FRIDAY (all times in PT)			
8:30 – 9:15	Alie	Bioethics/ Implementation	Bioethical principles in genetic epidemiology, deciding whether to implement genetic testing
BREAK			
9:30-10:15	Sara	Rare variation	Strategies to analyze rare genetic variants
BREAK			
10:30-11:15	Sara	GxE interactions	Gene x Environment interactions analyses
LUNCH BREAK			
11:45 – 12:30	Diane	PRS/Risk prediction	Polygenic risk scores and population screening
BREAK			
12:45-1:30	Sara	Mendelian Randomization	Mendelian Randomization studies
BREAK			
1:45 – 2:30	Sara/Alie	Office Hours	Stop by to ask questions from the day, or schedule time to discuss your own project.

EPIDEMIOLOGY
SCHOOL OF PUBLIC HEALTH

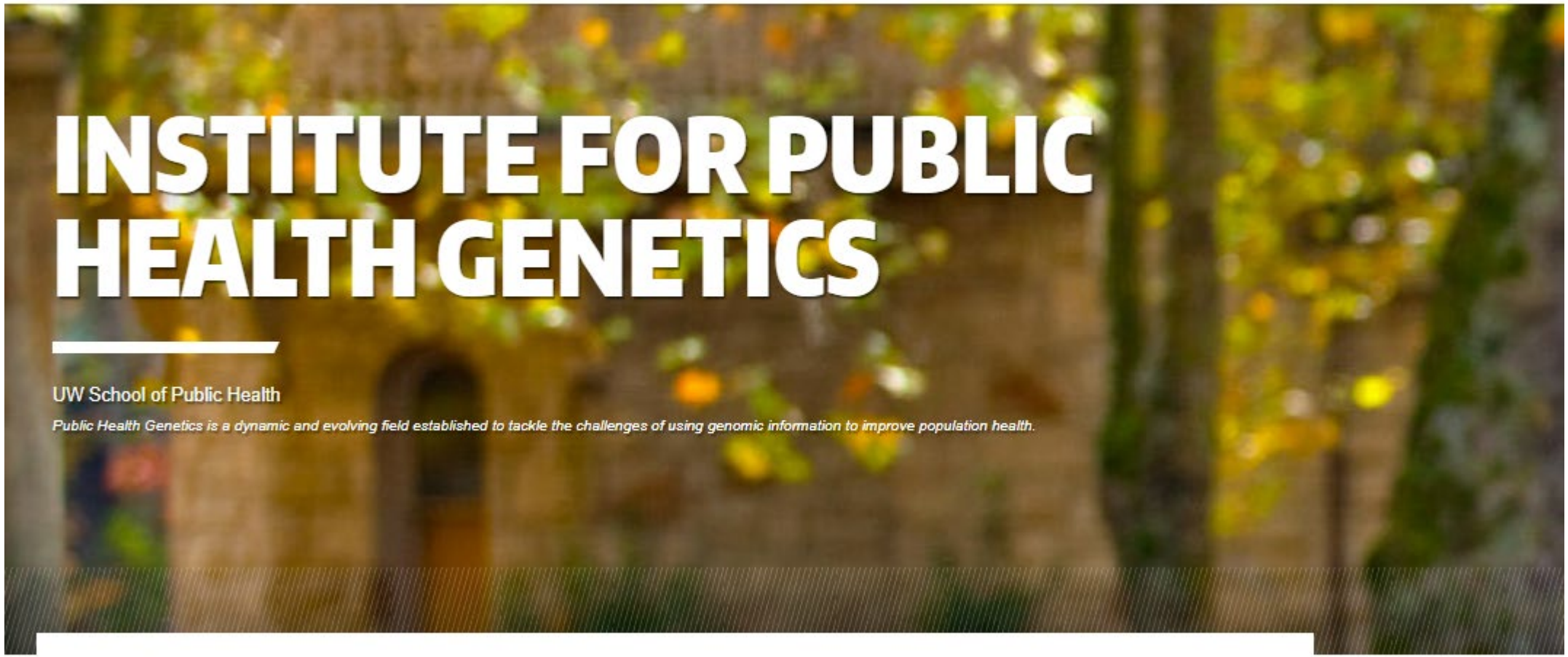
SCHOOL OF PUBLIC HEALTH
EPIDEMIOLOGY
UNIVERSITY of WASHINGTON

SISG 2022: Module 11 Genetic Epidemiology

Risk Prediction and Population Screening

Course survey:

https://si.biostat.washington.edu/user/login



INSTITUTE FOR PUBLIC HEALTH GENETICS

UW School of Public Health

Public Health Genetics is a dynamic and evolving field established to tackle the challenges of using genomic information to improve population health.

Apply for Autumn 2022

The admission cycle for Autumn 2022 admissions has closed. The Institute for Public Health Genetics will being accepting applications for the Autumn 2023 admission cycle on September 1, 2022. Applications for the Autumn 2023 cycle will be due by January 31, 2023, once the application

ANNOUNCEMENTS

- Questions about the program? Contact us: phgadmit@uw.edu.

Freakonomics Radio:

impact and utility of polygenic risk score for lipids

23:00-23:40; 27:30-28:20

http://freakonomics.com/podcast/23andme/

Polygenic Risk Scores

Genetic contribution to disease is complex

Only 1-10% of disease is thought to be driven by rare, high impact variants.

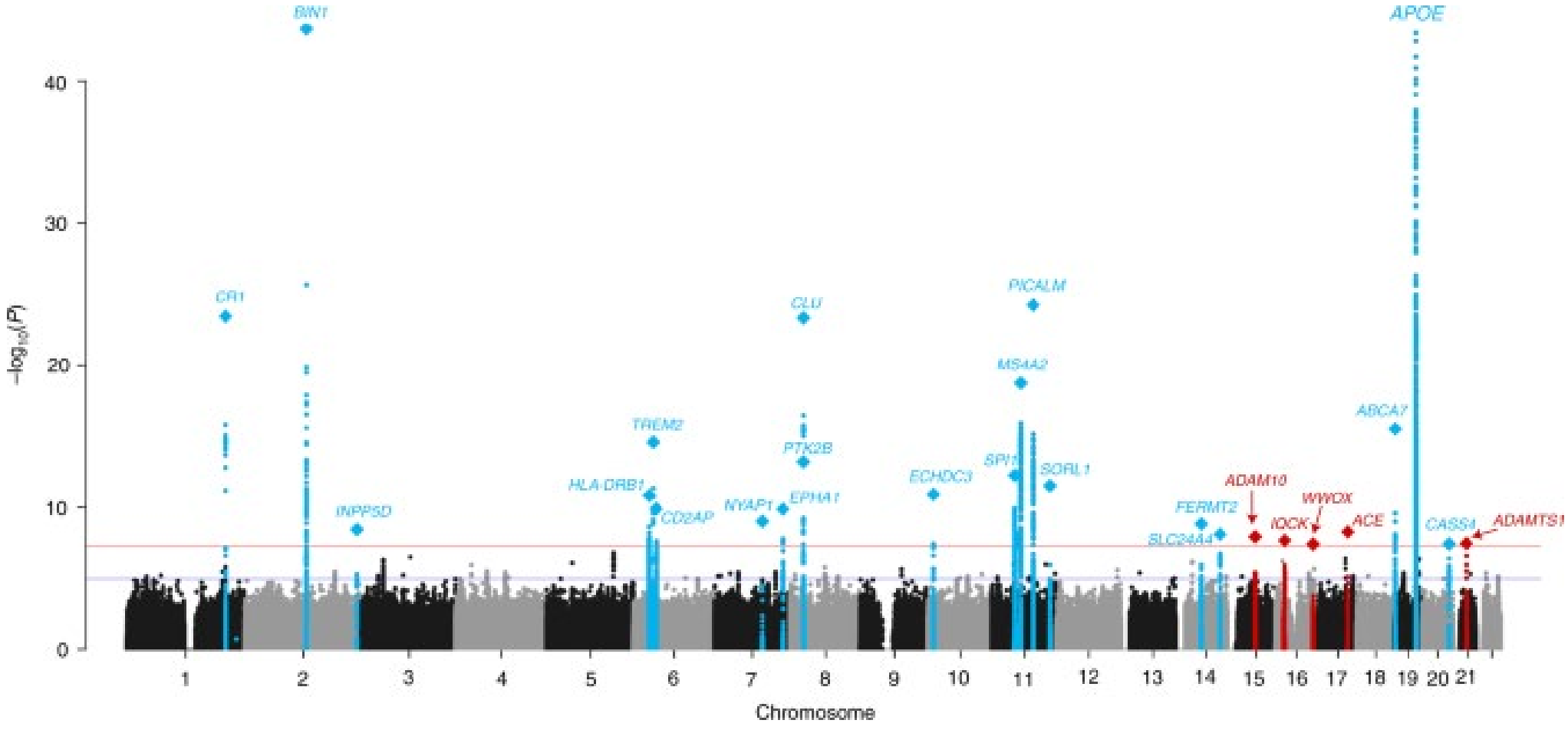
If you have these variants, you have a very high chance of developing the disease, but those variants account for a small amount of overall people with the disease

- ***BRCA1* in breast cancer**
  - 45% lifetime risk with the variant
  - only 5-10% of breast cancer is linked to *BRCA1*

- ***LDLR* in Familial hypercholesterolemia**
  - 66% risk of heart disease
  - only 2% of people with heart disease is linked to *LDLR*

What about the rest of disease?

GWAS of Alzheimer's disease



Kunkle *et al*. 2019 (Nature)

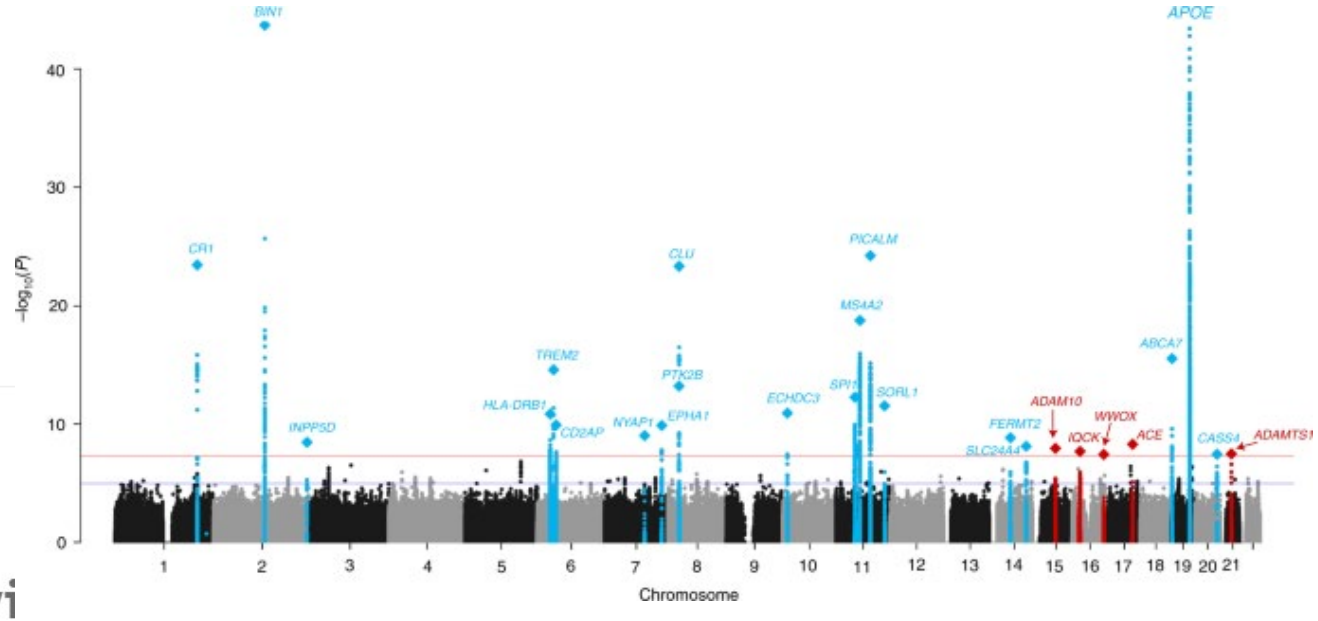
What about the rest of disease?

Explore more

You can find out whether you may have an increased risk of developing late-onset Alzheimer's disease based on your genetics with the 23andMe Late-Onset Alzheimer's Disease Genetic Health Risk report*. The report looks for the ε4 variant in the APOE gene associated with late-onset Alzheimer's disease. The report is available through the 23andMe Health + Ancestry Service.

Health + Ancestry Servi

Learn more



"Polygenic" risk = Sum of genome-wide risk

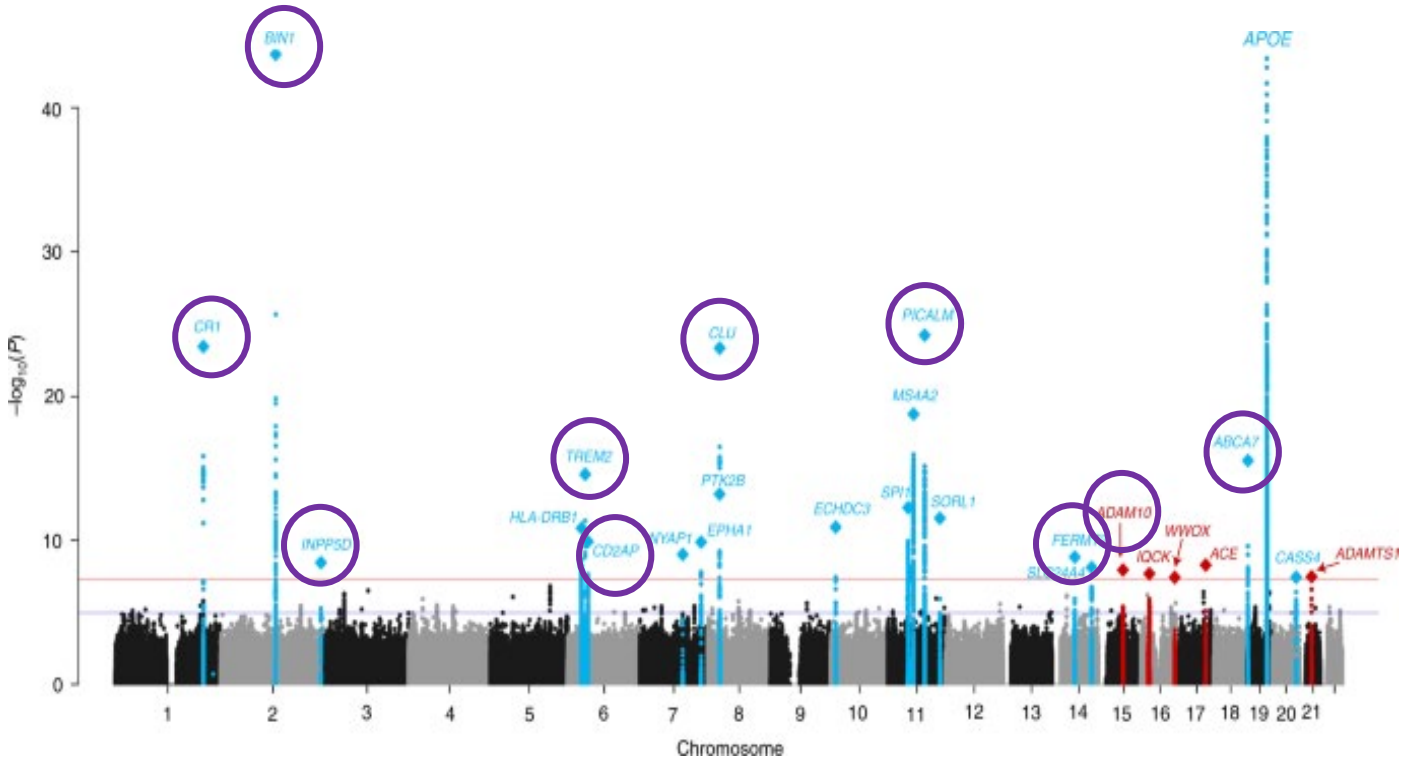
PRS

$$\hat{S} = \sum_{j=1}^{m} X_j \hat{\beta}_j$$

# of SNPs meeting the significance threshold

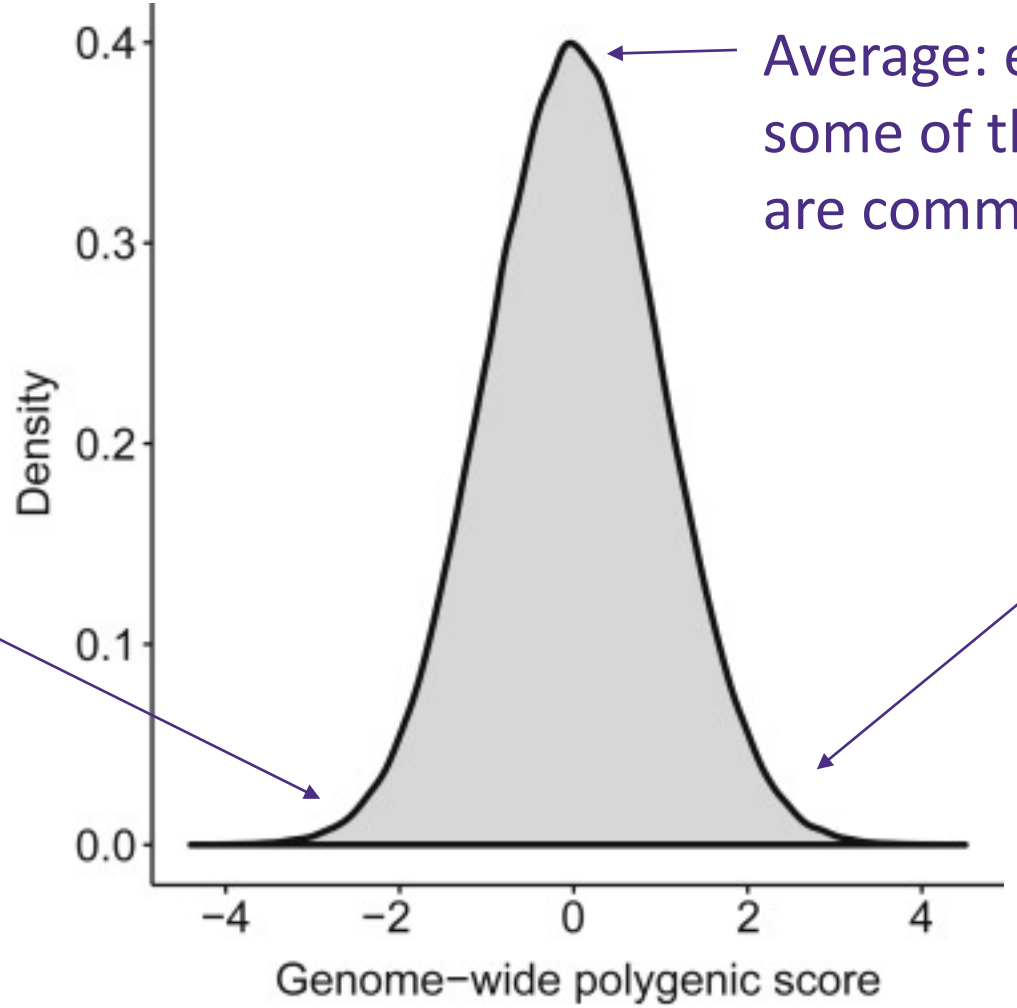
allele

effect size



Kunkle *et al*. 2019 (Nature)

Population polygenic risk score distribution

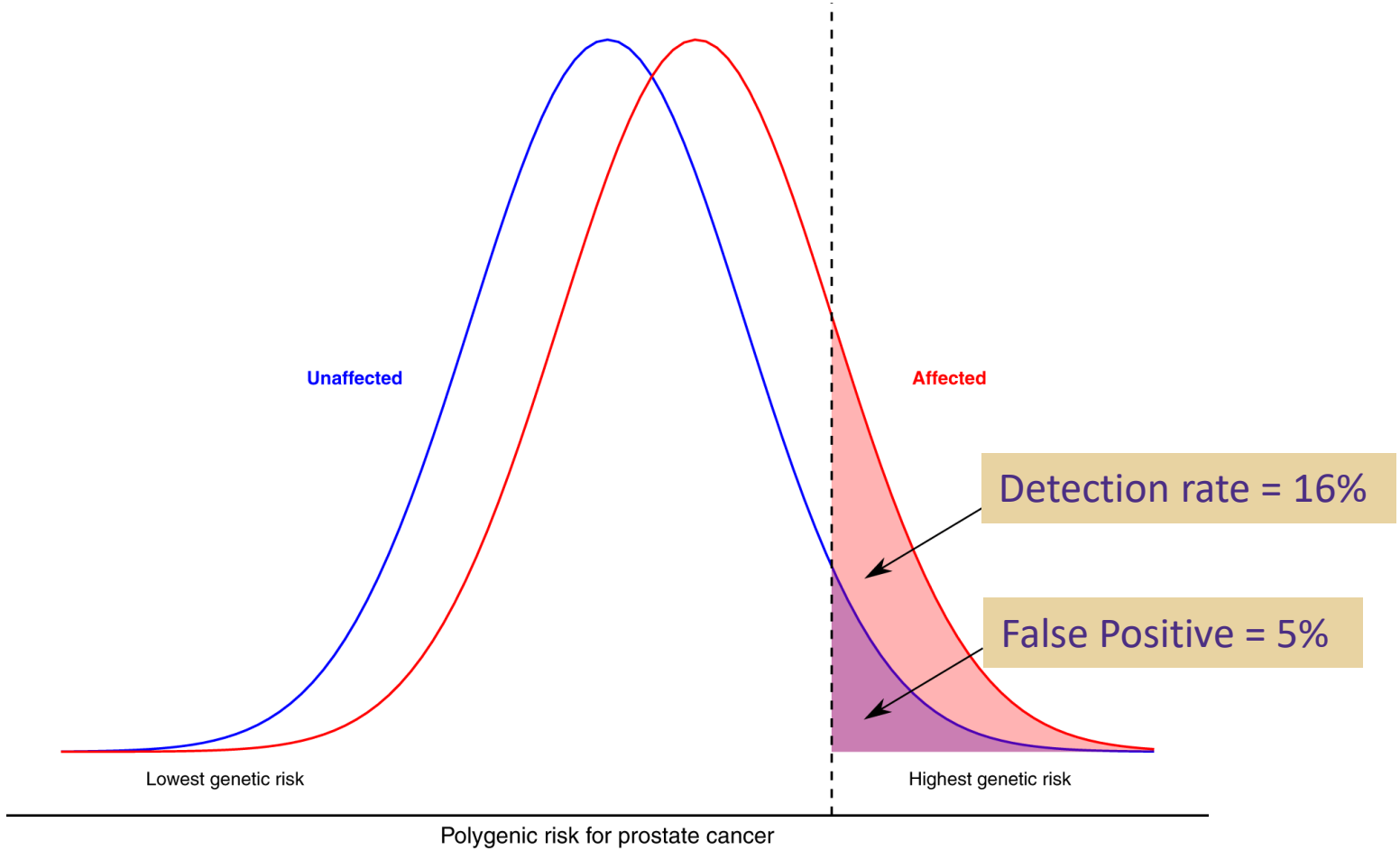


Average: everyone probably has some of these variants, since most are common in GWAS

Very few ‘risk’ variants or many protective variants

A lot of risk variants

Population polygenic risk score distribution



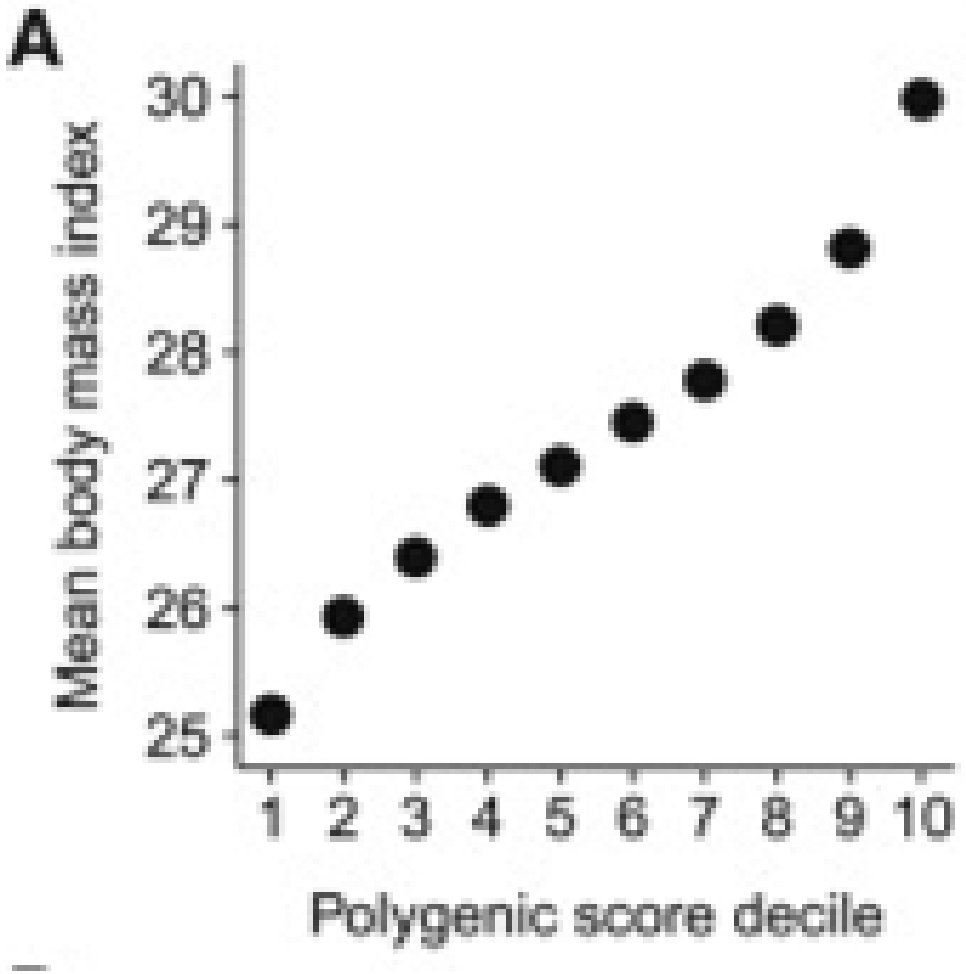
Is this amount of risk discrimination useful?

Sud *et al*. 2021 (Nature)

https://www.cell.com/cell/pdf/S0092-8674(19)30290-9.pdf

Another Example

Polygenic risk score for BMI

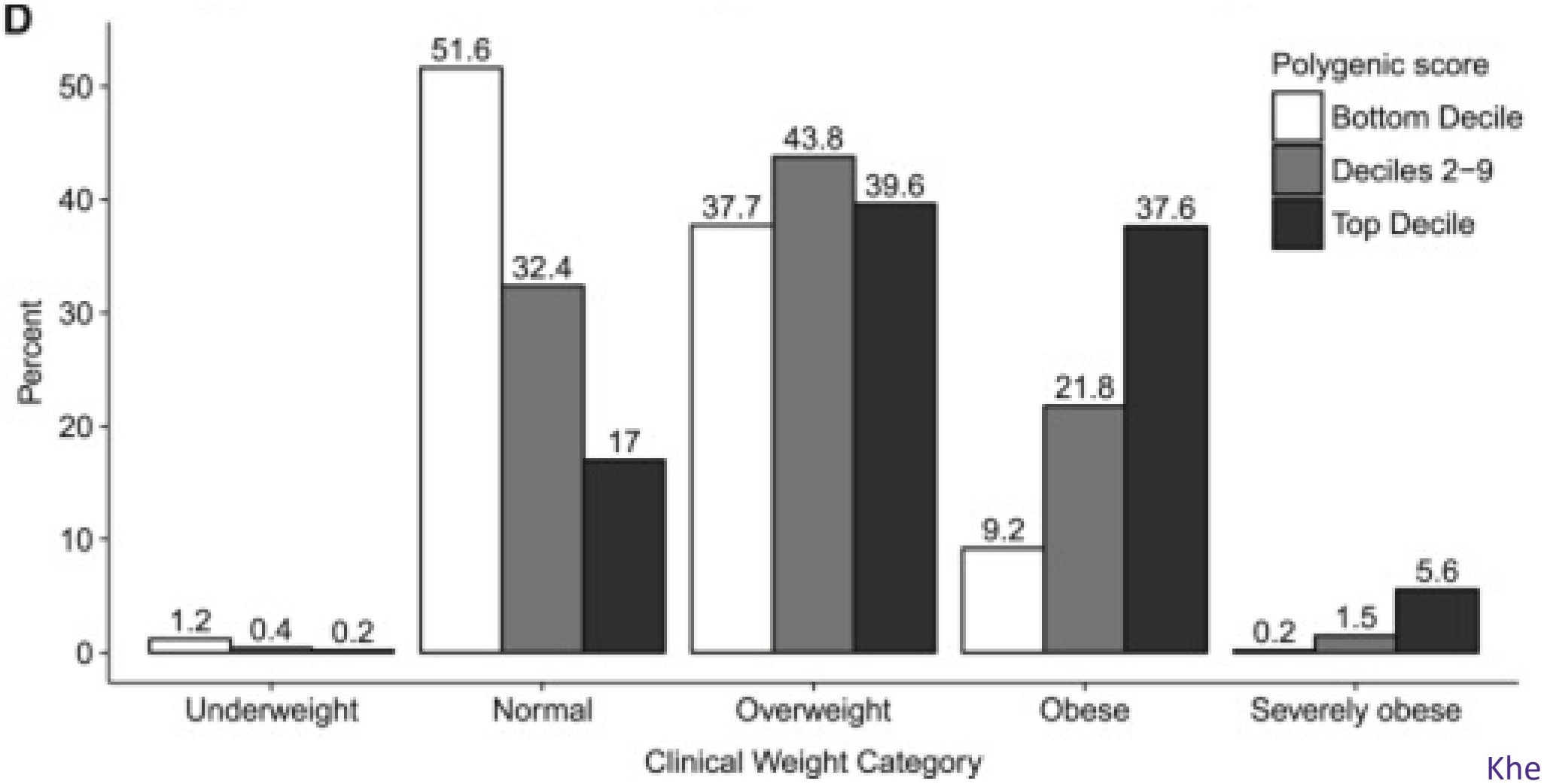


Cohort divided into people with each genetic score level

Khera, Cell. 2019

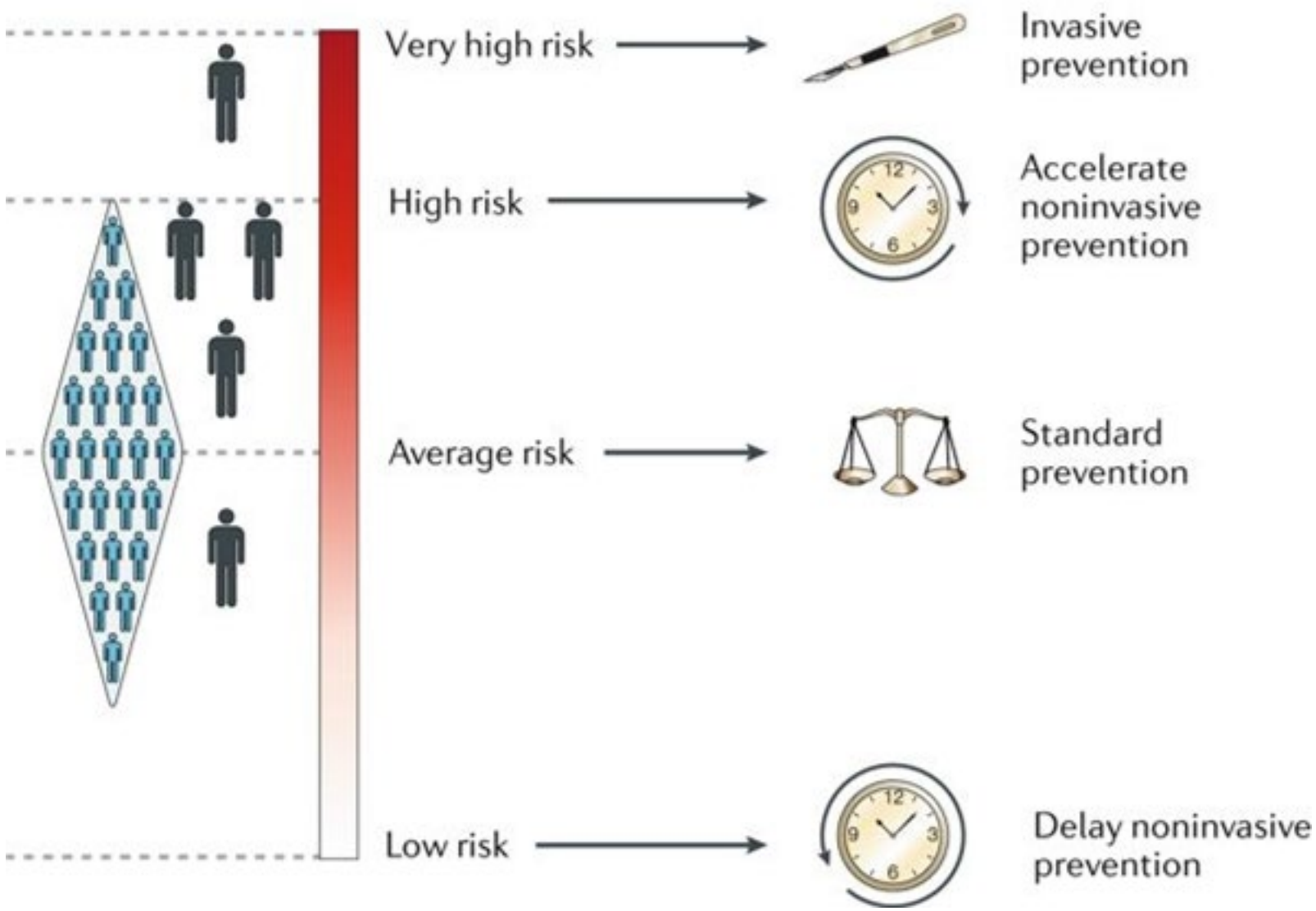
Polygenic risk scores are still probabilistic

Based on polygenic risk score category, what weight score do people have?



Khera, Cell. 2019

Polygenic scores as another data point → Precision health



Torkmani, Nature, 2018

From GWAS to PRS

From GWAS to PRS

1) Start with GWAS Summary Statistics

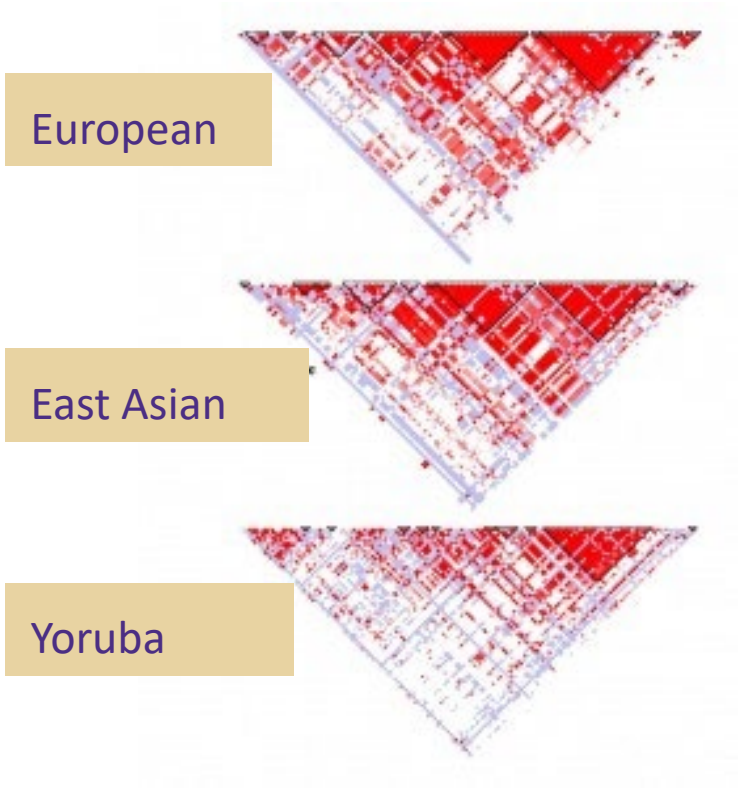
	CHR	SNP	BP	BETA	SE	R2	P
1	20	rs1418258	11799	0.031600	0.02845	4.353e-04	0.2669
2	20	rs6086616	16749	-0.012000	0.02998	5.650e-05	0.6891
3	20	rs6039403	17094	-0.001804	0.02754	1.520e-06	0.9478
4	20	rs6135141	22347	0.008119	0.02927	2.720e-05	0.7815
5	20	rs1935386	35416	0.010570	0.02786	5.090e-05	0.7043
6	20	rs6051659	39508	0.044900	0.03596	5.506e-04	0.2120

Effect Size

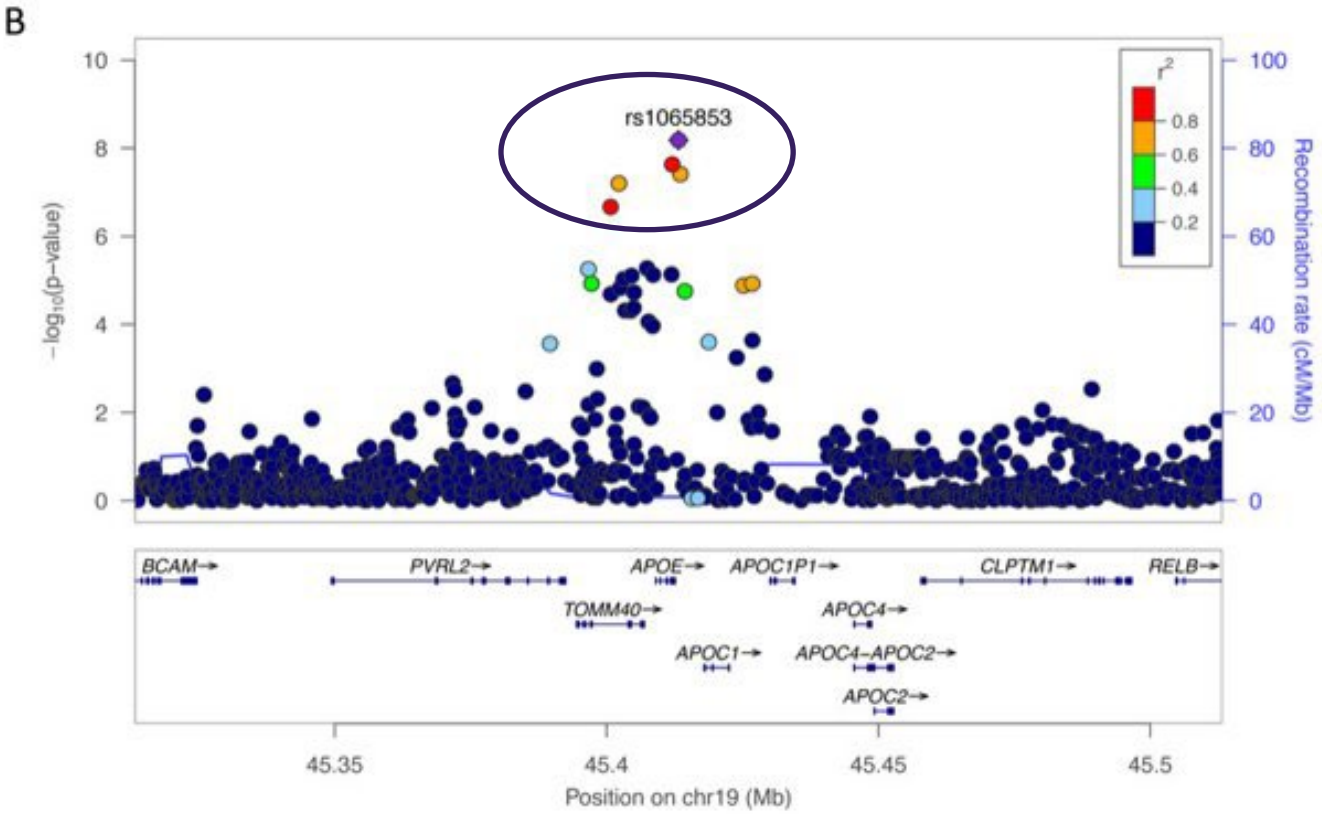
P-value (to determine cut-off)

From GWAS to PRS

2) Account for Linkage Disequilibrium (clumping)



You can see haplotypes clearly here

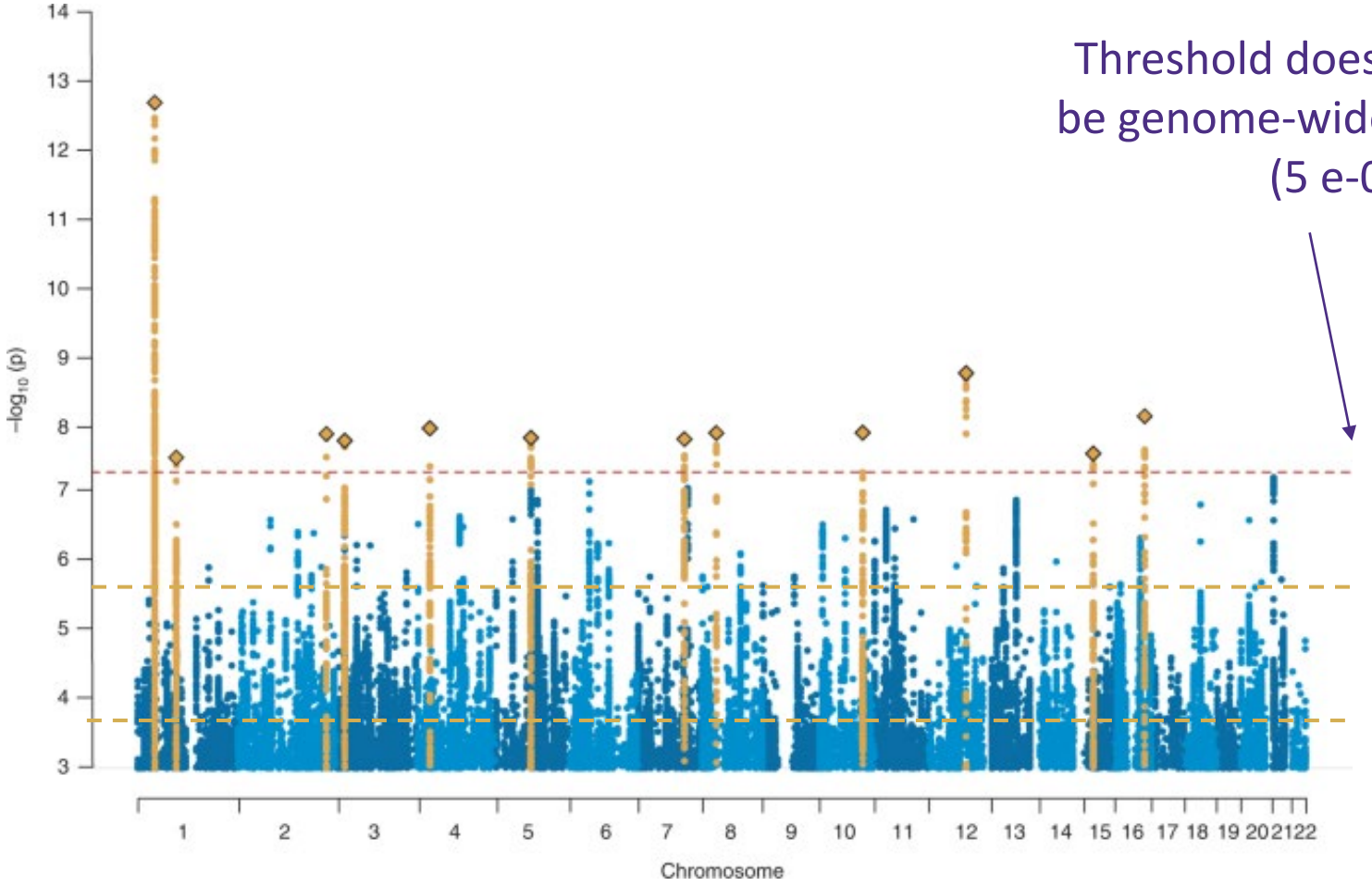


Locus zoom plot

Zoom: Why would a polygenic risk score developed in a European cohort be unreliable for a person with Yoruba ancestry?

From GWAS to PRS

3) Determine SNP inclusion threshold



Threshold does not have to be genome-wide significance (5 e-08)

Note: what is the reference allele?

4) Generate score model

BMI as the outcome

SNP	estimate	error	p-value
rs441084	1.20	0.89	$5x10^{-6}$
rs8783	0.50	0.22	$8x10^{-8}$
rs4699	-0.24	0.19	$6x10^{-7}$

Score = 1.20*(#alleles rs441084) + 0.50 *(#alleles rs8783) - 0.24*(#alleles rs4699)

This is calculated for each individual

Calculate polygenic risk score for Fred

			FRED
Genetic variant	*Effect allele*	*Effect size*	*Genotype*
rs12395	A	0.02	AA
rs44346	G	-0.04	GT
rs72557	C	-0.05	CG
rs18338	A	0.09	AT
rs29849	T	0.004	TT
rs43466	T	0.07	AA
rs29457	G	-0.01	CC
rs13458	C	0.015	AA

*Fred's outcome is not considered

Variant allele -- he gets points depending on how many copies of this allele he has

Calculate polygenic risk score for Fred

			FRED	
Genetic variant	*Effect allele*	*Effect size*	*Genotype*	*Effect*
rs12395	A	0.02	AA	+.02 (x2)
rs44346	G	-0.04	GT	-0.04
rs72557	C	-0.05	CG	-0.05
rs18338	A	0.09	AT	0.09
rs29849	T	0.004	TT	+.004 (x2)
rs43466	T	0.07	AA	
rs29457	G	-0.01	CC	
rs13458	C	0.015	AA	

*Fred's outcome is not considered

Calculate polygenic risk score for Fred

			FRED	
Genetic variant	*Effect allele*	*Effect size*	*Genotype*	*Effect*
rs12395	A	0.02	AA	+.02 (x2)
rs44346	G	-0.04	GT	-0.04
rs72557	C	-0.05	CG	-0.05
rs18338	A	0.09	AT	0.09
rs29849	T	0.004	TT	+.004 (x2)
rs43466	T	0.07	AA	
rs29457	G	-0.01	CC	
rs13458	C	0.015	AA	
		Polygenic score:	0.048	

*Fred's outcome is not considered

Zoom poll:
Calculate polygenic risk score for Alice

			FRED		ALICE
Genetic variant	*Effect allele*	*Effect size*	*Genotype*	*Effect*	*Genotype*
rs12395	A	0.02	AA	+.02 (x2)	TT
rs44346	G	-0.04	GT	-0.04	TT
rs72557	C	-0.05	CG	-0.05	CC
rs18338	A	0.09	AT	0.09	TT
rs29849	T	0.004	TT	+.004 (x2)	CT
rs43466	T	0.07	AA		TA
rs29457	G	-0.01	CC		CC
rs13458	C	0.015	AA		CA

5) Use polygenic score in regression model & Assess accuracy

Is our polygenic score associated with BMI?

R input:

```
Model1 <- lm(BMI ~ PRS + Age + Sex+… , data = BMIdata)
```

We put our summary score in the equation instead of a specific SNP

Q. We use a new dataset to develop this model: Why?

5) Use polygenic score in regression model

Is our polygenic score associated with BMI?

lm(BMI~PRS + Age + Sex, data = BMIdata)

Variable	estimate	error	p-value
Intercept	21.0		
PRS	8.9	0.05	$8x10^{-6}$
Age	0.02	0.01	0.004
SexF	1.0	0.17	0.006

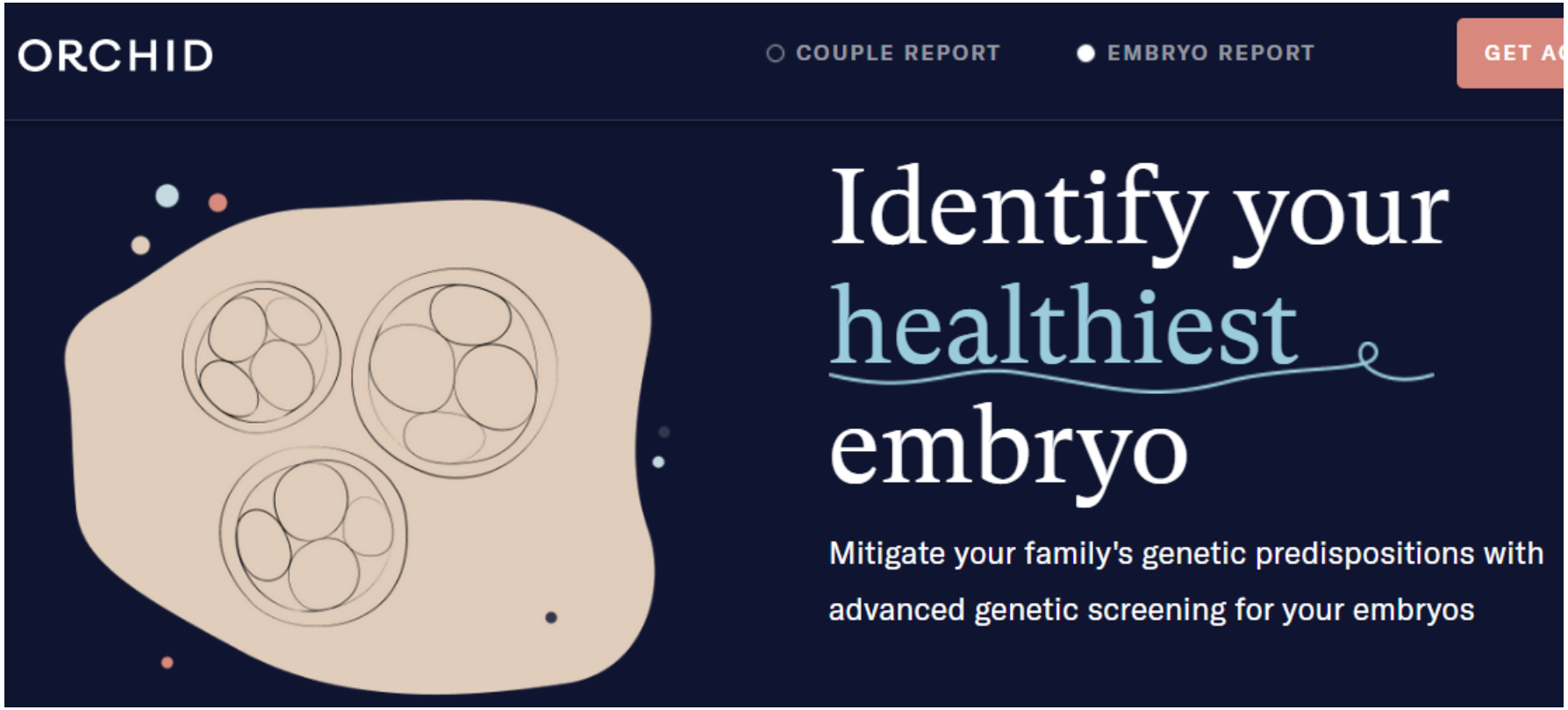
Popular PRS tools

- **PGS Catalog: https://www.pgscatalog.org/**

  - **Similar to GWAS catalog, contains summary information for PRS score development**
  - **You can download the GWAS Summary stats they used to calculate**

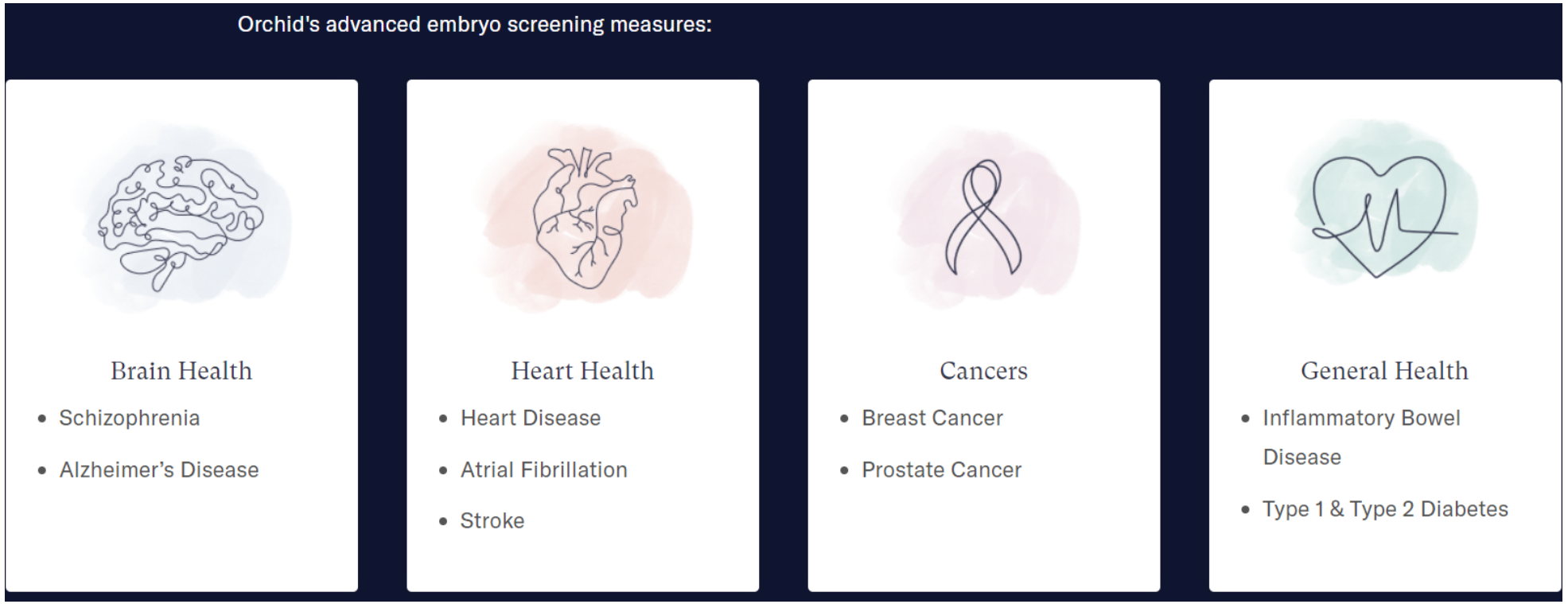
- **PRSsice : uses clumping+thresholding method we discussed**
- **LDPred2: uses LDMatrix and Bayesian method**

Ethical question



Ethical question

Orchid's report relies on what are called polygenic risk scores. These scores estimate the likelihood that an individual will develop a particular condition, based on an analysis of their genome. The data



Ethical question

- **In your zoom breakouts:**
  - Discuss the ethical and social implications of using polygenic risk scores for embryo selection
  - How should OrchidHealth handle rapid scientific developments? What happens if after an embryo is selected, new research comes out that shows that high PRS for one disease is inversely related to another disease?

New methods designed for application in multi-ethnic studies

- **PRS-CSx**
  - Bayesian method that combines GWAS results from multiple populations
    - Con: assumes that variants are mostly the same across ancestries
- **TL-PRS**
  - Transfer learning method to make use of large European GWAS
    - Con: works best with training data from only one homogenous ancestry group
- **TL-Multi**
  - Transfer learning method that uses summary statistics from GWAS of diverse ancestry
    - **Con: no clumping – uses ALL data (computationally inefficient)**

SCHOOL OF PUBLIC HEALTH
EPIDEMIOLOGY
UNIVERSITY of WASHINGTON

Population Screening

Implementing population screening

How do allele frequencies intersect with actionability and economics to make implementation decisions?

Determine how population allele frequencies affect implementation decisions for population screening

Zoom breakout.

Zoom breakout: population screening

Zoom breakout

- Expect 28/100,000 individuals in general population to have cystic fibrosis, but 4162 among people with Ashkenazi Jewish ancestry.
- Integrating bioethical decisions and practical decisions of budgetary constraints