Genotype-by-environment interactions inferred from genetic effects on phenotypic variability in the UK Biobank

Huanwei Wang1, Futao Zhang1, Jian Zeng1, Yang Wu1, Kathryn E. Kemper1, Angli Xue1, Min Zhang1, Joseph E. Powell1,2,3, Michael E. Goddard4,5, Naomi R. Wray1,6, Peter M. Visscher1,6, Allan F. McRae1, Jian Yang1,7*

Genotype-by-environment interaction (GEI) is a fundamental component in understanding complex trait variation. However, it remains challenging to identify genetic variants with GEI effects in humans largely because of the small effect sizes and the difficulty of monitoring environmental fluctuations. Here, we demonstrate that GEI can be inferred from genetic variants associated with phenotypic variability in a large sample without the need of measuring environmental factors. We performed a genome-wide variance quantitative trait locus (vQTL) analysis of ~5.6 million variants on 348,501 unrelated individuals of European ancestry for 13 quantitative traits in the UK Biobank and identified 75 significant vQTLs with \( P < 2.0 \times 10^{-9} \) for 9 traits, especially for those related to obesity. Direct GEI analysis with five environmental factors showed that the vQTLs were strongly enriched with GEI effects. Our results indicate pervasive GEI effects for obesity-related traits and demonstrate the detection of GEI without environmental data.

INTRODUCTION

Most human traits are complex because they are affected by many genetic and environmental factors as well as potential interactions between them (1, 2). Despite the long history of effort (3–5), there has been limited success in identifying genotype-by-environment interaction (GEI) effects in humans (5–8). This is likely because many environmental exposures are unknown or difficult to record during the life course and because the effect sizes of GEI are small, given the polygenic nature of most human traits (9–11), so that the sample sizes of most previous studies are not large enough to detect the small GEI effects. For model complex traits such as body mass index (BMI), GEI analyses have been limited to GEI tests at known BMI loci (12–14) or estimation of GEI variance captured by all common SNPs (15, 16).

GEI effect of a genetic variant on a quantitative trait could lead to differences in variance of the trait among groups of individuals with different variant genotypes (Fig. 1, A and B, and note SI). GEI can therefore be inferred from a variance quantitative trait locus (vQTL) analysis (17), although there are other explanations for an observed vQTL such as direct effect on phenotypic dispersion [e.g., induced by selection (18)], epistasis (17), and phantom vQTL (19, 20). Unlike the classical QTL analysis that tests the allelic substitution effect of a variant on the mean of a phenotype (Fig. 1C), vQTL analysis tests the allelic substitution effect on the trait variance (Fig. 1, B or D). In comparison to the analyses that perform direct GEI tests, vQTL analysis is more flexible because it does not require measures of environmental factors and thus can be performed in a very large sample where the environmental factors are unknown, unavailable, or incomplete (21).

Of course, the vQTL test is less powerful than the direct GEI test if the corresponding environmental factor has been measured on all the genotyped individuals in the sample (17). Although there had been empirical evidence for the genetic control of phenotypic variance in livestock for decades (22, 23), it was not until recent years that genome-wide vQTL analysis was applied in humans (17, 24, 25), and only a handful of vQTLs have been identified for a limited number of traits [e.g., the FTO locus for BMI (25)] owing to small effect sizes of vQTLs. The availability of data from large biobank-based genome-wide association studies (GWAS) (26, 27) provides an opportunity to interrogate the genome for vQTLs for a range of phenotypes in cohorts with unprecedented sample size.

On the other hand, statistical methods for vQTL analysis are not entirely mature (21). There have been a series of classical nonparametric methods (28), originally developed to detect violation of the homogeneous variance assumption in linear regression model, which can be used to detect vQTLs, including the Bartlett’s test (29), the Levene’s test (30, 31), and the Fligner-Killeen (FK) test (32). Recently, more flexible parametric models have been proposed, including the double generalized linear model (DGLM) (33–35) and the likelihood ratio test for variance effect (LRTV) (19). In addition, it has been suggested that the transformation of phenotype that alters phenotype distribution also has an influence on the power and/or false-positive rate (FPR) of a vQTL analysis (24, 36).

In this study, we calibrated the most commonly used statistical methods for vQTL analysis by extensive simulations. We then used the best performing method to conduct a genome-wide vQTL analysis for 13 quantitative traits in 348,501 unrelated individuals using the UK Biobank (UKB) data (26). We further investigated whether the detected vQTLs are enriched for GEI by conducting a direct GEI test for the vQTLs with five environmental factors (or covariates).

RESULTS

Evaluation of the vQTL methods by simulation

We used simulations to quantify the FPR and power (i.e., true-positive rate) for the vQTL methods and phenotype processing strategies.

1Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072, Australia. 2Garvan-Weizmann Centre for Cellular Genomics, Garvan Institute for Medical Research, Sydney, New South Wales 2010, Australia. 3Faculty of Medicine, University of New South Wales, Sydney, New South Wales 2052, Australia. 4Faculty of Veterinary and Agricultural Science, University of Melbourne, Parkville, Victoria, Australia. 5Biosciences Research Division, Department of Economic Development, Jobs, Transport and Resources, Bundoora, Victoria, Australia. 6Queensland Brain Institute, The University of Queensland, Brisbane, Queensland 4072, Australia. 7Institute for Advanced Research, Wenzhou Medical University, Wenzhou, Zhejiang 325027, China.

*Corresponding author. Email: jian.yang.Qt@gmail.com
(Methods). We first simulated a quantitative trait based on a simulated single-nucleotide polymorphism (SNP), i.e., a single-SNP model, under a number of different scenarios, namely, (i) five different distributions for the random error term (i.e., individual-specific environment effect) and (ii) four different types of SNP with or without the effect on mean or variance (Methods). We used the simulated data to compare the four most widely used vQTL methods, namely, Bartlett’s test (29), Levene’s test (30, 31), FK test (32), and DGLM (33–35). We observed no inflation in FPR for the Levene’s test under the null (i.e., no vQTL effect) regardless of the skew or kurtosis of the phenotype distribution or the presence or absence of SNP effect on the mean (fig. S1A). These findings are in line with the results from previous studies (24, 28, 37) that the Levene’s test is robust to the distribution of the phenotype. The FPR of the Bartlett’s test or DGLM was inflated if the phenotype distribution was skewed or heavy-tailed (fig. S1A). The FK test seemed to be robust to kurtosis but vulnerable to skewness of the phenotype distribution (fig. S1A). Because the Levene’s test performed the best in the simulations, for this test, we investigated the impact of nonlinear transformations of the phenotype by considering logarithm \(\log(y)\), square \(y^2\), cube \(y^3\), and rank-based inverse-normal transformation (RINT) and found that these nonlinear transformations could result in inflated FPR (fig. S1B).

To simulate more complex scenarios, we used a multiple-SNP model with two covariates (age and sex) and different numbers of SNPs (Fig. 2). The results were similar to those described above, although the power of the Levene’s test decreased with an increase of the number of causal SNPs (Fig. 2A). Nonlinear transformations led to an inflated FPR when the variance explained by a QTL effect (i.e., SNP effect on mean) was relatively large and a loss of power of vQTL detection when the per-QTL variance explained was relatively small, although logarithm transformation did not seem to affect the power (Fig. 2B). These results also suggested that pre-adjusting the phenotype by covariates slightly increased the power (Fig. 2B). On the basis of the results of these simulations, we used the Levene’s test, a one-way analysis of variance (ANOVA) to test for absolute deviations from the medians (Methods), for real data analysis, with the phenotypes pre-adjusted for covariates without any nonlinear transformation.

**Genome-wide vQTL analysis for 13 UKB traits**

We performed a genome-wide vQTL analysis using the Levene’s test with 5,554,549 genotyped or imputed common variants on 348,501 unrelated individuals of European ancestry for 13 quantitative traits in the UKB (Methods; table S1A and fig. S2A) (26). For each trait, we pre-adjusted the phenotype for age and the first 10 principal components (PCs; derived from SNP data) and standardized the residuals to \(z\) scores in each gender group (Methods). This process removed not only the effects of age and the first 10 PCs on the phenotype but also the differences in mean and variance between the two genders. We excluded individuals with adjusted phenotypes more than 5 SDs from the mean and removed SNPs with minor allele frequency (MAF) smaller than 0.05 to avoid potential false-positive associations due to the coincidence of a low-frequency variant with an outlier phenotype (see fig. S3A for an example). We acknowledge that this process could potentially result in a loss of power, but this can be compensated for by the use of a very large sample (\(n \sim 350,000\)).

With an experiment-wise significant threshold of \(2.0 \times 10^{-9}\) [i.e., \(1 \times 10^{-8}/5.0\), with \(1 \times 10^{-8}\) being a more stringent genome-wide significant threshold recommended by recent studies (38, 39) and 5.0 being the effective number of independent traits (note S4)], we identified 75 vQTLs [independent to linkage disequilibrium (LD) \(r^2 < 0.01\) within trait] across the nine traits (Fig. 3, Table 1, and table S2A). There was no vQTL for height, consistent with the observation in a previous study (25). We identified more than 15 vQTLs for each of the three obesity-related traits, i.e., BMI, waist circumference (WC), and
The 75 vQTLs were located at 41 near-independent loci after excluding one of each between-trait pair of top vQTL SNPs (i.e., the SNP with lowest vQTL P value at each vQTL association peak) with LD $r^2 > 0.01$, suggesting that some of the loci were associated with the phenotypic variance of multiple traits. For example, the FTO locus was associated with the phenotypic variance of WC, HC, BMI, body fat percentage (BFP), and basal metabolic rate (BMR) (fig. S3B), and the vQTL associations were likely to be driven by a shared causal variant with pleiotropic vQTL effects on multiple traits (fig. S3C). For the lung function–related traits, there was no significant

**Fig. 2. Evaluation of the statistical methods and phenotype processing strategies for vQTL analysis by simulation.** Phenotypes of 10,000 individuals were simulated on the basis of the different number of causal SNPs (i.e., 4, 40, or 80), two covariates (i.e., sex and age), and one error term in a multiple-SNP model (Methods). The SNP effects were simulated under four scenarios: (i) effect on neither mean nor variance (nei), (ii) effect on mean only (mean), (iii) effect on variance only (var), or (iv) effect on both mean and variance (both). The error term was generated from five different distributions: normal distribution, $t$-distribution with df = 10 or 3, or $\chi^2$ distribution with df = 15 or 1. (A) Four statistical test methods, i.e., the Bartlett’s test (Bart), the Levene’s test (Lev), the Fligner-Killeen test (FK), and the DGLM, were used to detect vQTLs. (B) The Levene’s test was used to analyze phenotypes processed using six strategies, i.e., raw phenotype (raw), raw phenotype adjusted for covariates (adj), RINT after adj (rint), logarithm transformation after adj (log), square transformation after adj (sq), and cube transformation after adj (cub). Positive rate is defined as the number of vQTLs with $P < 0.05$ divided by the total number of tests across 1000 simulations, which is the FPR under the null (“nei” and “mean”) and power under the alternative (“var” and “both”). The red horizontal line represents an FPR of 0.05.
vQTL for forced expiratory volume in 1 s (FEV1) and forced vital capacity (FVC) but there were three vQTLs for the FEV1/FVC ratio (FFR).

There was no evidence for an effect of MAF on vQTL test statistic at the 41 independent loci (fig. S3D), consistent with the observation in a previous study (25).

The Levene’s test assesses the difference in variance among three genotype groups free from the assumption about additivity (i.e., the vQTL effect of carrying two copies of the effect allele is not assumed to be twice that carrying one copy). We found two vQTLs (i.e., rs141783576 and rs10456362) potentially showing nonadditive genetic effect on the variance of HC and BMR, respectively (table S2A).

To demonstrate the vulnerability of vQTL analysis to nonlinear transformations in real data, we performed genome-wide vQTL analysis for height squared and cubed. There was no genome-wide significant vQTL for height squared but one genome-wide significant vQTL for height cubed, which was very likely to be driven by a strong...
Table 1. The number of experiment-wise significant vQTLs or QTLs for the 13 UKB traits.

| Trait       | Description                                     | Distribution of raw phenotype | Distribution of processed phenotype | Number of independent vQTLs for each trait | Number of independent QTLs for each trait |
|-------------|-------------------------------------------------|-------------------------------|------------------------------------|------------------------------------------|-------------------------------------------|
| HT          | Standing height                                 |                               |                                    | 0                                        | 1063                                      |
| FVC         | Forced vital capacity                           |                               |                                    | 0                                        | 325                                       |
| FEV1        | Forced expiratory volume in 1 s                 |                               |                                    | 0                                        | 266                                       |
| FFR         | FEV1 and FVC ratio                              |                               |                                    | 3                                        | 221                                       |
| BMD         | Heel bone mineral                               |                               |                                    | 6                                        | 267                                       |
| BW          | Birth weight                                    |                               |                                    | 1                                        | 57                                        |
| BMI         | Body mass index                                 |                               |                                    | 22                                       | 271                                       |
| WC          | Waist circumference                             |                               |                                    | 16                                       | 196                                       |
| HC          | Hip circumference                               |                               |                                    | 16                                       | 249                                       |
| WHR         | Waist-to-hip ratio                              |                               |                                    | 1                                        | 157                                       |
| WHRadjBMI   | WHR adjusted for BMI                            |                               |                                    | 0                                        | 187                                       |
| BFP         | Body fat percentage                            |                               |                                    | 5                                        | 249                                       |
| BMR         | Basal metabolic rate                            |                               |                                    | 5                                        | 465                                       |
| **Total**   |                                                 |                               |                                    | **75**                                   | **3973**                                  |
QTL signal for height \(P_{QTL(Height)} = 4.35 \times 10^{-150}\) (fig. S3, E and F), consistent with our simulation results that nonlinear transformations could inflate the vQTL test statistics in the presence of a strong QTL signal (Fig. 2B and fig. S1B). Although we have not applied any nonlinear transformation to the UKB traits, some of them are nonlinear functions of other traits, i.e., BMI (= WT/HT²), FFR (= FEV1/FVC), and WHR (= WC/HC). We therefore explored whether the BMI, FFR, and WHR vQTLs were driven by the nonlinear functions by testing the variance effects of the BMI, FFR, and WHR vQTLs on 1/HT², 1/FVC, and 1/HC, respectively. There were 26 tests in total, none of which reached the experiment-wise significance level (i.e., \(2.0 \times 10^{-9}\)) used to claim vQTLs in this study and 23 of which had a \(P\) value larger than 0.05 (table S2B), suggesting that the BMI, FFR, and WHR vQTLs were not driven by the nonlinear functions. Although the variance effect of an FFR vQTL (rs56077333) on 1/FVC was significant after correcting for 26 tests (\(P = 5.11 \times 10^{-6}\); table S2B), the effect of rs56077333 on the variance of 1/FVC was not large enough to drive the vQTL signal for FFR and rs56077333 has a known GEI effect on lung function (see below for more details).

**GWAS analysis for the 13 UKB traits**

To investigate whether the SNPs with effects on variance also have effects on mean, we performed GWAS (or genome-wide QTL) analyses for the 13 UKB traits described above. We identified 3973 QTLs at an experiment-wise significance level (i.e., \(P_{QTL} < 2.0 \times 10^{-9}\)) for the 13 traits in total, a much larger number than that of the vQTLs (Fig. 4 and Table 1). Among the 75 vQTLs, the top vQTL SNPs at nine loci did not pass the experiment-wise significance level in the QTL analysis (table S2A). For example, the CCDC92 locus showed a significant vQTL effect but no significant QTL effect on WC (table S2A and fig. S3G), whereas the FTO locus showed both significant QTL and vQTL effects on WC (fig. S3G). For the 66 vQTLs with both QTL and vQTL effects, the vQTL effects were all in the same directions as the QTL effects, meaning that, for any of these SNPs, the genotype group with a larger phenotypic mean also tends to have a larger phenotypic variance than the other groups. For the nine loci with vQTL effects only, it is equivalent to a scenario where a QTL has a GEI effect with no (or a substantially reduced) effect on average across different levels of an environmental factor (Fig. 1B).

**vQTL and GEI**

To further investigate whether the associations between vQTLs and phenotypic variance can be explained by GEI, we performed a direct GEI test based on an additive genetic model with an interaction term between a top vQTL SNP and one of five environmental factors/ covariates in the UKB data (Methods). The five environmental factors/ covariates are sex, age, physical activity (PA), sedentary behavior (SB), and ever smoking (note S5, fig. S2B, and table S1B). We observed 16 vQTLs showing a significant GEI effect with at least one of five environmental factors after Bonferroni correction for multiple tests \([P < 1.33 \times 10^{-4} = 0.05/(75 \times 5); \text{Fig.} 5A \text{and table S2C}]\).

To test whether the GEI effects are enriched among vQTLs in comparison with the same number of QTLs, we performed GEI test for the 75 top GWAS SNPs randomly selected from all the QTLs and repeated the analysis 1000 times. Of the 75 top SNPs with QTL effects, the number of SNPs with significant GEI effects was 2.25, averaged from the 1000 repeated samplings with an SD of 1.49 (Fig. 5B), significantly lower than the number \((16)\) observed for the vQTLs (the difference is larger than 9 SDs, equivalent to \(P = 1.4 \times 10^{-20}\)). This result shows that SNPs with vQTL effects are much more enriched with GEI effects compared to those with QTL effects. To exclude the possibility that the GEI signals were driven by phenotype processing (e.g., the adjustment of phenotype for sex and age), we repeated the GEI analyses using raw phenotype data without covariate adjustment; the results remain largely unchanged (fig. S5).

**DISCUSSION**

In this study, we leveraged the genetic effects associated with phenotypic variability to infer GEI. We calibrated the most commonly used vQTL methods by simulation. We found that the FPR of the Levene’s test was well calibrated across all simulation scenarios, whereas the other methods showed an inflated FPR if the phenotype distribution was skewed or heavy-tailed under the null hypothesis (i.e., no vQTL
effect), although the Levene’s test appeared to be less powerful than the other methods particularly when the per-variant vQTL effect was small (Fig. 2 and fig. S1). Parametric bootstrap or permutation procedures have been proposed to reduce the inflation in the test statistics of DGLM and LRTv, both of which are expected to be more powerful than the Levene’s test (19, 37), but bootstrap and permutation are computationally inefficient and thus are not practically applicable to biobank data such as the UKB. We observed inflated FPR for the Levene’s test in the absence of vQTL effects but in the presence of QTL effects if the phenotype was nonlinearly transformed (e.g., logarithm transformation or RINT). We therefore recommend the use of QTL effects if the phenotype was nonlinearly transformed (e.g., log).

We demonstrated in the analysis of the UKB data that a number of vQTLs with small variance effects yet to be discovered in larger samples in the future.

Fig. 5. Enrichment of GEI effects among the 75 vQTLs compared with a random set of QTLs. Five environmental factors/covariates, i.e., sex, age, physical activity (PA), sedentary behavior (SB), and smoking, were used in the GEI analysis. (A) Heatmap plot of GEI test statistics [-log10(P_GEI)] for the 75 top vQTL SNPs. ** denotes significant GEI effects after Bonferroni correction [P_GEI < 1.33 × 10^-4 = 0.05/(75 × 5)]. (B) Distribution of the number of significant GEI effects for the 75 top QTL SNPs randomly selected from all the top QTL SNPs with 1000 repeats (mean, 2.25; SD, 1.49). The red line represents the number of significant GEI effects for the 75 top vQTL SNPs (i.e., 16).

There are several vQTLs for which the GEI effect has been reported in previous studies. The first example is the interaction effect of the CHRNA5-A3-B4 locus (rs56077333) with smoking for lung function (as measured by FFR, i.e., FEV1/FVC), P_GEI = 1.1 × 10^-14 and P_GEI(smoking) = 4.6 × 10^-23 (table S3A). The CHRNA5-A3-B4 gene
cluster is known to be associated with smoking and nicotine dependence (43–45). However, results from recent GWAS (46–48) do not support the association of this locus with lung function. We hypothesize that the effect of the CHRNA5-A3-B4 locus on lung function depends on smoking (table S3A) (49). The vQTL signal at this locus remained ($P_{vQTL} = 5.2 \times 10^{-17}$) after adjusting the phenotype for array effect, which was reported to affect the QTQ association signal at this locus (26). The second example is the interaction of the WNT16-CPED1 locus with age for bone mineral density (BMD) ([rs10254825: $P_{vQTL} = 2.0 \times 10^{-45}$ and $P_{GEO(age)} = 1.2 \times 10^{-7}$]). The WNT16-CPED1 locus is one of the strongest BMI-associated loci identified from GWAS (50, 51). We observed a genotype-by-age interaction effect at this locus for BMD (table S3B), in line with the results from previous studies that the effect of the top SNP at WNT16-CPED1 on BMD in humans (52) and the knockout effect of Wnt16 on bone mass in mice (53) are age dependent. The third example is the interaction of the FTO locus with PA and SB for obesity-related traits [LD ($P_{vQTL} < 1 \times 10^{-10}$ for BMI, WC, HC, BFP, and BMR; $P_{GEO(PA)} = 1.3 \times 10^{-10}$ for BMI, $1.4 \times 10^{-7}$ for WC, $5.3 \times 10^{-7}$ for HC, and $2.6 \times 10^{-7}$ for BMR)]. The FTO locus was one of the first loci identified by the GWAS of obesity-related traits (54), although subsequent studies (55, 56) show that $IRX3$ and $IRX5$ (rather than $FTO$) are the functional genes responsible for the GWAS association. The top associated SNP at the FTO locus is not associated with PA, but its effect on BMI decreases with the increase of PA level (12, 57), consistent with the interaction effects of the FTO locus with PA or SB for obesity-related traits identified in this study (table S3, C and D). In addition, 5 of the 22 BMI vQTLs were in LD ($r^2 > 0.5$), with the variants (identified by a recently developed multiple-environment GEI test) showing significant interaction effects at a false discovery rate (FDR) of <5% (corresponding to $P < 1.16 \times 10^{-5}$) with at least 1 of 64 environmental factors for BMI in the UKB (58).

It should be noted that GEI is sufficient but not necessary to generate a vQTL. For the vQTLs that did not show a direct GEI effect in our GEI analysis, we cannot distinguish whether they are due to undetected GEI or direct effects on phenotypic dispersion, although GEI is a more likely explanation because of the enrichment of GEI (Fig. 5); hence, these traits and loci are candidates for follow-up studies to identify putative environmental risk factors that may be amendable to lifestyle modification. We also explored two other interpretations of the observed vQTLs, i.e., “phantom vQTLs” (19, 20) and epistasis (genotype-by-genotype interaction) (17). If the underlying causal QTL is not well imputed or not well tagged by a genotyped/imputed variant, then the untagged variation at the causal QTL will inflate the vQTL test statistic, potentially leading to a spurious vQTL association, i.e., the so-called phantom vQTL. We showed by theoretical deviations that the Levene’s test statistic due to the phantom vQTL effect was a function of sample size, effect size of the causal QTL, allele frequency of the causal QTL, allele frequency of the phantom vQTL, and LD between the causal QTL and the phantom vQTL (note S6 and fig. S7A). From our deviations, we computed the numerical distribution of the expected phantom vQTL $F$-statistics, given a number of parameters including the sample size ($n = 350,000$), variance explained by the causal QTL ($q^2 = 0.005, 0.01$, or $0.02$), and MAFs of the causal QTL and the phantom vQTL (MAF = 0.05 to 0.5). The result showed that, for a causal QTL with $q^2 < 0.005$ and MAF > 0.05, the largest possible phantom vQTL $F$-statistic was smaller than 2.69 (corresponding to a $P$ value of $6.8 \times 10^{-2}$; fig. S7, B to D). This explains why there were thousands of genome-wide significant QTLs but no significant vQTL for height (Fig. 3 and Table 1). This result also suggests that the vQTLs detected in this study are very unlikely to be phantom vQTLs because the estimated variance explained by their QTL effects were all smaller than 0.005, except for rs10254825 at the WNT16 locus on BMD ($q^2 = 0.014$) (fig. S7E). However, our numerical calculation also indicated that, for a QTL with MAF > 0.3 and $q^2 < 0.02$, the largest possible phantom vQTL $F$-statistic was smaller than 5.64 (corresponding to a $P$ value of $3.6 \times 10^{-3}$), suggesting that rs10254825 is also unlikely to be a phantom vQTL. Note that we used the variance explained estimated at the top GWAS SNP to approximate $q^2$ of the causal QTL so that $q^2$ was likely to be underestimated because of imperfect tagging.

In conclusion, we systematically quantified the FPR and power for four commonly used vQTL methods by extensive simulations and demonstrated the robustness of the Levene’s test. We also showed that, in the presence of QTQ effects, the Levene’s test statistic could be inflated if the phenotype was nonlinearly transformed. We implemented the Levene’s test as part of the OSCA software package (http://cnsgenomics.com/software/osca) (61) for efficient genome-wide vQTL analysis. We applied OSCA-vQTL to 13 quantitative traits in the UKB and identified 75 vQTL (at 41 near-independent loci) associated with 9 traits, 9 of which did not show a significant QTQ effect. As a proof of principle, we performed GEI analyses in the UKB with five environmental factors/covariates and demonstrated the enrichment of GEI effects among the detected vQTLs. Hence, the vQTL trait-loci combinations we have identified could be investigated for as-yet-undetermined but measurable environmental risk factors generating GEI, as these factors could be amenable to lifestyle change interventions. We further derived the theory to compute the expected "phantom vQTL" $F$-statistic due to untagged causal QTQ effect and showed by numerical calculation that our observed vQTLs were very unlikely to be driven by imperfectly tagged QTQ effects. Our theory is also consistent with the observation of pervasive phantom vQTLs for molecular traits with large-effect QTQs [e.g., DNA methylation (20)]. However, the conclusions from this study may only be applicable to quantitative traits of polygenic architecture. We caution vQTL analysis for binary or categorical traits, or molecular traits (e.g., gene expression or DNA methylation), for which the methods need further investigation.

**METHODS**

**Simulation study**

We used a DGLM (33–35) to simulate the phenotype based on two models with simulated SNP data in a sample of 10,000 individuals,
i.e., a single-SNP model and multiple-SNP model with two covariates (i.e., age and sex). The single-SNP model can be written as

\[ y = w_0 + e \text{ with } \log(\sigma^2_e) = 0 \]

and the multiple-SNP model can be expressed as

\[ y = \sum c_i \beta_j + \sum \phi_k + e \text{ with } \log(\sigma^2_e) \]

where \( y \) is a simulated phenotype; \( w \) or \( w_k \) is a standardized SNP genotype, \( i.e., w = (x - 2f)/\sqrt{2f(1 - f)} \), with \( x \) being the genotype indicator variable coded as 0, 1, or 2, generated from binomial(2, \( f \)) and \( f \) being the MAF generated from uniform(0.01, 0.5); \( c_i \) is a standardized covariate with \( c_i \) (sex) generated from binomial(1, 0.5) and \( c_2 \) (age) generated from uniform(20, 60); \( e \) is an error term with mean 0 and variance \( \sigma^2_e \). To simulate the error term with different levels of skewness and kurtosis, we generated \( e \) from five different distributions, including normal distribution, \( t \)-distribution with \( df = 10 \) or 3, and \( \chi^2 \) distribution with \( df = 15 \) or 1. \( \beta (\phi) \) is the effect on mean (variance) generated from \( N(1,1) \) if exists, 0 otherwise. \log(\sigma^2) is the intercept of the second linear model, which was set to 0. We recast the different components to control the variance explained, i.e., 0.1 and 0.9 for the genotype component and error term, respectively, in the single-SNP model, and 0.2, 0.4, and 0.4 for the covariate component, genotype component, and error term, respectively, in the multiple-SNP model. We simulated the SNP effects in four different scenarios: (i) effect on neither mean nor variance ( nei), (ii) effect on mean only (mean), (iii) effect on variance only (var), or (iv) effect on both mean and variance (both). We simulated only one causal SNP in the single-SNP model and 4, 40, or 80 causal SNPs in the multiple-SNP model.

We performed vQTL analyses using the simulated phenotype and SNP data to compare four vQTL methods, including the Bartlett’s test (29), the Levene’s test (31), the FK test (32), and the DGLM (note S2). We also performed the Levene’s test with six phenotype process strategies, including raw phenotype (raw), raw phenotype adjusted for covariates (adj), RINT after adj (rint) (note S3), logarithm transformation after adj (log), square transformation after adj (sq), and cube transformation after adj (cub). We repeated the simulation 1000 times and calculated the FPR and power at \( P < 0.05 \) at a single-SNP level.

The UKB data

The full release of the UKB data consisted of genotype and phenotype data for ~500,000 participants across the United Kingdom (26). The genotype data were cleaned and imputed to the Haploype Reference Consortium (59) and UK10K (62) reference panels by the UKB team. Genotype probabilities from imputation were converted to hard-call genotypes using PLINK2 (--hard-call 0.1) (63). We excluded genetic variants with MAF < 0.05, Hardy-Weinberg equilibrium test \( P \) value < \( 1 \times 10^{-5} \), missing genotype rate > 0.05, or imputation INFO score < 0.3 and retained 5,554,549 variants for further analysis.

We identified a subset of individuals of European ancestry (\( n = 456,422 \)) by projecting the UKB PCs onto those of the 1000 Genome Project (1KG) (64). We then removed one of each pair of individuals with SNP-derived (based on HapMap 3 SNPs) genomic relatedness >0.05 using GCTA-GRM (65) and retained 348,501 unrelated European individuals for further analysis.

We selected 13 quantitative traits for our analysis (table S1A and fig. S2A). We adjusted the raw phenotype values for age and the first 10 PCs, excluded from the analysis phenotype values that were more than 5 SDs from the mean, and then standardized to \( z \) scores with mean 0 and variance 1 in each gender group.

**Genome-wide vQTL analysis**

The genome-wide vQTL analysis was conducted using the Levene’s test implemented in the software tool OSCA (http://cnggenomics.com/software/oscA) (61). The Levene’s test used in the study [also known as the median-based Levene’s test or the Brown-Forsyth test (31)] is a modified version of the original Levene’s test (30) developed in 1960 that is essentially a one-way ANOVA test of the variable \( z_{ij} = |y_{ij} - \bar{y}_j| \), where \( y_{ij} \) is the phenotype of the \( j \)th individual in the \( i \)th group and \( \bar{y}_j \) is the median of the \( i \)th group. The Levene’s test statistic

\[
\frac{(n - k) \sum_{i=1}^{k} n_i (z_{ij} - \bar{z})^2}{(k - 1) \sum_{i=1}^{k} \sum_{j=1}^{n_i} (z_{ij} - \bar{z})^2}
\]

approximately follows an \( F \) distribution with \( k - 1 \) and \( n - k \) degrees of freedom under the null hypothesis, where \( n \) is the total sample size, \( k \) is the number of groups (\( k = 3 \) in vQTL analysis), \( n_i \) is the sample size of the \( i \)th group, and \( \bar{z} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} z_{ij} \).

The experiment-wise significance level was set to 2.0 \( \times 10^{-6} \), which is the genome-wide significance level (i.e., \( 1 \times 10^{-6} \)) (38, 39) divided by the effective number of independent traits (i.e., 5.00 for our 13 traits). The effective number of independent traits was estimated on the basis of the phenotypic correlation matrix (note S4) (66). To determine the number of near-independent vQTLs, we performed an LD clumping analysis for each trait using PLINK2 (--clump option with parameters --clump-p1 2.0e-9 --clump-p2 2.0e-9 --clump-r2 0.01 and --clump-kb 5000) (63). To visualize the results, we generated the Manhattan and regional association plots using the ggplot2 package in R.

**GWAS analysis**

The GWAS (or genome-wide QTL) analysis was conducted using PLINK2 (63) (--assoc option) using the same data as used in the vQTL analysis (note that the phenotype had been pre-adjusted for covariates and PCs). The other analyses, including LD clumping and visualization, were performed using the same pipelines as those for genome-wide vQTL analysis described above.

**GEl analysis**

Five environmental factors/covariates (i.e., sex, age, PA, SB, and smoking) were used for the direct GEl tests. Sex was coded as 0 or 1 for female or male. Age was an integer number ranging from 40 to 74. PA was assessed by a three-level categorical score (i.e., low, intermediate, and high) based on the short form of the International Physical Activity Questionnaire (IPAQ) guideline (67). SB was an integer number defined as the combined time (hours) spent driving, non–work–related computer using, or TV watching. The smoking factor “ever smoked” was coded as 0 or 1 for never or ever smoker. More details about the definition and derivation of environmental factor PA, SB, and smoking can be found in note S5, fig. S2B, and table S1B.
We performed a GEI analysis to test the interaction effect between the top vQTL SNP and one of the five environmental factors based on the following model

\[ y = \mu + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + e \]

where \( y \) is the phenotype, \( \mu \) is the mean term, \( x_1 \) is the mean-centered SNP genotype indicator, and \( x_2 \) is the mean-centered environmental factor. We used a standard ANOVA analysis to test for \( \beta_{12} \) and applied a stringent Bonferroni-corrected threshold 1.33 \times 10^{-4} \ [i.e., 0.05/(75 \times 5)] to claim a significant GEI effect.

**SUPPLEMENTARY MATERIALS**

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/5/8/eaaw3538/DC1

Note 1. Theoretical derivation of vQTL as a consequence of GEI

Note 2. The Bartlett's test, the FK test, and the DGLM test

Note 3. Rank-based inverse-normal transformation

Note 4. Theoretical derivation of vQTL as a consequence of GEI

Note 5. Definitions of the three environmental factors—PA, SB, and smoking

Note 6. Expected inflation in the Lewen's test statistic due to phantom vQTL effect

Note 7. Acknowledgments

Fig. 1. Evaluation of statistical methods and phenotype processing strategies for the vQTL analysis by simulation based on a single-SNP model.

Fig. 2. Phenotypic correlations among 13 quantitative traits, and PA and SB measures in the UKB.

Fig. 3. Genome-wide vQTL and QTL analyses for 13 traits in the UKB.

Fig. 4. Quantile-quantile plots of vQTL associations for the 13 UKB traits.

Fig. 5. Comparison of the Young et al. method with the Lewen's test by vQTL simulation.

Fig. 6. Excluding two alternative explanations for vQTL signals: Phantom vQTLs and epistasis.

Table S1. Descriptive summary of (A) the quantitative traits and (B) the environmental data used in this study from the UKB.

Table S2. Seventy-five experiment-wise significant vQTLs for nine UKB traits.

Table S3. GEI examples.

References (68-75)

**REFERENCES AND NOTES**

1. D. S. Falconer, T. C. F. Mackay, *Introduction to Quantitative Genetics* (Longman, ed. 4, 1996).

2. M. Lynch, B. Walsh, *Genetics and Analysis of Quantitative Traits* (Sinauer Associates, 1998).

3. A. E. Garrod, The incidence of alkaptonuria: A study in chemical individuality. *Lancet* 160, 1616–1620 (1902).

4. J. Haldane, *Heredity and Politics* (WW Norton & Co., 1938).

5. P. Kraft, D. Hunter, Integrating epidemiology and genetic association: The challenge of gene-environment interaction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1609–1616 (2005).

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

7. H. Ashard, S. Lutz, B. Maus, E. J. Duell, T. E. Fingerlin, N. Chatterjee, P. Kraft, K. van Steen, Challenges and opportunities in genome-wide environmental interaction (GWEI) studies. *Hum. Genet.* 131, 1591–1613 (2012).

8. K. McAllister, L. E. Mechanic, C. Amos, H. Ashard, I. A. Blair, N. Chatterjee, D. Conti, W. J. Gauderman, L. Hsu, C. M. Hutter, M. M. Jankowski, J. Kerr, P. Kraft, S. B. Montgomery, B. Mukherjee, G. J. Papanicolaou, C. J. Patel, D. M. Ritchie, B. R. Ritz, D. C. Thomas, P. Wei, J. S. Witte, Current challenges and new opportunities for gene-environment interaction study of complex diseases. *Am. J. Epidemiol.* 186, 753–761 (2017).

9. J. Yang, T. Lee, J. Kim, M. C. Cho, B. G. Han, J. Y. Lee, H. J. Lee, S. Cho, H. Kim, Ubiquitous polygenicity of human complex traits: Genome-wide analysis of 49 traits in Koreans. *PLOS Gen.* 9, e1003355 (2013).

10. H. Shi, G. Richaev, B. Pasunluc, Contrasting the genetic architecture of 3 complex traits from summary association data. *Am. J. Hum. Genet.* 99, 139–153 (2016).

11. R. M. Maier, P. M. Visscher, M. R. Robinson, N. R. Wray, Embracing polygenicity: A review of methods and tools for psychiatric genetics research. *Psychol. Med.* 48, 1055–1067 (2018).

12. T. O. Kilpeläinen, L. Qi, S. Brage, S. J. Sharp, E. Sonestedt, E. Derenstew, T. Ahmad, S. Mora, M. Kaaks, C. H. Sandholz, C. Holzapfel, C. S. Autenrieth, E. Hyypönen, S. Cauchi, M. He, Z. Kutalik, M. Kumari, A. Stančáková, K. Meidtmann, B. Balkau, J. T. Tan, M. Manginno, N. J. Timpson, Y. Song, M. C. Zillikens, K. A. Jablonski, M. E. Garcia, S. Johansson, J. L. Bragg-Gresham, Y. Wu, J. V. van Vliet-Ostapchuk, N. C. Onland-Moret, E. Zimmermann, N. V. Rivera, T. Tanaka, H. M. Stringham, G. Silber Nagel, S. Kanoni, M. F. Feltsos, S. Snitker, J. R. Ruiz, J. Metter, M. T. M. Lamad, M. Atalay, M. Hakanen, N. Amin, C. Cavalcanti-Proenca, A. Grøntved, G. Halfmans, J. O. Jansson, J. Kuusisto, M. Kähönen, P. L. Lubset, J. J. Nolan, L. Paala, O. Pedersen, L. Pérusse, F. Renström, R. A. Scott, D. Shungin, U. Sovio, T. H. Tammelin, T. Rönemaa, T. A. Lakka, M. Uusitupa, M. S. Riiss, F. Li, C. Bjornard, A. Meuhaeghe, M. Fu, M. Walker, I. B. Boeckli, G. V. Dedoussis, A. Fritsche, C. Ohlsson, M. Boehnke, S. Bandinelli, C. M. van Duijn, S. Ebrahim, D. A. Lawlor, V. Gudnason, T. B. Harris, T. I. A. Sørensen, K. L. Mohlke, A. Hofman, A. G. Uitterlinden, T. Juoniluoto, T. Lehtimäki, O. Raitakari, B. Isomaa, P. R. Njølstad, J. F. Loree, S. Liu, A. Ness, T. D. Spector, E. S. Tai, P. Frogleg, H. Boeing, M. Laakso, M. Marmot, S. Bergmann, C. Power, K. T. Khaw, D. Chasman, P. Rickler, T. Hansen, K. L. Monda, T. Ilbag, M. R. Jänväril, N. J. Wareham, F. B. Hu, L. C. Groop, M. Ørha-Melander, U. Ekfeld, P. W. Franks, R. F. J. Loos, Physical activity attenuates the influence of TTO variants on obesity risk: A meta-analysis of 218,166 adults and 19,268 children. *PLOS Med.* 8, e1001116 (2011).
Wang et al., Sci. Adv. 2019;5:eaaq3538
14 August 2019
12 of 12
