PENDERGRASS LAB

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

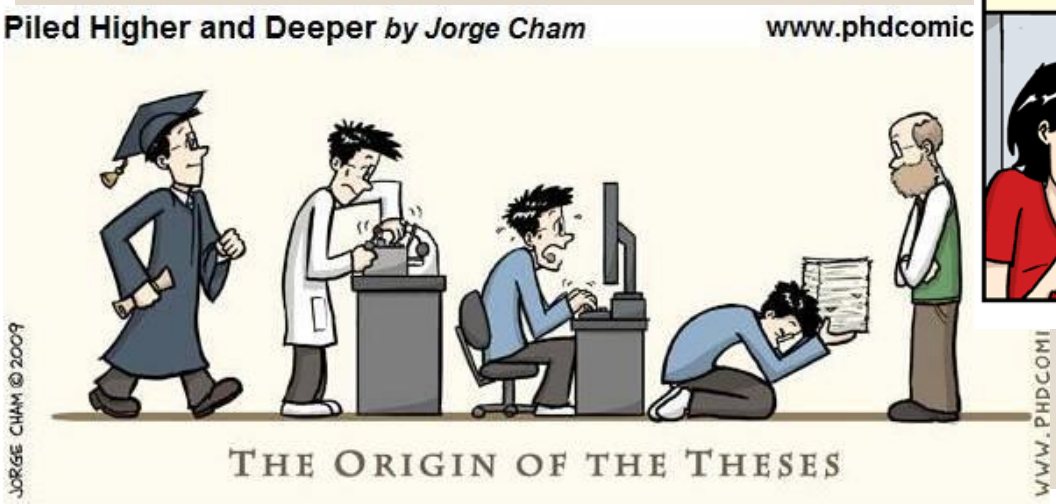
Sarah A. Pendergrass PhD, MS
Assistant Professor
Biomedical and Translational Informatics
Geisinger Health System

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
- Emerging world of bioinformatics and using these tools to study disease
- Genetic epidemiology and bioinformatics – how can we leverage complex multifaceted data for discovery??



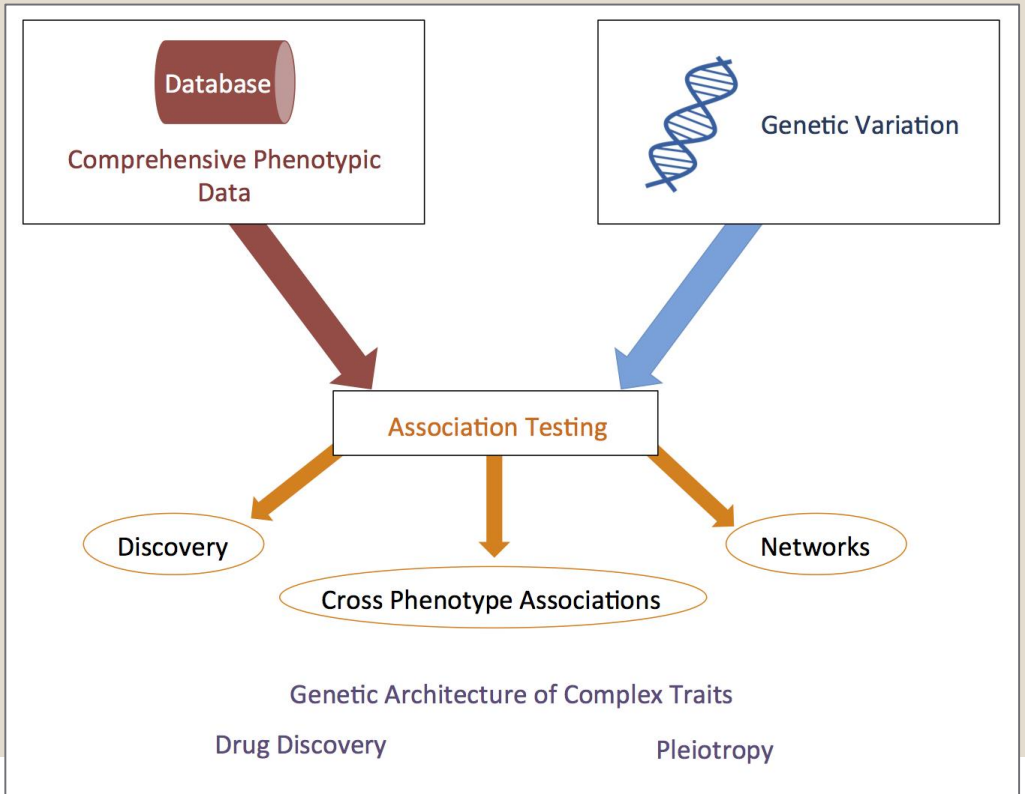
PheWAS

- 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

PheWAS

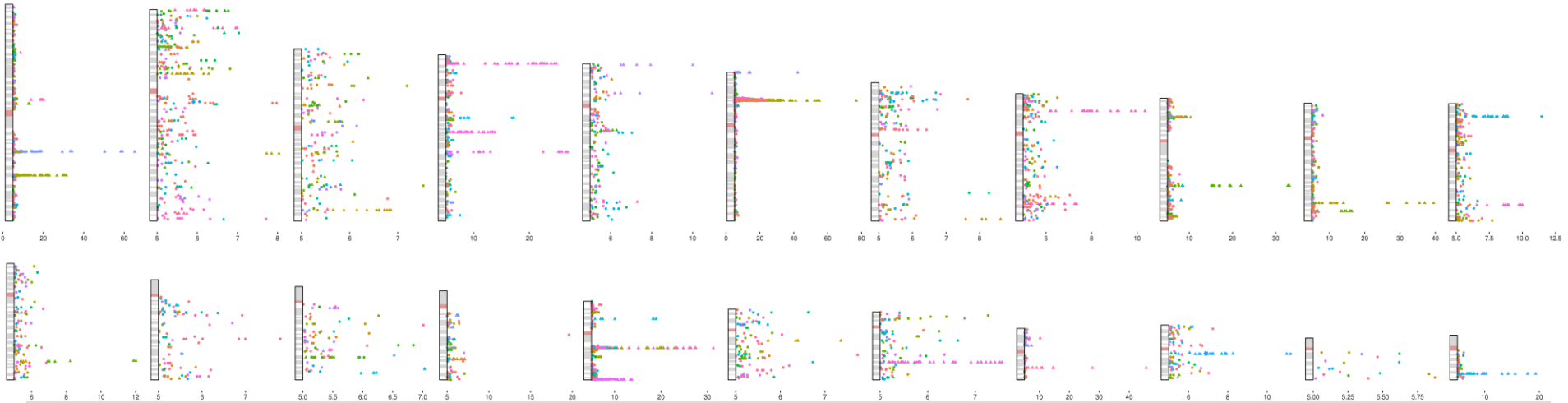
- Phenome Wide Association Studies
  - Datasets with thousands of collected phenotypes
  - Unlikely that time would be spent to carefully investigate each phenotype for use
  - Hypothesis generation
  - Can we identify SNPs and/or phenotypes for further research?

- Also exploring cross-phenotype associations
  - Exploring the relationship between associations for the same genetic variant and multiple phenotypes



MegaPheWAS

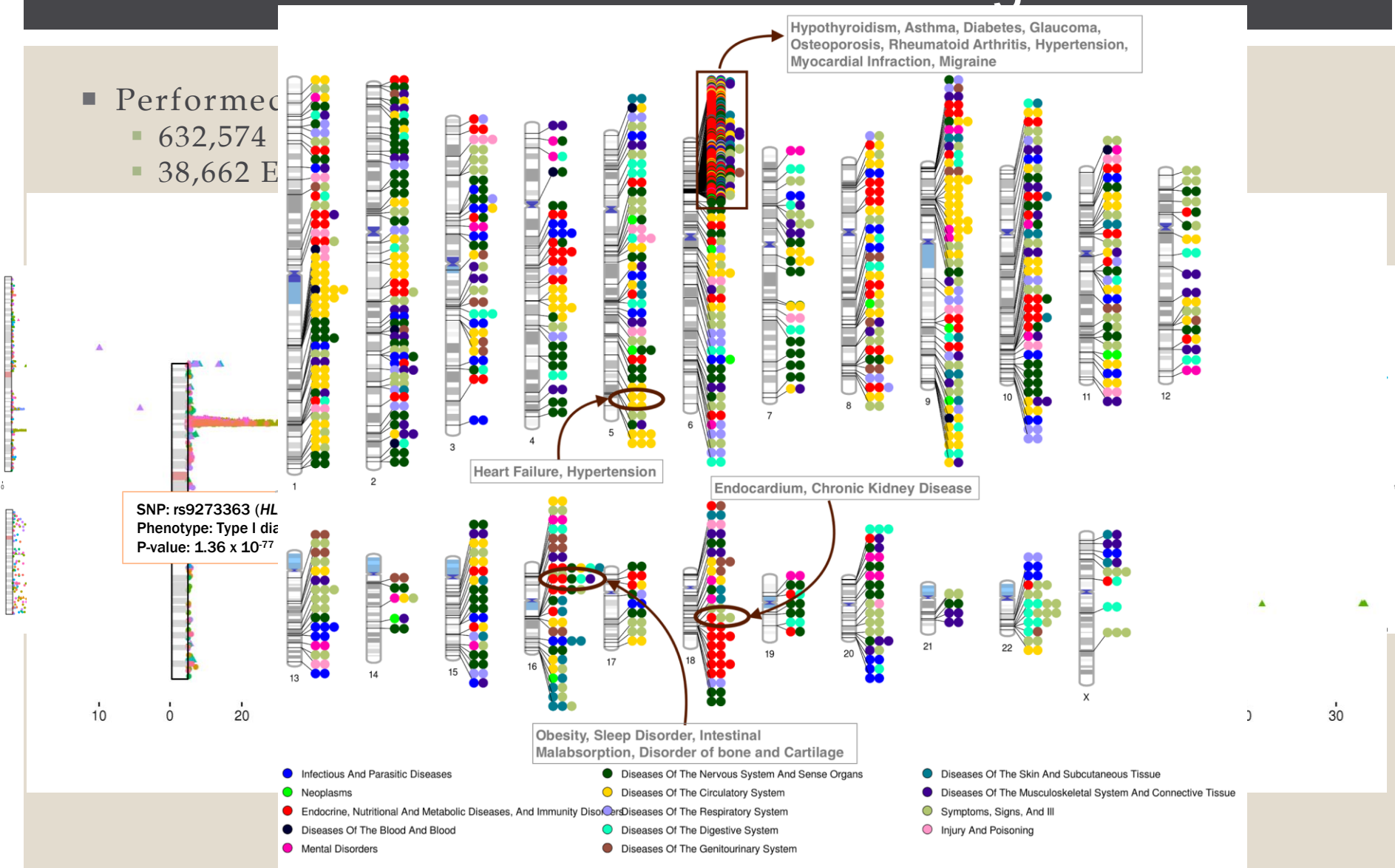
- Performed a comprehensive genome-wide PheWAS
  - 632,574 genetic variants, MAF > 1%
  - 38,662 EA
  - 541 ICD-9 derived case-control diagnoses and a series of clinical lab measures



Anurag Verma, Anastasia Lucas

MegaPheWAS

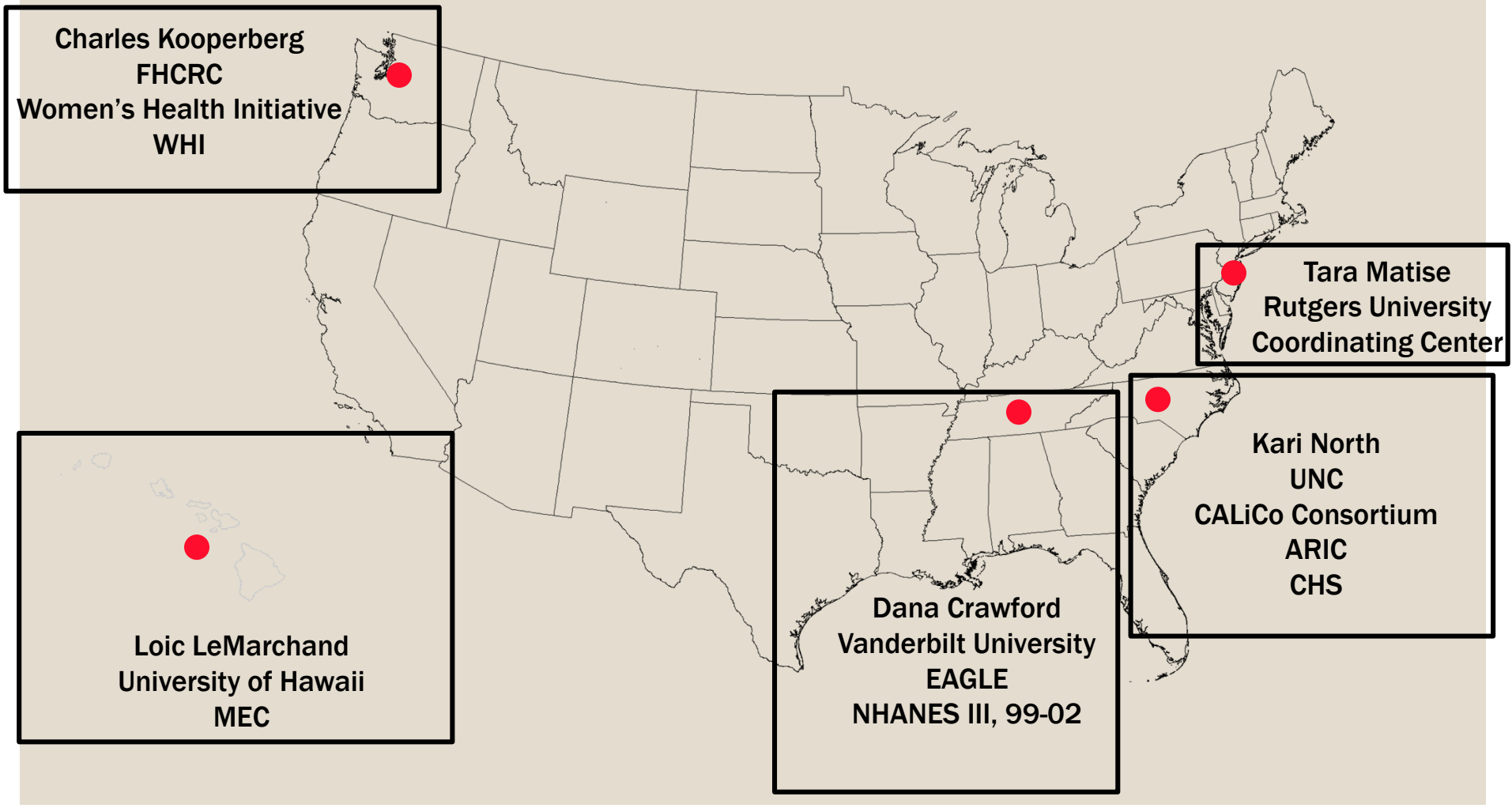
- Performed
  - 632,574
  - 38,662 E



Anurag Verma, Anastasia Lucas

PAGE PheWAS

Population Architecture using Genomics and Epidemiology

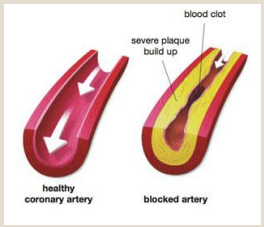


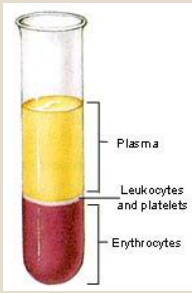
PAGE PheWAS

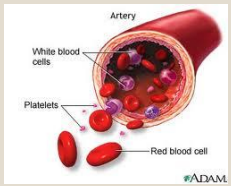
- Characterizing GWAS-identified variants in diverse population-based studies across
  - European Americans
  - African Americans
  - Hispanics/Mexican Americans
  - Asian/Pacific Islanders
- First epidemiological study based PheWAS

PAGE PheWAS

Stenosis and Plaque







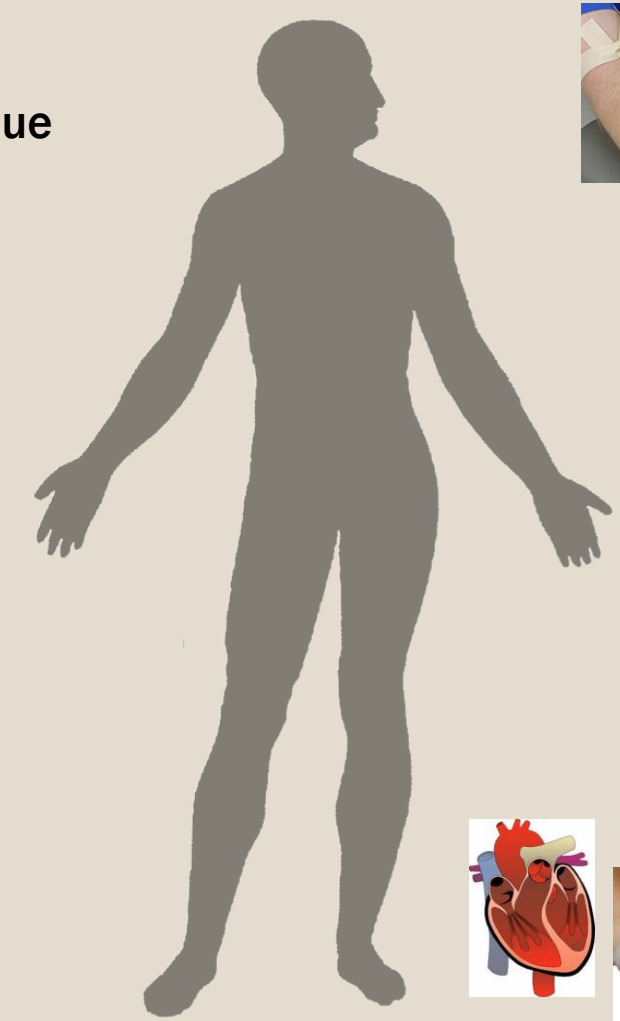
Circulating Blood Cells

Hematocrit

Plasma Fibrinogen

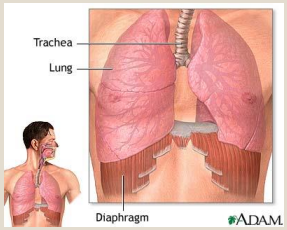
C-reactive Protein

Dietary Exposure



Cardiac Measurements

Disease Status
Prevalent/Incident
Longitudinal Information

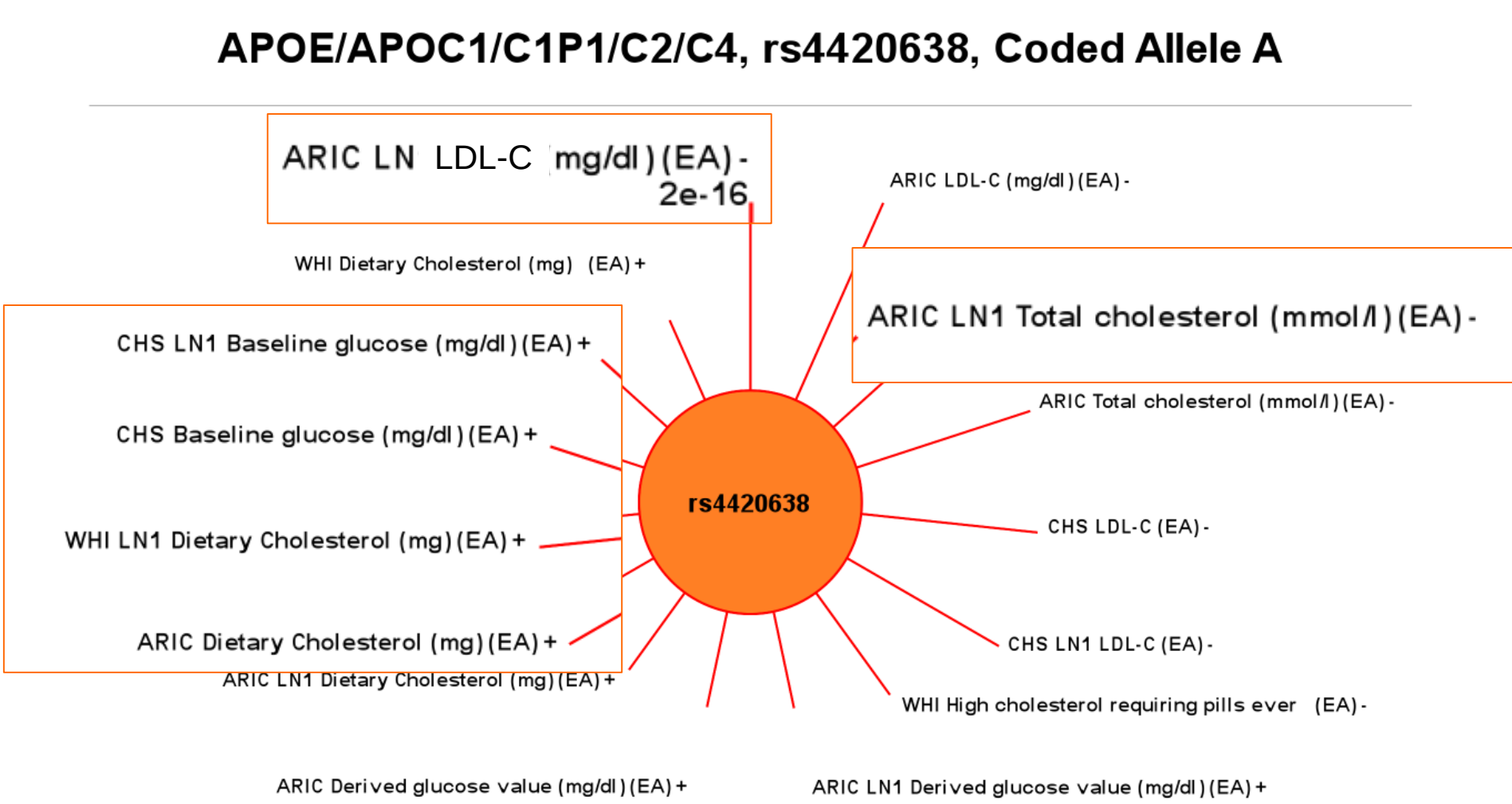


Lung Measurements

PAGE PheWAS

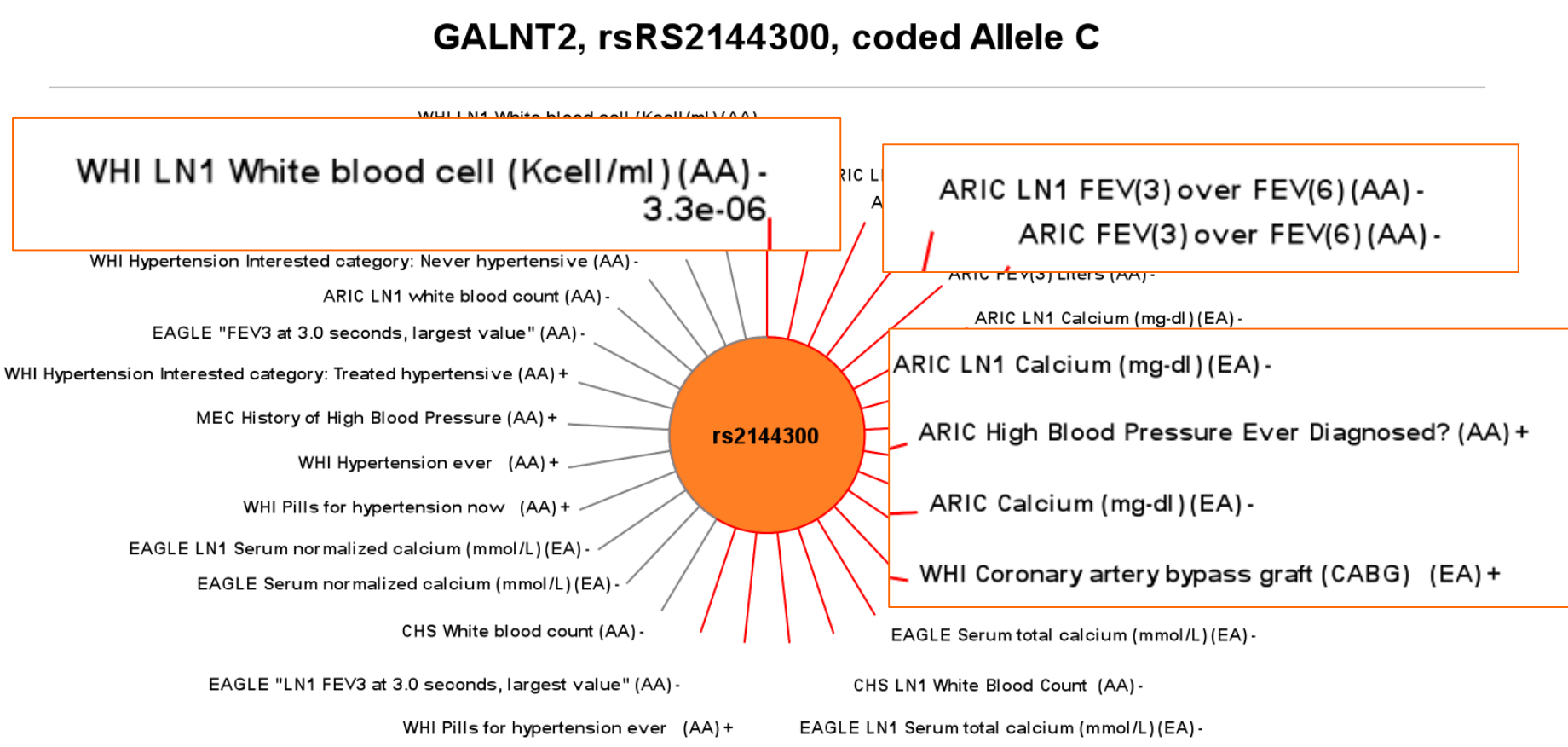
- **52 replicated previously reported associations**
  - NHGRI GWAS catalog matches
  - Showed our high-throughput approach was functional
- **26 represented phenotypes closely related to known associations**
- **33 represented novel-genotype phenotype associations**

PAGE PheWAS



Replicating Result: LDL-C levels

PAGE PheWAS



Previously associated with HDL-C Levels

FEV: Forced expired volume lung spirometry measurement

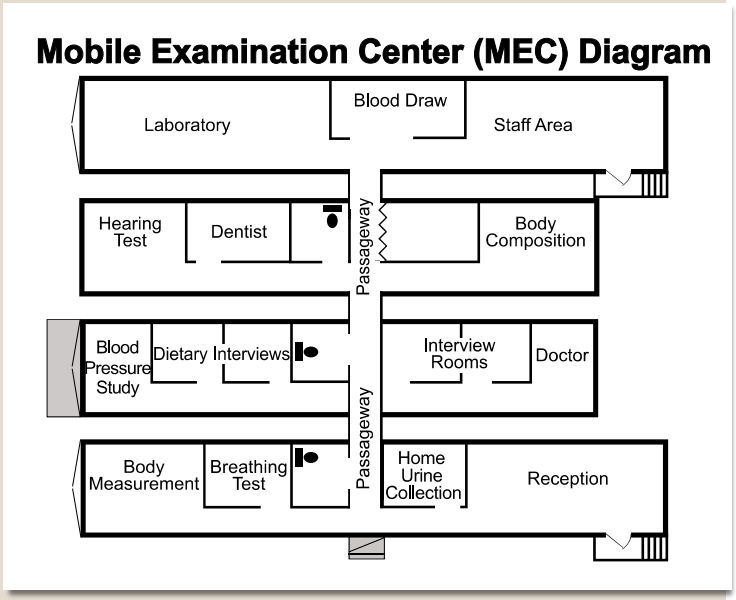
PAGE PheWAS

- 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

Phenome-wide association study (PheWAS) for detection of pleiotropy within the Population Architecture using Genomics and Epidemiology (PAGE) Network. Pendergrass et al., *PLoS Genetics*, 2013

NHANES PheWAS

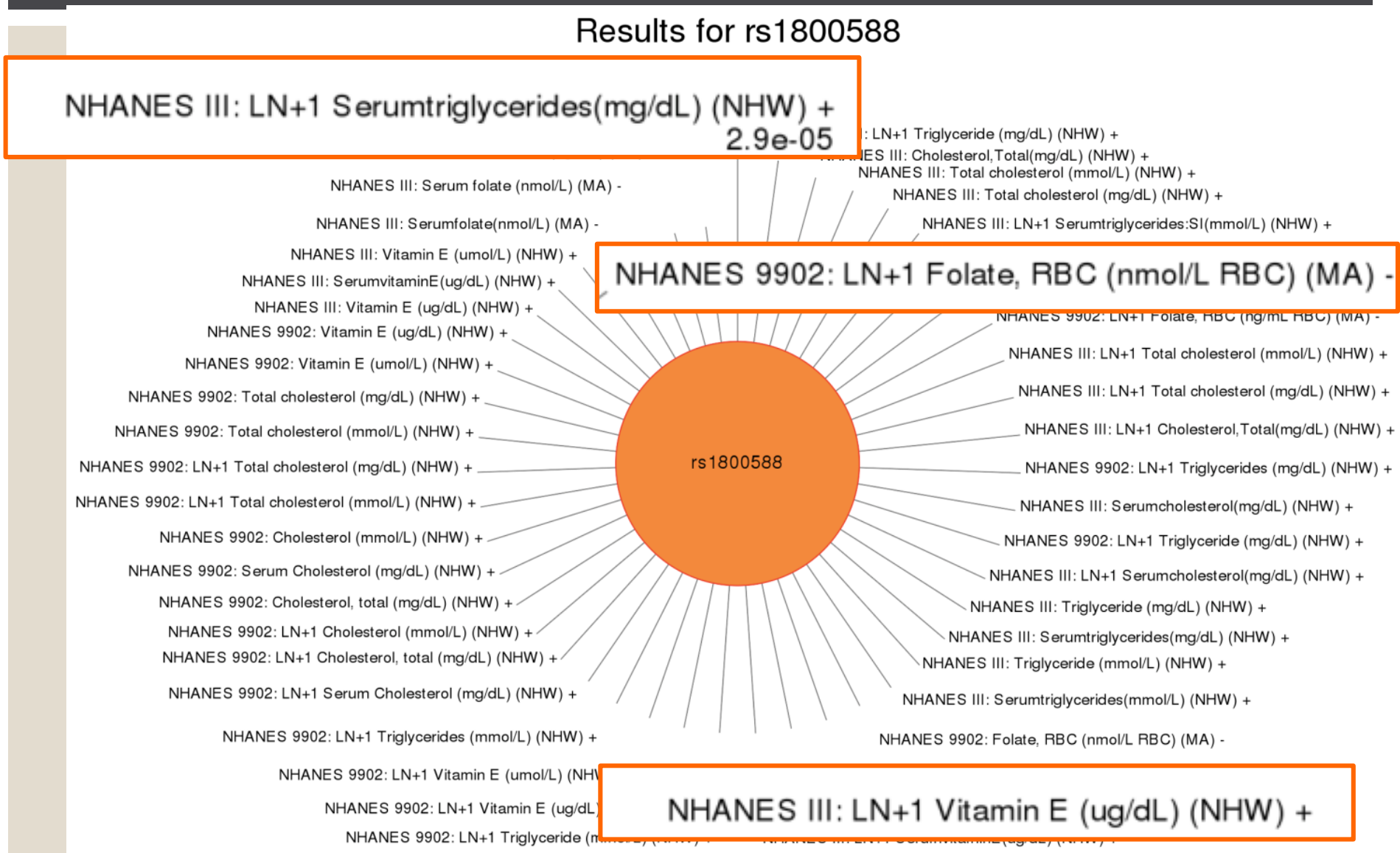
- National Health and Nutrition Examination Surveys (NHANES)
  - Conducted by Centers for Disease Control and Prevention (CDC)
  - Health and nutritional habits
  - Via questionnaire, direct laboratory measures, and a physical exam



NHANES PheWAS

- 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

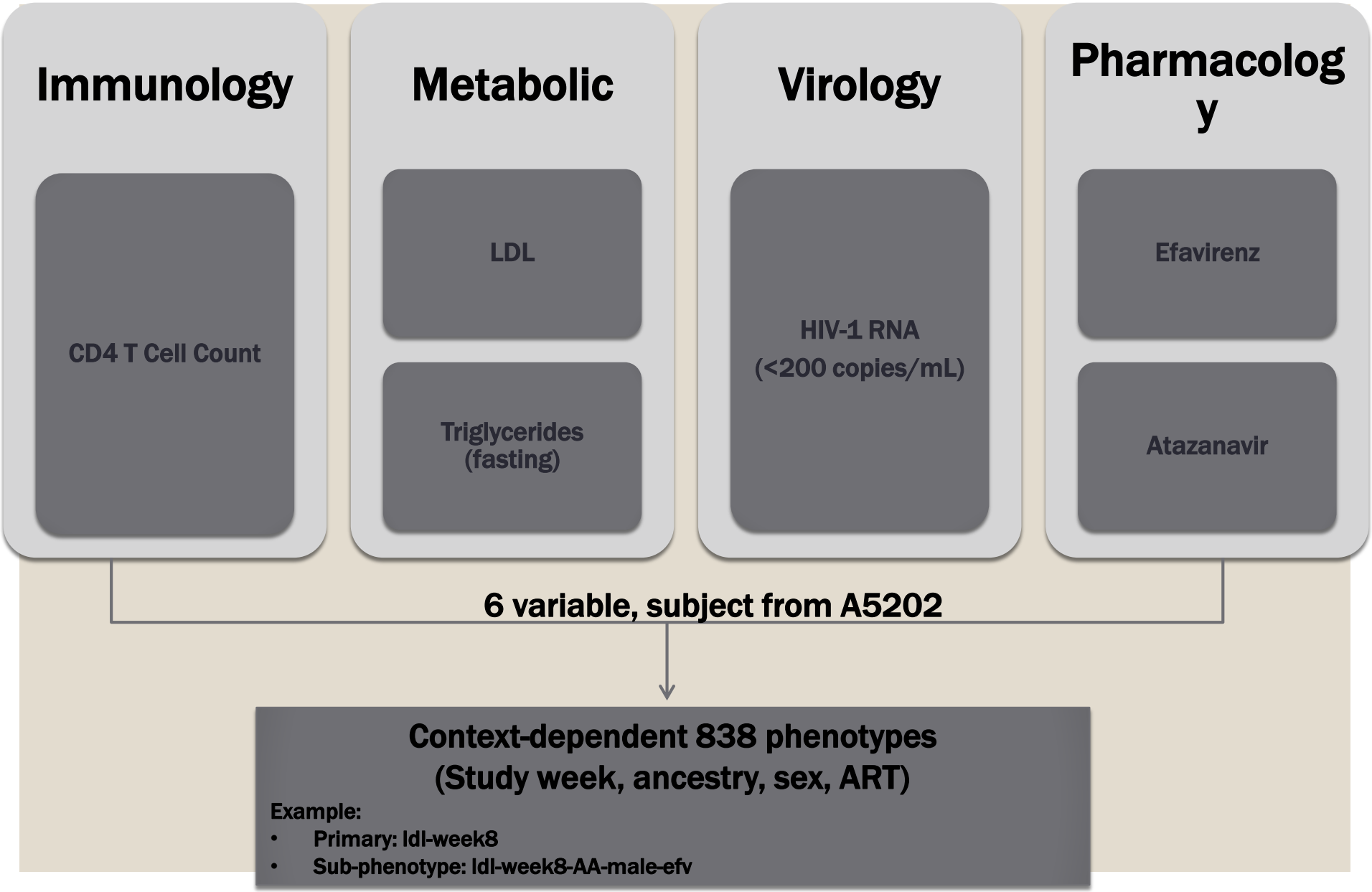


Previously associated with HDL-C Levels

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

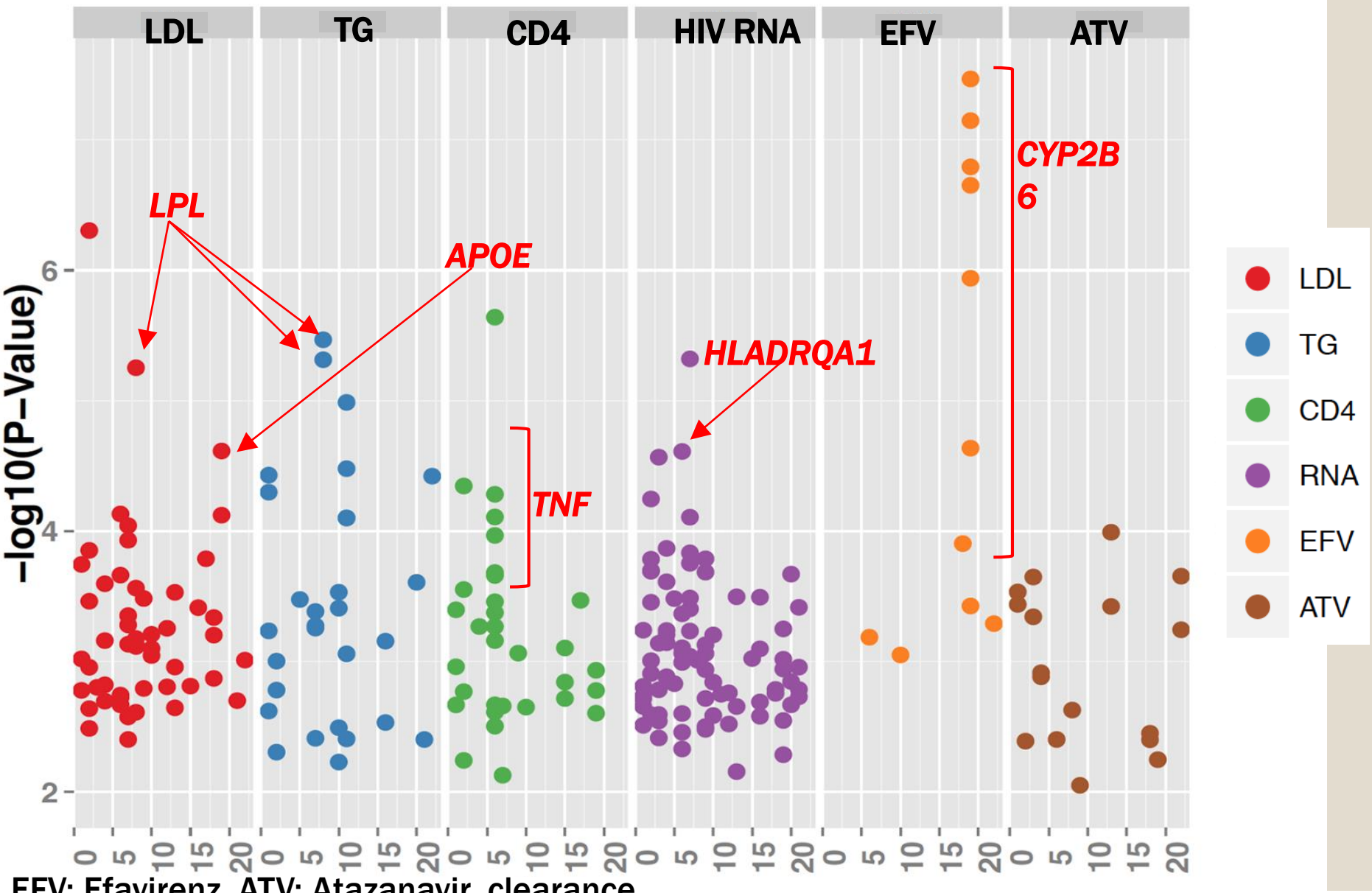
Clinical Trials PheWAS



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



EFV: Efavirenz, ATV: Atazanavir, clearance

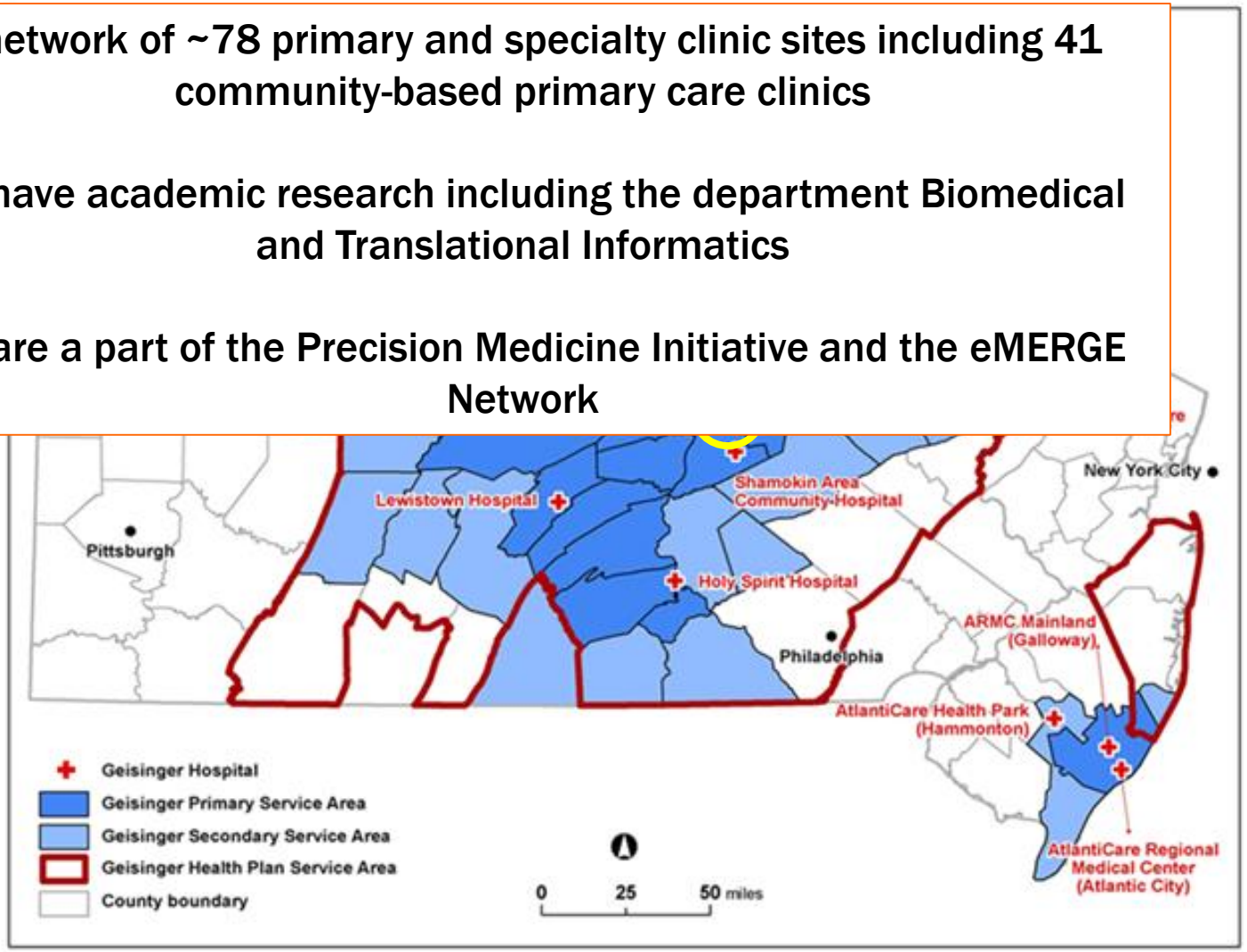
A Great Place for Further Research

- 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



MyCode

- 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)*

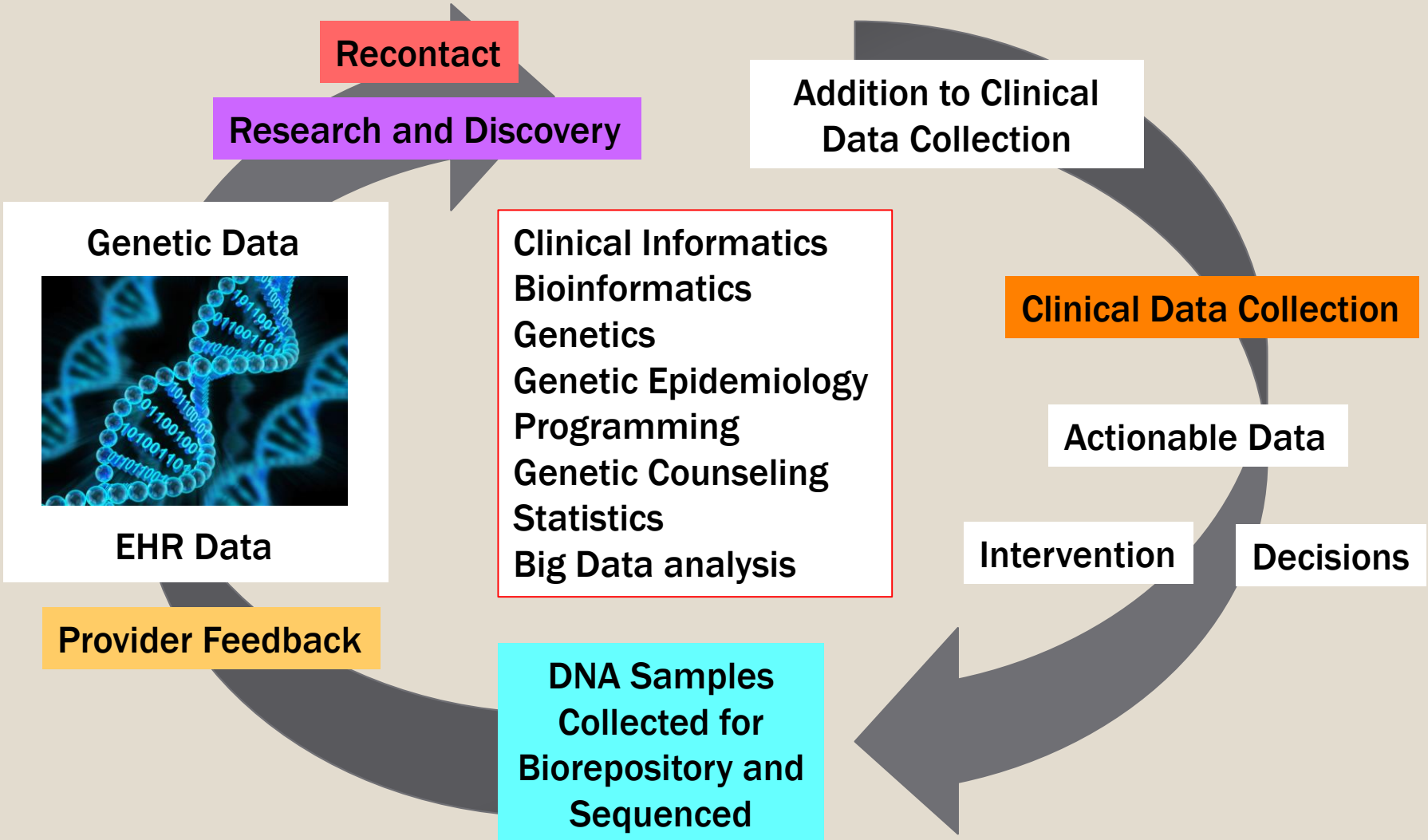
Geisinger 76

- 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



PENDERGRASS LAB

Acknowledgements

- **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 Verm
- Shefali Verma
- Anna Basil

We are seeking undergraduate and graduate interns for next summer
Also seeking postdoctoral fellows!

Geisinger

mycode