

# GWAS significance thresholds in large cohorts of European ancestry

Evans K. Cheruiyot <sup>1,\*</sup> Tingyan Yang<sup>2,3</sup> Allan F. McRae<sup>1</sup>

<sup>1</sup>Institute of Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072, Australia

<sup>2</sup>Translational Neurogenomics Group, QIMR Berghofer Medical Research Institute, Herston, QLD 4006, Australia

<sup>3</sup>Faculty of Medicine, The University of Queensland, Herston, QLD 4006, Australia

\*Corresponding author: Email: e.cheruiyot@uq.edu.au

While the  $P$ -value threshold of  $5.0 \times 10^{-8}$  remains the standard for genome-wide association studies (GWAS) in humans and other species, it still needs to be updated to reflect the current era of large-scale GWAS, where tens of thousands of sample sizes are used to discover genetic associations at loci with smaller minor allele frequencies. In this study, we used a dataset of 348,501 individuals of European ancestry from the UK Biobank to determine the GWAS thresholds required for multiple testing corrections when considering rare and common variants in additive and dominant GWAS models. Additionally, we employed conditional and joint analysis to quantify the proportion of false significant hits in the GWAS results for 72 traits in the UK Biobank when applying the traditional GWAS cutoff vs our newly proposed  $P$ -value thresholds. Overall, the results indicate that the conventional GWAS significance threshold of  $5.0 \times 10^{-8}$  yields a false-positive rate of between 20% and 30% in GWAS studies that utilize large sample sizes and less common variants. Instead, a more stringent GWAS  $P$ -value threshold of  $5.0 \times 10^{-9}$  is needed when rare variants (with minor allele frequency  $> 0.1\%$ ) are included in the association test for both additive and dominance models within the European ancestry population. However, further validation across diverse datasets and study designs, is needed to evaluate the broader applicability of this proposed threshold.

**Keywords:** GWAS; additive; dominance; threshold; UK Biobank;  $P$ -value; false discovery

## Introduction

The widely accepted  $P$ -value threshold for genome-wide association studies (GWAS) of common variants in humans is  $5.0 \times 10^{-8}$  (Panagiotou et al. 2012; Fadista et al. 2016; Marees et al. 2018). This threshold was developed based on smaller sample sizes and marker densities available during the early days of GWAS (International HapMap Consortium 2005; Dudbridge and Gusnanto 2008) and has been highly successful in discovering many reproducible genetic variants associated with common traits and diseases. However, it is important to note that the choice of GWAS thresholds can be context-dependent, varying based on study-specific factors such as sample size, the number of variants tested, and population structure (Fadista et al. 2016; Kanai et al. 2016; Asif et al. 2021).

Two remarkable advances in the GWAS era necessitate revisiting and updating this  $P$ -value threshold. First, the sample sizes used in GWAS have increased steadily over the years, reaching over 5 million individuals in recent studies (Yengo et al. 2022). Second, with the advances in large-scale genotyping and next-generation sequencing technologies, GWA studies have focused not only on common variants ( $MAF > 5\%$ ) but also on low-frequency ( $1\% < MAF < 5\%$ ) and rare ( $MAF < 1\%$ ) variants, which have been shown to contribute substantially to complex trait heritability (Lee et al. 2014).

Several studies have investigated GWAS significance threshold in the context of deep phenotyping and wide allele frequency spectrum in GWAS (Xu et al. 2014; Fadista et al. 2016; Kanai et al. 2016; Asif et al. 2021; Chen et al. 2021). However, these studies have used relatively small sample sizes in their simulations to derive GWAS thresholds and are thus outdated when performing GWAS in large biobank cohorts. In addition, large sample sizes are providing power to investigate nonadditive modes of association. In particular, the traditional GWAS traditional cutoff ( $5.0 \times 10^{-8}$ ) has also been used for dominance models (Palmer et al. 2023; Zhu et al. 2023). However, different GWAS thresholds for additive and dominance models would be expected since the extent of linkage disequilibrium (LD) tagging across variants differs between the 2 models (Zhu et al. 2023).

Various statistical approaches have been used to determine GWAS significance thresholds: the Bonferroni correction (which assumes independent tests and thus overly conservative considering that the genetic variants are in LD) (Risch and Merikangas 1996), permutation and bootstrapping (resampling) (International HapMap Consortium 2005; Kanai et al. 2016; Lin 2019), and Bayesian methods (Wellcome Trust Case Control Consortium 2007). Other methods involving calculating the “effective number of tests” based on LD pruning or eigenvalue decomposition of phenotypes and genotypes (Cheverud 2001; Duggal et al. 2008; Sobota

Received on 21 January 2025; accepted on 17 February 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of The Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact [reprints@oup.com](mailto:reprints@oup.com) for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com).

et al. 2015) have also been proposed. However, while the permutation-based method is computationally demanding, it remains a “gold standard” in GWAS multiple testing correction because it preserves correlation structure for the data and is robust to the assumption of marker independence (Dudbridge and Gusnanto 2008; Hoggart et al. 2008; Asif et al. 2021).

## Materials and methods

We simulated 2 sets of 1,000 phenotypes to reflect: (1) the ideal statistical scenario where the phenotype is normally distributed (or has undergone rank-based inverse normal transformation) and (2) a trait with skewness. A normally distributed phenotype serves as a common proxy for the assumption of normally distributed residuals in the GWAS regression model. To simulate skewness, we used body mass index (BMI) measurements from the UK Biobank participants (regardless of inclusion in the genetic data). BMI data were standardized within sex to have a mean of zero and unit variance and values with scaled BMI > 6 (~5 inter-quartile range) were removed. Then, we randomly sampled the scaled BMI without replacement to generate 1,000 independent traits.

To estimate GWA significance thresholds across different minor allele frequency (MAF) scenarios, the genotype data were pruned based on 4 MAF cutoffs [ $>0.1\%$ ,  $>0.5\%$ ,  $>1\%$ , and  $>5\%$ ]. A total of 13,035,294, 9,603,317, 8,545,459, and 6,098,223 autosomal variants remained following MAF filtering at these thresholds, respectively.

For each of the simulated phenotypes [ $N = 1,000 \times 2$  models], we ran additive and dominance GWAS analysis. We used the fastGWA method implemented in GCTA software [option `-fastGWA-lr`] (Yang et al. 2011) to run additive linear models, while the dominance GWAS was performed using PLINK version 2 software [`-glm dominant`] (Purcell et al. 2007). No covariates were included in all the GWAS models.

To identify genome-wide significance thresholds for each of the MAF cutoffs, we examined the distribution of the minimum  $P$ -values  $-\log_{10}P_{\min}$  for each simulated trait, separately for additive and dominance models. The 95% percentile of the  $-\log_{10}P_{\min}$  distribution was defined as the genome-wide significance threshold for each MAF filtering scenario, reflecting a 5% genome-wide significance  $P$ -value. The false-positive rate when using the traditional GWAS significance threshold for each scenario was calculated as the proportion of traits for which  $-\log_{10}P_{\min}$  surpassed the genome-wide significance threshold of  $5.0 \times 10^{-8}$ .

## Results and discussion

We used a permutation procedure to estimate the GWAS significance thresholds using the UK Biobank (version 3) dataset (Bycroft et al. 2018). We restricted our analyses to a cohort of 348,501 unrelated individuals with European ancestry. Our results revealed that a more stringent GWAS cutoff than the commonly used threshold ( $5.0 \times 10^{-8}$ ) is required to control false-positive rate if variants with MAF of  $<5\%$  are included in the association test (Table 1). The traditional GWAS  $P$ -value cutoff of  $5.0 \times 10^{-8}$  is still valid when considering common ( $>5\%$  MAF) variants in the additive GWA model using large sample sizes. However, a stringent GWA  $P$ -value threshold of  $8.9 \times 10^{-9}$  is needed when rare variants (MAF  $> 0.1\%$ ) are included in an additive GWA model (Fig. 1 and Table 1). A more stringent GWA correction is required to guard against false positives when the phenotype is not

**Table 1.** Estimated GWAS significance thresholds for different MAF cutoff obtained from additive and dominance models.

Model	MAF cutoff <sup>a</sup>	P-value <sub>sig</sub> threshold	M <sub>eff</sub>
Additive GWA (normal)	$>0.1\%$	$8.9 \times 10^{-9}$	5,636,978
	$>0.5\%$	$1.7 \times 10^{-8}$	2,873,563
	$>1\%$	$2.8 \times 10^{-8}$	1,785,714
	$>5\%$	$4.9 \times 10^{-8}$	1,014,198
Additive GWA (raw BMI)	$>0.1\%$	$5.9 \times 10^{-9}$	8,417,508
	$>0.5\%$	$1.5 \times 10^{-8}$	3,378,378
	$>1\%$	$1.7 \times 10^{-8}$	2,873,563
	$>5\%$	$3.3 \times 10^{-8}$	1,497,005
Dominance (normal)	$>0.1\%$	$7.4 \times 10^{-9}$	6,720,430
	$>0.5\%$	$1.5 \times 10^{-8}$	3,246,753
	$>1\%$	$1.9 \times 10^{-8}$	2,564,102
	$>5\%$	$3.5 \times 10^{-8}$	1,404,494
Dominance (raw BMI)	$>0.1\%$	$5.7 \times 10^{-8}$	8,756,567
	$>0.5\%$	$1.3 \times 10^{-8}$	3,937,007
	$>1\%$	$1.4 \times 10^{-8}$	3,597,122
	$>5\%$	$2.7 \times 10^{-8}$	1,818,181

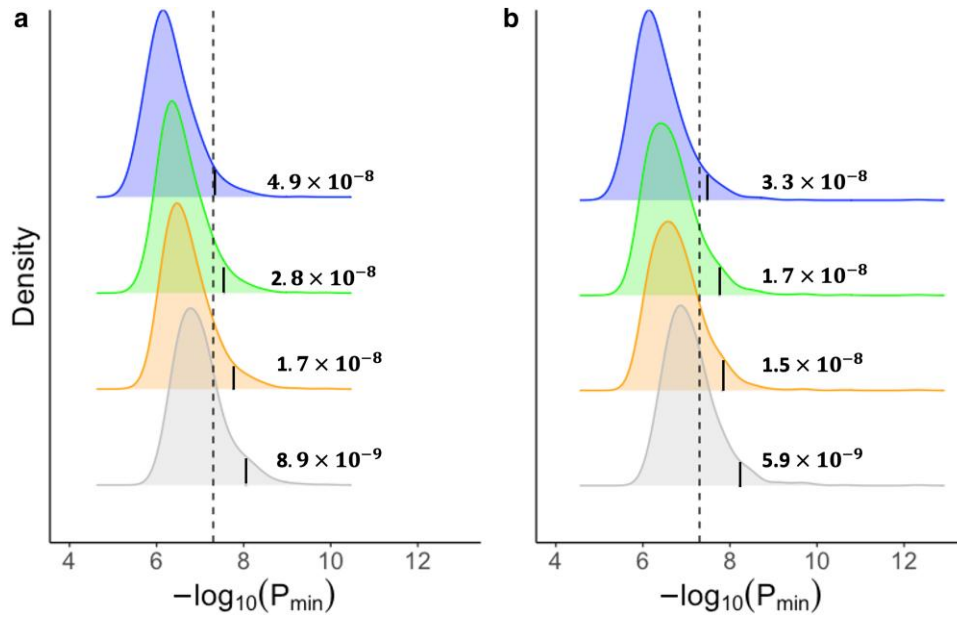
BMI, body mass index; GWA, genome-wide association.

<sup>a</sup>After pruning SNPs based on the 0.1%, 0.5, 1, and 5% MAF cutoff there were 13,035,294, 9,603,317, 8,545,459, and 6,098,223 SNPs remaining, respectively; Effective number of test ( $M_{\text{eff}}$ ) was calculated as  $0.05/P\text{-value}_{\text{sig}}$  threshold.

normally distributed. This was demonstrated using our BMI based trait, where the  $P$ -value cutoff was determined to be  $5.9 \times 10^{-9}$  for MAF  $> 0.1\%$  (Fig. 1). Using the traditional threshold of  $5.0 \times 10^{-8}$  leads to substantial Type-I error, particularly with low MAF thresholds (Table 2). When rare variants (MAF  $> 0.1\%$ ) are included in the association tests, we observed a false-positive rate of approximately 22% for a perfectly normal trait in the additive model, increasing to 30% with some phenotype skewness.

These findings are consistent with previous studies showing that a more stringent  $P$ -value cutoff is required when the MAF pruning is relaxed to accommodate rare variants in GWAS (Fadista et al. 2016). Our results are comparable with simulations using data (~140 k) from European ancestry that obtained a GWA significance  $P$ -value of  $6.6 \times 10^{-9}$  for MAF  $\geq 0.1\%$  (Kemp et al. 2017). This indicates that for MAF  $> 0.1\%$ , the effect of increasing sample size is attenuated by the time sample sizes reach hundreds of thousands (Asif et al. 2021). As such, our point estimates are likely applicable even for large GWAS (Yengo et al. 2022). Taken together, we recommend a more stringent GWAS  $P$ -value cutoff of  $5.0 \times 10^{-9}$  if rare variants (MAF  $> 0.1\%$ ) are considered in the association test for an additive model—equivalent to 10 million independent tests at 5% alpha (Table 1). This conservative threshold is motivated by our use of skewed BMI data in the simulations, as other traits may exhibit even greater skewness, resulting in higher false-positive rates (Supplementary Tables S1 and S2). While this threshold will provide stringent control of type-1 error for GWAS in European individuals for many traits, it is important to note that GWAS thresholds are context-dependent, and studies may benefit from the establishment of study-specific thresholds if they deviate from the conditions used in our simulations.

We also investigated the  $P$ -value threshold when applying a dominance model in GWAS (Fig. 2). It is common for gene mapping studies to assume a typical GWA  $P$ -value ( $5.0 \times 10^{-8}$ ) for both additive and dominance GWAS models, particularly for common variants (Palmer et al. 2023). However, a more stringent threshold is needed to minimize false-positive results under the dominance model—even for common variants—because dominance LD tagging captures less variation than additive LD tagging



**Fig. 1.** The distribution of the minimum P-value ( $-\log_{10}P_{\min}$ ) from the additive GWAS models. We extracted  $-\log_{10}P_{\min}$  from a GWA results for each of the (a) 1,000 simulated normally distributed phenotypes and (b) permuted raw BMI phenotypes from UK Biobank. The genotypes used in GWA were from the European cohort in the UK Biobank ( $N = 348,501$ ). The dotted line corresponds to the traditional genome-wide significance cutoff of ( $5.0 \times 10^{-8}$ ), while the vertical solid lines represent the new empirical GWAS cutoff after pruning genotypes based on the  $>5\%$  (blue),  $>1\%$  (green),  $>0.5\%$  (orange), and  $>0.1\%$  (gray) minor allele thresholds. The new cutoff GWAS thresholds represents 95% percentile of  $P_{\min}$  distribution ( $\alpha = 0.05$ ).

**Table 2.** Type I error rates associated with including common (MAF  $> 5\%$ ) and rare variants in GWAS based on the traditional threshold of  $5.0 \times 10^{-8}$ .

Model	Trait	MAF cutoff			
		$>0.1\%$	$>0.5\%$	$>1\%$	$>5\%$
Additive GWA	Simulated normal	22.2	13.1	9.5	5.1
	Raw BMI	29.5	16.6	13.3	7.3

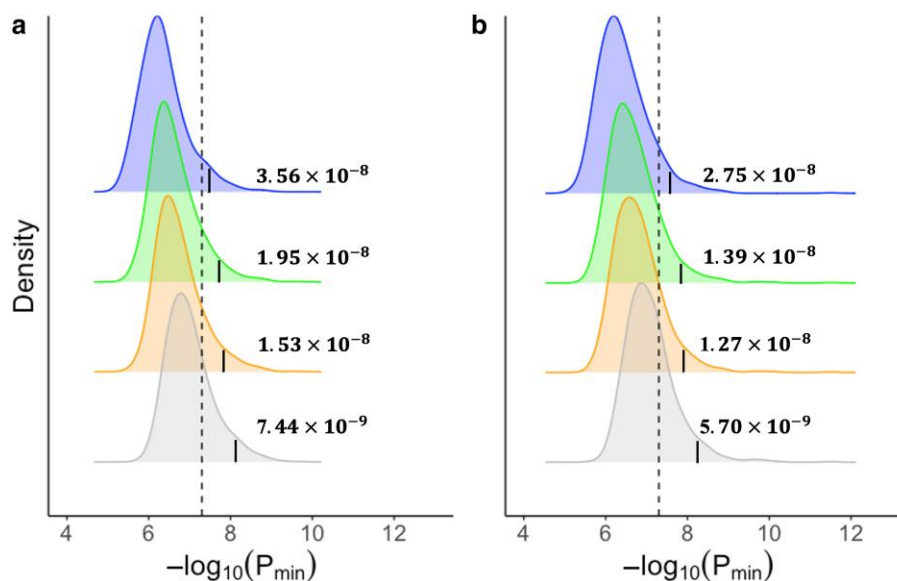
GWA, genome-wide association; MAF, minor allele frequency; simulated normal, normal distributed trait; Raw BMI, body mass index exhibiting skewness.

in the genome (Zhu et al. 2015). Our results show that a lower P-value threshold of  $3.5 \times 10^{-8}$  is needed when mapping common variants (MAF  $> 5\%$ ) in a dominance GWAS (Fig. 2). Even a smaller P-value cutoff is ideal if the phenotype is skewed, as shown from our simulations based on the UK Biobank BMI trait where we found a P-value threshold of  $2.7 \times 10^{-8}$  (Fig. 2 and Table 1). When rare variants (MAF  $> 0.1\%$ ) are included the GWAS, we found a P-value cutoff of  $7.4 \times 10^{-9}$  (for normally distributed trait and a  $5.7 \times 10^{-9}$  for a slightly skewed phenotypes (Fig. 2 and Table 1).

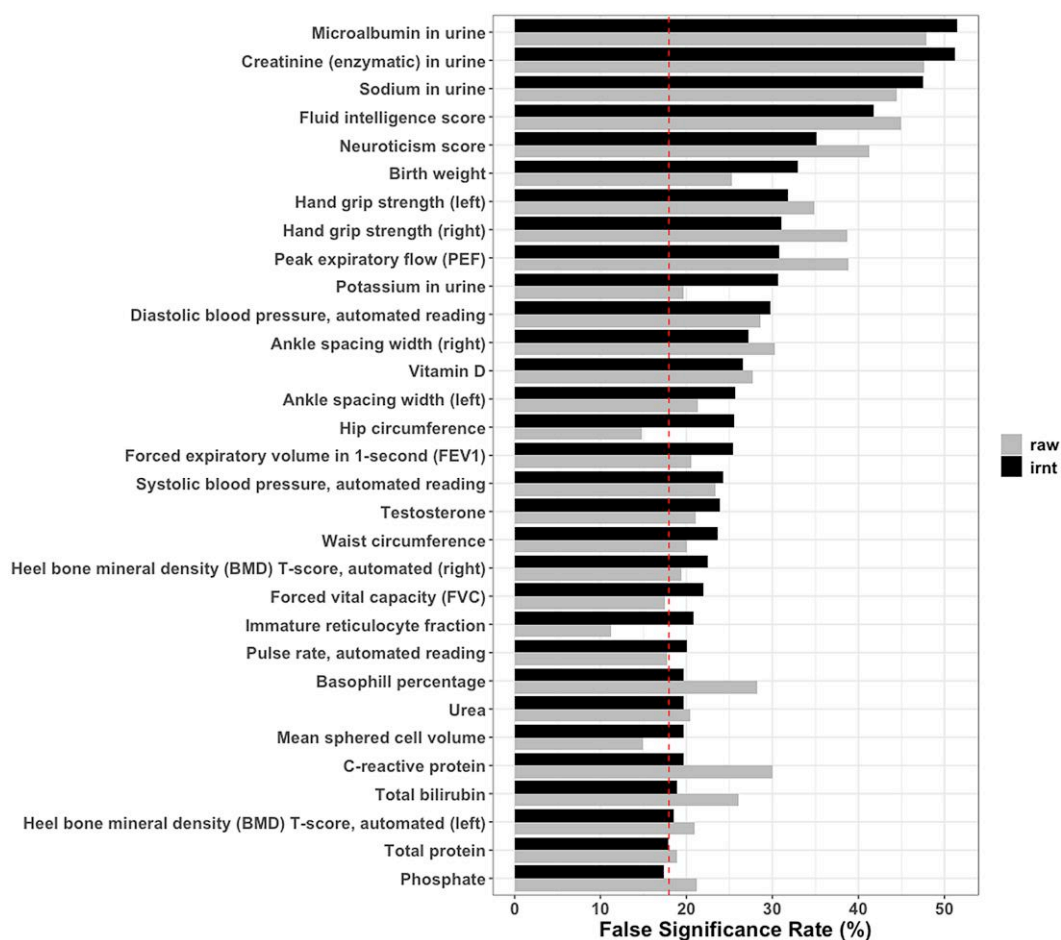
To quantify the number of independent significant GWAS signals, we employed PLINK “clumping” approach (Purcell et al. 2007) and conditional and joint (COJO) analysis (Yang et al. 2012) implemented in the GCTA software (Yang et al. 2011). COJO was performed using the following command: `gcta64 -bfile [LD reference] -maf 0.001 -cojo-file [trait name] -cojo-slc -cojo-p [GWAS cutoff] -thread-num 5 -out [file name]`, while PLINK was run with these parameters: `-clump [trait name], -clump-p1 [GWAS cutoff], -clump-r2 0.01, -clump-kb 250 -out [file name]`. We quantified the “false significance rate (FSR)”—defined as the proportion of tests in the GWA results from UK Biobank that passed traditional significance threshold ( $5.0 \times 10^{-8}$ ) but failed to reach the new empirical GWAS threshold from our simulations. Here, we

quantified the FSR from the most stringent empirical GWAS thresholds obtained from filtering genotypes based on the MAF cutoff of  $>0.1\%$  for only the additive GWAS simulations. COJO performs conditional analysis of GWA summary statistics and does not require individual-level genotypes. To run COJO, we downloaded GWA summary statistics for 72 quantitative traits in the UK Biobank (GWAS round 2). The summary statistics for the raw and inverse-rank normalized (irnt) phenotypes were available for these traits. We used a random subset of 10,000 unrelated Europeans from the UK Biobank as LD reference in the COJO analysis. We ran COJO per chromosome for the GWA summary data for both sexes using the default setting. We calculated the FSR for each trait as the percentage difference in the number of independent GWAS counts between the traditional GWA significance threshold ( $5.0 \times 10^{-8}$ ) and the new empirical GWA threshold.

By using the most stringent empirical GWA correction from the additive model based on the MAF  $> 0.1\%$  cutoff ( $8.9 \times 10^{-9}$ ) COJO analysis (see methods), we found the proportion of false significant results for the inverse normalized traits ranged from 5.56% (Glycated hemoglobin) to 51.43% (microalbumin in urine), with an mean of 18.26% across traits (Fig. 3 and Supplementary Table S1). Similarly, applying the updated P-value correction from the simulations for the raw phenotypes ( $5.9 \times 10^{-9}$ ) in COJO, yielded a slightly higher mean proportion of false significant results of 20% across non-normalized traits. Besides microalbumin in urine, the other normalized traits with the highest FSR included creatinine and sodium in urine and fluid intelligence score (Fig. 3). We found similar levels of FSR when using the software plink “clumping” approach, with the mean false significant results across traits of 20.95 and 23.29% for normalized (irnt) and non-normalized (raw) traits, respectively (Supplementary Table S2). Given the move to large meta-analysis leaving few suitable cohorts with power for replication, these results highlight the need for properly controlled type-I error to avoid follow-up studies on potential false-positive associations.



**Fig. 2.** The distribution of minimum P-value ( $-\log_{10}P_{\min}$ ) from the dominance GWAS models. The  $-\log_{10}P_{\min}$  was tracted from a GWA results based on the (a) 1,000 simulated normally distributed phenotypes and (b) permuted raw BMI phenotypes from UK Biobank. The genotypes used were from the European cohort in the UK Biobank ( $N = 348,501$ ). The dotted line corresponds to the traditional genome-wide significance cutoff of ( $5.0 \times 10^{-8}$ ), while the vertical solid lines represent the new empirical GWAS cutoff after pruning genotypes based on the >5% (blue), >1% (green), >0.5% (orange), and >0.1% (gray) minor allele thresholds. The new cutoff GWAS thresholds represents 95% percentile of  $P_{\min}$  distribution ( $\alpha = 0.05$ ).



**Fig. 3.** Top 30 traits with the highest proportion of associations (FSR) that passed traditional threshold ( $5.0 \times 10^{-8}$ ) but failed to reach the updated threshold for the additive GWAS. The FSR for the irnt traits and their corresponding raw untransformed phenotypes are represented by black and gray colors, respectively. The FSR was calculated as the percentage of minimum P-values ( $-\log_{10}P_{\min}$ ) passing the typical GWAS cutoff ( $5.0 \times 10^{-8}$ ) from the GWAS of the 1,000 simulated traits. The red vertical dashed line represents mean FSR for the irnt traits.



The debate over the most appropriate GWAS thresholds for different scenarios has persisted for decades (Dudbridge and Gusnanto 2008; Panagiotou et al. 2012; Fadista et al. 2016; Chen et al. 2021). While we have proposed increasing the stringency of  $P$ -value thresholds to reduce false positives, this approach could inadvertently increase the false-negative rate. Empirical data suggest that relaxing the traditional GWAS threshold ( $5.0 \times 10^{-8}$ ) could yield a substantial fraction of true “borderline” discoveries, albeit at the cost of a higher false discovery rate may reach up to 50% (Panagiotou et al. 2012). This observation aligns with our findings on the FSR. Although higher thresholds may mitigate false negatives, the relative cost and wasted effort associated with follow-up research on false discoveries could be substantial, which may justify the call for stringent GWAS thresholds (Ioannidis et al. 2011; Panagiotou et al. 2012). Future studies that integrate replication efforts, functional validation, and context-specific thresholds will be essential for balancing discovery potential with robust and replicable findings (Panagiotou et al. 2012; Abdellaoui et al. 2023).

Notably, our empirical GWA threshold applies to only the European ancestry population, which we used in the simulations because a large sample size was available for GWAS. Future studies on other populations, e.g. Africans, are required given that LD structure is population specific and, therefore, different GWAS correction thresholds are needed (Kanai et al. 2016; Pulit et al. 2017). However, we recognize that the choice of GWAS thresholds may be context-dependent, and further validation across diverse datasets, populations, and study designs is warranted to evaluate the broader applicability of this proposed threshold.

In summary, using simulations with large sample sizes, we have established that the traditional GWAS significance  $P$ -value of  $5.0 \times 10^{-8}$  is insufficient for multiple testing corrections in current GWAS that use large sample sizes and less common variants. Instead, a stringent GWAS correction  $P$ -value of  $5.0 \times 10^{-9}$  is needed when rare variants ( $MAF > 0.1\%$ ) are considered in the association test for both additive and dominance models in the European ancestry population. Adopting the new  $P$ -value threshold for rare variants is critical for guarding against type 1 errors in future GWAS for European ancestry.

## Data availability

This work used genotype and phenotype data from UK Biobank Resource under project 12505. UKB data can be accessed upon request once a research project has been submitted and approved by the UKB committee. The data analyses were conducted using publicly available tools (GCTA and PLINK), with results visualized using the R programming language. Detailed descriptions of the analytical procedures are provided in the Materials and Methods section to facilitate reproducibility of the results. The scripts for running GWAS and generating visualizations are available at: [https://github.com/mcrae/gwas\\_threshold](https://github.com/mcrae/gwas_threshold). UK Biobank GWAS data, <https://www.nealelab.is/uk-biobank> (Neale Lab).

Supplemental material available at GENETICS online.

## Funding

Allan McRae is the recipient of an Australian Research Council Australian Fellowship (project number FT200100837) funded by the Australian Government. The views expressed herein are those of the authors and are not necessarily those of the Australian Government or Australian Research Council. This research has been conducted using the UK Biobank Resource under project 12505.

## Conflicts of interest

The authors declare no competing interests.

## Author contributions

AFM conceived and designed the study. EKC, TY, and AFM conducted data analysis. EKC wrote the manuscript. All authors reviewed, revised and approved the final manuscript for publication.

## Literature cited

- Abdellaoui A, Yengo L, Verweij KJ, Visscher PM. 2023. 15 years of GWAS discovery: realizing the promise. *Am J Hum Genet.* 110(2):179–194. doi:10.1016/j.ajhg.2022.12.011.
- Asif H, Alliey-Rodriguez N, Keedy S, Tamminga CA, Sweeney JA, Pearlson G, Clementz BA, Keshavan MS, Buckley P, Liu C, et al. 2021. GWAS significance thresholds for deep phenotyping studies can depend upon minor allele frequencies and sample size. *Mol Psychiatry.* 26(6):2048–2055. doi:10.1038/s41380-020-0670-3.
- Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O’Connell J, et al. 2018. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 562(7726):203–209. doi:10.1038/s41586-018-0579-z.
- Chen Z, Boehnke M, Wen X, Mukherjee B. 2021. Revisiting the genome-wide significance threshold for common variant GWAS. *G3 (Bethesda).* 11(2):jkaa056. doi:10.1093/g3journal/jkaa056.
- Cheverud JM. 2001. A simple correction for multiple comparisons in interval mapping genome scans. *Heredity (Edinb).* 87(Pt 1):52–58. doi:10.1046/j.1365-2540.2001.00901.x.
- Dudbridge F, Gusnanto A. 2008. Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol.* 32(3):227–234. doi:10.1002/gepi.20297.
- Duggal P, Gillanders EM, Holmes TN, Bailey-Wilson JE. 2008. Establishing an adjusted  $p$ -value threshold to control the family-wide type 1 error in genome wide association studies. *BMC Genomics.* 9(1):516. doi:10.1186/1471-2164-9-516.
- Fadista J, Manning AK, Florez JC, Groop L. 2016. The (in)famous GWAS  $P$ -value threshold revisited and updated for low-frequency variants. *Eur J Hum Genet.* 24(8):1202–1205. doi:10.1038/ejhg.2015.269.
- Hoggart CJ, Clark TG, De Iorio M, Whittaker JC, Balding DJ. 2008. Genome-wide significance for dense SNP and resequencing data. *Genet Epidemiol.* 32(2):179–185. doi:10.1002/gepi.20292.
- International HapMap Consortium. 2005. A haplotype map of the human genome. *Nature.* 437(7063):1299–1320. doi:10.1038/nature04226.
- Ioannidis JP, Tarone R, McLaughlin JK. 2011. The false-positive to false-negative ratio in epidemiologic studies. *Epidemiology.* 22(4):450–456. doi:10.1097/EDE.0b013e31821b506e.
- Kanai M, Tanaka T, Okada Y. 2016. Empirical estimation of genome-wide significance thresholds based on the 1000 Genomes Project data set. *J Hum Genet.* 61(10):861–866. doi:10.1038/jhg.2016.72.
- Kemp JP, Morris JA, Medina-Gomez C, Forgetta V, Warrington NM, Youten SE, Zheng J, Gregson CL, Grundberg E, Trajanoska K, et al. 2017. Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. *Nat Genet.* 49(10):1468–1475. doi:10.1038/ng.3949.
- Lee S, Abecasis GR, Boehnke M, Lin X. 2014. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet.* 95(1):5–23. doi:10.1016/j.ajhg.2014.06.009.

- Lin DY. 2019. A simple and accurate method to determine genome-wide significance for association tests in sequencing studies. *Genet Epidemiol.* 43(4):365–372. doi:[10.1002/gepi.22183](https://doi.org/10.1002/gepi.22183).
- Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, Derks EM. 2018. A tutorial on conducting genome-wide association studies: quality control and statistical analysis. *Int J Methods Psychiatr Res.* 27(2):e1608. doi:[10.1002/mpr.1608](https://doi.org/10.1002/mpr.1608).
- Neale Lab. UK Biobank GWAS Results. <http://www.nealelab.is/uk-biobank>.
- Palmer DS, Zhou W, Abbott L, Wigdor EM, Baya N, Churchhouse C, Seed C, Poterba T, King D, Kanai M, et al. 2023. Analysis of genetic dominance in the UK Biobank. *Science.* 379(6639):1341–1348. doi:[10.1126/science.abn8455](https://doi.org/10.1126/science.abn8455).
- Panagiotou OA, Ioannidis JP; Genome-Wide Significance Project. 2012. What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. *Int J Epidemiol.* 41(1):273–286. doi:[10.1093/ije/dyr178](https://doi.org/10.1093/ije/dyr178).
- Pulit SL, de With SA, de Bakker PI. 2017. Resetting the bar: statistical significance in whole-genome sequencing-based association studies of global populations. *Genet Epidemiol.* 41(2):145–151. doi:[10.1002/gepi.22032](https://doi.org/10.1002/gepi.22032).
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 81(3):559–575. doi:[10.1086/519795](https://doi.org/10.1086/519795).
- Risch N, Merikangas K. 1996. The future of genetic studies of complex human diseases. *Science.* 273(5281):1516–1517. doi:[10.1126/science.273.5281.1516](https://doi.org/10.1126/science.273.5281.1516).
- Sobota RS, Shriner D, Kodaman N, Goodloe R, Zheng W, Gao Y-T, Edwards TL, Amos CI, Williams SM, et al. 2015. Addressing population-specific multiple testing burdens in genetic association studies. *Ann Hum Genet.* 79(2):136–147. doi:[10.1111/ahg.12095](https://doi.org/10.1111/ahg.12095).
- Wellcome Trust Case Control Consortium. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* 447(7145):661–678. doi:[10.1038/nature05911](https://doi.org/10.1038/nature05911).
- Xu C, Tachmazidou I, Walter K, Ciampi A, Zeggini E, Greenwood CMT; UK10K Consortium. 2014. Estimating genome-wide significance for whole-genome sequencing studies. *Genet Epidemiol.* 38(4):281–290. doi:[10.1002/gepi.21797](https://doi.org/10.1002/gepi.21797).
- Yang J, Ferreira T, Morris AP, Medland SE; Genetic Investigation of Anthropometric Traits (GIANT) Consortium; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium; Madden PA, Heath AC, Martin NG, Montgomery GW, et al. 2012. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet.* 44(4):369–375, S1–3. doi:[10.1038/ng.2213](https://doi.org/10.1038/ng.2213).
- Yang J, Lee SH, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 88(1):76–82. doi:[10.1016/j.ajhg.2010.11.011](https://doi.org/10.1016/j.ajhg.2010.11.011).
- Yengo L, Vedantam S, Marouli E, Sidorenko J, Bartell E, Sakaue S, Graff M, Eliassen AU, Jiang Y, Raghavan S, et al. 2022. A saturated map of common genetic variants associated with human height. *Nature.* 610(7933):704–712. doi:[10.1038/s41586-022-05275-y](https://doi.org/10.1038/s41586-022-05275-y).
- Zhu C, Ming MJ, Cole JM, Edge MD, Kirkpatrick M, Harpak A. 2023. Amplification is the primary mode of gene-by-sex interaction in complex human traits. *Cell Genom.* 3(5):100297. doi:[10.1016/j.xgen.2023.100297](https://doi.org/10.1016/j.xgen.2023.100297).
- Zhu Z, Bakshi A, Vinkhuyzen AA, Hemani G, Lee SH, Nolte IM, van Vliet-Ostaptchouk JV, Snieder H, Esko T, Milani L. 2015. Dominance genetic variation contributes little to the missing heritability for human complex traits. *Am J Hum Genet.* 96, 377–385.

Editor: M. Sillanpää