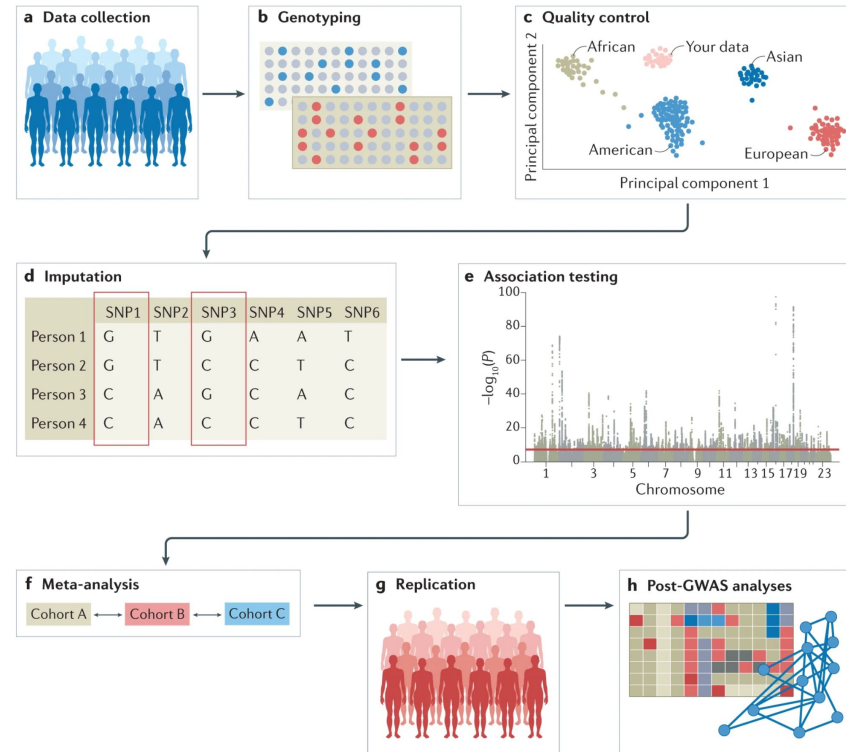


# Genetic Risk Scores

stratification & disease trajectories

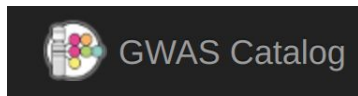
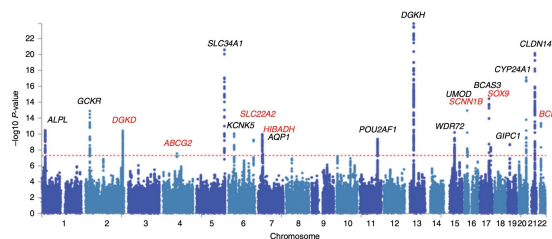
# Genome-wide Association studies (GWAS)

**Single nucleotide polymorphism (SNP):** This is a variation in a single nucleotide (i.e., **A**, **C**, **G**, or **T**) that occurs at a specific position in the genome. A SNP usually exists as two different forms (e.g., **A** vs. **T**). These different forms are called alleles. A SNP with two alleles has three different genotypes (e.g., **AA**, **AT**, and **TT**).



# Data sources & repositories

## Summary statistics

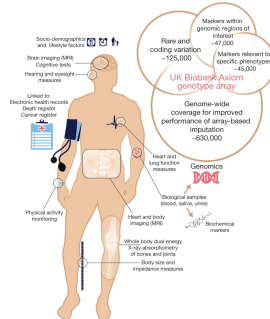


general GWAS data repository



specific for GWAS Chronic Kidney Disease

## Individual-level

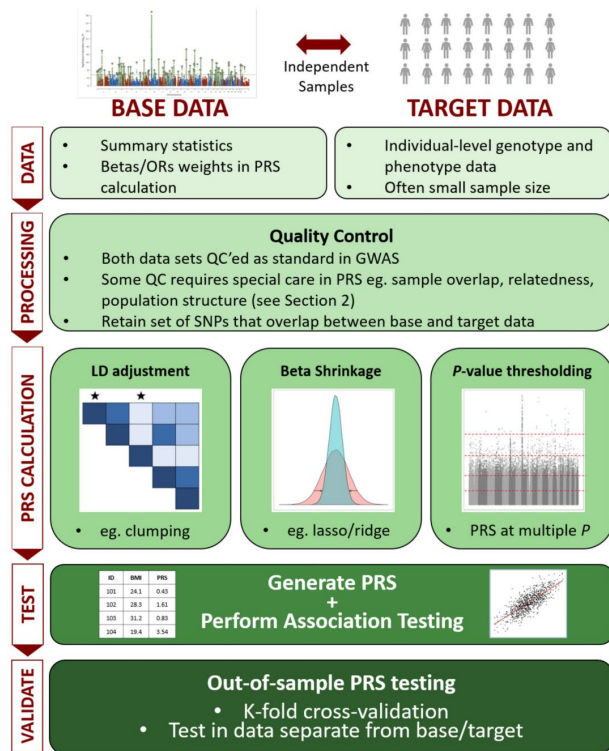


### Genotype data

	SNP1	SNP2	SNP3	SNP4
Individual 1	AT	CG	TT	CC
Individual 2	TA	GG	GT	CA
Individual 3	TT	CC	GT	CA
Individual 4	TT	CC	GG	AA



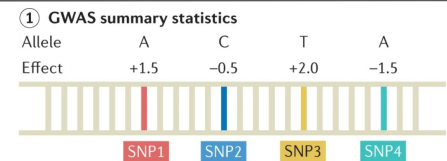
# Common workflow - single-trait PRS



## Polygenic risk score (PRS)

$$PRS_i = \sum_{j \in J} \beta_j G_{ij}$$

$i$ -th individual  
 $j$ -th variant  
 $G$ : genotype  
 $\beta$ : effect size

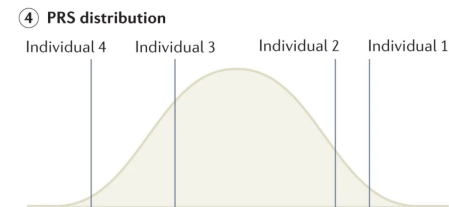


② Genotype data

	SNP1	SNP2	SNP3	SNP4
Individual 1	AT	CG	TT	CC
Individual 2	TA	GG	GT	CA
Individual 3	TT	CC	GT	CA
Individual 4	TT	CC	GG	AA

③ Polygenic risk score

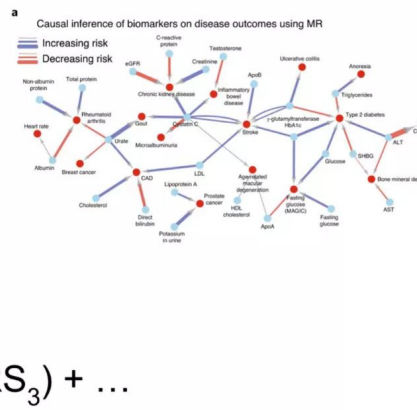
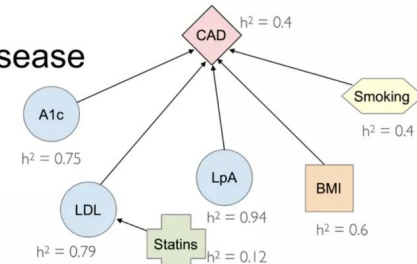
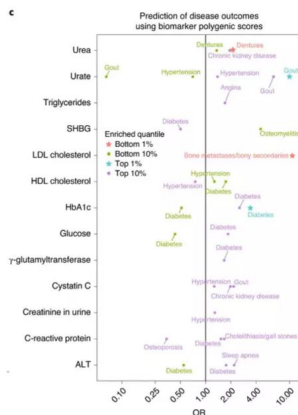
Individual 1	1.5	-	0.5	+	4.0	-	0.0	=	5.0
Individual 2	1.5	-	0.0	+	2.0	-	1.5	=	2.0
Individual 3	0.0	-	1.0	+	2.0	-	1.5	=	-0.5
Individual 4	0.0	-	1.0	+	0.0	-	3.0	=	-4.0



# Common workflow - multi-trait PRS

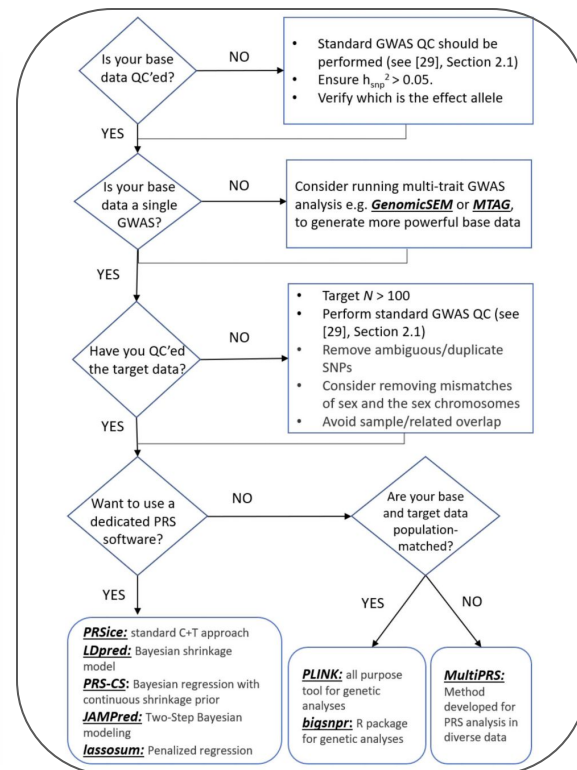
## - Multiple observations suggest “biomarkers → disease” links

- PRS-PheWAS analysis
- Biomarkers are more heritable than disease
- Mendelian Randomization



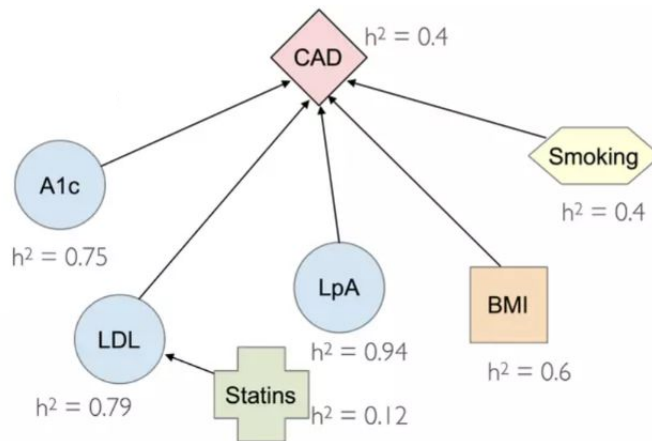
## - Multi-PRS is a weighted sum of PRSs

$$\text{i.e. } w_1(\text{PRS}_1) + w_2(\text{PRS}_2) + w_3(\text{PRS}_3) + \dots$$

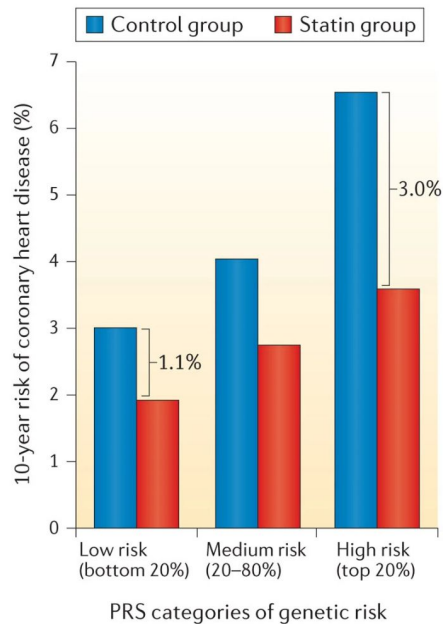


# Common workflow - multi-trait PRS

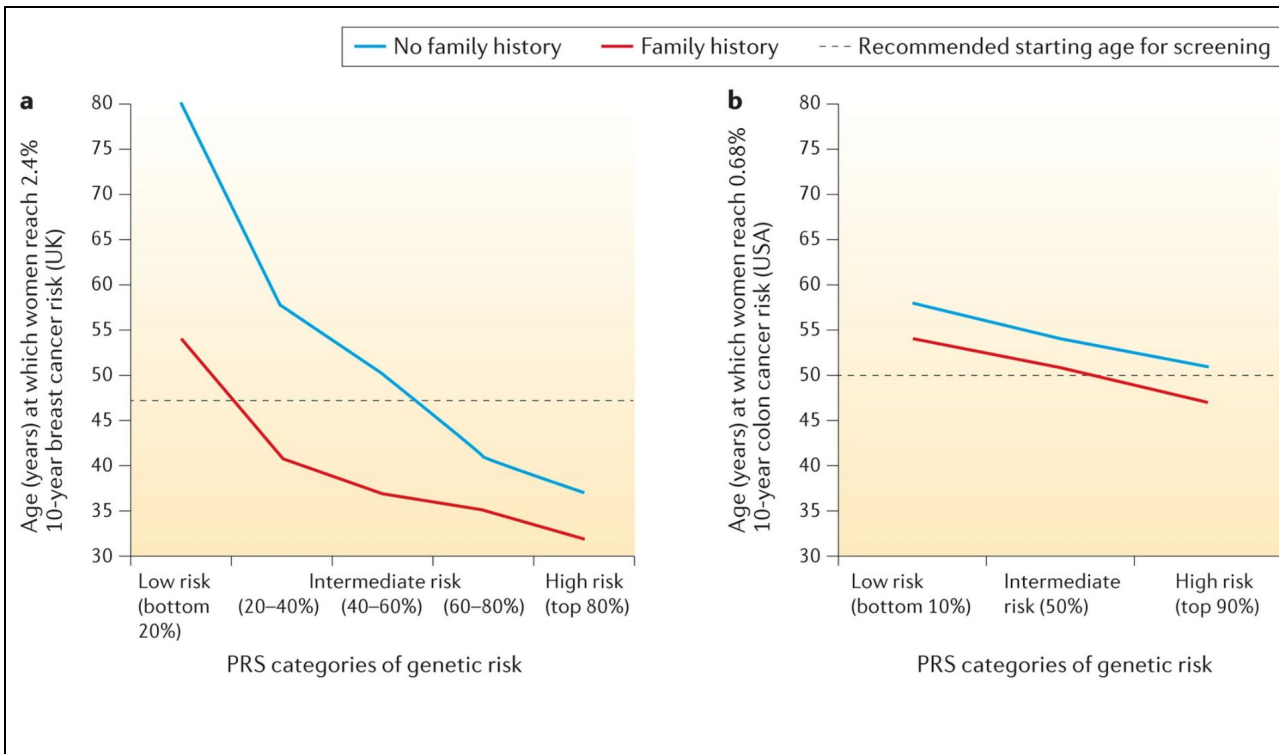
- Multi-PRS is a weighted sum of PRSs  
i.e.  $w_1(\text{PRS}_1) + w_2(\text{PRS}_2) + w_3(\text{PRS}_3) + \dots$



# Case study - disease stratification

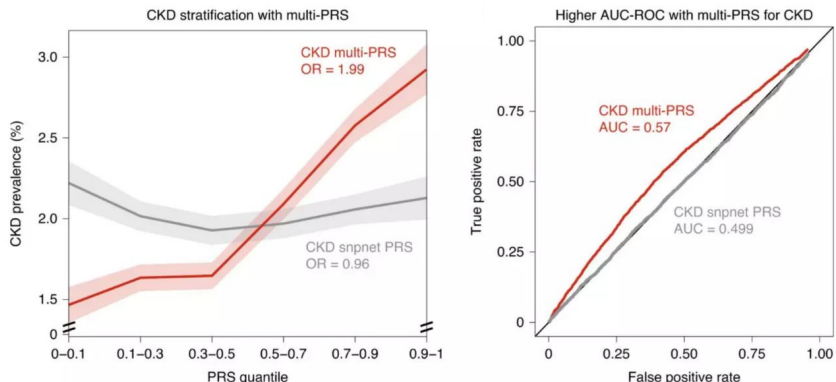


ARR (%)	1.1	1.3	3.0
RRR	0.36	0.32	0.46

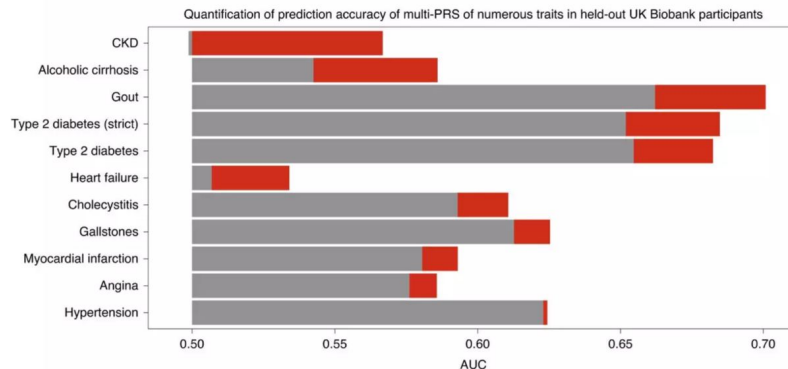


# Case study - multi-trait PRS improves disease prevalence prediction

## Chronic kidney disease (CKD)

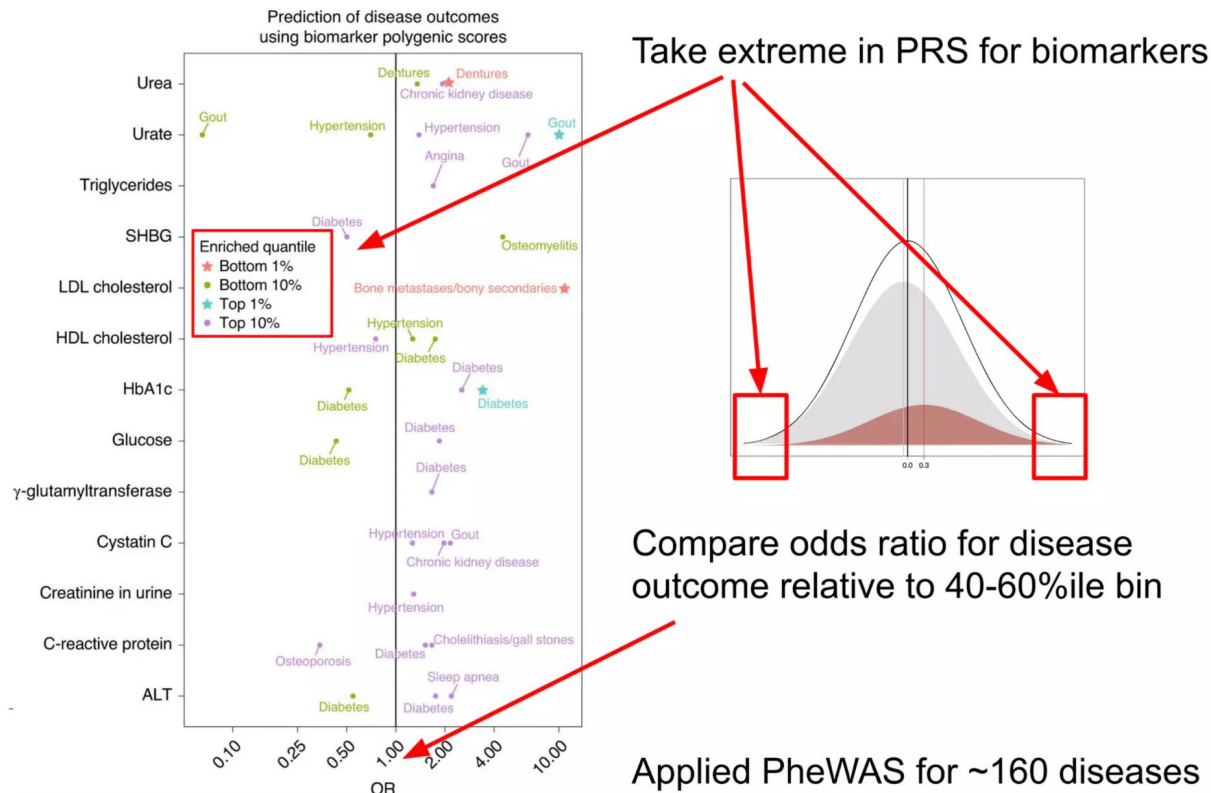


## Other diseases in UK Biobank

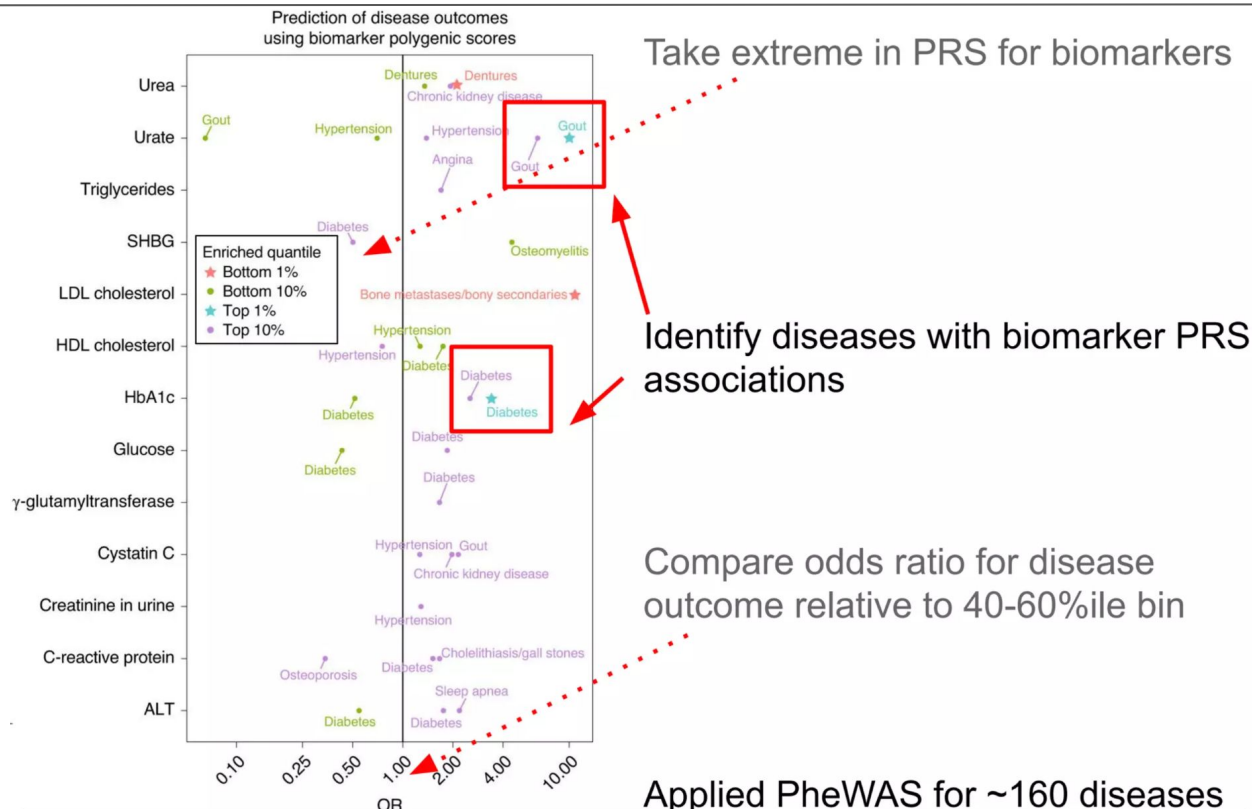




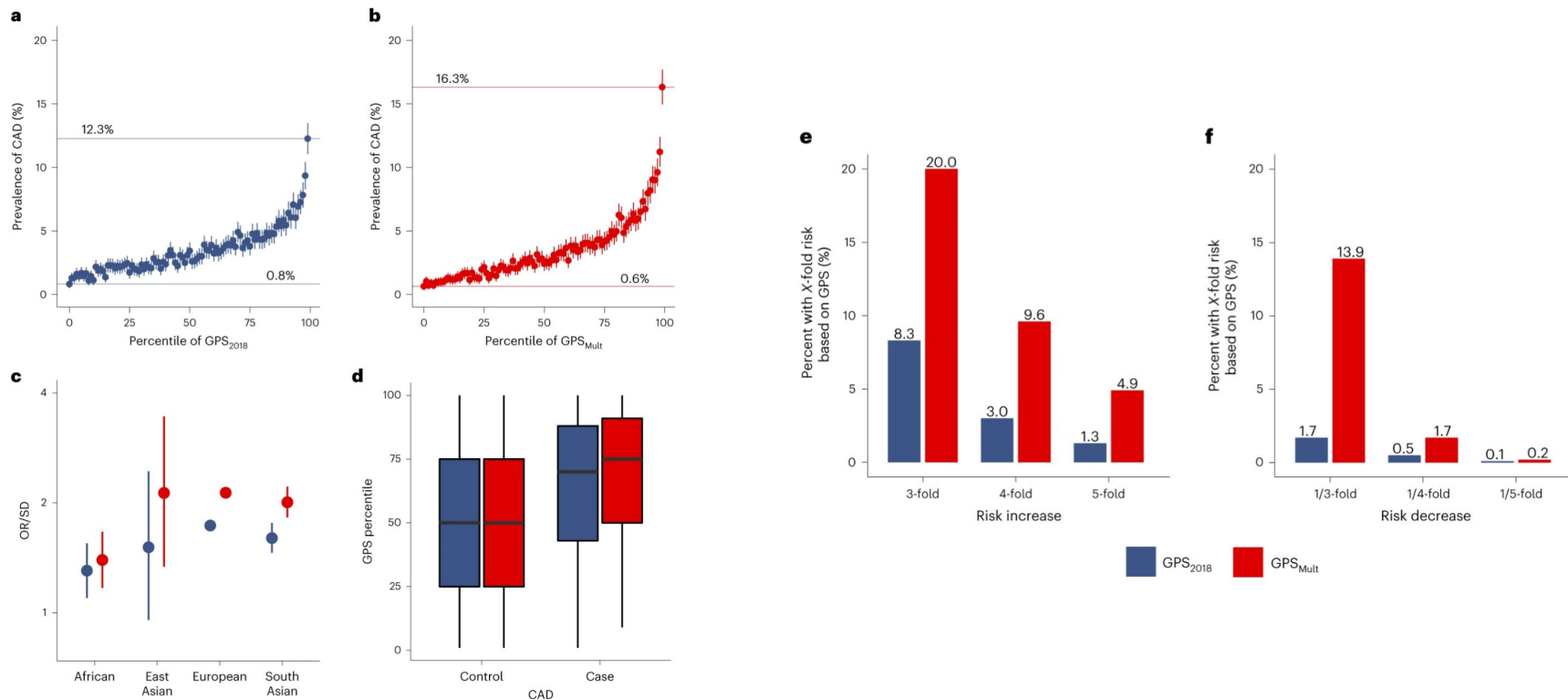
# Use of PRS for trait / disease prediction



# Use of PRS for trait / disease prediction

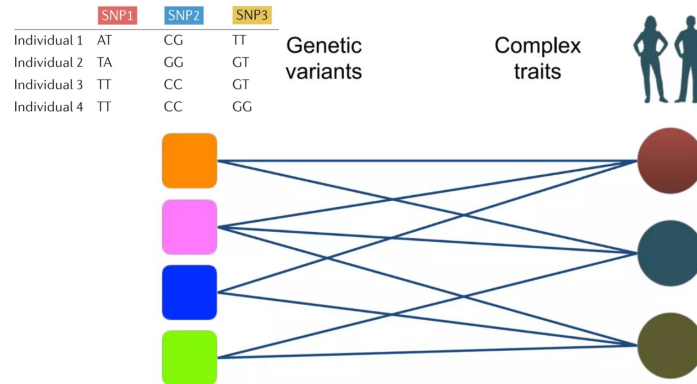


# Limitations - ethnicity / ancestry



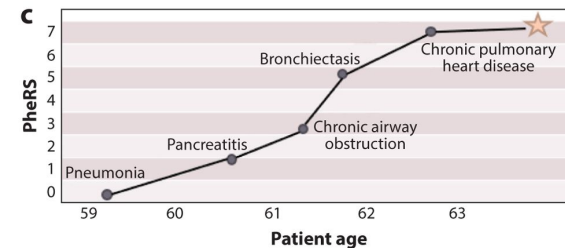
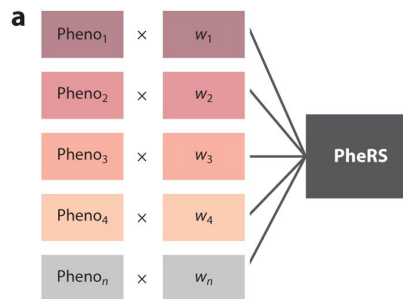
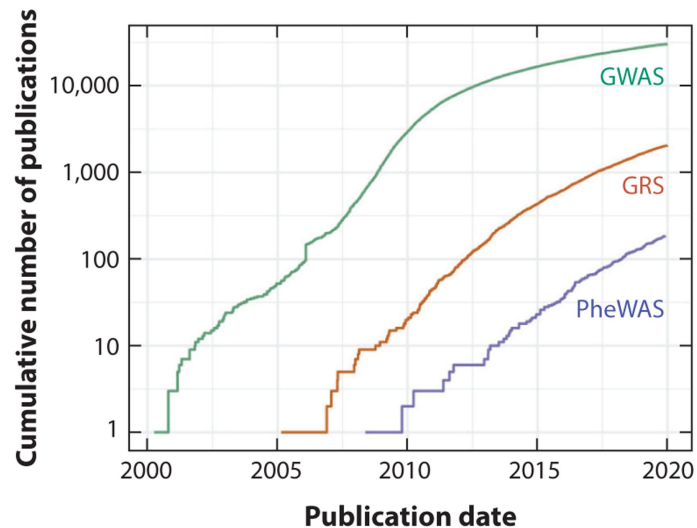
# Limitations - polygenicity & pleiotropy

- Polygenicity: many variants - one trait
- Pleiotropy: one variant - many traits



- Large number of associations in population-based cohorts
- Can we group them together for enhanced interpretation?

# Going further - EHR's & PRS



**b**

Respiratory		HPO	Phecodes and weights		
Airways					
	Chronic bronchopulmonary infection	6,538	483	Acute bronchitis & bronchiolitis	1.00
	Bronchiectasis	2,110	496.3	Bronchiectasis	1.80
	Asthma	2,099	495	Asthma	0.98
Abdomen					
Pancreas					
	Pancreatic insufficiency	1,738	577	Diseases of pancreas	1.42
Growth					
Other					
	Failure to thrive	1,508	264.2	Failure to thrive	1.62

# Summary

---

Two complementary approaches to improve predictive performance:

- Sample size → increase in **statistical power**
  - Multi-trait PRS analysis

Why does multi-PRS work?

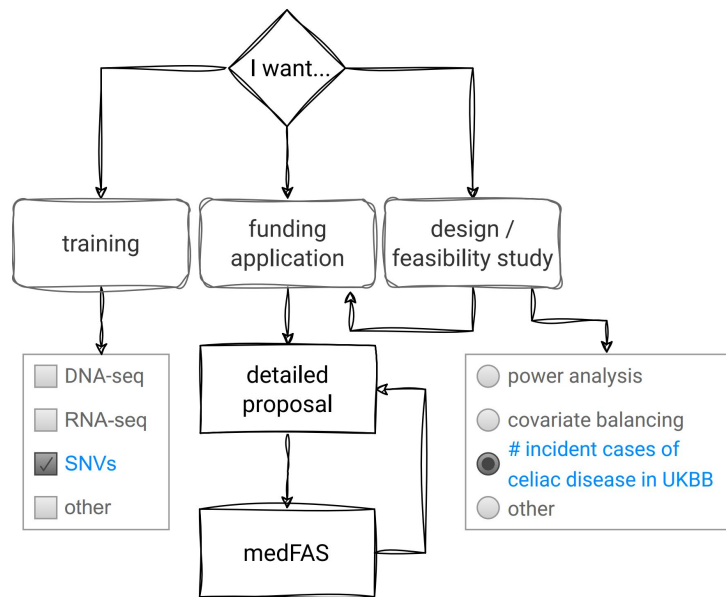
- Quantitative traits have more power
- **Genetic correlation** between biomarkers and disease

The multi-trait PRS model:

Genetics → Biomarkers (molecular traits) → Disease

# Bioinformatics Core

Reach us by filling the form



University of St Andrews 1413

## Bioinformatics Core

Hi, Marco. When you submit this form, the owner will see your name and email address.

\* Required

### Motivation for contact

1. I'm a... \*

☐ undergrad student

☐ postgrad student

☐ PhD student

☐ staff

☐ Other

2. Department / School. \*

Select your answer

3. I'm looking for... \*

☐ training

☐ help with experimental design

☐ help with funding application