Web-based Supplementary Materials

FOR

Methods for Meta-analysis of Multiple Traits using GWAS Summary Statistics

BY

Debashree Ray, Michael Boehnke

Department of Biostatistics and Center for Statistical Genetics,

University of Michigan, Ann Arbor, MI, USA

Supplementary S1

Details of asymptotic metaUSAT p-value computation

Without loss of generality let us assume a single study with K traits. Recall the meta-MANOVA and the SSU test statistics:

$$T_{\text{metaMANOVA}} = \boldsymbol{Z}'\hat{\boldsymbol{R}}^{-1}\boldsymbol{Z} \overset{a}{\sim} \chi^2_K$$

$$T_{\text{SSU}} = \boldsymbol{Z}'\boldsymbol{Z} \overset{approx}{\sim} a\chi^2_d + b$$

For metaUSAT, we first consider the weighted statistic $T_\omega = \omega T_{\text{metaMANOVA}} + (1-\omega)T_{\text{SSU}}$, where $\omega \in [0,1]$ is the weight. Both metaMANOVA and SSU are special cases of the class of statistics T_ω. Under H_0, for a given weight ω, T_ω is approximately a linear combination of chi-squared distributions. The computation of p-value p_ω of the test statistic T_ω does not require independence of the statistics $T_{\text{metaMANOVA}}$ and T_{SSU}. A detailed explanation of the determination of p_ω is provided below.

Observe that, for a given weight ω, one can write

$$T_\omega = \omega T_{\text{metaMANOVA}} + (1-\omega)T_{\text{SSU}} = \boldsymbol{Z}'\left(\omega\hat{\boldsymbol{R}}^{-1} + (1-\omega)\boldsymbol{I}_K\right)\boldsymbol{Z}$$

where $\boldsymbol{I}_K$ is the identity matrix of order K and the vector of univariate scores $\boldsymbol{Z}$ has mean $\boldsymbol{0}$, estimated variance $\text{Cov}(\boldsymbol{Z}) = \hat{\boldsymbol{R}}$, and has an asymptotic K-variate normal

distribution. Denote $\boldsymbol{A}_\omega = \omega\hat{\boldsymbol{R}}^{-1} + (1-\omega)\boldsymbol{I}_K$. Let $\boldsymbol{P}$ be a $K \times K$ orthonormal matrix that converts $\boldsymbol{B}_\omega = \text{Cov}(\boldsymbol{Z})^{1/2}\boldsymbol{A}_\omega\text{Cov}(\boldsymbol{Z})^{1/2} = \omega\boldsymbol{I}_K + (1-\omega)\text{Cov}(\boldsymbol{Z})$ to the diagonal form $\boldsymbol{\Gamma}_\omega = \text{diag}(\lambda_1, ...\lambda_K)$, where $\lambda_1 \geq 0, ..., \lambda_K \geq 0$ (see Liu et al., 2009; Ray et al., 2016, for example). Essentially, $\lambda_1, ..., \lambda_K$ are the non-negative eigen values of $\boldsymbol{B}_\omega$, i.e., $\boldsymbol{\Gamma}_\omega = \boldsymbol{P}\boldsymbol{B}_\omega\boldsymbol{P}'$. The weighted statistic T_ω can, then, be expressed as a non-negative quadratic form:

$$T_\omega = \boldsymbol{Z}'\boldsymbol{A}_\omega\boldsymbol{Z} = \boldsymbol{V}'\boldsymbol{\Gamma}_\omega\boldsymbol{V} = \sum_{j=1}^{K}\lambda_j\chi^2_{h_j}(\delta_j) \tag{S1}$$

where $\boldsymbol{V} = \boldsymbol{P}\text{Cov}(\boldsymbol{Z})^{-1/2}\boldsymbol{Z} \overset{a}{\sim} N(\boldsymbol{0}, \boldsymbol{I}_K)$, and $h_j = 1$, $\delta_j = 0$ for all $j = 1, 2, ..., K$. For a given $\omega \in [0, 1]$, the p-value p_ω of the statistic T_ω can, thus, be calculated numerically by Davies (1980) algorithm or by moment-matching using Liu et al. (2009) algorithm. Both of these algorithms are implemented in the `R` package `CompQuadForm` (Duchesne and de Micheaux, 2010). Particularly, Liu et al. (2009) approximates the upper tail probability as

$$p_\omega = \text{P}\left(T_\omega > t_\omega\right) \approx \text{P}\left(\chi^2_l(\delta) > t^*_\omega\sigma_\chi + \mu_\chi\right) \tag{S2}$$

where t_ω is the observed value of T_ω statistic, $t^*_\omega = (t_\omega - \text{E}(T_\omega))/\sqrt{\text{Var}(T_\omega)}$, $\mu_\chi = \text{E}\left(\chi^2_l(\delta)\right) = l + \delta$, $\sigma_\chi = \sqrt{\text{Var}\left(\chi^2_l(\delta)\right)} = \sqrt{2(l+2\delta)}$. The parameters δ and l are chosen such that the skewness of T_ω and $\chi^2_l(\delta)$ are same and the difference between the kurtoses of T_ω and $\chi^2_l(\delta)$ is minimized. Lee et al. (2012) modified the Liu et al.'s algorithm to match the kurtosis instead of skewness. We used Davies' algorithm implemented in `R` function `davies()` from `CompQuadForm`.

A priori the choice of weight ω is not known. We propose our unified test metaUSAT:

$$T_{\text{metaUSAT}} = \min_{0\leq\omega\leq1} p_\omega$$

Thus, metaUSAT is not exactly the best weighted combination of metaMANOVA and SSU but is the minimum of the p-values of the different weighted combinations. For

practical implementations of metaUSAT, we use a grid of 11 ω values: $\{\omega_1 = 0, \omega_2 = 0.1, ..., \omega_{10} = 0.9, \omega_{11} = 1\}$. A finer grid of ω values does not make any meaningful difference in metaUSAT's p-value.

To find the p-value of our metaUSAT test statistic, we need the null distribution, which does not have a closed-form. We propose an approximate p-value calculation using a one-dimensional numerical integration, similar to what Ray et al. (2016) proposed for USAT. This makes metaUSAT suitable for application on a GWAS scale. Observe that the p-value of statistic T_{metaUSAT} is

$$
\begin{aligned}
p_{\text{metaUSAT}} &= P(T_{\text{metaUSAT}} \leq t_{\text{metaUSAT}}) = 1 - P(T_{\text{metaUSAT}} \geq t_{\text{metaUSAT}}) \\
&= 1 - P\left(\min_{\omega} p_\omega \geq t_{\text{metaUSAT}}\right) = 1 - P\left(1 - \min_{\omega} p_\omega < 1 - t_{\text{metaUSAT}}\right) \\
&= 1 - P\left(\max_{\omega} (1 - p_\omega) < 1 - t_{\text{metaUSAT}}\right) \\
&= 1 - P\left(\{1 - p_{\omega_1} < 1 - t_{\text{metaUSAT}}\}, \ldots, \{1 - p_{\omega_{11}} < 1 - t_{\text{metaUSAT}}\}\right) \\
&= 1 - P\Big(\{(1 - p_{\omega_1})^{th} \text{ quantile} < (1 - t_{\text{metaUSAT}})^{th} \text{ quantile}\}, \ldots, \\
&\quad \{(1 - p_{\omega_{11}})^{th} \text{ quantile} < (1 - t_{\text{metaUSAT}})^{th} \text{ quantile}\}\Big) \\
&= 1 - P\left(T_{\omega_1} < q_{\min}(\omega_1), ..., T_{\omega_{11}} < q_{\min}(\omega_{11})\right) \\
&= 1 - P\left(T_{\text{SSU}} < \min_{\omega} \frac{q_{\min}(\omega) - \omega T_{\text{metaMANOVA}}}{1 - \omega}\right) \\
&= 1 - \int F_{T_{\text{SSU}}|T_{\text{metaMANOVA}}}\big(\delta_\omega(x)|x\big) f_{T_{\text{metaMANOVA}}}(x) dx
\end{aligned}
$$

where t_{metaUSAT} is the observed value of metaUSAT test statistic for a given dataset, $q_{\min}(\omega_b)$ is the $(1 - t_{\text{metaUSAT}})$-th percentile of the distribution of T_{ω_b} for a given $\omega = \omega_b$, $F_{T_{\text{SSU}}|T_{\text{metaMANOVA}}}(.|x)$ is the conditional cdf of SSU statistic given metaMANOVA statistic, $f_{T_{\text{metaMANOVA}}}(.)$ is the pdf of metaMANOVA test statistic, and $\delta_\omega(x) = \min_{\omega \in \{\omega_1, ..., \omega_{11}\}} \frac{q_{\min}(\omega) - \omega x}{1 - \omega}$. The evaluation of $q_{\min}(\omega)$ for a given ω can be done numerically using Liu et al. (2009)'s algorithm or Lee et al. (2012)'s modified Liu algorithm. Wu and Pankow (2016) showed that both these methods can poorly approximate tail probabilities when it comes to significance level of the order of 10^{-6}. So, Wu et al. (2016) proposed minimizing the standardized $(J-1)$-th and J-th moment differences. Wu et al. found that $J = 12$ can

accurately compute small tail probabilities under the null hypothesis while $J = 6$ is a suitable choice for approximating tail probabilities under the alternative. The cost of higher accuracy for these methods is time. To balance accuracy and computation time, we use a combination of two algorithms. When $t_{\text{metaUSAT}} > 10^{-4}$ (scenario where we expect relatively large tail probability), we use Liu et al.'s method while for the other, we use Wu et al.'s algorithm with $J = 6$.

Recall that T_{SSU} and $T_{\text{metaMANOVA}}$ are two quadratic forms (QF), which are not independently distributed. The exact joint distribution of T_{SSU} and $T_{\text{metaMANOVA}}$ is too complicated to compute (Khatri, 1980; Khatri et al., 1977). Our literature search did not yield any computationally feasible method for approximating the distribution $F_{T_{\text{SSU}}|T_{\text{metaMANOVA}}}(.|T_{\text{metaMANOVA}} = x)$ required to calculate p_{metaUSAT}. In such a scenario, a simple and straightforward approximation seems to be the assumption of independence and thereby we get the approximate p-value

$$p_{\text{metaUSAT}} \approx 1 - \int_0^\infty F_{T_{\text{SSU}}}\big(\delta_\omega(x)|x\big) f_{T_{\text{metaMANOVA}}}(x)dx$$

where $F_{T_{\text{SSU}}}(.)$ is the cdf of SSU test statistic. As demonstrated by metaUSAT type I error analysis, this approximation works well for low error levels. However, at very stringent error levels, this approximation, combined with numerical approximation involved in the afore-mentioned algorithms for evaluating integrals, can slightly inflate metaUSAT's type I error. In this context, it is worth noting that we have not assumed T_{SSU} and $T_{\text{metaMANOVA}}$ to be independent throughout. For example, the information on their dependence has been incorporated in the calculation of p_ω (p-value of weighted statistic T_ω). The independence assumption has been made only in the last step of metaUSAT p-value calculation.

Supplementary S2

Details of empirical metaUSAT p-value calculation

In our type I error analysis, we observe slightly inflated type I errors for all methods (Table 1) except S_{Hom}. We note that type I error of metaUSAT is worst at $\alpha = 5 \times 10^{-7}$. For real data, this can be a concern for SNPs that only metaUSAT detects as significant by a narrow margin with respect to the significance threshold. For example, in our analysis of the combined METSIM and T2D-GENES datasets, metaUSAT exclusively reports 4 significant SNPs (of which 3 are independent) that metaMANOVA fails to find (Table 4). To alleviate any concern that these metaUSAT findings may be a consequence of inflated type I error at the stringent genome-wide threshold, we describe here a Monte Carlo simulation based approach for estimating empirical p-values of metaUSAT.

Without loss of generality, we explain the method for a single study with K traits, which can be straightforwardly extended to multiple studies. Observe that the vector of univariate summary statistics $\boldsymbol{Z}$ has an asymptotic multivariate normal distribution under the null: $\boldsymbol{Z} \overset{a}{\sim} N_K(\boldsymbol{0}, \boldsymbol{R})$. Firstly, we draw B independent samples of vector $\boldsymbol{Z}$ under the null. Denote the b-th sample (a vector) as $\boldsymbol{Z}_b = (Z_{1,b}, \ldots, Z_{K,b})'$, $b = 1, 2, ..., B$. Note that the null distribution of $\boldsymbol{Z}$ is not variant-specific. Hence, we need to draw these B samples only once and use them for any number of variants/SNPs for a given study. If the sample sizes for the traits vary in the original dataset, we recommend weighting the summary statistics by sample size (as explained in the main text). Consequently, we need to pre-multiply and post-multiply the covariance matrix $\boldsymbol{R}$ by the diagonal weight matrix $\boldsymbol{W} = \text{diag}\{\sqrt{n_1}, ..., \sqrt{n_K}\}$, where n_k is the sample size for the k-th trait, $k = 1, 2, ..., K$. Let the weighted covariance matrix be $\boldsymbol{R}^w$, and the weighted samples be $\boldsymbol{Z}_1^w, ..., \boldsymbol{Z}_b^w, ..., \boldsymbol{Z}_B^w$.

Secondly, for the b-th sample, we calculate only the metaUSAT statistic: $T_{\text{metaUSAT},b} = \min_{\omega \in \{0, 0.1, ..., 0.9, 1\}} p_{\omega,b}$, where $p_{\omega,b}$ is the p-value of $T_{\omega,b} = \omega \boldsymbol{Z}_b^{w'} (\boldsymbol{R}^w)^{-1} \boldsymbol{Z}_b^w + (1-\omega) \boldsymbol{Z}_b^{w'} \boldsymbol{Z}_b^w$ for a given choice of $\omega \in \{0, 0.1, 0.2, ..., 0.8, 0.9, 1\}$. Finally, if the metaUSAT statistic for the test of association of a particular variant/SNP in the original dataset is T_{metaUSAT},

then the empirical p-value of metaUSAT for this variant/SNP is calculated as

$$p^{emp}_{\text{metaUSAT}} = \frac{1 + \sum_{b=1}^{B} I\big(|T_{\text{metaUSAT},b}| < |T_{\text{metaUSAT}}|\big)}{1 + B}$$

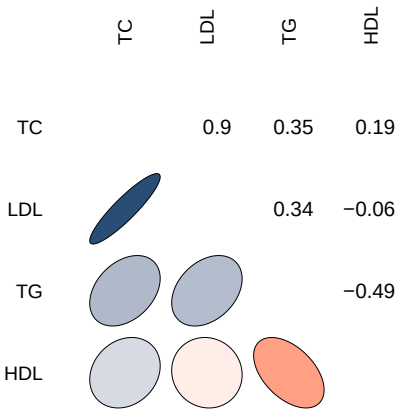
where $I(.)$ is the indicator function.

For computing empirical metaUSAT p-values for the 3 independent SNPs detected solely by metaUSAT from the combined METSIM and T2D-GENES studies (Table 4), we chose $B = 8.5 \times 10^9$ (the asymptotic metaUSAT p-values are $> 10^{-9}$). For the stringent genome-wide threshold 5×10^{-8}, we observe a p-value inflation (defined as the ratio of empirical p-value to asymptotic p-value) of the order of $1.5\times$ to $3\times$ − something we also observed for error level $\alpha = 5 \times 10^{-7}$ in our simulation study (Table 1). It looks like we need not calculate empirical metaUSAT p-values for all variants to alleviate concerns of inflated association signals; instead we can just focus on the handful of variants that have p-values just crossing the chosen threshold.

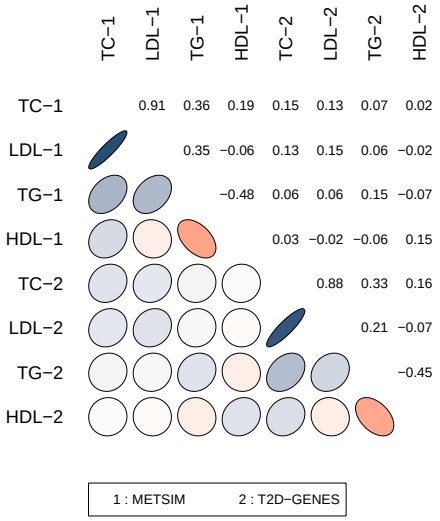
Supplementary S3

Simulation Experiments: Additional Details, Figures & Tables

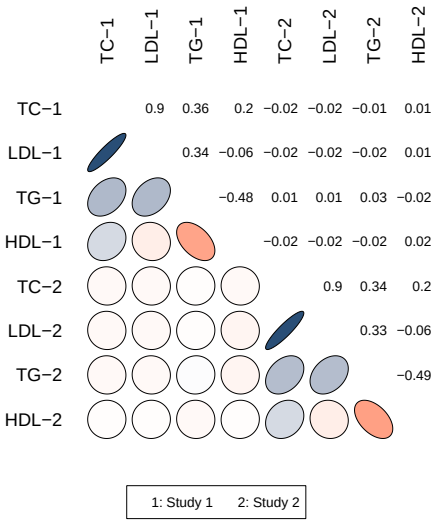
Calculation of empirical power based on corrected critical value. To compute corrected critical threshold for each method at significance level α, we generated 10^5 null replicates and obtained univariate summary statistics. Using these null summary statistics, we also obtain 10^5 null statistics for each of S_{Hom}, S_{Het}, minP, aSPU, metaMANOVA and metaUSAT methods. For a given method, these null statistics give the empirical null distribution of the test statistic. Specifically, for S_{Hom}, S_{Het} and metaMANOVA, the $(1-\alpha)$-th sample quantile is calculated from the 10^5 null statistics. Empirical power at level α is calculated as the proportion of non-null datasets with test statistics exceeding the $(1-\alpha)$-th sample quantile. On the other hand, for minP, aSPU and metaUSAT (methods with p-value type test statistics), the α-th sample quantile is calculated from the 10^5 null statistics. Empirical power at level α is, then, calculated as the proportion of non-null datasets with test statistics that could not exceed the α-th sample quantile.



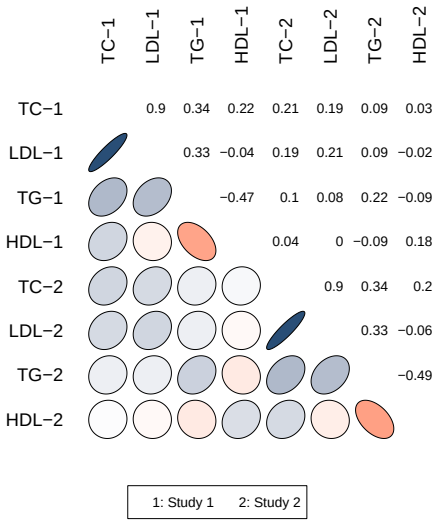
(a) Trait correlation structure: METSIM



(b) Trait correlation structure: METSIM+T2D-GENES



(c) Trait correlation structure: Simulation 2



(d) Trait correlation structure: Simulation 3

Figure S1: (a) The correlation matrix of the 4 lipid traits from METSIM Study. This correlation matrix $\boldsymbol{R}_{\text{metsim}}$ is used to simulate 4 correlated traits in Simulations 2 and 3. (b) Estimated correlation matrix of the lipid traits from the METSIM and the T2D-GENES studies. The non-zero correlations between studies are induced by overlapping samples. (c) The estimated correlation matrix for Simulation 2 with two independent studies. (d) The estimated correlation matrix for Simulation 3 (two studies with overlapping samples), which is similar to the correlation structure 1(b) of the METSIM and T2D-GENES studies.

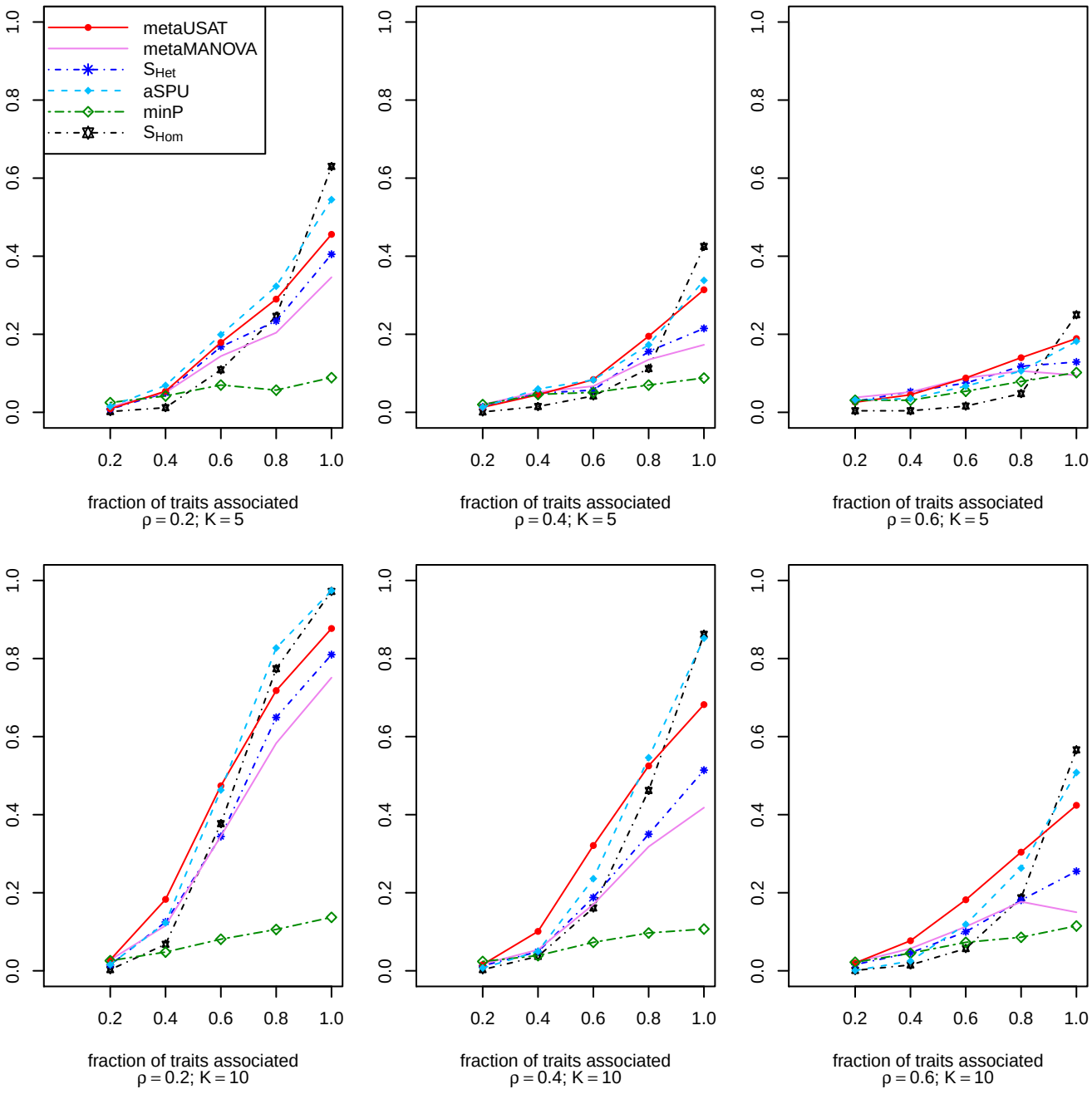


Figure S2: Simulation 1: Nominal power curves of S_{Hom}, S_{Het}, metaMANOVA, metaUSAT, minP and aSPU at significance level $\alpha = 10^{-4}$ when an AR1(ρ) correlation structure of the traits is considered. Power estimates are based on $1,000$ datasets with $1,000$ unrelated samples. Each sample has $K = 5$ or 10 traits with correlation parameter $\rho = 0.2$, 0.4 or 0.6.

Table S1: Simulation 2: Type I error estimates at various levels α. This table lists the type I error estimates divided by the significance level α. The ideal value for any cell is 1. Estimates are based on 10^7 null datasets. Each dataset consists of two independent studies having sample size $1,000$. Estimates for aSPU are not provided since it is computationally very intensive for 10^7 datasets.

Method	α				
	10^{-2}	10^{-3}	10^{-4}	10^{-5}	2.5×10^{-6}
S_{Hom}	1.02	1.03	1.08	1.19	1.48
S_{Het}	1.04	1.13	1.32	1.62	2.16
minP	1.04	1.10	1.15	1.37	1.44
metaMANOVA	1.03	1.08	1.14	1.30	1.32
metaUSAT	0.89	1.00	1.21	1.42	1.48

Table S2: Simulation 3: Type I error estimates at various levels α. This table lists the type I error estimates divided by the significance level α. The ideal value for any cell is 1. Estimates are based on 10^7 null datasets. Each dataset consists of two studies having sample size $1,000$, where 200 individuals are common to both studies. Estimates for aSPU are not provided since it is computationally very intensive for 10^7 datasets.

Method	α				
	10^{-2}	10^{-3}	10^{-4}	10^{-5}	2.5×10^{-6}
S_{Hom}	1.03	1.11	1.16	1.15	1.32
S_{Het}	1.09	1.22	1.49	1.71	1.64
minP	1.03	1.09	1.14	1.30	1.49
metaMANOVA	1.07	1.16	1.22	1.15	1.20
metaUSAT	0.96	1.11	1.40	1.58	2.04

Supplementary S4

Additional Simulation Experiment I: Misspecified $\boldsymbol{R}$

In this experiment, we aim to study the effects of misspecified estimates of $\boldsymbol{R}$ on the type I error behavior of all the summary statistic based tests discussed in the main manuscript (except aSPU). We believe that validity of test statistics directly incorporating $\boldsymbol{R}$ will be heavily affected while test statistics that indirectly incorporate $\boldsymbol{R}$ only through the null distribution will be much less affected. We assume 5 continuous traits for a single study (Simulation 1). In particular, each study has $n = 1,000$ unrelated individuals, each measured for $K = 5$ traits and a bi-allelic SNP with MAF 0.2 at Hardy-Weinberg Equilibrium. We use a multivariate normal model with exchangeable correlation $\boldsymbol{R}(\rho)$ with pairwise correlation $\rho \in \{0.2, 0.4, 0.6\}$. Since we are focusing on type I error behavior of tests, the genetic effects are 0 for all 5 traits. We simulate 10^7 such datasets/replicates.

To implement metaUSAT and the other tests, we need to estimate the trait correlation matrix $\boldsymbol{R}$. In order to study the effects of misspecified $\boldsymbol{R}$, we simulate an extreme scenario: 10^4 non-null datasets where the SNP is associated with all 5 traits. We assumed that $h\%$ of the total trait variance is explained by the SNP, where we varied $h \in \{0.5, 5, 10\}$. Using the univariate summary statistics for these 10^4 non-null SNPs, we calculated the sample covariance matrix $\hat{\boldsymbol{R}}$ and used it as the estimate for $\boldsymbol{R}$. This estimate is subsequently used in the test statistics for metaUSAT and the other tests to analyze the 10^7 null datasets described earlier. Table S3 shows ratio of type I error estimates from misspecified $\boldsymbol{R}$ estimated from SNPs with $h > 0$ to those estimated from null SNPs with $h = 0$ (the correct way of estimating $\boldsymbol{R}$).

Table S3 shows us that the extent to which misspecification of $\boldsymbol{R}$ affects the validity of summary-statistic-based multivariate tests depends on the strength of association of the non-null SNPs used to estimate $\boldsymbol{R}$ as well as on the test statistic. In particular, we observe that the type I error estimates for metaUSAT and minP are largely unaffected (when type I error estimates for non-zero h are compared against $h = 0$). metaMANOVA and S_{Het} have inflated type I errors, where the degree of inflation increases with the strength of association (of SNPs used to estimate $\boldsymbol{R}$). On the other hand, S_{Hom} is increasingly

conservative (i.e., deflated type I error) with increasing strength of association of the non-null SNPs used to estimate $\boldsymbol{R}$.

In this context, it is important to note that our conclusion is based on a limited simulation study, where we varied only a single parameter (the proportion of trait variance explained by the SNP for all traits). It is beyond the scope of this paper to explore the effect of misspecification of $\boldsymbol{R}$ in any more detail. For traits that are not highly polygenic, we expect most SNPs in a GWAS to have no effect on any trait under study, and hence the possibility of misspecification of $\boldsymbol{R}$ (and its ramification) seems negligible for variants that are not rare. For more polygenic traits, we expect validity of metaUSAT and minP to be mostly unaffected compared to other existing tests.

Additional Simulation Experiment II: Binary traits

Apart from continuous traits, we also simulated binary traits for two independent studies (Simulation 2) and for two studies with shared individuals (Simulation 3). In both scenarios, we first simulate continuous traits as before and then convert them to binary traits using a chosen threshold described below.

First, we consider two independent studies of $3,000$ independent individuals. We increased the sample size compared to Simulation 2 with continuous traits to have meaningful power differences between methods when analyzing binary traits. Each individual has measurements on a single SNP with MAF 0.1 and 4 traits inspired by the METSIM lipids data on total cholesterol (TC), high-density lipoprotein (HDL), low-density lipoprotein (LDL) and triglycerides (TG). We use the trait correlation matrix $\boldsymbol{R}_{\text{metsim}}$ (Figure S1(a)) to simulate 4 traits using the model described in Simulation 1. We consider 5 association scenarios: (i) only TC is associated, (ii) TC and LDL are associated, (iii) TC, LDL and TG are associated, (iv) all 4 traits are associated, and (v) none of the traits is associated. As before, the SNP explains 0.5% of the trait variance when associated. We assume TC, LDL and TG have negative genetic effects while HDL has positive effect when associated. We simulate two study types: "homogeneous" and "heterogeneous". For "homogeneous" studies, the association pattern of the traits is same across both s-

tudies. For "heterogeneous" studies, we assume association scenarios (i)-(iv) in the first study while the traits are not associated (scenario (v)) in the second study. We converted these traits to binary traits as follows: for trait 1, we picked a cutoff corresponding to its 50^{th} quantile; if the continuous trait value exceeds this cutoff, the binary trait value is 1, otherwise 0. Similarly, for traits 2, 3 and 4, we chose cutoffs corresponding to 60^{th}, 70^{th} and 80^{th} quantiles respectively. We analyze each binary trait in each study using generalized linear model and obtain the summary statistics. Thus, for the given SNP and a given replication, we have 4 summary statistics (corresponding to 4 binary traits) from each of the 2 studies. As before, we consider $1,000$ replications (non-null datasets) for computing statistical powers.

Next, we consider two studies of $3,000$ independent individuals out of which 600 individuals are common to both studies. We keep everything the same as in the aforementioned scenario (two independent studies with binary traits) except that the two studies now have 600 overlapping individuals. For "homogeneous" studies, we assume the association pattern is same across the two studies. For "heterogeneous" studies, excluding the overlap, we assume the association scenarios (i)-(iv) in one study while the traits are not associated (scenario (v)) in the other study. For individuals common to both studies, we assume scenario (v). As before we convert the continuous traits to binary traits, analyze each trait from each study individually using generalized linear model (logistic model) to obtain summary statistics. $1,000$ such replicates are considered for computing statistical powers.

Table S3: Simulation 1: Change in type I error estimates at various levels α for misspecified trait correlation matrix $\boldsymbol{R}$. To misspecify $\boldsymbol{R}$, non-null datasets (with the SNP being associated with all traits) are simulated. The degree of misspecification is varied by varying the signal-to-noise ratio (or, proportion of trait variance explained by SNP, $h\%$). This table lists the ratio of type I error estimates of methods when $\boldsymbol{R}$ is estimated from non-null SNPs (with $h > 0$) compared to when $\boldsymbol{R}$ is estimated from null SNPs (with $h = 0$). The ideal value for any cell is 1implying there is no change in type I error with misspecification of $\boldsymbol{R}$. Any cell value > 1 indicates inflation of type I error with misspecification. Estimates are based on 10^7 null datasets (where, for analysis, the trait correlation matrix $\boldsymbol{R}$ is estimated from non-null datasets with varying h), each with sample size $1,000$.

α	Method	$\rho = 0.2$			$\rho = 0.4$			$\rho = 0.6$		
		$\frac{h=0.5}{h=0}$	$\frac{h=5}{h=0}$	$\frac{h=10}{h=0}$	$\frac{h=0.5}{h=0}$	$\frac{h=5}{h=0}$	$\frac{h=10}{h=0}$	$\frac{h=0.5}{h=0}$	$\frac{h=5}{h=0}$	$\frac{h=10}{h=0}$
10^{-2}	S_{Hom}	1.03	0.91	0.79	1.00	0.95	0.90	0.97	0.95	0.94
	S_{Het}	0.98	1.04	1.12	1.00	1.07	1.13	1.07	1.11	1.17
	minP	1.00	1.00	1.01	1.00	1.01	1.01	1.00	1.01	1.01
	metaMANOVA	0.98	1.05	1.12	1.00	1.06	1.13	1.08	1.12	1.17
	metaUSAT	0.99	1.00	1.04	1.00	1.02	1.04	1.03	1.05	1.07
10^{-3}	S_{Hom}	1.05	0.86	0.70	1.00	0.92	0.85	0.94	0.93	0.90
	S_{Het}	0.97	1.05	1.20	1.00	1.09	1.21	1.10	1.17	1.26
	minP	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.01
	metaMANOVA	0.97	1.06	1.17	1.00	1.08	1.19	1.12	1.18	1.26
	metaUSAT	1.00	1.00	1.03	0.99	1.02	1.05	1.04	1.07	1.10
10^{-4}	S_{Hom}	1.07	0.80	0.60	1.00	0.89	0.77	0.95	0.91	0.87
	S_{Het}	0.98	1.10	1.28	1.02	1.13	1.28	1.12	1.21	1.35
	minP	1.00	1.00	1.00	1.01	1.02	1.00	1.03	1.01	1.03
	metaMANOVA	0.95	1.06	1.24	1.01	1.12	1.25	1.13	1.19	1.29
	metaUSAT	0.99	0.98	1.01	1.00	1.00	1.02	1.06	1.09	1.13
10^{-5}	S_{Hom}	1.04	0.81	0.56	1.00	0.91	0.83	0.92	0.90	0.87
	S_{Het}	1.00	1.06	1.35	0.98	1.18	1.41	1.15	1.19	1.36
	minP	1.01	1.02	1.02	1.00	1.10	0.98	1.04	1.00	0.96
	metaMANOVA	0.95	1.14	1.40	0.99	1.11	1.39	1.15	1.19	1.35
	metaUSAT	1.00	0.96	0.99	1.01	0.99	1.00	1.11	1.10	1.14

Table S4: Simulation 2 (binary traits): Comparison of empirical powers (based on corrected critical values) for two independent studies at level $\alpha = 10^{-4}$ when the traits are binary. Power is estimated based on 10^4 non-null datasets. For a given association scenario, the method with highest power is bold-faced and the method with lowest power is italicized.

Study type	No. of traits associated	Meta-analysis method					
		S_{Hom}	S_{Het}	minP	aSPU	metaMANOVA	metaUSAT
Homogeneous	1	*0.000*	0.718	0.203	0.074	**0.916**	0.848
	2	*0.030*	0.486	0.271	0.463	0.500	**0.526**
	3	0.735	0.795	*0.345*	**0.889**	0.763	0.822
	4	*0.011*	**0.967**	0.443	0.937	0.956	0.960
Heterogeneous	1	*0.000*	0.213	0.104	0.020	**0.360**	0.235
	2	*0.002*	0.073	0.148	0.088	0.097	**0.100**
	3	*0.050*	0.179	0.185	0.211	0.196	**0.247**
	4	*0.001*	**0.482**	0.261	0.389	0.472	0.475

Table S5: Simulation 3 (binary traits): Comparison of empirical powers (based on corrected critical values) for two studies with overlapping samples at level $\alpha = 10^{-4}$ when the traits are binary. Power is estimated based on 10^4 non-null datasets. For a given association scenario, the method with highest power is bold-faced and the method with lowest power is italicized.

Study type	No. of traits associated	Meta-analysis method					
		S_{Hom}	S_{Het}	minP	aSPU	metaMANOVA	metaUSAT
Homogeneous	1	*0.000*	0.520	0.188	0.091	0.800	**0.726**
	2	*0.018*	0.263	0.258	0.458	0.313	**0.465**
	3	0.586	0.577	*0.340*	**0.813**	0.575	0.772
	4	*0.009*	0.895	0.422	0.930	0.875	**0.934**
Heterogeneous	1	*0.000*	0.059	0.030	0.006	0.111	**0.073**
	2	*0.000*	0.015	**0.043**	0.025	0.027	0.027
	3	*0.009*	0.041	0.061	0.062	0.061	**0.071**
	4	*0.000*	0.158	0.088	0.135	0.177	**0.176**

Supplementary S5

METSIM Study: Additional Figures and Tables

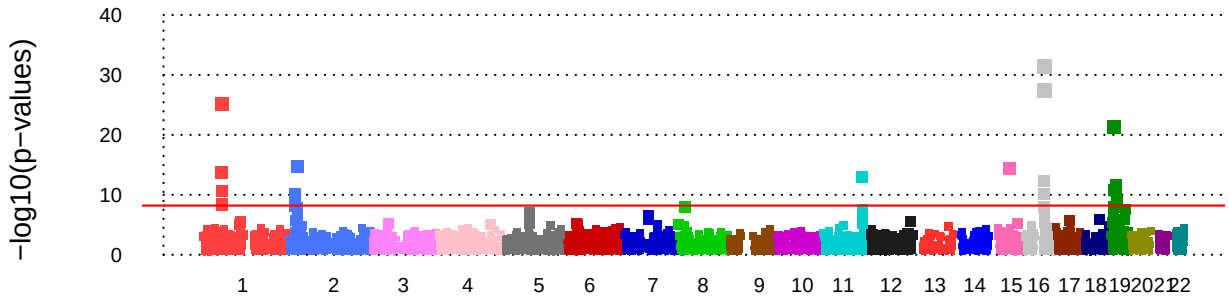
Table S6: METSIM Study: Joint meta-analysis of all lipid traits. This table lists the SNPs detected by both metaMANOVA and metaUSAT. Only the independent SNPs (pairwise distance > 500 kb and $r^2 < 0.1$) are listed. Genome-wide p-value threshold of 5×10^{-8} has been used to declare significance. The known association results are based on previously reported GWAS associations within 500 kb of and $r^2 > 0.7$ with any of these SNPs from the NHGRI GWAS catalog and our in-house GWAS catalog.

chr	position	rsID	p-value		Known association result
			meta-USAT	meta-MANOVA	
1	55085141	rs17395160	2.4×10^{-10}	1.6×10^{-10}	LIPO-fractions, Lipids
1	55224773	rs116816976	7.4×10^{-14}	1.3×10^{-13}	LIPO-fractions, Lipids
1	55505647	rs11591147	2.3×10^{-33}	3.7×10^{-32}	Lipids
1	55518467	rs2495477	5.4×10^{-14}	7.3×10^{-14}	LIPO-fractions, Lipids
1	55538552	rs10493176	3.1×10^{-10}	2.0×10^{-10}	LIPO-fractions, Lipids
1	55953290	rs2649629	9.5×10^{-9}	5.9×10^{-9}	APOA1B, LIPO-fractions, Lipids
1	62980607	rs10889337	5.2×10^{-9}	3.3×10^{-9}	Lipids
2	21225281	rs1042034	3.0×10^{-8}	1.8×10^{-8}	Amino Acids, LIPO-fractions, Lipids
2	21277922	rs6548010	1.3×10^{-12}	1.0×10^{-10}	APOA1B, LIPO-fractions, Lipids
2	27730940	rs1260326	5.1×10^{-14}	6.6×10^{-14}	Amino Acids, LIPO-fractions, Lipids
8	19816934	rs301	4.1×10^{-10}	2.7×10^{-10}	Lipids
8	126484526	rs2954026	4.8×10^{-12}	1.1×10^{-11}	Lipids
11	116648917	rs964184	2.2×10^{-21}	1.2×10^{-20}	Cardiovascular endpoints, Lipids
15	58678512	rs10468017	7.3×10^{-34}	1.2×10^{-32}	APOA1B, LIPO-fractions, Lipids
15	58723675	rs1800588	4.2×10^{-29}	4.4×10^{-28}	APOA1B, LIPO-fractions, Lipids
15	58855748	rs113298164	9.9×10^{-9}	6.2×10^{-9}	Lipids
16	56988044	rs173539	8.3×10^{-60}	1.6×10^{-57}	APOA1B, LIPO-fractions, Lipids
16	56992017	rs6499863	5.4×10^{-9}	3.4×10^{-9}	APOA1B, LIPO-fractions, Lipids
16	57006590	rs7499892	8.0×10^{-38}	1.9×10^{-36}	APOA1B, Lipids
16	57017319	rs1800777	1.9×10^{-8}	1.2×10^{-8}	APOA1B, LIPO-fractions, Lipids
19	10396336	rs3093032	5.1×10^{-9}	8.0×10^{-9}	Height
19	11202306	rs6511720	1.6×10^{-29}	1.8×10^{-28}	Lipids
19	19379549	rs58542926	3.8×10^{-14}	4.0×10^{-12}	Lipids, T2D
19	45242173	rs1531517	7.9×10^{-10}	5.1×10^{-10}	APOA1B, LIPO-fractions, Lipids
19	45373565	rs395908	2.3×10^{-13}	5.1×10^{-13}	Lipids
19	45395266	rs157580	4.5×10^{-13}	1.0×10^{-12}	APOA1B, LIPO-fractions, Lipids
19	45410002	rs769449	6.6×10^{-14}	1.1×10^{-13}	APOA1B, Fatty acids, LIPO-fractions
19	45412079	rs7412	3.0×10^{-76}	2.7×10^{-73}	APOA1B, LIPO-fractions, Lipids

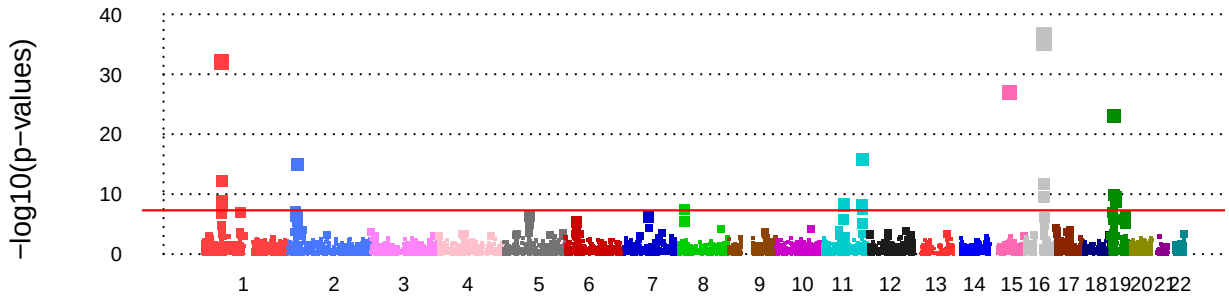
'LIPO-fractions' stands for 'lipoprotein fractions'.

Supplementary S6

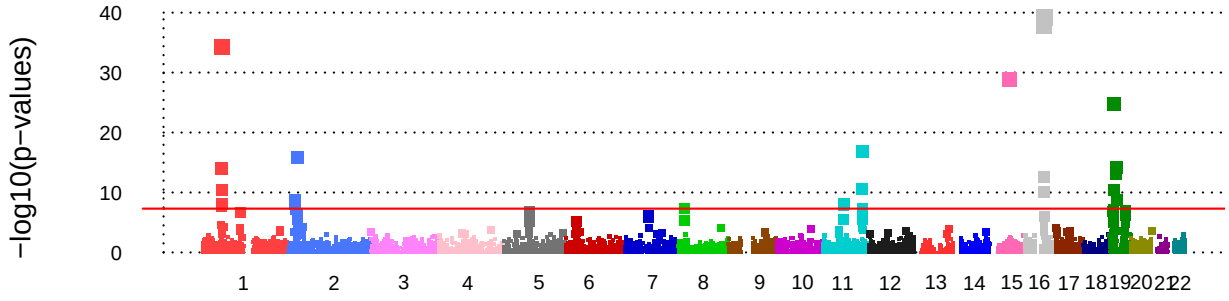
METSIM + T2D-GENES Studies: Additional Figures and Tables



(a) Most significant p-value among univarite trait p-values



(b) metaMANOVA p-values



(c) metaUSAT p-values

Figure S3: METSIM+T2D-GENES Studies: Manhattan Plots. In figure (a), $-\log_{10}(p_{\min})$ is plotted, where for a given SNP, $p_{\min}$ is the minimum of p-values across all 4 lipids traits and across both studies. In figures (b) and (c), $-\log_{10}(p_{\text{metaMANOVA}})$ and $-\log_{10}(p_{\text{metaUSAT}})$ are plotted respectively, where all 4 lipid traits from both studies are jointly meta-analyzed. Red solid line corresponds to GWAS threshold for significance. Bonferroni adjusted threshold is used for the univariate analysis.

Table S7: T2D-GENES + METSIM Studies: Meta-analysis of Total Cholesterol (TC). This table lists the SNPs detected by both metaMANOVA and metaUSAT. Only the independent SNPs (pairwise distance > 500 kb and $r^2 < 0.1$) are listed. Genome-wide p-value threshold of 5×10^{-8} has been used to declare significance. The known association results are based on previously reported GWAS associations within 500 kb of and $r^2 > 0.7$ with any of these SNPs from the NHGRI GWAS catalog and our in-house GWAS catalog.

chr	position	rsID	p-value		Known association result
			meta-USAT	meta-MANOVA	
1	55505647	rs11591147	1.4×10^{-20}	1.2×10^{-19}	Cardiovascular endpoints, LIPO-fractions, Lipids
1	55518467	rs2495477	4.1×10^{-15}	6.4×10^{-14}	Cardiovascular endpoints, LIPO-fractions, Lipids
1	55538552	rs10493176	1.8×10^{-10}	4.9×10^{-10}	Cardiovascular endpoints, LIPO-fractions, Lipids
2	21263900	rs1367117	3.3×10^{-9}	1.1×10^{-8}	Glycemic, Height, LIPO-fractions, Lipids, T2D
5	74651084	rs3846662	1.8×10^{-8}	1.1×10^{-8}	LIPO-fractions, Lipids
11	116662579	rs651821	1.2×10^{-8}	2.0×10^{-8}	APOA1B, LIPO-fractions, Lipids
19	11210912	rs2228671	9.6×10^{-20}	6.4×10^{-19}	Lipids
19	19379549	rs58542926	1.7×10^{-12}	2.3×10^{-11}	Lipids, T2D

'LIPO-fractions' stands for 'lipoprotein fractions'.

Details of variants associated with TC found by both metaUSAT and metaMANOVA (Table S7): rs11591147 ($p_{\text{metaUSAT}} = 1.4 \times 10^{-20}$), a missense variant on *PCSK9* gene in strongly associated with LDL (Willer et al., 2013) and is near known GWAS hit for TC (Surakka et al., 2015). Two other variants on *PCSK9* gene, rs2495477 ($p_{\text{metaUSAT}} = 4.1 \times 10^{-15}$) and rs10493176 ($p_{\text{metaUSAT}} = 1.8 \times 10^{-10}$), are near known GWAS hits for cardiovascular endpoints (Kathiresan et al., 2009) and lipoprotein fractions (Kettunen et al., 2012). rs1367117 ($p_{\text{metaUSAT}} = 3.3 \times 10^{-9}$), a missense variant on *APOB* gene, is strongly associated with TC (Willer et al., 2013). rs3846662 ($p_{\text{metaUSAT}} = 1.8 \times 10^{-8}$), a non coding transcript exon variant on *HMGCR* gene, and rs2228671 ($p_{\text{metaUSAT}} = 9.6 \times 10^{-20}$), a missense variant on *LDLR* gene, are associated with TC (Aulchenko et al., 2009). rs651821 ($p_{\text{metaUSAT}} = 1.2 \times 10^{-8}$), a 5'-UTR variant on *APOA5* gene, is associated with lipid metabolism (Kettunen et al., 2012). Yet another SNP found to be associated with TC is rs58542926 ($p_{\text{metaUSAT}} = 1.7 \times 10^{-12}$), a missense variant on *CILP2* gene (Surakka et al., 2015).

Table S8: T2D-GENES + METSIM Studies: Meta-analysis of all 4 lipid traits. This table lists the SNPs detected by both metaMANOVA and metaUSAT. Only the independent SNPs (pairwise distance > 500 kb and $r^2 < 0.1$) are listed. Genome-wide p-value threshold of 5×10^{-8} has been used to declare significance. The known association results are based on previously reported GWAS associations within 500 kb of and $r^2 > 0.7$ with any of these SNPs from the NHGRI GWAS catalog and our in-house GWAS catalog.

chr	position	rsID/exmID	p-value		Known association result
			meta-USAT	meta-MANOVA	
1	55505647	rs11591147	1.7×10^{-36}	5.0×10^{-34}	Cardiovascular endpoints, LIPO-fractions, Lipids
1	55518467	rs2495477	5.5×10^{-15}	2.9×10^{-13}	Cardiovascular endpoints, LIPO-fractions, Lipids
1	55538552	rs10493176	4.6×10^{-11}	5.6×10^{-10}	Cardiovascular endpoints, LIPO-fractions, Lipids
2	27730940	rs1260326	5.7×10^{-17}	5.9×10^{-16}	Glycemic, Height, LIPO-fractions, Lipids, T2D
8	19819724	rs328	3.1×10^{-8}	2.9×10^{-8}	LIPO-fractions, Lipids
11	61570783	rs174547	3.7×10^{-9}	2.4×10^{-9}	Glycemic, LIPO-fractions, Lipids
11	116655600	rs35120633	2.8×10^{-8}	2.2×10^{-8}	APOA1B, LIPO-fractions, Lipids
11	116662579	rs651821	7.7×10^{-18}	9.1×10^{-17}	APOA1B, LIPO-fractions, Lipids
15	58723675	rs1800588	6.4×10^{-30}	6.1×10^{-28}	APOA1B, Fatty acids, LIPO-fractions, Lipids
16	56995236	rs1800775	9.2×10^{-41}	5.8×10^{-38}	APOA1B, LIPO-fractions, Lipids
16	57015091	rs5880	1.3×10^{-13}	1.0×10^{-12}	APOA1B, LIPO-fractions, Lipids
19	11210912	rs2228671	4.5×10^{-26}	2.2×10^{-24}	Lipids
19	19379549	rs58542926	7.2×10^{-15}	3.4×10^{-10}	Lipids, T2D

'LIPO-fractions' stands for 'lipoprotein fractions'.

Details of variants associated with lipids found by metaUSAT alone (Table 4): rs2483205 ($p_{\text{metaUSAT}} = 2.5 \times 10^{-8}$) is strongly associated with LDL and TC in the GLGC data. It is near many known GWAS hits, which are associated with lipids (Surakka et al., 2015), lipoprotein fractions (Kettunen et al., 2012), cardiovascular endpoints (Kathiresan et al., 2009). The second SNP rs1367117, a missense variant on *APOB* gene ($p_{\text{metaUSAT}} = 1.5 \times 10^{-9}$), is strongly associated with 3 out 4 lipids traits (HDL, TC and TG) in the GLGC data (Willer et al., 2013), and is well-known for its association with lipids (Surakka et al., 2015; Teslovich et al., 2010). Lastly, the SNP rs2304130, a splice region variant on *ZNF101* gene ($p_{\text{metaUSAT}} = 1.5 \times 10^{-9}$), is very strongly associated with all 4 lipid traits in the GLGC data and is well known for its association with lipids (Kristiansson et al., 2012). Observe that $p_{\text{metaMANOVA}}$ for these SNPs are quite weak.

References

Aulchenko, Y. S., Ripatti, S., Lindqvist, I., Boomsma, D., Heid, I. M., Pramstaller, P. P., Penninx, B. W., Janssens, A. C. J., Wilson, J. F., Spector, T. et al. (2009), 'Loci influencing lipid levels and coronary heart disease risk in 16 european population cohorts', *Nat Genet* **41**(1), 47–55.

Davies, R. (1980), 'Algorithm AS 155: The distribution of a linear combination of chi-square random variables', *J R Stat Soc Series C Appl Stat* **29**(3), 323–333.

Duchesne, P. and de Micheaux, P. L. (2010), 'Computing the distribution of quadratic forms: Further comparisons between the Liu-Tang-Zhang approximation and exact methods', *Comput Stat Data Anal* **54**, 858–862.

Kathiresan, S., Voight, B. F., Purcell, S., Musunuru, K., Ardissino, D., Mannucci, P. M., Anand, S., Engert, J. C., Samani, N. J., Schunkert, H. et al. (2009), 'Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants', *Nat Genet* **41**(3), 334–341.

Kettunen, J., Tukiainen, T., Sarin, A.-P., Ortega-Alonso, A., Tikkanen, E., Lyytikäinen, L.-P., Kangas, A. J., Soininen, P., Würtz, P., Silander, K. et al. (2012), 'Genome-wide association study identifies multiple loci influencing human serum metabolite levels', *Nat Genet* **44**(3), 269–276.

Khatri, C. (1980), 'The necessary and sufficient conditions for dependent quadratic forms to be distributed as multivariate gamma', *J Multivar Anal* **10**, 233–242.

Khatri, C., Krishnaiah, P. and Sen, P. (1977), 'A note on the joint distribution of correlated quadratic forms', *J Stat Plan Inference* **1**, 299–307.

Kristiansson, K., Perola, M., Tikkanen, E., Kettunen, J., Surakka, I., Havulinna, A. S., Stančáková, A., Barnes, C., Widen, E., Kajantie, E. et al. (2012), 'Genome-wide screen for metabolic syndrome susceptibility loci reveals strong lipid gene contribution but no

evidence for common genetic basis for clustering of metabolic syndrome traits', *Circ Cardiovasc Genet* **5**(2), 242–249.

Lee, S., Wu, M. and Lin, X. (2012), 'Optimal tests for rare variant effects in sequencing association studies', *Biostatistics* **13**, 762–775.

Liu, H., Tang, Y. and Zhang, H. (2009), 'A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables', *Comput Stat Data Anal* **53**, 853–856.

Ray, D., Pankow, J. S. and Basu, S. (2016), 'USAT: a unified score-based association test for multiple phenotype-genotype analysis', *Genet Epidemiol* **40**(1), 20–34.

Surakka, I., Horikoshi, M., Mägi, R., Sarin, A.-P., Mahajan, A., Lagou, V., Marullo, L., Ferreira, T., Miraglio, B., Timonen, S. et al. (2015), 'The impact of low-frequency and rare variants on lipid levels', *Nat Genet* **47**(6), 589–597.

Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki, M., Pirruccello, J. P., Ripatti, S., Chasman, D. I., Willer, C. J. et al. (2010), 'Biological, clinical and population relevance of 95 loci for blood lipids', *Nature* **466**(7307), 707–713.

Willer, C. J., Schmidt, E. M., Sengupta, S., Peloso, G. M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M. L. et al. (2013), 'Discovery and refinement of loci associated with lipid levels', *Nat Genet* **45**(11), 1274–1283.

Wu, B., Guan, W. and Pankow, J. S. (2016), 'On efficient and accurate calculation of significance p-values for sequence kernel association testing of variant set', *Ann Hum Genet* **80**, 123–135.

Wu, B. and Pankow, J. (2016), 'Sequence kernel association test of multiple continuous phenotypes', *Genet Epidemiol* **40**(2), 91–100.