Gene Expression Profiles:

Biomarkers of Inter-Individual Susceptibility to Environmental Agents and Indicators of Exposure

Rebecca Fry, Ph.D.

Environmental exposures to potentially harmful environmental agents

Harmful agents

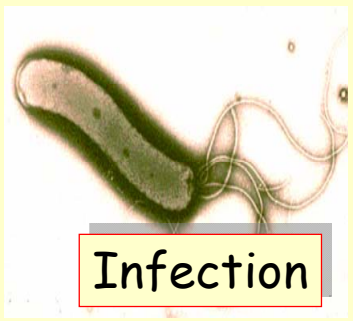
Eat

Medicine

Drink

Absorbed

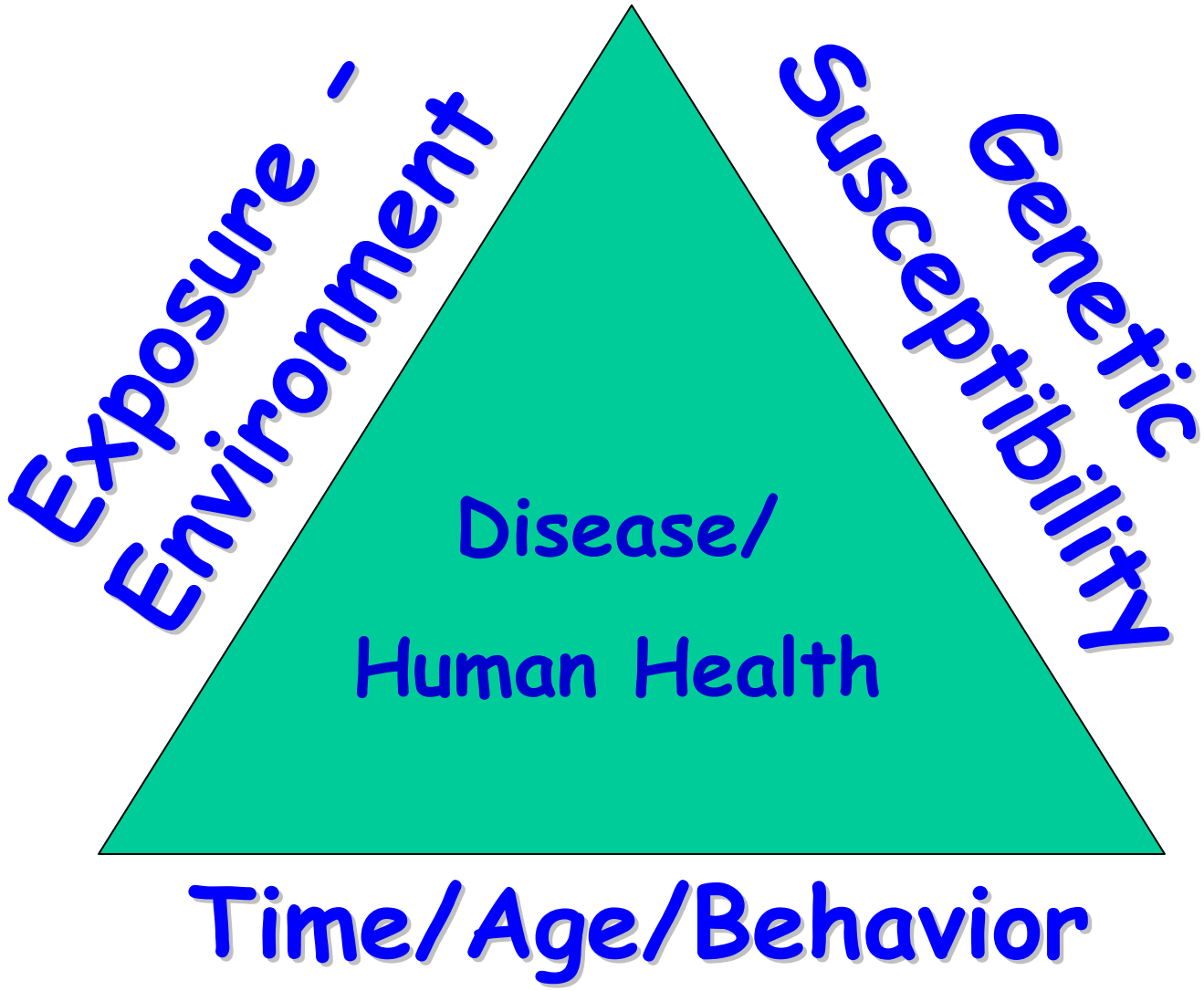
Breathe


Infection

People have different exposures

People have different responses

Complex Gene-Environment Interactions Influence Human Health



Our Mission

Explore biological effects of exposure to environmental agents:

Understand impact on human health:
Molecular basis for disease/carcinogenesis

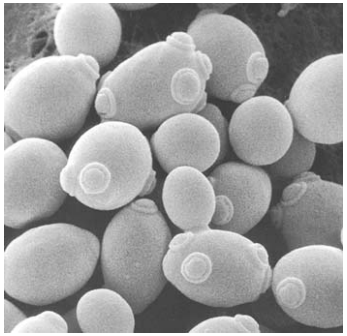
Develop Biomarkers
Biomarkers of population exposure, biomarkers of disease state

Predict inter-individual differences in susceptibility to disease

Prevent detrimental health effects from exposure

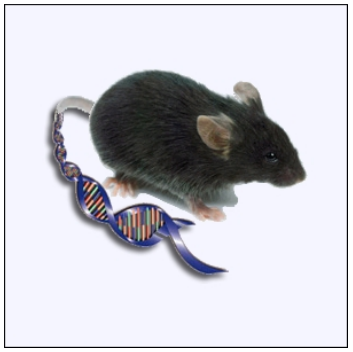
Systems and Tools

Human Populations
Human Cell lines

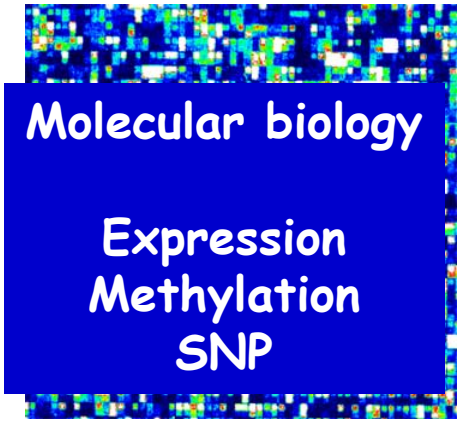


Yeast

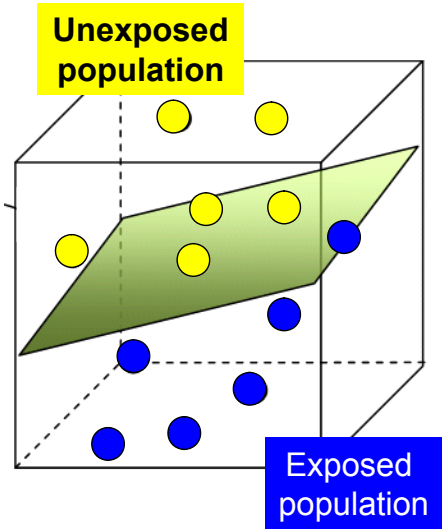
C. elegans



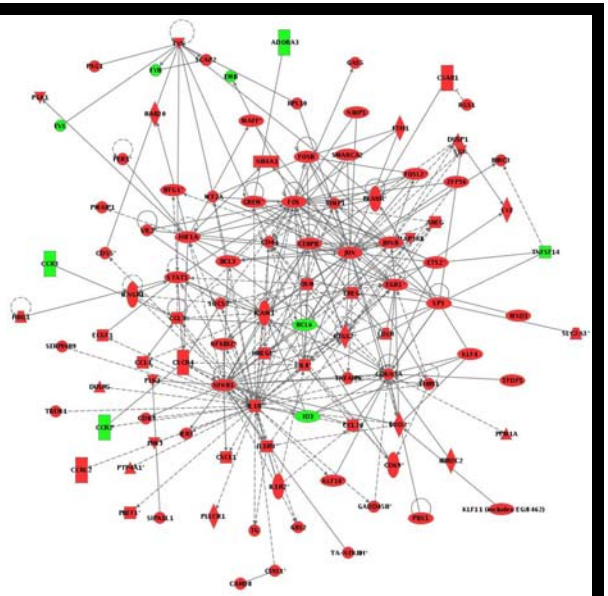
Mice



Genomics



Computational Biology



Systems Biology

Using genomics to predict and classify population responses to exposures

Classify: Who has been exposed to a damaging agent?

Using genomics to predict and classify population responses to exposures

Classify: Who has been exposed to a damaging agent?

Can we identify genetic biomarkers of exposure

Arsenic:

Inorganic arsenic is a ubiquitous environmental pollutant and known human carcinogen

SCIENCE VOL 315 23 MARCH 2007

A Sluggish Response to Humanity's Biggest Mass Poisoning

Chronic exposure results in many cancers:
skin, bladder, lung, liver, prostate and kidney

1987: Classified as Group 1 Carcinogen by International Agency for Research on Cancer (IARC)

1994: W.H.O. Recommended Guideline Value of 10 µg/L arsenic in drinking water

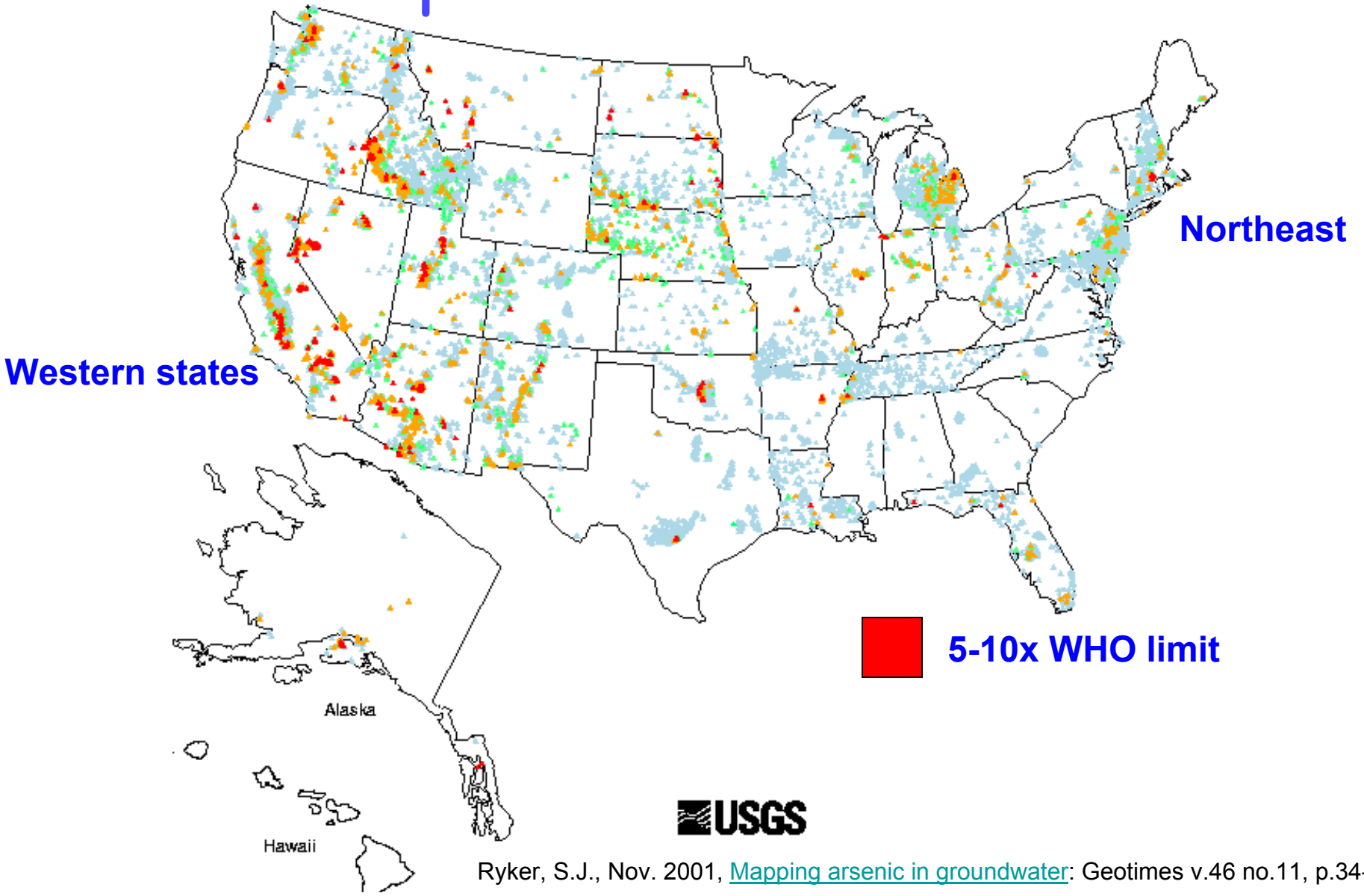
In Bangladesh ~30 million exposed to levels that FAR exceed WHO limit

Up to 100x

SCIENCE VOL 315 23 MARCH 2007

A Sluggish Response to Humanity's Biggest Mass Poisoning

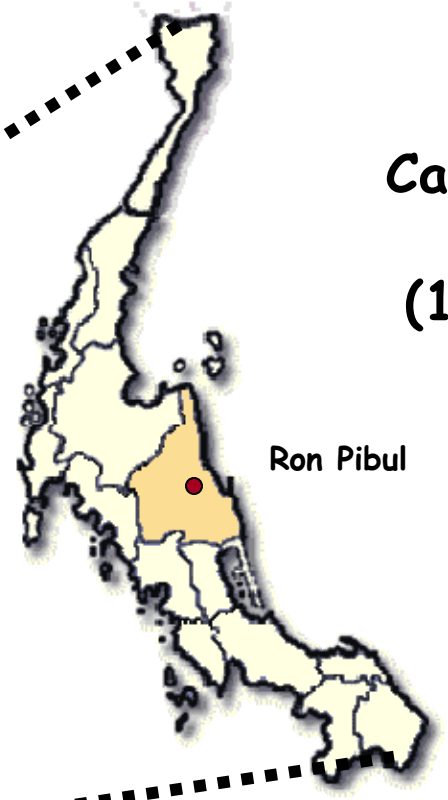
Arsenic contamination is ALSO a problem in USA



Ryker, S.J., Nov. 2001, Mapping arsenic in groundwater: Geotimes v.46 no.11, p.34-36.

Study site: Ron Pibul District, Thailand

Nakhon Si Thammarat Province



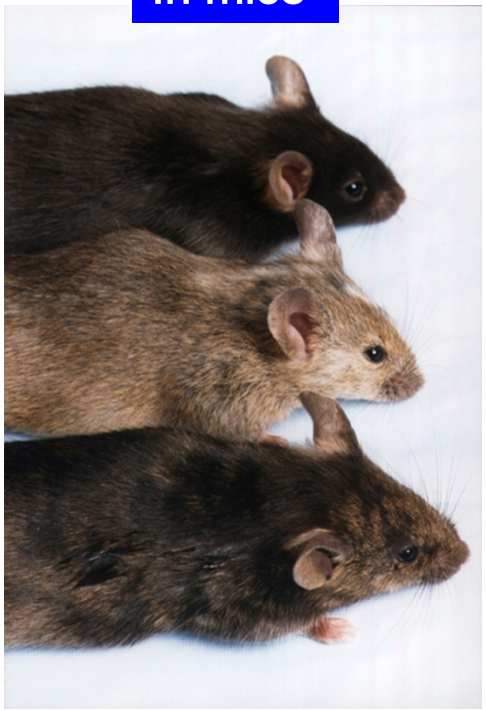
Caused by 30 years of tin-mining (1950's to 1980's)

up to 100X WHO limit (10ppb)

~Bangladesh exposures

In utero arsenic exposure in rodents – recent alarming findings

In mice



exposure to arsenic during gestation results in 5-fold increase in hepatocellular carcinomas

Gene expression changes in livers of offspring exposed to arsenic *in utero* when reach adulthood

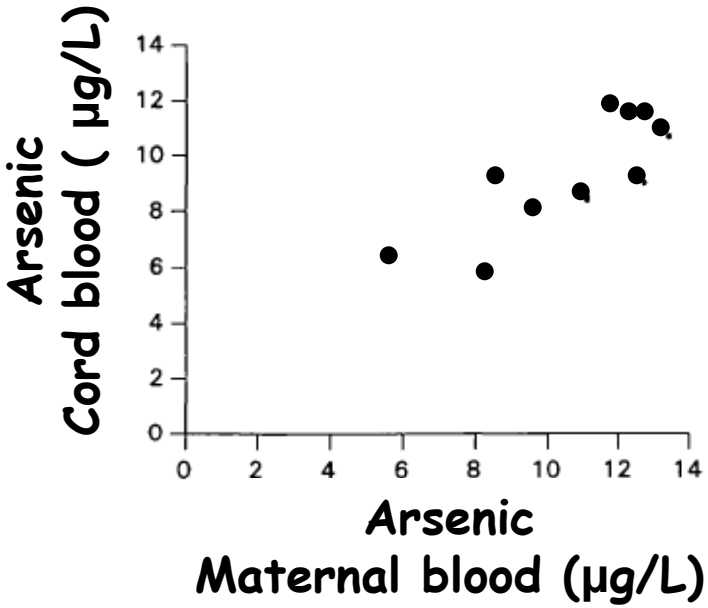
Waalkes et al., Tox Sci 2003

Gene expression changes evident in offspring when reach adulthood..
Could this be epigenetic reprogramming of gene expression?

Prenatal arsenic exposure in humans

Arsenic crosses the placenta in humans

Concha et al, Tox Sci 1998



Long term health effects

Increased mortality from
lung cancer and
liver cancer from
prenatal and early
childhood arsenic
exposures

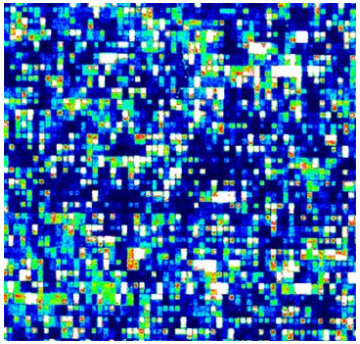
Smith et al, EHP 2006
Cancer Epidemiology
Biomarkers & Prevention
2008.

Expression profiling of blood from 32 newborns from Thailand whose mothers were exposed to varying levels of arsenic

Cord blood from 32 newborns

Gene Expression Profiling and Pathway mapping

Affymetrix HG-U133 Plus 2.0 Full Genome, Duplicate



What is the genome-wide impact of prenatal exposure??

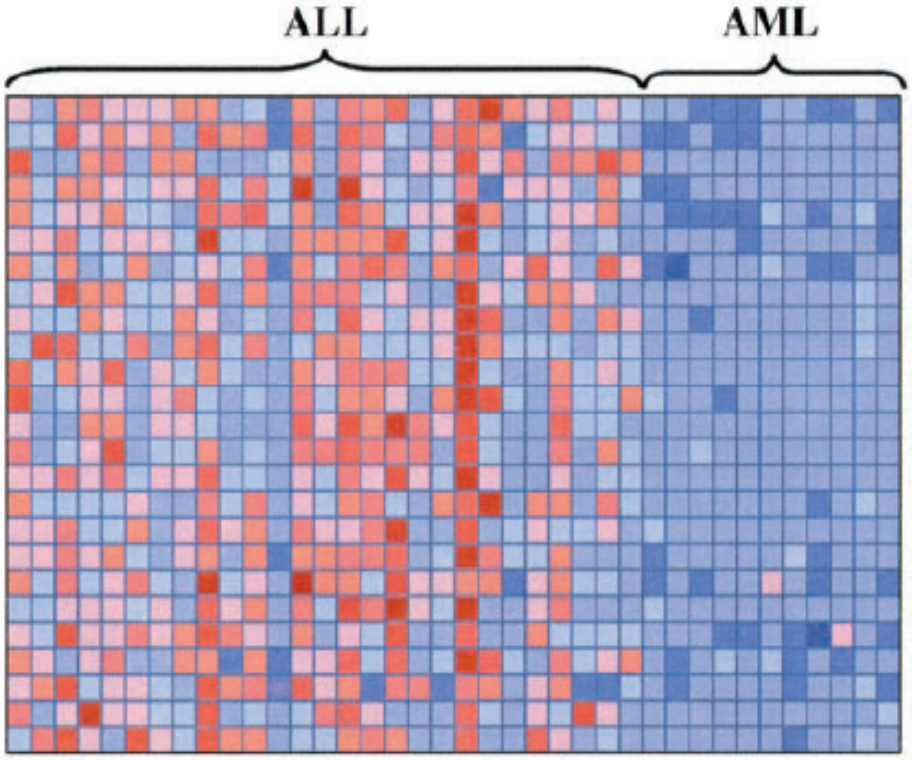
Can we identify genes as biomarkers of prenatal exposure to arsenic??

Can we use gene expression signatures from a **training** population of newborns

to classify arsenic exposure in a **test** population??

Application of Two-Class Prediction Algorithm

Classic Example of Two-Class Prediction – to Distinguish tumor types



www.sciencemag.org SCIENCE VOL 286 15 OCTOBER 1999

Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring

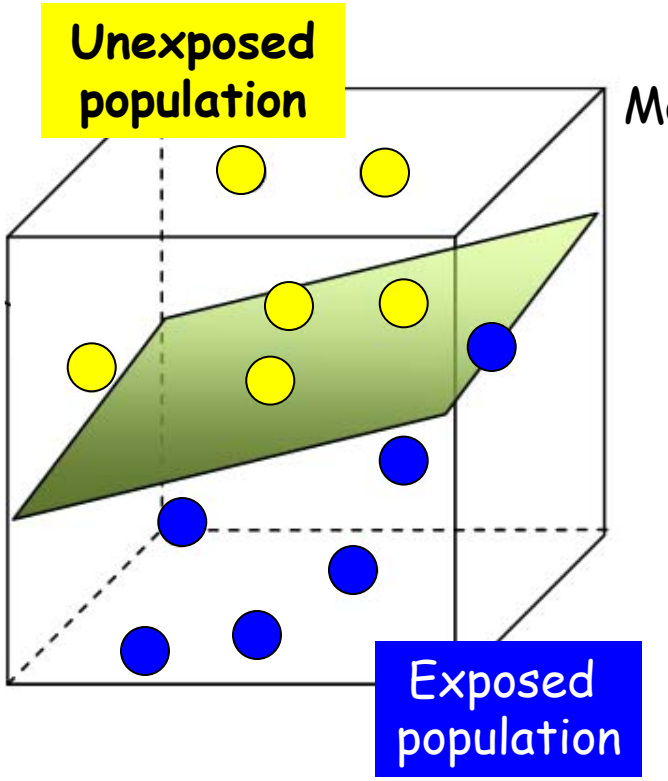
T. R. Golub,[1,2*†] D. K. Slonim,[1†] P. Tamayo,[1] C. Huard,[1] M. Gaasenbeek,[1] J. P. Mesirov,[1] H. Coller,[1] M. L. Loh,[2] J. R. Downing,[3] M. A. Caligiuri,[4] C. D. Bloomfield,[4] E. S. Lander[1,5*]

1) Use a training population to identify expression patterns that distinguish between two classes to create class predictor gene set

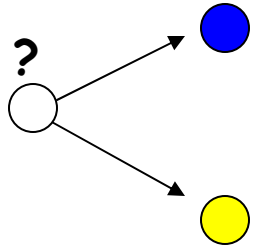
2) predictor then used to classify leukemia subtypes (85% accuracy)

Two-class prediction algorithm: Support Vector Machine

Training population of two groups with known environmental exposures used to build gene set predictor



Mathematical hyperplane

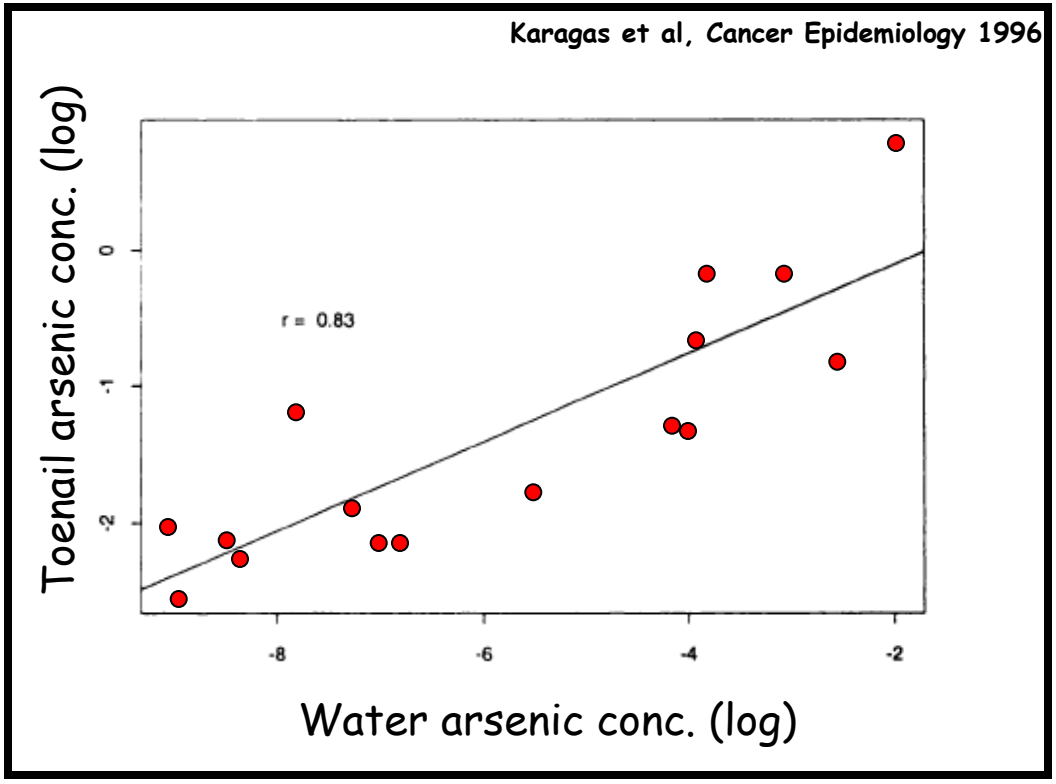


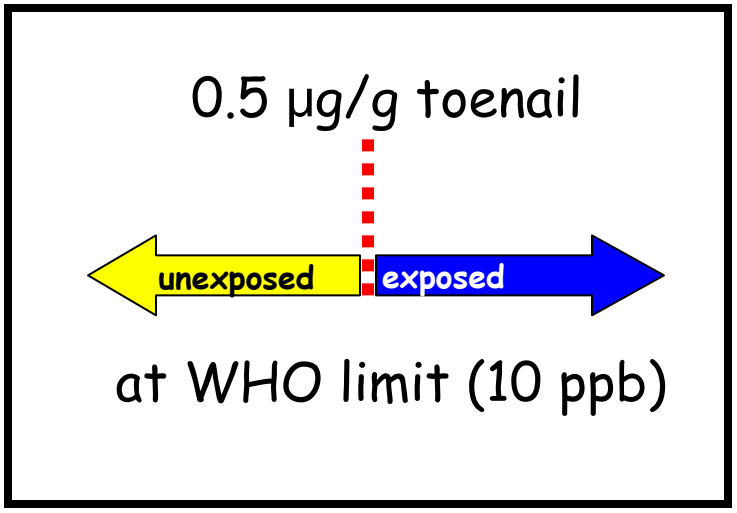
Predictor used to classify test samples with unknown exposures

1st **Training** Population
13 Newborns randomly selected
From population of 32

Newborn subject	1	2	3	4	5	6	7	8	9	10	11	12	13

Maternal exposure determined using toenail arsenic concentration: an indicator of chronic exposure



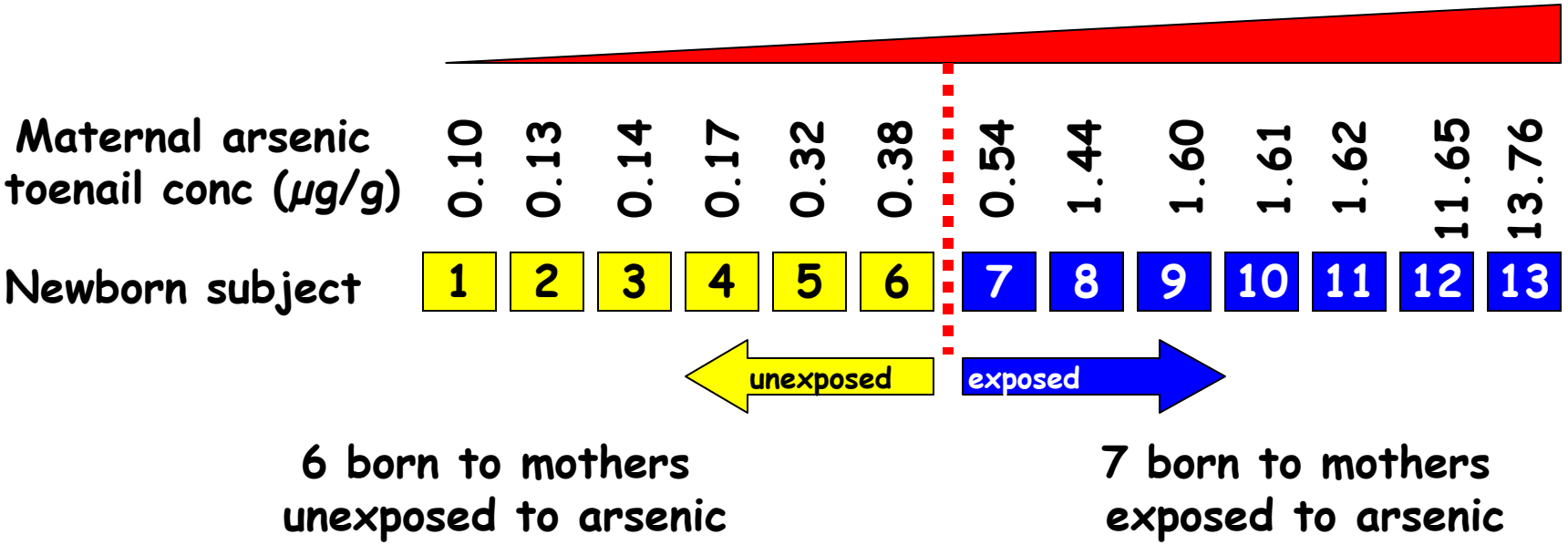


Kile et al, Cancer Epi Biomarker Prev 200
Karagas et al, Am J Epi 2000

1st **Training** Population
13 Newborns randomly selected
From population of 32

Maternal arsenic toenail conc (μg/g)	**0.10**	**0.13**	**0.14**	**0.17**	**0.32**	**0.38**	**0.54**	**1.44**	**1.60**	**1.61**	**1.62**	**11.65**	**13.76**
Newborn subject	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**

1st **Training** Population
13 Newborns randomly selected
From population of 32

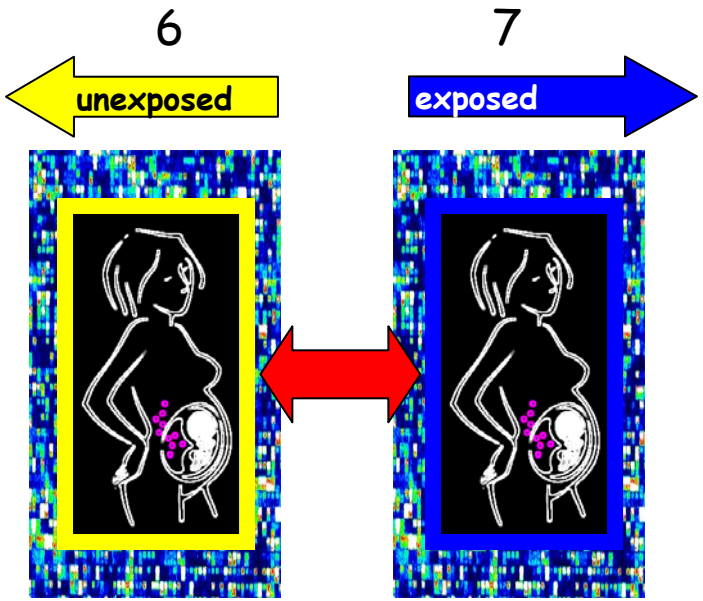


6 born to mothers unexposed to arsenic

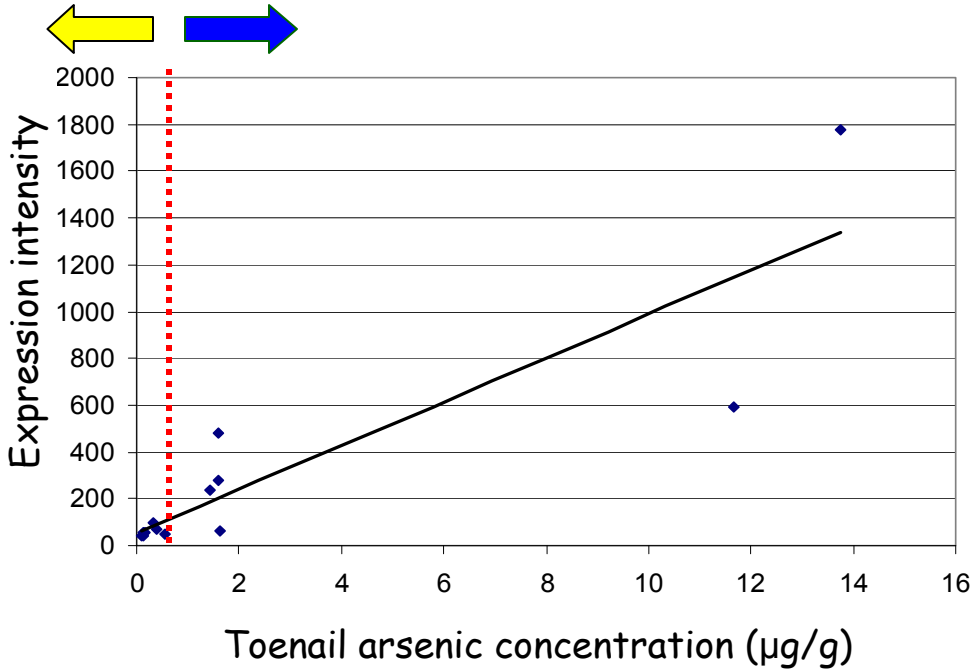
7 born to mothers exposed to arsenic

Integrate with gene expression data

Arsenic-Associated Gene Sets Identified by Integrating Arsenic Exposure and Gene Expression

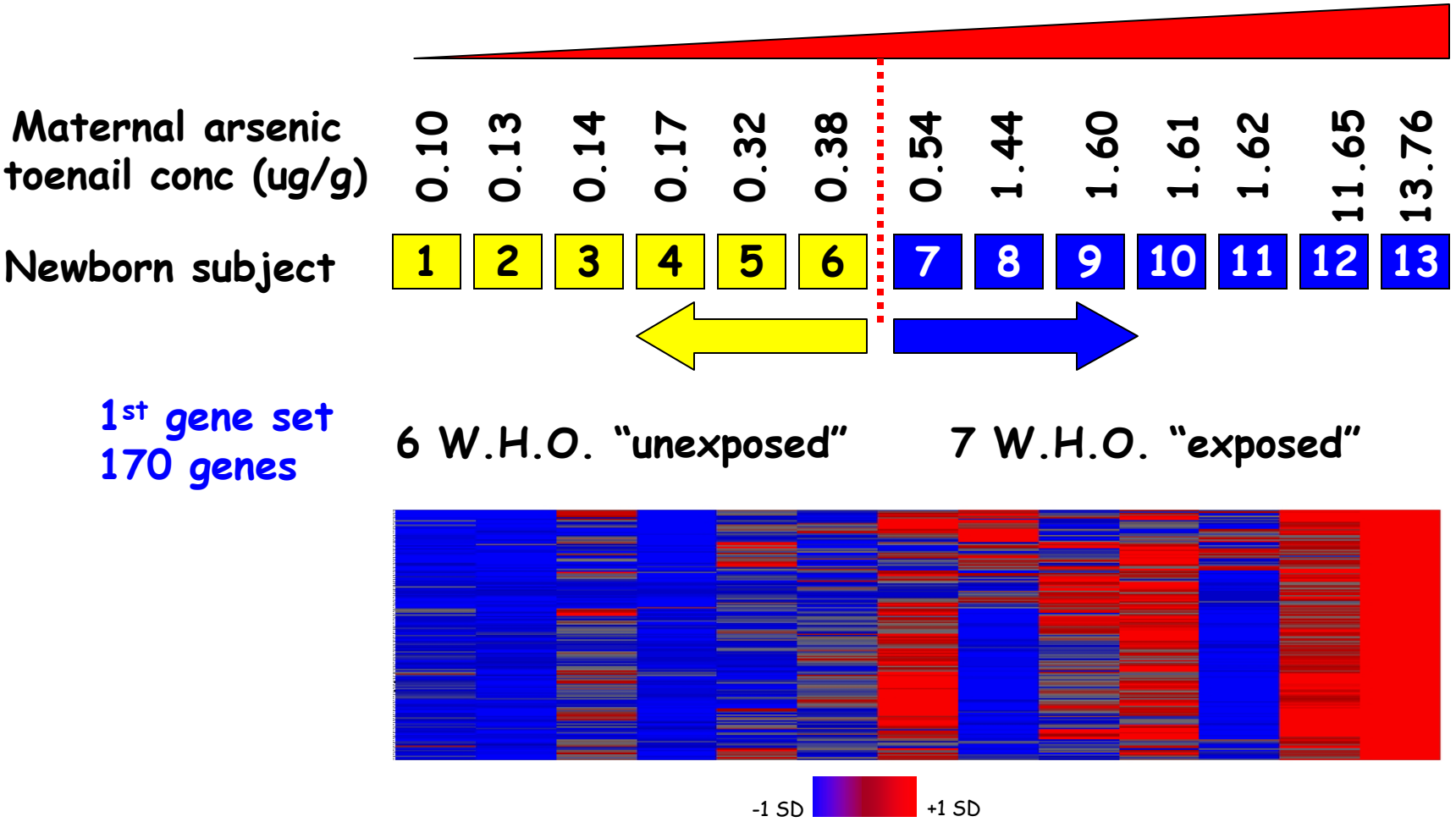


Comparative expression analysis:
Statistically Significant
Differential Expression
(unexposed/exposed)
1.5 FC , p< 0.05



Statistically significant association
(+ or -) (p<0.01) of exposure and
expression

Arsenic-Associated Gene Set Identified from Newborn Training Population

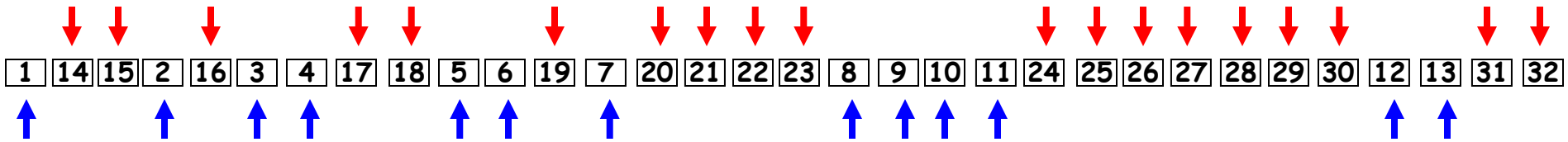


Expression signature identified from training population:

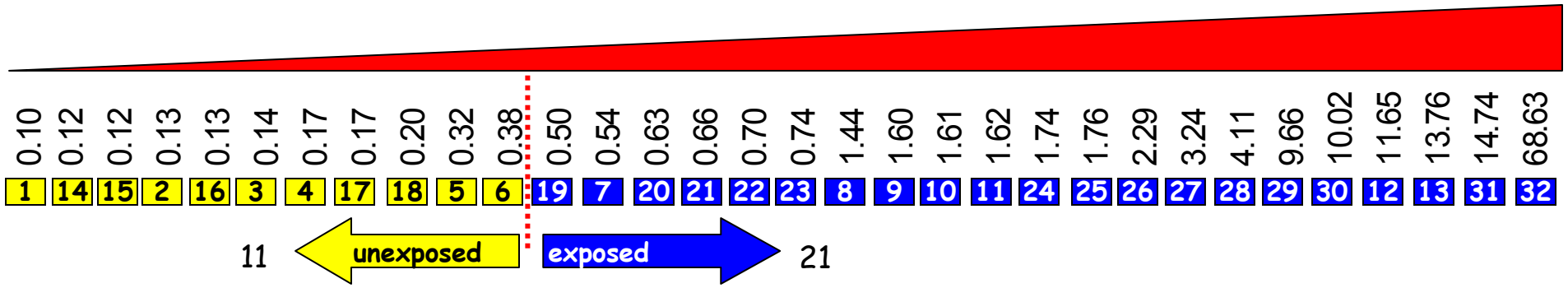
can we classify maternal exposure (unexposed or exposed) of remaining newborn population?

Newborn Training Population (13 newborns)

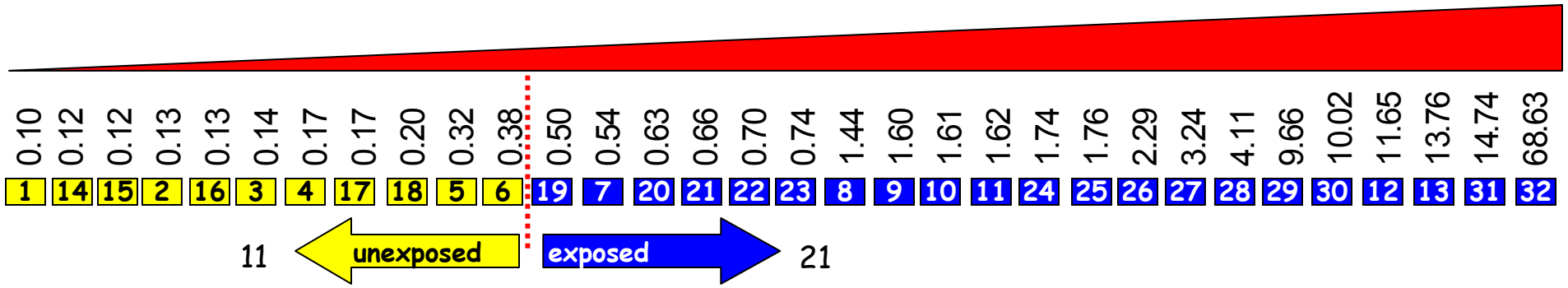
Newborn Test Population (19 newborns)



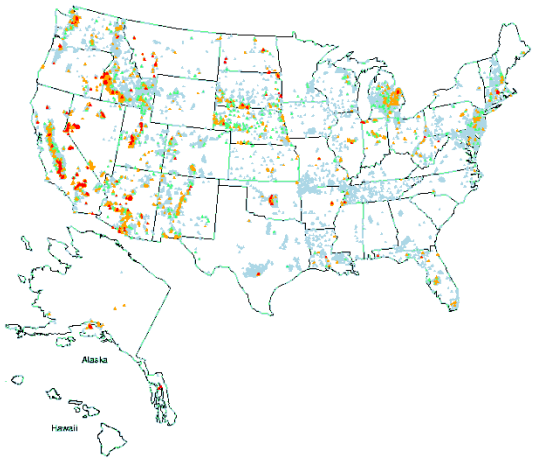
Reveal Maternal Exposure of Population



Reveal Maternal Exposure of Population



USA exposures

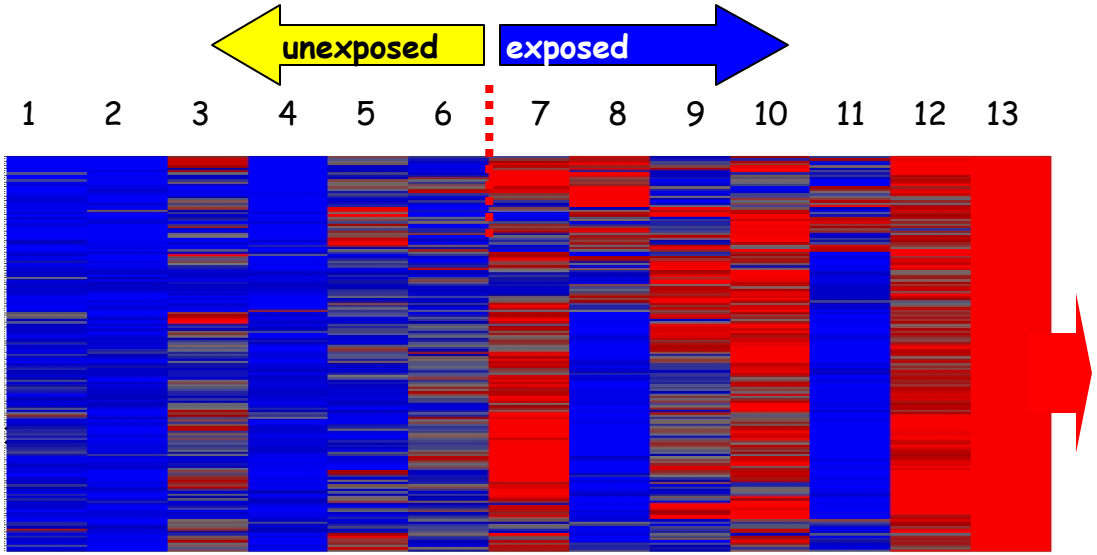


Bangladesh exposures

Apply Support Vector Machine Algorithm:

Classify Maternal Exposure in Test Population

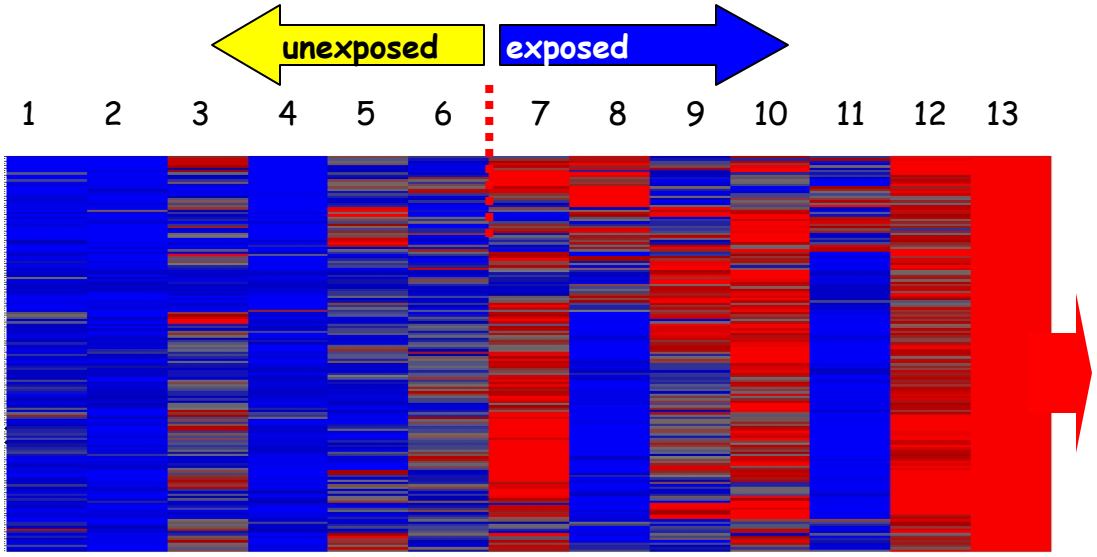
1st gene set
170 genes

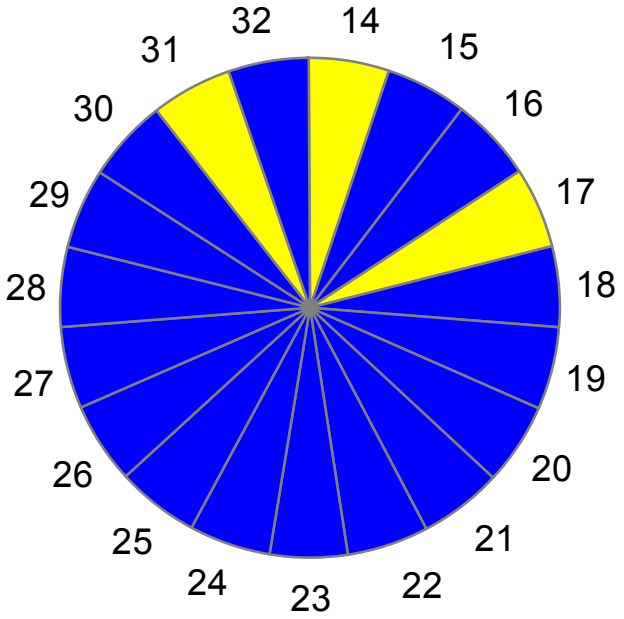


Apply Support Vector Machine Algorithm:

Classify Maternal Exposure in Test Population

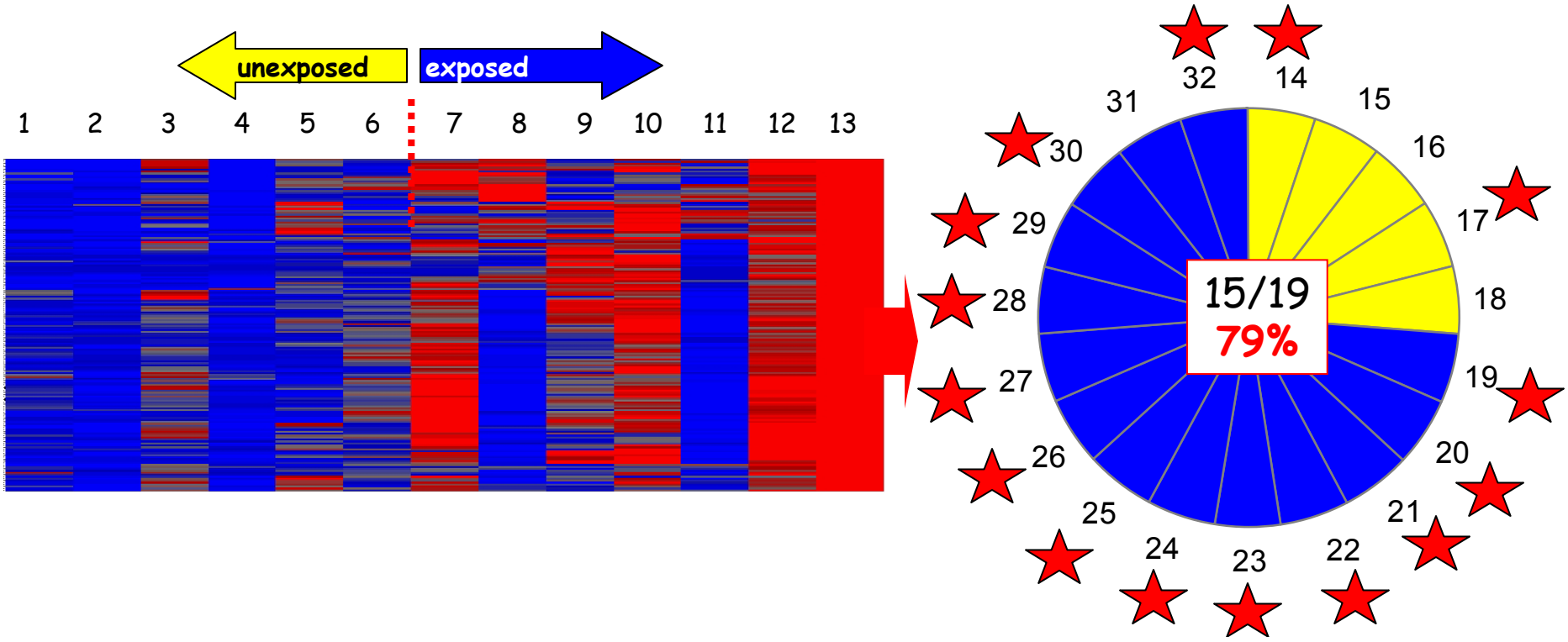
1st gene set
170 genes



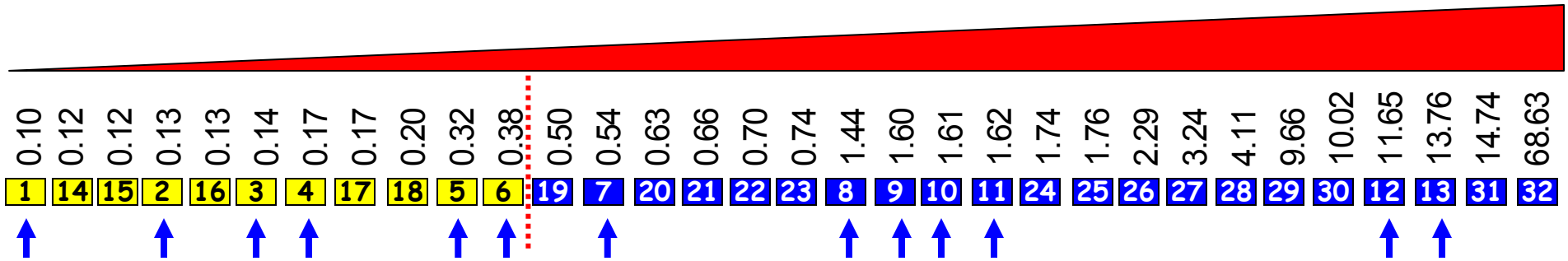


Robust predictor of maternal exposure in newborn test population

1st gene set
170 genes

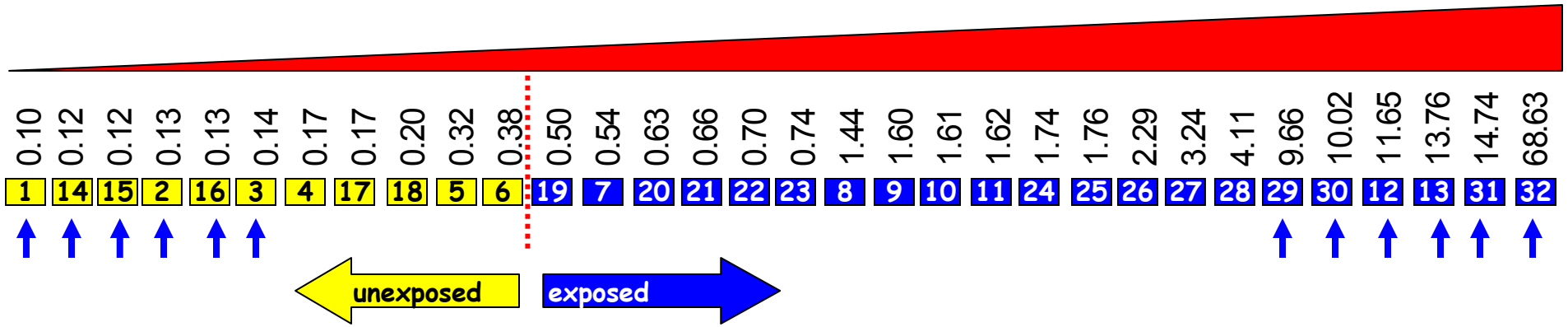


Distributed Exposure of Newborn Training Population



Would class prediction be higher using a training population of newborns whose mothers were at the extremes of arsenic exposure?

Extreme Exposure Based Training Population

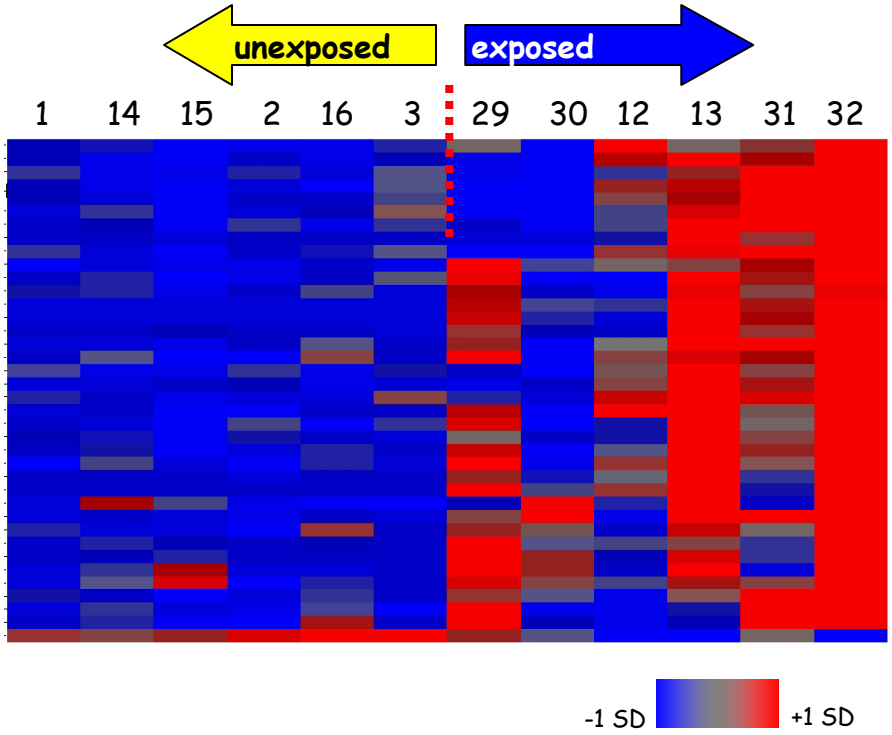


6 newborns
at low end of exposure

6 newborns with mothers
at high end of exposure

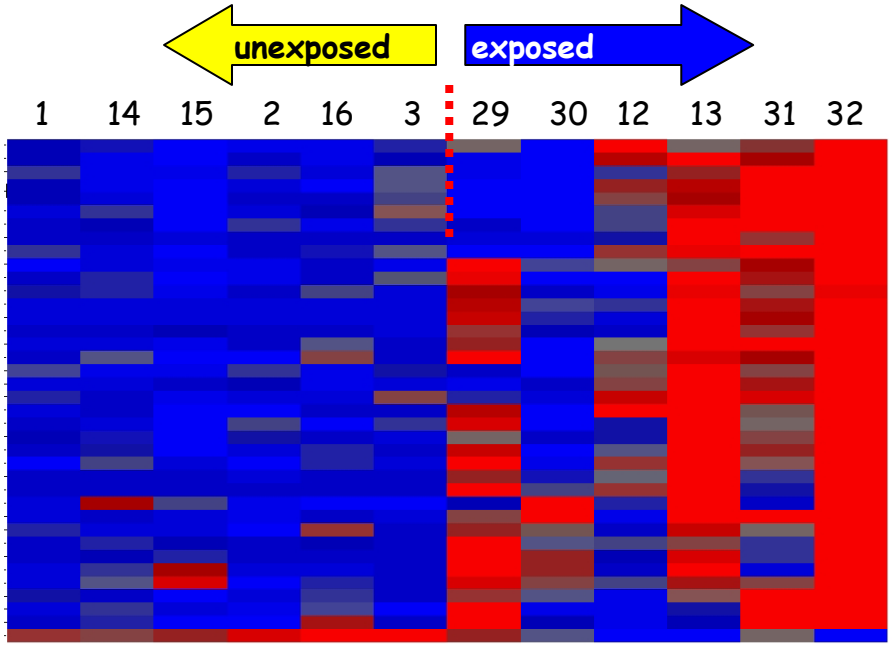
Extreme Exposure Based Class Predictor

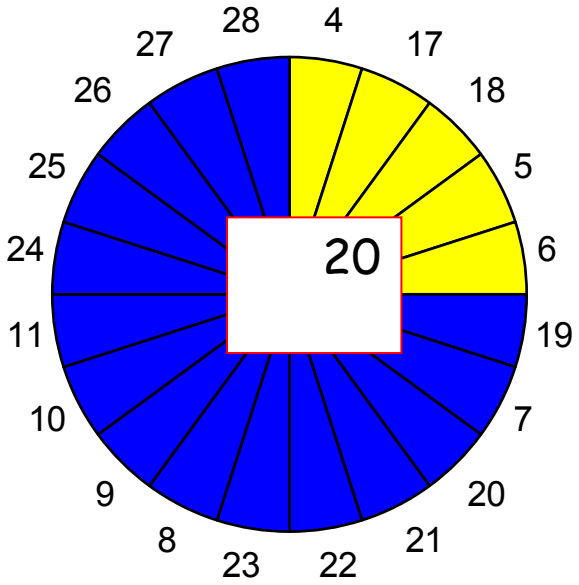
2nd gene set
38 genes



How well can the 2nd gene set (38 genes) classify maternal exposures of 20 newborns in test population?

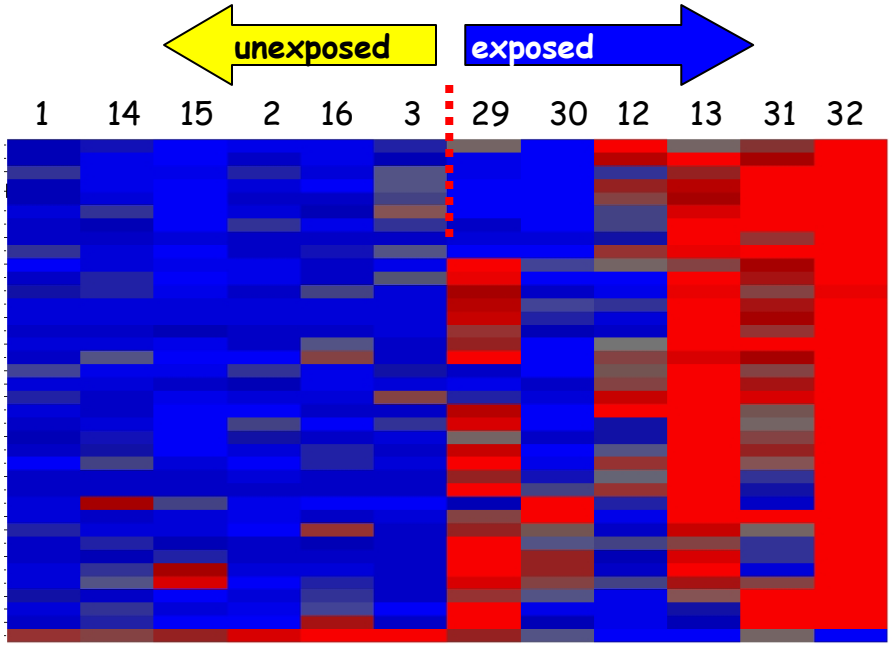
2nd gene set
38 genes

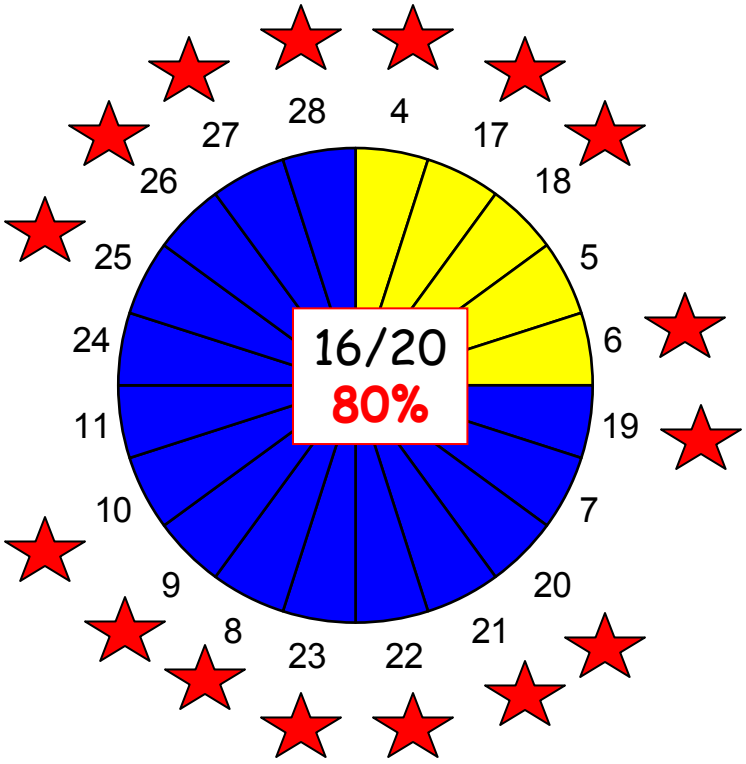




Robust Class Predictor

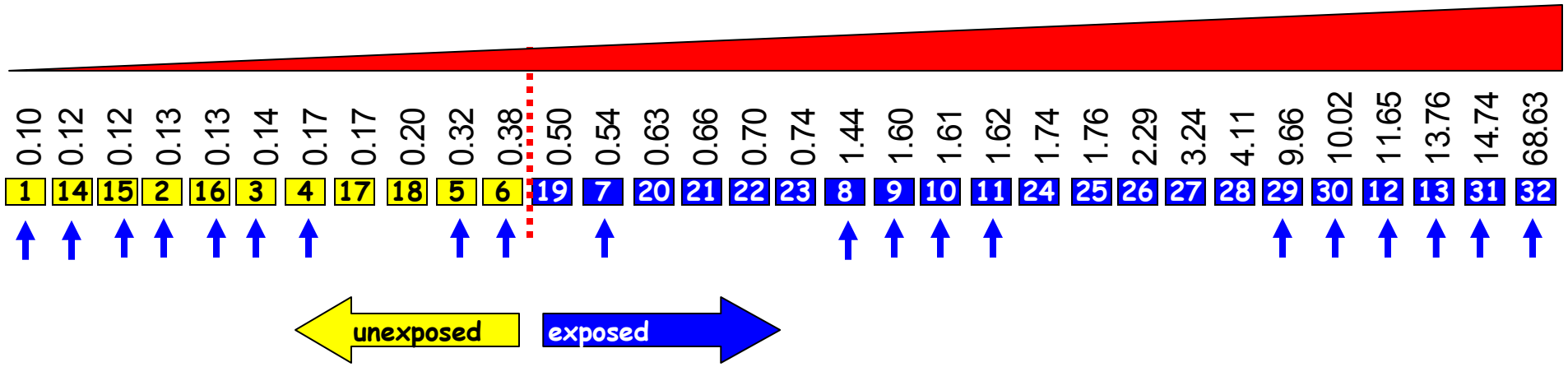
2nd gene set
38 genes





How would prediction change using a combination of the 1st two newborn training populations?

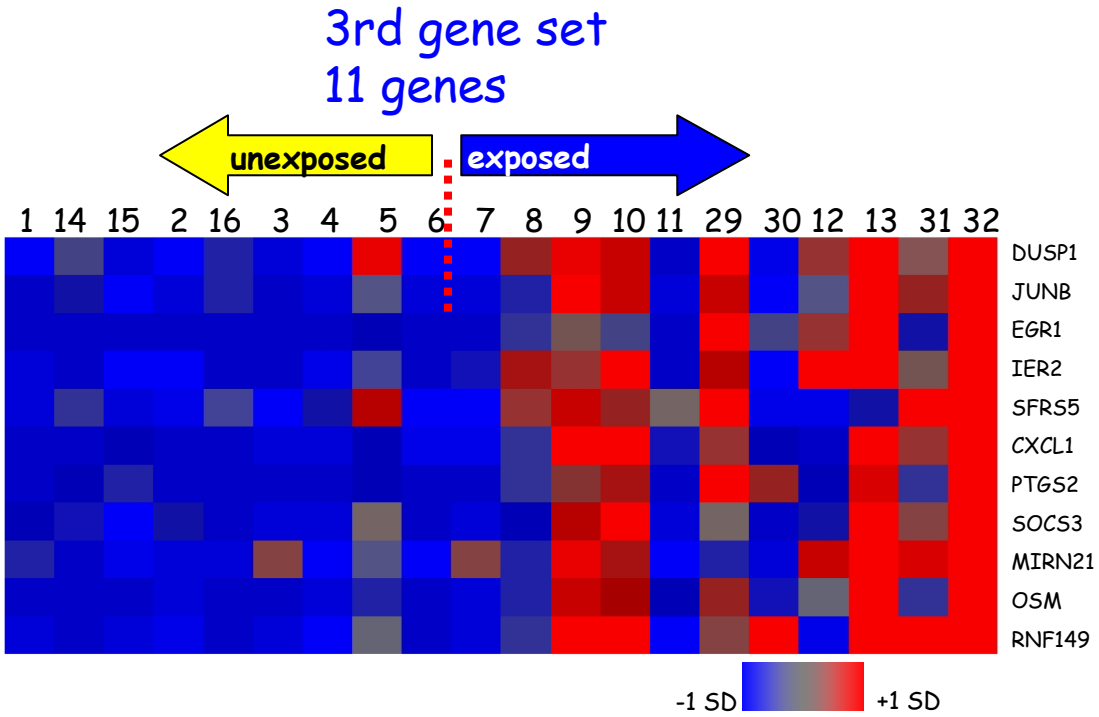
Combined training population Class Predictor

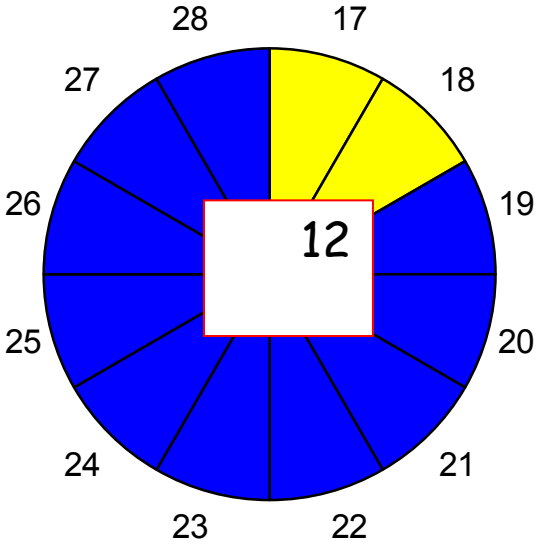


9 newborns
With unexposed mothers

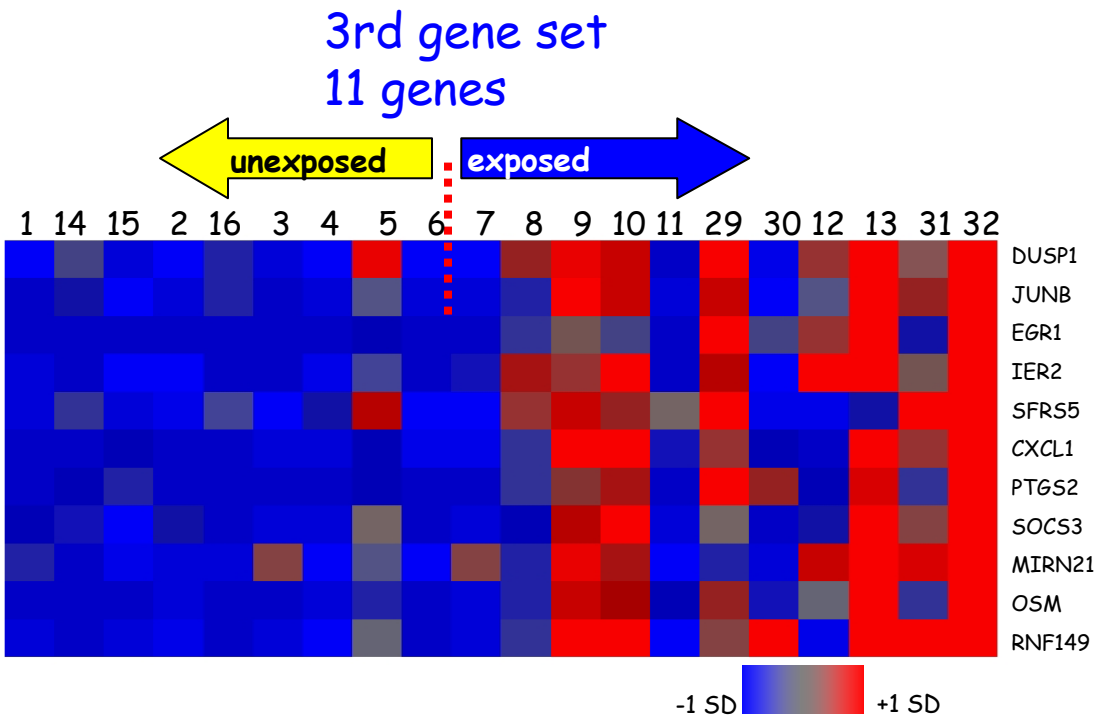
11 newborns with
Exposed mothers

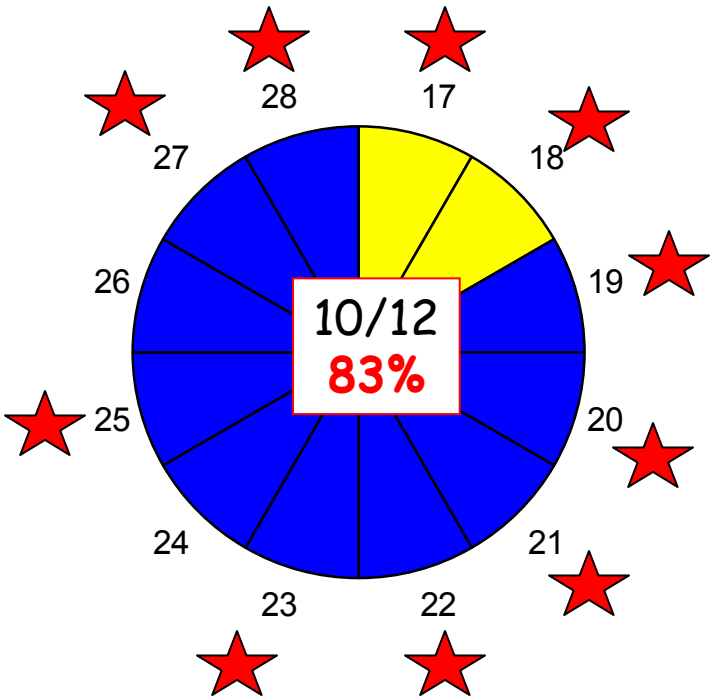
How well can the 3rd gene set (11 genes) classify maternal exposures of 12 newborns in test population?



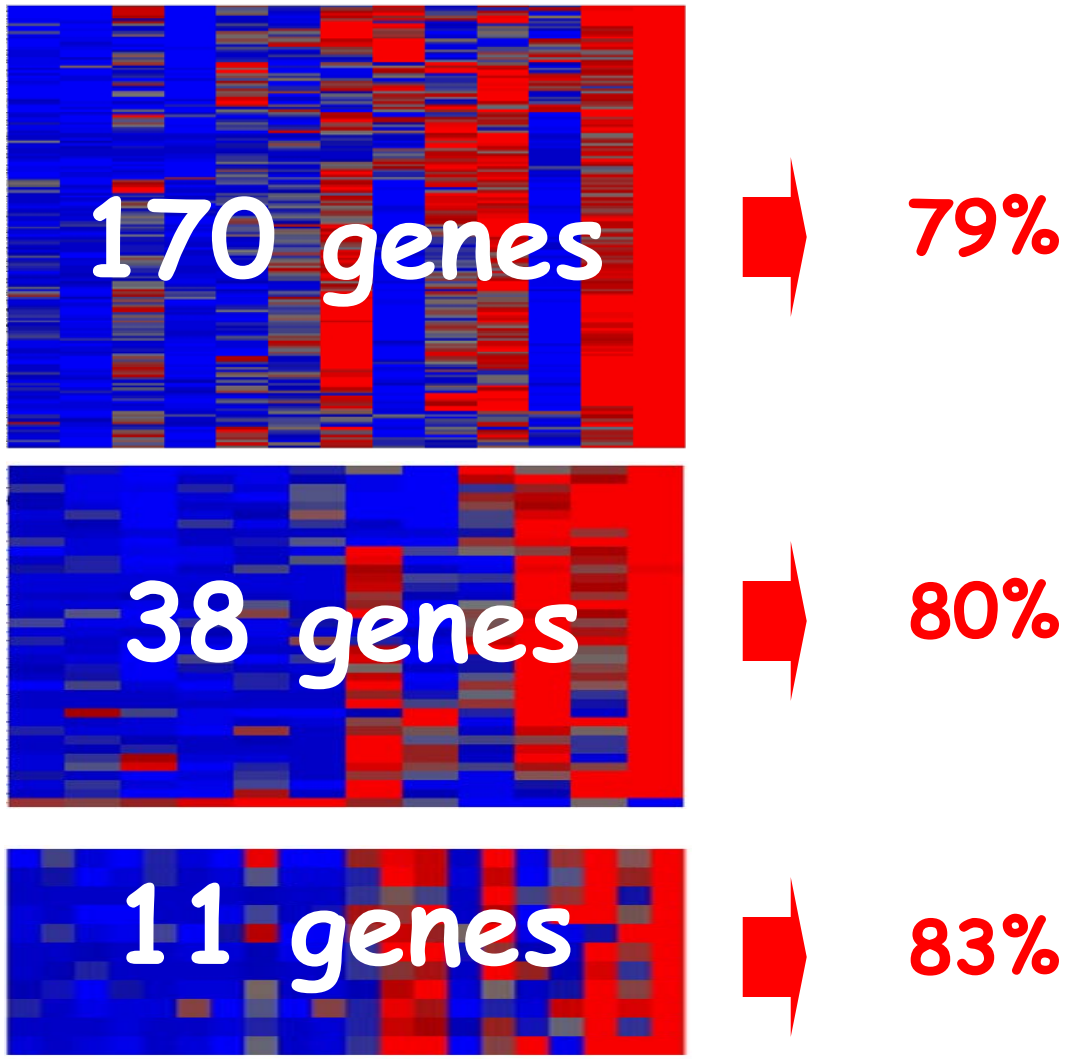


Robust Class Predictor



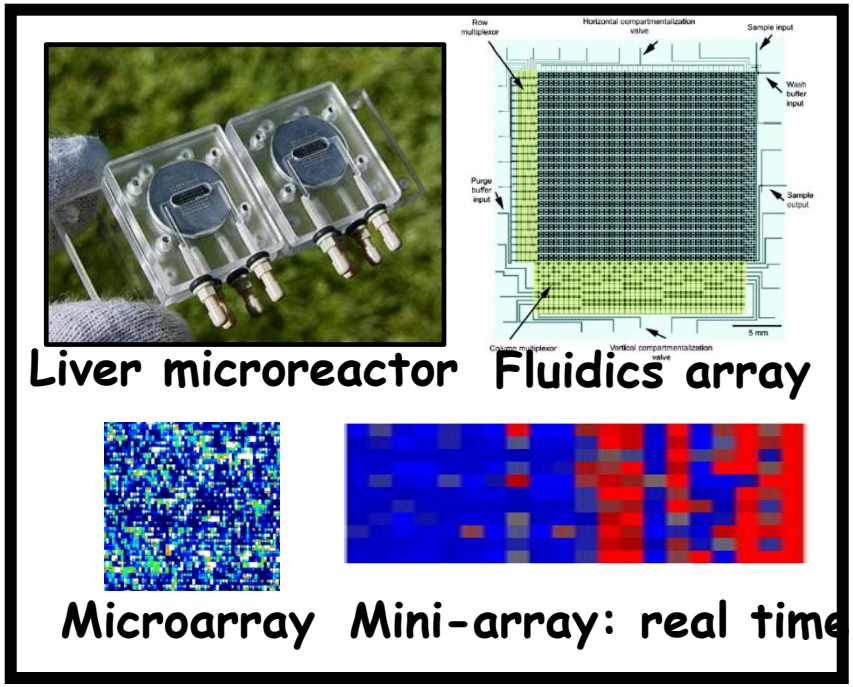


Equally predictive gene set of much smaller size: potential biomarkers



New tool development: measure exposures in the population

tools

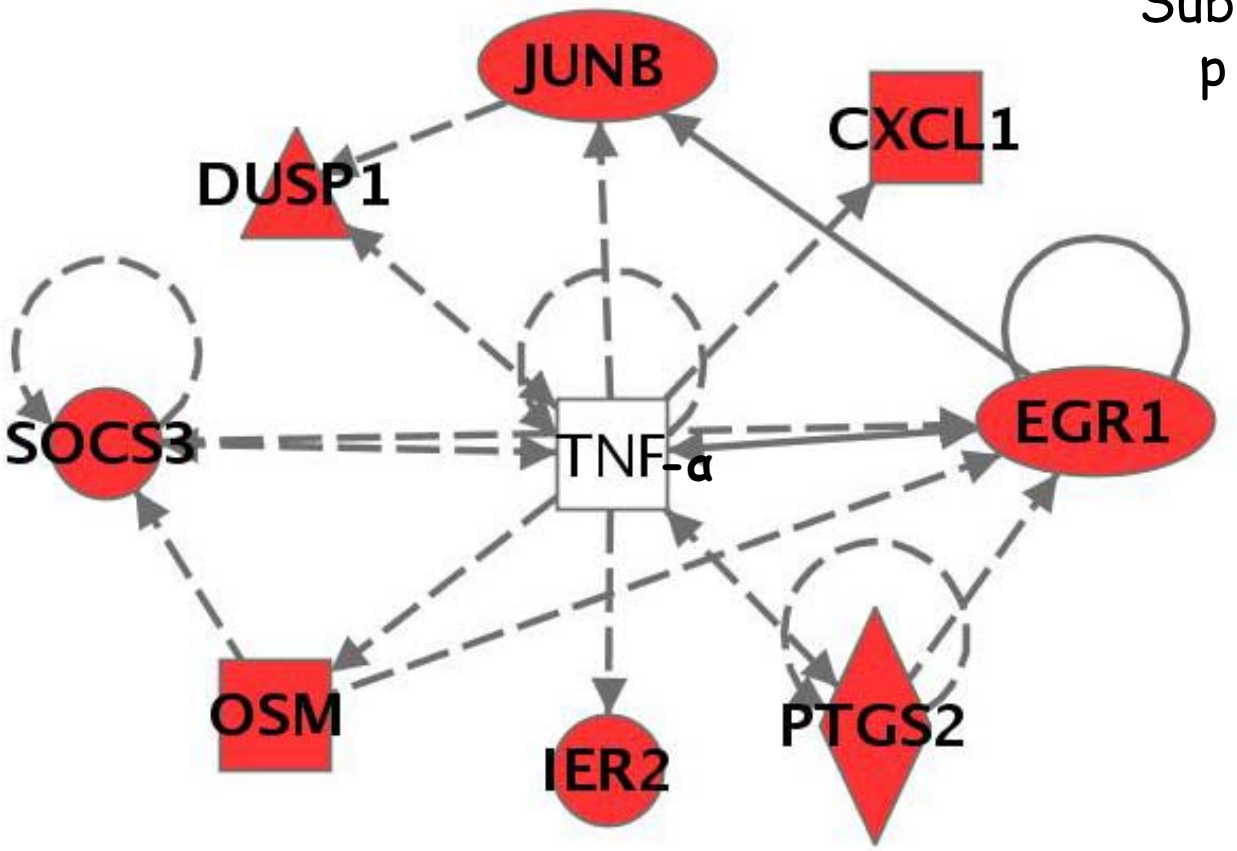


Biomarkers of environmental exposures

Are there known molecular interactions among the 11 biomarkers?

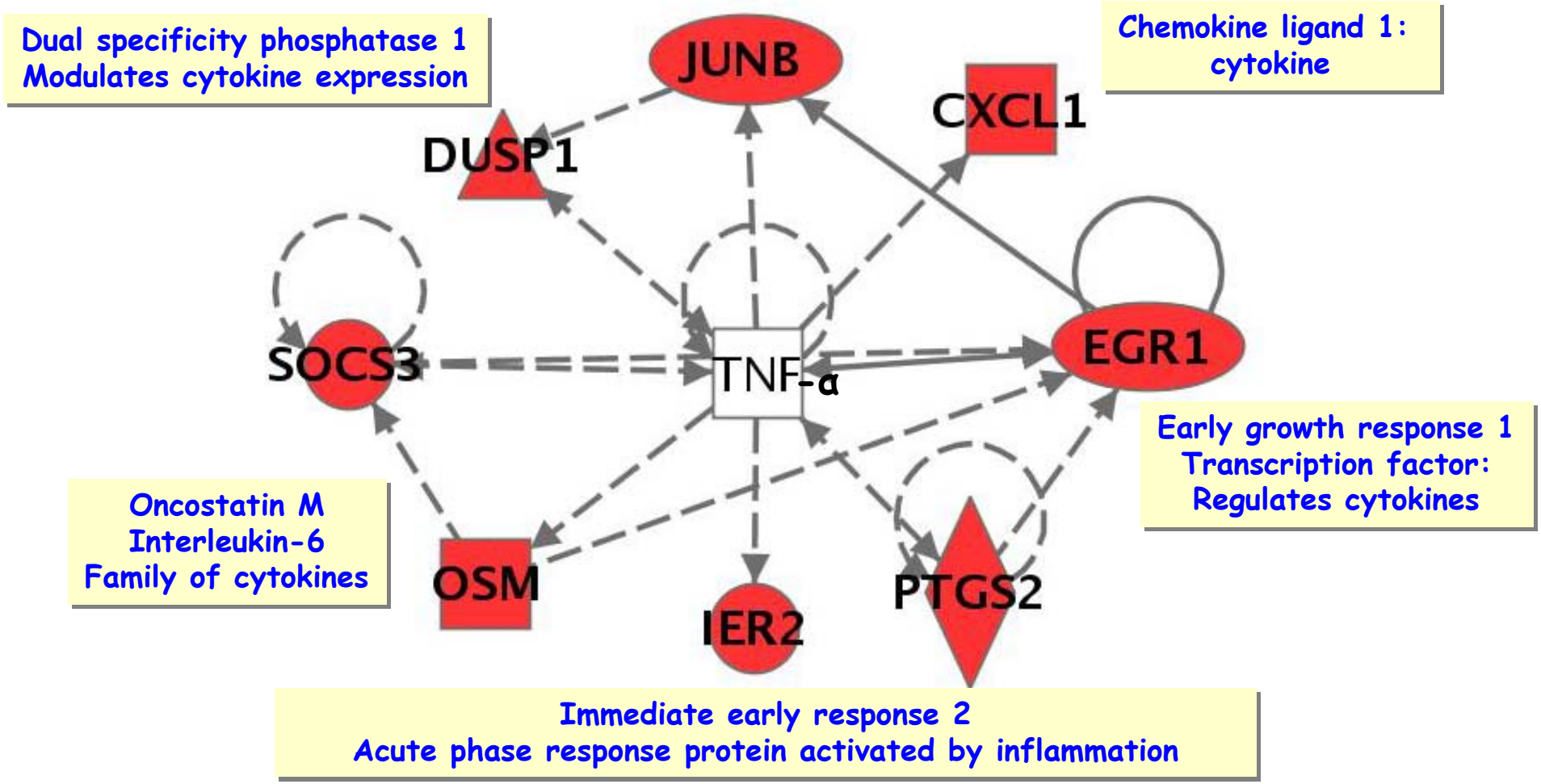
8 of 11 Potential Biomarker Genes are associated with the cytokine, TNF-α

Subnetwork
$p < 10^{-22}$



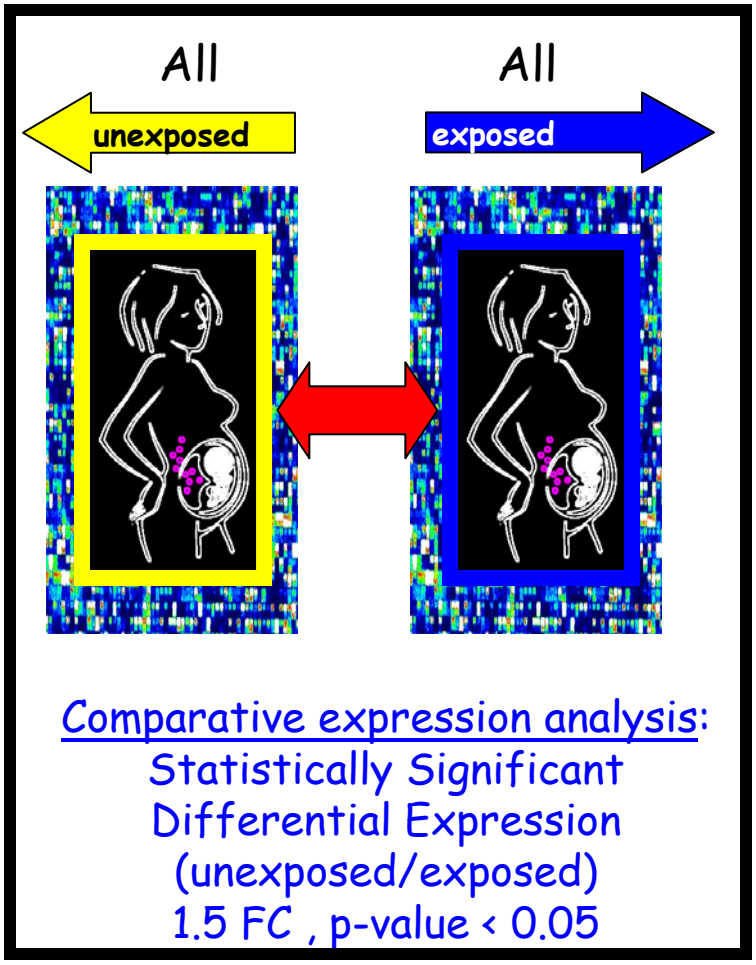
TNF-α known to be induced by arsenic in animal models (Germolec et al 1996, 1997, 1998)

8 of 11 Potential Biomarker Genes are associated with the cytokine, TNF-α



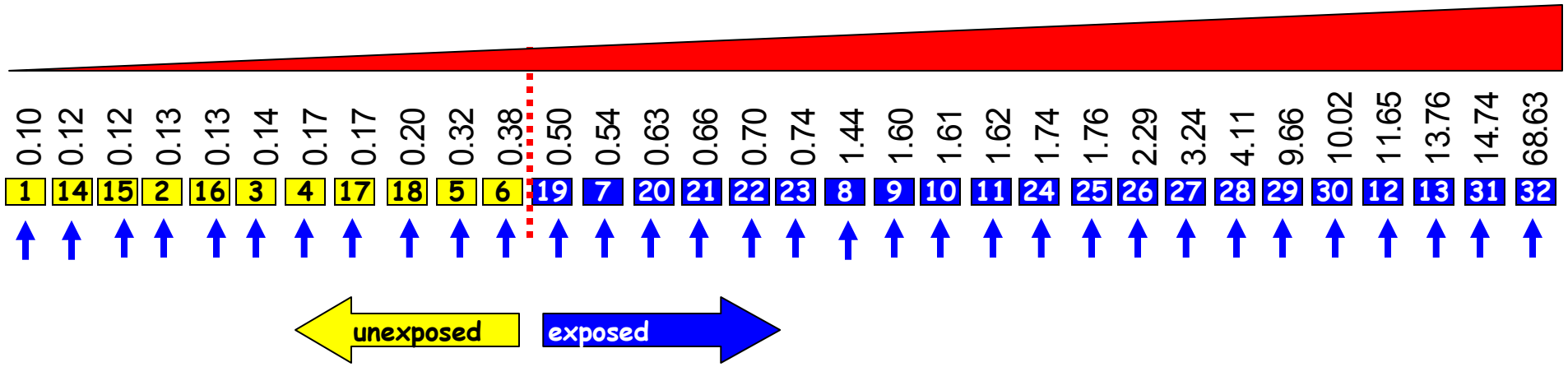
Stress Response: cytokines: inflammatory response

What is genome-wide impact of prenatal arsenic exposure?



Identify genes differentially expressed between the unexposed and exposed populations

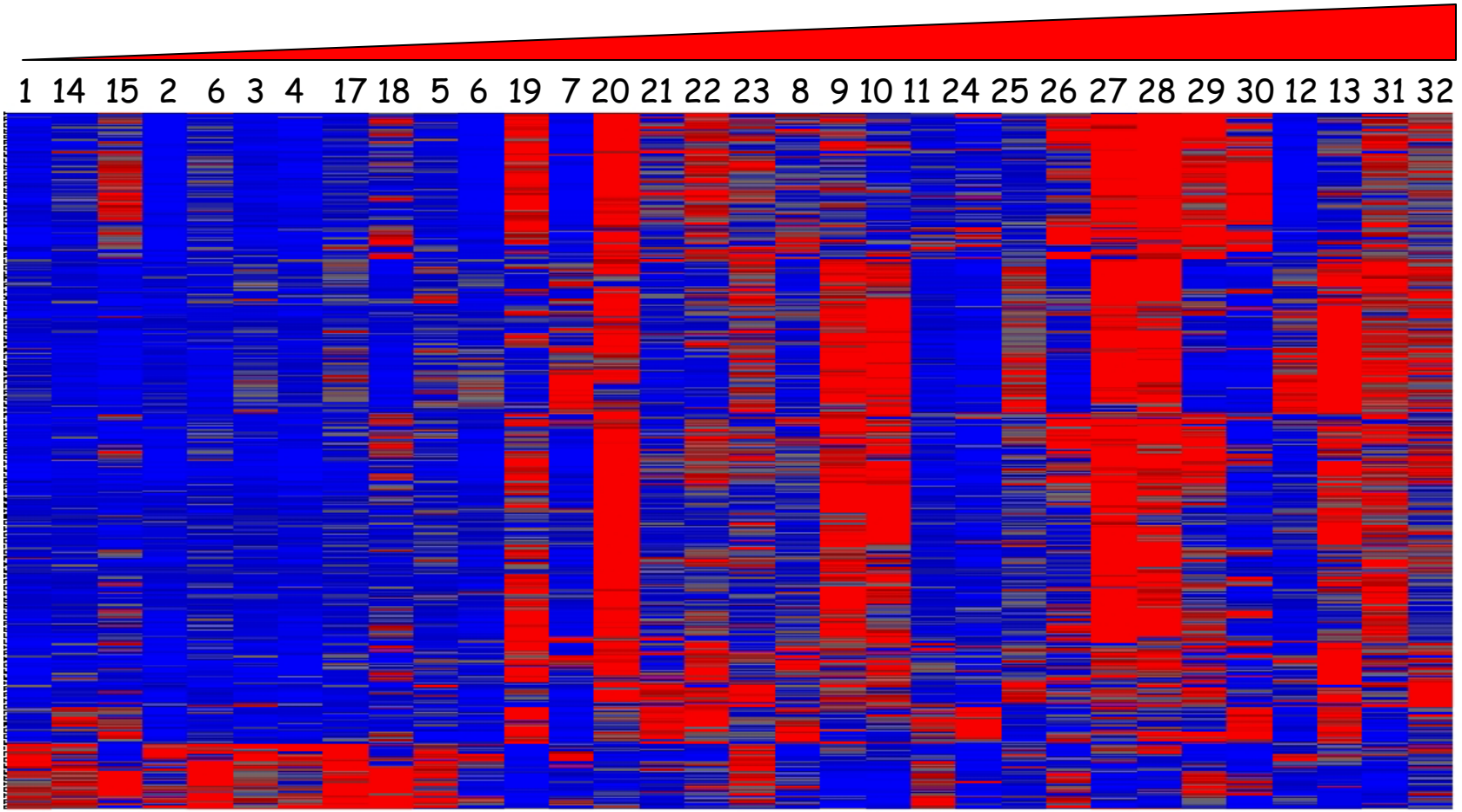
Identifying Differentially Expressed Genes Between the Two Populations



11 newborns born to unexposed mothers

21 newborns born to exposed mothers

Robust Genome Wide Changes of PRENATAL Arsenic Exposure



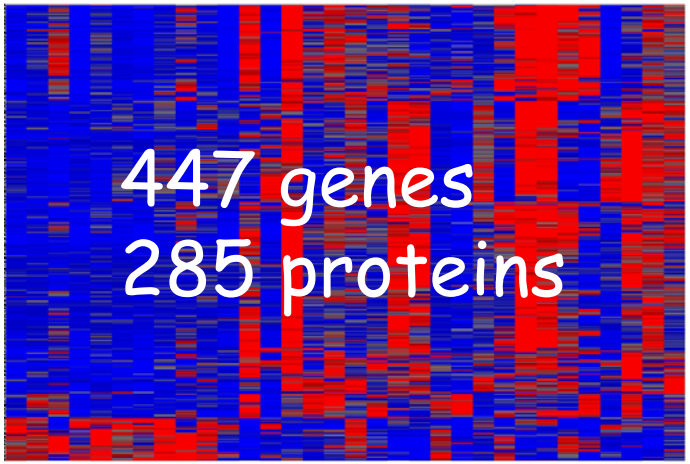
447 genes

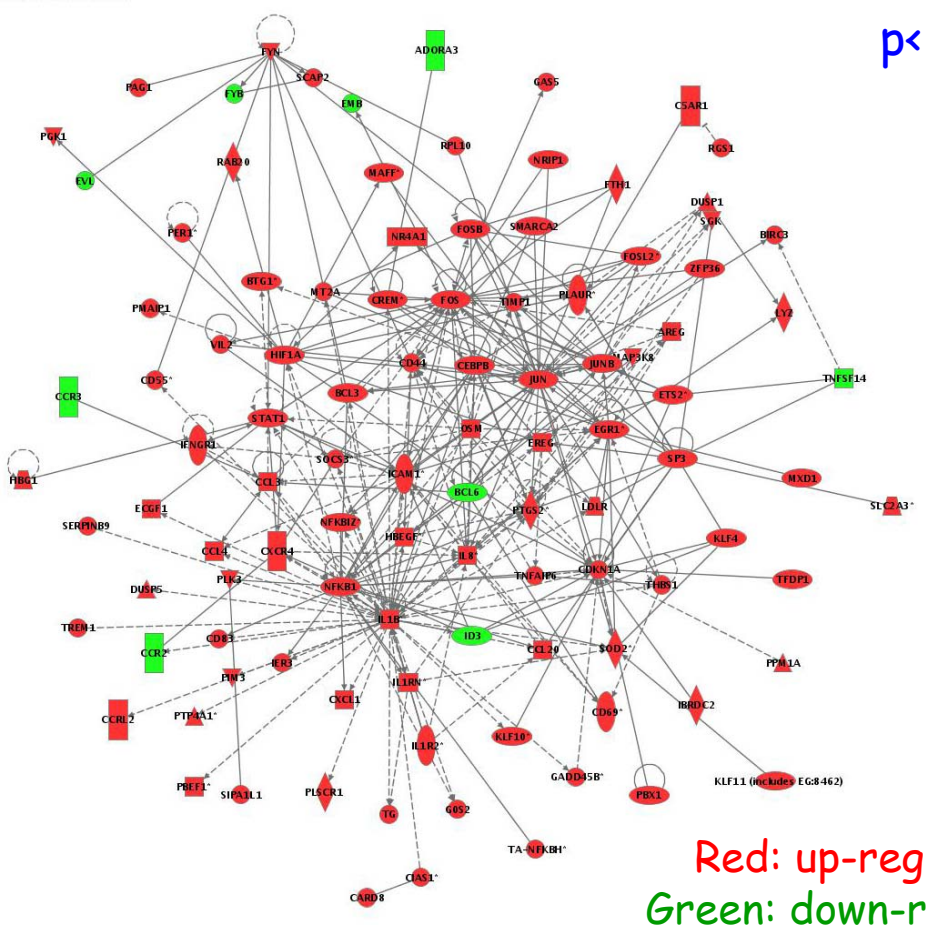
90% up-regulated

Are there known molecular interactions among the arsenic modulated genes??

Which biological pathways are modulated upon exposure to arsenic??

Large Arsenic-Modulated Interactome



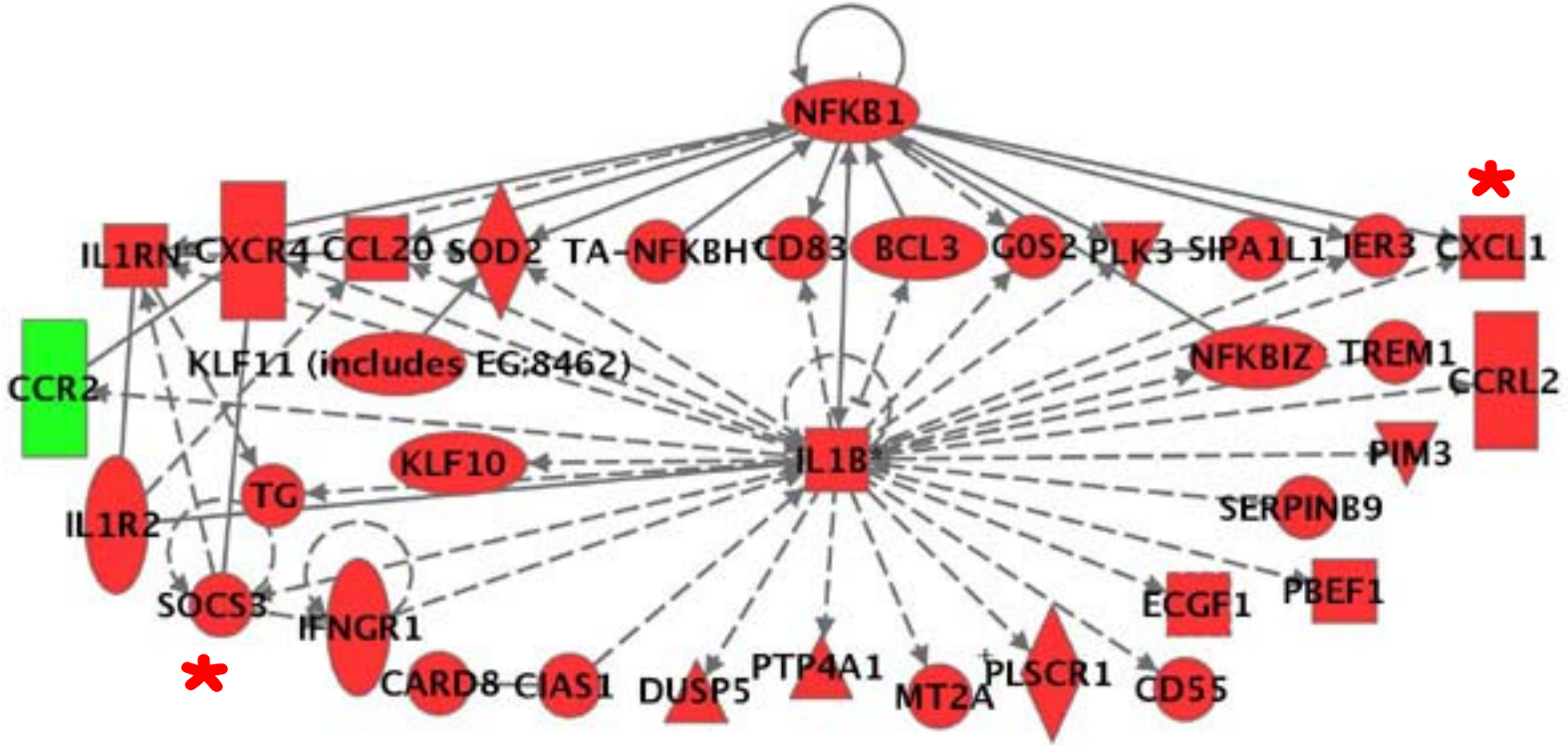


$p<10^{-55}$

Red: up-regulated
Green: down-regulated

105 proteins interact

Subnetwork 1 integrates 2 biomarkers with nuclear transcription factor, NF-κB



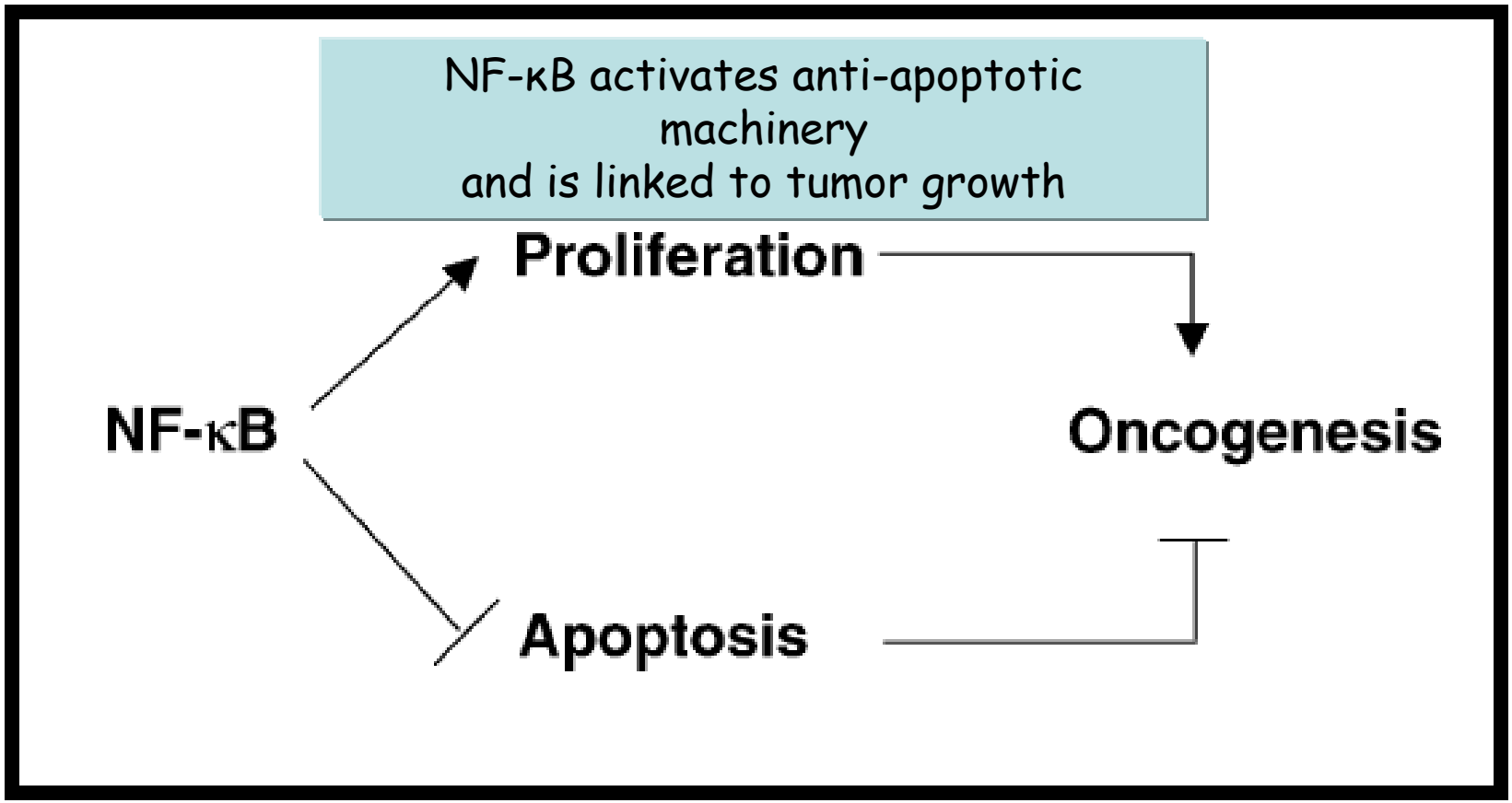
* Biomarkers of prenatal arsenic exposure

NF-κB regulates inflammation-related molecules
Note: all targets are increased

IL1-B is acute phase protein that increases in response to inflammation

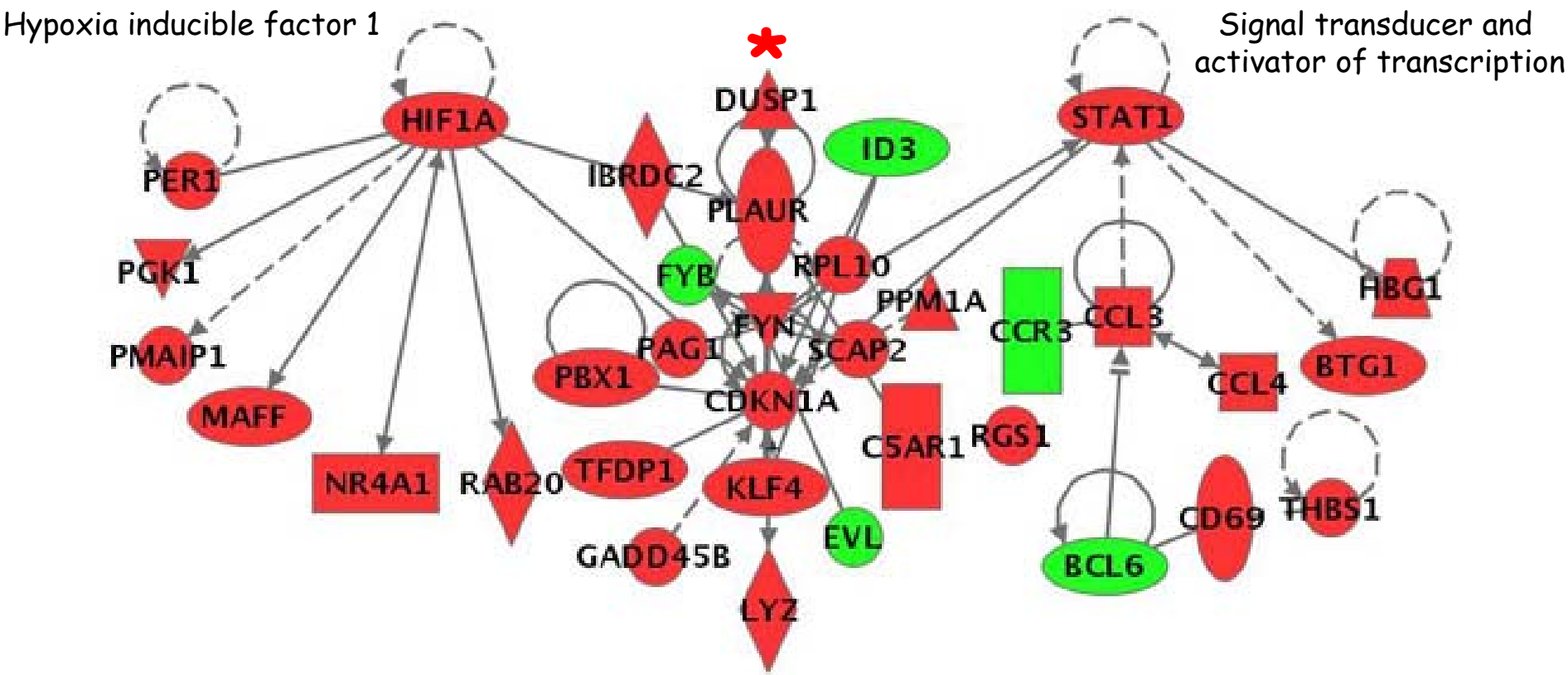
NF-κB activation has been identified in cell culture treated with arsenic (Huang et al, MCB 2001)

NF-κB is a key regulator of oncogenesis



Transient gene expression changes or epigenetic reprogramming?

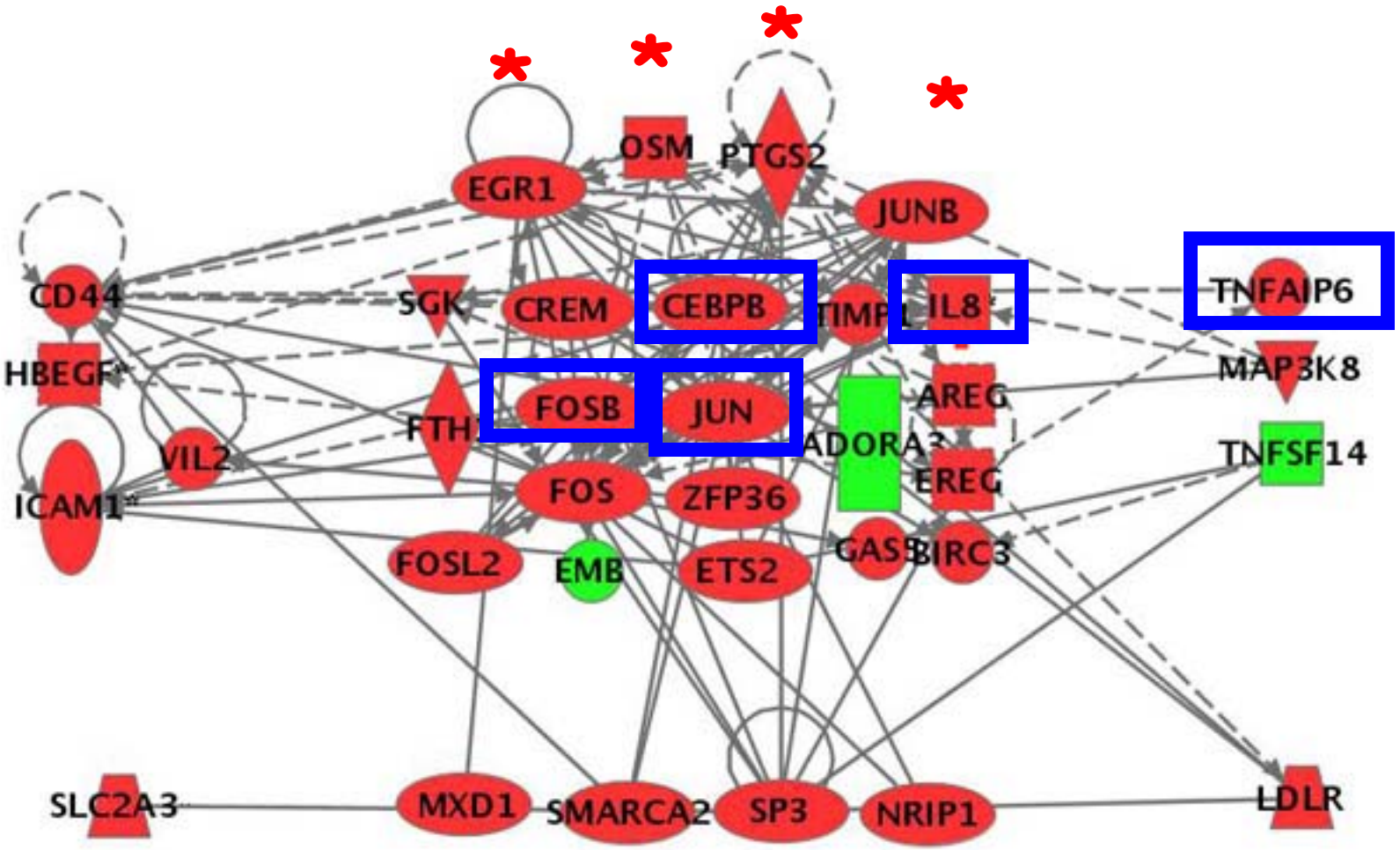
Subnetwork 2 integrates DUSP1 with stress activated transcription factors



STAT1 involved in cytokine signal transduction
Activation of both transcriptions factors linked to arsenic exposure in cell culture and animal models

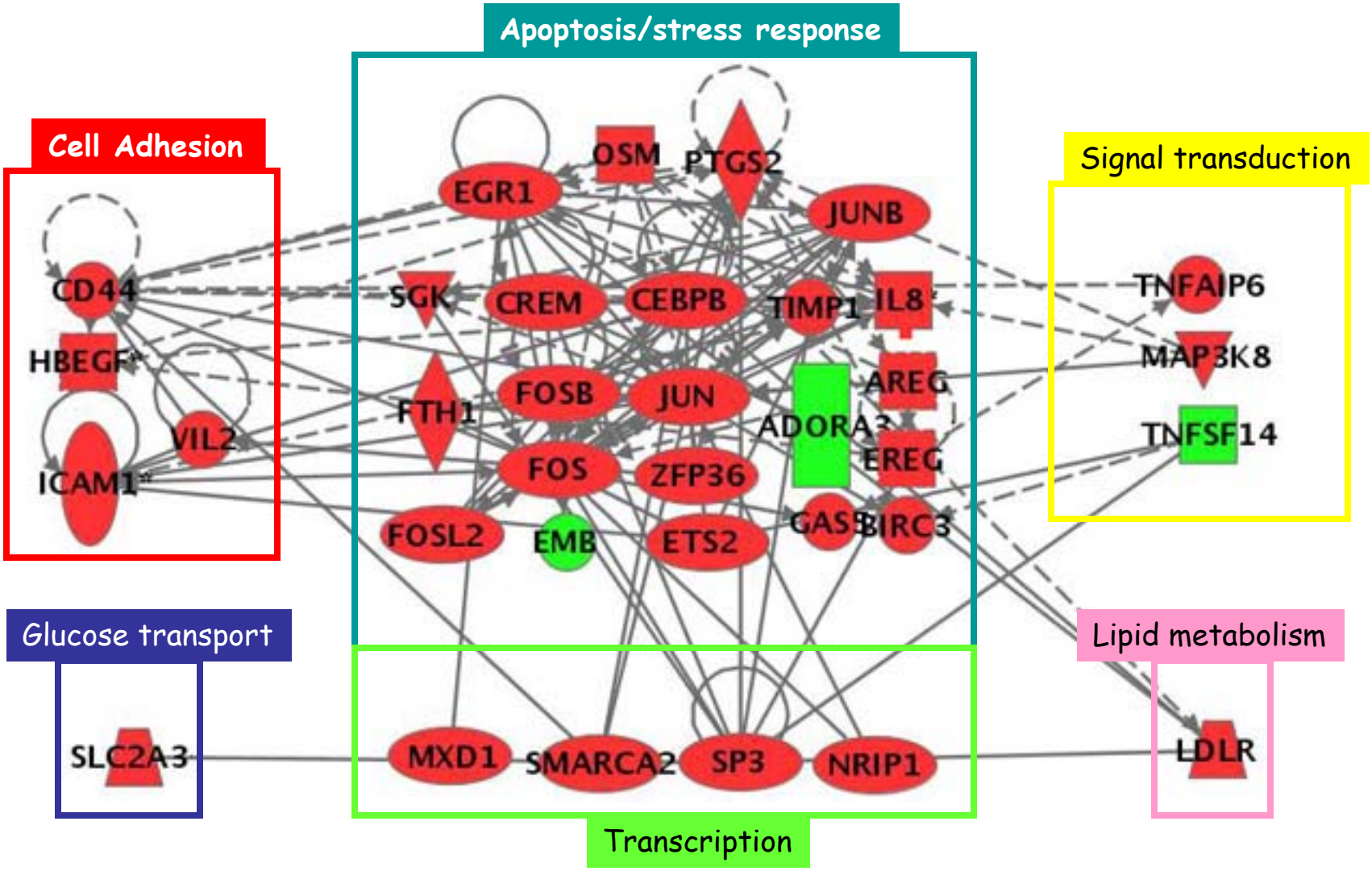
Chelbi-alix et al, Oncogene 2003 Kamat et al, Tox Sci 2005

Subnetwork 3 integrates 4 biomarkers with five members of the IL-6 Signaling Pathway



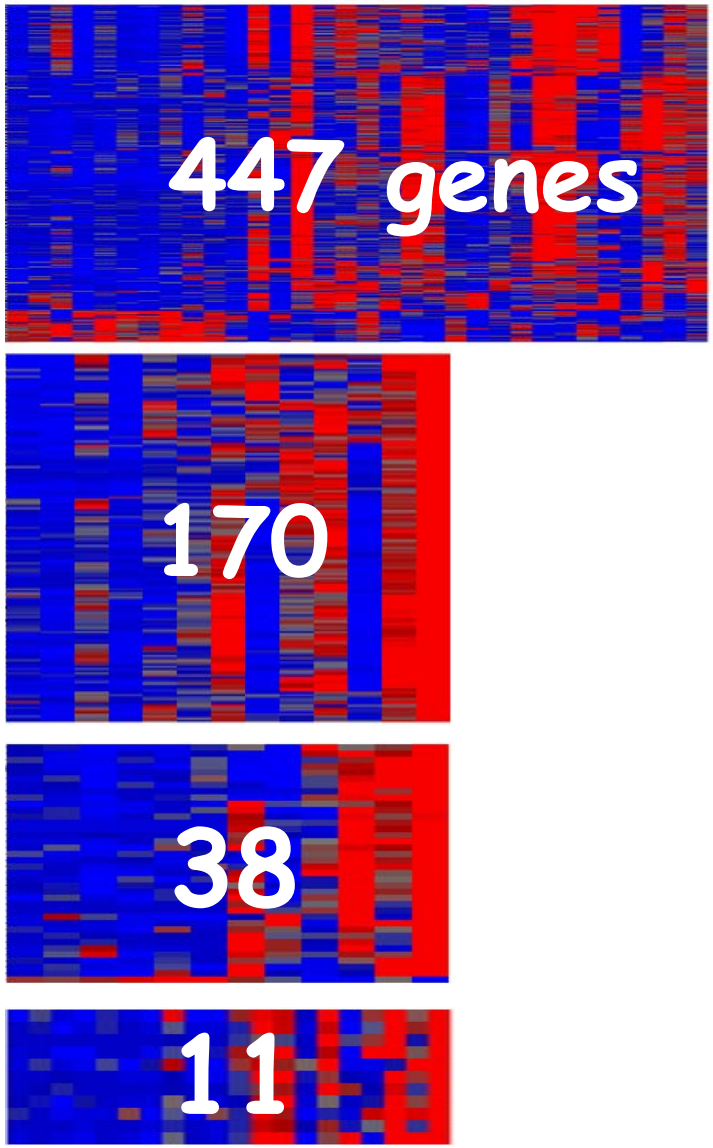
IL-6 is mediator of Acute-phase proteins

Many biological pathways are modulated in response to prenatal arsenic exposure

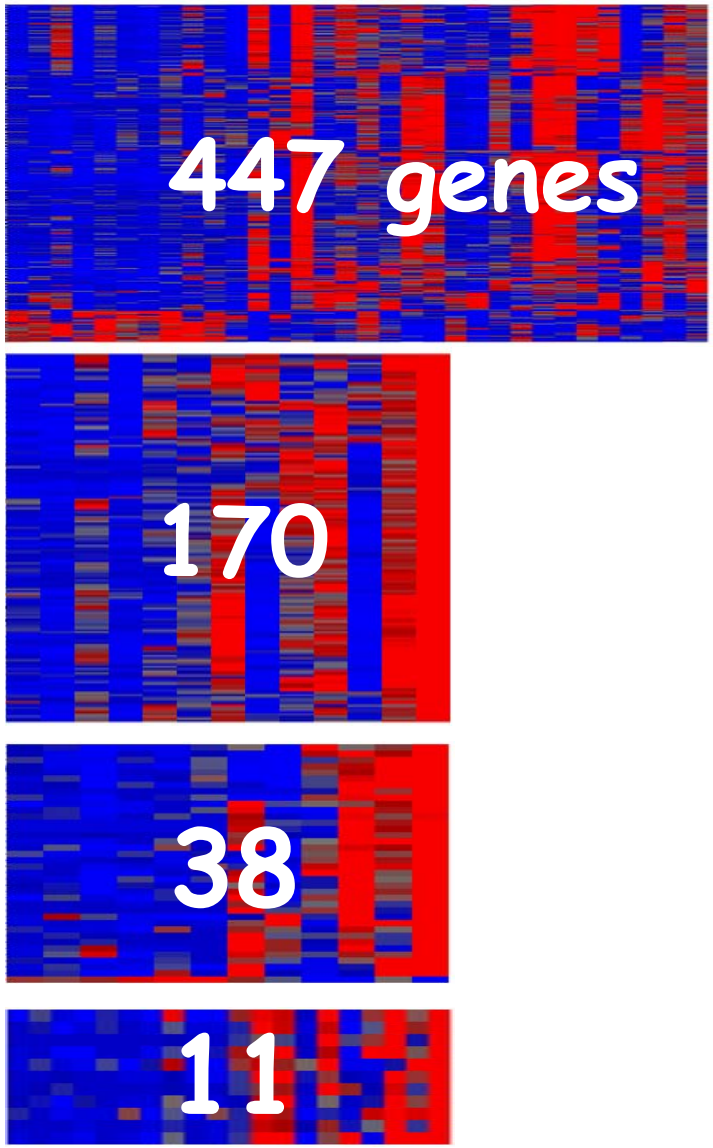


Evidence for Common Regulatory Control of the Arsenic-Associated genes??

Transcription factor binding site analysis



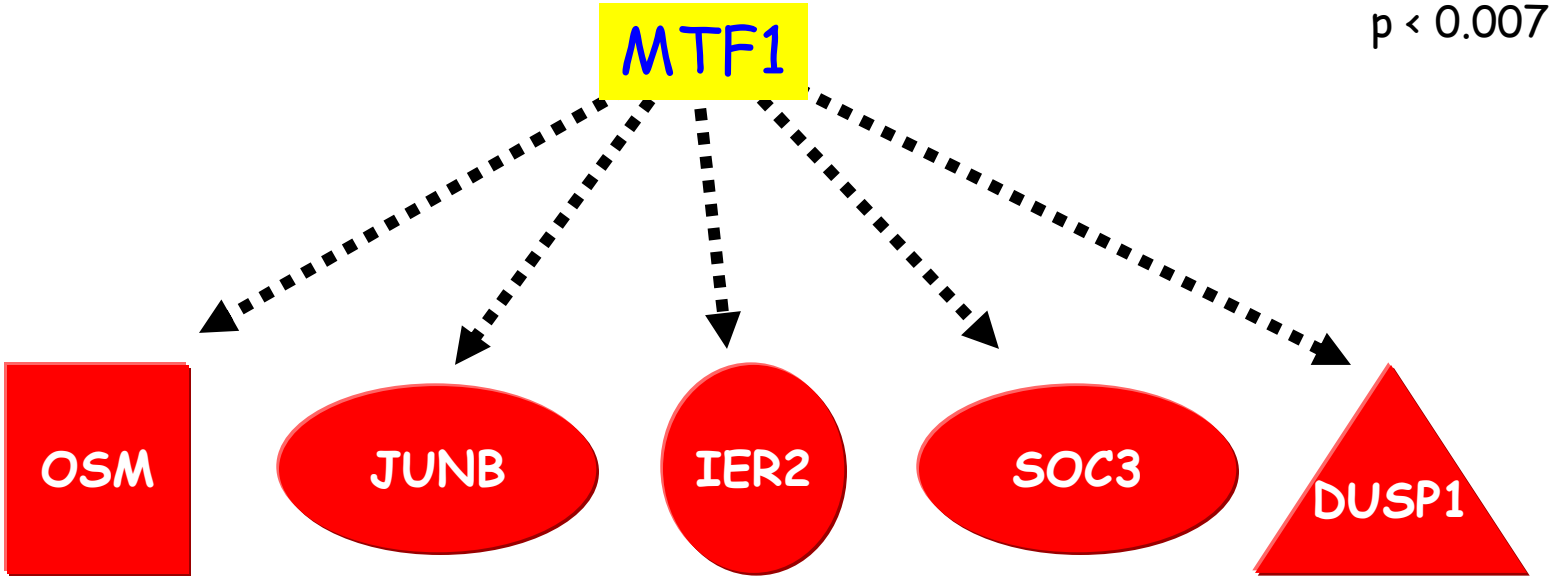
Binding sites for three transcription factors show significant enrichment



SRF	NF-κB	MTF1
$p<1.7\times10^{-5}$	$p<8.52\times10^{-6}$	$p<0.02$
$p<8.46\times10^{-4}$	$p<1.39\times10^{-4}$	$p<0.02$
$p<0.01$	$p<0.01$	$p<0.054$
$p<0.01$	0.01	$p<0.007$

Metal Responsive Transcription Factor-1

5 of 11 Biomarkers have MRE binding sites
Metal responsive element (MRE) TGCRCNC



MTF1 known to be activated by arsenic in animal models

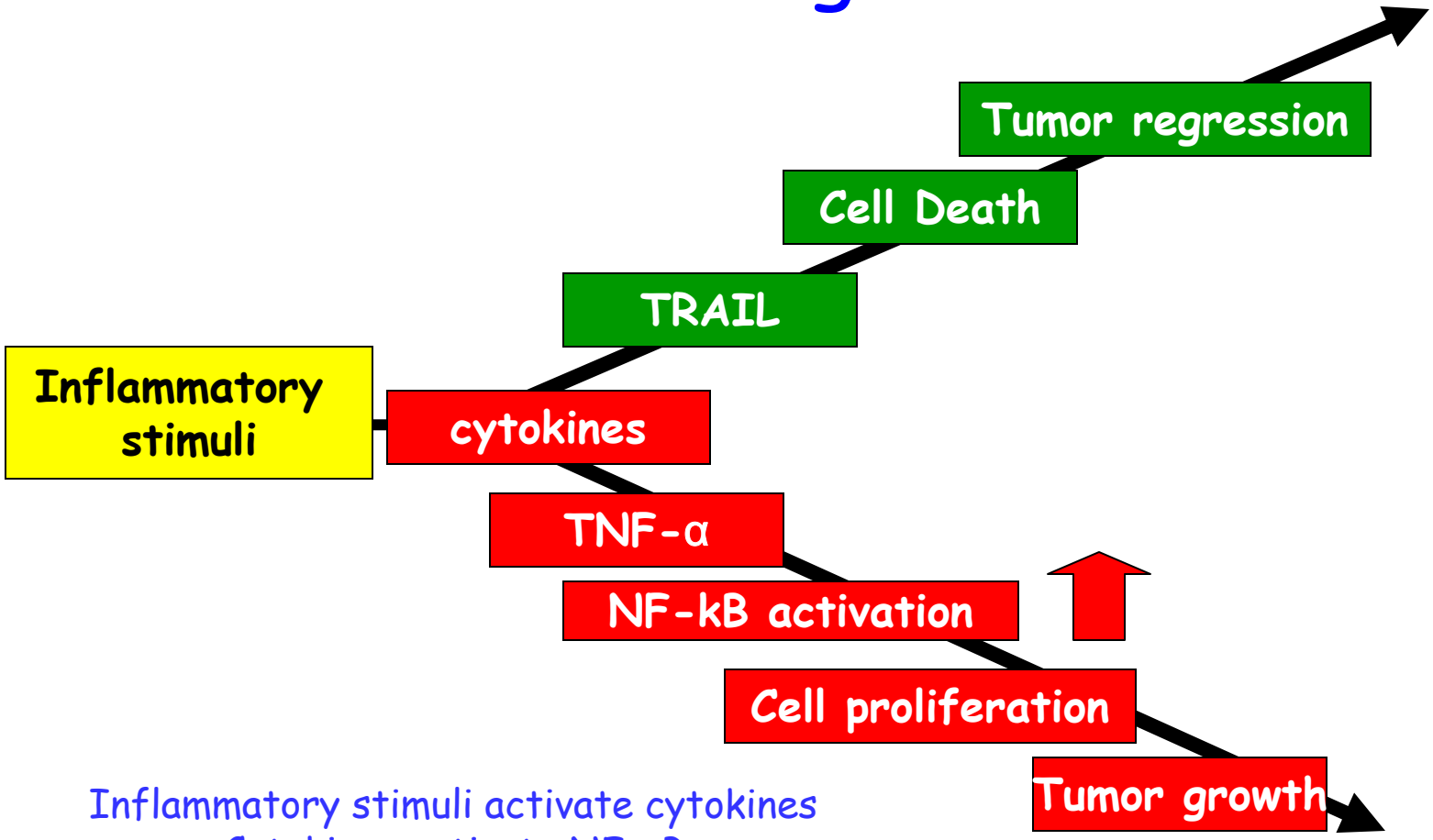
Liu et al Tox Sci 2001

Kumagai, Ann Rev Pharm Tox 2006

SRF: Serum Response Factor, binds to a serum response element (SRE) associated with immediate early genes such as c-fos, fosB, junB*, egr-1*

NF-κB, or Nuclear Factor kappaB, is involved in cellular responses to stress

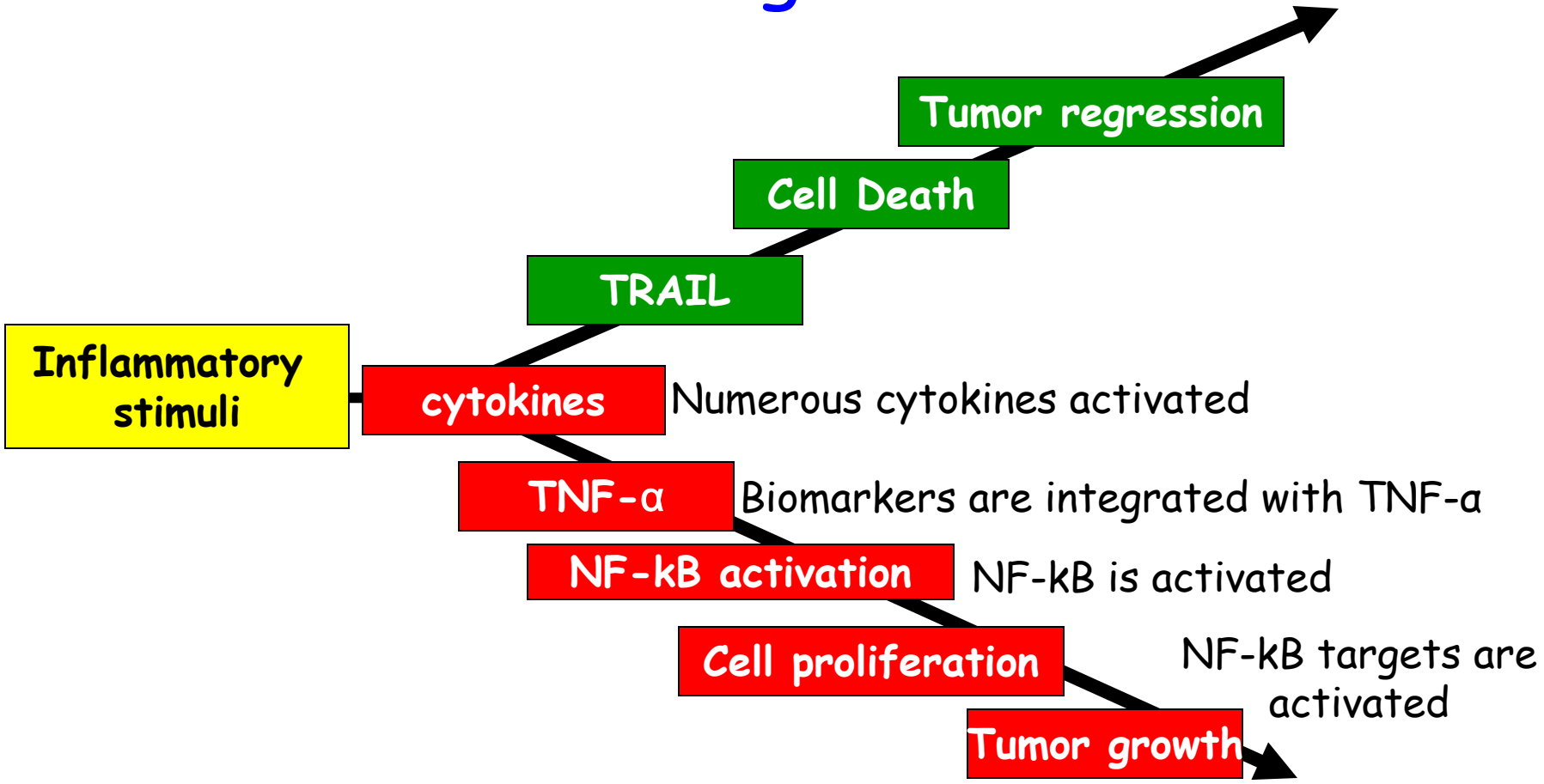
NF-κB activation is associated with tumorigenesis



Inflammatory stimuli activate cytokines
Cytokines activate NF-κB
NF-κB activates anti-apoptotic machinery
and is linked to tumor growth

Luo, Kamata, and Karin
The Journal of Clinical Investigation, 2005

NF-κB activation is associated with tumorigenesis



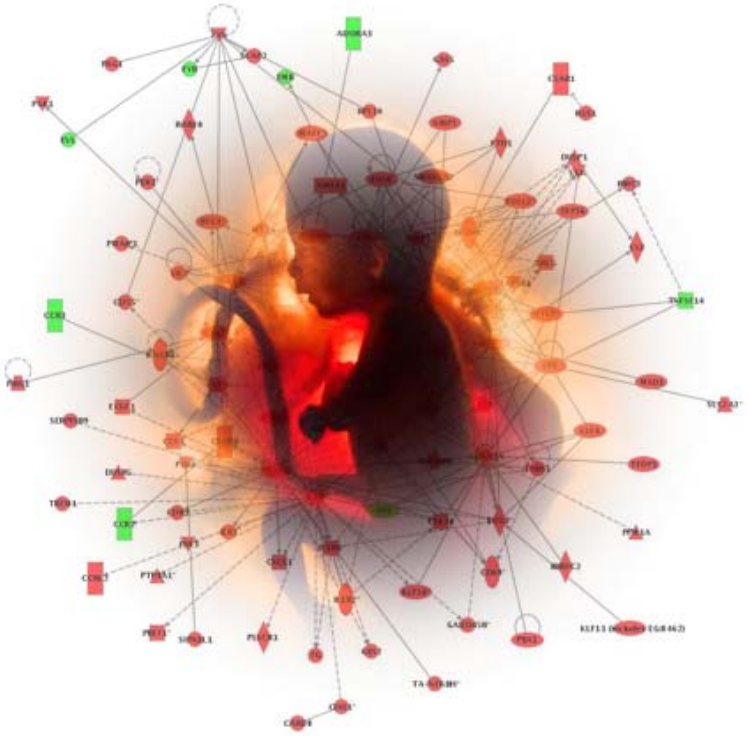
Luo, Kamata, and Karin
The Journal of Clinical Investigation, 2005

PRENATAL Arsenic Exposure Modulates Genes Involved in Inflammatory Response and Activates NF-kB Cascade

Robust genome-wide response to prenatal arsenic exposure

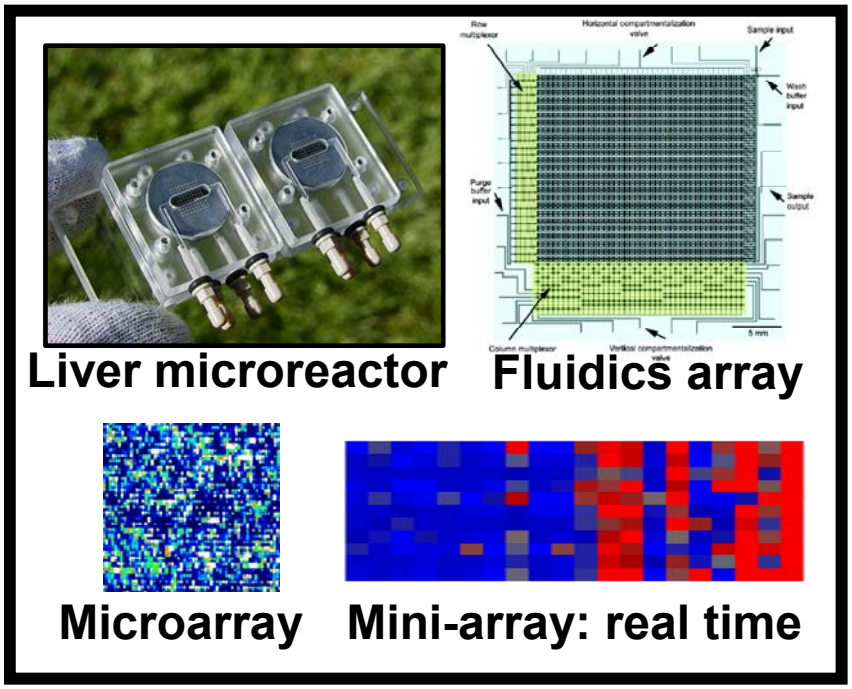
We can identify arsenic-associated gene sets that classify prenatal arsenic exposure

These genes map onto ontologies that include numerous processes including cell signaling, stress response and apoptosis



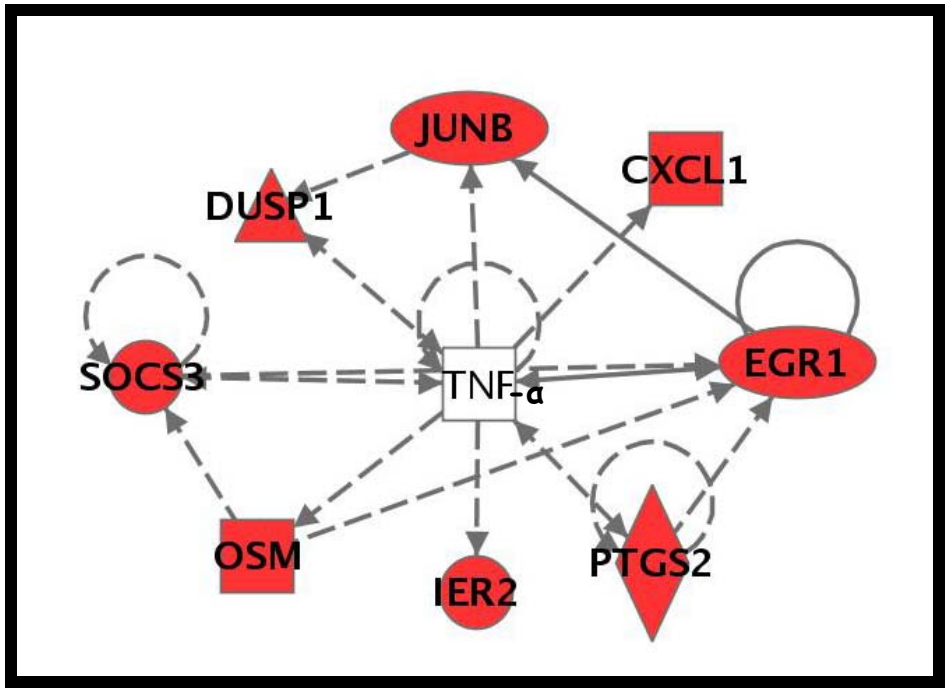
Increasing understanding of biological outcome of exposure and tool development

tools



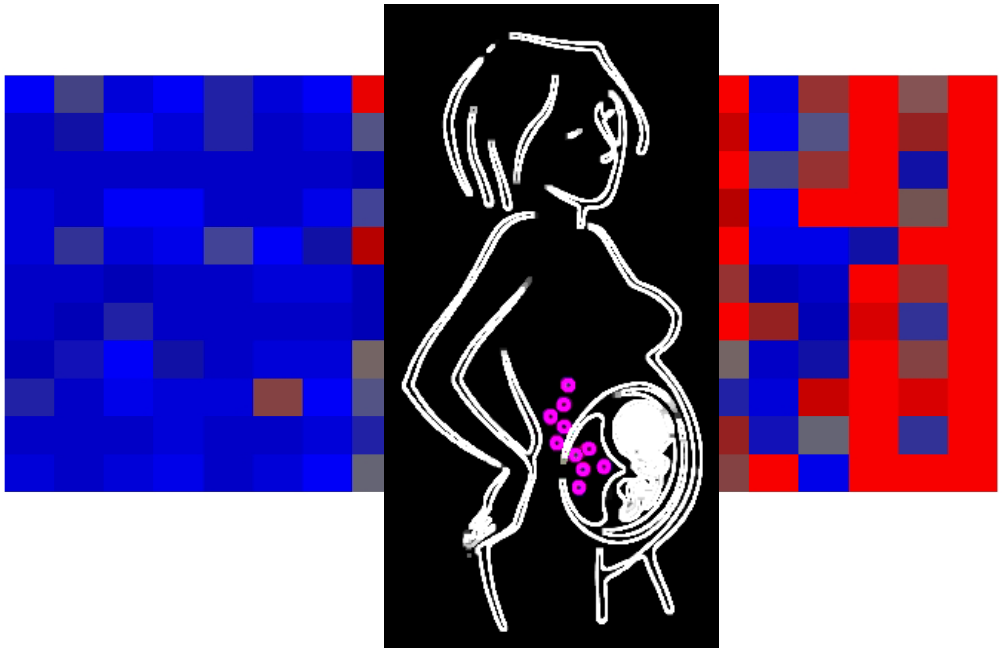
Biomarkers of Environmental exposures

Biological outcome of exposure



Mechanistic Insight

Can the Arsenic Biomarker genes classify prenatal arsenic exposure in a separate population??



Acknowledgements

Chandni Valiathan
Peter Svensson
Brad Hogan
Manlin Luo
Sanchita Bhattacharya
Leona Samson

Panida Navasumrit
Krittinee Kandjanapa
Sumitri Soontararuks
Sumonotha Nookabkaew
Chulabhorn Mahidol
Mathuros Ruchirawat

OPEN ACCESS Freely available online

PLoS GENETICS

Activation of Inflammation/NF-κB Signaling in Infants Born to Arsenic-Exposed Mothers

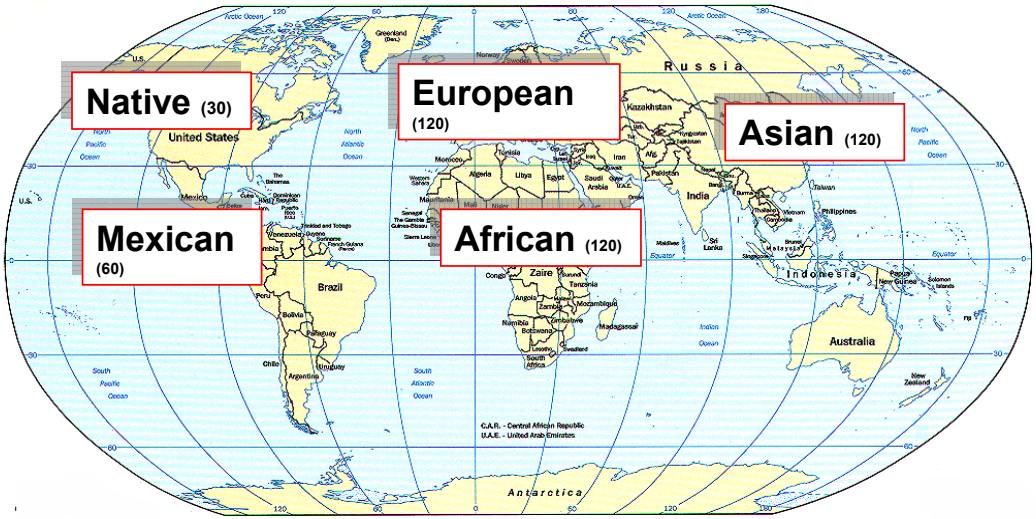
Rebecca C. Fry[1,2☯], Panida Navasumrit[3☯], Chandni Valiathan[1,2☯], J. Peter Svensson[1,2], Bradley J. Hogan[1,2], Manlin Luo[1,2], Sanchita Bhattacharya[1,2¤], Krittinee Kandjanapa[3], Sumitra Soontararuks[3], Sumontha Nookabkaew[3], Chulabhorn Mahidol[3], Mathuros Ruchirawat[3*], Leona D. Samson[1,2*]

Using genomics to predict population responses to exposures

Predict: How an individual will respond upon exposure?

Who will be sensitive to exposure?

Cell Lines Represent Healthy Genetically Diverse Population

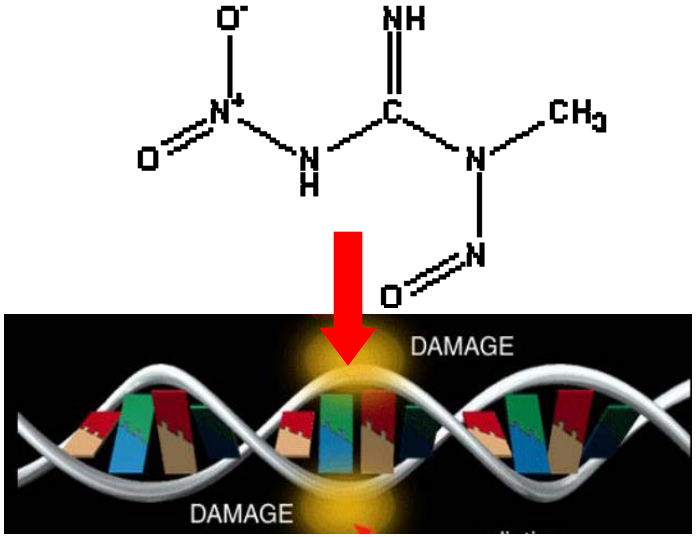


450 healthy, unrelated individuals

24 lymphoblastoid cell lines

Model DNA damaging agent

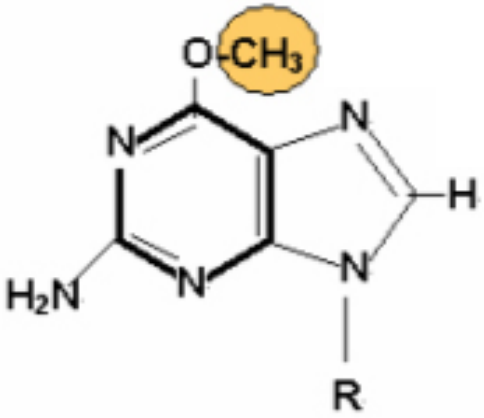
N-methyl *N*-Nitro *N*-Nitrosoguanidine



Damage induced mimics some chemotherapeutics

Damage induced mimics environmental exposures

O^6-meG can mispair with thymine
G/C to A/T transitions

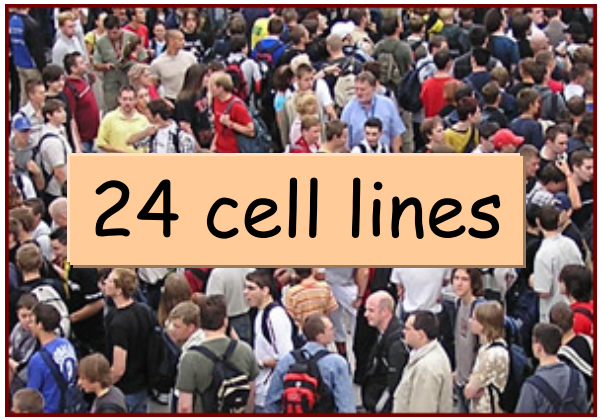


guanine —— **cytosine**

thymine

Can be cytotoxic or mutagenic lesion

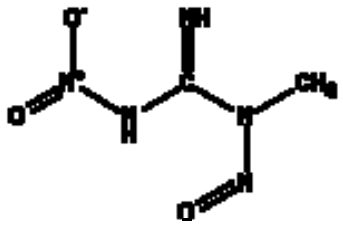
What is the extent of inter-individual variation in sensitivity to a DNA damaging agent?



Exposed or unexposed

+/-

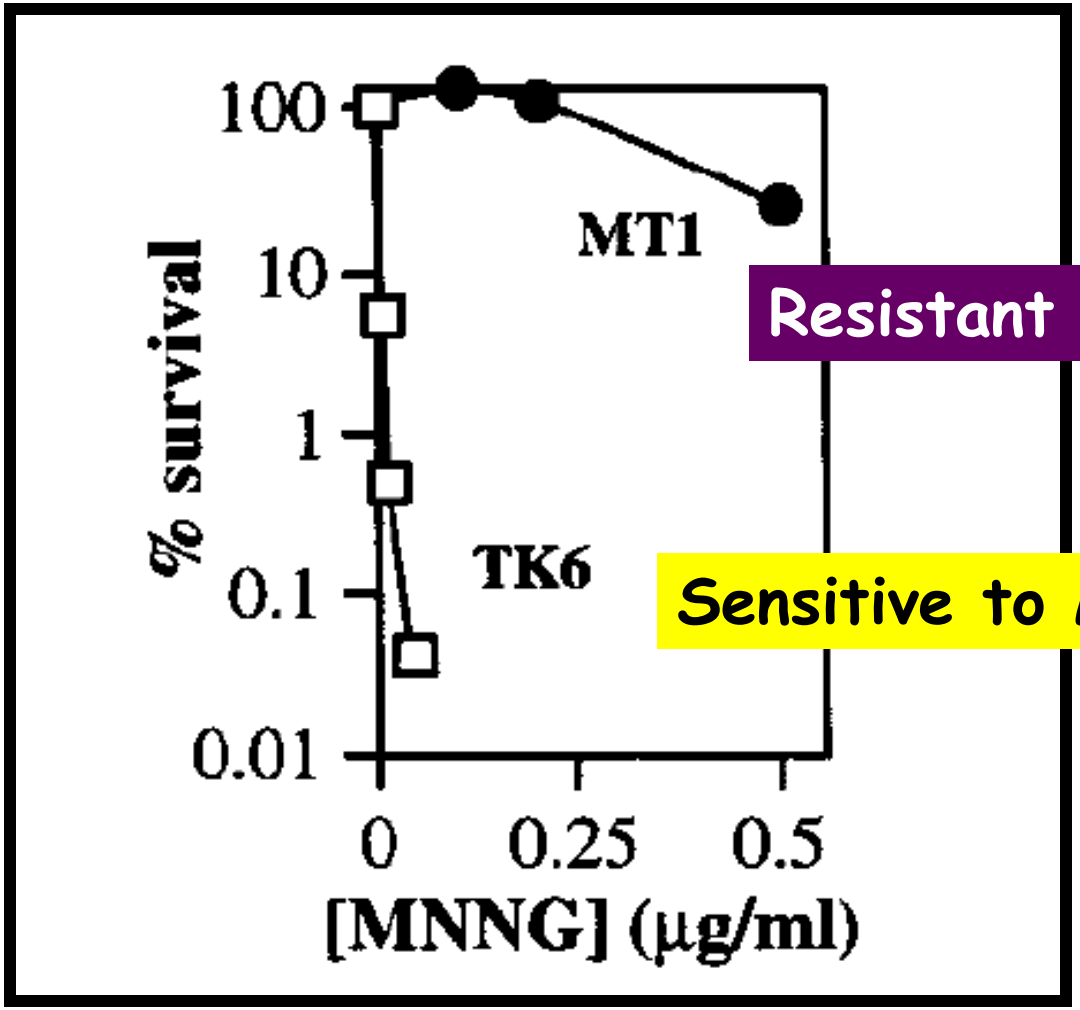
MNNG



0.5 ug/ml, 72hrs

Growth inhibition (sensitivity)

Control cell lines included in screen

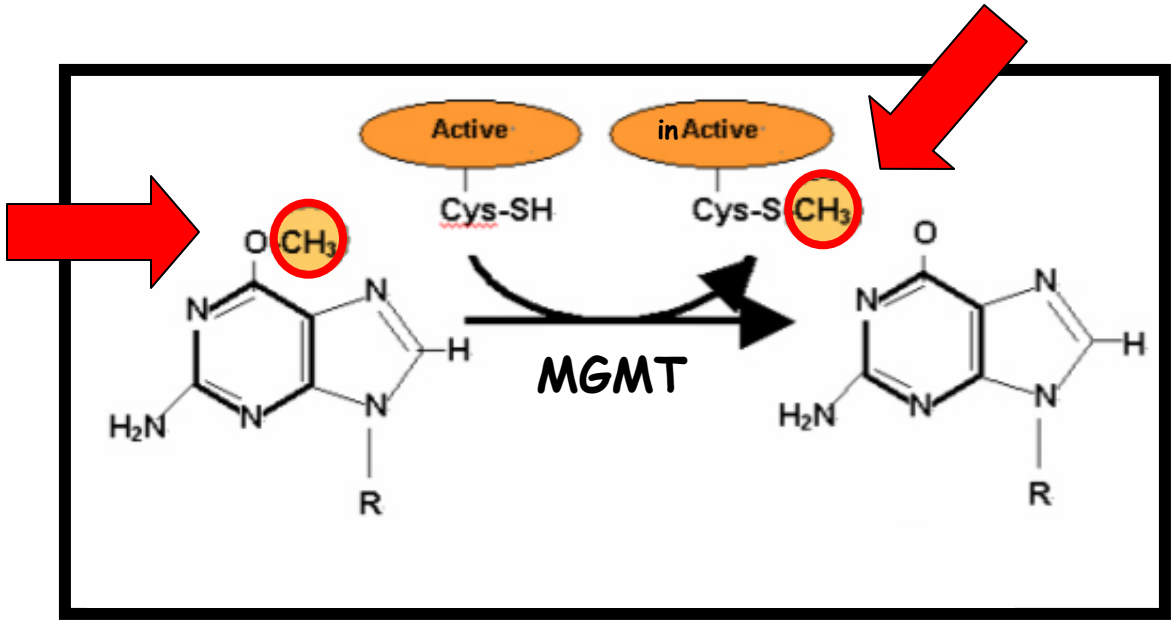


Resistant to MNNG

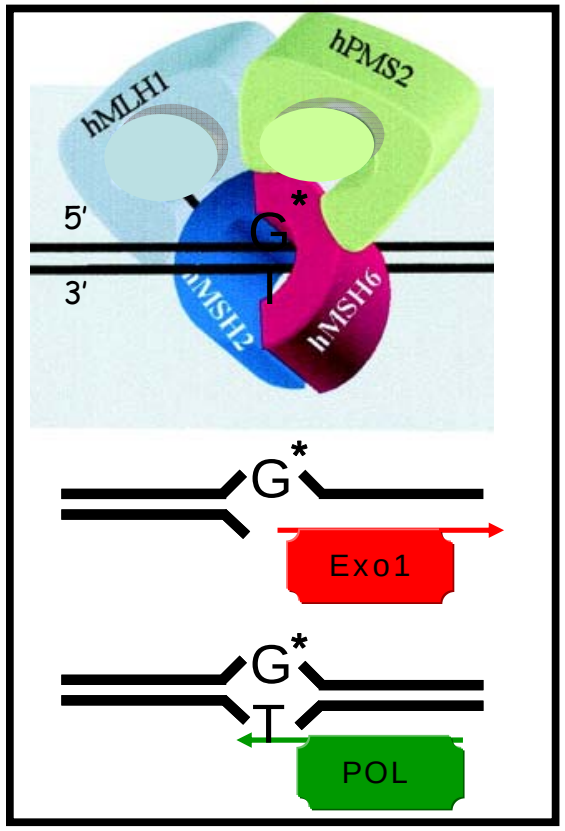
Sensitive to MNNG

Hickman and Samson, PNAS 1999

BOTH control cell lines lack MGMT: Direct Reversal of Damage



Different Mismatch Repair Status of control cell lines



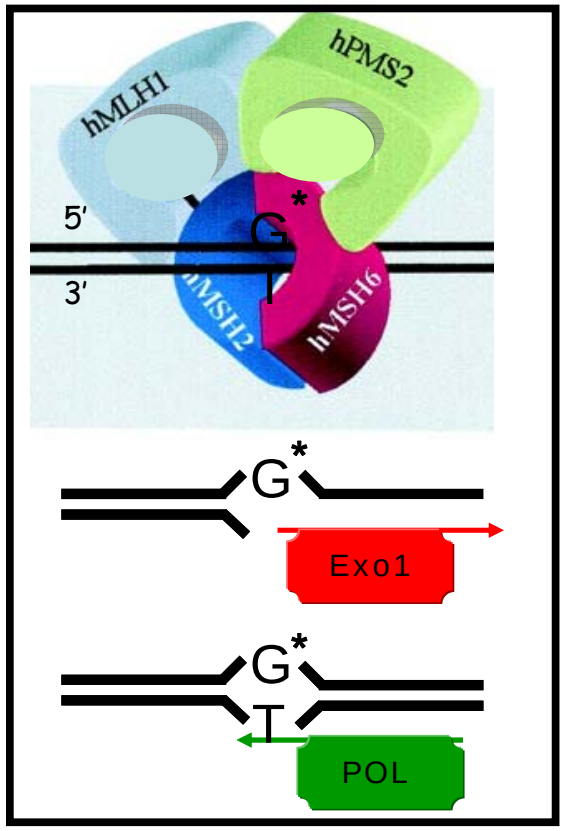
hMutLα
(MLH1 & PMS2)

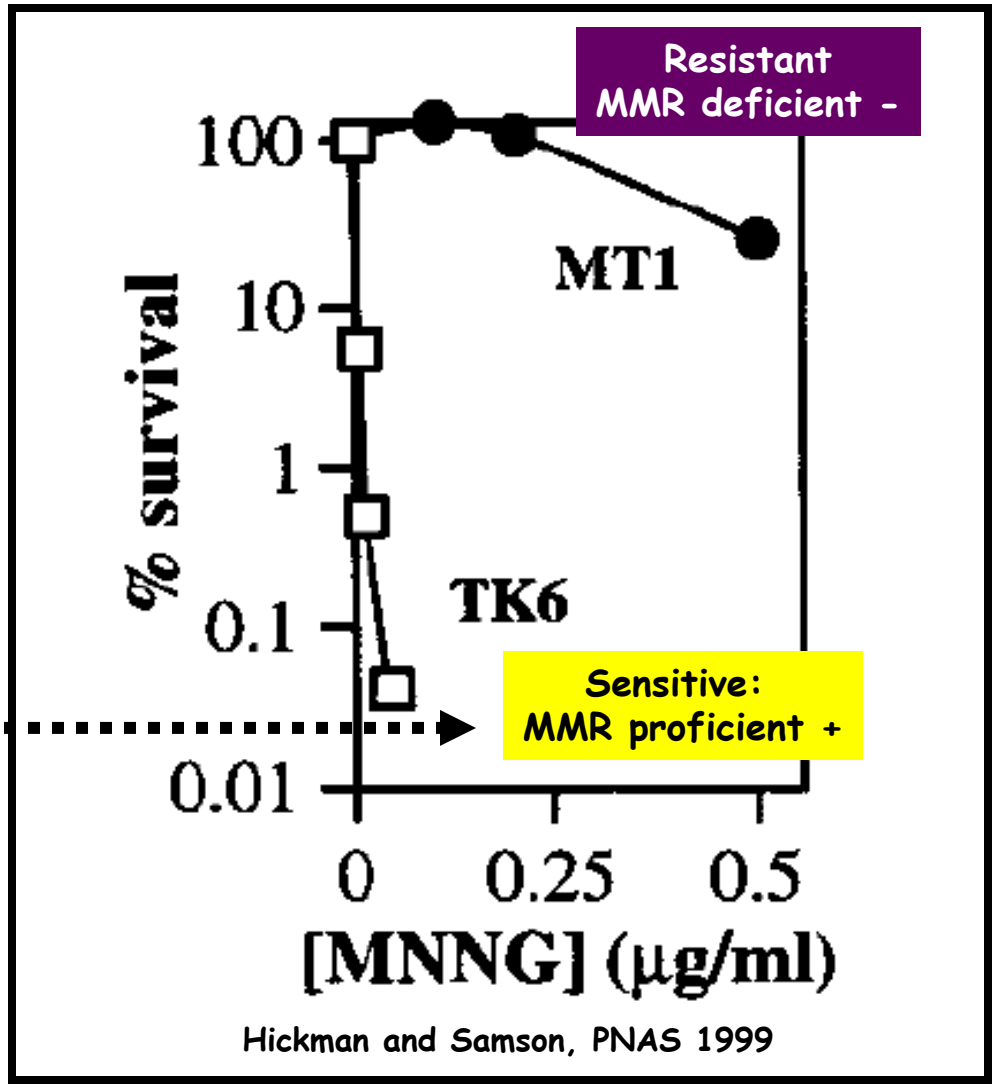
hMutSα
(MSH6 & MSH2)

excision via ExoI

repair synthesis via Pol delta

Paradox of Mismatch Repair





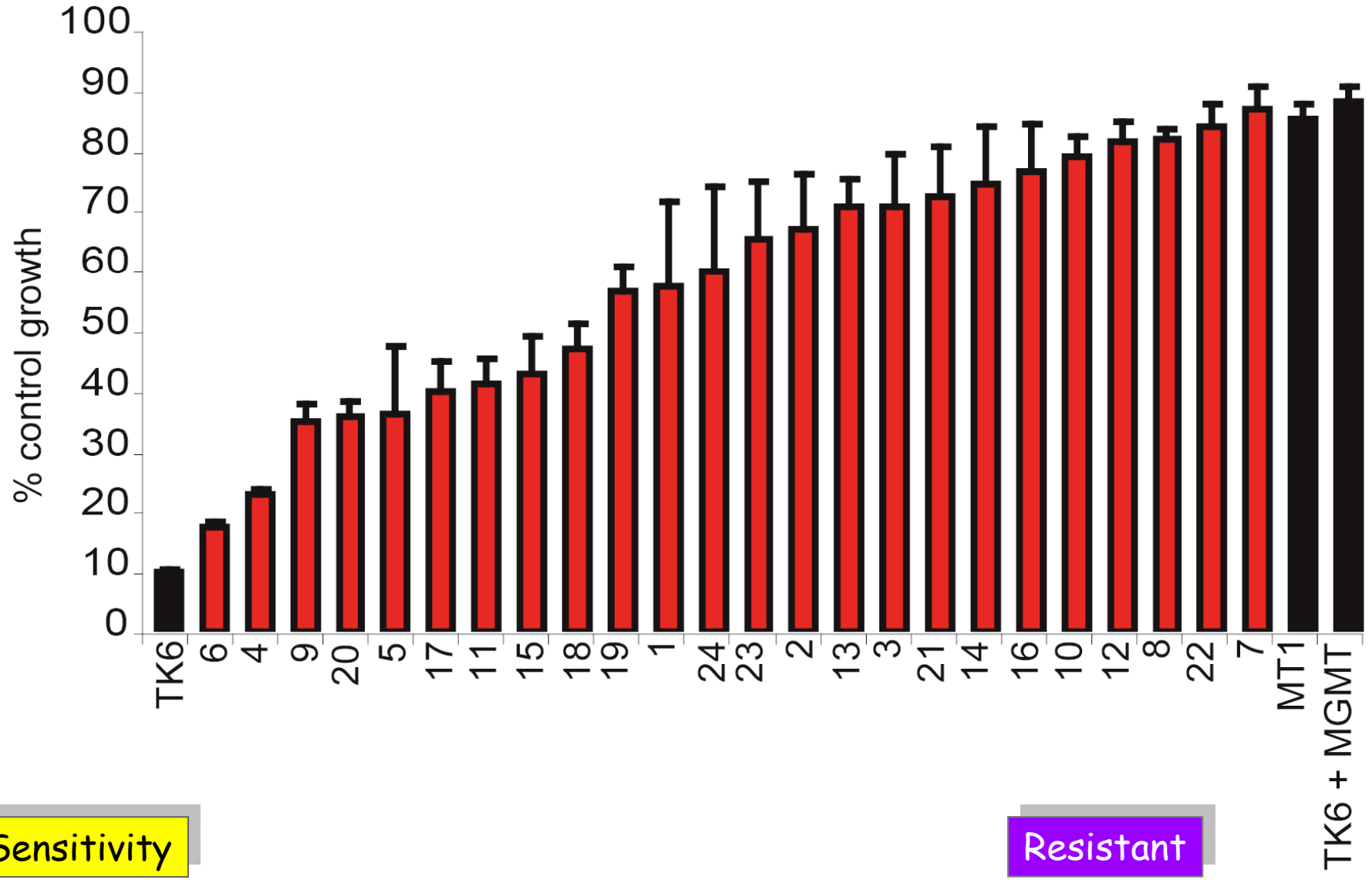
FUNCTIONAL MMR processing
O⁶MeG:T mismatches
Leads to cell death!!

Establish Range of Sensitivity in Cells Exposed to MNNG

Sensitivity

Resistant

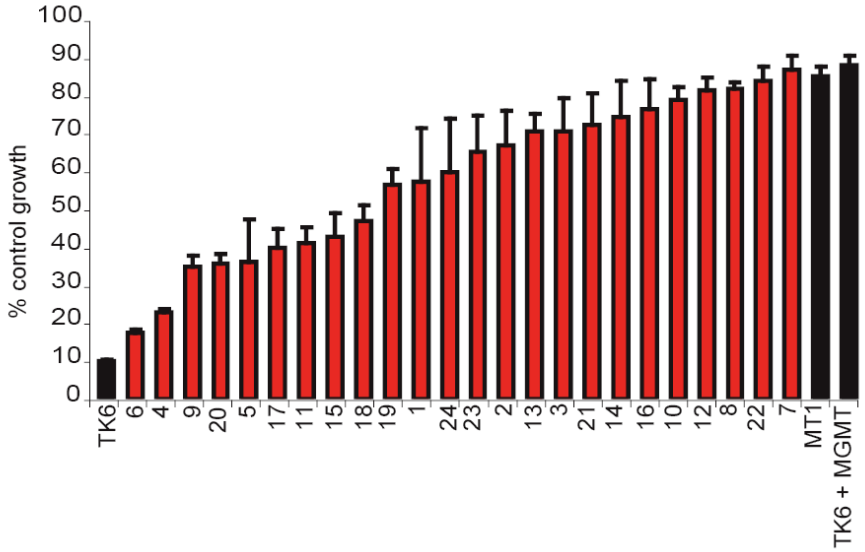
Establish Range of Sensitivity in Cells Exposed to MNNG

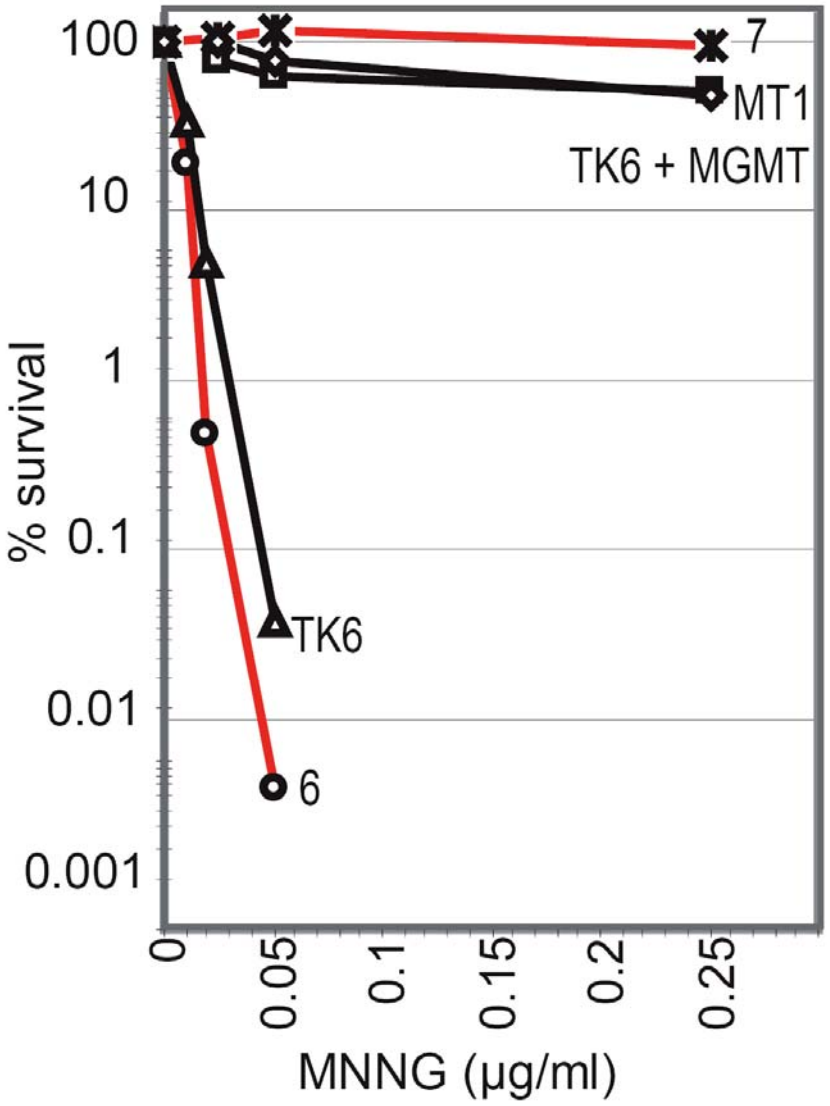


Sensitivity

Resistant

Killing Curve Establishes Range of Sensitivity

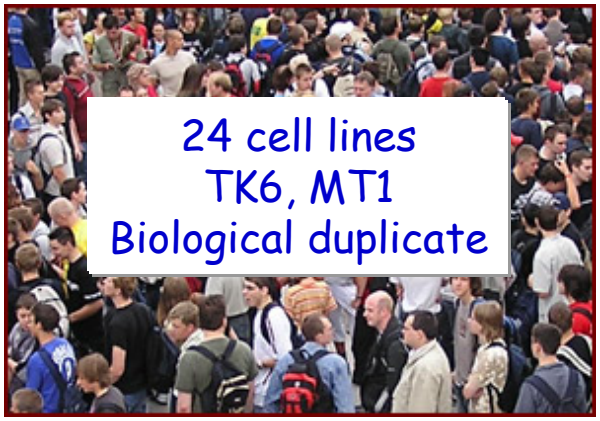




Can we predict the sensitivity of the cells upon exposure to MNNG ??

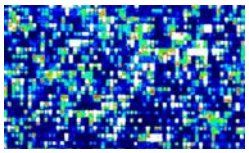
Two-class prediction algorithm

First step towards prediction: Integrate genome-wide expression data with sensitivity data



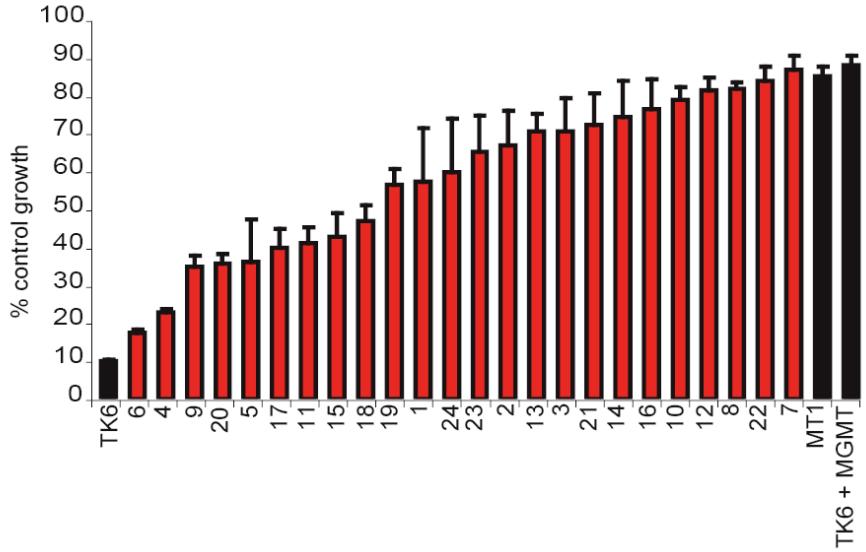
MNNG
0.5 ug/ml, 72hrs

– or –



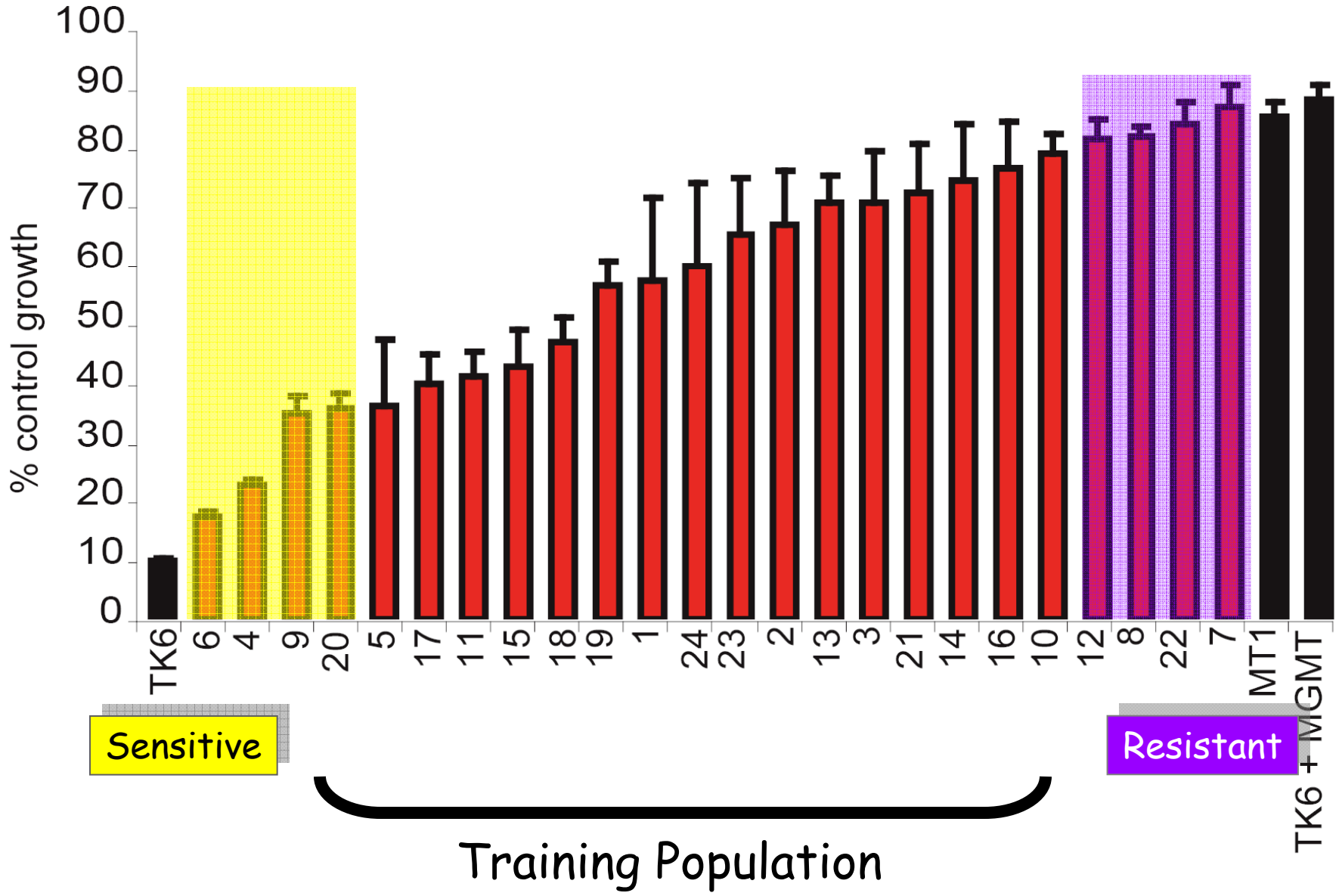
Unexposed (basal)
Exposed (treated)

Sensitivity data

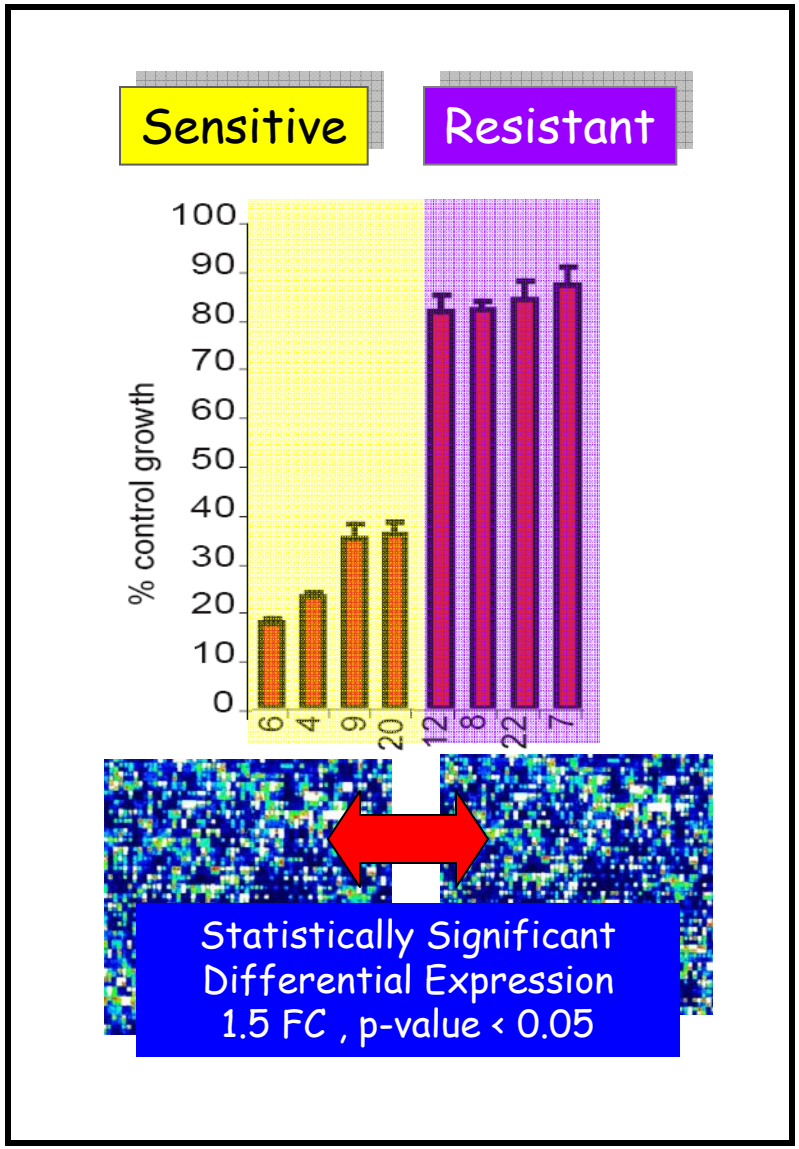


Establish a training population

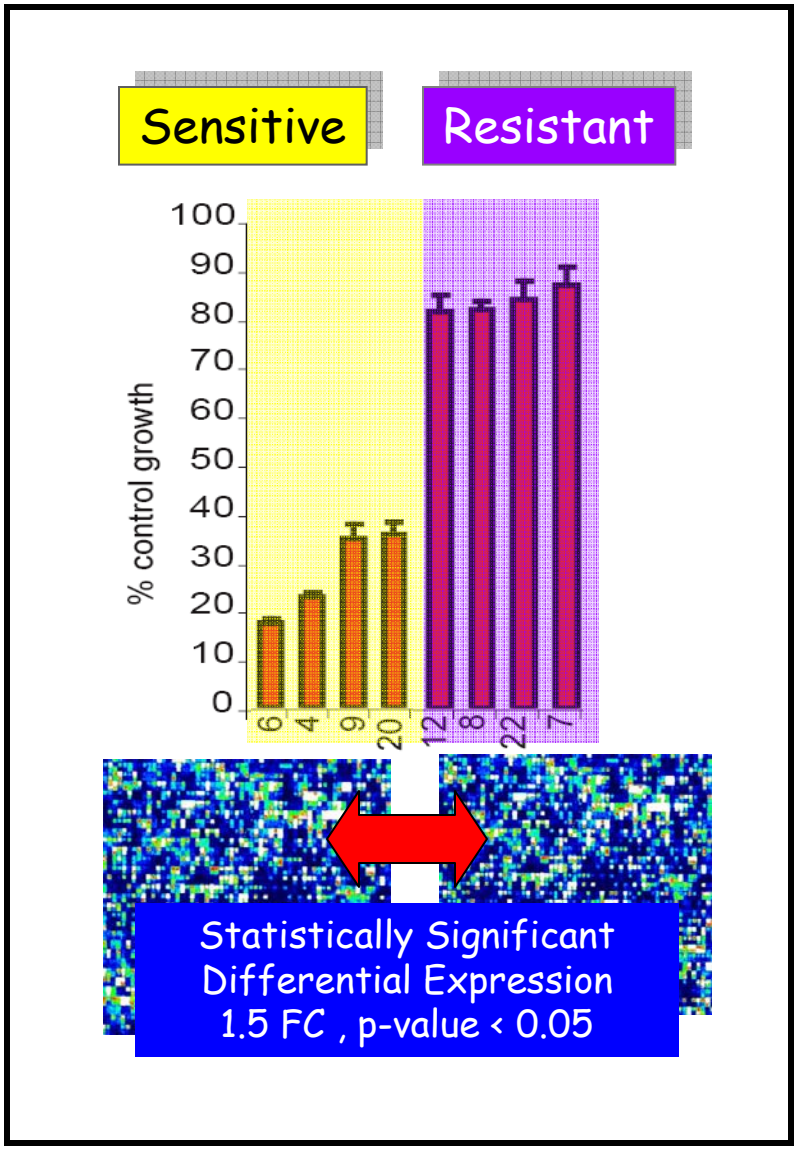
Establish Training Population based on extreme responders

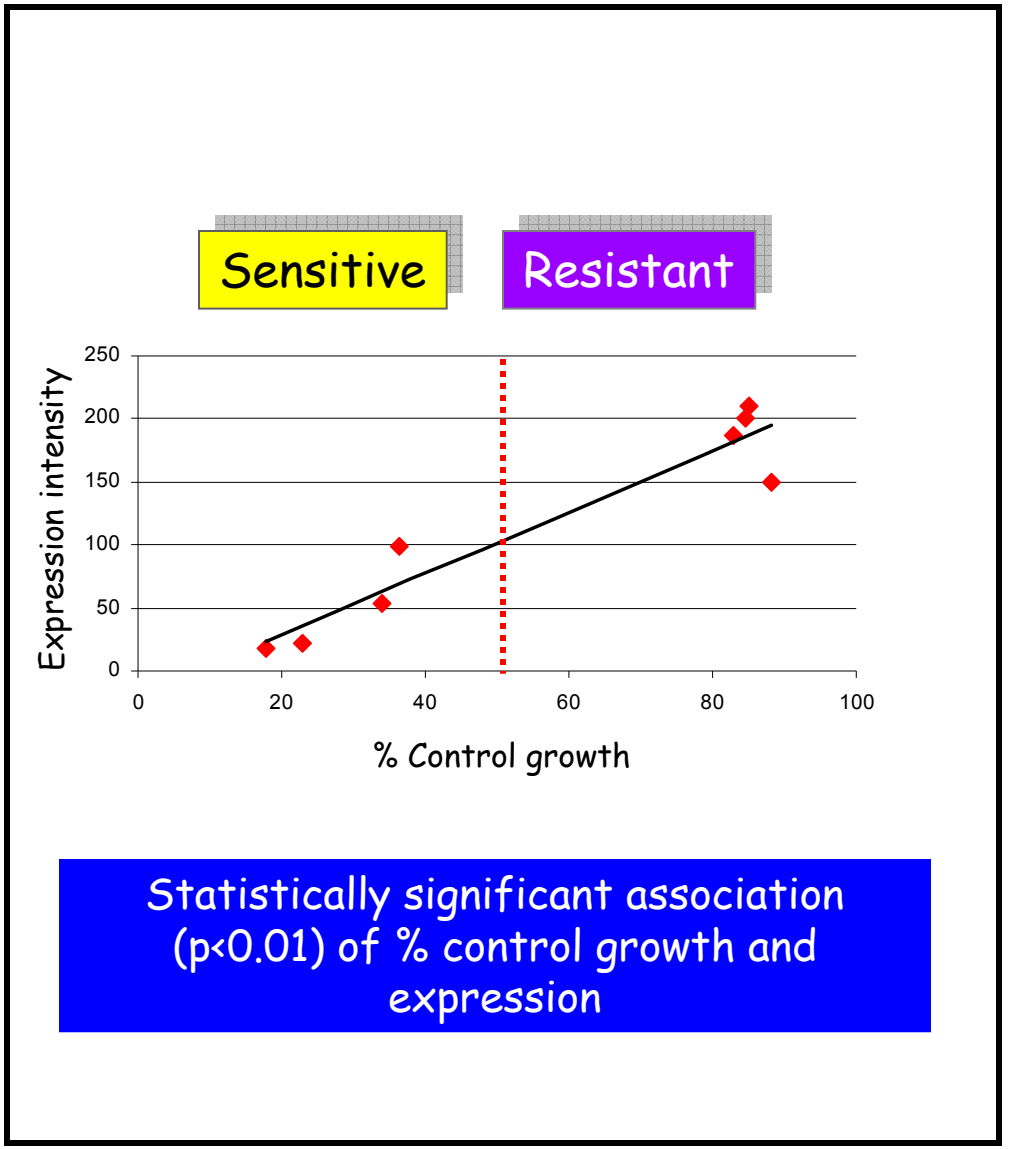


Alkylation-Sensitivity-Associated Gene Sets Identified by Integrating % Control Growth and Gene Expression



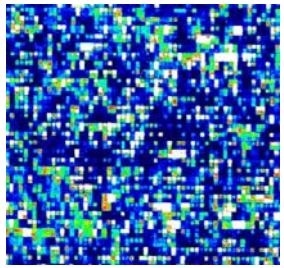
Alkylation-Sensitivity-Associated Gene Sets Identified by Integrating % Control Growth and Gene Expression



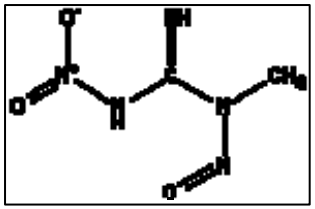


Basal and MNNG-induced transcripts were linked with sensitivity data

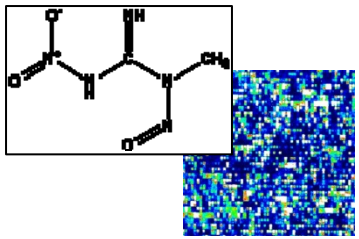
Alkylation-Sensitivity-Associated Gene Sets



Basal

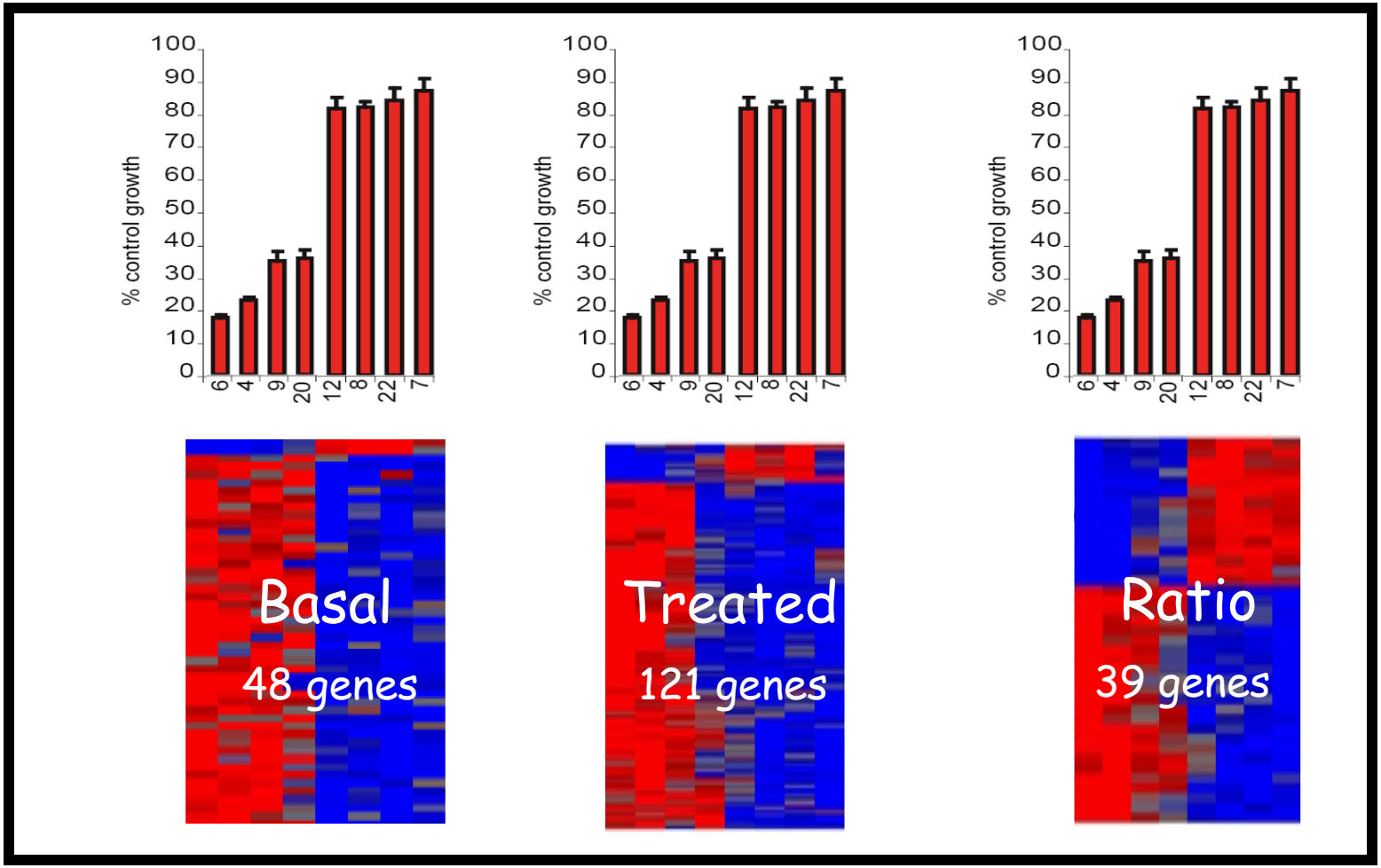


Treated

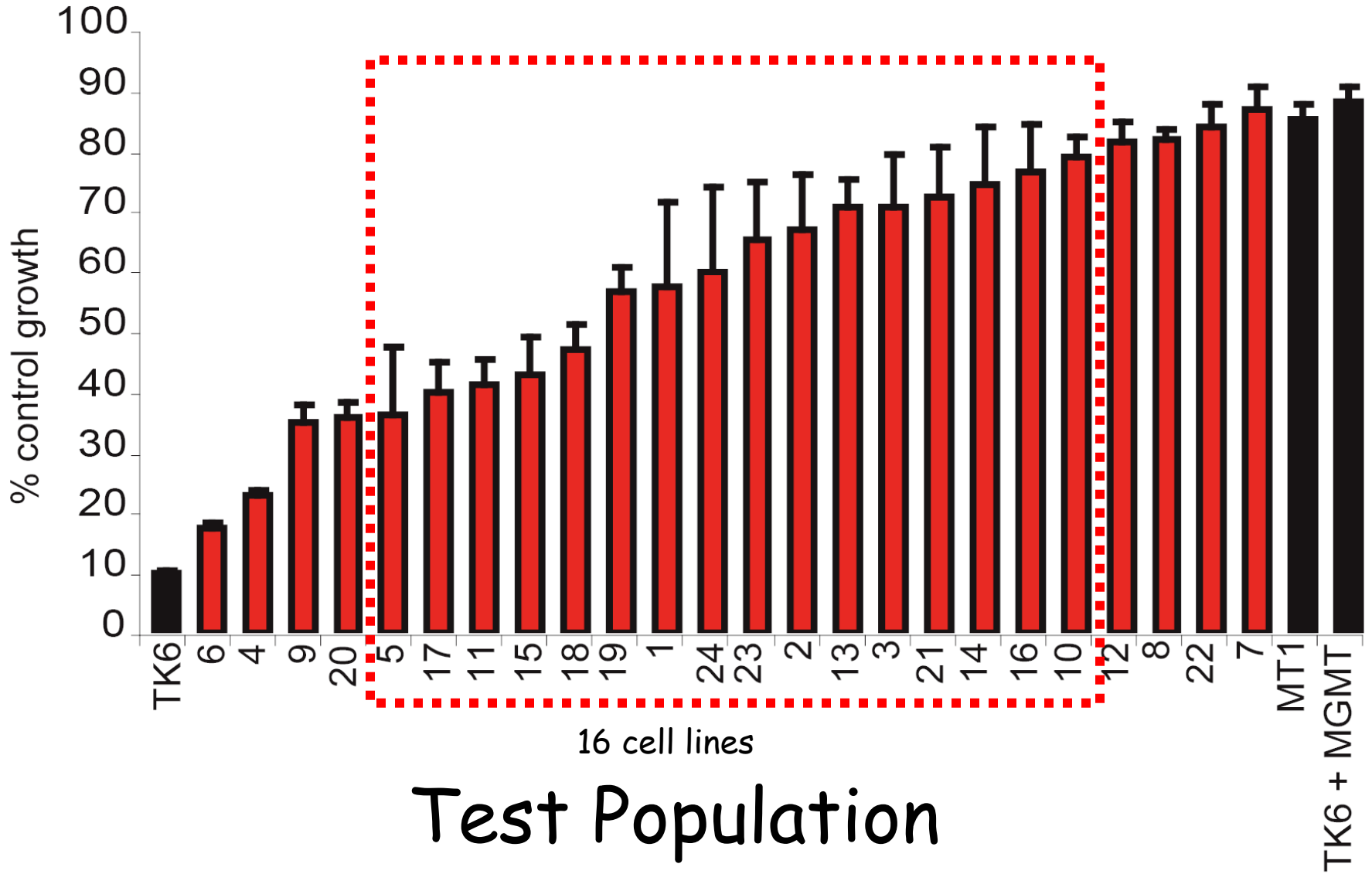


Ratio

Three ASA Gene Sets Identified

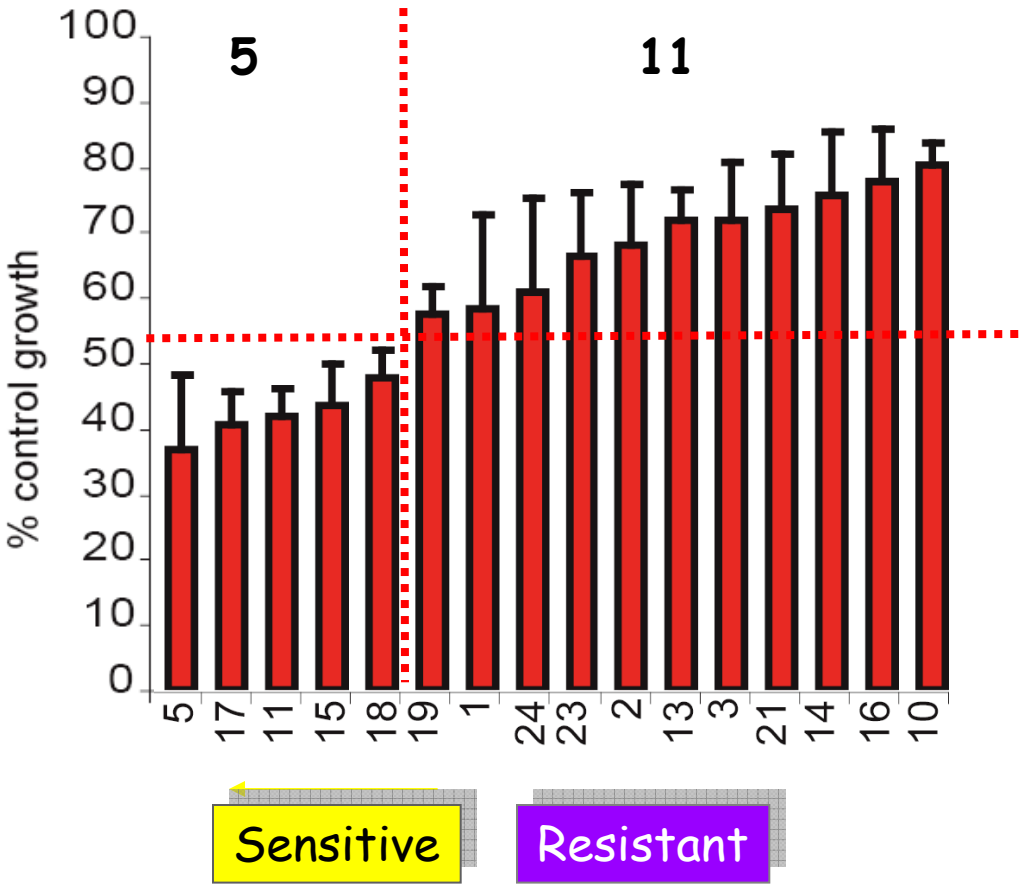


Can the ASA gene sets PREDICT MNNG sensitivity in the Test Population?



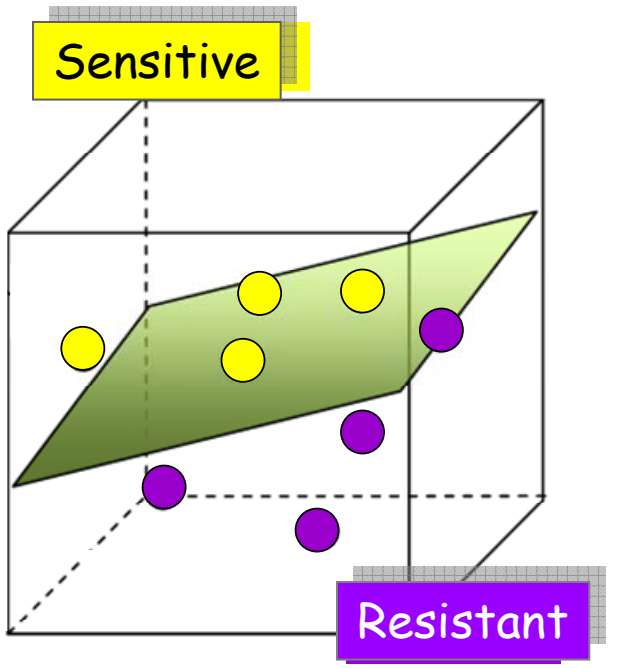
Test Population

Can the ASA gene sets PREDICT MNNG sensitivity in the Test Population?

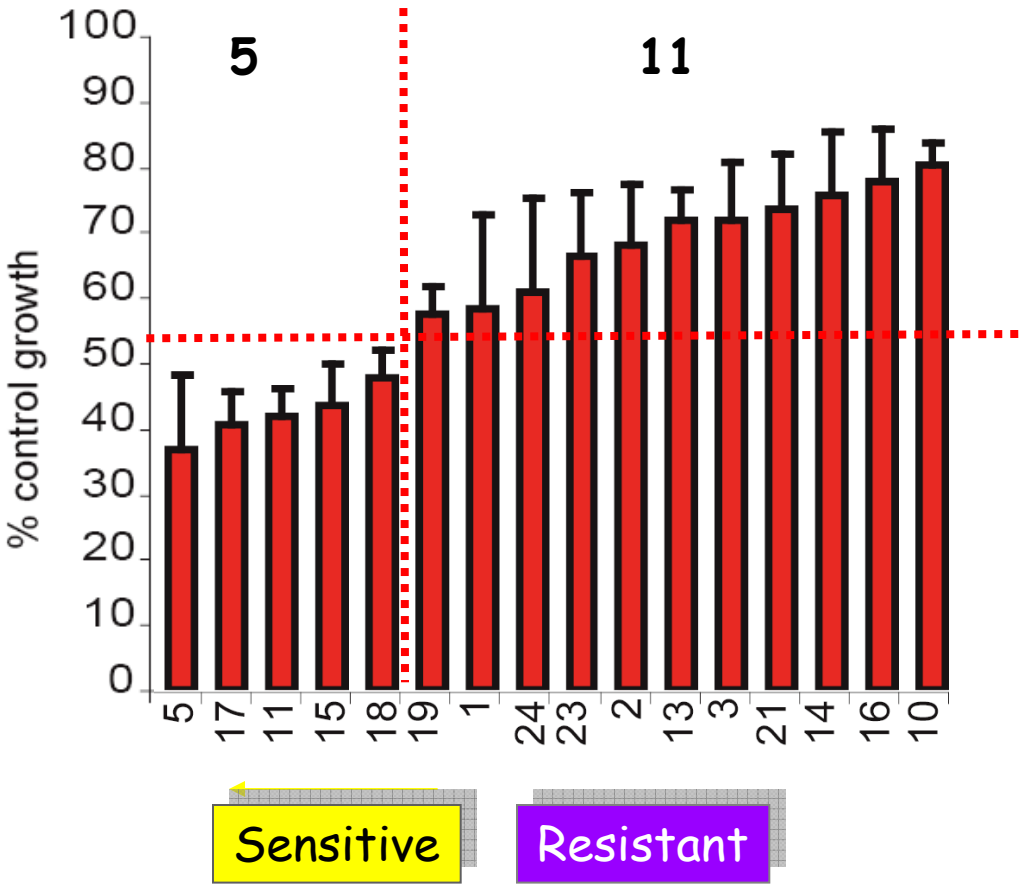


16 cell lines: Test Population

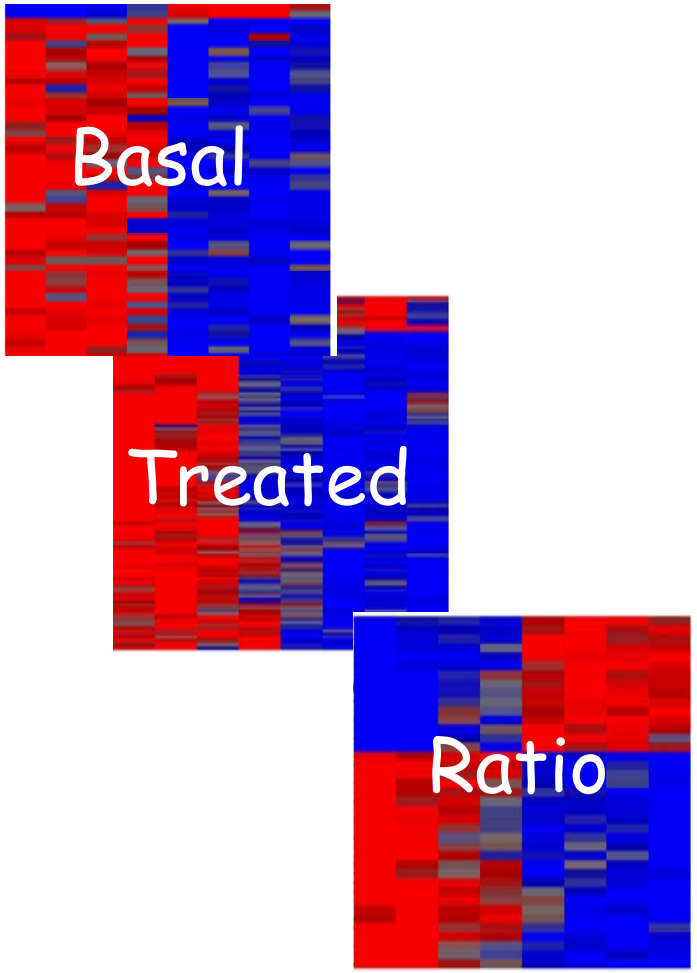
Support Vector Machine



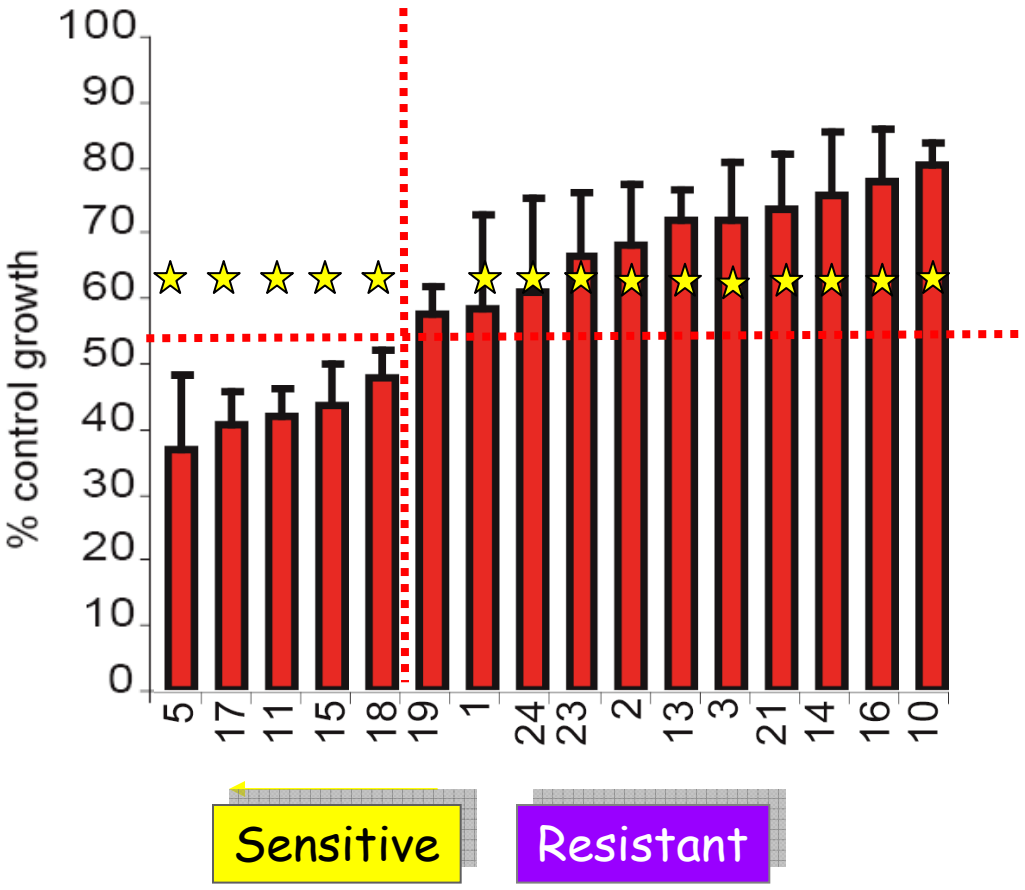
Can the ASA gene sets PREDICT MNNG sensitivity in the Test Population?



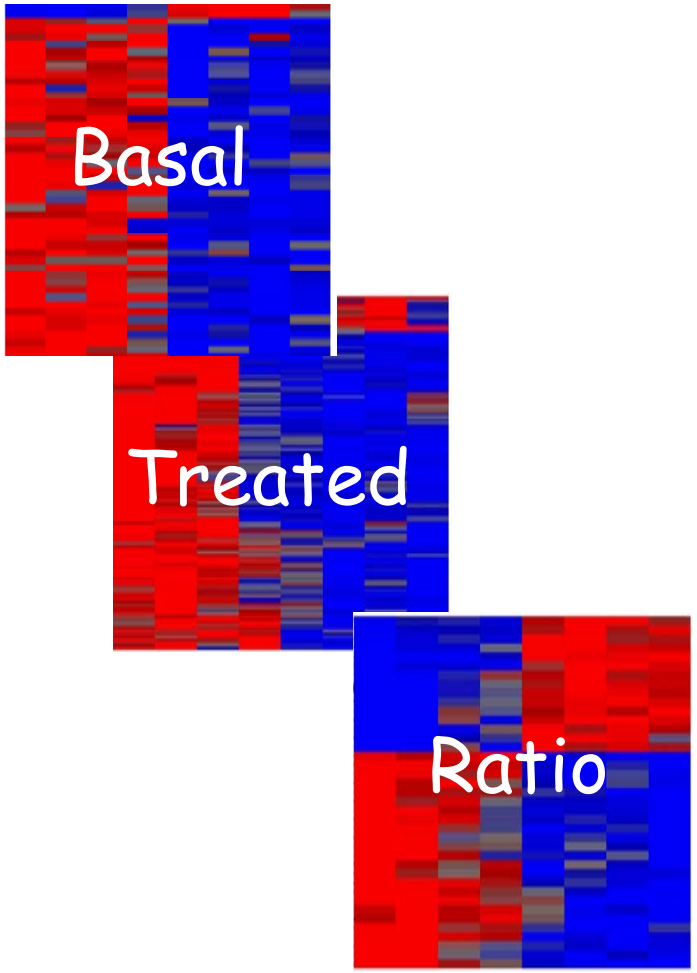
16 cell lines: Test Population



Can the ASA gene sets PREDICT MNNG sensitivity in the Test Population?



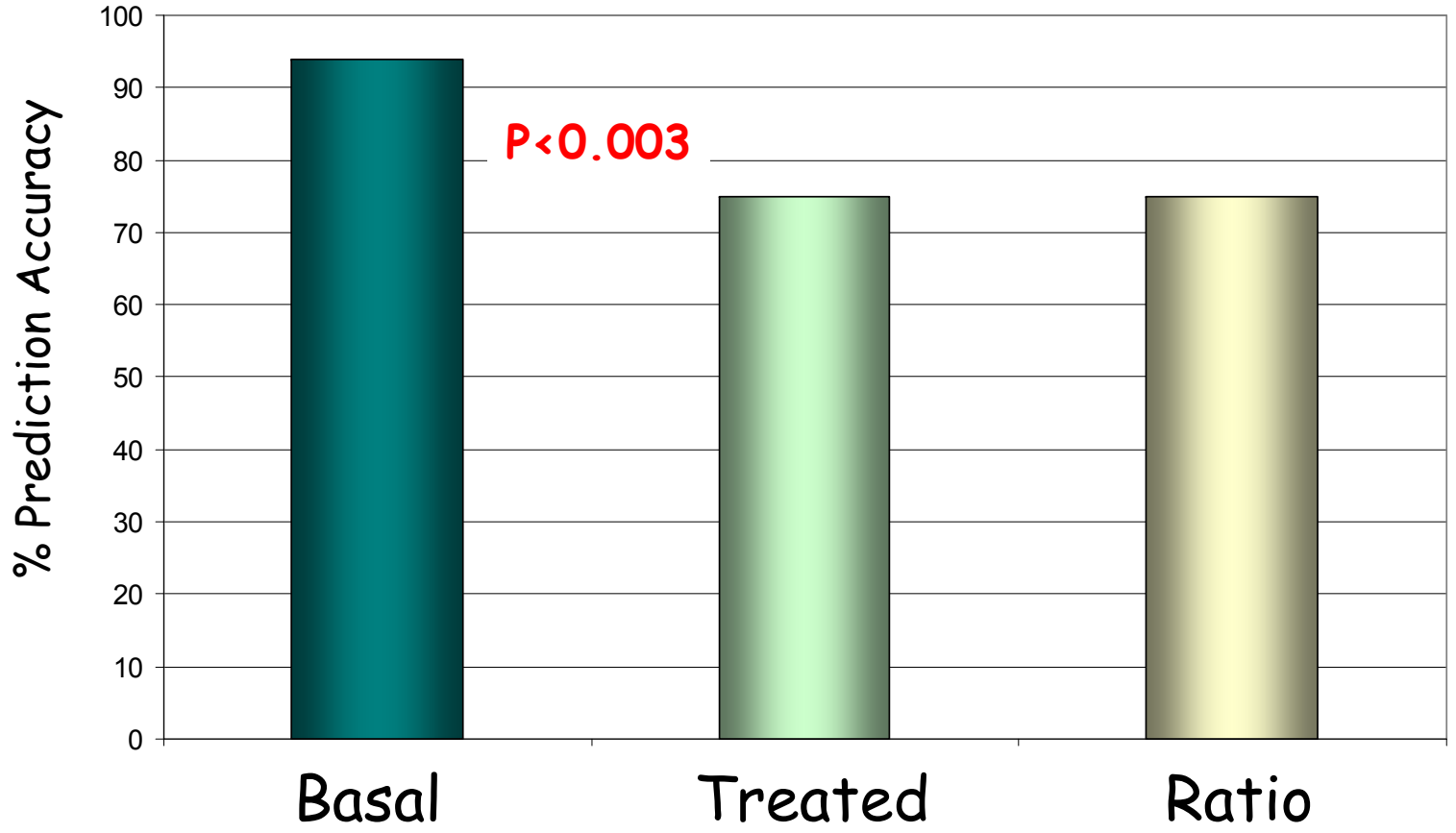
16 cell lines: Test Population



Basal gene expression is most predictive of sensitivity

94% accuracy

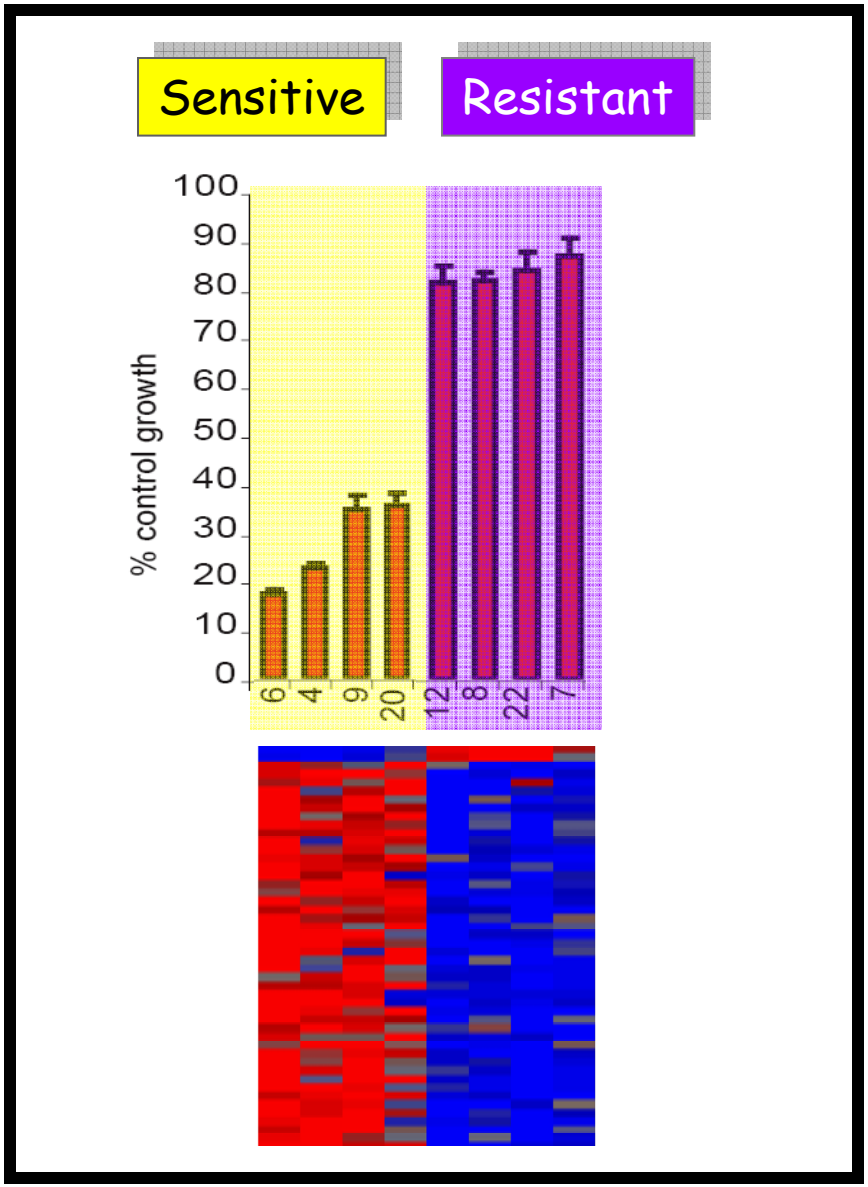
15/16 cell lines accurately predicted



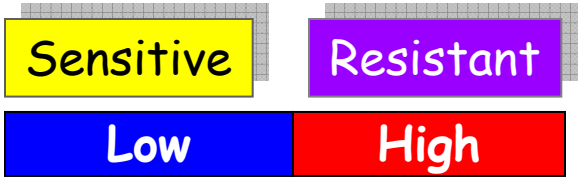
Which genes are contained in the highly predictive basal gene set?

Can we identify a potential molecular basis for the sensitivity?

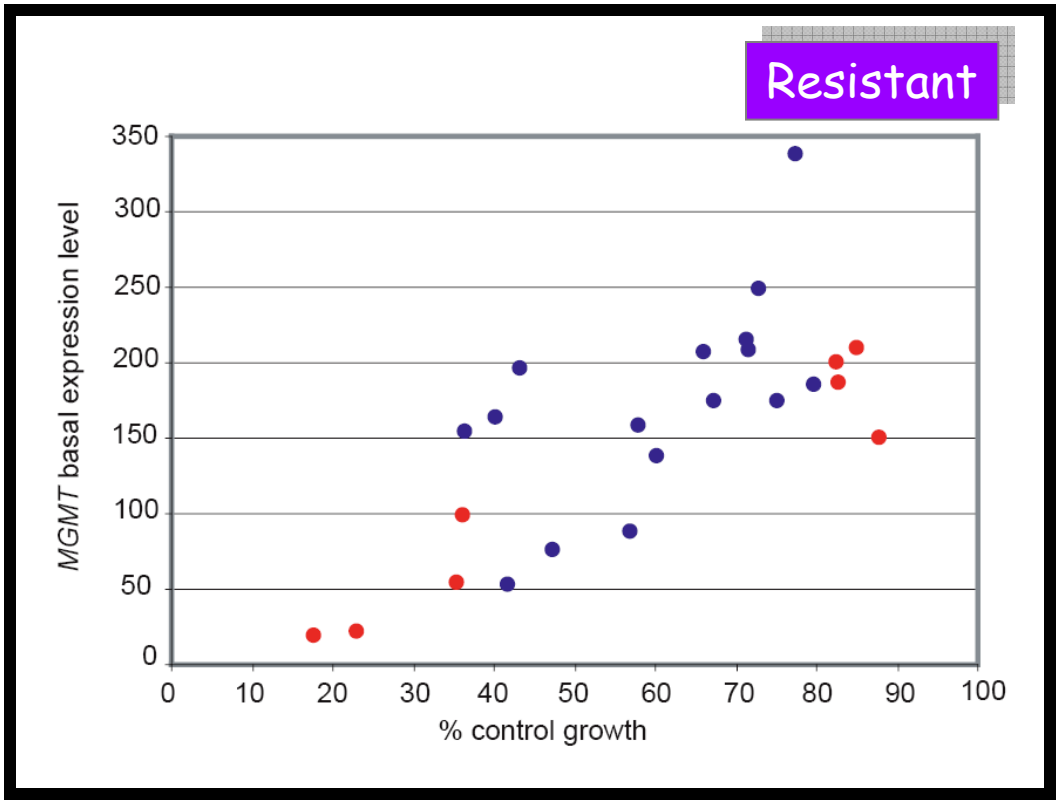
Two expression patterns in basal set



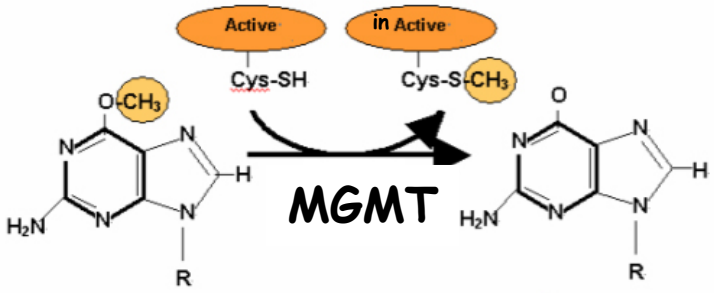
1: High expression in resistant cells



THE most significant positive association of *MGMT* expression with low sensitivity



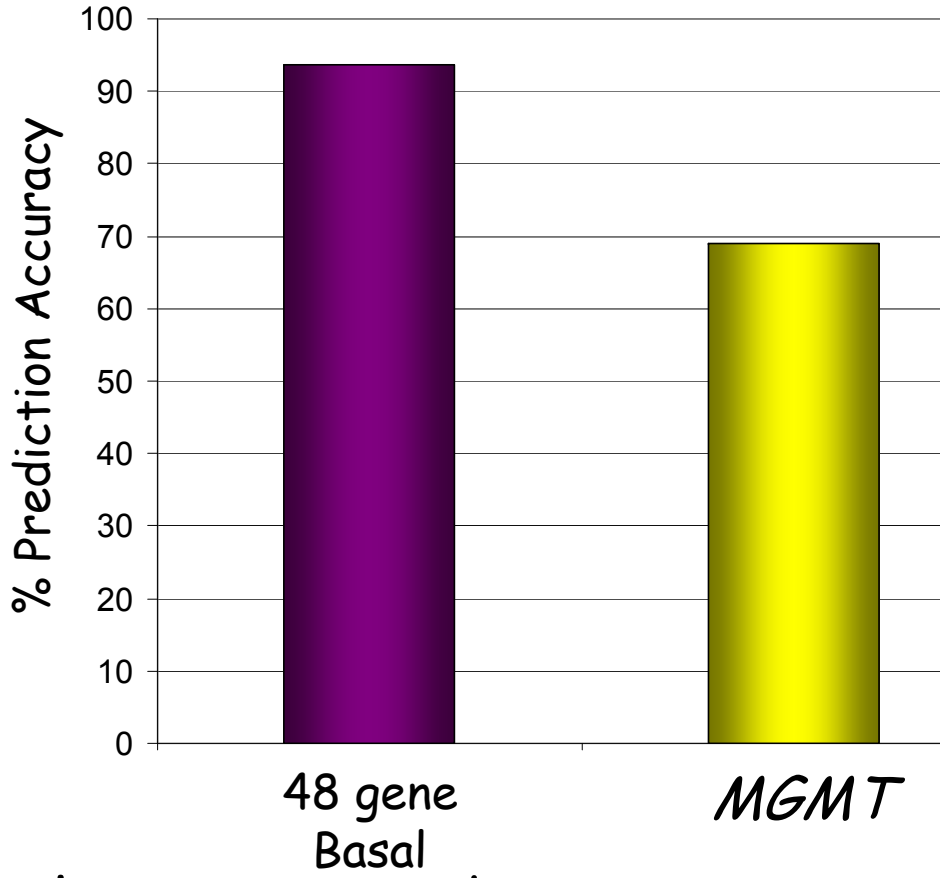
Unsupervised analysis uncovers DNA repair protein known to repair MNNG-induced damage



MGMT **activity known to vary among individuals**
MGMT **expression associated with resistance to certain chemotherapeutics**

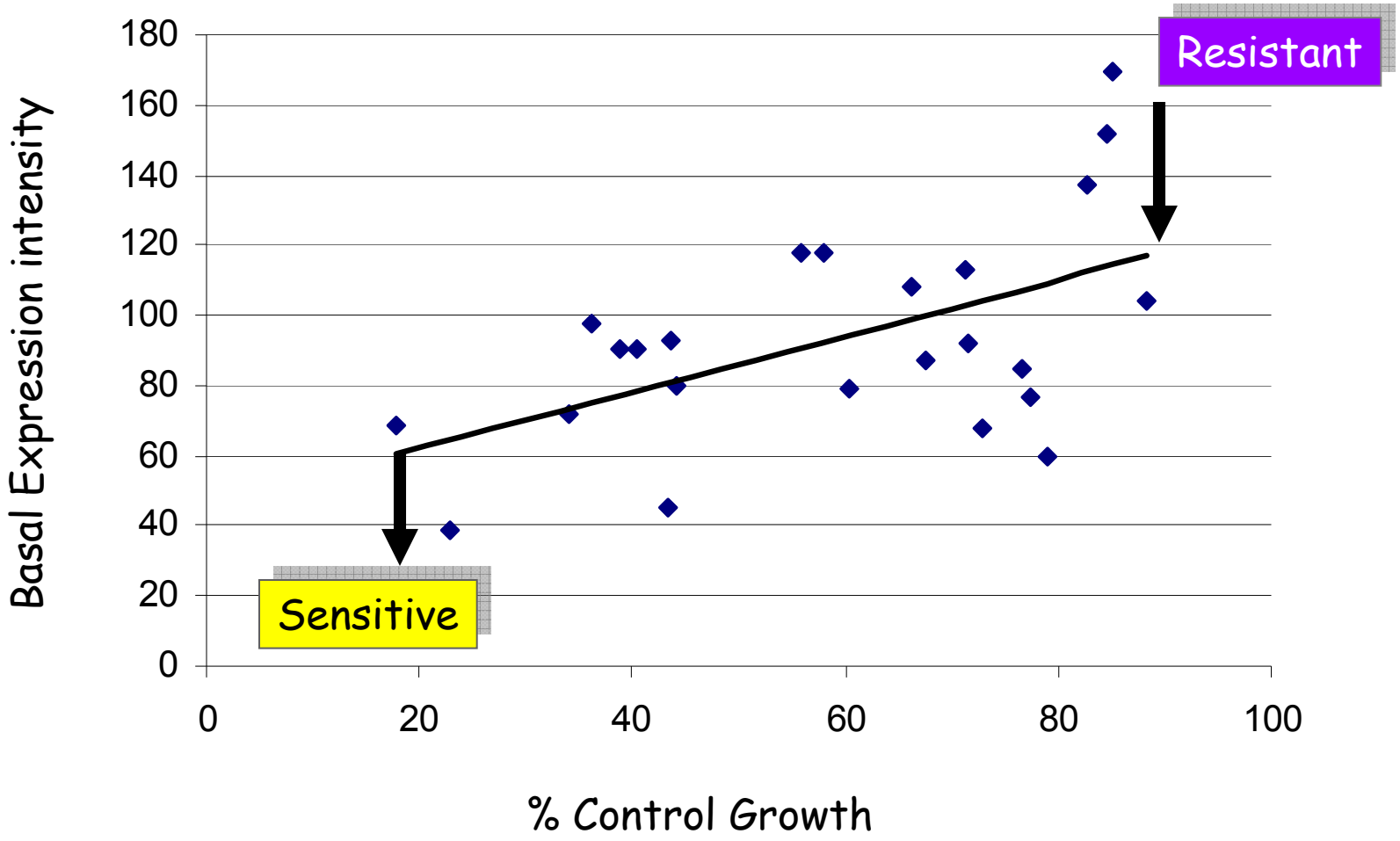
Is the existence of *MGMT* in the basal set the sole reason for high prediction??

The Basal gene set is a better predictor of alkylation sensitivity than *MGMT* alone

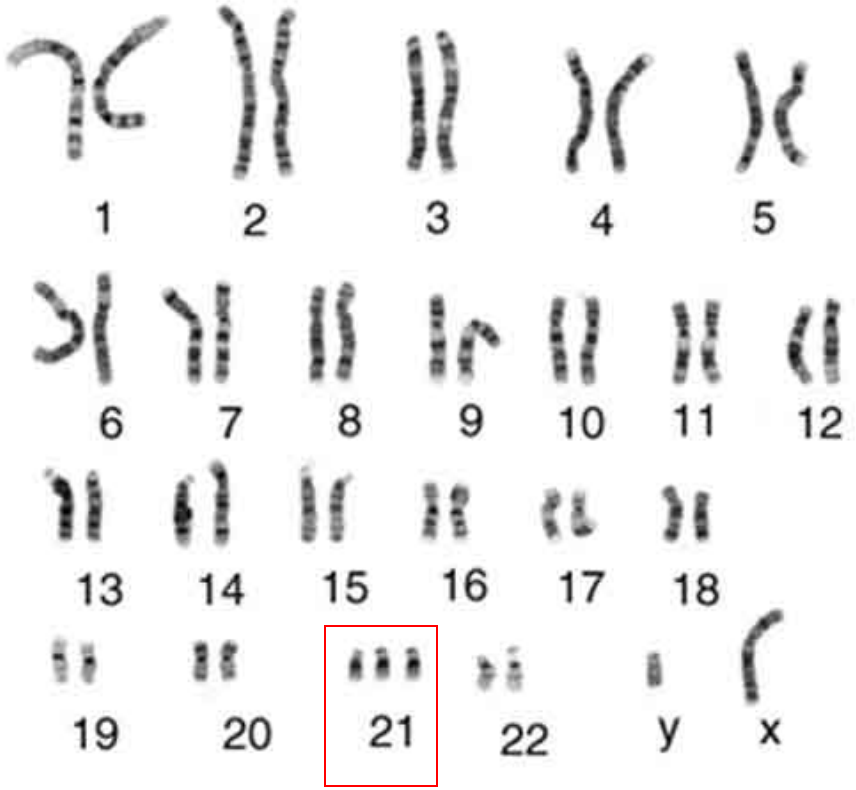


MGMT silencing is used as prognostic indicator of successful alkylation chemotheraory for glioblastoma-and only gene set member known to repair methylated DNA

Second most significant positive association: *C21ORF56*



Chromosome 21: Associated with Numerous Diseases

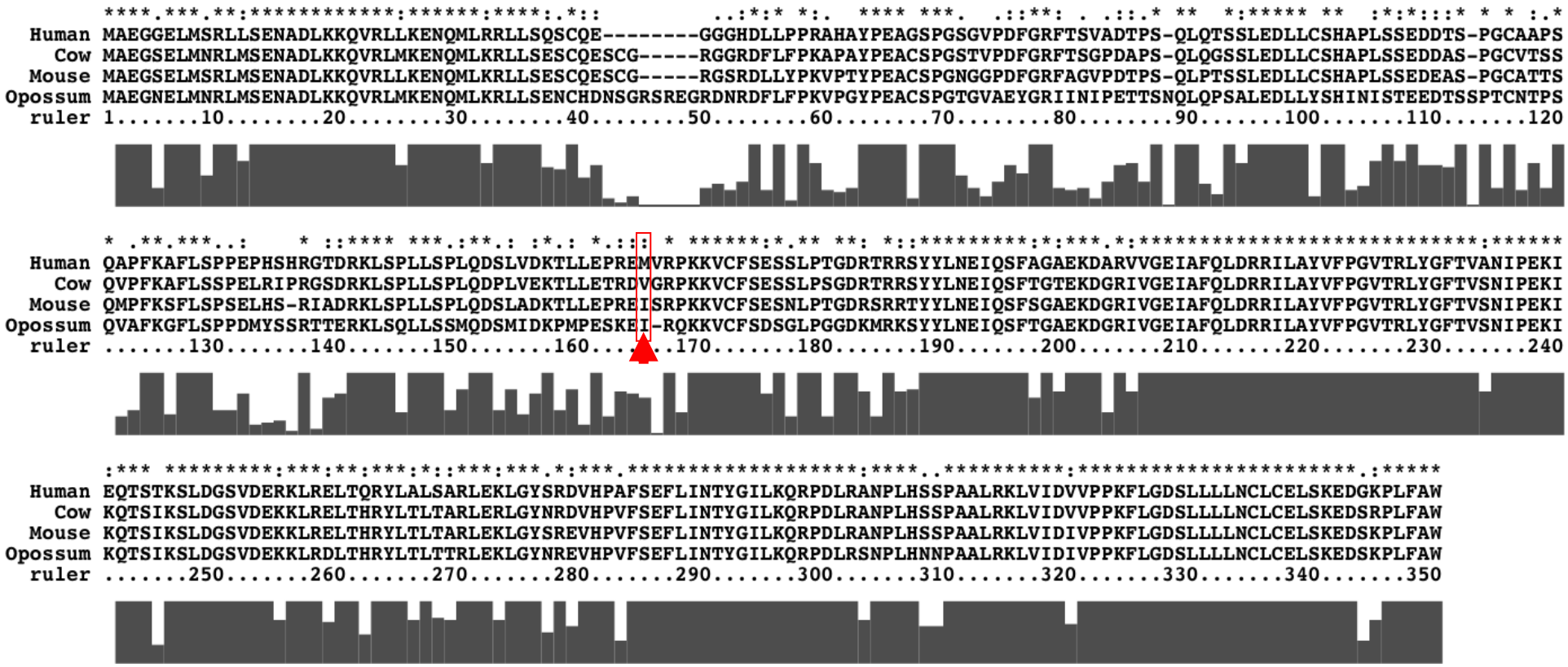


biotechnologyonline.gov

Chromosome 21
associated with disease:

Down Syndrome
Alzheimer's
Acute Myeloid Leukemia
Autoimmune Disease

C21ORF56 conserved across mammals



340 AA long form, 186 AA short form

Homologous to SPATC1: spermatogenesis and centriole-associated 1: Known to interact with CDC20, a regulator of cell cycle

C21ORF56 shows variation in expression in CEPH cell lines

OPEN ACCESS Freely available online PLoS GENETICS

Genome-Wide Associations of Gene Expression Variation in Humans

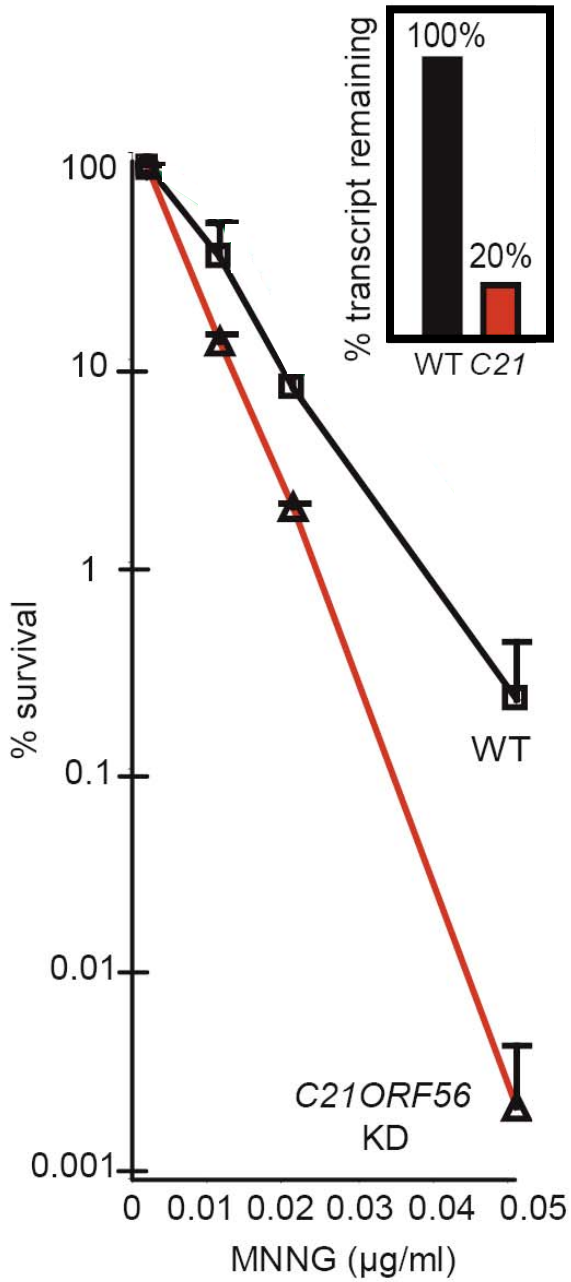
Barbara E. Stranger[1☯], Matthew S. Forrest[1☯], Andrew G. Clark[2], Mark J. Minichiello[1], Samuel Deutsch[3], Robert Lyle[3], Sarah Hunt[1], Brenda Kahl[4], Stylianos E. Antonarakis[3], Simon Tavaré[5,6], Panagiotis Deloukas[1*], Emmanouil T. Dermitzakis[1*]

1 Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom, 2 Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York, United States of America, 3 Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland, 4 Illumina, Inc., San Diego, California, United States of America, 5 Department of Oncology, University of Cambridge, Hutchison/MRC Research Centre, Cambridge, United Kingdom, 6 Program in Molecular and Computational Biology, University of Southern California, Los Angeles, California, United States of America

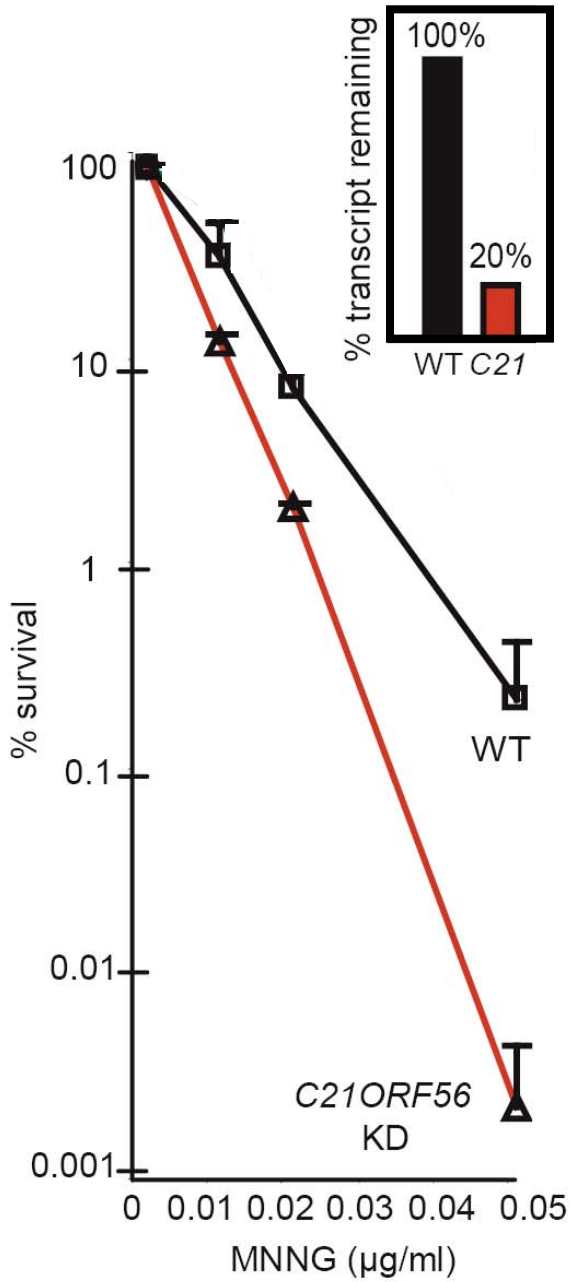
CEU samples: Centre d'Etude du Polymorphisme Humain
derived from 60 unrelated Utah residents descended from Northern and Western Europe

Does C21ORF56 influence MNNG sensitivity??

Does C21ORF56 influence MNNG sensitivity??

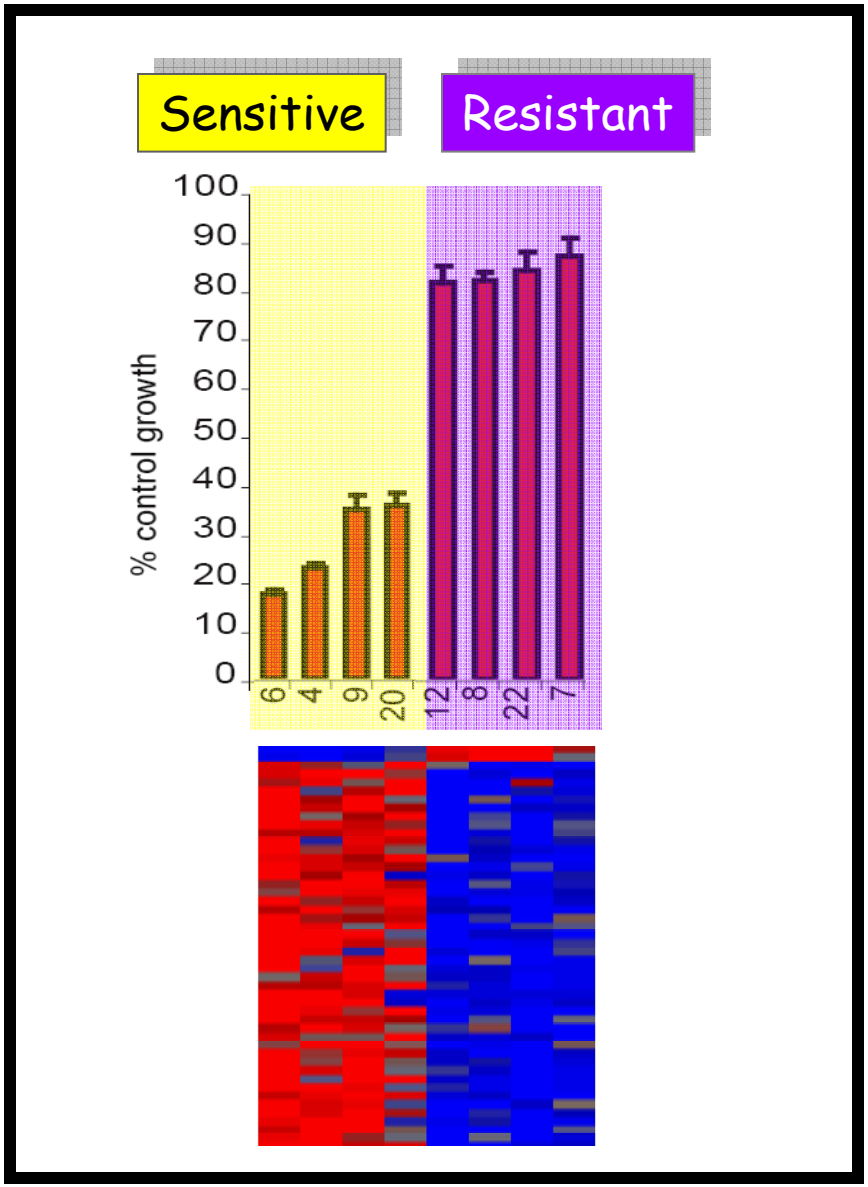


Does C21ORF56 influence MNNG sensitivity??

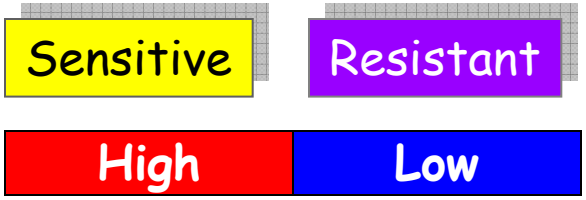


YES!!!!!!!

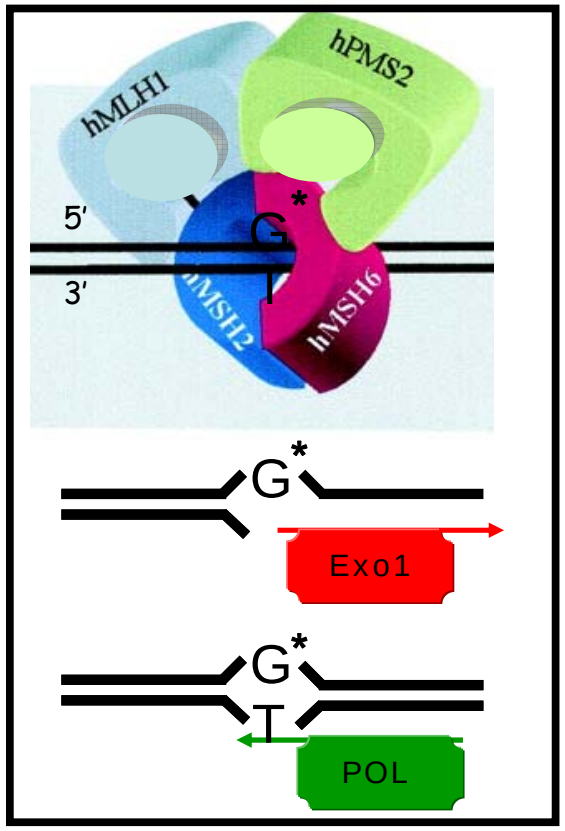
Two expression patterns in basal set

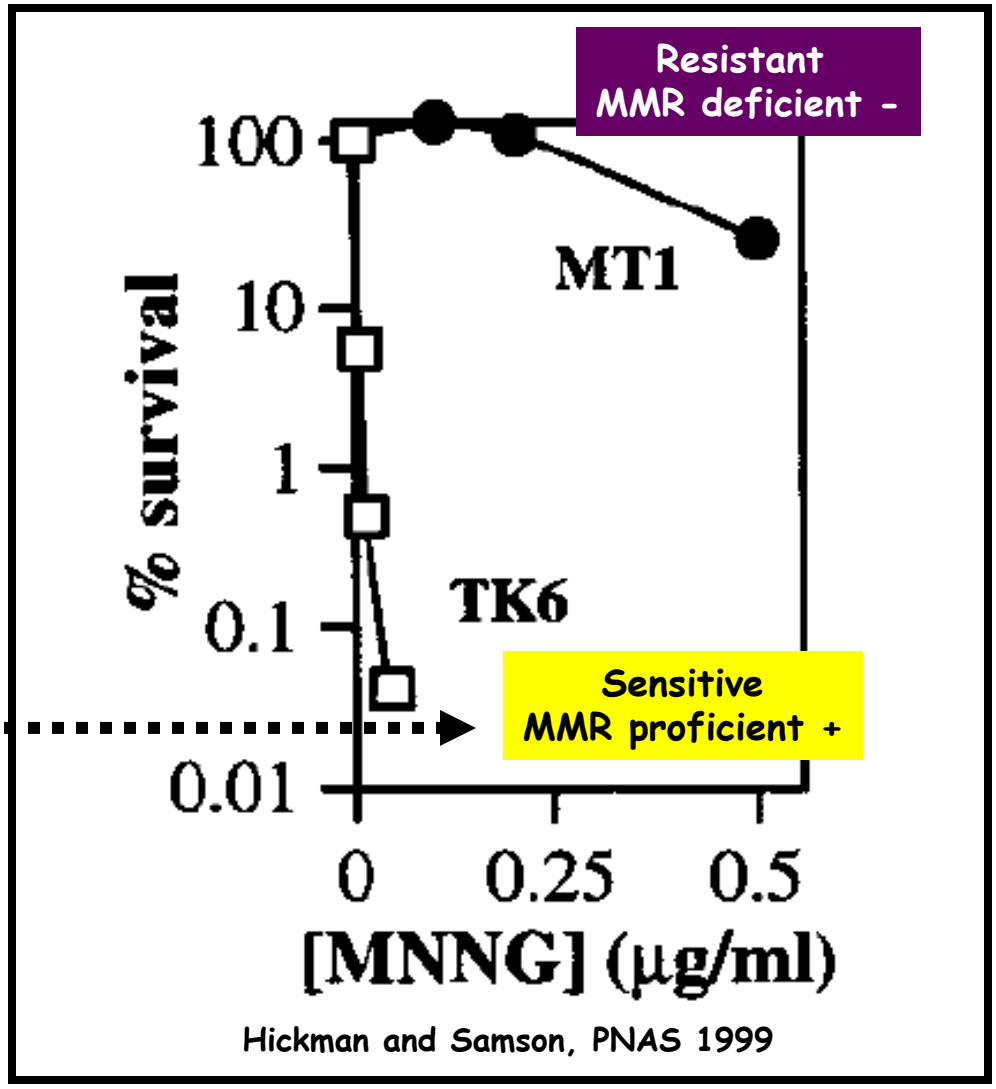


2: High expression in Cells with high sensitivity



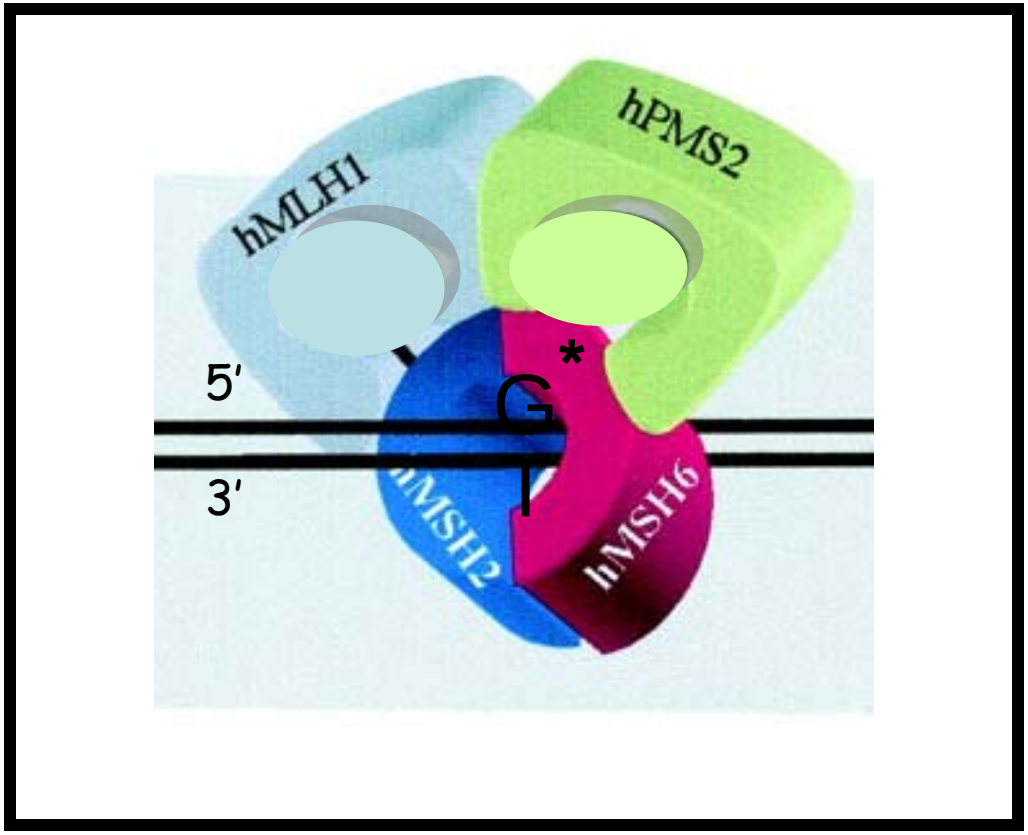
Paradox of Mismatch Repair



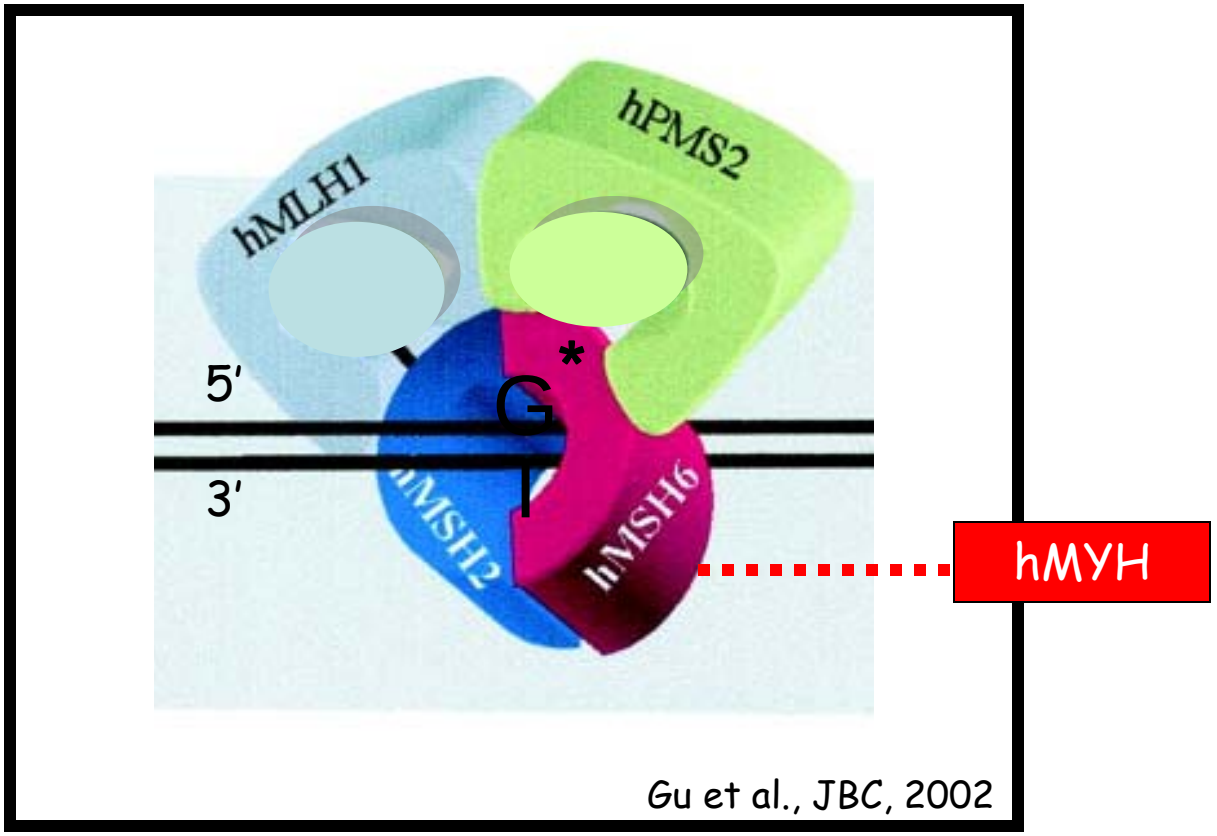


FUNCTIONAL MMR processing
O⁶MeG:T mismatches
Leads to cell death!!

Predicted high expression of Mismatch Repair transcripts in sensitive cell lines



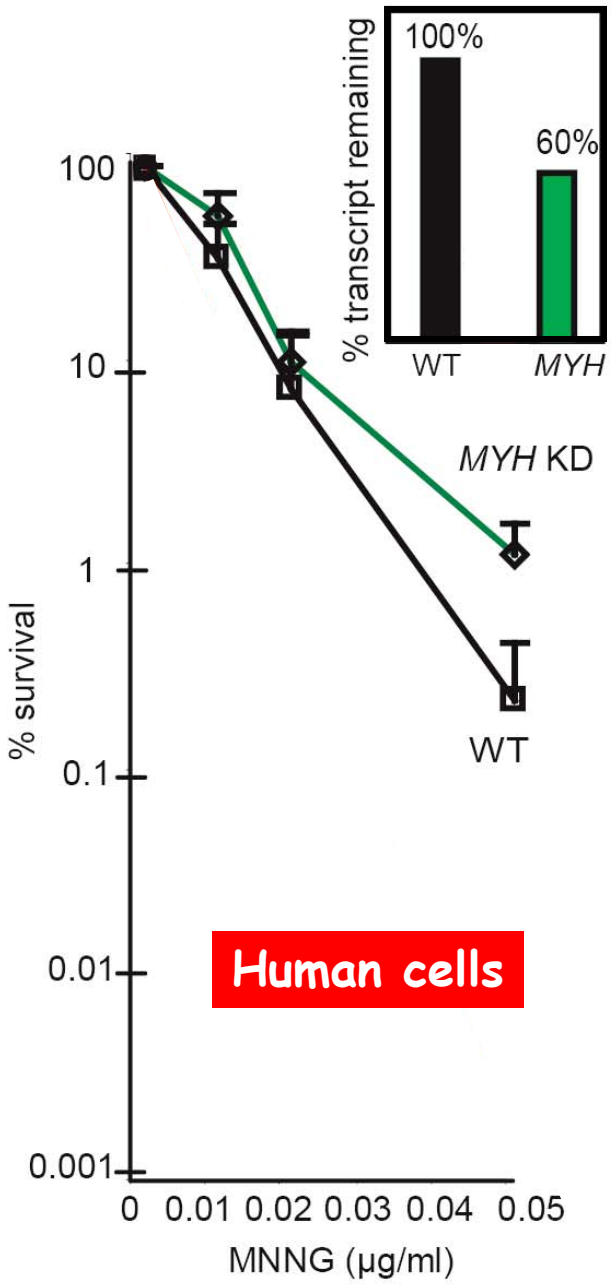
Base excision repair protein MYH has high expression in sensitive cell lines



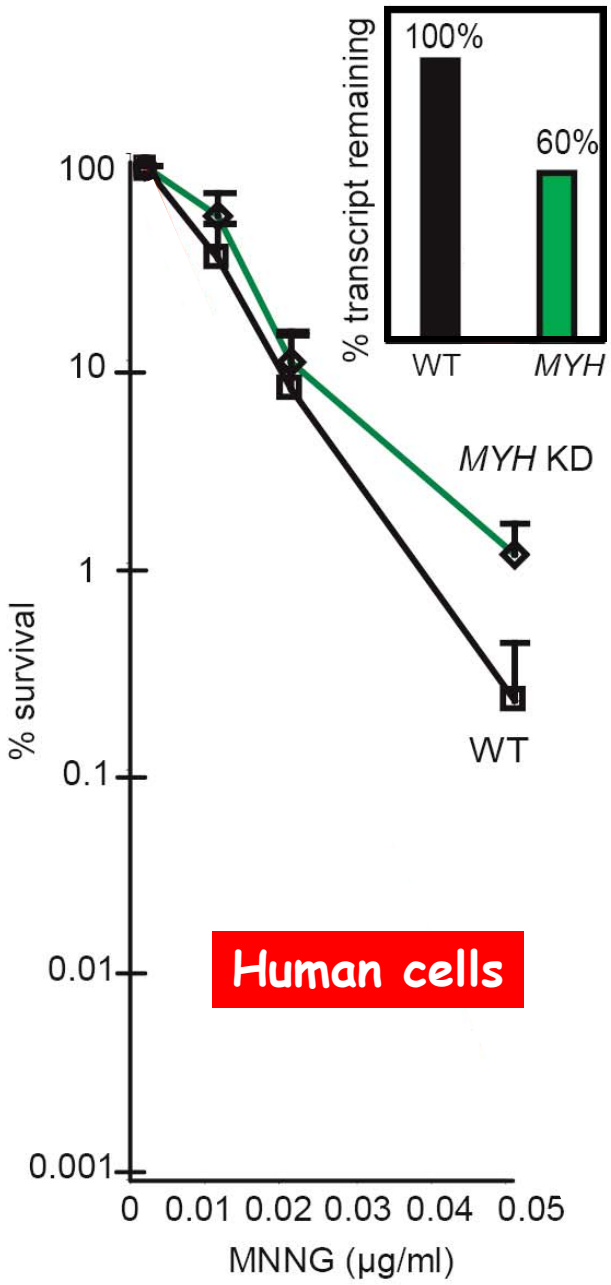
Does deficiency for MYH confer resistance to MNNG in the same manner as MMR deficiency?

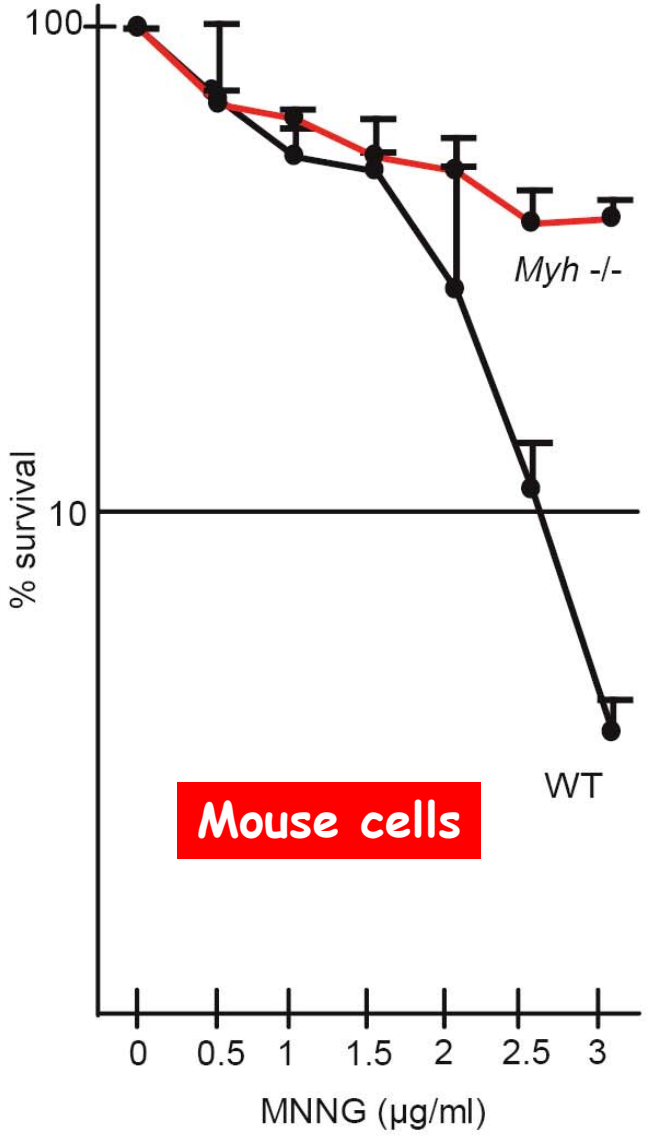
Does MYH influence MNNG sensitivity??

Does MYH influence MNNG sensitivity??



Does MYH influence MNNG sensitivity??





Summary

- Broad range of variation in sensitivity across cell lines derived from healthy genetically diverse individuals
- Basal gene expression predicts variation in sensitivity upon exposure to DNA damaging agent
- *MGMT*-currently used as predictor of tumor response to chemotherapy-is a member of our *ASA* set
- Tested and verified that two members of *ASA* set influence cellular sensitivity: *MYH* (DNA repair) and *C21ORF56* (unknown function)