- 1 Hypothesis-free detection of gene-interaction effects on
- 2 biomarker concentration in UK Biobank using variance

3 prioritisation

- 4 Matthew S. Lyon (0000-0002-2500-1013),^{1,2}* Louise A. C. Millard,² George Davey Smith,^{1,2} 5 Tom R. Gaunt,^{1,2}⁺ Kate Tilling^{1,2}⁺ 6 7 8 1. National Institute for Health Research Bristol Biomedical Research Centre, University of 9 Bristol, Oakfield House, Bristol, BS8 2BN, UK 10 2. Medical Research Council (MRC) Integrative Epidemiology Unit (IEU), Bristol Medical School (Population Health Sciences), University of Bristol, Oakfield House, Bristol, BS8 11 2BN, UK 12 13 14 ⁺ These authors contributed equally to this work 15 * To whom correspondence should be addressed. Tel: +44 (0) 117 331 4094; Email:
- 16 <u>matt.lyon@bristol.ac.uk;</u> Twitter: @matt_s_lyon

17 Abstract

18

19	Blood biomarkers include disease intervention targets that may interact with genetic and
20	environmental factors resulting in subgroups of individuals who respond differently to
21	treatment. Such interactions may be observed in genetic effects on trait variance. Variance
22	prioritisation is an approach to identify genetic loci with interaction effects by estimating
23	their association with trait variance, even where the modifier is unknown or unmeasured.
24	Here, we develop and evaluate a regression-based Brown-Forsythe test and variance effect
25	estimate to detect such interactions. We provide scalable open-source software (varGWAS)
26	for genome-wide association analysis of SNP-variance effects
27	(https://github.com/MRCIEU/varGWAS) and apply our software to 30 blood biomarkers in
28	UK Biobank. We find 468 variance quantitative trait loci across 24 biomarkers and follow up
29	findings to detect 82 gene-environment and six gene-gene interactions independent of
30	strong scale or phantom effects. Our results replicate existing findings and identify novel
31	epistatic effects of TREH rs12225548 x FUT2 rs281379 and TREH rs12225548 x ABO
32	rs635634 on alkaline phosphatase and ZNF827 rs4835265 x NEDD4L rs4503880 on gamma
33	glutamyltransferase. These data could be used to discover possible subgroup effects for a
34	given biomarker during preclinical drug development.
35	

36 Introduction

37

38 Blood biomarkers provide valuable information for diagnosis and prognosis of disease¹,

insight into biological mechanisms², and a source of causal modifiable risk factors which may

40 be intervened upon to create therapies¹. For example, lipids, glucose, and urate have

41	become successful therapeutic targets for cardiovascular disease ³ , type 2 diabetes ⁴ , and
42	gout ⁵ , respectively, among others. However, as biomarkers are complex traits they are
43	affected by genetic and environmental factors which may interact producing gene-gene
44	(GxG, epistasis) or gene-environment (GxE) effects ⁶ . Intervening on biomarkers which have
45	an interaction effect on disease outcome will produce subgroup effects with individual
46	variation in response to treatment dependent on the modifier ⁷ . Identifying these
47	interactions may contribute to stratified medicine which aims to provide optimum
48	treatments and preventative advice for disease based on individual characteristics ⁶ .
49	
50	Detecting interaction effects can be challenging. Statistical power to detect an interaction is
51	lower than for main effects; for randomised control trials the sample size needed to detect
52	an interaction with equal sized subgroups is around four times the size needed to detect the
53	main effect of the same magnitude ^{7,8} . Low power is exacerbated by multiple testing
54	correction that is essential to account for evaluating the large numbers of candidate
55	modifiers. To reduce multiple testing, pairwise interaction analyses of SNPs with moderate
56	main effects can be performed. However, this approach could miss subgroups with an effect
57	in only one group or opposing effect directionality (known as qualitative interaction effects ⁷)
58	hence weaker overall effects, yet these offer the most potential for stratified medicine. An
59	alternative approach to select SNPs for GxG/GxE testing is variance prioritisation ^{9,10} which
60	identifies differences in outcome variance across genotype levels (variance quantitative trait
61	loci, vQTL). Although not conclusive evidence, this observation is consistent with a SNP-
62	interaction effect ¹¹ and detection of vQTLs does not require the modifier to be measured ¹¹ .
63	

64	Variance QTLs can arise as a consequence of heterogeneous mean effects that could occur
65	from changing environment, background genetics and temporal regulation ¹¹ . Among the
66	first reported vQTL effects in humans was rs7202116 (FTO locus), associated with a large
67	change in variance (as well as mean) of body mass index (BMI) ¹² . More recently, systematic
68	testing of vQTL effects on 13 quantitative traits in UK Biobank and subsequent GxE testing
69	identified 16 GxE effects modified by age, sex, physical activity, sedentary behaviour, and
70	smoking ¹³ . Variance QTLs have also been identified for gene expression ¹⁴ , DNA
71	methylation ¹⁵ , Vitamin D ¹⁶ and facial morphology ¹⁷ . To date, gene-interaction studies have
72	mostly focused on testing a small number of candidate interactions, but hypothesis-free
73	testing of vQTL effects on blood biomarkers could lead to the identification of unanticipated
74	intervention targets with subgroup effects.
75	
76	Existing studies of vQTLs have employed a range of methods ^{13, 18–20} . Wang <i>et al</i> compared
77	the power and type I error of four widely used variance tests and found the median variant
78	of Levene's test ²¹ also known as the Brown-Forsythe test ^{13,22} to be most robust. However,
79	this test does not allow for inclusion of covariates or continuous genotype data (i.e.,
80	imputed allelic dose) and does not provide an effect estimate, all of which are limitations
81	when applied in a GWAS. However, the Brown-Forsythe test can be reformulated using
82	least-absolute deviation 15,23 (LAD) regression using the same structure as the Glejser test 24 .
83	Regression-based variance tests offer greater flexibility to overcome these limitations.
83 84	Regression-based variance tests offer greater flexibility to overcome these limitations. Recent developments in LAD regression have vastly reduced the computational burden for

- 87 In this study we compare the utility of the original Brown-Forsythe test and our LAD
- regression-based reformulation of the Brown-Forsythe test (LAD-BF) to detect SNP-
- 89 interaction effects under simulation and develop scalable open-source software
- 90 (https://github.com/MRCIEU/varGWAS) for performing variance GWAS using the latter. We
- 91 apply our regression-based model to estimate SNP effects on the variance of 30 blood
- 92 biomarkers in ~337K UK Biobank participants and follow up vQTLs with formal interaction
- 93 tests to detect GxG and GxE interactions.
- 94
- 95 Material and methods
- 96
- 97 Original Brown-Forsythe test
- 98
- 99 The Brown-Forsythe²² test (median variant of Levene's test²¹) refers to the original
- 100 published non-parametric test and will be used throughout. We applied the Brown-Forsythe
- 101 test to detect differences in trait variability across the three genotypic groups.
- 102
- 103 The test statistic W is F-distributed F(2, N 3) given by:
- 104

$$W = \frac{(N-3)}{2} * \frac{\sum_{i=0}^{2} N_i (\bar{X}_i - \bar{X})^2}{\sum_{i=0}^{2} \sum_{j=1}^{N_i} (X_{ij} - \bar{X}_i)^2}$$

105

106 Where N is the total number of observations. N_i is the number of observations with the *i*th 107 genotype group {0, 1, 2}. X_{ij} is the absolute residual of the outcome for the *j*th observation

- in the *i*th genotype group from the median. \overline{X}_i is the mean of X_{ij} for the *i*th genotype group
- 109 and \overline{X} is the mean of X_{ij} across genotype groups.

110

- 111 All analyses of the original Brown-Forsythe test used the omic-data-based complex trait
- analysis (OSCA) software package^{13,26} which additionally produces a variance effect estimate
- derived from the test P-value assuming linearity between the SNP and outcome variance²⁷.

114

115 LAD-BF test

116

117 Our reformulated regression-based Brown-Forsythe test uses LAD regression of outcome Y

118 on independent variable *X* to estimate the residuals adjusting for any covariates:

119

$$Y = \widehat{\beta_0} + \widehat{\beta_1}X + \widehat{U}$$

120

121 Where X is the genotype measured by continuous (expected value from genotype 122 imputation) or ordinal (directly genotyped) variable and \widehat{U} is the residual of this first-stage

123 model.

124

125 A second-stage ordinary least squares (OLS) model regressed the absolute residuals $|\hat{U}|$ of

126 the first-stage model on the genotype values coded as dummy variables (genotype expected

127 values were rounded to the nearest whole number resulting in some loss of precision)

including any covariates given in the first-stage model:

$$|U| = \hat{\gamma}_0 + \hat{\gamma}_1 X 1 + \hat{\gamma}_2 X 2 + \hat{E}$$

130

- 131 The test P-value was estimated from an F-test comparing the second-stage residual sum of
- 132 squares to an intercept-only model to test the null hypothesis of variability homogeneity
- across genotypes.
- 134

135 SNP effects on trait variance were calculated from second-stage regression coefficients

- 136 which are estimates of mean-absolute deviation. This transformation assumes trait
- 137 normality.
- 138 The var(Y|G == 1) was estimated using:

$$2 * \hat{\gamma_0} + \hat{\gamma_1} + \hat{\gamma_1^2} / (\frac{2}{pi})$$

139 The var(Y|G == 2) was estimated using:

$$2 * \hat{\gamma_0} + \hat{\gamma_2} + \hat{\gamma_2}^2 / (\frac{2}{pi})$$

140 The standard error of the variance effect was estimated using the delta method²⁸ and

141 heteroscedastic-consistent standard errors for the second-stage model coefficients²⁹.

142

143 LAD regression was implemented using the majorise-minimisation^{25,30} (MM) model with

default values for iterations (200) and tolerance (0.001) and first-stage OLS regression

145 coefficients provided as initial values.

146

147 Software

148

149 The LAD-BF test was implemented in varGWAS available in C++ v1.2.3 and R v1.0.0 (refer to

the code and data availability section). The MM model used functionality from the cqrReg R-

151	package ²⁵ (<u>https://cran.r-project.org/web/packages/cqrReg/index.html</u>). OLS and general
152	matrix functionality were provided with Eigen v3.4.0 ³¹ . BGEN file processing used the BGEN
153	library ³² v1.1.6.
154	
155	The original Brown-Forsythe test used the OSCA software package v0.46 ^{13,26} . Simulations
156	and follow up UK Biobank analyses were performed using R v3.6.0.
157	
158	Simulations
159	
160	The bias and statistical power of the two Brown-Forsythe tests were evaluated through a
161	series of simulation studies reported using the ADEMP structure ³³ (Table 1 & Supplemental
162	Material and Methods).
163	
163 164	Participants
163 164 165	Participants
163 164 165 166	Participants UK Biobank is a large prospective cohort study of approximately 500,000 UK participants
163 164 165 166 167	Participants UK Biobank is a large prospective cohort study of approximately 500,000 UK participants aged 37-73 at recruitment ³⁴ . Recruitment took place between 2006-2010 from across the
163 164 165 166 167 168	Participants UK Biobank is a large prospective cohort study of approximately 500,000 UK participants aged 37-73 at recruitment ³⁴ . Recruitment took place between 2006-2010 from across the UK. Measures were collected on lifestyle, socio-demographics, physical parameters, health-
163 164 165 166 167 168 169	Participants UK Biobank is a large prospective cohort study of approximately 500,000 UK participants aged 37-73 at recruitment ³⁴ . Recruitment took place between 2006-2010 from across the UK. Measures were collected on lifestyle, socio-demographics, physical parameters, health- related factors, and biological samples for genetic testing and biomarker measurements.
163 164 165 166 167 168 169 170	Participants UK Biobank is a large prospective cohort study of approximately 500,000 UK participants aged 37-73 at recruitment ³⁴ . Recruitment took place between 2006-2010 from across the UK. Measures were collected on lifestyle, socio-demographics, physical parameters, health- related factors, and biological samples for genetic testing and biomarker measurements. Ethical approval for the UK Biobank study was granted by the National Research Ethics
163 164 165 166 167 168 169 170	Participants UK Biobank is a large prospective cohort study of approximately 500,000 UK participants aged 37-73 at recruitment ³⁴ . Recruitment took place between 2006-2010 from across the UK. Measures were collected on lifestyle, socio-demographics, physical parameters, health- related factors, and biological samples for genetic testing and biomarker measurements. Ethical approval for the UK Biobank study was granted by the National Research Ethics Service (NRES) Committee North West (ref 11/NW/0382). All analyses were performed
163 164 165 166 167 168 169 170 171	Participants UK Biobank is a large prospective cohort study of approximately 500,000 UK participants aged 37-73 at recruitment ³⁴ . Recruitment took place between 2006-2010 from across the UK. Measures were collected on lifestyle, socio-demographics, physical parameters, health- related factors, and biological samples for genetic testing and biomarker measurements. Ethical approval for the UK Biobank study was granted by the National Research Ethics Service (NRES) Committee North West (ref 11/NW/0382). All analyses were performed under approved UK Biobank project 15825 (dataset ID 33352).

174 Genetic data

175

176	Genetic array data were available on 488,377 participants measured using a combination of
177	UK Biobank Axiom TM array (n=438,398) and UK BiLEVE array (n=49,979). Genotype
178	imputation was performed using a reference set combined with UK10K haplotypes and HRC
179	reference panels with the IMPUTE2 ³⁵ software as described ³⁶ . The following SNPs were
180	removed from analysis leaving a total of 6,812,700: multi-allelic loci, minor allele frequency
181	< 5%, Hardy-Weinberg violations (P < 1 x 10^{-5}), genotype missing rate >5%, low imputation
182	score (INFO < 0.3) and HLA locus (hg19/GRCh37 chr6:23477797-38448354).
183	
184	Quality control
185	
186	We applied standard exclusion criteria (Figure S1) to remove genotype-phenotype sex
187	mismatches, aneuploidies, and outliers for missingness or heterozygosity as previously
188	described ³⁶ leaving n=486,565 participants. To ensure data independence, closely related
189	subjects were removed as described elsewhere ³⁶ leaving n=407,176 participants. Finally,
190	'non-white British' participants defined using published methodology ³⁶ were removed to
191	avoid confounding by population stratification providing a total sample size of n=377,076.
192	
193	Phenotypes
194	
195	UK Biobank measures of 30 serum biochemistry markers were available for approximately
196	500k participants. Each measure was chosen based on being an established risk factor for
197	disease, a clinical diagnostic measure or because it characterises a phenotype that is not

198	well assessed by other approaches as described in the UK Biobank documentation ³⁷ .
199	Quantification and quality control was performed as previously described ³⁷ . Total physical
200	activity was calculated by summing self-reported duration of walking, moderate and
201	vigorous activity collected using the International Physical Activity Questionnaire as
202	described ³⁸ . For each analysis participants with missing data were removed. All continuous
203	outcomes were SD normalised.
204	
205	Genome-wide association studies (GWAS)
206	
207	GWAS of biomarker variability were performed using our LAD regression-based Brown-
208	Forsythe test adjusted for age, sex, and the top ten genetic principal components in first and
209	second regression models. We removed outlier biomarker values with a Z-score > 5SD from
210	the mean to control type I error inflation as previously described ¹³ . Quality control was
211	undertaken visually using Q-Q plots to check for a departure of P-value distribution from
212	that expected under the null. Independent vQTLs were identified by clumping GWAS loci
213	that passed the experiment-wise genome-wide evidence threshold P < 1.67 $ imes$ 10 ⁻⁹
214	(Bonferroni correction of standard GWAS threshold: $p = 5 \times 10^{-8} / 30$) using the OpenGWAS
215	API ³⁹ with default R ² threshold of 0.001 and 1000 genomes phase 3 European ancestry ⁴⁰ .
216	
217	Gene interaction test
218	
219	Independent vQTLs (see above) were tested for interaction effects on additive and
220	multiplicative scales using heteroscedasticity-consistent standard errors ²⁹ adjusted for age,

sex, and top ten genetic principal components. To ensure effects were robust to phantom

222	effects ⁴¹ , we performed sensitivity analyses adjusting for fine-mapped main effects
223	identified using SuSiE ⁴² (Supplemental Material and Methods). Interactions surpassing
224	genome-wide association significance (P < 5 x 10^{-8}) on additive and multiplicative scales that
225	did not strongly attenuate with adjustment for fine-mapped main effects were prioritised
226	for subgroup analyses. GxG effects were identified through interaction testing with
227	independent (R^2 < 0.001) vQTLs excluding pairwise combinations of vQTLs within a 10Mb
228	window as previously described ¹³ . GxE testing was performed using candidate modifiers:
229	age, sex, body mass index, alcohol intake, smoking status, physical activity, daily sugar
230	intake, and daily fat intake.
231	
232	Subgroup analyses
233	
234	Subgroup effects of top interaction effects were presented by estimating the SNP effect on
235	the outcome stratified by modifier using heteroscedasticity-consistent standard errors ²⁹
236	adjusted for age, sex and top ten genetic principal components. Modifiers were rounded
237	genetic dosage values or prepared by dichotomisation as follows: below or above the
238	median value for continuous variables (group [G] 1, below median; G2, median or greater),
239	ever (G1) vs never (G2) smoker, alcohol intake once a week or more (G1) vs less than once a
240	week on average (G2), males (G1) vs females (G2). Subgroup effects are presented along
241	with the SNP-variance estimates adjusted for age, sex and top ten genetic principial
242	components with and without adjustment for the interaction term (variance effects were
243	not adjusted for sex when sex was the modifier).

245 Gene annotation

- 246
- 247 Variance QTLs were annotated with the nearest gene using the closest function of
- 248 bedtools⁴³ (v2.3.0) and Ensembl v104 (GRCh37) protein-coding features which were filtered
- 249 to retain HUGO⁴⁴ valid identifiers. The following annotations were recoded based on
- 250 expression QTL evidence^{45,46}: rs4530622 *SLC2A9*, rs11244061 *ABO*, rs71633359 *HSD17B13*,
- 251 rs28413939 TREH, rs281379 FUT2, rs635634 ABO, rs964184 APOA5.
- 252
- 253 Results
- 254

255 Simulated power and type I error to detect interaction effects by change in variance256

- 257 The power to detect a difference in trait variability due to an interaction effect was low and
- 258 equivalent for both methods (Figure S2). Suppose a SNP has a main effect on a normally
- distributed outcome detectable with 80% power, then 10x the sample size needed to detect
- the main effect was required to detect the interaction with only 50% power assuming the
- 261 interaction was half the size of the main effect. Positive skew and kurtosis reduced power.
- 262 Both methods had equally well controlled type I error (Figure S3).
- 263

264 Simulated variance effect estimate and confidence interval coverage

- 265
- 266 Under a simulated linear effect of genotype on outcome variance both methods gave the
- correct effect estimate and 95% confidence interval coverage (Figure 1). However, when the
- 268 difference in variance was a consequence of an interaction effect, the relationship between

269	the genotype and outcome variance was non-linear and dependent on the modifier. Under
270	these conditions, the variance effect estimate produced using $OSCA^{26,27}$ from the Brown-
271	Forsythe test P-value gave the incorrect effect size while LAD-BF produced the correct
272	estimate albeit with slightly elevated coverage.
273	
274	Adjusting the LAD-BF test for an interaction effect through simulation
275	
276	We simulated an interaction effect and compared the LAD-BF test P-value distributions with
277	and without adjusting for the simulated interaction (Figure S4). Including the interaction
278	term in the first-stage regression model completely attenuated the variance test statistic.
279	After identifying an interaction at a variance locus this approach could be applied to
280	determine if additional strong interaction effects exist and could be used in a stepwise
281	regression fashion until all interaction effects are identified.
282	
283	Runtime performance
284	
285	Increasing the number of CPU threads reduced the total runtime of both methods to
286	process 1000 SNPs (Figure S5). For the C++ implementation of LAD-BF in varGWAS, the
287	lowest average runtime was 13.6 second (95% CI 13.5, 13.7) using four threads of an Intel(R)
288	Xeon(R) CPU E5-2680 v4 @ 2.40GHz. Under the same conditions, the original Brown-
289	Forsythe test implemented in OSCA was 1.78x faster (7.61 seconds [95% CI 7.60, 7.63]).
290	
291	GWAS of variance effects in UK Biobank
292	

293	We identified 468 independent (R ² < 0.001) vQTLs influencing 24 biomarkers (Figure S6,
294	Figure S7 & Table S1) using an experiment-wise P-value threshold of 1.67 x 10^{-9} (5 x 10^{-8} /
295	30) and no variance effects for albumin, calcium, oestradiol, phosphate, rheumatoid factor,
296	or total protein. Oestradiol and rheumatoid factor were measured on a subset of n=76,674
297	and n=41,315 participants respectively and therefore were less well powered to detect
298	effects. Of these vQTLs, 270 (57.7%) had suggestive evidence for a variance effect on the log
299	scale (P < 5 x 10^{-5}) and 453 (96.8%) had a mean effect (P < 5 x 10^{-8}). The low concordance
300	between natural and log scales and high concordance between mean and variance effects
301	suggests the presence of mean-variance relationships which is a likely consequence of
302	extreme non-normality for some of the trait distributions (Figure S8).
303	
304	Gene-environment interaction effects (GxE)
305	
306	We detected 139 additive and 104 multiplicative GxE effects (P < 5 x 10^{-8} ; Figure S9 & Figure
307	S10). Adjusting the additive effects for fine-mapped main effects (Figure S11) led to a small
308	increase in <i>UGT1A8</i> rs2741047 x sex on direct bilirubin to 0.037 SD (95% Cl 0.032, 0.042)
309	from 0.028 SD (95% CI 0.023, 0.033) and minor attenuation of <i>MAP3K4</i> rs1247295 x sex on
310	lipoprotein a to -0.011 SD (95% Cl -0.015, -0.007) from -0.016 SD (95% Cl -0.021, -0.010).
311	These findings could reflect the presence of large main effects in imperfect linkage
212	0
512	disequilibrium with the index SNP which is known to inflate/deflate test statistics ⁴¹ .
313	disequilibrium with the index SNP which is known to inflate/deflate test statistics ⁴¹ .
313 314	disequilibrium with the index SNP which is known to inflate/deflate test statistics ⁴¹ . We prioritised 82 GxE effects with evidence on both scales (P < 5 x 10^{-8}) to avoid spurious
313 314 315	disequilibrium with the index SNP which is known to inflate/deflate test statistics ⁴¹ . We prioritised 82 GxE effects with evidence on both scales ($P < 5 \times 10^{-8}$) to avoid spurious interactions dependent on scale (Table S2). Of these BMI (n=35), sex (n=27) and age (n=17)

317	tested for interaction by physical activity, and sugar and fat intake but identified little
318	evidence of interactions. The largest effects (Figure 2) were: PNPLA3 rs738409 x BMI on
319	alanine aminotransferase (ALT; 0.08 SD [95% CI 0.08, 0.09]), SLC2A9 rs938555 x sex on urate
320	(-0.08 SD [95% CI -0.09, -0.08]), APOE rs1065853 x sex on low-density lipoprotein (LDL; 0.06
321	SD [95% CI 0.05, 0.07]), SHBG rs1799941 x sex on testosterone (0.06 SD [95% CI 0.06, 0.06])
322	and <i>TM6SF2</i> rs58542926 x BMI on ALT (0.05 SD [95% CI 0.04, 0.06]). Adjusting the variance
323	effect for the interaction term (Figure 2) led to attenuation of PNPLA3 rs738409 and
324	TM6SF2 rs58542926 on ALT and SHBG rs1799941 on testosterone but strong variance
325	effects on ALT remained at <i>PNPLA3</i> rs738409 (LAD-BF P_adjust = 1.0×10^{-73}) and <i>TM6SF2</i>
326	rs58542926 (LAD-BF P_adjust = 1.84 x 10 ⁻⁸). There was no strong variance attenuation of
327	APOE rs1065853 on LDL or SLC2A9 rs938555 on urate following adjustment for the
328	interaction (Figure 2).
329	
330	Gene-gene interaction effects (GxG)
331	
332	We detected eight GxG effects on the additive scale (P < 5 x 10^{-8} ; Figure S12), six of which
333	were also associated on the multiplicative scale (P < 5 x 10^{-8} ; Figure S13). There was no

strong attenuation following adjustment for fine-mapped main effects (Figure S14)

suggesting phantom epistasis^{41,47} was not a major source of bias. *ZNF827* rs4835265 x

336 NEDD4L rs4503880 was inversely associated with -0.04 SD (95% CI -0.05, -0.03) gamma

337 glutamyltransferase (GGT), ABO rs635634 x FUT2 rs281379, ABO rs635634 x TREH

338 rs12225548, and TREH rs12225548 x FUT2 rs281379 were associated with 0.08 SD (95% CI

339 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

340 phosphatase (ALP) respectively, HSD17B13 rs71633359 x PNPLA3 rs738409 and HSD17B13

341 rs71633359 x PNPLA3 rs3747207 were associated with -0.04 SD (95% CI -0.05, -0.03) and -

342 0.04 SD (95% CI -0.05, -0.03) decrease in ALT and aspartate aminotransferase (AST)

respectively (Figure 3). Adjusting the variance effects for the interaction term had no strong

- impact on the variance estimate (Figure 3).
- 345
- 346 Discussion
- 347
- 348 Here we demonstrate the value of variance GWAS in identifying 468 independent vQTLs
- 349 with evidence of interaction on 24 serum biochemistry phenotypes in UK Biobank and
- subsequently identify 82 GxE and six GxG scale independent effects. To facilitate this large-

351 scale analysis on ~337K UK Biobank participants we developed an efficient C++

- implementation of a LAD regression-based Brown-Forsythe test²² (implemented in
- varGWAS) with functionality to reliably estimate variance effects and compared the test
- with the original non-parametric version (implemented in OSCA^{13, 26}) through a series of
- 355 simulations.

356

357 Although the power to detect genetic interaction effects using variance prioritisation was 358 low, when applied to large sample sizes such as UK Biobank strong evidence for association can be identified as demonstrated in this study and by Wang *et al*¹³. We found LAD-BF had 359 360 several advantages over the original non-parametric test when applied to GWAS. First, LAD-361 BF directly supports adjustment for covariates (although this could be achieved using the original test if applied to pre-adjusted phenotypes¹³). Second, LAD-BF can test effects of 362 363 continuous genotypes which enables application to the expected genotype value ("dose") 364 from imputed SNP array data. Third, our model provides a variance effect estimate which is

365	valid when there is a SNP interaction effect, unlike the implementation of the original
366	Brown-Forsythe test in OSCA which provides an incorrect variance effect estimate derived
367	from the test P-value ²⁷ . We also demonstrate through simulation that adjusting the variance
368	effect for the interaction term causes attenuation which is useful to determine if other
369	interactions exist and could potentially be applied using stepwise regression until all
370	interaction effects are discovered, subject to sufficient power. However, there are some
371	disadvantages. The runtime was 75% longer than the original test implemented in OSCA,
372	although this is still fast enough to allow large-scale analyses. Second, the effect estimate
373	(but not test statistic) is based on normality assumptions which may be violated in practice.
374	
375	The largest GxE effects replicate existing findings: <i>PNPLA3</i> rs738409 x BMI on ALT levels ^{48,49} ,
376	<i>SLC2A9</i> rs938555 x sex on urate ⁵⁰ , <i>APOE</i> rs1065853 x sex on LDL ⁵¹ , <i>SHBG</i> rs1799941 x sex on
377	testosterone ⁵² , and <i>TM6SF2</i> rs58542926 x BMI on ALT ⁴⁸ . Adjusting the variance effect for
378	the interaction led to attenuation of <i>PNPLA3</i> rs738409 and <i>TM6SF2</i> rs58542926 on ALT and
379	SHBG rs1799941 on testosterone, however strong evidence of variance effects remained for
380	ALT at PNPLA3 rs738409 and TM6SF2 rs58542926 suggesting other interaction effects may
381	exist at these loci. The variance effect of SHBG rs1799941 on testosterone was weak after
382	adjusting for rs1799941 x sex suggesting no strong evidence of further interaction effects on
383	testosterone at this locus, but the test may be underpowered to detect additional effects.
384	
385	We replicated previous GxG effects of <i>ABO</i> rs635634 x <i>FUT2</i> rs281379 on ALP ^{53,54} and
386	<i>HSD17B13</i> rs71633359 x <i>PNPLA3</i> rs738409/rs3747207 on ALT and AST ^{55,56} and find no
387	strong evidence of 'phantom epistasis' ^{41,47} as a potential explanation. Additionally, we

388 identified novel effects of TREH rs12225548 x FUT2 rs281379 and ABO rs635634 x TREH

389	rs12225548 on ALP and ZNF827 rs4835265 x NEDD4L rs4503880 on GGT. ABO blood group
390	antigens and secretion status are thought to influence ALP clearance ^{57,58} . <i>TREH</i> rs12225548
391	has a strong main effect on ALP $^{ m 39,59,60}$ and interactions of these loci may be explained by
392	interplay of ALP production and clearance mechanisms. ZNF827 and NEDD4L loci have
393	previously been reported to influence GGT levels in independent populations but the
394	mechanism is unclear ^{61,62} .
395	
396	None of the GxG loci variance effects strongly attenuated after adjusting for the interaction
397	term. This could be a consequence of low power since the interaction effect likely explains a
398	very small amount of the trait variance but could also indicate the presence of other
399	interaction effects involving the same SNP not included in the variance model. Indeed, we
400	found strong GxE evidence at some of these loci: ABO rs635634 x sex on ALP, HSD17B13
401	rs71633359 x BMI and PNPLA3 rs738409/rs3747207 x BMI on ALT and AST.
402	
403	Evidence of gene-interaction effects could suggest the protein product also has an
404	interaction effect. In which case interventions developed to target the protein will show
405	differential effects on the indication and could have low or no efficacy in some subgroups ⁶³ .
406	Such evidence could be important to support developments in stratified medicine.
407	Therefore, vQTL evidence may have a role in preclinical drug development to deprioritise
408	targets given the possibility of target-outcome heterogeneous effects. Further research is
409	needed to appraise the utility of vQTLs in the drug development pipeline as has been done
410	for protein QTLs ⁶⁴ .

412	However, there are other explanations for vQTLs that are not in terms of biology. First, loci
413	that are weakly correlated with a SNP having a strong main effect can introduce a phantom
414	vQTL ^{65,66} . In this situation variance is introduced through variability in LD between the
415	artefactual vQTL and QTL. Second, vQTLs could signify fluctuation of a trait measurement
416	within an individual over time ¹⁰ and may originate from normal biological processes such as
417	circadian rhythm. Third, we assume homogeneity of variance within each genotype group
418	which could be violated by the mean-variance relationship and observed low concordance
419	of vQTL effects on the log and natural scales are evidence for this. Additionally, our
420	interactions could be explained by non-linear relationships between the exposure and
421	outcome or scale artefacts ⁶⁷ . We sought to reduce the latter by replicating effects on
422	additive and multiplicative scales.
423	
424	Through this work we performed hypothesis-free analyses of genetic interaction effects on
425	30 blood biomarkers in UK Biobank using variance prioritisation and found evidence for 88
426	effects. Many of our top findings replicate previously reported associations, but we also
427	report first evidence of TREH rs12225548 x FUT2 rs281379 and TREH rs12225548 x ABO
428	rs635634 on ALP and ZNF827 rs4835265 x NEDD4L rs4503880 on GGT. Additionally, we
429	show variance attenuation of PNPLA3 rs738409 and TM6SF2 rs58542926 on ALT and SHBG
430	rs1799941 on testosterone after adjusting for the interaction indicating these effects were
431	contributing to the variance association, but the ALT effects were still strong suggesting
432	additional interactions may exist at these loci. These data could be used to discover possible
433	subgroup effects for a given biomarker during preclinical drug development. To facilitate our
434	analysis, we developed C++ variance GWAS software that implements a LAD-regression
435	based Brown-Forsythe test, provide a convenient R-package based on this software and

- 436 introduce methodology to estimate the variance effects which can be applied to other
- 437 studies.
- 438
- 439 Supplemental information description
- 440
- 441 Supplemental Material and Methods
- 442 Figure S1. UK Biobank participant inclusion criteria
- 443 Figure S2. Power to detect SNP-interaction effects using variance testing under simulation
- 444 Figure S3. Type I error of Brown-Forsythe tests
- 445 Figure S4. Effect of adjustment for the interaction effect on variance test P-value
- 446 distribution
- 447 Figure S5. Runtime performance of varGWAS and OSCA
- 448 Figure S6. Manhattan plots of biomarker variance GWAS using regression-based Brown-
- 449 Forsythe test
- 450 Figure S7. Q-Q plots of biomarker variance GWAS using regression-based Brown-Forsythe
- 451 test
- 452 Figure S8. Biomarker distribution
- 453 Figure S9. Top gene-by-environment interaction effects (P < 5 x 10⁻⁸) on biomarker
- 454 concentration using additive scale
- 455 Figure S10. Top gene-by-environment interaction effects ($P < 5 \times 10^{-8}$) on biomarker
- 456 concentration using multiplicative scale
- 457 Figure S11. Top gene-by-environment interaction effects ($P < 5 \times 10^{-8}$) on biomarker
- 458 concentration using additive scale adjusted for fine-mapped main effect

- 459 Figure S12. Top gene-by-gene interaction effects (P < 5 x 10⁻⁸) on biomarker concentration
- 460 using additive scale
- 461 Figure S13. Top gene-by-gene interaction effects ($P < 5 \times 10^{-8}$) on biomarker concentration
- 462 using multiplicative scale
- 463 Figure S14. Top gene-by-gene interaction effects ($P < 5 \times 10^{-8}$) on biomarker concentration
- 464 using additive scale adjusted for fine-mapped main effects
- 465 Table S1. GWAS summary statistics for top vQTLs identified through this study
- 466 Table S2. Top GxG/GxE effect summary statistics
- 467 Table S3. Fine-mapped loci covariates
- 468
- 469 Declaration of Interests
- 470
- 471 T.R.G receives funding from Biogen for unrelated research. K.T has been paid for
- 472 consultancy for CHDI.
- 473
- 474 Acknowledgements
- 475
- 476 This study was funded by the NIHR Biomedical Research Centre at University Hospitals
- 477 Bristol and Weston NHS Foundation Trust and the University of Bristol. The views expressed
- are those of the author(s) and not necessarily those of the NIHR or the Department of
- 479 Health and Social Care. This work was also funded by the UK Medical Research Council as
- 480 part of the MRC Integrative Epidemiology Unit (MC_UU_00011/1, MC_UU_00011/3 and
- 481 MC_UU_00011/4). L.A.C.M is funded by a University of Bristol Vice-Chancellor's fellowship.
- 482

483 Code and data availability

- 484
- 485 Software to perform variance GWAS using the LAD Brown-Forsythe model is available from
- 486 https://github.com/MRCIEU/varGWAS and R-package for *ad hoc* analyses is available from
- 487 https://github.com/MRCIEU/varGWASR. Code for performing simulation studies is available
- 488 from https://github.com/MRCIEU/varGWAS/sim. Code for running the UK Biobank analysis
- 489 is available from https://github.com/MRCIEU/varGWAS-ukbb-biomarkers. Full variance
- 490 GWAS summary statistics are available from the OpenGWAS project^{39,68}. All code
- 491 repositories are available under the GPL v3 license.

492 References

- 493
- 494 1. Holmes, M. V., Richardson, T.G., Ference, B.A., Davies, N.M., and Davey Smith, G. (2021).
- 495 Integrating genomics with biomarkers and therapeutic targets to invigorate cardiovascular
- 496 drug development. Nat. Rev. Cardiol. 18, 435–453.
- 497 2. Chen, V.L., Du, X., Chen, Y., Kuppa, A., Handelman, S.K., Vohnoutka, R.B., Peyser, P.A.,
- 498 Palmer, N.D., Bielak, L.F., Halligan, B., et al. (2021). Genome-wide association study of serum
- 499 liver enzymes implicates diverse metabolic and liver pathology. Nat. Commun. 2021 121 12,
- 500 1**-**13.
- 3. Pekkanen, J., Linn, S., Heiss, G., Suchindran, C.M., Leon, A., Rifkind, B.M., and Tyroler, H.A.
- 502 (2010). Ten-Year Mortality from Cardiovascular Disease in Relation to Cholesterol Level
- among Men with and without Preexisting Cardiovascular Disease. NEJM *322*, 1700–1707.
- 4. Diabetes Prevention Program Research Group (2002). Reduction of the incidence of type
- 505 2 diabetes with lifestyle intervention or metformin. N. Engl. J. Med. *34*, 162–163.
- 506 5. Seth, R., Kydd, A.S., Buchbinder, R., Bombardier, C., and Edwards, C.J. (2014). Allopurinol
- 507 for chronic gout. Cochrane Database Syst. Rev. 2014.
- 508 6. Hunter, D.J. (2005). Gene–environment interactions in human diseases. Nat. Rev. Genet.
 509 2005 64 6, 287–298.
- 510 7. Brookes, S.T., Whitely, E., Egger, M., Davey Smith, G., Mulheran, P.A., and Peters, T.J.
- 511 (2004). Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power
- and sample size for the interaction test. J. Clin. Epidemiol. *57*, 229–236.
- 513 8. Smith, P.G., and Day, N.E. (1984). The Design of Case-Control Studies: The Influence of
- 514 Confounding and Interaction Effects. Int. J. Epidemiol. *13*, 356–365.
- 515 9. Deng, W.Q., and Paré, G. (2011). A fast algorithm to optimize SNP prioritization for gene-

- 516 gene and gene-environment interactions. Genet. Epidemiol. *35*, 729–738.
- 517 10. Paré, G., Cook, N.R., Ridker, P.M., and Chasman, D.I. (2010). On the Use of Variance per
- 518 Genotype as a Tool to Identify Quantitative Trait Interaction Effects: A Report from the
- 519 Women's Genome Health Study. PLoS Genet. 6, e1000981.
- 520 11. Rönnegård, L., and Valdar, W. (2012). Recent developments in statistical methods for
- 521 detecting genetic loci affecting phenotypic variability. BMC Genet. 2012 131 13, 1–7.
- 522 12. Yang, J., Loos, R.J.F., Powell, J.E., Medland, S.E., Speliotes, E.K., Chasman, D.I., Rose, L.M.,
- 523 Thorleifsson, G., Steinthorsdottir, V., Mägi, R., et al. (2012). FTO genotype is associated with
- 524 phenotypic variability of body mass index. Nature *490*, 267–272.
- 13. Wang, H., Zhang, F., Zeng, J., Wu, Y., Kemper, K.E., Xue, A., Zhang, M., Powell, J.E.,
- 526 Goddard, M.E., Wray, N.R., et al. (2019). Genotype-by-environment interactions inferred
- 527 from genetic effects on phenotypic variability in the UK Biobank. Sci. Adv. 5, eaaw3538.
- 528 14. Brown, A.A., Buil, A., Viñuela, A., Lappalainen, T., Zheng, H.F., Richards, J.B., Small, K.S.,
- 529 Spector, T.D., Dermitzakis, E.T., and Durbin, R. (2014). Genetic interactions affecting human
- 530 gene expression identified by variance association mapping. Elife 2014, e01381.
- 531 15. Staley, J.R., Windmeijer, F., Suderman, M., Lyon, M.S., Davey Smith, G., and Tilling, K.
- 532 (2021). A robust mean and variance test with application to high-dimensional phenotypes.
- 533 Eur. J. Epidemiol. 1, 1–11.
- 16. Revez, J.A., Lin, T., Qiao, Z., Xue, A., Holtz, Y., Zhu, Z., Zeng, J., Wang, H., Sidorenko, J.,
- 535 Kemper, K.E., et al. (2020). Genome-wide association study identifies 143 loci associated
- with 25 hydroxyvitamin D concentration. Nat. Commun. 2020 111 11, 1–12.
- 17. Liu, D., Ban, H.-J., El Sergani, A.M., Lee, M.K., Hecht, J.T., Wehby, G.L., Moreno, L.M.,
- 538 Feingold, E., Marazita, M.L., Cha, S., et al. (2021). PRICKLE1 × FOCAD Interaction Revealed by
- 539 Genome-Wide vQTL Analysis of Human Facial Traits. Front. Genet. 0, 1112.

540	18. Young, J	A.I., Wauthie	r, F.L., ar	nd Donnelly, I	P. (2018). Identifying	loci affecting	g trait
			.,,					

- 541 variability and detecting interactions in genome-wide association studies. Nat. Genet. 50,
- 542 1608-1614.
- 543 19. Dumitrascu, B., Darnell, G., Ayroles, J., and Engelhardt, B.E. (2019). Statistical tests for
- 544 detecting variance effects in quantitative trait studies. Bioinformatics 35, 200–210.
- 545 20. Corty, R.W., and Valdar, W. (2018). QTL mapping on a background of variance
- 546 heterogeneity. G3 Genes, Genomes, Genet. *8*, 3767–3782.
- 547 21. Levene, H. (1960). Robust testes for equality of variances. Contrib. to Probab. Stat. 278-
- 548 292.
- 549 22. Brown, M.B., and Forsythe, A.B. (1974). Robust tests for the equality of variances. J. Am.
- 550 Stat. Assoc. *69*, 364–367.
- 551 23. Soave, D., and Sun, L. (2017). A generalized Levene's scale test for variance
- heterogeneity in the presence of sample correlation and group uncertainty. Biometrics 73,

553 <u>960–</u>971.

- 554 24. Glejser, H. (1969). A New Test for Heteroskedasticity. J. Am. Stat. Assoc. 64, 316.
- 555 25. Pietrosanu, M., Gao, J., Kong, L., Jiang, B., and Niu, D. (2020). Advanced algorithms for
- penalized quantile and composite quantile regression. Comput. Stat. 2020 361 36, 333–346.
- 557 26. Zhang, F., Chen, W., Zhu, Z., Zhang, Q., Nabais, M.F., Qi, T., Deary, I.J., Wray, N.R.,
- 558 Visscher, P.M., McRae, A.F., et al. (2019). OSCA: a tool for omic-data-based complex trait
- 559 analysis. Genome Biol. 2019 201 20, 1–13.
- 560 27. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W.,
- 561 Goddard, M.E., Wray, N.R., Visscher, P.M., et al. (2016). Integration of summary data from
- 562 GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. 48, 481–487.
- 563 28. Oehlert, G.W. (1992). A Note on the Delta Method. Source Am. Stat. 46, 27–29.

- 564 29. White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a
- 565 Direct Test for Heteroskedasticity. Econometrica *48*, 817.
- 566 30. Hunter, D.R., and Lange, K. (2000). Quantile Regression via an MM Algorithm. J. Comput.
- 567 Graph. Stat. *9*, 60.
- 568 31. Guennebaud, G., Jacob, B., and others (2010). Eigen v3.
- 569 32. Band, G., and Marchini, J. (2018). BGEN: a binary file format for imputed genotype and
- 570 haplotype data. BioRxiv 308296.
- 571 33. Morris, T.P., White, I.R., and Crowther, M.J. (2019). Using simulation studies to evaluate
- 572 statistical methods. Stat. Med. *38*, 2074–2102.
- 573 34. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A.,
- 574 Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep
- 575 phenotyping and genomic data. Nature *562*, 203–209.
- 576 35. Howie, B., Marchini, J., and Stephens, M. (2011). Genotype imputation with thousands
- 577 of genomes. G3 Genes, Genomes, Genet. 1, 457–470.
- 578 36. Mitchell, R.E., Hemani, G., Dudding, T., Corbin, L., Harrison, S., and Paternoster, L. UK
- 579 Biobank Genetic Data: MRC-IEU Quality Control, version 2, 18/01/2019.
- 580 37. Fry, D., Almond, R., Moffat, S., Gordon, M., and Singh, P. (2019). UK Biobank Biomarker
- 581 Project Companion Document to Accompany Serum Biomarker Data.
- 582 38. Cassidy, S., Chau, J.Y., Catt, M., Bauman, A., and Trenell, M.I. (2016). 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.
- 585 BMJ Open *6*, e010038.
- 586 39. Elsworth, B., Lyon, M., Alexander, T., Liu, Y., Matthews, P., Hallett, J., Bates, P., Palmer,
- 587 T., Haberland, V., Davey Smith, G., et al. (2020). The MRC IEU OpenGWAS data

- 588 infrastructure. BioRxiv 2020.08.10.244293.
- 40. Auton, A., Abecasis, G.R., Altshuler, D.M., Durbin, R.M., Bentley, D.R., Chakravarti, A.,
- 590 Clark, A.G., Donnelly, P., Eichler, E.E., Flicek, P., et al. (2015). A global reference for human
- 591 genetic variation. Nat. 2015 5267571 526, 68–74.
- 41. Hemani, G., Powell, J.E., Wang, H., Shakhbazov, K., Westra, H.-J., Esko, T., Henders, A.K.,
- 593 McRae, A.F., Martin, N.G., Metspalu, A., et al. (2021). Phantom epistasis between unlinked
- 594 loci. Nat. 2021 5967871 *596*, E1–E3.
- 42. Wang, G., Sarkar, A., Carbonetto, P., and Stephens, M. (2020). A simple new approach to
- variable selection in regression, with application to genetic fine mapping. J. R. Stat. Soc. Ser.
- 597 B (Statistical Methodology). *82*, 1273–1300.
- 43. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing
- 599 genomic features. Bioinforma. Appl. NOTE 26, 841–842.
- 44. Tweedie, S., Braschi, B., Gray, K., Jones, T.E.M., Seal, R.L., Yates, B., and Bruford, E.A.
- 601 (2021). Genenames.org: The HGNC and VGNC resources in 2021. Nucleic Acids Res. 49,
- 602 D939–D946.
- 45. Võsa, U., Claringbould, A., Westra, H.-J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H.,
- Saha, A., Kreuzhuber, R., Yazar, S., et al. (2021). Large-scale cis- and trans-eQTL analyses
- 605 identify thousands of genetic loci and polygenic scores that regulate blood gene expression.
- 606 Nat. Genet. 2021 539 *53*, 1300–1310.
- 607 46. GTEx Portal, https://gtexportal.org.
- 47. Campos, G. de los, Sorensen, D.A., and Toro, M.A. (2019). Imperfect Linkage
- 609 Disequilibrium Generates Phantom Epistasis (& Perils of Big Data). G3 Genes, Genomes,
- 610 Genet. 9, 1429–1436.
- 48. Stender, S., Kozlitina, J., Nordestgaard, B.G., Tybjaerg-Hansen, A., Hobbs, H.H., and

- 612 Cohen, J.C. (2017). Adiposity Amplifies the Genetic Risk of Fatty Liver Disease Conferred by
- 613 Multiple Loci HHS Public Access Author manuscript. Nat Genet 49, 842–847.
- 49. Viitasalo, A., Pihlajamaki, J., Lindi, V., Atalay, M., Kaminska, D., Joro, R., and Lakka, T.A.
- 615 (2015). 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,
- 617 84–90.
- 50. Döring, A., Gieger, C., Mehta, D., Gohlke, H., Prokisch, H., Coassin, S., Fischer, G., Henke,
- 619 K., Klopp, N., Kronenberg, F., et al. (2008). SLC2A9 influences uric acid concentrations with
- 620 pronounced sex-specific effects. Nat. Genet. 2008 404 40, 430–436.
- 51. Ferrières, J., Sing, C.F., Roy, M., Davignon, J., and Lussier-Cacan, S. (1994).
- 622 Apolipoprotein E polymorphism and heterozygous familial hypercholesterolemia. Sex-
- 623 specific effects. Arterioscler. Thromb. a J. Vasc. Biol. 14, 1553–1560.
- 52. Ruth, K.S., Day, F.R., Tyrrell, J., Thompson, D.J., Wood, A.R., Mahajan, A., Beaumont,
- 625 R.N., Wittemans, L., Martin, S., Busch, A.S., et al. (2020). Using human genetics to
- 626 understand the disease impacts of testosterone in men and women. Nat. Med. 2020 262 26,
- 627 252-258.
- 53. Masuda, M., Okuda, K., Ikeda, D.D., Hishigaki, H., and Fujiwara, T. (2015). Interaction of
- 629 genetic markers associated with serum alkaline phosphatase levels in the Japanese
- 630 population. Hum. Genome Var. 2015 21 2, 1–6.
- 631 54. Langman, M.J.S., Leuthold, E., Robson, E.B., Harris, J., Luffman, J.E., and Harris, H. (1966).
- 632 Influence of diet on the "intestinal" component of serum alkaline phosphates in people of
- 633 different ABO blood groups and secretor status. Nature *212*, 41–43.
- 55. Abul-Husn, N.S., Cheng, X., Li, A.H., Xin, Y., Schurmann, C., Stevis, P., Liu, Y., Kozlitina, J.,
- 635 Stender, S., Wood, G.C., et al. (2018). A Protein-Truncating HSD17B13 Variant and

- 636 Protection from Chronic Liver Disease. NEJM *378*, 1096–1106.
- 637 56. Gellert-Kristensen, H., Richardson, T.G., Davey Smith, G., Nordestgaard, B.G., Tybjærg-
- Hansen, A., and Stender, S. (2020). Combined Effect of PNPLA3, TM6SF2, and HSD17B13
- 639 Variants on Risk of Cirrhosis and Hepatocellular Carcinoma in the General Population.
- 640 Hepatology 72, 845–856.
- 57. Bayer, P.M., Hotschek, H., and Knoth, E. (1980). Intestinal alkaline phosphatase and the
- ABO blood group system--a new aspect. Clin. Chim. Acta. 108, 81-87.
- 58. Nakano, T., Shimanuki, T., Matsushita, M., Koyama, I., Inoue, I., Katayama, S., Alpers,
- D.H., and Komoda, T. (2006). Involvement of intestinal alkaline phosphatase in serum
- apolipoprotein B-48 level and its association with ABO and secretor blood group types.
- 646 Biochem. Biophys. Res. Commun. 341, 33–38.
- 59. Neale, B., et al. UK Biobank GWAS, http://www.nealelab.is/uk-biobank
- 648 60. Kanai, M., Akiyama, M., Takahashi, A., Matoba, N., Momozawa, Y., Ikeda, M., Iwata, N.,
- 649 Ikegawa, S., Hirata, M., Matsuda, K., et al. (2018). Genetic analysis of quantitative traits in
- 650 the Japanese population links cell types to complex human diseases. Nat. Genet. 50, 390-
- 651 400.
- 652 61. Young, K.A., Palmer, N.D., Fingerlin, T.E., Langefeld, C.D., Norris, J.M., Wang, N., Xiang,
- A.H., Guo, X., Williams, A.H., Chen, Y.D.I., et al. (2019). Genome-Wide Association Study
- 654 Identifies Loci for Liver Enzyme Concentrations in Mexican-Americans: The GUARDIAN
- 655 Consortium. Obesity (Silver Spring). 27, 1331.
- 656 62. Chambers, J.C., Zhang, W., Sehmi, J.S., Li, X., Wass, M.N., Van der Harst, P., Holm, H.,
- 657 Sanna, S., Kavousi, M., Baumeister, S.E., et al. (2011). Genome-wide association study
- 658 identifies loci influencing concentrations of liver enzymes in plasma. Nat. Genet. 2011 4311
- 659 *43*, 1131–1138.

- 660 63. Xu, Z.M., and Burgess, S. (2020). Polygenic modelling of treatment effect heterogeneity.
- 661 Genet. Epidemiol. 44, 868–879.
- 662 64. Zheng, J., Haberland, V., Baird, D., Walker, V., Haycock, P.C., Hurle, M.R., Gutteridge, A.,
- 663 Erola, P., Liu, Y., Luo, S., et al. (2020). Phenome-wide Mendelian randomization mapping the
- 664 influence of the plasma proteome on complex diseases. Nat. Genet. 2020 5210 52, 1122-
- 665 1131.
- 666 65. Cao, Y., Wei, P., Bailey, M., Kauwe, J.S.K., and Maxwell, T.J. (2014). A versatile omnibus
- test for detecting mean and variance heterogeneity. Genet. Epidemiol. 38, 51–59.
- 668 66. Ek, W.E., Rask-Andersen, M., Karlsson, T., Enroth, S., Gyllensten, U., and Johansson, Å.
- 669 (2018). Genetic variants influencing phenotypic variance heterogeneity. Hum. Mol. Genet.
- 670 *27*, 799–810.
- 671 67. Rees, J.M.B., Foley, C.N., and Burgess, S. (2020). Factorial Mendelian randomization:
- using genetic variants to assess interactions. Int. J. Epidemiol. 49, 1147–1158.
- 673 68. Lyon, M.S., Andrews, S.J., Elsworth, B., Gaunt, T.R., Hemani, G., and Marcora, E. (2021).
- 674 The variant call format provides efficient and robust storage of GWAS summary statistics.
- 675 Genome Biol. 22, 32.

677 Figure 1. Variance effect estimate accuracy and confidence interval coverage

- 678 Variance effect estimate accuracy (A, B) and 95% confidence interval coverage (C, D) of
- 679 simulated genotypes with linear effect on outcome variance (A, C) or interaction effect (B,
- 680 D). LAD-BF, least-absolute deviation regression Brown-Forsythe. OSCA-BF, original Brown-
- 681 Forsythe test implemented in OSCA²⁶ including effect estimate derived from the test P-
- 682 value²⁷. CI, confidence interval.

Figure 2. Effect of top gene-environment interaction loci on trait mean and variance

- 684 Per-allele effect of SNP stratified by modifier on outcome mean estimated with
- 685 heteroscedastic-consistent standard errors²⁹ and unstratified effect of SNP on variance
- estimated using LAD-BF (genotype 0 vs 1 and 0 vs 2) with or without adjustment for the
- 687 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
- 689 components. SD, standard deviation. CI, confidence interval. ALT, alanine aminotransferase.
- 690 LDL, low-density lipoprotein. BMI, body mass index. Low BMI, <= 26.7 kg/m². High BMI, >
- 691 26.7 kg/m².

Figure 3. Effect of top gene-gene interaction loci on trait mean and variance

- 693 Per allele effect of SNP stratified by modifier on outcome mean estimated with
- 694 heteroscedastic-consistent standard errors²⁹ and unstratified effect of SNP on variance
- 695 estimated using LAD-BF (genotype 0 vs 1 and 0 vs 2) with or without adjustment for the
- 696 interaction term. All estimates were adjusted for age, sex, and top ten genetic principal
- 697 components. SD, standard deviation. CI, confidence interval. ALP, alkaline phosphatase. ALT,
- 698 alanine aminotransferase. AST, aspartate aminotransferase. GGT, gamma
- 699 glutamyltransferase.

		1
Figure	Aim	Conditions
1	Determine variance effect estimate accuracy and confidence interval coverage	Increasing interaction effect size.
S2	Estimate power of variance tests to detect interaction effects using normal/non-normal outcomes	Increasing interaction effect size. Normal, log-normal, and t-distribution residuals
S3	Estimate type I error of variance tests with normal/non-normal outcomes	Increasing SNP minor allele frequency. Normal, log-normal, t-distribution, and mixed- normal residuals
S4	Adjusting for interaction effect on variance effect estimate	With/without adjustment for modifier and interaction term in first-stage model
S5	Estimate model runtime performance	Increasing number of CPU threads

700 Table 1. Simulation studies of the Brown-Forsythe test

SNP, single-nucleotide polymorphism. Log, natural logarithm. CPU, central processing unit.





