Integrative Analysis of Exposome and Transcriptome for Asthma

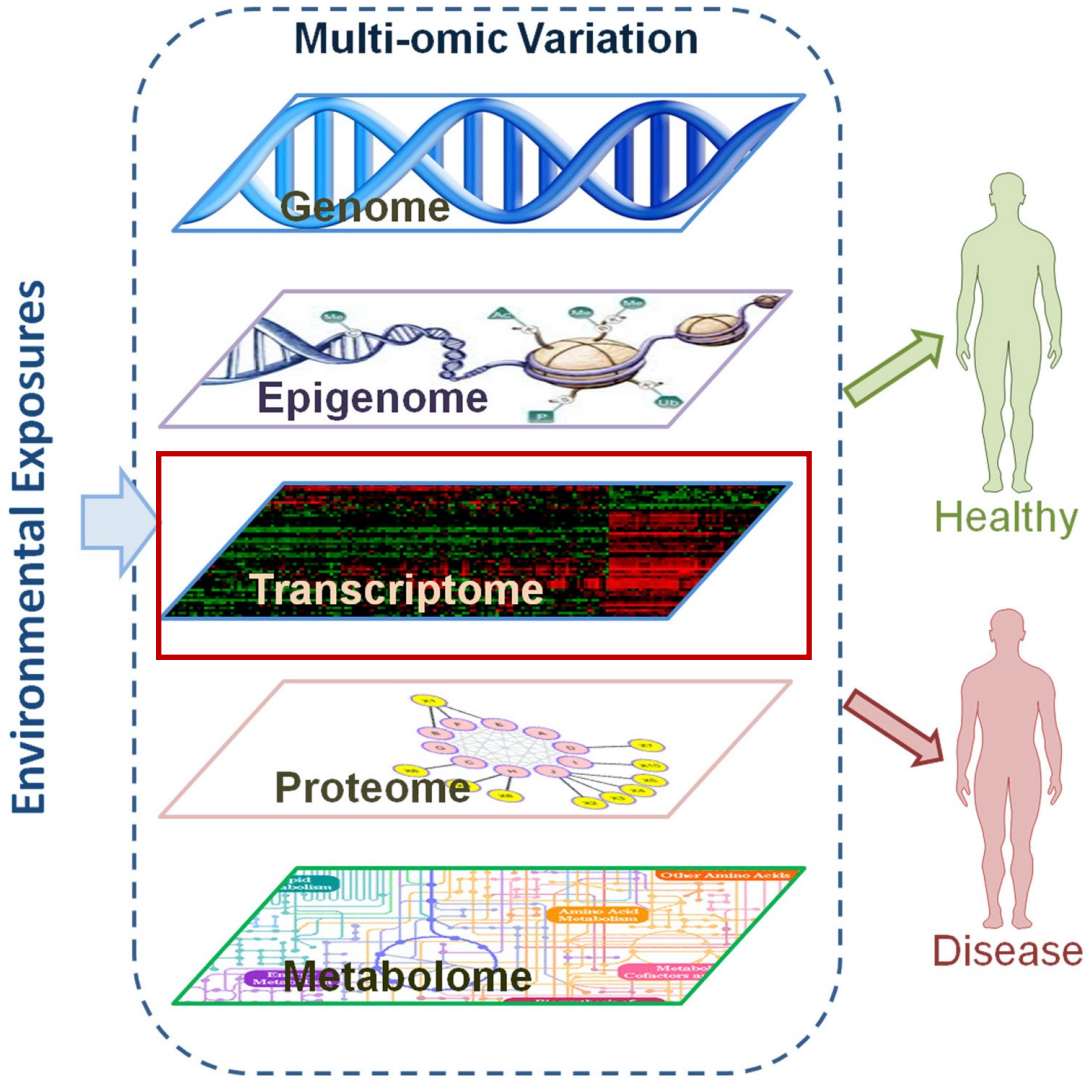
Ziyue Wang, Farida Akhtari, Dillon Lloyd

Biostatistics & Computational Biology Branch, National Institute of Environmental Health Science

2021.4.28

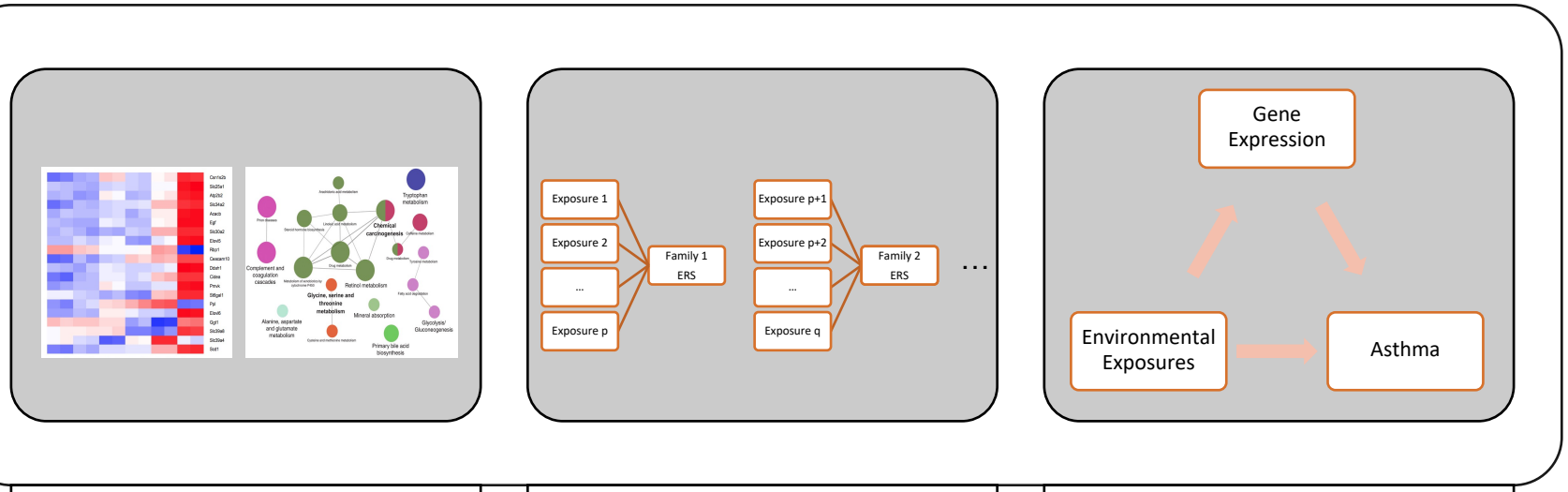
Introduction

- The development of human disease involve a highly dynamic and interactive system of molecular layers and are influenced by many environmental factors.
- Asthma remains an important public health problem worldwide because of high morbidity and inadequate disease control.
- Assessing the effect of various exposures on gene expression would be useful to discover potential cellular mechanisms through which exposures may influence the development of Asthma in human populations.
- Environmental exposures affected gene expressions associated with Asthma much more than genetic ancestry (Favé, M. J *et al.* (2018)).



Sun, Y. V., & Hu, Y. J. (2016)

Analytical Framework



Aim:

Discovering gene expression profiles and their functional roles associated with Asthma.

Methods:

Differential expressed gene analysis

Gene set enrichment analysis

Aim:

Dimension reduction of Asthma-associated exposures.

Methods:

Variable selection by elastic net regularization.

Calculating environmental risk scores (ERS) for all selected exposures and each exposure family.

Aim:

Assessing causality of environmental exposures on Asthma by perturbing gene expression levels.

Methods:

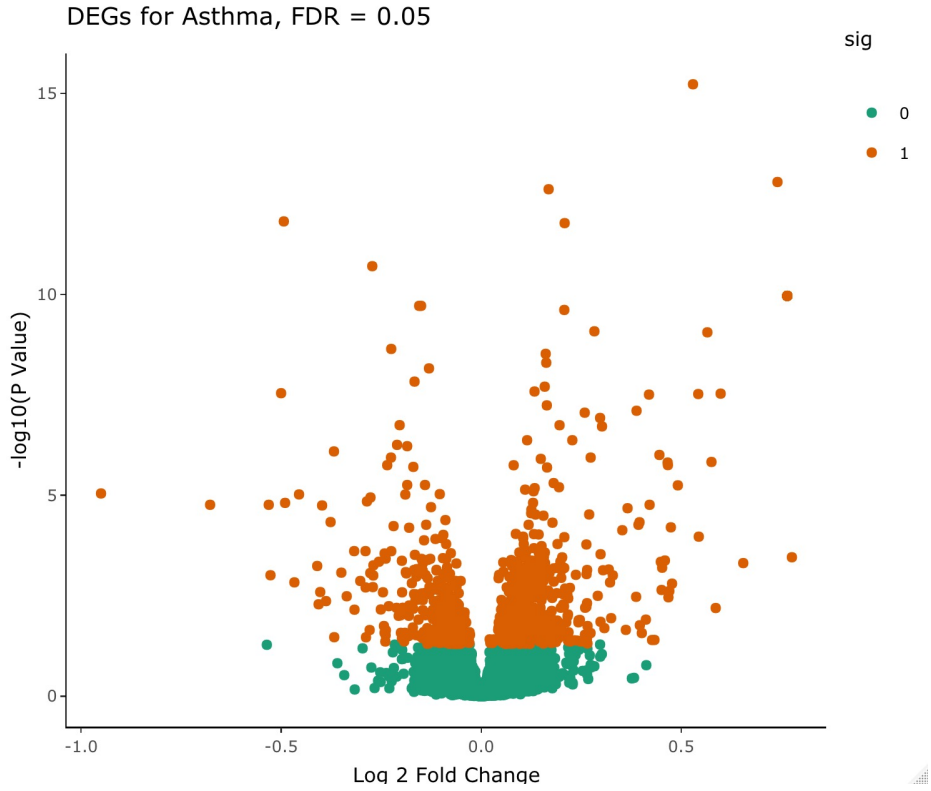
High-dimensional mediation analysis for evaluation all genes simultaneously.

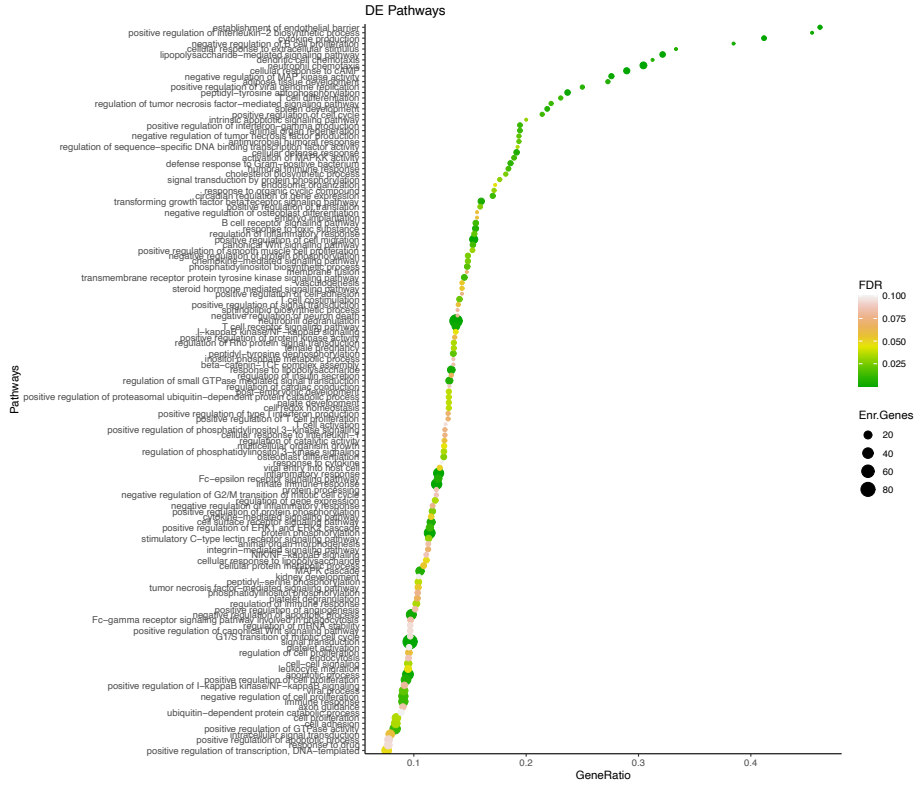
a). for single exposure

b). for ERS in each family

Gene Expression Profiles and Functionality

- Summarize to gene level when multiple transcripts map to the same gene
- 863 DE genes associated with Asthma adjusting covariates (sex, age and ethnicity)
- Gene Set Enrichment Analysis
  - REACTOME pathway; Gene Ontology





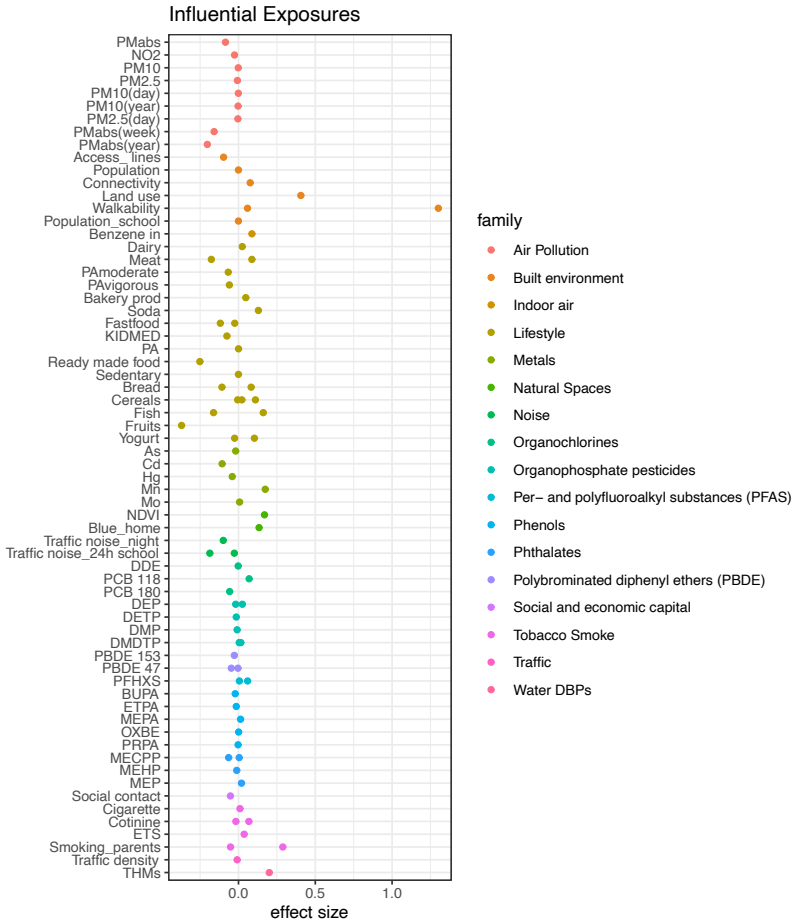
Variable Selection

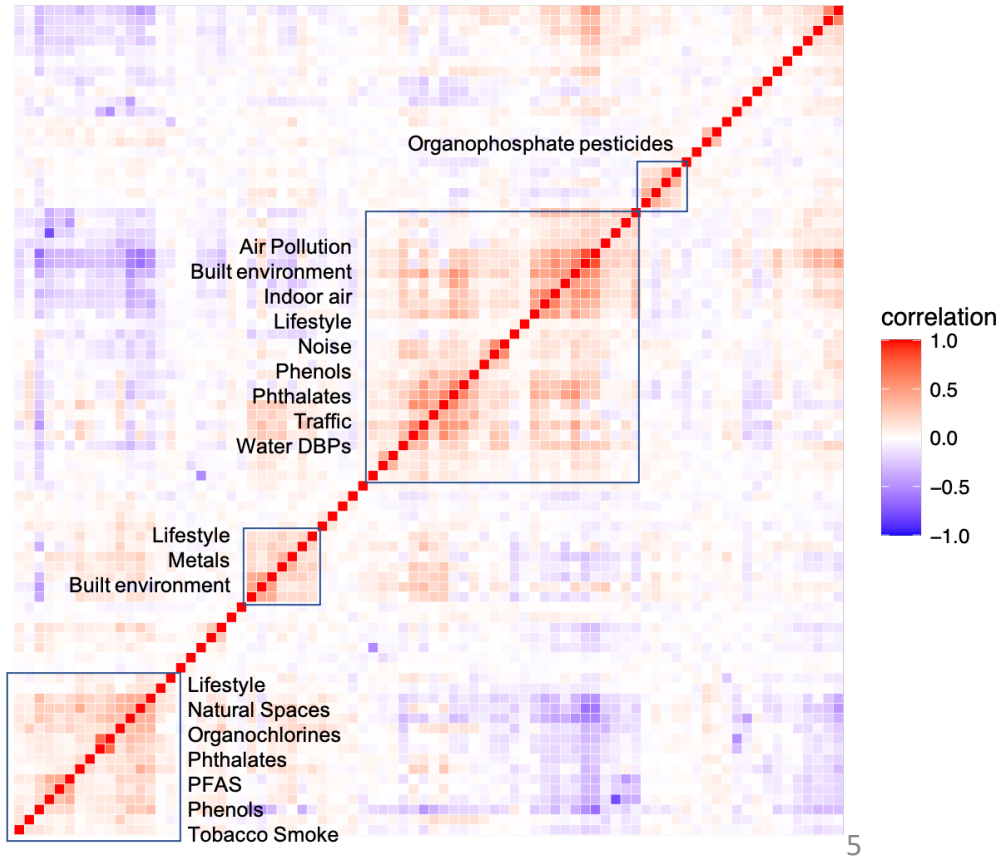
- Elastic Net
  - Combination of lasso and ridge, deal with correlated intra-family and inter-family exposures
  - Covariates: sex, age, cohort, birth weight, BMI of child and ethnicity
  - 76 exposures in 17 families

Asthma	N	0, N = 905[1]	1, N = 102[1]	p-value[2]
Sex	1,007			0.022
female		436 (48%)	37 (36%)	
male		469 (52%)	65 (64%)	
Age	1,007	7.88 (1.48)	7.95 (1.67)	0.9
Cohort	1,007			<0.001
1		133 (15%)	26 (25%)	
2		89 (9.8%)	20 (20%)	
3		177 (20%)	8 (7.8%)	
4		143 (16%)	11 (11%)	
5		220 (24%)	25 (25%)	
6		143 (16%)	12 (12%)	
Birth Weight	1,007	3,386 (510)	3,393 (589)	0.8
BMI of Child	1,007	0.41 (1.15)	0.38 (1.22)	0.7
Ethnicity	1,007			0.11
no		87 (9.6%)	15 (15%)	
yes		818 (90%)	87 (85%)	

[1] n (%); Mean (SD)

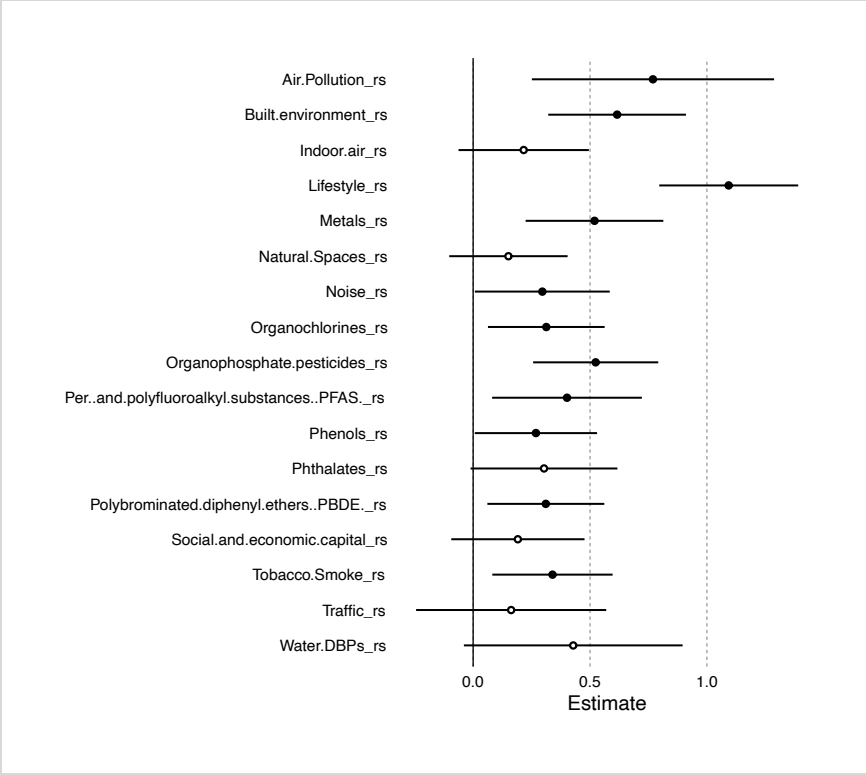
[2] Pearson's Chi-squared test; Wilcoxon rank sum test

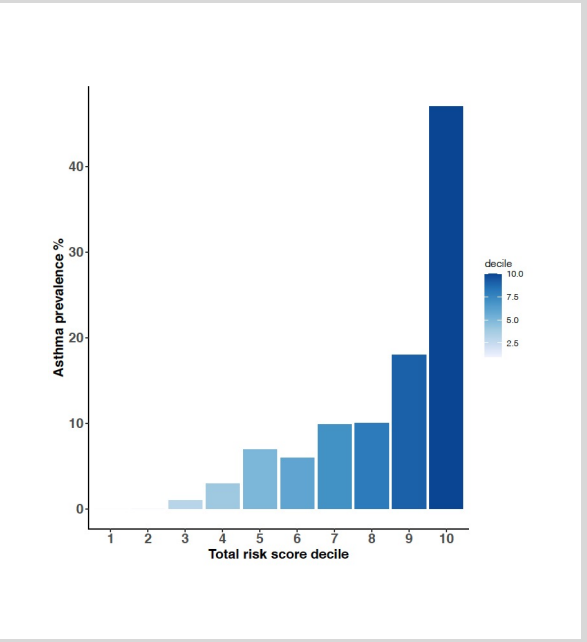




Variable Selection

- Define ERS for all exposures (total risk score) and for each family (family-based risk score)
  - Most family-based risk score are significantly associated with Asthma
  - Total risk score is a well-defined index for predicting Asthma status
- Reduce the collinearity across intra-family exposures
  - Absolute correlation ranging from 1e-5 to 0.57



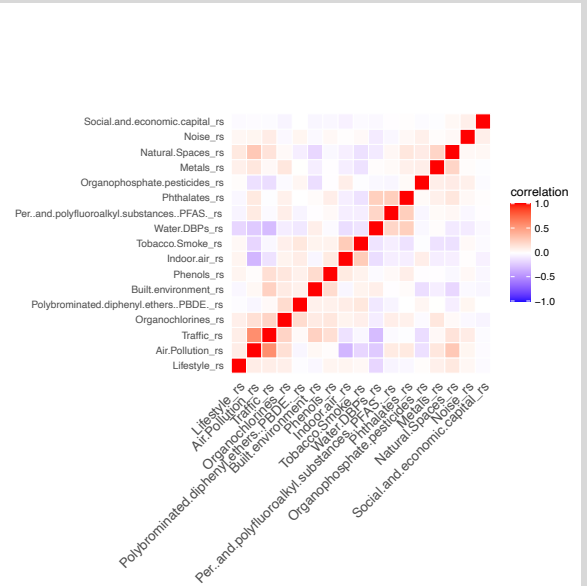


Environmental risk score (ERS)

For sample i and exposure family j,
suppose there are k_j exposures in each family,

$$\text{ERS}_{ij} = \hat{\beta}_j X_{ij},$$

where $\hat{\beta}_j = (\beta_1, \ldots, \beta_{k_j})$ is the vector of coefficient for k_j exposures estimated from elastic net.



Mediation Analysis

- Build a mediation model for individual exposure variable
- Model all mediators simultaneously
- 33 exposures reveal significant mediation signals

Model (1)

$$Y = c^* + \gamma^* X + \epsilon_1,$$
$$M_k = c_k + \alpha_k X + e_k, \quad k = 1, \cdots, p$$
$$Y = c + \gamma X + \beta_1 M_1 + \cdots + \beta_p M_p + \epsilon_2,$$

where X is exposure, M_k are potential mediators and Y is outcome;

γ^* represents the total effect of X on Y;

γ represents the direct effect of X on Y after adjusting mediators of interest;

$(\alpha_1\beta_1, \ldots, \alpha_p\beta_p)^T$ represents the mediation effect by the path $X \to M \to Y$.

Family	Exposure	Mediator Gene(s)
Air Pollution	h_abs_ratio_preg_Log	**NEO1**
	h_pm10_ratio_preg_None	SNORD15A
	hs_pm10_yr_hs_h_None	LRRC28
	hs_pm25abs_wk_hs_h_Log	**CES1**
	hs_pm25abs_yr_hs_h_Log	VPS41
Built environment	h_accesslines300_preg_dic0	TOMM40
	h_walkability_mean_preg_None	**NEO1**
Lifestyle	hs_fastfood_Ter	CES1, NDUFA6
	hs_mvpa_prd_alt_None	TLR10
	hs_total_bread_Ter	LRRC28
	hs_total_fish_Ter	SNORD15A
	hs_total_fruits_Ter	TMA7
	hs_total_yog_Ter	TMA7
Metals	hs_as_c_Log2	**CES1**
	hs_hg_m_Log2	TLR10
	hs_mn_c_Log2	TMA7
	hs_mo_c_Log2	VPS41
Natural Spaces	h_ndvi100_preg_None	**NEO1**
	hs_blueyn300_h_None	TREM1
Noise	hs_lden_cat_s_None	TLR10
Organophosphate pesticides	hs_detp_cadj_Log2	ZNF786
	hs_dmp_cadj_Log2	**CES1**
Polybrominated diphenyl ethers (PBDE)	hs_pbde47_cadj_Log2	PRDX2
	hs_pbde47_madj_Log2	SNORA10
Per- and polyfluoroalkyl substances (PFAS)	hs_pfhxs_c_Log2	VPS41
	hs_pfhxs_m_Log2	VPS41
Phenols	hs_bupa_cadj_Log2	CYB5B
	hs_etpa_madj_Log2	BIRC3
	hs_mepa_cadj_Log2	ZCCHC14
Phthalates	hs_mecpp_madj_Log2	PRDX2
Social and economic capital	hs_contactfam_3cat_num_None	**NEO1**
Traffic	h_trafnear_preg_pow1over3	**CES1**
Water DBPs	h_thm_preg_Log	SNORA10

* Significant mediated effect with FDR < 0.1

Mediation Analysis

- Build a mediation model for each family-based ERS
- Model all mediators simultaneously
- 7 exposure families reveal significant mediation signals

Model (2)

$$Y = c^* + \gamma^* X + \epsilon_1,$$
$$M_k = c_k + \alpha_k X + e_k, \quad k = 1, \cdots, p$$
$$Y = c + \gamma X + \beta_1 M_1 + \cdots + \beta_p M_p + \epsilon_2,$$

where X is ERS for an exposure family,
M_k are potential mediators and Y is outcome;
γ^* represents the total effect of X on Y;
γ represents the direct effect of X on Y after adjusting mediators of interest;
$(\alpha_1\beta_1, \ldots, \alpha_p\beta_p)^T$ represents the mediation effect by the path $X \to M \to Y$.

Family-based ERS	Mediator Gene(s)
Air Pollution	**CES1**
Organophosphate pesticides	ZNF786
Polybrominated diphenyl ethers (PBDE)	SNORA10
Per- and polyfluoroalkyl substances (PFAS)	VPS41
Social and economic capital	**NEO1**
Traffic	**CES1**
Water DBPs	SNORA10

* Significant mediated effect with FDR < 0.1


THE LANCET
Volume 383, Issue 9928, 3–9 May 2014, Pages 1581-1592

Series

Outdoor air pollution and asthma

Michael Guarnieri MD [a, b], Dr John R Balmes MD [a, b]


Natural Products as Modulators of CES1 Activity

Yuli Qian and John S. Markowitz

Drug Metabolism and Disposition October 2020, 48 (10) 993-1007; DOI: https://doi.org/10.1124/dmd.120.000065


Vol. 114, No. 5 | Research

Traffic, Susceptibility, and Childhood Asthma

is companion of

Rob McConnell, Kiros Berhane, Ling Yao, Michael Jerrett, Fred Lurmann, Frank Gilliland, Nino Künzli, Jim Gauderman, Ed Avol, Duncan Thomas, and John Peters

Published: 1 May 2006 | https://doi.org/10.1289/ehp.8594 | Cited by: 322


Journal of Allergy and Clinical Immunology

Volume 144, Issue 1, July 2019, Pages 94-108.e11

Asthma and lower airway disease

Blockade of RGMb inhibits allergen-induced airways disease

Sanhong Yu MD, PhD [a, b], Krystle M. Leung BA [b], Hye-Young Kim PhD [b, c], Sarah E. Umetsu MD, PhD [f], Yanping Xiao PhD [d], Lee A. Albacker PhD [b, e, i], Hyun-Jun Lee PhD [b, j], Dale T. Umetsu MD, PhD [b, g], Gordon J. Freeman PhD [a, ¶], Rosemarie H. DeKruyff PhD [b, c, g, ¶] ✉


Journal of Allergy and Clinical Immunology

Volume 117, Issue 5, May 2006, Pages 1014-1020

Asthma diagnosis and treatment

Socioeconomic status and inflammatory processes in childhood asthma: The role of psychological stress


Presence and inter-individual variability of carboxylesterases (CES1 and CES2) in human lung

Morena Gabriele [a], Paola Puccini [b], Marco Lucchi [c], Anna Vizziello [c], Pier Giovanni Gervasi [a], Vincenzo Longo [a] ✉

Existing Studies for CES1 and NEO1 with Asthma

- Previous studies have shown associations of air pollution, CES1 or NEO1 independently (separately) with asthma
- Our mediation analysis results hypothesize that air or traffic pollution could possibly alter the expression levels of the CES1 gene in the human lung putatively affecting the asthma phenotype, thus suggesting a potential pathway or cellular mechanism
- Lung injury results in induction of NEO1 expression and activates the BMP pathway in bronchial epithelial cells, potentially promoting lung inflammation.

Existing Studies for other significant exposure families with Asthma

- Previous studies have also shown that Organophosphate pesticides, Polybrominated diphenyl ethers (PBDE), Per- and polyfluoroalkyl substances (PFAS) and Water DBPs are are associated with an increasing risk of asthma.
- Further studies are needed to elucidate the mechanism by which the above families influence the asthma phenotype via the mediator genes.

Discussion

- Our proposed analysis pipeline jointly models the exposome and transcriptome data using the framework of mediation analysis
  - Account for multiple mediators in high-dimensional case
  - Reduce exposure dimensionality by introducing family-based environmental risk scores and estimated their cumulative effect

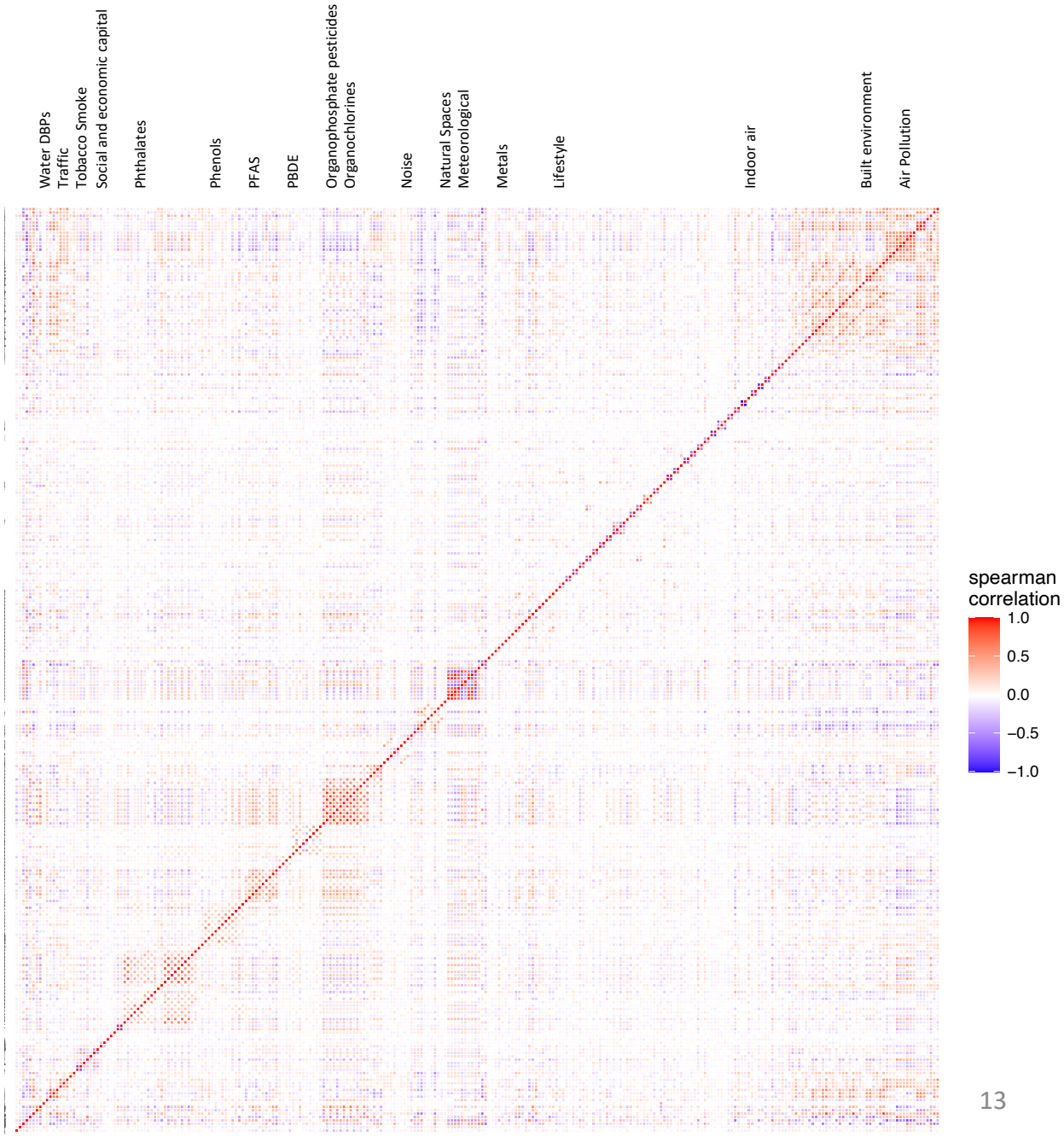
- Future work
  - Gene pathway/clusters mediation analysis
    - Reducing the dimensions of the mediator matrices can be a powerful tool for causal inference, especially when coupled with literature derived organization of the mediator matrix by biological pathways
  - Incorporate more omics as mediator layers
    - Incorporating other omics, e.g., methylome, metabolome, and their interactions with the transcriptome that are known to affect health outcomes could further explain the etiology

Acknowledge

- NIEHS BCBB Branch
- Dr. Alison Motsinger-Reif Group
  - Farida Akhtari
  - John House
  - Eunice Lee
  - Dillon Lloyd
  - Xiaoran Tong

Supplementary

- Correlation heatmap for all exposures
  - Clear pattern within each exposure family



Small Mediation Effect