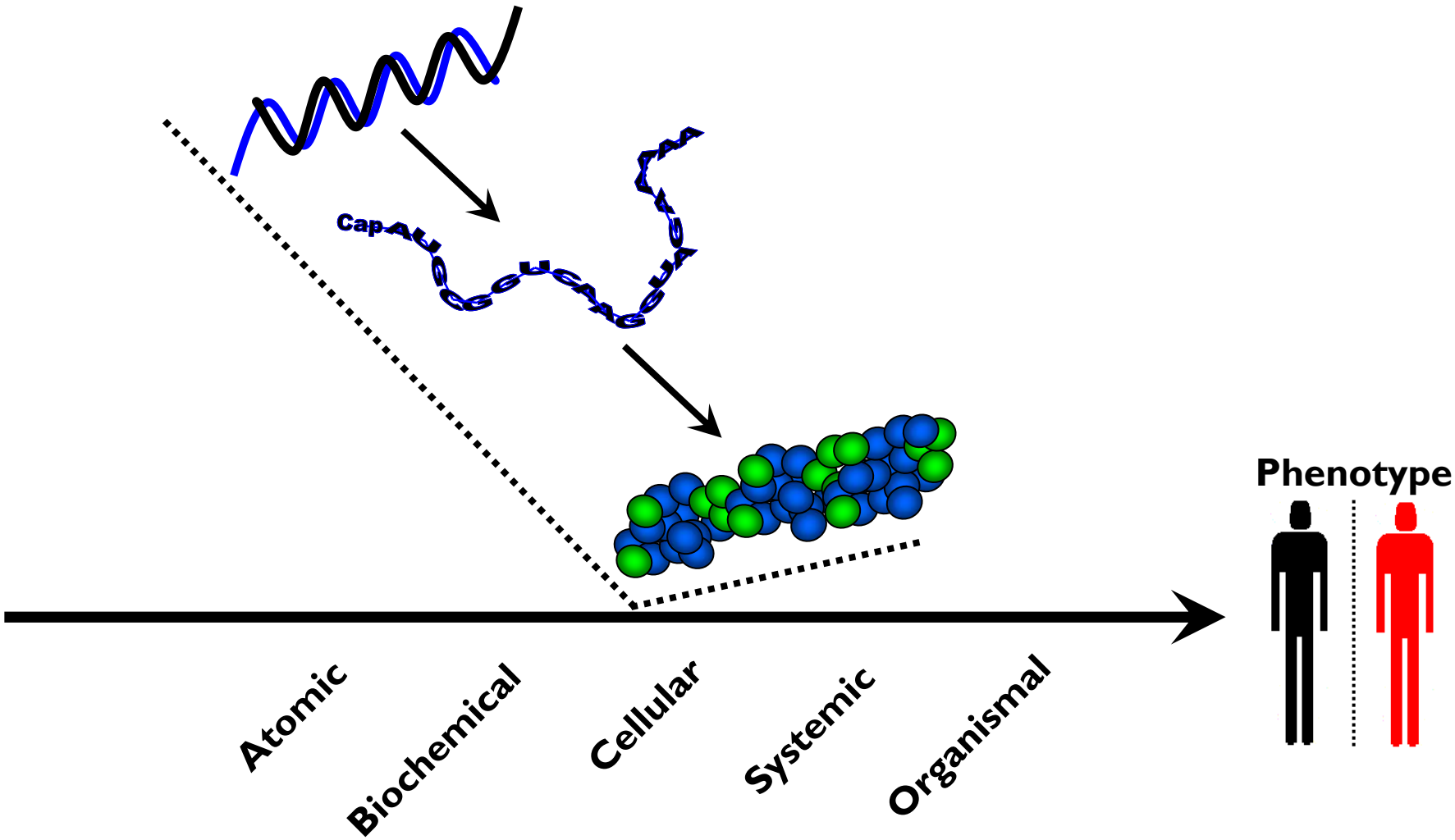
Integrating demographic, clinical, and environmental exposure information to identify genomic biomarkers associated with subtypes of childhood asthma

David Reif, Ph.D.

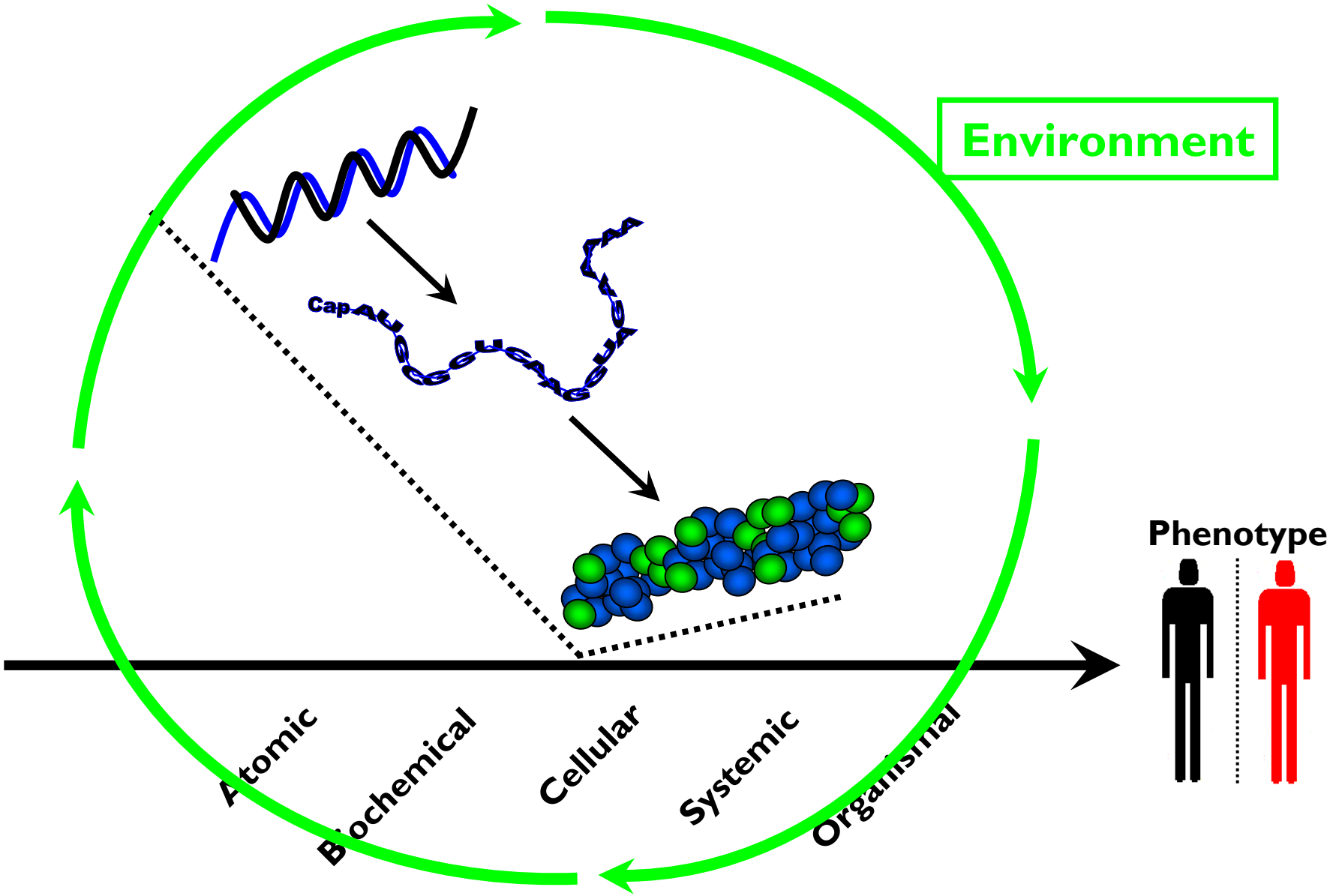
United States Environmental Protection Agency

Office of Research and Development
National Center for Computational Toxicology

We analyze only a slice of the information related to complex phenotypes



We analyze only a slice of the information related to complex phenotypes



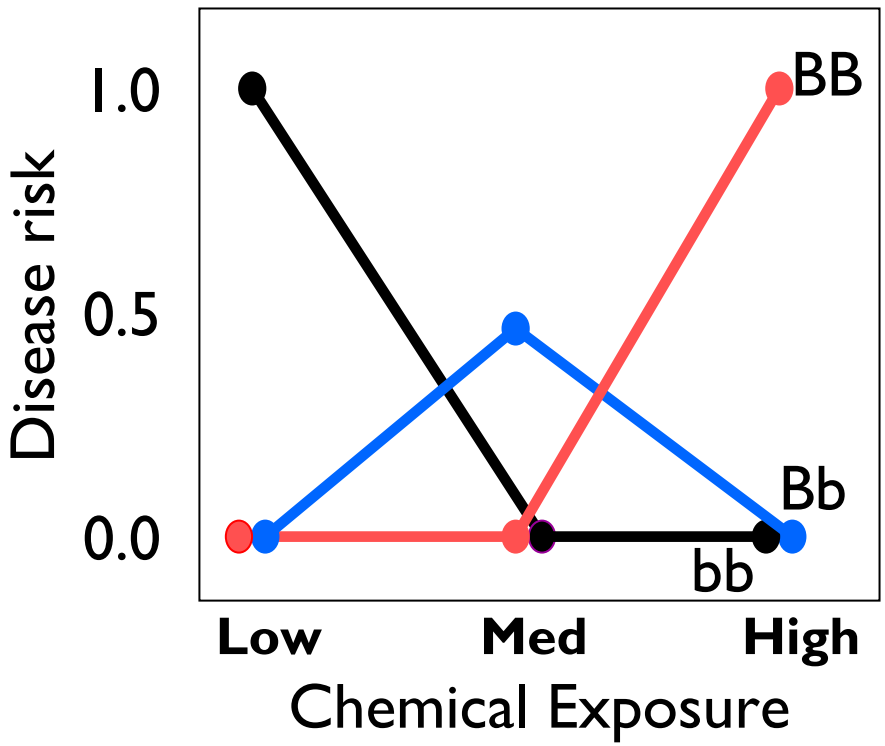
Ecogenetics

"genetic determinants that dictate susceptibility to environmentally influenced adverse health effects"

[Costa and Eaton (2006)]

"Genes load the gun. The environment pulls the trigger."

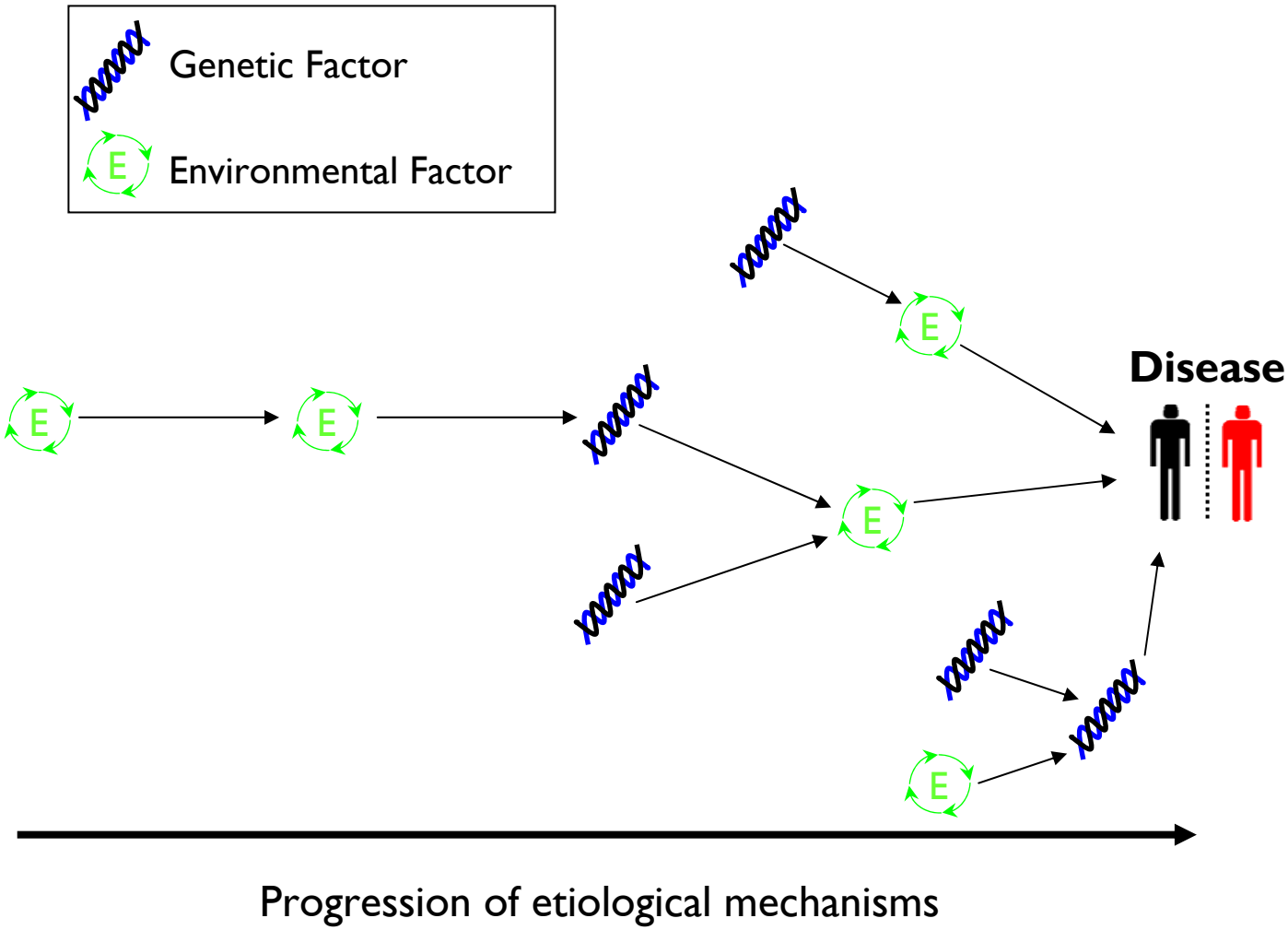
[Bray (1998)]



SNP Genotypes = {BB, Bb, bb}

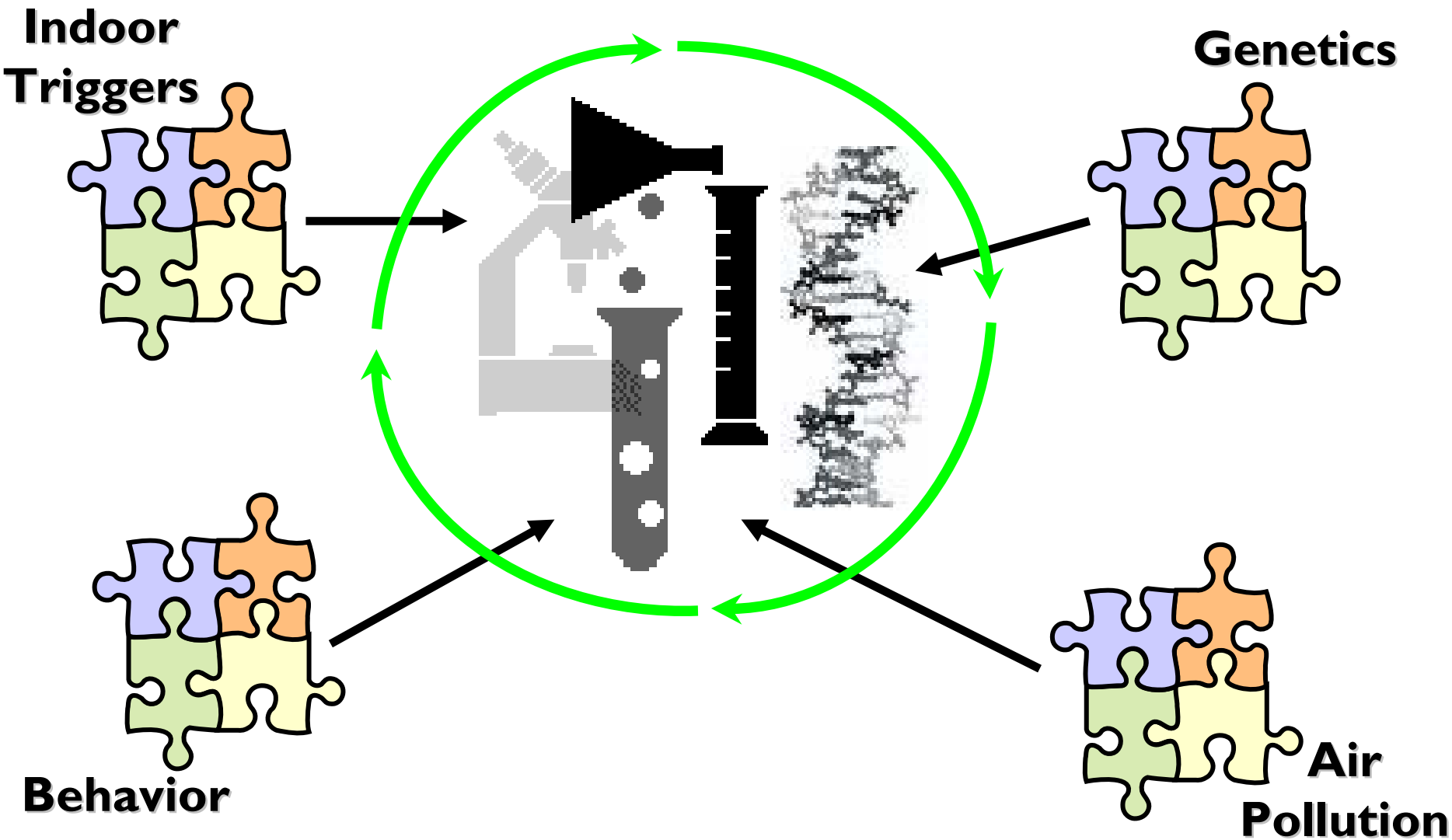
[Figure adapted from Jason Moore]

Complex diseases involve multiple etiological pathways



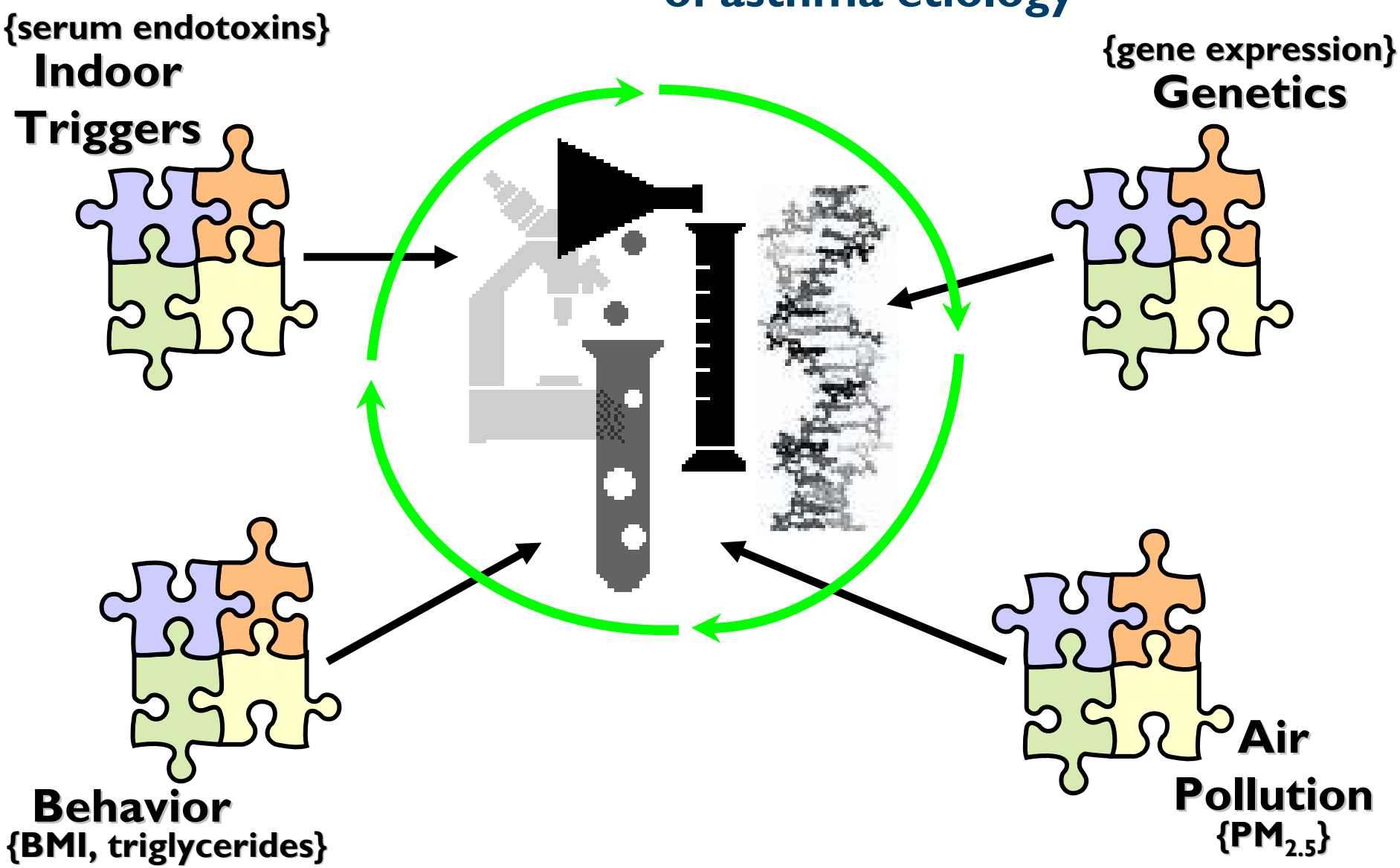
EPA United States Environmental Protection Agency

Asthma etiology



[Figure adapted from Jane Gallagher]

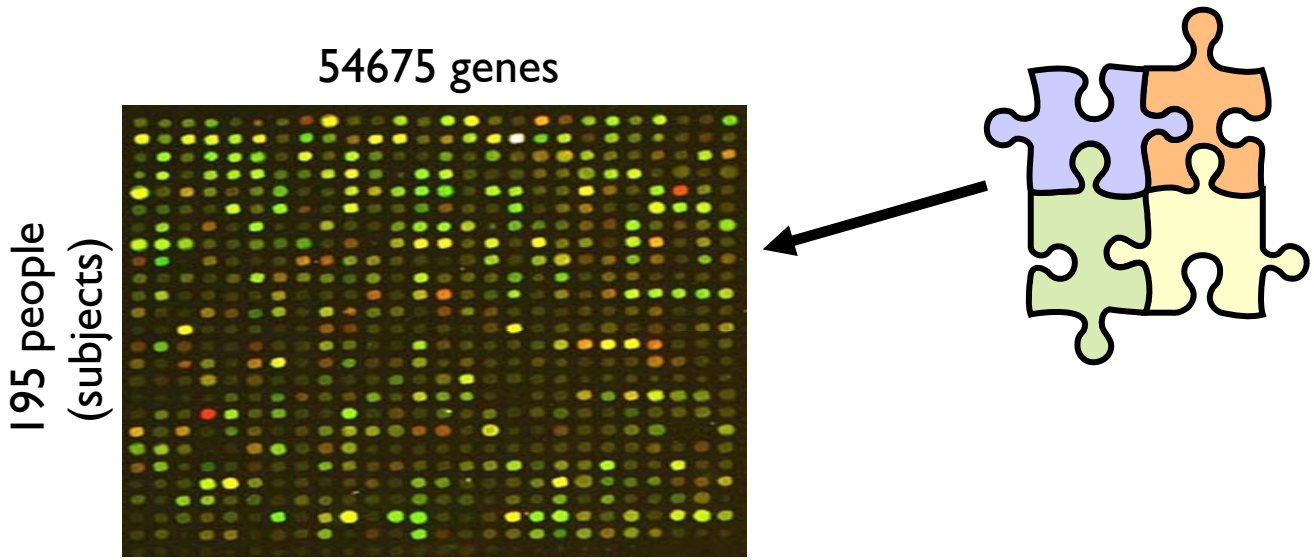
Mechanistic Indicators of Childhood Asthma (MICA) explores multiple aspects of asthma etiology



Gene expression measured using oligonucleotide microarrays

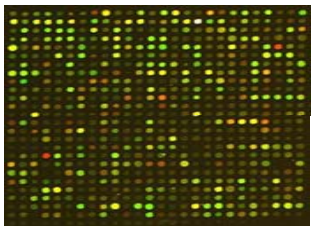
{gene expression}

Genetics



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

Worst. Slide. Ever.



Generic analysis via my favorite analysis method

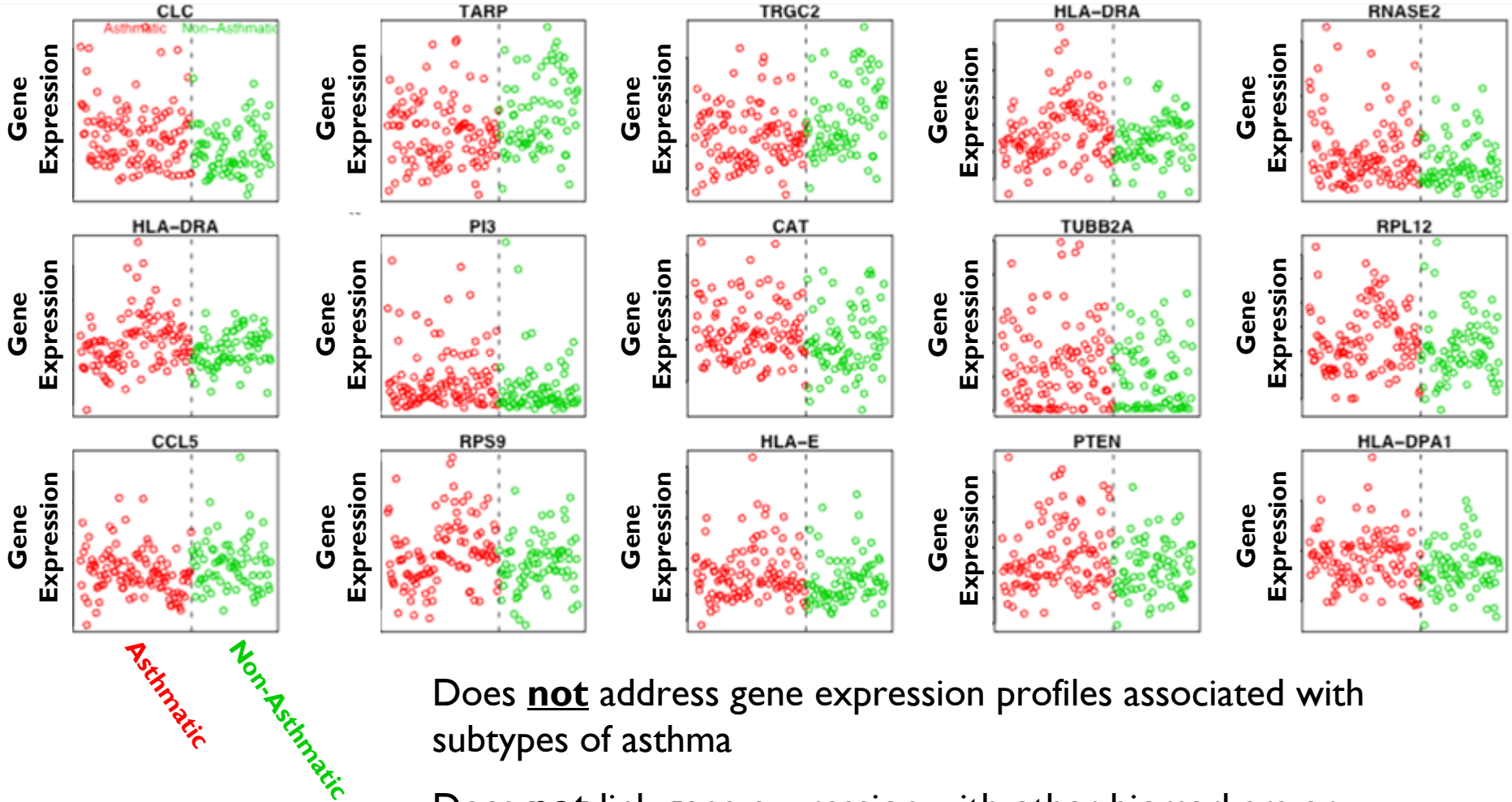
Gene Symbol	Probe ID	p-value (Asthmatic vs. Non-Asthmatic)
CLC	254666_s_at	0.0046
TARP	1552398_a_at	0.0067
TRGC2	1553993_s_at	0.0087
HLA-DRA	1554899_s_at	0.0074
RNASE2	1555349_a_at	0.0050
HLA-DRA	1555759_a_at	0.0014
PI3	200059_s_at	0.0075
CAT	200063_s_at	0.0013
TUBB2A	200065_s_at	0.0092
RPL12	200074_s_at	0.0061
CCL5	200080_s_at	0.0010
PRS9	200086_s_at	0.0030
HLA-E	200088_x_at	0.0021
•	•	•
•	•	•
•	•	•
PTEN	200091_s_at	0.0026
HLA-DPA1	200094_s_at	0.0045

Provides proof of my ability to generate thousands of p-values

Gives a long list of results that are essentially presented in a contextual vacuum

EPA
United States
Environmental Protection
Agency

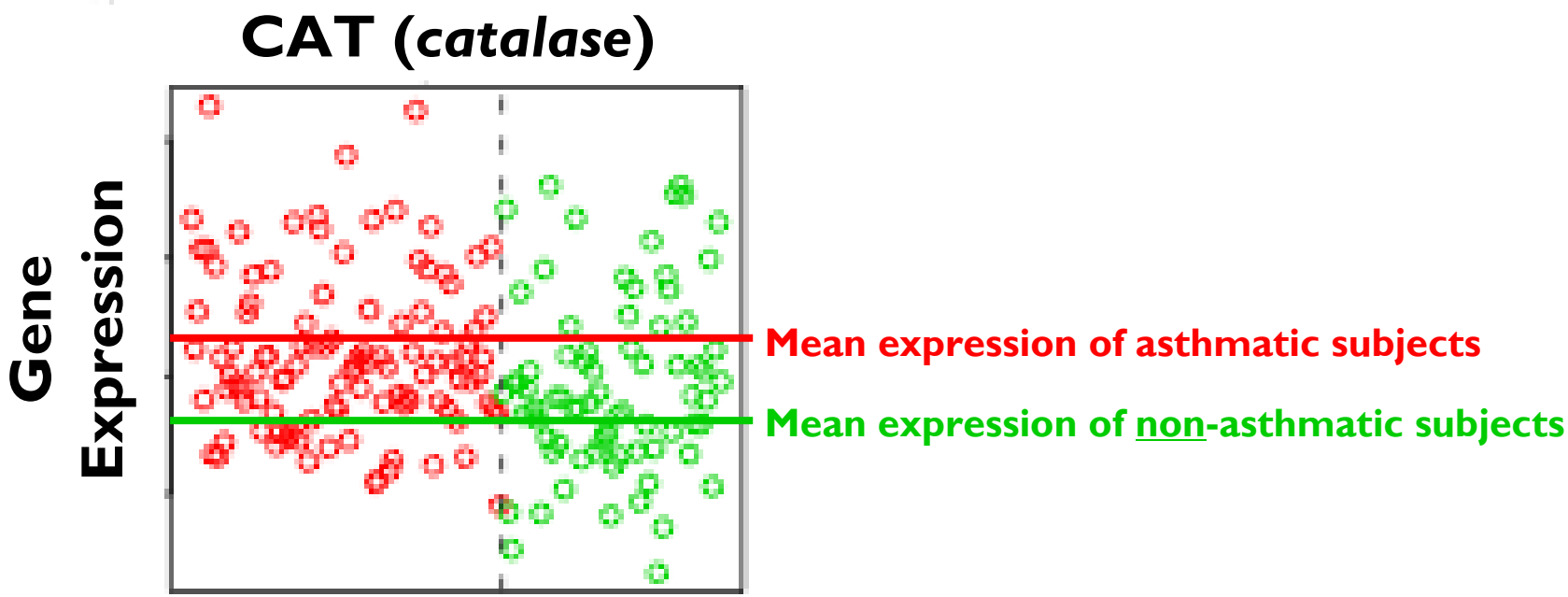
Why not just do the usual "here are some main-effect genes that discriminate asthmatics versus non-asthmatics"?



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

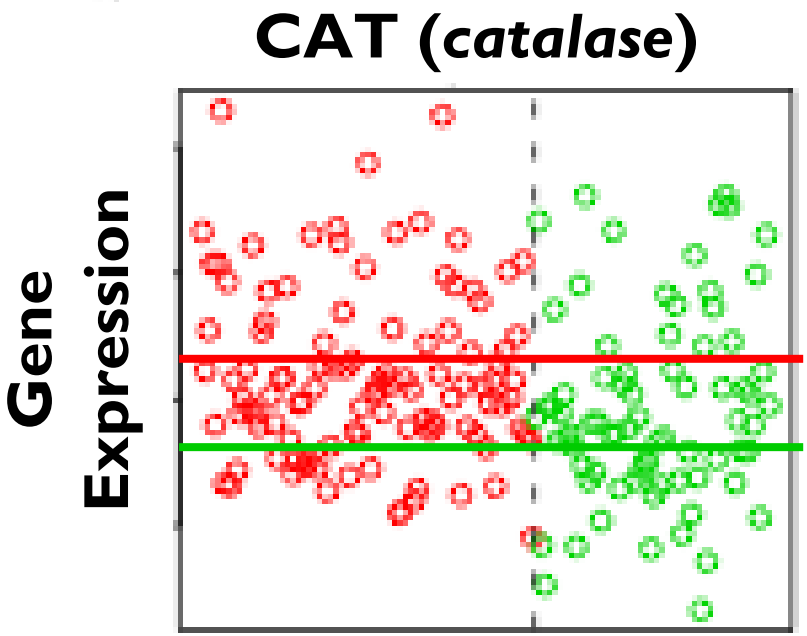
Does **not** link gene expression with other biomarkers or covariates (Where is the context?)

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

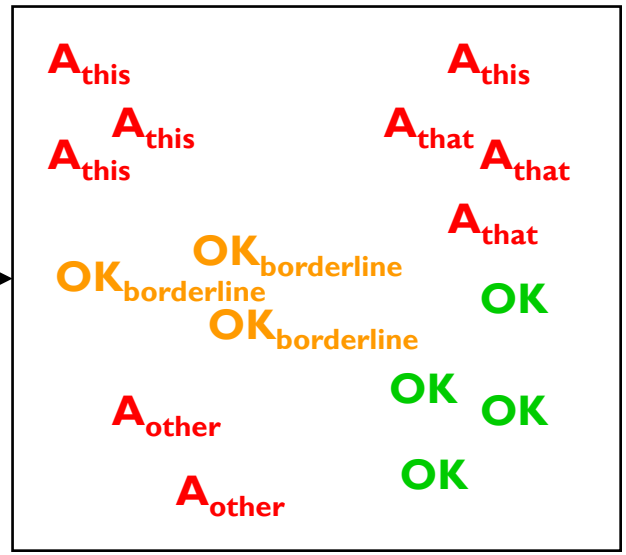


Reliance on group means **ignores** the complexity of asthma etiology

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

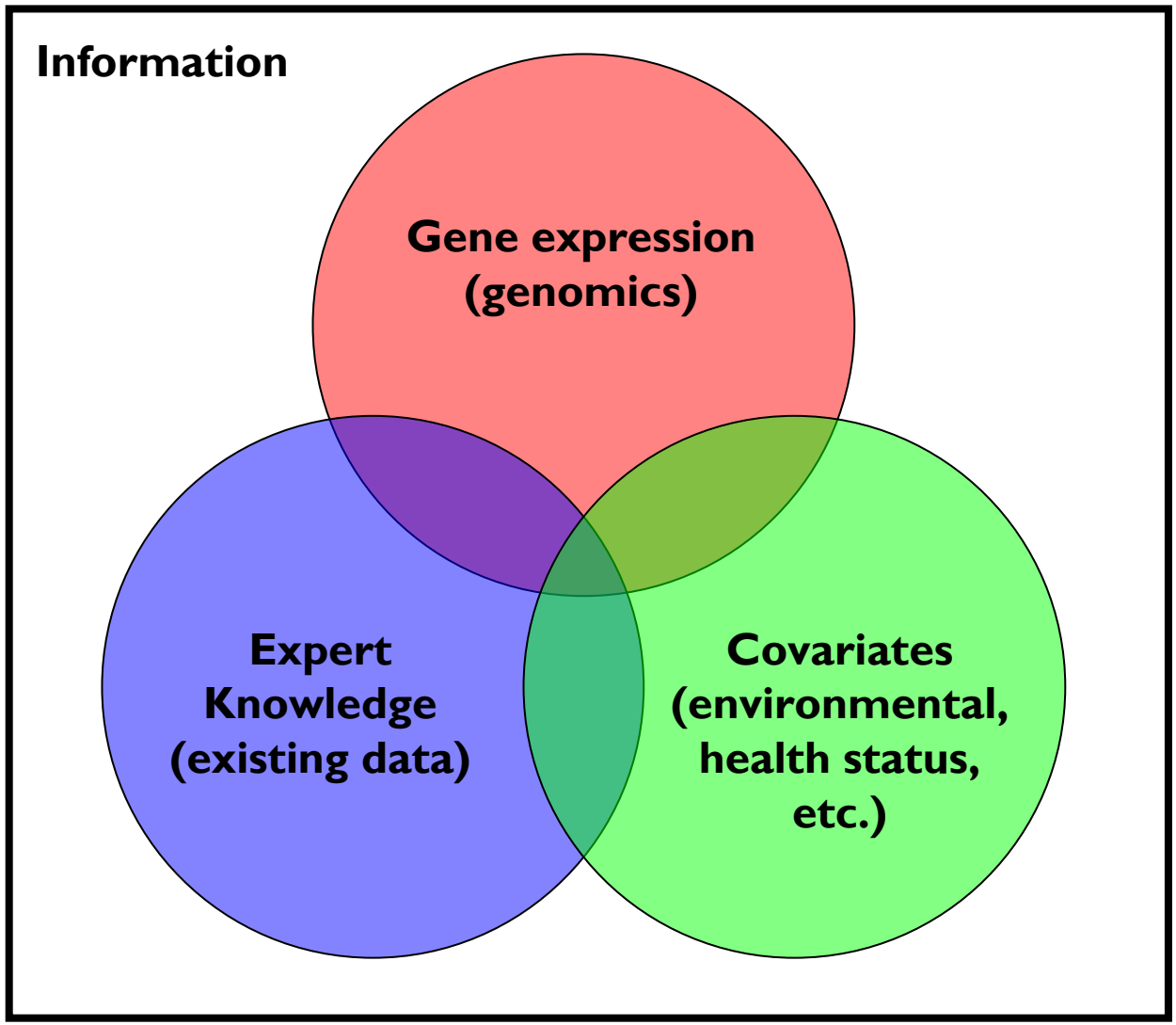


Reliance on group means ignores the **complexity** of asthma etiology



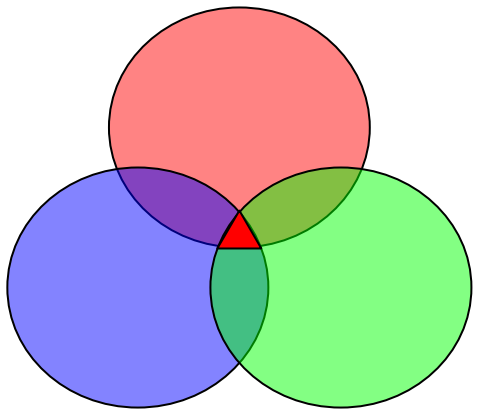
The ultimate goal is to glean mechanistic information regarding asthma subtypes

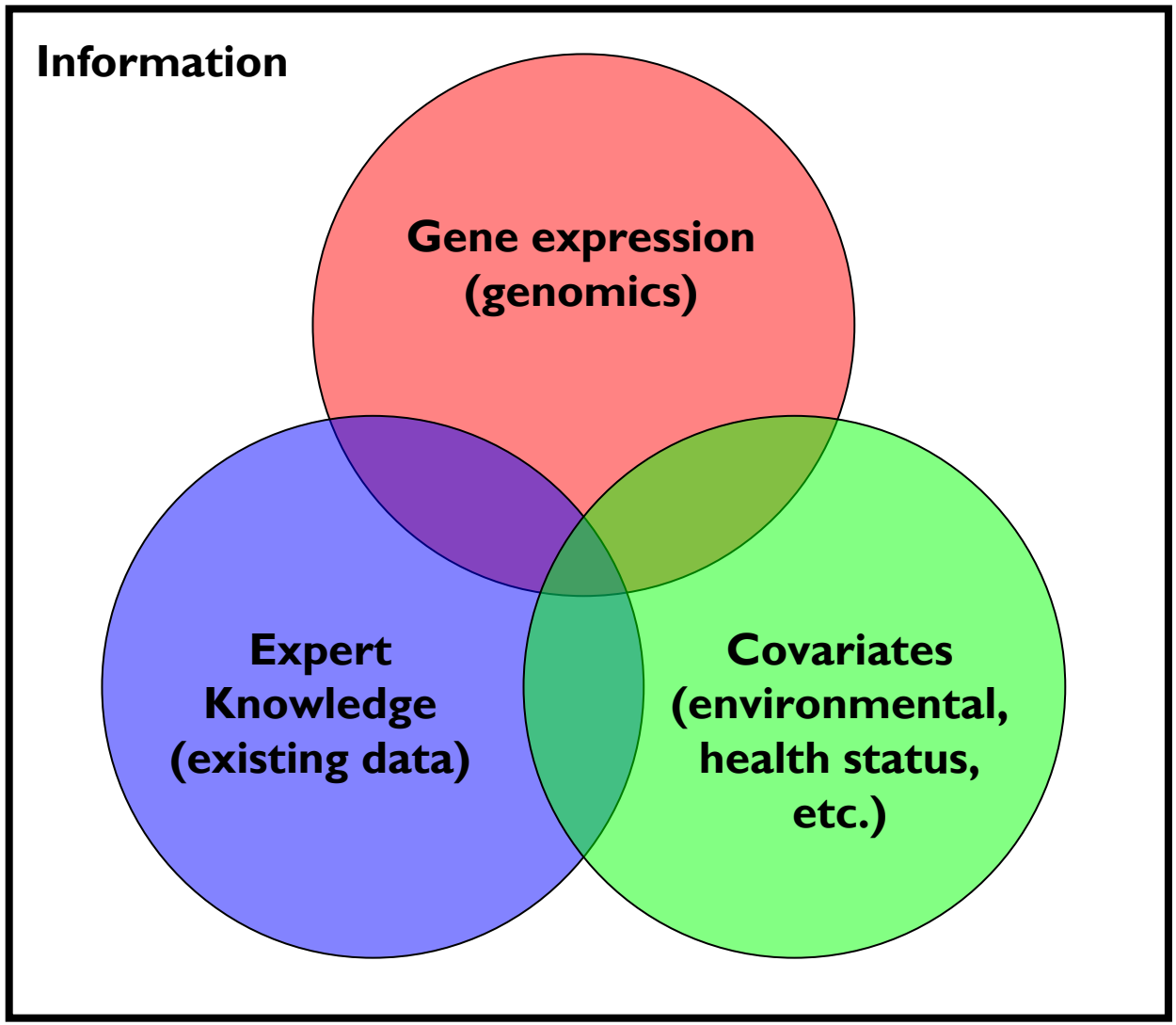
Is diagnosis as "asthmatic" a strong, homogeneous phenotype?



The ultimate goal is to glean mechanistic information regarding asthma subtypes

The intersection of these information sources can provide mechanistic **context**

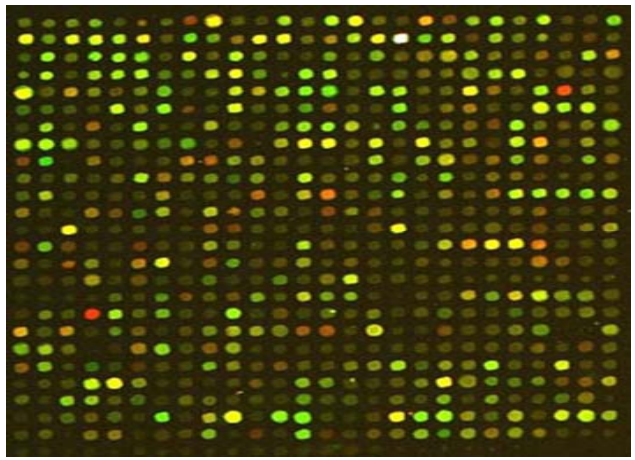




How do we leverage information from MICA covariates for the gene expression analysis?

54675 genes

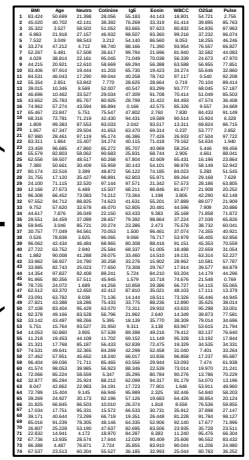
195 people (samples)



?

MICA covariates

195 people (samples)



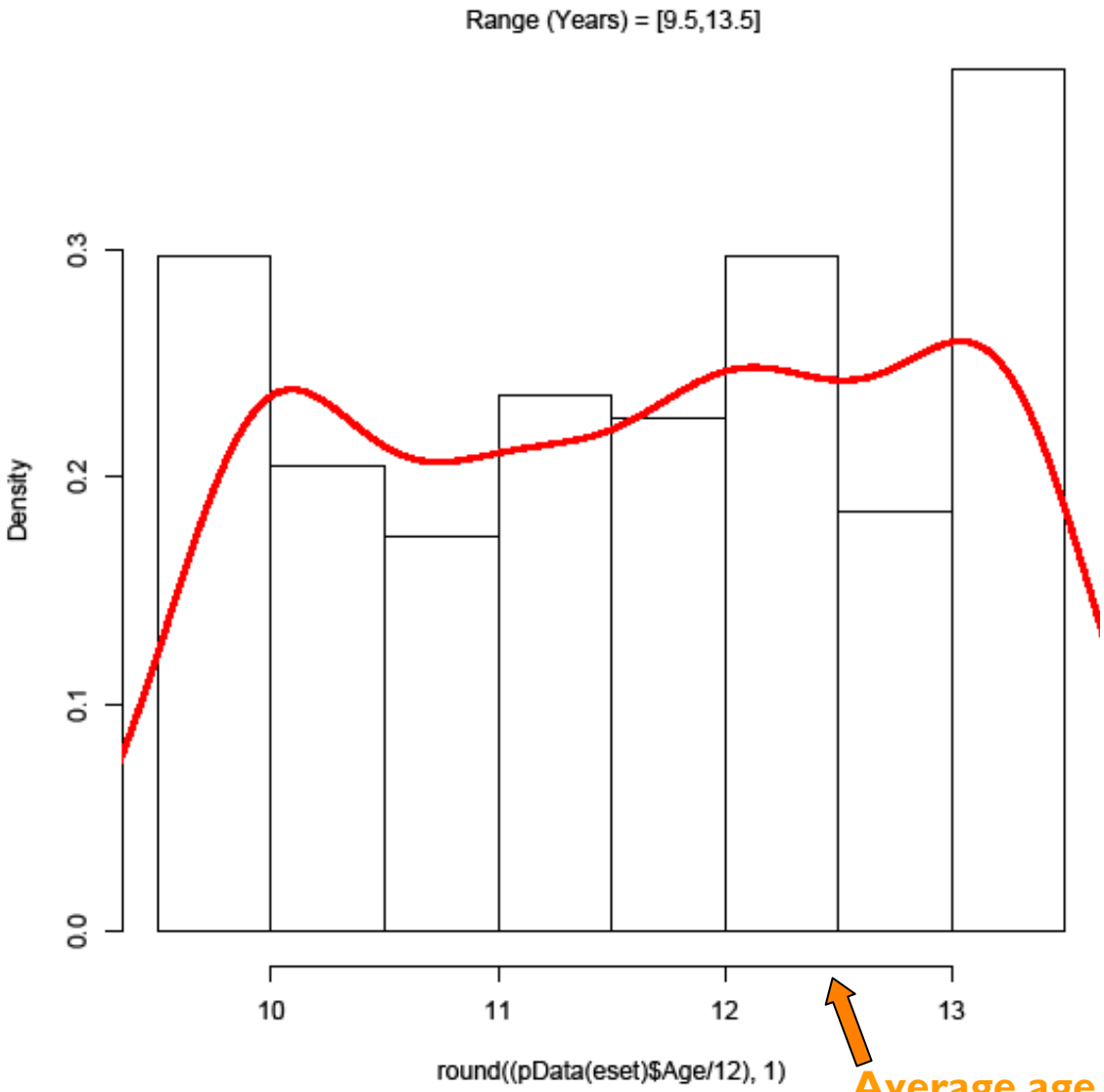
Partial list of MICA covariates

VariableName	Label
AgeYrs	Age at DrawDate in years
BMIPCT	BMI percentile
BMIZ	BMI z-score
HAZ	Height z-score
HTPCT	Height percentile
WAZ	Weight z-score
WHPCT	Weight-for-height percentile
WHZ	Weight-for-height z-score
WTPCT	Weight percentile
O2Sat	BP O2 Saturation (%)
Pulse	BP Pulse, Beats/min

• • •

D5AnyMedUse	Any Asthma Med Use
D6DrDiagnosis	Dr. Diagnosed Asthama (Q37)
D8Symptoms	Asthma symptoms from Q37 and Q38
D9IllWithAsthma	Ill With Asthma in last 12 months (L048)
L_JROS	Log10(JROS)
MeanDia	Mean of first two diastolic measurements
MeanSys	Mean of first two systolic measurements
PctI_Baso	Basophil percent of sum WBC
PctI_Eosino	Eosinophil percent of sum WBC
PctLymph	Lymphocyte percent of sum WBC
PctMono	Monocyte percent of sum WBC
PctNeut	Neutrophil percent of sum WBC

What are the 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**

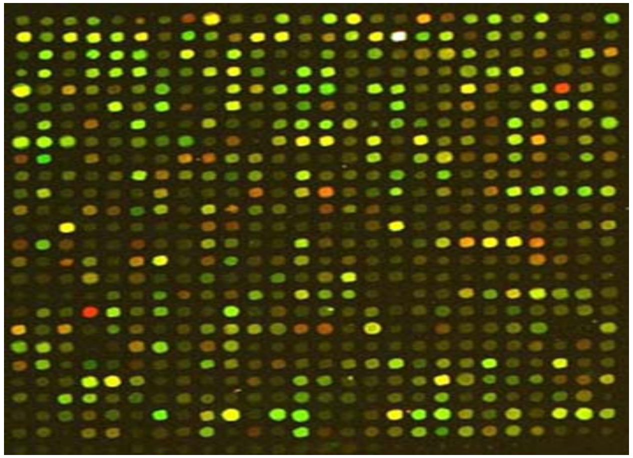
Distribution of age across all categories is fairly uniform

Average age at menarche (nation-wide)

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

54675 genes

195 people (samples)



?

MICA covariates

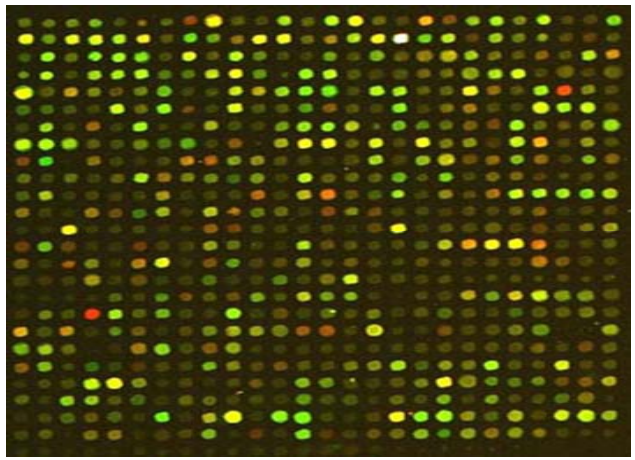
195 people (samples)

	BMI	Age	Neutro	Cotinine	IgE	Eosin	WBCC	O2Sat	Pulse
1	83.424	50.699	21.398	28.056	55.183	44.143	18.801	54.721	2.756
2	45.020	40.702	42.141	38.392	76.269	33.319	61.414	39.895	85.763
3	35.322	3.527	58.502	51.052	83.665	97.623	80.832	54.835	87.963
4	6.983	21.918	27.157	46.932	98.507	93.360	99.216	37.232	96.074
5	7.532	3.049	98.543	3.212	54.140	86.560	8.053	18.255	46.246
6	33.274	47.212	4.712	98.740	88.166	71.390	93.954	76.557	66.827
7	52.267	5.481	57.508	36.617	99.794	21.696	81.940	32.582	44.083
8	4.029	38.814	22.161	95.045	71.049	70.038	56.339	24.673	47.970
9	44.215	20.921	12.610	58.669	69.294	56.388	63.588	56.955	77.851
10	83.406	97.914	19.444	41.203	65.730	19.423	34.274	55.646	32.860
11	84.531	46.043	17.290	99.044	40.258	78.742	97.117	0.540	92.208
12	55.354	2.851	53.842	7.770	28.625	28.664	0.719	70.104	89.414
13	39.015	10.346	8.569	52.007	40.547	83.299	93.777	68.045	57.187
14	46.696	10.462	33.527	29.034	47.309	91.708	70.414	51.049	85.503
15	43.652	25.783	85.767	80.925	28.799	31.422	84.493	47.574	36.408
16	74.962	57.274	43.594	86.894	0.166	42.575	85.326	9.557	34.669
17	65.467	23.847	5.751	75.547	47.802	2.760	7.830	84.433	91.149
18	68.316	72.781	71.219	42.430	94.431	19.589	90.514	15.924	6.328
19	1.809	99.383	87.553	93.033	2.042	93.517	13.311	69.824	86.715
20	1.957	67.347	29.504	41.653	63.470	69.314	0.237	53.777	2.852
21	87.980	28.461	97.119	95.174	46.386	77.426	26.933	47.504	97.722
22	83.311	1.864	15.407	34.274	40.115	71.418	79.162	54.634	1.940
23	23.459	96.685	47.860	85.272	85.707	40.969	58.254	5.446	99.456
24	55.579	82.803	58.062	90.683	25.831	68.744	2.905	4.661	88.041
25	62.556	59.507	48.517	60.268	67.804	42.609	65.431	16.196	23.864
26	7.380	50.661	30.409	55.959	30.143	54.101	98.978	58.149	52.942
27	80.174	22.516	3.399	48.872	56.122	74.185	84.023	5.283	51.545
28	31.755	17.120	35.427	98.891	62.803	55.971	89.264	29.168	7.628
29	24.100	71.115	32.520	97.144	97.571	21.342	57.573	28.186	63.805
30	12.166	27.673	6.469	15.507	68.211	86.846	81.677	21.908	20.252
31	96.308	66.452	73.083	74.212	73.364	1.198	3.805	42.290	62.148
32	97.552	94.712	88.825	74.623	41.631	55.201	37.889	89.077	21.909
33	9.752	57.620	32.678	46.070	52.805	20.481	44.596	7.909	20.886
34	44.617	7.876	36.049	22.150	63.433	9.383	35.168	71.858	71.872
35	29.551	34.459	37.089	28.657	79.392	98.864	37.224	27.039	65.826
36	59.945	3.046	85.721	20.274	23.386	2.473	75.578	38.732	60.041
37	30.757	77.049	94.561	70.053	1.500	96.461	37.074	74.355	40.921
38	0.526	78.838	3.825	96.065	9.066	76.717	53.978	56.405	62.856
39	96.062	42.434	46.484	68.965	80.308	88.416	91.151	45.353	70.559
40	27.722	63.752	2.940	25.584	68.337	51.005	18.488	22.659	31.054
41	1.882	90.008	41.288	29.075	33.460	14.510	19.131	63.314	52.227
42	33.962	58.927	24.790	30.258	33.276	15.902	28.962	10.581	57.787
43	33.985	82.743	25.023	77.650	73.308	29.767	17.914	39.577	84.879
44	14.354	97.837	82.408	89.241	5.724	84.210	93.204	14.179	44.296
45	91.965	80.356	57.511	45.945	1.579	10.718	74.676	21.548	0.769
46	78.725	24.072	1.689	44.256	10.858	29.386	66.727	54.115	55.065
47	62.512	63.370	12.650	42.412	87.810	35.021	48.103	17.111	13.379
48	23.091	63.782	8.038	71.136	14.144	18.511	73.326	56.446	44.945
49	27.821	43.388	18.286	75.433	33.775	88.236	12.890	35.625	38.014
50	27.108	83.404	95.429	63.070	73.311	29.932	43.633	95.298	33.001
51	92.378	49.166	83.528	56.796	21.962	2.640	14.349	39.873	77.581
52	33.142	43.497	98.266	5.395	18.139	35.770	38.309	79.014	34.696
53	5.751	15.764	93.527	31.950	9.311	3.138	83.967	53.615	21.794
54	14.053	50.860	3.905	57.539	89.398	49.216	79.412	83.137	76.640
55	11.218	19.453	44.108	11.702	69.152	11.149	95.328	13.192	17.844
56	21.321	17.768	85.187	56.433	62.839	72.475	19.329	34.535	34.331
57	74.531	49.641	32.365	84.269	42.298	52.458	31.040	7.401	33.264
58	27.462	57.951	45.652	18.240	66.017	10.836	96.858	17.237	95.687
59	96.404	69.036	71.711	85.465	63.556	29.944	53.093	7.474	61.938
60	41.574	98.053	39.985	56.923	88.346	22.539	73.014	19.970	21.241
61	72.066	85.224	58.559	5.347	35.295	80.784	90.276	13.786	70.229
62	32.877	85.294	25.924	88.212	62.099	94.317	91.179	54.070	13.196
63	8.047	42.862	22.983	34.191	17.723	77.801	1.646	53.911	48.960
64	72.789	15.404	9.145	66.946	95.989	2.325	92.446	54.840	58.203
65	39.269	24.927	30.173	82.196	57.126	19.683	64.426	38.854	20.124
66	31.825	66.845	86.503	10.010	26.374	1.818	8.556	76.536	59.855
67	17.034	17.751	95.331	15.572	66.533	30.731	35.912	37.898	27.147
68	39.171	40.644	73.299	66.719	19.351	26.448	81.228	91.784	98.127
69	85.016	91.239	78.305	48.146	64.335	52.906	92.140	17.677	71.996
70	26.807	25.228	53.190	47.637	82.695	83.506	23.935	35.728	23.511
71	22.832	14.941	4.172	18.970	84.287	8.383	11.240	95.476	66.304
72	67.736	13.935	28.574	17.644	12.029	90.409	25.606	96.552	83.432
73	86.388	4.487	76.871	2.724	35.855	83.910	90.044	41.206	24.980
74	67.537	23.513	60.304	55.527	36.185	32.993	25.044	80.783	36.252

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

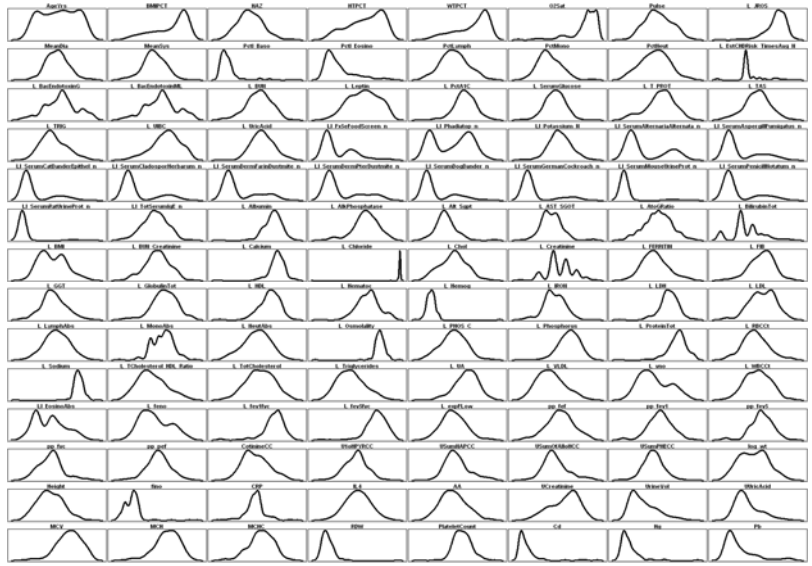
54675 genes

195 people (samples)

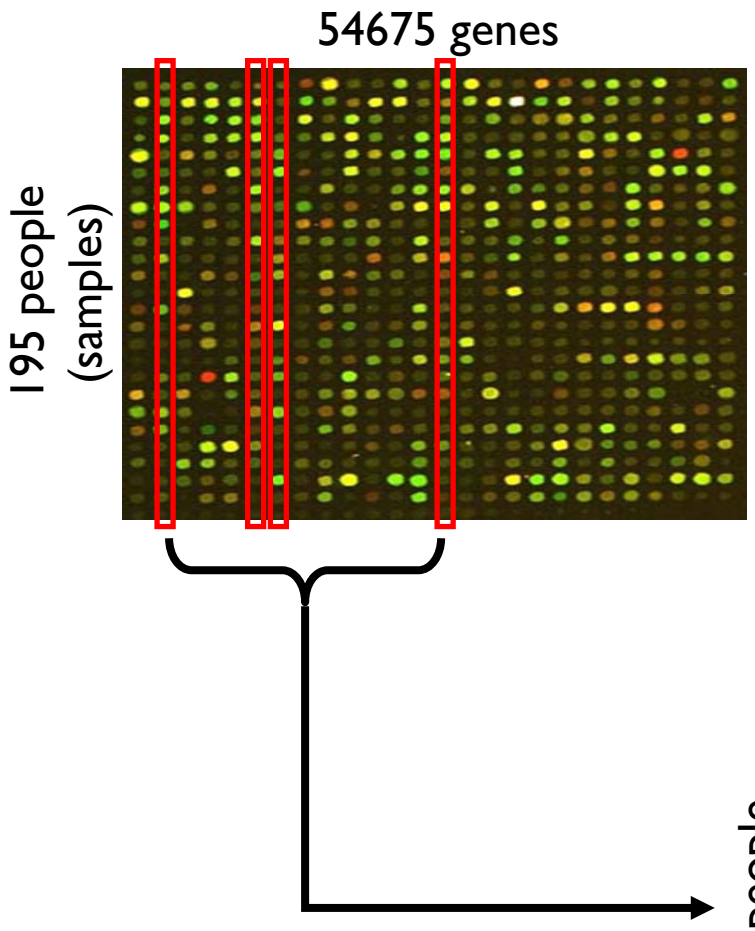


ρ

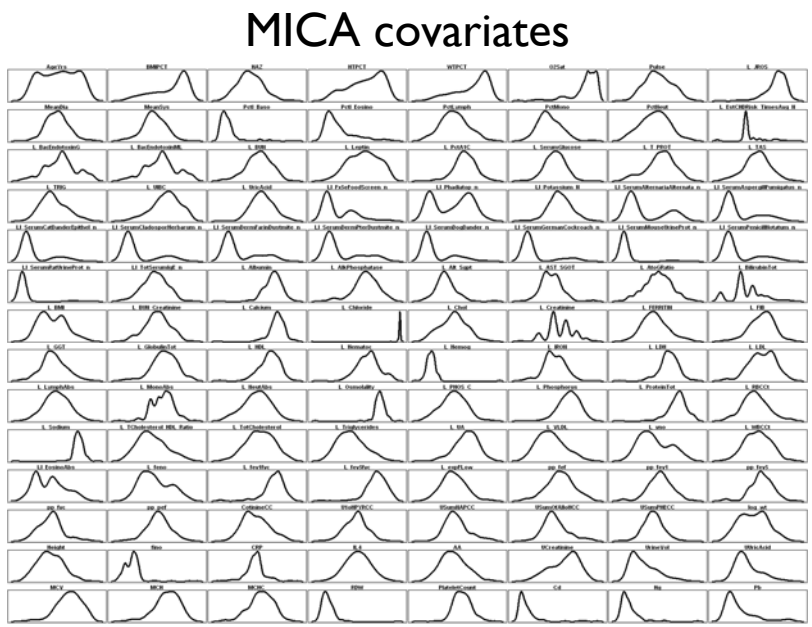
MICA covariates

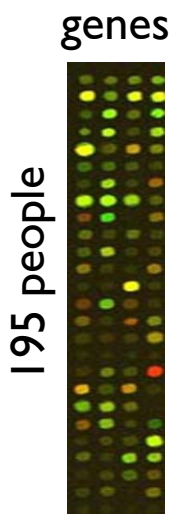


How do we leverage MICA covariate information for the gene expression analysis?



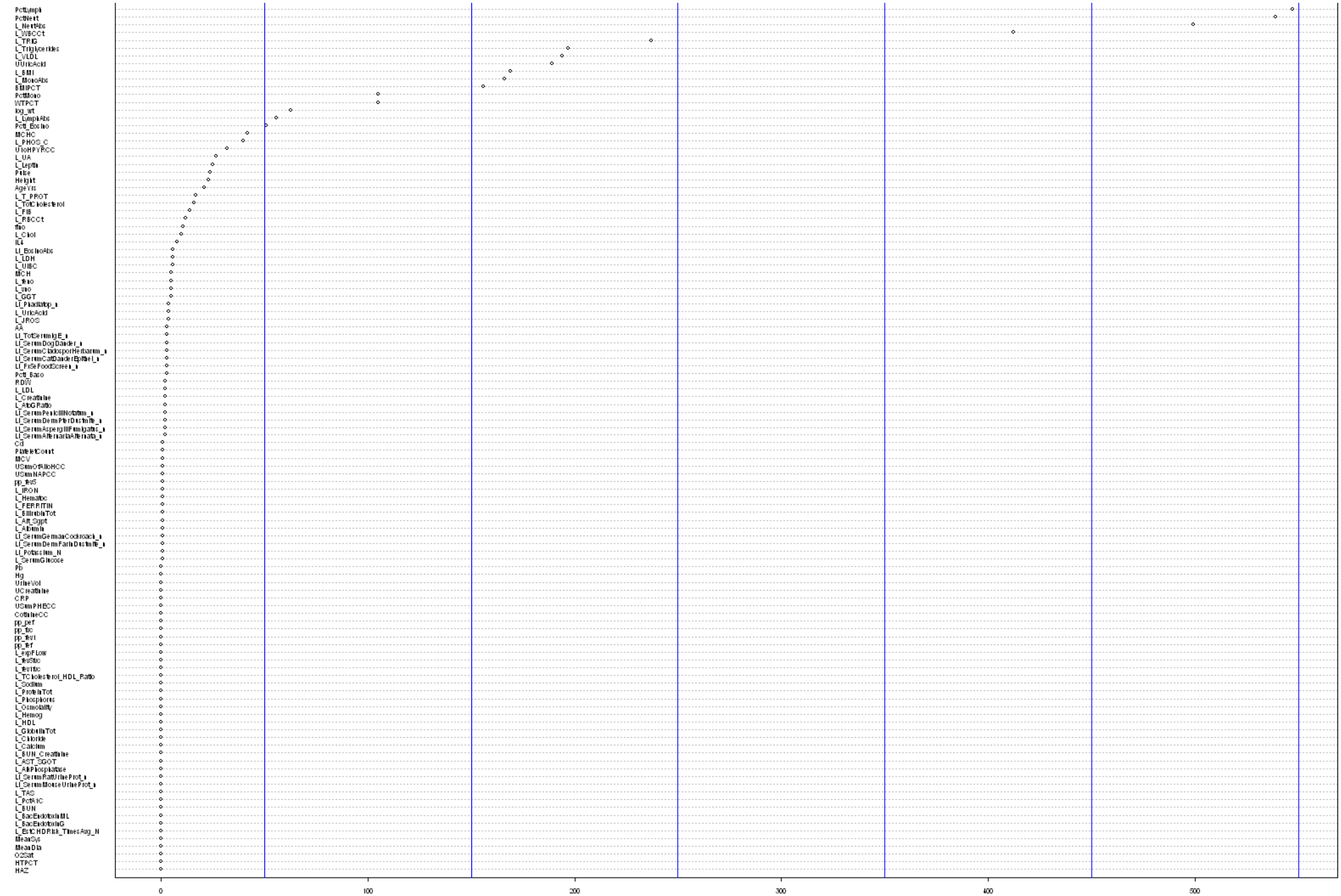
ρ



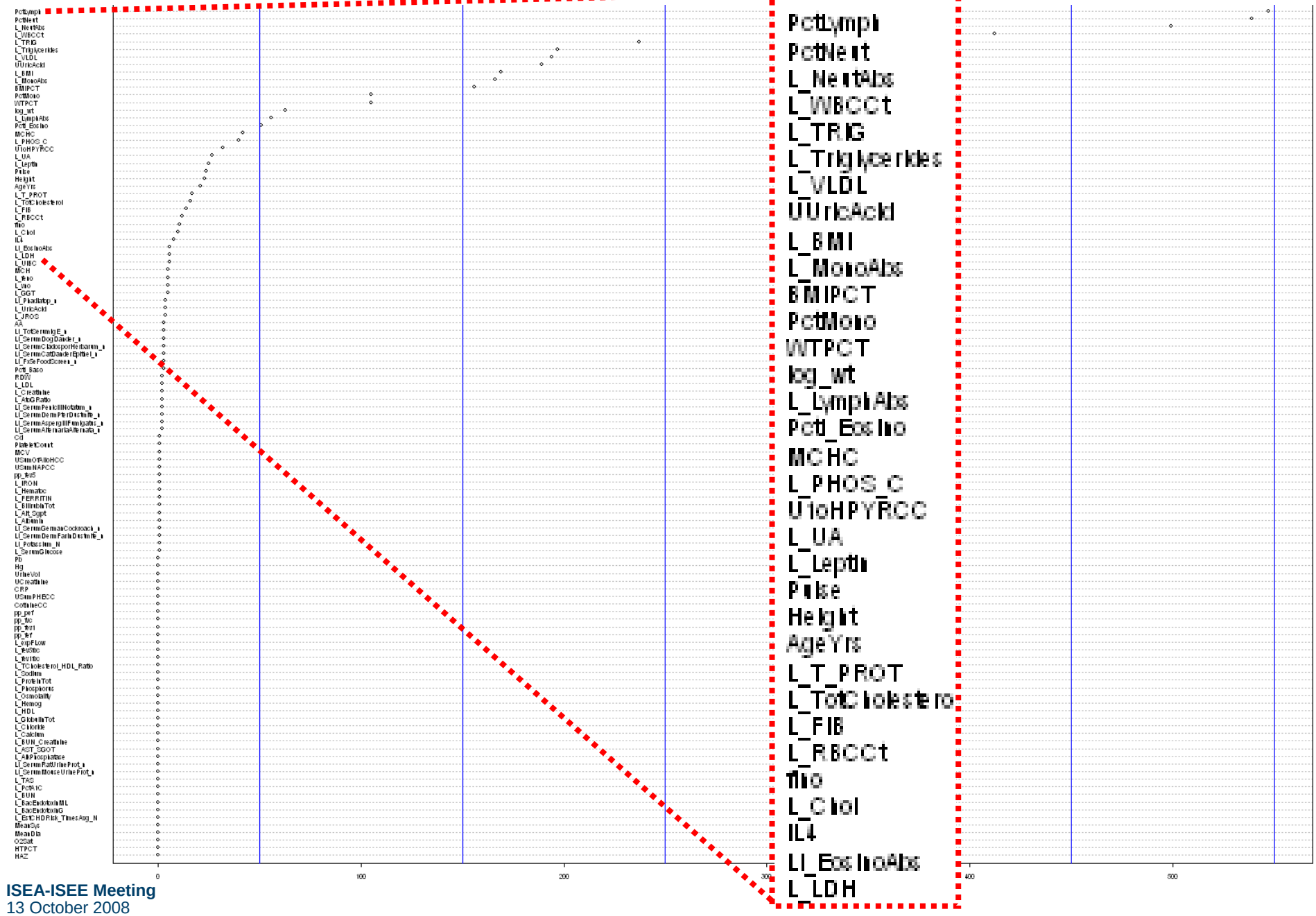


Results in a list of genes having significant correlation with at least one MICA covariate

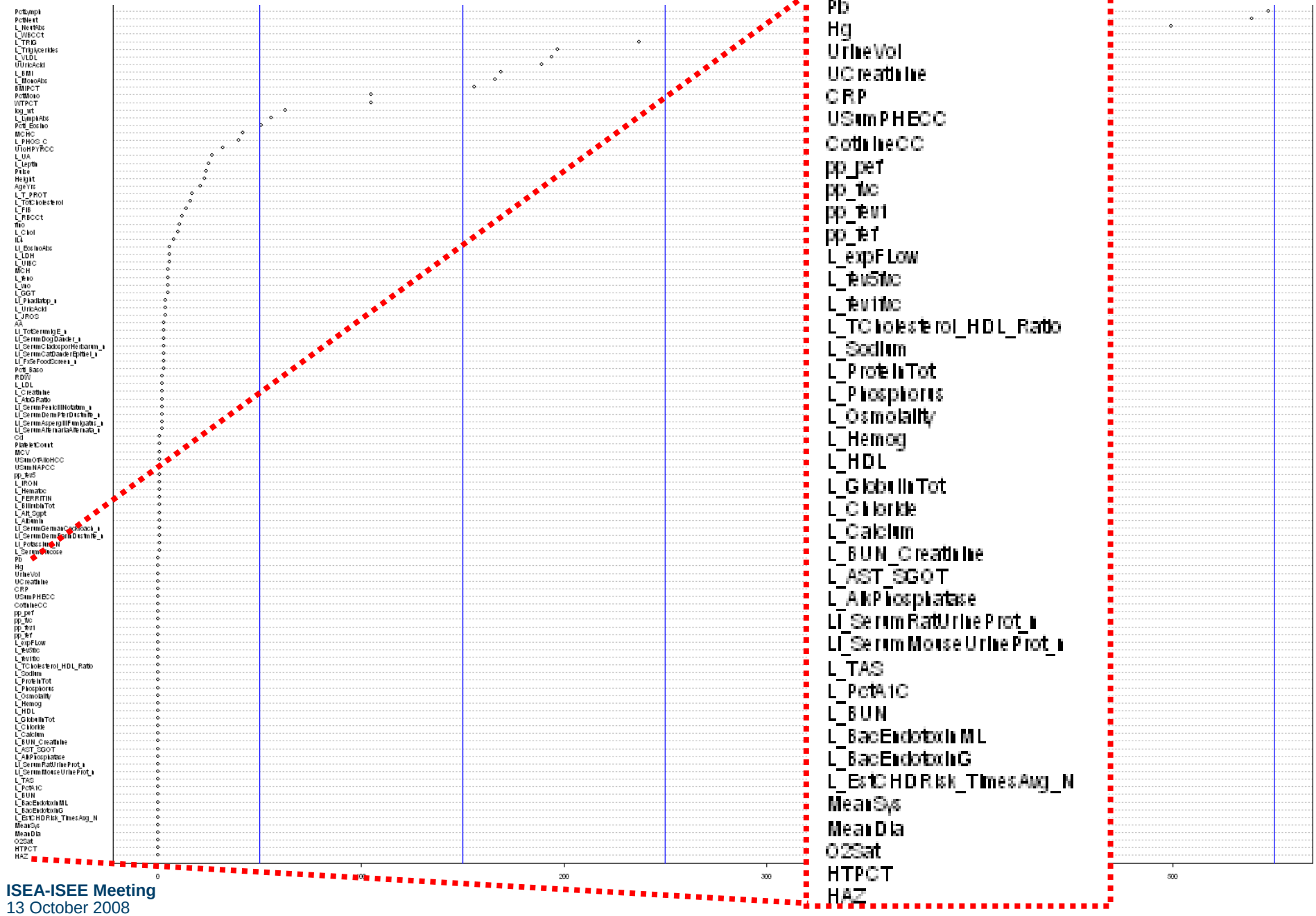
Number of significant gene-covariate correlations



Number of significant gene-covariate correlations

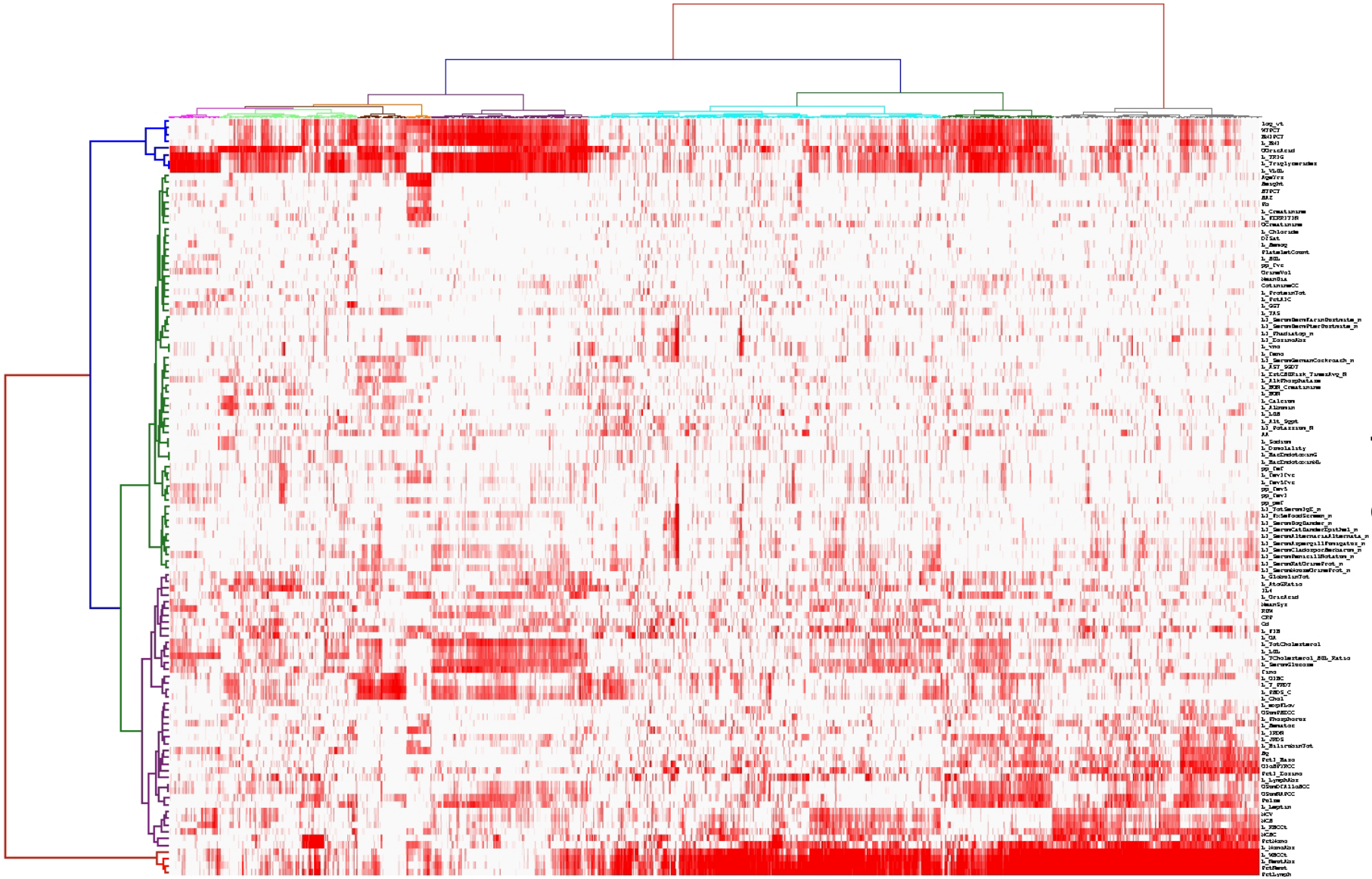


Number of significant gene-covariate correlations

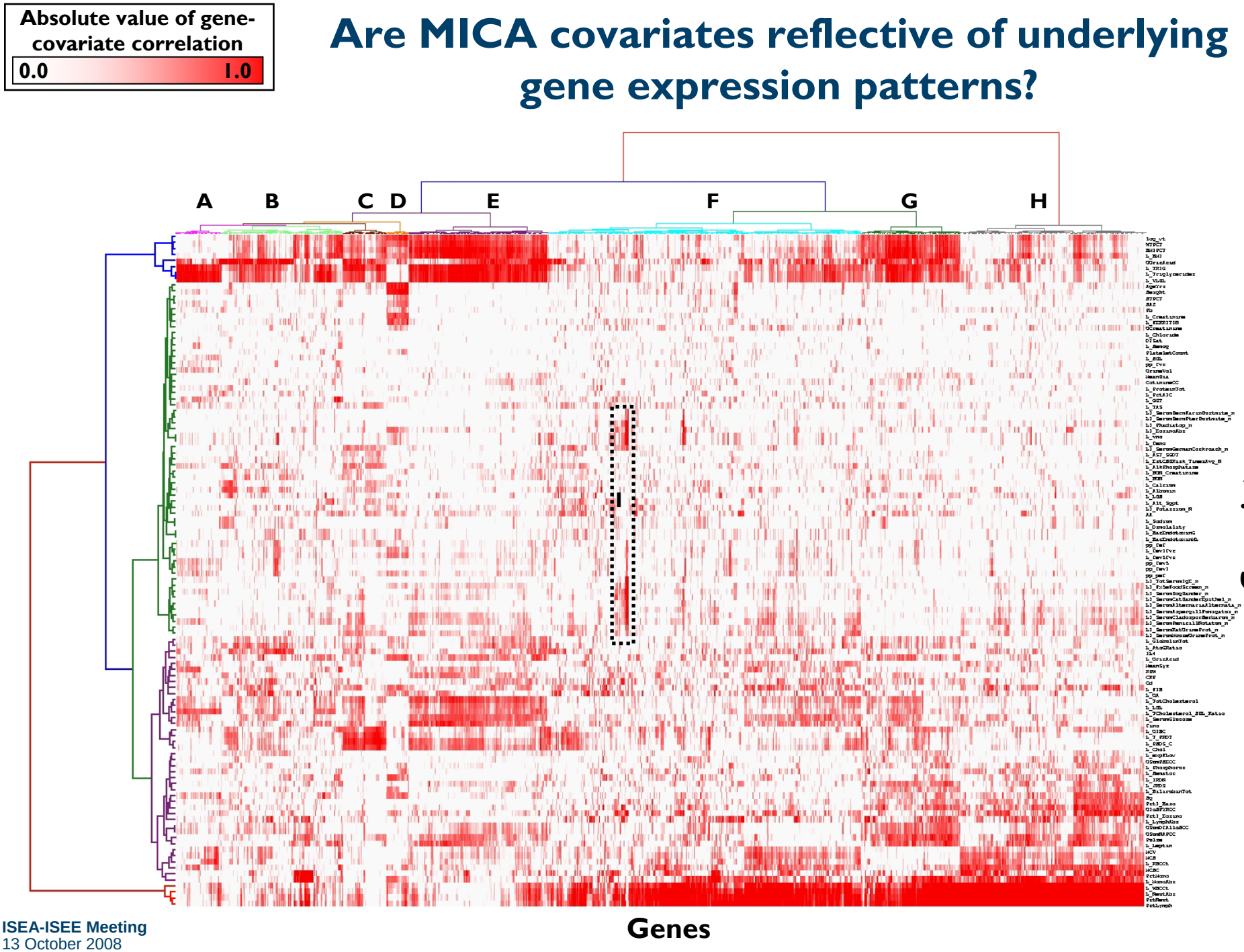


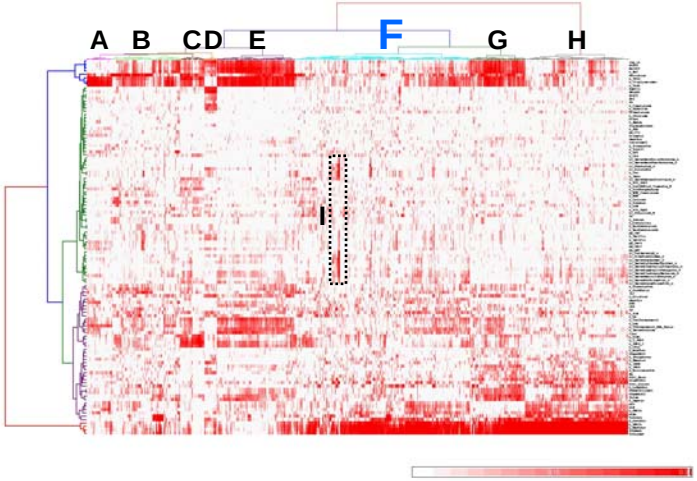
United States Environmental Protection Agency

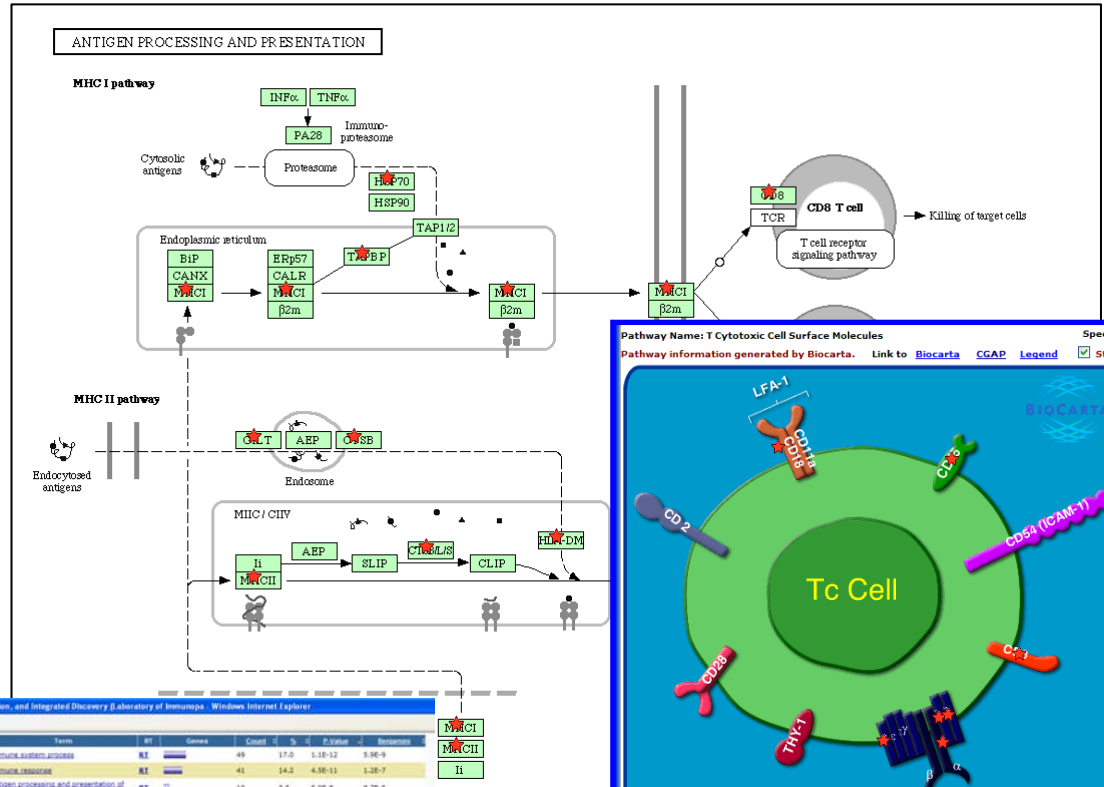
Are MICA covariates reflective of underlying gene expression patterns?

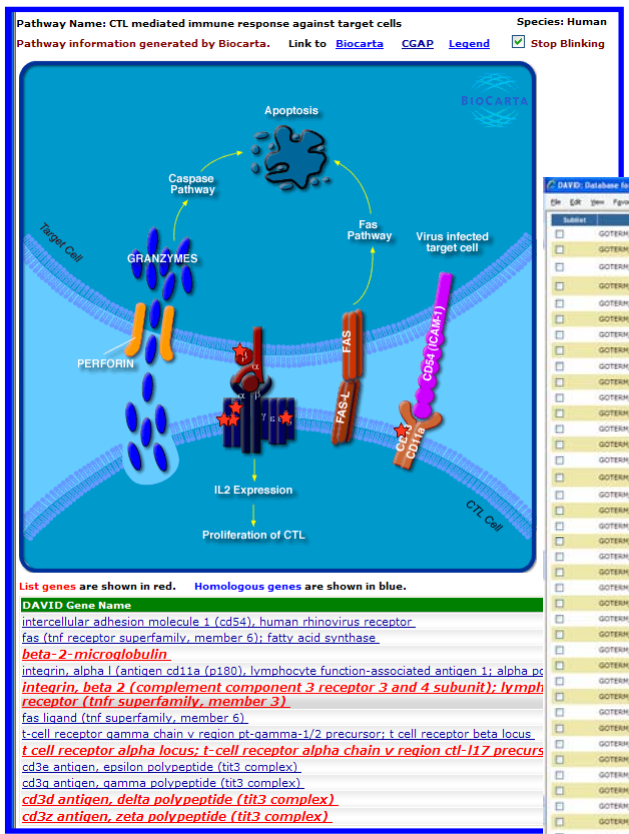


Are MICA covariates reflective of underlying gene expression patterns?

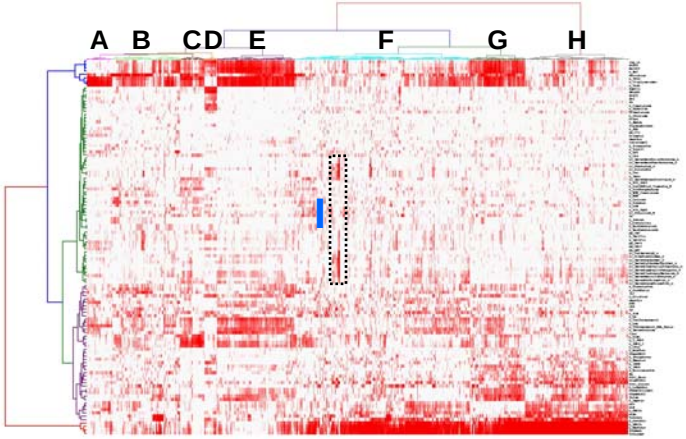


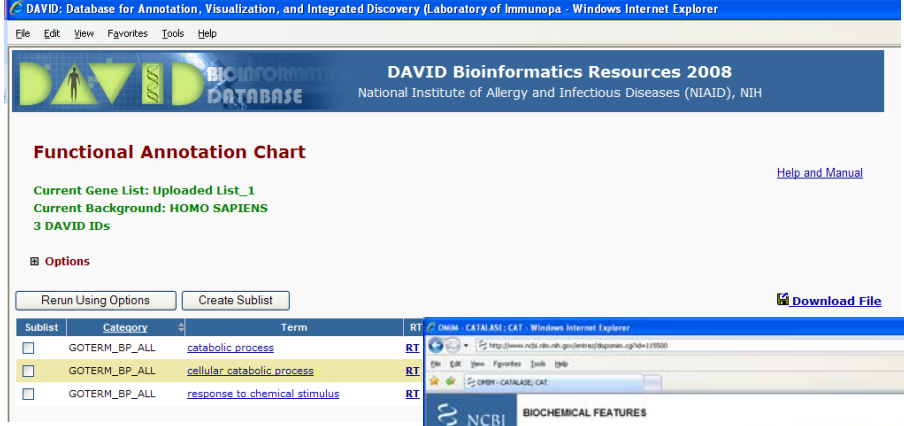


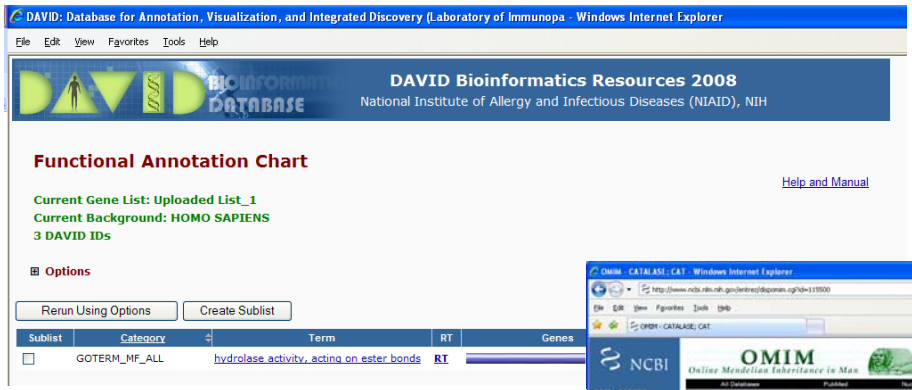


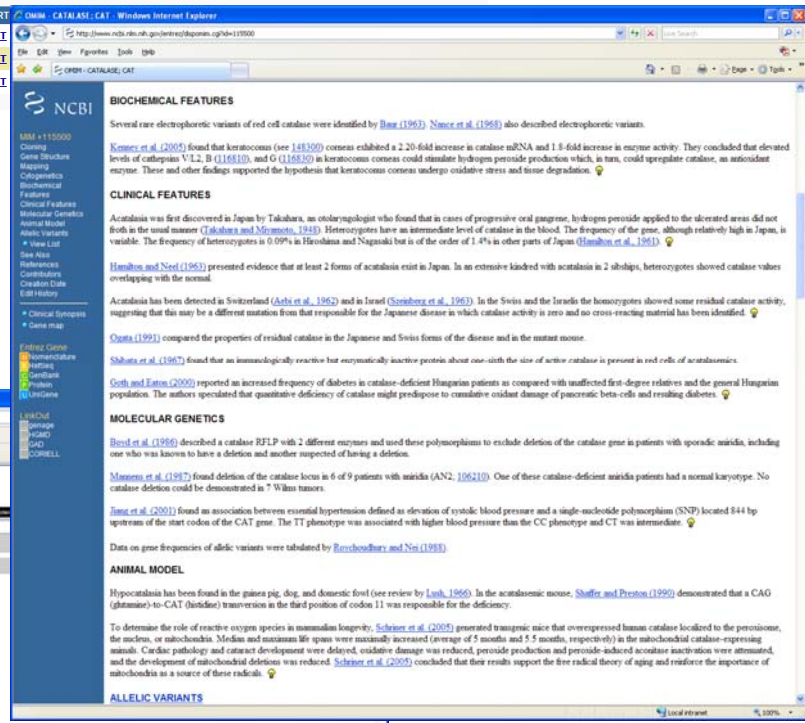


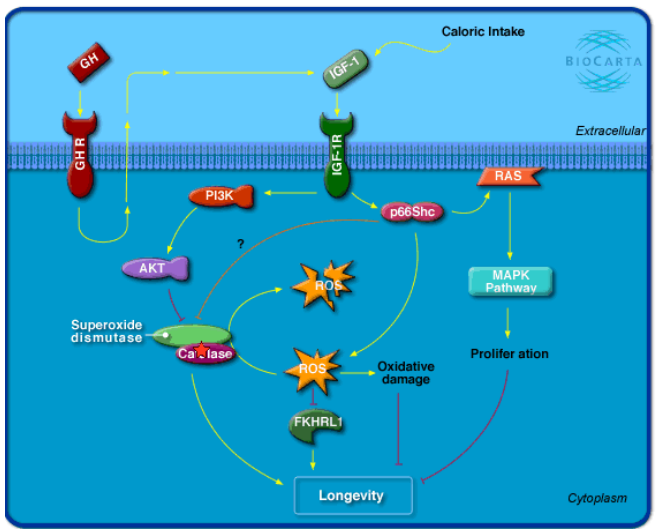
ISEA-ISEE Meeting
13 October 2008

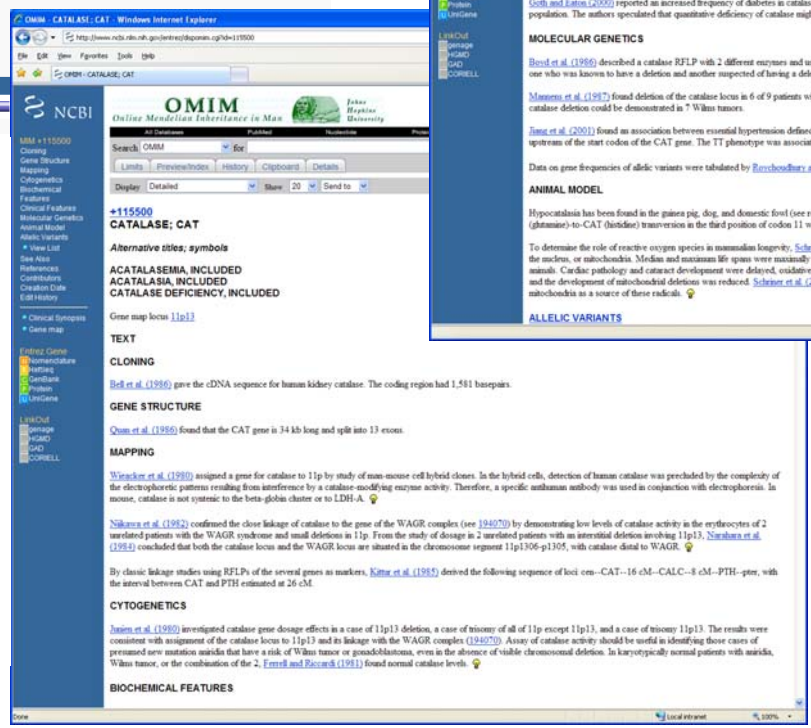

A B CD E F G H

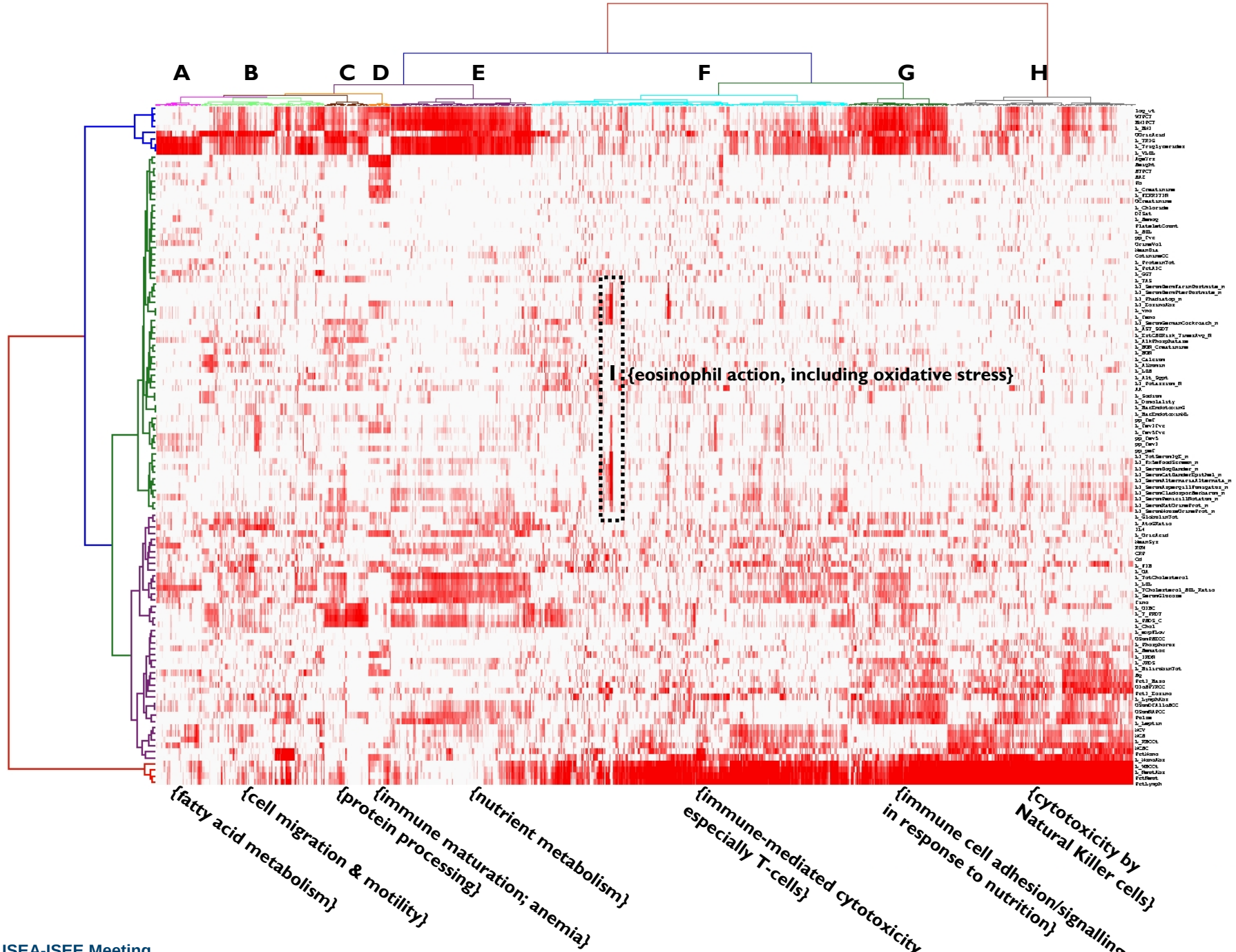

DAVID: Database for Annotation, Visualization, and Integrated Discovery (Laboratory of Immunopa - Windows Internet Explorer
File Edit View Favorites Tools Help
DAVID Bioinformatics Resources 2008
National Institute of Allergy and Infectious Diseases (NIAID), NIH
Functional Annotation Chart
Current Gene List: Uploaded List_1
Current Background: HOMO SAPIENS
3 DAVID IDs
Options
Help and Manual
Rerun Using Options
Create Sublist
Download File
Sublist
Category
Term
RT
GOTERM_BP_ALL
catabolic process
cellular catabolic process
response to chemical stimulus


DAVID: Database for Annotation, Visualization, and Integrated Discovery (Laboratory of Immunopa - Windows Internet Explorer
File Edit View Favorites Tools Help
DAVID Bioinformatics Resources 2008
National Institute of Allergy and Infectious Diseases (NIAID), NIH
Functional Annotation Chart
Current Gene List: Uploaded List_1
Current Background: HOMO SAPIENS
3 DAVID IDs
Options
Help and Manual
Rerun Using Options
Create Sublist
Sublist
Category
Term
RT
Genes
GOTERM_MF_ALL
hydrolase activity, acting on ester bonds

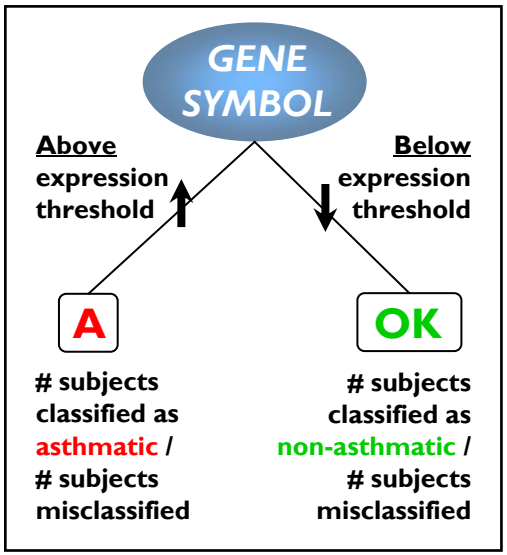

OMIM - CATALASE; CAT - Windows Internet Explorer
NCBI
BIOCHEMICAL FEATURES
CLINICAL FEATURES
MOLECULAR GENETICS
ANIMAL MODEL
ALLELIC VARIANTS

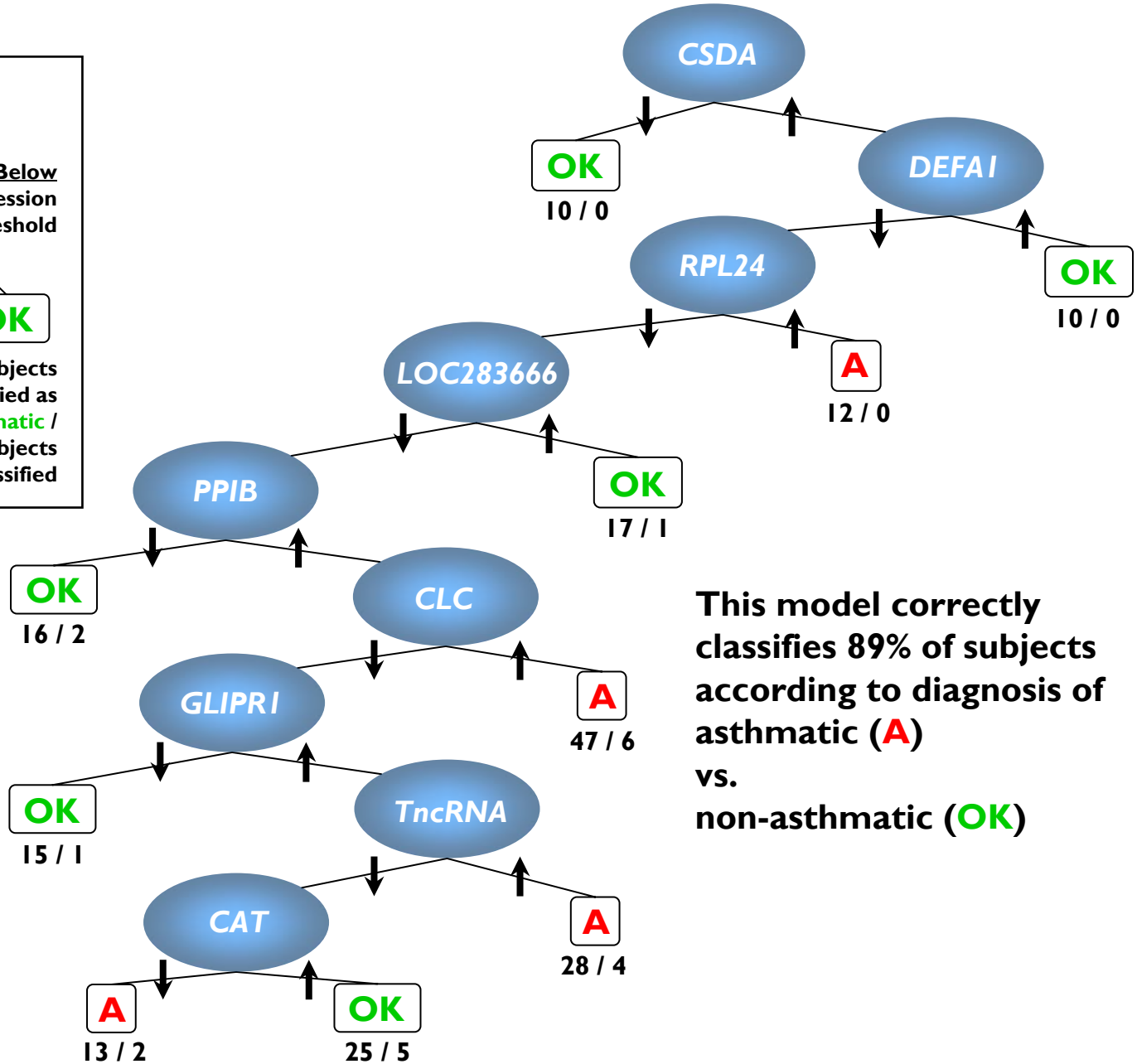

Caloric Intake
GH
IGF-1
BIOCARTA
Extracellular
GH R
IGF-1R
PI3K
RAS
p66Shc
?
AKT
ROS
MAPK Pathway
Superoxide dismutase
Catalase
ROS
Oxidative damage
Prolifer ation
FKHRL1
Longevity
Cytoplasm


OMIM - CATALASE; CAT - Windows Internet Explorer
NCBI
OMIM
Online Mendelian Inheritance in Man
+115500
CATALASE; CAT
Alternative titles; symbols
ACATALASEMIA, INCLUDED
ACATALASIA, INCLUDED
CATALASE DEFICIENCY, INCLUDED
TEXT
CLONING
GENE STRUCTURE
MAPPING
CYTOGENETICS
BIOCHEMICAL FEATURES


A
B
C
D
E
F
G
H
I {eosinophil action, including oxidative stress}
{fatty acid metabolism}
{cell migration & motility}
{protein processing}
{immune maturation; anemia}
{nutrient metabolism}
{immune-mediated cytotoxicity, especially T-cells}
{immune cell adhesion/signalling in response to nutrition}
{cytotoxicity by Natural Killer cells}

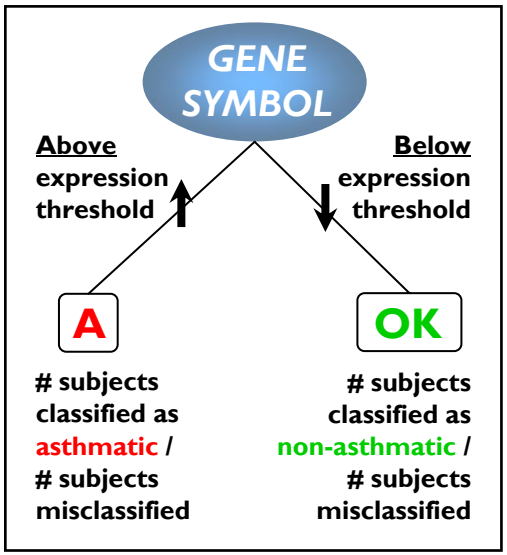
Can we discriminate between subtypes of asthma?

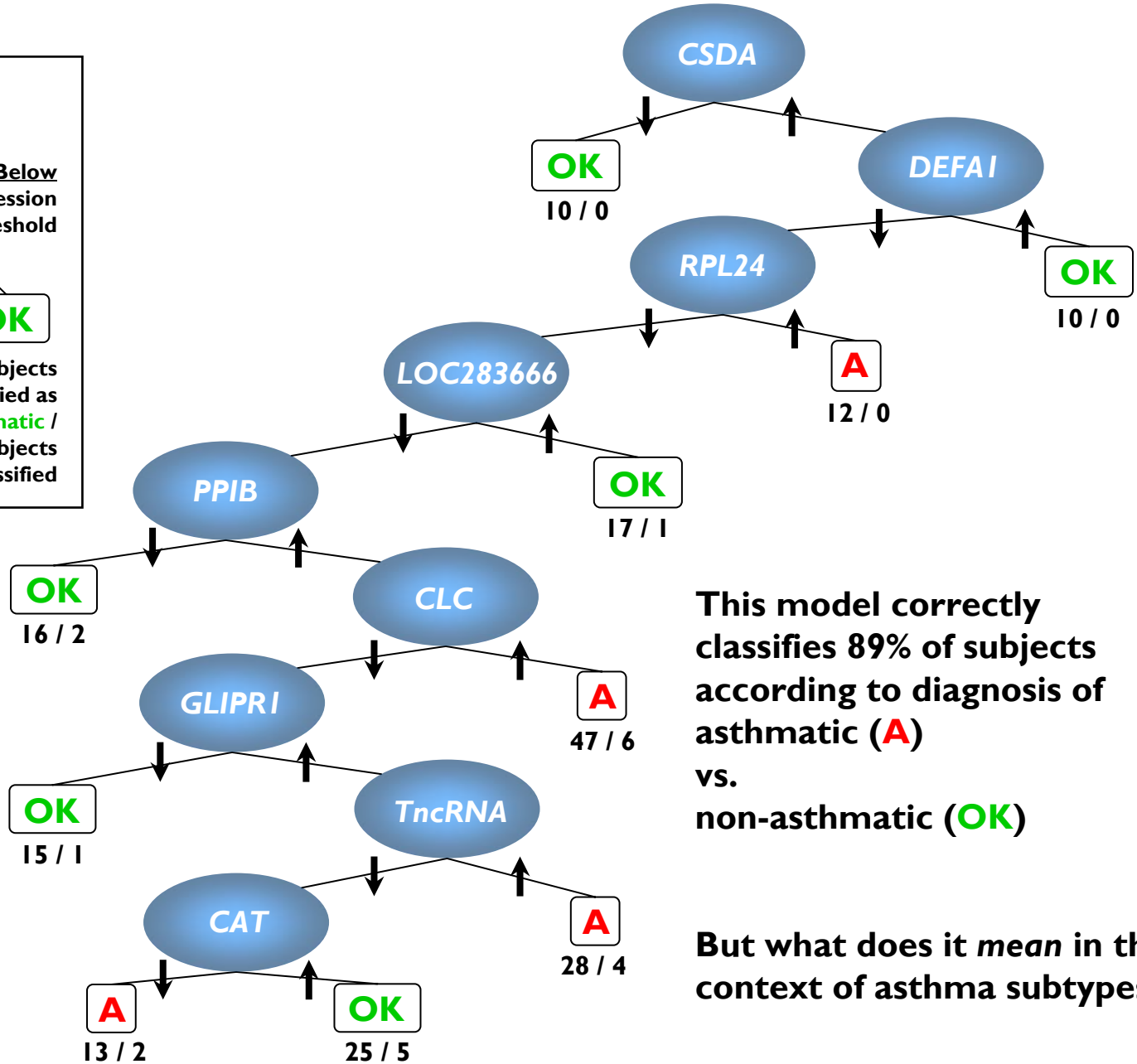




This model correctly classifies 89% of subjects according to diagnosis of asthmatic (A) vs. non-asthmatic (OK)

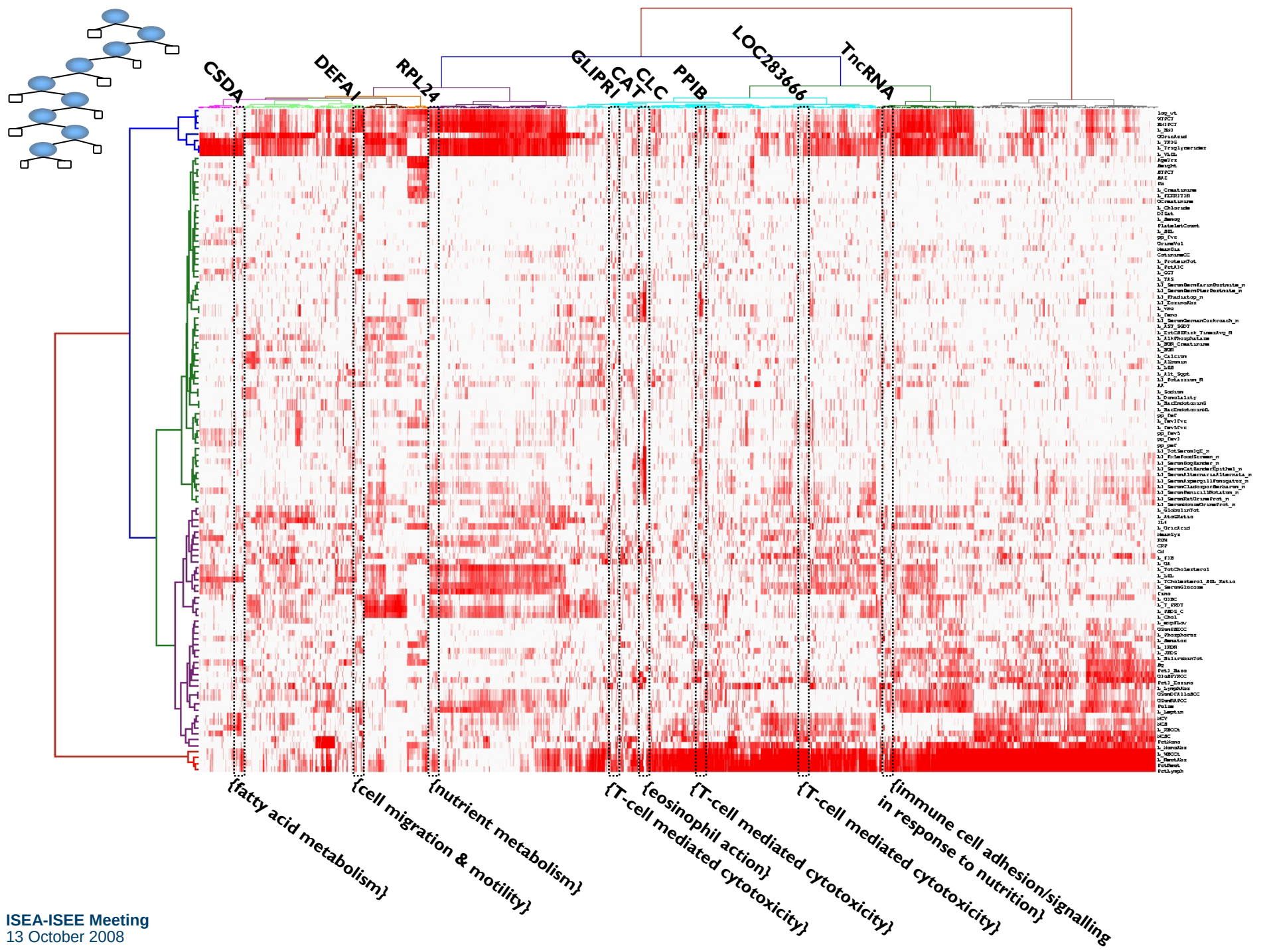
Can we discriminate between subtypes of asthma?





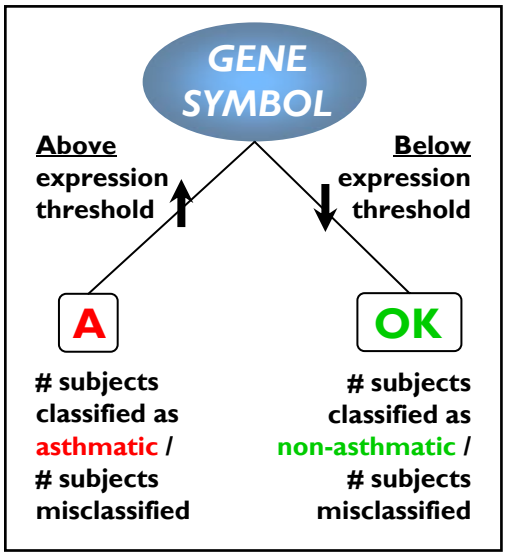
This model correctly classifies 89% of subjects according to diagnosis of asthmatic (A) vs. non-asthmatic (OK)

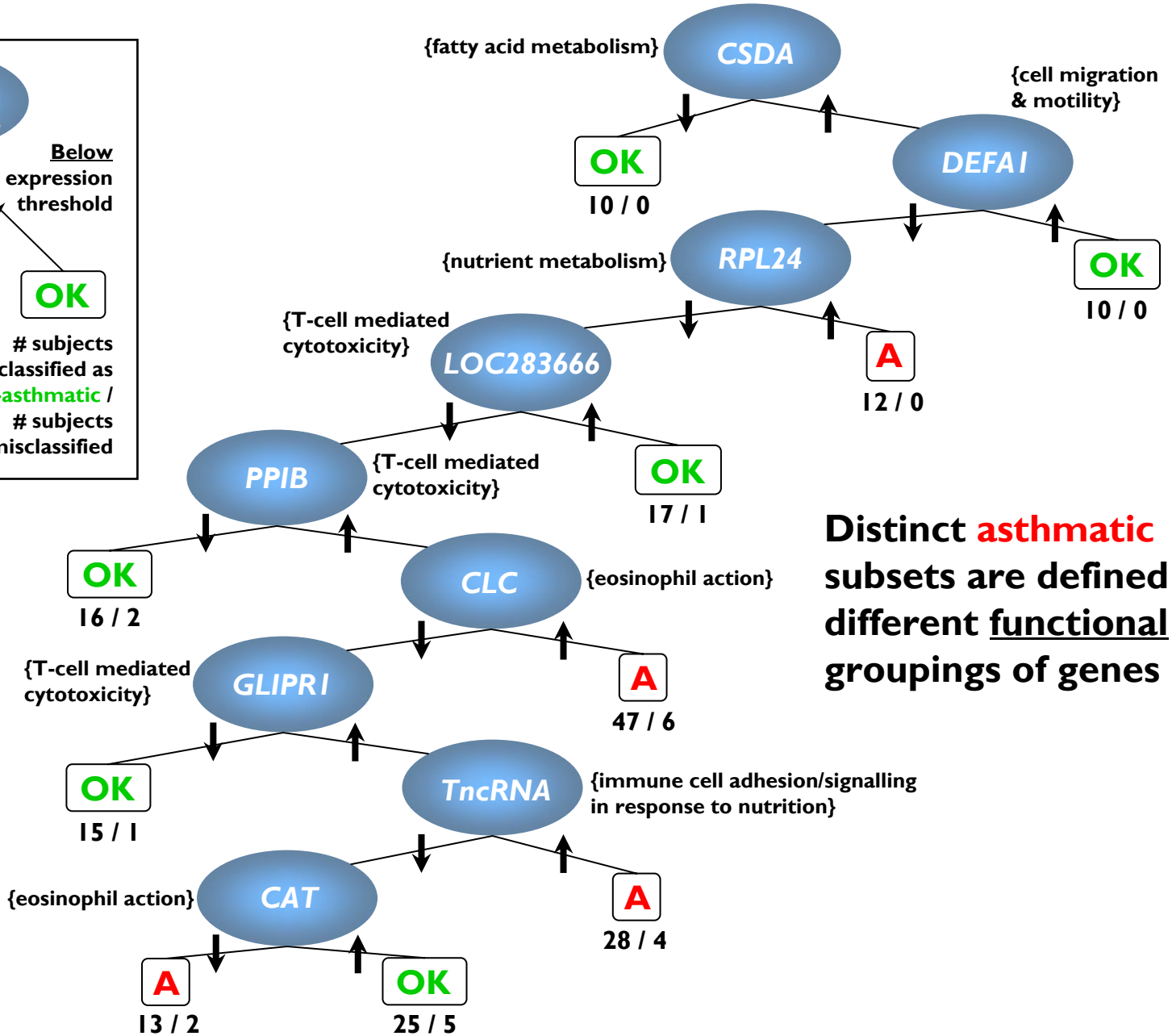
But what does it *mean* in the context of asthma subtypes?


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}

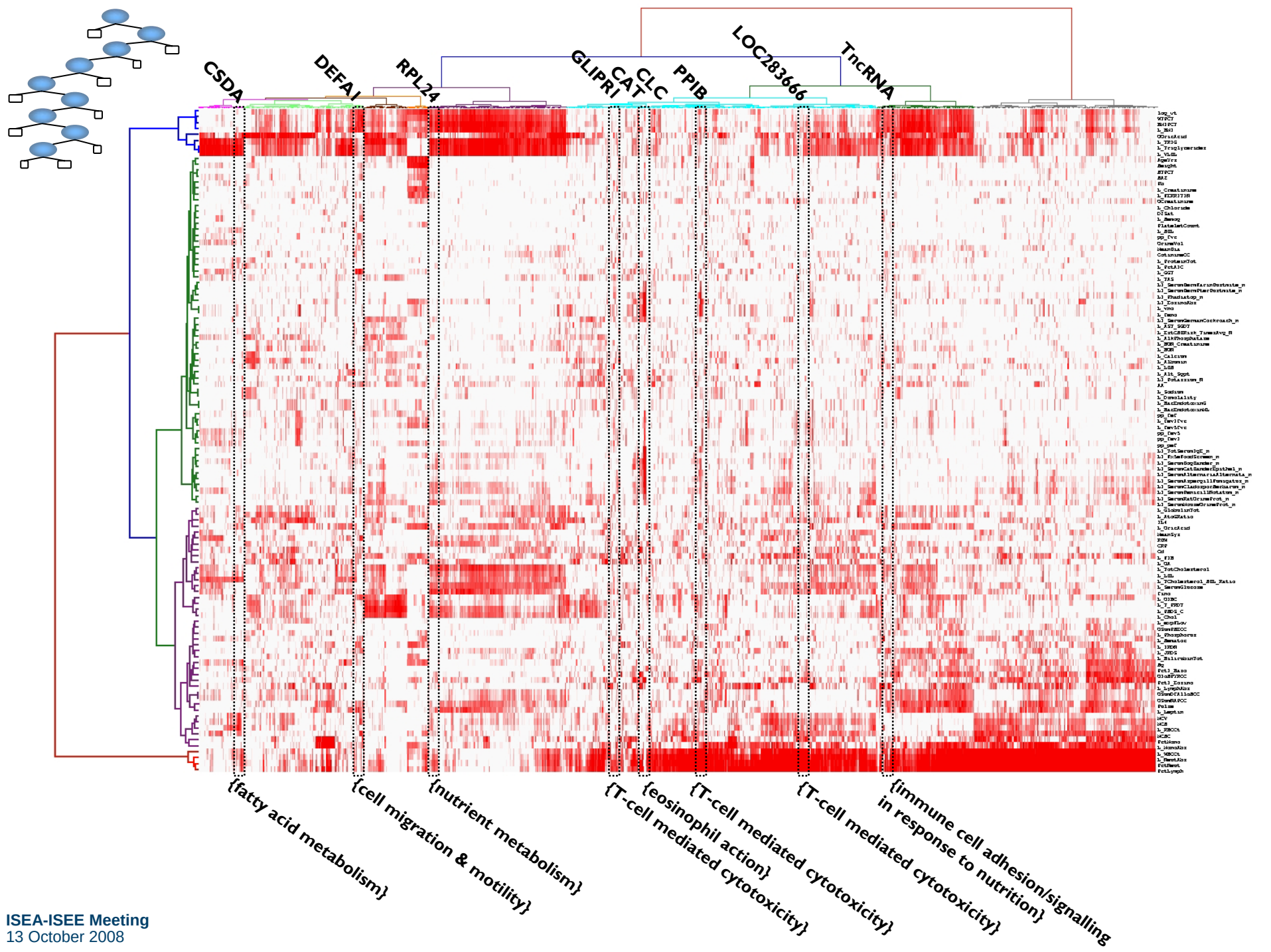
Leveraging covariate information to put gene expression asthma classifier results in context

United States Environmental Protection Agency



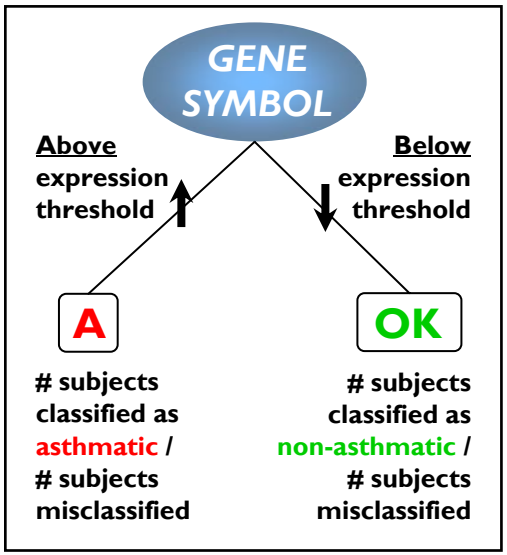


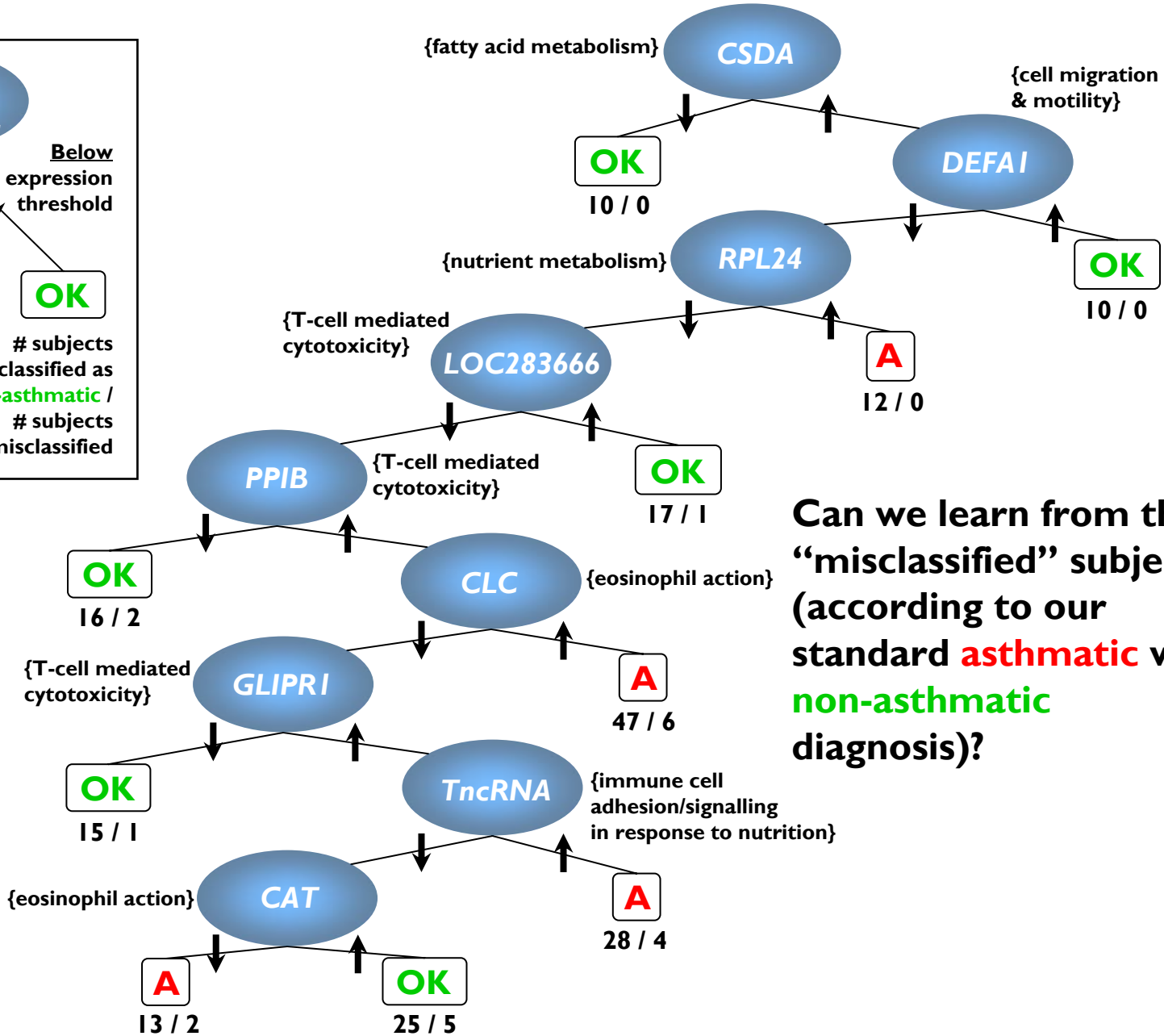
Distinct asthmatic subsets are defined by different functional groupings of genes


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}

Leveraging covariate information to put gene expression asthma classifier results in context

United States Environmental Protection Agency

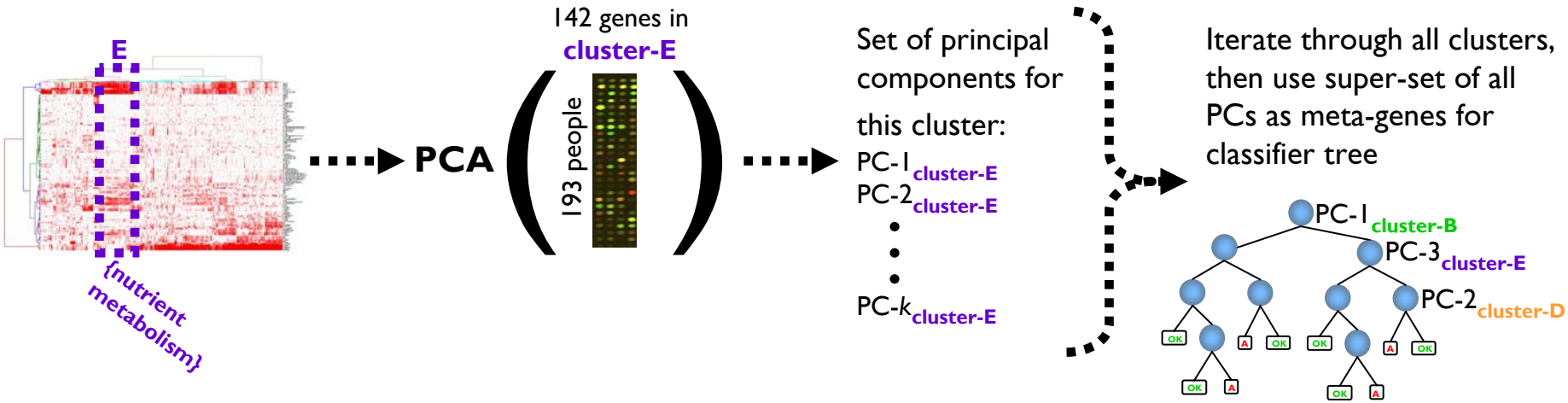


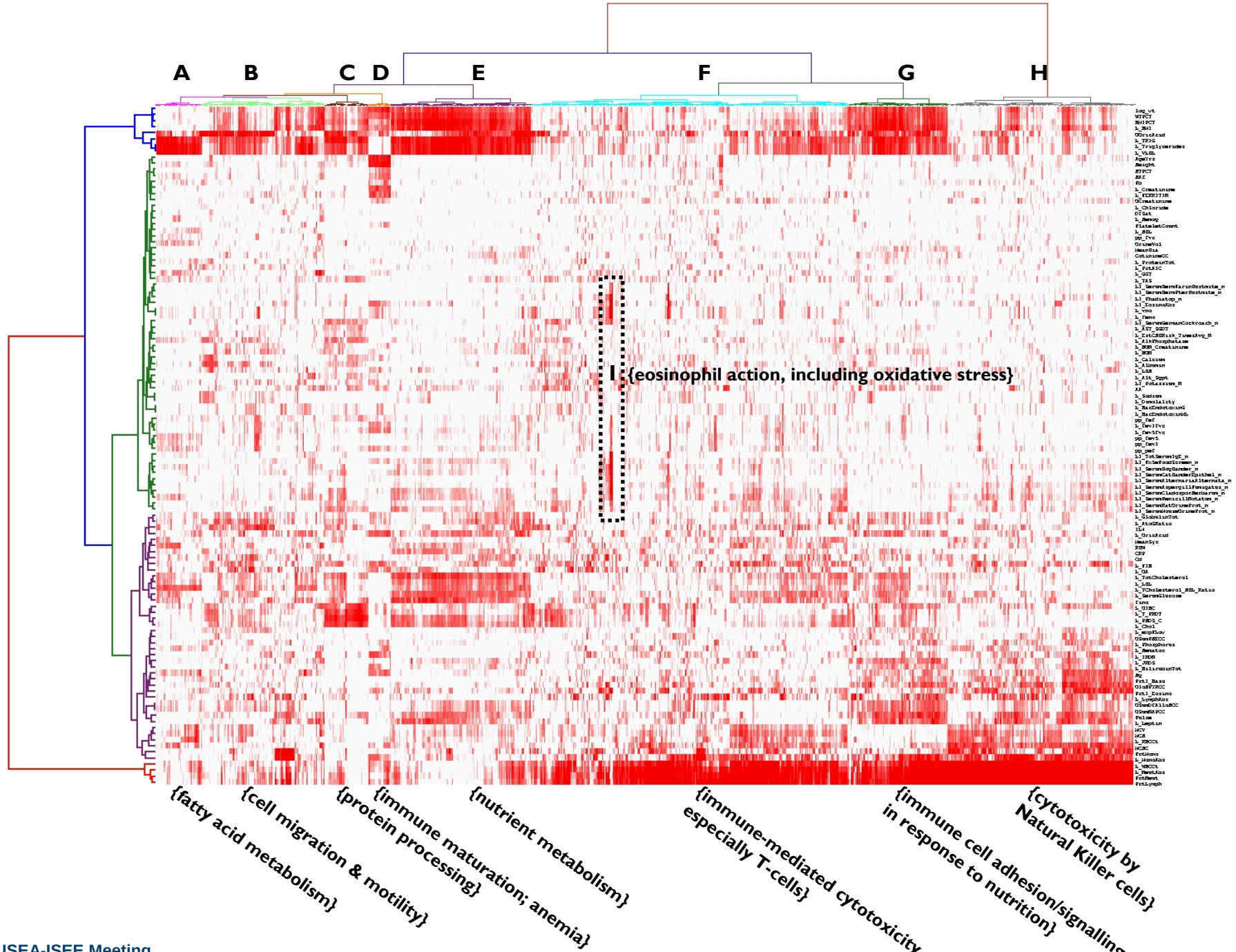


Can we learn from the "misclassified" subjects (according to our standard asthmatic vs. non-asthmatic diagnosis)?

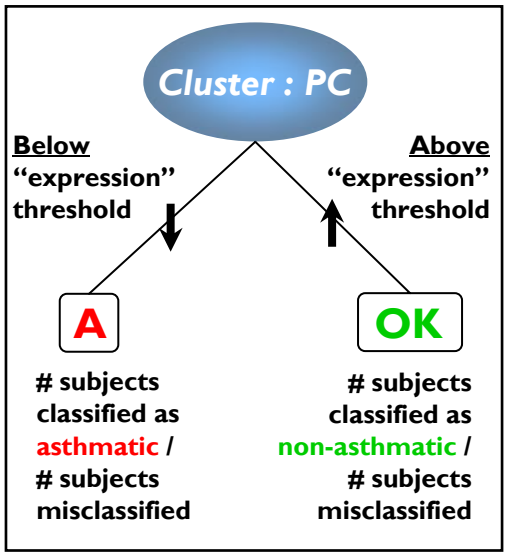
Leveraging covariate information to put gene expression asthma classifier results in context

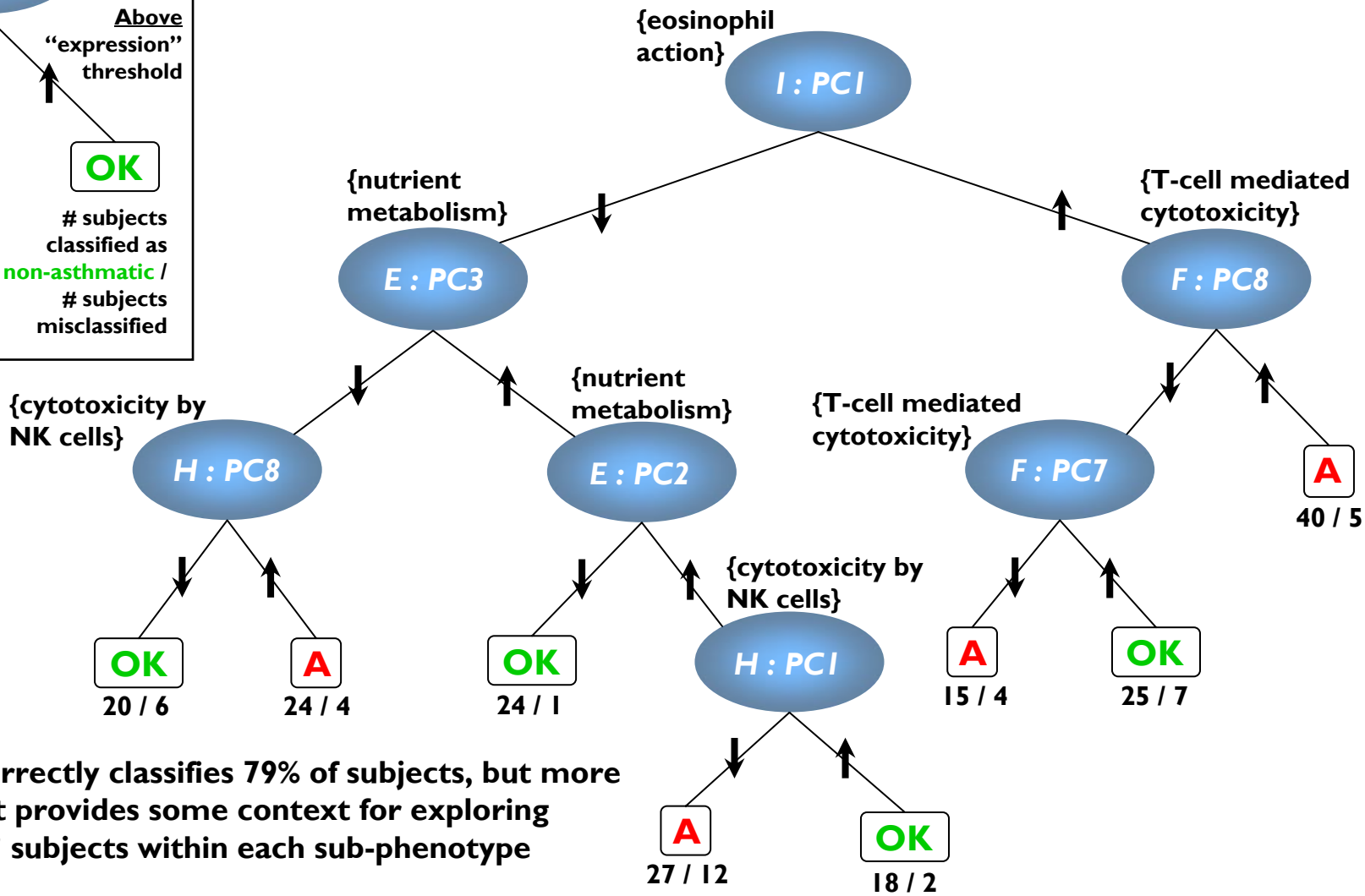
Deriving meta-genes to summarize information within covariate clusters:




A
B
C
D
E
F
G
H
I {eosinophil action, including oxidative stress}
{fatty acid metabolism}
{cell migration & motility}
{protein processing}
{immune maturation; anemia}
{nutrient metabolism}
{immune-mediated cytotoxicity, especially T-cells}
{immune cell adhesion/signalling in response to nutrition}
{cytotoxicity by Natural Killer cells}

Can we discriminate subtypes of asthma by summarizing information from gene clusters?





This model correctly classifies 79% of subjects, but more importantly, it provides some context for exploring "misclassifed" subjects within each sub-phenotype

Can we use the MICA covariates {clinical, exposure, demographic, etc.} to put gene expression signatures in a meaningful context?

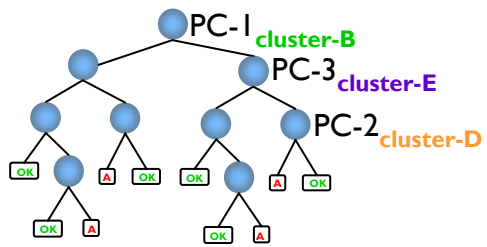
Advantages of this approach:

Avoids reliance on single statistics for single genes
- "Everything depends on everything, but nothing depends on any one thing"
- Covariate clusters are robust to perturbation

Readily extensible to continuous and categorical covariates (see Supplemental Slides)

Links gene expression with other biomarkers or covariates (provides context)
- We gain mechanistic insight from genes that cluster with certain types of covariates
- Provides an "internal" annotation of gene expression

Identifies covariate-associated expression profiles that discriminate asthma sub-phenotypes
- Targets subjects for further scrutiny



Acknowledgments

Jane Gallagher
(U.S. EPA / HSD)

Stephen Edwards
(U.S. EPA / NHEERL)

Lucas Neas
(U.S. EPA / HSD)

Edward Hudgens
(U.S. EPA / HSD)

John Wambaugh
(U.S. EPA / NCCT)

Wendell Jones
(Expression Analysis)

Elaine Cohen Hubal
(U.S. EPA / NCCT)

Brooke Heidenfelder
(Previously at U.S. EPA / HSD;
Currently at Almac Pharamceuticals)

Ann Williams
(U.S. EPA / HSD)

John Rogers
(Westat)

Alison Motsinger-Reif
(North Carolina State University)

DISCLAIMER: The contents of this presentation do not necessarily reflect EPA policy.