Outline

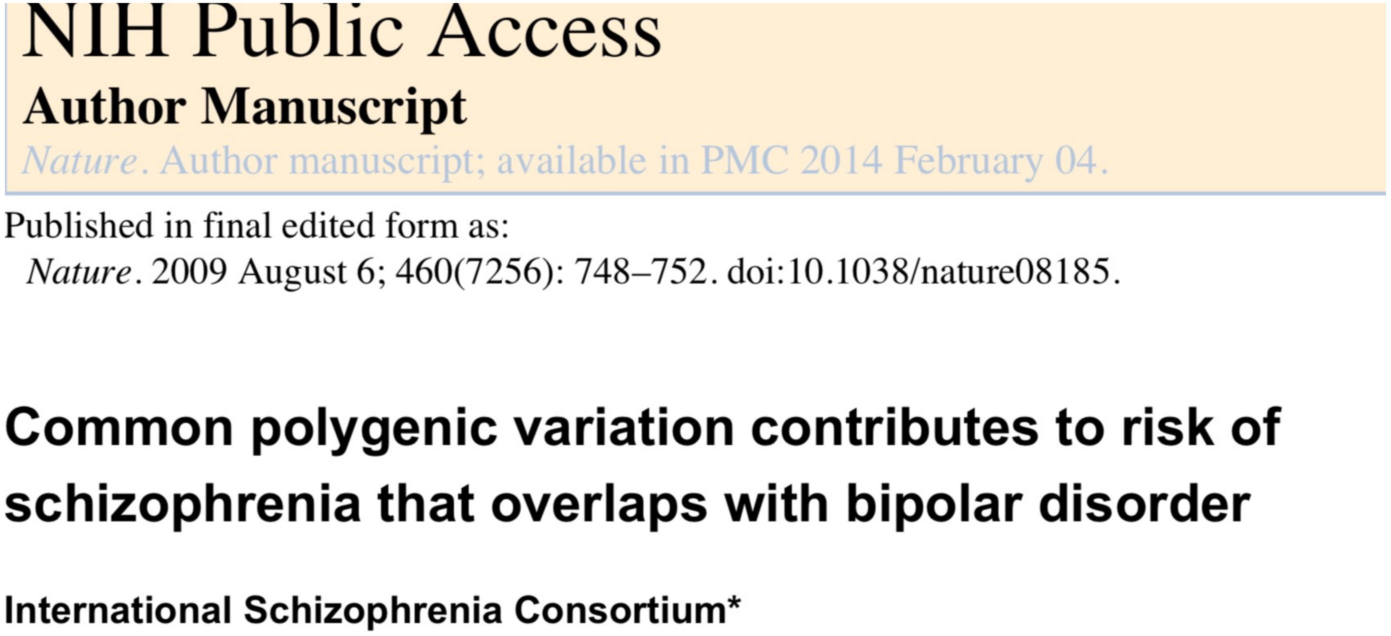
- Definition and calculation of polygenic scores
    - from GWAS to PGS
- Approaches to compute polygenic scores
    - standard vs advanced approaches
- Applications of PGS
    - trio design

Definition and calculation of polygenic scores

Polygenic (risk) Scores - **PRS**

- Polygenic scores - **PGS**
- Genetic (risk) Scores - **GRS**
- Genome-wide Polygenic Scores - **GPS**
- Polygenic Indices - **PGI**

- Individual **indices of the genetic predisposition, or burden, that an individual carries for a particular** (quantitative or case/control) **trait**



NIH Public Access

Author Manuscript

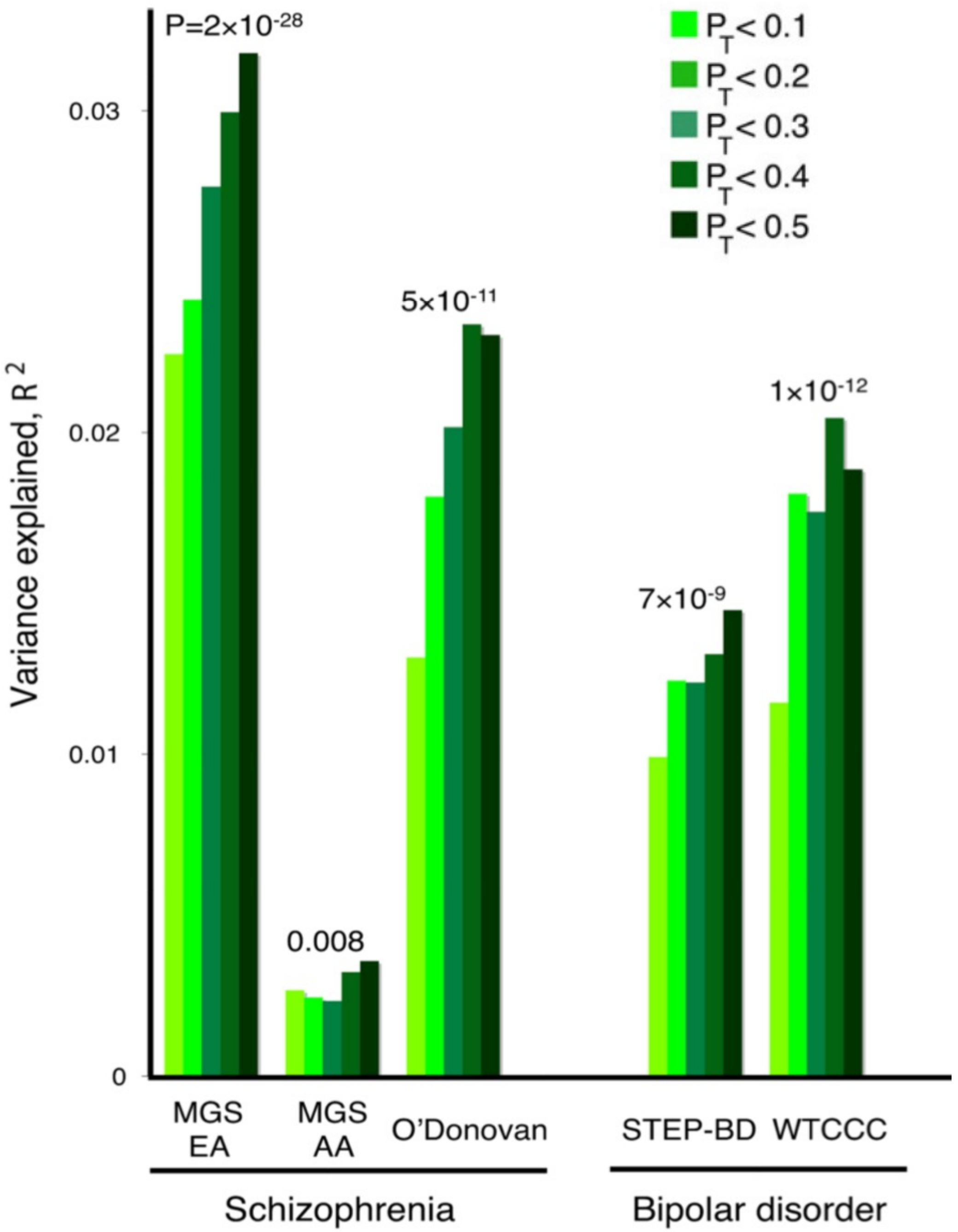
Nature. Author manuscript; available in PMC 2014 February 04.

Published in final edited form as:
Nature. 2009 August 6; 460(7256): 748–752. doi:10.1038/nature08185.

Common polygenic variation contributes to risk of schizophrenia that overlaps with bipolar disorder

International Schizophrenia Consortium*

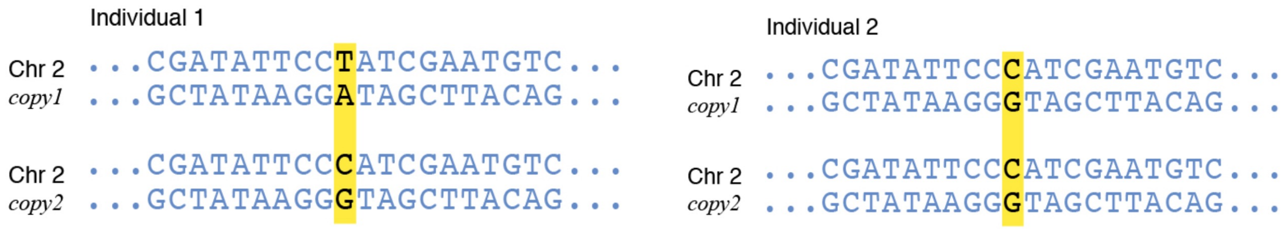
- **Landmark study in psychiatric genetics**
- **First application PGS Schizophrenia to infer genetic overlap with Bipolar Disorder**
- **Since then PGS have become common downstream analyses in GWAS analyses**



- *Measured genetic variation*

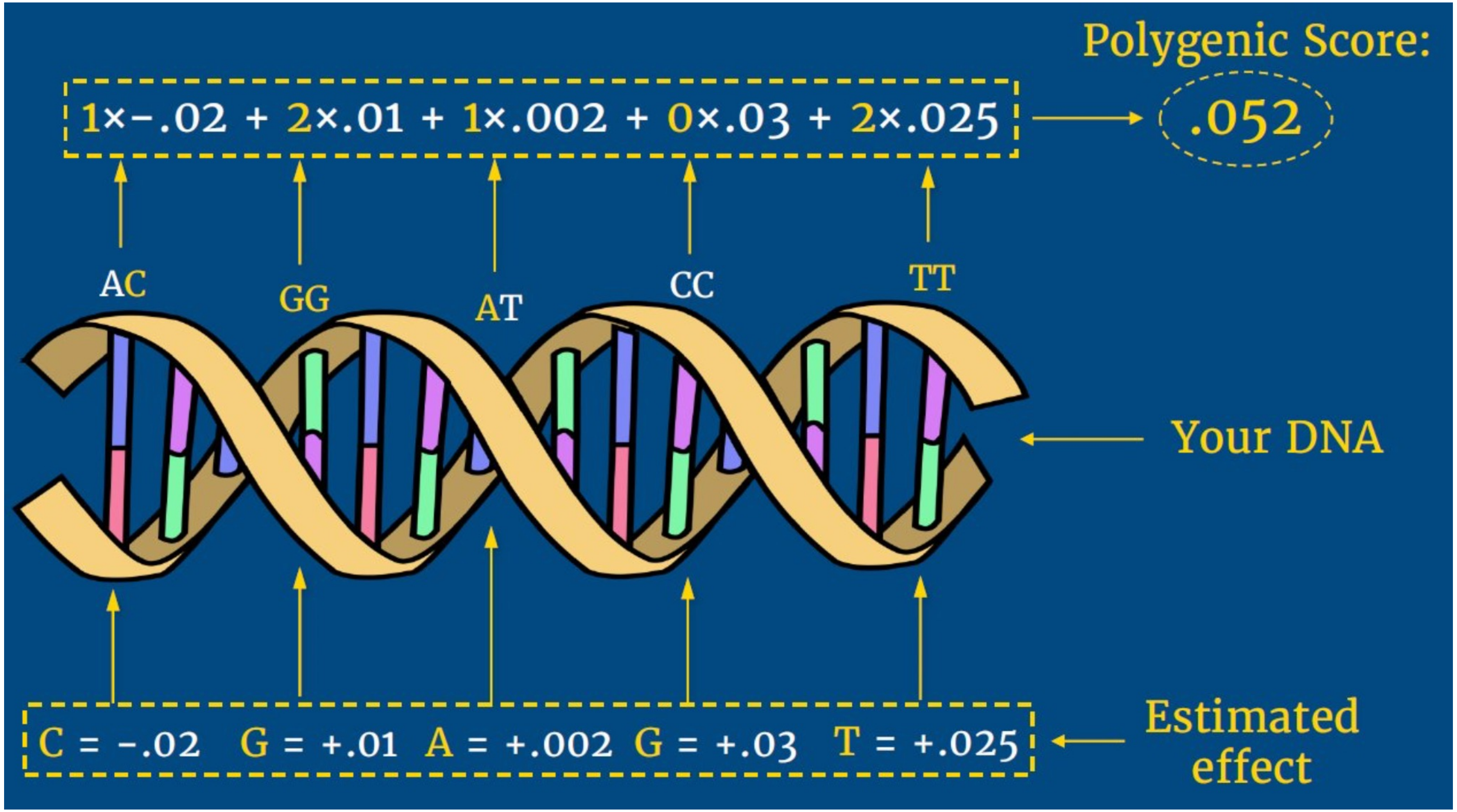
Most commonly:

- *Single nucleotide polymorphisms (**SNPs**)*: Common variation between individuals at a single position in the genetic code happening in at least 1% of the population.


Individual 1
Chr 2
copy1
...CGATATTCCTATCGAATGTC...
...GCTATAAGGATAGCTTACAG...
Chr 2
copy2
...CGATATTCCCATCGAATGTC...
...GCTATAAGGGTAGCTTACAG...
Individual 2
Chr 2
copy1
...CGATATTCCCATCGAATGTC...
...GCTATAAGGGTAGCTTACAG...
Chr 2
copy2
...CGATATTCCCATCGAATGTC...
...GCTATAAGGGTAGCTTACAG...

Calculation

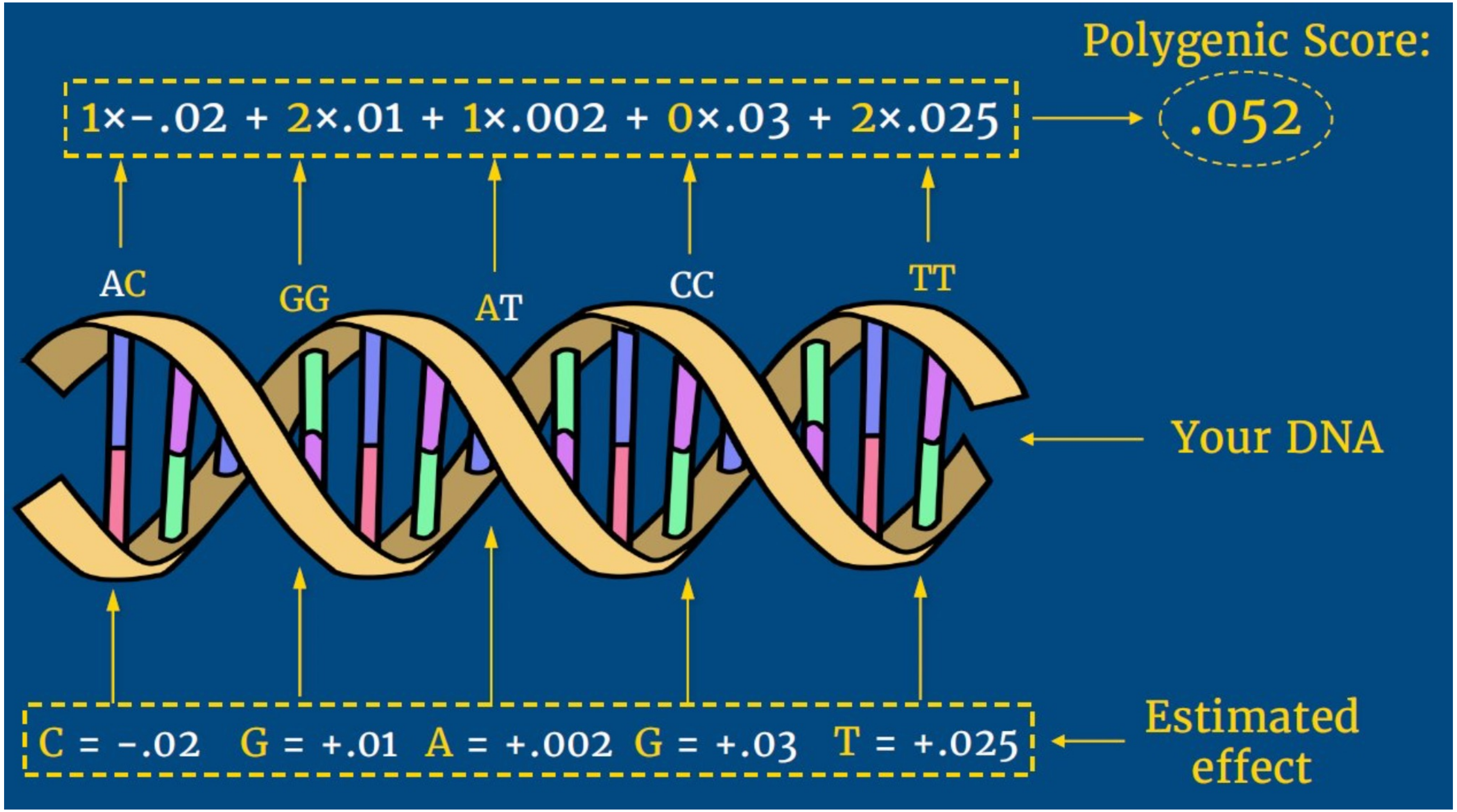
$$\hat{\text{PGS}} = \begin{bmatrix} \text{snp}_{11} & \cdots & \text{snp}_{1p} \\ \vdots & \vdots & \vdots \\ \text{snp}_{n1} & \cdots & \text{snp}_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} = \begin{bmatrix} \text{PGS}_1 \\ \vdots \\ \text{PGS}_n \end{bmatrix}$$



Calculation

$$\mathrm{P\hat{G}S} = \begin{bmatrix} \mathrm{snp}_{11} & \cdots & \mathrm{snp}_{1p} \\ \vdots & \vdots & \vdots \\ \mathrm{snp}_{n1} & \cdots & \mathrm{snp}_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} = \begin{bmatrix} \boxed{\mathrm{PGS}_1} \\ \vdots \\ \mathrm{PGS}_n \end{bmatrix} \longrightarrow \mathrm{P\hat{G}S}_i = \sum_{j=1}^{p} \left(x_{ij} \beta_j \right)$$

$$x_{ij} \in \{0, 1, 2\}$$



PGS are approximately normally distributed in the population with people varying on a continuum from low to high polygenic burden for a particular trait

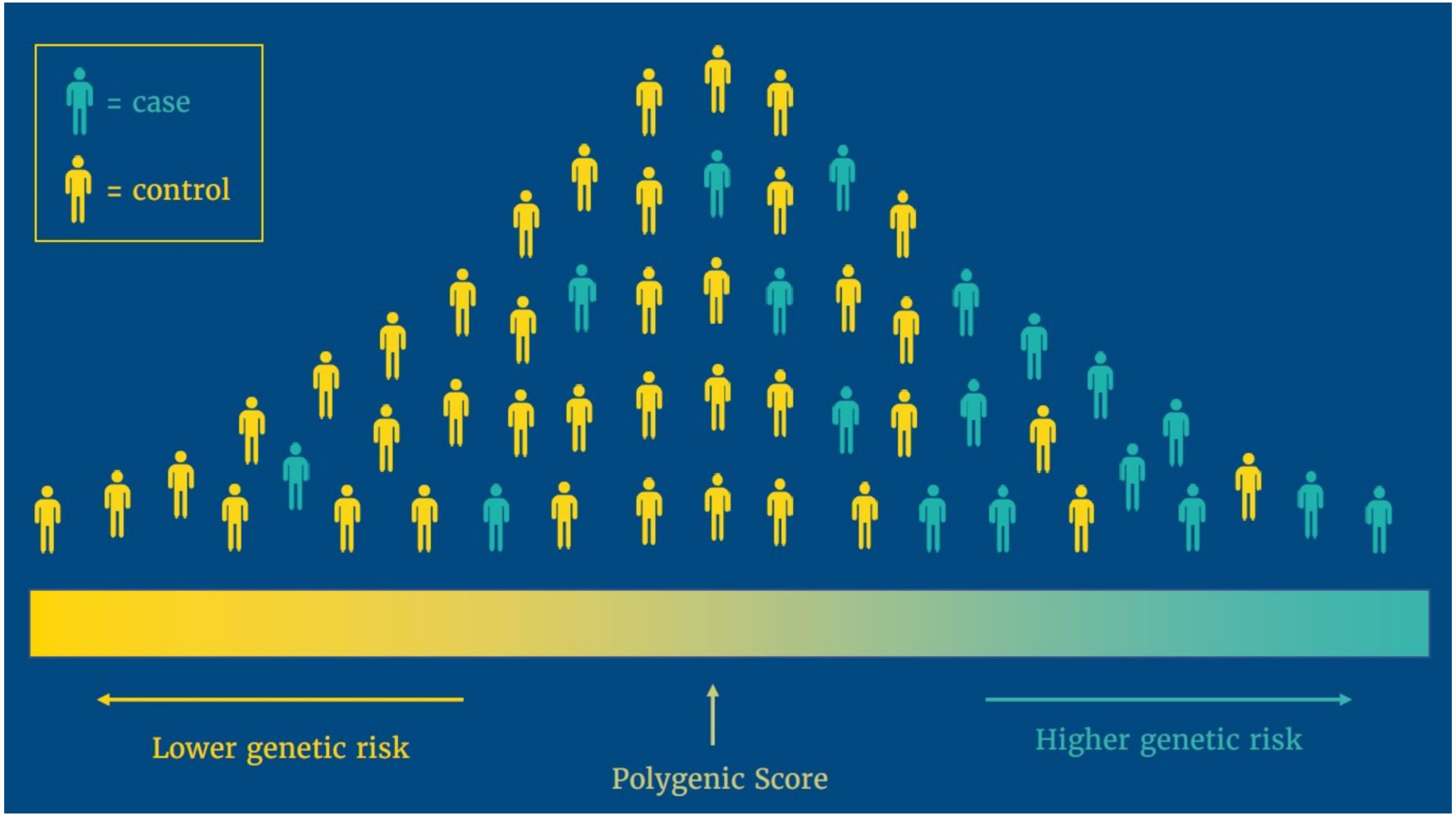
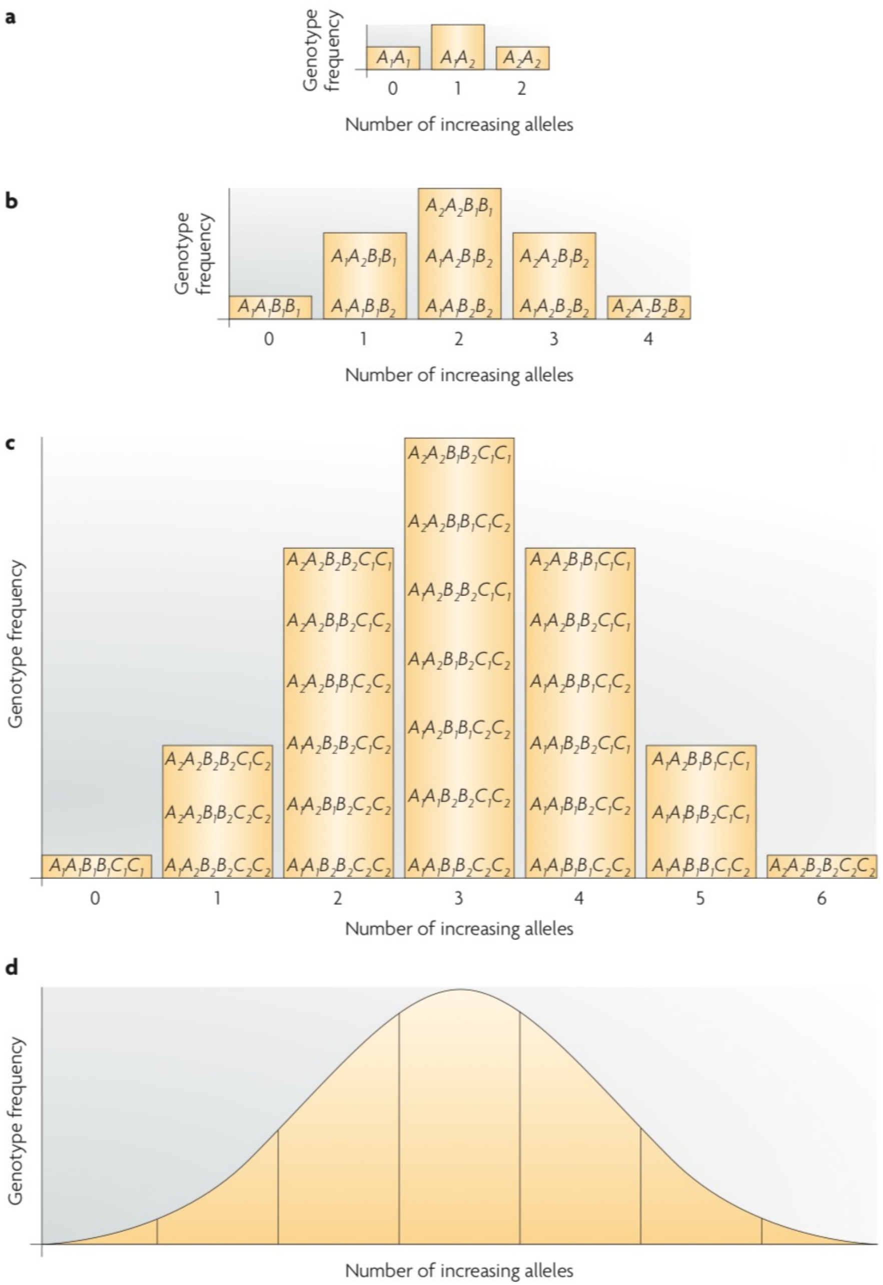
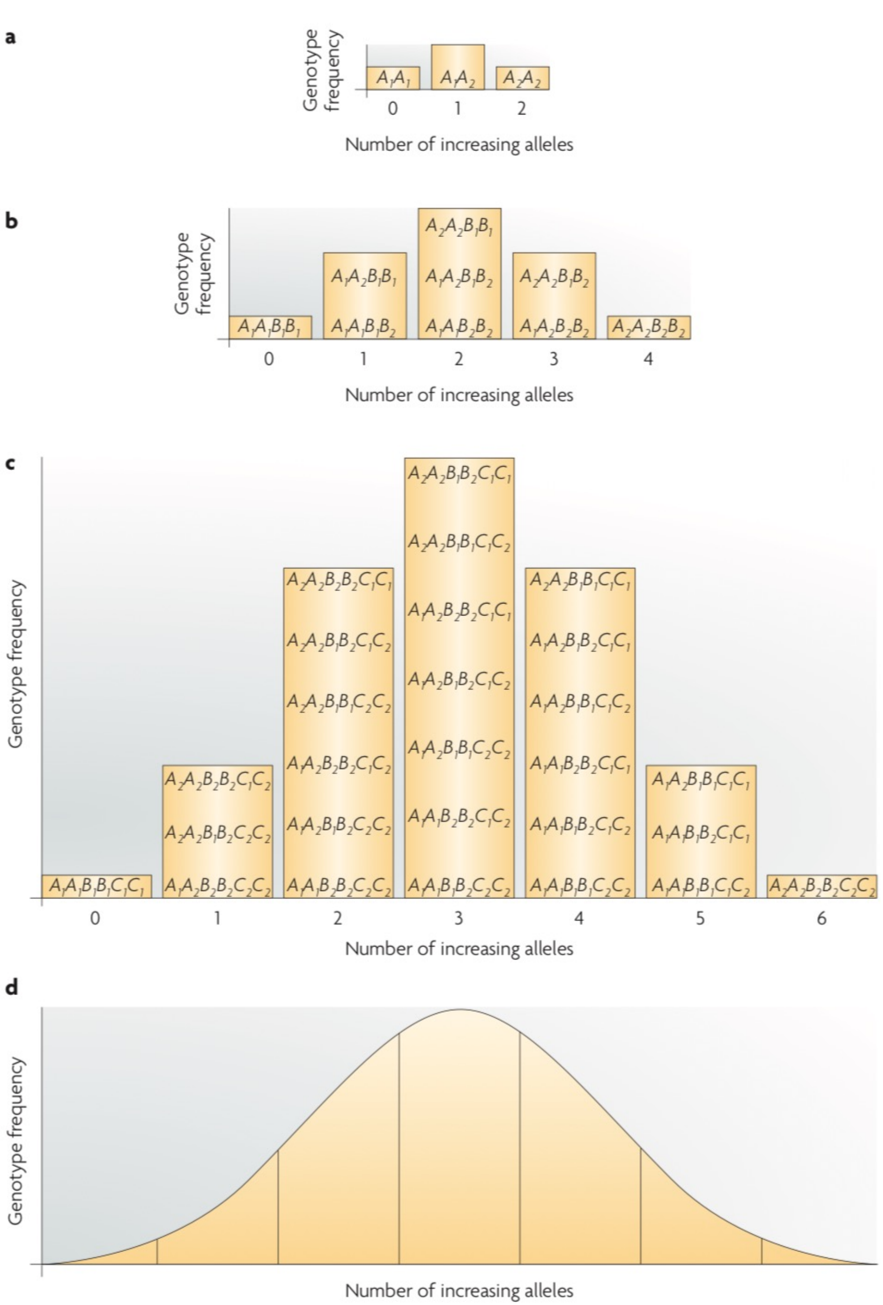


Figure: Abdellaoui, A. [@dr_Appie]. (2023, Mar 26). Twitter



Plomin et al ., 2009. *Nature reviews genetics*, *10*(12), 872-878



Plomin et al ., 2009. *Nature reviews genetics*, *10*(12), 872-878

REVIEW

15 years of GWAS discovery: Realizing the promise

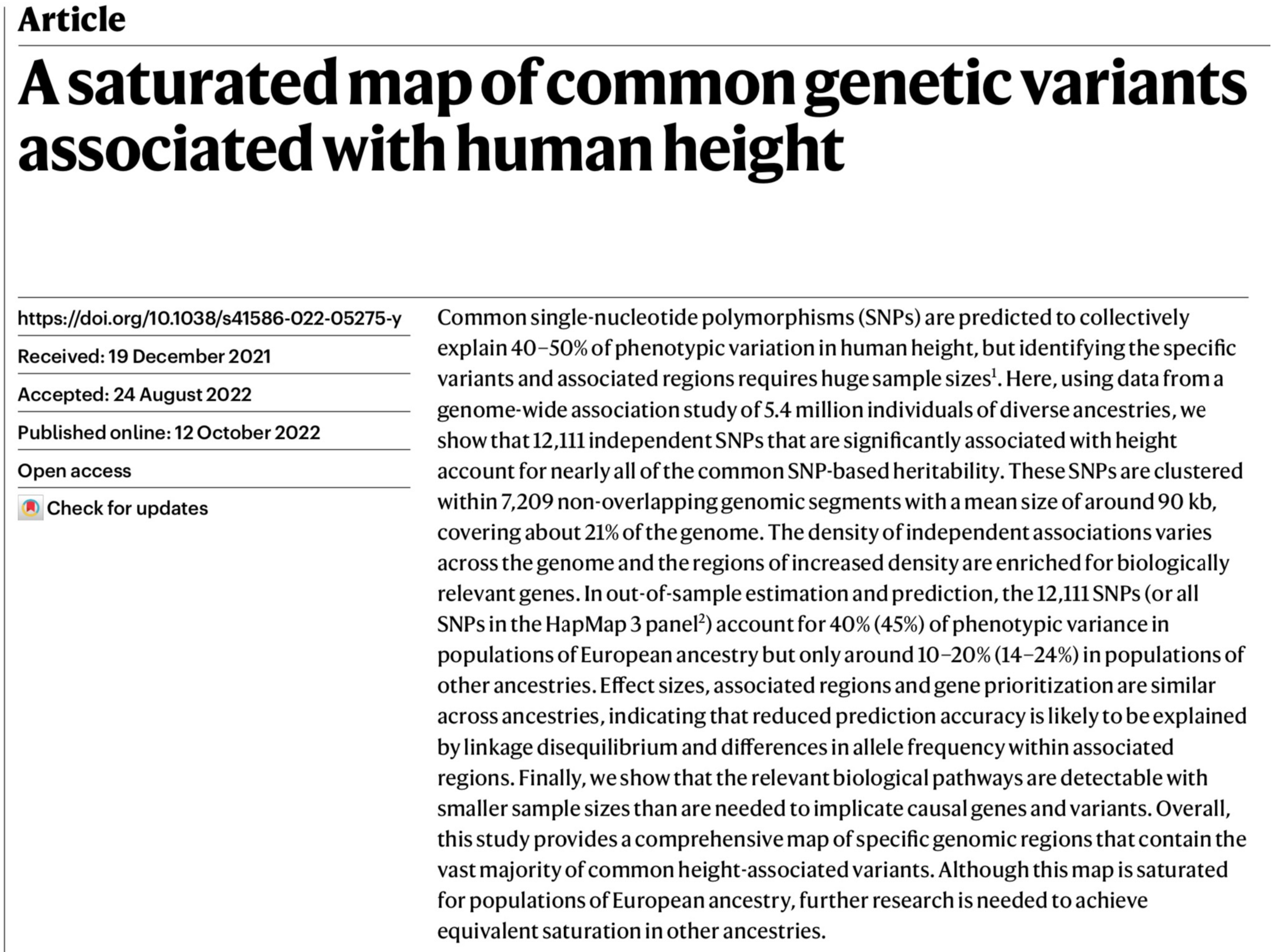
Abdel Abdellaoui,[1,*] Loic Yengo,[2] Karin J.H. Verweij,[1] and Peter M. Visscher[2]

Height served as the model complex trait when Mendel’s laws of inheritance were reconciled with the inheritance of quantitative traits

- Saturation of the common-variant architecture among European-ancestry genomes

- Approximately **12,000 SNPs** jointly explain **40% of variation** in out-of-sample prediction

- Approaches the *common* SNP-based heritability



Article

A saturated map of common genetic variants associated with human height

https://doi.org/10.1038/s41586-022-05275-y

Received: 19 December 2021

Accepted: 24 August 2022

Published online: 12 October 2022

Open access

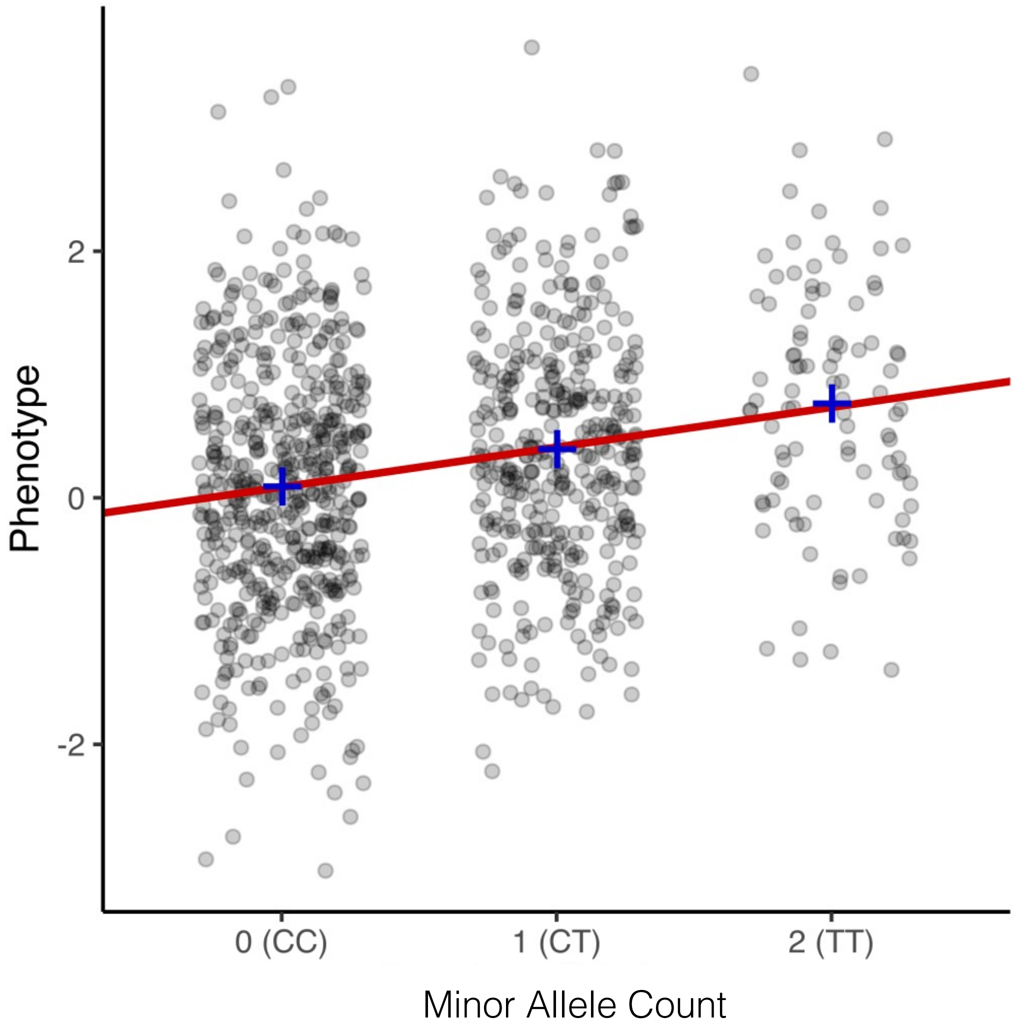
Check for updates

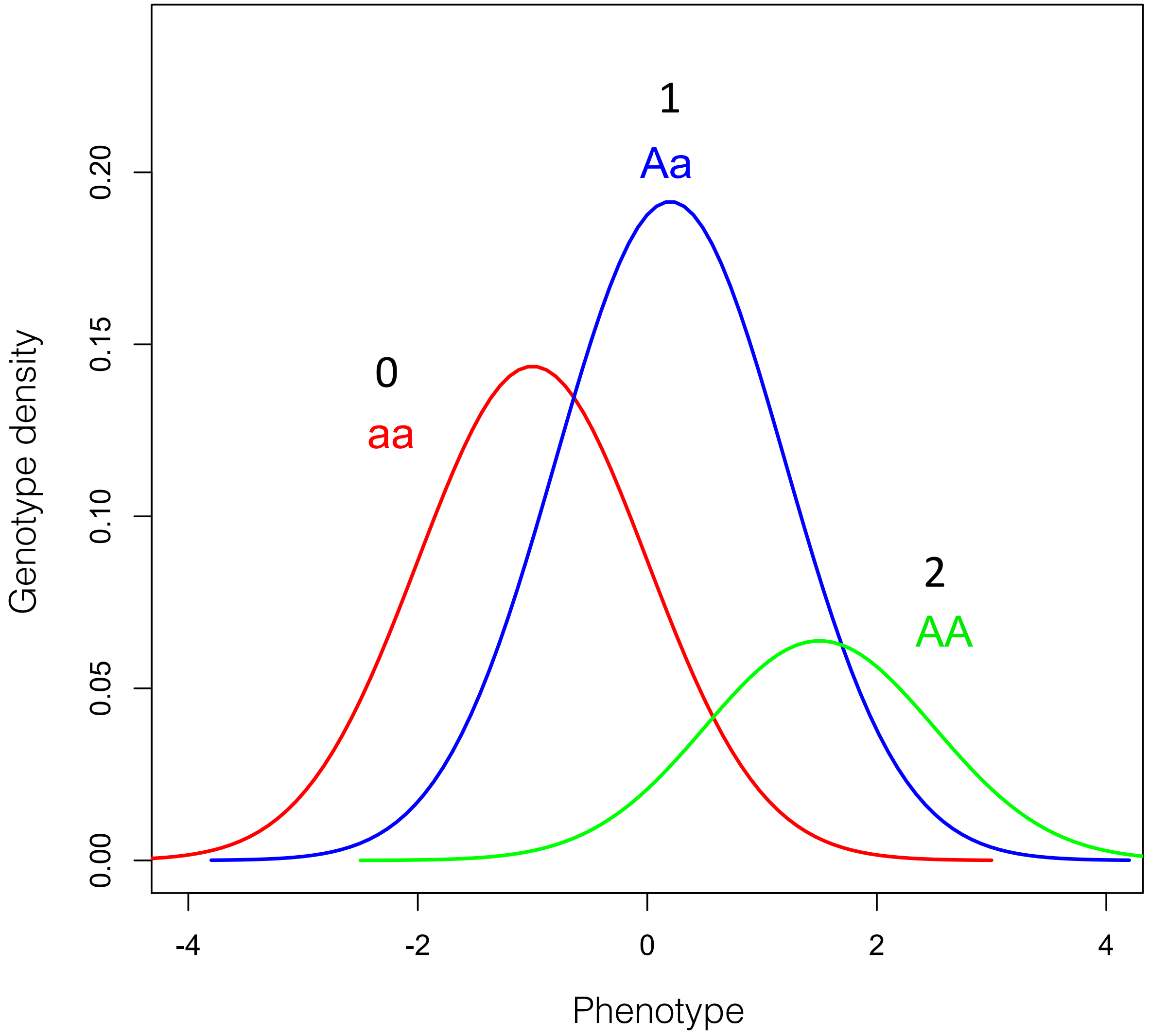
Common single-nucleotide polymorphisms (SNPs) are predicted to collectively explain 40–50% of phenotypic variation in human height, but identifying the specific variants and associated regions requires huge sample sizes[1]. Here, using data from a genome-wide association study of 5.4 million individuals of diverse ancestries, we show that 12,111 independent SNPs that are significantly associated with height account for nearly all of the common SNP-based heritability. These SNPs are clustered within 7,209 non-overlapping genomic segments with a mean size of around 90 kb, covering about 21% of the genome. The density of independent associations varies across the genome and the regions of increased density are enriched for biologically relevant genes. In out-of-sample estimation and prediction, the 12,111 SNPs (or all SNPs in the HapMap 3 panel[2]) account for 40% (45%) of phenotypic variance in populations of European ancestry but only around 10–20% (14–24%) in populations of other ancestries. Effect sizes, associated regions and gene prioritization are similar across ancestries, indicating that reduced prediction accuracy is likely to be explained by linkage disequilibrium and differences in allele frequency within associated regions. Finally, we show that the relevant biological pathways are detectable with smaller sample sizes than are needed to implicate causal genes and variants. Overall, this study provides a comprehensive map of specific genomic regions that contain the vast majority of common height-associated variants. Although this map is saturated for populations of European ancestry, further research is needed to achieve equivalent saturation in other ancestries.

From GWAS to PGS

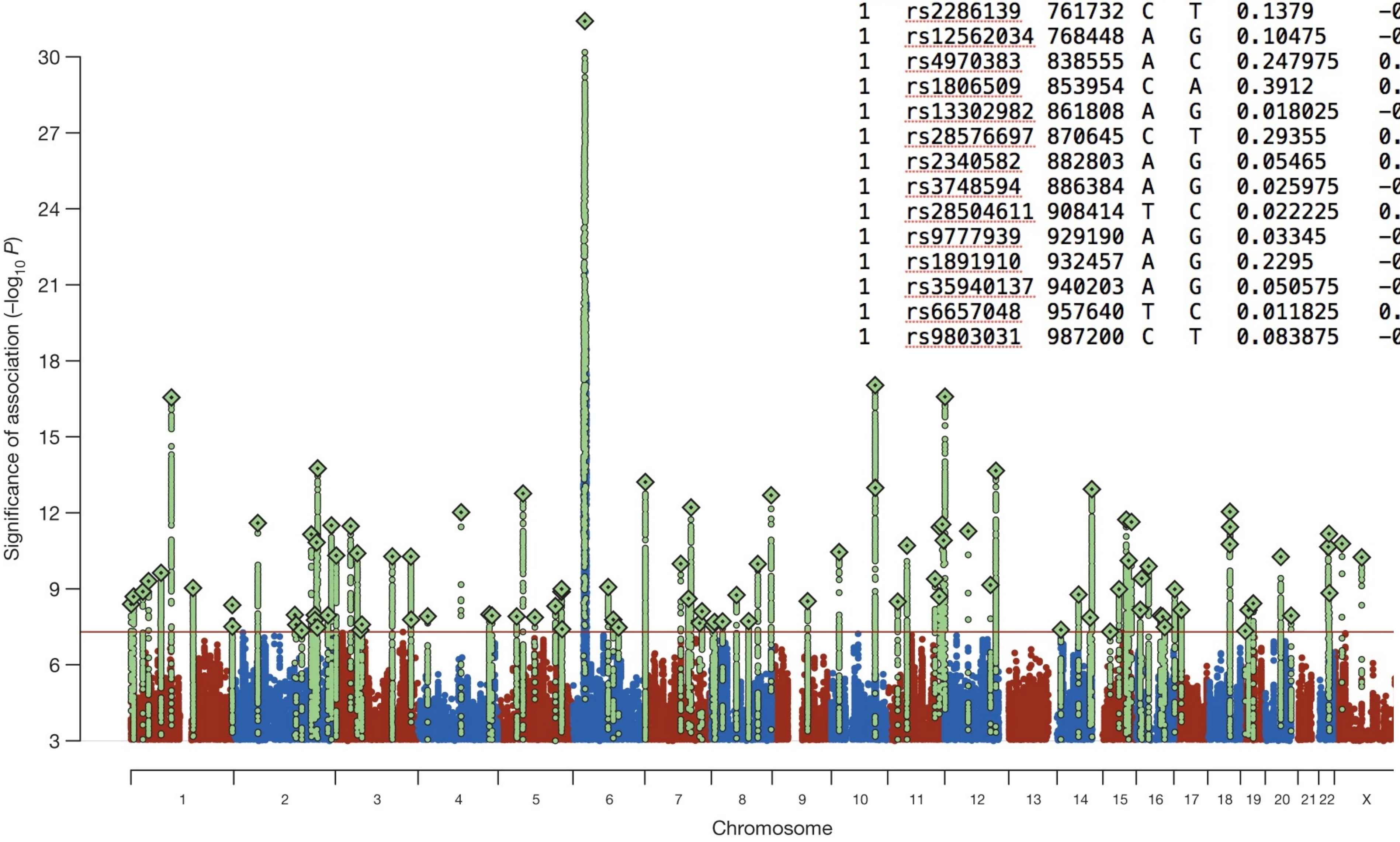
Association test

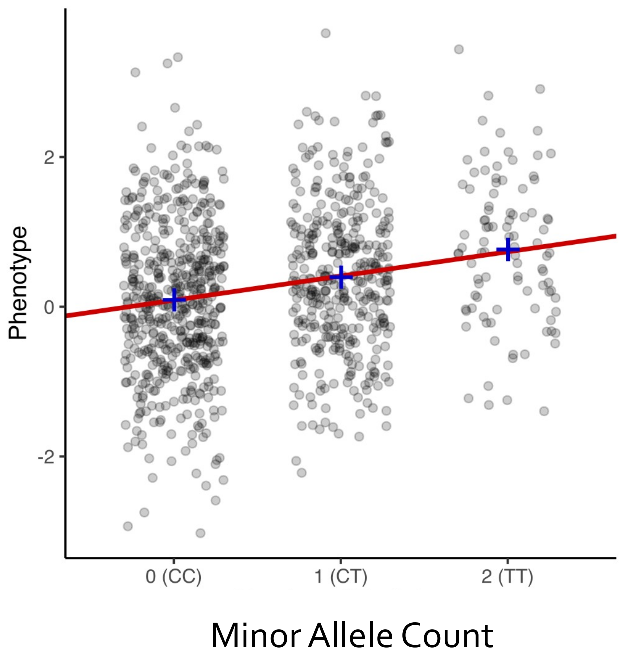
$$Y_i = a + bX_i + e_i$$





```
Chr SNP        bp     A1 A2 Freq      b           se         p
1   rs2286139  761732 C  T  0.1379    -0.0104056  0.00732416 0.155397
1   rs12562034 768448 A  G  0.10475   -0.00627592 0.00827054 0.447955
1   rs4970383  838555 A  C  0.247975  0.00946201  0.00587444 0.107243
1   rs1806509  853954 C  A  0.3912    0.0152744   0.00523012 0.00349507
1   rs13302982 861808 A  G  0.018025  -0.0180122  0.0189517  0.341895
1   rs28576697 870645 C  T  0.29355   0.0116486   0.00556379 0.0362916
1   rs2340582  882803 A  G  0.05465   0.0119371   0.0111055  0.282426
1   rs3748594  886384 A  G  0.025975  -0.01244    0.0158797  0.433401
1   rs28504611 908414 T  C  0.022225  0.00388796  0.0171623  0.820781
1   rs9777939  929190 A  G  0.03345   -0.00446279 0.0141522  0.752502
1   rs1891910  932457 A  G  0.2295    -0.00647527 0.00605864 0.285175
1   rs35940137 940203 A  G  0.050575  -0.10689    0.0115935  2.97533e-20
1   rs6657048  957640 T  C  0.011825  0.0892934   0.0233322  0.000129691
1   rs9803031  987200 C  T  0.083875  -0.00434284 0.0091306  0.634334
```





Manhattan plot from: Biological insights from 108 schizophrenia-associated genetic loci

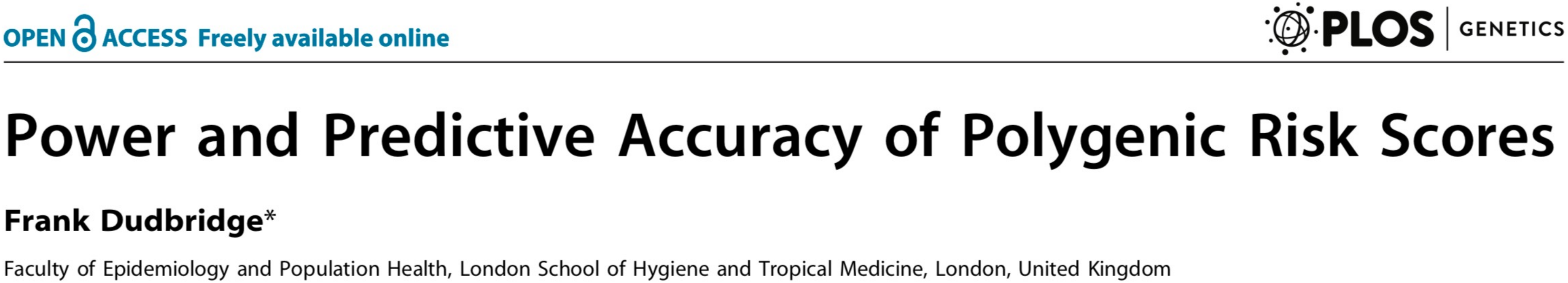
36,989 cases and 113,075 controls

Important (1)

- (SNP)h^2 is spread across thousands of loci of very small effect

- Sample size of GWAS is of central importance for the discovery and estimation of SNP effects and, in turn, for the predictive power of PGS

- See Dudbridge, 2013



OPEN ACCESS Freely available online

PLOS | GENETICS

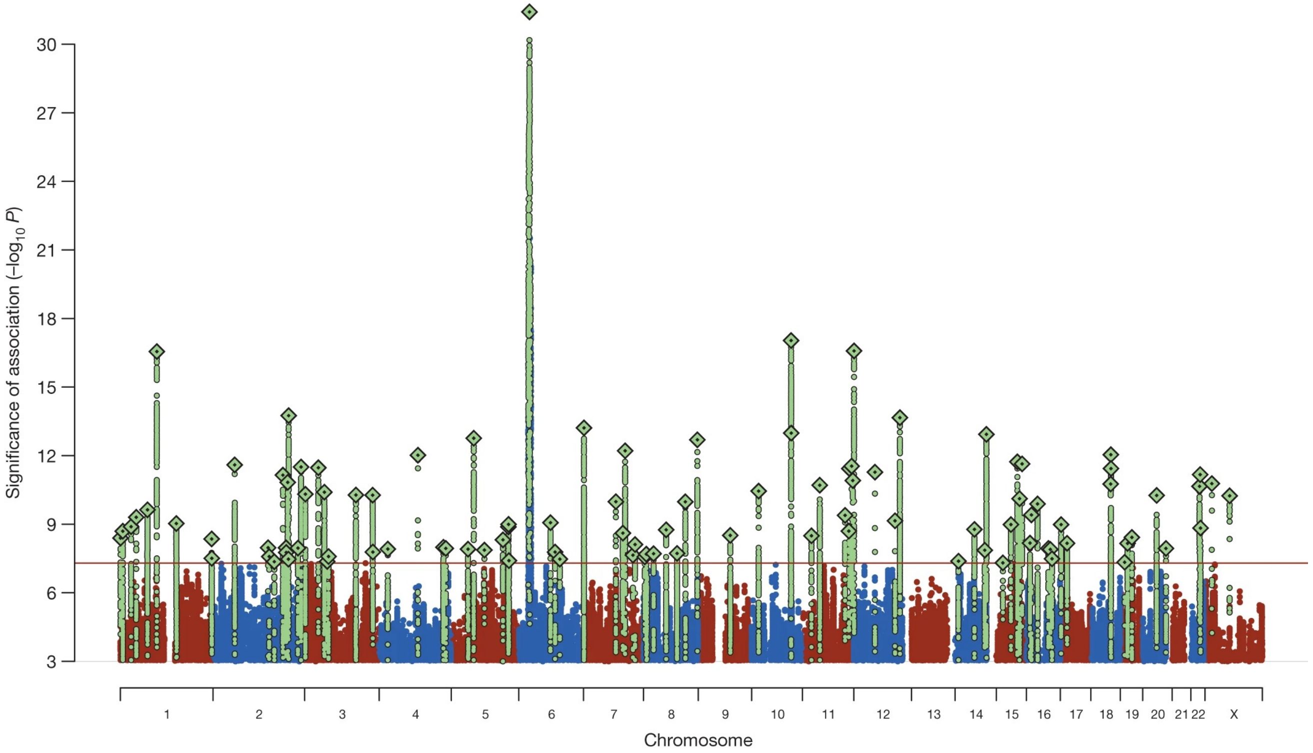
Power and Predictive Accuracy of Polygenic Risk Scores

Frank Dudbridge*

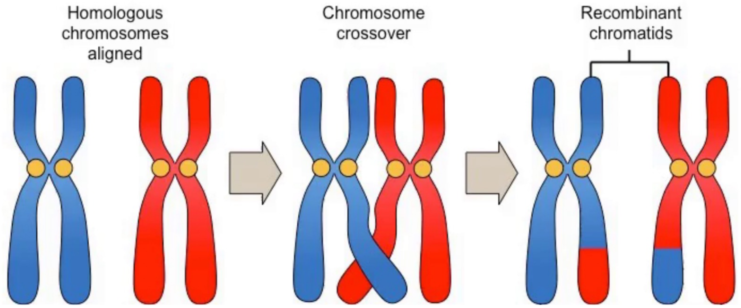
Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom

Important (2)

- We test for the effects of millions of variants on a phenotype but these are not independent due to linkage disequilibrium (LD)

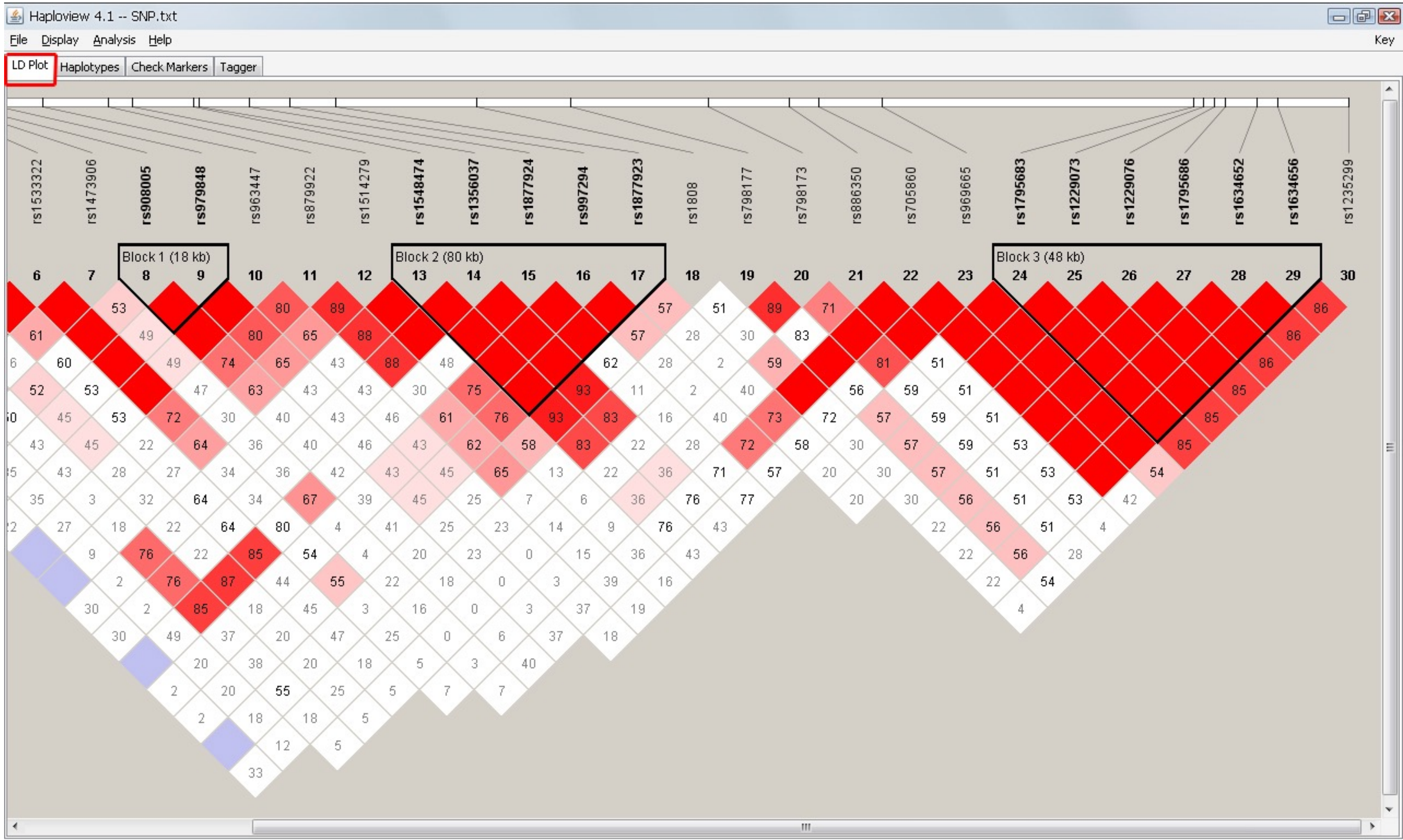


Linkage disequilibrium (LD)



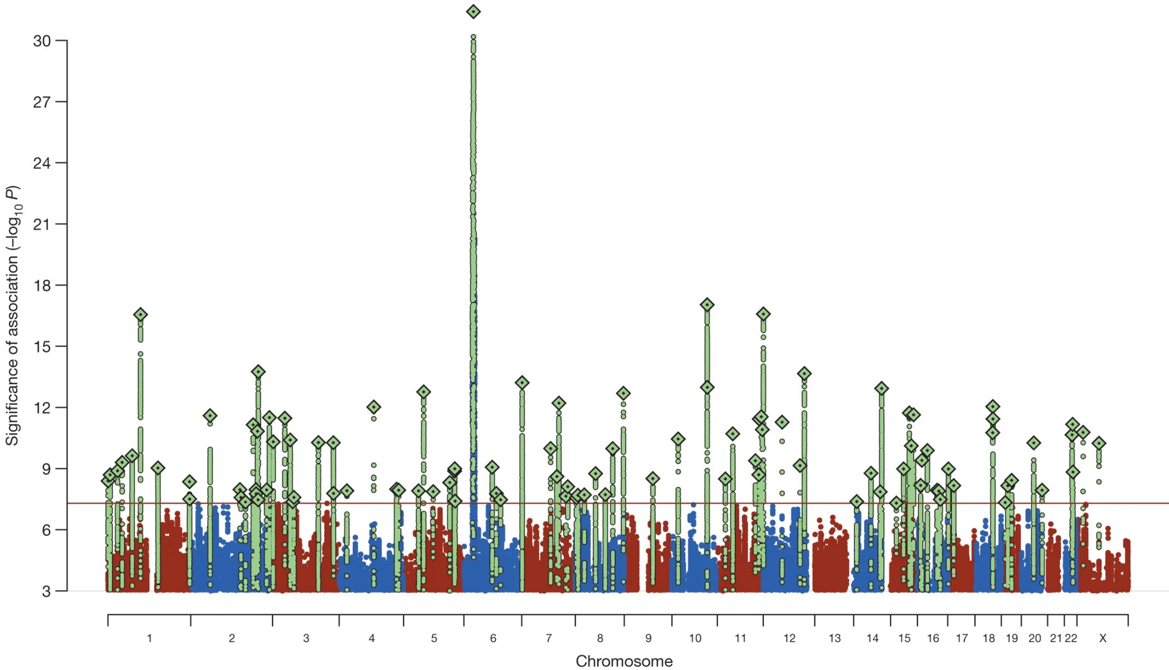
- *Linkage disequilibrium (LD)*: the correlation between nearby variants on the same chromosome

- Genetic variants near one another tend to be inherited together over generations



Important (2)

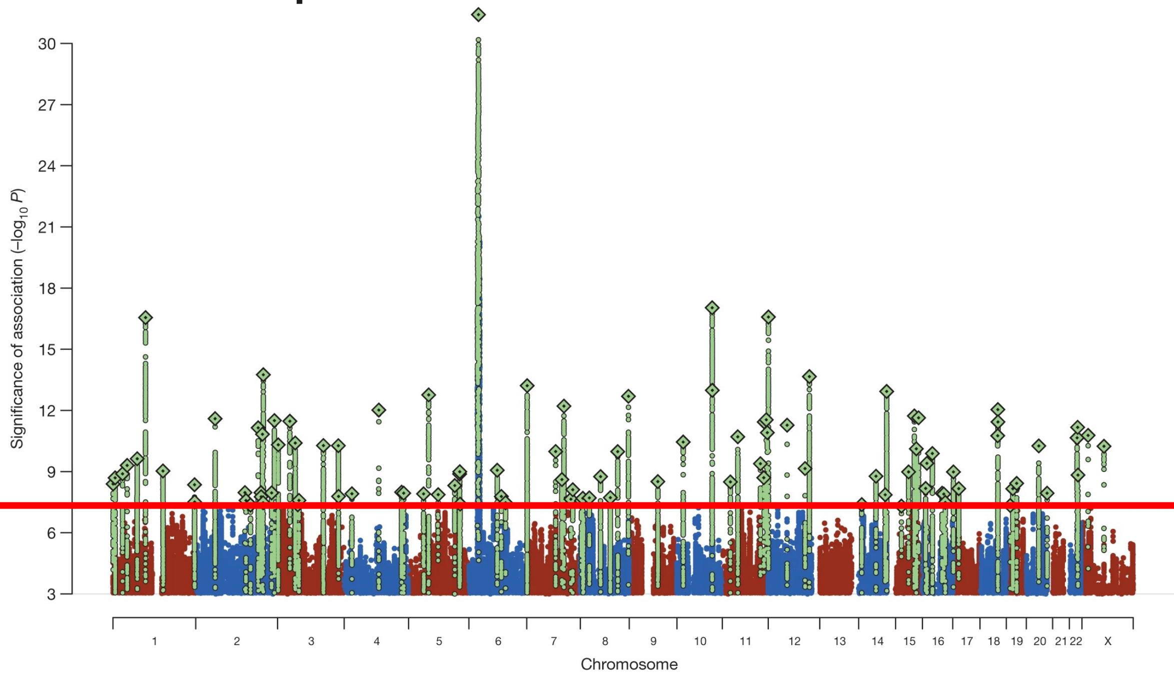
- We test for the effects of millions of variants on a phenotype but these are not independent due to linkage disequilibrium (LD)



Important (2)

- We test for the effects of millions of variants on a phenotype but these are not independent due to linkage disequilibrium (LD)

- We are conducting the equivalent of 1 million independent tests (Risch & Merikangas, 1996)



genome-wide significance threshold of 5 × 10^{-8}

Important (3)

- We need to account/adjust for LD when aggregating SNPs in PGS!

- Otherwise individual contributions of the specific loci included will be overestimated

- Optimally, we would estimate joint effects of all SNPs in a multivariable framework, but this is not feasible!

So you have your discovery GWAS sumstats…

PGC Psychiatric Genomics Consortium

ieu open gwas project

GWASATLAS

CAREFUL! **these need to be independent of your target set!**
(i.e. where you are performing PGS-phenotype analyses)

What next?

Approaches to compute polygenic scores

Approaches to compute polygenic scores

- which SNPs to include?
- how do you adjust for LD?

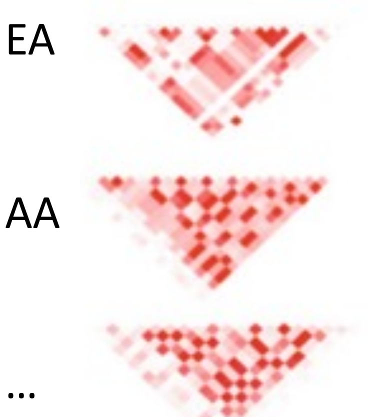
The standard approach

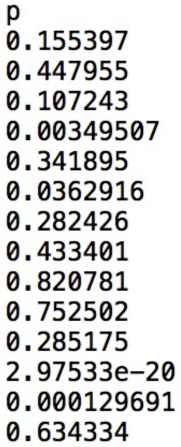
- Clumping and thresholding (C + T) approach

The standard approach

- Clumping and thresholding (C + T) approach

- **LD-clumping: obtaining a set of quasi-independent SNPs**
  - information from (ancestry-matched) LD reference panel (can be your target set)
  - prioritizing on *p*-values from GWAS summary statistics





```
p
0.155397
0.447955
0.107243
0.00349507
0.341895
0.0362916
0.282426
0.433401
0.820781
0.752502
0.285175
2.97533e-20
0.000129691
0.634334
```

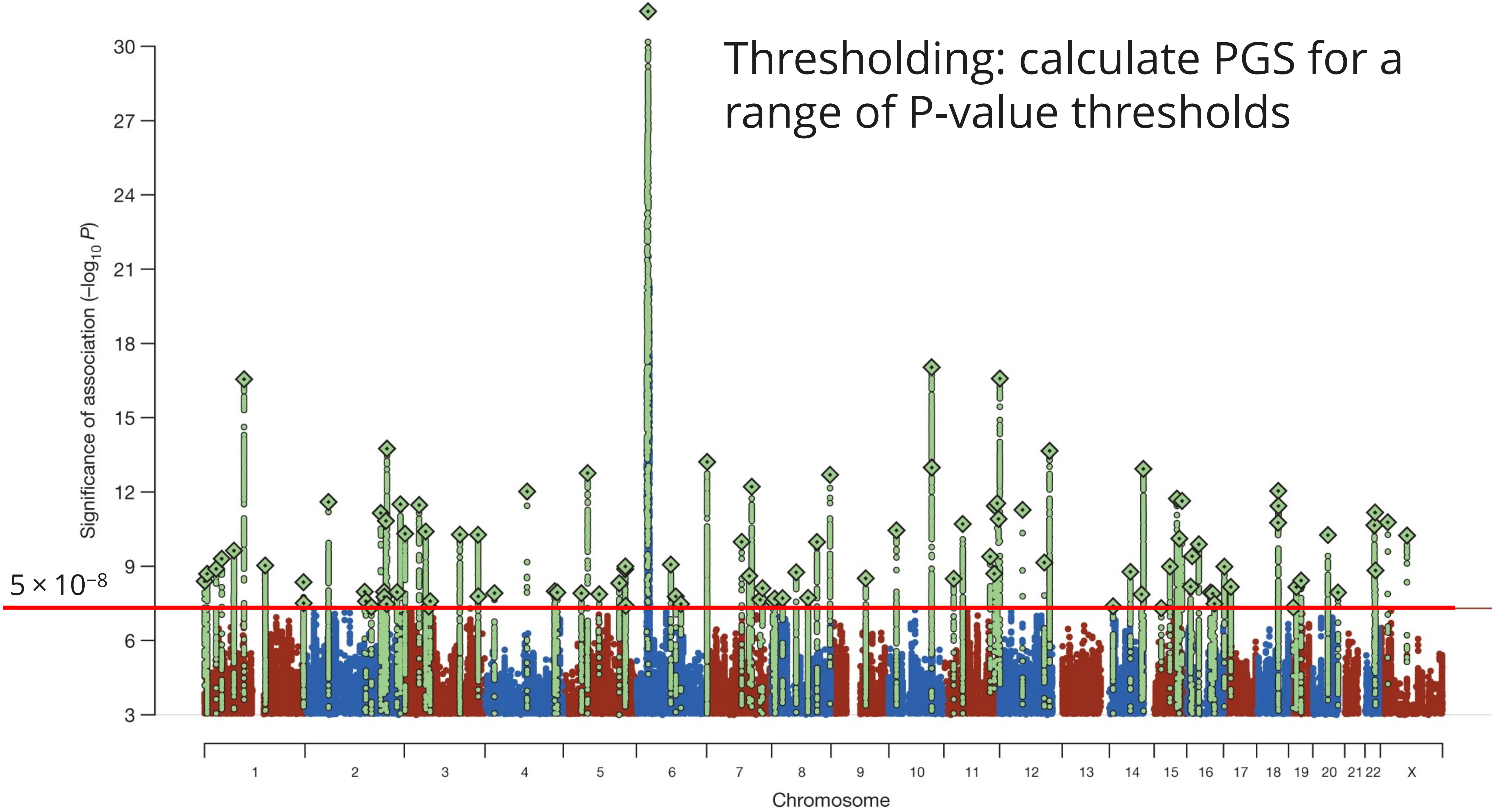
The standard approach

- Clumping and thresholding (C + T) approach

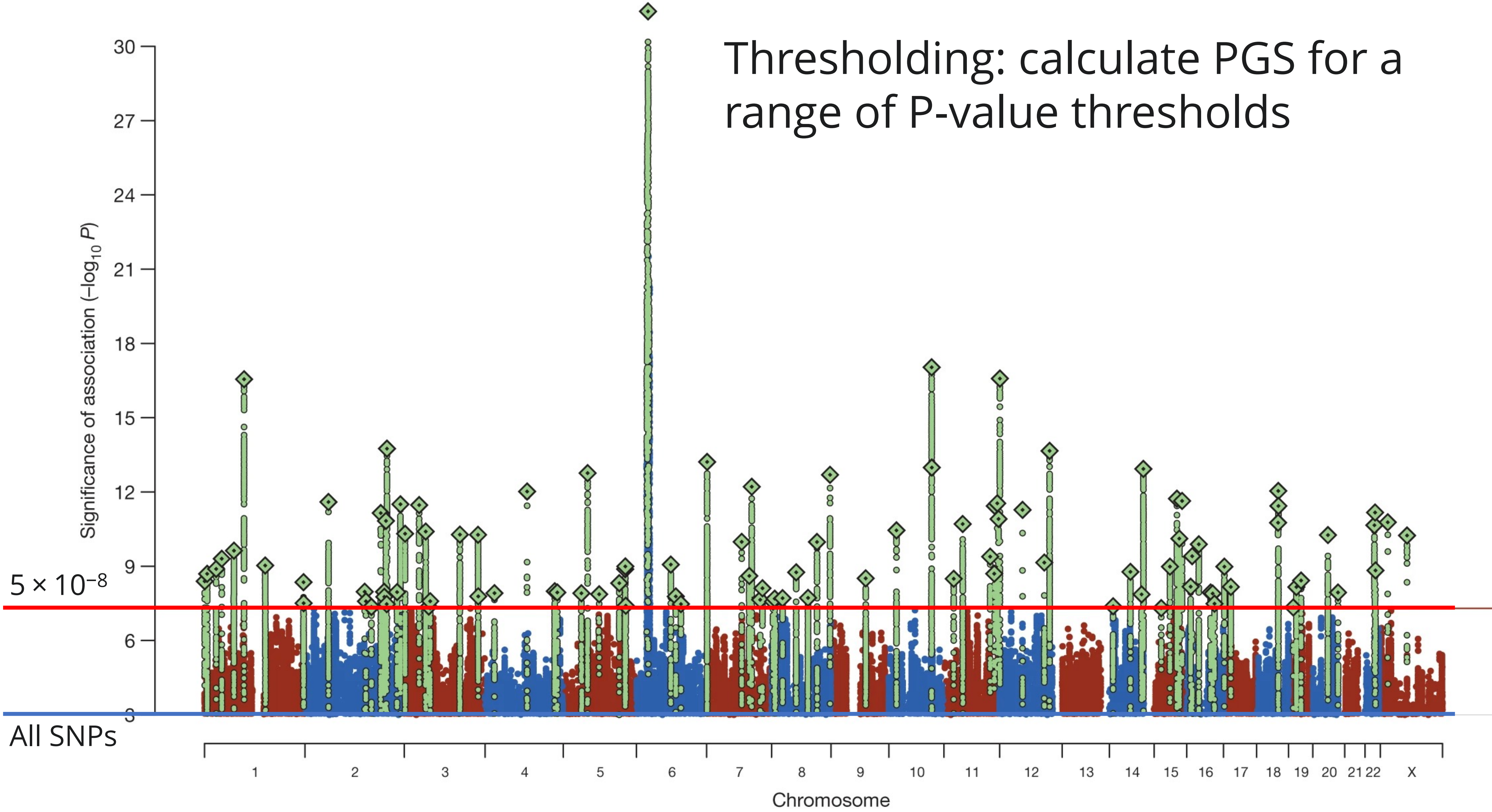
- LD-clumping: obtaining a set of quasi-independent SNPs
  - information from (ancestry-matched) LD reference panel (can be your target set)
  - prioritizing on *p*-values from GWAS summary statistics

- **Thresholding: calculate PGS for a range of P-value thresholds**

Thresholding: calculate PGS for a range of P-value thresholds



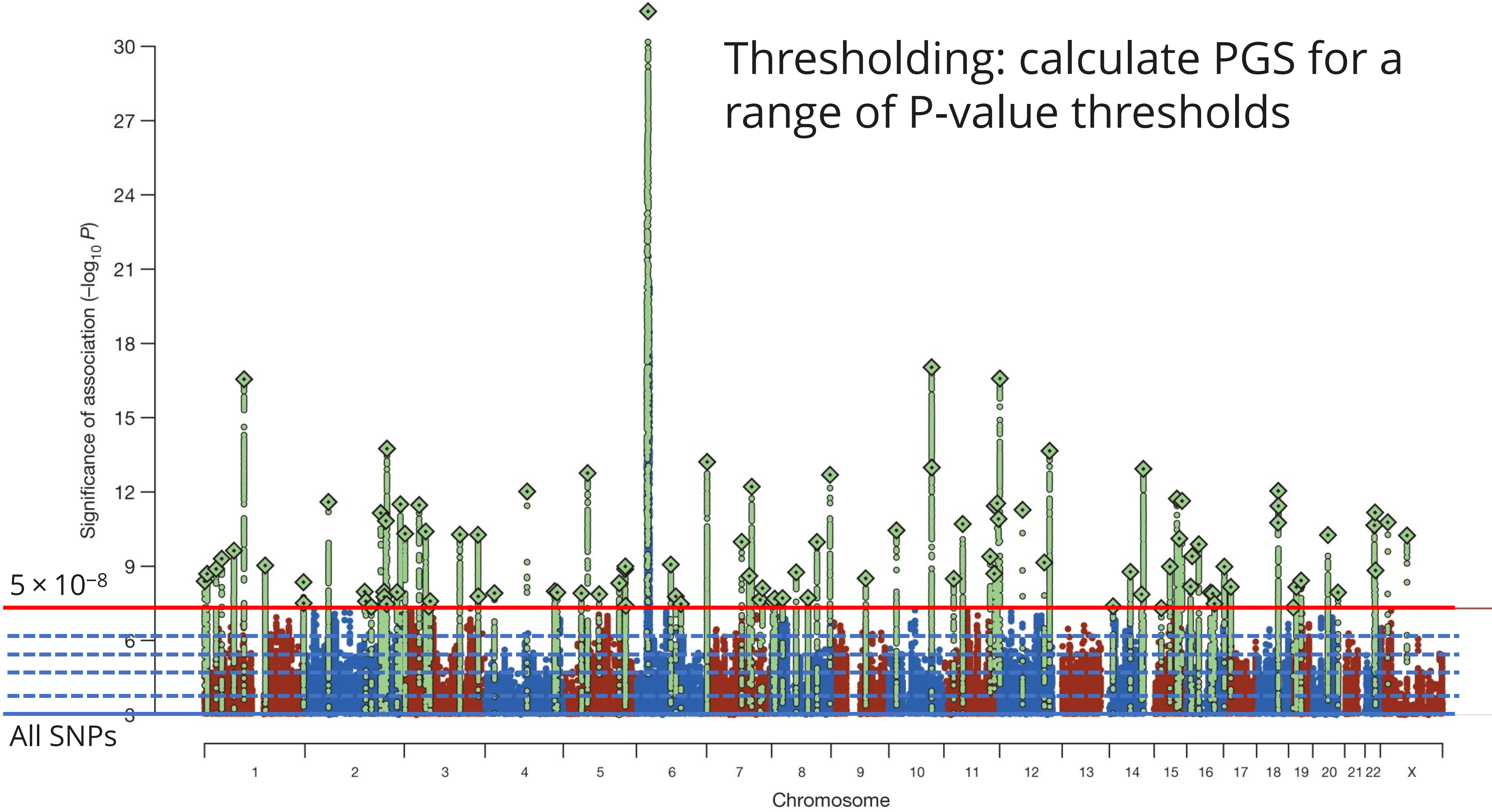
Thresholding: calculate PGS for a range of P-value thresholds



5×10^{-8}

All SNPs

Thresholding: calculate PGS for a range of P-value thresholds



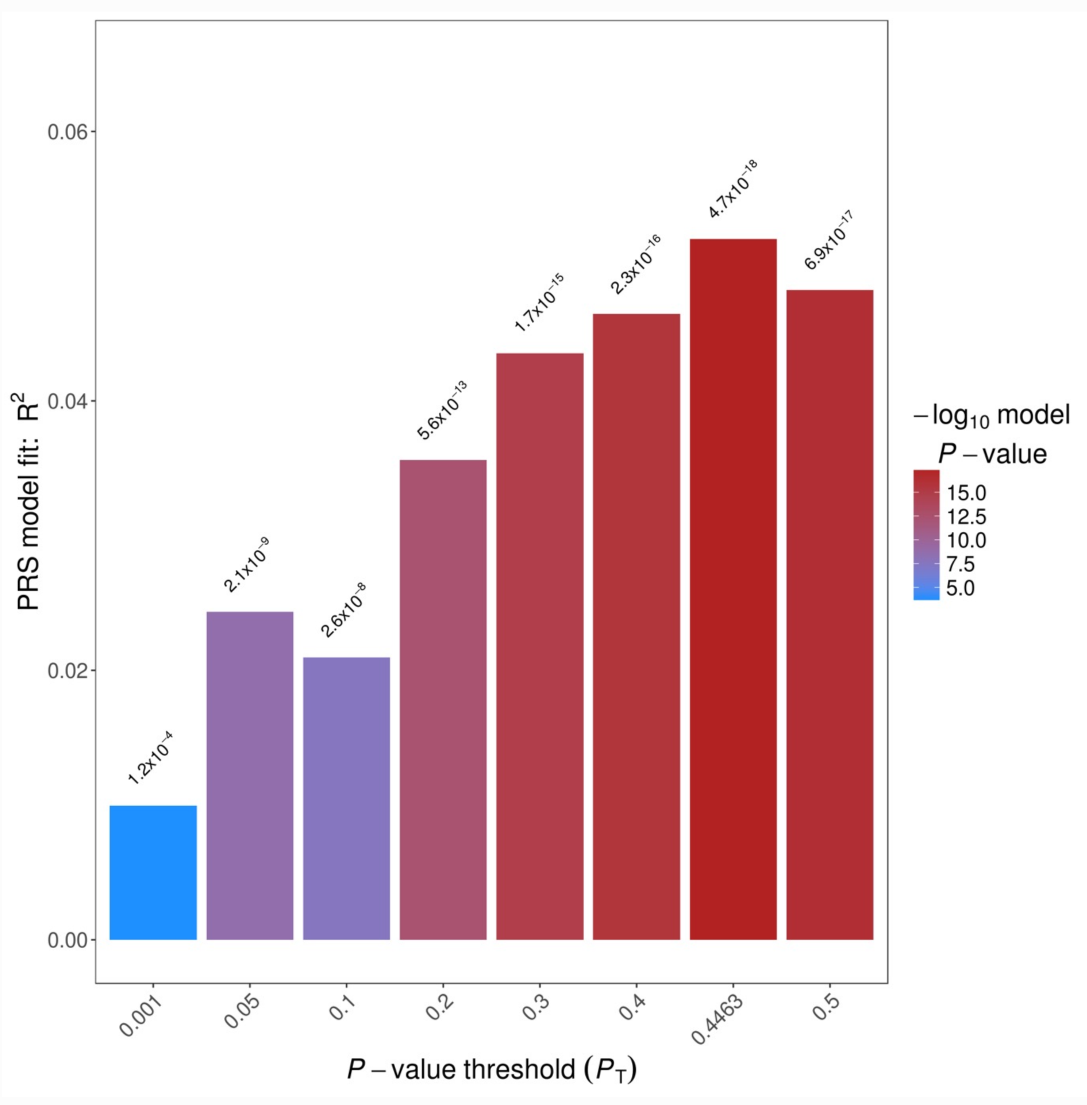
The standard approach

which SNPs to include?

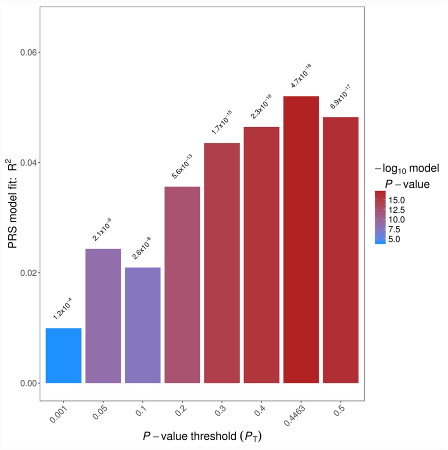
- Standard GWAS threshold to select SNPs is often too restrictive for the purpose of PGS construction

- PGS derived from all SNPs can also be suboptimal due to added noise across many false positives SNPs included

(PRSice2 bar plot)



GigaScience, Volume 8, Issue 7, July 2019, giz082



What options do you have?

Pick your poison:

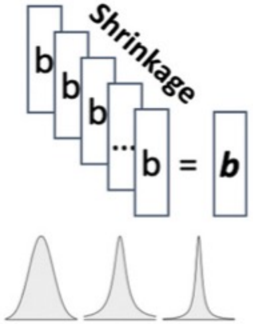
- Choose an a-priori threshold for inclusion (suboptimal in most cases!)
- Report all scores (biased upward!)
- Optimize score in a validation set (need separate sample from your discovery and target sets!)
- Obtain a unique score from first Principal Component across thresholds [see Coombes et al., 2020] (in most realistic scenarios this =~ threshold 1)
- **Use advanced 'single-score' methods approaches**

Advanced approaches

Advanced approaches

- **Two broad themes:** which SNPs are included in the scores, and the assumed distribution of SNP effect sizes

- **Depending on these:** reweighting of GWAS estimates



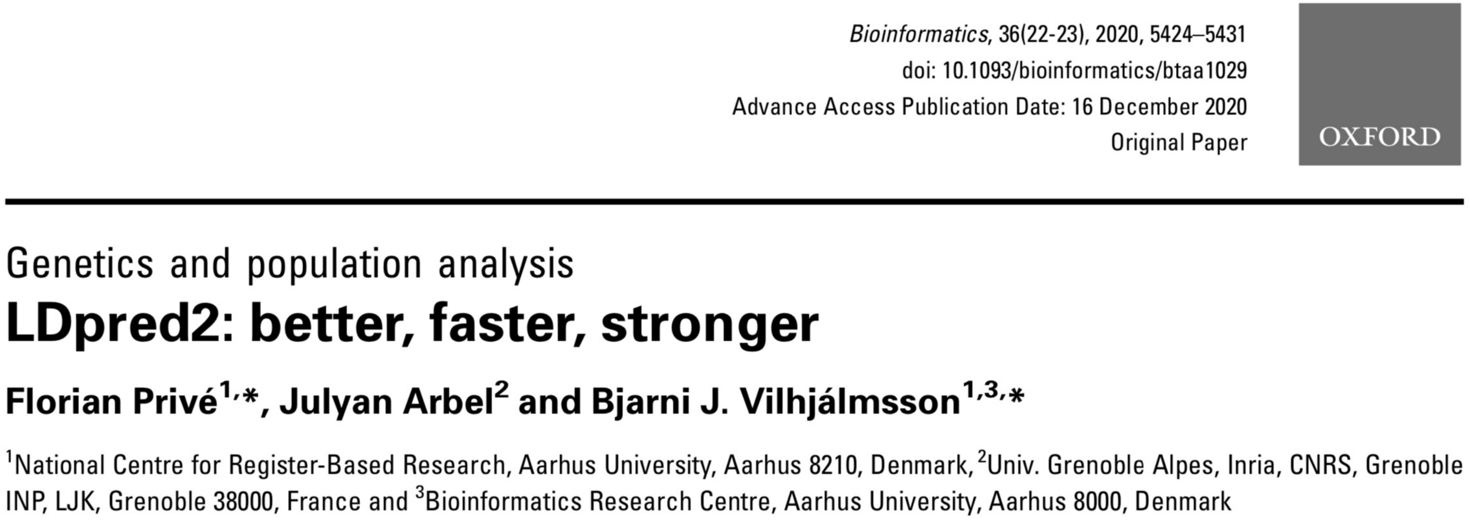
- **In general:** methods differ in terms of how they attempt to model genetic architecture to improve prediction accuracy

Advanced approaches

- The underlying trait distributions are in practice unknown, hence the optimal (tuning) parameters will need to be validated

- Unless pseudo-validation / ’single score’ methods are available…

Example: LDpred2



Bioinformatics, 36(22-23), 2020, 5424–5431
doi: 10.1093/bioinformatics/btaa1029
Advance Access Publication Date: 16 December 2020
Original Paper

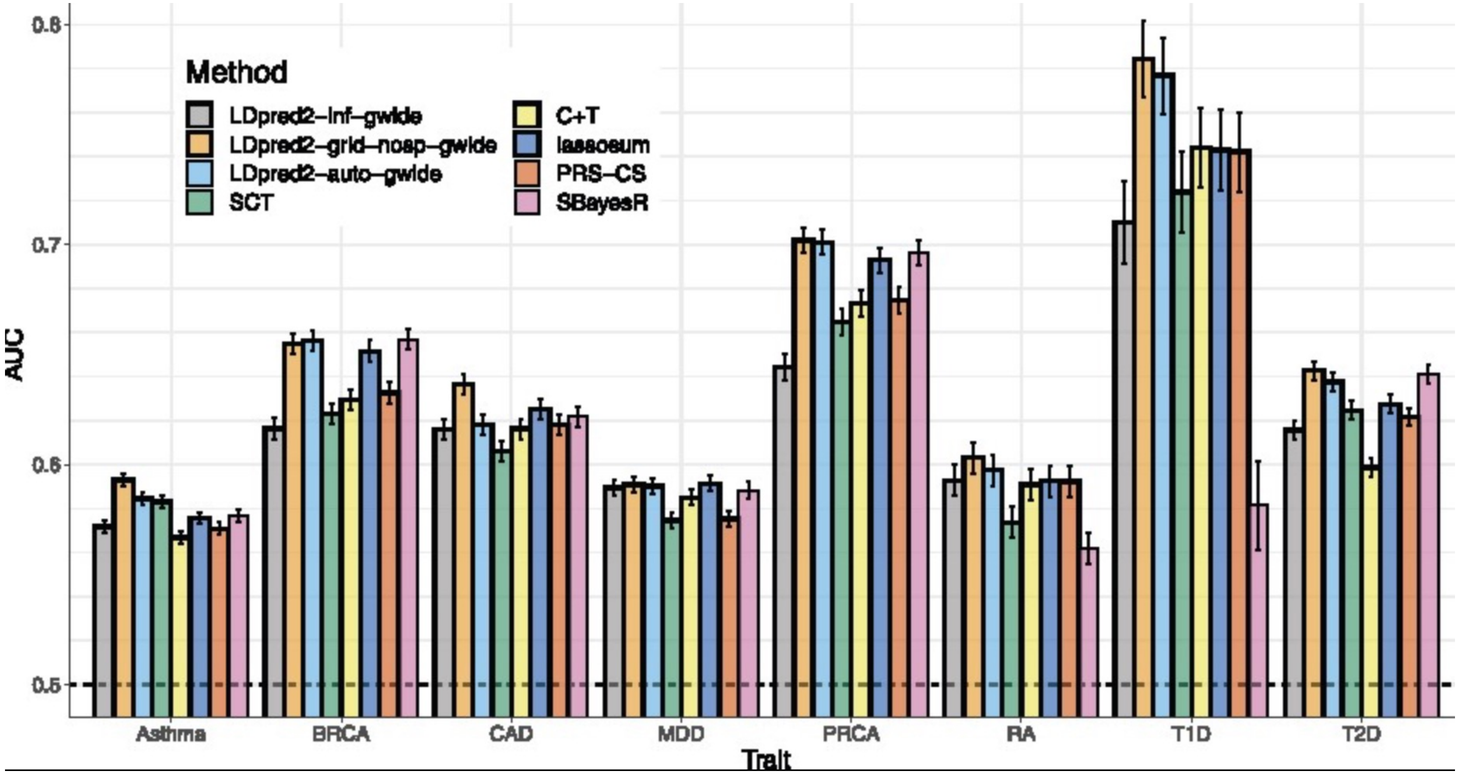
OXFORD

Genetics and population analysis

LDpred2: better, faster, stronger

Florian Privé[1,*], Julyan Arbel[2] and Bjarni J. Vilhjálmsson[1,3,*]

[1]National Centre for Register-Based Research, Aarhus University, Aarhus 8210, Denmark, [2]Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, Grenoble 38000, France and [3]Bioinformatics Research Centre, Aarhus University, Aarhus 8000, Denmark



Example: LDpred2

joint effects given marginal effects and correlation between SNPs:

$$\widehat{\gamma}_{\text{joint}} = \boldsymbol{S}^{-1}\boldsymbol{R}^{-1}\boldsymbol{S}\widehat{\gamma}_{\text{marg}} \,.$$

R =

only a specific fraction of markers is assumed to be involved in the trait and drawn from a normal distribution, while the rest is fixed to 0:

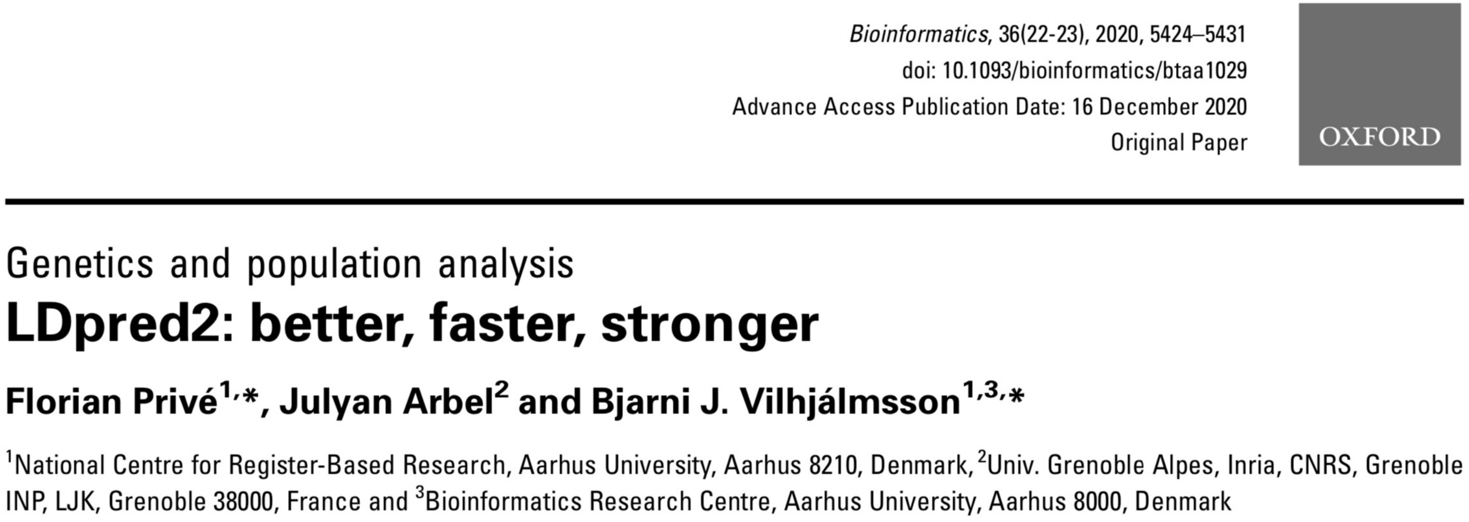
$$\beta_j = S_{j,j}\gamma_j \sim \begin{cases} \mathcal{N}\left(0, \dfrac{h^2}{Mp}\right) & \text{with probability p,} \\ 0 & \text{otherwise,} \end{cases}$$

M = variants

p = fraction of causal variants

h^2 = SNPh^2

hyper-parameter *p* (1, 0.3, 0.1, 0.03, 0.01, 0.003 and 0.001)



Bioinformatics, 36(22-23), 2020, 5424–5431
doi: 10.1093/bioinformatics/btaa1029
Advance Access Publication Date: 16 December 2020
Original Paper

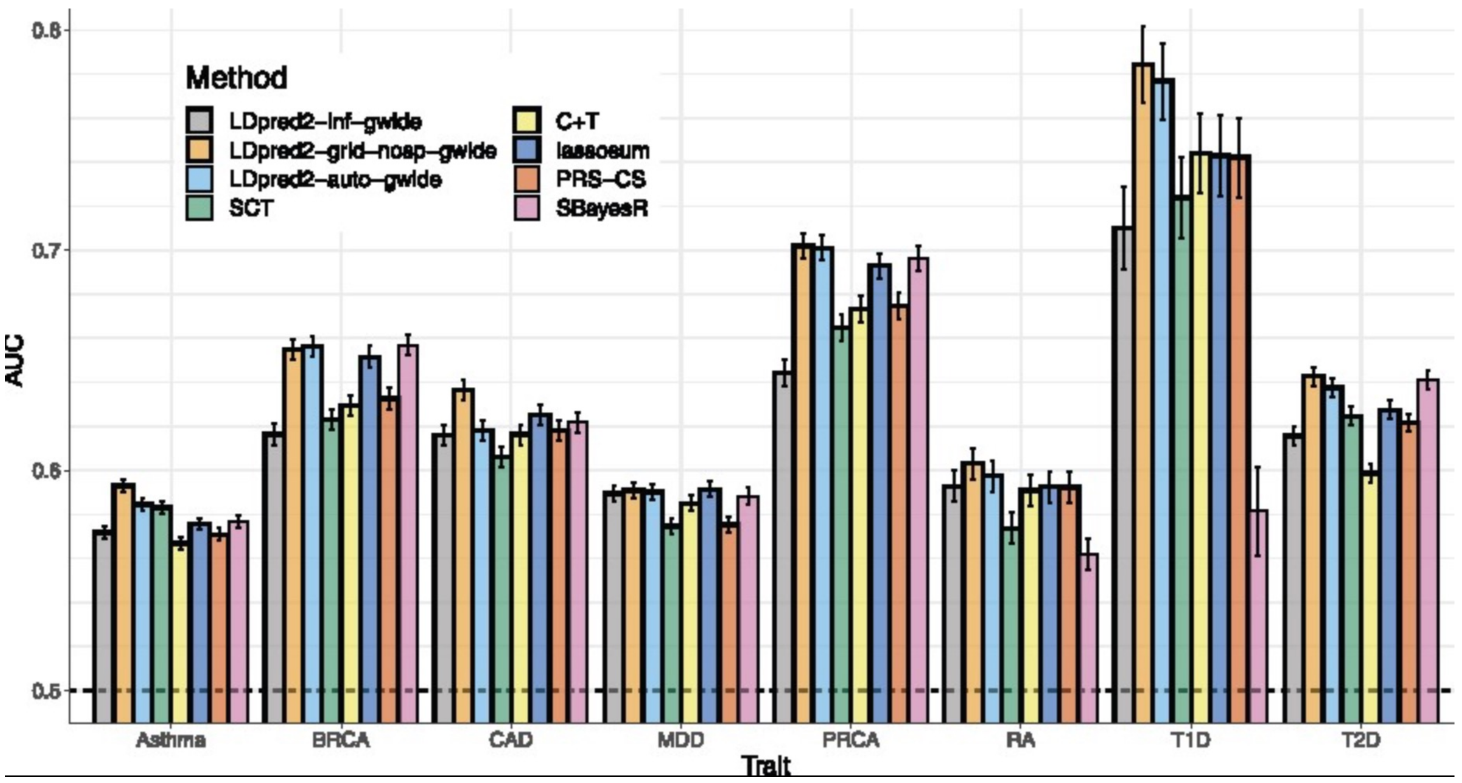
OXFORD

Genetics and population analysis

LDpred2: better, faster, stronger

Florian Privé[1,*], Julyan Arbel[2] and Bjarni J. Vilhjálmsson[1,3,*]

[1]National Centre for Register-Based Research, Aarhus University, Aarhus 8210, Denmark, [2]Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, Grenoble 38000, France and [3]Bioinformatics Research Centre, Aarhus University, Aarhus 8000, Denmark



Example: LDpred2

joint effects given marginal effects and correlation between SNPs:

$$\widehat{\boldsymbol{\gamma}}_{\text{joint}} = \boldsymbol{S}^{-1}\boldsymbol{R}^{-1}\boldsymbol{S}\widehat{\boldsymbol{\gamma}}_{\text{marg}} \, .$$

R =

only a specific fraction of markers is assumed to be involved in the trait and drawn from a normal distribution, while the rest is fixed to 0:

$$\beta_j = S_{j,j}\gamma_j \sim \begin{cases} \mathcal{N}\left(0, \dfrac{h^2}{Mp}\right) & \text{with probability p,} \\ 0 & \text{otherwise,} \end{cases}$$

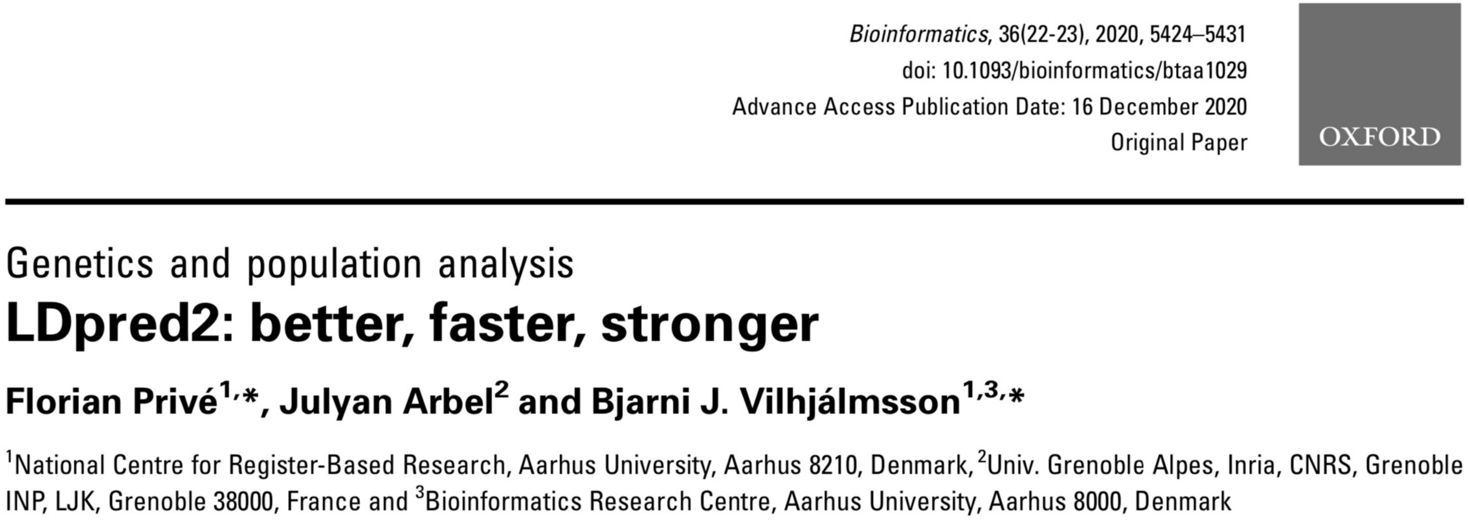
M = variants

p = fraction of causal variants

h^2 = SNPh^2

hyper-parameter p (1, 0.3, 0.1, 0.03, 0.01, 0.003 and 0.001)

LDpred2-auto: doesn't require explicit validation!



Bioinformatics, 36(22-23), 2020, 5424–5431
doi: 10.1093/bioinformatics/btaa1029
Advance Access Publication Date: 16 December 2020
Original Paper

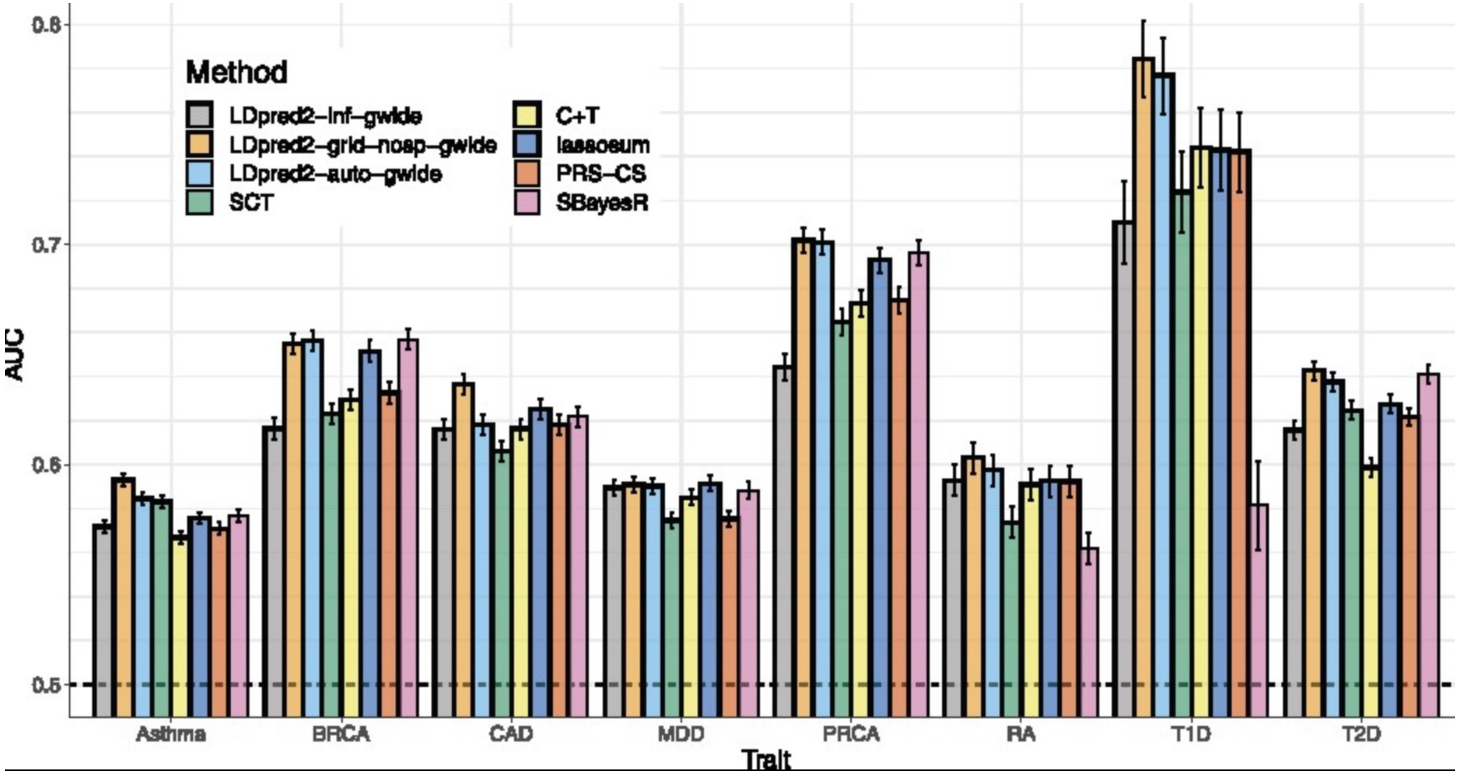
OXFORD

Genetics and population analysis

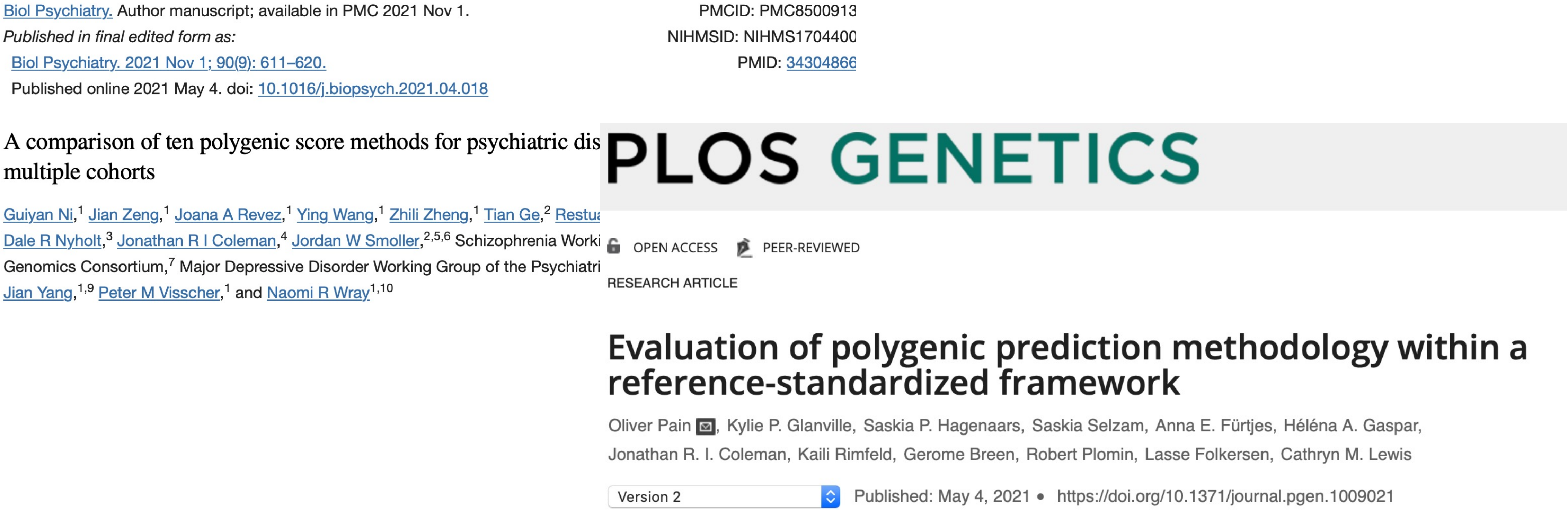
LDpred2: better, faster, stronger

Florian Privé[1,*], Julyan Arbel[2] and Bjarni J. Vilhjálmsson[1,3,*]

[1]National Centre for Register-Based Research, Aarhus University, Aarhus 8210, Denmark, [2]Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, Grenoble 38000, France and [3]Bioinformatics Research Centre, Aarhus University, Aarhus 8000, Denmark



What method works best?



Biol Psychiatry. Author manuscript; available in PMC 2021 Nov 1.
Published in final edited form as:
Biol Psychiatry. 2021 Nov 1; 90(9): 611–620.
Published online 2021 May 4. doi: 10.1016/j.biopsych.2021.04.018

PMCID: PMC8500913
NIHMSID: NIHMS1704400
PMID: 34304866

A comparison of ten polygenic score methods for psychiatric dis
multiple cohorts

Guiyan Ni,[1] Jian Zeng,[1] Joana A Revez,[1] Ying Wang,[1] Zhili Zheng,[1] Tian Ge,[2] Restu
Dale R Nyholt,[3] Jonathan R I Coleman,[4] Jordan W Smoller,[2,5,6] Schizophrenia Worki
Genomics Consortium,[7] Major Depressive Disorder Working Group of the Psychiatri
Jian Yang,[1,9] Peter M Visscher,[1] and Naomi R Wray[1,10]

PLOS GENETICS

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

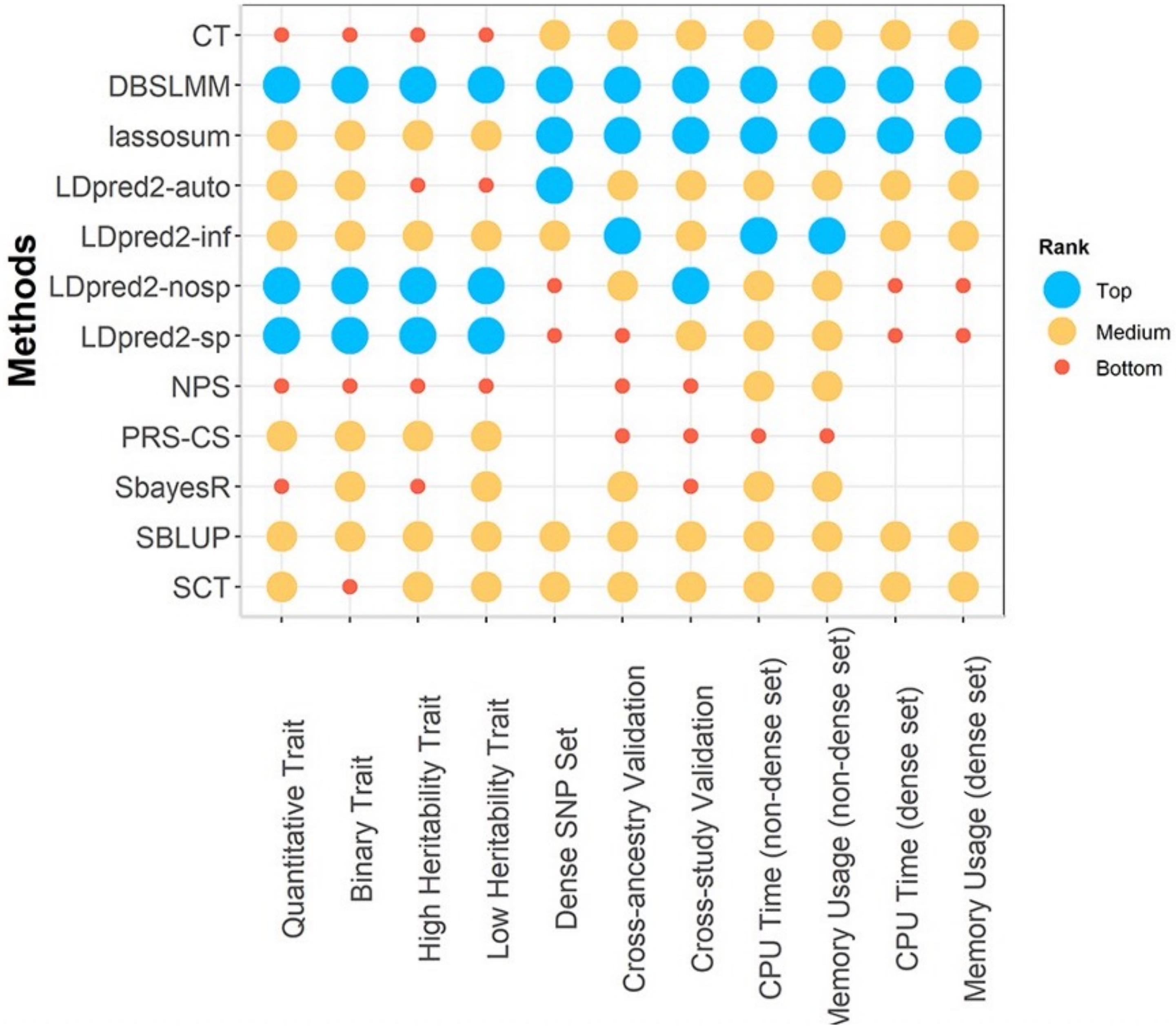
Evaluation of polygenic prediction methodology within a reference-standardized framework

Oliver Pain, Kylie P. Glanville, Saskia P. Hagenaars, Saskia Selzam, Anna E. Fürtjes, Héléna A. Gaspar, Jonathan R. I. Coleman, Kaili Rimfeld, Gerome Breen, Robert Plomin, Lasse Folkersen, Cathryn M. Lewis

Version 2

Published: May 4, 2021 • https://doi.org/10.1371/journal.pgen.1009021

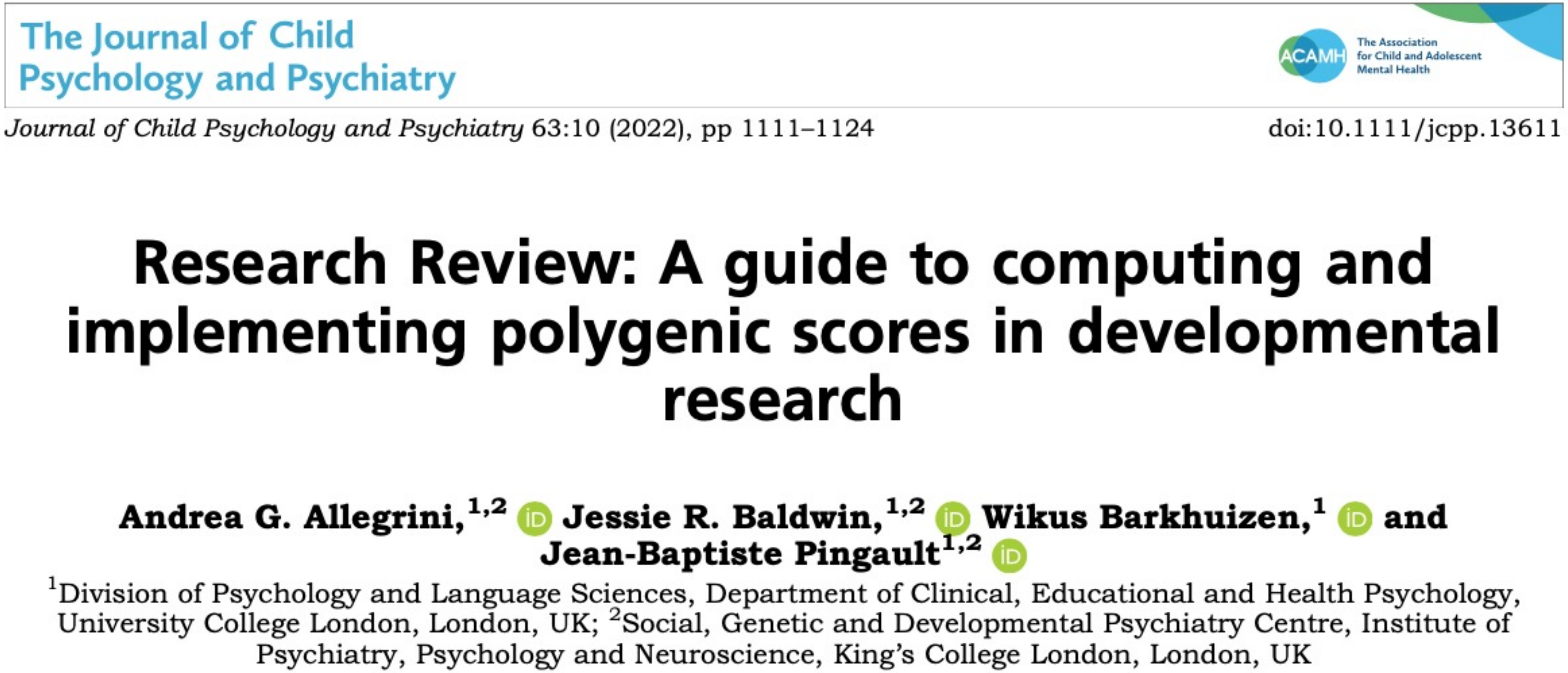
In general: **no dramatic differences**, although more nuanced results emerged depending on specific applications and settings (e.g. diverse genetic architectures).



Yang, S., & Zhou, X. (2022) *Briefings in Bioinformatics*, **23**, 39.

Table S2. General resources for PGS workflow

Resource	Reference	Brief description	link
Polygenic score catalog	(Lambert et al., 2021)	Database of PGS employed in published work providing relevant metadata to develop and evaluate them in different datasets.	https://www.pgscatalog.org
Polygenic index repository	(Becker et al., 2021)	PGS repository providing metadata to reproduce PGS, or already constructed PGS for a number of cohorts. PGS are obtained from a reference standardized and optimized pipeline.	https://www.thessgac.org/pgi-repository
Open GWAS	(Elsworth et al., 2020)	A curated collection of GWAS summary statistics.	https://gwas.mrcieu.ac.uk
GWAS catalog	(MacArthur et al., 2017)	A curated catalog of GWAS results and summary statistics.	https://www.ebi.ac.uk/gwas/
GWAS Atlas	(Watanabe et al., 2019)	Database of GWAS results and downstream analyses.	https://atlas.ctglab.nl
PGS Atlas	(Richardson et al., 2019)	Atlas of PGS – phenotype associations across 162 PGS and 551 traits.	http://mrcieu.mrsoftware.org/PRS_atlas/
GenoPred	(Pain et al., 2021)	A workflow for evaluating PGS methods within a reference standardized framework.	https://opain.github.io/GenoPred/ https://github.com/opain/GenoPred/tree/master/GenoPredPipe



The Journal of Child Psychology and Psychiatry

ACAMH The Association for Child and Adolescent Mental Health

Journal of Child Psychology and Psychiatry 63:10 (2022), pp 1111–1124 doi:10.1111/jcpp.13611

Research Review: A guide to computing and implementing polygenic scores in developmental research

Andrea G. Allegrini,[1,2] Jessie R. Baldwin,[1,2] Wikus Barkhuizen,[1] and Jean-Baptiste Pingault[1,2]

[1]Division of Psychology and Language Sciences, Department of Clinical, Educational and Health Psychology, University College London, London, UK; [2]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

Applications of PGS

- Research tools!

- Inferring genetic overlap between traits

- Risk stratification

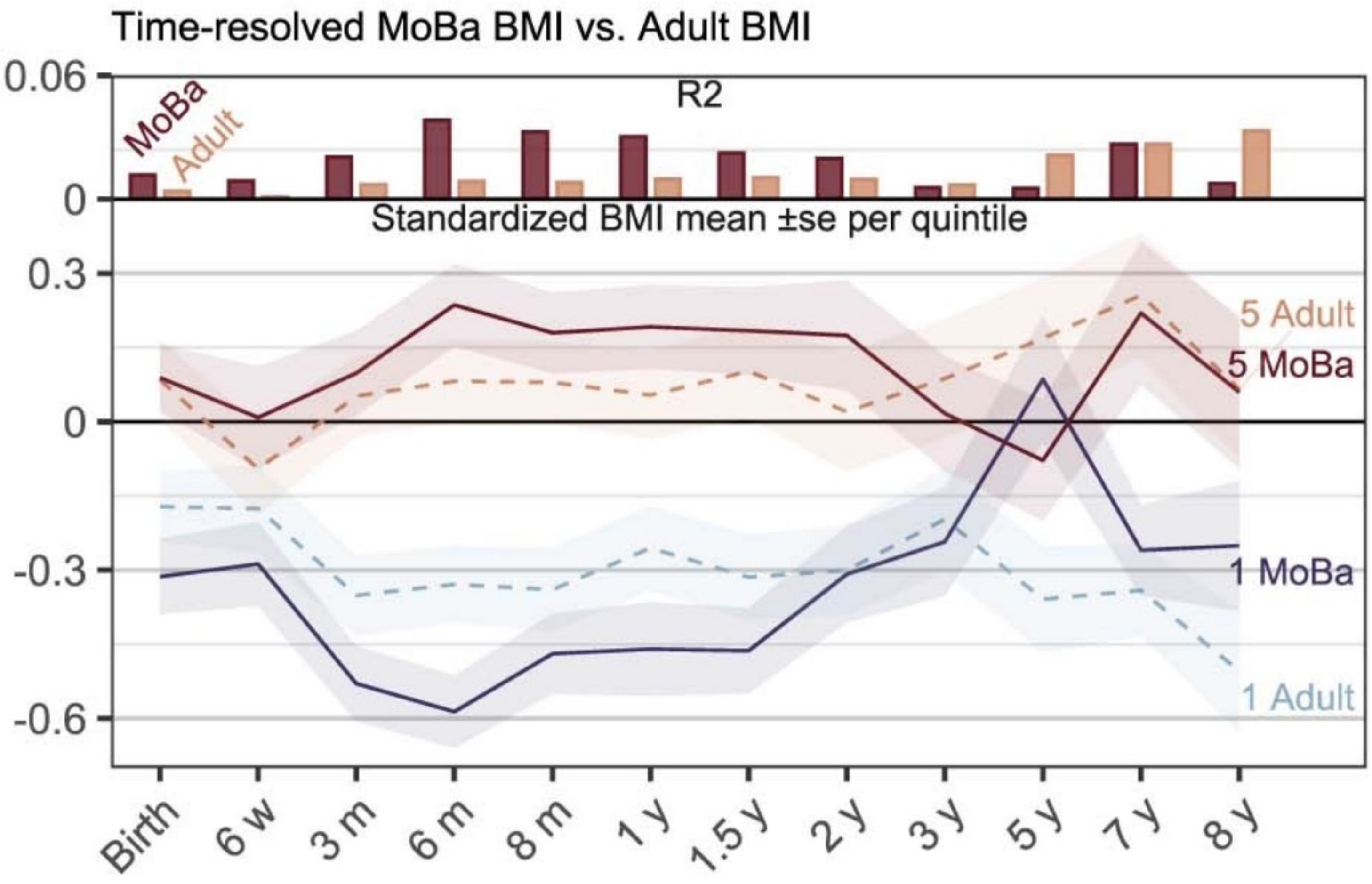
- Can be employed in (clinical) prediction models – utility still limited at present

Applications of PGS

- GWAS N is central, but bigger is not always better

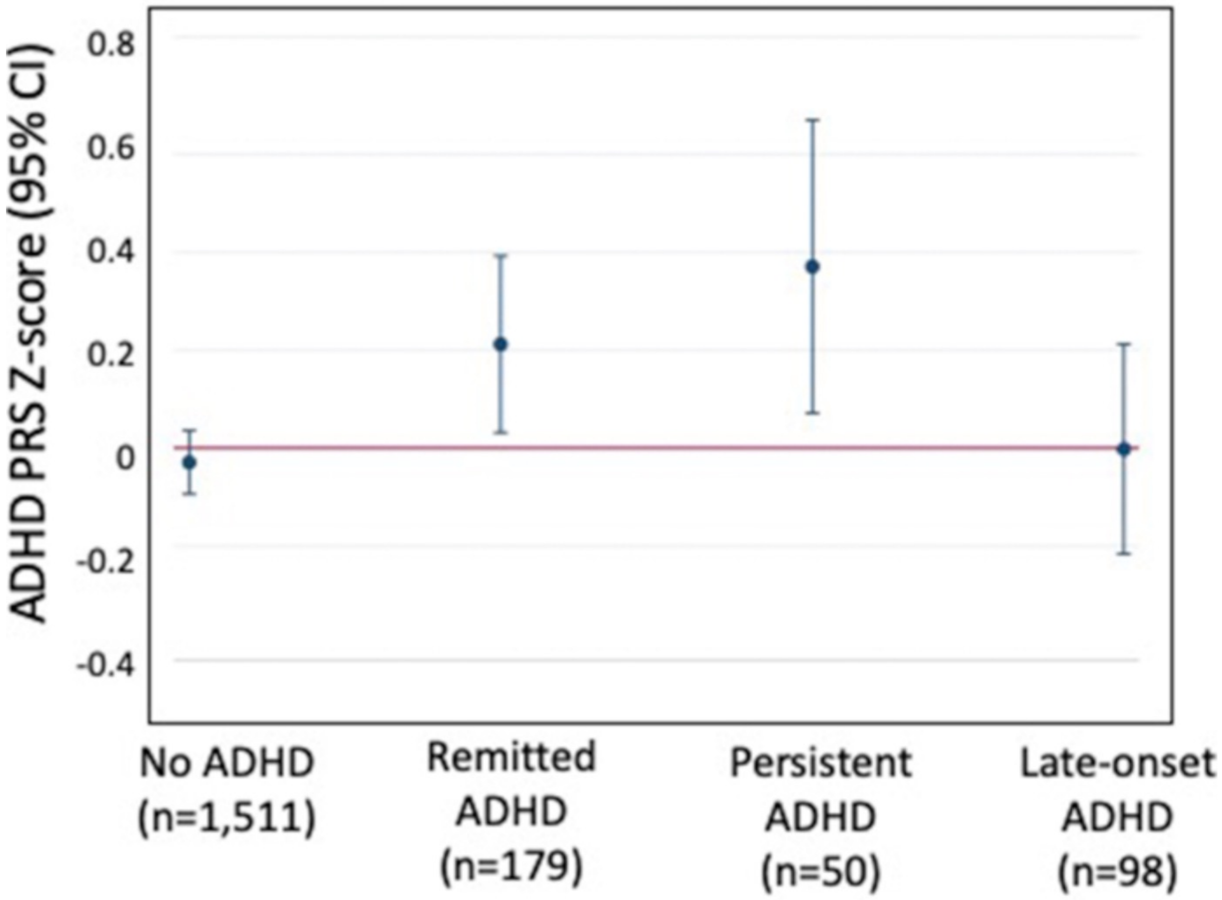
- The meaning of the PGS depends on the phenotypic definition employed in GWAS!

Example: PRS from developmental specific BMI GWASes (max N ~30k) perform better than adult BMI GWAS (N~700k)



Helgeland et al., 2022. Nature Metabolism volume 4, pages 344–358 (2022)

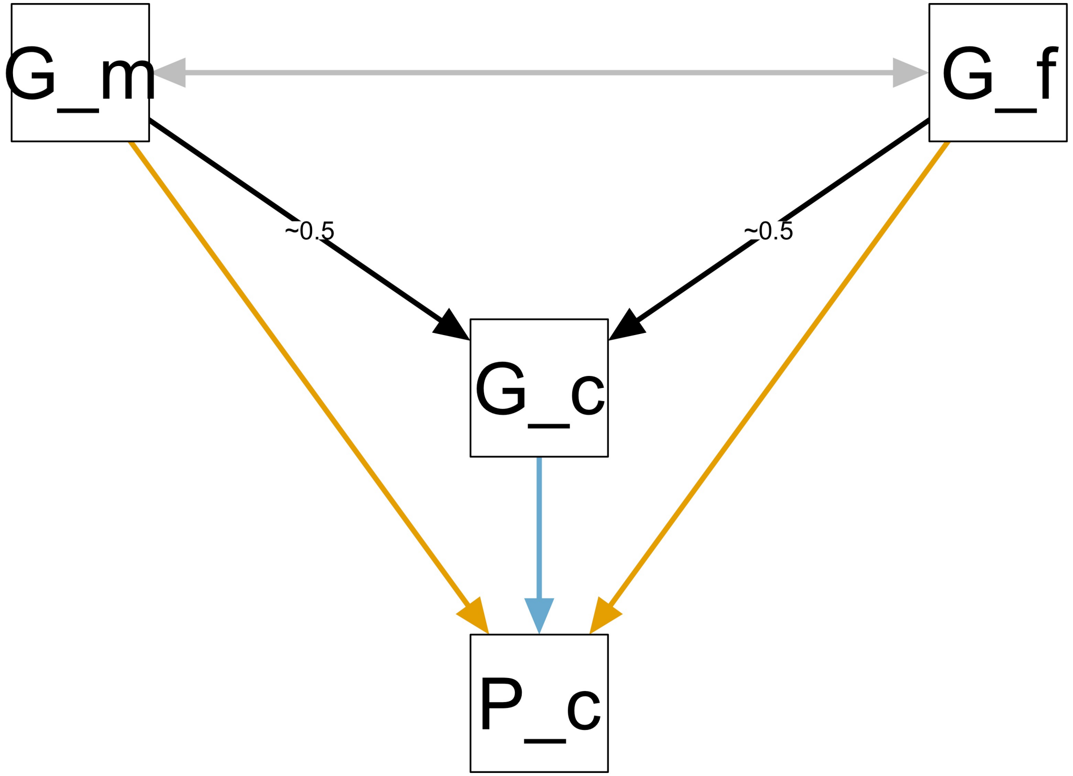
Example: PRS based on child case–control diagnosis of ADHD misses the full (genetic) complexity of the disorder across the lifespan



Agnew-Blais et al., 2021. J Am Acad Child Adolesc Psychiatry. 2021 Sep; 60(9): 1147–1156.

The trio design

The trio design

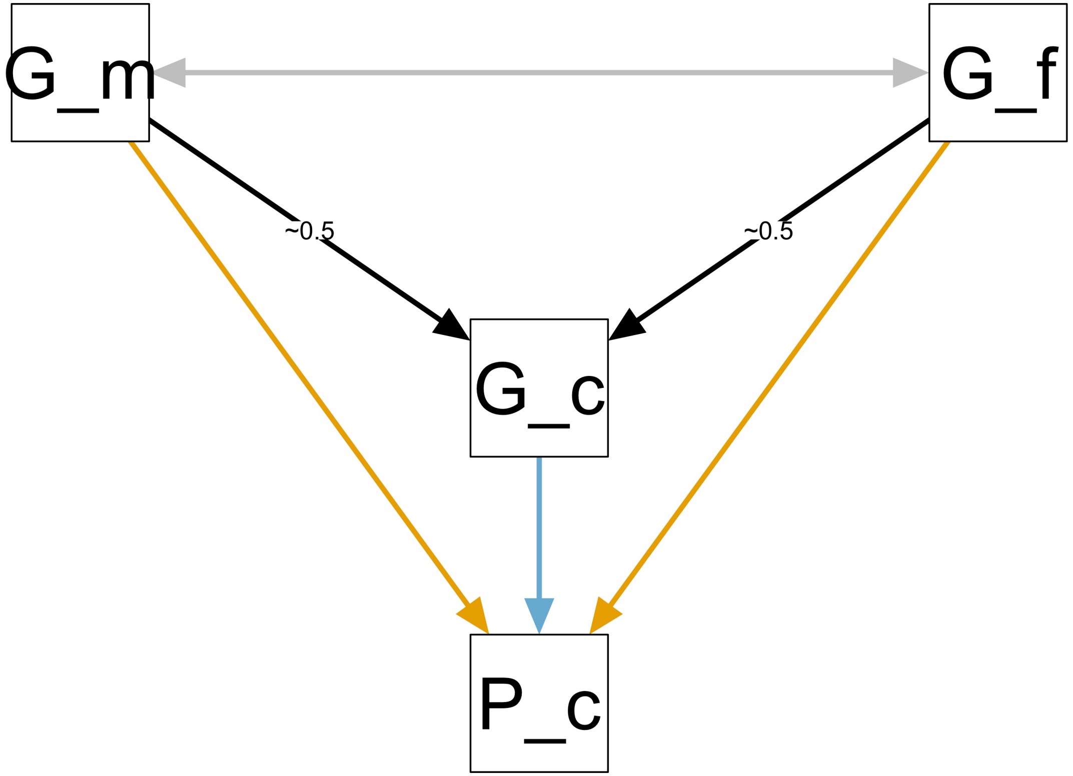


The trio design

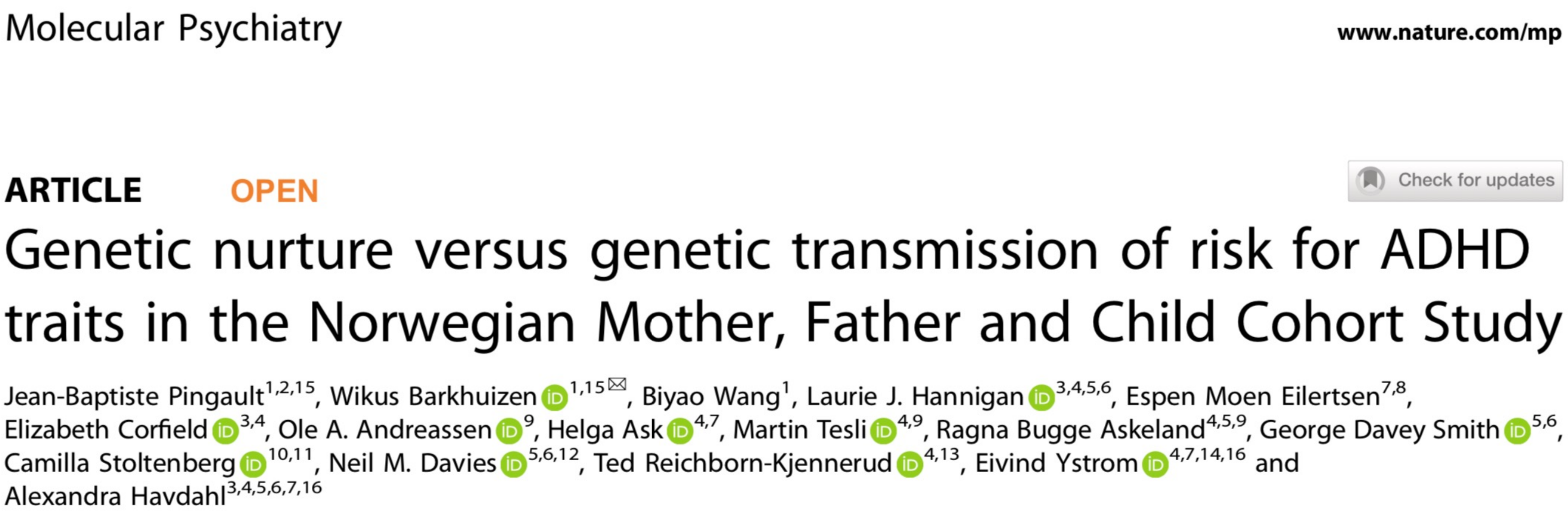
- “Direct” genetic effects

- “Indirect” genetic effects: rearing environments, dynastic effects, population structure

- PGS-phenotype associations potentially inflated, or completely accounted for, by indirect processes (Veller and Coop, 2023)



Example



Molecular Psychiatry

www.nature.com/mp

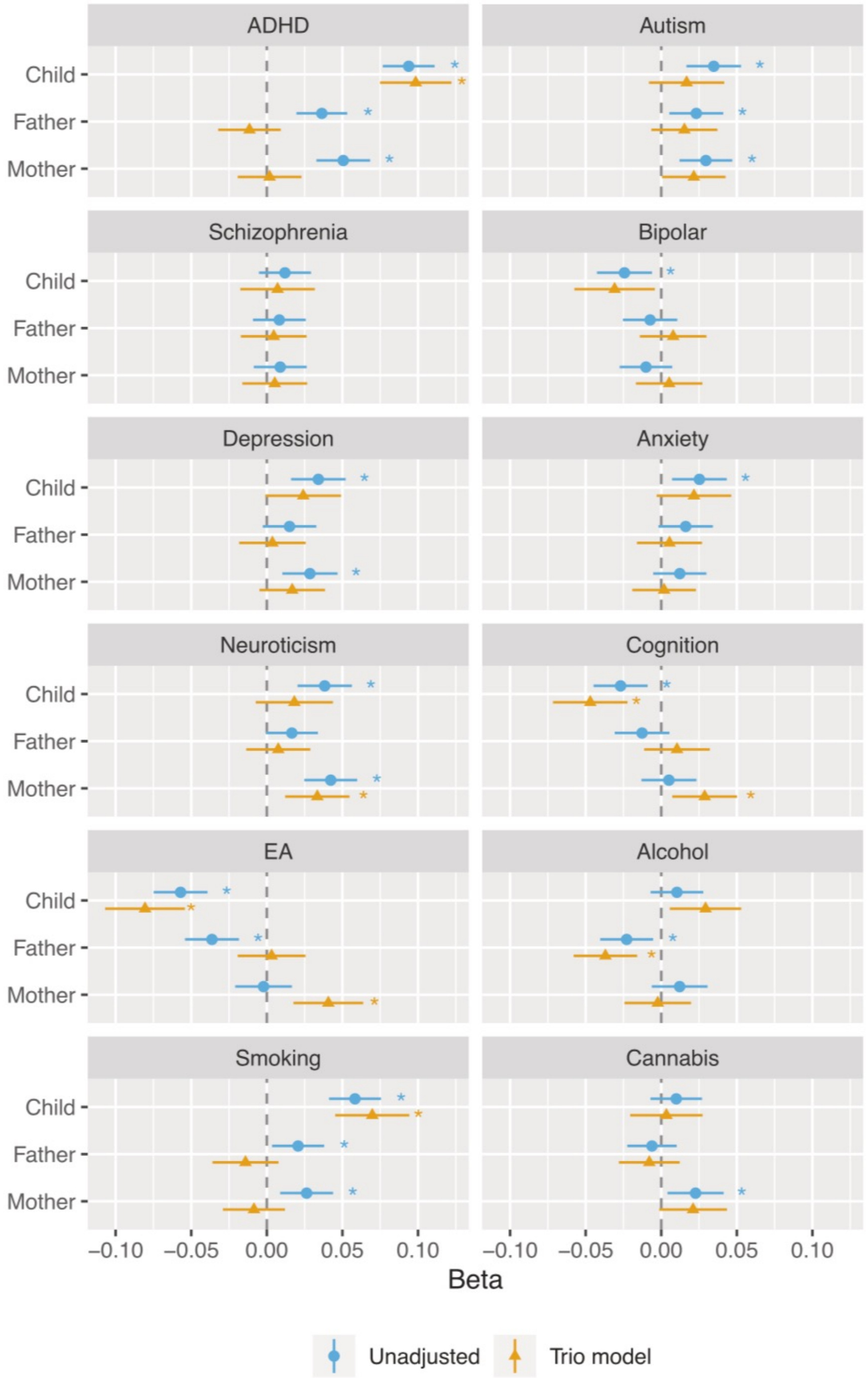
ARTICLE OPEN

Check for updates

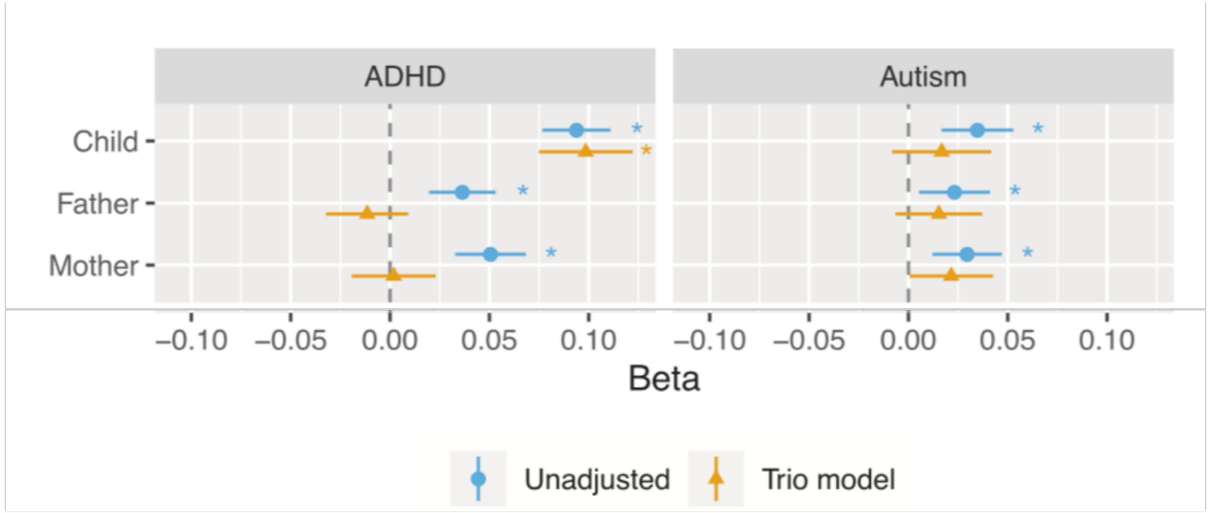
Genetic nurture versus genetic transmission of risk for ADHD traits in the Norwegian Mother, Father and Child Cohort Study

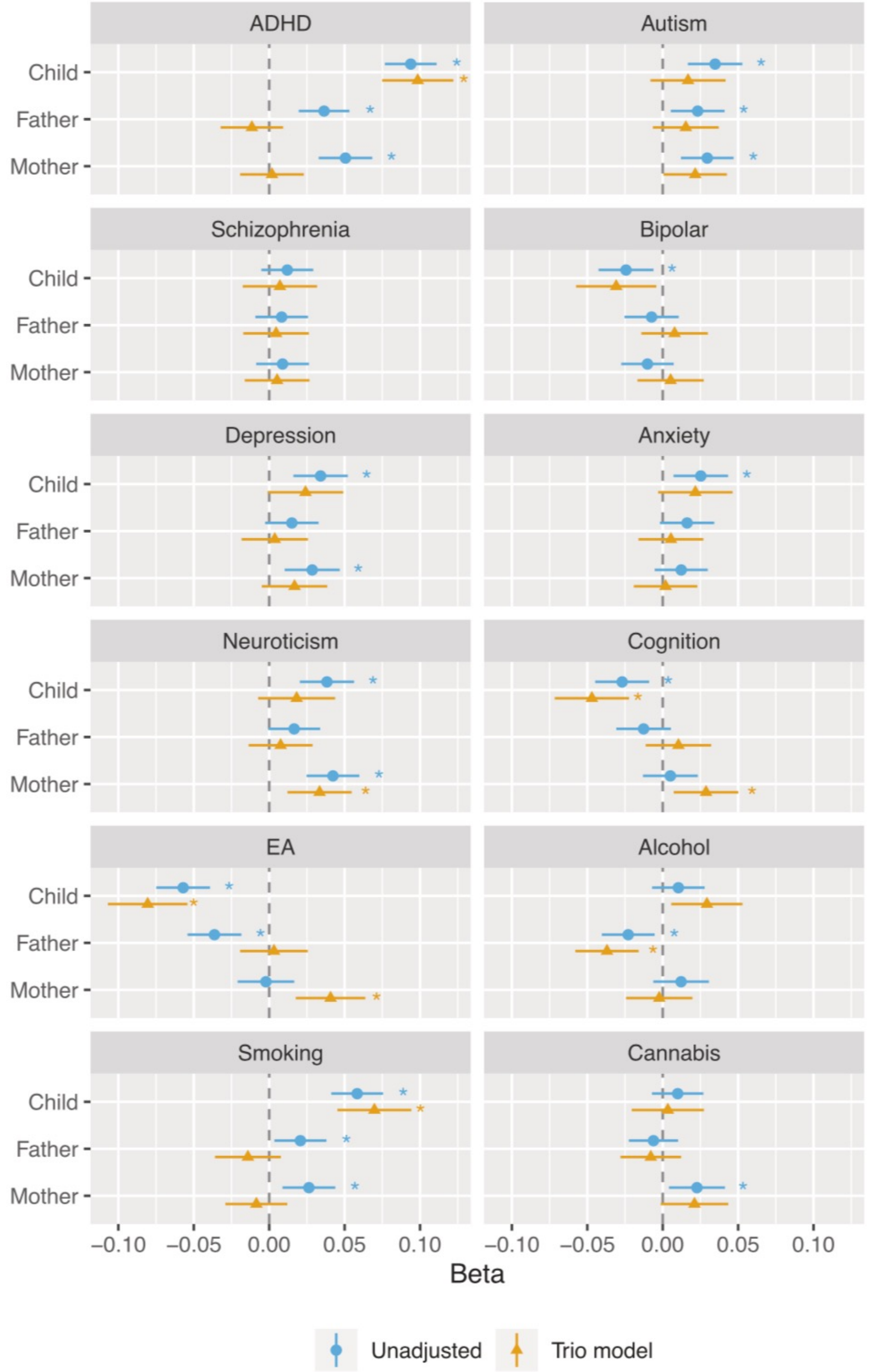
Jean-Baptiste Pingault[1,2,15], Wikus Barkhuizen[1,15] ✉, Biyao Wang[1], Laurie J. Hannigan[3,4,5,6], Espen Moen Eilertsen[7,8], Elizabeth Corfield[3,4], Ole A. Andreassen[9], Helga Ask[4,7], Martin Tesli[4,9], Ragna Bugge Askeland[4,5,9], George Davey Smith[5,6], Camilla Stoltenberg[10,11], Neil M. Davies[5,6,12], Ted Reichborn-Kjennerud[4,13], Eivind Ystrom[4,7,14,16] and Alexandra Havdahl[3,4,5,6,7,16]

Example


ADHD
Autism
Schizophrenia
Bipolar
Depression
Anxiety
Neuroticism
Cognition
EA
Alcohol
Smoking
Cannabis
Child
Father
Mother
*
−0.10
−0.05
0.00
0.05
0.10
Beta
Unadjusted
Trio model

Example

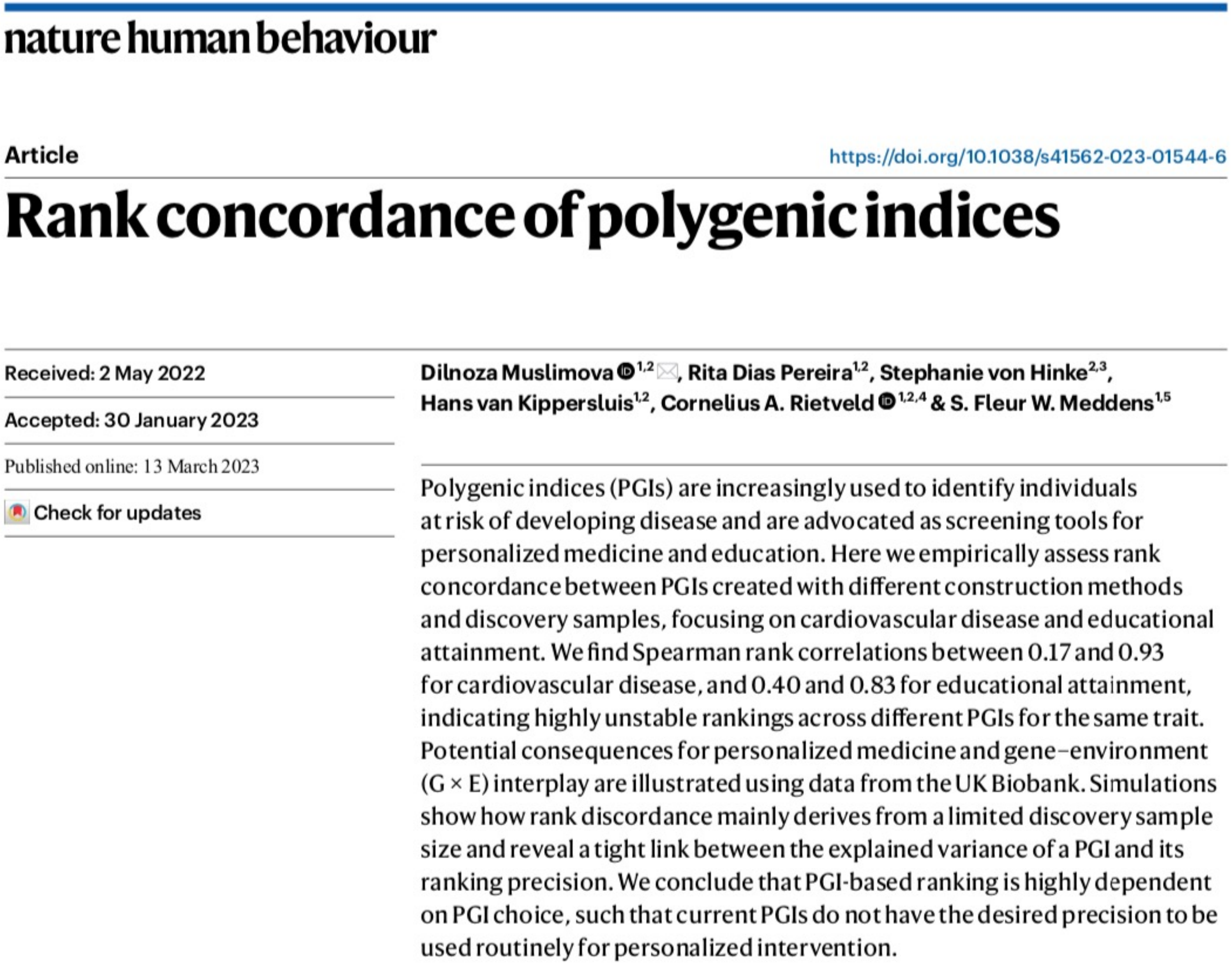

ADHD
Autism
Child
Father
Mother
*
−0.10
−0.05
0.00
0.05
0.10
Beta
Unadjusted
Trio model


ADHD
Autism
Schizophrenia
Bipolar
Depression
Anxiety
Neuroticism
Cognition
EA
Alcohol
Smoking
Cannabis
Child
Father
Mother
*
−0.10
−0.05
0.00
0.05
0.10
Beta
Unadjusted
Trio model

Caveats/considerations

- Personalized intervention

- Cross-ancestry portability



nature human behaviour

Article

https://doi.org/10.1038/s41562-023-01544-6

Rank concordance of polygenic indices

Received: 2 May 2022

Accepted: 30 January 2023

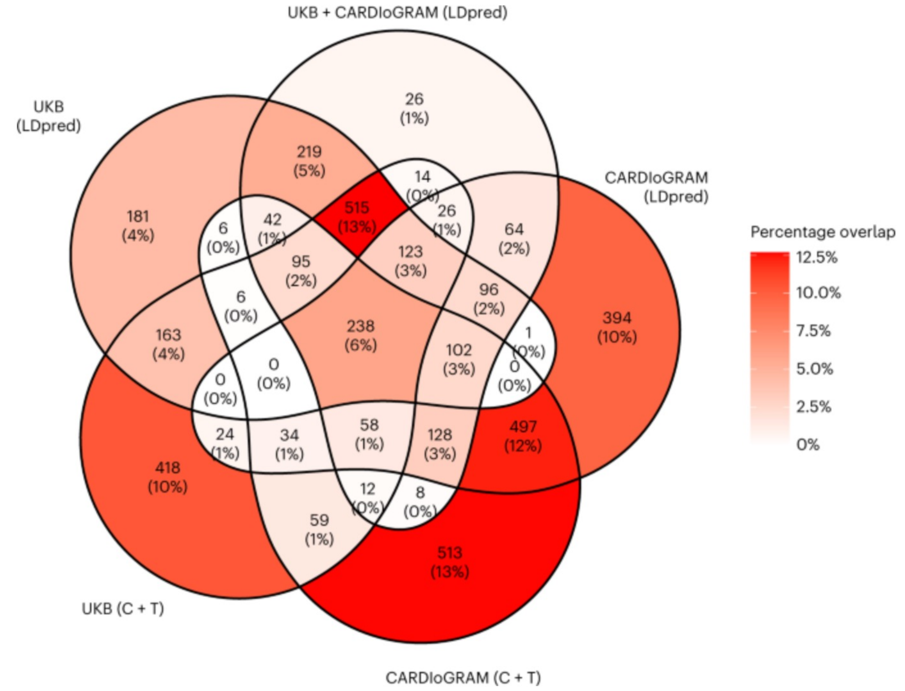
Published online: 13 March 2023

Check for updates

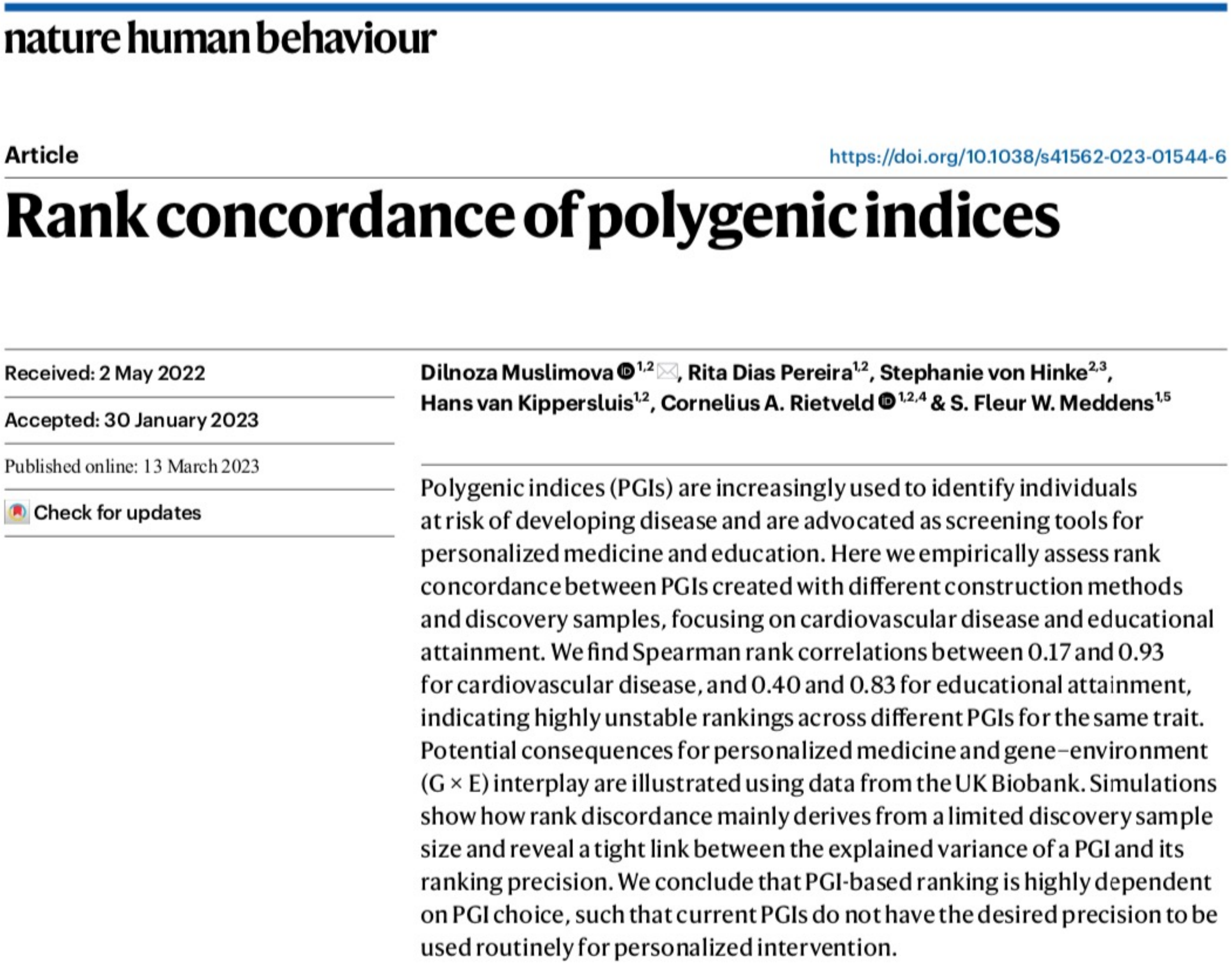
Dilnoza Muslimova[1,2], Rita Dias Pereira[1,2], Stephanie von Hinke[2,3], Hans van Kippersluis[1,2], Cornelius A. Rietveld[1,2,4] & S. Fleur W. Meddens[1,5]

Polygenic indices (PGIs) are increasingly used to identify individuals at risk of developing disease and are advocated as screening tools for personalized medicine and education. Here we empirically assess rank concordance between PGIs created with different construction methods and discovery samples, focusing on cardiovascular disease and educational attainment. We find Spearman rank correlations between 0.17 and 0.93 for cardiovascular disease, and 0.40 and 0.83 for educational attainment, indicating highly unstable rankings across different PGIs for the same trait. Potential consequences for personalized medicine and gene–environment (G × E) interplay are illustrated using data from the UK Biobank. Simulations show how rank discordance mainly derives from a limited discovery sample size and reveal a tight link between the explained variance of a PGI and its ranking precision. We conclude that PGI-based ranking is highly dependent on PGI choice, such that current PGIs do not have the desired precision to be used routinely for personalized intervention.

Fig. 3: Venn diagram depicting the overlap in individuals ranked in the top quintiles of five CVD PGIs (N = 4,061).



Individuals included in this figure are potential candidates for statin therapy[22]: they have an intermediate 10 year ASCVD risk (≥5%); have no (self-reported) history of CVD; are not statin users; and are not yet candidates according to current ACC/AHA guidelines.



nature human behaviour

Article https://doi.org/10.1038/s41562-023-01544-6

Rank concordance of polygenic indices

Received: 2 May 2022

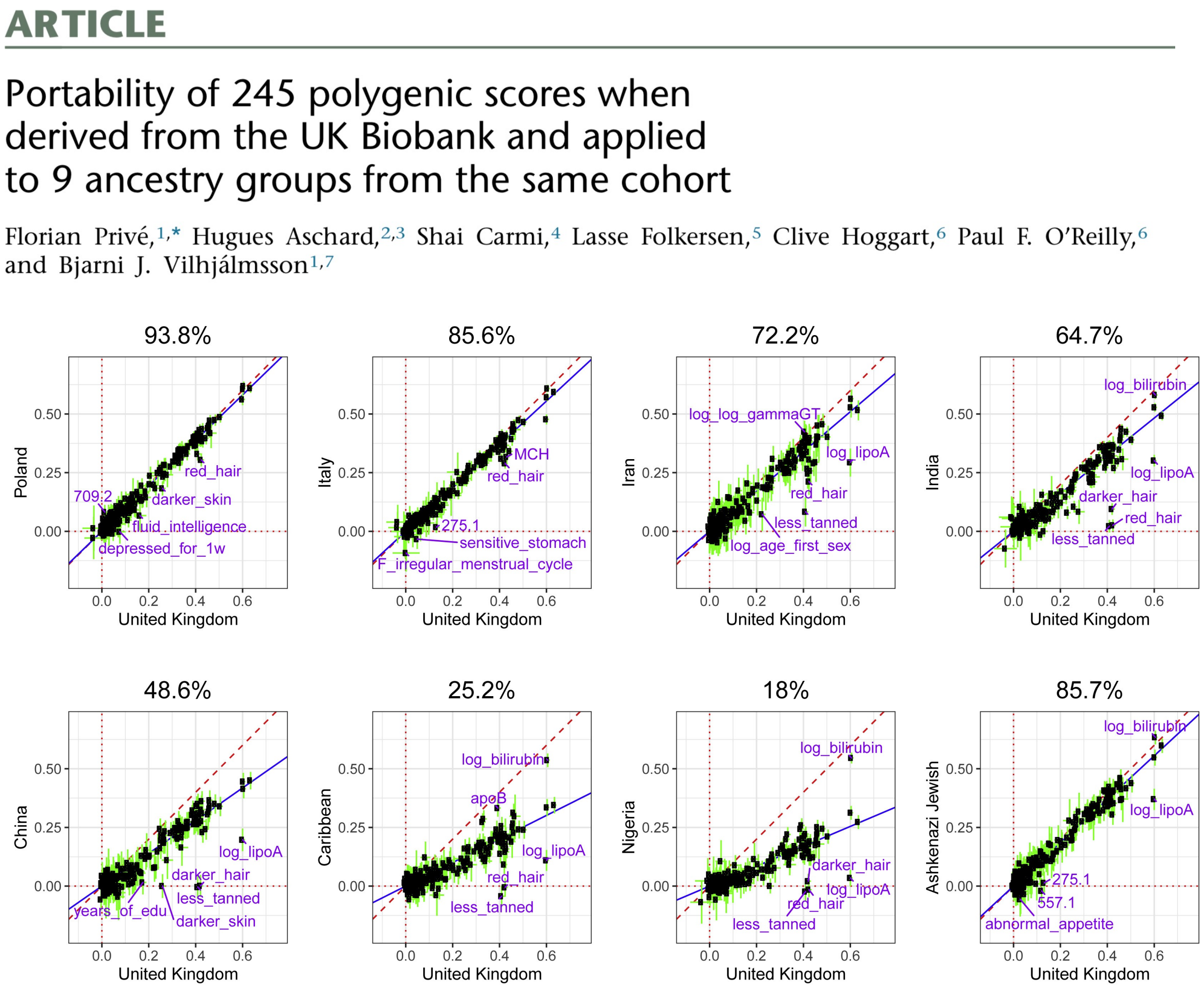
Accepted: 30 January 2023

Published online: 13 March 2023

Check for updates

Dilnoza Muslimova[1,2], Rita Dias Pereira[1,2], Stephanie von Hinke[2,3], Hans van Kippersluis[1,2], Cornelius A. Rietveld[1,2,4] & S. Fleur W. Meddens[1,5]

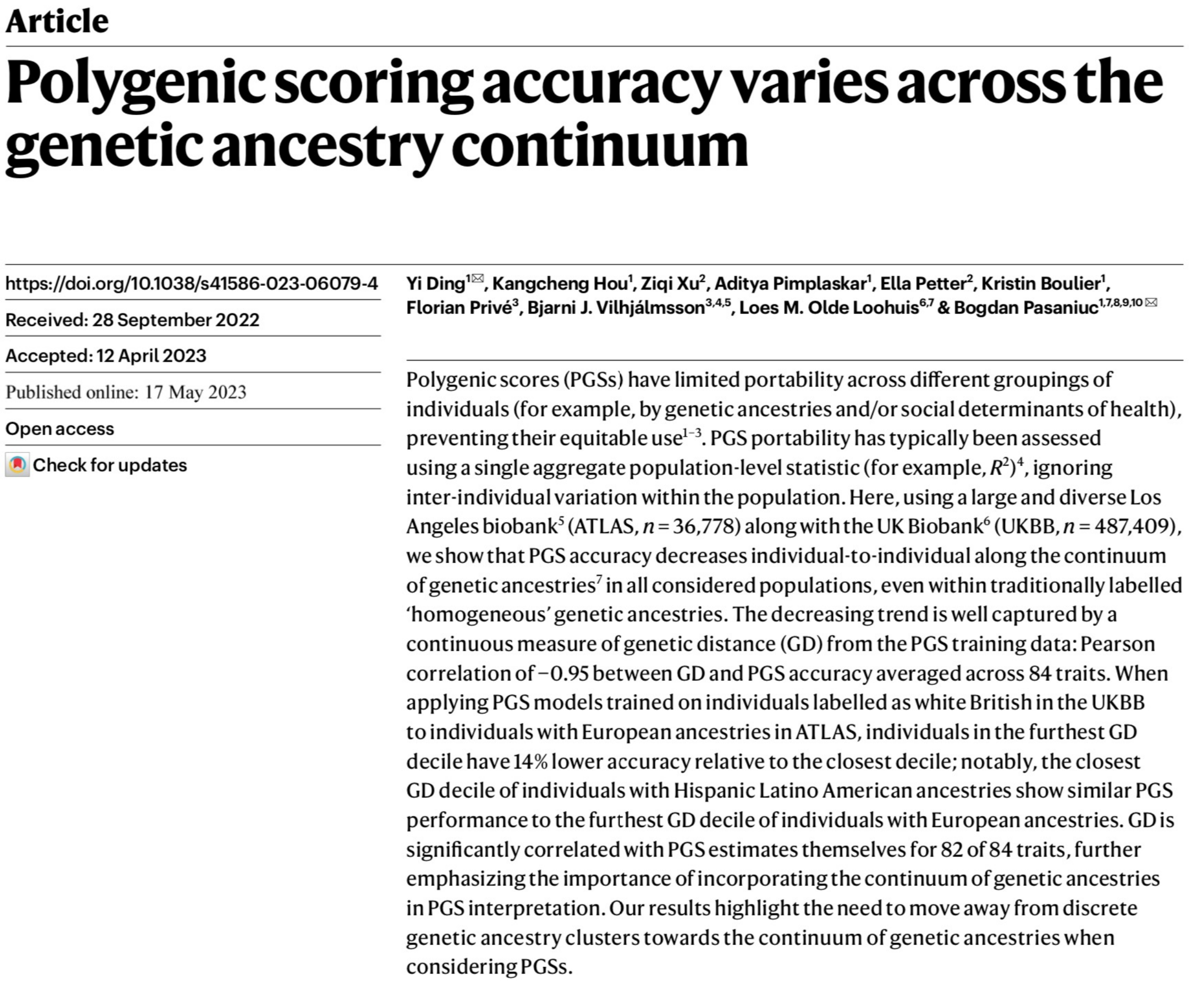
Polygenic indices (PGIs) are increasingly used to identify individuals at risk of developing disease and are advocated as screening tools for personalized medicine and education. Here we empirically assess rank concordance between PGIs created with different construction methods and discovery samples, focusing on cardiovascular disease and educational attainment. We find Spearman rank correlations between 0.17 and 0.93 for cardiovascular disease, and 0.40 and 0.83 for educational attainment, indicating highly unstable rankings across different PGIs for the same trait. Potential consequences for personalized medicine and gene–environment (G × E) interplay are illustrated using data from the UK Biobank. Simulations show how rank discordance mainly derives from a limited discovery sample size and reveal a tight link between the explained variance of a PGI and its ranking precision. We conclude that PGI-based ranking is highly dependent on PGI choice, such that current PGIs do not have the desired precision to be used routinely for personalized intervention.



ARTICLE

Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort

Florian Privé,[1,*] Hugues Aschard,[2,3] Shai Carmi,[4] Lasse Folkersen,[5] Clive Hoggart,[6] Paul F. O'Reilly,[6] and Bjarni J. Vilhjálmsson[1,7]



Article

Polygenic scoring accuracy varies across the genetic ancestry continuum

https://doi.org/10.1038/s41586-023-06079-4

Received: 28 September 2022

Accepted: 12 April 2023

Published online: 17 May 2023

Open access

Check for updates

Yi Ding[1]✉, Kangcheng Hou[1], Ziqi Xu[2], Aditya Pimplaskar[1], Ella Petter[2], Kristin Boulier[1], Florian Privé[3], Bjarni J. Vilhjálmsson[3,4,5], Loes M. Olde Loohuis[6,7] & Bogdan Pasaniuc[1,7,8,9,10]✉

Polygenic scores (PGSs) have limited portability across different groupings of individuals (for example, by genetic ancestries and/or social determinants of health), preventing their equitable use[1–3]. PGS portability has typically been assessed using a single aggregate population-level statistic (for example, R^2)[4], ignoring inter-individual variation within the population. Here, using a large and diverse Los Angeles biobank[5] (ATLAS, n = 36,778) along with the UK Biobank[6] (UKBB, n = 487,409), we show that PGS accuracy decreases individual-to-individual along the continuum of genetic ancestries[7] in all considered populations, even within traditionally labelled 'homogeneous' genetic ancestries. The decreasing trend is well captured by a continuous measure of genetic distance (GD) from the PGS training data: Pearson correlation of −0.95 between GD and PGS accuracy averaged across 84 traits. When applying PGS models trained on individuals labelled as white British in the UKBB to individuals with European ancestries in ATLAS, individuals in the furthest GD decile have 14% lower accuracy relative to the closest decile; notably, the closest GD decile of individuals with Hispanic Latino American ancestries show similar PGS performance to the furthest GD decile of individuals with European ancestries. GD is significantly correlated with PGS estimates themselves for 82 of 84 traits, further emphasizing the importance of incorporating the continuum of genetic ancestries in PGS interpretation. Our results highlight the need to move away from discrete genetic ancestry clusters towards the continuum of genetic ancestries when considering PGSs.

Lewis and Vassos *Genome Medicine* (2020) 12:44
https://doi.org/10.1186/s13073-020-00742-5

Genome Medicine

REVIEW Open Access

Polygenic risk scores: from research tools to clinical instruments

Cathryn M. Lewis[1,2*] and Evangelos Vassos[1]

Abstract

Genome-wide association studies have shown unequivocally that common complex disorders have a polygenic genetic architecture and have enabled researchers to identify genetic variants associated with diseases. These variants can be combined into a polygenic risk score that captures part of an individual's susceptibility to diseases. Polygenic risk scores have been widely applied in research studies, confirming the association between the scores and disease status, but their clinical utility has yet to be established. Polygenic risk scores may be used to estimate an individual's lifetime genetic risk of disease, but the current discriminative ability is low in the general population. Clinical implementation of polygenic risk score (PRS) may be useful in cohorts where there is a higher prior probability of disease, for example, in early stages of diseases to assist in diagnosis or to inform treatment choices. Important considerations are the weaker evidence base in application to non-European ancestry and the challenges in translating an individual's PRS from a percentile of a normal distribution to a lifetime disease risk. In this review, we consider how PRS may be informative at different points in the disease trajectory giving examples of progress in the field and discussing obstacles that need to be addressed before clinical implementation.

Keywords: Genetics, Common disorders, Polygenic risk scores, Prediction, Pharmacogenetics, Risk

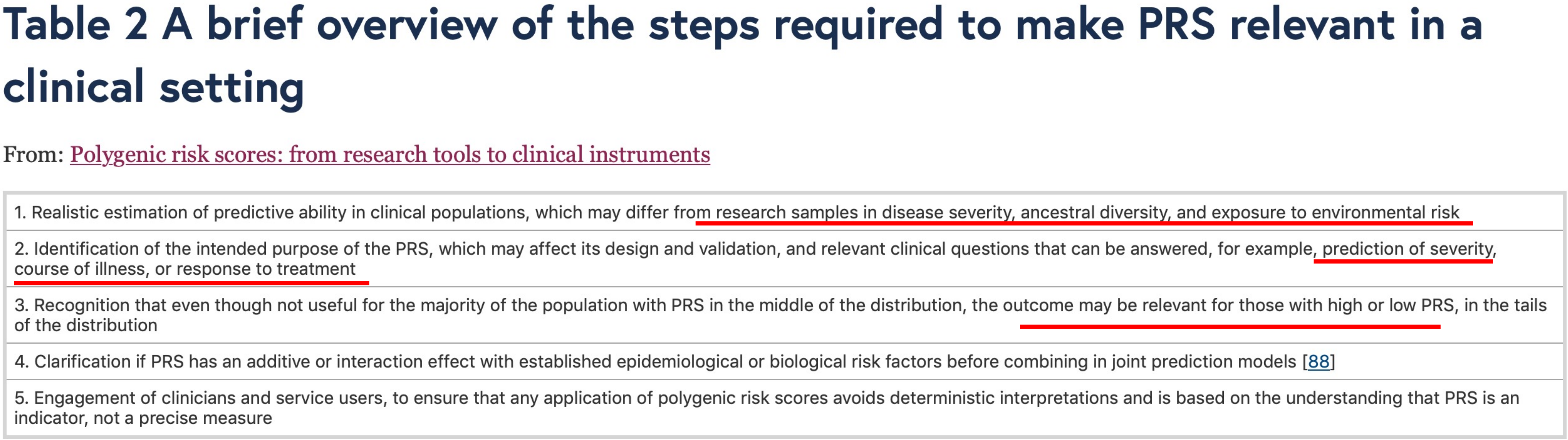


Table 2 A brief overview of the steps required to make PRS relevant in a clinical setting

From: Polygenic risk scores: from research tools to clinical instruments

1. Realistic estimation of predictive ability in clinical populations, which may differ from research samples in disease severity, ancestral diversity, and exposure to environmental risk
2. Identification of the intended purpose of the PRS, which may affect its design and validation, and relevant clinical questions that can be answered, for example, prediction of severity, course of illness, or response to treatment
3. Recognition that even though not useful for the majority of the population with PRS in the middle of the distribution, the outcome may be relevant for those with high or low PRS, in the tails of the distribution
4. Clarification if PRS has an additive or interaction effect with established epidemiological or biological risk factors before combining in joint prediction models [88]
5. Engagement of clinicians and service users, to ensure that any application of polygenic risk scores avoids deterministic interpretations and is based on the understanding that PRS is an indicator, not a precise measure