nature portfolio

Peer Review File

Genetic architecture of lumbar spinal stenosis

Corresponding Author: Mr Ville Salo

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

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Lumbar spinal stenosis (LSS) is an important cause of musculoskeletal morbidity which is clearly age-related and poorly understood. This excellent study contributes greatly to our understanding of the genetics underlying the condition. A large sample was assembled using data from 3 robust population studies (FinnGen, Estonian Biobank and UK Biobank) and utilised ICD 10 codes for categorising cases (n>30,000) and controls (n>700,000). Overall SNP based heritability of LSS was estimated at 18% and 54 genome wide significant loci were identified. Of these 7 replicated previously associated SNPs in LSS and 17 were previously associated with similar phenotypes such as back pain or disc herniation. A sensitivity analysis based on FinnGen samples of individuals coming to spine surgery for LSS revealed higher SNP heritability of 21%.

The manuscript is very well written, the study is comprehensive and limited only by the poor availability of relevant tissues (vertebral bone and intervertebral disc) in the expression datasets such as GTEx. Age of onset study showed that LSS diagnoses started around 40 and majority accumulate between 50 and 70 years, with 5 variants showing cumulative incidence for homozygotes already at age 40, and these findings were replicated in a second dataset Genetic correlations were highly significant for general health and mood-related conditions.

The limitations paragraph is well versed on the ethnic limitations of the study which is confined to Northern Europeans but the other main limitation is the tissues available for functional studies in GTeX. My comments are minor only and are largely related to the rather scant discussion following such extensive genome-wide significant findings:-

1. Are there heritability estimates of LSS which could be provided by way of background in the introduction?
2. Would it be possible to give prevalence values for symptomatic LSS in the general population, by age? This gives a scale of the clinical problem, rather than imaging prevalence estimates
3. In discussion it would be interesting to explore the pain vs spine structure genes, in particular NGF is already a drug target and merits further discussion.
4. The discussion would also benefit from causal relationship discussion regarding whole body fat-free mass and BMI - does this suggest that both bone/ muscle mass and adiposity both contribute to risk of LSS?
5. Does the FinnGen surgical dataset reflect enrichment for pain-mediating genes, as might be expected? Did the effect size of genetic correlations for general health and mood-related conditions change substantially in this subset?

This manuscript makes a very substantial contribution to the spine and back pain fields.

Reviewer #2

(Remarks to the Author)
Salo et al. conducted a GWAS meta-analysis of lumbar spinal stenosis (LSS) followed by some post-GWAS analyses. Several genetic association signals were identified, their related genes, pathways were validated using multiple methods. The novelty included a sub-sample GWAS analysis in LSS patients requiring surgical treatment and a follow-up analysis of the cumulative incidence difference of LSS in between two homozygotes for the lead variants. Below are our detailed comments.
Major comments
1. Structure,
a. Is the manuscript a short article? The manuscript does not have introduction, results and discussion.
b. abstract did not clearly explain the results. Half of the paragraph explained the motivation of the analysis.

2. GWAS meta-analysis
a. population stratification is an important consideration here. this is the basis of conduct a GWAS meta-analysis. This is specifically the case here since both Finnish and Estonians were reported to have population stratification when compared with other Europeans. Is this the reason why the LDSC intercept term also showed evidence of inflation?
b. Following the previous comment, this study relied on PanUKBB GWAS, which included non-European participants. Please specify the percentage of European participants, and evaluate how would this influence the reliability of the GWAS results.
c. Some SNPs (e.g. rs9438857) showed heterogenic effects across biobanks. Any SNP may need to report random effect meta-analysis results?
d. At least, for top hits, need to check whether the results were consistent among three biobanks. e.g. check effect size and allele freq of the top hits among three biobanks.
e. Table 1 and across the whole manuscript, p value need to be reported in scientific format (rather than 1e-8). The FinnGen Enrichment analysis is a useful analysis. However, it is not clear how the analysis was conducted, what does the value mean?
f. This study relied on FinnGen R9-version. Is it possible to update the results using FinnGen R12-version to increase the number of cases? “Meta-analysis data consists only SNPs that have been observed in at least two of the datasets.” It is unclear how many SNPs only available in FinnGen were dropped in meta-analysis. However, FinnGen contributes >50% of LSS cases, and SNPs only available in FinnGen should not be dropped.

3. The sensitivity analysis of LSS patients requiring surgical treatment, this is a quite useful analysis. Please clarify why this analysis was only been conducted in FinnGen?
4. Gene-based analysis, please specify how many genes were the same as reported by the GWAS and how many were new genes that were not reported by the GWAS? Please cite GCTA, MAGMA and other methods across the whole manuscript. What kind of pathway enrichment analysis been used?
5. Colocalization, are the eQTLs from GTEX? Does COLOC + Susie been considered?
6. Longitudinal analysis, this is an interesting analysis. Similar as a Cox regression based GWAS using tools such as GATE. Will be interesting to test the top hits using GATE and see the differences. Fig S4, seems have not massively different, since they are common variants.
7. Fig 3. I think it is not a good idea to predict medication usage using a genetic model. The variants been identified will likely to be those associated with the indication (e.g. blood pressure for anti-hypertensive drugs) rather than the drug usage itself. Similarly, I don't think operations is related to genetics. More likely to identify variants associated with major diseases.
8. Mendelian randomization
a. How were the factors been selected in the MR analysis? Please explain.
b. Only IVW and ME-Egger been applied? There are many more novel methods coming out recently.
c. For those reported in Fig4, does the author consider reverse MR results? Need to make sure there is no bi-directional causality.
d. This study identified 54 loci, but only 33 SNPs were used as instruments for LSS, as shown in Fig 4B. How would this influence the reliability of the MR results?

Minor comments:
1. Any gene reported before not showing up in this meta-analysis?
2. Sensitivity analysis of LSS patients requiring surgical treatment, need to report the number of cases in the main text. The genes need to be reported in Italic format. The comparison between the main and sensitivity effects. Need to report what kind of statistical analysis been used here to show the differences statistically? The confidence interval seems overlap with each other, so is there any statistical differences.
3. PheWAS, which tool and database been used? How many traits been included?
4. The current code availability statement does not help improve the reproducibility of this paper. Please provide the scripts used exactly for this study (including both GWAS and post-GWAS analyses) either in supplementary or on GitHub.

Reviewer #3

(Remarks to the Author)
In the present study, the authors investigated the genetic architecture of lumbar spinal stenosis (LSS). They conducted a genome-wide association study using data from FinnGen, Estonian, and UK Biobank, combining the results through a genome-wide meta-analysis, which identified 47 novel loci associated with LSS. Downstream analyses revealed that LSS is an age-related condition and identified pathways associated with more severe forms of LSS. Overall, this study has deepened our understanding of the genetic architecture of LSS. Specific comments are included below:
1.Providing the genetic background of LSS and the motivation for this study will help readers understand the novelty of this study.
2.The study lacks detailed inclusion and exclusion criteria for the EstBB and UKBB samples. Table S1 presents only the criteria for case and control exclusions in FinnGen, failing to specify the quality control details for the other two samples. In addition, a description of the procedures for genotyping, imputation, and quality control of variants in the UKB database is lacking in the methods.
3.I noticed that only FinnGen and EstBB were analyzed in the methods section, so how was the GWAS of LSS conducted in UKB, and whether Regenie software was also used, which I know is very helpful for extremely imbalanced case-control ratio samples.
4.This study conducted a MAGMA analysis; however, it does not specify the critical p-value used in this analysis.
5.This study employed Two-Sample MR to investigate the bidirectional causal relationship between LSS and its associated risk factors, which are detailed in Table S11. Please provide robust references to justify the selection of these nine risk

factors.
6.In sensitivity analyses, the results of the surgically treated GWAS compared with the original FinnGen GWAS should be provided.
7.In the results section, the authors show that the variants identified at the LSS-associated loci are also associated with many other spine-related disorders, but Table S3 did not reflect this information, please check carefully. In addition, the description of the PheWAS analysis was also not found in the Methods section; please add this information.
8.In the survival analysis of LSS-related variants, it is not clear how the difference in cumulative LSS diagnosis rate between homozygotes of lead variants is reflected in different age groups. Is the survival analysis performed by age group?
9.The expressions of scientific notation in this article lack sufficient rigor; for example, powers of ten are not represented using superscripts. Please review and correct this.

Reviewer #4

(Remarks to the Author)
Salo et al. present a large-scale meta-analysis of GWAS for lumbar spinal stenosis, comprising three major biobank resources. While a number of novel loci are identified, the study design, as well as the structure of the main text, can be further improved. To improve the clarity of the writing, it is suggested to add subtitles in each results section. Additionally, at the beginning of each paragraph, it would be better to include the rationale for why this analysis was done.

Below are some comments and suggestions:
Page #2
The authors mentioned that they “successfully replicated seven loci previously associated with LSS”. Are these seven loci the only signals reported in previous GWAS studies? Are there any previous signals not identified in the current study?

The last paragraph should be an overall design of the study rather than mixing cohort information and some analysis results.

In conditional analysis, what is the added value to understanding the genetic architecture of the disease? In those loci where more than one independent variant is identified, what are the functions of those variants? Have those variants linked to the regulatory effects of disease gene expression?

Page #3
It’ll be more straightforward to say that the heritability estimates are based on the liability scale in the main text.

“Our estimates were calculated using a larger sample, which is likely the main reason for the difference.” From the evidence presented in this section, this conclusion is unconvincing. There is no complementary analysis to support such a claim.

Page #7
“The pathway responsible for sequence-specific DNA binding showed the most significant enrichments in the MAGMA gene-set analysis”. What's the interpretation of this enrichment? Is this a unique pathway that is enriched just for this disease?

The colocalization analysis between GWAS and eQTL signals could be further enriched to provide deeper insights into the biological relevance of the findings. For instance, is there any gene only identified in the cervical spinal cord? If so, is that due to the SNP only showing eQTL in this tissue compared to other tissues tested, or the gene is only expressed in the cervical spinal cord?

Page #8
“Significant genetic correlations were found between 517 traits and LSS”. This sentence reads like all the 517 traits tested are significantly correlated with LSS.

Page #9
Why only IVW and MR-Egger methods were selected for the MR analysis? IVW is not robust to the pleiotropy effect and MR-Egger lacks statistical power compared to other recent MR methods.
For the LSS -> Back pain analysis, given the IVW is significant and MR-Egger is not, have the authors checked if there are any instrumental variables that violate any of the three key assumptions of the MR framework?

Page #14
In the meta-analysis, what is the total number of SNPs tested after selection? Were there any SNPs showing significant heterogeneity effects?

In the heritability analysis, have the authors estimated it in each cohort separately?

Page #16
In the clumping analysis, it is suggested to reduce the window size because the stringent r2 threshold already ensures that the clumped SNPs are independent. A window size of 1 or 2Mb would return more number of instrumental variables.

Reviewer #5

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #6

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Version 1:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
Salo et al have performed a whole set of additional analyses and strengthen the reliability and depth of the study. The updates have yielded some new findings too.

I just have one additional comment:
1. Re bidirectional MR, thanks for trying to use Steiger filtering. You can do a MR using LSS as exposure and risk factors as outcomes. That will directly inform existence of reverse causality.

Reviewer #3

(Remarks to the Author)
I thank the authors for their thorough and thoughtful responses to my comments. The revised manuscript has been significantly improved and adequately addresses all the concerns raised. I have no further comments and recommend acceptance in its present form.

Reviewer #4

(Remarks to the Author)
The authors have addressed most of my comments well.

One remaining question is, in the newly added text
"We estimated LD score regression-derived SNP-based heritability on the liability scale to be 0.1969-0.2499, suggesting genetic factors account for 19.7-25% of the common variation in LSS risk depending on the study dataset utilized (Table S6)."
What is exactly 0.1969-0.2499? Is this an estimate with 95% confidence interval?

Reviewer #5

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #6

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Lumbar spinal stenosis (LSS) is an important cause of musculoskeletal morbidity which is clearly age-related and poorly understood. This excellent study contributes greatly to our understanding of the genetics underlying the condition. A large sample was assembled using data from 3 robust population studies (FinnGen, Estonian Biobank and UK Biobank) and utilised ICD 10 codes for categorising cases (n>30,000) and controls (n>700,000). Overall SNP based heritability of LSS was estimated at 18% and 54 genome wide significant loci were identified. Of these 7 replicated previously associated SNPs in LSS and 17 were previously associated with similar phenotypes such as back pain or disc herniation. A sensitivity analysis based on FinnGen samples of individuals coming to spine surgery for LSS revealed higher SNP heritability of 21%.

The manuscript is very well written, the study is comprehensive and limited only by the poor availability of relevant tissues (vertebral bone and intervertebral disc) in the expression datasets such as GTEx. Age of onset study showed that LSS diagnoses started around 40 and majority accumulate between 50 and 70 years, with 5 variants showing cumulative incidence for homozygotes already at age 40, and these findings were replicated in a second dataset Genetic correlations were highly significant for general health and mood-related conditions.

The limitations paragraph is well versed on the ethnic limitations of the study which is confined to Northern Europeans but the other main limitation is the tissues available for functional studies in GTeX. My comments are minor only and are largely related to the rather scant discussion following such extensive genome-wide significant findings:-

1. Are there heritability estimates of LSS which could be provided by way of background in the introduction?
Response: Thank you for the suggestion. We now describe the heritability estimate in the introduction as follows: "*A previous MRI-based twin study[1] estimated the heritability of LSS to be 0.67, indicating a substantial genetic contribution to LSS risk*".

2. Would it be possible to give prevalence values for symptomatic LSS in the general population, by age? This gives a scale of the clinical problem, rather than imaging prevalence estimates
Response: Thank you for your feedback. We have revised the introduction to include also prevalence values for symptomatic LSS. The introduction was modified to:"*In the general population, the majority of radiological LSS cases are asymptomatic. The prevalence of diagnostic LSS is around 11%, increasing significantly with age.*".

We can presume that the LSS cases in our study were symptomatic because they were clinically defined using ICD codes. Patients without symptoms are unlikely to have sought medical care. We note in the manuscript: "*We can presume that diagnoses are symptomatic since they are clinically determined. Depending on the study cohort, the prevalence of LSS in our study varied from 2.5-8.4%; in the meta-analysis, it was 5.2%.*". Because we only have access to summary statistics from the Estonian and UK biobanks, we are unable to report age-specific prevalences from those datasets. Nevertheless, in FinnGen, we examined LSS prevalence of the across 5-year age groups, and the results are shown in the table below. We did not include this table in our manuscript, as most findings in the manuscript rely on meta-

analytic results rather than cohort-specific data. We utilized For patients over 75, it is also possible that registry data is lacking if they were diagnosed prior to registrations, which only started in Finland in the 1980s.

Age group	Controls	Cases	N in age group	Prevalence (%)
15-20	1746	9	1755	0,5
20-25	8112	29	8141	0,4
25-30	14085	68	14153	0,5
30-35	19374	153	19527	0,8
35-40	25030	321	25351	1,3
40-45	19616	717	20333	3,5
45-50	20732	1486	22218	6,7
50-55	22489	2672	25161	10,6
55-60	32986	3618	36604	9,9
60-65	31932	3736	35668	10,5
65-70	35772	4125	39897	10,3
70-75	37840	4031	41871	9,6
75-80	36168	3054	39222	7,8
80-85	24002	1551	25553	6,1
85-90	14576	441	15017	2,9
90-95	5452	69	5521	1,2

3. In discussion it would be interesting to explore the pain vs spine structure genes, in particular NGF is already a drug target and merits further discussion.
Response: Thank you for the suggestion. We now describe the pain vs spine structure phenomenon in the manuscript as follows:"*In addition, we identified and replicated previously identified loci near genes related to the nervous system, such as DCC (DCC netrin 1 receptor) and NGF (nerve growth factor), the latter of which is also already a known potential medical target for pain. The identification of nervous system-related genes suggests that, in addition to structural changes, nervous system-related pathways are potentially involved in the underlying symptomatic LSS and may thus explain the differences in symptoms experienced by patients. Similar findings have also been made in GWAS studies investigating the genetic background of LDH and chronic back pain*.".

4. The discussion would also benefit from causal relationship discussion regarding whole body fat-free mass and BMI - does this suggest that both bone/ muscle mass and adiposity both contribute to risk of LSS?
Response: Thank you for your thoughtful comment. Our findings may indeed suggest that both fat-free mass and adiposity play a role in LSS risk. While our results do not directly indicate the underlying mechanism, the result may indicate that higher body mass increases the risk for LSS. We have now included the following note in our manuscript: "*In light of our results, higher body mass seems to increase the risk of LSS.*".

5. Does the FinnGen surgical dataset reflect enrichment for pain-mediating genes, as might be expected? Did the effect size of genetic correlations for general health and mood-related conditions change substantially in this subset?
Response: We compared the original meta-analysis and the surgical meta-analysis in terms of gene-set enrichments. In fact, the MAGMA gene-set analysis for surgical meta-analysis showed the most significant enrichment of TGF-β signalling-related pathways. Subsequent top-ranked gene-sets were those linked to pain that we had observed in the LSS meta-analysis. It is important to consider that patients undergoing surgery have well-characterized

structural changes that can be surgically treated. Therefore, enrichment in structural gene sets is expected in the surgical meta-analysis. Moreover, while symptomatic LSS does not always cause pain, its symptoms, like sporadic walking, have usually a big impact on a patient's quality of life. If surgery is deemed likely to improve the patient's condition significantly, they will be operated, regardless of pain severity. In other words, pain is not usually the leading symptom when considering surgery for a symptomatic LSS patient. In the surgical meta-analysis, genetic correlations with traits related to general health and mood were slightly weaker than in the LSS meta-analysis. For example, major depressive syndrome (ICD-10 coded) id:ebi-a-GCST005903 had rg=0.40, pFDR=$1.10x10^{-19}$ in the meta-analysis and rg=0.32, pFDR=$2.34x10^{-9}$ in the surgical meta-analysis. However, the smaller sample size in the surgical subset likely contributes to these differences. These results do not necessarily reflect differences between the study populations.

This manuscript makes a very substantial contribution to the spine and back pain fields.

Reviewer #2 (Remarks to the Author):
Salo et al. conducted a GWAS meta-analysis of lumbar spinal stenosis (LSS) followed by some post-GWAS analyses. Several genetic association signals were identified, their related genes, pathways were validated using multiple methods. The novelty included a sub-sample GWAS analysis in LSS patients requiring surgical treatment and a follow-up analysis of the cumulative incidence difference of LSS in between two homozygotes for the lead variants. Below are our detailed comments.

Major comments
1. Structure,
a. Is the manuscript a short article? The manuscript does not have introduction, results and discussion.
Response: Yes, the manuscript was originally submitted in a letter format. During the review process, we have made substantial revisions to the structure and added subheadings, which we hope will clarify the structure of the manuscript.

b. abstract did not clearly explain the results. Half of the paragraph explained the motivation of the analysis.
Response: Thank you for your feedback. We have revised the abstract to describe our findings more comprehensively.

2. GWAS meta-analysis
a. population stratification is an important consideration here. this is the basis of conduct a GWAS meta-analysis. This is specifically the case here since both Finnish and Estonians were reported to have population stratification when compared with other Europeans. Is this the reason why the LDSC intercept term also showed evidence of inflation?
Response: Yes, residual population stratification in combination with polygenicity likely causes the observed inflation, as we noted in the manuscript: *"The genomic inflation factor lambda GC varied between 1.14-1.54 and suggested inflation in the test statistics. Considering the observed intercept values varied between of 1.00-1.14, the inflation is likely attributed to a polygenic signals".* For further context, Bulik-Sullivan et al.[4] demonstrated that the LDSC intercept approximates lambda under population stratification and remains close to 1 under polygenicity. Since our intercepts are considerably closer to 1 than the lambda values, we consider that the polygenicity remains a plausible explanation, while acknowledging some residual stratification.

b. Following the previous comment, this study relied on PanUKBB GWAS, which included non-European participants. Please specify the percentage of European participants, and evaluate how would this influence the reliability of the GWAS results.
Response: While the PanUKBB project includes various ethnicities, in our study, we used only the European subset: "*In our research, we used a subset of European ancestry, consisting of 5134 LSS patients and 401 917 controls from the UK Biobank, more precisely summary statistics from the PanUKBB project*." We acknowledge that the Finnish and Estonian populations have unique histories that differ from other European populations[5]. Considering that most lead variants identified in our meta-analysis showed consistent effect sizes across all three studies, our results can be considered robust (Fig S1 & Table S2).

c. Some SNPs (e.g. rs9438857) showed heterogenic effects across biobanks. Any SNP may need to report random effect meta-analysis results?
Response: Thank you for this suggestion. We have now performed a random-effect meta-analysis for the lead variants we identified in the meta-analysis. We have added a mention of this in our manuscript and reported the results of the analysis in Fig. S1 and Table S4. The estimates from the random-effects model were highly consistent with those from the fixed-effects model.

d. At least, for top hits, need to check whether the results were consistent among three biobanks. e.g. check effect size and allele freq of the top hits among three biobanks.
Response: We have checked the effect allele frequencies and effect sizes for all lead variants, and results presented in the supplementary materials (Fig S1 & Table S2). The observed effect sizes were consistent across all study populations.

e. Table 1 and across the whole manuscript, p value need to be reported in scientific format (rather than 1e-8). The FinnGen Enrichment analysis is a useful analysis. However, it is not clear how the analysis was conducted, what does the value mean?
Response: Thank you for your suggestion. We have now converted p-values to scientific format. The method for calculating Finnish enrichments was described in the table legend: "*Fin Enric., enrichment in Finns (calculated Finnish EAF/European (non-Finnish) EAF in the Genome Aggregation Database [gnomAD, v.3.1.2], where FIN EAF is the allele frequency in Finns and European (non-Finnish) EAF is the allele frequency in non-Finnish*". This value indicates whether the allele is more enriched in the Finnish population compared to the European population that did not include Finnish or Estonian participants. In the revised manuscript, the Finnish enrichments are given in Table S2, and we have clarified the interpretation in its legend:"*Values >1 mean that the EAF is larger in Finnish population than on non-Finnish Europeans and <1 that EAF is larger in non-Finnish European*".

f. This study relied on FinnGen R9-version. Is it possible to update the results using FinnGen R12-version to increase the number of cases? "Meta-analysis data consists only SNPs that have been observed in at least two of the datasets." It is unclear how many SNPs only available in FinnGen were dropped in meta-analysis. However, FinnGen contributes >50% of LSS cases, and SNPs only available in FinnGen should not be dropped.
Response: Our manuscript has now been updated from the FinnGen R9 version to FinnGen R12, and we have also incorporated updated datasets from Estonian and UK biobanks. This has resulted in both an increase in case numbers and a number of novel associations. While we agree that the SNPs identified in the FinnGen study could be reported and that the

FinnGen contribution to LSS cases is substantial, we chose to exclude these variants to improve the reproducibility and generalizability of our findings.

3. The sensitivity analysis of LSS patients requiring surgical treatment, this is a quite useful analysis. Please clarify why this analysis was only been conducted in FinnGen?
Response: We asked our collaborators to replicate the surgical GWAS in the Estonian Biobank dataset. In the resubmitted manuscript, we now report results from meta-analysis combining FinnGen and Estonian Biobank-based surgical GWASs.

4. Gene-based analysis, please specify how many genes were the same as reported by the GWAS and how many were new genes that were not reported by the GWAS? Please cite GCTA, MAGMA and other methods across the whole manuscript. What kind of pathway enrichment analysis been used?
Response: Thank you for these questions. Regarding the first point, GWAS reports associations at SNP level, whereas MAGMA provides gene-level results, so the two are not directly comparable. Since MAGMA analyses are based on our GWAS summary statistics, the genes it identifies should correspond to loci highlighted in the GWAS. If the Reviewer is referring to the overlap between MAGMA-identified genes and the genes we considered most likely to explain LSS associations at each locus, the number is 25. Regarding the 2nd point, our enrichment results refer to the MAGMA gene-set analysis[6] as stated in the manuscript. We also edited the manuscript to add citations for methods that we utilized across the manuscript.

5. Colocalization, are the eQTLs from GTEX? Does COLOC + Susie been considered?
Response: Thank you for the suggestion. In the revised manuscript, we have performed colocalization analyses using the coloc SuSiE approach, utilizing the FinnGen pipeline (https://github.com/FINNGEN/coloc.susie.direct) and leveraging a broad set of data from 571 resources, such as GWAS endpoints from UKBB and FinnGen, eQTL data from GTEx and eQTL Catalogue, and proteomics, metabolomics and lipidomics data from various sources. These enhancements have substantially improved the depth and quality of our colocalization results compared to the original version.

6. Longitudinal analysis, this is an interesting analysis. Similar as a Cox regression based GWAS using tools such as GATE. Will be interesting to test the top hits using GATE and see the differences. Fig S4, seems have not massively different, since they are common variants.
Response: Thank you for this excellent suggestion. We have now performed GATE, which yielded interesting results and validated several findings from our survival analysis. Many of the loci identified are located near the same genes that we observed in our meta-analysis, however, we also observed some age-of-onset specific associations (Table S13). Additionally, Fig S5 (previously Fig S4) has been revised due to the updated datasets. The variant near the *OTUD7B* gene now stands out more prominently (Table S12), and the exact same variant was also identified in GATE analysis, supporting its potential role in LSS onset. Additionally, as you pointed out, since the majority of the variants we observed are common, the differences between homozygotes are minimal.

7. Fig 3. I think it is not a good idea to predict medication usage using a genetic model. The variants been identified will likely to be those associated with the indication (e.g. blood pressure for anti-hypertensive drugs) rather than the drug usage itself. Similarly, I don't think operations is related to genetics. More likely to identify variants associated with major diseases.

Response: We agree that medication and operation-related traits are likely to reflect the underlying indication rather than the treatment itself. In response, we have removed the medication and operation-related traits from the genetic correlation analysis.

8. Mendelian randomization

a. How were the factors been selected in the MR analysis? Please explain.

Response: We have revised our manuscript to describe better the rationale for selecting the risk factors in the MR analysis: "*Prior research has shown associations between anthropometric traits and LSS*[7–9]*; however, the causal relationships between these factors have not been thoroughly examined.*" We also cited previous MR studies that have studied causal relationships between the same risk factors and LSS: "*Our findings for BMI and hip circumference were consistent with earlier reports*[10]*, and additionally, our study also highlights previously unreported potential causal relationships.".*

b. Only IVW and ME-Egger been applied? There are many more novel methods coming out recently.

Response: In the revised version of the manuscript, we have have expanded the methodological framework by also applying the weighted median model and conducting outlier-corrected causal estimates using MRPRESSO[11]. These results were consistent with those obtained using IVW and MR-Egger (Fig 4).

c. For those reported in Fig4, does the author consider reverse MR results? Need to make sure there is no bi-directional causality.

Response: We explored the possibility of bi-directional causality using the Steiger directionality test (mr_steiger; (https://mrcieu.github.io/TwoSampleMR/reference/mr_steiger.html). However, its results were inconclusive and did not provide clear evidence favoring one causal direction over the other, and the associations were not significant. Given this ambiguity, we interpreted the results in the context of previously observed associations[7–9] and known biological functions. We acknowledge the limitations of this approach and have added a note in the study limitations: "*While our MR analyses support potential causal relationships, the findings should be interpreted with caution, as the possibility of bi-directional causality between the endpoints cannot be excluded*."

d. This study identified 54 loci, but only 33 SNPs were used as instruments for LSS, as shown in Fig 4B. How would this influence the reliability of the MR results?

Response: Sample overlap in two-sample MR could lead to substantial bias and inflated Type I error[12]. To avoid sample overlap, we used summary statistics from a meta-analysis based on FinnGen and EstBB to extract instruments for LSS, as the summary statistics for most other traits include UKBB data. Therefore, the number of instrument SNPs is smaller than the number of loci identified. We had described this in our manuscript: "*To avoid sample overlap, since many of the GWAS data provided by the MRC-IEU are UKBB-based, we derived LSS instruments from a meta-analysis of FinnGen and EstBB summary statistics*."

Minor comments:

1. Any gene reported before not showing up in this meta-analysis?

Response: Yes, five loci that were not associated with LSS in our study have been identified in earlier studies. The manuscript now includes this information: "*Seventy-three of the loci were previously unreported. We also successfully replicated fifteen loci previously associated with LSS, while 5 loci did not replicate*.".

2. Sensitivity analysis of LSS patients requiring surgical treatment, need to report the number of cases in the main text. The genes need to be reported in Italic format. The comparison between the main and sensitivity effects. Need to report what kind of statistical analysis been used here to show the differences statistically? The confidence interval seems overlap with each other, so is there any statistical differences.
Response: Our manuscript now includes the numbers of cases and controls in the surgical meta-analysis. Gene names are now consistently formatted in italics. The captions of Figure 1B and Table S5 describe the statistical method (z-test) used to compare effect sizes. The differences in the effect sizes are statistically significant ($p<0.05$) if the effect size from one analysis falls outside the confidence interval of the other, even if the confidence intervals themselves partially overlap.

3. PheWAS, which tool and database been used? How many traits been included?
Response: Thank you for pointing this out. The description of the PheWAS analysis was unintentionally omitted from the original methods section. We have now added this information to the manuscript: "*In FinnGen, we also conducted a PheWAS analysis for lead variants associated with LSS, excluding those located in the HLA region. In the analysis we used 124 clinician-curated core endpoints to test the association of lead variants, and associations with* $p<5x10^{-8}$ *were considered significant. Endpoint definitions are available at https://www.finngen.fi/en/researchers/clinical-endpoints, and the PheWAS protocol has been described at https://github.com/FINNGEN/pheweb/.*".

4. The current code availability statement does not help improve the reproducibility of this paper. Please provide the scripts used exactly for this study (including both GWAS and post-GWAS analyses) either in supplementary or on GitHub
Response: Thank you for your comment. We would like to note that we have not developed any custom code specific to this study. All analyses were conducted using publicly available, open-source tools, as cited in Methods. We believe our current code availability statement "*The citations and URLs listed in the Methods section contain the code for every tool used in the analyses. All tools used in our study were open source*" is consistent with *Nature Communications'* guidelines. To further support transparency and reproducibility, we will make the GWAS summary statistics from the LSS meta-analysis and the surgical meta-analysis publicly available in FinnGen's public cloud bucket upon publication.

Reviewer #3 (Remarks to the Author):

In the present study, the authors investigated the genetic architecture of lumbar spinal stenosis (LSS). They conducted a genome-wide association study using data from FinnGen, Estonian, and UK Biobank, combining the results through a genome-wide meta-analysis, which identified 47 novel loci associated with LSS. Downstream analyses revealed that LSS is an age-related condition and identified pathways associated with more severe forms of LSS. Overall, this study has deepened our understanding of the genetic architecture of LSS. Specific comments are included below:

1.Providing the genetic background of LSS and the motivation for this study will help readers understand the novelty of this study.
Response: Thank you for the suggestion. We have revised the manuscript to include heritability estimates obtained in twin studies that demonstrate a significant genetic

component. We also emphasize that, despite this evidence, the genetic basis remains poorly characterized, and our study contributes to filling this gap.

2.The study lacks detailed inclusion and exclusion criteria for the EstBB and UKBB samples. Table S1 presents only the criteria for case and control exclusions in FinnGen, failing to specify the quality control details for the other two samples. In addition, a description of the procedures for genotyping, imputation, and quality control of variants in the UKB database is lacking in the methods.
Response: We have modified Table S1 to also describe the inclusion and exclusion criteria for the Estonian and UK biobanks. Furthermore, we have expanded the methods section to describe in more detail how genotyping, imputation, and quality control were conducted in the UKBB. We hope these revisions clarify the data processing steps across all cohorts.

3.I noticed that only FinnGen and EstBB were analyzed in the methods section, so how was the GWAS of LSS conducted in UKB, and whether Regenie software was also used, which I know is very helpful for extremely imbalanced case-control ratio samples.
Response: We have added a detailed description of the GWAS conducted in the UKBB to the methods section. In UKBB, GWAS was performed using an additive genetic model and the SAIGE pipeline.

4.This study conducted a MAGMA analysis; however, it does not specify the critical p-value used in this analysis.
Response: Thank you for pointing this out. We have added the p-value threshold used in the analyses to both the figure captions and within the main text of the manuscript: "*Overall, 105 genes with genome-wide significance ($p<5x10^{-8}$) were identified by GCTA-fastBAT analysis, and 148 genes were identified by the MAGMA gene-based test. Some of the genes were identified in both analyses. The pathway responsible for neurotrophin receptor binding showed the most significant enrichments in the MAGMA gene-set analysis (Fig. 2B), which used the pFDR<0.05 significance threshold.*".

5.This study employed Two-Sample MR to investigate the bidirectional causal relationship between LSS and its associated risk factors, which are detailed in Table S11. Please provide robust references to justify the selection of these nine risk factors.
Response: We have narrowed the list of used risk factors to the ones that we listed previously. Additionally, we have revised the manuscript text to better justify the selection of these traits: "*Prior research has shown associations between anthropometric traits and LSS[7–9]; however, the causal relationships between these factors have not been thoroughly examined.*" We have also cited previous MR studies investigating causal relationships between the same risk factors and LSS: "*Our findings for BMI and hip circumference were consistent with earlier reports[10], and additionally, our study also highlights previously unreported potential causal relationships.".*

6.In sensitivity analyses, the results of the surgically treated GWAS compared with the original FinnGen GWAS should be provided.
Response: We have now included a genome-wide meta-analysis of the FinnGen and EstBB surgical GWASs in the revised manuscript. The results of the surgical meta-analysis are compared with those from the original LSS meta-analysis, and in addition, we present the cohort-specific results (Fig S2, Fig S3, Table S7).

7.In the results section, the authors show that the variants identified at the LSS-associated loci

are also associated with many other spine-related disorders, but Table S3 did not reflect this information, please check carefully. In addition, the description of the PheWAS analysis was also not found in the Methods section; please add this information.
Response: Thank you for your comment. We apologize for the unintentional omission of the description of the PheWAS analysis in the original manuscript – this information has been added to the revised manuscript. In addition, we have revised the PheWAS results description: "*Several loci near genes that function in biological pathways, such as TGF-β, BMP or WNT-signalling pathways, which are known to play a role in musculoskeletal disorders were identified (Table S3). Some of observed loci have also been associated with other musculoskeletal disorders, such as osteoarthritis, lumbar disc herniation, or back pain. The same was observed for some loci in our PheWAS analysis, where we found that many LSS-associated lead variants were also associated with other musculoskeletal disorders. On the other hand, some of the associations appear to be LSS-specific based on our results (Table S3).*".

8.In the survival analysis of LSS-related variants, it is not clear how the difference in cumulative LSS diagnosis rate between homozygotes of lead variants is reflected in different age groups. Is the survival analysis performed by age group?
Response: Thank you for your comment. Our survival analysis was designed to study the cumulative incidence of LSS diagnoses across the full age range of the FinnGen sample, with age as the underlying time scale. This approach allows us to evaluate whether certain lead variants are associated with an earlier onset of LSS. Stratifying the analysis by age groups would limit the ability to capture lifetime risk and would reduce the power to detect differences in age of onset. We therefore chose to analyze genotype groups across the full cohort rather than within specific age bands. We agree that age-stratified analyses could be informative for other research questions, but such analyses are beyond the scope of the current study.

9.The expressions of scientific notation in this article lack sufficient rigor; for example, powers of ten are not represented using superscripts. Please review and correct this.
Response: Thank you for pointing this out. We have modified scientific notations to use proper formatting.

Reviewer #4 (Remarks to the Author):

Salo et al. present a large-scale meta-analysis of GWAS for lumbar spinal stenosis, comprising three major biobank resources. While a number of novel loci are identified, the study design, as well as the structure of the main text, can be further improved. To improve the clarity of the writing, it is suggested to add subtitles in each results section. Additionally, at the beginning of each paragraph, it would be better to include the rationale for why this analysis was done.

Below are some comments and suggestions:
Page #2
The authors mentioned that they "successfully replicated seven loci previously associated with LSS". Are these seven loci the only signals reported in previous GWAS studies? Are there any previous signals not identified in the current study?
Response: Thank you for your comment. Indeed, the number of replicated loci is seven. We have now identified five loci from earlier GWASs that were not associated with LSS in our study and added this information to the manuscript.

The last paragraph should be an overall design of the study rather than mixing cohort information and some analysis results.
Response: Thank you for your suggestion. We have revised the final paragraph of the introduction to describe the overall design and motivation of the study. It now reads: "*Further research is needed to fully understand the hereditary background of LSS, which is currently poorly characterized. Using data from FinnGen, Estonian Biobank, and UK Biobank, we perform a genome-wide meta-analysis and identify previously unreported loci. We also show that examining LSS by utilizing surgical procedure codes provides further information regarding a more severe form of the disease that requires surgical treatment.*".

In conditional analysis, what is the added value to understanding the genetic architecture of the disease? In those loci where more than one independent variant is identified, what are the functions of those variants? Have those variants linked to the regulatory effects of disease gene expression?
Response: Conditional analysis enables the identification of multiple independent association signals within the same locus. While its added value to a better understanding of the genetic architecture may be modest, it can reveal additional signals that are masked in standard GWAS. As with most GWAS findings, it remains challenging to identify the true association-causing variants and their functions. Most of the variants identified in the conditional analyses, however, are located in genomic regions that potentially regulate gene expression. We have added the variants identified in the conditional analysis and their annotations to Table S3.

Page #3
It'll be more straightforward to say that the heritability estimates are based on the liability scale in the main text.
Response: We have now clarified the main text that the heritability estimates are reported on the liability scale. While this was already mentioned in the methods, we agree it should also be mentioned in the main text.

"Our estimates were calculated using a larger sample, which is likely the main reason for the difference." From the evidence presented in this section, this conclusion is unconvincing. There is no complementary analysis to support such a claim.
Response: We have revised the wording regarding the observed heritability estimates and removed this claim. It now reads: "*We estimated LD score regression-derived SNP-based heritability on the liability scale to be 0.1969-0.2499, suggesting genetic factors account for 19.7-25% of the common variation in LSS risk depending on the study dataset utilized (Table S6). The genomic inflation factor lambda GC varied between 1.14-1.54 and suggested inflation in the test statistics. Considering the observed intercept values varied between of 1.00-1.14, the inflation is likely attributed to a polygenic signals*[10]*. Our heritability estimates were higher compared to the previous SNP-based heritability estimates, which have estimated that 10.1-14.9% of the common variation in LSS risk originates from the genetic factors (https://nealelab.github.io/UKBB_ldsc/h2_summary_M13_SPINSTENOSIS.html).*".

Page #7
"The pathway responsible for sequence-specific DNA binding showed the most significant enrichments in the MAGMA gene-set analysis". What's the interpretation of this enrichment? Is this a unique pathway that is enriched just for this disease?

Response: Since updating from FinnGen R9 to R12, this pathway is no longer the most significantly enriched gene-set. More broadly, gene-set enrichments should be interpreted with caution, as statistical enrichment does not necessarily indicate that the pathway in question is involved in the disease under study, but describes which biological pathways are overrepresented among associated genes. We have avoided making functional claims and emphasize the need for further functional research to determine the role of specific biological pathways in LSS. Regarding the sequence-specific DNA binding gene-set, the gene set is broad: "*Binding to DNA of a specific nucleotide composition, e.g. GC-rich DNA binding, or with a specific sequence motif or type of DNA e.g. promotor binding or rDNA binding*." Therefore, it is unlikely to be an LSS-specific signal.

The colocalization analysis between GWAS and eQTL signals could be further enriched to provide deeper insights into the biological relevance of the findings. For instance, is there any gene only identified in the cervical spinal cord? If so, is that due to the SNP only showing eQTL in this tissue compared to other tissues tested, or the gene is only expressed in the cervical spinal cord?
Response: We have updated our colocalization analysis by switching to coloc SuSiE in the revised manuscript. This has substantially expanded our analysis from four eQTL datasets to over 500 resources. In the revised version, we have attempted to provide a more profound understanding of the biological significance of the findings. In the revised colocalization results paragraph say the following: "*In total, we detected 758 colocalizations (posterior probability [PP] for the shared variant 0.8) between LSS GWAS and 322 different resources that contained numerous QTL datasets and other FinnGen R12 endpoints. Unsurprisingly, several colocalizations were observed with other back-related disorders in FinnGen. For example, LSS association signals on chr12 near gene SOX5 (SRY-box transcription factor 5) and chr15 near SMAD3 (SMAD family member 3) were colocalized with association signals of several back-related disorders (Table S12). These findings, together with our PheWAS findings, suggest that some of the biological pathways involved in the pathogenesis of LSS overlap with other back-related disorders. The LSS association signals on chr2 near GFPT1 (glutamine-fructose-6-phosphate transaminase 1) and chr7 near IL6 (interleukin 6) colocalized specifically with several immune-related eQTLs. For both variants, colocalizations were observed in eQTL datasets related to monocytes and macrophages, and also in possible structural eQTL datasets such as fibroblasts and tibial artery (Table S12).*".

Page #8
"Significant genetic correlations were found between 517 traits and LSS". This sentence reads like all the 517 traits tested are significantly correlated with LSS.
Response: Thank you for pointing this out. To clarify, genetic correlations were tested across 517 different traits, and not all of them were significant. We have revised the wording in the manuscript as follows: "*Genetic correlations were studied between 849 traits and LSS.*".

Page #9
Why only IVW and MR-Egger methods were selected for the MR analysis? IVW is not robust to the pleiotropy effect and MR-Egger lacks statistical power compared to other recent MR methods. For the LSS -> Back pain analysis, given the IVW is significant and MR-Egger is not, have the authors checked if there are any instrumental variables that violate any of the three key assumptions of the MR framework?
Response: In the revised manuscript, we expanded the MR analysis to include weighted median model and obtained outlier-corrected causal estimates using MR-PRESSO[11].
Regarding the possible causal relationship between LSS and back pain, we conducted more

sensitivity analyses. We found no indication that the fundamental MR principles were being violated but have added emphasis that MR Egger did not confirm our main findings. We have described our findings as follows: "*The only outcome for which we found LSS to be potentially causal was back pain (OR=1.027, pFDR=5.92x10^{-13}, Fig. 4B, Fig. S10, Table S16). Pain is one of the main reasons for seeking care and sometimes also the main symptom*[2]*. However, the sensitivity analysis with MR Egger (OR=1.025, pFDR=0.3156), was not statistically significant. The fact that we did not observe significant directional pleiotropy (Table S16) suggests that the null result we obtained with MR Egger could be the consequence of the known reduced statistical power of the method compared with IWV. Although pleiotropy was not observed, some causal estimates were heterogeneous (Table S16), and thus, so these results should be interpreted with cautious.*". We also performed the MR Steiger test (https://mrcieu.github.io/TwoSampleMR/reference/mr_steiger.html), a statistical analysis to determine the validity of the exposure-outcome hypothesis. However, the test results were inconclusive, as neither direction reached statistical significance.

Page #14
In the meta-analysis, what is the total number of SNPs tested after selection? Were there any SNPs showing significant heterogeneity effects?
Response: In the meta-analysis, the total number of SNPs tested was 15,144,278 and in the surgical meta-analysis 13,994,491. We have now added this information to the meta-analysis descriptions in the methods section of our manuscript. Regarding heterogeneity, some of the SNPs showed evidence of heterogeneity across cohorts. To account for this, in the revised manuscript, we also report results from a random-effect meta-analysis for the lead variants we identified in the fixed-effect meta-analysis. The results were highly consistent.

In the heritability analysis, have the authors estimated it in each cohort separately?
Response: Thank you for this suggestion. In the revised manuscript, we have also estimated heritability in each cohort separately and describe the following: "*We estimated LD score regression-derived SNP-based heritability on the liability scale to be 0.1969-0.2499, suggesting genetic factors account for 19.7-25% of the common variation in LSS risk depending on the study dataset utilized (Table S5). The genomic* inflation *factor lambda GC varied between 1.14-1.54 and suggested inflation in the test statistics. Considering the observed intercept values varied between of 1.00-1.14, the inflation is likely attributed to a polygenic signal and residual population stratification not fully controlled by principal component adjustments in SNP-based heritability estimations performed in FinnGen and Estonian Biobank*[4]*.*". In addition, the results of heritability estimation are presented in Table S5 in the supplementary materials.

Page #16
In the clumping analysis, it is suggested to reduce the window size because the stringent r2 threshold already ensures that the clumped SNPs are independent. A window size of 1 or 2Mb would return more number of instrumental variables.
Response: Thank you for the helpful suggestion. In the revised manuscript, we have decreased the clumping window from 10,000 KB to 1000 KB, which resulted in a greater number of instrumental variables without increasing pleiotropy or heterogeneity.

Reviewer #5 (Remarks to the Author):
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to

provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #6 (Remarks to the Author):
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

References

1. Battié, M. C., Ortega-Alonso, A., Niemelainen, R., Gill, K., Levalahti, E., Videman, T., & Kaprio, J. Brief report: Lumbar spinal stenosis is a highly genetic condition partly mediated by disc degeneration. *Arthritis and Rheumatology* **66**, 3505–3510 (2014).
2. Verma, A., Huffman, J. E., Rodriguez, A., Conery, M., Liu, M., Ho, Y. L., Kim, Y., Heise, D. A., Guare, L., Panickan, V. A., Garcon, H., Linares, F., Liao, K. P., *et al.* Diversity and scale: Genetic architecture of 2068 traits in the VA Million Veteran Program. *Science* **385**, eadj1182 (2024).
3. Salo, V., Määttä, J., Sliz, E., Palotie, A., Daly, M., Riley-Gills, B., Jacob, H., Paul, D., Petrovski, S., Runz, H., John, S., Okafo, G., Kettunen, J., *et al.* Genome-wide meta-analysis conducted in three large biobanks expands the genetic landscape of lumbar disc herniations. *Nat Commun* **15**, 9424 (2024).
4. Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., Daly, M. J., Price, A. L., & Neale, B. M. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291–295 (2015).
5. Kurki, M. I., Karjalainen, J., Palta, P., Sipilä, T. P., Kristiansson, K., Donner, K. M., Reeve, M. P., Laivuori, H., Aavikko, M., Kaunisto, M. A., Loukola, A., Lahtela, E., Palotie, A., *et al.* FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
6. de Leeuw, C. A., Mooij, J. M., Heskes, T., & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput Biol* **11**, e1004219 (2015).
7. Lurie, J., & Tomkins-Lane, C. Management of lumbar spinal stenosis. *BMJ* h6234 (2016) doi:10.1136/bmj.h6234.
8. Kalichman, L., Guermazi, A., Li, L., & Hunter, D. J. Association between age, sex, BMI and CT-evaluated spinal degeneration features. *J Back Musculoskelet Rehabil* **22**, 189–95 (2009).
9. Ou, C.-Y., Lee, T.-C., Lee, T.-H., & Huang, Y.-H. Impact of body mass index on adjacent segment disease after lumbar fusion for degenerative spine disease. *Neurosurgery* **76**, 396–401; discussion 401-2; quiz 402 (2015).
10. Sobczyk, M. K., Faber, B. G., Southam, L., Frysz, M., Hartley, A., Zeggini, E., Tang, H., & Gaunt, T. R. Causal relationships between anthropometric traits, bone mineral density, osteoarthritis and spinal stenosis: A Mendelian randomisation investigation. *Osteoarthritis Cartilage* **32**, 719–729 (2024).
11. Verbanck, M., Chen, C.-Y., Neale, B., & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet* **50**, 693–698 (2018).
12. Burgess, S., Davies, N. M., & Thompson, S. G. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol* **40**, 597–608 (2016).

REVIEWERS' COMMENTS

Reviewer #2 (Remarks to the Author):

Salo et al have performed a whole set of additional analyses and strengthen the reliability and depth of the study. The updates have yielded some new findings too.

I just have one additional comment:
1. Re bidirectional MR, thanks for trying to use Steiger filtering. You can do a MR using LSS as exposure and risk factors as outcomes. That will directly inform existence of reverse causality.
Response: Thank you. We performed the MR analysis as you suggested using prior exposures as an outcome, but we were unable to identify any statistically significant (pFDR<0.05) causal relationships. However, we discovered a statistically significant potential causal relationship when we examined the possible causal relationship between back pain and LSS. The results of these analyses are included in our manuscript as Table S16 and Fig. S84. We described these results and potential two-way causal relationship of back pain in LSS as follows:"*We also investigated the possibility of reverse causality, mapping potential causal relationships for the same traits in the other direction as well. Back pain was the only trait for which we observed a statistically significant potential causal relationship in both directions (Fig. S84, Table S16). The observed potential two-way causal relationship is intriguing and may be partly explained by the fact that those suffering from back pain usually seek treatment and are diagnosed*.". Based on these results, we also removed the previous limitation regarding potential reverse causality from the limitations section of our manuscript.

Reviewer #3 (Remarks to the Author):

I thank the authors for their thorough and thoughtful responses to my comments. The revised manuscript has been significantly improved and adequately addresses all the concerns raised. I have no further comments and recommend acceptance in its present form.

Reviewer #4 (Remarks to the Author):

The authors have addressed most of my comments well.

One remaining question is, in the newly added text
"We estimated LD score regression-derived SNP-based heritability on the liability scale to be 0.1969-0.2499, suggesting genetic factors account for 19.7-25% of the common variation in LSS risk depending on the study dataset utilized (Table S6)."
What is exactly 0.1969-0.2499? Is this an estimate with 95% confidence interval?
Response: We appreciate you bringing this to our attention. The sentence in question is a little ambiguous, so we changed it in the manuscript to say: "*Depending on the study dataset used, the LD score regression-derived SNP-based heritability on the liability scale ranged between 0.1969-0.2499, indicating that genetic factors explain 19.7-25% of the common variation in LSS risk.*" .

Reviewer #5 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #6 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.