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
© Copyright by Wei Q. Deng, August 2013
All Rights Reserved
TITLE: A statistical framework to detect gene-environment interactions 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

Abstract iii

Acknowledgements v

List of abbreviations vi

1 Introduction 1

2 Statistical Methods to Prioritize SNPs for Genetic Interactions 9

2.1 Preliminaries 10

2.2 A Mathematical Representation of Variance Conditional on the Genotype 13

2.2.1 Quantitative Trait Variance per Genotype 13

2.2.2 Characterization of Ordered Variances Using a Ratio Parameter 14

2.3 A Likelihood Ratio Test for Variance Heterogeneity 16

2.3.1 Derivation of the Likelihood Ratio Test Statistic 17

2.3.2 Calculation of Non-centrality Parameter and Statistical Power 21

2.4 Other Variance Heterogeneity Tests 23

2.4.1 Levene’s Test 23
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.1</td>
<td>Regularity Conditions</td>
<td>94</td>
</tr>
<tr>
<td>C.2</td>
<td>The Approximation of Ordered Variances Using a Ratio and a Difference Parameter</td>
<td>95</td>
</tr>
<tr>
<td>C.3</td>
<td>Second Partial Derivative Test</td>
<td>97</td>
</tr>
<tr>
<td>C.4</td>
<td>Tables of Type I Error Inflation Rates Associated with Non-normal Distributions When MAF is 5%, 10%, or 40%</td>
<td>101</td>
</tr>
</tbody>
</table>
## List of Tables

1. Trends in Conditional Variances According to the Directions of Interaction and Interacting Covariate Main Effects .......................... 14
2. Estimated VP Power to Detect a Gene-Environment Interaction in a Sample of 2,000 Individuals ....................................................... 31
3. Estimated VP Power to Detect a Gene-Environment Interaction in a Sample of 10,000 Individuals ...................................................... 32
4. Sample Size at 80% Power to Detect a Gene-Environment Interaction using VP and an Exhaustive Search ........................................ 35
5. Type I Error Rates Associated with Non-normal Distributions (MAF = 20%) .................................................................................. 40
6. Summary of SNP Quality Controls .............................................................................................................................. 44
7. Summary Statistics of Selected Traits and Interacting Covariates ................................................................................. 45
8. Summary of Candidate Models for VP ................................................................. 46
9. Top Ten SNPs Involved in Variance Heterogeneity Ranked by LRT p-values .................................................................................... 50
10. Top Ten SNPs Involved in Variance Heterogeneity Ranked by Levene’s Test p-values ............................................................... 52
11 Top Ten SNPs per Model Ranked by Gene-Environment Interaction $p$-values ................................................................. 54
12 SNPs per Model with Gene-Environment Interaction $p$-values Passing Bonferroni Correction Using Variance Prioritization .................. 57
13 Concordance in the Direction of Effects ................................................. 66
14 Concordance in the Direction of Effects for Residualized Traits .... 67
15 Hessian Determinant Calculated from Simulated Data ................. 100
16 Type I Error Rates Associated with Non-normal Distributions (MAF = 5%) .......................................................... 101
17 Type I Error Rates Associated with Non-normal Distributions (MAF = 10%) .......................................................... 102
18 Type I Error Rates Associated with Non-normal Distributions (MAF = 40%) .......................................................... 103
List of Figures

1  Variance Prioritization Procedure ............................................. 6
2  Comparison of VP Power with Variance Heterogeneity Tests and an
   Exhaustive Search in a Sample of 2,000 Individuals ...................... 29
3  Comparison of VP Power with Variance Heterogeneity Tests and an
   Exhaustive Search in a Sample of 10,000 Individuals .................. 30
4  Sample Size at 80% Power to Detect a Gene-Environment Interaction
   using VP and an Exhaustive Search on a log 10 scale .................... 36
5  Probability Density Plots of Student’s t and Skew-Normal Distributions 38
6  Quantile-Quantile Plots of Variance Heterogeneity Test p-values using
   Untransformed Traits ............................................................. 48
7  Quantile-Quantile Plots of Variance Heterogeneity Test p-values using
   Transformed Traits .................................................................. 49
8  Manhattan Plots of Gene-Environment Interaction p-values for TG .... 53
9  Manhattan Plots of Gene-Environment Interaction p-values for HDL-C 55
10 Manhattan Plots of Gene-Environment Interaction p-values for FG ..... 56
11 Venn Diagrams Showing the Number of SNPs Prioritized using Lev-
    ene’s test and LRT (TG) ............................................................ 61
12 Venn Diagrams Showing the Number of SNPs Prioritized using Levene’s test and LRT (HDL-C) ................................. 61
13 Venn Diagrams Showing the Number of SNPs Prioritized using Levene’s test and LRT (FG) ........................................ 62
14 Quantile-Quantile Plots of Gene-Environment Interaction \( p \)-values for SNPs Prioritized with LRT (TG) ............................. 63
15 Quantile-Quantile Plots of Gene-Environment Interaction \( p \)-values for SNPs Prioritized with LRT (HDL-C) ........................ 64
16 Quantile-Quantile Plots of Gene-Environment Interaction \( p \)-values for SNPs Prioritized with LRT (FG) ............................. 65
17 Distribution of Variance Heterogeneity Test \( p \)-values for TG before and after Adjusting for the Interacting Covariate ............... 70
18 Distribution of Variance Heterogeneity Test \( p \)-values for HDL-C before and after Adjusting for the Interacting Covariate .... 71
19 Distribution of Variance Heterogeneity Test \( p \)-values for FG before and after Adjusting for the Interacting Covariate .......... 72
20 Gene-Environment Interaction \( p \)-value Before and After Adjusting for the Interacting Covariate ................................. 73
21 Quantile-Quantile Plots of Gene-Environment Interaction \( p \)-values of SNPs Prioritized Using LRT on TG Adjusted for the Interacting Covariates ................................................................. 74
22 Quantile-Quantile Plots of Gene-Environment Interaction \( p \)-values of SNPs Prioritized Using LRT on HDL-C Adjusted for the Interacting Covariates ................................................................. 75
23 Quantile-Quantile Plots of Gene-Environment Interaction $p$-values of
SNPs Prioritized Using LRT on FG Adjusted for the Interacting Co-
variates ................................................................. 76
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
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 ($\eta_0$) 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 $\eta$ 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.
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 TGAA to TGGA, 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:

\[
N_0 = (1 - p)^2 N \quad \text{(2.1)}
\]
\[
N_1 = 2(1 - p)(p)N \quad \text{(2.2)}
\]
\[
N_2 = (p)^2 N \quad \text{(2.3)}
\]

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 $G \sim B(2, p)$. The statistical interaction between the genotype variable $G$ and an environmental covariate $C$ can be tested by the linear regression model:

$$Y = \beta_0 + \beta_1 G + \beta_2 C + \beta_3 GC + \epsilon \quad (2.4)$$

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, $\epsilon \sim N(0,1)$. In addition, assume the error term $\epsilon$ 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):

$$Var(Y) = Var(\beta_0 + \beta_1 G + \beta_2 C + \beta_3 GC + \epsilon)$$

$$= Var(\beta_1 G) + Var(\beta_2 C) + Var(\beta_3 GC) + Var(\epsilon)$$

$$+ 2Cov(\beta_1 G, \beta_2 C) + 2Cov(\beta_2 C, \beta_3 GC) + 2Cov(\beta_1 G, \beta_3 GC)$$

$$= Var(\beta_1 G) + Var(\beta_2 C) + Var(\beta_3 GC) + Var(\epsilon)$$

$$= 2p(1 - p)\beta_1^2 + \beta_2^2 Var(C) + \beta_3^2 Var(G) Var(C) + Var(\epsilon)$$

$$= 2p(1 - p)\beta_1^2 + \beta_2^2 + 2p(1 - p)\beta_3^2 + 1$$

(2.5)

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

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

(2.6)

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

$$\theta_{E} = \frac{Var(\beta_2 C)}{Var(Y)} = \frac{\beta_2^2}{2p(1 - p)\beta_1^2 + \beta_2^2 + 2p(1 - p)\beta_3^2 + 1}$$

(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:

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

$$= \text{Var}(\beta_2 C + \beta_3 GC + \varepsilon|G = g)$$

$$= (\beta_2 + \beta_3 g)^2 \text{Var}(C|G = g) + \text{Var}(\varepsilon|G = g)$$

$$= (\beta_2 + \beta_3 g)^2 + \text{Var}(\varepsilon|G = g) \quad (2.8)$$

Assume the error term has constant variances across genotypes:

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

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

$$\sigma_0^2 = \beta_2^2 + 1 \quad (2.10)$$

$$\sigma_1^2 = (\beta_2 + \beta_3)^2 + 1 \quad (2.11)$$

$$\sigma_2^2 = (\beta_2 + 2\beta_3)^2 + 1 \quad (2.12)$$
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:


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. $|\beta_2| > \frac{3|\beta_3|}{2}$ & $\beta_2\beta_3 < 0$ (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

<table>
<thead>
<tr>
<th></th>
<th>$\beta_2 &gt; 0$</th>
<th>$\beta_2 &lt; 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_3 &gt; 0$</td>
<td>$\sigma_0^2 &lt; \sigma_1^2 &lt; \sigma_2^2$</td>
<td>$\sigma_0^2 &gt; \sigma_1^2 &gt; \sigma_2^2$</td>
</tr>
<tr>
<td>$\beta_3 &lt; 0$</td>
<td>$\sigma_0^2 &gt; \sigma_1^2 &gt; \sigma_2^2$</td>
<td>$\sigma_0^2 &lt; \sigma_1^2 &lt; \sigma_2^2$</td>
</tr>
</tbody>
</table>

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 $r_2$.

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

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

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 $d_2$.

$$d_1 = \sigma_1^2 - \sigma_0^2 = 2\beta_2\beta_3 + \beta_3^2 \quad (2.15)$$

$$d_2 = \sigma_2^2 - \sigma_1^2 = 2\beta_2\beta_3 + 3\beta_3^2 \quad (2.16)$$

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 Heterogeneity

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:

$$H_0 : r = 1 \text{ versus } H_1 : r \neq 1$$

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)

$$\Lambda = 2(l_1 - l_o) \quad (2.17)$$
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 $\sigma^2_g$. 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 $\sigma^2_g$ (Hogg et al., 2005):

$$
\frac{(N_g - 1)s_g^2}{\sigma^2_g} \sim \chi^2(N_g - 1) \quad (2.18)
$$

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 $\sigma^2_g$ and ratio parameter $r$, the probability distribution function of the major homozygote group variance is

$$
f(s_0^2, \sigma_1^2, r) = \frac{1}{2^{\frac{N_0-1}{2}} \Gamma\left(\frac{N_0-1}{2}\right)} \left(\frac{(N_0 - 1)s_0^2}{2\sigma_1^2/r}\right)^{\frac{N_0-3}{2}} \exp\left(-\frac{(N_0 - 1)s_0^2}{2\sigma_1^2/r}\right), \quad (2.19)
$$
the density function for the heterozygote group variance is

\[
f(s_1^2; \sigma_1^2) = \frac{1}{2^{(N_1-1)/2} \Gamma\left(\frac{N_1-1}{2}\right)} \left(\frac{(N_1 - 1)s_1^2}{2\sigma_1^2}\right)^{\frac{N_1-3}{2}} \exp\left(-\frac{(N_1 - 1)s_1^2}{2\sigma_1^2}\right), \tag{2.20}
\]

and the density function for the minor homozygote group variance is

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

**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 \(\sigma_1^2\) are estimated by maximizing the log-likelihood function:

\[
l_1(\sigma_1^2, r | s_0^2, s_1^2, s_2^2) = \sum_{g=0}^2 \log(f(s_g^2; \sigma_1^2, r))
\]

\[
= -\log(A) + \frac{N_0-3}{2} \left(\log(N_0 - 1) + \log(s_0^2) + \log(r) - \log(\sigma_1^2)\right)
\]

\[
+ \frac{N_1-3}{2} \left(\log(N_1 - 1) + \log(s_1^2) - \log(\sigma_1^2)\right)
\]

\[
+ \frac{N_2-3}{2} \left(\log(N_2 - 1) + \log(s_2^2) - \log(\sigma_2^2) - \log(r)\right)
\]

\[
= \frac{(N_0 - 1)s_0^2 r + (N_1 - 1)s_1^2 + (N_2 - 1)s_2^2/r}{2\sigma_1^2}
\]

where \(A = 2^{\frac{N_0-1}{2}} \Gamma\left(\frac{N_0-1}{2}\right) 2^{\frac{N_1-1}{2}} \Gamma\left(\frac{N_1-1}{2}\right) 2^{\frac{N_2-1}{2}} \Gamma\left(\frac{N_2-1}{2}\right)\)
The first partial derivatives of the log-likelihood function with respect to \( \sigma_1^2 \) and \( r \) are

\[
\frac{d l_1(\sigma_1^2, r | s_0^2, s_1^2, s_2^2)}{d \sigma_1^2} = -\frac{N_0 + N_1 + N_2 - 9}{2\sigma_1^2} + \frac{(N_0 - 1)s_0^2r + (N_1 - 1)s_1^2 + (N_2 - 1)s_2^2 / r}{2(\sigma_1^2)^2}
\]

(2.22)

and

\[
\frac{d l_1(\sigma_1^2, r | s_0^2, s_1^2, s_2^2)}{d r} = \frac{N_0 - 3}{2} \left( \frac{1}{r} \right) + \frac{N_2 - 3}{2} \left( -\frac{1}{r} \right) - \frac{(N_0 - 1)s_0^2 - (N_2 - 1)s_2^2 / r^2}{2(\sigma_1^2)^2}.
\]

(2.23)

Set the above partial derivatives to zero and the MLEs of \( \sigma_1^2 \) and \( r \) are obtained by solving the equations simultaneously (select only the positive root as \( \hat{r} \))

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

(2.24)

and

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

(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 $\sigma^2$. The log-likelihood function under the null is:

$$l_0(\sigma^2 | s_0^2, s_1^2, s_2^2) = \sum_{g=0}^{2} \log(f(s_g^2; \sigma^2))$$

$$= -\log(A) + \frac{N_0 - 3}{2} (\log(N_0 - 1) + \log(s_0^2) - \log(\sigma^2))$$

$$+ \frac{N_1 - 3}{2} (\log(N_1 - 1) + \log(s_1^2) - \log(\sigma^2))$$

$$+ \frac{N_2 - 3}{2} (\log(N_2 - 1) + \log(s_2^2) - \log(\sigma^2))$$

$$- \frac{(N_0 - 1)s_0^2 + (N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{2\sigma^2}$$

(2.26)

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

$$\frac{dl_0(\sigma^2 | s_0^2, s_1^2, s_2^2)}{d\sigma^2} = \frac{N_0 + N_1 + N_2 - 9}{2\sigma^2} + \frac{(N_0 - 1)s_0^2 + (N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{2\sigma^4}$$

(2.27)

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

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

(2.28)
The second partial derivative of the log-likelihood function is strictly negative:

\[
\frac{d^2 l_0(\sigma^2 | s_0^2, s_1^2, s_2^2)}{d^2 \sigma^2} = \frac{N_0 + N_1 + N_2 - 9}{2} \left( \frac{1}{\sigma^4} \right) + \frac{(N_0 - 1)s_0^2 + (N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{\sigma^6}
\]

\[
= \frac{1}{\sigma^4} \left( \frac{N - 9}{2} - \frac{(N_0 - 1)s_0^2 + (N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{\sigma^2} \right)
\]

\[
= \frac{1}{\sigma^4} \left( \frac{N - 9}{2} - (N - 9) \right)
\]

\[
= -\frac{1}{\sigma^4} \left( \frac{N - 9}{2} \right) < 0
\]

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 $\lambda$ is computed as the expected difference of the test statistic $\Lambda$ 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:

\[
\begin{align*}
E(\Lambda) &= 2E(l_1(\sigma^2, r|\sigma_0^2, \sigma_1^2, S_0^2, S_1^2, S_2^2) - l_0(\sigma^2|\sigma_0^2, \sigma_1^2, S_0^2, S_1^2, S_2^2)) \\
 &= 2E(- \log(A) + \frac{N_0 - 3}{2} (\log(N_0 - 1) + \log(S_0^2) + \log(r) - \log(\sigma_1^2)) \\
&\quad + \frac{N_1 - 3}{2} (\log(N_1 - 1) + \log(S_1^2) - \log(\sigma_1^2)) + \frac{N_2 - 3}{2} (\log(N_2 - 1) \\
&\quad + \log(S_2^2) - \log(\sigma_2^2) - \log(r)) - \frac{(N_0 - 1)S_0^2r + (N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{2\sigma_1^2} \\
&\quad - (\log(A) + \frac{N_0 - 3}{2} (\log(N_0 - 1) + \log(S_0^2) - \log(\sigma^2)) \\
&\quad + \frac{N_1 - 3}{2} (\log(N_1 - 1) + \log(S_1^2) - \log(\sigma^2)) + \frac{N_2 - 3}{2} (\log(N_2 - 1) + \log(S_2^2) - \log(\sigma^2)) \\
&\quad - \frac{(N_0 - 1)S_0^2 + (N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{2\sigma^2}) \\
&= 2E(\frac{N_0 - 3}{2} (\log(r) - \log(\sigma_1^2) + \log(\sigma^2)) + \frac{N_1 - 3}{2} (- \log(\sigma_1^2) + \log(\sigma^2)) \\
&\quad + \frac{N_2 - 3}{2} (- \log(\sigma_2^2) - \log(r) + \log(\sigma^2)) - \frac{(N_0 - 1)S_0^2r + (N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{2\sigma_1^2} \\
&\quad + \frac{(N_0 - 1)S_0^2 + (N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{2\sigma^2}) \\
&= (N - 9) \log \frac{\sigma^2}{\sigma_1^2} + (N_0 - N_2) \log(r) - \frac{(N - 9)\sigma_1^2}{\sigma_1^2} - \frac{(N - 9)\sigma^2}{\sigma^2} \\
&= (N - 9) \log \frac{\sigma^2}{\sigma_1^2} + (N_0 - N_2) \log(r) \\
&\quad (2.30)
\end{align*}
\]

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

\[
\begin{align*}
E(S_0^2) &= \sigma_0^2 = \frac{1}{2}\sigma_1^2/r \\
E(S_1^2) &= \sigma_1^2 \\
E(S_2^2) &= \sigma_2^2 = \sigma_1^2 r
\end{align*}
\]

(2.31) (2.32) (2.33)

The non-centrality parameter is thus \((N - 9) \log \left(\frac{\sigma^2}{\sigma_1^2}\right) + (N_0 - N_2) \log(r) - 1.\)
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 \( \bar{Z} \) the overall mean. Levene’s test statistic to assess whether the variances conditional on the genotype are equal is:
\[
L = \frac{(N - 3) \sum_{g=0}^{2} N_g (\bar{Z}_g - \bar{Z})^2}{(3 - 1) \sum_{g=0}^{2} \sum_{j=1}^{N_g} (z_{gj} - \bar{Z}_g)^2} \tag{2.34}
\]

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):

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

(2.35)

where

\[
S_p^2 = \frac{\sum_{g=0}^{2} (N_g - 1) S_g^2}{N - 3}
\]  

(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...
correcting for \( M = 500,000 \) hypotheses simultaneously (interaction \( p \)-value < \( 0.05/M \)). For each VP \( p \)-value threshold (\( \eta \)), 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 (\( \eta_o \)) 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 \( \eta \) 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 \( \eta \). 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 \( \eta_o \) 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 \( \eta_o \), 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.
### Table 2: Estimated VP Power to Detect a Gene-Environment Interaction in a Sample of 2,000 Individuals

<table>
<thead>
<tr>
<th>Simulation Parameters</th>
<th>Simulation Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_2$</td>
<td>$\beta_3$</td>
</tr>
<tr>
<td>0.05</td>
<td>0.1 (0.040%, 10.9%)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>0.35</td>
<td>0.4 (0.107%, 10.9%)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>0.08</td>
<td>0.1 (0.103%, 10.9%)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.4 (0.273%, 10.9%)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td>0.1 (0.036%, 20.0%)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>0.08</td>
<td>0.4 (0.096%, 20.0%)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.1 (0.096%, 20.0%)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.
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 $\sim 10\%$ 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.
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,
<table>
<thead>
<tr>
<th>$\theta_{GE}$ (%)</th>
<th>Exhaustive Search</th>
<th>$\theta_E = 5%$</th>
<th>$\theta_E = 10%$</th>
<th>$\theta_E = 15%$</th>
<th>$\theta_E = 20%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>76073</td>
<td>76068</td>
<td>75043</td>
<td>69108</td>
<td>67558</td>
</tr>
<tr>
<td>0.1</td>
<td>38025</td>
<td>38023</td>
<td>37285</td>
<td>34642</td>
<td>33815</td>
</tr>
<tr>
<td>0.15</td>
<td>25342</td>
<td>25339</td>
<td>24604</td>
<td>23092</td>
<td>22661</td>
</tr>
<tr>
<td>0.2</td>
<td>19000</td>
<td>18984</td>
<td>18472</td>
<td>17370</td>
<td>16680</td>
</tr>
<tr>
<td>0.25</td>
<td>15196</td>
<td>15191</td>
<td>14767</td>
<td>13786</td>
<td>13427</td>
</tr>
<tr>
<td>0.3</td>
<td>12659</td>
<td>12651</td>
<td>12252</td>
<td>11402</td>
<td>11161</td>
</tr>
<tr>
<td>0.35</td>
<td>10847</td>
<td>10809</td>
<td>10588</td>
<td>9766</td>
<td>9582</td>
</tr>
<tr>
<td>0.4</td>
<td>9488</td>
<td>9386</td>
<td>9180</td>
<td>8549</td>
<td>8368</td>
</tr>
<tr>
<td>0.45</td>
<td>8431</td>
<td>8425</td>
<td>8154</td>
<td>7609</td>
<td>7376</td>
</tr>
<tr>
<td>0.5</td>
<td>7586</td>
<td>7505</td>
<td>7321</td>
<td>6825</td>
<td>6662</td>
</tr>
<tr>
<td>0.55</td>
<td>6894</td>
<td>6804</td>
<td>6649</td>
<td>6207</td>
<td>6093</td>
</tr>
<tr>
<td>0.6</td>
<td>6318</td>
<td>6166</td>
<td>6113</td>
<td>5694</td>
<td>5541</td>
</tr>
<tr>
<td>0.65</td>
<td>5830</td>
<td>5744</td>
<td>5622</td>
<td>5278</td>
<td>5114</td>
</tr>
<tr>
<td>0.7</td>
<td>5412</td>
<td>5272</td>
<td>5222</td>
<td>4869</td>
<td>4750</td>
</tr>
<tr>
<td>0.75</td>
<td>5049</td>
<td>4929</td>
<td>4865</td>
<td>4587</td>
<td>4416</td>
</tr>
<tr>
<td>0.8</td>
<td>4732</td>
<td>4600</td>
<td>4589</td>
<td>4280</td>
<td>4145</td>
</tr>
<tr>
<td>0.85</td>
<td>4452</td>
<td>4340</td>
<td>4294</td>
<td>4022</td>
<td>3910</td>
</tr>
<tr>
<td>0.9</td>
<td>4204</td>
<td>4080</td>
<td>4050</td>
<td>3784</td>
<td>3683</td>
</tr>
<tr>
<td>0.95</td>
<td>3981</td>
<td>3854</td>
<td>3823</td>
<td>3594</td>
<td>3469</td>
</tr>
<tr>
<td>1</td>
<td>3781</td>
<td>3685</td>
<td>3613</td>
<td>3394</td>
<td>3202</td>
</tr>
</tbody>
</table>

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 log10 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 log10 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 log10 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 $\eta_0$ 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.

![Student’s t distribution and Skew-Normal distribution](image)

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 \( \alpha \)-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 \( \alpha \)-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
### 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 $\alpha$-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.

<table>
<thead>
<tr>
<th>Type of Distribution</th>
<th>Student’s $t$ Distribution</th>
<th>Skew-Normal Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Degrees of Freedom</td>
<td>Shape Parameter</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Theoretical Skewness</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Mean Sample Skewness</td>
<td>-0.013</td>
<td>-0.001</td>
</tr>
<tr>
<td>Standard Error</td>
<td>0.0072</td>
<td>0.0016</td>
</tr>
<tr>
<td>Mean Sample Kurtosis</td>
<td>7.699</td>
<td>3.983</td>
</tr>
<tr>
<td>Standard Error</td>
<td>0.1320</td>
<td>0.0067</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Variance Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>LRT 0.0648</td>
</tr>
<tr>
<td></td>
<td>Bartlett 0.0918</td>
</tr>
<tr>
<td></td>
<td>Levene 0.0020</td>
</tr>
<tr>
<td></td>
<td>LV Linear 0.0014</td>
</tr>
<tr>
<td></td>
<td>LV Square 0.0018</td>
</tr>
<tr>
<td>0.01</td>
<td>LRT 0.1414</td>
</tr>
<tr>
<td></td>
<td>Bartlett 0.1892</td>
</tr>
<tr>
<td></td>
<td>Levene 0.0122</td>
</tr>
<tr>
<td></td>
<td>LV Linear 0.0118</td>
</tr>
<tr>
<td></td>
<td>LV Square 0.0100</td>
</tr>
<tr>
<td>0.05</td>
<td>LRT 0.2562</td>
</tr>
<tr>
<td></td>
<td>Bartlett 0.3306</td>
</tr>
<tr>
<td></td>
<td>Levene 0.0534</td>
</tr>
<tr>
<td></td>
<td>LV Linear 0.0494</td>
</tr>
<tr>
<td></td>
<td>LV Square 0.0468</td>
</tr>
<tr>
<td>0.1</td>
<td>LRT 0.3398</td>
</tr>
<tr>
<td></td>
<td>Bartlett 0.4238</td>
</tr>
<tr>
<td></td>
<td>Levene 0.1002</td>
</tr>
<tr>
<td></td>
<td>LV Linear 0.1088</td>
</tr>
<tr>
<td></td>
<td>LV Square 0.0950</td>
</tr>
</tbody>
</table>
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 anal-
ysis 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.

43
<table>
<thead>
<tr>
<th>Quality Control Filters</th>
<th>SNPs Lost</th>
<th>SNPs Remained</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP probes</td>
<td>934,940</td>
<td></td>
</tr>
<tr>
<td>Genotyping center filters</td>
<td>25,318</td>
<td>909,622</td>
</tr>
<tr>
<td>HWE $&lt; 1E-6$ in Caucasian samples</td>
<td>3,279</td>
<td></td>
</tr>
<tr>
<td>Missing rate per SNP $&gt; 5%$</td>
<td>7,312</td>
<td></td>
</tr>
<tr>
<td>MAF $&lt; 5%$</td>
<td>223,242</td>
<td></td>
</tr>
<tr>
<td>Percentage of SNPs removed due to filters</td>
<td></td>
<td>27.2%</td>
</tr>
<tr>
<td>Non-autosomal SNPs</td>
<td>24,546</td>
<td>680,550</td>
</tr>
<tr>
<td></td>
<td></td>
<td>656,004</td>
</tr>
</tbody>
</table>

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.
Table 7: Summary Statistics of Selected Traits and Interacting Covariates

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

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 of 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).

<table>
<thead>
<tr>
<th>Response Variable</th>
<th>Interacting Covariate</th>
<th>Pearson’s Correlation</th>
<th>Variance Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>TG</td>
<td>BMI</td>
<td>0.319</td>
<td>10.07%</td>
</tr>
<tr>
<td></td>
<td>WC</td>
<td>0.3228</td>
<td>10.39%</td>
</tr>
<tr>
<td>HDL-C</td>
<td>TG</td>
<td>-0.439</td>
<td>19.86%</td>
</tr>
<tr>
<td></td>
<td>WC</td>
<td>-0.369</td>
<td>13.54%</td>
</tr>
<tr>
<td></td>
<td>BMI</td>
<td>-0.316</td>
<td>9.907%</td>
</tr>
<tr>
<td>FG</td>
<td>WC</td>
<td>0.3475</td>
<td>11.59%</td>
</tr>
<tr>
<td></td>
<td>BMI</td>
<td>0.283</td>
<td>8.223%</td>
</tr>
</tbody>
</table>

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.
<table>
<thead>
<tr>
<th>Trait</th>
<th>SNP</th>
<th>Nearest Gene*</th>
<th>CHR Position</th>
<th>MAFSD per Genotype</th>
<th>Ratio</th>
<th>p-values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>S0</td>
<td>S1</td>
<td>S2</td>
<td>LRT</td>
</tr>
<tr>
<td>TG</td>
<td>rs1934101</td>
<td>ELAVL2</td>
<td>9</td>
<td>23220746</td>
<td>0.37</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td>rs10757400</td>
<td>ELAVL2</td>
<td>9</td>
<td>23187358</td>
<td>0.36</td>
<td>1.07</td>
</tr>
<tr>
<td></td>
<td>rs2651976</td>
<td>TMCC3</td>
<td>12</td>
<td>93554392</td>
<td>0.46</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td>rs7742224</td>
<td>AKAP12</td>
<td>6</td>
<td>151618963</td>
<td>0.49</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td>rs1339142</td>
<td>ELAVL2</td>
<td>9</td>
<td>23209125</td>
<td>0.24</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>rs2651985</td>
<td>TMCC3</td>
<td>12</td>
<td>93559305</td>
<td>0.45</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>rs10811869</td>
<td>ELAVL2</td>
<td>9</td>
<td>23190174</td>
<td>0.34</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>rs1005872</td>
<td>CDH12</td>
<td>5</td>
<td>21868337</td>
<td>0.26</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>rs7874120</td>
<td>ELAVL2</td>
<td>9</td>
<td>23192654</td>
<td>0.35</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>rs6878249</td>
<td>ITGA1</td>
<td>5</td>
<td>52210980</td>
<td>0.21</td>
<td>0.95</td>
</tr>
<tr>
<td>HDL-C</td>
<td>rs12932428</td>
<td>FTO</td>
<td>16</td>
<td>52518028</td>
<td>0.48</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>rs1775757</td>
<td>MMD</td>
<td>17</td>
<td>50964405</td>
<td>0.12</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td>rs1253426</td>
<td>ARHGAP19</td>
<td>10</td>
<td>98960900</td>
<td>0.08</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>rs61150</td>
<td>PARP11</td>
<td>12</td>
<td>3875013</td>
<td>0.18</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>rs10887396</td>
<td>FAM190B</td>
<td>10</td>
<td>86652805</td>
<td>0.40</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td>rs1767676</td>
<td>GAD2</td>
<td>10</td>
<td>26564812</td>
<td>0.08</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>rs10019557</td>
<td>LPHN3</td>
<td>4</td>
<td>61367329</td>
<td>0.15</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>rs9364097</td>
<td>SMAP1</td>
<td>6</td>
<td>71418475</td>
<td>0.30</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>rs9972377</td>
<td>MCTP2</td>
<td>15</td>
<td>92716873</td>
<td>0.22</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>rs10852208</td>
<td>MCTP2</td>
<td>15</td>
<td>92717430</td>
<td>0.22</td>
<td>1.04</td>
</tr>
<tr>
<td>FG</td>
<td>rs9905394</td>
<td>HS3ST3B1</td>
<td>17</td>
<td>14216969</td>
<td>0.45</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td>rs7700535</td>
<td>COX7C</td>
<td>5</td>
<td>85671585</td>
<td>0.13</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>rs11193637</td>
<td>SORCS1</td>
<td>10</td>
<td>109460510</td>
<td>0.18</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>rs7727064</td>
<td>COX7C</td>
<td>5</td>
<td>85669119</td>
<td>0.12</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>rs637278</td>
<td>CELF2</td>
<td>10</td>
<td>11259469</td>
<td>0.07</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>rs636416</td>
<td>CELF2</td>
<td>10</td>
<td>11259469</td>
<td>0.07</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>rs1195842</td>
<td>COX7C</td>
<td>5</td>
<td>85799252</td>
<td>0.11</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>rs6725114</td>
<td>EPC2</td>
<td>2</td>
<td>149083368</td>
<td>0.09</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>rs4792477</td>
<td>HS3ST3B1</td>
<td>17</td>
<td>14215811</td>
<td>0.42</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>rs642349</td>
<td>CELF2</td>
<td>10</td>
<td>11260140</td>
<td>0.07</td>
<td>1.06</td>
</tr>
</tbody>
</table>

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
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 \( (r^2 = 0.119; D' = 0.823; \text{Distance} = 69.07 \text{ KB}) \) and rs12543154 \( (r^2 = 0.112; D' = 0.820; \text{Distance} = 70.40 \text{KB}) \) 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.
<table>
<thead>
<tr>
<th>Trait</th>
<th>SNP</th>
<th>Nearest Gene*</th>
<th>CHR Position</th>
<th>MAFSD per Genotype</th>
<th>Ratio</th>
<th>p-values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 interaction.
<table>
<thead>
<tr>
<th>Trait</th>
<th>Covariate</th>
<th>SNP</th>
<th>Nearest Gene*</th>
<th>CHR</th>
<th>MAF</th>
<th>Ratio</th>
<th>Variance Het</th>
<th>p-value</th>
<th>G-E Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDLC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 11: Top Ten SNPs per Model Ranked by Gene-Environment Interaction

* 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 $-\log_{10}$ 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 $-\log_{10}$ 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 $1E^{-05}$ and the genome-wide significance threshold of $5E^{-08}$, respectively.
<table>
<thead>
<tr>
<th>Trait</th>
<th>Covariate</th>
<th>SNP</th>
<th>Nearest Gene*</th>
<th>CHR</th>
<th>MAF</th>
<th>η Threshold</th>
<th>Ratio</th>
<th>Variance Het p-value</th>
<th>G-E Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>CRPA2</td>
<td>2</td>
<td>0.68</td>
<td>0.01</td>
<td>1.10</td>
<td>2.05E-05</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs3851306</td>
<td>CTNNB1</td>
<td>2</td>
<td>0.49</td>
<td>0.01</td>
<td>1.10</td>
<td>2.18E-02</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs712819</td>
<td>VOPP1</td>
<td>7</td>
<td>0.09</td>
<td>0.001 &amp; 0.01</td>
<td>1.29</td>
<td>2.92E-04</td>
<td>-0.07951</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs1684901</td>
<td>CCDC6</td>
<td>10</td>
<td>0.46</td>
<td>0.001 &amp; 0.01</td>
<td>0.87</td>
<td>1.61E-03</td>
<td>2.59E-03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs2647389</td>
<td>ST6GALNAC3</td>
<td>1</td>
<td>0.15</td>
<td>0.001</td>
<td>1.27</td>
<td>6.16E-05</td>
<td>4.09E-05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs11780061</td>
<td>ZFAT</td>
<td>8</td>
<td>0.11</td>
<td>0.001</td>
<td>0.79</td>
<td>9.28E-04</td>
<td>2.82E-03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs2647389</td>
<td>ST6GALNAC3</td>
<td>1</td>
<td>0.15</td>
<td>0.001</td>
<td>1.27</td>
<td>6.16E-05</td>
<td>4.09E-05</td>
</tr>
</tbody>
</table>

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 $\eta$.

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 $\eta$ dropped from 0.01 to 0.001 and increased 5.4% when $\eta$ dropped from 0.05 to 0.01.
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 $\eta$. The right circle represents the SNPs with LRT $p$-value lower than the VP threshold of $\eta$. SNPs with both LRT and Levene’s test $p$-values lower than the VP threshold of $\eta$ correspond to number in the area where the two circles overlap.

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 $\eta$. The right circle represents the SNPs with LRT $p$-value lower than the VP threshold of $\eta$. SNPs with both LRT and Levene’s test $p$-values lower than the VP threshold of $\eta$ correspond to number in the area where the two circles overlap.
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 $\eta$. The right circle represents the SNPs with LRT $p$-value lower than the VP threshold of $\eta$. SNPs with both LRT and Levene’s test $p$-values lower than the VP threshold of $\eta$ 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.
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.
<table>
<thead>
<tr>
<th>Trait</th>
<th>Covariate</th>
<th>Pearson’s Correlation</th>
<th>LRT $\eta = 0.001$</th>
<th>LRT $\eta = 0.01$</th>
<th>LRT $\eta = 0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Concordant (Perc.)</td>
<td>Discordant</td>
<td>Concordant (Perc.)</td>
</tr>
<tr>
<td>TG</td>
<td>BMI</td>
<td>0.319</td>
<td>293 (93.0%)</td>
<td>22</td>
<td>3634 (85.0%)</td>
</tr>
<tr>
<td></td>
<td>WC</td>
<td>0.323</td>
<td>299 (94.9%)</td>
<td>16</td>
<td>3685 (86.2%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20223 (80.3%)</td>
</tr>
<tr>
<td>HDL-C</td>
<td>BMI</td>
<td>-0.316</td>
<td>326 (91.6%)</td>
<td>30</td>
<td>4193 (87.7%)</td>
</tr>
<tr>
<td></td>
<td>WC</td>
<td>-0.369</td>
<td>343 (96.3%)</td>
<td>13</td>
<td>4365 (91.3%)</td>
</tr>
<tr>
<td></td>
<td>TG</td>
<td>-0.439</td>
<td>340 (95.5%)</td>
<td>16</td>
<td>4497 (94.0%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>23397 (88.6%)</td>
</tr>
<tr>
<td>FG</td>
<td>BMI</td>
<td>0.293</td>
<td>296 (91.6%)</td>
<td>27</td>
<td>3657 (86.6%)</td>
</tr>
<tr>
<td></td>
<td>WC</td>
<td>0.348</td>
<td>306 (94.7%)</td>
<td>17</td>
<td>3837 (90.8%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21326 (85.3%)</td>
</tr>
</tbody>
</table>

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 ($\eta$) of 0.001, 0.01 and 0.05, the total numbers of concordant SNPs and discordant SNPs were recorded for each linear regression model.
#### 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 ($\eta$) of 0.001, 0.01 and 0.05, the total numbers of concordant SNPs and discordant SNPs were recorded for each linear regression model.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Covariate</th>
<th>Pearson's Correlation</th>
<th>LRT $\eta = 0.001$</th>
<th>LRT $\eta = 0.01$</th>
<th>LRT $\eta = 0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Concordant (Perc.)</td>
<td>Discordant</td>
<td>Concordant (Perc.)</td>
<td>Discordant</td>
</tr>
<tr>
<td>Residualized TG</td>
<td>BMI</td>
<td>-0.009</td>
<td>332 (66.9%)</td>
<td>174</td>
<td>2033 (36.3%)</td>
</tr>
<tr>
<td></td>
<td>WC</td>
<td>-0.011</td>
<td>326 (64.3%)</td>
<td>181</td>
<td>2145 (39.5%)</td>
</tr>
<tr>
<td>Residualized HDL-C</td>
<td>BMI</td>
<td>0.0002</td>
<td>145 (39.5%)</td>
<td>222</td>
<td>2088 (44.8%)</td>
</tr>
<tr>
<td></td>
<td>WC</td>
<td>0.001</td>
<td>183 (42.8%)</td>
<td>244</td>
<td>2356 (43.0%)</td>
</tr>
<tr>
<td></td>
<td>TG</td>
<td>-0.008</td>
<td>155 (50.5%)</td>
<td>152</td>
<td>2181 (49.3%)</td>
</tr>
<tr>
<td>Residualized FG</td>
<td>BMI</td>
<td>0.006</td>
<td>183 (47.3%)</td>
<td>204</td>
<td>2047 (43.8%)</td>
</tr>
<tr>
<td></td>
<td>WC</td>
<td>0.006</td>
<td>265 (62.6%)</td>
<td>158</td>
<td>2078 (41.9%)</td>
</tr>
</tbody>
</table>
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.
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 when SNPs were prioritized at LRT p-value thresholds of 0.001 (A), 0.01 (C), 0.05 (M), and 0.1 (V). 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

```
#### PLINK SCRIPT TO TEST FOR GENE-ENVIRONMENT ####

### 1. Generate Ethnic Specific tped
for filename in Caucasian Asian African Hispanic
do
  plink --noweb \
    --remove ./Pheno/Diabetic/$filename_* \
    --bfile $bfile \
    --keep ./TransPheno/$filename_excluded* \
    --maf 0.05 \
    --hwe 1e-6 \
    --mind 0.05 \
    --geno 0.05 \
    --recode \
    --transpose \
    --out ./Geno/$filename_CLEAN_GENO

### 2. Performs Sex check
plink --bfile $bfile \
  --keep $pheno_caucasian \
  --check-sex --out $QCdir/Report_SexCheck

gawk '5!="OK" print $1, $2' Report_SexCheck.sexcheck > Report_Sex_discrepancy
# 2 IIDs with mismatched sex removed ##

### 3. Run Interactions
plink --noweb \
  --bfile $Cbfile \
  --pheno $Caucasian_Pheno \
  --pheno=name trig1_logw \
  --linear \
  --covar $Caucasian_Pheno \
  --covar=name age1c,gender1,waistcm1_logw \
  --interaction \
  --parameters 1,2,3,4,7 \
  --standard-beta \
  --missing-phenotype NA \
  --out $outdir/Interaction/trig_logw_x_waistcm1_logw
```
Listing B.2: Likelihood ratio test

```
LRRatio <- function(variance, obs, verbose = F){
  ### check inputs
  if (dim(variance)[2] != 3)
    stop("number of rows of variance should be 3")
  if (dim(obs)[2] != 3)
    stop("number of rows of observed genotype counts should be 3")
  if (dim(obs)[2] != dim(variance)[2])
    stop("observed genotype counts and variance should have the same number of rows")
  n1 <- obs[,1]
  n2 <- obs[,2]
  n3 <- obs[,3]
  N <- rowSums(obs, na.rm = T)
  s1 <- variance[,1]
  s2 <- variance[,2]
  s3 <- variance[,3]

  ## constants:
  log_A <- -lgamma(n1/2 - 1/2) + lgamma(n2/2 - 1/2)
  + lgamma(n3/2 - 1/2) + log(2) * (n1/2 - 1/2) + log(2) * (n2/2 - 1/2) + log(2) * (n3/2 - 1/2)
  C <- (n1 - n3) / (N - 9)

  ########### NULL MLE
  opt_theo <- ((n1 - 1) * s1 + (n2 - 1) * s2 + (n3 - 1) * s3) / (N - 9)
  likelihood.null <- -log_A + (n1/2 - 3/2) * log(n1 - 1) +
    log(s1) - log(opt_theo) + (n2 - 2 - 3/2) * log(n2 - 1) + log(s2) -
    log(opt_theo) + (n3 - 2 - 3/2) * log(n3 - 1) + log(s3) - log(opt_theo) -
    (N - 9) / 2

  ########### ALTERNATIVE
  opt_ratio <- ((n2 - 1) * s2 * C + sqrt(((n2 - 1) * s2 * C) - 2 *
    4 * (n1 - 1) * s1 * s3 * (1 - C)) / (2 * (n1 - 1) * s1) * (1 - C))
  opt_variance <- ((n1 - 1) * s1 * opt_ratio + (n2 - 1) * s2 + (n3 - 1) *
    s3 / opt_ratio) / (N - 9)
  likelihood.alter <- (-log_A + (n1/2 - 3/2) * log(n1 - 1) + log(s1) -
    log(opt_variance) + log(opt_ratio)) + (n2 - 2 - 3/2) * log(n2 - 1) +
    log(s2) - log(opt_variance) + (n3 - 2 - 3/2) * log(n3 - 1) + log(s3) -
    log(opt_variance) - log(opt_ratio) - (N - 9) / 2

  ##### p_values
  MLR <- 2 * likelihood.alter - 2 * likelihood.null
  pval <- pchisq(MLR, 1, lower.tail = F)
  if (verbose == F) return(list("LRT_pvalue" = pval))
  else return(list("Test_Statistics" = MLR, "RatioPar" = opt_ratio))
}
```

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

```
# Calculate Simulation Parameters
b2 <- c(0.35, 0.5)
b3 <- c(0.05, 0.08)
p <- c(0.1, 0.2, 0.4)
calc_ve <- function(input){
b2 <- input[1]
b3 <- input[2]
p <- input[3]
ve_e <- (b2^2)/(b2^2 + 2*p*(1-p)*b3^2 + 1)
ve_ge <- (2*p*(1-p)*b3^2)/(2*p*(1-p)*b3^2 + b2^2 + 1)
return(c(ve_e, ve_ge)*100)
}
datamat <- data.frame(rep(b2,each=6),
                      rep(rep(b3,each=3),2),
                      rep(rep(p, each=1),4))

# Function to Calculate the probability of MAF
genotype.gen <- function(maf,n){
n0 <- round(n*(1-maf)^2)
n1 <- round(n*maf*(1-maf)*2)
n2 <- n-n0-n1
c(n0,n1,n2)}

### Step 1. Set Simulation Parameters
K <- 5000
M <- 500000
datamat <- data.frame(rep(b2,each=6),
                      rep(rep(b3,each=3),2),
                      rep(rep(p, each=1),4))
N <- c(2000, 5000, 10000, 20000)

### Step 2. Set Output Lists
Outputlist <- list()
for (k in 1:length(N)){
k <- 1
n <- N[k]
Outputlist[[k]] <- list()
for (t in 1:dim(datamat)[1]){p <- datamat[t, 3]
b2 <- datamat[t, 1]
b3 <- datamat[t, 2]
```
b1 <- 0

#################### Step 3. Simulate Data ####################
## simulate genotype and covariate independently
Genotype <- sample(rep(0:2, genotype.gen(p, n)))
COV <- rnorm(n)

## Test P-values
test_pvalues <- data.frame(levene_p = NA, levene_t1 = NA,
                           levene_t2 = NA, bart = NA, intp = NA)
mafvar <- data.frame(major = NA, heter = NA, minor = NA)

for (j in 1:K) {
  error <- rnorm(n)
  Geno <- sample(Genotype)
  trait <- b2 * COV + b3 * COV * Geno + error
  mafvar[j,] <- c(var(trait[Geno == 0]), var(trait[Geno == 1]),
    var(trait[Geno == 2]))
  interaction_pval <- summary(lm(trait ~ COV * Geno))$coef[4,4]

  valid_cases <- complete.cases(trait, as.factor(Geno))
  mean_per_geno <- tapply(trait[valid_cases], as.factor(Geno)[valid_cases], mean)
  resp <- abs(trait - mean_per_geno[as.factor(Geno)])
  levene_pval <- anova(lm(resp ~ as.factor(Geno)))[1, 3]

  lvt1_pval <- summary(lm(resp ~ Geno))$coef[2, 4]
  lvt2_pval <- summary(lm(resp ~ sqrt(Geno)))$coef[2, 4]

  bart <- bartlett.test(trait ~ Geno)$p.value
  test_pvalues[j,] <- c(levene_pval, lvt1_pval, lvt2_pval,
    pv_pval, bart, interaction_pval)
}

LRTratio(mafvar, obs = data.frame(rep(genotype.gen(p, n)[1], K),
  rep(genotype.gen(p, n)[2], K), rep(genotype.gen(p, n)[3], K)))$LRT

Outputlist[[k]][[t]] <- list(mafvar, test_pvalues)
}

save.image("POWERsim_K1.rdata")
gewistLRT <- function(p, N, obs=NA, theta_gc, theta_c, M, K = 20000,verbose=FALSE){
  ## Input the known parameters (variance explained ~ (0,1))
  ## Assume Gene x Environment
  ### check inputs
  if (length(obs) == 3) {
    n1 <- obs[1]
    n2 <- obs[2]
    n3 <- obs[3]
    N <- sum(n1, n2, n3, na.rm=TRUE)
  }else {
    if (p > 0.5 | p <= 0)
      stop("minor allele frequency should be a number between 0 and 0.5")
    if (!(N > 0 | M > 0 | K > 0 | theta_gc >0 | theta_c > 0))
      stop("negative input values are not allowed")
    N <- round(N)
    n1 <- round(N * (1 - p)^2)
    n2 <- round(N * (1 - p)*p*2)
    n3 <- N - n1 - n2
  }
  if (!(theta_gc < 1 | theta_c < 1 ))
    stop("variance explained should be a number between 0 and 1 ")
  M <- round(M)
  K <- round(K)
  ### calculate beta coefficients
  b2 <- sqrt(( theta_c )/( 1 - theta_gc - theta_c ))
  b3 <- sqrt(theta_gc/(2*p*(1 - p)*( 1 - theta_gc - theta_c )))
  # sample the variance of covariate and error per genotype
  var_C_G_1 <- rchisq(K,n1 - 1)/(n1 - 1)
  var_C_G_2 <- rchisq(K,n2 - 1)/(n2 - 1)
  var_C_G_3 <- rchisq(K,n3 - 1)/(n3 - 1)
  error_G_1 <- rchisq(K,n1 - 1)/(n1 - 1)
  error_G_2 <- rchisq(K,n2 - 1)/(n2 - 1)
  error_G_3 <- rchisq(K,n3 - 1)/(n3 - 1)
  # sample the b_x's
  b_x1 <- rnorm(K,b2 + b3*(- 2*p),sqrt(1/(var_C_G_1*(n1 - 1))))
  b_x2 <- rnorm(K,b2 + b3*(1 - 2*p),sqrt(1/(var_C_G_2*(n2 - 1))))
  b_x3 <- rnorm(K,b2 + b3*(2 - 2*p),sqrt(1/(var_C_G_3*(n3 - 1))))
}

Listing B.4: Statistical Power of VP with Five Variance Heterogeneity Tests
the covariance of error term and covariate per genotype

\[
\text{cov_error_C_1} <- (b_x1 - (b_2 + b_3*( -2*p))) \times \text{var_C_G_1}
\]
\[
\text{cov_error_C_2} <- (b_x2 - (b_2 + b_3*(1 - 2*p))) \times \text{var_C_G_2}
\]
\[
\text{cov_error_C_3} <- (b_x3 - (b_2 + b_3*(2 - 2*p))) \times \text{var_C_G_3}
\]

\[
\text{sum_of_C_square_G} <- \text{var_C_G_1} + 2*(\text{cov_error_C_1} + \text{cov_error_C_1}) + \text{var_C_G_2} + 2*(\text{cov_error_C_1} + \text{cov_error_C_2}) + \text{var_C_G_3} + 2*(\text{cov_error_C_1} + \text{cov_error_C_2} + \text{cov_error_C_3})
\]

\[
\text{total_cov} <- \text{var_C_G_1} + \text{var_C_G_2} + \text{var_C_G_3}
\]

\[
\text{total_cov_error} <- (\text{total_cov}^2 - 2*\text{var_C_G_1} * \text{var_C_G_1} - 2*\text{var_C_G_2} * \text{var_C_G_2} - 2*\text{var_C_G_3} * \text{var_C_G_3})
\]

\[
\text{lrt_p} <- \text{LRTratio(data.frame(var_group_1, var_group_2, var_group_3)}, \quad \text{obs = data.frame(rep(n1, length(var_group_1)), rep(n2, length(var_group_2)), rep(n3, length(var_group_3))), verbose=F})$LRT
\]

\[
\text{RSS1} <- -2*\text{beta2}^2 + 2*\text{var_C_G_1} - 2*\text{beta3}^2 + 2*\text{var_C_G_1} - 2*\text{beta4}^2 + 2*\text{var_C_G_1}
\]
\[
\text{RSS2} <- -2*\text{beta2}^2 + 2*\text{var_C_G_2} - 2*\text{beta3}^2 + 2*\text{var_C_G_2} - 2*\text{beta4}^2 + 2*\text{var_C_G_2}
\]
\[
\text{RSS3} <- -2*\text{beta2}^2 + 2*\text{var_C_G_3} - 2*\text{beta3}^2 + 2*\text{var_C_G_3} - 2*\text{beta4}^2 + 2*\text{var_C_G_3}
\]

\[
\text{f_stats} <- 2*\text{RSS2}^2*2p*(1 - p)*(N - 4)^2/(\text{RSS1}^2*(n1 - 1) + \text{RSS2}*(n2 - 1) + \text{RSS3}*(n3 - 1))
\]
interaction_p <- 1 - pf(f_stats, df1 = 1, df2 = N - 4)

##### Calculate powers
result <- data.frame("p-value_Cut-offs" = NA, "VP_power" = NA)
for (i in 1:1000){
  power <- mean(lrt_p < i/1000 & interaction_p<0.05/(M*i/1000))
  result[i,] <- c(i/1000, power)
}
conv_power <- power
optimal_power <- max(result[,2])
optimal_p_threshold <- which.max(result[,2])/1000
if (verbose){return(result)
} else {
  return(list("Conventional_power"=conv_power,
               "Optimal_VP_power"=optimal_power,
               "Optimal_pval_threshold"=optimal_p_threshold))
}
##### End of Script

f2.power <- function(theta_ge, theta_e, N, power, alpha=0.05){
  f2 <- theta_ge/(1-theta_ge-theta_e)*N
  power <- 1 + pf(qf(1-alpha, df1=1, df2=N-4), df1=1, df2=N-4, ncp=f2)
}
### model 1 has 4 parameters including intercept
### model 2 has 3 parameters including intercept
### partial F test statistics df1 = 1, df2 = N-4

calc_beta <- function(input){
  theta_c <- input[1]
  theta_gc <- input[2]
  p <- input[3]
  b2 <- sqrt(theta_c/(1-theta_c-theta_gc))
  b3 <- sqrt(theta_gc/(1-theta_c-theta_gc)/(2*p*(1-p)))
}
#####

VP_Power <- function(n, effect.size, p.cut=0.05, power, cov_exp, MAF, var.method, m=500000){
  f2 <- effect.size/(1-effect.size-cov_exp)
  NCP <- (effect.size)/(1-effect.size)*n
  if (var.method=="LRT"){
    optimal.p <- gewistLRT(MAF, n, theta_gc = effect.size,
                           theta_c = cov_exp, M = m)$Optimal_pval_threshold
  } else if (var.method=="Levene") {
    optimal.p <- gewistLevene(MAF, n, theta_gc = effect.size,
                               theta_c = cov_exp, M = m)$Optimal_pval_threshold
  } else {optimal.p=1
}
}
\[ \text{power} = 1 + \frac{1}{\text{pf}([0, \text{optimal} \times \text{m}], \text{df1} = 1, \text{df2} = n - 2), \text{df1} = 1, \text{df2} = n - 4, \text{ncp} = \text{NCP}} \]

\[
\text{int}\_\text{exp} < - \text{seq}(0.05, 1, \text{by} = 0.05)/100
\]

\[
\text{Conditions} < - \text{list}([0.05, 0.05], [0.05, 0.2], [0.05, 0.4],
\text{[0.1, 0.05], [0.1, 0.2], [0.1, 0.4],
\text{[0.2, 0.05], [0.2, 0.2], [0.2, 0.4],
\text{[0.25, 0.05], [0.25, 0.2], [0.25, 0.4]])}
\]

\[
\text{sample.size.list} < - \text{list}()
\]

\[
\text{for} (j \text{ in} 1: \text{length}(\text{Conditions})){
\text{MAF} < - \text{Conditions}[j][2]
\text{COV} < - \text{Conditions}[j][1]
\text{sample.size} < - \text{data.frame}("\text{Interaction}" = NA, "\text{VP}\_\text{Lv}" = NA, "\text{VP}\_\text{LRT}" = NA)
\text{for} (i \text{ in} 1: \text{length}(\text{int}\_\text{exp})){
\text{int} < - \text{uniroot}(\text{VP}\_\text{Power}, \text{c}(1000, 10e10), \text{effect.size} = \text{int}\_\text{exp}[i],
\text{p.cut} = 0.05, \text{power} = 0.8, \text{cov}\_\text{exp} = \text{COV}, \text{var.method} = "\text{None}")$\text{root}$
\text{LVint} < - \text{uniroot}(\text{VP}\_\text{Power}, \text{c}(1000, 10e10), \text{effect.size} = \text{int}\_\text{exp}[i],
\text{p.cut} = 0.05, \text{power} = 0.8, \text{cov}\_\text{exp} = \text{COV}, \text{MAF} = \text{MAF}, \text{var.method} = "\text{Levene}")$\text{root}$
\text{LRTint} < - \text{uniroot}(\text{VP}\_\text{Power}, \text{c}(1000, 10e10), \text{effect.size} = \text{int}\_\text{exp}[i],
\text{p.cut} = 0.05, \text{power} = 0.8, \text{cov}\_\text{exp} = \text{COV}, \text{MAF} = \text{MAF}, \text{var.method} = "\text{LRT}")$\text{root}$
\text{sample.size[i,]} < - \text{c}(\text{int}, \text{LVint}, \text{LRTint})
\text{sample.size.list}[j] < - \text{sample.size}
\}
\]

\[
\text{setwd}(\text{"/home/dengw/LRT/Thesis/Simulations"})
\]

\[
\text{save.image}(\text{"SAMPLESIZE.rdata"})
\]
Listing B.5: PLINK R plugin Script to Produce Summary Statistics for LRT and Levene’s Test

```r
######## PLINK R plugin Function for Variance Het P−values ####
Rplink <- function(PHENO,GENO,CLUSTER,COVAR)
{
  f1 <- function(x)
  {
    ######### LRT−Variance Calculation
    PHENO <- (PHENO−mean(PHENO, na.rm=T))/sd(PHENO,na.rm=T)
    mafvar <- c(var(PHENO|x==0,na.rm=T), var(PHENO|x==1,na.rm=T),var(PHENO|x==2,na.rm=T))
    geno.counts <- c(sum(x==0, na.rm=T), sum(x==1, na.rm=T),sum(x==2, na.rm=T))
    valid_cases <- complete.cases(PHENO, as.factor(x))
    mean_per_geno <- tapply(PHENO[valid_cases], as.factor(x)[valid_cases],mean)
    responses <- abs(PHENO−mean_per_geno[as.factor(x)])
    z_avg <- c(mean(responses|x==0,na.rm=T),
                mean(responses|x==1,na.rm=T),
                mean(responses|x==2,na.rm=T))
    z_var <- c(var(responses|x==0,na.rm=T),
               var(responses|x==1,na.rm=T),
               var(responses|x==2,na.rm=T))
    r <- c(mafvar, geno.counts, z_avg, z_var)
    c( length(r) , r )
  }
apply(GENO, 2 , f1)
}
```
Listing B.6: R Script to Produce Variance Prioritization Results

```r
##### Make Figure Prioritiation #######

traits <- c("trig1_logw", "hdl1_logw", "glucos1c_logw")
VarHet_files <- c("trig1_logwVarHet_LeveneLRT_SAVE.txt",
                 "hdl1_logwVarHet_LeveneLRT_SAVE.txt",
                 "glucos1c_logwVarHet_LeveneLRT_SAVE.txt")

interaction_traits <- list(c("trig_logw_x_bmi1c_logw", "trig_logw_x_waistcm1_logw"),
                           c("hdl1_logw_x_bmi1c_logw", "hdl1_logw_x_waistcm1_logw",
                             "hdl1_logw_x_trig1_logw"),
                           c("glucos1c_logw_x_bmi1c_logw", "glucos1c_logw_x_waistcm1_logw"))

p_threshold=c(0.001, 0.01, 0.05, 0.1)

# Original Analysis
for (t in 1:length(traits)){
  trait <- traits[t]
  subset_analyzed_var <- read.table(paste(datadir, VarHet_files[t], sep=""), head=T)

  tiff(paste(trait, "VarPLRT_IntP_QQplot.tiff", sep=""),
       width = 1000*length(interaction_traits[[t]]), height = 2000,
       units = "px", res = 200, compression = "lzw")
  par(mfcol=c(4, length(interaction_traits[[t]])*2), mar=c(4,4,3,1), oma=c(0,0,1,0))

  for (c in 1:length(interaction_traits[[t]])){
    #### READ IN DATA FROM PLINK ####
    subset_int_data <- read.table(paste(datadir, interaction_traits[[t]][c],
                                   ".assoc.linear_InteractionP_SAVE.txt", sep=""), head=F)
    names(subset_int_data) <- c("CHR", "SNP", "BP", "N", "BETA", "TEST", "P")

    # MERGE VARIANCE Het output and INTERACTION output
    vp_data <- merge(subset_int_data, subset_analyzed_var,
                     by.x="SNP", by.y="MarkerName", sort=F)

    for (p in 1:length(p_threshold)){
      vp <- subset(vp_data, vp_data$LRT < p_threshold[p])
      nonvp <- subset(vp_data, vp_data$LRT >= p_threshold[p])
      produce_qqplot_fun(vp, which.p=which(names(vp_data)=="P"))
      abline(h=-log10(0.05/dim(vp)[1]), line=1)
      title(paste(fig_LET[[t]][p+8*(c-1)], ""), line=1)
    }
  }
}
```

rm(list=ls())

traits <- c("trig1_logw", "hdl1_logw", "glucos1c_logw")
p_threshold=c(0.001, 0.01, 0.05, 0.1)

VarHet_adj_files <- list(c("trig_logw_adj_bmi1c_logwVarHet_LeveneLRT_SAVE.txt", 
                       "trig_logw_adj_waistcm1_logwVarHet_LeveneLRT_SAVE.txt"),
                       c("hdl1_logw_adj_bmi1c_logwVarHet_LeveneLRT_SAVE.txt", 
                         "hdl1_logw_adj_waistcm1_logwVarHet_LeveneLRT_SAVE.txt", 
                         "hdl1_logw_adj_trig1_logwVarHet_LeveneLRT_SAVE.txt"),
                       c("glucos1c_logw_adj_bmi1c_logwVarHet_LeveneLRT_SAVE.txt", 
                         "glucos1c_logw_adj_waistcm1_logwVarHet_LeveneLRT_SAVE.txt"))

adjusted_traits <- list(c("trig_logw_adj_bmi1c_logwInteractionP_SAVE.txt", 
                           "trig_logw_adj_waistcm1_logwInteractionP_SAVE.txt"),
                        c("hdl1_logw_adj_bmi1c_logwInteractionP_SAVE.txt", 
                          "hdl1_logw_adj_waistcm1_logwInteractionP_SAVE.txt", 
                          "hdl1_logw_adj_trig1_logwInteractionP_SAVE.txt"),
                        c("glucos1c_logw_adj_bmi1c_logwInteractionP_SAVE.txt", 
                          "glucos1c_logw_adj_waistcm1_logwInteractionP_SAVE.txt"))

for (t in 1:length(traits)){
  trait <- traits[t]
  tiff(paste(trait, "_VarPLRT_IntP_Adjusted_QQplot.tiff", sep=""),
       width = 1000+length(adjusted_traits[[t]]), height = 2000,
       units = "px", res = 200, compression = "lzw")
  par(mfcol=c(4, length(adjusted_traits[[t]])*2), mar=c(4,4,3,1), oma=c(0,0,1,0))
  for (c in 1:length(adjusted_traits[[t]])){
    subset_analyzed_var_adj <- read.table(paste(data_dir2, VarHet_adj_files[[t]][[c]], sep=""), head=T)
    subset_int_data_adj <- read.table(paste(data_dir2, adjusted_traits[[t]][[c]], sep=""), head=T)
    vp_data <- merge(subset_int_data_adj,subset_analyzed_var_adj,
                      by.x="SNP", by.y="MarkerName", sort=F)
    for (p in 1:length(p_threshold)){
      vp <- subset(vp_data, vp_data$LRT < p_threshold[p])
      #nonvp <- subset(vp_data, vp_data$LRT >= p_threshold[p])
      produce_qqplot_fun(vp, which.p=which(names(vp_data)=="P"))
      abline(h=-log10(0.05/dim(vp)[1]))
      title(paste(fig_LET[[t]][p+8*(c-1)], ""), line=1)
    }
  }
}

#vp <- subset(vp_data, vp_data$LRT < p_threshold[p])
nonvp <- subset(vp_data, vp_data$LRT >= p_threshold[p])
produce_qqplot.fun(nonvp, which.p=which(names(vp_data)=="P"))
abline(h=-log10(0.05/dim(nonvp)[1]))
title(paste(fig_LET[[t]][[p]+8*(c-1)+4], ", line=1))
}
}
dev.off()
}

produce_qqplot.fun <- function(data, which.p, titles=NULL){
pval_data <- data[order(data[,which.p], decreasing=T, na.last=NA),]
pval_data$pvalref <- ((dim(pval_data)[1]:1)/dim(pval_data)[1])
pval <- pval_data[,which.p]
if (length(pval) < 100000){
pval_qq11 <- pval
ordering <- length(pval)+1 - 1:length(pval)
pval_ref <- pval_data$pvalref
}else{
pval_qq11 <- pval[c(seq(1, sum(pval > 0.01, na.rm=T), 100),
seq(sum(pval > 0.01,na.rm=T),sum(pval > 0.001,na.rm=T),10),
sum(pval> 0.001,na.rm=T):length(pval))]
ordering <- length(pval)+1 - c(seq(1, sum(pval > 0.01, na.rm=T), 100),
seq(sum(pval > 0.01,na.rm=T),sum(pval > 0.001,na.rm=T),10),
sum(pval> 0.001,na.rm=T):length(pval))
pval_ref <- pval_data$pvalref[c(seq(1, sum(pval > 0.01, na.rm=T), 100),
seq(sum(pval > 0.01,na.rm=T),sum(pval > 0.001,na.rm=T),10),
sum(pval> 0.001,na.rm=T):length(pval))]
}
c05 <- NA
c95 <- NA
for(i in 1:length(ordering)){
c95[i] <- qbeta(0.95,ordering[i],length(pval)-ordering[i]+1)
c05[i] <- qbeta(0.05,ordering[i],length(pval)-ordering[i]+1)
}
plot(-log10(pval_ref), -log10(pval_qq11), xlim=c(0,max(-log10(pval_ref))),
ylim=c(0,7.5), col="blue", xlab=NA, ylab=NA)
title(xlab=expression(Expected~~-log[10](italic(p))),
ylab=expression(Observed~~-log[10](italic(p))),main=titles, cex.lab=1, cex.axis=1)
polygon(x=c(-log10(pval_ref), rev(-log10(pval_ref))),
y=c(-log10(c95), rev(-log10(c05))), col="lightgrey", border=NA)
points(-log10(pval_ref), -log10(pval_qq11), pty=1, col="blue")
abline(0,1,col=2)
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:

\[ \sigma_0^2 = \beta_2^2 + 1 \]  \hspace{1cm} (C.1)

\[ \sigma_1^2 = (\beta_2 + \beta_3)^2 + 1 \]  \hspace{1cm} (C.2)

\[ \sigma_2^2 = (\beta_2 + 2\beta_3)^2 + 1 \]  \hspace{1cm} (C.3)
Increasing variances with the number of minor alleles, or $\sigma_0^2 < \sigma_1^2 < \sigma_2^2$ implies:

\[
\begin{align*}
\beta_2^2 + 1 &< (\beta_2 + \beta_3)^2 + 1 < (\beta_2 + 2\beta_3)^2 + 1 \iff \\
\beta_2^2 &< \beta_2^2 + 2\beta_2\beta_3 + \beta_3^2 < \beta_2^2 + 4\beta_2\beta_3 + 4\beta_3^2 \iff \\
0 &< 2\beta_2\beta_3 + \beta_3^2 < 4\beta_2\beta_3 + 4\beta_3^2 \iff \\
(2\beta_2 + \beta_3)\beta_3 &> 0 \quad \& \quad (2\beta_2 + 4\beta_3)\beta_3 > 0 \iff \\
\beta_3 &> 0 \quad \& \quad 2\beta_2 + \beta_3 > 0 \quad \& \quad 2\beta_2 + 3\beta_3 > 0 \quad \text{or} \quad \beta_3 < 0 \quad \& \quad 2\beta_2 + \beta_3 < 0 \quad \& \quad 2\beta_2 + 3\beta_3 < 0 \iff \\
\beta_3 &> 0 \quad \& \quad \beta_2 > -\frac{\beta_3}{2} \quad \& \quad \frac{\beta_2}{2} > -\frac{3\beta_3}{2} > 0 \quad \text{or} \quad \beta_3 < 0 \quad \& \quad \beta_2 < -\frac{\beta_3}{2} \quad \& \quad \frac{\beta_2}{2} < -\frac{3\beta_3}{2} \iff \\
\beta_3 &> 0 \quad \& \quad \beta_2 > -\frac{\beta_3}{2} > -\frac{3\beta_3}{2} > 0 \quad \text{or} \quad \beta_3 < 0 \quad \& \quad \beta_2 < -\frac{\beta_3}{2} < -\frac{3\beta_3}{2} \iff \\
|\beta_2| &> \left|\frac{3\beta_3}{2}\right| \quad \& \quad \beta_2\beta_3 < 0
\end{align*}
\]

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:

\[
|r_2 - r_1| = \left| 1 + \frac{2\beta_2 \beta_3 + 3 \beta_3^2}{(\beta_2 + \beta_3)^2 + 1} - \left(1 + \frac{2\beta_2 \beta_3 + \beta_3^2}{\beta_2^2 + 1}\right) \right|
\]

\[
= \left| \frac{2\beta_2 \beta_3 + 3 \beta_3^2}{(\beta_2 + \beta_3)^2 + 1} - \frac{2\beta_2 \beta_3 + \beta_3^2}{\beta_2^2 + 1} \right|
\]

\[
= \left| \frac{(2\beta_2 \beta_3 + 3 \beta_3^2)(\beta_2^2 + 1) - (2\beta_2 \beta_3 + \beta_3^2)((\beta_2 + \beta_3)^2 + 1)}{(\beta_2^2 + 1)((\beta_2 + \beta_3)^2 + 1)} \right|
\]

\[
= \left| \frac{(2\beta_2 \beta_3 + 3 \beta_3^2)(\beta_2^2 + 1) - (\beta_2 + \beta_3)^2 + 1 + 2\beta_3^2(\beta_2 + 1)}{(\beta_2^2 + 1)((\beta_2 + \beta_3)^2 + 1)} \right|
\]

\[
= \left| \frac{-2(\beta_2 \beta_3 + \beta_3^2)^2 + 2\beta_3^2(\beta_2^2 + 1)}{(\beta_2^2 + 1)((\beta_2 + \beta_3)^2 + 1)} \right|
\]

\[
= \left| \frac{2\beta_3^2}{((\beta_2 + \beta_3)^2 + 1)} \right| \frac{1 - (2\beta_2 + \beta_3)^2}{2(\beta_2^2 + 1)}
\]

(C.4)

And the difference between the two differences is:

\[
|d_2 - d_1| = 2\beta_3^2 > 0
\]

(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 \(1 - \frac{(2\beta_2 + \beta_3)^2}{2(\beta_2^2 + 1)}\) 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;

\[
|d_2 - d_1| = 2\beta_3^2 > |r_2 - r_1|
\]

(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 $(\beta_2 + \beta_3)^2 + 2$.

\[
(\beta_2 + \beta_3)^2 + 2 - \frac{(2\beta_2 + \beta_3)^2}{2(\beta_2^2 + 1)} = (\beta_2 + \beta_3)^2 + 2 - \frac{4\beta_2^2 + 4\beta_2\beta_3 + \beta_3^2}{2(\beta_2^2 + 1)} \\
= (\beta_2 + \beta_3)^2 + 2 - \frac{4\beta_2^2 + 4 - 4 + 4\beta_2\beta_3 + \beta_3^2}{2(\beta_2^2 + 1)} \\
= (\beta_2 + \beta_3)^2 - \frac{4\beta_2\beta_3 + \beta_3^2 - 4}{2(\beta_2^2 + 1)} \\
= \frac{2(\beta_2^2 + 1)(\beta_2 + \beta_3)^2}{2(\beta_2^2 + 1)} - \frac{4\beta_2\beta_3 + \beta_3^2 - 4}{2(\beta_2^2 + 1)} \\
= \frac{2\beta_2^2\beta_3^2 + 2\beta_3^2 + 4\beta_2^3\beta_3 + 4\beta_2^2\beta_3 + 2\beta_2^4 + 2\beta_2^2 - 4\beta_2\beta_3 - \beta_3^2 + 4}{2(\beta_2^2 + 1)} \\
= \frac{2\beta_2^2\beta_3^2 + 2\beta_3^2 + 4\beta_2^3\beta_3 + 2\beta_2^4 + 2\beta_2^2 + 4}{2(\beta_2^2 + 1)} \\
= \frac{2\beta_2^2(\beta_3 + \beta_2)^2 + \beta_3^2 + 2\beta_2^2 + 4}{2(\beta_2^2 + 1)} > 0 \quad \text{(C.7)}
\]

C.3 Second Partial Derivative Test

The second order partial derivatives of the log-likelihood function under the alternative model are:
\[
\frac{d^2 l_1}{d(\sigma_1^2)^2} = \frac{N - 9}{2(\sigma_1^2)^2} - \frac{(N_0 - 1)s_0^2r + (N_1 - 1)s_1^2 + (N_2 - 1)s_2^2/r}{(\sigma_1^2)^3}, \tag{C.8}
\]

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

\[
\frac{d^2 l_1}{dr^2} = \frac{N_0 - N_2}{2} \left(- \frac{1}{r^2}\right) - \frac{(N_2 - 1)s_2^2/r^3}{(\sigma_1^2)^2} < 0, \tag{C.10}
\]

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

The second order derivative with respect to \(\sigma_1^2\) evaluated at the MLEs, \(\hat{r}\) and \(\hat{\sigma}_1^2\), is negative:

\[
\frac{d^2 l_1}{d(\hat{r}, \hat{\sigma}_1^2)^2} = \frac{N - 9}{2(\hat{\sigma}_1^2)^2} - \frac{(N_0 - 1)s_0^2\hat{r} + (N_1 - 1)s_1^2 + (N_2 - 1)s_2^2/\hat{r}}{(\hat{\sigma}_1^2)^3} = -\frac{N - 9}{2(\hat{\sigma}_1^2)^2} < 0 \tag{C.12}
\]

The determinant of the Hessian matrix evaluated at \(\hat{r}\) and \(\hat{\sigma}_1^2\), is then:

\[
D(\hat{r}, \hat{\sigma}_1^2) = \text{det}(H((\hat{r}, \hat{\sigma}_1^2)))
\]

\[
= \frac{d^2 l_1}{d(\hat{\sigma}_1^2)^2} \left. \frac{d^2 l_1}{d\hat{r} d\hat{\sigma}_1^2} (\hat{r}, \hat{\sigma}_1^2) \right|^2 - \frac{d^2 l_1}{d\hat{r} d\hat{\sigma}_1^2} (\hat{r}, \hat{\sigma}_1^2)^2
\]

\[
= -\frac{N - 9}{2(\hat{\sigma}_1^2)^2} \left(\frac{N_0 - N_2}{2} \left(- \frac{1}{\hat{r}^2}\right) - \frac{(N_2 - 1)s_2^2/\hat{r}^3}{(\hat{\sigma}_1^2)^2}\right) - \frac{(N_0 - 1)s_0^2 + (N_2 - 1)s_2^2/\hat{r}^2}{2(\hat{\sigma}_1^2)^2}
\]
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 (\( \sigma_i^2 \)) 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).
<table>
<thead>
<tr>
<th>MAF ($p$)</th>
<th>N</th>
<th>$r = 0.8$</th>
<th>$r = 0.9$</th>
<th>$r = 1$</th>
<th>$r = 1.1$</th>
<th>$r = 1.2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$(\hat{r}, \hat{\sigma}_1^2)$</td>
<td>Hessian Det</td>
<td>$(\hat{r}, \hat{\sigma}_1^2)$</td>
<td>Hessian Det</td>
<td>$(\hat{r}, \hat{\sigma}_1^2)$</td>
</tr>
<tr>
<td>0.05</td>
<td>2000</td>
<td>(0.82,1.03)</td>
<td>137938.95</td>
<td>(0.93,1.03)</td>
<td>108989.45</td>
<td>(1.03, 1.03)</td>
</tr>
<tr>
<td>0.05</td>
<td>5000</td>
<td>(0.81,1.01)</td>
<td>904563.76</td>
<td>(0.91,1.01)</td>
<td>714690.25</td>
<td>(1.01,1.01)</td>
</tr>
<tr>
<td>0.05</td>
<td>10000</td>
<td>(0.80,1.01)</td>
<td>3657255.74</td>
<td>(0.91,1.01)</td>
<td>2899665.07</td>
<td>(1.01,1.01)</td>
</tr>
<tr>
<td>0.05</td>
<td>20000</td>
<td>(0.80,1.00)</td>
<td>14736155.49</td>
<td>(0.90,1.00)</td>
<td>11643298.31</td>
<td>(1.00,1.00)</td>
</tr>
<tr>
<td>0.1</td>
<td>2000</td>
<td>(0.81, 1.01)</td>
<td>271449.00</td>
<td>(0.91, 1.01)</td>
<td>214478.17</td>
<td>(1.01, 1.01)</td>
</tr>
<tr>
<td>0.1</td>
<td>5000</td>
<td>(0.80, 1.01)</td>
<td>1733110.52</td>
<td>(0.90, 1.01)</td>
<td>1369370.77</td>
<td>(1.01, 1.01)</td>
</tr>
<tr>
<td>0.1</td>
<td>10000</td>
<td>(0.80, 1.00)</td>
<td>6981721.42</td>
<td>(0.90, 1.00)</td>
<td>5516419.79</td>
<td>(1.00, 1.00)</td>
</tr>
<tr>
<td>0.1</td>
<td>20000</td>
<td>(0.80, 1.00)</td>
<td>28025852.07</td>
<td>(0.90, 1.00)</td>
<td>22143874.88</td>
<td>(1.00, 1.00)</td>
</tr>
<tr>
<td>0.2</td>
<td>2000</td>
<td>(0.80, 1.01)</td>
<td>491710.54</td>
<td>(0.91, 1.01)</td>
<td>385515.34</td>
<td>(1.01, 1.01)</td>
</tr>
<tr>
<td>0.2</td>
<td>5000</td>
<td>(0.80, 1.00)</td>
<td>3104203.09</td>
<td>(0.90, 1.00)</td>
<td>2452726.18</td>
<td>(1.00, 1.00)</td>
</tr>
<tr>
<td>0.2</td>
<td>10000</td>
<td>(0.80, 1.00)</td>
<td>12458372.77</td>
<td>(0.90, 1.00)</td>
<td>9843745.44</td>
<td>(1.00, 1.00)</td>
</tr>
<tr>
<td>0.2</td>
<td>20000</td>
<td>(0.80, 1.00)</td>
<td>49916800.01</td>
<td>(0.90, 1.00)</td>
<td>3940788.49</td>
<td>(1.00, 1.00)</td>
</tr>
<tr>
<td>0.4</td>
<td>2000</td>
<td>(0.80, 1.00)</td>
<td>743401.43</td>
<td>(0.90, 1.00)</td>
<td>588089.65</td>
<td>(1.00, 1.00)</td>
</tr>
<tr>
<td>0.4</td>
<td>5000</td>
<td>(0.80, 1.00)</td>
<td>4673243.90</td>
<td>(0.90, 1.00)</td>
<td>3692437.29</td>
<td>(1.00, 1.00)</td>
</tr>
<tr>
<td>0.4</td>
<td>10000</td>
<td>(0.80, 1.00)</td>
<td>18721520.61</td>
<td>(0.90, 1.00)</td>
<td>14792303.25</td>
<td>(1.00, 1.00)</td>
</tr>
<tr>
<td>0.4</td>
<td>20000</td>
<td>(0.80, 1.00)</td>
<td>74943221.43</td>
<td>(0.90, 1.00)</td>
<td>59214359.90</td>
<td>(1.00, 1.00)</td>
</tr>
</tbody>
</table>

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%

<table>
<thead>
<tr>
<th>Type of Distribution</th>
<th>Student’s t-Distribution</th>
<th>Skew-Normal Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Degrees of Freedom</td>
<td>Shape Parameter</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>α</td>
<td>Variance Tests</td>
<td></td>
</tr>
<tr>
<td>LRT</td>
<td>0.0626</td>
<td>0.0088</td>
</tr>
<tr>
<td>Bartlett</td>
<td>0.0572</td>
<td>0.0058</td>
</tr>
<tr>
<td>Levene</td>
<td>0.0080</td>
<td>0.0024</td>
</tr>
<tr>
<td>LV Linear</td>
<td>0.0014</td>
<td>0.0012</td>
</tr>
<tr>
<td>LV Square</td>
<td>0.0030</td>
<td>0.0006</td>
</tr>
<tr>
<td>LRT</td>
<td>0.1346</td>
<td>0.0390</td>
</tr>
<tr>
<td>Bartlett</td>
<td>0.1278</td>
<td>0.0334</td>
</tr>
<tr>
<td>Levene</td>
<td>0.0202</td>
<td>0.0124</td>
</tr>
<tr>
<td>LV Linear</td>
<td>0.0118</td>
<td>0.0134</td>
</tr>
<tr>
<td>LV Square</td>
<td>0.0182</td>
<td>0.0100</td>
</tr>
<tr>
<td>LRT</td>
<td>0.2340</td>
<td>0.1174</td>
</tr>
<tr>
<td>Bartlett</td>
<td>0.2398</td>
<td>0.1084</td>
</tr>
<tr>
<td>Levene</td>
<td>0.0536</td>
<td>0.0504</td>
</tr>
<tr>
<td>LV Linear</td>
<td>0.0578</td>
<td>0.0532</td>
</tr>
<tr>
<td>LV Square</td>
<td>0.0520</td>
<td>0.0478</td>
</tr>
<tr>
<td>LRT</td>
<td>0.3082</td>
<td>0.1852</td>
</tr>
<tr>
<td>Bartlett</td>
<td>0.3284</td>
<td>0.1762</td>
</tr>
<tr>
<td>Levene</td>
<td>0.0984</td>
<td>0.1016</td>
</tr>
<tr>
<td>LV Linear</td>
<td>0.1066</td>
<td>0.1020</td>
</tr>
<tr>
<td>LV Square</td>
<td>0.0952</td>
<td>0.1006</td>
</tr>
</tbody>
</table>

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.
<table>
<thead>
<tr>
<th>Type of Distribution</th>
<th>Student’s $t$-Distribution</th>
<th>Skew-Normal Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Degrees of Freedom</td>
<td>Shape Parameter</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Variance Tests</td>
<td></td>
</tr>
<tr>
<td>0.001</td>
<td>LRT</td>
<td>0.0602</td>
</tr>
<tr>
<td></td>
<td>Bartlett</td>
<td>0.0784</td>
</tr>
<tr>
<td></td>
<td>Levene</td>
<td>0.0016</td>
</tr>
<tr>
<td></td>
<td>LV Linear</td>
<td>0.0008</td>
</tr>
<tr>
<td></td>
<td>LV Square</td>
<td>0.0008</td>
</tr>
<tr>
<td>0.01</td>
<td>LRT</td>
<td>0.1324</td>
</tr>
<tr>
<td></td>
<td>Bartlett</td>
<td>0.1682</td>
</tr>
<tr>
<td></td>
<td>Levene</td>
<td>0.0142</td>
</tr>
<tr>
<td></td>
<td>LV Linear</td>
<td>0.0102</td>
</tr>
<tr>
<td></td>
<td>LV Square</td>
<td>0.0098</td>
</tr>
<tr>
<td>0.05</td>
<td>LRT</td>
<td>0.2476</td>
</tr>
<tr>
<td></td>
<td>Bartlett</td>
<td>0.3030</td>
</tr>
<tr>
<td></td>
<td>Levene</td>
<td>0.0556</td>
</tr>
<tr>
<td></td>
<td>LV Linear</td>
<td>0.0484</td>
</tr>
<tr>
<td></td>
<td>LV Square</td>
<td>0.0448</td>
</tr>
<tr>
<td>0.1</td>
<td>LRT</td>
<td>0.3272</td>
</tr>
<tr>
<td></td>
<td>Bartlett</td>
<td>0.4040</td>
</tr>
<tr>
<td></td>
<td>Levene</td>
<td>0.1062</td>
</tr>
<tr>
<td></td>
<td>LV Linear</td>
<td>0.1034</td>
</tr>
<tr>
<td></td>
<td>LV Square</td>
<td>0.0968</td>
</tr>
</tbody>
</table>

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 $\alpha$-level.
### 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 $\alpha$-level.
Bibliography


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


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