Risk Prediction and Population Screening

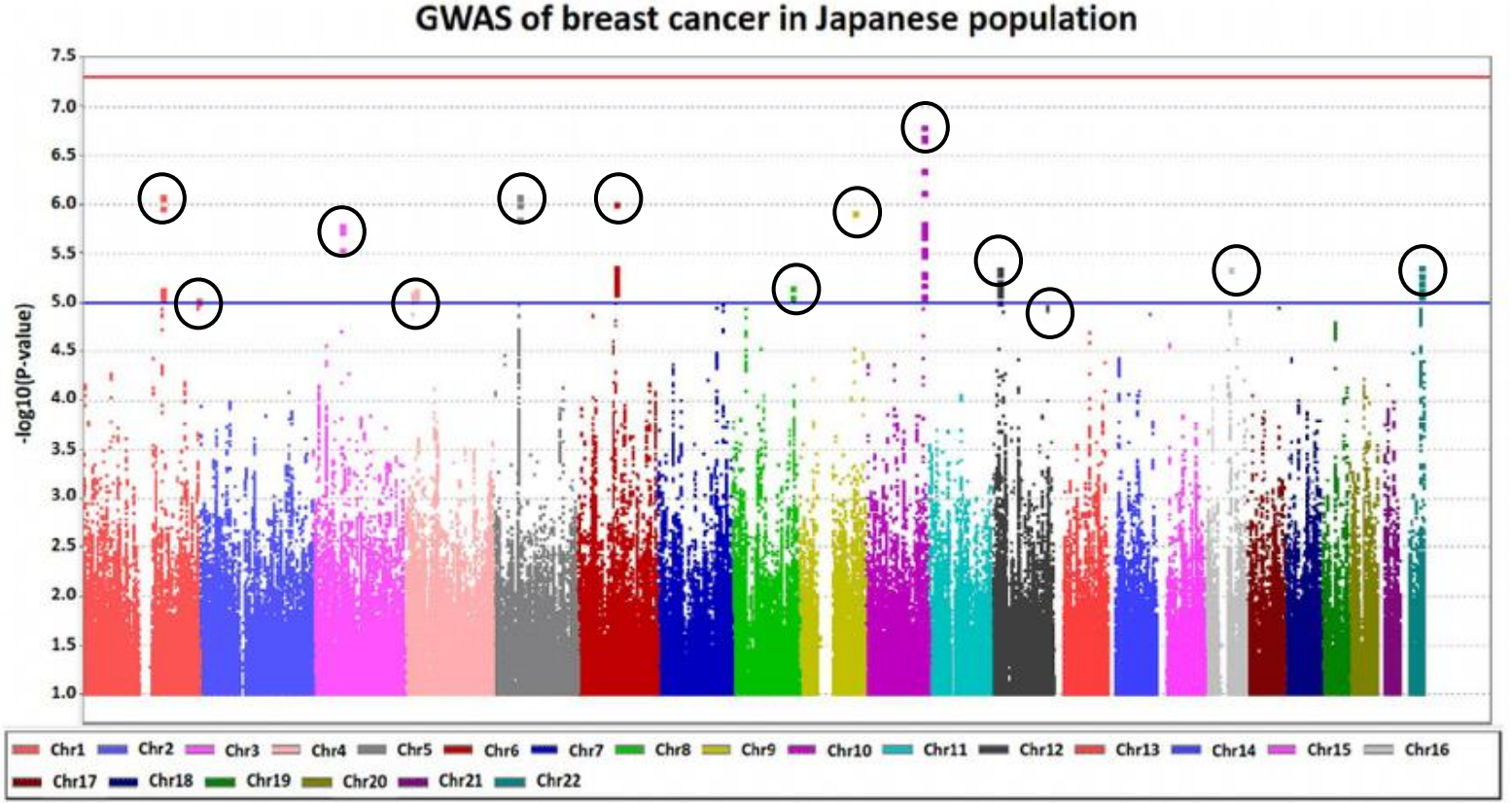
Session 12

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

What about the rest of disease?



Small contributions of many genetic locations

Each variant in itself is inherited, but the specific combinations can be different, making it look like a disease is not family-based.

Each increases risk a small amount: odds ratio 1.01-1.5

All together, these variants influence overall risk of an outcome: Polygenic Risk

Based on what variants an individual has: Polygenic Risk Score

(also referred to as genetic risk score, polygenic score, genome-wide score)

Many SNPs, small cumulative impacts

~80 loci explain ~20% of coronary artery disease heritability

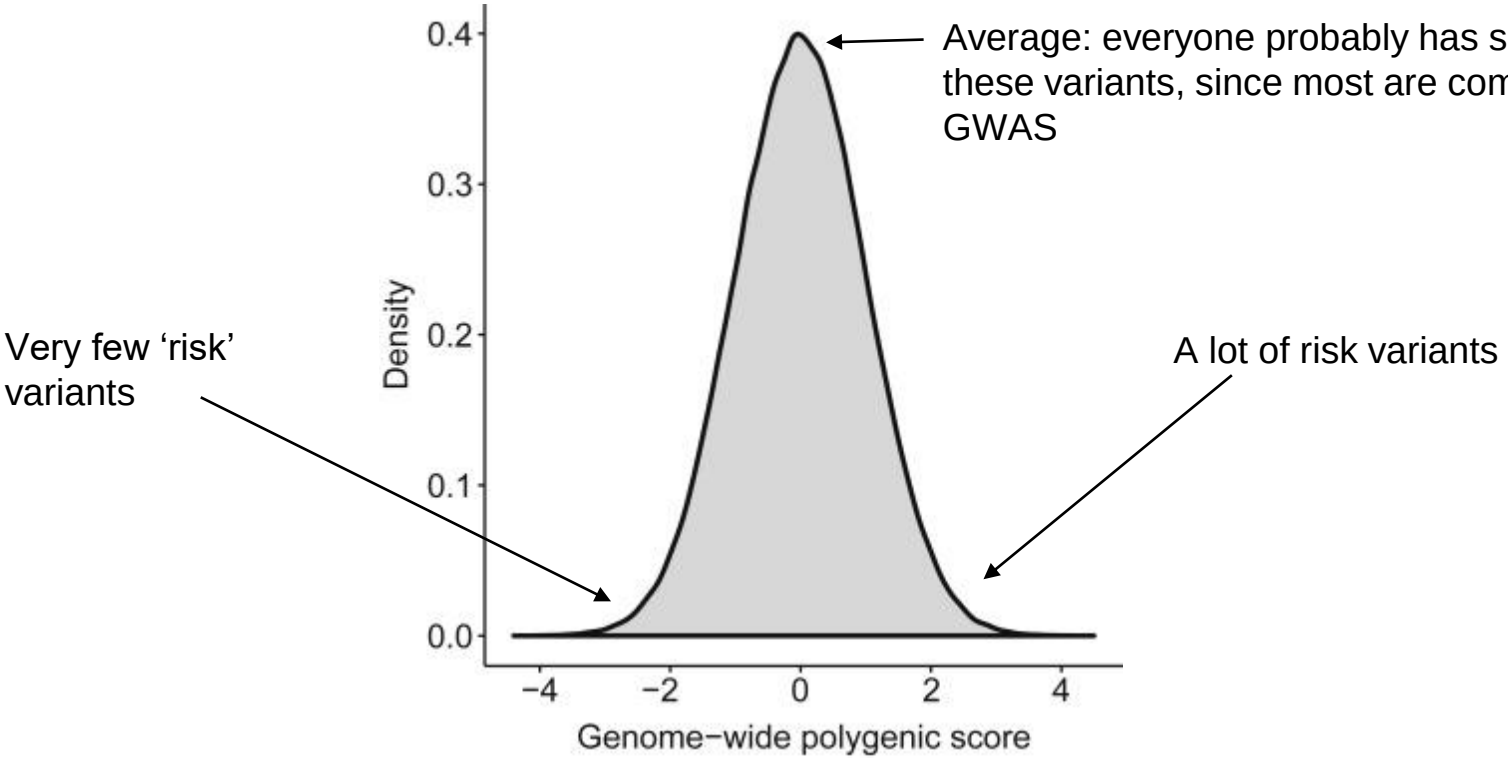
~100 loci explain ~20% of type 2 diabetes heritability

~150 loci explain ~20% of familial breast cancer

~100 loci explain ~33% of familial prostate cancer

~20 loci explain ~30% of Alzheimer disease heritability

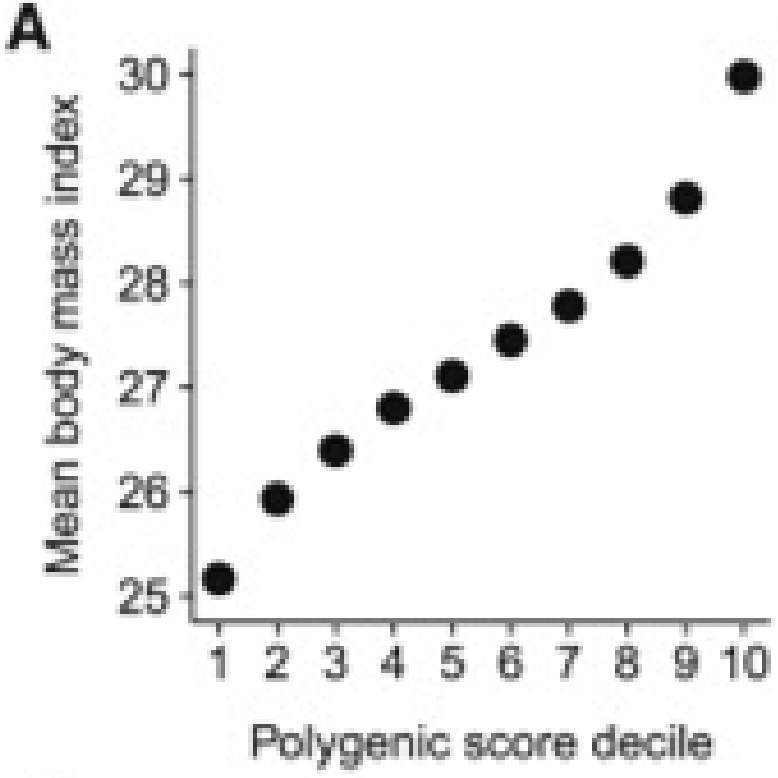
Population polygenic risk score distribution



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

Consider that polygenic score in risk of outcome

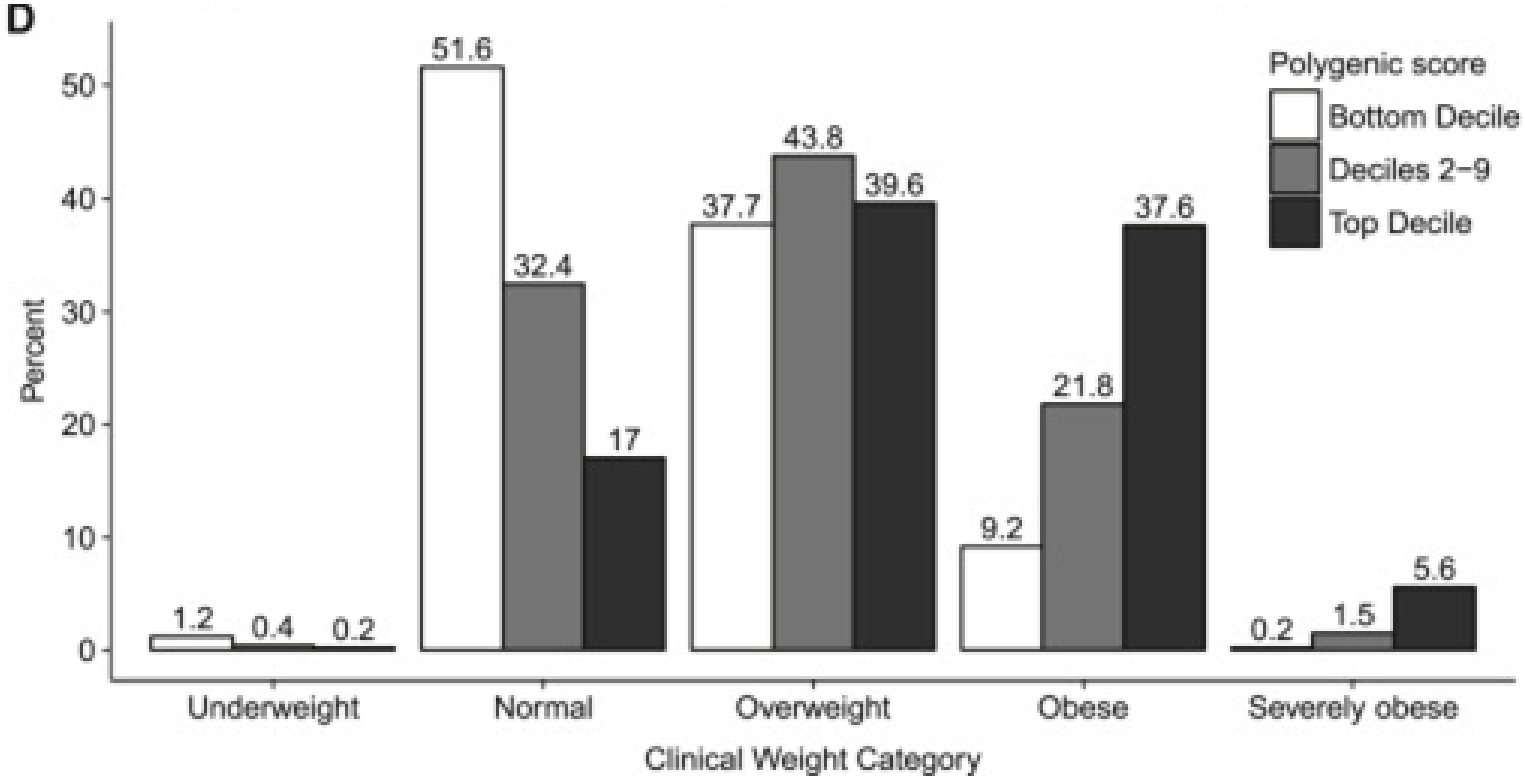
Polygenic risk score for BMI



Cohort divided into people with each genetic score level

Polygenic risk scores are still probabilistic

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



glm(outcome~SNP + Age+ Sex) <-each SNP is modeled alone

How do we calculate polygenic risk scores?

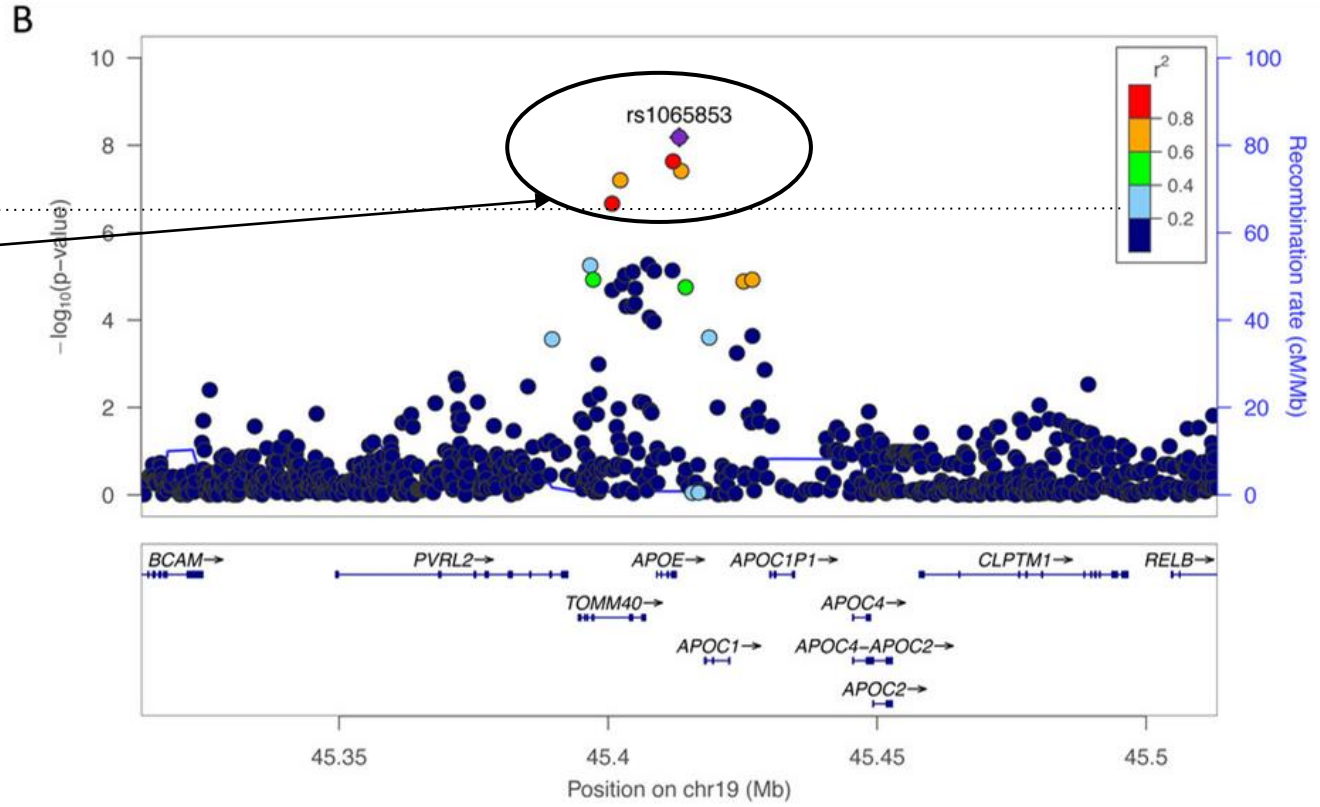
1) Start with our GWAS and the individual estimates for each SNP

	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

BMI study started with 2.1 million SNPs for 300,000 individuals

2) Account for linkage disequilibrium in SNPs

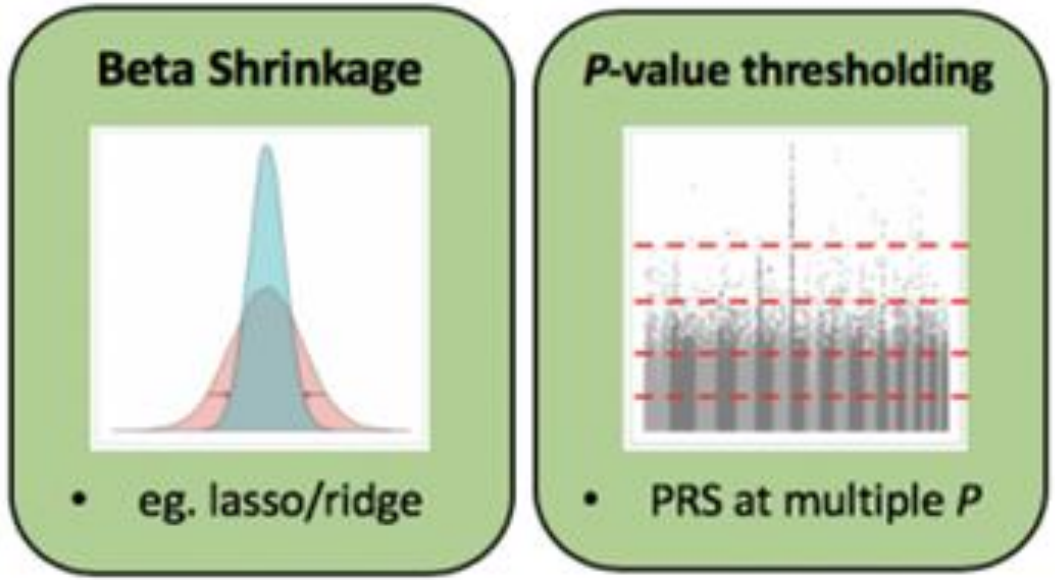
LD-clumping $r^2 > 0.50$



https://www.biorxiv.org/content/10.1101/416545v1.full

3) Determine which SNPs to include in the score

Various methods: penalized regression LASSO, p-value thresholding



Threshold set a little lower than Bonferroni GWAS: $1x\ 10^{-5}$

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	$6x10^{-7}$

SNPs in additive form

Loading value from GWAS output

Score Formula =

each loading value multiplied by number of alleles for that SNP, summed across each SNP

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)

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

Is our polygenic score associated with BMI?

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

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

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

Remember if the outcome is disease/no disease, we use the ‘glm’ odds ratio

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

Why is a polygenic score helpful?

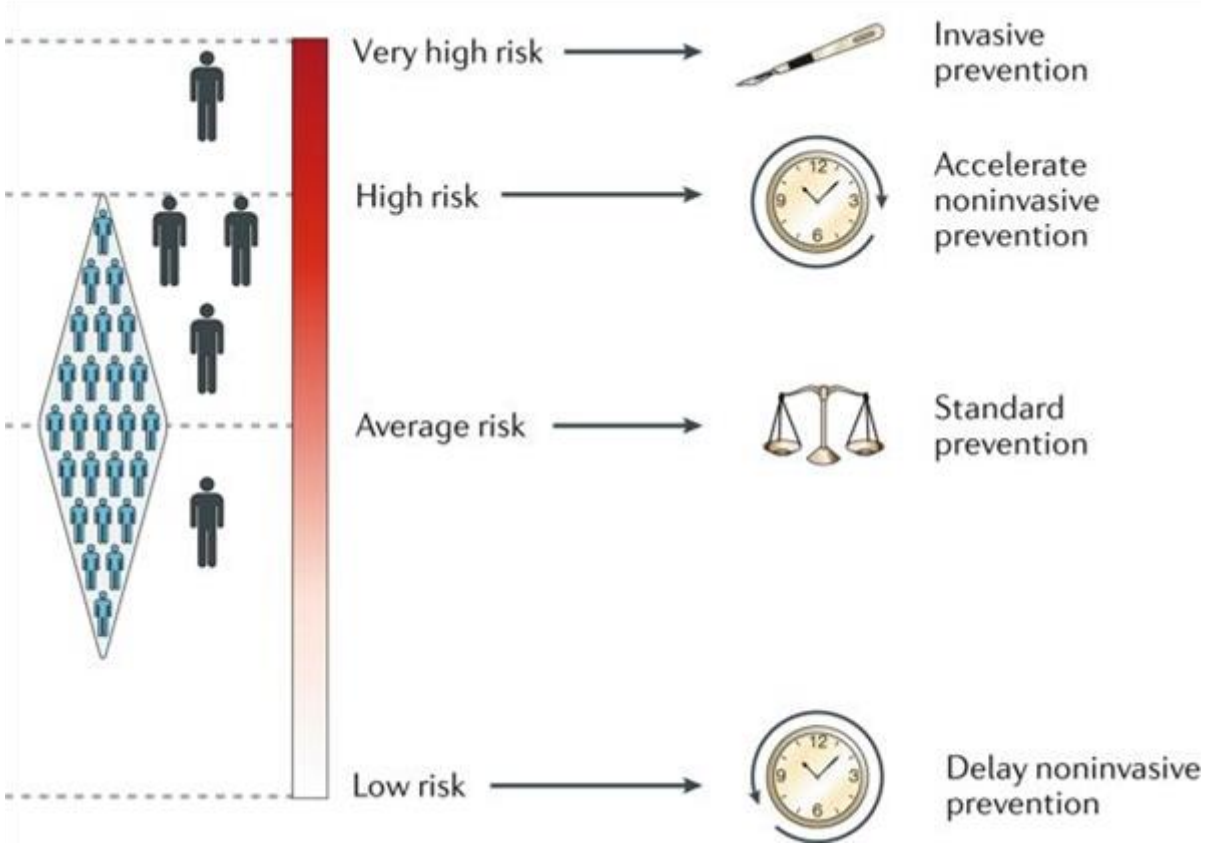
You are born with your genetic variants, so it is one of the earliest and more stable predictors possible. The score can include rare, yet high impact variants and common, low impact variants.

We can identify people early for screening and lifestyle modifications.

Utility still depends on strength of the association and the possible interventions (the more invasive an intervention, the bigger the association needs to be).

Ask: Does it allow useful stratification of patients into risk groups?

Polygenic scores as another data point



Racial/ethnic ancestry in polygenic scores

Why would a polygenic risk score developed in a European cohort be unreliable for a person who does not have recent European ancestors?

Zoom breakout

Freakonomics Radio:
impact and utility of polygenic risk score for lipids
23:00-23:40; 27:30-28:20
http://freakonomics.com/podcast/23andme/

Implementing population screening

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

Zoom breakout.