



**HKU  
Med** School of Clinical Medicine  
Department of Surgery  
香港大學外科學系



Dr Li Dak-Sum Research Centre  
The University of Hong Kong - Karolinska Institutet  
Collaboration in Regenerative Medicine  
香港大學 - 卡羅琳斯卡學院再生醫學合作計劃  
李達三博士研究中心

# Achievements, opportunities and challenges in Asian GWAS

**Clara Tang**

9<sup>th</sup> June 2022

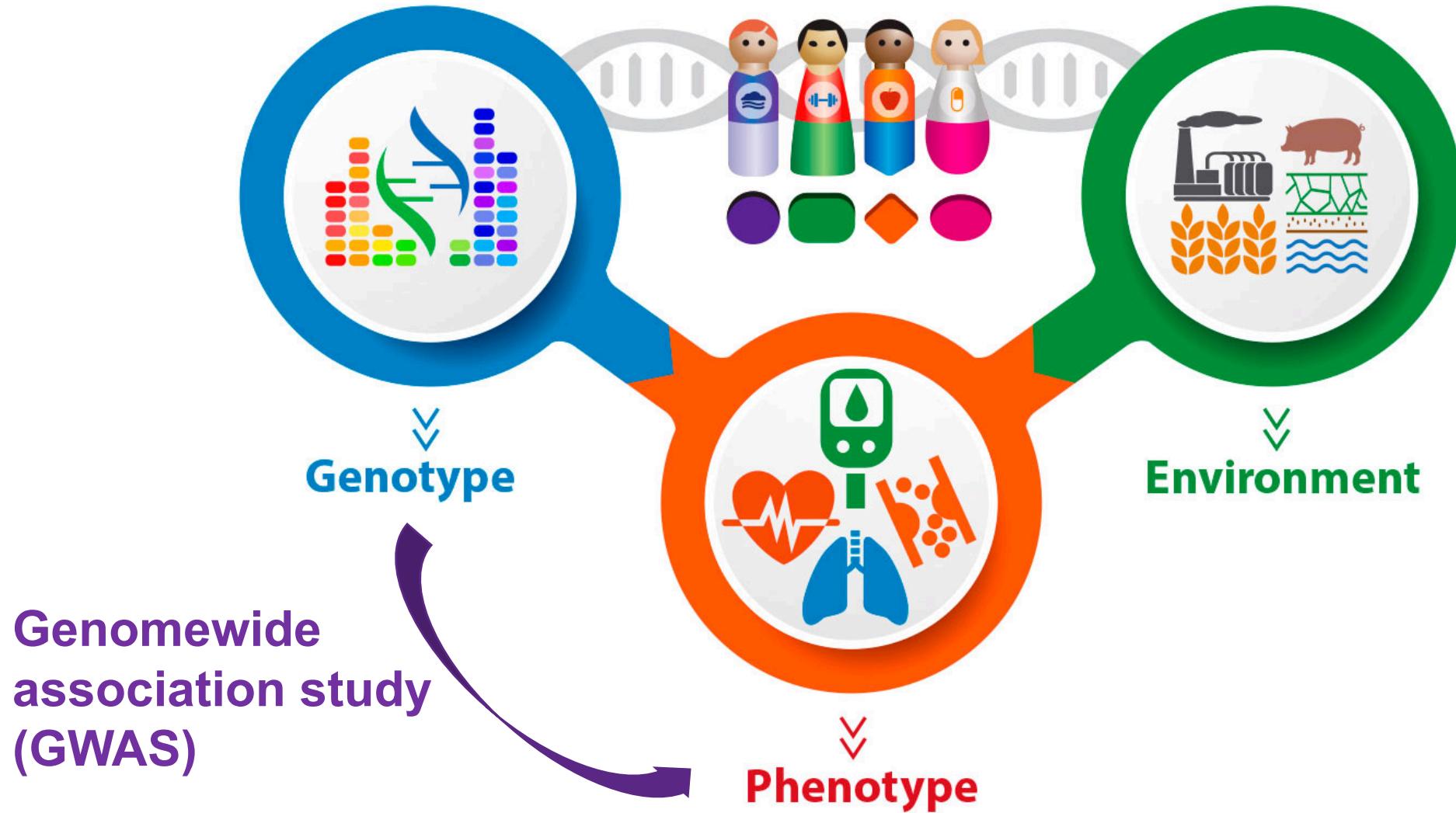


# Outline

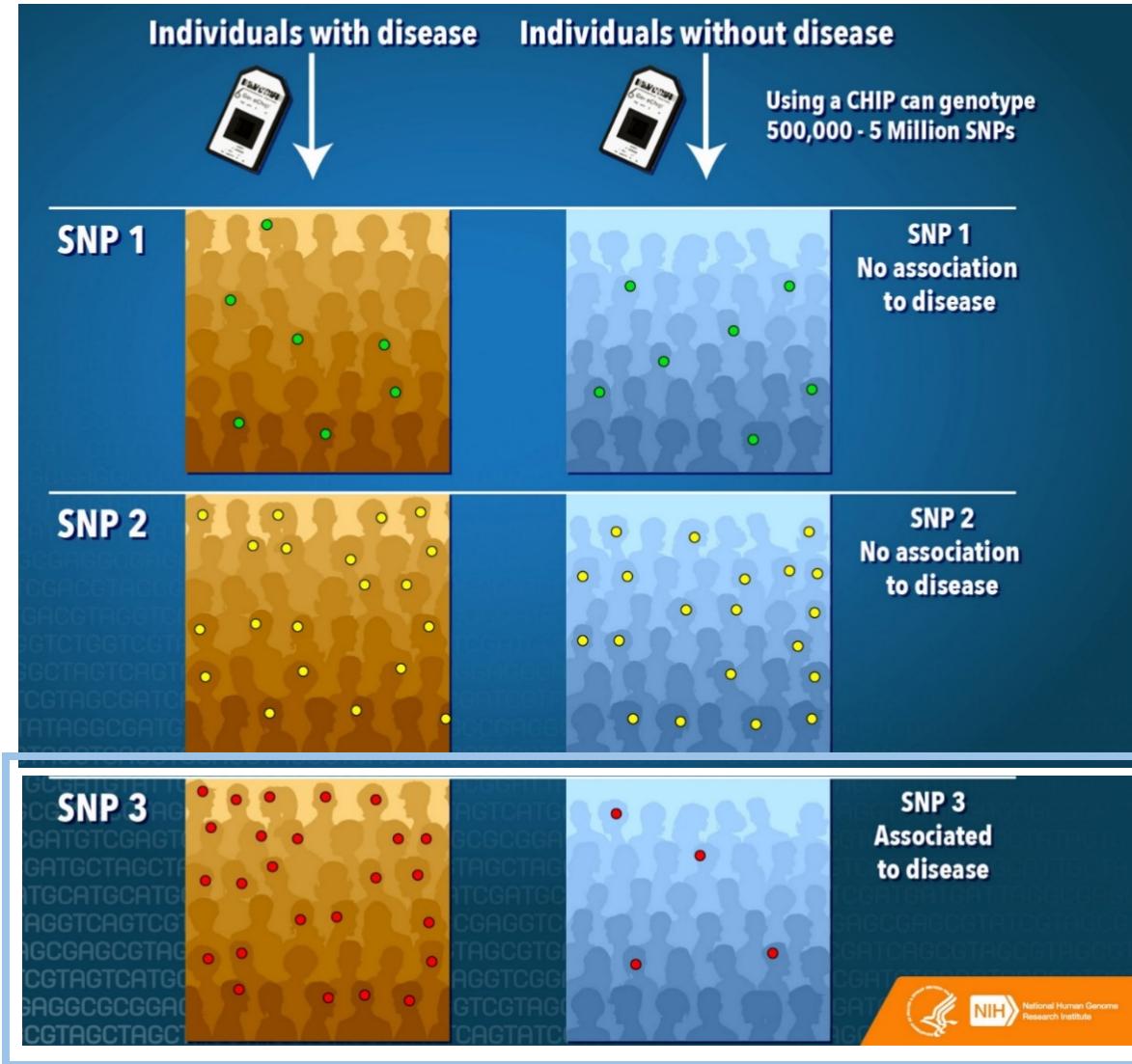
- **Genetic variations and GWAS**
- **Genetic diversity: Asian population**
- **Opportunities and challenges of Asian GWAS**



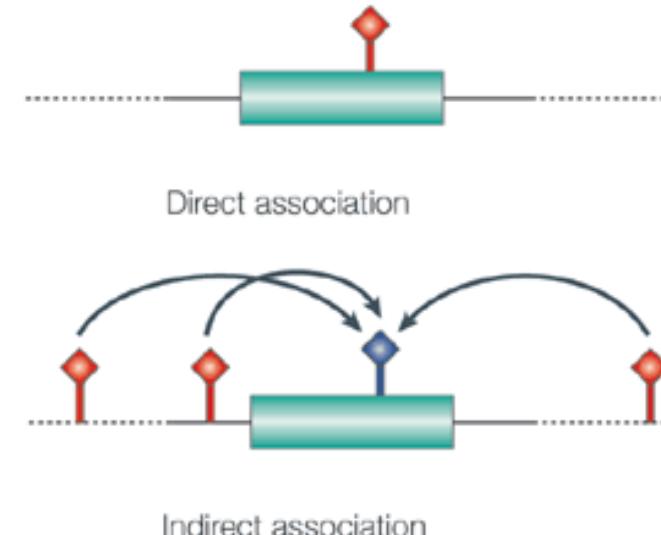
# Genetics, environment and phenotypes



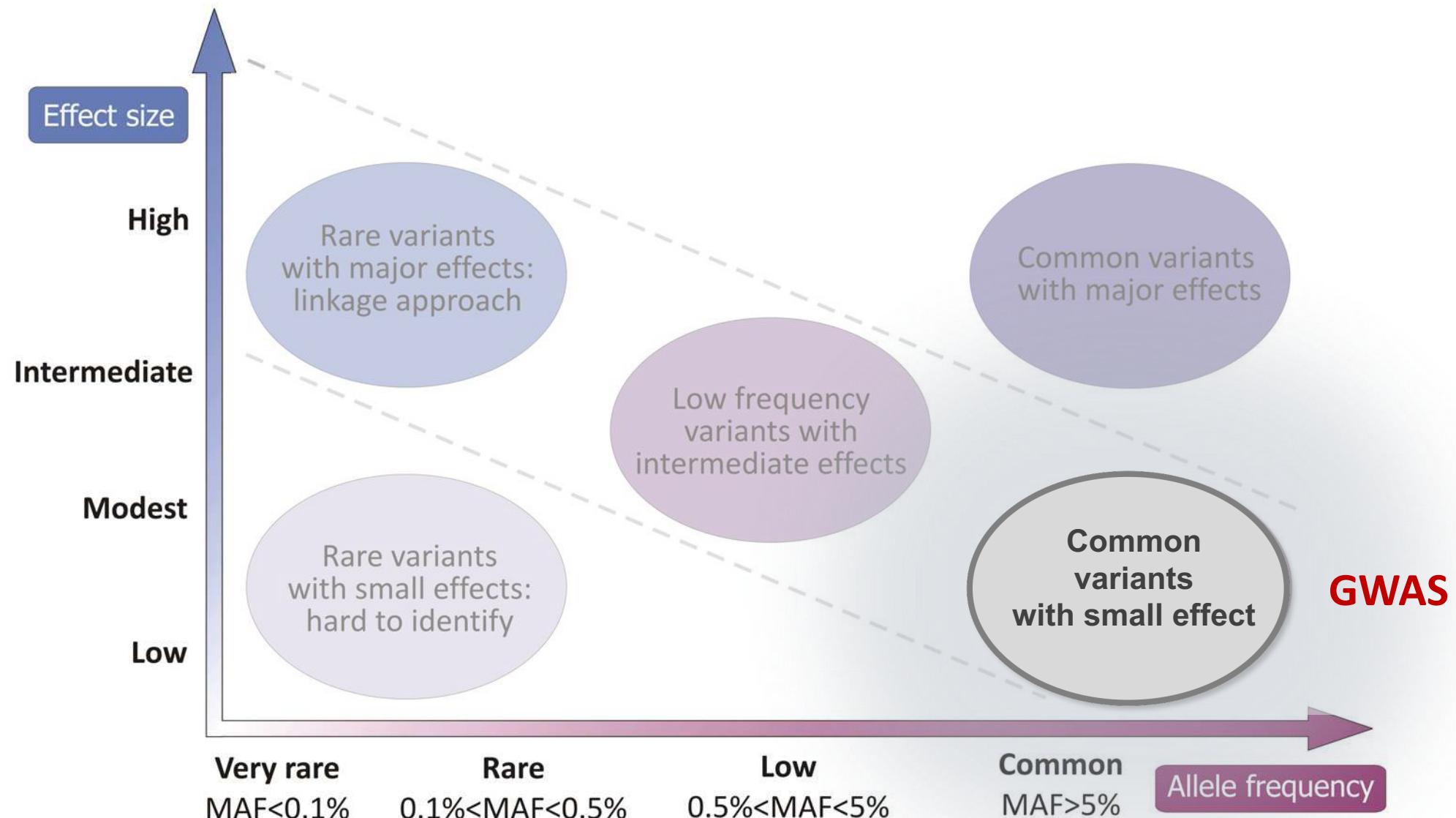
# Genomewide association study (GWAS)



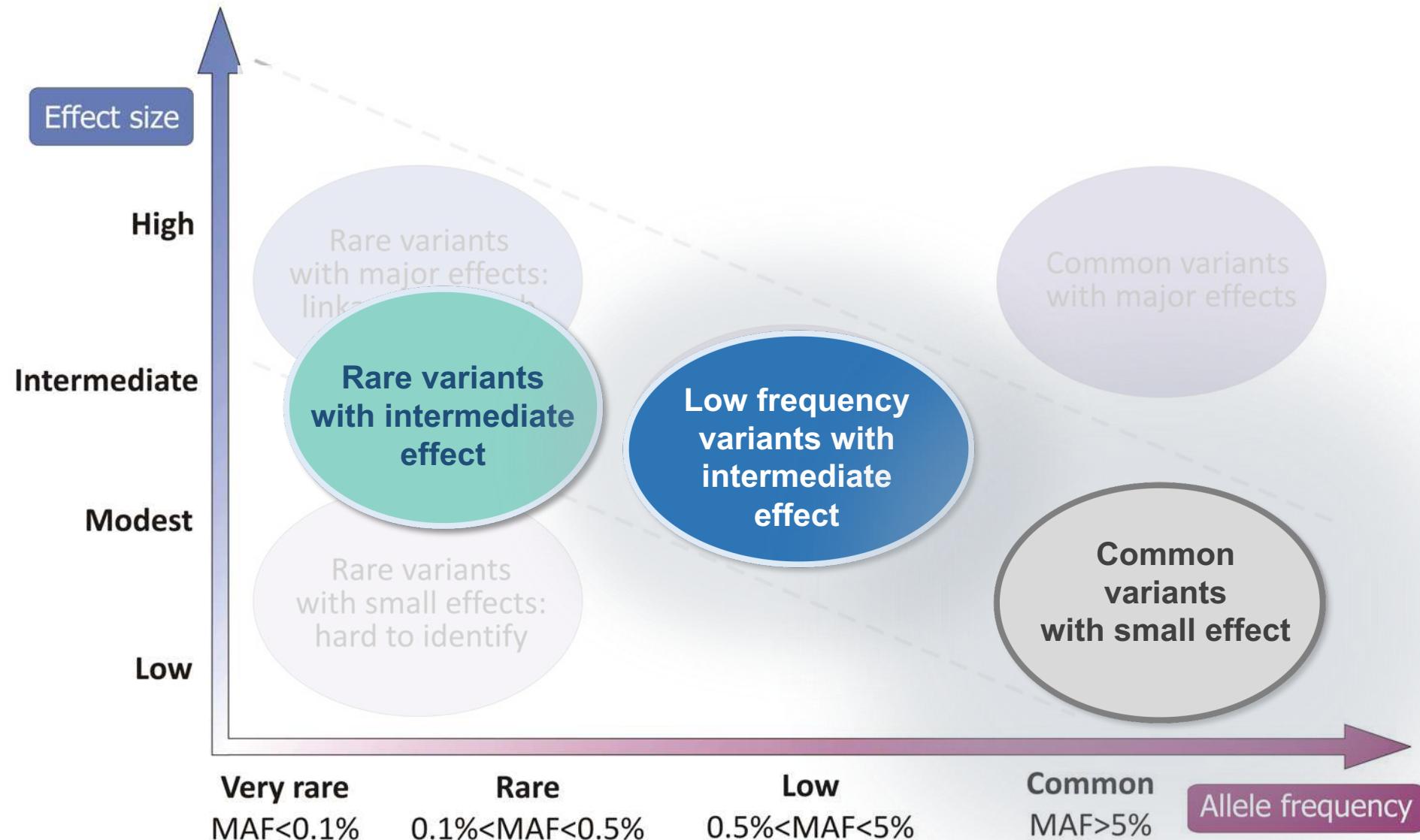
- Comparing frequency of alleles or genotypes between cases and controls/across quantitative traits
- An unbiased approach to detect association of variants in **linkage disequilibrium (LD) with the causal variants**



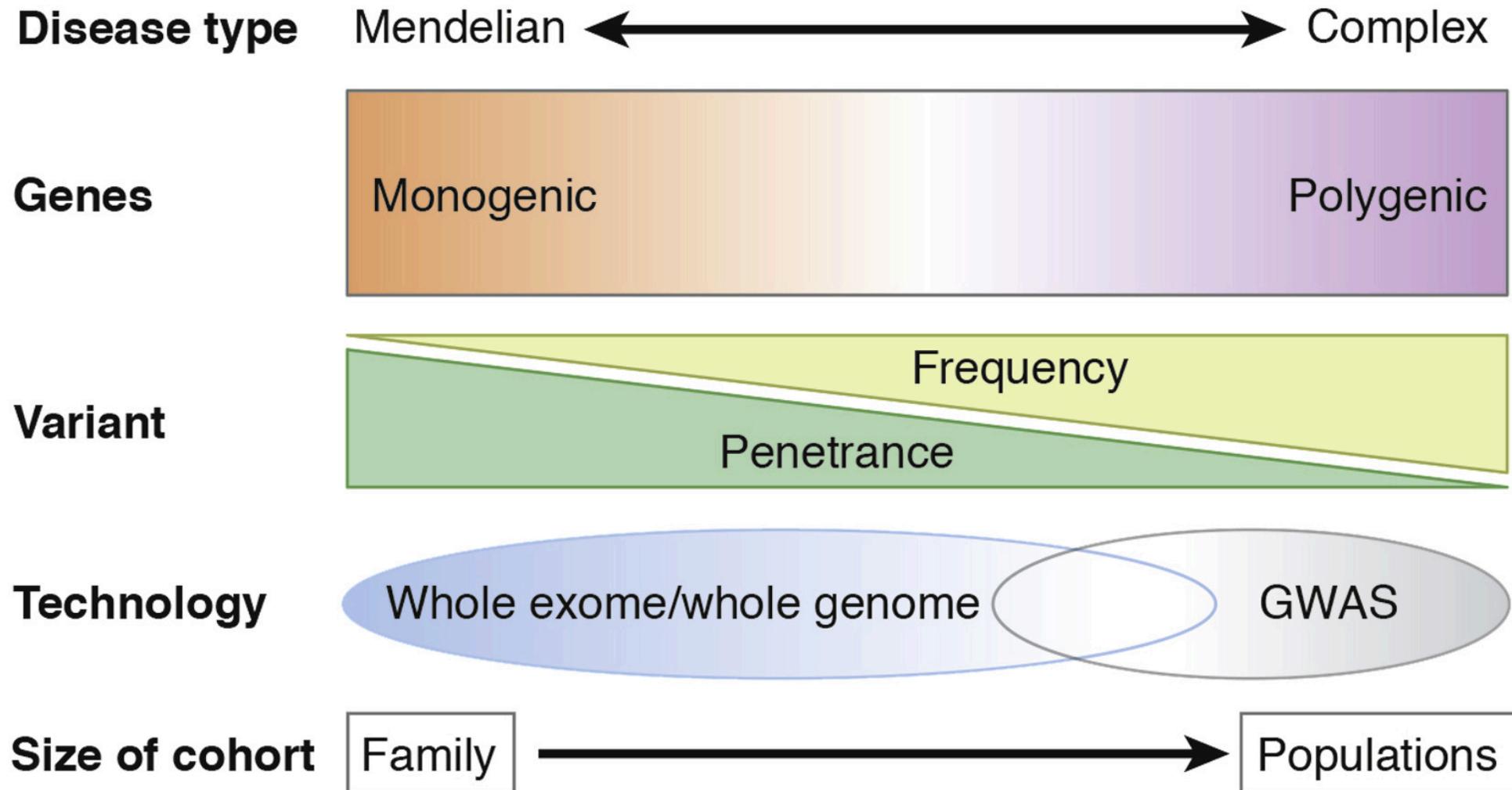
# GWAS of common and low frequency variants



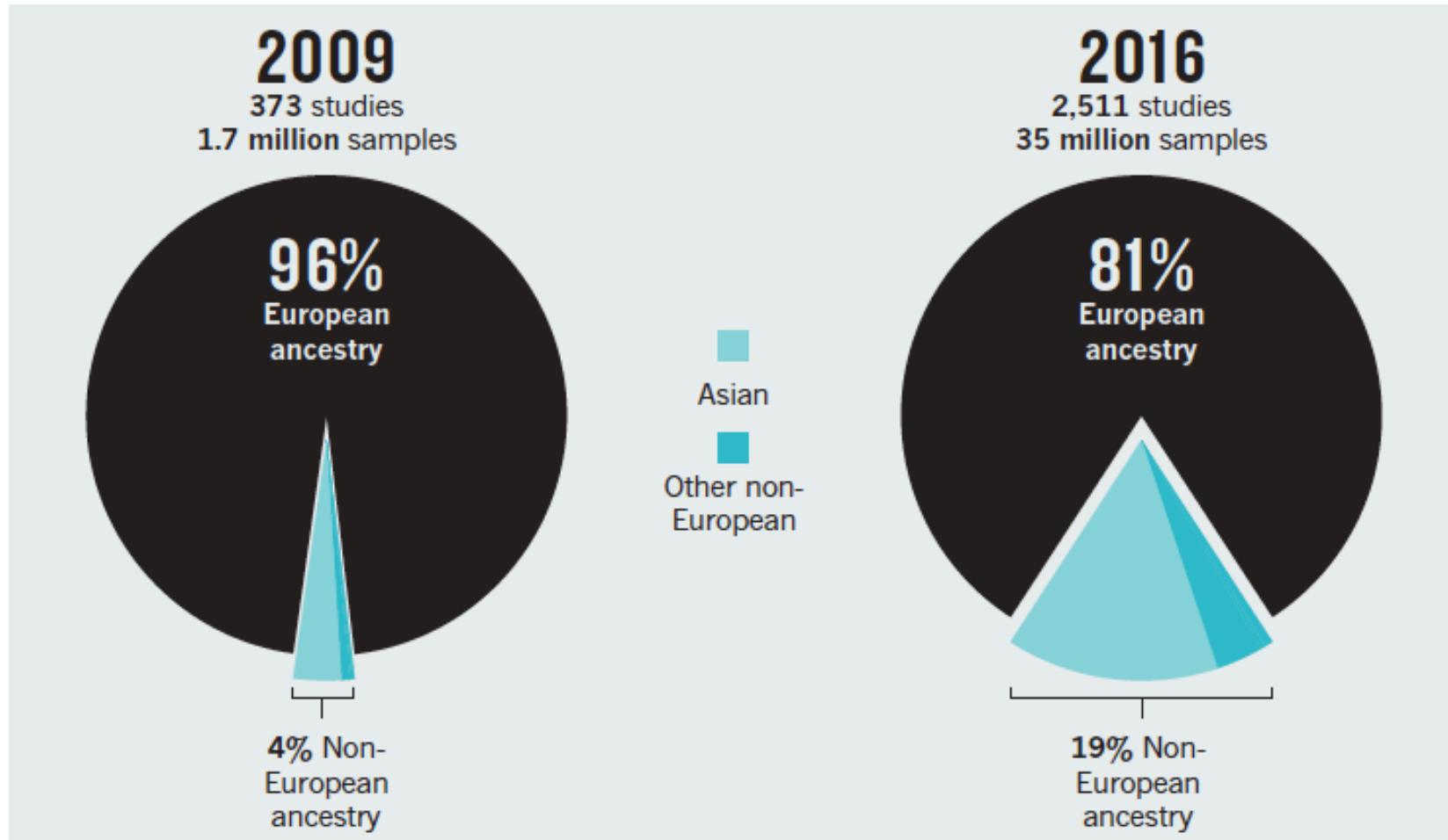
# GWAS of common and low frequency variants



# GWAS for complex traits and disorders



# Under-representation of Asian GWAS



3% are Asian samples

11-14% are Asian samples

# Under-representation of Asian GWAS

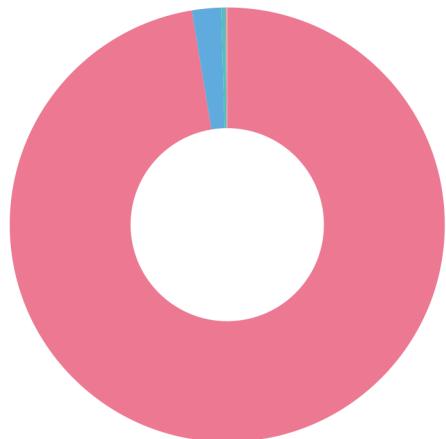


GWAS

Diversity Monitor

Participants by ancestry

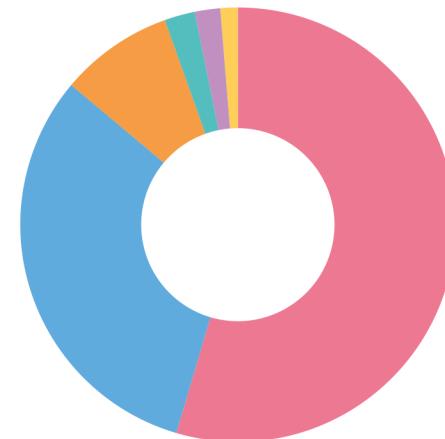
Discovery Stage - All parent terms - 2021



European  
97.39% (Asso. 54.57%)  
Asian  
2.17% (Asso. 31.55%)  
African  
0.03% (Asso. 8.44%)

Count of all associations discovered

Discovery Stage - All parent terms - 2021



African American or Afro-Caribbean 0.02% (Asso. 1.3%)  
Hispanic or Latin American 0.08% (Asso. 1.89%)  
Other/Mixed 0.31% (Asso. 2.26%)

Please cite this as: Mills, M.C and Rahal, C., (2020). 'The GWAS Diversity Monitor Tracks diversity by disease in real time'. Nature Genetics, 52, 242-243. doi: 10.1038/s41588-020-0580-y

<https://gwasdiversitymonitor.com/>

# Impact of lack of genetic diversity in GWAS

## Discovery

- Missing **association of population-specific variants** or variants relatively more common in non-European populations
- Missing **gene-by-environment interaction** for variant with different responses to environmental factors

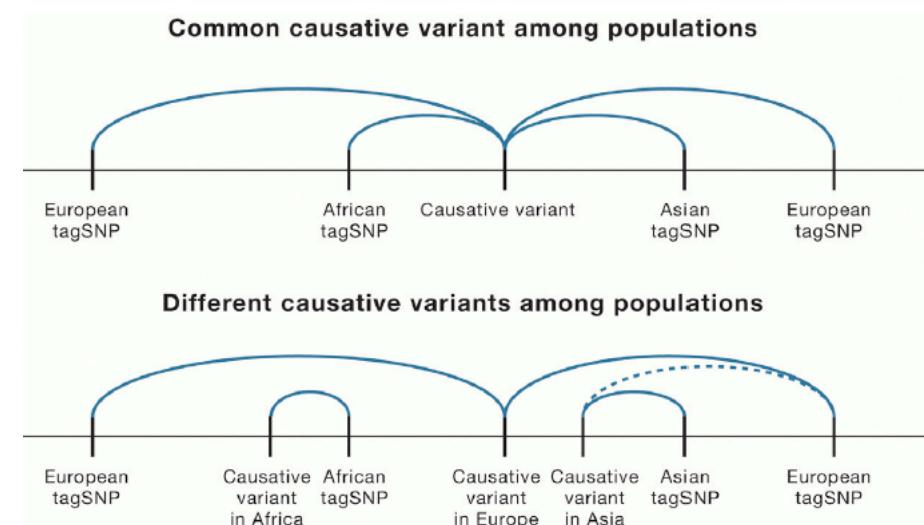


Genetic signatures of high-altitude adaptation in the Himalayas

Hackinger et al. Hum Genet 2106; 135:393–402.

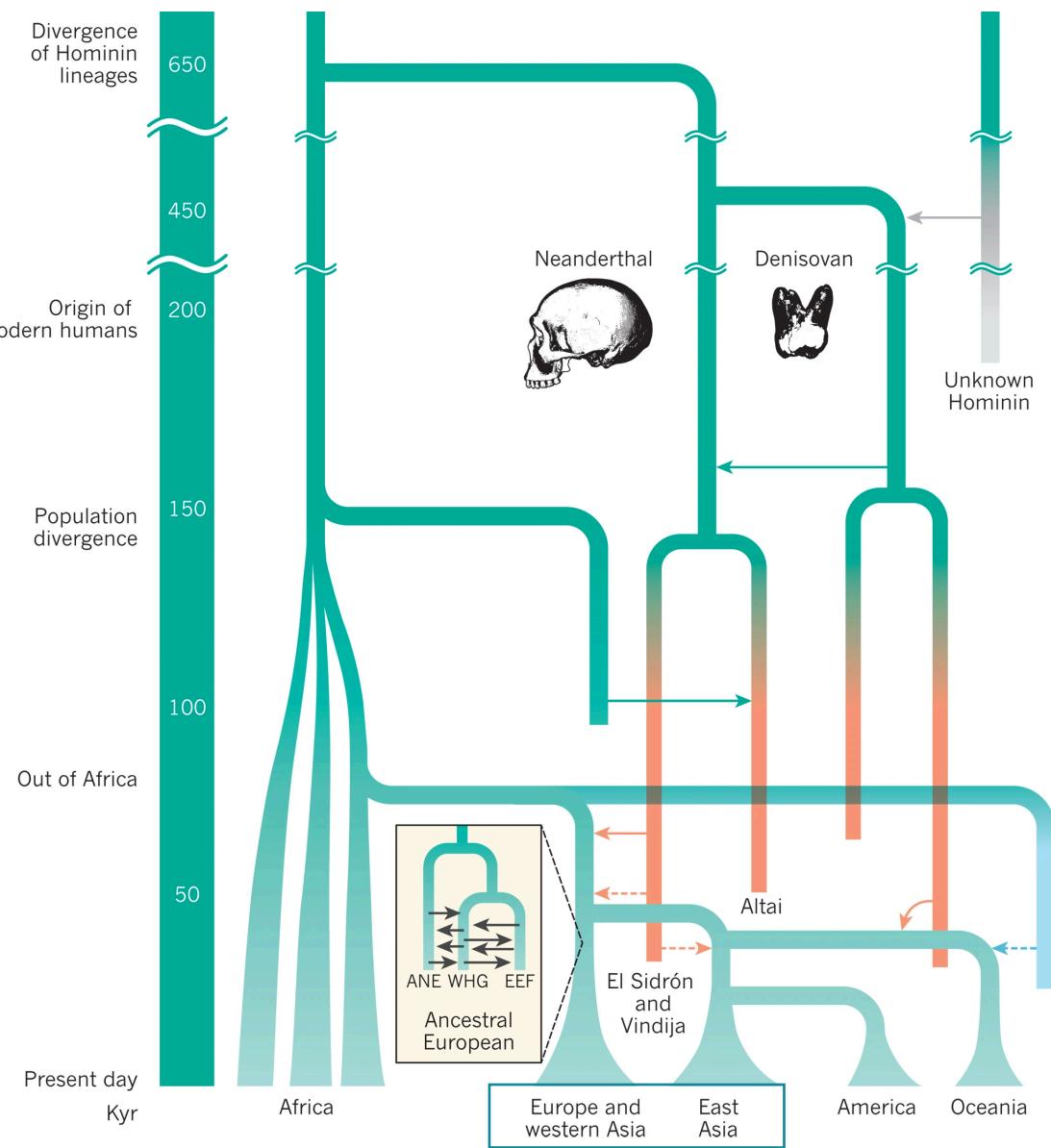
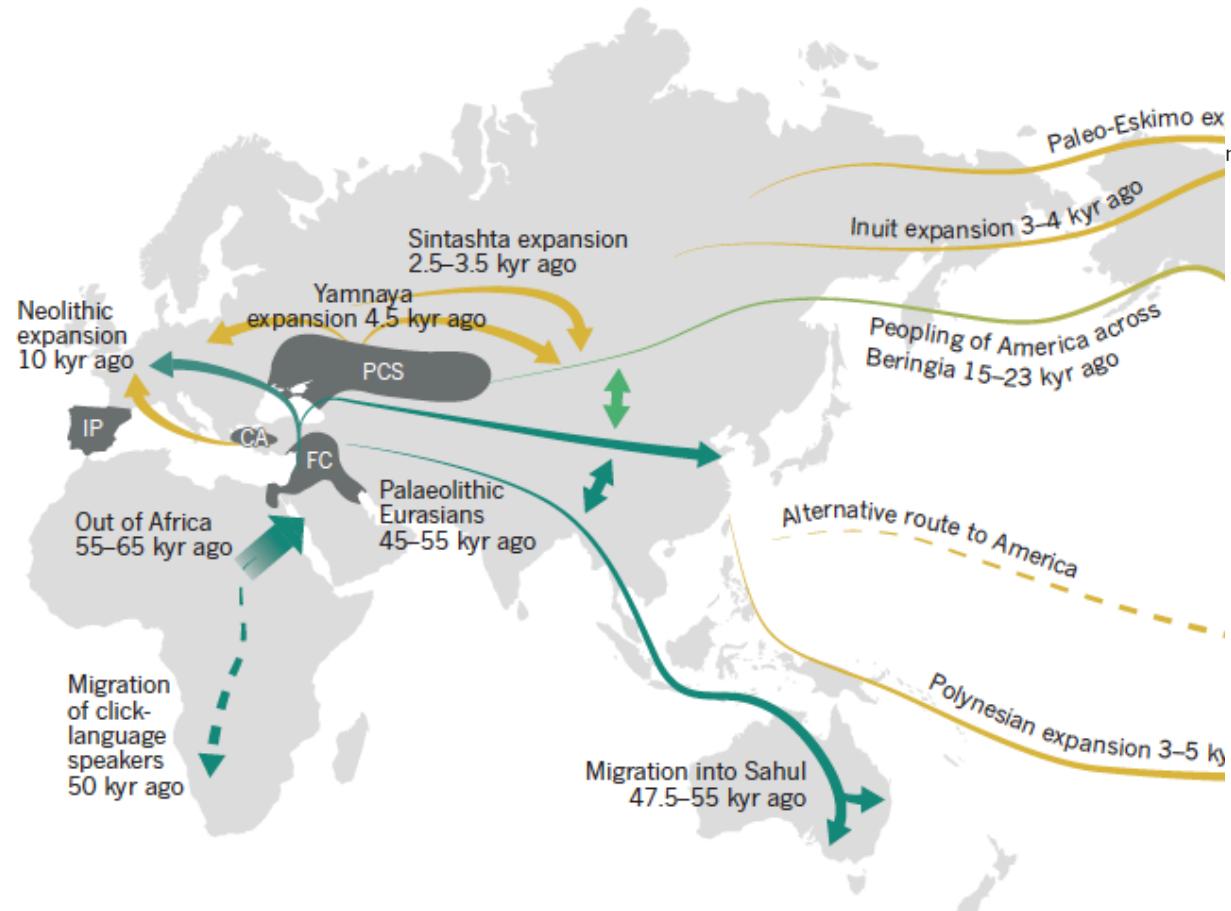
## Clinical application

- **Reduced translatability** of genetic research into clinical practice
  - pharmacogenetic variants associated with drug response
- **Reduced accuracy of polygenic risk score (PRS) estimation**

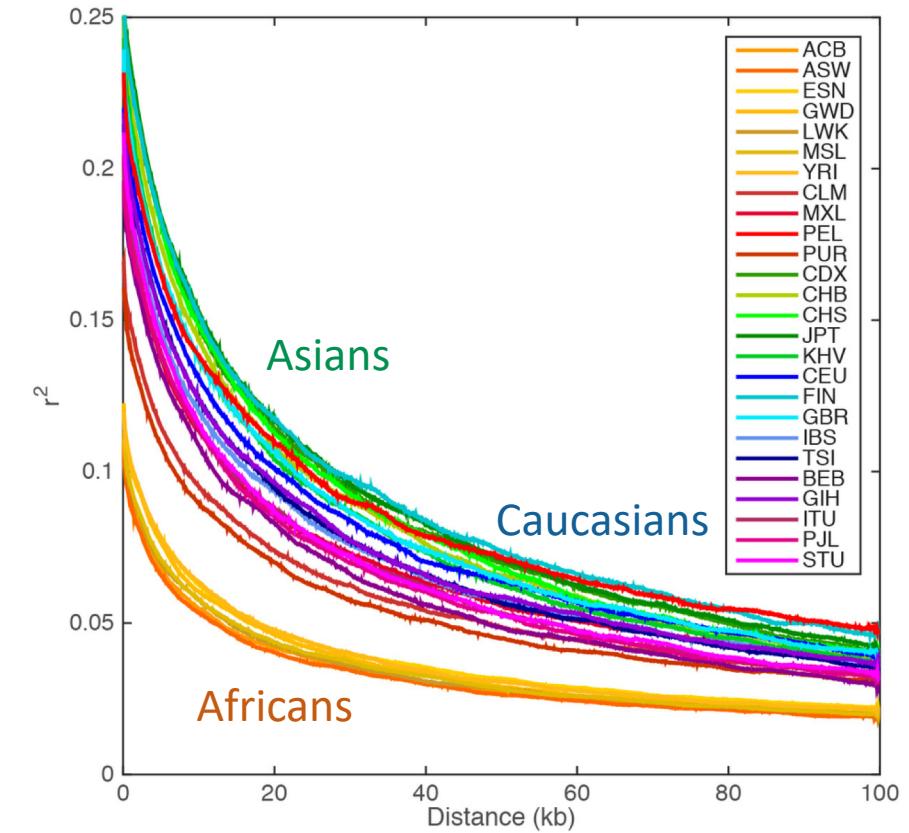
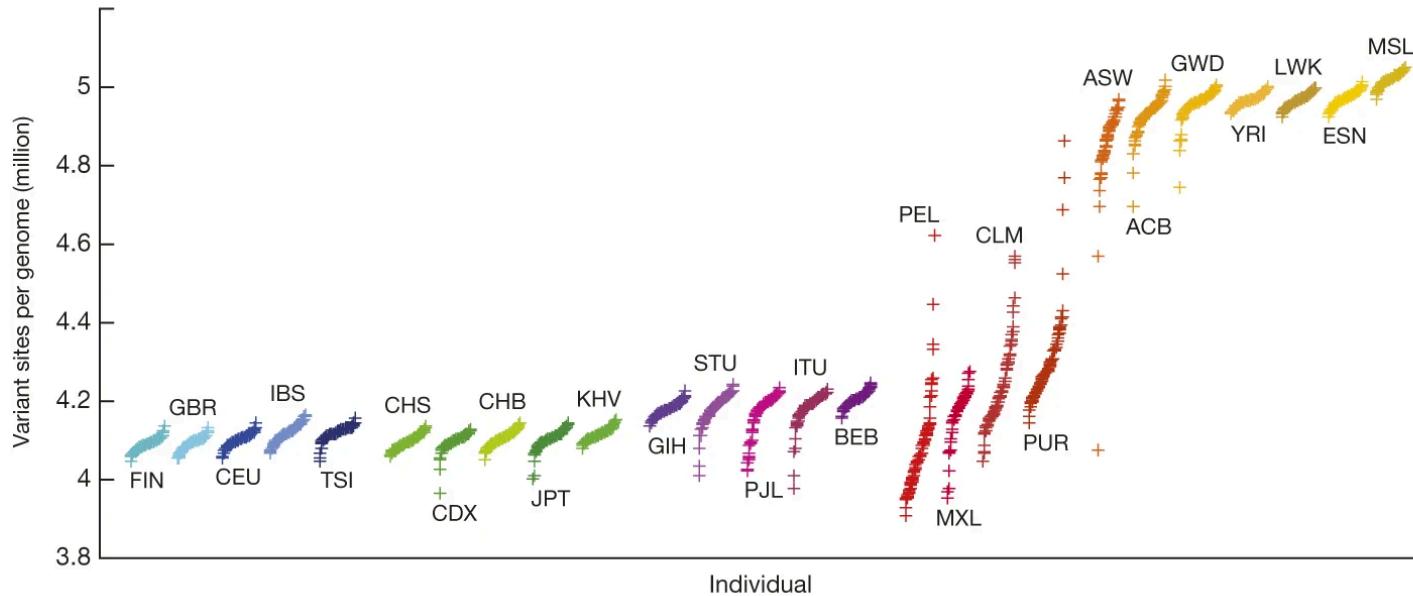


Sirugo et al. Cell. 2019;177(1):26-31.

# Human migration history



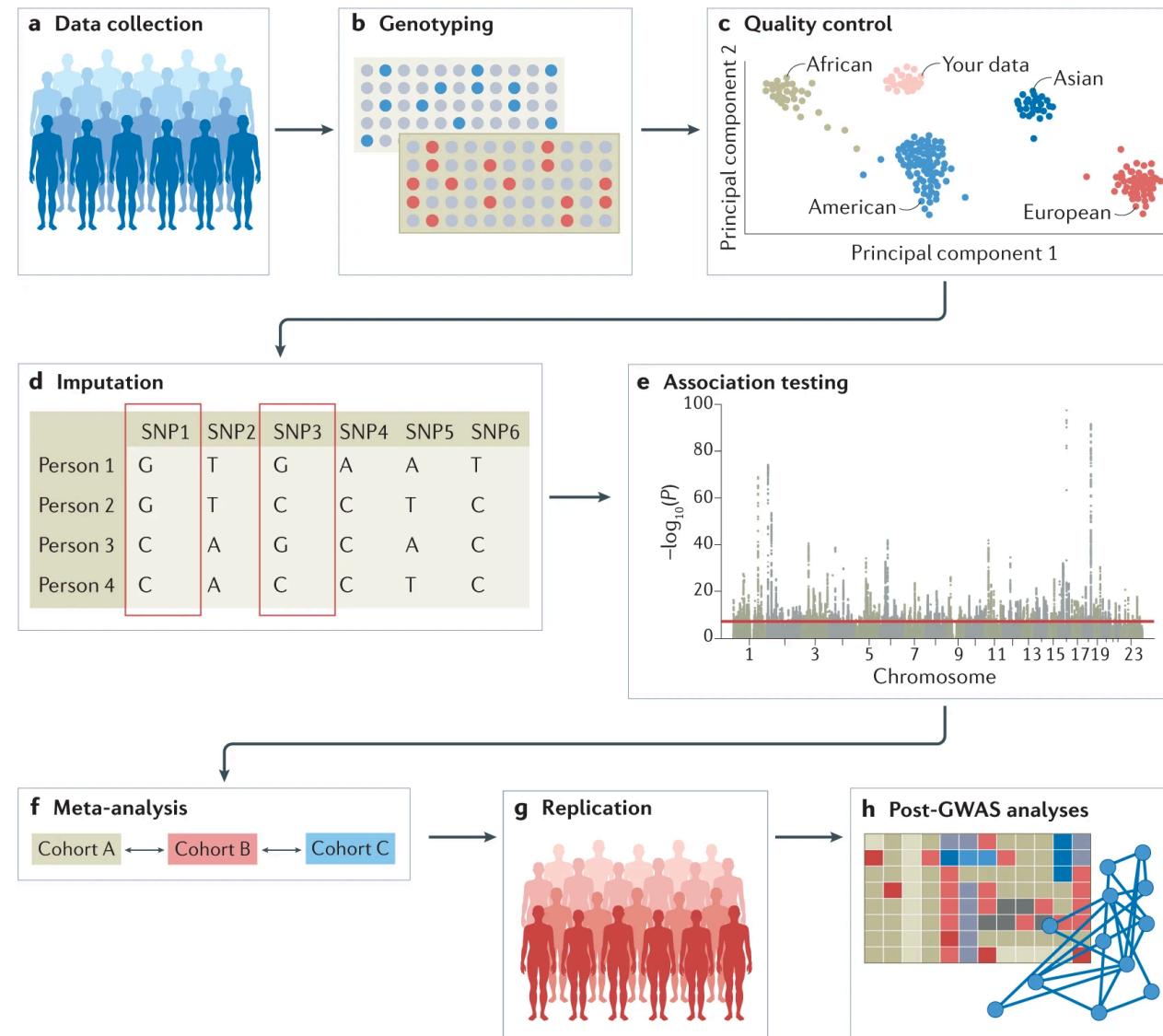
# Genomic signatures of Asians



Asians have **fewer variants** on average than African and African American populations

Asians have **stronger LD** and **longer LD block**

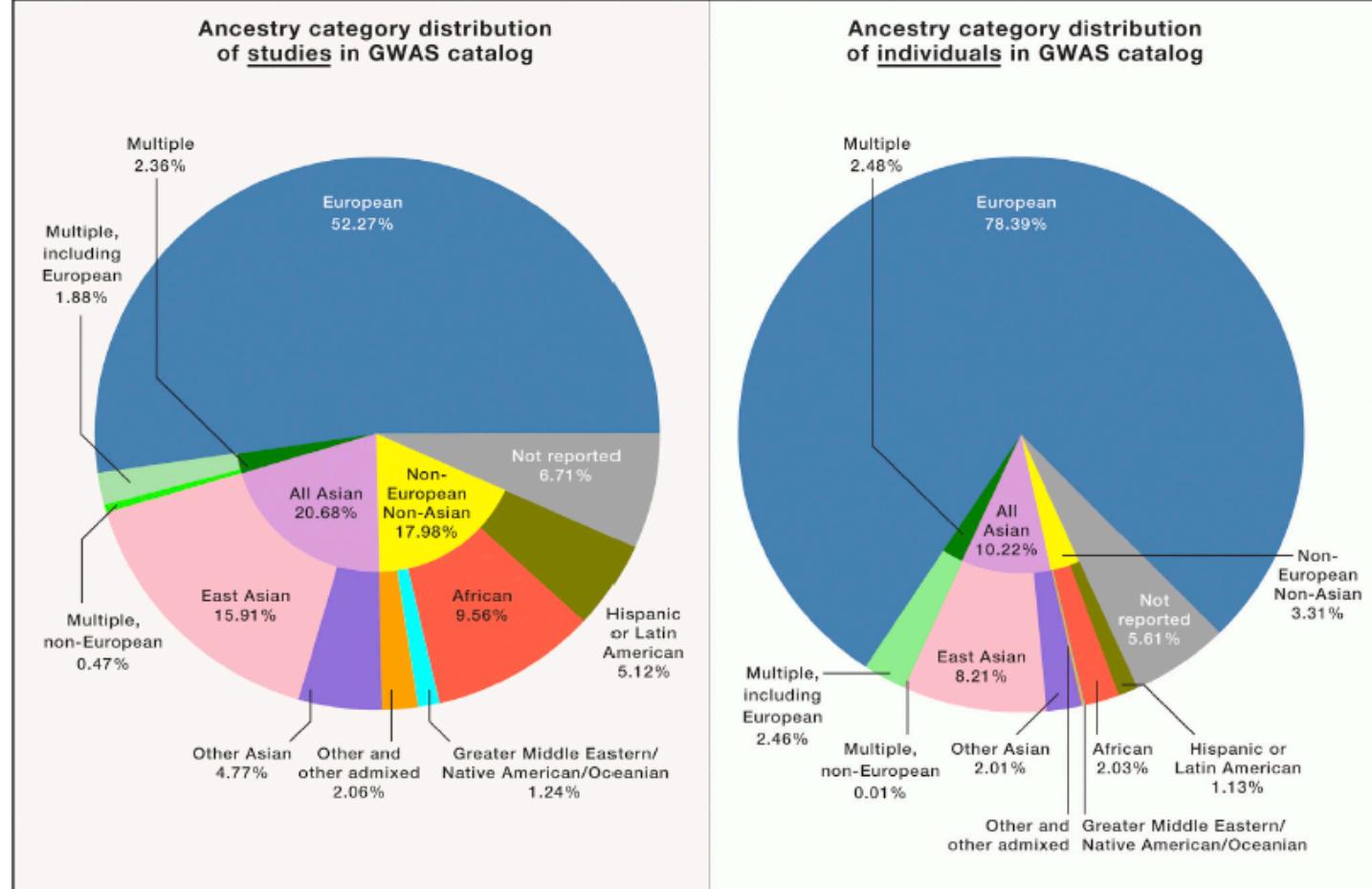
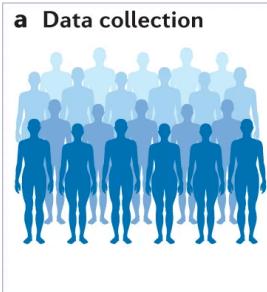
# Challenges and opportunities of Asian GWAS



# Challenges and opportunities of Asian GWAS

## Challenges

- Small sample size

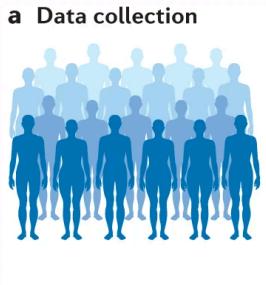


## Opportunities

# Challenges and opportunities of Asian GWAS

## Challenges

- Small sample size



**Table 7 | Number of selected tag SNPs to capture all observed common SNPs in the Phase I HapMap**

$r^2$ threshold*	YRI	CEU	CHB + JPT
$r^2 \geq 0.5$	324,865	178,501	159,029
$r^2 \geq 0.8$	474,409	293,835	259,779
$r^2 = 1.0$	604,886	447,579	434,476

Tag SNPs were picked to capture common SNPs in HapMap release 16c1 using the software program Haploview.

\*Pairwise tagging at different  $r^2$  thresholds.

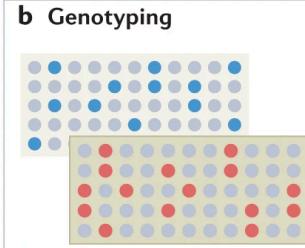
**Table 8 | Proportion of common SNPs in Phase I captured by sets of tag SNPs**

Tag SNP set size	Common SNPs captured (%)		
	YRI	CEU	CHB + JPT
10,000	12.3	20.4	21.9
20,000	19.1	30.9	33.2
50,000	32.7	50.4	53.6
100,000	47.2	68.5	72.2
250,000	70.1	94.1	98.5

As in Table 7, tag SNPs were picked to capture common SNPs in HapMap release 16c1 using Haploview, selecting SNPs in order of the fraction of sites captured. Common SNPs were captured by fixed-size sets of pairwise tags at  $r^2 \geq 0.8$ .

## Opportunities

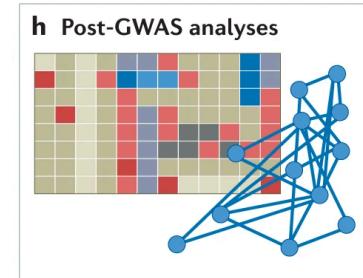
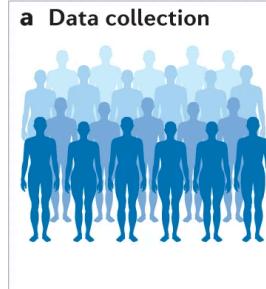
- Fewer variants needed for good coverage of array



# Challenges and opportunities of Asian GWAS

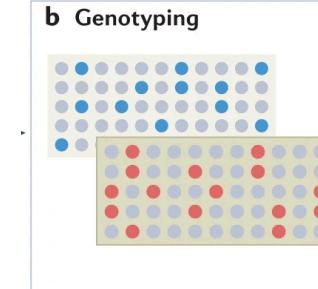
## Challenges

- Small sample size
- **Lower resolution in fine-mapping**



## Opportunities

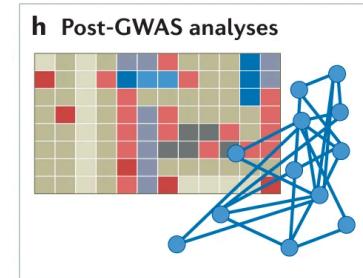
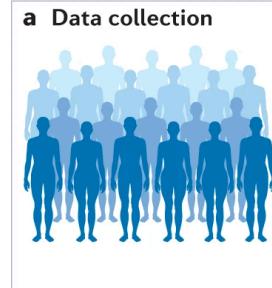
- Fewer variants needed for good coverage of array



# Challenges and opportunities of Asian GWAS

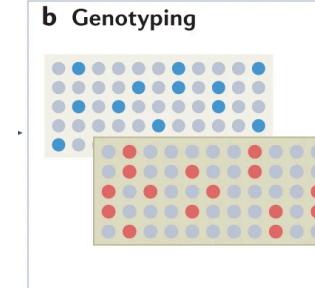
## Challenges

- Small sample size
- Lower resolution in fine-mapping
- **Poor imputation for rare variants**



## Opportunities

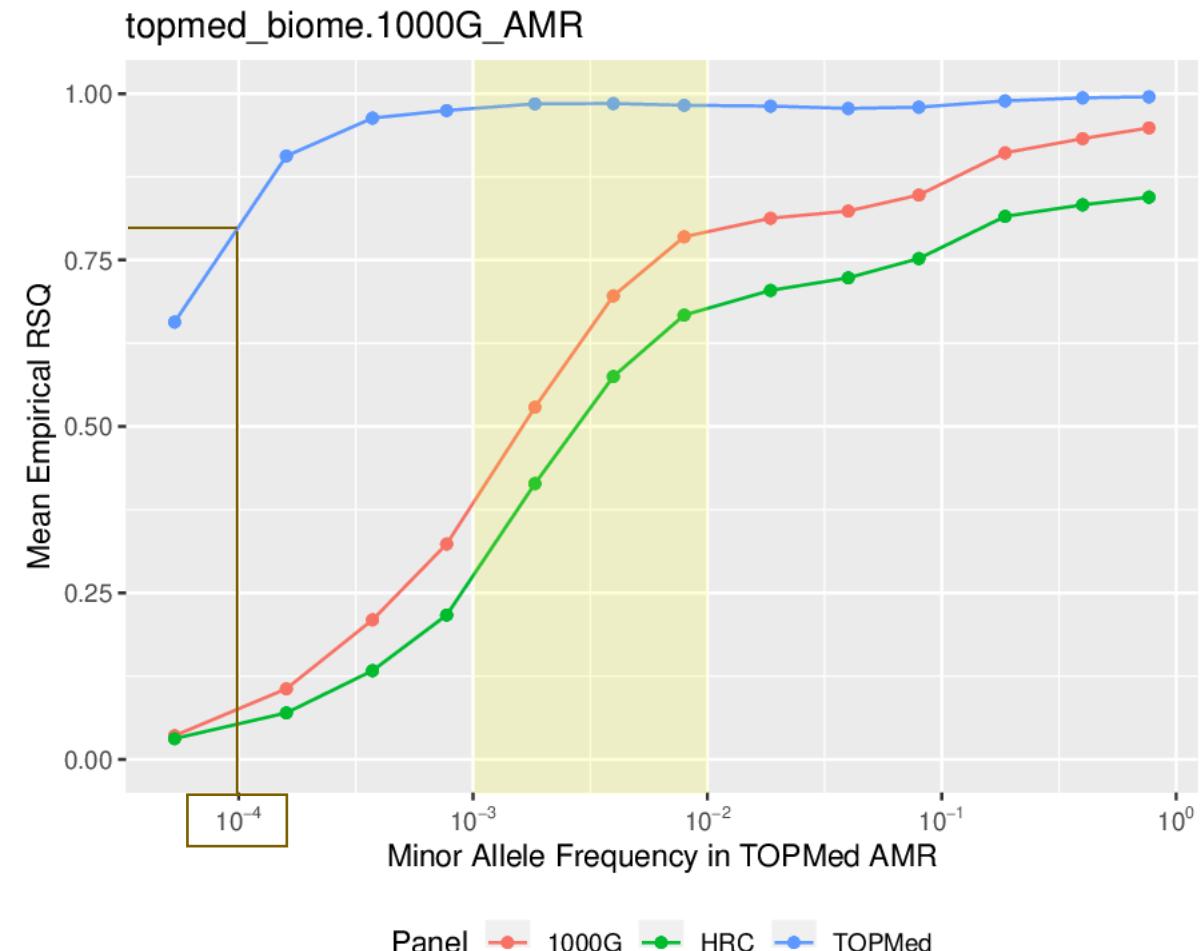
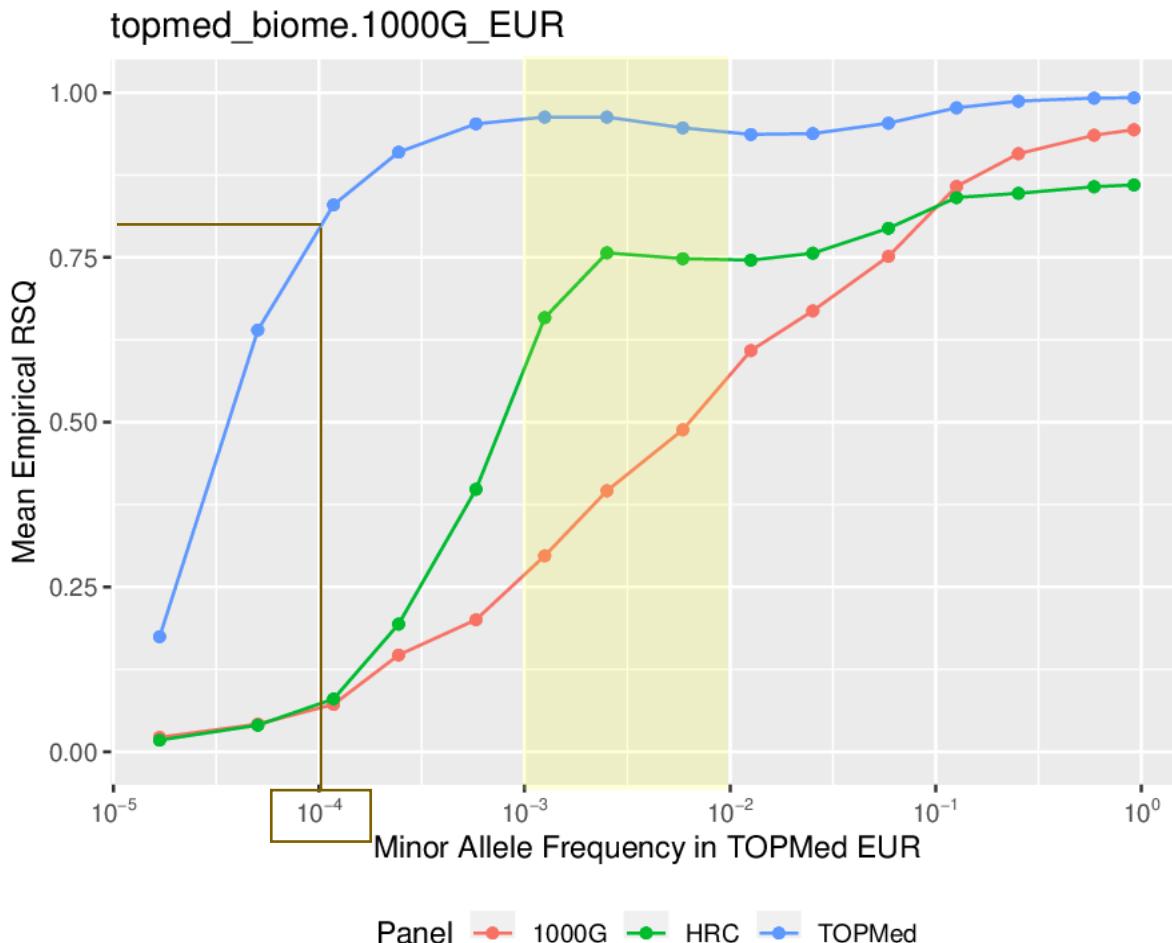
- Fewer variants needed for good coverage of array



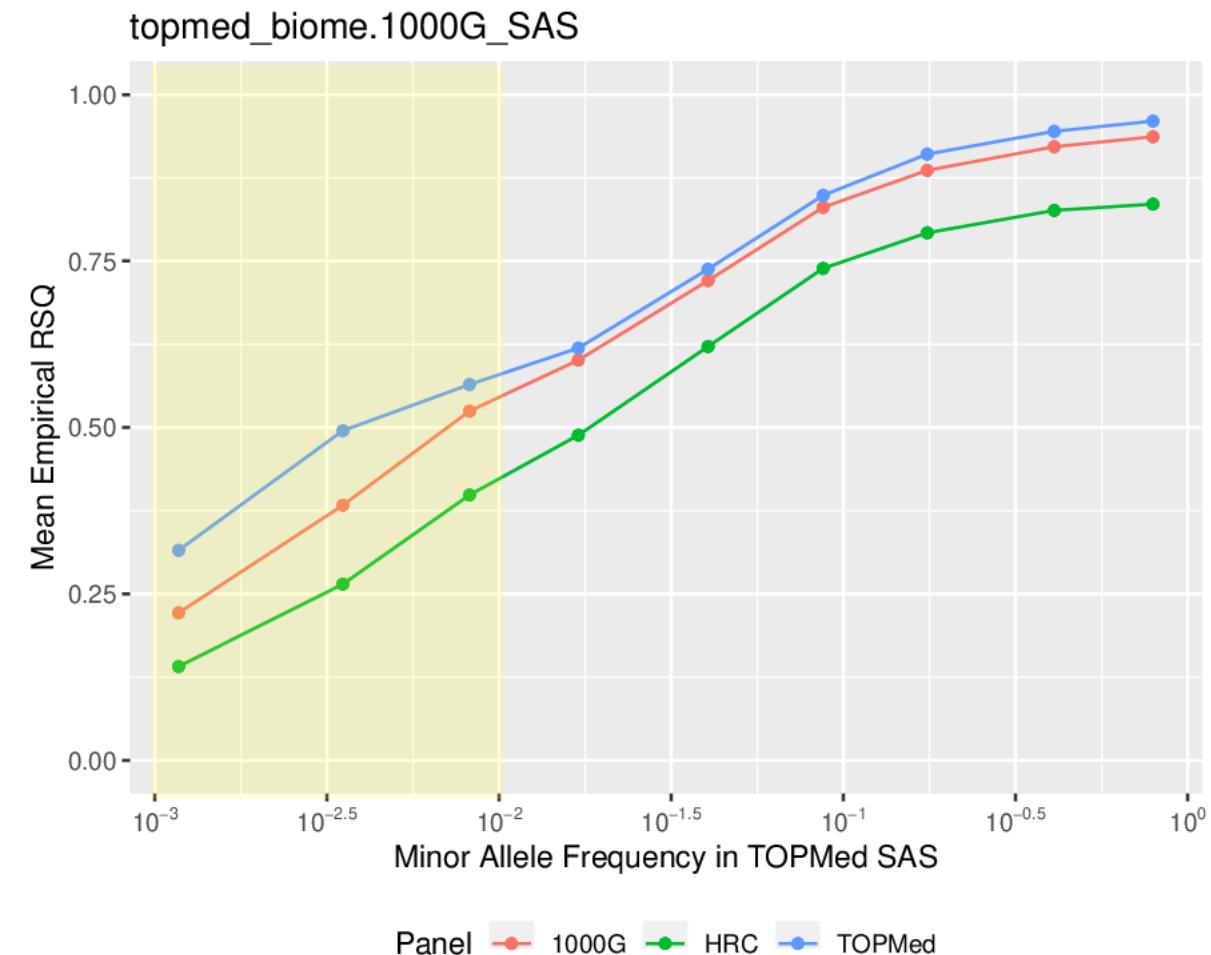
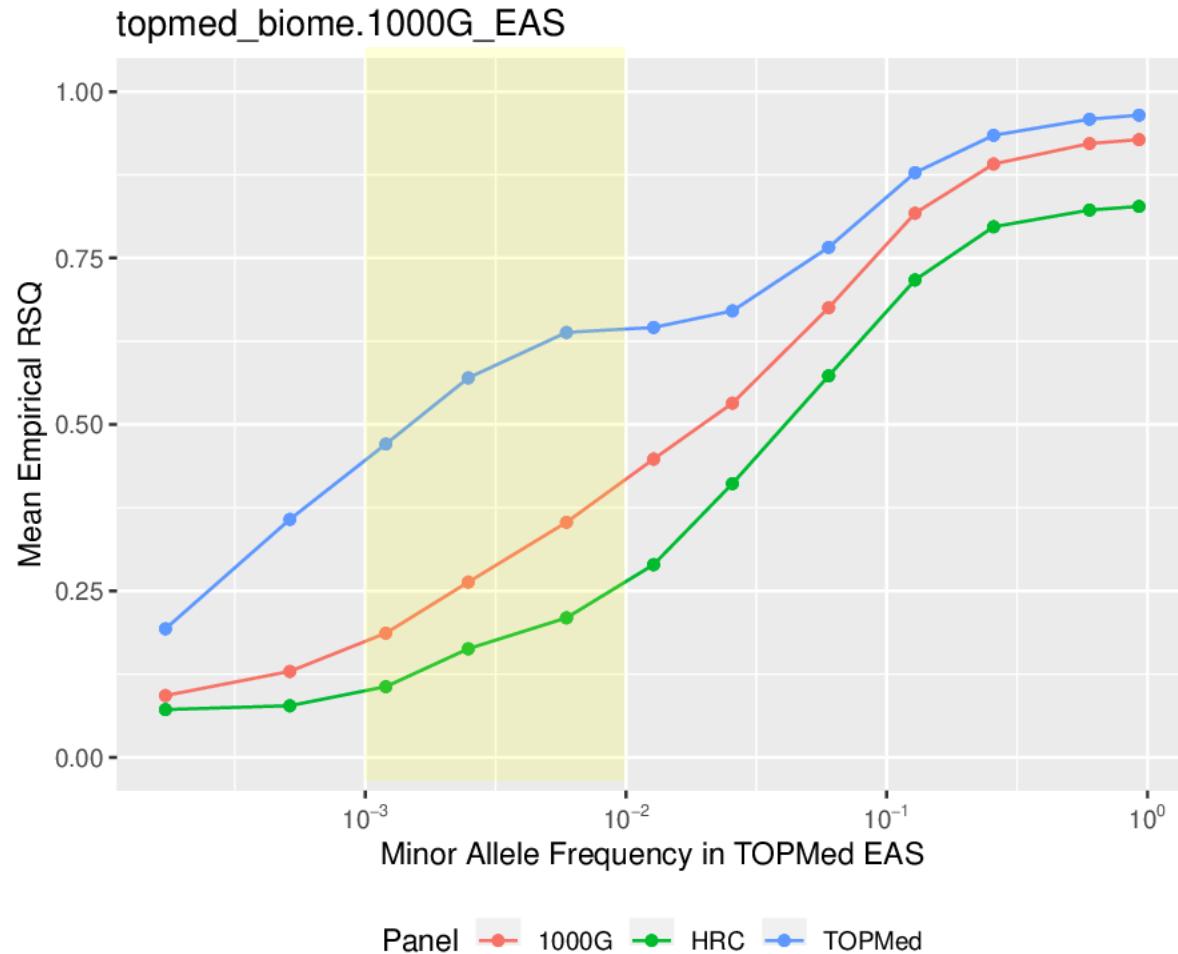
**d Imputation**

A table showing SNP genotype data for four individuals across six SNPs. Some SNPs (SNP1, SNP2, SNP3) are highlighted in red boxes.

# Imputation accuracy using global reference panel



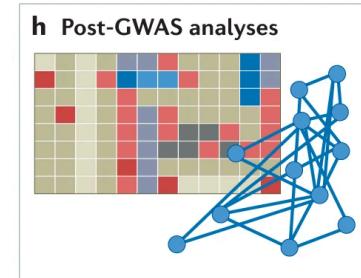
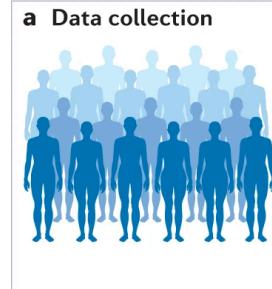
# Imputation accuracy using global reference panel



# Challenges and opportunities of Asian GWAS

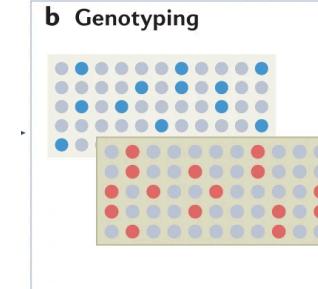
## Challenges

- Small sample size
- Lower resolution in fine-mapping
- **Poor imputation for rare variants**



## Opportunities

- Fewer variants needed for good coverage of array



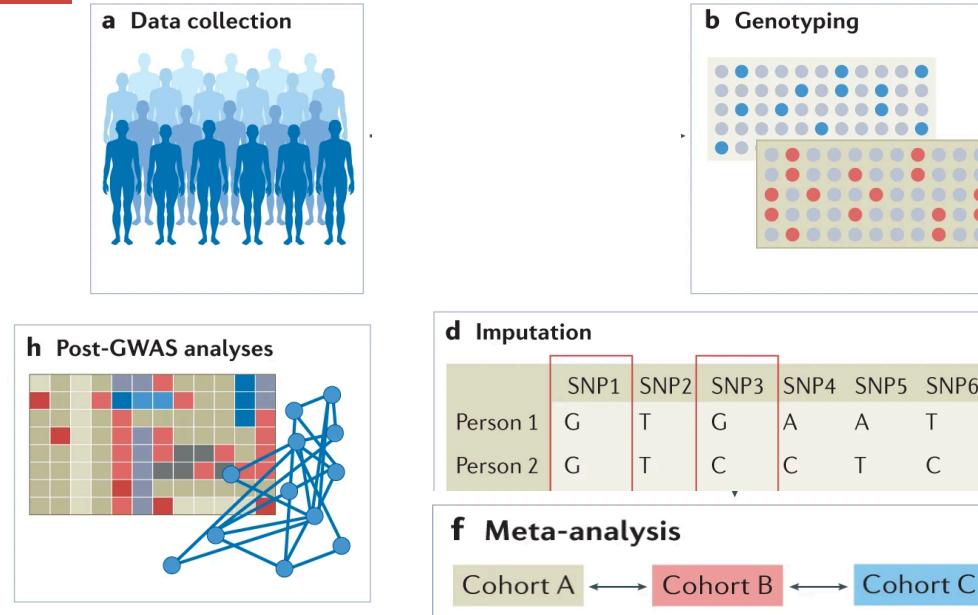
**d Imputation**

	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6
Person 1	G	T	G	A	A	T
Person 2	G	T	C	C	T	C
Person 3	C	A	G	C	A	C
Person 4	C	A	C	C	T	C

# Challenges and opportunities of Asian GWAS

## Challenges

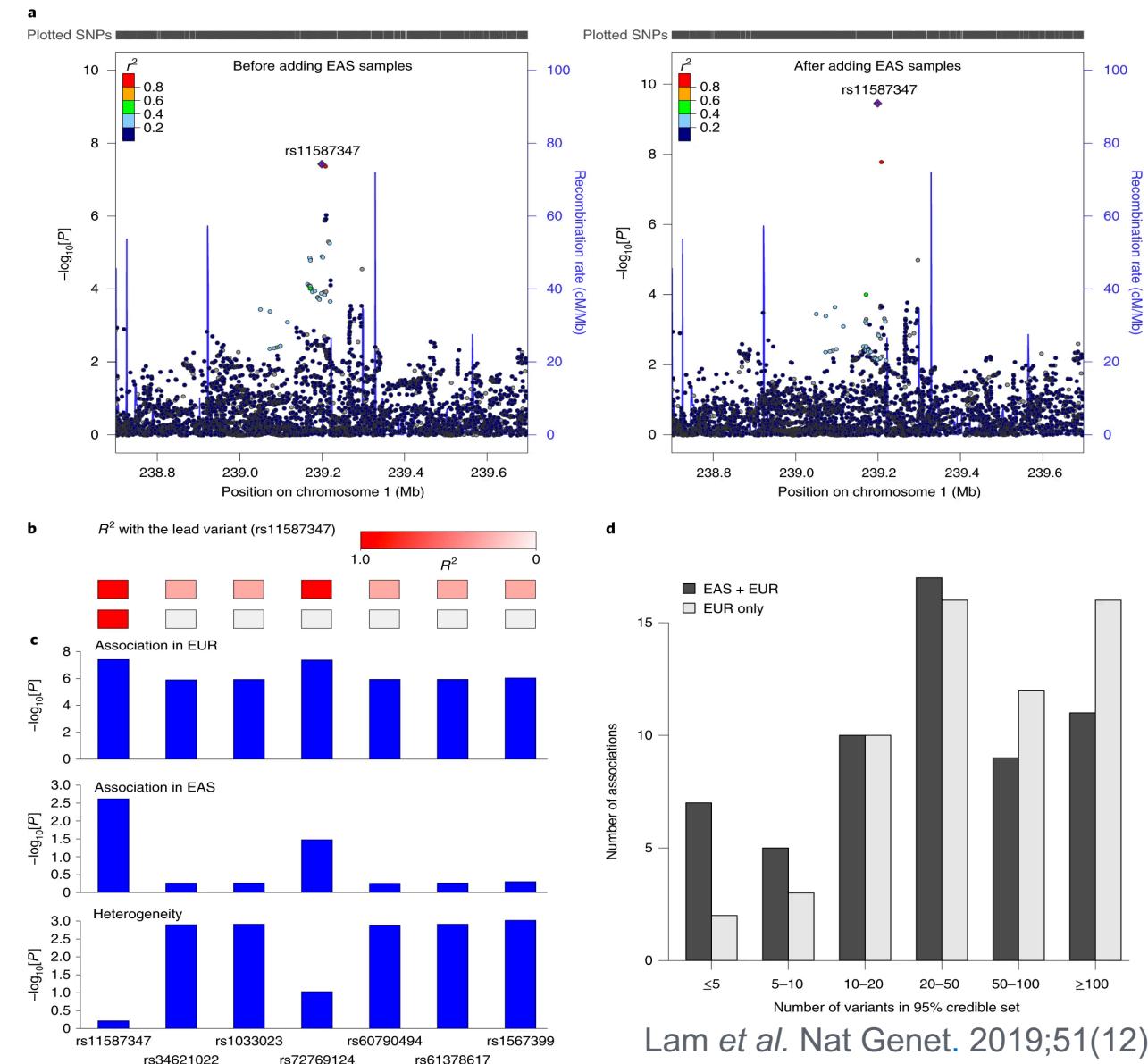
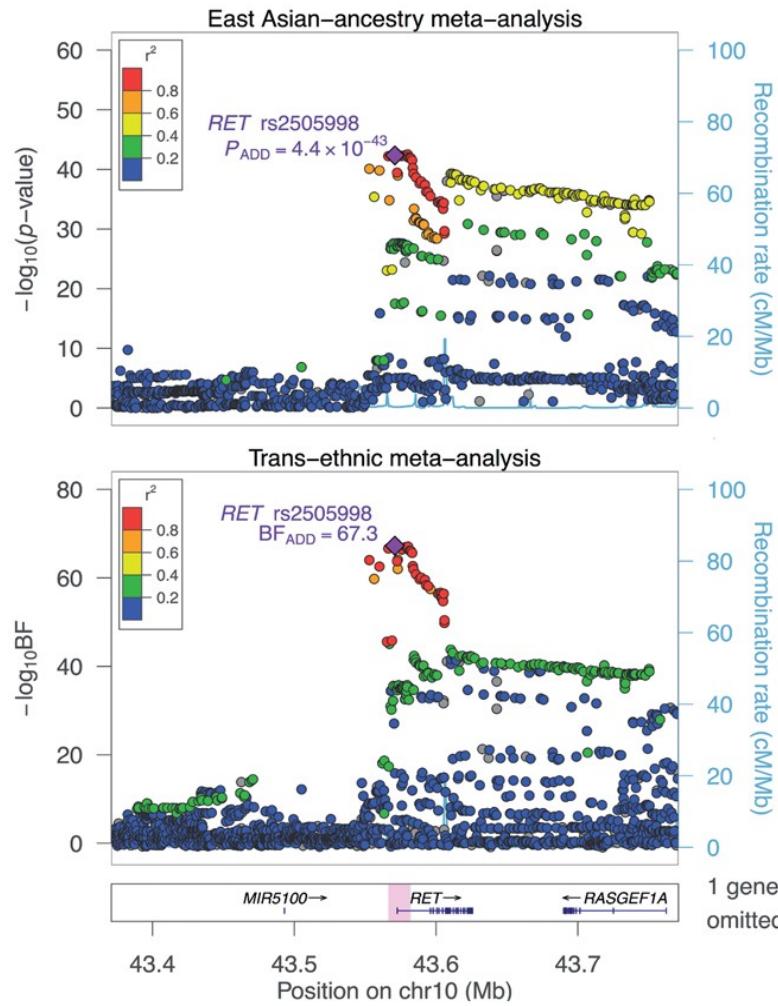
- Small sample size
- Lower resolution in fine-mapping
- Poor imputation for rare variants



## Opportunities

- Fewer variants needed for good coverage of array
- **Multi-ethnic meta-analysis to improve fine-mapping resolution**

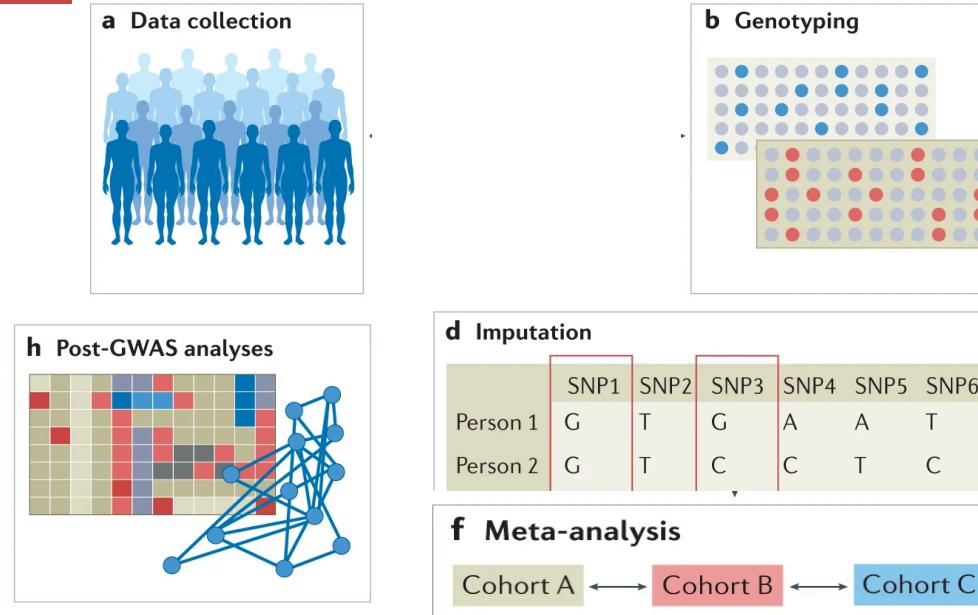
# Multi-ethnic meta-analysis to improve fine-mapping



# Challenges and opportunities of Asian GWAS

## Challenges

- Small sample size
- Lower resolution in fine-mapping
- Poor imputation for rare variants



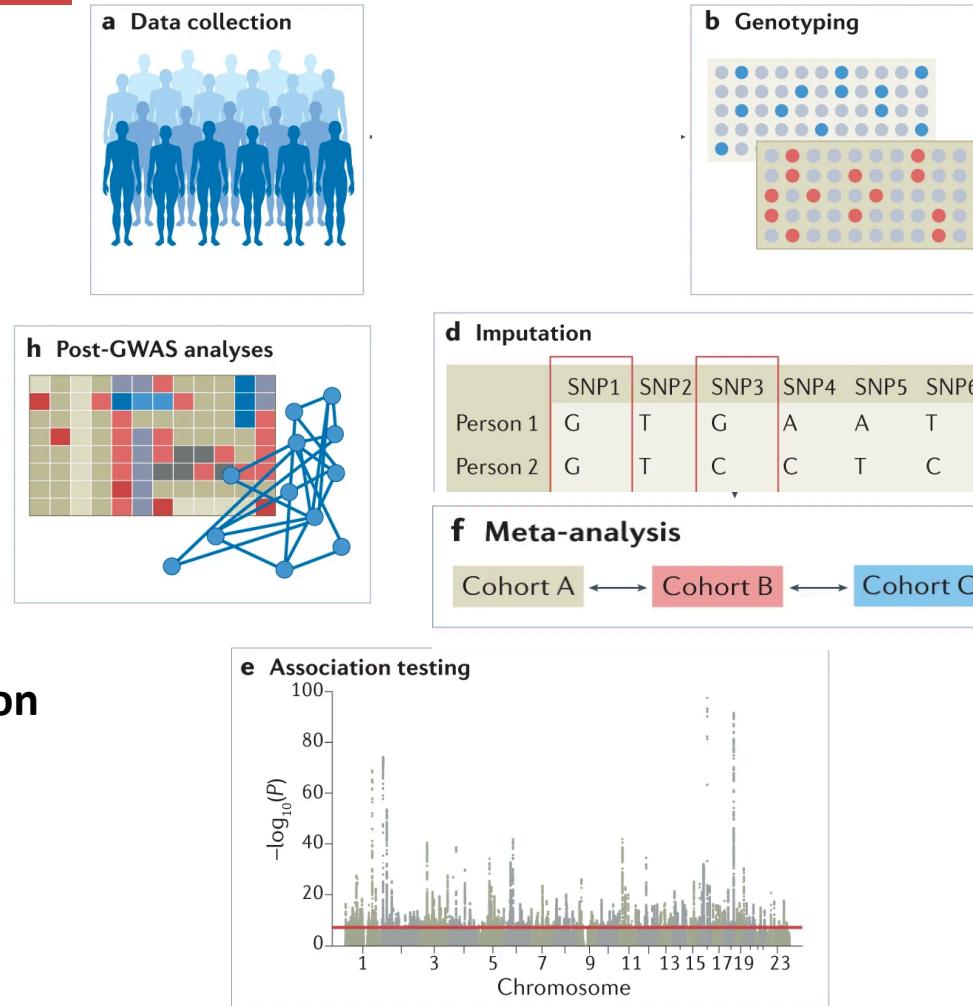
## Opportunities

- Fewer variants needed for good coverage of array
- **Multi-ethnic meta-analysis to improve fine-mapping resolution**

# Challenges and opportunities of Asian GWAS

## Challenges

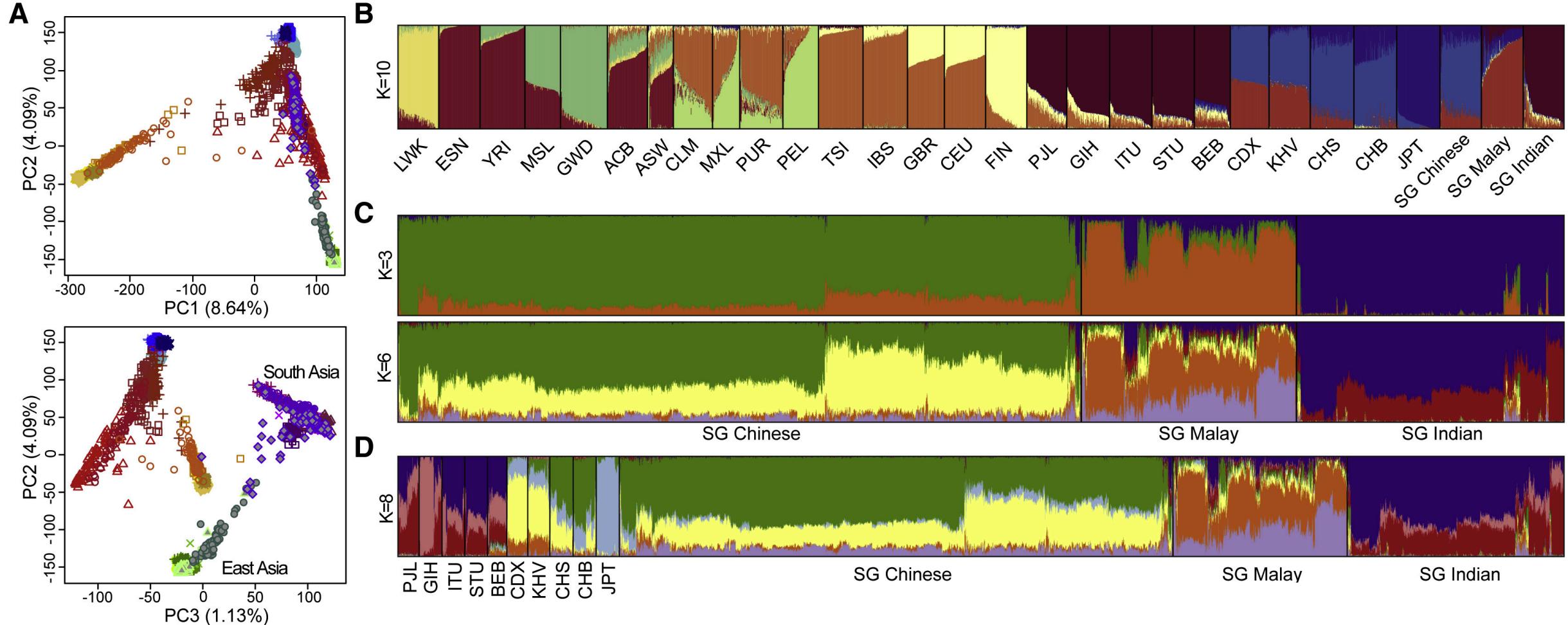
- Small sample size
- Lower resolution in fine-mapping
- Poor imputation for rare variants
- Population stratification



## Opportunities

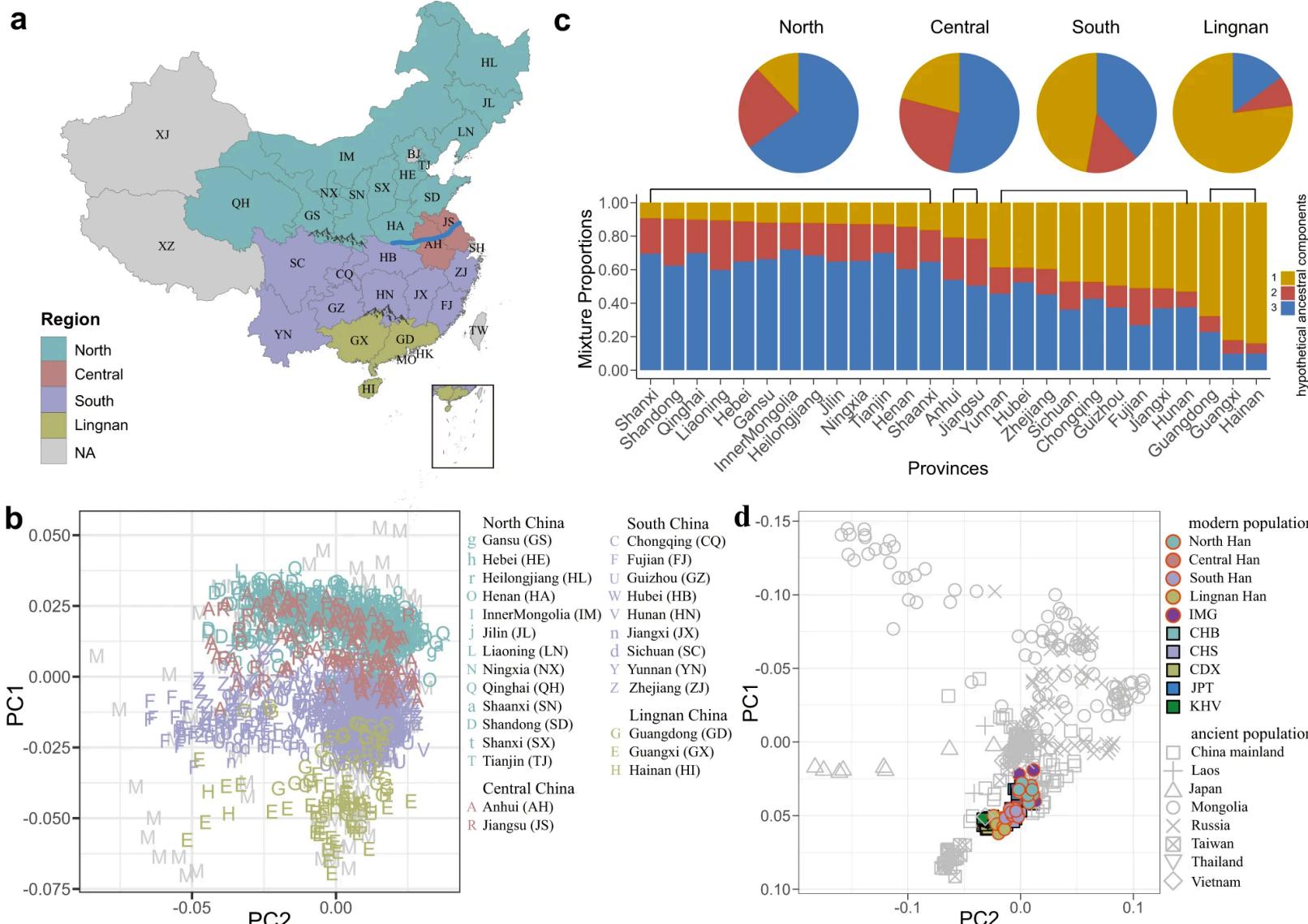
- Fewer variants needed for good coverage of array
- Multi-ethnic meta-analysis to improve fine-mapping resolution
- Novel discoveries for population-specific variants or variants with differentially high frequency in Asians

# Population structure among Singaporean populations

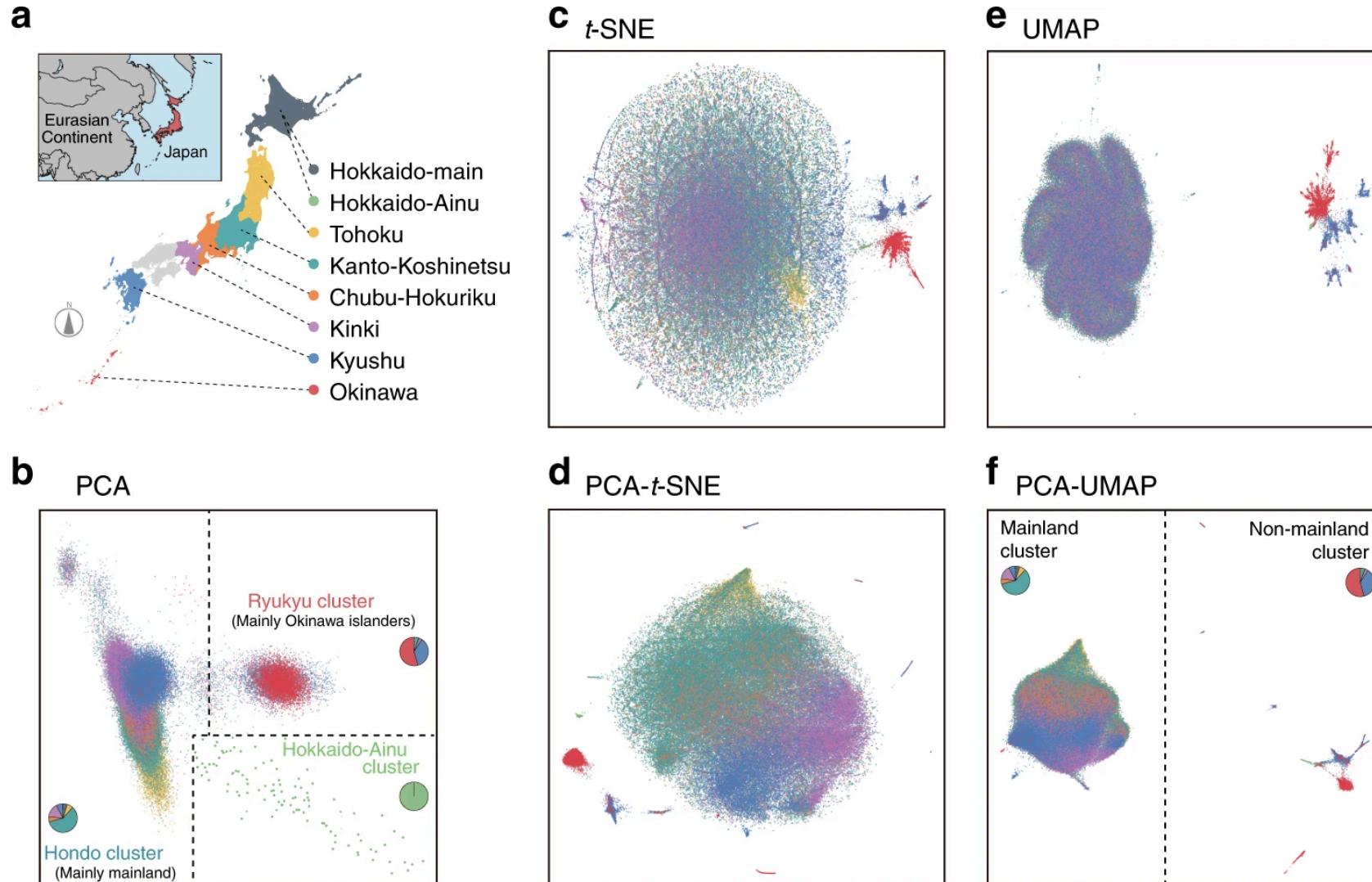


SG10K	ASW	America	Europe	South Asia	East Asia	
SG Chinese	○	ASW	■	CEU	□	CDX
SG Malay	▲	ESN	□	CEU	○	CHB
SG Indian	◆	GWD	○	FIN	○	CHS
Africa	■	LWK	○	GBR	△	CHB
ACB	□	CLM	○	PEL	△	CHS
MSL	◇	MLX	○	PEL	▲	CHB
PUR	+	PUR	+	IBS	+	JPT
PEL	+	PUR	+	IBS	+	JPT
TSI	×	TSI	×	TSI	×	KHV

# Population substructure among Chinese



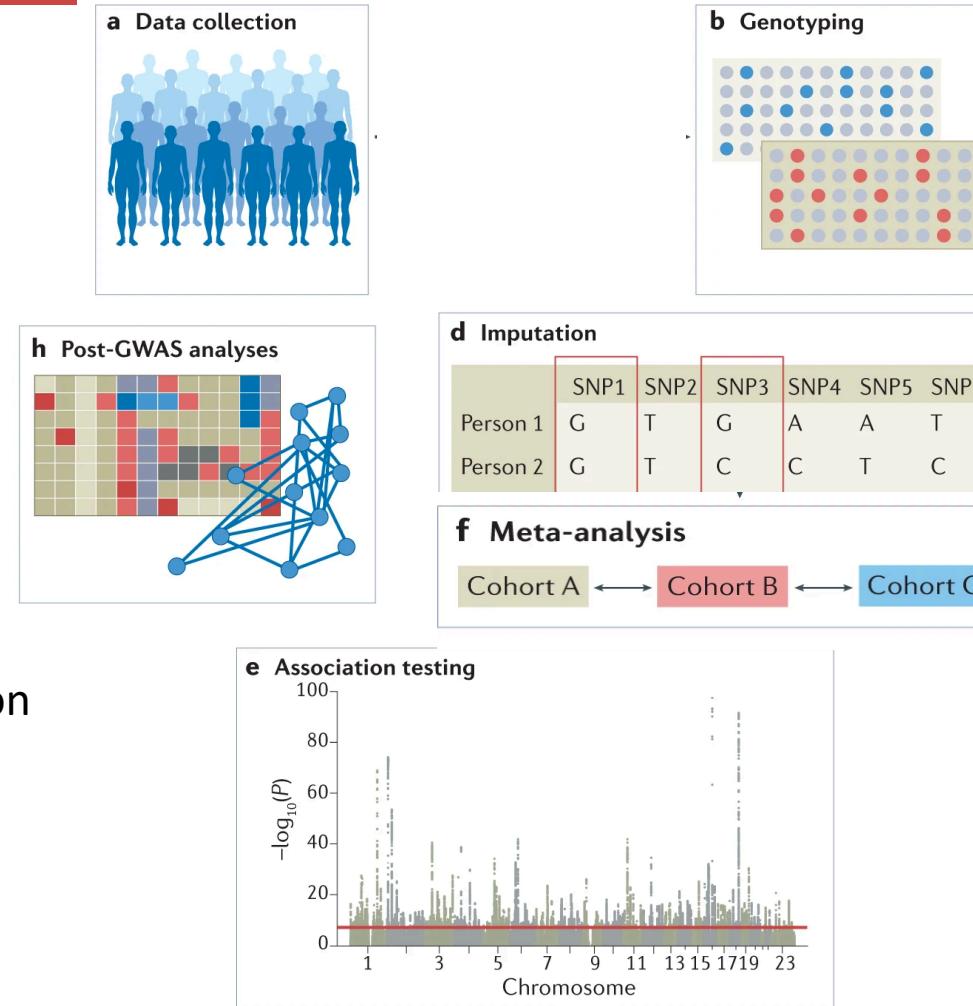
# Population substructure among Japanese



# Challenges and opportunities of Asian GWAS

## Challenges

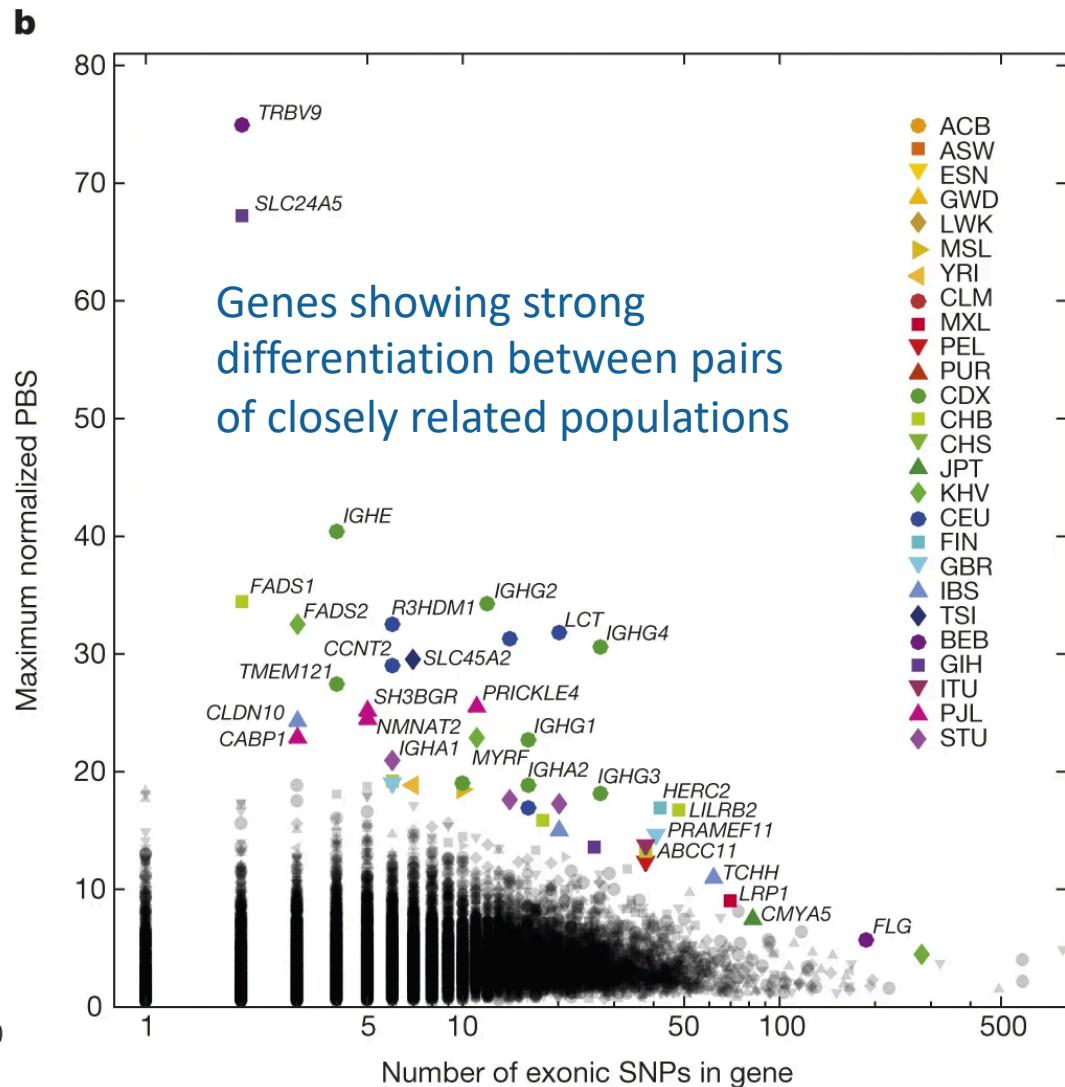
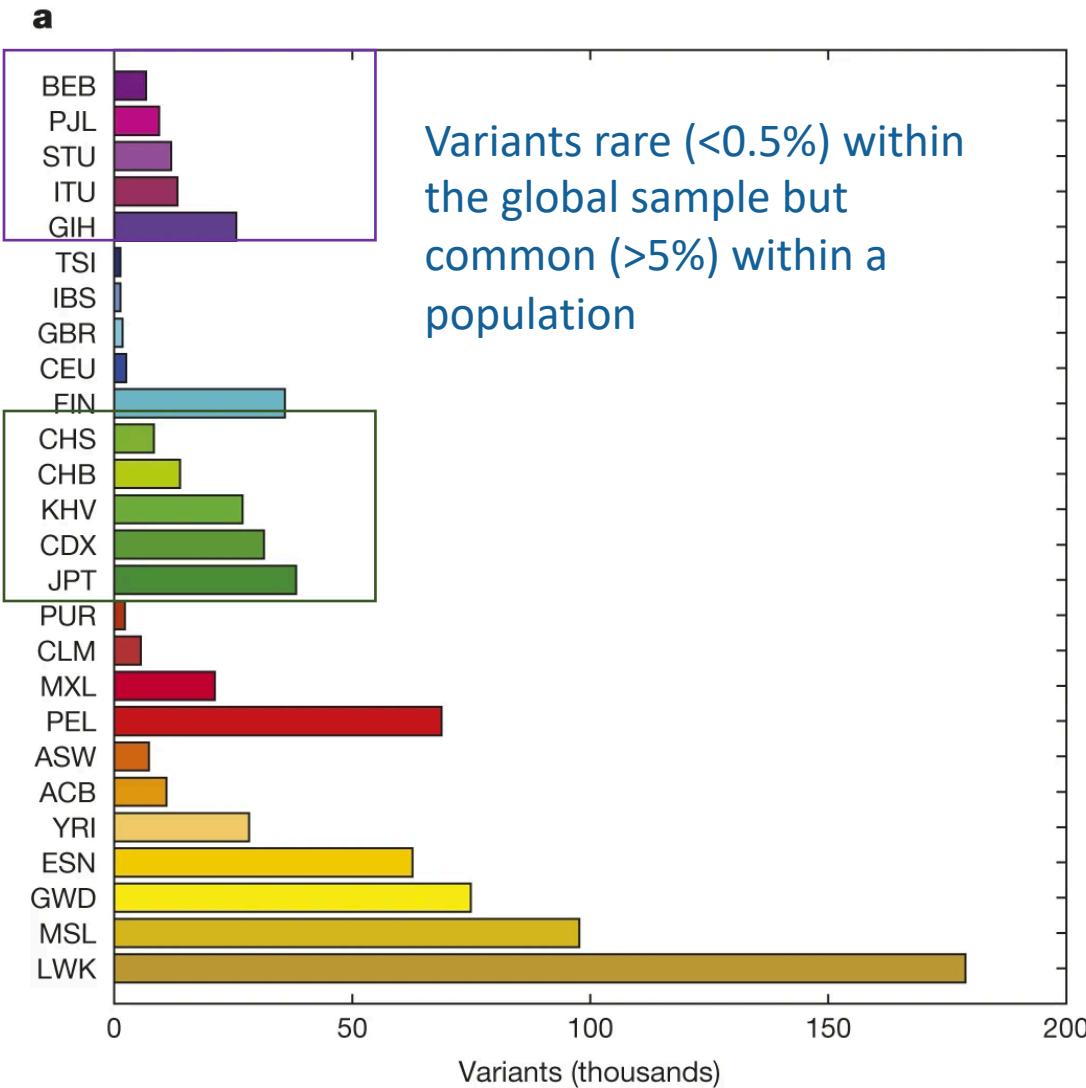
- Small sample size
- Lower resolution in fine-mapping
- Poor imputation for rare variants
- Population stratification



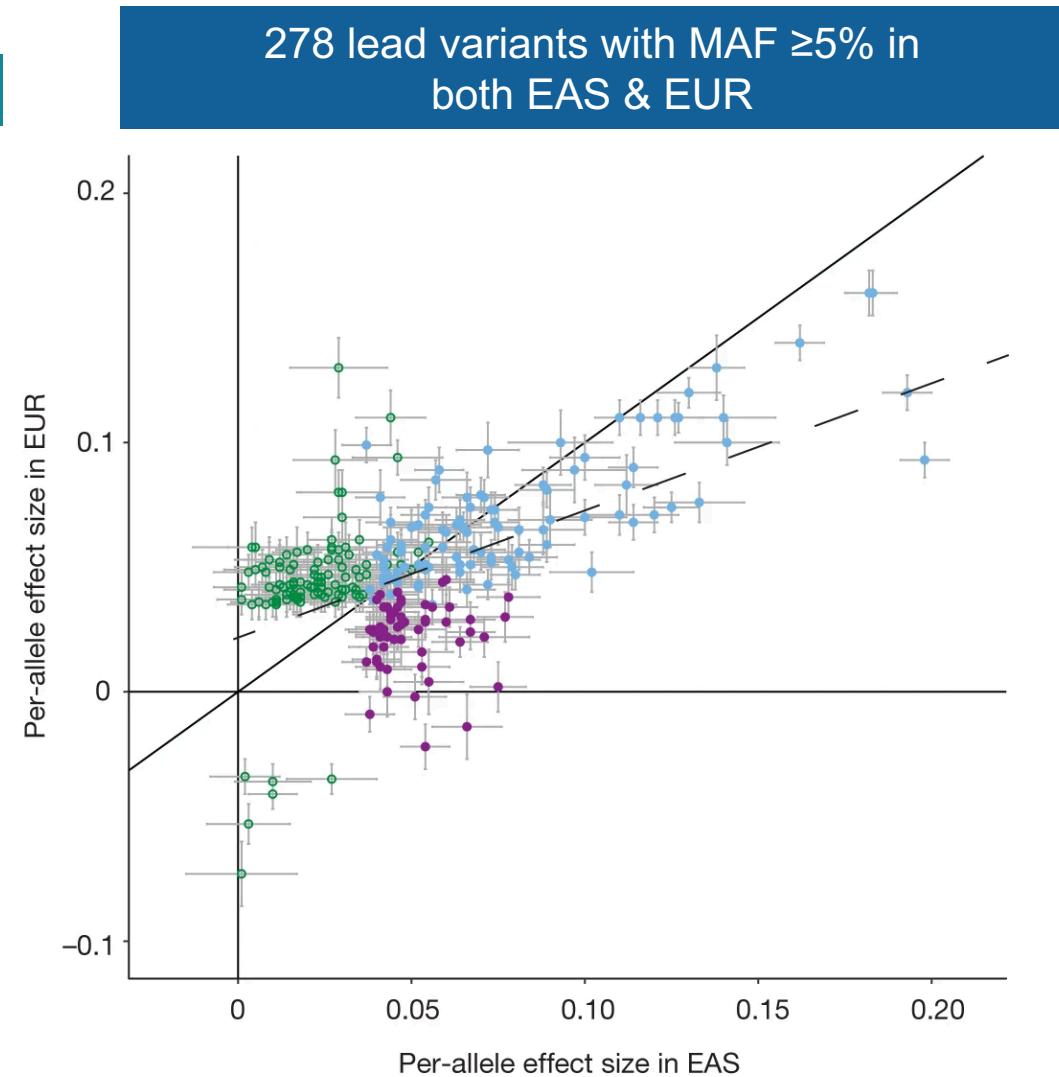
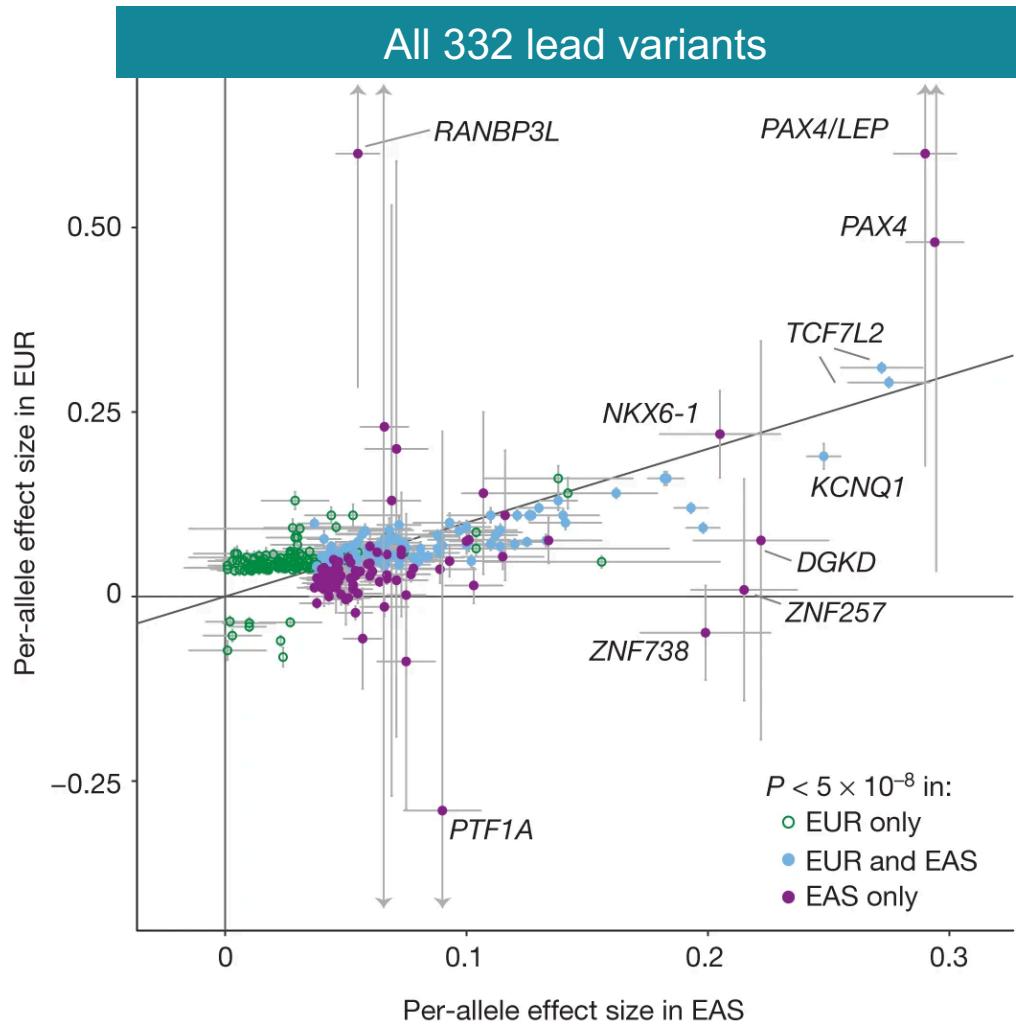
## Opportunities

- Fewer variants needed for good coverage of array
- Population-specific polygenic risk score model
- Multi-ethnic meta-analysis to improve fine-mapping resolution
- **Novel discoveries for population-specific variants or variants with differentially high frequency in Asians**

# Differential frequencies of variants across populations



# Opportunities of Asian GWAS: Population-specific association with type 2 diabetes



# Opportunities of Asian GWAS: 14 novel East Asian-specific association with blood lipids

Gene	Lipid type	East Asian (n=47,532)			GLGC (>300,000 Europeans)		
		Freq (%)	Effect	P	Freq (%)	Effect	P
<i>EVI5</i> (R354C)	TC	0.69	0.21	1.4x10 <sup>-7</sup>	0.03	0.10	0.25
<i>APOB</i> (I3768T) (C478Y)	TC	0.15 0.09	-0.66 -0.88	8.4x10 <sup>-12</sup> 2.1x10 <sup>-10</sup>			
		12.4	-0.11	1.5x10 <sup>-19</sup>	0.19	-0.08	6.7x10 <sup>-3</sup>
<i>HMGCR</i> (Y311S)	LDL	1.7	-0.19	2.2x10 <sup>-13</sup>	0.04	-0.12	0.08
<i>CD36</i> (R386W)	HDL	0.31	0.34	3.2x10 <sup>-9</sup>	0.02	0.22	0.01
<i>APOA1</i> (A61T)	HDL	3.3	-0.12	5.5x10 <sup>-10</sup>	0.02	0.08	0.45
<i>ACACB</i> (V2141I)	TG	74.3	0.04	4.0x10 <sup>-8</sup>	80.2	0.01	5.3x10 <sup>-4</sup>
<i>ALDH2</i> (Q457K) (N459G)	HDL	20.4 2.23	-0.05 0.41	1.2x10 <sup>-8</sup> 7.5x10 <sup>-62</sup>	0.08 0.02	-0.01 0.38	0.93 3.2x10 <sup>-5</sup>
<i>PKD1L3</i> (R1572H)	LDL	5.4	0.09	2.1x10 <sup>-8</sup>	24.4	-0.01	8.5x10 <sup>-5</sup>
<i>LDLR</i> (R257W)	TC	0.09	0.68	5.6x10 <sup>-10</sup>	0.001	1.90	1.6x10 <sup>-4</sup>
<i>PPARA</i> (V227A)	TG	4.2	-0.09	3.2x10 <sup>-7</sup>	0.15	-0.06	0.12

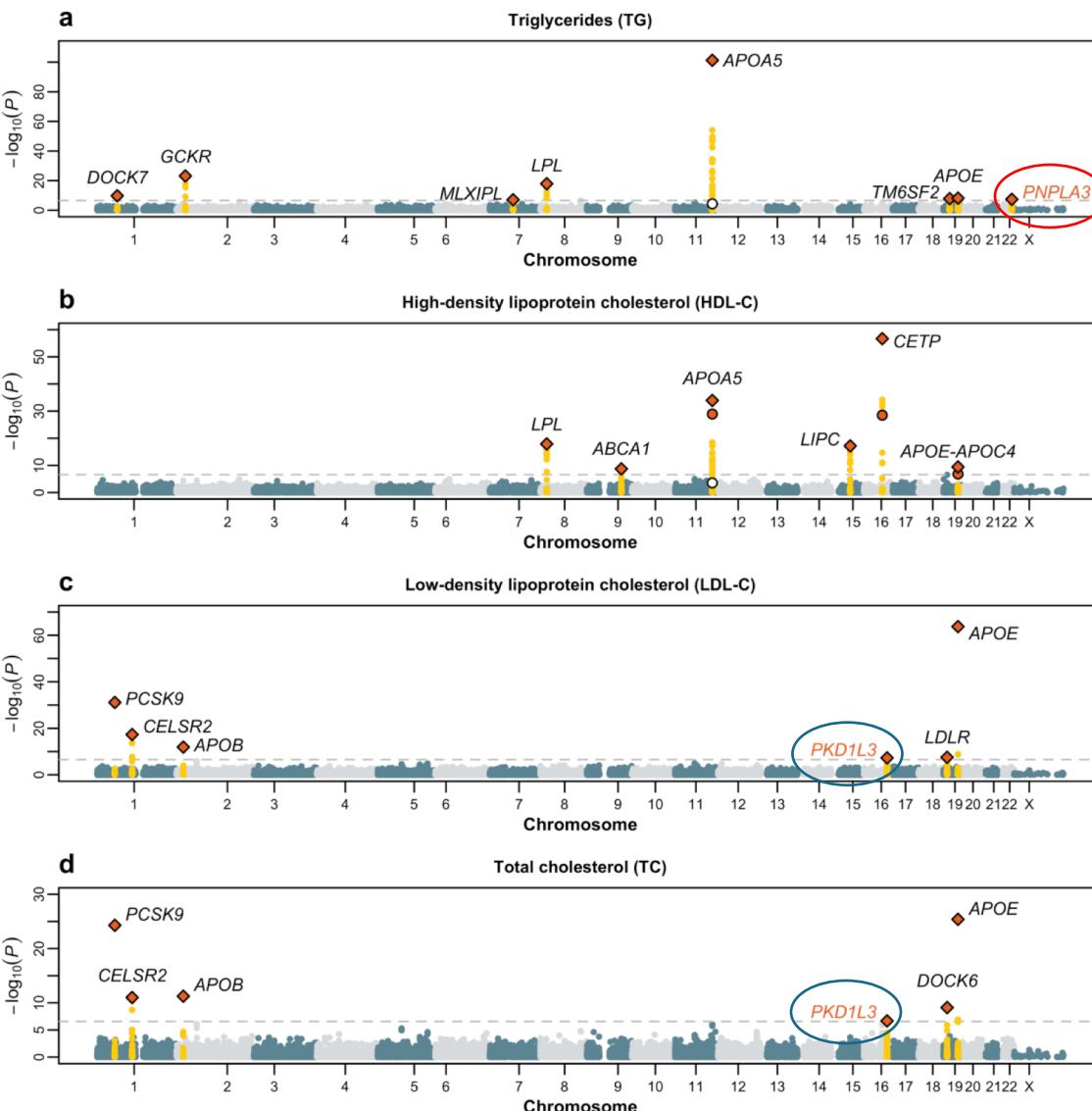
# Opportunities of Asian GWAS: Association of *PNPLA3* variants with triglyceride in Asian GWAS

TG

HDL

LDL

TC



Exome-wide  
association analysis of  
blood lipid levels on  
12,685 East Asians

***PNPLA3***



***PKD1L3***



**LDL**

**TC**

# Opportunities of Asian GWAS: Replication of association of *PNPLA3* with TG in European GWAS

LETTERS

nature  
genetics

## Exome-wide association study of plasma lipids in >300,000 individuals

We screened variants on an exome-focused genotyping array in >300,000 participants (replication in >280,000 participants) and identified 444 independent variants in 250 loci significantly associated with total cholesterol (TC), high-density-lipoprotein cholesterol (HDL-C), low-density-lipoprotein cholesterol (LDL-C), and/or triglycerides (TG). At two loci (*JAK2* and *A1CF*), experimental analysis in mice showed lipid changes consistent with the human data. We also found that: (i) beta-thalassemia trait carriers displayed lower TC and were protected from coronary artery disease (CAD); (ii) excluding the *CETP* locus, there was not a predictable relationship between plasma HDL-C and risk for age-related macular degeneration; (iii) only some mechanisms of lowering LDL-C appeared to increase risk for type 2 diabetes (T2D); and (iv) TG-lowering alleles involved in hepatic production of TG-rich lipoproteins (*TM6SF2* and *PNPLA3*) tracked with higher liver fat, higher risk for T2D, and lower risk for CAD, whereas TG-lowering alleles involved in peripheral lipolysis (*LPL* and *ANGPTL4*) had no effect on liver fat but decreased risks for both T2D and CAD.

Global Lipid Genetics  
Consortium (GLGC)

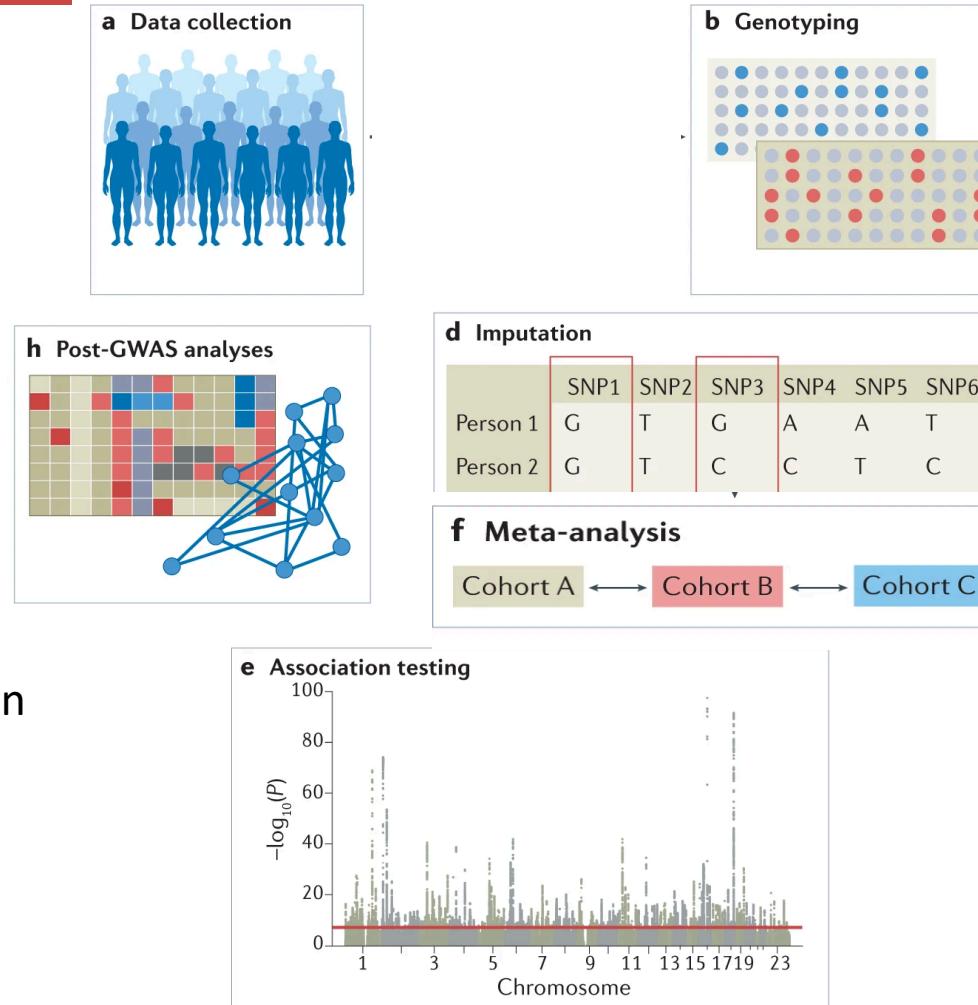
>300,000 samples  
(84% Europeans)



# Challenges and opportunities of Asian GWAS

## Challenges

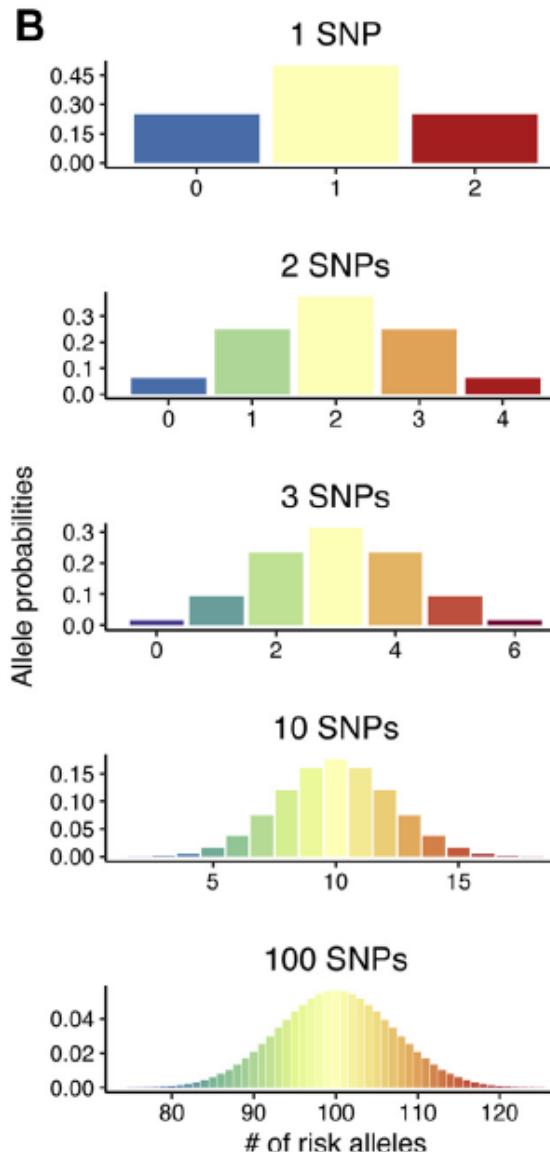
- Small sample size
- Lower resolution in fine-mapping
- Poor imputation for rare variants
- Population stratification



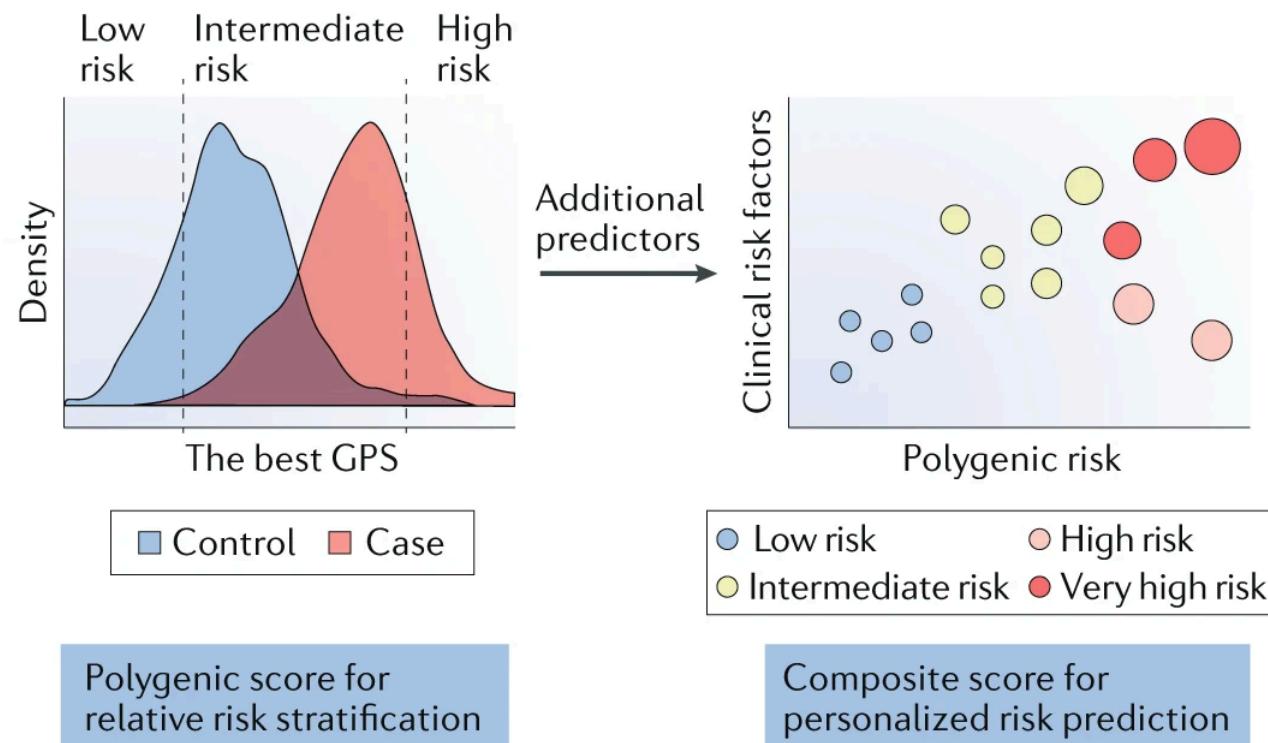
## Opportunities

- Fewer variants needed for good coverage of array
- **Population-specific polygenic risk score model**
- Multi-ethnic meta-analysis to improve fine-mapping resolution
- Novel discoveries for population-specific variants or variants with differentially high frequency in Asians

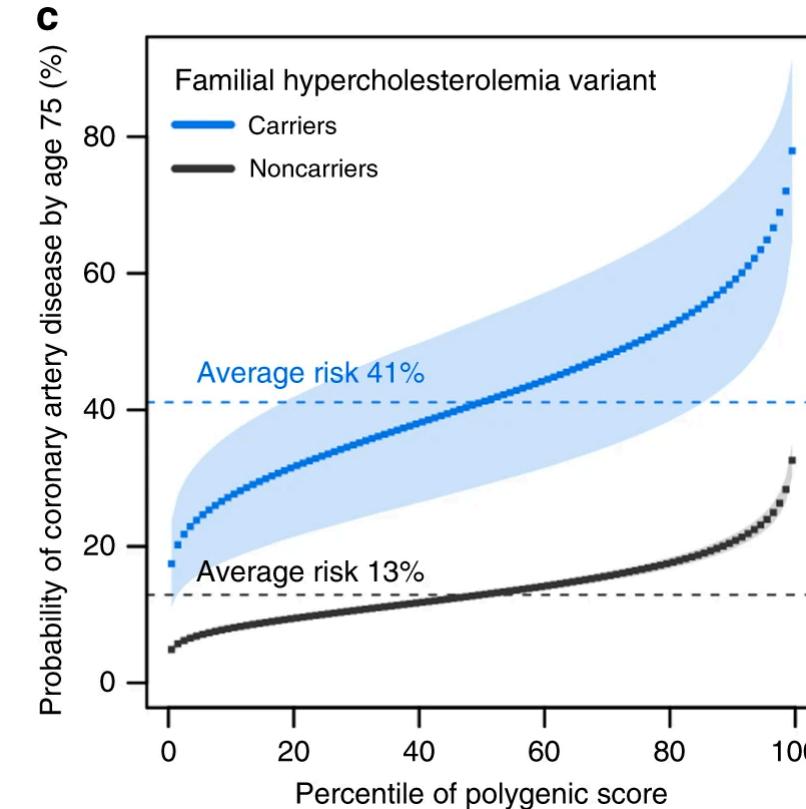
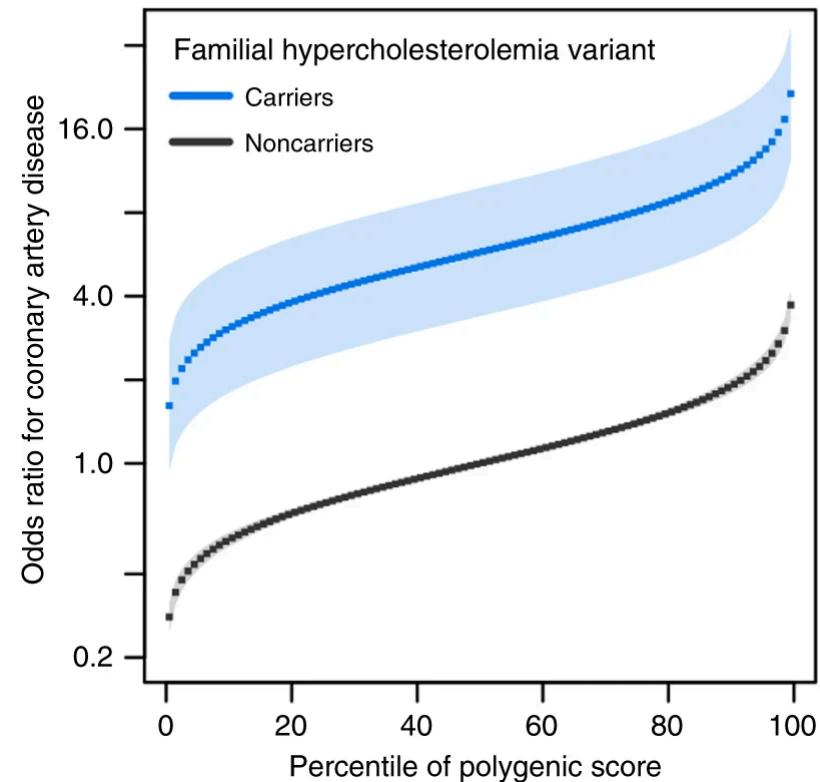
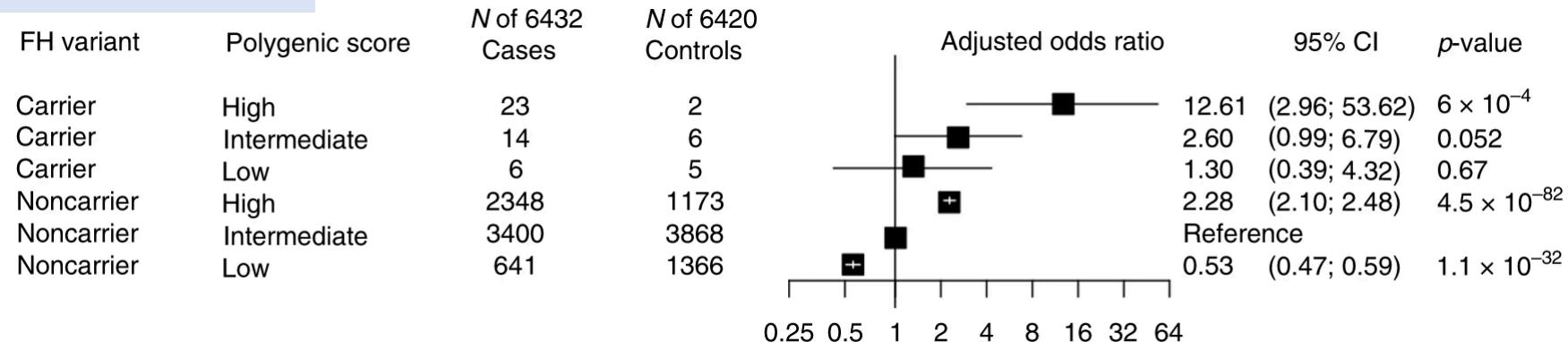
# Polygenic risk score (PRS)



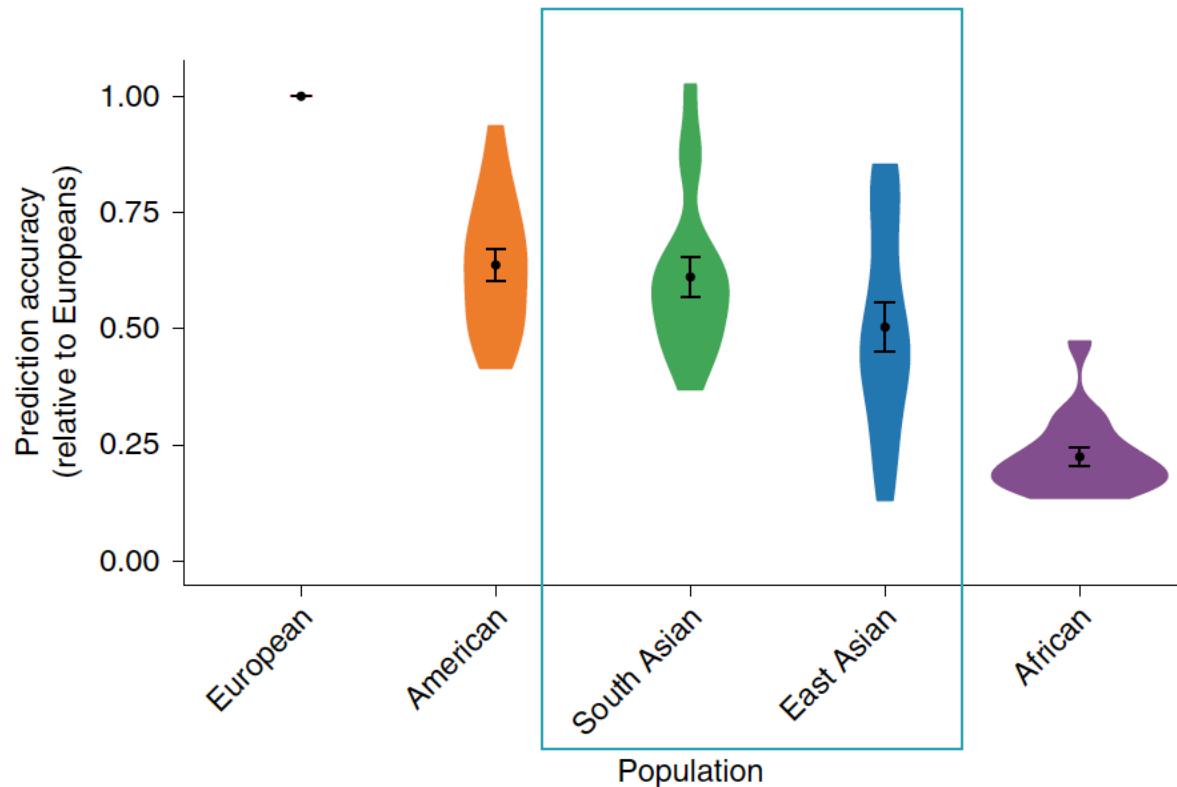
- Also be referred to as **genomic risk scores (GRS)**
- PRS** aggregates the effects of hundreds to millions of genetic variants which are combined using a weighted sum of **allele dosages** multiplied by their **effect sizes**



## PRS using 6,630,150 variants

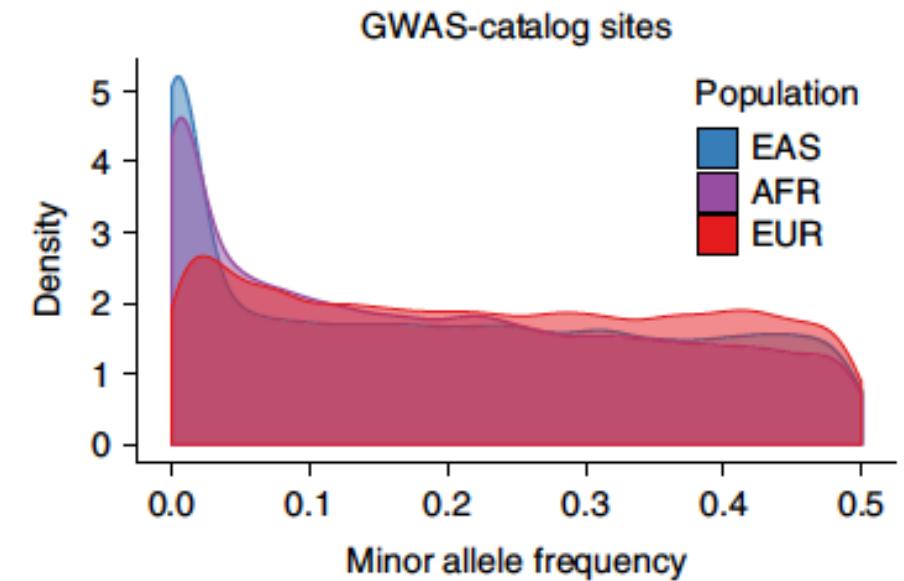


# Low generalizability of polygenic risk score (PRS)

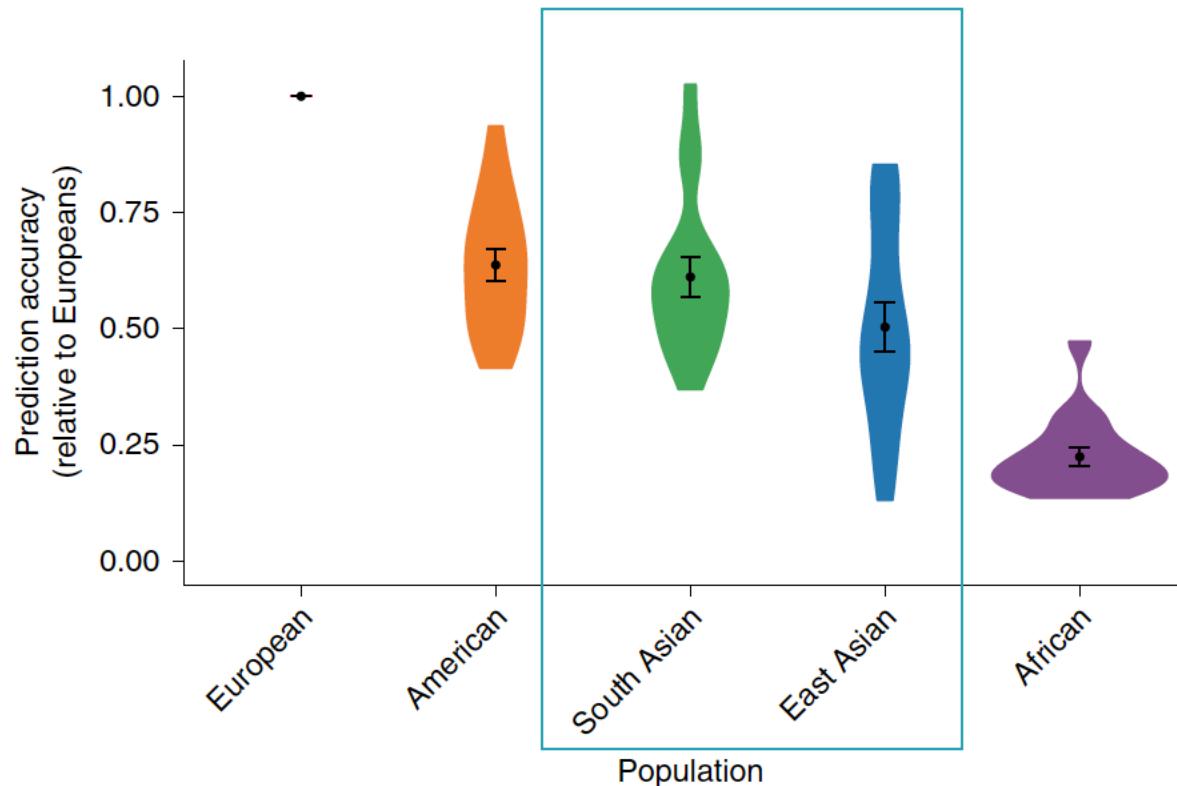


Prediction accuracy across 17 quantitative traits  
in the UK Biobank

- Discovery of associated variants are biased towards those **common in study population**



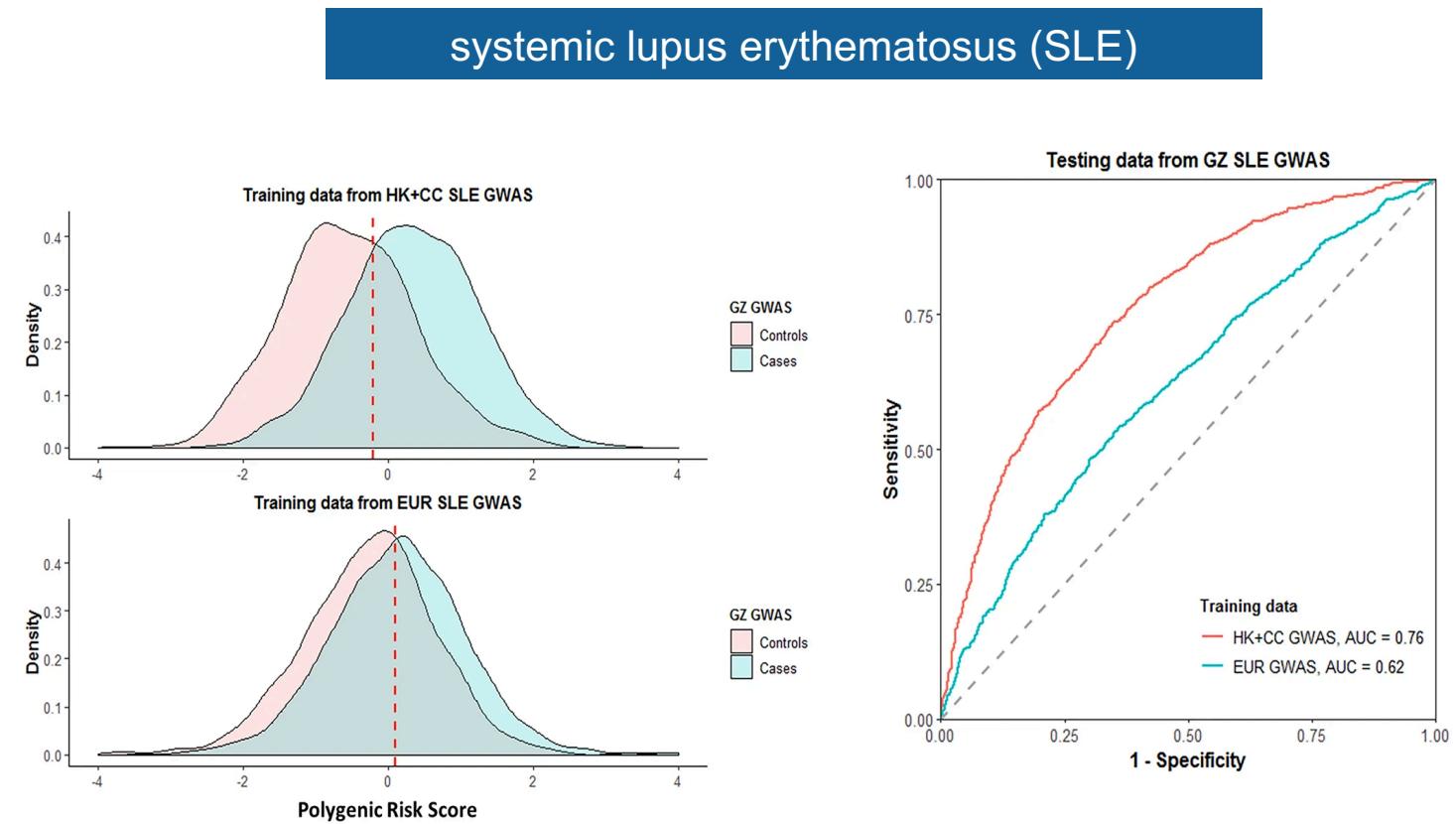
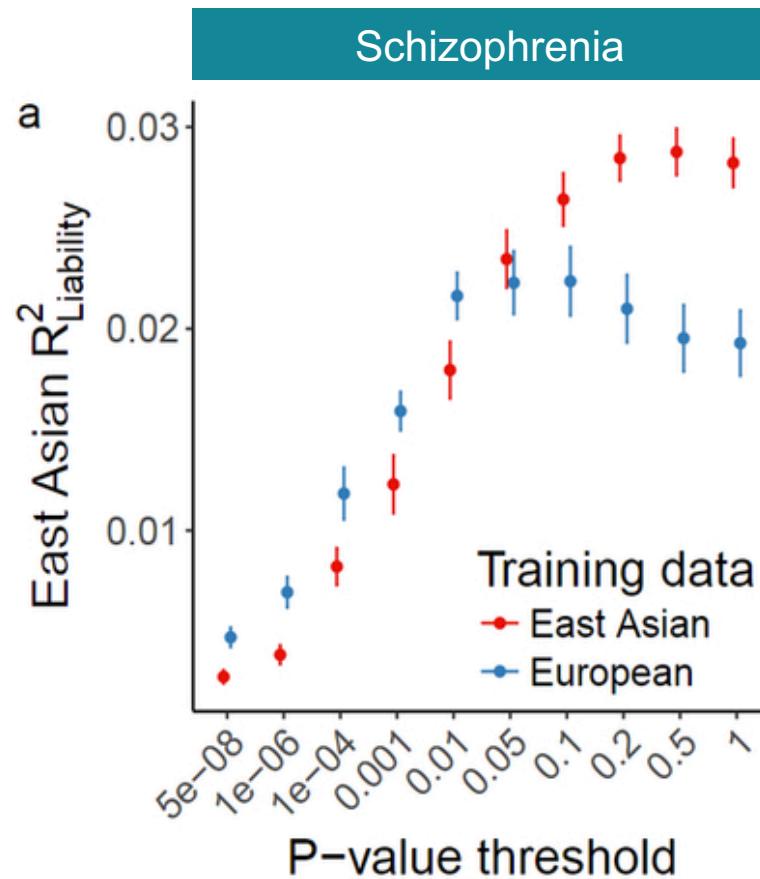
# Low generalizability of polygenic risk score (PRS)



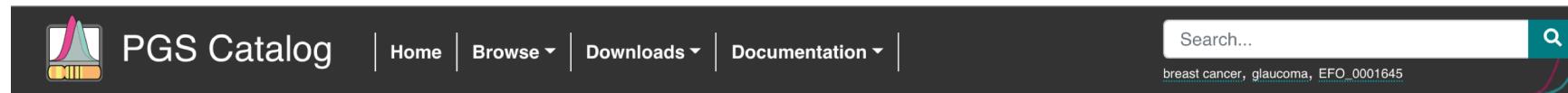
Prediction accuracy across 17 quantitative traits  
in the UK Biobank

- Discovery of associated variants are biased towards those **common in study population**
- **Linkage disequilibrium (LD) structure** between top and causal variants varies across population and drives **difference in effect size estimates**
- **Differences in environmental** and demographical factors

# Improved accuracy of PRS using GWAS with matched ancestry



# The Polygenic Score (PGS) Catalog



Latest release: Sept. 17

## The Polygenic Score (PGS) Catalog

An open database of polygenic scores and the relevant metadata required for accurate application and evaluation.

Search...  
breast cancer, glaucoma, EFO\_0001645

	PGP000006 » Khera AV et al. Nat Genet (2018)	Type 2 diabetes	type II diabetes mellitus	6,917,436	G D E	Terms/Lice
	PGP000006 » Khera AV et al. Nat Genet (2018)	Inflammatory bowel disease	inflammatory bowel disease	6,907,112	G D E	Terms/Lice
	PGP000006 » Khera AV et al. Nat Genet (2018)	Atrial fibrillation	atrial fibrillation	6,730,541	G D E	Terms/Lice
	PGP000006 » Khera AV et al. Nat Genet (2018)	Coronary artery disease	coronary artery disease	6,630,150	G D E	Terms/Lice
	PGP000090 » Wang M et al. J Am Coll Cardiol (2020)	Coronary artery disease	coronary artery disease	6,630,150	G D E	Terms/Lice
	PGP000100 » Mars N et al. Nat Med (2020)	Prostate cancer	prostate carcinoma	6,606,785	G D E	
	PGP000100 » Mars N et al. Nat Med (2020)	Type 2 diabetes	type II diabetes mellitus	6,437,380	G D E	

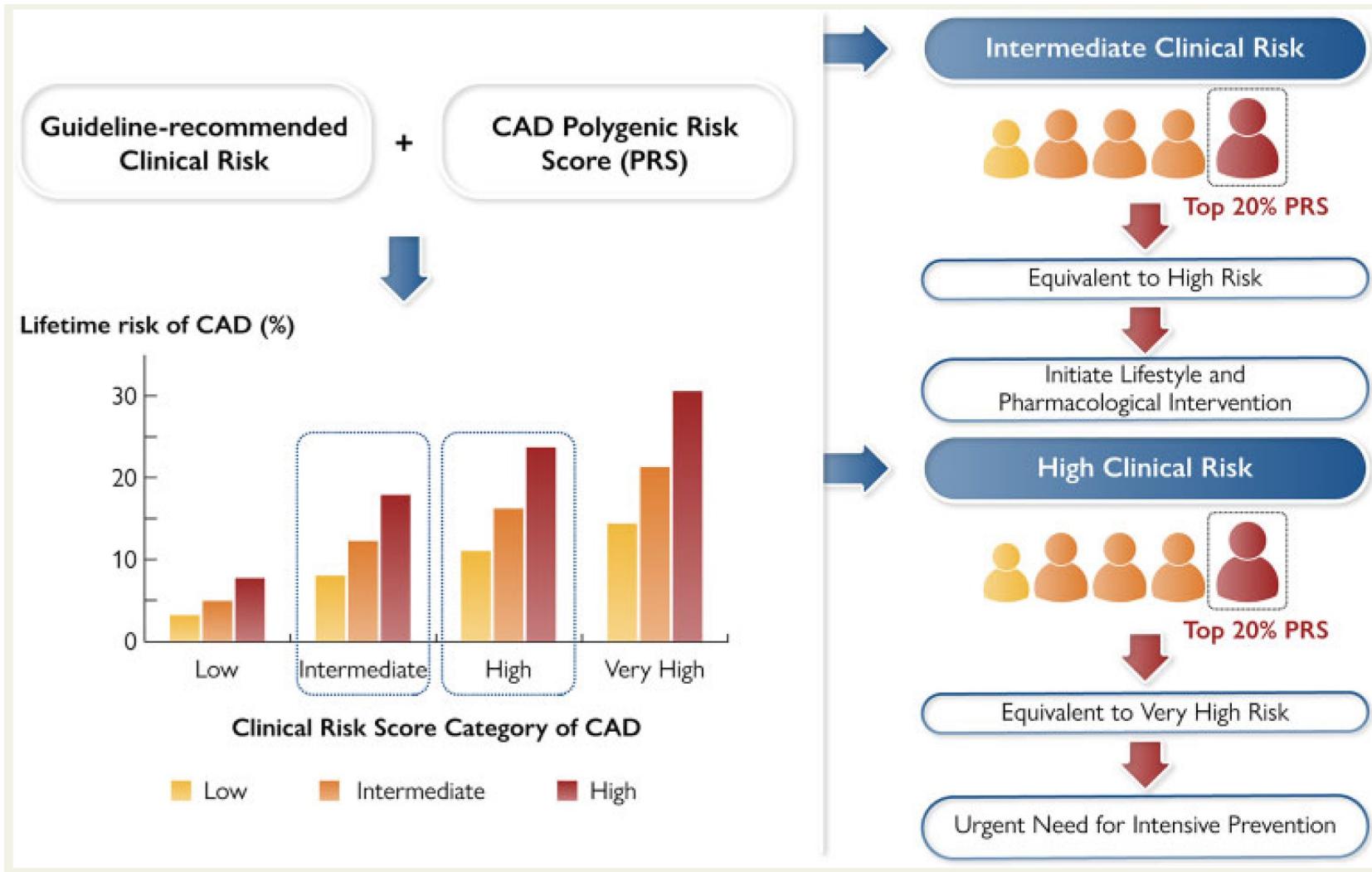
Search...  
breast cancer, glaucoma, EFO\_0001645

	PGP000006 » Khera AV et al. Nat Genet (2018)	Type 2 diabetes	type II diabetes mellitus	6,917,436	G D E	Terms/Lice
	PGP000006 » Khera AV et al. Nat Genet (2018)	Inflammatory bowel disease	inflammatory bowel disease	6,907,112	G D E	Terms/Lice
	PGP000006 » Khera AV et al. Nat Genet (2018)	Atrial fibrillation	atrial fibrillation	6,730,541	G D E	Terms/Lice
	PGP000006 » Khera AV et al. Nat Genet (2018)	Coronary artery disease	coronary artery disease	6,630,150	G D E	Terms/Lice
	PGP000090 » Wang M et al. J Am Coll Cardiol (2020)	Coronary artery disease	coronary artery disease	6,630,150	G D E	Terms/Lice
	PGP000100 » Mars N et al. Nat Med (2020)	Prostate cancer	prostate carcinoma	6,606,785	G D E	
	PGP000100 » Mars N et al. Nat Med (2020)	Type 2 diabetes	type II diabetes mellitus	6,437,380	G D E	

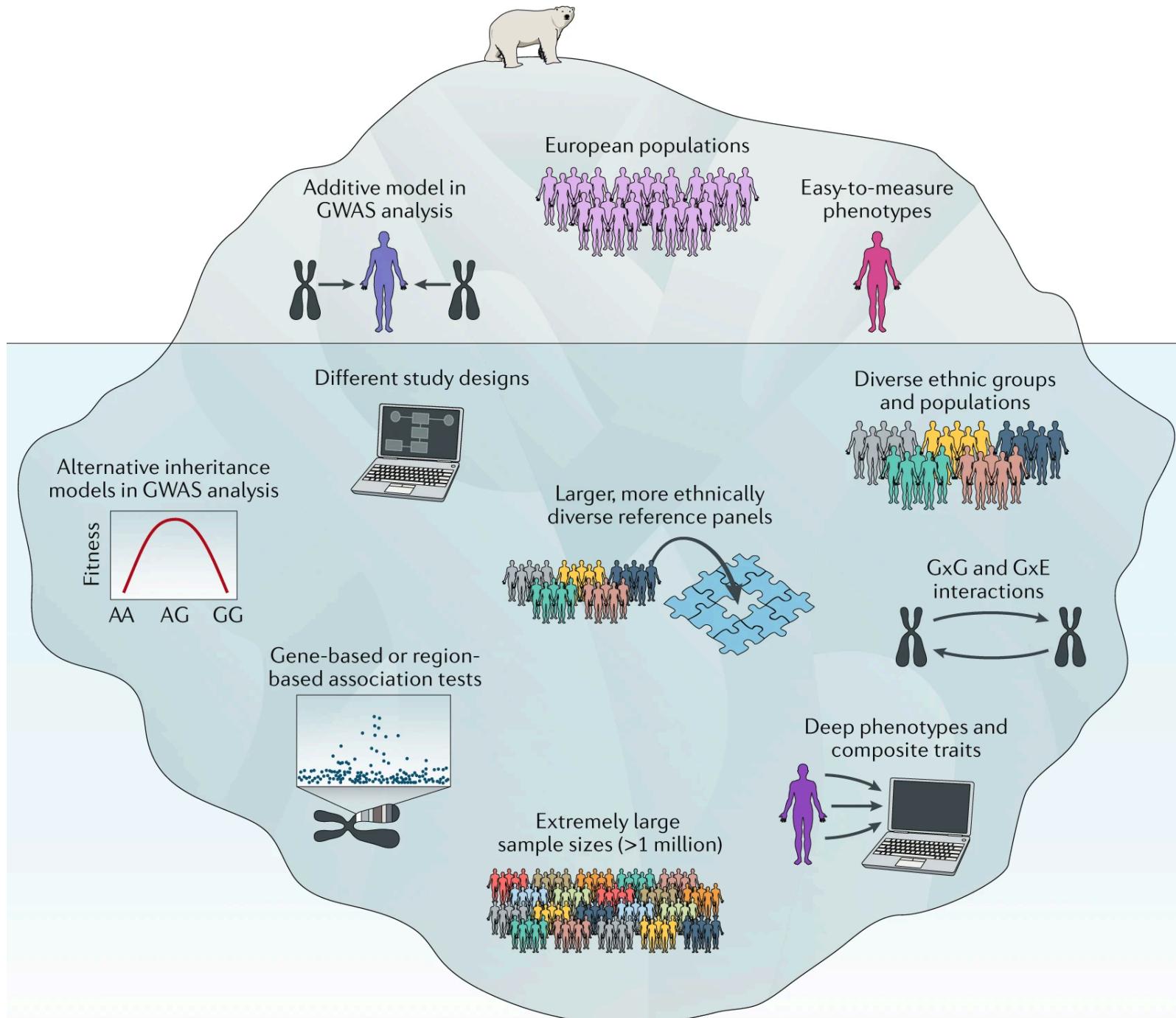
Ancestry legend ⓘ

- Multi-ancestry (including European)
- Multi-ancestry (excluding European)
- African
- East Asian
- South Asian
- European
- Greater Middle Eastern
- Hispanic or Latin American
- Additional Diverse Ancestries
- Not Reported
- Additional Asian Ancestries

# Asian-specific PRS on coronary artery disease (CAD)



- Asian-specific PRS generated using 540 variants
- PRS trained on 2800 patients with CAD and 2055 controls
- Asian-specific PRS outperforms European-based PRS of 1.7M and 6.6M variants trained on thousands of European samples



# Precision genomics

## Understanding

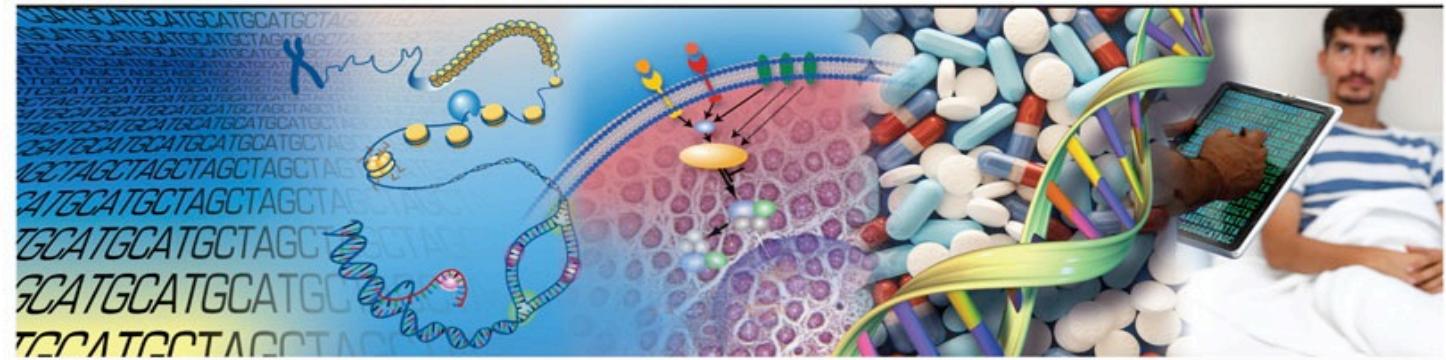
structure of genome

biology of genome

biology of disease

Advancing science of medicine

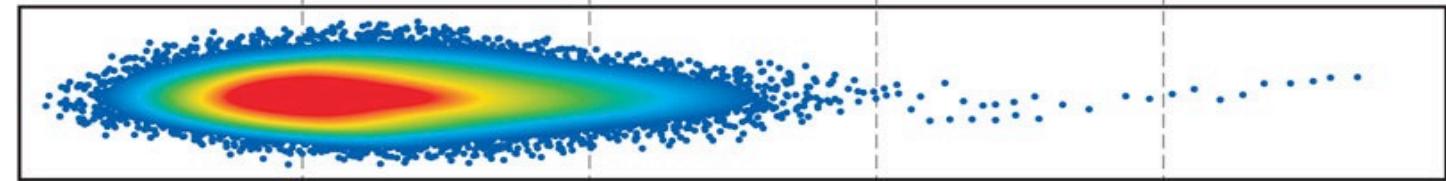
Improving effectiveness of healthcare



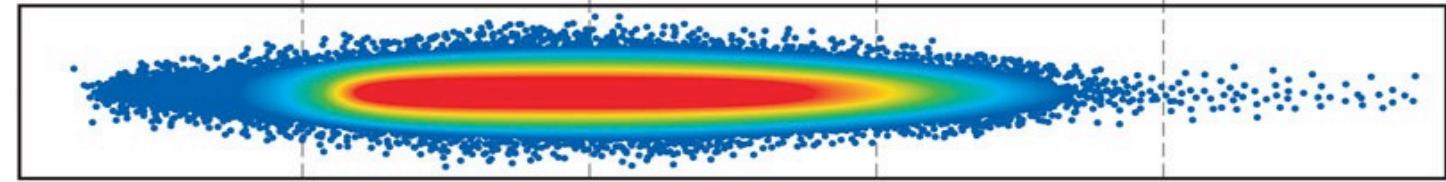
Before 2003  
Human Genome Project



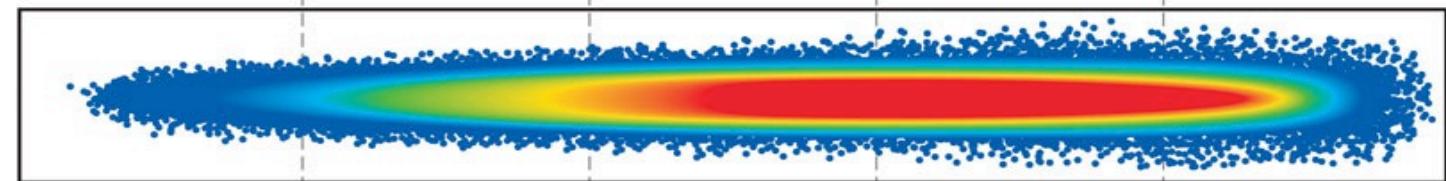
2004 -  
2010



2011 -  
2020



Beyond  
2020



Where were we?

Where are we?

Where will we be?

# Thank you

