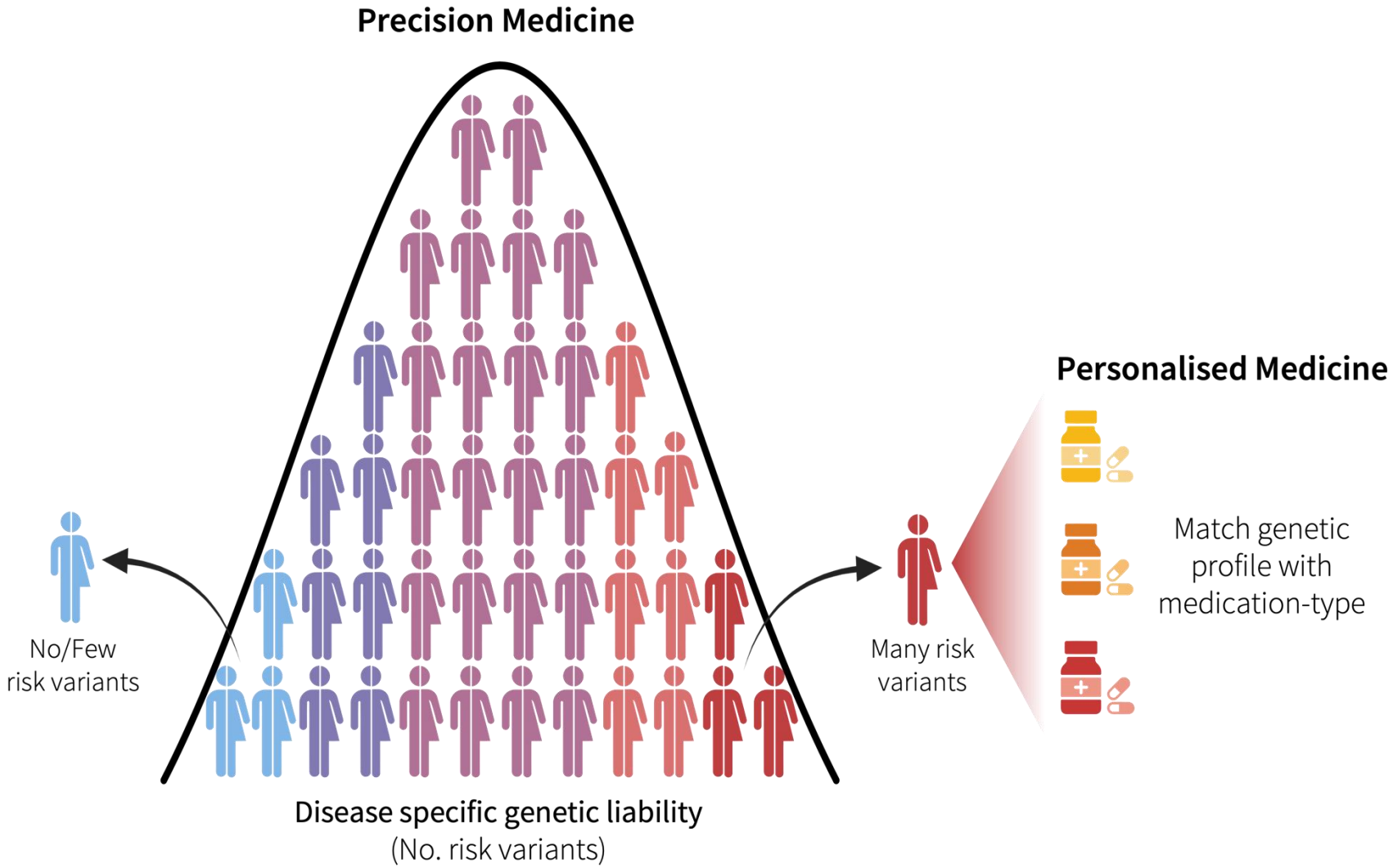
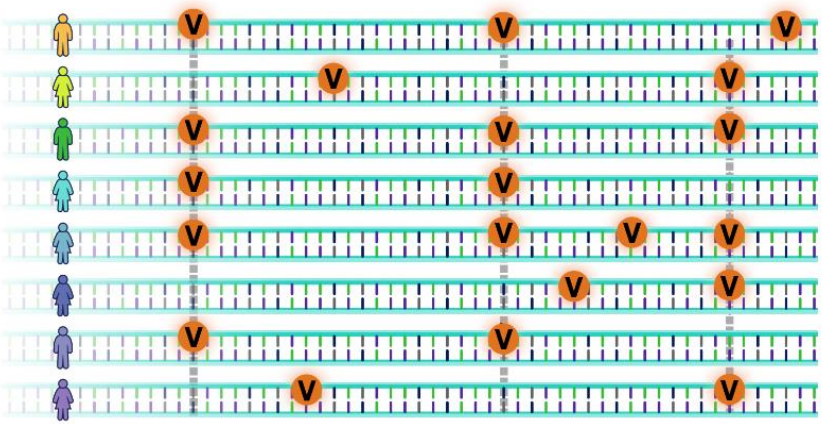
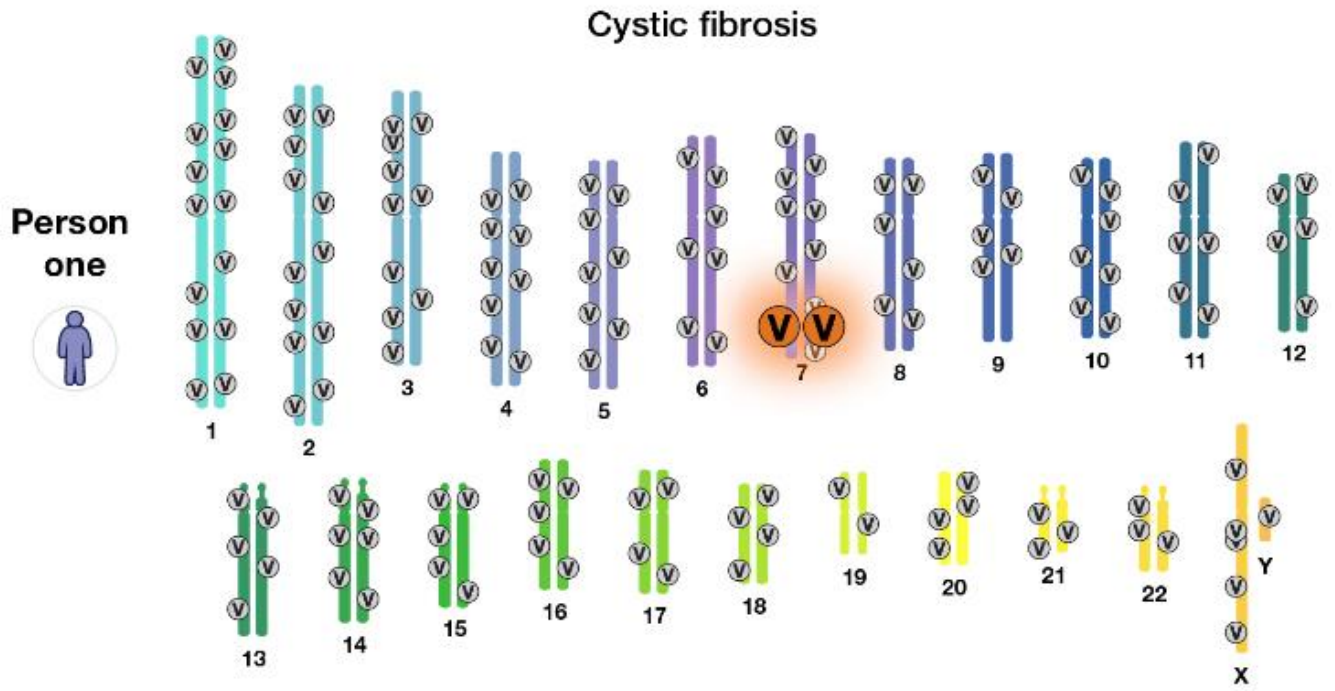
Polygenic Scores
RECAP



POLYGENIC SCORES



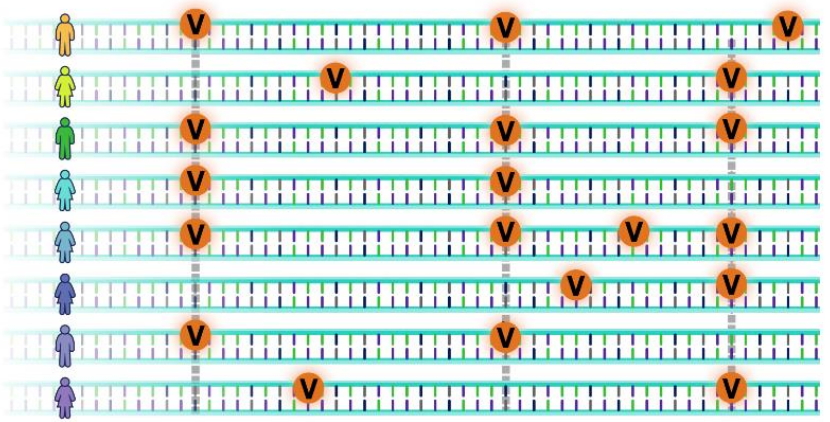


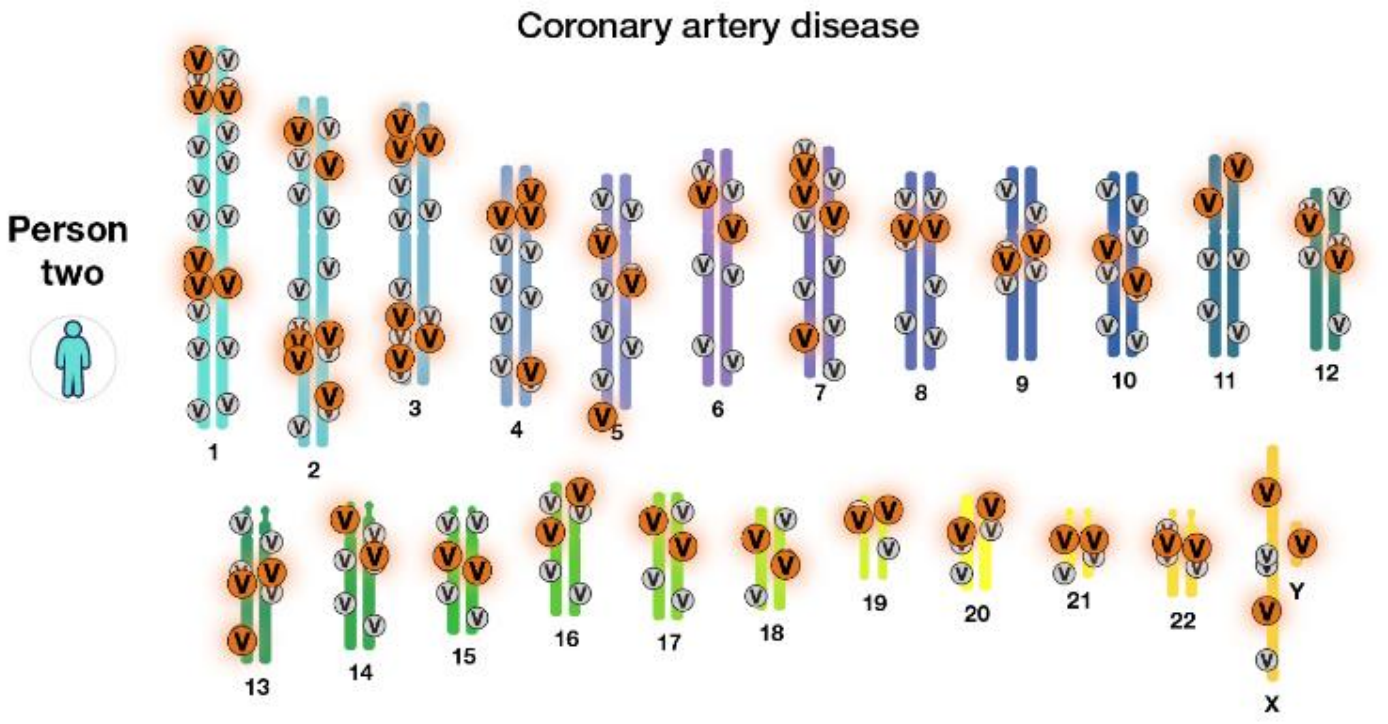
Each small "v" represents a genomic variant that is present in an individual's genome but are not associated with cystic fibrosis.

Each larger "V" represents a CFTR gene mutation.

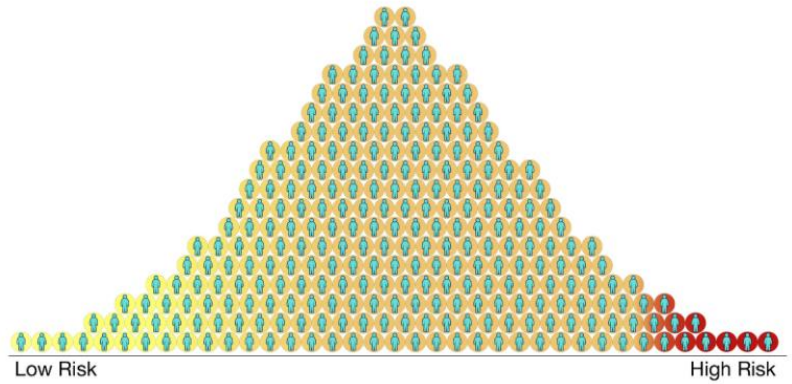
https://www.genome.gov/Health/Genomics-and-Medicine/Polygenic-risk-scores

POLYGENIC SCORES





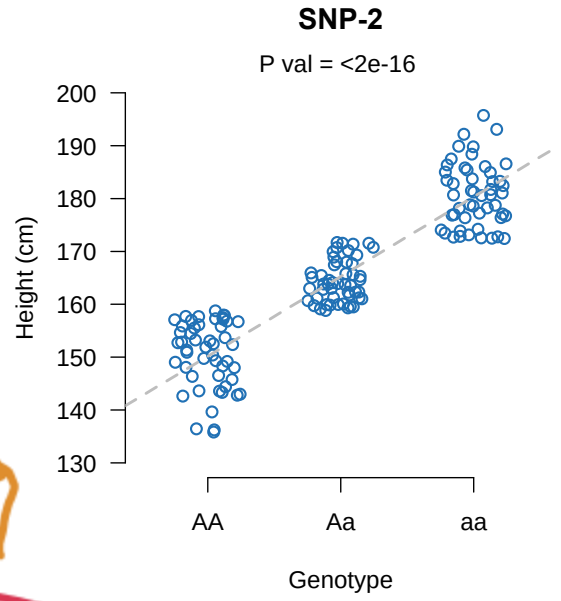
Each red "v" represents variants in an individual's genome that is associated with coronary artery disease. Each smaller gray "v" is a variant that is also present in the person's genome but is not implicated in disease.



A polygenic score is the sum of disease-specific risk genetic variants an individual has.

https://www.genome.gov/Health/Genomics-and-Medicine/Polygenic-risk-scores

WHAT IS A PGS



$$PGS = \sum X_i b_i$$

b is the slope (effect size) from reggresion/GWAS

The genotype of the individual for SNP i
(0, 1, 2 – counting the number of the alternative allele)

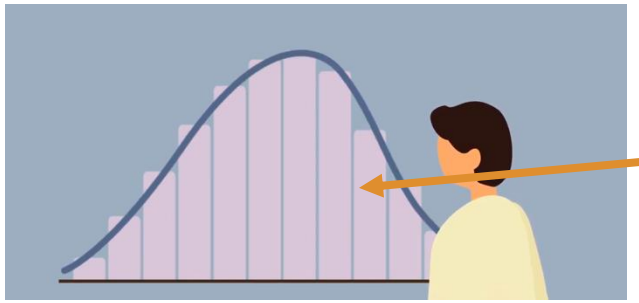
AA = 0
Aa = 1
aa = 2

i represents the SNP; thus, if you have 10 SNPs, *i* will take the values 1, 2,..,10 iteratively

$$PGS = \sum X_i b_i$$

HOW TO COMPUTE A (simple) PGS?

SNPs	Adams Genotypes	Ref allele	Alt allele	X	b	Xb
SNP-1	TC	T	C	1	0.04	0.04
SNP-2	GG	G	T	0	0.02	0.00
SNP-3	CC	A	C	2	0.05	0.10
SNP-4	TG	T	G	1	0.02	0.02
SNP-5	AA	A	G	0	0.06	0.00



PGS = 0.16

CLUMPING AND THRESHOLDING (C+T)

0: Set LD (=0.8) and *P* values (0.01)

SNP	b	p
1	0.21	0.005
2	0.22	0.0048
3	0.25	0.0003
4	0.1	0.04
5	0.05	0.15
6	0.02	0.49
7	0.03	0.87
8	0.12	0.003
9	0.14	0.0034
10	0.18	0.0004
11	0.21	0.00003
12	0.12	0.15
13	0.14	0.12
14	0.03	0.84
15	0.02	0.32

1: Sort by P-value

SNP	b	p
11	0.21	0.00003
3	0.25	0.0003
10	0.18	0.0004
8	0.12	0.003
9	0.14	0.0034
2	0.22	0.0048
1	0.21	0.005
4	0.1	0.04
13	0.14	0.12
5	0.05	0.15
12	0.12	0.15
15	0.02	0.32
6	0.02	0.49
14	0.03	0.84
7	0.03	0.87

2: Compute LD and select variants based of thresholds

SNP	b	p	r^2
11	0.21	0.00003	
3	0.25	0.0003	0.96
10	0.18	0.0004	0.93
8	0.12	0.003	0.88
9	0.14	0.0034	0.74
2	0.22	0.0048	0.4
1	0.21	0.005	0.03
4	0.1	0.04	0.04
13	0.14	0.12	0.05
5	0.05	0.15	0.03
12	0.12	0.15	0.04
15	0.02	0.32	0.01
6	0.02	0.49	0.01
14	0.03	0.84	0.01
7	0.03	0.87	0.01

1st variant in LD-pair

Have LD>r^2 – ignore those

CLUMPING AND THRESHOLDING (C+T)

0: Set LD (=0.8) and *P* values (0.01)

SNP	b	p
1	0.21	0.005
2	0.22	0.0048
3	0.25	0.0003
4	0.1	0.04
5	0.05	0.15
6	0.02	0.49
7	0.03	0.87
8	0.12	0.003
9	0.14	0.0034
10	0.18	0.0004
11	0.21	0.00003
12	0.12	0.15
13	0.14	0.12
14	0.03	0.84
15	0.02	0.32

1: Sort by P-value

SNP	b	p
11	0.21	0.00003
3	0.25	0.0003
10	0.18	0.0004
8	0.12	0.003
9	0.14	0.0034
2	0.22	0.0048
1	0.21	0.005
4	0.1	0.04
13	0.14	0.12
5	0.05	0.15
12	0.12	0.15
15	0.02	0.32
6	0.02	0.49
14	0.03	0.84
7	0.03	0.87

2: Compute LD and select variants based of thresholds

SNP	b	p	r^2
11	0.21	0.00003	
3	0.25	0.0003	
10	0.18	0.0004	
8	0.12	0.003	
9	0.14	0.0034	
2	0.22	0.0048	0.98
1	0.21	0.005	0.96
4	0.1	0.04	0.96
13	0.14	0.12	0.52
5	0.05	0.15	0.34
12	0.12	0.15	0.10
15	0.02	0.32	0.04
6	0.02	0.49	0.01
14	0.03	0.84	0.01
7	0.03	0.87	0.01

1st variant in LD-pair

Have LD>r^2 – ignore those

CLUMPING AND THRESHOLDING (C+T)

0: Set LD (=0.8) and *P* values (0.01)

SNP	b	p
1	0.21	0.005
2	0.22	0.0048
3	0.25	0.0003
4	0.1	0.04
5	0.05	0.15
6	0.02	0.49
7	0.03	0.87
8	0.12	0.003
9	0.14	0.0034
10	0.18	0.0004
11	0.21	0.00003
12	0.12	0.15
13	0.14	0.12
14	0.03	0.84
15	0.02	0.32

1: Sort by P-value

SNP	b	p
11	0.21	0.00003
3	0.25	0.0003
10	0.18	0.0004
8	0.12	0.003
9	0.14	0.0034
2	0.22	0.0048
1	0.21	0.005
4	0.1	0.04
13	0.14	0.12
5	0.05	0.15
12	0.12	0.15
15	0.02	0.32
6	0.02	0.49
14	0.03	0.84
7	0.03	0.87

2: Compute LD and select variants based of thresholds

SNP	b	p	r^2	
11	0.21	0.00003		
3	0.25	0.0003		
10	0.18	0.0004		
8	0.12	0.003		
9	0.14	0.0034		
2	0.22	0.0048		
1	0.21	0.005		
4	0.1	0.04		
13	0.14	0.12		1st variant in LD-pair
5	0.05	0.15	0.86	
12	0.12	0.15	0.82	
15	0.02	0.32	0.81	
6	0.02	0.49	0.85	Have LD>r^2 – ignore those
14	0.03	0.84	0.85	
7	0.03	0.87	0.81	

CLUMPING AND THRESHOLDING (C+T)

0: Set LD (=0.8) and *P* values (0.01)

SNP	b	p
1	0.21	0.005
2	0.22	0.0048
3	0.25	0.0003
4	0.1	0.04
5	0.05	0.15
6	0.02	0.49
7	0.03	0.87
8	0.12	0.003
9	0.14	0.0034
10	0.18	0.0004
11	0.21	0.00003
12	0.12	0.15
13	0.14	0.12
14	0.03	0.84
15	0.02	0.32

1: Sort by P-value

SNP	b	p
11	0.21	0.00003
3	0.25	0.0003
10	0.18	0.0004
8	0.12	0.003
9	0.14	0.0034
2	0.22	0.0048
1	0.21	0.005
4	0.1	0.04
13	0.14	0.12
5	0.05	0.15
12	0.12	0.15
15	0.02	0.32
6	0.02	0.49
14	0.03	0.84
7	0.03	0.87

2: Compute LD and select variants based on LD

SNP	b	p	r^2
11	0.21	0.00003	
3	0.25	0.0003	
10	0.18	0.0004	
8	0.12	0.003	
9	0.14	0.0034	
2	0.22	0.0048	
1	0.21	0.005	
4	0.1	0.04	
13	0.14	0.12	
5	0.05	0.15	
12	0.12	0.15	
15	0.02	0.32	
6	0.02	0.49	
14	0.03	0.84	
7	0.03	0.87	

3: Compute PGS based on effect sizes (b) and *P*-values

$$PGS = \sum X_i b_i$$

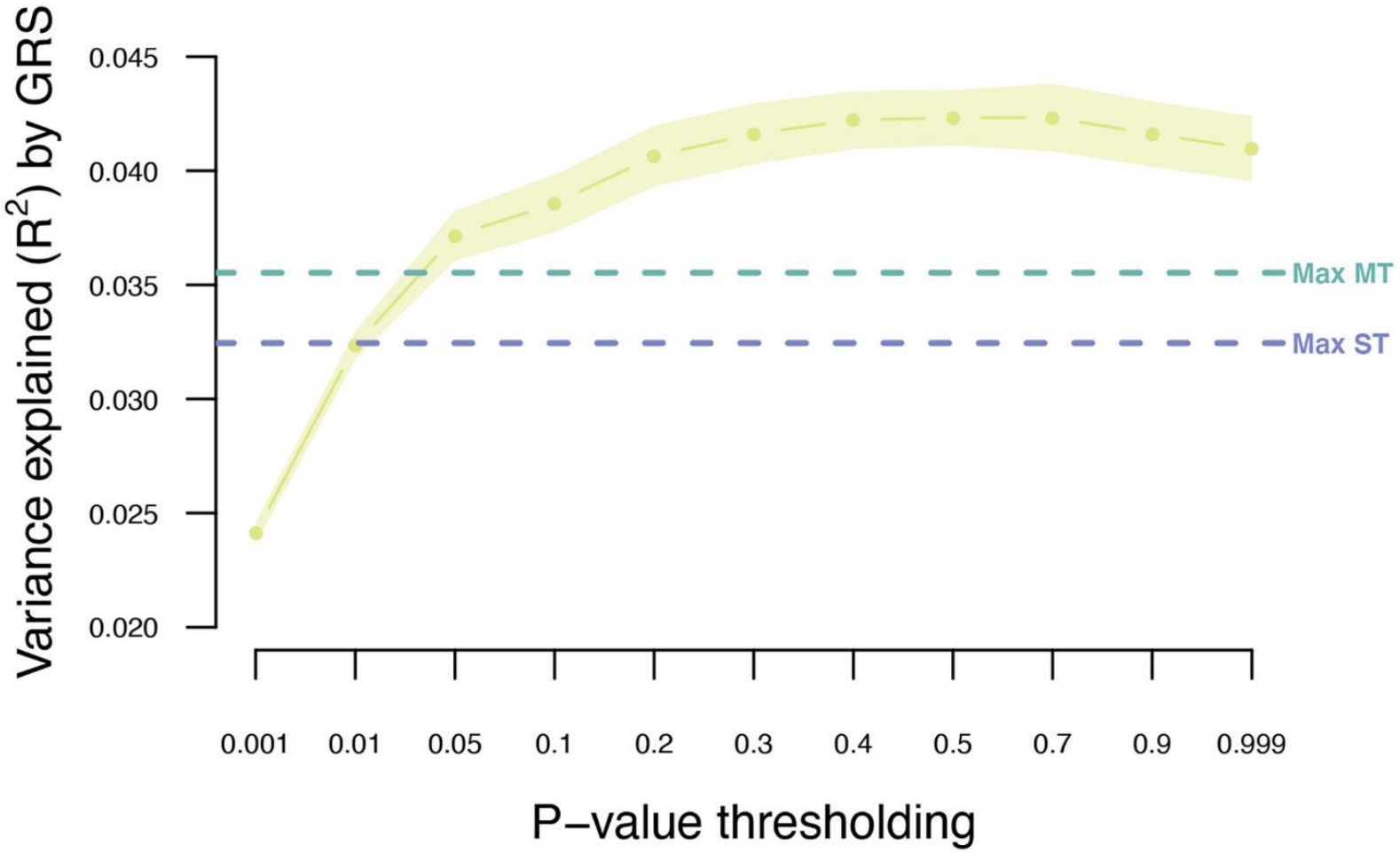
$$= \boldsymbol{X}_{11} \times 0.21 + \boldsymbol{X}_{9} \times 0.14$$

CLUMPING AND THRESHOLDING (C+T)

Repeat for other *P*-value cutoffs (and LD values)

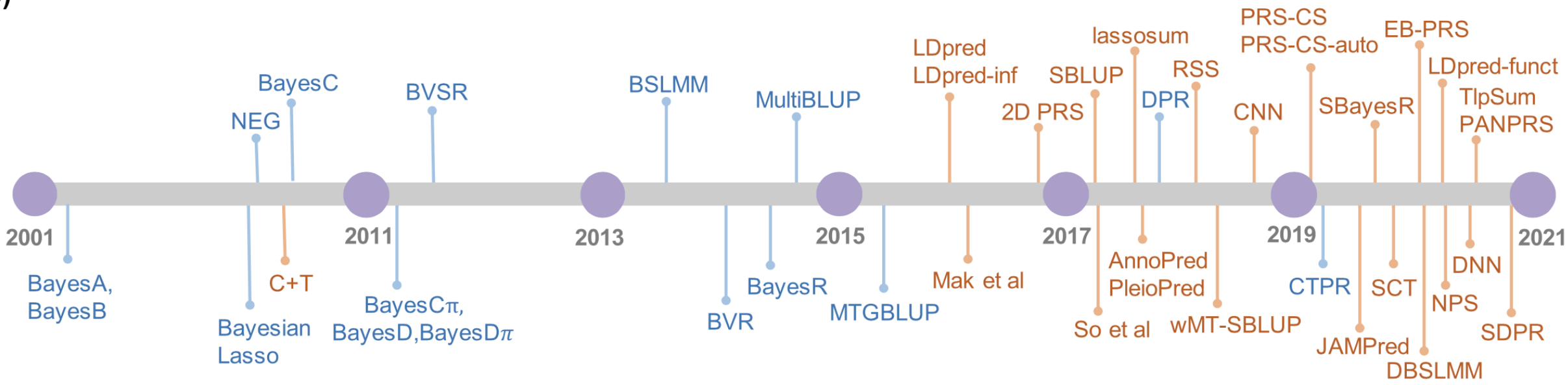
How does the PGS associate with the disease

$$y_{disease} = PGS + \varepsilon$$



MORE SOPHISTICATED METHODS EXISTS

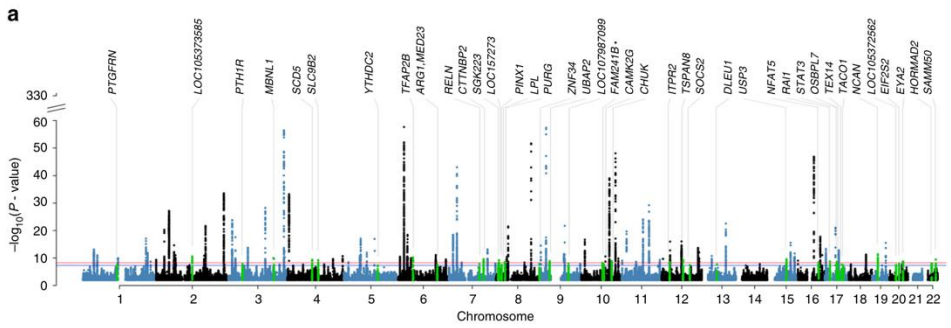
(B)



Ma, Y., & Zhou, X. (2021). *Trends in Genetics*, *37*(11), 995–1011.

WHAT DO YOU NEED?

1. A large well-powered GWAS for your trait of interest



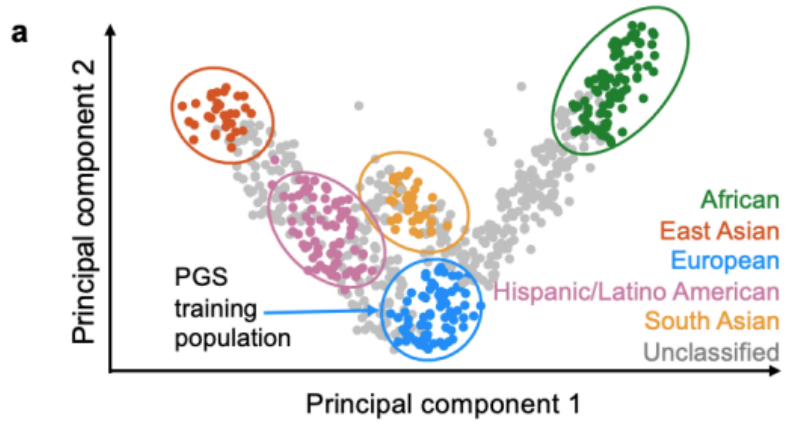
2. An independent cohort that has been genotyped

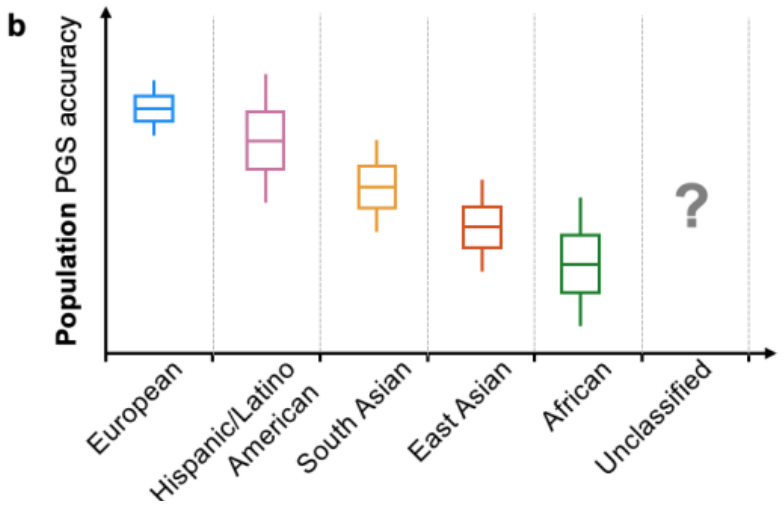
(3. That some individuals in the cohort has the phenotype)

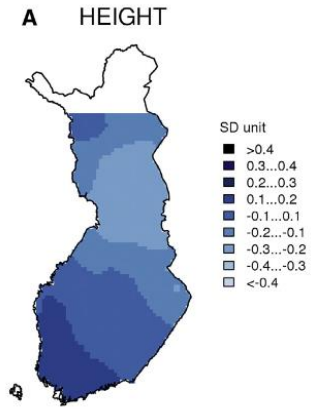
IMPORTANT CONSIDERATIONS

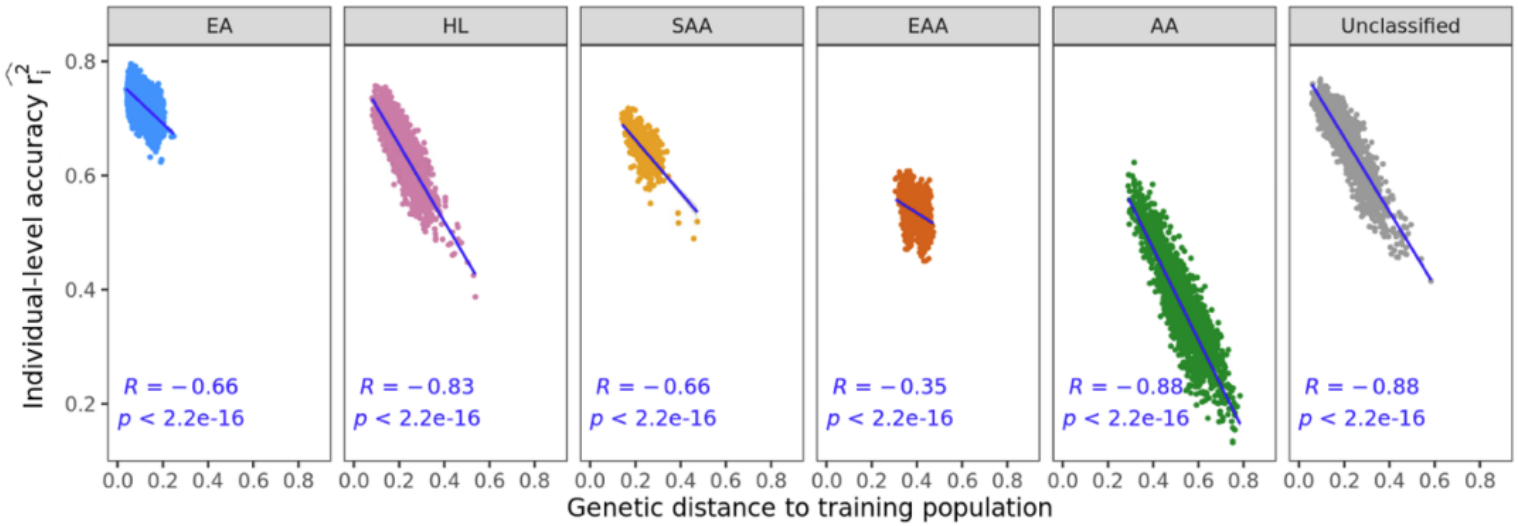
1. **You are ”*born*” with your polygenic score**
   - Everytime a new and better
     - GWAS is released
     - Scoring method is developed
   - we can recalculate a persons score
2. **The accuracy of the polygenic score depents on ancestry**

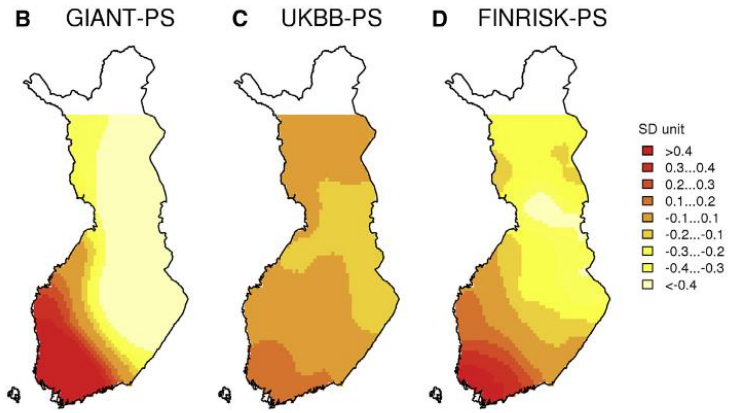
ACCURACY OF PGS AND ANCESTRY











IMPORTANT CONSIDERATIONS

1. **You are ”*born*” with your polygenic score**
   - Everytime a new and better
     - GWAS is released
     - Scoring method is developed
   - we can recalculate a persons score
2. **The accuracy of the polygenic score depents on ancestry**
3. **Rare genetic variants in concert with the polygenic burden may modulate the disease risk**

RARE VARIANTS AND PGS – MODULATION OF RISK

Two frameshift mutations strongly associated with breast cancer in Finland

Table 2 Risk for breast cancer events in the population in carriers of the *PALB2* and *CHEK2* frameshift mutations, and in the top decile of the polygenic risk score (PRS).

	PALB2	*CHEK2*	PRS > 90%
Number of individuals	336	1648	12,298
Number of cases	84	214	1821
Lifetime risk of breast cancer, % (95% CI)	56.1 (50.8–61.4)	31.7 (29.5–33.9)	32.5 (31.6–33.4)
Mean age at disease onset in cases (SD)	53.1 (10.4)	56.5 (12.0)	57.8 (11.3)

Lifetime risk was estimated by age 80. The variants were rs180177102 (c.[illegible]592delT) for *PALB2* and rs555607708 (c.110[illegible]0delC) for *CHEK2*. The *PALB2* analysis was done in 109,371 women, and the *CHEK2* and PRS analyses in 122,978 women.
CI confidence interval, *SD* standard deviation.

HR=4.99 HR=2.19

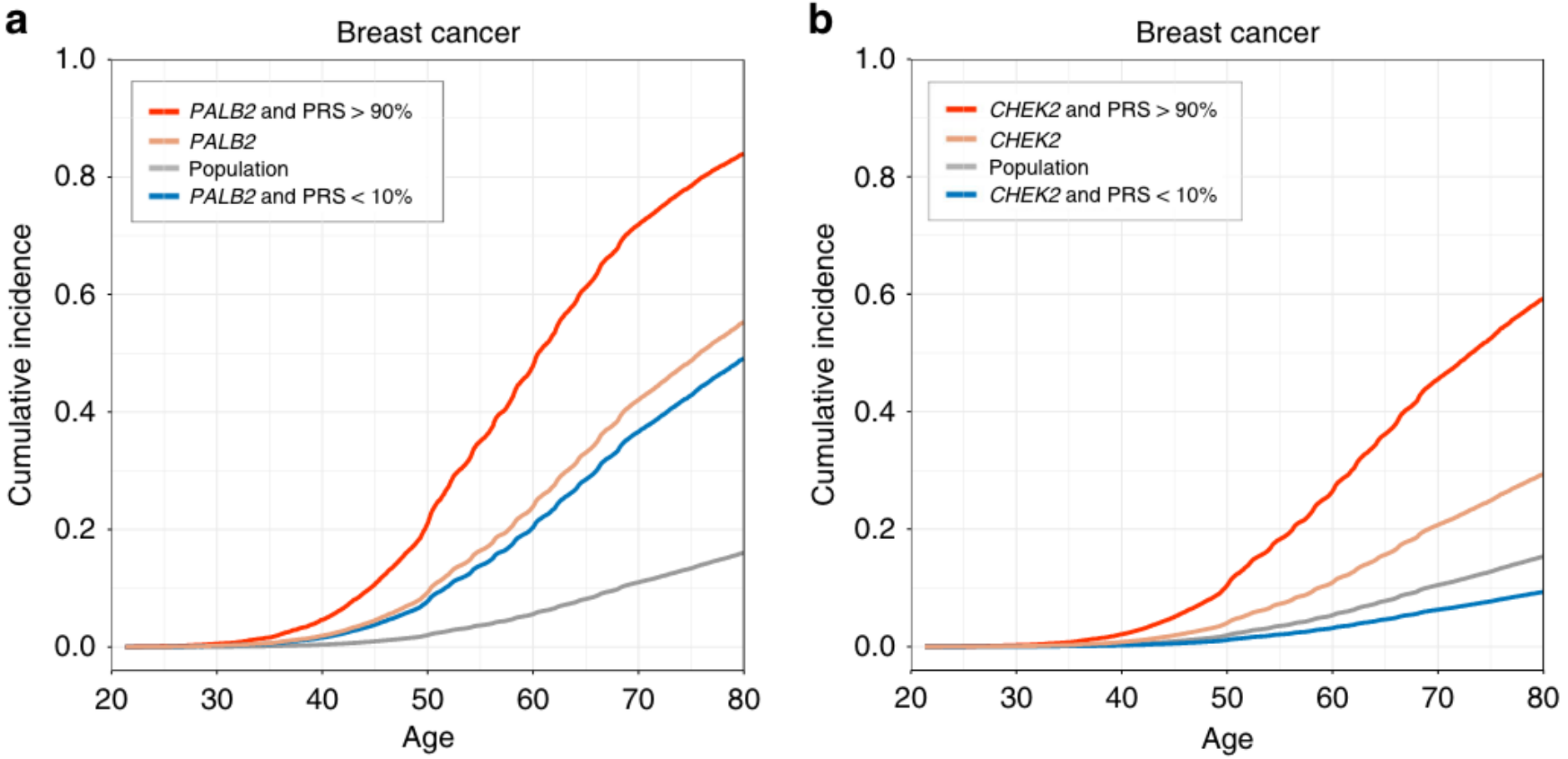
HR = Hazard ratio

A hazard ratio tells us whether a subject in the treatment group who is unaffected at any given time has a greater, equal, or lower probability (i.e., hazard rate) of experiencing the event during the next unit of time than an unaffected subject in the control group.

Mars, N., et al., (2020). *Nature Communications*, *11*(1), 6383. https://doi.org/10.1038/s41467-020-19966-5

RARE VARIANTS AND PGS – MODULATION OF RISK

→ high breast cancer PGS comes with a comparable risk profile to frameshift mutations in breast cancer susceptibility genes *PALB2* and *CHEK2*, and that the PGS strongly modifies breast cancer risk in the mutation carriers



Mars, N., et al., (2020). *Nature Communications, 11*(1), 6383. https://doi.org/10.1038/s41467-020-19966-5