

Polygenic Risk Scores

Tinashe Chikowore, PhD

tinashe.Chikowore@wits.ac.za

Outline

Background of PRS

Key concepts of computing PRS using the Pruning and Thresholding Approach

- Independence of Discovery and Target data set
- Ambiguous SNPs
- Clumping distance and LD pruning
- Adjustment for population structure and study specific covariates
- Evaluation of Predictivity using R^2
- PRS prediction visualizations
- Applications of PRS

Other PRS approaches

Background

- During the first GWAS wave from 2004 to 2009
- Shaun Purcell suggested combining SNPs into a score
- Initially GWAS significant hits used
- Now many SNPs are now used in scores . A PRS score can have 2million SNPs

LETTERS

Common polygenic variation contributes to risk of schizophrenia and bipolar disorder

The International Schizophrenia Consortium*



Predictive utility of a genetic risk score of common variants associated with type 2 diabetes in a black South African population

Tinashe Chikowore ^{a,*}, Tertia van Zyl ^a, Edith J.M. Feskens ^b, Karin R. Conradie ^a

^a Centre for Excellence in Nutrition, North-West University, Potchefstroom, North West Province 2520, South Africa

^b Wageningen University, Division of Human Nutrition, P.O. Box 17, 6700 AA Wageningen, The Netherlands

2172

Diabetes Volume 63, June 2014

Jason L. Vassy,^{1,2,3} Marie-France Hivert,^{1,4,5} Bianca Porneala,⁶ Marco Dauriz,^{1,6,7} Jose C. Florez,^{1,8,9} Josée Dupuis,^{10,11} David S. Siscovick,¹² Myriam Fornage,¹³ Laura J. Rasmussen-Torvik,¹⁴ Claude Bouchard,¹⁵ and James B. Meigs^{1,6}

Polygenic Type 2 Diabetes Prediction at the Limit of Common Variant Detection



New test predicts the risk of serious disease months before symptoms appear

[Download PDF Copy](#)



By [Sally Robertson, BSc](#)

Aug 13 2018

Researchers from the [Broad Institute of MIT and Harvard](#) have developed a new genetic test that can identify people who are at an increased risk of serious disease, long before they develop any symptoms.

A Harvard Scientist Thinks He Has a Gene Test for Heart Attack Risk. He Wants to Give It Away Free.



Matthew Herper Forbes Staff
Healthcare

I cover science and medicine, and believe this is biology's century.



[bes.com/matthewherper/files/2018/08/sekar_kathiresan-](#)

Online attention



Altmetric score (what's this?)

- Tweeted by 900
- Blogged by 5
- On 10 Facebook pages
- Picked up by 113 news outlets

Clues to Your Health Are Hidden at 6.6 Million Spots in Your DNA

With a sophisticated new algorithm, scientists have found a way to forecast an individual's risks for five deadly diseases.



Venk Murthy MD PhD
[@venkmurthy](#)



Replies to [@ewanbirney](#) [@skathire](#)

Media-ites, perhaps also talk to [@cecilejanssens](#) or [@f2harrell](#) about the deep flaws in this line of thinking and the limitations rather than just blindly [#hyping](#).

[twitter.com/cecilejanssens...](#)

Cecile Janssens [@cecilejanssens](#)

My take home message from the latest polygenic risk paper: 6.6 million snps improve AUC by 0.015. In the derivation sample. It is hard to believe that a 6.6M SNP model performs noticeably better than a 74 SNP model. [nature.com/articles/s4158...](#)

N variants available / N variants in score (%)	Tuning parameters	AUC (95%CI)
74 / 74 (100%)	$p < 5 \times 10^{-8}$, $r^2 < 0.2$	0.791 (0.785 – 0.798)
105,942 / 105,595 (99.67%)	$p < 0.05$, $r^2 < 0.8$	0.799 (0.793 – 0.806)
6,629,369 / 6,630,150 (99.99%)	$p = 0.001$	0.806 (0.800 – 0.813)
55 / 55 (100%)	$p < 5 \times 10^{-8}$, $r^2 < 0.2$	0.766 (0.757 – 0.776)

Perspective | Published: 29 March 2019

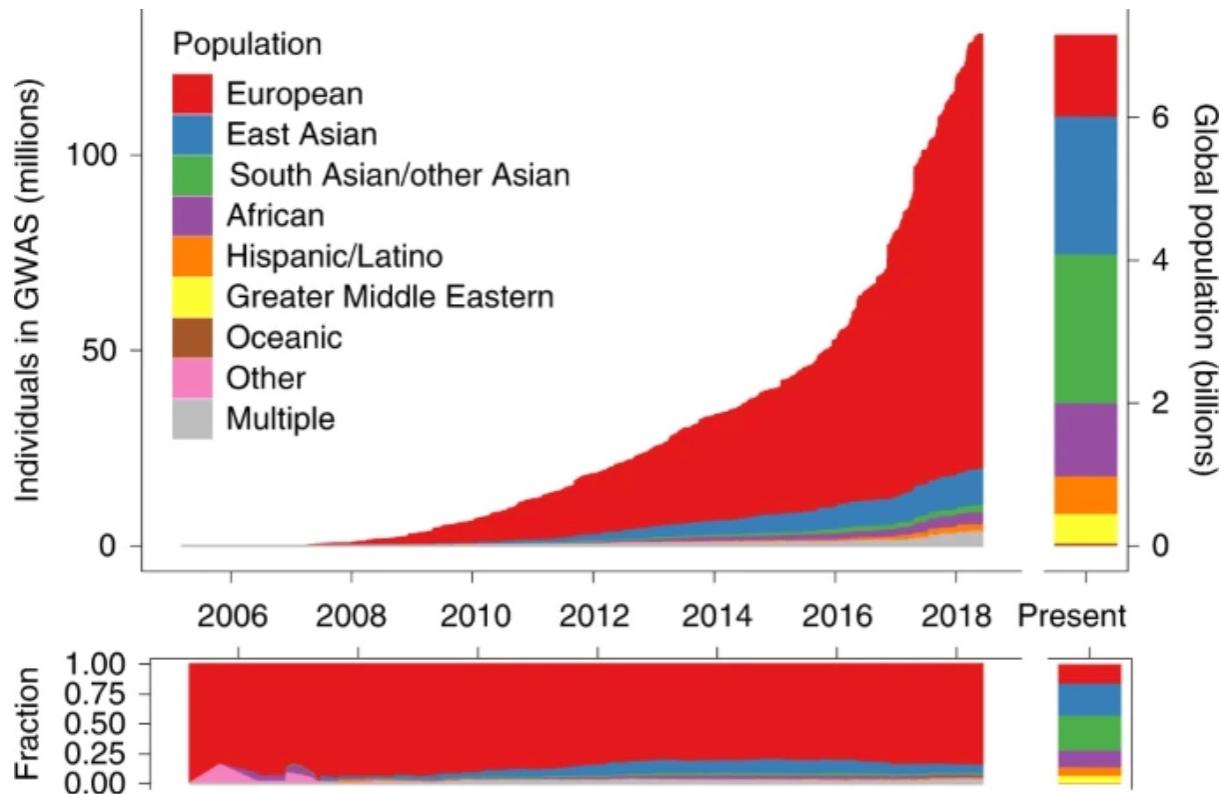
Clinical use of current polygenic risk scores may exacerbate health disparities

Alicia R. Martin , Masahiro Kanai, Yoichiro Kamatani, Yukinori Okada, Benjamin M. Neale & Mark J. Daly

Nature Genetics 51, 584–591(2019) | Cite this article

17k Accesses | 202 Citations | 577 Altmetric | Metrics

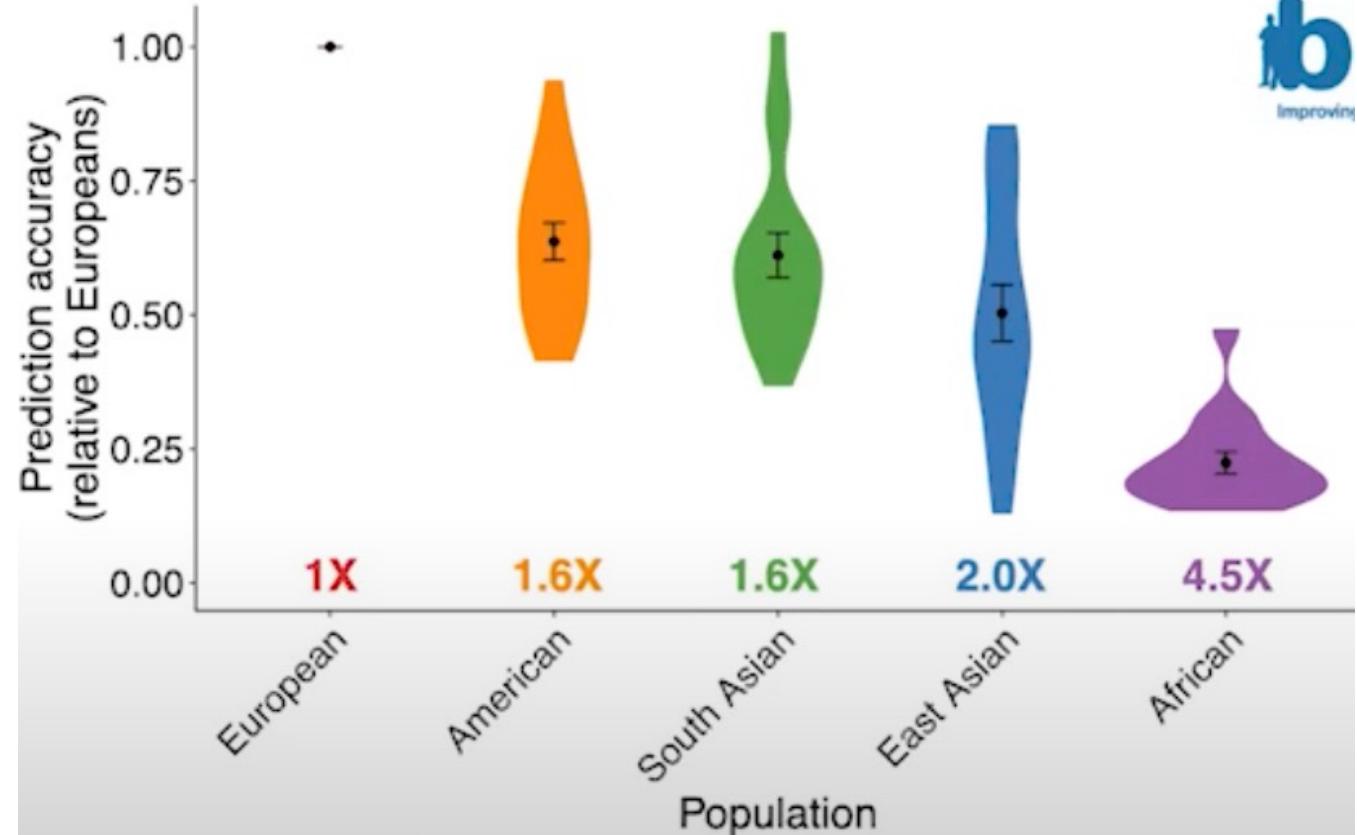
Fig. 1: Ancestry of GWAS participants over time, as compared with the global population.



Cumulative data, as reported by the GWAS catalog⁷⁶. Individuals whose ancestry is ‘not reported’ are not shown.

European derived PRS predict 2.0 fold less in East Asians

- Differences in allele frequency , LD patterns
- Differences in environmental factors that might interact with the genetic factors



Martin et al. 2019 *Nat. Genet* **51**, 584-591

Quest to improve PRS accuracy in diverse populations

- Many tools being developed to enhance accuracy

nature > nature genetics > articles > article

Article | Published: 05 May 2022

Improving polygenic prediction in ancestrally diverse populations

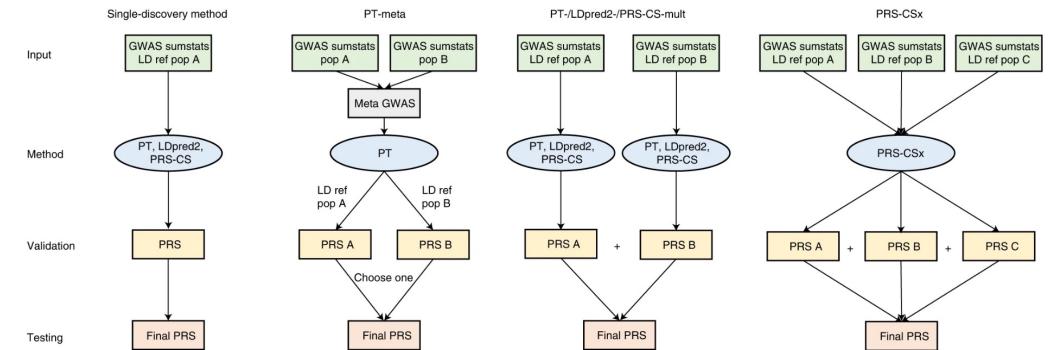
Yunfeng Ruan, Yen-Feng Lin, Yen-Chen Anne Feng, Chia-Yen Chen, Max Lam, Zhenglin Guo, Stanley Global Asia Initiatives, Lin He, Akira Sawa, Alicia R. Martin, Shengying Qin✉, Hailiang Huang✉ & Tian Ge✉

Nature Genetics 54, 573–580 (2022) | Cite this article

4199 Accesses | 1 Citations | 156 Altmetric | Metrics

Fig. 1: Overview of polygenic prediction methods.

From: [Improving polygenic prediction in ancestrally diverse populations](#)



The predictive performances of three representative single-discovery (PT, LDpred2 and PRS-CS) and five multi-discovery (PT-meta, PT-mult, LDpred2-mult, PRS-CS-mult and PRS-CSx) methods are compared in this study. LDpred2-mult and PRS-CS-mult depicted here are not published methods but are helpful for comparing potential improvements from PRS-CSx, which uses a coupled CS prior for the effect sizes of genetic variants. The discovery samples (to generate GWAS summary statistics (sumstats)), validation samples (to tune hyperparameters in PRS construction methods) and testing samples (to assess prediction accuracy) are nonoverlapping. LD ref, LD reference panel; pop A/B/C, Population A/B/C.

Comparisons

Fig. 2: Prediction accuracy of single-discovery and multi-discovery polygenic prediction methods in simulations.

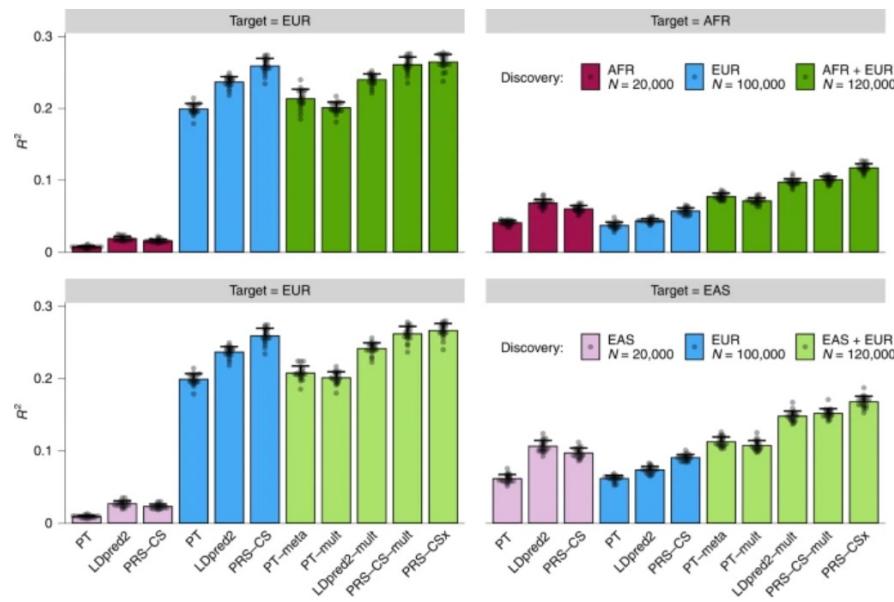
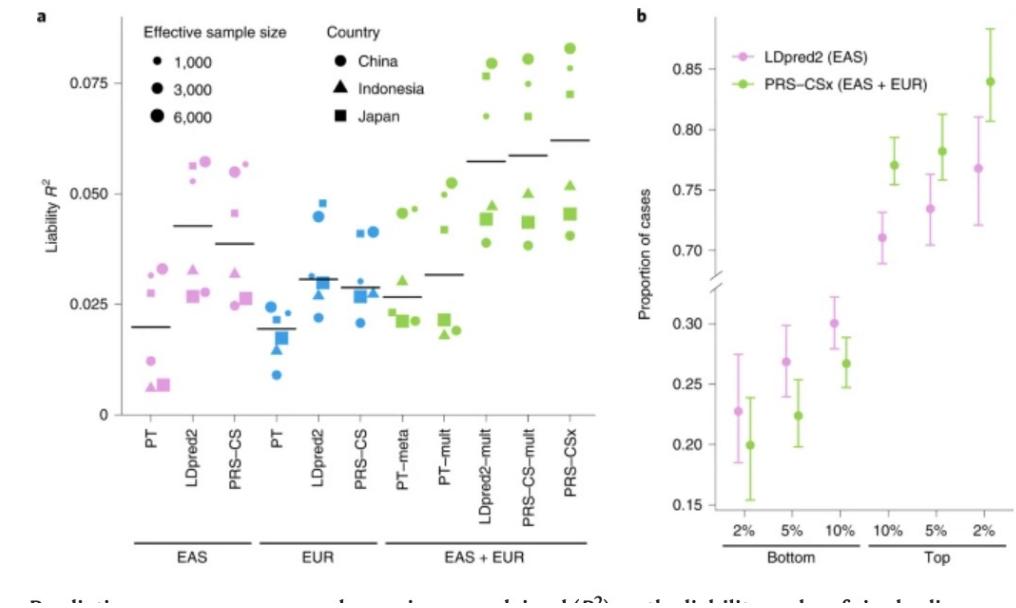
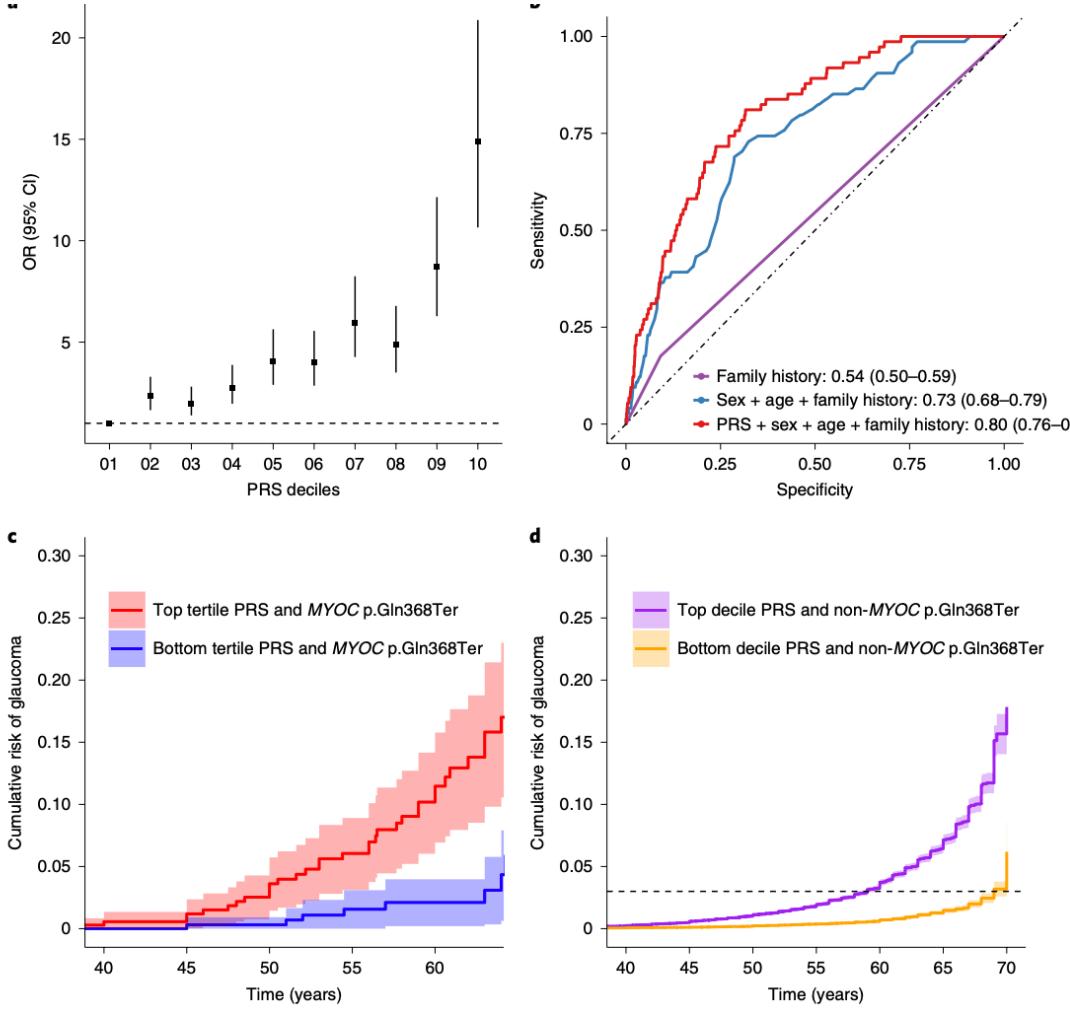


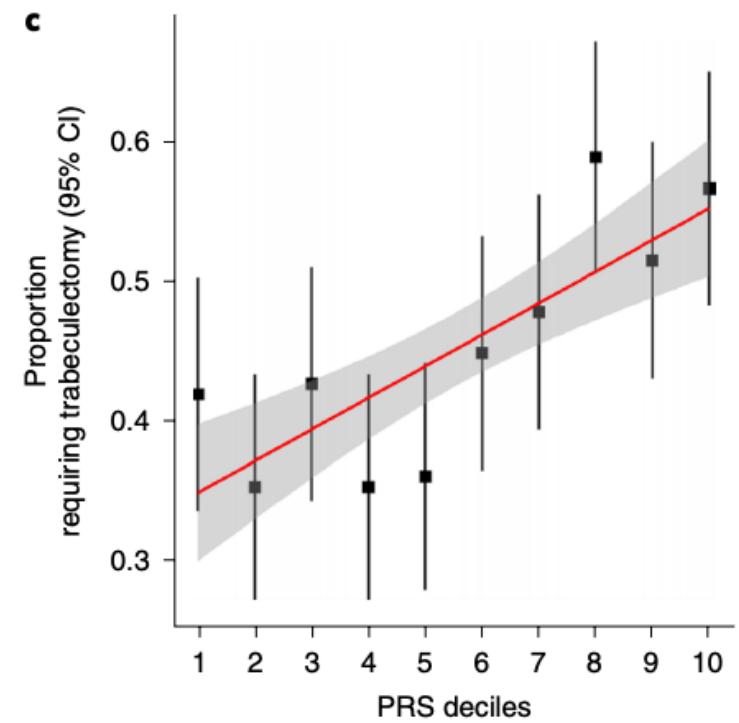
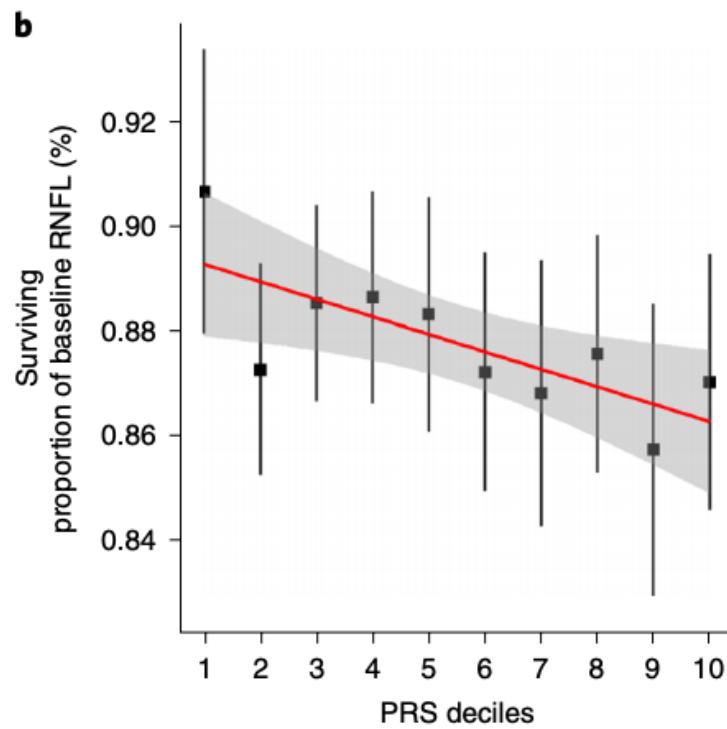
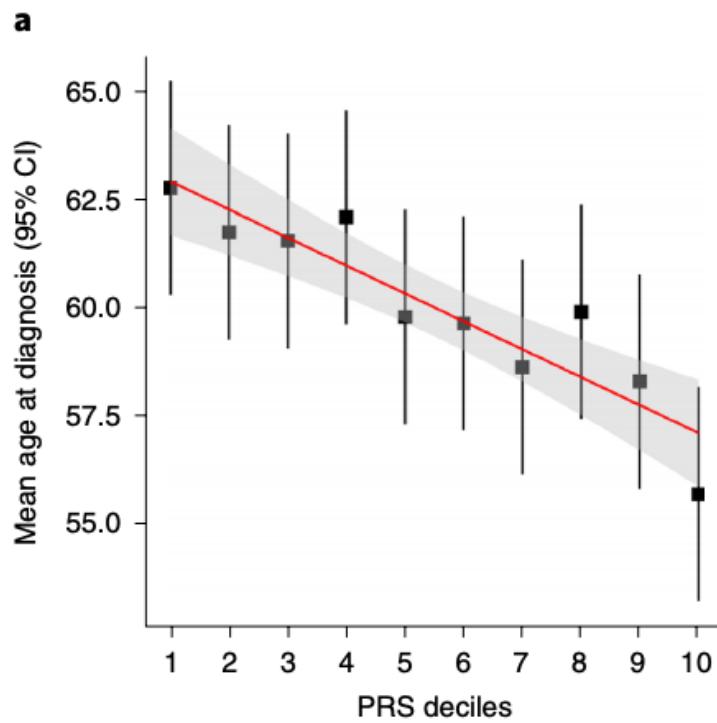
Fig. 4: Prediction accuracy of schizophrenia risk in EAS cohorts.



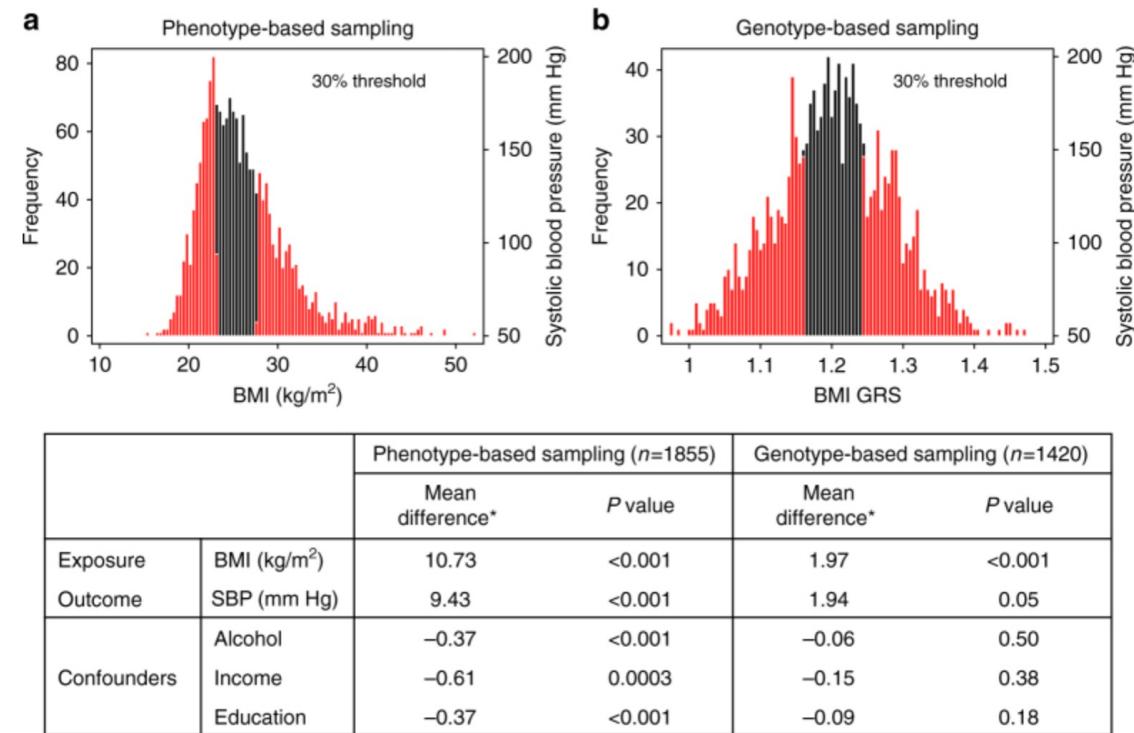
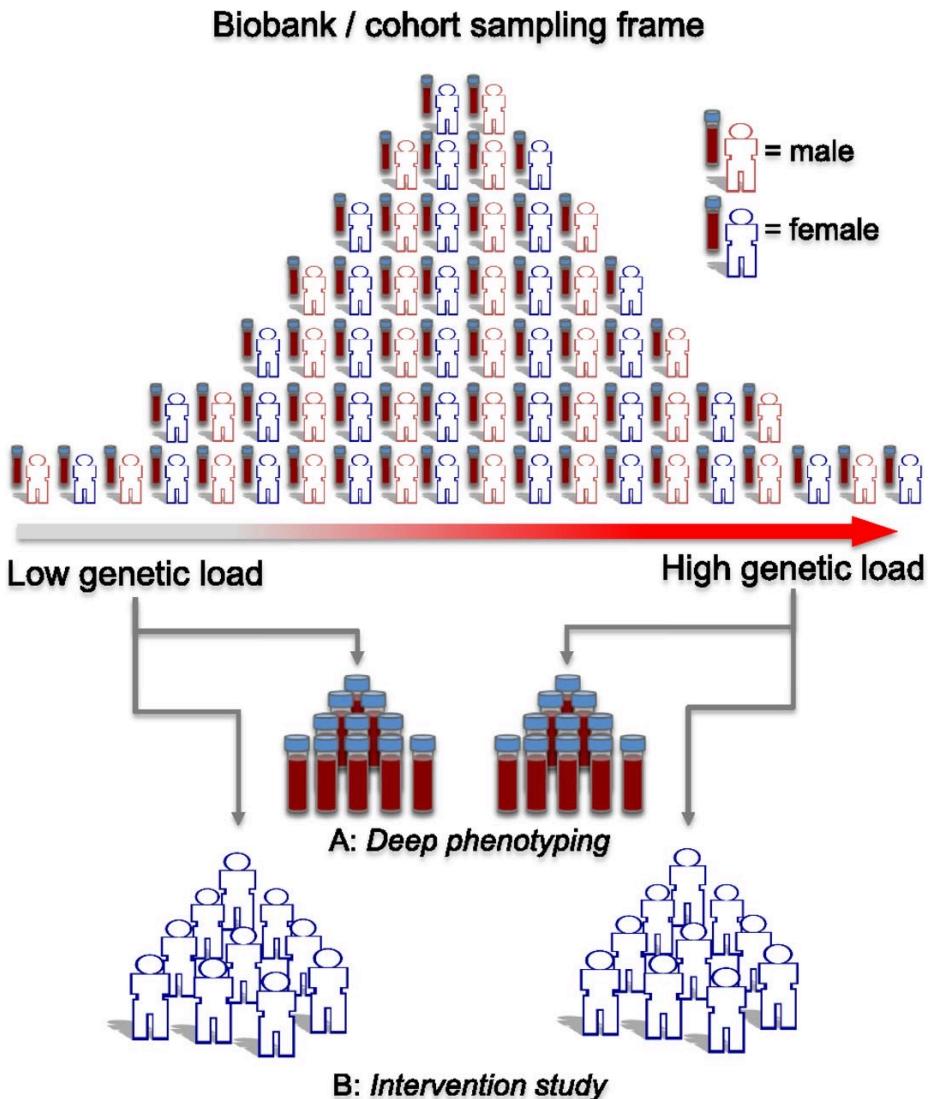
Predictivity and risk stratification of PRS



Clinical Utility of PRS

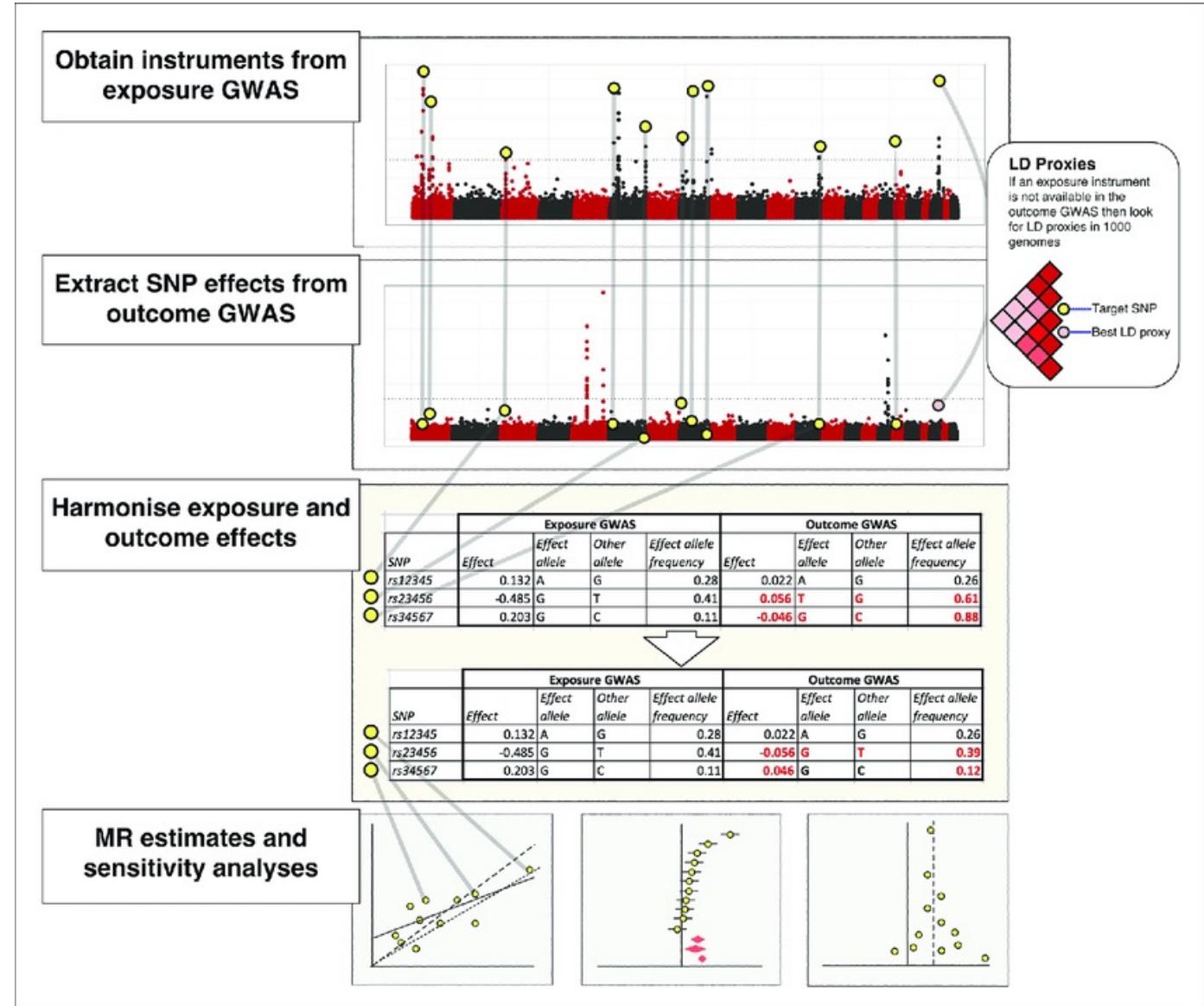


Recall by genotype study design controls for confounding and reverse causation and PRS



BMI, body mass index; SBP, systolic blood pressure; * difference in means between upper and lower groups

Well powered
GWAS
required for
better MR
instruments



Recall by genotype designs and statistical power

Open Access | Published: 25 November 2016

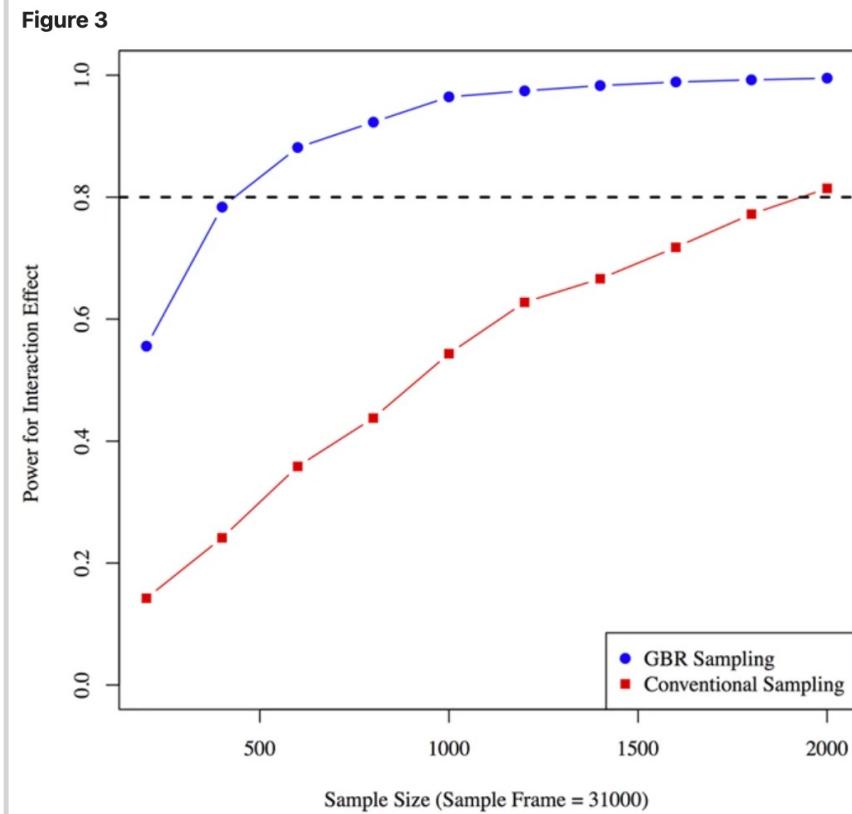
Statistical power considerations in genotype-based recall randomized controlled trials

Naeimeh Atabaki-Pasdar, Mattias Ohlsson, Dmitry Shungin, Azra Kurbasic, Erik Ingelsson, Ewan R. Pearson, Ashfaq Ali & Paul W. Franks

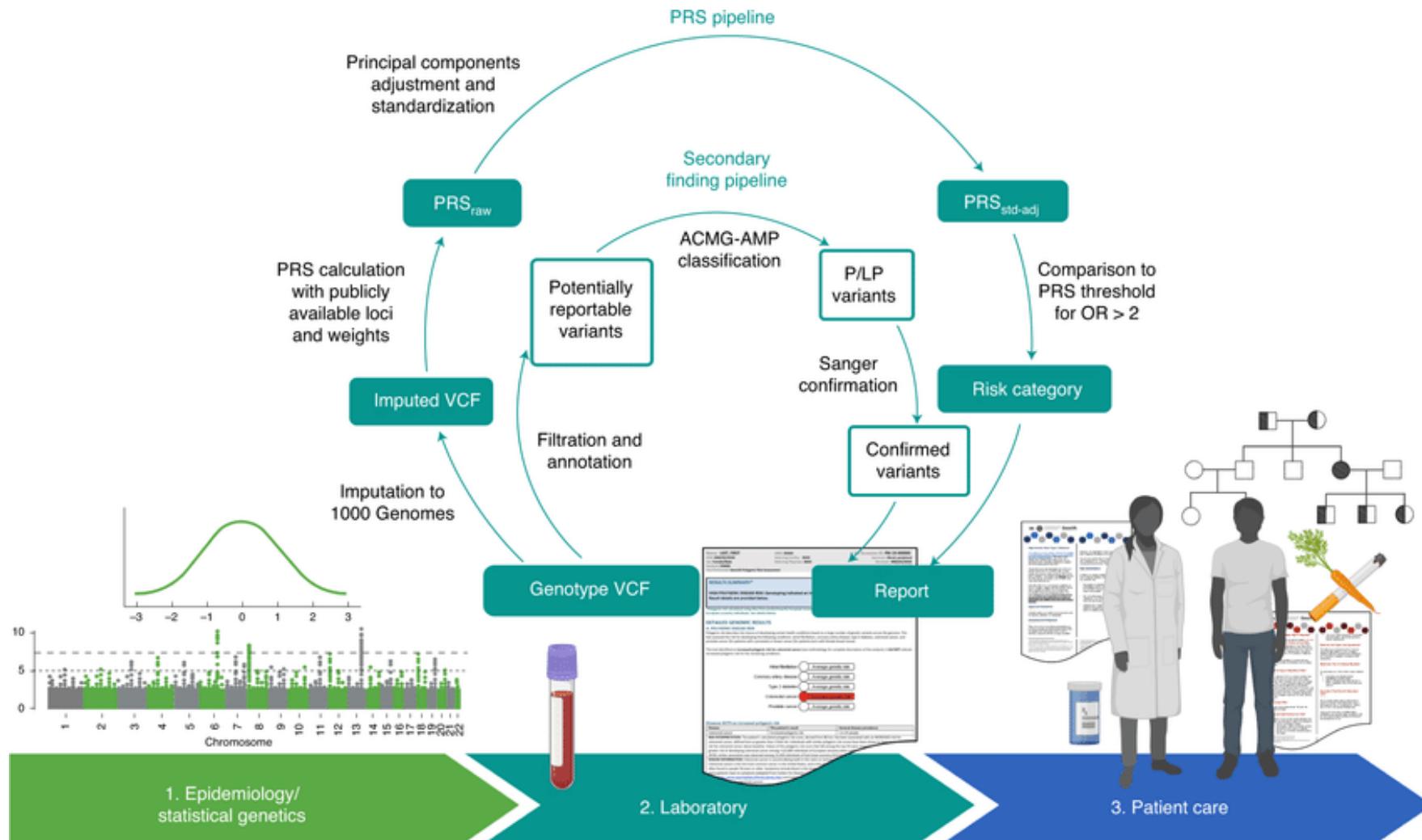
Scientific Reports 6, Article number: 37307 (2016) | [Cite this article](#)

1313 Accesses | 6 Citations | 6 Altmetric | [Metrics](#)

To obtain 80% power , ~400 participants required using the GBR vs ~1,900 with conventional recruitment in RCT



Statistical power to observe GRS × lifestyle interactions on 1-year small LDL particle levels (based on DPP clinical trial parameters).

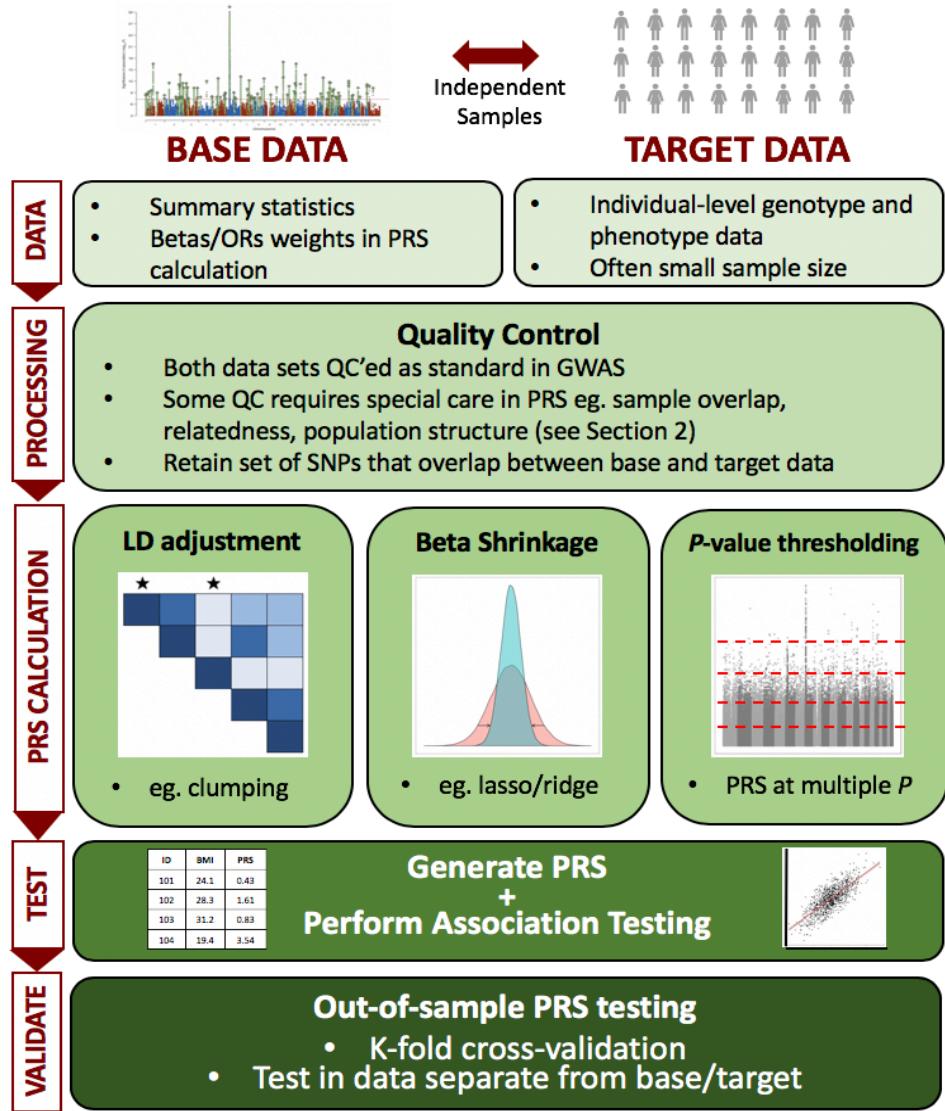


Computational methods are applied to GWAS data to develop and validate PRS in large populations

Laboratory develops analytically and clinically valid assay and pipeline to calculate, interpret, and report PRS for the individual

Physician contextualizes PRS results with the patient's other risk factors, comorbidities, and patient preferences, to make medical management decisions

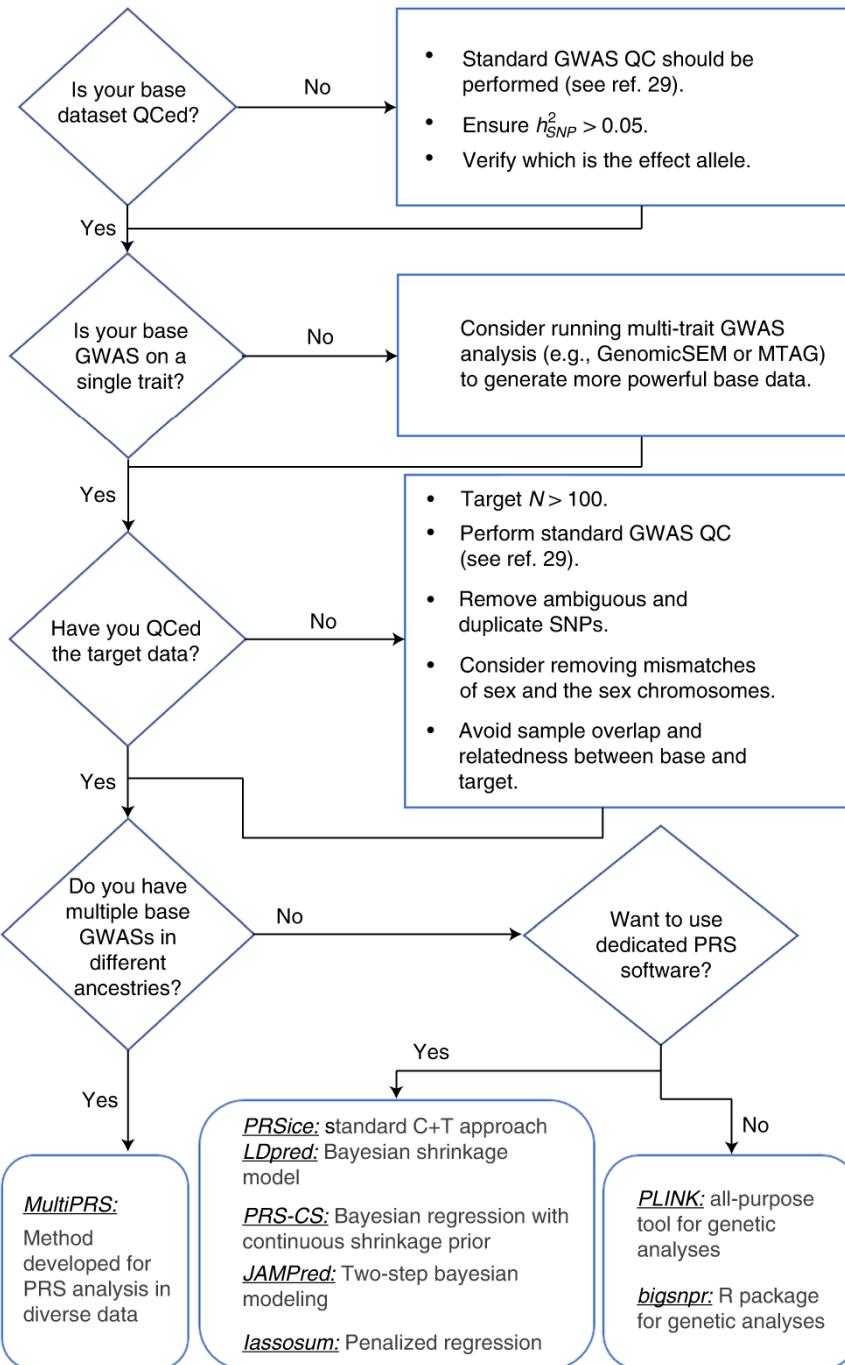
Overview of PRS computation



$$PRS_j = \sum [\beta_{i,\text{discovery}} * SNP_{ij}]$$

$\beta_{i,\text{discovery}}$ = effect size in discovery sample from (continuous trait) or logistic reg (binary trait; $\beta = \log(\text{OR})$)

SNP_{ij} = # alleles (0,1,2) for SNP i of person j in target sample



Tutorial: a guide to performing polygenic risk score analyses

Shing Wan Choi^{1,2}, Timothy Shin-Heng Mak¹ and Paul F. O'Reilly^{1,2}✉

- The target data set and the base data should be different
- To avoid strand misspecification ambiguous SNPs are removed
- Other methods include, GRABLD a machine learning approach