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

David Reif, Ph.D.
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\}
Complex diseases involve multiple etiological pathways

Genetic Factor

Environmental Factor

Progression of etiological mechanisms
Asthma etiology

Indoor Triggers

Genetics

Behavior

Air Pollution

[Figure adapted from Jane Gallagher]
Mechanistic Indicators of Childhood Asthma (MICA) explores multiple aspects of asthma etiology.

**Indoor Triggers**
- Serum endotoxins

**Behavior**
- BMI, triglycerides

**Air Pollution**
- PM$_{2.5}$

**Genetics**
- Gene expression

---

**Gene Expression**

**PM$_{2.5}$**
Gene expression measured using oligonucleotide microarrays

54675 genes

195 people (subjects)

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

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

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
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?

- Gene expression (genomics)
- Expert Knowledge (existing data)
- Covariates (environmental, health status, etc.)
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?

195 people (samples)

54675 genes

MICA covariates

195 people (samples)
Partial list of MICA covariates

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

| D5AnyMedUse                   | Any Asthma Med Use                          |
| D6DrDiagnosis                 | Dr. Diagnosed Asthma (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     |
| Pctl_Baso                     | Basophil percent of sum WBC                 |
| Pctl_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

MICA covariates

195 people (samples)
How do we leverage MICA covariate information for the gene expression analysis?

195 people (samples)

54675 genes

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

54,675 genes

195 people (samples)

MICA covariates

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
Are MICA covariates reflective of underlying gene expression patterns?
Are MICA covariates reflective of underlying gene expression patterns?
(fatty acid metabolism)

(cell migration & motility)

(nutrient metabolism)

(immune maturation, anemia)

(protein processing)

(immune-mediated cytotoxicity, especially T-cells)

(immune cell adhesion/signalling in response to nutrition)

(cytotoxicity by Natural Killer cells)

(eosinophil action, including oxidative stress)
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, DE, FA1, RPL24, GLIPRT, CAT, PPIB, LOC283666, TncRNA

{fatty acid metabolism}
{cell migration & motility}
{nutrient metabolism}
{T-cell mediated cytotoxicity}

(immune cell adhesion/signalling in response to nutrition)
(eosinophil action)
(T-cell mediated cytotoxicity)
(T-cell mediated cytotoxicity)
(eosinophil action)
Leveraging covariate information to put gene expression asthma classifier results in context

Distinct asthmatic subsets are defined by different functional groupings of genes.
CSDA
DEFA1
RPL24
GLIPR1
CAT
LOC283666
TncRNA

{fatty acid metabolism}
{cell migration & motility}
{nutrient metabolism}
{T-cell mediated cytotoxicity}

(immune cell adhesion/signalling in response to nutrition)
(eosinophil action)
(T-cell mediated cytotoxicity)
Leveraging covariate information to put gene expression asthma classifier results in context

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:

142 genes in cluster-E

193 people

Set of principal components for this cluster:
PC-1_{cluster-E}
PC-2_{cluster-E}
PC-k_{cluster-E}

Iterate through all clusters, then use super-set of all PCs as meta-genes for classifier tree
(fatty acid metabolism)
(cell migration & motility)
(nutrient metabolism)
(immune maturation, anemia)
(immune-mediated cytotoxicity, especially T-cells)
(immune cell adhesion/signalling in response to nutrition)
(cytotoxicity by Natural Killer cells)
(eosinophil action, including oxidative stress)
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 “misclassified” 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.