Distinct explanations underlie gene-environment interactions in the UK Biobank

Arun Durvasula[1,2,3,4,5] and Alkes L. Price[4,5,6]

1. Center for Genetic Epidemiology, Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
2. Department of Genetics, Harvard Medical School, Cambridge, MA, USA
3. Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA
4. Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA
5. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
6. Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

Correspondence: arun.durvasula@med.usc.edu or aprice@hsph.harvard.edu

Abstract

The role of gene-environment (GxE) interaction in disease and complex trait architectures is widely hypothesized, but currently unknown. Here, we apply three statistical approaches to quantify and distinguish three different types of GxE interaction for a given disease/trait and E variable. First, we detect locus-specific GxE interaction by testing for genetic correlation (r_g) < 1 across E bins. Second, we detect genome-wide effects of the E variable on genetic variance by leveraging polygenic risk scores (PRS) to test for significant PRSxE in a regression of phenotypes on PRS, E, and PRSxE, together with differences in SNP-heritability across E bins. Third, we detect genome-wide proportional amplification of genetic and environmental effects as a function of the E variable by testing for significant PRSxE with no differences in SNP-heritability across E bins. Simulations show that these approaches achieve high sensitivity and specificity in distinguishing these three GxE scenarios. We applied our framework to 33 UK Biobank diseases/traits (average N=325K) and 10 E variables spanning lifestyle, diet, and other environmental exposures. First, we identified 19 trait-E pairs with r_g significantly < 1 (FDR<5%) (average r_g=0.95); for example, white blood cell count had r_g=0.95 (s.e. 0.01) between smokers and non-smokers. Second, we identified 28 trait-E pairs with significant PRSxE and significant SNP-heritability differences across E bins; for example, type 2 diabetes had a significant PRSxE for alcohol consumption (P=1e-13) with 4.2x larger SNP-heritability in the largest versus smallest quintiles of alcohol consumption (P<1e-16). Third, we identified 15 trait-E pairs with significant PRSxE with no SNP-heritability differences across E bins; for example, triglyceride levels had a significant PRSxE effect for composite diet score (P=4e-5) with no SNP-heritability differences. Analyses using biological sex as the E variable produced additional significant findings in each of the three scenarios. Overall, we infer a substantial contribution of GxE and GxSex effects to disease and complex trait variance.

Introduction

Although gene-environment (GxE) interactions have long been thought to impact the genetic architecture of diseases and complex traits[1–4], the overall contribution of these effects remains unclear. Previous studies have detected GxE at a limited number of specific loci[5–7] (including studies that associated genotype to phenotypic variance without knowing the underlying E variable[8–12]). Previous studies have also proposed variance components methods for detecting genome-wide contributions of GxE to disease heritability[13–18], but these methods have not been applied at biobank scale across a broad range of disease/traits. Thus, the overall contribution of GxE to disease/trait architectures is currently unknown. In addition, the relative importance of different types of GxE (e.g., locus-specific GxE, genome-wide effects of E on genetic variance, genome-wide effects of E on both genetic and environmental variance) is currently unclear. Studies of GxSex interaction face similar challenges[19–24].

Here, we apply three statistical approaches to quantify and distinguish three different types of GxE interaction for a given disease/trait and E variable. First, we detect locus-specific GxE interaction by testing for genetic correlation[25] (r_g) < 1 across E bins. Second, we detect genome-wide effects of the E variable on genetic variance by leveraging polygenic risk scores[26,27] (PRS) to test for significant PRSxE[28,29] in a regression of phenotypes on PRS, E, and PRSxE, together with differences in SNP-heritability[30–34] across E bins. Third, we detect genome-wide proportional amplification of genetic and environmental effects as a function of the E variable by testing for significant PRSxE with no differences in SNP-heritability across E bins. We analyze 33 diseases/traits from the UK Biobank[35] (average N=325K), quantifying the contributions of each type of GxE effect across 10 E variables spanning lifestyle, diet, and other environmental exposures, as well as contributions of GxSex effects.

Results

Overview of methods

We aim to detect genome-wide GxE, i.e., GxE effects aggregated across the genome. We consider three potential scenarios that give rise to genome-wide GxE for a given disease/trait and E variable (**Figure 1a**). In the first scenario (Imperfect genetic correlation), there is an imperfect genetic correlation across E bins due to different SNP effect sizes in different E bins. In the second scenario (Varying genetic variance), there are differences in SNP-heritability across E bins due to uniform amplification of SNP effect sizes across E bins; the environmental variance may either remain constant or vary across E bins. In the third scenario (proportional amplification), the genetic and environmental variance vary proportionately across E bins due to proportionate scaling of SNP effect sizes and environmental effect sizes across E bins, so that SNP-heritability remains the same across E bins. We conceptualize these three scenarios as acting at different levels in a hierarchy that leads from genetic variants to pathways to disease (see **Discussion**).

The three scenarios can be formalized under the following model:

$$y_j = \sum_i x_{ij}\beta_i + \sum_i \gamma_i x_{ij} E_j + \sum_i \xi x_{ij}\beta_i E_j + \varepsilon_j + \eta\varepsilon_j E_j \,, \#(1)$$

where y_j denotes the phenotype for individual j, x_{ij} denotes the genotype of individual j at SNP i, β_i denotes the effect size of SNP i, γ_i denotes SNP-specific GxE effects, E_j denotes the E variable value for individual j, ξ quantifies the amplification of genetic effects across E values, ε_j denotes environmental effects, and η quantifies the amplification of environmental effects across E values. In Scenario 1, γ_i will be nonzero. In Scenario 2, ξ will be nonzero. In Scenario 3, ξ and η will be nonzero and equal.

In this study, we apply three statistical approaches to UK Biobank data to detect genome-wide GxE, analyzing 33 diseases/traits (average N=325K) and 10 environmental variables as well as biological sex (**Figure 1b**). First, we detect Imperfect genetic correlation (Scenario 1) by estimating the genetic correlation of effect sizes between sets of individuals binned on their E variables using cross-trait LD Score regression[25] (LDSC) (**Methods**). For non-binary E variables, we estimate the genetic correlation between the most extreme quintiles of the E variable; for binary E variables, we estimate the genetic correlation between individuals in each E bin. Second, we employ PRSxE regression[28,29], defined as a regression of the phenotype on the PRS[26,27] multiplied by the E variable across individuals, to detect both Varying genetic variance (Scenario 2) and Proportional amplification (Scenario 3) (**Methods**); we note that PRSxE regression is not sensitive to changes in environmental variance only (**Methods**). We use PRS computed by PolyFun-pred[27] for all analyses involving PRS. We do not standardize the E variables, and we correct for main and interaction effects of several covariates (**Methods**). Finally, we distinguish between Scenario 2 and Scenario 3 by estimating the SNP-heritability within each E bin using BOLT-REML[33] and testing for significant differences between E bins (most extreme quintiles for non-binary E variables; each bin for binary E variables).

We assign a trait-E pair to Scenario 1 if it has a genetic correlation across E bins < 1 (regardless of whether it differs in SNP-heritability or has a significant PRSxE regression term); we assign a trait-E pair to Scenario 2 if it has both a significant PRSxE regression term and a significant difference in SNP-heritability across E bins; finally, we assign a trait-E pair to Scenario 3 if it has a significant PRSxE regression term with no significant difference in SNP-heritability across E bins (**Figure 1c**). We note that for some trait-E pairs, we detected both locus-dependent GxE (Scenario 1) and non-locus-dependent GxE (Scenario 2 or Scenario 3). We estimate the excess disease/trait variance explained by genome-wide GxE as follows. In Scenario 1, we transform the estimate of genetic correlation across E bins to the variance scale (**Methods**; **Supplementary Note**). In Scenario 2 and Scenario 3, we approximate the relative amount of trait variance explained by GxE effects (relative to the genetic variance) as the trait variance explained by PRSxE effects divided by the trait variance explained by the PRS; this approximation is valid under a model in which the PRSxE effects are proportional to the GxE effects (**Methods**). All reported variances are transformed to the liability scale for disease traits. We have released open-source software implementing the above approaches (see Code Availability), as well as their output from this study (see Data Availability).

Simulations

We performed simulations of the three Scenarios to evaluate the properties of the three statistical approaches. We assigned individuals to one of two E bins and simulated genetic effects at 10,000 causal SNPs based on the Scenario and E bin. We simulated sample sizes specific to each statistical approach to match our real data analyses (see below). In Scenario 1, we set the SNP-heritability to 25% and varied the genetic correlation from 99% to 94%. In Scenario 2, we set the genetic correlation to 100%, set the SNP-heritability to 25% in one E bin, and varied the SNP-heritability from 26% to 30% in the other E bin. In Scenario 3, we amplified the (genetic and environmental components of) phenotypes in one E bin by a range of values from 1.025 to 1.1. In each Scenario, we report the proportion of significant tests ($P<0.05$, which is fairly similar to our significance threshold for real traits; see below) for each of our three approaches: Genetic correlation (N=67K individuals per E bin), PRSxE regression (training N=337K, testing N=49K), and SNP-heritability by E (N=67K individuals per E bin). Because linkage disequilibrium (LD) does not impact GxE effects, we simulated genotypes without LD. We adjusted the methods used in our simulations accordingly. For Genetic correlation, we used cross-trait LD score regression in the special case of no LD[25]. For PRSxE regression, we used a simple shrinkage estimator in the special case of no LD to compute PRS. For SNP-heritability by E, we estimated SNP-heritability using LD score regression in the special case of no LD[36]. Further details of the simulation framework are provided in the **Methods** section.

In Scenario 1, the Genetic correlation approach reported a significant test in 93% of simulations when the true genetic correlation was 97% or smaller, whereas the PRSxE regression and SNP-heritability by E approaches were well-calibrated (**Figure 2a** and **Supplementary Table 1**). In Scenario 2, the PRSxE regression approach reported a significant test in 88% of simulations when the SNP-heritability difference was 4% or larger, and the SNP-heritability by E approach reported a significant test in more than 88% of simulations when the SNP-heritability difference was 2% or larger, whereas the Genetic correlation approach was well-calibrated (**Figure 2b** and **Supplementary Table 1**). In Scenario 3, the PRSxE regression approach reported a significant test in 88% of simulations when the proportional amplification was 1.075 or larger, whereas the Genetic correlation and SNP-heritability by E approaches were well-calibrated (**Figure 2c** and **Supplementary Table 1**). In null simulations (heritable trait with no GxE), all three statistical approaches were well-calibrated (**Supplementary Figure 1**).

We compared our framework with GxEMM[16], a variance components-based framework that implements two GxE tests: 1) a test for polygenic GxE under homoskedasticity (GxEMM-Hom), and 2) a test for polygenic GxE under heteroskedasticity (GxEMM-Het). We note that GxEMM-Hom and GxEMM-Het do not precisely map to the 3 scenarios that we study here. In addition, because GxEMM is a variance components-based framework, it is currently unable to scale to biobank-sized datasets. We evaluated the performance of GxEMM on a sample size of 10,000 individuals, as in the simulations of ref. [16]. We evaluated our statistical approaches using matched sample sizes, with 5,000 individuals per binary E bin and 10,000 test individuals for PRSxE regression. We kept the training data set size the same as in our main simulations (N=337K). In Scenario 1, the GxEMM-Hom test reported a similar proportion of significant tests as the Genetic correlation approach, whereas the GxEMM-Het test reported roughly half as many significant tests (**Supplementary Figure 2**). In Scenario 2, the GxEMM-Het test was less powerful than the PRSxE regression and SNP-heritability by E approaches, whereas the GxEMM-Hom test was well-calibrated (**Supplementary Figure 2**). In Scenario 3, the GxEMM-

Het test was less powerful than the PRSxE regression approach, whereas the GxEMM-Hom test was well-calibrated (**Supplementary Figure 2**). Thus, at sample sizes that permit computational tractability, GxEMM is generally less powerful than our framework (and cannot distinguish Scenario 2 and Scenario 3).

Our framework also estimates the excess disease/trait variance explained by GxE effects, beyond what is explained by additive effects (for brevity, we refer to this as variance explained). We determined that estimates of disease/trait variance explained were accurate in each of Scenario 1 (regression slope = 0.98; **Supplementary Figure 3a**), Scenario 2 (regression slope = 0.85; **Supplementary Figure 3b**), and Scenario 3 (regression slope = 1.05; **Supplementary Figure 3c**). We note that in both Scenario 2 and Scenario 3, G effects are correlated with GxE effects, as a correlation between genetic variance (G^2) and the E variable implies a correlation between G and GxE. Current variance components methods do not account for this correlation and may therefore produce biased estimates of variance explained by GxE; we have verified this in simulations (**Supplementary Table 2**). Here, we report the difference in variance explained by a model including an interaction term (PRS+PRSxE terms) over a base model (PRS only) that does not include an interaction term, which is robust to this correlation (**Methods** and **Supplementary Figure 3**).

In summary, our simulations indicate that our statistical approaches attain high sensitivity and specificity in classifying trait-E pairs into the distinct scenarios of GxE considered here and produce accurate estimates of excess trait variance explained by GxE.

Identifying gene-environment interactions across 33 diseases/complex traits and 10 E variables

We analyzed individual-level data for *N*=384K unrelated European-ancestry individuals from the UK Biobank[35]. We selected 33 highly heritable (z-score for nonzero SNP-heritability[36] > 6) and relatively independent (squared genetic correlation[25] < 0.5) diseases and traits (**Supplementary Table 3**). In addition, we selected 10 relatively independent E variables spanning lifestyle, diet, and other environmental exposures (r^2 < 0.1; primarily from ref. [14]; **Supplementary Figure 4**; see **Methods**). We note that these E variables are all significantly heritable, although the heritability tends to be low (mean SNP-heritability = 6%, max SNP-heritability = 15%; **Supplementary Table 4**). We assessed statistical significance using a threshold of FDR<5% across traits and E variables for a given statistical test (see **Methods**); in practice, this FDR threshold corresponded to a P-value threshold of ≈0.01, which is fairly similar to our simulations.

Trait-E pairs assigned to Scenario 1 (Imperfect genetic correlation) are reported in **Figure 3a** and **Supplementary Table 5**. We identified 19 trait-E pairs with genetic correlation significantly less than 1 (FDR<5%; average genetic correlation: 0.95), implicating 12 of 33 traits and 9 of 10 E variables tested. The implicated traits included 9 blood cell and biochemistry traits, as well as height, BMI, and asthma. On average, these interactions explained 0.27% of disease/trait variance across all traits analyzed. The lowest significant genetic correlation was 0.85 (se=0.06) for asthma x time spent watching television, explaining 1.5% of trait variance. The significant GxE interaction for BMI and smoking status (explaining 0.4% of trait variance) was consistent with results from ref. [14]. Trait-E pairs assigned to Scenario 2 (Varying genetic

variance) are reported in **Figure 3b** and **Supplementary Table 5**. We identified 28 trait-E pairs with significant PRSxE interaction (FDR<5%) and a significant SNP-heritability by E test (FDR<5%), implicating 13 of 33 traits and 9 of 10 E variables tested. On average, these interactions explained 2.6% of disease/trait variance across all traits analyzed; the variance explained by GxE effects was larger for binary traits than for quantitative traits (see **Discussion**). Because standard interaction tests can be anti-conservative due to unmodeled heteroskedasticity[37], we repeated our PRSxE interaction analysis using Huber-White variance estimators[38,39] (**Methods**). We determined that results were highly concordant with our primary PRSxE interaction analysis (mean Pearson correlation in p-values for interaction across trait-E pairs: 97%; **Supplementary Table 6**), suggesting that our findings are not driven by unmodeled heteroskedasticity. Trait-E pairs assigned to Scenario 3 (Proportional amplification) are reported in **Figure 3c** and **Supplementary Table 5**. We identified 15 trait-E pairs with significant PRSxE interaction (FDR<5%) but a non-significant SNP-heritability by E test (FDR<5%), implicating 11 of 33 traits and 9 of 10 E variables tested. On average, these interactions explained 0.13% of disease/trait variance across all traits analyzed.

We checked whether any trait-E pairs were assigned to more than one Scenario. We determined that 2 trait-E pairs were assigned to both Scenario 1 and Scenario 2 (BMI x alcohol consumption and BMI x Townsend deprivation index); 0 trait-E pairs were assigned to both Scenario 1 and Scenario 3; and 0 trait-E pairs were assigned to both Scenario 2 and Scenario 3 (which is not possible based on their definition). We also identified 108 trait-E pairs with a significant SNP-heritability by E test but non-significant PRSxE interaction (**Supplementary Table 7**); our primary interpretation is that this is due to changes in environmental variance rather than GxE interaction (**Methods**), but we cannot exclude the possibility that this is due to GxE interaction that we have incomplete power to detect.

Examples of trait-E pairs assigned to each scenario are reported in **Figure 4** and **Supplementary Table 8**. First, white blood cell count x smoking status was assigned to Scenario 1 (**Figure 4a**). The Genetic correlation approach estimated a genetic correlation between smokers and non-smokers of 0.95, which is significantly less than 1 (P=6.7e-7; FDR < 5%), explaining 0.5% of the variance of white blood cell count (vs. SNP-heritability of 30%). On the other hand, the PRSxE regression approach (P=0.46) and SNP-heritability x E approach (P=0.39) produced non-significant results. We note that smokers had 0.09 s.d. higher mean white blood cell count than non-smokers (T-test P<1e-16), as previously reported[40]. Second, type 2 diabetes x alcohol consumption was assigned to Scenario 2 (**Figure 4b**). The PRSxE regression approach (P=1e-13) and SNP-heritability x E approach (SNP-heritability (liability scale) of 0.45 for highest E quintile vs. 0.38 for lowest E quintile; P<1e-16) both produced significant results (FDR < 5%), explaining 18% of the variance of type 2 diabetes (vs. SNP-heritability of 35%). On the other hand, the genetic correlation approach produced a non-significant result (P=0.15). We note that the prevalence of type 2 diabetes varied with alcohol consumption (6% in highest E quintile vs. 3% in lowest E quintile; P<2e-16), as previously reported[41]. Third, WHRadjBMI x time spent watching television (TV time) was assigned to Scenario 3 (**Figure 4c**). The PRSxE regression approach produced a significant result (P=5e-3; FDR < 5%), explaining 0.95% of the variance of WHRadjBMI. On the other hand, the genetic correlation approach (P=0.29) and SNP-heritability x E approach (P=0.08) produced non-significant results. We note that WHRadjBMI and TV time were correlated (r = 0.08, P<1e-16).

In summary, we detected GxE interaction in each of the three scenarios across the 33 diseases/traits and 10 E variables analyzed. We estimate that these GxE effects explain 3.0% of disease/trait variance across all traits analyzed (s.e. 1.5% across traits), compared to SNP-heritability of 29% (s.e. 3% across traits).

Identifying gene-sex interactions across 33 diseases/complex traits

We analyzed the same 33 diseases/traits for GxSex interaction using the same 3 statistical approaches. Traits assigned to Scenario 1 (Imperfect genetic correlation) are reported in **Figure 5a** and **Supplementary Table 9**. We identified 22 traits with cross-sex genetic correlation significantly less than 1 (FDR<5%; average genetic correlation: 0.92), consistent with previous results[22]. On average, these interactions explained 2.6% of trait variance across all traits analyzed. The lowest significant genetic correlation was 0.66 for WHRadjBMI[22], explaining 17% of trait variance. Traits assigned to Scenario 2 (Varying genetic variance) are reported in **Figure 5b** and **Supplementary Table 9**. We identified 12 traits with significant PRSxSex interaction (FDR<5%) and a significant SNP-heritability by Sex test (FDR<5%). On average, these interactions explained 1.4% of trait variance across all traits; the variance explained by GxSex effects was larger for binary traits than for quantitative traits (see **Discussion**). The largest PRSxSex interaction was for type 2 diabetes, explaining 0.39% of trait variance. Traits assigned to Scenario 3 (Proportional amplification) are reported in **Figure 5c** and **Supplementary Table 9**. We identified 8 traits with significant PRSxSex interaction (FDR<5%) but a non-significant SNP-heritability by Sex test (FDR<5%). On average, these interactions explained 0.05% of trait variance across all traits analyzed (a very small contribution). Of the 30 traits implicated across three scenarios, we identified 7 traits assigned to both Scenario 1 and Scenario 2, and 5 traits assigned to both Scenario 1 and Scenario 3 (**Supplementary Table 9**). We also identified 2 traits with a significant SNP-heritability x Sex test but non-significant PRSxSex interaction (**Supplementary Table 10**); our primary interpretation is that this is due to changes in environmental variance rather than GxSex interaction (**Methods**).

Examples of traits with significant GxSex assigned to each Scenario are reported in **Figure 6** and **Supplementary Table 11**. First, neuroticism was assigned to Scenario 1 (**Figure 6a**). The Genetic correlation approach estimated a cross-sex genetic correlation of 0.90, which is significantly less than 1 (P=3.5e-9; FDR < 5%), explaining 5.0% of the variance of neuroticism. On the other hand, the PRSxSex regression approach (P=0.58) and SNP-heritability by Sex approach (P=0.45) produced non-significant results. We note that males had lower prevalence of neuroticism than females (1.6% vs. 2.3% in top score for neuroticism, P<1e-16), as previously reported[42]. Second, All autoimmune disease was assigned to Scenario 2 (**Figure 6b**). The PRSxSex regression approach (P=2e-15) and SNP-heritability by Sex approach (SNP-heritability (liability scale) of 23% for males and 18% for females; P<1e-16) both produced significant results (FDR < 5%), explaining 4.9% of the variance of All autoimmune disease (vs. SNP-heritability of 19%). On the other hand, the genetic correlation approach produced a non-significant result (P=0.03). We note that males had lower prevalence of All autoimmune disease than females (7.2% vs. 16%, P<1e-16), as previously reported[43]. Third, HDL cholesterol was assigned to Scenario 3 (**Figure 6c**). The PRSxSex regression approach produced a significant result (P<2e-16; FDR < 5%), explaining 0.4% of the variance of HDL cholesterol. On the other

hand, the SNP-heritability by Sex approach (P=0.09) was not significant. However, the genetic correlation approach estimated a cross-sex genetic correlation of 0.93, which is significantly less than 1 (P=5e-6; FDR < 5%), explaining 3.5% of the variance of HDL cholesterol (Scenario 1); this implies that multiple types of GxSex interaction impact HDL cholesterol. We note that males had 0.83 s.d. lower HDL cholesterol than females (P<1e-16).

In summary, we detected GxSex interaction in each of the three scenarios across the 33 diseases/traits analyzed. We estimate that these GxSex effects explain 4.0% of disease/trait variance across all traits analyzed (s.e. 1.1% across traits), compared to SNP-heritability of 29% (s.e. 3%).

Discussion

We have applied three statistical approaches to detect, quantify, and distinguish the genome-wide contributions of three different types of GxE interaction (**Figure 1a**) across 33 UK Biobank diseases/traits, analyzing 10 E variables spanning lifestyle, diet, and other environmental exposures as well as biological sex. We determined that GxE interactions (involving these E variables) and GxSex interactions each explained a significant fraction of phenotypic variance, representing an appreciable contribution to disease/trait architectures. It is possible that GxE interactions involving E variables not studied here could explain even more phenotypic variance.

Our finding of distinct explanations underlying GxE interactions (**Figure 1a**) motivates a unified model consistent with this finding. We propose a model in which GxE occurs at different levels of a hierarchy that leads from genetic variants to pathways to disease (**Supplementary Figure 5**). In this model, Scenario 1 (Imperfect genetic correlation) occurs when an E variable modifies the effects of individual variants (or sets of variants), differentially impacting different parts of the genome; Scenario 2 (Varying genetic variance) occurs when an E variable modifies all of the pathways underlying genetic risk, uniformly impacting genetic variance; and Scenario 3 (Proportional amplification) occurs when an E variable modifies all aspects of disease biology, proportionately impacting both genetic and environmental variance. Under this model, an E variable can modify any point along the hierarchy from genetic variants to pathways to disease. Further investigation and validation of this model is a direction for future research.

Our study represents an advance over previous studies investigating genome-wide GxE. First, we distinguish three different types of GxE interaction: Imperfect genetic correlation, Varying genetic variance, and Proportional amplification (**Figure 1a**; also see **Supplementary Figure 5**). Second, most variance components methods for detecting genome-wide GxE[13–15,17,18] cannot detect genome-wide GxE unless SNP-heritability varies across E bins (Scenario 2). An exception is GxEMM[16], which detects other types of GxE by explicitly modeling genetic and environmental variance that varies with the E variable; however, GxEMM is less computationally tractable and generally less powerful than our framework (**Supplementary Figure 2**). Third, variance components methods that assume independence between G and GxE effects are susceptible to bias if G and GxE effects are correlated, but our statistical approaches are robust to this possibility (**Supplementary Figure 3**. Fourth, previous methods have not been applied at biobank scale across a broad range of disease/traits; the statistical approaches that we propose are computationally scalable to very large data sets (see **Methods**), enabling our

biobank-scale analyses implicating 60 trait-E pairs with significant GxE and 30 traits with significant GxSex. Fifth, a recent study reported that GxSex acts primarily through amplification[24] (Scenario 2 and Scenario 3), but our analyses of GxSex determined that Imperfect genetic correlation (Scenario 1) explained a larger proportion of trait variance than amplification; in addition, ref. [24] did not estimate contributions to trait variance and did not distinguish between Scenario 2 and Scenario 3, as we do here.

Our study has several implications. First, our results narrow the search space of disease/traits and E variables for which genome-wide association studies of GxE interactions are most likely to be fruitful; in particular, trait-E pairs with substantial trait variance explained by Scenario 1 (Imperfect genetic correlation) (**Supplementary Table 5**) should be prioritized for locus-specific analyses, in preference to trait-E pairs with trait variance explained by Scenario 2 or Scenario 3. Second, our results imply that there is broad potential to improve polygenic risk scores (PRS) by leveraging information on E variables in training and/or test samples[44]. Third, there is broad potential to prioritize individuals for which a lifestyle intervention to modify an E variable would be most effective based on their genetic profile. Fourth, previous work has suggested that population-specific causal effect sizes in functionally important regions may be caused by GxE[45], motivating efforts to partition the imperfect genetic correlations across E bins that we have identified across functionally important regions. Fifth, the significant contribution of GxE to disease/trait architectures—even when restricting to the limited set of E variables that we analyzed here—implicates GxE effects as a factor in "missing heritability", defined as the gap between estimates of SNP-heritability[30] and estimates of narrow-sense heritability[46] (e.g. from twin studies[47]); although GxE effects are not included in the *definition* of narrow-sense heritability, they can inflate twin-based *estimates* of narrow-sense heritability, analogous to GxG effects[48]. All of these implications motivate directions for future research.

Our study has several limitations. First, our analyses assess GxE and GxSex interaction for binary traits on the observed scale (and then transform estimates to the liability scale), consistent with prevailing approaches for variance component analysis of binary traits[31–34]. This approach leads to much larger variance explained by GxE and GxSex for binary traits vs. quantitative traits in Scenario 2 (**Supplementary Table 12;** also see **Figure 3b** and **Figure 5b**). These statistical interactions may not be indicative of biological interactions, as previously noted in the context of both GxE interaction[2] and GxG interaction[49]. Directly modeling GxE interaction on the liability scale[50,51] is an important direction for future research, and may produce different findings. Second, the E variables that we analyzed comprise an extremely limited subset of the set of E variables that may contribute to GxE effects (and their values may be subject to measurement error); even when GxE effects are detected, the implicated E variable may be tagging an unmeasured causal E variable with larger GxE effects. Third, our use of PRSxE regression to detect GxE is limited by the accuracy of PRS and may require larger training sample sizes (enabling more accurate PRS) to be well-powered, particularly for less heritable diseases/traits. The average accuracy of the PRS across traits in the held-out set of 49K individuals was 9.2%, as measured by r^2 between predicted and true phenotypes (**Supplementary Table 3**). Fourth, our estimates of the trait variance explained by GxE effects detected via PRSxE analyses assume that PRSxE effects extrapolate linearly to GxE effects (**Methods**); we believe that this is a reasonable assumption, but we cannot formally exclude the possibility that genetic effects captured by PRS interact differently with an E variable than genetic effects not captured by PRS. Fifth, most of the

E variables that we study are weakly heritable (**Supplementary Table 4**), raising the possibility of GxG (rather than GxE) effects; we consider GxG to be an unlikely explanation given the E variables' low SNP-heritabilities, but we cannot formally exclude this possibility. Sixth, our use of PRSxE regression to detect GxE may be anti-conservative due to unmodeled heteroskedasticity[37]; however, we obtained nearly identical results using Huber-White variance estimators (also known as robust regression[38,39]) (**Supplementary Table 6**), suggesting that this does not impact our findings. We note that we observe many instances of differences in trait variance across E variables (**Supplementary Table 7**), but these alone are not indicative of GxE interactions. Seventh, our use of PRSxE regression to detect GxE may produce false positives if there is a nonlinear relationship between E and trait value; we included an E^2 term in PRSxE regressions to ameliorate this possibility but determined that inclusion or exclusion of the E^2 term had little impact on our results (**Supplementary Table 13**), suggesting that nonlinear effects do not greatly impact our findings. Eighth, we have analyzed British-ancestry samples from the UK Biobank, but an important future direction is to extend our analyses to cohorts of diverse genetic ancestry[52,53], which may differ in their distributions of E variables, tagging of causal E variables by measured E variables, and/or causal GxE effects (analogous to differences in main G effects[45,54]). Eighth, we do not analyze GxAge interaction (and we note the limited age variation in UK Biobank samples; age = 55 $\pm$ 8 years), but we highlight GxAge interaction and longitudinal data as important directions for future research[51,55,56]. Despite these limitations, our work quantifies and distinguishes three different types of GxE interaction across a broad set of diseases/traits and E variables.

Code Availability

Cross trait LDSC: https://github.com/bulik/ldsc
BOLT-LMM: https://alkesgroup.broadinstitute.org/BOLT-LMM/downloads/
PRSxE regression: Will be added upon publication.
Code to reproduce analysis: Will be added upon publication.

Data Availability

We will make the results of the three statistical approaches we use here publicly available upon publication.

Acknowledgements

We are grateful to Martin Zhang, Ben Strober, Xilin Jiang, and Jordan Rossen for helpful discussions and Sriram Sankararaman and Ali Pazokitoroudi for comments on an earlier version of this manuscript. This research was conducted using the UK Biobank resource under application no. 16549 and funded by National Institutes of Health (NIH) grants R01 MH101244, R37 MH107649 and R01 HG006399. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Methods

Data sources and preprocessing

We used data from the UK Biobank in all our analyses. For polygenic score-based analyses that required a training and testing dataset, we used a set of 337K unrelated white British individuals for training[27]. For testing, we used a set of 49K European individuals who are unrelated to each other and to the training cohort[27]. Note that while "testing" typically refers to a setting where the ultimate goal is to assess PRS accuracy, here we use it to refer to the set of samples we in which we run a regression of phenotype on PRSxE and covariates. We used polygenic scores generated by ref. [27]. We used the linear scoring function in Plink v1.9[57] to compute polygenic scores in the set of 49K test individuals.

Choice of diseases/traits and environmental variables

We chose a set of 33 diseases/traits with SNP heritability Z scores > 6 and squared genetic correlation less than 0.5 (**Supplementary Table 3, Supplementary Table 4**). We chose a set of 10 E variables, including 5 previously analyzed E variables from ref. [14] and 5 additional E variables (Air pollution, time spent napping, sleeplessness, Diet, wheat consumption) (**Supplementary Figure 4)**.

To compute the Diet variable, we performed PCA on a covariance matrix consisting of several diet variables: cooked vegetable intake, salad intake, fresh fruit intake, processed meat intake, poultry intake, beef intake, pork intake, coffee intake (**Supplementary Figure 6**). We used the function prcomp in R and extracted the first PC.

Genetic correlation approach to detecting GxE

We performed GWAS using BOLT-LMM[26] within bins of E variables. Then, we used bivariate LD Score regression[36] to estimate the genetic correlation between the top and bottom quintiles of E variables; for binary E variables, we estimated the genetic correlation between individuals in each E bin. We used imputed SNPs with MAF > 0.01% and used the --no-intercept option to increase our power. Computed a Z score testing against the null hypothesis that the genetic correlation is 1 as:

$$Z = \frac{1 - \widehat{r_g}}{\widehat{se}}$$

PRSxE regression approach to detecting GxE

Our PRSxE regression takes the following form:

$$Y = PRS + E + PRS * E + C\#$$

where Y is the trait value, PRS is the polygenic score for the trait (see *Data sources and preprocessing*), E is the environment variable, and C is a set of covariates. For all analyses, we correct for the following covariates: age, sex, 10 genetic PCs computed in the held-out set, the squared E variable: E^2 , age*sex, E*age, E*sex. We carried out this regression using the Python package statsmodels v0.14. We also compute a 'base' model, which is the same regression

without the PRSxE term. We use the p-value associated with the PRSxE term in the interaction model to assess significance.

To test whether our results were driven by heteroskedasticity, we performed the same analysis using robust standard errors as implemented in statsmodels using the 'H1' covariance matrix (**Supplementary Table 12**).

We note the PRSxE regression test is not expected to produce a significant finding if the environmental variance changes as a function of E but the genetic variance does not change as a function of E, because the PRS does not measure changes in environmental variance.

SNP-heritability by E approach to detecting GxE

We used BOLT-REML[33] v2.3.6 to compute heritability in bins of E variables.
To test for a significant difference in heritability between two bins, we computed a Z score as:

$$Z = \frac{h_1^2 - h_2^2}{\sqrt{\sigma_1^2 + \sigma_2^2}}$$

where 1 and 2 index the E bins.

False Discovery Rate (FDR) control

We chose a 5% FDR control separately for each statistical approach (Genetic correlation, PRSxE regression, and SNP-heritability by E) using the qvalue R package[58]. We ensured our one-sided test against a null genetic correlation of 1 did not produce a skewed P-value distribution, which could indicate improper choice of a one-sided test. We chose to control the FDR separately for GxE and GxSex analyses because we expected the proportional of truly null tests to be different between GxE and GxSex. In particular, we expected to find more truly positive GxSex tests given previous studies[22,24]. Consistent with this, we found the qvalue procedure for estimating the proportion of truly null hypotheses failed in the GxSex analyses and we had to set the proportion of true null tests (π_0) to 1, which is equivalent to the Benjamini-Hochberg procedure[59]. Story and Tibshirani[60] argue this is much more conservative than the qvalue procedure. Our choice to control each E variable together is conservative, but accounts for non-zero correlations between E variables.

Classification of trait-E pairs into Scenarios

We combined the results of the three statistical approaches to classify trait-E pairs into 3 distinct scenarios. We classified trait-E pairs into Scenario 1 if the Genetic correlation was significantly less than 1. We classified trait-E pairs into Scenario 2 if the SNP-heritability by E and PRSxE regression approaches were significant. We classified trait-E pairs into Scenario 3 if the PRSxE regression approach was significant but the SNP-heritability by E approach was not significant. It is possible that the SNP-heritability by E approach is significant but the PRSxE regression approach is not significant, which should not be viewed as an instance of GxE because the SNP-heritability difference may be driven by changes to the environmental variance rather than the genetic variance. In addition, it is possible for trait-E pairs to be classified into both Scenario 1

and Scenario 2, or both Scenario 1 and Scenario 3, but not both Scenario 2 and Scenario 3 because significance or non-significance of the SNP-heritability by E approach are mutually exclusive.

Scalability of statistical approaches

We consider the scalability of the three statistical approaches we use here. First, there is the computational cost of producing the input to our statistical approaches. For the genetic correlation test, this consists of running GWAS in bins of E variables. There are many scalable approaches for this, including BOLT-LMM[26], regenie[61], and fastGWA[62]. For PRSxE regression, this consists of computing PRS weights. There are many scalable approaches for this including BOLT-LMM[26], PRScs[63], SBayesR[64], and LDpred2[65]. SNP-heritability by E does not require generating additional input. In these analyses, we use BOLT-LMM for GWAS, which has a runtime that scales with O(MN), where M is the number of SNPs and N is the sample size of individuals. For PRSxE regression we use weights computed by Weissbrod et al 2022[27], who did not publish an analysis of runtime. Second, there is the computational cost of the statistical approaches themselves. For genetic correlation, we use cross-trait LDSC[25], which runs in seconds (< 30s for the SNP sets that we analyze here). For PRSxE regression, we use a multiple regression, which also runs in seconds (< 30s for the sample size that we analyze here). For SNP-heritability by E, we use BOLT-REML[33], which has a runtime that scales with O(MN).

Simulations

To test the power of each approach, we simulated 1,000 replicates of each scenario. In all cases, we simulated two E bins and varied the parameters according to the respective generative models. For each replicate, we simulated M=10,000 causal SNPs with effect sizes drawn from a specified distribution. We generated unlinked genotypes with binomial sampling from an allele frequency of 0.5.

We simulated causal effect sizes for each scenario as follows:

Scenario 1

$$\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix} \sim N\left(\begin{bmatrix}0\\ 0\end{bmatrix}, \begin{bmatrix}\sigma_g^2/M & \gamma/M\\ \gamma/M & \sigma_g^2/M\end{bmatrix}\right)$$

We simulated $\sigma_g^2 = 0.25$ and varied γ to produce genetic correlations $r_g \in \{1, 0.99, 0.98, 0.97, 0.96, 0.95, 0.94\}$.

Scenario 2

$$\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix} \sim N\left(\begin{bmatrix}0\\ 0\end{bmatrix}, \begin{bmatrix}\sigma_{g,1}^2/M & \sigma_{g1}\sigma_{g2}/M\\ \sigma_{g1}\sigma_{g2}/M & \sigma_{g,2}^2/M\end{bmatrix}\right)$$

We simulated σ_{g1}^2=0.25 and set σ_{g2}^2 to produce a difference in heritability: $\{0, 0.01, 0.02, 0.03, 0.04, 0.05\}$. Our choice of covariance ensures the genetic correlation is one.

Scenario 3

$$\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix} \sim N\left(\begin{bmatrix}0\\ 0\end{bmatrix}, \begin{bmatrix}\sigma_g^2/M & \sigma_g^2/M\\ \sigma_g^2/M & \sigma_g^2/M\end{bmatrix}\right)$$

We set $\sigma_g^2 = 0.25$. To simulate proportional amplification, we multiplied the phenotypes for individuals in environment 2 by a constant: {1.0,1.025,1.05,1.075,1.1}.

Using the simulated causal effect sizes, we simulated GWAS effect size estimates as:

$$\widehat{\beta_1} \sim N(\beta_1, \frac{(1-h_1^2)/M}{N})$$

$$\widehat{\beta_2} \sim N(\beta_2, \frac{(1-h_2^2)/M}{N})$$

where 1 and 2 index the environments and *N* denotes GWAS sample size. We estimate h_g^2 from the simulated causal effect sizes by first computing the χ^2 statistic ($\chi^2 := N * \hat{\beta}^2$), then computing $h_g^2 = \frac{M}{N} E[\chi^2 - 1]$, where E denotes the mean computed over the independent SNPs[25].

We compute the genetic correlation as:

$$r_g = \frac{\hat{\beta}_1^T \hat{\beta}_2}{\sqrt{\hat{h}_{g,1}^2 * \hat{h}_{g,2}^2}}$$

where *T* denotes the transpose. We compute standard errors for the estimates using a jackknife over SNPs, where we leave out one SNP at a time because they are independent.

To simulate PRSxE, we first simulated causal effect sizes for 10,000 independent SNPs. Then, we compute PRS weights analytically as:

$$\beta_{PRS} = \left(\frac{h_g^2}{h_g^2 + \frac{M}{N}}\right)\hat{\beta}.$$

This simple shrinkage estimator can be interpreted as the posterior mean causal effect size under a normal prior (in the special case of no LD), and is similar to the posterior mean causal effect size under a point-normal prior (in the special case of no LD) when the genetic architecture is highly polygenic[66], as simulated here.

We estimate h_g^2 without knowledge of the E bins, mimicking estimation of SNP-heritability across the 337K individuals; we estimate h_g^2 as the sum of squared standardized effect sizes (averaged across E bins).

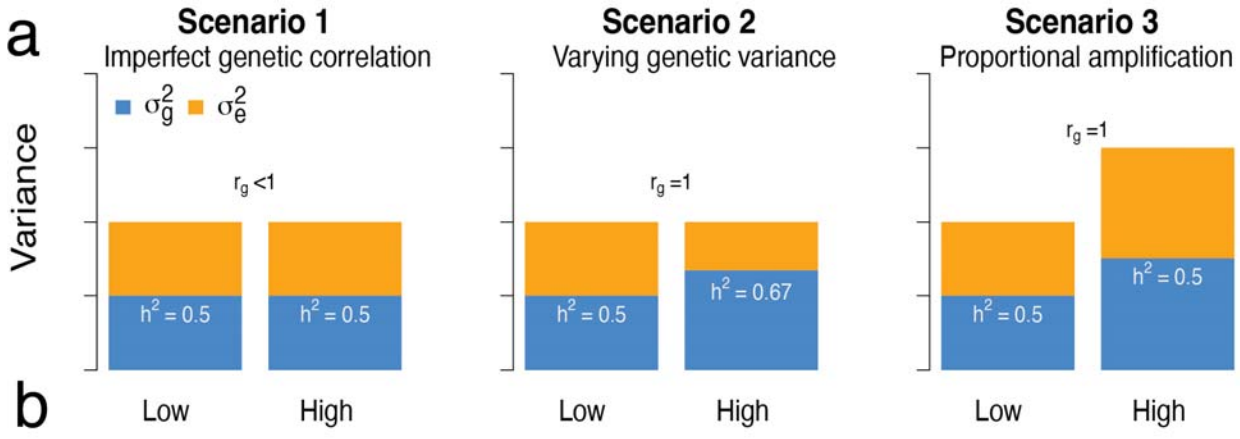
We also evaluated the performance of GxEMM in detecting GxE in Scenarios 1, 2, and 3. We followed the simulation framework in the original publication and simulated 1,000 causal SNPs and 10,000 individuals. We simulated a binary E variable and drew SNP effects according to each Scenario. We performed two tests within the GxEMM framework: 1) IID versus Hom, which tests for polygenic GxE under homoscedasticity, and 2) Free versus Hom, which test for polygenic GxE allowing for heteroskedasticity. We performed a Wald test as implemented in GxEMM and compared the point estimates of heritability in the free model to the simulated heritability in each of the environments. We performed 100 simulation replicates.

To compare GxEMM with our tests, we simulated data under the same framework with matched sample sizes. Specifically, for genetic correlation and SNP-heritability x E, we simulated 5,000 individual per E bin (total N=10,000). For PRSxE, we used a training set sample size of 337K, which matches the real data, and a held out set of N=10,000.

Estimation of trait variance explained

For trait-E pairs in Scenario 1, we compute the trait variance explained by GxE as $(1 - r_g)/2$ for binary E variables (where r_g is the genetic correlation between the two E bins) (**Supplementary Note**) and $\frac{1-r_g}{10}$ for continuous E variables (where r_g is the genetic correlation between the top and bottom quintiles of E values). To obtain the transformation for continuous E variables, we used our simulations (see above) to examine the relationship between estimated genetic correlation and the variance explained by GxE. We found when we binned the E variable into 5 bins and computed the genetic correlation between the top and bottom bins, the transformation $\frac{1-r_g}{10}$ produced accurate estimates of the variance explained by GxE. For trait-E pairs in Scenarios 2 or 3, we divide the variance explained by the PRSxE regression term by the variance explained by the PRS and multiply by the SNP-heritability. We verified these scaling procedures produce accurate estimates of the excess variance explained by GxE in simulations (**Supplementary Figure S3**). For Scenario 1, we simulated a continuous E variable with mean 0 and variance 1 for 337K individuals. We simulated main genetic effects drawn from a normal distribution with mean 0 and variance 0.25 and environment interaction effects from a normal distribution with mean 0 and variance across a range of parameters (1e-1 to 1e-5) for 5,000 SNPs. We binned individuals into 5 bins and ran a GWAS in the top and bottom bins and compute the genetic correlation between the bins. Then, we scaled the estimates according to the formula above (**Supplementary Figure S3a)**. For Scenarios 2 and 3, we simulated 1,000 causal SNPs from a normal distribution with mean 0 and variance 0.25. We simulated a continuous E variable with mean 0 and variance 1 for 49K individuals. We set the amplification parameter to 0.1 and generated phenotypes according to Eq. 1 (*Overview of methods*). We performed GWAS and estimated PRS weights as in the *Simulations* section. Then, we ran the PRSxE test and computed the variance explained. We then compared this to the true variance explained (**Supplementary Figure S3b, S3c**). This scaling assumes that PRSxE effects linearly extrapolate to GxE effects. We do not use the estimates of differences in SNP-heritability by E to estimate the variance explained by GxE. When reporting average variance explained per trait, we computed the R^2 for each trait using a model including all marginally significant (FDR < 5%) interaction terms for that trait (**Supplementary Table 14**).

a



b

Scenario \ Statistical approach	Genetic correlation	PRSxE regression	SNP-heritability by E
1. Imperfect genetic correlation	✓	✗	✗
2. Varying genetic variance	✗	✓	✓
3. Proportional amplification	✗	✓	✗

c

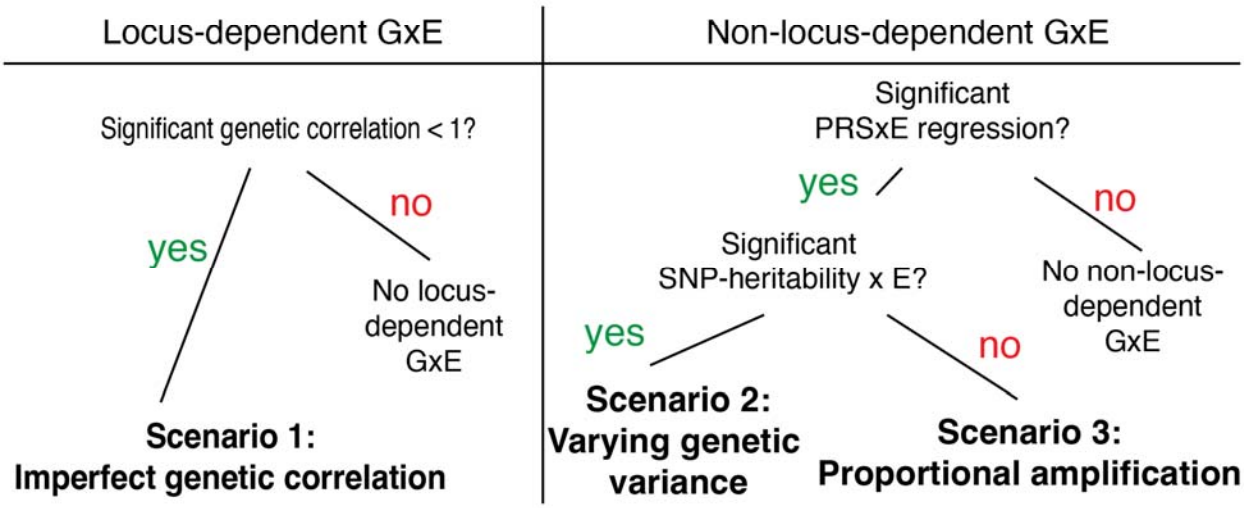


Figure 1. Overview of 3 GxE Scenarios and statistical approaches to detect and distinguish between them. (a) Relative values of genetic (blue) and environmental (orange) variance in each Scenario. (b) Statistical approaches to detect and distinguish between each Scenario. (c) Flow chart for classifying results into Scenarios.

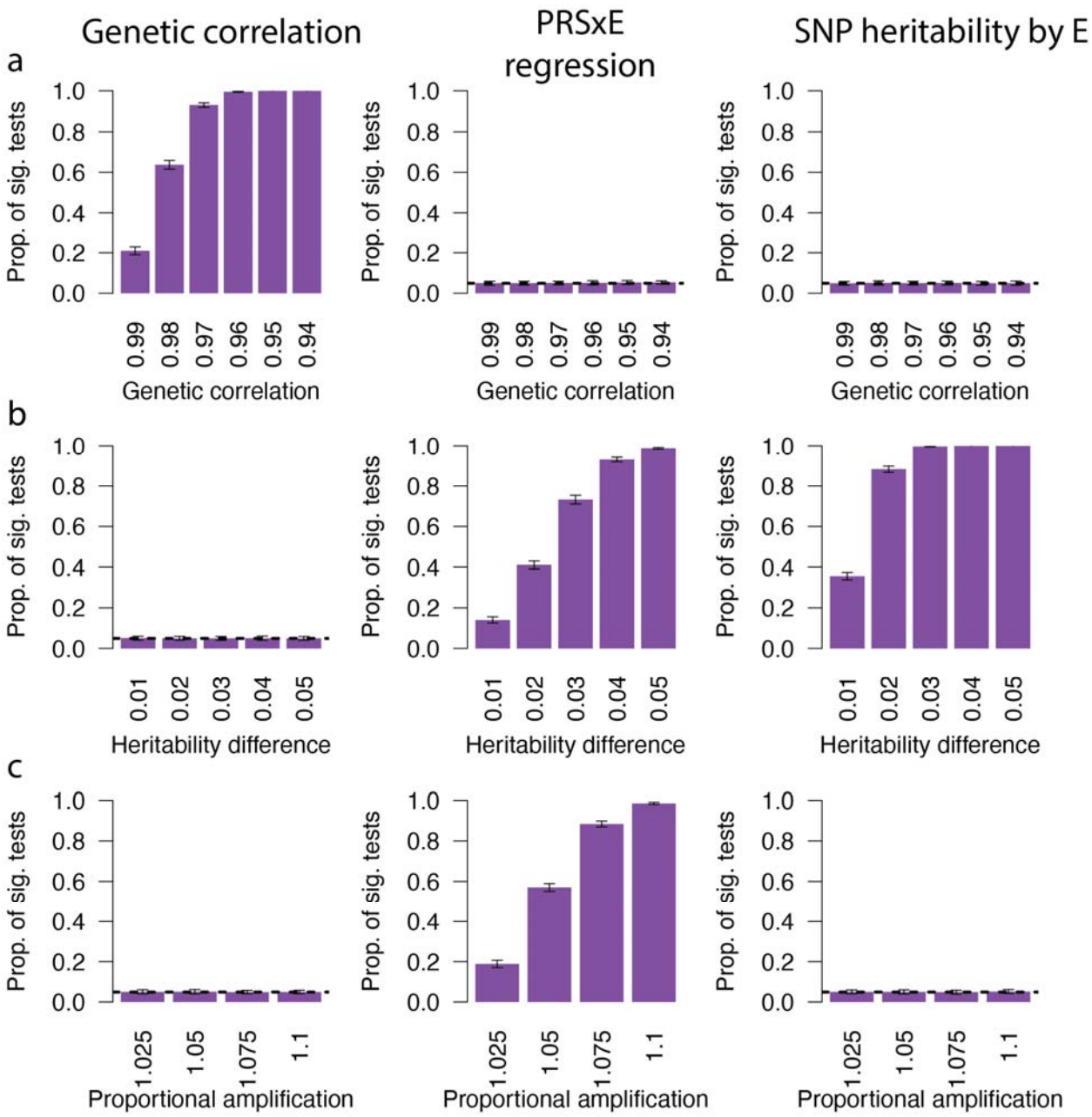


Figure 2. Detecting and distinguishing between 3 Scenarios of GxE interaction in simulations. Rows denote 3 Scenarios (1-3), and columns denote 3 statistical approaches. (a) Proportion of significant tests for Scenario 1 (Imperfect genetic correlation) across 3 statistical approaches. (b) Proportion of significant tests for Scenario 2 (varying genetic variance) across 3 statistical approaches. (c) Proportion of significant tests for Scenario 3 (proportional amplification) across 3 statistical approaches. Error bars denote standard deviations across 100 simulation replicates. Numerical results are reported in **Supplementary Table 1**.

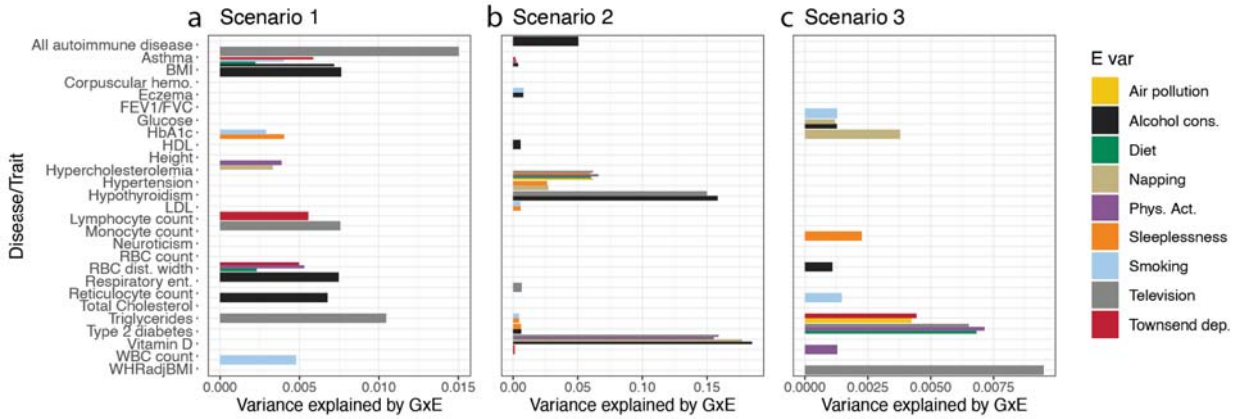


Figure 3. Detecting, quantifying, and distinguishing between 3 Scenarios of GxE interaction across 33 diseases/traits and 10 E variables. Diseases/traits are reported on the y-axis and estimates of excess variance explained by GxE are reported on the x-axis. Only significant results are reported (FDR < 5% across traits and E variables, computed separately for each Scenario). For diseases/traits with multiple significant E variables in a given Scenario, results for each significant E variable are reported separately using bars with smaller thickness. (a) Results for trait-E pairs in Scenario 1: Imperfect genetic correlation. (b) Results for trait-E pairs in Scenario 2: Varying genetic variance; we note that BMI has significant GxE for Townsend deprivation (red), physical activity (purple), and alcohol consumption (black). (c) Results for trait-E pairs in Scenario 3: Proportional amplification. Numerical results are reported in **Supplementary Table S5**.

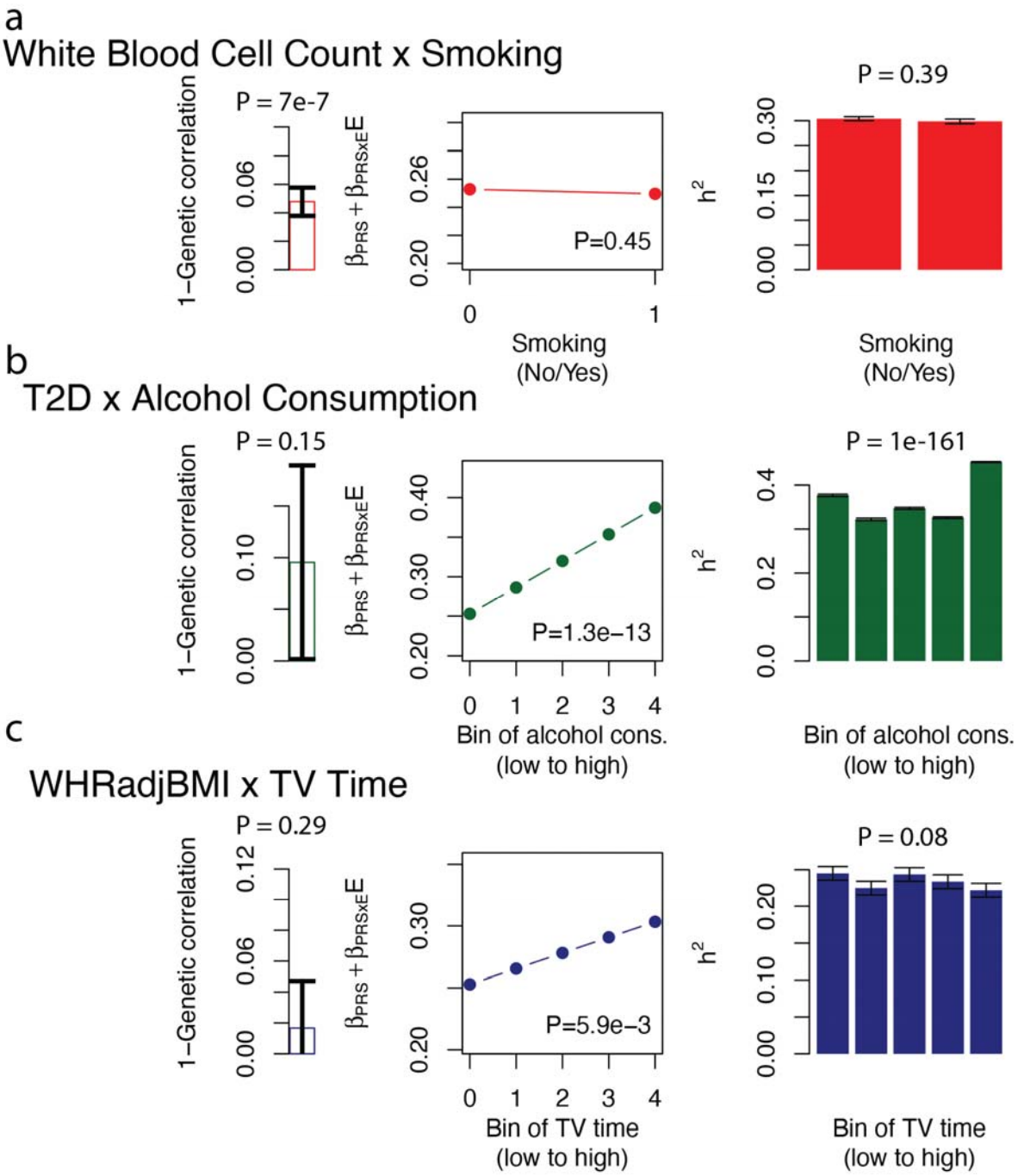


Figure 4. Examples of 3 Scenarios of GxE interaction. (a) White blood cell count x smoking status is consistent with Scenario 1: Imperfect genetic correlation. (b) Type 2 diabetes x alcohol consumption is consistent with Scenario 2: Varying genetic variance. (c) Waist-to-hip ratio adjusted for BMI x Time spent watching TV is consistent with Scenario 3: Proportional amplification. Numerical results are reported in **Supplementary Table S8**.

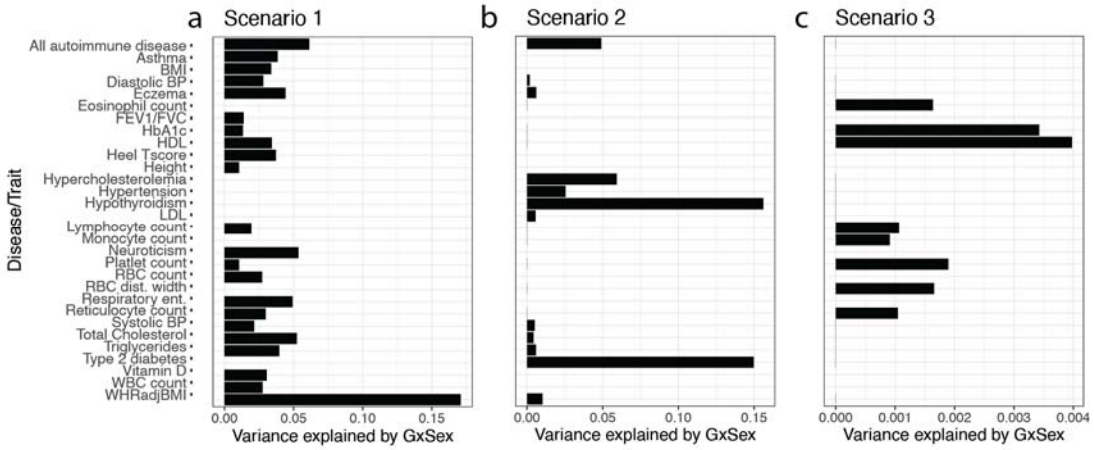


Figure 5. Detecting, quantifying, and distinguishing between 3 Scenarios of GxSex interaction across 33 diseases/traits. Diseases/traits are reported on the y-axis and estimates of excess variance explained by GxSex are reported on the x-axis. Only significant results are reported (FDR < 5% across traits, computed separately for each Scenario). (a) Results for traits in Scenario 1: Imperfect genetic correlation. (b) Results for traits in Scenario 2: Varying genetic variance. (c) Results for traits in Scenario 3: Proportional amplification. Numerical results are reported in **Supplementary Table S9**.

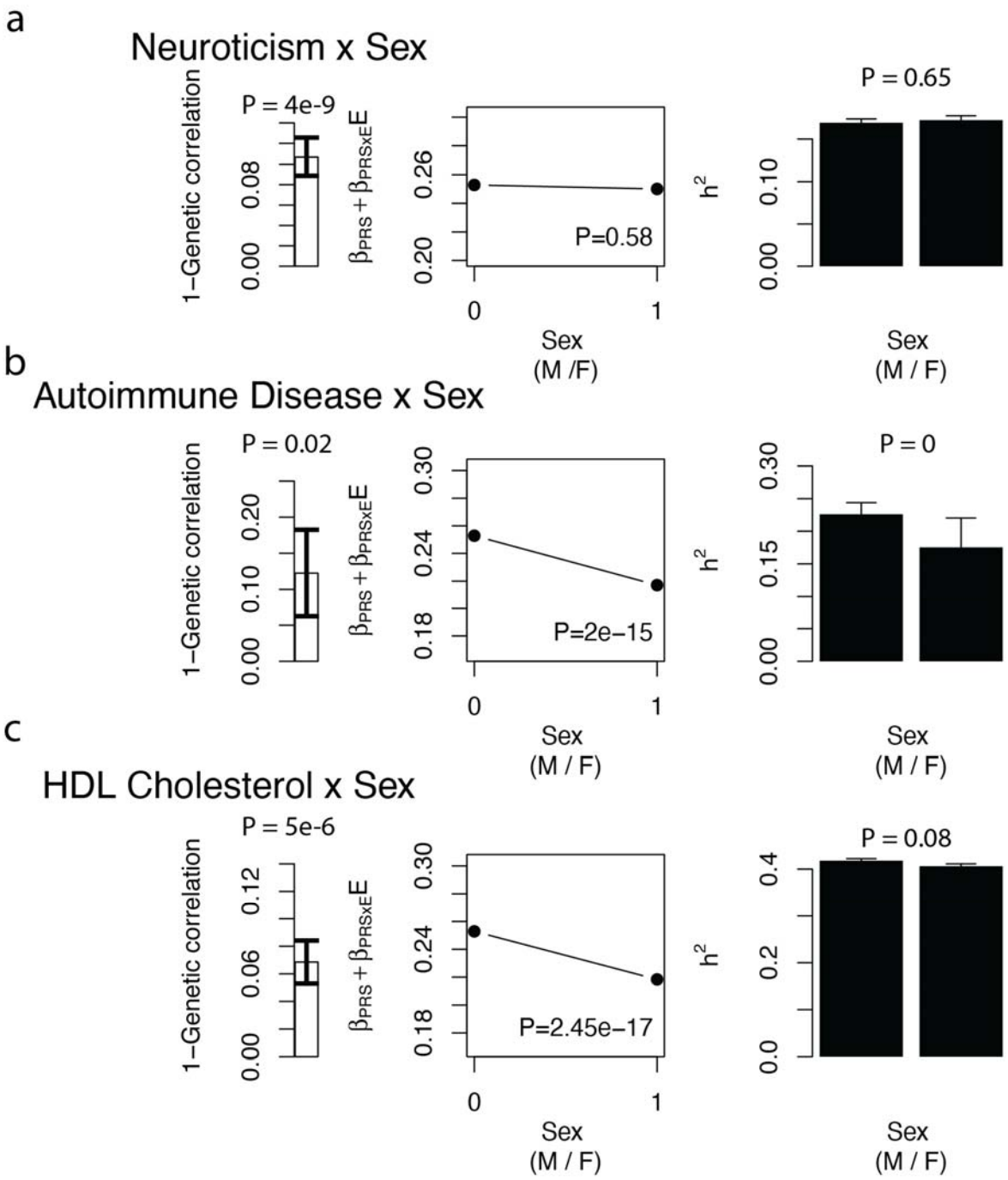


Figure 6. Examples of 3 Scenarios of GxSex interaction. (a) Neuroticism x Sex is consistent with Scenario 1: Imperfect genetic correlation. (b) All autoimmune disease x Sex is consistent with Scenario 2: Varying genetic variance. (c) HDL Cholesterol x Sex is consistent with Scenario 1 and Scenario 3. Numerical results are reported in **Supplementary Table S11**.

References

1. Hunter, D. J. Gene–environment interactions in human diseases. *Nat. Rev. Genet.* **6**, 287–298 (2005).
2. Thomas, D. Gene-Environment-Wide Association Studies: Emerging Approaches. *Nat. Rev. Genet.* **11**, 259–272 (2010).
3. Franks, P. W. & McCarthy, M. I. Exposing the exposures responsible for type 2 diabetes and obesity. *Science* **354**, 69–73 (2016).
4. Li, J., Li, X., Zhang, S. & Snyder, M. Gene-Environment Interaction in the Era of Precision Medicine. *Cell* **177**, 38–44 (2019).
5. Hu, Z. *et al.* A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. *Nat. Genet.* **43**, 792–796 (2011).
6. Wu, C. *et al.* Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify multiple susceptibility loci and gene-environment interactions. *Nat. Genet.* **44**, 1090–1097 (2012).
7. Young, A. I., Wauthier, F. & Donnelly, P. Multiple novel gene-by-environment interactions modify the effect of FTO variants on body mass index. *Nat. Commun.* **7**, 12724 (2016).
8. Yang, J. *et al.* FTO genotype is associated with phenotypic variability of body mass index. *Nature* **490**, 267–272 (2012).
9. Shungin, D. *et al.* Ranking and characterization of established BMI and lipid associated loci as candidates for gene-environment interactions. *PLOS Genet.* **13**, e1006812 (2017).
10. Young, A. I., Wauthier, F. L. & Donnelly, P. Identifying loci affecting trait variability and detecting interactions in genome-wide association studies. *Nat. Genet.* **50**, 1608–1614 (2018).

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

12. Westerman, K. E. *et al.* Variance-quantitative trait loci enable systematic discovery of gene-environment interactions for cardiometabolic serum biomarkers. *Nat. Commun.* **13**, 3993 (2022).

13. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).

14. Robinson, M. R. *et al.* Genotype–covariate interaction effects and the heritability of adult body mass index. *Nat. Genet.* **49**, 1174–1181 (2017).

15. Moore, R. *et al.* A linear mixed-model approach to study multivariate gene–environment interactions. *Nat. Genet.* **51**, 180–186 (2019).

16. Dahl, A. *et al.* A Robust Method Uncovers Significant Context-Specific Heritability in Diverse Complex Traits. *Am. J. Hum. Genet.* **106**, 71–91 (2020).

17. Kerin, M. & Marchini, J. Inferring Gene-by-Environment Interactions with a Bayesian Whole-Genome Regression Model. *Am. J. Hum. Genet.* **107**, 698–713 (2020).

18. Di Scipio, M. *et al.* A versatile, fast and unbiased method for estimation of gene-by-environment interaction effects on biobank-scale datasets. *Nat. Commun.* **14**, 5196 (2023).

19. Traglia, M. *et al.* Genetic Mechanisms Leading to Sex Differences Across Common Diseases and Anthropometric Traits. *Genetics* **205**, 979–992 (2017).

20. Khramtsova, E. A., Davis, L. K. & Stranger, B. E. The role of sex in the genomics of human complex traits. *Nat. Rev. Genet.* **20**, 173–190 (2019).

21. Kamitaki, N. *et al.* Complement genes contribute sex-biased vulnerability in diverse disorders. *Nature* **582**, 577–581 (2020).

22. Bernabeu, E. *et al.* Sex differences in genetic architecture in the UK Biobank. *Nat. Genet.* **53**, 1283–1289 (2021).

23. Khramtsova, E. A. *et al.* Quality control and analytic best practices for testing genetic models of sex differences in large populations. *Cell* **186**, 2044–2061 (2023).

24. Zhu, C. *et al.* Amplification is the primary mode of gene-by-sex interaction in complex human traits. *Cell Genomics* **3**, (2023).

25. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).

26. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat. Genet.* **50**, 906–908 (2018).

27. Weissbrod, O. *et al.* Leveraging fine-mapping and multipopulation training data to improve cross-population polygenic risk scores. *Nat. Genet.* **54**, 450–458 (2022).

28. Rask-Andersen, M., Karlsson, T., Ek, W. E. & Johansson, Å. Gene-environment interaction study for BMI reveals interactions between genetic factors and physical activity, alcohol consumption and socioeconomic status. *PLoS Genet.* **13**, e1006977 (2017).

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

30. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).

31. Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* **88**, 294–305 (2011).

32. Lee, S. H. *et al.* Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat. Genet.* **44**, 247–250 (2012).

33. Loh, P.-R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. Genet.* **47**, 1385–1392 (2015).

34. Pazokitoroudi, A. *et al.* Efficient variance components analysis across millions of genomes. *Nat. Commun.* **11**, 4020 (2020).

35. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203 (2018).

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

37. Voorman, A., Lumley, T., McKnight, B. & Rice, K. Behavior of QQ-Plots and Genomic Control in Studies of Gene-Environment Interaction. *PLOS ONE* **6**, e19416 (2011).

38. Huber, P. J. The behavior of maximum likelihood estimates under nonstandard conditions. in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics* vol. 5.1 221–234 (University of California Press, 1967).

39. White, H. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica* **48**, 817–838 (1980).

40. Pedersen, K. M. *et al.* Smoking and Increased White and Red Blood Cells. *Arterioscler. Thromb. Vasc. Biol.* **39**, 965–977 (2019).

41. Baliunas, D. O. *et al.* Alcohol as a Risk Factor for Type 2 Diabetes. *Diabetes Care* **32**, 2123–2132 (2009).

42. Wendt, F. R. *et al.* Sex-Specific Genetic and Transcriptomic Liability to Neuroticism. *Biol. Psychiatry* **93**, 243–252 (2023).

800 43. Whitacre, C. C. Sex differences in autoimmune disease. *Nat. Immunol.* **2**, 777–780
801 (2001).
802 44. Mostafavi, H. *et al.* Variable prediction accuracy of polygenic scores within an ancestry
803 group. *eLife* **9**, e48376 (2020).
804 45. Shi, H. *et al.* Population-specific causal disease effect sizes in functionally important
805 regions impacted by selection. *Nat. Commun.* **12**, 1098 (2021).
806 46. Visscher, P. M., Hill, W. G. & Wray, N. R. Heritability in the genomics era--concepts
807 and misconceptions. *Nat. Rev. Genet.* **9**, 255–266 (2008).
808 47. Polderman, T. J. C. *et al.* Meta-analysis of the heritability of human traits based on fifty
809 years of twin studies. *Nat. Genet.* **47**, 702–709 (2015).
810 48. Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability:
811 Genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 1193–
812 1198 (2012).
813 49. Cordell, H. J. Detecting gene-gene interactions that underlie human diseases. *Nat. Rev.*
814 *Genet.* **10**, 392–404 (2009).
815 50. Zaitlen, N. *et al.* Informed Conditioning on Clinical Covariates Increases Power in Case-
816 Control Association Studies. *PLOS Genet.* **8**, e1003032 (2012).
817 51. Jiang, X., Holmes, C. & McVean, G. The impact of age on genetic risk for common
818 diseases. *PLOS Genet.* **17**, e1009723 (2021).
819 52. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links
820 cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).
821 53. All of Us Research Program Investigators *et al.* The 'All of Us' Research Program. *N.*
822 *Engl. J. Med.* **381**, 668–676 (2019).

54. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584 (2019).

55. Dey, R. *et al.* Efficient and accurate frailty model approach for genome-wide survival association analysis in large-scale biobanks. *Nat. Commun.* **13**, 5437 (2022).

56. Pedersen, E. M. *et al.* ADuLT: An efficient and robust time-to-event GWAS. *Nat. Commun.* **14**, 5553 (2023).

57. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, s13742-015-0047–8 (2015).

58. qvalue: Q-value estimation for false discovery rate control. (2023).

59. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).

60. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9440–9445 (2003).

61. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.* **53**, 1097–1103 (2021).

62. Jiang, L. *et al.* A resource-efficient tool for mixed model association analysis of large-scale data. *Nat. Genet.* **51**, 1749–1755 (2019).

63. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).

64. Lloyd-Jones, L. R. *et al.* Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat. Commun.* **10**, 5086 (2019).

65. Privé, F., Arbel, J. & Vilhjálmsson, B. J. LDpred2: better, faster, stronger. *Bioinformatics* **36**, 5424–5431 (2021).

66. Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).

Supplementary Material

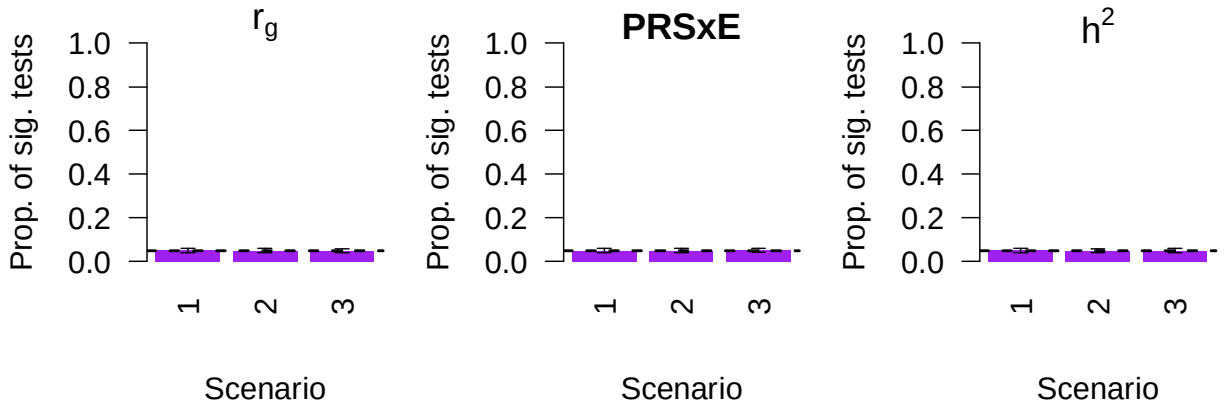


Figure S1 Results of 3 statistical approaches for detecting GxE in null simulations with no GxE. (a) Proportion of significant tests for genetic correlation across 3 scenarios. (b) Proportion of significant tests for PRSxE regression across 3 scenarios. (c) Proportion of significant tests for SNP-heritability by E across 3 scenarios. Error bars denote standard deviations across 100 simulation replicates.

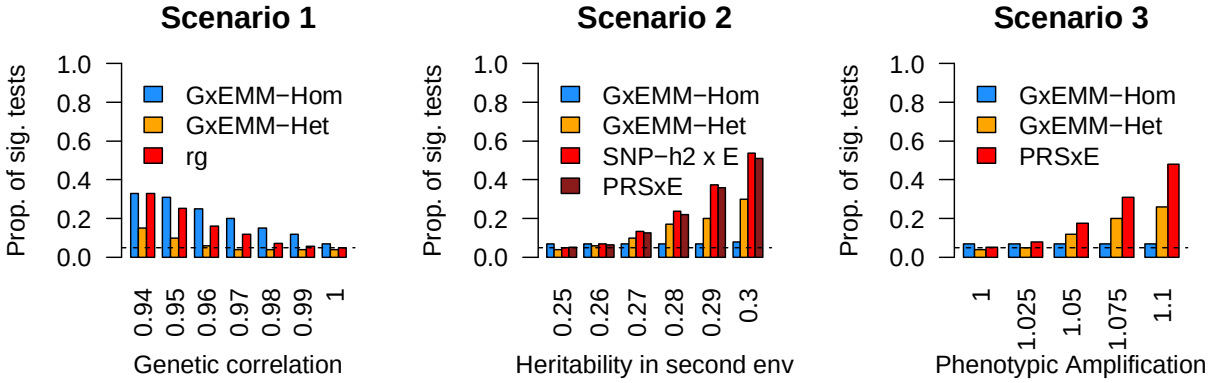


Figure S2 Comparison of three statistical approaches for detecting GxE to ExEMM in simulations. a) Scenario 1 with varying true genetic correlation across E, b) Scenario 2 with varying heritability in the second environment, c) Scenario 3 with phenotypic amplification across E bins.

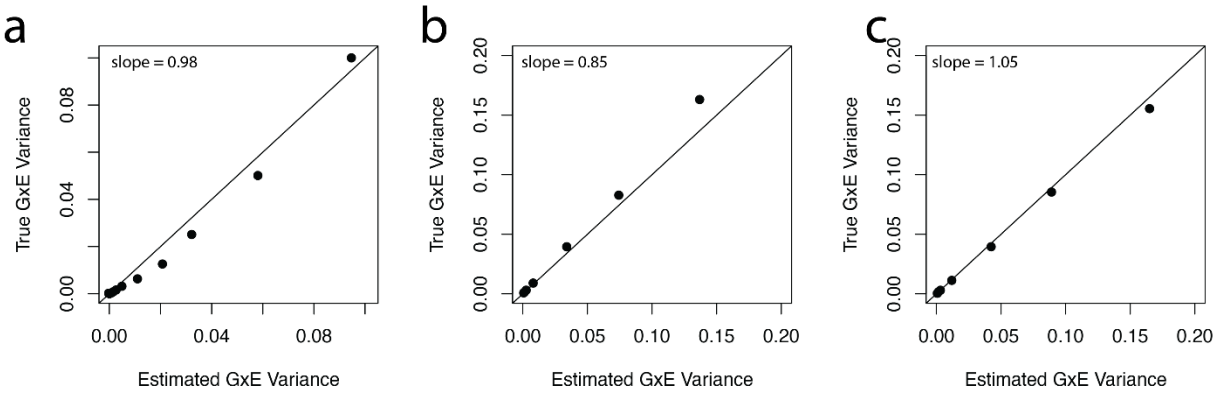
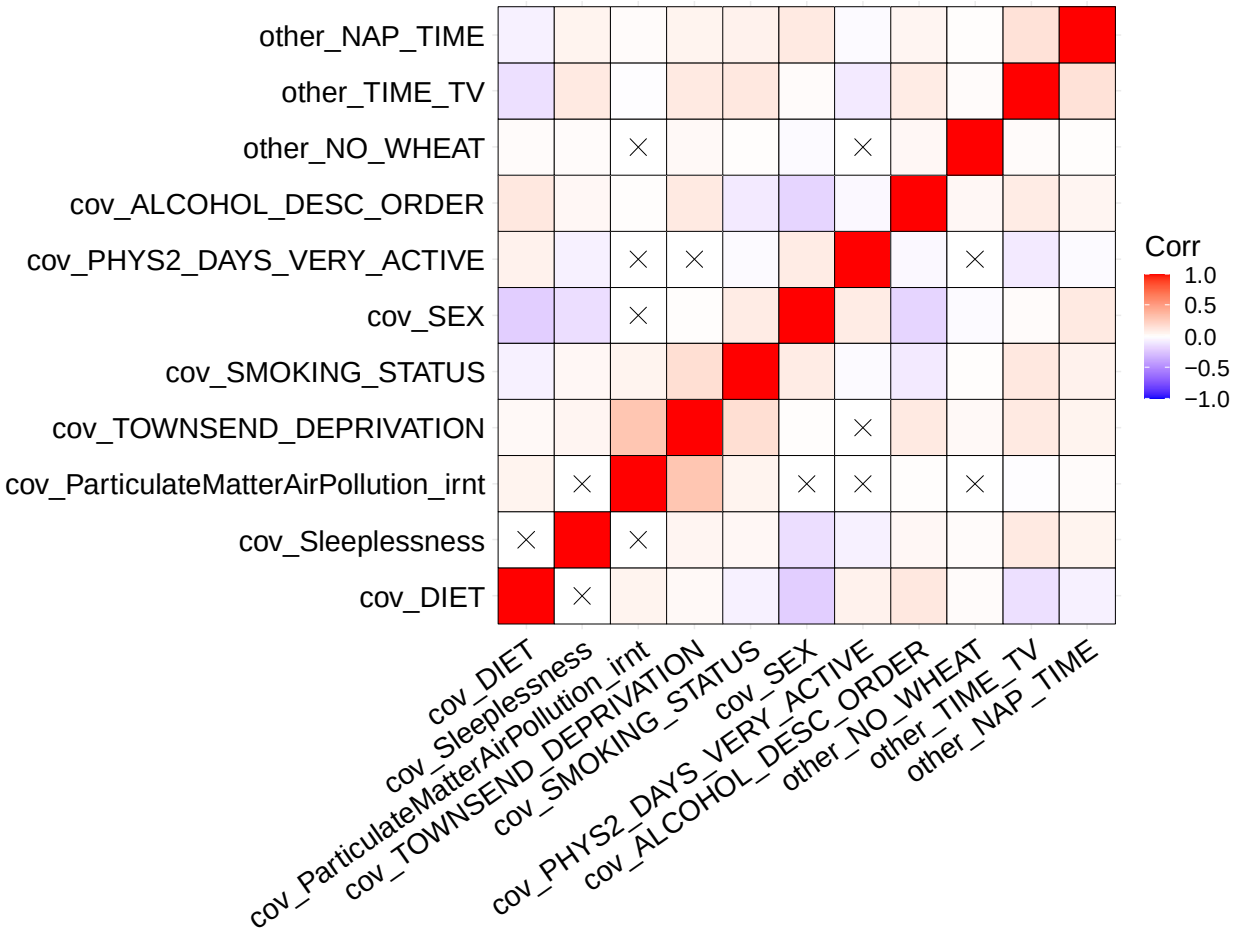


Figure S3 Accuracy of estimates of excess trait variance explained by GxE interaction in simulations. a) genetic correlation in Scenario 1, b) PRSxE in Scenario 2, and c) PRSxE in Scenario 3. For all plots, the black line corresponds to the y=x line and the x and y axes are both on a log scale.



873
874 **Figure S4 Phenotypic correlations between E variables.** X denotes non-significant
875 comparisons at a p-value threshold of 0.05/11.

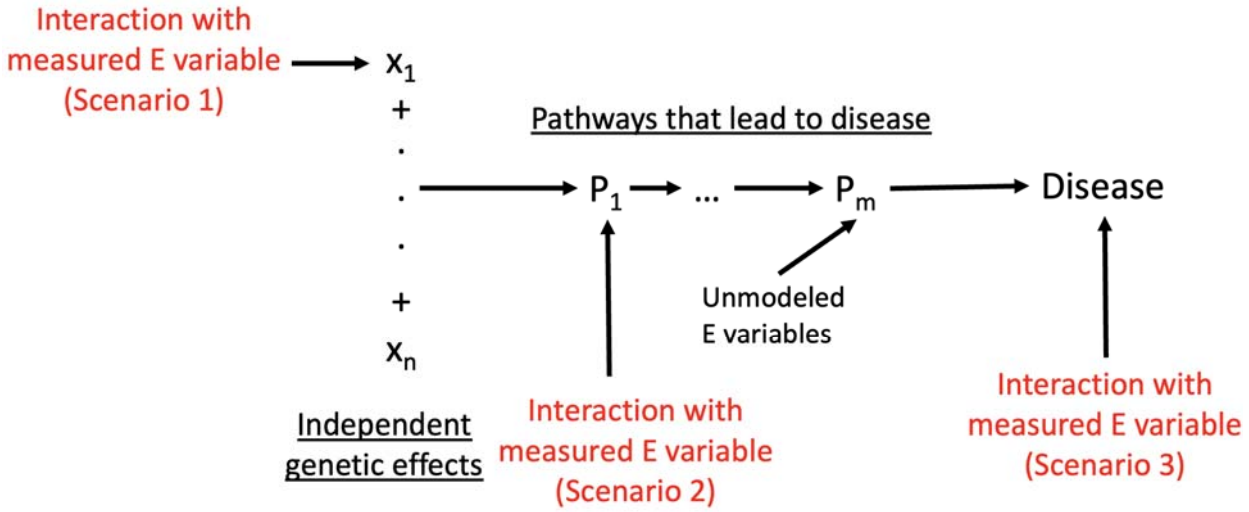


Figure S5 Conceptual model linking three scenarios of GxE. Scenario 1 can be conceptualized as E variables modifying the effects of independent loci. Scenario 2 can be conceptualized as modifying pathways which aggregate the genetic effects of many loci, resulting in a scaling of genetic effects. Finally, Scenario 3 can be conceptualized as modifying the total genetic liability.

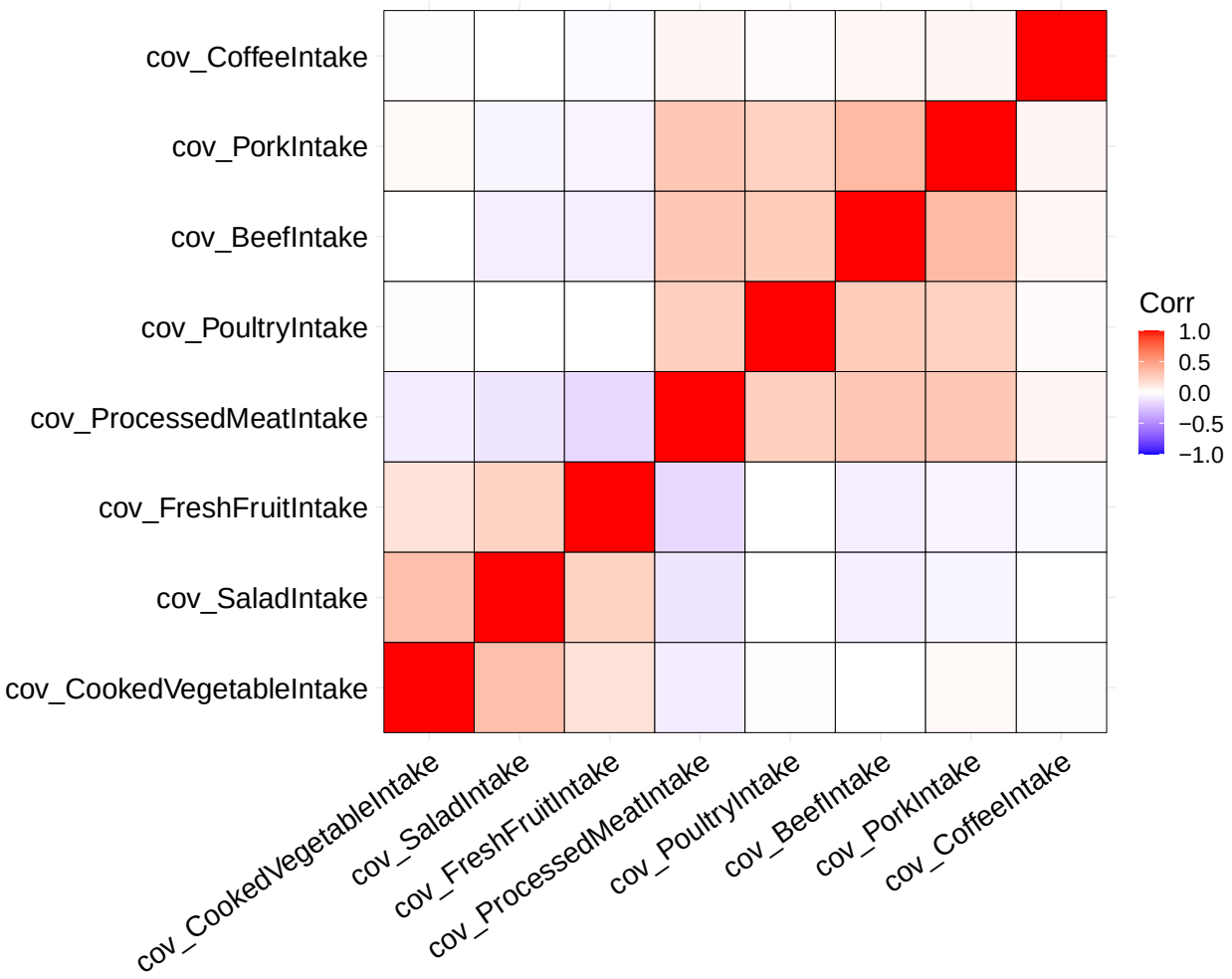


Figure S6 Phenotypic correlations of diet variables used to construct a composite Diet variable. Each cell in the heatmap shows the correlation between the measured diet variable on the X axis and the measured diet variable on the Y axis. All correlations are significant at a Bonferroni corrected p-value threshold of 0.05.

Supplementary Table 1 Numerical results of Detecting and distinguishing between 3 Scenarios of GxE interaction in simulations. For each statistical approach and scenario, we report the proportion of significant tests and standard deviation across replicates.

Supplementary Table 2 Simulations showing bias induced by correlated G and GxE effects. We tested the impact of correlated G and GxE effects on variance component estimates when assuming that G and GxE effects are not correlated. We set the true variance of G effects to 0.1 and the true variance of GxE effects to 0.1. We varied the correlation of the G and GxE effects

and simulated values for 100,000 individuals. We estimate the variance explained by G and GxE using ANOVA in R and report the bias (estimated effect - true effect).

Supplementary Table 3 Description of the 33 UK Biobank traits analyzed. For each trait we report a detailed name, the GWAS sample size (including number of cases for binary traits), the SNP-heritability (liability scale for binary traits), and the PRS accuracy (R^2; observed scale for binary traits).

Supplementary Table 4 SNP-heritability of E variables studied here. We estimated SNP-heritability using LDSC. For the composite diet variable, we report the SNP-heritability for each of the underlying variables that make up the composite diet variable. P-values test against a null of zero SNP-heritability.

Supplementary Table 5 Numerical results of Detecting, quantifying, and distinguishing between 3 Scenarios of GxE interaction across 33 diseases/traits and 10 E variables. For each trait-E pair (A, B), we report (C) the excess variance explained by PRSxE and (D) the associated q value, (E) the difference in heritability between the top and bottom bins of the E variables and (F) the associated q value, (G) the genetic correlation between the top and bottom bin of the E variable and (H) the associated q value. We also assign each trait-E pair to the three scenarios (I, J, K).

Supplementary Table 6 P-values using robust regression in PRSxE regression analysis compared to P-value from the main PRSxE regression analysis. For each trait-E pair, we report (A) p-value and (B) effect size for the main PRSxE regression and (C, D) using the Huber-White variance estimator (robust regression).

Supplementary Table 7 SNP-heritability differences for trait-E pairs with no PRSxE interaction. For each trait-E pair with a significant difference in SNP-heritability and no significant PRSxE interaction we report the SNP-heritability difference and q-value at 5% FDR control.

Supplementary Table 8 Numerical results of Examples of 3 Scenarios of GxE interaction. We report detailed results for 3 trait-E pairs reported in Figure 4. For each trait-E pair (A, B), we report (C) the genetic correlation and (D) p-value, (E) PRSxE regression coefficient and (F) p-value, (G, H, I, J, K) SNP-heritability across bins of the E variable with associated standard error and (L) the p-value testing for a difference between the top and bottom bins of the E variable.

Supplementary Table 9 Numerical results of Detecting, quantifying, and distinguishing between 3 Scenarios of GxSex interaction across 33 diseases/traits. For each trait (A, B), we report (C) the excess variance explained by PRSxSex and (D) the associated q value, (E) the difference in heritability between males and females and (F) the associated q value, (G) the genetic correlation between males and females and (H) the associated q value. We also assign each trait-E pair to the three scenarios (I, J, K).

Supplementary Table 10 SNP-heritability differences for trait-sex pairs with no PRSxSex interaction. For each trait-sex pair with a significant difference in SNP-heritability and no

significant PRSxSex interaction we report the SNP-heritability difference and q-value at 5% FDR control.

Supplementary Table 11 Numerical results of Examples of 3 Scenarios of GxSex interaction. We report detailed results for 3 trait-sex pairs reported in Figure 4. For each trait (A, B), we report (C) the genetic correlation and (D) p-value, (E) PRSxE regression coefficient and (F) p-value, (G, H, I, J, K) SNP-heritability across sex with associated standard error and (L) the p-value testing for a difference between the top and bottom bins of the E variable.

Supplementary Table 12 Average variance explained by binary and quantitative traits. For each scenario, we report the average trait variance explained by binary, quantitative, and all traits for both GxE and GxSex interactions.

Supplementary Table 13 PRSxE regression results including a non-linear E term. For each trait-E pair, we report (A) P-value and (B) effect size including E^2 as a covariate and (C, D) not including E^2 as a covariate.

Supplementary Table 14 Trait variance explained by GxE interactions with multiple E variables. For traits with multiple marginally significant E variable interactions, we report the variance explained by a joint model with all marginally significant E variables.

Supplementary Note for "Distinct explanations underlie gene-environment interactions in the UK Biobank"