Author:
Lyon, Matt S

Title:
Variance quantitative trait loci

development of novel software and methodology to facilitate discovery, analysis and sharing

Variance quantitative trait loci: development of novel software and methodology to facilitate discovery, analysis and sharing

Matthew S. Lyon

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Bristol Medical School

NIHR Bristol Biomedical Research Centre &

MRC Integrative Epidemiology Unit

Department of Population Health Sciences

Bristol Medical School

University of Bristol

Bristol, UK

28 December 2022

Word count: 36,200

Abstract

Genetic epidemiological studies have largely focused on SNP mean effects, but variance effects may also exist that can indicate the presence of SNP interaction effects. Identification of these effects may be useful for improving understanding of disease mechanisms, prediction of disease outcomes, and in combination with other data may provide opportunities for developments in precision medicine. This thesis aims to develop methodology and software to identify and analyse variance loci applied to serum biomarker concentration. To achieve these aims, a regression-based Brown-Forsythe variance test was evaluated and implemented in C++ and R which enables adjustment of covariates and provides an unbiased variance effect estimate for normally distributed traits (**Chapter 4**). This model was subsequently applied in variance genome-wide association studies (vGWAS) of 30 serum biomarkers in UK Biobank identifying 468 variance loci of 209 million SNPs tested. These loci were investigated to detect 82 gene-environment and six gene-gene interactions including three novel epistatic effects (**Chapter 5**). The utility of these vGWAS summary statistics in detecting violation of Mendelian randomization homogeneity assumptions was explored through a series of simulation studies. This approach was subsequently applied to investigate the impact of homogeneity violation of low-density lipoprotein, urate and glucose on cardiovascular disease, gout, and type 2 diabetes, respectively. There was no strong evidence of difference in causal estimate after removing instruments associated with exposure variance. These findings are consistent with the main analysis targeting the population average causal effect (**Chapter 6**). To facilitate sharing and future analyses of vGWAS summary statistics, an efficient and robust storage format was developed using the variant call format that can be used for any GWAS analysis along with Python packages, web-interface and data processing pipeline which are widely used and embedded within the MRC-IEU OpenGWAS infrastructure (**Chapter 7**).

Acknowledgments

To my supervisors Professor Tom Gaunt, Professor Kate Tilling, Dr Louise Millard, and Professor George Davey Smith for their support, guidance and mentorship and research funders for supporting this research.

This research was funded by the NIHR Biomedical Research Centre at University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol. This research was also funded by the UK Medical Research Council as part of the MRC Integrative Epidemiology Unit (MC_UU_00011/1, MC_UU_00011/3, MC_UU_00011/4 and MC_UU_00011/6). The views expressed are those of the author and not necessarily those of the NIHR, The Department of Health and Social Care or Medical Research Council. Chapter 7 was also supported by Wellcome Trust and Royal Society [208806/Z/17/Z].

I thank all study participants and UK Biobank team without whom this work would not have been possible.

Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ……. DATE: …….

Research outputs and engagement

Chapter 4

Staley, J. R., Windmeijer, F., Suderman, M., **Lyon, M. S.**, Davey Smith, G., & Tilling, K. (2021). A robust mean and variance test with application to high-dimensional phenotypes. European Journal of Epidemiology, 1, 1–1

I performed type I error rate simulations of the Brown-Forsythe test included in this publication which were adapted for inclusion in **Chapter 4**.

Chapter 4 & 5

Lyon, M. S., Millard, L. A. C., Davey Smith, G., Gaunt, T. R., & Tilling, K. (2022). Hypothesis-free detection of gene-interaction effects on biomarker concentration in UK Biobank using variance prioritisation. MedRxiv, 2022.01.05.21268406

I led this analysis and wrote the manuscript which was edited by PhD supervisors. Findings from this manuscript are included in **Chapter 4** and **Chapter 5**. I also developed C++ variance GWAS software implementing the LAD-BF model described in **Chapter 4** (https://github.com/MRCIEU/varGWAS) and an R-package for estimating SNP-variance effects using LAD-BF model (https://github.com/MRCIEU/varGWASR). Professor Tilling proposed the LAD-BF test and derived an expression for the relationship between exposure and outcome variance under interaction effect and the formula for calculating variance from mean-absolute deviation in this context.

Zheng, J., Tang, H., **Lyon, M.**, Davies, N. M., Walker, V., Floyd, J. S., Austin, T. R., Shojaie, A., Psaty, B. M., Gaunt, T. R., & Davey Smith, G. (2021). Genetic effect modification of cis-acting C-reactive protein variants in cardiometabolic disease status. BioRxiv, 2021.09.23.461369

I advised on interaction analyses and provided variance QTL analyses of CRP loci to support findings presented in this paper.

Chapter 7

Lyon, M. S., Andrews, S. J., Elsworth, B., Gaunt, T. R., Hemani, G., & Marcora, E. (2021). The variant call format provides efficient and robust storage of GWAS summary statistics. Genome Biology, 22(1), 1–10

Findings from this paper are included in **Chapter 7** which I drafted and was edited by Dr Shea Andrews, Dr Ben Elsworth, Professor Tom Gaunt, Dr Gibran Hemani and Professor Edoardo Marcora. I also developed Python software for harmonising GWAS summary statistics to GWAS-VCF (https://github.com/MRCIEU/gwas2vcf), a web-application for converting GWAS summary statistics to GWAS-VCF (https://github.com/MRCIEU/gwas2vcfweb) and Python library for reading GWAS-VCF files (https://github.com/MRCIEU/pygwasvcf).

Hayhurst, J., Buniello, A., Harris, L., Mosaku, A., Chang, C., Gignoux, CR., Hatzikotoulas, K., Karim, M. A., Lambert, S. A., **Lyon, M**., McMahon, A., Okada, Y., Pirastu, N., Rayner, N. W., Schwartzentruber, J., Vaughan, R., Verma, S., Wilder, S. P., Cunningham, F., Hindorff, L., Wiley, K., Parkinson, H., Barroso, I. (2022). A community driven GWAS summary statistics standard. BioRxiv, 2022.07.15.500230

Following from Lyon *et al*, 2021 (above) I was invited to participate in the NHGRI-EBI GWAS Catalog data format and content working group. This working group developed a community standard for sharing GWAS summary statistics which I contributed towards.

Liu, Y., Elsworth, B., Erola, P., Haberland, V., Hemani, G., **Lyon, M.**, Zheng, J., Lloyd, O., Vabistsevits, M., & Gaunt, T. R. (2021). EpiGraphDB: a database and data mining platform for health data science. Bioinformatics, 37(9), 1304–1311

I contributed to data harmonisation for Mendelian randomization estimates included in EpigraphDB, described in this paper. The harmonisation pipeline I developed is described in **Chapter 7**.

Elsworth, B., **Lyon, M.**, Alexander, T., Liu, Y., Matthews, P., Hallett, J., Bates, P., Palmer, T., Haberland, V., Davey Smith, G., Zheng, J., Haycock, P., Gaunt, T. R., & Hemani, G. (2020). The MRC IEU OpenGWAS data infrastructure. BioRxiv, 2020.08.10.244293

I co-developed the MRC-IEU OpenGWAS API (https://github.com/MRCIEU/opengwas-api) which provides back-end functionality to the MRC-IEU OpenGWAS website (https://gwas.mrcieu.ac.uk) and developed data harmonisation pipeline described in **Chapter 7**. I was also a member of MRC-IEU OpenGWAS consortium which managed development of this resource.

Table of Contents

List of Tables

List of Figures

List of Equations

Abbreviations

Abbreviation	Term
ACE	Average causal effect
ADEMP	Aims, data-generating mechanism, estimand, methods, performance
ALB	Albumin
ALK	Alkaline phosphatase
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
API	Applied programming interface
ApoA	Apolipoprotein A
ApoB	Apolipoprotein B
AST	Aspartate aminotransferase
BAM	Binary alignment format
BCF	Binary call format
BF	Brown-Forsythe
BGEN	Binary GEN
BGZIP	Block GNU Zip
BMI	Body mass index
BR	Bilirubin
CHD	Coronary heart disease
CI	Confidence interval
CPU	Central processing unit
CRP	C-reactive protein
dbGAP	Database of genotypes and phenotypes
dbSNP	Database of single nucleotide polymorphism
DGLM	Double generalized linear model
DOI	Digital object identifier
DRM	Deviation regression model
EBI	European Bioinformatics Institute
EFO	Experimental factor ontology
eQTL	Expression quantitative trait locus
ES	Effect size
FAIR	Findability, accessibility, interoperability, reusability
FEV1	Forced expiratory volume in 1-second
FVC	Forced vital capacity
GATK	Genome analysis tool kit
GDPR	General Data Protection Regulation
GEN	Oxford genotype file format
GGT	Gamma glutamyltransferase
GHz	Gigahertz
GNU	GNU's Not Unix!
GPL	GNU General Public License

GPU	Graphics processing unit
GRC	Genome reference consortium
GWAS	Genome-wide association study
GxE	Gene-environment interaction
GxG	Gene-gene interaction
H^2	Broad-sense heritability
h^2	Narrow-sense heritability
HbA1C	Glycated haemoglobin
HDL	High density lipoprotein
HLM	Heteroskedastic linear model
HTSJDK	High throughput sequencing Java development toolkit
HTSLIB	High throughput sequencing library
HUGO	Human genome organisation
HWE	Hardy-Weinberg equilibrium
IGF-1	Insulin growth factor
IV	Instrumental variable
IVW	Inverse-variance weighted
JLSsc	Joint location and scale score test
LAD	Least absolute deviation
LAD-BF	Least absolute deviation Brown-Forsythe
LD	Linkage disequilibrium
LDL	Low-density lipoprotein
LipoA	Lipoprotein A
LP	-log10(P)
LRT	Likelihood ratio test
LRTmv	Likelihood ratio test of mean and variance
MAD	Median absolute deviation
MAF	Minor allele frequency
MR	Mendelian Randomization
MRC-IEU	Medical Research Council Integrative Epidemiology Unit
NHGRI	National Human Genome Research Institute
NOSH	NO Simultaneous Heterogeneity
OLS	Ordinary least squares
OR	Odds ratio
OSCA	Omic-data-based complex trait analysis
PACE	Population average causal effect
PIP	Posterior inclusion probability
PMID	PubMed identifier
PRS	Polygenic risk score
Q-Q	Quantile-quantile
QTL	Quantitative trait loci
RCT	Randomised control trial
RF	Rheumatic factor

RSIDX	Reference SNP index
SD	Standard deviation
SE	Standard error
SHBG	Sex-hormone binding globulin
SMR	Summary data–based Mendelian randomization
SNP	Single nucleotide polymorphism
SuSiE-RSS	Sum of single effects regression summary statistics
SVLM	Squared residual value linear modelling
T2DM	Type 2 diabetes mellitus
TC	Total cholesterol
TG	Triglycerides
TSSR	Two-step squared residual
VCF	Variant call format
vGWAS	Variance genome-wide association study
vQTL	Variance quantitative trait loci

Chapter 1: Introduction

1.1 Contribution statement

Parts of the introduction were taken from papers or manuscripts that I wrote but that were contributed towards by others.

Background on genome-wide association studies including sharing of summary statistics was taken from Lyon *et al*, 2021[1], a paper which I drafted and was edited by Dr Shea Andrews, Dr Ben Elsworth, Professor Tom Gaunt, Dr Gibran Hemani and Professor Edoardo Marcora.

Background on Mendelian randomization and instrumental variable assumptions was taken from a manuscript in preparation which I wrote and was edited PhD supervisors and Dr Fernando Hartwig (University of Pelotas).

Background on interaction effects, variance QTLs, and biomarkers forms part of a manuscript I wrote that was edited by PhD supervisors available as a preprint on MedRxiv (Lyon *et al*, 2022)[2]

1.2 Background

1.2.1 Complex traits

Complex traits are phenotypes influenced by many small genetic effects at loci throughout the genome (also known as polygenic traits) in combination with environmental factors[3,4]. This is in contrast with Mendelian traits (also known as monogenic traits) which are affected by highly penetrant variation in a single or small group of genes[3].

The genetic contribution of a phenotype P can be described in terms of heritability which is defined as the proportion of phenotype variance $\sigma^2 P$ explained by genetic contribution[3,5] $\sigma^2 G$. Narrow-sense heritability h^2 concerns only the variance explained by

additive genetic effects $\sigma^2 A$ (**Equation 1.2.1.1**)[3,5]. Meanwhile, broad-sense heritability H^2 is the total variance explained by the genetic contribution $\sigma^2 G$ (**Equation 1.2.1.2**)[3,5] which includes additive $\sigma^2 A$, dominant $\sigma^2 D$ and interaction $\sigma^2 I$ effects (**Equation 1.2.1.3**)[3,5]. In addition, phenotypes are affected by environmental factors E as well as interaction of genetics G and environment GE (**Equation 1.2.1.4**)[5].

Equation 1.2.1.1 Narrow-sense heritability

Narrow-sense heritability h^2 is defined by:

$$h^2 = \sigma^2 A / \sigma^2 P$$

Where $\sigma^2 A$ is the phenotypic variance explained by the additive genetic component and $\sigma^2 P$ is the total phenotypic variance[3,5].

Equation 1.2.1.2 Broad-sense heritability

Broad-sense heritability H^2 is defined by:

$$H^2 = \sigma^2 G / \sigma^2 P$$

Where $\sigma^2 G$ is the phenotypic variance explained by the genetic contribution and $\sigma^2 P$ is the total phenotypic variance[3,5].

Equation 1.2.1.3 Genetic variance components

Phenotypic variance explained by the genetic contribution $\sigma^2 G$ is the sum of:

$$\sigma^2 G = \sigma^2 A + \sigma^2 D + \sigma^2 I$$

Where phenotypic variance explained by additive, dominant and interaction genetic effects are denoted by $\sigma^2 A$, $\sigma^2 D$ and $\sigma^2 I$, respectively[3,5].

Equation 1.2.1.4 Phenotype definition

The phenotype of an organism is expressed as the sum of:

$$P = G + E + GE$$

Where P is the phenotype, G and E are the genetic and environmental contributions and GE are interactions between genotype and environment[5].

1.2.2 Genome-wide association study

The genome-wide association study (GWAS) is a powerful method for identifying genetic loci associated with any trait, including case-control studies for binary disease outcomes and quantitative measures such as height and gene expression[6,7]. A GWAS is performed using a statistical test of the relationship between SNP allele count for genetic variants throughout the genome and phenotype in the study population[7].

GWAS has implicit assumptions that individuals under study have similar genetic ancestry and vary only because of the SNP under investigation[7]. Violation of this assumption may induce spurious associations hence quality control procedures are essential[7].

Some commonly used quality control steps including restricting to high quality SNPs by removing variants that depart from Hardy-Weinberg equilibrium (HWE)[7], have high missingness[7], or are poorly imputed[7]. Secondly, participants with SNP-phenotype sex mismatches are removed to avoid sample errors and aneuploidies[7]. Third, the sample can be restricted to a homogeneous and unrelated population (unless performing family-based analyses[8]) using measures of genetic ancestry and relatedness[7]. Inspection of GWAS test statistics compared with the null distribution is important to look for systematic inflation of test statistics which may indicate results are unreliable[7].

Since common SNPs that are physically close are often inherited together and highly correlated (known as linkage disequilibrium [LD])[7], GWAS provides evidence of association for

the genetic locus and trait, but the strongest associated SNP may not be causally related to the outcome[9] (**Chapter 1.2.4**).

1.2.3 Population stratification

Genetic association studies that correlate variant allele count with disease occurrence or quantitative trait are especially susceptible to confounding (**Chapter 1.3.2**) by differences in ancestry across the study population, a situation known as population stratification[7]. This occurs because allele frequencies vary between populations regardless of the trait of interest[10] and can bias SNP-trait estimates through confounding (**Chapter 1.3.2**). However, there are several approaches that may be used to mitigate this type of bias including restriction to an ethnically homogeneous population[10,11], adjustment for self-reported or genetically[11] measured ancestry, family-based analyses[8,10–12] and random polygenic effect[13]. Secondly, replication studies should be performed to increase robustness of findings[10].

1.2.4 Fine mapping of causal loci

GWAS (**Chapter 1.2.2**) identifies loci associated with a trait, but the top associated SNP (i.e., lead SNP) at a locus may not be causally (**Chapter 1.3.4**) responsible for the observed effect on trait, instead it may just be correlated with the causal SNP[9]. This is because either the causal SNP was not available for testing (i.e., not typed or imputed) or due to low power and sampling variability a related but non-causal SNP gave rise to a lower p-value[9]. The goal of identifying causal SNP(s) at a locus is known as fine mapping[9].

Three main approaches to fine mapping have been suggested[9]. The heuristic approach prioritises candidate causal SNPs in high LD with the lead SNP by applying a correlation threshold but this approach does not provide any measures of confidence in the candidate

causal SNPs[9]. Penalised regression has been applied to identify potential causal SNPs by modelling all SNPs in a joint model and then shrinking or removing weak predictors[9]. Bayesian methods have been proposed that are advantageous because they provide probabilistic evidence for each candidate SNP[9]. 1. Candidate causal SNPs can be further refined using functional annotation[9].

1.2.5 Sharing of GWAS results

Results from GWAS analyses (**Chapter 1.2.2**) have led to a wide range of secondary research applications including causal gene and functional variant prioritisation[14], pathway analysis[6], causal inference[15], risk prediction[6], genetic correlation[16] and heritability estimation[17] among others. To facilitate these applications, development of new methodologies and replication of findings it is vital that GWAS results are made freely available for research purposes[7]. This message is widely accepted, with many large consortia providing access to data[7] and large databases such as the MRC-IEU OpenGWAS platform[18] and GWAS catalog[19] aggregating and distributing these data. Sharing of data has also become a condition of future research funding[7].

European General Data Protection Regulation (GDPR) laws set out strong requirements for data sharing to avoid de-identification and ensure appropriate consent is in place which make sharing of individual level data challenging[7]. However, GWAS results are summary statistics that describe the association between SNP and trait (i.e., variant, effect size, standard error, test p-value etc) without including individual identifiable information[7] and may be shared freely often without restriction. But there is currently no common agreed format for storing and sharing of GWAS summary statistics.

1.2.6 Issues with non-standard GWAS summary statistics

Various processing issues are typically encountered during secondary analysis of GWAS summary statistics. First, there is often inconsistency and ambiguity of which allele relates to the effect size estimate[20] (the "effect" allele). Confusion over the effect allele can have disastrous consequences on the interpretation of GWAS findings and the validity of post-GWAS analyses[20]. For example Mendelian randomization (MR; **Chapter 1.3.7**) studies may provide causal estimates with incorrect effect directionality[20]. Likewise, prediction models based on polygenic risk scores[21] might predict disease wrongly or suffer reduced power if some of the effect directionalities are incorrect. Second, the schema (i.e., which columns/fields are included and how they are named) of these tabular formats varies greatly[22]. Absent fields can limit analyses and although approaches exist to estimate the values of some of these missing columns (e.g., standard error from P value[23]) imprecision may be introduced reducing subsequent test power. Varying field names are easily addressed in principle, but the process can be cumbersome and error prone. Third, data are frequently distributed with no or insufficient metadata describing the study, trait(s), and variants (e.g., trait measurement units, variant id/annotation sources, etc.) which can lead to errors, impede integration of results from different studies and hamper reproducibility[22]. Fourth, querying unindexed text files is slow and memory inefficient[24], making some potential applications computationally infeasible (e.g., systematic hypothesis-free analyses).

1.2.7 Existing GWAS summary statistics formats

Some proposals for a standard tabular format have been made. The NHGRI-EBI GWAS Catalog developed a tab-separated values text format with a minimal set of required (and

optional) columns along with standardised headings and separate metadata file[19,25]. The SMR tool[26] introduced a binary format for rapid querying of quantitative trait loci. These approaches are adequate for storing variant level summary statistics but do not enforce allele consistency or support embedding of essential metadata within the GWAS file. There is a need to develop a GWAS summary statistics file format that can address these limitations.

1.3 Experimental design

1.3.1 Observational analysis

Observational studies aim to measure the association between an exposure and outcome using a study sample of the population[27] observed in the data. However, as the exposure cannot be controlled during the experiment, observational studies do not provide evidence on cause and effect owing to the interplay of many other factors (known as confounders **Chapter 1.3.2**) which may not be known or measured[27].

1.3.2 Confounding

A confounder is, or represents, a common cause of both exposure and outcome creating a pathway between exposure and outcome leading to biased associations[28]. Accounting for the confounder effect either by adjustment or stratification can remove this bias and produce valid estimates of the exposure-outcome effect[28]. However, these procedures require that all confounders are hypothesised, measured (without measurement error) and adjusted in the model but this is difficult to achieve[28]. Instead, certain study designs are employed which minimise confounding such as the randomised control trial[29] (RCT) and MR[15] (**Chapter 1.3.7**).

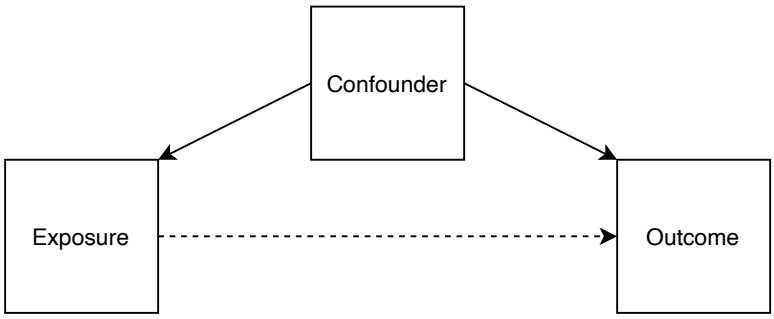
1.3.3 Reverse causation

Reverse causation is another form of bias that is especially problematic for observational studies and is said to occur when the outcome at an earlier timepoint influences the exposure under investigation[30]. For example, early on in the disease process the disease itself may influence changes in the exposure[30].

1.3.4 Causal effect

In contrast with observation studies (**Chapter 1.3.1**), experimental studies such as the RCT aim to estimate the causal effect of an exposure-outcome relationship[27] (**Figure 1.3.4.1**). Since interventions developed to target an outcome will only be successful if they change the level of an exposure that is causally related to said outcome[27]. This requires knowledge of the exposure-outcome causal effect which can be obtained through experimental studies in which the researchers' have control over the exposure and can limit confounding variables[27].

Figure 1.3.4.1 Confounding causal diagram



Causal diagram of confounded relationship.

1.3.5 Instrumental variable analyses

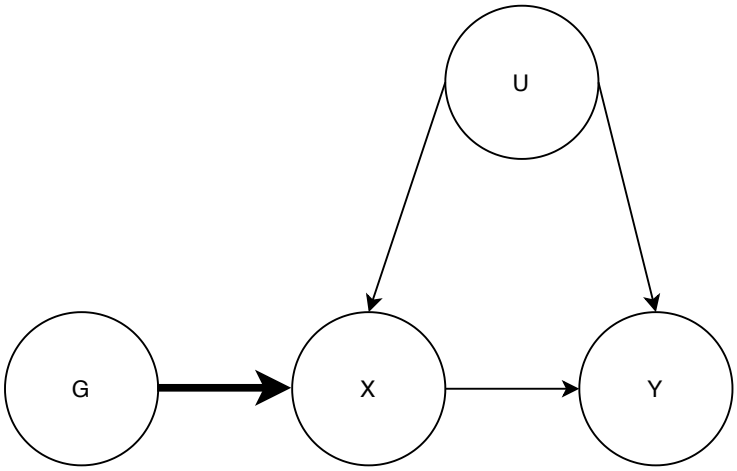
Instrumental variable (IV) analyses provides causal evidence for an exposure-outcome effect that is less susceptible to confounding (**Chapter 1.3.2**) and reverse causation (**Chapter 1.3.3**) than conventional observational epidemiological associations[15,31,32] (**Chapter 1.3.1**). IV analyses employ a third variable known as an IV or instrument that is not affected by confounding of the outcome[15,31,32], strongly predicts the exposure[15,31,32] and is unrelated to the outcome except via the exposure[15,31,32]. This instrument may then be used to test for a causal effect of exposure on outcome. Examples of suitable instruments include changes to and regional variation in public policy (e.g., rising of school leaving age[33] and variation in prescribing policies[34]) and genetic variants[15].

1.3.6 Instrumental variable assumptions

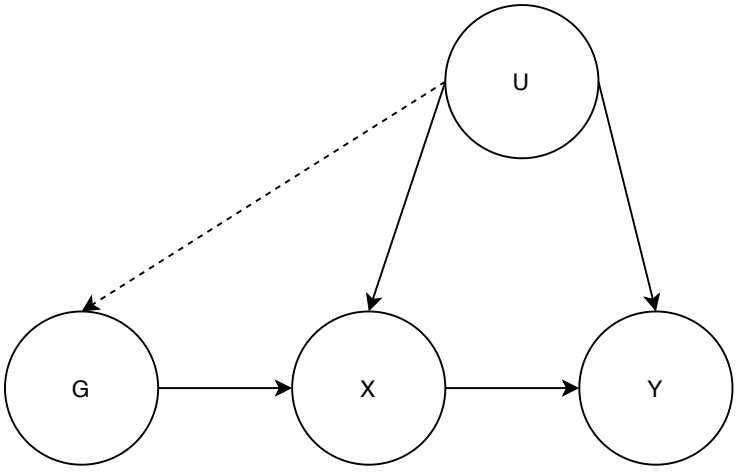
Formally, IV analyses require three core assumptions (IV1-IV3)[15,31,32] (**Figure 1.3.6.1**). IV1, the instrument is robustly associated with the exposure (relevance assumption)[15,31,32]. IV2, there are no confounders of the instrument-outcome relationship (exchangeability)[15,31,32]. IV3, the instrument only affects the outcome via the exposure (exclusion restriction)[15,31,32]. However, only the relevance assumption (IV1) can be proven to hold true[31,32]. The plausibility of these assumptions must be considered to draw appropriate inference[31,32].

Figure 1.3.6.1 Instrumental variable assumptions diagram

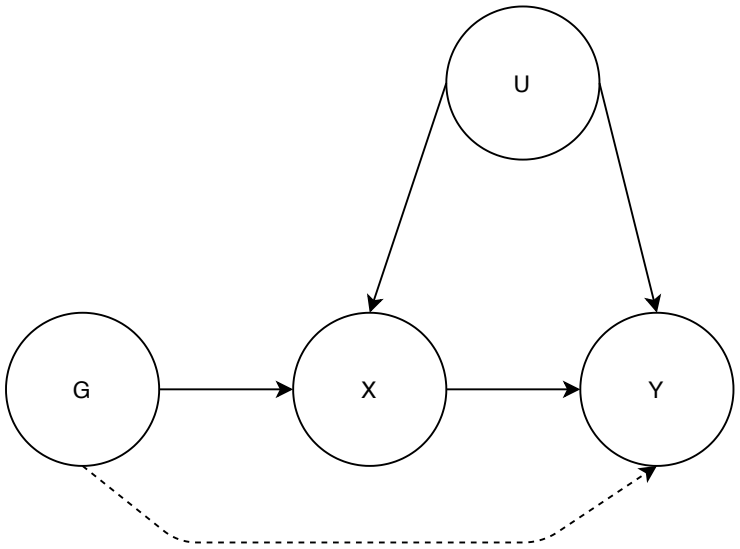
A



B



C



Causal diagrams of instrumental variable analysis assumptions. X, exposure. G, instrument, Y, outcome. U, unmeasured/unknown confounder. A, Relevance assumption (IV1) showing bold edge between genotype and exposure indicating the requirement for a robust association. B, exchangeability (IV2) showing dotted line indicating no unmeasured confounders of the genotype-outcome relationship. C, exclusion restriction (IV3) and dotted line indicating no direct effect of genotype on outcome independent of the exposure. Adapted from Sanderson *et al*[31] with permission from Springer Nature (License Number 5367640337820).

IV1-3 assumptions are sufficient to test the sharp null hypothesis that the exposure does not have an effect on the outcome for any individual in the population[31,35]. At least one additional assumption – these are often collectively referred to as IV4 assumptions – is needed to produce a clearly defined causal estimand point estimate and confidence interval[31,35,36].

Several IV4 assumptions have been proposed and the choice of assumption influences interpretation of the estimate[35,36]. The causal estimand of interest is typically the average causal effect[35] (ACE). For a binary exposure, ACE is the average difference in outcome between exposure groups[35,37]. For a continuous exposure, ACE defines the average difference for a one unit increase in exposure[38].

Homogeneity of the IV-exposure[39] and/or exposure-outcome[40] effect is necessary to estimate the population ACE (PACE) which is ACE over the whole population under study. A weaker assumption of IV-exposure monotonicity (i.e., "no defiers") will produce a local ACE (LACE), which is the ACE in a subgroup of the population (i.e. "compliers"), but this subgroup may be unidentifiable[35]. This is important since an intervention developed to target a LACE may only be effective among the compliers[41]. Meanwhile, under PACE an intervention will be effective across the entire population[41].

Recently, the NO Simultaneous Heterogeneity (NOSH) assumption was proposed[35] which implies PACE can be identified even in the presence of effect modification of either IV association with exposure or exposure-outcome association, provided that effect modifiers are independent (NOSH assumption one) and the exposure-outcome relationship is additive linear (NOSH assumption two).

Although IV4 cannot be proven to hold, hypothesised testing of IV-exposure interaction effects to evaluate homogeneity assumptions has been suggested[34]. But this approach may miss unanticipated interaction effects, cannot be used if the modifier is unmeasured, and potentially incurs a large multiple testing burden[4,42]. Alternatively, IV association with exposure variance has been suggested as an approach to assess IV4 assumptions[31,35,43] but to my knowledge this has not yet been explored.

1.3.7 Mendelian randomization

Mendelian randomization (MR) is a form of IV analysis that employs genetic variants to proxy for an exposure in order to estimate the causal exposure-outcome effect free of observational confounding and reverse causation[15]. MR framework requires the IV three core assumptions described above (**Chapter 1.3.5**; **Figure 1.3.6.1**)[31,32].

MR can provide causal evidence of the effect of modifiable risk factors on disease[31,32] even when clinical trials are ethically infeasible or impractical. For example MR has been applied to study the effects of alcohol consumption on cardiovascular disease[31] but an RCT of alcohol consumption would not be ethical. MR is less susceptible to confounding (**Chapter 1.3.2**) and reverse causation (**Chapter 1.3.3**) than conventional observational epidemiological analyses[31,32] (**Chapter 1.3.1**).

MR estimation may be performed using individual level data using two-stage least squares provided that the exposure and outcome are available for each observation[31,32]. Alternatively, two-sample MR may be used in which the instrument-exposure and instrument-outcome associations are estimated from two separate samples derived from the same

population[31,32]. Two-sample MR may be performed using freely available data in the form of GWAS summary statistics[31,32] (**Chapter 1.2.2**).

Violation of IV1 produces weak instrument bias[44] (**Chapter 1.3.6**). Under one-sample MR, weak instrument bias produces estimates that are biased in the direction of the observational association[45]. Meanwhile, weak instrument biased with two-sample MR will produces estimates closer to the null[45] provided there is no sample overlap between exposure and outcome datasets[45]. Where there is sample overlap, weak instruments may produce biased estimates towards the observational association[45] (**Chapter 1.3.1**).

One potential source of error in two-sample MR is poor harmonisation of GWAS data (**Chapter 1.2.2**) that result in mismatching of effect alleles[20]. This is problematic as, incorrect effect alleles between studies will produce an MR estimate with the wrong sign[20].

One approach to evaluate the plausibility of MR assumptions is through the use of positive and negative controls[31,46]. Controls could be outcomes which are expected (positive) or unexpected (negative) to be causally affected by the exposure using evidence from other studies or epidemiological domain knowledge[31,46]. Alternatively, the instrument-outcome association may be evaluated in subgroups of the population where the instrument is either anticipated (positive control) or not anticipated (negative control; also known as the no-relevance group[47]) to associate with the exposure[31,46]. For example, the MR effect of alcohol intake on cardiovascular disease was evaluated in a Chinese population[31]. In this population, women are less likely to consume alcohol, and the negative control was assessed by estimating the effect among women only where the instrument-exposure effect is expected to be close to the null[31].

1.4 Effect modification

1.4.1 Gene-interaction effects

SNPs may interact with other SNPs (gene-gene interaction, also known as epistasis[4]; GxG) or the environment (gene-environment; GxE). The presence of genetic interaction effects may imply perturbation of protein abundance or function that varies by the level of another variable known as the modifier[4].

Genetic interaction effects are important to study for several reasons. First, this evidence can aid in elucidating disease mechanisms and improving our understanding of disease biology by using genetic information to study implicated pathways[4,48]. Second, to improve prediction of disease outcomes using genetic data[4,48] and by implication improve our understanding of the heritable components of disease[3,4]. This may help to explain 'missing heritability'[3,4] (**Chapter 1.2.1**). Third, to identify therapeutic targets which may be used to develop drugs or preventative advice[48,49] supporting developments in precision medicine (**Chapter 1.4.2**).

1.4.2 Precision medicine

Precision medicine is a treatment paradigm that aims to develop tailored treatments for patients within subgroups defined by characteristics of individual level biology with the aim of enhancing efficacy and reducing unwanted side-effects[50–52]. This approach has been applied to improve efficacy of treatment for cystic fibrosis patients with specific sodium channel mutations and cancers with certain somatic mutations[52]. Patient subgroups have also been defined by variation in *VKORC1* and *CYP2C9* which affect warfarin metabolism enabling tailored dosing strategies to improve efficacy and reduce side-effects[52].

1.4.3 Identifying genetic interaction effects

Genetic interaction effects (**Chapter 1.4.1**) have traditionally been identified assuming linear model using linear or logistic regression for continuous and binary outcomes, respectively[4]. This may be accomplished by comparing a full regression model containing all possible interaction terms (GxE: $additive \times environment$ and $dominant \times environment$; GxG: $additive \times additive$ and $additive \times dominant$ and $dominant \times additive$ and $dominant \times dominant$[53]) with a restricted model containing only main effects by contrasting model fit for example using an F-test[53] (**Chapter 2.2.6**) or likelihood ratio test[4]. Alternatively, when only linear additive interaction effects are anticipated then a single interaction term (GxE: $additive \times environment$; GxG: $additive \times additive$) may be estimated and the coefficient taken as evidence for interaction[4] (**Chapter 2.2.5**). The latter consumes fewer degrees of freedom and therefore has greater power if only linear additive interaction effects are present[4].

The search space for interaction effects is potentially very large and for example could involve exhaustively testing every SNP in the genome against every other SNP yielding hundreds of billions of tests[4]. The same may be potentially true for testing of GxE effects if modifiers measured across the entire phenome are considered[42]. The number of modifiers tested in a GxE experiment is usually lower than for GxG analysis but could include high dimensional measurements such as continuous monitoring sensors[54]. This will lead to large numbers of tests and consequently elevated type I error rate[7]. Type I error rate can be controlled using multiple testing correction, for example with Bonferroni-correction by dividing significance threshold by the number of tests performed[7]. However, controlling type I error rates comes at the cost of reduced power to detect effects (i.e., type II errors)[7]. Additionally,

when statistical power is low, findings that are statistically 'significant' are more likely to be overestimated compared with the true value[55]. To reduce multiple testing, pairwise interaction analyses of SNPs with moderate main effects can be performed[56]. However, this approach may miss weaker overall (main) effects that are strong in subgroups of the population or opposing effect directionality between population subgroups[57], yet these may offer the greatest potential for precision medicine.

Aside from multiple testing correction, testing for an interaction term itself has lower power than for main effects[57]. For example, under an RCT the sample size needed to detect an interaction with equal sized subgroups is around four times the size needed to detect the main effect of equal magnitude[57,58]. Before interaction findings may be considered robust it is essential for independent replication and an appraisal of biological plausibility by considering gene function and affected biological pathway(s)[48].

1.4.4 Previously reported genetic interaction effects

A search of PubMed (https://pubmed.ncbi.nlm.nih.gov/) with the term *"gene-environment interaction" OR "gene-gene interaction" OR "epistasis"* identified a sizable 20,476 results (retrieved 3rd August 2022). However, the validity of previous gene-interaction studies has been questioned[59,60]. It is thought low replicability is due to low power and increased type I error rate as a consequence of multiple testing and publication bias[59,60]. Following are a small subset of gene-interaction findings that have been reported with evidence from more than one study and are therefore more likely to be valid[48].

The *M1CR* locus encodes a melanocortin receptor that controls the level of melanin found in skin which affects skin colour[61]. Melanin has a protective role against skin cancer from

UV exposure[61]. Previously studies have identified a gene-environment interaction effect of variation at *M1CR,* and skin cancer modified by sun light exposure[61]. Additionally, a second skin cancer predisposition gene, *CDK2NA* involved in cell cycling, has been reported to have a gene-gene interaction effect with *M1CR* on melanoma reducing the age of onset by up to 20 years[62,63].

Low dietary intake of folate and methionine combined with high intake of alcohol is a risk factor for colorectal cancer[64]. The enzyme methylenetetrahydrofolate reductase encoded by the *MTHFR* gene is responsible for metabolism of folate and methionine which are substrates for DNA synthesis[64]. Variation in *MTHFR* associated with reduced enzyme activity has an interaction effect on colorectal cancer risk modified by dietary levels of folate and methionine[64].

Circulatory low-density lipoprotein (LDL) concentration is a risk factor for cardiovascular disease (CVD)[65]. Apolipoprotein E (ApoE) activates lipoprotein receptors which leads to LDL uptake and removal by the liver[65]. LDL levels and CVD risk are also affected by dietary intake. Effect of dietary intake of LDL on CVD was found to vary by ApoE genotype[65] which highlights a production and clearance interaction effect.

Alzheimer's disease (AD) is associated with chronic inflammation mediated by microglia and astrocytes that release inflammatory cytokines such as IL-6 and IL-10[66]. Variation in the genes encoding these cytokines *IL6* and *IL10*, respectively are independently associated with AD risk[66]. Gene-gene interaction effects of *IL6* and *IL10* were shown to increase AD risk[66].

1.4.5 Qualitative interaction effects

In addition to estimating any genetic interaction effect(s) (**Chapter 1.4.1**), it is important to understand the qualitative nature of the interaction[57]. Qualitative interactions describe the phenomenon where the exposure has opposing effects on the outcome between subgroups which may be of equal or differing magnitude[57]. This could be across groups of a categorical modifier or quantiles of a continuous modifier. These can be identified by examining the effect of a SNP on outcome across levels of the modifier[57]. As discussed above, a SNP with an interaction effect on an outcome may only have an effect on an outcome within subgroups of individuals who are exposed to an environmental factor or possess specific genotypes[57]. Alternatively, a SNP may act in the same direction on an outcome within all subgroups, but with different sizes of effect[57].

1.4.6 Interaction effects on trait variance

Genetic loci with interaction effects (**Chapter 1.4.1**) on a trait may observed by their association with trait variance[56]. The effect of an interaction on trait variance can be demonstrated using simulated data, as shown in **Figure 1.4.6.1**. Here, series A has a SNP main effect only and the outcome is linearly associated with a unit increase in exposure. This produces an effect with constant variance (also known as homoscedasticity). Meanwhile, series B has a SNP main and interaction effect and consequently the SNP is associated with outcome mean and variance (also known as heteroscedasticity). Loci that associate with trait variance are known as variance quantitative trait loci[67] (vQTL; **Chapter 1.5**).

Figure 1.4.6.1 SNP effect on trait mean and variance under interaction

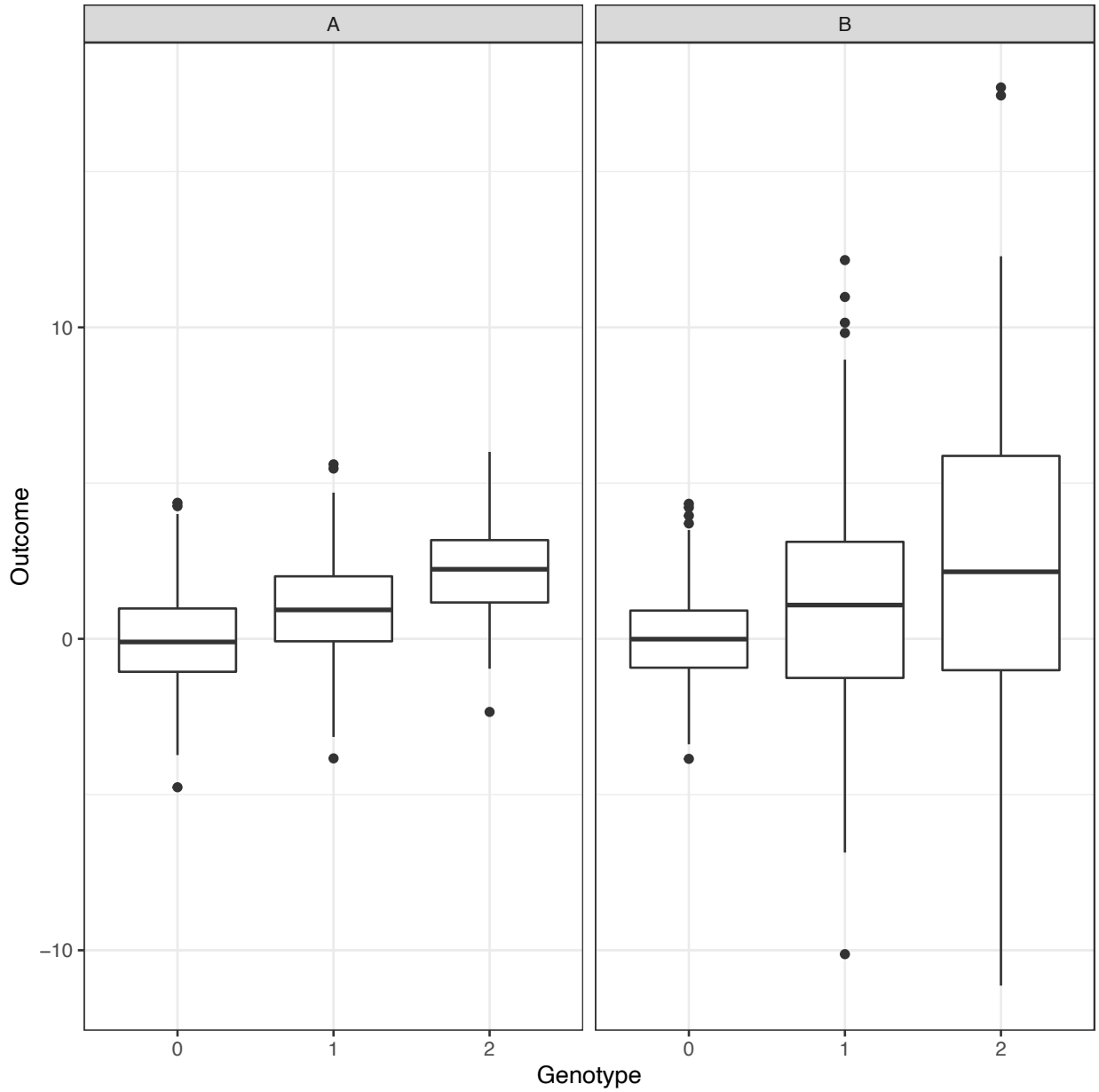


Illustration of SNP effect on trait median and interquartile range under main effect only (A) or main and interaction effect (B). In both cases, the simulated SNP x was drawn from $binom(2,0.33')$ and effect modifier u from $N(0,1)$ set to have main effects of one. The outcome was simulated to have main effects only (A) with $y = x + u + \epsilon$ or main and interaction effects (B) with $y = x + u + x \times u \times 2 + \epsilon$ where the error term ϵ was drawn from

the standard Normal distribution. A, SNP has main effect only and homoscedastic error. B, SNP has main and interaction effect with heteroscedastic errors. SNP, single nucleotide polymorphism allele dosage.

1.5 Variance QTL analysis

1.5.1 Statistical tests for detecting vQTLs

Whereas GWAS (**Chapter 1.2.2**) estimates the SNP mean effect on trait outcome typically applying linear or logistic regression[7], a range of statistical tests have been proposed to detect SNP association with trait variance[68–70] (henceforth variance GWAS [vGWAS]; **Table 1.5.2.1**). Several popular variance tests applied to GWAS are described as follows and discussed in relation to previous studies in Chapter 1.5.2 with summary in Table 1.5.2.1.

Equation 1.5.1.1 The Brown-Forsythe test

The Brown-Forsythe test[71] evaluates trait variance across categorical groups producing test statistic W which follows an F-distribution with $k-1$ and $N-k$ degrees of freedom.

$$W = \frac{(N-k)}{k-1} \times \frac{\sum_{j=1}^{k} N_j \left(\bar{Y}_j - \bar{Y}\right)^2}{\sum_{j=1}^{k} \sum_{i=1}^{N_j} \left(Y_{ji} - \bar{Y}_j\right)^2}$$

Where N is the sample size, k is the number of groups, N_j is the sample size in the jth group. Y_{ji} is the outcome absolute residual for the ith observation in the jth group from within group median. $\bar{Y}_j$ is the mean of Y_{ji} for the jth group and $\bar{Y}$ is the mean of $\bar{Y}_j$ across groups. Since the Brown-Forsythe test compares outcome variability among a categorical exposure, the test cannot be applied to imputed genotype data and cannot adjust for covariates (although the outcome may be pre-adjusted or stratified to account for confounding). The test also has no variance effect estimate and does not assume linearity between exposure and outcome. The Brown-Forsythe test is very similar to Levene's test[72], which estimates the outcome residual from the outcome mean rather than median as by Brown and Forsythe.

Equation 1.5.1.2 Bartlett's test

Bartlett's test[73] evaluates the null hypothesis of homoscedasticity of outcome variance among categorical groups indexed by i.

$$T = \frac{(N-k)\ ln\left(\hat{\sigma}_p^2\right) - \sum_{i=1}^{k}(N_i - 1)\ ln(\sigma_i^2)}{1 + \frac{1}{3(k-1)}\left(\sum_{i=1}^{k}\left(\frac{1}{N_i - 1}\right) - \frac{1}{N-k}\right)}$$

Where N is the sample size, k is the number of groups, σ_p^2 is the population variance and σ_i^2 is the within group variance. The test statistic T follows the chi-square distribution with $k-1$ degrees of freedom. As with the Brown-Forsythe test, Bartlett's test cannot adjust for covariates or be applied to imputed genotype data where the genotype takes a continuous value. This test requires no linearity assumptions and does not produce a variance effect estimate.

Equation 1.5.1.3 Fligner-Killeen test

The Fligner-Killeen test[68,74] is a rank score based test of homoscedasticity among categorical groups producing a chi-squared statistic with $k-1$ degrees of freedom.

$$\chi_{k-1}^2 \sim \frac{\Sigma_{i=1}^{k} N_i(\bar{A}_i - \bar{a})^2}{V^2}$$

Where N_i is the sample size for the ith group across k number of groups, α is a rank score obtained using Φ^{-1}, the standard Normal quantile function: $\Phi^{-1}(\frac{1+\frac{j}{n+1}}{2})$. Where j is the rank of the absolute residual of $|y_{ij} - \tilde{y}_i|$ and y_{ij} is the outcome for the jth individual belonging to the ith group and $\tilde{y}_i$ is the outcome median for the ith group. $\bar{A}_i$ is the mean of a for the ith group. $\bar{a}$ is the mean of a across all groups. As with the Brown-Forsythe test, the Fligner-Killeen test cannot be applied to imputed genotypes or adjust for covariates. The test provides no variance effect estimate and has no linearity assumptions.

Equation 1.5.1.4 Double generalized linear model

The double generalized linear model[75,76] (DGLM) is a regression-based test of the exposure effect on trait mean and variance assuming linearity. Given a normally distributed outcome vector Y and single genotype locus vector X, the DGLM maximises the likelihood to solve:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

$$\epsilon \sim N(0, \exp(\gamma_0 + \gamma_1 X))$$

Where β_0 and γ_0 are the mean and variance intercept terms, and β_1 and γ_1 are the estimated SNP effects on trait mean and variance. The effect of γ_1 can be interpreted as the log-linear effect of X on the variance of Y and tested using the Wald ratio to produce a chi-squared test statistic with one degree of freedom. As the DGLM is set within the regression framework, the test can be applied to imputed genotype data and adjust for confounding by providing additional covariates in the mean and/or variance portion of the model. As DGLM assumes linearity, dominant inheritance could introduce heteroscedasticity, as the effect of X on the mean of Y is non-linear producing a vQTL but this does not imply the presence of an interaction.

Equation 1.5.1.5 Heteroskedastic linear model

The heteroskedastic linear model[69] (HLM) extends the DGLM (**Equation 1.5.1.4**) to test linear and non-linear effects of genotype vector X on the variance of outcome vector Y.

$$Y = \beta_0 + \beta_1 X + \epsilon$$

$$\epsilon \sim N(0, \exp(\gamma_0 + \gamma_1 X + \delta(X - 2f)^2))$$

As above (**Equation 1.5.1.4**), where f is the genotype frequency of X and δ is the quadratic effect of X on the variance of Y. The HLM test is performed to jointly test for both mean and variance effects of X on Y by comparing the model likelihood to a null model using the likelihood ratio test.

The HLM further extends the DGLM to model the case where Y is non-normal. First, the outcome is inverse-rank normal transformed. As this transformation induces a mean-variance relationship (**Chapter 1.5.4**), SNPs with a mean effect will have an apparent variance effect leading false-positive vQTL effects that are not related to interaction effects. Second, the mean-variance confounding relationship is removed to estimate the genotype variance effect independent of the mean-variance relationship as follows:

$$\hat{\phi} = \hat{\gamma}_1 - \rho\hat{\beta}_1, \hat{\gamma}_1\hat{\beta}_1$$

Where $\hat{\phi}$ is an estimate of the effect of X on the variance of Y independent of mean-variance confounding, and $\rho\hat{\beta}_1, \hat{\gamma}_1$ is the correlation between estimated SNP mean and variance effects on the outcome.

Equation 1.5.1.6 Deviation regression model

The deviation regression model (DRM)[70] uses a linear model to regress the genotype on the absolute deviation of the phenotype from the within-genotype median. For the ith genotype from the jth individual the outcome is denoted with Y_{ij} and the within-genotype outcome median with $\tilde{Y}_i$. The deviation is estimated using $Z_{ij} = |Y_{ij} - \tilde{Y}_i|$ which is regressed on the genotype to provide an estimate of the genotype effect β on outcome variability with $Y_i = \beta Z_{ij} + \epsilon$. Covariates are adjusted by regressing Y on a set of covarites and taking the predicted value before applying the DRM. As with DGLM (**Equation 1.5.1.4**), the DRM assumes additive

linearity and produces a variance effect estimate. However, the DRM cannot be applied to imputed genotype dosages due to the requirement of a categorical exposure.

Equation 1.5.1.7 Two-step squared residual

The two-step squared residual (TSSR) approach[77] first adjusts the outcome Y for covariates to estimate the predicted value $\hat{Y}$ which is inverse rank normalised and then squared to produce Z^2. The value of Z^2 is regressed on the genotype using a second linear model:

$$Z^2 = \beta_0 + X\beta_1 + \epsilon$$

Where β_0 is the intercept, ϵ is the residual variance and β_1 is the effect of the genotype on trait variance. As with the DGLM, the TSSR test can be applied to imputed genotype data, adjust for covariates, and produces a variance effect estimate. The test also assumes linearity of the effect of X on the variance of Y.

Equation 1.5.1.8 Breusch-Pagan test

The Breusch-Pagan test[78] (also described as the squared residual value linear modelling [SVLM][79] test) is a two-stage regression-based test for heteroscedasticity and may adjust for covariates in each model.

$$Y = \beta_0 + \beta_1 X + U$$

$$\widehat{U}^2 = \gamma_0 + \gamma_1 X + E$$

Where Y is a continuous outcome vector and X is an exposure vector that may be continuous or categorical and is suitable for application to imputed dosage genotypes. U and E denote the residual variance of first and second-stage regression models. β_0 and γ_0 are the intercept terms estimating the mean and variance of Y when $X = 0$. β_1 and γ_1 are the effects of a one unit

increase in X on the mean and variance of Y. This test assumes a linear relationship between exposure and outcome variance.

Equation 1.5.1.9 Omnibus likelihood ratio test

The likelihood ratio test (LRT) first estimates the linear effect of genotype exposure X on outcome Y:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

Where X_1 and X_2 are indicator (dummy) variables for the SNP minor allele count taking a value of $X_1 = 1$ and $X_2 = 2$.

Second, the residual variance is parameterised as follows:

$$\epsilon \sim N(0, \gamma_1 X_1 + \gamma_2 X_2)$$

The likelihood ratio test evaluates the joint null H_0 effect of:

$$H_0: \beta_1 = \beta_2 = \gamma_1 = \gamma_2 = 0$$

The LRT employs the regression framework which facilitates adjusting of covariates and produces an effect estimate. Since the genotypes are used as dummy variables, the test makes no linearity assumptions of X on Y but cannot be applied to imputed dosage values.

Equation 1.5.1.10 Joint location-scale score test

The joint location-scale score test (JLSc)[80] simultaneously compares the exposure effect on outcome mean and variance using a score test[80]. The mean and variance models follow a similar structure to the Breusch-Pagan test[78].

$$Y = \beta_0 + \beta_1 X + U$$

$$\widehat{U}^2 = \gamma_0 + \gamma_1 X + E$$

Where Y is the outcome, X is the genotype and β_1 is the exposure effect on trait mean and $\gamma_1$is the exposure effect on trait variance. Then, a score test is used to evaluate the joint effect of β_1 and γ_1 on the outcome evaluating the null hypothesis of:

$$H_0: \beta_1 = \gamma_1 = 0$$

As with the DGLM (**Equation 1.5.1.4**), the JLSc assumes linearity of X on Y, may be applied to imputed dosage genotype values, produces an effect estimate on trait mean and variance and can be adjusted for covariates. As this is a joint-test, SNPs with mean or variance effect on the outcome will reject the null hypothesis.

1.5.2 Previous studies of statistical tests for detecting vQTLs

Two major studies evaluated the performance of several variance tests through simulation[68,70]. One of these, Wang *et al*[68], compared the power and type I error rate of the Brown-Forsythe test[71] (**Equation 1.5.1.1**), Bartlett's test[73] (**Equation 1.5.1.2**), Fligner-Killeen test[74] (**Equation 1.5.1.3**) and double generalised linear model[75] (**Equation 1.5.1.4**; DGLM) to detect interaction effects[68].

Brown-Forsythe, Bartlett's, and Fligner-Killeen tests compare trait variance among genotypic groups using non-parametric statistical tests. These tests do not allow adjustment for covariates, continuous genotypes (i.e., imputed) or provide an effect estimate which are limitations when applied to GWAS[7].

Meanwhile, DGLM jointly models the effect of exposure on the mean and dispersion of an outcome using the generalised linear model framework[75]. Mean and variance effects are estimated using the maximum likelihood estimation that iterates until convergence to first estimate the mean effects and then the variance effects[75].

Wang *et al* found the Brown-Forsythe test was most robust to non-normality having no type I error inflation even in the presence of trait skew and kurtosis with or without mean effects[68]. This may be because the median is a more robust measure of central tendency in the presence of skew or kurtosis[81]. Meanwhile, Bartlett's test, DGLM and Fligner-Killeen test had high type I error rate when the trait was skewed[68]. Another study also replicated this finding for Bartlett's test[82].

Since Bartlett's, DGLM and Fligner-Killeen tests are susceptible to type I error rate under non-normality, Wang *et al* also investigated a range of non-linear scale transformations which aimed to mitigate skew and kurtosis. However, non-linear scale transformations are also known to induce a relationship between the mean and variance and therefore induce mean-variance confounding[68] (**Chapter 1.5.4**). The non-linear scale transformations led to increased type I error rate particularly when the SNP had a strong mean effect[68] leading the authors to suggest that these transformations should be avoided[68].

An alternative strategy uses the heteroskedastic linear model (HLM; **Equation 1.5.1.5**)[69] which implements a log-linear model using DGLM on inverse-rank normal transformed traits (to avoid problems with trait non-normality), and then applies an additional model to remove the mean-variance relationship induced by scale transformation and test for the remaining variance effect (known as dispersion effect)[69]. The mean-variance relationship is estimated and removed by regressing the genome-wide variance effect on the additive effects[69]. This approach was shown to provide similar type I error rate and power to the Brown-Forsythe test[68].

A second major study conducted by Marderstein *et al*[70] compared the performance of a range of tests through simulation including the deviation regression model[70] (DRM; **Equation**

1.5.1.6), Brown-Forsythe test, Levene's test[72], regression-based Levene's test[83], two-step squared residual (TSSR; **Equation 1.5.1.7**)[84], squared residual value linear modelling (SVLM; **Equation 1.5.1.8**)[79], DGLM (**Equation 1.5.1.4**), Bartlett's test (**Equation 1.5.1.2**), and Fligner-Killeen test (**Equation 1.5.1.3**).

DRM is a regression-based implementation of the Brown-Forsythe test which first regresses the outcome on covariates using ordinary least squares (OLS) and then takes forward the residuals of this fit. The second step involves estimating the within-genotype median absolute deviation (MAD) using the first-stage fit residuals which is then regressed on to the SNP to estimate the effect of the SNP on outcome variability[70].

TSSR first estimates the Z-score for each individual which is a measure of standard deviation units from the mean and then regresses the square Z-score on the SNP using an OLS model to estimate the effect of the SNP on trait variance[84].

SVLM[79] estimates the effect of a SNP on trait variance by first regressing the trait on the SNP and then regressing the SNP on the squared residuals of the first model through a second linear model. This model is also known as the Breusch-Pagan test[78] and regression-based Levene's test[72].

Marderstein *et al*[70] found Levene's test[72], DGLM, Fligner-Killeen test and Bartlett's test had elevated type I error rate when applied to non-normal outcomes. SVLM was shown to have lower power than the Brown-Forsythe and DRM tests[70]. TSSR had elevated type I error rate when the SNP had a mean effect[70]. This study also found that DGLM and regression based Levene's test had the slowest runtime and may be difficult to scale[70]. HLM is also based on the DGLM which may be difficult to scale genome-wide to multiple traits.

Joint location and scale tests have also been proposed (**Equation 1.5.1.9** and **Equation 1.5.1.10**)[80,85] but these could potentially lead to large numbers of variants for downstream testing of genetic interaction effects obfuscating any benefits in power obtained compared with an exhaustive pairwise analysis of all possible interaction analyses.

One such example of a joint test, the omnibus likelihood ratio test (LRT; **Equation 1.5.1.9**)[86] evaluates the SNP joint effect on trait mean and variance through comparison with a null model. The authors suggest performing a joint test for mean and variance effects as an approach to find SNPs that may be involved in genetic interaction. However, this approach has high type I error rate in the presence of non-normality[86].

Of the previously reported tests described in this section, the Brown-Forsythe test, DRM and HLM performed best with the lowest type I error rates and highest power.

Table 1.5.2.1 Statistical tests to identify exposure effect on trait variance

Test	Model	Assumptions	Limitations
Bartlett's test[73]	Non-parametric test for variance among genotypic groups	Continuous outcome and categorical exposure	Cannot model covariates or imputed genotypes. Elevated type I error rate for non-normal outcomes[68,70,82]
Breusch-Pagan test[78]	Linear regression	Outcome normality. Linear additivity of first stage-model.	Lower power than the Brown-Forsythe and deviation regression model tests[70]
Brown-Forsythe test[71]	Non-parametric test for variance among genotypic groups	Continuous outcome and categorical exposure	Cannot model covariates or imputed dosage genotypes. No variance effect estimate.
Deviation regression model[70]	Linear regression	Continuous outcome and categorical exposure. Linear exposure-outcome effect	Variance effect estimate makes linearity assumptions[70]
Double generalised linear model[75]	Generalised linear regression	Outcome normality. Linear exposure-outcome effect	Elevated type I error rate for non-normal outcomes[68,70]. Slow runtime and may be difficult to scale[70]
Fligner-Killeen test[74]	Non-parametric test for variance among genotypic groups	Continuous outcome and categorical exposure	Cannot model covariates or imputed genotypes. Elevated type I error rate for non-normal outcomes[68,70]
Heteroskedastic linear model[69]	Log-linear regression	Assumes linear mean-variance relationship after applying rank normal transformation	Based on double generalised linear model which has slow runtime and may be difficult to scale[70]
Joint location-and-scale score (JLSsc) test[80]	Least absolute deviation regression	Linear exposure-outcome effect. Assumes P-values for	Cannot separate mean and variance effects without

		location and scale test are independent	additional test. May result in large numbers of loci for follow up
Levene's test[72]	Non-parametric test for variance among genotypic groups	Outcome normality and categorical exposure	Cannot model covariates or imputed genotypes. Elevated type I error rate for non-normal outcomes[70]
Likelihood ratio test[86]	Linear regression model	Linear exposure-outcome effect for variance estimate	Elevated type I error rate in the presence of non-normailty[86]
Regression-based Levene's test[83]	Linear regression model	Outcome normality. Linear exposure-outcome effect	Slow runtime and may be difficult to scale[70]
Squared residual value linear modelling[79]	Linear regression model	Outcome normality. Linear exposure-outcome effect	Lower power than the Brown-Forsythe and deviation regression model tests[70]
Two-step squared residual test[84]	Linear regression model	Assumes linearity of exposure-outcome effect	Elevated type I error when the SNP has main effect[70]

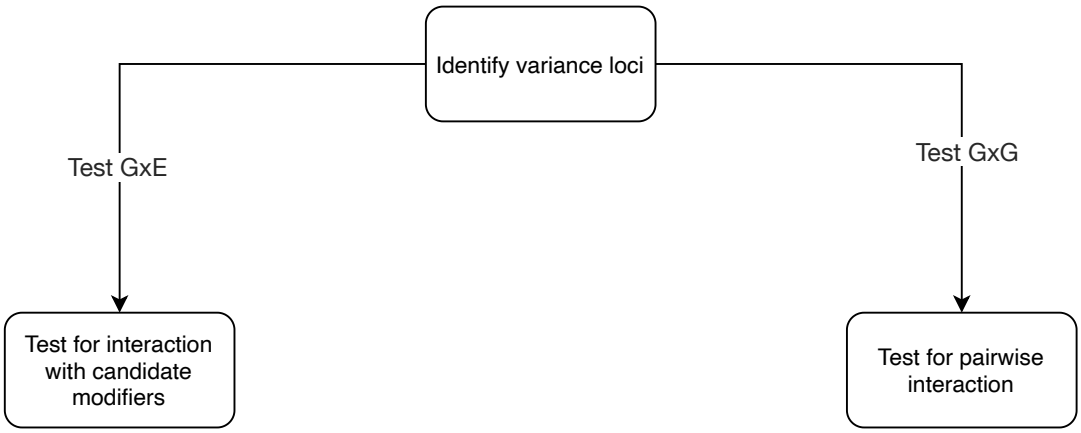
1.5.3 Application of vQTLs

To my knowledge, the only current application of vQTLs is in the detection of genetic interaction effects. This approach involves screening for vQTLs to direct further analyses using formal interaction tests using a set of candidate modifiers, and is known as variance prioritisation[56,70,87] (**Figure 1.5.3.1**). A vQTL finding is not conclusive evidence for interaction but it is consistent with a SNP-interaction effect[67] and since detection of vQTLs does not require the modifier to be measured or hypothesised[67] this approach could lead to unanticipated findings and detection of novel biology[56]. This approach can narrow the search for testing of genetic interaction effects, thereby increasing power compared with exhaustive pairwise testing of every possible interaction effect[87].

In humans the seminal paper by Yang *et al*[77] identified strong positive effects on body mass index (BMI) mean and variance of *FTO* rs7202116. More recently, Wang *et al*[68] performed systematic testing of 13 physical traits in UK Biobank and identified 75 vQTLs. These were investigated and led to the detection of 16 GxE effects modified by age, sex, physical activity, sedentary behaviour, and smoking[68]. Among these was an effect of *CHRNA5-A3-B4* rs56077333 on lung function. SNP rs56077333 was strongly associated with smoking heaviness but only weakly with smoking initiation[88] and is anticipated to adversely affect lung function only among those who smoke[68]. Wang *et al* also reported vQTL effects of *WNT16-CPED1* rs10254825 on bone mineral density which interacts with age, and this has been supported by studies in mice[68]. Thirdly, Wang *et al* reported strong vQTL evidence at the *FTO* locus on measures of adiposity which was shown to interact with physical activity and sedentary behaviour[68]. Variance QTL effects have also been identified for gene expression[89], DNA methylation[80],

Vitamin D[90] and facial morphology[91]. During thesis preparation, a study reported variance GWAS of 20 serum biomarkers in UK Biobank identifying 182 vQTLs which were tested for interaction with 2,380 candidate modifiers identifying 846 GxE effects[42].

Figure 1.5.3.1 Flow diagram for detecting interaction effects using variance prioritisation



Schema for detecting genetic interaction effects from vQTLs. GxE, Gene-environment interaction effect. GxG, gene-gene interaction effect.

1.5.4 Mean-variance confounding

Variance QTL effects are susceptible to bias by mean-variance confounding **(Chapter 1.3.2**) which can occur when the trait variance is related to its mean and is typical for non-normally distributed phenotypes[69]. This implies that a variant that has an effect on the trait mean will also appear to have an effect on the variance[69]. One approach to determine if vQTLs are due to mean-variance confounding is to repeat the analysis after applying a scale transformation to the trait[92]. If the effect on variance disappears this could suggest mean-variance confounding is responsible for the association[92]. If the effect remains, then the association may not be entirely driven by mean-variance confounding[92]. For example as a sensitivity analysis the log-scale transformation may be applied to the outcome (e.g., natural logarithm) to consider scale dependency of variance effects, but non-linear transformations induce a mean-variance relationship[68] and should be avoided for main analyses.

1.5.5 Phantom effects

Phantom effects are another source of bias that can affect genetic testing of variance or interaction effects[53,68,93,94]. Suppose SNP $G1$ is causally associated with outcome Y and imperfectly correlated $0 > R^2 < 1$ with SNP $G2$ not having a causal effect on Y, then $G2$ will have an apparent vQTL effect on Y even in the absence of interaction effects[93]. Furthermore, another SNP $G3$ that also has no causal effect on Y but is in imperfect LD $0 > R^2 < 1$ with $G1$ will have an apparent interaction effect with $G2$ on Y even though all effects are purely additive[93,94]. This is because under phantom effects, the residual error of Y is a mixture of normal and binomial distributions causing type I error inflation of parametric tests[53]. Simply testing for correlation between $G2$ and $G3$ as potential evidence of phantom interaction is

insufficient as $G2$ and $G3$ may both be more strongly correlated with $G1$ than each other[94]. Nevertheless, adjusting $G1$ in the interaction model will attenuate the interaction effect of $G2 \times G3$[94]. Therefore, one suggested approach to mitigate phantom effects is to adjust for fine mapped main effects[53,68] (**Chapter 1.2.4**) of $G2$ and $G3$ in an attempt to capture $G1$ (or a highly correlated proxy) which will attenuate phantom effects on Y. However, if there is measurement error of $G1$ this adjustment will only partially attenuate phantom effects[53]. Another approach to avoid this bias in GxG testing which may be used in combination with adjusting for fine mapped effects is to perform testing for GxG effects at least 10Mb apart[68] because, in most cases, variants at this distance will be in minimal LD[95].

1.5.6 Variance confounding by population stratification

Association of SNP and outcome mean may be biased by population stratification which describes the situation where ancestry confounds the SNP-outcome relationship (**Chapter 1.2.3**). A similar situation could potentially affect the variance association. Suppose the ith individual has SNP G_i and outcome Y_i which is confounded by ancestry A_i. Further, AU_i has an interaction effect on Y_i which is the interaction of A_i and effect modifier U_i. E_{1_i} and E_{2_i} are the residual variance. Under this situation which I describe as variance confounding by population stratification, the SNP is associated with the variance of Y_i without having an interaction effect.

Equation 1.5.6.1 Variance confounding by population stratification

$$G_i = A_i + E_{1i}$$

$$Y_i = A_i + U_i + AU_i + G_i + E_{2i}$$

1.5.7 Existing variance GWAS software

The OmicS-data-based Complex trait Analysis (OSCA) software package[96] has functionality to perform variance GWAS using a range of commonly used variance tests providing routines for performing Levene's test[72] (based on the mean or median [Brown-Forsythe test[71]]), Bartlett's test[73] and Fligner-Killeen test[74].

As discussed above, the Brown-Forsythe test has among the lowest type I error rate when applied to non-normal traits (**Chapter 1.5.2**). However, the Brown-Forsythe test also has some limitations. First, the test requires comparison of trait variability among categorical genotype groups[71]. This prevents adjustment for covariates such as genetic ancestry to remove confounding by population stratification (**Chapter 1.2.3**) or age and sex to reduce unexplained variance in order to increase test power[67]. But this method could be applied to a pre-adjusted trait[68]. Secondly, this approach prevents application to imputed genotypes unless values are rounded to a whole number which may result in loss of information[67]. Third, the Brown-Forsythe test provides no effect estimate[71]. The OSCA implementation of the Brown-Forsythe test provides an effect estimate derived from the test p-value[26,96] but this assumes linearity between SNP and trait variance, which may not hold (**Chapter 9.1**). Fourth, the model cannot include a random effect to allow modelling of polygenic effects. To overcome this, the phenotype could potentially be pre-adjusted using a model containing a random subsample of SNPs that capture genetic ancestry[13,67].

The DRM addresses some of these limitations by providing functionality to adjust for covariates and produces an effect estimate but this approach is similar to the effect derived from the test p-value using OSCA in that both assume linearity between SNP and outcome

variance, which may not hold[56] (**Chapter 9.1**). the DRM model is implemented in R and may not easily scale to genome-wide analyses of multiple traits[70].

From my review of the literature, there is a need for a regression-based Brown-Forsythe test that produces an unbiased variance effect estimate and is scalable to enable vGWAS analyses of multiple traits.

1.5.8 Limitations of previous vQTL analyses

Following identification of a vQTL, current studies apply formal interaction testing using a set of candidate modifiers and report on potential genetic interaction effects[68,69] (**Chapter 1.5.3**). However, an obvious question remains – do other interaction effects exist or have all interaction effects been identified (subject to power). One way to investigate this further could be to adjust the variance model for the identified interaction effect(s) and measure attenuation of the vQTL. If the vQTL only partially attenuates then there could be other interactions that remain, but these are currently not identified.

A second important question that has received little attention in the literature – does adjustment for confounding (**Chapter 1.3.2**) of the mean effect also control confounding of the variance effect. It may be because the variance effect cannot be adjusted using many of the proposed variance tests (**Table 1.5.2.1**). Often variance testing is applied to outcomes that are pre-adjusted using linear regression, but this only adjusts the association with outcome mean and not also variance.

Third, do vQTLs have additional utility beyond the detection of interaction effects. These signals provide evidence of the net effect of effect modification at a locus and aside from characterising the exact interaction effect, this evidence could be used as a general measure of

the presence of effect modification. In RCTs effect modification of the treatment-outcome relationship limits generalisability of findings to other populations where the interaction effect does not hold[97]. One area vQTL evidence may be useful is in evaluating instrumental variable homogeneity assumptions which affect interpretation of the causal estimate[31,35] (**Chapter 1.3.5**).

Finally, previous vQTL studies have largely focused on physical measures such as BMI, height, lung function, and bone mineral density[68,69]. However, variance studies of molecular biomarkers may provide findings of increased translational value (**Chapter 1.6**).

1.6 Molecular biomarkers

1.6.1 Background

Biomarkers (biological markers) describe a group of measures that provide accurate, objective and reproducible evidence on biological systems and processes[98]. Molecular biomarkers are a subgroup of these measurements that provide evidence on molecular processes and include gene expression levels and metabolite and protein concentration[99].

Biomarker measures are often used as 'surrogate endpoints' in clinical studies meaning that they provide evidence in place of outcomes of interest that are more difficult to measure such as stroke, myocardial infarction, and diabetes[98]. While many biomarkers are causally related to their surrogate outcome they need not be[98]. Where a biomarker is causally related (**Chapter 1.3.4**) to the outcome, then therapeutic interventions may be developed that target the biomarker to reduce incidence or improve prognosis[98]. Meanwhile, where a biomarker is only predictive of the outcome, then interventions will not impact on the outcome[98].

1.6.2 Application of biomarkers in drug development

Studies on the causative nature (**Chapter 1.3.4**) of biomarkers in disease can provide valuable information on disease biology and lead to the identification of risk factors along the causal pathway which may be useful for developing interventions[99]. This approach has led to the development of therapies for lipids, glucose, and urate in the treatment of cardiovascular disease[100], type 2 diabetes[101], and gout[102], respectively, among others. Usually drugs target the abundance of a protein and drug development studies are beginning to incorporate genetic evidence[99]. The aim of which is to identify proteins that modulate risk factors having a causal effect on disease (such as blood biomarkers)[99]. For example, statins act to inhibit HMG-CoA reductase lowering serum LDL cholesterol levels which is protective against cardiovascular disease[99].

One approach is the use of MR (**Chapter 1.3.7**) to estimate the causal effect of protein concentration on a disease outcome[103]. To avoid violation of the exclusion restriction assumption by horizontal pleiotropy, genetic instruments for protein concentration are selected near to the gene coding region (also known as *cis*-acting instruments)[103]. These instruments are identified through GWAS (**Chapter 1.2.2**) of protein concentration[103].

1.6.3 Utility of detecting gene-interaction effects on biomarker concentration

GWAS of mean effects (**Chapter 1.2.2**) on biomarker concentration have been used to identify genes and therefore proteins that may be useful targets for drug development[99]. Identification of loci with biomarker interaction effects may provide evidence of drug targets, that upon intervention, produce subgroup effects with individual variation in response to treatment dependent on the modifier[57]. However, evidence of such may be difficult to obtain

since gene-interaction effects are generally small and studies are often underpowered for detection[57].

1.7 Thesis aims

Genetic epidemiological studies have largely focused on SNP mean effects while comparatively few studies have investigated variance effects (i.e., vQTLs)[56]. This thesis aims to develop methodology and software to improve the discovery, analysis and sharing of vQTL data to promote and facilitate usage of this type of genetic evidence in future studies.

These tools and methods will be applied to studies of 30 serum biomarkers in UK Biobank as an exemplar but could be applied to any continuous trait. I chose these outcomes because they act as surrogate endpoints for disease outcomes (**Chapter 1.6.1**), and they are continuous which is a requirement for the described variance tests (**Chapter 1.5.1**).

First, I aim to implement and evaluate a regression-based implementation of the Brown-Forsythe test (LAD-BF) for testing variance effects which can estimate the variance effect size and adjust for covariates. Second, I aim to apply LAD-BF to identify vQTLs and genetic interaction effects on 30 serum biomarkers in UK Biobank. Third, I aim to develop methods to use LAD-BF for falsification tests of the assumption of homogeneity in MR (**Chapter 1.3.7**). Fourth, I aim to develop a data sharing standard to facilitate and promote sharing and secondary analysis of vQTLs and GWAS summary statistics more widely.

Chapter 2: Methods

2.1 Contribution statement

The Brown-Forsythe test, LAD-BF test and gene-interaction test forms part of a manuscript I wrote that was edited by PhD supervisors available as a preprint on MedRxiv (Lyon *et al*, 2022)[2].

2.2 Statistical analysis

Unless stated, the threshold for statistical significance was set to $\alpha = 0.05$ throughout.

2.2.1 Simulation studies

Simulation studies are *in silico* experiments performed using data drawn from known probability distributions[104]. In this thesis I used simulation studies to test assumptions and evaluate and characterise the performance and limitations of methodologies including an appraisal of when approaches 'break' or fail[104]. Throughout this thesis I evaluated analysis methodology using simulation and evaluated several parameters as follows. Power was defined as the proportion of tests that rejected the null hypothesis when an alternative hypothesis is true[55]. Type I error rate was defined as the proportion of tests that rejected the null hypothesis when the null hypothesis was true[55]. Coverage was defined as the proportion of confidence intervals that contained the true value (which may be theoretically known from the data generating mechanism, or may be estimated in the simulation process)[104]. Absolute bias was defined as the residual of the estimate from its expected (true) value[104].

I followed best practise guidance for planning, programming, analysis, and presentation of simulation results[104]. I reported simulation study designs using the aims, data-generating mechanism, estimand, methods, performance (ADEMP) structure[104].

2.2.2 Brown-Forsythe test

The Brown-Forsythe test[71] (also known as the median variant of Levene's test[72]; **Equation 2.2.2.1**) refers to the original published non-parametric test and will be used throughout. I applied the Brown-Forsythe test to detect differences in outcome variability across the three genotypic groups.

All analyses of the original Brown-Forsythe test were conducted using the OSCA software package[68,96] which additionally produces a variance effect estimate derived from the test P-value[26] (**Equation 2.2.2.1**). This derived estimate assumes linearity between the SNP and outcome variance[26] although the test itself does not make linearity assumptions.

Equation 2.2.2.1 The Brown-Forsythe test

The Brown-Forsythe test evaluates trait variance across genotype groups.

$$W = \frac{(N-3)}{2} \times \frac{\sum_{G=0}^{2} N_G(\bar{Y}_G - \bar{Y})^2}{\sum_{G=0}^{2}\sum_{i=1}^{N_G}(Y_{Gi} - \bar{Y}_G)^2}$$

Where N is the total number of observations. N_G is the number of observations in the Gth genotype group where $G \in \{0, 1, 2\}$ is the count of the minor allele. Y_{Gi} is the absolute residual of the outcome for the ith observation in the Gth genotype group from the outcome median in that group. $\bar{Y}_G$ is the mean of Y_{Gi} for the Gth genotype group and $\bar{Y}$ is the mean of $\bar{Y}_G$ across genotype groups. The test statistic W is F-distributed $F(2, N-3)$.

All analyses of the original Brown-Forsythe test were conducted using the OSCA software package[68,96] which additionally produces a variance effect estimate derived from the test P-value[26] (**Equation 2.2.2.2**). This derived estimate assumes linearity between the SNP and outcome variance[26] although the test itself does not make linearity assumptions.

Equation 2.2.2.2 Brown-Forsythe effect estimate

First, the Brown-Forsythe test P-value was converted to a Z-score, Z, then the linear effect β of G on the variance of Y was calculated using the following formula along with its standard error.

$$\beta = Z/\sqrt{2MAF(1-MAF)(N+Z^2)}$$

$$SE(\beta) = 1/\sqrt{2MAF(1-MAF)(N+Z^2)}$$

Where MAF is the SNP minor allele frequency and N is the sample size.

2.2.3 Breusch-Pagan test

The Breusch-Pagan test[78] (**Equation 2.2.3.1**) was applied to test for an effect of a SNP on the variance of a continuous outcome through the use of two OLS regression models.

Equation 2.2.3.1 Breusch-Pagan test

First, the vector of outcomes Y was regressed on the vector of minor allele counts G adjusting for any covariates to estimate the residuals U_{OLS} and per minor allele average effect β_{OLS_1} of G on Y. The intercept effect was denoted by β_{OLS_0}.

$$Y = \beta_{OLS_0} + \beta_{OLS_1}G + U_{OLS}$$

A second OLS model then regressed the squared residual $\widehat{U}^2{}_{OLS}$ of the first-stage model on G and the square of the genotypes G^2 including any covariates to estimate the average variance effect $g_1$and g_2 of G on Y and second-stage model residual variance E_{OLS}. The intercept effect was denoted by g_0.

$$\widehat{U}^2{}_{OLS} = g_0 + g_1G + g_2G^2 + E_{OLS}$$

Significance testing was performed using a F-test comparing the second-stage residual sum of squares to a restricted model lacking G and G^2.

2.2.4 LAD-BF test

The least-absolute deviation Brown-Forsythe test (LAD-BF) was proposed by Professor Tilling. I implemented and evaluated this test throughout this thesis. The test used the same structure as the Breusch-Pagan test[78] (**Chapter 2.2.3**). Briefly, the Breusch-Pagan test estimates the variance effect by regressing the outcome on exposure and then regressing the squared residuals of this fit back on the exposure through a second model.

LAD-BF uses least-absolute deviation (LAD) regression[105] for the first-stage model which estimates the exposure effect on outcome median (rather than mean used in the Breusch-Pagan test). Therefore LAD-BF measures variability from the median which is more robust to non-normality[81] and consistent with the Brown-Forsythe test (**Equation 2.2.4.1**).

Equation 2.2.4.1. LAD-BF test

The vector outcome Y was regressed on vectors of dummy genotypes $G1$ and $G2$ representing one or two minor allele counts, respectively. This test used LAD regression adjusting for any covariates to estimate the residual U_{LAD} and average effects β_{LAD_1} and β_{LAD_2} of $G1$ and $G2$ on the median of Y. The intercept effect was denoted by β_{LAD_0}.

$$Y = \beta_{LAD_0} + \beta_{LAD_1} G1 + \beta_{LAD_2} G2 + U_{LAD}$$

Using OLS, a second regression model regressed the vector of absolute residuals $|\widehat{U}_{LAD}|$ estimated using this first-stage fit on $G1$ and $G2$ including any covariates. Genotypes were coded as dummy variables $G1$ and $G2$ to accommodate potential non-linearity with first-stage model residuals. The per allele effect estimates g_{OLS_1} and g_{OLS_2} measure the outcome mean absolute deviation from the median. The vector of second-stage model residuals was denoted with E_{OLS} and intercept effect denoted with g_{OLS_0}.

$$\left|\widehat{U}_{LAD}\right| = g_{OLS_0} + g_{OLS_1}G1 + g_{OLS_2}G2 + E_{OLS}$$

LAD-BF test P-values were estimated using an F-test comparing the second-stage residual sum of squares to a restricted model without genotypes to test the joint null hypothesis of outcome homogeneity across genotypic groups.

Each per allele genotype coefficient g_{OLS} was transformed to variance units $var(Y|G)$ as follows. First, the sum of covariance between the intercept g_{OLS_0} and coefficient term g_{OLS} and squared coefficient g_{OLS}^2 was estimated to produce a per-allele effect on the mean-absolute deviation of Y. Second, this measure of mean-absolute deviation was transformed to variance units.

$$var(Y|G) = 2 \times g_0 + g + g^2 \,/\, (2/\pi)$$

Throughout this study the transformation used was specific to the normal distribution (i.e., $2/\pi$ is used as the denominator as part of the Normal distribution probability density function) but this could potentially be extended to other accommodate other distributions.

The standard error of $var(Y|G)$ was calculated using the delta method[106] from second-stage model heteroscedastic-consistent standard errors (aka White standard errors)[107].

2.2.5 SNP interaction test

To estimate the interaction effect of SNP-by-modifier I used OLS regression including a single additive interaction term[4] (**Equation 2.2.5.1**).

Equation 2.2.5.1 Additive linear interaction test

The additive linear interaction effect was estimated using OLS regressing the vector of outcomes Y on vector of minor allele counts G, modifier vector E and interaction of genotype-by-modifier GE. The residual variance vector is denoted with U and intercept denoted by β_0.

$$Y = \beta_0 + \beta_1 G + \beta_2 E + \beta_3 GE + U$$

The genotype main effect β_1, modifier main effect β_2 and interaction effect β_3 were estimated along with their heteroscedastic-consistent standard errors using White *et al*[107] (**Chapter 2.2.7**). The interaction test P-value was obtained from a t-statistic of β_3 using heteroscedastic-consistent standard errors and compared to the expected null distribution.

2.2.6 F-test for comparing model fits

The fit of two nested regression models may be compared using an F-test (**Equation 2.2.6.1**) which tests for a difference in the residual sum of squares between models[108]. The residual sum of squares SS_1 SS_2 of the first and second regression models, respectively, are estimated by summing the squared difference between observed Y and predicted $\hat{Y}$ outcome. Since the number of parameters differ between models the degrees of freedom of both models df_1 and df_2 must be estimated by $df = N - V$ where N is the sample size and V is the number of parameters in the model. The F statistic is estimated by comparing model fits (**Equation 2.2.6.1**) and used to obtain a P-value for the difference in model fit.

Equation 2.2.6.1 F-test for comparing model fit

$$F = \frac{(SS_1 - SS_2)/(df_1 - df_2)}{SS_2/df_2}$$

2.2.7 Heteroscedastic consistent standard errors

A vQTL is a genotype that has an association with trait variance, that is, trait variance is non-constant among levels of the genotype[67]. This implies that regression analyses of the vQTL-outcome relationship will violate the homoscedasticity assumption of constant trait variance[107]. The exception is where the model correctly defines the term that fully explains the heteroscedasticity, but it is not possible to know this in advance. Therefore, it is important to

account for potential heteroscedasticity in regression models of vQTL effects. This can be achieved using heteroscedastic-consistent standard errors[107] (**Equation 2.2.7.1**).

Equation 2.2.7.1. Heteroscedastic-consistent standard errors

The unbiased effect β of exposure vector X on outcome vector Y can be estimated using OLS, but under heteroscedasticity the standard error of estimate $SE(\hat{\beta})$ will be biased. This bias may be attenuated using heteroscedastic consistent standard errors estimated as follows.

$$SE(\hat{\beta}_i) = \sqrt{\frac{1}{N} . \frac{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2 \hat{u}_i^2}{\left[\frac{1}{N}\sum_{i=1}^{N}(X - \bar{X})^2\right]^2}}$$

Where N is the sample size and $\hat{u_i}^2$ equal the squared residual of Y regressed on X.

2.2.8 Mendelian randomization

MR was implemented using the two-sample framework[20] where instrument-exposure and instrument-outcome associations were selected from GWAS summary statistics estimated in separate samples from the same population[20] (**Chapter 3**). Causal estimates were obtained using Wald ratio or inverse-variance weighting using the TwoSampleMR R-package[109].

2.2.9 Confidence interval

All confidence intervals produced in this thesis were estimated using the t.test function in R (v.3.6.0)[110].

$$CI = \bar{X} \pm Z\frac{S}{\sqrt{N}}$$

Where $\bar{X}$ is the mean of variable X, Z is the significance level set to $\alpha = 0.05$ throughout, S is the standard deviation of X and N is the sample size of X.

2.3 Software

The original Brown-Forsythe test used the OSCA software package v0.46[68,96]. Simulations and follow-up UK Biobank analyses were performed using R v3.6.0.

2.4 Code availability

The LAD-BF test implemented in C++ is available from https://github.com/MRCIEU/varGWAS and R version is available from https://github.com/MRCIEU/varGWASR. R code for performing the simulation studies in **Chapter 4** is available from https://github.com/MRCIEU/varGWAS/tree/master/sim.

R code for running the UK Biobank analysis in **Chapter 5** available from https://github.com/MRCIEU/varGWAS-ukbb-biomarkers.

R code for simulations, MR studies and UK Biobank analyses in **Chapter 6** are available from https://github.com/MRCIEU/variance-iv4-violation.

Python code to convert GWAS summary statistics to GWAS-VCF (**Chapter 7**) is available from https://github.com/MRCIEU/gwas2vcf. Python code for reading GWAS-VCF files is available from https://github.com/MRCIEU/pygwasvcf. Python code for running the webservice to convert GWAS summary statistics to GWAS-VCF is available from https://github.com/MRCIEU/gwas2vcfweb. R code for performing the GWAS-VCF simulations is available from https://github.com/MRCIEU/gwas-vcf-performance.

Chapter 3: Data sources

3.1 Contribution statement

Background on UK Biobank (including genetic data, quality control, and biomarkers) forms part of a manuscript I wrote that was edited by PhD supervisors available as a preprint on MedRxiv (Lyon *et al*, 2022)[2].

3.2 Introduction

In **Chapter 4**, **Chapter 5** and **Chapter 6** I apply analyses to data from UK Biobank. In **Chapter 6** I also use GWAS summary data from large consortia that were estimated using non-overlapping samples from UK Biobank. In **Chapter 7** I used publicly available GWAS summary statistics from UK Biobank.

3.3 UK Biobank

3.3.1 Background

UK Biobank is a large prospective cohort study of approximately 500,000 UK participants recruited between 2006-2010 from across the UK aged 37-73 at recruitment[111]. Phenotypic measures were made available on lifestyle, socio-demographics, health-related factors, and physical parameters including blood pressure, lung function, anthropometry, bone density, hearing and eye measures and cardiorespiratory fitness among others[111]. Blood, urine, and saliva samples were also collected which have been assayed to quantify metabolite and protein concentration and for genetic profiling including genotyping[111] and more recently exome[112] and whole genome sequencing[113]. Furthermore, participants' healthcare records are linked so that ongoing medical records can be obtained and included for research purposes[111]. This data resource is available to any researcher that wishes to undertake health-related research to

improve outcomes for the public[111]. All analyses were performed under UK Biobank application number 15825.

3.3.2 Genetic data

Genetic array data were available on n=488,377 consenting participants measured using a combination of UK Biobank Axiom™ array (n=438,398) and UK BiLEVE array (n=49,979). Genotype imputation was performed by UK Biobank using a reference set combined with UK10K haplotypes and HRC reference panels with the IMPUTE2[114] software as described in their companion paper[111].

I removed the following SNPs from analyses leaving a total of n=6,812,700: multi-allelic loci, loci with a minor allele frequency < 5%, Hardy-Weinberg violations ($P < 1 \times 10^{-5}$), genotype missing rate >5%, or a low imputation score (INFO < 0.3), and *HLA* loci (hg19/GRCh37 chr6:23477797-38448354).

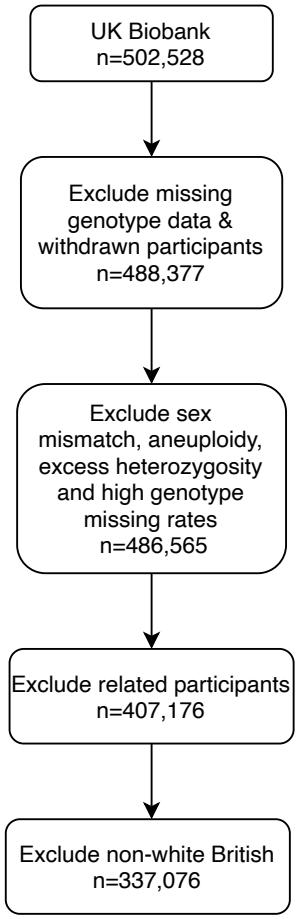
Forty genetic principal components were estimated by UK Biobank using 407,219 unrelated participants and 147,604 independent genotypes[111]. These were prepared using fastPCA[115].

3.3.3 Quality control

I applied standard exclusion criteria (**Figure 3.3.3.1**) using pre-calculated variables created by the MRC Integrative Epidemiology Unit to remove SNP-phenotype sex mismatches, aneuploidies, and outliers for missingness or heterozygosity as described in the published protocol[116] leaving n=486,565 participants. To ensure data independence, I removed closely related subjects using pre-calculated variables by the MRC Integrative Epidemiology Unit as described[116] leaving n=407,176 participants. Finally, I excluded 'non-white British' participants

defined by the MRC Integrative Epidemiology Unit using published methodology[116] to minimise confounding by population stratification providing a final sample size of n=337,076.

Figure 3.3.3.1. UK Biobank participant inclusion criteria



Flowchart of UK Biobank participant quality control procedure applied to all UK Biobank analyses

3.3.4 Serum biomarkers

This thesis applied vQTL methods to 30 serum biomarkers measured in UK Biobank participants using a single serum sample collected at baseline without fasting[117] (**Table 3.3.4.1**; **Figure 3.2.4.1**). These measures were chosen by UK Biobank[117] to include disease risk factors (such as lipids, glycaemic measures and urate), diagnostic measures (such as cystatin C, alanine/aspartate aminotransferase and gamma glutamyltransferase) and measures of phenotypes that are less well assessed using other means (such as rheumatoid factor, oestradiol and testosterone)[117]. After restricting to a white British subset and performing sample quality control (**Chapter 3.3.3**), measures without missing data were available on up to 321,260 participants (see **Table 3.3.4.1** for sample size for each measure). Oestradiol and rheumatoid factor had high levels of missingness due to values reported below the assay limit of detection[118].

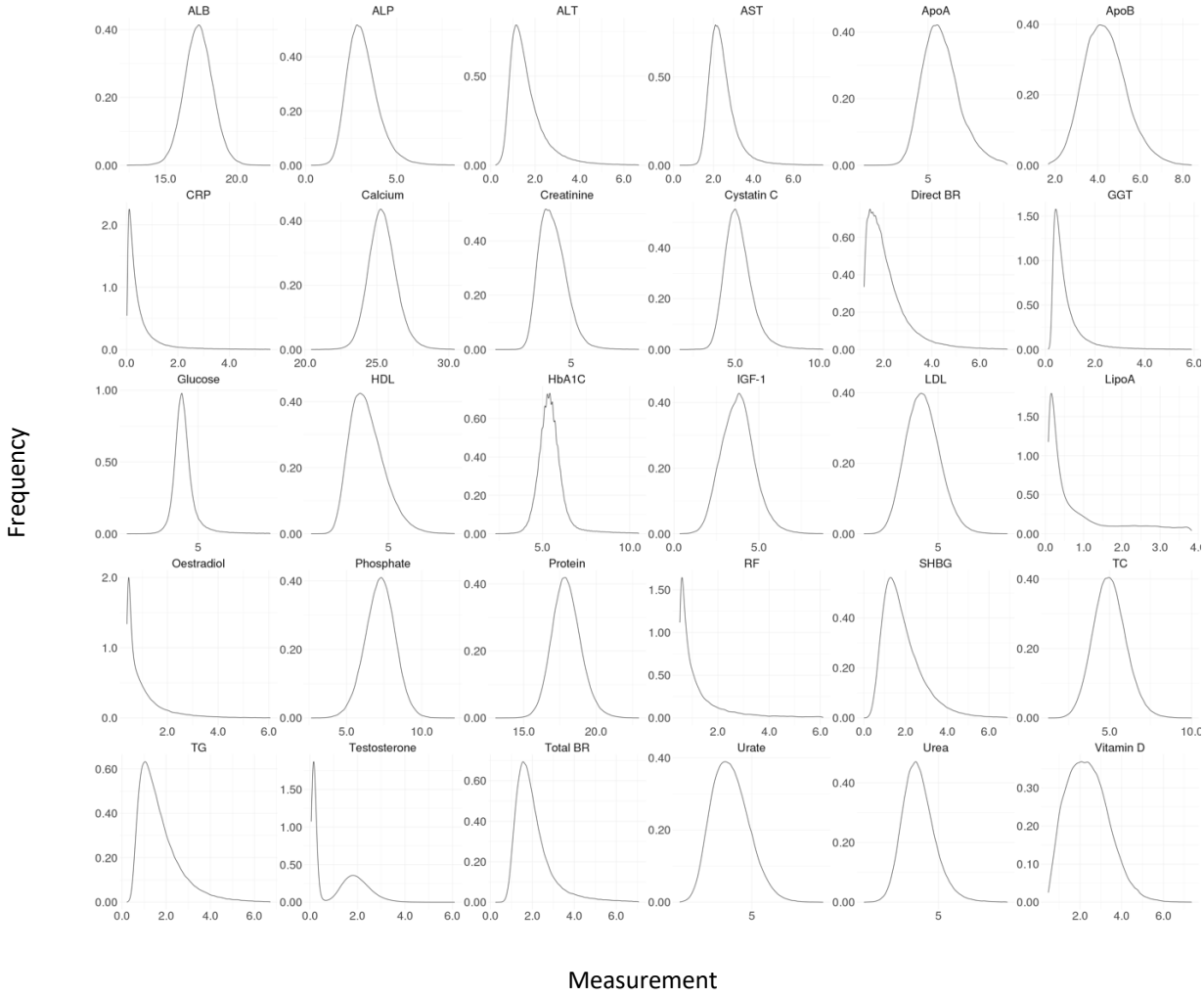
Table 3.3.4.1. UK Biobank serum biochemistry biomarkers and sample size with measures

UK Biobank Field ID	Biomarker	Abbreviation	N	Mean	SD
30620	Alanine aminotransferase	ALT	319,817	23.55	14.18
30600	Albumin	ALB	294,114	45.21	2.63
30610	Alkaline phosphatase	ALP	320,661	83.67	26.46
30630	Apolipoprotein A	ApoA	292,384	1.54	0.27
30640	Apolipoprotein B	ApoB	319,725	1.03	0.24
30650	Aspartate aminotransferase	AST	318,847	26.23	10.66
30710	C-reactive protein	CRP	318,256	2.60	4.36
30680	Calcium	Calcium	293,851	2.38	0.09
30690	Total cholesterol	TC	321,260	5.69	1.14
30700	Creatinine	Creatinine	320,650	72.31	18.55
30720	Cystatin C	Cystatin C	320,423	0.91	0.18
30660	Direct bilirubin	Direct BR	272,719	1.83	0.85
30730	Gamma glutamyltransferase	GGT	319,210	37.39	42.09
30740	Random glucose	Glucose	291,579	5.12	1.24
30750	Glycated haemoglobin	HbA1c	318,931	36.13	6.78
30760	HDL cholesterol	HDL	293,951	1.45	0.38
30770	IGF-1	IGF-1	319,365	21.40	5.70
30780	LDL cholesterol	LDL	320,678	3.56	0.87
30790	Lipoprotein A	LipoA	255,575	44.65	49.21
30800	Oestradiol	Oestradiol	50,380	461.17	431.16
30810	Phosphate	Phosphate	293,580	1.16	0.16
30820	Rheumatoid factor	RF	28,680	24.56	19.86
30830	Sex hormone binding globulin	SHBG	290,600	51.63	27.78
30850	Testosterone	Testosterone	291,163	6.56	6.05
30840	Total bilirubin	Total BR	318,577	9.13	4.43
30860	Total protein	Protein	293,758	72.51	4.12
30870	Triglycerides	TG	320,016	1.75	1.03
30880	Urate	Urate	320,848	309.21	80.43
30670	Urea	Urea	320,479	5.40	1.40
30890	Vitamin D	Vitamin D	307,091	48.61	21.11

Table of 30 serum biomarkers measured in UK Biobank under study in this thesis. ID, UK

Biobank biomarker identifier.

Figure 3.3.4.1. Biomarker distributions



UK Biobank serum biomarker distribution. ALB, albumin. ALP, alkaline phosphatase. ALT, alanine aminotransferase. AST, aspartate aminotransferase. ApoA, Apolipoprotein A. ApoB, apolipoprotein B. CRP, C-reactive protein. Direct BR, direct bilirubin. GGT, Gamma glutamyltransferase. HDL, high-density lipoprotein. HbA1C, glycated haemoglobin. IGF-1, insulin growth factor. LDL, low-density lipoprotein. LipoA, lipoprotein A. RF, rheumatic factor. SHBG, sex-hormone binding globulin. TC, total cholesterol. TG, triglycerides. Total BR, total bilirubin. SD units. Biomarker outliers with Z-score > 5SD from the mean were removed.

3.3.5 Ethical approval and consent

Ethical approval for the UK Biobank study was granted (date 17/06/2011) by the National Research Ethics Service Committee Northwest (ref 11/NW/0382). Informed consent was obtained from all subjects to prior to participation in the UK Biobank study[111]. All analyses were performed under approved UK Biobank project 15825 (dataset ID 33352).

3.3.6 Strengths and limitations

The UK Biobank data resource is large and contains rich and diverse measures of individual genetic and phenotypic characteristics as described above. However, UK Biobank also has some limitations. As with all cohort studies, participants may not be representative of the population from which they are sampled[119].

Nine million individuals were invited to take part in the UK Biobank study but only 500,000 participants were recruited, representing at 5.5% response rate[119]. These individuals differ from the wider UK population in several ways, for example, compared to the national average, smoking status is low, educational attainment is high, and mortality is low[119] which are features of higher socio-economic status[119]. Study selection bias is known to induce biased estimates[119]. Evidence for a selected sample may also be observed in the finding that 30.3% of UK Biobank participants are related to one or more individuals in the study which is double that expected due to chance[111]. Selection bias may also be exacerbated by attrition which is unlikely to be random and may be related to socioeconomic factors[119] resulting in a study population that becomes less representative over time[119] and also by restricting future investigations to participants with complete data[119].

3.4 Secondary data sources

3.4.1 Background

In this thesis I used GWAS summary statistics to perform MR of selected biomarkers on disease outcomes (**Chapter 6**) and for evaluating the performance of a new GWAS summary statistics format (**Chapter 7**).

3.4.2 Measurements

Summary statistics were obtained from large case-control GWAS (**Table 3.4.2.1**) of type 2 diabetes[120], gout[121] and cardiovascular disease[122] using samples not thought to overlap with UK Biobank. I also used mean and variance GWAS summary statistics of LDL[2,123], glucose[2,123] and urate[2,123] estimated in UK Biobank. These data were extracted from the MRC-IEU OpenGWAS platform[18]. I obtained GWAS summary statistics of BMI estimated in UK Biobank from Neale *et al*[123].

3.4.3 Ethical approval and consent

Ethical approval for research involving GWAS summary statistics was not required as the data were publicly available.

3.4.4 Strengths and limitations

Use of GWAS summary statistics circumvents the need to share individual level data which is problematic due to privacy and consent laws[7] (**Chapter 1.2.5**). However, analyses with GWAS summary statistics can be more limited than using individual level data. For example, using summary statistics the researcher has no control over the model used to estimate the SNP-trait association or which covariates were adjusted.

Table 3.4.2.1. Sources and characteristics of GWAS summary statistics

Outcome	Source	Sex	Population	N	Cases	Controls
Type 2 diabetes	DIAGRAMplusMetabochip[120]	Males and females	Mostly European	149,821	34,840	114,981
Coronary heart disease	CARDIoGRAMplusC4D[122]	Males and females	Mostly European	184,305	60,801	123,504
Gout	Global Urate Genetics Consortium[121]	Males and females	European	69,374	2,115	67,259
LDL cholesterol (mean)	Neale *et al (UK Biobank)*[123]	Males and females	White British	343,621	-	-
Random glucose (mean)	Neale *et al (UK Biobank)*[123]	Males and females	White British	314,916	-	-
Urate (mean)	Neale *et al (UK Biobank)*[123]	Males and females	White British	343,836	-	-
LDL cholesterol (variance)	Lyon *et al (UK Biobank)*[2]	Males and females	White British	320,678	-	-
Random glucose (variance)	Lyon *et al (UK Biobank)*[2]	Males and females	White British	291,579	-	-
Urate (variance)	Lyon *et al (UK Biobank)*[2]	Males and females	White British	320,848	-	-
Body mass index (mean)	Neale *et al (UK Biobank)*[123]	Males and females	White British	359,983	-	-

Table of GWAS summary statistics under study in this thesis. MRC-IEU, Medical Research Council Integrative Epidemiology Unit. LDL, low-density lipoprotein cholesterol.

3.5 Data availability

Variance GWAS summary statistics produced in **Chapter 5** are available from the MRC-IEU OpenGWAS platform (https://gwas.mrcieu.ac.uk)[18]. Top vQTLs (**Table 9.2.1**) and GxG/GxE (**Table 9.2.2**) summary statistics produced in **Chapter 5** are available in the appendix (**Chapter 9**). GWAS summary statistics of BMI used in **Chapter 7** are available from: https://broad-ukb-sumstats-us-east-1.s3.amazonaws.com/round2/additive-tsvs/21001_raw.gwas.imputed_v3.both_sexes.tsv.bgz.

UK Biobank data are available from https://www.ukbiobank.ac.uk.

Chapter 4: Evaluation of LAD-BF for estimating SNP effects on trait variance and implementation in variance GWAS software

4.1 Overview

The presence of genetic interaction effects may be detected by testing for association between the locus and trait variance[56] (**Chapter 1.5.3**). Of the multitude of variance tests that have been proposed (**Chapter 1.5.1**), the Brown-Forsythe test provides among the best type I error rate and power[68,70]. However, the Brown-Forsythe test has limitations when applied to GWAS (**Chapter 1.5.7**). Here, I implemented a regression-based Brown-Forsythe test (LAD-BF) proposed by Professor Tilling and evaluated the test through a series of simulation studies and application to positive and negative controls in UK Biobank to detect loci with variance effects. In contrast with existing implementations of the Brown-Forsythe test, LAD-BF provides an unbiased variance effect estimate when the trait is normally distributed. This can be useful for determining if a variance effect is driven by an interaction through adjusting for the interaction and considering attenuation of the variance effect size. Additionally, LAD-BF enables adjustment for genetic confounding (**Chapter 1.3.2**) of the variance model which can be useful to avoid bias by population stratification (**Chapter 1.2.3**). I developed fast open-source software (varGWAS) for scalable genome-wide association analysis of SNP-variance effects (https://github.com/MRCIEU/varGWAS) and an R-package for smaller scale analyses (https://github.com/MRCIEU/varGWASR) to facilitate future research.

4.2 Contribution statement

Work in **Chapter 4** forms part of a manuscript I wrote that was edited by PhD supervisors available as a preprint on MedRxiv (Lyon *et al*, 2022)[2]. I also contributed type I error

simulations of the Brown-Forsythe test presented here to a journal article published in European Journal of Epidemiology (Staley *et al*, 2021)[80].

Professor Tilling proposed the LAD-BF test and derived an expression for the relationship between exposure and outcome variance under interaction effect and the formula for calculating variance from mean-absolute deviation in this context (**Chapter 9.1**). Professor Davey Smith proposed the positive and negative controls. I performed the simulation studies, implemented the model in C++ and R and tested the approach using positive and negative controls (**Chapter 1.3.7**) with data from UK Biobank.

4.3 Introduction

Genetic interaction effects can provide valuable information on disease mechanisms[4,48], improve prediction of disease outcomes[4,48], and identify drug targets for precision medicines[48] (**Chapter 1.4.1**). However, detection of these effects may incur large multiple testing burden[4]. One approach to reduce multiple testing is through prioritisation of loci with effects on trait variance[56] which would be anticipated under an interaction effect (**Chapter 1.4.6**). Previous studies[68,70] have evaluated a number of statistical tests for detecting variance effects and have suggested the Brown-Forsythe test[71] due to low type I error rate and comparable power with other methods (**Chapter 1.5.2**).

4.3.1 Limitations of the Brown-Forsythe test

The original implementation of the Brown-Forsythe test[71] does not estimate the size of the variance association (**Chapter 1.5.1**). This feature could be useful for enabling comparisons of association magnitude with and without adjustment for a candidate interaction to determine the extent to which the included interaction drives the observed effect on outcome variance.

The Brown-Forsythe test also cannot adjust for covariates, and, while it is possible to pre-adjust outcomes[68], it is unclear if this strategy can account for confounding of the variance association or just the mean association.

4.3.2 Regression-based implementation of the Brown-Forsythe test

Levene's test has been reformulated using the regression framework[83] (**Chapter 1.5.1**) which provides greater flexibility to overcome the aforementioned limitations. This approach uses the same structure as the Breusch-Pagan test[78,83] (**Chapter 1.5.1**) which estimates the exposure-outcome variance effect through two independent regression models. The first model regresses the outcome on exposure. The second regresses the squared residuals of the first-stage model on the exposure providing an estimate of the exposure on outcome variance. However, as Levene's test detects variability from the mean it is susceptible to elevated type I error rate in the presence of non-normality.

The deviation regression model (DRM)[70] was proposed as a regression-based implementation of the Brown-Forsythe test. The DRM pre-adjusts an outcome using OLS and then estimates the absolute deviation from the outcome median within each SNP group. This deviation is then regressed on the SNP in a second OLS model and has near identical performance to the Brown-Forsythe test in terms of power and type I error rate. However, the DRM implies linearity of the SNP effect on outcome variability which may not hold[56] (**Chapter 9.1**).

4.3.3 Aims

In this chapter I aim to implement and evaluate the least-absolute deviation (LAD) regression Brown-Forsythe test (LAD-BF) proposed by Professor Tilling. This approach uses the

same structure as the Breusch-Pagan test. However, the first-stage regression model uses least-absolute deviation (LAD) regression[105] in instead of OLS and takes the absolute rather than squared residuals as in the case of Breusch and Pagan. LAD regression estimates the mean exposure effect on outcome median (rather than mean as with OLS) providing robustness to trait non-normality. The LAD-BF test can adjust for covariates and provides a variance effect estimate that does not make linearity assumptions.

I aim to compare LAD-BF with Brown-Forsythe and Breusch-Pagan tests through simulation to detect SNP-interaction effects. I also aim to develop scalable open-source software for performing variance GWAS using LAD-BF and an R-package for evaluation and smaller scale analyses.

4.4 Materials and methods

4.4.1 Software implementation

The LAD-BF test was implemented in varGWAS available in C++ v1.2.3 (https://github.com/MRCIEU/varGWAS) and R v1.0.0 (https://github.com/MRCIEU/varGWASR). As GWAS studies are highly computationally intensive, I decided to use C++ which is a performant language designed for efficiency[124]. Many existing GWAS tools are also developed in C++ including BOLT-LMM[125], PLINK[126] and SAIGE[127]. To enable parallel processing of genetic loci I used the OpenMP multithreading library[128].

Both implementations used LAD regression model from the cqrReg R-package[105] (https://cran.r-project.org/web/packages/cqrReg/index.html). The C++ implementation also used Eigen v3.4.0[129] and BGEN library[130] v1.1.6 for general matrix functionality and BGEN file processing, respectively.

4.4.2 Simulation study overview

The following section describes a comprehensive range of simulation studies (**Chapter 2.2.1**) undertaken throughout **Chapter 4** which aim to evaluate the LAD-BF test (**Table 4.4.2.1**). First, I verified the relationship between exposure and outcome variance under interaction effect (**Simulation 4.4.3**). Professor Tilling derived an algebraic expression for this relationship (**Chapter 9.1**) indicating outcome variance was conditional on the exposure and the square of the exposure. Second, I evaluated LAD-BF, Brown-Forsythe and Breusch-Pagan tests based on this expression for type I error rate under a range of outcome distributions (**Simulation 4.4.4**). Third, I performed simulations for LAD-BF and Brown-Forsythe tests for power (**Simulation 4.4.5**) and bias and coverage (**Simulation 4.4.6**). Fourth, I explored the consequences of confounding by population stratification on the variance effect with various covariate adjustment strategies (**Simulation 4.4.7**). Fifth, I compared variance test P-value distributions under interaction effect with/without adjustment for the interaction (**Simulation 4.4.8**). Sixth, I compared the power of detecting an interaction effect using linear regression and a series of candidate modifiers against testing for an effect on outcome variance (**Simulation 4.4.9**). Finally, I compared the runtime performance of LAD-BF and the Brown-Forsythe test implemented in the OSCA[96] software package with increasing CPU thread count (**Simulation 4.4.10**).

Table 4.4.2.1. Simulation study summary

Simulation	Description
4.4.3	Verify the relationship between SNP and outcome variance when the SNP has an interaction effect on the outcome
4.4.4	Measure the LAD-BF, Brown-Forsythe, and Breusch-Pagan test type I error rate under a range of outcome distributions
4.4.5	Estimate the power of LAD-BF and Brown-Forsythe tests under interaction effect of SNP on outcome
4.4.6	Estimate bias and coverage of variance effect estimate and confidence interval for LAD-BF and Brown-Forsythe test
4.4.7	Estimate the null hypothesis rejection rate for LAD-BF and Brown-Forsythe tests under variance confounding with and without adjustment
4.4.8	Compare LAD-BF and Brown-Forsythe test P-value distributions with and without adjustment for a simulated interaction effect
4.4.9	Compare the power of exhaustive testing using a set of simulated modifiers with LAD-BF to detect the presence of an interaction effect
4.4.10	Comparison of runtime performance for LAD-BF and Brown-Forsythe tests with increasing CPU threads

SNP, single nucleotide polymorphism. LAD-BF, least-absolute deviation regression Brown-

Forsythe test. CPU, central processing unit.

4.4.3 Simulation to verify the relationship between SNP and outcome variance under SNP interaction effect

Aim: To verify algebraic expression derived by Professor Tilling for the relationship between SNP and outcome variance when the SNP has an interaction effect on the outcome (**Chapter 9.1**).

Data-generating mechanisms: Data were simulated for N=1000 independent observations within each simulated dataset. For the ith observation, I simulated a SNP G_i in HWE with a MAF of 0.4 and standard Normal modifier U_i. I used the approach of Brookes *et al*[57] to set the effect sizes, defined as follows: the outcome Y_i was simulated to have main effects of G_i and modifier U_i detectable with 80% power. The interaction effect GU_i of G_i and U_i was set $\epsilon\{0, 0.5, 1, 1.5..6\}$ times the size of the main effect of G_i.

$$Y_i = \beta_1 G_i + \beta_2 U_i + \beta_3 G U_i + E_i$$

Where E_i is the residual variance of Y_i drawn from the standard Normal distribution.

Estimand: The outcome variance conditional on SNP.

Methods: The outcome variance was estimated within each SNP group.

Performance measures: Bias of estimated variance compared with calculated variance with N=1000 replications. This value was chosen to ensure confidence intervals derived from Monte Carlo standard errors were sufficiently precise while optimising computing resources.

Open-source code: https://github.com/MRCIEU/varGWAS/blob/master/sim/sim19.R

4.4.4 Simulation to estimate type I error rate of variance tests under the null hypothesis

Aim: To compare type I error rate of LAD-BF, Brown-Forsythe, and Breusch-Pagan tests under the null of no SNP effect on continuous outcomes. This was performed using a range of

outcome distributions to determine how deviation from Normality affect the test type I error rate.

Data-generating mechanisms: Data were simulated for N=100,000 independent observations within each simulated dataset. For the ith observation, I simulated a SNP G_i in HWE with MAF of 0.05. This sample size was selected to simulate the real-world application to biobank data. This large sample and low MAF were chosen, as previous research indicated a high false-positive rate at lower MAF with non-normal outcomes[68] and I aim for the findings to inform applied analyses. The outcome Y_i was randomly generated from either standard Normal, t (df=4), log standard Normal or mixed Normal $0.9\ N(0,1), 0.1\ N(5,1)$ distributions. These distributions were chosen to evaluate type I error rate for distributions with skew (log Normal) and kurtosis (t and mixed Normal).

$$Y_i = \beta_1 G_i + E_i$$

Where the effect β_1 of G_i on Y_i was set to null and E_i was the residual variance drawn from a range of distributions.

Estimand: The test statistic for the null hypothesis of variance homogeneity.

Methods: The effect of the SNP on outcome variance was tested using the Brown-Forsythe test, LAD-BF (**Chapter 2.2.2**; **Chapter 2.2.4**) and Breusch-Pagan test (**Chapter 2.2.3)**.

Performance measures: Type I error rate compared with the expected null with N=1000 replications.

Open-source code: https://github.com/MRCIEU/varGWAS/blob/master/sim/sim2b.R

4.4.5 Simulation to estimate the power of Brown-Forsythe tests to detect interaction effects

Aim: To estimate and compare statistical power of LAD-BF and Brown-Forsythe tests to detect variance effects produced by a simulated SNP interaction effect on continuous outcome.

Data-generating mechanisms: Data were simulated for N=200, N=2000, N=200,000 and N=2,000,000 independent observations within each simulated dataset. I chose these large sample sizes to ensure high power across all methods and parameters. For the ith observation, I simulated a SNP G_i in HWE with a MAF of 0.4, a standard Normal modifier U_i and outcome Y_i with residual drawn from either standard Normal, t (df=4) or log standard Normal distributions. I used the approach of Brookes *et al*[57] to set the effect sizes. The main effects β_1 and β_2 of G_i and U_i, respectively were set to have 80% power when then sample size was N=200 (assuming normally distributed residuals). The interaction effect β_3 of G_i and U_i denoted by GU_i was varied $\epsilon\{0, 0.5, 1, 1.5..6\}$ times the size of the main effect of G_i.

$$Y_i = \beta_1 G_i + \beta_2 U_i + \beta_3 GU_i + E_i$$

Estimand: Test statistic for the null hypothesis of variance homogeneity.

Methods: The effect of G_i on the variance of Y_i was tested using LAD-BF and Brown-Forsythe tests (**Chapter 2.2.2**; **Chapter 2.2.4**).

Performance measures: Power was defined as the percentage of tests with P < 0.05. Each configuration of parameters was evaluated using N=200 replications. This value was chosen to ensure confidence intervals derived from Monte Carlo standard errors were sufficiently precise while optimising computing resources.

Open-source code: https://github.com/MRCIEU/varGWAS/blob/master/sim/sim1.R

4.4.6 Simulation to estimate bias of the variance effect and confidence interval coverage

Aim: To evaluate the bias of variance effect estimates for LAD-BF and Brown-Forsythe tests (using effect estimate derived by Zhu *et al*[26]). This was performed using a linear association of exposure with outcome variance or interaction effect.

Data-generating mechanisms: Data were simulated for N=10,000 independent observations within each simulated dataset. For the ith observation, I simulated a SNP G_i in HWE with a MAF of 0.4 and standard Normal modifier U_i. G_i was set to either have linear effect β_4 on the variance of outcome Y_i or an interaction effect β_3 on Y_i. I adapted the approach of Brookes *et al*[57] to set the effect sizes. The main effects β_1 and β_2 of G_i and U_i on Y_i were fixed across all simulations and set to have 95% power. Where the variance was generated by the interaction effect, the magnitude of the interaction effect size β_3 was varied and set relative to the main effect β_1 of X_i ranging from $\epsilon\{0,1..12\}$ while $\beta_4 = 0$. Where the variance effect of G_i was linear $\beta_3 = 0$ and instead β_4 was varied and set relative to the main effect β_1 of X_i ranging from $\epsilon\{0,1..12\}$. Finally, Y_i was scaled to have zero mean and unit variance so that variance effect estimates were on the same scale.

$$Y_i = \beta_1 G_i + \beta_2 U_i + \beta_3 G U_i + E_i$$

$$E_i \sim N(0, 1 + \beta_4 G_i)$$

Where E_i is the residual variane of Y_i drawn from the Normal distribution. I chose a large sample size to avoid asymptotic bias[104].

Estimand: Variance effect.

Methods: The difference in variance between SNP dosage zero and one and one and two was estimated using LAD-BF and Brown-Forsythe tests. For Brown-Forsythe the additive

variance effect size was estimated using Zhu *et al*[26] and multiplied to obtain the estimate for one or two allele increases in variance. Bias was estimated using the difference between the expected and observed variance effect estimate. Coverage was estimated using the proportion of replicates where the estimate 95% confidence included the true difference in outcome variance.

Performance measures: Bias of variance effect size estimates. Coverage of variance estimate 95% confidence intervals. Each configuration of parameters was evaluated using N=1000 replications. This value was chosen to ensure confidence intervals derived from Monte Carlo standard errors were sufficiently precise while optimising computing resources.

Open-source code: https://github.com/MRCIEU/varGWAS/blob/master/sim/sim12.R

4.4.7 Simulation to compare the rejection rate of the null hypothesis under ancestry variance confounding and strategies for adjustment of population stratification

Aim: To explore the null hypothesis rejection rate of LAD-BF and Brown-Forsythe tests in the presence of variance confounding by population stratification (i.e., a main effect of ancestry on genotype frequencies and interaction of ancestry on outcome; **Chapter 1.2.3**). Additionally, I explored different adjustment strategies to control confounding by population stratification.

Data-generating mechanisms: Data were simulated for N=1000 independent observations within each simulated dataset. I assigned the ith observation to a simulated ancestral group A_i with $\epsilon\{1,2,3,4,5\}$ levels, within each A_i I simulated a SNP G_i in HWE with varying minor allele frequency (which were selected randomly from the uniform distribution) for each A_i group with values of $\epsilon\{0.14, 0.39, 0.20, 0.44, 0.47\}$. I also simulated two modifiers U_{1i} and U_{2i} drawn from the standard Normal distribution which interacted with A_i and G_i,

respectively denoted by AU_{1i} and GU_{2i}. The outcome Y_i was simulated to have main effects β_1 β_2 β_3 β_4 for A_i U_{1i} G_i and U_{2i} and interaction effects β_5 and β_6 for AU_{1i} and GU_{2i}. The residual variance of Y_i was drawn from the standard Normal distribution.

$$Y_i = \beta_1 A + \beta_2 U_{1i} + \beta_3 G_i + \beta_4 U_{2i} + \beta_5 AU_{1i} + \beta_6 GU_{2i} + E_i$$

The main effects of A_i and U_{1i} were set to explain 25% and 10% of the variance of Y_i, respectively. The main effects of G_i and U_{2i} were set to explain 5% and 10% of the variance of Y_i, respectively. The interaction effects β_5 and β_6 were varied to explain 0-20% and 0-2% of the variance of Y_i respectively. Under this simulation, both A_i and G_i are expected to associate with the variance of Y_i.

Estimand: The test statistic for the null hypothesis of variance homogeneity.

Methods: The effect of G_i on variance of Y_i was tested under varying conditions:

- Brown-Forsythe test applied to unadjusted outcome (BF_1)
- Brown-Forsythe test applied to pre-adjusted outcome for A_i using OLS (BF_2)
- LAD-BF unadjusted (LAD-BF_1)
- LAD-BF adjusted for A_i in the first-stage regression model (LAD-BF_2)
- LAD-BF adjusted for A_i in the first and second-stage regression models (LAD-BF_3)
- LAD-BF adjusted for A_i in the first-stage model and A_i^2 in the second-stage model (LAD-BF_4)
- LAD-BF adjusted for A_i in the first-stage regression model and $A_i + A_i^2$ in the second-stage model (LAD-BF_5)

Performance measures: Null hypothesis rejection rate $\alpha = 0.05$ with N=1000 replications.

Open-source code: https://github.com/MRCIEU/varGWAS/blob/master/sim/sim16.R

4.4.8 Simulation to compare the rejection rate of the null hypothesis under interaction effect with/without adjustment of the interaction effect

Aim: To compare the variance test null hypothesis rejection rate under simulated SNP with interaction effect on outcome with/without adjustment for interaction effect.

Data-generating mechanisms: Data were simulated for N=1000 independent observations within each simulated dataset. For the ith observation, I simulated a SNP G_i in HWE with MAF of 0.4 and standard Normal modifier U_i. G_i was simulated to have a main β_1 and interaction β_3 effect of G_i and U_i explaining 6.5% and 20% of the variance of the outcome, respectively. The main effect β_2 of U_i was set to null. The outcome Y_i residual variance E_i was drawn from standard Normal distribution. I chose these large effect sizes so that the simulation would clearly show that adjusting for the interaction led to strong change in LAD-BF test p-value distribution.

$$Y_i = \beta_1 G_i + \beta_2 U_i + \beta_3 G U_i + E_i$$

Estimand: The test statistic for the null hypothesis of variance homogeneity.

Methods: LAD-BF with/without adjustment for U_i and GU_i in the first-stage regression model.

Performance measures: Null hypothesis rejection rate with N=1000 replications.

Open-source code: https://github.com/MRCIEU/varGWAS/blob/master/sim/sim13.R

4.4.9 Comparison of statistical power to detect the presence of an interaction with LAD-BF and linear regression applied to multiple hypothesised modifiers

Aim: To compare the power of exhaustive interaction effect testing using linear regression in comparison with variance testing to detect the presence of effect modification.

Data-generating mechanisms: Data were simulated for N=1000, N=5000, N=25,000 independent observations selected empirically to achieve full power using each testing approach. For the ith observation I simulated a SNP G_i in HWE with a MAF of 0.4 and five modifiers $U_{1:5i}$ drawn from the standard Normal distribution. The outcome Y_i was simulated with main effects β_1 and β_2 of X_i and U_{1i} and interaction effect β_3 of XU_{1i} each detectable with 80% power using linear regression when the sample size was set to N=1000. Modifiers $U_{2:5i}$ had no effect on Y_i. The residual variance E_i of Y_i was drawn from the standard Normal distribution.

$$Y_i = \beta_1 G_i + \beta_2 U_{1i} + \beta_3 G U_{1i} + E_i$$

Estimand: The test statistic for the null hypothesis of variance homogeneity or for the null hypothesis of linear regression interaction effect.

Methods: The effect of G_i on the variance of Y_i was tested using LAD-BF test (without requiring any modifier). Between one and five interaction effects $GU_{1:5i}$ of G_i and $U_{1:5i}$ were tested using linear regression within each simulation (only one of which was non-null).

Performance measures: Power estimates of LAD-BF and up to five linear regression interaction term(s) with N=200 replications.

Open-source code: https://github.com/MRCIEU/varGWAS/blob/master/sim/sim20.R

4.4.10 Simulation to estimate runtime performance of variance tests with increasing CPU threads

Aim: To estimate and compare runtime performance of LAD-BF and Brown-Forsythe tests across increasing numbers of CPU threads.

Data-generating mechanisms: Data were simulated for N=100,000 independent observations within each simulated dataset. I chose this large sample size to simulate analysis from large biobank datasets. For the ith observation, I simulated 1000 independent SNPs in HWE with a MAF of 0.4 and random outcome Y_i drawn from standard Normal distribution.

Estimand: Test runtime.

Methods: The effect of the SNP on outcome variance was tested using LAD-BF and Brown-Forsythe tests with increasing CPU threads $\epsilon\{1,2,4,8\}$ using an Intel Xeon CPU E5-2680 v4 @ 2.40GHz.

Performance measures: Difference in runtime between methods with N=200 replications.

Open-source code: https://github.com/MRCIEU/varGWAS/blob/master/sim/sim4.R

4.4.11 Positive and negative control in UK Biobank

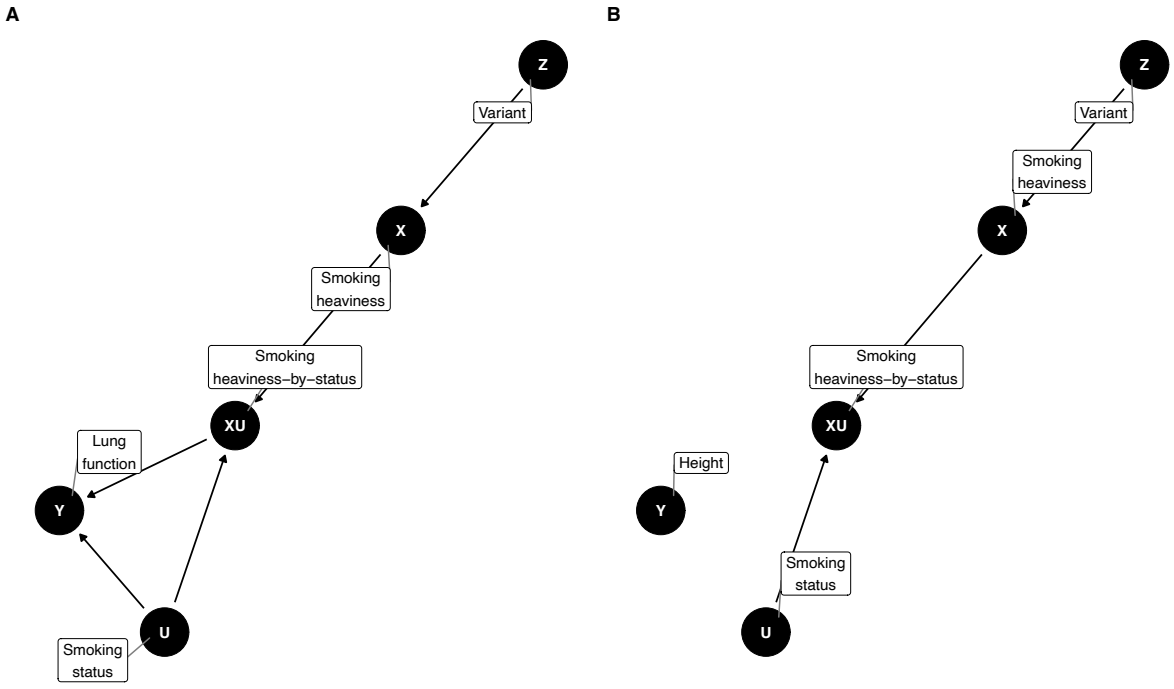
Following comprehensive characterisation of the LAD-BF test through simulation, I aimed to apply LAD-BF to positive and negative controls using real data from UK Biobank. These controls were proposed by Professor Davey Smith.

SNP *CHRNA3* rs1051730-A has a strong positive effect on smoking heaviness but weak effect on smoking initiation[88,131]. It follows that rs1051730-A will strongly affect lung function only among current/former smokers implying a rs1051730-by-smoking status interaction

(**Figure 4.5.8.1**). This interaction effect would be expected to increase the variance of lung function measures. SNP rs1051730 would not be expected to influence adult height (**Figure 4.5.8.1**) and therefore have no variance effect.

I tested for a variance effect of rs1051730 using LAD-BF (**Chapter 2.2.4**) on adult standing height (negative control), forced expiratory volume in 1-second (FEV1; positive control) and forced vital capacity (FVC; positive control) in 337k unrelated white British UK Biobank participants (**Chapter 3.3**). These models were adjusted for age at recruitment, sex, and top ten genetic principal components. The imputed variant dosage was rounded to the nearest whole number so that it could be included as a dummy variable. The variance effect of the variant on these outcomes was reported as an average difference in variance standardised to SD units for an increase of one or two alleles compared with no alleles.

Figure 4.4.11.1. Causal diagram of interaction positive and negative controls



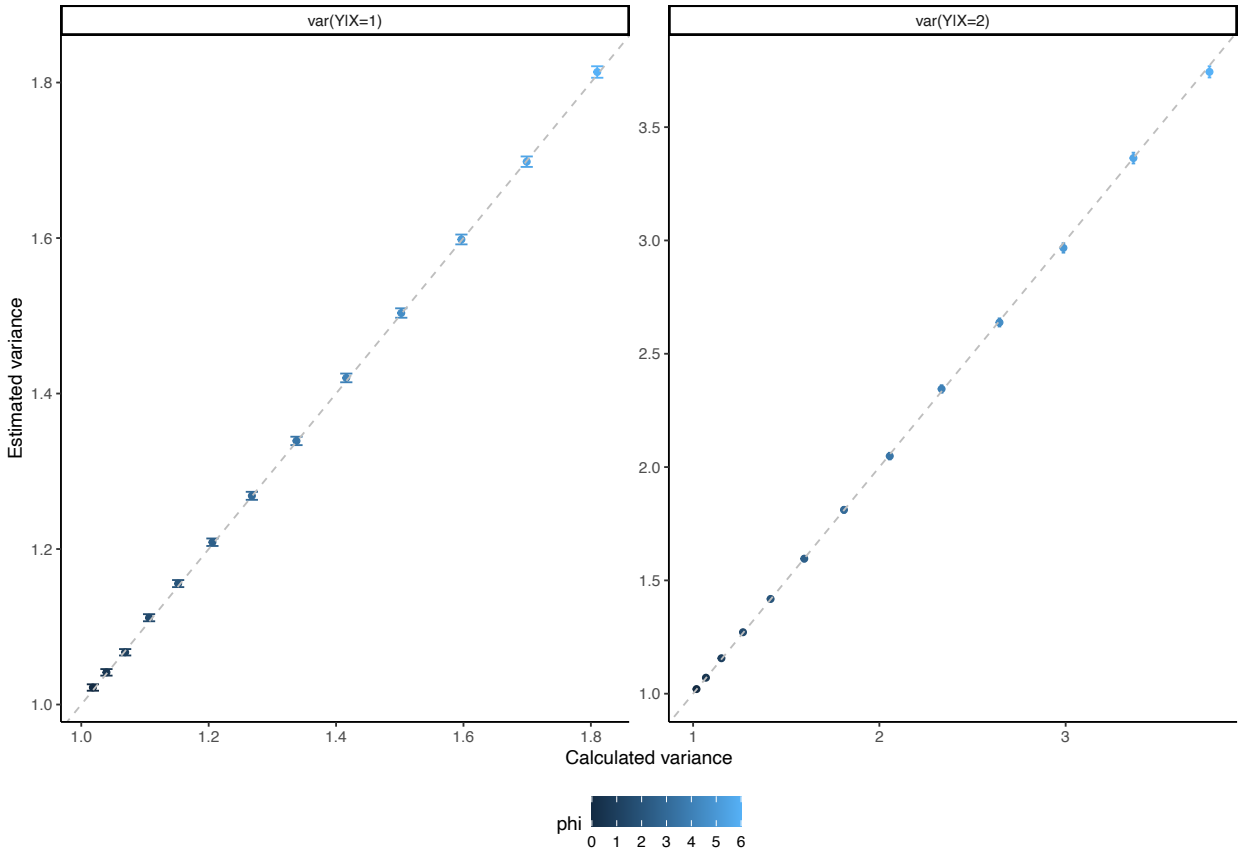
Causal diagram of UK Biobank positive and negative controls. Z, genetic variant. X, exposure. XU, interaction effect of exposure on outcome. U, modifier. Y, outcome. A, the effect of smoking heaviness on lung function which only acts in those who smoke. B, the effect of smoking heaviness on adult standing height which is anticipated to have no effect in either smokers or non-smokers.

4.5 Results

4.5.1 Verifying the relationship between exposure and outcome variance under an interaction effect (Simulation 4.4.3)

Under an interaction effect of SNP on outcome, outcome variance was proportional to the SNP and its square (**Figure 4.5.1.1**). This finding confirms the formula derived by Professor Tilling (**Chapter 9.1**).

Figure 4.5.1.1. Variance of outcome across levels of SNP, under an interaction effect



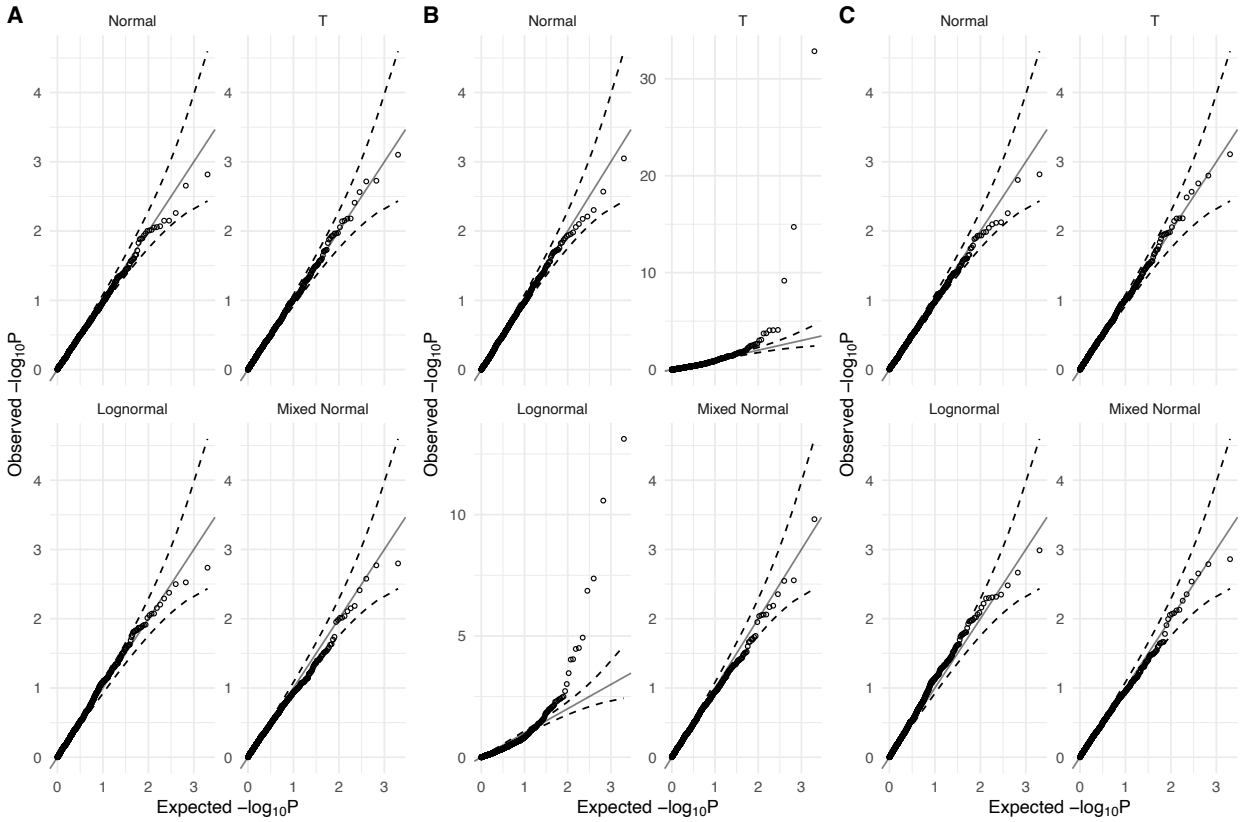
Outcome variance among genotype groups for a SNP with an interaction effect against calculated variance using formula derived by Professor Tilling (**Chapter 9.1**). Confidence intervals were produced from 1000 simulation replications using the Monte Carlo standard errors. The SNP and continuous modifier were set to have 80% power. The SNP-by-modifier interaction effect was set 0-6x the size of the SNP main effect. CI, confidence interval.

4.5.2 Simulated type I error rate and power to detect interaction effects by difference in trait variance under a range of outcome distributions (Simulation 4.4.4 and Simulation 4.4.5)

The LAD-BF and Brown-Forsythe tests were equally well controlled for type I error rate in comparison with the Breusch-Pagan test which was adversely affected by log-normal and t-distributions (**Figure 4.5.2.1**) and not considered further.

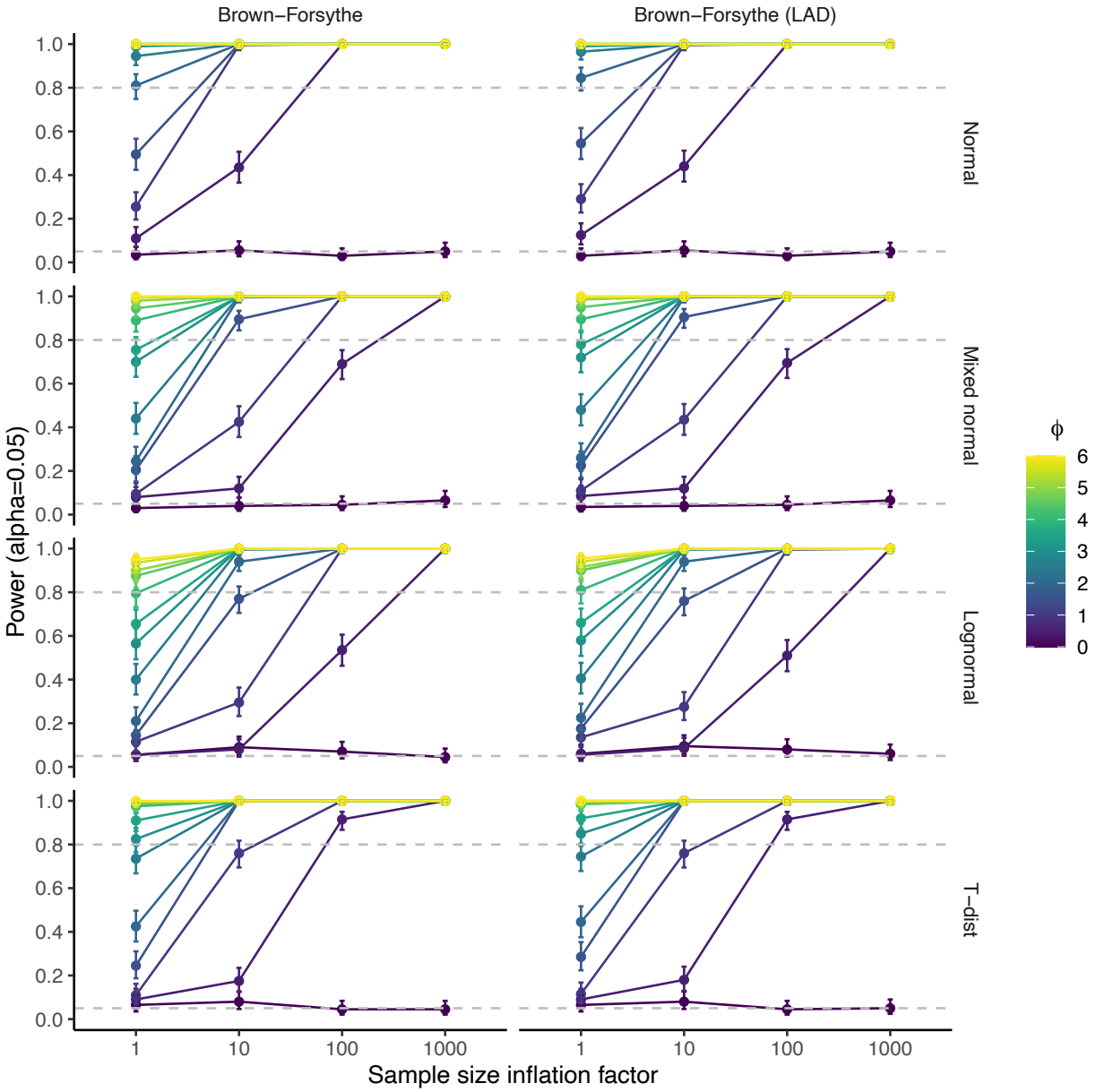
The power to detect a difference in trait variance due to an interaction effect was low and equal for LAD-BF and Brown-Forsythe tests (**Figure 4.5.2.2**). However, the approach showed utility to detect larger effects when applied to biobank scale sample sizes such as UK Biobank. For example, 78% power (95% CI 0.71, 0.83) was obtained for a normally distributed outcome when the SNP main and interaction effects explained 5% variance of the outcome, and the sample size was n=500,000. However, a SNP explaining 5% trait variance is unlikely for complex traits but could be applicable for molecular phenotypes such as protein expression where *cis*-SNPs explain on average 5.8% trait variance[132]. Compared with normally distributed outcomes, non-normal outcomes with positive skew (mixed-normal and log-normal distributions) and kurtosis (t-distribution) had lower power (**Figure 4.5.2.2**).

Figure 4.5.2.1. Type I error rate of tests for effect on outcome variance, across simulation repetitions



Variance test P-value distributions under the null hypothesis of no effect on outcomes simulated from the: Normal, standard Normal distribution. Lognormal, standard log Normal distribution. T-dist, distribution with 4 degrees of freedom. Mixed normal, distribution produced with $0.9\ N(0,1), 0.1\ N(5,1)$. A, Brown-Forsythe test. B, Breusch-Pagan test. C, LAD-BF test. Simulations were produced with 1000 repetitions and 100,000 observations.

Figure 4.5.2.2. Power to detect SNP-interaction effects using variance testing under simulation



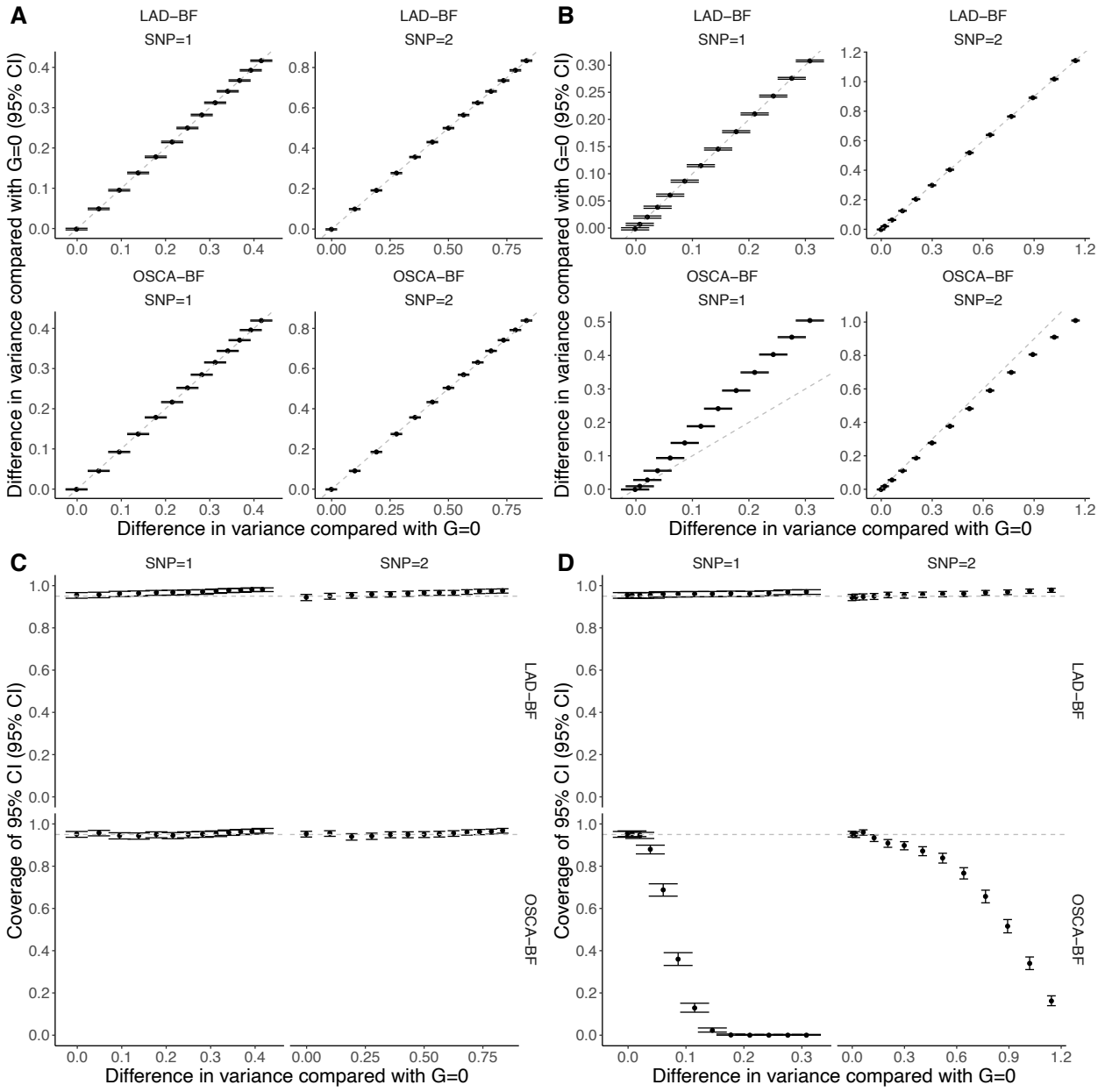
Power to detect interaction effects using SNP association with trait variance using LAD-BF and Brown-Forsythe tests and a range of outcome distributions. Phi, interaction effect size relative to main effect. Inflation factor, sample size relative to the size required to detect the main effect with 80% power. Normal, distribution with mean of 0 and variance of 1. Mixed normal,

distribution with 90% Normal with mean of 0 and variance of 1 and 10% Normal with mean of 5 and variance of 1. Lognormal, distribution with mean of 0 and variance of 1. T-dist, distribution with 4 degrees of freedom. SNP, single-nucleotide polymorphism simulated with minor allele frequency of 0.4 in Hardy-Weinberg equilibrium. All simulations had a fixed main effect detectable with 80% power when the sample size inflation factor was equal to 1. Simulation was performed with 200 repetitions. Sample size inflation factor of 1 was set to 200 observations. Error bars represent the 95% confidence interval. LAD, least absolute deviation.

4.5.3 Bias and confidence interval coverage of simulated variance effect estimate (Simulation 4.4.6)

Under a simulated linear effect of SNP dosage on outcome variance LAD-BF and Brown-Forsythe tests gave the correct effect estimate and 95% confidence interval coverage (**Figure 4.5.3.1**). However, when the difference in variance was a consequence of an interaction effect, the relationship between the SNP and outcome variance was non-linear, dependent on the modifier. Under these conditions, the variance effect estimate produced by Brown-Forsythe from test P-value[26,96] gave the incorrect effect size while LAD-BF produced the correct estimate.

Figure 4.5.3.1. Variance effect estimate accuracy and confidence interval coverage



Comparison of variance effect estimate accuracy and confidence interval coverage for LAD-BF and OSCA tests with one (SNP=1) or two (SNP=2) copies of the minor allele when the SNP is simulated to have a linear effect on trait variance or variance effect produced through interaction effect on outcome. In each plot the dotted line shows expected value. Variance effect estimate accuracy (A, B) and 95% confidence interval coverage (C, D) of simulated SNPs

(G) with proportional (linear) effect on outcome variance (A, C) or interaction effect (B, D). LAD-BF, least-absolute deviation regression Brown-Forsythe. OSCA-BF, Brown-Forsythe test implemented in OSCA[96] including effect estimate derived from the test P-value[26]. CI, confidence interval.

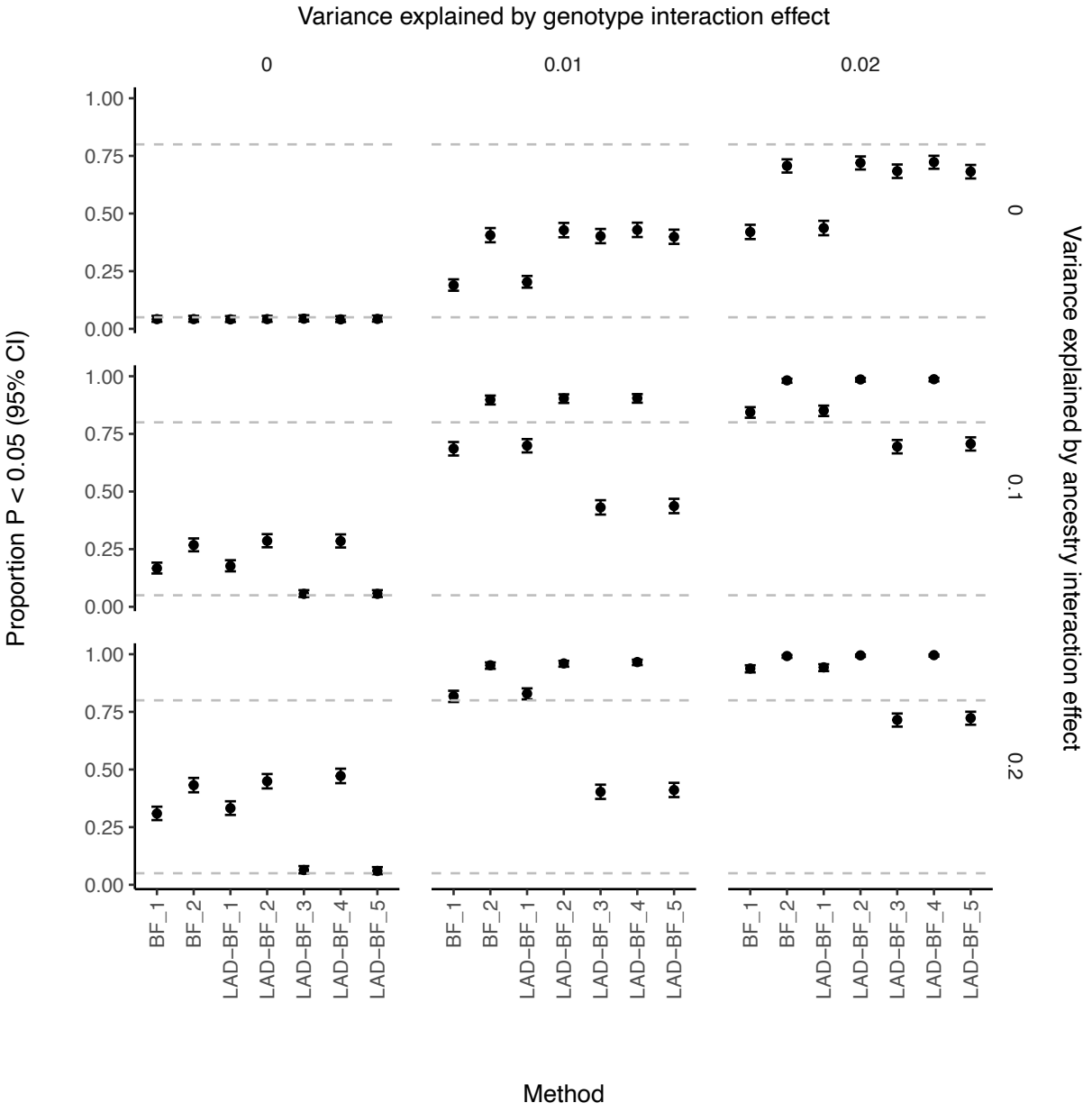
4.5.4 Simulating the effect of variance confounding by population stratification and adjustment of Brown-Forsythe and LAD-BF tests (Simulation 4.4.7)

Ancestry may influence both genotype frequencies and outcome in which case it is a confounder (**Chapter 1.3.2**), and association studies are susceptible to bias by population stratification (**Chapter 1.2.3**). Suppose ancestry also interacts with another variable to influence an outcome. If it does, the variance effect may be susceptible to population stratification as well as the mean association. Here, I explored this situation using simulation to consider the scenarios where adjustment for ancestry correctly controlled type I error rate from population stratification. I also explored the possibility of adjustment for variance confounding leading to gains in power to detect gene-interaction effects not biased by population stratification (**Figure 4.5.4.1**).

Figure 4.5.4.1 shows the null hypothesis rejection rate for variance homogeneity using LAD-BF and Brown-Forsythe tests. This was performed across a range of testing scenarios with increasing variance explained by the SNP-interaction effect (top) and ancestry-interaction effect (right). When the ancestry-interaction effect variance explained was zero, I found no differences in null hypothesis rejection rate between formal modelling of covariates in the first-stage LAD-BF model (LAD-BF_2) compared with applying the Brown-Forsythe test to pre-adjusted outcomes (BF_2). Meanwhile, under ancestry confounding, LAD-BF adjusted for ancestry in both regression models (LAD-BF_3) produced correctly controlled type I error rate with slight reduction in power due to reduced degrees of freedom. In contrast, the Brown-Forsythe test (BF_2) applied to pre-adjusted outcomes using OLS regression could not control

for this type of confounding and had elevated type I error rate despite the null SNP interaction effect.

Figure 4.5.4.1. Effect of adjustment for variance confounding by ancestry on test type I error rate



Type I error rate (first column) and statistical power (second and third columns) of original and LAD-regression based Brown-Forsythe tests under genetic confounding by ancestry and range of confounding adjustment approaches. BF 1, Brown-Forsythe test without adjustment. BF_2, Brown-Forsythe test on preadjusted outcome for main effect of ancestry. LAD-BF_1, LAD-BF

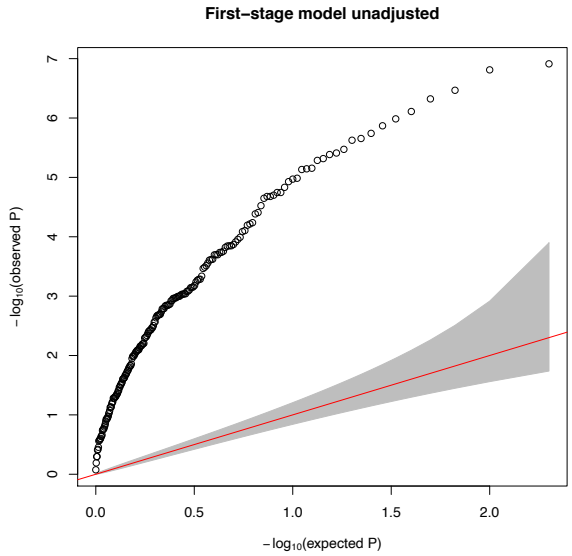
model unadjusted. LAD-BF_2, LAD-BF model adjusted for ancestry in the first-stage model. LAD-BF_3, LAD-BF model adjusted for ancestry in both models. LAD-BF_4, LAD-BF model adjusted for ancestry in the first-stage model and squared ancestry in the second-stage model. LAD-BF_5, LAD-BF adjusted for ancestry in both models and squared ancestry in the second-stage model. CI, confidence interval.

4.5.5 LAD-BF null hypothesis rejection-rate under interaction effect when adjusting for an interaction effect through simulation (Simulation 4.4.8)

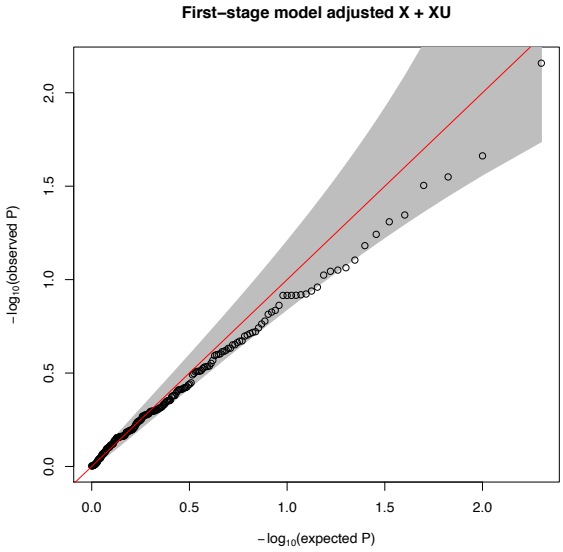
I simulated an interaction effect and compared the LAD-BF test P-value distributions with and without adjusting for the simulated interaction (**Figure 4.5.5.1**). Including the main effect of the modifier and SNP-by-modifier interaction term in the first-stage regression model completely attenuated the SNP effect on outcome variance test statistic.

Figure 4.5.5.1. Effect of adjustment for the interaction effect on variance test P-value distribution

A



B



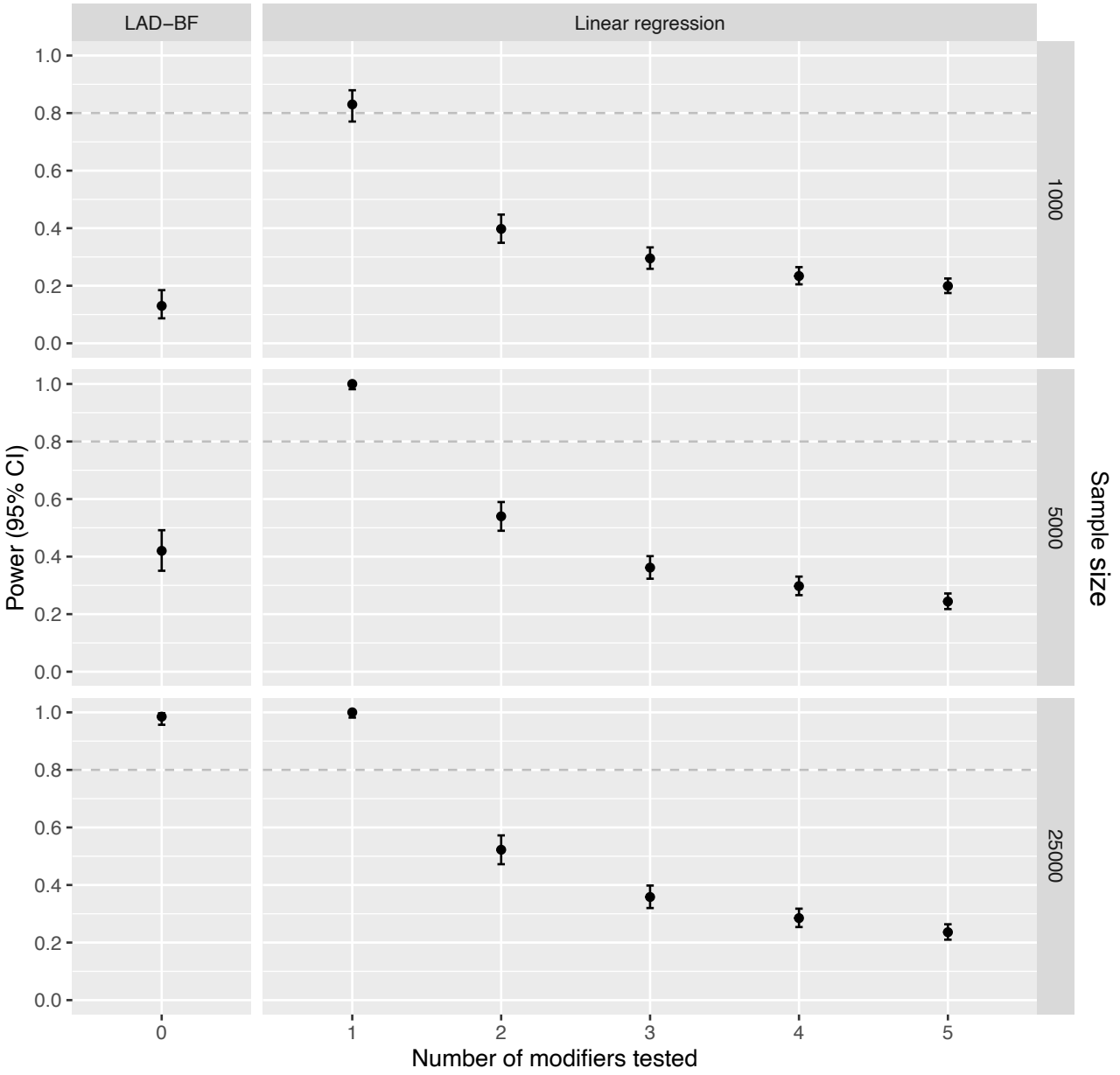
Distribution of LAD-BF test P-values under interaction effect with and without adjustment for interaction. A, No adjustment. B, Adjustment for interaction in the first-stage regression model. The SNP was simulated to have main and interaction effects explaining 6.5% and 20% of the variance of a standard Normal outcome, respectively.

4.5.6 Effect of exhaustive interaction analyses compared with variance prioritisation approach on power to detect an interaction effect when the modifier is unknown (Simulation 4.4.9)

When the sample size was N=1000 and a single interaction effect was tested (i.e., the interaction was known), linear regression had 83% power (95% CI 77%-88%) while LAD-BF had only 13% power (95% CI 9%-19%) at the same sample size. However, as the sample size increased so did power. For linear regression this was 100% power (95% CI 98%-100%) with N=5000 and 99% power (95% CI 96%-100%) for LAD-BF with N=25,000.

LAD-BF was better powered than exhaustive pairwise interaction analyses when the modifier was unknown (**Figure 4.5.6.1**). For example, under the conditions of a single true interaction and four null interactions (i.e., five tests), statistical power at N=25,000 for linear regression was 24% (95% CI 21%-26%). Under the same conditions LAD-BF power remained at 99% power (95% CI 96%-100%) as only one test was performed.

Figure 4.5.6.1. Power of linear regression and LAD-BF to detect the presence of effect modification



Power simulation to detect the presence of an interaction effect when the modifier is unknown using either linear regression or LAD-BF test with an increasing number of modifiers evaluated. All simulations contained a single true interaction effect detectable with 80% power using linear

regression when the sample size was N=1000 and up to four modifiers which had no effect. All simulations had n=200 repetitions. CI, confidence interval.

4.5.7 Runtime performance (Simulation 4.4.10)

Increasing the number of CPU threads reduced the total runtime of both methods to process 1000 SNPs (**Figure 4.5.7.1**). For the C++ implementation of LAD-BF in varGWAS, the lowest average runtime was 13.6 second (95% CI 13.5, 13.7) using four threads of an Intel Xeon CPU E5-2680 v4 @ 2.40GHz. Under the same conditions, the original Brown-Forsythe test implemented in OSCA was 1.78x faster (7.61 seconds [95% CI 7.60, 7.63]). For LAD-BF, runtime performance was slightly worse with eight threads than with four threads, this may be due to the overhead of creating and producing thread with the program. These results suggest the LAD-BF performance may be optimal with four threads.

Table 4.5.7.1. Runtime performance of varGWAS and OSCA

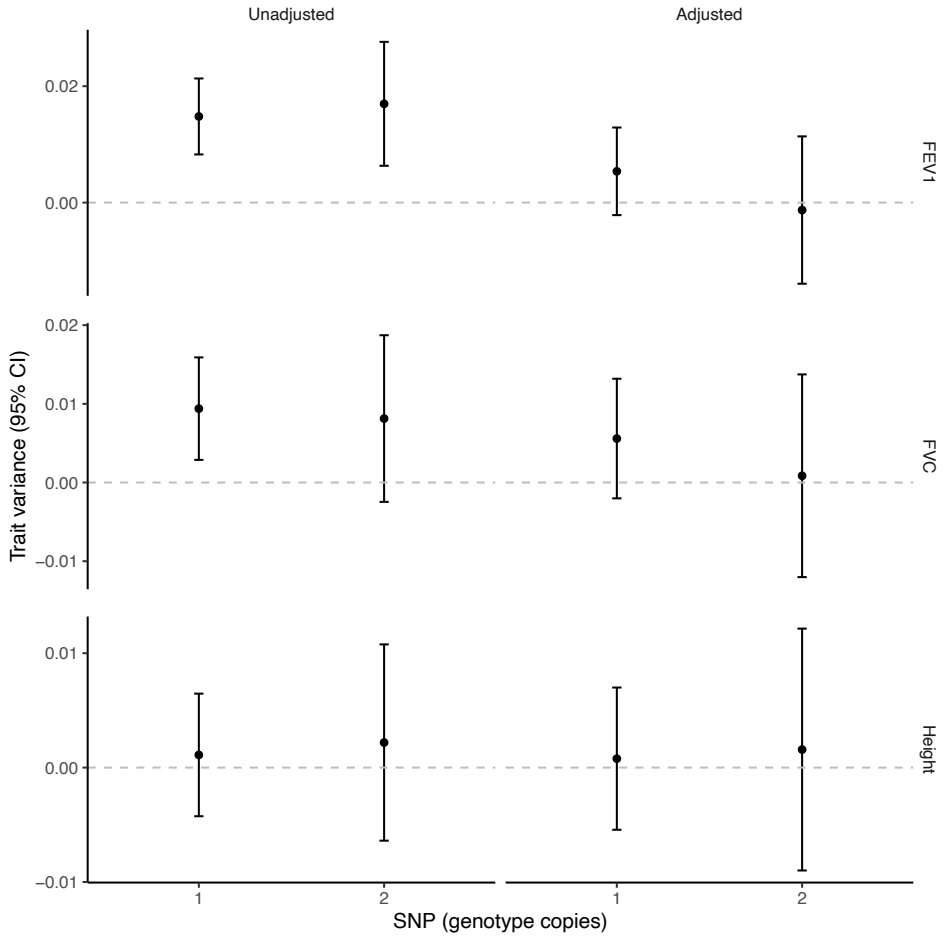
Threads	Duration	95% CI		Test
1	29.18	28.92	29.43	LAD-BF (varGWAS)
2	23.75	23.26	24.24	LAD-BF (varGWAS)
4	13.56	13.47	13.65	LAD-BF (varGWAS)
8	14.61	14.56	14.67	LAD-BF (varGWAS)
1	9.18	9.13	9.23	Brown-Forsythe (OSCA)
2	8.87	8.75	8.99	Brown-Forsythe (OSCA)
4	7.61	7.60	7.63	Brown-Forsythe (OSCA)
8	7.35	7.34	7.36	Brown-Forsythe (OSCA)

Average runtime for effect of 1000 SNPs tested on outcome variance using LAD-BF implemented in varGWAS and original Brown-Forsythe implemented in OSCA with increasing CPU threads. CI, confidence interval.

4.5.8 Positive and negative controls using data from UK Biobank

CHRNA3 rs1051730 was strongly associated **(Figure 4.5.8.1)** with a 0.015 SD (95% CI 0.008, 0.021) increase in FEV1 variance for one SNP dosage increase and 0.017 SD for two SNP dosage increase (95% CI 0.006, 0.028). SNP rs1051730 was less strongly associated with FVC variance; one SNP dosage increase of 0.009 SD (95% CI 0.003, 0.016) and two SNP dosage increase 0.008 SD (95% CI -0.002, 0.019) in FVC variance. SNP rs1051730 was not strongly associated with adult standing height variance (one SNP dosage increase of 0.001 SD [95% CI -0.004, 0.006] and two SNP dosage increase of 0.002 SD [95% CI -0.006, 0.011]). These results are consistent with the expected causal diagram shown in **Figure 4.4.11.1**, highlighting a pathway between *CHRNA3* rs1051730 and lung function modified by smoking status but no effect on adult standing height**.** Adjusting the variance effects for the interaction of *CHRNA3* rs1051730 × smoking status led to complete attenuation of the effects on FEV1 (one SNP dosage increase of 0.005 SD [95% CI -0.002, 0.013] and two SNP dosage increase of -0.001 [95% CI -0.014, 0.011]) and FVC (0.006 SD [95% CI -0.002, 0.013] and two SNP dosage increase of 0.001 SD [95% CI -0.012, 0.014]).

Figure 4.5.8.1. Per allele effect of CHRNA3 rs1051730 on variance of lung function and adult standing height



Effect of smoking heaviness variant (*CHRNA3* rs1051730) on variance of standardized lung function (FEV1 and FVC) and standardised adult standing height adjusted for age, sex and top ten genetic principal components estimated using LAD-BF. Unadjusted, the effect of rs1051730 on trait variance without adjustment for interaction effect. Adjusted, LAD-BF variance effect adjusted for interaction of rs1051730 × smoking status. FEV1, forced expiratory volume in 1-second. FVC, forced vital capacity. CI, confidence interval.

4.6 Discussion

I implemented and evaluated the LAD-BF test, a LAD regression-based[105] Brown-Forsythe test[71] with functionality to estimate an unbiased variance effect (under trait normality) and control for ancestry confounding of the variance effect. I compared this test with the Brown-Forsythe test implemented in OSCA[68,96] through a series of simulations and evaluated the test using positive and negative controls with data from UK Biobank. I provided C++ and R implementations of LAD-BF available in varGWAS and R-package, that are open-source and freely available for other researchers to use.

I obtained high correlation between the per-genotype estimated variance and expected variance. The association between the SNP and outcome variance under an interaction effect was proportional to the exposure and its square. This finding confirms the expression derived by Professor Tilling (**Chapter 9.1**) and is consistent with a previous study[56].

I compared the type I error rate of LAD-BF, Brown-Forsythe, Breusch-Pagan tests. The Brown-Forsythe and LAD-BF tests were robust to non-normality giving a null hypothesis rejection rate close to the expected error rate. But the Breusch-Pagan test showed elevated type I error rate when applied to non-normal outcomes consistent with previous research[133] and was not considered further. This is in line with previous studies which demonstrated the Brown-Forsythe test has lower type I error rate compared with other tests when applied to non-normal outcomes[68,70]. This is because the median is a more robust measure of central tendency in the presence of skew or kurtosis[81].

The power to detect genetic interaction effects using variance prioritisation was low but comparable for LAD-BF and Brown-Forsythe test. However, when applied to a large sample size

such as UK Biobank strong evidence for association of larger effects can be identified as demonstrated in **Chapter 5** and by Wang *et al*[68].

LAD-BF provides an unbiased variance effect estimate of the SNP on outcome when there is a SNP interaction effect unlike the variance estimate from OSCA and DRM which incorrectly assume linear association between SNP and outcome variance. I showed how this estimate can be adjusted for an interaction to determine if the interaction is responsible for the variance signal and if additional interactions are likely to exist and could potentially be applied using stepwise regression until all interaction effects are discovered, subject to sufficient power. I also demonstrated through simulation that adjusting the variance effect for ancestry can reduce confounding by population stratification when ancestry has main and interaction effects on the outcome. These results suggest that ancestry covariates should be included in both regression models to mitigate bias from population stratification (**Chapter 1.2.3**) of either mean or variance effects. However, this is not possible using the Brown-Forsythe test which must be applied to pre-adjusted outcomes and this approach does not control for population stratification of the variance effect.

Through simulation I also compared the value of variance prioritisation in comparison with exhaustive pairwise interaction testing to detect the presence of effect modification. Variance prioritisation had greater power than exhaustive pairwise interaction testing of hypothesised modifiers when considering five or more modifiers. However, the interaction test provides evidence on the exact interaction while variance testing can only indicate the presence of effect modification. But variance testing does not require a hypothesised or measured modifier which may lead to the detection of unanticipated findings.

In addition to extensive simulation studies, I also evaluated positive and negative controls using data from UK Biobank (**Chapter 1.3.7**). I was advised to use the known association of *CHRNA3* rs1051730 with smoking heaviness[88] to evaluate the genotypic effects on variance of lung function and adult standing height. *CHRNA3* rs1051730 is strongly associated with smoking heaviness and therefore would be anticipated to have an effect on lung function only among those who smoke[131]. Meanwhile, *CHRNA3* rs1051730 is anticipated to have no strong effect on adult standing height, although it is possible the variant may have a weak effect as *CHRNA3* rs1051730 status is predictive of parental SNPs whose smoking heaviness could influence exposure to smoke during development leading to growth restriction[134]. However, *CHRNA3* rs1051730 did not show strong association with adult standing height variance but did have a strong effect on FEV1 variance and weaker association with FVC variance. Adjusting the lung function effects for interaction with own smoking status led to complete attenuation of variance effects suggesting this interaction was driving the association of *CHRNA3* rs1051730 with lung function variance. These results validate the analysis approach and are consistent with findings of the simulation study.

4.7 Limitations

First, the test cannot be applied to imputed genotype dosage values without rounding. This is needed for modelling non-linearities in the second-stage model between SNP and outcome variance. Second, the runtime of LAD-BF was 75% longer than the Brown-Forsythe test implemented in OSCA but was still fast enough to allow large-scale analyses such as application to UK Biobank. Third, the effect estimate (but not test P-value) is based on normality assumptions which may be violated in practice. Fourth, the LAD-BF approach does

not account for imprecision in first-stage regression model residuals in the variance effect estimate standard error. However, when applied to large sample sizes such as UK Biobank, residuals may be reliably estimated[67].

4.8 Conclusions

Through extensive simulation studies and application to positive and negative controls in UK Biobank I evaluated the LAD-BF test for detection of interaction effects. To facilitate variance GWAS analyses I implemented the LAD-BF test in C++ and R.

Chapter 5: Genome-wide detection of gene-interaction effects on 30 serum biomarkers in UK Biobank using variance prioritisation

5.1 Overview

Variance prioritisation is an approach to identify genetic loci with interaction effects by estimating their association with trait variance, even where the modifier is unknown or unmeasured[56]. In **Chapter 5** I applied LAD-BF software evaluated and implemented in **Chapter 4** to 30 serum biomarkers in UK Biobank and found evidence for 468 variance quantitative trait loci across 24 biomarkers and followed up findings to detect 82 gene-environment and six gene-gene interactions independent of strong scale or phantom effects. Among these results include replication of existing findings and identify novel epistatic effects of *TREH* rs12225548 × *FUT2* rs281379 and *TREH* rs12225548 × *ABO* rs635634 on alkaline phosphatase and *ZNF827* rs4835265 × *NEDD4L* rs4503880 on gamma glutamyltransferase. These findings may help to improve our understanding of biological mechanisms underpinning biomarker concentration, weakly increase prediction of disease outcomes and in combination with other evidence support the identification of therapeutic targets for drug development[48].

5.2 Contribution statement

Work in **Chapter 5** forms part of a manuscript I wrote that was edited by PhD supervisors available as a preprint on MedRxiv (Lyon *et al*, 2022)[2]. I performed the variance GWAS and quality control analyses, tested for interaction effects of vQTLs, and performed subgroup analyses and sensitivity analyses.

5.3 Introduction

5.3.1 Variance prioritisation

Variance prioritisation[56,68,70,87] is an approach to select SNPs for GxG/GxE testing which identifies differences in outcome variance across SNP levels (variance quantitative trait loci, vQTL; **Chapter 1.5.3**). Variance QTLs arise as a consequence of heterogeneous mean effects that could occur from differences in environment or background genetic profile[67] (**Chapter 1.4.6**). Although detection of a vQTL is not conclusive evidence for interaction (**Chapter 1.5.3**), this observation is consistent with SNP-interaction effects[67] and detection of vQTLs does not require the modifier to be measured[67].

5.3.2 LAD-BF

In **Chapter 4** I implemented and evaluated the least-absolute deviation regression[105] Brown-Forsythe test[71] (LAD-BF) which provides greater flexibility to enable adjustment of covariates and provide an unbiased variance effect estimate. This model used the same structure as the Breusch-Pagan test[78] (**Chapter 2.2.3**) which evaluates the presence of heteroscedasticity through two independent regression models but using least-absolute deviation[105] (LAD) regression in the first-stage instead of OLS. LAD regression estimates the mean exposure effect on outcome median (rather than mean as with OLS) providing robustness to trait non-normality[105].

5.3.3 Aims

This chapter aims to apply the LAD-BF model evaluated in **Chapter 4** to estimate SNP effects on the variance of 30 serum biomarkers in approximately 300K UK Biobank participants

and follow up vQTLs with formal interaction tests to detect GxG/GxE interactions as well as adjusting vQTL signals for these interactions to consider impact on variance effect attenuation.

5.4 Materials and methods

5.4.1 Phenotypes

All 30 serum biomarkers measured in UK Biobank were evaluated in this study as described in the documentation[117] (**Table 3.3.4.1**; **Chapter 1.6; Chapter 3.3.4**). For each GWAS and follow up analysis participants with missing data were excluded. All continuous outcomes were placed on the same scale by dividing each outcome by its standard deviation irrespective of distribution shape.

5.4.2 Variance genome-wide association studies (vGWAS)

Biomarker vGWAS was performed using the LAD-BF test (**Chapter 2.2.4**) evaluated in **Chapter 4** adjusted for age, sex, and the first ten genetic principal components in both regression models. I removed outlier biomarker values with a Z-score > 5 SD from the mean and restricted the analysis to MAF > 5% to control type I error inflation as previously described[68]. Qualitative quality control was undertaken using Q-Q plots of each GWAS to check for a departure of P-value distribution from that expected under the null. Independent vQTLs were identified by clumping GWAS loci that passed the experiment-wise genome-wide evidence threshold $P < 1.67 \times 10^{-9}$ (Bonferroni correction of standard GWAS threshold: $p = 5 \times 10^{-8} / 30$) using the OpenGWAS API[18] with default R^2 threshold of 0.001 and the 1000 genomes phase 3 European ancestry reference panel[135].

5.4.3 Gene-interaction tests

Independent vQTLs were tested for interaction effects on additive and multiplicative scales using heteroscedasticity-consistent standard errors[107] adjusted for age, sex, and first ten genetic principal components (**Chapter 2.2.5**). To ensure effects were robust to phantom effects[53,68], I performed sensitivity analyses adjusting for fine-mapped main effects identified using SuSiE-RSS[136] (**Chapter 1.2.4**). Interactions surpassing genome-wide association significance ($P < 5 \times 10^{-8}$) on both scales that did not attenuate to null with adjustment for fine-mapped main effects were prioritised for subgroup analyses. GxG effects were identified through interaction testing with independent ($R^2 < 0.001$) vQTLs excluding pairwise combinations of vQTLs within a 10Mb window as previously described[68]. GxE testing was performed using a set of candidate modifiers: age (SD), sex (SD), BMI (SD), alcohol intake (SD), smoking status (SD), total physical activity (SD), daily sugar intake (SD), and daily fat intake (SD). These were chosen to include the modifiers evaluated in Wang *et al*[68] (sex, age, physical activity, and smoking) supplemented with additional related phenotypes (BMI, alcohol intake, daily sugar and fat intake). Total physical activity was calculated by summing self-reported duration of walking, moderate and vigorous activity collected using the International Physical Activity Questionnaire as described[137]. Alcohol intake was self-reported from the question "About how often do you drink alcohol?" with six possible responses ranging from never to daily. Smoking status was derived by UK Biobank from several questions with possible values of "never", "current" and "previous".

5.4.4 Subgroup analyses

To determine if top interaction effects have a qualitative interaction effect (**Chapter 1.4.5**), I performed subgroup analyses estimating the SNP-outcome mean effect across levels of the modifier (dichotomising continuous measures). The aim of this analysis was to determine if the SNP had opposing effects between subgroups or no effect in one subgroup which may be of clinical relevance[57]. These stratified estimates were adjusted for age, sex and top ten genetic principal components (except for age and sex modifiers where these covariates were removed) and used heteroscedasticity-consistent standard errors[107].

For GxG analysis, categorical modifiers were rounded genetic dosage values. GxE subgroup analyses were performed using dichotomous modifier groups $k1$ and $k2$ as follows: below or above the median value for continuous variables ($k1$ below median; $k2$ median or greater), ever ($k1$) vs never ($k2$) smoker, alcohol intake once a week or more ($k1$) vs less than once a week on average ($k2$), males ($k1$) vs females ($k2$). Subgroup effects are presented along with the SNP-variance estimates adjusted for age, sex and first ten genetic principal components with and without adjustment for the interaction term in both models.

5.4.5 Gene annotation

Variance QTLs were annotated with the nearest gene using the closest function of bedtools[138] (v2.3.0) and gene coordinates from Ensembl[139] v104 (GRCh37) protein-coding features which were filtered to retain HUGO[140] valid identifiers. For the top interactions reported, I recoded the gene annotation using expression QTL evidence in blood[141,142]: rs4530622 *SLC2A9*, rs11244061 *ABO*, rs71633359 *HSD17B13*, rs28413939 *TREH*, rs281379 *FUT2*, rs635634 *ABO*, rs964184 *APOA5*.

5.4.6 Fine mapping of main effect SNPs

Fine-mapping was performed between natural linkage disequilibrium break points identified in European populations[95] containing the interacting variant using SuSiE-RSS[136] assuming at most 10 causal variants. The data were processed using gwasglue R-package[18]. Summary statistics for the SNP-biomarker main effects were obtained from Neale *et al*[18,123]. European 1000 genomes[135] phase 3 linkage disequilibrium matrices were obtained from OpenGWAS containing bi-allelic SNPs with MAF > 0.01[18]. This methodology was obtained from the gwasglue R-package[18] documentation.

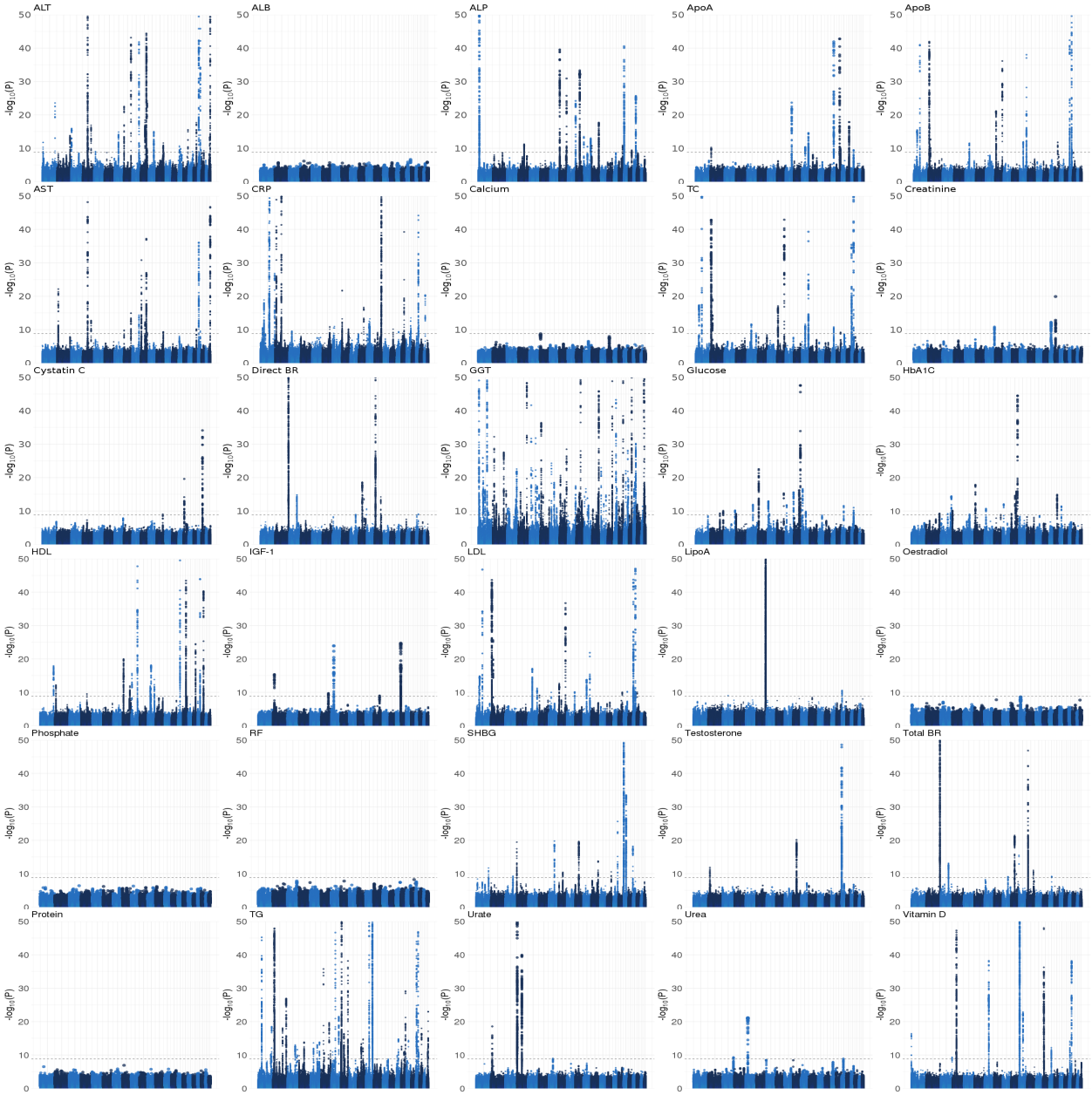
5.5 Results

5.5.1 GWAS of variance effects in UK Biobank

In UK Biobank, biomarkers have different levels of missingness with sample sizes ranging from 28,680 (rheumatoid factor) to 321,260 (total cholesterol) (**Table 3.3.4.1**). There was also considerable departure from normality for many biomarkers. Both sample size and non-normality were shown to contribute to lower power in my simulation study (**Chapter 4**). Non-normality was also associated with elevated type I error rate (**Chapter 4**) and vQTL findings of non-Normal outcomes are likely to include false positives. However, as this approach is a screening method to prioritise loci for further analysis some false positives are acceptable. I found evidence of 468 independent ($R^2 < 0.001$) vQTLs influencing 24 biomarkers (**Figure 5.5.1.1**; **Figure 5.5.1.2**) below the experiment-wise P-value threshold (1.67×10^{-9}) and no variance effects for albumin, calcium, oestradiol, phosphate, rheumatoid factor, or total protein. Oestradiol and rheumatoid factor were only available on a small subset of 50,380 and 28,680 participant, respectively resulting in lower power. Of the identified vQTLs, 183 (39.1%)

had evidence for a variance effect on the multiplicative scale ($P < 5 \times 10^{-8}$) and 453 (96.8%) had a mean effect ($P < 5 \times 10^{-8}$). The low concordance between additive and multiplicative scales and high concordance between mean and variance effects suggests the presence of mean-variance relationships which is a likely consequence of extreme non-normality for some of the trait distributions (**Figure 3.3.4.1**). GGT and TG had very strong evidence for variance association across the genome (**Figure 5.5.1.1**) but were left-skewed (**Figure 3.3.4.1**) implying elevated type I error rate rather than findings of biological significance. Nevertheless, I found evidence for GxE on GGT and TG at these variance loci (**Chapter 5.5.2**). During thesis development, another study reported vQTL analyses of 20 serum biomarkers in UK Biobank[42]. Of the top vQTLs reported here, N=293 (68.5%) were also identified by this study[42]. Additionally, another study performed a vGWAS for serum vitamin D[90] also in UK Biobank which detected 11 of the 15 vQTLs reported here (73.3%).

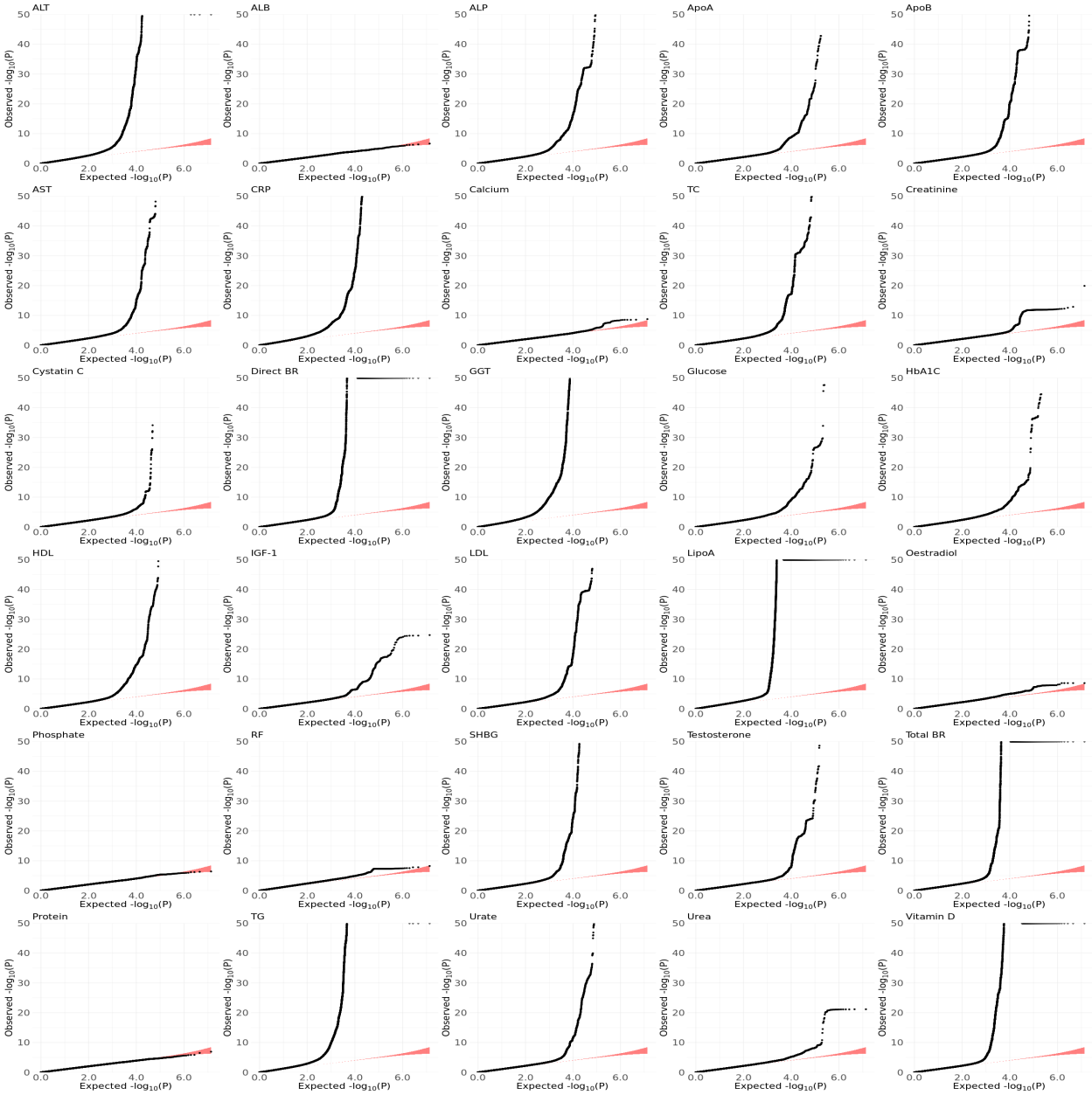
Figure 5.5.1.1. Manhattan plots of biomarker variance GWAS using regression-based Brown-Forsythe test



Manhattan plots for GWAS of SNP effects on biomarker variance in UK Biobank using the LAD-BF test adjusted for age, sex, and top ten genetic principal components. ALB, albumin. ALP, alkaline phosphatase. ALT, alanine aminotransferase. AST, aspartate aminotransferase. ApoA, Apolipoprotein A. ApoB, apolipoprotein B. CRP, C-reactive protein. Direct BR, direct bilirubin.

GGT, Gamma glutamyltransferase. HDL, high-density lipoprotein. HbA1C, glycated haemoglobin. IGF-1, insulin growth factor 1. LDL, low-density lipoprotein. LipoA, lipoprotein A. RF, rheumatic factor. SHBG, sex-hormone binding globulin. TC, total cholesterol. TG, triglycerides. Total BR, total bilirubin. Biomarker outliers with Z-score > 5SD from the mean were removed to control type I error rate. Y axis is capped at Y=50. Horizontal dashed line marks experiment-wide P-value threshold ($P < 1.67 \times 10^{-9}$).

Figure 5.5.1.2. Q-Q plots of biomarker variance GWAS using LAD regression-based Brown-Forsythe test



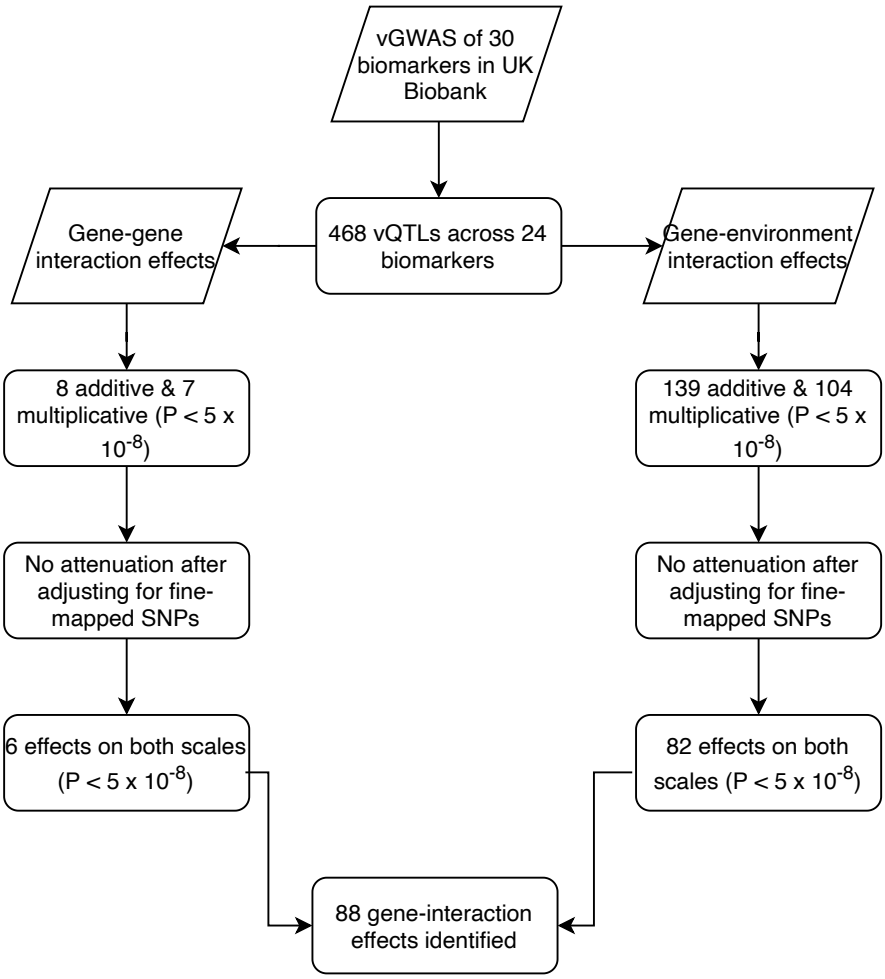
Q-Q plots for GWAS of SNP effects on biomarker variance in UK Biobank using the LAD-BF test adjusted for age, sex, and top ten genetic principal components. ALB, albumin. ALP, alkaline phosphatase. ALT, alanine aminotransferase. AST, aspartate aminotransferase. ApoA, Apolipoprotein A. ApoB, apolipoprotein B. CRP, C-reactive protein. Direct BR, direct bilirubin.

GGT, Gamma glutamyltransferase. HDL, high-density lipoprotein. HbA1C, glycated haemoglobin. IGF-1, insulin growth factor. LDL, low-density lipoprotein. LipoA, lipoprotein A. RF, rheumatic factor. SHBG, sex-hormone binding globulin. TC, total cholesterol. TG, triglycerides. Total BR, total bilirubin. Biomarker outliers with Z-score > 5SD from the mean were removed to control type I error rate. Y axis is capped at Y=50.

5.5.2 Gene-environment interaction effects (GxE)

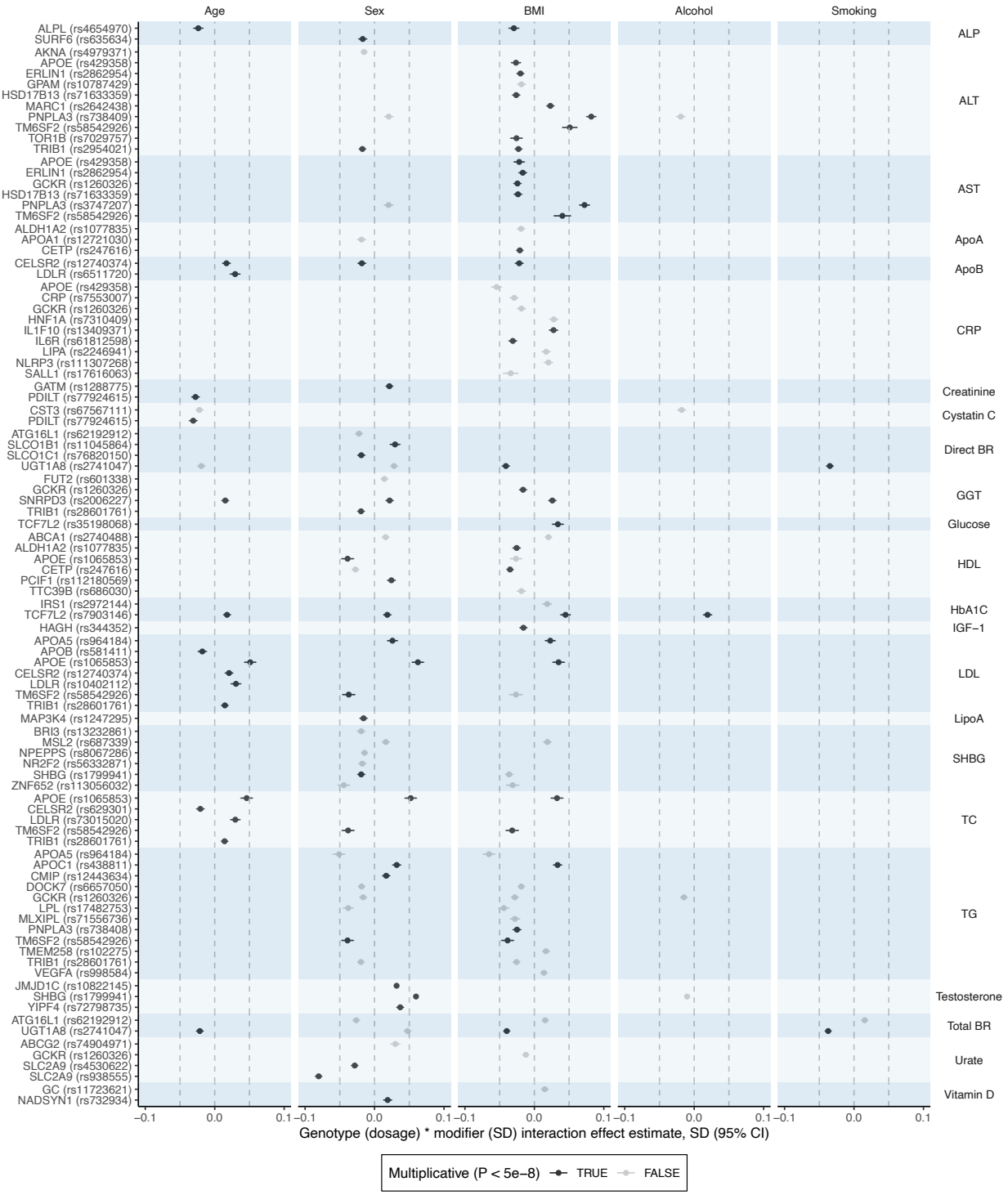
I detected 139 additive and 104 multiplicative GxE effects (i.e., using natural and log scales) using the standard genome-wide significance threshold ($P < 5 \times 10^{-8}$; **Figure 5.5.2.1**; **Figure 5.5.2.2**; **Figure 9.2.1**). This threshold was chosen to prioritise effects for further analysis while acknowledging that far fewer SNP-interaction tests have been performed than would be required under a Bonferroni corrected P-value at this threshold. These findings include evidence of effect modification by all phenotypes except physical activity and sugar/fat intake. Self-reported diet[143] and exercise[144] phenotypes have high measurement error leading to lower power which may explain the lack of association. Adjusting the additive effects for fine-mapped main effects (**Figure 9.2.2**) led to a small increase in effect of *UGT1A8* rs2741047 × sex on direct bilirubin to 0.037 SD (95% CI 0.032, 0.042) from 0.028 SD (95% CI 0.023, 0.033) and minor attenuation of *MAP3K4* rs1247295 × sex on lipoprotein a to -0.011 SD (95% CI -0.015, -0.007) from -0.016 SD (95% CI -0.021, -0.010). These findings could reflect the presence of large main effects in imperfect linkage disequilibrium ($R^2 < 1$) with the index SNP (i.e., phantom effects) which is known to elevate type I error rate[53,68,93].

Figure 5.5.2.1. UK Biobank analysis flowchart



Flow diagram of variance GWAS study investigated in **Chapter 5** and numbers of interaction/vQTL effects obtained.

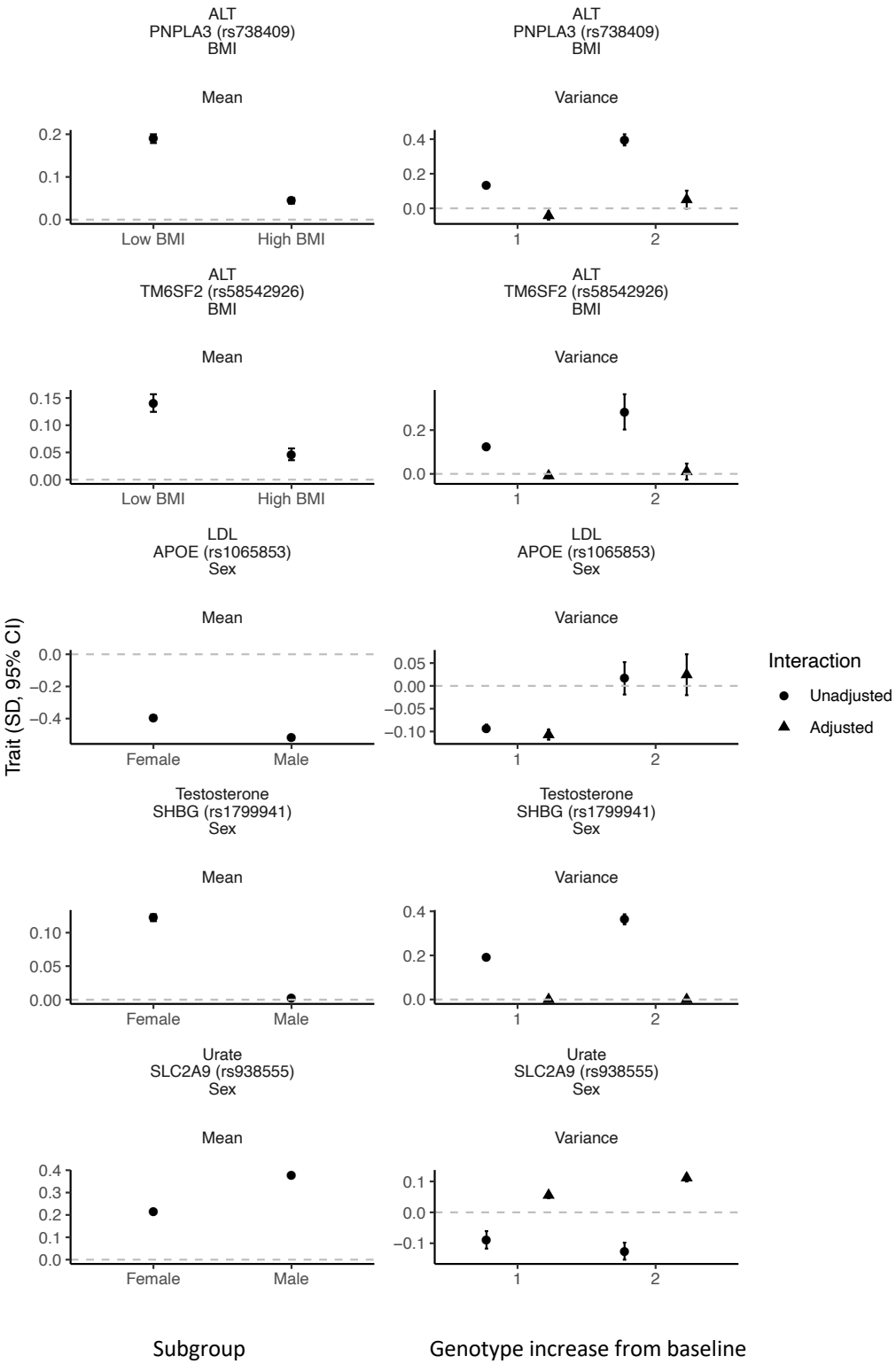
Figure 5.5.2.2. Top gene-by-environment interaction effects ($P < 5 \times 10^{-8}$) on biomarker concentration using additive scale



GxE effects using additive scale and heteroscedasticity consistent standard errors[107] ($P < 5 \times 10^{-8}$). ALP, alkaline phosphatase. ALT, alanine aminotransferase. AST, aspartate aminotransferase. ApoA, Apolipoprotein A. ApoB, apolipoprotein B. CRP, C-reactive protein. Direct BR, direct bilirubin. GGT, Gamma glutamyltransferase. HDL, high-density lipoprotein. HbA1c, glycated haemoglobin. IGF-1, insulin-like growth factor 1. LDL, low-density lipoprotein. LipoA, lipoprotein A. SHBG, sex-hormone binding globulin. TC, total cholesterol. TG, triglycerides. Total BR, total bilirubin. BMI, body mass index. Smoking, smoking status. Alcohol, intake. PA, physical activity. All measures reported on SD scale. All estimates were adjusted for the main effect, age, sex, and top ten genetic principal components. Gene name is the nearest protein coding gene HGNC name by chromosomal position. Vertical dashed lines are present at -0.05, 0 and 0.05 SD. The legend indicates if effect estimates were also strong on the multiplicative i.e., log scale. SD, standard deviation. CI, confidence interval.

I prioritised 82 GxE effects with evidence on both scales ($P < 5 \times 10^{-8}$) to avoid spurious interactions dependent on scale (**Figure 5.5.2.1**). Of these, BMI (n=35), sex (n=27) and age (n=17) were responsible for modification of most effects while smoking status (n=2) and alcohol intake (n=1) modified fewer effects. This could be because smoking and alcohol measures were self-reported and categorical and therefore having high measurement error leading to lower power to detect effects[143]. The largest interaction effects (**Figure 5.5.2.3**) were: *PNPLA3* rs738409 × BMI on alanine aminotransferase (ALT; 0.08 SD [95% CI 0.08, 0.09]), *SLC2A9* rs938555 × sex on urate (-0.08 SD [95% CI -0.09, -0.08]), *APOE* rs1065853 × sex on low-density lipoprotein (LDL; 0.06 SD [95% CI 0.05, 0.07]), *SHBG* rs1799941 × sex on testosterone (0.06 SD [95% CI 0.06, 0.06]) and *TM6SF2* rs58542926 × BMI on ALT (0.05 SD [95% CI 0.04, 0.06]). Adjusting the variance effect for the interaction term (**Figure 5.5.2.3**) led to attenuation of *PNPLA3* rs738409 and *TM6SF2* rs58542926 on ALT and *SHBG* rs1799941 on testosterone but strong variance effects on ALT remained at *PNPLA3* rs738409 (LAD-BF P_adjust = 1.0×10^{-73}) and *TM6SF2* rs58542926 (LAD-BF P_adjust = 1.84×10^{-8}). There was no strong variance attenuation of *APOE* rs1065853 on LDL or *SLC2A9* rs938555 on urate following adjustment for the interaction (**Figure 5.5.2.3**).

Figure 5.5.2.3. Effect of top gene-environment interaction loci on trait mean and variance

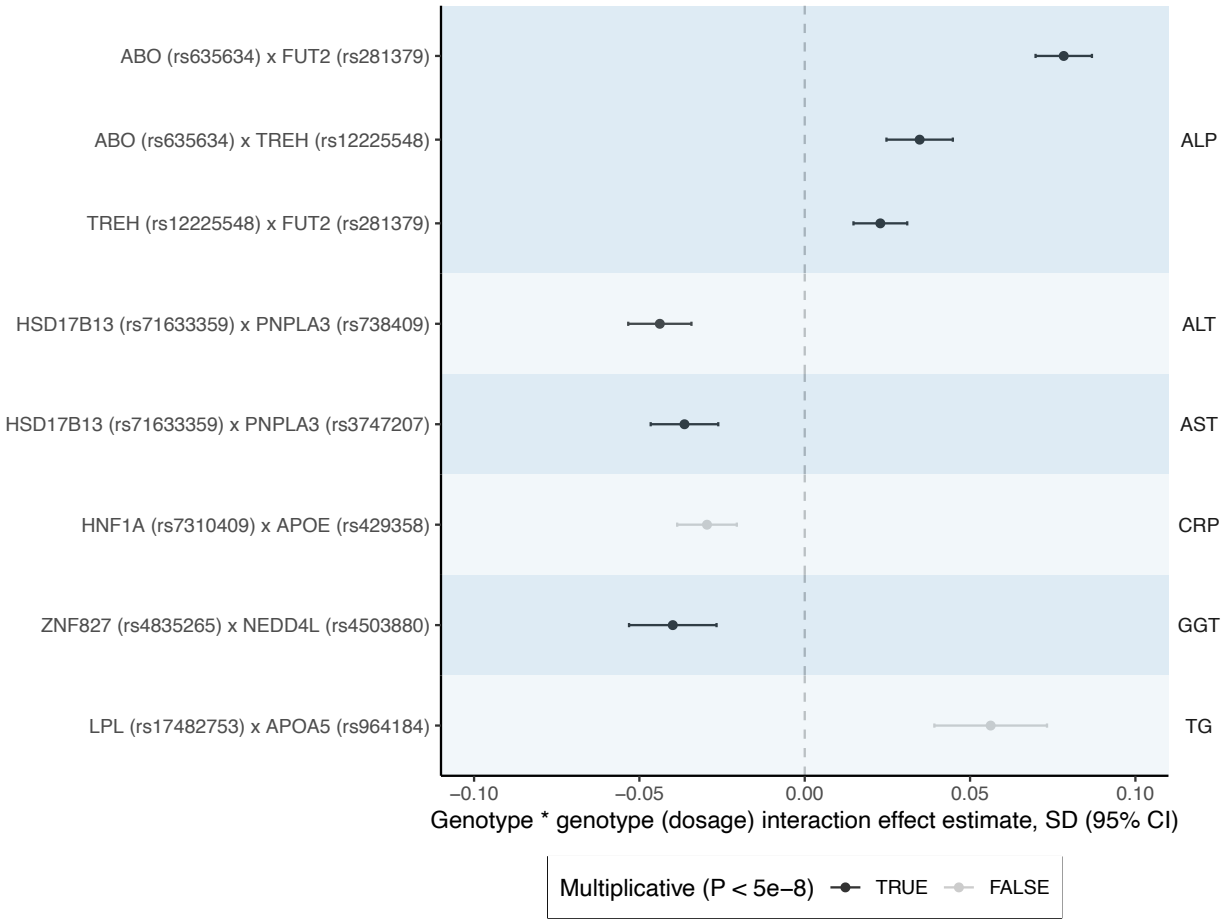


Effect of SNP stratified by modifier on outcome mean (per-allele) estimated with heteroscedastic-consistent standard errors[107] and unstratified effect of SNP on variance estimated using LAD-BF (SNP dosage 0 vs 1 and 0 vs 2) with or without adjustment for the interaction term. All estimates were adjusted for age, sex (except for rs1065853, rs1799941 and rs938555 on variance as the modifier was sex) and top ten genetic principal components. Horizontal dashed line marks null association. These estimates were selected as the largest five interaction effect sizes. SD, standard deviation. CI, confidence interval. ALT, alanine aminotransferase. LDL, low-density lipoprotein. BMI, body mass index. Low BMI, <= 26.7 kg/m^2. High BMI, > 26.7 kg/m^2.

5.5.3 Gene-gene interaction effects (GxG)

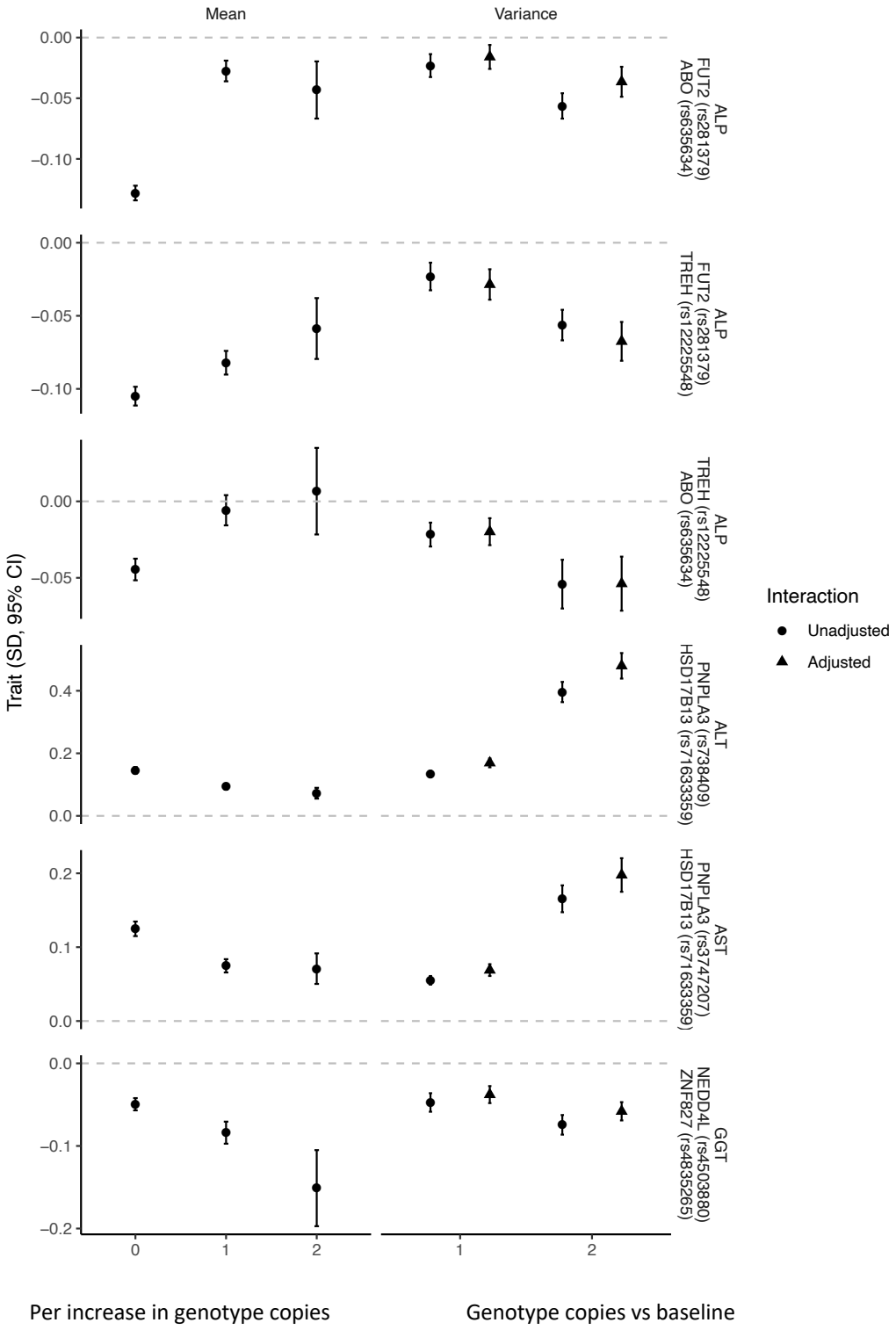
I detected eight GxG effects on the additive scale (**Figure 5.5.3.1**), six of which were also associated on the multiplicative scale (**Figure 9.2.3**) using standard genome-wide significance threshold to prioritise effects ($P < 5 \times 10^{-8}$). There was no strong attenuation following adjustment for fine-mapped main effects (**Figure 9.2.4**) which does not support but cannot exclude phantom epistasis[53,68,93] as a major source of bias. *ZNF827* rs4835265 × *NEDD4L* rs4503880 was associated with -0.04 SD (95% CI -0.05, -0.03) gamma glutamyltransferase (GGT), *ABO* rs635634 × *FUT2* rs281379, *ABO* rs635634 × *TREH* rs12225548, and *TREH* rs12225548 × *FUT2* rs281379 were associated with 0.08 SD (95% CI 0.07, 0.09), 0.04 SD (95% CI 0.03, 0.05) and 0.02 SD (95% CI 0.02, 0.03) increase in alkaline phosphatase (ALP) respectively, *HSD17B13* rs71633359 x *PNPLA3* rs738409 and *HSD17B13* rs71633359 × *PNPLA3* rs3747207 were associated with -0.04 SD (95% CI -0.05, -0.03) and -0.04 SD (95% CI -0.05, -0.03) decrease in ALT and aspartate aminotransferase (AST) respectively (**Figure 5.5.3.2**). Adjusting the variance effects for the interaction term had no strong impact on the variance estimate (**Figure 5.5.3.2**).

Figure 5.5.3.1. Top gene-by-gene interaction effects ($P < 5 \times 10^{-8}$) on biomarker concentration using additive scale



GxG effects using additive scale and heteroscedasticity consistent standard errors[107] ($P < 5 \times 10^{-8}$) adjusted for the main effect, age, sex, and top ten genetic principal components. ALP, alkaline phosphatase. ALT, alanine aminotransferase. AST, Aspartate aminotransferase. CRP, C-reactive protein. GGT, Gamma glutamyltransferase. TG, triglycerides. All measures reported on SD scale. Gene name is the nearest protein coding gene HGNC name by chromosomal position. Vertical dashed line marks null association. SD, standard deviation. CI, confidence interval.

Figure 5.5.3.2. Effect of top gene-gene interaction loci on trait mean and variance



Effect of SNP stratified by genetic modifier on outcome mean (per-allele) estimated with heteroscedastic-consistent standard errors[107] and unstratified effect of SNP on variance estimated using LAD-BF (SNP dosage 0 vs 1 and 0 vs 2) with or without adjustment for the interaction term (indicated by the legend). All estimates were adjusted for age, sex, and top ten genetic principal components. SD, standard deviation. CI, confidence interval. ALP, alkaline phosphatase. ALT, alanine aminotransferase. AST, aspartate aminotransferase. GGT, gamma glutamyltransferase. Horizontal dashed line marks null association.

5.5.4 Replication

The largest GxE and GxG effects overlapped with previously reported interaction effects (**Table 5.5.4.1**).

Table 5.5.4.1. Replication of top gene-interaction effects

Association	Sign	P	Studies
PNPLA3 rs738409 × BMI on ALT	+++++	3.37 x 10^{-119}; 6 x 10^{-5}; 3 x 10^{-10}; 7 x 10^{-15}; 0.02	UK Biobank; Dallas Heart study[145]; Dallas Biobank[145]; Copenhagen cohort[145]; PANIC study[146]
SLC2A9 rs938555 × sex on urate	-----	1.05 x10^{-232}; 3.93 x 10^{-10}; 8.79 x 10^{-31}; 5.56 x 10^{-18}; 3.51 x 10^{-6}	UK Biobank; KORA F3 500K[147]*; KORA S4[147]*; SAPHIR[147]*; FHS[147]**, ARIC[148]**, CARDIA[148]**, CHS[148]**
TM6SF2 rs58542926 × BMI on ALT	+??+	4.52 x 10^{-21}; 0.14; 0.39; 1 x 10^{-4}	UK Biobank; Dallas Heart study[145]; Dallas Biobank[145]; Copenhagen cohort[145]
HSD17B13 rs71633359 × *PNPLA3* rs738409 on ALT	--	2.57 x 10^{-19}; 0.002	UK Biobank; DiscovEHR***[149]
HSD17B13 rs71633359 × *PNPLA3* rs3747207 on AST	--	3.03 x 10^{-12}; 0.004	UK Biobank; DiscovEHR***[149]

BMI, body mass index. ALT, alanine aminotransferase. LDL, low-density lipoprotein. ALP, alkaline phosphatase. AST, aspartate aminotransferase. PANIC, Physical Activity and Nutrition in Children. KORA, Kooperative Gesundheitsforschung in der Region Augsburg. SAPHIR, Salzburg Atherosclerosis Prevention Program in Subjects at High Individual Risk. ARIC, Atherosclerosis

Risk in Communities study. FHS, Framingham Heart Study. CARDIA, Coronary Artery Risk Development in Young Adults study. CHS, Cardiovascular Health Study. *Used proxy SNP rs6855911. **Used proxy rs6449173. ***Used proxy rs72613567 and rs738409.

5.6 Discussion

I demonstrated the value of variance GWAS (**Chapter 1.2.2**; **Chapter 1.5.3**) in identifying 468 independent vQTLs implying potential evidence of interaction on 24 serum biomarkers in UK Biobank and subsequently identified evidence for 82 GxE and six GxG scale-independent effects.

Of the vQTL effects, 15 influenced vitamin D in comparison with 25 identified in a previous study[90] using OSCA and UK Biobank but the authors stratified analyses by self-reported vitamin supplement intake which may reduce residual variance increasing power to detect effects. A second study[42] performed variance GWAS of 20 biomarkers and detected 182 vQTLs (in contrast with 468 vQTLs I found) for 30 biomarkers presented in this analysis. An important difference is that the authors of this study[42] performed analyses on the log-scale while I chose to use the natural scale. Use of the log-scale introduces mean-variance confounding and should be avoided[68] (unless the natural scale is the log-scale) but may be useful for sensitivity studies to confirm the associations are not scale dependent[92].

The largest GxE effects replicated existing findings: *PNPLA3* rs738409 × BMI on ALT levels[146,150], *SLC2A9* rs938555 × sex on urate[147,148], and *TM6SF2* rs58542926 × BMI on ALT[150]. Association of *SHBG* rs1799941 × sex on testosterone[151] was also consistent with previous work performed in UK Biobank. Adjusting the variance effect for the identified interaction led to attenuation of *PNPLA3* rs738409 and *TM6SF2* rs58542926 on ALT and *SHBG* rs1799941 on testosterone suggesting no further interaction effects exist at these loci, but the test may be underpowered to detect such effects. Strong evidence of variance effects remained for ALT at

PNPLA3 rs738409 and *TM6SF2* rs58542926 suggesting other interaction effects may exist at these loci.

I replicated previous GxG effects *HSD17B13* rs71633359 × *PNPLA3* rs738409/rs3747207 on ALT and AST[149,152] and found no strong evidence of 'phantom epistasis'[53,68,93] as a potential explanation but cannot exclude its presence. Additionally, I identified novel effects of *TREH* rs12225548 × *FUT2* rs281379 and *ABO* rs635634 × *TREH* rs12225548 on ALP and *ZNF827* rs4835265 × *NEDD4L* rs4503880 on GGT. ABO blood group antigens and secretion status are thought to influence ALP clearance[153,154]. *TREH* rs12225548 was previously reported to have a strong main effect on ALP[18,123,155] and interactions of these loci may be explained by interplay of ALP production and clearance mechanisms. *ZNF827* and *NEDD4L* loci have previously been reported to influence GGT levels in independent populations but the mechanism remains unclear[156,157].

None of the GxG variance effects strongly attenuated after adjusting for the interaction term. This is likely a consequence of low power since the GxG effects explained a very small amount of the trait variance but could also indicate the presence of other interaction effects involving the same SNP not included in the variance model. Indeed, I found strong GxE evidence at some of these loci: *ABO* rs635634 × sex on ALP, *HSD17B13* rs71633359 × BMI and *PNPLA3* rs738409/rs3747207 × BMI on ALT and AST.

These interaction findings may help to improve our understanding of disease mechanisms and biology influencing biomarker concentration[4,48]. This evidence may also contribute towards developing prediction models for biomarker concentration from genetic and environmental factors and may help to predict disease outcomes[4], although the size of

these interaction effects are small and may only weakly increase explained variance. Finally, interaction loci may be used in combination with other evidence to characterise patient subgroups in whom therapies have a differential effect which is important for developments in precision medicine[48,57].

5.7 Limitations

There are other explanations for these potential vQTLs that are not biological. First, loci that are weakly correlated with a SNP having a strong main effect can introduce a phantom vQTL[53,68,93] (**Chapter 1.54**). In this situation variance is introduced through variability in LD between the supposed vQTL and QTL. Second, I assumed homogeneity of variance within each SNP group which could be violated by the mean-variance relationship[68] (**Chapter 1.5.4**) and observed low concordance of vQTL effects on the multiplicative and additive scales are evidence for this. Additionally, these interactions could be explained by non-linear relationships between the exposure and outcome or scale artefacts[158] (**Chapter 1.5.4**). I sought to reduce the latter by replicating effects on additive and multiplicative scales. Finally, novel interaction effects require independent replication studies for confirmation[48].

5.8 Conclusions

Through this work I performed a genome-wide screen for genetic interaction effects on 30 serum biomarkers in UK Biobank using variance prioritisation[56] and found evidence for 88 interaction effects. Many of the top findings replicated previously reported associations, but I also reported first evidence of *TREH* rs12225548 × *FUT2* rs281379 and *TREH* rs12225548 × *ABO* rs635634 on ALP and *ZNF827* rs4835265 × *NEDD4L* rs4503880 on GGT. Additionally, I showed variance attenuation of *PNPLA3* rs738409 and *TM6SF2* rs58542926 on ALT and *SHBG* rs1799941

on testosterone after adjusting for the interaction indicating these effects were contributing to the variance association, but the ALT effects were still strong suggesting additional interactions may exist at these loci. These data may add to existing knowledge in understanding of disease biology[48], weakly improve disease prediction[48] and in combination with other data, help to identify opportunities for drug development[48].

Chapter 6: Examining the evidence for Mendelian randomization homogeneity assumption violation using instrument association with exposure variance

6.1 Overview

Mendelian randomization is an instrumental variable (IV) technique for evaluating the causal effect of an exposure on an outcome using genetic variants[15] and requires three core assumptions of relevance (IV1), exchangeability (IV2), and exclusion restriction (IV3; **Chapter 1.3.5**; **Chapter 1.3.6**). Identification of a well-defined causal point estimate requires additional assumptions of homogeneity for the IV-exposure and/or exposure-outcome relationships[35]. While it is not possible to verify if homogeneity assumptions hold, empirical evidence against this assumption may be observed in the data[34]. Previous research has suggested testing for IV interaction effects on exposure using a set of effect modifiers[34]. However, this requires that modifiers are hypothesised and measured[34]. In this chapter, I evaluate the utility of testing for IV-exposure variance effects to provide evidence against homogeneity assumptions[35]. Secondly, I evaluate the utility of removing IVs from the MR analysis that show strong association with exposure variance (hence are likely to have heterogeneous effects). I apply these approaches to investigate the effects of LDL, urate and glucose on cardiovascular disease, gout, and type 2 diabetes, respectively finding no strong evidence of violation of the IV-exposure homogeneity assumption. These approaches could be applied in the future when larger sample sizes are available to gain improved understanding of the MR causal estimand.

6.2 Contribution statement

Work in **Chapter 6** is from a manuscript in preparation I wrote and was edited by PhD supervisors and Dr Fernando Hartwig (University of Pelotas). I performed the simulation studies and applied analyses described in this chapter.

6.3 Introduction

6.3.1 Background

Testing for IV-exposure effect modification may be used as an empirical approach to detect violation of the IV homogeneity assumption[34]. The NO Simultaneous Heterogeneity[35] (NOSH) assumption implies the population average causal effect (PACE) can be identified if there is effect modification of either the instrumental variable (IV) effect on the exposure or exposure effect on the outcome, provided that effect modifiers for the IV-exposure and exposure-outcome relationships are independent (NOSH assumption one) and the exposure-outcome relationship is additive linear (NOSH assumption two).

Hypothesised testing of candidate IV-exposure interaction effects to evaluate homogeneity assumptions has been suggested[34] but this approach may miss unanticipated interaction effects, cannot be used if the modifier is unmeasured, and potentially incurs a large multiple testing burden[4,42]. Alternatively, the presence of effect modification can be identified by testing the association of the IV with exposure variance provided that the exposure is continuous[2,43,68]. This evidence could be used to evaluate homogeneity assumptions to draw conclusions on departure from PACE. Secondly, in a multi-IV setting such as MR[15] (**Chapter 1.3.6**), IVs with strong exposure variance effects could be removed from the analysis to produce a causal estimate closer to PACE.

6.3.2 Aims

In this study I explore the utility of testing for IV-exposure variance effects to provide empirical evidence of IV-exposure homogeneity violation using simulation studies. Secondly, I apply this approach to MR and examples with data from UK Biobank and large GWAS consortia. First, I propose a falsification strategy where IV-exposure variance effects are used to provide evidence against homogeneity. Second, I demonstrate that evidence of IV-exposure variance effects can be used in sensitivity analyses that remove instruments with the strongest variance effects from MR estimation.

6.4 Methods

6.4.1 Summary of simulation studies

The following section describes a series of simulation studies (**Table 6.4.1.1**) undertaken throughout **Chapter 6** which aim to evaluate the utility of testing IV-exposure variance effects in estimating the PACE. First, I varied the IV-exposure and exposure-outcome interaction effect size to determine the consequences of NOSH assumption one violation on PACE bias (**Chapter 6.4.2**). I also tested the IV-exposure variance effect to determine the conditions under which this evidence can provide information against targeting PACE. Second, I extended this simulation to estimate the relative bias of PACE and related the magnitude of this bias to estimates for the strength of the IV-exposure variance effect (**Chapter 6.4.3**). Third, I explored the utility of removing instruments with evidence for IV-exposure variance effects from the IVW analysis on PACE bias and IVW causal test efficiency (**Chapter 6.4.4**).

Table 6.4.1.1. Simulation study summary

Simulation	Aim	Description
6.4.2	Estimate bias of PACE under NOSH assumption one violation	Simulating NOSH assumption one violation and estimating the causal effect. Determine if the IV-exposure variance test can detect NOSH assumption violation.
6.4.3	Estimate relative bias of PACE for increasing interaction effect size and compare this with IV-exposure variance test power	Estimate relative bias of PACE for increasing IV-exposure interaction effect size and fixed exposure-outcome interaction effect. Relate the magnitude of this bias to IV-exposure variance testing.
6.4.4	Determine if removing instruments with interaction effects on exposure can attenuate IVW bias	Estimate the IVW effect using instruments with/without IV-exposure interaction effects. Apply IV-exposure variance test to remove instruments with interaction effects from the MR estimate.

NOSH, NO Simultaneous Heterogeneity. IV, instrumental variable. MR, Mendelian randomization. PACE, Population average causal effect.

6.4.2 Simulated bias of PACE under NOSH assumption one violation and rejection rate of IV-exposure variance test null hypothesis

Aim: To estimate PACE bias under interaction effect of IV-exposure and exposure-outcome by a common modifier (thus violating NOSH assumption one) and relate PACE bias to IV-exposure variance test null hypothesis rejection rate. The effect of exposure on outcome was additive linear (satisfying NOSH assumption two).

Data-generating mechanisms: Data were simulated for N=10,000 independent observations. For the ith observation, I simulated a SNP G_i in Hardy-Weinberg equilibrium (HWE) with a minor allele frequency (MAF) of 0.25 scaled to have mean of zero and unit variance. I also simulated a modifier U_i using the standard Normal distribution. A standard Normal exposure X_i was simulated with SNP main effect α_1 and modifier main effect α_2 each explaining 5% of the exposure variance and SNP-by-modifier GU_i interaction effect α_3 explaining $\epsilon\{0, 0.02, 0.04, 0.06, 0.08, 0.1\}$ of exposure variance. The outcome Y_i was simulated to have main effects of the exposure β_1 and modifier β_2 both explaining 5% variance, and 0-10% variance explained by the interaction effect β_3 and residual drawn from the standard Normal distribution. Note that the GU_i and XU_i varied by the same modifier U_i violating the first NOSH assumption[35]. The residual variance for X_i and Y_i was denoted with E_{1i} and E_{2i}, respectively.

$$X_i = \alpha_1 G_i + \alpha_2 U_i + \alpha_3 GU_i + E_{1i}$$

$$Y_i = \beta_1 X_i + \beta_2 U_i + \beta_3 XU_i + E_{2i}$$

Estimand: The IV PACE of X_i on Y_i.

Methods: The effect of G_i on $var(X)$ was tested using the least-absolute deviation regression based Brown-Forsythe test (LAD-BF)[2] (**Chapter 2.2.4**). The causal effect of X_i on Y_i was estimated using the Wald ratio[109].

Performance measures: LAD-BF null hypothesis rejection rate was defined as the percentage of repetitions with P < 0.05. PACE bias was defined as the difference between Wald ratio estimate and simulated causal effect i.e., $\widehat{\beta_1} - \beta_1$. Each configuration of parameters was evaluated using N=500 repetitions.

Open-source code: https://github.com/MRCIEU/variance-iv4-violation/blob/master/sim7.R

6.4.3 Simulated relative bias of PACE under NOSH assumption one violation and rejection rate of IV-exposure variance test null hypothesis

Aim: To estimate the relative bias of the PACE from violation of the first NOSH assumption and the IV-exposure variance test null hypothesis rejection rate. This simulation fixed the exposure-outcome interaction effect and varied the IV-exposure interaction effect. Both interactions were varied by the same modifier thus violating the first NOSH assumption[35]. The effect of exposure on outcome was additive linear satisfying NOSH assumption two[35]. This is distinct from **Chapter 6.4.2** in that relative bias is used instead of absolute bias. This decision was made to inform future studies by enabling lookups of expected PACE bias and IV-exposure variance test power given known IV-exposure association and anticipated interaction effect sizes. Relative bias was chosen so that all estimates are on the same scale.

Data-generating mechanisms: Data were simulated for N=500, N=1000, N=2000 and N=4000 independent observations for continuous outcomes and N=1000, N=2000, N=4000 and

N=6000 independent observations for binary outcomes. These sample sizes were chosen to show a range of relative biases across the IV-exposure and exposure-outcome variance explained. For the ith observation, I simulated a SNP G_i in HWE with a MAF of 0.25 scaled to have mean of zero and unit variance and standard Normal modifier U_i. The standard Normal exposure X_i was simulated with G_i main effect α_1 explaining 1-5% of the variance, U_i main effect α_2 explaining 20% variance, and SNP-by-modifier GU_i interaction effect α_3 0-2x the size of α_1. Exposures with sample size of N=500 and 1% SNP main effect variance explained were not presented as these were susceptible to weak instrument bias (F-statistic less than 10). For the continuous outcome, a standard Normal outcome Y_{1i} was simulated to have 20% variance explained by X_i and 20% variance explained by U_i and 10% variance explained by the interaction effect β_3 of exposure-by-modifier XU_i. For the binary outcome Y_{2i}, the intercept γ_0 was set to $\log\mathrm{OR}\ (1.1)$ and the main effects of X_i and U_i were set to $\log\mathrm{OR}\ (1.1)$ per 1 SD increase denoted by γ_1 and γ_2. The exposure-by-modifier XU_i interaction effect γ_3 was set to half the size of γ_1. Note that GU_i and XU_i interaction effects vary by U_i violating NOSH assumption one[35].

$$X_i = \alpha_1 G_i + \alpha_2 U_I + \alpha_3 G U_i + E_{1i}$$

$$Y_{1i} = \beta_1 X_i + \beta_2 U_i + \beta_3 X U_i + E_{2i}$$

$$logit(Y_{2i}) = \gamma_0 + \gamma_1 X_i + \gamma_2 U_i + \gamma_3 X U_i$$

Estimand: Relative PACE bias of X_i on Y_{1i} or Y_{2i}.

Methods: The effect of G_i on $var(X_i)$ was tested using LAD-BF[2] (**Chapter 2.2.4**). The causal effect of X_i on Y_{1i} and Y_{2i} was estimated using the Wald ratio[109].

Performance measures: LAD-BF test null rejection rate was defined as the percentage of tests with P < 0.05. PACE relative bias was defined as the simulated Wald ratio[109] divided by the true causal effect for binary and continuous outcomes, respectively. Each configuration of parameters was evaluated using N=500 replications.

Open-source code: https://github.com/MRCIEU/variance-iv4-violation/blob/master/sim9.R and https://github.com/MRCIEU/variance-iv4-violation/blob/master/sim11.R

6.4.4 Simulated PACE and IVW test efficiency under NOSH assumption one violation using subsets of instruments ranked by exposure-variance association

Aim: To estimate PACE and IVW[109] causal test efficiency under violation of the first NOSH assumption for only a subset of instruments. I explored the consequences of progressively removing instruments from the IVW analysis on PACE bias and IVW test efficiency by ranking instruments by their association with exposure variance. The effect of exposure on outcome was additive linear holding NOSH assumption two[35].

Data-generating mechanisms: Data were simulated for N=100,000 independent observations within each simulated dataset. This large sample size was chosen to obtain precise causal estimates with small numbers of instruments. For the ith observation, I simulated six uncorrelated SNPs G_j indexed by j, each in HWE and with a MAF of 0.25 scaled to have mean of zero and unit variance. I simulated a single modifier U_i drawn from the standard Normal distribution. The standard Normal exposure X_i was simulated to have $G_{i,j}$ $\alpha_{1,j}$ main effects drawn from the uniform distribution with sizes of $u(0.02, 0.06)$ which varied across simulation repetitions. These values were chosen to represent typical GWAS effect sizes. Half of the

instruments $GU_{i,j=1:3}$ had an interaction effect half the size of the main effect $\alpha_{3,j=1:3}$ on X_i while the remaining instruments $GU_{i,j=4:6}$ had no interaction effect $\alpha_{3,j=4:6} = 0$ on X_i. The main effect of U_i and X_i on Y_i explained 20% of the total variance. The interaction effect of XU_i explaining 10% of the variance Y_i. The residual error of X_i and Y_i were drawn from the standard Normal distribution denoted by E_{1_i} and E_{2_i}, respectively. Note that U_i modified the instrument-exposure and exposure-outcome relationships violating NOSH assumption one.

$$X_i = \sum_{j=1:6} \alpha_{1j} G_{ij} \quad + \alpha_2 U_i + \sum_{j=1:6} \alpha_{3j} GU_{ij} + E_{1i}$$

$$Y_i = \beta_1 X_i + \beta_2 U_i + \beta_3 XU_i + E_{2i}$$

Estimand: PACE of exposure-outcome relationship.

Methods: The PACE was estimated using IVW first using all instruments and then by progressively removing 5%, 10%, 25%, 50%, and 75% of instruments with strongest evidence for IV-exposure variance effect by the LAD-BF test p-value[2]. This was compared to the 'oracle' method which removed all the instruments with a known interaction effect on exposure (i.e., without using the IV-exposure variance test statistic). I anticipated the oracle method to differ only when the IV-exposure interaction effect is incorrectly identified using variance analysis.

Performance measures: PACE for the exposure-outcome effect. IVW test efficiency, this was estimated using the mean of the IVW standard error between replicates. Each configuration of parameters was evaluated using N=500 replications.

Open-source code: https://github.com/MRCIEU/variance-iv4-violation/blob/master/sim12.R

6.4.5 Effect of serum metabolites on disease outcomes

Genetic instruments for mean and variance of randomly sampled urate, glucose and low-density lipoprotein (LDL) cholesterol were extracted from the MRC-IEU OpenGWAS platform[18] estimated in UK Biobank (**Table 3.4.2.1**). These were applied in a two-sample MR framework[109] to estimate the causal effect of these traits on gout, type 2 diabetes mellitus (T2DM) and coronary heart disease (CHD), respectively using outcome datasets from large consortia (**Table 3.4.2.1**) with non-overlapping samples. The main analysis used all available instruments. Sensitivity analyses were performed by removing 5%, 10%, 25%, 50% and 75% instruments with the strongest IV-exposure variance effects estimated using LAD-BF as potential evidence for NOSH assumption one violation. IV-exposure variance associations were estimated in UK Biobank (**Chapter 5**). To ensure any differences in causal estimate between sensitivity analyses was not produced by selecting for weaker instruments (since mean and variance may be correlated; **Chapter 1.5.4; Chapter 1.3.6**), I estimated the mean instrument strength (F-statistic) for each subset of instruments calculated independently and compared this with the complete set of instruments. This was accomplished by comparing the mean IV-exposure F-statistic for each subset with the mean of all instruments using the Mann-Whitney U test to determine if IVW estimates were affected by differing instrument strength.

6.4.6 Estimation of SNP variance explained and F-statistic from GWAS summary data

The F-statistic was used as a measure of IV strength and was estimated from GWAS summary statistics (**Equation 6.4.6.1**) as previously described[159].

Equation 6.4.6.1. Estimation of the F-statistic from R^2

$$F = R^2 \times (N - 1 - k)/((1 - R^2) \times k)$$

Where N is the sample size, k is the number of SNPs included in the model ($k = 1$ in this analysis), and R^2 is the variance explained by the SNP which may be estimated using **Equation 6.4.6.2**[159].

Equation 6.4.6.2. Estimation of R^2 from GWAS summary statistics

$$R^2 = 2\beta^2 \times (MAF) \times (1 - MAF)/(2\beta^2 \times (MAF) \times (1 - MAF) + \big(se(\beta)\big)^2 \times 2N \times MAF \times (1 - MAF))$$

Where β is the average effect of the SNP on the trait, MAF is the minor allele frequency and se is the standard error of β.

6.4.7 Software

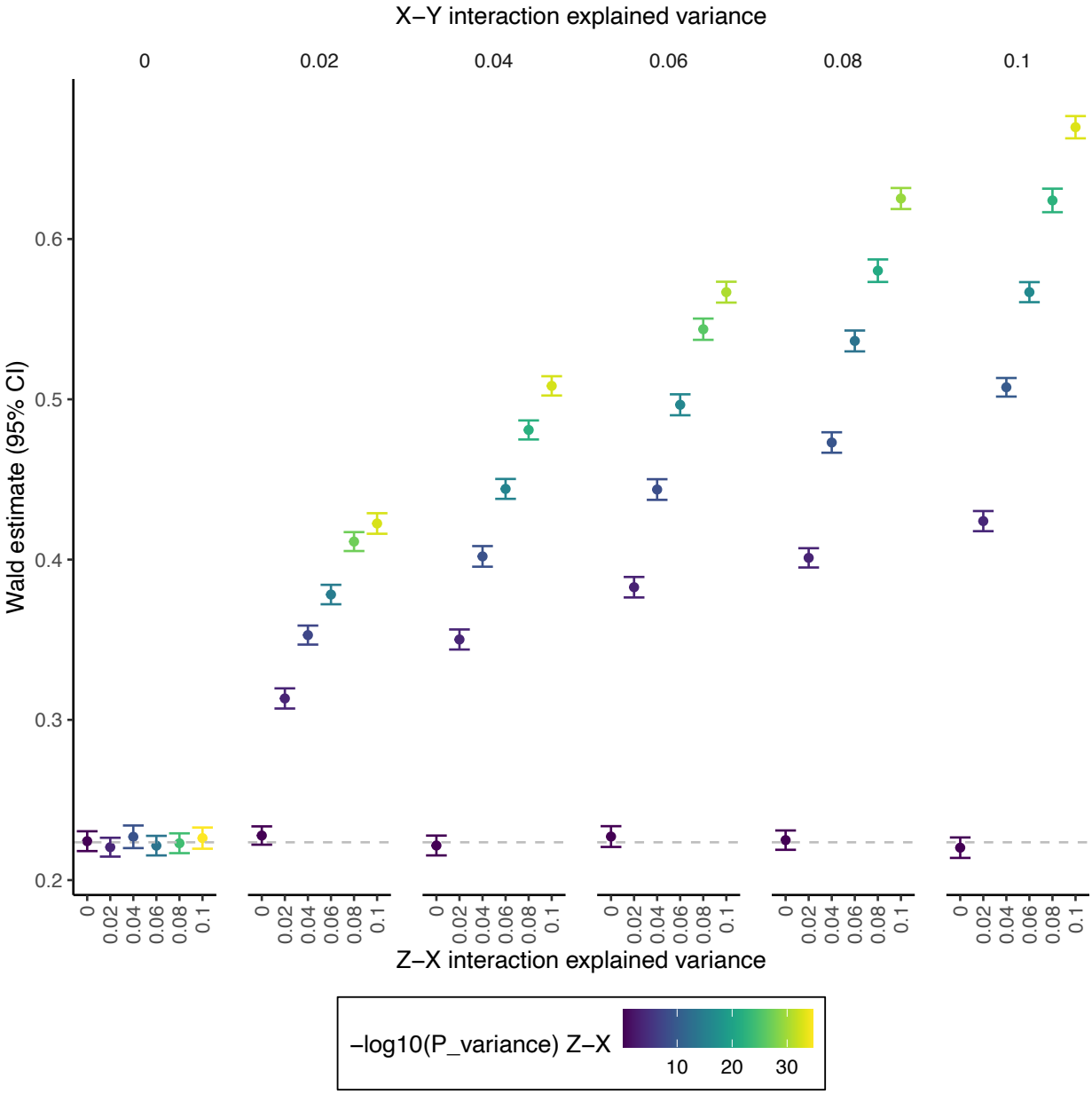
All MR estimates were produced using the TwoSampleMR R-package[109] (v0.5.5). Variant association with trait variance was estimated using the varGWAS R-package developed in **Chapter 4** (v1.0.0). All analyses and simulation studies were conducted using R (v3.6.0).

6.5 Results

6.5.1 Simulated evidence for NOSH assumption one violation using IV-exposure variance test statistics

The PACE of continuous exposure on a continuous outcome was unbiased when either IV-exposure or exposure-outcome interaction effects were null (**Figure 6.5.1.1**). Increasing the IV-exposure and exposure-outcome interaction effect size was associated with increased bias of estimated PACE when both relationships were modified by a single variable. This is consistent with violation of the first NOSH assumption[35]. Increasing IV-exposure interaction effect size was also associated with increased strength of IV-exposure variance association (**Figure 6.5.1.1**).

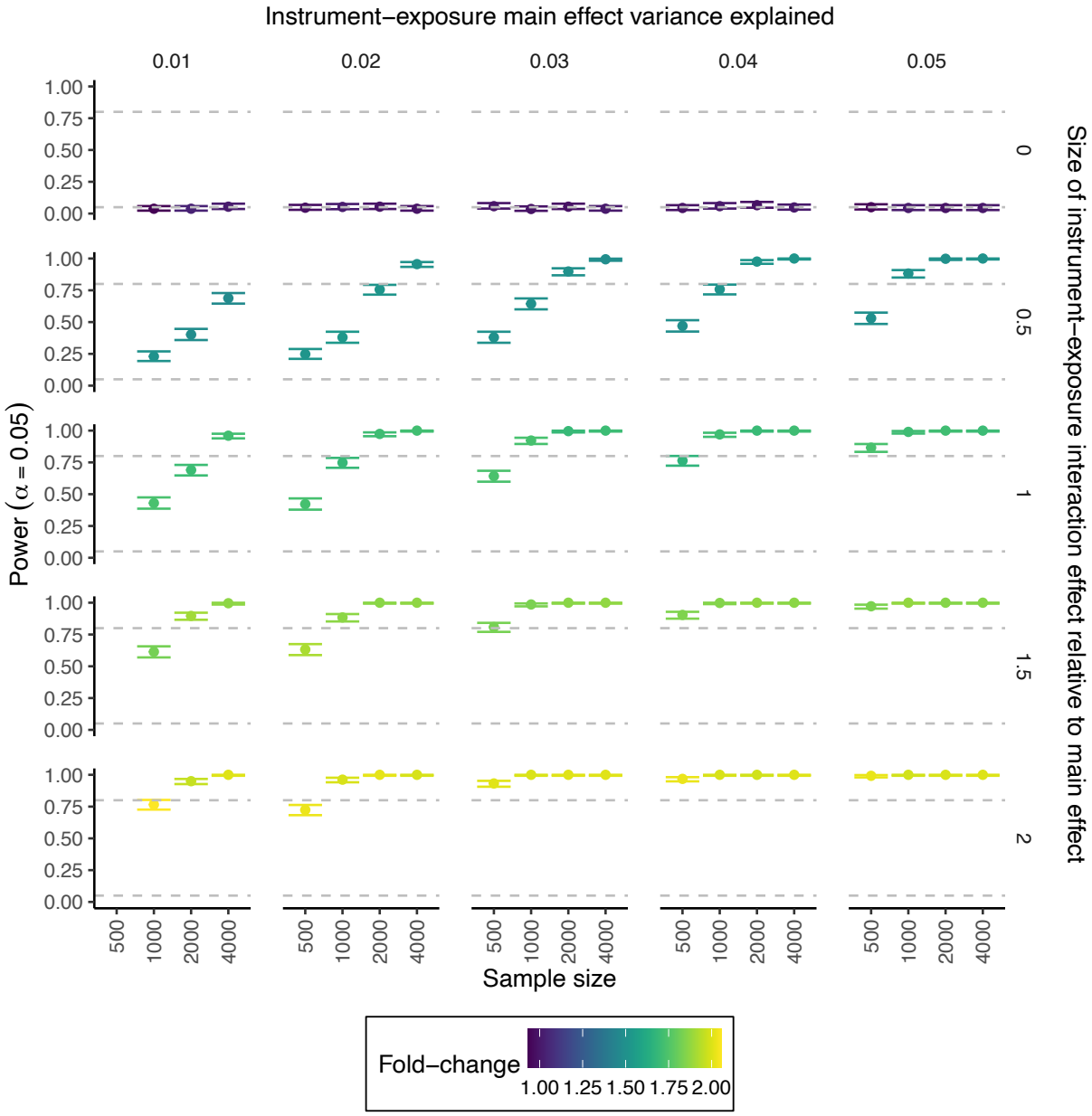
Figure 6.5.1.1. PACE bias under homogeneity assumption violation



Effect of NOSH assumption on violation on PACE bias. Dashed line is expected PACE effect. Interactions of Z-X and X-Y are modified by the same binary variable (violating NOSH assumption one) but exposure-outcome effect is additive linear (NOSH assumption two holds). Z, instrumental variable. X, exposure. Y, outcome. CI, confidence interval.

Next, I explored the relative bias of PACE and related the magnitude of this bias to evidence for IV-exposure variance effects. The strength of IV-exposure variance association was used to indicate potential NOSH assumption one violation. Fixing the continuous exposure-outcome variance explained to 20%, under IV-exposure main and interaction effects of 2% and 1% variance explained respectively, the PACE was on average 1.50x the size of the expected effect and the IV-exposure variance test rejected the null in 96% of repetitions (95% CI 93%, 97%) with a sample size of N=4000 (**Figure 6.5.1.2**). Fixing the continuous exposure binary outcome effect of 1.1 OR per 1 SD, the magnitude of PACE bias was on average 1.28x times the size of the expected effect and the IV-exposure variance test null was rejected in 96% of repetitions (95% CI 94%, 98%) given a sample size of N=4000 (**Figure 6.5.1.3**).

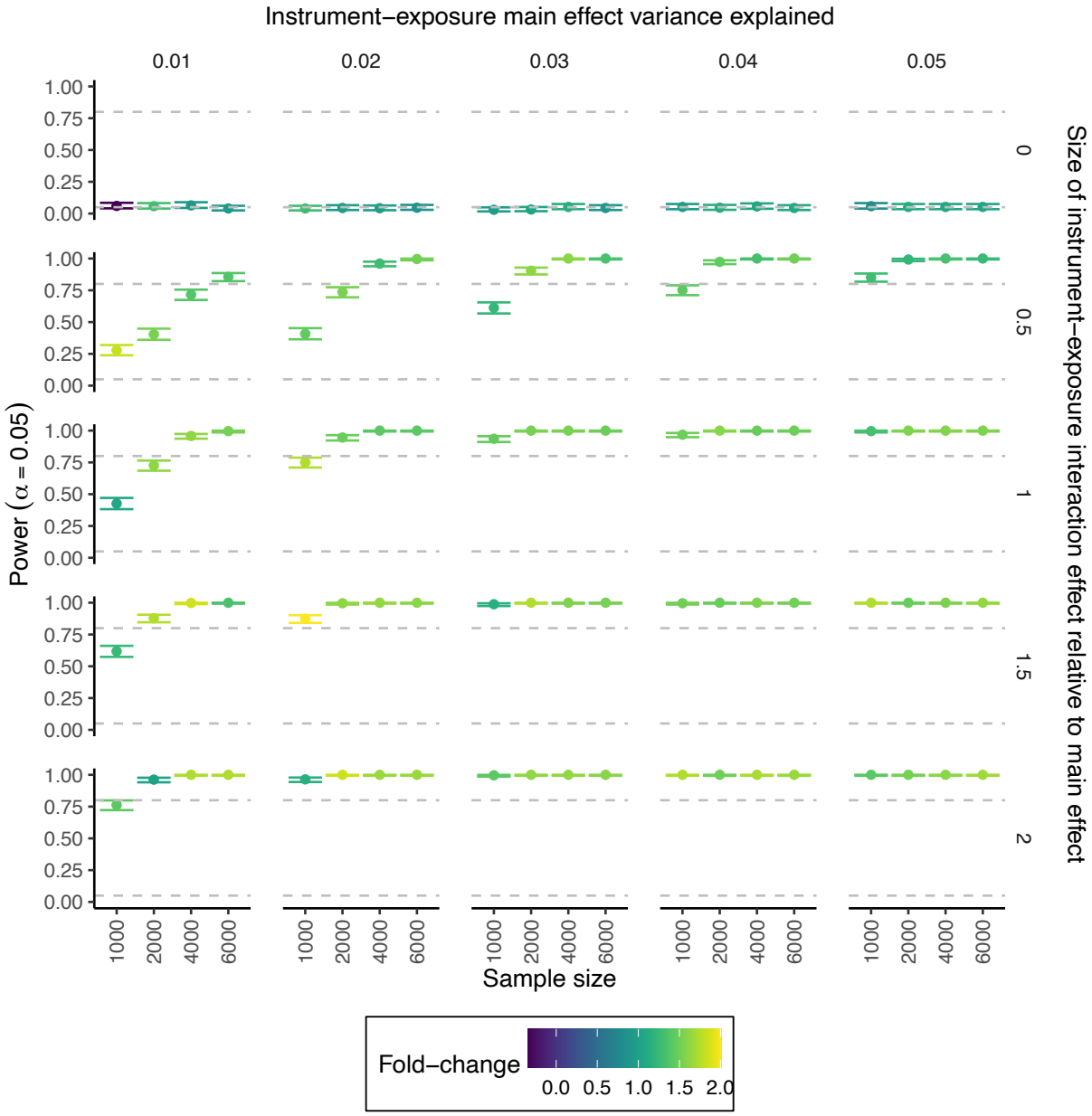
Figure 6.5.1.2. Simulated power to detect IV-exposure variance effect and relative PACE bias under NOSH assumption one violation



Power, proportion of repetitions where the IV-exposure variance test P < 0.05. Fold-change, relative bias of PACE compared with simulated causal effect. IV-exposure main and interaction effects were varied. Exposure-outcome main and interaction effects were fixed to 20% and 10% variance explained, respectively. Simulations with N=500 with 1% variance explained by the IV-

exposure relationship produced an F-statistic less than 10 and were not shown. Dashed lines represent 80% power.

Figure 6.5.1.3. Power to detect IV-exposure variance effect and relative PACE bias of binary outcomes from NOSH assumption one violation

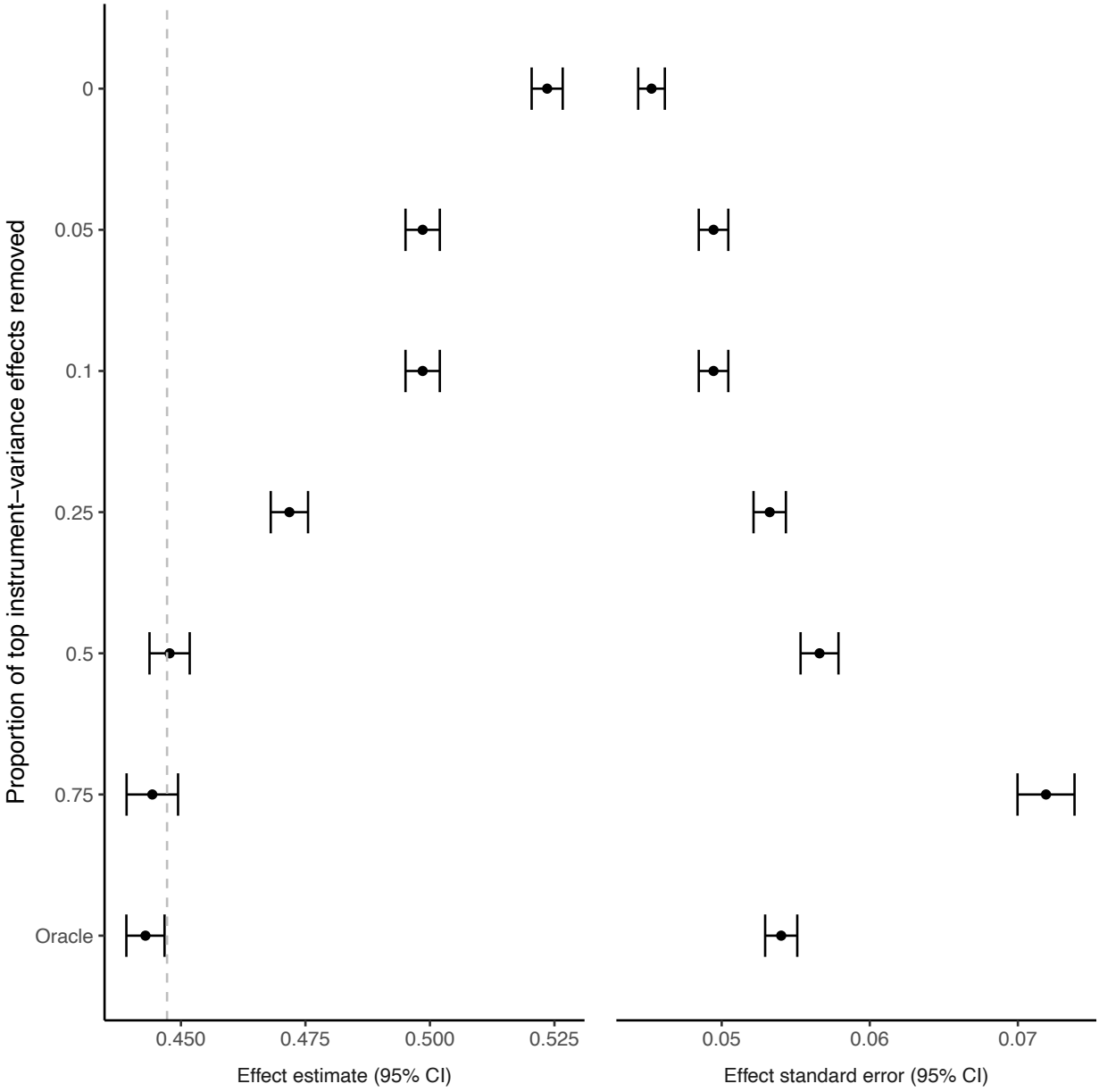


Power, proportion of repetitions where the IV-exposure variance test P < 0.05. Fold-change, relative bias of PACE compared with simulated causal effect on the log-odds scale. IV-exposure main and interaction effects were varied. Exposure-outcome main and interaction effects were fixed to 1.1 OR. Dashed lines represent 80% power.

6.5.2 Simulated effects on PACE bias and statistical efficiency of removing instruments by strength of association with exposure variance

Under simulation, I explored the consequences on PACE bias and statistical efficiency of IVW by removing instruments from analysis which were associated with exposure variance (**Figure 6.5.2.1**). Half of the instruments were simulated to have an interaction effect on the exposure and the exposure was simulated to have an interaction effect on the outcome. All interaction effects had the same single modifier, violating NOSH assumption one. Instruments were progressively removed from the IVW analysis using IV-exposure variance test strength by P-value estimated with LAD-BF (**Chapter 3 and Chapter 4**). IVW estimates were less biased when instruments with exposure variance effects were removed but this also led to larger IVW standard errors. For example, using all the instruments including 50% simulated with an interaction effect on the exposure, the PACE estimate was 0.53 SD (95% CI of 0.52-0.54) per 1 SD exposure in contrast to the simulated effect of 0.447 SD. Removing the top 50% of instrument-exposure variance effects produced a PACE estimate of 0.45 SD (95% CI 0.44-0.46) in line with the simulated effect. This estimate was also consistent with the oracle method which removed instruments simulated to have non-zero exposure interaction effect (0.45 SD [95% CI 0.44-0.47]). The oracle results may differ if the IV-exposure variance test incorrectly removed/retained SNPs from the model. However, fewer instruments led to reduced statistical efficiency of IVW (**Figure 6.5.2.1**). When all instruments were employed the average standard error for the causal effect estimate was 0.05 (95% CI 0.04-0.05) but this increased to 0.07 (95% CI 0.07-0.07) after removing 75% of instruments with variance effects.

Figure 6.5.2.1. Simulated effect of removing instruments with exposure variance effects on PACE bias and statistical efficiency



IVW effect of simulated exposure on outcome using subsets of instruments by instrument-exposure variance effect strength and mean of IVW effect standard error. The effect estimate subplot provides the mean causal estimate for each analysis. The effect standard error subplot

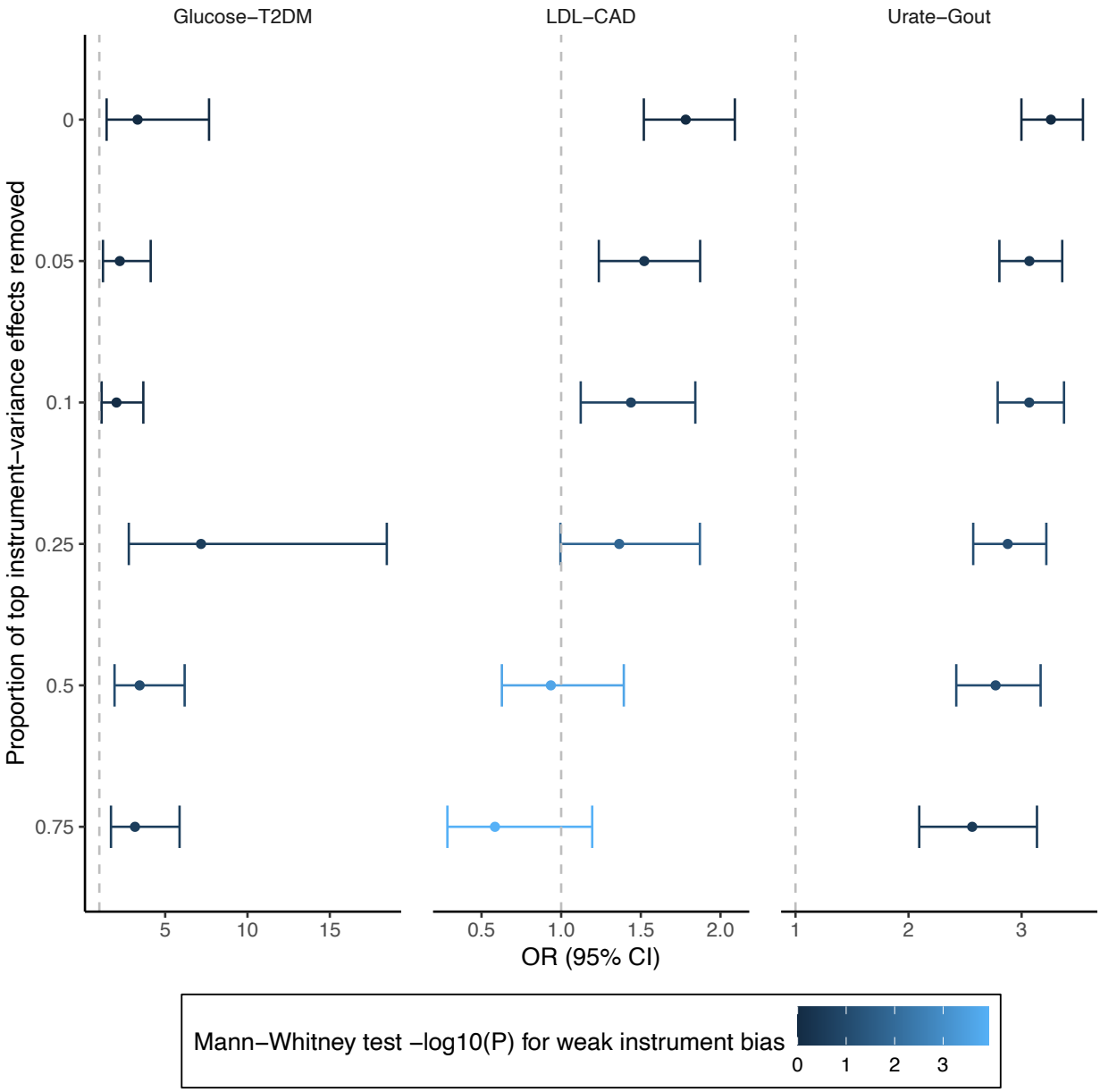
shows the mean standard error for each analysis. SD, standard deviation. CI, confidence interval. SE, mean IVW standard error of 500 replicates.

6.5.3 IV-exposure variance testing to detect potential NOSH assumption one and attenuate IV estimate bias on disease outcomes

IVW effect estimates of serum metabolites on disease outcomes were produced using instruments stratified by instrument-exposure variance effect strength (**Figure 6.5.3.1**). Starting with the complete set of instruments, the per SD exposure causal effect estimates were: 3.33 OR (95% CI 1.44-7.66) for glucose effect on T2DM, 1.78 OR (95% CI 1.52-2.09) LDL on CHD and 3.26 OR (95% CI 3.00-3.54) urate on gout. As instruments were removed from the analysis, causal estimates attenuated towards the null. This was most extreme for LDL-CHD which reversed sign giving evidence for a protective effect of 0.58 OR (95% CI 0.29-1.20) after removing the top 75% of instruments ranked by instrument-exposure variance strength. However, this group of instruments was also weaker (F-statistic mean of 40.56 [95% CI 35.50-46.73]) compared with the full set (F-statistic mean of 71.62 [95% CI 59.69-88.91]) suggesting weaker instruments may be responsible for biasing these effects[44]. Where there was little evidence of weaker instruments, I detected little attenuation of point estimates compared with the full set of instruments and overlapping confidence intervals suggested no strong difference. For example, the effect of LDL-CHD excluding the top 10% of instrument-variance effects gave a causal estimate of 1.44 OR (95% CI 1.22-1.84) and removing the top 75% of instruments led to estimates of 3.17 OR (95% CI 1.71-5.87) and 2.56 OR (95% CI 2.10-3.14) for glucose on T2DM and urate on gout, respectively. As these estimates were consistent with the full set of instruments, there was no strong evidence for violation of NOSH suggesting the estimand is targeting PACE. However, it is also possible that instruments with the weakest evidence for

exposure variance effect have an interaction effect on the exposure, but the LAD-BF test power was too low to detect an effect.

Figure 6.5.3.1. IVW sensitivity analysis removing instruments with exposure variance effects



Inverse-variance weighted effect of serum metabolite concentration on binary outcomes. OR, odds ratio. CI, confidence interval. LDL, low-density lipoprotein. CAD, coronary artery disease. T2DM, type II diabetes mellitus. Mann-Whitney U test is for the comparison of instrument-exposure F-statistic mean effect for all instruments vs the subset as a measure for conditioning on instrument strength.

6.6 Discussion

Under NOSH the IV estimand targets PACE[35], but the two NOSH assumptions cannot be completely assessed using the observed data. Here I demonstrated via simulation the potential in testing of IV-exposure variance effects for continuous exposures as an empirical approach to evaluate NOSH assumption one. Testing of IV-exposure variance effects cannot prove that NOSH assumption one holds as power may be too low to detect a variance effect of IV-exposure where a true interaction exists. Further, even if there is an interaction effect of IV-exposure then NOSH assumption one is only violated if the exposure-outcome effect is modified by this same variable[35]. Secondly, I show how this evidence could be used to mitigate bias from PACE by removing instruments with strong exposure variance effects. This methodology was applied to GWAS summary statistics generated in UK Biobank and non-overlapping large consortia.

Simulations showed that the approach was well powered to detect PACE bias using IV-exposure variance effects when using the large sample sizes which are now readily available from biobanks. However, IV-exposure variance association cannot specifically identify NOSH violation. The first NOSH assumption requires effect modification of the exposure-outcome and IV-exposure relationships to be independent[35] but this scenario is not specifically evaluated using measures of IV effect on exposure variance. Lack of an IV-exposure variance effect could suggest that the IV estimand targets PACE subject to sufficient power to detect an effect. Identification of IV-exposure variance effects could enable follow up studies to identify the precise exposure-modifier interaction effect as has been shown in **Chapter 5**. This could be useful to consider if this variable also modifies the exposure-outcome effect which would then

imply NOSH assumption one is violated. Conversely, NOSH violation may occur without IV-exposure variance association through non-linearity of the exposure-outcome relationship (NOSH assumption two)[35].

This approach of testing for violation of homogeneity is similar in principle with Brookhart *et al*[34] who suggest testing for IV-exposure interaction effects. However, Brookhart *et al* require effect modifiers to be hypothesised and measured while the approach outlined here is based on IV-exposure variance association and does not.

I explored the utility of eliminating IVs based on their association with exposure variance to determine if the IV estimate returns to PACE through simulation studies. I observed reduced departure from PACE but also wider confidence intervals which is anticipated because fewer instruments were included in the causal model. Nevertheless, this approach could be useful as a sensitivity analysis to determine if the main analysis (i.e., using all instruments) produces an estimate strongly different from a subset of instruments with the least exposure variance association. While I used the strength of association between IV and exposure variance, future studies could explore the magnitude of effect (i.e., using the variance effect size) but with LAD-BF two coefficients are produced for each allele copy and are difficult to rank.

This approach was applied to evaluate the effects of serum LDL, glucose, and urate on CHD, T2DM, and gout, respectively. IVW estimates did not robustly differ after removing IV-exposure variance for glucose-T2DM and urate-gout suggesting no strong evidence for violation of the first NOSH assumption. Meanwhile, the effect directionally of LDL-CHD was reversed when the top IV-exposure variance effects were removed. Two possibilities may explain this finding.

First, instruments may be acting via distinct causal pathways, for example a previous study found a cluster of instruments for BMI that had protective effects on cardiovascular disease[160] whereas the remaining instruments were associated with adverse effects. Furthermore, evidence for IV-exposure variance association could arise under horizontal pleiotropy[161], for example if the instrument is acting on the exposure through several pathways and some of these are influenced by effect modification. This is in contrast with, for example, an instrument within the *cis* region of a protein coding gene which is less likely to be affected by horizontal pleiotropy[103].

Second, weak instrument bias may be introduced by conditioning on weaker IV-exposure variance effects (since both mean and variance effects can be correlated[69,92]). While the mean F-statistic was above the rule of thumb of ten[44] these estimates could be inflated due to chance which is known to introduce bias to causal estimates[44]. In the two-sample MR framework weak instrument bias causes estimates to attenuate towards the null in contrast with two-stage least squares approach which introduces bias towards the observational association[162]. One way to avoid this is to use two-sample MR with second-order weights which incorporate imprecision in the IV-exposure relationship into the causal estimate[163].

Testing for IV-exposure variance effects could be applied to future MR studies as a sensitivity analysis to determine if effects deviate from PACE. However, this strategy would require larger sample sizes than are available today to have sufficient power to detect a variance effect of the IV on exposure. This approach could be developed further using meta-regression[164] techniques to add less weight to IVs that have strong variance effects rather than simply removing these IVs and may preserve statistical efficiency.

6.7 Limitations

However, this work also has some limitations. First, identification of a variance effect of the IV on exposure has low power and would require larger sample size than are currently available before routine implementation could be considered. Secondly, non-normality of the exposure may reduce power and increase type I error rate of the IV-exposure variance association[68] (**Chapter 4**). Third, removing instruments from IV analyses (e.g., by the association of IV with exposure mean or variance) may bias the causal estimate standard errors leading to type I error rate inflation[161]. This could also lead to conditioning on instruments that exhibit horizontal pleiotropy introducing bias into the estimate[161].

6.8 Conclusions

Through this work I evaluated the strength of the IV-exposure variance association as an empirical approach to evaluate IV-exposure homogeneity assumptions which may be used in falsification studies to determine if the causal estimand departs from PACE, and for sensitivity analyses to provide evidence against PACE as the target estimand. I applied these methods to evaluate the effects of LDL-CHD, urate-gout and glucose-T2DM but found no strong evidence for departure from PACE. This approach could be applied to future IV studies when sample sizes are much larger to improve the interpretability of causal estimates.

Chapter 7: Developing a robust and efficient file format for sharing GWAS summary statistics

7.1 Overview

Genome-wide association study summary statistics are an important resource for a variety of secondary research applications (**Chapter 1.2.5**). Yet despite their widespread utility, no common storage format has been widely adopted hindering tool development and data sharing (**Chapter 1.2.6**). Existing tabular formats lack approaches for robustly storing variants and essential metadata increasing the possibility of errors in data interpretation[22]. Additionally, data are typically provided unindexed requiring the file be read line-by-line to extract specific SNP-trait associations which is slow and computationally inefficient[24]. To address these issues, this chapter proposes storing GWAS summary statistics using the variant call format[165] (VCF) known as GWAS-VCF and introduces open-source tools for producing and reading these data. Simulations of query performance using Tabix[24] and standard UNIX tools suggested VCF is 8.6-45.5x faster to extract variant(s) by genomic position. I converted variance GWAS data produced in **Chapter 5** using LAD-BF developed in **Chapter 4** to GWAS-VCF for sharing, downstream analysis, and rapid querying. This format has also been used by colleagues at the MRC-IEU to provide open access to >10,000 complete GWAS summary statistics as part of the OpenGWAS platform[18] (gwas.mrcieu.ac.uk).

7.2 Contribution statement

Work in **Chapter 7** was published in Genome Biology (Lyon *et al*, 2021[1]), a paper which I drafted and was edited by Dr Shea Andrews (Mount Sinai), Dr Ben Elsworth, Professor Tom Gaunt, Dr Gibran Hemani and Professor Edoardo Marcora (Mount Sinai). The gwas2vcf and

pygwasvcf software I developed in present here was also included in a manuscript available as a preprint on BioRxiv (Elsworth *et al*, 2020[18]). I produced **Figure 7.5.4.1** which was included in Elsworth *et al*.

I and researchers at the School of Medicine at Mount Sinai separately proposed a VCF file specification for storing and distributing GWAS summary statistics. Dr Hemani (MRC-IEU), Professor Tom Gaunt (MRC-IEU) and Professor Marcora (Mount Sinai) agreed to collaborate to agree a single consistent format (as VCF is flexible in how data are stored).

All co-authors had input on the user requirements and final file format. I developed the first version of the gwas2vcf Python software to harmonise GWAS summary statistics and automate file conversion processes. This codebase was subsequently developed further by Dr Gibran Hemani and members of the community. I performed the simulations and converted variance GWAS summary statistics produced in **Chapter 5** into GWAS-VCF available from the MRC-IEU OpenGWAS database[18].

7.3 Introduction

7.3.1 Background

In **Chapter 5** I estimated the variance effect of 290M loci on biomarker concentration in UK Biobank and used these data to perform follow up interaction analyses. In **Chapter 6** I also applied these variance GWAS summary statistics to evaluate MR homogeneity assumptions. Variance GWAS (vGWAS) may have other uses for secondary research application (**Chapter 5** and **Chapter 6**). Therefore, it is vital that these data are shared to support re-analysis and development of new analysis methodology (**Chapter 1.2.5**). However, the utility of GWAS summary statistics is hampered by the absence of a universally adopted storage format and

associated tools[166]. Historic lack of a common standard has resulted in GWAS analysis tools outputting summary statistics in different tabular formats (e.g. plink[126], BOLT-LMM[125], and METAL[167]). The VCF[165] is easily adapted for storing a range of genomic data but there is flexibility in how this information is stored which can impact on file size and read/write performance. The VCF is also capable of storing GWAS data from multiple traits (samples)[165] in a single file which may be advantageous for distributing summary statistics on a collection of closely related traits such as biomarkers[168], gene expression[141], and protein[169] concentration.

7.3.2 Aims

The aims of this chapter are to develop a set of requirements for a suitable universal format, adapt the variant call format (VCF)[165] for storing GWAS summary statistics, demonstrate how the VCF meets these requirements, showcase the capabilities of this medium, and introduce tools and resources for working with this format. Finally, I prepare variance GWAS summary statistics from **Chapter 5** in GWAS-VCF.

7.4 Materials and methods

7.4.1 File indexing

Two file indexing approaches were used to support a range of different queries. First, the VCF file was indexed using chromosome position by Tabix[24] which is a karyotypically sorted list of chromosome intervals including their offset file position. Second, a custom SQLite (www.sqlite.org) database was created for each GWAS-VCF that contained a unique list of dbSNP rsIDs[170] and their corresponding chromosomal position so that records could be retrieved using the Tabix[24] index. The dbSNP (SQLite) index was adapted from a previous project (rsidx)[171].

7.4.2 Query performance simulation

Aim: To compare the query runtime performance of tab-separated value (TSV) and GWAS-VCF file formats to extract GWAS results under a range of conditions.

Data-generating mechanisms: Densely imputed GWAS summary statistics (N=13,791,467 variants) of BMI using data from UK Biobank were obtained from Neale *et al*[123]. From this data, two sets of GWAS-VCF files were produced containing either one or five trait(s) and with varying number of SNPs by combining randomly subsampled summary statistics with either $\epsilon\{0.5, 2.5, 10\}$ million rows. TSV files were prepared from the GWAS-VCF to replicate a typical storage medium currently used for distributing summary statistics[22].

Estimand: Result retrieval performance

Methods: GWAS summary statistics were mapped to GWAS-VCF using gwas2vcf v1.1.1 (**Chapter 7.5.4**) and processed using bcftools v1.10[172] to remove multiallelic variants or records with missing dbSNP[170] identifiers. Query runtime performance was compared between Tabix v1.10.2[24] (using file index) with bcftools[172] and rsidx[171] and standard Ubuntu v18.04 UNIX commands (which read line-by-line) using AWK and grep for the following queries: single variant selection using dbSNP identifier[170] or chromosome position, multi-variant selection by association P value (thresholds: $P < 5 \times 10^{-8}$, $P < 0.2$, $P < 0.4$, $P < 0.6$, $P < 0.8$) or 1Mb genomic interval.

Performance measures: Queries were performed with 100 repetitions using BGZIP[172] GWAS-VCF or unindexed TSV with and without GZIP compression on an Ubuntu v18.04 server with Intel Xeon® 2.0 Ghz processor. All comparisons were performed using single thread

operations and therefore differences in runtime performance were due to query software and/or file index usage.

Open-source code: https://github.com/MRCIEU/gwas-vcf-performance

7.5 Results

7.5.1 Requirements

Requirements for a universal GWAS summary statistics format specification were developed through collaboration between the MRC-IEU and Ronald M. Loeb Center for Alzheimer's Disease (**Table 7.5.1.1**). These features place emphasis on consistency and robustness, capacity for metadata to provide a full audit trail, efficient querying, and file storage, ensuring data integrity, interoperability with existing open-source tools and across multiple datasets to support data sharing and integration.

Table 7.5.1.1. Requirements for a summary statistics storage format and solutions offered by the VCF

Requirement	Solution using the variant call format
Human readable	Read with any text viewer
Easy to parse	Mature open-source parsing libraries are available (HTSLIB[173] and HTSJDK[173]) and implemented in most modern programming languages, for example: VariantAnnotation[174] R-package is available from Bioconductor[175,176] and python package pysam[173,177]. Bcftools[172], GATK[178], bedtools[138] and others provides user-friendly functionality from the command line.
Unambiguous interpretation of the data	Data field descriptions, value types and number of values are required and defined in the file header[165]. File validity is enforced during each read/write[173].
Unambiguous representation of bi-allelic, multi-allelic and insertion-deletion variants	Every variant substitution is represented by reference and alternative allele haplotypes defining the exact base change on the forward strand[165]. The reference allele is required to match genome sequences defined in the file header[165]. The alternative allele is always the effect allele allowing consistency between studies for ease of comparison[165].
Genomic information can be validated	The file header contains information about reference genome assembly and chromosomes[165]. Reference alleles must match the sequence in the referenced genome build[165] (in FASTA format). GATK[178] ValidateVariants can be used to verify file format validity and compare reference allele information against the corresponding genome reference sequence.
Flexibility on which GWAS fields are recorded and enforcement of essential fields	All fields are defined in the file header and can be set optional or required as desired[165]. The specification contains essential fields and their reserved names[165].
Capacity to store metadata about the study and trait(s)	The file header contains information about the source and date of summary statistics, study IDs (e.g., PMID/DOI of publication describing the study, or accession number and repository of individual-level data), description of the trait(s) studied (e.g., type, association test used, and measurement unit) as well as the source and version of trait IDs (e.g., MRC-IEU OpenGWAS database[18], Experimental Factor Ontology[179], Human Phenotyping Ontology[180], Medical Subject Headings[181] IDs for clinical and other traits, Ensembl[139] Gene IDs for eQTL datasets, or any other ontology to describe the data).
Allows multiple traits to be stored together	The SAMPLE column[165] was chosen to store variant-trait association data to allow for storage of multiple traits in a single VCF file, or as individual files if desired.
Rapid querying by variant identifier, genomic position interval or GWAS	The file is sorted karyotypically and indexed by chromosome position using Tabix[24] to enable fast queries by genomic position. Secondary indexing on dbSNP[170] identifier is also provided using rsidx[171]. Refer to performance comparisons of indexed VCF files and standard UNIX tools.

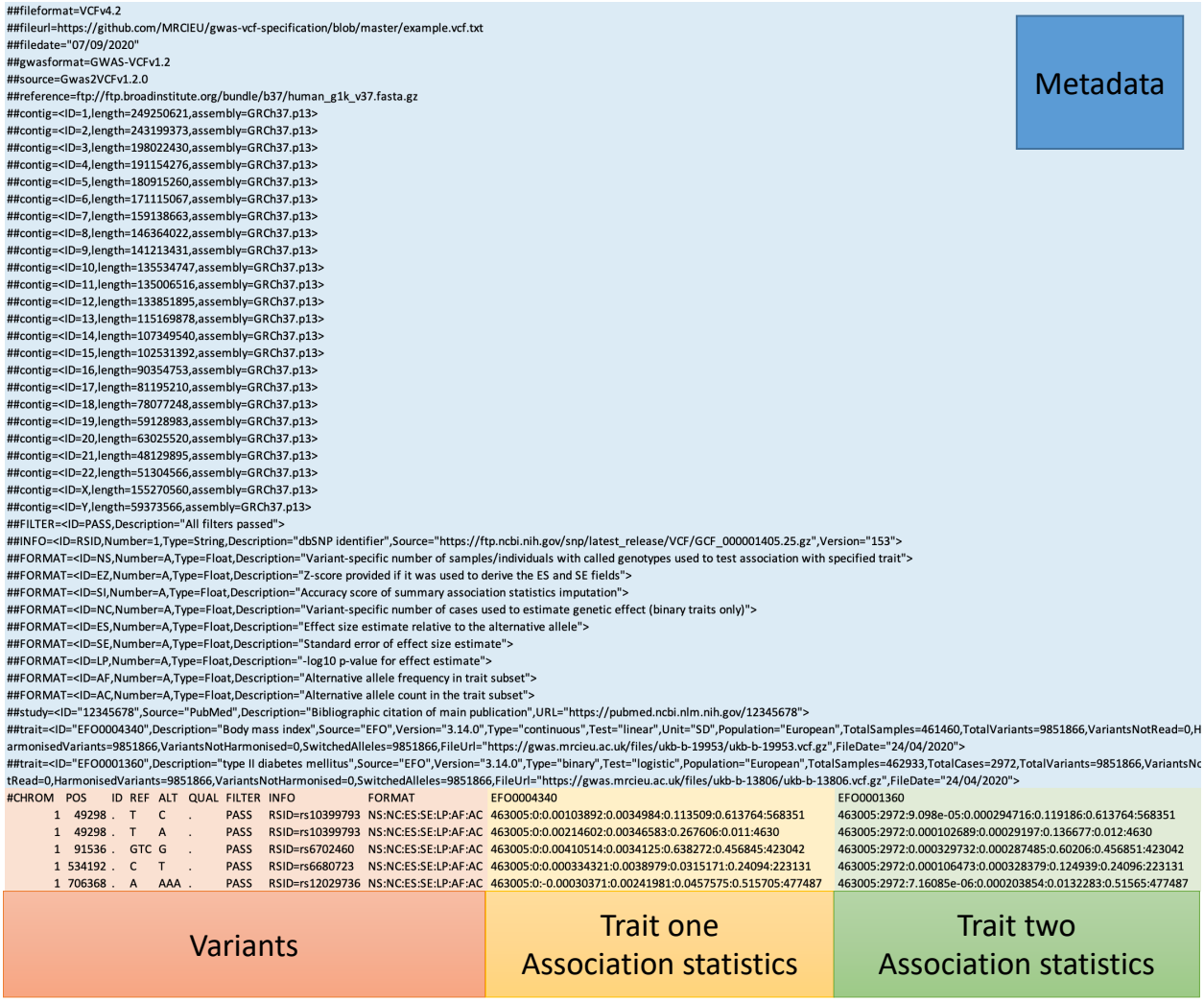
summary statistics value (range or exact value)	
File compression	VCF files may be compressed with block GZIP (BGZIP)[172] or converted to a binary call file which is a binary VCF companion format[172].
Readable by existing open-source tools	A large number of tools support VCF files including: GATK[178], Picard[182], bcftools[172], bedtools[138], vcftools[165] and plink[126]. Bcftools[172] can also provide a tabular extract for use with non-compatible tools.
Amenable to cloud-based streaming and database storage	Genomic intervals may be extracted over a network using a range-request which extracts file segments without transferring the whole file[165]. This enables rapid streaming of queries over the internet. For high-throughput and distributed storage and querying, VCF[165] files can be easily imported into GenomicsDB[183].

GWAS, genome-wide association study. dbSNP, database of single-nucleotide polymorphisms. HTSLIB, high-throughput sequencing data library. HTSJDK, high-throughput sequencing data Java development kit. GATK, genome-analysis toolkit. dbSNP, single nucleotide polymorphism database. eQTL, expression quantitative trait loci.

7.5.2 File format

The VCF is organised into three components[165]: a flexible file header containing metadata (lines beginning with '#'), and a file body containing variant- (one locus per row with one or more alternative alleles/variants) and sample-level information (one sample per column). The VCF was adapted to include GWAS-specific metadata and utilise the sample column (one per GWAS trait) to store variant-trait association data (**Figure 7.5.2.1**; **Table 7.5.2.1**).

Figure 7.5.2.1. VCF format adapted to store GWAS summary statistics (GWAS-VCF)




```
##fileformat=VCFv4.2
##fileurl=https://github.com/MRCIEU/gwas-vcf-specification/blob/master/example.vcf.txt
##filedate="07/09/2020"
##gwasformat=GWAS-VCFv1.2
##source=Gwas2VCFv1.2.0
##reference=ftp://ftp.broadinstitute.org/bundle/b37/human_g1k_v37.fasta.gz
##contig=<ID=1,length=249250621,assembly=GRCh37.p13>
##contig=<ID=2,length=243199373,assembly=GRCh37.p13>
##contig=<ID=3,length=198022430,assembly=GRCh37.p13>
##contig=<ID=4,length=191154276,assembly=GRCh37.p13>
##contig=<ID=5,length=180915260,assembly=GRCh37.p13>
##contig=<ID=6,length=171115067,assembly=GRCh37.p13>
##contig=<ID=7,length=159138663,assembly=GRCh37.p13>
##contig=<ID=8,length=146364022,assembly=GRCh37.p13>
##contig=<ID=9,length=141213431,assembly=GRCh37.p13>
##contig=<ID=10,length=135534747,assembly=GRCh37.p13>
##contig=<ID=11,length=135006516,assembly=GRCh37.p13>
##contig=<ID=12,length=133851895,assembly=GRCh37.p13>
##contig=<ID=13,length=115169878,assembly=GRCh37.p13>
##contig=<ID=14,length=107349540,assembly=GRCh37.p13>
##contig=<ID=15,length=102531392,assembly=GRCh37.p13>
##contig=<ID=16,length=90354753,assembly=GRCh37.p13>
##contig=<ID=17,length=81195210,assembly=GRCh37.p13>
##contig=<ID=18,length=78077248,assembly=GRCh37.p13>
##contig=<ID=19,length=59128983,assembly=GRCh37.p13>
##contig=<ID=20,length=63025520,assembly=GRCh37.p13>
##contig=<ID=21,length=48129895,assembly=GRCh37.p13>
##contig=<ID=22,length=51304566,assembly=GRCh37.p13>
##contig=<ID=X,length=155270560,assembly=GRCh37.p13>
##contig=<ID=Y,length=59373566,assembly=GRCh37.p13>
##FILTER=<ID=PASS,Description="All filters passed">
##INFO=<ID=RSID,Number=1,Type=String,Description="dbSNP identifier",Source="https://ftp.ncbi.nih.gov/snp/latest_release/VCF/GCF_000001405.25.gz",Version="153">
##FORMAT=<ID=NS,Number=A,Type=Float,Description="Variant-specific number of samples/individuals with called genotypes used to test association with specified trait">
##FORMAT=<ID=EZ,Number=A,Type=Float,Description="Z-score provided if it was used to derive the ES and SE fields">
##FORMAT=<ID=SI,Number=A,Type=Float,Description="Accuracy score of summary association statistics imputation">
##FORMAT=<ID=NC,Number=A,Type=Float,Description="Variant-specific number of cases used to estimate genetic effect (binary traits only)">
##FORMAT=<ID=ES,Number=A,Type=Float,Description="Effect size estimate relative to the alternative allele">
##FORMAT=<ID=SE,Number=A,Type=Float,Description="Standard error of effect size estimate">
##FORMAT=<ID=LP,Number=A,Type=Float,Description="-log10 p-value for effect estimate">
##FORMAT=<ID=AF,Number=A,Type=Float,Description="Alternative allele frequency in trait subset">
##FORMAT=<ID=AC,Number=A,Type=Float,Description="Alternative allele count in the trait subset">
##study=<ID="12345678",Source="PubMed",Description="Bibliographic citation of main publication",URL="https://pubmed.ncbi.nlm.nih.gov/12345678">
##trait=<ID="EFO0004340",Description="Body mass index",Source="EFO",Version="3.14.0",Type="continuous",Test="linear",Unit="SD",Population="European",TotalSamples=461460,TotalVariants=9851866,VariantsNotRead=0,H
armonisedVariants=9851866,VariantsNotHarmonised=0,SwitchedAlleles=9851866,FileUrl="https://gwas.mrcieu.ac.uk/files/ukb-b-19953/ukb-b-19953.vcf.gz",FileDate="24/04/2020">
##trait=<ID="EFO0001360",Description="type II diabetes mellitus",Source="EFO",Version="3.14.0",Type="binary",Test="logistic",Population="European",TotalSamples=462933,TotalCases=2972,TotalVariants=9851866,VariantsNo
tRead=0,HarmonisedVariants=9851866,VariantsNotHarmonised=0,SwitchedAlleles=9851866,FileUrl="https://gwas.mrcieu.ac.uk/files/ukb-b-13806/ukb-b-13806.vcf.gz",FileDate="24/04/2020">
#CHROM  POS     ID  REF  ALT  QUAL  FILTER  INFO              FORMAT                 EFO0004340                                                  EFO0001360
1       49298   .   T    C    .     PASS    RSID=rs10399793   NS:NC:ES:SE:LP:AF:AC   463005:0:0.00103892:0.0034984:0.113509:0.613764:568351      463005:2972:9.098e-05:0.000294716:0.119186:0.613764:568351
1       49298   .   T    A    .     PASS    RSID=rs10399793   NS:NC:ES:SE:LP:AF:AC   463005:0:0.00214602:0.00346583:0.267606:0.011:4630          463005:2972:0.000102689:0.00029197:0.136677:0.012:4630
1       91536   .   GTC  G    .     PASS    RSID=rs6702460    NS:NC:ES:SE:LP:AF:AC   463005:0:0.00410514:0.0034125:0.638272:0.456845:423042      463005:2972:0.000329732:0.000287485:0.60206:0.456851:423042
1       534192  .   C    T    .     PASS    RSID=rs6680723    NS:NC:ES:SE:LP:AF:AC   463005:0:0.000334321:0.0038979:0.0315171:0.24094:223131      463005:2972:0.000106473:0.000328379:0.124939:0.24096:223131
1       706368  .   A    AAA  .     PASS    RSID=rs12029736   NS:NC:ES:SE:LP:AF:AC   463005:0:-0.00030371:0.00241981:0.0457575:0.515705:477487    463005:2972:7.16085e-06:0.000203854:0.0132283:0.51565:477487
```

The GWAS-VCF file contains study and trait(s) metadata, variant-level data, and variant-trait association summary statistics. Each field is defined in the file header including variable type and number of values. The format can store the GWAS results of one or more traits in a single file.

Table 7.5.2.1. Data fields in the GWAS-VCF

Field	Description
VCF Header	
Study	
ID*	Study identifier e.g., publication or data repository identifier e.g., 12345678 (PubMed https://pubmed.ncbi.nlm.nih.gov/) or phs001997.v1.p1 (dbGaP[184])
Source*	Source of study identifier e.g., PubMed or dbGaP[184]
Version	Version of study ID source used to describe study
Description	Study description
URL	Web link to study
Trait	
ID*	Trait identifier e.g., an ontology or metadata repository identifier e.g., EFO0004340 (EFO[179]), ieu-a-835 (MRC IEU OpenGWAS database[18]) or any other ontology
Source*	Source of trait identifier e.g. EFO[179] or MRC IEU OpenGWAS database[18]
Description*	Trait description e.g., Body mass index
Version	Version of trait ID source used to describe trait
Type	Outcome variable type (continuous or binary)
Test	Statistical test for association data e.g., linear regression
Unit	Phenotype units e.g., kg/m^2 or SD
Population	Participant ancestry (or mixed ancestry) using the standardised framework[185]
FileUrl	URL of GWAS summary statistics file
FileDate	Date GWAS summary statistics were produced
TotalSamples	Total number of samples/individuals in the study
TotalCases	Total number of cases in the study (if case-control)
TotalVariants	Total number of variants tested in the study
VariantsNotRead	Number of variants that could not be read
VariantsHarmonised	Number of harmonised variants
VariantsNotHarmonised	Number of variants that could not be harmonised
SwitchedAlleles	Number of variants strand switched
VCF Body	
Per trait variant-level information	
NS	Variant-specific number of samples/individuals with called genotypes used to test association with specified trait

EZ	Z-score provided if it was used to derive the ES and SE fields
SI	Accuracy score of association statistics imputation
NC	Variant-specific number of cases used to estimate genetic effect (binary traits only)
ES*	Effect size estimate relative to the alternative allele
SE*	Standard error of effect size estimate
LP*	-log10 p-value for effect estimate
AF	Alternative allele frequency for the trait GWAS
AC	Alternative allele count for the trait GWAS

ID, identifier. EFO, Experimental Factor Ontology. VCF, variant call format. * Required fields.

According to the VCF specification[165], the file header consists of metadata lines containing 1) the specification version number, 2) information about the reference genome assembly and contigs, and 3) information (ID, number, type, description, source, and version) about the fields used to describe variants and samples (or variant-trait associations in the case of GWAS-VCF) in the file body. The VCF file header is used to store additional information about the GWAS including 1) source and date of summary statistics, 2) study IDs (e.g., PMID/DOI of publication describing the study, or accession number and repository of individual-level data), 3) description of the trait(s) studied (e.g., type, association test used, sample size, ancestry and measurement unit) as well as the source and version of trait IDs (e.g., Experimental Factor Ontology[179], Human Phenotyping Ontology[180], Medical Subject Headings[181] IDs for clinical and other traits, Ensembl[139] Gene IDs for expression quantitative trait loci (eQTL) datasets, or any other ontology or identifier).

While VCF can contain information about multiple alternative alleles in a single row observed at the same site/locus[165], the GWAS-VCF specification requires that each variant is stored in a separate row of the file body. Each row contains eight mandatory fields: chromosome name (CHROM), base-pair position (POS), unique variant identifier (ID), reference/non-effect allele (REF), alternative/effect allele (ALT), quality (QUAL), filter (FILTER) and variant information (INFO). The ID, QUAL and FILTER fields can contain a null value (represented by a dot). Importantly, the ID value (unless null) should not be present in more than one row. The FILTER field may be used to flag poor quality variants for exclusion in downstream analyses. The INFO field is a flexible data store for additional variant-level key-value pairs (fields) and may be used to store for example: allele frequency for the entire

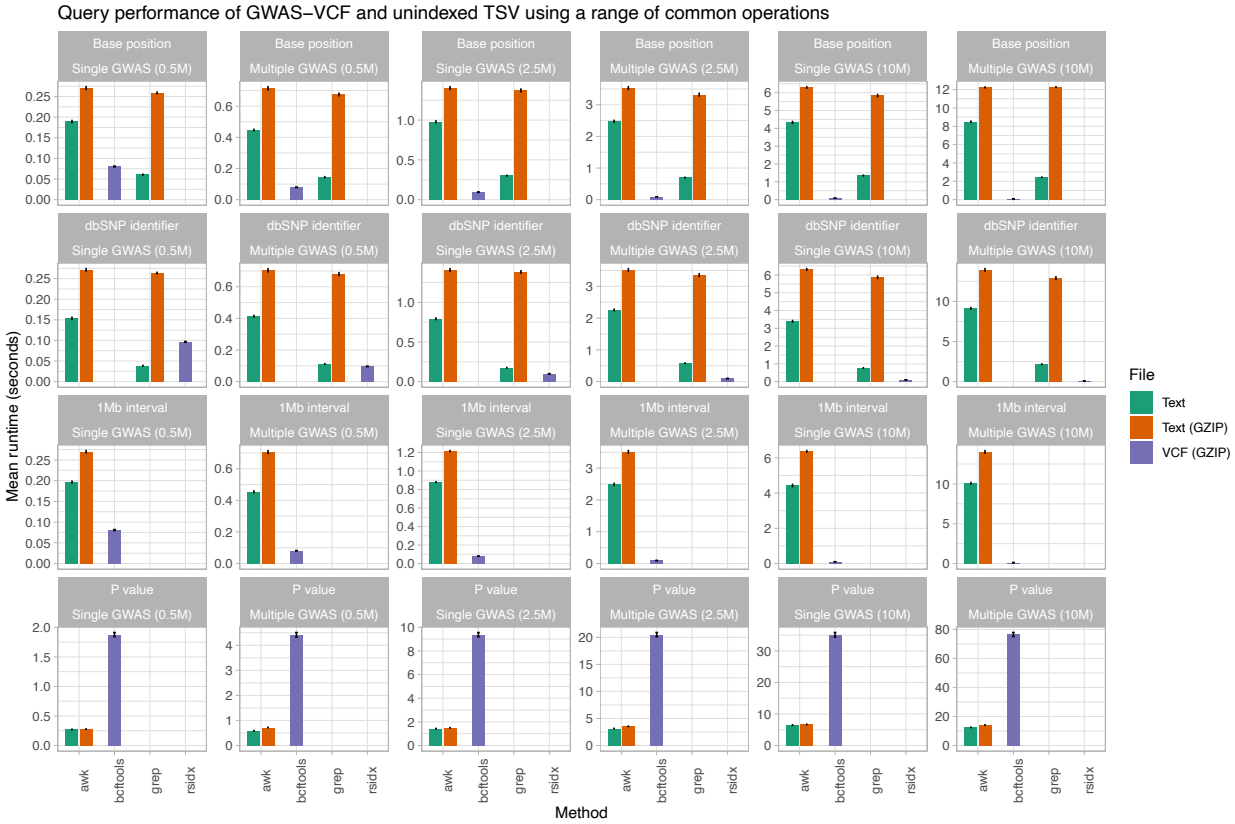
population AF), genomic annotations and variant functional effects. The INFO field is used to store the dbSNP[170] locus identifier (rsid; instead of ID field) for the site at which the variant resides. This is because (despite their common usage as variant identifiers) rsids[170] uniquely identify loci and thus cannot be used in the ID field which is required to contain a unique identifier for each row as per VCF[165] specification. Following the INFO column is a format field (FORMAT) and one or more sample columns which were used to store variant-trait association data, with values for the fields listed in the FORMAT column for example: effect size (ES), standard error (SE) and -log10 P-value (LP).

7.5.3 Query performance

Simulations of query performance demonstrate compressed GWAS-VCF is substantially quicker than unindexed and uncompressed TSV format for querying by genomic position when the GWAS is densely imputed (**Figure 7.5.3.1**). The greatest improvements were seen when the GWAS-VCF contained a single trait with 10 million variants where on average GWAS-VCF was 15x faster to extract a single variant using chromosome position (mean query duration of GWAS-VCF was 0.09 seconds [95% CI 0.08, 0.09] vs mean 1.35 seconds [95% CI 1.34, 1.37] for TSV) and 8x quicker using the rsid (0.1 seconds [95% CI 0.1, 0.1] vs 0.76 seconds [95% 0.75, 0.78]). Extracting a 1Mb window of variants GWAS-VCF was 44x quicker (0.1 seconds [95% CI 0.1, 0.11] vs 4.43 seconds [95% CI 4.36, 4.5]). However, querying on association P value was over 5x faster using TSV (mean query duration in TSV 6.48 seconds [95% CI 6.38, 6.57] vs mean query duration in GWAS-VCF 35.11 seconds [95% CI 34.35, 35.86]). When the number of variants stored in the GWAS-VCF was 0.5 million, uncompressed text was faster for single position and rsid lookups but not interval queries (**Figure 7.5.3.1**). Additionally, storing multiple

traits in a single GWAS-VCF reduced the P value query performance but had little impact on the positional queries (**Figure 7.5.3.1**).

Figure 7.5.3.1. Performance comparison for querying summary statistics in plain text and GWAS-VCF

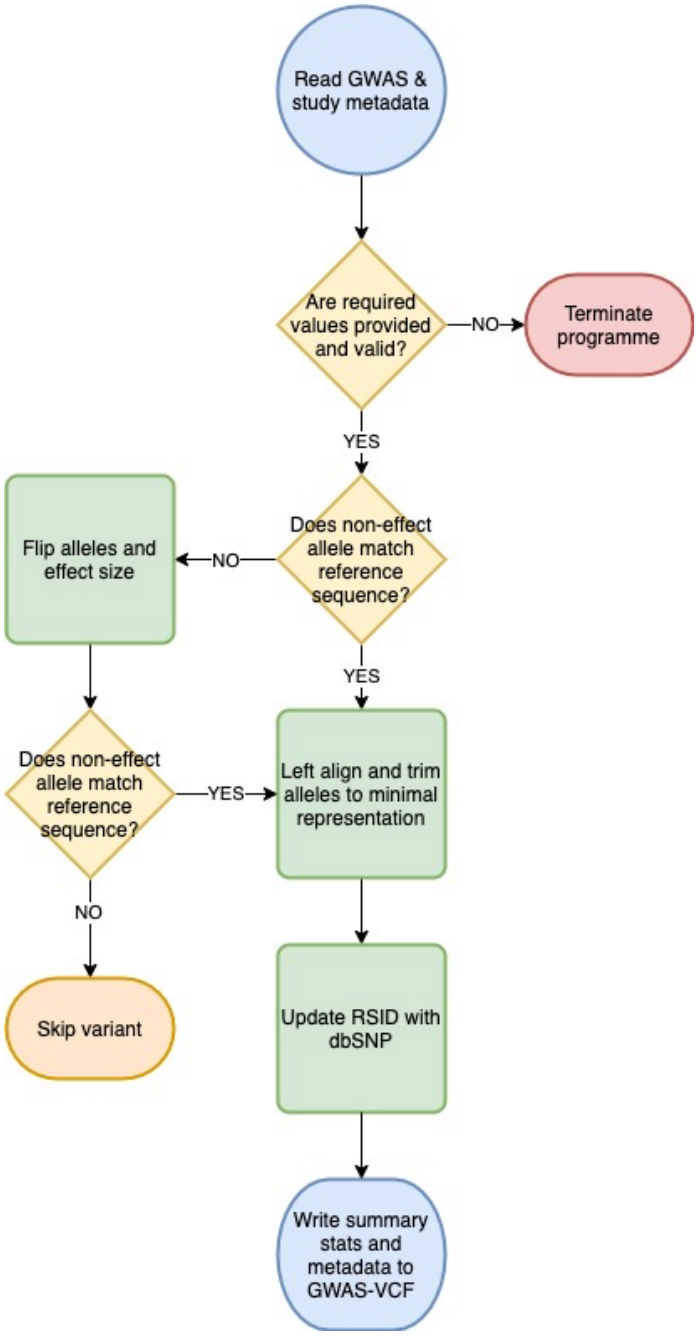


Mean query time (seconds, lower is quicker; repetitions n=100) to extract either: a single variant using the chromosome position or dbSNP[170] identifier or multiple variants using a 1 Mb interval or association P value. AWK, grep, bcftools[172] and rsidx[171] were evaluated using uncompressed/GZIP compressed TSV and BGZIP[172] compressed VCF. The summary statistics files contained one (single) or five (multiple) GWAS studies which were prepared by subsampling variants (n=0.5M, 2.5M, 10M) obtain from Neale *et al*[123]. Error bars represent the 95% confidence interval.

7.5.4 Software

To automate the conversion and harmonisation of existing summary statistics files to the GWAS-VCF format I developed gwas2vcf software. This software reads in metadata and SNP-trait association data using a user-defined schema requiring the chromosome base-position to start at one. During processing, variants are harmonised using a supplied reference genome file to ensure the non-effect allele matches the reference sequence enabling consistent directionality of allelic effects across studies. Insertion-deletion variants are left-aligned and trimmed for consistent representation using the vgraph library[186] (**Figure 7.5.4.1**). Finally, the GWAS-VCF is indexed using Tabix[24] and rsidx[171] which enable rapid queries by genomic position and rsid[170], respectively. I developed a freely available web application providing a user-friendly interface for this implementation (https://github.com/MRCIEU/gwas2vcfweb).

Figure 7.5.4.1. Workflow for gwas2vcf



Flow diagram of gwas2vcf theory of operation. GWAS, genome-wide association study. VCF, variant call format.

Once stored in a GWAS-VCF file, summary statistics can be read and queried using R (developed by Dr Gibran Hemani, MRC-IEU), Python (which I developed; https://github.com/MRCIEU/pygwasvcf), or from the command line using for example: bcftools[172], GATK[178] or bedtools[138]. These tools also enable variant annotation and filtering (e.g. allele frequency, functional effect, gene and pathway), mapping between reference genome assemblies, file validation and converting to any other tabular format including the NHGRI-EBI GWAS Catalog format[19]. I have provided tutorials on how to perform downstream analysis of GWAS-VCF files (https://mrcieu.github.io/gwas2vcf/downstream).

7.5.5 Variance GWAS summary statistics in GWAS-VCF

I converted 30 complete variance GWAS summary statistics prepared in **Chapter 5** to GWAS-VCF and made these available for download through the MRC IEU OpenGWAS database[18].

7.6 Discussion

The GWAS-VCF format has several advantages over existing solutions. First, the VCF[165] provides a consistent and robust approach to storing genetic variants, annotations and metadata enabling interoperability and reusability consistent with the FAIR principles[187]. Furthermore, variable type and number requirements[165] reduce parsing errors and missing data, preventing unexpected program operation. Second, the VCF is well established and scalable to support GWAS of whole-genome sequencing studies[165]. Many mature tools have been developed providing a range of functions for querying, annotating, transforming, and analysing genetic data in VCF (**Table 7.5.1.1**). Third, the GWAS-VCF file header stores comprehensive metadata about the GWAS including necessary information to understand the

analysis and interpret the data (**Table 7.5.2.1**). Fourth, a GWAS-VCF file can store individual or multiple traits (in one or more sample columns) in a single file which is beneficial for the distribution of GWAS datasets where genotypes of each sample/individual have been tested for association with multiple traits (e.g., QTL datasets[141]).

The simulation studies demonstrated GWAS-VCF was substantially quicker when the GWAS was densely imputed (8-44x) than TSV using standard UNIX tools for extracting records by genomic position. Although the GWAS-VCF was slower for extracting records by association P value this could be improved by using variant flags (i.e., in the INFO field) to highlight records below prespecified thresholds. For example, all variants below genome-wide significance ($P < 5 \times 10^{-8}$) or a more relaxed threshold (e.g., $P < 5 \times 10^{-5}$).

7.7 Limitations

A limitation of GWAS-VCF is the lack of a widely adopted and stable representation of variants that can be used as a universal unique identifier. Published summary statistics often use rsids[170] to identify the variant substitution but this practice is inappropriate because rsids are locus identifiers and do not distinguish between multiple alternative alleles observed at the same site[188]. Moreover, rsids are not stable as they can be merged and retired over time[170]. The reason this is a problem is that in GWAS summary statistics every record represents the effect of a specific allele on one or more traits (**Chapter 1.2.5**), and if a record identifier is used that is not unique for each allelic substitution, then the association statistics cannot be correctly interpreted for a specific allele. An alternative approach is to concatenate chromosome, base position, reference, and alternative allele field values into a single string[22], but this is non-standardised, genome build specific and unwieldy for long insertion-deletion variants. In the

current version of the GWAS-VCF specification it is suggested to query variants by chromosome and base-position and filtering the output to retain the target substitution, but this approach can be cumbersome and difficult to interoperate with other software. The ideal solution would be for the community to adopt universally accepted variant identifiers that can then be used in the ID column of GWAS-VCF files.

7.8 Conclusions

The VCF specification for GWAS summary statistics (GWAS-VCF) was defined to be amenable to high-throughput analyses and robust data sharing and integration. I implemented open-source Python tools to convert existing summary statistics formats into this format and to query the file to extract subsets of data. I also produced examples of integrating these data with existing analysis tools. Finally, I provided 30 variance GWAS summary statistics datasets from **Chapter 5** in GWAS-VCF. These resources enable convenient and efficient secondary analyses of GWAS summary statistics and support future tool development.

Chapter 8: Discussion

8.1 Overview

Variance QTLs are genetic loci that associate with the variance of a trait and can indicate the presence of gene-interaction effects[56,68] (**Chapter 1.5.3**). Previous research has used vQTL evidence to prioritise loci for interaction testing using a series of candidate modifiers[68,69]. While many variance tests exist, few tools are available for GWAS identification[69,70,96] (**Chapter 1.5.7**) and analysis of vQTLs. In addition, efficient, and robust downstream analysis of GWAS summary statistics including variance GWAS requires adoption of a common data format[22,166]. Furthermore, vQTL evidence may have other applications not previously explored for example in testing of MR[15] homogeneity assumptions[35] (**Chapter 6**). The aims of this thesis were to develop software tools and methodology to support the discovery, analysis and sharing of vQTLs. As an exemplar these methods were applied to study serum biomarker concentration in UK Biobank which include causal modifiable risk factors for disease[99] with the aim of producing findings of translational value (**Chapter 1.4.1**; **Chapter 1.6**). The key methodology advances and findings of this thesis are presented in **Table 8.1.1**.

I implemented a regression-based Brown-Forsythe test for robustly detecting vQTLs with equal power and type I error rate to the original method. Although the power of this approach was low and may only be suitable in cases where variants explain a large proportion of trait variance such as with molecular QTLs[132]. A second limitation of LAD-BF was the inability to use imputed dosage values directly, instead these should be rounded to whole numbers which is necessary for detecting non-linear effects on trait variance in the second-stage regression model. Through simulation, I demonstrated the ability to estimate an unbiased variance effect

and to adjust this effect for confounding. I showed how these features are useful to determine if an interaction term is contributing to a variance effect by using the novel approach of adjusting the model for candidate interactions and measuring attenuation of variance effects under simulation. Secondly, I demonstrated that when ancestry has an interaction effect on an outcome adjusting both LAD-BF regression models for ancestry can reduce bias due to population stratification (**Chapter 1.2.3**). I implemented the LAD-BF model in R (github.com/MRCIEU/varGWASR) and scalable C++ GWAS software (github.com/MRCIEU/varGWAS) available open-source for future research at scale. Type I error rate simulations produced in **Chapter 4** were included in a publication in the European Journal of Epidemiology (2021)[80], on which I am a co-author.

I applied methodology developed in **Chapter 4** to identify vQTLs influencing 30 serum biomarkers in UK Biobank (**Chapter 5**). These results included association for 290 million SNPs which I have made freely available for download through the MRC-IEU OpenGWAS platform[18] to support secondary research applications such as in **Chapter 5** and **Chapter 6**. Among these vGWAS I identified 468 loci robustly associated with the variance of serum biomarker concentrations which I further investigated by testing for interaction effects leading to the identification of six gene-gene and 82 gene-environment interaction effects. I prepared a manuscript containing findings from **Chapter 4** and **Chapter 5**[2]. Expertise gained through this work was also employed to contribute to another study investigating effect modification of C-reactive protein variants in cardiometabolic disease, which is under review with Human Molecular Genetics[189], on which I am a co-author.

I used simulations to show how instrument-exposure variance effects estimated using LAD-BF developed in **Chapter 4** may be used to assess instrumental variable homogeneity assumptions (**Chapter 6**). This approach could be applied as a sensitivity analysis to remove instruments with strong exposure variance effects to minimise bias of the population average causal effect[35] or as a falsification strategy to test if homogeneity assumptions are violated. However, this approach requires detection of instrument-exposure variance effect which is low powered and may challenging to implement with current sample sizes. While applied in an MR (**Chapter 1.3.7**) setting this methodology could have wider utility for any instrumental variable analysis applied, for example, in health[15], economics[190], and social science[191]. I used this approach to investigate the effects of LDL, glucose and urate on cardiovascular disease, type 2 diabetes, and gout, respectively using variance GWAS summary statistics produced in **Chapter 5**. Although some instruments for these exposures had variance effects on the exposure, removing these instruments had little impact on causal estimates. These findings suggest NOSH assumption one[35] was not violated which could imply either the variance effect was not a consequence of an interaction of the instrument-exposure relationship, or the exposure-outcome relationship was not modified by the same variable as the instrument-exposure. I am currently preparing a manuscript for publication of this work.

The variant call format[165] (VCF) was extended in collaboration with researchers at the MRC-IEU and School of Medicine at Mount Sinai, New York to develop a robust and efficient file format for storing and sharing of GWAS summary statistics (GWAS-VCF) including vQTLs (**Chapter 7**). This format facilitates downstream analyses and has already been used widely with the paper amassing 34 citations to date. I published the full variance GWAS summary statistics

produced in **Chapter 5** using this format through the MRC-IEU OpenGWAS platform[18]. I found GWAS-VCF was 8.6-45.5x faster to extract GWAS results by chromosome position than extracting records by reading the file line-by-line which can be used to improve the performance of downstream analyses supporting large-scale hypothesis-free analyses. I produced open-source Python software to convert GWAS summary statistics to GWAS-VCF (github.com/MRCIEU/gwas2vcf) and web interface to automate this process (https://github.com/MRCIEU/gwas2vcfweb). I also developed an open-source Python library for reading these files (github.com/MRCIEU/pygwasvcf). The GitHub repository has received 25 stars and between 18th June - 1st July 2022 was visited and downloaded by 60 and 16 unique users, respectively. This work was published in Genome Biology (2021)[1]. Following expertise gained through this project, I participated in the NHGRI-EBI GWAS Catalog data format and content working group (2021) which has led to the development of another GWAS summary statistics standard and preprint on which I am a co-author[25]. This work was also used in the development of the MRC-IEU OpenGWAS platform[18] currently under review with eLife, and in producing MR estimates for the EpigraphDB platform[192] published in Bioinformatics (2020). I contributed towards development of these manuscripts on which I was included as co-author.

Table 8.1.1. Summary of results and methodological advances

Section	Findings
	Chapter 4
4.5.3	Implemented LAD-BF in C++ and R which produces an unbiased variance effect estimate
4.5.4	vQTL confounding by ancestry can be controlled using LAD-BF but not with the Brown-Forsythe test
4.5.5	Adjusting LAD-BF for the interaction effect will attenuate the variance effect
4.5.6	LAD-BF has greater power to detect the presence of an interaction effect compared with exhaustive testing of linear regression and a series of candidate modifiers
	Chapter 5
5.5.1	Detected 468 independent vQTLs influencing 30 serum biomarkers in UK Biobank
5.5.2	Detected 82 scale-independent gene-environment interaction effects
5.5.3	Detected 6 scale-independent gene-gene interaction effects including possible novel effects of *TREH* rs12225548 x *FUT2* rs281379 and *ABO* rs635634 x *TREH* rs12225548 on ALP and *ZNF827* rs4835265 x *NEDD4L* rs4503880 on GGT
	Chapter 6
6.5.1	Instrument-exposure variance effects may be used to partially assess NOSH assumption one violation[35]
6.5.2	Removing instruments with strong exposure variance effect may reduce bias of the population average causal effect.
6.5.3	No strong evidence for departure from the population average causal effect for the effects of low-density lipoprotein on coronary heart disease, random glucose on type 2 diabetes or urate on gout. There may exist a pathway of low-density lipoprotein on coronary heart disease that is protective
	Chapter 7
7.5.1	GWAS-VCF is a robust solution for distributing GWAS summary statistics
7.5.3	GWAS-VCF is up to 46x faster to query plain text files used
7.5.4	I developed gwas2vcf software to support cataloguing of GWAS summary statistics
7.5.5	Deposited variance GWAS summary statistics for 30 biomarkers in MRC-IEU OpenGWAS

LAD, least absolute deviation. vQTL, variance quantitative trait loci. GWAS, genome-wide association study. NOSH, NO Simultaneous Heterogeneity. VCF, variant call format.

8.2 Contribution statement

Chapter 8.3 contains future work from a manuscript I wrote that was edited by PhD supervisors available as a preprint on MedRxiv (Lyon *et al*, 2022)[2].

8.3 Future work

8.3.1 Variance GWAS studies of protein measurements

Throughout this thesis I have focused on variance studies of serum biomarker concentration in UK Biobank as an exemplar (**Chapter 1.6**). There were two key reasons for selecting these traits. First, these measures are continuous surrogate endpoints for disease outcomes that are easier to measure and apply to variance QTL studies[98] and findings may extend to disease outcomes. Secondly, serum biomarkers include causal risk factors for disease that are potential intervention targets[99] and interaction findings in combination with other evidence may be useful for drug development (**Chapter 8.3.2**). This is in contrast with previous studies that have largely investigated physical measures such as BMI, height, lung function, and bone mineral density[68,69].

One key limitation of this work is low power to detect vQTLs, this is because individual variants explain only a small amount of biomarker variance[168]. Future studies should not apply these methods to polygenic traits because power to detect vQTLs is too low to reliably identify such loci. The simulation in Chapter 4.5.2 suggests a SNP explaining 5% of trait variance is needed to obtain approximately 80% power to detect the vQTL given samples sizes available in biobanks today.

Meanwhile, plasma proteins (as well as other molecular phenotypes such as gene expression) are strongly affected by SNPs in the *cis*-coding region that explain on average 5.8%

of trait variance[132]. Future vQTL studies could be extended to investigate these molecular phenotypes (e.g. in UK Biobank protein concentration measurements are due to be made available this year as part of the UK Biobank Pharma Proteomics Project[193]) which are adequately powered and may lead to the identification of novel biology or drug targets with interaction effects on disease outcomes.

8.3.2 Drug target prioritisation

It is not normally possible to intervene on biomarker concentration directly[99]. Instead, drugs may be developed to modulate proteins that regulate biomarker synthesis or metabolism[99], for example, statins act to inhibit HMG-CoA reductase lowering serum LDL cholesterol levels[99]. GWAS of mean biomarker concentration and follow up analyses can help identify protein targets for drug development[99]. However, as biomarkers are complex traits[99], they are affected by genetic and environmental factors which may interact producing gene-gene or gene-environment interaction effects[48] (**Chapter 1.4.1**). Identification of loci with interaction effects on biomarker concentration may provide evidence of drug targets that, when intervened produce subgroup effects with individual variation in response to treatment dependent on the modifier[57]. However, this approach is likely to be limited by low power, requiring a large sample size and SNPs explaining a large proportion of trait variance such as with protein or gene expression traits as discussed above.

The results in **Chapter 5** could be combined with other available evidence to identify drug targets with potential subgroup effects on a given biomarker during preclinical drug development. Intervention on such targets may produce differential effects on the indication[194] that could have low, no or opposing efficacy in some subgroups[57]. Variance QTLs may have a

potential role in preclinical drug development to identify subgroups where efficacy is higher, although their detection is low powered and complicated by the presence of mean-variance confounding (**Chapter 1.5.4**). As an example of this, consider the effect of *SLC2A9* rs938555 x sex on urate (**Chapter 5**), suppose rs938555 is acting on urate concentration via the SLC2A9 protein abundance (and not by horizontal pleiotropy[161]). Then a drug developed to target SLC2A9 protein abundance may also show differing effects by sex on urate and by extension gout.

8.3.3 Use of joint test to identify loci involved in genetic interaction

I applied the LAD-BF model to detect effects on trait variance with follow up studies to identify genetic interaction effects (**Chapter 4**). Among these findings were the observation that almost all SNPs with interaction effects also affected the mean and prioritisation of either mean or variance effects may be fruitful in detecting interactions. Therefore, future studies could use a joint test for mean and variance effects, using for example JLSsc[80] or LRTmv[86] in order to prioritise loci for interaction testing which may have greater power than just screening for variance effects only.

8.3.4 Improved control for population stratification

GWAS are susceptible to genetic confounding (**Chapter 1.3.2; Chapter 1.2.3**) which may produce spurious results[7]. Family-based designs are advantageous in that they can attenuate bias from population stratification, as well as dynastic effects and assortative mating[8,12]. Another approach to reduce effects of population stratification (**Chapter 1.2.3**) is the use of a population random effect. In future work I aim to implement a population random effect in the LAD-BF model, which could be achieved using the GRAMMAR approach[13]. This method first

estimates the population random effect[13] which can be included in the LAD-BF model as a covariate.

8.3.5 Estimating the causal effect of an exposure on the variance of an outcome

I showed a limitation of the variance GWAS approach is low power to detect vQTLs and consequently low power to detect interaction effects operating at a locus (**Chapter 4**). One way this could be improved is to combine multiple SNPs using for example polygenic risk scores[7] (PRS). The PRS could serve as a genetic instrument to proxy for an exposure[195] such as complex trait or molecular phenotype to test for a variance effect of exposure on an outcome which may imply the presence of an interaction between exposure and outcome. An alternative strategy is to estimate the effect of each instrument on outcome variance separately and then meta-analyse the effects[196]. The use of dummy SNPs in the second-stage LAD-BF model means that two coefficients are produced and would require the use of bivariate meta-analysis[197].

8.3.6 Comparison of vQTL evidence with randomised control trials

Studies of variance effects have also been investigated using RCTs[43,50] where outcome variance was compared between trial arms as evidence for interaction. For example, Cortés *et al*[50] systematically reviewed 208 trials and found 7.2% had evidence of increased variance in the treatment group which is consistent with effect modification. Other studies have used trial data to investigate variance effects on brain volume[198], biomarker concentration[199] pain[200], depression[43] and schizophrenia[201]. This evidence could be integrated with genetic variance effects in order to triangulate findings[202], although the latter is low powered which may make this approach challenging.

8.3.7 Variance QTL evidence for fine mapping of causal loci

Phantom vQTL effects (**Chapter 1.5.5**) occur when a SNP is in imperfect linkage disequilibrium with the causal SNP having a strong mean effect but no interaction effect[53,94]. The strength of the phantom vQTL effect is a function of the effect size and allele frequency of the causal SNP effect, correlation between SNPs, and allele frequency of the non-causal SNP[68]. Thus, vQTL evidence may have a role in fine mapping of causal loci (**Chapter 1.2.4**) by prioritising SNPs with the largest mean effect and smallest variance effect. However, in situations where a true interaction effect exists this approach may instead deprioritise the causal SNP. This evidence could be integrated in existing Bayesian models[9] to strengthen evidence and improve resolution for causal SNP prioritisation applied to continuous traits.

8.3.8 Colocalization of vQTL and QTL

Colocalization is a method to determine if shared causal variant(s) exist between two traits at a single genetic locus which would be anticipated if one trait has a causal effect on the other[203].

Colocalization of vQTLs with GWAS of other traits including molecular phenotypes could provide insight into the causal mechanisms underlying the association with trait variance. For example, suppose a vQTL is a consequence of an interaction effect, then colocalization with other traits could be used to identify interacting exposures which could then be tested using a formal genetic interaction test.

Colocalization of vQTLs with other GWAS traits has already been performed using *FTO* locus vQTL to check for shared causal variant[68] and Westerman *et al*[42] but these analyses were conducted using the OSCA effect estimate which is biased under an interaction effect (**Chapter**

4) and may reduce colocalization performance. Further studies should be conducted to explore colocalization of vQTLs using a model that makes no linearity assumptions, for example using LAD-BF (**Chapter 4**). How to do this given two coefficients are provided is left to future work.

8.3.9 Systematic testing of homogeneity assumptions in Mendelian randomization

I developed methodology to detect violation of instrumental variable homogeneity assumptions (**Chapter 6**). As an exemplar this work was applied to investigate the effects of selected biomarker traits (**Chapter 1.6**) on disease outcomes where findings may yield high translational value. Future studies could apply this methodology systematically to determine if there are causal effects where homogeneity assumptions are strongly violated and to estimate effects without vQTL instruments. For example, this could be applied using all continuous exposures in UK Biobank on all outcomes in the MRC-IEU OpenGWAS platform[18] as has been done for systematic MR analyses of 'everything against everything'[161]. These results could be used to populate a database that researchers are able to inspect and would aid in widespread adoption of this evidence. For effects where homogeneity assumptions are violated a statement could be included in the report to suggest the population average causal effect may not be targeted which would aid interpretation. Additionally, an effect estimate produced without vQTL instruments could be provided which may be closer to the population average causal effect.

8.3.10 Runtime performance improvements for varGWAS

I developed variance GWAS software (varGWAS) to implement the LAD-BF model (**Chapter 4**). I used C++ and multithreading to improve scalability which enabled successful application to 30 traits in over 300k UK Biobank participants (**Chapter 5**). However, this analysis

was very computationally demanding taking around 2,400 CPU hours per trait which would cost approximately £160 (ex VAT) per GWAS to perform using Microsoft Azure (azure.microsoft.com) cloud computing (pricing February 2022). These computational requirements represent a barrier to adoption for future studies applied to large numbers of traits. There are several performance improvements that could be applied to reduce runtime requirements. First, specialist computing hardware such as the Graphics Processing Unit (GPU) which are widely available and could be employed to improve the number of simultaneous loci tested for a variance effect reducing computing time and resources. GPUs have already been applied to genetic epidemiology with success. For example, epiGPU[204] was developed to perform exhaustive testing for gene-gene interaction effects and was 92x faster than using a single CPU core[204]. Second, a low-density coverage first pass could be performed which would involve testing a random subsample of the directly genotyped SNPs and then performing focused testing around associated loci to identify lead SNPs. Alternatively, a genome-wide pre-screen could be applied using a less computationally intensive model to identify loci which may be associated with trait variance for analysis by LAD-BF. For example, one of the first variance GWAS used a standard GWAS linear model applied to the square of the standardised trait[77] but this approach assumes linearity between SNP and trait variance which I show (**Chapter 4**) does not hold under an interaction effect. Although this approach could be followed up by LAD-BF.

8.3.11 Developing binary format for storing of GWAS summary statistics

While the GWAS-VCF format[1] (**Chapter 7**) solves a range of issues working with these data such as ensuring consistency of metadata and allele and effect orientation it is uncompressed hence inefficient for large numbers of GWAS in terms of storage requirements.

One way this might be addressed is to develop a binary format which enables storage of fields using the smallest number of bytes possible as has been done with other formats such as BGEN[130], BCF[165] and BAM[173]. Secondly, these data could be stored along with file indexes based on chromosome position or RS identifier in a single file using the HDF5 format[205] to avoid the need to distribute multiple files.

8.4 Summary

In this thesis I discovered vQTL effects on serum biomarker concentration in UK Biobank, performed comprehensive analyses to identify potential gene-interaction effects and applied these data to test instrumental variable homogeneity assumptions. I also developed a robust and performant format to share these data to support secondary research applications. To facilitate these analyses, I produced open-source software that may be valuable in future research studies.

Chapter 9: Appendix

9.1 Algebraic expression of exposure interaction effect on outcome variance

Professor Tilling derived the following expression to calculate the expected variance of an outcome conditional on an exposure under an interaction effect on the outcome.

Equation 9.1.1 Variance of outcome conditional on genotype with interaction effect

For the ith observation suppose exposure X_i and modifier U_i interact XU_i on outcome Y_i such that:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 U_i + \beta_3 XU_i + E_i$$

Where E_i is the residual variance following Normal distribution with mean of zero and unit variance. Then the relationship between X_i and Y_i variance can be calculated as:

$$var(Y|X) = \gamma_0 + \gamma_1 X_i + \gamma_2 X_i^2$$

Where:

$$\gamma_0 = \beta_3^2 var(U_i) + var(E_i)$$

$$\gamma_1 = 2\beta_2\beta_3 var(U_i)$$

$$\gamma_2 = \beta_3^2 var(U_i)$$

9.2 Supplemental tables

Table 9.2.1. Effect of top vQTLs on standardised biomarker variance in UK Biobank

SNP	Outcome	Gene	phi_x1	95% CI		phi_x2	95% CI		P_{Lyon}	$P_{Westerman}$
rs4654970	ALP	*ALPL*	-0.093	-0.102	-0.085	-0.153	-0.179	-0.127	6.04E-109	1.93E-45
rs12449427	ALP	*CLDN7*	0.048	0.04	0.056	0.094	0.077	0.111	1.81E-52	1.43E-11
rs9987289	ALP	*PPP1R3B*	-0.08	-0.125	-0.034	-0.141	-0.183	-0.099	2.67E-40	1.95E-04
rs10740131	ALP	*REEP3*	0.027	0.019	0.036	0.065	0.054	0.076	3.16E-33	-
rs167537	ALP	*ASGR2*	0.043	0.034	0.052	0.095	0.068	0.122	2.68E-32	3.51E-21
rs112875651	ALP	*TRIB1*	-0.032	-0.04	-0.024	-0.064	-0.075	-0.054	1.21E-31	6.74E-03
rs281379	ALP	*FUT2*	-0.023	-0.032	-0.014	-0.056	-0.066	-0.046	2.15E-26	6.51E-05
rs4654748	ALP	*NBPF3*	-0.029	-0.037	-0.02	-0.057	-0.067	-0.047	2.38E-26	-
rs635634	ALP	*ABO*	-0.041	-0.049	-0.034	-0.001	-0.023	0.021	4.18E-24	1.44E-68
rs1169312	ALP	*C12orf43*	0.025	0.017	0.032	0.051	0.038	0.063	1.96E-18	-
rs1058935	ALP	*IFITM2*	0.014	0.005	0.023	0.041	0.03	0.051	3.29E-14	-
rs12225548	ALP	*TREH*	-0.022	-0.029	-0.014	-0.054	-0.07	-0.038	9.06E-14	1.81E-06
rs1154416	ALP	*ADH5*	-0.014	-0.024	-0.003	-0.037	-0.047	-0.026	5.74E-12	-
rs74697591	ALP	*TXNL4B|HP|HPR*	0.023	0.014	0.031	0.071	0.043	0.099	1.36E-11	5.10E-03
rs12825673	ALP	*B4GALNT3*	0.024	0.014	0.034	0.039	0.028	0.05	1.86E-11	-
rs174570	ALP	*FADS2*	0.02	0.011	0.029	0.088	0.056	0.12	2.94E-11	1.19E-03
rs4940697	ALP	*NEDD4L*	-0.032	-0.048	-0.017	-0.047	-0.062	-0.032	4.60E-10	4.19E-04
rs2576452	ALP	*TMC4*	0.017	0.009	0.025	0.035	0.024	0.046	8.68E-10	-
rs738409	ALT	*PNPLA3*	0.133	0.124	0.143	0.396	0.365	0.427	0	1.78E-257
rs58542926	ALT	*TM6SF2*	0.123	0.109	0.137	0.283	0.203	0.363	4.57E-94	6.72E-40
rs2862954	ALT	*ERLIN1*	-0.058	-0.068	-0.047	-0.123	-0.134	-0.111	8.95E-88	5.28E-35
rs71633359	ALT	*HSD17B13*	-0.07	-0.078	-0.061	-0.104	-0.118	-0.091	3.98E-71	3.22E-45
rs429358	ALT	*APOE*	-0.068	-0.077	-0.058	-0.085	-0.11	-0.06	1.45E-46	3.45E-24
rs2954021	ALT	*TRIB1*	-0.053	-0.064	-0.042	-0.087	-0.099	-0.075	7.30E-44	1.77E-15

rs4979371	ALT	*AKNA*	-0.04	-0.05	-0.029	-0.086	-0.097	-0.074	1.48E-42	4.89E-20
rs10787429	ALT	*GPAM*	-0.046	-0.064	-0.028	-0.079	-0.096	-0.062	6.97E-25	5.60E-14
rs2642438	ALT	*MARC1*	0.041	0.026	0.056	0.073	0.058	0.088	2.61E-24	4.56E-08
rs4841133	ALT	*PPP1R3B*	-0.05	-0.103	0.003	-0.106	-0.156	-0.055	2.98E-23	1.70E-06
rs1262002	ALT	*DLG5*	0.021	0.008	0.034	0.055	0.042	0.068	3.76E-19	4.50E-09
rs4503880	ALT	*NEDD4L*	-0.039	-0.063	-0.016	-0.073	-0.096	-0.051	2.08E-18	4.78E-05
rs4835265	ALT	*ZNF827*	0.036	0.026	0.046	0.081	0.05	0.111	1.24E-17	2.07E-06
rs1801282	ALT	*PPARG*	-0.039	-0.05	-0.029	-0.085	-0.116	-0.053	1.13E-16	9.76E-12
rs4782568	ALT	*OSGIN1*	-0.025	-0.035	-0.015	-0.053	-0.065	-0.041	2.66E-16	1.92E-11
rs57562692	ALT	*PANX1*	-0.035	-0.045	-0.025	-0.102	-0.134	-0.069	9.53E-16	9.02E-07
rs35348663	ALT	*TBL2*	0.028	0.019	0.037	0.057	0.04	0.073	1.31E-15	2.15E-12
rs13423088	ALT	*NYAP2*	-0.023	-0.032	-0.014	-0.056	-0.07	-0.042	1.45E-14	2.61E-03
rs132639	ALT	*APOL3*	0.048	0.025	0.072	0.075	0.052	0.098	4.50E-14	-
rs36086195	ALT	*ARHGEF19*	0.017	0.005	0.029	0.044	0.031	0.057	1.72E-12	7.24E-03
rs1169288	ALT	*HNF1A*	0.02	0.011	0.029	0.053	0.037	0.069	2.55E-12	5.23E-08
rs72731415	ALT	*SRP14*	-0.023	-0.032	-0.014	-0.046	-0.059	-0.032	2.00E-11	1.46E-04
rs7029757	ALT	*TOR1B*	-0.038	-0.049	-0.027	-0.055	-0.098	-0.012	2.20E-11	7.72E-07
rs4973550	ALT	*EFHD1*	0.03	0.017	0.042	0.045	0.032	0.058	8.16E-11	2.07E-04
rs3816873	ALT	*MTTP*	-0.022	-0.031	-0.013	-0.051	-0.068	-0.034	1.01E-10	3.05E-05
rs339969	ALT	*RORA*	0.011	-0.001	0.023	0.038	0.025	0.051	1.65E-10	2.37E-05
rs641738	ALT	*TMC4*	0.018	0.009	0.028	0.042	0.029	0.055	2.69E-10	2.00E-04
rs4810880	ALT	*PREX1*	-0.028	-0.038	-0.019	-0.034	-0.046	-0.021	5.33E-10	2.21E-04
rs10075805	ALT	*CDH6*	-0.024	-0.033	-0.015	-0.041	-0.057	-0.025	9.03E-10	5.51E-03
rs1477066	ALT	*SOX9*	0.009	-0.005	0.023	0.035	0.02	0.049	9.40E-10	5.16E-05
rs28413626	ALT	*SETD8*	-0.026	-0.035	-0.017	-0.043	-0.063	-0.024	9.91E-10	1.71E-03
rs247616	ApoA	*CETP*	0.056	0.046	0.065	0.125	0.108	0.141	1.02E-68	3.09E-04

rs1077835	ApoA	*ALDH1A2|LIPC*	0.043	0.034	0.053	0.131	0.107	0.155	1.10E-42	7.72E-05
rs13284054	ApoA	*ABCA1*	-0.047	-0.057	-0.036	-0.114	-0.146	-0.083	2.08E-24	5.42E-06
rs2043085	ApoA	*ALDH1A2*	-0.033	-0.046	-0.02	-0.061	-0.075	-0.048	9.15E-19	-
rs1943973	ApoA	*LIPG*	0.037	0.007	0.066	0.078	0.048	0.107	1.39E-18	3.67E-04
rs12721030	ApoA	*APOA1*	0.034	0.025	0.044	0.064	0.039	0.089	2.62E-15	-
rs174528	ApoA	*MYRF|TMEM258*	-0.022	-0.032	-0.012	-0.046	-0.059	-0.033	4.05E-11	1.31E-04
rs1042034	ApoA	*APOB*	-0.025	-0.049	-0.001	-0.053	-0.076	-0.03	5.93E-11	-
rs112310696	ApoA	*DUS2*	0.033	0.022	0.044	0.066	0.028	0.104	1.46E-10	-
rs5167	ApoA	*APOC4|APOC4-APOC2*	0.018	0.009	0.028	0.047	0.032	0.062	3.11E-10	5.78E-03
rs72654473	ApoA	*APOE*	0.037	0.026	0.049	0.038	-0.004	0.08	3.49E-10	3.36E-03
rs438811	ApoB	*APOC1*	0.115	0.105	0.124	0.309	0.284	0.333	1.74E-294	8.68E-10
rs12740374	ApoB	*CELSR2*	-0.07	-0.078	-0.061	-0.124	-0.141	-0.107	1.47E-78	6.04E-04
rs6511720	ApoB	*LDLR*	-0.084	-0.093	-0.074	-0.143	-0.171	-0.115	1.71E-72	-
rs562338	ApoB	*APOB*	0.039	0.017	0.061	0.102	0.079	0.125	1.34E-52	2.86E-03
rs964184	ApoB	*APOA5*	-0.051	-0.087	-0.016	-0.112	-0.145	-0.079	9.10E-39	-
rs28601761	ApoB	*TRIB1*	-0.042	-0.051	-0.032	-0.077	-0.088	-0.065	7.14E-37	-
rs3208305	ApoB	*LPL*	-0.036	-0.045	-0.027	-0.06	-0.074	-0.045	6.46E-22	7.99E-04
rs369599	ApoB	*PVRL2*	-0.028	-0.037	-0.019	-0.057	-0.071	-0.042	4.74E-17	3.34E-43
rs780093	ApoB	*GCKR*	-0.025	-0.038	-0.013	-0.052	-0.065	-0.04	1.11E-16	5.42E-03
rs626787	ApoB	*USP1*	-0.025	-0.034	-0.016	-0.053	-0.066	-0.041	4.57E-16	-
rs4245791	ApoB	*ABCG8*	-0.025	-0.04	-0.011	-0.05	-0.064	-0.036	1.73E-13	-
rs174576	ApoB	*FADS2*	-0.015	-0.024	-0.006	-0.05	-0.063	-0.037	1.12E-12	2.06E-03
rs56228609	ApoB	*CETP*	-0.015	-0.024	-0.007	-0.053	-0.067	-0.039	1.44E-12	-
rs56174528	ApoB	*ANKRD31*	0.027	0.017	0.036	0.072	0.043	0.1	2.77E-12	-
rs2569550	ApoB	*LDLR*	0.024	0.013	0.036	0.041	0.029	0.054	2.84E-10	-
rs2618566	ApoB	*SNX5*	-0.003	-0.017	0.011	-0.03	-0.044	-0.016	1.45E-09	-

rs3747207	AST	*PNPLA3*	0.055	0.05	0.061	0.165	0.148	0.183	5.82E-208	4.02E-92
rs71633359	AST	*HSD17B13*	-0.035	-0.04	-0.03	-0.044	-0.051	-0.036	6.45E-49	6.44E-23
rs2862954	AST	*ERLIN1*	-0.02	-0.026	-0.014	-0.045	-0.051	-0.039	5.86E-38	7.53E-20
rs58542926	AST	*TM6SF2*	0.042	0.034	0.05	0.107	0.065	0.149	8.37E-37	1.63E-16
rs56278466	AST	*TMEM236*	0.024	0.016	0.031	0.044	0.036	0.052	1.50E-31	-
rs555045010	AST	*TMEM236*	0.023	0.018	0.028	0.041	0.03	0.051	7.23E-27	-
rs1260326	AST	*GCKR*	-0.026	-0.033	-0.018	-0.038	-0.045	-0.031	6.97E-23	4.24E-21
rs2954038	AST	*TRIB1*	-0.019	-0.028	-0.009	-0.033	-0.042	-0.024	1.14E-15	6.19E-10
rs4979371	AST	*AKNA*	-0.012	-0.018	-0.006	-0.028	-0.035	-0.021	4.51E-15	1.23E-08
rs7682289	AST	*ZNF827*	0.016	0.011	0.022	0.042	0.026	0.059	1.76E-13	1.44E-07
rs754465	AST	*DLG5*	0.012	0.005	0.018	0.026	0.019	0.033	5.37E-13	5.91E-07
rs4245267	AST	*NEDD4L*	-0.031	-0.044	-0.018	-0.04	-0.052	-0.027	6.83E-11	7.80E-06
rs2126259	AST	*PPP1R3B*	-0.004	-0.029	0.022	-0.025	-0.049	0	1.40E-10	3.27E-04
rs429358	AST	*APOE*	-0.018	-0.023	-0.013	-0.017	-0.032	-0.002	2.76E-10	6.99E-09
rs2701175	AST	*HNF1A*	-0.016	-0.024	-0.008	-0.026	-0.034	-0.018	4.41E-10	2.49E-04
rs10787429	AST	*GPAM*	-0.009	-0.019	0.001	-0.023	-0.033	-0.014	7.07E-10	2.82E-05
rs77924615	Creatinine	*PDILT*	-0.007	-0.009	-0.005	-0.014	-0.018	-0.01	1.23E-20	6.10E-03
rs1288775	Creatinine	*GATM*	0.005	0.004	0.007	0.01	0.006	0.013	1.08E-12	-
rs10254101	Creatinine	*PRKAG2*	0.005	0.003	0.006	0.01	0.006	0.013	1.40E-11	-
rs429358	CRP	*APOE*	-0.057	-0.06	-0.054	-0.095	-0.099	-0.09	6.72E-307	-
rs7553007	CRP	*CRP*	-0.039	-0.042	-0.036	-0.07	-0.074	-0.066	5.19E-207	7.56E-03
rs7310409	CRP	*HNF1A*	0.034	0.03	0.038	0.071	0.065	0.076	2.36E-206	-
rs61812598	CRP	*IL6R*	-0.021	-0.025	-0.018	-0.046	-0.05	-0.042	2.04E-85	-
rs17616063	CRP	*SALL1*	-0.032	-0.036	-0.028	-0.069	-0.08	-0.059	4.61E-57	-
rs13409371	CRP	*IL1F10*	0.014	0.01	0.017	0.041	0.035	0.046	3.87E-56	1.80E-10
rs1260326	CRP	*GCKR*	-0.02	-0.025	-0.015	-0.038	-0.043	-0.033	3.45E-55	-

rs111307268	CRP	*NLRP3*	0.011	0.008	0.015	0.027	0.021	0.032	4.79E-27	1.70E-06
rs2972558	CRP	*PVRL2*	-0.014	-0.02	-0.009	-0.026	-0.031	-0.021	2.36E-24	-
rs7012637	CRP	*PPP1R3B*	0.007	0.003	0.011	0.023	0.018	0.028	1.83E-22	-
rs2836883	CRP	*PSMG1*	-0.015	-0.018	-0.012	-0.019	-0.025	-0.013	4.36E-21	1.37E-05
rs3811452	CRP	*ATP8B2*	-0.013	-0.017	-0.01	-0.035	-0.043	-0.026	3.49E-20	-
rs4655537	CRP	*LEPR*	-0.011	-0.016	-0.006	-0.022	-0.027	-0.018	1.26E-19	6.41E-70
rs2246941	CRP	*LIPA*	0.011	0.008	0.014	0.021	0.015	0.027	2.13E-17	4.37E-08
rs11868378	CRP	*SOCS3*	0.002	-0.008	0.011	-0.012	-0.021	-0.004	3.04E-14	-
rs1037171	CRP	*RAB37|CD300LF*	0.01	0.004	0.016	0.02	0.014	0.026	2.63E-13	-
rs2269434	CRP	*MYBPC3*	-0.012	-0.015	-0.008	-0.015	-0.02	-0.01	2.94E-13	-
rs56189574	CRP	*MS4A6A*	-0.008	-0.012	-0.004	-0.018	-0.022	-0.013	5.94E-13	-
rs3027063	CRP	*DARC*	0.011	0.007	0.014	0.016	0.01	0.021	1.93E-12	1.76E-03
rs2393794	CRP	*SPPL3*	0.011	0.008	0.015	0.021	0.011	0.031	2.28E-12	-
rs728538	CRP	*SALL1*	0.009	0.005	0.013	0.031	0.019	0.042	3.13E-12	-
rs2700938	CRP	*EEPD1*	0.008	0.005	0.012	0.017	0.012	0.022	1.10E-11	-
rs72959041	CRP	*RSPO3*	-0.018	-0.023	-0.013	-0.035	-0.061	-0.009	1.28E-11	-
rs11983782	CRP	*TOMM7*	-0.008	-0.012	-0.005	-0.016	-0.02	-0.012	2.96E-11	-
rs61781391	CRP	*HEYL*	0.011	0.008	0.015	0.015	0.007	0.023	2.97E-11	-
rs7828742	CRP	*TRPS1*	0.011	0.006	0.015	0.017	0.012	0.021	6.15E-11	-
rs2283371	CRP	*RGS6*	0.008	0.004	0.011	0.018	0.012	0.023	7.19E-11	-
rs2280406	CRP	*MST1R*	0.007	0.003	0.011	0.015	0.011	0.02	3.00E-10	-
rs340005	CRP	*RORA*	0.009	0.004	0.014	0.016	0.011	0.021	4.02E-10	-
rs10783792	CRP	*RBMS2*	0.004	-0.006	0.013	0.015	0.005	0.025	6.92E-10	-
rs75777234	CRP	*PRKG1*	0.01	0.004	0.015	0.092	0.043	0.14	1.17E-09	5.73E-04
rs10410651	CRP	*PVR*	0.007	0.004	0.011	0.019	0.012	0.026	1.25E-09	-
rs11668719	CRP	*LRRC25*	0.005	0.001	0.009	0.015	0.01	0.02	1.31E-09	-

rs11145763	CRP	*CARD9*	0.008	0.004	0.011	0.015	0.01	0.02	1.60E-09	1.33E-03
rs67567111	Cystatin C	*CST3*	-0.004	-0.005	-0.004	-0.007	-0.007	-0.006	2.84E-68	-
rs77924615	Cystatin C	*PDILT*	-0.004	-0.004	-0.003	-0.005	-0.007	-0.004	2.49E-20	6.72E-05
rs73102387	Cystatin C	*CST3*	0.002	0.002	0.003	0.005	0.002	0.007	3.33E-10	-
rs7310615	Cystatin C	*SH2B3*	-0.002	-0.003	-0.001	-0.003	-0.004	-0.003	1.03E-09	-
rs62192912	Direct BR	*ATG16L1*	-0.189	-0.199	-0.18	-0.298	-0.311	-0.285	0	1.00E-300
rs2741047	Direct BR	*UGT1A8|UGT1A10|UGT1A9*	0.133	0.124	0.141	0.697	0.682	0.712	0	3.27E-19
rs6712540	Direct BR	*TRPM8*	-0.13	-0.14	-0.12	-0.186	-0.22	-0.151	1.68E-113	9.48E-70
rs11045864	Direct BR	*SLCO1B1*	0.083	0.072	0.094	0.185	0.151	0.22	1.87E-74	3.25E-20
rs474242	Direct BR	*MROH2A*	-0.059	-0.07	-0.048	-0.105	-0.118	-0.092	1.93E-52	2.04E-34
rs9750891	Direct BR	*INPP5D*	0.089	0.072	0.107	0.236	0.124	0.348	5.46E-32	7.25E-27
rs76820150	Direct BR	*SLCO1C1*	-0.053	-0.068	-0.038	-0.082	-0.096	-0.067	1.01E-26	7.11E-07
rs10761737	Direct BR	*JMJD1C*	-0.025	-0.036	-0.014	-0.066	-0.079	-0.053	2.24E-19	3.21E-11
rs1070232	Direct BR	*STAG1*	-0.019	-0.039	0.001	-0.056	-0.075	-0.036	1.55E-15	1.23E-07
rs10774624	Direct BR	*SH2B3*	-0.036	-0.048	-0.024	-0.055	-0.069	-0.042	6.15E-15	3.73E-05
rs2068888	Direct BR	*CYP26A1*	0.033	0.022	0.044	0.052	0.038	0.066	2.20E-14	1.79E-05
rs80284120	Direct BR	*NGEF*	-0.035	-0.045	-0.024	-0.05	-0.075	-0.026	6.62E-12	4.95E-09
rs450244	Direct BR	*SLC22A18*	-0.158	-0.217	-0.098	-0.18	-0.237	-0.122	2.30E-11	8.16E-06
rs113041162	Direct BR	*HK1*	0.03	0.019	0.041	0.077	0.041	0.113	1.12E-10	4.29E-05
rs7412	Direct BR	*APOE*	0.038	0.025	0.052	0.107	0.039	0.174	7.93E-10	2.49E-07
rs6479336	Direct BR	*AUH*	0.028	0.017	0.038	0.059	0.031	0.088	1.46E-09	1.28E-05
rs2006227	GGT	*SNRPD3|GGT1*	0.035	0.032	0.038	0.08	0.074	0.087	1.34E-283	4.83E-58
rs4835265	GGT	*ZNF827*	0.039	0.035	0.042	0.085	0.071	0.099	5.94E-169	7.53E-107
rs4503880	GGT	*NEDD4L*	-0.047	-0.056	-0.038	-0.074	-0.082	-0.067	5.92E-129	6.48E-85
rs1497406	GGT	*ARHGEF19*	0.02	0.016	0.023	0.046	0.042	0.05	4.20E-124	3.41E-72
rs11624282	GGT	*EXOC3L4*	0.031	0.028	0.034	0.047	0.04	0.054	5.87E-124	5.29E-42

rs754466	GGT	*DLG5*	0.02	0.017	0.023	0.06	0.052	0.068	1.10E-103	1.92E-82
rs10075805	GGT	*CDH6*	-0.025	-0.028	-0.022	-0.035	-0.04	-0.031	5.32E-86	1.05E-69
rs10908456	GGT	*EFNA1*	-0.02	-0.023	-0.016	-0.036	-0.039	-0.033	4.91E-76	3.67E-50
rs28650012	GGT	*DYNLRB2*	-0.021	-0.027	-0.015	-0.036	-0.041	-0.031	1.29E-49	4.89E-34
rs35645198	GGT	*MICAL3*	-0.017	-0.02	-0.014	-0.03	-0.034	-0.025	5.80E-47	3.00E-20
rs7310409	GGT	*HNF1A*	0.015	0.011	0.018	0.029	0.025	0.033	1.30E-46	-
rs339969	GGT	*RORA*	0.012	0.008	0.016	0.027	0.023	0.032	5.18E-44	2.26E-14
rs12190285	GGT	*SOX4*	-0.013	-0.016	-0.011	-0.028	-0.032	-0.023	3.59E-37	5.70E-23
rs1260326	GGT	*GCKR*	-0.01	-0.014	-0.006	-0.023	-0.027	-0.019	5.85E-33	2.96E-12
rs6879279	GGT	*EFNA5*	-0.017	-0.019	-0.014	-0.028	-0.035	-0.021	1.48E-32	5.19E-23
rs601338	GGT	*FUT2*	0.006	0.002	0.009	0.022	0.018	0.026	6.73E-31	1.17E-15
rs28601761	GGT	*TRIB1*	-0.012	-0.015	-0.009	-0.022	-0.026	-0.019	3.26E-29	1.04E-10
rs9913936	GGT	*SOX9*	0.008	0.004	0.013	0.022	0.017	0.026	7.57E-29	7.66E-29
rs35596292	GGT	*MYO1B*	0.013	0.01	0.016	0.023	0.017	0.028	2.74E-28	3.63E-17
rs3811468	GGT	*LPHN2*	-0.011	-0.014	-0.008	-0.025	-0.03	-0.02	1.15E-26	6.29E-24
rs7780562	GGT	*NFE2L3*	0.005	-0.002	0.011	0.019	0.012	0.026	5.08E-25	1.38E-14
rs1778793	GGT	*PDX1*	-0.01	-0.014	-0.007	-0.02	-0.024	-0.017	1.15E-23	2.20E-21
rs5402	GGT	*SLC2A2*	0.017	0.013	0.02	0.021	0.008	0.034	2.74E-23	2.11E-07
rs4795218	GGT	*HNF1B*	0.01	0.007	0.013	0.029	0.021	0.037	9.17E-23	1.80E-13
rs4074793	GGT	*ITGA1*	0.018	0.014	0.023	0.054	0.029	0.078	1.01E-22	2.49E-09
rs900776	GGT	*DMTN*	0.013	0.01	0.016	0.025	0.016	0.034	2.27E-21	1.69E-12
rs7247349	GGT	*PEPD*	-0.009	-0.013	-0.006	-0.019	-0.022	-0.015	6.42E-21	5.08E-11
rs10424333	GGT	*RHPN2*	0.016	0.012	0.02	0.036	0.017	0.055	2.72E-19	1.32E-08
rs33951980	GGT	*MLXIPL*	-0.012	-0.015	-0.009	-0.028	-0.036	-0.02	3.64E-19	5.09E-04
rs62375243	GGT	*HSPA4*	-0.008	-0.011	-0.006	-0.022	-0.026	-0.017	3.33E-18	5.17E-09
rs115478735	GGT	*SURF6*	0.013	0.01	0.016	0.013	0.005	0.022	5.10E-18	-

rs7551732	GGT	*PKN2*	0.01	0.007	0.014	0.018	0.014	0.022	7.64E-18	5.97E-10
rs35149321	GGT	*CD276*	0.008	0.005	0.011	0.017	0.013	0.021	1.84E-17	6.92E-07
rs17358295	GGT	*EHF*	-0.014	-0.017	-0.011	-0.025	-0.036	-0.015	2.26E-17	9.76E-11
rs12979186	GGT	*MAP1S*	0.008	0.004	0.011	0.017	0.013	0.021	2.60E-17	2.61E-12
rs2641352	GGT	*ADAM30*	0.011	0.007	0.014	0.046	0.029	0.064	3.93E-17	3.74E-11
rs4973550	GGT	*EFHD1*	0.011	0.007	0.015	0.018	0.013	0.022	7.91E-17	1.54E-15
rs4822983	GGT	*CHEK2*	-0.01	-0.013	-0.007	-0.016	-0.02	-0.012	1.94E-16	2.10E-08
rs6855886	GGT	*KLB*	-0.009	-0.012	-0.006	-0.017	-0.021	-0.013	2.27E-16	5.06E-07
rs1649079	GGT	*BICC1*	-0.009	-0.012	-0.005	-0.017	-0.02	-0.013	2.88E-16	2.20E-11
rs7314285	GGT	*CUX2*	0.017	0.013	0.022	0.025	0.001	0.048	7.48E-16	8.28E-09
rs4921915	GGT	*NAT2*	-0.011	-0.018	-0.004	-0.021	-0.027	-0.014	1.18E-15	1.70E-12
rs4242221	GGT	*TENM2*	-0.01	-0.016	-0.005	-0.019	-0.024	-0.014	1.31E-15	3.06E-07
rs625899	GGT	*MLIP*	-0.012	-0.017	-0.007	-0.019	-0.024	-0.014	1.63E-15	2.17E-10
rs3861491	GGT	*C14orf182*	0.005	0.001	0.009	0.015	0.011	0.02	2.06E-15	1.50E-10
rs72655725	GGT	*COL4A1*	-0.01	-0.012	-0.007	-0.02	-0.026	-0.014	3.08E-15	1.68E-07
rs11022131	GGT	*DKK3*	-0.007	-0.01	-0.004	-0.02	-0.025	-0.015	3.80E-15	1.35E-07
rs93075	GGT	*SEPT9*	0.013	0.008	0.017	0.019	0.014	0.024	5.71E-15	9.12E-06
rs10994838	GGT	*A1CF*	0.007	0.004	0.01	0.017	0.012	0.021	5.85E-15	2.55E-09
rs17145884	GGT	*AHNAK*	-0.01	-0.013	-0.007	-0.021	-0.028	-0.014	1.28E-14	2.13E-05
rs10936201	GGT	*SMC4*	0.007	0.004	0.01	0.02	0.014	0.027	2.71E-14	9.31E-09
rs2904889	GGT	*NDUFAF6|TP53INP1*	0.009	0.006	0.012	0.026	0.016	0.036	3.43E-14	3.25E-11
rs9456946	GGT	*SYNJ2*	0.009	0.006	0.012	0.019	0.012	0.026	6.18E-14	4.31E-06
rs72840109	GGT	*DNMBP*	-0.016	-0.02	-0.012	-0.036	-0.054	-0.018	6.44E-14	1.28E-07
rs77666400	GGT	*S1PR1*	0.008	0.005	0.011	0.019	0.013	0.025	6.61E-14	3.22E-07
rs4850046	GGT	*RPS7*	-0.028	-0.04	-0.016	-0.036	-0.048	-0.025	6.67E-14	5.15E-06
rs62030794	GGT	*TMEM8A*	0.016	0.011	0.02	0.027	0.003	0.051	1.92E-13	6.91E-11

rs55931203	GGT	*BPTF*	0.011	0.008	0.014	0.012	0.005	0.02	2.48E-13	9.79E-08
rs10104003	GGT	*SOX17*	0.007	0.004	0.01	0.023	0.015	0.03	4.33E-13	2.29E-06
rs123698	GGT	*PTBP1*	0.009	0.005	0.013	0.015	0.011	0.019	4.83E-13	7.26E-11
rs13026184	GGT	*SERTAD2*	-0.008	-0.011	-0.005	-0.017	-0.022	-0.012	5.14E-13	1.86E-08
rs73220641	GGT	*KLF5*	-0.01	-0.013	-0.008	-0.011	-0.015	-0.006	7.14E-13	1.99E-08
rs12928392	GGT	*MMP15*	-0.009	-0.012	-0.006	-0.016	-0.022	-0.011	7.60E-13	2.08E-05
rs2024924	GGT	*MACROD2*	0.007	0.002	0.011	0.015	0.01	0.02	9.49E-13	5.45E-08
rs636672	GGT	*TENM4*	0.009	0.006	0.012	0.013	0.008	0.018	1.69E-12	1.52E-10
rs11114042	GGT	*CORO1C*	-0.006	-0.009	-0.003	-0.015	-0.018	-0.011	1.79E-12	3.56E-08
rs5757252	GGT	*GTPBP1*	0.007	0.004	0.01	0.016	0.011	0.02	2.31E-12	1.02E-06
rs909537	GGT	*ASAP3*	0.01	-0.003	0.022	0.021	0.008	0.034	2.56E-12	1.67E-07
rs11644920	GGT	*LITAF*	0.007	0.004	0.01	0.016	0.011	0.021	3.72E-12	4.54E-04
rs7260785	GGT	*DDRGK1*	0.008	0.005	0.011	0.023	0.014	0.033	4.06E-12	2.34E-08
rs4816700	GGT	*DSCAM*	-0.01	-0.012	-0.007	-0.012	-0.017	-0.006	4.11E-12	9.20E-08
rs38849	GGT	*MET*	-0.007	-0.013	-0.001	-0.016	-0.021	-0.01	1.07E-11	1.70E-03
rs62241682	GGT	*RBMS3*	0.008	0.005	0.011	0.016	0.01	0.022	1.28E-11	1.74E-06
rs10450314	GGT	*PARD3*	-0.009	-0.011	-0.006	-0.013	-0.018	-0.008	1.49E-11	2.68E-06
rs13089831	GGT	*TM4SF4*	0.005	-0.001	0.012	0.014	0.008	0.021	1.54E-11	1.08E-09
rs11709077	GGT	*PPARG*	-0.009	-0.013	-0.006	-0.024	-0.033	-0.015	2.21E-11	1.17E-06
rs67588707	GGT	*SLCO1B3|SLCO1B7*	0.01	0.007	0.013	0.02	0.009	0.031	2.71E-11	1.32E-05
rs56013261	GGT	*SETD2*	-0.006	-0.009	-0.003	-0.014	-0.018	-0.01	2.95E-11	1.05E-05
rs273506	GGT	*MAST3*	0.006	0.003	0.009	0.014	0.01	0.018	4.36E-11	9.26E-04
rs11635675	GGT	*USP3*	0.006	0.003	0.009	0.015	0.01	0.02	5.24E-11	2.24E-07
rs9980195	GGT	*PTTG1IP*	-0.007	-0.011	-0.004	-0.013	-0.017	-0.01	6.37E-11	2.07E-08
rs11543269	GGT	*ATP8B1*	-0.011	-0.014	-0.008	-0.013	-0.023	-0.002	8.93E-11	7.82E-06
rs6072249	GGT	*TOP1*	0.006	0.003	0.01	0.013	0.009	0.017	9.46E-11	-

rs112375685	GGT	*LPP*	0.008	0.004	0.011	0.013	0.009	0.017	1.02E-10	6.21E-06
rs9379084	GGT	*RREB1*	0.012	0.008	0.015	0.009	-0.002	0.021	1.08E-10	6.37E-08
rs11753995	GGT	*SLC22A1*	-0.007	-0.01	-0.005	-0.02	-0.027	-0.014	1.18E-10	2.78E-04
rs73620883	GGT	*TBC1D13*	-0.01	-0.013	-0.007	-0.014	-0.024	-0.004	2.12E-10	2.27E-04
rs530939	GGT	*TENM4*	-0.008	-0.012	-0.004	-0.014	-0.018	-0.01	2.47E-10	9.71E-08
rs2811290	GGT	*C1orf220*	-0.004	-0.007	0	-0.013	-0.016	-0.009	2.59E-10	3.98E-05
rs11114664	GGT	*ACSS3*	-0.006	-0.012	-0.001	-0.014	-0.019	-0.009	2.59E-10	7.34E-05
rs2941465	GGT	*HNF4G*	-0.005	-0.009	-0.002	-0.013	-0.017	-0.009	4.02E-10	8.83E-06
rs13112099	GGT	*UGT2B15*	-0.007	-0.011	-0.004	-0.013	-0.016	-0.009	4.34E-10	6.48E-03
rs2059988	GGT	*TM4SF1*	0.004	0	0.009	0.012	0.007	0.017	4.46E-10	3.59E-08
rs12243124	GGT	*FFAR4*	-0.008	-0.011	-0.006	-0.013	-0.019	-0.007	5.74E-10	1.49E-06
rs3102990	GGT	*EZR*	-0.007	-0.011	-0.003	-0.013	-0.017	-0.009	6.19E-10	3.65E-05
rs34346558	GGT	*PROSER2*	0.009	0.006	0.012	0.012	0.005	0.019	1.02E-09	1.12E-04
rs114484444	GGT	*TNFSF10*	0.016	0.01	0.021	0.013	-0.015	0.041	1.14E-09	4.28E-06
rs2981451	GGT	*FGFR2*	-0.008	-0.012	-0.005	-0.012	-0.016	-0.008	1.20E-09	3.88E-05
rs112038040	GGT	*ARIH1*	0.004	0	0.008	0.012	0.007	0.016	1.25E-09	5.88E-04
rs12480190	GGT	*ZBTB46*	0.009	0.006	0.012	0.022	0.009	0.034	1.58E-09	1.43E-05
rs4911256	GGT	*DNMT3B*	0.007	0.004	0.011	0.012	0.008	0.016	1.62E-09	7.80E-06
rs35198068	Glucose	*TCF7L2*	0.015	0.012	0.017	0.035	0.03	0.04	3.36E-81	1.54E-17
rs7756992	Glucose	*CDKAL1*	0.006	0.004	0.008	0.021	0.016	0.027	2.88E-23	2.36E-09
rs11187138	Glucose	*HHEX*	-0.006	-0.009	-0.004	-0.013	-0.016	-0.01	5.79E-17	6.23E-10
rs7928810	Glucose	*NCR3LG1*	-0.009	-0.012	-0.006	-0.014	-0.017	-0.011	7.38E-17	8.82E-11
rs77684335	Glucose	*GPSM1*	0.009	0.005	0.012	0.015	0.011	0.018	2.63E-16	1.21E-09
rs11558471	Glucose	*SLC30A8*	-0.007	-0.009	-0.005	-0.013	-0.016	-0.01	4.99E-16	6.20E-03
rs113042771	Glucose	*DGKB*	0.005	0.003	0.008	0.011	0.008	0.014	1.17E-13	3.45E-03
rs35658696	Glucose	*PAM*	0.011	0.007	0.015	0.051	0.014	0.088	1.18E-12	2.66E-06

rs10811660	Glucose	*CDKN2B*	-0.008	-0.01	-0.006	-0.01	-0.015	-0.005	1.69E-12	4.69E-03
rs11651755	Glucose	*HNF1B*	-0.007	-0.01	-0.005	-0.011	-0.013	-0.008	3.09E-12	-
rs2237895	Glucose	*KCNQ1*	0.005	0.003	0.008	0.01	0.007	0.013	4.32E-11	1.83E-03
rs74889068	Glucose	*QPCTL*	0.006	0.003	0.008	0.021	0.012	0.031	7.25E-11	9.63E-10
rs2396316	Glucose	*IRS1*	0.008	0.005	0.011	0.011	0.008	0.014	7.71E-11	7.80E-05
rs9859406	Glucose	*IGF2BP2*	0.006	0.003	0.008	0.011	0.007	0.015	1.13E-10	3.79E-03
rs11603349	Glucose	*ARAP1*	-0.007	-0.009	-0.005	-0.012	-0.018	-0.007	2.57E-10	2.88E-03
rs491443	Glucose	*SPC25*	-0.005	-0.007	-0.002	-0.01	-0.012	-0.007	4.83E-10	2.54E-20
rs849138	Glucose	*JAZF1*	-0.005	-0.008	-0.003	-0.01	-0.012	-0.007	6.56E-10	-
rs4234731	Glucose	*WFS1*	0.003	0	0.006	0.009	0.006	0.012	6.86E-10	7.99E-03
rs7903146	HbA1C	*TCF7L2*	0.02	0.017	0.022	0.048	0.042	0.054	4.00E-107	2.42E-17
rs9368222	HbA1C	*CDKAL1*	0.005	0.003	0.008	0.023	0.017	0.029	1.35E-18	8.56E-05
rs5015480	HbA1C	*HHEX*	-0.006	-0.009	-0.004	-0.015	-0.018	-0.012	1.28E-16	6.22E-06
rs1421085	HbA1C	*FTO*	0.007	0.004	0.01	0.015	0.011	0.019	1.50E-15	-
rs10823346	HbA1C	*HK1*	0.005	0.002	0.007	0.021	0.015	0.027	2.45E-15	1.27E-73
rs1470580	HbA1C	*IGF2BP2*	0.007	0.005	0.01	0.016	0.011	0.02	3.99E-15	-
rs10965250	HbA1C	*CDKN2B*	-0.009	-0.012	-0.007	-0.013	-0.019	-0.007	7.98E-13	-
rs11263763	HbA1C	*HNF1B*	-0.008	-0.011	-0.005	-0.013	-0.016	-0.009	3.68E-12	-
rs703978	HbA1C	*ZMIZ1*	-0.008	-0.012	-0.005	-0.013	-0.017	-0.01	7.68E-12	4.44E-03
rs849134	HbA1C	*JAZF1*	-0.009	-0.011	-0.006	-0.012	-0.015	-0.008	2.50E-11	-
rs881796	HbA1C	*WFS1*	0.003	-0.001	0.006	0.011	0.007	0.014	9.25E-11	-
rs5398	HbA1C	*SLC2A2*	-0.005	-0.008	-0.003	-0.013	-0.017	-0.009	3.62E-10	-
rs2972144	HbA1C	*IRS1*	0.005	0.002	0.009	0.012	0.008	0.015	4.97E-10	4.03E-04
rs7895525	HbA1C	*CDC123*	0.007	0.004	0.01	0.012	0.006	0.018	6.52E-10	-
rs72964564	HbA1C	*ADCY5*	-0.006	-0.009	-0.004	-0.013	-0.017	-0.008	6.59E-10	2.50E-03
rs74567345	HbA1C	*PAM*	0.013	0.008	0.017	0.03	0	0.06	9.48E-10	-

rs35859536	HbA1C	*SLC30A8*	-0.004	-0.007	-0.002	-0.013	-0.017	-0.009	1.24E-09	-
rs247616	HDL	*CETP*	0.085	0.076	0.093	0.214	0.198	0.231	6.98E-217	8.69E-06
rs1077835	HDL	*ALDH1A2|LIPC*	0.047	0.038	0.056	0.141	0.118	0.163	1.21E-55	9.44E-04
rs2740488	HDL	*ABCA1*	-0.054	-0.063	-0.046	-0.092	-0.108	-0.077	1.80E-48	5.33E-13
rs1065853	HDL	*APOE*	0.081	0.068	0.094	0.15	0.087	0.212	1.16E-44	2.44E-22
rs112180569	HDL	*PCIF1*	-0.053	-0.061	-0.044	-0.101	-0.121	-0.081	5.21E-41	2.89E-14
rs2043085	HDL	*ALDH1A2*	-0.039	-0.052	-0.027	-0.068	-0.08	-0.055	2.25E-25	-
rs1943973	HDL	*LIPG*	0.041	0.014	0.069	0.086	0.058	0.113	3.19E-25	2.70E-03
rs4922118	HDL	*LPL*	0.025	0.016	0.034	0.059	0.046	0.071	1.11E-20	-
rs174560	HDL	*FADS2|FADS1*	-0.019	-0.028	-0.011	-0.065	-0.079	-0.052	7.27E-19	-
rs10779836	HDL	*GALNT2*	0.046	0.025	0.067	0.076	0.056	0.097	1.23E-18	-
rs686030	HDL	*TTC39B*	0.012	-0.017	0.041	0.052	0.022	0.081	2.24E-16	5.70E-07
rs11789603	HDL	*ABCA1*	0.04	0.029	0.051	0.096	0.052	0.14	2.41E-16	-
rs1132899	HDL	*APOC4|APOC4-APOC2*	0.021	0.011	0.031	0.05	0.038	0.062	6.70E-16	7.16E-04
rs13338063	HDL	*NUP93*	0.028	0.019	0.038	0.048	0.035	0.06	3.45E-15	-
rs964184	HDL	*APOA5*	0.039	0.009	0.069	0.073	0.043	0.103	1.47E-14	-
rs6544366	HDL	*APOB*	0.029	0.02	0.038	0.05	0.03	0.07	6.77E-13	-
rs2954031	HDL	*TRIB1*	0.016	0.006	0.025	0.045	0.033	0.058	1.22E-12	1.54E-03
rs7105282	HDL	*NUP160*	-0.02	-0.029	-0.011	-0.048	-0.061	-0.035	3.18E-12	-
rs4490856	HDL	*LPL*	0.02	0.009	0.031	0.043	0.03	0.055	1.05E-11	-
rs13107325	HDL	*SLC39A8*	-0.035	-0.046	-0.023	-0.098	-0.149	-0.047	2.81E-10	-
rs112310696	HDL	*DUS2*	0.031	0.021	0.042	0.042	0.008	0.077	1.45E-09	-
rs344352	IGF-1	*HAGH*	0.056	0.037	0.075	0.094	0.075	0.113	1.91E-25	-
rs700753	IGF-1	*TNS3*	0.046	0.03	0.063	0.084	0.067	0.101	1.16E-24	-
rs1260326	IGF-1	*GCKR*	0.041	0.026	0.055	0.066	0.051	0.082	4.30E-16	-
rs112166936	IGF-1	*CENPW*	0.018	0.006	0.03	0.051	0.035	0.066	2.02E-10	-

rs35766	IGF-1	*IGF1*	-0.032	-0.068	0.005	-0.065	-0.1	-0.031	9.51E-10	-
rs1065853	LDL	*APOE*	-0.104	-0.111	-0.097	0.026	-0.015	0.068	9.75E-141	4.48E-254
rs10402112	LDL	*LDLR*	-0.084	-0.091	-0.077	-0.151	-0.171	-0.13	4.68E-118	-
rs12740374	LDL	*CELSR2*	-0.05	-0.057	-0.043	-0.093	-0.107	-0.079	5.78E-64	-
rs581411	LDL	*APOB*	0.031	0.013	0.049	0.082	0.064	0.101	9.44E-53	2.98E-04
rs28601761	LDL	*TRIB1*	-0.036	-0.044	-0.029	-0.061	-0.07	-0.052	1.94E-37	-
rs4299376	LDL	*ABCG8*	-0.034	-0.046	-0.022	-0.059	-0.071	-0.048	2.30E-26	-
rs964184	LDL	*APOA5*	-0.034	-0.062	-0.005	-0.07	-0.097	-0.044	1.38E-22	-
rs2738447	LDL	*LDLR*	0.024	0.015	0.034	0.045	0.035	0.055	8.04E-19	-
rs10045497	LDL	*HMGCR*	0.017	0.009	0.024	0.047	0.036	0.058	8.38E-18	-
rs1168114	LDL	*DOCK7*	0.016	0.005	0.027	0.04	0.029	0.051	1.38E-15	-
rs1535	LDL	*FADS2*	-0.012	-0.019	-0.005	-0.044	-0.054	-0.033	1.60E-14	7.16E-03
rs3208305	LDL	*LPL*	-0.022	-0.029	-0.015	-0.037	-0.049	-0.025	2.26E-13	-
rs2495477	LDL	*PCSK9*	-0.018	-0.026	-0.011	-0.035	-0.045	-0.025	7.26E-12	-
rs4704727	LDL	*TIMD4*	0.016	0.005	0.027	0.035	0.024	0.046	1.17E-11	-
rs58542926	LDL	*TM6SF2*	-0.032	-0.042	-0.023	-0.051	-0.091	-0.012	2.29E-11	3.49E-06
rs406315	LDL	*PVRL2*	-0.018	-0.028	-0.007	-0.035	-0.045	-0.024	5.31E-11	-
rs113120414	LDL	*ABCA8*	0.033	0.022	0.045	0.123	0.05	0.196	6.23E-11	-
rs2000999	LDL	*TXNL4B|HPR*	0.015	0.008	0.023	0.056	0.036	0.076	9.39E-11	-
rs2618566	LDL	*SNX5*	-0.009	-0.02	0.003	-0.03	-0.041	-0.019	9.56E-11	-
rs532436	LDL	*SURF6*	0.023	0.015	0.03	0.037	0.017	0.056	1.01E-10	2.48E-03
rs2072183	LDL	*NPC1L1*	0.015	0.008	0.022	0.049	0.032	0.066	1.06E-10	-
rs7746081	LDL	*MYLIP*	-0.014	-0.021	-0.007	-0.037	-0.048	-0.025	7.73E-10	9.38E-03
rs12209724	LipoA	*MAS1*	0.26	0.244	0.275	0.516	0.468	0.564	0	3.23E-146
rs688359	LipoA	*IGF2R*	-0.238	-0.258	-0.219	-0.513	-0.529	-0.497	0	1.62E-246
rs402219	LipoA	*SLC22A3*	-0.218	-0.23	-0.206	-0.426	-0.443	-0.408	0	1.43E-23

rs1247295	LipoA	*MAP3K4*	-0.264	-0.278	-0.25	-0.489	-0.503	-0.475	0	-
rs1247336	LipoA	*MAP3K4*	0.123	0.11	0.137	0.255	0.225	0.284	1.85E-127	9.56E-99
rs9458188	LipoA	*AGPAT4*	-0.073	-0.097	-0.049	-0.159	-0.181	-0.136	2.52E-60	2.78E-37
rs911844	LipoA	*SOD2*	-0.05	-0.064	-0.037	-0.112	-0.13	-0.095	4.77E-34	5.83E-13
rs687183	LipoA	*HS3ST3B1*	-0.033	-0.046	-0.02	-0.064	-0.083	-0.045	2.82E-11	2.29E-04
rs66987859	LipoA	*MTRNR2L12*	0.049	0.033	0.065	0.051	-0.002	0.103	9.49E-10	2.48E-05
rs1799941	SHBG	*SHBG*	0.154	0.142	0.167	0.334	0.307	0.361	6.38E-240	-
rs113056032	SHBG	*ZNF652*	0.093	0.076	0.11	0.215	0.134	0.295	2.96E-34	-
rs56332871	SHBG	*NR2F2*	0.048	0.035	0.06	0.111	0.086	0.135	2.34E-26	-
rs8067286	SHBG	*NPEPPS*	0.037	0.023	0.051	0.081	0.064	0.098	1.48E-20	-
rs10822145	SHBG	*JMJD1C*	0.036	0.022	0.049	0.081	0.064	0.098	2.80E-20	-
rs13108218	SHBG	*HGFAC*	-0.037	-0.055	-0.019	-0.081	-0.099	-0.063	3.67E-20	-
rs13232861	SHBG	*BRI3|BAIAP2L1*	0.016	-0.016	0.048	0.073	0.041	0.105	1.84E-19	-
rs2537856	SHBG	*ZNF554*	0.05	0.038	0.063	0.073	0.05	0.096	6.83E-19	-
rs7979473	SHBG	*HNF1A*	0.044	0.027	0.062	0.072	0.054	0.09	1.74E-14	-
rs62062620	SHBG	*DNAH2*	0.032	0.018	0.045	0.067	0.051	0.084	2.39E-14	-
rs10864086	SHBG	*PROX1*	-0.002	-0.027	0.024	-0.046	-0.07	-0.021	2.06E-12	-
rs7149605	SHBG	*SERPINA1*	0.049	0.033	0.065	0.122	0.047	0.196	2.87E-11	-
rs61557287	SHBG	*ZBTB10*	0.043	0.029	0.056	0.07	0.03	0.109	3.61E-11	-
rs687339	SHBG	*MSL2*	0.004	-0.023	0.032	-0.036	-0.063	-0.009	5.09E-10	-
rs1065853	TC	*APOE*	-0.089	-0.096	-0.082	0.096	0.047	0.146	1.63E-114	7.51E-95
rs73015020	TC	*LDLR*	-0.073	-0.08	-0.066	-0.125	-0.145	-0.105	2.83E-96	1.13E-03
rs629301	TC	*CELSR2*	0.035	0.021	0.049	0.079	0.064	0.094	7.07E-54	-
rs28601761	TC	*TRIB1*	-0.038	-0.045	-0.031	-0.062	-0.071	-0.053	1.16E-43	2.15E-03
rs581411	TC	*APOB*	0.024	0.007	0.041	0.068	0.051	0.086	1.28E-43	5.07E-05
rs964184	TC	*APOA5*	-0.054	-0.082	-0.027	-0.1	-0.125	-0.074	4.76E-40	1.17E-06

rs4299376	TC	*ABCG8*	-0.026	-0.037	-0.015	-0.048	-0.058	-0.037	8.31E-20	-
rs2131925	TC	*DOCK7*	0.019	0.009	0.029	0.042	0.031	0.052	4.80E-18	-
rs3208305	TC	*LPL*	-0.025	-0.031	-0.018	-0.041	-0.052	-0.03	8.73E-18	3.07E-07
rs2569550	TC	*LDLR*	0.024	0.015	0.033	0.04	0.031	0.05	3.09E-16	-
rs58542926	TC	*TM6SF2*	-0.032	-0.041	-0.024	-0.057	-0.094	-0.02	8.38E-13	1.68E-03
rs10045497	TC	*HMGCR*	0.011	0.004	0.018	0.037	0.027	0.047	2.31E-12	-
rs1535	TC	*FADS2*	-0.009	-0.016	-0.002	-0.037	-0.047	-0.028	7.37E-12	7.32E-04
rs35853021	TC	*ALDH1A2*	0.014	0.007	0.021	0.036	0.025	0.046	1.65E-11	-
rs406315	TC	*PVRL2*	-0.017	-0.027	-0.007	-0.033	-0.043	-0.023	3.58E-11	-
rs140798831	TC	*APOB*	-0.017	-0.027	-0.006	-0.033	-0.043	-0.023	7.59E-11	-
rs780093	TC	*GCKR*	-0.017	-0.026	-0.007	-0.032	-0.042	-0.023	8.92E-11	-
rs2000999	TC	*TXNL4B|HPR*	0.013	0.006	0.02	0.052	0.033	0.071	6.37E-10	-
rs1799941	Testosterone	*SHBG*	0.004	0.003	0.005	0.013	0.011	0.014	1.75E-98	-
rs10822145	Testosterone	*JMJD1C*	0.001	0.001	0.002	0.004	0.003	0.005	7.93E-21	-
rs72798735	Testosterone	*YIPF4*	0.003	0.002	0.004	0.008	0.001	0.015	1.52E-12	-
rs964184	TG	*APOA5*	-0.296	-0.34	-0.252	-0.511	-0.547	-0.476	0	7.96E-77
rs17482753	TG	*LPL*	-0.145	-0.154	-0.137	-0.251	-0.275	-0.227	1.47E-221	8.10E-33
rs1260326	TG	*GCKR*	-0.087	-0.099	-0.075	-0.168	-0.18	-0.157	8.58E-179	6.16E-38
rs28601761	TG	*TRIB1*	-0.081	-0.09	-0.072	-0.139	-0.15	-0.129	1.24E-137	9.16E-19
rs438811	TG	*APOC1*	0.079	0.071	0.088	0.173	0.152	0.194	1.89E-129	4.86E-26
rs71556736	TG	*MLXIPL*	-0.089	-0.097	-0.08	-0.187	-0.209	-0.165	2.33E-107	1.67E-15
rs6657050	TG	*DOCK7*	-0.053	-0.061	-0.044	-0.121	-0.132	-0.11	3.50E-87	1.48E-11
rs102275	TG	*TMEM258*	0.046	0.038	0.055	0.106	0.092	0.12	2.77E-63	2.13E-21
rs5112	TG	*APOC1*	0.04	0.03	0.05	0.082	0.07	0.093	8.59E-45	2.05E-05
rs58542926	TG	*TM6SF2*	-0.073	-0.083	-0.063	-0.158	-0.195	-0.121	1.06E-44	1.26E-08
rs998584	TG	*VEGFA*	0.034	0.025	0.043	0.073	0.062	0.085	1.72E-36	5.18E-15

rs673548	TG	*APOB*	-0.044	-0.053	-0.036	-0.079	-0.097	-0.062	3.67E-33	-
rs12443634	TG	*CMIP*	-0.031	-0.047	-0.015	-0.07	-0.085	-0.055	7.22E-30	9.56E-07
rs2222018	TG	*IRS1*	0.033	0.021	0.045	0.066	0.053	0.079	1.19E-27	-
rs738408	TG	*PNPLA3*	-0.028	-0.037	-0.02	-0.082	-0.099	-0.066	1.06E-23	2.57E-22
rs34282904	TG	*KLF14*	-0.023	-0.032	-0.013	-0.056	-0.067	-0.045	3.27E-22	3.30E-03
rs632057	TG	*CITED2*	-0.028	-0.04	-0.015	-0.056	-0.068	-0.044	2.17E-20	1.70E-06
rs4846914	TG	*GALNT2*	-0.032	-0.044	-0.02	-0.056	-0.067	-0.044	3.34E-19	-
rs13389219	TG	*COBLL1*	-0.025	-0.034	-0.017	-0.053	-0.065	-0.042	8.52E-19	-
rs6073958	TG	*PLTP*	0.034	0.026	0.043	0.06	0.038	0.082	1.06E-18	-
rs948690	TG	*BUD13*	-0.021	-0.03	-0.013	-0.056	-0.069	-0.044	7.58E-18	1.30E-05
rs1390357	TG	*NAT2*	-0.027	-0.057	0.003	-0.064	-0.092	-0.036	9.40E-18	3.61E-04
rs7005978	TG	*UBXN2B*	-0.033	-0.047	-0.02	-0.056	-0.069	-0.043	4.56E-17	7.26E-03
rs2068888	TG	*CYP26A1*	-0.025	-0.034	-0.016	-0.047	-0.058	-0.036	1.73E-15	4.02E-03
rs55829990	TG	*USP3*	0.024	0.016	0.033	0.05	0.037	0.064	1.73E-15	3.97E-04
rs851057	TG	*SOST*	-0.023	-0.057	0.011	-0.059	-0.091	-0.028	4.09E-15	2.66E-04
rs11134475	TG	*TIMD4*	0.03	0.018	0.042	0.049	0.037	0.062	1.36E-14	-
rs3775228	TG	*AFF1*	0.018	0.009	0.027	0.048	0.036	0.061	1.41E-14	-
rs40270	TG	*ANKRD55*	0.036	0.018	0.054	0.06	0.042	0.077	2.36E-14	2.52E-03
rs2945247	TG	*ZNF705B*	-0.017	-0.026	-0.008	-0.045	-0.056	-0.034	3.43E-14	9.22E-04
rs7924036	TG	*JMJD1C*	-0.02	-0.03	-0.011	-0.045	-0.056	-0.034	4.17E-14	-
rs78058190	TG	*PRKAG3*	0.055	0.04	0.07	0.078	-0.016	0.172	8.05E-14	-
rs1045242	TG	*TNFAIP8*	-0.028	-0.036	-0.019	-0.044	-0.058	-0.029	8.74E-14	1.32E-06
rs9757777	TG	*STAG1*	0.03	0.014	0.045	0.052	0.036	0.067	1.99E-13	1.00E-03
rs12928099	TG	*PDXDC1*	-0.021	-0.029	-0.013	-0.05	-0.064	-0.037	2.59E-13	-
rs4821767	TG	*TMEM184B*	0.029	0.018	0.039	0.042	0.031	0.054	1.43E-12	-
rs114484444	TG	*TNFSF10*	0.047	0.033	0.062	0.117	0.02	0.215	2.39E-12	2.73E-03

rs7826687	TG	*TRIB1*	0.022	0.013	0.03	0.048	0.033	0.064	2.66E-12	-
rs11751347	TG	*LPA*	0.035	0.024	0.045	0.065	0.023	0.106	5.21E-12	-
rs4784741	TG	*CETP*	-0.019	-0.028	-0.01	-0.041	-0.052	-0.03	1.27E-11	-
rs4722551	TG	*NFE2L3*	-0.026	-0.035	-0.018	-0.058	-0.081	-0.035	1.32E-11	-
rs2902745	TG	*ZNF579*	0.028	0.018	0.038	0.082	0.043	0.121	1.85E-11	3.28E-03
rs34682685	TG	*TXNL4B|HPR*	0.021	0.01	0.031	0.117	0.072	0.161	4.74E-11	-
rs1417066	TG	*SLC30A10*	0.016	0.005	0.026	0.038	0.027	0.05	5.69E-11	1.68E-04
rs13179413	TG	*MAP3K1*	0.02	0.012	0.028	0.046	0.03	0.062	6.33E-11	2.21E-03
rs141783576	TG	*RSPO3*	0.036	0.024	0.048	0.095	0.032	0.158	8.97E-11	2.01E-03
rs62020701	TG	*UBR1*	0.036	0.025	0.047	0.02	-0.02	0.061	1.37E-10	2.83E-04
rs13108218	TG	*HGFAC*	-0.024	-0.036	-0.012	-0.041	-0.053	-0.029	1.58E-10	-
rs1178982	TG	*FZD9*	-0.037	-0.048	-0.026	-0.038	-0.088	0.013	4.76E-10	-
rs9817452	TG	*LEKR1*	-0.018	-0.027	-0.009	-0.039	-0.05	-0.027	5.17E-10	8.72E-03
rs28446899	TG	*EYA1*	0.032	0.021	0.044	0.093	0.034	0.152	1.03E-09	-
rs2908806	TG	*TP53*	-0.036	-0.062	-0.01	-0.059	-0.084	-0.033	1.05E-09	-
rs12808829	TG	*EML3*	0.021	0.013	0.03	0.036	0.023	0.048	1.24E-09	6.13E-04
rs62192912	Total BR	*ATG16L1*	-0.233	-0.242	-0.223	-0.357	-0.369	-0.344	0	3.76E-03
rs2741047	Total BR	*UGT1A8|UGT1A10|UGT1A9*	0.14	0.133	0.147	0.841	0.826	0.856	0	8.11E-32
rs6712540	Total BR	*TRPM8*	-0.145	-0.155	-0.134	-0.212	-0.246	-0.179	3.07E-136	1.34E-73
rs11045864	Total BR	*SLCO1B1*	0.085	0.073	0.096	0.161	0.125	0.196	1.52E-64	1.12E-04
rs474242	Total BR	*MROH2A*	-0.067	-0.078	-0.056	-0.113	-0.126	-0.1	7.67E-59	7.70E-39
rs9750891	Total BR	*INPP5D*	0.11	0.093	0.128	0.29	0.168	0.411	1.93E-45	9.44E-25
rs9414801	Total BR	*JMJD1C*	0.038	0.026	0.051	0.07	0.056	0.084	3.43E-22	6.30E-12
rs76820150	Total BR	*SLCO1C1*	-0.044	-0.059	-0.029	-0.072	-0.087	-0.057	4.16E-20	-
rs450244	Total BR	*SLC22A18*	-0.194	-0.261	-0.128	-0.224	-0.289	-0.159	3.18E-16	4.68E-05
rs687339	Total BR	*MSL2*	-0.009	-0.033	0.015	-0.047	-0.07	-0.025	7.31E-14	1.91E-06

rs13031505	Total BR	*EFHD1*	-0.02	-0.031	-0.01	-0.055	-0.069	-0.041	1.29E-12	1.08E-07
rs10774624	Total BR	*SH2B3*	-0.03	-0.042	-0.017	-0.05	-0.064	-0.037	9.43E-12	-
rs2068888	Total BR	*CYP26A1*	0.028	0.016	0.039	0.047	0.032	0.061	1.01E-10	-
rs11635675	Total BR	*USP3*	0.025	0.015	0.036	0.047	0.031	0.064	6.47E-10	1.57E-05
rs6479336	Total BR	*AUH*	0.026	0.016	0.037	0.068	0.039	0.097	9.27E-10	1.73E-04
rs74904971	Urate	*ABCG2*	0.053	0.046	0.06	0.11	0.082	0.138	1.45E-69	-
rs938555	Urate	*SLC2A9*	0.029	0.019	0.04	0.067	0.056	0.077	4.46E-65	-
rs1260326	Urate	*GCKR*	-0.02	-0.027	-0.012	-0.036	-0.044	-0.028	2.82E-19	-
rs4530622	Urate	*SLC2A9*	-0.023	-0.028	-0.018	-0.032	-0.042	-0.021	3.03E-19	-
rs12056034	Urate	*BAZ1B*	-0.018	-0.024	-0.012	-0.037	-0.056	-0.017	1.11E-09	-
rs11956741	Urea	*PTGER4*	-0.029	-0.035	-0.023	-0.053	-0.078	-0.029	7.10E-22	-
rs9880162	Urea	*LPP*	-0.011	-0.015	-0.006	-0.022	-0.029	-0.014	4.92E-10	-
rs2138733	Urea	*HNF1B*	-0.005	-0.012	0.001	-0.018	-0.024	-0.011	1.28E-09	-
rs11023212	Vitamin D	*COPB1*	-0.191	-0.202	-0.179	-0.377	-0.392	-0.361	0	-
rs11723621	Vitamin D	*GC*	-0.162	-0.173	-0.151	-0.326	-0.343	-0.309	3.30E-301	-
rs146128209	Vitamin D	*PDE3B*	-0.129	-0.144	-0.113	-0.345	-0.401	-0.289	4.02E-69	-
rs8022510	Vitamin D	*SEC23A*	-0.096	-0.132	-0.06	-0.19	-0.224	-0.156	5.25E-66	-
rs4757226	Vitamin D	*RRAS2*	-0.072	-0.084	-0.06	-0.139	-0.157	-0.122	2.92E-59	-
rs1057868	Vitamin D	*POR*	0.064	0.052	0.076	0.115	0.092	0.137	5.93E-39	-
rs10426201	Vitamin D	*SULT2A1*	-0.091	-0.13	-0.052	-0.161	-0.198	-0.124	6.42E-39	-
rs732934	Vitamin D	*NADSYN1*	0.035	0.008	0.063	0.09	0.063	0.118	1.33E-23	-
rs6685829	Vitamin D	*RER1*	-0.025	-0.039	-0.011	-0.07	-0.086	-0.054	3.95E-17	-
rs58542926	Vitamin D	*TM6SF2*	0.036	0.02	0.053	0.273	0.183	0.364	2.08E-14	-
rs56287450	Vitamin D	*PDE3B*	0.072	0.053	0.091	0.115	0.015	0.216	3.39E-14	-
rs261290	Vitamin D	*ALDH1A2*	0.044	0.026	0.063	0.069	0.051	0.088	5.83E-13	-
rs11933459	Vitamin D	*UGT2B7*	-0.027	-0.04	-0.014	-0.059	-0.075	-0.043	9.32E-12	-

rs12574800	Vitamin D	*FAR1*	-0.055	-0.071	-0.038	-0.105	-0.186	-0.025	1.27E-10	-
rs12494636	Vitamin D	*PAK2*	0.004	-0.008	0.016	-0.078	-0.102	-0.055	5.05E-10	-

Phi, SNP effect on standardised trait variance. X1, effect of one SNP dosage increase. X2, effect of two SNP dosage increase. P_{Lyon}, test for effect on trait variance using LAD-BF. $P_{Westerman}$ test for effect on trait variance from Westerman *et al*[42].

Table 9.2.2 Top GxG/GxE effects on biomarker concentration in UK Biobank

SNP	Modifier	Outcome	Beta	95% CI		P
***ABO*(rs635634)**	*FUT2*(rs281379)	ALP	0.078	0.070	0.087	1.62E-72
***ALPL*(rs4654970)**	BMI	ALP	-0.029	-0.037	-0.022	8.54E-16
***ABO*(rs635634)**	*TREH*(rs12225548)	ALP	0.035	0.025	0.045	1.17E-11
***ALPL*(rs4654970)**	Age	ALP	-0.024	-0.031	-0.017	1.27E-11
***ABO*(rs635634)**	Sex	ALP	-0.017	-0.023	-0.011	2.44E-08
***TREH*(rs12225548)**	*FUT2*(rs281379)	ALP	0.023	0.015	0.031	3.50E-08
***PNPLA3*(rs738409)**	BMI	ALT	0.082	0.075	0.089	3.37E-119
***TM6SF2*(rs58542926)**	BMI	ALT	0.051	0.040	0.062	4.52E-21
***HSD17B13*(rs71633359)**	BMI	ALT	-0.026	-0.032	-0.021	7.57E-21
***HSD17B13*(rs71633359)**	*PNPLA3*(rs738409)	ALT	-0.044	-0.053	-0.034	2.57E-19
***TRIB1*(rs2954021)**	BMI	ALT	-0.023	-0.028	-0.018	1.09E-17
***MARC1*(rs2642438)**	BMI	ALT	0.023	0.018	0.028	1.12E-16
***APOE*(rs429358)**	BMI	ALT	-0.026	-0.033	-0.020	4.14E-14
***ERLIN1*(rs2862954)**	BMI	ALT	-0.020	-0.025	-0.015	4.93E-14
***TRIB1*(rs2954021)**	Sex	ALT	-0.017	-0.022	-0.013	2.95E-12
***TOR1B*(rs7029757)**	BMI	ALT	-0.026	-0.034	-0.017	2.98E-09
***CETP*(rs247616)**	BMI	ApoA	-0.021	-0.026	-0.016	2.77E-17
***LDLR*(rs6511720)**	Age	ApoB	0.029	0.022	0.037	1.23E-15
***CELSR2*(rs12740374)**	BMI	ApoB	-0.022	-0.028	-0.016	2.51E-13
***CELSR2*(rs12740374)**	Sex	ApoB	-0.018	-0.024	-0.013	4.50E-10
***CELSR2*(rs12740374)**	Age	ApoB	0.017	0.011	0.023	6.82E-09
***PNPLA3*(rs3747207)**	BMI	AST	0.072	0.065	0.080	1.95E-82
***GCKR*(rs1260326)**	BMI	AST	-0.024	-0.030	-0.019	1.42E-18
***HSD17B13*(rs71633359)**	BMI	AST	-0.024	-0.029	-0.018	8.62E-16
***HSD17B13*(rs71633359)**	*PNPLA3*(rs3747207)	AST	-0.036	-0.047	-0.026	3.03E-12

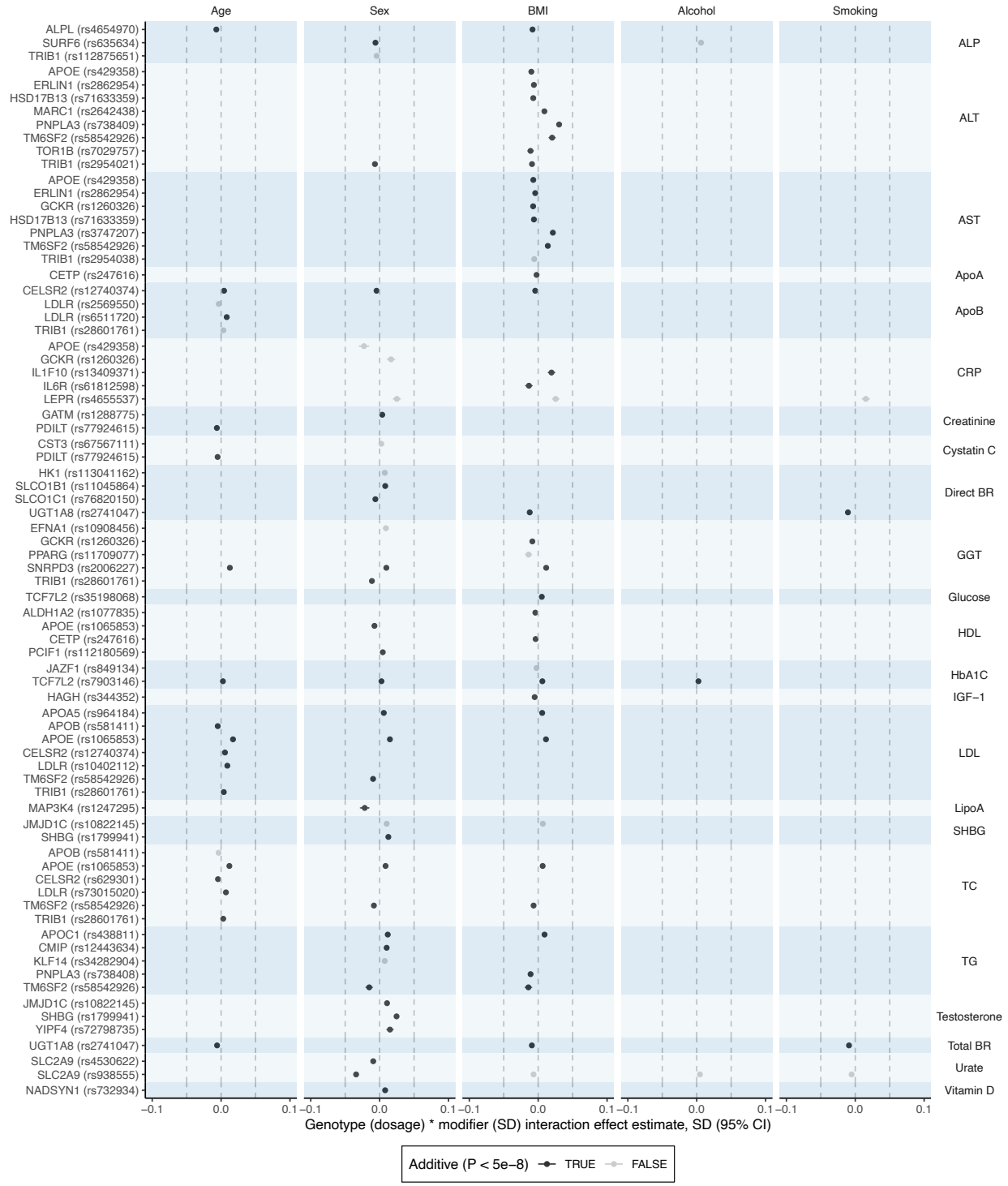
***TM6SF2*(rs58542926)**	BMI	AST	0.040	0.028	0.052	5.79E-11
***ERLIN1*(rs2862954)**	BMI	AST	-0.017	-0.022	-0.011	1.02E-08
***APOE*(rs429358)**	BMI	AST	-0.022	-0.029	-0.014	1.91E-08
***PDILT*(rs77924615)**	Age	Creatinine	-0.028	-0.033	-0.022	5.44E-22
***GATM*(rs1288775)**	Sex	Creatinine	0.021	0.016	0.027	3.12E-16
***IL6R*(rs61812598)**	BMI	CRP	-0.031	-0.037	-0.025	1.51E-25
***IL1F10*(rs13409371)**	BMI	CRP	0.028	0.021	0.034	1.97E-18
***PDILT*(rs77924615)**	Age	Cystatin C	-0.031	-0.037	-0.025	3.00E-26
***UGT1A8*(rs2741047)**	BMI	Direct BR	-0.041	-0.046	-0.036	1.37E-51
***UGT1A8*(rs2741047)**	Smoking	Direct BR	-0.035	-0.040	-0.030	5.56E-41
***SLCO1B1*(rs11045864)**	Sex	Direct BR	0.030	0.022	0.037	1.15E-15
***SLCO1C1*(rs76820150)**	Sex	Direct BR	-0.019	-0.024	-0.014	1.76E-12
***SNRPD3*(rs2006227)**	BMI	GGT	0.026	0.020	0.031	6.18E-20
***SNRPD3*(rs2006227)**	Sex	GGT	0.022	0.016	0.027	7.79E-15
***TRIB1*(rs28601761)**	Sex	GGT	-0.019	-0.025	-0.014	2.98E-14
***GCKR*(rs1260326)**	BMI	GGT	-0.016	-0.022	-0.011	1.80E-09
***ZNF827*(rs4835265)**	*NEDD4L*(rs4503880)	GGT	-0.040	-0.053	-0.027	3.41E-09
***SNRPD3*(rs2006227)**	Age	GGT	0.015	0.010	0.020	6.92E-09
***TCF7L2*(rs35198068)**	BMI	Glucose	0.034	0.026	0.042	1.64E-16
***TCF7L2*(rs7903146)**	BMI	HbA1C	0.045	0.038	0.052	9.70E-35
***TCF7L2*(rs7903146)**	Age	HbA1C	0.018	0.013	0.023	2.87E-11
***TCF7L2*(rs7903146)**	Sex	HbA1C	0.018	0.013	0.024	1.06E-10
***TCF7L2*(rs7903146)**	Alcohol	HbA1C	0.019	0.013	0.025	7.44E-10
***CETP*(rs247616)**	BMI	HDL	-0.035	-0.040	-0.030	8.52E-47
***ALDH1A2*(rs1077835)**	BMI	HDL	-0.026	-0.031	-0.020	1.42E-19
***APOE*(rs1065853)**	Sex	HDL	-0.039	-0.048	-0.030	7.67E-18
***PCIF1*(rs112180569)**	Sex	HDL	0.024	0.019	0.030	1.44E-16
***HAGH*(rs344352)**	BMI	IGF-1	-0.016	-0.021	-0.010	7.14E-09

***APOE*(rs1065853)**	Sex	LDL	0.062	0.054	0.071	3.34E-48
***APOE*(rs1065853)**	Age	LDL	0.051	0.043	0.059	4.25E-33
***LDLR*(rs10402112)**	Age	LDL	0.031	0.024	0.038	2.53E-17
***TM6SF2*(rs58542926)**	Sex	LDL	-0.037	-0.046	-0.028	6.52E-16
***APOE*(rs1065853)**	BMI	LDL	0.035	0.026	0.043	6.98E-16
***CELSR2*(rs12740374)**	Age	LDL	0.021	0.015	0.026	1.24E-12
***APOA5*(rs964184)**	Sex	LDL	0.026	0.018	0.033	5.70E-12
***APOA5*(rs964184)**	BMI	LDL	0.023	0.016	0.030	1.06E-09
***TRIB1*(rs28601761)**	Age	LDL	0.015	0.010	0.020	6.83E-09
***APOB*(rs581411)**	Age	LDL	-0.018	-0.024	-0.012	1.46E-08
***MAP3K4*(rs1247295)**	Sex	LipoA	-0.016	-0.021	-0.010	1.19E-08
***SHBG*(rs1799941)**	Sex	SHBG	-0.019	-0.024	-0.014	1.25E-12
***APOE*(rs1065853)**	Sex	TC	0.052	0.044	0.061	6.48E-33
***APOE*(rs1065853)**	Age	TC	0.046	0.038	0.055	3.19E-26
***TM6SF2*(rs58542926)**	Sex	TC	-0.038	-0.047	-0.029	4.43E-17
***LDLR*(rs73015020)**	Age	TC	0.030	0.023	0.037	1.27E-16
***APOE*(rs1065853)**	BMI	TC	0.033	0.024	0.041	2.33E-13
***CELSR2*(rs629301)**	Age	TC	-0.021	-0.026	-0.015	6.92E-13
***TM6SF2*(rs58542926)**	BMI	TC	-0.032	-0.041	-0.023	5.70E-12
***TRIB1*(rs28601761)**	Age	TC	0.014	0.009	0.019	1.20E-08
***SHBG*(rs1799941)**	Sex	Testosterone	0.060	0.057	0.062	0
***JMJD1C*(rs10822145)**	Sex	Testosterone	0.032	0.029	0.034	6.11E-169
***YIPF4*(rs72798735)**	Sex	Testosterone	0.037	0.032	0.042	1.74E-45
***APOC1*(rs438811)**	BMI	TG	0.033	0.027	0.040	6.28E-27
***APOC1*(rs438811)**	Sex	TG	0.032	0.026	0.038	1.33E-25
***TM6SF2*(rs58542926)**	Sex	TG	-0.039	-0.047	-0.030	7.44E-19
***TM6SF2*(rs58542926)**	BMI	TG	-0.039	-0.047	-0.030	1.19E-17
***PNPLA3*(rs738408)**	BMI	TG	-0.025	-0.031	-0.019	3.34E-17

***CMIP*(rs12443634)**	Sex	TG	0.017	0.011	0.022	2.11E-09
UGT1A8(rs2741047)	BMI	Total BR	-0.040	-0.044	-0.035	1.18E-66
***UGT1A8*(rs2741047)**	Smoking	Total BR	-0.037	-0.042	-0.033	2.52E-60
***UGT1A8*(rs2741047)**	Age	Total BR	-0.021	-0.026	-0.016	5.64E-17
***SLC2A9*(rs938555)**	Sex	Urate	-0.080	-0.085	-0.076	1.05E-232
***SLC2A9*(rs4530622)**	Sex	Urate	-0.029	-0.033	-0.024	8.01E-31
***NADSYN1*(rs732934)**	Sex	Vitamin D	0.019	0.013	0.025	6.22E-10

Beta, interaction effect of SNP on outcome. CI, confidence interval. P, p-value for interaction slope.

Figure 9.2.1. Top gene-by-environment interaction effects ($P < 5 \times 10^{-8}$) on biomarker concentration using multiplicative scale



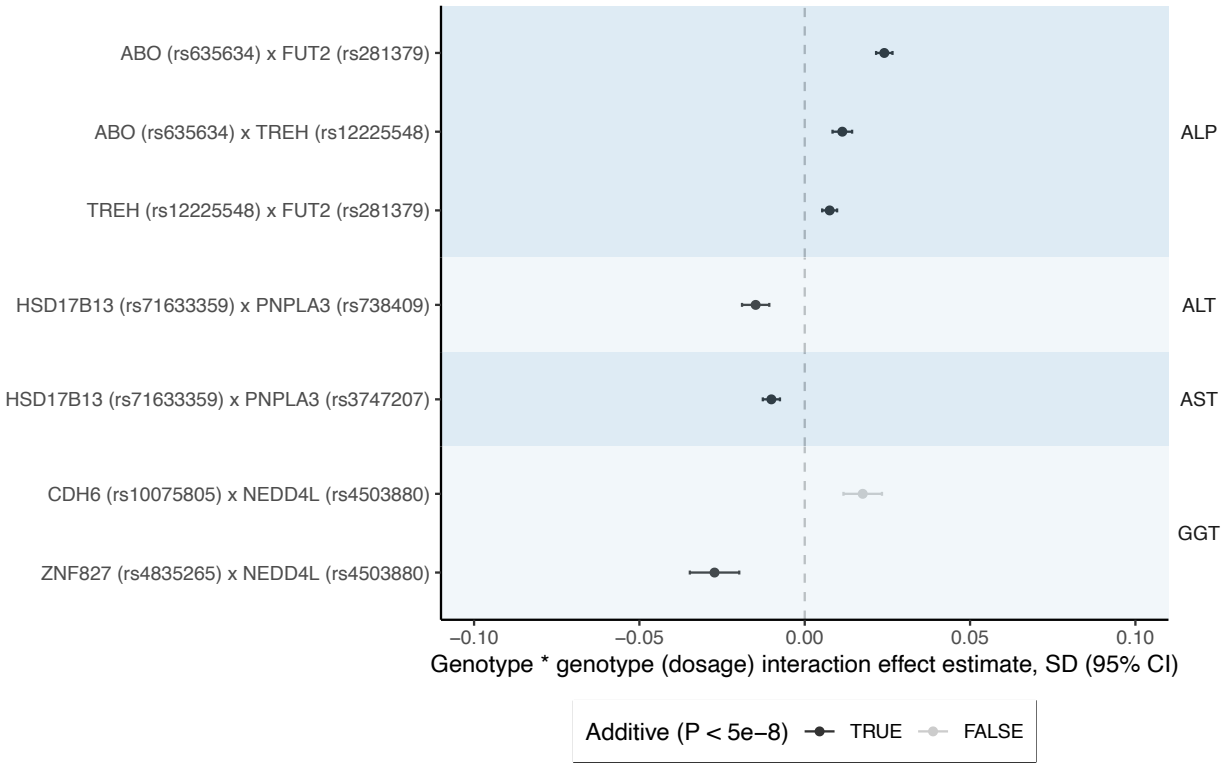
GxE effects using multiplicative scale and heteroscedasticity consistent standard errors[107] ($P < 5 \times 10^{-8}$). ALP, alkaline phosphatase. ALT, alanine aminotransferase. AST, aspartate aminotransferase. ApoA, Apolipoprotein A. ApoB, apolipoprotein B. CRP, C-reactive protein. Direct BR, direct bilirubin. GGT, Gamma glutamyltransferase. HDL, high-density lipoprotein. HbA1c, glycated haemoglobin. LDL, low-density lipoprotein. LipoA, lipoprotein A. IGF-1, insulin-like growth factor 1. SHBG, sex-hormone binding globulin. TC, total cholesterol. TG, triglycerides. Total BR, total bilirubin. BMI, body mass index. Smoking, smoking status. Alcohol, intake. PA, physical activity. All measures reported on SD scale. All estimates were adjusted for the main effect, age, sex, and top ten genetic principal components. Vertical dashed lines are present at -0.05, 0 and 0.05 SD. Gene name Is the nearest protein coding gene HGNC name by chromosomal position. SD, standard deviation. CI, confidence interval.

Figure 9.2.2. Top gene-by-environment interaction effects ($P < 5 \times 10^{-8}$) on biomarker concentration using additive scale adjusted for fine-mapped main effect

GxE effects using additive scale and heteroscedasticity consistent standard errors[107] ($P < 5 \times 10^{-8}$) adjusted for fine-mapped main effects. ALP, alkaline phosphatase. ALT, alanine

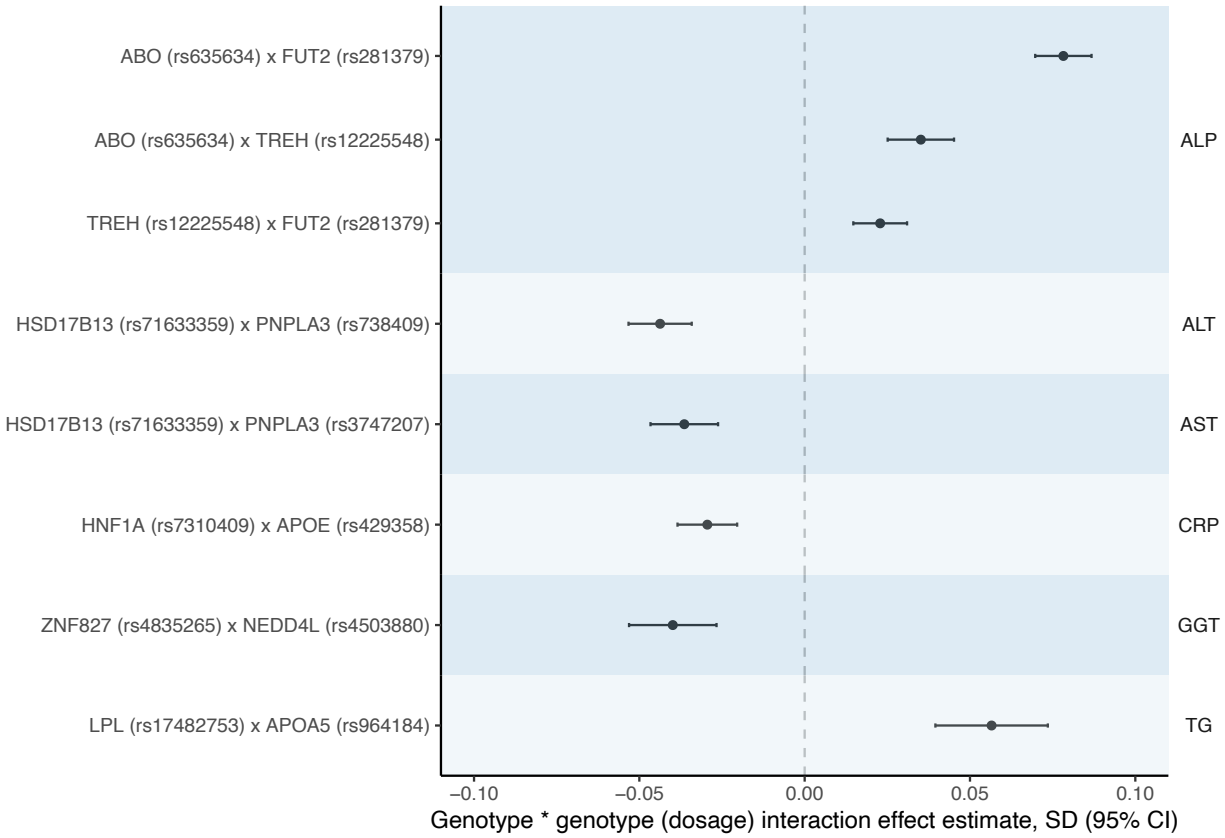
aminotransferase. AST, aspartate aminotransferase. ApoA, Apolipoprotein A. ApoB, apolipoprotein B. CRP, C-reactive protein. Direct BR, direct bilirubin. GGT, Gamma glutamyltransferase. HDL, high-density lipoprotein. HbA1c, glycated haemoglobin. LDL, low-density lipoprotein. LipoA, lipoprotein A. IGF-1, insulin-like growth factor 1. SHBG, sex-hormone binding globulin. TC, total cholesterol. TG, triglycerides. Total BR, total bilirubin. BMI, body mass index. Smoking, smoking status. Alcohol, intake. PA, physical activity. All measures reported on SD scale. All estimates were adjusted for the main effect, age, sex, and top ten genetic principal components. Vertical dashed lines are present at -0.05, 0 and 0.05 SD. Gene name is the nearest protein coding gene HGNC name by chromosomal position. SD, standard deviation. CI, confidence interval.

Figure 9.2.3. Top gene-by-gene interaction effects ($P < 5 \times 10^{-8}$) on biomarker concentration using multiplicative scale



GxG effects using multiplicative scale and heteroscedasticity consistent standard errors[107] ($P < 5 \times 10^{-8}$) adjusted for the main effect, age, sex, and top ten genetic principal components. ALP, alkaline phosphatase. ALT, alanine aminotransferase. AST, Aspartate aminotransferase. GGT, Gamma glutamyltransferase. All measures reported on SD scale. Gene name is the nearest protein coding gene HGNC name by chromosomal position. Vertical dashed line marks null association. SD, standard deviation. CI, confidence interval.

Figure 9.2.4. Top gene-by-gene interaction effects ($P < 5 \times 10^{-8}$) on biomarker concentration using additive scale adjusted for fine-mapped main effects



GxG effects using additive scale and heteroscedasticity consistent standard errors[107] ($P < 5 \times 10^{-8}$) adjusted for the main effect, age, sex, top ten genetic principal components and fine-mapped main effects. ALP, alkaline phosphatase. ALT, alanine aminotransferase. AST, Aspartate aminotransferase. CRP, C-reactive protein. GGT, Gamma glutamyltransferase. TG, triglycerides. All measures reported on SD scale. Gene name is the nearest protein coding gene HGNC name by chromosomal position. Vertical dashed line marks null association. SD, standard deviation. CI, confidence interval.

References

1. Lyon, M. S. *et al.* The variant call format provides efficient and robust storage of GWAS summary statistics. *Genome Biol.* **22**, 32 (2021).

2. Lyon, M. S., Millard, L. A. C., Smith, G. D., Gaunt, T. R. & Tilling, K. Hypothesis-free detection of gene-interaction effects on biomarker concentration in UK Biobank using variance prioritisation. *medRxiv* (2022) doi:10.1101/2022.01.05.21268406.

3. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nat. 2009 4617265* **461**, 747–753 (2009).

4. Wei, W. H., Hemani, G. & Haley, C. S. Detecting epistasis in human complex traits. *Nat. Rev. Genet.* **15**, 722–733 (2014).

5. Visscher, P. M., Hill, W. G. & Wray, N. R. Heritability in the genomics era - Concepts and misconceptions. *Nature Reviews Genetics* vol. 9 255–266 (2008).

6. Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).

7. Uffelmann, E. *et al.* Genome-wide association studies. *Nat. Rev. Methods Prim.* **1**, 1–21 (2021).

8. Howe, L. J. *et al.* Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects. *Nat. Genet.* **54**, 581–592 (2022).

9. Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* **19**, 491–504 (2018).

10. Cardon, L. R. & Palmer, L. J. Population stratification and spurious allelic association. *Lancet* **361**, 598–604 (2003).

11. Price, A. L., Zaitlen, N. A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* **11**, 459–463 (2010).

12. Brumpton, B. *et al.* Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. *Nat. Commun.* **11**, 3519 (2020).

13. Aulchenko, Y. S., De Koning, D. J. & Haley, C. Genomewide Rapid Association Using Mixed Model and Regression: A Fast and Simple Method For Genomewide Pedigree-Based Quantitative Trait Loci Association Analysis. *Genetics* **177**, 577–585 (2007).

14. Hou, L. & Zhao, H. A review of post-GWAS prioritization approaches. *Front. Genet.* **4**, 280 (2013).

15. Smith, G. D. & Ebrahim, S. 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* **31**, 1–22 (2003).

16. Bulik-Sullivan, B. *et al.* LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

17. Yang, J., Zeng, J., Goddard, M. E., Wray, N. R. & Visscher, P. M. Concepts, estimation and interpretation of SNP-based heritability. *Nat. Genet.* **49**, 1304–1310 (2017).

18. Elsworth, B. *et al.* The MRC IEU OpenGWAS data infrastructure. *bioRxiv* (2020) doi:10.1101/2020.08.10.244293.

19. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).

20. Hartwig, F. P., Davies, N. M., Hemani, G. & Smith, G. D. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. *Int. J. Epidemiol.* **45**, 1717–1726 (2016).

21. Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).

22. Murphy, A. E., Schilder, B. M. & Skene, N. G. MungeSumstats: a Bioconductor package for the standardization and quality control of many GWAS summary statistics. *Bioinformatics* **37**, 4593–4596 (2021).

23. Altman, D. G. & Bland, J. M. How to obtain the P value from a confidence interval. *BMJ* **343**, (2011).

24. Li, H. Tabix: fast retrieval of sequence features from generic TAB-delimited files. *Bioinforma. Appl. NOTE* **27**, 718–719 (2011).

25. Hayhurst, J. *et al.* A community driven GWAS summary statistics standard. *bioRxiv* (2022) doi:10.1101/2022.07.15.500230.

26. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).

27. Bland, M. *An introduction to medical statistics*. (Oxford University Press, 2000).

28. Pearl, J. & Paz, A. Confounding equivalence in causal inference. *Proc. 26th Conf. Uncertain. Artif. Intell. UAI 2010* **2**, 433–441 (2010).

29. Akobeng, A. K. Understanding randomised controlled trials. *Arch. Dis. Child.* **90**, 840–844 (2005).

30. Burgess, S., Swanson, S. A. & Labrecque, J. A. Are Mendelian randomization investigations immune from bias due to reverse causation? *Eur. J. Epidemiol.* **36**, 253–257 (2021).

31. Sanderson, E. *et al.* Mendelian randomization. *Nat. Rev. Methods Prim.* **2**, 1–21 (2022).

32. Davies, N. M., Holmes, M. V. & Davey Smith, G. Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. *BMJ* **362**, 601 (2018).

33. Davies, N. M., Dickson, M., Davey, G., Windmeijer, S. F. & Van Den Berg, G. J. The Causal Effects of Education on Adult Health, Mortality and Income: Evidence from Mendelian Randomization and the Raising of the School Leaving Age. *Inst. Labor Econ.* (2019).

34. Brookhart, M. A. & Schneeweiss, S. Preference-based instrumental variable methods for the estimation of treatment effects: Assessing validity and interpreting results. *Int. J. Biostat.* **3**, (2007).

35. Hartwig, F. P., Bowden, J., Wang, L., Smith, G. D. & Davies, N. M. Average causal effect estimation via instrumental variables: the no simultaneous heterogeneity assumption. *arXiv* (2020) doi:https://doi.org/10.48550/arXiv.2010.10017.

36. Labrecque, J. & Swanson, S. A. Understanding the Assumptions Underlying Instrumental Variable Analyses: a Brief Review of Falsification Strategies and Related Tools. *Curr. Epidemiol. Reports* **5**, 214–220 (2018).

37. Angrist, J. D., Imbens, G. W. & Rubin, D. B. Identification of Causal Effects Using Instrumental Variables. *Source J. Am. Stat. Assoc.* **91**, 444–455 (1996).

38. Sheehan, N. A. & Didelez, V. Epidemiology, genetic epidemiology and Mendelian randomisation: more need than ever to attend to detail. *Hum. Genet.* **139**, 121–136 (2020).

39. Wang, L. & Tchetgen Tchetgen, E. Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **80**, 531–550 (2018).

40. Hernán, M. A. & Robins, J. M. Instruments for causal inference: An epidemiologist's dream? *Epidemiology* **17**, 360–372 (2006).

41. Swanson, S. A. & Hernán, M. A. Commentary: How to report instrumental variable analyses (suggestions welcome). *Epidemiology* **24**, 370–374 (2013).

42. Westerman, K. E. *et al.* Variance-quantitative trait loci enable systematic discovery of gene-environment interactions for cardiometabolic serum biomarkers. *Nat. Commun.* **13**, 1–11 (2022).

43. Mills, H. L. *et al.* Detecting Heterogeneity of Intervention Effects Using Analysis and Meta-analysis of Differences in Variance between Trial Arms. *Epidemiology* **32**, 846–854 (2021).

44. Burgess, S. & Thompson, S. G. Avoiding bias from weak instruments in mendelian randomization studies. *Int. J. Epidemiol.* **40**, 755–764 (2011).

45. Burgess, S., Davies, N. M. & Thompson, S. G. Bias due to participant overlap in two-sample Mendelian randomization. *Genet. Epidemiol.* **40**, 597–608 (2016).

46. Sanderson, E., Richardson, T. G., Hemani, G. & Davey Smith, G. The use of negative control outcomes in Mendelian randomization to detect potential population stratification. *Int. J. Epidemiol.* **50**, 1350–1361 (2021).

47. Spiller, W., Hartwig, F. P., Sanderson, E., Davey Smith, G. & Bowden, J. Interaction-based Mendelian randomization with measured and unmeasured gene-by-covariate interactions. *PLoS One* **17**, e0271933 (2022).

48. Hunter, D. J. Gene–environment interactions in human diseases. *Nat. Rev. Genet.* **6**, 287–298 (2005).

49. Ochoa, D. *et al.* Human genetics evidence supports two-thirds of the 2021 FDA-approved drugs. *Nat. Rev. Drug Discov.* (2022) doi:10.1038/d41573-022-00120-3.

50. Cortés, J. *et al.* Does evidence support the high expectations placed in precision medicine? A bibliographic review. *F1000Research* **7**, 30 (2019).

51. Senn, S. Mastering variation: variance components and personalised medicine. *Stat. Med.* **35**, 966–977 (2016).

52. Ashley, E. A. Towards precision medicine. *Nat. Rev. Genet.* **17**, 507–522 (2016).

53. Hemani, G. *et al.* Phantom epistasis between unlinked loci. *Nature* **596**, E1–E3 (2021).

54. Doherty, A. *et al.* Large Scale Population Assessment of Physical Activity Using Wrist Worn Accelerometers: The UK Biobank Study. *PLoS One* **12**, e0169649 (2017).

55. Button, K. S. *et al.* Power failure: Why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* **14**, 365–376 (2013).

56. Paré, G., Cook, N. R., Ridker, P. M. & Chasman, D. I. On the Use of Variance per Genotype as a Tool to Identify Quantitative Trait Interaction Effects: A Report from the Women's Genome Health Study. *PLoS Genet.* **6**, e1000981 (2010).

57. Brookes, S. T. *et al.* Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *J. Clin. Epidemiol.* **57**, 229–236 (2004).

58. Smith, P. G. & Day, N. E. The Design of Case-Control Studies: The Influence of Confounding and Interaction Effects. *Int. J. Epidemiol.* **13**, 356–365 (1984).

59. Colhoun, H. M., McKeigue, P. M. & Smith, G. D. Problems of reporting genetic associations with complex outcomes. *Lancet* **361**, 865–872 (2003).

60. Duncan, L. E. & Keller, M. C. A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *Am. J. Psychiatry* **168**, 1041–1049 (2011).

61. Rees, J. L. The Genetics of Sun Sensitivity in Humans. *Am. J. Hum. Genet.* **75**, 739–751 (2004).

62. Box, N. F. *et al.* MC1R Genotype Modifies Risk of Melanoma in Families Segregating

CDKN2A Mutations. *Am. J. Hum. Genet.* **69**, 765–773 (2001).

63. Van der Velden, P. A. *et al.* Melanocortin-1 Receptor Variant R151C Modifies Melanoma Risk in Dutch Families with Melanoma. *Am. J. Hum. Genet.* **69**, 774–779 (2001).

64. Chen, J., Giovannucci, E. L. & Hunter, D. J. MTHFR Polymorphism, Methyl-Replete Diets and the Risk of Colorectal Carcinoma and Adenoma among U.S. Men and Women: An Example of Gene-Environment Interactions in Colorectal Tumorigenesis. *J. Nutr.* **129**, 560S-564S (1999).

65. Lehtimaki, T. *et al.* Association between serum lipids and apolipoprotein E phenotype is influenced by diet in a population-based sample of free-living children and young adults: the Cardiovascular Risk in Young Finns Study. *J. Lipid Res.* **36**, 653–661 (1995).

66. Combarros, O. *et al.* Replication by the Epistasis Project of the interaction between the genes for IL-6 and IL-10 in the risk of Alzheimer's disease. *J. Neuroinflammation* **6**, 22 (2009).

67. Rönnegård, L. & Valdar, W. Recent developments in statistical methods for detecting genetic loci affecting phenotypic variability. *BMC Genet.* **13**, 1–7 (2012).

68. Wang, H. *et al.* Genotype-by-environment interactions inferred from genetic effects on phenotypic variability in the UK Biobank. *Sci. Adv.* **5**, eaaw3538 (2019).

69. Young, A. I., Wauthier, F. L. & Donnelly, P. Identifying loci affecting trait variability and detecting interactions in genome-wide association studies. *Nat. Genet.* **50**, 1608–1614 (2018).

70. Marderstein, A. R. *et al.* Leveraging phenotypic variability to identify genetic interactions in human phenotypes. *Am. J. Hum. Genet.* **108**, 49–67 (2021).

71. Brown, M. B. & Forsythe, A. B. Robust tests for the equality of variances. *J. Am. Stat. Assoc.* **69**, 364–367 (1974).

72. Levene, H. Robust testes for equality of variances. *Contrib. to Probab. Stat.* 278–292 (1960).

73. Properties of sufficiency and statistical tests. *Proc. R. Soc. London. Ser. A - Math. Phys. Sci.* **160**, 268–282 (1937).

74. Fligner, M. A. & Killeen, T. J. Distribution-free two-sample tests for scale. *J. Am. Stat. Assoc.* **71**, 210–213 (1976).

75. Smyth, G. K. Generalized Linear Models with Varying Dispersion. *Source J. R. Stat. Soc. Ser. B* **51**, 47–60 (1989).

76. Corty, R. W. & Valdar, W. QTL mapping on a background of variance heterogeneity. *G3 Genes, Genomes, Genet.* **8**, 3767–3782 (2018).

77. Yang, J. *et al.* FTO genotype is associated with phenotypic variability of body mass index. *Nature* **490**, 267–273 (2012).

78. Breusch, T. S. & Pagan, A. R. A Simple Test for Heteroscedasticity and Random Coefficient Variation. **47**, 1287–1294 (1979).

79. Struchalin, M. V, Amin, N., Eilers, P. H., Duijn, C. M. van & Aulchenko, Y. S. An R package 'VariABEL' for genome-wide searching of potentially interacting loci by testing genotypic variance heterogeneity. *BMC Genet.* **13**, 1–7 (2012).

80. Staley, J. R. *et al.* A robust mean and variance test with application to high-dimensional phenotypes. *Eur. J. Epidemiol.* **1**, 1–11 (2021).

81. Manikandan, S. Measures of central tendency: Median and mode. *J. Pharmacol. Pharmacother.* **2**, 214–215 (2011).

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

83. Soave, D. & Sun, L. A generalized Levene's scale test for variance heterogeneity in the presence of sample correlation and group uncertainty. *Biometrics* **73**, 960–971 (2017).

84. Yang, J. *et al.* FTO genotype is associated with phenotypic variability of body mass index. *Nature* **490**, 267–72 (2012).

85. Soave, D. *et al.* A Joint Location-Scale Test Improves Power to Detect Associated SNPs, Gene Sets, and Pathways. *Am. J. Hum. Genet.* **97**, 125–138 (2015).

86. Cao, Y., Wei, P., Bailey, M., Kauwe, J. S. K. & Maxwell, T. J. A Versatile Omnibus Test for Detecting Mean and Variance Heterogeneity. *Genet. Epidemiol.* **38**, 51–59 (2014).

87. Deng, W. Q. & Paré, G. A fast algorithm to optimize SNP prioritization for gene-gene and gene-environment interactions. *Genet. Epidemiol.* **35**, 729–738 (2011).

88. Lassi, G. *et al.* The CHRNA5–A3–B4 Gene Cluster and Smoking: From Discovery to Therapeutics. *Trends Neurosci.* **39**, 851–861 (2016).

89. Brown, A. A. *et al.* Genetic interactions affecting human gene expression identified by variance association mapping. *Elife* e01381 (2014) doi:10.7554/eLife.01381.

90. Revez, J. A. *et al.* Genome-wide association study identifies 143 loci associated with 25 hydroxyvitamin D concentration. *Nat. Commun.* **11**, 1–12 (2020).

91. Liu, D. *et al.* PRICKLE1 × FOCAD Interaction Revealed by Genome-Wide vQTL Analysis of Human Facial Traits. *Front. Genet.* **12**, 1112 (2021).

92. Sun, X., Elston, R., Morris, N. & Zhu, X. What is the significance of difference in phenotypic variability across SNP genotypes? *Am. J. Hum. Genet.* **93**, 390–397 (2013).

93. Campos, G. de los, Sorensen, D. A. & Toro, M. A. Imperfect Linkage Disequilibrium Generates Phantom Epistasis (& Perils of Big Data). *G3 Genes, Genomes, Genet.* **9**, 1429–1436 (2019).

94. Wood, A. R. *et al.* Another explanation for apparent epistasis. *Nature* **514**, E3–E5 (2014).

95. Berisa, T. & Pickrell, J. K. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* **32**, 283 (2016).

96. Zhang, F. *et al.* OSCA: A tool for omic-data-based complex trait analysis. *Genome Biol.* **20**, 1–13 (2019).

97. Schmidt, A. F. *et al.* Tailoring treatments using treatment effect modification. *Pharmacoepidemiol. Drug Saf.* **25**, 355–362 (2016).

98. Strimbu, K. & Tavel, J. A. What are biomarkers? *Curr. Opin. HIV AIDS* **5**, 463–466 (2010).

99. Holmes, M. V., Richardson, T. G., Ference, B. A., Davies, N. M. & Davey Smith, G. Integrating genomics with biomarkers and therapeutic targets to invigorate cardiovascular drug development. *Nat. Rev. Cardiol.* **18**, 435–453 (2021).

100. Pekkanen, J. *et al.* Ten-Year Mortality from Cardiovascular Disease in Relation to Cholesterol Level among Men with and without Preexisting Cardiovascular Disease. *NEJM* **322**, 1700–1707 (2010).

101. Group, D. P. P. R. Reduction of the incidence of type 2 diabetes with lifestyle intervention or metformin. *N. Engl. J. Med.* **34**, 162–163 (2002).

102. Seth, R., Kydd, A. S., Buchbinder, R., Bombardier, C. & Edwards, C. J. Allopurinol for chronic gout. *Cochrane Database Syst. Rev.* (2014) doi:10.1002/14651858.CD006077.PUB3.

103. Zheng, J. *et al.* Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat. Genet.* **52**, 1122–1131 (2020).

104. Morris, T. P., White, I. R. & Crowther, M. J. Using simulation studies to evaluate statistical methods. *Stat. Med.* **38**, 2074–2102 (2019).

105. Pietrosanu, M., Gao, J., Kong, L., Jiang, B. & Niu, D. Advanced algorithms for penalized quantile and composite quantile regression. *Comput. Stat.* **36**, 333–346 (2020).

106. Oehlert, G. W. A Note on the Delta Method. *Source Am. Stat.* **46**, 27–29 (1992).

107. White, H. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica* **48**, 817 (1980).

108. Allen, M. P. *Understanding Regression Analysis*. *Understanding Regression Analysis* (Springer US, 1997). doi:10.1007/978-0-585-25657-3_24.

109. Hemani, G. *et al.* The MR-base platform supports systematic causal inference across the human phenome. *Elife* **7**, (2018).

110. Hazra, A. Using the confidence interval confidently. *J. Thorac. Dis.* **9**, 4125 (2017).

111. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data.

Nature **562**, 203–209 (2018).

112. Backman, J. D. *et al.* Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* **599**, 628–634 (2021).

113. Halldorsson, B. V. *et al.* The sequences of 150,119 genomes in the UK Biobank. *Nature* **607**, 732–740 (2022).

114. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3 Genes, Genomes, Genet.* **1**, 457–470 (2011).

115. Galinsky, K. J. *et al.* Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. *Am. J. Hum. Genet.* **98**, 456–472 (2016).

116. Mitchell, R. E. *et al. UK Biobank Genetic Data: MRC-IEU Quality Control, version 2, 18/01/2019*.

117. Fry, D., Almond, R., Moffat, S., Gordon, M. & Singh, P. UK Biobank Biomarker Project Companion Document to Accompany Serum Biomarker Data. (2019).

118. UK Biobank Biomarker assay quality procedures: approaches used to minimise systematic and random errors (and the wider epidemiological implications). (2019).

119. Munafò, M. R., Tilling, K., Taylor, A. E., Evans, D. M. & Smith, G. D. Collider scope: when selection bias can substantially influence observed associations. *Int. J. Epidemiol.* **47**, 226–235 (2018).

120. Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).

121. Köttgen, A. *et al.* Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nat. Genet.* **45**, 145–154 (2013).

122. Nikpay, M. *et al.* A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).

123. UK Biobank — Neale lab. http://www.nealelab.is/uk-biobank/.

124. Goldthwaite, L. *Technical Report on C++ Performance*. https://www.open-std.org/Jtc1/SC22/wg21/docs/papers/2004/n1666.pdf (2004).

125. Loh, P. R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).

126. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

127. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).

128. DagumLeonardo & MenonRamesh. OpenMP. *IEEE Comput. Sci. Eng.* **5**, 46–55 (1998).

129. Guennebaud, G., Jacob, B. & others. Eigen v3. (2010).

130. Band, G. & Marchini, J. BGEN: a binary file format for imputed genotype and haplotype data. *bioRxiv* (2018) doi:10.1101/308296.

131. Millard, L. A. C., Munafò, M. R., Tilling, K., Wootton, R. E. & Davey Smith, G. MR-pheWAS with stratification and interaction: Searching for the causal effects of smoking heaviness identified an effect on facial aging. *PLOS Genet.* **15**, e1008353 (2019).

132. Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).

133. Evans, M. Robustness of size of tests of autocorrelation and heteroscedasticity to nonnormality. *J. Econom.* **51**, 7–24 (1992).

134. Yang, Q., Millard, L. A. C. & Davey Smith, G. Proxy gene-by-environment Mendelian randomization study confirms a causal effect of maternal smoking on offspring birthweight, but little evidence of long-term influences on offspring health. *Int. J. Epidemiol.* **49**, 1207–1218 (2020).

135. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

136. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. Ser. B (Statistical Methodol.* **82**, 1273–1300 (2020).

137. Cassidy, S., Chau, J. Y., Catt, M., Bauman, A. & Trenell, M. I. Cross-sectional study of diet, physical activity, television viewing and sleep duration in 233 110 adults from the UK Biobank; the behavioural phenotype of cardiovascular disease and type 2 diabetes. *BMJ Open* **6**, e010038 (2016).

138. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma. Appl. NOTE* **26**, 841–842 (2010).

139. Cunningham, F. *et al.* Ensembl 2022. *Nucleic Acids Res.* **50**, D988–D995 (2022).

140. Tweedie, S. *et al.* Genenames.org: The HGNC and VGNC resources in 2021. *Nucleic Acids Res.* **49**, D939–D946 (2021).

141. Võsa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* **53**, 1300–1310 (2021).

142. GTEx Portal. https://www.gtexportal.org/home/.

143. Paeratakul, S. *et al.* Measurement error in dietary data: Implications for the epidemiologic study of the diet-disease relationship. *Eur. J. Clin. Nutr.* **52**, 722–727 (1998).

144. Tooze, J. A., Troiano, R. P., Carroll, R. J., Moshfegh, A. J. & Freedman, L. S. A

Measurement Error Model for Physical Activity Level as Measured by a Questionnaire With Application to the 1999–2006 NHANES Questionnaire. *Am. J. Epidemiol.* **177**, 1199–1208 (2013).

145. Stender, S. *et al.* Adiposity Amplifies the Genetic Risk of Fatty Liver Disease Conferred by Multiple Loci. *Nat. Genet.* **49**, 842 (2017).

146. Viitasalo, A. *et al.* Associations of I148M variant in PNPLA3 gene with plasma ALT levels during 2-year follow-up in normal weight and overweight children: The PANIC Study. *Pediatr. Obes.* **10**, 84–90 (2015).

147. Döring, A. *et al.* SLC2A9 influences uric acid concentrations with pronounced sex-specific effects. *Nat. Genet.* **40**, 430–436 (2008).

148. Topless, R. K. *et al.* Association of SLC2A9 genotype with phenotypic variability of serum urate in pre-menopausal women. *Front. Genet.* **6**, (2015).

149. Abul-Husn, N. S. *et al.* A Protein-Truncating HSD17B13 Variant and Protection from Chronic Liver Disease. *NEJM* **378**, 1096–1106 (2018).

150. Stender, S. *et al.* Adiposity Amplifies the Genetic Risk of Fatty Liver Disease Conferred by Multiple Loci HHS Public Access Author manuscript. *Nat Genet* **49**, 842–847 (2017).

151. Ruth, K. S. *et al.* Using human genetics to understand the disease impacts of testosterone in men and women. *Nat. Med.* **26**, 252–258 (2020).

152. Gellert-Kristensen, H. *et al.* Combined Effect of PNPLA3, TM6SF2, and HSD17B13 Variants on Risk of Cirrhosis and Hepatocellular Carcinoma in the General Population. *Hepatology* **72**, 845–856 (2020).

153. Bayer, P. M., Hotschek, H. & Knoth, E. Intestinal alkaline phosphatase and the ABO blood group system--a new aspect. *Clin. Chim. Acta.* **108**, 81–87 (1980).

154. Nakano, T. *et al.* Involvement of intestinal alkaline phosphatase in serum apolipoprotein B-48 level and its association with ABO and secretor blood group types. *Biochem. Biophys. Res. Commun.* **341**, 33–38 (2006).

155. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).

156. Young, K. A. *et al.* Genome-Wide Association Study Identifies Loci for Liver Enzyme Concentrations in Mexican-Americans: The GUARDIAN Consortium. *Obesity (Silver Spring).* **27**, 1331 (2019).

157. Chambers, J. C. *et al.* Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat. Genet.* **43**, 1131–1138 (2011).

158. Rees, J. M. B., Foley, C. N. & Burgess, S. Factorial Mendelian randomization: using genetic variants to assess interactions. *Int. J. Epidemiol.* **49**, 1147–1158 (2020).

159. Yarmolinsky, J. *et al.* Circulating selenium and prostate cancer risk: A mendelian

randomization analysis. *J. Natl. Cancer Inst.* **110**, 1035–1038 (2018).

160. Grant, A. J., Gill, D., Kirk, P. D. W. & Burgess, S. Noise-augmented directional clustering of genetic association data identifies distinct mechanisms underlying obesity. *PLoS Genet.* **18**, e1009975 (2022).

161. Hemani, G., Bowden, J. & Davey Smith, G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum. Mol. Genet.* **27**, R195–R208 (2018).

162. Lawlor, D. A. Commentary: Two-sample Mendelian randomization: opportunities and challenges. *Int. J. Epidemiol.* **45**, 908 (2016).

163. Bowden, J. *et al.* Improving the accuracy of two-sample summary-data Mendelian randomization: Moving beyond the NOME assumption. *Int. J. Epidemiol.* **48**, 728–742 (2019).

164. Higgins, J. P. T. & Thompson, S. G. Controlling the risk of spurious findings from meta-regression. *Stat. Med.* **23**, 1663–1682 (2004).

165. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

166. MacArthur, J. A. L. *et al.* Workshop proceedings: GWAS summary statistics standards and sharing. *Cell Genomics* **1**, 100004 (2021).

167. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinforma. Appl. NOTE* **26**, 2190–2191 (2010).

168. Sinnott-Armstrong, N. *et al.* Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat. Genet. 2021 532* **53**, 185–194 (2021).

169. Pietzner, M. *et al.* Synergistic insights into human health from aptamer- and antibody-based proteomic profiling. *Nat. Commun. 2021 121* **12**, 1–13 (2021).

170. Sherry, S. T., Ward, M. & Sirotkin, K. dbSNP—Database for Single Nucleotide Polymorphisms and Other Classes of Minor Genetic Variation. *Genome Res.* **9**, 677–679 (1999).

171. bioforensics/rsidx: Library for indexing VCF files for random access searches by rsID. https://github.com/bioforensics/rsidx.

172. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–93 (2011).

173. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma. Appl. NOTE* **25**, 2078–2079 (2009).

174. Obenchain, V. *et al.* Sequence analysis VariantAnnotation: a Bioconductor package for exploration and annotation of genetic variants. **30**, 2076–2078 (2014).

175. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, (2004).

176. Huber, W. *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12**, 115–121 (2015).

177. pysam-developers/pysam: Pysam is a Python module for reading and manipulating SAM/BAM/VCF/BCF files. It's a lightweight wrapper of the htslib C-API, the same one that powers samtools, bcftools, and tabix. https://github.com/pysam-developers/pysam.

178. McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–303 (2010).

179. Malone, J. *et al.* Databases and ontologies Modeling sample variables with an Experimental Factor Ontology. **26**, 1112–1118 (2010).

180. Köhler, S. *et al.* Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Res.* **47**, D1018–D1027 (2019).

181. Medical Subject Headings - Home Page. https://www.nlm.nih.gov/mesh/meshhome.html.

182. broadinstitute/picard: A set of command line tools (in Java) for manipulating high-throughput sequencing (HTS) data and formats such as SAM/BAM/CRAM and VCF. https://github.com/broadinstitute/picard.

183. GenomicsDB/GenomicsDB: Highly performant data storage in C++ for importing, querying and transforming variant data with Java/Spark. Used in gatk4. https://github.com/GenomicsDB/GenomicsDB.

184. Mailman, M. D. *et al.* The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.* **39**, 1181 (2007).

185. Morales, J. *et al.* A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS Catalog. *Genome Biol.* **19**, 21 (2018).

186. bioinformed/vgraph: vgraph is a command line application and Python library to compare genetic variants using variant graphs. ``vgraph`` utilizes a graph representation of genomic variants in to precisely compare complex variants that are refractory to comparison by conventional comparison methods. https://github.com/bioinformed/vgraph.

187. Wilkinson, M. D. *et al.* Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 1–9 (2016).

188. Cezard, T. *et al.* The European Variation Archive: a FAIR resource of genomic variation for all species. *Nucleic Acids Res.* **50**, D1216–D1220 (2022).

189. Zheng, J. *et al.* Genetic effect modification of cis-acting C-reactive protein variants in

cardiometabolic disease status. *bioRxiv* (2021) doi:10.1101/2021.09.23.461369.

190. Imbens, G. W. Instrumental Variables: An Econometrician's Perspective. (2014) doi:10.3386/W19983.

191. Bollen, K. A. Instrumental Variables in Sociology and the Social Sciences. *https://doi.org/10.1146/annurev-soc-081309-150141* **38**, 37–72 (2012).

192. Liu, Y. *et al.* EpiGraphDB: A database and data mining platform for health data science. *Bioinformatics* **37**, 1304–1311 (2021).

193. UK Biobank project extended to Explore 3072 - Olink. https://www.olink.com/news/uk-biobank-pharma-proteomics-project-extended-to-explore-3072-platform/.

194. Xu, Z. M. & Burgess, S. Polygenic modelling of treatment effect heterogeneity. *Genet. Epidemiol.* **44**, 868–879 (2020).

195. Burgess, S. & Thompson, S. G. Use of allele scores as instrumental variables for Mendelian randomization. *Int. J. Epidemiol.* **42**, 1134–1144 (2013).

196. Higgins, J. P. T. *et al.* Synthesising quantitative evidence in systematic reviews of complex health interventions. *BMJ Glob. Heal.* **4**, e000858 (2019).

197. Van Houwelingen, H. C., Zwinderman, K. H. & Stijnen, T. A bivariate approach to meta-analysis. *Stat. Med.* **12**, 2273–2284 (1993).

198. Brugger, S. P. & Howes, O. D. Heterogeneity and Homogeneity of Regional Brain Structure in Schizophrenia: A Meta-analysis. *JAMA psychiatry* **74**, 1104–1111 (2017).

199. Pillinger, T. *et al.* A Meta-Analysis of Immune Parameters, Variability, and Assessment of Modal Distribution in Psychosis and Test of the Immune Subgroup Hypothesis. *Schizophr. Bull.* **45**, 1120–1133 (2019).

200. Gewandter, J. S. *et al.* Demonstrating Heterogeneity of Treatment Effects Among Patients: An Overlooked but Important Step Toward Precision Medicine. *Clin. Pharmacol. Ther.* **106**, 204–210 (2019).

201. Winkelbeiner, S., Leucht, S., Kane, J. M. & Homan, P. Evaluation of Differences in Individual Treatment Response in Schizophrenia Spectrum Disorders: A Meta-analysis. *JAMA Psychiatry* **76**, 1063–1073 (2019).

202. Lawlor, D. A., Tilling, K. & Smith, G. D. Triangulation in aetiological epidemiology. *Int. J. Epidemiol.* **45**, 1866–1886 (2016).

203. Wallace, C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLOS Genet.* **17**, e1009440 (2021).

204. Hemani, G., Theocharidis, A., Wei, W. & Haley, C. EpiGPU: Exhaustive pairwise epistasis scans parallelized on consumer level graphics cards. *Bioinformatics* **27**, 1462–1465 (2011).

205. The HDF Group. Hierarchical Data Format, version 5.