Published in final edited form as:

Genet Epidemiol. 2014 February ; 38(2): 162–171. doi:10.1002/gepi.21766.

A Robust Method for Genome-wide Association Meta-Analysis with the Application to Circulating Insulin-like Growth Factor I Concentrations

Tao Wang¹, Baiyu Zhou¹, Tingwei Guo¹, Martin Bidlingmaier², Henri Wallaschofski³, Alexander Teumer⁴, Ramachandran S. Vasan⁵, and Robert C. Kaplan¹

¹ Albert Einstein College of Medicine, Bronx, NY 10461, USA

² Institute of Epidemiology, Helmholtz Zentrum München, Neuherberg 85764, Germany

³ Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, 17475 Germany

⁴ Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, 17475 Germany

⁵ Boston University School of Medicine, Boston, MA, 02118, USA

Abstract

Genome-wide association studies (GWAS) offer an excellent opportunity to identify the genetic variants underlying complex human diseases. Successful utilization of this approach requires a large sample size to identify single nucleotide polymorphisms (SNPs) with subtle effects. Meta-analysis is a cost-efficient means to achieve large sample size by combining data from multiple independent GWAS; however, results from studies performed on different populations can be variable due to various reasons, including varied linkage equilibrium structures as well as gene-gene and gene-environment interactions. Nevertheless, one should expect effects of the SNP are more similar between similar populations than those between populations with quite different genetic and environmental backgrounds. Prior information on populations of GWAS is often not considered in current meta-analysis methods, rendering such analyses less optimal for the detecting association. This paper describes a test that improves meta-analysis to incorporate variable heterogeneity among populations. The proposed method is remarkably simple in computation and hence can be performed in a rapid fashion in the setting of GWAS. Simulation results demonstrate the validity and higher power of the proposed method over conventional methods in the presence of heterogeneity. As a demonstration, we applied the test to real GWAS data to identify SNPs associated with circulating Insulin-like growth factor I concentrations.

Keywords

Genome-wide association study; Meta-analysis; Variance-component model; Insulin-like growth factor I

Introduction

Genome-wide association studies (GWAS) are now a standard method for disease gene discovery[2007; Saxena, et al. 2007; Scott, et al. 2007; Sladek, et al. 2007; Steinthorsdottir, et al. 2007; Zeggini, et al. 2007a]. Most array-based GWAS are designed to detect common variations (i.e., single nucleotide polymorphisms (SNPs) with minor allele frequencies greater than 1%). Although an increasing number of genetic variants have been identified for many traits such as body mass index (BMI)[Lindgren, et al. 2009; Scherag, et al. 2010; Willer, et al. 2009] as well as diseases such as Type II diabetes[Frayling 2007] and breast cancer[Fanale, et al. 2012], the GWAS findings to date explain only a small proportion of heritability for most complex human traits and diseases. Attention is now shifting to rare variants and common variants that interact with other genetic variants and environmental exposures[Manolio, et al. 2009; McCarthy and Hirschhorn 2008; Pritchard 2001]. Identification of such variants requires a very large sample size to achieve sufficient statistical power.

Meta-analysis is a cost-efficient method to boost sample size by combining data from multiple independent studies, and has already successfully identified many significant genome-wide associations that otherwise would have been missed by single studies' [Barrett, et al. 2009] [Scott, et al. 2007; Zeggini, et al. 2007b] [Scott, et al. 2009] [Franke, et al. 2010]. Nevertheless, conventional methods of meta-analysis, such as the fixed effects model and random effects model were developed in the setting of clinical trials, in which one expects effects of a treatment to be consistent across individual trials. Inconsistent, even opposite, effects observed in different trials often cause concern about random variations or systematic biases. Hence, random effects model is preferred in clinical trial as this method accounts for unexpected heterogeneity of treatment effects, and fixed effects model is often used only after testing for heterogeneity[Normand 1999; van Houwelingen, et al. 2002].

Rather than estimating effect sizes of a treatment, GWAS aim to identify new disease genes and hence have many different features compared to clinical trials. One important difference is that in GWAS, we expect some “genuine” heterogeneity of effects of a genetic marker between studies, which does not disappear after excluding potential biases, such as population stratification and genotyping errors. There are several sources of “genuine” heterogeneity. Often, GWAS are conducted in populations with varying genetic and environmental backgrounds. One major sources of heterogeneity may be gene-gene and gene-environment interactions. In the presence of interactions, the effect sizes of a genetic variant vary depending on the genetic and environmental backgrounds[Hirschhorn and Daly 2005; Hunter 2005; Ioannidis, et al. 2007; Moonesinghe, et al. 2008]. Another source of heterogeneity is due to the fact that SNPs in genotyping panels of GWAS may not be the underlying causal variants, but may be linked to causal variants in variable linkage disequilibrium across different populations[Han and Eskin 2011; Hirschhorn and Daly 2005]. Heterogeneity can also occur because traits are measured or diseases are defined differently in different studies.

To account for heterogeneity in GWAS meta-analysis, a new likelihood ratio test (LRT) based on the random-effects model was recently proposed [Han and Eskin 2011; Lebrech, et

al. 2010]. This method improves the power of conventional random effects model by allowing effects of a SNP of interest to vary under the alternative hypothesis. Under this approach, the effects of a SNP are modeled by a normal distribution with the mean of μ and the variance of τ^2 . Assuming there is no bias in any single study, both the mean and variance of effects of a SNP is zero under the null hypothesis of no association. As such, this test tests $\mu=0$ and $\tau^2=0$ instead of testing only the average effect as conventional random effects method, and hence can improve power in the presence of high levels of heterogeneity, i.e. τ^2 is not small. By simulations, this new test has been shown to be substantially more powerful than conventional fixed effects model and random effects model in the presence of heterogeneity[KY 1987]. Asymptotically, the LRT is distributed as $0.5\chi^2_{df=1} + 0.5\chi^2_{df=2}$ under the null hypothesis, because $\tau^2=0$ is on the boundary of the parameter space. However, the asymptotic results of this test hold only when the number of GWAS is large, which is often not the case. As a result, the asymptotic p-value is overly conservative and has to be evaluated empirically by extensive computational simulations or tabulated values, if available[Han and Eskin 2011]. More importantly, this method implies the same level of heterogeneity between individual GWAS in meta-analysis. However, one may expect that causal SNPs have more similar effects in populations with similar known genetic and environment factors, e.g. sex, age and ethnicity, etc. than in populations with different background. For example, in the GWAS meta-analysis of meiotic recombination, the similarity of populations may be defined by sex as effects of causal SNPs in males and females are likely to be different [Begum, et al. 2012].

In this report, we propose a new test to account for heterogeneity of SNP effects between GWAS for meta-analysis. In this method, the effects of a SNP are also modeled as a random variable, but the association is examined via testing the corresponding variance component only. In general, the efficiency of the variance component test is challenged by the large number of degrees of freedom that increases with the number of studies. To balance the trade-off between the gain in power by accounting for heterogeneity and penalty on efficiency because of extra degrees of freedom in the variance component test, we propose to smooth SNP effects between different studies using a correlation matrix that is designed based on prior population information. That is, the design matrix of the random effects of a SNP between studies does not necessarily have a diagonal structure as usual variance-component analysis. With this design matrix, we are able to incorporate prior population information into meta-analysis. One additional advantage of testing the variance component is that the proposed test asymptotically has a weighted sum of chi-square distributions and this asymptotic result holds regardless of the number of GWAS included in analysis. As such, the p-value of the proposed method can be quickly and accurately evaluated and does not rely on computationally extensive simulations.

We refer to the proposed method as the “VC” test (for variance components). This method is easy to implement and computationally fast yet sufficiently feasible to accommodate varied heterogeneity of SNP effects between GWAS. Through simulation, we demonstrate the validity of the proposed method and compared its power performance with the conventional fixed effects model, random effects model and the new LRT methods in general scenarios. We also apply the new method to a real GWAS dataset (with data concerning circulating

Insulin-like growth factor-I (IGF-I)), comprising 1,712 participants in the Cardiovascular Health Study (CHS), 3,507 participants in the Framingham Heart Study (FHS), 1,607 participants in the Cooperative Research in the Region of Augsburg (KORA) study and 3,454 participants in the Study of Health in Pomerania (SHIP)[Kaplan, et al. 2011]. We verify the findings from the proposed method by comparing them with the results reported by Kaplan et al. [Kaplan, et al. 2011]

Methods

We consider a setting where there are m GWAS. We assume the parameters of the effect of a SNP in m studies are $\beta_1, \dots, \beta_m$ and the corresponding estimates are $\hat{\beta}_1, \dots, \hat{\beta}_m$, which can be the estimated regression coefficients of linear regression or logistic regression models. Let $SE(\hat{\beta}_j)$ be the standard error of $\hat{\beta}_j$ and $w_j = SE(\hat{\beta}_j)$ be the inverse of the standard error, where $j \in (1, \dots, m)$. In genetic association studies, we consider the hypothesis of $H_{0j}: \beta_j = 0$ for each study. For GWAS meta-analysis, the entire parameter vector of the SNP effect is $\beta = (\beta_1, \dots, \beta_m)^T$, and the null hypothesis can be written as $H_0: \beta_1 = \dots = \beta_m = 0$ and the alternative hypothesis $H_A: \beta_j \neq 0$ for at least one study.

To allow for heterogeneity between studies, we consider a random effects model assuming that β_j follows a normal distribution with a common variance as $\beta_j \sim (0, \tau^2)$. Under the null hypothesis, effects of all studies are expected to be zero, leading to $\tau^2 = 0$. Hence, the null hypothesis can be tested by examining a single variance component parameter, i.e. $H_0: \tau^2 = 0$ versus $H_1: \tau^2 > 0$.

With an assumption that the effects of a SNP in different GWAS studies are independent, β has a diagonal correlation matrix. We have $\hat{\tau}^2 = \sum_{j=1}^m \hat{\beta}_j^2 / m = \hat{\beta}^T \hat{\beta} / m$ as an empirical estimate of τ^2 . Hence, a natural test statistic of meta-analysis for testing the hypothesis of $H_0: \tau^2 = 0$ is

$$T_\tau = \hat{\beta}^T \hat{\beta}.$$

However, this test may not be efficient as $\hat{\beta}_j$ s are estimated from data and hence subject to different levels of random variation. One may consider weighting $\hat{\beta}_j$ to account for the variation in the estimates by defining

$$T_\tau = \hat{\beta}^T W W^T \beta$$

where W is an $m \times m$ diagonal weight matrix with the diagonal elements being the corresponding weights. Different weights may be considered. For example, $\hat{\beta}_j$ may be weighted by its inverse standard error, w_j .

The above test statistic may still not result in optimal power in the setting of GWAS meta-analysis because they assume effects of a specific SNP are completely independent, leading

to a large number of degrees of freedom of the test. Indeed, the effects of a marker, to some extent, are expected to be consistent at least in similar populations. Inconsistent results from studies performed in populations with close genetic and environmental background are unlikely due to a genuine genetic association, but more likely due to random variations or systematic errors in some studies. Instead of assuming completely independent effects among studies, we want to define a more parsimonious test statistic with greater power against the alternative hypotheses such that effects of individual studies with similar populations have a certain degree of consistency compared to alternatives in which individual studies have uncorrelated, even opposite effects.

To do this, we consider that SNP effects have a multivariate normal (MN) prior, i.e. $\beta \sim MN(0, \tau^2 R_\beta)$, where R_β is the correlation matrix of effects of a SNP between studies. As such, “genuine” heterogeneity of effects of a SNP between two studies is defined by the corresponding correlation. That is the lower a correlation coefficient the higher level of heterogeneity between two studies. Still, the null hypothesis of no association can be examined by testing the variance component of τ^2 . Given a pre-specified correlation structure R_β , the corresponding variance-component test is given by

$$T_{VC} = \hat{\beta}^T W R_\beta W^T \beta.$$

When the weight of a study is w_j , this test can be simplified as

$$T_{VC} = Z^T R_\beta Z,$$

where Z is the vector of Z scores of studies. Conceptually, the (j,k) element of the $m \times m$ correlation matrix R_β , denoted by $r_{j,k}$, should be defined by the similarity of effects of a SNP between two studies, which is usually unknown. However, one may specify $r_{j,k}$ based on prior information about population similarity of two studies in terms of genetic and environmental backgrounds.

Regardless the choice of R_β it can be easily demonstrated that the proposed tests have the

weighted χ^2 random variables defined by $\sum_{j=1}^k \lambda_j \chi_1^2$, where χ_1^2 is an χ^2 random variable with 1 df, and λ_j is the ordered nonzero eigenvalues of the semipositive definite matrix of R_β , with $\lambda_1 \lambda_2 \dots \lambda_k > 0$ ($k = m$). This VC test reject the hypothesis of $H_0: \tau^2 = 0$ if the tests exceed the threshold of T_α , the (1- α)th quantile of the weighted χ^2 distribution. Given that T_{VC} has a weighted sum of χ_1^2 distributions, we can evaluate the p-value of T_{VC} by simulations. However, because the statistical significance level of GWAS is often very low, e.g. $\alpha = 10^{-8}$, extensive computation may be required to obtain an accurate p-value. Based on the work of Kuonen, we propose use of the “saddle point” approximation. The saddle point approximation of quadratic forms in normal variables at q is

$$Pr(T_{VC} > q) \approx 1 - \Phi \left\{ \omega + \frac{1}{\omega} \log \left(\frac{v}{\omega} \right) \right\},$$

Where $\omega = \text{sign}(\hat{\xi}) [2\hat{\xi}q - K(\hat{\xi})]^{1/2}$ and $v = \hat{\xi} \{K''(\hat{\xi})\}$ with $K(\hat{\xi})$ being the cumulant generating function and $\hat{\xi}$, known as the saddle point, being the value of ξ satisfying the equation $K'(\hat{\xi}) = q$ (Primes denote derivatives). As their relative errors of saddle point approximations stay bounded in the extreme tails, it could provide accurate estimates of p-values even out in the very far upper tail of the distribution [Kuonen 1999].

Figure 1 is a simple illustration of the rejection regions of fixed effects model, random effects model, the VC method assuming various correlations of effects between two studies. We can see that fixed effects model and random effects model are very sensitive to heterogeneity of effects between studies. This is particularly prominent in random effects model, in which power diminishes very rapidly with increasing heterogeneity. In contrast, the efficiency of the VC test is more robust in cases with heterogeneity of effects. Particularly for the VC test assuming $R_\beta = I$, if one study already yields a test statistic exceeding the significance threshold, the results of meta-analysis will maintain significance regardless of whether the results of another study are similar or opposing.

Although the choice of R_β does not affect the validity of the proposed VC test, it is clear that different prior correlation structures of R_β yield different weights for χ^2 distributions, leading to tests that are sensible to different alternative hypotheses. In this sense, there is no uniformly best design correlation matrix for all associated SNPs. To understand the VC test, it may be useful to rewrite it in the form of

$$T_{VC} = \left(\sum_j Z_j \right)^2 - 2 \sum_{j,k} (1 - r_{jk}) Z_j Z_k,$$

where $j=1, \dots, m$ and $k \neq j$. The first part of $(\sum_j Z_j)^2$ can be looked upon as a statistic for the average effects across studies, and the second part of $-2 \sum_{j,k} (1 - r_{jk}) Z_j Z_k$ is a statistic for heterogeneity of effects in which r_{jk} is the tuning parameter to determine the contribution from heterogeneity. Hence, the VC test assuming $R_\beta = I$ ($r_{jk} = 0$) with the most contribution from heterogeneity may not be powerful in the case of no heterogeneity as it unnecessarily involves a large number of extra degrees of freedom (or more generally, extra variance) of the test statistic. To obtain a robust test, it is helpful to balance the contribution from the average effect and heterogeneity. Based on the similarity of populations of different studies, we may define r_{jk} so that the extra variance due to testing heterogeneity is 0.5-1 fold of that testing the average effect. By simulations, we found the test with r_{jk} in this range had a good performance under a wide range of scenarios. In our simulations and data analysis below, the extra variance for testing heterogeneity is fixed at 0.75 fold of that testing the average effect.

Results

Simulation Studies

We study performance of the proposed test of association by simulating a case-control meta-analysis. Assuming rare disease and a SNP with Hardy-Weinberg equilibriums, we directly sampled the genotypes of controls and cases from a binomial distribution with the event rate of $p_{0j} = p_j$ and $p_{1j} = \frac{\gamma p_j}{(\gamma-1)p_j+1}$ where p_j is the allele frequency and γ_j is the relative risk of population j . The estimated log odds ratios and their standard errors are then used for meta-analysis.

We first performed a simulation study to confirm that the proposed test can maintain a good control of type I error rate. For each scenario, we performed 100,000 repeats and estimate the empirical type I error rate by the proportion of repetitions with p-values greater than α when $\alpha = 0.05$ and 0.01 . In our simulations, we assumed the minor allele frequency was $p_{0j} = p_{1j} = 0.2$, the number of GWAS in meta-analysis was 3, 5, 10 and 20, and studies either have an equal sample size (the sample size was fixed at 1000) or unequal sample sizes (the sample sizes were randomly sampled from 500, 1,500, and 2,000). In the proposed VC method, all pairs of studies were assigned to have a common prior correlation.

Table 1 shows empirical type I error rates of the proposed VC method at significance levels of 0.05 and 0.01. We can see that the VC test was around the nominal level, indicating that our approximation of the null distribution is adequate. As expected, fixed effects performed well in maintaining the nominal significance level, while random effects model tended to be conservative because the variance between studies were inaccurately estimated; this was particularly true when the number of studies was small.

We then examined if different prior correlations has an impact on empirical type I error rates of the proposed VC method. Figure 2 shows the empirical type I error rates of the proposed VC method with different prior correlations from 0 to 1 in a setting of five studies with equal and unequal sample sizes. We can see the VC method maintained good control of type I error rates regardless of the prior correlation. The proposed method also maintained type I error rates when varied prior correlations were assigned to different pairs of studies (data not shown).

In GWAS, the type I error rate is usually determined at a very stringent threshold to control for multiple testing. To examine if the proposed approximation is adequate when the p-value is examined in the far right tail of the null distribution, we performed 50 million replicates under the null hypothesis with simulation settings as described above. Table 2 shows the empirical type I error rate of the VC method at increasing significance levels. The VC method maintained good control of the type I error rate, although they tended to be slightly conservative in some settings.

Power

To evaluate the empirical power of the different methods, we simulated data for five studies under various settings. Empirical power was estimated by using 10,000 replicates as the

proportion of the replicates whose p-value exceeds a genome wide threshold of 10^{-6} . Because the asymptotic p-values of random effects model and LRT are overly conservative, cut-offs of the significance for those methods were determined by the empirical null distribution of 50 million simulations. Significance thresholds for the proposed method were obtained by the asymptotic-based method, which requires approximation of weighted χ^2 distributions.

To simulate heterogeneous effects, we first evaluated the empirical power of different methods by assuming that the true effect sizes of different studies were sampled from a normal distribution with various variances. Specifically, we sampled the log odds ratios γ of different studies from a common normal distribution of $N(\log(\gamma), k\log(\gamma))$, where γ is the average odds ratio across studies and k is a parameter of heterogeneity. We set k between 0 to 1 to mimic heterogeneity from zero to a high level. We simulated datasets under two different settings. In the first setting, we assumed studies had an equal sample size of 1,000. In the second setting, we assumed studies had unequal sample sizes randomly sampled from 500, 1,500 or 2,000 individuals. In the proposed VC method, a common prior correlation was specified for all pairs of studies.

Figure 3 (a) shows the empirical power of different methods at the statistical significance level of 10^{-6} when the effect sizes of studies were sampled from a common normal distribution and studies had an equal sample size. As expected, random effects model was the worst performer, having particular difficulty in the presence of a high level of heterogeneity. In the setting with no heterogeneity among studies ($k=0$), fixed effects performed best, but its power decreased with increasing heterogeneity. Compared with fixed effects method, the performance of the VC method had a robust performance. In the setting of no or low heterogeneity, the proposed VC method, although not the best, was a close second. With increasing heterogeneity, the VC method greatly outperformed fixed effects method, in particular when heterogeneity is high. Compared LRT, the VC method was more powerful under most conditions, except when heterogeneity is very high ($k>0.8$).

Figure 3 (b) shows the empirical power of different methods at the statistical significance level of 10^{-6} when the effect sizes of studies were sampled from a common normal distribution and studies had unequal sample sizes. When the VC methods was compared to fixed effects method and LRT similar trends of performance were observed as in the case of equal samples.

The major advantage of the proposed VC method over LRT is the feasibility to incorporate prior information of effects between studies through assigning the correlation structure in meta-analysis. To verify this, we performed simulation of five studies with equal sample size of 1,000. Among five studies, we assumed subjects of two studies were sampled in one population, and subjects of the other three studies were sampled from another population. We assumed the genotyped SNP marker was variable in terms of linkage disequilibrium (LD) with an untyped causal variant in two populations. We considered three settings (Table 3), corresponding to the following conditions: (1) the model with no heterogeneity, which assumes that the effects of the genotyped SNP marker are consistent across studies because of the same linkage disequilibrium between the SNP marker and the causal variant in both

populations; (2) the heterogeneous model with the same direction, which assumes the effects of the genotypes are variable but having the same direction; and (3) the cross-over model, which assumes the effects of the genotypes have opposite directions in two populations. To account for potential heterogeneity both within a population and between two populations, we defined the prior correlation design matrix of the VC test by fixing: (1) the correlation within a population to be five times of that between two populations; and (2) the total extra-variance due to testing heterogeneity to be 0.75 fold of that for testing the average effect..

Figure 4 shows the empirical power of different methods for testing the association of the disease with a SNP marker at a significance level of 10^{-6} under different settings of LD structures between two populations. When the effects of the SNP marker were purely homogeneous, fixed effects method, which assumes no heterogeneity, had the highest power. In this setting, both the VC and LRT test were only slightly less powerful than fixed effects method. When the effects of the SNP marker were variable, but in the same direction for two populations, the VC method, which account for heterogeneity between two populations, clearly outperformed fixed effects method and LET. Finally, under the crossover model, in which the opposite linkage disequilibrium coefficients between the marker and the causal variant in two populations leads to SNP effects in opposite directions, fixed effects model was far inferior to other methods because it tests the average effect only. In contrast, LRT and the VC method maintained power. Because random effects model performed significant worse in all settings, its power is not described here.

Application to GWAS data of IGF-I

Insulin-like growth factor I (IGF-I) is a central mediator of metabolic, endocrine and anabolic effects of growth hormone. IGF-I promotes cell proliferation and differentiation, inhibits apoptosis and has insulin-like metabolic effects[Le Roith 1997]. In population-based studies, circulating IGF-I is associated with longevity[Kenyon 2010], cancer[Rehman, et al. 2004; Samani, et al. 2007] and common chronic diseases[Friedrich, et al. 2009; Juul, et al. 2002]. Prior work has shown high heritability for blood IGF-I levels (about 40–60%) [Harrela, et al. 1996; Hong, et al. 1996; Souren, et al. 2007]. Thus, there is interest in identification of genetic variants associated with circulating IGF-I concentrations. We reanalyzed a real GWAS dataset of IGF-1 using the proposed VC methods. This GWAS dataset includes 10,280 men and women, comprising 1,712 participants in the Cardiovascular Health Study (CHS), 3507 participants in the Framingham Heart Study (FHS), 1,607 participants in the Cooperative Research in the Region of Augsburg (KORA), and 3454 participants in the study of Health in Pomerania[Kaplan, et al. 2011]. For the comparison, we also included results from fixed effects model. We did not include LRT because the asymptotic p-values are not accurate and simulation of the empirical null distribution is computationally prohibitive for a large number of SNPs at the stringent significance level in GWAS. In this dataset, all subjects were of European or Caucasian origin as ascertained by self-report. Further, since IGF-1 measurements were performed by each cohort study group using different assays[Kaplan, et al. 2011], we expected a certain amount of heterogeneity between studies and hence assigned a common prior correlation to allow heterogeneity.

Figure 5 and 6 present the q-q plot and Manhattan plot of the genome-wide distribution of results from the VC method. Table 4 list all SNPs variants identified in meta-analysis with a p value $< 1 \times 10^{-6}$. Compared to fixed effects model, the VC method yielded the most SNPs with $p < 1 \times 10^{-6}$. Specifically, the VC method yielded three SNPs (rs700753, $p=8.92\text{e-}10$; rs700752, $p=1.52\text{e-}9$; and rs700750, $p=1.63\text{e-}8$) located at the same gene region that achieved genome-wide association, while fixed effects model yielded two significant SNPs with less significant p-values (rs700753, $p=2.18\text{e-}9$ and rs700752, $p=3.19\text{e-}8$).

Three additional regions achieved $p < 10^{-6}$ in relation to IGF-1 concentrations: the FOXO3 gene region on Chromosome 6; a region on Chromosome 10 close to SPOCK2, and a region on Chromosome 10 close to RPA3. The VC method tended to yield more SNPs with $p < 10^{-6}$ (In the FOXO3 gene region, VC yielded 13 SNPs, while fixed effects model yielded 5 SNPs; In the region close to SPOCK2, VC yielded 7 SNPs while fixed effects model yielded four SNPs). The results of our analysis of the real dataset was consistent with the simulation results in that VC had the better performance when the sample sizes varied for different studies in the presence of heterogeneity of effect sizes of SNPs.

Discussion

In the present study, we extend conventional methods of meta-analysis under the framework of a variance-component model to utilize prior information on study populations for improving power of GWAS meta-analysis. There are various analytic methods that were developed for GWAS meta-analysis [Begum, et al. 2012]. Both fixed effects and random effects models rely on the average effect of multiple studies, ignoring second order information about variance of effects. Therefore, they tend to lose power in the presence of heterogeneity, especially when the effects of a SNP of interest have opposite directions among multiple studies. It may be useful for testing variance of effects for improving power, while it leads to additional degrees of freedom. Thus, there is a trade-off between the gain in information from heterogeneity and the loss of efficiency due to larger degrees of freedom. Nevertheless, one may have prior information on population similarities and expect that effect sizes of a SNP are more similar in studies performed in the similar populations than those for studies performed in populations that are quite different in genetic and environment background. Currently, to incorporate heterogeneity, LRT depends on a likelihood function, which models both average effect and heterogeneity [Han and Eskin 2011; Lebrech, et al. 2010]. One drawback of the current LRT method is that it does not utilize such prior information about populations in meta-analysis, and uses the same number of degrees of freedom for testing heterogeneity regardless of whether studies are performed in the same population or different populations. The proposed method in this paper naturally incorporates such prior information by assigning a correlation structure for effects between studies, and in simulations, we show that this substantially improve powers over conventional methods.

Another advantage of the proposed method is its remarkably simplicity in computation. First, the VC method does not require estimating the average effect of a SNP and the variance introduced by heterogeneity. More importantly, the asymptotic null distribution of

the test statistic can be approximated quickly and simply and it avoids extensive simulation to evaluate the empirical null distribution.

This proposed test is constructed based on estimates of effect sizes and standard errors of individual studies. When standard errors are not available, the methodology can be easily generalized to use only p-values by transforming p-values to z-scores for calculation of the VC test statistic. As such, the proposed method can also be considered as an approach to combine p-values. The conventional Fisher's method transforms p-values to χ^2 scores with 2 degrees of freedom by a scaled negative natural logarithm transformation, and uses the sum of χ^2 scores for testing association[FISHER 1932]. Essentially, Fisher's method assumes effects of a SNP in all studies are independent. Although this test is always valid as the assumption of independency is true under the null hypothesis of no association, it may not always be optimal in terms of power as the assumption is often not true under the alternative hypothesis. The proposed method is also similar to Stouffer's Z score test, which is based on turning the p-values into Z-scores and uses the un-weighted or weighted average of Z-scores as the test statistic[STOUFFER 1949]. In contrast to Fisher's method, Stouffer's Z score test completely ignores information on heterogeneity. Conceptually, we can apply the same idea to assign an appropriate correlation structure to Z-scores to allow heterogeneity. Nevertheless, further work is needed to study how to assign the correlation structure for developing more powerful tests based on p-values.

The proposed method, just like other meta-analysis methods, has its limitations. The use of information on heterogeneity can improve the power of meta-analysis, but at the same time makes meta-analysis vulnerable to biased results of one single study than conventional fixed effects and random effects methods. As such, particular caution is required regarding quality control of all studies involved in meta-analysis. As found in our simulations, the gain in power of the proposed methods depends on the similarity of true effects of SNPs in different studies, which is generally unknown. In practice, we have to define the correlation matrix of the random effects of SNPs based on known factors in study populations, such as age, sex, ethnicity and methods of phenotype measurement. In addition, the true effect models of different SNPs can be variable because of varied local linkage disequilibrium structures. As such, the designed matrix may sometime be poorly chosen, potentially leading to loss in power. Hence, it is important to have a method that is robust to misspecifications. In this paper, we suggest using a correlation design matrix balancing the contribution from testing the average effects and heterogeneity across studies, so that the method can have a robust performance, even when the correlation is completely misspecified, as the number of degrees of freedom due to heterogeneity is limited. Our simulations indicated this approach performed well under a wide range of true models from no to very high level of heterogeneity (Fig 3).

Although the method is proposed to address the issue of heterogeneity in GWAS meta-analysis, it offers a potential approach to perform tests of significance across multiple correlated and possibly heterogeneous effects in general. Particularly, the proposed test statistic may be used for testing the association of a group of pre-defined SNPs. Indeed, with an independent correlation matrix, this method is close to the gene-based test statistic based on the sum of the chi-squared 1 degree freedom statistics of individual variants[Liu, et al.

2010]. The proposed test provides an approach for evaluating p-values without extensive simulations. Moreover, it is flexible to incorporate prior functional information of variants, e.g. computational predictions of functional effect of amino acid changes, into a correlation structure for improving power. Extensions of the proposed test will be further studied. Another issue that warrants further studies is how to assign weights in the proposed VC method. In this paper, the effect estimate $\hat{\beta}_j$ is weighted by its inverse standard error. This weight has been shown by simulation had a good performance when the prior correlation is small. It is known that the weight of the inverse-variance is optimal for fixed effects model in the sense that the p-values approximate the p-value of the test based on raw data [Zaykin 2011]. When the prior correlation is assigned as 1, the VC model is essentially equivalent to fixed effects model and hence the inverse-variance should be more powerful [Zhou, et al. 2011]. However, this is not the case when a small prior correlation is assigned to studies. To further understand this concept, it may be useful to examine the case where the prior correlation is 0. In this case, VC essentially weights chi-square variables and the inverse-variance weight leads to a test statistic dominated by studies with larger sample sizes. In a future study, we will examine the optimal weights for the VC method under various prior correlation design matrices.

In summary, we propose a method to improve the power of meta-analysis by considering heterogeneity in effects between studies. The new test is simple and fast, and hence feasible for large scale genetic studies with millions of genetic markers.

Acknowledgments

We thank two reviewers for their valuable comments. T.W was support by the following NIH grants: R21HG006150 from the National Human Genome Research Institute (NHGRI) and P01HD070454 from the National Heart, Lung, and Blood Institute (NHLBI). Cardiovascular Health Study (CHS): This research was supported by grants and contracts HHSN268201200036C, HHSN268200800007C, N01 HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, N01HC35129, N01HC15103, N01HC55222, N01HC75150, N01HC45133, U01HL080295, R01HL087652 and HL080295 from the National Heart, Lung, and Blood Institute (NHLBI), with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided by AG023629 and AG031890 from the National Institute on Aging (NIA). DNA handling and genotyping was supported in part by National Center for Research Resources [M01-RR00425] to the Cedars-Sinai General Clinical Research Center Genotyping core and National Institute of Diabetes and Digestive and Kidney Diseases [DK063491] to the Southern California Diabetes Endocrinology Research Center. FHS cohort: supported by the National Heart, Lung and Blood Institute's Framingham Heart Study [N01-HC-25195] and its contract with Affymetrix, Inc for genotyping services [N02-HL-6-4278]. A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. KORA cohort: initiated and financed by the Helmholtz Center Munich, German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Part of this work was financed by the German National Genome Research Network [NGFN-2 and NGFNPlus: 01GS0823]. Our research was supported within the Munich Center of Health Sciences (MC Health) as part of LMUinnovativ. This study was in part supported by a grant from the German Federal Ministry of Education and Research (BMBF) to the German Center for Diabetes Research (DZD e.V.). SHIP cohort: SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg - West Pomerania. Genome- wide data have been supported by the Federal Ministry of Education and Research (03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg-West Pomerania. The University of Greifswald is a member of the Center of Knowledge Interchange' program of the Siemens AG. Pfizer provided partial grant support for the determination of serum samples and data analysis. Computing resources have been made available by the Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities (HLRB project h1231).

References

- Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007; 447(7145):661–78. [PubMed: 17554300]
- Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, Julier C, Morahan G, Nerup J, Nierras C. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet*. 2009; 41(6):703–7. others. [PubMed: 19430480]
- Begum F, Ghosh D, Tseng GC, Feingold E. Comprehensive literature review and statistical considerations for GWAS meta-analysis. *Nucleic Acids Res*. 2012; 40(9):3777–84. [PubMed: 22241776]
- Fanale D, Amodeo V, Corsini LR, Rizzo S, Bazan V, Russo A. Breast cancer genome-wide association studies: there is strength in numbers. *Oncogene*. 2012; 31(17):2121–8. [PubMed: 21996731]
- FISHER, RA. *Statistical Methods for Research Workers*. Oliver and Boyd; Edinburgh: 1932.
- Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, Lees CW, Balschun T, Lee J, Roberts R. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet*. 2010; 42(12):1118–25. others. [PubMed: 21102463]
- Frayling TM. Genome-wide association studies provide new insights into type 2 diabetes aetiology. *Nat Rev Genet*. 2007; 8(9):657–62. [PubMed: 17703236]
- Friedrich N, Haring R, Nauck M, Ludemann J, Rosskopf D, Spilcke-Liss E, Felix SB, Dorr M, Brabant G, Volzke H. Mortality and serum insulin-like growth factor (IGF)-I and IGF binding protein 3 concentrations. *J Clin Endocrinol Metab*. 2009; 94(5):1732–9. others. [PubMed: 19223521]
- Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am J Hum Genet*. 2011; 88(5):586–98. [PubMed: 21565292]
- Harrela M, Koistinen H, Kaprio J, Lehtovirta M, Tuomilehto J, Eriksson J, Toivanen L, Koskenvuo M, Leinonen P, Koistinen R. Genetic and environmental components of interindividual variation in circulating levels of IGF-I, IGF-II, IGFBP-1, and IGFBP-3. *J Clin Invest*. 1996; 98(11):2612–5. others. [PubMed: 8958225]
- Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*. 2005; 6(2):95–108. [PubMed: 15716906]
- Hong Y, Pedersen NL, Brismar K, Hall K, de Faire U. Quantitative genetic analyses of insulin-like growth factor I (IGF-I), IGF-binding protein-1, and insulin levels in middle-aged and elderly twins. *J Clin Endocrinol Metab*. 1996; 81(5):1791–7. [PubMed: 8626837]
- Hunter DJ. Gene-environment interactions in human diseases. *Nat Rev Genet*. 2005; 6(4):287–98. [PubMed: 15803198]
- Ioannidis JP, Patsopoulos NA, Evangelou E. Heterogeneity in meta-analyses of genome-wide association investigations. *PLoS One*. 2007; 2(9):e841. [PubMed: 17786212]
- Juul A, Scheike T, Davidsen M, Gyllenborg J, Jorgensen T. Low serum insulin-like growth factor I is associated with increased risk of ischemic heart disease: a population-based case-control study. *Circulation*. 2002; 106(8):939–44. [PubMed: 12186797]
- Kaplan RC, Petersen AK, Chen MH, Teumer A, Glazer NL, Doring A, Lam CS, Friedrich N, Newman A, Muller M. A genome-wide association study identifies novel loci associated with circulating IGF-I and IGFBP-3. *Hum Mol Genet*. 2011; 20(6):1241–51. others. [PubMed: 21216879]
- Kenyon CJ. The genetics of ageing. *Nature*. 2010; 464(7288):504–12. [PubMed: 20336132]
- Kuonen D. Saddlepoint approximation for distribution of quadratic forms in normal variables. *Biometrika*. 1999; 86(4):929–935.
- SSaL KY. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*. 1987; 82:605–610.
- Le Roith D. Seminars in medicine of the Beth Israel Deaconess Medical Center. Insulin-like growth factors. *N Engl J Med*. 1997; 336(9):633–40. [PubMed: 9032050]
- Lebre JJ, Stijnen T, van Houwelingen HC. Dealing with heterogeneity between cohorts in genomewide SNP association studies. *Stat Appl Genet Mol Biol*. 2010; 9 Article 8.

- Lindgren CM, Heid IM, Randall JC, Lamina C, Steinthorsdottir V, Qi L, Speliotes EK, Thorleifsson G, Willer CJ, Herrera BM. Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution. *PLoS Genet*. 2009; 5(6):e1000508. others. [PubMed: 19557161]
- Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, Hayward NK, Montgomery GW, Visscher PM, Martin NG. A versatile gene-based test for genome-wide association studies. *Am J Hum Genet*. 2010; 87(1):139–45. others. [PubMed: 20598278]
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A. Finding the missing heritability of complex diseases. *Nature*. 2009; 461(7265):747–53. others. [PubMed: 19812666]
- McCarthy MI, Hirschhorn JN. Genome-wide association studies: potential next steps on a genetic journey. *Hum Mol Genet*. 2008; 17(R2):R156–65. [PubMed: 18852205]
- Moonesinghe R, Khoury MJ, Liu T, Ioannidis JP. Required sample size and nonreplicability thresholds for heterogeneous genetic associations. *Proc Natl Acad Sci U S A*. 2008; 105(2):617–22. [PubMed: 18174335]
- Normand SL. Meta-analysis: formulating, evaluating, combining, and reporting. *Stat Med*. 1999; 18(3):321–59. [PubMed: 10070677]
- Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet*. 2001; 69(1):124–37. [PubMed: 11404818]
- Renahan AG, Zwahlen M, Minder C, O'Dwyer ST, Shalet SM, Egger M. Insulin-like growth factor (IGF)-I, IGF binding protein-3, and cancer risk: systematic review and meta-regression analysis. *Lancet*. 2004; 363(9418):1346–53. [PubMed: 15110491]
- Samani AA, Yakar S, LeRoith D, Brodt P. The role of the IGF system in cancer growth and metastasis: overview and recent insights. *Endocr Rev*. 2007; 28(1):20–47. [PubMed: 16931767]
- Saxena R, Voight BF, Lyssenko V, Burt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*. 2007; 316(5829):1331–6. others. [PubMed: 17463246]
- Scherag A, Dina C, Hinney A, Vatin V, Scherag S, Vogel CI, Muller TD, Grallert H, Wichmann HE, Balkau B. Two new Loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and German study groups. *PLoS Genet*. 2010; 6(4):e1000916. others. [PubMed: 20421936]
- Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*. 2007; 316(5829):1341–5. others. [PubMed: 17463248]
- Scott LJ, Muglia P, Kong XQ, Guan W, Flickinger M, Upmanyu R, Tozzi F, Li JZ, Burmeister M, Absher D. Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. *Proc Natl Acad Sci U S A*. 2009; 106(18):7501–6. others. [PubMed: 19416921]
- Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*. 2007; 445(7130):881–5. others. [PubMed: 17293876]
- Souren NY, Paulussen AD, Loos RJ, Gielen M, Beunen G, Fagard R, Derom C, Vlietinck R, Zeegers MP. Anthropometry, carbohydrate and lipid metabolism in the East Flanders Prospective Twin Survey: heritabilities. *Diabetologia*. 2007; 50(10):2107–16. [PubMed: 17694296]
- Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters GB, Styrkarsdottir U, Gretarsdottir S, Emilsson V, Ghosh S. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nature Genetics*. 2007; 39(6):770–5. others. [PubMed: 17460697]
- STOUFFER, SA.; SUCHMAN, EA.; DEVINNEY, LC.; STAR, SA.; WILLIAMS, RM, JR. Adjustment during Army life 1. Princeton Univ. Press; Princeton, NJ: 1949. The American soldier..
- van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med*. 2002; 21(4):589–624. [PubMed: 11836738]

- Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, Heid IM, Berndt SI, Elliott AL, Jackson AU, Lamina C. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet.* 2009; 41(1):25–34. others. [PubMed: 19079261]
- Zaykin DV. Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. *J Evol Biol.* 2011; 24(8):1836–41. [PubMed: 21605215]
- Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science.* 2007a; 316(5829):1336–41. others. [PubMed: 17463249]
- Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science.* 2007b; 316(5829):1336–41. others. [PubMed: 17463249]
- Zhou B, Shi J, Whittemore AS. Optimal methods for meta-analysis of genome-wide association studies. *Genet Epidemiol.* 2011; 35(7):581–91. [PubMed: 21922536]

Figure 1.

A simplified illustration of reject regions (blue area) for different meta-analysis methods for a pair studies with a similar design. The upper row, from left to right, are FE and RE; the lower row, from left to right are VC with assumption of independent effects between two studies. Abbreviations: FE, fixed effects; RE random effects and VC, variance components

Figure 2.

The empirical type I error rate of the proposed VC method at the nominal significance levels of $\alpha=0.05$ and $\alpha=0.01$ under various prior correlations. Abbreviations: FE, fixed effects; RE random effects and VC, variance components.

Figure 3.

The empirical power of different methods at the statistical significance level of 10^{-6} when the effect sizes of studies were sampled from a common normal distribution. (a) Studies had an equal sample size. The average OR was fixed at 1.3. (b) Studies had unequal sample sizes. The average OR was fixed at 1.3. Abbreviations: OR, odds ratio; FE, fixed effects; RE random effects, LRT likelihood ratio test, and VC, variance components

Figure 4.

The empirical power of different methods at the statistical significance level of 10^{-6} for three settings. The three settings correspond three settings correspond to (1) no heterogeneity, (2) heterogeneous with the same direction, and (3) heterogeneous with the opposite directions. The OR of the underlying causal variants was set at 1.4 and five studies had equal sample sizes. Abbreviations: OR, odds ratio; FE, fixed effects; RE, random effects; LRT, likelihood ratio test; and VC, variance components.

Figure 5.
The quantile-quantile plot of the GWAS results from the VC method.

Figure 6.
Manhattan plot of the GWAS results from the VC method

Table 1

Empirical Type I Error Rate of FE, RE, and the New VC at $\alpha=0.05$ and 0.01.

# studies	Sample size	$\alpha = 0.05$			$\alpha = 0.01$		
		FE	RE	VC	FE	RE	VC
3	Equal	0.0500	0.0373	0.0517	0.0102	0.0072	0.0101
	Unequal	0.0505	0.0377	0.0526	0.0097	0.0068	0.0101
5	Equal	0.0505	0.0385	0.0518	0.0101	0.0073	0.0098
	Unequal	0.0487	0.0357	0.0509	0.0091	0.0067	0.0091
10	Equal	0.0491	0.0379	0.0506	0.0095	0.0070	0.0094
	Unequal	0.0508	0.0390	0.0528	0.0102	0.0072	0.0098
20	Equal	0.0491	0.0408	0.0490	0.0095	0.0071	0.0096
	Unequal	0.0496	0.0402	0.0503	0.0100	0.0075	0.0094

Abbreviations: FE, fixed effects; RE random effects and VC, variance components

Table 2

Empirical Type I Error rate of the VC method at Increasing Significance Levels.

Sample size	Significance Level			
	1e-3	1e-4	1e-5	1e-6
Equal	0.90×10^{-3}	0.94×10^{-4}	0.83×10^{-5}	1.1×10^{-6}
Unequal	0.91×10^{-3}	0.90×10^{-4}	0.96×10^{-5}	0.90×10^{-6}

Abbreviations: VC, variance components

Table 3

Hypothetical correlation coefficients between the genotyped SNP marker and an untyped causal variant in two populations.

Setting	Population 1	Population 2	Note
1	0.8	0.8	No heterogeneity
2	0.2	1	Heterogeneity with same direction
3	-0.6	1	Heterogeneity with opposite direction

Abbreviation: SNP, single nucleotide polymorphism

Table 4

Loci associated with IGF-I concentrations at the genome-wide significant ($P < 5 \times 10^{-8}$) and $P < 1 \times 10^{-6}$ level.

Chromosome	RS number	FE	VC	Nearby Genes
6	rs2153960	6.32e-7	1.86e-7	FOXO3
6	rs1935949	6.71e-7	1.95e-7	FOXO3
6	rs3800229	6.87e-7	1.97e-7	FOXO3
6	rs1935952	7.33e-7	2.14e-7	FOXO3
6	rs9398172	9.65e-7	2.99e-7	FOXO3
6	rs9398171	1.18e-6	3.77e-7	FOXO3
6	rs4946936	1.53e-6	4.83e-7	FOXO3
6	rs4946935	1.50e-6	6.13e-7	FOXO3
6	rs4946932	2.91e-6	6.31e-7	FOXO3
6	rs1268165	1.01e-6	8.01e-7	FOXO3
6	rs1268162	1.03e-6	9.11e-7	FOXO3
6	rs1268164	1.18e-6	9.53e-7	FOXO3
6	rs2207731	1.11e-6	9.90e-7	FOXO3
7	rs700753	2.18e-9	8.92e-10	IGFBP3/TNS3
7	rs700752	3.19e-8	1.52e-9	IGFBP3/TNS3
7	rs700750	3.61e-7	1.63e-8	IGFBP3/TNS3
7	rs856563	1.43e-7	5.19e-8	IGFBP3/TNS3
7	rs856560	2.71e-7	1.28e-7	IGFBP3/TNS3
7	rs856540	2.96e-7	1.98e-7	IGFBP3/TNS3
7	rs7780564	6.03e-7	5.11e-7	IGFBP3/TNS3
7	rs12533001	8.01e-7	6.71e-7	IGFBP3/TNS3
7	rs957755	8.84e-7	8.59e-7	IGFBP3/TNS3
7	rs11769597	1.62e-7	9.58e-7	RPA3
10	rs1245541	5.38e-7	1.73e-7	SPOCK2
10	rs1245543	5.44e-7	1.75e-7	SPOCK2
10	rs2394851	5.68e-7	1.87e-7	SPOCK2
10	rs1245542	7.26e-6	2.04e-7	SPOCK2
10	rs1668156	1.15e-6	6.63e-7	SPOCK2
10	rs1245578	1.06e-6	8.32e-7	SPOCK2

Abbreviations: FE, fixed effects; RE random effects and VC, variance components, IGF-I, insulin-growth like factor I