SCHOOL OF PUBLIC HEALTH
EPIDEMIOLOGY
UNIVERSITY of WASHINGTON

Session 9: Risk Prediction

Complex traits are often influenced by many variants

- The genetic architecture of polygenic complex traits
  - A large number genetic variants contribute to disease risk
  - Each risk variant typically has a small effect, ranging from OR 1.01-1.2

- Collectively, these variants can lead to better risk prediction than any given variant on its own

What are Polygenic Risk Scores?

Definition[1]: Polygenic scores **quantify genetic predisposition** to a *heritable* trait, calculated as a sum of genetic alleles weighted by corresponding variant-specific effect sizes.

- **Polygenic Risk Score (PRS)/Genetic Risk Score (GRS)**: Estimates genetic contribution to disease risk
- **Polygenic Score (PGS)**: Estimates genetic contribution to any phenotype (e.g., disease, biomarkers, height)
- **Integrated Risk Model**: Estimates disease risk by combining PRS/PGS with other established clinical risk factors

Potential Clinical Utility[2]:

1. **PRS-informed disease screening**: Decision to initiate and the interpretation of screening (*disease risk prediction/stratification*)
2. **PRS-informed therapeutic intervention**: Selection of interventions to treat or prevent disease (*disease subtyping, prediction of prognostic outcomes/response to therapy*)
3. **PRS-informed life planning**: Personal utility of PRS, even in the absence of clinical intervention (e.g., Alzheimer's disease PRS: may inform financial, legal, and care planning)

[1] Wand, Lambert et al., *Nature* 2021, Improving reporting standards for polygenic scores in risk prediction studies.
[2] Torkamani et al., *Nature Reviews Genetics* 2018, The personal and clinical utility of polygenic risk scores.

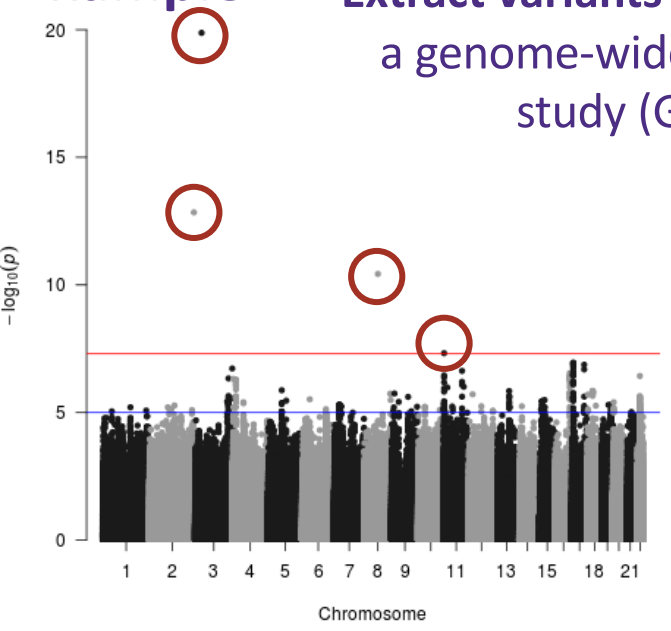
Calculating Polygenic Risk Scores

PRS is calculated as a weighted sum of genetic *risk* alleles, weighted by corresponding variant-specific effect sizes → Higher PRS = Higher genetic risk

$$PRS_i = \sum_{m=1}^{M} \beta_m G_{im}$$

β = weight of variant m
G = # risk alleles variant m in individual i
M = total # variants

Example



Extract variants & effects from a genome-wide association study (GWAS)

	Risk Allele	Non-Risk Allele	β (risk allele)	P
SNP1	G	C	0.50	5.3E-20
SNP2	A	C	0.13	2.2E-13
SNP3	T	A	0.28	5.6E-10
SNP4	T	C	0.10	4.0E-08

Use identified variants and weights to **calculate PRS** in an *external* sample

$PRS_{person1}$ = 0.50*1 + 0.13*1 + 0.28*0 + 0.10*2 = 0.83
$PRS_{person2}$ = 0.50*2 + 0.13*0 + 0.28*0 + 0.10*1 = 1.10
$PRS_{person3}$ = …

** This is an example of a genome-wide significant PRS, but there are many other ways to develop a PRS*

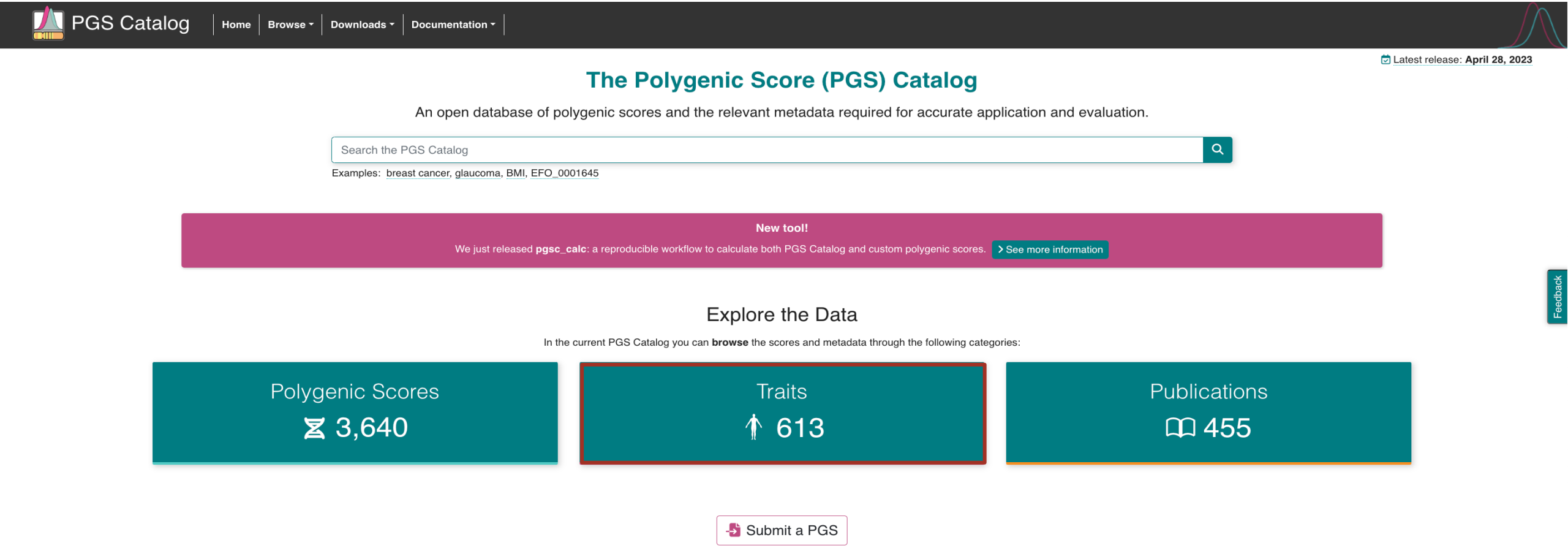
How to identify weights and variants?

Selecting Polygenic Risk Scores

Using an existing PRS based on large GWAS is a simple and powerful approach

pgscatalog.org

Lambert et al., *Nature Genetics* 2021



Selecting Polygenic Risk Scores

Using an existing PRS based on large GWAS is a simple and powerful approach

pgscatalog.org

Lambert et al., *Nature Genetics* 2021

PGS Catalog | Home | Browse | Downloads | Documentation

Search...

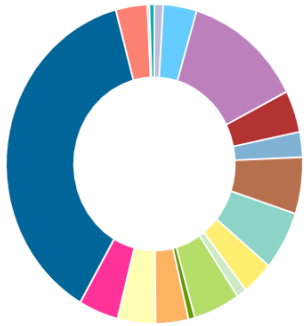
breast cancer, glaucoma, BMI, EFO_0001645

PGS Catalog / Browse / Traits

Traits

Browse PGS by Trait Category

Reset view



Biological process	37 PGS
Body measurement	164 PGS
Cancer	598 PGS
Cardiovascular disease	191 PGS
Cardiovascular measurement	112 PGS
Digestive system disorder	261 PGS
Hematological measurement	271 PGS
Immune system disorder	155 PGS
Inflammatory measurement	43 PGS
Lipid or lipoprotein measurement	230 PGS
Liver enzyme measurement	25 PGS
Metabolic disorder	160 PGS
Neurological disorder	186 PGS
Other disease	195 PGS
Other measurement	1749 PGS
Other trait	153 PGS
Response to drug	3 PGS
Sex-specific PGS	18 PGS

melanoma	40 PGS
multiple myeloma	6 PGS
nasopharyngeal neoplasm	1 PGS
neuroblastoma	1 PGS
non-Hodgkins lymphoma	4 PGS
non-melanoma skin carcinoma	1 PGS
ocular cancer	1 PGS
oral cavity cancer	4 PGS
ovarian carcinoma	9 PGS
ovarian neoplasm	21 PGS
ovarian serous carcinoma	2 PGS
pancreatic carcinoma	8 PGS
pancreatic ductal adenocarcinoma	1 PGS
pharynx cancer	2 PGS
polycythemia vera	1 PGS
prostate adenocarcinoma	1 PGS
prostate cancer	2 PGS
prostate carcinoma	62 PGS

Feedback

Selecting Polygenic Risk Scores

Using an existing PRS based on large GWAS is a simple and powerful approach

pgscatalog.org

Lambert et al., *Nature Genetics* 2021

PGS Catalog | Home | Browse | Downloads | Documentation

Search...

breast cancer, glaucoma, BMI, EFO_0001645

PGS Catalog / Traits / EFO_0001663

Trait: **prostate carcinoma**

Experimental Factor Ontology (EFO) Information	
Identifier	EFO_0001663
Description	A carcinoma that arises from epithelial cells of the prostate gland.
Trait category	Cancer
Synonyms	5 synonyms
Mapped terms	7 mapped terms

Associated Polygenic Score(s)

Filter PGS by Participant Ancestry

Individuals included in: All Stages combined [G, D, E]

G - Source of Variant Associations (GWAS)
D - Score Development/Training
E - PGS Evaluation

List of ancestries includes: --

Display options:
Show European ancestry data
Show only Multi-ancestry data

Ancestry legend

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

Search

Polygenic Score ID & Name	PGS Publication ID (PGP)	Reported Trait	Mapped Trait(s) (Ontology)	Number of Variants	Ancestry distribution	Scoring File (FTP Link)
	nat					
PGS000030 (PrCa)	PGP000019 » Schumacher FR *et al.* Nat Genet (2018)	Prostate cancer	prostate carcinoma	147	G - E	
PGS000084 (CC_Prostate)	PGP000050 » Graff RE *et al.* Nat Commun (2021)	Prostate cancer	prostate carcinoma	161	G - E	
PGS000333 (PRS_PC)	PGP000100 » Mars N *et al.* Nat Med (2020)	Prostate cancer	prostate carcinoma	6,606,785	G D E	
PGS000662 (GRS.PCa.269)	PGP000122 » Conti DV *et al.* Nat Genet (2021)	Prostate Cancer	prostate carcinoma	269	G - E	

Feedback

Selecting Polygenic Risk Scores

Using an existing PRS based on large GWAS is a simple and powerful approach

pgscatalog.org

Lambert et al., *Nature Genetics* 2021

PGS Catalog | Home | Browse | Downloads | Documentation

Search...

breast cancer, glaucoma, BMI, EFO_0001645

PGS Developed By This Publication

Search

Polygenic Score ID & Name	PGS Publication ID (PGP)	Reported Trait	Mapped Trait(s) (Ontology)	Number of Variants	Ancestry distribution	Scoring File (FTP Link)
PGS000662 (GRS.PCa.269)	PGP000122 » Conti DV *et al.* Nat Genet (2021)	Prostate Cancer	prostate carcinoma	269	G - E	

Showing 1 to 1 of 1 rows

Performance Metrics

Disclaimer: The performance metrics are displayed as reported by the source studies. It is important to note that metrics are not necessarily comparable with each other. For example, metrics depend on the sample characteristics (described by the PGS Catalog Sample Set [PSS] ID), phenotyping, and statistical modelling. Please refer to the source publication for additional guidance on performance.

Search

PGS Performance Metric ID (PPM)	Evaluated Score	PGS Sample Set ID (PSS)	Performance Source	Trait	PGS Effect Sizes (per SD change)	Classification Metrics	Other Metrics	Covariates Included in the Model	PGS Performance: Other Relevant Information
PPM001365	PGS000662 (GRS.PCa.269)	PSS000596 European Ancestry 199,969 individuals	PGP000122 » Conti DV *et al.* Nat Genet (2021)	Reported Trait: Prostate Cancer	–	AUROC: 0.833	Odds ratio (OR, top 10% versus 40-60% GRS): 4.17 Overall Net Reclassification Index (NRI [%]): 59.4	Age, 10 PCs	–
PPM001366	PGS000662 (GRS.PCa.269)	PSS000595 African Ancestry 2,633 individuals	PGP000122 » Conti DV *et al.* Nat Genet (2021)	Reported Trait: Prostate Cancer	–	AUROC: 0.679	Odds ratio (OR, top 10% versus 40-60% GRS): 3.53	Age, 10 PCs, study	–

Showing 1 to 2 of 2 rows

Feedback

Selecting Polygenic Risk Scores

Using an existing PRS based on large GWAS is a simple and powerful approach

Conti, Darst et al. *Nature Genetics* 2021

	A	B	C	D	E	F	G	H	I	J
1	###PGS CATALOG SCORING FILE - see https://www.pgscatalog.org/downloads/#dl_ftp_scoring for additional information									
2	#format_version=2.0									
3	##POLYGENIC SCORE (PGS) INFORMATION									
4	#pgs_id=PGS000662									
5	#pgs_name=GRS.PCa.269									
6	#trait_reported=Prostate Cancer									
7	#trait_mapped=prostate carcinoma									
8	#trait_efo=EFO_0001663									
9	#weight_type=beta									
10	#genome_build=GRCh37									
11	#variants_number=269									
12	##SOURCE INFORMATION									
13	#pgp_id=PGP000122									
14	#citation=Conti DV, Darst BF et al. Nat Genet (2020). doi:10.1038/s41588-020-00748-0									
15	rsID	chr_name	chr_position	effect_allele	other_allele	effect_weight	allelefrequency_effect_European	allelefrequency_effect_African	allelefrequency_effect_Asian	allelefrequency_effect_Hispanic
16	rs7542260	1	5743196	T	C	0.102298257	0.067	0.439	0.113	0.157
17	rs2847344	1	10564675	A	G	0.042411273	0.692	0.703	0.69	0.54
18	rs10803412	1	16376831	C	T	0.055506528	0.176	0.594	0.021	0.126
19	rs544780844	1	46251655	T	C	0.07282201	0.188	0.098	0.015	0.1
20	rs56391074	1	88210715	AT	A	0.048255598	0.37	0.722	0.751	0.5
21	rs1811698	1	150772613	C	T	0.080240037	0.895	0.448	0.81	0.887
22	rs607518	1	150954671	A	G	0.067047369	0.209	0.068	0.042	0.116
23	rs10127983	1	153923276	T	C	0.066274137	0.312	0.223	0.29	0.399

Convert genomic coordinates between builds:

UCSF LiftOver https://genome.ucsc.edu/cgi-bin/hgLiftOver

NCI Remap: https://www.ncbi.nlm.nih.gov/genome/tools/remap

Selecting Polygenic Risk Scores

How to select the optimal PRS for your study?

- Matching your trait
  - Closely matching the trait under investigation in the testing data (i.e., the study the PRS is being applied to) to the one used in the training data (i.e., the GWAS used to develop the PRS) will improve the accuracy of the PRS
- Sample Size
  - The larger the training data, the more accurate the PRS
- Population
  - Matching the training population to the ancestral background of the testing population could optimize results
  - ***However***, large training sample sizes are important for PRS accuracy

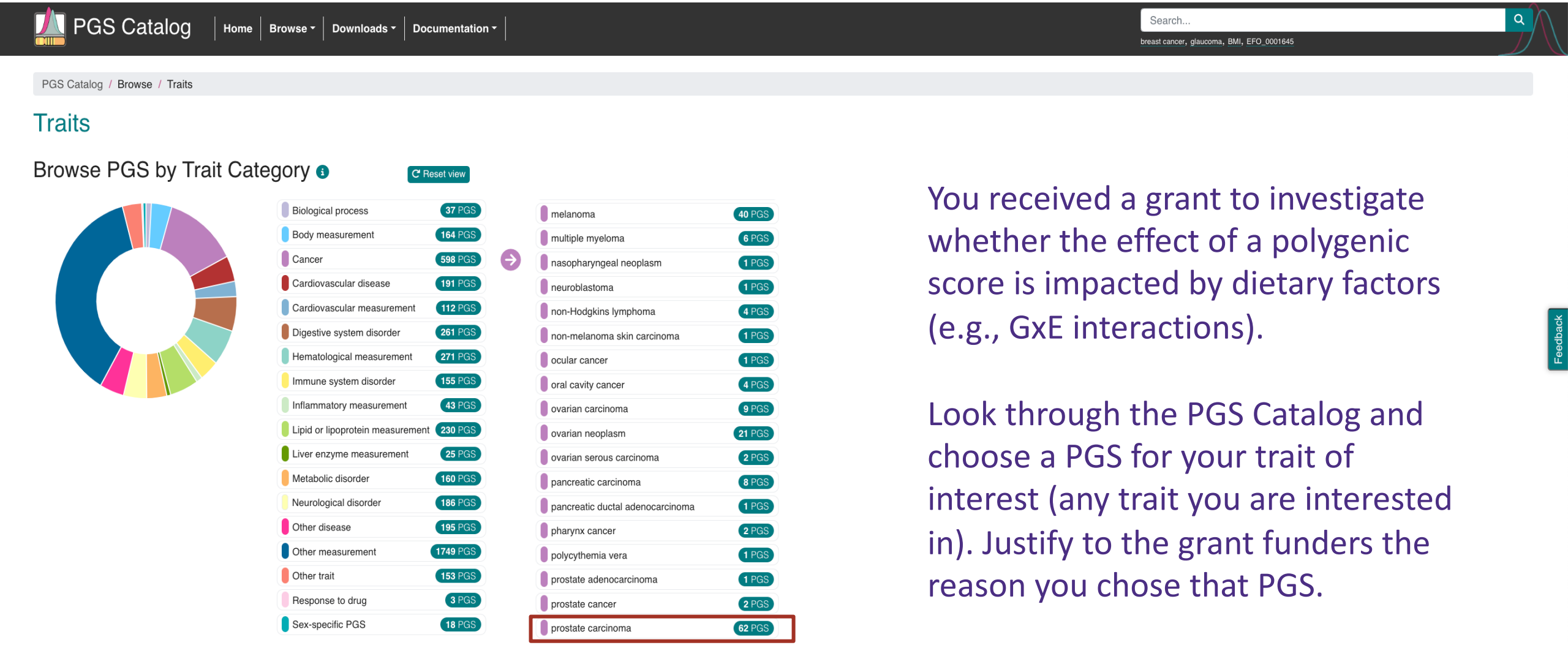
PRS accuracy is <u>also</u> dependent on trait heritability (h^2)

- Traits with low h^2 typically lead to poor predictive models

Breakout Activity

pgscatalog.org

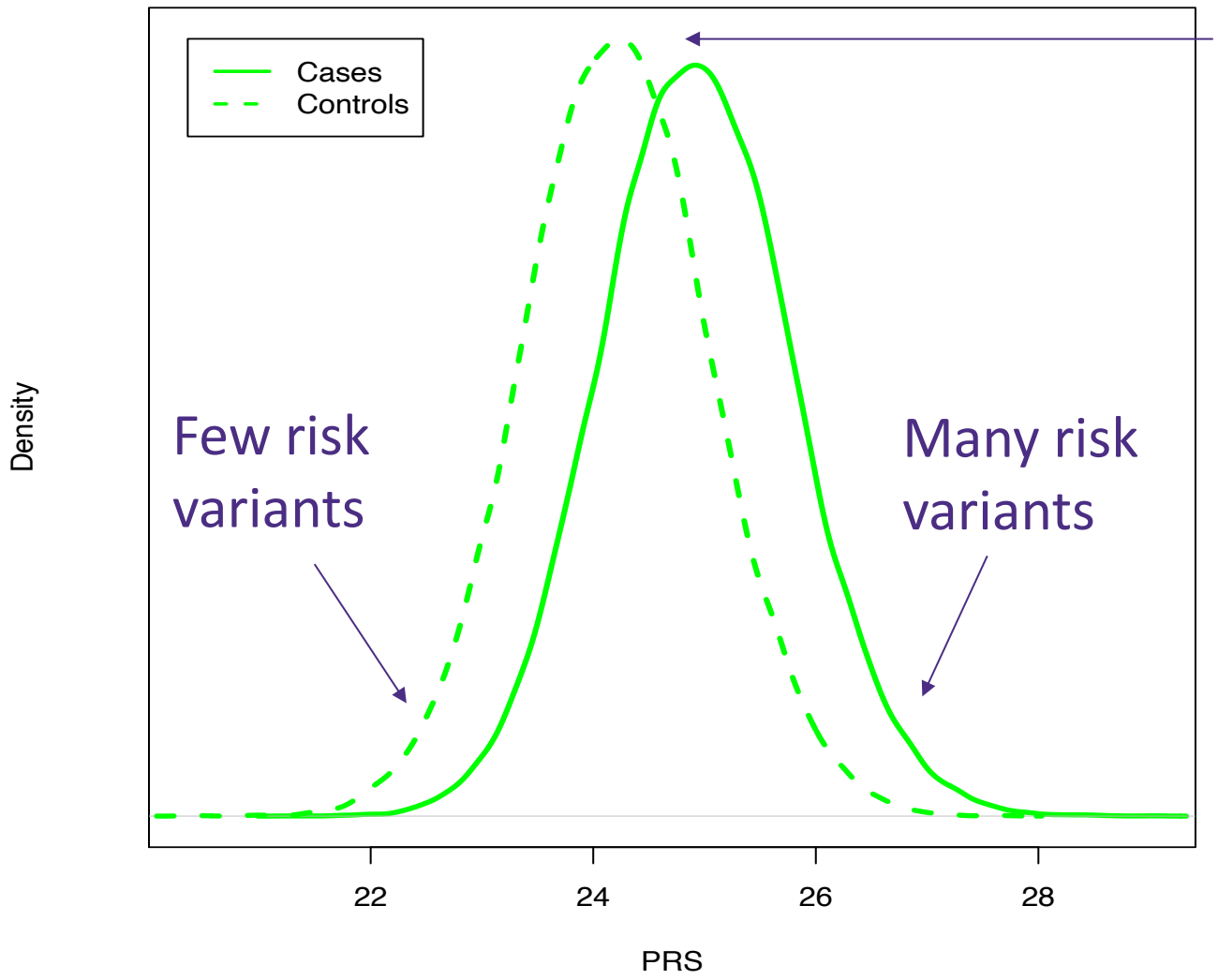
Lambert et al., *Nature Genetics* 2021



You received a grant to investigate whether the effect of a polygenic score is impacted by dietary factors (e.g., GxE interactions).

Look through the PGS Catalog and choose a PGS for your trait of interest (any trait you are interested in). Justify to the grant funders the reason you chose that PGS.

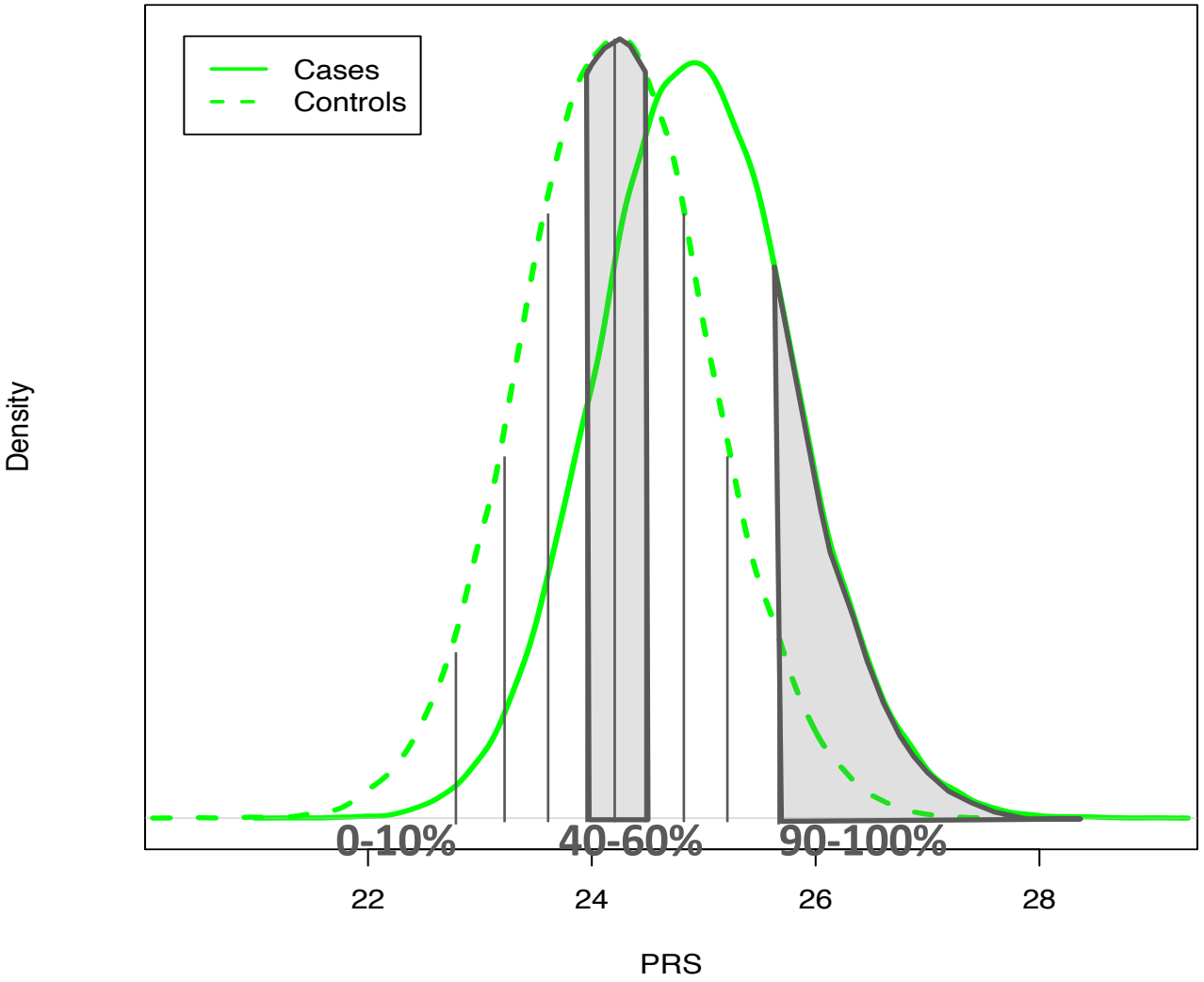
Evaluating Polygenic Risk Scores



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

[1] Choi et al., *Nature Protocols* 2020, A guide to performing polygenic risk score analyses.
[2] Chatterjee et al., *Nature Reviews Genetics* 2016, Developing and evaluating polygenic risk prediction models for stratified disease prevention.

Evaluating Polygenic Risk Scores

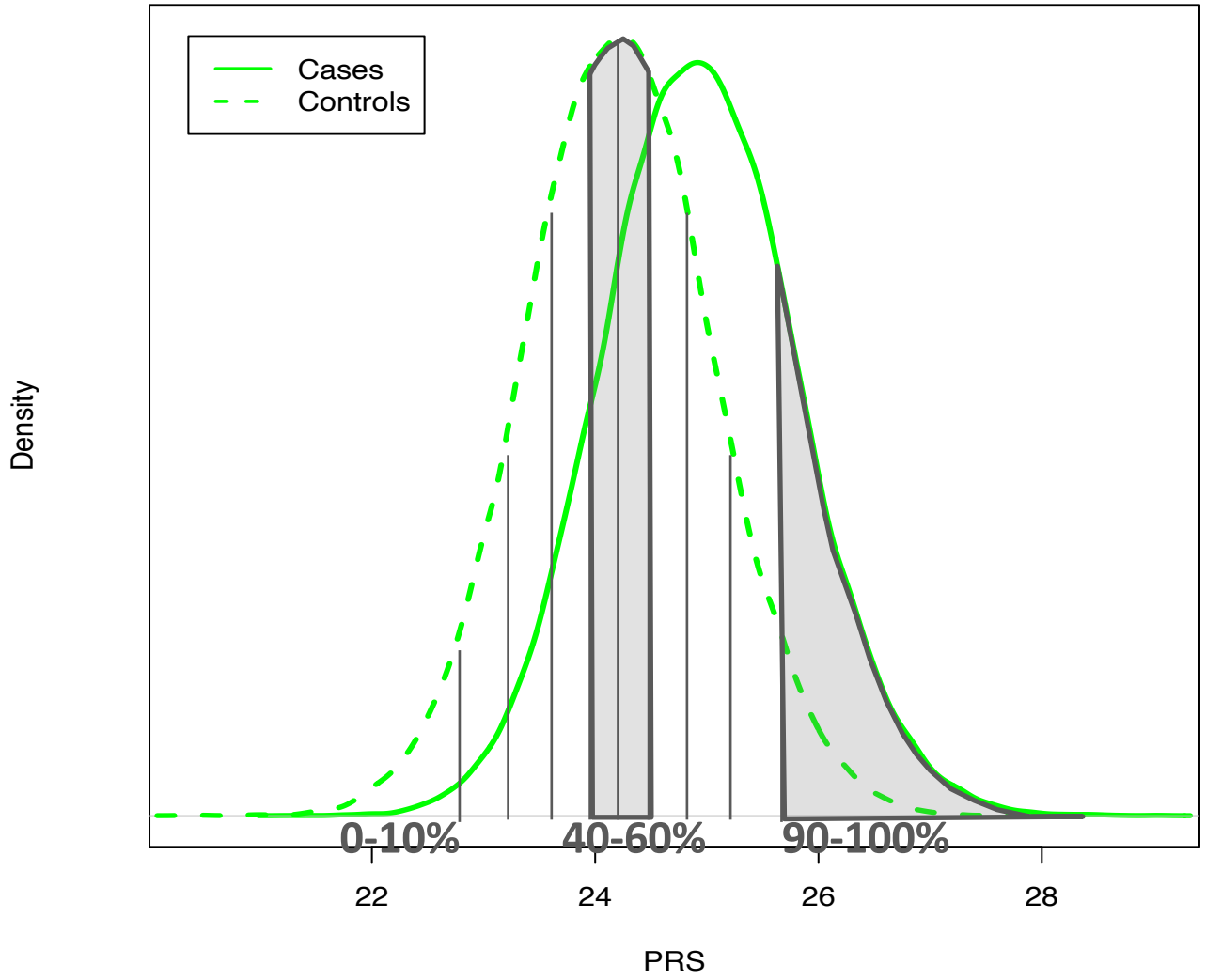


PRS could be analyzed as a:

- *Continuous predictor*
  - Standardize PRS (subtract mean & divide by SD for each individual)
  - Interpretation: "1 SD increase in PRS associated with …"
- *Categorical predictor*
  - Compare individuals between deciles/quantiles (e.g., Disease OR for individuals in 90-100% vs 40-60% PRS)
  - Interpretation: "Individuals in the 90-100% PRS category have x-fold higher odds of disease than those in the 40-60% PRS category"

[1] Choi et al., *Nature Protocols* 2020, A guide to performing polygenic risk score analyses.
[2] Chatterjee et al., *Nature Reviews Genetics* 2016, Developing and evaluating polygenic risk prediction models for stratified disease prevention.

Evaluating Polygenic Risk Scores



Evaluate performance for dichotomous trait

Logistic regression

- Effect on trait (OR & P-value)
- Area under the curve (AUC)
- Net reclassification index (NRI)
- Variance explained (Pseudo R^2)

Evaluate performance for continuous traits

Linear regression

- Effect on trait (beta & P-value)
- Variance explained (R^2)

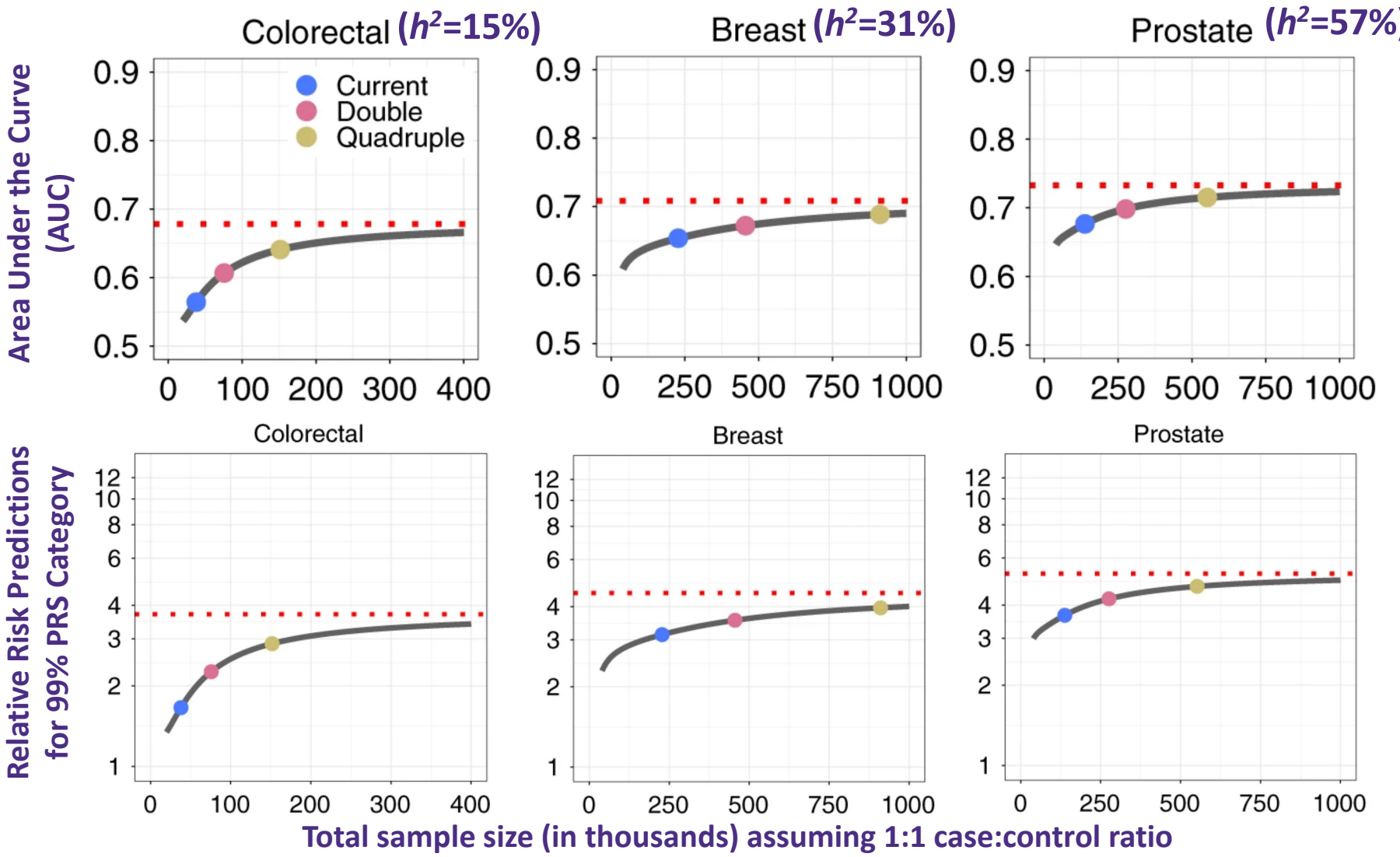
Adjust for age, sex, population stratification (principal components of ancestry)

Few other factors can be true "confounders" but additional adjustment may be necessary

[1] Choi et al., *Nature Protocols* 2020, A guide to performing polygenic risk score analyses.

[2] Chatterjee et al., *Nature Reviews Genetics* 2016, Developing and evaluating polygenic risk prediction models for stratified disease prevention.

Factors that Impact Polygenic Risk Score Performance



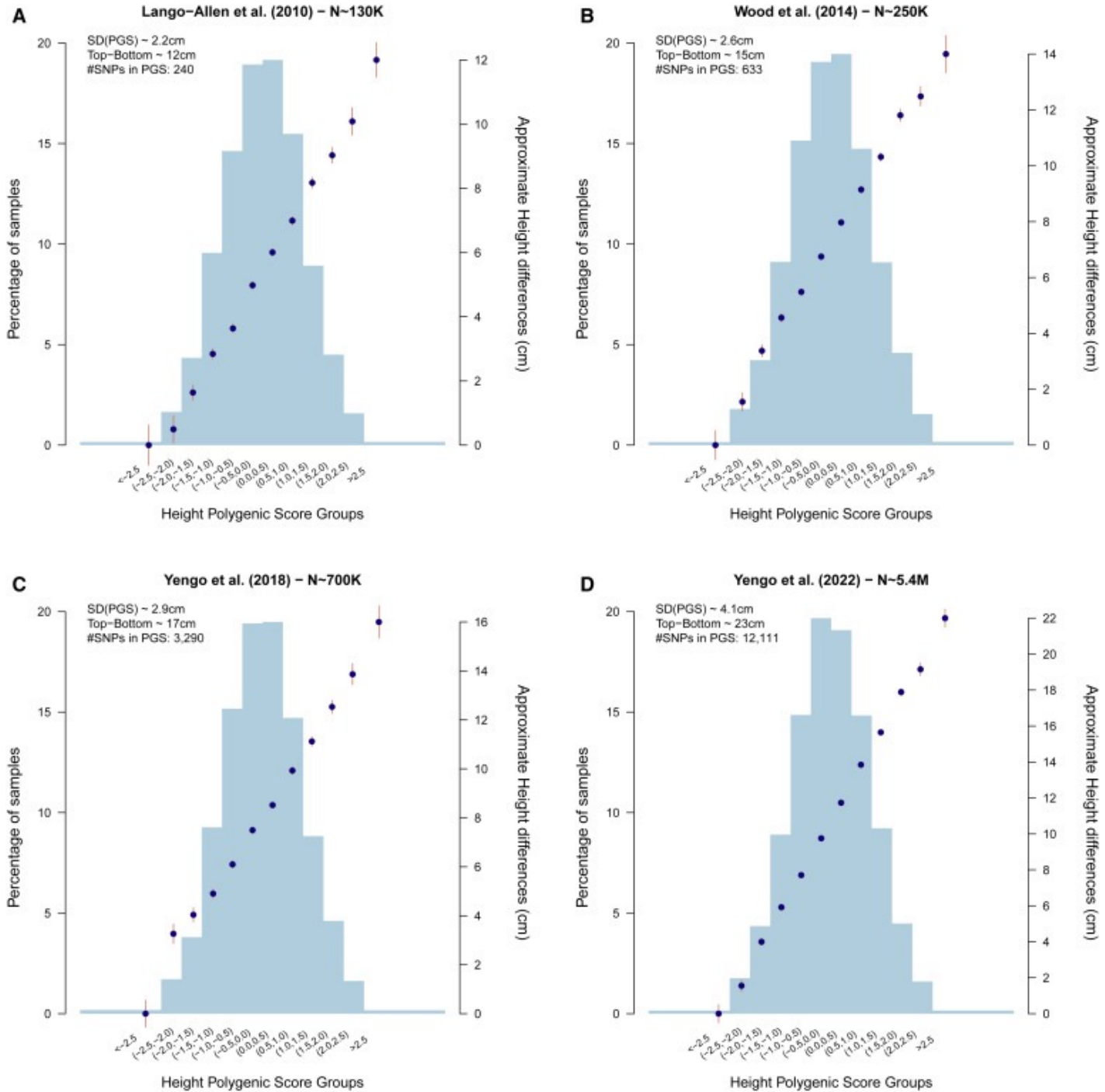
PRS accuracy is highly dependent on sample size and trait heritability (h^2)

[1] Zhang et al., *Nature Communications* 2020, Assessment of polygenic architecture and risk prediction based on common variants across fourteen cancers.
[2] Mucci et al., *JAMA* 2016, Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. (provided heritability estimates above)

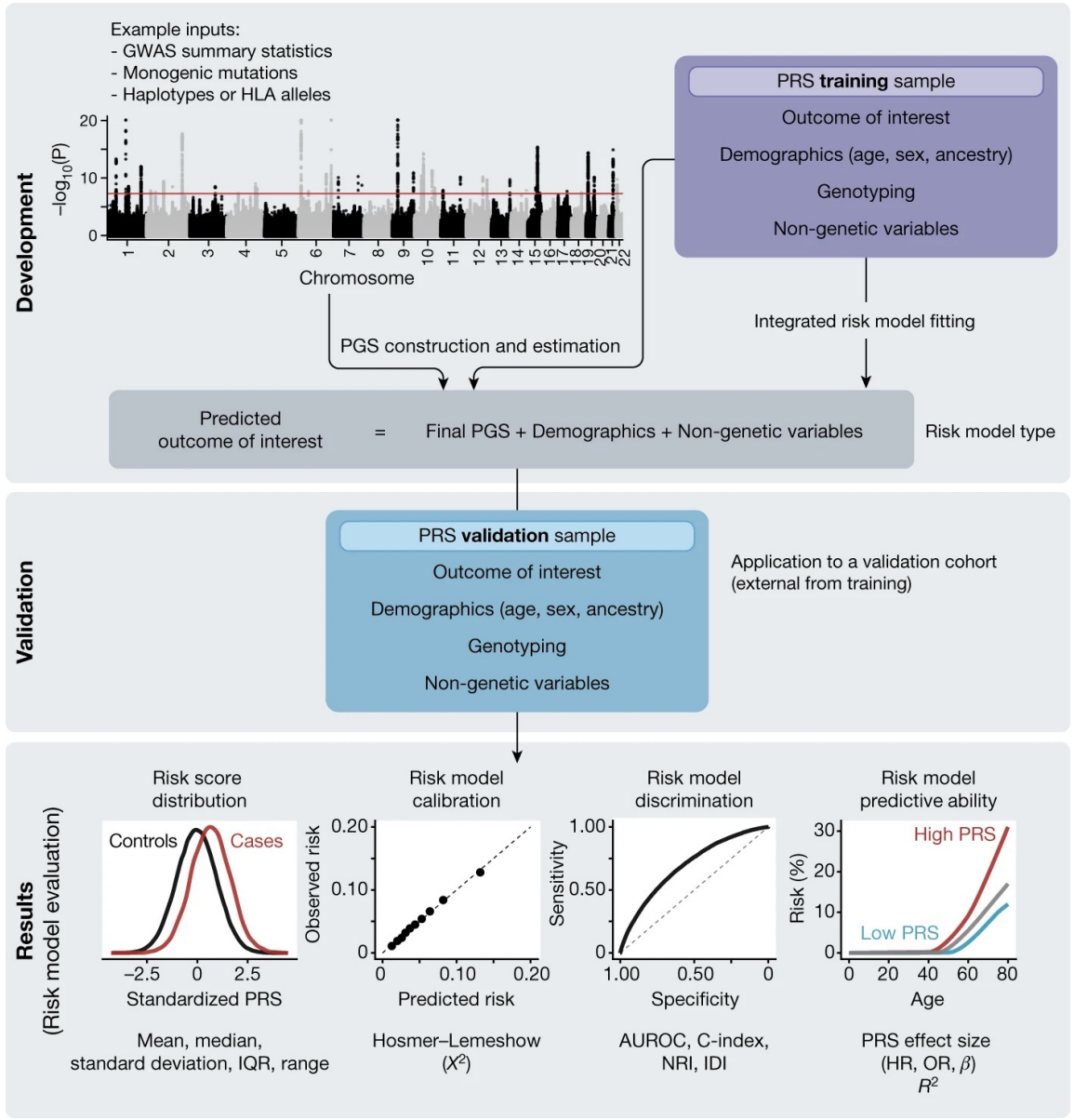
Impact of sample size on PGS performance

As the sample size of height GWAS increases, the predictive ability of polygenic scores (PGS) have improved

Abdellaoui, AJHG 2023



Summary: Developing & Evaluating Polygenic Risk Scores



PRS development, testing, and validation process

Wand, Lambert et al., *Nature* 2021

(Choi et al., Nature Protocols 2020 has a similar diagram)

Training sample. Used to develop the PRS: run GWAS, optimize the PRS (i.e., test many different PRS models and determine which has the best performance).

Validation sample. Independent data used to evaluate the optimized PRS.

How can you newly construct and evaluate PRS?

- 1) Generate/obtain GWAS summary statistics
- 2) Determine variant inclusion threshold
  - Often based on P-values (e.g., $5x10^{-8}$)
  - For “genome-wide” PRS, could include 1.2M HapMap3 variants
- 3) Account for LD
  - Often limit to independent variants (e.g., $r^2<0.10$)
  - Genome-wide PRS approaches often reweight variants to account for LD rather than limiting to independent variants
- 4) Calculate PRS in dataset independent of that used to develop the PRS
- 5) Evaluate PRS performance

Breakout Activity: Calculate PRS

*Participant 1's outcome is not considered

Genetic Variant	Risk allele	Risk weight	Participant 1 Genotype	Participant 1 PRS
rs1234	A	0.02	AA	
rs2345	G	0.04	GT	
rs3456	C	0.05	CT	
rs4567	A	0.09	AC	
rs5678	T	0.004	TT	
rs6789	T	0.07	CC	
rs7891	G	0.01	TT	
rs8912	C	0.015	AA	
			Polygenic risk score	

$$PRS_i = \sum_{m=1}^{M} \beta_m G_{im}$$

β = weight of variant m
G = # risk alleles variant m in individual i
M = total # variants

EPIDEMIOLOGY
SCHOOL OF PUBLIC HEALTH

Breakout Activity: Calculate PRS

*Participant 1's outcome is not considered

Genetic Variant	Risk allele	Risk weight	Participant 1 Genotype	Participant 1 PRS
rs1234	A	0.02	AA	+0.02 * 2
rs2345	G	0.04	GT	+0.04 * 1
rs3456	C	0.05	CT	+0.05 * 1
rs4567	A	0.09	AC	+0.09 * 1
rs5678	T	0.004	TT	+0.004 * 2
rs6789	T	0.07	CC	0
rs7891	G	0.01	TT	0
rs8912	C	0.015	AA	0
			Polygenic risk score	**0.228**

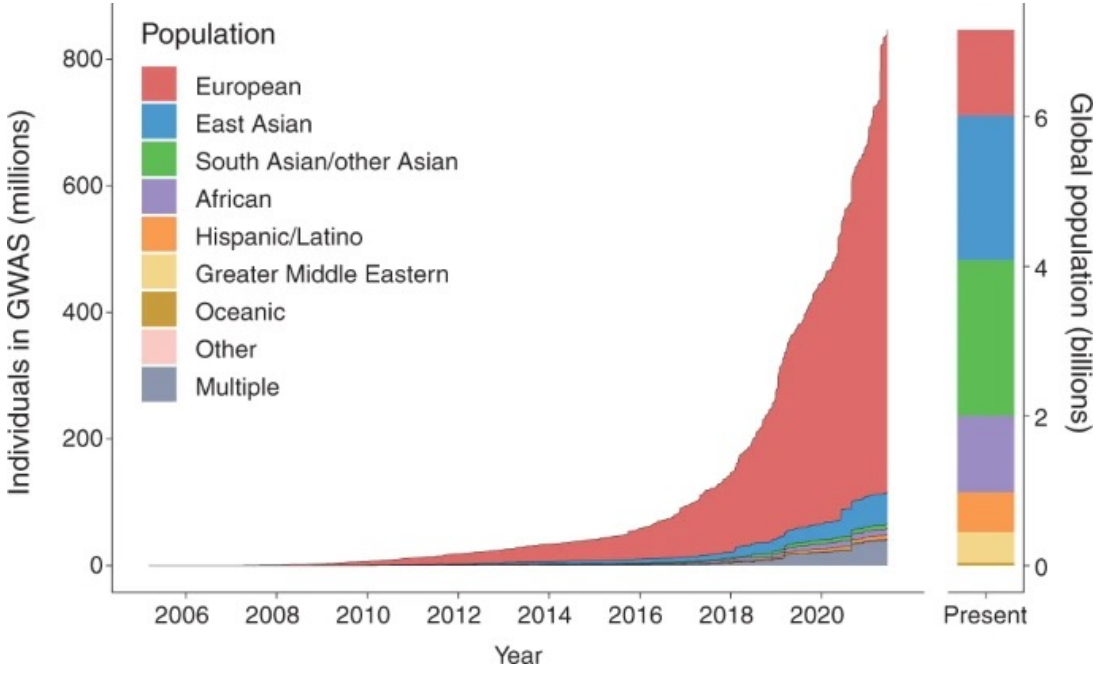
$$PRS_i = \sum_{m=1}^{M} \beta_m G_{im}$$

β = weight of variant m
G = # risk alleles variant m in individual i
M = total # variants

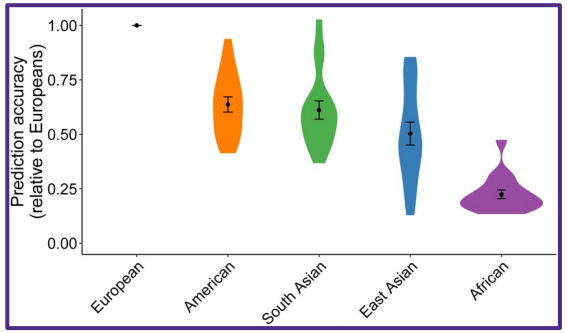
EPIDEMIOLOGY
SCHOOL OF PUBLIC HEALTH

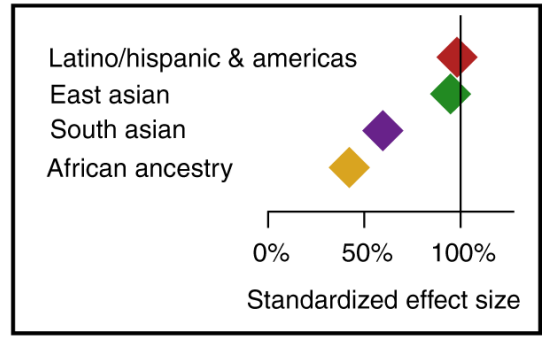
PRS Across Diverse Populations

Lack of Diversity in GWAS Could Contribute to Health Disparities

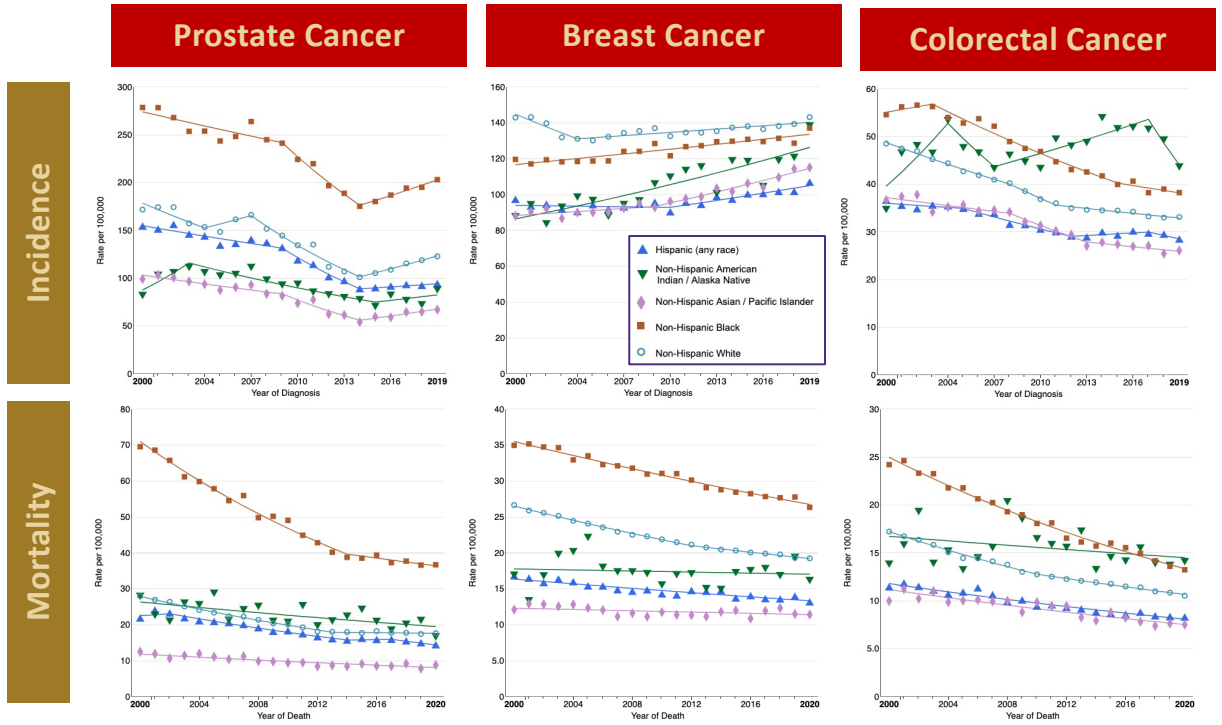


Diversity of GWAS over time relative to the global population





Polygenic predictive ability relative to European ancestry individuals, ~20 traits



Martin et al., *Nature Genetics* 2019
Duncan et al., *Nature Communications* 2019
Fatumo et al., *Nature Medicine* 2022

Established Risk Factors of Common Cancers

	Prostate Cancer Risk Factors	Breast Cancer Risk Factors	Colorectal Cancer Risk Factors
Age	70% cases >65	40% cases >65	56% cases >65
Modifiable/ Other Non-Genetic	Exercise, adiposity, smoking, tomato intake	Exercise, adiposity, alcohol use, HRT, reproductive history	Exercise, adiposity, smoking, alcohol use, red and processed meat intake, fiber intake
Germline Genetics	**h^2~58%**	**h^2~31%**	**h^2~15%**

Heritability (h^2) estimates: Mucci et al., *JAMA* 2016

Current State of GWAS of Common Cancers

Population	Prostate Cancer[1-3]			Breast Cancer[4-7]			Colorectal Cancer[8-11]		
	Cases	Controls	Total	Cases	Controls	Total	Cases	Controls	Total
African	19,391	61,608	80,999	9,241	10,193	19,434	1,894	4,703	6,597
East Asian	10,809	95,790	106,599	27,116	112,407	139,523	25,395	220,368	245,763
European	122,188 (78.2%)	604,640	726,828	133,384 (77.9%)	113,789	247,173	78,706 (74.5%)	170,949	247,655
Hispanic	3,931	26,405	30,336	1,497	3,212	4,709	1,611	4,330	5,941
Total	156,319	788,443	944,762[1]	132,218	116,167	410,839	105,606	400,350	505,956
Largest Published GWAS to Date*	**107,247**	**127,006**	**234,253[3]**	**160,500**	**226,196**	**386,696[2]**	**100,204**	**154,587**	**254,791[5]**

* Based on number of cases

[1] Multi-ancestry: Wang et al., *under revision* → **451 variants**

[2] African: Chen et al., *Eur Urol* 2023 → **9 novel variants**

[3] Multi-ancestry: Conti*, Darst* et al., *Nat Gen* 2021 → **269 variants**

[4] European & East Asian: Jia*, Ping* et al., *AJHG* 2022 → **222 variants**

[5] European & African: Adedokun et al., *Nat Comm* 2021 → **6 loci**

[6] Hispanic: Fejerman et al., *Nat Comm* 2014 → **1 loci**

[7] European & East Asian: Michailidou et al., *Nature* 2017 → **180 loci**

[A] Mavaddat et al., *BCR* 2010

[8] European & East Asian: Fernandez-Rozadilla et al., *Nat Gen* 2023 → **205 variants**

[9] European & East Asian: Xin et al., *Genome Med* 2023 → **48 variants**

[10] African: Wang et al., *IJC* 2017 → **1 loci**

[11] Hispanic: Schmit et al., *Carcinogenesis* 2016 → **4 loci**

Current State of GWAS of Common Cancers

Population	Prostate Cancer[1-3] Cases	Controls	Total	Breast Cancer[4-7] Cases	Controls	Total	Colorectal Cancer[8-11] Cases	Controls	Total
African	19,391	61,608	80,999	9,241	10,193	19,434	1,894	4,703	6,597
East Asian	10,809	95,790	106,599	27,116	112,407	139,523	25,395	220,368	245,763
European	122,188 78.2%	604,640	726,828	133,384 77.9%	113,789	247,173	78,706 74.5%	170,949	247,655
Hispanic	3,931	26,405	30,336	1,497	3,212	4,709	1,611	4,330	5,941
Total	156,319	788,443	944,762[1]	132,218	116,167	410,839	105,606	400,350	505,956
Largest Published GWAS to Date*	**107,247**	**127,006**	**234,253[3]**	**160,500**	**226,196**	**386,696[4]**	**100,204**	**154,587**	**254,791[8]**

* Based on number of cases

Proportion of genetic variation accounted for by known genetic risk variants

	Prostate Cancer **269 known common risk variants[3]**	Breast Cancer **180 known common risk variants[7]**	Colorectal Cancer **205 known common risk variants[8]**
African	43.2%	--	--
East Asian	33.6%	11%	19.7%
European	42.6%	18%	19.7%
Hispanic	39.3%	--	--

[1] Multi-ancestry: Wang et al., *under revision* → **451 variants**

[2] African: Chen et al., *Eur Urol* 2023 → **9 novel variants**

[3] Multi-ancestry: Conti*, Darst* et al., *Nat Gen* 2021 → **269 variants**

[4] European & East Asian: Jia*, Ping* et al., *AJHG* 2022 → **222 variants**

[5] European & African: Adedokun et al., *Nat Comm* 2021 → **6 loci**

[6] Hispanic: Fejerman et al., *Nat Comm* 2014 → **1 loci**

[7] European & East Asian: Michailidou et al., *Nature* 2017 → **180 loci**

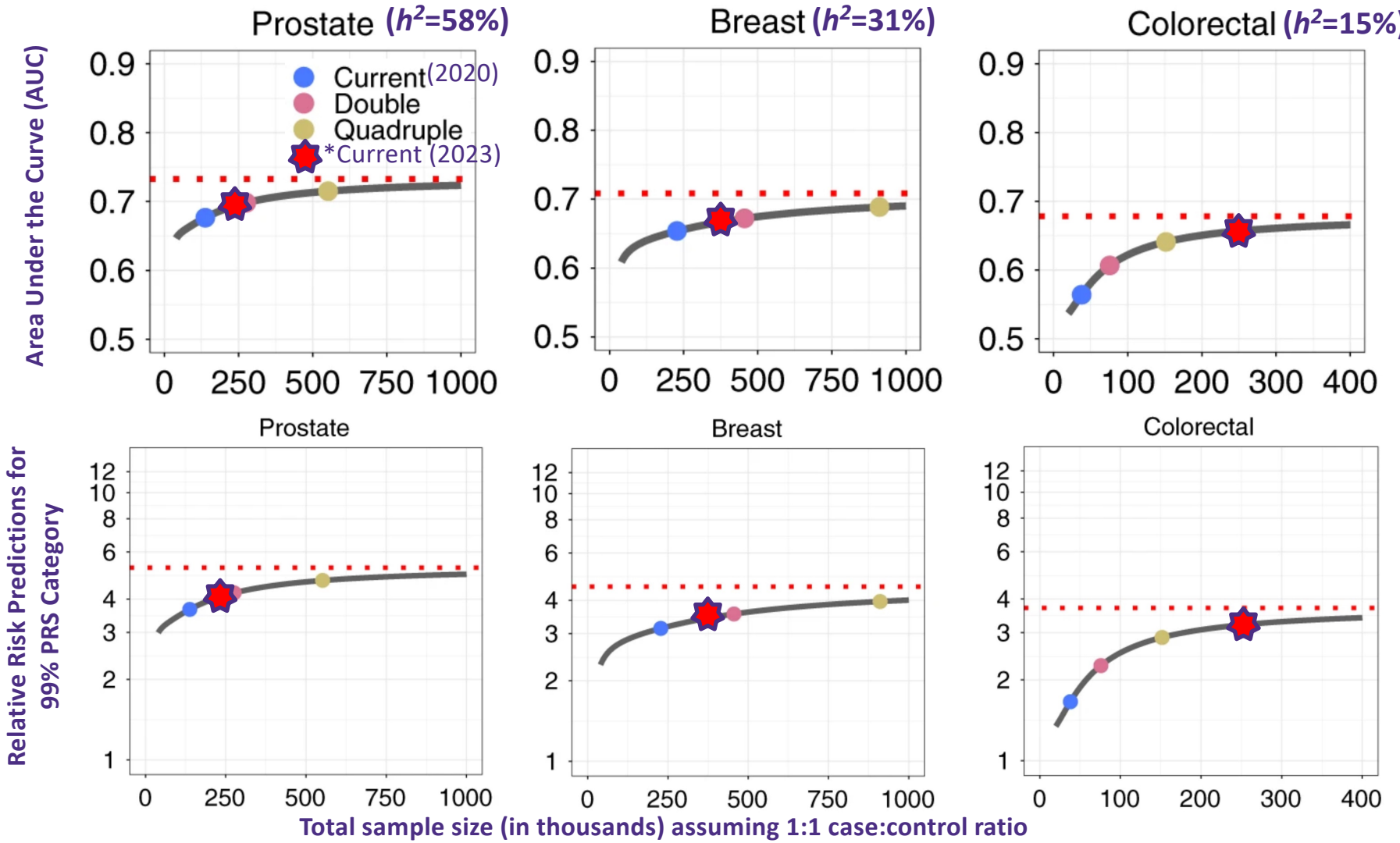
[8] European & East Asian: Fernandez-Rozadilla et al., *Nat Gen* 2023 → **205 variants**

[9] European & East Asian: Xin et al., *Genome Med* 2023 → **48 variants**

[10] African: Wang et al., *IJC* 2017 → **1 loci**

[11] Hispanic: Schmit et al., *Carcinogenesis* 2016 → **4 loci**

Estimated PRS Predictive Ability Trajectories for Common Cancers



PRS accuracy is highly dependent on sample size and trait heritability (h^2)

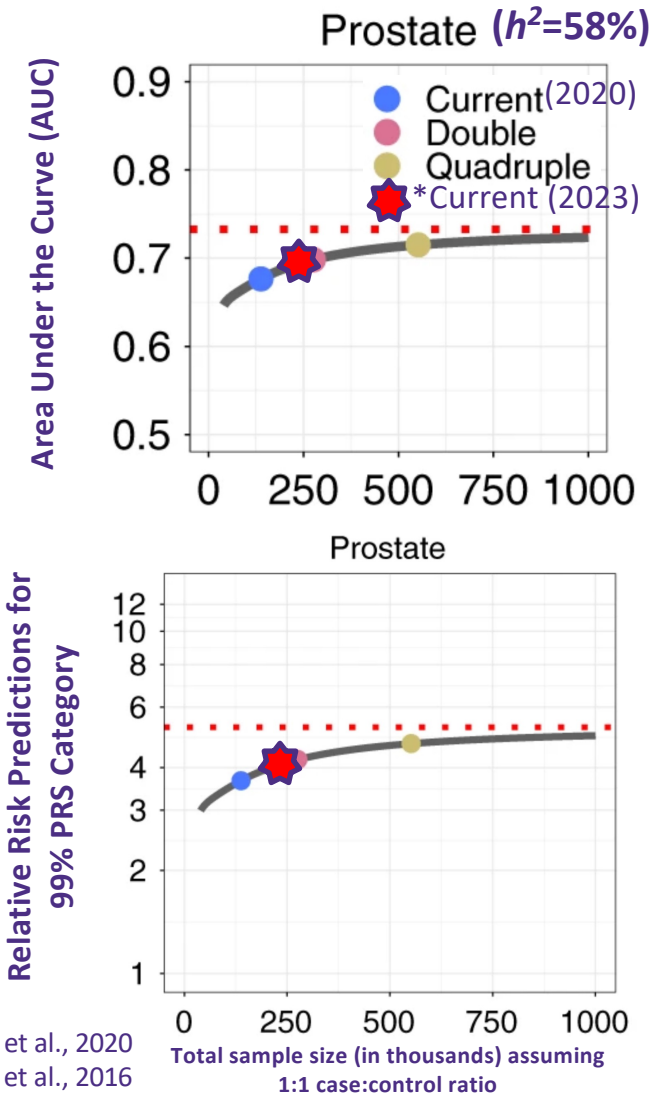
Limitation: Estimates based on European ancestry individuals

Zhang et al., *Nat Comm* 2020
Mucci et al., *JAMA* 2016

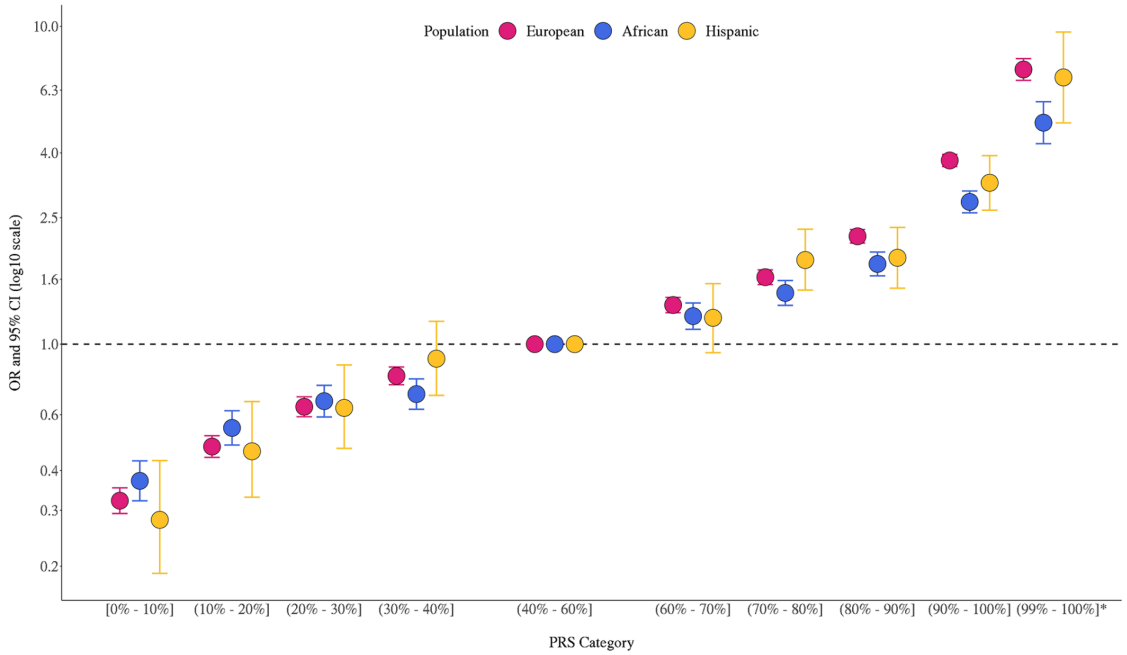
Current PRS Predictive Ability

Prostate Cancer

PRS developed from genome-wide significant variants, with fine-mapping to identify variants most likely to be causal.



Zhang et al., 2020
Mucci et al., 2016



Pop	AUC (95% CI)[a]	ΔAUC[b]	99-100% OR (95% CI), P
African	0.66 (0.65-0.66)	+0.14	4.98 (4.27-5.79), P=5x10^{-95}
Hispanic	0.68 (0.67-0.70)	+0.15	6.91 (4.97-9.60), P=1x10^{-30}
European	0.69 (0.69-0.70)	+0.11	7.32 (6.76-7.92), P=<5x10^{-324}

[a] AUC includes age, PCs, and PRS
[b] AUC adding PRS to age and PCs

PRS Training Data

269 Variants and Weights (Conti*, Darst* et al., *Nat Gen* 2021)

Population	Cases	Controls
African	10,368	10,986
East Asian	8,611	18,809
European	85,554	91,972
Hispanic	2,714	5,239

PRS Testing Data[1]

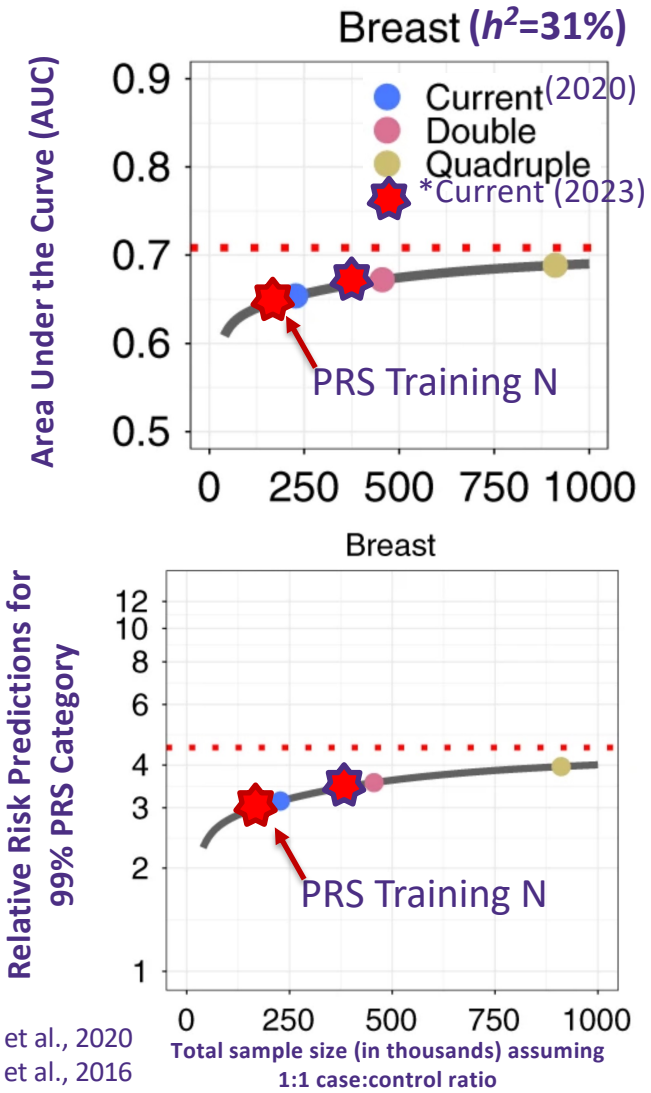
Population	Cases	Controls
African	8,794	55,657
European	22,049	414,249
Hispanic	1,082	20,601

[1] Chen*, Darst* et al., *eLife* 2022

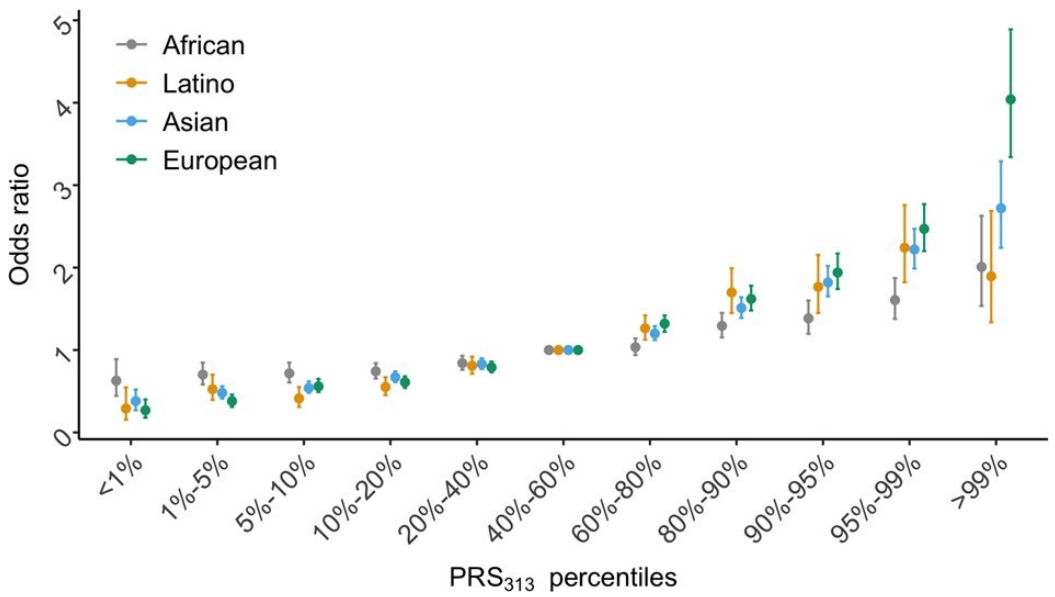
Current PRS Predictive Ability

Breast Cancer

PRS developing using hard-thresholding (P<10-5) stepwise forward regression



Zhang et al., 2020
Mucci et al., 2016



Pop	AUC (95% CI)[a]	99-100% OR (95% CI)
Asian	0.62 (0.60-0.63)	2.72 (2.24-3.29)
African	0.57 (0.56-0.58)	2.01 (1.53-2.63)
European	0.65 (--)	4.37 (3.59-5.33)
Hispanic	0.63 (0.62-0.64)	1.90 (1.41-2.65)

[a] AUC adjusts for PCs and study

PRS Training Data

313 Variants and Weights
(Mavaddat et al., *AJHG* 2019)

Population	Cases	Controls
European	94,075	75,017

PRS Testing Data[1-4]

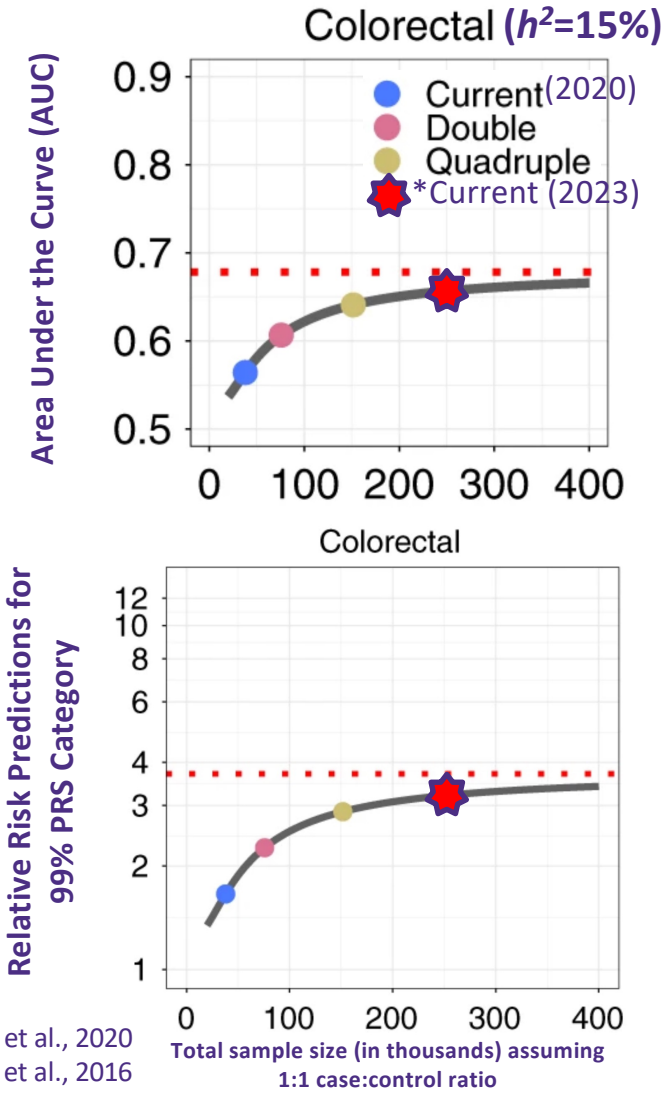
Population	Cases	Controls
African	9,241	10,192
Asian	17,262	17,695
European	11,428	18,323
Latino	4,658	7,622

[1] Du et al., *JNCI* 2021
[2] Ho et al., *Nat Comm* 2020
[3] Shieh et al., *JNCI* 2020
[4] Mavaddat et al., *AJHG* 2019

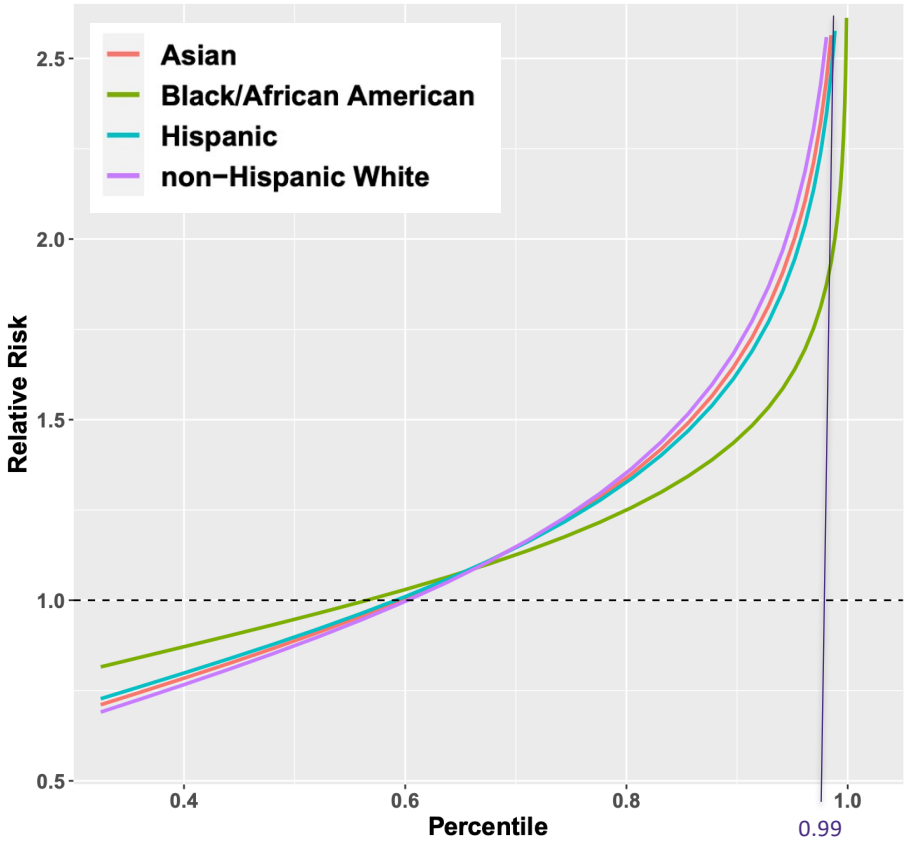
Current PRS Predictive Ability

Colorectal Cancer

PRS developed using PRS-CSx (Ruan et al., Nature Genetics 2022) with 1.2M HapMap3 variants.



Zhang et al., 2020
Mucci et al., 2016



[1] Thomas et al., medRxiv 2023

PRS Training Data

Genome-wide PRS (Fernandez-Rozadilla et al., *Nat Gen* 2023)

Population	Cases	Controls
European	78,473	107,143
Asian	21,731	47,444

PRS Testing Data[1]

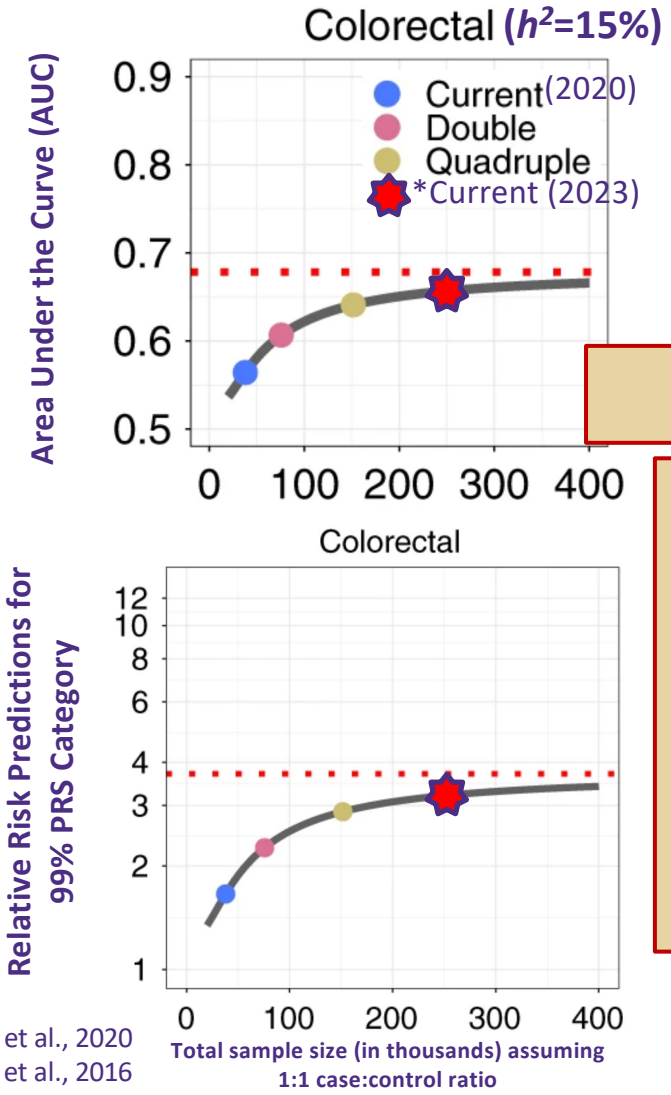
Population	Cases	Controls
African	1,954	11,869
Asian	2,420	9,605
European	3,651	115,105
Hispanic	1,681	8,696

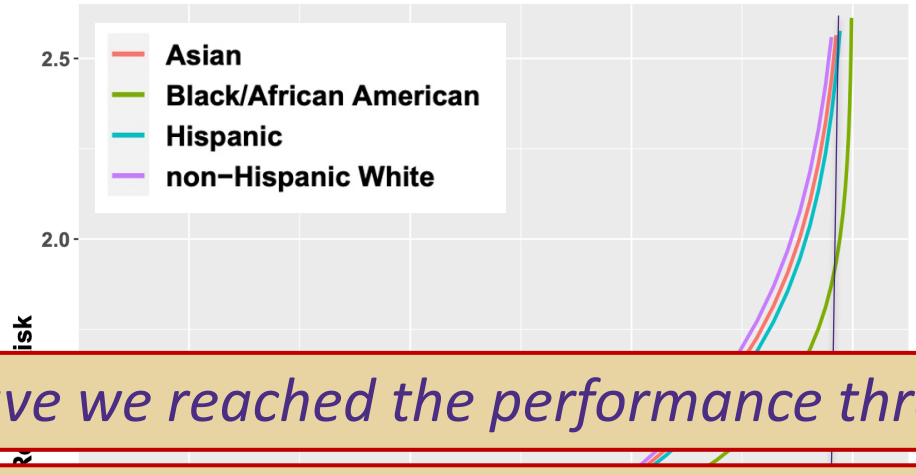
Pop	AUC (95% CI)	99-100% OR (approx.)
African	0.59 (0.57-0.61)	2.0
Asian	0.63 (0.62-0.64)	2.6
European	0.65 (0.64-0.66)	2.7
Hispanic	0.62 (0.60-0.63)	2.5

[a] AUC adjust for age, sex, and PCs

Current PRS Predictive Ability

Colorectal Cancer





PRS Training Data

Genome-wide PRS (Fernandez-Rozadilla et al., *Nat Gen* 2023)

Population	Cases	Controls
European	78,473	107,143
Asian	21,731	47,444

Have we reached the performance threshold for PRS?

Unlikely!

- LD differences between populations → opportunity for improved performance within and across populations
- *Crucial* next step: increase the representation of non-European ancestry individuals

ata[1]

	Controls
African	11,869
	9,605
	115,105
	8,696

		99-100% OR (approx.)
		2.0
		2.6
		2.7
Hispanic	0.62 (0.60-0.63)	2.5

Zhang et al., 2020
Mucci et al., 2016

[1] Thomas et al., medRxiv 2023

[a] AUC adjust for age, sex, and PCs

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