United States Environmental Protection Agency

Mechanistic Indicators of Childhood Asthma (MICA) Study

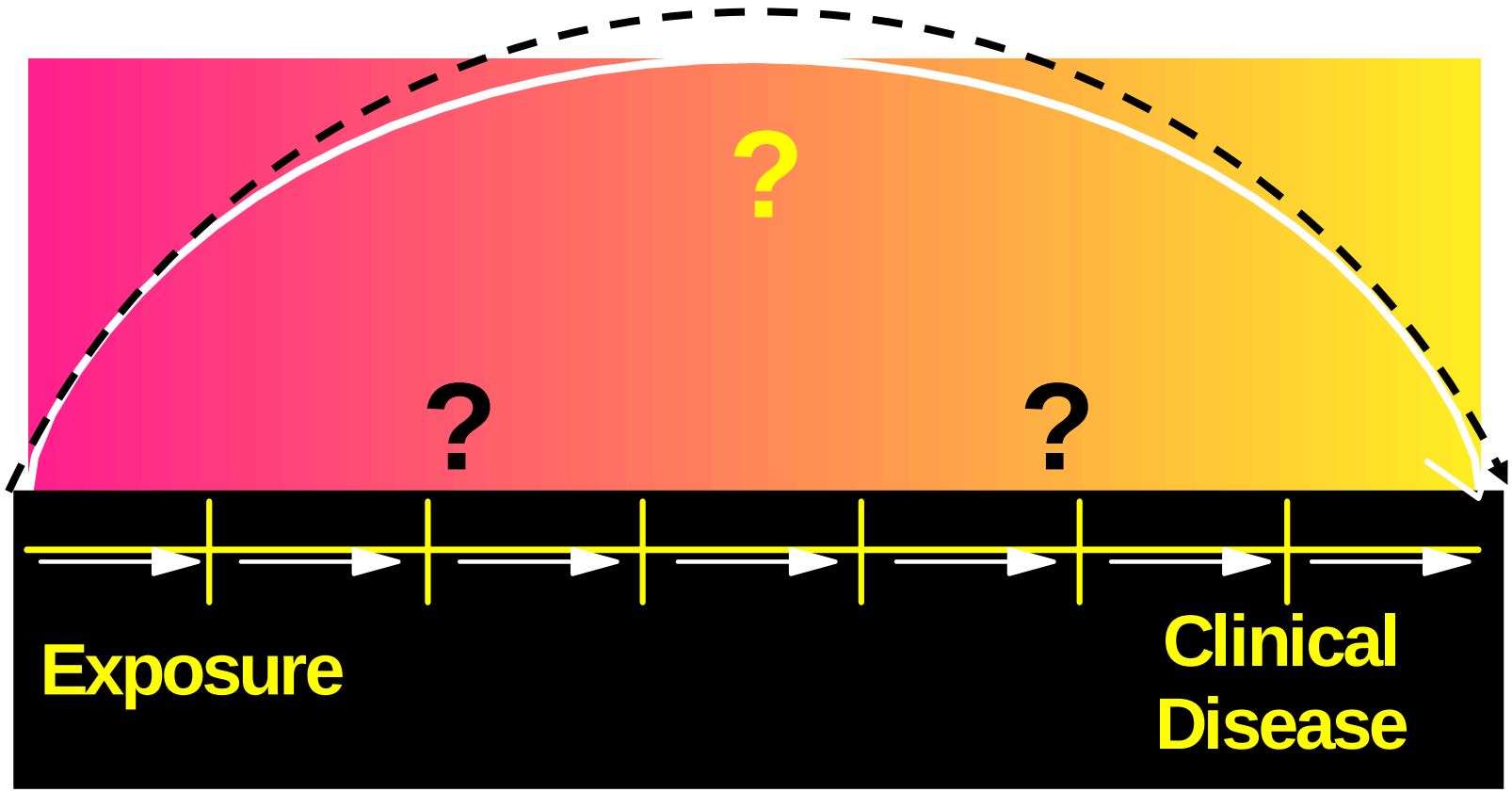
Elaine Cohen Hubal,
David Reif, Stephen Edwards, Lucus Neas, Ed Hudgens, Jane Gallagher

11th SAC Seminar: NEW TRENDS IN CHEMICAL TOXICOLOGY
22 – 25 September 2008
Danilovskaya Hotel, Moscow, Russian Federation

United States Environmental Protection Agency

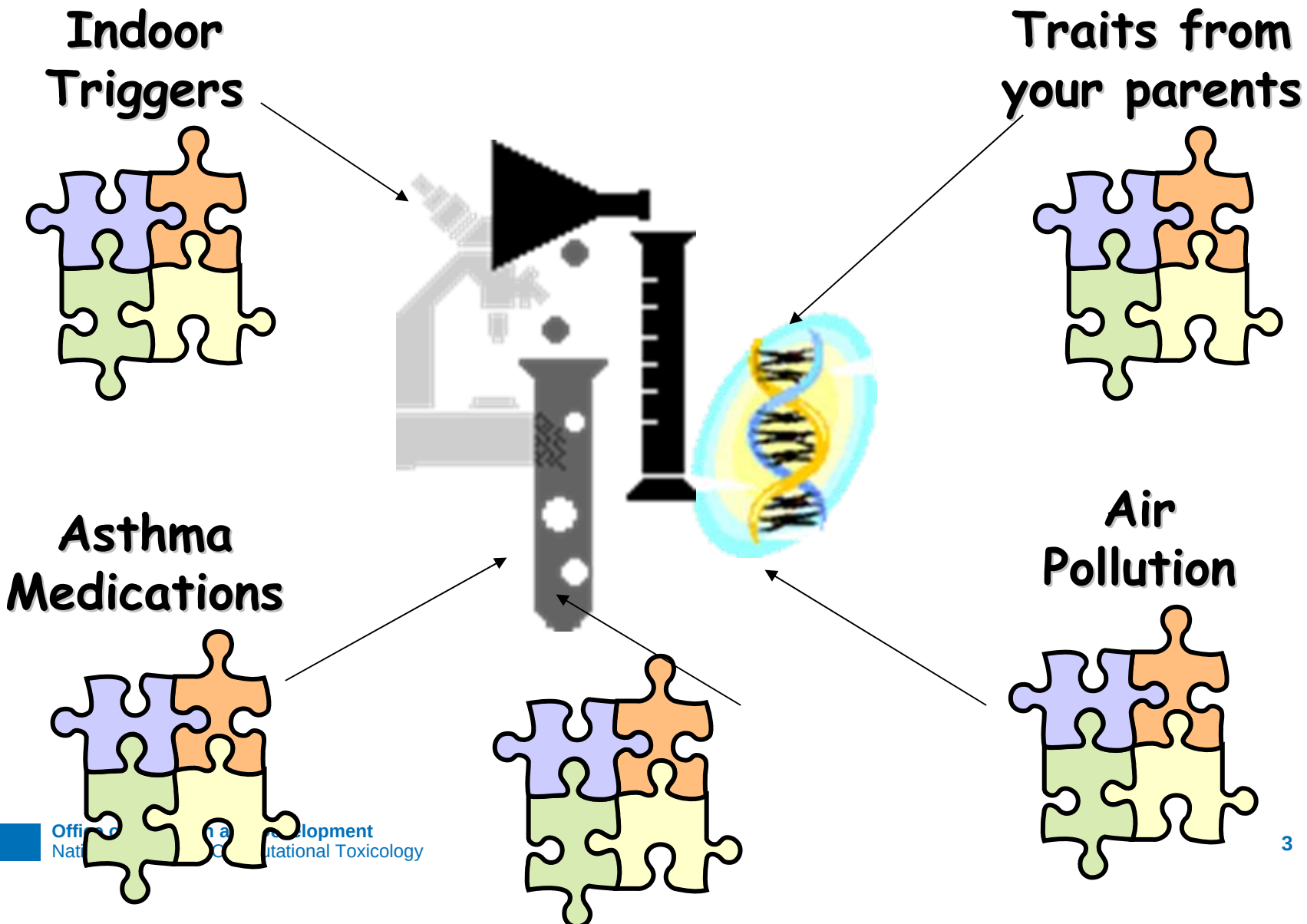
Mechanistic Indicators of Childhood Asthma (MICA)

- Apply state-of-the-art technologies to examine the interplay between environmental and genetic factors affecting asthma
- 100 asthmatic and 100 non-asthmatic children, ages 9-12 years (subset of Detroit Children's Health Study cohort)
- Collected multiple types of clinical, demographic, exposure, and gene expression data
- Consider markers of susceptibility, exposure, and effects to analyze and characterize combined risk factors that relate to asthma severity
- http://www.epa.gov/dears/studies.htm



Jane Gallagher

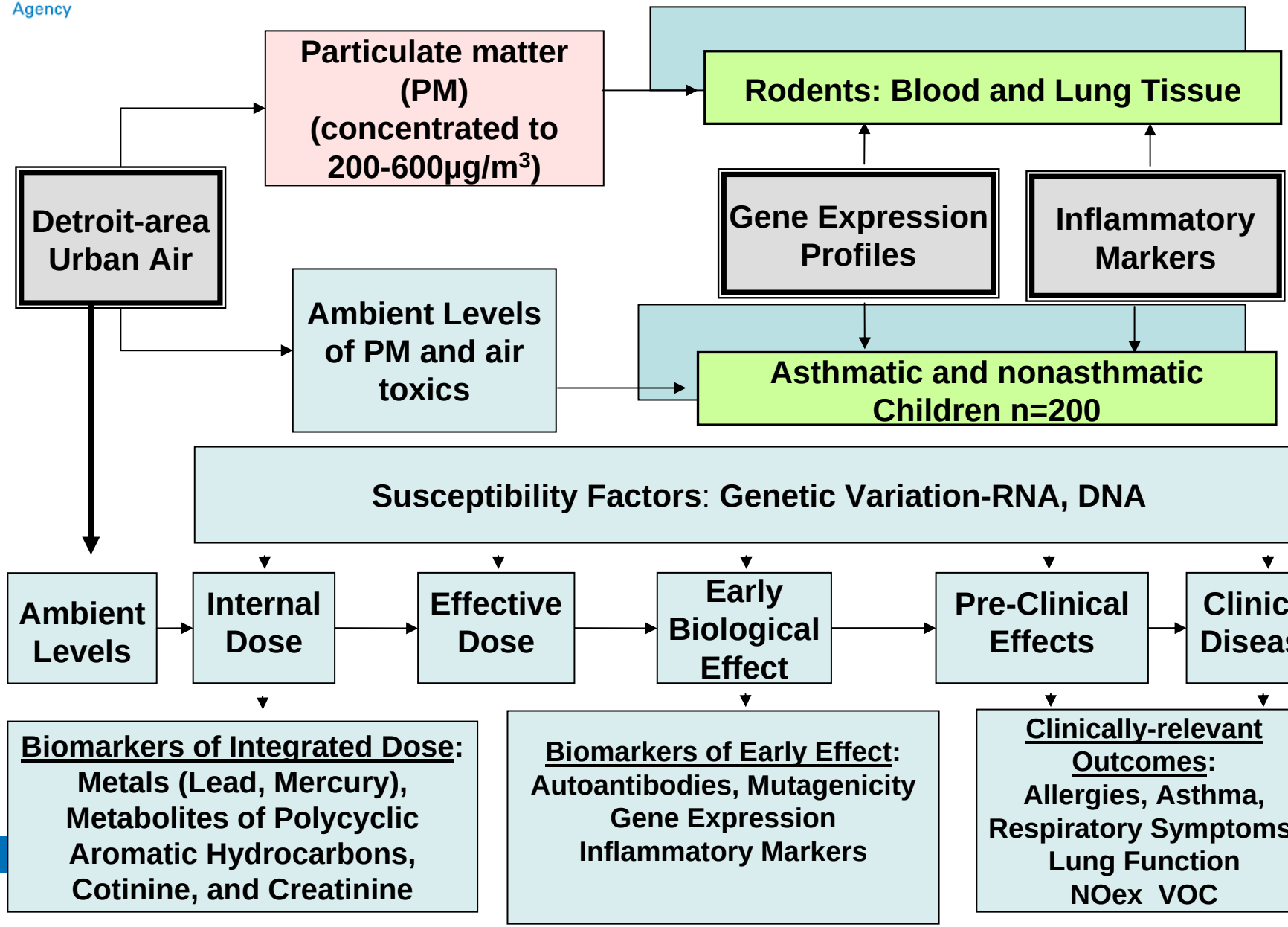
Slide taken from the **MICA** educational presentation used in our study in Detroit



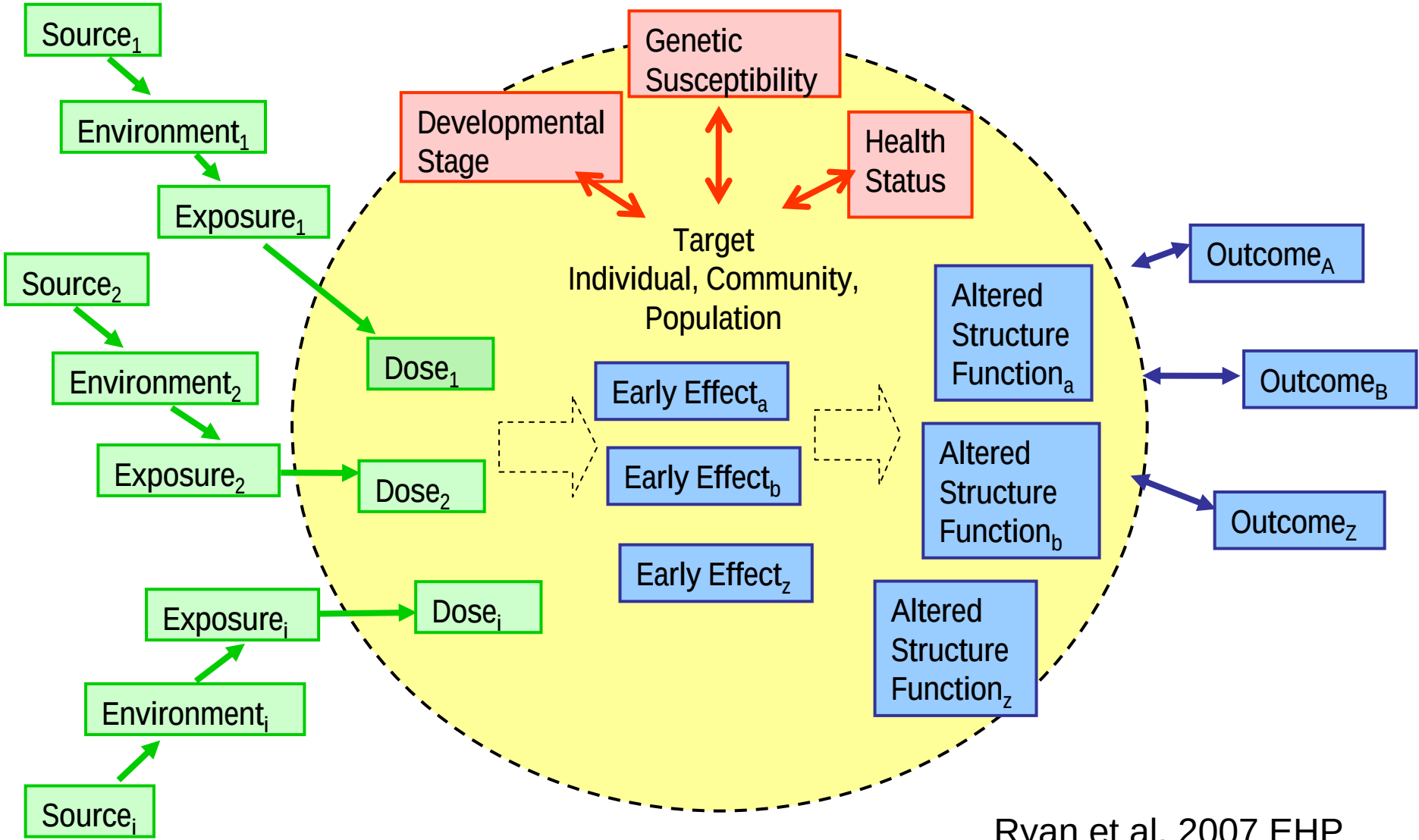
MICA Study Design

Jane Gallagher

United States Environmental Protection Agency

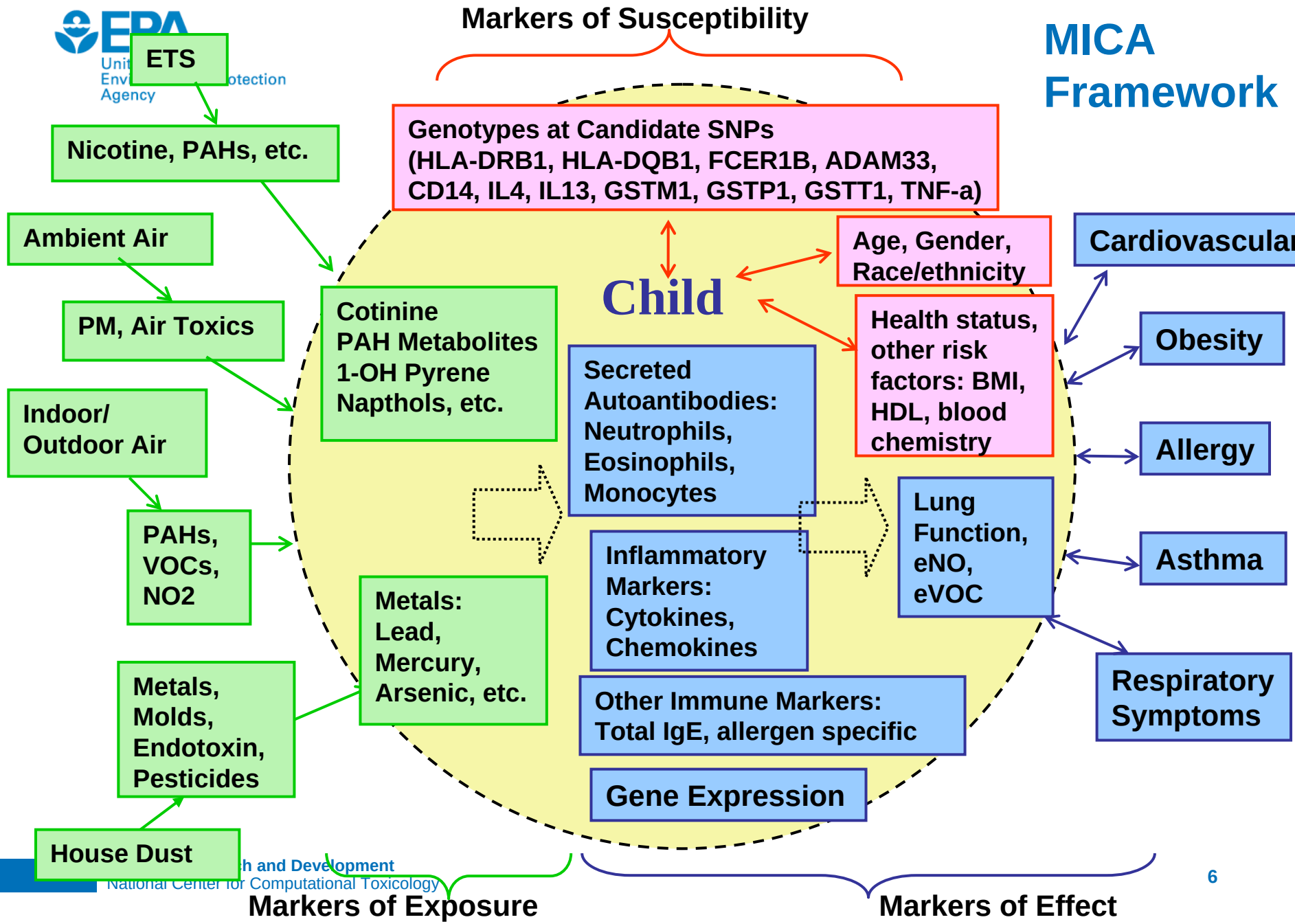


Biomarker Framework

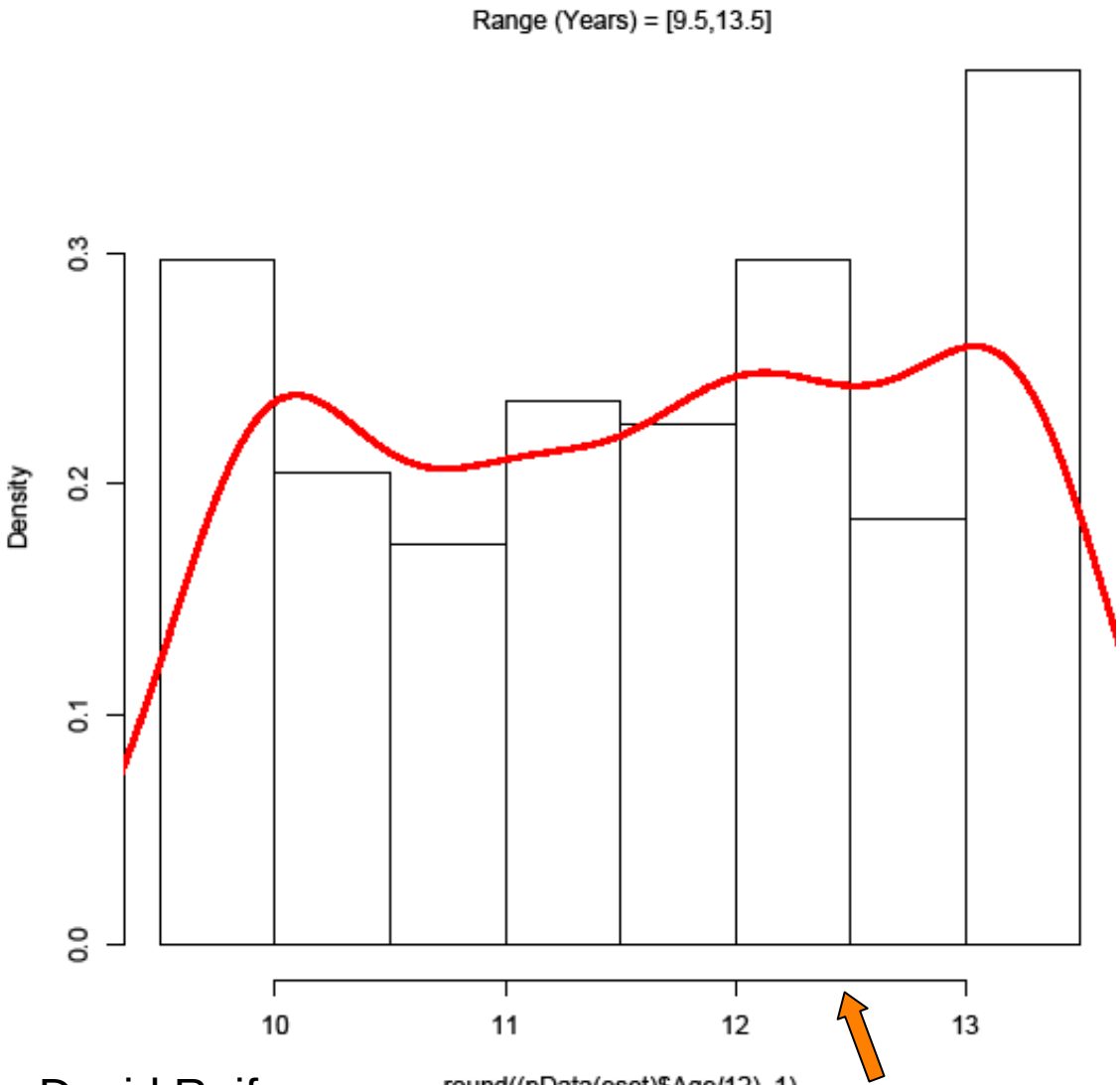


Ryan et al. 2007 EHP

MICA Framework



Characteristics of our MICA Study Sample



Study sample includes both boys and girls

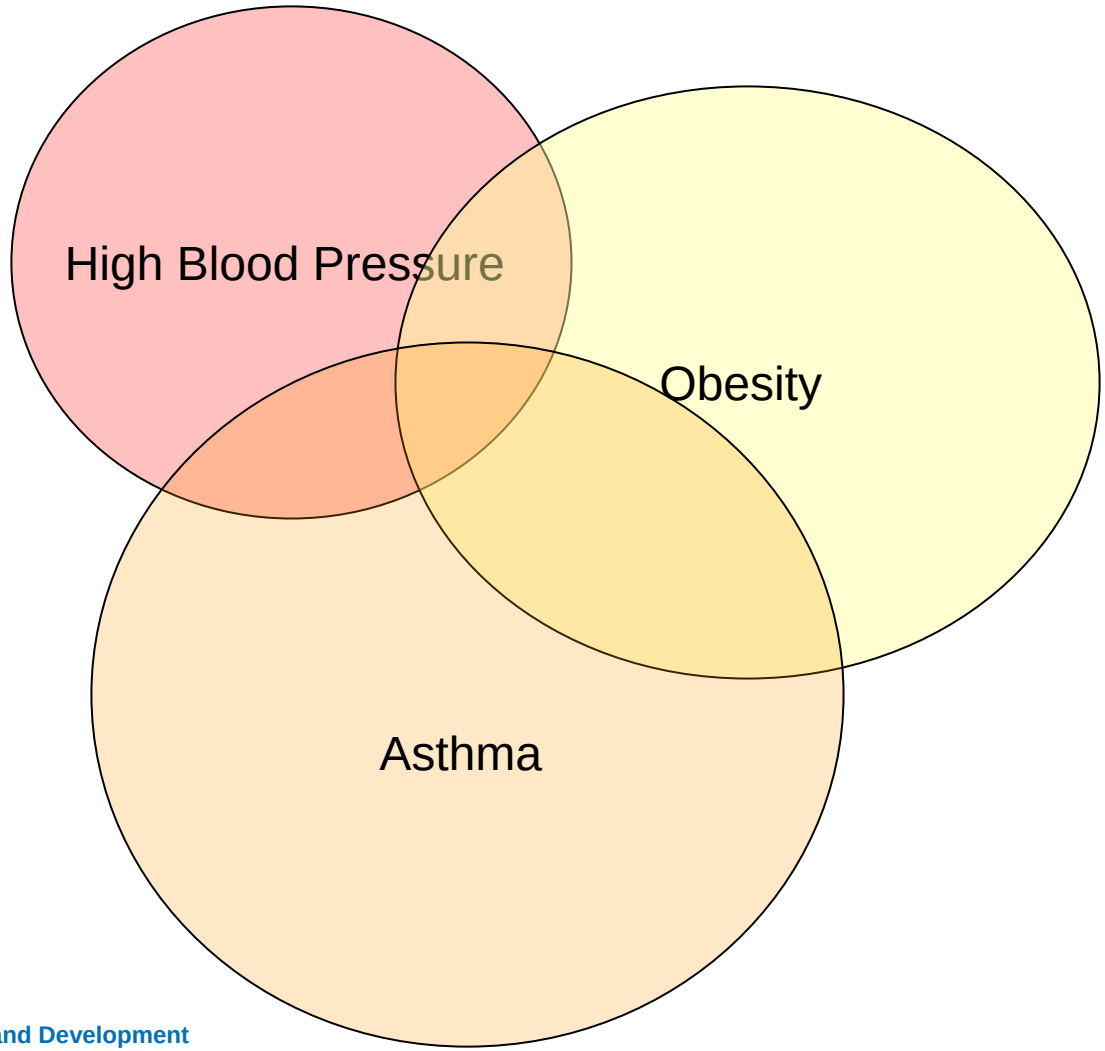
Study sample is predominantly African-American (>80%)

Study sample includes children diagnosed as **asthmatic** and **non-asthmatic**

Distribution of age across all categories is fairly uniform

Average age at menarche (nation-wide)

MICA Childhood Study Multiple Risk Factors



Jane Gallagher

United States Environmental Protection Agency

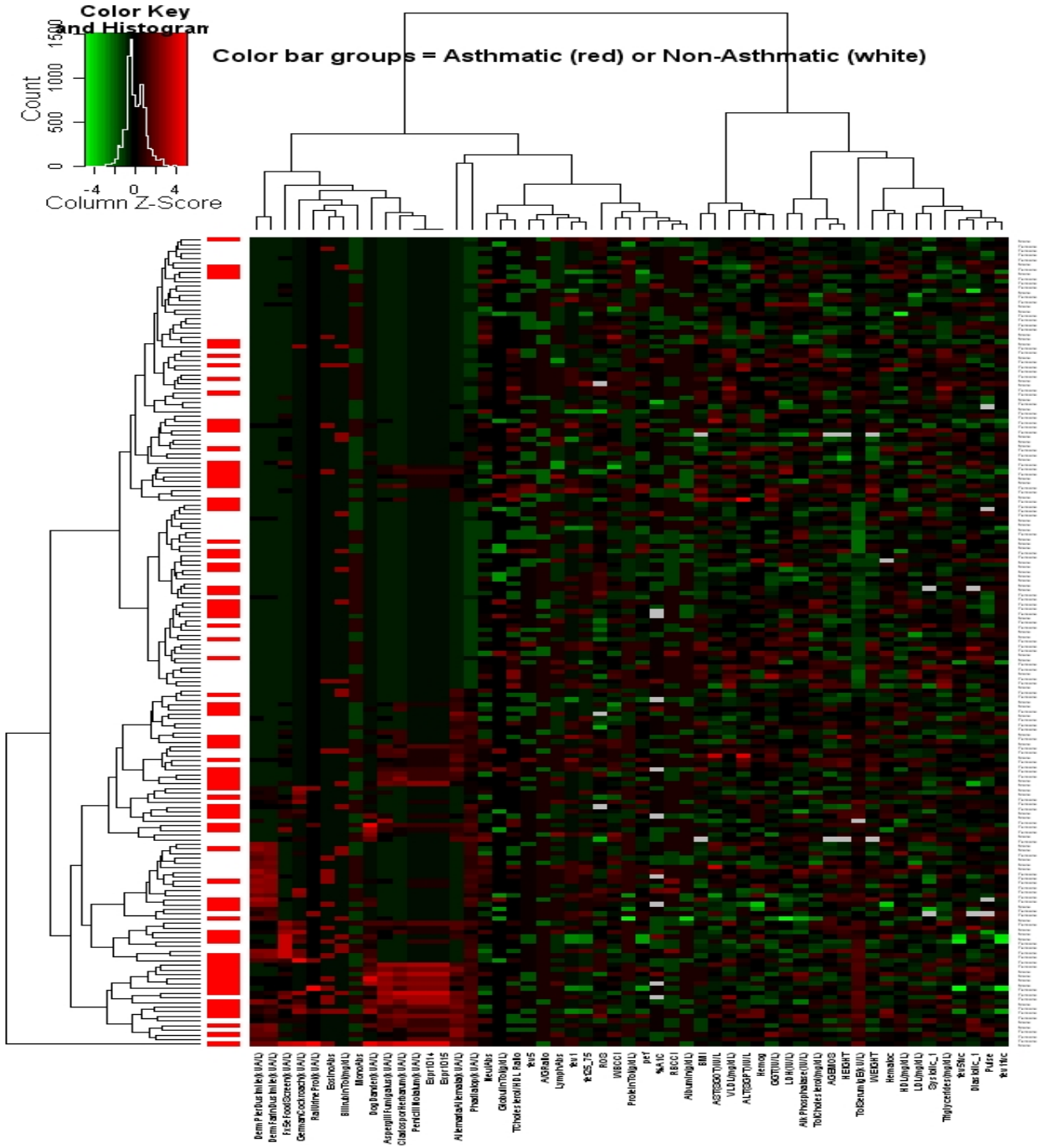
Collaborative NCCT-MICA Goals

- The objectives of the NCCT include advancement of a systems approach to evaluate complex relationships between
  - environmental factors
  - physiological biomarkers
  - health outcomes.
- NCCT collaborating to apply advanced statistical and machine learning methods to evaluate biomarker data collected in MICA
- Contribution of the NCCT component:
  - analyze genetic and gene expression data
  - use a systems biology approach to put data into framework for evaluating ecogenetics

United States Environmental Protection Agency

Data Analysis Methods

- Traditional statistics: linear regression, logistic regression, ANOVA, linear discriminant analysis.
- Machine learning: recursive partitioning trees, bootstrap aggregation (bagging) techniques, evolutionary computation-optimized classifiers, multifactor dimensionality reduction, random forests.
- Bioinformatics: protein interaction databases, knowledge (literature) mining tools, biological pathway database and inference software.
- Graphical approaches: cluster diagrams, expression “heat” maps, dendrograms, overlaid scatter plots (both exploratory and summary), distributional “violin” plots, regression plots.



MICA: Physiological Markers/ Health Status

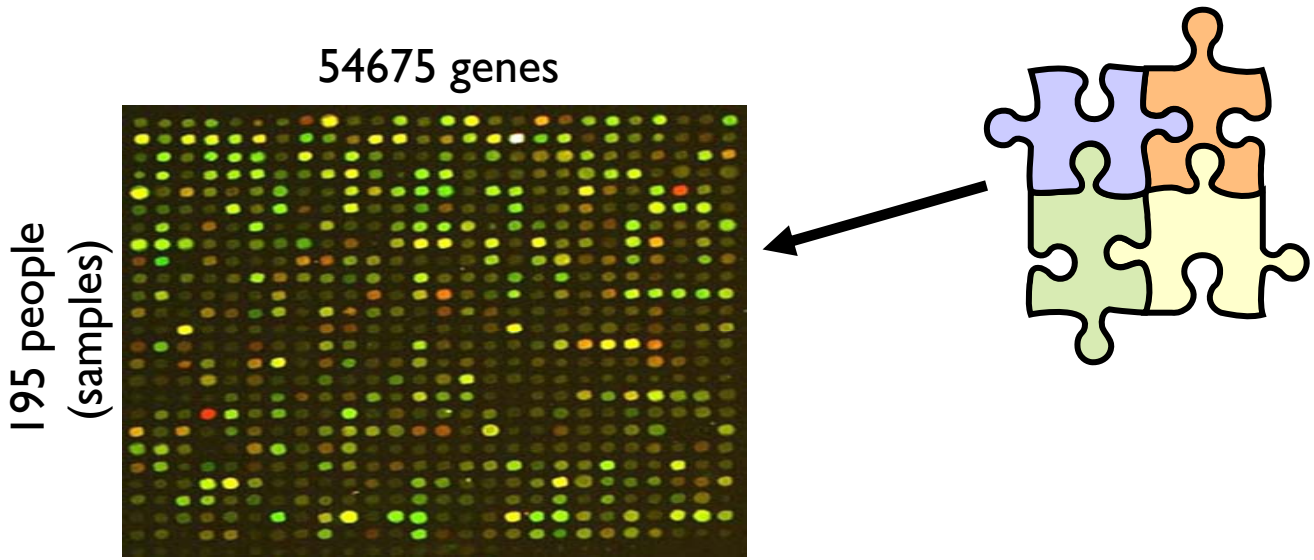
Red – Asthmic

White – Non Asthmatic

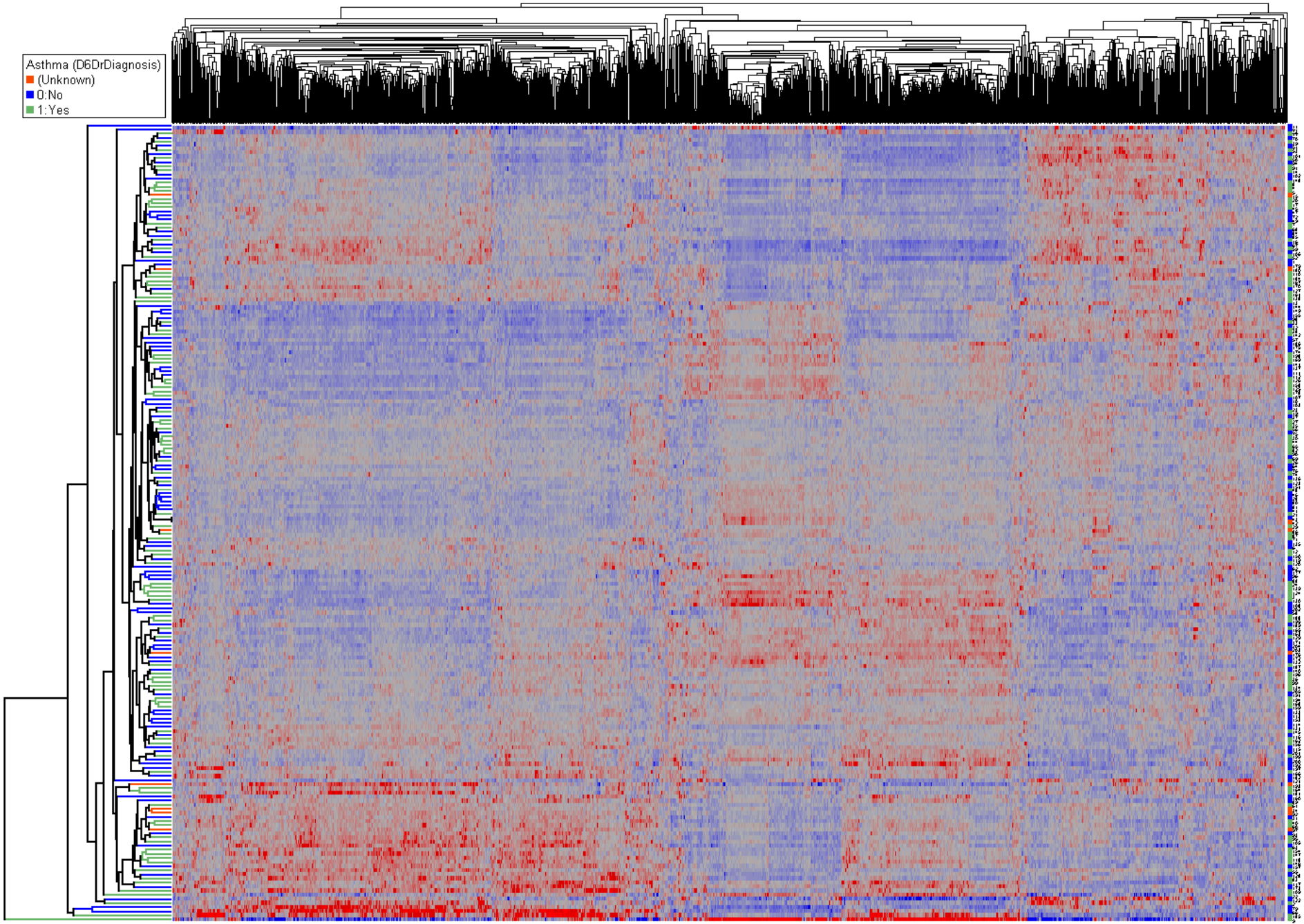
Gene expression measured using oligonucleotide microarrays

{gene expression}

Genetics



IMPORTANT: Total RNA was taken from whole blood samples in the absence of any deliberate experimental perturbation.



David Reif

Clustering of Subjects (Rows) by Gene Expression (Columns)

-6 0 6

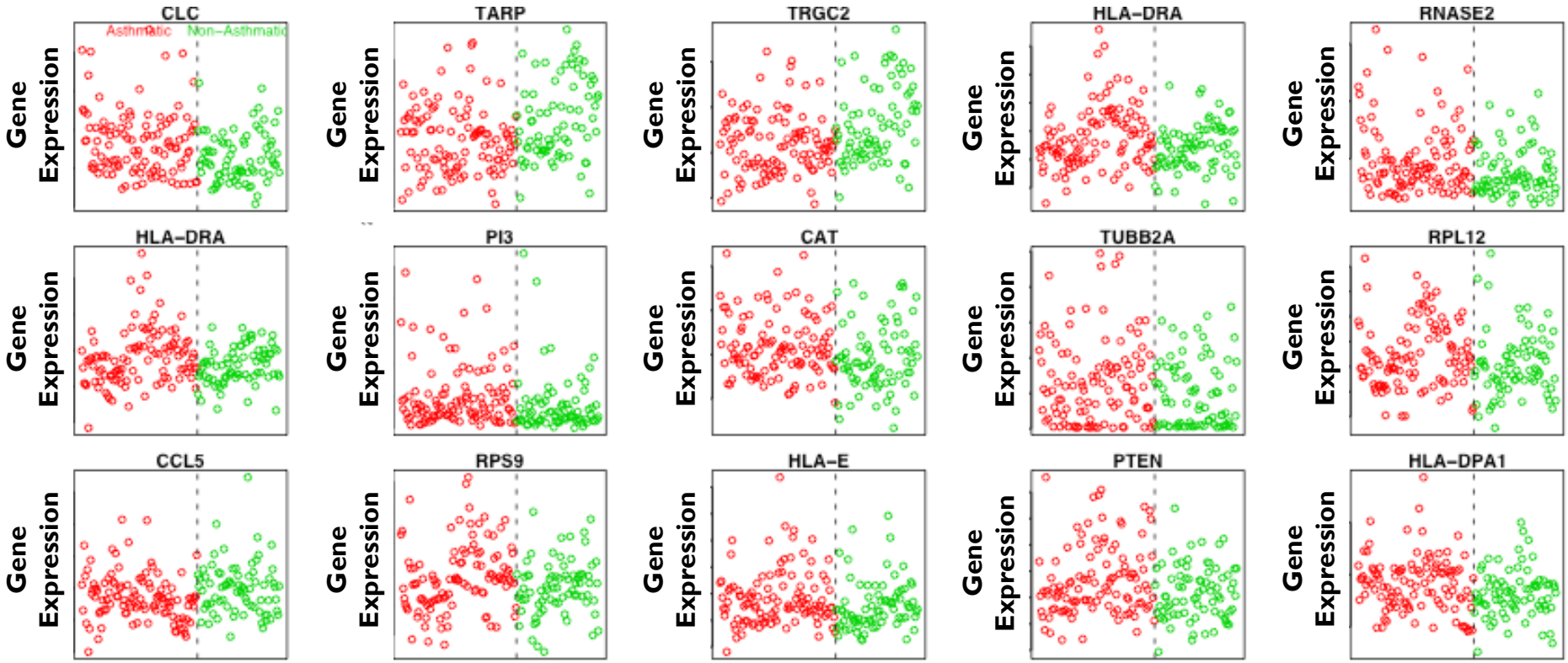
Column-wise standardized gene expression levels

United States Environmental Protection Agency

Can Gene Expression Distinguish Subtypes of Asthmatics?

- To identify subtypes, first applying an unbiased (i.e. without knowledge of asthma status) analysis to assess association between gene expression data and information on clinical, demographic, and exposure indicators.
- Next, select only gene expression probe sets that are significantly correlated with at least one of the demographic, clinical, or exposure indicators.
  - This filtering method prevents selecting only genes whose expression is associated with broadly-defined, imperfect asthma diagnoses.
- Examination of the genes differentiating asthma subtypes in this context highlights mechanistic genomic etiologies underlying the disease. These include subtypes of asthma characterized by
  - Patterns of gene expression associated with immune over-stimulation and household allergy exposures
  - Combinations of genomic biomarkers with demographic factors such as gender

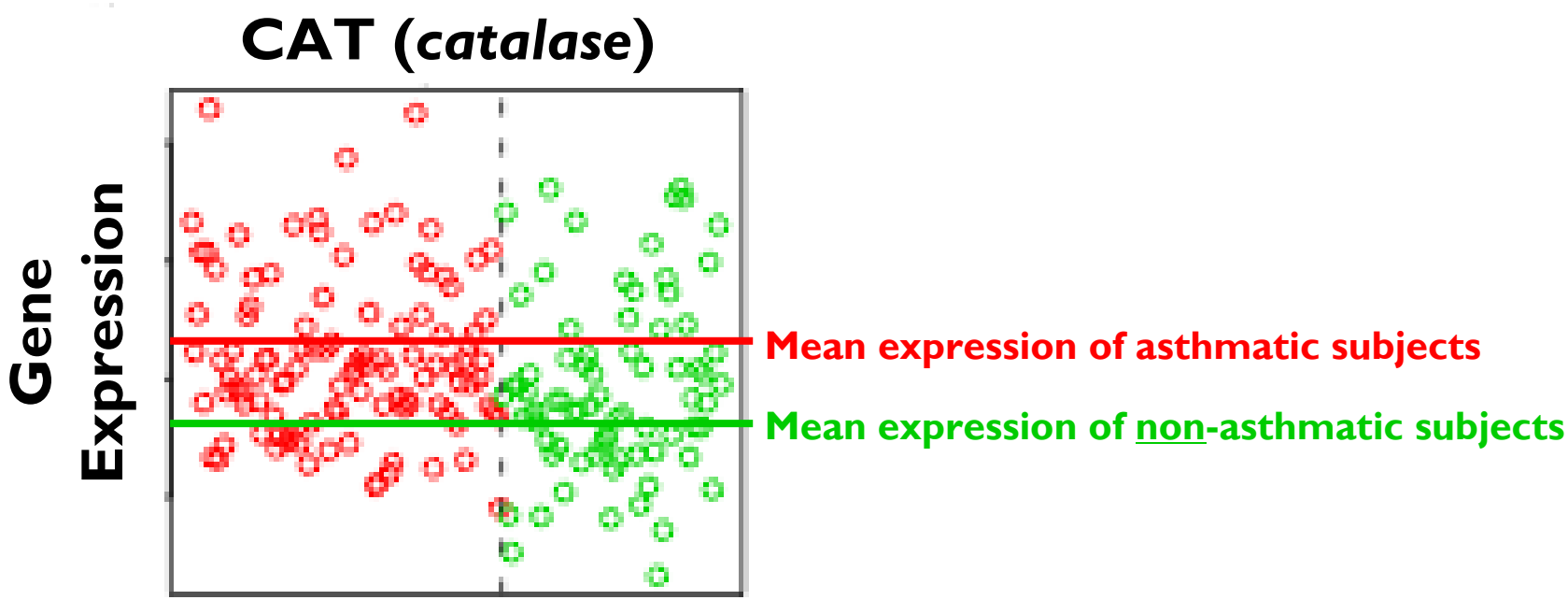
Why not just do the usual "here are some main-effect genes that discriminate asthmatics versus non-asthmatics"?



Does **not** address gene expression profiles associated with subtypes of asthma

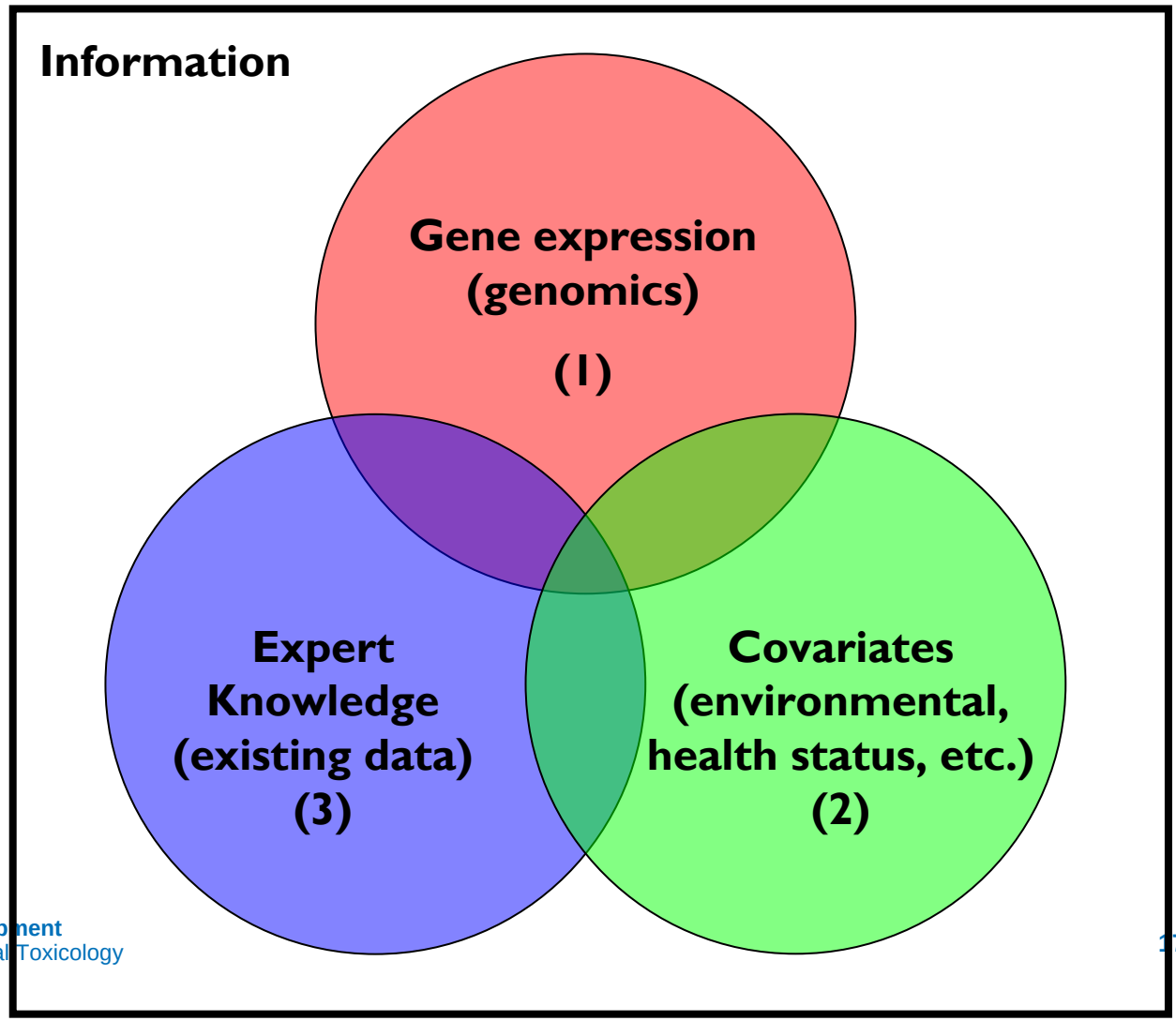
Does **not** link gene expression with other biomarkers or covariates (Where is the context?)

Why not just do the usual "here are some main-effect genes that discriminate asthmatics versus non-asthmatics"?



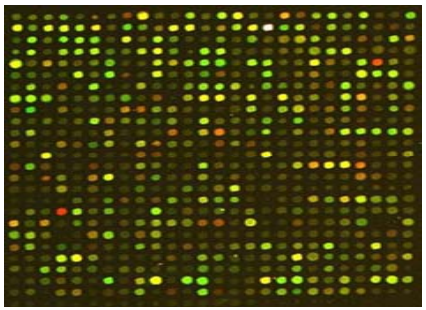
Reliance on group means **ignores** the complexity of asthma etiology

The ultimate goal is to glean mechanistic information regarding asthma subtypes

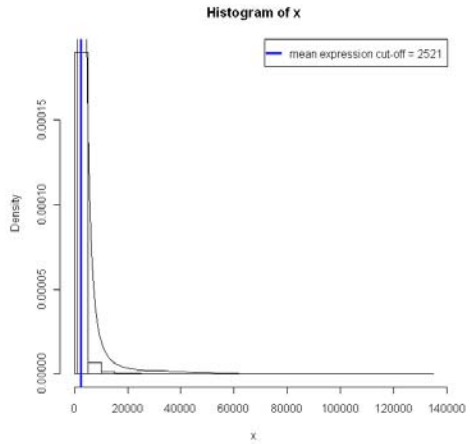


David Reif

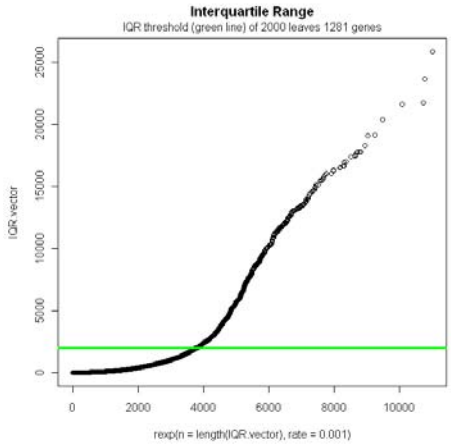
Analysis pipeline for the gene expression data

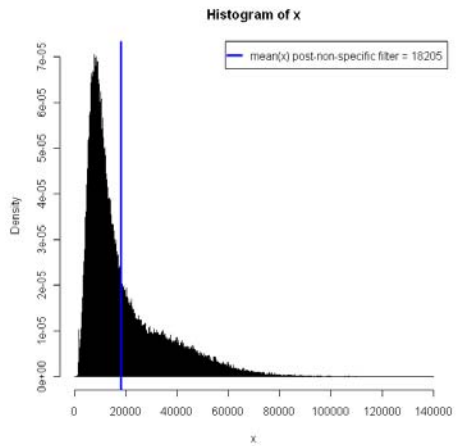


Gender adjustment

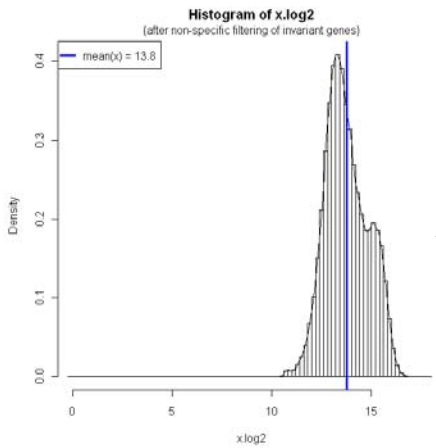


IQR filtering

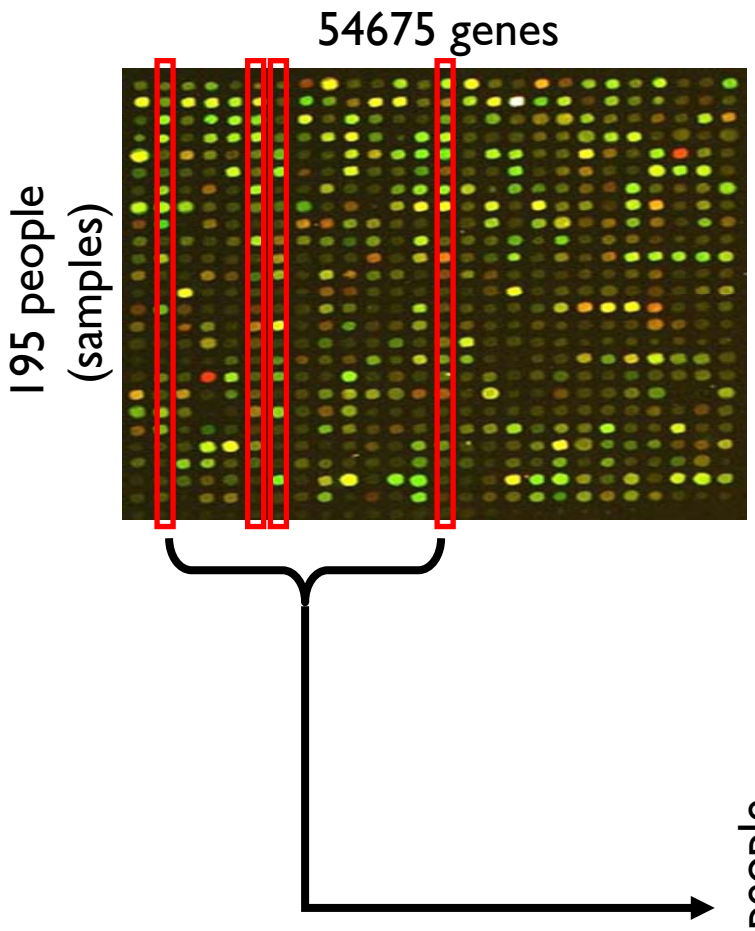




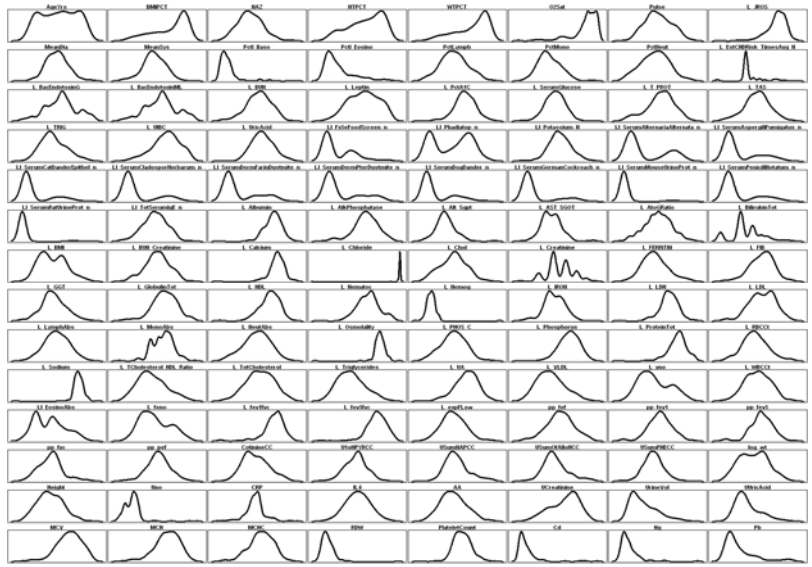
Log_2 transformation



How do we leverage MICA covariate information for the gene expression analysis?

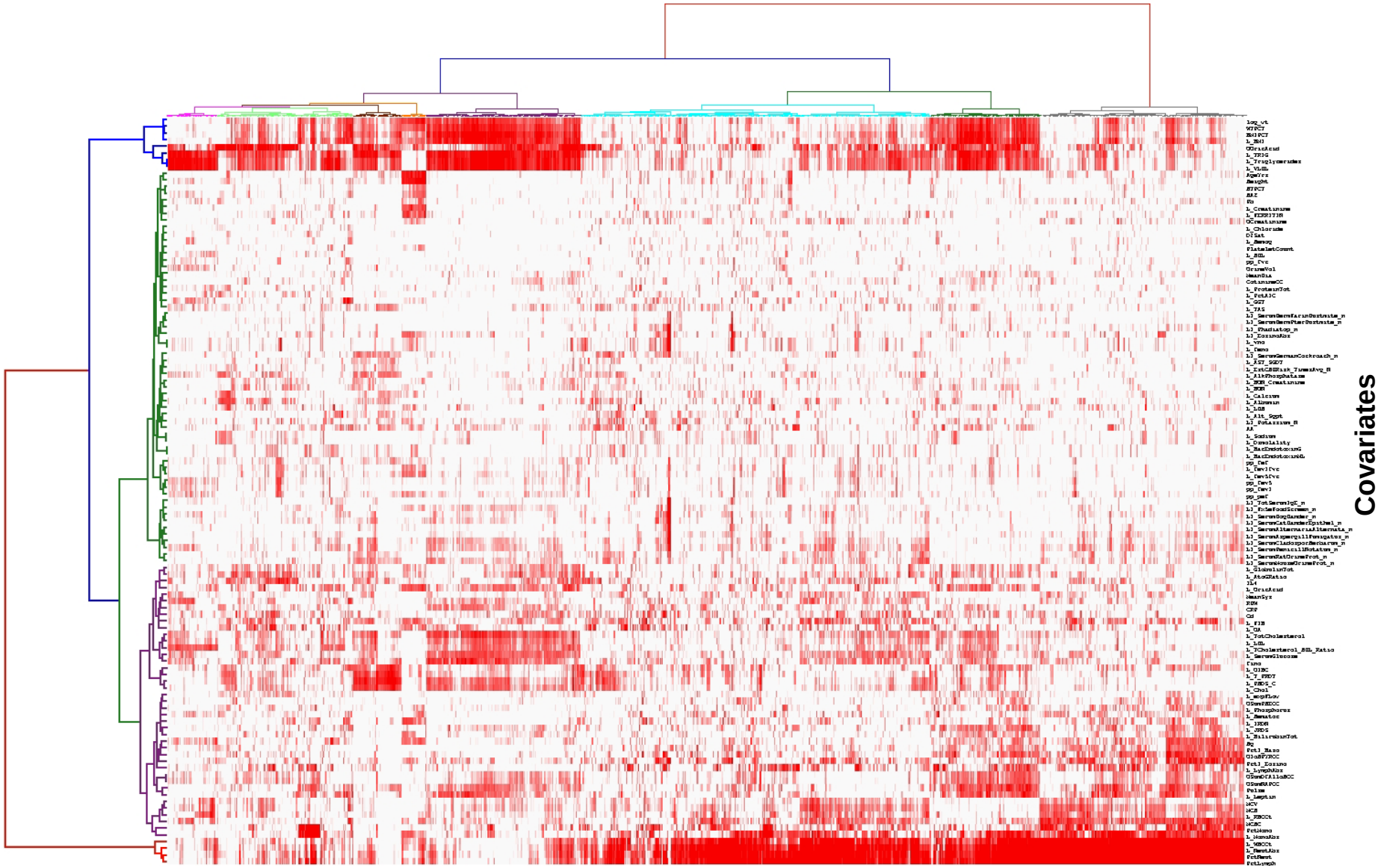


ρ



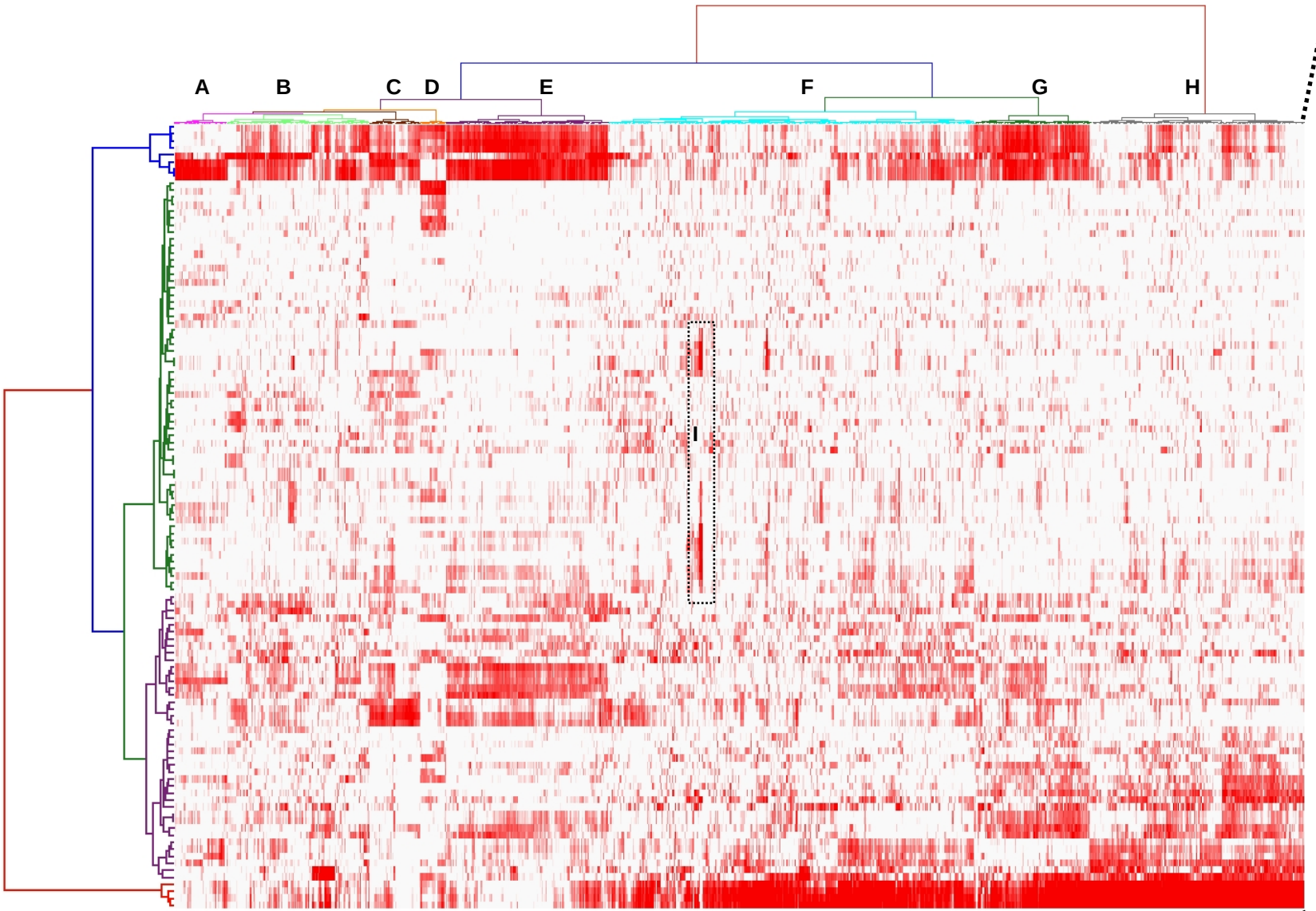
Results in a list of genes having significant correlation with at least one MICA covariate

Are MICA covariates reflective of underlying gene expression patterns?



David Reif

Finding context for gene expression patterns

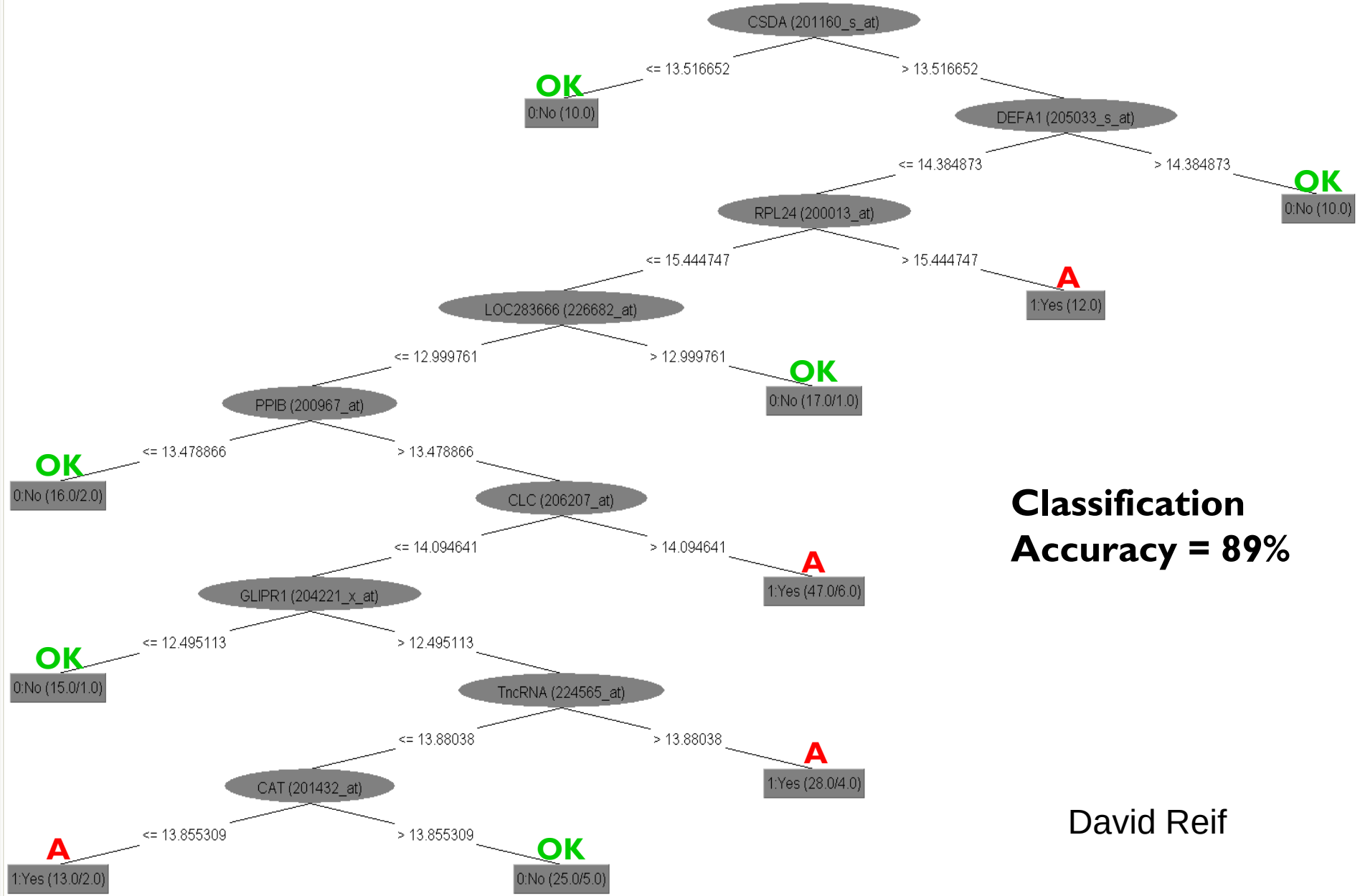


Absolute value of gene-covariate correlation

0.0 — 1.0

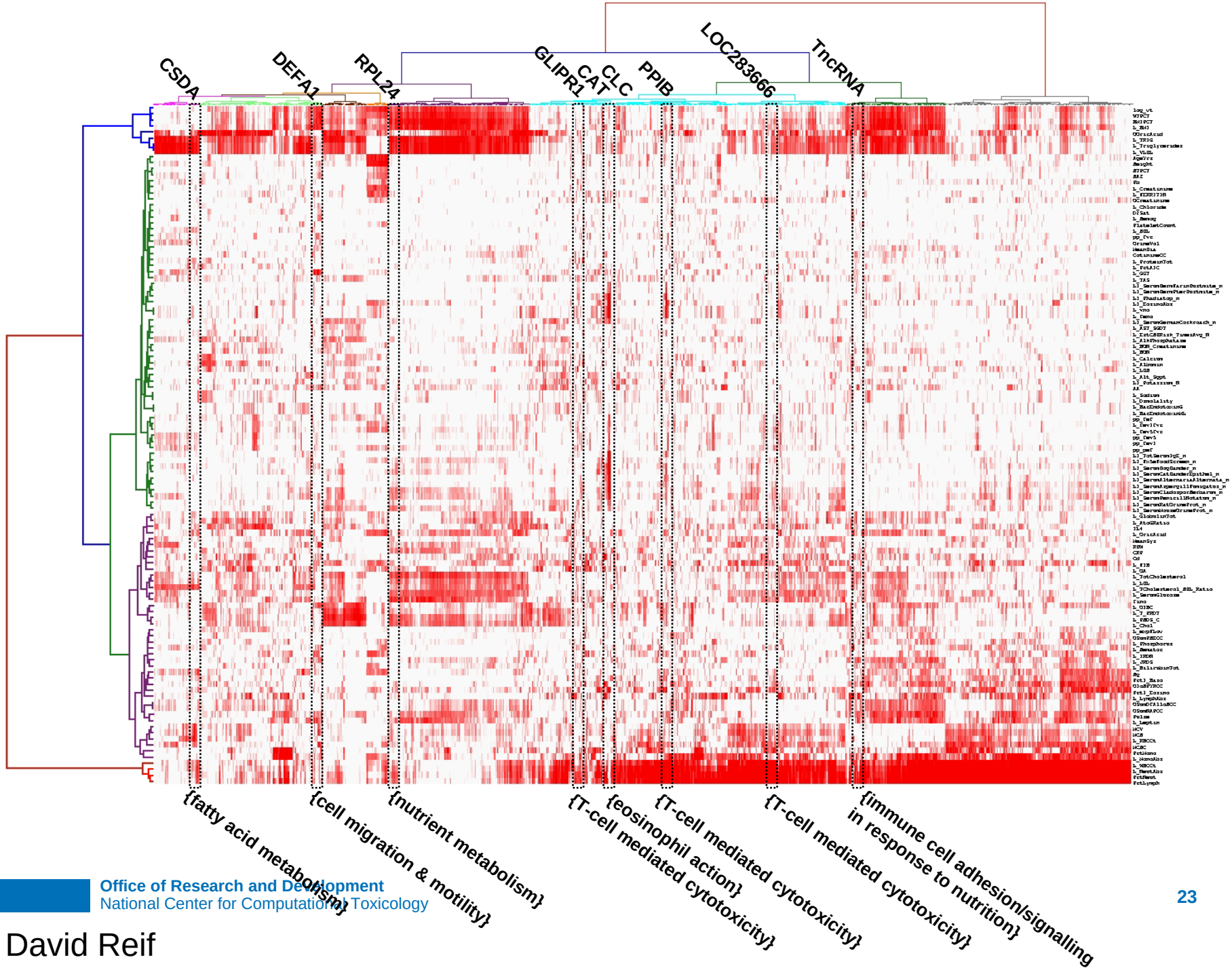
David Reif

Can we discriminate between subtypes of asthma?



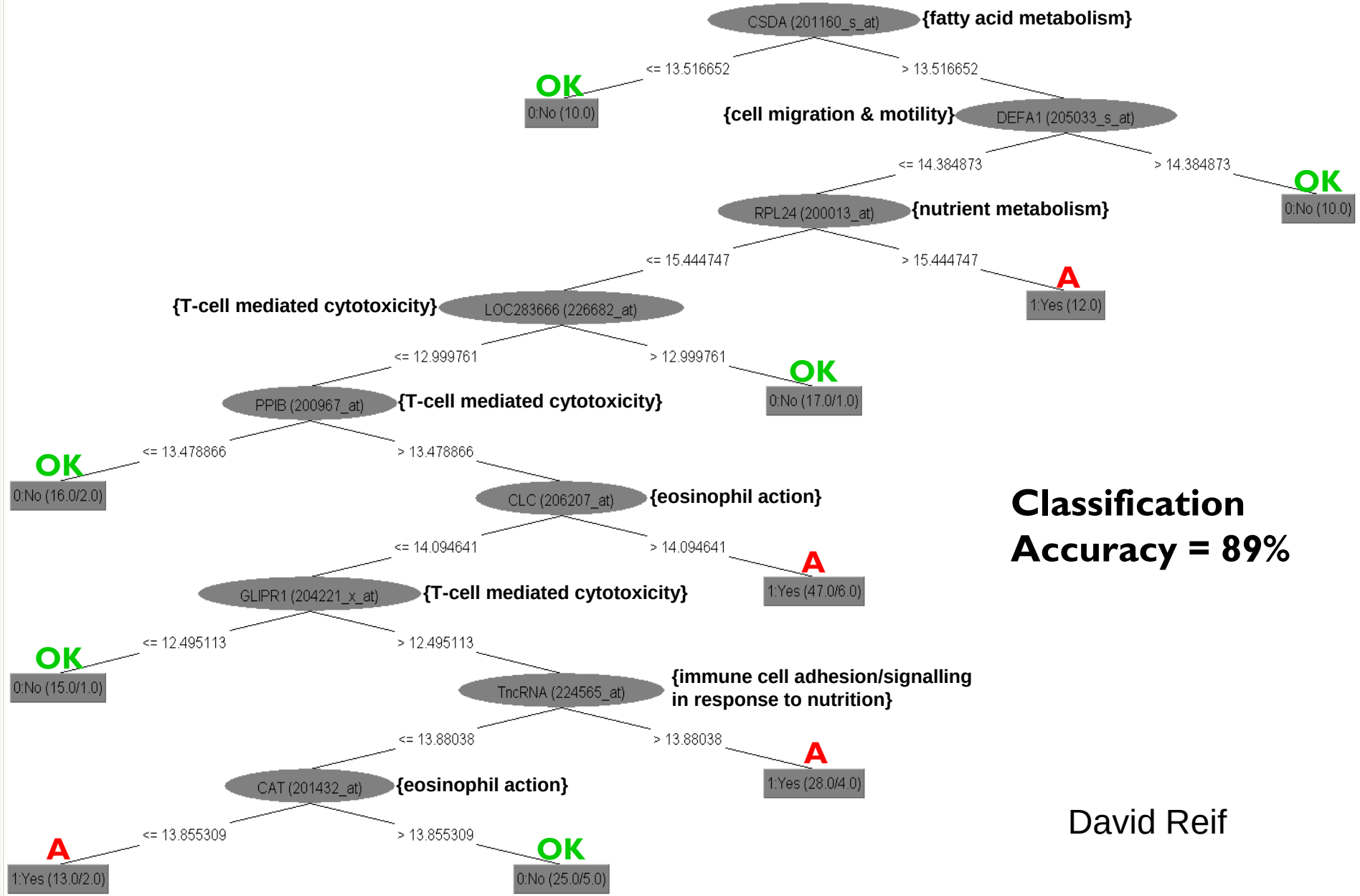
Classification Accuracy = 89%

David Reif


CSDA
DEFA1
RPL24
GLIPR1
CAT
CLC
PPIB
LOC283666
TncRNA
{fatty acid metabolism}
{cell migration & motility}
{nutrient metabolism}
{T-cell mediated cytotoxicity}
{eosinophil action}
{T-cell mediated cytotoxicity}
{T-cell mediated cytotoxicity}
{immune cell adhesion/signalling in response to nutrition}

David Reif

Can we leverage covariate information to put gene expression asthma classifier results in context?



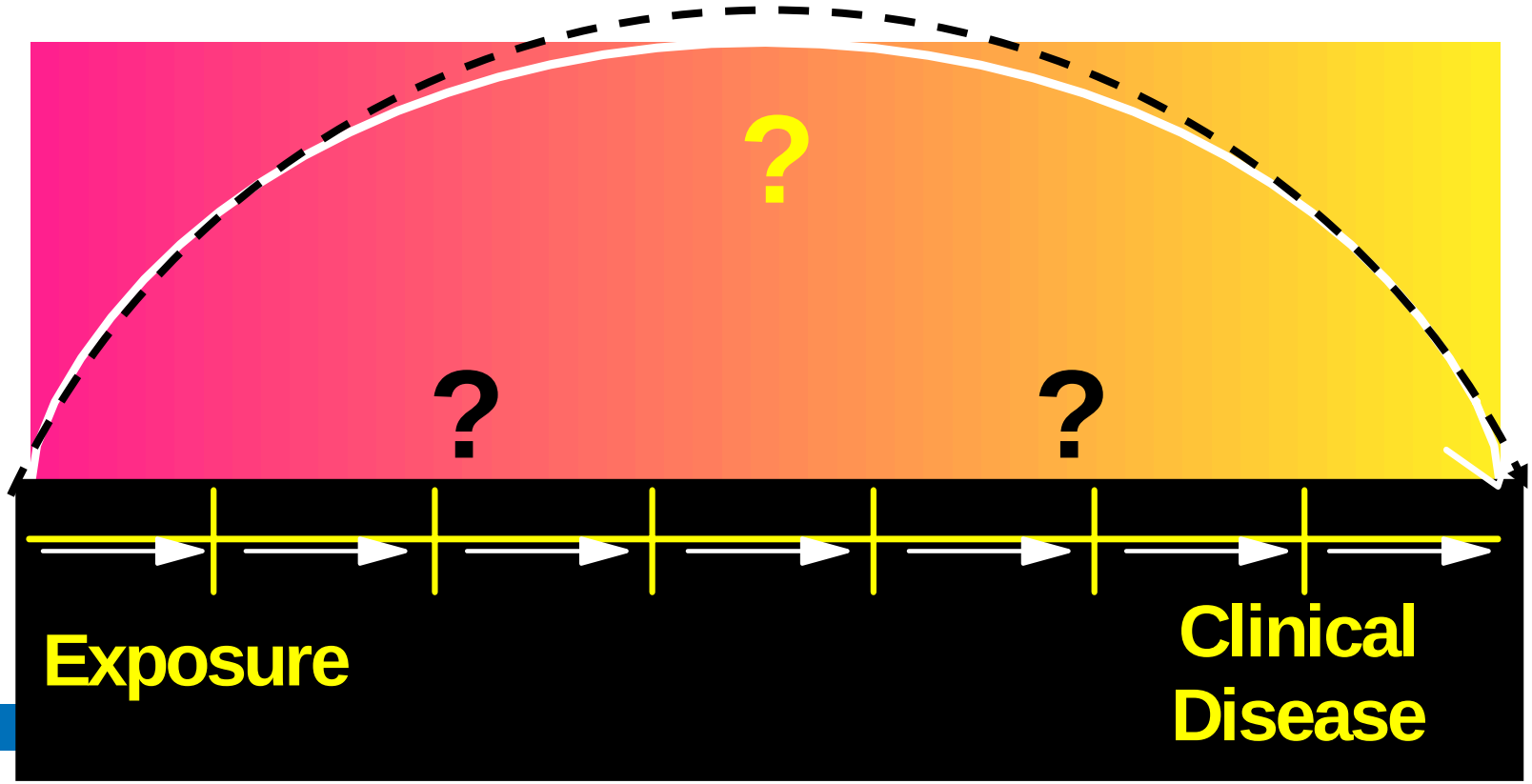
Classification Accuracy = 89%

David Reif

United States Environmental Protection Agency

Summary

- Integration of diverse set of exposure, effects and susceptibility measures
- High-data content technologies, elucidating the genetic and environmental basis for toxicity and disease



United States Environmental Protection Agency

Acknowledgements

- **Jane Gallagher**, Human Studies Division, NHEERL
- **Stephen Edwards**, National Human and Environmental Effects Laboratory
- **David Reif**, National Center for Computational Toxicology
- **Wendell Jones**, Expression Analysis
- **John Wambaugh**, National Center for Computational Toxicology

Disclaimer

Although this work was reviewed by EPA and approved for presentation, it may not necessarily reflect official Agency policy.