

Estimation and Mapping of the Missing Heritability of Human Phenotypes

Corresponding Author: Professor Loïc Yengo

Any redactions in this file are there to maintain patient confidentiality, the confidentiality of unpublished data, or to remove third-party material.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Referee #1

(Remarks to the Author)

This paper contains important new results, the analysis seems of a high standard and the presentation is generally clear. It deserves to be published in Nature but I propose a number of revisions.

Major comments:

Perhaps the most important is that fastGWA generally performs poorly for LMAA (see e.g. fig 3 of <https://doi.org/10.1101/2024.07.25.24311005>). Regenie is better and also computationally fast. BOLT-LMM is better again, but presents bigger computational challenges. It seems a shame to publish analysis of such an important dataset that uses suboptimal inference software. At least you could redo the association analysis using Regenie and report differences in results.

Another concern is that GCTA-LDMS has not been shown to perform well for rare variants. The citation given is to the original 2015 paper that did not address rare variants. Also, how is GCTA-LDMS affected by AM?

Specific issues:

There has been so much nonsense published about "missing heritability" that many of us have an allergic reaction to the term. Although the presentation here is generally good, I would prefer to stick more to the science and reduce the journalistic hyperbole such as "recovers the missing heritability" and "myriad". Nothing has been "recovered" (which means you get back something you previously had - not true here). Similarly RHR is poor terminology, "explained HR" would be a better alternative.

I find the 2nd paragraph (L58 to 79) a waste of space, the distinction between h2gwas and h2snf is arbitrary (depends on the significance threshold; they are based on different analyses of the same data and so its just differences in analysis). The topic is not worthy of such a prominent position - get on to the interesting results, and if you like put this in an appendix. The later comparison of h2WGS with h2ped is interesting, but overall I think the paper could be improved by trimming 5-10% of the wording on some of the less interesting analyses.

On L79 the spurious use of the term "equity" made me wince. There's no justification for its use in a study almost entirely of Europeans.

The authors are generally admirably careful about possible biases, but I didn't see an explicit mention in the text that h2PED values can be inflated by shared environment. I guess you are relying on this being minimal for older study participants but is there evidence for this? I see there is an adjustment in the analysis - it seems plausible but hard to assess whether it is adequate.

L116 "pairwise genetic relation matrix (GRM)" is unnecessarily obscure, say what you are computing (i.e. an allelic

correlation) and refer to the formula given in Methods.

L254 "... RVA to explain on average 0.021% of phenotypic variance ... 0.023% observed for CVA". Better to clarify that this per variant as I was initially confused, but also why do we need to talk about "phenotypic variance" rather than being consistent with discussing heritability? (Similarly elsewhere, e.g. L293 "per-SNP variance explained" why not talk about per SNP h^2 ? Again at L323 "... phenotypic variance (and thus heritability)" why do we need to mention both?

L266 "CVAs and RVAs accounted for ... of the average common-variant and rare-variant heritability" I think you need to explain how the latter quantities are computed.

L331 "replaced sequenced variants with imputed SNPs with MAF>0.01%" this needs a little explanation, if all sequence variants were removed how was the imputation done?

L414 I don't understand "the entire mutational target".

L544 and 558 unnecessary repeat of "452,618 samples of European ancestry" but also not an exact repeat which may confuse.

L560 (also L570) "estimated relationship" you don't "estimate" relatedness you compute a coefficient that measures relatedness. For an estimate there has to be a meaningful true value, there isn't for relatedness (e.g. pedigree-based values are arbitrary, depending on the choice of pedigree; GWAS-based coefficients are also arbitrary, there is no "true value").

L568 "median LD distribution" doesn't make sense to me. Firstly, LD is not a numerical quantity so doesn't have a median, but also do you mean a distribution of median values? If so distribution over what?

L668 typo "not affect to the"

L687 I guess you mean "(see under Code availability)"

L714 I don't understand "... 100 k-means for birth coordinates"

(Remarks on code availability)

Referee #2

(Remarks to the Author)

Wainschtein and colleagues present results of analyses in the UK Biobank to evaluate the impact of rare genetic variation of complex human traits and diseases. They demonstrate that whole-genome sequence (WGS) variants (MAF >0.01%) capture ~91% of pedigree-based narrow sense heritability, with ~21% attributable to rare variants (MAF <1%). They assess the relative contribution of coding and non-coding variation to heritability amongst rare and common genetic variation and conduct GWAS analyses across phenotypes using WGS and imputed variants that highlight rare variant associations that account for a substantial proportion of the heritability of lipid traits. Overall, the manuscript is well written, providing insights to the key questions I have about the contribution of rare variants to complex phenotypes, and the methods used generally robust and adequately motivated (see some comments below). However, my enthusiasm for the manuscript is somewhat dampened by the focus of analyses only on individuals of European ancestry, which the authors highlight as a limitation of the work.

Comments:

1. The methods state that quantitative traits were standardised, but were they inverse rank normalised? Extreme outliers were removed, but with such a large sample size, subtle deviations from a Normal distribution will impact on findings and increase false positive associations with rare variants in GWAS analyses.

2. Does fastGWA-GLMM allow for case-control imbalance. If not, what impact would you expect this to have on your analyses?

3. The definition of "independence" using LD clumping is not ideal. With large sample sizes, a large effect size and highly significant association may still have a "shadow" at $r^2 < 0.01$. Given that individual-level data are available, were these independent associations confirmed through formal conditional analyses? The use of 1Mb window may also not be sufficient for the MHC.

4. Presumably individuals from BioVU for these analyses were of European ancestry – this should be emphasized in the main text methods. Were sample sizes sufficiently large in BioVU to allow investigation of rare variant effects in other ancestry groups? Could the authors speculate on whether they expect their findings to be consistent in other ancestry groups?

5. In the imputed data analysis, were imputed genotypes first converted to hard genotype calls (I notice that a "missing genotype rate" is used for QC). If so, this is not optimal, and the dosage should be used instead.

6. The comparison of WGS and imputed variants is very interesting. Like the authors, I was not expecting that common variant associations would be missed and such a large proportion of rare variant associations captured with imputed data. Does WGS offer any advantages in terms of fine-mapping? This is difficult to quantify without knowing causal variants, but I wonder if credible set sizes around common and rare variants are smaller with WGS than imputed data, which would reflect improved fine-mapping resolution?

7. The authors highlight that there is limited power, even in UK Biobank, to detect associations with rare variants with MAF <0.1%. Strategies to analyse rare coding variants through aggregate gene-based tests are well developed, but could the authors suggest approaches that could be applied for non-coding variants?

Minor points:

1. The definition of rare and common variants was not entirely clear – if MAF <1% and MAF >1%, respectively, is used, what about MAF =1%?

2. Paragraph beginning line 94. Emphasize that all of these analyses are undertaken in European ancestry individuals.

(Remarks on code availability)

Referee #3

(Remarks to the Author)

Wainschtein and colleagues undertake a series of analyses using whole-genome sequencing data from ~350,000 individuals in the UK Biobank. Looking at 36 complex traits (predominantly lipid traits) the authors estimate the heritability of these traits, the proportion of this heritability captured by the WGS data, and the relative contribution of rare and common variation, plus coding and non-coding variation, to the heritability explained by the WGS data. The authors also undertake GWAS analyses for these 36 phenotypes.

Excluded variants: In the Discussion, you rightly list various potential sources of missing heritability for the traits where $h^2_{WGS} < h^2_{PED}$. These include ultra-rare variants, structural variants, genome build gaps and non-additive effects. Given these exclusions, it surprises me that you have identified 15 traits where the WGS data explains 100% of h^2_{PED} . Do you believe this means that ultra-rare or structural variants, genome-build gaps, non-additive effects or sex-chromosome variants play no role in these traits/diseases? I find this hard to believe, and it makes me wonder if you have biased estimates of either h^2_{WGS} or h^2_{PED} . If you don't believe this is the case, then how should one interpret your results and their impact on genetic studies of these phenotypes moving forward?

Robustness of findings: Following on from the above, and based on your description of the limitations of the study (a section which I appreciated), I am concerned that the findings of the work are not robust. The healthy volunteer bias, the exclusion of many variant classes, the low power to accurately quantify rare variant effects, the variation in power across traits and diseases confounding comparisons make me question the robustness of the findings.

Utility and interpretation of findings: In the introduction, you state that quantifying the contribution of different classes of genetic variants to variation in complex traits is crucial to designing optimal experiments, identifying causal variants, and delivering precision medicine and drug discovery. Perhaps this was the case previously when one had to decide between a common variant microarray, whole-exome sequencing or whole-genome sequencing. However, WGS will now be performed on most cohorts, so which subset of variants to assay is no longer a relevant study design question. I would have liked to see more discussion about how your results impact the study design of future genetic studies of the 36 traits and affect the delivery of precision medicine and drug target identification for these. How does this differ depending on the proportion of heritability explained, and where in the genome this heritability lies?

Portability of findings: Building on my concerns around the utility of these findings for the 36 traits and diseases included directly in this work, it wasn't clear to me how the results from this study are relevant to other complex diseases and traits. What general insights, if any, have been made about common complex traits and diseases? Also, what effect does the restriction of your analyses to the 36 traits with a marginally significant rare-variant heritability estimate have on the portability of your findings across other traits? The complex diseases were also dropped from some of your analyses because of the larger standard error around the h^2_{PED} estimate - does this impact negatively the portability of your findings?

The novelty of GWAS findings: Analyses of the UK Biobank WGS data are emerging, not least by the UK Biobank WGS Consortium. The novelty of the GWAS analyses presented in this work was not clear to me, and it wasn't clearly outlined in the paper (or Supplementary Table 11, as far as I could tell - it would be good if a legend could be supplied explaining the column headings). I suggest the authors compare their analyses to those undertaken in <https://www.medrxiv.org/content/10.1101/2023.12.06.23299426v1.full> and presented publically at <https://azphewas.com/>. I tried to do this for some variants myself, but due to the limited information supplied in the main text of the manuscript, I was not able to confidently find the indel downstream of HBQ1 associated with mean corpuscular volume in Supp Table 11 (for example). See also a previous analysis of the UK biobank WGS data for telomere length: <https://www.nature.com/articles/s41588-024-01884-7>.

Clustering of independent GWAS signals: If I understand correctly, you clumped together lead SNPs within 1Mb that had an $r^2 > 0.01$ (if this is correct then the inequality is incorrect in the Methods section). Is 1Mb broad enough given the number of very strong associations that will be detected by very weak tags? In my experience, association peaks in biobank-scale analyses can have very broad shoulders that extend beyond 1Mb. Not appropriately accounting for association signals driven by weak tagging of known effects could also drive the observation about the colocalisation of common and rare variants. Again, in my experience, LD-based approaches (LDSC etc) struggle to account for (multiple) large effects in a region and this can result in spurious residual association signals (rare or common). Perhaps try fitting a full model of lead variants in a region to see how this compares to your standard approach?

(Remarks on code availability)

Version 1:

Reviewer comments:

Referee #1

(Remarks to the Author)

I'm generally happy with the revisions made in response to reviewer comments. However it's very disappointing that the revised version I received is marred by dozens of "Error! Reference source not found", there was even one of these in the response letter. It seems nobody looked at the submitted version, otherwise these errors should have been obvious.

Apologies I did not state my point well about LDMS not having been shown to "perform well". I won't pursue the point further, I don't have major concerns but there are limitations in Evans et al. and Wainschtein et al (2022) particularly due to unrealistic simulations that assign equal heritability to causal SNPs, and in any case these papers are not cited to support use of LDMS.

I'm glad you've adopted TetraHer and Regenie. The preprint I previously cited is now published in Nature Genetics <https://www.nature.com/articles/s41588-025-02286-z> showing that LDAK-KVIK is more powerful than Regenie and slightly faster. I won't pursue this point either, Regenie is at least better than fastGWA but it's a shame not to have used the best available software.

The response to my comment at previous L116 missed the point: "genomic relationship coefficient" is still ambiguous, there are lots of ways to measure relatedness from SNP data. I understand that an allelic correlation has become widely used and I guess is what you've used but I have not seen a good argument that this is an optimal choice of relatedness coefficient and in any case you need to be explicit what you have computed - you do give what I presume is the relevant formula at L645, you need to say that at first use.

L96 typo "account of" should be "for"

L391 "fine-mapping resolution is improved ..." I am confused whether the improvement is from use of imputation or use of WGS.

(Remarks on code availability)

Referee #2

(Remarks to the Author)

The authors have carried out extensive additional analyses to address the comments raised by the reviewers. I believe the use of REGENIE as an alternative and more robust analysis tool has been an important addition. I appreciated the attempt to look up results of rare variant associations in other ancestry groups, which highlighted some interesting results. No further comments.

(Remarks on code availability)

Referee #3

(Remarks to the Author)

Thank you for your thoughtful and diligent responses to my comments and those of my fellow reviewers.

I think the changes you have made to the manuscript have further strengthened it. In particular, the use of SUSIE for fine-mapping, the joint modelling of all fine-mapped SNPs to ensure independence, and the switch to REGENIE increase my confidence in the robustness of the findings.

I congratulate all the authors on a very nice paper that neatly addresses important questions in the field of complex disease genetics.

(Remarks on code availability)

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Summary of main changes to the manuscript

Heritability analyses and trait selection

We initially estimated pedigree-based heritability for binary traits on the 0-1 observed scale (so, analysing those traits the same way we analysed quantitative traits) then converted our estimates to the liability scale using a linear transformation. While this approach gives a valid approximation for analyses based on distant relatives, it tends to overestimate heritability from close relatives. Speed and Evans, AJHG (2024) recently proposed a new method, *TetraHer*, to estimate heritability of disease directly on the liability scale. We used this approach in our revised manuscript.

To be consistent with our previous criterion to only retain phenotypes with a marginally significant pedigree-based heritability, this re-estimation led us to exclude two diseases (i.e., varicose veins and gastro-oesophageal reflux disease) from our main analyses. Therefore, our final list of traits was reduced from 36 to 34.

Finally, we added a new Supplementary Note in which we estimate the contribution of heritability from common variants currently missed in the hg38 genome build. Results from that note are also briefly reported in our Discussion section.

GWAS analyses

- Software: **as suggested by referees**, we have now repeated all our GWAS analyses using REGENIE, instead of fastGWA.
- Criterion to determine independent associations: **as suggested by the referees**, we added to our clumping criterion a second step in which we fit all clumped associations jointly, then only retain those that remain genome-wide significant in the joint analysis.
- Number of variants: The REGENIE analysis overall identified more variants than fastGWA, consistent with reports from the literature. However, given that we applied a stricter definition of independent signals, we ended up with approximately the same number of associations as previously reported, although our number of rare variant associations was almost halved.
- Replication in AGD: We have updated those analyses using a larger sample size available. Also, we have extended the replication (beyond LDL) to HDL and alkaline phosphatase levels in two ancestry groups (European and African).
- Fine-mapping: **As suggested by the referees**, we performed fine-mapping analyses for all associations detected using WGS and imputed variants (so ~30,000 analyses) and show evidence of increased fine-mapping resolution for WGS over imputation, especially for rare variant associations.

Authorship

We have swapped the positions of two co-authors (namely, Irfahan Kassam and Petko P. Fiziev) to better reflect contributions made during the revision. This has been approved by all co-authors.

Referee #1 (Remarks to the Author):

This paper contains important new results; the analysis seems of a high standard and the presentation is generally clear. It deserves to be published in Nature but I propose a number of revisions.

We thank this referee for acknowledging the importance of our results and the high standard of our analyses and presentation. We are grateful for all the revisions suggested, each of which is addressed below.

Major comments:

Perhaps the most important is that fastGWA generally performs poorly for LMAA (see e.g. fig 3 of <https://doi.org/10.1101/2024.07.25.24311005>). Regenie is better and also computationally fast. BOLT-LMM is better again, but presents bigger computational challenges. It seems a shame to publish analysis of such an important dataset that uses suboptimal inference software. At least you could redo the association analysis using Regenie and report differences in results.

We thank the referee for this suggestion. We have redone all our GWAS analyses using Regenie. Overall, we detected more associations using Regenie as compared to fastGWA, consistent with the expected increased statistical power for the former approach. However, we also applied more stringent criteria to determine independence between associations such that the final numbers are somewhat similar to those reported before. We have updated our manuscript using Regenie results. Our main conclusions remain unchanged.

Another concern is that GCTA-LDMS has not been shown to perform well for rare variants. The citation given is to the original 2015 paper that did not address rare variants. Also, how is GCTA-LDMS affected by AM?

We thank the referee for this point although we are not sure what they mean by “perform well” in this context. The LDMS method was specifically developed for analysing WGS data (typically rare variants) and extensively assessed in previous publications from Evans et al. Nat. Genet. (2018) and Wainschein et al. Nat. Genet. (2022) beyond the initial 2015 paper by Yang and colleagues. A subgroup of authors of this manuscript (Border et al. Nat. Comm. 2022) has previously shown through theory and simulations that assortative mating (AM) upwardly bias GREML estimates of heritability and that this bias decreases with sample size. The same authors also showed that AM also biases heritability estimates from methods of moments like Haseman-Elston (HE) regression, but these biases can be corrected if the spousal correlation is known. Importantly, AM-induced biases are mainly caused by cross-chromosomal correlations between trait-increasing alleles. The LDMS method focuses on correcting biases associated with heterogeneity in within-chromosome LD between SNPs used to calculate relationship matrices and causal variants. Therefore, the LDMS method would be affected by AM in the same way as standard GREML (see Border et al. Nat. Comm. 2022). We compared our LDMS estimates with AM-corrected estimates of heritability from HE regression (Ext. Data. Fig. 2). We showed little difference between those two estimates, suggesting that AM-related biases in our LDMS results are likely small if not negligible. We attribute

this to the fact that we performed all our analyses using the largest possible sample size and that theory (and empirical analyses) predicts the magnitude of biases to decrease with sample size.

Specific issues:

There has been so much nonsense published about "missing heritability" that many of us have an allergic reaction to the term. Although the presentation here is generally good, I would prefer to stick more to the science and reduce the journalistic hyperbole such as "recovers the missing heritability" and "myriad". Nothing has been "recovered" (which means you get back something you previously had - not true here). Similarly RHR is poor terminology, "explained HR" would be a better alternative.

This point is well-taken. We have replaced the word "myriad" with "thousands". We understand that nothing is actually missing (as explained in our introduction) but kept the "missing heritability" phrase for historical reference. Also, we have replaced RHR with explained heritability ratio (EHR).

I find the 2nd paragraph (L58 to 79) a waste of space, the distinction between h^2_{gwas} and h^2_{snp} is arbitrary (depends on the significance threshold; they are based on different analyses of the same data and so its just differences in analysis). The topic is not worthy of such a prominent position - get on to the interesting results, and if you like put this in an appendix. The later comparison of h^2_{WGS} with h^2_{ped} is interesting, but overall I think the paper could be improved by trimming 5-10% of the wording on some of the less interesting analyses.

We understand that this section may seem a bit long but wanted to use that part of the introduction to recontextualize what the missing heritability debate has been about and (hopefully) clarify the terminology before presenting our results. We believe that has been a lot of confusion in the literature, partly because of a lack of understanding of certain nuances around h^2_{gwas} , h^2_{snp} and h^2_{ped} . We fully agree with the referee's point that some of these distinctions are dependent on statistical power as previously highlighted in our sentence: "The gap between h^2_{GWAS} and h^2_{SNP} , also referred to as "hiding heritability"⁹, has been predicted to vanish as GWAS sample sizes increase". We have now slightly rephrased that sentence as: "The gap between h^2_{GWAS} and h^2_{SNP} was previously referred to as "hiding heritability"⁹ and is expected to vanish as GWAS sample sizes increase."

On l79 the spurious use of the term "equity" made me wince. There's no justification for its use in a study almost entirely of Europeans.

We thank the referee of this comment. We used the word equity in this context not to refer to ancestry diversity alone but to sources of interactions in general. Our sentence followed an enumeration of factors potentially influencing heritability estimates, which included non-additive genetic effects such as gene-by-environment (GxE) interactions. Accounting for GxE interactions (especially when G and E are correlated) is important even when applications (e.g., genetic risk prediction) are done within a homogeneous ancestry group. However, we acknowledge that this point was not fully fleshed out in our submitted manuscript and apologize for that. Therefore, we have 1) added between brackets what those non-additive genetic effects might be ("(for example, interactions between genetic variants or between genetic variants and shared environment between relatives)"), and 2) removed the word equity from that sentence to minimize confusion.

The authors are generally admirably careful about possible biases, but I didn't see an explicit mention in the text that h2PED values can be inflated by shared environment. I guess you are relying on this being minimal for older study participants but is there evidence for this? I see there is an adjustment in the analysis - it seems plausible but hard to assess whether it is adequate.

We thank the referee for this point, which relates to our answer above. We do recognize that shared environmental effects can inflate h2PED values. We now explicitly mentioned this in our introduction: "Those factors include, genetic variation not well tagged by common SNPs (including rare variants or structural variants), **shared environmental effects** and non-additive genetic effects (for example, interactions between genetic variants or between genetic variants and shared environment between relatives), which may have inflated estimates of additive genetic variation from pedigree-based studies.". We did include in our model for estimating h2ped a term to account non-additive genetic effects, which would also capture shared-environmental effects that would be correlated with genetic similarity between relatives.

L116 "pairwise genetic relation matrix (GRM)" is unnecessarily obscure, say what you are computing (i.e. an allelic correlation) and refer to the formula given in Methods.

We thank the referee for this point. We have simplified that sentence by using "genomic relationship coefficient" instead.

L254 "... RVA to explain on average 0.021% of phenotypic variance ... 0.023% observed for CVA". Better to clarify that this per variant as I was initially confused, but also why do we need to talk about "phenotypic variance" rather than being consistent with discussing heritability? (Similarly elsewhere, e.g. L293 "per-SNP variance explained" why not talk about per SNP h2? Again at L323 "... phenotypic variance (and thus heritability)" why do we need to mention both?

We apologize for the confusion. The numbers 0.021% and 0.023% referred to averages over sets of trait-associated variants and, therefore, can be interpreted as per-variant statistics (here, proportion of phenotypic variance explained per-SNP). To help clarify this we have rephrase the sentence by adding "each": "After winner's curse correction,^{26,27} we found **each** RVA to explain on average 0.027% of phenotypic variance as compared to 0.023% for CVAs." In this paragraph, we prefer to report the proportion of phenotypic variance explained by SNPs rather than the proportion of heritability. However, Fig. 3a normalize those values (cumulated across all associations) relative to the heritability of each trait. We found both complementary and important to report independently.

L266 "CVAs and RVAs accounted for ... of the average common-variant and rare-variant heritability" I think you need to explain how the latter quantities are computed.

We apologize for the lack clarity. We have rephrased it as: "On average across traits, the cumulative proportion of phenotypic variance explained by CVAs and RVAs represents 31% (range across traits: 1.9% - 56%) and 12% (range across traits: 0.2% - 50%) of the average common-variant and rare-variant

heritability, respectively.”. For example, we detected 1,712 CVAs and 134 RVAs for height, which cumulatively explain 27.5% and 1.4% of height variance. We then compared these proportions of variance explained with our estimates of rare-variant and common-variant heritability from WGS data, that is 12.5% and 58.4% respectively. This means that to $27.5\%/58.4\%=47\%$ and $1.4\%/12.5\%=11\%$ of the estimated common-variant and rare-variant heritability of height is already explained by CVAs and RVAs, respectively.

L331 "replaced sequenced variants with imputed SNPs with MAF>0.01%" this needs a little explanation, if all sequence variants were removed how was the imputation done?

We apologize for the confusion here. We simply reran the same analysis pipelines using imputed SNPs instead of sequenced ones. The corrected sentence now reads as: “Therefore, we performed similar GWAS analyses as described above using imputed SNPs with MAF>0.01%.”

L414 I don't understand "the entire mutational target".

By “mutational target” we mean the population genetics concept of the set of DNA loci, which, if mutated, can causally affect a given phenotype and such that mutations outside of the target do not have any effect. We can convey the same idea using “causal variants” instead. Therefore, we have rephrased the second half of the sentence as “suggesting that the identified loci might contain all causal variants for height”.

L544 and 558 unnecessary repeat of "452,618 samples of European ancestry" but also not an exact repeat which may confuse.

Thank you. We have removed the second instance.

L560 (also L570) "estimated relationship" you don't "estimate" relatedness you compute a coefficient that measures relatedness. For an estimate there has to be a meaningful true value, there isn't for relatedness (e.g. pedigree-based values are arbitrary, depending on the choice of pedigree; GWAS-based coefficients are also arbitrary, there is no "true value").

We thank the referee for this point. We have rephrased that part of the sentence as : “We extracted a sparse GRM with non-zero entries for pairs of relatives with a genomic relationship coefficient above 0.05”.

L568 "median LD distribution" doesn't make sense to me. Firstly, LD is not a numerical quantity so doesn't have a median, but also do you mean a distribution of median values? If so distribution over what?

We apologize for the confusion there. We defined the groups using the median of the LD score statistic (i.e., the sum for squared correlations between a focal SNP and SNPs within a 1 Mb distance). We have rephrased this sentence as: “and further assigned a LD bin (based on the median LD score statistic within each MAF bin) (Error! Reference source not found.). LD score statistics were calculated for each SNP as the sum of squared correlation between allele counts at that SNP and that of all nearby SNPs within 1 Mb radius.”.

L668 typo "not affect to the" – [Corrected](#).

L687 I guess you mean "(see under Code availability)"

[Yes, that is what we mean. We have corrected that part accordingly.](#)

L714 I don't understand "... 100 k-means for birth coordinates"

[We mean that we created 100 k-means clusters using birth coordinates of UK Biobank participants. We now clarify this as: "\(including 100 k-means clusters determined from birth coordinates of UK Biobank participants\)".](#)

Referee #2 (Remarks to the Author):

Wainschtein and colleagues present results of analyses in the UK Biobank to evaluate the impact of rare genetic variation of complex human traits and diseases. They demonstrate that whole-genome sequence (WGS) variants (MAF >0.01%) capture ~91% of pedigree-based narrow sense heritability, with ~21% attributable to rare variants (MAF <1%). They assess the relative contribution of coding and non-coding variation to heritability amongst rare and common genetic variation and conduct GWAS analyses across phenotypes using WGS and imputed variants that highlight rare variant associations that account for a substantial proportion of the heritability of lipid traits. Overall, the manuscript is well written, providing insights to the key questions I have about the contribution of rare variants to complex phenotypes, and the methods used generally robust and adequately motivated (see some comments below). However, my enthusiasm for the manuscript is somewhat dampened by the focus of analyses only on individuals of European ancestry, which the authors highlight as a limitation of the work.

We thank this referee for their accurate summary of our work and for acknowledging that our manuscript is well written and provide key insights using robust methods. We did acknowledge that the ancestry diversity in the UK Biobank is limited because >90% of participants have European ancestry. However, robustly quantifying the contribution of rare variants to the heritability of complex traits, which was one of the main goals of our study, requires large sample sizes, which to date, are still skewed towards populations of European ancestry. We have expanded our discussion to mention a recent study that showed high consistency in SNP-based heritability estimates (based on common SNPs) between ancestry groups. We acknowledge that larger multi-ancestry studies are needed to bridge the gap for rare variants.

(Lines 492 – 499): “First, our analyses were restricted to individuals with European ancestries because of the limited sample sizes of other ancestry groups in the UK Biobank ($N < 12,000$), especially for studying rare variants. To date, there is a crucial need for heritability studies in other ancestry groups to better benchmark the accuracy of polygenic predictors of complex traits (including risk of disease) and refine understanding of their genetic architectures. Recent studies focusing on common variants have shown consistent heritability estimates between ancestry groups⁵⁰. However, future large scale and multi-ancestries studies using WGS data and family-based designs are still needed to bridge the gap.”

Comments:

1. The methods state that quantitative traits were standardised, but were they inverse rank normalised? Extreme outliers were removed, but with such a large sample size, subtle deviations from a Normal distribution will impact on findings and increase false positive associations with rare variants in GWAS analyses.

Quantitative traits were not rank-based transformed. However, we did exclude outliers (values >6 SD) to minimize skewness in trait distribution (METHODS: line 683). Finally, we applied a strict adjustment for population stratification by fitting (on top of other covariates) both common- and rare-variant PCs, as well as birth regions, which most GWAS in the UK Biobank have not done (METHODS lines 665 – 679). Ultimately, we replicated some of our findings in an independent cohort

for lipid traits (known to be skewed), which brings some reassurance on the robustness of our analyses (lines 268 – 283).

2. Does fastGWA-GLMM allow for case-control imbalance. If not, what impact would you expect this to have on your analyses?

Yes, fastGWA allows for case-control imbalance using a saddle point approximation. As suggested by Referee 1, we have repeated our GWAS analyses using Regenie and now report the latter analyses as our primary results. Our conclusions are largely unchanged.

3. The definition of “independence” using LD clumping is not ideal. With large sample sizes, a large effect size and highly significant association may still have a “shadow” at $r^2 < 0.01$. Given that individual-level data are available, were these independent associations confirmed through formal conditional analyses? The use of 1Mb window may also not be sufficient for the MHC.

We thank the referee for these comments. We have performed a formal conditional analysis by jointly fitting all associations detected on a chromosome and only keeping those that remain genome-wide significant after that step, which has indeed removed a few associations across phenotypes. We now only report and focus downstream analyses on this set of conditionally independent associations.

4. Presumably individuals from BioVU for these analyses were of European ancestry – this should be emphasized in the main text methods. Were sample sizes sufficiently large in BioVU to allow investigation of rare variant effects in other ancestry groups? Could the authors speculate on whether they expect their findings to be consistent in other ancestry groups?

We thank the referee for this question. The replication in the AGD cohort indeed focused on individuals of European ancestry. The replication analysis in AGD was previously only described in **Supplementary Note 2** but we now also emphasize in the main text that “We replicated this last result in an independent sample of approximately 67,000 unrelated European ancestry individuals in the Alliance for Genomic Discovery (AGD) cohort (METHODS, **Supplementary Note 2**).”. We initially did not attempt replication in other ancestry groups as the main goal of that analysis was to quantify how much LDL variance was explained by GWAS-associations in a European ancestry sample, compare estimate with our results in the UK Biobank, and most importantly benchmark it against our heritability estimates.

We have now extended our replication analyses for three traits (HDL, LDL and Alkaline phosphatase (ALK) levels) to a sample of 15,690 African ancestry individuals in AGD. In particular, we highlight 2 SNPs showing strong replication ($P < 5e-9$, and highly consistent effect sizes), which is partly explained by the fact that these 2 SNPs segregate at higher frequency (here, $> 3\%$) in the African ancestry group while their MAF was $< 1\%$ in European ancestry groups. We have now added the following sentence in our result section: “Interestingly, two ALK-associated RVAs (rs79257782 and rs73728135) had frequencies larger than 3% in African ancestry individuals in AGD (N=15,690) and

showed significant association with ALK ($P < 5 \times 10^{-8}$) with effect sizes highly consistent with those observed in European ancestry individuals from both UKB and AGD (**Supplementary Table 15**).

It is difficult to draw firm conclusions about other SNPs beyond these specific examples because the sample size in the AGD African ancestry group remains limited (N=15,690). More generally, the question of consistency of effect sizes between ancestry groups has been addressed in recent studies (Hu et al. Nat. Genet. 2025: PMID 39901012; Hou et al. Nat. Genet. 2023: PMID 36941441), which have estimated large correlations of effect sizes, typically >0.8 on average across traits. However, those estimated correlations are still noisy for specific traits (e.g., behavioural traits) and most analyses were focused on common variants. Therefore, the question remains open for specific traits and for rare variants.

5. In the imputed data analysis, were imputed genotypes first converted to hard genotype calls (I notice that a “missing genotype rate” is used for QC). If so, this is not optimal, and the dosage should be used instead.

We thank the referee for this question. We present below an overview of our revised analyses using Regenie.

We applied a missing call rate threshold of 0.1 to filter out variants with a large fraction of dosages too far from 0, 1 or 2. This step removed $\sim 0.57\%$ variants with uncertain dosage information. We subsequently kept a subset of the remaining variants with an imputation accuracy (INFO score statistic) larger than 0.3, which is classically used in GWAS studies. Finally, we ran the association analyses with Regenie using dosages of QC'ed variants.

We have now explicitly mentioned this in the Methods section named “GWAS of imputed SNPs from HRC+UK10K and TOPMed panels” where we further detail that the variant missingness filtering was performed on hardcalls while the association analysis was performed on the dosage genotypes.

6. The comparison of WGS and imputed variants is very interesting. Like the authors, I was not expecting that common variant associations would be missed and such a large proportion of rare variant associations captured with imputed data. Does WGS offer any advantages in terms of fine-mapping? This is difficult to quantify without knowing causal variants, but I wonder if credible set sizes around common and rare variants are smaller with WGS than imputed data, which would reflect improved fine-mapping resolution?

We thank the referee for this question. The benefit of WGS in terms of fine-mapping has been explored in previous studies using simulations or real data. For example, Wu et al. used simulations (based on a simple genetic architecture with one causal variant per locus) to show that $>90\%$ of causal variants are mapped within 100 kb to GWAS hits using WGS, while the same proportion remains $<60\%$ using imputation. The preprint from the UK Biobank WGS consortium also suggests a gain in fine-mapping resolution (<https://www.medrxiv.org/content/10.1101/2023.12.06.23299426v1.full.pdf>), although the paper does not report any critical data to support this claim.

Therefore, we compared fine-mapping resolution between imputed and sequenced variants across all our GWAS results. Overall, we observed a significant reduction in 95% credible sets sizes (with respect to both number of SNPs in the credible set and genomic range covered by the set) using WGS as compared to imputed data. As intuited by this referee, the improvement in mapping

resolution was stronger (~1.2-fold) for rare variant associations than for common variant associations. We also observed a larger posterior probability of inclusion (PIP) per credible sets for those identified using WGS data as compared to imputed data, which is a corollary of the former observation about credible sets.

We summarise the results of these secondary analyses in a short paragraph at the end the GWAS section and generated a summary figure reported as **Extended Data Fig. 6**.

7. The authors highlight that there is limited power, even in UK Biobank, to detect associations with rare variants with MAF <0.1%. Strategies to analyse rare coding variants through aggregate gene-based tests are well developed, but could the authors suggest approaches that could be applied for non-coding variants?

We thank the referee for this comment. For rare variants, our analyses showed that the genomic annotation that correlates best with per-SNP variance explained is the proximity of common variant association nearby. Therefore, we believe this could be a powerful mask to use in burden test analyses of non-coding rare variants as highlighted in our Discussion: “Moreover, consistent with previous studies, we found a strong colocalization between rare- and common-variant associations. Importantly, this aspect of the genetic architecture of complex traits can be utilised to improve GWAS discovery for rare (and ultra-rare) non-coding variants, for example, by aggregating pathogenic variants within loci containing common-variant associations in burden test analyses⁴².”

Minor points:

1. The definition of rare and common variants was not entirely clear – if MAF <1% and MAF >1%, respectively, is used, what about MAF =1%?

Apologies for the confusion. We define common variants as **MAF ≥ 1%**. We have corrected the abstract accordingly.

2. Paragraph beginning line 94. Emphasize that all of these analyses are undertaken in European ancestry individuals.

We now specify in that paragraph that our analyses are conducted in European ancestry individuals.

Referee #3 (Remarks to the Author):

Wainschtein and colleagues undertake a series of analyses using whole-genome sequencing data from ~350,000 individuals in the UK Biobank. Looking at 36 complex traits (predominantly lipid traits) the authors estimate the heritability of these traits, the proportion of this heritability captured by the WGS data, and the relative contribution of rare and common variation, plus coding and non-coding variation, to the heritability explained by the WGS data. The authors also undertake GWAS analyses for these 36 phenotypes.

We thank Reviewer 3 for summarizing our study. We would like to highlight that only four of the traits we examined are lipid-related—namely, LDL cholesterol, HDL cholesterol, triglycerides, and dyslipidaemia.

Excluded variants: In the Discussion, you rightly list various potential sources of missing heritability for the traits where $h^2_{WGS} < h^2_{PED}$. These include ultra-rare variants, structural variants, genome build gaps and non-additive effects. Given these exclusions, it surprises me that you have identified 15 traits where the WGS data explains 100% of h^2_{PED} . Do you believe this means that ultra-rare or structural variants, genome-build gaps, non-additive effects or sex-chromosome variants play no role in these traits/diseases? I find this hard to believe, and it makes me wonder if you have biased estimates of either h^2_{WGS} or h^2_{PED} . If you don't believe this is the case, then how should one interpret your results and their impact on genetic studies of these phenotypes moving forward?

We thank the Reviewer for these questions. We highlight a subset of traits for which h^2_{WGS} and h^2_{PED} estimates are not statistically different from each other. This does not imply that all these factors are not contributing at all — indeed, we cannot confirm the null hypothesis — but rather that, for these traits, any differences lie within the margins of standard errors (SE). Therefore, it remains possible that more precise estimates of h^2_{PED} (currently the less precise of the two estimates) could reveal statistically significant differences, although our results suggest such differences would likely be small. Typically, we focused on these 15 traits because the SEs of their h^2_{PED} estimates were $< 3\%$. Therefore, the 95% confidence interval (based on a one-sided test, given that h^2_{PED} is expected to be larger than h^2_{WGS}) for the difference between h^2_{WGS} and h^2_{PED} would be $\sim 5\%$ (that is, $qnorm(0.95) * \sqrt{0.01^2 + 0.03^2}$, where 0.01 and 0.03 the conservative SEs of h^2_{WGS} and h^2_{PED} estimates, respectively). Note that we conservatively used two-sided tests in the main text. We also reported a few traits (e.g., height, HDL, or number of children), for which the differences between h^2_{WGS} and h^2_{PED} estimates remain statistically significant ($P < 0.05$), thus underscoring differences in genetic architectures between complex phenotypes.

We have clarified this point by adding the following sentence to the discussion (lines 397 – 400): “Importantly, while more precise pedigree-based estimates from future studies may still reveal statistical differences from WGS-based estimates for those 15 traits, our results suggest that any such differences are likely to be small.”.

Robustness of findings: Following on from the above, and based on your description of the limitations of the study (a section which I appreciated), I am concerned that the findings of the work are not robust. The healthy volunteer bias, the exclusion of many variant classes, the low power to

accurately quantify rare variant effects, the variation in power across traits and diseases confounding comparisons make me question the robustness of the findings.

We thank the reviewer for this comment and for their positive feedback on this section of our Discussion. We largely share these concerns, which are indeed potential limitations. We emphasize the word *potential* because, at present, it remains unclear how much these factors might bias our results. Importantly, as shown in **Table 2**, our WGS-based heritability estimates for height, BMI, and smoking initiation are consistent with those from previous studies using TOPMed data, lending support to the robustness of our estimates. Unfortunately, comparable estimates for other traits were not available, which underscores that our study fills a major gap in the literature. Using a subset of traits and common variants, Tsuo et al. (<https://www.biorxiv.org/content/10.1101/2024.08.06.606846v2.full.pdf>) recently showed high consistency of SNP-based heritability estimates between the UK Biobank (UKB) and All Of Us (AoU) European ancestry sub-cohorts (regression slope UKB ~ AoU: 0.81). This also emphasizes the robustness of UKB-based results although analyses of rare variants were not available in that study. Interestingly, Tsuo et al. also report high consistency of estimates obtained in European (EUR) and African (AFR) ancestry sub-cohorts in AoU (regression slope AFR ~ EUR: 0.96), which addresses another concern raised by the referee.

Finally, while ultra-rare variants were excluded from our primary analyses, we later showed and discussed (**Supplementary Figure 11**) that this class of variants is unlikely to have a substantial impact. We fully acknowledge that further work is needed to address remaining gaps, but we believe such future efforts will complement, rather than undermine, the conclusions of our study.

We have added the following sentence to the discussion (lines 496 – 499): “Recent studies focusing on common variants have shown consistent heritability estimates between ancestry groups⁴⁴. However, future large scale and multi-ancestries studies using WGS data and family-based designs are still needed to bridge the gap.”.

Utility and interpretation of findings: In the introduction, you state that quantifying the contribution of different classes of genetic variants to variation in complex traits is crucial to designing optimal experiments, identifying causal variants, and delivering precision medicine and drug discovery. Perhaps this was the case previously when one had to decide between a common variant microarray, whole-exome sequencing or whole-genome sequencing. However, WGS will now be performed on most cohorts, so which subset of variants to assay is no longer a relevant study design question. I would have liked to see more discussion about how your results impact the study design of future genetic studies of the 36 traits and affect the delivery of precision medicine and drug target identification for these. How does this differ depending on the proportion of heritability explained, and where in the genome this heritability lies?

We thank the reviewer for this comment. We agree that WGS is becoming a gold standard although its cost-effectiveness is still debated relative to other approaches combining WES with SNP-array followed by imputation (Gaynor et al., 2024 Nat. Genet. PMID – 39322778; Regeneron approach). We highlight specific examples of rare variants explaining notable amounts of phenotypic variance, which would have been missed by WES or imputation. Overall, choosing the right genomic technology is a critical part of designing genetic studies but determining an optimal sample size is equally important. By providing precise estimates of rare-variant heritability (and therefore, of

average per-SNP variance explained), our study informs sample size calculation to detect rare-variant associations (RVAs) in future WGS-based GWAS. Finally, by quantifying the degree of colocalization between common- and rare-variant associations, our study also provides a key indication of how often RVAs can be expected near common-variant associations, which could also inform optimal statistical methods to detect these RVAs. For example, we previously mentioned creating rare variants aggregates around CVA loci to improve statistical power to detect RVA.

We acknowledge that the path from heritability estimation to drug discovery is not straightforward. Therefore, we have removed this from our introduction. Finally, we have now expanded our Discussion as follows (lines 443 – 460):

“The cost-effectiveness of WGS for GWAS discovery is still debated relative to alternatives combining WES with SNP-array genotyping followed by imputation⁵⁰. Our study shows that while imputation captures the bulk of genetic signals, it misses a small fraction of common-variant associations and, expectedly, a larger fraction of rare-variant associations. Although many rare-variant associations missed by imputation can be detected by WES, we nevertheless highlighted multiple variants explaining notable amounts of phenotypic variance, which were detected outside of genomic regions covered by WES technologies. Altogether, these observations support that WGS remains a future-proof option, providing costs continue to drop.

Moreover, while choosing the right genomic technology is a critical part of designing future genetic studies, determining an optimal sample size is equally important. By providing precise estimates of rare-variant heritability (and therefore, of average per-SNP variance explained), our study informs the expected accuracy of WGS-based polygenic scores and sample size calculation to detect rare-variant associations in future WGS-based GWAS. Finally, by quantifying the degree of colocalization between common- and rare-variant associations, our study also provides a key indication of how often rare-variant associations can be expected near common-variant associations, which could also inform optimal statistical methods to detect these rare-variant associations.”.

Portability of findings: Building on my concerns around the utility of these findings for the 36 traits and diseases included directly in this work, it wasn't clear to me how the results from this study are relevant to other complex diseases and traits. What general insights, if any, have been made about common complex traits and diseases? Also, what effect does the restriction of your analyses to the 36 traits with a marginally significant rare-variant heritability estimate have on the portability of your findings across other traits? The complex diseases were also dropped from some of your analyses because of the larger standard error around the h2PED estimate - does this impact negatively the portability of your findings?

We thank the referee for this comment. A key aim of our study was to assess how much familial heritability can be reliably explained by sequenced variants. Reliability is critical here as most studies that have attempted to answer this question previously were largely underpowered. Therefore, we focused on traits and diseases for which h2PED was also precisely estimated. This might have skewed our trait selection towards more heritable traits and diseases. However, this should not impact negatively the generalizability of our findings to other heritable traits or diseases.

Our findings predict that there is substantial rare genetic variation explaining inter-individual differences in complex traits and disease susceptibility, which is currently untapped because most GWAS (and more generally genetic studies) to date still focus on imputed genotypes. Our study quantifies how much gain in trait and disease risk prediction one can hope to achieve when GWAS

studies will start to use WGS routinely and sample sizes continue to grow. We report that the answer varies across traits but on average a ~20% gain is to be expected.

Although limited, our selected set of phenotypes displays a range of genetic architectures with h2PED varying between 0.1 and 0.9 and a normalized number of trait-associated loci (normalized by SNP-based heritability) varying between 110 and 2,600 (interquartile range: 458 – 1,453). Across these different genetic architectures, we show that WGS captures ~88% of heritability, on average (interquartile range: 78% - 99%). We believe this finding is generalizable across traits. In addition, observing highly similar genetic correlations between common and rare variants is also likely to be generalizable across traits. Finally, our quantification of the degree of colocalization between rare and common variant associations provides a strong prior for using common variant GWAS loci as a genomic unit for aggregating rare variants outside of coding regions.

The novelty of GWAS findings: Analyses of the UK Biobank WGS data are emerging, not least by the UK Biobank WGS Consortium. The novelty of the GWAS analyses presented in this work was not clear to me, and it wasn't clearly outlined in the paper (or Supplementary Table 11, as far as I could tell - it would be good if a legend could be supplied explaining the column headings). I suggest the authors compare their analyses to those undertaken in <https://www.medrxiv.org/content/10.1101/2023.12.06.23299426v1.full> and presented publically at <https://azphewas.com/>. I tried to do this for some variants myself, but due to the limited information supplied in the main text of the manuscript, I was not able to confidently find the indel downstream of HBQ1 associated with mean corpuscular volume in Supp Table 11(for example). See also a previous analysis of the UK biobank WGS data for telomere length: <https://www.nature.com/articles/s41588-024-01884-7>.

We thank the Reviewer for these questions and suggestions, which give us an opportunity to clarify the motivations of our GWAS analyses. Our main goal was to assess how much of the estimated heritability could already be captured by a GWAS design as previously mentioned in our submitted manuscript (Line 242 – 244): “We performed GWAS analyses of all 34 phenotypes to assess how much of their rare-variant heritability can already be mapped to single loci using WGS data from 452,618 genomes.”). We acknowledge that more and in-depth GWAS analysis have been and will continue to be conducted using WGS data in the UK Biobank. Therefore, we mainly focused on high-level genetic architecture statistics such as co-localization between common and rare variants associations as well as proportions of variance explained by SNPs identified. We have now added the following sentence to emphasize that more targeted analyses have been done in other studies: “More comprehensive and trait-focused GWAS analyses using WGS data in the UK Biobank have been conducted in previous studies²⁶⁻²⁹.”.

The case of the HBQ1 variant.

We initially highlighted the association near *HBQ1* (SNPID: 16:182397:AGAGT:A), as an example of associations explaining a large amount of phenotypic variance, which would have been missed by WES technologies. We are not sure of why this variant cannot be queried through the online platform. We also tried but could not find it. However, we do confirm that this variant is reported as associated with mean corpuscular volume (MCV) in the pre-print released by the UK Biobank WGS consortium (Supplementary Table 8), with an effect size of -1.399 (s.e. 0.0279) MCV standard

deviation (SD) per A allele, while we report -1.382 (s.e. 0.026) MCV SD per A allele. Note that this variant is also associated with red blood cell count (RBC) in our two studies with very similar effect sizes (i.e., ~0.85 RBC SD per A allele.).

Comparison with UK Biobank WGS consortium

We performed a systematic comparison of our results with those from the UK Biobank WGS consortium across 24 phenotypes (4 diseases and 20 quantitative traits) in common between our two studies. We focused on their Supplementary Table 8, which highlights 26,478 loci (7,859 loci across the 24 phenotypes in common) identified in European ancestry samples. Importantly, their definition of loci was only based on physical distance between variants (i.e., all associations within 500 kb of each other were grouped into a single locus) and did not account for linkage disequilibrium. Other important differences between our analyses are that 1) they used a p-value threshold of $5e-8$ for genome-wide significance, while we used $5e-9$; 2) they used a relatively sparse set of covariates (less than 25) while we fitted over 300 covariates, including common and rare variants PCs; and 3) we focused on variants with a frequency larger than 0.01%.

Overall, we found that >95% of the unique trait-variant associations identified in our study (across the 24 phenotypes in common) were located within the 7,859 loci identified by the UK Biobank WGS consortium. There were 2,536 trait-variant associations for which the lead SNP was identical between our two studies and the correlation of estimated effect sizes across those 2,536 trait-variant associations was >0.99 (**Fig. R1a** below). While this very high correlation is not utterly surprising given that we used very similar data and analytical methods, this observation still highlights the robustness of our results (and those of the UK Biobank WGS consortium).

Update on clumping results.

During the revision process, we identified an issue in our clumped results because too small p-values were automatically set to 0. Therefore, we could not identify the correct lead SNP when multiple SNPs at a locus had their p-values set to 0. After updating our results (using the logarithm of p-values), we identified another rare indel (rs754165241, intronic to ASGR1) explaining an even larger amount of phenotypic variance (i.e., >3% of variance in alkaline phosphatase levels) and still located outside of WES-covered regions. Rs754165241, was associated with 9 different traits and previously associated with alkaline phosphatase levels in another study. As a result, we have updated our abstract and corresponding sections in the manuscript to now highlight rs754165241 instead of 16:182397:AGAGT:A, initially mentioned. Importantly, the comparison with the UK Biobank WGS consortium above is based upon our updated clumped results.

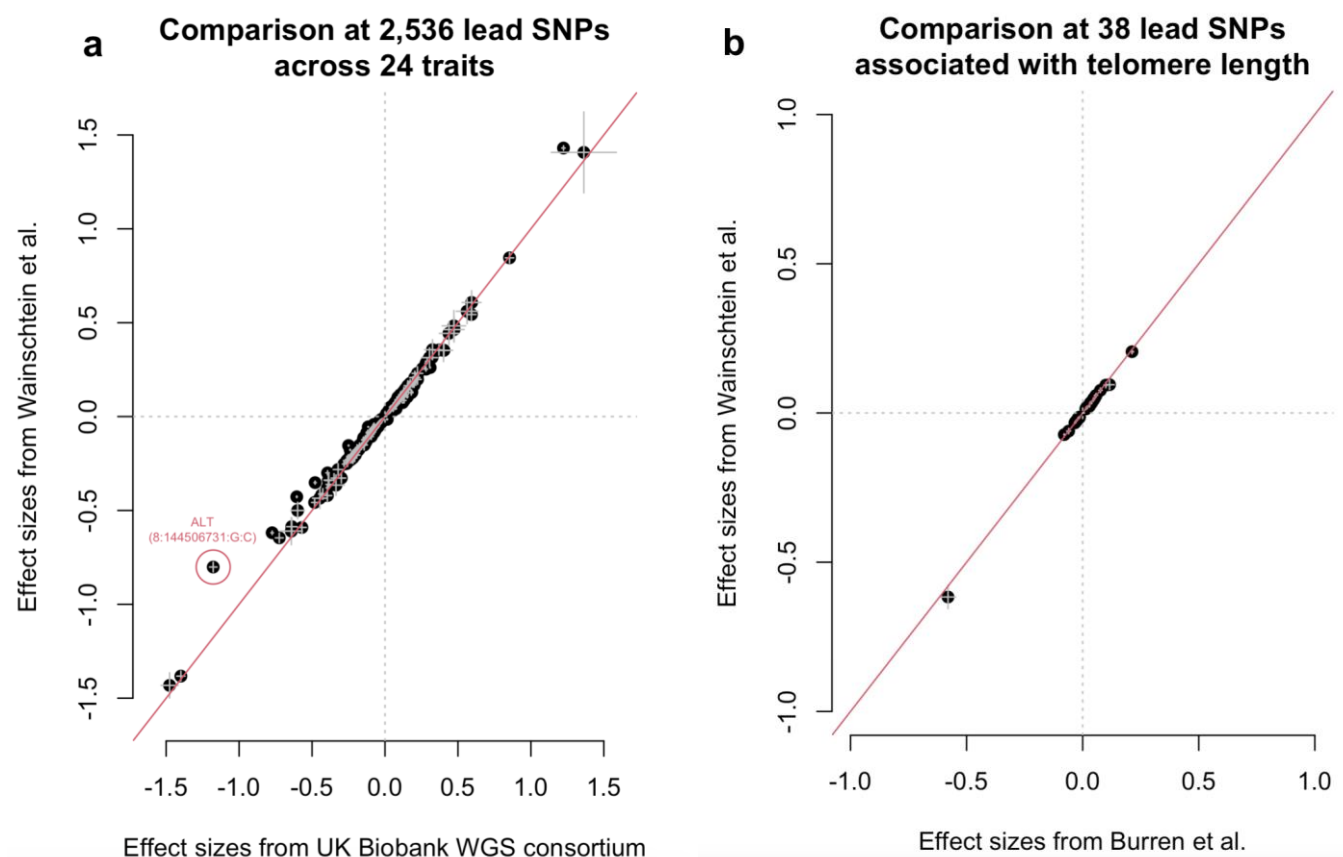


Fig. R1. Comparison of SNP effects across lead SNPs identified in our study and those identified in other studies using WGS data in the UK Biobank. Panel a: comparison with the UK Biobank WGS consortium. Note that the effect size of 8:144506731:G:C associated with alanine amino transferase levels is slightly attenuated in our study as compared to the UK Biobank WGS consortium. Panel b: comparison with that of Burren et al.

Comparison with Burren et al. (telomere length).

Regarding telomere length, we previously cited Burren et al. (2024) in our submitted manuscript when highlighting a rare variant association downstream *TINF2*, which was likely missed by imputation. Importantly, Burren et al. (2024) mainly used WGS to call some of their phenotypes while their GWAS analyses focused on 1) imputed SNPs with $MAF > 0.1\%$ and 2) only considered rare exonic sequenced SNPs. We are unsure why the authors did not use all the WGS-called genotypes for their GWAS analysis. Nevertheless, we compared our GWAS results with that from Burren et al. (2024), focusing on the qPCR-based phenotype in common between our two studies). For this phenotype, we detected 184 independent associations while Burren et al. only detected 134 associations (using imputed SNPs with $MAF > 0.01\%$ and imputation quality $INFO > 0.7$).

Thirty-eight (38) associations detected across our two studies had the same lead SNP. The correlation of estimated effect sizes between studies for these 38 SNPs was > 0.99 (Fig. R1b below). Moreover, over $> 78\%$ of our detected associations had a signal detected in Burren et al. within less than 100 kb (the median distance between signals detected across studies was 16.5 kb). However, 25 signals (that is $\sim 14\%$) detected in our analysis did not have any association from Burren et al., within 1 Mb. Interestingly, 5 of these 25 signals were detected in Burren et al. using a more heritable telomere length phenotype, aggregating WGS-based and qPCR-based phenotypes.

Conclusions for all comparisons

Altogether, our results are highly consistent with those from Burren et al. and from the UK Biobank WGS consortium which is expected given that we used very similar data and methods. Therefore, we prefer to keep these comparisons within the rebuttal document and not mention them in our revised manuscript as they are expected and somewhat peripheral to our main aim of our study. Nevertheless, we now clearly highlight that more comprehensive and targeted GWAS analyses using WGS data available in the UK Biobank have been conducted elsewhere and cite those studies accordingly.

Clustering of independent GWAS signals: If I understand correctly, you clumped together lead SNPs within 1Mb that had an $r^2 > 0.01$ (if this is correct then the inequality is incorrect in the Methods section). Is 1Mb broad enough given the number of very strong associations that will be detected by very weak tags? In my experience, association peaks in biobank-scale analyses can have very broad shoulders that extend beyond 1Mb. Not appropriately accounting for association signals driven by weak tagging of known effects could also drive the observation about the colocalisation of common and rare variants. Again, in my experience, LD-based approaches (LDSC etc) struggle to account for (multiple) large effects in a region and this can result in spurious residual association signals (rare or common). Perhaps try fitting a full model of lead variants in a region to see how this compares to your standard approach?

We thank the referee for this comment. We have performed a formal conditional analysis and now report our results accordingly. This conditional analysis we performed after selecting clumped SNPs with $r^2 < 0.01$ to ensure no residual LD between associations.

Referee #1 (Remarks to the Author):

I'm generally happy with the revisions made in response to reviewer comments. However it's very disappointing that the revised version I received is marred by dozens of "Error! Reference source not found", there was even one of these in the response letter. It seems nobody looked at the submitted version, otherwise these errors should have been obvious.

We thank Referee #1 for their overall positive feedback on our revised manuscript. We apologise for the issues with the references, which seems to have been created by Nature file management system, but we acknowledge that we should have been more careful about that.

Apologies I did not state my point well about LDMS not having been shown to "perform well". I won't pursue the point further, I don't have major concerns but there are limitations in Evans et al. and Wainschtein et al (2022) particularly due to unrealistic simulations that assign equal heritability to causal SNPs, and in any case these papers are not cited to support use of LDMS.

We thank Referee #1 for this point. We agree that these methods have limitations but also flexible enough to capture a wide range of genetic architectures.

I'm glad you've adopted TetraHer and Regenie. The preprint I previously cited is now published in Nature Genetics <https://www.nature.com/articles/s41588-025-02286-z> showing that LDK-KVIK is more powerful than Regenie and slightly faster. I won't pursue this point either, Regenie is at least better than fastGWA but it's a shame not to have used the best available software.

We thank Referee #1 for this point. We agree this is very fast-moving area and that new methods like LDK-KVIK or Quickdraws have been developed recently. Note that LDK-KVIK was published in August 2025, when our manuscript was under review, which had limited our ability to use it in our paper.

The response to my comment at previous L116 missed the point: "genomic relationship coefficient" is still ambiguous, there are lots of ways to measure relatedness from SNP data. I understand that an allelic correlation has become widely used and I guess is what you've used but I have not seen a good argument that this is an optimal choice of relatedness coefficient and in any case you need to be explicit what you have computed - you do give what I presume is the relevant formula at L645, you need to say that at first use.

We agree with Referee #1 on this point. We have added in the main text and METHODS section before providing the formula that genomic relationship was calculated as an allelic correlation (line 165 – 167): "Pairs of individuals were labelled as relatives when their genomic relationship coefficient (estimated as an allelic correlation; METHODS) exceeded 0.05."

L96 typo "account of" should be "for"

This has been fixed. Thank you.

L391 "fine-mapping resolution is improved ..." I am confused whether the improvement is from use of imputation or use of WGS.

We apologize for the confusion. We have rephrased the sentence as (Line 285 – 287): "We notably quantify the improvement in fine-mapping resolution from using WGS instead of imputed SNPs, while highlighting that existing imputation panels may still be missing common haplotypes in European ancestry populations".

Referee #2 (Remarks to the Author):

The authors have carried out extensive additional analyses to address the comments raised by the reviewers. I believe the use of REGENIE as an alternative and more robust analysis tool has been an important addition. I appreciated the attempt to look up results of rare variant associations in other ancestry groups, which highlighted some interesting results. No further comments.

We thank Referee #2 for their positive comments on our revised manuscript. We are grateful for their suggestions (and those of other referees), which have helped improve our manuscript.

Referee #3 (Remarks to the Author):

Thank you for your thoughtful and diligent responses to my comments and those of my fellow reviewers.

I think the changes you have made to the manuscript have further strengthened it. In particular, the use of SUSIE for fine-mapping, the joint modelling of all fine-mapped SNPs to ensure independence, and the switch to REGENIE increase my confidence in the robustness of the findings.

I congratulate all the authors on a very nice paper that neatly addresses important questions in the field of complex disease genetics.

We thank Referee #3 for acknowledging that the strength of our revised manuscript. We are thankful for their comments and suggestions.