Cell Genomics Transparent Peer Review Record

Genome-wide study on 72,298 individuals in Korean biobank data for 76 traits

Kisung Nam, Jangho Kim, and Seunggeun Lee

Summary

Initial submission:	Received : 2/23/2022
	Scientific editor: Rita Gemayel, Laura Zahn
First round of review:	Number of reviewers: 2 Revision invited : 5/31/2022 Revision received : 8/4/2022
Second round of review:	Number of reviewers: 2 Accepted : 9/9/2022
Data freely available:	Yes
Code freely available:	N/A

Referees' reports, first round of review

Reviewer#1

This study has further demonstrated the importance of performing additional genetic association studies in non-European population. The authors identified a good number of novel associations with human diseases and traits and, for a number of these associations, highlighted frequency difference between Asian and European populations - those that are rare in Europeans but more common in Asians may have improved power to identify such variants. There are a few suggestions and clarifications.

1. A large number of GWAS analyses have been performed in the study, and multiple testing issue needs to be handled in a more robust way. While estimating FDR can help to understand the true significance of all the results, it will be more helpful to use a more stringent threshold, such as conservative Bonferroni correction for "claiming significant associations". For example, giving that GWAS analysis has been performed by using 77 phenotypes, the threshold of genome-wide significance should be adjusted by using Bonferroni correct (to be conservative). Similarly, for the validation analysis in BBJ's cohort, it is inappropriate to use a p value of 0.05 as evidence for successful validation. In addition, the FDR calculation was based on the assumption of 1 M independent loci. How was this number of independent loci determined?

2. It is also comforting that these hits were additionally followed-up in independent EAS samples (BBJ) for validation. More information is needed to explain why only 53 top SNPs for 18 phenotypes (of the 117 novel associations for 28 traits) were presented in the BBJ dataset. Is this due to the absence of traits or SNPs or both in BBJ? At the same time, it is somewhat concerning that almost 30% of the genome-wide hits from the Korean analysis did not replicate in the Japanese population at a p value of 0,05. If using a threshold corrected for multiple testing, a much smaller number of loci were validated in the BBJ cohort. Could the authors discuss this further - was this due to differences in allele frequency between the two EAS populations and statistical power? Differences in disease definitions and trait measurements? The low validation rate in another Asian cohort also raised concern about the threshold of significance for "claiming significant loci" (above). It will be important to evaluate these variants/associations in other EAS GWAS studies.

3. Similarly, the lack of replication for a number of the identified novel genome-wide hits that were common in Europeans (MAF_NFE > 0.01 from sup fig 3) need to be discussed. Were these from traits not previously evaluated in European GWAS studies? For variants such as those in ZEB1 and CES1, which were common in European populations, did these also show at least nominal significance in the European studies?

4. The genetic correlation data could be investigated further. Besides the expected correlations were there any trait correlations that were unexpected or novel? For those traits with strong correlations, could the related set of variants and potential gene/pathways explaining this correlation be identified?

5. Especially for dietary variables it would be difficult to evaluate associations without additional transformations (overall DASH, Mediterranean diet, etc) or adjustments with overall caloric intakes. The strong correlations between the dietary variables suggest that these variables may be grouped together through such scores and it may be more appropriate to evaluate the genetic association with these overall dietary intake scores. Were such evaluations performed and if not, how would the authors interpret the novel genetic associations identified for these individual dietary variables?

6. In the pleiotropy analysis it is interesting that expected known genes such as GCKR and ALDH2 mapped to multiple traits. It is however unclear how the authors mapped the single most significant variant from each GWAS analysis to 1 specific gene in FUMA. Was this through a gene-level association analysis? If so, would it be more appropriate to utilize all regional genetic associations to determine gene-level associations? Also were there additional considerations on potential functionality of the top variant (and all SNPs in LD in the EAS reference population) in terms of eqtl and roadmap data as well as appropriate tissue/cell-type specific effects when determining the most appropriate gene?

For the CYP3A4, CYP3A7 and ANXA3 what were the additional variance explained for the corresponding traits in the EAS/Korean population? Were any potentially functional variants (either top GWAS hit or through LD analysis) that are truly EAS specific and may explain a substantial proportion of variance picked up for any of these traits?

Don't follow figure 3 - what do the x-axis and connections between dots of BBJ, KoGES and meta indicate?

7. The PRS analysis indicates slight improvements when utilizing EAS risk estimates. Will the improvement be statistically significant? Could the authors make comparisons with PRS derived with risk estimates from European studies? This may better highlight the importance for incorporating EAS data in overall genetic scores for non-European populations.

Reviewer#2

This reviewer thinks this manuscript is useful as an addition to existing East Asian GWAS studies. It does not contain any novel biological and clinical discoveries, but it can be used for future studies on some frequency related genetic studies.
They did not add much biological/functional studies as they are not experimentalists who can perform validations and further biochemical analyses.

The title can be more specific in noting numbers (instead of saying hundreds of novel loci).

They need to put more Korean GWAS studies in the past in their introduction so that readers can know the extent and novelty of this study.

>Among associations, 117 were novel, and more than
>70 percent of novel associations with corresponding phenotypes and genetic variants in BBJ were
>replicated at a nominal p-value of 0.05.

0.05 seems too high. What if it was 0.001?

>We identified 379 novel loci for 25 traits,

In how many genes?

>Korean chip genotyped and imputed were used in our analysis.

A bit of explanation on what "Korean chip" is will be good.

>To avoid false positive findings, a genetic correlation was treated
>as zero when the p-value was greater than 0.05

P value of 0.05 is too high. It will be good to have a calibration table using 0.0001, 0.001, 0.001, 0.01, and 0.05.

>131 variants in chromosome 12 were associated with more than 10 traits. SNP rs11066132 and
>rs116873087, intron variants in NAA25, were the most pleiotropic variants (23 traits).

Are these new? (these variansts and the pleiotropy). Why they cause such pleiotropy?

>the potential function of Annexin A3 (ANXA3)22,23, our result may provide a link
>between HDL level and the ANXA3 locus.

Does this make sense physiologially?

Authors' response to the first round of review

We thank the reviewers and editors for their thoughtful and constructive comments that helped improve the manuscript. Below are our detailed responses. Quoted text from the manuscript is highlighted in blue. To distinguish comments and responses, comments are

italicized. In the main manuscript and supplementary materials, the changes are marked in track changes.

Reviewer #1

1. A large number of GWAS analyses have been performed in the study, and multiple testing issue needs to be handled in a more robust way. While estimating FDR can help to understand the true significance of all the results, it will be more helpful to use a more stringent threshold, such as conservative Bonferroni correction for "claiming significant associations". For example, giving that GWAS analysis has been performed by using 77 phenotypes, the threshold of genome-wide significance should be adjusted by using Bonferroni correct (to be conservative). Similarly, for the validation analysis in BBJ's cohort, it is inappropriate to use a p value of 0.05 as
evidence for successful validation. In addition, the FDR calculation was based on the assumption
of 1 M independent loci. How was this number of independent loci determined?

RE: Thanks for the helpful comments. Previously we used the genome-wide $\alpha = 5 \times 10^{-8}$ and showed that the estimated FDR was 0.0017, which is much lower than 0.05, a commonly used significance level for the FDR control. Considering the number of tests, now we added analysis results with the more stringent criterion for adjusting the number of phenotypes in the genome-wide α, which results in $\alpha = 5 \times 10^{-8} / 76 = 6.58 \times 10^{-10}$. With this p-value cutoff, the number of significantly associated loci was reduced from 2233 loci for 47 phenotypes to 1,455 for 42 phenotypes. We also applied the more stringent threshold ($=0.05 / 53 = 9.43 \times 10^{-4}$) for the validation.

We now added: "When a more stringent criterion adjusted for the number of phenotypes at the top of the genome-wide significant level ($p < 5 \times 10^{-8} / 76 = 6.58 \times 10^{-10}$) was used, the number of significant loci was 1,455 for 42 phenotypes."

And for successful validations, we added: "With a more stringent threshold for replication by Bonferroni correction ($p < 0.05 / 53 = 9.43 \times 10^{-4}$), 25 top SNPs (47.2%) were replicated."

We note that the genome-wide $\alpha = 5 \times 10^{-8}$ is the Bonferroni corrected α with assuming 1 million independent variants. This genome-wide significant level has been a standard in GWAS studies. There have been several studies that estimated the empirical genome-wide threshold, which is closely related to the number of independent variants, and showed that the genomewide $\alpha = 5 \times 10^{-8}$ can control for type I error rates. For example, by using the 1000 Genome Project data, Kanai et al. (2016)[1] showed that the empirical genome-wide significant level for East Asian is 1.61×10^{-7}, which is larger than 5×10^{-8}. The corresponding number of

independent variants is $0.05 / (1.61 \times 10^{-7}) = 311{,}275$, smaller than 1 million independent variants we assumed. The result clearly shows that the assumption of 1 million independent variants would not inflate type I error rates.

We now added an explanation in the Results section with citations to the related studies: "The estimated false discovery rate (FDR) is 0.0017 (FDR = 76 (# of traits) × 10^6 (# of independent Loci) × 5×10^{-8} (genome-wide significance threshold) / 2223 (# of significant loci)), with assuming 1 million independent loci that correspond to genome-wide $\alpha = 5 \times 10^{-8}$."

2. It is also comforting that these hits were additionally followed-up in independent EAS samples (BBJ) for validation. More information is needed to explain why only 53 top SNPs for 18 phenotypes (of the 117 novel associations for 28 traits) were presented in the BBJ dataset. Is

this due to the absence of traits or SNPs or both in BBJ? At the same time, it is somewhat concerning that almost 30% of the genome-wide hits from the Korean analysis did not replicate in the Japanese population at a p value of 0,05. If using a threshold corrected for multiple testing, a much smaller number of loci were validated in the BBJ cohort. Could the authors discuss this further - was this due to differences in allele frequency between the two EAS populations and statistical power? Differences in disease definitions and trait measurements? The low validation rate in another Asian cohort also raised concern about the threshold of significance for "claiming significant loci" (above). It will be important to evaluate these variants/associations in other EAS GWAS studies.

RE: Thanks for the comment. Among 122 novel associations for 32 traits, 57 loci for 13 phenotypes were not compared due to the absence of similar traits (eg. Blood or glucose level in urine, alcohol drinking frequencies, nutrition intake, etc.). Of the remaining 65 top SNPs, 12 (18.5%) were not present in the BBJ. This may be because different genotyping chips and imputation panels were used in these two studies. As is mentioned in the response to the first comment, we now added results with the more stringent thresholds. As a result, 25 variants (47.2%) were successfully replicated under Bonferroni corrected threshold ($0.05 / 53 = 9.43 \times 10^{-4}$).

We acknowledge that the validation rate of novel variants in BBJ was lower than expected, and it did not substantially change even when we applied the Bonferroni corrected threshold. For variants not replicated in BBJ, allele frequencies in KoGES and BBJ were not substantially different. The low validation is probably due to the difference between these two biobanks, such as cohort characteristics, phenotype definition, and measurement.

We now added a sentence in the Discussion section to discuss it: "We acknowledge that the validation rate of novel variants in BBJ was lower than expected, and it did not substantially change even when we applied the Bonferroni corrected threshold. For variants not replicated in BBJ, allele frequencies in KoGES and BBJ were not substantially different. The low validation is probably due to the difference between these two biobanks, such as cohort characteristics, phenotype definition, and measurement."

3. Similarly, the lack of replication for a number of the identified novel genome-wide hits that were common in Europeans (MAF_NFE > 0.01 from sup fig 3) need to be discussed. Were these from traits not previously evaluated in European GWAS studies? For variants such as those in ZEB1 and CES1, which were common in European populations, did these also show at least nominal significance in the European studies?

RE: Among 122 novel top SNPs, 38 were from KoGES specific phenotypes, such as blood or glucose level in urine and some nutrition intake. We couldn't find European GWAS with the corresponding phenotypes. And 6 variants were completely monomorphic among Europeans. We note that 61 variants were common ($MAF_{NFE} > 0.01$) in European samples, among 84 remaining novel variants. When we narrow the definition of common variants to $MAF_{NFE} > 0.05$, 48 variants were common among them.

Interestingly, for variants in ZEB1 and CES1, there was no reported genetic association of ZEB1 for body weight and CES1 for LDL cholesterol in European. In UK Biobank data analysis among

Europeans, p-values of those associations > 0.05.

We added a paragraph in the Discussion section: "Although we highlighted the novel loci with low MAF among Europeans, there exist many novel loci that are not rare among Europeans. For example, rs1314013 ($MAF_{EUR} = 0.0492$) for body weight and rs9921399 ($MAF_{EUR} = 0.2646$)

for LDL cholesterol did not show a signal for association among Europeans (p-value in EUR = 0.75 and 0.17, respectively). It supports the necessity for further investigation of the genetic difference between ancestry groups."

4. The genetic correlation data could be investigated further. Besides the expected correlations were there any trait correlations that were unexpected or novel? For those traits with strong correlations, could the related set of variants and potential gene/pathways explaining this correlation be identified?

RE: We updated the genetic correlation table with more stringent p-value thresholds in Supplementary Figure 1. From the table with the most stringent cutoff (p-value < 10^{-4}), we have found the unexpected negative correlation between retinol intake and sugar intake (both after adjusting energy intake), while genetic correlations between most nutrition intake phenotypes are strongly positive.

We added a related discussion in the Discussion section: "It is not surprising that most dietaryrelated

traits showed a high genetic correlation. In the UK Biobank study, there were several clusters with strong correlations among food liking phenotypes. Interestingly, we found a negatively strong correlation between sugar intake and retinol (Vitamin A1) intake ($r_g = -0.90$, p-value = 3.4×10^{-7}), both adjusted for overall energy intake. Since there were no genomewide significant loci for both phenotypes, we could not identify the set of variants or genes to explain this correlation."

5. Especially for dietary variables it would be difficult to evaluate associations without additional

transformations (overall DASH, Mediterranean diet, etc) or adjustments with overall caloric intakes. The strong correlations between the dietary variables suggest that these variables may be grouped together through such scores and it may be more appropriate to evaluate the genetic association with these overall dietary intake scores. Were such evaluations performed and if not, how would the authors interpret the novel genetic associations identified for these individual dietary variables?

RE: Thank you for providing us helpful advice. We now conducted analysis for phenotypes related to nutrition intake with adjusting for the overall energy (calorie) intake. We still observed strong (positive) genetic correlations between most nutrition intake phenotypes even after adjusting the energy intake. We updated GWAS results (supplementary table 2), LDSC results (supplementary table 3), novel associations (supplementary table 7), and genetic correlation heatmap (supplementary figure 1) with the adjustment.

And we added a comment for this adjustment in the Methods section: "For nutrition intake phenotypes, we additionally adjusted for the total energy intake since most nutrients are closely correlated with caloric intake."

Based on the comment, we considered combining multiple nutrition intake phenotypes into a single score. There are several studies that suggest methods for calculating Diet Quality Index (DQI)[2-4]. However, components required to obtain the Diet Quality Index such as overall food group variety, intake of vegetables, fruit, and grain, and level of saturated fat, were not available in the KoGES data set. Therefore, we decided to provide the results for individual nutrition intake phenotypes for those interested in each separate phenotype.

6. In the pleiotropy analysis it is interesting that expected known genes such as GCKR and ALDH2 mapped to multiple traits. It is however unclear how the authors mapped the single most significant variant from each GWAS analysis to 1 specific gene in FUMA. Was this through a

gene-level association analysis? If so, would it be more appropriate to utilize all regional genetic associations to determine gene-level associations? Also were there additional considerations on potential functionality of the top variant (and all SNPs in LD in the EAS reference population) in terms of eqtl and roadmap data as well as appropriate tissue/cell-type specific effects when determining the most appropriate gene?

RE: We used the software FUMA to map a top variant into genes. For the mapping, FUMA uses multiple information including linkage disequilibrium (LD), eQTL, and chromatin interaction. It first characterizes independent significant SNPs and surrounding genomic loci based on LD structure. Next, independent significant SNPs are annotated for functional consequences on gene functions using ANNOVAR, CADD scores, RegulomeDB score, chromatin state, eQTL, and 3D structure of chromatin interactions. Subsequently, SNP2GENE function of FUMA maps using functional annotations of SNPs and positional, eQTL, and chromatin interaction information of SNPs. Since GWAS is mainly done with a single variant test, and due to LD, the gene-based test still cannot pinpoint the causal genes, we did not carry out additional gene-level association analyses.

We now added more details on the FUMA in the Methods section: "FUMA is a bioinformatic tool that uses multiple sources of information, including LD structure, functional score, and chromatin interaction, to link associated variants to relevant genes. FUMA first characterizes independent significant variants and surrounding genomic loci based on LD structure. Next, those variants are annotated using various tools and databases such as ANNOVAR, CADD, RegulomeDB, and Hi-C data. Then annotated variants are mapped to genes using position, eQTL association, and chromatin interaction."

7. For the CYP3A4, CYP3A7 and ANXA3 what were the additional variance explained for the corresponding traits in the EAS/Korean population? Were any potentially functional variants (either top GWAS hit or through LD analysis) that are truly EAS specific and may explain a substantial proportion of variance picked up for any of these traits?

RE: Thanks for the comments. We estimated the heritability of (single) top variants to show the proportion of variance explained by those variants. The single variant heritability can be calculated as $h^2_{GWAS} = \beta^2 \times 2 \times MAF \times (1 - MAF)$*, where* β *is the estimated effect size of the variant and MAF is the minor allele frequencies*[5]*. For rs939955, a variant associated with TG,* $h^2_{GWAS} = 4.7 \times 10^{-4}$*. And* $h^2_{GWAS} = 3.7 \times 10^{-4}$ *for rs118190473, a variant associated with HDLC. These variants only explain the modest amount of heritability compared to the most significant SNPs (*$h^2_{GWAS} = 0.0193$ *for rs74368849 with TG and* $h^2_{GWAS} = 0.0170$ *for rs72786786 with HDLC). However, it appears to be comparable to other known variants. For example, rs56156922 is associated with TG in both KoGES and BBJ (p-value in KoGES =* 9.5×10^{-9}*, and pvalue in BBJ =* 1.5×10^{-13}*), and the heritability explained by it was* 4.0×10^{-4} *and* 4.5×10^{-4} *in KoGES and BBJ, respectively.*

We added this result in the Discussion section: "We further estimated the heritability of (single) top variants to show the proportion of variance explained by those variants. The single variant heritability can be calculated as $h^2_{GWAS} = \beta^2 \times 2 \times MAF \times (1 - MAF)$, where β is the estimated effect size of the variant and MAF is the minor allele frequencies[5]. For rs939955, a variant associated with TG, $h^2_{GWAS} = 4.7 \times 10^{-4}$. And $h^2_{GWAS} = 3.7 \times 10^{-4}$ for rs118190473, a variant associated with HDLC. These variants only explain the modest amount of heritability compared to the most significant SNPs ($h^2_{GWAS} = 0.0193$ for rs74368849 with TG and $h^2_{GWAS} = 0.0170$ for rs72786786 with HDLC). However, it appears to be comparable to other known variants. For example, rs56156922 is associated with TG in both KoGES and BBJ (p-value in KoGES = 9.5×10^{-9}, and p-value in BBJ = 1.5×10^{-13}), and the heritability explained by it was 4.0×10^{-4} and 4.5×10^{-4} in KoGES and BBJ, respectively."

8. Don't follow figure 3 - what do the x-axis and connections between dots of BBJ, KoGES and meta indicate?

RE: Sorry for the unclear legend in the figure. In Figure 3, black dots indicate significance in each study. For example, 120 loci were significant only in KoGES (did not show association signal in BBJ and meta-analysis across BBJ and KoGES). A line connected between dots means simultaneous significance. For instance, 2,108 loci are associated with corresponding phenotypes both in BBJ and meta-analysis (but not significant in KoGES).

We now added a more detailed explanation of Figure 3 to clarify the meaning of the figure: "Black dots indicate significance in the analysis, and a line connected between dots represents simultaneous significance in multiple cohorts. The number of loci is counted based on the meta-analysis summary statistics after clumping for the variants with p-values less than 5×10^{-8}, window size of 5Mb, and linkage disequilibrium threshold R^2 of 0.1."

9. The PRS analysis indicates slight improvements when utilizing EAS risk estimates. Will the improvement be statistically significant? Could the authors make comparisons with PRS derived with risk estimates from European studies? This may better highlight the importance for incorporating EAS data in overall genetic scores for non-European populations.

RE: We thank the reviewer for the constructive suggestion. We added a comparison with PRS derived from European samples from UKBB summary statistics. PRS model based on a metaanalysis performed better than BBJ-based and EUR-based models for three lipid phenotypes (HDLC, LDLC, and TG). However, the EUR-based PRS model had a higher performance for two blood pressure traits (SBP and DBP). Since both EUR-based and EAS-based PRS exist, we additionally conducted a multi-ethnic PRS analysis that linearly combines EUR-based and EASbased PRS models. For all 5 phenotypes we tested, the PRS model based on EUR + metaanalysis performed better than the model constructed by EUR + BBJ. We added this in the Results and Discussion section and Supplementary Table 9.

Results section: "For the five phenotypes we tested, PRS based on the East Asian meta-analysis

($PRS_{EAS\text{-}Meta}$) provided better predictive performance, in terms of R-squared, compared to BBJbased
PRS (PRS_{BBJ}) in all models (Supplementary Table 9 a). Interestingly, the European-based PRS model (PRS_{EUR}) performed better than two East Asian-based PRS (i.e., $PRS_{EAS\text{-}Meta}$ and PRS_{BBJ})
for two blood pressure traits (SBP and DBP). We also conducted a multi-ethnic PRS analysis[6], which linearly combines PRSs from Europeans and East Asians (Supplementary Table 9 b). For all five phenotypes, the multi-ethnic PRS model based on PRS_{EUR} and $PRS_{EAS\text{-}Meta}$ performed better than the model constructed by PRS_{EUR} and PRS_{BBJ}. "

Discussion section: "For SBP and DBP, European-based PRS showed better prediction performance than East Asian-based PRS. There may be two possible reasons. As the UK Biobank
data were used for constructing EUR-based PRS, the phenotype definition and genotyping platform were identical to the test set (East Asians in UK Biobank data), while KoGES and Biobank Japan were not. It is also possible that the genetics of blood pressure may be less varying across ancestry groups than lipid phenotypes. In this case, predictive performance can be more affected by the sample sizes. The EUR-based PRS models were built using GWAS of 400K samples, while sample sizes of BBJ and meta-analysis were 140K and 210K samples, respectively."

Regarding the statistical significance, it is challenging to directly evaluate the level of significance of the improvement. Instead, we compared the significance of PRS in three models: (1) $Y \sim PRS_{BBJ}$, (2) $Y \sim PRS_{EAS\text{-}Meta}$, and (3) $Y \sim PRS_{BBJ} + PRS_{EAS\text{-}Meta}$. When comparing the first two models, the p-value of PRS_{META} in model 2 was much smaller than that of PRS_{BBJ} in model 1. Moreover, PRS_{BBJ} was no longer statistically significant in model 3 for all 5 phenotypes we tested. These results suggest that PRS based on meta-analysis explains the phenotype much better than BBJ-based PRS.

We added these results in the Results section and Supplementary Table 9: "To evaluate whether the improvement of the use of $PRS_{EAS\text{-}Meta}$ over PRS_{BBJ} is significant, we fitted the models with $PRS_{EAS\text{-}Meta}$ and PRS_{BBJ} (Supplementary Table 9 c). The first two models included each PRS only, and the third model had both $PRS_{EAS\text{-}Meta}$ over PRS_{BBJ}. When these two PRSs were
included in the model, only $PRS_{EAS\text{-}Meta}$ was statistically significant for all 5 phenotypes we tested.
In addition, the R-squared values of the model with two PRSs were not substantially different from the R-squared values of the model with $PRS_{EAS\text{-}Meta}$. It suggests that PRS based on metaanalysis
explains the phenotype better than PRS_{BBJ}."

And we added details for these analysis in the Methods section: "In addition, we conducted a multi-ethnic PRS analysis, which combines PRS from Europeans and East Asians. Multi-ethnic PRS is defined as the linear combination of two PRS: $PRS_{multi} = w_1 PRS_{EUR} + w_2 PRS_{EAS}$. We used half of the East Asian samples in UK Biobank to estimate w_1 and w_2, and the other half was used as a test set. To evaluate the improvement, we compared the significance of PRS in three linear regression models: (1) $Y \sim PRS_{BBJ}$, (2) $Y \sim PRS_{EAS\text{-}Meta}$, and (3) $Y \sim PRS_{BBJ} + PRS_{EAS\text{-}Meta}$."

Reviewer #2

1. The title can be more specific in noting numbers (instead of saying hundreds of novel loci).

RE: We thank the reviewer for the suggestion. We contemplated providing a specific number in the title. However, readers may have a different opinion on the significant p-value cut-off (reviewer 1 comment). So we decided to remove the phrase "hundreds of novel loci". The revised title is

"Genome-wide study on 72,298 individuals in Korean biobank data for 76 traits"

2. They need to put more Korean GWAS studies in the past in their introduction so that readers can know the extent and novelty of this study.

RE: We added details for existing Korean GWAS studies in the Introduction section. There are several GWAS on the Korean population[7-9], but they are limited to one or a few phenotypes of interest.

We added in the Introduction section: "Previously several GWAS were performed using KoGES data, including GWAS for anthropometric traits and some metabolites. However, these studies mainly focused on one or a few traits of interest. Recently, significant efforts have been made to catalog genetic associations in East Asians by analyzing a large number of phenotypes, including phenome-wide analysis of Biobank Japan (BBJ) and Taiwan Biobank (TWB)."

3. Among associations, 117 were novel, and more than 70 percent of novel associations with corresponding phenotypes and genetic variants in BBJ were replicated at a nominal p-value of 0.05. -> 0.05 seems too high. What if it was 0.001?

RE: When we apply a more stringent threshold for successful replication, 25 variants (47.2%) were replicated under Bonferroni corrected threshold ($0.05 / 53 = 9.43 \times 10^{-4} \approx 0.001$).

In the Results section, we added: "With a more stringent threshold for replication by Bonferroni correction ($p < 0.05 / 53 = 9.43 \times 10^{-4}$), 25 top SNPs (47.2%) were replicated.".

4. We identified 379 novel loci for 25 traits, -> In how many genes?

RE: Thanks for the comments. Since the most significant SNP can be mapped to multiple genes, and the SNP to gene mapping approach we used (i.e. FUMA) uses more information (functional annotation, eQTL, chromatin interaction, etc) than just the location information used in the GWAS catalog, it is very challenging to determine whether a particular gene is novel or not. Note that the GWAS catalog maps a SNP to a gene only using location information: A SNP is mapped to a gene when the gene includes the SNP or the gene is the closest upstream and downstream gene within 50kb of the SNP. So naively comparing the mapped gene in the GWAS catalog and our results would produce large numbers of novel genes that may be because of the different mapping approaches. For instance, 2,223 significant loci in KoGES were mapped to 4,923 genes by FUMA. Among them, only 1,483 genes (30%) were previously known, while 3,440 genes were not in the current GWAS catalog. When looking at the meta-analysis result, only 3,016 (24%) were previously identified. Due to this challenge, we decide not to provide the number of novel genes in our manuscript. Instead, we added gene mapping results from FUMA to meta-analysis summary statistics (Supplementary Table 8-2). We note that there were 12,635 associated genes for 30 phenotypes in total.

5. Korean chip genotyped and imputed were used in our analysis. -> A bit of explanation on what "Korean chip" is will be good.

RE: Thanks for the comment. We added a brief description of KoreanChip in the Methods section: "All samples in the analysis were genotyped with KoreanChip. KoreanChip is a customized array optimized for the Korean population. It has 833K variants selected using 2,576 Korean sequencing data (397 WGS and 2,179 WES). Among them, 600K variants are tagging variants for genome-wide coverage. The details of the KoreanChip can be found elsewhere."

6. To avoid false positive findings, a genetic correlation was treated as zero when the p-value

was greater than 0.05 -> P value of 0.05 is too high. It will be good to have a calibration table using 0.0001, 0.001, 0.001, 0.01, and 0.05.
RE: Thanks for the suggestion. We updated the genetic correlation table for more stringent pvalue
thresholds (0.0001, 0.001, 0.005, 0.01, and 0.05) in Supplementary Figure 1.
7. 131 variants in chromosome 12 were associated with more than 10 traits. SNP rs11066132 and rs116873087, intron variants in NAA25, were the most pleiotropic variants (23 traits). -> Are
these new? (these variants and the pleiotropy). Why they cause such pleiotropy?
RE: Thanks for the comment. We note that such a large degree of pleiotropy is identified previously. The most pleiotropic variants in KoGES, rs11066132 and rs116873087, were also pleiotropic in BBJ, associated with 48 phenotypes when applying conventional genome-wide significant threshold ($p < 5 \times 10^{-8}$). The variants are close to the ALDH2 gene, which is wellknown
for its pleiotropy and alcohol metabolism pathway, and the LD structure near this gene seems to affect such pleiotropy.
8. The potential function of Annexin A3 (ANXA3), our result may provide a link between HDL level and the ANXA3 locus. -> Does this make sense physiologically?
RE: Thanks for the comment. ANXA3 encodes a member of the annexin family and is predicted to be involved in several functions, including phospholipase A2 (PLA2s) inhibitor activity. Secretory PLA2s are known to be associated with HDL, and a mouse study has shown that overexpression
of secretory PLA2 caused the decrease in serum HDL.
Now we have added it in the Results section to discuss it: "ANXA3 encodes a member of the annexin family and is predicted to be involved in several functions, including phospholipase A2 (PLA2s) inhibitor activity. Secretory PLA2s are known to be associated with HDL, and a mouse study has shown that over-expression of secretory PLA2 caused the decrease in serum HDL[10,11]."
References

1. Kanai, M., Tanaka, T. & Okada, Y. Empirical estimation of genome-wide significance thresholds based on the 1000 Genomes Project data set. J Hum Genet 61, 861-866 (2016).
2. Patterson, R.E., Haines, P.S. & Popkin, B.M. Diet quality index: capturing a multidimensional behavior. J Am Diet Assoc 94, 57-64 (1994).
3. Haines, P.S., Siega-Riz, A.M. & Popkin, B.M. The Diet Quality Index revised: a measurement instrument for populations. J Am Diet Assoc 99, 697-704 (1999).
4. Kim, S., Haines, P.S., Siega-Riz, A.M. & Popkin, B.M. The Diet Quality Index-International (DQI-I) provides an effective tool for cross-national comparison of diet quality as illustrated by China and the United States. J Nutr 133, 3476-84 (2003).
5. Shim, H. et al. A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. PLoS One 10, e0120758 (2015).
6. Marquez-Luna, C., Loh, P.R., South Asian Type 2 Diabetes, C., Consortium, S.T.D. & Price, A.L. Multiethnic polygenic risk scores improve risk prediction in diverse populations. Genet Epidemiol 41, 811-823 (2017).
7. Kim, T., Park, A.Y., Baek, Y. & Cha, S. Genome-Wide Association Study Reveals Four Loci for Lipid Ratios in the Korean Population and the Constitutional Subgroup. PLoS One 12,

e0168137 (2017).
8. Cho, H.W., Jin, H.S. & Eom, Y.B. A Genome-Wide Association Study of Novel Genetic Variants Associated With Anthropometric Traits in Koreans. Front Genet 12, 669215 (2021).
9. Park, J.S., Kim, Y. & Kang, J. Genome-wide meta-analysis revealed several genetic loci associated with serum uric acid levels in Korean population: an analysis of Korea Biobank data. J Hum Genet 67, 231-237 (2022).
10. Tietge, U.J. et al. Human secretory phospholipase A2 mediates decreased plasma levels of HDL cholesterol and apoA-I in response to inflammation in human apoA-I transgenic mice. Arterioscler Thromb Vasc Biol 22, 1213-8 (2002).
11. Curcic, S. et al. Secretory phospholipase A2 modified HDL rapidly and potently suppresses platelet activation. Sci Rep 7, 8030 (2017).

Referees' report, second round of review

Reviewer#1

All the comments have been addressed. No further comments/suggestions.

Reviewer#2

The authors have improved the manuscript and it is acceptable.

Authors' response to the second round of review

Thanks for the positive response!