



Published in final edited form as:

*Genet Epidemiol.* 2020 January ; 44(1): 16–25. doi:10.1002/gepi.22267.

## Trans-ethnic Meta-analysis of Metabolic Syndrome in a Multi-ethnic Study

Emileigh L. Willems<sup>1</sup>, Jia Y. Wan<sup>2</sup>, Trina M. Norden-Krichmar<sup>2</sup>, Karen L. Edwards<sup>2</sup>,  
Stephanie A. Santorico<sup>1,3,4</sup>

<sup>1</sup>Department of Mathematical and Statistical Sciences, University of Colorado Denver

<sup>2</sup>Department of Epidemiology, University of California Irvine

<sup>3</sup>Human Medical Genetics and Genomics Program, University of Colorado Denver

<sup>4</sup>Department of Biostatistics & Informatics, University of Colorado Denver

### Abstract

Genome-wide association studies (GWAS) have been used to establish thousands of genetic associations across numerous phenotypes. To improve the power of GWAS and generalize associations across ethnic groups, trans-ethnic meta-analysis methods are used to combine the results of several GWAS from diverse ancestries. The goal of this study is to identify genetic associations for eight quantitative metabolic syndrome (MetS) traits through a meta-analysis across four ethnic groups. Traits were measured in the GENetics of Non-Insulin dependent Diabetes mellitus (GENNID) Study which consists of African-American (families=73, individuals=288), European-American (families=79, individuals=519), Japanese-American (families=17, individuals=132), and Mexican-American (families=113, individuals=610) samples. Genome-wide association results from these four ethnic groups were combined using four meta-analysis methods: fixed effects, random effects, TransMeta, and MR-MEGA. We provide an empirical comparison of the four meta-analysis methods from the GENNID results, discuss which types of loci (characterized by allelic heterogeneity) appear to be better detected by each of the four meta-analysis methods in the GENNID Study, and validate our results using previous genetic discoveries. We specifically compare the two trans-ethnic methods, TransMeta and MR-MEGA, and discuss how each trans-ethnic method's framework relates to the types of loci best detected by each method.

### Keywords

meta-analysis; trans-ethnic; genome-wide association; metabolic syndrome; allelic heterogeneity

---

**Corresponding Author:** Emileigh Willems, Campus Box 170, P.O. Box 173364, 1201 Larimer St., Room 4000, Denver, CO 80217, emileigh.willems@ucdenver.edu.

Data Availability Statement

Genotype and phenotype data will be deposited in the database of Genotypes and Phenotypes (dbGaP). Submission is in process.

## Introduction

Meta-analysis methods are commonly used to combine the results of multiple genome-wide association studies (GWAS) and can boost power by increasing overall sample size. Performing a meta-analysis on studies from diverse populations can begin to address the issue of whether associations present within one population are consistent over other populations. These trans-ethnic meta-analysis methods have been used to understand the role of allelic heterogeneity across diverse ethnic groups. This knowledge will be necessary for applications such as precision medicine, where correct diagnoses are tailored to the patient's genetics and ancestry. A major benefit of meta-analysis methods is that they only require summary statistics from each GWAS rather than raw genotype data, which gives preference to meta-analysis when using publicly available resources or when privacy concerns arise.

Trans-ethnic meta-analysis methods are recent extensions of traditional meta-analysis methods designed to allow heterogeneity between studies, while also leveraging information between studies of similar ethnicity or ancestry. When performing a meta-analysis on studies from diverse backgrounds, we expect studies from similar ethnic backgrounds to have more homogeneous allelic effects, while we expect studies from diverse ethnic backgrounds to more likely have heterogeneous allelic effects. Thus, trans-ethnic methods model the between-study correlation of each variant's effect through some metric of genetic distance.

The GENNID Study (Raffel, Robbins, Norris, Boerwinkle, et al., 1996) will be used to empirically compare four meta-analysis methods. This data set is representative of trans-ethnic effects because it consists of samples from four diverse backgrounds: African-American, European-American, Japanese-American, and Mexican-American. We compare the results of four existing meta-analysis methods when applied to the GENNID data set (Raffel et al., 1996). We compare the two standard methods, fixed effects (FE) (Stram, 2014) and random effects (RE2) (Han & Eskin, 2011) meta-analysis, along with two trans-ethnic methods, TransMeta (Shi & Lee, 2016) and MR-MEGA (Magi et al., 2017). The two trans-ethnic methods, TransMeta and MR-MEGA, are both relatively new and have not been compared to each other, either through real data application or simulation. Thus, our overarching goal is to gain a better understanding of each method's empirical performance when applied to the GENNID data set. In particular, we will consider the validity of each test, evaluate the types of loci best found by each method in the GENNID data set, and discuss validation of each method's observed findings from previous genetic studies.

## Methods

### - Study subjects and data

The GENNID data set (Raffel et al., 1996) contains samples from Type 2 Diabetes multiplex families across four diverse ethnic groups: African-American (AA, families = 73, individuals = 288), European-American (EA, families = 79, individuals = 519), Japanese-American (JA, families = 17, individuals = 132), and Mexican-American (MA, families = 113, individuals = 610). Phenotypic information for the following metabolic traits were collected for all GENNID samples: body weight (kg), waist circumference (cm), systolic and diastolic blood pressure (mmHg), triglyceride (mg/dl), hdl cholesterol (mg/dl), fasting

glucose (mg/dl), and fasting insulin (mg/dl). Genotypic data was obtained for all samples using the Infinium Multi-Ethnic Global BeadChip (v1.0) with the following number of single nucleotide polymorphisms (SNPs) passing post-genotyping quality control (QC) and a minor allele frequency (MAF) threshold of 0.01 in each ethnic group: 864,805 SNPs in AA, 704,357 SNPs in EA, 583,902 SNPs in JA, and 761,432 SNPs in MA (see Appendix A for full QC information).

Samples were phased using EAGLE2 by the Sanger Imputation Service (Loh et al., 2016). The Haplotype Reference Consortium (HRC) panel (McCarthy et al., 2016) was used to phase the EA sample, and the 1000 Genomes Phase 3 (1000G) panel (Gibbs et al., 2015) was used to phase the AA, MA, and JA samples, respectively. The choice for reference panels during phasing was based on Supplementary Table 5 of Loh et al. (2016). Since EAGLE2 cannot account for familial relatedness, the software DuoHMM was used to correct phasing switch errors based on pedigree structures (O'Connell et al., 2014). The samples were then imputed separately using Minimac3 by the Michigan Imputation Server using the HRC panel (Das et al., 2016; McCarthy et al., 2016). Since the HRC panel contains 1000 Genomes, the HRC panel was chosen to align with recommendations to use a cosmopolitan approach when imputing admixed multi-ethnic populations (Howie, Marchini, & Stephens, 2011; Medina-Gomez et al., 2015). After imputation, imputed data from detected chromosomal anomalies was removed (see Appendix A). SNPs with an imputed R-squared  $\leq 0.7$  and MAF  $\leq 0.01$  were used in association analyses. Genotypes for all Mendelian errors (MEs) were blanked using PLINK1.9, and SNPs with  $> 2$  MEs in each ethnic group were removed. Hardy-Weinberg equilibrium (HWE) checks were performed for all remaining SNPs, and variants with HWE p-value  $< 1e-6$  were removed (Chang et al., 2015). Final SNP counts used in each ethnic group for association analyses were 13,042,663 in AA, 7,681,619 in EA, 5,495,615 in JA, and 7,907,815 in MA. Only SNPs that passed QC and had MAF  $\geq 0.05$  in all four ethnic groups were included in the meta-analyses (3,153,931 SNPs).

#### - Association Analyses

A linear mixed-effects model (LMM) was performed separately for the EA, AA, and MA samples using GCTA (Yang, Lee, Goddard, & Visscher, 2011). The phenotypes hdl, triglycerides, waist, and insulin average were log transformed, and a rank-based inverse normal transformation was applied to the phenotypes weight, diastolic average, systolic average, and glucose average. The same transformations were applied across ethnic groups, so the effect size estimates are on the same scale. A genetic relationship matrix (GRM) was created separately for each of the three samples in LDAK using all SNPs with MAF  $\geq 0.01$  that were either genotyped SNPs or had Imputed R-squared  $\geq 0.99$  (Speed, Hemani, Johnson, & Balding, 2012). The covariates age, sex, and self-reported diabetes status (yes/no) were modeled as fixed effects in the LMM, and the effect of the GRM,  $K$ , was modeled as a random effect,  $g$ . An association p-value for each SNP was generated by testing the coefficient  $\beta_{SNP} = 0$  for each LMM described below. The vector  $\underline{SNP}$  in the equation below is the additively coded genotype for each subject at the tested SNP. The vector  $\underline{\epsilon}$  is additive noise due to statistical variability (modeled by  $\sigma_{\epsilon}^2$ ), while  $g$  models the

variability between individuals due to genetic relatedness (modeled by  $\sigma_g^2$  and  $K$ ). In the equation below, all underlined variables represent  $m \times 1$  vectors, where  $m$  is the number of subjects included in the LMM.

$$\underline{Phenotype} = \beta_0 + \beta_{Age} * \underline{Age} + \beta_{Sex} * \underline{Sex} + \beta_D * \underline{DiabetesStatus} + \beta_{SNP} * \underline{SNP} + \underline{g} + \underline{\epsilon}$$

$$\underline{g} \sim N(0, \sigma_g^2 \cdot K), \quad \underline{\epsilon} \sim N(0, \sigma_e^2 \cdot I), \quad \text{where } \sigma_g^2 \text{ and } \sigma_e^2 \text{ are estimated by GCTA}$$

Since the validity of an LMM depends on asymptotics due to large sample size, association analysis using unconditional gene-dropping (GD) was performed in the JA sample (families = 15, individuals = 125). Unconditional GD data sets were generated using Merlin (100,000–20,000,000 simulations per SNP, Appendix B), and additive association effect sizes were found for each GD data set using phenotypes pre-adjusted for age, sex and diabetes status (Abecasis, Cherny, Cookson, & Cardon, 2001). The effect sizes from the 100,000–20,000,000 data sets per SNP were used to generate the empirical distributions of the effect size estimates and produce empirical p-values for each SNP (Appendix B).

#### - Meta-analysis Methods

The summary statistics from the association analyses in each sample were combined in meta-analyses separately by phenotype using Fixed Effects (FE) (Stram, 2014), Random Effects (RE2) (Han & Eskin, 2011), TransMeta (Shi & Lee, 2016), and MR-MEGA (Magi et al., 2017). FE and RE2 were implemented using METASOFT (Han & Eskin, 2011).

The FE implementation performed by METASOFT produces an FE p-value using the typical inverse-variance weighted estimator (Stram, 2014):

$$\hat{\beta}_{FE, \ell} = \frac{\sum_{i=1}^n \frac{\hat{\beta}_{i\ell}}{\hat{\sigma}_{i\ell}^2}}{\sum_{i=1}^n \frac{1}{\hat{\sigma}_{i\ell}^2}}$$

where  $n$  is the number of studies combined in the meta-analysis ( $n = 4$  in our case),  $\hat{\beta}_{i\ell}$  is the effect size estimate for SNP  $\ell$  in study  $i$ , and  $\hat{\sigma}_{i\ell}^2$  is the estimated variance of  $\hat{\beta}_{i\ell}$  (found using the LMM or GD distributions). The test statistic:

$$T_{FE, \ell} = \left( \hat{\beta}_{FE, \ell} \cdot \sqrt{\sum_{i=1}^n \frac{1}{\hat{\sigma}_{i\ell}^2}} \right)$$

is distributed  $N(0,1)$  under the null hypothesis of no association and can be used to produce a p-value for testing for association at SNP  $\ell$  (Han & Eskin, 2011). Note that FE assumes homogeneous effects between all studies and is thus ideal when all studies come from the same ethnic background.

The RE2 implementation in METASOFT produces a p-value using Han & Eskin’s corrected random effects meta-analysis method (2011). The method performs a likelihood ratio test where the likelihoods assume that the estimators,  $\hat{\beta}_{i\ell}$ , are normally distributed with mean  $\beta_\ell$  and variance differing between the two hypotheses. The parameter  $\beta_\ell$  is the expected effect size across all populations at SNP  $\ell$ . The variance of  $\hat{\beta}_{i\ell}$  is equal to  $\hat{\sigma}_{i\ell}^2$  under  $H_0$  and equal to  $\hat{\sigma}_{i\ell}^2 + \tau_\ell^2$  under  $H_1$ , where  $\hat{\sigma}_{i\ell}^2$  is the variability attributed to estimation error and  $\tau_\ell^2$  is the between study variability. This method tests the hypotheses:  $H_0: \beta_\ell = 0, \tau_\ell = 0$  vs.  $H_1: \beta_\ell \neq 0, \tau_\ell \neq 0$  by finding the maximum likelihood estimates,  $\hat{\beta}_{\ell, MLE}$  and  $\hat{\tau}_{\ell, MLE}^2$ , and performing a likelihood ratio test of the following likelihoods (Han & Eskin, 2011).

$$\mathcal{L}_0: \prod_{i=1}^n \frac{1}{\sqrt{2\pi\hat{\sigma}_{i\ell}^2}} e^{-\frac{\hat{\beta}_{i\ell}^2}{2\hat{\sigma}_{i\ell}^2}}, \mathcal{L}_1: \prod_{i=1}^n \frac{1}{\sqrt{2\pi(\hat{\sigma}_{i\ell}^2 + \tau_\ell^2)}} e^{-\frac{(\hat{\beta}_{i\ell} - \beta_\ell)^2}{2(\hat{\sigma}_{i\ell}^2 + \tau_\ell^2)}}$$

Therefore, the two hypotheses can be interpreted as:  $H_0$ : *no effect* ( $\beta_\ell = 0$ ) and *no heterogeneity present* ( $\tau_\ell = 0$ ) vs.  $H_1$ : *Test for non-zero effect* ( $\beta_\ell \neq 0$ ) or *heterogeneity between studies* ( $\tau_\ell \neq 0$ ). Since RE2 models additional heterogeneity between studies through the parameter  $\tau_\ell^2$ , RE2 is ideal when studies come from very diverse backgrounds or there is a strong environmental effect.

The trans-ethnic method, TransMeta, follows a hierarchical framework that incorporates a kernel matrix  $K$  into the covariance structure of the effect size estimates (Shi & Lee, 2016). The hierarchy described below is in multivariate form where every underlined element is an  $n \times 1$  vector and  $MVN(\cdot, \cdot)$  is the  $n$ -dimensional multivariate normal distribution.

$$\underline{\hat{\beta}}_\ell | \underline{\beta}_\ell \sim MVN(\underline{\beta}_\ell, \Sigma_\ell)$$

$$\underline{\beta}_\ell | \tau \sim MVN(\underline{0}, \tau \cdot V_\rho)$$

$$V_\rho = (1 - \rho)K + \rho \underline{1}\underline{1}'$$

The vector  $\underline{\hat{\beta}}_\ell$  contains the estimated effect sizes  $\hat{\beta}_{i\ell}$ ,  $\underline{\beta}_\ell$  is the vector of true effect sizes in each study, and  $\Sigma_\ell = \text{diag}(\hat{\sigma}_{1\ell}^2, \dots, \hat{\sigma}_{n\ell}^2)$ . The term  $\underline{1}\underline{1}'$  is an  $n \times n$  matrix of ones, while the  $n \times n$  matrix,  $K$ , describes the pairwise correlation between studies using the genetic distance metric  $F_{st}$  (matrix form described by the Transmeta.Fst method (Shi & Lee, 2016)). For this matrix,  $F_{st}$  was calculated using a pruned set of 54,431 SNPs (using PLINK1.9’s --indep-pairwise flag with 10 Mb sliding windows and LD  $r^2 = 0.2$ ) and the R package BEDASSLE (Bradburd, Ralph, & Coop, 2013; Purcell et al., 2007). TransMeta tests the

hypotheses,  $H_0: \tau = 0$  vs.  $H_1: \tau \neq 0$ , by computing Wald test p-values for the values  $\rho \in \{0, 0.09, 0.25, 1\}$ . The minimum of the four Wald test p-values is used as the final p-value for this method which is corrected for multiple testing over  $\rho$ .

The final trans-ethnic method, MR-MEGA (Magi et al., 2017), follows a meta-regression framework that includes Multi-Dimensional Scaling (MDS) axes as covariates to adjust for trans-ethnic effects between studies. The MDS axes are found from an  $n \times n$  pairwise

genetic distance matrix whose  $ij$ th entry is  $\frac{\sum_{\ell} I_{\ell} (p_{i\ell} - p_{j\ell})^2}{\sum_{\ell} I_{\ell}}$ , where  $I_{\ell}$  is the indicator function

for inclusion of the  $k$ th variant and  $p_{i\ell}$  and  $p_{j\ell}$  are the allele frequency estimates of the  $k$ th variant in the  $i$ th and  $j$ th population. MR-MEGA fits a weighted least squares regression with the estimated study specific effect sizes,  $\hat{\beta}_{i\ell}$ , as the response variable and the inverse of

the estimated variance for each  $\hat{\beta}_{i\ell}$ ,  $\frac{1}{\hat{\sigma}_{i\ell}^2}$ , as the weight for each study. Mathematically, for the  $k$ th variant, where  $I \in \{1, 2, \dots, n\}$ , indexes the  $i$ th GWAS study, we model (Magi et al., 2017):

$$E(\hat{\beta}_{i\ell}) = \alpha_{\ell} + \sum_{t=1}^T \gamma_{t\ell} x_{it}$$

where  $x_{it}$  is the  $t$ th MDS axis. The MDS axes utilized in MR-MEGA are automatically calculated by the MR-MEGA software, and no alterations were specified for this calculation (Magi et al., 2017). One MDS axis was included in the MR-MEGA model during analyses ( $T = 1$ ) given the constraint of  $T < n - 2$  discussed in (Magi et al., 2017). The test of overall association at the  $k$ th variant compares the unconstrained model (equation above) to the model with  $\alpha_{\ell} = \gamma_{1\ell} = \dots = \gamma_{T\ell} = 0$  using a likelihood ratio test. This results in a test statistic asymptotically distributed  $\chi_{T+1}^2$  under the null hypothesis that can be used to calculate the association p-value for each variant.

## Results

We assessed the validity of the four methods by constructing QQplots of the  $-\log_{10}$  p-values from a LD-pruned set of 48,624 SNPs, where these SNPs were the intersection of the list of LD-pruned SNPs used to calculate  $F_{st}$  and the list of SNPs included for meta-analysis (see Methods). Figure 1 shows the QQplots and inflation factors ( $\lambda$ ) for all four methods and eight phenotypes. The expected line is shown in black with dashed black lines for the 95% confidence interval around this expectation. Figure 1 shows that the results of FE appear to be well controlled (both through the plot and  $\lambda$  values), while the results of RE2 appear to be deflated (maximum RE2  $\lambda = 0.815$ ). The  $\lambda$  values for TransMeta show deflation as well (maximum TransMeta  $\lambda = 0.887$ ), and the QQplot depicts an excess of p-values equal to 1 (bottom left of TransMeta QQplot), which skews the rest of the distribution from expectation. MR-MEGA appears to be a better controlled test with slight deflation (minimum MR-MEGA  $\lambda = 0.965$ ). Inflation factors calculated at the 10<sup>th</sup> percentile, 1<sup>st</sup> percentile, and 0.1<sup>th</sup> percentile (instead of the median) for each method's results are shown

in Appendix C. These results show that FE is well-controlled using all metrics, and that the deflation observed for RE2 improves for the 1<sup>st</sup> and 0.1<sup>th</sup> percentile results, where we only observe slight deflation in the tail for RE2. The TransMeta results in Supplemental Table C.3 show that while the method is slightly deflated at the median, we observe slight inflation in the tail (1<sup>st</sup> and 0.1<sup>th</sup> percentile results). With Supplemental Table C.4, we observe slight deflation across all four metrics for MR-MEGA, illustrating this test to be slightly conservative.

Table 1 summarizes the results of the four methods using a suggestive threshold of  $1e-6$  and a genome-wide significant threshold of  $5e-8$ . The table provides the number of the 3,153,931 SNPs at least one meta-analysis method found to be suggestively significant ( $p\text{-value} < 1e-6$ ). There were 78 such SNPs in total (further information for the 78 SNPs found in Appendix D). TransMeta found the most SNPs (62 SNPs) to be suggestive ( $p\text{-value} < 1e-6$ ), while MR-MEGA found the least to be suggestive (31 SNPs). Although TransMeta found more SNPs to be suggestive than FE and RE2, results among FE, RE2, and TransMeta are comparable at the genome-wide significance level ( $p\text{-value} < 5e-8$ ). All methods besides MR-MEGA found at least one SNP to be genome-wide significant. Two of the three SNPs found to be genome-wide significant by TransMeta were found to be genome-wide significant by RE2, and all three genome-wide significant SNPs found by TransMeta were found to be genome-wide significant by FE. The additional SNPs found to be genome-wide significant by FE and RE2 did not overlap between the methods. TransMeta had the smallest p-value of the four methods most frequently (48 times). We notice that although MR-MEGA does not find as many SNPs to be suggestive or genome-wide significant as FE or RE2, MR-MEGA has the lowest p-value of the four methods more frequently than FE or RE2. This result appears to be driven by the fact that there was a locus detected only by MR-MEGA.

A Venn Diagram of the 78 SNPs detected by the four methods is displayed in Figure 2. We see that there is a large amount of overlap between the SNPs detected by each of the methods. FE and RE2 only found 2 SNPs each that were undetected by the other methods, while MR-MEGA and TransMeta both found 12 SNPs each that were undetected by any other method (Figure 2). Ten of the 12 SNPs detected by only MR-MEGA make up the locus that was only detected by MR-MEGA, while the additional 12 SNPs detected by only TransMeta were scattered across multiple loci.

The results of the four meta-analysis methods suggest that each method may be better able to detect loci with specific types of effects (either as suggestive or genome-wide significant) than the other methods (column 2 of Table 1). The three loci shown in Figures 3, 4, and 5 depict the types of loci best captured by RE2, MR-MEGA, and TransMeta, respectively, in the GENNID data set. Figure 3 shows a locus (for triglycerides) only found to be genome-wide significant by RE2 on chromosome 2. The table in Figure 3 shows that this locus is primarily driven by the EA sample, with some moderate influence from the AA sample. We note that the effect size estimates in the EA and AA samples (0.254 & 0.205, respectively) are very different than the effect size estimates in the MA and JA samples (0.059 & 0.009, respectively) at the lead SNP in this locus. Figure 4 shows the locus (for triglycerides) detected only by MR-MEGA ( $p\text{-value} < 1e-6$ ) on chromosome 19. The individual



sample p-values in the table of Figure 4 show that this locus is solely driven by the association observed in the AA sample. Figure 5 shows a locus where TransMeta had the lowest p-value of the four methods (chromosome 2, weight association). The table in Figure 5 demonstrates that this suggestive locus is driven by moderate effects in the EA, AA, and MA samples. The effect size estimates in the three samples vary slightly ( $-0.206$ ,  $-0.230$ , and  $-0.265$ , respectively), while the JA effects size estimate is quite different at this locus ( $0.187$ ). We characterize and discuss the types of loci that appear to be best captured by each method in the Discussion.

## Discussion

The QQplots of the results from the four methods show that the FE meta-analysis appears to be the only method with a well-controlled Type 1 error rate when applied to the multi-ethnic GENNID data set (Figure 1). The slight deflation observed in the tail of the RE2 results (Appendix C) can likely be attributed to the small number of samples used for the meta-analysis. Han & Eskin stress in the methodological paper for RE2 that when a meta-analysis is comprised of a small number of studies, their table of pre-tabulated p-values should be used (2011). However, these pre-tabulated p-values were created based on the assumption of equal sample sizes between studies and deviations from this assumption (like in the GENNID data set) will lead to slightly conservative results (Han & Eskin, 2011). An excess of p-values equal to 1 in the TransMeta QQplot leads to deviations from the expected distribution and deflated  $\lambda$  values at the median (Figure 1). The authors of TransMeta have communicated that the excess of p-values equal to 1 is due to the algorithm's use of numerical integration when generating final p-values for the method (personal correspondence). However, since we observe slight inflation in the tail of the results (Appendix C) it is unclear whether some of the significant or suggestive results found by TransMeta are spurious. Slight deflation is observed in MR-MEGA's QQplot and by all four metrics explored in Appendix C, and this deflation can likely be attributed to the small number of samples combined in the GENNID meta-analysis ( $n = 4$ ). MR-MEGA follows a weighted least squares regression framework, and it may be that four studies are insufficient to meet the assumptions of the model. Simulations and applications on MR-MEGA in Magi et al. (2017) use at least 9 studies, so we are unsure from the existing literature of MR-MEGA as to how many studies are required to produce valid results. In order to compare the results found by each method, we will assume that the suggestive and genome-wide significant results produced by each method are unlikely to be spurious for the rest of the Discussion.

As expected, we observed FE to better detect loci characterized by homogeneous effects across studies and observed RE2 to better detect loci characterized by completely heterogeneous/independent effects across studies. The locus described in Figure 3 is best described by heterogeneous effects since the effect size estimates for the MA and JA samples ( $MA = 0.059$ ,  $JA = 0.009$ ) are quite different from the effect size estimates in the EA and AA samples ( $EA = 0.254$ ,  $AA = 0.205$ ). RE2 was the only method able to declare this locus to be genome-wide significant since the other three meta-analysis methods cannot adequately accommodate completely heterogeneous effects. Even though the two trans-ethnic methods modeled as much heterogeneity between the samples as possible at the locus



described in Figure 3 ( $\rho = 0$  in TransMeta, 1 MDS axis included in MR-MEGA), the trans-ethnic methods could not capture the same effects as RE2 at this locus. This locus is located in the intronic and exonic regions of the gene GCKR, and SNPs in this region have been previously reported in multiple ethnicities for association with metabolic traits (Coram et al., 2013; Jason et al., 2017; Kamatani et al., 2010; Scott et al., 2012). The lead SNP at this locus was also found to be associated with triglycerides in the Global Lipids Genetics Consortium (GLGC) with a p-value =  $2.29e-239$  (Willer et al., 2013). Figure 3 shows that this locus is driven primarily by the EA samples, and this matches with the locus identified by GLGC since the consortium is made up of primarily European samples. This prior evidence supports the validity of the locus shown in Figure 3 and gives us a better understanding of how the trans-ethnic methods behave with loci characterized by heterogeneity.

While TransMeta found more SNPs to be suggestive than the other methods, the additional SNPs detected by TransMeta were typically not independent from the SNPs detected by other methods. TransMeta found more SNPs to be suggestive because it detected additional SNPs at loci not captured by the other methods. This is because TransMeta would produce slightly lower p-values than the other methods at a SNP, leading to a suggestive result by TransMeta while the other method's p-values were only on the order of  $10^{-6}$  or  $10^{-5}$ . However, it should again be noted that slight inflation was observed in the tail of the TransMeta distribution (Appendix C), so it is unclear whether some of the additional suggestive results found by TransMeta are spurious. Since TransMeta contains FE in its hierarchical framework, TransMeta detected most of the loci detected by FE; there was also a large overlap between the loci detected by RE2 and TransMeta (Figure 2). In addition to being able to adequately capture homogeneous and heterogeneous effects, the trans-ethnic approach of TransMeta allows it to more easily detect loci characterized by moderate trans-ethnic effects, such as the locus described in Figure 5. We define the locus described in Figure 5 as having moderate trans-ethnic effects because neither homogeneity nor complete heterogeneity is observed since the effect size estimates of the EA, AA, and MA samples are similar (EA =  $-0.206$ , AA =  $-0.230$ , MA =  $-0.265$ ) and the effect size estimate of JA is quite different (JA =  $0.187$ ). While the other three methods had p-values on the order of  $10^{-6}$  at this locus, TransMeta was able to declare this locus as suggestively significant because it was better able to capture the moderate trans-ethnic effects seen at this locus than the other methods. While this locus was primarily driven by the MA sample in the GENNID sample, this locus showed evidence of replication in the UK BioBank (UKBB) (Mahajan et al., 2018), with the lead SNP at this locus being associated with weight in the UKBB with a p-value =  $1.88e-4$ . This passes a replication threshold of  $\frac{0.05}{17} = 2.94e-3$ , where 17 is the number of independent loci in our results with meta-analysis p-value  $1e-6$ . Thus, there is evidence for replication at this locus in UKBB, providing support for the validity of the locus described in Figure 5.

MR-MEGA appears to be slightly more conservative than the other methods and detected half as many SNPs to be suggestive compared to TransMeta and did not detect any SNPs to be genome-wide significant (Table 1). However, MR-MEGA may be better able to detect loci driven by a single ethnic group compared to the other methods. This observation is also

supported by (Magi et al., 2017) where simulations found that MR-MEGA had the greatest increases in power over FE and RE2 in a simulation scenario characterized by a single ethnic group driven locus. Figure 6 shows that the MDS axis used by MR-MEGA when analyzing the GENNID data set separates the African-American group from the other three ethnic groups. This explains why MR-MEGA best detects loci driven by African-American effects (Figure 4), where all other methods had  $p\text{-value} > 1e-4$  at this locus. Although replication through an independent data set is needed, this locus may represent a novel finding that resulted from increased power from using the trans-ethnic method MR-MEGA.

Generally, the same set of variants were detected by the four meta-analysis methods; however, there were some differences. While it is unclear why the two trans-ethnic methods seem to detect different types of loci, it is possible that the use of slightly different genetic distance metrics to model trans-ethnic effects (see Methods) or how each method integrates trans-ethnic effects into its model structure contributes to these differences. Simulations could help to answer these questions and provide additional information on which to judge performance. Since TransMeta can adaptively choose the appropriate amount of heterogeneity to model and is able to collapse to an FE model, TransMeta may have more power to detect loci across a wider range of effects than MR-MEGA (homogeneous to moderate trans-ethnic effects). However, TransMeta does not seem able to capture the same heterogeneity as MR-MEGA in the GENNID data. This is likely because MR-MEGA always models heterogeneity by permanently including the MDS axes as covariates, where the MDS axes explain the directions of most variation between studies. Thus, MR-MEGA is better able to capture scenarios of heterogeneity characterized by the MDS axes than TransMeta (as seen in the GENNID data set with African-American driven signals). However, since the framework of MR-MEGA is not as flexible as TransMeta, the types of trans-ethnic signals MR-MEGA can detect is limited to the MDS axes included. This can be an issue for meta-analyses with a small number of samples (like the GENNID data set), since the maximum number of MDS axes that can be included is  $n - 3$ .

Additionally, since MR-MEGA always models heterogeneity between studies, it is poorly powered to capture homogeneous effects, and the authors recommend implementing FE in conjunction to MR-MEGA to help address this shortcoming of the method (Magi et al., 2017). Since the process of applying MR-MEGA and FE would call for a multiple testing correction, more study is needed to discern if this dual approach is advantageous to TransMeta in practice. Note that we did not compare the trans-ethnic meta-analysis method MANTRA (Morris, 2011) because TransMeta and MR-MEGA have been shown to be more powerful or comparable to MANTRA across all simulation comparisons (Magi et al., 2017; Shi & Lee, 2016).

In conclusion, we empirically evaluated four trans-ethnic methods and found apparent differences between the methods when applied to the GENNID data set. Simulations are needed to adequately compare the efficiency of the methods under a wider range of ethnic group combinations not captured by the GENNID data. Simulations could better address differences between TransMeta and MR-MEGA by allowing better characterization as to when each method is optimal, which could lead to a better understanding of when each trans-ethnic method appears to be better powered in different scenarios.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

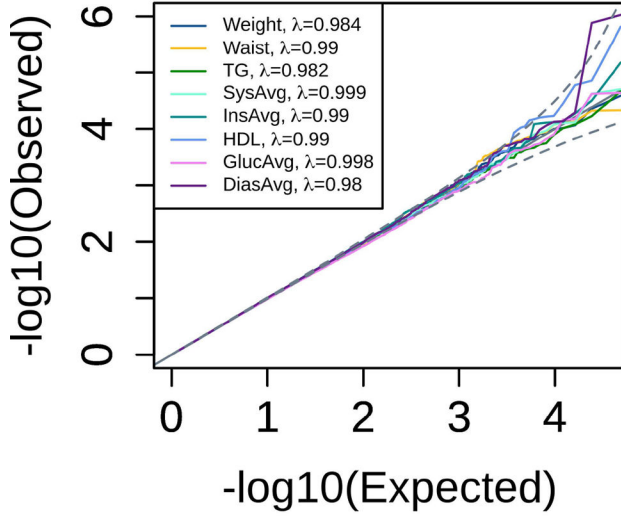
Genotyping services were provided by the Northwest Genomics Center. This study was funded by the National Institutes of Health (NIH R01 HL113189). The GENNID study was supported by the American Diabetes Association.

## References

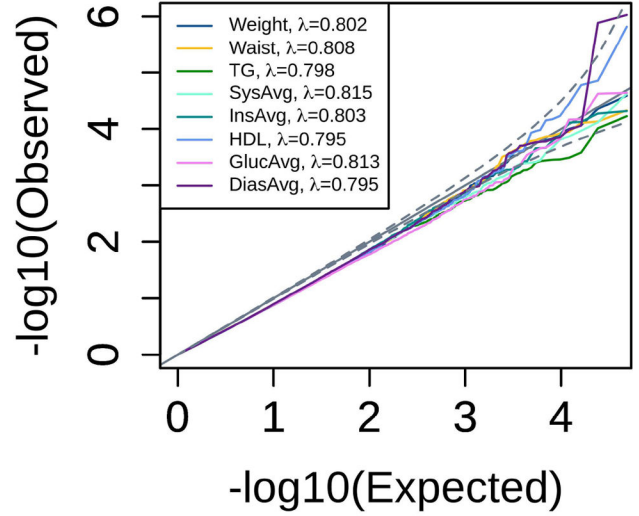
- Abecasis GR, Cherny SS, Cookson WO, & Cardon LR (2001). Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genetics*, 30(1), 97–101. 10.1038/ng786 [PubMed: 11731797]
- Bradburd GS, Ralph PL, & Coop GM (2013). Disentangling the Effects of Geographic and Ecological Isolation on Genetic Differentiation. *Evolution*, 67(11), 3258–3273. 10.1111/evo.12193 [PubMed: 24102455]
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, & Lee JJ (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4(1), 7. 10.1186/s13742-015-0047-8 [PubMed: 25722852]
- Coram MA, Duan Q, Hoffmann TJ, Thornton T, Knowles JW, Johnson NA, ... Tang H (2013). Genome-wide Characterization of Shared and Distinct Genetic Components that Influence Blood Lipid Levels in Ethnically Diverse Human Populations. *The American Journal of Human Genetics*, 92(6), 904–916. 10.1016/j.ajhg.2013.04.025 [PubMed: 23726366]
- Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, ... Fuchsberger C (2016). Next-generation genotype imputation service and methods. *Nature Genetics*, 48(10), 1284–1287. 10.1038/ng.3656 [PubMed: 27571263]
- [dataset] Willems EL, Wan JY, Norden-Krichmar TM, Edwards KL, Santorico SA; 2019; Dataset title TBD; dbGaP; doi TBD
- Gibbs RA, Boerwinkle E, Doddapaneni H, Han Y, Korchina V, Kovar C, ... Rasheed A (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68–74. 10.1038/nature15393 [PubMed: 26432245]
- Han B, & Eskin E (2011). Random-Effects Model Aimed at Discovering Associations in Meta-Analysis of Genome-wide Association Studies. *The American Journal of Human Genetics*, 88, 586–598. 10.1016/j.ajhg.2011.04.014 [PubMed: 21565292]
- Howie B, Marchini J, & Stephens M (2011). Genotype Imputation with Thousands of Genomes. *G3*, 1, 457–469. Retrieved from <http://www.g3journal.org/content/ggg/1/6/457.full.pdf> [PubMed: 22384356]
- Jason F, Fuchsberger C, Mahajan A, Teslovich TM, Agarwala V, Gaulton KJ, ... McCarthy MI (2017). Sequence data and association statistics from 12,940 type 2 diabetes cases and controls. *Scientific Data*, 4, 170179. 10.1038/sdata.2017.179 [PubMed: 29257133]
- Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, Daigo Y, ... Kamatani N (2010). Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nature Genetics*, 42(3), 210–215. 10.1038/ng.531 [PubMed: 20139978]
- Loh P-R, Danecek P, Palamara PF, Fuchsberger C, A Reshef Y, K Finucane H, ... L Price A (2016). Reference-based phasing using the Haplotype Reference Consortium panel. *Nature Genetics*, 48(11), 1443–1448. 10.1038/ng.3679 [PubMed: 27694958]
- Magi R, Horikoshi M, Sofer T, Mahajan A, Kitajima H, Franceschini N, ... Morris AP (2017). Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Human Molecular Genetics*, 0(0), 1–12. 10.1093/hmg/ddx280
- Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, ... McCarthy MI (2018). Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and

- islet-specific epigenome maps. *Nature Genetics*, 50(11), 1505–1513. 10.1038/s41588-018-0241-6 [PubMed: 30297969]
- McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood A, & Al. E (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics*, 48(10), 1279–1283. 10.1038/ng.3643 [PubMed: 27548312]
- Medina-Gomez C, Felix JF, Estrada K, Peters MJ, Herrera L, Kruithof CJ, ... Rivadeneira F (2015). Challenges in conducting genome-wide association studies in highly admixed multi-ethnic populations: the Generation R Study. *European Journal of Epidemiology*, 30(4), 317–330. 10.1007/s10654-015-9998-4 [PubMed: 25762173]
- Morris AP (2011). Transethnic meta-analysis of genomewide association studies. *Genetic Epidemiology*, 35(8), 809–822. 10.1002/gepi.20630 [PubMed: 22125221]
- O’Connell J, Gurdasani D, Delaneau O, Pirastu N, Ulivi S, Cocca M, ... Marchini J (2014). A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness. *PLoS Genetics*, 10(4), e1004234 10.1371/journal.pgen.1004234 [PubMed: 24743097]
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, ... Sham PC (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, 81(3), 559–575. 10.1086/519795 [PubMed: 17701901]
- Raffel LJ, Robbins DC, Norris JN, Boerwinkle E, & et al. (1996). The GENNID Study. A resource for mapping the genes that cause NIDDM. *Diabetes Care*, 19(8), 864–872. Retrieved from <https://pubmed.ncbi.nlm.nih.gov/8842605/> [PubMed: 8842605]
- Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, ... Barroso I (2012). Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nature Genetics*, 44(9), 991–1005. 10.1038/ng.2385 [PubMed: 22885924]
- Shi J, & Lee S (2016). A novel random effect model for GWAS meta-analysis and its application to trans-ethnic meta-analysis. *Biometrics*, 72(3), 945–954. 10.1111/biom.12481 [PubMed: 26916671]
- Speed D, Hemani G, Johnson MR, & Balding DJ (2012). Improved Heritability Estimation from Genome-wide SNPs. *The American Journal of Human Genetics*, 91(6), 1011–1021. 10.1016/j.ajhg.2012.10.010 [PubMed: 23217325]
- Stram DO (2014). Design, Analysis, and Interpretation of Genome-Wide Association Scans. New York: Springer Retrieved from <http://www.springer.com/series/2848>
- Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, ... Abecasis GR (2013). Discovery and refinement of loci associated with lipid levels. *Nature Genetics*, 45(11), 1274–1283. 10.1038/ng.2797 [PubMed: 24097068]
- Yang J, Lee SH, Goddard ME, & Visscher PM (2011). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, 88(1), 76–82. 10.1016/j.ajhg.2010.11.011 [PubMed: 21167468]

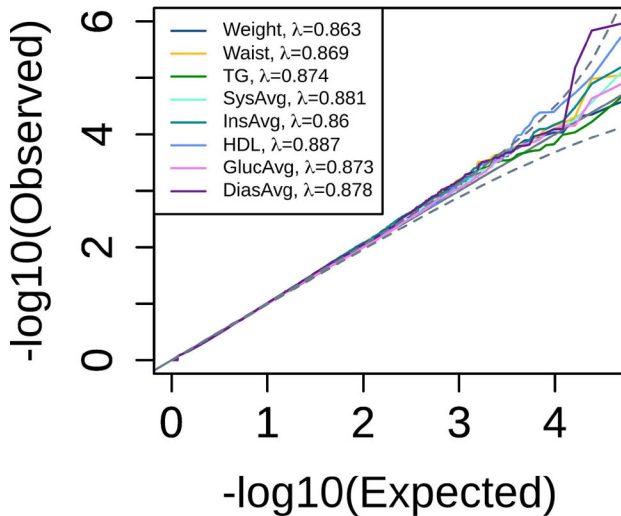
### Fixed Effects



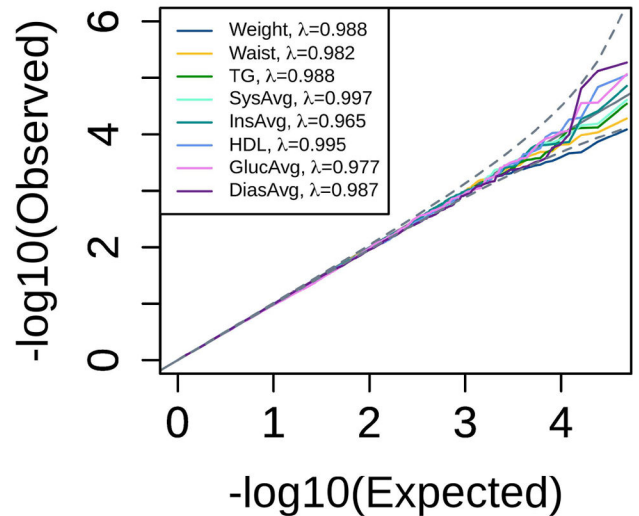
### Random Effects



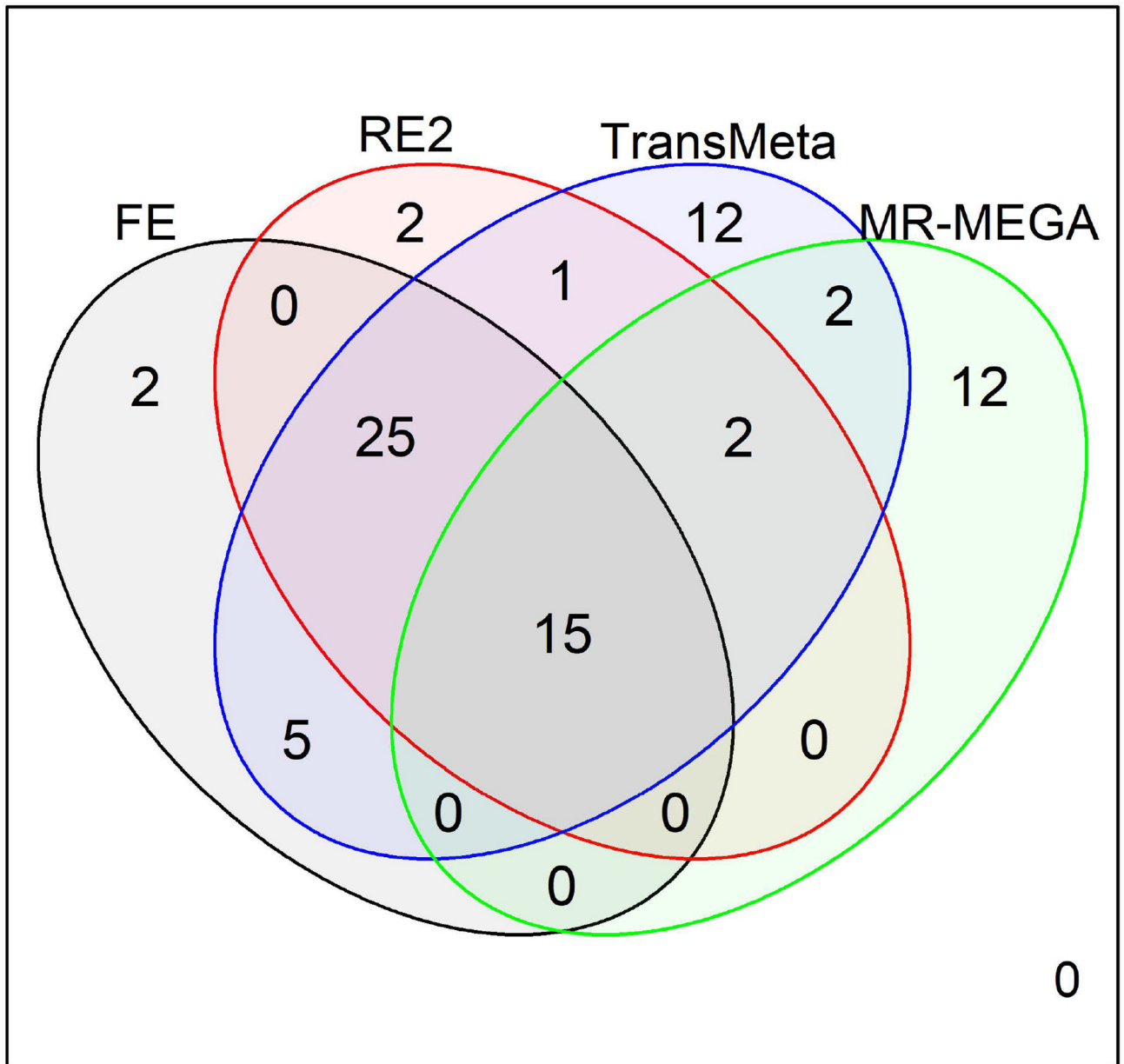
### TransMeta



### MR-MEGA



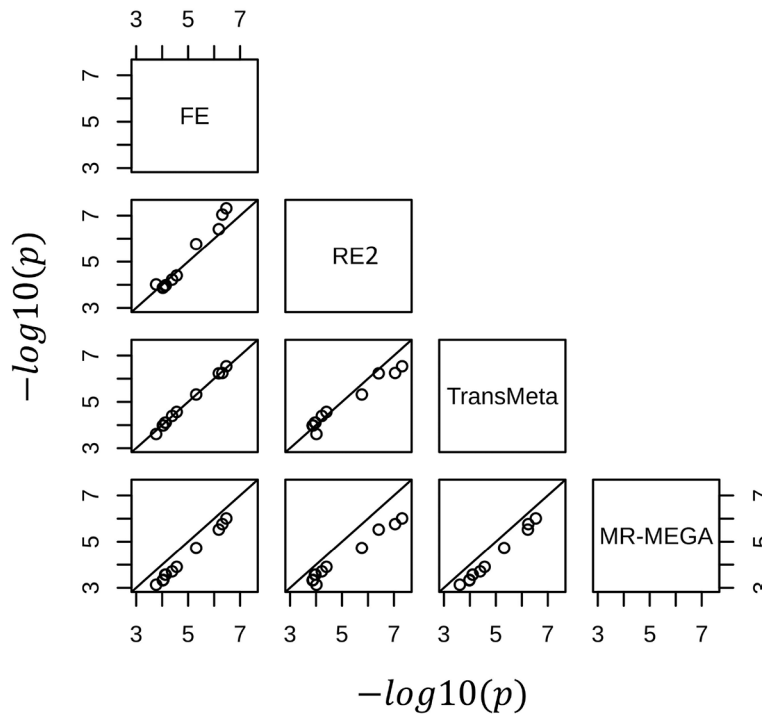
**Figure 1:** QQplots of  $-\log_{10}(p\text{-value})$  for FE, RE, TransMeta, and MR-MEGA meta-analysis methods. The expected line is shown in black and confidence bands are shown in dashed black lines. All 8 phenotypes used are shown using different colored lines (legend). QQplots were generated using a genome-wide LD pruned set of 48,624 SNPs.



**Figure 2:** Venn Diagram of the 78 suggestively significant ( $p$ -value  $\leq 1 \times 10^{-6}$ ) SNPs found by the four meta-analysis methods: FE (black), RE2 (red), TransMeta (blue), and MR-MEGA (green).



### Chr 2, Triglycerides

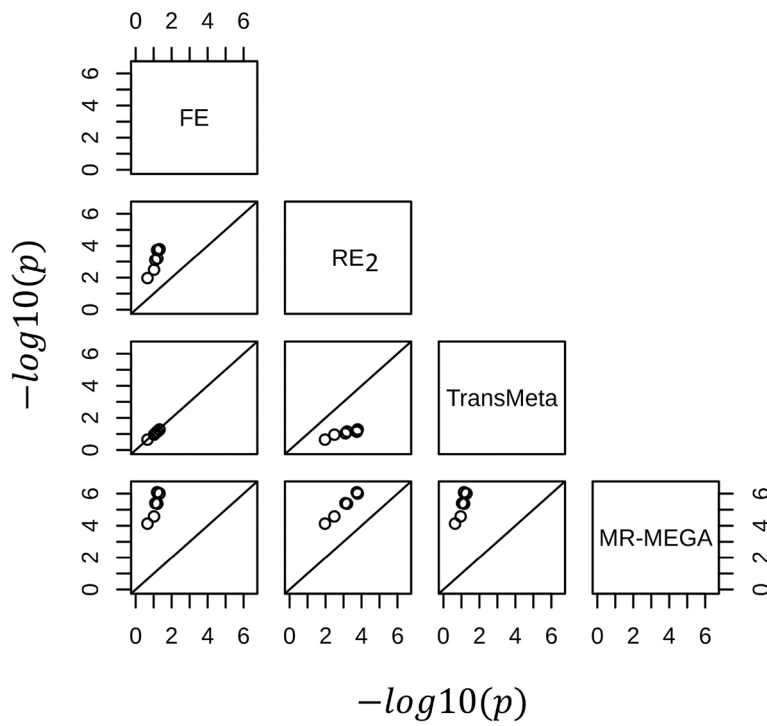


FE p-value	3.39e-7
RE2 p-value	<b>4.86e-8</b>
TransMeta p-value	2.91e-7
MR-MEGA p-value	9.83e-7
EA Effect Size	0.254
AA Effect Size	0.205
MA Effect Size	0.059
JA Effect Size	0.009
EA p-value	1.40e-7
AA p-value	0.005
MA p-value	0.120
JA p-value	0.944

**Figure 3:**

The left figure shows pairwise scatterplots of the  $-\log_{10}(p\text{-value})$  for all SNPs with  $p\text{-value} < 1e-4$  by at least one meta-analysis method for a locus at 27,598,097 – 28,774,081 bp on chromosome 2 associated with triglycerides (11 SNPs total). RE2 is better able to detect this locus at the genome-wide significance threshold of  $5e-8$  than the trans-ethnic methods (TransMeta, MR-MEGA). The right table shows the summary statistics (effect sizes and p-values) from each ethnic group for the SNP with the lowest p-value by any of the four meta-analysis methods. The meta-analysis p-values for each method are also shown for the leading SNP, with the lowest meta-analysis p-value (by RE2) in bold. Base pair coordinates are in Build 37. See Supplemental Figure S1 and Supplemental Table D.1 for more information about this locus.

### Chr 19, Triglycerides

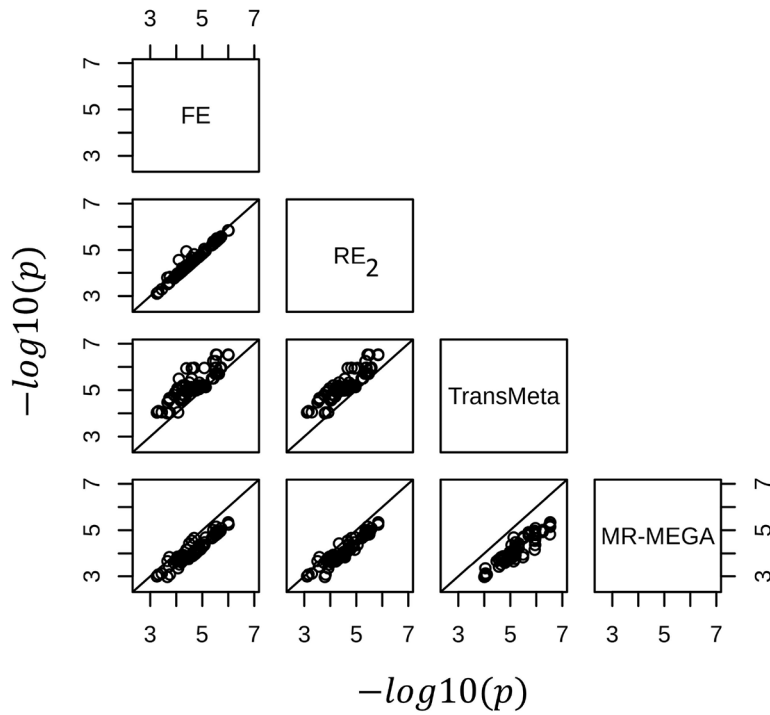


FE p-value	0.064
RE2 p-value	1.83e-4
TransMeta p-value	0.072
MR-MEGA p-value	<b>8.29e-7</b>
EA Effect Size	0.013
AA Effect Size	0.633
MA Effect Size	0.024
JA Effect Size	-0.112
EA p-value	0.816
AA p-value	1.60e-7
MA p-value	0.593
JA p-value	0.532

**Figure 4:**

The left figure shows pairwise scatterplots of the  $-\log_{10}(p\text{-value})$  for all SNPs with  $p\text{-value} < 1e-4$  by at least one meta-analysis method for a locus at 38,030,933 – 38,097,313 bp on chromosome 19 associated with triglycerides (14 SNPs total). MR-MEGA is better able to detect the locus at the suggestive threshold of  $1e-6$  than TransMeta. The right table shows the summary statistics (effect sizes and p-values) from each ethnic group for the SNP with the lowest p-value by any of the four meta-analysis methods. The meta-analysis p-values for each method are also shown for this leading SNP, with the lowest meta-analysis p-value (by MR-MEGA) in bold. Base pair coordinates are in Build 37. See Supplemental Figure S2 and Supplemental Table D.1 for more information about this locus.

### Chr 2, Weight

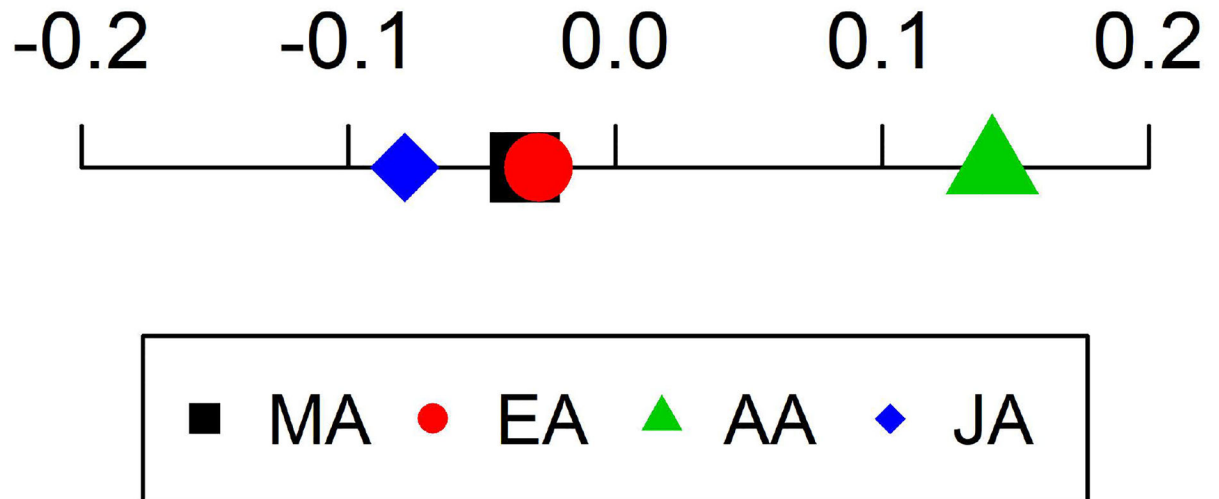


FE p-value	1.01e-6
RE2 p-value	1.43e-6
TransMeta p-value	<b>2.93e-7</b>
MR-MEGA p-value	4.60e-6
EA Effect Size	-0.206
AA Effect Size	-0.230
MA Effect Size	-0.265
JA Effect Size	0.187
EA p-value	0.008
AA p-value	0.097
MA p-value	5.62e-6
JA p-value	0.208

**Figure 5:**

The left figure shows pairwise scatterplots of the  $-\log_{10}(p\text{-value})$  for all SNPs with  $p\text{-value} < 1e-4$  by at least one meta-analysis method for a locus at 161,435,255 – 161,916,409 bp on chromosome 2 associated with weight (107 SNPs total). TransMeta is better able to detect the locus at the suggestive threshold of  $1e-6$  than MR-MEGA. The right table shows the summary statistics (effect sizes and p-values) from each ethnic group for the SNP with the lowest p-value by any of the four meta-analysis methods. The meta-analysis p-values for each method are also shown for this leading SNP, with the lowest meta-analysis p-value (by TransMeta) in bold. Base pair coordinates are in Build 37. See Supplemental Figure S3 and Supplemental Table D.1 for more information about this locus.

# First MDS axis used by MR-MEGA



**Figure 6:** First MDS axis of the genetic distance matrix calculated by MR-MEGA. The axis separates the African-American group from the other three groups, allowing MR-MEGA to better detect loci driven by African-American effects.

**Table 1:**

## Summary of Meta-analyses Results

Method	Type of Effects Best Detected	# SNPs w/P $1e - 6$ (suggestive significance)	# SNPs w/P $5e - 8$ (genome-wide significant)	# of times method had the lowest P
<b>Fixed Effects</b>	Homogeneous Effects	47	4	11
<b>Random Effects</b>	Independent Effects	45	3	6
<b>TransMeta</b>	Trans-ethnic effects	62	3	48
<b>MR-MEGA</b>	Trans-ethnic effects	31	0	13

\* Table counts are found by counting how many of the 3,153,931 SNPs reached suggestive or genome-wide significance by each method. Of the 3,153,931 SNPs, 78 SNPs were found to be suggestive by at least one method. Column 5 counts the number of times each method had the lowest p-value for the 78 SNPs.