nature research

Peer Review Information

Journal: Nature Genetics
Manuscript Title: Multivariate genomic analysis of 5 million people elucidates the genetic architecture of shared components of the metabolic syndrome
Corresponding author name(s): Woojae Myung, Hong-Hee Won

Reviewer Comments & Decisions:

Decision Letter, initial version:

9th January 2024

Dear Professor Won,

Your Article "Multivariate genomic analysis of 5 million people elucidates the genetic architecture of the metabolic syndrome" has been seen by three referees. You will see from their comments below that, while they find your work of potential interest, they have raised substantial concerns that must be addressed. In light of these comments, we cannot accept the manuscript for publication at this time, but we would be interested in considering a suitably revised version that addresses the referees' concerns.

We hope you will find the referees' comments useful as you decide how to proceed. If you wish to submit a substantially revised manuscript, please bear in mind that we will be reluctant to approach the referees again in the absence of major revisions.

To guide the scope of the revisions, the editors discuss the referee reports in detail within the team, including with the chief editor, with a view to identifying key priorities that should be addressed in revision, and sometimes overruling referee requests that are deemed beyond the scope of the current study. In this case, we ask that you thoroughly address all technical queries related to the multivariate association analyses and their interpretation, including further assessment of the degree of overlap with the single-trait association signals for each component trait using stringent thresholds for claiming independence, and extend the downstream analyses where feasible to provide further biological insights. We hope you will find this prioritized set of referee points to be useful when revising your study. Please do not hesitate to get in touch if you would like to discuss these issues further.

If you choose to revise your manuscript taking into account all reviewer and editor comments, please highlight all changes in the manuscript text file. At this stage, we will need you to upload a copy of the manuscript in MS Word .docx or similar editable format.

We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

If revising your manuscript:

*1) Include a "Response to referees" document detailing, point-by-point, how you addressed each referee comment. If no action was taken to address a point, you must provide a compelling argument. This response will be sent back to the referees along with the revised manuscript.

*2) If you have not done so already, please begin to revise your manuscript so that it conforms to our Article format instructions, available here.
Refer also to any guidelines provided in this letter.

*3) Include a revised version of any required Reporting Summary:
https://www.nature.com/documents/nr-reporting-summary.pdf
It will be available to referees (and, potentially, statisticians) to aid in their evaluation if the manuscript goes back for peer review.
A revised checklist is essential for re-review of the paper.

Please be aware of our guidelines on digital image standards.

You may use the link below to submit your revised manuscript and related files:

[redacted]

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

If you wish to submit a suitably revised manuscript, we hope to receive it within 3-6 months. If you cannot send it within this time, please let us know. We will be happy to consider your revision so long as nothing similar has been accepted for publication at Nature Genetics or published elsewhere. Should your manuscript be substantially delayed without notifying us in advance and your article is eventually published, the received date would be that of the revised, not the original, version.

Please do not hesitate to contact me if you have any questions or would like to discuss the required revisions further.

Nature Genetics is committed to improving transparency in authorship. As part of our efforts in this direction, we are requesting that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information, please visit www.springernature.com/orcid.

Thank you for the opportunity to review your work.

Sincerely,

Kyle

Kyle Vogan, PhD
Senior Editor
Nature Genetics
https://orcid.org/0000-0001-9565-9665

Referee expertise:

Referee #1: Genetics, cardiometabolic diseases, statistical methods

Referee #2: Genetics, cardiometabolic diseases, statistical methods

Referee #3: Genetics, cardiometabolic diseases, clinical translation

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:

Multivariate genomic analysis of 5 million people elucidates the genetic architecture of the metabolic syndrome.

Park et al. performed a large-scale multivariate GWAS of metabolic syndrome using genomic structural equation modelling.

I have some comments here for the authors to consider:

1. The authors used seven components of MetS including BMI, WC, T2D, FG, HTN, HDL-C and TG. Is T2D considered a component of MetS?

2. It would be helpful to have a study design figure to link the different analyses in a pictorial summary.

3. For the 1,270 variants that were independent of previous MetS GWAS, do they show stronger association with glycemic or dyslipidemia or both? Line 172 listed three numbers for four traits? How does these compare to the Qsnp? Some overlay tabulation or Venn diagram will be helpful to discuss the findings with respect to solely MetS, heterogenous effect, and which had been associated previously with the individual or multiple traits.

4. For the MetS polygenic scores analysis, it seems surprising that FG has the lowest incremental R2 and that the East Asian incremental R2 is relatively higher than European for HDL. Also, the incremental R2 is almost 30% for the MetS PRS in East Asian compared to European (62K for KoGES and 11K for European). This transferability seems much better than previously observed. Can the authors provide more insights to the observations? Also, what are F2, F1 and F3? Have the authors compared the PRS from PRS-CS with a simple genome-wide threshold lead variant PRS?

Minor comments:

1. Methods Line 421: "...non-overlapping GWAS samples and the METAL.". Sentence needs revision. Please review.

Reviewer #2:
Remarks to the Author:

Park and colleagues conducted multivariate GWAS of seven metabolic syndrome (MetS) components (body mass index (BMI), waist circumference, type 2 diabetes, fasting glucose, hypertension, HDL-C, and triglycerides). Their data included European population summary statistics from large research consortia as well as from three biobanks (UK Biobank, FinnGen, MVP) yielding an observed sample size of 4.9M. They utilized genomic structural equation modelling (Genomic-SEM) to run their multitrait GWAS based on a hierarchical factor model with three latent factors that clustered to obesity, insulin resistance or hypertension, and dyslipidemia. Their multitrait GWAS based on Genomic-SEM identified 1,650 lead SNPs for MetS across 939 loci (~40% novel compared to component signals; 77% novel for MetS).

The authors conducted a range of GWAS follow-up analyses, including genetic correlation analysis (method: LD score regression, LDSC; 96 significant genetic correlations with external factors; including brain grey matter volume), analysis of heritability (method: LDSC; SNP h2=11%), enrichment of heritability in functional categories (method: LDSC-SEG; genes at MetS loci enriched for expression in brain tissues and cells), gene prioritization (summary data-based Mendelian randomization, SMR; 6 genes prioritized robustly based on consistent effects in independent expression data), polygenic risk score analysis (method: PRS-CS; MetS PRS explains 19.3% of variation in UKB; performs equally well in ~60K independent East Asian individuals), PRS-PheWAS (method: logistic regression; 350 significant association with health outcomes in UKB) and causal association analyses (method: two-sample Mendelian randomization; MetS causally associated with 29 of 350 health outcomes that were associated with the PGS).

This is the largest multivariate GWAS of MetS to date. The data used and the Genomic-SEM method are appropriate to conduct a multivariate GWAS for MetS. Their main Genomic-SEM analysis included UKB. They excluded UKB to construct the PRS, which was then applied to an independent UKB data set, which is good. The methods for the follow-up analyses are state-of-the art, robust and the authors conducted useful sensitivity analyses. They did not conduct fine mapping for variant prioritization.

While I think that the paper is well done and robust, I still have the following two major comments:

1) Novelty of SNP associations: The authors identify 1,650 MetS lead variants. They claim that 704 (42.7%) were independent of GWAS signals of the seven MetS components based on squared correlation between variants ($r^2<0.1$) estimated from 1000 Genomes. Can the authors please include single trait GWAS results in their supplement and show what their multi-trait GWAS has identified in addition to the single trait GWAS's? I cannot believe that Genomic-SEM identified >40% additional signals compared to single trait GWAS of the MetS components. For example, for BMI, there are >12,000 associations in GWAS Catalog and ~1,000 truly independent variants have been identified by Yengo and colleagues (PMID: 30124842) based on conditional analyses (which are state of the art to derive "independence"). Please detail what the reference of known variants was. Also, a $r^2<0.1$ criterion is too liberal to claim independence; it refers to a $r\sim0.3$ and does not account for the haplotype. Many "uncorrelated" variants based on $r^2<0.1$ will not be truly independent signals in conditional analyses. Ideally, the authors conduct conditional analyses. If that is not possible, it would

be good if they revise their r2 criterion to be more stringent in their comparison with the single component GWAS's.

2) Specificity of results: The authors have identified an enrichment of expression effects in brain tissues for the genes harboring the MetS loci. How is that specific to MetS? This type of enrichment is well known for BMI genetic loci and I wonder whether the authors particularly picked up BMI genetics by their multitrait GWAS? Same for the PRS-PheWAS: a plethora of outcomes coming up would certainly also come up for BMI PRS. As with the first point, I am questioning on what is truly novel identified by this multitrait analysis (in comparison to single trait GWAS's)?

In addition, I have the following minor comments:

3) In line 79, the authors write: "For example, a GWAS conducted by the Global Lipids Genetics Consortium (GLGC), including approximately 1.3 million European individuals, identified 380 and 388 genetic variants associated with high-density lipoprotein cholesterol (HDL-C) and triglycerides (TG), respectively5". However, in the reference paper, there were >1700 distinct signals identified. It would be good if the authors could clarify the numbers.

4) Line 116/120 Supplementary Table 2/3 should be exchanged in the xlsx.

5) Line 186: Can the authors explain what type of pleiotropy is exerted via the MetS factor?

6) The text states 80 significant genetic correlations among 117 external traits but the respective figure shows 96.

7) The authors conducted robust MR analyses. Yet 22 of their 28 significant causal effects shown in Table 2 show significant heterogeneity. Even MR-PRESSO can be problematic for instrument selection in this case, and I suggest using a constrained maximum-likelihood method to select instruments (PMID 34214446).

8) Their PRS explains ~20% of MetS in UKB but heritability was estimated at only ~11%. I know the outcomes are not completely comparable but some discussion on what can be expected from ever larger MetS GWAS would be helpful.

Reviewer #3:
Remarks to the Author:

Enjoyed reading this paper on the genetic architecture of the metabolic syndrome (MetS) in European populations. Found 6 "genes" associated. PRS translatable to East Asians.

A major limitation is the focus on European populations especially since most data suggests other populations are higher risk (East and South Asian). This is acknowledged by the authors.

The title is a little misleading. The term Metabolic Syndrome has been used by various clinical organizations to define a clinical entity with the goal of determining who may be at greater risk of ASCVD, T2D and other adverse outcomes. Some of these organizations have published their own definitions, which can differ slightly and undergo revisions periodically (some of these are below). In this case, the title needs to say something like the "shared genetic architecture of components of the metabolic syndrome"

• Grundy, S. M., Brewer, H. B. Jr., Cleeman, J. I., Smith, S. C. Jr. & Lenfant, C. Definition of metabolic syndrome: Report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition. Circulation. 109(3), 433–438 (2004).
• Einhorn, D. American College of Endocrinology position statement on the insulin resistance syndrome. Endocr. Pract. 9, 5–21 (2003).
• Eckel, R. H., Grundy, S. M. & Zimmet, P. Z. The metabolic syndrome. Lancet. 365(9468), 1415–1428 (2005).

Lind did a GWAS for MetS based on criteria as above in the UKBB and was cited.

However, in this case, the authors have not performed a GWAS for MetS but rather looked at many large GWAS for MetS broadly related traits. This was a reasonable choice, but it really needs to be clear how the authors defined MetS.

It seems the motivation of this paper is a search for the genes underlying the overlapping part of the Venn diagram for these traits. This kind of clustering work has been done previously: Lotta, Dimas, Udler, Gloudemans, O'Rahilly etc. None of this prior work has been cited or used to place the current paper in context.

Also, if the authors are using a loose way of defining MetS, why not be even more broad? The underlying common feature for MetS is thought to largely be related to insulin resistance. Indeed, Reaven originally used the Insulin Resistance Syndrome to describe to concepts that underlie the idea of the MetS. While no surrogates capture insulin resistance perfectly, what is the correlation and overlap with GWAS signals for fasting insulin (although this was not measured in the UKBB has been assessed many other large studies)? What about for NAFLD?

While prioritizing MetS genes based on brain QTL data is reasonable especially for phenotypes related to BMI, the decision to use blood QTL data makes less sense. Why not use tissues known to be more strongly related to the phenotypes of the components of MetS (fat, liver, vascular tissue, muscle)?

A PRS for MetS again is a reasonable thing to try to develop (and show it functions ok in non-White populations) and it is not surprising that the MetS PRS explained the largest variance for prediction of MetS. But was it better at predicting adverse outcomes (cardiovascular or diabetes or death) than PRS using things like HTN, TG, T2D, LDL? I don't think a PRS for disease prediction will be that useful since we do not have interventions for MetS while we have many interventions for specific components of the MetS.

The results and discussion of the likely causal genes (FEZ2, STRA13, RFT1, MED23, SP1, HM13) could be further developed. What do these genes do? Do they act through common pathways? What is known about the non-brain genes in other tissues? Are there rare variants in these genes in human populations?

Marked limitations in literature review to build on the comments above. On a very cursory review, see that some but not all relevant papers were cited.

Cited:
Genome-Wide Association Study of the Metabolic Syndrome in UK Biobank.
Lind L. Metab Syndr Relat Disord. 2019 Dec;17(10):505-511. doi: 10.1089/met.2019.0070. Epub 2019 Oct 7. PMID: 31589552

Not cited:
A bivariate genome-wide approach to metabolic syndrome: STAMPEED consortium. Kraja AT...Borecki IB. Diabetes. 2011 Apr;60(4):1329-39. Epub 2011 Mar 8. PMID: 21386085

Transethnic meta-analysis of metabolic syndrome in a multiethnic study. Willems EL, Wan JY, Norden-Krichmar TM, Edwards KL, Santorico SA. Genet Epidemiol. 2020 Jan;44(1):16-25. doi: 10.1002/gepi.22267. Epub 2019 Oct 24. PMID: 31647587

Genome-wide association analysis of metabolic syndrome quantitative traits in the GENNID multiethnic family study. Wan JY, Goodman DL, Willems EL, Freedland AR, Norden-Krichmar TM, Santorico SA, Edwards KL; American Diabetes GENNID Study Group. Diabetol Metab Syndr. 2021 Jun 1;13(1):59. doi: 10.1186/s13098-021-00670-3. PMID: 34074324

Author Rebuttal to Initial comments

Manuscript Number: NG-A63720

Manuscript: Multivariate genomic analysis of 5 million people elucidates the genetic architecture of shared components of the metabolic syndrome

Responses from the Authors for Review Comments:

We appreciate the reviewers for their insightful and constructive comments. We believe that our study has been significantly improved and strengthened by revising the manuscript in response to these comments.

Reviewers' Comments:

Reviewer #1:

Remarks to the Author: Multivariate genomic analysis of 5 million people elucidates the genetic architecture of the metabolic syndrome.

Park et al. performed a large-scale multivariate GWAS of metabolic syndrome using genomic structural equation modelling.

I have some comments here for the authors to consider:

1. The authors used seven components of MetS including BMI, WC, T2D, FG, HTN, HDL-C and TG. Is T2D considered a component of MetS?

Response: Thank you for your valuable insight and comments. We appreciate it for pointing this out. In defining MetS, we followed the clinical criteria outlined by the American Heart Association/National Heart, Lung, and Blood Institute (AHA/NHLBI). As per these criteria, one of the five components is elevated fasting glucose (FG) ≥100 mg/dL or treatment for diabetes. While type 2 diabetes (T2D) is often viewed as a risk factor for MetS, the AHA/NHLBI definition allows individuals with T2D exhibiting elevated FG levels (≥100 mg/dL) or receiving diabetes medication to be considered as meeting this component. This interpretation is supported by Alberti *et al*. (2009), who note that the FG threshold of ≥100 mg/dL captures most T2D cases. Therefore, in our study, we have included T2D as a component of MetS, in line with the widely accepted AHA/NHLBI guidelines. Additionally, there is a significant correlation between MetS components and T2D, both clinically and genetically, suggesting shared genetic factors. This is particularly relevant to our study's aim of exploring the common genetic liability among these metabolic traits/diseases.

Acknowledging that we lacked an explanation of the MetS definition and the rationale behind the selection of MetS components, we have explained this in the **Supplementary Note.**

[Added to the Supplementary Note, pages 4–6, lines 69–116]

"**1. Summary of the clinical definitions of metabolic syndrome**

Metabolic syndrome (MetS) is a collection of risk factors that increase the risk of cardiovascular disease and type 2 diabetes (T2D). Despite its seemingly straightforward definition, it is still a challenge to diagnose MetS clinically[1-3].

The initial definition of MetS was established in 1998 by … (*omitted*)"

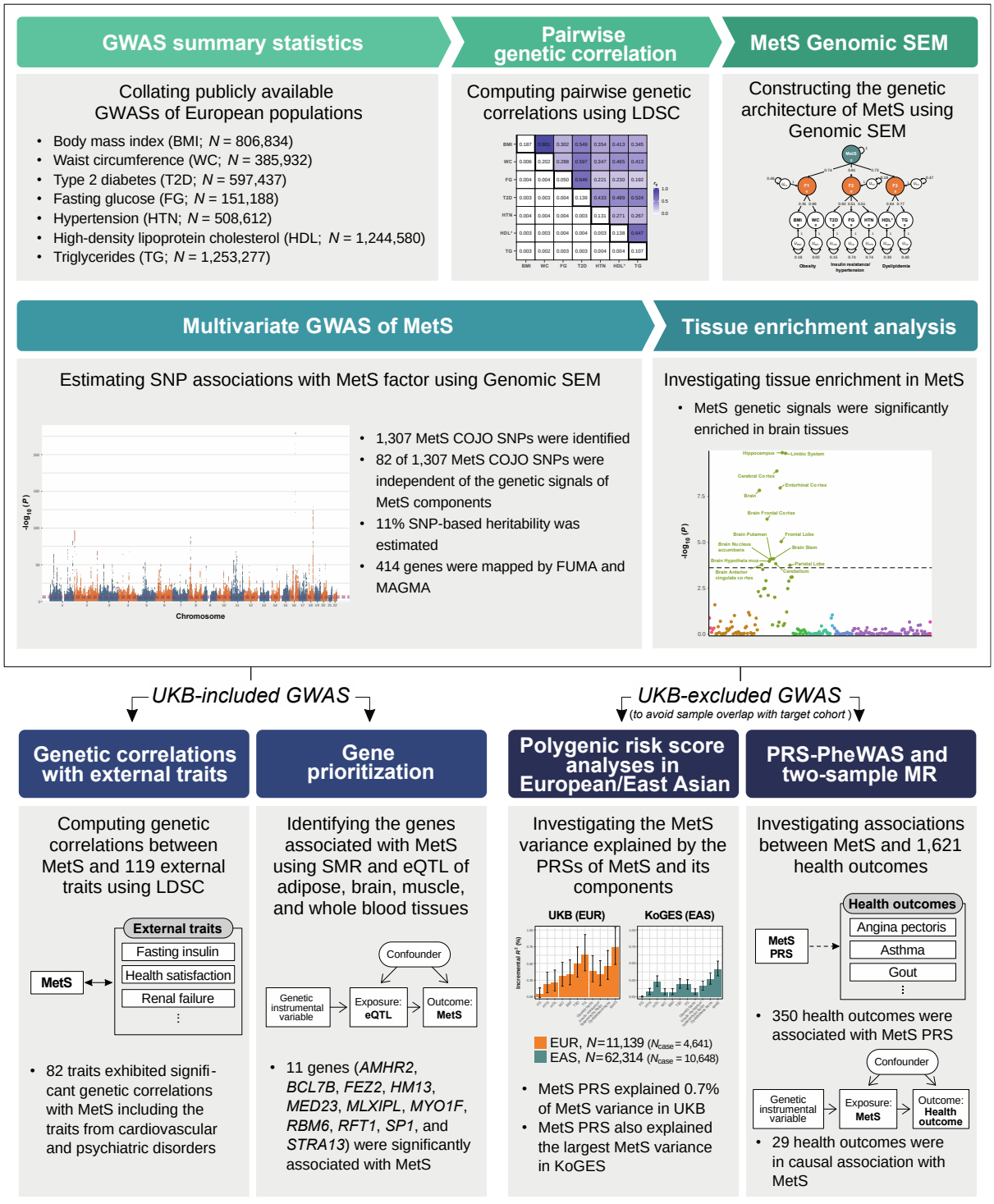
2. It would be helpful to have a study design figure to link the different analyses in a pictorial summary.

Response: Thank you for the comment. We created an overall study design figure and added it to **Supplementary Figure 1**.

[Added to the Supplementary Figure]

Supplementary Figure 1. Overview of multivariate GWAS of MetS

The workflow of this study is illustrated.



[Added to the Main, page 7, lines 121–123]

"Collectively, our results provided new insights into the complex genetic structure of MetS (**Supplementary Fig. 1**)."

3. For the 1,270 variants that were independent of previous MetS GWAS, do they show stronger association with glycemic or dyslipidemia or both?

Line 172 listed three numbers for four traits?

How does these compare to the Qsnp? Some overlay tabulation or Venn diagram will be helpful to discuss the findings with respect to solely MetS, heterogenous effect, and which had been associated previously with the individual or multiple traits.

Response: Thank you for this comment. In this revised version, we report independent signals identified through conditional analysis and the independence of SNPs determined with more stringent criteria in response to the comment from reviewer 2. Among the 1,307 MetS SNPs identified through conditional analysis (hereafter referred to as COJO MetS SNPs), 854 COJO MetS SNPs were independent of previous MetS GWASs. Among these, 245 (28.7%) COJO MetS SNPs exhibited genome-wide significant (GWS) association with dyslipidemia traits (triglyceride [TG] or high-density lipoprotein cholesterol [HDL] GWASs), and 143 (16.7%) COJO MetS SNPs showed GWS association with glycemic traits (fasting glucose [FG], type 2 diabetes [T2D], or hypertension [HTN]). Moreover, 66 (7.8%) COJO MetS SNPs were associated with GWS in both dyslipidemia and glycemic traits, highlighting the presence of a shared genetic architecture between them. Although a larger number of SNPs showed a significant genome-wide association with dyslipidemia compared with glycemic traits, it is insufficient to conclude that these COJO MetS SNPs have a stronger association with dyslipidemia relative to other constituent traits of MetS. This is because a simple comparison of p-values is subject to the statistical power of the GWAS for each trait.

In the previous version, we noted the number of independent SNPs identified for three first-order factors: F1 (obesity), F2 (insulin resistance/hypertension), and F3 (dyslipidemia) (line 172 in the previous version, line 191 in the revised version). We have rephrased this sentence to improve clarity and avoid confusion.

In response to the reviewer's suggestion, we have modified **Supplementary Table 12**, which now presents an overlay tabulation of comparisons of our findings with heterogeneous effects, MetS components, and previous MetS GWASs to facilitate easier comparisons. In addition, we have included **Supplementary Table 16** to allow for the query that signals from the GWAS in comparison were independent of MetS. Furthermore, we have added **Supplementary Figure 6** to illustrate the number of COJO MetS SNPs that were independent of the GWASs used in the comparison. **Table 1** has been updated accordingly.

[Modified Table 1]

Table 1. Summary of multivariate GWAS for MetS factor model

Factor	n_{eff}	h^2 (s.e.)	Mean χ^2	λ_{GC}	LDSC intercept (s.e.)	Attenuation ratio (s.e.)	*N* COJO SNP	Independent from Q_{SNP}	Independent from corresponding MetS components	Independent from previous studies	Unreported in GWAS catalog
MetS	1,384,348	0.1109 (0.003)	4.2501	2.8971	1.1208 (0.0259)	0.0372 (0.008)	1,307	811	82	848	159
F1	679,472	0.1721 (0.0048)	3.515	2.5641	1.1097 (0.0191)	0.0436 (0.0076)	718	677	14	-	40
F2	728,556	0.1136 (0.0039)	2.7305	1.9923	1.0529 (0.0208)	0.0306 (0.012)	496	346	57	-	57
F3	1,086,560	0.0993 (0.0069)	3.0111	1.6831	0.8766 (0.0234)	<0	608	329	0	-	85

Abbreviations: n_{eff}, effective sample size; s.e., standard error; λ_{GC}, genomic inflation factor; SNP, single-nucleotide polymorphism; F1, obesity factor; F2, insulin resistance/hypertension factor; F3, dyslipidemia factor

[a]Previous studies include van Walree *et al*. and Lind.

[Modified the Supplementary Table]

Supplementary Table 12. *MetS GWAS lead SNPs and COJO SNPs*

Genomic Locus	uniqID	rsID	CHR	POS	P	COJO	bJ	bJ_se	pJ	TG SNP in LD	HDL SNP in LD	HTN SNP in LD	FG SNP in LD	T2D SNP in LD	WC SNP in LD	Genomic Locus	
1	1:1865298:A:G	rs2803316	1	1865298	9.05E-10	Yes	-0.00766041	0.00125115	9.20E-10	rs2076327	rs2377037	NA	NA	NA	rs10907195	1	
2	1:21880032:A:G	rs7546430	1	21880032	3.20E-08	Yes	-0.0138359	0.00243557	1.34E-08	rs6605083;rs12731820	rs2377037	NA	NA	NA	rs10907195	2	
2	1:2725475:A:C	rs7537581	1	2725475	7.48E-11	Yes	-0.00828681	0.00124947	3.31E-11	NA	NA	NA	NA	NA	NA	2	
3	1:67153900:G:T	rs6577584	1	67153900	2.74E-22	Yes	-0.0124303	0.00131927	4.42E-21	NA	NA	rs11892	NA	rs6577586	rs6577585	3	
3	1:70614300:C:T	rs7525119	1	70614300	3.36E-08	No	NA	NA	NA	NA	NA	NA	NA	NA	NA	3	*(omitted)*
4	1:77283911:G:T	rs1891216	1	77283911	4.92E-09	Yes	-0.00752791	0.00127012	3.09E-09	NA	NA	NA	NA	NA	rs12128526	4	
4	1:77668166:C:G	rs12120907	1	77668166	2.23E-08	No	NA	NA	NA	NA	NA	NA	NA	NA	NA	4	
5	1:85522199:A:G	rs4908761	1	85522199	4.20E-11	Yes	-0.00875015	0.00128603	1.02E-11	NA	NA	rs301802	NA	NA	NA	5	
6	1:16851440:C:T	rs10907231	1	16851440	9.86E-09	Yes	0.0077867	0.0013276	4.49E-09	rs7538833	NA	NA	NA	NA	NA	6	
7	1:22249589:A:G	rs9426785	1	22249589	8.56E-11	Yes	0.00823814	0.00127868	1.17E-10	NA	rs11587362	rs12081298	NA	NA	NA	7	
8	1:23294671:A:G	rs637168	1	23294671	2.56E-09	Yes	0.0100178	0.00170072	3.85E-09	rs7551124;rs2742967	rs10753556; rs2298632	NA	NA	NA	rs506004	8	
(omitted)																	

[Added to the Supplementary Table]

Supplementary Table 16. *Lead SNPs of MetS components, Qsnp, previous MetS GWAS that are independent from GWAS of this study*

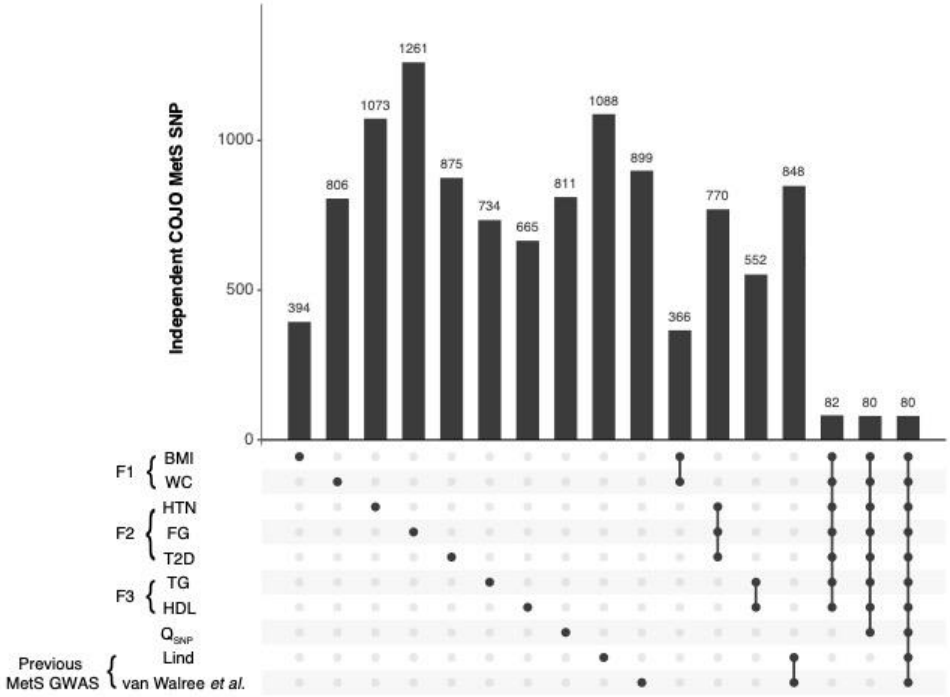
Trait	GenomicLocus	uniqID	SNP	chr	pos	p	nIndSigSNPs	IndSigSNPs	MetS COJO SNP in LD
TG	1	1:1686962:C:T	rs2076327	1	1686962	4.69E-08	1	rs2076327	rs2803316
TG	2	1:2147162:C:T	rs6605083	1	2147162	5.19E-13	3	rs6605083;rs10910028;rs262680	rs7546430
TG	2	1:2222368:A:G	rs12731820	1	2222368	3.18E-08	1	rs12731820	rs7546430
TG	3	1:15259476:C:T	rs12563724	1	15259476	3.59E-08	1	rs12563724	NA
TG	4	1:16504381:C:T	rs7538833	1	16504381	3.72E-15	9	rs7538833;rs11260930;rs6704422;rs34713890;rs61769918;rs28791627;rs7534979;rs12084478;rs4661712	rs10907231
TG	5	1:23785760:C:T	rs7551124	1	23785760	2.89E-17	5	rs7551124;rs1986133;rs2811964;rs2275355;rs10753556	rs637168
TG	5	1:23827720:G:T	rs2742967	1	23827720	3.25E-09	3	rs2742967;rs2275355;rs1986133	rs637168
TG	6	1:26232356:A:C	rs213641	1	26232356	1.18E-11	1	rs213641	NA
TG	7	1:26879792:C:T	rs4970489	1	26879792	2.16E-15	4	rs4970489;rs364977;rs4274112;rs71640328	rs6666121
TG	7	1:26919414:C:T	rs12408288	1	26919414	4.31E-09	3	rs12408288;rs11809021;rs4589135	rs11810321;rs6666121
TG	7	1:27029551:G:T	rs6666121	1	27029551	2.19E-12	4	rs6666121;rs4589135;rs12727590;rs11809021	rs11810321;rs6666121
TG	7	1:27262545:C:T	rs182050989	1	27262545	1.21E-38	7	rs182050989;rs12125238;rs12410656;rs186042737;rs34517168;rs11810321;rs12727590	rs11810321;rs6666121
TG	8	1:32197257:A:G	rs3766823	1	32197257	3.48E-08	1	rs3766823	rs3766823

(omitted)

[Added to the Supplementary Figure]

Supplementary Figure 6. COJO MetS SNPs independent from MetS components, Q_{SNP}, and previous MetS studies

The UpSet plot presents the number of COJO MetS SNPs that were independent of the lead SNPs of GWAS from MetS components, Q_{SNP}, and previous MetS studies. Bar chart shows the number of independent COJO MetS SNPs for each trait. The connected dots in the bottom panel represent traits that were considered simultaneously to determine the independence of the COJO MetS SNPs.



[Rephrased in the Results, page 9, lines 171–174]

"We further conducted a conditional and joint analysis (COJO)[21] on 1,650 MetS lead SNPs to report statistically significant and independent SNPs, of which 1,307 COJO MetS SNPs were identified (**Supplementary Table 12**)."

[Rephrased in the Results, page 9, lines 181–187]

"We then assessed the independence of the COJO MetS SNPs, with a window size of 500 kb and r^2 threshold of <0.01, from previously reported signals. Among the 1,307 COJO MetS SNPs, 82

(6.3%) were independent of the GWAS signals of the seven MetS components included in the Genomic SEM, 848 (64.9%) were independent of previous MetS GWAS (Lind[10] and van Walree *et al.*[22]), and 159 (12.2%) were previously unreported in the NHGRI-EBI GWAS Catalog[23] using FUMA (**Table 1**, **Supplementary Tables 11 and 16**, **Supplementary Fig. 6**)."

[**Added to the Methods, page 24, lines 534–537**]
"The identified lead SNPs were subjected to GCTA-COJO[21] v1.94.1 and determined whether they were conditionally and jointly associated SNPs through a stepwise model selection procedure with default parameters (i.e., *P*-value $<5 \times 10^{-8}$, window of 10 Mb, and collinearity <0.9)."

4. For the MetS polygenic scores analysis, it seems surprising that FG has the lowest incremental R2 and that the East Asian incremental R2 is relatively higher than European for HDL. Also, the incremental R2 is almost 30% for the MetS PRS in East Asian compared to European (62K for KoGES and 11K for European). This transferability seems much better than previously observed. Can the authors provide more insights to the observations? Also, what are F2, F1 and F3? Have the authors compared the PRS from PRS-CS with a simple genome-wide threshold lead variant PRS?

Response: We appreciate the reviewer's keen comment regarding the PRS analysis. We acknowledge that our initial methodology, which used the relative increase in the incremental R^2, may have led to misunderstandings. This approach notably resulted in higher incremental R^2 values in the EAS populations than in the EUR populations for certain traits. In this revised version, we have calculated the incremental R^2 by determining the difference in R^2 values between the baseline and PRS models, rather than assessing the relative increase with respect to the baseline. We are grateful to the reviewer for the meticulous review and valuable insights that have allowed us to enhance our analysis and reduce the possibility of misinterpretations.

In addition, we wish to clarify that the primary aim of conducting PRS analysis within the East Asian (EAS) population was to demonstrate the applicability of PRS derived from European GWAS to both the UKB and KoGES cohorts and not to compare PRS performance across

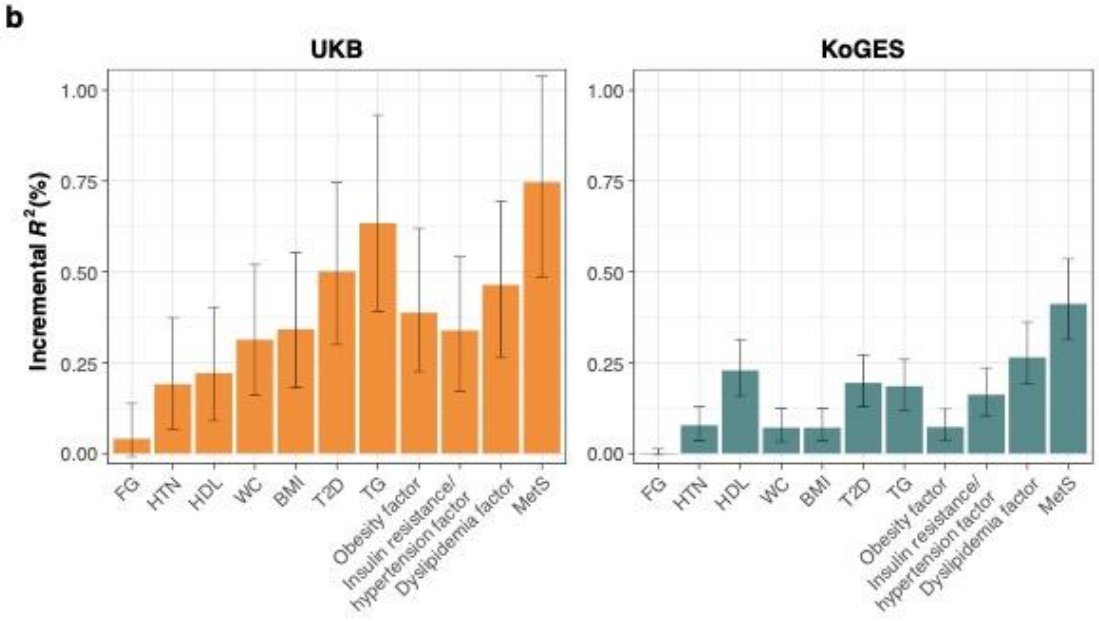
European (EUR) and EAS populations. In pursuit of this objective and to enhance clarity, we have modified **Figure 4b** to present separate bar plots for each target cohort, ensuring a more precise representation of our findings.

F1, F2, and F3 are the first-order factors from the multivariate Genomic SEM model that represent obesity, insulin resistance/hypertension, and dyslipidemia, respectively. We have added the full forms of F1, F2, and F3 to **Figure 4b**.

Following the reviewer's suggestion, we have used PRSice–2 (described in **Supplementary Note**), which adopts a pruning and thresholding strategy to compute the PRS. Using the PRS from PRSice–2 confirmed that the MetS PRS explained the largest variance in both the UKB and KoGES cohorts (**Supplementary Table 39**). In addition, FG PRS showed the lowest PRS performance in both cohorts, which could be partly due to the FG GWAS summary statistics with a small sample size of 151,188 compared with other traits (n = 232,101–888,227).

[Modified Figure 4]

Figure 4b*. Bar plot illustrating the incremental proportion of variance explained (ΔR^2) by the polygenic risk score of seven MetS components and four latent factors for predicting MetS in UKB and KoGES as target cohorts. The error bars indicate 95% CIs for ΔR^2, and they were computed using the percentile method of bootstrapping with 1,000 iterations.*


b
UKB
KoGES
Incremental R^2(%)
1.00
0.75
0.50
0.25
0.00
FG
HTN
HDL
WC
BMI
T2D
TG
Obesity factor
Insulin resistance/ hypertension factor
Dyslipidemia factor
MetS

[Added to the Supplementary Table]

Supplementary Table 39. *European and East Asian MetS polygenic risk score incremental R2 analyses using PRSice–2*

PRS	Target cohort	N observed	PRS estimate	SE	P	PRS estimate lower bound	PRS estimate upper bound	OR	OR lower bound	OR upper bound	Null R2	PRS R2	Incremental R2 (%)	Incremental R2 (%) lower bound	Incremental R2 (%) upper bound
BMI	UKB	11139	0.191464301	0.020130365	1.88E-21	0.152009511	0.230919091	1.2110216	1.164171309	1.25975731	0.038733652	0.044656975	0.59%	0.36%	0.86%
FG	UKB	11139	0.088188453	0.019819369	8.60E-06	0.049343203	0.127033702	1.092193929	1.050580852	1.135455284	0.038733652	0.039913854	0.12%	0.02%	0.25%
HDL	UKB	11139	-0.1197418	0.019863342	1.66E-09	-0.158673236	-0.080810364	0.887149469	0.853275133	0.922368589	0.038733652	0.041015848	0.23%	0.10%	0.41%
HTN	UKB	11139	0.115860411	0.019780147	4.70E-09	0.077092035	0.154628787	1.122839125	1.080141482	1.167224592	0.038733652	0.040879843	0.21%	0.09%	0.40%
T2D	UKB	11139	0.192234847	0.020065152	9.65E-22	0.152907871	0.231561823	1.211955107	1.165217623	1.260567256	0.038733652	0.044741318	0.60%	0.37%	0.86%
TG	UKB	11139	0.255869444	0.020599424	2.01E-35	0.215495316	0.296243572	1.291584093	1.240476173	1.344797673	0.038733652	0.049019056	1.03%	0.71%	1.39%
WC	UKB	11139	0.197332953	0.020062205	7.87E-23	0.158011753	0.236654152	1.218149559	1.17117996	1.267002852	0.038733652	0.045083683	0.64%	0.39%	0.91%
F1	UKB	11139	0.217311426	0.020154613	4.18E-27	0.177809111	0.256813742	1.242731061	1.194597264	1.292804309	0.038733652	0.046411435	0.77%	0.52%	1.09%
F2	UKB	11139	0.195056937	0.01998725	1.69E-22	0.155882646	0.234231228	1.215380185	1.168689045	1.263936716	0.038733652	0.044978565	0.62%	0.39%	0.90%
F3	UKB	11139	0.215311714	0.020249989	2.09E-26	0.175628086	0.255006194	1.240255169	1.191994657	1.290469614	0.038733652	0.046188609	0.75%	0.49%	1.03%
MetS	UKB	11139	0.307192368	0.020644061	4.42E-50	0.266730752	0.347653984	1.359602487	1.305688845	1.415742296	0.038733652	0.053689833	1.50%	1.12%	1.86%
MetS	KoGES	62314	0.305136282	0.011005274	3.36E-169	0.283566341	0.326706223	1.356809899	1.327856968	1.386394126	0.014404896	0.028106846	1.37%	1.18%	1.56%
BMI	KoGES	62314	0.146219448	0.010779425	6.49E-42	0.125092164	0.167346733	1.157450161	1.133252893	1.182164089	0.014404896	0.017606128	0.32%	0.24%	0.42%
WC	KoGES	62314	0.125002827	0.010742174	2.68E-31	0.103948552	0.146057102	1.133151657	1.10954337	1.157262268	0.014404896	0.016749869	0.23%	0.16%	0.32%
FG	KoGES	62314	0.071756353	0.010713345	2.12E-11	0.050758583	0.092754124	1.074393354	1.052068875	1.097191929	0.014404896	0.015156116	0.08%	0.03%	0.13%
T2D	KoGES	62314	0.150278752	0.0108729	1.89E-43	0.128968259	0.171589246	1.162158152	1.137654013	1.187190089	0.014404896	0.017730316	0.33%	0.25%	0.44%
HTN	KoGES	62314	0.085508167	0.010761469	1.93E-15	0.064416076	0.106600259	1.089270458	1.066536066	1.112489458	0.014404896	0.015477371	0.11%	0.06%	0.17%

HDL	KoGES	62314	-0.266202162	0.010943998	1.09E-130	-0.287652004	-0.244752319	0.766284199	0.750022552	0.782898423	0.014404896	0.024911739	1.05%	0.89%	1.22%
TG	KoGES	62314	0.265555193	0.011005914	1.26E-128	0.243983998	0.287126388	1.304154832	1.276323907	1.332592626	0.014404896	0.024727538	1.03%	0.88%	1.20%
F1	KoGES	62314	0.141592833	0.010767736	1.71E-39	0.120488457	0.162697208	1.152107452	1.128047720	1.176680347	0.014404896	0.017410745	0.30%	0.22%	0.40%
F2	KoGES	62314	0.151905577	0.010852301	1.61E-44	0.130635459	0.173175695	1.164050318	1.139552292	1.189075002	0.014404896	0.017816874	0.34%	0.25%	0.45%
F3	KoGES	62314	0.289180654	0.010976137	5.66E-153	0.267667821	0.310693488	1.335332941	1.306912938	1.364370961	0.014404896	0.026752941	1.23%	1.06%	1.42%

[Added to the Supplementary Note, pages 19–20, lines 414–425]

"**9. Polygenic risk score computation using PRSice–2**

The PRSice–2[51] employs a pruning and thresholding (P+T) strategy using various subsets of independent SNPs at different *P*-value cutoffs to calculate the polygenic risk score (PRS). First, SNPs underwent a process called clumping (also known as pruning) to identify independent SNPs (using a 500 kb window and an r^2 of 0.1). Subsequently, these independent SNPs were categorized into subsets according to *P*-value thresholds (0.001, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, and 1), and the PRS was calculated for each group. The best *P*-value threshold is determined by identifying the PRS that accounts for the largest variance in the target phenotype (i.e., MetS) within the validation cohort. Although using the same validation data for the best *P*-value selection as the target data for the PRS analysis of incremental R^2 might predispose to overfitting, this potential bias is substantially mitigated when the validation set is sufficiently large[52]."

[Rephrased in the Results, page 13, lines 283–285]

"The MetS PRS explained the largest variance for predicting MetS (ΔR^2 = 0.75%, 95% CI = 0.49–1.04%), followed by the TG PRS (ΔR^2 = 0.63%, 95% CI = 0.39–0.93%) and the T2D PRS (ΔR^2 = 0.5%, 95% CI = 0.3–0.75%)."

[Rephrased in the Results, page 14, lines 306–310]

"Similar to the UKB target, the MetS PRS demonstrated the largest ΔR^2 in KoGES (UKB ΔR^2 = 0.75%, 95% CI = 0.49–1.04%; KoGES ΔR^2 = 0.41%, 95% CI = 0.31–0.54%), and a similar ΔR^2 pattern was evident in PRS computed using PRSice–2[37] (**Supplementary Note**, **Supplementary Table 39**). These findings suggest the potential transferability of European MetS GWAS findings to diverse populations."

[Rephrased in the Discussion, page 19, lines 411–418]

"The MetS PRS demonstrated superior predictive power for dichotomized MetS in both cohorts compared with the PRSs of its components, which is consistent with MetS exhibiting the highest PAT. In contrast, the FG accounted for the least MetS variance in both cohorts. This may be

attributed to the fact that the performance of PRS depends on the GWAS sample size[59], and the sample size of the FG GWAS was comparatively smaller than that of the other components. These findings highlight a promising scope for wider application of the MetS PRS across different populations, yet they stress the need for GWAS with larger sample sizes."

Minor comments:

Methods Line 421: "...non-overlapping GWAS samples and the METAL.". Sentence needs revision. Please review.

Response: We appreciate your comment. We have rephrased the sentence accordingly.

[Rephrased in the Methods, page 21, lines 462–463]

"We conducted a fixed-effects meta-analysis of T2D and HTN using METAL[62] to increase the sample size of the corresponding GWAS."

Reviewer #2:

Remarks to the Author: Park and colleagues conducted multivariate GWAS of seven metabolic syndrome (MetS) components (body mass index (BMI), waist circumference, type 2 diabetes, fasting glucose, hypertension, HDL-C, and triglycerides). Their data included European population summary statistics from large research consortia as well as from three biobanks (UK Biobank, FinnGen, MVP) yielding an observed sample size of 4.9M. They utilized genomic structural equation modelling (Genomic-SEM) to run their multitrait GWAS based on a hierarchical factor model with three latent factors that clustered to obesity, insulin resistance or hypertension, and dyslipidemia. Their multitrait GWAS based on Genomic-SEM identified 1,650 lead SNPs for MetS across 939 loci (~40% novel compared to component signals; 77% novel for MetS).

The authors conducted a range of GWAS follow-up analyses, including genetic correlation analysis (method: LD score regression, LDSC; 96 significant genetic correlations with external factors; including brain grey matter volume), analysis of heritability (method: LDSC; SNP h2=11%), enrichment of heritability in functional categories (method: LDSC-SEG; genes at MetS loci enriched for expression in brain tissues and cells), gene prioritization (summary data-based Mendelian randomization, SMR; 6 genes prioritized robustly based on consistent effects in independent expression data), polygenic risk score analysis (method: PRS-CS; MetS PRS explains 19.3% of variation in UKB; performs equally well in ~60K independent East Asian individuals), PRS-PheWAS (method: logistic regression; 350 significant association with health outcomes in UKB) and causal association analyses (method: two-sample Mendelian randomization; MetS causally associated with 29 of 350 health outcomes that were associated with the PGS).

This is the largest multivariate GWAS of MetS to date. The data used and the Genomic-SEM method are appropriate to conduct a multivariate GWAS for MetS. Their main Genomic-SEM analysis included UKB. They excluded UKB to construct the PRS, which was then applied to an independent UKB data set, which is good. The methods for the follow-up analyses are state-of-the

art, robust and the authors conducted useful sensitivity analyses. They did not conduct fine mapping for variant prioritization.

While I think that the paper is well done and robust, I still have the following two major comments:

1) Novelty of SNP associations: The authors identify 1,650 MetS lead variants. They claim that 704 (42.7%) were independent of GWAS signals of the seven MetS components based on squared correlation between variants (r2<0.1) estimated from 1000 Genomes. Can the authors please include single trait GWAS results in their supplement and show what their multi-trait GWAS has identified in addition to the single trait GWAS's? I cannot believe that Genomic-SEM identified >40% additional signals compared to single trait GWAS of the MetS components. For example, for BMI, there are >12,000 associations in GWAS Catalog and ~1,000 truly independent variants have been identified by Yengo and colleagues (PMID: 30124842) based on conditional analyses (which are state of the art to derive "independence"). Please detail what the reference of known variants was. Also, a r2<0.1 criterion is too liberal to claim independence; it refers to a r~0.3 and does not account for the haplotype. Many "uncorrelated" variants based on r2<0.1 will not be truly independent signals in conditional analyses. Ideally, the authors conduct conditional analyses. If that is not possible, it would be good if they revise their r2 criterion to be more stringent in their comparison with the single component GWAS's.

Response: We appreciate the reviewer's insightful comment. As the reviewer pointed out, deciding on the independence of discovered genetic signals compared with the previously reported GWAS using r^2 <0.1 could be a lenient threshold and may lead to reporting an inflated number of independent SNPs. Following the reviewer's advice, to report the truly independent MetS genetic signal, we performed conditional analysis on the MetS GWAS from our study using conditional and joint analysis (COJO) implemented in the GCTA software and determined their independence from the GWASs of MetS components with a more stringent r^2 criterion of <0.01.

By clumping using FUMA, we initially identified 1,650 MetS lead SNPs associated with MetS. Applying the default stepwise model selection procedure to the lead SNPs, 1,307 SNPs remained significantly associated with MetS. These SNPs, refined through COJO analysis, are henceforth referred to as COJO MetS SNPs.

To ensure novelty and independence, we have applied a stringent r^2 threshold of <0.01, comparing these COJO MetS SNPs with previously reported SNPs in a MetS-component GWAS. Among the 1,307 COJO MetS SNPs, we identified 734, 665, 1,073, 1,261, 875, 806, and 394 independent COJO MetS SNPs from the TG, HDL, HTN, FG, T2D, WC, and BMI GWAS, respectively. Overall, 82 COJO MetS SNPs were independent of all MetS component GWASs, and we report that these genetic signals are truly significant, independent, and specific to MetS. We have added **Supplementary Figure 6** to illustrate the number of COJO MetS SNPs that were independent of the GWASs used in the comparison. In addition, we have added **Supplementary Table 12** to allow readers to easily distinguish which COJO MetS SNPs were independent or identified from previous GWASs of MetS components, and **Supplementary Table 16** for the opposite. Owing to the changes in the number of SNPs reported, we have updated the results of all relevant subsequence analyses, as shown in **Table 1**.

[Modified Table 1]

Table 1. Summary of multivariate GWAS for MetS factor model

Factor	n_{eff}	h^2 (s.e.)	Mean χ^2	λ_{GC}	LDSC intercept (s.e.)	Attenuation ratio (s.e.)	*N* COJO SNP	Independent from Q_{SNP}	Independent from corresponding MetS components	Independent from previous studies	Unreported in GWAS catalog
MetS	1,384,348	0.1109 (0.003)	4.2501	2.8971	1.1208 (0.0259)	0.0372 (0.008)	1,307	811	82	848	159
F1	679,472	0.1721 (0.0048)	3.515	2.5641	1.1097 (0.0191)	0.0436 (0.0076)	718	677	14	-	40
F2	728,556	0.1136 (0.0039)	2.7305	1.9923	1.0529 (0.0208)	0.0306 (0.012)	496	346	57	-	57
F3	1,086,560	0.0993 (0.0069)	3.0111	1.6831	0.8766 (0.0234)	<0	608	329	0	-	85

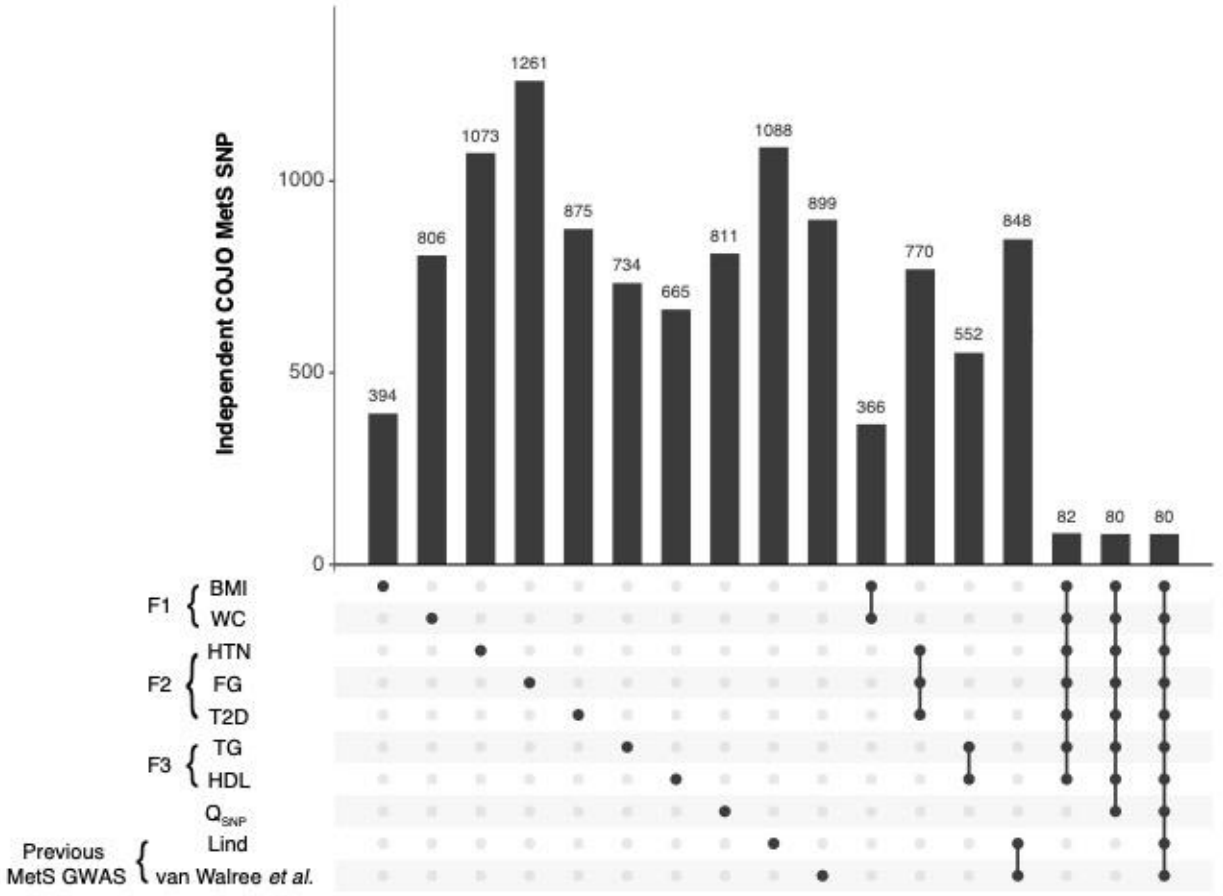
Abbreviations: n_{eff}, effective sample size; s.e., standard error; λ_{GC}, genomic inflation factor; SNP, single-nucleotide polymorphism; F1, obesity factor; F2, insulin-resistance/hypertension factor; F3, dyslipidemia factor

[a]Previous studies include van Walree *et al*. and Lind.

[Added to the Supplementary Figure]

Supplementary Figure 6. COJO MetS SNPs independent from MetS components, Q_{SNP}, and previous MetS studies

The UpSet plot presents the number of COJO MetS SNPs that were independent of the lead SNPs of GWAS from MetS components, Q_{SNP}, and previous MetS studies. Bar chart shows the number of independent COJO MetS SNPs for each trait. The connected dots in the bottom panel represent traits that were considered simultaneously to determine the independence of the COJO MetS SNPs.



[Modified the Supplementary Table]

Supplementary Table 12. *MetS GWAS lead SNPs and COJO SNPs*

Genomic Locus	uniqID	rsID	CHR	POS	P	COJO	bJ	bJ_se	pJ	TG SNP in LD	HDL SNP in LD	HTN SNP in LD	FG SNP in LD	T2D SNP in LD	WC SNP in LD	Genomic Locus	
1	1:1865298:A:G	rs2803316	1	1865298	9.05E-10	Yes	-0.00766041	0.00125115	9.20E-10	rs2076327	rs2377037	NA	NA	NA	rs10907195	1	
2	1:21880032:A:G	rs7546430	1	21880032	3.20E-08	Yes	-0.0138359	0.00243557	1.34E-08	rs6605083;rs12731820	rs2377037	NA	NA	NA	rs10907195	2	
2	1:27254755:A:C	rs7537581	1	27254755	7.48E-11	Yes	-0.00828681	0.00124947	3.31E-11	NA	NA	NA	NA	NA	NA	2	
3	1:67153900:G:T	rs6577584	1	67153900	2.74E-22	Yes	-0.0124303	0.00131927	4.42E-21	NA	NA	rs11892	NA	rs6577586	rs6577585	3	*(omitted)*
3	1:70614300:C:T	rs7525119	1	70614300	3.36E-08	No	NA	NA	NA	NA	NA	NA	NA	NA	NA	3	
4	1:77283911:G:T	rs1891216	1	77283911	4.92E-09	Yes	-0.00752791	0.00127012	3.09E-09	NA	NA	NA	NA	NA	rs12128526	4	
4	1:77668166:C:G	rs12120907	1	77668166	2.23E-08	No	NA	NA	NA	NA	NA	NA	NA	NA	NA	4	
5	1:85522199:A:G	rs4908761	1	85522199	4.20E-11	Yes	-0.00875015	0.00128603	1.02E-11	NA	NA	rs301802	NA	NA	NA	5	
6	1:16851440:C:T	rs10907231	1	16851440	9.86E-09	Yes	0.0077867	0.00113276	4.49E-09	rs7538833	NA	NA	NA	NA	NA	6	
7	1:22249589:A:G	rs9426785	1	22249589	8.56E-11	Yes	0.00823814	0.00127868	1.17E-10	NA	rs11587362	rs12081298	NA	NA	NA	7	
8	1:23294671:A:G	rs637168	1	23294671	2.56E-09	Yes	0.0100178	0.00170072	3.85E-09	rs7551124;rs2742967	rs10753556;rs2298632	NA	NA	NA	rs506004	8	
(omitted)																	

[Added to the Supplementary Table]

Supplementary Table 16. *Lead SNPs of MetS components, Qsnp, and previous MetS GWAS that are independent from GWAS of this study*

Trait	GenomicLocus	uniqID	SNP	chr	pos	p	nIndSigSNPs	IndSigSNPs	MetS COJO SNP in LD
TG	1	1:1686962:C:T	rs2076327	1	1686962	4.69E-08	1	rs2076327	rs2803316
TG	2	1:2147162:C:T	rs6605083	1	2147162	5.19E-13	3	rs6605083;rs10910028;rs262680	rs7546430
TG	2	1:2222368:A:G	rs12731820	1	2222368	3.18E-08	1	rs12731820	rs7546430
TG	3	1:15259476:C:T	rs12563724	1	15259476	3.59E-08	1	rs12563724	NA
TG	4	1:16504381:C:T	rs7538833	1	16504381	3.72E-15	9	rs7538833;rs11260930;rs6704422;rs34713890;rs61769918;rs28791627;rs7534979;rs12084478;rs4661712	rs10907231
TG	5	1:23785760:C:T	rs7551124	1	23785760	2.89E-17	5	rs7551124;rs1986133;rs2811964;rs2275355;rs10753556	rs637168
TG	5	1:23827720:G:T	rs2742967	1	23827720	3.25E-09	3	rs2742967;rs2275355;rs1986133	rs637168
TG	6	1:26232356:A:C	rs213641	1	26232356	1.18E-11	1	rs213641	NA
TG	7	1:26879792:C:T	rs4970489	1	26879792	2.16E-15	4	rs4970489;rs364977;rs4274112;rs71640328	rs6666121
TG	7	1:26919414:C:T	rs12408288	1	26919414	4.31E-09	3	rs12408288;rs11809021;rs4589135	rs11810321;rs6666121
TG	7	1:27029551:G:T	rs6666121	1	27029551	2.19E-12	4	rs6666121;rs4589135;rs12727590;rs11809021	rs11810321;rs6666121
TG	7	1:27262545:C:T	rs182050989	1	27262545	1.21E-38	7	rs182050989;rs12125238;rs12410656;rs186042737;rs34517168;rs11810321;rs12727590	rs11810321;rs6666121
TG	8	1:32197257:A:G	rs3766823	1	32197257	3.48E-08	1	rs3766823	rs3766823

(omitted)

[Added to the Results, page 9, lines 171–174]

"We further conducted a conditional and joint analysis (COJO)[21] on 1,650 MetS lead SNPs to report statistically significant and independent SNPs, of which 1,307 COJO MetS SNPs were identified (**Supplementary Table 12**)."

[Updated in the Results, page 9, lines 174–180]

"Among them, 26 were non-synonymous, 44 had a RegulomeDB score of 1 (indicating a high likelihood of having a regulatory function), and 19 were located in the active transcription start site with the highest accessibility. Moreover, 414 genes were mapped using three gene-mapping strategies, including positional mapping, expression quantitative trait loci mapping (eQTL) mapping, and chromatin interaction mapping, in the FUMA and MAGMA gene-based analyses (**Supplementary Note**, **Supplementary Tables 13–15**, **Supplementary Figs. 4–5**)."

[Updated in the Results, page 9, lines 181–187]

"We then assessed the independence of the COJO MetS SNPs, with a window size of 500 kb and r^2 threshold of <0.01, from previously reported signals. Among the 1,307 COJO MetS SNPs, 82 (6.3%) were independent of the GWAS signals of the seven MetS components included in the Genomic SEM, 848 (64.9%) were independent of previous MetS GWAS (Lind[10] and van Walree *et al.*[22]), and 159 (12.2%) were previously unreported in the NHGRI-EBI GWAS Catalog[23] using FUMA (**Table 1**, **Supplementary Tables 11 and 16**, **Supplementary Fig. 6**)."

[Updated in the Results, page 9, lines 189–191]

"We identified a substantial number of COJO SNPs (718, 496, and 608) associated with obesity (F1), insulin resistance/hypertension (F2), and dyslipidemia (F3), respectively."

[Updated in the Results, page 10, lines 195–205]

"Among the 1,307 COJO MetS SNPs, 62.1% (*n* COJO SNPs = 811) were independent of Q_{SNP}, and 20.9% (*n* COJO SNPs = 273) were genome-wide significant (GWS) in the heterogeneity test of Q_{SNP}. When we directly compared the direction of the effects of the COJO MetS SNPs with the

corresponding SNPs in the GWAS of MetS components, we observed that 59% of the COJO MetS SNPs (n = 772) had a perfect match in terms of effect direction. Additionally, 37.9% of the COJO MetS SNPs (n = 495) showed consistency in the effect direction with five or six MetS components. The Cohen's kappa (κ) test indicated agreement in the effect direction between MetS and each of the seven MetS components, with κ values ranging from 0.47 to 0.96 (**Supplementary Table 17**). These findings suggest that the identified COJO MetS SNPs exhibit consistent and horizontal pleiotropic effects across MetS components via shared genetic liability (i.e., MetS factor)."

[Updated in the Discussion, page 17, lines 365–368]

"Genomic SEM identified genomic loci associated with MetS, of which 1,307 were significant COJO SNPs even after conditional analysis. Furthermore, 6.3% (n = 82) of the COJO MetS SNPs were independent of the genomic loci of the MetS components, and only 21% (n = 273) were GWS in the Q_{SNP} heterogeneity test."

[Added to the Methods, page 24, lines 534–543]

"The identified lead SNPs were subjected to GCTA-COJO[21] v1.94.1 and determined whether they were conditionally and jointly associated SNPs through a stepwise model selection procedure with default parameters (i.e., P-value $<5 \times 10^{-8}$, window of 10 Mb, and collinearity <0.9). To identify MetS COJO SNPs independent of other GWAS, we compared the lead SNPs with their independent significant SNPs within the 500 kb window of the MetS COJO SNP. A MetS COJO SNP was classified as "independent" if any lead SNPs and their independent significant SNPs with P-value $<5 \times 10^{-8}$ and r^2 >0.01 were not identified. A MetS COJO SNP was classified as "previously reported" if the SNP was mapped from FUMA using NHGRI-EBI GWAS catalog[23] (last updated on 2023.08.02)."

2) Specificity of results: The authors have identified an enrichment of expression effects in brain tissues for the genes harboring the MetS loci. How is that specific to MetS? This type of enrichment is well known for BMI genetic loci and I wonder whether the authors particularly picked up BMI genetics by their multitrait GWAS? Same for the PRS-PheWAS: a plethora of outcomes coming

up would certainly also come up for BMI PRS. As with the first point, I am questioning on what is truly novel identified by this multitrait analysis (in comparison to single trait GWAS's)?

Response: We appreciate the reviewer for this valuable feedback. To address the reviewer's concerns regarding the specificity of our findings for MetS in tissue enrichment and PRS-PheWAS, we have conducted detailed comparisons at the first-order factor levels: F1 (obesity factor), F2 (insulin resistance/hypertension factor), and F3 (dyslipidemia factor).

Upon conducting LDSC-SEG and MAGMA analyses to assess tissue enrichment for each of the three first-order factors, we observed that not only F1 (the obesity factor that included BMI) demonstrated a significant presence of brain tissue enrichment, but F2 (the insulin resistance/hypertension factor) exhibited indications of brain tissue enrichment as well (**Supplementary Tables 27–28**). This finding suggests that the observed brain tissue enrichment in MetS may not be solely attributable to the genetic signals of BMI.

In our PRS-PheWAS analysis, we investigated the association between each of the three first-order factors and various health outcomes. Through this analysis, we observed that the PRS for F2 and F3, which did not include BMI, were also associated with a range of health outcomes (**Supplementary Table 41**). We have identified 350 outcomes associated with PRS in patients with MetS. Among these, 201, 213, and 281 outcomes were significantly associated with the F1, F2, and F3 PRS, respectively. Moreover, of the 201 outcomes significantly associated with both MetS and F1 PRS, 170 and 182 were associated with F2 and F3 PRS, respectively. These findings revealed that most outcomes related to MetS and F1 were associated with F2 and F3, suggesting that BMI, represented within F1, is not the sole factor influencing these associations.

In terms of the novelty of the PRS-PheWAS results from the MetS PRS, we further tested the association between significantly associated outcomes with MetS PRS and F1, F2, and F3 PRS. Among the 350 outcomes that were associated with MetS-PRS, 43 were solely associated with

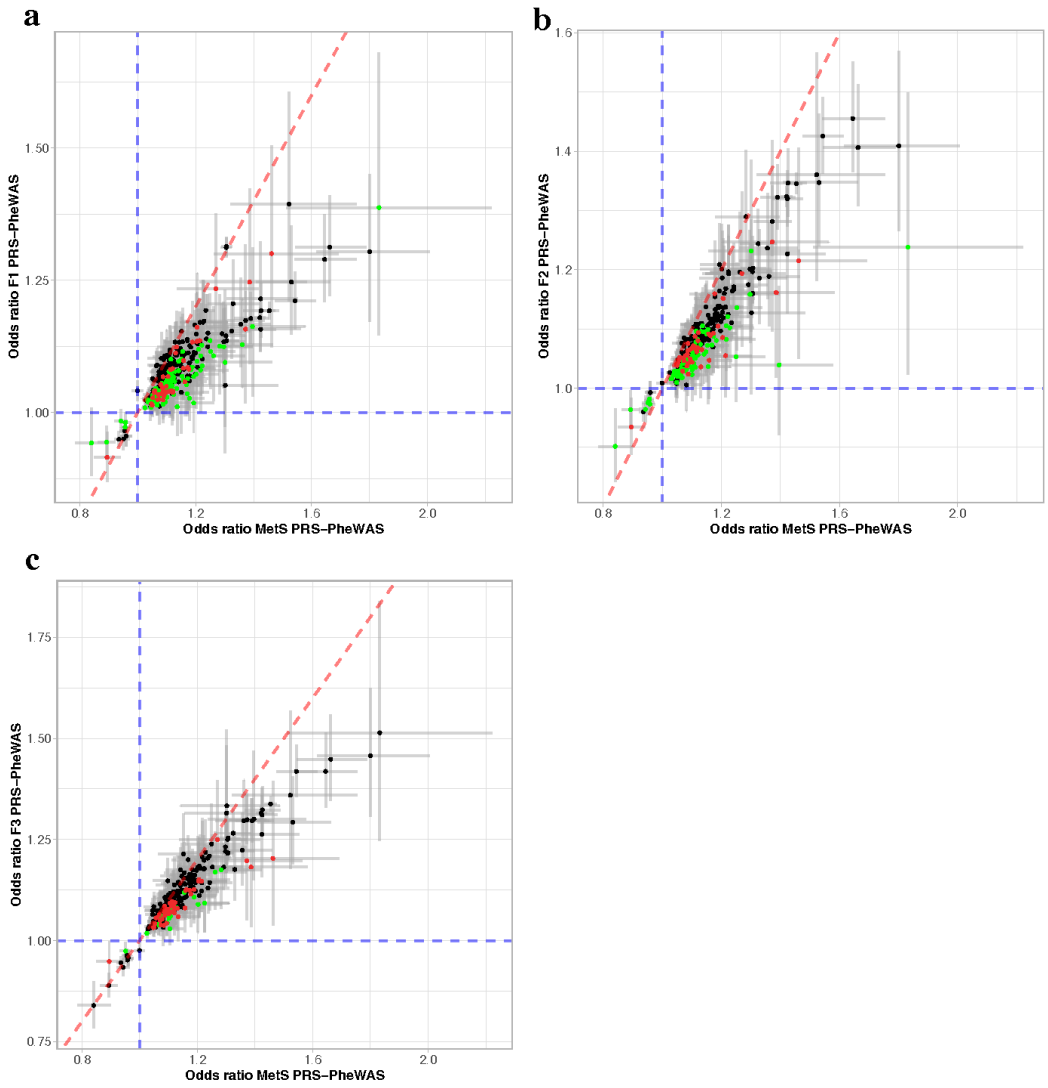
MetS-PRS, including paroxysmal tachycardia (Phecode = 427.1, OR = 1.09) and delirium (Phecode = 290.2, OR = 1.10).

We are thankful for the reviewer's comments, which have helped highlight the novel findings from our research and will enable a better explanation of the impact of MetS on health for our readers.

[Added to the Supplementary Figure]

Supplementary Figure 12. Scatter plot of odds ratio (OR) of PRS-PheWAS results that showed significant association with MetS PRS

a*, Comparison of OR from PRS-PheWAS between MetS PRS and F1 PRS.* ***b****, Comparison of OR from PRS-PheWAS between MetS PRS and F2 PRS.* ***c****, Comparison of OR from PRS-PheWAS between MetS PRS and F3 PRS. In all panels, green dots represent the outcomes that were significantly associated with MetS and the corresponding factor PRS in comparison; red dots represent the outcomes that were significantly associated with only MetS PRS; grey bars are error bars with 95% confidence intervals; the blue dashed line represents OR = 1; and the red dashed line represents the identity line.*


a
Odds ratio F1 PRS–PheWAS
Odds ratio MetS PRS–PheWAS
b
Odds ratio F2 PRS–PheWAS
Odds ratio MetS PRS–PheWAS
c
Odds ratio F3 PRS–PheWAS
Odds ratio MetS PRS–PheWAS

[Added to the Supplementary Table]

Supplementary Table 27. *LDSC-SEG enrichment in multi-tissue gene expression for F1–3 (FDR-correction significant results)*

Factor	Name	Coefficient	Coefficient_std_error	Coefficient_P_value	Bonferroni P-value
F1	A08.186.211.464.405.Hippocampus	1.17E-08	1.97E-09	1.32E-09	2.71E-07
F1	A08.186.211.464.Limbic.System	1.21E-08	2.09E-09	3.66E-09	7.51E-07
F1	A08.186.211.730.885.287.500.Cerebral.Cortex	1.12E-08	2.00E-09	1.15E-08	2.35E-06
F1	A08.186.211.Brain	1.14E-08	2.05E-09	1.63E-08	3.35E-06
F1	Brain_Frontal_Cortex_(BA9)	9.88E-09	1.93E-09	1.43E-07	2.93E-05
F1	A08.186.211.464.710.225.Entorhinal.Cortex	1.20E-08	2.37E-09	2.18E-07	4.47E-05
F1	Brain_Nucleus_accumbens_(basal_ganglia)	8.56E-09	1.90E-09	3.45E-06	0.000707971
F1	Brain_Putamen_(basal_ganglia)	8.16E-09	1.86E-09	5.70E-06	0.001167591
F1	A08.186.211.132.Brain.Stem	9.07E-09	2.08E-09	6.75E-06	0.001384703
F1	A08.186.211.132.810.428.200.Cerebellum	8.63E-09	2.01E-09	8.55E-06	0.001752272
F1	Brain_Caudate_(basal_ganglia)	8.12E-09	1.89E-09	8.86E-06	0.001816765
F1	A08.186.211.730.885.287.500.270.Frontal.Lobe	9.81E-09	2.31E-09	1.05E-05	0.002150155
F1	Brain_Anterior_cingulate_cortex_(BA24)	8.13E-09	1.92E-09	1.12E-05	0.0023029
F1	A08.186.211.865.428.Metencephalon	8.08E-09	1.99E-09	2.58E-05	0.005293623
F1	Brain_Hypothalamus	7.55E-09	1.92E-09	4.10E-05	0.00841122
F1	Brain_Cerebellar_Hemisphere	7.32E-09	1.88E-09	5.01E-05	0.010266094
F1	Brain_Cerebellum	7.52E-09	1.97E-09	6.73E-05	0.013787566
F1	A09.371.729.Retina	9.64E-09	2.60E-09	0.000106628	0.021858662
F1	A08.186.211.730.885.287.500.571.735.Visual.Cortex	7.40E-09	2.13E-09	0.00025821	0.052932959

F1	Brain_Cortex	6.99E-09	2.04E-09	0.000304534	0.062429532
F1	A08.186.211.730.885.287.500.670.Parietal.Lobe	7.09E-09	2.26E-09	0.000853082	0.174881711
F1	A08.186.211.730.885.287.249.Basal.Ganglia	6.41E-09	2.12E-09	0.001236649	0.253513053
F1	Brain_Hippocampus	5.97E-09	2.03E-09	0.001658895	0.34007355
F1	Brain_Amygdala	5.07E-09	1.97E-09	0.005124746	1
F3	Adipose_Visceral_(Omentum)	1.39E-08	4.06E-09	0.0003005	0.061602511
F3	A03.620.Liver	1.62E-08	4.73E-09	0.000303685	0.062255351
F3	Liver	1.54E-08	4.72E-09	0.000546571	0.112047122

Supplementary Table 28. *MAGMA gene-property analysis for average gene expression per tissue types in the GTEx v8 for F1–3 (FDR-correction significant results)*

Factor	Tissue	N genes	Beta	Beta STD	SE	P-value	Bonferroni P-value
F1	Brain Cerebellar Hemisphere	16312	0.078874	0.15723	0.0096002	1.14E-16	6.04E-15
F1	Brain Cerebellum	16312	0.081215	0.16005	0.0099135	1.38E-16	7.34E-15
F1	Brain Frontal Cortex BA9	16312	0.066812	0.12189	0.010873	4.10E-10	2.17E-08
F1	Brain Cortex	16312	0.067854	0.1225	0.011286	9.38E-10	4.97E-08
F1	Brain Anterior cingulate cortex BA24	16312	0.059261	0.10361	0.011437	1.12E-07	5.91E-06
F1	Brain Hypothalamus	16312	0.058327	0.099223	0.012879	2.99E-06	0.000158507
F1	Brain Nucleus accumbens basal ganglia	16312	0.051681	0.088693	0.012052	9.06E-06	0.0004801
F1	Pituitary	16312	0.059025	0.1087	0.014146	1.51E-05	0.000802685
F1	Brain Hippocampus	16312	0.048437	0.081107	0.012668	6.61E-05	0.003501498
F1	Brain Amygdala	16312	0.046279	0.077994	0.012466	0.00010304	0.00546112
F1	Brain Caudate basal ganglia	16312	0.042943	0.073383	0.012491	0.00029406	0.01558518

F1	Brain Putamen basal ganglia	16312	0.036253	0.06105	0.012584	0.0019858	0.1052474
F1	Brain Spinal cord cervical c-1	16312	0.033565	0.059735	0.0135	0.0064608	0.3424224
F2	Brain Cerebellum	16312	0.035196	0.06936	0.0096114	0.00012561	0.00665733
F2	Brain Cerebellar Hemisphere	16312	0.031212	0.062219	0.0093085	0.0004006	0.0212318
F2	Uterus	16312	0.051749	0.10448	0.01663	0.00093169	0.04937957
F3	Uterus	16312	0.0858	0.17323	0.018564	1.92E-06	0.000103567
F3	Adipose Subcutaneous	16312	0.066687	0.13195	0.018319	0.00013664	0.00737856
F3	Ovary	16312	0.052444	0.10541	0.016499	0.00074153	0.04004262
F3	Cervix Endocervix	16312	0.061778	0.12099	0.019752	0.00088279	0.04767066
F3	Fallopian Tube	16312	0.056708	0.10881	0.020174	0.0024727	0.1335258
F3	Adipose Visceral Omentum	16312	0.050863	0.097896	0.019024	0.0037562	0.2028348
F3	Breast Mammary Tissue	16312	0.054727	0.10418	0.021443	0.0053581	0.2893374
F3	Thyroid	16312	0.039844	0.078137	0.016351	0.0074133	0.4003182
F3	Liver	16312	0.025786	0.046076	0.010708	0.0080215	0.433161
F3	Nerve Tibial	16312	0.040022	0.079767	0.017103	0.0096455	0.520857
F3	Cervix Ectocervix	16312	0.048457	0.094008	0.020724	0.0096948	0.5235192

[**Added to the Results, page 12, lines 241–243**]

"Furthermore, both F1 and F2 were enriched in brain tissues, whereas F3 was enriched in adipose and liver tissues (**Supplementary Tables 27–28**)."

[Added to the Results, page 15, lines 319–323]

"Additionally, we conducted PRS-PheWAS for F1, F2, and F3, and 205, 213, and 294 health outcomes exhibited significant associations, respectively (**Supplementary Table 41**). Moreover, 43 health outcomes showed significant associations solely with the MetS PRS, including paroxysmal tachycardia (OR = 1.09) and delirium (OR = 1.10) (**Supplementary Fig. 12**)."

In addition, I have the following minor comments:

3) In line 79, the authors write: "For example, a GWAS conducted by the Global Lipids Genetics Consortium (GLGC), including approximately 1.3 million European individuals, identified 380 and 388 genetic variants associated with high-density lipoprotein cholesterol (HDL-C) and triglycerides (TG), respectively5". However, in the reference paper, there were >1700 distinct signals identified. It would be good if the authors could clarify the numbers.

Response: Thank you for the detailed examination of these numbers. Graham *et al*. (2021) conducted a multi-ancestral meta-analysis of lipid traits from the Global Lipids Genetics Consortium (GLGC). They reported 1,765 distinct index variants across various ancestry groups and lipid traits. In this study, we focused specifically on European populations. Hence, in the main text, we refer to the subset of these signals that were distinctly identified within the European cohort, as reported by Graham *et al*. (2021). In the revised version, to clarify this point, we have specified that the results pertain to European ancestry.

[Rephrased in the Main, page 5, lines 80–83]

"For example, a GWAS conducted by the Global Lipids Genetics Consortium has identified 380 and 388 genetic variants associated with high-density lipoprotein cholesterol (HDL) and triglycerides (TG), respectively, in the European population[5]."

4) Line 116/120 Supplementary Table 2/3 should be exchanged in the xlsx.

Response: Thank you for your thorough review. We have revised the content of **Supplementary Table 2** and **Supplementary Table 3**. We have updated the indices in the **Supplementary Tables**.

5) Line 186: Can the authors explain what type of pleiotropy is exerted via the MetS factor?

Response: Thank you for the feedback. The meaning of "pleiotropy" denoted in the text aligns with the context of horizontal pleiotropy. Observations from the genetic correlation analysis of various GWASs show the presence of pervasive pleiotropy across different phenotypes. In other words, genetic variants have significant effects on different traits. Genomic SEM uses the genetic covariance between traits to construct a common factor. A significant SNP association with this factor was considered to be an SNP exerting a consistent effect across the traits involved with the factor. In the context of a MetS GWAS, an SNP that is significantly associated with the MetS factor is considered to have a consistent, pleiotropic effect on the seven MetS components (Grotzinger *et al*. [2019] and Karlsson Linnér *et al*. [2021]). We have rephrased the sentence to clarify this point.

[Rephrased in the Results, page 10, lines 203–205]
"These findings suggest that the identified COJO MetS SNPs exhibit consistent and horizontal pleiotropic effects across MetS components via shared genetic liability (i.e., MetS factor)."

6) The text states 80 significant genetic correlations among 117 external traits but the respective figure shows 96.

Response: Thank you for the comment. We aimed to highlight the results of genetic correlations between MetS and external traits that have passed the Bonferroni correction, which are 80 genetic

correlation results. However, we have incorrectly written the results that passed the Benjamini-Hochberg false discovery rate (FDR) threshold below 0.05 in the **Results**.

Meanwhile, we added three traits (fasting insulin, HOMA-IR, and non-alcoholic fatty liver disease) in response to the feedback from reviewer 3. Hence, the total number of external traits analyzed was 119, with 82 traits exhibiting significant genetic correlations with MetS. We have updated **Figure 2**, **Results**, and **Supplementary Tables 18–19**.

[Updated Figure 2]

Figure 2. Genetic correlation between MetS and external traits

*The genetic correlation (r_g) between MetS and external traits was estimated using LD score regression. Among the 119 external traits, 82 Bonferroni significant r_g values are illustrated (**Supplementary Table 19** reports all r_g values with 119 traits). Error bars represent 95% confidence intervals (CIs) for r_g, calculated as 1.96 times the standard error. The black dotted line indicates a r_g of 0.*


Verbal numerical reasoning
Cognitive performance
Fluid intelligence score
Intelligence
Fluid intelligence 7: synonym
Fluid intelligence 8: chained arithmetic
Fluid intelligence 3: word interpolation
Fluid intelligence 5: family relationship calculation
Fluid intelligence 6: conditional arithmetic
Education years
HOMA−IR
Body fat percentage
Fasting insulin
C−reactive protein
Urate
White blood cell (leukocyte) count
Haemoglobin concentration
Total serum protein
Vitamin D
Testosterone
Time spent watching television
Duration of walks
Coffee intake
Never alcohol drinking
Sleep duration
Ever taken cannabis
Never smoking
Leisure social activities: sports club or gym
Leisure social activities: religious group
Leisure social activities: adult education class
Ever highly irritable argumentative for 2 days
Ever panic attacks
Ever manic hyper for 2 days
Ever unenthusiastic disinterested for a whole week
Anxiety disorder
Ever depressed for a whole week
Financial situation satisfaction
Health satisfaction
Small vessel stroke
Any stroke
Any ischemic stroke
Parkinson's disease
Migraine
Fed−up
Lonely
Mood instability
Miserable
Extraversion
Irritable
Hurt
Worry
Worry after embarrasing
Nervous
Non−alcoholic fatty liver disease
Renal failure
Sleep apnea
Sleep disorders
Urinary tract infection
Angina pectoris
Cerebrovascular disease
Ischemic heart disease
Myocardial infarction
Coronary atherosclerosis
Iron deficiency anemias
Rheumatoid arthritis and other inflammatory polyarthropathies
Gastritis and duodenitis
Functional digestive disorders
Osteoarthrosis
Esophagitis, GERD and related diseases
Urinary incontinence
Asthma
Cardiac dysrhythmias
COVID−19
Constipation
Hypothyroidism
Atrial fibrillation and flutter
Attention deficit hyperactivity disorder
Major depression
Schizophrenia
Biopolar disorder
Obesessive−compulsive disorder
Anorexia nervosa
−0.5
0.0
0.5
1.0
r_g
Category
Cognitive function
Laboratory and physical findings
Life style
Mental health
Neurological diseases
Neuroticism
Physical health
Psychiatric disorders

[Modified the Supplementary Table]

Supplementary Table 18. *Summary of the external traits used in genetic correlation analysis*

Category	Trait	Reference	Trait type	*N* case	*N* control	*N* total
Laboratory and physical findings	Fasting insulin	[28]	continuous	-	-	105,056
Laboratory and physical findings	HOMA-IR	[29]	continuous	-	-	51,750
Physical health	Non-alcoholic fatty liver disease	[30]	binary	8,434	770,180	778,614
ROI101	101 GWAS summary statistics of ROI	[27]	continuous	-	-	19,629
DTI110	110 GWAS summary statistics of DTI	[26]	continuous	-	-	17,706
Neurological diseases	Alzheimer's disease	[1]	binary	71,880	383,378	455,258
(omitted)						

Supplementary Table 19. *Genetic correlation between MetS and external traits*

Category	Trait	r_g	SE	Z-score	*P*	95% CI lower bound	95% CI upper bound	Bonferroni *P*	Bonferroni significant
Physical health	Non-alcoholic fatty liver disease (NAFLD)	0.8441	0.1492	5.659	1.52E-08	0.551668	1.136532	1.82E-06	Yes
Physical health	Renal failure	0.825	0.2126	3.8805	1.00E-04	0.408304	1.241696	0.012	Yes
Laboratory and physical findings	HOMA-IR	0.816	0.0641	12.7262	4.23E-37	0.690364	0.941636	5.08E-35	Yes
Laboratory and physical findings	Body fat percentage	0.7836	0.0127	61.9461	0	0.758708	0.808492	0	Yes
Laboratory and physical findings	Fasting insulin	0.7344	0.0444	16.5492	1.62E-61	0.647376	0.821424	1.94E-59	Yes
(omitted)									

[Updated the Results, pages 10–11, lines 208–217]

"We estimated the genetic correlation (r_g) between MetS and 119 external traits categorized into eight groups using LDSC regression (**Fig. 2**, **Supplementary Table 18**). Among the 119 examined traits, 82 exhibited significant associations after Bonferroni correction. Given that the link between MetS and CVD is widely acknowleged[24], our findings reveal a moderate genetic overlap with angina pectoris (r_g = 0.51, 95% CI = 0.46–0.56) and ischemic heart disease (r_g = 0.50, 95% CI = 0.46–0.55).

Besides, we observed a substantial positive genetic correlation between MetS and non-alcoholic fatty liver disease (NAFLD) (r_g = 0.84, 95% CI = 0.55–1.14) and a negative correlation with health satisfaction (r_g = −0.53, 95% CI = −0.57–−0.49) (**Supplementary Table 19**)."

7) The authors conducted robust MR analyses. Yet 22 of their 28 significant causal effects shown in Table 2 show significant heterogeneity. Even MR-PRESSO can be problematic for instrument selection in this case, and I suggest using a constrained maximum-likelihood method to select instruments (PMID 34214446).

Response: We thank the reviewer for suggesting a new method for robust MR analysis. We agree that MR-PRESSO may result in an inflated type-I error when there is significant heterogeneity. By contrast, the proposed method (i.e., constrained maximum likelihood and model-averaging-based MR [cML-MA]) can control type I errors because it is robust against invalid IVs with correlated and uncorrelated pleiotropic effects. Therefore, we performed MR using cML-MA for all 29 significant results, as shown in **Table 3**.

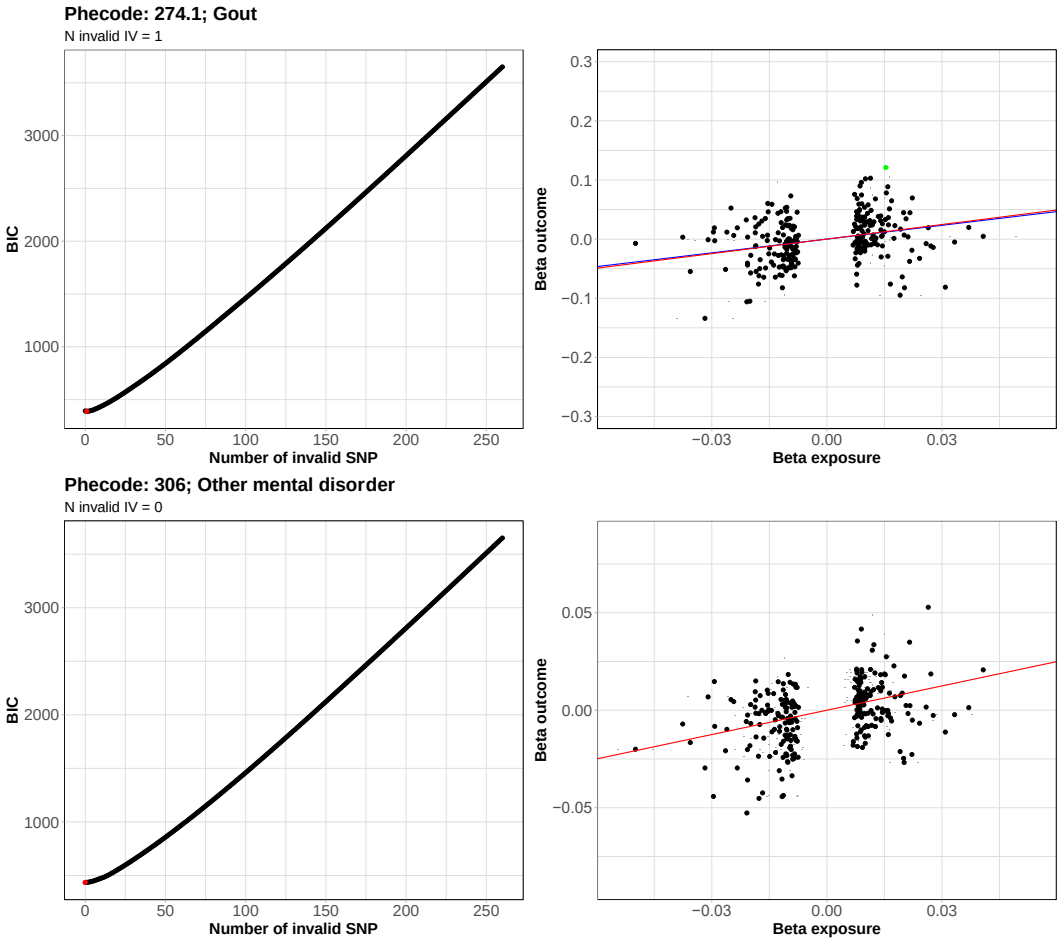
Using cML-MA, we identified 0–7 instrumental variables (IVs) to be removed depending on the outcome of interest. The significance and direction of the causal associations were identical to the results of our TSMR analysis. Peripheral angiopathy in diseases classified elsewhere showed a higher OR compared to previous results, which could be due to a small number of cases (N = 385) compared to other outcomes whose sample size ranged from 1,333 to 38,715. We have added the results from cML-MA to **Supplementary Table 44**. We have added BIC and scatter plot pairs for

29 health outcomes and compared the TSMR and cML-MA MR figures in **Supplementary Figure 13** and **Supplementary Figure 14**, respectively. The cML-MA method is described in the **Supplementary Note**.

[Added to the Supplementary Figure]

Supplementary Figure 13. BIC and scatter plot pairs for 29 health outcomes were tested for their causal association with MetS using constrained maximum likelihood and model averaging-based Mendelian randomization (cML-MA MR)

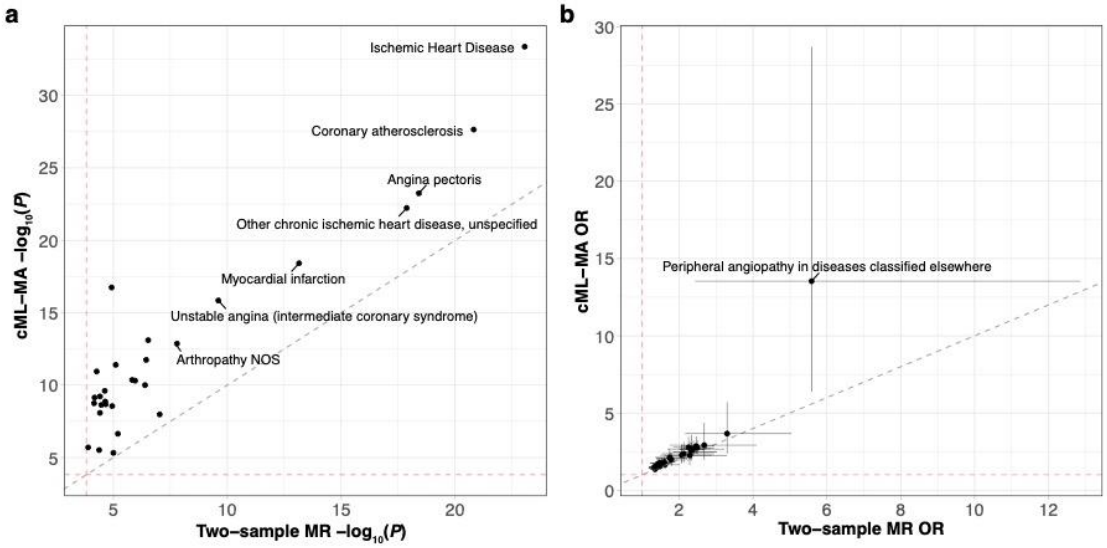
The left panel shows the number of invalid instrumental variables (IVs) on the x-axis and BIC values on the y-axis. The right panel shows the beta from exposure (i.e., MetS) and outcome (i.e., health outcome) on the x-axis and y-axis, respectively, with their respective error bars. The green dots represent invalid IVs detected using cML-MA, the black dashed lines represent beta = 0, the blue line represents the causal estimate from the inverse variance weighted (IVW) method, and the red line represents the causal estimate from either cML-MA-BIC or cML-MA-BIC-DP MR.



(omitted)

Supplementary Figure 14. Scatter plot of comparison between two-sample Mendelian randomization (TSMR) and constrained maximum likelihood and model averaging-based Mendelian randomization (cML-MA MR) results for 29 health outcomes

***a**, Scatter plot of -log$_{10}$(P) comparison with the x-axis from TSMR and the y-axis from cML-MA MR. The red dashed line represents the Bonferroni correction threshold of 0.05/29, and the gray dashed line represents the identity line. **b**, Scatter plot of odds ratio (OR) comparison with x-axis from TSMR and y-axis from cML-MA MR. The red dashed line represents OR = 1, and the gray dashed line represents the identity line.*


a
Ischemic Heart Disease
Coronary atherosclerosis
Angina pectoris
Other chronic ischemic heart disease, unspecified
Myocardial infarction
Unstable angina (intermediate coronary syndrome)
Arthropathy NOS
cML–MA –log10(P)
Two–sample MR –log10(P)
b
Peripheral angiopathy in diseases classified elsewhere
cML–MA OR
Two–sample MR OR

[Added to the Supplementary Table]

Supplementary Table 44. *cML-MA MR result*

Phecode	Outcome	approach_selection	BIC_invalid_N	BIC_invalid_SNP	MA_BIC_theta	MA_BIC_se	MA_BIC_p	MA_BIC_DP_theta	MA_BIC_DP_se	MA_BIC_DP_p	GOF1_p	GOF2_p
274.1	Gout	MA_BIC	1	rs12476704	0.820	0.137	2.32E-09	0.826	0.164	4.57E-07	3.04.E-01	2.94.E-01
306	Other mental disorder	MA_BIC	0		0.416	0.049	1.81E-17	0.394	0.058	1.06E-11	5.37.E-01	5.76.E-01
318	Tobacco use disorder	MA_BIC_DP	4	rs10853981;rs13396091;rs205262;rs2820311	0.368	0.058	2.41E-10	0.368	0.077	1.97E-06	1.20.E-02	5.82.E-03
411	Ischemic Heart Disease	MA_BIC_DP	5	rs1558902;rs2575876;rs4599845;rs6857;rs9987289	0.979	0.052	1.13E-77	0.967	0.079	4.29E-34	1.02.E-06	3.97.E-07
411.1	Unstable angina (intermediate coronary syndrome)	MA_BIC	2	rs4755720;rs6857	0.910	0.110	1.43E-16	0.956	0.132	5.14E-13	6.08.E-02	6.14.E-02
411.2	Myocardial infarction	MA_BIC_DP	3	rs1558902;rs17821274;rs6857	1.024	0.091	1.12E-29	1.021	0.114	3.83E-19	2.73.E-03	5.97.E-03
411.3	Angina pectoris	MA_BIC_DP	2	rs1558902;rs6857	1.026	0.075	6.03E-43	1.052	0.104	5.67E-24	3.44.E-04	1.11.E-03
411.4	Coronary atherosclerosis	MA_BIC_DP	5	rs1558902;rs2575876;rs4599845;rs6857;rs9987289	1.037	0.063	5.94E-61	1.009	0.091	2.28E-28	1.58.E-07	3.02.E-06
411.8	Other chronic ischemic heart disease, unspecified	MA_BIC_DP	1	rs6857	0.959	0.072	7.76E-41	0.989	0.100	5.88E-23	8.12.E-03	4.42.E-03
414	Other forms of chronic heart disease	MA_BIC	0		0.813	0.178	4.72E-06	0.825	0.191	1.63E-05	7.15.E-01	7.43.E-01

(omitted)

[Added to the Supplementary Note, pages 21–22, lines 455–476]

"11. Causal effect estimation using constrained maximum likelihood and model averaging-based (cML-MA) Mendelian randomization (MR)

The cML-MA[54] method employs maximum likelihood and model averaging to select valid instrumental variables (IVs), addressing both correlated and uncorrelated pleiotropic effects. Unlike MR-Egger, it does not assume Instrument Strength Independent of Direct Effect (InSIDE) and shows improved control over type I errors. Utilizing the Bayesian information criterion (BIC), it assigns weights and performs model averaging to accommodate model selection uncertainties. Data perturbation … (*omitted*)**"**

[Added to the Results, page 16, lines 347–350]

"We observed robustness in the identified causal associations through constrained maximum likelihood and model averaging-based MR (cML-MA)[43], and the bi-directional TSMR suggested insufficient evidence for the potential reverse causation (**Supplementary Tables 44–46**, **Supplementary Figs. 13–14**)."

8) Their PRS explains ~20% of MetS in UKB but heritability was estimated at only ~11%. I know the outcomes are not completely comparable but some discussion on what can be expected from ever larger MetS GWAS would be helpful.

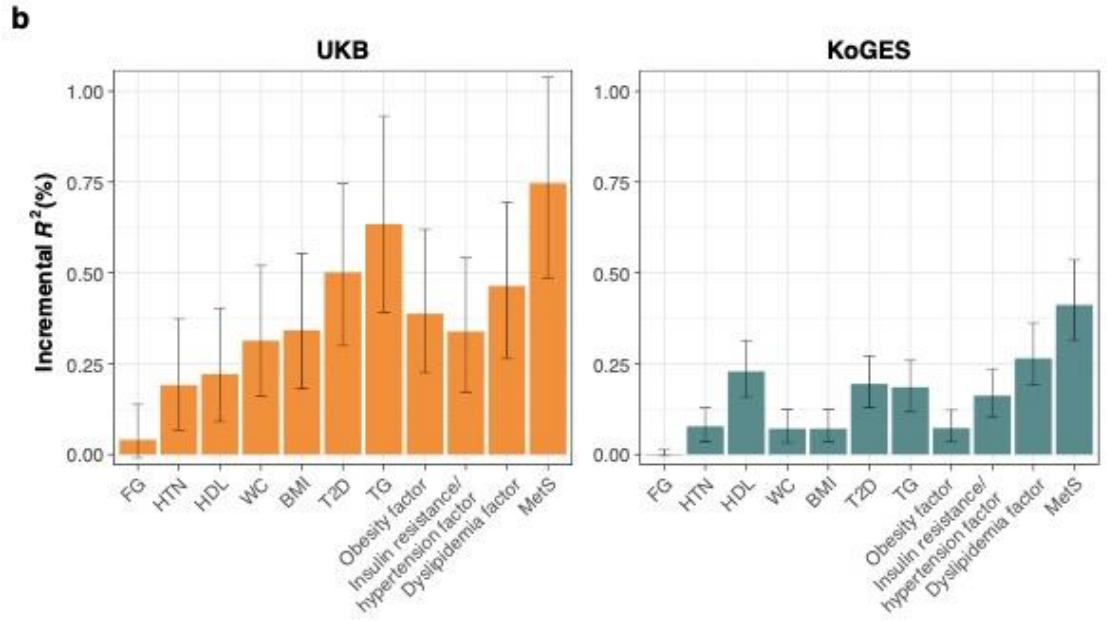
Response: We appreciate the reviewers' comments and critical insights. We acknowledge that our initial methodology, which used the relative increase in the incremental R^2, may have led to misunderstandings. In this revised version, based on previous studies (Turley *et al*. [2018], Lee *et al*. [2018], and Karlsson Linnér *et al*. [2021]), we have calculated incremental R^2 by determining the difference in R^2 values between the baseline and PRS models, rather than assessing the relative increase with respect to the baseline. Using this revised definition, we observed that MetS PRS accounts for a 0.75% increase in the variance of MetS within the UKB cohort, which is lower than the 11% explained by SNP heritability. However, these two values are not directly comparable because the PRS analysis excluded GWAS from the UKB, whereas SNP heritability was estimated

based on the GWAS, including the UKB. Nonetheless, this discrepancy highlights the importance of conducting further research on MetS, ideally with a larger sample size, to enhance our understanding of the unexplained genetic architecture of MetS.

[Modified Figure 4b]

Figure 4. MetS polygenic risk score analysis in European and East Asian populations

***b**, Bar plot illustrating the incremental proportion of variance explained (ΔR^2) by the polygenic risk score of seven MetS components and four latent factors for predicting MetS in UKB and KoGES as target cohorts. The error bars indicate 95% CIs for ΔR^2, and they were computed using the percentile method of bootstrapping with 1,000 iterations.*



[Added to the Discussion, page 19, lines 411–418]

"The MetS PRS demonstrated superior predictive power for dichotomized MetS in both cohorts compared with the PRSs of its components, which is consistent with MetS exhibiting the highest PAT. In contrast, the FG accounted for the least MetS variance in both cohorts. This may be attributed to the fact that the performance of PRS depends on the GWAS sample size[59], and the sample size of the FG GWAS was comparatively smaller than that of the other components. These

findings highlight a promising scope for wider application of the MetS PRS across different populations, yet they stress the need for GWAS with larger sample sizes."

Reviewer #3:

Remarks to the Author:

Enjoyed reading this paper on the genetic architecture of the metabolic syndrome (MetS) in European populations. Found 6 "genes" associated. PRS translatable to East Asians.

A major limitation is the focus on European populations especially since most data suggests other populations are higher risk (East and South Asian). This is acknowledged by the authors.

1. The title is a little misleading. The term Metabolic Syndrome has been used by various clinical organizations to define a clinical entity with the goal of determining who may be at greater risk of ASCVD, T2D and other adverse outcomes. Some of these organizations have published their own definitions, which can differ slightly and undergo revisions periodically (some of these are below). In this case, the title needs to say something like the "shared genetic architecture of components of the metabolic syndrome"

• Grundy, S. M., Brewer, H. B. Jr., Cleeman, J. I., Smith, S. C. Jr. & Lenfant, C. Definition of metabolic syndrome: Report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition. Circulation. 109(3), 433–438 (2004).

• Einhorn, D. American College of Endocrinology position statement on the insulin resistance syndrome. Endocr. Pract. 9, 5–21 (2003).

Eckel, R. H., Grundy, S. M. & Zimmet, P. Z. The metabolic syndrome. Lancet. 365(9468), 1415–1428 (2005).

Lind did a GWAS for MetS based on criteria as above in the UKBB and was cited.

However, in this case, the authors have not performed a GWAS for MetS but rather looked at many large GWAS for MetS broadly related traits. This was a reasonable choice, but it really needs to be clear how the authors defined MetS.

Response: Thank you for your thoughtful comment and suggestion. The definition of metabolic syndrome has been periodically updated, and the current title of our study may mislead readers. We have updated the title as follows to reflect the reviewer's comment: "Multivariate genomic analysis of 5 million people elucidates the genetic architecture of shared components of the metabolic syndrome".

We did not provide a detailed explanation for the definition of MetS or the selection of specific traits/diseases as components of MetS. These points have been thoroughly addressed in the **Supplementary Note**.

[Revised Title]

"Multivariate genomic analysis of 5 million people elucidates the genetic architecture of shared components of the metabolic syndrome"

[Added to the Supplementary Note, pages 4–6, lines 69–116]

"1. Summary of the clinical definitions of metabolic syndrome

Metabolic syndrome (MetS) is a collection of risk factors that increase the risk of cardiovascular disease and type 2 diabetes (T2D). Despite its seemingly straightforward definition, it is still a challenge to diagnose MetS clinically[1-3].

The initial definition of MetS was established in 1998 by … (*omitted*)"

2. It seems the motivation of this paper is a search for the genes underlying the overlapping part of the Venn diagram for these traits. This kind of clustering work has been done previously: Lotta, Dimas, Udler, Gloudemans, O'Rahilly etc. None of this prior work has been cited or used to place the current paper in context.

Response: Thank you for pointing this out. We reviewed the articles mentioned by the reviewer and referenced them accordingly.

[Added in the Main, pages 5–6, lines 85–101]

"The co-occurrence of unhealthy metabolic traits has prompted ongoing endeavors to unveil their common genetic underpinnings. For instance, five categories of T2D mechanistic pathways were identified using T2D-associated variants through clustering analyses, and the genes and traits associated with each cluster were investigated[6,7]. Similarly, colocalization analyses were conducted between cardiometabolic traits and quantitative trait loci (QTL) to identify shared genes across various traits[8]. However, these clustering analyses were constrained to T2D-associated variants, and the mere overlap between colocalized genes may lack substantial evidence of a shared genetic basis across all components of MetS.

Most previous MetS GWAS have focused on a binary definition of MetS. Kraja *et al*.[9] expanded on this by conducting a GWAS for MetS and pairwise combinations of its components and identified 29 common variants; however, these studies lacked robust evidence for a consistent association across all MetS components. Lind[10] examined 291,107 individuals from the UK Biobank (UKB) and identified 93 independent loci associated with MetS. Despite the different definitions of MetS, both studies categorized MetS based on the number of MetS criteria met. Such an approach may potentially introduce variability due to the different combinations of criteria met, thus limiting the representativeness of MetS and leading to an incomplete understanding of the genetic architecture of MetS[11]."

[Added in the Discussion, page 17, lines 376–377]

"MetS has strong connections with brain functions[45], and its association with various psychiatric disorders has been consistently suggested[46,47]."

3. Also, if the authors are using a loose way of defining MetS, why not be even more broad? The underlying common feature for MetS is thought to largely be related to insulin resistance. Indeed, Reaven originally used the Insulin Resistance Syndrome to describe to concepts that underlie the

idea of the MetS. While no surrogates capture insulin resistance perfectly, what is the correlation and overlap with GWAS signals for fasting insulin (although this was not measured in the UKBB has been assessed many other large studies)? What about for NAFLD?

Response: Thank you for your comment. Insulin resistance has been identified as a key factor in MetS, as evidenced by definitions proposed by the WHO (1998) and AACE (2003). However, other organizations, such as ATP III (2001), IDF (2005), and AHA/NHLBI (2005), have not made insulin resistance a compulsory criterion, instead placing a greater emphasis on central obesity or a combination of various conditions. We have summarized the definitions of MetS from various organizations in **Supplementary Note**. This variation in the criteria for defining MetS complicates the selection of a definitive set of traits or diseases that accurately represent MetS.

Nonetheless, this diversity in definitions provides an opportunity to explore additional traits or diseases that may share genetic links with MetS under a more flexible definition. However, in this study, we focused on the most commonly discussed components, thereby excluding other insulin resistance traits and NAFLD. Moreover, well-powered and well-phenotyped GWAS for these traits and disorders remain limited. FG GWAS included in our study, which had the smallest sample size (n = 151,188), indicates that traits associated with insulin resistance, such as fasting insulin (FI, n = 105,056; Lagou *et al*. [2021]) and HOMA-IR (n = 51,750; Manning *et al*. [2012]), were relatively underpowered compared with other metabolic traits (n = 385,932–1,253,277). Similarly, the largest GWAS for NAFLD (Ghodsian *et al*. [2021]) features a relatively small sample size (n_{case} = 8,434, n_{total} = 778,614), as the definitive method for diagnosing NAFLD requires a liver biopsy, leading to a scarcity of properly phenotyped samples. Again, we emphasize that, within the scope of this study, we concentrated on the MetS components commonly recognized by various organizations. Including a broader range of traits and diseases could be a potential direction for future MetS GWAS studies, as described in the **Discussion**.

In response to the reviewer's advice, we have investigated the genetic correlation among MetS, insulin resistance indicators (specifically FI and HOMA-IR), and NAFLD. We observed a

significant genetic correlation between these conditions and MetS, with FI (r_g = 0.73, *s.e.* = 0.04), HOMA-IR (r_g = 0.82, *s.e.* = 0.06), and NAFLD (r_g = 0.84, *s.e.* = 0.15). These results, together with the various MetS definitions, underscore the importance of adopting a more inclusive definition of MetS in future research. We have added these three traits to the genetic correlation analysis, comprising 119 analyzed traits and 82 traits that exhibited a significant genetic correlation with MetS after Bonferroni correction. Consequently, we have updated **Figure 2, Results**, and **Supplementary Tables 18–19**.

[Updated Figure 2]

Figure 2. Genetic correlation between MetS and external traits

*The genetic correlation (r_g) between MetS and external traits was estimated using LD score regression. Among the 119 external traits, 82 Bonferroni significant r_g values are illustrated (**Supplementary Table 19** reports all r_g values with 119 traits). Error bars represent 95% confidence intervals (CIs) for r_g, calculated as 1.96 times the standard error. The black dotted line indicates a r_g of 0.*


Verbal numerical reasoning
Cognitive performance
Fluid intelligence score
Intelligence
Fluid intelligence 7: synonym
Fluid intelligence 8: chained arithmetic
Fluid intelligence 3: word interpolation
Fluid intelligence 5: family relationship calculation
Fluid intelligence 6: conditional arithmetic
Education years
HOMA–IR
Body fat percentage
Fasting insulin
C–reactive protein
Urate
White blood cell (leukocyte) count
Haemoglobin concentration
Total serum protein
Vitamin D
Testosterone
Time spent watching television
Duration of walks
Coffee intake
Never alcohol drinking
Sleep duration
Ever taken cannabis
Never smoking
Leisure social activities: sports club or gym
Leisure social activities: religious group
Leisure social activities: adult education class
Ever highly irritable argumentative for 2 days
Ever panic attacks
Ever manic hyper for 2 days
Ever unenthusiastic disinterested for a whole week
Anxiety disorder
Ever depressed for a whole week
Financial situation satisfaction
Health satisfaction
Small vessel stroke
Any stroke
Any ischemic stroke
Parkinson's disease
Migraine
Fed–up
Lonely
Mood instability
Miserable
Extraversion
Irritable
Hurt
Worry
Worry after embarrasing
Nervous
Non–alcoholic fatty liver disease
Renal failure
Sleep apnea
Sleep disorders
Urinary tract infection
Angina pectoris
Cerebrovascular disease
Ischemic heart disease
Myocardial infarction
Coronary atherosclerosis
Iron deficiency anemias
Rheumatoid arthritis and other inflammatory polyarthropathies
Gastritis and duodenitis
Functional digestive disorders
Osteoarthrosis
Esophagitis, GERD and related diseases
Urinary incontinence
Asthma
Cardiac dysrhythmias
COVID–19
Constipation
Hypothyroidism
Atrial fibrillation and flutter
Attention deficit hyperactivity disorder
Major depression
Schizophrenia
Biopolar disorder
Obsessive–compulsive disorder
Anorexia nervosa
−0.5
0.0
0.5
1.0
r_g
Category
Cognitive function
Laboratory and physical findings
Life style
Mental health
Neurological diseases
Neuroticism
Physical health
Psychiatric disorders

[Modified the Supplementary Table]

Supplementary Table 18. *Summary of the external traits used in genetic correlation analysis*

Category	Trait	Reference	Trait type	*N* case	*N* control	*N* total
Laboratory and physical findings	Fasting insulin	[28]	continuous	-	-	105,056
Laboratory and physical findings	HOMA-IR	[29]	continuous	-	-	51,750
Physical health	Non-alcoholic fatty liver disease	[30]	binary	8,434	770,180	778,614
ROI101	101 GWAS summary statistics of ROI	[27]	continuous	-	-	19,629
DTI110	110 GWAS summary statistics of DTI	[26]	continuous	-	-	17,706
Neurological diseases	Alzheimer's disease	[1]	binary	71,880	383,378	455,258
(omitted)						

Supplementary Table 19. *Genetic correlation between MetS and external traits*

Category	Trait	r_g	SE	Z-score	*P*	95% CI lower bound	95% CI upper bound	Bonferroni *P*	Bonferroni significant
Physical health	Non-alcoholic fatty liver disease (NAFLD)	0.8441	0.1492	5.659	1.52E-08	0.551668	1.136532	1.82E-06	Yes
Physical health	Renal failure	0.825	0.2126	3.8805	1.00E-04	0.408304	1.241696	0.012	Yes
Laboratory and physical findings	HOMA-IR	0.816	0.0641	12.7262	4.23E-37	0.690364	0.941636	5.08E-35	Yes
Laboratory and physical findings	Body fat percentage	0.7836	0.0127	61.9461	0	0.758708	0.808492	0	Yes
Laboratory and physical findings	Fasting insulin	0.7344	0.0444	16.5492	1.62E-61	0.647376	0.821424	1.94E-59	Yes
(omitted)									

[Updated the Results, pages 10–11, lines 208–217]

"We estimated the genetic correlation (r_g) between MetS and 119 external traits categorized into eight groups using LDSC regression (**Fig. 2**, **Supplementary Table 18**). Among the 119 examined traits, 82 exhibited significant associations after Bonferroni correction. Given that the link between MetS and CVD is widely acknowleged[24], our findings reveal a moderate genetic overlap with angina pectoris (r_g = 0.51, 95% CI = 0.46–0.56) and ischemic heart disease (r_g = 0.50, 95% CI = 0.46–0.55).

Besides, we observed a substantial positive genetic correlation between MetS and non-alcoholic fatty liver disease (NAFLD) (r_g = 0.84, 95% CI = 0.55–1.14) and a negative correlation with health satisfaction (r_g = −0.53, 95% CI = −0.57–−0.49) (**Supplementary Table 19**)."

[Added in the Discussion, page 20, lines 436–438]

"Furthermore, considering that the definition of MetS undergoes periodic updates, encompassing a broader spectrum of metabolic traits or diseases, such as insulin resistance and NAFLD, will provide more comprehensive insights into the genetic underpinnings of MetS."

4. While prioritizing MetS genes based on brain QTL data is reasonable especially for phenotypes related to BMI, the decision to use blood QTL data makes less sense. Why not use tissues known to be more strongly related to the phenotypes of the components of MetS (fat, liver, vascular tissue, muscle)?

Response: Thank you for the comment. As you mentioned, conducting gene prioritization analysis using the eQTL of brain tissues is well established, given that our tissue enrichment analysis provided indications of MetS genetic signals being enriched in brain tissues.

There were three reasons for conducting gene prioritization analysis using blood tissue. First, the purpose of gene prioritization using SMR is to identify genes associated with MetS that may act as potential therapeutic targets for MetS. However, it might be challenging to deliver drugs to targeted tissues such as the brain, whereas it is more feasible and efficient to target them if we can

identify the genes that are causally involved in the etiology of the blood tissue as a medium. Secondly, the identified genes may serve as potential drug targets and biomarkers. Moreover, it is less invasive and more practical to perform biomarker measurements using blood tests. Third, the eQTL data for blood tissue from the eQTLGen cohort comprised a large number of samples (*N* = 31,684). The sample size is an essential component that determines the statistical power for identifying significant results from an exploratory perspective. This can be observed from Extended Data Figure 1 of Võsa *et al*. [2021], which compares the replicability of cis-eQTL from eQTLGen in GTEx. Although 94.9% of cis-eQTLs from eQTLGen showed a concordant allelic direction in GTEx, only 14.8% cis-eQTLs were replicated in GTEx because of the low statistical power of the small sample size (max *N* GTEx = 620). These points highlight the importance of conducting gene prioritization analyses in blood tissue (van Rheenen *et al*. [2021], Storm *et al*. [2021], Wang *et al*. [2022]). We have addressed these points in the **Discussion**.

In response to the reviewer's advice, we have conducted a gene prioritization analysis on five additional tissues, namely subcutaneous adipose, skeletal muscle, liver, aorta, and coronary artery, in addition to the brain and whole blood tissue. Owing to the limited public availability of eQTL data for such tissues, we have utilized GTEx eQTL for the aforementioned tissues as a discovery set. We used independent eQTL data available for replication in the adipose subcutaneous and skeletal muscles from METSIM and FUSION, respectively. For the liver, aorta, and coronary artery tissues, we have noted genes that were identified from the discovery set (i.e., GTEx). A summary of the gene prioritization results is provided below.

We have identified 14, 16, 5, 14, and 5 genes associated with MetS in the subcutaneous adipose, skeletal muscle, liver, aorta, and coronary artery tissues, respectively. Replication of genes identified in subcutaneous adipose and skeletal muscle tissues identified five and three genes, respectively. We have obtained five genes (*RBM6*, *BCL7B*, *MLXIPL*, *MYO1F*, and *AMHR2*) in addition to six genes identified in the brain and whole blood tissues (*STRA13*, *FEZ2*, *RFT1*, *MED23*, *SP1*, and *HM13*). To discuss the results for the four tissues, we have replaced **Figure 3c**

with a circular Manhattan plot. Thank you for your advice, which has helped expand and advance our research.

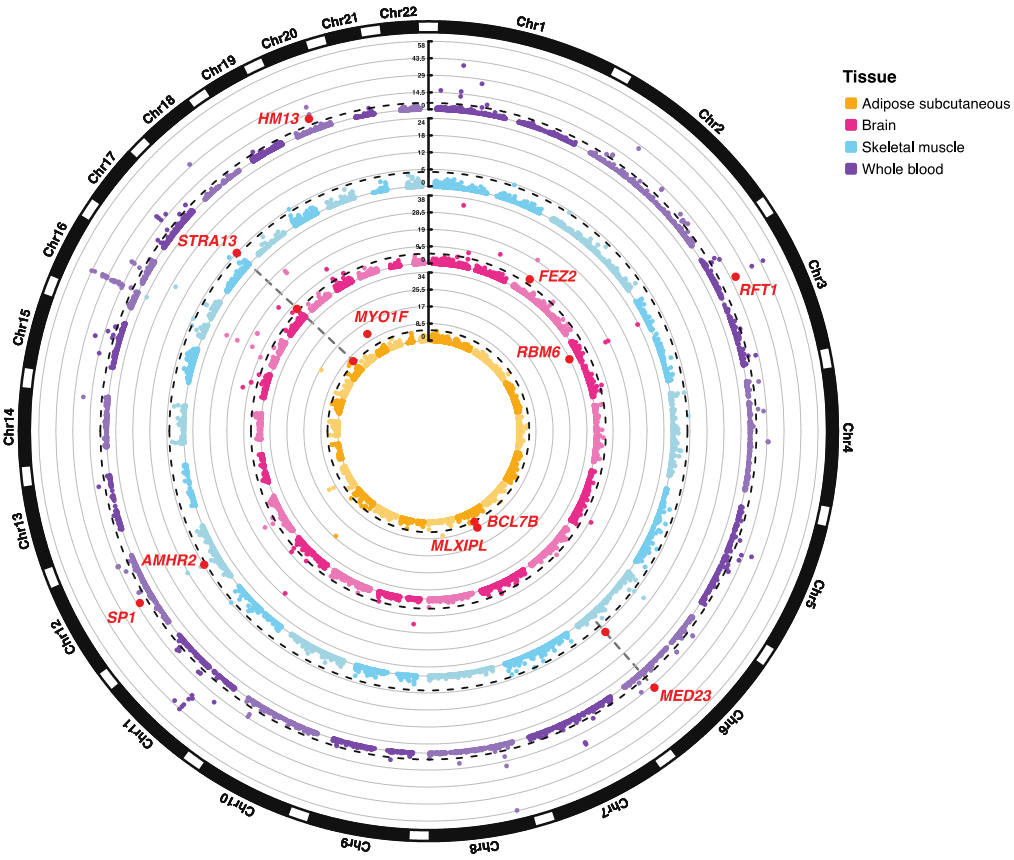
Summarized table of SMR analysis findings

Tissue	Discovery cohort	*N* genes	*N* genes significant	Replication cohort	N genes present	N genes replicated	Genes replicated	Unique genes replicated
Brain	BrainMeta v2	15,131	43	GTEx (cortex)	2	2	*STRA13, FEZ2*	*STRA13, FEZ2*
				GTEx (frontal cortex)	2	2		
				GTEx (anterior cingulate cortex)	2	2		
Whole blood	eQTLGen	15,155	136	GTEx	95	4	*RFT1, MED23, SP1, HM13*	*RFT1, MED23, SP1, HM13*
Adipose subcutaneous	GTEx	7,207	14	METSIM	12	5	*RBM6, BCL7B, MLXIPL, STRA13, MYO1F*	*RBM6, BCL7B, MLXIPL, STRA13, MYO1F*
Skeletal muscle	GTEx	6,245	16	FUSION	13	3	*MED23, AMHR2, STRA13*	*MED23, AMHR2, STRA13*
Liver	GTEx	2,157	5	-	-	-	-	-
Artery aorta	GTEx	5,692	14	-	-	-	-	-
Artery coronary	GTEx	2,378	5	-	-	-	-	-

[Modified Figure 3]

Figure 3c*. Circos plot[83] for gene prioritization using SMR in the subcutaneous adipose tissue, brain, skeletal muscle, and whole blood (inner to outer circles). The y-axis represents the -log$_{10}$(P) of the gene association. Red dots with annotations show replicated genes. The dashed black line represents the Bonferroni significance threshold for the corresponding tissue.*

c


Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17
Chr18
Chr19
Chr20
Chr21
Chr22
HM13
STRA13
MYO1F
FEZ2
RBM6
BCL7B
MLXIPL
AMHR2
SP1
RFT1
MED23
Tissue
Adipose subcutaneous
Brain
Skeletal muscle
Whole blood

[Added Table 2]

Table 2. MetS genes prioritized from the SMR analysis

Gene	Tissue	Chr	Top SNP	P_{eQTL}	P_{GWAS}	Beta	SE	P_{SMR}	P_{HEIDI}
RBM6	Adipose subcutaneous	3	rs9814664	1.27×10^{-116}	1.28×10^{-47}	0.034	0.003	1.54×10^{-34}	-
MLXIPL	Adipose subcutaneous	7	rs17145813	5.19×10^{-12}	1.03×10^{-34}	0.111	0.018	1.78×10^{-9}	6.78×10^{-2}
BCL7B	Adipose subcutaneous	7	rs11972595	1.54×10^{-14}	4.55×10^{-9}	-0.054	0.012	3.15×10^{-6}	5.69×10^{-1}
MYO1F	Adipose subcutaneous	19	rs4804311	2.59×10^{-20}	8.05×10^{-30}	0.095	0.013	7.98×10^{-13}	1.30×10^{-1}
STRA13 (a.k.a. *CENPX*)	Adipose subcutaneous	17	rs4995642	2.78×10^{-47}	1.61×10^{-7}	-0.016	0.003	8.41×10^{-7}	1.32×10^{-1}
STRA13 (a.k.a. *CENPX*)	Brain cortex	17	rs4995642	2.17×10^{-192}	1.61×10^{-7}	-0.007	0.001	2.48×10^{-7}	6.13×10^{-2}
FEZ2	Brain cortex	2	rs10172196	2.72×10^{-88}	1.05×10^{-9}	0.013	0.002	5.42×10^{-9}	1.00×10^{-1}
MED23	Skeletal muscle	6	rs2608954	7.35×10^{-16}	1.44×10^{-16}	0.048	0.008	7.89×10^{-9}	3.10×10^{-1}
AMHR2	Skeletal muscle	12	rs2272002	3.15×10^{-11}	6.98×10^{-12}	-0.046	0.010	1.84×10^{-6}	1.92×10^{-1}
STRA13 (a.k.a. *CENPX*)	Skeletal muscle	17	rs4995642	2.99×10^{-45}	1.61×10^{-7}	-0.016	0.003	9.00×10^{-7}	1.03×10^{-1}
HM13	Whole blood	20	rs6120704	0	1.41×10^{-11}	0.021	0.003	2.14×10^{-11}	7.32×10^{-2}
MED23	Whole blood	6	rs2245133	3.75×10^{-181}	2.04×10^{-19}	0.051	0.006	8.14×10^{-18}	6.95×10^{-2}
RFT1	Whole blood	3	rs2336725	6.08×10^{-263}	7.37×10^{-21}	-0.038	0.004	1.52×10^{-19}	-
SP1	Whole blood	12	rs10876447	1.66×10^{-139}	9.12×10^{-14}	-0.032	0.004	8.95×10^{-13}	9.21×10^{-1}

P_{eQTL} is the *P*-value for the top associated cis-eQTL in the eQTL, P_{GWAS} is the *P*-value for the top associated cis-eQTL in the GWAS, Beta is the effect estimate from SMR, SE is the corresponding standard error, P_{SMR} is the *P*-value for SMR, and P_{HEIDI} is the *P*-value for the HEIDI test.
Abbreviations: Chr, chromosome; HEIDI, heterogeneity in dependent instruments

[Added to the Supplementary Table]

Supplementary Table 30. *Overview of eQTL data used in discovery and replication set*

Tissue	Discovery or replication	Cohort	N sample	N gene	Reference
Brain cortex	Discovery	BrainMeta v2	2,865	15,131	[1]
Brain cortex	Replication	GTEx v8	205	3,348	[2]
Brain frontal cortex	Replication	GTEx v8	175	2,551	[2]
Brain anterior cingulate cortex	Replication	GTEx v8	147	1,737	[2]
Whole blood	Discovery	eQTLGen	31,684	15,155	[3]
Whole blood	Replication	GTEx v8	670	6,104	[2]
Subcutaneous adipose	Replication	METSIM	434	6,347	[4]
Subcutaneous adipose	Discovery	GTEx v8	581	7,207	[2]
Skeletal muscle	Replication	FUSION	301	6,452	[5]
Skeletal muscle	Discovery	GTEx v8	706	6,245	[2]
Liver	Discovery	GTEx v8	208	2,157	[2]
Artery aorta	Discovery	GTEx v8	387	5,692	[2]
Artery coronary	Discovery	GTEx v8	213	2,378	[2]
(omitted)					

Supplementary Table 31. *SMR analysis results in discovery (Bonferroni significant)*

Tissue	Cohort	N total gene	probeID	ProbeChr	Gene	Probe_bp	topSNP	topSNP_chr	topSNP_bp	A1	A2	Freq	b_GWAS	se_GWAS	p_GWAS	b_eQTL	se_eQTL	p_eQTL	b_SMR	se_SMR	p_SMR	p_HEIDI	nsnp_HEIDI
Brain cortex	BrainMeta v2	15131	ENSG00000004534	3	`RBM6	49977440	rs2245365	3	50097319	C	A	0.53795	-0.019717	0.001362	1.50E-47	-0.859674	0.029874	4.88E-182	0.022964	0.000177495	2.75E-38	NA	20
Adipose subc	GTEx	7207	ENSG00000	3	`RBM6	49977440	rs9814664	3	50078541	T	C	0.46235	0.019734	0.001361	1.28E-47	0.576733	0.025122	1.27E-116	0.034217	0.000279166	1.54E-34	NA	20

utaneous			04534																		
Whole blood	eQTLGen	15155	ENSG00000004534	3	`RBM6	49977440	rs224536	3	50097319	A	C	0.46205	0.019741	0.001362	1.50E-47	0.754537	0.007406	0	0.026164	0.001824	1.21E-46	NA	20
Whole blood	eQTLGen	15155	ENSG00000005955	17	`GGNBP2	34900737	rs1126377	17	34896877	A	G	0.47366	-0.012428	0.001277	2.31E-22	-0.609276	0.010720	0	0.020398	0.002127	9.02E-22	NA	20
Adipose subcutaneous	GTEx	7207	ENSG00000009950	7	`MLXIPL	73007524	rs1714581	7	73046758	A	C	0.22179	0.017164	0.001396	1.03E-34	0.154709	0.022421	5.19E-12	0.110959	0.018441	1.78E-09	0.067777	5
Whole blood	eQTLGen	15155	ENSG00000011523	2	`CEP68	65283500	rs100935	2	65276452	C	T	0.37605	0.014244	0.001253	6.44E-30	-0.277036	0.008767	4.14E-219	-0.051402	0.004807	1.11E-26	NA	20
Whole blood	eQTLGen	15155	ENSG00000025434	11	`NR1H3	47269851	rs375867	11	47278917	T	C	0.30638	0.018444	0.001347	1.21E-42	-0.370563	0.008285	0	-0.049772	0.003802	3.87E-39	NA	20
(omitted)																					

Supplementary Table 32. *SMR analysis results in replication*

Tissue	Cohort	N genes tested	probeID	ProbeChr	Gene	Probe_bp	topSNP	topSNP_chr	topSNP_bp	A1	A2	Freq	b_GWAS	se_GWAS	p_GWAS	b_eQTL	se_eQTL	p_eQTL	b_SMR	se_SMR	p_SMR	p_HEIDI	nsnp_HEIDI
Adipose subcutaneous	METSIM	12	ENSG00000004534	3	`RBM6	49977440	rs2245365	3	50097319	A	C	0.46205	0.019741	0.001362	1.50E-47	0.990568	0.055457	2.88E-71	0.019839	0.001763	2.38E-29	NA	20
Adipose subcutaneous	METSIM	12	ENSG00000009950	7	`MLXIPL	73007524	rs7995868	7	73050864	A	G	0.22108	0.016822	0.001388	8.86E-34	0.827656	0.082198	7.57E-24	0.020325	0.002624	9.67E-15	0.12190	7

Whole blood	GTEx	95	ENSG00000101294	20	`HM13	30102231	rs605969 0	20	30094128	T	C	0.15491 8	0.010876 4	0.00163581	2.95E-11	0.47146 4	0.0311229	7.76E-52	0.0230694	0.0037891 5	1.14E-09	0.1453296	20
Adipose subcutaneous	METSIM	12	ENSG00000106635	7	`BCL7B	72950686	rs1198398 1	7	72939341	G	C	0.07475 75	0.012644 4	0.0022753 2	2.74E-08	-1.13778	0.160671	1.43E-12	-0.011132	0.0025420 5	1.23E-05	0.9446905	3
Whole blood	GTEx	95	ENSG00000112282	6	`MED23	131895106	rs384399 5	6	131950798	G	A	0.16725 4	-0.014953 3	0.0016880 3	8.12E-19	-0.16695 6	0.0183724	1.02E-19	0.0895643	0.0141197	2.25E-10	0.2474603	13
(omitted)																							

[Added to the Results, page 12, lines 249–253]

"We used summary data-based Mendelian randomization (SMR) using gene expression data from the following tissues to prioritize MetS genes: brain cortex (BrainMeta v2[28], *n* = 2,865), whole blood (eQTLGen[29], *n* = 31,684), subcutaneous adipose (GTEx[30], *n* = 581), skeletal muscle (GTEx, *n* = 706), liver (GTEx, *n* = 208), aorta (GTEx, *n* = 387), and coronary artery (GTEx, *n* = 213) (**Supplementary Table 30**)."

[Added to the Results, pages 12–13, lines 255–267]

"Significant associations with MetS were identified in various tissues, including the brain cortex, whole blood, subcutaneous adipose, skeletal muscle, liver, aorta, and coronary artery, with 43, 136, 14, 16, 5, 14, and 5 genes, respectively, meeting the criteria after Bonferroni correction (*P*-value <0.05/*n* genes tested) and passing the HEIDI test (*P*-value >0.05) (**Fig. 3c**, **Table 2**, and **Supplementary Table 31**). Among these, 11 genes (*AMHR2*, *BCL7B*, *FEZ2*, *HM13*, *MED23*, *MLXIPL*, *MYO1F*, *RBM6*, *RFT1*, *SP1*, and *STRA13*) were replicated in an independent cohort across the corresponding tissues, except for the liver, aorta, and coronary artery because of the lack of available independent data (**Supplementary Table 32**). Nonetheless, among the 11 replicated genes, significant associations were observed for *STRA13* in the aorta and coronary artery tissues, and *MYO1F* in the coronary artery tissue. Besides, *HM13*, *AMHR2*, *RFT1*, and *SP1* were identified through positional, eQTL, and chromatin interaction mapping using FUMA, and demonstrated significant gene associations with MAGMA."

[Added to the Discussion, page 18, lines 386–390]

"We prioritized 11 potential genes (*AMHR2*, *BCL7B*, *FEZ2*, *HM13*, *MED23*, *MLXIPL*, *MYO1F*, *RBM6*, *RFT1*, *SP1*, and *STRA13*) strongly associated with MetS through SMR analyses across MetS-relevant tissues. Additionally, whole blood tissue from the eQTLGen cohort was included because of its drug delivery suitability, biomarker measurement convenience, and large sample size enhancing statistical power[50-52]."

[Added to the Methods, page 27, lines 605–614]

"We used publicly available cis-eQTL data from BrainMeta v2[28] (*n* = 2,865) for the brain, eQTLGen[29] (*n* = 31,684) for whole blood, and GTEx v8[30] for subcutaneous adipose tissue (*n* = 581), skeletal muscle (*n* = 706), liver (*n* = 208), aorta (*n* = 387), and coronary arteries (*n* = 213) as our discovery set. The results were then validated in brain cortical tissues (cortex [*n* = 205], frontal cortex [*n* = 175], anterior cingulate cortex [*n* = 147]), and whole blood (*n* = 670) using GTEx v8. Brain cortical tissues in the GTEx were considered specifically because the genotype and RNA sequencing data of BrainMeta v2 were obtained from brain cortical tissue. The results for subcutaneous adipose and skeletal muscles were validated using METSIM[78] (*n* = 434) and FUSION[79] (*n* = 301), respectively (an overview of the eQTL data used can be obtained in **Supplementary Table 30**)."

5. A PRS for MetS again is a reasonable thing to try to develop (and show it functions ok in non-White populations) and it is not surprising that the MetS PRS explained the largest variance for prediction of MetS. But was it better at predicting adverse outcomes (cardiovascular or diabetes or death) than PRS using things like HTN, TG, T2D, LDL? I don't think a PRS for disease prediction will be that useful since we do not have interventions for MetS while we have many interventions for specific components of the MetS.

Response: We appreciate your insight into the oversight in our work. In response to your advice, we have performed multivariable Cox regression analyses to evaluate the performance of various PRS in predicting cardiovascular disease (CVD) incidences in the UK Biobank (Yun *et al*. [2022]).

Briefly, individuals with ischemic stroke, hemorrhagic stroke, peripheral artery disease, heart failure, or arterial fibrillation/flutter were categorized as having CVD incidence. Collectively, among 352,781 individuals without prior CVD history, 35,711 individuals had CVD incidence with a median follow-up period of 11 years (interquartile range of 10.2–11.7). The overall CVD incidence rate was 9.67 incidence per 1,000 persons year (95% CI = 9.57–9.77).

Multivariable Cox regression was performed to evaluate the hazard ratio (HR) of each PRS for CVD incidence. The HR of the PRS were assessed using both quantitative and categorical variables. The resultant HR of the quantitative PRS was investigated using the scaled PRS as a predictor and interpreted as the CVD incidence rate per unit standard deviation. In addition, the PRS was categorized into four groups: low risk (1–20th percentile), intermediate risk (21–80th percentile), high risk (81–99th percentile), and very high risk (100th percentile). The HR for each group were computed using the low-risk group as a reference. All multivariable Cox regression analyses were performed after adjusting for age, sex, and the first 10 principal components of genetic ancestry.

First, we compared the hazard ratios (HR) for CVD risk among different PRS risk groups. A non-overlapping HR confidence interval (CI) between the PRS groups can be seen for MetS and HTN PRS, where clear discrimination in the cumulative CVD incidence rate between the groups can be observed. The largest HR, relative to the low risk, was observed as 1.33 (95% CI = 1.28–1.38) and 1.69 (95% CI = 1.53–1.87) in high-risk and very high-risk, respectively, for the MetS PRS. Second, although all quantitative PRS showed a significant association with CVD incidence rate, the MetS PRS showed the strongest association with HR = 1.11 per unit SD (95% CI = 1.10–1.13, *P*-value = 4.66×10^{-76}). These findings suggest that the MetS PRS provides additional predictive value beyond the predictive capabilities of the individual PRS components. A further description of this analysis is provided in the **Supplementary Note**.

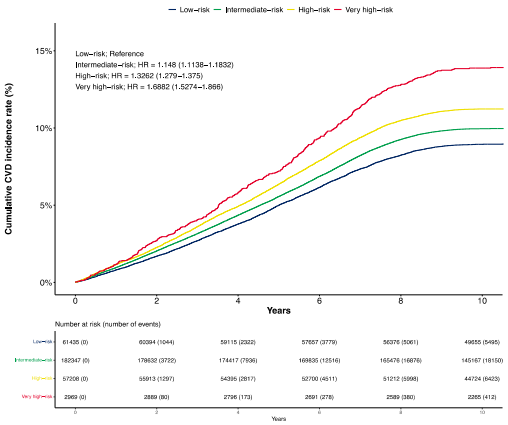
We concur that the utility of PRS for disease prediction may seem limited given the lack of direct interventions for MetS. However, identifying high-risk groups through PRS could facilitate the implementation of preventive measures, such as lifestyle modifications. Furthermore, if new treatments, such as GLP-1 inhibitors, are shown to be effective for MetS, PRS could help identify which groups are likely to gain the most clinical benefit from these interventions. In the revised manuscript, we have added a discussion on this topic. We are grateful for your advice, which has enriched the outcome of our research.

[Added Supplementary Figure]

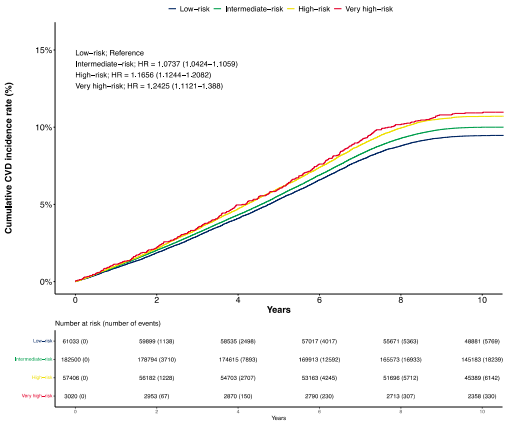
Supplementary Figure 9. Hazard plot for PRS of MetS and its components with cardiovascular disease incidence rate in the UK Biobank

Hazard plot for cardiovascular disease incidence rate in UK Biobank (UKB) with PRS stratified into four groups: low risk, intermediate risk, high risk, and very high risk. The HR of the intermediate-risk, high-risk, and very high-risk groups were annotated with low-risk as the reference group. ***a****, MetS PRS.* ***b****, BMI PRS.* ***c****, WC PRS.* ***d****, HTN PRS.* ***e****, FG PRS.* ***f****, T2D PRS.* ***g****, HDL* PRS.* ***h****, TG PRS. HR, hazard ratio; MetS, metabolic syndrome; BMI, body mass index; WC, waist circumference; FG, fasting glucose; HTN, hypertension; T2D, type 2 diabetes; HDL*: high-density lipoprotein; TG, triglyceride. *Reverse-coded.*

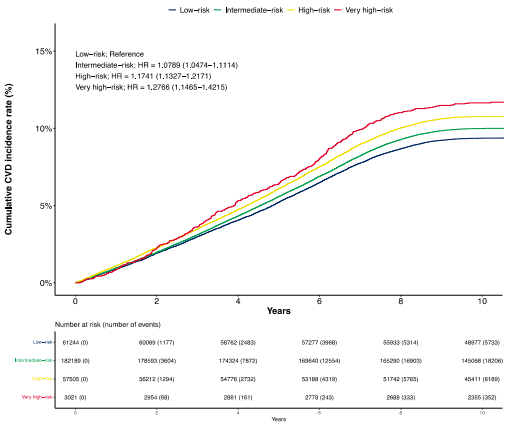
a. MetS PRS



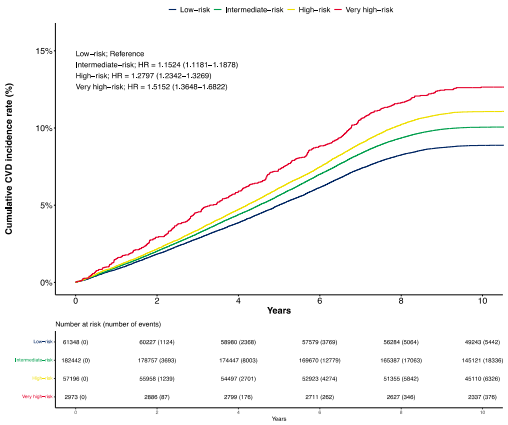
b. BMI PRS



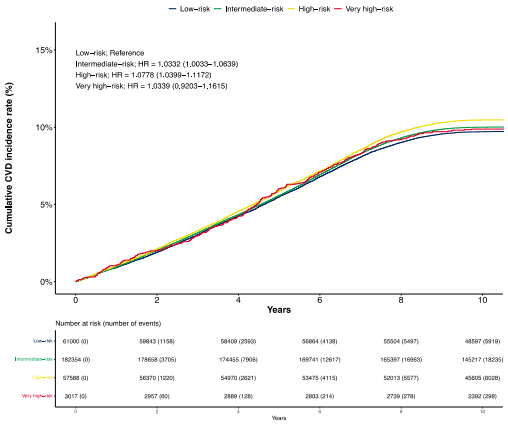
c. WC PRS



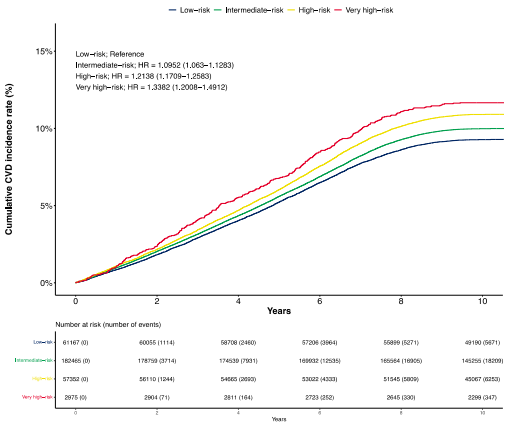
d. HTN PRS



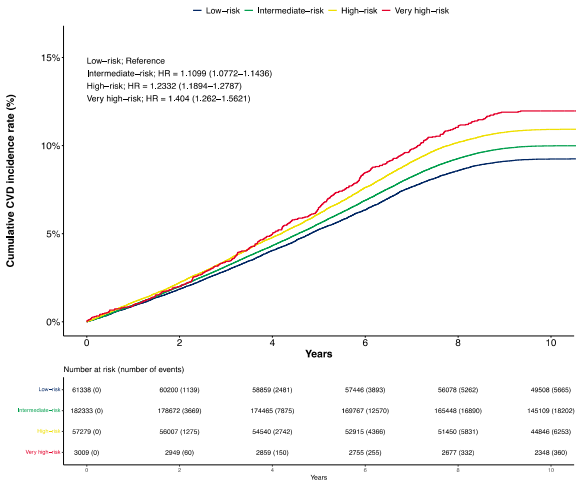
e. FG PRS



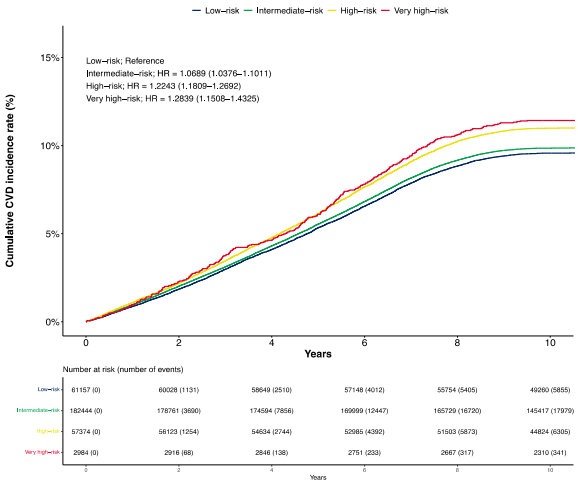
f. T2D PRS



g. HDL* PRS



h. TG PRS



[Added Supplementary Table]

Supplementary Table 35. *Multivariable Cox regression analysis for PRS of MetS and its components with cardiovascular disease incidence rate in UK Biobank*

PRS	Term	Beta	SE	P	Beta.95%CI_lo	Beta.95%CI_up	HR	HR.95%CI_lo	HR.95%CI_up
WC	PRS_scaled.WC	0.056236027	0.005725372	9.03E-23	0.045014299	0.067457756	1.057847335	1.046042817	1.069785067
TG	PRS_scaled.TG	0.072674067	0.005837153	1.39E-35	0.061233247	0.084114887	1.075379978	1.063146861	1.087753856
T2D	PRS_scaled.T2D	0.071325741	0.005740044	1.89E-35	0.060075255	0.082576227	1.073930992	1.061916458	1.086081459
MetS	PRS_scaled.MetS	0.10643637	0.005767016	4.66E-76	0.095133019	0.11773972	1.112307148	1.09980514	1.124951271
HTN	PRS_scaled.HTN	0.093738639	0.005733443	4.39E-60	0.082501091	0.104976187	1.098272663	1.085999859	1.110684161
HDL	PRS_scaled.HDL.rev	0.0803673	0.005776564	5.31E-44	0.069045234	0.091689366	1.083685032	1.071484676	1.096024307
FG	PRS_scaled.FG	0.027162899	0.005729991	2.13E-06	0.015932116	0.038393681	1.027535173	1.016059709	1.039140242
BMI	PRS_scaled.BMI	0.055744507	0.005735703	2.51E-22	0.044502529	0.066986485	1.057327509	1.04550762	1.069281027
(omitted)									

[Added to the Supplementary Note, pages 20–21, lines 427–453]

"**10. Association between polygenic risk score and cardiovascular disease incidence rate through multivariable Cox regression analysis**

MetS is a known risk factor for cardiovascular disease (CVD). We assessed whether the MetS PRS showed better stratification and a stronger association with the CVD incidence rate in the UKB than the PRS of MetS components. We followed the previous work by Yun *et al*.[53] to define CVD incidence and conducted multivariable Cox regression analyses.

Briefly, individuals with ischemic stroke … *(omitted)*"

[Added to the Results, pages 13–14, lines 286–290]

"We further investigated the relationship between MetS PRS and the incidence rate of CVD in the UKB. The MetS PRS showed distinct differences in CVD incidence rates across stratified genetic risk groups, and a notable association between the MetS PRS and CVD incidence rate (hazard ratio = 1.11, 95% CI = 1.10–1.13) was identified through multivariable Cox regression analysis (**Supplementary Fig. 9** and **Supplementary Table 35**)."

[Added to the Discussion, pages 18–19, lines 404–408]

"Furthermore, the MetS PRS exhibited slightly improved discrimination in identifying individuals at an elevated risk of developing CVD compared with the PRS of its components. These findings emphasize the utility of the MetS PRS in identifying individuals at high CVD risk and in implementing proactive lifestyle adjustments and clinical interventions."

6. The results and discussion of the likely causal genes (FEZ2, STRA13, RFT1, MED23, SP1, HM13) could be further developed. What do these genes do? Do they act through common pathways? What is known about the non-brain genes in other tissues? Are there rare variants in these genes in human populations?

Response: Thank you for the comment. In response to your previous feedback, we have identified five additional genes linked to MetS. Drawing on information from GeneCards, the Open Targets

Platform, the International Mouse Phenotyping Consortium, and pertinent previous studies, we have updated the **Supplementary Note** with a summary of each prioritized gene. We conducted over-representation analyses using the Web-based Gene Set Analysis Toolkit (WebGestalt) to investigate common pathways, such as biological processes, molecular functions, KEGG, and Reactome, among the prioritized MetS genes (Wang *et al.* [2017]). Although individual genes demonstrated an association with MetS, none of these pathways reached significance when collectively analyzed after adjusting for multiple tests. The key points have been summarized, and the **Discussion** has been revised accordingly.

[Added to the Supplementary Note, pages 12–17, lines 237–358]

"**6. Summary of 11 MetS genes prioritized using Summary-based Mendelian randomization**

A summary of the 11 genes associated with MetS is described below. This information was queried from GeneCards[39], the Open Targets Platform[40], and the International Mouse Phenotyping Consortium (IMPC)[41] (accessed on 2024.04.01).

AMHR2

The AMHR2 (Anti-Mullerian Hormone Receptor Type 2) encodes a receptor … *(omitted)*"

[Rephrased in the Discussion, page 18, lines 386–401]

"We prioritized 11 potential genes (*AMHR2*, *BCL7B*, *FEZ2*, *HM13*, *MED23*, *MLXIPL*, *MYO1F*, *RBM6*, *RFT1*, *SP1*, and *STRA13*) strongly associated with MetS through SMR analyses across MetS-relevant tissues. Additionally, whole blood tissue from the eQTLGen cohort was included because of its drug delivery suitability, biomarker measurement convenience, and large sample size enhancing statistical power[50-52]. Using resources such as GeneCards[53], the Open Targets Platform[54], and the International Mouse Phenotyping Consortium (IMPC)[55], the majority of these genes were linked to MetS components (**Supplementary Note**). *STRA13* regulates adipogenesis and lipogenesis[56] and *SP1* is crucial for the transcription of genes associated with hyperinsulinemia, T2D, and MetS in response to insulin levels[57]. Both *BCL7B* and *MLXIPL* are associated with MetS and inflammation, whereas *MLXIPL* is specifically associated with lipid metabolism[58]. Furthermore, *MED23* knockout mice had higher HDL and lean body mass, whereas

HM13 and *RBM6* deletions led to lower fasting glucose levels, which is concordant with the direction of the genetic effect from SMR analyses. The prioritized MetS genes demonstrated substantial evidence of their relevance to metabolic traits, supporting their use as potential targets for therapeutic interventions in MetS."

7. Marked limitations in literature review to build on the comments above. On a very cursory review, see that some but not all relevant papers were cited.

Cited: Genome-Wide Association Study of the Metabolic Syndrome in UK Biobank. Lind L. Metab Syndr Relat Disord. 2019 Dec;17(10):505-511. doi: 10.1089/met.2019.0070. Epub 2019 Oct 7. PMID: 31589552

Not cited:

A bivariate genome-wide approach to metabolic syndrome: STAMPEED consortium. Kraja AT...Borecki IB. Diabetes. 2011 Apr;60(4):1329-39. Epub 2011 Mar 8. PMID: 21386085

Transethnic meta-analysis of metabolic syndrome in a multiethnic study. Willems EL, Wan JY, Norden-Krichmar TM, Edwards KL, Santorico SA. Genet Epidemiol. 2020 Jan;44(1):16-25. doi: 10.1002/gepi.22267. Epub 2019 Oct 24. PMID: 31647587

Genome-wide association analysis of metabolic syndrome quantitative traits in the GENNID multiethnic family study. Wan JY, Goodman DL, Willems EL, Freedland AR, Norden-Krichmar TM, Santorico SA, Edwards KL; American Diabetes GENNID Study Group. Diabetol Metab Syndr. 2021 Jun 1;13(1):59. doi: 10.1186/s13098-021-00670-3. PMID: 34074324

Response: Thank you for providing the relevant papers. We have cited the references in the text in the appropriate context.

[Added to the Main, page 5, lines 93–96]

"Most previous MetS GWAS have focused on a binary definition of MetS. Kraja *et al*.[9] expanded on this by conducting a GWAS for MetS and pairwise combinations of its components and identified 29 common variants; however, these studies lacked robust evidence for a consistent association across all MetS components."

[Added to the Discussion, pages 19–20, lines 429–432]

"However, the potential heterogeneity in SNP effects on MetS components among various ancestral backgrounds suggests that genetic signals for MetS may vary across populations[60,61]. These findings highlight the importance of conducting analyses that include multiple ancestral groups."

Additional comments from the authors

We have updated the main figures and tables accordingly while addressing the reviewers' comments. The updates and rearrangements of the main figures and tables are listed below.

Figures and tables	Initially	After revision	Comment on changes in the structure
Figure 1	Panel A) Pair-wise genetic correlation Panel B) Genomic structural equation modeling of MetS	Panel A) Pair-wise genetic correlation Panel B) Genomic structural equation modeling of MetS Panel C) Manhattan plot of multivariate GWAS of MetS	Figure 1 has been merged with Figure 2.
Figure 2	Manhattan plot of multivariate GWAs of MetS	Genetic correlation of MetS with external traits	Initial Figure 3 is now Figure 2.
Figure 3	Genetic correlation of MetS with external traits	Panel A) Partitioned SNP-heritability analysis Panel B) Tissue enrichment analysis Panel C) Circos plot of SMR analyses	Initial Figure 4 is now Figure 3 with a modified figure for the Panel C.
Figure 4	Panel A) Partitioned SNP-heritability analysis Panel B) Tissue enrichment analysis Panel C) Miami plot of SMR analyses	Panel A) OR per MetS PRS decile Panel B) Incremental R^2 of PRSs in UKB and KoGES cohorts	Initial Figure 5 is now Figure 4. Panel B has been modified with respect to the reviewer's comments.
Figure 5	Panel A) OR per MetS PRS decile Panel B) Incremental R^2 of PRSs in UKB and KoGES cohorts	PRS-PheWAS of MetS	Initial Figure 6 is now Figure 5.
Figure 6	PRS-PheWAS of MetS	Deleted	There is no Figure 6 now.
Table 1	Summary of multivariate GWAS for MetS factor model	Summary of multivariate GWAS for MetS factor model	No change.
Table 2	Two-sample Mendelian Randomization analysis	MetS genes prioritized from the SMR analysis	Now added Table 2 for SMR analysis result listing 14 genes that have been replicated.
Table 3	There was no Table 3.	Two-sample Mendelian Randomization analysis	Initial Table 2 is now Table 3.

Response letter References

Alberti, K. G. M. M. et al. Harmonizing the Metabolic Syndrome: A Joint Interim Statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. Circulation 120, 1640–1645 (2009).

Ghodsian, N. et al. Electronic health record-based genome-wide meta-analysis provides insights on the genetic architecture of non-alcoholic fatty liver disease. CR Med 2, (2021).

Graham, S. E. et al. The power of genetic diversity in genome-wide association studies of lipids. Nature 600, 675–679 (2021).

Grotzinger, A. D. et al. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. Nature Human Behaviour 3, 513–525 (2019).

Lagou, V. et al. Sex-dimorphic genetic effects and novel loci for fasting glucose and insulin variability. Nat Commun 12, 24 (2021).

Lee, J. J. et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet 50, 1112–1121 (2018).

Manning, A. K. et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet 44, 659–669 (2012).

Yun, J.-S. et al. Polygenic risk for type 2 diabetes, lifestyle, metabolic health, and cardiovascular disease: a prospective UK Biobank study. Cardiovascular Diabetology 21, 131 (2022).

Karlsson Linnér, R. et al. Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction. Nat Neurosci 24, 1367–1376 (2021).

Storm, C. S. et al. Finding genetically-supported drug targets for Parkinson's disease using Mendelian randomization of the druggable genome. Nat Commun 12, 7342 (2021).

Turley, P. et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. Nat Genet 50, 229–237 (2018).

van Rheenen, W. et al. Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. Nat Genet 53, 1636–1648 (2021).

Võsa, U. et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat Genet 53, 1300–1310 (2021).

Wang, J., Vasaikar, S., Shi, Z., Greer, M. & Zhang, B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. Nucleic Acids Research 45, W130–W137 (2017).

Wang, X. et al. The association of metabolic syndrome and cognitive impairment in Jidong of China: a cross-sectional study. BMC Endocrine Disorders 21, 40 (2021).

Wang, Z. et al. Genome-wide association analyses of physical activity and sedentary behavior provide insights into underlying mechanisms and roles in disease prevention. Nat Genet 54, 1332–1344 (2022).

Decision Letter, first revision:

31st May 2024

Dear Dr. Won,

Your revised manuscript "Multivariate genomic analysis of 5 million people elucidates the genetic architecture of shared components of the metabolic syndrome" (NG-A63720R) has been seen by the original referees. As you will see from their comments below, they find that the paper has improved in revision, and they have no further requests. Therefore, we will be happy in principle to publish your study in Nature Genetics as an Article pending final revisions to comply with our editorial and formatting guidelines.

We are now performing detailed checks on your paper, and we will send you a checklist detailing our editorial and formatting requirements soon. Please do not upload the final materials or make any revisions until you receive this additional information from us.

Thank you again for your interest in Nature Genetics. Please do not hesitate to contact me if you have any questions.

Sincerely,
Kyle

Kyle Vogan, PhD

Senior Editor
Nature Genetics
https://orcid.org/0000-0001-9565-9665

Reviewer #1 (Remarks to the Author):

Thanks for the detailed responses to the comments. The updates to the manuscript have made it clearer and cognizant of other developments in the same area. I do not have further comments.

Reviewer #2 (Remarks to the Author):

I have thoroughly reviewed the revised version. I appreciate that the authors have well addressed my main concerns regarding the novel variants being claimed and specificity of results. I have no further concerns.

Reviewer #3 (Remarks to the Author):

The paper is massively improved. Thanks for the detailed responses.

Final Decision Letter:

29th August 2024

Dear Dr. Won,

I am delighted to say that your manuscript "Multivariate genomic analysis of 5 million people elucidates the genetic architecture of shared components of the metabolic syndrome" has been accepted for publication in an upcoming issue of Nature Genetics.

Over the next few weeks, your paper will be copyedited to ensure that it conforms to Nature Genetics style. Once your paper is typeset, you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required.

After the grant of rights is completed, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately.

You will not receive your proofs until the publishing agreement has been received through our system.

Due to the importance of these deadlines, we ask that you please let us know now whether you will be difficult to contact over the next month. If this is the case, we ask you provide us with the contact information (email, phone and fax) of someone who will be able to check the proofs on your behalf, and who will be available to address any last-minute problems.

Your paper will be published online after we receive your corrections and will appear in print in the next available issue. You can find out your date of online publication by contacting the Nature Press

Office (press@nature.com) after sending your e-proof corrections.

You may wish to make your media relations office aware of your accepted publication, in case they consider it appropriate to organize some internal or external publicity. Once your paper has been scheduled, you will receive an email confirming the publication details. This is normally 3-4 working days in advance of publication. If you need additional notice of the date and time of publication, please let the production team know when you receive the proof of your article to ensure there is sufficient time to coordinate. Further information on our embargo policies can be found here: https://www.nature.com/authors/policies/embargo.html

Before your paper is published online, we will be distributing a press release to news organizations worldwide, which may very well include details of your work. We are happy for your institution or funding agency to prepare its own press release, but it must mention the embargo date and Nature Genetics. Our Press Office may contact you closer to the time of publication, but if you or your Press Office have any enquiries in the meantime, please contact press@nature.com.

Acceptance is conditional on the data in the manuscript not being published elsewhere, or announced in the print or electronic media, until the embargo/publication date. These restrictions are not intended to deter you from presenting your data at academic meetings and conferences, but any enquiries from the media about papers not yet scheduled for publication should be referred to us.

Please note that Nature Genetics is a Transformative Journal (TJ). Authors may publish their research with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. Find out more about Transformative Journals

Authors may need to take specific actions to achieve compliance with funder and institutional open access mandates. If your research is supported by a funder that requires immediate open access (e.g. according to Plan S principles) then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route, the journal's standard licensing terms will need to be accepted, including <a href="https://www.nature.com/nature-portfolio/editorial-policies/self-archiving-and-license-to-publish. Those licensing terms will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

If you have posted a preprint on any preprint server, please ensure that the preprint details are updated with a publication reference, including the DOI and a URL to the published version of the article on the journal website.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

You can now use a single sign-on for all your accounts, view the status of all your manuscript

submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

An online order form for reprints of your paper is available at https://www.nature.com/reprints/author-reprints.html. Please let your coauthors and your institutions' public affairs office know that they are also welcome to order reprints by this method.

If you have not already done so, we strongly recommend that you upload the step-by-step protocols used in this manuscript to protocols.io. protocols.io is an open online resource that allows researchers to share their detailed experimental know-how. All uploaded protocols are made freely available and are assigned DOIs for ease of citation. Protocols can be linked to any publications in which they are used and will be linked to from your article. You can also establish a dedicated workspace to collect all your lab Protocols. By uploading your Protocols to protocols.io, you are enabling researchers to more readily reproduce or adapt the methodology you use, as well as increasing the visibility of your protocols and papers. Upload your Protocols at https://protocols.io. Further information can be found at https://www.protocols.io/help/publish-articles.

Sincerely,
Kyle

Kyle Vogan, PhD
Senior Editor
Nature Genetics
https://orcid.org/0000-0001-9565-9665