Mind the (diversity) gap: contributions of diverse populations to common disease studies

Lucia A. Hindorff, PhD, MPH

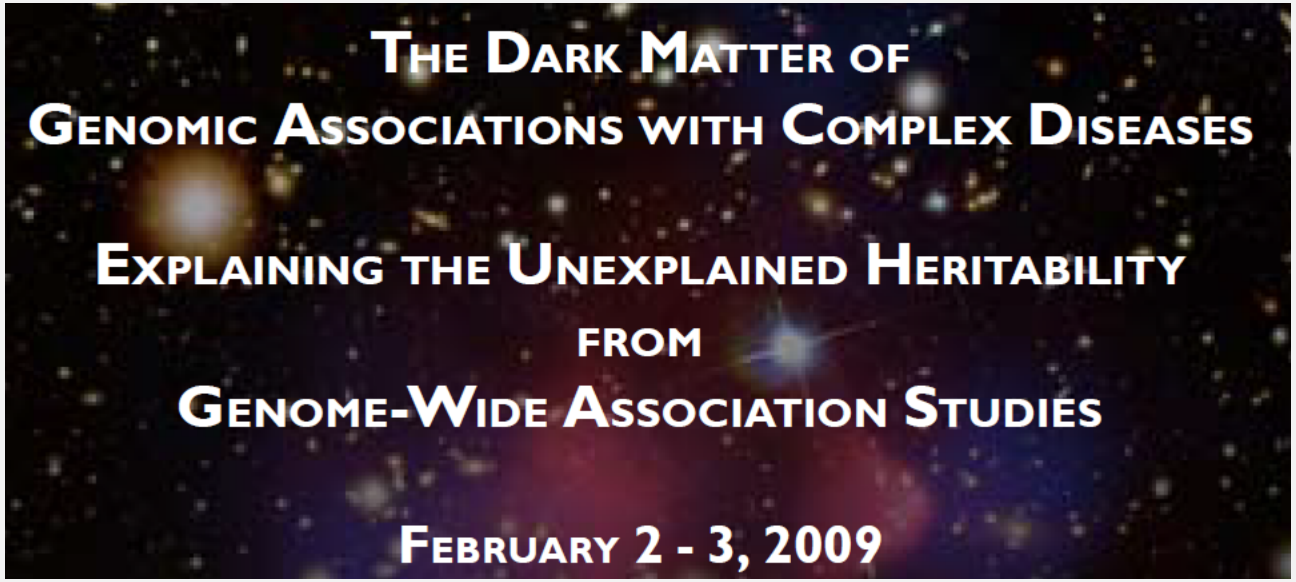
Program Director, Division of Genomic Medicine

Missing Heritability Workshop – May 2, 2018

NIH National Human Genome Research Institute

The **Forefront** of **Genomics**®

2009


The Dark Matter of
Genomic Associations with Complex Diseases

Explaining the Unexplained Heritability
from
Genome-Wide Association Studies

February 2 - 3, 2009

2018

What have we learned from GWAS in diverse populations?

- Landscape of GWAS
- Evolution of GWAS arrays
- Enhancing GWAS analyses

GWAS Catalog

The NHGRI-EBI Catalog of published genome-wide association studies

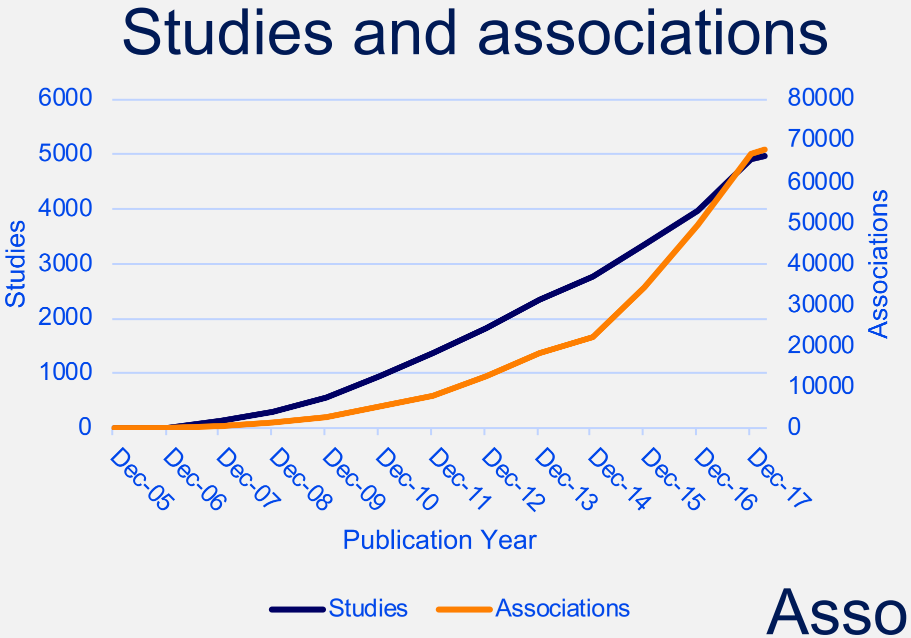
Examples: breast cancer, rs7329174, Yang, 2q37.1, HBS1L, 6:16000000-25000000

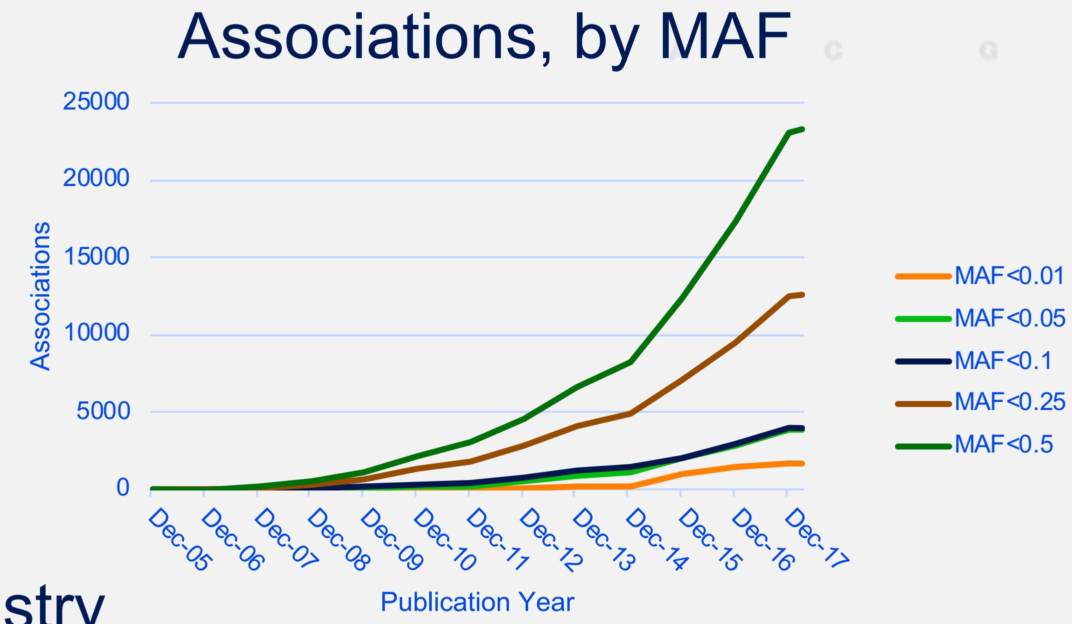
April, 2018:

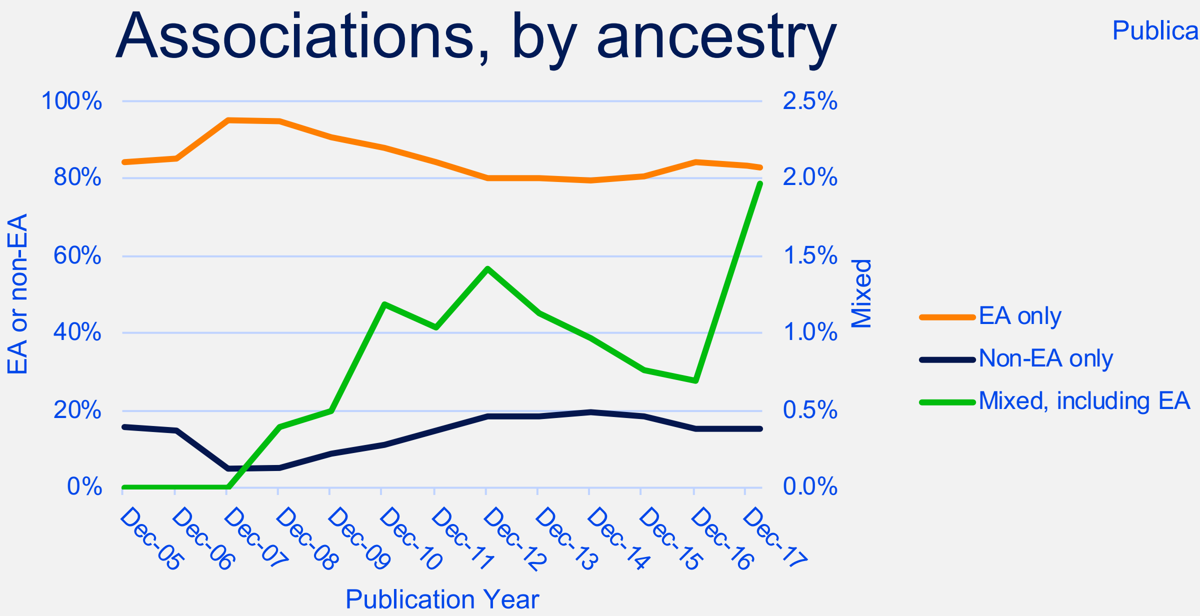
- 3,349 publications
- 59,967 unique SNP-trait associations

https://www.ebi.ac.uk/gwas/

Evolving landscape of GWAS



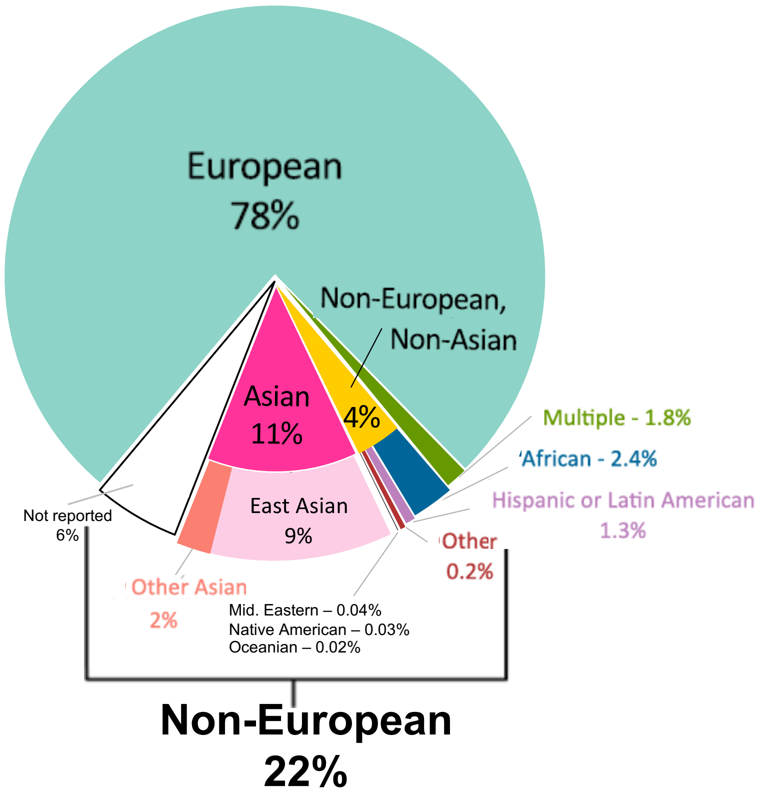




NIH
NHGRI

Studies vs. associations

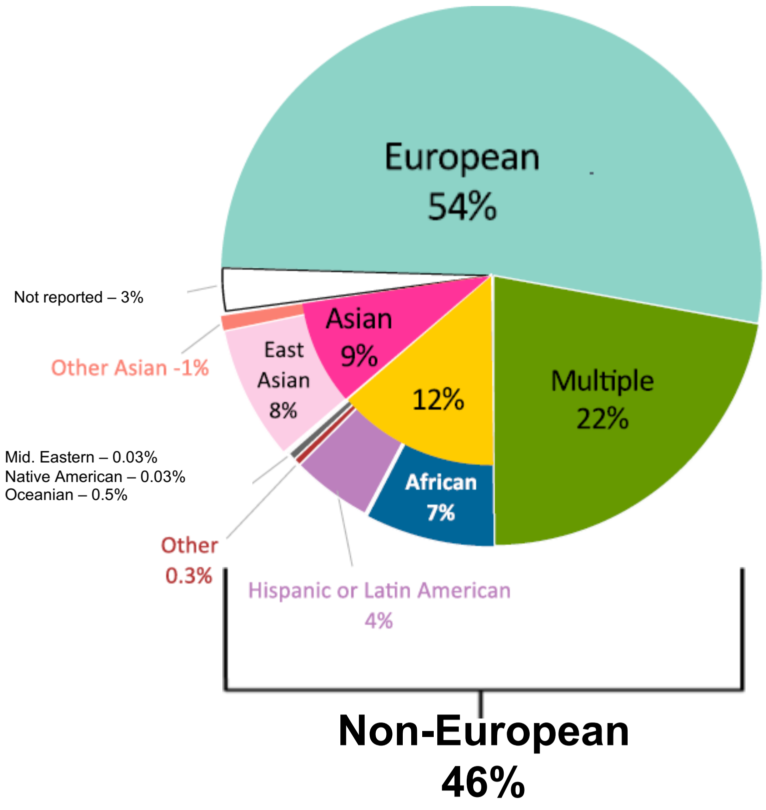
Individuals



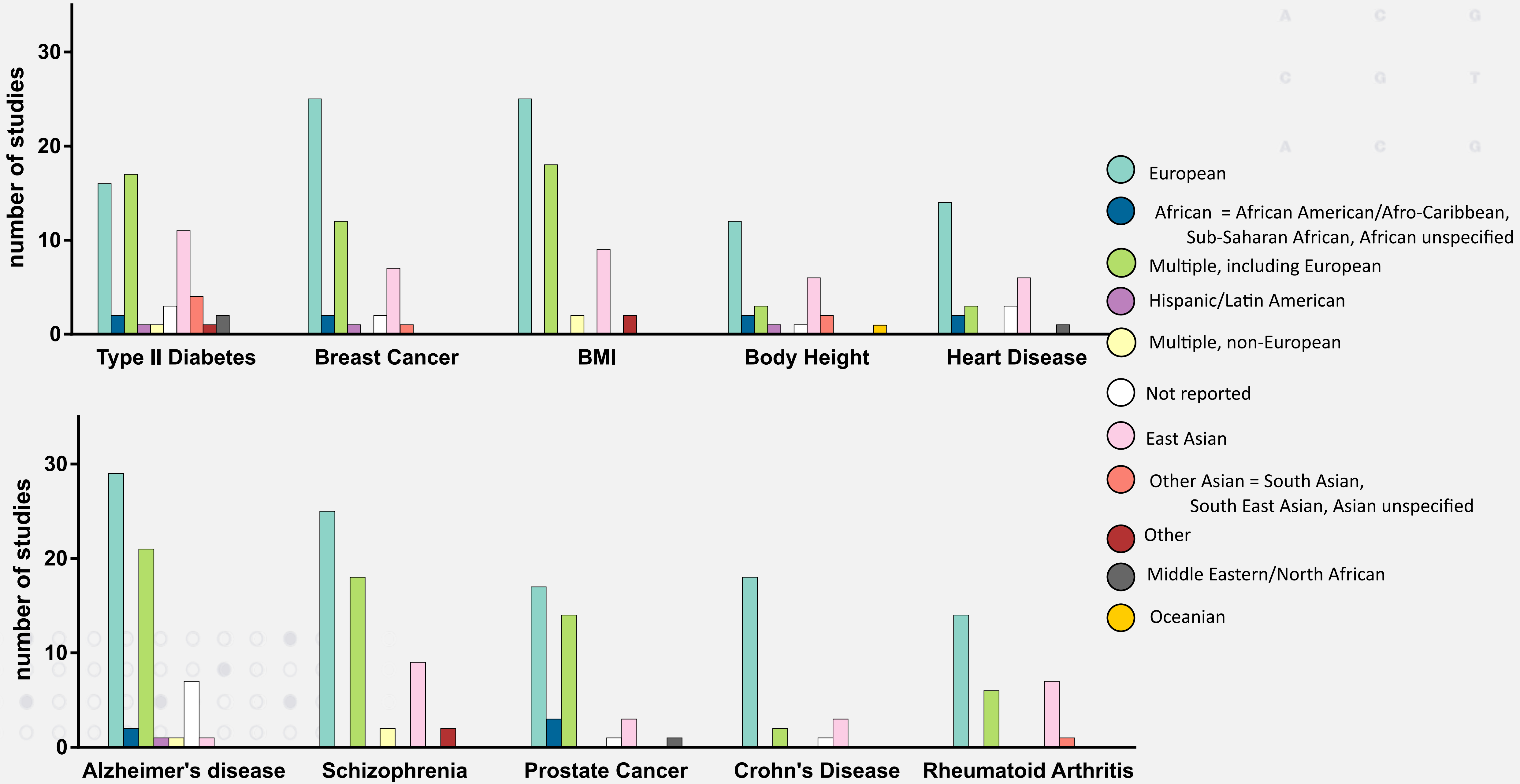
Legend

- ● Non-European, Non-Asian
- ○ Not reported
- ● European
- ● All Asian
- ● East Asian
- ● Other Asian
- ● Hispanic or Latin American
- ● African
- ● Middle Eastern
- ● Multiple
- ● Multiple, non-European
- ● Multiple, including European
- ● Other and other admixed

Associations

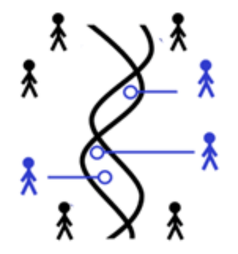


Morales, et al. *Genome Biology* 2018. PMID 29448949



Morales, et al. *Genome Biology* 2018. PMID 29448949

Population Architecture using Genomics and Epidemiology (PAGE)

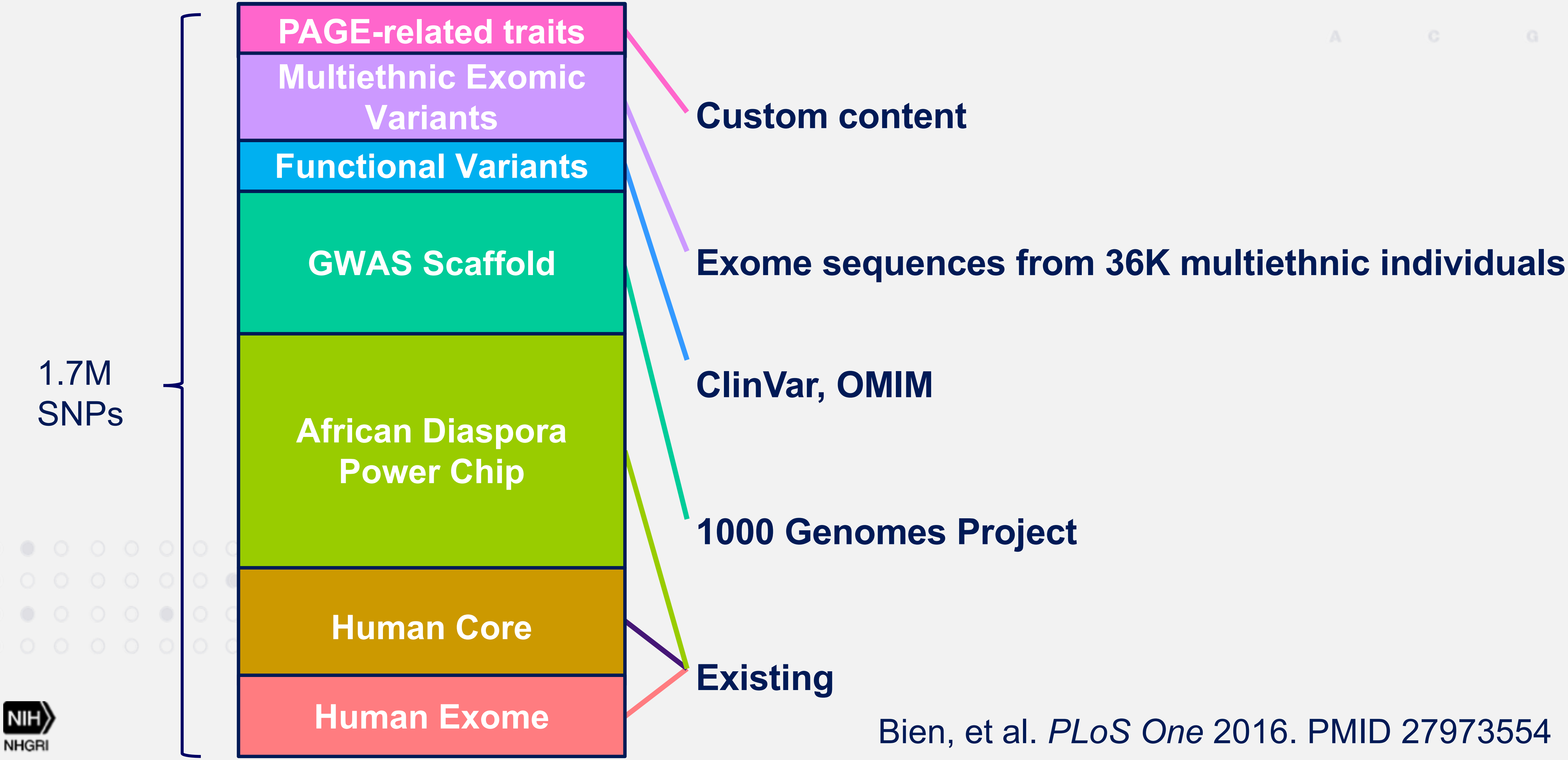


- Multi-cohort study
  - MEC
  - WHI
  - SOL
  - MSSM BioMe
- ~50,000 individuals
- Depth and breadth of phenotype information

Ancestry	N
African-American	17,299
Asian-American	4,680
Hispanic/Latino	22,216
Native Hawaiian	3,940
Native American	652
Other	1,052
Total	**49,796**

NIH
NHGRI

PAGE Multi-ethnic Genotyping (MEGA) Array



1.7M SNPs

PAGE-related traits

Multiethnic Exomic Variants

Functional Variants

GWAS Scaffold

African Diaspora Power Chip

Human Core

Human Exome

Custom content

Exome sequences from 36K multiethnic individuals

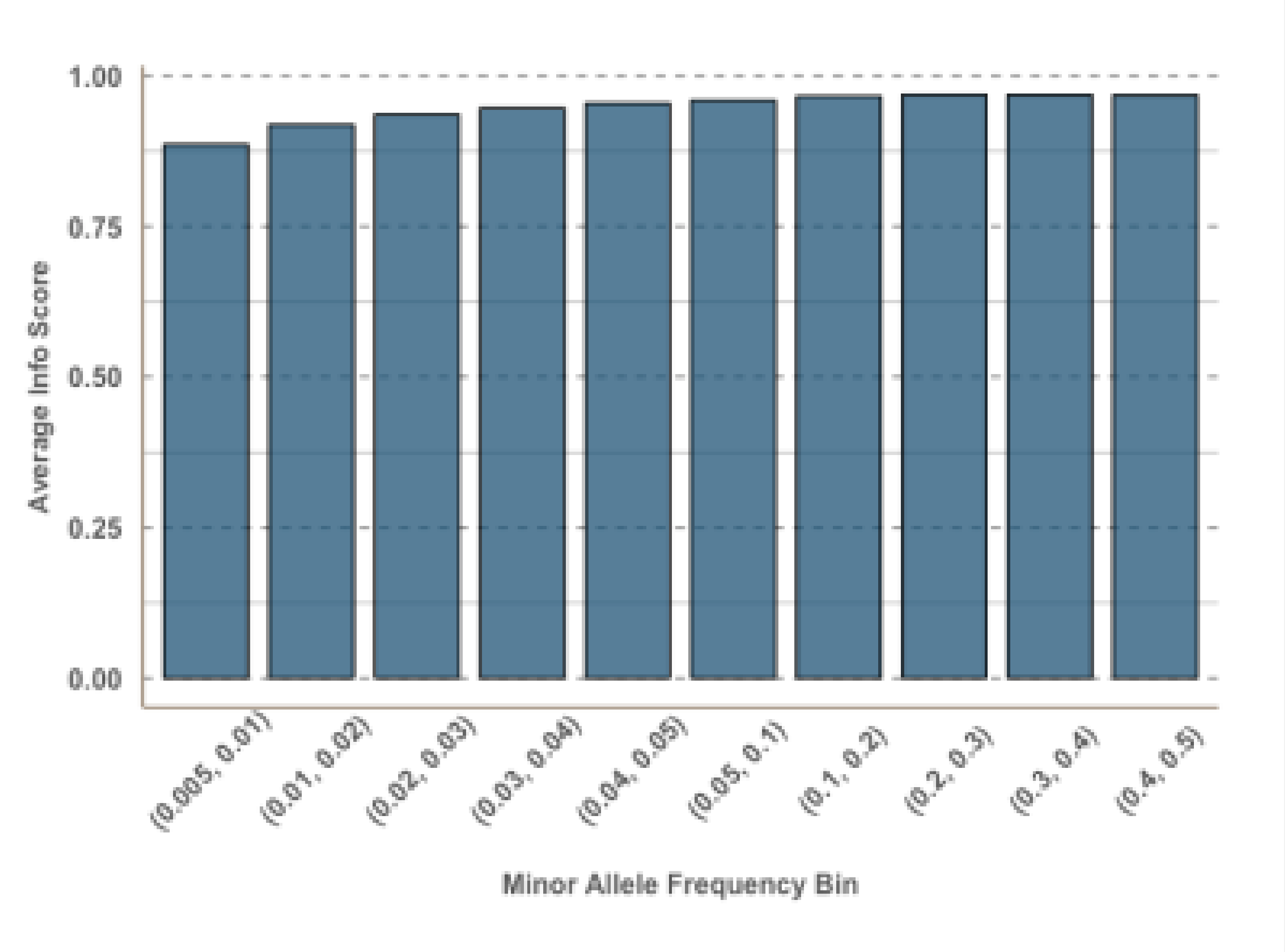
ClinVar, OMIM

1000 Genomes Project

Existing

Bien, et al. *PLoS One* 2016. PMID 27973554

GWAS scaffold tags common and low frequency variants



Info scores from PAGE MEGA array data, imputed to 1000 Genomes

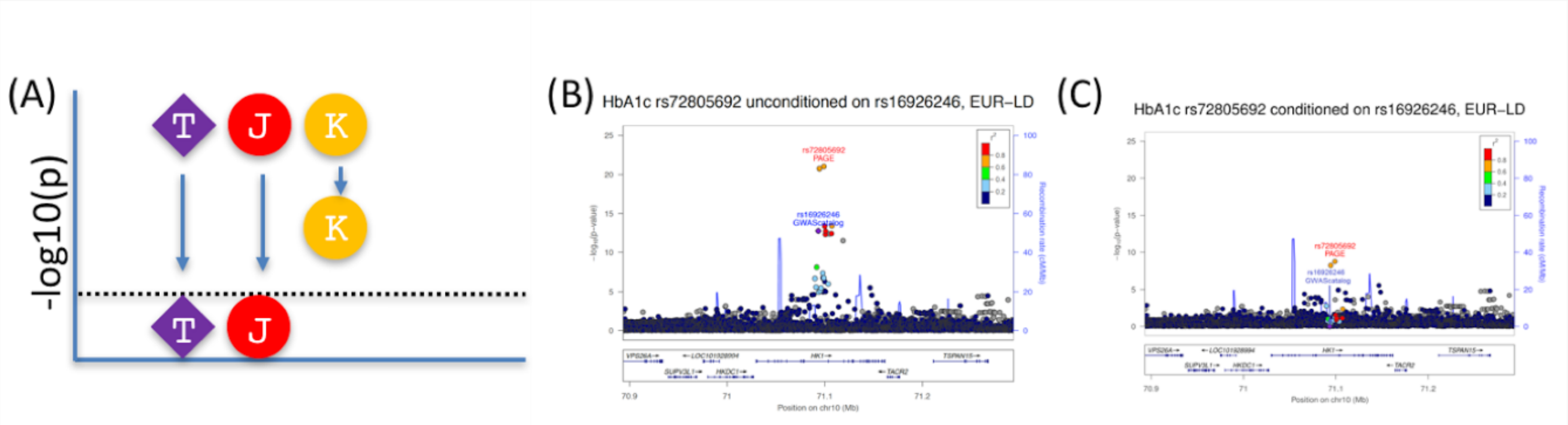
Wojcik, Graff, Kenny, Carlson, et al. Updated from https://www.biorxiv.org/content/early/2017/09/15/188094

Multiethnic data complement and add to prior GWAS results

	Largest GWAS catalog discovery population				PAGE
Phenotype	European	East Asian	African	Hispanic/ Latino	
HDL	99,900	12,545	7,917	4,383	33,063
LDL	94,595	12,545	7,861	4,383	32,221
TG	96,598	12,545	7,601	4,383	33,096
TC	100,184	8,344	6,480	4,383	33,185

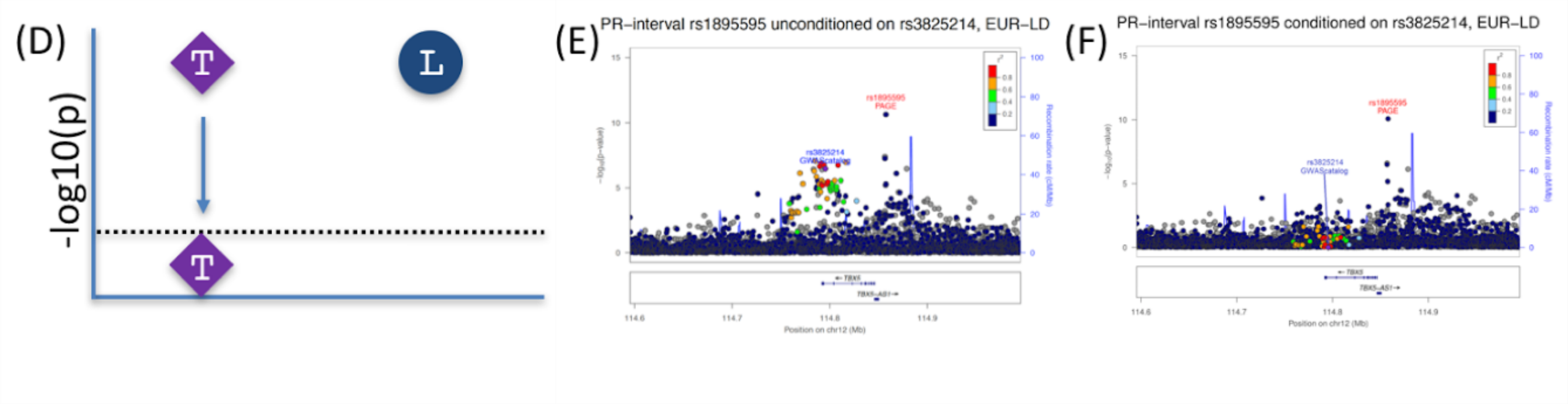
Wojcik, Graff, Kenny, Carlson, et al. Updated from https://www.biorxiv.org/content/early/2017/09/15/188094

Residual signal: fine mapping



Wojcik, Graff, Kenny, Carlson, et al. Updated from https://www.biorxiv.org/content/early/2017/09/15/188094

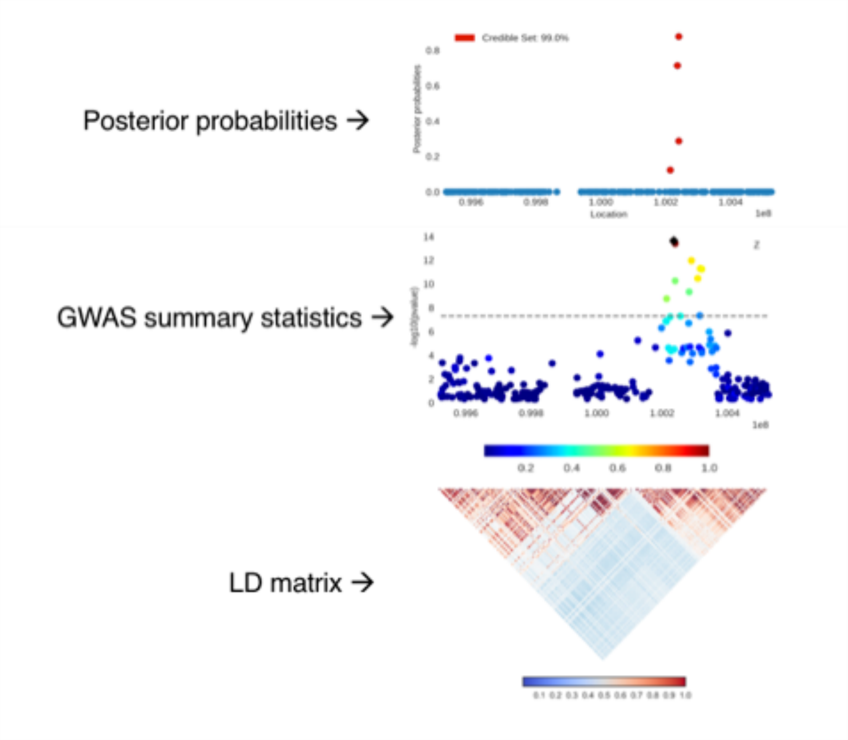
Residual signal: secondary allele



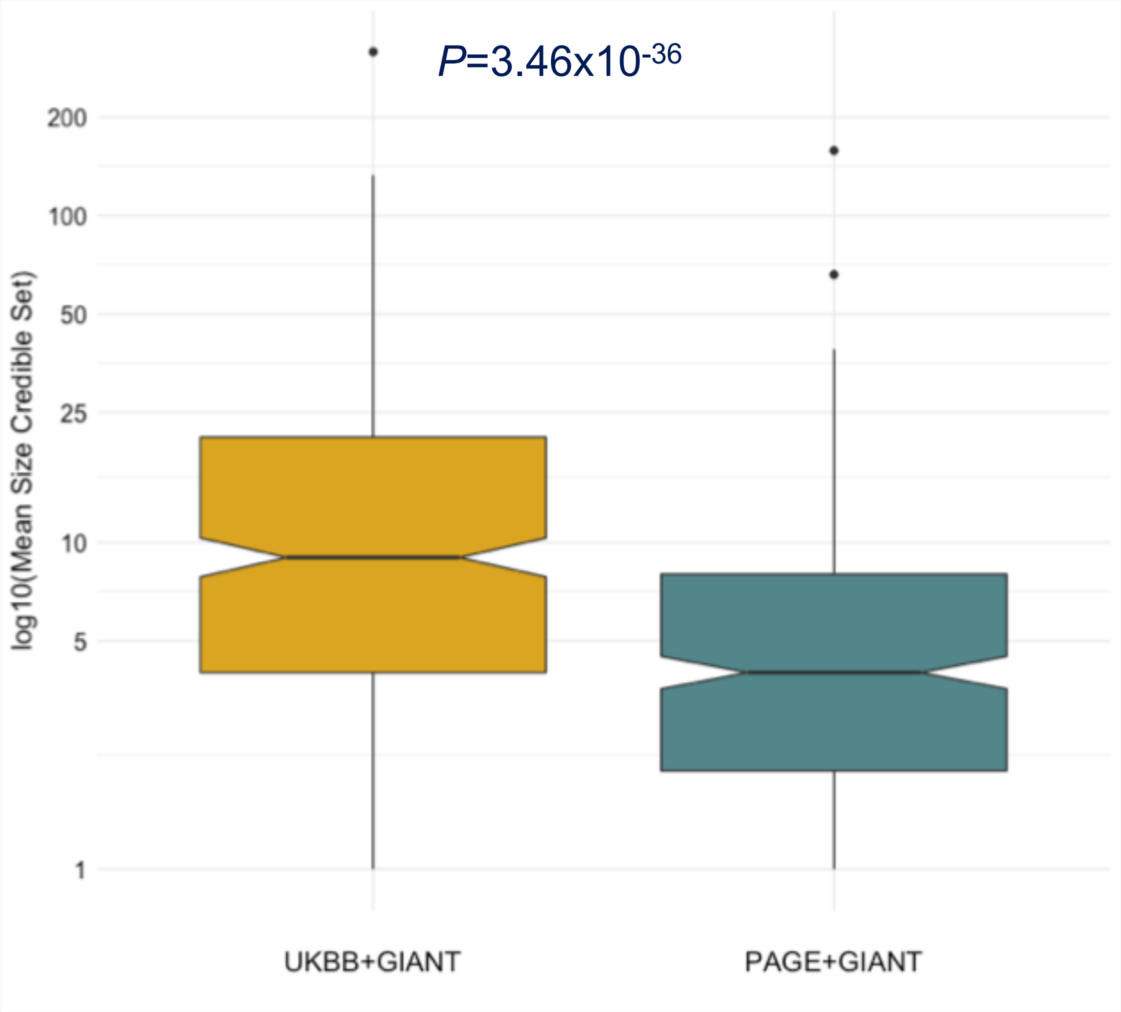
Wojcik, Graff, Kenny, Carlson, et al. Updated from https://www.biorxiv.org/content/early/2017/09/15/188094

What additional information is gained from diverse participants?

- “Why not just add 50K more Europeans?”
- Meta-analysis of GIANT+PAGE vs. GIANT+50K UKBB
- FINEMAP over both datasets with weighted LD matrices
- Compare: credible set size, posterior probabilities



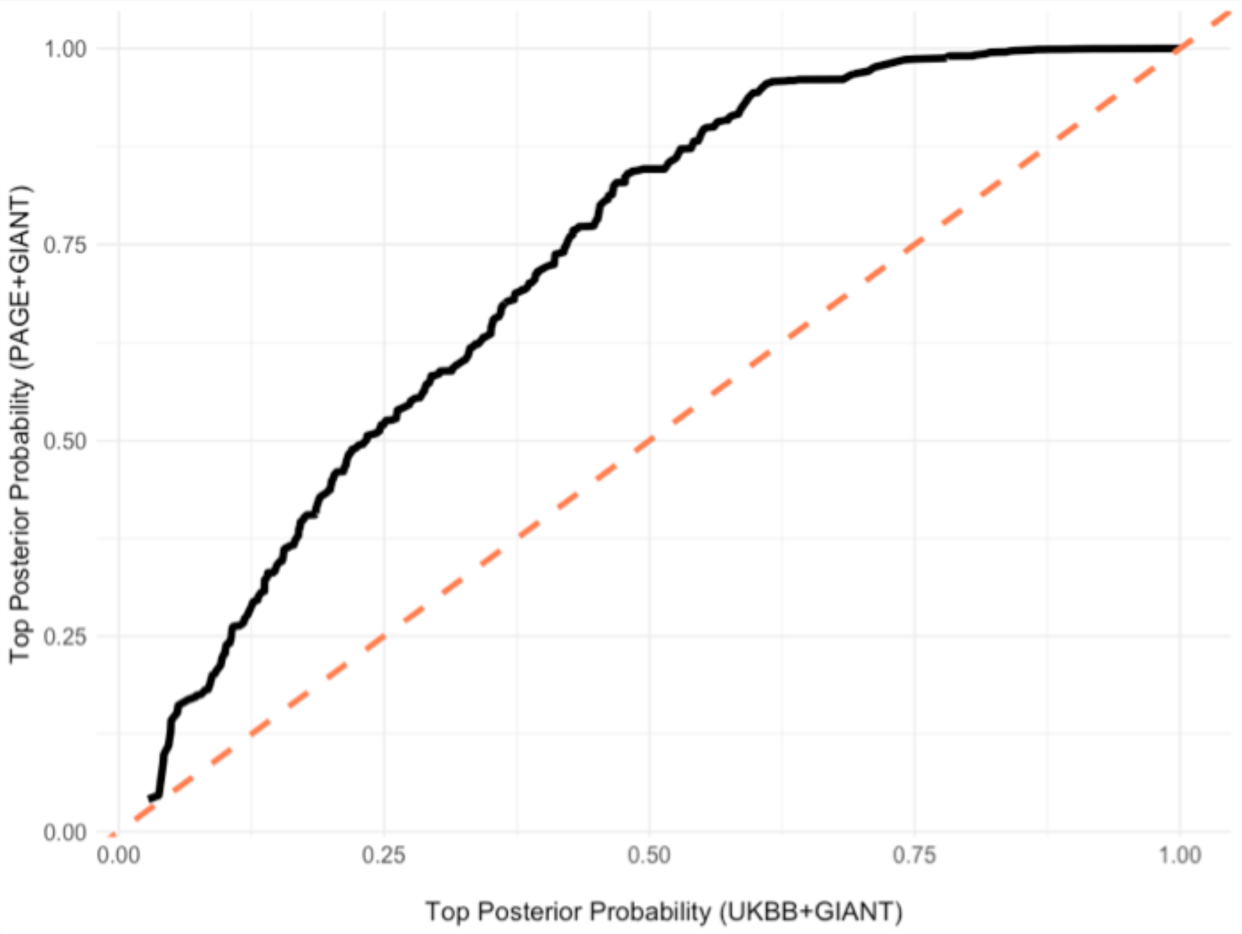
Comparing credible sets for height



- 366 index loci
- Credible set sizes are significantly smaller for height in meta-analysis of GIANT + PAGE, compared to GIANT + 50K UKBB Europeans

Courtesy G. Wojcik

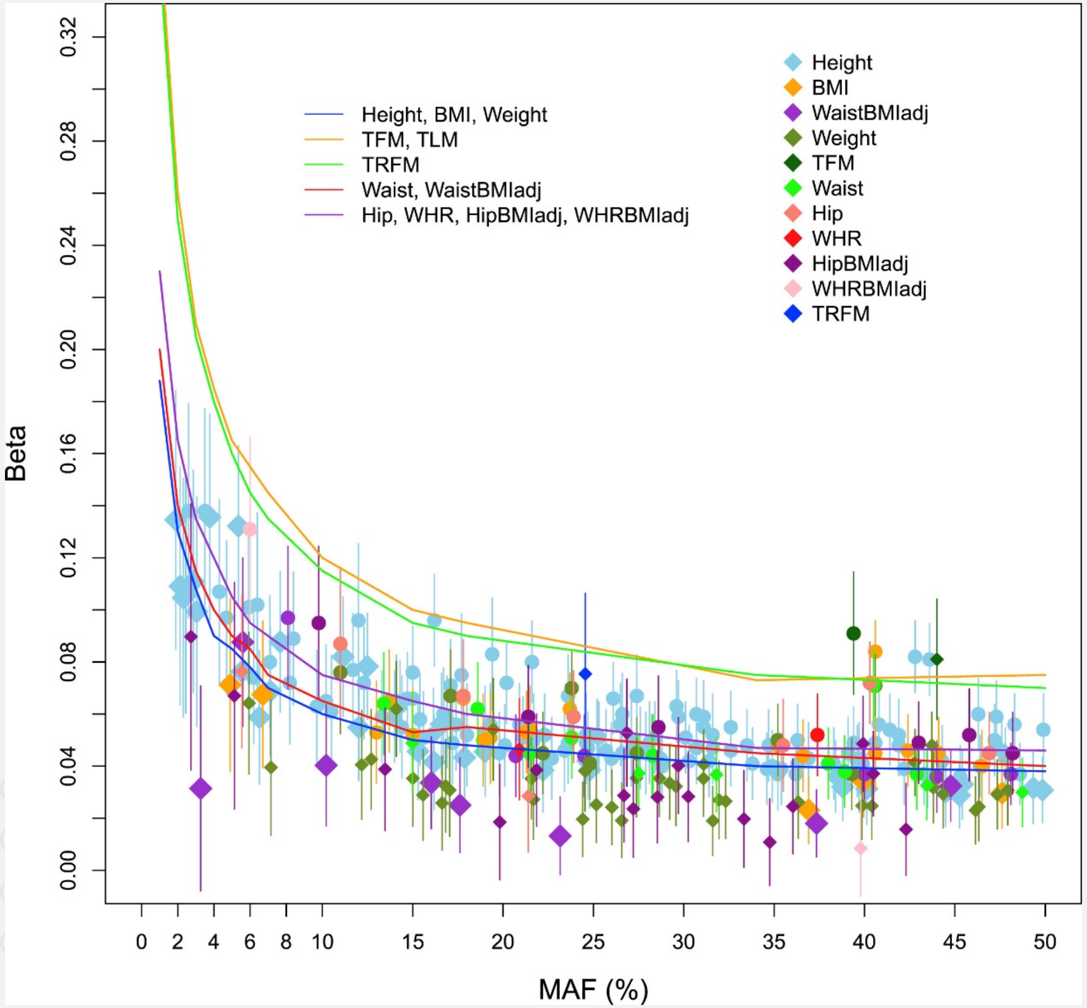
Comparing posterior probability



- 366 index loci
- Top ranked SNP has higher posterior probability in meta-analysis of GIANT + PAGE, compared to GIANT + 50K UKBB Europeans

Courtesy G. Wojcik

Towards rare variant association studies

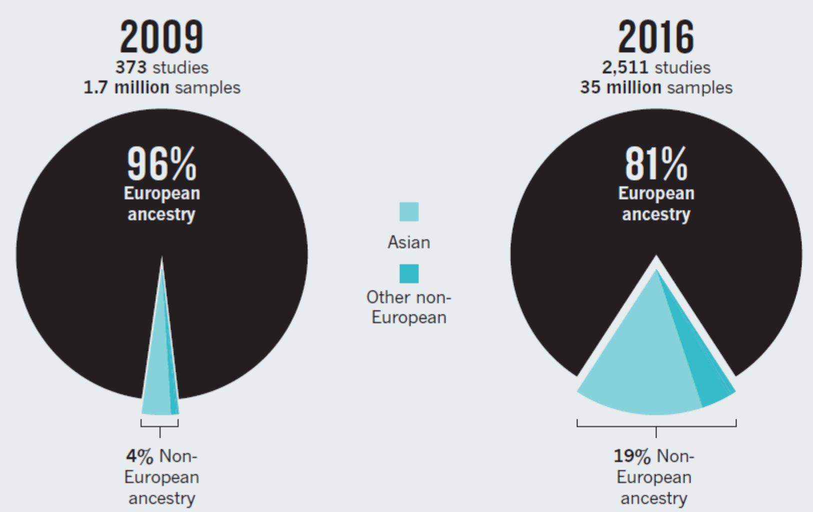


- WGS with deep imputation in up to 267,616 individuals
- 12 anthropometric traits
- 106 novel signals
  - 6 not implicated with related traits
  - 28 independent signals at previously reported regions
  - 72 previously reported signals for different traits
- All European ancestry

Tachmazidou, *AJHG* 2018; PMID 28552196

Information disparity

- Missing heritability: lack of scientific information
- Lack of information likely to differ among groups (eg, racial/ethnic minorities)
- To address this information disparity, need to look beyond the low-hanging fruit



Popejoy and Fullerton, Nature 2016. PMID 27734877

Acknowledgments

GWAS Catalog - EBI

- **Jackie MacArthur**
- **Joannella Morales**
- Helen Parkinson
- Paul Flicek
- Fiona Cunningham
- Tony Burdett
- Annalisa Buniello
- Maria Cerezo
- Laura Harris
- Emma Hastings
- Xin He
- Cinzia Malangone
- Aoife McMahon
- Annalisa Milano
- Zoe Pendlington
- Trish Wetzel

PAGE

- **Stephanie Bien**
- **Chris Carlson**
- **Sinead Cullina**
- **Chris Gignoux**
- **Misa Graff**
- **Jeffrey Haessler**
- **Heather Highland**
- **Eimear Kenny**
- **Katie Nishimura**
- **Yesha Patel**
- **Gen Wojcik**
- Steve Buyske
- Chris Haiman
- Charles Koopberberg
- Loic Le Marchand
- Ruth Loos
- Kari North
- Tara Matise
- Ulrike Peters
- CIDR/UW investigators
- Many PAGE co-investigators

NHGRI

- Heather Colley
- Maggie Ginoza
- Peggy Hall
- Teri Manolio
- Ken Wiley

NIMHD

- Regina James