Breakout Session 6: Track B

Using Ancestry-Agnostic Approaches for Genome-Wide Association Studies and Polygenic Risk Scores

Dr. Matteo D'Antonio
Assistant Professor, UC San Diego

Using ancestry-agnostic approaches for genome-wide association studies and polygenic risk scores

Matteo D'Antonio, Division of Biomedical Informatics, UC San Diego

Supplement award: Exploring the Ethics and Utility of Ancestry-Agnostic Approaches to Address Equity and Diversity Issues in Genome-Wide Association Studies

(PI Lucila Ohno-Machado, HG011558)



Project summary

- Overarching goals:
  - Improve the utility of genome science for all populations living in the United States
  - Make GWAS more informative for individuals of diverse backgrounds
  - Enhance equity in the distribution of benefits from genetics research

- Aim 1: Develop a novel ancestry-agnostic haplotyping method for improving GWAS and PRS accuracy (Matteo D'Antonio)

- Aim 2: Determine the feasibility of ancestry-agnostic approaches for GWAS and PRS research (Ramya Rajagopalan)

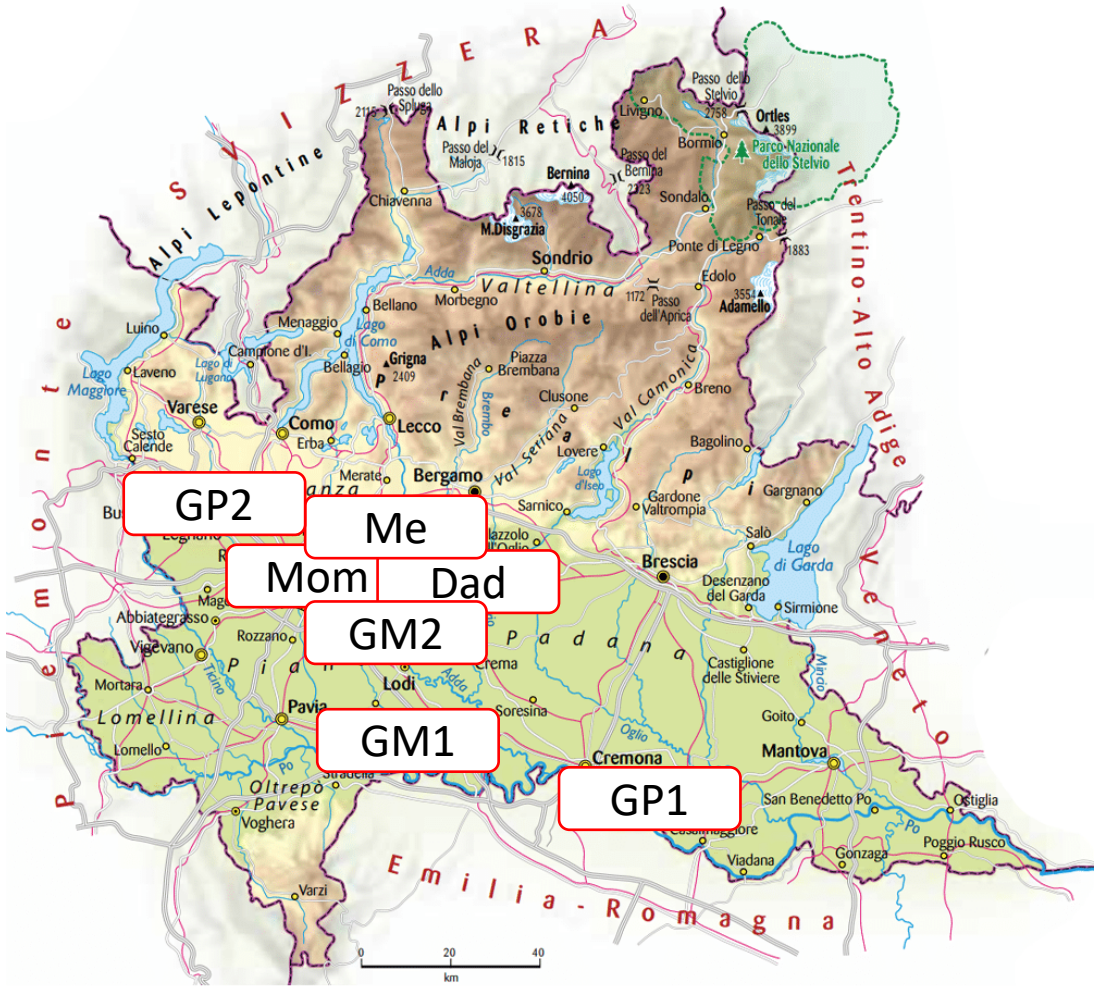
GWAS: genome-wide association studies
PRS: polygenic risk scores

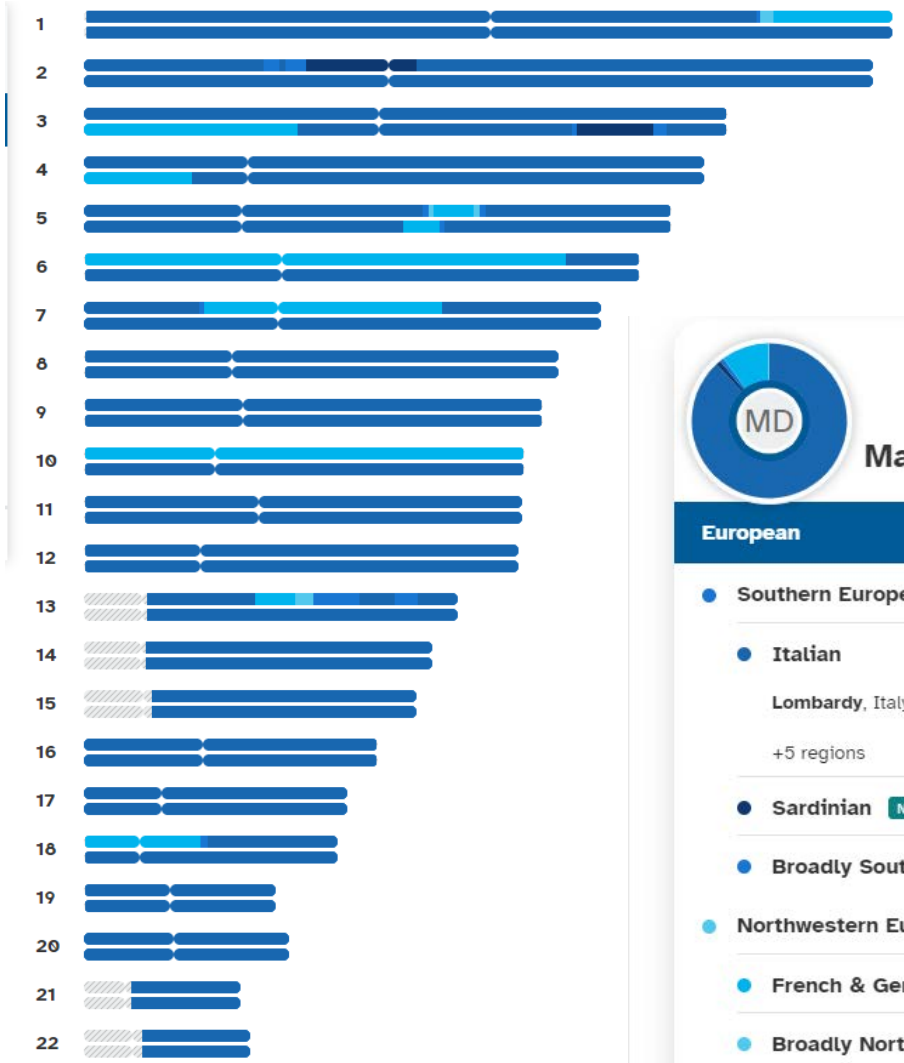
Research outputs

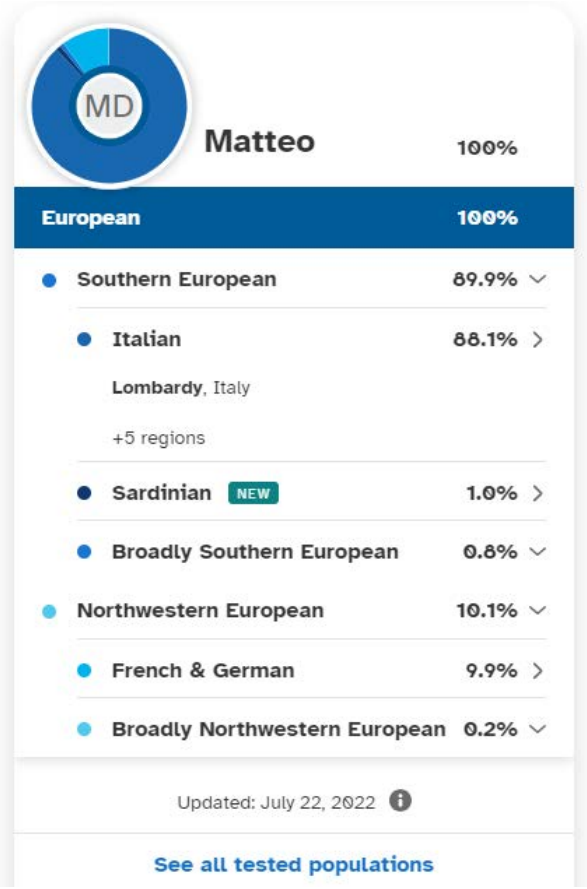
- Manuscripts
  - D'Antonio et al. Fine mapping spatiotemporal mechanisms of genetic variants underlying cardiac traits and disease (*Nat Comm* 2023)
  - Massarat A et al. Haptools: a toolkit for admixture and haplotype analysis (*Bioinformatics* 2023)
  - Nguyen et al. Complex regulatory networks influence pluripotent cell state transitions in human iPSCs (*Nat Comm* 2024)
  - Rajagopalan R et al. Enhancing Equity in Genomics: Incorporating Measures of Structural Racism and Discrimination as Social Determinants of Health within Genomics Research (*CERA-Hastings*, submitted)
  - D'Antonio M et al. Using an ancestry-agnostic approach to improve polygenic risk scores for breast cancer (in preparation)

- Conferences:
  - D'Antonio M et al. Chronic Kidney Disease and Kidney Function Genome-Wide Association Study Reveals Population-Specific Associations (*ASHG* 2023) – poster
  - D'Antonio M et al. Identifying variants associated with creatinine and cystatin C serum levels in the top 1% of the population provides insight into the genetic architecture of kidney function and chronic kidney disease (*MarketsAndMarkets* 2024) – oral presentation
  - Rajagopalan R et al. From Accounting for Race to Accounting for Racism: Strategies for Reimagining the Utility and Benefits of Genomics Research for Diverse Communities (*ELSIcon2024*, June 2024) - poster

We are all admixed

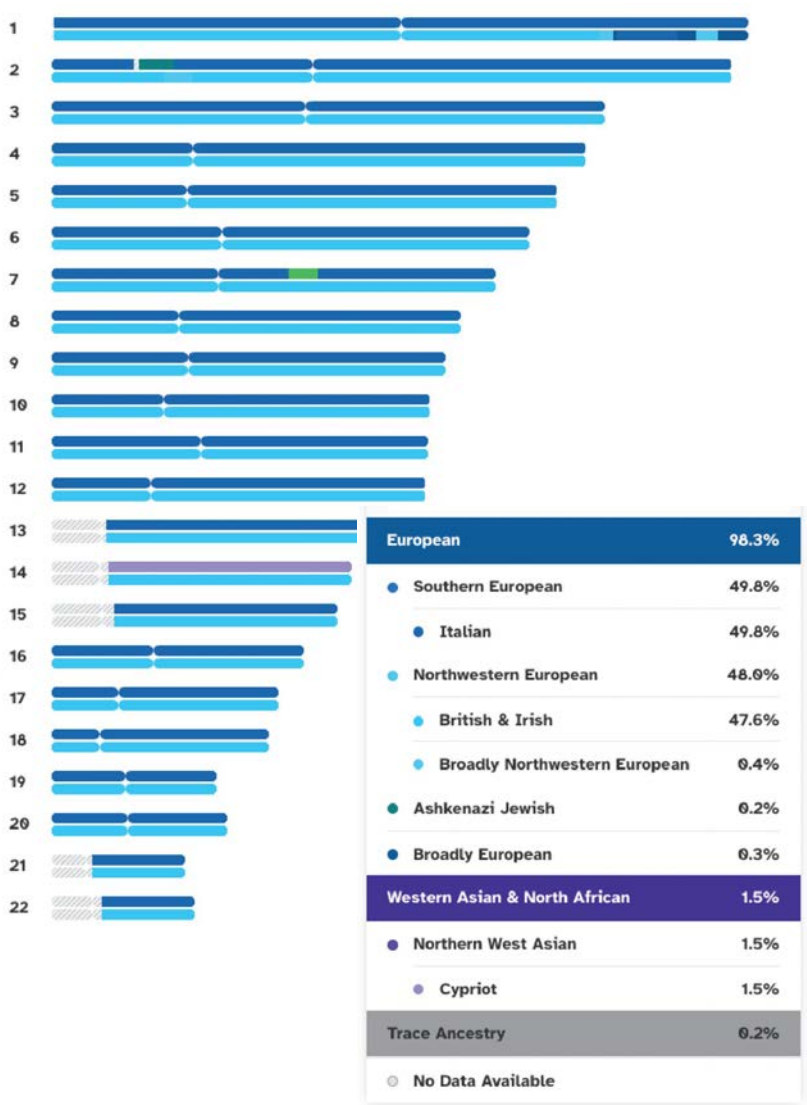


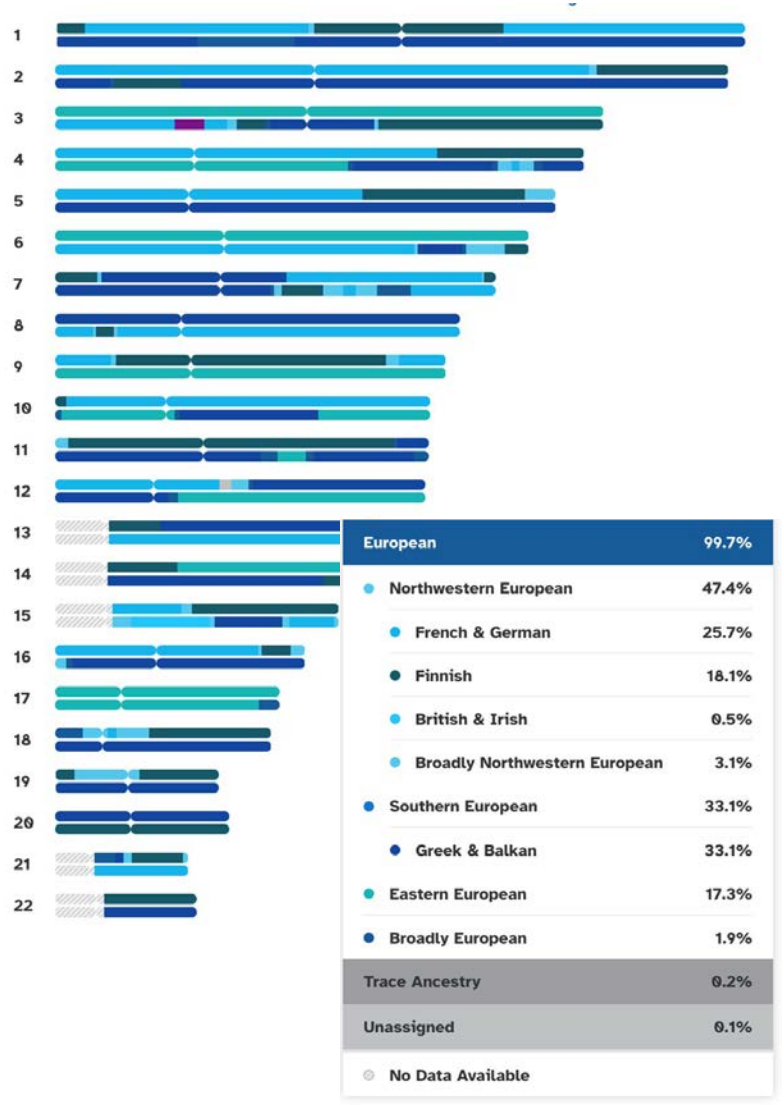


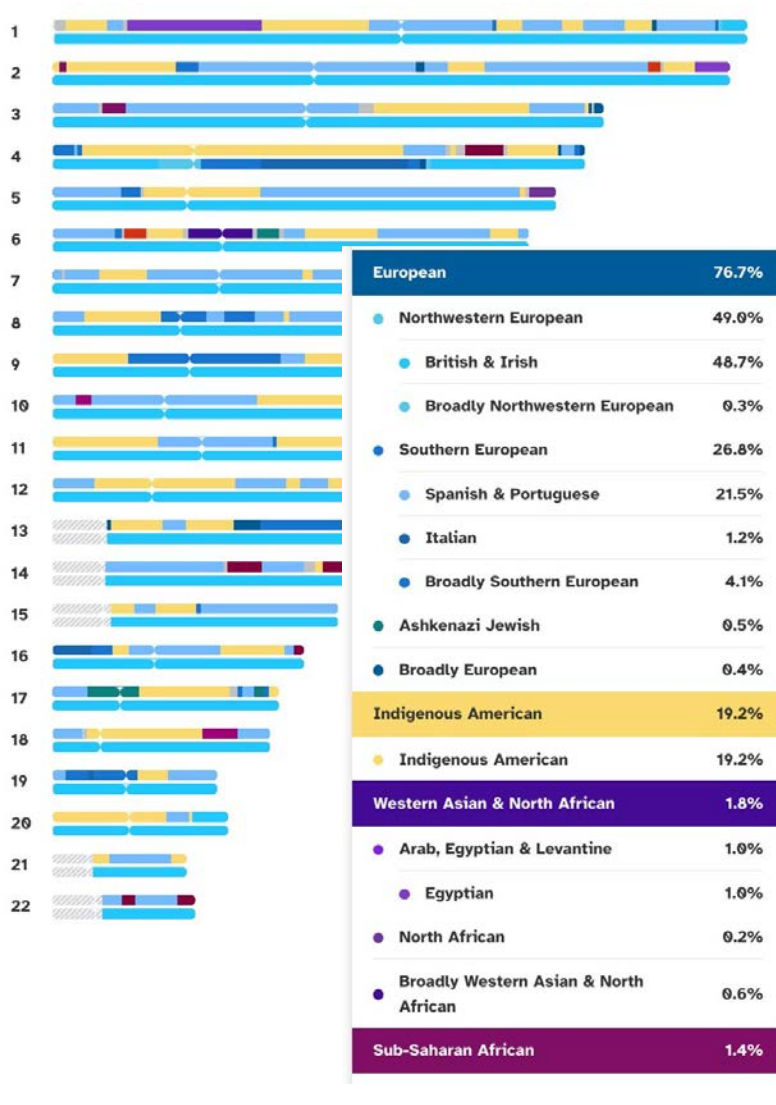


23andMe

The American population is admixed

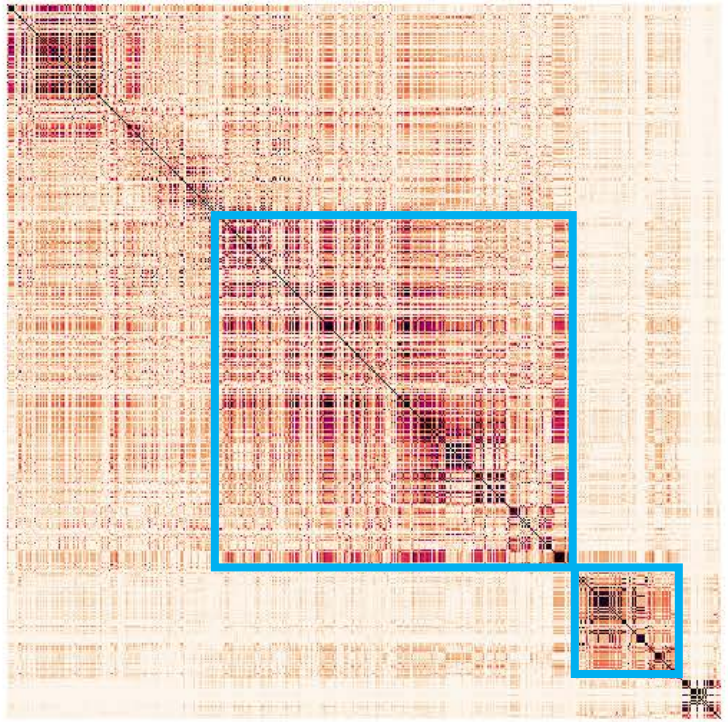




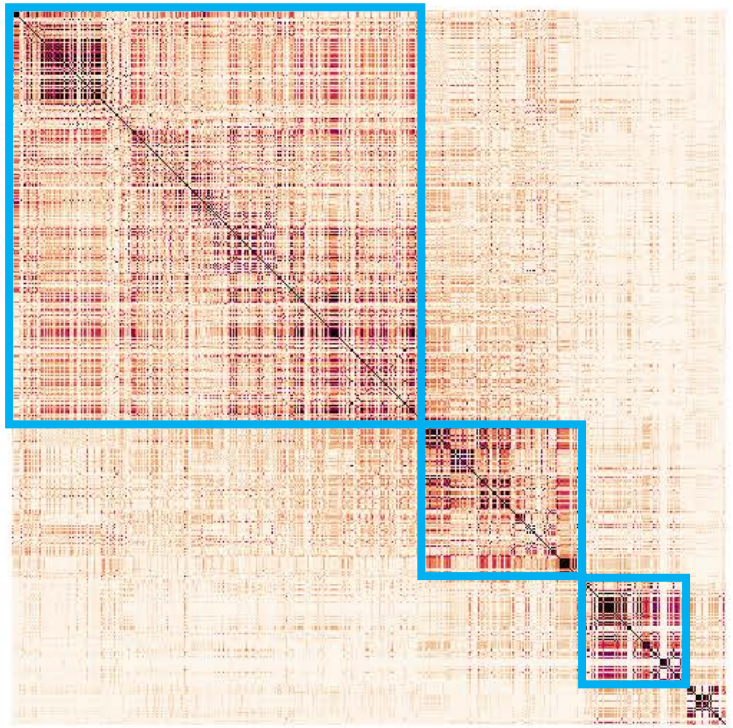


LD structure changes between populations

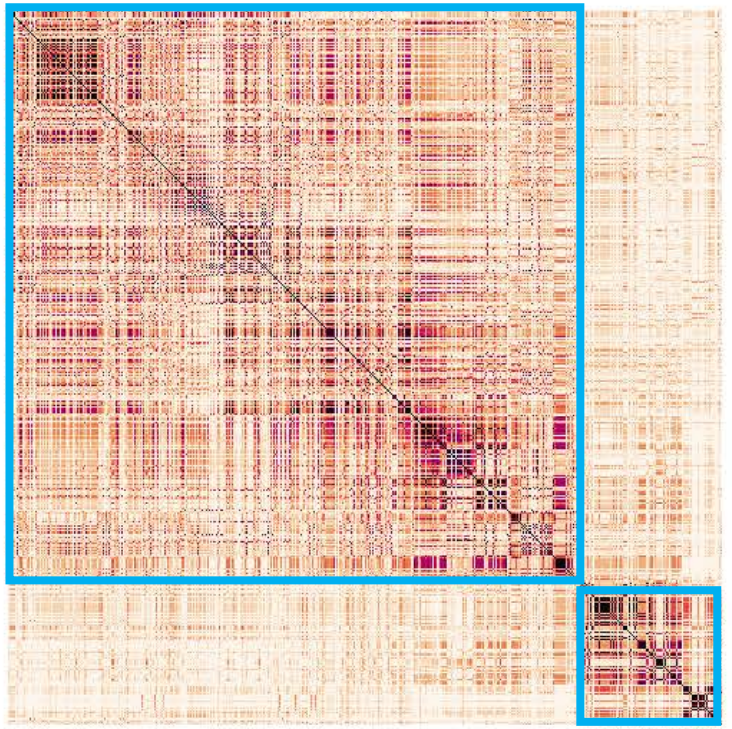
HLA-DQB1: AFR



HLA-DQB1: EAS



HLA-DQB1: EUR



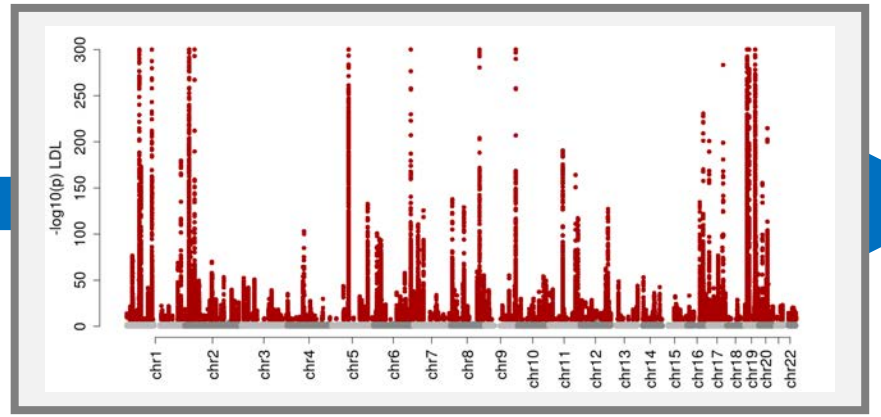
LD = linkage disequilibrium

Stratify individuals based on disease risk

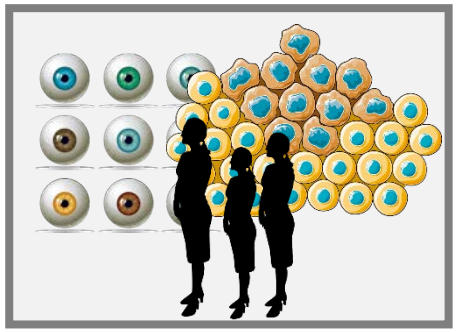
Genetic variation



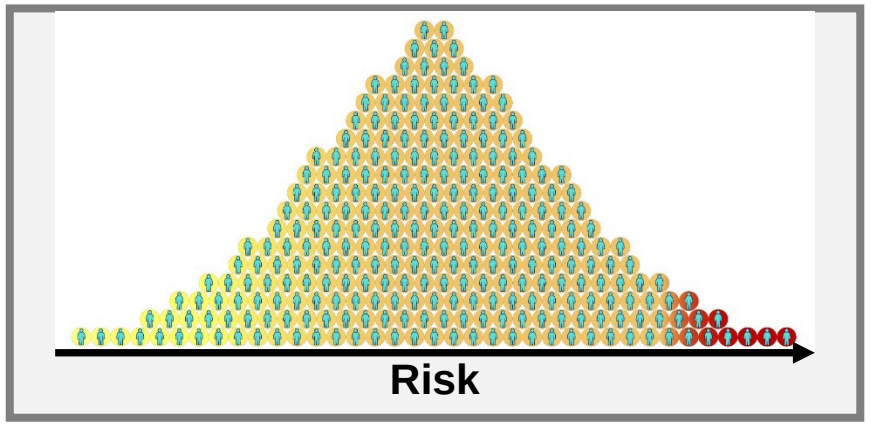
Genome-wide association studies (GWAS)



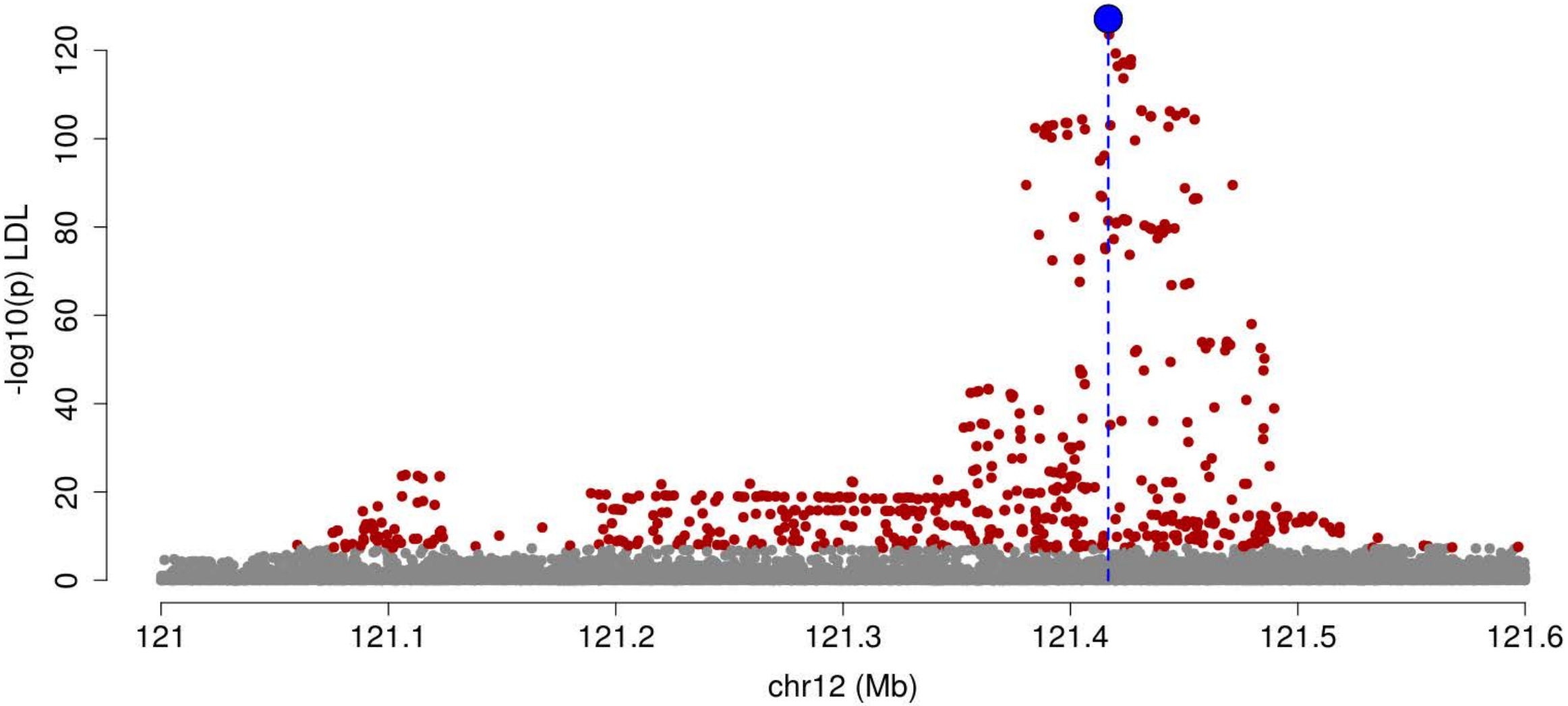
Phenotype



Stratify individuals based on risk
Polygenic risk score (PRS)



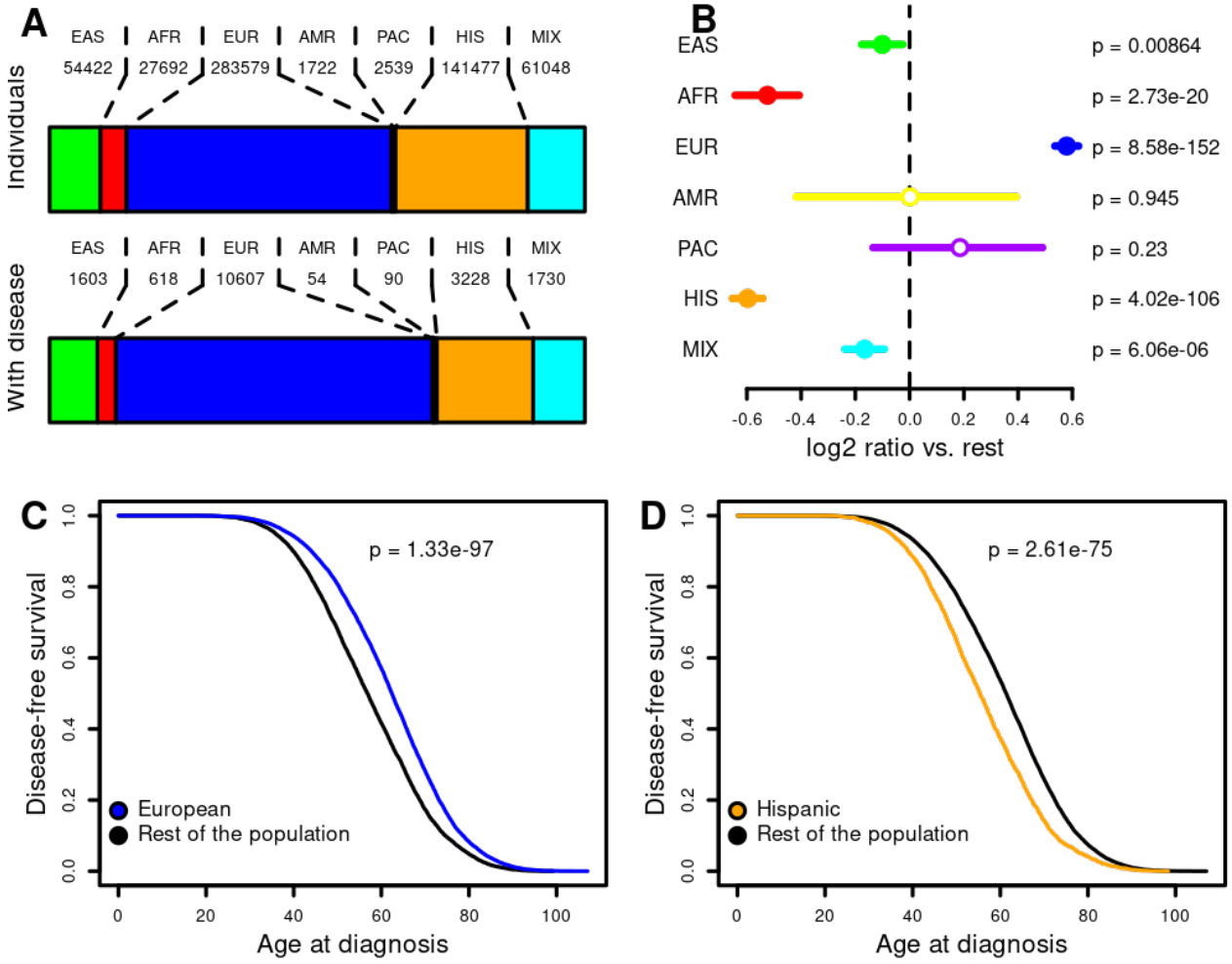
Genome-wide association studies (GWAS)

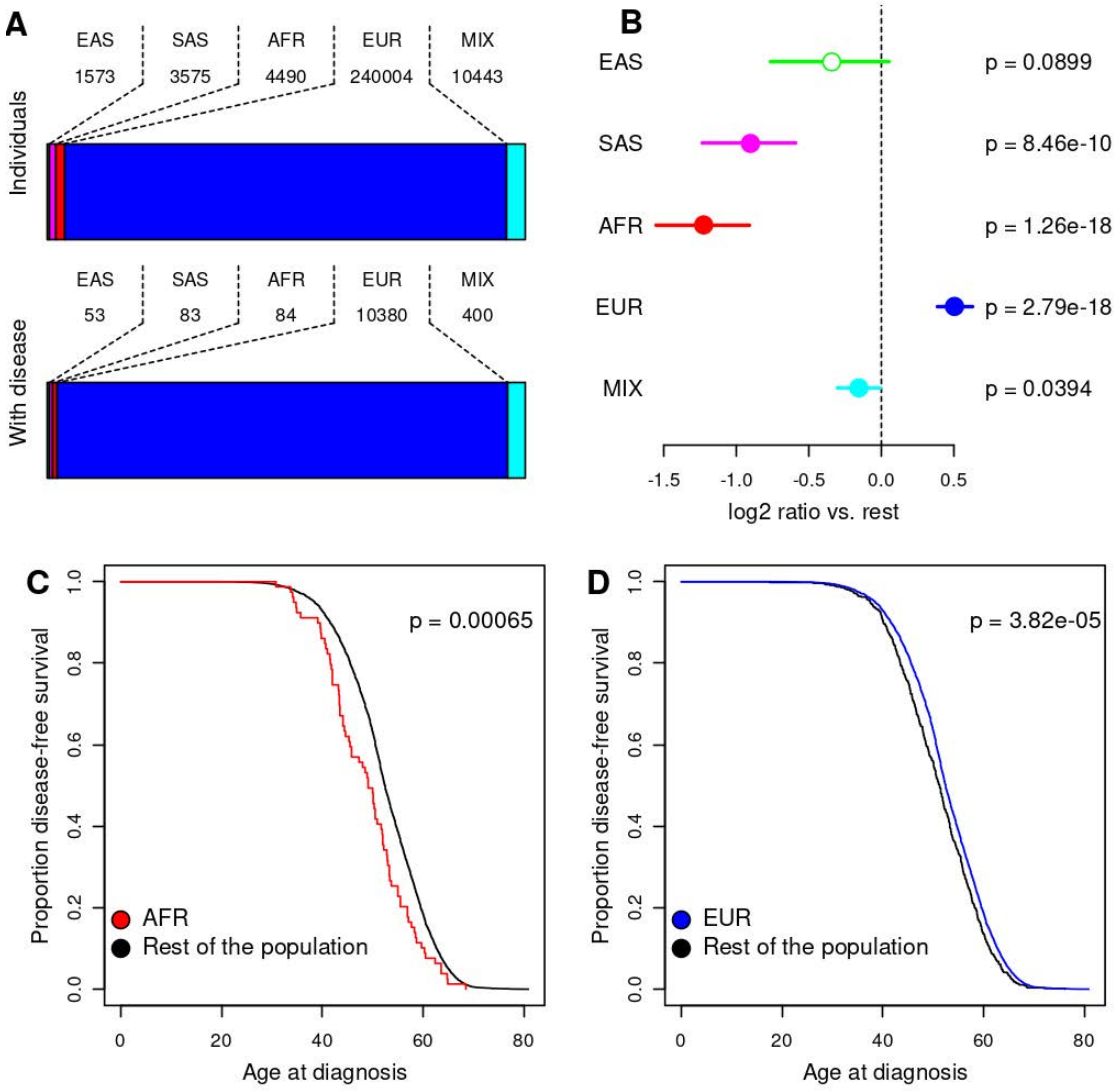


Multiethnic local genotype PCA improves polygenic risk scores for breast cancer

- Investigate the differences in breast cancer risk across populations
  - UCSD Epic EHR data
  - UK Biobank
- Perform a multiethnic GWAS
  - UK Biobank
- Develop local genotype PCA (lPCA) GWAS
- Test changes in PRS between standard and lPCA GWAS
  - UK Biobank
  - All Of Us

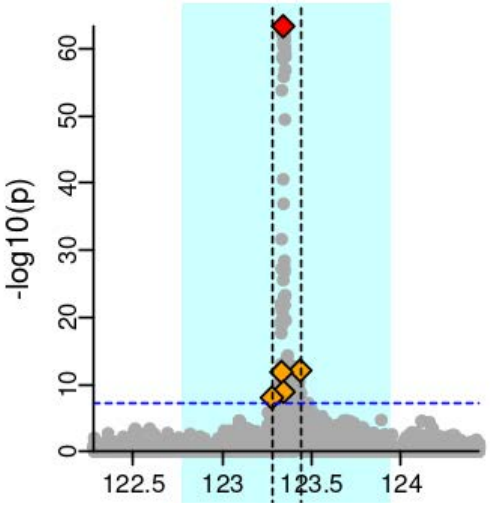
Individuals of European ancestry have higher incidence and a later age at diagnosis

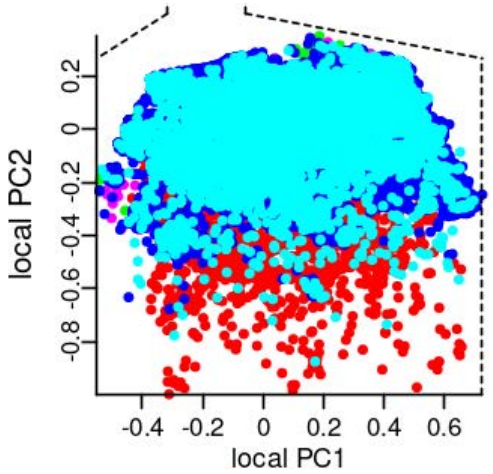




UCSD Epic: 662,859 females (18,905 breast cancer patients)
All Of Us: 247,453 females (11,897 breast cancer patients)
UK Biobank: 260,085 females (11,000 breast cancer patients)

Multiethnic GWAS on the UK Biobank identifies 27 signals in 18 loci





Traditional GWAS

$$p_i = \frac{1}{1 - exp\left(\beta_j X_{ij} + \sum_{p=1}^{P} \gamma_p C_{ip} + \epsilon_{ij}\right)}$$

p_i = probability of breast cancer for individual i
β_j = effect size (fixed effect) of SNP j

X_{ij} = genotype of individual i at SNP j

C_{ip} = value of the p^{th} covariate for individual i
γ_p = effect size of the p^{th} covariate
ϵ_{ij} is the error term for individual i at SNP j

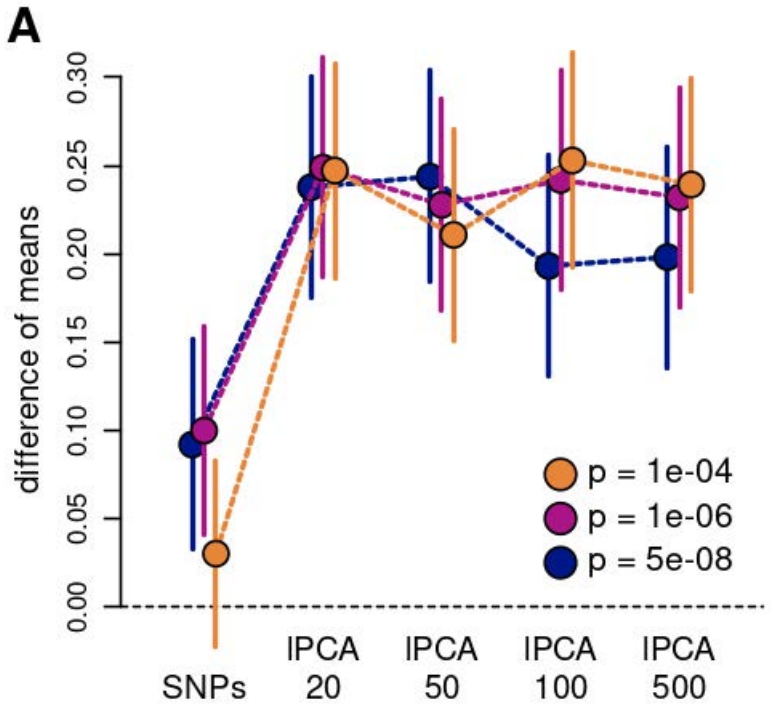
Our proposed approach

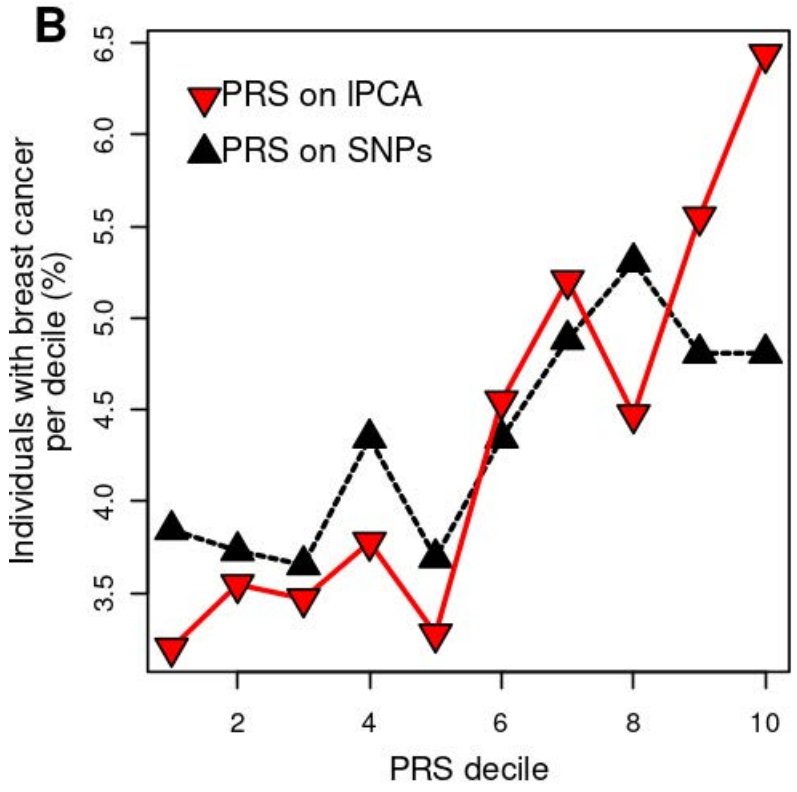
$$p_i = \frac{1}{1 - exp\left(\sum_{k=1}^{K} \beta_{jk} D_{ijk} + \sum_{p=1}^{P} \gamma_p C_{ip} + \epsilon_{ij}\right)}$$

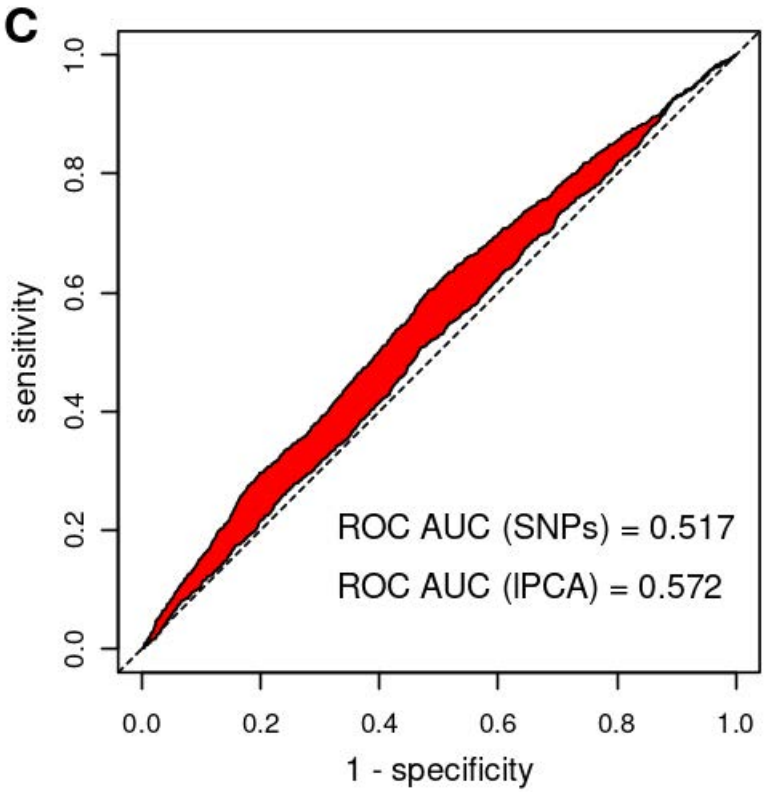
p_i = probability of breast cancer for individual i
β_{jk} = effect size (fixed effect) of dimension k in haplotype block j
D_{ijk}= Value of individual i for dimension k in haplotype block j
C_{ip} = value of the p^{th} covariate for individual i
γ_p = effect size of the p^{th} covariate
ϵ_{ij} is the error term for individual i at haplotype block j

Polygenic risk scores improve when using lPCA

- Standard GWAS: PRS calculated using pruning and thresholding (PLINK)
- lPCA: PRS calculated using the lPC with the strongest effect at each locus
- Tested different p-value thresholds







Summary and next steps

- EUR individuals have a higher breast cancer risk at a later age
- lPCA GWAS improves PRS calculation

Hypothesis: rich individuals have better access to care and are more likely to undergo hormone therapies, which may affect cancer risk

Combine SNPs with lPCA into PRS

Add environmental, lifestyle and SDOH factors to PRS

Combine all into a new PRS model

Acknowledgements

Gymrek Laboratory

Melissa Gymrek

Arya Massarat

Jonathan Margoliash

Michael Lamkin

Wilfredo Gonzalez Rivera

Frazer Laboratory

Kelly Frazer

Jennifer Nguyen

Tim Arthur

CAST

Center for Admixture Science and Technology

Ramya Rajagopalan

Lucila Ohno-Machado

Tiffany Amariuta

Alon Goren

Vineet Bafna

Amy Sitapati

Former CAST members

Daniela S. Perry

Andrew P. Hodges

Isaac Joshua

Woo-Yeong Park

Ximei Wu

BG113996

NIH

P30CA023100
HG011558

American Cancer Society®

IRG-19-230-48-IRG