



National Institute of Environmental Health Sciences
Your Environment. Your Health.

Discovering Gene-Environment Interactions

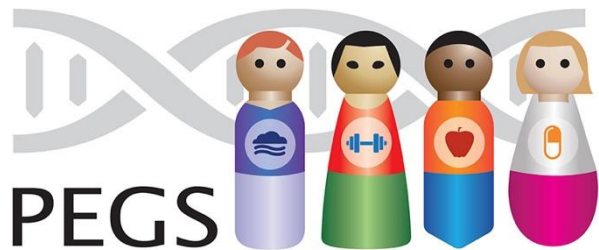
Alison Motsinger-Reif, PhD

Branch Chief, Senior Investigator

Biostatistics and Computational Biology Branch

Research Overview

- Applied and methodological research



Personalized Environment & Genes Study

Powerful science for integrating genetics and environmental data



Team Members



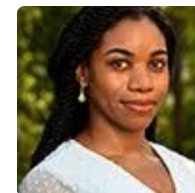
Dr. Farida Akhtari



Dr. Dillon Lloyd



Dr. John House



Jasmine Mack



Dr. Ryan Campbell



Dr. Joe Breeyear



Cristina Justice



Dr. Kwangmi Ahn

Other Key Co-Conspirators



Dr. Jan Hall



Dr. David Fargo



Dr. Charles Schmitt



Dr. Kyle Messier



Dr. Geoff Ginsburg



Dr. Rick Woychik

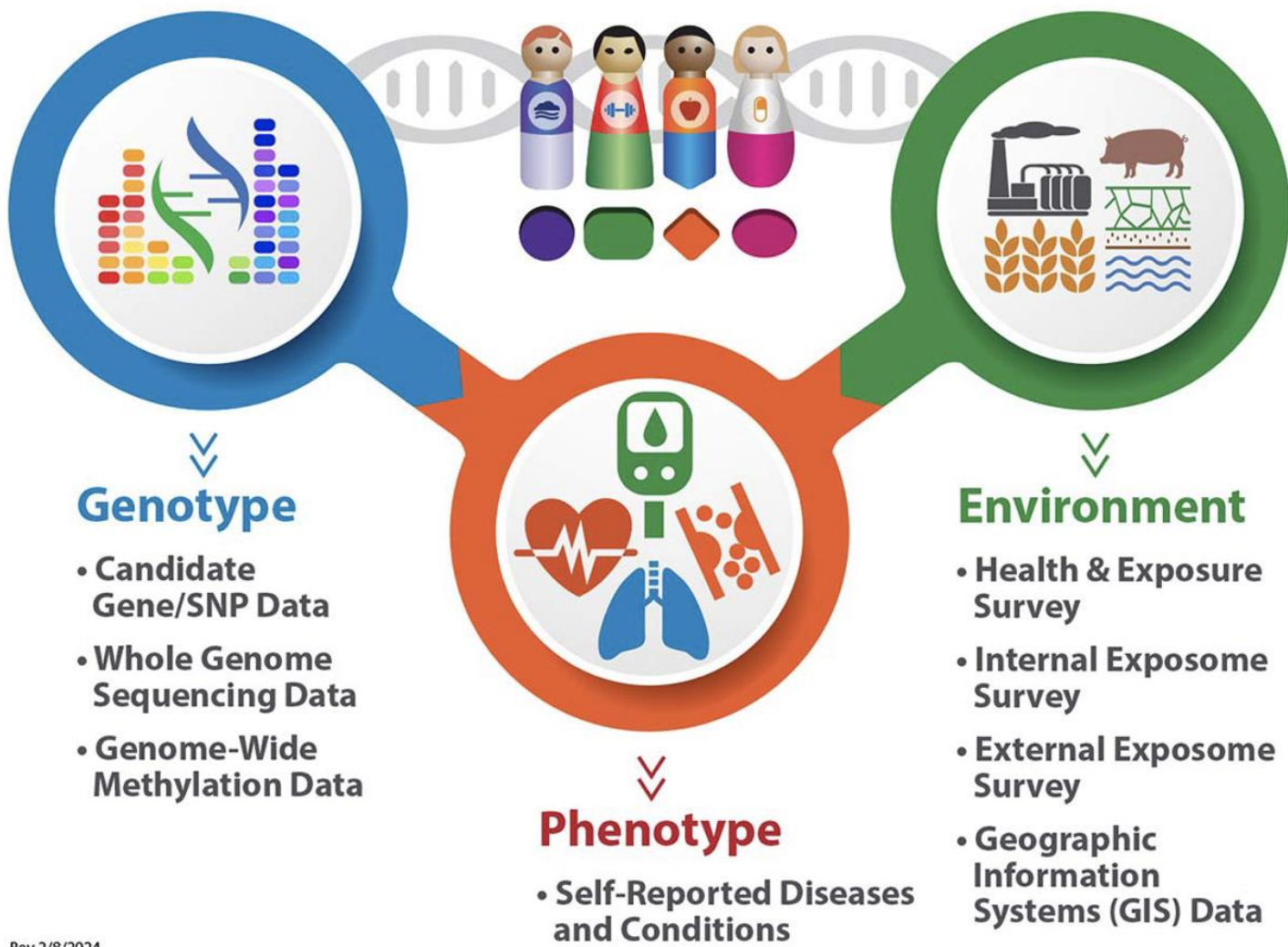


Dr. Larry Kirschne



Dr. Josh Denny

Personalized Environment and Genes Study



PEGS

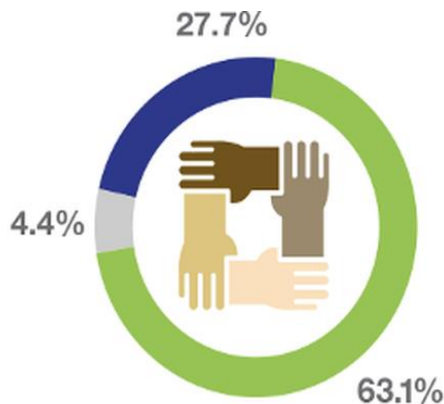
Demographics



Average Age: 49.7

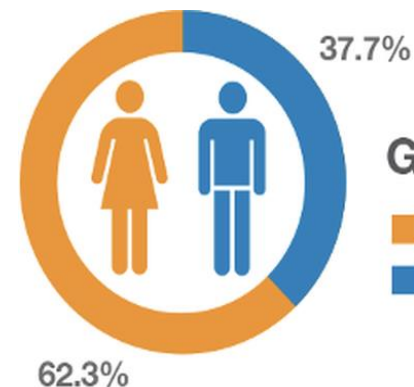
Race

- White
- Black
- Other

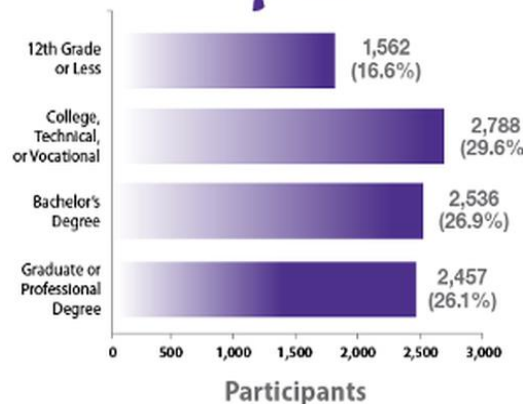


Gender

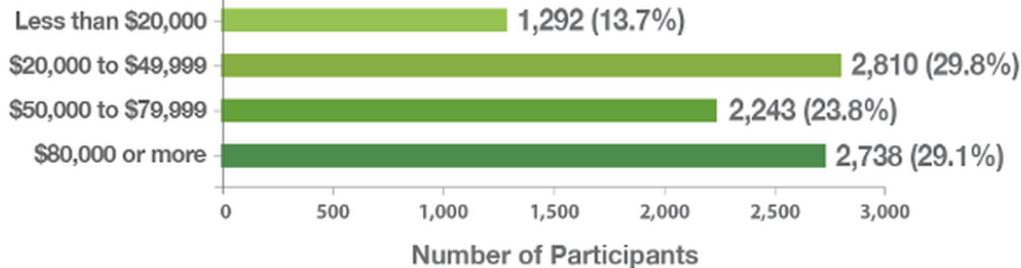
- Female
- Male



Education

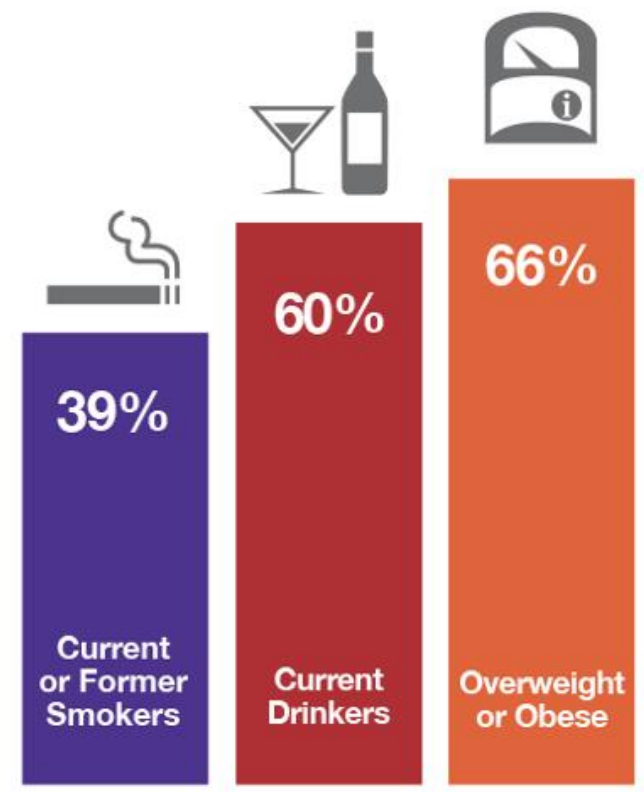
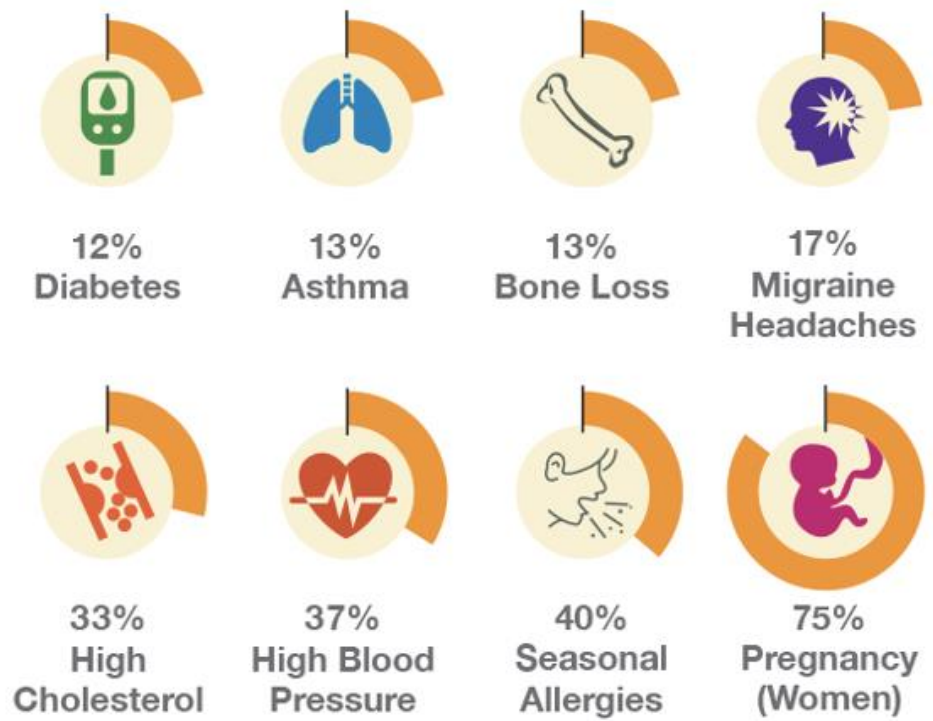


Income



Health and Exposure Survey

Self-Reported Diseases or Conditions | Number: 122



Lifestyle Factors

Exposome Surveys

Part A: 'External Exposome'

- Characteristics of current and past homes
- Workplace characteristics
- Chemical and metal exposures at work
- Hobby exposures
- Ultraviolet light exposures



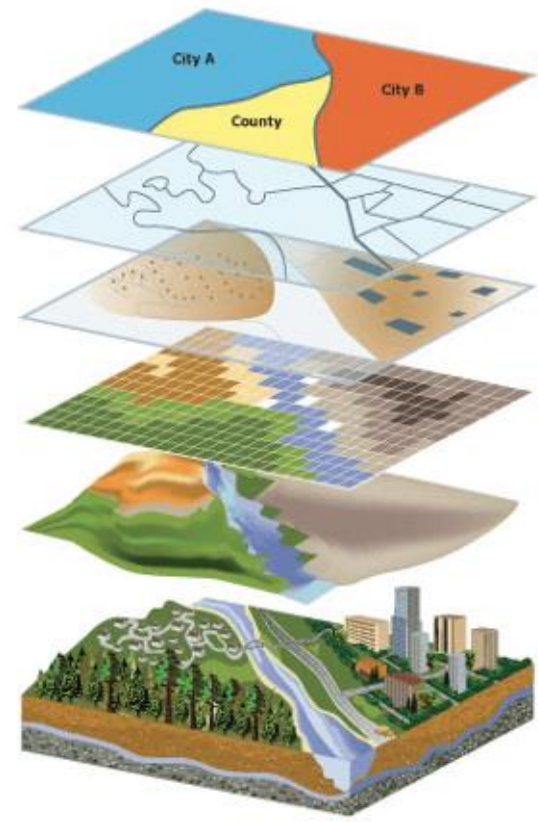
Part B: 'Internal Exposome'

- Medications
- Vitamins, minerals, dietary supplements
- Chemotherapy/radiation therapy
- Physical activity
- Stress
- Infection
- Sleep
- Dietary behavior
- Dietary intake
- Siblings/twins/birth order
- Genetic history

GIS Data

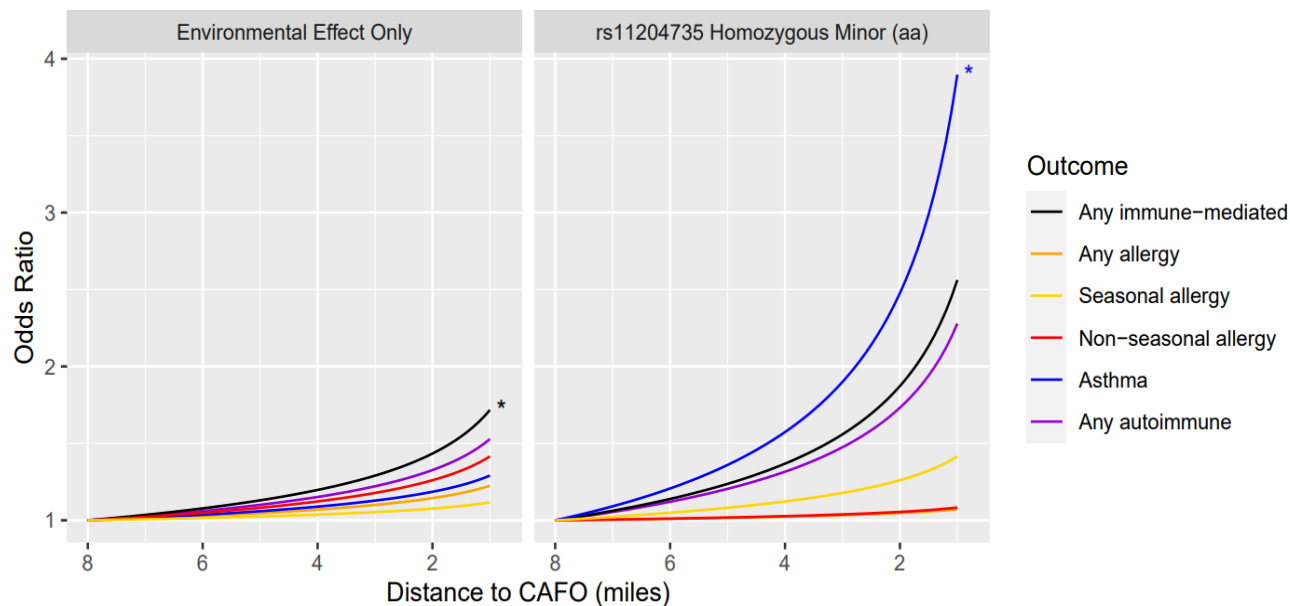
Growing Number of Data Layers

- Airports
- CAFOs
- Cellular towers
- Drinking water
- Dry cleaners
- Hazardous waste
- Highways
- Nuclear sites
- Wastewater
- Population info
- Power lines
- PR landfills
- Railroads
- Spills
- Sanitary landfills
- Superfund sites
- Toxic release sites,
- Etc.



Address at the time of survey completion and longest lived childhood address

Caged Animal Feeding Operations and IMD



Logistic regression adjusted for sex, race, age, income, and smoking status. * Indicates significant effect.

Summary

- CAFO proximity was associated with multiple IMD phenotypes suggesting possible shared disease mechanisms
- Multiple PTPN22 and AHR pathway gene interactions with CAFO exposure were identified

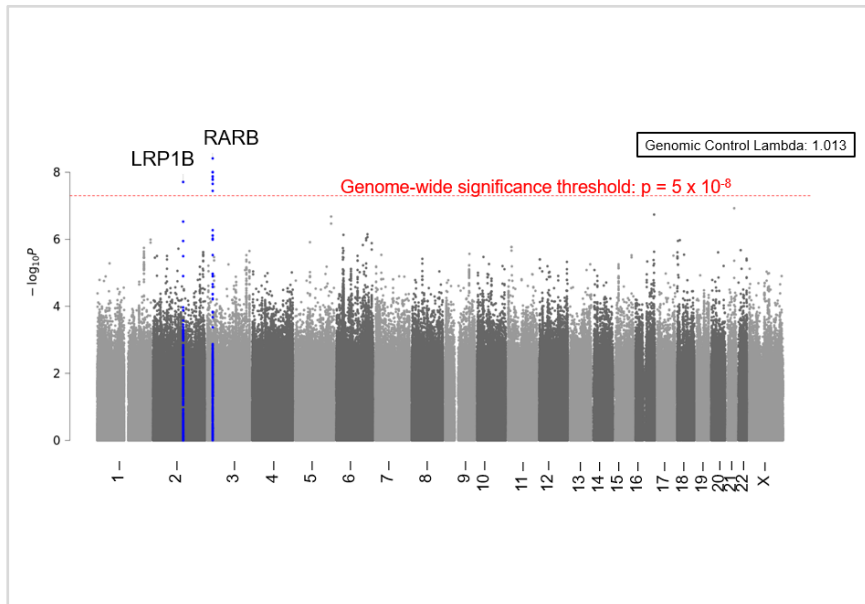
Exploring Landscape of Gestational Hypertension

Significant genetic variant association near *LRP1B* (chr2) and *RARB* (chr3)

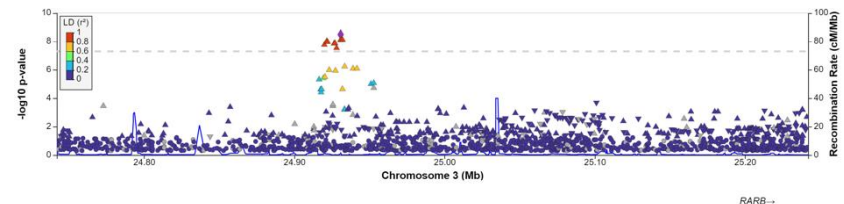
- 202 cases and 1,569 controls- Racial/ethnic groups: 71% White, 20% Black, 3% Hispanic/Latino, 2% Asian, and 4% Other
- Retinoic acid (RA) signaling gene *RARB* replicated in UKBB
- *RARB* Shown to be associated with the severity of proteinuria in the preeclamptic fetal genome



Jasmine Mack

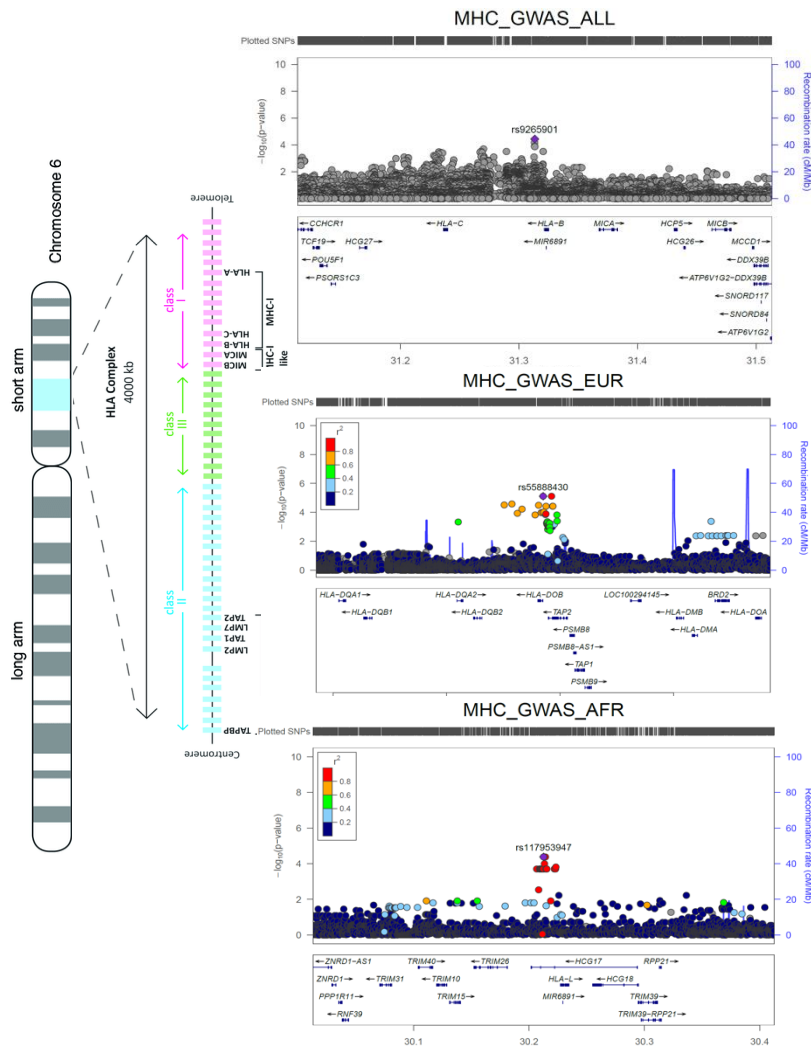


Regional Plot of Candidate Region near *RARB*



- The Lead SNP is at chromosome 3 (rs61176331): of cases, 20.9% are heterozygous at this locus, while 7.7% of controls are heterozygous.

MHC Region Mapping and Adult-Onset Asthma



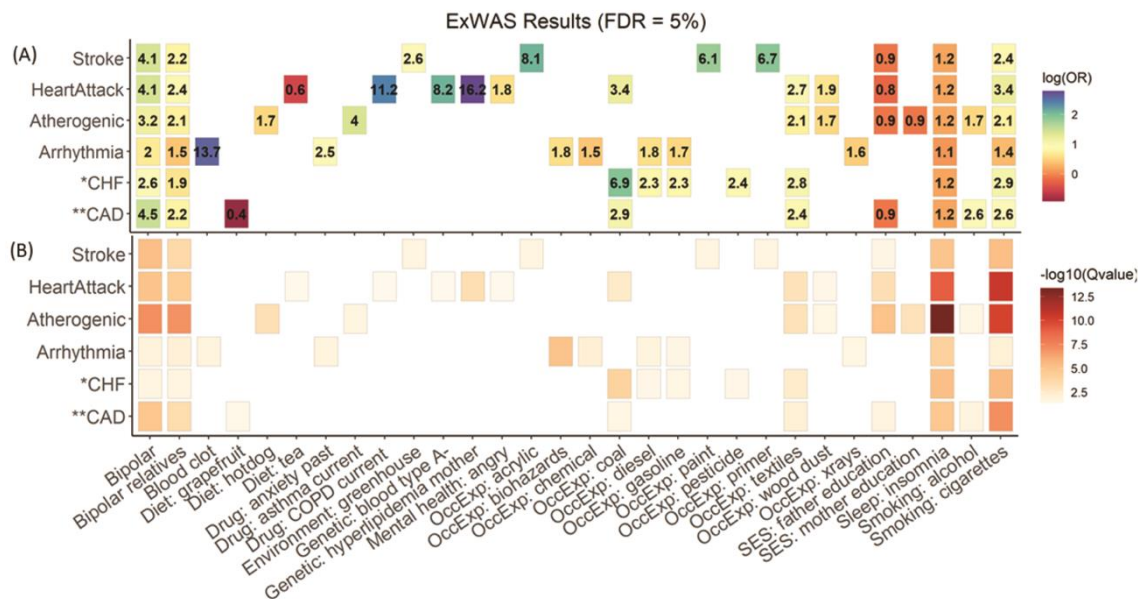
Study Objectives

- Tested population-specific SNP and HLA associations

Findings

- **MHC-wide Association Study (MWAS)**
 - SNPs with evidence of functional elements : *rs9265901/HLA-B (pooled)*, *rs55888430/HLA-DOB (EUR subjects)*, *rs117953947/HCG17 (AFR subjects)*
- **HLA Analysis**
 - Class I HLA genes (B and C) for participants with EUR ancestry
 - Class II HLA gene (DRB1) for participants with AFR ancestry – HLA-DRB1*0301 associated with childhood asthma (Hanchard et al., 2010)

ExWAS Analysis of Cardiovascular Disease



Study Objective

- Identify **unique** and **overlapping nongenetic** risk factors associated with **CVD-related** outcomes

Main Results

- Blood type A (Rh-): heart attack
- Paint-related exposure (acrylic paint, primer): stroke
- Biohazardous materials : arrhythmia
- Paternal education level : stroke, heart attack, CAD, combined atherogenic outcome



Dr. Eunice Lee

Exposome Wide Association Studies

PEGS



Health and Exposure Survey

Participants: 9,414
Questions: 496



- Alcohol
- Smoking
- Mood
- Family Health
- Socioeconomic Status
- Occupational Exposure
- Sleep

External Exposome Survey

Participants: 3,519
Questions: 607



- Occupational Exposure
- Workplace Characteristics
- Residence
- Hobby
- Ultraviolet Light

Internal Exposome Survey

Participants: 2,962
Questions: 719



- Supplements
- Medications
- Chemo/Radiation
- Stress
- Infectious Disease
- Sleep
- Dietary Behavior
- Exercise
- Genetic History

Disease Phenotyping

PEGS prevalence cutoff = 10%



- | | | |
|-------------------|-----------------------|---------------------------|
| Hypertension | Ovarian Cysts | Bone Loss |
| Cholesterol | Migraines | Type 2 Diabetes |
| Allergic Rhinitis | Lower GI Polyps | Atherogenic Heart Disease |
| Fibroids | Iron Deficient Anemia | |
| | Asthma | |

Exposome-Wide Association Study (ExWAS)

Single Exposure Modeling (Logistic Regression)

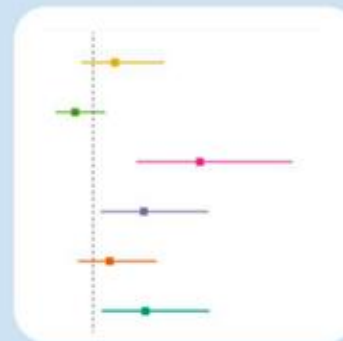
$$Disease\ Status = Exposure + Age + Sex + Race + \epsilon$$



Multi-Exposure Modeling (Deletion/Substitution/Addition)



Results Visualization



Correlations

Choose Correlation Type



Fit Model for Shrinkage



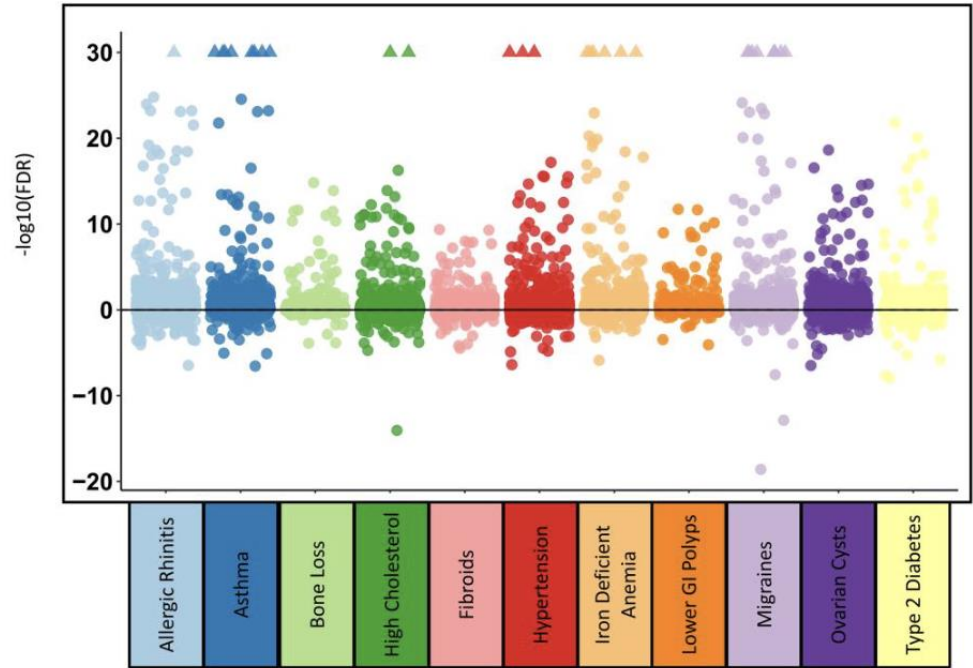
Results Visualization



PEGS Web Application

ExWAS of High-Prevalent Traits

- Significant associations were found with:
 - lifestyle and socioeconomic factors
 - exposures such as asbestos, dust, biohazards, and textiles
- Both shared and unique signals
- AI-driven literature review identified both known (e.g., smoking, sleep) and novel (especially occupational) exposure-disease associations



Miami plot of ExWAS results

Lloyd et al. Exposome 2024



Results Dissemination

The screenshot shows the PEGS Explorer website interface. The main content area is titled 'About Your General Health' and displays search results for 'Asthma'. The results are presented in a list format with columns for N, FDR, and OR(95%CI). A circular chord diagram on the right side of the results area visualizes the relationships between the different search results.

Search Result	N	FDR	OR(95%CI)
A003 Weight at birth Birth weight: 'Between 5 1/2 and 9 pounds' Ref Level: 1 = 'Less than 5 1/2 pounds'	6696	<.001	0.64(0.51,0.81)
A004a Weighed less than 3.5 pounds at birth At birth weighed less than 5.5lbs: 'Yes' Ref Level: 0 = 'No'	2021	0.00	
A004b Pre-term or premature at birth Born pre-term: 'Yes' Ref Level: 0 = 'No'	4675	0.00	

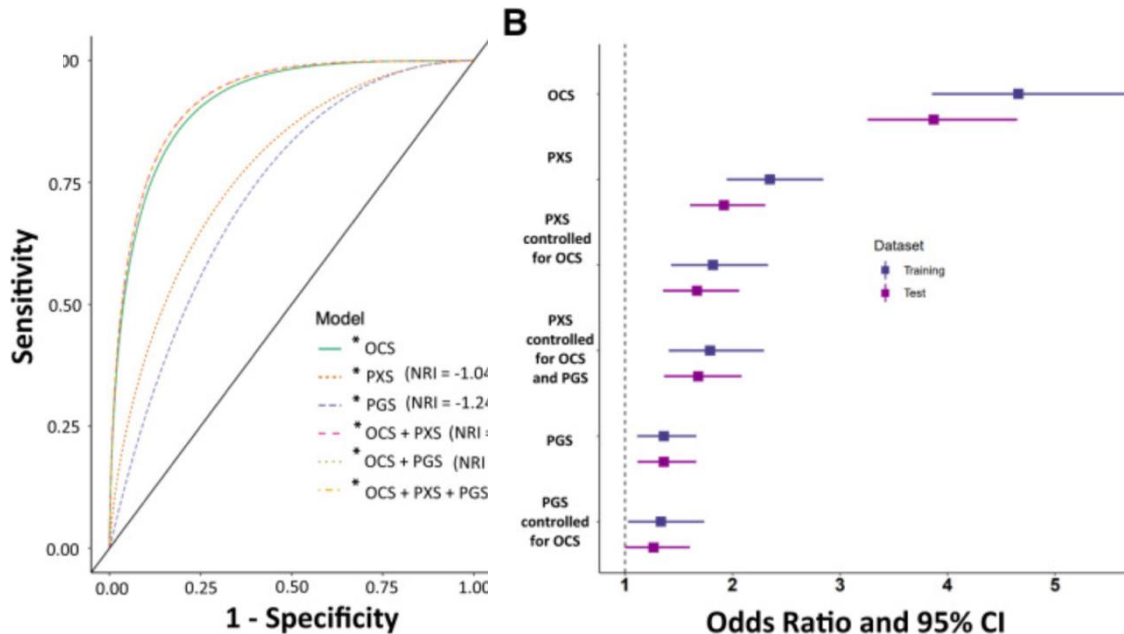
<https://pegsexplorer.niehs.nih.gov/>

Polyexposure Score (PXS)

- Aggregates multiple exposures into a single score
- Captures cumulative effects of diverse, potentially modifiable factors on health or other outcomes
- Emphasizes combined exposure patterns rather than any single risk source
- Includes intervenable exposures that can be modified through targeted action
- Examples of intervenable factors:
 - Occupational hazards (e.g., chemical exposure, ergonomic risks)
 - Hobbies or lifestyle choices (e.g., sports-related risks)
 - Stress exposures (e.g., work-related stress, chronic social stress)

Genetic and Environmental Scores and T2D

- Computed: Overall Clinical Score (OCS), polyexposure scores (PXS) and polygenic scores (PGS)
 - Rigorous train/test/validate approach
- PXS explained a larger proportion of variance compared to PRS
 - Added predictive value to OCS



Akhtari and Lloyd et al.
Diabetes Care 2023

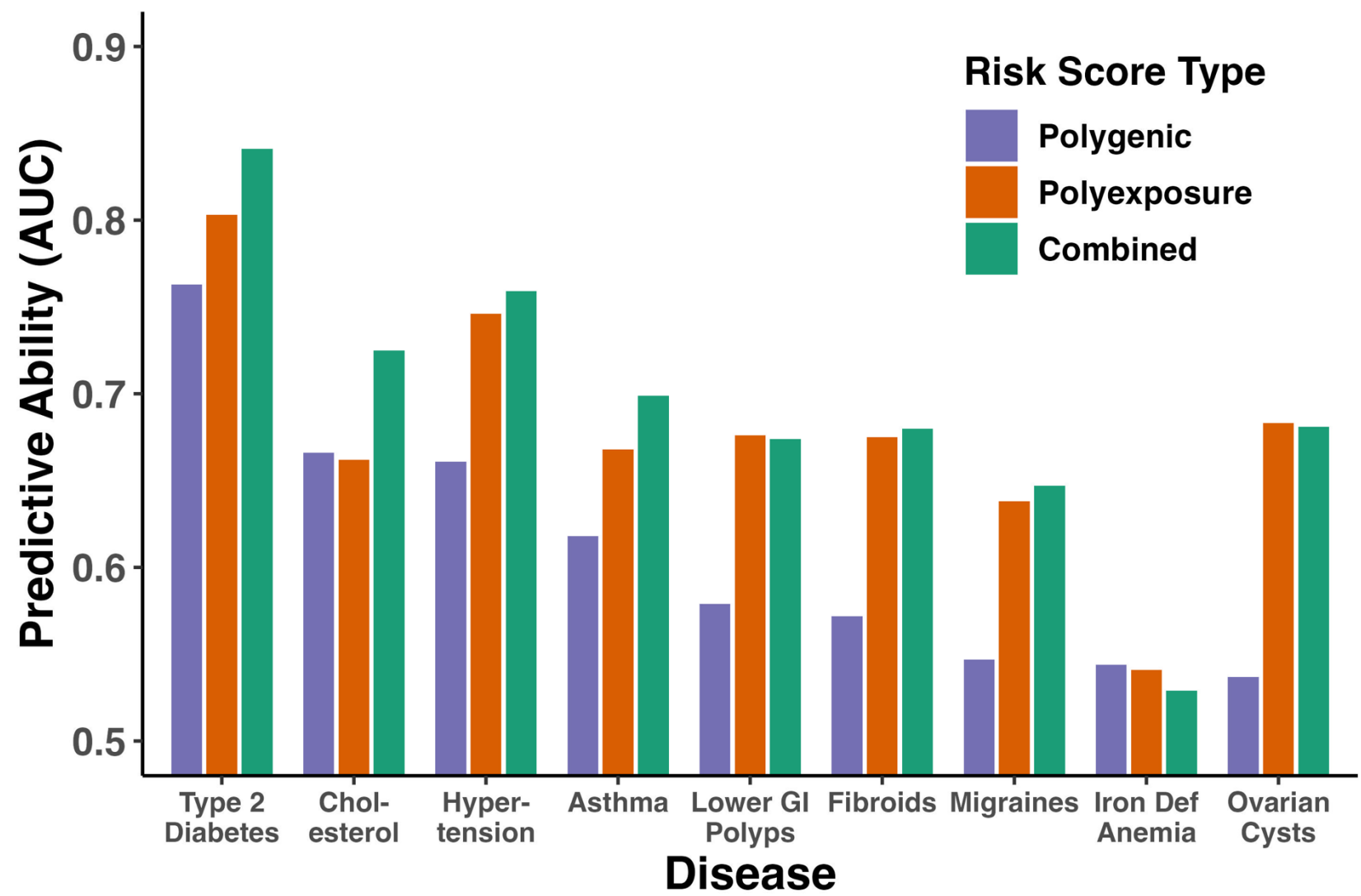


Dr. Farida Akhtari



Dr. Dillon Lloyd

Comparison of Genetic and Exposure Risk Scores in PEGS



Ongoing Work and Future Directions

- Separating exposures into two different polyexposure scores
- Incorporating advanced ML/AI methods to build exposure scores
- Understanding PGSs in Context of Exposure, building on Schaid et al. 2025

Polysocial Score (PSS)

- **Non-intervenable**, structural features – factors that are more difficult to change at an individual level

Examples:

- Income or socioeconomic status
- Housing type or living conditions
- Social vulnerability indices

Polyexposure Score (PXS)

- **Intervenable** exposures – factors that can be modified or addressed through targeted interventions.

Examples:

- Occupational hazards
- Hobbies or lifestyle choices
- Stress exposures



Solving Problems. Together.

- Pose Question
- Prepare Data
- Engage Solvers
- Evaluate Models
- Share Results

[LEARN MORE](#) [JOIN A CHALLENGE](#)

DREAM Challenges use crowd-sourcing to solve complex biomedical research questions

60+
Crowd-sourced DREAM Challenges have benchmarked informatic algorithms in biomedicine

30,000
Cross-disciplinary participants from around the world have volunteered as solvers.

105+
Academic journal publications have resulted from DREAM Challenges covering a range of disease areas

PEGS DREAM Challenge

Personalized Environment and Genes Study Challenge



Task 1: Disease Classification

- **Aim:** Classify hypercholesterolemia status using **any combinations** of the health, exposomics, geospatial, and genomic data available for the PEGS DREAM Challenge.

Task 2: Ideation Challenge

- **Aim:** Develop a novel hypothesis(-es) and/or generate model(s) using the **multi-dimensional** PEGS DREAM Challenge data to improve the understanding of the etiology of hypercholesterolemia beyond conventional clinical and genetic risk factors.

Ongoing Work

- GIS ExWAS
- EJI ExWAS
- Genetic correlation analysis
 - Genetic associations with exposures
- Epigenetic Wide Association Studies (EWAS)
- Epigenetic biomarkers of exposure and protein expression association studies
- Integrative 'omics analysis (ex. TWAS)
- Expansion of PXS vs. PGS comparisons

All of Us Ancillary Studies



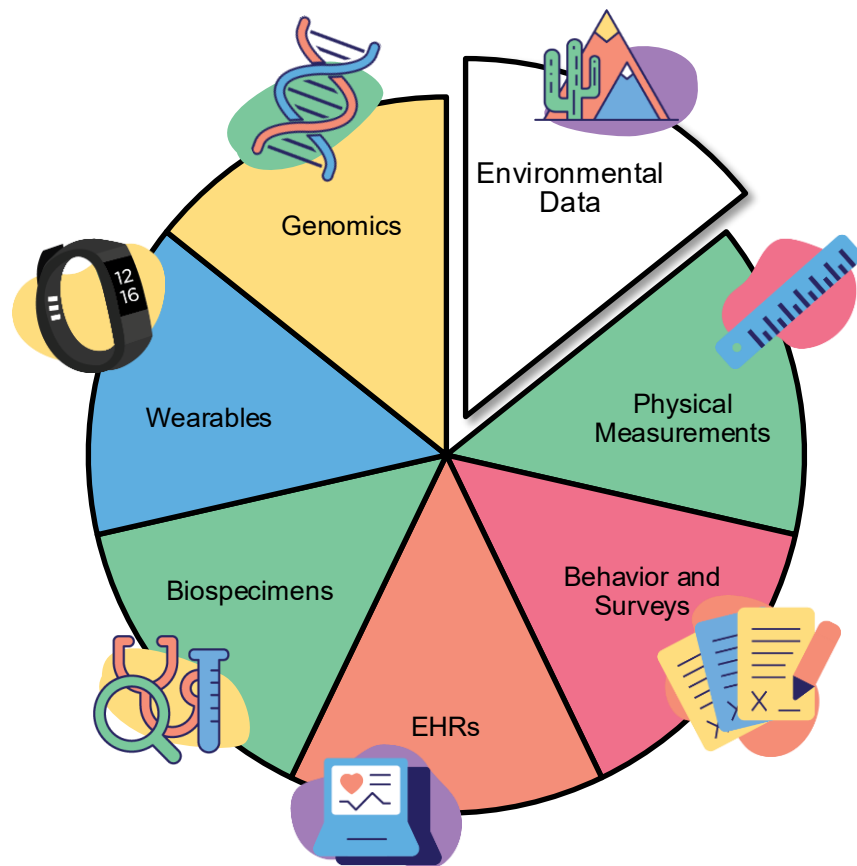
The *All of Us* Research Program is a historic, longitudinal effort to **gather data from one million or more people** living in the United States to **accelerate research and improve health**. By taking into account individual differences in **lifestyle, socioeconomics, environment, and biology**, researchers will uncover paths toward delivering **precision medicine – or individualized prevention, treatment, and care – for all of us**.



“All of Us is among the most ambitious research efforts that our nation has undertaken!”

NIH Director Francis Collins, M.D., Ph.D.

Integrating Environmental Data into All of Us



To meet scientific priorities and goals for precision medicine, All of Us will need to integrate individual location and environmental exposure data.

This collaboration will lay the groundwork for investigating the health implications of interactions between the exposome and other data types available in the All of Us Researcher Workbench.

Far-reaching research potential of integrating environmental data



Health Equity

How do socioeconomic factors interact with environmental exposure to increase disease risk?



Empowerment

How can returning information about environmental risks and exposures empower participants?

Risk & Prevention

What environmental biomarkers can help identify risk for future disease?



Diagnosis

What exposure signatures are typical of patients with different conditions?



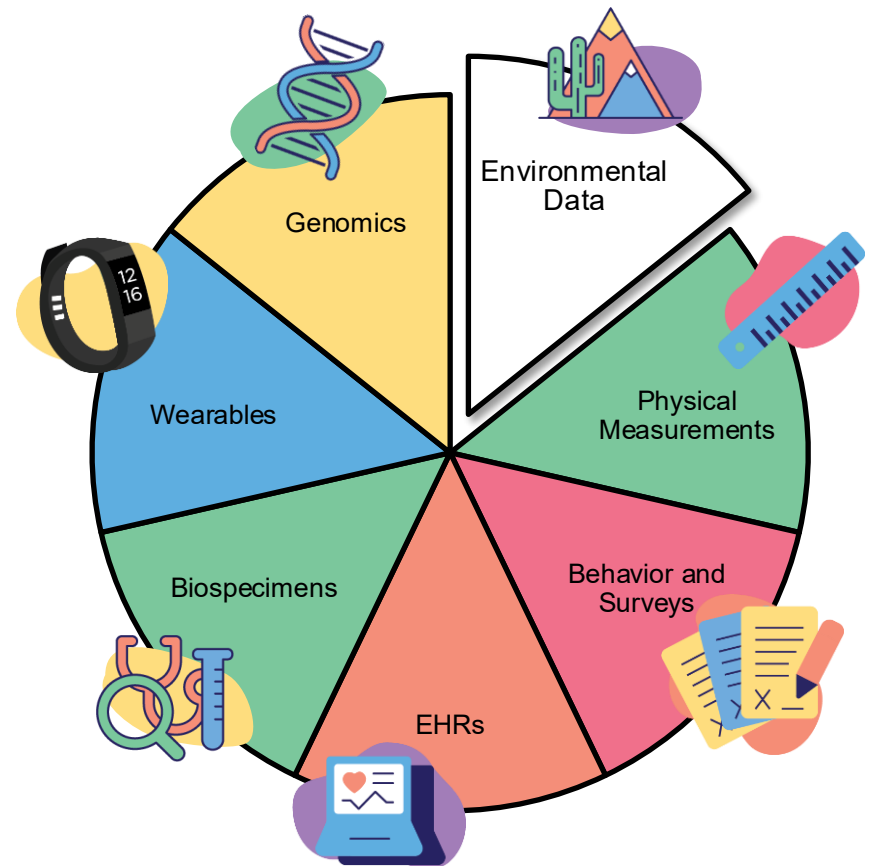
Treatment & Outcomes

How can we objectively assess whether a treatment or intervention will be effective?



All of Us Partnered Research Studies

- Exposomics
 - Case-cohort study of Type 2 Diabetes
 - Untargeted metabolomics on ~5,600 samples
- Environmental Exposures and Occupational Health Partnered Research Study
 - Two questionnaires
 - Residential history and exposures
 - Work history and occupational exposures



Encorporating the Environment into All of Us

Phase 1:

Location

Geocoded location information across the lifecourse

Exposure survey

For the entire cohort

Phase 2:

Geospatial Environment

Initial linkages to the Environmental Justice Index

Additional linkages with PCORTF tools

For the entire cohort

Phase 3:

Health

HHEAR untargeted assays on a case-cohort study of T2D

n~5600

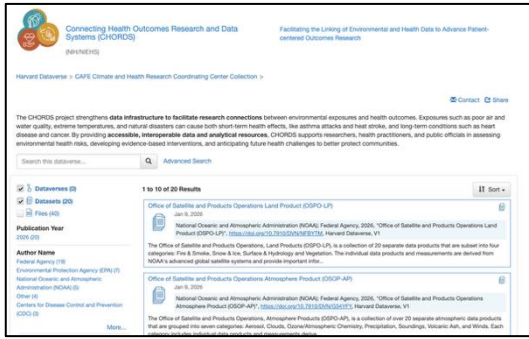
Supported by a contract through the Center for Linkage and Acquisition of Data (CLAD) and CHORDS

Supported by NIEHS



Connecting Health Outcomes Research Data Systems (CHORDS)

Deliverables



Data catalog with over 150 curated data resources.

Hosted by **HEW Café DataVerse** (>1000 data sets avail.)



Open source, validated, software packages and published data sets

6,979 Amadeus downloads, being used in CHORDS Use Cases & All of Us research project

Amadeus

Machine for Data, Environments, and User Setup for common environmental health datasets is an R package developed to improve and expedite users' access to large, publicly available geospatial datasets.

GeoIOx, or source-to-outcome, modeling framework with an S3 object-oriented approach. Facilitates the calculation and visualization of single and multiple chemical risk at individual and group levels.

Beethoven

Building an Extensible, Reproducible, Test-driven, Harmonized, Open-source, Versioned, Ensemble model for air quality is an R package developed to facilitate the development of ensemble models for air quality.

implements nested Generalized Concentration Addition: A geometric, piecewise inverse function for 3+ parameter sigmoidal models used in chemical mixture concentration-response modeling.

Chopin

Computation of Spatial Data by Hierarchical and Objective Partitioning of Inputs for Parallel Processing.

Scalable penalized regression on spatio-temporal outcomes using Gaussian processes. Designed for big data, large-scale geospatial exposure assessment, and geophysical modeling.

Demonstration Use Cases

- AHRQ Healthcare Cost and Utilization Project (HCUP) National & State Databases
- NIEHS Personalized Environment and Genes Study (PEGS) Cohort
- NHLBI Cardia Cohort

Online tutorials, vignettes, & publications

8 HCUP and Amadeus Smoke Plume Use Case

Integrating HCUP databases with Amadeus

Date Modified: April 26, 2025
Author: Darius M. Bost
Programming Language: R

8.1 Motivation

Understanding the relationship between external environmental guiding public health strategies and policy decisions. Integrating Healthcare Cost and Utilization Project (HCUP) with data from us to examine how elements such as air quality, wildfire emissions, hospital visits and healthcare utilization patterns.

Ultimately, linking HCUP and environmental exposure data enables researchers better quantify environmental health risks.

download_data Function

download_data

Motivation

The `download_data` function was developed to improve researcher access to available environmental data, although the data are already available online (https://www.epa.gov/air-quality-data). This function allows researchers to download data from a variety of sources. This wrapper function calls source-specific code and returns a unique combination of input parameters, host URL, source format.

download_data

Download Function	Data Source
14 download_garden	ChronicDiseaseData
13 download_largeriver	ChronicDiseaseData
6 download_smokeplume	Health Effects Research Center (HERC)
6 download_land	USGS National Wetlands Inventory (NWI)

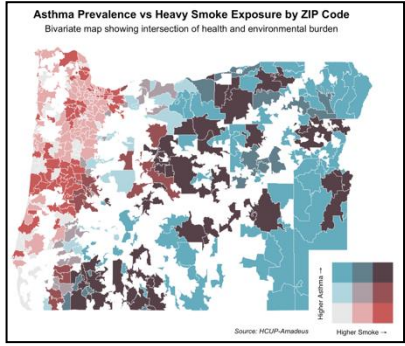
Review Article Open access Published: 06 September 2024

A review of geospatial exposure models and approaches for health data integration

Lara P. Clark, Daniel Zilber, Charles Schmitt, David C. Fargo, David M. Reif, Alison A. Motesinger-Reif & Kyle P. Messier

Journal of Exposure Science & Environmental Epidemiology 35, 131–148 (2025) | [Cite this article](#)

13k Accesses | 18 Citations | 2 Altmetric | [Metrics](#)



CHORDS Products

amadeus




A mechanism for environmental data and user setup in R for CHORDS exposure analysis.

Large-Scale Downloads

Downloads environmental data from 20+ sources including NASA, NOAA, EPA, and USGS.

One Consistent Workflow

Automatically handles source-specific URLs, file names, formats, and metadata.

 Amadeus ↗ repo status: Active Machine for Data, Environments, and User Setup for common environmental health datasets is an R package developed to improve and expedite users' access to large, publicly available geospatial datasets.	 Beethoven ↗ repo status: WIP Building an Extensible, Reproducible, Test-driven, Harmonized, Open-source, Versioned, Ensemble model for air quality is an R package developed to facilitate the development of ensemble models for air quality.	 Chopin ↗ repo status: WIP Computation of Spatial Data by Hierarchical and Objective Partitioning of Inputs for Parallel Processing.
---	---	--

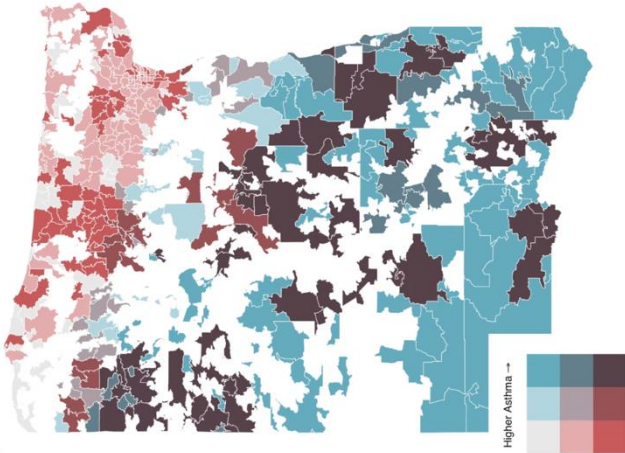


Dr. Kyle Messier

Asthma Demo: Wildfire Smoke Increases Rates

Using NOAA HMS smoke plumes (light, medium, heavy) for fire seasons **2014–2023**, linked daily at the ZIP3 level to **11,700–33,400 asthma participants/year** in *All of Us*.

Asthma Prevalence vs Heavy Smoke Exposure by ZIP Code
Bivariate map showing intersection of health and environmental burden



ZIP3-level exposure recovers plausible clinical signals in *All of Us* at population scale.

Key Findings

- **2017–2022:** consistent, significant increases in ED/UC rates
- **Per smoke day:** ~1–5% higher visit rates
- **Per 100-unit PM proxy:** ~10–43% higher visit rates
- **Largest effects:** 2019 and 2021

Exposure Construction

Intensity-weighted PM_{2.5} proxy: **5×light + 16×medium + 27×heavy** smoke days. Aggregated ZIP5 → ZIP3, linked by participant ZIP3 and event date.

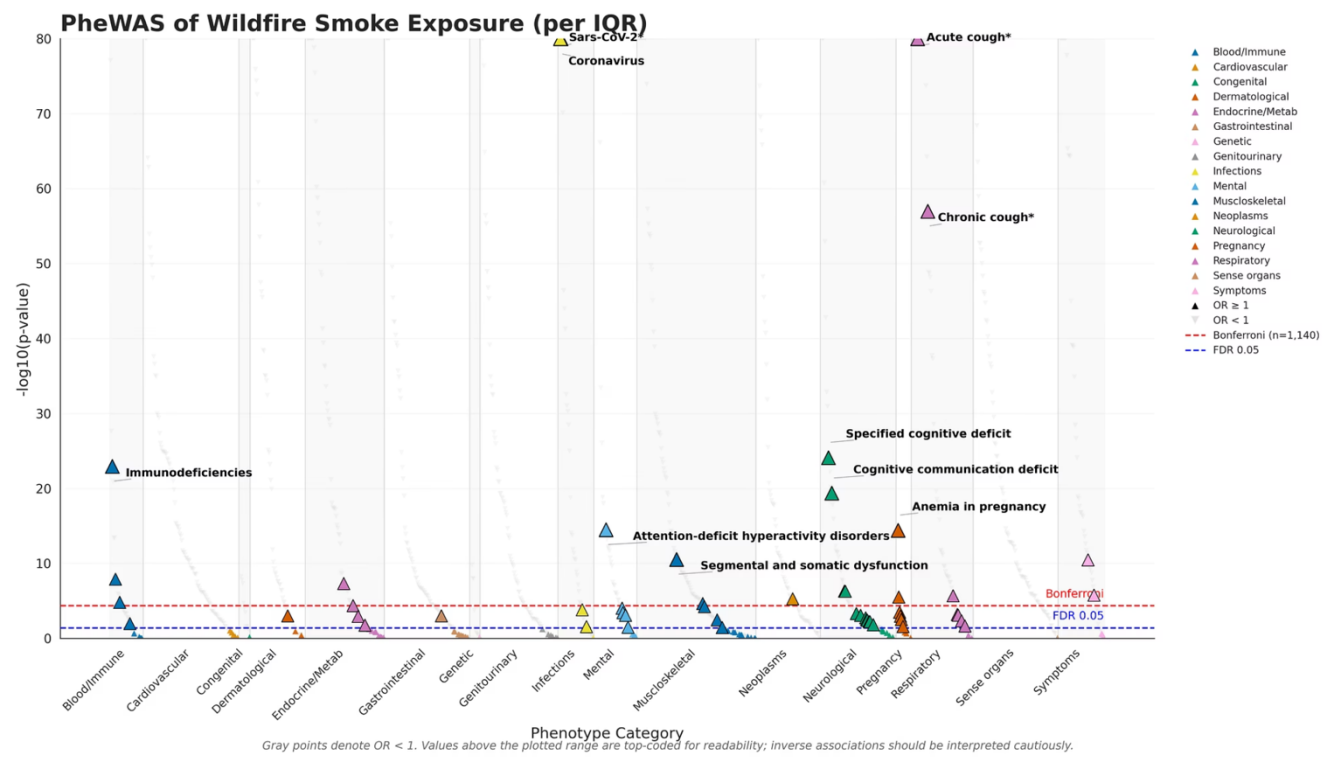
Outcome Model

Negative binomial model adjusted for age, sex, race/ethnicity, BMI, SVI, smoking, income, and EHR length offset.



PheWAS: Discovery-Scale Scan with Smoke

Using the same standardized pipeline, a phenome-wide association study tested **IQR-scaled PM_{2.5} proxy** (1-year lag, 2014–2023) against incident phecodes in the adult *All of Us* EHR cohort — going beyond a single clinical endpoint.



Study Design

Incident disease defined as ≥ 2 encounters with same phecode; prevalent cases excluded via 2-year washout. Logistic regression per phecode.

Covariates

Age, sex, race/ethnicity, region, smoking, BMI, SVI, and healthcare use. Effect scale: odds ratio per IQR increase in wildfire-smoke exposure.

Key Findings

Positive associations for **respiratory, infectious, neurological, and symptom phenotypes**. Inverse associations likely reflect confounding, healthcare use, or selection effects.

All of Us Exposomics Partnered Research Study

Launched in July, this study is performing **untargeted metabolomics** on **>5,600 diverse participant samples** to measure environmental exposures.

Data on **environmental, pharmaceutical, dietary, and endogenous metabolites** will be added to existing *All of Us* data in 2026.



New research on environment, gene, and behavior interactions will **drive discovery across health conditions.**

Samples were selected to reflect the high diversity of the *All of Us* cohort and for overlap with the multi-omics pilot. In this cohort 86% represent at least one community previously underrepresented in research and 62% represent an underrepresented race or ethnicity.

Untargeted Exposomics through the Human Health Exposure Analysis Resource (HHEAR)

Endogenous: Host Metabolism (~1,043)

Amino acids, carboxylic acids, biogenic amines, polyamines, bases, nucleosides and nucleotides, bile acids, carnitine, sugars, mono- and disaccharides, fatty acids, lipids, steroids, and hormones, as well as *coenzymes and vitamins*

Food Metabolome: Includes Subclasses, such as:

Phytoestrogens	Hippuric Acids
Aromatic Ketones	Hydroxytoluenes
Benzoic Acids	Phenylamines
Elegiac Acids	Stilbenes
Flavonoids	Urolithins
Caffeoylquinic Acids	Valerolactones
Catecholamines	Xanthonoids
Coumarins	

~Identified 550 compounds across ~50 subclasses via targeted analysis that are being transferred to untargeted platform

Environmentally Relevant Metabolites (~433)

AP Pesticides	Candidate compounds and metabolites were derived from the National Health and Nutrition Examination Survey (NHANES) and literature reviews based on known identifications using quantitative targeted methods.
PFAS	
Phenols	
Parabens	
Phthalates	
Tobacco-Related	
VOCs	
PBDEs, PCBs	Parabens: food additives
PAH Metabolites	Phthalates: food and water contaminants
Pesticides	

Additional Categories

Drugs and Medicines (48)
Choline Metabolism (24)
Folate Metabolism (17)
Vitamins and Vitamin-Like Compounds (30)
Carnitines and Acylcarnitines (39)
Amino Acids (24)
Omega 3 and Omega 6 Fatty Acids (5)

EEOH Partnered Research Study

In collaboration between NIEHS and *All of Us*, we will integrate participant-reported survey data on environmental, residential, and workplace exposures into the *All of Us* dataset. The environments where people live and work profoundly affect their health across the lifecourse. This survey will go out to all *All of Us* participants.



Home & Work Environment



Lifestyle & Behaviors



Social & Community Factors



Air & Environmental Pollution



Occupational Exposures



Diet & Chemical Exposures

Status

Questionnaires have been approved, translated, and are going through final rounds of burden testing.

Will launch to all *All of Us* participants in November 2026.

EEOH Partnered Research Study

In collaboration between NIEHS and *All of Us*, we will integrate participant-reported survey data on environmental, residential, and workplace exposures into the *All of Us* dataset. The environments where people live and work profoundly affect their health across the lifecourse. This survey will go out to all *All of Us* participants.



Home & Work Environment



Lifestyle & Behaviors



Social & Community Factors



Air & Environmental Pollution



Occupational Exposures



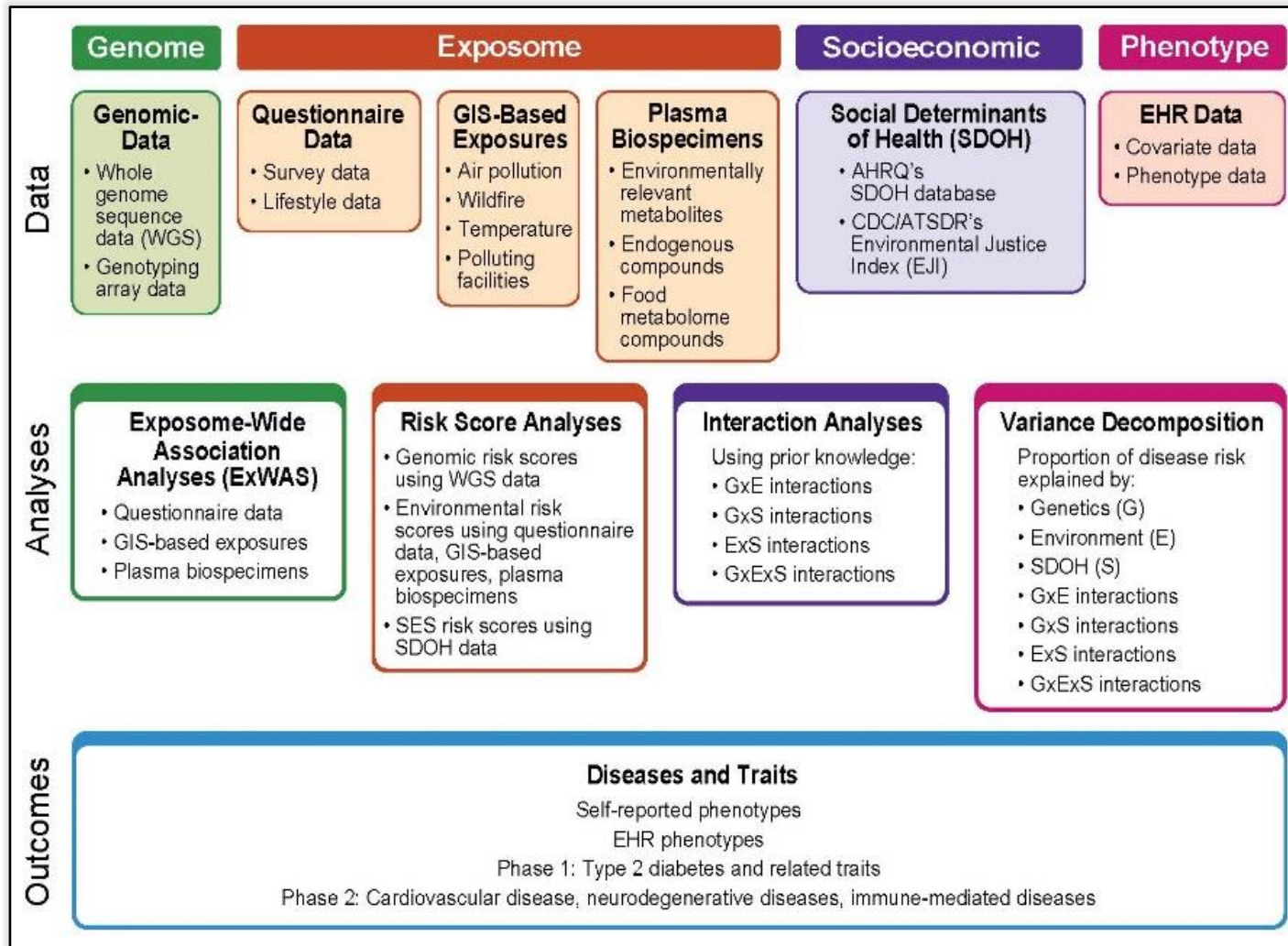
Diet & Chemical Exposures

Status

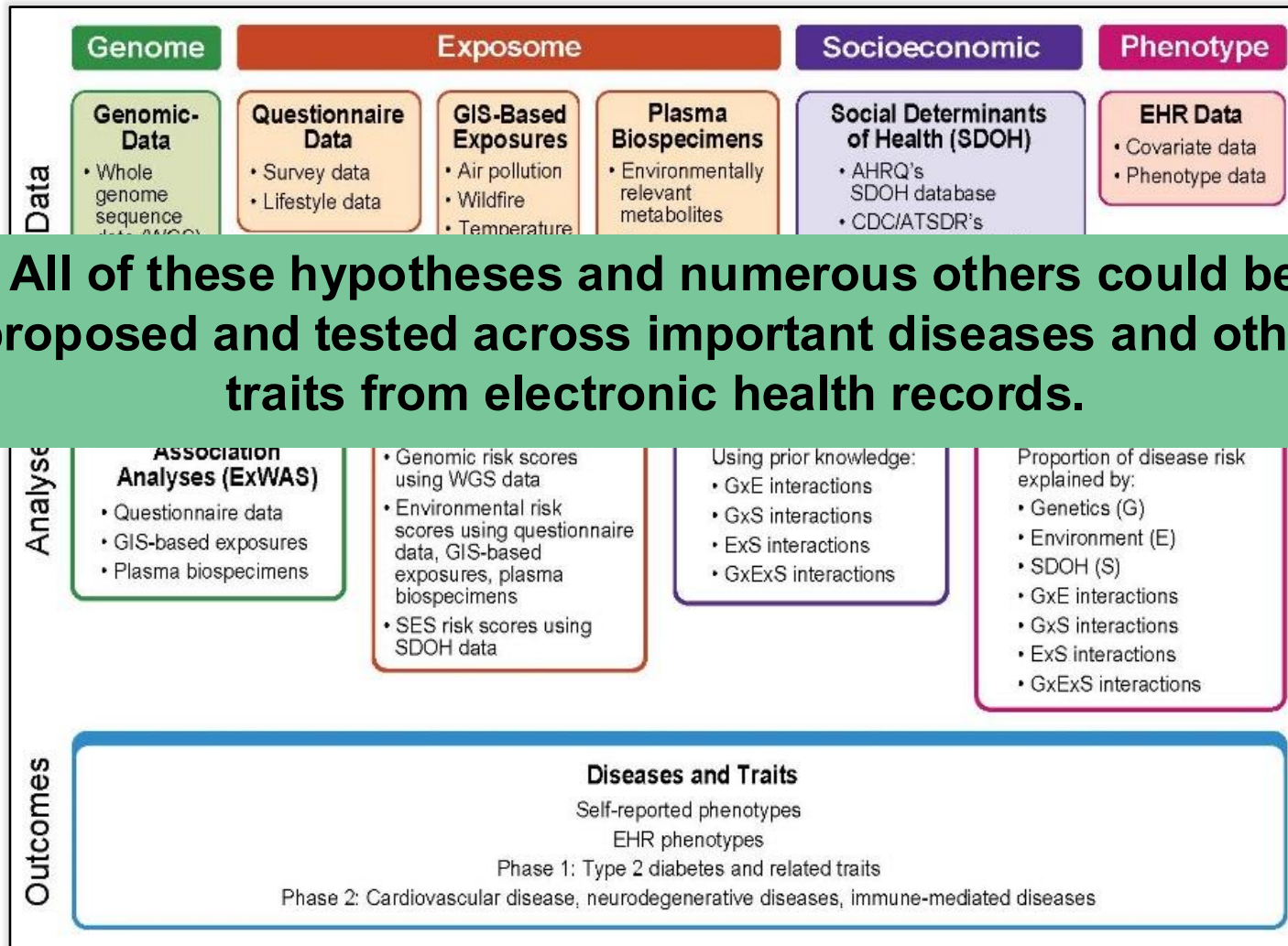
Questionnaires have been approved, translated, and are going through final rounds of burden testing.

Will launch to all *All of Us* participants in November 2026.

Supports a variety of hypotheses

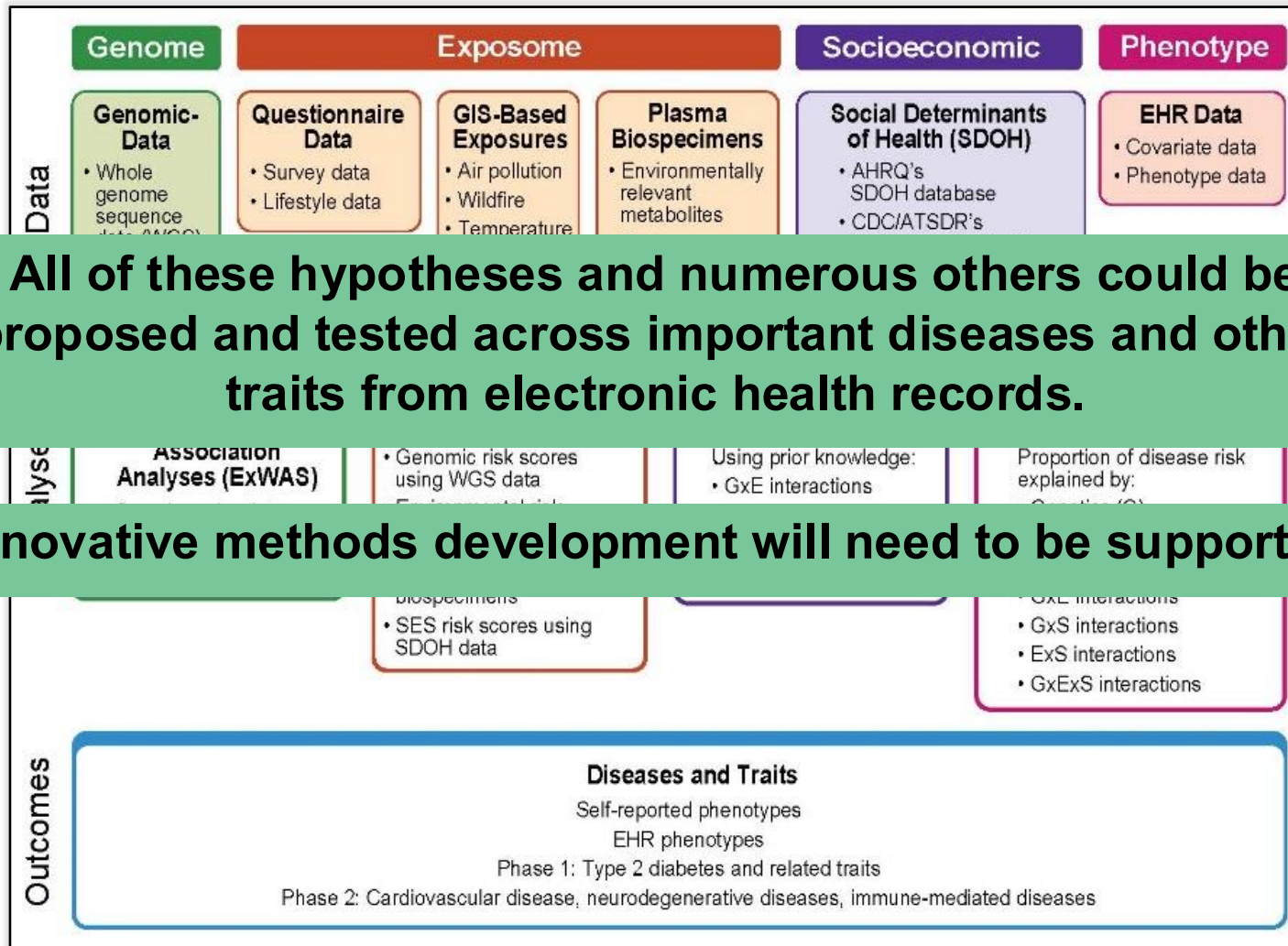


Supports a variety of hypotheses



All of these hypotheses and numerous others could be proposed and tested across important diseases and other traits from electronic health records.

Supports a variety of hypotheses



All of these hypotheses and numerous others could be proposed and tested across important diseases and other traits from electronic health records.

Innovative methods development will need to be supported.

A Whirlwind Tour



ChatGPT

Acknowledgements

- PEGS Leadership
 - Janet Hall (co-PI)
 - David Fargo
 - Charles Schmitt
- GIS Modeling
 - Kyle Messier
 - Melissa Lowe
 - Kristin Eccles
 - Nat McNell
- All of Us
 - Geoff Ginsburg
 - Erica Landis
 - Sheri Schully
 - Josh Denny
- Computational Work
 - Farida Akhtari
 - John House
 - Eunice Lee
 - Xiaron Tong
 - Dillon Lloyd
 - Jasmine Mack
 - Zongli Xu
- Bioinformatics
 - Adam Burkholder
 - Jianying Li
- Collaborators
 - Rick Woychick
 - Ross Simpson
 - John Buse
 - Gordon Smith
 - Les Beiseker
 - Rob Tighe
 - Ken Jacobson
 - Karamarie Fecho
 -
- Contract Support
 - S3
 - Sciome

PEGS Participants



Questions?

Alison.motsinger-reif@nih.gov