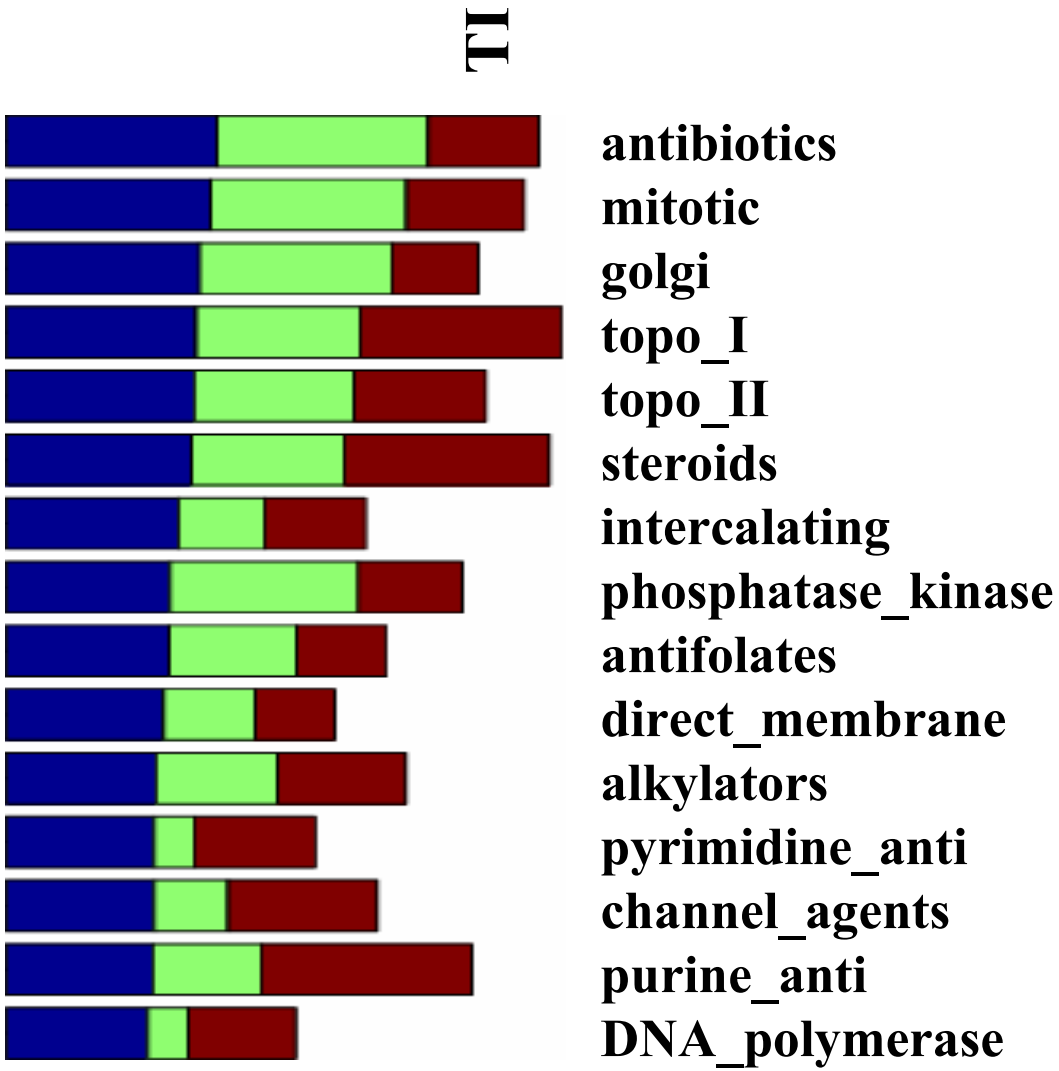




Thiosemicarbazone NSC73306

Xenograft Data





Phosphatase_kinase agents produce near maximal tumor weight reduction for modest values in GI50 and Therapeutic Index