Introduction to Genomic Prediction

Evangelina López de Maturana & Oscar González-Recio

Topics

GP in human genetics	Polygenic risk scores	GP in animal and plant breeding	Comparison between PRS and GS
Background	Background	Overview	Overview
Particularities	PRS analysis process	Comparison of plant and animal breeding approaches	
GWAS in human genetics	Accuracy	Accuracy	
	Limitations		
	Other considerations		
	Examples		
	Resources		
	PRS applications		
	Ethical considerations		

PATHOGENIC PERIOD

PREPATHOGENIC PERIOD

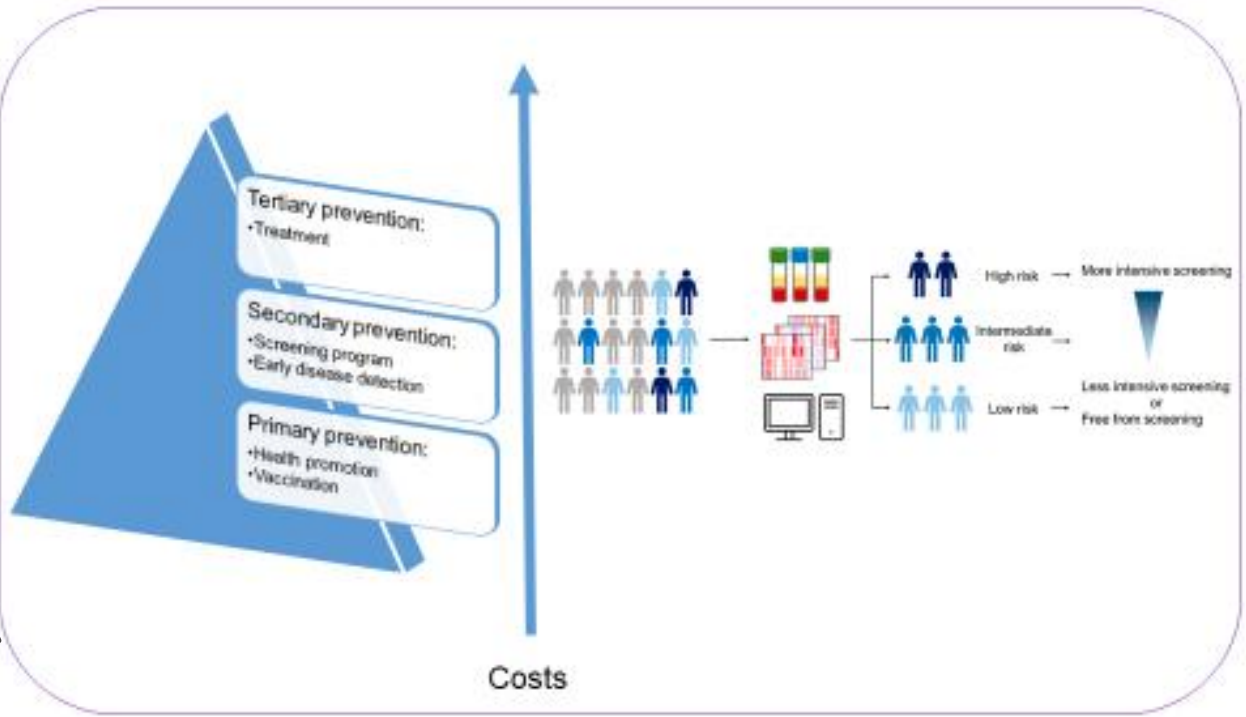
Restoring health and rehabilitation

Reduce the prevalence:

- Screening
- Early diagnosis
- Early therapy
- Reduce the consequences

Reduce the incidence

Preventive strategies against disease



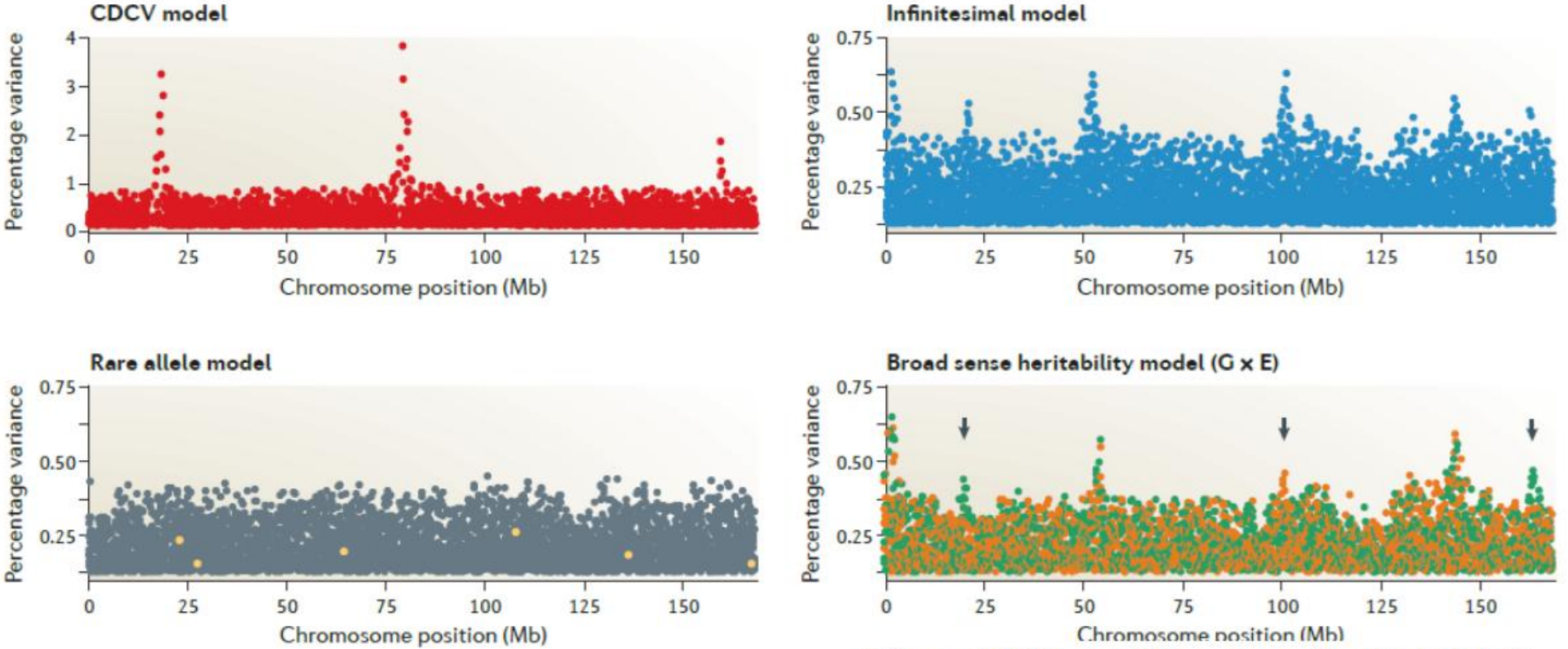
Genome-wide predictor

Genome-wide Prediction in Human Genetics

Prediction of disease risk is an essential part of preventative medicine, often guiding clinical management

Improving effective medical treatment and preventative interventions needs to know how modifiable social, behavioural and physiological factors influence risk of disease (Abraham et al., 2016), as well as the non-modifiable factors:

- Non-genetic risk factors: age, sex, family history of disease, lifestyle factors (smoking status, alcohol consumption ...), comorbidities (e.g., diabetes)
- Genetic risk factors: the genetic basis for many human traits and diseases has been established as polygenic (contributions of many genes each of them contributing very little to the trait), in contrast to Mendelian diseases (caused by variation in one or few genes with large effect)



Gibson (2012). Nat Rev Genetics 13,135-145

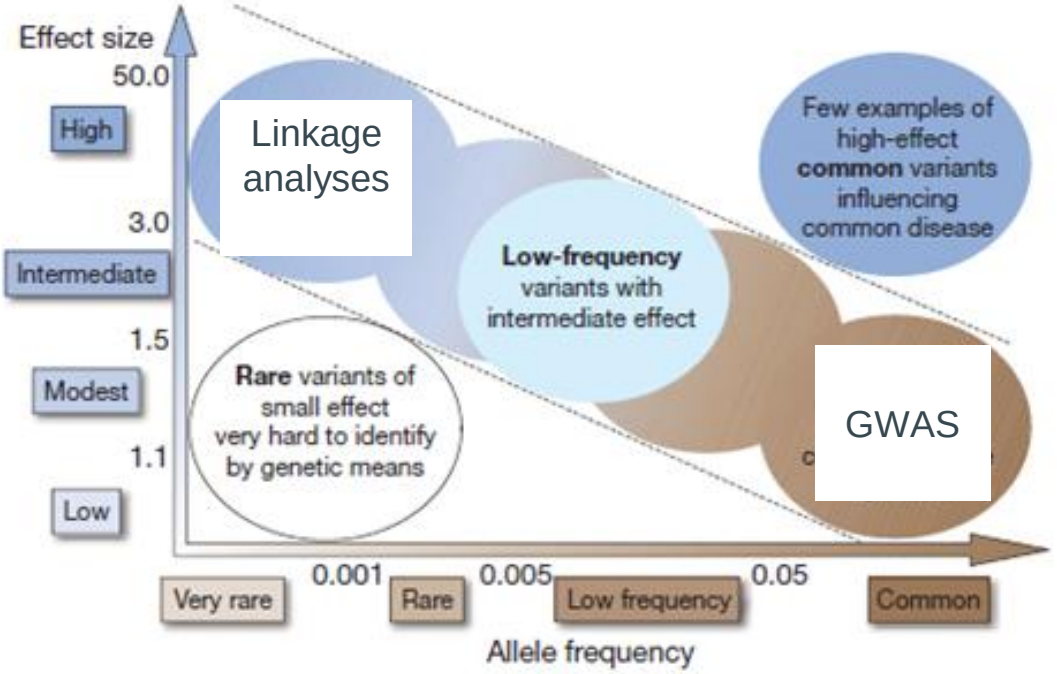
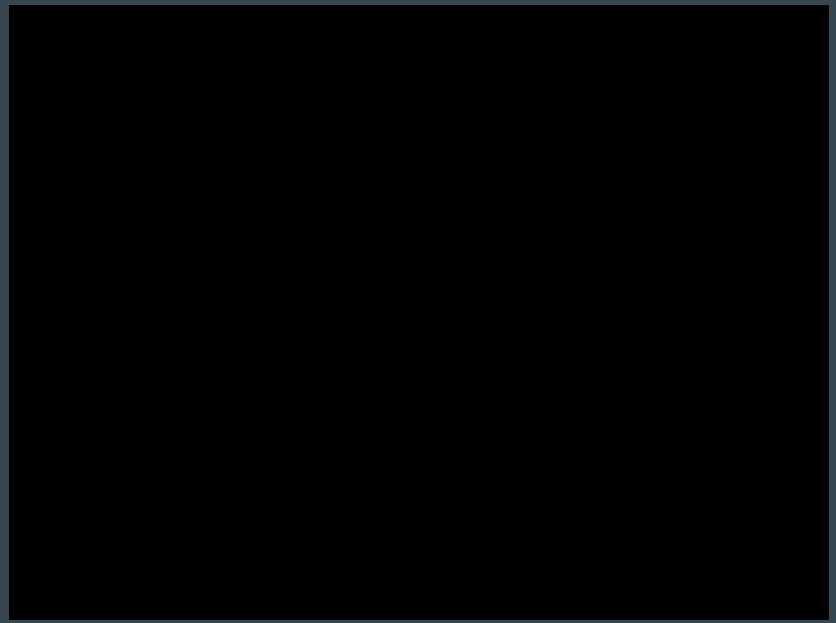


Figure 1 | Feasibility of identifying genetic variants by risk allele frequency and strength of genetic effect (odds ratio). Most emphasis and interest lies in identifying associations with characteristics shown within diagonal dotted lines. Adapted from ref. 42.

GWAS in Human Genetics

- Genome-wide association studies (GWAS) have identified many SNPs-trait associations
- GWAS catalog (https://www.ebi.ac.uk/gwas/home) contains a high-quality collection of all published (and since 2020 also unpublished) GWAS studies
- As of 2025-01-30, the GWAS Catalog contains 7139 publications, 782879 top associations and 102188 full summary statistics
- GWAS Catalog data is currently mapped to Genome Assembly GRCh38.p14 and dbSNP Build 156
- GWAS data are often made available only as summary statistics (Estimated Beta, *p*-value).
- GWAS Catalog

GWAS in Human Genetics



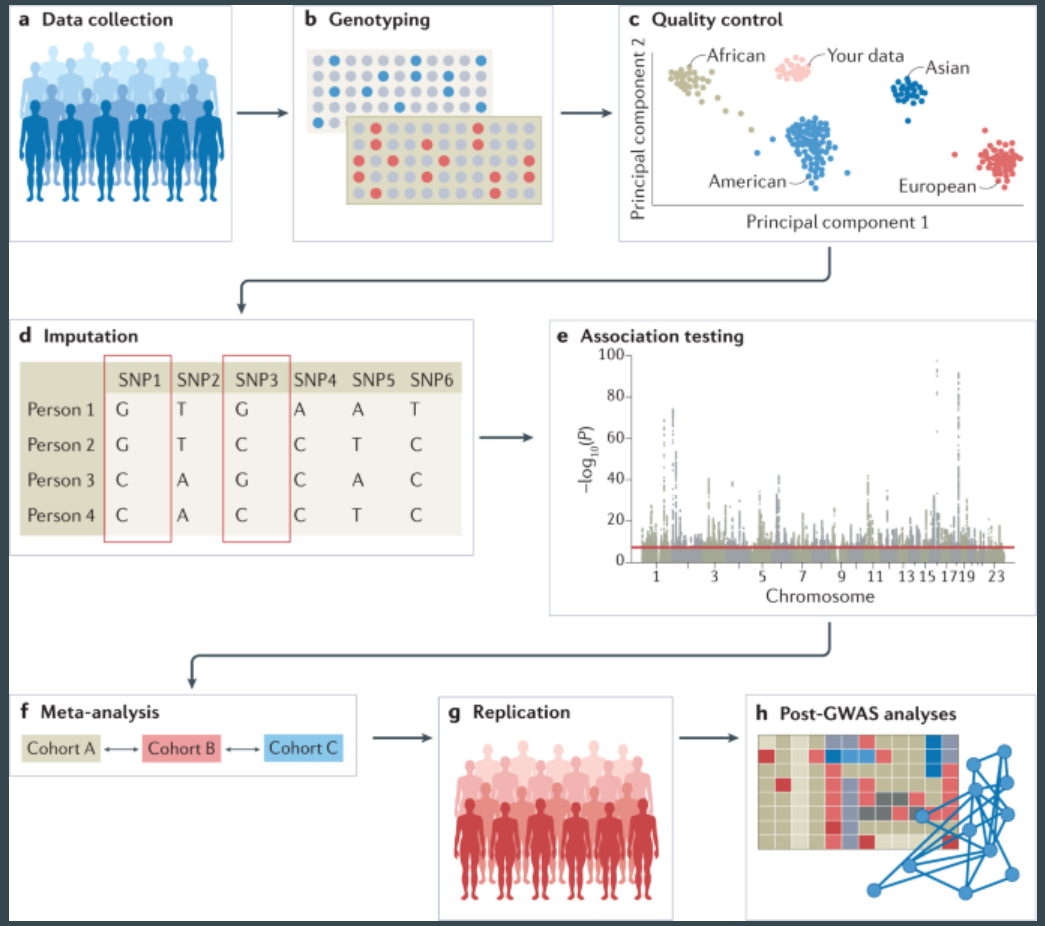
GWAS in Human Genetics

These GWAS SNP-trait associations have provided:

- insights into the genes and pathways that cause disease
- more recently the use of these data for disease risk prediction

Fundamentals:

- Comparison of the frequency of alleles, genotypes or haplotypes in candidate genes or anonymous genome regions between unrelated affected and unaffected individuals
- The alleles analyzed may be thought to contribute to the disease or be in linkage disequilibrium with any such causative variation
- It can provide sufficient power to distinguish slight variations in disease risks, being more sensitive than linkage methods when the genes of interest contribute to disease susceptibility but are neither necessary nor sufficient to cause disease



Uffelmann et al, 2021

Polygenic risk scores

Polygenic risk scores (PRS) (also referred as genomic risk scores) is a method to predict an individual's genetic predisposition for a given disease

It is a single value estimate of an individual's genetic liability to a phenotype

Simplest form:

$$PRS_i = \sum_{j=1}^{m} x_{ij}\hat{\beta}_j$$

Genotype for the *i* individual for the *j* SNP (allelic dosage of the minor or effect allele)

Estimated SNP effect (obtained from GWAS summary statistics)

The genotypes are typically those of common (minor allele frequency > 0.01) biallelic SNPs

Polygenic risk scores

PRS can be constructed from genome-wide significant SNPs ($p < 5x10^{-8}$):

Weakly predictive PRS when the set of GWAS hits is small

PRS with larger number of SNPs (e.g., ($p < 5x10^{-5}$)):

Large # of SNPs with increasingly less precise effect estimates

Small # of SNPs with a more precise effect estimates

Polygenic risk scores

Optimization of PRS:

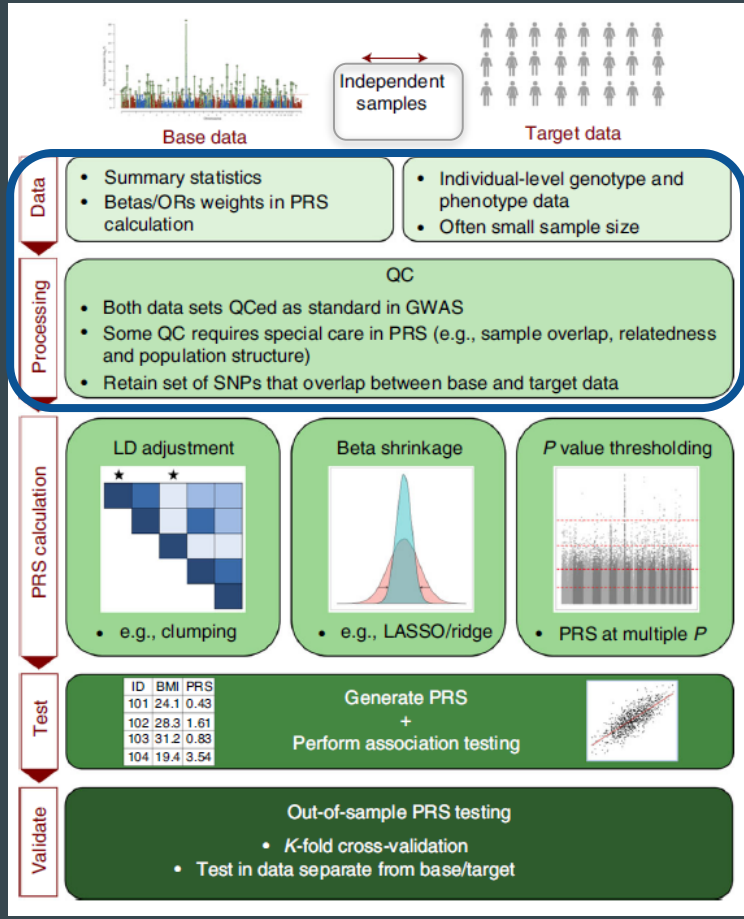
Accounting for LD:
LD pruning (remove one SNP from a pair in high LD)
LD cumpling (LD pruning + most significant SNPs)

Important: No overlap between training and testing datasets

PRS analysis process (https://choishingwan.github.io/PRS-Tutorial/base/)

Most powerful GWAS results available on the phenotype

- Important to check the effect allele:
  - Contact authors if not clear from the summary data
  - Ambiguous alleles (A/T, C/G): check MAF or discard them
  - Mismatching alleles: remove non-resolvable matching SNPs
- Target data with effective samples with >100 indiv
- Check if corrupted files
- Base and target data SNPs assigned to the same genome build
- Base and target data SNPs with good quality:
  - MAF, genotyping rate, HWE, heterozygosity, info



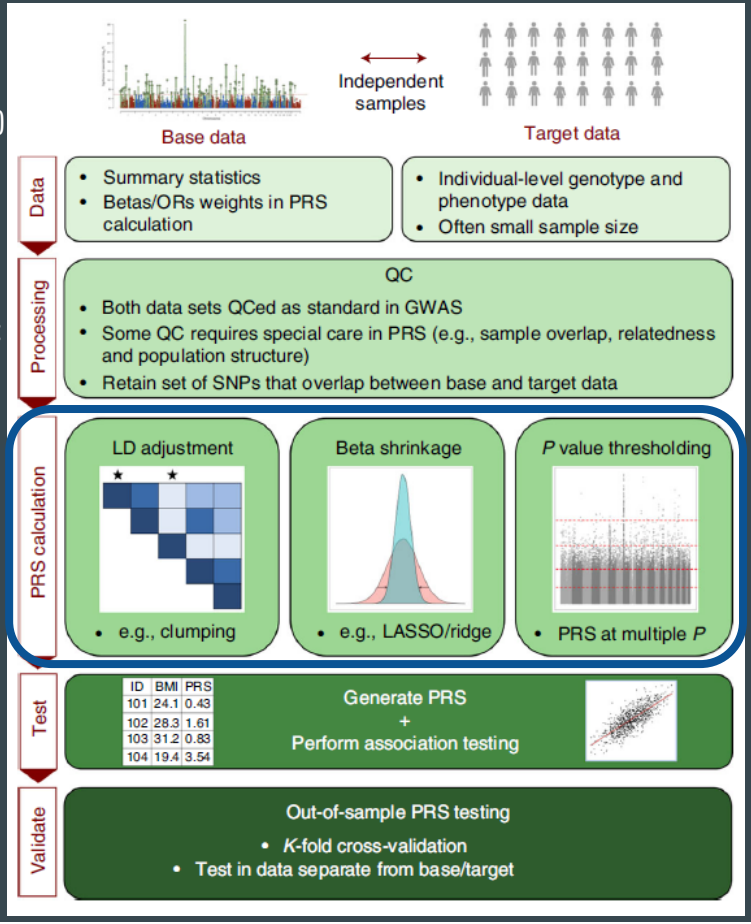
Choi et al, 2020

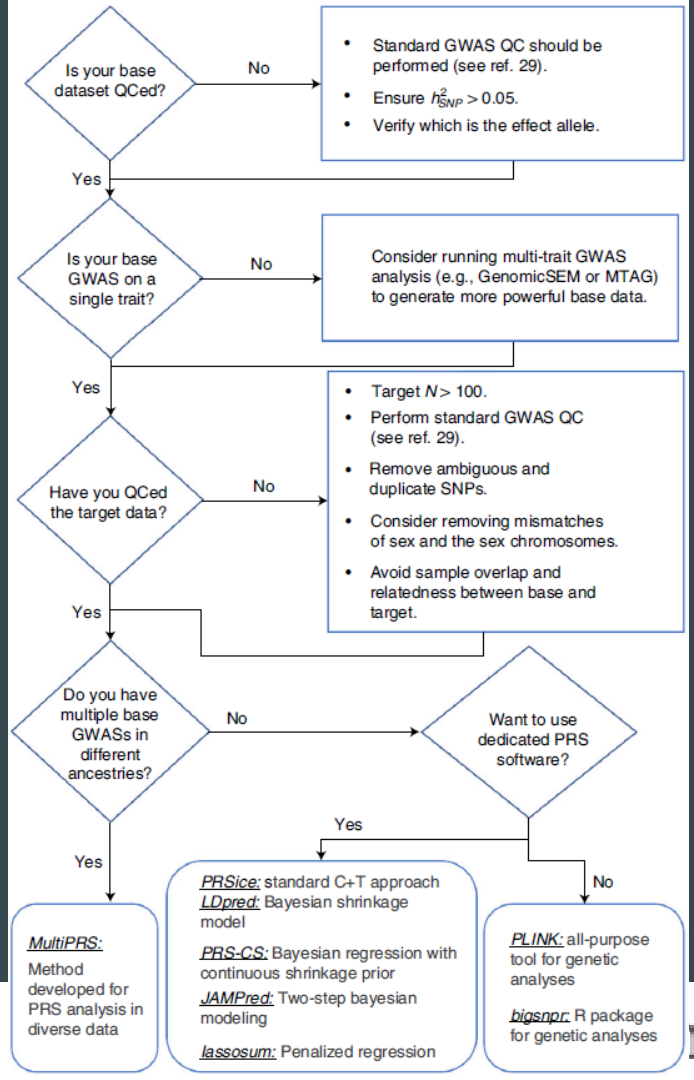
PRS analysis process (https://choishingwan.github.io/PRS-Tutorial/base/)

To calculate PRSs for all individuals in the base sample

Key factors in the development of methods for calculating PRSs:

1. Accounting for LD (if single SNP analysis was used):
   a. clumping (prioritization of SNPs in the locus based on their *p*-value)
   b. Inclusion of all SNPs accounting for LD among them
2. Potential adjustment of GWAS estimated effect sizes:
   a. shrinkage of the effect estimates of all SNPs via standard or tailored statistical techniques
   b. use of P value selection thresholds as inclusion criteria for SNPs into the score (Variable selection)
3. Tailoring of PRSs to target populations:
   a. standardization of the units
   b. Standardization (same scale)
   c. Transformation of the phenotype should be taken into account





- QC steps consist of filtering out of SNPs and individuals:
  - (1) **individual and SNP missingness**
  - (2) inconsistencies in assigned and genetic sex of subjects
  - (3) **minor allele frequency (MAF)**
  - (4) deviations from **Hardy–Weinberg equilibrium (HWE)**
  - (5) **heterozygosity** rate
  - (6) **relatedness**
  - (7) ethnic outliers (see **population stratification**).

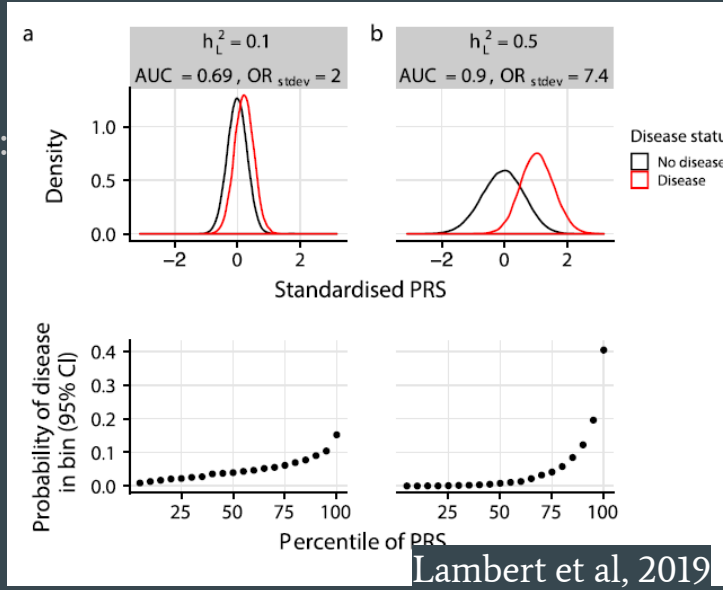
http://zzz.bwh. harvard.edu/plink
https://www.cog-genomics.org/ plink/1.9/

https://doi.org/10.1038/s41596-020-0353-1

ne-wide prediction

Accuracy of PRS

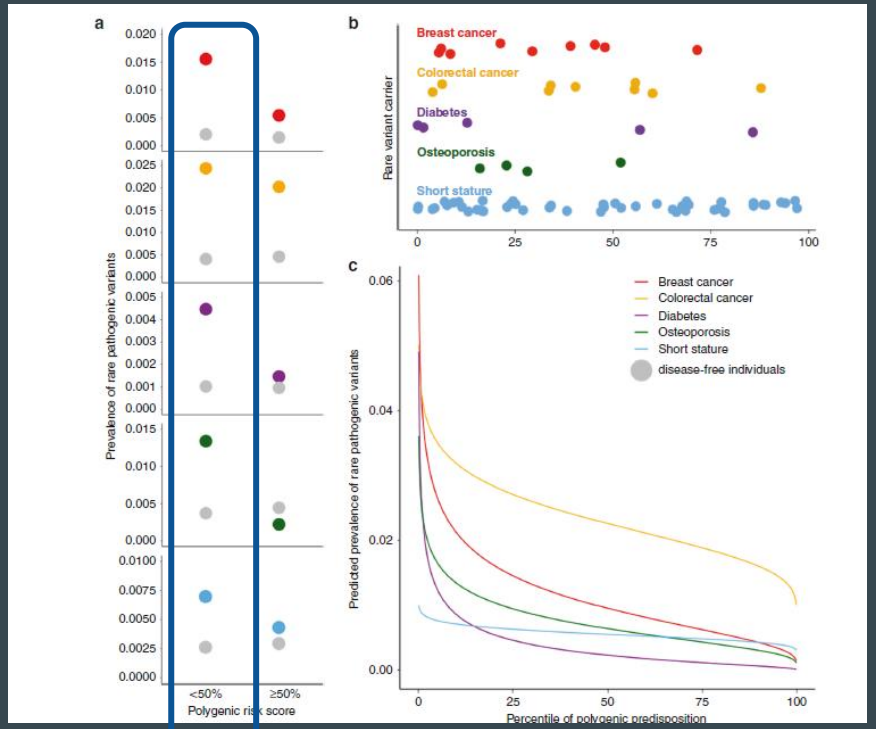
- Disease heritability ($h^2_{SNP} > 0.05$) :
  - Software to estimate h^2_{SNP} from GWAS sum stat:
    - LD score regression (Bullik-Sullivan, 2015)
    - SumHer (Speed and Balding, 2019)



Lambert et al, 2019

Accuracy of PRS

- Disease heritability
- Genetic architecture:
  - Rare genetic causes are more prevalent among patients with a low PRS
  - These patients may be prioritized for deep-depth sequencing of relevant genes



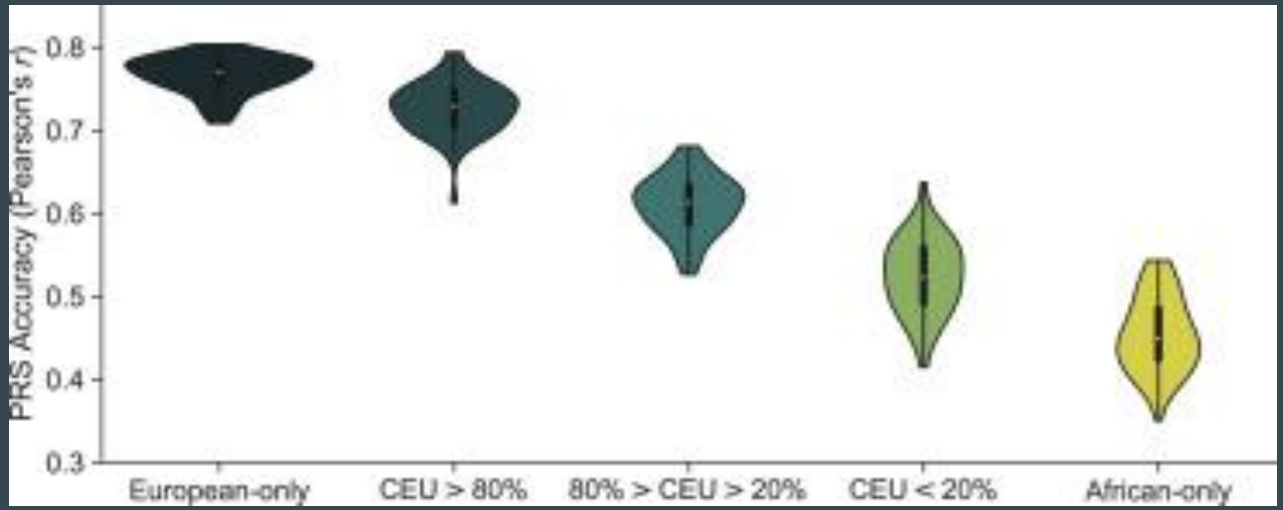
Lu et al, 2020

Accuracy of PRS

- Disease heritability and genetic architecture
- cor(PRS,PRS_estimated). It depends on:
  - Method used to construct the PRS
  - Sample size
- Update of PRSs:
  - GWASs expand in size
  - additional risk loci are identified
  - Alternative methods for score calculation
- Imputation variability in underrepresented populations → health disparities
- Different genetic background (one-third as informative for African ancestry individuals (Duncan et al, 2020)

Accuracy of PRS

Accuracy of PRSs, with variants and weights from a European GWAS, decreases linearly with increasing proportion of African ancestry



Cavazos and Witte, 2021

Polygenic Risk Scores - Limitations

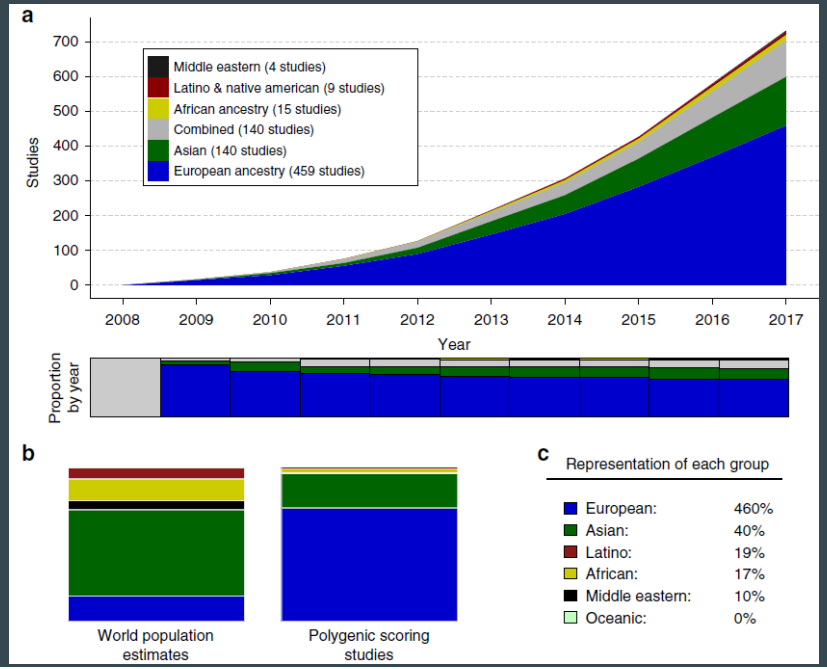
- Individualized PRS should be updated over time → Fluctuations in individual-level scores → may affect the application of PRSs in practice which relates to prioritization of preventative behaviors
- Sparse genotyping approaches (SNP arrays or low-pass WGS) → Need of imputing → Variability introduction at the individual level

Polygenic Risk Scores - Limitations

- Findings regarding genetic liability from resources such as UKBiobank or GWAS results may not be generalizable to individuals who are not of European descent:
    - Differences in variant frequencies
    - LD patterns artifactual differences due to uncorrected population stratification (Berg et al, 2019)

Polygenic Risk Scores - Limitations

Ancestry representation in the first decade of polygenic scoring studies (2008–2017; N = 733 studies)



Duncan et al, 2020

Polygenic Risk Scores - Limitations

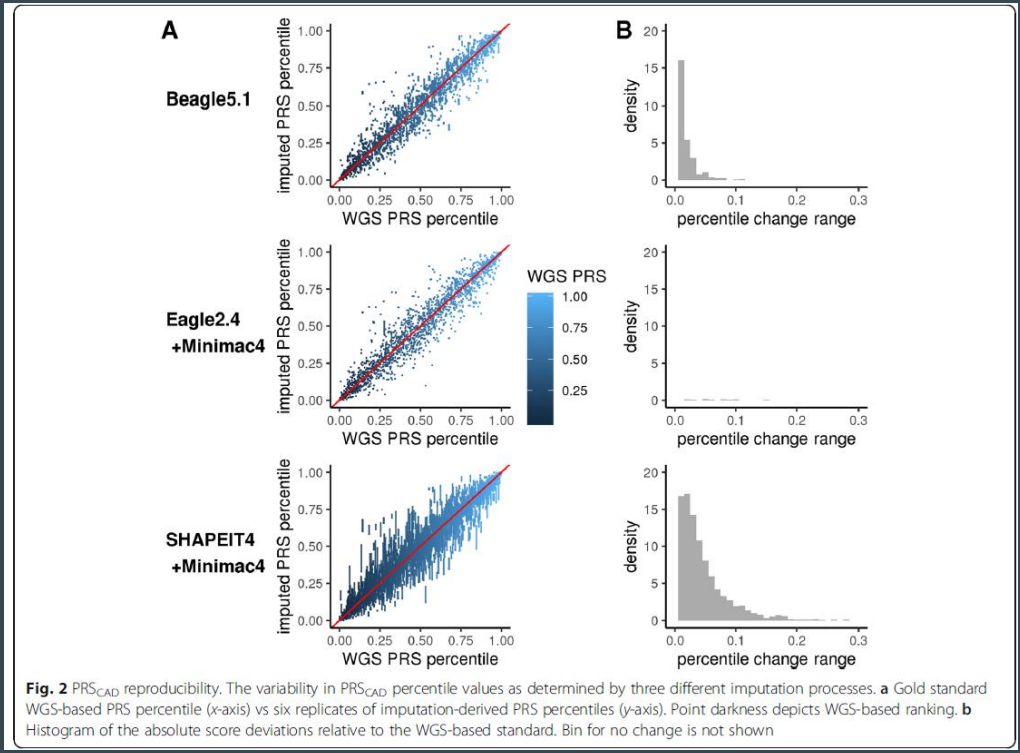


Fig. 2 PRS_{CAD} reproducibility. The variability in PRS_{CAD} percentile values as determined by three different imputation processes. **a** Gold standard WGS-based PRS percentile (*x*-axis) vs six replicates of imputation-derived PRS percentiles (*y*-axis). Point darkness depicts WGS-based ranking. **b** Histogram of the absolute score deviations relative to the WGS-based standard. Bin for no change is not shown

IMPUTATION:

- Pre-phasing step introduces the bulk of the stochasticity in imputation and PRS results
- SHAPEIT+Minimac leads to the most intra-individual variability, followed by Beagle and Eagle+Minimac
- This algorithm-level variability is observed regardless of the original approach used to derive the PRS and the number of SNPs included in the score

Chen et al, 2020

Polygenic Risk Scores - Limitations

- Variability of the PRS percentile is greatest in the middle of the distribution and lowest at the tails (Chen et al, 2020)
  Solution: deterministic imputation processes should be favored or stochastic imputation processes could be run multiple times in order to select the most common result

- Both the clumping and thresholding steps are arbitrary, and reporting the results from the P-value threshold that maximizes out-of-sample prediction in a single cohort is a form of Winner's curse (Wray et al, 2019)

- GWAS one-SNP-at-a-time regression may not the optimal way to estimate SNP effects for use in prediction (Wray et al, 2007):
  - Methods that fit all SNPs simultaneously usually generate more accurate out-of sample prediction than those fitting one SNP at a time

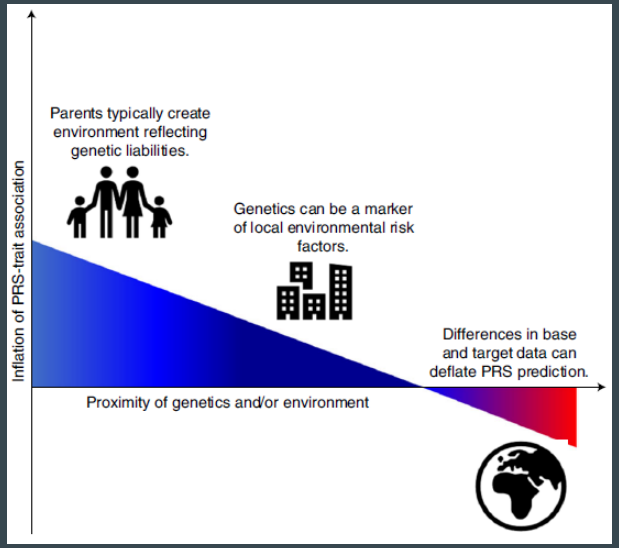
Chen et al, 2020

Polygenic Risk Scores - Other considerations

- Having relatives in the discovery sample will improve the prediction for an individual (Lee et al, 2017):
    - However, having first/second degree relatives may inflate the association between PRS and phenotype (target sample) → removal of those ind (Choi et al., 2020)
- Include Family History as predictor, because it captures genetic and not genetic factors not captured by PRS (Inouye et al, 2018)
    - PRS is an estimate of the aggregate genetic value of an individual, tracking only the genetic contribution to the trait tagged by common DNA polymorphisms.
    - Family history reflects the phenotypes of relatives of the individual.

Polygenic Risk Scores - Other considerations

Major sources of inflation/deflation of PRS-trait associations



Choi et al, 2020

Polygenic Risk Scores - Examples

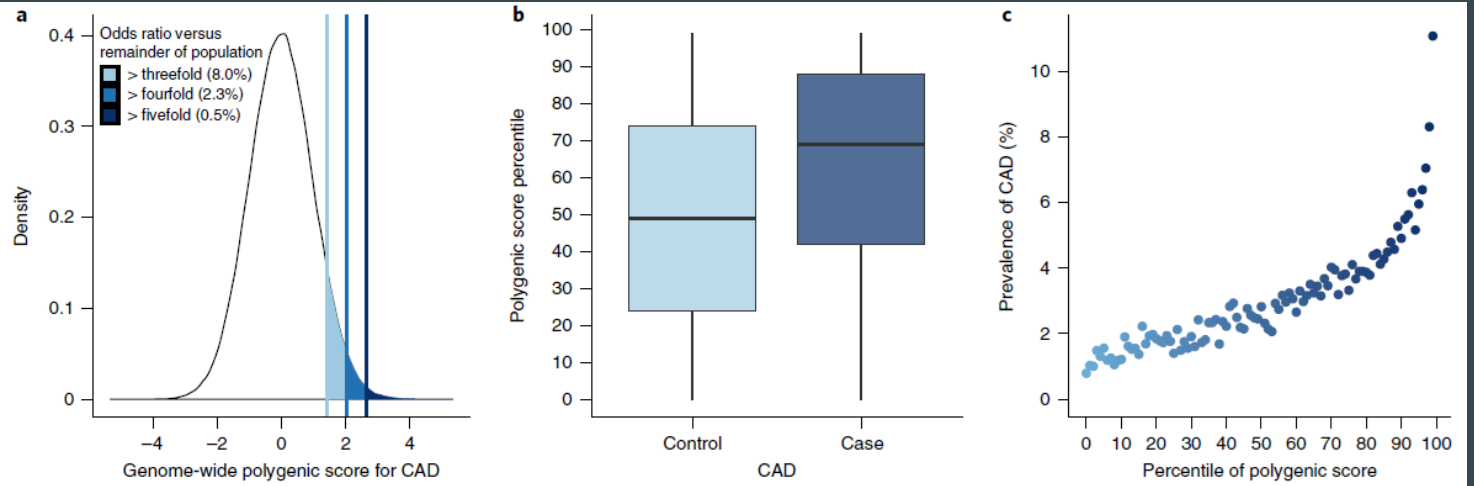


Fig. 2 | Risk for CAD according to GPS. a, Distribution of GPS_{CAD} in the UK Biobank testing dataset ($n = 288{,}978$). The *x* axis represents GPS_{CAD}, with values scaled to a mean of 0 and a standard deviation of 1 to facilitate interpretation. Shading reflects the proportion of the population with three-, four-, and fivefold increased risk versus the remainder of the population. The odds ratio was assessed in a logistic regression model adjusted for age, sex, genotyping array, and the first four principal components of ancestry. **b**, GPS_{CAD} percentile among CAD cases versus controls in the UK Biobank testing dataset. Within each boxplot, the horizontal lines reflect the median, the top and bottom of each box reflect the interquartile range, and the whiskers reflect the maximum and minimum values within each grouping. **c**, Prevalence of CAD according to 100 groups of the testing dataset binned according to the percentile of the GPS_{CAD}.

Khera et al, 2018

A

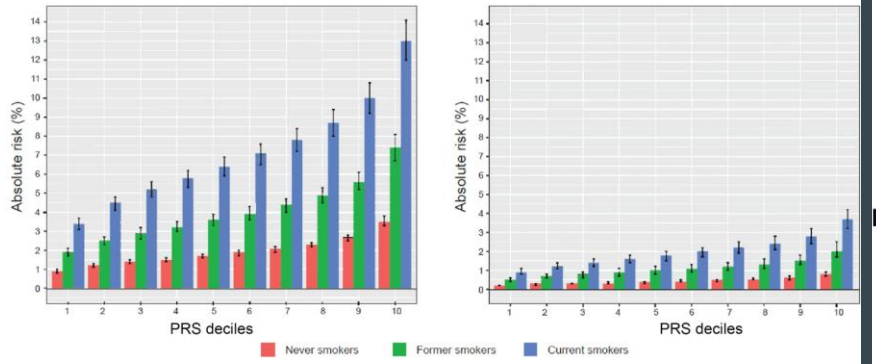
White non-Hispanic males
Absolute risk, %

	Never smoker	Former smoker	Current smoker
	Average risk, 95% CI	Average risk, 95% CI	Average risk, 95% CI
Overall	1.9 (1.8, 2.0)	4.0 (3.7, 4.4)	7.1 (6.6, 7.7)
PRS, decile 1	0.9 (0.8, 1.0)	1.9 (1.7, 2.1)	3.4 (3.1, 3.7)
PRS, decile 2	1.2 (1.1, 1.3)	2.5 (2.3, 2.7)	4.5 (4.1, 4.6)
PRS, decile 3	1.4 (1.3, 1.5)	2.9 (2.6, 3.2)	5.2 (4.8, 5.6)
PRS, decile 4	1.5 (1.4, 1.6)	3.2 (3.0, 3.5)	5.8 (5.3, 6.2)
PRS, decile 5	1.7 (1.6, 1.8)	3.6 (3.3, 3.9)	6.4 (5.9, 6.9)
PRS, decile 6	1.9 (1.7, 2.0)	3.9 (3.6, 4.3)	7.1 (6.5, 7.6)
PRS, decile 7	2.1 (1.9, 2.2)	4.4 (4.0, 4.7)	7.8 (7.2, 8.4)
PRS, decile 8	2.3 (2.2, 2.4)	4.9 (4.5, 5.3)	8.7 (8.0, 9.4)
PRS, decile 9	2.7 (2.5, 2.8)	5.6 (5.2, 6.1)	10 (9.2, 10.8)
PRS, decile 10	3.5 (3.3, 3.8)	7.4 (6.7, 8.1)	13 (12.0, 14.1)
Top 5%	4.0 (3.7, 4.2)	8.3 (7.5, 9.1)	14.6 (13.4, 15.8)
Top 1%	5.0 (4.6, 5.4)	10.3 (9.4, 11.4)	18.0 (16.4, 19.8)

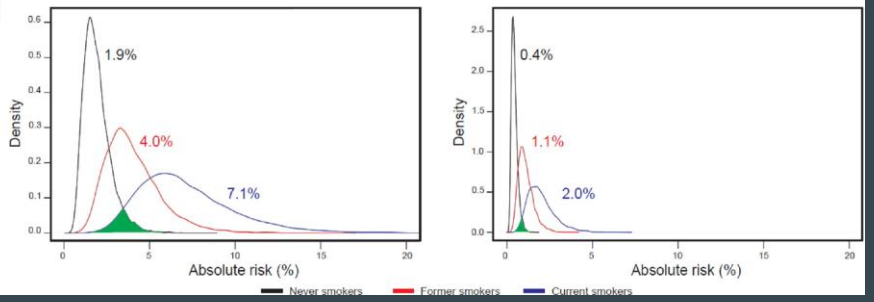
White non-Hispanic females
Absolute risk, %

	Never smoker	Former smoker	Current smoker
	Average risk, 95% CI	Average risk, 95% CI	Average risk, 95% CI
Overall	0.4 (0.4, 0.5)	1.1 (0.9, 1.3)	2.0 (1.7, 2.3)
PRS, decile 1	0.2 (0.2, 0.2)	0.5 (0.4, 0.6)	0.9 (0.8, 1.1)
PRS, decile 2	0.3 (0.2, 0.3)	0.7 (0.6, 0.8)	1.2 (1.1, 1.4)
PRS, decile 3	0.3 (0.3, 0.3)	0.8 (0.6, 0.9)	1.4 (1.2, 1.6)
PRS, decile 4	0.3 (0.3, 0.4)	0.9 (0.7, 1.1)	1.6 (1.4, 1.8)
PRS, decile 5	0.4 (0.3, 0.4)	1.0 (0.8, 1.2)	1.8 (1.5, 2.0)
PRS, decile 6	0.4 (0.4, 0.5)	1.1 (0.9, 1.3)	2.0 (1.7, 2.2)
PRS, decile 7	0.5 (0.4, 0.5)	1.2 (1.0, 1.4)	2.2 (1.9, 2.5)
PRS, decile 8	0.5 (0.5, 0.6)	1.3 (1.1, 1.6)	2.4 (2.1, 2.8)
PRS, decile 9	0.6 (0.5, 0.7)	1.5 (1.3, 1.8)	2.8 (2.4, 3.2)
PRS, decile 10	0.8 (0.7, 0.9)	2.0 (1.7, 2.5)	3.7 (3.2, 4.2)
Top 5%	0.9 (0.8, 1.0)	2.3 (1.9, 2.8)	4.2 (3.6, 4.8)
Top 1%	1.2 (1.0, 1.3)	2.9 (4.5, 6.0)	5.3 (4.5, 6.0)

B



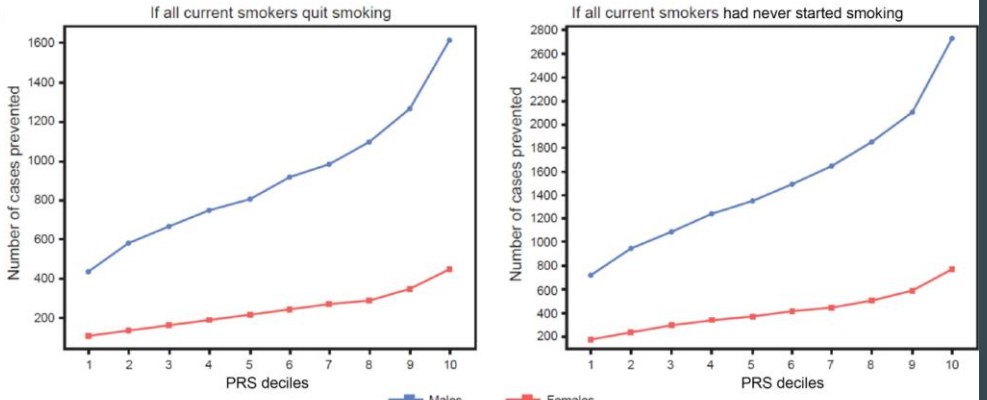
C



A

Population of 50-yr-old individuals	Males	Females
Resident population (census estimate, July 1, 2017), *n*	1 990 382	2 045 713
Non-Hispanic Whites (National Health Interview Survey 2015-2017), %	65.7	62.4
Resident population: non-Hispanic Whites, *n*	1 307 681	1 276 525
Never smokers, *n* (%) Average risk, % Number of cases	623 764 (47.7) 1.9 11 852	741 661 (58.3) 0.4 2,967
Former smokers, *n* (%) Average risk, % Number of cases	396 227 (30.3) 4.0 15 849	270 623 (21.2) 1.1 2,977
Current smokers, *n* (%) Average risk, % Number of cases	287 690 (22.0) 7.1 20 426	264 241 (20.7) 2.0 5,285

B



C

Decile	1	2	3	4	5	6	7	8	9	10
Males	431	576	662	748	805	921	978	1093	1266	1611
Females	106	132	159	185	212	237	264	290	344	450

Decile	1	2	3	4	5	6	7	8	9	10
Males	719	950	1093	1237	1352	1496	1640	1841	2100	2733
Females	185	238	291	344	370	422	449	502	581	767

Koutros*, ..., López de Maturana* et al, 2023

Polygenic Risk Scores - Examples

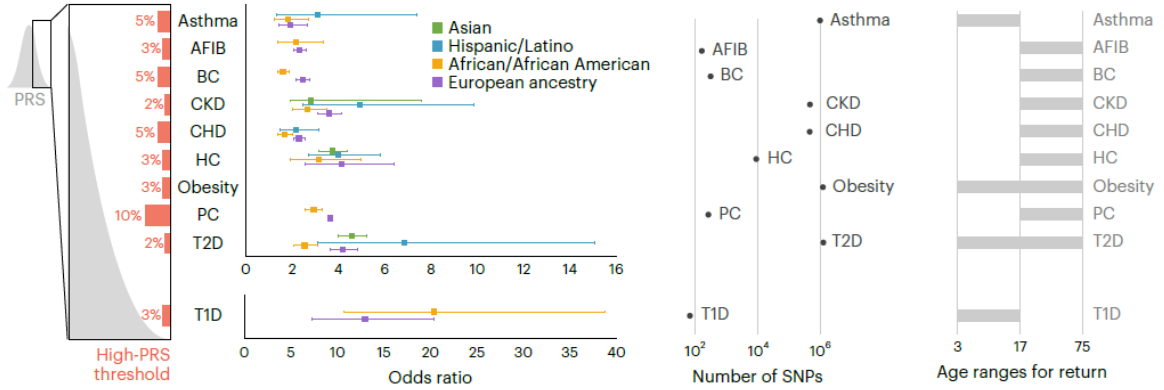


Fig. 2 | Summary of the ten conditions that were implemented. 'High-PRS threshold' represents the percentile that is deemed to be the cutoff for a specific condition above which a high-PRS result is reported for that condition. Odds ratios are reported as the mean odds ratios (square dot) associated with having a score above the specified threshold, compared to having a score below the specified threshold, along with 95% confidence intervals (CIs), shown in the whiskers. The number of case and control samples used to derive these odds ratios and CIs for each condition can be found in Supplementary Table 2. Note that the odds ratio for obesity is not reported here, as it will be published by the Genetic Investigation of ANthropometric Traits consortium (Smit et al., manuscript in preparation). 'Number of SNPs' represents the range of numbers or sites included in each score. 'Age ranges for return' indicates the participant ages at which a PRS is calculated for a given condition. AFIB, atrial fibrillation; BC, breast cancer; CKD, chronic kidney disease; CHD, coronary heart disease; HC, hypercholesterolemia; PC, prostate cancer; T1D, type 1 diabetes; T2D, type 2 diabetes.

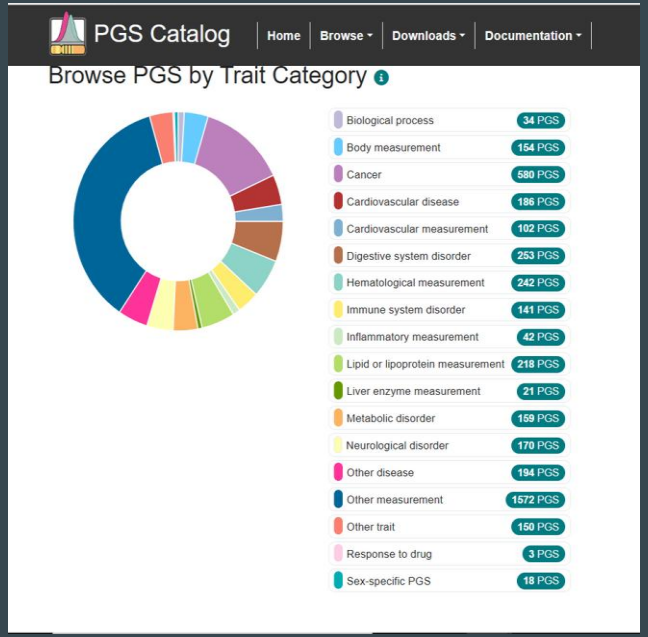
Lennon et al, 2024

Polygenic Risk Scores - Resources

An open database of polygenic scores and the relevant metadata

- PGS Catalog - The Polygenic Score Catalog

- https://pgsc-calc.readthedocs.io/en/latest/



Polygenic Risk Scores - Resources

Polygenic Risk Score software for calculating, applying, evaluating and plotting the results of polygenic risk scores (PRS) analyses (https://choishingwan.github.io/PRS-Tutorial/cal_prs/)

- Different methods (PLINK, PRSice-2, LDPred-2 and lassosum) with tutorials
- It handles both genotyped and imputed data
- it provides empirical association P-values free from inflation due to overfitting
- It supports different inheritance models
- it can evaluate multiple continuous and binary target traits simultaneously

PRS applications

PRS likely to be used in the near future due to:

- data sharing restrictions to individual-level data;
- heterogeneity across cohorts
- the largest sources of individual-level data—population cohorts, such as the UK Biobank—generally have relatively few individuals with specific diseases compared to dedicated case/control studies, for which there is typically only summary statistic data

Genetics in Medicine (2023) 25, 100803

ELSEVIER

Genetics in Medicine
An Official Journal of the ACMG
www.journals.elsevier.com/genetics-in-medicine

ACMG STATEMENT

The clinical application of polygenic risk scores: A points to consider statement of the American College of Medical Genetics and Genomics (ACMG)

Aya Abu-El-Haija[1,2], Honey V. Reddi[3], Hannah Wand[4], Nancy C. Rose[5], Mari Mori[6,7], Emily Qian[8], Michael F. Murray[9]; on behalf of the ACMG Professional Practice and Guidelines Committee[10,*]

Table 1 Clinical application of polygenic risk score

Point number	Points to consider
1	PRS test results do not provide a diagnosis, instead they provide a statistical prediction of increased clinical risk.
2	A low PRS does not rule out significant risk for the disease or condition in question.
3	If the risk prediction of a PRS is derived from a population that is different from the patient being tested, then the results may have a poor predictive value for the patient.
4	Isolated PRS testing is not the appropriate test for clinical scenarios in which monogenic etiology is known or suspected.
5	Before testing, a patient and provider should discuss the indications for the PRS test, and the patient should be informed how the PRS results will be used to guide medical management.
6	PRS-based medical management should be evidence-based; however, there is currently limited evidence to support the use of PRS to guide medical management.
7	Clinical follow-up for PRS should be consistent with best practices outlined by professional societies with appropriate expertise in instances when and where evidence-based practice guidelines exist.
8	The ACMG's position is that preimplantation PRS testing is not yet appropriate for clinical use and should not be offered at this time.[14]

ACMG, American College of Medical Genetics and Genomics; *PRS*, polygenic risk score.

PRS implications

PRSs have been shown to have some potential in disease risk identification, drug targeting and stratified medicine across a range of therapeutic areas including oncology, cardiovascular and psychiatry

The future of polygenic risk scores looks cautiously optimistic

Analysis

Heritable polygenic editing: the next frontier in genomic medicine?

https://doi.org/10.1038/s41586-024-08300-4

Received: 30 June 2022

Accepted: 29 October 2024

Published online: 8 January 2025

Open access

Check for updates

Peter M. Visscher[1,2]✉, Christopher Gyngell[3,4], Loic Yengo[1] & Julian Savulescu[3,5,6]✉

Polygenic genome editing in human embryos and germ cells is predicted to become feasible in the next three decades. Several recent books and academic papers have outlined the ethical concerns raised by germline genome editing and the opportunities that it may present[1–3]. To date, no attempts have been made to predict the consequences of altering specific variants associated with polygenic diseases. In this Analysis, we show that polygenic genome editing could theoretically yield extreme reductions in disease susceptibility. For example, editing a relatively small number of genomic variants could make a substantial difference to an individual's risk of developing coronary artery disease, Alzheimer's disease, major depressive disorder, diabetes and schizophrenia. Similarly, large changes in risk factors, such as low-density lipoprotein cholesterol and blood pressure, could, in theory, be achieved by polygenic editing. Although heritable polygenic editing (HPE) is still speculative, we completed calculations to discuss the underlying ethical issues. Our modelling demonstrates how the putatively positive consequences of gene editing at an individual level may deepen health inequalities. Further, as single or multiple gene variants can increase the risk of some diseases while decreasing that of others, HPE raises ethical challenges related to pleiotropy and genetic diversity. We conclude by arguing for a collectivist perspective on the ethical issues raised by HPE, which accounts for its effects on individuals, their families, communities and society[4].

Nature, vol 637, pages 637–647 (2025)

Genome-wide prediction

Genome editing

Embryo editing for disease is unsafe and unproven

Shai Carmi, Henry T. Greely & Kevin J. Mitchell

Mathematical modelling suggests that it is theoretically possible to reduce risk of common diseases using heritable genome editing. Scientists argue that the technology involves considerable risk and uncertain benefits. **See p.637**

54-556 (2025)

We need to talk about human genome editing

In a few decades, gene-editing technologies could reduce the likelihood of common human diseases. Societies must use this time to prepare for their arrival.

Genome Editing and Eugenics

The one hundred and third Take:

GREG GIBSON
JAN 16, 2025

6 1 1 Share

I've been pondering a reboot of Genome's Take for a while, but find myself kicked into action by the appearance of a disturbing eugenic article on human genome editing, "Heritable polygenic editing: the next frontier in genomic medicine?" this week (January 8, 2025) in Nature. That this has met with barely a whimper on social media is astonishing, so I feel the need to comment. I do so with a heavy heart, because nobody likes to be completely at odds with good friends and collaborators. I am in awe of the contributions of the two authors I know, Peter Visscher and Loic Yengo, and learn

https://genomestake.substack.com/p/genome-editing-and-eugenics

Genome-wide prediction

Time to discuss on the polygenic editing of embryos

- Visscher et al (2025) describe a mathematical model that estimates that handful editings can lead to a dramatic reduced risk of various disorders:
  - **Assumptions**: theoretical scenario where large-scale genome editing is **feasible** and **safe**
  - Selection of embryos based on PRSs is already a possibility offered by fertility clinics in USA since 2019 → polygenic editing goes even beyond:
    - Successful for rare conditions but in somatic tissue and not in reproductive cells or embryos (heritable)
    - Heritable editing may lead to off-target effects and unpredictable negative consequences

Nature **637**, 554-556 (2025)

Time to discuss on the polygenic editing of embryos

- Main claim is that disease risk can be reduced by introducing into the genomes of embryos ‘rare protective alleles’ — genetic variants that are uncommon in the population but are thought to protect against disease
  - This is, in fact, the strategy taken by He Jiankui in his attempts to grant embryos some protection against HIV

Nature **637**, 554-556 (2025)

Time to discuss on the polygenic editing of embryos

Time to discuss on the polygenic editing of embryos

- The model proposes that editing only ten rare protective alleles per disease is expected to lead to dramatic reductions in risk, between 2-fold and 60-fold, but…
  - Mathematical models are only as good as the assumptions they are based on, and this model depends on:
    1. Genome-editing techniques will be able to modify DNA with **perfect** accuracy:
       - This is not the case, yet
    2. The success of the proposed approach relies on accurate identification of genetic variants that have a causal effect:
       - GWAS studies do not necessarily identify causal variants, but common variants in LD with the causal ones

Nature **637**, 554-556 (2025)

Time to discuss on the polygenic editing of embryos

3. Visscher et al. assume that the protective effects of different variants are independent and will add up
   - This is a common assumption made at population level
   - If two protective variants belong to the same pathway→ can we assume that their effects sum up?
4. GxE: are the variants protective for a given trait in an environment also protective in every environment/exposure?
   - What if in the future an exposure disappears? Introducing a protective variant in an environment that no longer exists may not be useful

Nature **637**, 554-556 (2025)

Time to discuss on the polygenic editing of embryos

5. the model assumes that the degree of risk reduction would be the same across the relevant population, but the genetic reasons may be different
   - Some embryos might already be at low risk of disease because they are carrying target protective alleles
   - Others might be at low or high risk because of their burden of common variants.

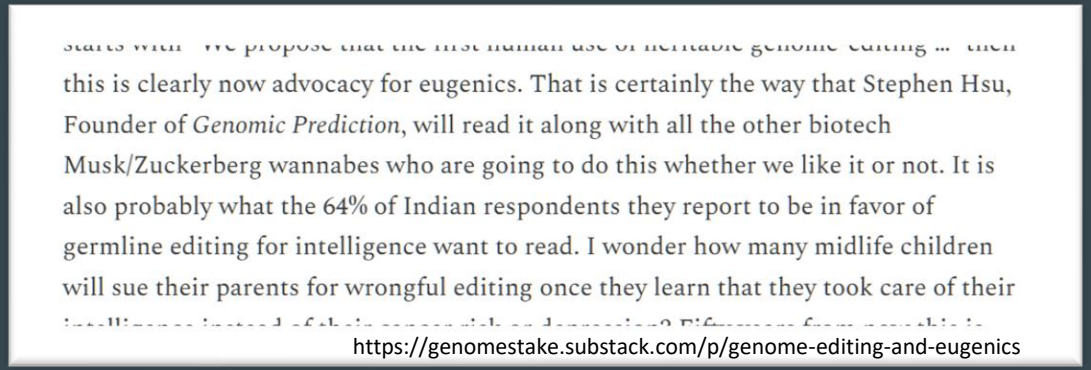
Time to discuss on the polygenic editing of embryos

- Other issues:
  - safety is far from guaranteed
  - risks to children must be taken especially seriously:
    - Unexpected outcomes
      - early use of genetic testing of embryos to screen for chromosomal abnormalities before implantation inadvertently worsened IVF outcomes (Yan et al, 2021)
      - Unlike somatic editing, any errors will affect every cell and organ in the future child:
        - misidentified causal alleles
        - Pleiotropy
        - Rare alleles may be rare for a reason
        - Unpredictable interactions arising from new combinations of variants

Nature 637: 554-556 (2025)

Time to discuss on the polygenic editing of embryos

- Ethical issues:
  - The authors calculate that this technology will be so low risk and effective in every individual that deploying it on a large scale might be justified, even for individuals with low absolute risk of disease ¿?
  - Eugenics: unedited genomes are intrinsically worth less than edited ones, it creates hierarchies of worth ¿?



starts with "We propose that the first human use of heritable genome editing ..." then this is clearly now advocacy for eugenics. That is certainly the way that Stephen Hsu, Founder of *Genomic Prediction*, will read it along with all the other biotech Musk/Zuckerberg wannabes who are going to do this whether we like it or not. It is also probably what the 64% of Indian respondents they report to be in favor of germline editing for intelligence want to read. I wonder how many midlife children will sue their parents for wrongful editing once they learn that they took care of their intelligence instead of their cancer risk or depression? Fifty years from now this is

https://genomestake.substack.com/p/genome-editing-and-eugenics

Genome-wide Prediction in Animal and Plant breeding

Quantitative genetics is a cornerstone of both plant and animal breeding for the last century

Genomic selection has led to the re-emergence of quantitative genetics as a framework for incorporating marker and sequence information to supplement and complement standard phenotypic descriptors and pedigree information (Hickey et al 2017)

Access to large-scale sequence and phenotypic information would provide opportunities to unify breeding methods and tools across several plant and animal species → Modernization of breeding programs (Hickey et al 2017)

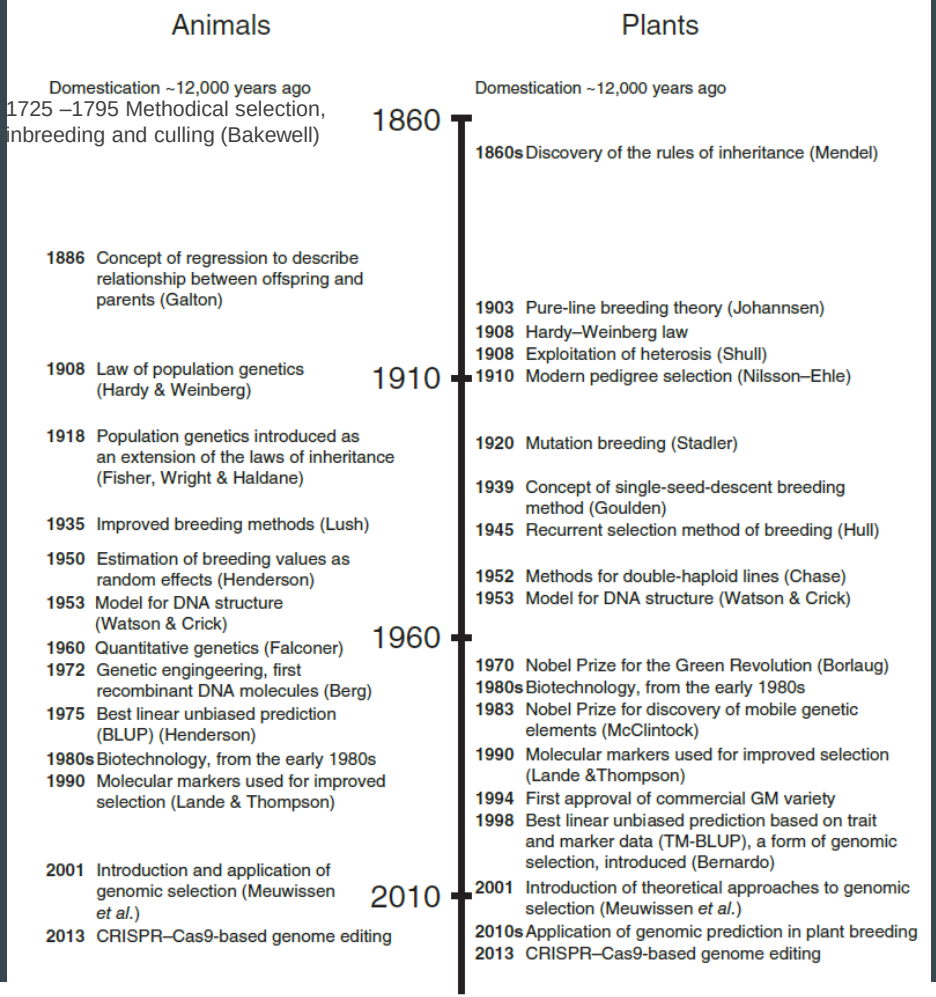


Figure 1 Some key milestones of selective animal and plant breeding.

(Hickey et al 2017)

Differences in plant and animal breeding

Although there are conceptual similarities between animal and plant breeding, breeding methods have diverged:

- Species specific characteristics: reproduction mode, # of progeny per cycle ...
- Plants: breeding since domestication; consisted mainly in selection (need for hybridization recognized in the last 250 years (Kingsbury, 2009))
- Animals: a more structured approach adopted earlier than in plants (Bakewell, 18 century → herdbooks)

Although both plant and animal breeders deal with complex traits, individual mutations with moderate to large effects have been exploited more importantly in plant breeding than in animal breeding (Hickey et al, 2017)

Differences in plant and animal breeding

Other differences:

- Plant breeders used well-designed trials to measure phenotype to inform their selection, and animal breeders use complex statistical methods to estimate breeding value
- Animal breeders use information from the relatives of selection candidates (milk yield in bulls), with low heritability or measured late in the breeding process (longevity); Plant breeders don't have the problem of 'sex-limited' traits and could increase the accuracy growing more plants from the same cultivar

Genomic selection in plant and animal species

GS was adopted rapidly in the more technologically developed livestock sectors (dairy cattle)

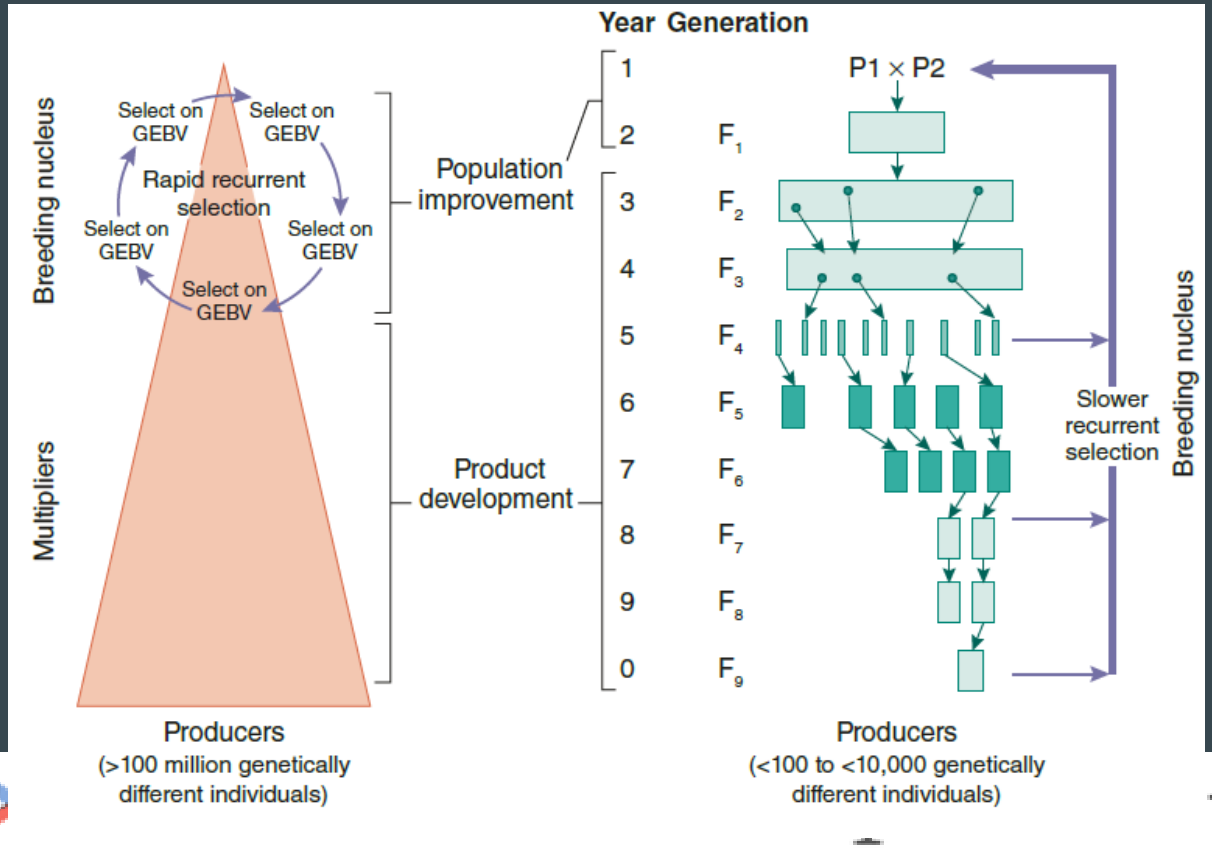
Major international seed companies are routinely using genomic selection

Many public-sector breeding programs are exploring this technology

Bottleneck:

- Computational and recording infrastructure
- Genotypic and phenotypic data to implement GS
- Complexity of genomes of many plant species

Comparison of plant (inbreeding cereal) and animal breeding approaches

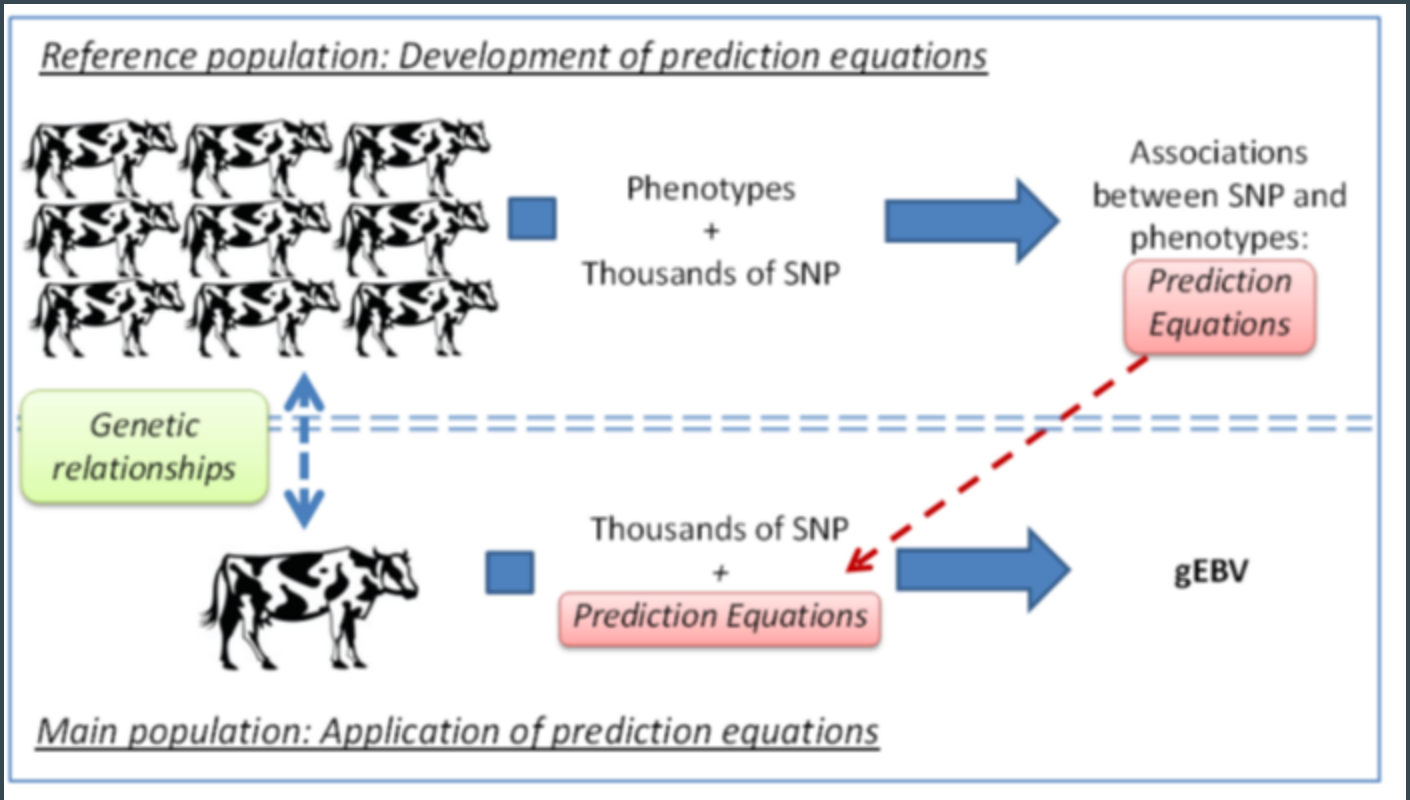


Hickey et al, 2017

Joining efforts between animal and plant breeders

Plant and animal breeders will benefit from working together to address problems that are common to the two disciplines, such as prediction of traits in structured populations (Schön and Simianer, 2015)

Examples of GS in animal and plant breeding



Groen Kennisnet (2017), Textbook animal breeding and genetics

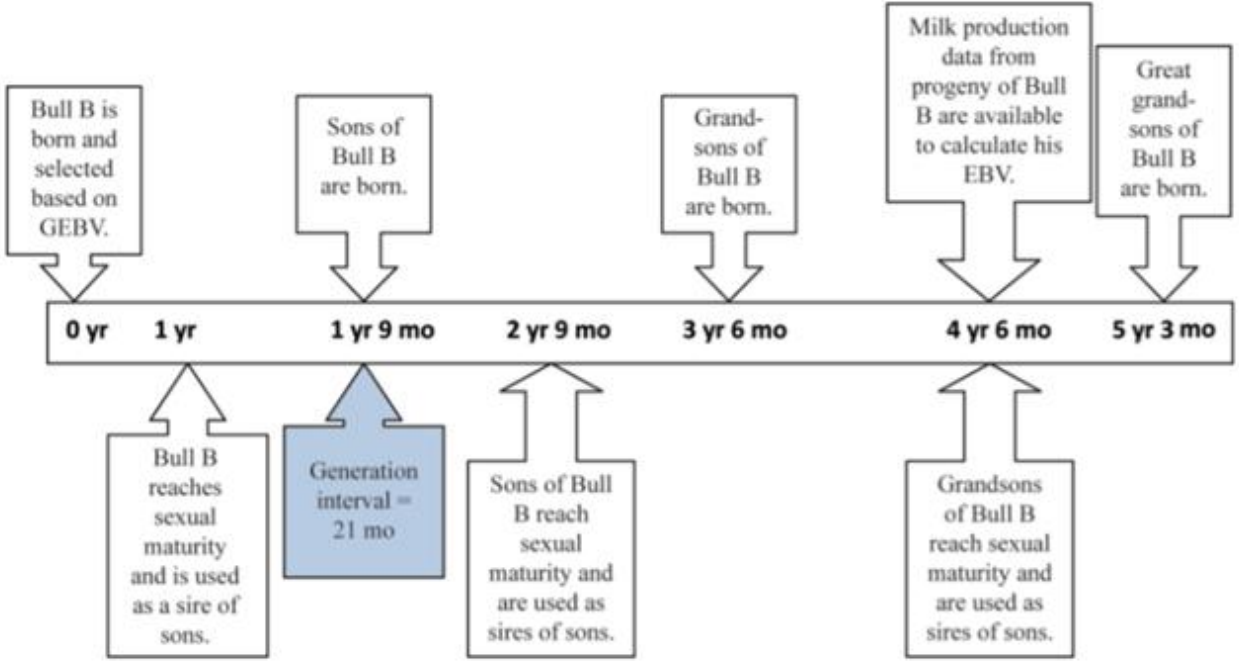
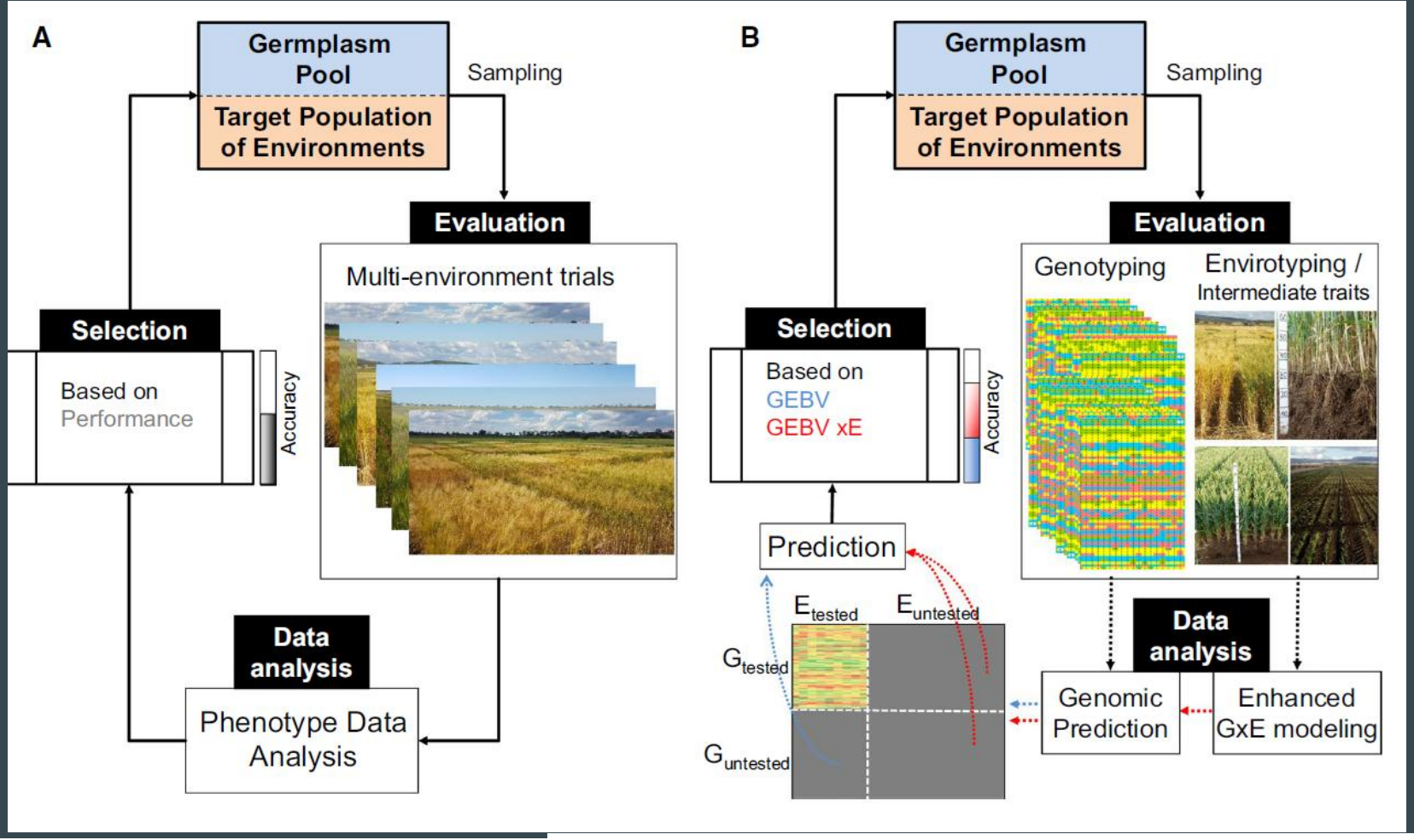


Figure 2. Timeline of an aggressive artificial insemination breeding program based on the use of genomic bulls as sires of sons. GEBV = genomic estimated breeding value; EBV = estimated breeding value.

Animal **Frontiers**

Schefers and Weigel 2012, Vol. 2, No. 1

Genome-wide prediction



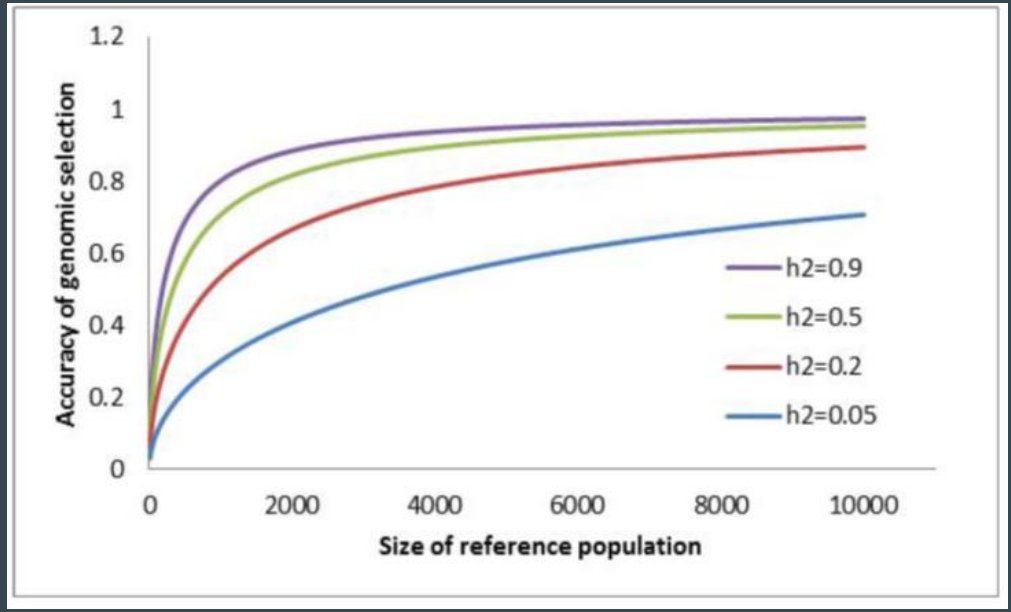
Voss-Fels et al, 2019 (https://doi.org/10.1007/s00122-018-3270-8)

Accuracy of Genomic predictions

- Quality of phenotypes in the training population
- Size of reference population
- Reference population must be updated (Poldich et al, 2004)
- Training and testing should be closely related

$$\sqrt{Nh^2/(Nh^2+M_e)}$$ Daetwyler et al, 2008

$$M_e = 2N_eL$$ Hayes et al, 2009b



Groen Kennisnet (2017), Textbook animal breeding and genetics

Similarities and differences in performance of PRS and GS

Both PRS and GEBV are estimates of the additive genetic value of a trait of an individual (Wray et al, 2019)

Higher proportion of genetic variance explained by SNPs in livestock than in humans is due to the greater LD in livestock

SNP-based heritability estimates in humans are lower than those in animal, due to differences in recent effective sample size:

- In animals, common SNPs tag causal variants at much greater physical distance, compared to in humans, and including across chromosomes

Similarities and differences in performance of PRS and GS

* Different in purpose:

- PRS: predict the future phenotype of an individual (efficacy depends on the SNP h2)
- GEBV: predict the average value of an animal's genetic material to its offspring

The use of summary statistic data for the genotype effect size estimates distinguishes PRS from phenotypic prediction approaches that exploit individual-level data only

In the latter, genotype effect sizes are usually estimated in joint models of multiple variants and prediction performed simultaneously, using approaches such as best linear unbiased prediction or (LASSO)

Similarities and differences in performance of PRS and GS

* Effect sizes estimation:

- PRS: usually, one SNPs at a time
- GEBV: all SNPS jointly fitted

Topics

GP in human genetics	Polygenic risk scores	GP in animal and plant breeding	Comparison between PRS and GS
Background	Background	Overview	Overview
Particularities	PRS analysis process	Comparison of plant and animal breeding approaches	
GWAS in human genetics	Accuracy	Accuracy	
	Limitations		
	Other considerations		
	Examples		
	Resources		
	PRS applications		