# Inclusive study design at every scale improves genomic analysis and clinical application for everyone.

Chani Hodonsky, PhD, MPH
Center for Public Health Genomics
University of Virginia
March 17, 2020

I have no conflict of interest to disclose

*Please feel free to email me* with any questions, or for links to any of the papers I mention:

ch2um@virginia.edu

# Background: CAD epidemiology

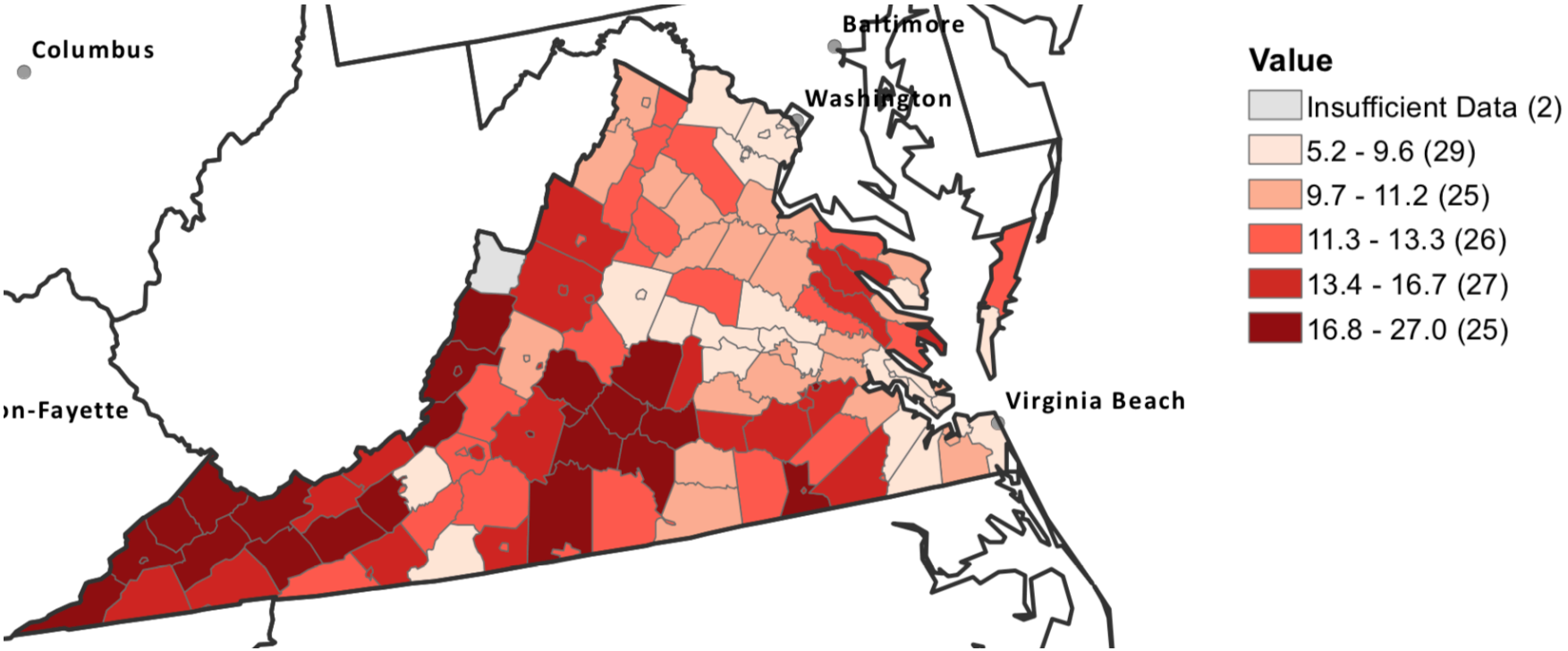
CAD is *one type of heart disease*

- Diagnosis often occurs at first heart attack (AKA “myocardial infarction”, ~600,000 events per year)

Risk factors

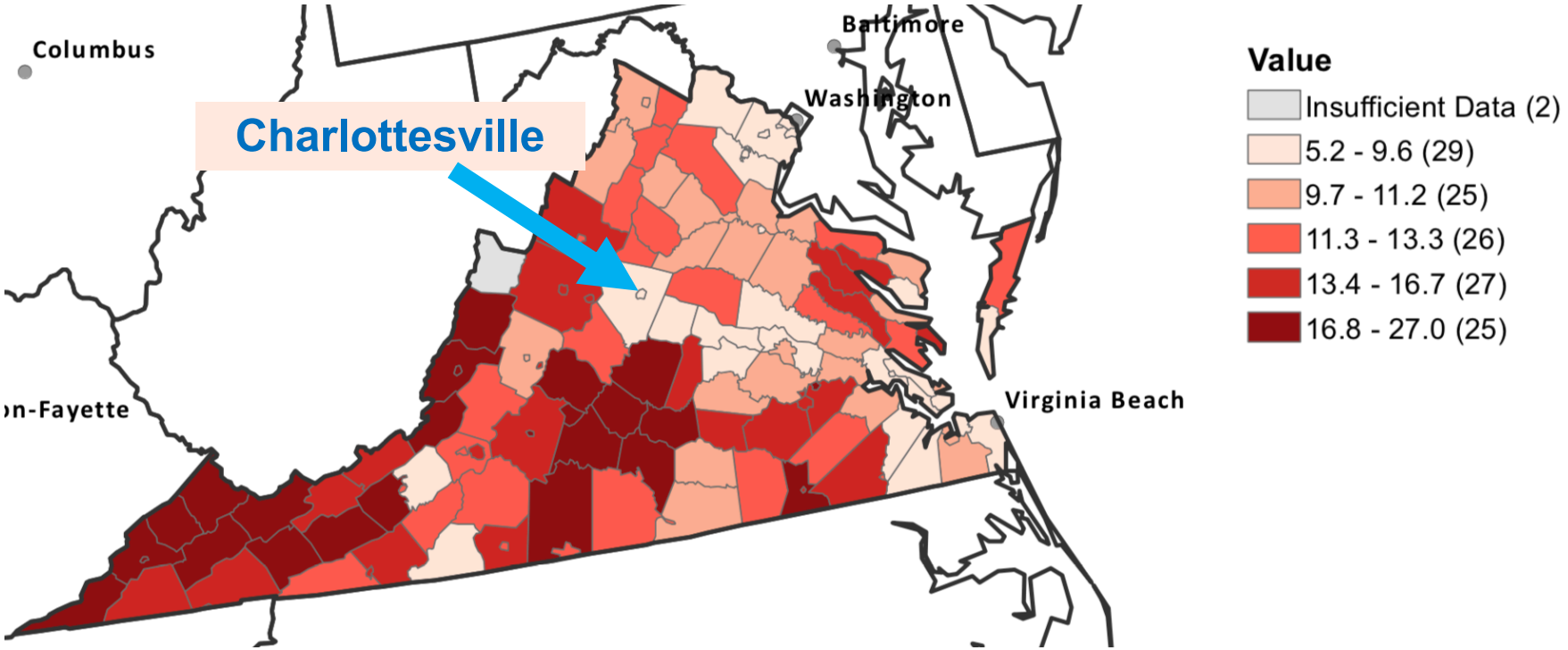
- Overweight
- Smoking
- Diet
- Sedentary behavior
- Type II Diabetes
- **Family history**

# Population estimates of MI…



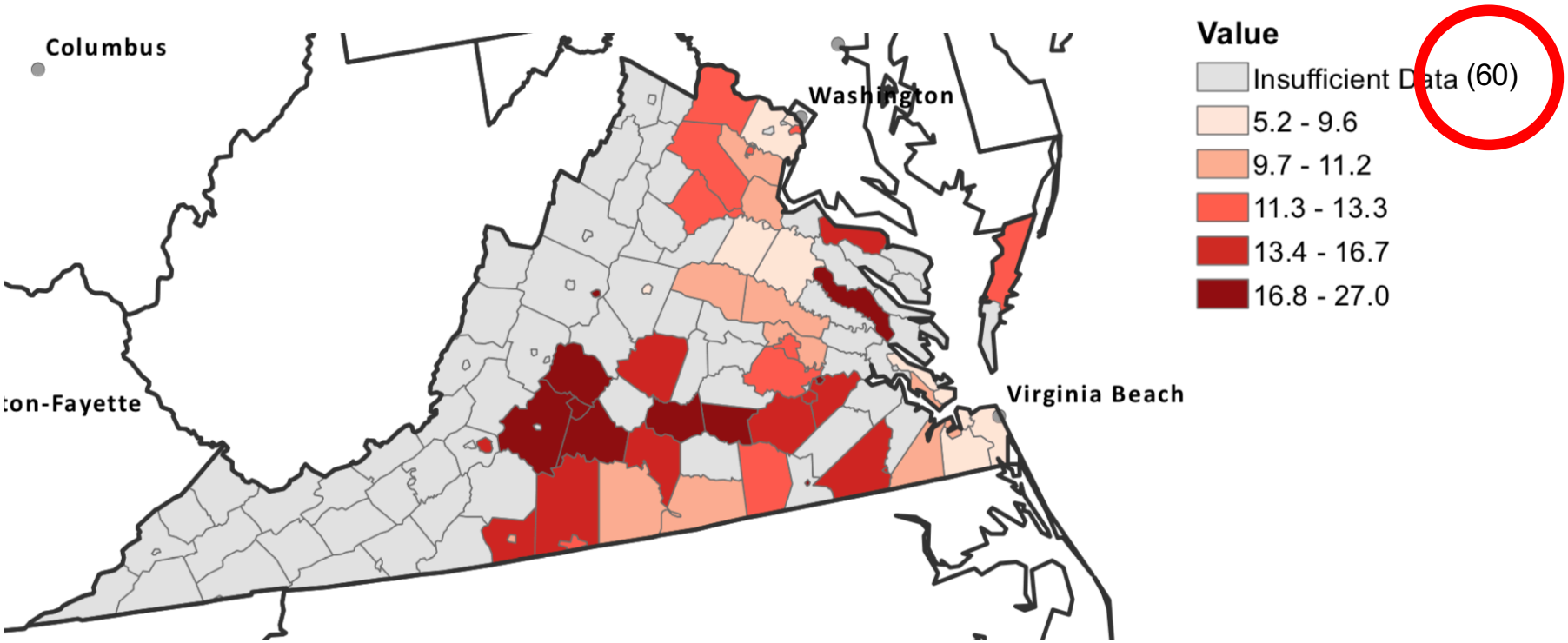

**Incidence of MI in whites per 1000 PY**

# Population estimates of MI…



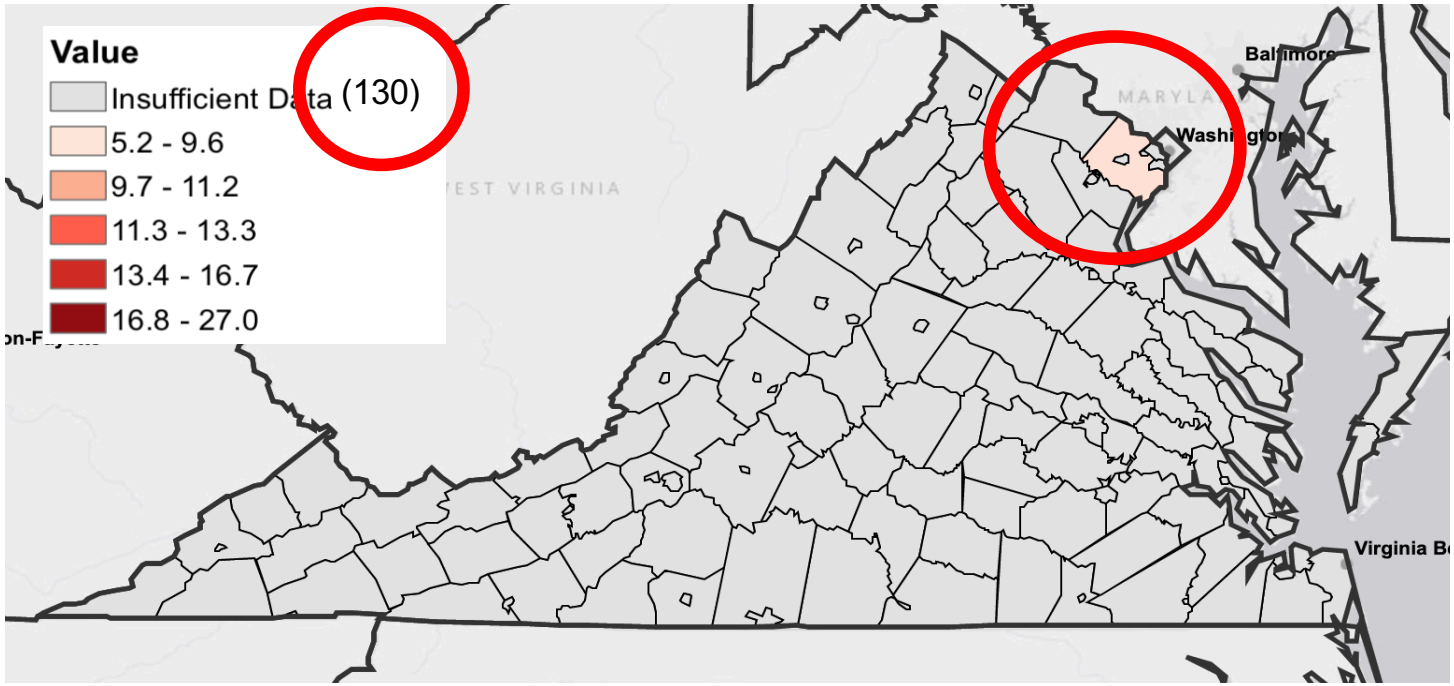

**Incidence of MI in whites per 1000 PY**

# …not as easy as we may think.




**Incidence of MI in African Americans per 1000 PY**

# *…particularly* for underrepresented populations




**Incidence of MI in Hispanics/Latinos per 1000 PY**

Luckily, we expect the same genes to play a role

in atherosclerosis (and CAD/MI)

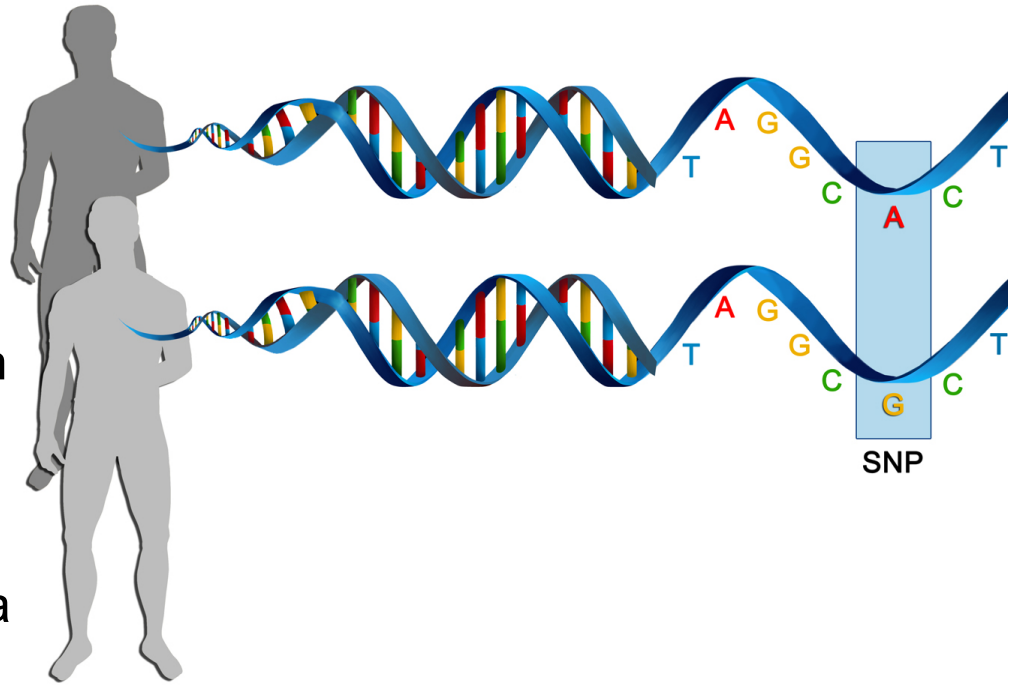
in everyone!

# Review: Genes vs genetic variants

***DNA*** refers to the >3 billion nucleotides (A/C/G/T) that comprise an individual's genetic make up. The sequence is the same in every cell.

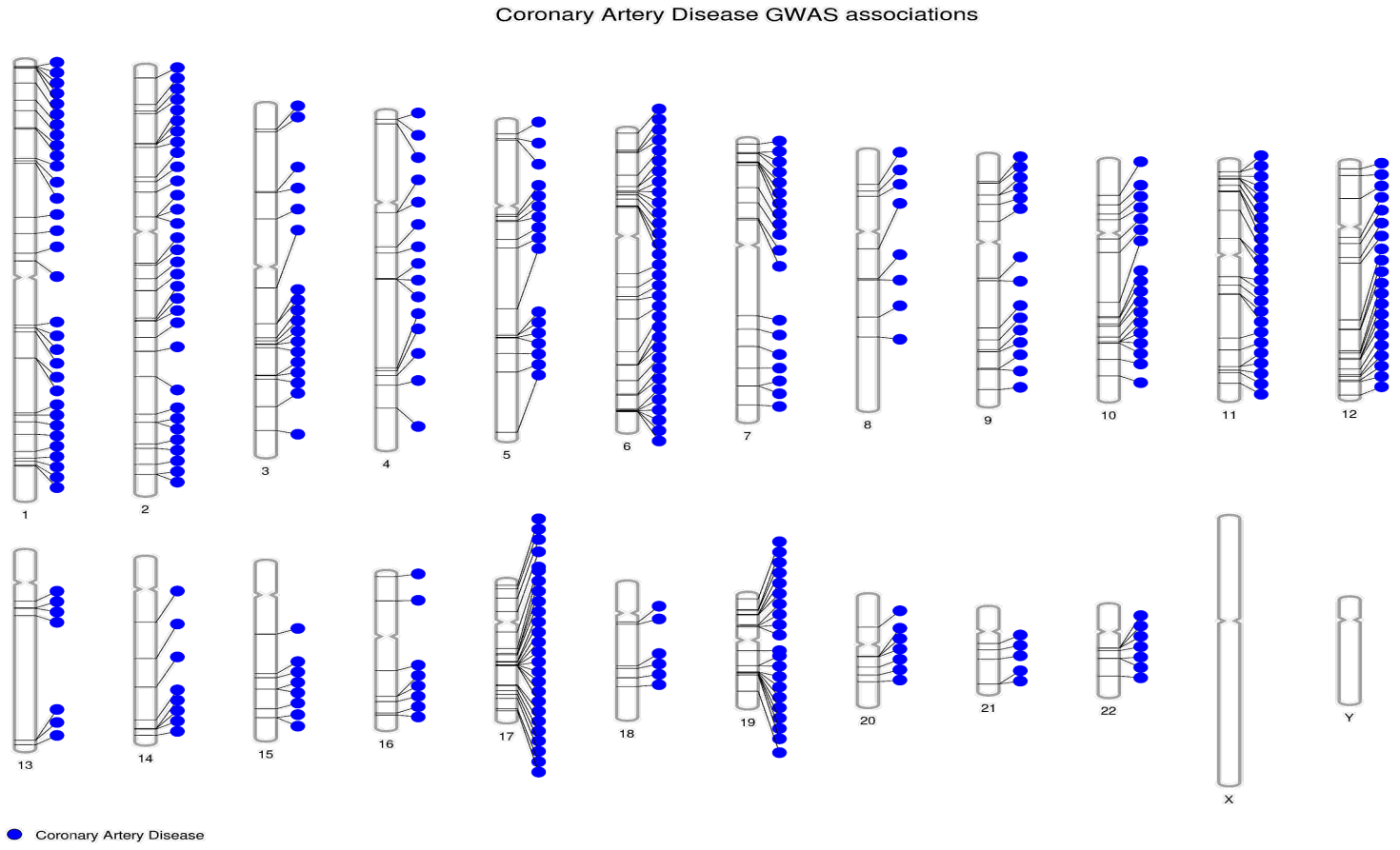
A ***gene*** refers to a specific region of DNA that contains information that can be transcribed into RNA, which may eventually be translated into a protein.

A ***genetic variant*** (sometimes called a SNP) is either one or a group of nucleotides that differs by individual.




http://atlasofscience.org/single-nucleotide-polymorphisms-as-genomic-markers-for-high-throughput-pharmacogenomic-studies/

# CAD Genome-wide Association Studies (GWAS)



# Eurocentric genomics has long been an issue

CLOCKWISE FROM TOP LEFT: O. STREWE/LONELY PLANET IMAGES; A. DISSANAYAKE/LONELY PLANET IMAGES; K. COOLE/LONELY PLANET IMAGES; K. CALVO VIA AP IMAGES; T. VOETEN/PANOS; R. I'ANSON/LONELY PLANET IMAGES; J. HAGLUND/LONELY PLANET IMAGES; P. ADAMS/GETTY IMAGES

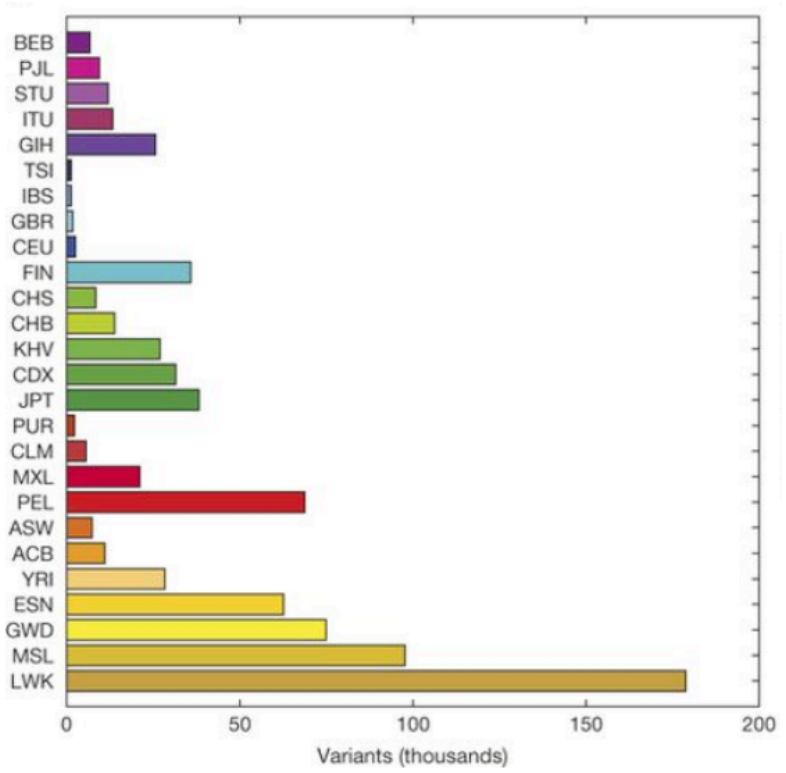
## Genomics for the world

Medical genomics has focused almost entirely on those of European descent. Other ethnic groups must be studied to ensure that more people benefit, say **Carlos D. Bustamante**, **Esteban González Burchard** and **Francisco M. De La Vega**.

- In 2011 Bustamante, González Burchard, & De La Vega published an editorial in Nature calling for increased representation
- They pointed out that findings cannot be equitably applied in translational research unless they are discovered in populations representing all of humanity

# Study population homogeneity is problematic

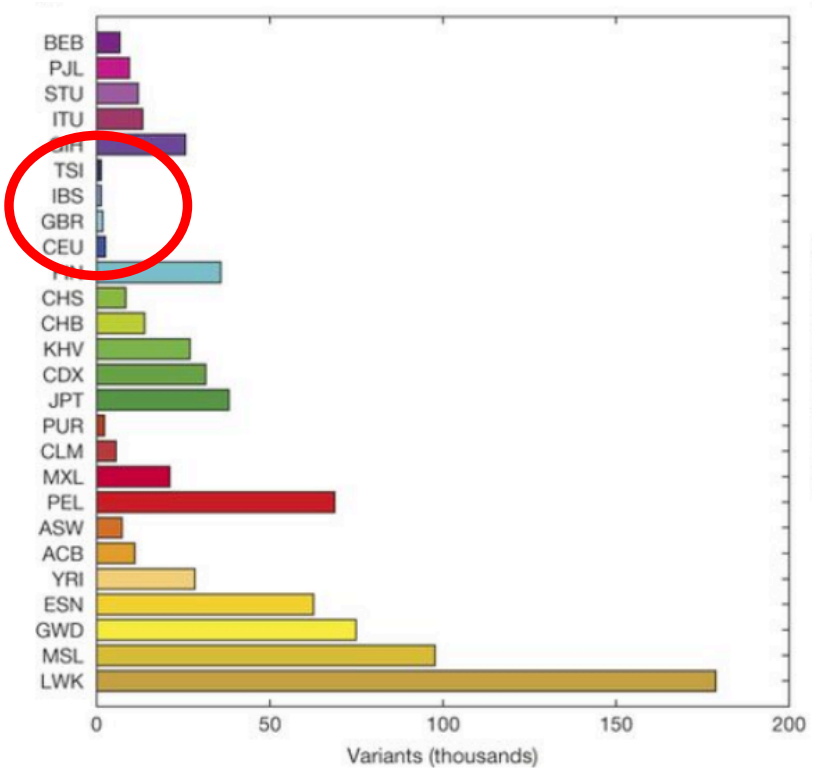
- The likelihood of two variants being inherited together differs across populations
- The large majority of human genetic variation is rare




1000G Consortium, 2015

# Study population homogeneity is problematic

- The likelihood of two variants being inherited together differs across populations
- The large majority of human genetic variation is rare
- Variants common in one ancestral population may be rare in others
- Tools developed for only one population have limited application to other ancestry groups




1000G Consortium, 2015

# Translational applications based on exclusive design have consequences

MENU nature genetics

Perspective | Published: 29 March 2019

## Clinical use of current polygenic risk scores may exacerbate health disparities

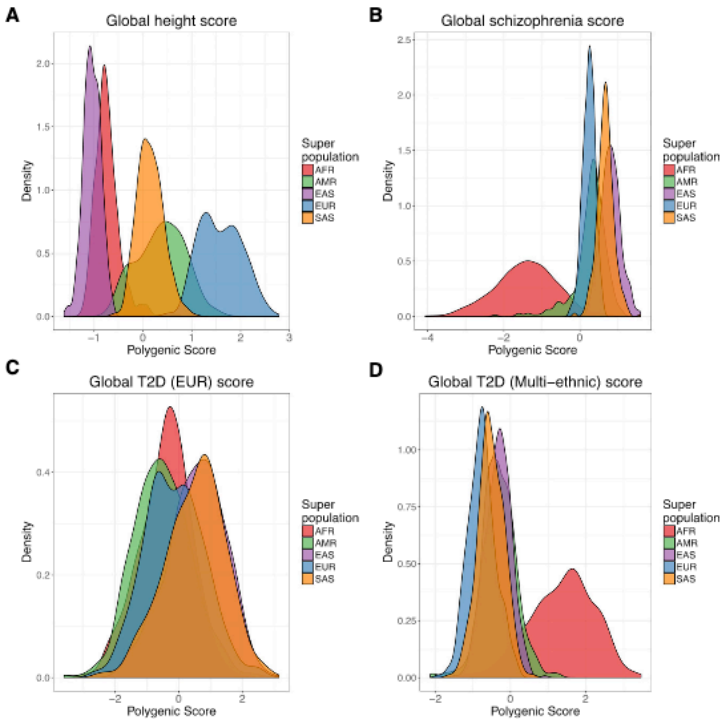
Alicia R. Martin, Masahiro Kanai, Yoichiro Kamatani, Yukinori Okada, Benjamin M. Neale & Mark J. Daly

*Nature Genetics* **51**, 584–591(2019) | Cite this article

**13k** Accesses | **69** Citations | **522** Altmetric | Metrics

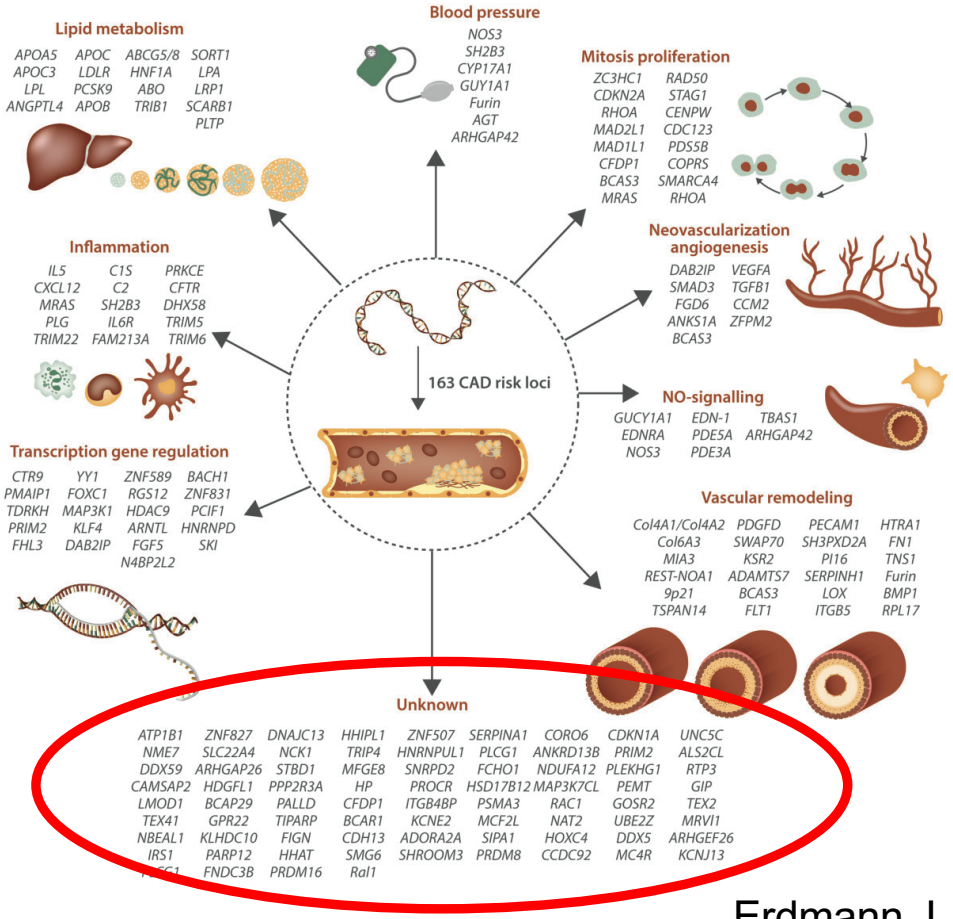
### Abstract

Polygenic risk scores (PRS) are poised to improve biomedical outcomes via precision medicine. However, the major ethical and scientific challenge surrounding clinical implementation of PRS is that those available today are several times more accurate in individuals of European ancestry than other ancestries. This disparity is an inescapable consequence of Eurocentric biases in genome-wide association studies, thus highlighting that—unlike clinical biomarkers and prescription drugs, which may individually work better in some populations but do not ubiquitously perform far better in European populations—clinical uses of PRS today would systematically afford greater improvement for European-descent populations. Early




Martin, et al, *Am J Hum Genet,* 2017

# CAD GWAS findings highlight many pathways




Erdmann J, et al, *Cardiovasc Res* 2018

eQTLs vs GWAS: same exposure, different outcomes

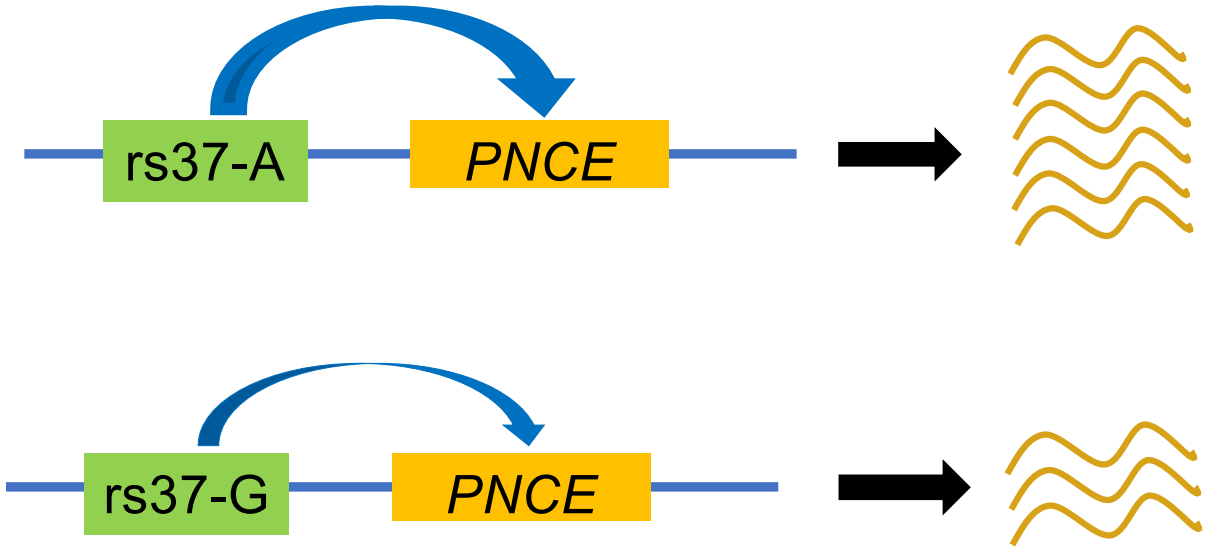
**GWAS**

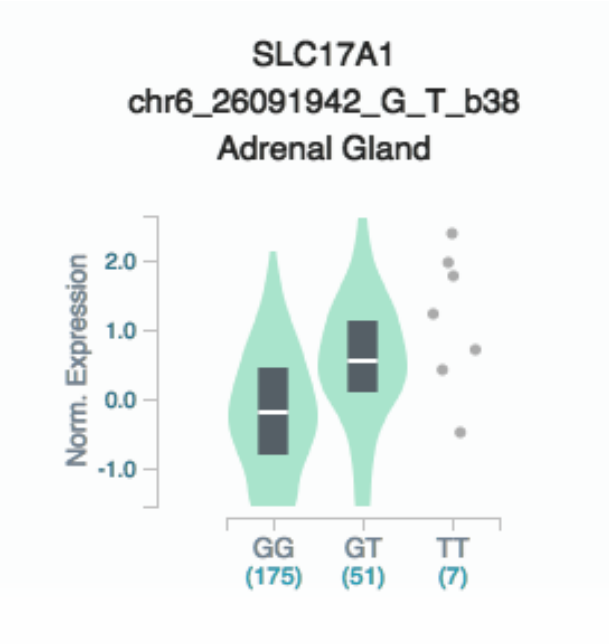
Genotypes → Disease / Trait

**eQTL analysis**

Genotypes → Gene expression

# eQTLs are measured by comparing expression across genotypes



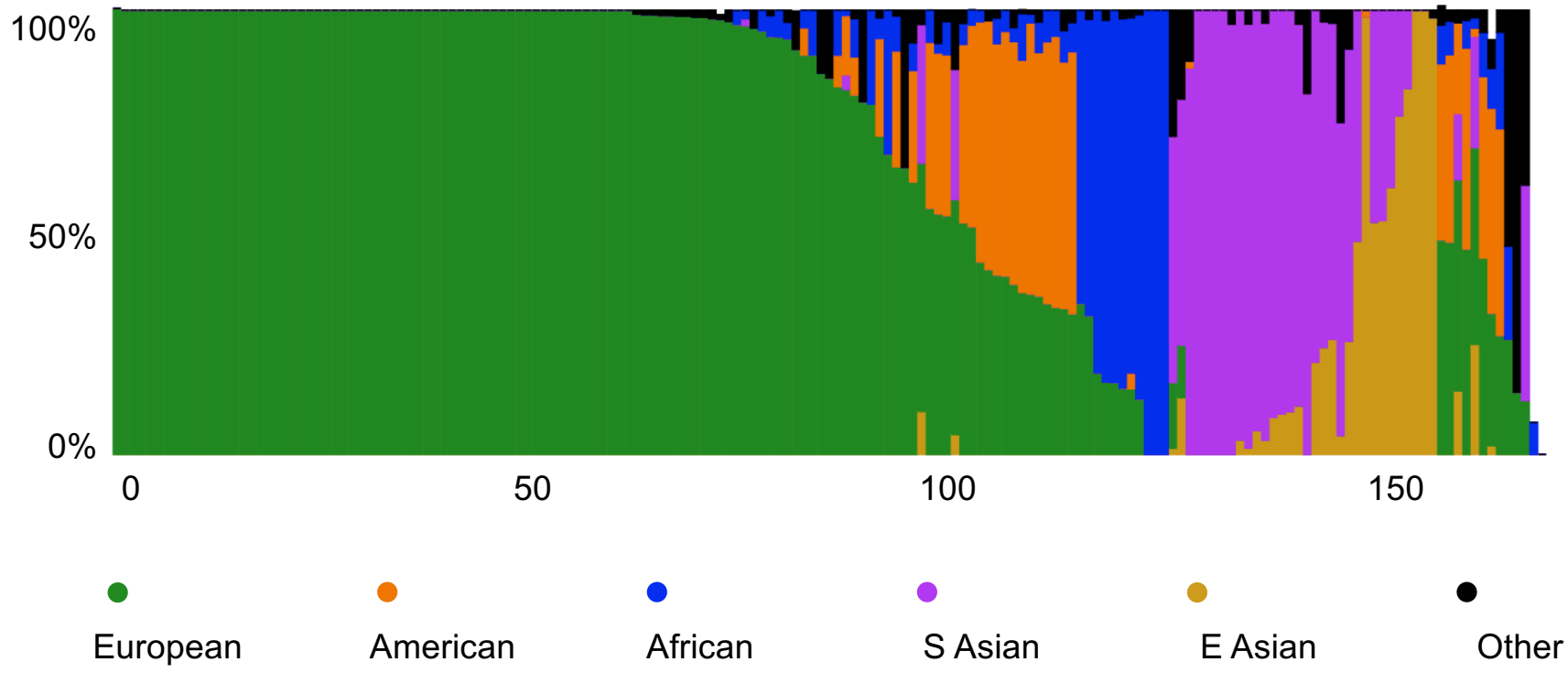



# Our study population

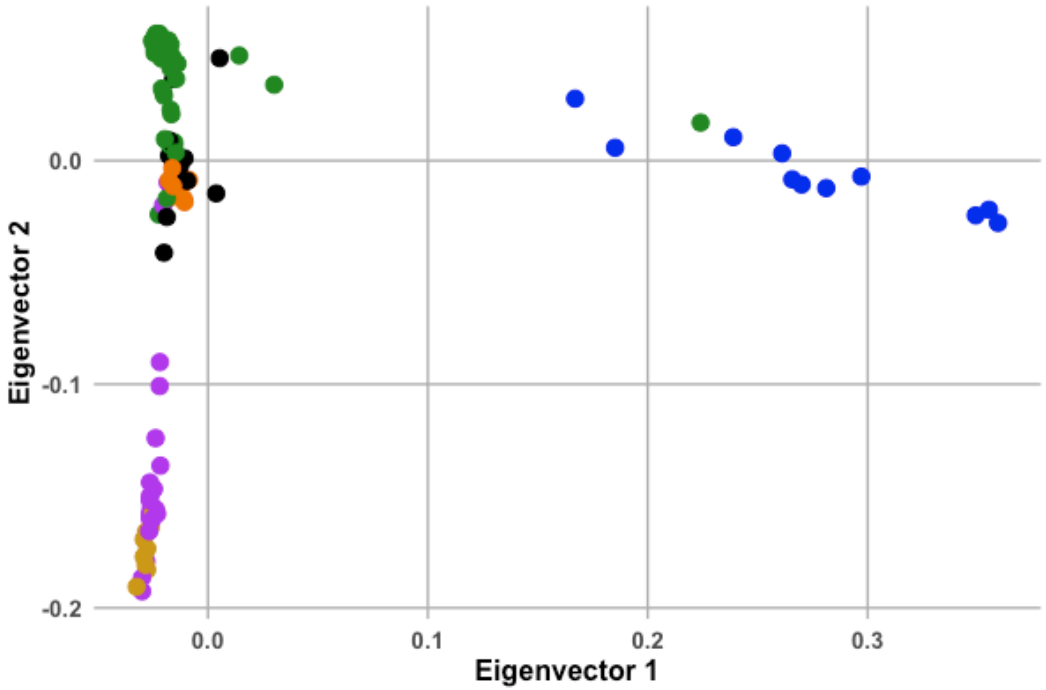
Patient demographics

- ~200 adults, either receiving a heart transplant or a donor organ which could not be transplanted; ~35% female sex
- ~150 individuals with coronary samples sent for RNA sequencing (outcome) as well as low-pass whole-genome sequencing (exposure)

# Broad ancestral representation

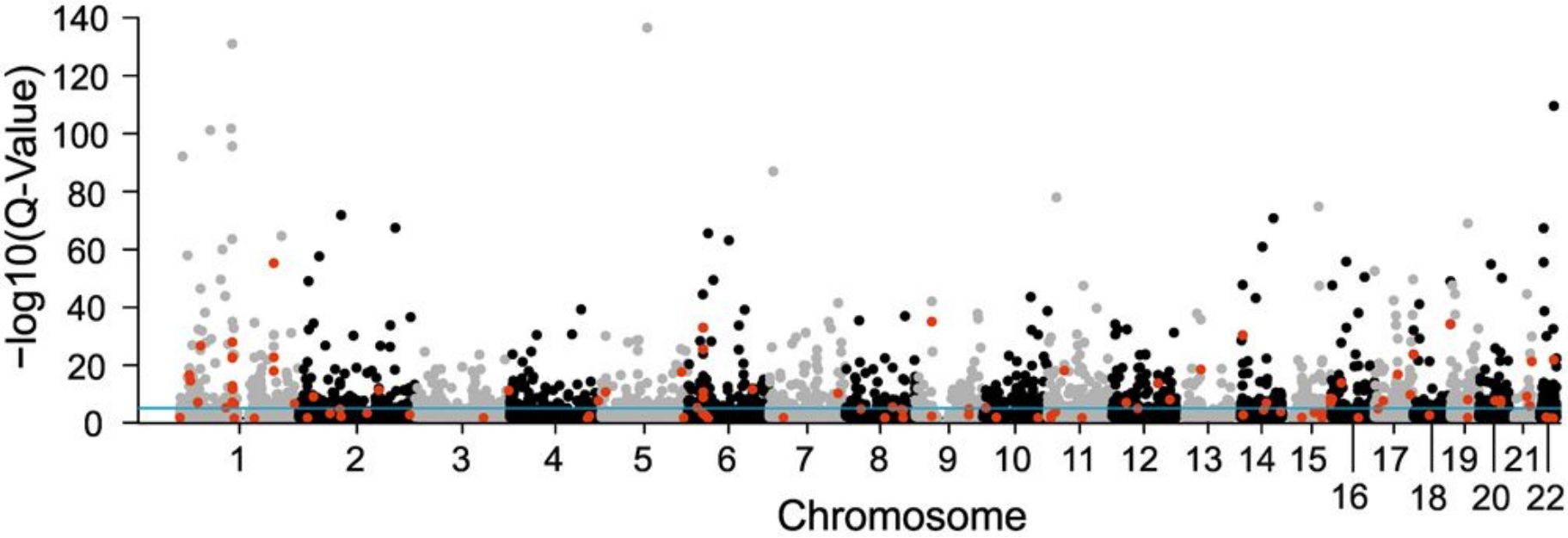


# PCs cluster by majority inferred ancestry




European
American
African
S Asian
E Asian
Other

# eQTLs are typically visualized in a Manhattan plot

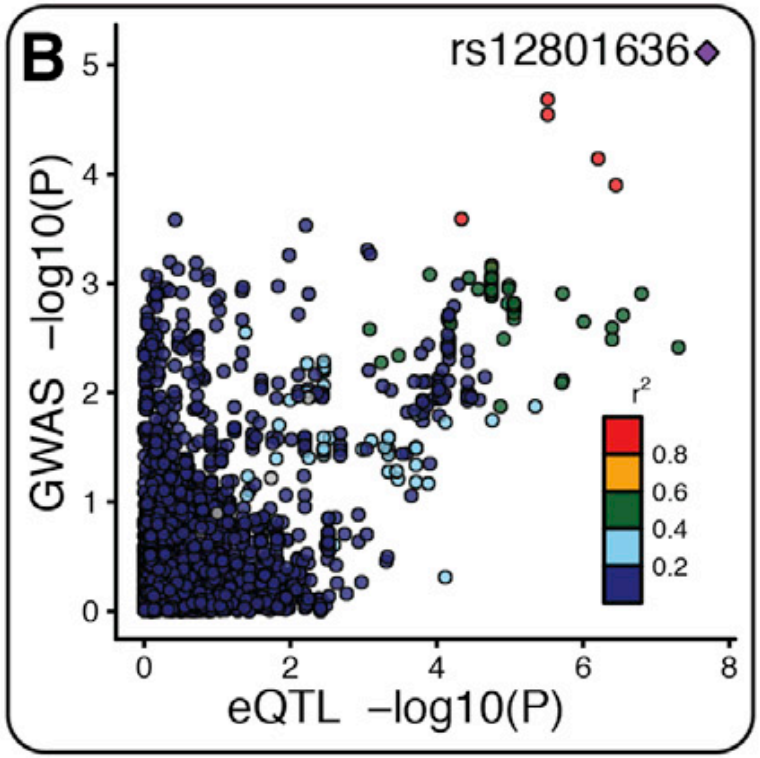



Strunz, et al, *Scientific Reports* 2018

# After eQTL mapping, colocalization helps identify causal variants

- If aforementioned problems can be ameliorated, comparing GWAS variants to eQTL variants improves interpretability
- Causal mechanism at a particular locus usually requires molecular experimentation
    - Preferable to limit number of variants (materials and time are expensive)

# Regional association plots help with interpretation




X-axis: GWAS significance

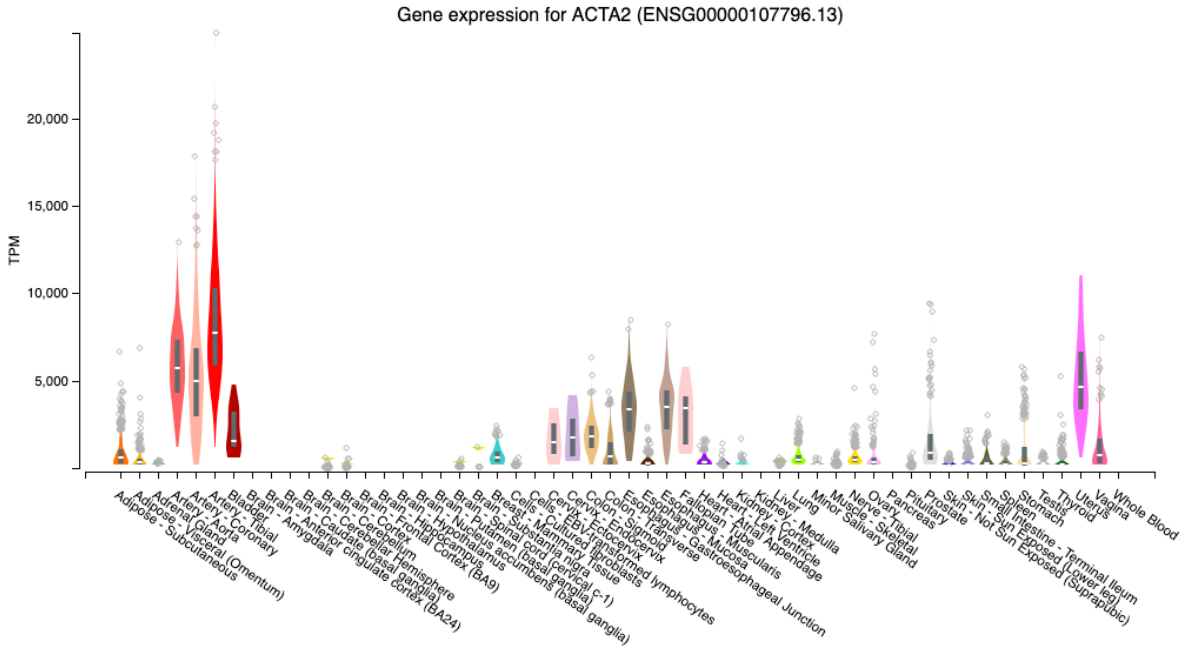
Y-axis: gene-expression significance

Color legend: LD (likelihood of variant being inherited together with the lead variant/purple diamond)

Liu B, et al, *AJHG* 2018

# Comparing results across studies is useful

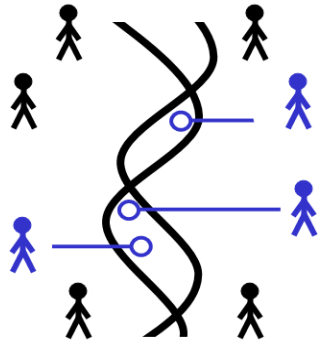
- Replication is important
- Meta-analysis can increase statistical power
- Comparing data across tissue types can improve interpretation—most studies only have one or two tissues



# Comparing results across studies is difficult

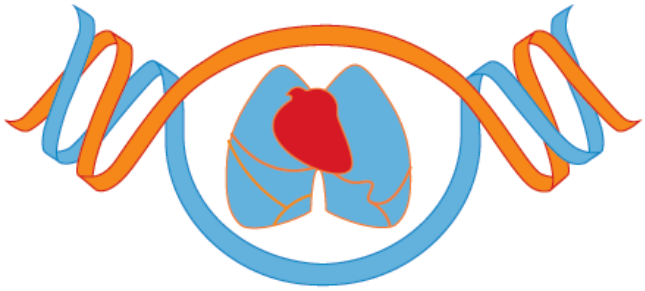
- GTEx and other publicly available data curation efforts have featured primarily or exclusively European-ancestry populations
- Rare alleles more likely to be causal or have large effect size, less likely to be globally present, less likely to be found in European-ancestry population
- EBI GWAS catalog faces similar representation issues

# Progress is being made



PAGE Study

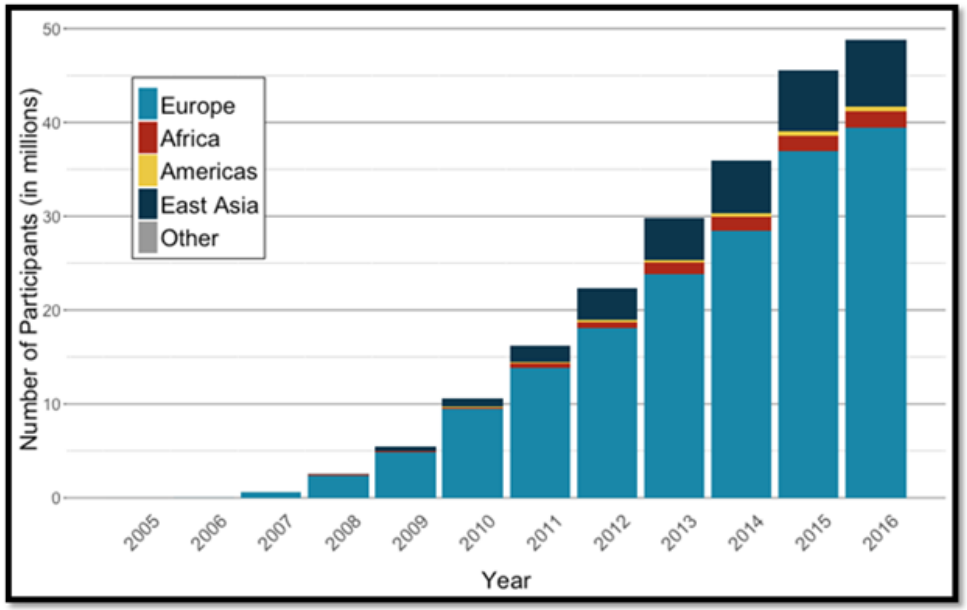




TOPMed

# However, research has a long way to go

- Journals continue to publish European-only studies
- Agencies continue to fund studies that have exclusionary designs
- Researchers may often feel like “this is someone else’s problem”
- Studies may use a population of convenience (UKBB) and think it justified to exclude non-European individuals




Wojcik, 2017

# Summary

- Ancestrally diverse samples are the best option for identifying globally relevant disease associations and assembling widely applicable translational tools such as PRS
- Despite evidence that inclusive study design improves ability to find genomic associations, it remains uncommon
- Consideration of the community contributing to your study will help allay justifiable concerns of communities that have been grossly mistreated in the past
- We need more funding for ancestrally diverse publicly available resources—we have to work together to achieve this!

# Nature Worldview: Angela Saini (March 9th)



WORLD VIEW · 09 MARCH 2020

## Want to do better science? Admit you're not objective

When science is viewed in isolation from the past and politics, it's easier for those with bad intentions to revive dangerous and discredited ideas.

Angela Saini

One of the world's leading universities — University College London (UCL) — has completed an inquiry into its support for the discredited pseudoscience of eugenics. Funds linked to Francis Galton, a racist who believed it was possible to improve the British population through selective breeding, and who founded the Eugenics Records Office at UCL in 1904, continue to line the university's coffers to the value of more than £800,000 (US$1 million).

PDF version

RELATED ARTICLES

How science has shifted our sense of identity

# Acknowledgments


**Miller Lab**

Adam Turner

Daud Khan

Doris Wong

Yipei Song

Nelson Barrientos

**UNC Chapel Hill**

Kari North

Christy Avery

Heather Highland

Misa Graff

Kristin Young

**PAGE investigators**

Eimear Kenny (ISMMS)

Chris Gignoux (Colorado)

Genevieve Wojcik (Hopkins)

**UVA CVRC & CPHG**

Mary Sheffer

Jennifer Dean

**Funding:**

**T32 HL007284-40S1**