Published in final edited form as:

Genet Epidemiol. 2020 October ; 44(7): 717–732. doi:10.1002/gepi.22333.

Leveraging Existing GWAS Summary Data of Genetically Correlated and Uncorrelated Traits to Improve Power for a New GWAS

Haoran Xue¹, Chong Wu², Wei Pan³

¹School of Statistics, University of Minnesota, Minneapolis, Minnesota 55455.

²Department of Statistics, Florida State University, Tallahassee, Florida, 32306.

³Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minnesota 55455.

Abstract

In spite of the tremendous success of genome-wide association studies (GWAS) in identifying genetic variants associated with complex traits and common diseases, many more are yet to be discovered. Hence, it is always desirable to improve the statistical power of GWAS. Paralleling with the intensive efforts of integrating GWAS with functional annotations or other omic data, we propose leveraging other published GWAS summary data to boost statistical power for a new/focus GWAS; the traits of the published GWAS may or may not be genetically correlated with the target trait of the new GWAS. Building on weighted hypothesis testing with a solid theoretical foundation, we develop a novel and effective method to construct SNP-specific weights based on 22 published GWAS datasets with various traits, detecting sometimes dramatically increased numbers of significant SNPs and independent loci as compared to the standard/unweighted analysis. For example, by integrating a schizophrenia GWAS summary dataset with 19 other GWAS summary datasets of non-schizophrenia traits, our new method identified 1585 genome-wide significant SNPs mapping to 15 LD-independent loci, largely exceeding 818 significant SNPs in 13 independent loci identified by the standard/unweighted analysis; furthermore, using a later and larger schizophrenia GWAS summary dataset as the validation data, 1423 (out of 1585) significant SNPs identified by the weighted analysis, compared to 705 (out of 818) by the unweighted analysis, were confirmed, while all 15 and 13 independent loci were also confirmed. Similar conclusions were reached with lipids and Alzheimer's disease traits. We conclude that the proposed approach is simple and cost-effective to improve GWAS power.

Keywords

Pleiotropy; SNP; Statistical power; Weighted hypothesis testing

Corresponding author. panxx014@umn.edu. Phone: 612-624-4655. Fax: 612-626-0660.

Supplementary Materials

In a Supplementary File we provide more details on other datasets and more analysis results.

1 Introduction

Genome-wide association studies (GWAS) have been quite successful in identifying genetic variants, mostly single nucleotide polymorphisms (SNPs), associated with complex traits and common diseases (Visscher et al., 2017). However, the identified variants can only explain a small proportion of heritability and many more are waiting to be discovered (Visscher et al., 2017). Accordingly, substantial efforts have been devoted to improving the power of GWAS. One obvious approach is to increase the sample size, which however is more costly and will eventually be limited by the number of available individuals. Alternatively, a major effort has been on integrating GWAS with SNP or gene functional annotations. As shown by many authors (Maurano et al., 2012; Schork et al., 2013; Finucane et al., 2015; Lu, Powles, Wang, He, & Zhao, 2016b), the SNPs are not created equal; some SNPs with certain functions are more likely to be trait-associated. Accordingly, many approaches have been proposed to integrate GWAS with various SNP functional annotations and other prior knowledge (Eskin, 2008; Darnell, Duong, Han, & Eskin, 2012; Chung, Yang, Li, Gelernter, & Zhao, 2014; Pickrell, 2014; Ignatiadis, Klaus, Zaugg, & Huber, 2016; Lu, Yao, Hu, & Zhao, 2016a; Sveinbjornsson et al., 2016; Kichaev et al., 2019), or with other omic data (Ho et al., 2014; Schmidt et al., 2015; Gamazon et al., 2015; Gusev et al., 2016; Xu, Wu, Wei, & Pan, 2017; Wu & Pan, 2019).

Our motivation, albeit related, is different: instead of using other genomic data or functional annotations directly, we use other published GWAS data to indirectly infer and thus use the prior information of SNPs. Since most GWAS trait-associated SNPs are not in the coding region of the genome, they are believed to play some regulatory roles, such as in transcriptional regulation. For example, it has been shown that GWAS trait-associated SNPs are more likely to be eQTLs (Nicolae et al., 2010). Hence, instead of using possibly incomplete or possibly non-informative functional annotations, we hypothesize that the association statistics of the SNPs from published GWAS summary data of various traits contain information about the regulatory roles of the SNPs; in other words, those more significant SNPs (with larger association statistics) are hypothesized to be more likely to have some regulatory roles, and thus more likely to be trait-associated in *another new/focus* GWAS of a target trait that may or may *not* be known to be related to the traits of the previous GWAS. Our approach is more general than those based on pleiotropy, which require that the GWAS traits are related with high genetic correlations or shared genetic components (Andreassen et al., 2013a; Andreassen et al., 2013b; Chung, Yang, Li, Gelernter, & Zhao, 2014; Watanabe et al., 2019). As more and more GWAS summary data are published, it offers an ample opportunity to take full use of them with our proposed approach as a cost-effective one to boost statistical power for other GWAS of interest. Our approach is based on the practice and theory of weighted hypothesis testing pioneered by Roeder and co-worker (Roeder, Devlin, & Wasserman, 2007; Genovese, Roeder, & Wasserman, 2006), which has been effectively applied in other contexts (Roeder, Bacanu, Wasserman, & Devlin, 2006; Ionita-Laza, McQueen, Laird, & Lange, 2007; Kang, Ye, Liu, Allison, & Gao, 2009; Gui, Tosteson, & Borsuk, 2012; Hsu et al., 2012; Ignatiadis, Klaus, Zaugg, & Huber, 2016; Sveinbjornsson et al., 2016). It is emphasized that there is a solid theoretical foundation of the adopted weighted hypothesis testing. For example, under mild

conditions, it is guaranteed to maintain correct type I error rates regardless of the weights being used. Of course, if the weights are informative (i.e. those trait-associated SNPs are given higher weights), the weighted hypothesis testing will boost statistical power, often substantially, over the usual or standard/unweighted testing. Furthermore, the weighted hypothesis testing is quite robust to mis-specified weights with no or barely any loss of power. An innovation with our approach is to develop a simple and efficient method to data-adaptively combine and borrow information across multiple GWAS datasets, which is both critical and, to our best knowledge, unique. We applied our proposed method to 22 GWAS traits, convincingly showing its sometimes dramatically improved power in identifying trait-associated SNPs and independent loci; furthermore, for several traits with more recent and large-scale GWAS summary datasets, we validated almost all of the newly discovery loci.

2 Methods

2.1 Data

2.1.1 Discovery samples—We used 22 large-scale GWAS summary datasets in our analysis from 11 publications (Ripke et al., 2011; Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013; Schunkert et al., 2011; Okbay et al., 2016; Zheng et al., 2015; Manning et al., 2012; Wheeler et al., 2017; Teslovich et al., 2010; Lambert et al., 2013; Day et al., 2017; Furberg et al., 2010) with the traits covering psychiatric/mental disorders (such as Schizophrenia (SCZ), attention deficit-hyperactivity disorder, bipolar disorder, major depressive disorder, depressive symptoms, autism spectrum disorder), neurodegenerative diseases (Alzheimer's disease (AD), coronary artery disease, blood lipids (high-density lipid cholesterol (HDL), low-density lipid cholesterol (LDL), triglycerides (TG) and total cholesterol (TC)), fasting glucose, bone mineral density (for forearm, femoral neck and lumbar spine), age at menarche, and smoking behavior (such as cigarettes smoked per day, ever smoker, former smoker and age at onset of smoking), and with the sample sizes ranging from tens of thousands to hundreds of thousands; the details are given in Supplementary Table S1.

The 22 datasets were from different studies with most (if not all) subjects of European ancestry, but genotyped on different chips with different sets of SNPs. So we imputed each GWAS summary dataset separately using software ImpG (Pasaniuc et al., 2014) with the 1000 Genomes phase 3 European samples as the reference panel (The 1000 Genomes Project Consortium, 2015). We also did some quality control on the imputed datasets, including deleting imputed SNPs with $R^2 < 0.3$. Finally, we took the common subset (i.e. intersection) of the available SNPs from all the 22 imputed datasets, resulting in $N = 8,329,510$ SNPs, each with an original or imputed Z-score for each of the 22 traits.

Independent datasets of all traits: Since weighted hypothesis testing requires the independence between the constructed weights and the test statistics of the SNPs being tested, our method can only use other GWAS samples independent of (i.e. non-overlapping with) the target GWAS. We thus used the (estimated) null SNPs not associated with the GWAS trait from each GWAS sample to calculate its Pearson correlation with each of other traits; if the correlation was larger than a threshold, we inferred that the corresponding two

GWAS samples might overlap, thus excluding the use of either sample for the other. Such a method has been widely used to adjust for possibly overlapping subjects (or cryptic relatedness) across multiple GWAS datasets (Lin & Sullivan, 2009; Kim, Bai, & Pan, 2015). To be conservative, we used multiple thresholds to select null SNPs for each GWAS. Specifically, for two traits T_1 and T_2 , and $\gamma = 1, 2, 3, 4, 5$, we did the following. First, we obtained all SNPs whose absolute values of the Z-scores for two traits T_1 and T_2 were both less than γ , and denote their corresponding Z-scores as $\mathbf{Z}_1 = (Z_{11}, \dots, Z_{1k})^T$ and $\mathbf{Z}_2 = (Z_{21}, \dots, Z_{2k})^T$ for the two traits respectively. Then we calculated $\text{corr}_\gamma(T_1, T_2) = \text{cor}(\mathbf{Z}_1, \mathbf{Z}_2)$. Second, we defined $\text{corr}(T_1, T_2) = \max_{1 \leq \gamma \leq 5} |\text{corr}_\gamma(T_1, T_2)|$. If $\text{corr}(T_1, T_2) > 0.05$, we did not use either one trait/dataset in the weighted analysis (including generating weights) for the other.

Supplementary Figure S1 shows the pairwise correlations for the 22 traits. It is not surprising that the traits from the same study/publication were often highly correlated, but there were also some exceptions. For example, since the four lipid traits were taken from the same study, for any one of them, all other three traits would be excluded. In addition, perhaps the high (null) correlations among traits schizophrenia, major depressive disorder and bipolar disorder were expected, partly because they were all from the GWAS conducted by the same Psychiatric Genomics Consortium (PGC). On the other hand, it is interesting to note a relatively high correlation of 0.054 between traits coronary artery disease and bipolar disorder, slightly higher than 0.051 between major depressive disorder and bipolar disorder. If a correlation was larger than 0.05, we excluded the use of either GWAS sample for the other. For each GWAS trait, the number of the other GWAS traits excluded ranged from 1 to 7; the detail is given in Supplementary Table S2.

Independent datasets of genetically uncorrelated traits: After excluding the GWAS datasets/traits that might have overlapped samples for each other, some remaining traits might be genetically correlated. We aimed to study whether/how much power might be improved by weighted testing using only genetically uncorrelated traits. Genetic correlations between any two of 49 traits were studied in (Bulik-Sullivan et al., 2015). Except Major Depressive Disorder (MDD) and Forearm Bone Mineral Density (FA), the other traits were covered by those 49 traits; we used Depression and Femoral Neck Bone Mineral Density (FN) as surrogates for MDD and FA respectively. Figure S24 shows the p-values for all 190 pairs out of 22 traits of their genetic correlations drawn from (Bulik-Sullivan et al., 2015); after a Bonferroni correction, we show the pairs with p-value $\leq 0.05/190$ as red cells. For any pair of traits being significant, we further excluded them for each other when conducting weighted testing with only genetically uncorrelated traits. Note that, even though the p-values between the following pairs were not significant (i.e. $> 0.05/190$): SCZ and ADHD, SCZ and AUT, CAD and LDL, CAD and TC, due to their possibly perceived close relationships, to be conservative, we also excluded these pairs. Details for all the exclusions are shown in Table S17.

2.1.2 Validation samples—For the four lipids traits (HDL, LDL, TC and TG), we used the GWAS data of year 2010 in the standard weighted analyses (as the discovery sample); we then used a more recent and larger (exome-focused) GWAS dataset of year 2017 for

these four traits (Lu et al., 2017) as the Gold Standard (for partial validation). Specifically, if one *SNP* was claimed to be significant in the weighted (or standard) analysis (of the 2010 lipids GWAS data, along with other GWAS traits), and it is also significant in the standard single-*SNP* analysis in the corresponding (2017) Gold Standard or validation data, we say that it is validated. For Schizophrenia (SCZ), we used the PGC GWAS data of year 2011 (Ripke et al., 2011) in our analysis for discovery, and used another more recent GWAS data of year 2014 (Ripke et al., 2014) as the Gold Standard data. For AD, we used the 2013 IGAP (Lambert et al., 2013) and 2019 (Jansen et al., 2019) GWAS datasets as the discovery and validation samples respectively. In general, the validation samples were much larger than the discovery samples as shown in Table 1. Since a discovery sample and a Gold Standard/validation sample might overlap (while it would be impossible to separate them with only GWAS summary statistics as we used), this consisted of only a partial (non-independent) validation.

2.2 Weighted Analysis

For a given target/focus trait, denoted T_0 , after excluding any other traits whose (null) correlations with T_0 was larger than 0.05 (as discussed above), we denoted the remaining m traits as T_1 to T_m , which were used to generate a weight for each *SNP* (with a power transformation). Recall that we had $N = 8,329,510$ SNPs in total. For each *SNP* i , we denoted its p-value for trait T_0 as p_i , its Z-score for trait T_j as Z_{ji} , $i = 1, \dots, N$, and $j = 1, \dots, m$. For a positive integer x as the power index, the un-normalized weight for *SNP* i was $u_i = u_i(x) = \sum_{j=1}^m |z_{ji}|^x$, then its normalized version was $w_i = Nu_i / \sum_{i=1}^N u_i$.

It has previously been shown that grouping p-values improves the performance of weighted testing (Roeder, Devlin, & Wasserman, 2007; Roeder & Wasserman, 2009). For our main analysis we therefore partitioned SNPs into (almost) 100 equally-sized groups, with the last group sized according to the remaining number of SNPs. We explored the effect of grouping versus non-grouping, i.e. just using p_i/w_i , further in Supplementary Material. As shown in Supplementary Section 4, directly using p_i/w_i without grouping for the weighted test could detect new signals, but grouping with more stable and smoother weights gave more robust results, and was thus preferred. For group k , $1 \leq k \leq 100$, we denoted the SNPs belonging to this group as $SNP_{k1}, \dots, SNP_{kn}$, where n was the size of the group. We calculated $W_k = \sum_{i=1}^n w_{ki}/n$ as the weight for group k . For each *SNP* i in group k , its weighted p-value was its original p-value divided by W_k , i.e. p_i/W_k . As in the standard analysis of GWAS, if the weighted p-value of an *SNP* was less than 5×10^{-8} , we claimed that the *SNP* was associated with trait T_0 .

Once we discovered significant SNPs, we mapped them into Linkage Disequilibrium (LD) independent loci. (Berisa & Pickrell, 2016) partitioned the whole human genome into 1,703 approximately LD independent loci, which were used throughout this paper.

The power index x serves as a tuning parameter to data-adaptively combine information across multiple GWAS datasets/traits, similar to that used in an adaptive gene-based association test to optimally combine association information across multiple SNPs (Pan et al., 2014). It is noted that $x = 0$ leads to the usual unweighted analysis (by giving each *SNP*

an equal weight while ignoring all other GWAS data). The optimal choice of the tuning parameter x depends on unknown relationships among the traits, thus is difficult to determine a priori. Accordingly, we proposed trying each of $x = 1, 2, \dots, 10$ (while the usual unweighted analysis is always conducted with $x = 0$). The main idea was to use a larger value of x to more heavily over-weight the traits/datasets with larger Z-scores. Let's consider SNP 1 and its corresponding Z-scores $Z_{.1} = (Z_{11}, Z_{21}, \dots, Z_{m1})'$ across the m GWA studies/traits. When $x = 1$, each trait j contributed to the weight w_1 proportional to its $|Z_{j1}|$; as x increased, a larger $|Z_{j1}|$, and thus the corresponding trait j , would contribute more to the weight w_1 ; when x was large enough, only the largest $|Z_{j1}|$ would contribute:

$$w_1 \propto \|Z_{.1}\|_x \rightarrow \|Z_{.1}\|_\infty = \max\{|Z_{11}|, |Z_{21}|, \dots, |Z_{m1}|\}$$

as $x \rightarrow \infty$. We would show the results for different values of power index x , and thus chose the one giving the largest number of independent locus-trait associations; if multiple x values gave the same largest number of associations, we chose the smallest of such x values. While this approach is the focus here to demonstrate the *potential* power gains in unraveling more trait-associated SNPs and loci with a weighted analysis, it is acknowledged that this approach does not adjust for multiple testing with various values of power index x .

2.3 Combining Results Across Multiple Sets of Weights

We propose combining multiple weighted analysis results across multiple sets of weights (corresponding to multiple values of power index x) using the Bonferroni adjustment. To alleviate the problem of being conservative with the Bonferroni adjustment, we estimated the effective number of independent tests since the multiple tests based on multiple x values were likely to be correlated (van der Sluis, Posthuma, & Dolan, 2013).

The procedure was based on the eigenvalues of the correlation matrix of the p-values from these multiple correlated tests (Li, Yeung, Cherny, & Sham, 2012). Specifically, after the weighted analysis with each candidate value of power index x , we obtained the p-values of all the SNP for each x ; next we calculated the Pearson's correlation coefficient of the p-values across all (or a subset of) SNPs for each pair of the x values, obtaining a correlation matrix. Then the effective number of the independent tests was estimated as

$$M_e = m - \sum_{i=1}^m I(\lambda_i > 1) \cdot (\lambda_i - 1),$$

where λ_i 's were the eigenvalues of the $m \times m$ correlation matrix of the p-values from m tests, and $I()$ was the indicator function. When the multiple tests are independent, we have all $\lambda_i = 1$ and thus $M_e = m$; on the other hand, if the tests are correlated, we expect $M_e < m$, leading to a lower loss of power by the Bonferroni adjustment.

We applied the Bonferroni adjustment with M_e : for any given SNP, if its minimum p-value across all of its weighted (and unweighted) analyses with multiple power index x values was smaller than $5 \times 10^{-8} / M_e$, it was claimed to be statistically significant.

2.4 Data Availability

All the GWAS summary data are publicly available (as indicated in their original publications), as shown in Supplementary Table S1.

3 Results

3.1 Main Analyses with All Traits

For each of the 22 traits as the target trait, we first calculated its correlations with the other 21 traits, and excluded those with correlations larger than 0.05. After that, we generated the weight for each SNP using a given power index value (x), then calculated its weighted p-value. If the weighted p-value of a SNP was less than 5×10^{-8} , it was claimed to be statistically significantly associated with the trait. We also mapped each significant *SNP* into one of 1,703 LD independent loci. Then we checked the reproducibility using validation data for some traits.

Table 2 shows the results for the six traits with (partially overlapping but) much larger validation samples. We chose the power index x that gave the most significant SNPs/loci. It is clear that, for each trait, the new method identified a much larger number of significant SNPs and that of independent loci, most of which were also confirmed in the larger validation data. It is noted that for the four lipids, since the validation GWAS was exome-focused (while most of the significant SNPs from the original GWAS were from non-coding regions), there were relatively fewer SNPs validated, but the numbers of the validated loci remained large. For schizophrenia (SCZ), it is striking that the number of significant SNPs was doubled by the new method, the overwhelming majority of which (and loci) were also validated.

As a reviewer pointed out, for SCZ in Table 2, the weighted analysis identified a much larger number of significant SNPs but only two more loci than that from the unweighted analysis. Figure 1 shows the numbers of significant SNPs in each of 16 loci discovered by the unweighted or weighted testing. It can be seen that the major differences came from three or four loci commonly identified by both methods, including a locus near the major histocompatibility complex (MHC) region on chromosome 6.

For the other 16 traits, the results are listed in Supplementary Table S3. Among the 16 traits, for seven, there were no changes with respect to both the numbers of significant SNPs and independent loci by both methods; for four traits, the new method identified more significant SNPs but the same numbers of independent loci; for other five traits, the new method identified more significant SNPs and 2–3 additional independent loci.

For traits SCZ, HDL and AD, we also searched and listed the leading SNPs in each significant independent locus, along with the corresponding leading SNP in the validation data or in the GWAS Catalog (Supplementary Tables S6–S8). If a leading SNP identified by our method was also significant in the Gold Standard (i.e. validation) data, we indicated its source with “Gold”; otherwise, we searched on the GWAS Catalog, and listed the SNP closest to it that was also reported to be related to the trait with its source indicated as “GC”. We also showed the physical distance and LD r^2 between the leading SNP and the reported

SNP. Almost all the leading SNPs identified in the weighted analysis were close to and/or in high LD with a significant SNP in the validation data or GWAS Catalog. Two exceptions were the following. For HDL, a leading SNP, rs13207673 on chromosome 6, with the un-weighted and weighted p-values of 9.3×10^{-7} and 2.5×10^{-8} respectively, was not close to or not in high LD with any known HDL-associated SNPs in the GWAS Catalog. For AD, the leading SNP rs17878252 on chromosome 19, with the un-weighted and weighted p-values of 7.23×10^{-8} and 3.05×10^{-8} respectively, was only 7686 bp away from the known AD-associated SNP rs76320948, but based on 270 Cognitive Normal individuals in the ADNI data, their LD r^2 was only 0.11.

Figures 2 and 3 show that for SCZ and HDL, how the numbers of significant SNPs or loci would change with the power index x in the weighted analysis, as compared to the standard **un-weighted** analysis; for other traits the figures (Figures S2 to S23) are in the Supplementary Materials. In general, the pattern was that, as the power index x increased, the numbers of the significant SNPs and loci would increase, then reach their peaks before decreasing. This observed phenomenon is not really surprising: it is expected that suitably borrowing information from other GWAS data would boost statistical power due to pleiotropy or other shared genetic mechanisms across the traits.

Figure 4 shows that for SCZ (with the power index $x = 4$), how using different subsets of other (independent) traits/datasets would affect the results. For SCZ, 19 traits/datasets (excluding traits SCZ, BIP and MDD) were used to generate weights in the weighted analysis. For each of the 19 traits, we could either use or not use it to generate weights, leading to $2^{19} = 524288$ possible subsets of the data to be used. For illustration (while controlling the number of subsets to be tried at a manageable level), we showed the results of using 0 (i.e. unweighting), 1, 2, 3, 16, 17, 18 and all 19 datasets/traits to generate weights; for each of the above positive number, we tried all its possible combinations. For example, if we used 2 datasets/traits, there would be $\binom{19}{2} = 171$ combinations, each of which was shown as a (black) point in Figure 4. For each number of the subsets used, we also showed the average number (of significant SNPs or loci) across all possible subsets. Figure 5 shows the similar results for HDL (with the power index $x = 4$). The general trend is the increasing average numbers of significant SNPs and loci as more datasets being used.

Figures 6–7 show the corresponding results for SCZ and HDL respectively as each GWAS trait/dataset was added sequentially into weighted analysis to maximize the number of the significant SNPs identified. It is noted that as more datasets were added, the numbers of significant SNPs or loci identified in the discovery or validation sample might go up or go down, but overall the maximum numbers, especially those of significant loci, were reached with the use of multiple datasets, demonstrating the informativeness of multiple GWAS datasets.

Finally, we show the results of combining multiple weighted analyses with multiple values of power index x . As shown in Table 3, compared to the original/unweighted analysis (Table 2), the combined weighted analysis yielded more significant SNPs and loci for SCZ and all four lipid traits, but not for AD. Since larger values of x were less informative in forming the

weights (as shown in Figures 2 and 3), using the smaller set of x values of $\{1, 2, \dots, 5\}$ was often more powerful than using the larger set of $\{1, 2, \dots, 10\}$. The benefit of estimating the effective number of independent tests was clear with $M_e = 2.3$ and 1.7, much smaller than 10 and 5 (correlated) tests corresponding to the two sets of the weights (or x values) respectively. In Tables S9 to S16 in the Supplementary Materials, we also show similar results for other traits and the following two observations. First, we could combine the unweighted test with the weighted test, which however did not show power gains over combining only the weighted analyses. Second, when calculating the correlation matrix of the p-values, instead of using all the SNPs, we used a subset of randomly selecting 100 SNPs from each independent locus before taking them together; the results of using the two sets of the SNPs were essentially the same.

3.2 Secondary Analyses with Only Genetically Uncorrelated Traits

After excluding datasets with possibly overlapping samples and genetically correlated traits, we performed the weighted testing for all 22 traits as in the main analyses. Similar to Table 2, Table 4 shows the improved power of weighted testing for the six traits with validation samples. For the other 16 traits, the results are listed in Supplementary Table S18. For all 22 traits, we show how the numbers of significant SNPs/loci would change with the power index x in Supplementary Figures S25 to S46. Similar to Table 3, for the six traits with validation samples, Table 5 again shows the improved results by weighted testing after combining multiple weighted analyses with multiple values of power index x . Supplementary Tables S19 to S26 list all the results for other set of x and other traits. Figures 8–9 show the results for SCZ and HDL respectively as each genetically uncorrelated GWAS trait/dataset was added one by one to maximize the number of the significant SNPs identified. It is clear that the maximum numbers were achieved with the use of and thus information gain from multiple datasets. In summary, with only genetically uncorrelated traits, we reached the same conclusion of improved performance by weighted testing as before.

4 Conclusions and Discussion

We conclude that the weighted analysis of a GWAS could gain substantial power by integrating it with other published GWAS summary data, while the type I error rate is guaranteed to be satisfactorily controlled by its statistical theory. This approach takes advantage of existing GWAS data, thus is cost-effective. We emphasize that our approach does not necessarily require the use of other GWAS data with related traits; our approach leverages not only genetic pleiotropy (Andreassen et al., 2013a; Chung, Yang, Li, Gelernter, & Zhao, 2014), but also other shared genetic mechanisms of trait-associated SNPs, e.g. in transcriptional regulation as expression QTL. As shown by our secondary analyses, using the GWAS data of traits genetically *uncorrelated* with the trait of a new GWAS could still boost statistical power for the latter. There are several possible reasons. First, due to the limitation of measuring only a linear relationship of the causal variant effect sizes between two traits by a genetic correlation, genetically uncorrelated traits may still share overlapping causal variants (Frei et al., 2019). Second, given the well hypothesized functional, especially regulatory, roles of GWAS loci, coupled with the emerging theory of the omnigenic model

for GWAS (Boyle et al., 2017), it is reasonable to assume that many GWAS loci for some seemingly unrelated GWAS traits may turn out to be shared. In anticipation of more and more published GWAS data of various traits, we expect that there will be higher and higher power gains by our proposed approach.

A major technical challenge in our approach is the requirement for using other independent GWAS data to control the Type I error rate satisfactorily, as imposed by the theory of weighted hypothesis testing (Genovese, Roeder, & Wasserman, 2006); relaxing such an assumption will significantly facilitate the easy use and expand the applicability of the approach, though it is unclear how to do so. Second, we treat all the prior GWAS traits or datasets equally a priori; due to varying genetic correlations among the traits, it is expected to be more powerful to weight the traits/datasets differently. However, such weights on the traits, e.g. based on estimated genetic correlations, cannot be derived from the same set of the GWAS datasets used for the weighted analysis; otherwise, it would violate the required assumption of the independence between the test statistics and the weights imposed by the weighted hypothesis testing as discussed above. We also note that such an analysis is related to the pleiotropy-informed approach to improving power using independent GWAS data of related traits (Andreassen et al., 2013a; Andreassen et al., 2013b; Chung, Yang, Li, Gelernter, & Zhao, 2014), which however is based on stratified analysis controlling the false discovery rate (FDR), instead of weighted analysis controlling the family-wise error rate as proposed here. Third, we propose using the Z-statistic to represent each SNP's possible functional role and thus constructing its weight, one may explore the use of the (estimated) effect size for each SNP, its MAF and/or its LD score (Speed et al., 2017). Finally, it would be worthwhile to combine our proposed approach with a more popular line of research in integrating GWAS with functional annotations (and possibly pleiotropy) (Chung, Yang, Li, Gelernter, & Zhao, 2014). These are interesting topics to be explored in the future.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank the reviewers for many helpful comments and suggestions. This work was supported by NIH grants R01AG065636, R21AG057038, R01HL116720, R01GM113250 and R01GM126002, and by the Minnesota Supercomputing Institute at the University of Minnesota.

References

- [1]. Andreassen OA, Djurovic S, Thompson WK, Schork AJ, Kendler KS, O'Donovan MC, ... Dale OM (2013a). Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. *The American Journal of Human Genetics*, 92(2), 197–209. 10.1016/j.ajhg.2013.01.001 [PubMed: 23375658]
- [2]. Andreassen OA, Thompson WK, Schork AJ, Ripke S, Mattingsdal M, Kelsoe JR, ... Dale OM (2013b). Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate. *PLoS genetics*, 9(4). 10.1371/journal.pgen.1003455

- [3]. Berisa T, & Pickrell J (2016). Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics*, 32(2), 283–285. 10.1093/bioinformatics/btv546 [PubMed: 26395773]
- [4]. Boyle EA, Li YI, & Pritchard JK (2017). An expanded view of complex traits: from polygenic to omnigenic. *Cell*, 169(7), 1177–1186. 10.1016/j.cell.2017.05.038 [PubMed: 28622505]
- [5]. Bulik-Sullivan B, Finucane H, Anttila V, Gusev A, Day F, Loh P, ... Neale B (2015). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, 47(11), 1236–1241. 10.1038/ng.3406 [PubMed: 26414676]
- [6]. Chung D, Yang C, Li C, Gelernter J, & Zhao H (2014). GPA: a statistical approach to prioritizing GWAS results by integrating pleiotropy and annotation. *PLoS genetics*, 10(11). 10.1371/journal.pgen.1004787
- [7]. Cross-Disorder Group of the Psychiatric Genomics Consortium. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *The Lancet*, 381(9875), 1371–1379. 10.1016/S0140-6736(12)62129-1
- [8]. Day FR, Thompson DJ, Helgason H, Chasman DI, Finucane H, Sulem P, ... Perry JRB (2017). Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nature genetics*, 49(6), 834–841. 10.1038/ng.3841 [PubMed: 28436984]
- [9]. Darnell G, Duong D, Han B, & Eskin E (2012). Incorporating prior information into association studies. *Bioinformatics*, 28(12), 1147–1153. 10.1093/bioinformatics/bts235 [PubMed: 22689754]
- [10]. Eskin E (2008). Increasing power in association studies by using linkage disequilibrium structure and molecular function as prior information. *Genome Research*, 18(4), 653–660. 10.1101/gr.072785.107 [PubMed: 18353808]
- [11]. Finucane H, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P, ... Price A (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics*, 47(11), 1228–1235. 10.1038/ng.3404 [PubMed: 26414678]
- [12]. Frei O, Holland D, Smeland OB, Shadrin AA, Fan CC, Maeland S, ... Andreassen OA (2019). Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. *Nature communications*, 10(1), 1–11. 10.1038/s41467-019-10310-0
- [13]. Furberg H, Kim Y, Dackor J, Boerwinkle E, Franceschini N, Ardisson D, ... Mauri F (2010). Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature genetics*, 42(5), 441 10.1038/ng.571 [PubMed: 20418890]
- [14]. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, ... Im HK (2015). A gene-based association method for mapping traits using reference transcriptome data. *Nature genetics*, 47(9), 1091 10.1038/ng.3367 [PubMed: 26258848]
- [15]. Genovese C, Roeder K, & Wasserman L (2006). False discovery control with p-value weighting. *Biometrika*, 93(3), 509–524. 10.1093/biomet/93.3.509
- [16]. Gui J, Tosteson T, & Borsuk M (2012). Weighted multiple testing procedures for genomic studies. *BioData Mining*, 5(1), 4 10.1186/1756-0381-5-4 [PubMed: 22676256]
- [17]. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, ... Pasaniuc B (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nature genetics*, 48(3), 245 10.1038/ng.3506 [PubMed: 26854917]
- [18]. Ho YY, Baechler EC, Ortmann W, Behrens TW, Graham RR, Bhangale TR, & Pan W (2014). Using gene expression to improve the power of genome-wide association analysis. *Human heredity*, 78(2), 94–103. 10.1159/000362837 [PubMed: 25096029]
- [19]. Hsu L, Jiao S, Dai J, Hutter C, Peters U, & Kooperberg C (2012). Powerful cocktail methods for detecting genome-wide gene-environment interaction. *Genetic Epidemiology*, 36(3), 183–194. 10.1002/gepi.21610 [PubMed: 22714933]
- [20]. Ignatiadis N, Klaus B, Zaugg JB, & Huber W (2016). Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nature methods*, 13(7), 577 10.1038/nmeth.3885 [PubMed: 27240256]
- [21]. Ionita-Laza I, McQueen M, Laird N, & Lange C (2007). Genomewide Weighted Hypothesis Testing in Family-Based Association Studies, with an Application to a 100K Scan. *The American Journal of Human Genetics*, 81(3), 607–614. 10.1086/519748 [PubMed: 17701906]

- [22]. Iotchkova V, Ritchie GR, Geijs M, Morganella S, Min JL, Walter K, ... Soranzo N (2019). GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nature genetics*, 51(2), 343–353. 10.1038/s41588-018-0322-6 [PubMed: 30692680]
- [23]. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, ... Posthuma D (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature genetics*, 51(3), 404–413. 10.1038/s41588-018-0311-9 [PubMed: 30617256]
- [24]. Kang G, Ye K, Liu N, Allison DB, & Gao G (2009). Weighted multiple hypothesis testing procedures. *Statistical applications in genetics and molecular biology*, 8(1), 1–22. 10.2202/1544-6115.1437
- [25]. Kichaev G, Bhatia G, Loh PR, Gazal S, Burch K, Freund MK, ... Price AL (2019). Leveraging polygenic functional enrichment to improve GWAS power. *The American Journal of Human Genetics*, 104(1), 65–75. 10.1016/j.ajhg.2018.11.008 [PubMed: 30595370]
- [26]. Kim J, Bai Y, & Pan W (2015). An adaptive association test for multiple phenotypes with GWAS summary statistics. *Genetic epidemiology*, 39(8), 651–663. 10.1002/gepi.21931 [PubMed: 26493956]
- [27]. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, ... Amouyel P (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature genetics*, 45(12), 1452 10.1038/ng.2802 [PubMed: 24162737]
- [28]. Li MX, Yeung JM, Cherny SS, & Sham PC (2012). Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Human genetics*, 131(5), 747–756. 10.1007/s00439-011-1118-2 [PubMed: 22143225]
- [29]. Lin DY, & Sullivan PF (2009). Meta-analysis of genome-wide association studies with overlapping subjects. *The American Journal of Human Genetics*, 85(6), 862–872. 10.1016/j.ajhg.2009.11.001 [PubMed: 20004761]
- [30]. Lu X, Peloso GM, Liu DJ, Wu Y, Zhang H, Zhou W, ... Willer CJ (2017). Exome chip meta-analysis identifies novel loci and East Asian-specific coding variants that contribute to lipid levels and coronary artery disease. *Nature genetics*, 49(12), 1722 10.1038/ng.3978 [PubMed: 29083407]
- [31]. Lu Q, Yao X, Hu Y, & Zhao H (2016a). GenoWAP: GWAS signal prioritization through integrated analysis of genomic functional annotation. *Bioinformatics*, 32(4), 542–548. 10.1093/bioinformatics/btv610 [PubMed: 26504140]
- [32]. Lu Q, Powles RL, Wang Q, He BJ, & Zhao H (2016b). Integrative tissue-specific functional annotations in the human genome provide novel insights on many complex traits and improve signal prioritization in genome wide association studies. *PLoS genetics*, 12(4). 10.1371/journal.pgen.1005947
- [33]. Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, ... Langenberg C (2012). A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nature genetics*, 44(6), 659–669. 10.1038/ng.2274 [PubMed: 22581228]
- [34]. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, ... Stamatoyannopoulos JA (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science*, 337(6099), 1190–1195. 10.1126/science.1222794 [PubMed: 22955828]
- [35]. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, & Cox NJ (2010). Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS genetics*, 6(4). 10.1371/journal.pgen.1000888
- [36]. Okbay A, Baselmans BM, De Neve JE, Turley P, Nivard MG, Fontana MA, ... Cesarini D (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature genetics*, 48(6), 624–633. 10.1038/ng.3552 [PubMed: 27089181]
- [37]. Pan W, Kim J, Zhang Y, Shen X, & Wei P (2014). A powerful and adaptive association test for rare variants. *Genetics*, 197(4), 1081–1095. 10.1534/genetics.114.165035 [PubMed: 24831820]

- [38]. Pasaniuc B, Zaitlen N, Shi H, Bhatia G, Gusev A, Pickrell J, ... Price AL (2014). Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. *Bioinformatics*, 30(20), 2906–2914. 10.1093/bioinformatics/btu416 [PubMed: 24990607]
- [39]. Pickrell JK (2014). Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *The American Journal of Human Genetics*, 94(4), 559–573. 10.1016/j.ajhg.2014.03.004 [PubMed: 24702953]
- [40]. Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA, ... Gejman PV (2011). Genome-wide association study identifies five new schizophrenia loci. *Nature genetics*, 43(10), 969. 10.1038/ng.940 [PubMed: 21926974]
- [41]. Ripke S, Neale BM, Corvin A, Walters JT, Farh KH, Holmans PA, ... O'Donovan MC (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, 511(7510), 421–427. 10.1038/nature13595 [PubMed: 25056061]
- [42]. Roeder K, Bacanu SA, Wasserman L, & Devlin B (2006). Using linkage genome scans to improve power of association in genome scans. *The American Journal of Human Genetics*, 78(2), 243–252. 10.1086/500026 [PubMed: 16400608]
- [43]. Roeder K, Devlin B, & Wasserman L (2007). Improving power in genome-wide association studies: weights tip the scale. *Genetic Epidemiology*, 31(7), 741–747. 10.1002/gepi.20237 [PubMed: 17549760]
- [44]. Roeder K, & Wasserman L (2009). Genome-wide significance levels and weighted hypothesis testing. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 24(4), 398. 10.1214/09-STS289 [PubMed: 20711421]
- [45]. Schmidt EM, Zhang J, Zhou W, Chen J, Mohlke KL, Chen YE, & Willer CJ (2015). GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. *Bioinformatics*, 31(16), 2601–2606. 10.1093/bioinformatics/btv201 [PubMed: 25886982]
- [46]. Schork AJ, Thompson WK, Pham P, Torkamani A, Roddey JC, Sullivan PF, ... Dale AM (2013). All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS genetics*, 9(4). 10.1371/journal.pgen.1003449
- [47]. Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, ... the CARDIoGRAM Consortium. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature genetics*, 43(4), 333–338. 10.1038/ng.784 [PubMed: 21378990]
- [48]. Speed D, Cai N, the UCLEB Consortium, Johnson MR, Nejentsev S, & Balding DJ (2017). Reevaluation of SNP heritability in complex human traits. *Nature genetics*, 49(7), 986–992. 10.1038/ng.3865 [PubMed: 28530675]
- [49]. Sveinbjornsson G, Albrechtsen A, Zink F, Gudjonsson SA, Oddson A, Måsson G, ... Stefansson K (2016). Weighting sequence variants based on their annotation increases power of whole-genome association studies. *Nature genetics*, 48(3), 314. 10.1038/ng.3507 [PubMed: 26854916]
- [50]. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, ... Kathiresan S (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, 466(7307), 707–713. 10.1038/nature09270 [PubMed: 20686565]
- [51]. The 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68–74. 10.1038/nature15393 [PubMed: 26432245]
- [52]. Trynka G, Westra HJ, Slowikowski K, Hu X, Xu H, Stranger BE, ... Raychaudhuri S (2015). Disentangling the effects of colocating genomic annotations to functionally prioritize non-coding variants within complex-trait loci. *The American Journal of Human Genetics*, 97(1), 139–152. 10.1016/j.ajhg.2015.05.016 [PubMed: 26140449]
- [53]. van der Sluis S, Posthuma D, & Dolan CV (2013). TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies. *PLoS genetics*, 9(1). 10.1371/journal.pgen.1003235
- [54]. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, & Yang J (2017). 10 years of GWAS discovery: biology, function, and translation. *The American Journal of Human Genetics*, 101(1), 5–22. 10.1016/j.ajhg.2017.06.005 [PubMed: 28686856]

- [55]. Watanabe K, Stringer S, Frei O, Mirkov MU, de Leeuw C, Polderman TJ, ... Posthuma D (2019). A global overview of pleiotropy and genetic architecture in complex traits. *Nature genetics*, 51(9), 1339–1348. 10.1038/s41588-019-0481-0 [PubMed: 31427789]
- [56]. Wheeler E, Leong A, Liu CT, Hivert MF, Strawbridge RJ, Podmore C, ... Meigs JB (2017). Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLoS medicine*, 14(9). 10.1371/journal.pmed.1002383
- [57]. Wu C, & Pan W (2019). Integration of methylation QTL and enhancer-target gene maps with schizophrenia GWAS summary results identifies novel genes. *Bioinformatics*, 35(19), 3576–3583. 10.1093/bioinformatics/btz161 [PubMed: 30850848]
- [58]. Xu Z, Wu C, Wei P, & Pan W (2017). A powerful framework for integrating eQTL and GWAS summary data. *Genetics*, 207(3), 893–902. 10.1534/genetics.117.300270 [PubMed: 28893853]
- [59]. Zheng HF, Forgetta V, Hsu YH, Estrada K, Rosello-Diez A, Leo PJ, ... Richards JB (2015). Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. *Nature*, 526(7571), 112–117. 10.1038/nature14878 [PubMed: 26367794]

Figure 1:
The numbers of the significant SNPs in each locus identified by the two methods for SCZ.

Figure 2:
The numbers of significant SNPs and loci identified by the un-weighted analysis and weighted analysis (left panels) and those confirmed by the validation data (right panels), as a function of the value of the power index x for SCZ.

Figure 3:

The numbers of significant SNPs and loci identified by the un-weighted analysis and weighted analysis (left panels) and those confirmed by the validation data (right panels), as a function of the value of the power index x for HDL.

Figure 4:

The numbers of significant SNPs and loci identified (left panels) and validated (right panels) by the weighted analysis as a function of the number of the GWAS traits/datasets being used for SCZ (with power index $\alpha = 4$). The average numbers are plotted as larger/red triangles.

Figure 5:

The numbers of significant SNPs and loci identified (left panels) and validated (right panels) by the weighted analysis as a function of the number of the GWAS traits/datasets being used for HDL (with power index $\alpha=4$). The average numbers are plotted as larger/red triangles.

Figure 6:

The numbers of significant SNPs and loci identified (left panels) and validated (right panels) by the weighted analysis as each GWAS trait/dataset was sequentially added for SCZ (with power index $x = 4$).

Figure 7:

The numbers of significant SNPs and loci identified (left panels) and validated (right panels) by the weighted analysis as each GWAS trait/dataset was sequentially added for HDL (with power index $x = 4$).

Figure 8:

The numbers of significant SNPs and loci identified (left panels) and validated (right panels) by the weighted analysis as each genetically uncorrelated GWAS trait/dataset was added for SCZ (with power index $\alpha = 4$).

Figure 9:

The numbers of significant SNPs and loci identified (left panels) and validated (right panels) by the weighted analysis as each genetically uncorrelated GWAS trait/dataset was added for HDL (with power index $x = 1$).

Table 1:

The sample sizes of the discovery and validation data for 6 traits.

Trait	Discovery data	Validation data
SCZ	21,856	82,315
HDL	99,900	~347,000
LDL	95,454	~347,000
TC	100,184	~347,000
TG	96,598	~347,000
AD	54,162	455,258

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2:

Results for 6 traits: the significant numbers of the SNPs and loci identified (with those confirmed by the validation samples in parentheses) by the original/unweighted test and our proposed weighted test (with a single selected value of power index x).

Trait	Original		Weighted		
	# SNPs	# Loci	# SNPs	# Loci	Power Index x
SCZ	818 (705)	13 (13)	1585 (1423)	15 (15)	4
HDL	5405 (189)	49 (45)	5563 (200)	52 (50)	4
LDL	4663 (177)	41 (39)	4763 (181)	44 (41)	3
TC	6079 (273)	57 (54)	6374 (282)	59 (55)	2
TG	4226 (226)	34 (32)	4462 (228)	38 (35)	4
AD	1230 (944)	14 (13)	1265 (957)	14 (13)	1

Table 3:

Results for 6 traits from the weighted analyses after combining over each of two sets of the weights with the corresponding values of power index x . The significant numbers of the SNPs and loci identified (with those confirmed by the validation samples in parentheses) are shown.

Trait	$x \in \{1,2, \dots, 10\}$			$x \in \{1,2, \dots, 5\}$		
	M_e	# SNPs	# Loci	M_e	# SNPs	# Loci
SCZ	2.273	1475 (1317)	14 (14)	1.646	1567 (1389)	16 (16)
HDL	2.287	5642 (199)	51 (49)	1.675	5673 (201)	52 (50)
LDL	2.280	4734 (180)	42 (40)	1.659	4787 (180)	42 (40)
TC	2.285	6391 (288)	57 (54)	1.666	6533 (292)	57 (54)
TG	2.284	4400 (225)	38 (36)	1.663	4533 (229)	38 (35)
AD	2.275	1192 (919)	12 (11)	1.651	1239 (929)	14 (13)

Table 4:

Results for the 6 traits after excluding GWAS data of genetically correlated traits: the significant numbers of the SNPs and loci identified (with those confirmed by the validation samples in parentheses) by the original/unweighted test and our proposed weighted test (with a single selected value of power index x).

Trait	Original		Weighted		
	# SNPs	# Loci	# SNPs	# Loci	Power Index x
SCZ	818 (705)	13 (13)	1510 (1353)	15 (15)	4
HDL	5405 (189)	49 (45)	5445 (189)	48 (45)	1
LDL	4663 (177)	41 (39)	4693 (180)	43 (41)	2
TC	6079 (273)	57 (54)	6358 (280)	59 (55)	2
TG	4226 (226)	34 (32)	4397 (230)	37 (33)	3
AD	1230 (944)	14 (13)	1265 (957)	14 (13)	1

Table 5:

Results for the 6 traits from the weighted analyses after combining over each of two sets of the weights with the corresponding values of power index x after excluding GWAS data of genetically correlated traits. The significant numbers of the SNPs and loci identified (with those confirmed by the validation samples in parentheses) are shown.

Trait	$x \in \{1,2, \dots, 10\}$			$x \in \{1,2, \dots, 5\}$		
	M_e	# SNPs	# Loci	M_e	# SNPs	# Loci
SCZ	2.282	1398(1262)	14(14)	1.664	1492(1329)	16(16)
HDL	2.29	5353(191)	48(46)	1.682	5488(192)	48(46)
LDL	2.282	4704(180)	42(40)	1.664	4761(180)	42(40)
TC	2.288	6351(287)	57(54)	1.671	6512(292)	57(54)
TG	2.285	4385(224)	37(35)	1.669	4503(228)	38(35)
AD	2.275	1192 (919)	12 (11)	1.651	1239 (929)	14 (13)