

1 **Genomic summary statistics and meta-analysis for set-based gene-** 2 **environment interaction tests in large-scale sequencing studies**

3 Xinyu Wang, Duy T. Pham, Kenneth E. Westerman, Cong Pan, Alisa K. Manning & Han Chen*

4 **Abstract**

5 We propose an efficient method to generate the summary statistics for set-based gene-environment
6 interaction tests, as well as a meta-analysis approach that aggregates the summary statistics across
7 different studies, which can be applied to large biobank-scale sequencing studies with related samples.
8 Simulations showed that meta-analysis is numerically concordant with the equivalent pooled analysis
9 using individual-level data. Moreover, meta-analysis accommodates heterogeneity between studies and
10 enhances power in multi-ethnic studies. We applied the meta-analysis approach to the whole-exome
11 sequencing data from the UK Biobank and successfully identified gene regions associated with waist-hip
12 ratio, as well as those with sex-specific genetic effects.

13

14 **Keywords**

15 Meta-analysis, genomic summary statistics, gene-environment interaction, joint test, rare variants, set-
16 based test, UK Biobank

17

18

19

20

21

22

23

24

25

26

27 **Background**

28 Complex diseases are influenced by the synergy of genes and the environment. The study of gene-
29 environment interactions (GEI) may shed light on disease etiology and help identify environmental risk
30 factors that modify the effects of disease-susceptible genes, as well as genetic variants that modify the
31 effects of environmental risk factors for complex diseases (1,2). Following rapid advances in next-
32 generation sequencing (NGS) technologies over the past years, a growing body of research focuses on
33 investigating GEI effects in rare variants (with minor allele frequency (MAF) < 5%), which provides
34 insights into additional disease risk and trait variability undiscovered in common variants from genome-
35 wide association studies (GWAS) (3,4). As in genetic main effect tests, single-variant tests are
36 underpowered in rare variant GEI tests (1,2). A wide variety of set-based GEI tests were developed to
37 improve the statistical power, which are commonly extensions of the burden test and sequence kernel
38 association test (SKAT) (5-7), or a combination of them (8-10). Additionally, the joint tests of both
39 genetic main effects and interaction effects for rare variants were developed to identify variants associated
40 with complex traits, accounting for heterogeneous genetic effects in different environmental exposures
41 (8,10). It is especially useful to perform joint tests when the main effects are small and GEI effects are
42 difficult to detect through interaction only tests (11,12).

43 Detecting rare variant GEI effects requires extremely large sample sizes given the low power as only a
44 few individuals carried a rare variant while exposed to the environment factor (2,12). Meta-analysis is a
45 well-established way to achieve the scale of sample size needed by rare variant GEI tests. By aggregating
46 summary statistics from different studies, meta-analysis not only increases the sample size, but also
47 avoids sharing individual-level data, which is usually restricted by study policies and confidentiality laws.
48 On the other hand, the file sizes of individual-level genotype data are often extremely large, making it
49 difficult to transfer across platforms with limited disk space and computational resources. The meta-
50 analysis approach can circumvent this problem easily as it does not require sharing individual-level data
51 but rather the genomic summary statistics which are more resource-efficient.

52 Meta-analysis tools have been developed for genetic main effect tests (13-21). In GWAS, the widely-used
53 inverse variance weighted meta-analysis for common variants usually combines the single-variant effect
54 estimates and their standard errors (21). Meta-analysis for rare variants cannot use the same approach, as
55 models estimating multiple genetic effect sizes with large variances for sparse data can fail to converge,
56 making the algorithm numerically unstable. For this reason, meta-analysis tools for rare variants usually
57 combine score statistics and covariance matrices of individual variants to recover the set-based tests,
58 which are better suited for low-frequency and rare variants. Among these tools, MetaSKAT (13), MASS
59 (17), and Meta-Qtest (19) were developed for unrelated samples. For related samples, RAREMETAL (15-
60 17) can retrieve either the burden test or SKAT, whereas SMMAT (20) can recover burden, SKAT, and a
61 unified test that combines both.

62 For set-based GEI tests and joint tests, however, very few meta-analysis methods have been developed.
63 Among them, ofGEM (22) introduces filtering statistics based on meta-analysis, but it is only applicable
64 to unrelated samples and no joint tests for genetic main effects and GEI effects were proposed. A recent
65 study proposed extending the rareGE framework (8) to meta-analysis, which can also be conducted for
66 unrelated samples only (23). It is not computationally efficient to generate summary statistics using this
67 method, as for every variant set, a separate statistical model accounting for genetic main effects needs to
68 be fitted. Moreover, methods have been proposed for meta-analysis of joint tests for genetic main effects
69 and GEI effects (24, 25), but they are only applicable to single variant tests. It remains a technical gap in
70 the field to efficiently generate and utilize genomic summary statistics for set-based GEI and joint tests.

71 Recently, we developed a computationally efficient method, Mixed-model Association test for Gene-
72 Environment interactions (MAGEE), for rare variant GEI and joint tests (10). The goal of this study is to
73 develop a general framework for genomic summary statistics and meta-analysis for set-based GEI and
74 joint tests, which are applicable to both unrelated and related samples as well as both quantitative and
75 binary traits. Since MAGEE does not require fitting a separate model adjusting for the genetic main
76 effects in each testing region, the genomic summary statistics for set-based GEI and joint tests can be
77 generated efficiently even for large samples in a whole-genome analysis.

78 **Results**

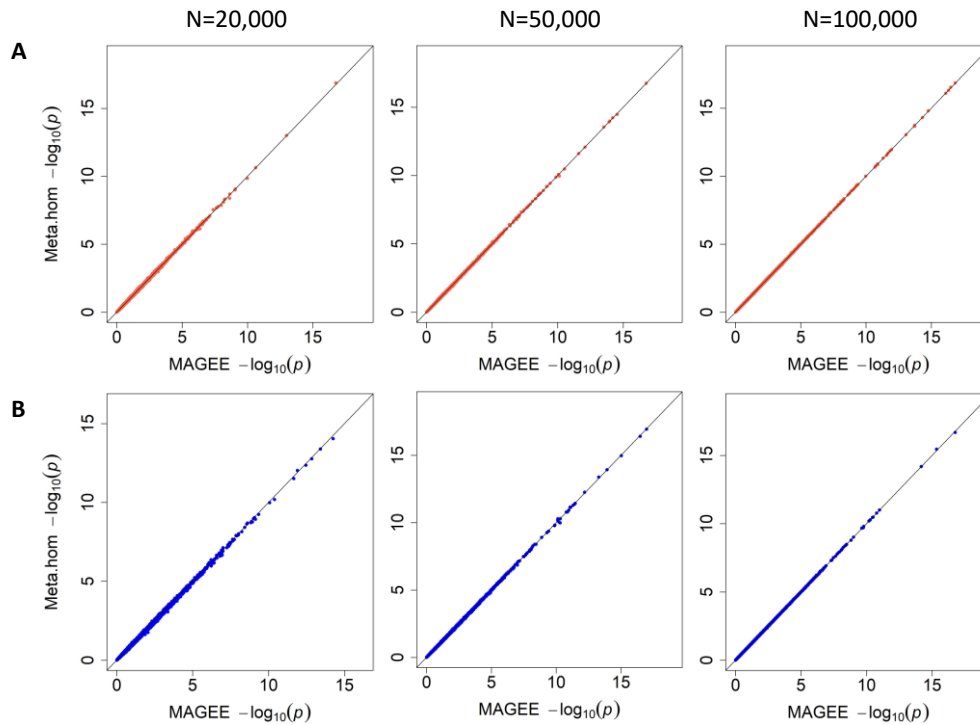
79 ***P* value benchmark**

80 The simulated individual-level genotype data included 100,000 samples and 100,000 genetic variants in
81 1,000 groups (100 variants per group). In the meta-analysis, a great amount of time can be saved when
82 the sample size is large as summary statistic-based calculations do not depend on the sample size. For
83 example, we directly computed the *p* values from individual-level data of 100,000 samples, which took
84 1,030 s on a single thread for each simulation replicate with 100,000 variants. In contrast, when using the
85 summary statistics, computing the *p* values of the same tests took only 9.7 s using a single thread on the
86 same computing server, which saved over 99% CPU time.

87 For both quantitative and binary traits with different sample sizes, we compared the *p* values from the
88 meta-analysis (assuming homogeneous genetic effects across studies) with those from pooled individual-
89 level data analysis. Each panel in Figure 1 and Figure 2 displays 10,000 *p* values from quantitative trait
90 analyses with related samples. In the MAGEE framework for individual-level data, we developed two
91 GEI tests, interaction variance component (IV) test and interaction hybrid test using Fisher's method (IF),
92 along with three joint tests, joint variance component (JV) test, joint hybrid test using Fisher's method
93 (JF), and joint hybrid test using double Fisher's procedures (JD) (10). We conducted extensive
94 simulations in MAGEE (10), which showed that each test was well-calibrated for type I error rates. As
95 illustrating examples, here we compared results from IF test as the GEI test, and JD test as the joint test,
96 for meta-analysis and MAGEE pooled analysis. The meta-analysis assuming homogeneous genetic effects
97 shared highly consistent empirical GEI and joint test *p* values with MAGEE pooled analysis (Figures 1
98 and 2), regardless of the simulation scenario of homogeneous or heterogeneous genetic effects, and the
99 precision of estimates increased with the sample size. Most importantly, for small *p* values around the
100 commonly used significance thresholds (5.0×10^{-8} to 2.5×10^{-6}), meta-analysis did not lose any accuracy
101 compared with the MAGEE pooled analysis while being more resource-efficient in saving the
102 computational time. We also found similar results for binary traits (Supplemental Figure S1 and S2).

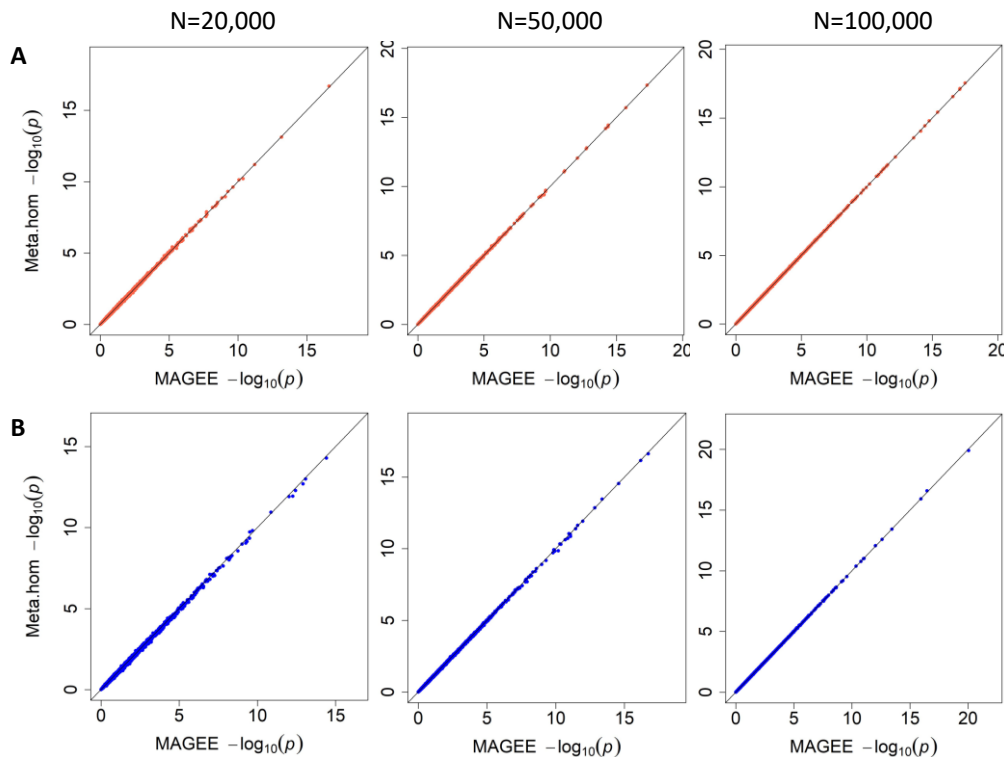
103

104 Figure 1. GEI test p value benchmark for meta-analysis assuming homogeneous genetic effects with
105 MAGEE in 20,000, 50,000, and 100,000 unrelated samples. (A) Scenario 1 (homogeneous scenario) in
106 quantitative traits. (B) Scenario 2 (heterogeneous scenario) in quantitative traits.



107

108 Figure 2. Joint test p value benchmark for meta-analysis assuming homogeneous genetic effects with
109 MAGEE in 20,000, 50,000, and 100,000 unrelated samples. (A) Scenario 1 (homogeneous scenario) in
110 quantitative traits. (B) Scenario 2 (heterogeneous scenario) in quantitative traits.



111

112 **Power comparison**

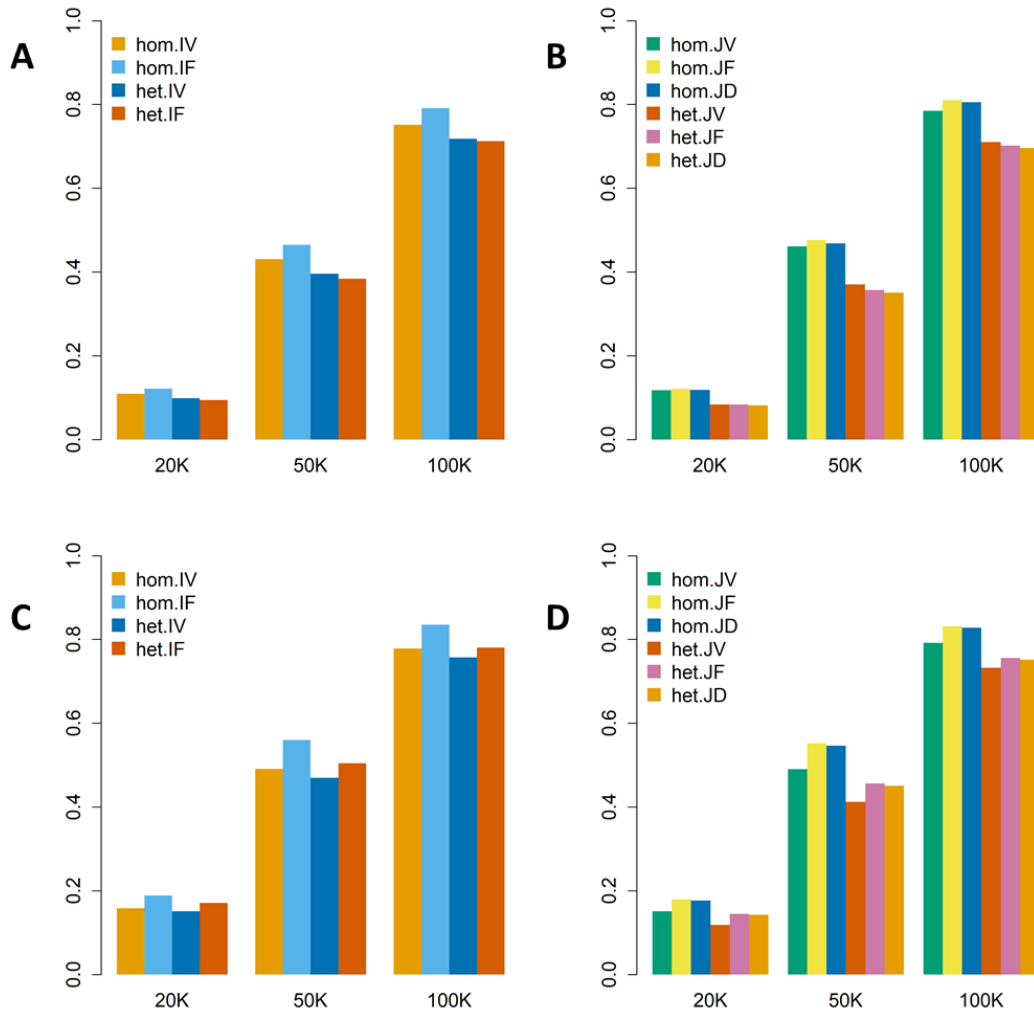
113 Figure 3 shows the empirical power of the meta-analysis GEI and joint tests for analyzing quantitative
114 traits from scenario 1, where the two studies had homogeneous covariate effects, genetic effects, and GEI
115 effects. The power was calculated at the significance level of 2.5×10^{-6} with 20,000, 50,000, and 100,000
116 related samples, respectively. The top panel displays results when the ratio of positive and negative causal
117 variants was 1:1, while the bottom panel displays results when the ratio of positive and negative causal
118 variants was a 4:1. Overall, the meta-analysis tests assuming homogeneous effects are more powerful for
119 both GEI and joint tests, with the IF test being more powerful than the IV test and the JF test being
120 slightly more powerful than the other 2 joint tests.

121 Figure 4 illustrated the empirical power of the meta-analysis GEI and joint tests for analyzing quantitative
122 traits from scenario 2, where the two studies had heterogeneous covariates effects, genetic effects, and
123 GEI effects. Same as in Figure 3, the ratio of positive to negative causal variants in Figure 4 was 1:1 on
124 the top panel and 4:1 on the bottom panel. In this scenario, as heterogeneity was introduced to these
125 studies, the overall performance of the meta-analysis tests assuming heterogeneous effects outperformed
126 the meta-analysis assuming homogeneous effects for both GEI and joint tests.

127 The same simulations were conducted for binary traits, and we summarized those results in Supplemental
128 Figures S3 and S4.

129

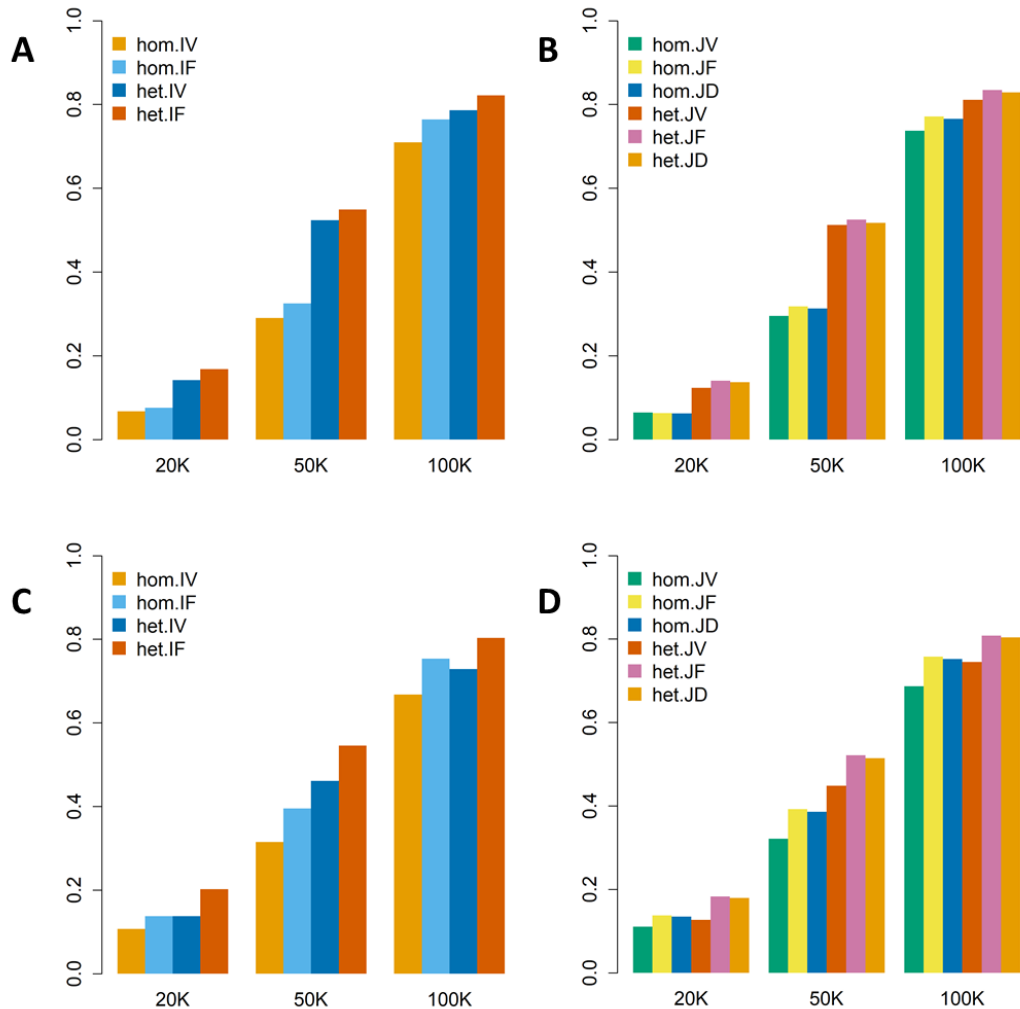
130 Figure 3. Empirical power of meta-analysis tests for scenario 1 (homogeneous scenario) on quantitative
131 traits in 20,000, 50,000, and 100,000 related samples. (A) GEI tests with 80% null variants, 10% causal
132 variants with positive effects, and 10% causal variants with negative effects. (B) Joint tests with 80% null
133 variants, 10% causal variants with positive effects, and 10% causal variants with negative effects. (C) GEI
134 tests with 80% null variants, 16% causal variants with positive effects, and 4% causal variants with
135 negative effects. (D) Joint tests with 80% null variants, 16% causal variants with positive effects, and 4%
136 causal variants with negative effects.



137

138

139 Figure 4. Empirical power of meta-analysis tests for scenario 2 (heterogeneous scenario) on quantitative
 140 traits in 20,000, 50,000, and 100,000 related samples. (A) GEI tests with 80% null variants, 10% causal
 141 variants with positive effects, and 10% causal variants with negative effects. (B) Joint tests with 80% null
 142 variants, 10% causal variants with positive effects, and 10% causal variants with negative effects. (C) GEI
 143 tests with 80% null variants, 16% causal variants with positive effects, and 4% causal variants with
 144 negative effects. (D) Joint tests with 80% null variants, 16% causal variants with positive effects, and 4%
 145 causal variants with negative effects.



146

147

148 **Meta-analysis of waist-hip ratio in the UK Biobank**

149 The original individual-level genotype data from whole-exome sequencing (WES) provided by the UK
 150 Biobank for 200,632 samples was 830 GB for 22 autosomes in the PLINK BED format, while the gene-
 151 based summary statistics were about 11 GB in total, including 5.4 GB intermediate files from the first
 152 batch and 5.4 GB intermediate files from the second batch. By sharing only the summary statistics files

153 for meta-analysis, we saved about 98.7% disk space compared to directly sharing individual-level data for
154 set-based GEI tests. In addition, when we directly computed p values from individual-level data, it took
155 25.48 h (91,732 s) CPU time with 10 threads to conduct all the GEI tests and joint tests in the MAGEE
156 framework. In contrast, we spent only 23.8 min (1,428 s) CPU time using a single thread on the same
157 computing server to compute the p values of the same GEI and joint tests when we had the summary
158 statistics, saving about 98.4% of computational time. One additional advantage for meta-analyzing the
159 UK Biobank WES data is that when new samples are released in future tranches, we do not have to spend
160 extra computational cost in rerunning the pooled analysis including previously released samples. Instead,
161 by combining the summary statistics from different releases in a meta-analysis, we can get similar results
162 as the pooled analysis while saving the computational resources.

163 The two batches of WES data from the UK Biobank were meta-analyzed in 18,668 protein-coding regions
164 to investigate gene-sex interactions effects on waist-hip ratio (WHR). The significance level was adjusted
165 by Bonferroni correction for multiple testing (26) at $0.05/18,668 = 2.7 \times 10^{-6}$. As shown in Figure 5, the
166 GEI test found three significant regions: *COBLL1* (p value = 5.8×10^{-10}) on chromosome 2, *HMGA1* (p
167 value = 2.8×10^{-7}) on chromosome 6, and *VEGFB* (p value = 1.3×10^{-6}) on chromosome 11. However, none
168 of them remained significant after conditioning on significant variants in the region, defined as those with
169 a single variant joint test p value $< 5.0 \times 10^{-8}$, MAF $> 1\%$ and genotype missing rate $< 5\%$ (Table 1). A
170 total of 13 significant protein-coding regions were found from the joint tests, including *TBX15*, *ACVR1C*,
171 *COBLL1*, *NISCH*, *PLXND1*, *CYTL1*, *HLA-B*, *HMGA1*, *KIAA0408*, *MLXIPL*, *VEGFB*, *SBNO1*, and
172 *PAMI6*, after excluding less significant protein-coding regions in 1 million base pair flanking regions. Of
173 note, *COBLL1* (p value = 1.3×10^{-8} after adjusting for 7 significant single variants) and *KIAA0408* (p
174 value = 9.3×10^{-9} after adjusting for 1 significant single variant) had significant joint test p values in the
175 conditional analysis (Table 1).

176 Previous GWAS have identified many loci in these genes for WHR, some of which have sex-dimorphic
177 effects. For example, there is some evidence that certain loci at *COBLL1*, *PLXND1*, *HMGA1*, and *VEGFB*
178 have more impact on BMI-adjusted WHR among women (27, 28). In our meta-analysis GEI tests,

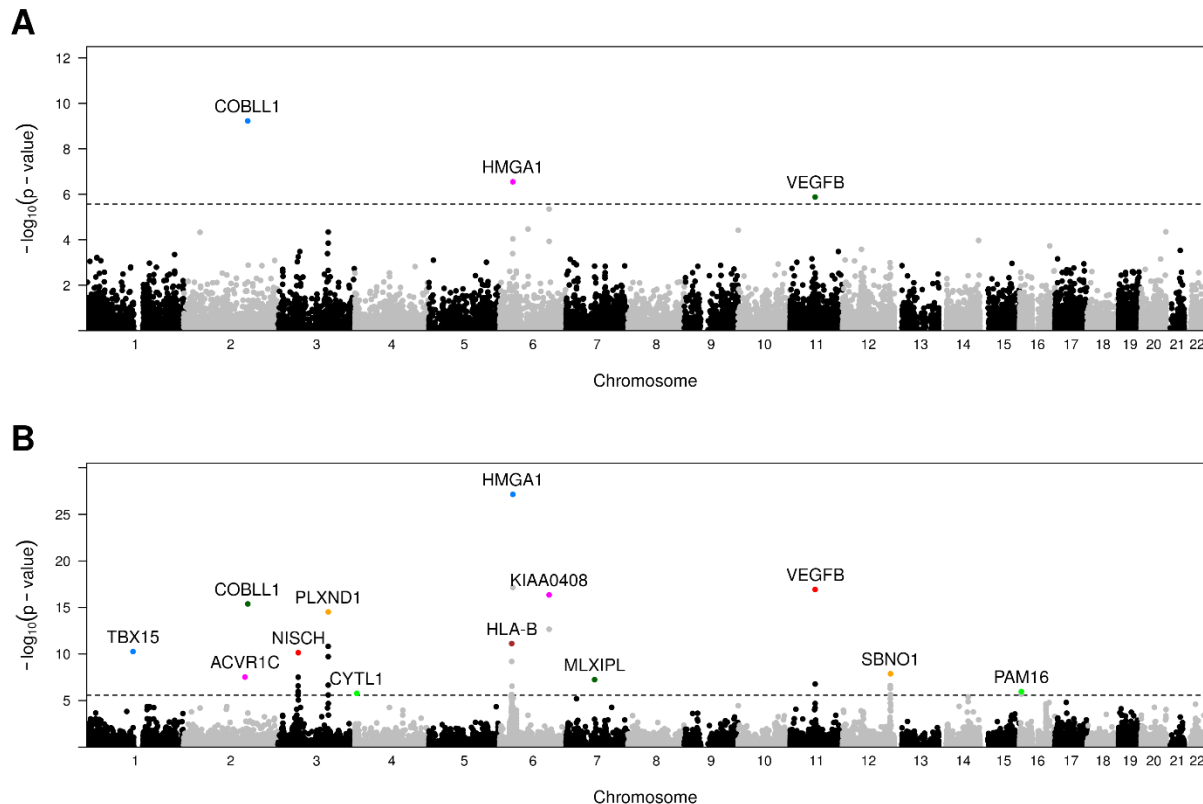
179 *PLXND1* did not show significance, however, we did identify gene-sex interactions in the other three
180 regions. Additionally, Justice *et al.* (29) reported gene-sex interactions in *COBLL1*, *PLXND1*, and
181 *KIAA0408*. There were no significant gene-sex interaction signals in *PLXND1* (p value = 4.6×10^{-5}) or
182 *KIAA0408* (p value = 4.5×10^{-6}), possibly because of the limited sample size in our analysis, and more
183 weights were added to rare variants using the beta function, which had potentially diminished the signals
184 from common variants.

185 We observed most signals driving our joint analyses were contributed by genetic main effects, as only
186 *COBLL1*, *HMGAI* and *VEGFB* were significant in the GEI tests (they were significant in the main effect
187 tests as well). Several loci in the genes identified in the joint tests were reported to be associated with
188 WHR in previous GWAS analysis (27-32) except for *HLA-B*. *HLA-B* is a member of the human leukocyte
189 antigen (*HLA*) complex gene family, which helps the immune system distinguish between proteins
190 created by the body and proteins made by outside invaders like viruses and bacteria (33). Further research
191 is needed to investigate if there are any sex-specific genetic impacts on WHR in *HLA-B*. The underlying
192 mechanisms for most of these genetic and sex differences have yet to be discovered.

193 We pooled individual-level data and ran MAGEE GEI and joint tests to validate the findings.
194 Supplemental Figure S5 shows that p values from the meta-analysis and the MAGEE pooled analysis
195 were highly concordant.

196

197 Figure 5. Manhattan plots of UK Biobank WES data analysis using meta-analysis for gene-sex interaction
 198 effects on WHR. (A) GEI test. (B) Joint test.



199
 200
 201 Table 1. Meta-analysis results for gene-sex interaction effects on WHR from the UK Biobank WES data
 202 analysis.

Gene	Chr	Start	End	N variants	N sig variants*	Unconditional GEI	Joint	Conditional GEI	Joint
<i>TBX15</i>	1	118883046	118989556	82	1	0.79	5.5×10^{-11}	0.67	0.011
<i>ACVR1C</i>	2	157526767	157628864	104	0	0.0089	3.1×10^{-8}	0.0089	3.1×10^{-8}
<i>COBLL1</i>	2	164653624	164843679	221	7	5.8×10^{-10}	4.2×10^{-16}	3.8×10^{-5}	7.4×10^{-8}
<i>NISCH</i>	3	52455118	52493068	348	11	5.6×10^{-4}	7.3×10^{-11}	5.0×10^{-4}	3.6×10^{-4}
<i>PLXND1</i>	3	129555175	129606818	423	14	4.6×10^{-5}	3.0×10^{-15}	0.037	8.3×10^{-4}
<i>CYTL1</i>	4	5014586	5019458	33	0	0.085	1.6×10^{-6}	0.085	1.6×10^{-6}
<i>HLA-B</i>	6	31269491	31357188	639	3	0.41	7.5×10^{-12}	0.45	2.0×10^{-5}
<i>HMGA1</i>	6	34236873	34246231	67	3	2.8×10^{-7}	7.3×10^{-28}	0.42	0.10
<i>KIAA0408</i>	6	127438406	127459389	76	1	4.5×10^{-6}	4.4×10^{-17}	0.0036	9.3×10^{-9}
<i>MLXIPL</i>	7	73593194	73624543	202	4	0.0014	5.6×10^{-8}	0.016	0.0022
<i>VEGFB</i>	11	64234538	64238793	81	2	1.3×10^{-6}	1.2×10^{-17}	0.22	0.16
<i>SBNO1</i>	12	123289109	123364847	344	0	0.021	1.3×10^{-8}	0.021	1.3×10^{-8}
<i>PAM16</i>	16	4331549	4355607	234	0	0.11	1.1×10^{-6}	0.11	1.1×10^{-6}

203 * Number of significant variants in the region, defined as those with a single variant joint test p value <
 204 5.0×10^{-8} , MAF > 1% and genotype missing rate < 5%. Conditional analysis p values were computed by
 205 adjusting for both the genetic main effect and gene-sex interaction terms of significant single variants.

207 **Discussion**

208 We present a framework for meta-analysis of rare variant GEI and joint tests for continuous and binary
209 traits. The meta-analysis is based on a generalized linear mixed effect model (GLMM), which can
210 account for relatedness in the samples. Single-variant scores and covariance matrices of individual studies
211 can be efficiently generated from MAGEE. The summary statistics for the meta-analysis can be easily
212 aggregated, which allows combining the evidence from millions of samples collected from multiple large-
213 scale sequencing studies to improve statistical power. Additionally, the proposed meta-analysis
214 framework provides flexible set-based joint tests for genetic main effects and GEI effects, including a
215 SKAT-type variance component test, and two unified tests combining burden and SKAT-type tests. In
216 population-based studies, heterogeneity is possible because different ethnic groups might have different
217 MAFs or effect sizes for the same variants, as well as different levels of environmental exposure.
218 Different ways of collecting and measuring data, genotyping methods, quality-control criterion, and
219 imputation methods make studies heterogeneous (34). The meta-analysis can easily account for the
220 heterogeneity across studies, which is difficult in pooled analyses using individual-level data.
221 We evaluated the performance of the proposed GEI and joint tests through computer simulations. We first
222 showed that meta-analysis assuming homogeneous genetic effects is numerically concordant with the
223 equivalent pooled analysis using individual-level data. Additionally, the proposed methodology can
224 account for heterogeneity across studies in terms of both covariate and genetic effects. GEI and joint tests
225 assuming heterogeneous genetic effects were generally more powerful than homogeneous-type meta-
226 analyses in the presence of simulated heterogeneity. Overall, IF tests were more powerful than IV tests,
227 and JF tests were the most efficient of the three joint tests, in accordance with what we determined from
228 the single-study numerical simulation of MAGEE.
229 We assessed gene-sex interaction effects impacting WHR, using GEI and joint tests within the proposed
230 meta-analysis framework on WES data from the UK Biobank. GEI tests identified *COBLL1*, *HMGAI* and
231 *VEGFB* that are already known to have loci with sex-dimorphic effects on WHR. The meta-analysis joint
232 tests identified 13 protein-coding regions, of which 10 had been previously linked to WHR in GWAS (27-

233 32), including *COBLL1*, *HMGAI*, *VEGFB*, *PLXND1* and *KIAA0408* that were reported to have sex-
234 specific genetic loci (27-29). The proposed meta-analysis GEI test did not reach significance for *PLXND1*
235 and *KIAA0408*, possibly due to the limited sample size. To identify rare variants that have effects
236 modified by the environment, extremely large sample sizes are required. Furthermore, we only examined
237 genetic variants in the exome, but noncoding regions may play a substantial role and may have integrated
238 effects with protein-coding genes in complex diseases and traits (27, 35). Once more WES data from the
239 UK Biobank and more whole-genome sequencing (WGS) data become available from other large-scale
240 consortia in the future, we expect that the meta-analysis will be used to identify more novel genetic
241 pathways contributing to a wide range of complex diseases.

242 Our summary statistics and meta-analysis framework have a few limitations. All participating studies in
243 the proposed meta-analysis must use a unified group definition file to ensure that the summary statistics
244 are generated consistently across studies. Although all component variants need not be present for all
245 studies, if different studies use different gene-level tests or group definition files, the summary statistics
246 would be incompatible (3). In addition, in certain case-control studies, case/control ratios could be
247 extremely unbalanced, leading to strong heterogeneity across studies (13). Future research could explore
248 these extremely heterogeneous cases in further depth.

249 **Conclusions**

250 We developed efficient and powerful meta-analysis approaches to overcome the challenges associated
251 with set-based GEI and joint tests for rare variants while allowing for sample relatedness. The meta-
252 analysis leverages study-specific genomic summary statistics, eliminating the need to share individual-
253 level data, and is applicable to both quantitative and binary traits. In addition, the meta-analysis can
254 accommodate heterogeneity among studies. These features allow the analysis of GEI effects across
255 multiple large-scale sequencing studies and can improve the power in identification of novel risk factors
256 for complex traits.

257

258 **Methods**

259 **Modeling summary statistics**

260 The meta-analysis of rare-variant GEI and joint tests involves two major steps. In the first step, we
261 perform a single-study analysis to calculate the summary statistics of the single-variant score vector and
262 covariance matrix for each variant. The second step consists of combining the summary statistics and
263 constructing the test statistic for the meta-analysis. The first step can be performed by using the MAGEE
264 software package. Below is a brief illustration of how to generate the summary statistics.

265 Suppose we have $m = 1, 2, \dots, M$ studies, first we need to generate the summary statistics for each
266 individual study. The full generalized linear mixed model (GLMM) for a single study m is in the
267 following form:

$$268 \quad g(\mu_{mi}) = \mathbf{X}_{mi}\boldsymbol{\alpha}_m + \mathbf{G}_{mi}\boldsymbol{\beta}_m + \mathbf{K}_{mi}\boldsymbol{\gamma}_m + r_{mi} \quad (1)$$

269 where $g(\cdot)$ is the link function of μ_{mi} , which is the conditional mean of the phenotype of individual i in
270 study m . Typically, for a quantitative trait, $g(\cdot)$ is an identity function, while for a binary trait, $g(\cdot)$ is a
271 logit function. \mathbf{X}_{mi} is a vector of p covariates including the intercept, and \mathbf{G}_{mi} is a vector of q variants,
272 and \mathbf{K}_{mi} is a vector of dq pairwise GEI terms for d environmental factors and q variants. Note that the
273 environmental factors are a subset of the covariates in \mathbf{X}_{mi} . Accordingly, $\boldsymbol{\alpha}_m$ is a $p \times 1$ vector for the
274 covariate effects, $\boldsymbol{\beta}_m$ is a $q \times 1$ vector for the genetic main effects, and $\boldsymbol{\gamma}_m$ is the $dq \times 1$ vector for GEI
275 effects. The vector for the random intercept $\mathbf{r}_m = (r_1 \ r_2 \ \dots \ r_N)^T \sim N(0, \sum_{l=1}^L \lambda_l \boldsymbol{\psi}_l)$, where N is the sample
276 size for study m , λ_l are the variance component parameters of L random effects, and $\boldsymbol{\psi}_l$ are $N \times N$
277 known relatedness matrices.

278 Through MAGEE, we made statistical inference that the GEI effects test statistic in model (1) can be
279 approximated by fitting a global null model without adjusting for the genetic main effects. Specifically,
280 the global null models are fitted to each individual study as follows:

$$281 \quad g(\mu_{mi}) = \mathbf{X}_{mi}\boldsymbol{\alpha}_m + r_{mi} \quad (2)$$

282 The score vector \mathbf{S}_{G_m} for genetic main effects and GEI effects \mathbf{S}_{K_m} are in the forms of $\mathbf{S}_{G_m} =$
 283 $\mathbf{G}_m^T(\mathbf{y}_m - \hat{\boldsymbol{\mu}}_{0m})/\hat{\phi}_m$ and GEI effects $\mathbf{S}_{K_m} = \mathbf{K}_m^T(\mathbf{y}_m - \hat{\boldsymbol{\mu}}_{0m})/\hat{\phi}_m$, where $\hat{\boldsymbol{\mu}}_{0m}$ is a vector of fitted values
 284 and $\hat{\phi}_m$ is the dispersion parameter estimated from model (2) in study m . Assuming the main effect of
 285 genetic variants $\boldsymbol{\beta}_m$ are small in a null model with genetic main effects, but without GEI effects
 286 $g(\mu_{mi}) = \mathbf{X}_{mi}\boldsymbol{\alpha}_m + \mathbf{G}_{mi}\boldsymbol{\beta}_m + r_{mi}$ (3), we can approximate the score vector for GEI effects accounting
 287 for the genetic main effects by $\tilde{\mathbf{S}}_{K_m} \approx \mathbf{S}_{K_m} - \mathbf{K}_m^T\hat{\mathbf{P}}_m\mathbf{G}_m(\mathbf{G}_m^T\hat{\mathbf{P}}_m\mathbf{G}_m)^{-1}\mathbf{S}_{G_m}$ where $\mathbf{G}_m =$
 288 $(\mathbf{G}_{m1}^T \mathbf{G}_{m2}^T \dots \mathbf{G}_{mN}^T)^T$ is a $N \times q$ matrix of genetic variants, $\mathbf{X}_m = (\mathbf{X}_{m1}^T \mathbf{X}_{m2}^T \dots \mathbf{X}_{mN}^T)^T$ is a $N \times p$ matrix
 289 of covariates, $\hat{\mathbf{P}}_m = \hat{\boldsymbol{\Sigma}}_m^{-1} - \hat{\boldsymbol{\Sigma}}_m^{-1}\mathbf{X}_m(\mathbf{X}_m^T\hat{\boldsymbol{\Sigma}}_m^{-1}\mathbf{X}_m)^{-1}\mathbf{X}_m^T\hat{\boldsymbol{\Sigma}}_m^{-1}$ is an $N \times N$ projection matrix from the global
 290 null model, where $\hat{\boldsymbol{\Sigma}}_m = \hat{\mathbf{V}}_m + \sum_{l=1}^L \hat{\lambda}_l \boldsymbol{\psi}_l$, $\hat{\mathbf{V}}_m = \hat{\phi}_m \mathbf{I}_N$ for quantitative traits and $\text{diag}\{\frac{1}{\hat{\mu}_{0mi}(1-\hat{\mu}_{0mi})}\}$ for
 291 binary traits, which we estimate from model (2).

292 In the meta-analysis context, we save the single-variant scores of genetic main effects and GEI effects,

293 $\mathbf{S}_{G_m}, \mathbf{S}_{K_m}$, as well as the covariance matrix for the score vectors $\boldsymbol{\Omega}_m = \begin{bmatrix} \mathbf{G}_m^T\hat{\mathbf{P}}_m\mathbf{G}_m & \mathbf{G}_m^T\hat{\mathbf{P}}_m\mathbf{K}_m \\ \mathbf{K}_m^T\hat{\mathbf{P}}_m\mathbf{G}_m & \mathbf{K}_m^T\hat{\mathbf{P}}_m\mathbf{K}_m \end{bmatrix} =$

294 $\begin{bmatrix} \mathbf{V}_{G_m} & \mathbf{C}_m \\ \mathbf{C}_m^T & \mathbf{V}_{K_m} \end{bmatrix}$ for each participant study m , where \mathbf{S}_{G_m} is a $q \times 1$ vector, \mathbf{S}_{K_m} is a $dq \times 1$ vector, $\mathbf{V}_{G_m} =$

295 $\mathbf{G}_m^T\hat{\mathbf{P}}_m\mathbf{G}_m$ is a $q \times q$ matrix, $\mathbf{V}_{K_m} = \mathbf{K}_m^T\hat{\mathbf{P}}_m\mathbf{K}_m$ is a $dq \times dq$ matrix, and $\mathbf{C}_m = \mathbf{G}_m^T\hat{\mathbf{P}}_m\mathbf{K}_m$ is a $q \times dq$

296 matrix. As $\boldsymbol{\Omega}_m$ is a large and symmetric matrix, we can save the lower triangle elements in a binary

297 format data to save space. Compared to saving the summary statistics as the double-precision floating

298 numbers that each takes 8 bytes, we used 4 bytes to save each floating-point number to further save the

299 space by 50%, without losing accuracy in the downstream meta-analysis.

300 The test statistics are recovered using two combination strategies based on the assumption that genetic

301 effects are homogeneous or heterogeneous across studies.

302 **Combine summary statistics assuming homogeneous effects**

303 We combine the test statistics by summing the scores and their covariance matrices if we assume the

304 genetic effects are homogeneous across all the studies. For example, for $m = 1, 2 \dots M$ studies, we sum

305 the score statistics by $\mathbf{S}_G = \sum_{m=1}^M \mathbf{S}_{G_m}$, $\mathbf{S}_K = \sum_{m=1}^M \mathbf{S}_{K_m}$ and covariance matrix $\mathbf{\Omega} = \sum_{m=1}^M \mathbf{\Omega}_m =$

306 $\begin{bmatrix} \sum_{m=1}^M \mathbf{V}_{G_m} & \sum_{m=1}^M \mathbf{C}_m \\ \sum_{m=1}^M \mathbf{C}_m^T & \sum_{m=1}^M \mathbf{V}_{K_m} \end{bmatrix} = \begin{bmatrix} \mathbf{V}_G & \mathbf{C} \\ \mathbf{C}^T & \mathbf{V}_K \end{bmatrix}$. Assuming all participating studies have q variants, \mathbf{S}_G is a $q \times 1$

307 vector, \mathbf{S}_K is a $dq \times 1$ vector, \mathbf{V}_G is a $q \times q$ matrix, \mathbf{V}_K is a $dq \times dq$ matrix, and \mathbf{C} is a $q \times dq$ matrix. In
308 practice, each variant does not need to be included in all the studies. If a variant is absent in study m , we
309 put a placeholder of 0 in place of this variant.

310 **Combine summary statistics assuming heterogeneous effects**

311 Suppose B ethnic groups are studied, some variants can have different allele frequencies depending on the
312 ancestry. Also, lifestyles (e.g., eating habits) and environmental exposures might modify the gene
313 differently, thus their effects on GEI might differ among ethnicities. By correctly accounting for
314 heterogeneity across studies, the power of the meta-analysis should be improved (4). Assuming

315 homogeneous within-ancestry but heterogeneous between-ancestry genetic and GEI effects, we sum the
316 score statistics by $\mathbf{S}_G = (\sum_{m=1}^{M_1} \mathbf{S}_{G1_m}^T \sum_{m=1}^{M_2} \mathbf{S}_{G2_m}^T \cdots \sum_{m=1}^{M_B} \mathbf{S}_{GB_m}^T)^T$, $\mathbf{S}_K =$

317 $(\sum_{m=1}^{M_1} \mathbf{S}_{K1_m}^T \sum_{m=1}^{M_2} \mathbf{S}_{K2_m}^T \cdots \sum_{m=1}^{M_B} \mathbf{S}_{KB_m}^T)^T$ and the covariance matrices by $\mathbf{\Omega} = \text{diag}\{\sum_{m=1}^{M_b} \mathbf{\Omega}_{b_m}\}$ over

318 the studies, where \mathbf{S}_{Gb_m} , \mathbf{S}_{Kb_m} , and $\mathbf{\Omega}_{b_m}$ are the genetic score vector, GEI score vector, and covariance
319 matrix of study m in ethnic group b , respectively. M_b is the number of studies in ethnic group b and

320 $\sum_{b=1}^B M_b = M$. $\mathbf{\Omega}$ can be partitioned to $\text{diag}\left\{\sum_{m=1}^{M_b} \begin{bmatrix} \mathbf{V}_{Gb_m} & \mathbf{C}_{b_m} \\ \mathbf{C}_{b_m}^T & \mathbf{V}_{Kb_m} \end{bmatrix}\right\}$, and the variance matrices for

321 genetic effects, GEI effects, and the covariance matrix of genetic and GEI effects are $\mathbf{V}_G =$

322 $\text{diag}\{\sum_{m=1}^{M_b} \mathbf{V}_{Gb_m}\}$, $\mathbf{V}_K = \text{diag}\{\sum_{m=1}^{M_b} \mathbf{V}_{Kb_m}\}$, and $\mathbf{C} = \text{diag}\{\sum_{m=1}^{M_b} \mathbf{C}_{b_m}\}$, respectively. The lengths are

323 then Bq for \mathbf{S}_G and Bdq for \mathbf{S}_K . \mathbf{V}_G , \mathbf{V}_K , and \mathbf{C} are block-diagonal matrices with B blocks of $q \times q$,

324 $dq \times dq$, and $q \times dq$ sub-matrices respectively for each ancestry.

325 **GEI tests**

326 In the MAGEE framework (10), we developed two GEI tests: interaction variance component (IV) test

327 and interaction hybrid test using Fisher's method (IF). The same GEI tests can be reconstructed for the

328 meta-analysis from the combined score vectors and the covariance matrices. IV test assumes the overall
 329 GEI effects $\boldsymbol{\gamma} \sim N(0, \tau \mathbf{W}_K^2)$, for which the test statistic is

$$330 \quad T_{\boldsymbol{\gamma}} = \tilde{\mathbf{S}}_K^T \mathbf{W}_K \mathbf{W}_K \tilde{\mathbf{S}}_K,$$

331 where $\tilde{\mathbf{S}}_K \approx \mathbf{S}_K - \mathbf{C}^T \mathbf{V}_G^{-1} \mathbf{S}_G$ and \mathbf{W}_K is a predefined diagonal weight matrix for interaction. $T_{\boldsymbol{\gamma}}$ follows a
 332 distribution of $\sum_{j=1}^{Bdq} \xi_{\boldsymbol{\gamma},j} \chi_{1,j}^2$, where $\chi_{1,j}^2$ are independent chi-square distributions with 1 degree of
 333 freedom (df), and $\xi_{\boldsymbol{\gamma},j}$ are the eigenvalues of $\mathbf{W}_K \boldsymbol{\Lambda} \mathbf{W}_K$, where $\boldsymbol{\Lambda} = \mathbf{V}_K - \mathbf{C}^T \mathbf{V}_G^{-1} \mathbf{C}$.

334 \mathbf{W}_K can be flexibly determined using annotation information or MAFs in the combined studies without
 335 having to access the individual-level data (36-38). One of the most popular ways to add weight to variants
 336 based on MAF is to use a beta density function with parameters 1 and 25 for the MAF (6). The dimension
 337 of \mathbf{W}_K is $Bdq \times Bdq$, in which $B = 1$ if we combine the summary statistics assuming homogeneous
 338 genetic effects.

339 IF test assumes the overall GEI effects $\boldsymbol{\gamma} \sim N(\mathbf{W}_K \mathbf{1}_{Bdq} \gamma_0, \tau \mathbf{W}_K^2)$, where $\mathbf{1}_{Bdq}$ is a vector of 1's with
 340 length Bdq . To test for $H_0: \gamma_0 = \tau = 0$, we need two test statistics, T_{γ_0} for the burden effects γ_0 and T_{τ} for
 341 the variance component effects τ . The burden effects test statistic

$$342 \quad T_{\gamma_0} = \tilde{\mathbf{S}}_K^T \mathbf{W}_K \mathbf{1}_{Bdq} \mathbf{1}_{Bdq}^T \mathbf{W}_K \tilde{\mathbf{S}}_K$$

343 follows the distribution of $\xi_{\gamma_0} \chi_1^2$ under the null hypothesis $H_0: \gamma_0 = 0$, where $\xi_{\gamma_0} = \mathbf{1}_{Bdq}^T \mathbf{W}_K \boldsymbol{\Lambda} \mathbf{W}_K \mathbf{1}_{Bdq}$.

344 The variance component effects test statistic is

$$345 \quad T_{\tau} = \tilde{\mathbf{S}}_{KV}^T \mathbf{W}_K \mathbf{W}_K \tilde{\mathbf{S}}_{KV},$$

346 where $\tilde{\mathbf{S}}_{KV} \approx \tilde{\mathbf{S}}_K - \boldsymbol{\Lambda} \mathbf{W}_K \mathbf{1}_{Bdq} (\mathbf{1}_{Bdq}^T \mathbf{W}_K \boldsymbol{\Lambda} \mathbf{W}_K \mathbf{1}_{Bdq})^{-1} \mathbf{1}_{Bdq}^T \mathbf{W}_K \tilde{\mathbf{S}}_K$. T_{τ} follows the distribution of

347 $\sum_{j=1}^{Bdq} \xi_{\tau,j} \chi_{1,j}^2$ under the null hypothesis $H_0: \tau = 0$, where $\xi_{\tau,j}$ are eigenvalues for $\mathbf{W}_K \boldsymbol{\Lambda} \mathbf{W}_K -$

348 $\mathbf{W}_K \boldsymbol{\Lambda} \mathbf{W}_K \mathbf{1}_{Bdq} (\mathbf{1}_{Bdq}^T \mathbf{W}_K \boldsymbol{\Lambda} \mathbf{W}_K \mathbf{1}_{Bdq})^{-1} \mathbf{1}_{Bdq}^T \mathbf{W}_K \boldsymbol{\Lambda} \mathbf{W}_K$. Since T_{γ_0} and T_{τ} are asymptotically independent

349 (proved by Wang *et al.* (10)), a Fisher's combination (39) can be used to find the p value for the IF test:

350 $p_{IF} = P(\chi_4^2 > -2 \log p_{\gamma_0} - 2 \log p_{\tau})$, where p_{γ_0} and p_{τ} are p values from the burden-type test $H_0: \gamma_0 =$
351 0 (under the assumption $\tau = 0$) and the adjusted variance component test $H_0: \tau = 0$, respectively.

352 **Joint tests**

353 In the joint test, which evaluates both genetic effects and GEI effects, β and γ , the test statistics for β
354 must also be reconstructed after summing the summary statistics across all studies. A meta-analysis
355 framework for genetic main effects tests has been developed by SMMAT, which supports both SKAT test
356 and unified burden and SKAT test. Wang *et al.* (10) has shown that those test statistics for the genetic
357 main effects and the GEI test are mutually independent. Assume that the p value from the meta-analysis
358 SKAT is p_{MV} and the p value from the meta-analysis IV test is p_{IV} , we can combine them through
359 Fisher's method as $p_{JV} = P(\chi_4^2 > -2 \log p_{MV} - 2 \log p_{IV})$ for the joint variance component (JV) test.
360 Similarly, assuming p_{β_0} and p_{σ} are the p values from the hybrid genetic main effects meta-analysis, we
361 can combine them with their GEI test counterparts p_{γ_0} and p_{τ} through $p_{JF} = P(\chi_8^2 > -2 \log p_{\beta_0} -$
362 $2 \log p_{\sigma} - 2 \log p_{\gamma_0} - 2 \log p_{\tau})$, which is the joint hybrid test using Fisher's method (JF). Alternatively,
363 we can combine the p values for genetic effects separately by $p_{MF} = P(\chi_4^2 > -2 \log p_{\beta_0} - 2 \log p_{\sigma})$,
364 and then combine the main effects and GEI p values by $p_{JD} = P(\chi_4^2 > -2 \log p_{MF} - 2 \log p_{IF})$, which
365 is the joint hybrid test using double Fisher's procedures (JD).

366 **Numerical Simulations**

367 Using related individuals from the same population and different populations, we conducted simulation
368 studies with two purposes: (1) benchmarking p values from the meta-analysis to the pooled analysis of
369 individual-level data from MAGEE; and (2) evaluating the performance of each test under the conditions
370 of heterogeneous genetic influence across populations.

371 We simulated the genotype replicates in two different populations using MS (40) and prepared two
372 scenarios based on the two populations. In scenario 1, we simulated 100,000 individuals from the same
373 population, which included 50,000 parents and 50,000 children in a family relatedness pattern with a

374 kinship matrix of $\begin{bmatrix} 0.5 & 0.0 & 0.25 & 0.25 \\ 0.0 & 0.5 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.5 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.5 \end{bmatrix}$. The 100,000 samples were split into two studies of 50,000

375 individuals (12,500 families) each, and the covariates were the same in the two studies. In scenario 2, we
 376 simulated 2 populations with a migration rate of 10. Each population had 50,000 individuals, including
 377 25,000 parents and 25,000 children who were related to each other in the same way as scenario 1. The
 378 covariates in these two studies were different. The simulation scenarios and covariates setup for each
 379 study are summarized in Table 2. Each scenario had 1,000 genotype replicates of 100 variants in a group.
 380 Both qualitative and binary traits were simulated, and the following section describes the details for each
 381 simulation.

382 Table 2. Summary of simulation scenarios.

	Population 1	Population 2	Covariates
Scenario 1: homogeneous genetic main effects and GEI effects			
Study 1	50,000	0	Age, sex, BMI
Study 2	50,000	0	Age, sex, BMI
Scenario 2: heterogeneous genetic main effects and GEI effects			
Study 1	50,000	0	Age, sex, BMI
Study 2	0	50,000	Sex, BMI

383

384 Comparison of p values and power

385 We simulated 10 phenotype replicates per genotype replicate. In each study, the quantitative traits of
 386 individual j in family i were simulated from

$$387 \quad y_{ij} = \alpha_1 Age_{ij} + \alpha_2 Sex_{ij} + \alpha_3 BMI_{ij} + \sum_{t=1}^q \beta_t (G_{ijt} - \bar{G}_t) + \sum_{t=1}^q \gamma_t (K_{ijt} - \bar{K}_t) + r_{ij} + \varepsilon_{ij},$$

388 where $Age_{ij} \sim N(50, 5)$ for parents and $Age_{ij} \sim N(50, 5)$ for children, $Sex_{ij} \sim Bernoulli(0.5)$ for both
 389 parents and children, BMI for family i follows a distribution with heritability (41) of 0.75 (44) $BMI_i \sim$

$$390 \quad N\left(\begin{bmatrix} 25 \\ 25 \\ 25 \\ 25 \end{bmatrix}, \begin{bmatrix} 4.0 & 0.0 & 1.5 & 1.5 \\ 0.0 & 4.0 & 1.5 & 1.5 \\ 1.5 & 1.5 & 4.0 & 1.5 \\ 1.5 & 1.5 & 1.5 & 4.0 \end{bmatrix}\right), \text{ the random effects for family } i \quad r_i \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1.0 & 0.0 & 0.5 & 0.5 \\ 0.0 & 1.0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1.0 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1.0 \end{bmatrix}\right), \text{ and the}$$

391 random error $\varepsilon_{ij} \sim N(0, 1)$. G_{ijt} was the t -th genetic variant for individual j in family i and \bar{G}_t was the

392 mean of the t -th variant, and $K_{ijt} = (G_{ijt} - \bar{G}_t)(BMI_{ij} - \bar{BMI})$ was the GEI term for the t -th variant with
393 BMI for individual j from family i and \bar{K}_t was the mean of the t -th interaction term in that study.

394 The binary traits were simulated using a logistic regression model from

$$395 \log\left(\frac{P(y_{ij}=1)}{1-P(y_{ij}=1)}\right) = \alpha_0 + \alpha_1(Age_{ij} - \bar{Age}) + \alpha_2(Sex_{ij} - \bar{Sex}) + \alpha_3(BMI_{ij} - \bar{BMI}) +$$
$$396 \sum_{t=1}^q \beta_t(G_{ijt} - \bar{G}_t) + \sum_{t=1}^q \gamma_t(K_{ijt} - \bar{K}_t) + r_{ij},$$

397 where Age , Sex , BMI , and the random effects for family i r_i followed the same distribution as the
398 quantitative traits, and \bar{Age} , \bar{Sex} , and \bar{BMI} were the population means in each study, α_0 was set to $\log\frac{0.4}{1-0.4}$
399 representing a prevalence rate of 0.4. G_{ijt} , \bar{G}_t and the calculation of K_{ijt} and \bar{K}_t were the same as the
400 quantitative traits.

401 For scenario 1, the parameters for age, sex, and BMI were $\alpha_1 = 0.1$, $\alpha_2 = 0.2$, and $\alpha_3 = 0.1$, the same
402 for study 1 and study 2. The effect sizes β_t for G and γ_t for GEI were also the same across in both
403 studies. In scenario 2, the parameters for the covariates in study 1 were the same as those for scenario 1,
404 while the parameters in study 2 were $\alpha_1 = 0$, $\alpha_2 = 0.4$, and $\alpha_3 = 0.12$, so that age was excluded from
405 the covariates. In addition, the effect sizes for G and GEI effects were different between the 2 studies in
406 scenario 2. Specifically, the effect sizes of a variant β_t and γ_t were proportional to its MAF as
407 $c \log_{10}(MAF_t)$, where c represents a constant. Detailed information of the constants c for β_t and γ_t in
408 each simulation can be found in supplemental Table S1 for the p value benchmark, supplemental Table
409 S2 for the quantitative trait power comparison, and supplemental Table S3 for the binary trait power
410 comparison.

411 For the p value benchmark, we used different parameters for three different sample sizes. In total, 20% of
412 the variants were randomly selected as causal, which included 10% positive effects and 10% negative
413 effects. For power comparisons, the same parameters were used with different sample sizes, but different
414 parameters were used for GEI and joint tests. We used two approaches to indicate the proportion of causal
415 variants: (1) we randomly chose 20% of causes, among which 10% were positive and 10% were negative;

416 (2) we randomly selected 20% of causes, among which 16% were positive and 4% were negative. The
417 gene-BMI interaction was tested both for binary and quantitative traits. According to Wu *et al.* (6), we
418 used a beta density weight function with parameters 1 and 25 on the MAF, so that rare variants had a
419 higher weight than common variants.

420 **Application to UK Biobank whole exome sequencing data**

421 We used the UK Biobank (42) whole exome sequencing (WES) data as an example of our proposed meta-
422 analysis. The UK Biobank has released the first tranche of WES data containing 49,959 samples in March
423 2019. The second tranche of WES data was released in October 2020 with 200,632 samples (including the
424 samples in the first tranche). We divided the dataset in the second tranche by the first and second batches
425 to meta-analyze the two datasets because there may be measurement errors or batch-specific
426 heterogeneities due to differences in quality control procedures. After excluding the subjects with missing
427 sex information, missing age, BMI or WHR, non-white British, and those who have withdrawn from the
428 study, a total of 43,190 subjects were eligible for batch 1 (23,372 women and 19,818 men), and 132,058
429 subjects were eligible for batch 2 (72,961 women and 59,097 men). The age distribution was mean =
430 56.95 ($sd = 7.89$) in batch 1 and mean = 56.75 ($sd = 8.04$) in batch 2. The BMI distribution was mean =
431 27.40 ($sd = 4.77$) in batch 1 and mean = 27.34 ($sd = 4.70$) in batch 2. We tested for gene-by-sex
432 interaction effects on WHR at the gene-level for the two batches. We first fitted a linear mixed model (43)
433 adjusting for sex, age, age^2 and the top ten ancestry principal components (PCs) for each batch using
434 `glmmkin` function from the GMMAT package,

$$435 \quad \text{WHR} \sim \text{sex} + \text{age} + \text{age}^2 + \text{BMI} + \text{PC1} + \text{PC2} + \dots + \text{PC10}.$$

436 The relatedness matrix was constructed using the kinship coefficients computed by KING software (44)
437 for third-degree and closer relatives, provided by the UK Biobank (see UK Biobank Resource 531).

438 We used the genotype data generated by the Functionally Equivalent (FE) pipeline (45). The variant
439 groups were defined as protein-coding regions for the WES data and a total of 19,449 protein-coding
440 regions with 17,549,650 variants were available for analysis. MAGEE was used to generate the score

441 statistics and the covariance matrices for each batch. In the next step, we combined and meta-analyzed the
442 scores and covariance matrices for the two batches, assuming homogeneous genetic effects. As IF test
443 was more powerful than the IV test, and JF test was slightly more powerful than JV and JD tests in our
444 power simulations, in the UK Biobank WES data analysis we reported IF and JF test p values as the GEI
445 and joint test results, respectively. The meta-analysis excluded variants with a minor allele count (MAC)
446 less than 5, or with a genotype missing rate greater than 5% in the combined dataset, and 18,668 protein-
447 coding regions with 2,328,550 variants passed these filters. We applied a beta density weight function
448 with parameters 1 and 25 on the MAF. In addition, we compared the meta-analysis results to the results of
449 a joint analysis pooling all the individual-level data using MAGEE. For protein-coding regions that
450 reached the Bonferroni-corrected significance level of $0.05/18,668 = 2.7 \times 10^{-6}$, we also performed
451 conditional analysis by adjusting for variants in the region that had a single variant joint test p value less
452 than 5.0×10^{-8} , MAF greater than 1% and genotype missing rate less than 5%. Both the genetic main effect
453 and gene-sex interaction terms of significant single variants were included in the conditional analysis.

454

455 **References**

- 456 (1) Hunter DJ. Gene-environment interactions in human diseases. *Nat Rev Genet* 2005;6(4):287-298.
- 457 (2) Thomas D. Gene-environment-wide association studies: emerging approaches. *Nat Rev Genet*
458 2010;11(4):259-272.
- 459 (3) Hu Y, Berndt SI, Gustafsson S, Ganna A, Hirschhorn J, North KE, et al. Meta-analysis of gene-level
460 associations for rare variants based on single-variant statistics. *Am J Hum Genet* 2013;93(2):236-48.
- 461 (4) Lee S, Abecasis GR, Boehnke M, Lin X. Rare-Variant Association Analysis: Study Designs and
462 Statistical Tests. *Am J Hum Genet* 2014;95(1):5-23.
- 463 (5) Li B, Leal SM. Methods for Detecting Associations with Rare Variants for Common Diseases:
464 Application to Analysis of Sequence Data. *The American Journal of Human Genetics* 2008;83(3):311-
465 321.
- 466 (6) Wu M, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-Variant Association Testing for Sequencing Data
467 with the Sequence Kernel Association Test. *The American Journal of Human Genetics* 2011;89(1):82-93.

- 468 (7) Price AL, Kryukov GV, de Bakker PIW, Purcell SM, Staples J, Wei L, et al. Pooled Association Tests
469 for Rare Variants in Exon-Resequencing Studies. *Am J Hum Genet* 2010;86(6):832-838.
- 470 (8) Chen H, Meigs JB, Dupuis J. Incorporating Gene-Environment Interaction in Testing for Association
471 with Rare Genetic Variants. *Hum Hered* 2014;78(2):81-90.
- 472 (9) Su Y, Di C, Hsu L. A unified powerful set-based test for sequencing data analysis of GxE interactions.
473 *Biostatistics (Oxford, England)* 2017;18(1):119.
- 474 (10) Wang X, Lim E, Liu C, Sung YJ, Rao DC, Morrison AC, et al. Efficient gene-environment
475 interaction tests for large biobank-scale sequencing studies. *Genet Epidemiol* 2020;44(8):908-923.
- 476 (11) Kraft P, Yen Y, Stram DO, Morrison J, Gauderman WJ. Exploiting Gene-Environment Interaction to
477 Detect Genetic Associations. *Hum Hered* 2007;63(2):111-119.
- 478 (12) Sung YJ, Winkler TW, Manning AK, Aschard H, Gudnason V, Harris TB, et al. An Empirical
479 Comparison of Joint and Stratified Frameworks for Studying $G \times E$ Interactions: Systolic Blood Pressure
480 and Smoking in the CHARGE Gene-Lifestyle Interactions Working Group. *Genet Epidemiol*
481 2016;40(5):404-415.
- 482 (13) Lee S, Teslovich TM, Boehnke M, Lin X. General Framework for Meta-analysis of Rare Variants in
483 Sequencing Association Studies. *Am J Hum Genet* 2013;93(1):42-53.
- 484 (14) Liu DJ, Peloso GM, Zhan X, Holmen OL, Zawistowski M, Feng S, et al. Meta-analysis of gene-level
485 tests for rare variant association. *Nat Genet* 2014;46(2):200-204.
- 486 (15) Feng S, Liu D, Zhan X, Wing MK, Abecasis GR. RAREMETAL: fast and powerful meta-analysis
487 for rare variants. *Bioinformatics* 2014;30(19):2828-2829.
- 488 (16) Yang J, Chen S, Abecasis G. Improved score statistics for meta-analysis in single-variant and gene-
489 level association studies. *Genet Epidemiol* 2018;42(4):333-343.
- 490 (17) Tang Z, Lin D. MASS: meta-analysis of score statistics for sequencing studies. *Bioinformatics*
491 2013;29(14):1803-1805.
- 492 (18) He Z, Xu B, Lee S, Ionita-Laza I. Unified Sequence-Based Association Tests Allowing for Multiple
493 Functional Annotations and Meta-analysis of Noncoding Variation in MetaboChip Data. *Am J Hum Genet*
494 2017;101(3):340-352.
- 495 (19) Ka J, Lee J, Kim Y, Oh B, Park T. Meta-Qtest: meta-analysis of quadratic test for rare variants. *BMC*
496 *medical genomics; BMC Med Genomics* 2019;12:102-102.
- 497 (20) Chen H, Huffman JE, Brody JA, Wang C, Lee S, Li Z, et al. Efficient Variant Set Mixed Model
498 Association Tests for Continuous and Binary Traits in Large-Scale Whole-Genome Sequencing Studies.
499 *The American Journal of Human Genetics* 2019;104(2):260-274.
- 500 (21) Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association
501 scans. *Bioinformatics* 2010;26(17):2190-2191.

- 502 (22) Wang J, Liu Q, Pierce BL, Huo D, Olopade OI, Ahsan H, et al. A meta-analysis approach with
503 filtering for identifying gene-level gene–environment interactions. *Genet Epidemiol* 2018;42(5):434-446.
- 504 (23) Jin X, Shi G. Variance-Component-Based Meta-Analysis of Gene-Environment Interactions for Rare
505 Variants. G3 (Bethesda, Md.) 2021.
- 506 (24) Aschard H, Hancock DB, London SJ, Kraft P. Genome-Wide Meta-Analysis of Joint Tests for
507 Genetic and Gene-Environment Interaction Effects. *Hum Hered* 2010;70(4):292-300.
- 508 (25) Manning AK, LaValley M, Liu C, Rice K, An P, Liu Y, et al. Meta-analysis of gene-environment
509 interaction: joint estimation of SNP and SNP \times environment regression coefficients. *Genet Epidemiol*
510 2011;35(1):11-18.
- 511 (26) Bland JM, Altman DG. Multiple Significance Tests: The Bonferroni Method. *BMJ: British Medical*
512 *Journal* 1995;310(6973):170-170.
- 513 (27) Lumish HS, O'Reilly M, Reilly MP. Sex Differences in Genomic Drivers of Adipose Distribution
514 and Related Cardiometabolic Disorders: Opportunities for Precision Medicine. *Arterioscler Thromb Vasc*
515 *Biol* 2020;40(1):45-60.
- 516 (28) Shungin D, Winkler TW, Croteau-Chonka D, Ferreira T, Locke AE, Mägi R, et al. New genetic loci
517 link adipose and insulin biology to body fat distribution. *Nature* 2015;518(7538):187-U378.
- 518 (29) Justice AE, Karaderi T, Highland HM, Young KL, Graff M, Lu Y, et al. Protein-coding variants
519 implicate novel genes related to lipid homeostasis contributing to body-fat distribution. *Nat Genet*
520 2019;51(3):452-469.
- 521 (30) Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, et al. Meta-analysis
522 identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis
523 of fat distribution. *Nat Genet* 2010;42(11):949-U160.
- 524 (31) Lotta LA, Wittemans L, Zuber V, Stewart ID, Sharp SJ, Luan J, et al. Association of genetic variants
525 related to gluteofemoral vs abdominal fat distribution With type 2 diabetes, coronary disease, and
526 cardiovascular risk factors. 2018.
- 527 (32) Zhu Z, Guo Y, Shi H, Liu C, Panganiban RA, Chung W, et al. Shared genetic and experimental links
528 between obesity-related traits and asthma subtypes in UK Biobank. *J Allergy Clin Immunol*
529 2020;145(2):537-549.
- 530 (33) Hill AVS, Allsopp CEM, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, et al. Common West
531 African HLA antigens are associated with protection from severe malaria. *Nature* 1991;352(6336):595-
532 600.
- 533 (34) Tang Z, Lin D. Meta-Analysis of Sequencing Studies With Heterogeneous Genetic Associations.
534 *Genet Epidemiol* 2014;38(5):389-401.
- 535 (35) Lee YH. Meta-analysis of genetic association studies. *Annals of laboratory medicine*
536 2015;35(3):283.

- 537 (36) Kircher M, Witten DM, Jain P, O'Roak B,J., Cooper GM, Shendure J. A general framework for
538 estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46(3):310-315.
- 539 (37) Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of
540 variants throughout the human genome. *Nucleic Acids Res* 2019;47:D886-D894.
- 541 (38) Rogers MF, Shihab HA, Mort M, Cooper DN, Gaunt TR, Campbell C. FATHMM-XF: accurate
542 prediction of pathogenic point mutations via extended features. *Bioinformatics* 2018;34(3):511-513.
- 543 (39) Fisher RA. *Statistical Methods for Research Workers*. Second Edition, Revised and enlarged.
544 Edinburgh: Oliver and Boyd. *Journal of Comparative Pathology and Therapeutics* 1928;41:261-262.
- 545 (40) Hudson RR. Generating samples under a Wright–Fisher neutral model of genetic variation.
546 *Bioinformatics* 2002;18(2):337-338.
- 547 (41) Elks CE, Den Hoed M, Zhao JH, Sharp SJ, Wareham NJ, Loos RJF, et al. Variability in the
548 heritability of body mass index: a systematic review and meta-regression. *Frontiers in endocrinology*
549 2012;3:29.
- 550 (42) Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource
551 with deep phenotyping and genomic data. *Nature* 2018;562(7726):203-209.
- 552 (43) Conomos MP, Laurie CA, Stilp AM, Gogarten SM, Mchugh CP, Nelson SC, et al. Genetic Diversity
553 and Association Studies in US Hispanic/Latino Populations: Applications in the Hispanic Community
554 Health Study/Study of Latinos. *Am J Hum Genet* 2016;98(1):165.
- 555 (44) Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W. Robust relationship inference
556 in genome-wide association studies. *Bioinformatics* 2010;26(22):2867-2873.
- 557 (45) Regier AA, Farjoun Y, Larson DE, Krasheninina O, Kang HM, Howrigan DP, et al. Functional
558 equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human
559 genetics projects. *Nature communications* 2018;9(1):4038.

560

561 **Declarations**

562 Ethics approval and consent to participate

563 This study was approved by the Institutional Review Board (IRB) of The University of Texas Health
564 Science Center at Houston (UTHealth) with project number HSC-SPH-18-0466. Waiver of informed
565 consent was granted.

566 Availability of data and materials

567 The individual-level data that support the findings of this study are available upon application to the UK
568 Biobank (<https://www.ukbiobank.ac.uk/register-apply/>).
569 The method is implemented in R package MAGEE, available at [https://github.com/large-scale-gxe-](https://github.com/large-scale-gxe-methods/MAGEE)
570 [methods/MAGEE](https://github.com/large-scale-gxe-methods/MAGEE).

571 Competing interests

572 The authors declare that they have no competing interests.

573 Funding

574 This research was supported by National Institutes of Health (NIH) grant R01 HL145025.

575 Authors' Information

576 *Affiliations*

577 **Department of Biostatistics and Data Science, School of Public Health, The University of Texas**
578 **Health Science Center at Houston, Houston, TX, USA**

579 Xinyu Wang

580 **Human Genetics Center, Department of Epidemiology, Human Genetics and Environmental**
581 **Sciences, School of Public Health, The University of Texas Health Science Center at Houston,**
582 **Houston, TX, USA**

583 Duy T. Pham, Cong Pan & Han Chen

584 **Clinical and Translational Epidemiology Unit, Mongan Institute, Massachusetts General Hospital,**
585 **Boston, MA, USA**

586 Kenneth E. Westerman & Alisa K. Manning

587 **Metabolism Program, Broad Institute of MIT and Harvard, Cambridge, MA, USA**

588 Kenneth E. Westerman & Alisa K. Manning

589 **Department of Medicine, Harvard Medical School, Boston, MA, USA**

590 Kenneth E. Westerman & Alisa K. Manning

591 **Center for Precision Health, School of Public Health and School of Biomedical Informatics, The**

592 **University of Texas Health Science Center at Houston, Houston, TX, USA**

593 Han Chen

594 *Corresponding author*

595 Correspondence to: Han Chen

596 **Acknowledgements**

597 This research has been conducted using the UK Biobank Resource under Application Number 42646.

598 This work was supported by National Institutes of Health (NIH) grant R01 HL145025.

599