

PENDERGRASS LAB

Genomic Research in Action: Phenome-Wide Association Studies in Diverse Populations

Sarah A. Pendergrass PhD, MS
Assistant Professor
Biomedical and Translational Informatics
Geisinger Health System

October 14, 2016

GEISINGER



How Did I Get Here?

Math and analytical thinking and programming, never ending problem

sets

Working on disease based problems with computational solutions

- PhD in Genetics from Dartmouth College
- Emerging world of bioinformatics and using these tools to study disease
- Postdoctoral work at Vanderbilt University

Genetic epidemiology and bioinformatics – how can we leverage complex multifaceted data for discovery??

Investigator at Geisinger Health System

School of

Engineering at Dartmouth

PhD in Genetics from Dartmouth College

Postdoctoral work at Vanderbilt University

Investigator at Geisinger Health System

WHAT TO DO WHEN YOU'RE OVERWHELMED WITH WORK

STEP 1: MAKE A LIST OF ALL THE THINGS YOU HAVE TO DO.



STEP 2: CATEGORIZE THEM BASED ON THEIR IMPORTANCE.



STEP 3: REALIZE YOU JUST MADE IT WORSE.



WWW.PHDCOMICS.COM

WWW.PHDCOMI

JORGE CHAM © 2013

WWW.PHDCOMICS.COM

Piled Higher and Deeper by Jorge Cham

www.phdcomic

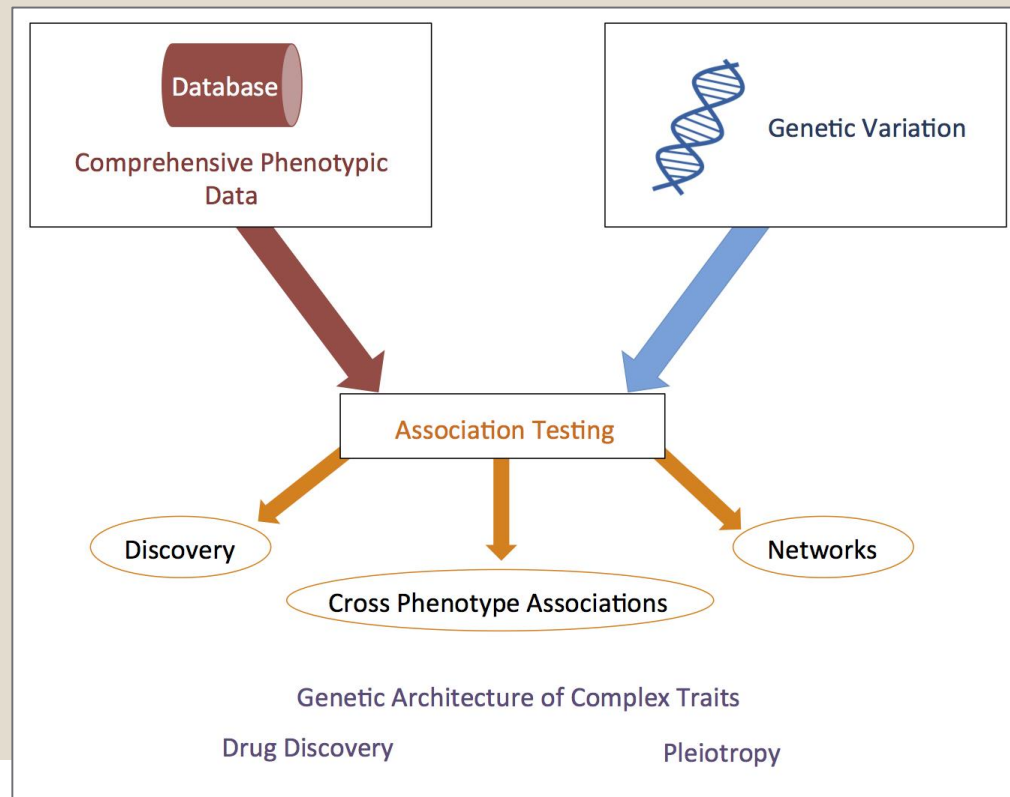


THE ORIGIN OF THE THESES

JORGE CHAM © 2009

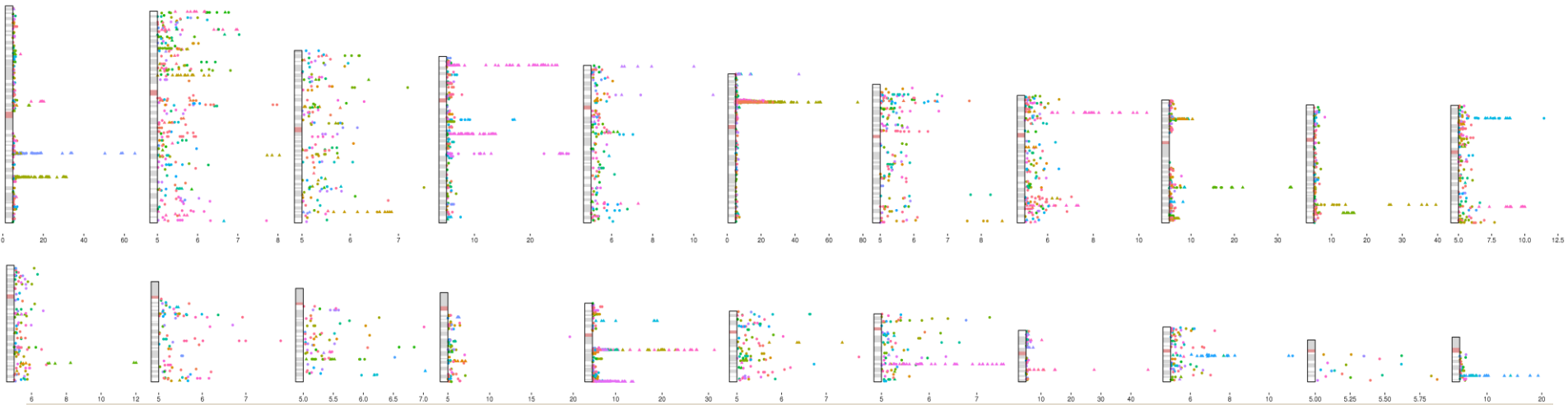
- What are Phenome Wide Association Studies (PheWAS)?
 - Genome Wide Association Studies (GWAS)
 - Single Nucleotide Polymorphisms (SNPs)
 - One phenotype, or a limited phenotypic domain
- Phenome Wide Association Studies (PheWAS)
 - Comprehensively calculating the association between
 - Small to large number of SNPs
 - Diverse range of phenotypes from phenotypically rich datasets
- We can also use other data types too!
 - Other genetic variants: copy number variants, rare variants, mitochondrial variants
 - Quantitative lab variables instead of SNPs for biomarker identification

- Phenome Wide Association Studies
 - Datasets with thousands of collected phenotypes
 - Unlikely that time would be spent to carefully investigate each phenotype for use
 - Hypothesis generation
 - Can we identify SNPs and/or phenotypes for further research?
- Also exploring cross-phenotype associations
 - Exploring the relationship between associations for the same genetic variant and multiple phenotypes



MegaPheWAS

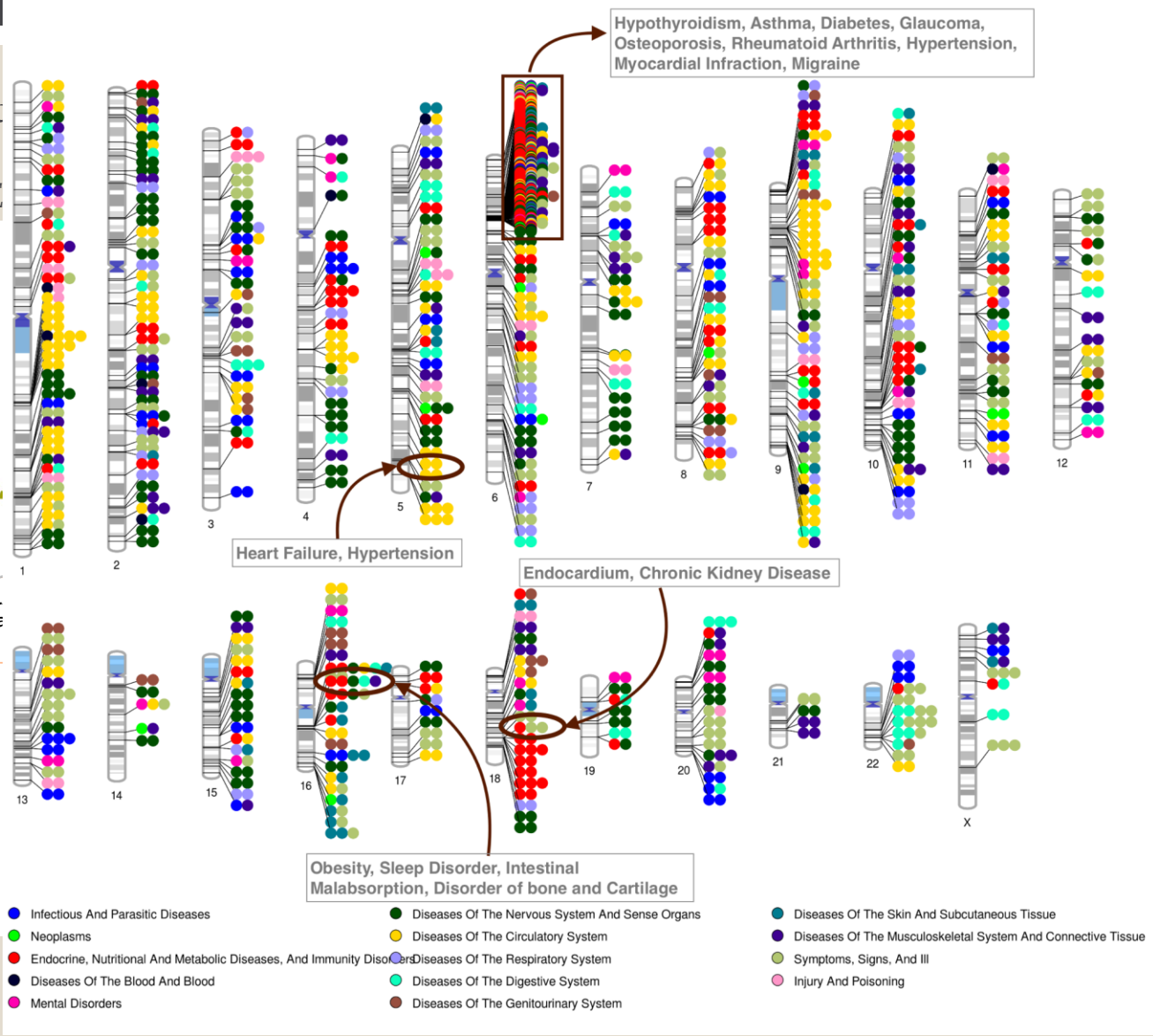
- Performed a comprehensive genome-wide PheWAS
 - 632,574 genetic variants, MAF > 1%
 - 38,662 EA
 - 541 ICD-9 derived case-control diagnoses and a series of clinical lab measures



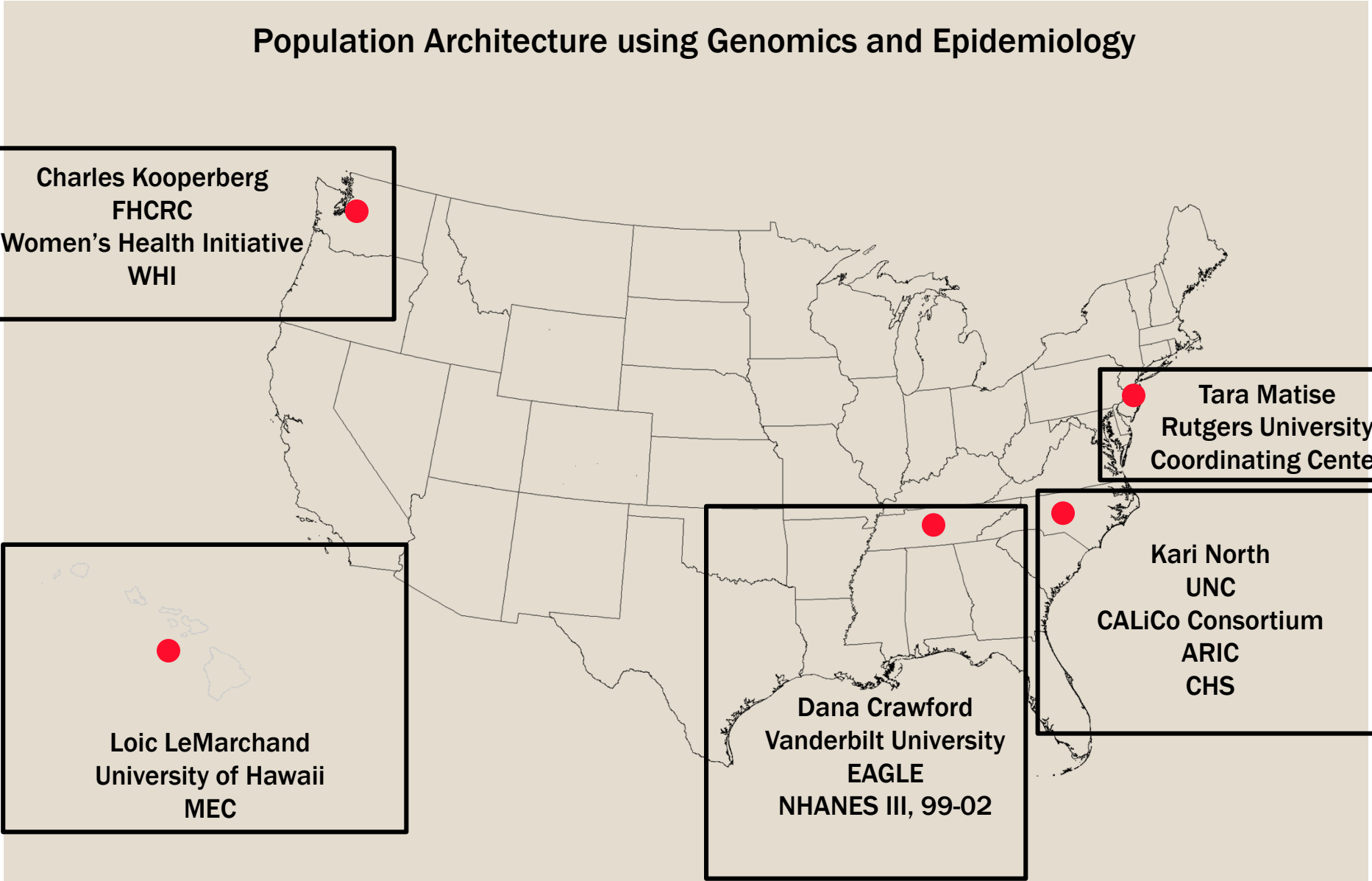
MegaPheWAS

- Performed
- 632,574
- 38,662 E

SNP: rs9273363 (*HLA*)
 Phenotype: Type I dia
 P-value: 1.36×10^{-77}



Population Architecture using Genomics and Epidemiology



Charles Kooperberg
FHCRC
Women's Health Initiative
WHI

Tara Matisse
Rutgers University
Coordinating Center

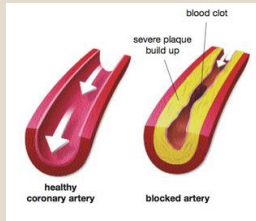
Loic LeMarchand
University of Hawaii
MEC

Dana Crawford
Vanderbilt University
EAGLE
NHANES III, 99-02

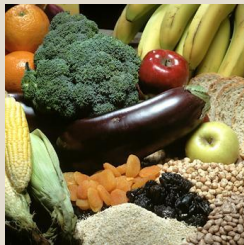
Kari North
UNC
CALiCo Consortium
ARIC
CHS

- Characterizing GWAS-identified variants in diverse population-based studies across
 - European Americans
 - African Americans
 - Hispanics/Mexican Americans
 - Asian/Pacific Islanders
- First epidemiological study based PheWAS

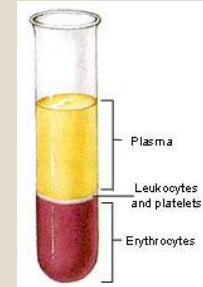
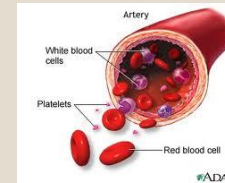
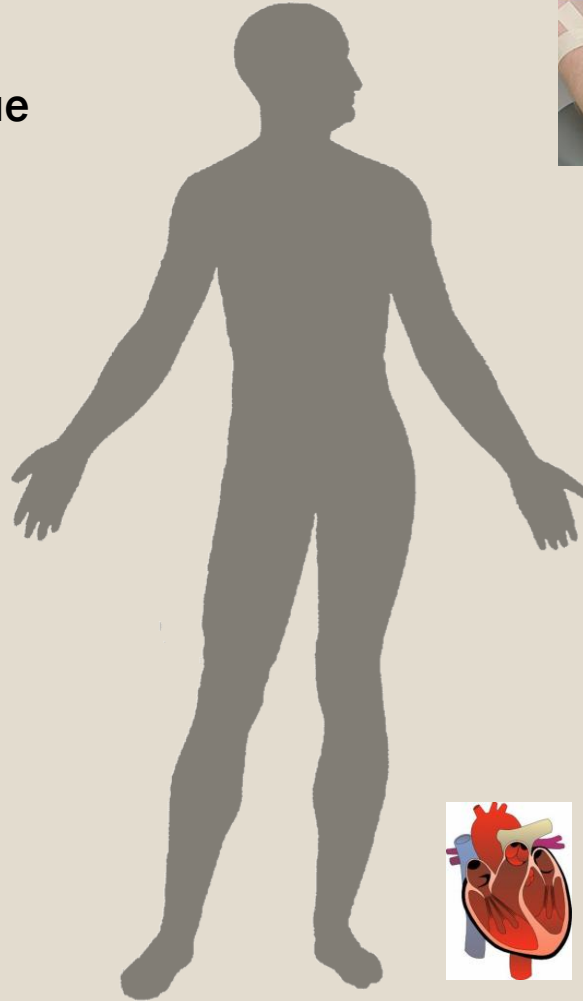
Stenosis and Plaque



Dietary Exposure



Disease Status
Prevalent/Incident
Longitudinal Information



Circulating Blood Cells

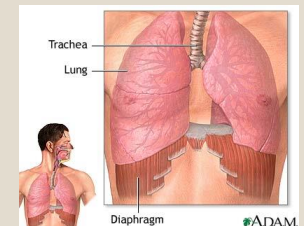
Hematocrit

Plasma Fibrinogen

C-reactive Protein



Cardiac Measurements



Lung Measurements

- **52 replicated previously reported associations**
 - NHGRI GWAS catalog matches
 - Showed our high-throughput approach was functional
- **26 represented phenotypes closely related to known associations**
- **33 represented novel-genotype phenotype associations**

APOE/APOC1/C1P1/C2/C4, rs4420638, Coded Allele A

ARIC LN LDL-C (mg/dl) (EA) -
2e-16

WHI Dietary Cholesterol (mg) (EA) +

ARIC LDL-C (mg/dl) (EA) -

ARIC LN1 Total cholesterol (mmol/l) (EA) -

CHS LN1 Baseline glucose (mg/dl) (EA) +
CHS Baseline glucose (mg/dl) (EA) +
WHI LN1 Dietary Cholesterol (mg) (EA) +
ARIC Dietary Cholesterol (mg) (EA) +



ARIC Total cholesterol (mmol/l) (EA) -
CHS LDL-C (EA) -
CHS LN1 LDL-C (EA) -
WHI High cholesterol requiring pills ever (EA) -

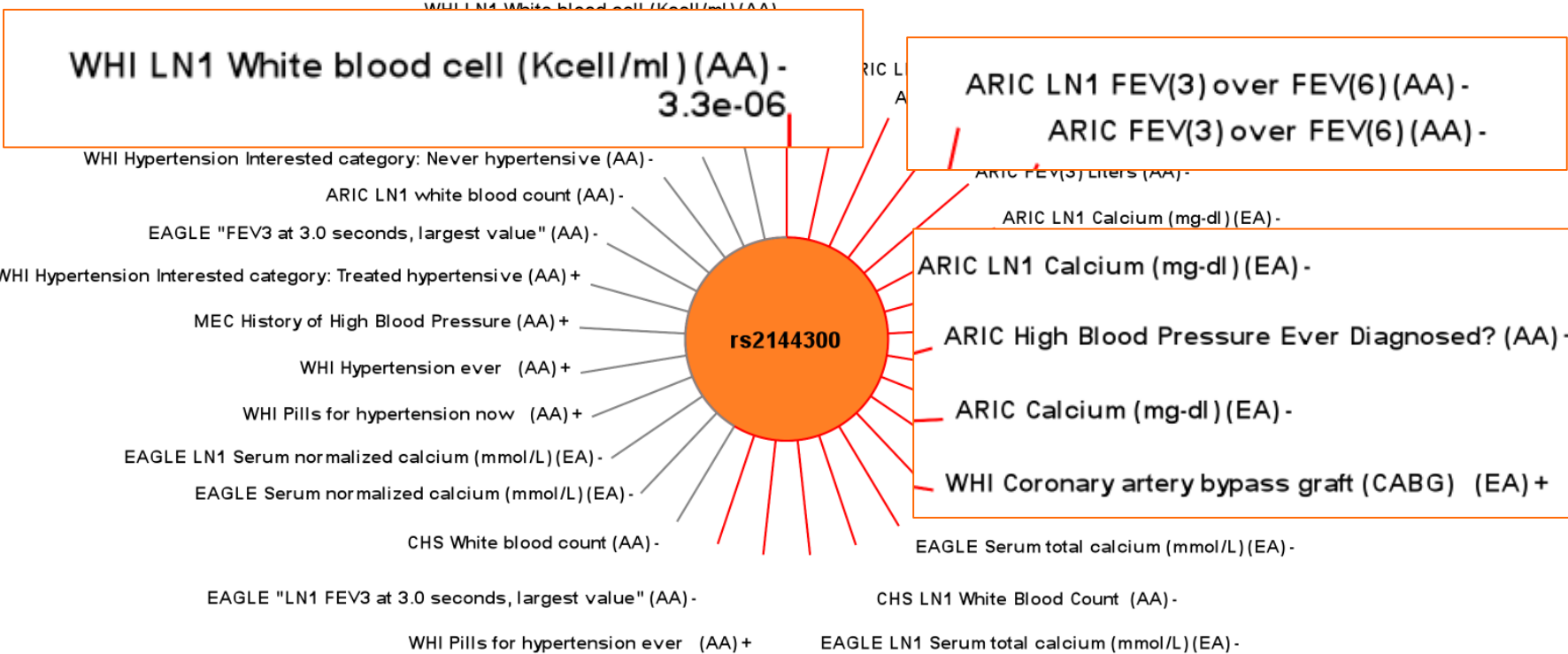
ARIC LN1 Dietary Cholesterol (mg) (EA) +

ARIC Derived glucose value (mg/dl) (EA) +

ARIC LN1 Derived glucose value (mg/dl) (EA) +

Replicating Result: LDL-C levels

GALNT2, rsRS2144300, coded Allele C



Previously associated with HDL-C Levels

FEV: Forced expired volume lung spirometry measurement

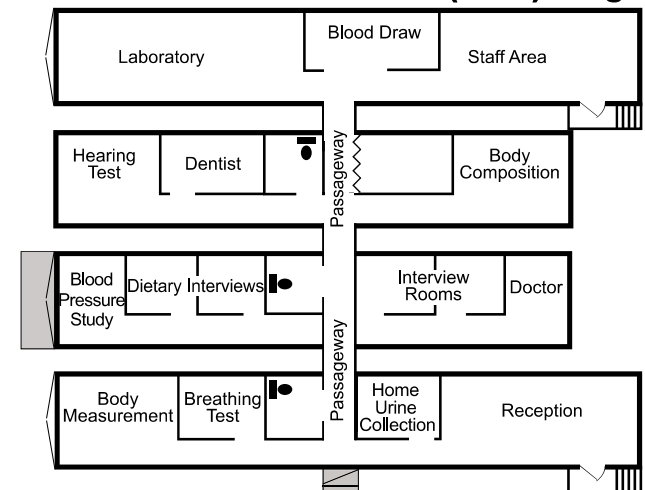
- Lipid trait SNPs
 - Age at menarche, EA
 - Thyroid/goiter, EA
 - Artery measurements, EA
 - Artery treatment, AA
 - Angina, EA
- Type 2 diabetes SNPs
 - Ever Smoked, EA
 - Hypertension, AA
- Inflammation SNPs
 - Carotene (dietary consumption)
- Myocardial infarction SNPs
 - Vitamin B12 (dietary consumption)
- Body mass index SNPs
 - Ever smoked, EA

NHANES PheWAS

- National Health and Nutrition Examination Surveys (NHANES)
 - Conducted by Centers for Disease Control and Prevention (CDC)
 - Health and nutritional habits
 - Via questionnaire, direct laboratory measures, and a physical exam



Mobile Examination Center (MEC) Diagram



- Genetic NHANES three surveys
 - NHANES III, 1999-2000, and 2001-2002
 - We merged 1999-2000 and 2001-2002 together
 - Three race-ethnicities
 - Non-Hispanic whites (n=6,634)
 - Non-Hispanic blacks (n=3,458)
 - Mexican Americans (n=3,950)
- 80 SNPs



NHANES PheWAS

Results for rs1800588

NHANES III: LN+1 Serumtriglycerides(mg/dL) (NHW) +
2.9e-05

NHANES 9902: LN+1 Folate, RBC (nmol/L RBC) (MA) -

rs1800588

NHANES III: LN+1 Vitamin E (ug/dL) (NHW) +

Previously associated with HDL-C Levels

Clinical Trials PheWAS

- ACTG Clinical Trial Data
 - On-treatment PheWAS
 - Smaller number of SNPs
 - More evidence for initial group SNP selection
 - Context-dependent sub-phenotypes
- What can we learn about genetic architecture and response to antiretroviral therapy drugs?
 - There is marked interindividual variability in HIV drug responses

Clinical Trials PheWAS

Immunology

CD4 T Cell Count

Metabolic

LDL

Triglycerides
(fasting)

Virology

HIV-1 RNA
(<200 copies/mL)

Pharmacology

Efavirenz

Atazanavir

6 variable, subject from A5202

Context-dependent 838 phenotypes
(Study week, ancestry, sex, ART)

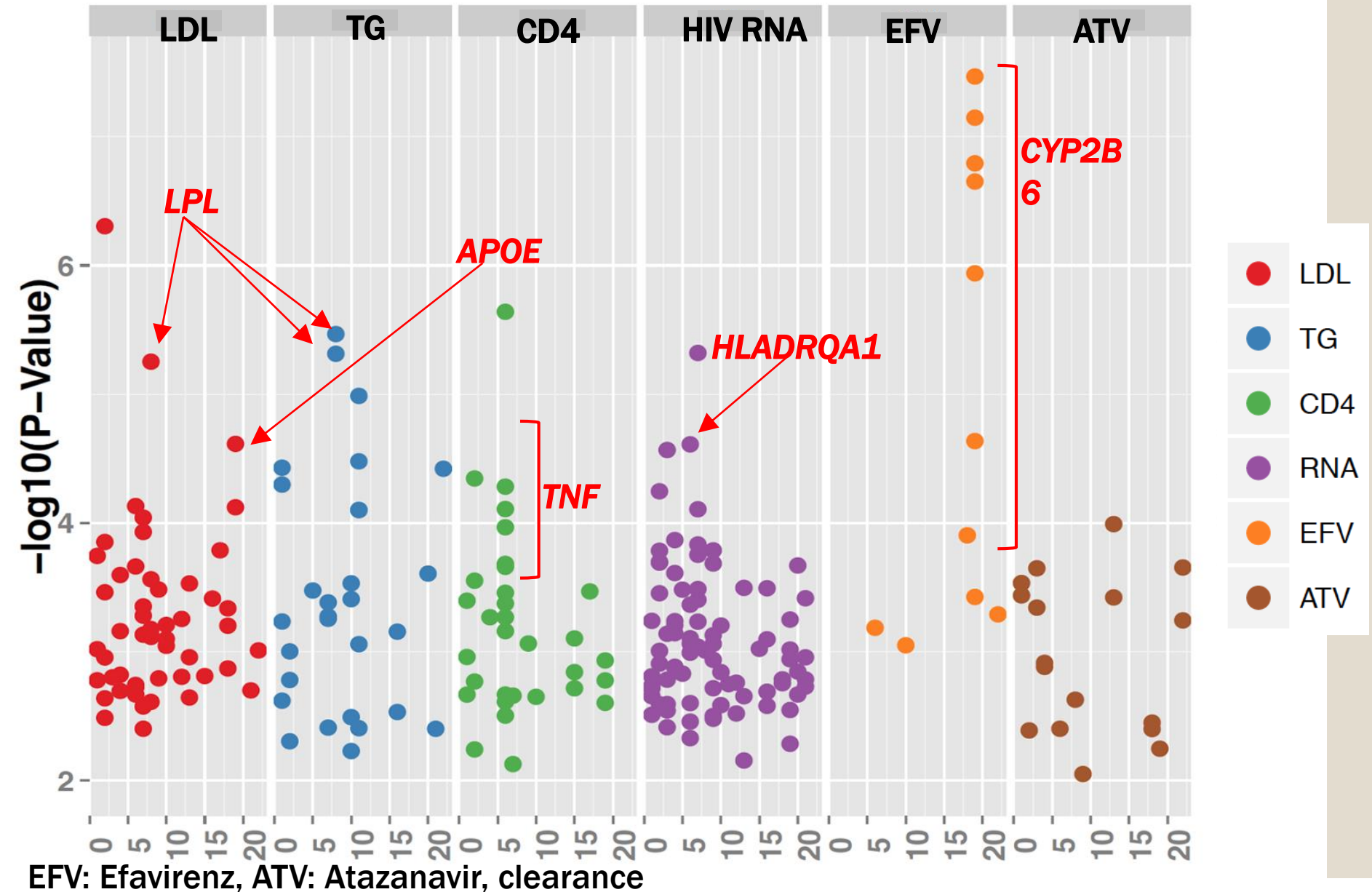
Example:

- Primary: ldl-week8
- Sub-phenotype: ldl-week8-AA-male-efv

Clinical Trials PheWAS

- Absorption, Distribution, Metabolism, and Excretion (ADME) genes
- SNPs selected for associations:
 - ~2000 annotated SNPs in drug response related genes from Pharmacogenomics Knowledgebase (PharmGKB)
 - 478 SNPs from NHGRI GWAS Catalog associated with lipid traits

PharmGKB Results



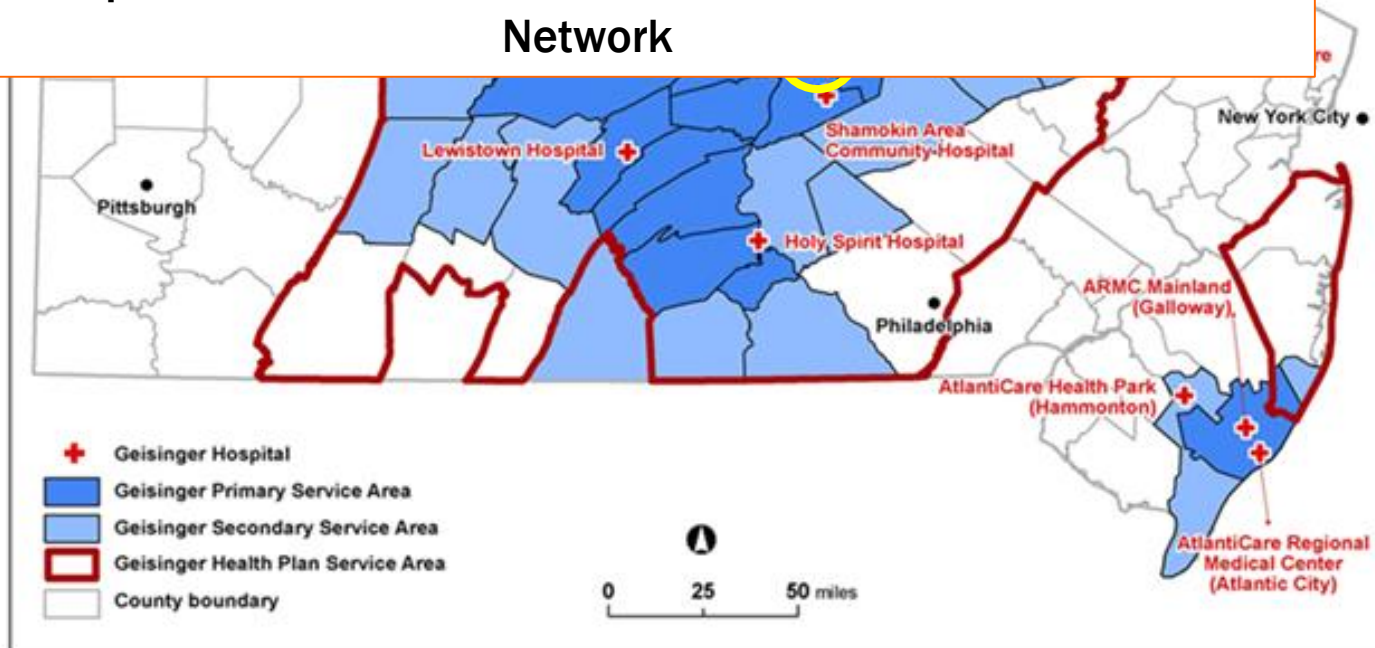
A Great Place for Further Research

■ Geisinger Health System

A network of ~78 primary and specialty clinic sites including 41 community-based primary care clinics

We have academic research including the department Biomedical and Translational Informatics

We are a part of the Precision Medicine Initiative and the eMERGE Network





- Geisinger Health System
- MyCode Community Health Initiative
 - A project with the goal of collecting whole exome sequencing data and whole genome array genotyping data on more than 250,000 individuals
 - Not adding in individuals based on any one disease type
- These data are linked to the EHR
- These data are de-identified for research
- *We are also contacting patients for return of results (ROR)*

Geisinger 76

- We are returning results for highly-actionable genetic variants
- Within 76 genes

Genomic Condition	Population Prevalence	Clinical Risk	Disease-altering Intervention
Hereditary Breast & Ovarian Cancer Syndrome	1 in 400	Early-onset breast, ovarian, & prostate cancers	Targeted screening & prophylactic surgical intervention
Familial Hypercholesterolemia	1 in 175	Early-onset coronary artery disease & stroke	Targeted screening & aggressive medical management
Lynch Syndrome	1 in 440	Early-onset colon & uterine cancers	Targeted screening & management of pre-cancerous changes
Total	> 1 in 100	Multiple cancers & cardiovascular diseases	Life-saving screening & intervention before development of disease

Life Cycle of the Data: An Opportunity

- Life cycle of the data - an opportunity

Recontact

Research and Discovery

Addition to Clinical
Data Collection

Genetic Data



EHR Data

Clinical Informatics
Bioinformatics
Genetics
Genetic Epidemiology
Programming
Genetic Counseling
Statistics
Big Data analysis

Clinical Data Collection

Actionable Data

Intervention

Decisions

Provider Feedback

DNA Samples
Collected for
Biorepository and
Sequenced

- Pendergrass Lab *www.pendergrasslab.com*
- Chris Bauer
- Navya Josyula
- Elliott Cha
- Crawford Lab
- Dr. Dana Crawford
- Dr. David Haas
- Ritchie Lab *www.ritchielab.psu.edu*
- Dr. Marylyn Ritchie
- Anurag Verma
- Shefali Verma
- Anna Basil

We are seeking undergraduate and graduate interns for next summer

Also seeking postdoctoral fellows!

Geisinger

