A statistical framework to detect gene-environment

interactions influencing complex traits

A STATISTICAL FRAMEWORK TO DETECT

GENE-ENVIRONMENT INTERACTIONS INFLUENCING

COMPLEX TRAITS

BY

WEI Q. DENG, B.Sc.

A THESIS

SUBMITTED TO THE DEPARTMENT OF CLINICAL EPIDEMIOLOGY & BIOSTATISTICS

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

MCMASTER UNIVERSITY

Master of Science (2013)   McMaster University

(Clinical Epidemiology & Biostatistics)   Hamilton, Ontario, Canada

TITLE: A statistical framework to detect gene-environment in-
                        teractions influencing complex traits

AUTHOR: Wei Q. Deng

                        B.Sc,

                        McMaster University, Hamilton, Canada

SUPERVISORS: Dr. Guillaume Paré

                        Dr. Angelo J. Canty

NUMBER OF PAGES: xiv, 113

Abstract

Advancements in human genomic technology have helped to improve our understanding of how genetic variation plays a central role in the mechanism of disease susceptibility. However, the very high dimensional nature of the data generated from large-scale genetic association studies has limited our ability to thoroughly examine genetic interactions.

A prioritization scheme – Variance Prioritization (VP) – has been developed to select genetic variants based on differences in the quantitative trait variance between the possible genotypes using Levene's test (Paré *et al.*, 2010). Genetic variants with Levene's test p-values lower than a pre-determined level of significance are selected to test for interactions using linear regression models. Under a variety of scenarios, VP has increased power to detect interactions over an exhaustive search as a result of reduced search space. Nevertheless, the use of Levene's test does not take into account that the variance will either monotonically increase or decrease with the number of minor alleles when interactions are present.

To address this issue, I propose a maximum likelihood approach to test for trends in variance between the genotypes, and derive a closed-form representation of the likelihood ratio test (LRT) statistic. Using simulations, I examine the performance of LRT in assessing the inequality of quantitative traits variance stratified by genotypes,

and subsequently in identifying potentially interacting genetic variants.

LRT is also used in an empirical dataset of 2,161 individuals to prioritize genetic variants for gene-environment interactions. The interaction p-values of the prioritized genetic variants are consistently lower than expected by chance compared to the non-prioritized, suggesting improved statistical power to detect interactions in the set of prioritized genetic variants. This new statistical test is expected to complement the existing VP framework and accelerate the process of genetic interaction discovery in future genome-wide studies and meta-analyses.

Acknowledgements

I would like to thank my supervisors Dr Guillaume Paré and Dr Angelo J. Canty for their supervision and continual academic support throughout my masters (and undergraduate) research. I am grateful to Dr Paré for introducing me to the field of genetic epidemiology, and constantly challenging me with interesting problems in statistical genetics. I am also privileged to have Dr Canty with a background in computational statistics in addition to his research interest in genetic epidemiology. Their valuable comments, constructive criticisms, and insightful suggestions are essential to the improvement of this thesis.

I would also like to express my gratitude to my thesis committee member Dr David Meyre, for the many casual conversations we had swapping research ideas and discussing emerging trends in genetics literature.

I also want to thank Dr Peter D. M. MacDonald, who agreed to be my external examiner, for his interest and commitment to this thesis.

I am indebted to my mother, Caren, who has always supported and encouraged me in my dark times.

Finally, I would like to extend my appreciation to Claude Debussy and Joseph-Maurice Ravel for their beautiful piano compositions that accompanied me during the completion of this thesis.

List of abbreviations

ANOVA: analysis of variance

BMI: body mass index

FG: Fasting glucose

GEWIST: gene-environment wide interaction search threshold

GWAS: genome-wide association study

HDL-C: high density lipoprotein cholesterol

HWE: Hardy-Weinberg equilibrium

LD: linkage disequilibrium

LDL-C: low density lipoprotein cholesterol

LRT: likelihood ratio test

MAF: minor allele frequency

MLE: maximum likelihood estimator

SNP: single nucleotide polymorphism

TG: triglyceride

VE: variance explained

VP: variance prioritization

WC: waist circumference

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The desire to understand the human genome and how genetic variations affect human health and physiology has largely driven the search for genetic determinants of complex traits. In less than a decade, the advent of rapid and inexpensive genotyping technology has led to waves of large-scale genome-wide association studies (GWASs), in which millions of genetic variants are genotyped and analyzed at once. GWAS interrogates the most common form of genetic variation – single nucleotide polymorphisms (SNPs) – to test for association with a disease status or a quantitative trait. The agnostic nature of GWAS could potentially lead to false positives if the large numbers of simultaneous hypotheses are not properly accounted. As a result, stringent statistical criteria are proposed: a p-value significance threshold of 5E-08 has been widely accepted to be the gold standard to claim a significant association (Pe'er *et al.*, 2008; Rice *et al.*, 2008).

Currently, the GWAS catalogue (`www.genome.gov/gwastudies`) maintained by the NIH characterizes over 900 SNPs that have been shown to be associated with

human disease phenotypes and quantitative traits from the previously published reports (Hindorff *et al.*, 2009). GWAS has been extremely successful in terms of linking single locus to human traits. However, these variants combined explain relatively little of the heritability for most complex traits. The complexity of human traits has inevitably stimulated interest in the search for gene-environment interactions as most complex disease and traits are likely to be a result of the interplay between genes and environmental factors. Identification of genetic interactions may help to address the so-called missing heritability problem. Heritability is an important concept in genetics since it establishes the relative contribution of genes compared to environmental factors in influencing human trait (Visscher *et al.*, 2008). The quantitative trait variance explained by genetic variants under an additive model (i.e. variance explained by a linear combination of SNPs) is defined as the narrow sense heritability, while the broad sense heritability encompasses both the additive and non-additive (i.e. gene-gene, gene-environment and higher order interactions) genetic components. Under the additive genetic assumption, the missing heritability refers to the inconsistency in the total heritability derived indirectly from population data and the total heritability according to the variance explained by all the known genetic variants detected in GWASs. It is suggested that the unexplained proportion of heritability could be due to an overestimate of the total heritability from population data without properly accounting for non-additive contributions from gene-gene and gene-environment interactions (Zuk *et al.*, 2012). That is, even if all the additive genetic variance were accounted for, the total heritability would still be less than that estimated from the population thus leading to an inflated proportion of unexplained heritability.

The secondary use of existing GWAS data offers a convenient platform to identify

genes that influence phenotypic traits under the modification of environmental factors (Bookman *et al.*, 2011). However, the very high dimensional nature of the data generated from large-scale GWASs has limited our ability to thoroughly examine genetic interactions. Furthermore, robust replications at the genome-wide significance threshold are required to confirm interactions reported in the literature. On the other hand, there is little consensus on how to handle borderline interactions as supposed to dealing with main effect associations (Panagiotou and Ioannidis, 2012). Further, within the context of genetic interactions, we face another challenge because the genetic architecture of interactions is usually not known *a priori* and choices of the interaction model (i.e. additive and multiplicative), are thus subjective to the investigators.

Despite these conundrums, there remains a pressing need to understand how the environment in synergy with individual genetic profiles influences human conditions. In the long run, it is hoped that gene-environment interactions can help improve targeted intervention and predict individual drug responses. There has been an ongoing effort to develop statistical methods to identify genetic interactions. Many sophisticated approaches such as Bayesian model selection, data mining, and machine learning methods have been proposed (for a thorough review refer to Cordell (2009)). These methods are typically designed to examine interactions in specific contexts, such as non-linear models of interaction effects or higher order interactions, and thus have limited utilities in genome-wide settings compared to more general methods such as generalized linear models. It should be noted that for the purpose of this thesis, statistical gene-environment interactions refer to the departure from a model that is linear in main effects (Fisher, 1958).

Performing an exhaustive search using linear regression models for interactions

on a whole genome has been shown to be computationally feasible (Cordell, 2009; Marchini *et al.*, 2005): the search time for pair-wise interactions increases linearly in M for gene-environment and quadratically in M for gene-gene interactions, where M is the number of SNPs. Rather, the rate-limiting factors in novel discovery of gene-environment and gene-gene interactions are likely to be a result of the small interaction effect sizes and the huge number of genetic variants to be corrected for in multiple hypothesis testing (Lindstrom *et al.*, 2009; Rice *et al.*, 2008; Thomas, 2010). To deal with these underlying issues, a large sample size is required. This may be obtained through meta-analyses of multiple studies or individually large studies with innovative designs (Bookman *et al.*, 2011). These statistical challenges make a strong case for novel prioritization methods to filter out genetic variants that are unlikely to be involved in genetic interactions. One emerging approach is to combine pathway information and prior GWAS results to define a group of SNPs at the gene or exon level for gene-gene and gene-environment interactions (Bush *et al.*, 2009). This approach has demonstrated utility in association studies (Baranzini *et al.*, 2009; Ritchie, 2009). However, prioritization of SNPs using this method is limited by the existing body of biological knowledge, overlooking the possibility of novel interactions between SNPs that are yet to be characterized. Another strategy to reduce multiple hypotheses leads to a two-step analysis, where SNPs are first selected based on a minimal main effect on the trait of interest, and subsequently tested for interaction effects (Evans *et al.*, 2006; Kooperberg and Leblanc, 2008; Marchini *et al.*, 2005; Millstein *et al.*, 2006). These methods have improved statistical power and computational efficiencies over an exhaustive search. Nevertheless, they make the assumption that interacting SNPs will necessarily show a main effect on the trait. Therefore, interacting SNPs with no

or weak main effects will be incorrectly classified as non-interacting, which inevitably results in a loss of overall statistical power (Culverhouse *et al.*, 2002; Murcray *et al.*, 2009; Thomas, 2010).

Paré *et al.* (2010) recently proposed a novel method – variance prioritization (VP) – to prioritize SNPs by leveraging the effect of genetic interactions manifested on the variance of a quantitative trait; an active area of research (Struchalin *et al.*, 2010, 2012). Prioritization of SNPs is achieved by comparing the quantitative trait variance conditional on the three possible genotypes of biallelic SNPs using Levene's test (Levene, 1960) for variance heterogeneity (Figure 1). That is, SNPs are first selected based on the quantitative trait variance inequality p-values at a pre-determined threshold (typically a nominal significance level of 0.05). Second, the subset of SNPs selected for that particular trait (the response variable) is tested for interaction effects against either categorical or continuous environmental covariates or other SNPs using linear regression models.

This two-step procedure exploits the fact that tests for variance heterogeneity and interaction effect are uncorrelated under the null hypothesis of no interaction, and thus can be used to select SNPs for interaction testing according to their variance heterogeneity p-values (Paré *et al.*, 2010). The novelty of the VP framework lies in the fact that it comprehensively searches through all SNPs in the first stage and bases the prioritization on the additional information acquired from the quantitative trait variability according to the theoretical model of statistical interactions under plausible scenarios. Furthermore, the use of VP is not restricted to the variance heterogeneity test. Depending on the underlying nature of the interaction effect, variance heterogeneity tests with a general alternative such as Levene's test are more
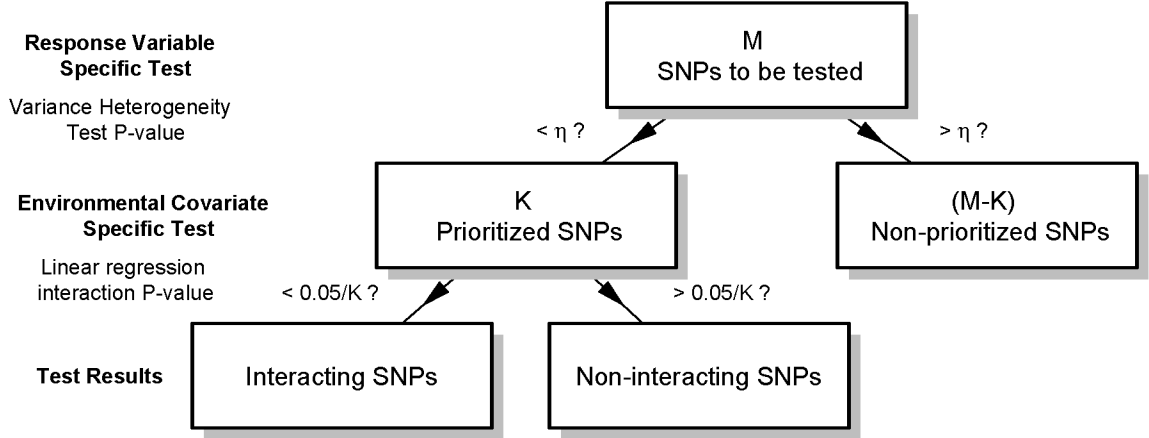
Figure 1: Variance Prioritization Procedure

suitable for detection of non-linear (qualitative) interactions. On the other hand, statistical tests with a specific alternative for ordered variances would perform better if the interaction effects mediated in an additive fashion.

Instead of using a pre-determined threshold for all SNPs, we introduced a fast algorithm – Gene Environment Wide Interaction Search Threshold (GEWIST) – to efficiently and accurately determine the optimal significance level (η_0) for the variance heterogeneity test on a per SNP basis (Deng and Paré, 2011, 2012). GEWIST enables implementation of optimal VP in more general genome-wide settings. The original steps of VP are then equivalent to conducting two independent statistical tests, such that the optimal choice of η for the variance heterogeneity test is conditional on multiple factors (i.e. minor allele frequency (MAF), sample size, variance explained by covariate, and interaction effect size) and the level of significance for interaction testing using linear regression is determined by the proportion of SNPs deemed significant

6

in the variance heterogeneity test. Facilitated by GEWIST, we were able to show that VP has increased statistical power to detect genetic interactions in a variety of scenarios, most strikingly when the interaction effect sizes are small. Though a powerful approach, effectiveness of VP depends upon the statistical procedures employed to quantify the heterogeneity of quantitative trait variance between genotype groups. The performance of VP could potentially be compromised if the variance inequality p-values computed for individual SNPs are overly conservative.

In this thesis, I propose a maximum likelihood approach to prioritize SNPs by taking advantage of the monotonic trends in quantitative trait variance stratified according to the genotype when interactions are present. Apart from leveraging for gene-gene or gene-environment interactions, examining genetic effects on phenotypic variance, or even higher order moments (Aschard *et al.*, 2013) may help add to our understanding of the phenotypic complexity. There has been on-going attention and investigation into the effects of genetic variants on phenotypic variability since the first publication of genetic variant affecting variance of C-Reactive Protein at genome-wide significance (Paré *et al.*, 2010). Yet it was not until recently that a consortium effort to meta-analyze heterogeneity of phenotypic variance discovered an *FTO* genetic variant associated with the variability of BMI at genome-wide significance (Yang *et al.*, 2012).

The resulting likelihood ratio test (LRT) requires only the set of quantitative trait variances per genotype and genotype counts to determine the variance inequality p-value for a given SNP. A closed-form representation of the LRT statistic is derived to expedite the computations for genome-wide studies. Using simulation studies, I will compare the performance of VP using LRT and alternative variance heterogeneity tests to a conventional exhaustive search in a variety of gene-environment scenarios.

In addition, sensitivity analysis will be conducted to establish the empirical type I error rates associated with LRT and alternative variance heterogeneity tests when the normality assumption is violated. To validate the VP method empirically, I will apply LRT to a genome-wide dataset in order to prioritize SNPs for interactions with a variety of adiposity measures on selected metabolic traits. These analyses will address two methodological issues in the search for novel genetic interactions using VP: first, is there an efficient and reliable way to rank all potentially interacting covariates based on the optimal gain in interaction power through VP? And secondly, how effective is VP in terms of selecting potentially interacting SNPs under the alternative hypothesis of variance heterogeneity?

Chapter 2

Statistical Methods to Prioritize SNPs for Genetic Interactions

The present statistical framework to test for interactions uses a linear regression model, where the interaction beta coefficient is either positive or negative under the alternative hypothesis of interaction. In this section, I present the quantitative trait variance conditional on genotype as a function of the interaction and interacting co-variate beta coefficients. Under plausible assumptions, the conditional variance would increase or decrease monotonically with the number of minor alleles if the interaction beta coefficient is not zero. This theoretical observation leads to a statistical test based on likelihood principles. The proposed likelihood ratio test (LRT) assesses the specific alternative of increasing or decreasing variance against the null hypothesis of variance homogeneity. I demonstrate that a closed form representation of the LRT statistics is available when the variances increase or decrease by a factor of r. To conclude this chapter, I describe four additional variance heterogeneity tests that will be examined in the simulation studies (Chapter 3).

2.1   Preliminaries

Among all forms of genetic variations, single nucleotide polymorphisms (SNPs) are considered the most common type. A SNP denotes a single base pair change in the DNA sequence, which contains four types of bases: adenine (A), cytosine (C), thymine (T), or guanine (G). For most of the common SNPs, the possible number of unique single nucleotides at a genetic locus is two. For example, in the sequence TG<u>A</u>A to TG<u>G</u>A, there is a SNP at the third position with two possible alleles, A or G. The allele most frequently observed in the human population is termed the major allele and the other one termed the minor allele. The paired chromosomes give rise to combinations of these two alleles and form genotypes. The genotype of a SNP can be considered as a categorical variable with three levels, i.e. AA, AG, or GG. The genotype consisting of two major alleles, two minor alleles, and one copy of each allele is classified as the major allele homozygote, the minor allele homozygote, and the heterozygote, respectively. The minor allele frequency (MAF) can be estimated from the observed genotype counts. For each biallelic genetic variant with MAF p, the expected number of individuals in each genotype group is determined according to the Hardy-Weinberg principle:

$$N_0 = (1-p)^2 N \tag{2.1}$$

$$N_1 = 2(1-p)(p)N \tag{2.2}$$

$$N_2 = (p)^2 N \tag{2.3}$$

where N is the overall sample size. Let g denote the genotype of a biallelic SNP with MAF p in Hardy-Weinberg Equilibrium (HWE), where g can take on the values

0, 1, or 2, denoting the number of minor alleles. Under HWE, the genotype G can be considered as a binomial random variable with probability p and size 2, written $G \sim \mathrm{B}(2, p)$. The statistical interaction between the genotype variable G and an environmental covariate C can be tested by the linear regression model:

$$Y = \beta_0 + \beta_1 G + \beta_2 C + \beta_3 GC + \varepsilon \tag{2.4}$$

where Y is the quantitative trait. In the example to follow (Chapter 4), Y was the triglyceride level, C was the waist circumference, and the interaction between 656,004 predictor variables (G) and waist circumference was investigated one at a time. Since the distribution of the interacting covariate C above is not specified, Equation 2.4 can be generalized to different types of genetic interactions with categorical environmental exposures, continuous environmental covariate or the three genotype classes of a second biallelic SNP. For simplicity, assume both the covariate C and the genotype variable G are standardized by subtracting the mean and dividing by the standard deviation, so they have mean 0 and variance 1. However, as the population values are usually unknown, the sample mean and variance are used in practice. It is also assumed that the quantitative trait Y conditional on the genotype G and covariate C has a standard normal distribution with mean 0 and variance 1, that is, $\varepsilon \sim \mathrm{N}(0,1)$. In addition, assume the error term ε is independent of the genotype G and covariate C.

Under the assumption of gene-environment independence (Lindstrom et al., 2009),

the total variance of Y is (Paré $et\ al.$, 2010):

$$
\begin{aligned}
\mathrm{Var}(Y) &= \mathrm{Var}(\beta_0 + \beta_1 G + \beta_2 C + \beta_3 GC + \varepsilon) \\
&= \mathrm{Var}(\beta_1 G) + \mathrm{Var}(\beta_2 C) + \mathrm{Var}(\beta_3 GC) + \mathrm{Var}(\varepsilon) \\
&\quad + 2\mathrm{Cov}(\beta_1 G, \beta_2 C) + 2\mathrm{Cov}(\beta_2 C, \beta_3 GC) + 2\mathrm{Cov}(\beta_1 G, \beta_3 GC) \\
&= \mathrm{Var}(\beta_1 G) + \mathrm{Var}(\beta_2 C) + \mathrm{Var}(\beta_3 GC) + \mathrm{Var}(\varepsilon) \\
&= 2p(1-p)\beta_1^2 + \beta_2^2 \mathrm{Var}(C) + \beta_3^2 \mathrm{Var}(G)\mathrm{Var}(C) + \mathrm{Var}(\varepsilon) \\
&= 2p(1-p)\beta_1^2 + \beta_2^2 + 2p(1-p)\beta_3^2 + 1
\end{aligned}
\tag{2.5}
$$

Effect size θ_{GE} of a gene-environment interaction is defined by the proportion of quantitative trait variance explained by the interaction (Paré $et\ al.$, 2010):

$$
\theta_{GE} = \frac{\mathrm{Var}(\beta_3 GC)}{\mathrm{Var}(Y)} = \frac{2p(1-p)\beta_3^2}{2p(1-p)\beta_1^2 + \beta_2^2 + 2p(1-p)\beta_3^2 + 1}
\tag{2.6}
$$

Similarly, the proportion of variance explained by an environmental covariate C is (Paré $et\ al.$, 2010):

$$
\theta_E = \frac{\mathrm{Var}(\beta_2 C)}{\mathrm{Var}(Y)} = \frac{\beta_2^2}{2p(1-p)\beta_1^2 + \beta_2^2 + 2p(1-p)\beta_3^2 + 1}
\tag{2.7}
$$

2.2 A Mathematical Representation of Variance Conditional on the Genotype

2.2.1 Quantitative Trait Variance per Genotype

The conditional variance of the quantitative trait Y given a genotype g is:

$$\sigma_g^2 = \text{Var}(Y|G=g) = \text{Var}(\beta_0 + \beta_1 G + \beta_2 C + \beta_3 GC + \varepsilon | G = g)$$
$$= \text{Var}(\beta_2 C + \beta_3 GC + \varepsilon | G = g)$$
$$= (\beta_2 + \beta_3 g)^2 \text{Var}(C|G=g) + \text{Var}(\varepsilon|G=g)$$
$$= (\beta_2 + \beta_3 g)^2 + \text{Var}(\varepsilon|G=g) \tag{2.8}$$

Assume the error term has constant variances across genotypes:

$$\text{Var}(\varepsilon|G=0) = \text{Var}(\varepsilon|G=1) = \text{Var}(\varepsilon|G=2) = 1 \tag{2.9}$$

The variance can be stratified according to the three possible genotypes:

$$\sigma_0^2 = \beta_2^2 + 1 \tag{2.10}$$

$$\sigma_1^2 = (\beta_2 + \beta_3)^2 + 1 \tag{2.11}$$

$$\sigma_2^2 = (\beta_2 + 2\beta_3)^2 + 1 \tag{2.12}$$

The conditional variances increase or decrease with the number of minor alleles when the interaction effect is present ($\beta_3 \neq 0$) and the following conditions hold true:

1. Gene-environment independence.

2. The error term satisfies the constant variance assumption when stratified by the genotypes.

3. The absolute value of the covariate term beta-coefficient must be at least 1.5 times greater than that of the interaction term coefficient when the two beta coefficients are opposite in sign, i.e. $|\beta_2| > \left|\frac{3\beta_3}{2}\right|$ & $\beta_2\beta_3 < 0$ (see Appendix C.1 for details).

Under these assumptions, the relationship between the two beta-coefficients and the trend in conditional variances given the genotypes is summarized in Table 1. Note that the direction of the interaction is usually not known *a priori*.

2.2.2 Characterization of Ordered Variances Using a Ratio Parameter

Let $\sigma_0{}^2$, $\sigma_1{}^2$, and $\sigma_2{}^2$ denote the unknown population variance parameters conditional on the three possible genotypes. I can specify either a multiplicative model

	$\beta_2 > 0$	$\beta_2 < 0$
$\beta_3 > 0$	$\sigma_0^2 < \sigma_1^2 < \sigma_2^2$	$\sigma_0^2 > \sigma_1^2 > \sigma_2^2$
$\beta_3 < 0$	$\sigma_0^2 > \sigma_1^2 > \sigma_2^2$	$\sigma_0^2 < \sigma_1^2 < \sigma_2^2$

Table 1: Trends in Conditional Variances According to the Directions of Interaction and Interacting Covariate Main Effects

where the relative increase or decrease in variance is approximated by a ratio of size r, or an additive model where the absolute increase or decrease in variance is approximated by a difference of size d. Mathematically, my motives for adopting a ratio parameter are 1) the ratio parameter provides a better approximation to the increase or decrease in variance between genotypes than the difference parameter, and 2) close-form representation of the maximum likelihood estimator (MLE) of the ratio parameter is available under the additional assumption of normality of the trait given the genotypes.

Let the ratio of conditional variances given the heterozygote and the major homozygote groups be r_1, and the ratio of conditional variances given the minor homozygote and the heterozygote groups be r_2.

$$r_1 = \frac{\sigma_1^2}{\sigma_0^2} = \frac{(\beta_2 + \beta_3)^2 + 1}{\beta_2^2 + 1} = 1 + \frac{2\beta_2\beta_3 + \beta_3^2}{\beta_2^2 + 1} \tag{2.13}$$

$$r_2 = \frac{\sigma_2^2}{\sigma_1^2} = \frac{(\beta_2 + 2\beta_3)^2 + 1}{(\beta_2 + \beta_3)^2 + 1} = 1 + \frac{2\beta_2\beta_3 + 3\beta_3^2}{(\beta_2 + \beta_3)^2 + 1} \tag{2.14}$$

Alternatively, let the difference in conditional variances given the heterozygote and the major homozygote groups be d_1, and the difference in conditional variances given the minor homozygote and the heterozygote groups be d_2.

$$d_1 = \sigma_1^2 - \sigma_0^2 = 2\beta_2\beta_3 + \beta_3^2 \tag{2.15}$$

$$d_2 = \sigma_2^2 - \sigma_1^2 = 2\beta_2\beta_3 + 3\beta_3^2 \tag{2.16}$$

The difference and the ratio are both functions of the two beta-coefficients obtained from the linear regression model. I then show that the absolute difference

between the two ratio parameters is smaller than the two difference parameters, and thus a one parameter model with $r = r_1 = r_2$ is a better approximation than a one-parameter model with $d = d_1 = d_2$ (Appendix C.2). Though, it has also been shown elsewhere that results from a multiplicative and an additive model of the SNP effects on phenotypic variance converge when the effect is small (Visscher and Posthuma, 2010).

2.3   A Likelihood Ratio Test for Variance Hetero-geneity

Consider the null hypothesis that the conditional variances are all equal and the ratio is 1. The alternative hypothesis states that the conditional variances increase or decrease by a factor of r, or in mathematical terms:

$$H_o : r = 1 \quad \text{versus} \quad H_1 : r \neq 1$$

The null model aims to estimate a common unknown variance parameter assuming all variances share the same distribution, while the alternative model suggests a class of distributions with three unknown variance parameters where each is a function of the ratio and the variance of the heterozygote group. Recall that the likelihood ratio statistic (Hogg *et al.*, 2005)

$$\Lambda = 2(l_1 - l_o) \tag{2.17}$$

asymptotically follows a central chi-squared distribution with 1 degree of freedom ($\sim \chi^2(1)$) under the null hypothesis of $r = 1$, where l_1 and l_0 denote the log-likelihoods under the null and alternative model, respectively. In the following section, I will calculate the log-likelihoods under both the null and alternative hypotheses, and derive closed-form solutions for the maximum likelihood estimators (MLEs).

2.3.1 Derivation of the Likelihood Ratio Test Statistic

Assume the quantitative trait Y given a genotype class g follows a normal distribution with unknown mean and population variance σ_g^2. Let s_0^2, s_1^2, and s_2^2 denote the observed variances conditional on the three possible genotypes. The observed variance (or the sample variance) s_g^2 has a scaled chi-squared distribution with the population variance σ_g^2 (Hogg $et\ al.$, 2005):

$$\frac{(N_g - 1)s_g^2}{\sigma_g^2} \sim \chi^2(N_g - 1) \tag{2.18}$$

where N_g denotes the sample size in the genotype group g.

Probability Density Function of the Quantitative Trait Variance Conditional on the Genotype

Given the heterozygote group variance σ_g^2 and ratio parameter r, the probability distribution function of the major homozygote group variance is

$$f(s_0^2; \sigma_1^2, r) = \frac{1}{2^{\frac{N_0-1}{2}}\Gamma\left(\frac{N_0-1}{2}\right)} \left(\frac{(N_0-1)s_0^2}{2\sigma_1^2/r}\right)^{\frac{N_0-3}{2}} \exp\left(\frac{-(N_0-1)s_0^2}{2\sigma_1^2/r}\right), \tag{2.19}$$

the density function for the heterozygote group variance is

$$f(s_1^2; \sigma_1^2) = \frac{1}{2^{\frac{N_1-1}{2}}\Gamma\left(\frac{N_1-1}{2}\right)} \left(\frac{(N_1-1)s_1^2}{2\sigma_1^2}\right)^{\frac{N_1-3}{2}} \exp\left(\frac{-(N_1-1)s_1^2}{2\sigma_1^2}\right), \qquad (2.20)$$

and the density function for the minor homozygote group variance is

$$f(s_2^2; \sigma_1^2, r) = \frac{1}{2^{\frac{N_2-1}{2}}\Gamma\left(\frac{N_2-1}{2}\right)} \left(\frac{(N_2-1)s_2^2}{2\sigma_1^2 r}\right)^{\frac{N_2-3}{2}} \exp\left(\frac{-(N_2-1)s_2^2}{2\sigma_1^2 r}\right). \qquad (2.21)$$

Log-Likelihood Function under the Alternative of Ordered Variance

Given the observed set of variances $(s_0^2,\ s_1^2,\ s_2^2)$ under the alternative hypothesis, the ratio r and variance parameter σ_1^2 are estimated by maximizing the log-likelihood function:

$$\begin{aligned}
l_1(\sigma_1^2, r | s_0^2, s_1^2, s_2^2) &= \sum_{g=0}^{2} \log(f(s_g^2; \sigma_1^2, r)) \\
&= -\log(A) + \frac{N_0-3}{2}(\log(N_0-1) + \log(s_0^2) + \log(r) - \log(\sigma_1^2)) \\
&\quad + \frac{N_1-3}{2}(\log(N_1-1) + \log(s_1^2) - \log(\sigma_1^2)) \\
&\quad + \frac{N_2-3}{2}(\log(N_2-1) + \log(s_2^2) - \log(\sigma_2^2) - \log(r)) \\
&\quad - \frac{(N_0-1)s_0^2 r + (N_1-1)s_1^2 + (N_2-1)s_2^2/r}{2\sigma_1^2}
\end{aligned}$$

where $A = 2^{\frac{N_0-1}{2}}\Gamma(\frac{N_0-1}{2})2^{\frac{N_1-1}{2}}\Gamma(\frac{N_1-1}{2})2^{\frac{N_2-1}{2}}\Gamma(\frac{N_2-1}{2})$

The first partial derivatives of the log-likelihood function with respect to σ_1^2 and r are

$$\frac{dl_1(\sigma_1^2, r|s_0^2, s_1^2, s_2^2)}{d\sigma_1^2} = -\frac{N_0 + N_1 + N_2 - 9}{2\sigma_1^2} + \frac{(N_0 - 1)s_0^2 r + (N_1 - 1)s_1^2 + (N_2 - 1)s_2^2/r}{2(\sigma_1^2)^2} \tag{2.22}$$

and

$$\frac{dl_1(\sigma_1^2, r|s_0^2, s_1^2, s_2^2)}{dr} = \frac{N_0 - 3}{2}\left(\frac{1}{r}\right) + \frac{N_2 - 3}{2}\left(-\frac{1}{r}\right) - \frac{(N_0 - 1)s_0^2 - (N_2 - 1)s_2^2/r^2}{2(\sigma_1^2)}. \tag{2.23}$$

Set the above partial derivatives to zero and the MLEs of σ_1^2 and r are obtained by solving the equations simultaneously (select only the positive root as $\hat{r}$)

$$\hat{r} = \frac{(N_1 - 1)s_1^2 c + \sqrt{((N_1 - 1)s_1^2 c)^2 + 4(N_0 - 1)(N_2 - 1)(1 - c^2)s_2^2 s_0^2}}{2(N_0 - 1)(1 - c)s_0^2} \tag{2.24}$$

and
$$\hat{\sigma}_1^2 = \frac{(N_0 - 1)s_0^2 \hat{r} + (N_1 - 1)s_1^2 + (N_2 - 1)s_2^2/\hat{r}}{N - 9}, \tag{2.25}$$

where $c = \frac{N_0 - N_2}{N - 9}$.

It is analytically difficult to show the Hessian determinant to be positive. Instead, I show under plausible scenarios that the hessian determinants for the simulated data are positive. I then conclude from the second partial derivative test that these MLEs indeed maximize the log-likelihood under the ordered alternative for reasonable values of ratio (Appendix C.3).

Log-Likelihood Function under the Null Hypothesis

Under the null model, the log-likelihood is reduced to a function of a single unknown variance parameter σ^2. The log-likelihood function under the null is:

$$
\begin{aligned}
l_0(\sigma^2|s_0^2,s_1^2,s_2^2) &= \sum_{g=0}^{2} \log(f(s_g^2;\sigma^2)) \\
&= -\log(A) + \frac{N_0-3}{2}(\log(N_0-1)+\log(s_0^2)-\log(\sigma^2)) \\
&\quad + \frac{N_1-3}{2}(\log(N_1-1)+\log(s_1^2)-\log(\sigma^2)) \\
&\quad + \frac{N_2-3}{2}(\log(N_2-1)+\log(s_2^2)-\log(\sigma^2)) \\
&\quad - \frac{(N_0-1)s_0^2+(N_1-1)s_1^2+(N_2-1)s_2^2}{2\sigma^2}
\end{aligned}
\tag{2.26}
$$

The unknown variance parameter σ^2 is then estimated by maximizing the above log-likelihood function. The first partial derivative of the log-likelihood function with respect to σ^2 is:

$$
\frac{\mathrm{d}l_0(\sigma^2|s_0^2,s_1^2,s_2^2)}{\mathrm{d}\sigma^2} = \frac{N_0+N_1+N_2-9}{2\sigma^2} + \frac{(N_0-1)s_0^2+(N_1-1)s_1^2+(N_2-1)s_2^2}{2\sigma^4}
\tag{2.27}
$$

Set the above partial derivative to zero and the MLE of σ^2 is obtained accordingly:

$$
\hat{\sigma}^2 = \frac{(N_0-1)s_0^2+(N_1-1)s_1^2+(N_2-1)s_2^2}{N-9}
\tag{2.28}
$$

The second partial derivative of the log-likelihood function is strictly negative:

$$
\begin{aligned}
\frac{dl_0(\sigma^2|s_0^2,s_1^2,s_2^2)^2}{d^2\sigma^2} &= \frac{N_0+N_1+N_2-9}{2}\left(\frac{1}{\sigma^4}\right) + \frac{(N_0-1)s_0^2+(N_1-1)s_1^2+(N_2-1)s_2^2}{\sigma^6} \\
&= \frac{1}{\sigma^4}\left(\frac{N-9}{2} - \frac{(N_0-1)s_0^2+(N_1-1)s_1^2+(N_2-1)s_2^2}{\sigma^2}\right) \\
&= \frac{1}{\sigma^4}\left(\frac{N-9}{2} - (N-9)\right) \\
&= -\frac{1}{\sigma^4}\left(\frac{N-9}{2}\right) < 0
\end{aligned}
\tag{2.29}
$$

so $\hat{\sigma}^2$ is indeed the MLE. Note that the MLEs derived above are biased. However, considering GWASs usually have sample sizes in the order of thousands for the detection of SNP main effects, the use of biased estimates should not present a major problem.

2.3.2   Calculation of Non-centrality Parameter and Statistical Power

Finally, I derive the non-centrality parameter of our LRT for variance heterogeneity to perform power calculation. The non-centrality parameter λ is computed as the expected difference of the test statistic Λ under the alternative and null hypotheses (Sham *et al.*, 2000). Let S_0^2, S_1^2, and S_2^2 denote the observed variance random variables for the major allele homozygote ($G=0$), heterozygote ($G=1$) and minor allele homozygote ($G=2$) groups, respectively. The expected value of the test statistic under the null is simply 1 since the test statistic follows a chi-square distribution with 1 degree of freedom. The expected value of the test statistic under the alternative

hypothesis can be derived:

$$
\begin{aligned}
\mathrm{E}(\Lambda) &= 2\mathrm{E}(l_1(\sigma_1^2, r|S_0^2, S_1^2, S_2^2) - l_0(\sigma^2|S_0^2, S_1^2, S_2^2)) \\
&= 2\mathrm{E}\Big(-\log(A) + \frac{N_0 - 3}{2}(\log(N_0 - 1) + \log(S_0^2) + \log(r) - \log(\sigma_1^2)) \\
&\quad + \frac{N_1 - 3}{2}(\log(N_1 - 1) + \log(S_1^2) - \log(\sigma_1^2)) + \frac{N_2 - 3}{2}(\log(N_2 - 1) \\
&\quad + \log(S_2^2) - \log(\sigma_2^2) - \log(r)) - \frac{(N_0 - 1)S_0^2 r + (N_1 - 1)S_1^2 + (N_2 - 1)S_2^2/r}{2\sigma_1^2} \\
&\quad - (-\log(A) + \frac{N_0 - 3}{2}(\log(N_0 - 1) + \log(S_0^2) - \log(\sigma^2)) \\
&\quad + \frac{N_1 - 3}{2}(\log(N_1 - 1) + \log(S_1^2) - \log(\sigma^2)) + \frac{N_2 - 3}{2}(\log(N_2 - 1) + \log(S_2^2) - \log(\sigma^2)) \\
&\quad - \frac{(N_0 - 1)S_0^2 + (N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{2\sigma^2})) \\
&= 2\mathrm{E}\Big(\frac{N_0 - 3}{2}(\log(r) - \log(\sigma_1^2) + \log(\sigma^2)) + \frac{N_1 - 3}{2}(-\log(\sigma_1^2) + \log(\sigma^2)) \\
&\quad + \frac{N_2 - 3}{2}(-\log(\sigma_1^2) - \log(r) + \log(\sigma^2)) - \frac{(N_0 - 1)S_0^2 r + (N_1 - 1)S_1^2 + (N_2 - 1)S_2^2/r}{2\sigma_1^2} \\
&\quad + \frac{(N_0 - 1)S_0^2 + (N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{2\sigma^2}\Big) \\
&= (N - 9)\log\frac{\sigma^2}{\sigma_1^2} + (N_0 - N_2)\log(r) - (\frac{(N - 9)\sigma_1^2}{\sigma_1^2} - \frac{(N - 9)\sigma^2}{\sigma^2}) \\
&= (N - 9)\log\frac{\sigma^2}{\sigma_1^2} + (N_0 - N_2)\log(r) \qquad\qquad (2.30)
\end{aligned}
$$

From Equation 2.18, the expected values of the sample variance variables are:

$$
\mathrm{E}(S_0^2) = \sigma_0^2 = \sigma_1^2/r \qquad\qquad (2.31)
$$

$$
\mathrm{E}(S_1^2) = \sigma_1^2 \qquad\qquad (2.32)
$$

$$
\mathrm{E}(S_2^2) = \sigma_2^2 = \sigma_1^2 r \qquad\qquad (2.33)
$$

The non-centrality parameter is thus $(N - 9)\log(\frac{\sigma^2}{\sigma_1^2}) + (N_0 - N_2)\log(r) - 1$.

2.4   Other Variance Heterogeneity Tests

A number of statistical tests have been proposed to test whether the variances of subgroups from the same population are equal (Bartlett, 1937; Brown and Forsythe, 1974; Levene, 1960), among which Levene's test has been frequently referenced for its robustness to violations of the normality assumption and other irregularities. Despite its popularity, Levene's test has been shown to be conservative (Keyes and Levy, 1997; O'Neill and Mathews, 2000) under unbalanced designs, in particular when the group size varies dramatically. This would very likely be the case if the study population is stratified by the observed genotypes of a genetic variant, which will almost always lead to unbalanced groups. A more serious disadvantage, pertaining to the specific context of prioritizing SNPs for genetic interactions, is that Levene's test does not take into account the theoretical observation that the conditional variance of the quantitative trait will either monotonically increase or decrease with the number of minor alleles when interactions are present. Alternative trend tests for increasing or decreasing phenotypic variability have been proposed (Fujino, 1979; Hines and Hines, 2000), yet are too computationally intensive to be conveniently adapted to the unique setting of genome-wide interaction testing. Here, I briefly describe four variance heterogeneity tests that will be compared against LRT.

2.4.1   Levene's Test

The original Levene's test is equivalent to performing a one-way ANOVA (analysis of variance) on the transformed variable $z_{gj} = |y_{gj} - \bar{Y}_g|$, where $\bar{Y}_g$ is the sample mean in the genotype group g ($= 0, 1, 2$) and $\bar{\bar{Z}}$ the overall mean. Levene's test statistic to assess whether the variances conditional on the genotype are equal is:

$$L = \frac{(N-3)\sum_{g=0}^{2} N_g (\bar{Z}_g - \bar{\bar{Z}})^2}{(3-1)\sum_{g=0}^{2}\sum_{j=1}^{N_g}(z_{gj} - \bar{Z}_g)^2} \tag{2.34}$$

Under the null hypothesis of variance homogeneity, Levene's test statistic follows an F-distribution with 2 (number of genotype groups minus one) and $N-3$ (N minus number of genotype groups) degrees of freedom.

2.4.2   Levene-type Trend Tests

Levene-type trend tests have been introduced as a generalization to assess alternative hypotheses that feature an increasing or decreasing trend in variance amongst ordered groups (Levene, 1960; Gastwirth *et al.*, 2009). Specifically, these tests assign a weighted score w_g to the genotype group g and examine whether the beta coefficient obtained by regressing Z_{gj} to the order weights w_g is zero or not under the null hypothesis of variance equality. The ordered weights can be defined by g for a linear trend and $\sqrt{g}$ for a quadratic trend (Gastwirth *et al.*, 2009). The slopes of these regression models can then be used to assess the statistical significance of a linear or quadratic trend in variance.

2.4.3   Bartlett's Test

Similar to the original Levene's test, Bartlett's test assesses the general alternative hypothesis that at least two subgroups differ in variance (Bartlett, 1937). Following the notation in Chapter 2, given the observed set of variances (s_0^2, s_1^2, s_2^2), the

Bartlett's test statistic is (Bartlett, 1937; Snedecor and Cochran, 1980):

$$B = \frac{(N-3)\ln\left(S_p{}^2\right) - \sum_{g=0}^2 (N_g - 1)\ln\left(S_g{}^2\right)}{1 + \frac{1}{3(3-1)}\left(\sum_{g=0}^2 \left(\frac{1}{N_g-1}\right) - \frac{1}{N-3}\right)} \qquad (2.35)$$

where

$$S_p^2 = \frac{\sum_{g=0}^2 (N_g - 1)S_g^2}{N-3} \qquad (2.36)$$

The test statistic asymptotically follows a chi-square distribution with 2 degrees of freedom. Bartlett's test is frequently criticized for its sensitivity to violations of the normality assumption. However, when the normality assumption does hold, Bartlett's test has been shown to provide better statistical power compared to Levene's test (Levene, 1960).

Chapter 3

Simulation Studies

Consider a GWAS with N individuals genotyped on M genetic markers. To maintain consistent genotype counts between simulation runs for a given SNP, the observed genotypes were forced to be in exact HWE (Equations 2.1, 2.2, and 2.3). The quantitative trait variable Y was simulated according to Equation 2.4. Without loss of generality, assume the absence of main effect from genetic variants in all scenarios considered. The main objectives of these simulation studies are to assess the performance of all five variance heterogeneity tests in terms of statistical power to prioritize genetic variants for interactions using Variance Prioritization (VP) and how robust they are against non-normality of the continuous traits. All simulation studies were carried out using statistical programming software R version 2.14.1 in the Linux environment (R Core Team, 2010).

3.1 Performance of VP to Detect Gene-Environment Interactions with Variance Heterogeneity Tests

3.1.1 Statistical Power

In this simulation study, I examined the power to detect gene-environment interactions using VP with five options of variance heterogeneity test compared to an exhaustive search in the same sets of simulated data. Two studies of small ($N =$ 2,000) and moderate ($N =$10,000) sample sizes were considered. For each sample size specification, combinations of MAFs (10%, 20%, and 40%), interaction beta coefficients ($\beta_3 =$ 0.05 and 0.08) and covariate beta coefficients ($\beta_2 =$ 0.35 and 0.5) were used to simulate the outcome phenotype. The choice for MAFs of the simulated genetic variants was motivated by the search for gene-environment interactions among common variants (MAF > 5%). The interaction and covariate main effect beta coefficients were also chosen to reflect the effect sizes that have been currently observed or projected in the literature (Goldstein, 2009). It is commonly assumed that the interaction effect sizes are likely to be an order of magnitude smaller than the SNP main effects. While environmental covariates, such as life style factors, usually capture a large proportion of variance in the phenotype. For instance, smoking status explains up to 13% of the phenotypic variability in sICAM-1 levels (Paré *et al.*, 2010).

For a single SNP, each simulation run returned a variance heterogeneity *p*-value and an interaction *p*-value, which were then used to generate empirical statistical power. Unless otherwise specified, each simulation run was repeated 5,000 times. The conventional power using an exhaustive search corresponded to the proportion of SNPs with interaction *p*-values passing a nominal significance level of 0.05 while

27

correcting for $M=$ 500,000 hypotheses simultaneously (interaction p-value $< 0.05/M$). For each VP p-value threshold (η), the corresponding VP power was defined as the proportion of SNPs that passed the Bonferroni correction at a nominal level of 0.05 for interaction testing (interaction p-value $< 0.05/K$), where K is the number of SNPs passing the variance heterogeneity test (variance heterogeneity p-value $< \eta$; see Figure 1). The optimal VP p-value threshold (η_o) was determined empirically using simulations and indicated by the VP p-value threshold at which the VP power was maximized.

Figures 2 and 3 depict the VP power as a function of the VP p-value threshold η in each scenario, where the optimal VP power is marked by the peak of the curve and the conventional power is represented by a flat line invariant to the choice of η. The same set of results is alternatively reported in Tables 2 and 3, showing the performance of all variance heterogeneity test to prioritize SNPs as reflected by their respective optimal VP thresholds η_o and optimal VP powers. In the range of effects considered for this simulation study, the variance explained by the environmental covariate was robustly related to its main effect beta-coefficient but the variance explained by interaction was driven by both the MAFs and the interaction beta coefficients. This is also mathematically evident from Equations 2.6 and 2.7 presented in Chapter 2.

These observations support the conclusion that VP is superior to an exhaustive search in all scenarios considered, irrespective of the chosen variance heterogeneity test. The interaction effect size, MAF of the SNP, sample size, and interplay of all of these factors influence the optimal VP threshold η_o, and naturally the optimal VP power (Deng and Paré, 2011). For a given sample size and comparable interaction effect sizes, the exhaustive search power stayed relatively fixed while the optimal
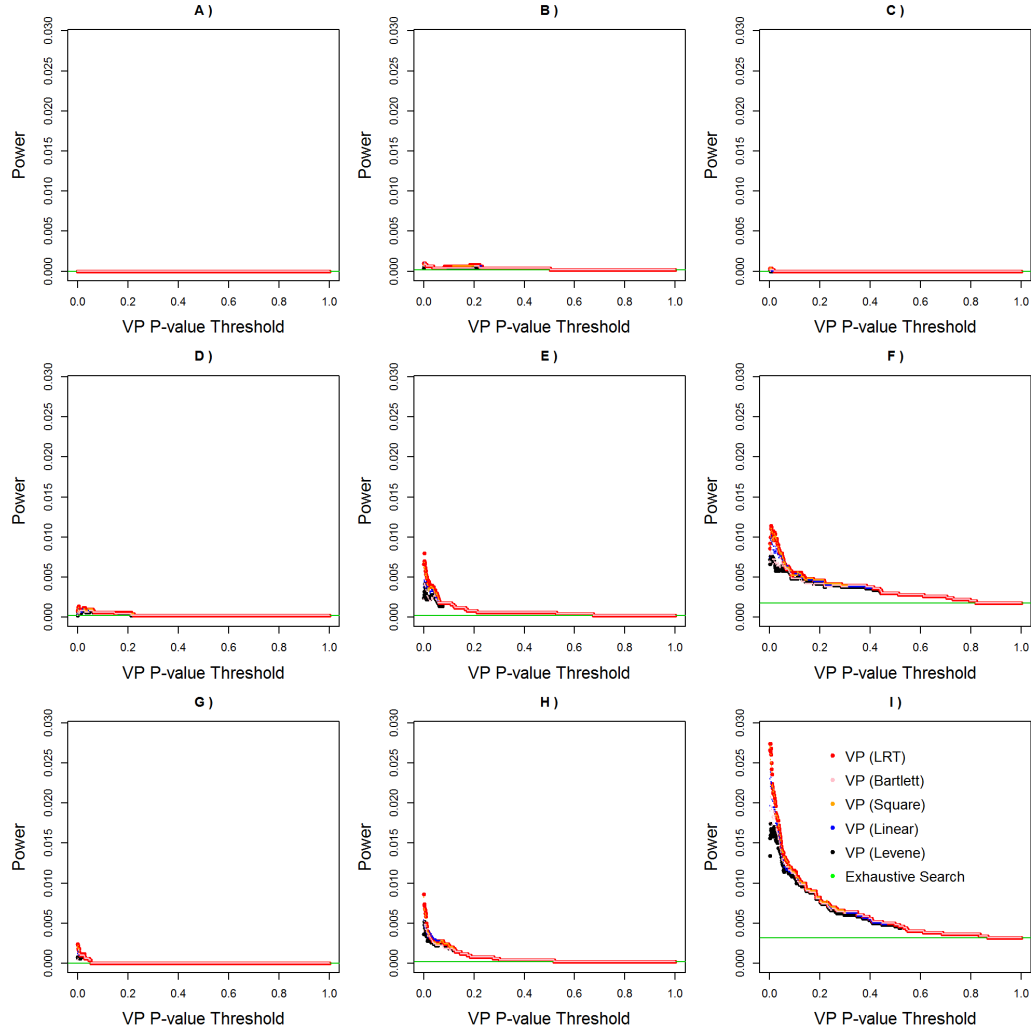
Figure 2: **Comparison of VP Power with Variance Heterogeneity Tests and an Exhaustive Search in a Sample of 2,000 Individuals** The proportion of variance explained by covariate was set at two levels: 10% (A-C) and 20% (D-I). Within each level, the interaction beta coefficient was set at 0.05 (A-C, D-F) and 0.08 (G-I). In addition, the MAF was fixed at 10% (A, D, G), 20% (B, E, H), and 40% (C, F, I). Each condition was simulated 5,000 times with 2,000 individuals. The horizontal line in green represents the power to detect an interaction with linear regression after correcting for $M = 500,000$ SNPs (p-value $< 0.05/M$). The coloured curves represent the power of VP at each variance heterogeneity p-value threshold ranging from 0.001 to 1 with 0.001 incremental increases. The power of VP is maximized at the optimal p-value threshold represented by the peak in the curve.
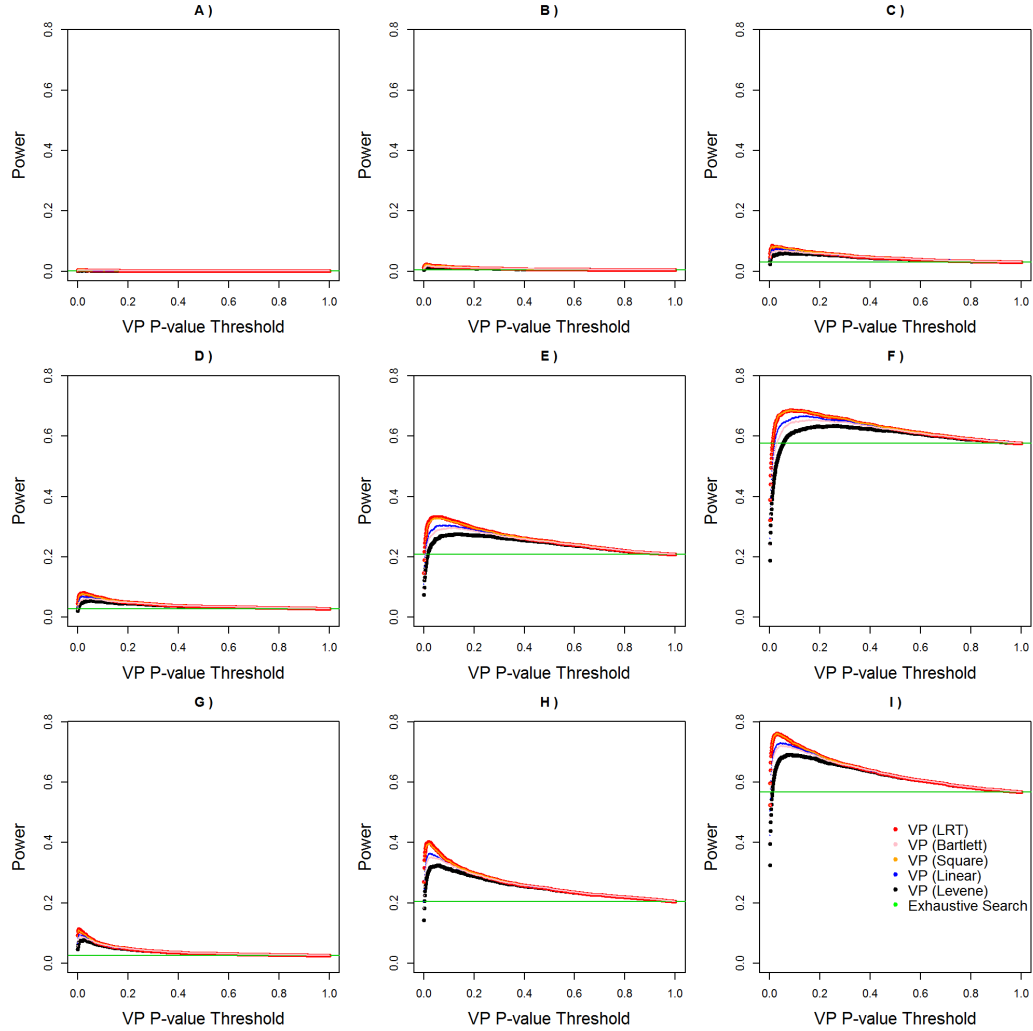
Figure 3: **Comparison of VP Power with Variance Heterogeneity Tests and an Exhaustive Search in a Sample of 10,000 Individuals** The proportion of variance explained by covariate was set at two levels: 10% (A-C) and 20% (D-I). Within each level, the interaction beta coefficient was set at 0.05 (A-C) and 0.08 (D-I). In addition, the MAF was fixed at 10% (A, D, G), 20% (B, E, H), and 40% (C, F, I). Each condition was simulated 5,000 times with 10,000 individuals. The horizontal line in green represents the power to detect an interaction with linear regression after correcting for $M = 500,000$ SNPs (p-value $< 0.05/M$). The coloured curves represent the power of VP at each variance heterogeneity p-value threshold ranging from 0.001 to 1 with 0.001 incremental increases. The power of VP is maximized at the optimal p-value threshold represented by the peak in the curve.

Simulation Parameters				Simulation Results			
β_2	β_3	MAF	(θ_{GE}, θ_E)	ES Power	VP Tests	η_0	Optimal VP Power
0.35	0.05	0.1	(0.040%, 10.9%)	0.00%	LRT	0.001	0.00%
					Levene's Test	0.001	0.00%
					LV Linear	0.001	0.00%
					LV Square	0.001	0.00%
					Bartlett	0.001	0.00%
		0.4	(0.107%, 10.9%)	0.00%	LRT	0.001	0.04%
					Levene's Test	0.001	0.02%
					LV Linear	0.001	0.02%
					LV Square	0.001	0.04%
					Bartlett	0.001	0.02%
	0.08	0.1	(0.103%, 10.9%)	0.02%	LRT	0.004	0.14%
					Levene's Test	0.019	0.10%
					LV Linear	0.005	0.10%
					LV Square	0.004	0.12%
					Bartlett	0.002	0.08%
		0.4	(0.273%, 10.9%)	0.18%	LRT	0.005	1.14%
					Levene's Test	0.004	0.76%
					LV Linear	0.008	1.02%
					LV Square Trend	0.006	1.10%
					Bartlett	0.008	0.86%
0.50	0.05	0.1	(0.036%, 20.0%)	0.00%	LRT	0.001	0.06%
					Levene's Test	0.002	0.02%
					LV Linear	0.001	0.02%
					LV Square	0.001	0.06%
					Bartlett	0.001	0.04%
		0.4	(0.096%, 20.0%)	0.04%	LRT	0.001	0.28%
					Levene's Test	0.016	0.18%
					LV Linear	0.002	0.24%
					LV Square	0.001	0.28%
					Bartlett	0.010	0.20%
	0.08	0.1	(0.096%, 20.0%)	0.00%	LRT	0.001	0.24%
					Levene's Test	0.018	0.10%
					LV Linear	0.002	0.14%
					LV Square	0.002	0.22%
					Bartlett	0.005	0.16%
		0.4	(0.245%, 20.0%)	0.32%	LRT	0.002	2.74%
					Levene's Test	0.004	1.74%
					LV Linear	0.004	2.38%
					LV Square	0.001	2.70%
					Bartlett	0.004	2.08%

Table 2: **Estimated VP Power to Detect a Gene-Environment Interaction in a Sample of 2,000 Individuals** The main effect of the SNP was set to 0 throughout. Variance explained by covariate and interaction was calculated using beta-coefficients and also to reflect effect sizes as a function of MAF. SNPs with variance heterogeneity p-values less than their respective optimal VP p-value thresholds were then selected for interaction with a continuous environmental covariate explaining either $\sim$10% or 20% of the phenotypic variance. Each condition was simulated 5,000 times with 2,000 individuals. The power to detect an interaction with linear regression using an exhaustive search after correcting for $M = 500{,}000$ SNPs (p-value $< 0.05/M$) was recorded for each scenario (ES Power). Optimal VP power represents the power at the optimal VP p-value threshold for each variance heterogeneity test. Levene's test with a linear trend alternative was denoted LV Linear and Levene's test with a square trend alternative was denoted LV Square.

Simulation Parameters				Simulation Results			
β_2	β_3	MAF	(θ_{GE}, θ_E)	ES Power	VP Tests	η_0	Optimal VP Power
0.35	0.05	0.1	(0.040%, 10.9%)	0.12%	LRT	0.014	0.52%
					Levene's Test	0.007	0.30%
					LV Linear	0.014	0.38%
					LV Square	0.013	0.50%
					Bartlett	0.007	0.34%
		0.4	(0.107%, 10.9%)	3.02%	LRT	0.010	8.48%
					Levene's Test	0.056	6.14%
					LV Linear	0.027	7.32%
					LV Square	0.012	8.24%
					Bartlett	0.037	6.76%
	0.08	0.1	(0.103%, 10.9%)	2.84%	LRT	0.022	8.06%
					Levene's Test	0.055	5.60%
					LV Linear	0.023	6.78%
					LV Square	0.022	7.72%
					Bartlett	0.056	6.16%
		0.4	(0.273%, 10.9%)	57.68%	LRT	0.085	68.70%
					Levene's Test	0.266	63.48%
					Linear Trend	0.136	66.54%
					Square Trend	0.85	68.60%
					Bartlett	0.189	65.28%
0.50	0.05	0.1	(0.036%, 20.0%)	0.04%	LRT	0.002	0.96%
					Levene's Test	0.003	0.40%
					LV Linear	0.002	0.64%
					LV Square	0.002	0.90%
					Bartlett	0.004	0.58%
		0.4	(0.096%, 20.0%)	2.70%	LRT	0.009	11.04%
					Levene's Test	0.019	8.22%
					LV Linear	0.009	9.74%
					LV Square	0.007	10.92%
					Bartlett	0.019	9.02%
	0.08	0.1	(0.096%, 20.0%)	2.52%	LRT	0.005	11.34%
					Levene's Test	0.023	7.76%
					LV Linear	0.009	9.32%
					LV Square	0.005	10.92%
					Bartlett	0.009	8.90%
		0.4	(0.245%, 20.0%)	56.72%	LRT	0.029	76.18%
					Levene's Test	0.078	69.16%
					LV Linear	0.043	72.84%
					LV Square	0.028	75.96%
					Bartlett	0.048	71.96%

Table 3: **Estimated VP Power to Detect a Gene-Environment Interaction in a Sample of 10,000 Individuals** The main effect of the SNP was set to 0 throughout. Variance explained by covariate and interaction was calculated using beta-coefficients and also to reflect effect sizes as a function of MAF. SNPs with variance heterogeneity p-values less than their respective optimal VP p-value thresholds were then selected for interaction with a continuous environmental covariate explaining either ∼10% or 20% of the phenotypic variance. Each condition was simulated 5,000 times with 2,000 individuals. The power to detect an interaction with linear regression using an exhaustive search after correcting for $M = 500,000$ SNPs (p-value $< 0.05/M$) was recorded for each scenario (ES Power). Optimal VP power represents the power at the optimal VP p-value threshold for each variance heterogeneity test. Levene's test with a linear trend alternative was denoted LV Linear and Levene's test with a square trend alternative was denoted LV Square.

VP powers using variance heterogeneity tests increased with the effect sizes of the covariate (Tables 2 and 3). For instance, when the interaction explained 0.182% and 0.163% of the variance, the exhaustive search power was 20.9% and 20.48%, respectively. However, the optimal VP power using LRT increased from 33.30% to 40.16%, a result of an almost 10% increase in variance explained by the covariate (Figure 3-E, H). For the gene-environment interaction effect sizes considered above, LRT consistently outperformed not only the original Levene's test but also the trend tests at their respective optimal VP p-value thresholds. The optimal VP p-value thresholds determined using LRT were lower than that by Levene-type tests and Bartlett's test, which suggest that LRT is more sensitive to variance heterogeneity induced by gene-environment interactions. This observation reflects the increased power of VP using LRT to identify gene-environment interactions in various scenarios.

3.1.2   Sample Size

The sample size required to detect gene-environment interactions at 80% power was calculated for VP using Levene's test, using LRT, and an exhaustive search. The following simulation parameters were considered to calculate the sample size: the main effect of the environmental covariate (= 5%, 15%, and 20%), and the interaction effect (= 0.05% to 1% with 0.05% incremental increases). The MAF of the interacting SNP influenced the optimal VP power only marginally, and was consequently fixed at 20% for simplicity. The sample size required to detect an interaction using a linear regression model alone was determined at a nominal significance level of 0.05 while correcting for simultaneous hypothesis testing of $M = 500{,}000$ SNPs. For sample size calculations using VP with Levene's test or LRT, the nominal level was also

fixed at 0.05; however, correcting only for the proportion of SNPs prioritized at their respective optimal VP p-value thresholds.

Since the optimal p-value threshold is influenced by many factors, namely, variance explained by interaction, environmental covariate, MAF, and sample size (Deng and Paré, 2011), the sample size required to detect an interaction using VP was iteratively determined by the optimal VP p-value threshold that maximized the power for a given effect size using LRT or Levene's test. When the interaction effect is small (0.1% to 0.4%), the use of VP with LRT lead to significantly reduced sample size to detect genetic interactions compared to with Levene's test (Table 4). For both LRT and Levene' test, the greatest reduction in sample size occurred when the interacting covariate explained a large proportion of the phenotypic variance. For instance, when the covariate explained 20% of the phenotypic variance, VP using LRT and Levene's test led to a 18.4% and 10.5% reduction in sample size to detect an interaction that explained 0.1% of the total variance (Table 4), respectively. This observation could be used to guide covariate selection for the most gain in power when a variety of interacting covariates are available.

As previously observed (Paré *et al.*, 2010), the sample size required to detect an interacting SNP using an exhaustive search drops substantially with large interaction effect sizes (0.5% to 1%), such that the reduction in sample sizes provided by VP using either LRT or Levene's test tended to be minimal (Figure 4). Nevertheless, while the relative advantage of VP decreases with increasing interaction effect sizes, optimal VP power is always superior or at least equivalent to an exhaustive search. In particular, if the interaction effect sizes followed the infinitesimal model (Fisher, 1958; Visscher *et al.*, 2008), such that there are many genetic variants of small effects,

θ_{GE} (%)	Exhaustive Search	$\theta_E = 5\%$		$\theta_E = 10\%$		$\theta_E = 15\%$		$\theta_E = 20\%$	
		Levene	LRT	Levene	LRT	Levene	LRT	Levene	LRT
0.05	76073	76068	76072	75043	71472	69108	64533	67558	61793
0.1	38025	38023	38019	37285	35304	34642	31917	33815	30804
0.15	25342	25339	25334	24604	23642	23092	21221	22661	20533
0.2	19000	19000	18984	18472	17506	17370	15894	16680	15435
0.25	15196	15192	15191	14767	14002	13786	12637	13427	12259
0.3	12659	12658	12651	12252	11625	11402	10495	11161	10127
0.35	10847	10846	10809	10588	9950	9766	9052	9582	8717
0.4	9488	9487	9386	9180	8675	8549	7924	8368	7632
0.45	8431	8431	8425	8154	7764	7609	6903	7376	6830
0.5	7586	7586	7505	7321	6839	6825	6250	6662	6128
0.55	6894	6894	6804	6649	6323	6207	5714	6093	5496
0.6	6318	6317	6166	6113	5719	5694	5197	5541	5091
0.65	5830	5829	5744	5622	5316	5278	4823	5114	4671
0.7	5412	5410	5272	5222	4911	4869	4491	4750	4314
0.75	5049	5047	4929	4865	4587	4560	4137	4416	4028
0.8	4732	4732	4600	4589	4280	4222	3863	4145	3770
0.85	4452	4451	4340	4294	4022	4011	3633	3910	3543
0.9	4204	4198	4080	4050	3784	3761	3429	3683	3375
0.95	3981	3980	3854	3823	3594	3586	3270	3469	3162
1	3781	3781	3685	3613	3394	3403	3052	3320	3007

Table 4: **Sample Size at 80% Power to Detect a Gene-Environment Interaction using VP and Exhaustive Search** MAF of the interacting SNP is set at 20%. The exhaustive search sample size at 80% power to detect an interaction with linear regression after correcting for $M = 500,000$ SNPs (p-value $< 0.05/M$) as a function of interaction effect size alone. The VP sample size at 80% power to detect an interaction with linear regression as a function of interaction effect size and covariate explained, accounting for SNPs acting on variance heterogeneity using either Levene's test or LRT (p-value $< 0.05/K$), where K is the number of SNPs prioritized.
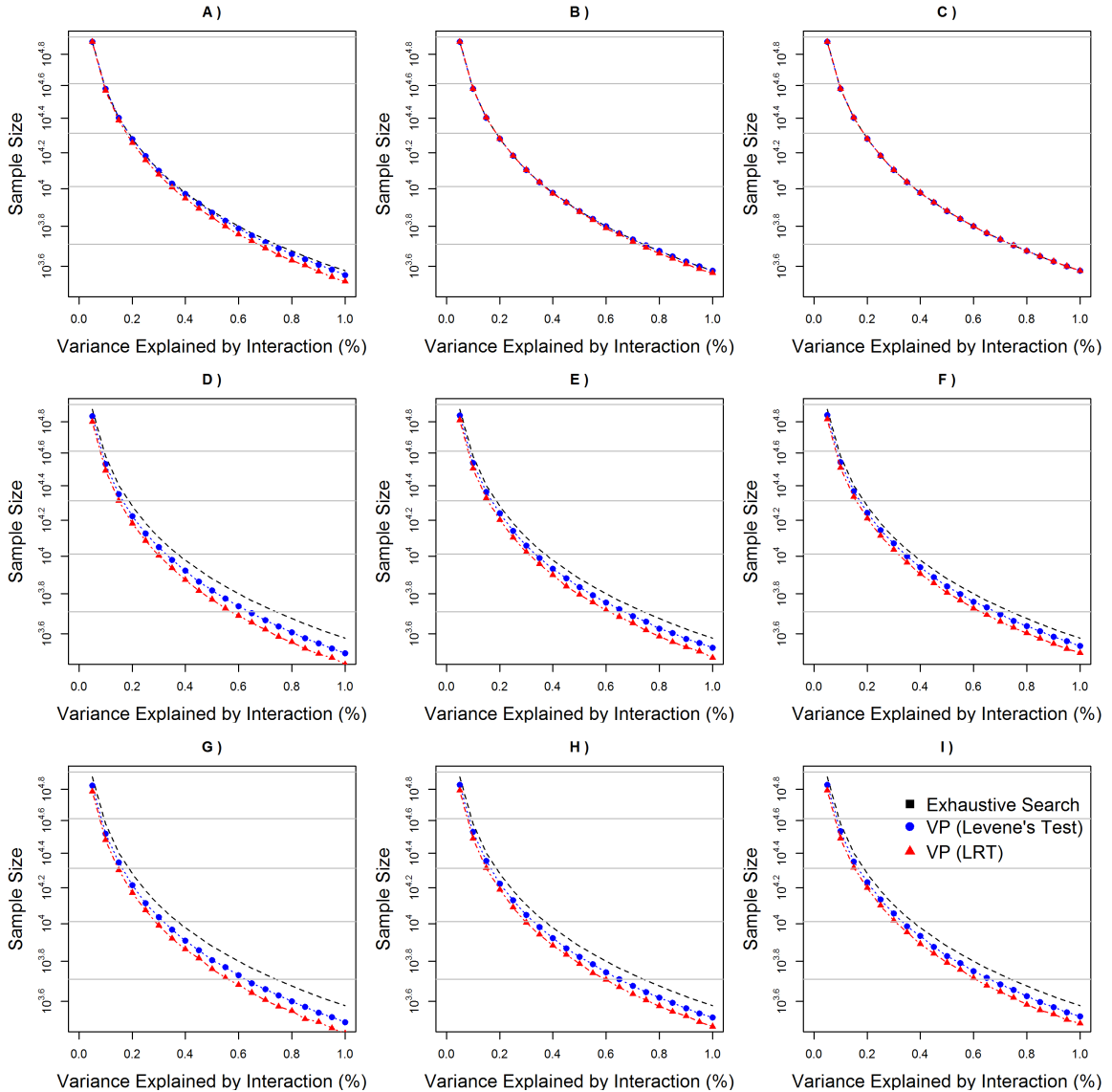
Figure 4: **Sample Size at 80% Power to Detect a Gene-Environment Interaction using VP and an Exhaustive Search on a** $\log 10$ **scale** The variance explained by the covariate was set at four levels: 5% (A-C), 15% (D-F), and 20% (G-I). Within each level, the MAF was set at 5% (A, D, G), 20% (B, E, H), and 40% (C, F, I). The black squares represent the sample size on a $\log 10$ scale at 80% power to detect an interaction with linear regression after correcting for $M = 500,000$ SNPs (p-value $< 0.05/M$, i.e. an exhaustive search) at each interaction effect size. The blue circles and the red triangles represent the VP sample size on a $\log 10$ scale at 80% power to detect an interaction with linear regression accounting for SNPs acting on variance heterogeneity using Levene's test and LRT, respectively. The Bonferroni corrected interaction p-value threshold is $0.05/(M\eta_0)$, where η_0 is the optimal VP Levene's test or LRT p-value threshold.

VP would be advantageous in prioritizing the thousands of weakly interacting SNPs underlying the genetic architecture of complex traits.

3.1.3   Computational Efficiency

In all simulated scenarios, LRT performed at least equally well when compared to Bartlett's test and the trend versions of Levene's test in terms of statistical power. Further, if prioritization were to be performed genome-wide, LRT would strike a better balance between computational efficiency and statistical performance. The same set of simulated data with 500,000 SNPs genotyped on 15,000 individuals was used to compare the computational time used by each variance heterogeneity test. Under the same computing environment (Red Hat Enterprise Linux 5 with dual Intel CPUs with 16 cores), it took 5.84 hours, 4.64 hours, 4.76 hours and 0.44 hours, respectively, to carry out Levene's test, its two modifications against ordered alternatives and Bartlett's test, whereas LRT was completed in 0.28 hours. In comparison, an exhaustive search of gene-environment interactions using linear regression under the same computing environment was completed in 3.7 hours. The closed-form solutions to the maximum likelihood estimators of LRT greatly speeded up the computation, an evident advantage in genome-wide settings.

3.2 Impact of non-normality on Variance Heterogeneity Tests

The proposed LRT statistic relies on distributional assumptions to derive the probability density functions of the sample variances; therefore its performance is contingent upon normality of the continuous trait. Similarly, Bartlett's test has also been shown to be sensitive to departure from normality (Conover *et al.*, 1981). In this simulation study, I examined the empirical type I error rate associated with each variance heterogeneity test when the error term followed: 1) a Student's t-distribution with k degrees of freedom, where $k = 5, 10, 20, 50$; and 2) a skew-normal distribution with a shape parameter of 0.5, 1, 2, and 4 under the null hypothesis of no interaction (Figure 5). To evaluate this simulation study in reference to the empirical analyses presented in Chapter 4, a sample of 2,000 individuals was generated. In addition, four scenarios including a MAF of 5%, 10%, 20% and 40% were considered.
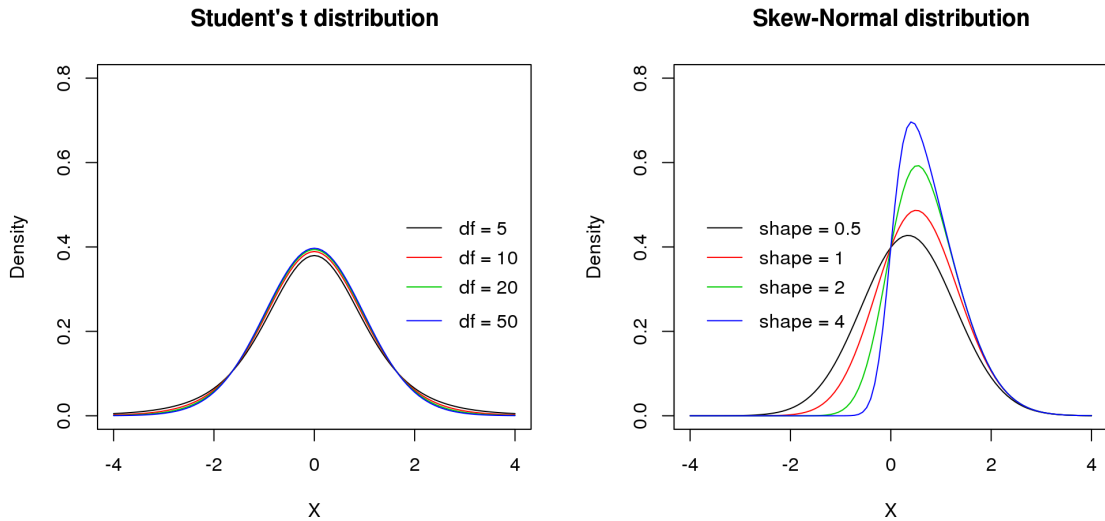
Figure 5: Probability Density Plots of Student's t and Skew-Normal Distributions

Since no systematic difference was observed across the spectrum of MAFs (tables with other choices of MAFs are available in Appendix C.4), only results for SNPs with a MAF of 20% are reported. The empirical type I error rate was calculated as the proportion of SNPs with variance heterogeneity test p-values less than the nominal α-level.

Type I error rates associated with different combinations of skewness and kurtosis estimated through simulations are shown in Table 5. No noticeable inflation was detected in all five tests when the theoretical skewness and kurtosis were 0 and 3.13 (a Student's t-distribution with 50 df, mesokurtosis designated by 3), respectively. However, LRT and Bartlett's test were more sensitive to large values of kurtosis than Levene-type tests. Specifically, LRT and Bartlett's test both had various levels of inflation in type I error rates when the error term followed a Student's t-distribution with degrees of freedom less than 20, whereas the others had satisfactory type I error rates. For example, when the error term followed a Student's t-distribution with 5 degrees of freedom, the mean sample skewness was approximately zero and mean sample kurtosis was 7.6992 (± 0.1320), the empirical type I error rate of Levene and its trend tests did not deviate considerably from the nominal levels of 0.01, 0.05 and 0.001, while LRT was associated with higher type I error rates than the nominal α-levels. Nevertheless, when the distribution had moderate values of kurtosis (< 3.5) and sample skewness was less than 0.5; all five tests provided satisfactory results. For example, when the error followed a skew-normal distribution with shape parameter less or equal to 2, type I error rates of all tests were adherent to the null distribution.

Simulated data with imperfections helped to further evaluate the effect violation

Type of Distribution		Student's t Distribution Degrees of Freedom				Skew-Normal Distribution Shape Parameter			
		5	10	20	50	0.5	1	2	4
Theoretical Skewness		0.000	0.000	0.000	0.000	0.0239	0.137	0.454	0.784
Theoretical Kurtosis		9.000	4.000	3.375	3.130	3.006	3.062	3.305	3.633
Mean Sample Skewness		-0.013	-0.001	-0.001	-0.001	0.024	0.137	0.454	0.781
Standard Error		0.0072	0.0016	0.0010	0.0008	0.0008	0.0008	0.0009	0.0009
Mean Sample Kurtosis		7.699	3.983	3.369	3.125	3.000	3.057	3.300	3.617
Standard Error		0.1320	0.0067	0.0027	0.0018	0.0015	0.0017	0.0025	0.0035
α	Variance Tests								
0.001	LRT	0.0648	0.0078	0.0024	0.0016	0.0008	0.0008	0.0014	0.0050
	Bartlett	0.0918	0.0082	0.0032	0.0010	0.0016	0.0012	0.0018	0.0060
	Levene	0.0020	0.0020	0.0006	0.0014	0.0016	0.0008	0.0010	0.0036
	LV Linear	0.0014	0.0016	0.0002	0.0012	0.0012	0.0008	0.0008	0.0022
	LV Square	0.0018	0.0014	0.0012	0.0008	0.0008	0.0004	0.0004	0.0010
0.01	LRT	0.1414	0.0332	0.0150	0.0132	0.0074	0.0090	0.0160	0.0254
	Bartlett	0.1892	0.0410	0.0198	0.0130	0.0098	0.0098	0.0202	0.0330
	Levene	0.0122	0.0082	0.0100	0.0092	0.0094	0.0096	0.0124	0.0222
	LV Linear	0.0118	0.0084	0.0078	0.0092	0.0070	0.0074	0.0106	0.0192
	LV Square	0.0100	0.0084	0.0080	0.0110	0.0064	0.0074	0.0082	0.0106
0.05	LRT	0.2562	0.1020	0.0728	0.0568	0.0464	0.0500	0.0714	0.0828
	Bartlett	0.3306	0.1260	0.0816	0.0588	0.0466	0.0536	0.0750	0.1032
	Levene	0.0534	0.0442	0.0504	0.0490	0.0454	0.0542	0.0604	0.0826
	LV Linear	0.0494	0.0478	0.0486	0.0570	0.0438	0.0516	0.0540	0.0724
	LV Square	0.0468	0.0466	0.0506	0.0496	0.0444	0.0432	0.0504	0.0502
0.1	LRT	0.3398	0.1706	0.1330	0.1156	0.0912	0.1018	0.1308	0.1544
	Bartlett	0.4238	0.2020	0.1556	0.1152	0.0960	0.1100	0.1354	0.1738
	Levene	0.1002	0.0930	0.0986	0.1002	0.1000	0.1028	0.1154	0.1492
	LV Linear	0.1088	0.0936	0.1024	0.1076	0.0930	0.0982	0.1096	0.1368
	LV Square	0.0950	0.0936	0.0990	0.1068	0.0890	0.0954	0.1024	0.0964

Table 5: **Type I Error Rates Associated with Non-normal Distributions (MAF = 20%)** Each condition was simulated 5,000 times with 2,000 individuals. A MAF of 20% was used throughout. The empirical type I error rate was calculated as the proportion of SNPs with variance heterogeneity p-values less than the nominal α-level. Levene's test with a linear trend alternative was denoted LV Linear and Levene's test with a square trend alternative was denoted LV Square.

and the nature of the violation has on the type I error rate. Normality of the continuous quantitative trait is often assumed in practice. However, in actuality, this assumption is almost always violated to some degree and transformation may be necessary. Although sample kurtosis and skewness offer useful insight in assessing deviation from normality, it is more important to inspect the empirical distribution of the sample population for additional irregularities. I conclude that Bartlett's test is the most sensitive to non-normally distributed data while the modified Levene's tests for trend alternatives are comparably more robust choices. Such irregularities in the non-transformed continuous trait may cause the LRT statistic to significantly deviate from the null distribute of variance homogeneity and thus induce inflated type I error rates. In the context of VP, this inflation leads to a lower optimal VP p-value threshold in the first stage with more SNPs prioritized, and as a result, compromises the optimal VP power to identify interactions in the second stage.

Chapter 4

Empirical Analysis of Variance Heterogeneity and Variance Prioritization

Genome-wide analyses were conducted to identify SNPs with variance heterogeneity using LRT, and subsequently whether these SNPs could potentially be interacting with environmental covariates using linear regression models. The genome-wide dataset is accessible through the database of Genotype and phenotype (dbGap) (Mailman *et al.*, 2007) as a part of the Multi Ethnic Study of Atherosclerosis (MESA) study depository. The MESA cohort (Study accession: phs000209.v10.p2, `http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000209.v10.p2`) consists of a population-based sample of 7,258 asymptomatic individuals, about 30% of which are of Caucasian origin (Bild *et al.*, 2002). Considering the genetic and lifestyle variations among different ethnicities, my investigation was restricted to an ethnically and geographically homogeneous subgroup of European Caucasians

according to self reported with ancestry confirmed through principal component analysis in the study report. Statistical analyses were primarily conducted using PLINK (`http://pngu.mgh.harvard.edu/~purcell/plink/`); (Purcell *et al.*, 2007) and the statistical program R version 2.14.1 (R Core Team, 2010) in the Linux environment.

Genotype Data Genotyping was performed on Affymetrix Human SNP array 6.0. The genotype data were initially cleaned prior to being deposited to dbGap, excluding monomorphic SNPs (that is, SNPs with only one observed genotype in all samples), SNPs with high missing rate (greater than 5% in total or within each race), and SNPs with observed heterozygosity greater than 53% (that is, SNPs with an excess of heterozygote genotypes than expected under HWE). For the subset of Caucasian individuals, I performed additional quality controls to the cleaned genotype data based on missing rate per SNP, MAF, and HWE. SNPs with high missing rates ($>$ 5%) were removed out of consideration for genotyping quality. I also decided to exclude SNPs with MAFs below 5%, which is higher than the conventional threshold of 1% for common variants, out of consideration for the calculation of per genotype variance in the minor allele homozygote genotype group (at least two individuals are needed to calculate variance). Additionally, SNPs with observed genotypes that deviated from the expected distribution under HWE were removed based on an α-level of 1E-06. 656,004 autosomal SNPs remained after the quality control (Table 6), the α-level with Bonferroni correction for declaring a statistical significant interaction is 0.05/656,004= 7.62E-08.

Quality Control Filters	SNPs Lost	SNPs Remained
SNP probes		934,940
Genotyping center filters	25,318	909,622
HWE $< 1E-6$ in Caucasian samples	3,279	
Missing rate per SNP $> 5\%$	7,312	
MAF $< 5\%$	223,242	
Percentage of SNPs removed due to filters	27.2%	
Non-autosomal SNPs	24,546	680,550
		656,004

Table 6: Summary of SNP Quality Controls

Phenotype Data Only unrelated individuals were retained based on self-report. Individuals with self-reported sex contradicting that estimated by X chromosome heterozygosity were excluded. In addition, I excluded individuals with diabetes as defined by plasma fasting glucose level greater than or equal to 7 mmol/L. Genotyping rates per subject in the remaining 2,166 European Caucasians were greater than 95%. A common set of quantitative trait variables including body mass index (BMI), waist circumference (WC), hip circumference, height, high density lipoprotein cholesterol (HDL-C), low density lipoprotein cholesterol (LDL-C), triglyceride (TG), fasting glucose (FG) and total cholesterol are available. Only directly measured biological traits with less than 20% missing data were selected for analysis. All continuous variables underwent log-transformation and were standardized to have mean 0 and variance 1. To eliminate outlier effects and reduce possible false positive findings, winsorization at three standard deviations was performed. Summary statistics of these transformed traits can be found in Table 7.

Transformation	Traits	Mean	Std Dev	Skewness	Kurtosis
No transformation	BMI (kg/m2)	27.54	4.96	0.86	4.18
	WC (cm)	97.44	14.22	0.42	4.27
	HDL-C (mg/dl)	53.03	15.91	1.00	4.45
	TG (mg/dl)	128.32	74.72	2.22	13.47
	FG (mg/dl)	87.25	8.98	0.26	3.25
log transformed, winsorized at 3 standard deviations	log(BMI)	0.00	1.00	0.28	3.03
	log(WC)	0.00	0.99	-0.12	3.04
	log(HDL-C)	0.00	1.00	0.16	2.81
	log(TG)	0.00	1.00	0.12	2.78
	log(FG)	0.00	0.98	0.09	2.78

Table 7: Summary Statistics of Selected Traits and Interacting Covariates

Model Selection Among the phenotypic variables passing quality control, I was primarily interested in TG, HDL-C, and FG because these traits are highly heritable and have established clinical roles in relation to cardiovascular risks (Teslovich *et al.*, 2010; Edmondson and Rader, 2008; Willer *et al.*, 2008; Kathiresan *et al.*, 2007). In addition, these traits tend to fluctuate over time and thus are more susceptible to either changes in the environment or the interplay between environment and individual genetic profiles. BMI and WC were included as potentially interacting covariates, as it has been suggested that adiposity such as BMI modifies the influence of genetic variants on metabolic traits (Manning *et al.*, 2011). In light of the simulation results, I constructed linear regression models the form:

$$Y = \text{Age} + \text{Sex} + G + C + G \times C, \tag{4.1}$$

where G is the genotype of a biallelic SNP and C is the potentially interacting covariate. These models are selected for the VP procedure according to Pearson's

correlation coefficients calculated between the log-transformed response variables and log-transformed covariates (Table 8).

Response Variable	Interacting Covariate	Pearson's Correlation	Variance Explained
TG	BMI	0.319	10.07%
	WC	0.3228	10.39%
HDL-C	TG	-0.439	19.86%
	WC	-0.369	13.54%
	BMI	-0.316	9.907%
FG	WC	0.3475	11.59%
	BMI	0.283	8.223%

Table 8: Summary of Candidate Models for VP

4.1 Genome-Wide Variance Heterogeneity and Interaction Analyses

Two genome-wide analyses were conducted independently to examine the variance heterogeneity p-values of all autosomal SNPs and the effects of those SNPs interacting with adiposity measures on the quantitative traits.

4.1.1 Trait Specific Tests (Step 1 in Figure 1)

Variance Heterogeneity Levene's test and LRT were applied separately to each of the three outcome traits, i.e. TG, HDL-C, and FG. It was observed from the analysis using empirical data that the optimal performance of LRT relied on the normality assumption, as type I error rate inflation was associated with TG and HDL-C: both had skewness greater than 0.3 and kurtosis greater than 4 when no transformation was applied (Table 7, Figure 6). Similarly, type I error rates in Levene's test p-values for TG and HDL-C were also inflated due to the high kurtosis values. Transformation was necessary to maintain skewness and kurtosis measures similar to that of a normal distribution (Table 7). Consequently, the quantile-quantile plots of variance heterogeneity p-values for each of three transformed traits suggest the type I error rates are well controlled (Figure 7).
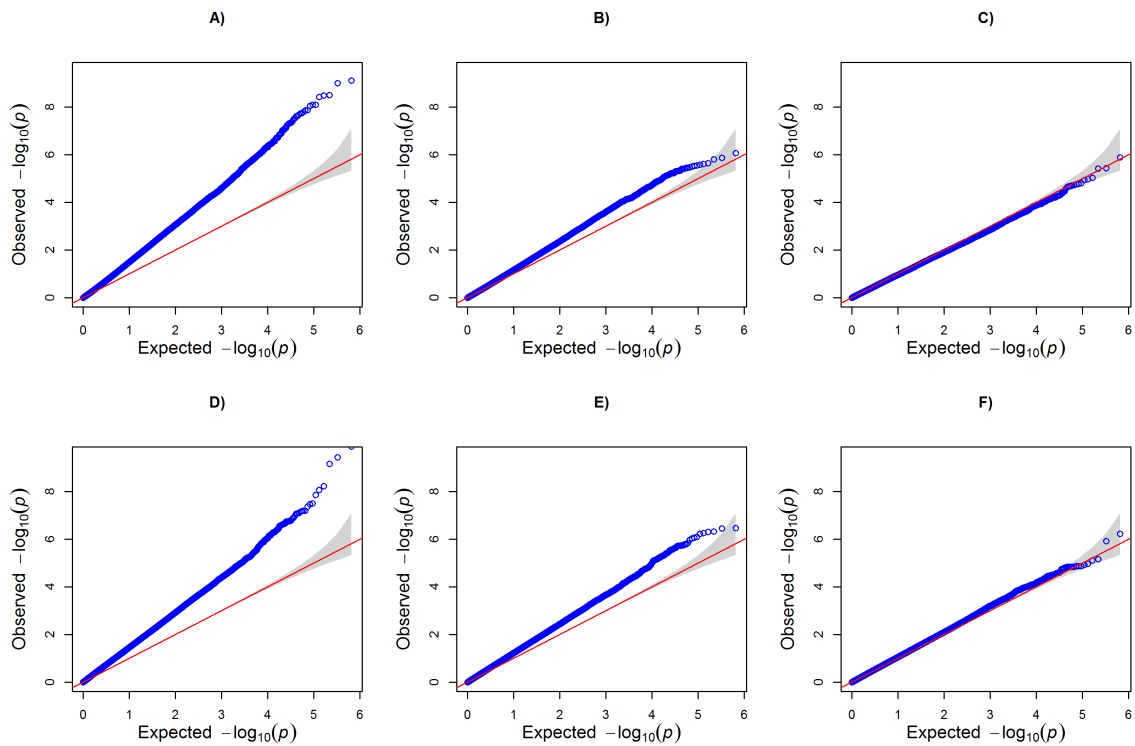
Figure 6: **Quantile-Quantile Plots of Variance Heterogeneity Test p-values using Untransformed Traits** Illustrated in the upper row are the quantile-quantile plots of LRT p-values for untransformed traits: TG (A), HDL-C (B), and FG (C). Illustrated in the lower row are the quantile-quantile plots of Levene's test p-values for untransformed traits: TG (D), HDL-C (E), and FG (F).
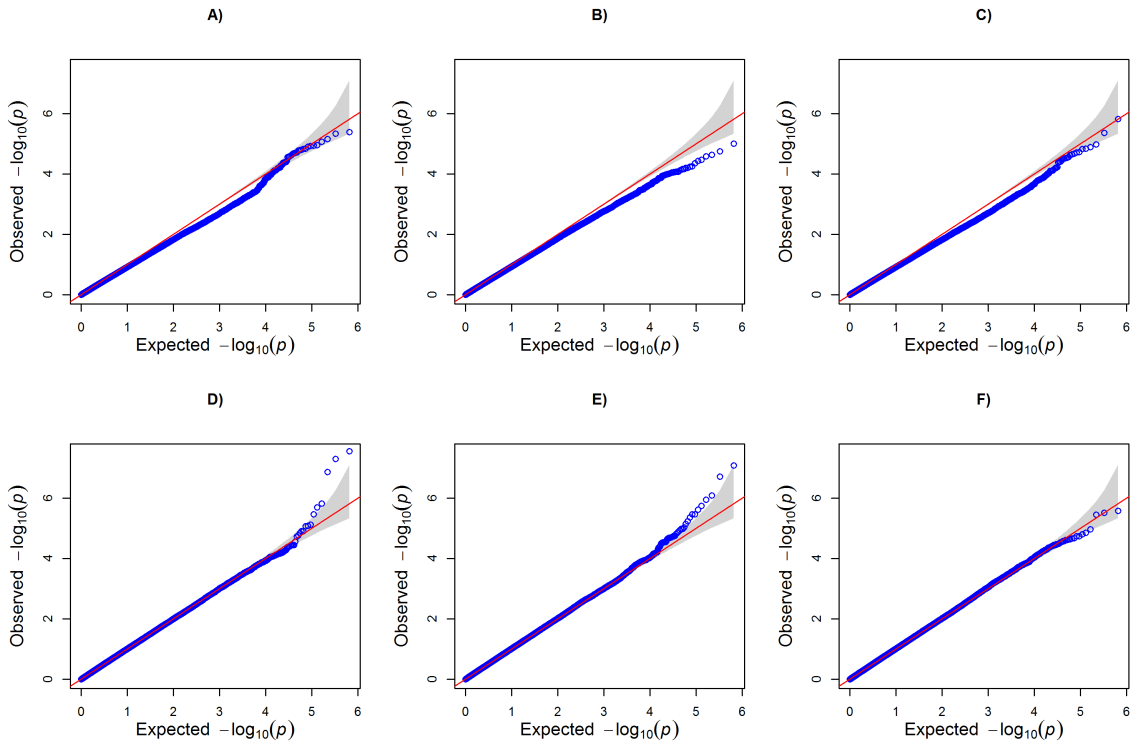
Figure 7: **Quantile-Quantile Plots of Variance Heterogeneity Test p-values using Transformed Traits** Illustrated in the upper row are the quantile-quantile plots of LRT p-values for TG (A), HDL-C (B), and FG (C) after log-transformation and winsorization at three standard deviations. Illustrated in the lower row are the quantile-quantile plots of Levene's test p-values for TG (D), HDL-C (E) and FG (F) after log-transformation and winsorization at three standard deviations.

Trait	SNP	Nearest Gene*	CHR	Position	MAFSD	per Genotype			Ratio	p-values	
						S_0	S_1	S_2		LRT	Levene
	rs1934101	ELAVL2	9	23220746	0.37	1.08	0.95	0.89	0.81	4.11E-06	9.74E-05
	rs10757400	ELAVL2	9	23187358	0.36	1.07	0.96	0.87	0.81	4.64E-06	9.47E-05
	rs2651976	TMCC3	12	93554392	0.46	1.08	0.99	0.88	0.82	6.82E-06	1.06E-04
	rs7742224	AKAP12	6	151618963	0.49	1.08	1.00	0.89	0.83	8.56E-06	3.62E-05
TG	rs1339142	ELAVL2	9	23209125	0.24	1.04	0.95	0.80	0.80	1.11E-05	1.88E-04
	rs2651985	TMCC3	12	93559305	0.45	0.91	1.00	1.09	1.21	1.15E-05	1.50E-04
	rs10811869	ELAVL2	9	23190174	0.34	1.06	0.95	0.88	0.82	1.18E-05	2.63E-04
	rs10058472	CDH12	5	21868337	0.26	1.06	0.91	0.90	0.81	1.23E-05	1.52E-04
	rs7874120	ELAVL2	9	23192654	0.35	1.06	0.95	0.88	0.82	1.45E-05	6.55E-04
	rs6878249	ITGA1	5	52210980	0.21	0.95	1.05	1.20	1.25	1.48E-05	3.42E-06
	rs12932428	FTO	16	52518028	0.48	1.05	1.03	0.85	0.82	9.81E-06	8.23E-07
	rs17757577	MMD	17	50964405	0.12	1.03	0.88	0.83	0.74	1.79E-05	2.95E-04
	rs1253426	ARHGAP19	10	98960900	0.08	1.02	0.85	0.82	0.71	2.28E-05	1.77E-05
	rs611150	PARP11	12	3875013	0.18	1.04	0.93	0.75	0.78	2.57E-05	4.01E-04
HDL-C	rs10887396	FAM190B	10	86652805	0.40	1.08	0.96	0.91	0.83	3.33E-05	2.96E-05
	rs17667367	GAD2	10	26564812	0.08	0.97	1.15	1.19	1.38	3.72E-05	1.09E-03
	rs10019557	LPHN3	4	61367329	0.15	1.04	0.88	0.95	0.78	4.34E-05	3.92E-05
	rs9364097	SMAP1	6	71418475	0.30	0.93	1.06	1.06	1.21	5.59E-05	1.03E-04
	rs9972377	MCTP2	15	92716873	0.22	1.04	0.94	0.85	0.81	5.74E-05	1.03E-03
	rs10852208	MCTP2	15	92717430	0.22	1.04	0.94	0.85	0.81	6.24E-05	1.27E-03
	rs9905394	HS3ST3B1	17	14216969	0.45	1.08	0.96	0.88	0.81	1.50E-06	4.39E-05
	rs7700535	COX7C	5	85671585	0.13	0.94	1.10	1.18	1.33	4.37E-06	2.93E-05
	rs11193637	SORCS1	10	109460510	0.18	0.94	1.06	1.20	1.29	1.04E-05	2.28E-05
	rs7727064	COX7C	5	85669119	0.12	0.95	1.09	1.18	1.32	1.30E-05	5.90E-05
FG	rs637278	CELF2	10	11259659	0.07	0.96	1.13	1.38	1.42	1.44E-05	7.83E-05
	rs636416	CELF2	10	11259496	0.07	0.96	1.13	1.38	1.42	1.50E-05	6.20E-05
	rs11958742	COX7C	5	85799252	0.11	0.95	1.10	1.12	1.32	1.86E-05	1.54E-04
	rs6725114	EPC2	2	149083368	0.09	0.95	1.12	1.16	1.36	1.94E-05	1.02E-04
	rs4792477	HS3ST3B1	17	14215811	0.42	1.06	0.96	0.89	0.83	2.10E-05	4.62E-04
	rs642349	CELF2	10	11260140	0.07	0.96	1.12	1.38	1.40	2.25E-05	1.76E-04

Table 9: **Top Ten SNPs Involved in Variance Heterogeneity Ranked by LRT p-values** Positions according to Human Genome Build 36.3. * Nearest genes are based on a maximum of 500KB distance

50

The top ten SNPs with the lowest LRT or Levene's test p-values for each of the three traits are presented in Tables 9 and 10, respectively. None of the SNPs showed genome-wide significant LRT or Levene's test p-value, however, a strong agreement between the LRT and Levene's test p-values is consistently observed. To bring perspective to these results, I searched for nearby SNPs associated at the genome-wide significance with any traits or disease in the catalogue of published GWAS (http://www.genome.gov/gwastudies/), filtering based on a maximum distance of 500KB and $r^2 > 0.8$ or $D' > 0.8$. I found that rs2197089 near the LPL gene, which is known to be associated with metabolic syndrome and lipid traits such as TG and HDL-C (Kraja $et\ al.$, 2011), was in weak linkage disequilibrium (LD) with rs1441771 ($r^2 = 0.119$; D' = 0.823; Distance = 69.07 KB) and rs12543154 ($r^2 = 0.112$; $D' = 0.820$; Distance = 70.40KB) that showed highly suggestive Levene's test p-value for TG. Nevertheless, this particular known SNP (rs2197089) did not show any evidence of variance heterogeneity (Levene's test and LRT p-value > 0.1) for neither TG nor HDL-C. None of the other top variance heterogeneous SNPs for TG, HDL-C or FG were linked to genetic variants associated with other traits or diseases in their neighbouring regions. It should be noted that an FTO variant (rs12932428) was identified with the lowest variance heterogeneity p-value using both LRT and Levene's test for HDL-C. Despite the weak LD between this variant and any of the known FTO genetic variants, the FTO gene has been shown to be functional and associated with obesity and type 2 diabetes (Meyre $et\ al.$, 2009; WTCC, 2007), and thus the identified variant may be of biological relevance.

Trait	SNP	Nearest Gene*	CHR	Position	MAF	SD per Genotype			Ratio	p-values	
						S_0	S_1	S_2		LRT	Levene
	rs6878249	ITGA1	5	52210980	0.21	0.95	1.05	1.20	1.25	1.48E-05	3.42E-06
	rs7732839	ITGA1	5	52209527	0.22	0.95	1.04	1.17	1.22	7.65E-05	3.62E-05
	rs7742224	AKAP12	6	151618963	0.49	1.08	1.00	0.89	0.83	8.56E-06	3.62E-05
	rs1005707	ITGA1	5	52210413	0.32	0.94	1.02	1.13	1.20	7.60E-05	4.10E-05
TG	rs12543154	LPL	8	19800257	0.18	1.03	0.92	0.76	0.79	4.24E-05	5.55E-05
	rs7742218	AKAP12	6	1516189448	0.50	1.07	1.00	0.89	0.83	1.62E-05	5.78E-05
	rs1441771	LPL	8	19801580	0.18	1.03	0.92	0.77	0.79	2.78E-05	6.05E-05
	rs7693763	LPHN3	4	62101599	0.39	0.90	1.03	1.07	1.21	2.03E-05	6.10E-05
	rs4236876	COL22A1	8	140002685	0.06	0.97	1.14	1.29	1.39	7.68E-05	6.98E-05
	rs5999369	-	22	33162649	0.20	0.95	1.06	1.15	1.24	6.49E-05	7.19E-05
	rs12932428	FTO	16	52518028	0.48	1.05	1.03	0.85	0.82	9.81E-06	8.23E-07
	rs13169679	FER	5	108039802	0.24	1.05	0.91	0.94	0.82	6.74E-05	1.11E-06
	rs869287	FER	5	108025195	0.23	1.05	0.91	0.94	0.82	7.12E-05	1.80E-06
	rs1355079	FER	5	108027404	0.24	1.05	0.91	0.94	0.82	9.67E-05	2.43E-06
HDL-C	rs10051840	FER	5	107991845	0.28	1.06	0.94	0.90	0.83	9.94E-05	1.17E-05
	rs1253426	ARHGAP19	10	98960900	0.08	1.02	0.85	0.82	0.71	2.28E-05	1.77E-05
	rs10887396	FAM190B	10	86652805	0.40	1.08	0.96	0.91	0.83	3.33E-05	2.96E-05
	rs10019557	-	4	61367329	0.15	1.04	0.88	0.95	0.78	4.34E-05	3.92E-05
	rs4360643	FAM190B	10	86628627	0.43	1.08	0.97	0.92	0.84	8.25E-05	5.41E-05
	rs17090285	-	4	61378175	0.15	1.04	0.88	0.95	0.79	7.60E-05	7.25E-05
	rs6877686	ERGIC1	5	172164368	0.09	0.95	1.08	1.46	1.33	5.84E-05	2.65E-06
	rs13277300	OXR1	8	107131995	0.42	1.04	0.99	0.84	0.83	9.77E-05	1.41E-05
	rs11193637	-	10	109460510	0.18	0.94	1.06	1.20	1.29	1.04E-05	2.28E-05
	rs11956702	APC	5	112010672	0.25	0.93	1.04	1.12	1.22	7.66E-05	2.28E-05
FG	rs3748712	ST6GALNAC3	1	76871071	0.10	0.95	1.11	1.10	1.32	7.84E-05	2.42E-05
	rs11193622	-	10	1094447952	0.18	0.94	1.06	1.17	1.27	2.71E-05	2.81E-05
	rs7700535	COX7C	5	85671585	0.13	0.94	1.10	1.18	1.33	4.37E-06	2.93E-05
	rs2647389	ST6GALNAC3	1	76845046	0.15	0.95	1.06	1.25	1.27	6.16E-05	4.09E-05
	rs9905394	HS3ST3B1	17	14216969	0.45	1.08	0.96	0.88	0.81	1.50E-06	4.39E-05
	rs12724872	ST6GALNAC3	1	76899334	0.10	0.95	1.10	1.17	1.33	6.21E-05	4.80E-05

Table 10: **Top Ten SNPs Involved in Variance Heterogeneity Ranked by Levene's Test p-values** Positions according to Human Genome Build 36.3. * Nearest genes are based on a maximum of 500KB distance

Figure 8: **Manhattan Plot of Gene-Environment Interaction p-values for TG**
Genome-wide results showing $-\log 10$ of the interaction p-value from linear regression models using TG as the response variable and BMI (A) or WC (B) as the interacting covariate. The blue and red horizontal lines represent a nominal p-value threshold at $1E-05$ and the genome-wide significance threshold of $5E-08$, respectively.

4.1.2   Covariate Specific Tests (Step 2 in Figure 1)

Gene-Environment Interaction To identify gene-environment interactions acting on these selected traits, I considered linear regression models with TG as the response variable, BMI or WC as the interacting covariate. Age and sex were included as covariates but their interactions with the genes were not explored. Similar regression models were examined with HDL-C or FG as the response variable (Table 7). Additionally, when the response variable was HDL-C, I also investigated the linear regression model with TG as the interacting covariate. The interaction results were presented using Manhattan plots (Figures 8 to 10), where the x-axis denotes the chromosomal location of each SNP and y-axis the p-value for a gene-environment

53

Trait	Covariate	SNP	Nearest Gene*	CHR	MAF	Ratio	Variance Het p-value		G-E Interaction	
							LRT	Levene's Test	β_3	Interaction p-value
TG	BMI	rs2882974	CTLA4	2	0.48	1.10	2.18E-02	1.43E-01	0.172	9.78E-07
		rs6767746	-	3	0.18	1.07	2.41E-01	1.24E-01	0.119	1.24E-06
		rs10179639	CTLA4	2	0.49	1.10	2.66E-02	1.42E-01	0.170	1.46E-06
		rs12485894	-	3	0.18	1.07	1.89E-01	7.42E-02	0.117	1.59E-06
		rs1447262	MPPED2	11	0.20	0.96	4.69E-01	3.07E-01	-0.120	1.94E-06
		rs7933755	MPPED2	11	0.18	0.97	6.20E-01	6.61E-01	-0.117	2.41E-06
		rs294363	MPPED2	11	0.20	0.96	4.89E-01	3.56E-01	-0.118	2.84E-06
		rs294365	MPPED2	11	0.20	0.97	5.20E-01	3.50E-01	-0.118	2.89E-06
		rs7896207	ARL5B	10	0.09	0.99	8.99E-01	4.41E-01	-0.104	4.08E-06
		rs12354939	ARL5B	10	0.08	1.00	9.86E-01	5.70E-01	-0.101	6.62E-06
	WC	rs2882974	CTLA4	2	0.48	1.10	2.18E-02	1.43E-01	0.170	1.09E-06
		rs10179639	CTLA4	2	0.49	1.10	2.66E-02	1.42E-01	0.166	2.09E-06
		rs9790882	ENC1	5	0.20	0.89	3.81E-02	1.26E-01	-0.116	3.44E-06
		rs7933755	MPPED2	11	0.18	0.97	6.20E-01	6.61E-01	-0.113	3.89E-06
		rs4879523	-	9	0.13	1.18	9.40E-01	1.10E-02	0.105	4.77E-06
		rs10813599	-	9	0.13	1.15	2.47E-02	2.97E-02	0.103	5.67E-06
		rs10970405	-	9	0.13	1.15	2.61E-02	2.58E-02	0.103	5.79E-06
		rs9791149	ENC1	5	0.20	0.90	5.62E-02	1.60E-01	-0.113	6.74E-06
		rs10970371	-	9	0.13	1.15	2.86E-02	4.01E-02	0.103	6.80E-06
		rs11515349	-	9	0.13	1.15	2.90E-02	3.83E-02	0.102	7.36E-06
HDL-C	BMI	rs4019375	PTPRN2	7	0.26	0.88	1.53E-02	1.14E-01	0.112	3.04E-06
		rs8069454	PRKCA	17	0.19	1.01	8.39E-01	9.69E-01	-0.101	4.24E-06
		rs1684901	CCDC6	10	0.46	0.87	1.61E-03	2.59E-03	0.136	5.47E-06
		rs8069142	PRKCA	17	0.19	1.01	8.40E-01	9.76E-01	-0.099	5.61E-06
		rs510321	NOTCH4	6	0.19	1.11	6.36E-02	2.79E-01	-0.097	7.02E-06
		rs9891403	PRKCA	17	0.19	1.01	8.37E-01	9.74E-01	-0.097	8.67E-06
		rs10993842	VAV2	9	0.14	0.91	1.40E-01	1.56E-01	0.090	1.02E-05
		rs11067772	MED13L	12	0.33	1.10	4.15E-02	9.36E-02	-0.113	1.09E-05
		rs6931344	TINAG	6	0.12	1.14	4.39E-02	2.44E-01	-0.089	1.11E-05
		rs6569294	SMPDL3A	6	0.45	0.94	1.17E-01	1.59E-01	0.131	1.12E-05
	WC	rs12002290	VAV2	9	0.16	0.89	4.44E-02	4.07E-02	0.099	3.22E-06
		rs1684901	CCDC6	10	0.46	0.87	1.61E-03	2.59E-03	0.140	3.46E-06
		rs11067772	MED13L	12	0.33	1.10	4.15E-02	9.36E-02	-0.115	6.04E-06
		rs10993842	VAV2	9	0.14	0.91	1.40E-01	1.56E-01	0.093	6.46E-06
		rs4271376	API5	11	0.46	0.96	3.51E-01	2.89E-01	0.133	7.00E-06
		rs12002767	VAV2	9	0.15	0.91	9.54E-02	1.23E-01	0.094	7.97E-06
		rs7122883	API5	11	0.49	0.96	3.92E-01	6.50E-01	0.137	9.16E-06
		rs2110910	ZDHHC22	14	0.29	1.04	3.79E-01	1.30E-01	-0.109	1.07E-05
		rs10196056	MAP2	2	0.28	1.15	2.95E-03	4.55E-02	-0.105	1.39E-05
		rs10821533	VAV2	9	0.15	0.90	8.19E-02	1.04E-01	0.091	1.48E-05
FG	TG	rs13437130	POU3F2	6	0.06	0.95	6.29E-01	4.59E-01	0.091	1.69E-07
		rs203466	AKAP10	17	0.38	0.91	3.26E-02	1.38E-02	0.130	3.28E-07
		rs203457	AKAP10	17	0.38	0.91	3.11E-02	1.30E-02	0.130	3.44E-07
		rs2108978	AKAP10	17	0.38	0.91	3.20E-02	1.25E-02	0.129	3.65E-07
		rs119672	AKAP10	17	0.38	0.91	3.62E-02	1.54E-02	0.126	7.47E-07
		rs17604987	AKAP10	17	0.26	0.94	2.48E-01	6.71E-02	0.106	1.60E-06
		rs1638527	AKAP10	17	0.26	0.94	2.27E-01	6.47E-02	0.105	1.72E-06
		rs6813	AKAP10	17	0.26	0.94	2.19E-01	5.64E-02	0.105	1.78E-06
		rs11950959	DTWD2	5	0.42	0.94	1.69E-01	6.12E-01	0.119	3.09E-06
		rs12361016	-	11	0.22	0.91	6.75E-02	1.24E-01	0.095	3.42E-06
	BMI	rs10098680	ASAP1	8	0.20	1.05	3.30E-01	4.52E-01	0.114	1.94E-06
		rs16904325	ADCY8	8	0.16	1.02	7.51E-01	9.64E-01	0.108	2.52E-06
		rs12549418	ADCY8	8	0.16	1.02	7.05E-01	9.32E-01	0.108	2.56E-06
		rs16904315	ASAP1	8	0.16	1.03	5.95E-01	9.43E-01	0.107	2.90E-06
		rs2491395	C4BPA	1	0.24	1.04	4.71E-01	6.51E-01	0.117	4.55E-06
		rs9963062	CTIF	18	0.25	0.94	2.11E-01	2.72E-02	-0.117	5.13E-06
		rs16904323	ADCY8	8	0.17	1.03	5.63E-01	9.02E-01	0.105	6.61E-06
		rs12543088	ASAP1	8	0.16	1.01	8.59E-01	8.35E-01	0.102	8.44E-06
		rs12195946	SPACA1	6	0.29	1.07	1.47E-01	2.97E-01	0.115	1.64E-05
		rs8090560	NFATC1	18	0.24	1.06	2.74E-01	8.97E-03	0.105	1.70E-05
	WC	rs2815551	MKX	10	0.41	1.09	3.67E-02	4.05E-02	0.139	3.48E-06
		rs2642295	MKX	10	0.41	1.09	4.24E-02	3.96E-02	0.134	7.54E-06
		rs11780061	ZFAT	8	0.11	0.79	9.28E-04	2.82E-03	-0.098	8.97E-06
		rs1835812	TCF7L1	2	0.46	1.05	2.94E-01	1.93E-01	0.143	1.01E-05
		rs2647389	ST6GALNAC3	1	0.15	1.27	6.16E-05	4.09E-05	0.099	1.39E-05
		rs6984278	XKR4	8	0.12	1.10	1.47E-01	1.41E-01	0.095	1.42E-05
		rs7014798	GSDMC	8	0.30	0.95	3.14E-01	8.10E-01	-0.117	1.50E-05
		rs1900576	SYNPR	3	0.12	1.17	1.97E-02	1.91E-02	0.096	1.52E-05
		rs7824281	XKR4	8	0.12	1.10	1.53E-01	6.87E-02	0.094	1.69E-05
		rs1570857	UBR4	1	0.43	0.94	1.59E-01	3.12E-01	-0.130	2.08E-05

Table 11: **Top Ten SNPs per Model Ranked by Gene-Environment Interaction p-values** * Nearest genes are based on a maximum of 500KB distance
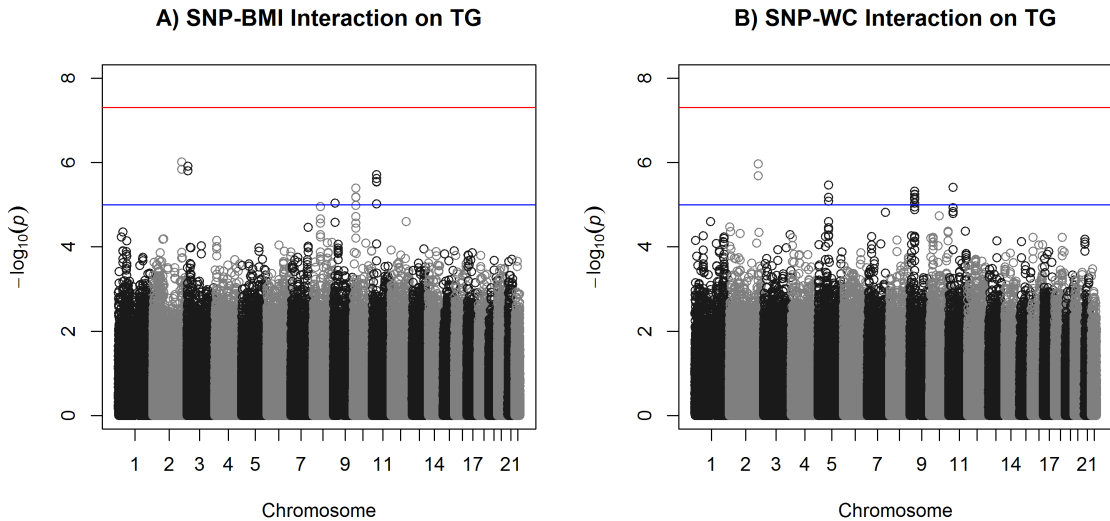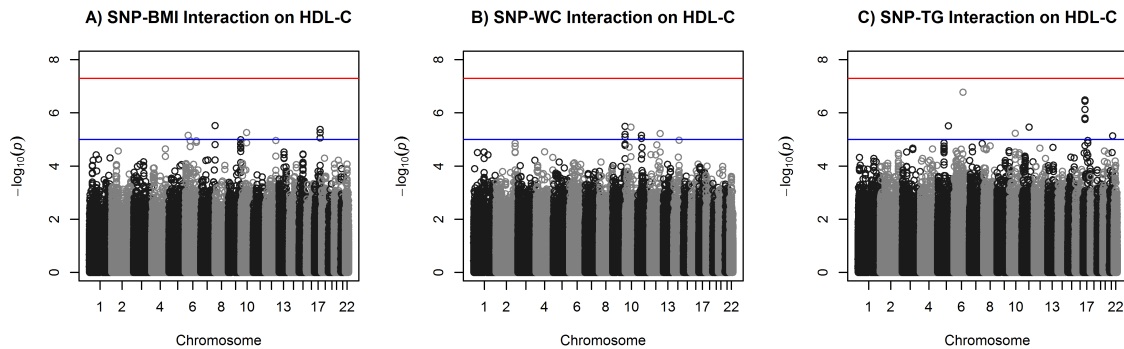
Figure 9: **Manhattan Plots of Gene-Environment Interaction p-values for HDL-C** Genome-wide results showing $-\log10$ of the interaction p-value from linear regression models using HDL-C as the response variable and BMI (A), WC (B) or TG (C) as the interacting covariate. The blue and red horizontal lines represent a nominal p-value threshold at $1E{-}05$ and the genome-wide significance threshold of $5E{-}08$, respectively.

interaction on a log10 scale. An exhaustive search on the entire autosomal SNPs revealed no genome-wide significant interactions (Table 11).
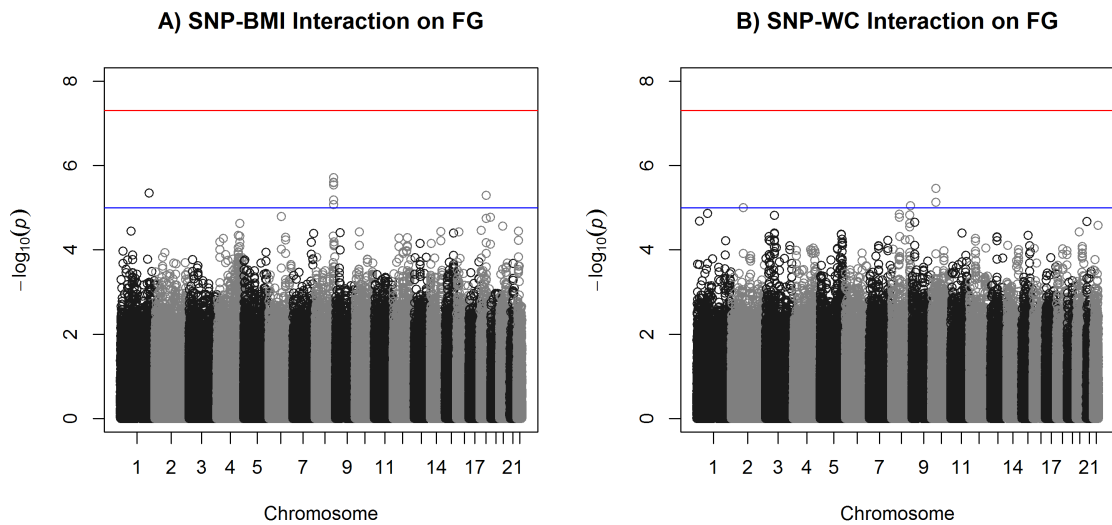
Figure 10: **Manhattan Plots of Gene-Environment Interaction p-values for FG** Genome-wide results showing $-\log10$ of the interaction p-value from linear regression models using FG as the response variable and BMI (A) or WC (B) as the environmental covariate. The blue and red horizontal lines represent a nominal p-value threshold at $1\text{E}-05$ and the genome-wide significance threshold of $5\text{E}-08$, respectively.

Trait	Covariate	SNP	Nearest Gene*	CHR	MAF	η Threshold	Ratio	Variance Het p-value		G-E Interaction	
								LRT	Levene's Test	β_3	Interaction p-value
TG	BMI	rs2882974	CTLA4	2	0.48	0.05	1.10	2.18E-02	1.43E-01	0.172	9.78E-07
		rs10179639	CTLA4	2	0.49	0.05	1.10	2.66E-02	1.42E-01	0.170	1.46E-06
	WC	rs4262060	ITGA1	5	0.32	0.001	1.19	1.5E-04	8.01E-05	0.11	9.98E-05
		rs2882974	CTLA4	2	0.48	0.05	1.10	2.18E-02	1.43E-01	0.170	1.09E-06
		rs4879523	-	9	0.13	0.01	1.18	9.40E-03	1.10E-02	0.105	4.77E-06
HDL-C	BMI	rs712819	VOPP1	7	0.09	0.001 & 0.01	1.29	2.92E-04	7.43E-04	-0.07951	6.12E-05
	WC	rs1684901	CCDC6	10	0.46	0.001	0.87	1.61E-03	2.59E-03	0.140	3.46E-06
	TG	rs1684901	CCDC6	10	0.46	0.01	0.87	1.61E-03	2.59E-03	0.125	5.87E-06
		rs203466	AKAP10	17	0.38	0.05	0.91	3.26E-02	1.38E-02	0.130	3.28E-07
		rs203457	AKAP10	17	0.38	0.05	0.91	3.11E-02	1.30E-02	0.130	3.44E-07
		rs2108978	AKAP10	17	0.38	0.05	0.91	3.20E-02	1.25E-02	0.129	3.65E-07
		rs119672	AKAP10	17	0.38	0.05	0.91	3.62E-02	1.54E-02	0.126	7.47E-07
FG	BMI	rs2647389	ST6GALNAC3	1	0.15	0.001	1.27	6.16E-05	4.09E-05	0.093	3.65E-05
		rs1780061	ZFAT	8	0.11	0.001	0.79	9.28E-04	2.82E-03	-0.088	7.42E-05
	WC	rs2647389	ST6GALNAC3	1	0.15	0.001	1.27	6.16E-05	4.09E-05	0.099	1.39E-05
		rs1780061	ZFAT	8	0.11	0.001	0.79	9.28E-04	2.82E-03	-0.098	8.97E-06

Table 12: **SNPs per Model with Gene–Environment Interaction p-values Passing Bonferroni Correction Using Variance Prioritization** * Nearest genes are based on a maximum of 500KB distance

4.1.3   Variance Prioritization Result

The interaction p-values of SNPs with all possible interacting covariates as well as the variance heterogeneity p-values (from LRT and Levene's test) of SNPs were cross examined according to the response variable for each linear regression model. Although none of the interactions were genome-wide significant, there is a clear and consistent enrichment in the lowest gene-environment interaction p-values among SNPs with low variance heterogeneity p-values (Table 11). In addition, see Figures 11 to 13 for a summary of number of SNPs prioritized for subsequent interaction testing at different VP thresholds using Levene's test and LRT. It is of note that half of the SNPs interacting with BMI on FG would have been prioritized for interactions at a VP p-value threshold of 0.05 using either LRT or Levene's test. For TG and HDL-C, SNPs that demonstrated the lowest interaction p-value with either BMI or WC also had LRT p-values less than a nominal level of 0.05. Furthermore, some interactions turned out significant after Bonferroni correction when accounting for only the prioritized SNPs at various VP p-value threshold (Table 12). These results suggest that VP using LRT or Levene's test selects potentially interacting SNPs for gene-environment interactions yet larger sample sizes are required to positively identify these interactions.

4.2   Enrichment Analysis

In this section, I will address the second methodological question of whether VP provides an effective filter to select a subset of interacting SNPs. To evaluate the contribution of variance heterogeneity to the enrichment of low interaction p-values genome-wide, all autosomal SNPs were divided into the prioritized and the non-prioritized sets based on LRT VP p-value thresholds of 0.001, 0.01, 0.05, and 0.1.The corresponding Bonferroni corrected thresholds for claiming a significant interaction were determined for each VP threshold according to the number of SNPs prioritized (lower than the respective LRT VP p-value thresholds).

Across the four choices of VP p-value thresholds, there existed at least one prioritized SNP with interaction p-value less than its corresponding Bonferroni correction threshold for all linear regression models tested (Figures 14 to 16). The quantile-quantile plots also suggest that the collections of prioritized SNPs have generally lower interaction p-values or larger interaction effects than expected when compared to the non-prioritized SNPs. I also observed distinct prioritization patterns in the seven linear models: 1) For the same quantitative trait, there were clearer cases of enrichment depending on the strength of correlation between the trait and interacting covariates as the theory suggested; 2) for the same interacting covariate, the choice of prioritization p-value threshold heavily influenced the enrichment signal as marked by the level of deviation from the null distribution in the quantile-quantile plots (such as in Figure 16-A, E, I and M). It was evident that as the VP p-value threshold increased, the deviation from the null became more moderate. These results also confirm empirically that effective prioritization depends on the choice of prioritization p-value threshold for the specific interaction model under consideration. In addition,

this excess of lower interaction p-values in the prioritized set is robustly related to the strength of correlation between the trait and interacting covariates as well as the choice of the prioritization threshold η.

Enrichment was further quantified for low interaction p-values in the set of prioritized SNPs by counting the number of concordant SNPs in terms of direction of effects between log-transformed ratio and interaction beta coefficient. Concordance in direction of effects are based on the theoretical model presented in Table 1, where the beta coefficients of the covariate and interaction terms together determine the trend in variance to be monotonically increasing (ratio greater than 1) or decreasing (ratio less than 1). For example, when the interacting covariate and trait are negatively correlated, increasing variance or an estimated ratio of greater than 1 indicates a negative interaction effect and vice versa. For each LRT VP p-value threshold (0.001, 0.01, and 0.05), the direction of effect for interaction beta coefficient and log-transformed ratio were highly concordant in the prioritized sets (Table 13). For the same quantitative trait, the stronger the correlation between the trait and the interacting covariate, the further away the interaction p-values of the prioritized SNPs deviate from the null distribution. In addition, the percentage of concordant SNPs was also inversely related to the prioritization threshold selected, i.e. lower VP thresholds provided higher concordant rates in the original linear models examined. For instance, the percentage of concordant SNPs for HDL-C increased 1.5% when η dropped from 0.01 to 0.001 and increased 5.4% when η dropped from 0.05 to 0.01.

$\eta = 0.001$ $\eta = 0.01$ $\eta = 0.05$

Figure 11: **Venn Diagrams Showing the Number of SNPs Prioritized using Levene's test and LRT (TG)** VP prioritization thresholds of $\eta = 0.001, 0.01, 0.05$ were considered. The left circle represents the SNPs with Levene's test p-value lower than the VP threshold of η. The right circle represents the SNPs with LRT p-value lower than the VP threshold of η. SNPs with both LRT and Levene's test p-values lower than the VP threshold of η correspond to number in the area where the two circles overlap.
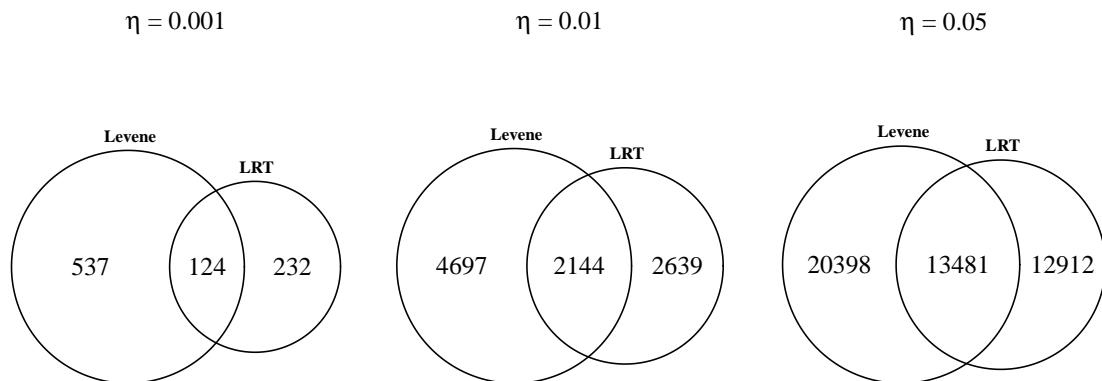
$\eta = 0.001$ $\eta = 0.01$ $\eta = 0.05$

Figure 12: **Venn Diagrams Showing the Number of SNPs Prioritized using Levene's test and LRT (HDL-C)** VP prioritization thresholds of $\eta = 0.001, 0.01, 0.05$ were considered. The left circle represents the SNPs with Levene's test p-value lower than the VP threshold of η. The right circle represents the SNPs with LRT p-value lower than the VP threshold of η. SNPs with both LRT and Levene's test p-values lower than the VP threshold of η correspond to number in the area where the two circles overlap.

$\eta = 0.001$ $\eta = 0.01$ $\eta = 0.05$
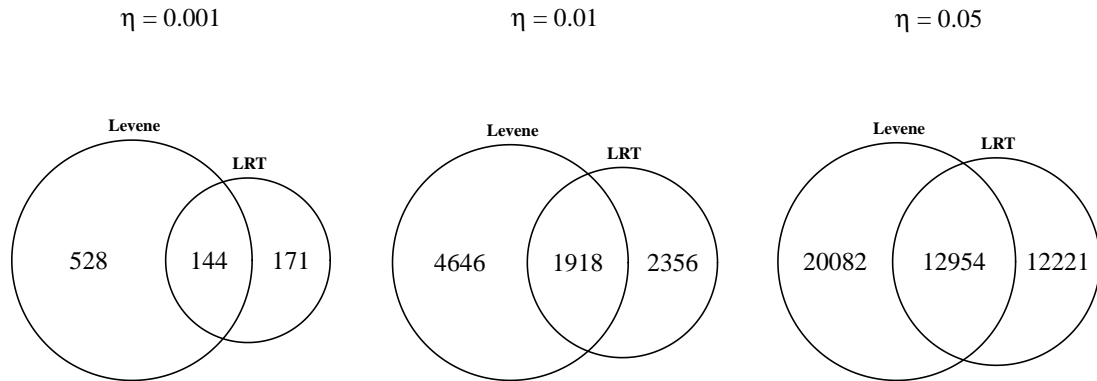

Figure 13: **Venn Diagrams Showing the Number of SNPs Prioritized using Levene's test and LRT (FG)** VP prioritization thresholds of $\eta = 0.001, 0.01, 0.05$ were considered. The left circle represents the SNPs with Levene's test p-value lower than the VP threshold of η. The right circle represents the SNPs with LRT p-value lower than the VP threshold of η. SNPs with both LRT and Levene's test p-values lower than the VP threshold of η correspond to number in the area where the two circles overlap.
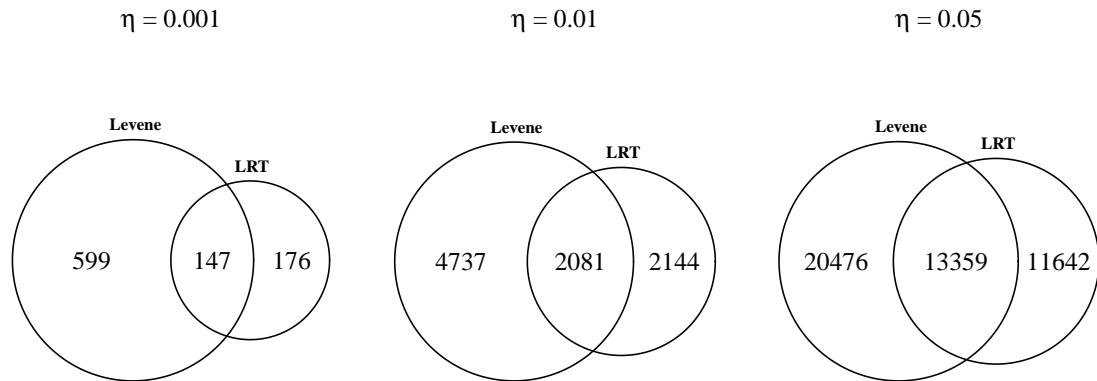
Figure 14: **Quantile-Quantile Plots of Gene-Environment Interaction *p*-values for SNPs Prioritized with LRT (TG)** Illustrated in the first column are the quantile-quantile plots of gene-BMI interaction *p*-values when SNPs were prioritized at LRT *p*-value thresholds of 0.001 (A), 0.01 (E), 0.05 (I) and 0.1 (M). In contrast, the second column illustrates the quantile-quantile plots of gene-BMI interaction *p*-values when SNPs were not prioritized at LRT *p*-value thresholds of 0.001 (B), 0.01 (F), 0.05 (J) and 0.1 (N). Similarly illustrated in the third column are the quantile-quantile plots of gene-WC interaction *p*-values when SNPs were prioritized at LRT *p*-value thresholds of 0.001 (C), 0.01 (G), 0.05 (K) and 0.1 (O). Finally, the fourth column illustrates the quantile-quantile plots of gene-WC interaction *p*-values when SNPs were not prioritized at LRT *p*-value thresholds of 0.001 (D), 0.01 (H), 0.05 (L) and 0.1 (P). The horizontal line represents the Bonferroni correction *p*-value threshold accounting for the number of SNPs in each scenario.
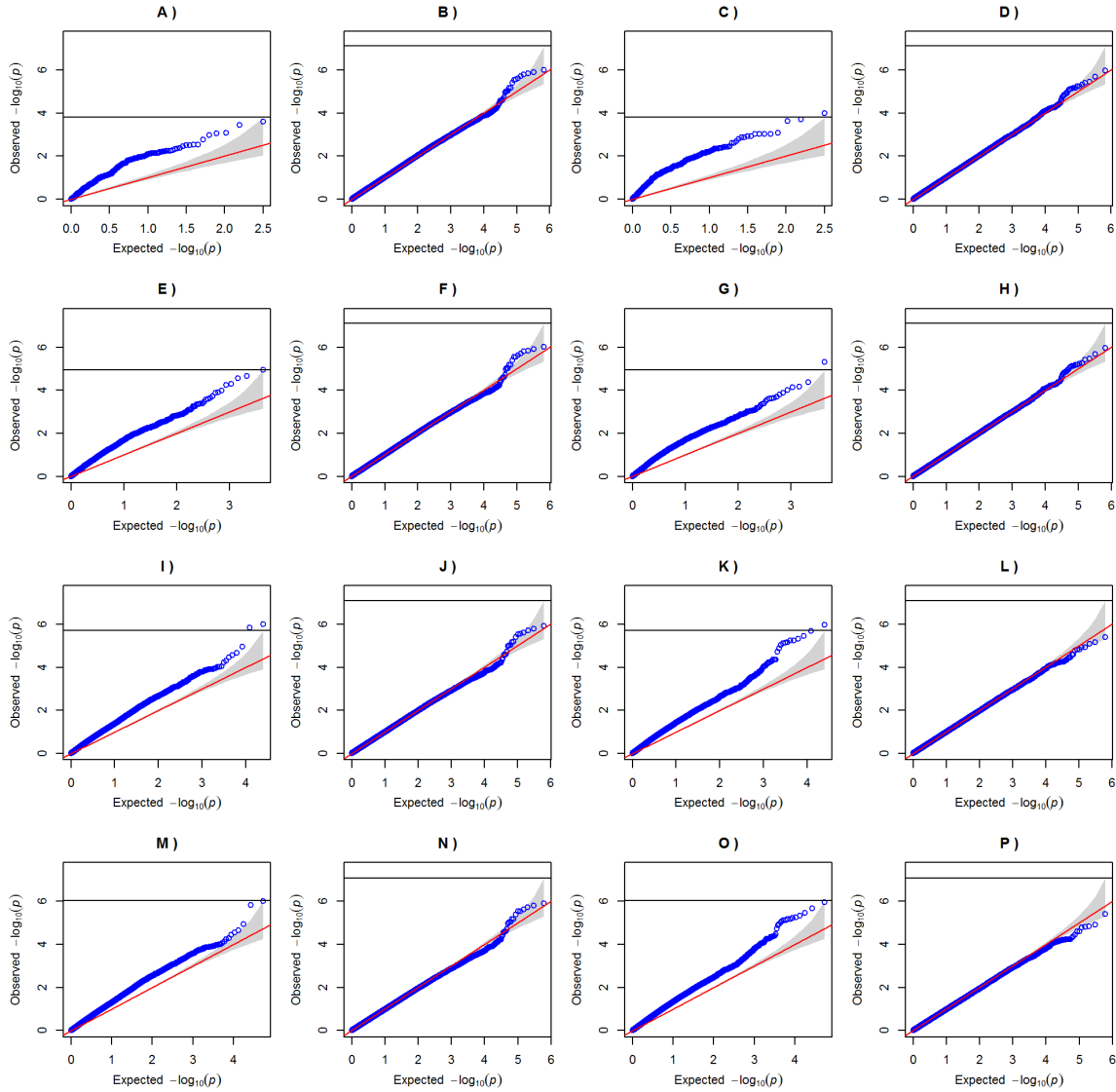
Figure 15: **Quantile-Quantile Plots of Gene-Environment Interaction *p*-values for SNPs Prioritized with LRT (HDL-C)** Illustrated in the first column are the quantile-quantile plots of gene-BMI interaction *p*-values when SNPs were prioritized at LRT *p*-value thresholds of 0.001 (A), 0.01 (G), 0.05 (M) and 0.1 (S). In contrast, the second column illustrates the quantile-quantile plots of gene-BMI interaction *p*-values when SNPs were not prioritized at LRT *p*-value thresholds of 0.001 (B), 0.01 (H), 0.05 (N) and 0.1 (T). Similarly illustrated in the third column are the quantile-quantile plots of gene-WC interaction *p*-values when SNPs were prioritized at LRT *p*-value thresholds of 0.001 (C), 0.01 (I), 0.05 (O) and 0.1 (U). The fourth column illustrates the quantile-quantile plots of gene-WC interaction *p*-values when SNPs were not prioritized at LRT *p*-value thresholds of 0.001 (D), 0.01 (J), 0.05 (P) and 0.1 (V). The fifth column are the quantile-quantile plots of gene-triglyceride interaction *p*-values when SNPs were prioritized at LRT *p*-value thresholds of 0.001 (E), 0.01 (K), 0.05 (Q) and 0.1 (W). The sixth column shows the quantile-quantile plots of gene-triglyceride interaction *p*-values when SNPs were not prioritized at LRT *p*-value thresholds of 0.001 (F), 0.01 (L), 0.05 (R) and 0.1 (X). The horizontal line represents the Bonferroni correction *p*-value threshold accounting for the number of SNPs in each scenario.

64

Figure 16: **Quantile-Quantile Plots of Gene-Environment Interaction p-values for SNPs Prioritized with LRT (FG)** Illustrated in the first column are the quantile-quantile plots of gene-BMI interaction p-values when SNPs were prioritized at LRT p-value thresholds of 0.001 (A), 0.01 (E), 0.05 (I) and 0.1 (M). In contrast, the second column illustrates the quantile-quantile plots of gene-BMI interaction p-values when SNPs were not prioritized at LRT p-value thresholds of 0.001 (B), 0.01 (F), 0.05 (J) and 0.1 (N). Similarly illustrated in the third column are the quantile-quantile plots of gene-WC interaction p-values when SNPs were prioritized at LRT p-value thresholds of 0.001 (C), 0.01 (G), 0.05 (K) and 0.1 (O). Finally, the fourth column illustrates the quantile-quantile plots of gene-WC interaction p-values when SNPs were not prioritized at LRT p-value thresholds of 0.001 (D), 0.01 (H), 0.05 (L) and 0.1 (P). The horizontal line represents the Bonferroni correction p-value threshold accounting for the number of SNPs in each scenario.
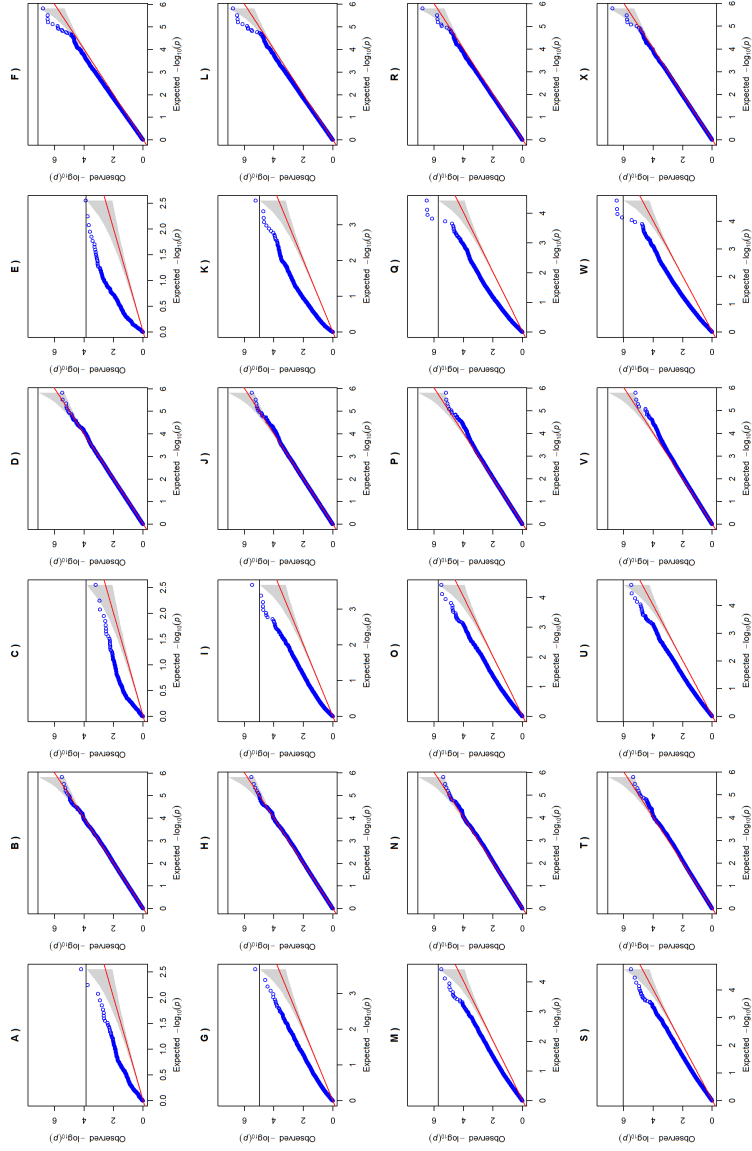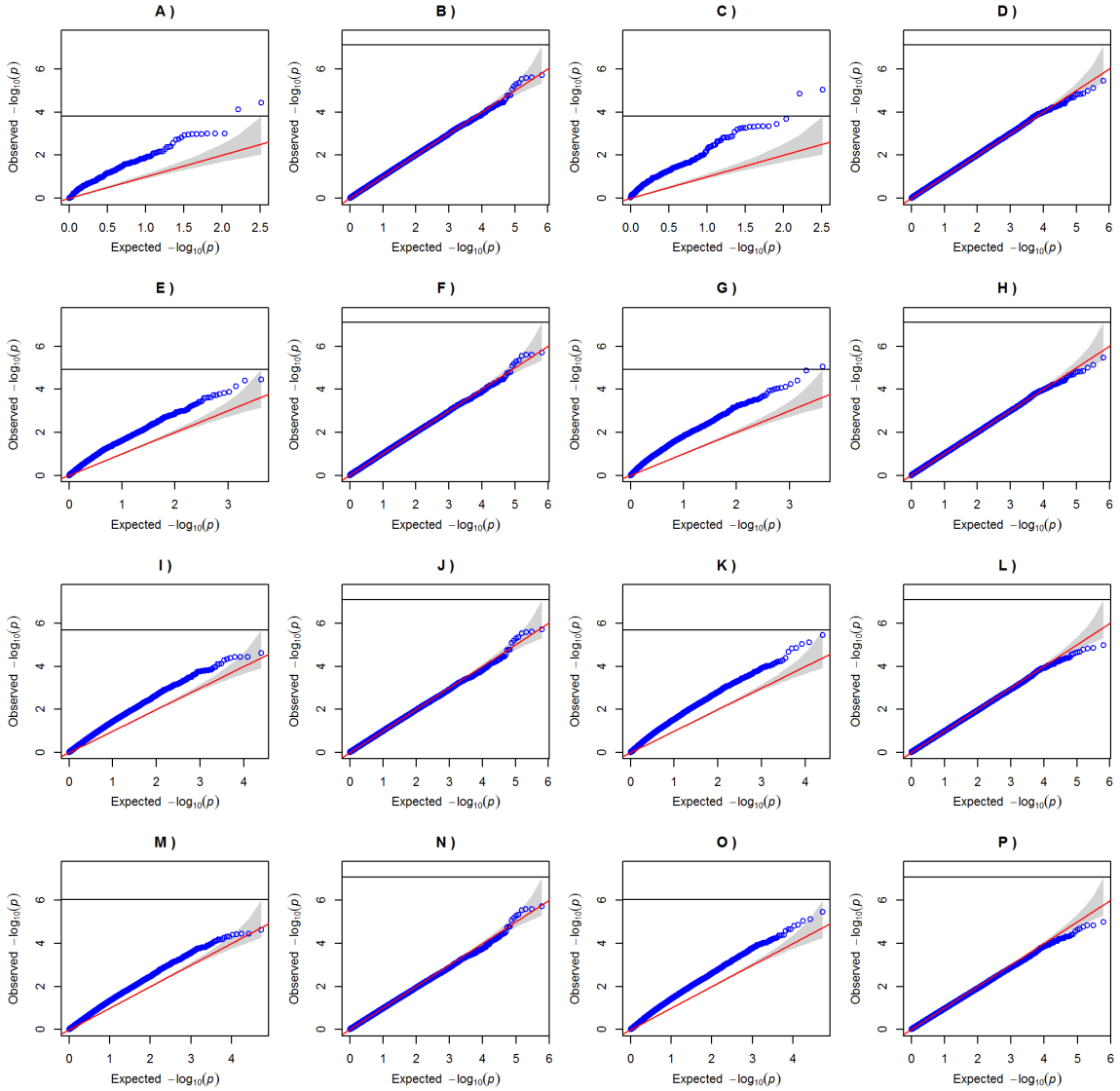
Trait	Covariate	Pearson's Correlation	LRT $\eta = 0.001$ Concordant (Perc.)	Discordant	LRT $\eta = 0.01$ Concordant (Perc.)	Discordant	LRT $\eta = 0.05$ Concordant (Perc.)	Discordant
TG	BMI	0.319	293 (93.0%)	22	3634 (85.0%)	640	20223 (80.3%)	4952
	WC	0.323	299 (94.9%)	16	3685 (86.2%)	589	20584 (81.8%)	4591
HDL-C	BMI	-0.316	326 (91.6%)	30	4193 (87.7%)	590	21517 (81.5%)	4876
	WC	-0.369	343 (96.3%)	13	4365 (91.3%)	418	22502 (85.5%)	3891
	TG	-0.439	340 (95.5%)	16	4497 (94.0%)	286	23397 (88.6%)	2996
FG	BMI	0.293	296 (91.6%)	27	3657 (86.6%)	568	20296 (81.2%)	4705
	WC	0.348	306 (94.7%)	17	3837 (90.8%)	388	21326 (85.3%)	3675

Table 13: **Concordance in the Direction of Effects** Concordance in direction of effects are based on the theoretical prediction presented in Table 1, where the beta coefficients of the covariate and interaction terms determine the trend in variance to be increasing (a ratio greater than 1) or decreasing (a ratio less than 1). For the prioritized subsets at VP thresholds (η) of 0.001, 0.01 and 0.05, the total numbers of concordant SNPs and discordant SNPs were recorded for each linear regression model.

Trait	Covariate	Pearson's Correlation	LRT η = 0.001 Concordant (Perc.)	Discordant	LRT η = 0.01 Concordant (Perc.)	Discordant	LRT η = 0.05 Concordant (Perc.)	Discordant
Residualized TG	BMI	-0.009	352 (66.9%)	174	2033 (36.3%)	3574	12358 (40.6%)	18079
	WC	-0.011	326 (64.3%)	181	2145 (39.5%)	3284	12297 (41.5%)	17343
Residualized HDL-C	BMI	0.0002	145 (39.5%)	222	2088 (44.8%)	2575	12489 (46.1%)	14615
	WC	0.001	183 (42.8%)	244	2356 (43.0%)	3124	12874 (43.7%)	16554
	TG	-0.008	155 (50.5%)	152	2181 (49.3%)	2244	12685 (49.5%)	12966
Residualzied FG	BMI	0.006	183 (47.3%)	204	2047 (43.8%)	2629	12066 (45.4%)	14512
	WC	0.006	265 (62.6%)	158	2078 (41.9%)	2884	11722 (43.0%)	15551

Table 14: **Concordance in the Direction of Effects for Residualized Traits** Concordance in direction of effects are based on the theoretical prediction presented in Table 1, where the beta coefficients of the covariate and interaction terms determine the trend in variance to be increasing (a ratio greater than 1) or decreasing (a ratio less than 1). For the prioritized subsets at VP thresholds (η) of 0.001, 0.01 and 0.05, the total numbers of concordant SNPs and discordant SNPs were recorded for each linear regression model.

4.3 Sensitivity Analysis of Variance Prioritization

The enrichment signals were indicated by the deviation of interaction p-values in the prioritized subsets from the null distribution, which could be perceived as inflation and therefore question the validity of our method in two possible ways. These concerns have been addressed previously (Paré *et al.*, 2010) in theory using mathematical proofs. First, quantitative trait variance per genotype and interaction beta coefficients are uncorrelated under the null hypothesis of no interaction. Second, when variance heterogeneity is invoked by factors other than interactions, correct type I error rate will be maintained. To empirically exclude the possibility that the enrichment could reflect inflation of type I error in disguise for any of the reasons above, I repeated the analyses using covariate-adjusted traits. In the context of VP, adjusting the trait of interest for the interacting covariate removes any heterogeneity of variance caused by the interaction. Therefore, heteroscedasticity is eliminated as a possible source of inflation owing to the absence of type I error inflation in interaction testing of SNPs based solely on the heterogeneity of variance of adjusted traits.

For each of the seven linear regression models, the quantitative trait was adjusted for its respective interacting covariate, and similar analyses were repeated. On one hand, the variance heterogeneity p-values distribution of the SNPs changed after adjustment of interacting covariate was applied as suggested by the scatter of p-values on the log10 scale (Figures 17 to 19). In particular, a small proportion of SNPs had reduced variance heterogeneity p-values; some showed increased p-values, while the majority maintained similar ranks. On the other hand, the interaction p-values remained the same when adjustment was applied (Figure 20). Statistically, the proportion of variance explained by the covariate would be accounted for in either the

full model with covariate in the regression model or the adjusted model so that the interaction term beta coefficient was indifferent to the inclusion of the covariate main effect. This observation suggests that a biologically relevant correlation between the trait and interacting covariate is unimportant for statistical interaction testing.

Adjusting the trait of interest for the interacting covariate did not change the interaction p-values, however, it did influence the variance heterogeneity p-values. Since interaction testing does not depend on the correlation between the trait and interacting covariate but prioritization of SNPs using LRT does, while the set of SNPs prioritized was not different from any random sets in terms of enrichment for low interaction p-values. For the adjusted analyses, I did not observe any inflation of type I error of interaction p-values in the prioritized and non-prioritized sets of SNPs at various prioritization p-value thresholds (Figures 21 to 23). More importantly, there was very little concordance between log-transformed ratio and interaction beta coefficient in the adjusted models (Table 14). These results suggest that selecting SNPs solely on the basis of variance heterogeneity did not lead to inflated type I error rate, and thus the enrichment in the original analysis was indeed encouraging.
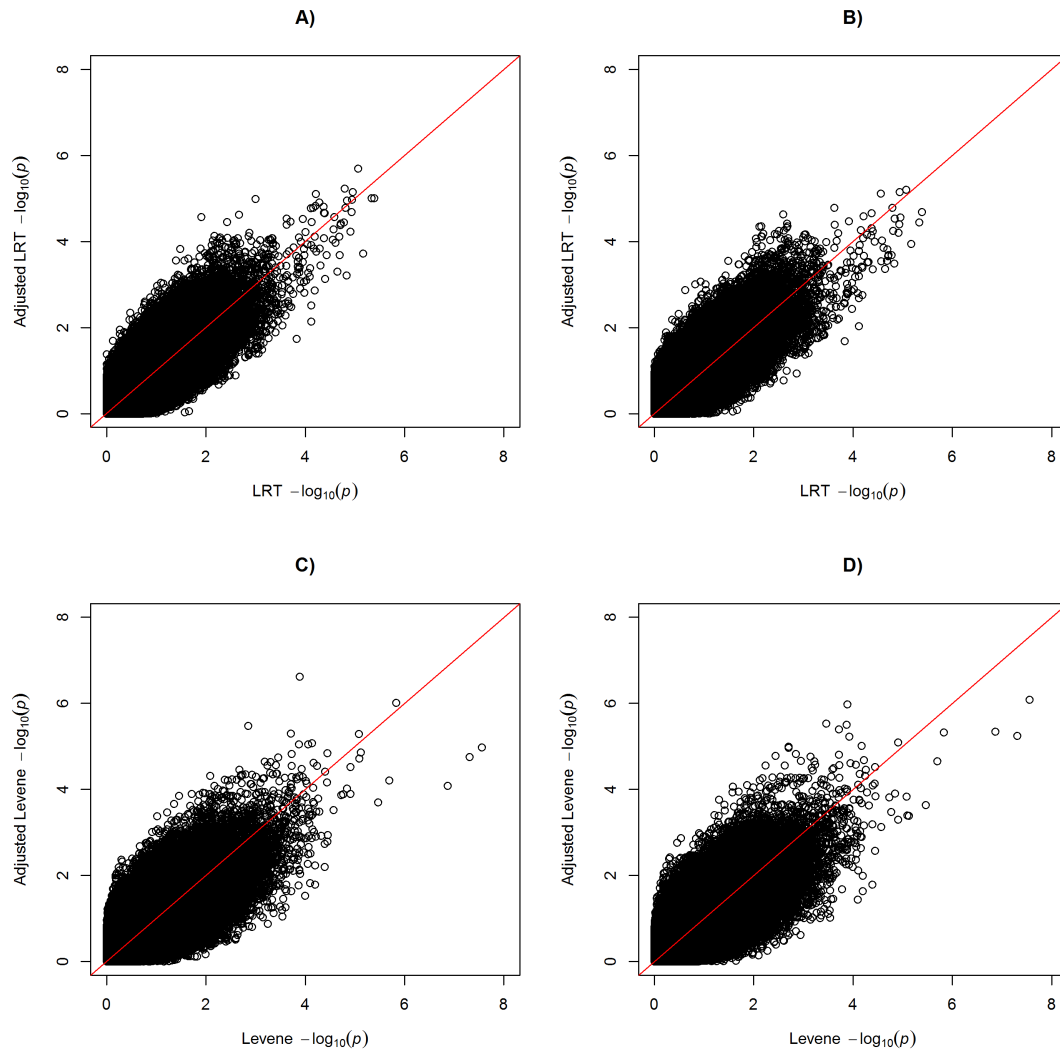
Figure 17: **Distribution of Variance Heterogeneity Test *p*-values for TG before and after Adjusting for Interacting Covariate** These scatterplots show the variance heterogeneity tests *p*-values for TG before and after adjusted for BMI (A, C) and WC (B, D). The top row shows the changes in LRT *p*-value distribution while the bottom row shows the changes in Levene's test *p*-values.
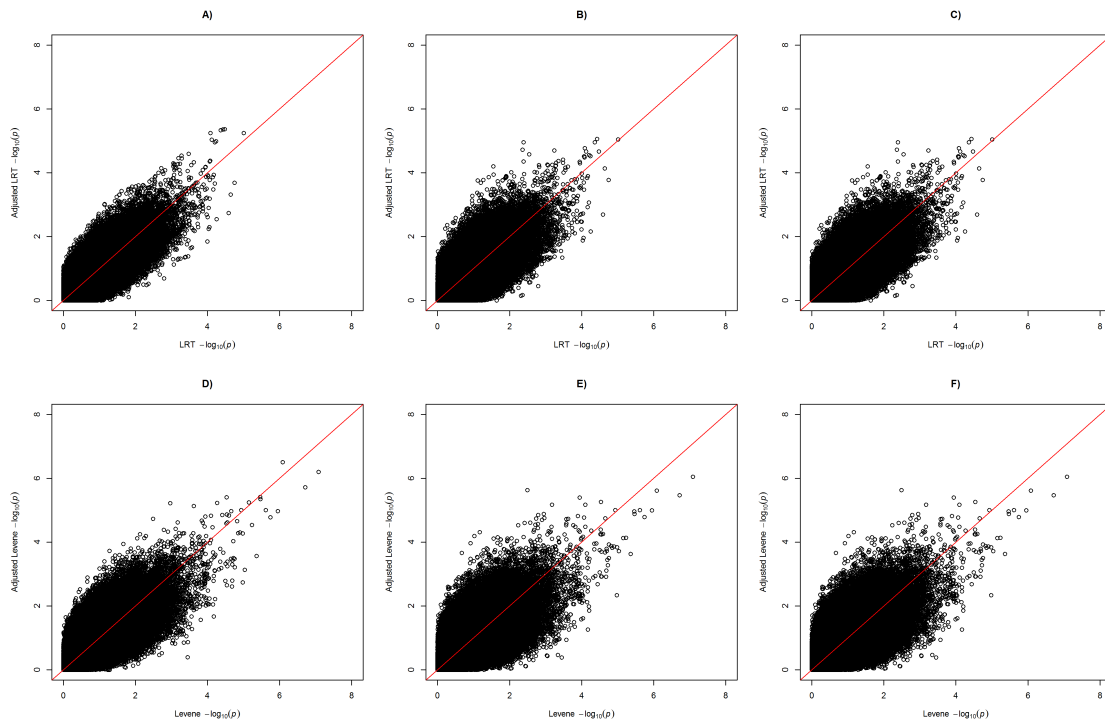
Figure 18: **Distribution of Variance Heterogeneity Test p-values for HDL-C before and after Adjusting for Interacting Covariate** These scatterplots show the variance heterogeneity tests p-values for HDL-C before and after adjusted for BMI (A, D), WC (B, E), and TG (C, F). The top row shows the changes in LRT p-value distribution while the bottom row shows the changes in Levene's test p-values.
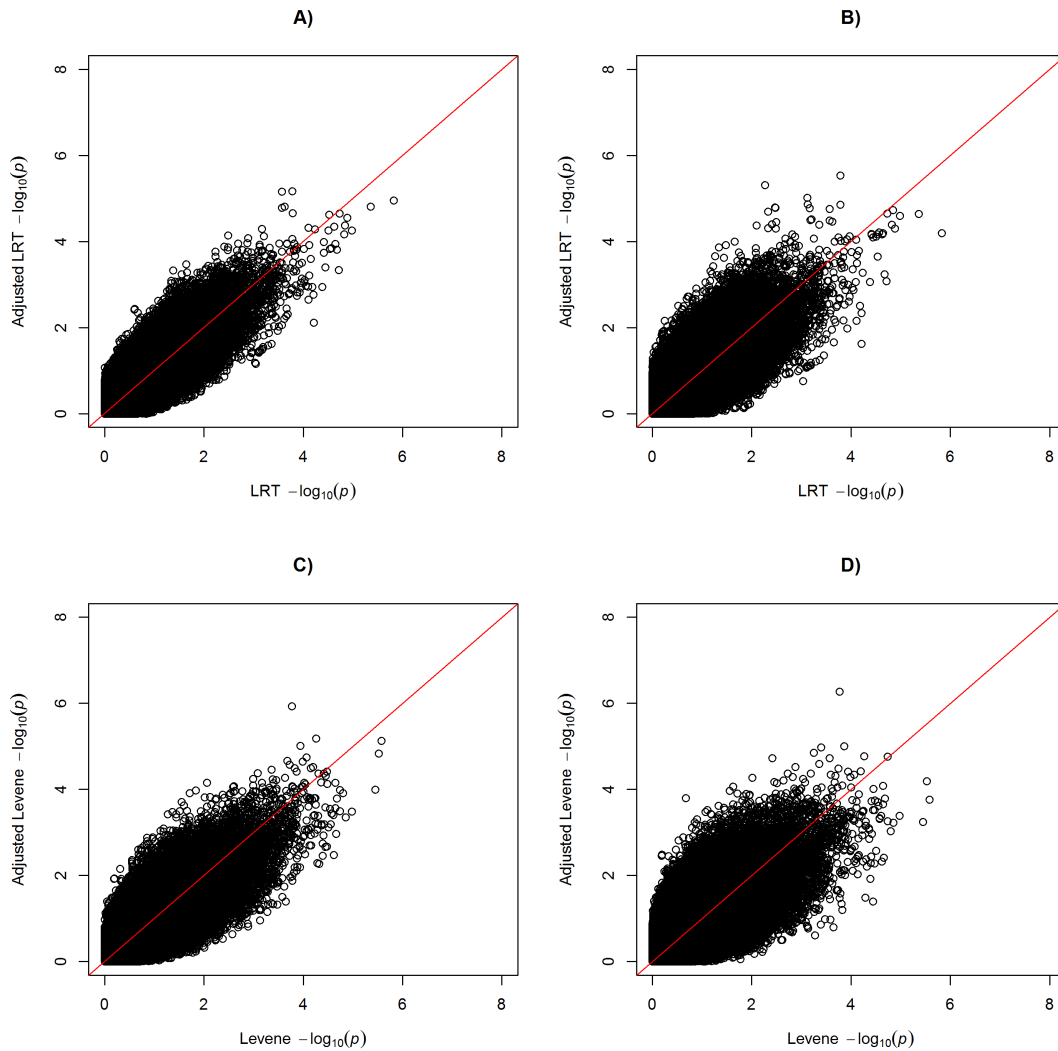
Figure 19: **Distribution of Variance Heterogeneity Test *p*-values for FG before and after Adjusting for Interacting Covariate** These scatterplots show the variance heterogeneity tests *p*-values for FG before and after adjusted for BMI (A, C) and WC (B, D). The top row shows the changes in LRT *p*-value distribution while the bottom row shows the changes in Levene's test *p*-values.
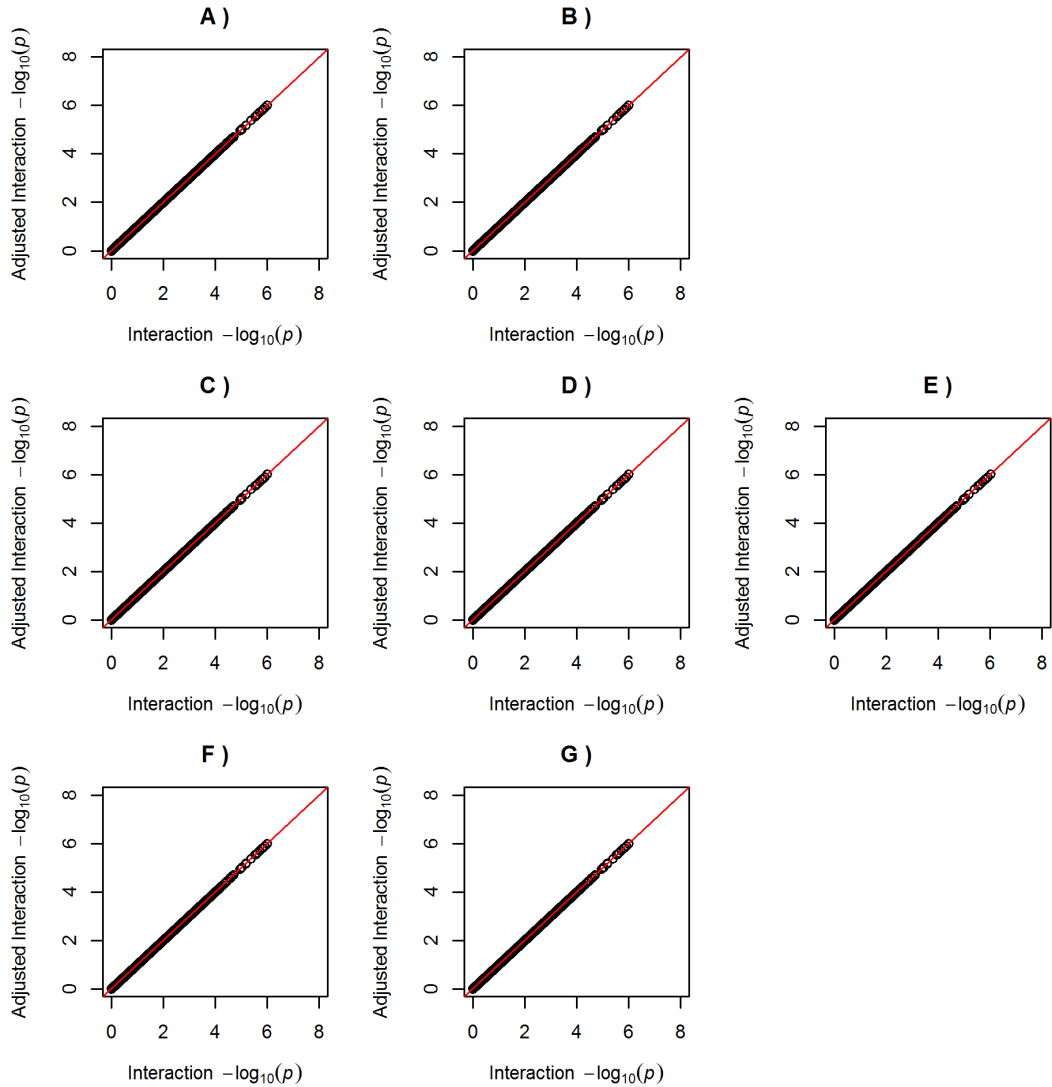
Figure 20: **Gene-Environment Interaction p-value Before and After Adjusting for Interacting Covariate** The top row shows the gene-environment interaction p-values for TG before and after adjusted for BMI (A) and WC (B). The second row shows the gene-environment interaction p-values for HDL-C before and after adjusted for BMI (C), WC (D), and TG (E). The bottom row shows the gene-environment interaction p-values for FG before and after adjusted for BMI (F) and WC (G).

Figure 21: **Quantile-Quantile Plots of Gene-Environment Interaction p-values of SNPs Prioritized Using LRT on TG Adjusted for Interacting Covariates** Illustrated in the first column are the quantile-quantile plots of gene-BMI interaction p-values for BMI adjusted TG when SNPs were prioritized at LRT p-value thresholds of 0.001 (A), 0.01 (E), 0.05 (I) and 0.1 (M). In contrast, the second column illustrates the quantile-quantile plots of gene-BMI interaction p-values when SNPs were not prioritized at LRT p-value thresholds of 0.001 (B), 0.01 (F), 0.05 (J) and 0.1 (N). Similarly illustrated in the third column are the quantile-quantile plots of gene-WC interaction p-values for WC adjusted TG when SNPs were prioritized at LRT p-value thresholds of 0.001 (C), 0.01 (G), 0.05 (K) and 0.1 (O). Finally, the fourth column illustrates the quantile-quantile plots of gene-WC interaction p-values when SNPs were not prioritized at LRT p-value thresholds of 0.001 (D), 0.01 (H), 0.05 (L) and 0.1 (P). The horizontal line represents the Bonferroni correction p-value threshold accounting for the number of SNPs in each scenario.

74
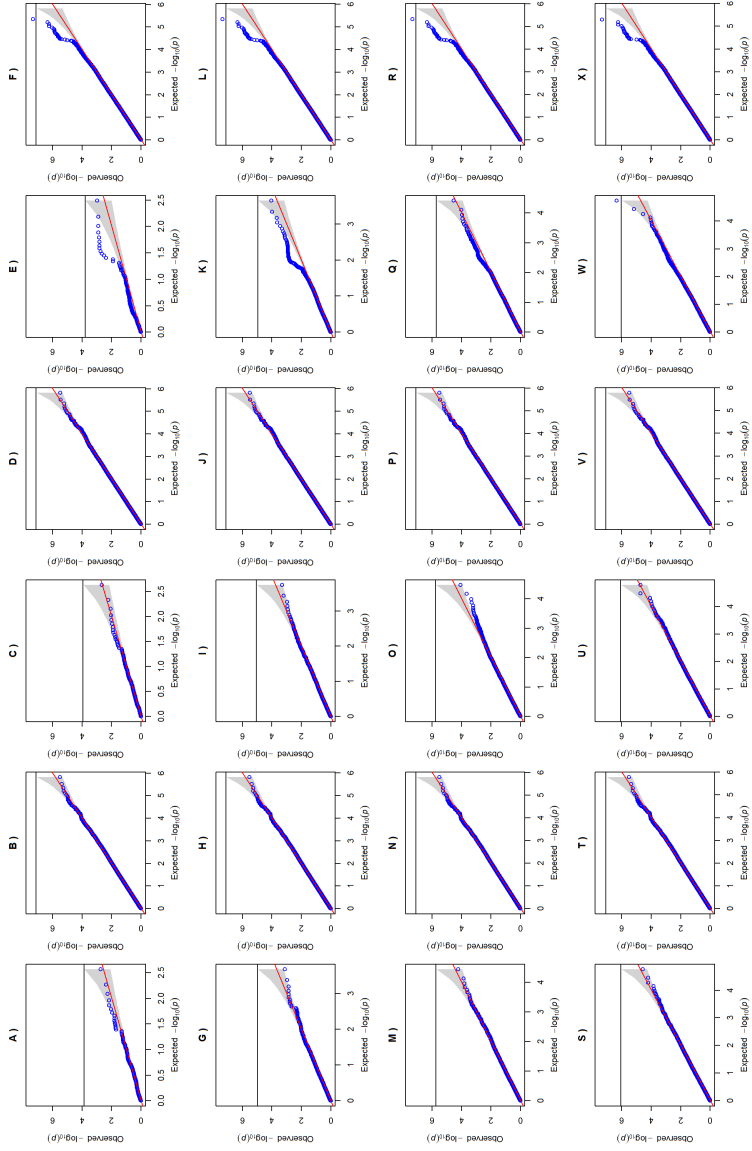
Figure 22: **Quantile-Quantile Plots of Gene-Environment Interaction *p*-values of SNPs Prioritized Using LRT on HDL-C Adjusted for Interacting Covariates** Illustrated in the first column are the quantile-quantile plots of gene-BMI interaction *p*-values for BMI adjusted HDL-C when SNPs were prioritized at LRT *p*-value thresholds of 0.001 (A), 0.01 (G), 0.05 (M) and 0.1 (S). In contrast, the second column illustrates the quantile-quantile plots of gene-BMI interaction *p*-values when SNPs were not prioritized at LRT *p*-value thresholds of 0.001 (B), 0.01 (H), 0.05 (N) and 0.1 (T). Similarly illustrated in the third column are the quantile-quantile plots of gene-WC interaction *p*-values for WC adjusted HDL-C when SNPs were prioritized at LRT *p*-value thresholds of 0.001 (C), 0.01 (I), 0.05 (O) and 0.1 (U). The fourth column illustrates the quantile-quantile plots of gene-WC interaction *p*-values when SNPs were not prioritized at LRT *p*-value thresholds of 0.001 (D), 0.01 (J), 0.05 (P) and 0.1 (V). The fifth column are the quantile-quantile plots of gene-triglyceride interaction *p*-values for triglyceride adjusted HDL-C when SNPs were prioritized at LRT *p*-value thresholds of 0.001 (E), 0.01 (K), 0.05 (Q) and 0.1 (W). The sixth column shows the quantile-quantile plots of gene-triglyceride interaction *p*-values when SNPs were not prioritized at LRT *p*-value thresholds of 0.001 (F), 0.01 (L), 0.05 (R) and 0.1 (X). The horizontal line represents the Bonferroni correction *p*-value threshold accounting for the number of SNPs in each scenario.
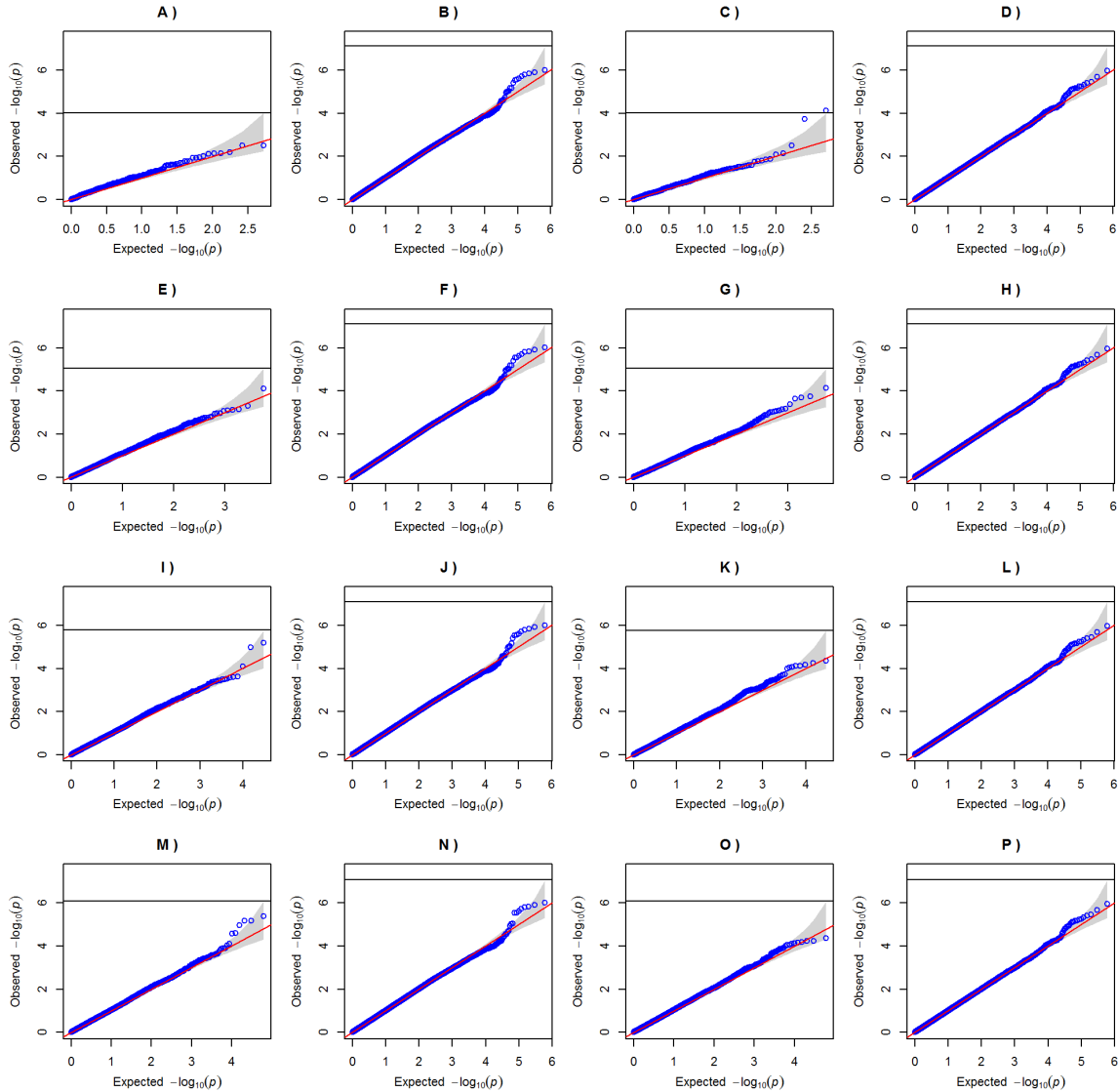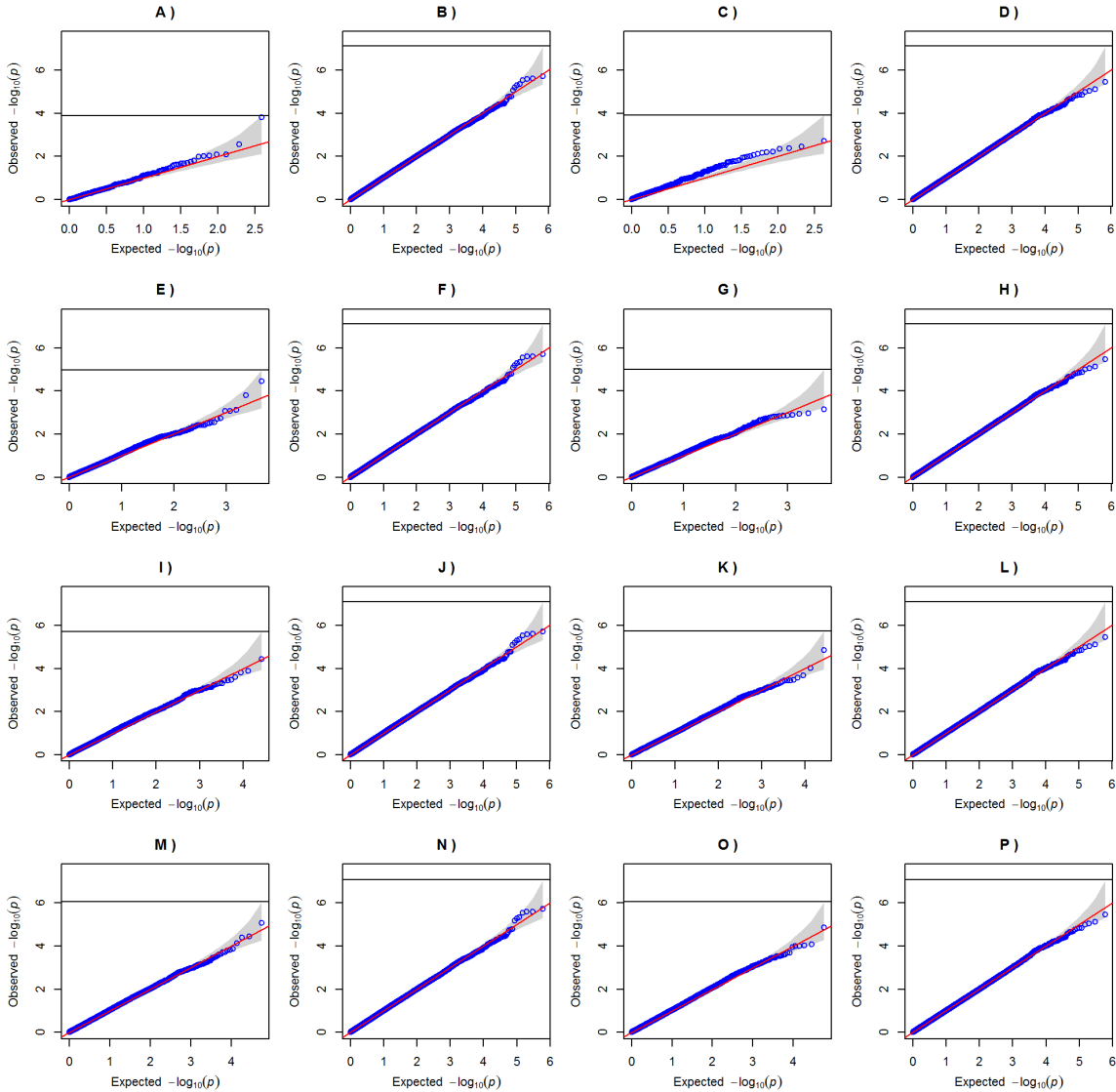
Figure 23: **Quantile-Quantile Plots of Gene-Environment Interaction *p*-values of SNPs Prioritized Using LRT on FG Adjusted for Interacting Covariates** Illustrated in the first column are the quantile-quantile plots of gene-BMI interaction *p*-values for BMI adjusted FG when SNPs were prioritized at LRT *p*-value thresholds of 0.001 (A), 0.01 (E), 0.05 (I) and 0.1 (M). In contrast, the second column illustrates the quantile-quantile plots of gene-BMI interaction *p*-values when SNPs were not prioritized at LRT *p*-value thresholds of 0.001 (B), 0.01 (F), 0.05 (J) and 0.1 (N). Similarly illustrated in the third column are the quantile-quantile plots of gene-WC interaction *p*-values for WC adjusted FG when SNPs were prioritized at LRT *p*-value thresholds of 0.001 (C), 0.01 (G), 0.05 (K) and 0.1 (O). Finally, the fourth column illustrates the quantile-quantile plots of gene-WC interaction *p*-values when SNPs were not prioritized at LRT *p*-value thresholds of 0.001 (D), 0.01 (H), 0.05 (L) and 0.1 (P). The horizontal line represents the Bonferroni correction *p*-value threshold accounting for the number of SNPs in each scenario.

Chapter 5

Concluding Remarks

In this thesis, I proposed a statistical test to detect trends in variances of subgroups using a maximum likelihood approach, and illustrated its application in genetics to optimize the Variance Prioritization (VP) p-value threshold for selection of potentially interacting SNPs. The most commonly used statistical framework to test interactions employs a linear regression model. Under plausible conditions, the quantitative trait variance conditional on genotype either monotonically increases or decreases with the number of minor alleles when interactions are present. Based on the ratio parameterization, I introduce a LRT for variance heterogeneity. The proposed LRT requires only the set of quantitative trait variances per genotype and genotype counts to determine the variance inequality p-value for a given SNP.

The closed-form representation requires only the quantitative variance conditional on the three genotypes and observed genotype counts to accurately and quickly compute the variance inequality p-values for genome-wide dataset. The use of an ordered alternative hypothesis as compared to that of a general alternative leads to an improvement in statistical power. Indeed, the simulation studies suggest that

LRT outperforms the original Levene's test, Bartlett's test, Levene type trend tests and the conventional exhaustive search across a variety of interaction scenarios and demonstrates computational feasibility.

Inflated type I error rates usually arise when the normality assumption is not satisfied. Although LRT was less robust to non-normally distributed quantitative trait, as long as transformation was applied to maintain acceptable measures of skewness and kurtosis, the performance of LRT was not compromised.

Furthermore, I also explored the utility of LRT using a genome-wide dataset. There were no genome-wide significant interactions detected when correcting for all SNPs tested. However, some turned out significant after Bonferroni correction when accounting for only the prioritized SNPs at various VP p-value threshold. Moreover, there are strong enrichment signals and a good agreement in the direction of interaction effects at various VP p-value thresholds in the linear regression models investigated. I further conducted the same set of analyses on covariate-adjusted traits to ensure the enrichment signals were not a result of inflation of type I error rate due to heteroscedasticity.

To summarize, I have demonstrated the effectiveness of LRT in prioritizing individual SNPs for genetic interactions using only the variance per genotype and genotype counts. An R Bioconductor package containing the statistical functions will be made publicly available online (`http://www.bioconductor.org/packages/2.12/bioc/html/GEWIST.html`). This new statistical test is expected to complement the existing VP framework and accelerate the process of genetic interaction discovery in future genome-wide studies and meta-analyses. Further, the novel gene-environment

and gene-gene interactions identified will improve our understanding of disease susceptibility, and combined with clinical predictors, help to target personalized disease prevention.

Appendix A

Ethical Considerations

This thesis work involved only secondary use of the genome-wide datasets that are available from the database of Genotype and Phenotype (dbGap) portal at the National Institute of Health (NIH). No research ethics board review was required for the datasets applied. Only participants who consented to General Research Use are included for analysis. Data source has been managed centrally. All computers and servers are password protected and only members of the Paré research group associated with the project have access to the data.

Appendix B

List of Computer Codes

Listing B.1: PLINK code to test for gene-environment interactions

```
1   ###### PLINK SCRIPT TO TEST FOR GENE−ENVIRONMENT ######
2
3   ### 1. Generate Ethnic Specific tped
4   for filename in Caucasian Asian African Hispanic
5   do
6   plink −−noweb \\
7           −−remove ./Pheno/Diabetic/$filename_* \\
8           −−bfile $bfile \\
9           −−keep ./TransPheno/$filename_excluded* \\
10          −−maf 0.05 \\
11          −−hwe 1e−6 \\
12          −−mind 0.05 \\
13          −−geno 0.05 \\
14          −−recode\\
15          −−transpose\\
16          −−out ./Geno/$filename_CLEAN_GENO
17
18  ### 2. Performs Sex check
19
20  plink −−bfile $bfile \\
21          −−keep $pheno_caucasian \\
22          −−check−sex −−out $QCdir/Report_SexCheck
23
24  gawk '$5!="OK" print $1, $2' Report_SexCheck.sexcheck > Report_Sex_discrepancy
25  # 2 IIDs with mismatched sex removed ##
26
27  ### 3. Run Interactions
28
29  plink  −−noweb \\
30              −−bfile $Cbfile \\
31              −−pheno $Caucasian_Pheno \\
32              −−pheno−name trig1_logw \\
33              −−linear \\
34              −−covar $Caucasian_Pheno \\
35              −−covar−name age1c,gender1,waistcm1_logw \\
36              −−interaction \\
37              −−parameters 1,2,3,4,7 \\
38              −−standard−beta \\
39              −−missing−phenotype NA \\
40              −−out $outdir/Interaction/trig_logw_x_waistcm1_logw
```

Listing B.2: Likelihood ratio test

```
1   LRTratio<− function(variance, obs ,verbose=F){
2   ### check inputs
3   if (dim(variance)[2]!=3)
4           stop("number of rows of variance should be 3")
5
6   if (dim(obs)[2]!=3)
7           stop("number of rows of observed genotype counts should be 3")
8
9   if (dim(obs)[2]!=dim(variance)[2])
10          stop("observed genotype counts and variance should have the same number of rows")
11          n1 <− obs[,1]
12          n2 <− obs[,2]
13          n3 <− obs[,3]
14          N <− rowSums(obs, na.rm=T)
15          s1 <− variance[,1]
16          s2 <− variance[,2]
17          s3 <− variance[,3]
18
19  ## constants:
20  log_A <− lgamma(n1/2−1/2)+ lgamma(n2/2−1/2)
21          + lgamma(n3/2−1/2) + log(2)*(n1/2−1/2)+ log(2)*(n2/2−1/2)+ log(2)*(n3/2−1/2)
22  C <− (n1−n3)/(N−9)
23
24  ############ NULL MLE
25
26          opt_theo <− ((n1−1)*s1+(n2−1)*s2+(n3−1)*s3)/(N−9)
27          likelihood.null <− −log_A +(n1/2−3/2)*(log(n1−1)+
28          log(s1)−log(opt_theo))+(n2/2−3/2)*(log(n2−1)+log(s2)−
29          log(opt_theo))+(n3/2−3/2)*(log(n3−1)+log(s3)−log(opt_theo))−
30          (N−9)/2
31
32  ############# ALTERNATIVE
33
34          opt_ratio <− ((n2−1)*s2*C+sqrt(((n2−1)*s2*C)^2+
35                          4*(n1−1)*s1*(n3−1)*s3*(1−C^2)))/(2*((n1−1)*s1)*(1−C))
36          opt_variance <− ((n1−1)*s1*opt_ratio+(n2−1)*s2+(n3−1)*
37                          s3/opt_ratio)/(N−9)
38          likelihood.alter <− −log_A +(n1/2−3/2)*(log(n1−1)+log(s1)−
39          log(opt_variance)+log(opt_ratio))+(n2/2−3/2)*(log(n2−1)+
40          log(s2)−log(opt_variance)) +(n3/2−3/2)*(log(n3−1)+log(s3)−
41          log(opt_variance)−log(opt_ratio))−(N−9)/2
42
43  ##### p_values
44
45          MLR <− 2*likelihood.alter−2*likelihood.null
46          pval <− pchisq(MLR,1,lower.tail=F)
47          if (verbose == F)return(list("LRT_pvalue"=pval))
48          else return(list("Test_Statistics" = MLR, "RatioPar" = opt_ratio))
49                          }
```

Listing B.3: Statistical Power of VP with Five Variance Heterogeneity Tests

```
1   ########### Calculate Simulation Parameters #############
2   b2 <- c(0.35, 0.5)
3   b3 <- c(0.05, 0.08)
4   p <- c(0.1, 0.2, 0.4)
5
6   calc_ve <- function(input){
7   b2 <- input[1]
8   b3 <- input[2]
9   p <- input[3]
10  ve_e <- (b2^2)/(b2^2 + 2*p*(1-p)*b3^2 + 1)
11  ve_ge <- (2*p*(1-p)*b3^2)/(2*p*(1-p)*b3^2 + b2^2 + 1)
12  return(c(ve_e, ve_ge)*100)
13  }
14
15  datamat <- data.frame(rep(b2,each=6),
16                  rep(rep(b3,each=3),2),
17                  rep(rep(p, each=1),4))
18
19  ## Function to Calculate the probability of MAF
20
21          genotype.gen <- function(maf,n){
22          n0 <- round(n*(1-maf)^2)
23          n1 <- round(n*maf*(1-maf)*2)
24          n2 <- n-n0-n1
25          c(n0,n1,n2)}
26
27  ################# Step 1. Set Simulation Parameters ###############
28
29  #@# number of simualtions K and number of SNPs corrected M
30          K <- 5000
31          M <- 500000
32
33          datamat <- data.frame(rep(b2,each=6),
34                  rep(rep(b3,each=3),2),
35                  rep(rep(p, each=1),4))
36
37          N <- c(2000, 5000, 10000, 20000)
38
39  #################### Step 2. Set Output Lists ###################
40
41          Outputlist <- list()
42          for (k in 1:length(N)){
43                  k <- 1
44                  n <- N[k]
45
46          Outputlist[[k]] <- list()
47
48          for (t in 1:dim(datamat)[1]){
49                  p <- datamat[t, 3]
50                  b2 <- datamat[t, 1]
51                  b3 <- datamat[t, 2]
```

```
52                  b1 <- 0
53
54    #################### Step 3. Simulate Data ####################
55    ## simulate genotype and covariate independently
56            Genotype <- sample(rep(0:2,genotype.gen(p,n)))
57            COV <- rnorm(n)
58
59    ## Test P-values
60    test_pvalues <- data.frame(levene_p = NA,levene_t1 = NA,
61                        levene_t2 = NA,bart=NA, intp = NA)
62    mafvar <- data.frame (major=NA, heter=NA, minor=NA)
63
64    for (j in 1:K){
65            error <- rnorm(n)
66            Geno <- sample(Genotype)
67            trait <- b2*COV + b3*COV*Geno + error
68            mafvar[j,] <- c(var(trait[Geno==0]),var(trait[Geno==1]),
69                                var(trait[Geno==2]))
70            interaction_pval <- summary(lm(trait ~ COV*Geno))$coef[4,4]
71
72    ############### Levene
73
74            valid_cases <- complete.cases(trait, as.factor(Geno))
75            mean_per_geno <- tapply(trait[valid_cases], as.factor(Geno)[valid_cases],mean)
76            resp <- abs(trait-mean_per_geno[as.factor(Geno)])
77            levene_pval<- anova(lm(resp ~ as.factor(Geno)))[, c(1, 4, 5)][1,3]
78
79    ############### Levene Trend 1
80            lvt1_pval <- summary(lm(resp~Geno))$coef[2,4]
81
82    ############### Levene Trend 2
83            lvt2_pval <- summary(lm(resp~sqrt(Geno)))$coef[2,4]
84
85    ############### Bartlett Test
86            bart <- bartlett.test(trait ~ Geno)$p.value
87            test_pvalues[j,] <- c(levene_pval, lvt1_pval, lvt2_pval,
88                                pv_pval, bart, interaction_pval)
89            }
90    ############### LRT
91    test_pvalues$lrt_p <- LRTratio(mafvar,
92            obs=data.frame(rep(genotype.gen(p,n)[1],K),
93            rep(genotype.gen(p,n)[2],K),rep(genotype.gen(p,n)[3],K)))$LRT
94
95            Outputlist[[k]][[t]] <- list(mafvar, test_pvalues)
96    }
97    }
98    save.image("POWERsim_K1.rdata")
```

Listing B.4: Statistical Power of VP with Five Variance Heterogeneity Tests

```
1
2  gewistLRT <- function(p, N, obs=NA, theta_gc, theta_c, M, K = 20000,verbose=FALSE){
3
4  ## Input the known parameters (variance explained ~ (0,1))
5  ## Assume Gene x Environment
6
7  ### check inputs
8          if (length(obs) == 3) {
9                  n1 <- obs[1]
10                 n2 <- obs[2]
11                 n3 <- obs[3]
12                 N <- sum(n1, n2, n3, na.rm=TRUE)
13         }else {
14         if (p > 0.5 | p <= 0)
15         stop("minor allele frequency should be a number between 0 and 0.5")
16
17         if (!(N > 0 | M > 0 | K > 0 | theta_gc >0 | theta_c > 0))
18                 stop( "negative input values are not allowed")
19
20         N <- round(N)
21         n1 <- round(N*(1 - p)^2)
22         n2 <- round(N*(1 - p)*p*2)
23         n3 <- N - n1 - n2
24         }
25         if (!(theta_gc < 1 | theta_c < 1 ))
26         stop(" variance explained should be a number between 0 and 1 ")
27
28    M <- round(M)
29    K <- round(K)
30
31  ### calculate beta coefficients
32
33         b2 <- sqrt(( theta_c )/( 1 - theta_gc - theta_c ))
34         b3 <- sqrt(theta_gc/(2*p*(1 - p)*( 1 - theta_gc - theta_c )))
35
36  ############# sample the variance of covariate and error per genotype
37
38         var_C_G_1 <- rchisq(K,n1 - 1)/(n1 - 1)
39         var_C_G_2 <- rchisq(K,n2 - 1)/(n2 - 1)
40         var_C_G_3 <- rchisq(K,n3 - 1)/(n3 - 1)
41
42         error_G_1 <- rchisq(K,n1 - 1)/(n1 - 1)
43         error_G_2 <- rchisq(K,n2 - 1)/(n2 - 1)
44         error_G_3 <- rchisq(K,n3 - 1)/(n3 - 1)
45
46  ########### sample the b_x's
47
48         b_x1 <- rnorm(K,b2 + b3*( - 2*p),sqrt(1/(var_C_G_1*(n1 - 1))))
49         b_x2 <- rnorm(K,b2 + b3*(1 - 2*p),sqrt(1/(var_C_G_2*(n2 - 1))))
50         b_x3 <- rnorm(K,b2 + b3*(2 - 2*p),sqrt(1/(var_C_G_3*(n3 - 1))))
51
```

```
52  ############ the covariance of error term and covariate per genotype
53
54          cov_error_C_1 <- (b_x1 − (b2 + b3*( − 2*p)))*var_C_G_1
55          cov_error_C_2 <- (b_x2 − (b2 + b3*(1 − 2*p)))*var_C_G_2
56          cov_error_C_3 <- (b_x3 − (b2 + b3*(2 − 2*p)))*var_C_G_3
57
58          sum_of_C_square_G <- var_C_G_1*n1*( − 2*p) +
59                  var_C_G_2*n2*(1 − 2*p) + var_C_G_3*n3*(2 − 2*p)
60          sum_of_C_G_square <- var_C_G_1*n1*( − 2*p)^2 +
61                  var_C_G_2*n2*(1 − 2*p)^2 + var_C_G_3*n3*(2 − 2*p)^2
62          sum_C_2 <- var_C_G_1*n1 + var_C_G_2*n2 + var_C_G_3*n3
63
64          total_cov_error <- ( − 2*p)*n1*cov_error_C_1 +
65                          (1 − 2*p)*n2*cov_error_C_2 + (2 − 2*p)*n3*cov_error_C_3
66          total_cov <- n1*cov_error_C_1 + n2*cov_error_C_2 + n3*cov_error_C_3
67
68  ############# observed b2s
69
70          beta2 <- b2 + (sum_of_C_square_G*total_cov_error + total_cov*
71                  sum_of_C_G_square)/(sum_C_2*sum_of_C_G_square +
72                  sum_of_C_square_G^2)
73          beta3 <- b3 + ((b2 − beta2)*sum_of_C_square_G +
74                  total_cov_error)/(sum_of_C_G_square)
75
76  ############ Thus we can obtain the total variance per genotype
77
78          var_group_1 <- abs(var_C_G_1*(b2 + b3*( − 2*p))^2 +
79                  error_G_1 + 2*cov_error_C_1*(b2 + b3*( − 2*p)))
80          var_group_2 <- abs(var_C_G_2*(b2 + b3*(1 − 2*p))^2 +
81                  error_G_2 + 2*cov_error_C_2*(b2 + b3*(1 − 2*p)))
82          var_group_3 <- abs(var_C_G_3*(b2 + b3*(2 − 2*p))^2 +
83                  error_G_3 + 2*cov_error_C_3*(b2 + b3*(2 − 2*p)))
84
85  lrt_p <- LRTratio(data.frame(var_group_1,var_group_2,var_group_3),
86                  obs = data.frame(rep(n1, length(var_group_1)),
87                  rep(n2, length(var_group_2)), rep(n3, length(var_group_3))),
88                  verbose=F)$LRT
89
90  ############# Calculate interation p values
91          RSS1 <- error_G_1 + (b2 + b3*( − 2*p)− beta2 −
92                  beta3*( −2*p))^2*var_C_G_1 + 2*(b2 + b3*( −2*p)
93                  − beta2 − beta3*( − 2*p))*cov_error_C_1
94          RSS2 <- error_G_2 + (b2 + b3*(1 − 2*p) − beta2 −
95                  beta3*(1 − 2*p))^2*var_C_G_2 + 2*(b2 + b3*(1 − 2*p)
96                  − beta2 − beta3*(1 − 2*p))*cov_error_C_2
97          RSS3 <- error_G_3 + (b2 + b3*(2 − 2*p) − beta2 −
98                  beta3*(2 − 2*p))^2*var_C_G_3 + 2*(b2 + b3*(2 − 2*p)
99                  − beta2 − beta3*(2 − 2*p))*cov_error_C_3
100
101         f_stats <- 2*beta3^2*p*(1 − p)*(N − 4)^2/(RSS1*(n1 − 1)
102                 + RSS2*(n2 − 1) + RSS3*(n3 − 1))
103
```

```
104            interaction_p <- 1 - pf(f_stats,df1 = 1,df2 = N - 4)
105   ########## Calculate powers
106
107   result <- data.frame("p-value_Cut-offs" = NA, "VP_power" = NA)
108
109   for (i in 1:1000){
110           power <- mean(lrt_p < i/1000 & interaction_p<0.05/(M*i/1000))
111           result[i,] <- c( i/1000,power)
112                                   }
113           conv_power <- power
114           optimal_power <- max(result[,2])
115           optimal_p_threshold <- which.max(result[,2])/1000
116
117           if (verbose){return(result)
118                                   }else {
119                   return(list("Conventional_power"=conv_power,
120                           "Optimal_VP_power"=optimal_power,
121                           "Optimal_pval_threshold"=optimal_p_threshold))
122                                   }
123                                   }
124   ############### End of Script
125
126   f2.power <- function(theta_ge, theta_e, N, power, alpha=0.05){
127     f2 <- theta_ge/(1-theta_ge-theta_e)*N
128     power-1+pf(qf(1-alpha,df1=1,df2=N-4), df1=1,df2=N-4,ncp=f2)
129   }
130   ### model 1 has 4 parameters - including intercept
131   ### model 2 has 3 parametesr - including intercept
132   ### partial F test statistics df1 = 1, df2 = N-4
133
134   calc_beta <- function(input){
135   theta_c <- input[1]
136   theta_gc <- input[2]
137   p <- input[3]
138
139   b2 <- sqrt(theta_c/(1-theta_c-theta_gc))
140   b3 <- sqrt(theta_gc/(1-theta_c-theta_gc)/(2*p*(1-p)))
141   }
142   ##########################
143   VP_Power <- function(n, effect.size, p.cut=0.05, power, cov_exp, MAF , var.method, m=500000){
144
145   f2 <- effect.size/(1-effect.size-cov_exp)
146   NCP <- (effect.size)/(1-effect.size)*n
147
148   if (var.method=="LRT"){
149   optimal.p <- gewistLRT(MAF, n, theta_gc = effect.size,
150                           theta_c = cov_exp, M = m)$Optimal_pval_threshold
151   } else if (var.method=="Levene") {
152   optimal.p <- gewistLevene(MAF, n, theta_gc = effect.size,
153                           theta_c = cov_exp, M = m)$Optimal_pval_threshold
154   } else {optimal.p=1
155                   }
```

```
156  power−1+pf(qf(1−p.cut/(optimal.p∗m),df1=1,df2=n−2), df1=1,df2=n−4,ncp=NCP)
157  }
158  #############################
159          int_exp <− seq(0.05,1, by = 0.05)/100
160          Conditions <− list(c(0.05, 0.05),c(0.05, 0.2), c(0.05, 0.4),
161                                      c(0.1, 0.05),c(0.1, 0.2), c(0.1, 0.4),
162                                      c(0.2, 0.05),c(0.2, 0.2), c(0.2, 0.4),
163                                      c(0.25, 0.05),c(0.25, 0.2), c(0.25, 0.4))
164
165          sample.size.list <− list()
166          for (j in 1:length(Conditions)){
167                  MAF <− Conditions[[j]][2]
168                  COV <− Conditions[[j]][1]
169
170          sample.size <− data.frame("Interaction" = NA, "VP_Lv" = NA, "VP_LRT" = NA)
171          for (i in 1:length(int_exp)){
172          int <− uniroot(VP_Power, c(1000, 10e10), effect.size = int_exp[i],
173                  p.cut = 0.05, power= 0.8, cov_exp= COV, var.method="None")$root
174          LVint <− uniroot(VP_Power, c(1000, 10e10), effect.size = int_exp[i],
175                  p.cut = 0.05, power= 0.8, cov_exp= COV, MAF = MAF, var.method="Levene")$root
176          LRTint <− uniroot(VP_Power, c(1000, 10e10), effect.size = int_exp[i],
177                  p.cut = 0.05, power= 0.8, cov_exp= COV, MAF = MAF, var.method="LRT")$root
178      sample.size[i,] <− c(int, LVint, LRTint)
179          }
180          sample.size.list[[j]] <− sample.size
181          }
182  setwd("/home/dengw/LRT/Thesis/Simulations")
183  save.image("SAMPLESIZE.rdata")
```

Listing B.5: PLINK R plugin Script to Produce Summary Statistics for LRT and

Levene's Test

```
1   ######## PLINK R plugin Function for Variance Het P−values ####
2   Rplink <− function(PHENO,GENO,CLUSTER,COVAR)
3   {
4     f1 <− function(x)
5     {
6   ######### LRT−Variance Calculation
7
8     PHENO <− (PHENO−mean(PHENO, na.rm=T))/sd(PHENO,na.rm=T)
9
10    mafvar <− c(var(PHENO[x==0],na.rm=T), var(PHENO[x==1],na.rm=T),var(PHENO[x==2],na.rm=T))
11    geno.counts <− c(sum(x==0, na.rm=T), sum(x==1, na.rm=T),sum(x==2, na.rm=T))
12
13   valid_cases <− complete.cases(PHENO, as.factor(x))
14   mean_per_geno <− tapply(PHENO[valid_cases], as.factor(x)[valid_cases],mean)
15
16   responses <− abs(PHENO−mean_per_geno[as.factor(x)])
17
18   z_avg <− c(mean(responses[x==0],na.rm=T),
19                  mean(responses[x==1],na.rm=T),
20                  mean(responses[x==2],na.rm=T))
21
22   z_var <− c(var(responses[x==0],na.rm=T),
23                    var(responses[x==1],na.rm=T),
24                    var(responses[x==2],na.rm=T))
25
26     r <− c(mafvar, geno.counts, z_avg, z_var)
27     c( length(r) , r )
28     }
29   apply(GENO, 2 , f1)
30   }
```

Listing B.6: R Script to Produce Variance Prioritization Results

```
1   ###### Make Figure Prioritiation ########
2
3   traits <- c("trig1_logw", "hdl1_logw", "glucos1c_logw")
4   VarHet_files <- c("trig1_logwVarHet_LeveneLRT_SAVE.txt",
5                     "hdl1_logwVarHet_LeveneLRT_SAVE.txt",
6                     "glucos1c_logwVarHet_LeveneLRT_SAVE.txt")
7
8   interaction_traits <- list(c("trig_logw_x_bmi1c_logw", "trig_logw_x_waistcm1_logw"),
9                     c("hdl1_logw_x_bmi1c_logw", "hdl1_logw_x_waistcm1_logw",
10                    "hdl1_logw_x_trig1_logw"),
11                    c("glucos1c_logw_x_bmi1c_logw", "glucos1c_logw_x_waistcm1_logw"))
12
13  p_threshold=c(0.001, 0.01, 0.05, 0.1)
14
15  ####################### Original Analysis #######################
16  for (t in 1:length(traits)){
17  trait <- traits[t]
18  subset_analyzed_var <- read.table(paste(datadir, VarHet_files[t] ,sep=""), head=T)
19
20  tiff(paste(trait, "_VarPLRT_IntP_QQplot.tiff", sep=""),
21  width = 1000*length(interaction_traits[[t]]), height = 2000,
22  units = "px", res = 200, compression = "lzw")
23  par(mfcol=c(4, length(interaction_traits[[t]])*2), mar=c(4,4,3,1), oma=c(0,0,1,0))
24
25  for (c in 1:length(interaction_traits[[t]])){
26  ####### READ IN DATA FROM PLINK ########
27  subset_int_data <- read.table(paste(datadir, interaction_traits[[t]][c],
28  ".assoc.linear_InteractionP_SAVE.txt", sep=""), head=F)
29  names(subset_int_data) <- c("CHR", "SNP", "BP", "N", "BETA", "TEST", "P")
30
31  ######### MERGE VARIANCE Het output and INTERACTION output #########
32  vp_data <- merge(subset_int_data,subset_analyzed_var,
33                    by.x="SNP", by.y="MarkerName", sort=F)
34
35  for (p in 1:length(p_threshold)){
36  vp <- subset(vp_data, vp_data$LRT < p_threshold[p])
37  #nonvp <- subset(vp_data, vp_data$LRT >= p_threshold[p])
38  produce_qqplot_fun(vp, which.p=which(names(vp_data)=="P"))
39  abline(h=-log10(0.05/dim(vp)[1]))
40  title(paste(fig_LET[[t]][p+8*(c-1)], ")"), line=1)
41  }
42
43  for (p in 1:length(p_threshold)){
44  #vp <- subset(vp_data, vp_data$LRT < p_threshold[p])
45  nonvp <- subset(vp_data, vp_data$LRT >= p_threshold[p])
46  produce_qqplot_fun(nonvp, which.p=which(names(vp_data)=="P"))
47  abline(h=-log10(0.05/dim(nonvp)[1]))
48  title(paste(fig_LET[[t]][p+8*(c-1)+4], ")"), line=1)
49  }
50  }
51  dev.off()
```

```
52   }
53   ################ END:: Original P−values ########################
54   rm(list=ls())
55
56   ###################### Adjusted P−values #####################
57
58   traits <− c("trig1_logw", "hdl1_logw", "glucos1c_logw")
59
60   p_threshold=c(0.001, 0.01, 0.05, 0.1)
61
62   VarHet_adj_files <− list(c("trig_logw_adj_bmi1c_logwVarHet_LeveneLRT_SAVE.txt",
63   "trig_logw_adj_waistcm1_logwVarHet_LeveneLRT_SAVE.txt"),
64   c( "hdl1_logw_adj_bmi1c_logwVarHet_LeveneLRT_SAVE.txt",
65   "hdl1_logw_adj_waistcm1_logwVarHet_LeveneLRT_SAVE.txt",
66   "hdl1_logw_adj_trig1_logwVarHet_LeveneLRT_SAVE.txt"),
67   c( "glucos1c_logw_adj_bmi1c_logwVarHet_LeveneLRT_SAVE.txt",
68   "glucos1c_logw_adj_waistcm1_logwVarHet_LeveneLRT_SAVE.txt"))
69
70   adjusted_traits <− list(c("trig_logw_adj_bmi1c_logwInteractionP_SAVE.txt",
71   "trig_logw_adj_waistcm1_logwInteractionP_SAVE.txt"),
72   c("hdl1_logw_adj_bmi1c_logwInteractionP_SAVE.txt",
73    "hdl1_logw_adj_waistcm1_logwInteractionP_SAVE.txt",
74    "hdl1_logw_adj_trig1_logwInteractionP_SAVE.txt"),
75   c("glucos1c_logw_adj_bmi1c_logwInteractionP_SAVE.txt",
76   "glucos1c_logw_adj_waistcm1_logwInteractionP_SAVE.txt"))
77
78
79   for (t in 1:length(traits)){
80   trait <− traits[t]
81
82   tiff(paste(trait, "_VarPLRT_IntP_Adjusted_QQplot.tiff", sep=""),
83    width = 1000*length(adjusted_traits[[t]]), height = 2000,
84    units = "px", res = 200, compression = "lzw")
85   par(mfcol=c(4, length(adjusted_traits[[t]])*2), mar=c(4,4,3,1), oma=c(0,0,1,0))
86
87   for (c in 1:length(adjusted_traits[[t]])){
88
89   subset_analyzed_var_adj <− read.table(paste(data_dir2, VarHet_adj_files[[t]][c] ,sep=""), head=T)
90   subset_int_data_adj <− read.table(paste(data_dir2, adjusted_traits[[t]][c], sep=""), head=T)
91
92   vp_data <− merge(subset_int_data_adj,subset_analyzed_var_adj,
93                    by.x="SNP", by.y="MarkerName", sort=F)
94
95   for (p in 1:length(p_threshold)){
96   vp <− subset(vp_data, vp_data$LRT < p_threshold[p])
97   #nonvp <− subset(vp_data, vp_data$LRT >= p_threshold[p])
98   produce_qqplot_fun(vp, which.p=which(names(vp_data)=="P"))
99   abline(h=−log10(0.05/dim(vp)[1]))
100  title(paste(fig_LET[[t]][p+8*(c−1)], ")"), line=1)
101  }
102  for (p in 1:length(p_threshold)){
103  #vp <− subset(vp_data, vp_data$LRT < p_threshold[p])
```

```
104  nonvp <− subset(vp_data, vp_data$LRT >= p_threshold[p])
105  produce_qqplot_fun(nonvp, which.p=which(names(vp_data)=="P"))
106  abline(h=−log10(0.05/dim(nonvp)[1]))
107  title(paste(fig_LET[[t]][p+8*(c−1)+4], ")"), line=1)
108  }
109  }
110  dev.off()
111  }
112  ################### END:: Adjusted P−values ###################
113
114   produce_qqplot_fun <− function(data, which.p, titles=NULL){
115              pval_data <− data[order(data[,which.p], decreasing=T, na.last=NA ),]
116              pval_data$pvalref <− ((dim(pval_data)[1]:1)/dim(pval_data)[1])
117
118              pval <− pval_data[,which.p]
119
120              if (length(pval) < 100000){
121              pval_qq11 <− pval
122              ordering <− length(pval)+1 − 1:length(pval)
123              pval_ref <− pval_data$pvalref
124              }else{
125              pval_qq11 <− pval[c(seq(1, sum(pval > 0.01, na.rm=T), 100),
126              seq(sum(pval > 0.01,na.rm=T),sum(pval > 0.001,na.rm=T),10),
127                         sum(pval> 0.001,na.rm=T):length(pval))]
128
129              ordering <− length(pval)+1 − c(seq(1, sum(pval > 0.01, na.rm=T), 100),
130              seq(sum(pval > 0.01,na.rm=T),sum(pval > 0.001,na.rm=T),10),
131                         sum(pval> 0.001,na.rm=T):length(pval))
132              pval_ref <− pval_data$pvalref[c(seq(1, sum(pval > 0.01, na.rm=T), 100),
133              seq(sum(pval > 0.01,na.rm=T),sum(pval > 0.001,na.rm=T),10),
134                         sum(pval> 0.001,na.rm=T):length(pval))]
135                     }
136         c05 <− NA
137         c95 <− NA
138         for(i in 1:length(ordering)){
139                     c95[i] <− qbeta(0.95,ordering[i],length(pval)−ordering[i]+1)
140                     c05[i] <− qbeta(0.05,ordering[i],length(pval)−ordering[i]+1)
141                                 }
142  plot(−log10(pval_ref), −log10(pval_qq11), xlim=c(0,max(−log10(pval_ref))),
143  ylim=c(0,7.5), col="blue", xlab=NA, ylab=NA)
144  title(xlab=expression(Expected~~−log[10](italic(p))),
145   ylab=expression(Observed~~−log[10](italic(p))),line=2.6,
146   main=titles, cex.lab=1, cex.axis=1)
147  polygon(x=c(−log10(pval_ref), rev(−log10(pval_ref))),
148   y=c(−log10(c95), rev(−log10(c05))), col="lightgrey", border=NA)
149  points(−log10(pval_ref), −log10(pval_qq11), pty=1, col="blue")
150  abline(0,1,col=2)
151     }
```

Appendix C

Additional Materials

C.1 Regularity Conditions

The set of variances stratified according to the three possible genotypes are expressed in terms of the linear regression beta coefficients:

$$\sigma_0^2 = \beta_2^2 + 1 \tag{C.1}$$

$$\sigma_1^2 = (\beta_2 + \beta_3)^2 + 1 \tag{C.2}$$

$$\sigma_2^2 = (\beta_2 + 2\beta_3)^2 + 1 \tag{C.3}$$

Increasing variances with the number of minor alleles, or $\sigma_0^2 < \sigma_1^2 < \sigma_2^2$ implies:

$$\beta_2^2 + 1 < (\beta_2 + \beta_3)^2 + 1 < (\beta_2 + 2\beta_3)^2 + 1 \iff$$

$$\beta_2^2 < \beta_2^2 + 2\beta_2\beta_3 + \beta_3^2 < \beta_2^2 + 4\beta_2\beta_3 + 4\beta_3^2 \iff$$

$$0 < 2\beta_2\beta_3 + \beta_3^2 < 4\beta_2\beta_3 + 4\beta_3^2 \iff$$

$$(2\beta_2 + \beta_3)\beta_3 > 0 \quad \& \quad (2\beta_2 + 4\beta_3)\beta_3 > 0 \iff$$

$$\beta_3 > 0 \, \& \, 2\beta_2 + \beta_3 > 0 \, \& \, 2\beta_2 + 3\beta_3 > 0 \quad \text{or} \quad \beta_3 < 0 \, \& \, 2\beta_2 + \beta_3 < 0 \, \& \, 2\beta_2 + 3\beta_3 < 0 \iff$$

$$\beta_3 > 0 \, \& \, \beta_2 > -\frac{\beta_3}{2} \, \& \, \beta_2 > -\frac{3\beta_3}{2} > 0 \quad \text{or} \quad \beta_3 < 0 \, \& \, \beta_2 < -\frac{\beta_3}{2} \, \& \, \beta_2 < -\frac{3\beta_3}{2} \iff$$

$$\beta_3 > 0 \, \& \, \beta_2 > -\frac{\beta_3}{2} > -\frac{3\beta_3}{2} > 0 \quad \text{or} \quad \beta_3 < 0 \, \& \, \beta_2 < -\frac{\beta_3}{2} < -\frac{3\beta_3}{2} \iff$$

$$|\beta_2| > \left|\frac{3\beta_3}{2}\right| \quad \& \quad \beta_2\beta_3 < 0$$

The results for a decreasing trend in variance with the number of minor alleles $(\sigma_0^2 > \sigma_1^2 > \sigma_2^2)$ can be derived analogously.

C.2   The Approximation of Ordered Variances Using a Ratio and a Difference Parameter

To show that a one parameter model with $r = r_1 = r_2$ is a better approximation than a one-parameter model with $d = d_1 = d_2$, I just need to demonstrate the absolute difference between the two ratio parameters is smaller than that between the two difference parameters.

The difference between the two ratios is:

$$
\begin{aligned}
|r_2 - r_1| &= \left| 1 + \frac{2\beta_2\beta_3 + 3\beta_3^2}{(\beta_2+\beta_3)^2+1} - (1 + \frac{2\beta_2\beta_3+\beta_3^2}{\beta_2^2+1}) \right| \\[2mm]
&= \left| \frac{2\beta_2\beta_3+3\beta_3^2}{(\beta_2+\beta_3)^2+1} - \frac{2\beta_2\beta_3+\beta_3^2}{\beta_2^2+1} \right| \\[2mm]
&= \left| \frac{(2\beta_2\beta_3+3\beta_3^2)(\beta_2^2+1) - (2\beta_2\beta_3+\beta_3^2)((\beta_2+\beta_3)^2+1)}{(\beta_2^2+1)((\beta_2+\beta_3)^2+1)} \right| \\[2mm]
&= \left| \frac{(2\beta_2\beta_3+\beta_3^2)(\beta_2^2+1-(\beta_2+\beta_3)^2-1) + 2\beta_3^2(\beta_2^2+1)}{(\beta_2^2+1)((\beta_2+\beta_3)^2+1)} \right| \\[2mm]
&= \left| \frac{-(2\beta_2\beta_3+\beta_3^2)^2 + 2\beta_3^2(\beta_2^2+1)}{(\beta_2^2+1)((\beta_2+\beta_3)^2+1)} \right| \\[2mm]
&= \left| \frac{2\beta_3^2}{((\beta_2+\beta_3)^2+1)} \frac{2(\beta_2^2+1)-(2\beta_2+\beta_3)^2}{2(\beta_2^2+1)} \right| \\[2mm]
&= \frac{2\beta_3^2}{((\beta_2+\beta_3)^2+1)} \left| 1 - \frac{(2\beta_2+\beta_3)^2}{2(\beta_2^2+1)} \right| \qquad\qquad \text{(C.4)}
\end{aligned}
$$

And the difference between the two differences is:

$$
|d_2 - d_1| = 2\beta_3^2 > 0 \qquad\qquad \text{(C.5)}
$$

To demonstrate that the absolute difference between the ratios is always less than that between the differences, I only need to show the maximum of $\left| 1 - \frac{(2\beta_2+\beta_3)^2}{2(\beta_2^2+1)} \right|$ is bounded by $(\beta_2+\beta_3)^2+1$, or equivalently, $\frac{(2\beta_2+\beta_3)^2}{2(\beta_2^2+1)} > 0$ is bounded by $(\beta_2+\beta_3)^2+2$ (Appendix C.2).

So the difference between the two difference parameters is greater than or equal to the difference between the two ratio parameters;

$$
|d_2 - d_1| = 2\beta_3^2 > |r_2 - r_1| \qquad\qquad \text{(C.6)}
$$

To demonstrate under the regularity conditions that the multiplicative model is superior than the additive model, I need to show the maximum of $\left| 1 - \frac{(2\beta_2 + \beta_3)^2}{2(\beta_2^2 + 1)} \right|$ is bounded by $(\beta_2 + \beta_3)^2 + 1$, or equivalently, $\frac{(2\beta_2 + \beta_3)^2}{2(\beta_2^2 + 1)} > 0$ is bounded by $(\beta_2 + \beta_3)^2 + 2$.

$$
\begin{aligned}
(\beta_2 + \beta_3)^2 + 2 - \frac{(2\beta_2 + \beta_3)^2}{2(\beta_2^2 + 1)} &= (\beta_2 + \beta_3)^2 + 2 - \frac{4\beta_2^2 + 4\beta_2\beta_3 + \beta_3^2}{2(\beta_2^2 + 1)} \\
&= (\beta_2 + \beta_3)^2 + 2 - \frac{4\beta_2^2 + 4 - 4 + 4\beta_2\beta_3 + \beta_3^2}{2(\beta_2^2 + 1)} \\
&= (\beta_2 + \beta_3)^2 - \frac{4\beta_2\beta_3 + \beta_3^2 - 4}{2(\beta_2^2 + 1)} \\
&= \frac{2(\beta_2^2 + 1)(\beta_2 + \beta_3)^2}{2(\beta_2^2 + 1)} - \frac{4\beta_2\beta_3 + \beta_3^2 - 4}{2(\beta_2^2 + 1)} \\
&= \frac{2\beta_2^2\beta_3^2 + 2\beta_3^2 + 4\beta_2^3\beta_3 + 4\beta_2\beta_3 + 2\beta_2^4 + 2\beta_2^2 - 4\beta_2\beta_3 - \beta_3^2 + 4}{2(\beta_2^2 + 1)} \\
&= \frac{2\beta_2^2\beta_3^2 + \beta_3^2 + 4\beta_2^3\beta_3 + 2\beta_2^4 + 2\beta_2^2 + 4}{2(\beta_2^2 + 1)} \\
&= \frac{2\beta_2^2(\beta_3^2 + 2\beta_2\beta_3 + \beta_2^2) + \beta_3^2 + 2\beta_2^2 + 4}{2(\beta_2^2 + 1)} \\
&= \frac{2\beta_2^2(\beta_3 + \beta_2)^2 + \beta_3^2 + 2\beta_2^2 + 4}{2(\beta_2^2 + 1)} > 0
\end{aligned}
\tag{C.7}
$$

C.3 Second Partial Derivative Test

The second order partial derivatives of the log-likelihood function under the alternative model are:

$$\frac{\mathrm{d}^2 l_1}{\mathrm{d}(\sigma_1^2)^2} = \frac{N-9}{2(\sigma_1^2)^2} - \frac{(N_0-1)s_0^2 r + (N_1-1)s_1^2 + (N_2-1)s_2^2/r}{(\sigma_1^2)^3}, \tag{C.8}$$

$$\frac{\mathrm{d}^2 l_1}{\mathrm{d}(\sigma_1^2)\mathrm{d}r} = \frac{(N_0-1)s_0^2 + (N_2-1)s_2^2/r^2}{2(\sigma_1^2)^2}, \tag{C.9}$$

$$\frac{\mathrm{d}^2 l_1}{\mathrm{d}r^2} = \frac{N_0-N_2}{2}(-\frac{1}{r^2}) - \frac{(N_2-1)s_2^2/r^3}{(\sigma_1^2)} < 0, \tag{C.10}$$

$$\frac{\mathrm{d}^2 l_1}{\mathrm{d}r\mathrm{d}\sigma_1^2} = \frac{(N_0-1)s_0^2 + (N_2-1)s_2^2/r^2}{2(\sigma_1^2)^2}. \tag{C.11}$$

The second order derivative with respect to σ_1^2 evaluated at the MLEs, $\hat{r}$ and $\hat{\sigma_1}^2$, is negative:

$$\frac{\mathrm{d}^2 l_1}{\mathrm{d}(\sigma_1^2)^2}(\hat{r}, \hat{\sigma_1}^2) = \frac{N-9}{2(\hat{\sigma_1}^2)^2} - \frac{(N_0-1)s_0^2 r + (N_1-1)s_1^2 + (N_2-1)s_2^2/r}{(\hat{\sigma_1}^2)^3} = -\frac{N-9}{2(\hat{\sigma_1}^2)^2} < 0 \tag{C.12}$$

The determinant of the Hessian matrix evaluated at $\hat{r}$ and $\hat{\sigma_1}^2$, is then:

$$\begin{aligned}
\mathrm{D}(\hat{r}, \hat{\sigma_1}^2) &= \det(H((\hat{r}, \hat{\sigma_1}^2))) \\
&= \frac{\mathrm{d}^2 l_1}{\mathrm{d}(\sigma_1^2)^2}\frac{\mathrm{d}^2 l_1}{\mathrm{d}r^2}(\hat{r}, \hat{\sigma_1}^2) - \left(\frac{\mathrm{d}^2 l_1}{\mathrm{d}r\mathrm{d}\sigma_1^2}(\hat{r}, \hat{\sigma_1}^2)\right)^2 \\
&= -\frac{N-9}{2(\hat{\sigma_1}^2)^2}(\frac{N_0-N_2}{2}(-\frac{1}{\hat{r}^2}) - \frac{(N_2-1)s_2^2/\hat{r}^3}{(\hat{\sigma_1}^2)}) - (\frac{(N_0-1)s_0^2 + (N_2-1)s_2^2/\hat{r}^2}{2(\hat{\sigma_1}^2)^2})^2
\end{aligned}$$

98

It is extremely difficult to work out the expression in terms of $\hat{r}$ alone and show the minimum of that function is strictly positive. Instead, I simulate plausible range of ratio values and show the Hessian determinants are indeed positive. The conditional variance given the heterozygote genotype (σ_1^2) was assumed to be 1 for simplicity. I considered ratio of size 0.8, 0.9, 1, 1.1, 1.2, encompassing the range of ratios observed empirically (presented in Chapter 4). In addition, considering that the likelihood is also a function of genotype counts, combinations of MAF (0.05, 0.1, 0.2, 0.4) and sample size (2,000, 5,000, 10,000, 20,000) were also investigated (Table 15).

MAF (p)	N	r = 0.8 $(\hat{r}, \hat{\sigma_1}^2)$	Hessian Det	r = 0.9 $(\hat{r}, \hat{\sigma_1}^2)$	Hessian Det	r = 1 $(\hat{r}, \hat{\sigma_1}^2)$	Hessian Det	r = 1.1 $(\hat{r}, \hat{\sigma_1}^2)$	Hessian Det	r = 1.2 $(\hat{r}, \hat{\sigma_1}^2)$	Hessian Det
0.05	2000	(0.82,1.03)	137938.95	(0.93,1.03)	108989.45	(1.03, 1.03)	88282.33	(1.13 ,1.03)	72959.85	(1.24,1.03)	61306.15
0.05	5000	(0.81,1.01)	904563.76	(0.91,1.01)	714690.25	(1.01,1.01)	578911.02	(1.11,1.01)	478433.07	(1.21,1.01)	402013.13
0.05	10000	(0.80,1.01)	3657255.74	(0.91,1.01)	2889665.07	(1.01,1.01)	2340681.59	(1.11,1.01)	1934423.94	(1.21, 1.01)	1625436.46
0.05	20000	(0.80,1.00)	14736155.49	(0.90,1.00)	11643298.31	(1.00,1.00)	9431294.56	(1.10,1.00)	7794362.11	(1.20, 1.00)	6549355.79
0.1	2000	(0.81, 1.01)	271449.00	(0.91, 1.01)	214478.17	(1.01, 1.01)	173727.27	(1.12,1.01)	143576.39	(1.22, 1.01)	120644.47
0.1	5000	(0.80, 1.01)	1733110.52	(0.90, 1.01)	1369370.77	(1.01, 1.01)	1109189.97	(1.11, 1.01)	916707.50	(1.21, 1.01)	770274.42
0.1	10000	(0.80, 1.00)	6981721.42	(0.90, 1.00)	5516419.79	(1.00, 1.00)	4468298.83	(1.10, 1.00)	3692908.83	(1.20, 1.00)	3103000.77
0.1	20000	(0.80, 1.00)	28025852.07	(0.90, 1.00)	22143874.88	(1.00, 1.00)	17936533.62	(1.10, 1.00)	14824001.61	(1.20, 1.00)	12455987.40
0.2	2000	(0.80, 1.01)	491710.54	(0.91, 1.01)	388515.34	(1.01, 1.01)	314695.94	(1.11, 1.01)	260078.25	(1.21, 1.01)	218537.97
0.2	5000	(0.80, 1.00)	3104203.09	(0.90, 1.00)	2452726.18	(1.00, 1.00)	1986698.86	(1.10, 1.00)	1641892.07	(1.20, 1.00)	1379645.54
0.2	10000	(0.80, 1.00)	12458372.77	(0.90, 1.00)	9843745.44	(1.00, 1.00)	7973395.64	(1.10, 1.00)	6589550.48	(1.20, 1.00)	5537053.46
0.2	20000	(0.80, 1.00)	49916800.01	(0.90, 1.00)	39440788.49	(1.00, 1.00)	31946884.16	(1.10, 1.00)	26402250.07	(1.20, 1.00)	22185226.93
0.4	2000	(0.80, 1.00)	744301.43	(0.90, 1.00)	588089.65	(1.00, 1.00)	476352.95	(1.10, 1.00)	393679.67	(1.20, 1.00)	330800.48
0.4	5000	(0.80, 1.00)	4673243.90	(0.90, 1.00)	3692437.29	(1.00, 1.00)	2990876.04	(1.10, 1.00)	2471796.09	(1.20, 1.00)	2076996.26
0.4	10000	(0.80, 1.00)	18721520.61	(0.90, 1.00)	14792303.25	(1.00, 1.00)	11981772.59	(1.10, 1.00)	9902283.15	(1.20, 1.00)	8320671.71
0.4	20000	(0.80, 1.00)	74943221.43	(0.90, 1.00)	59214359.90	(1.00, 1.00)	47963658.55	(1.10, 1.00)	39639355.84	(1.20, 1.00)	33308081.73

Table 15: Hessian Determinant Calculated from Simulated Data

C.4 Tables of Type I Error Inflation Rates Associated with Non-normal Distributions When MAF is 5%, 10%, or 40%

Type of Distribution		Student's *t*-Distribution Degrees of Freedom				Skew-Normal Distribution Shape Parameter			
		5	10	20	50	0.5	1	2	4
α	Variance Tests								
0.001	LRT	0.0626	0.0088	0.0032	0.0028	0.0006	0.0012	0.0044	0.0060
	Bartlett	0.0572	0.0058	0.0018	0.0014	0.0006	0.0008	0.0022	0.0028
	Levene	0.0080	0.0024	0.0016	0.0016	0.0014	0.0014	0.0022	0.0022
	LV Linear	0.0014	0.0012	0.0008	0.0014	0.0006	0.0010	0.0016	0.0030
	LV Square	0.0030	0.0006	0.0014	0.0016	0.0004	0.0010	0.0008	0.0012
0.01	LRT	0.1346	0.0390	0.0188	0.0158	0.0128	0.0122	0.0216	0.0260
	Bartlett	0.1278	0.0334	0.0146	0.0118	0.0106	0.0100	0.0170	0.0208
	Levene	0.0202	0.0124	0.0074	0.0094	0.0108	0.0090	0.0150	0.0204
	LV Linear	0.0118	0.0134	0.0080	0.0104	0.0112	0.0102	0.0148	0.0192
	LV Square	0.0182	0.0100	0.0084	0.0102	0.0092	0.0090	0.0114	0.0094
0.05	LRT	0.2340	0.1174	0.0700	0.0600	0.0558	0.0634	0.0740	0.0878
	Bartlett	0.2398	0.1084	0.0670	0.0554	0.0500	0.0536	0.0668	0.0812
	Levene	0.0536	0.0504	0.0446	0.0498	0.0498	0.0512	0.0562	0.0694
	LV Linear	0.0578	0.0532	0.0470	0.0524	0.0524	0.0514	0.0606	0.0704
	LV Square	0.0520	0.0478	0.0432	0.0454	0.0512	0.0486	0.0474	0.0446
0.1	LRT	0.3082	0.1852	0.1304	0.1144	0.1106	0.1144	0.1328	0.1514
	Bartlett	0.3284	0.1762	0.1330	0.1144	0.0978	0.1054	0.1242	0.1468
	Levene	0.0984	0.1016	0.1000	0.1060	0.1050	0.1046	0.1086	0.1368
	LV Linear	0.1066	0.1020	0.0922	0.1018	0.1000	0.1010	0.1152	0.1322
	LV Square	0.0952	0.1006	0.0878	0.0976	0.0970	0.0964	0.1004	0.0888

Table 16: **Type I Error Rates Associated with Non-normal Distributions (MAF = 5%)** Each condition was simulated 5,000 times with 2,000 individuals. A MAF of 5% was used throughout. The empirical type I error rate was calculated as the proportion of SNPs with variance heterogeneity p-values less than the nominal α-level.

Type of Distribution		Student's t-Distribution Degrees of Freedom				Skew-Normal Distribution Shape Parameter			
		5	10	20	50	0.5	1	2	4
α	Variance Tests								
0.001	LRT	0.0602	0.0062	0.0030	0.0012	0.0008	0.0014	0.0026	0.0044
	Bartlett	0.0784	0.0088	0.0018	0.0016	0.0008	0.0014	0.0014	0.0048
	Levene	0.0016	0.0024	0.0008	0.0010	0.0014	0.0014	0.0014	0.0024
	LV Linear	0.0008	0.0016	0.0012	0.0008	0.0008	0.0008	0.0022	0.0030
	LV Square	0.0008	0.0012	0.0004	0.0002	0.0010	0.0004	0.0014	0.0014
0.01	LRT	0.1324	0.0328	0.0226	0.0118	0.0104	0.0140	0.0154	0.0282
	Bartlett	0.1682	0.0368	0.0212	0.0128	0.0106	0.0128	0.0154	0.0272
	Levene	0.0142	0.0094	0.0104	0.0082	0.0112	0.0100	0.0102	0.0204
	LV Linear	0.0102	0.0086	0.0098	0.0096	0.0100	0.0096	0.0116	0.0174
	LV Square	0.0098	0.0086	0.0088	0.0088	0.0100	0.0110	0.0092	0.0090
0.05	LRT	0.2476	0.1080	0.0732	0.0568	0.0526	0.0588	0.0702	0.0966
	Bartlett	0.3030	0.1182	0.0846	0.0606	0.0538	0.0576	0.0712	0.0978
	Levene	0.0556	0.0458	0.0538	0.0466	0.0522	0.0518	0.0544	0.0782
	LV Linear	0.0484	0.0480	0.0478	0.0510	0.0516	0.0558	0.0570	0.0784
	LV Square	0.0448	0.0444	0.0516	0.0502	0.0508	0.0512	0.0494	0.0520
0.1	LRT	0.3272	0.1758	0.1274	0.1122	0.1054	0.1116	0.1346	0.1592
	Bartlett	0.4040	0.1982	0.1496	0.1162	0.1020	0.1086	0.1364	0.1684
	Levene	0.1062	0.0970	0.1064	0.1008	0.1036	0.1080	0.1122	0.1480
	LV Linear	0.1034	0.0960	0.0948	0.1006	0.0988	0.1072	0.1158	0.1360
	LV Square	0.0968	0.0944	0.0982	0.0980	0.0970	0.1072	0.0994	0.1024

Table 17: **Type I Error Rates Associated with Non-normal Distributions (MAF = 10%)** Each condition was simulated 5,000 times with 2,000 individuals. A MAF of 10% was used throughout. The empirical type I error rate was calculated as the proportion of SNPs with variance heterogeneity p-values less than the nominal α-level.

Type of Distribution		Student's t-Distribution Degrees of Freedom				Skew-Normal Distribution Shape Parameter			
		5	10	20	50	0.5	1	2	4
α	Variance Tests								
0.001	LRT	0.0678	0.0104	0.0028	0.0010	0.0012	0.0004	0.0012	0.0040
	Bartlett	0.1078	0.0112	0.0026	0.0008	0.0020	0.0010	0.0012	0.0060
	Levene	0.0010	0.0010	0.0010	0.0006	0.0020	0.0006	0.0012	0.0028
	LV Linear	0.0004	0.0026	0.0012	0.0014	0.0008	0.0004	0.0010	0.0024
	LV Square	0.0008	0.0018	0.0012	0.0008	0.0010	0.0004	0.0006	0.0002
0.01	LRT	0.1418	0.0344	0.0148	0.0098	0.0096	0.0084	0.0168	0.0242
	Bartlett	0.2176	0.0496	0.0212	0.0118	0.0108	0.0108	0.0198	0.0312
	Levene	0.0096	0.0134	0.0106	0.0094	0.0118	0.0094	0.0116	0.0232
	LV Linear	0.0098	0.0106	0.0084	0.0108	0.0112	0.0092	0.0140	0.0216
	LV Square	0.0084	0.0110	0.0072	0.0088	0.0088	0.0064	0.0098	0.0110
0.05	LRT	0.2538	0.1130	0.0742	0.0544	0.0478	0.0552	0.0724	0.0870
	Bartlett	0.3704	0.1314	0.0836	0.0528	0.0496	0.0566	0.0756	0.1036
	Levene	0.0490	0.0516	0.0514	0.0476	0.0516	0.0542	0.0648	0.0846
	LV Linear	0.0526	0.0540	0.0494	0.0466	0.0490	0.0514	0.0632	0.0726
	LV Square	0.0512	0.0490	0.0510	0.0450	0.0490	0.0500	0.0574	0.0506
0.1	LRT	0.3342	0.1738	0.1260	0.1094	0.0962	0.1076	0.1254	0.1490
	Bartlett	0.4620	0.2136	0.1494	0.1080	0.1000	0.1098	0.1442	0.1768
	Levene	0.1022	0.1032	0.1080	0.0930	0.1004	0.1034	0.1176	0.1490
	LV Linear	0.1024	0.0978	0.0992	0.0938	0.0958	0.1034	0.1154	0.1322
	LV Square	0.1040	0.1042	0.1004	0.0994	0.0946	0.1030	0.1040	0.1010

Table 18: **Type I Error Rates Associated with Non-normal Distributions (MAF = 40%)** Each condition was simulated 5,000 times with 2,000 individuals. A MAF of 40% was used throughout. The empirical type I error rate was calculated as the proportion of SNPs with variance heterogeneity p-values less than the nominal α-level.

Bibliography

Aschard, H., Zaitlen, N., Tamimi, R. M., Lindstrom, S., and Kraft, P. (2013). A non-parametric test to detect quantitative trait loci where the phenotypic distribution differs by genotypes. *Genet Epidemiol*, **37**(4), 323–33.

Baranzini, S. E., Galwey, N. W., Wang, J., Khankhanian, P., Lindberg, R., Pelletier, D., Wu, W., Uitdehaag, B. M., Kappos, L., Polman, C. H., Matthews, P. M., Hauser, S. L., Gibson, R. A., Oksenberg, J. R., and Barnes, M. R. (2009). Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. *Hum Mol Genet*, **18**(11), 2078–90.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London Series A - Mathematical and Physical Sciences*, **160**(901), 268–282.

Bild, D. E., Bluemke, D. A., Burke, G. L., Detrano, R., Diez Roux, A. V., Folsom, A. R., Greenland, P., Jacob, D. R., J., Kronmal, R., Liu, K., Nelson, J. C., O'Leary, D., Saad, M. F., Shea, S., Szklo, M., and Tracy, R. P. (2002). Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol*, **156**(9), 871–81.

Bookman, E. B., McAllister, K., Gillanders, E., Wanke, K., Balshaw, D., Rutter,

J., Reedy, J., Shaughnessy, D., Agurs-Collins, T., Paltoo, D., Atienza, A., Bierut, L., Kraft, P., Fallin, M. D., Perera, F., Turkheimer, E., Boardman, J., Marazita, M. L., Rappaport, S. M., Boerwinkle, E., Suomi, S. J., Caporaso, N. E., Hertz-Picciotto, I., Jacobson, K. C., Lowe, W. L., Goldman, L. R., Duggal, P., Gunnar, M. R., Manolio, T. A., Green, E. D., Olster, D. H., Birnbaum, L. S., and for the NIH G-E Interplay Workshop participants (2011). Gene-environment interplay in common complex diseases: forging an integrative modelÑrecommendations from an NIH workshop. *Genetic Epidemiology*, **35**(4), 217–225.

Brown, M. B. and Forsythe, A. B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, **69**(346), 364–367.

Bush, W. S., Dudek, S. M., and Ritchie, M. D. (2009). Biofilter: a knowledge-integration system for the multi-locus analysis of genome-wide association studies. *Pac Symp Biocomput*, pages 368–79.

Conover, W. J., Johnson, M. E., and Johnson, M. M. (1981). A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics*, **23**(4), 351–361.

Cordell, H. J. (2009). Detecting gene-gene interactions that underlie human diseases. *Nature*, **10**, 392–404.

Culverhouse, R., Suarez, B. K., Lin, J., and Reich, T. (2002). A perspective on epistasis: limits of models displaying no main effect. *Am J Hum Genet*, **70**(2), 461–71.

Deng, W. and Paré, G. (2011). A fast algorithm to optimize SNP prioritization for

gene-gene and gene-environment interactions. *Genetic epidemiology*, **35**(7), 729–738.

Deng, W. and Paré, G. (2012). *GEWIST: Gene Environment Wide Interaction Search Threshold*. Bioconductor R package version 1.5.0.

Edmondson, A. C. and Rader, D. J. (2008). Genome-wide approaches to finding novel genes for lipid traits: the start of a long road. *Circ Cardiovasc Genet*, **1**(1), 3–6.

Evans, D. M., Marchini, J., Morris, A. P., and Cardon, L. R. (2006). Two-stage two-locus models in genome-wide association. *PLoS Genet*, **2**(9), e157.

Fisher, R. A. (1958). *The Genetical Theory of Natural Selection*. Dover books on science. Dover Publications, New York,, 2d rev. edition.

Fujino, Y. (1979). Tests for the homogeneity of a set of variances against ordered alternatives. *Biometrika*, **66**(1), 133–139.

Gastwirth, J. L., Gel, Y. R., and Miao, W. (2009). The impact of Levene's test of equality of variances on statistical theory and practice. *Statistical Science*, **24**(3), 343–360.

Goldstein, D. B. (2009). Common genetic variation and human traits. *N Engl J Med*, **360**(17), 1696–8.

Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., and Manolio, T. A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA*, **106**(23), 9362–7.

Hines, W. G. and Hines, R. J. (2000). Increased power with modified forms of the Levene (med) test for heterogeneity of variance. *Biometrics*, **56**(2), 451–4.

Hogg, R. V., McKean, J. W., and Craig, A. T. (2005). *Introduction to Mathematical Statistics*. Pearson Prentice Hall, Upper Saddle River, N.J., 6th edition.

Kathiresan, S., Manning, A. K., Demissie, S., D'Agostino, R. B., Surti, A., Guiducci, C., Gianniny, L., Burtt, N. P., Melander, O., Orho-Melander, M., Arnett, D. K., Peloso, G. M., Ordovas, J. M., and Cupples, L. A. (2007). A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. *BMC Med Genet*, **8 Suppl 1**, S17.

Keyes, T. K. and Levy, M. S. (1997). Analysis of Levene's test under design imbalance. *Journal of Educational and Behavioral Statistics*, **22**(2), 227–236.

Kooperberg, C. and Leblanc, M. (2008). Increasing the power of identifying gene x gene interactions in genome-wide association studies. *Genet Epidemiol*, **32**(3), 255–63.

Kraja, A. T., Vaidya, D., Pankow, J. S., Goodarzi, M. O., Assimes, T. L., Kullo, I. J., Sovio, U., Mathias, R. A., Sun, Y. V., Franceschini, N., Absher, D., Li, G., Zhang, Q., Feitosa, M. F., Glazer, N. L., Haritunians, T., Hartikainen, A. L., Knowles, J. W., North, K. E., Iribarren, C., Kral, B., Yanek, L., O'Reilly, P. F., McCarthy, M. I., Jaquish, C., Couper, D. J., Chakravarti, A., Psaty, B. M., Becker, L. C., Province, M. A., Boerwinkle, E., Quertermous, T., Palotie, L., Jarvelin, M. R., Becker, D. M., Kardia, S. L., Rotter, J. I., Chen, Y. D., and Borecki, I. B. (2011). A bivariate genome-wide approach to metabolic syndrome: Stampeed consortium. *Diabetes*, **60**(4), 1329–39.

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin, editor, *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, pages 278–292. Stanford University Press, Stanford, CA.

Lindstrom, S., Yen, Y. C., Spiegelman, D., and Kraft, P. (2009). The impact of gene-environment dependence and misclassification in genetic association studies incorporating gene-environment interactions. *Hum Hered*, **68**(3), 171–81.

Mailman, M. D., Feolo, M., Jin, Y., Kimura, M., Tryka, K., Bagoutdinov, R., Hao, L., Kiang, A., Paschall, J., Phan, L., Popova, N., Pretel, S., Ziyabari, L., Lee, M., Shao, Y., Wang, Z. Y., Sirotkin, K., Ward, M., Kholodov, M., Zbicz, K., Beck, J., Kimelman, M., Shevelev, S., Preuss, D., Yaschenko, E., Graeff, A., Ostell, J., and Sherry, S. T. (2007). The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet*, **39**(10), 1181–6.

Manning, A. K., LaValley, M., Liu, C. T., Rice, K., An, P., Liu, Y., Miljkovic, I., Rasmussen-Torvik, L., Harris, T. B., Province, M. A., Borecki, I. B., Florez, J. C., Meigs, J. B., Cupples, L. A., and Dupuis, J. (2011). Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP x environment regression coefficients. *Genet Epidemiol*, **35**(1), 11–8.

Marchini, J., Donnelly, P., and Cardon, L. R. (2005). Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat Genet*, **37**(4), 413–7.

Meyre, D., Delplanque, J., Chevre, J. C., Lecoeur, C., Lobbens, S., Gallina, S., Durand, E., Vatin, V., Degraeve, F., Proenca, C., Gaget, S., Korner, A., Kovacs, P., Kiess, W., Tichet, J., Marre, M., Hartikainen, A. L., Horber, F., Potoczna, N., Hercberg, S., Levy-Marchal, C., Pattou, F., Heude, B., Tauber, M., McCarthy,

M. I., Blakemore, A. I., Montpetit, A., Polychronakos, C., Weill, J., Coin, L. J., Asher, J., Elliott, P., Jarvelin, M. R., Visvikis-Siest, S., Balkau, B., Sladek, R., Balding, D., Walley, A., Dina, C., and Froguel, P. (2009). Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nat Genet*, **41**(2), 157–9.

Millstein, J., Conti, D. V., Gilliland, F. D., and Gauderman, W. J. (2006). A testing framework for identifying susceptibility genes in the presence of epistasis. *Am J Hum Genet*, **78**(1), 15–27.

Murcray, C. E., Lewinger, J. P., and Gauderman, W. J. (2009). Gene-environment interaction in genome-wide association studies. *Am J Epidemiol*, **169**(2), 219–26.

O'Neill, M. E. and Mathews, K. (2000). Theory & methods: A weighted least squares approach to Levene's test of homogeneity of variance. *Australian & New Zealand Journal of Statistics*, **42**(1), 81–100.

Panagiotou, O. A. and Ioannidis, J. P. (2012). What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. *Int J Epidemiol*, **41**(1), 273–86.

Paré, G., Cook, N., Ridker, P., and DI., C. (2010). On the use of variance per genotype as a tool to identify quantitative trait interaction effects: a report from the Women's Genome Health Study. *PLoS Genet*, **6**(6).

Pe'er, I., Yelensky, R., Altshuler, D., and Daly, M. J. (2008). Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol*, **32**(4), 381–5.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., and Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, **81**(3), 559–75.

R Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rice, T. K., Schork, N. J., and Rao, D. C. (2008). Methods for handling multiple testing. *Adv Genet*, **60**, 293–308.

Ritchie, M. D. (2009). Using prior knowledge and genome-wide association to identify pathways involved in multiple sclerosis. *Genome Med*, **1**(6), 65.

Sham, P. C., Lin, M. W., Zhao, J. H., and Curtis, D. (2000). Power comparison of parametric and nonparametric linkage tests in small pedigrees. *Am J Hum Genet*, **66**(5), 1661–8.

Snedecor, G. W. and Cochran, W. G. (1980). *Statistical Methods*. Iowa State University Press, Ames, Iowa, 7th edition.

Struchalin, M. V., Dehghan, A., Witteman, J. C., van Duijn, C., and Aulchenko, Y. S. (2010). Variance heterogeneity analysis for detection of potentially interacting genetic loci: method and its limitations. *BMC Genet*, **11**, 92.

Struchalin, M. V., Amin, N., Eilers, P. H., van Duijn, C. M., and Aulchenko, Y. S. (2012). An R package "VariABEL" for genome-wide searching of potentially interacting loci by testing genotypic variance heterogeneity. *BMC Genet*, **13**, 4.

Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki, M., Pirruccello, J. P., Ripatti, S., Chasman, D. I., Willer, C. J., Johansen, C. T., Fouchier, S. W., Isaacs, A., Peloso, G. M., Barbalic, M., Ricketts, S. L., Bis, J. C., Aulchenko, Y. S., Thorleifsson, G., Feitosa, M. F., Chambers, J., Orho-Melander, M., Melander, O., Johnson, T., Li, X., Guo, X., Li, M., Shin Cho, Y., Jin Go, M., Jin Kim, Y., Lee, J. Y., Park, T., Kim, K., Sim, X., Twee-Hee Ong, R., Croteau-Chonka, D. C., Lange, L. A., Smith, J. D., Song, K., Hua Zhao, J., Yuan, X., Luan, J., Lamina, C., Ziegler, A., Zhang, W., Zee, R. Y., Wright, A. F., Witteman, J. C., Wilson, J. F., Willemsen, G., Wichmann, H. E., Whitfield, J. B., Waterworth, D. M., Wareham, N. J., Waeber, G., Vollenweider, P., Voight, B. F., Vitart, V., Uitterlinden, A. G., Uda, M., Tuomilehto, J., Thompson, J. R., Tanaka, T., Surakka, I., Stringham, H. M., Spector, T. D., Soranzo, N., Smit, J. H., Sinisalo, J., Silander, K., Sijbrands, E. J., Scuteri, A., Scott, J., Schlessinger, D., Sanna, S., Salomaa, V., Saharinen, J., Sabatti, C., Ruokonen, A., Rudan, I., Rose, L. M., Roberts, R., Rieder, M., Psaty, B. M., Pramstaller, P. P., Pichler, I., Perola, M., Penninx, B. W., Pedersen, N. L., Pattaro, C., Parker, A. N., Pare, G., Oostra, B. A., O'Donnell, C. J., Nieminen, M. S., Nickerson, D. A., Montgomery, G. W., Meitinger, T., McPherson, R., McCarthy, M. I., *et al.* (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, **466**(7307), 707–13.

Thomas, D. (2010). Gene-environment-wide association studies: emerging approaches. *Nat Rev Genet*, **11**(4), 259–272.

Visscher, P. M. and Posthuma, D. (2010). Statistical power to detect genetic loci affecting environmental sensitivity. *Behav Genet*, **40**(5), 728–33.

Visscher, P. M., Hill, W. G., and Wray, N. R. (2008). Heritability in the genomics era–concepts and misconceptions. *Nat Rev Genet*, **9**(4), 255–66.

Willer, C. J., Sanna, S., Jackson, A. U., Scuteri, A., Bonnycastle, L. L., Clarke, R., Heath, S. C., Timpson, N. J., Najjar, S. S., Stringham, H. M., Strait, J., Duren, W. L., Maschio, A., Busonero, F., Mulas, A., Albai, G., Swift, A. J., Morken, M. A., Narisu, N., Bennett, D., Parish, S., Shen, H., Galan, P., Meneton, P., Hercberg, S., Zelenika, D., Chen, W. M., Li, Y., Scott, L. J., Scheet, P. A., Sundvall, J., Watanabe, R. M., Nagaraja, R., Ebrahim, S., Lawlor, D. A., Ben-Shlomo, Y., Davey-Smith, G., Shuldiner, A. R., Collins, R., Bergman, R. N., Uda, M., Tuomilehto, J., Cao, A., Collins, F. S., Lakatta, E., Lathrop, G. M., Boehnke, M., Schlessinger, D., Mohlke, K. L., and Abecasis, G. R. (2008). Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet*, **40**(2), 161–9.

WTCC (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**(7145), 661–78.

Yang, J., Loos, R. J., Powell, J. E., Medland, S. E., Speliotes, E. K., Chasman, D. I., Rose, L. M., Thorleifsson, G., Steinthorsdottir, V., Magi, R., Waite, L., Smith, A. V., Yerges-Armstrong, L. M., Monda, K. L., Hadley, D., Mahajan, A., Li, G., Kapur, K., Vitart, V., Huffman, J. E., Wang, S. R., Palmer, C., Esko, T., Fischer, K., Zhao, J. H., Demirkan, A., Isaacs, A., Feitosa, M. F., Luan, J., Heard-Costa, N. L., White, C., Jackson, A. U., Preuss, M., Ziegler, A., Eriksson, J., Kutalik, Z., Frau, F., Nolte, I. M., Van Vliet-Ostaptchouk, J. V., Hottenga, J. J., Jacobs, K. B., Verweij, N., Goel, A., Medina-Gomez, C., Estrada, K., Bragg-Gresham,

J. L., Sanna, S., Sidore, C., Tyrer, J., Teumer, A., Prokopenko, I., Mangino, M., Lindgren, C. M., Assimes, T. L., Shuldiner, A. R., Hui, J., Beilby, J. P., McArdle, W. L., Hall, P., Haritunians, T., Zgaga, L., Kolcic, I., Polasek, O., Zemunik, T., Oostra, B. A., Junttila, M. J., Gronberg, H., Schreiber, S., Peters, A., Hicks, A. A., Stephens, J., Foad, N. S., Laitinen, J., Pouta, A., Kaakinen, M., Willemsen, G., Vink, J. M., Wild, S. H., Navis, G., Asselbergs, F. W., Homuth, G., John, U., Iribarren, C., Harris, T., Launer, L., Gudnason, V., O'Connell, J. R., Boerwinkle, E., Cadby, G., Palmer, L. J., James, A. L., Musk, A. W., Ingelsson, E., Psaty, B. M., Beckmann, J. S., Waeber, G., Vollenweider, P., Hayward, C., Wright, A. F., Rudan, I., *et al.* (2012). FTO genotype is associated with phenotypic variability of body mass index. *Nature*, **490**(7419), 267–72.

Zuk, O., Hechter, E., Sunyaev, S. R., and Lander, E. S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci USA*, **109**(4), 1193–8.