nature genetics

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41588-018-0271-0

In the format provided by the authors and unedited.

A linear mixed-model approach to study multivariate gene–environment interactions

Rachel Moore[1,2,3,9], Francesco Paolo Casale[4,9], Marc Jan Bonder[2], Danilo Horta[2], BIOS Consortium[5], Lude Franke[6], Inês Barroso[1*] and Oliver Stegle[2,7,8*]

[1]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK. [2]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK. [3]University of Cambridge, Cambridge, UK. [4]Microsoft Research New England, Cambridge, Massachusetts, USA. 5A full list of members and affiliations appears at the end of the paper. [6]University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, the Netherlands. [7]European Molecular Biology Laboratory, Genome Biology Unit, Heidelberg, Germany. [8]Division of Computational Genomics and Systems Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany. [9]These authors contributed equally: Rachel Moore, Francesco Paolo Casale. *e-mail: ib1@sanger.ac.uk; oliver.stegle@embl.de

Supplementary tables

Main simulations										
N	-	-	-	1,000	2,000	**5,000**	-	-	-	-
L	-	2	10	20	40	**60**	80	100	-	-
π	-	-	3.3%	16.7%	33.3%	**50%**	66.7%	100%	-	-
v_g	-	-	-	-	-	**0.6%**	-	-	-	-
ρ	-	0.0	0.1	0.3	0.5	**0.7**	0.8	0.9	1.0	-
v_e	-	-	-	0	0.1	**0.2**	0.3	0.4	-	-
v_{pop}	-	-	-	-	-	**0.4**	-	-	-	-
L_{unobs}	-	-	-	-	-	**0**	10	20	30	40
Heritable environments										
r^2	0-0.1	0.1-0.2	0.2-0.5	0.5-0.8	0.8-1	**1**	-	-	-	-
v_x	-	0.01	0.02	0.05	0.10	**0.20**	-	-	-	-
Skewed (Gamma-distributed) environments										
a	-	100	5	3	2	1	0.5	0.2	0.1	-
Binary environments										
f_B	-	0	0.25	0.50	0.75	**1**	-	-	-	-
ν	-	0.50	0.20	0.10	0.05	**0.02**	0.01	0.005	-	-

Supplementary Table 1 | Parameters used for the simulation experiments. Shown are the default parameter values and ranges as considered in the calibration and power experiments. Specifically, simulation experiments shown in (**Fig. 1**,**2**, **Supp. Fig. 2**) are based on empirical environments from UK Biobank, varying the sample size (N), the number of environments (L), the percentage of environments that contribute to GxE (π), sample variance explained by the total genetic effect (G+GxE effect, v_g), fraction of the genetic variance due to GxE (ρ) and the variance explained by additive environmental effects (v_e), population structure (v_{pop}), residual noise (v_n) and number of unobserved environments contributing to GxE (L_{unobs}). Default parameter values are shown in bold and are left constant when varying other parameters. In additional simulations based on synthetic environments (**Supp. Fig. 3**,**4**), we vary LD between variants that affect environments and variants with G or GxE effects (r^2), the average fraction of variance explained by the variant across environments (v_x, heritability). For simulated skewed environments, we vary the shape parameter of the gamma distribution (a) and finally, for binary environments, we vary both the fraction of environments that are binary (f_B) and the event frequency (ν). All remaining genetic parameters were set to default parameters.

Method name	GxE	Additive E	Test type	Number parameters for additive E	Number DoF Interaction test	Number DoF association test
Multi-environment tests						
StructLMM	**Random**	**Random**	**Score**	**1**	**1**	**2**
MultiEnv-Renv-LRT	Fixed	Random	LRT	1	60	61
MultiEnv-Fenv-LRT	Fixed	Fixed	LRT	60	60	61
MultiEnv-Renv-Score	Fixed	Random	Score	1	60	61
MultiEnv-Fenv-Score	Fixed	Fixed	Score	60	60	61
Single-environment tests						
SingleEnv-Renv	**Fixed**	**Random**	**LRT**	**1**	**1**	**2**
SingleEnv-Fenv	Fixed	Fixed	LRT	60	1	2
SingleEnv-Senv	Fixed	Fixed	LRT	1	1	2
Linear association tests						
LMM-Renv	**None**	**Random**	**LRT**	**1**	**NA**	**1**
LM-Fenv	None	Fixed	LRT	60	NA	1
LM	None	None	LRT	0	NA	1

Supplementary Table 2 | Tabular overview of considered methods. Shown is the name of the method (column 1; name of the association method displayed and '-int' appended to the single- and multi-environment test names for the corresponding interaction tests), whether a random or fixed effect term is used to model GxE under the alternative hypothesis (column 2), whether a random or fixed effect term is used to model the additive environment (column 3), the statistical test used to assess the alternative hypothesis (column 4), how many parameters are used to model the additive environment (column 5), the number of additional parameters used to model the alternative hypothesis versus the null hypothesis for interaction and association tests (columns 6 and 7 respectively). The number of model parameters in the final three columns assume that 60 environmental variables are used in the test (default simulation setting; see **Supp. Table 1**). The tests are grouped into multi-environment tests, single-environment tests, and linear association tests. Representative methods that are considered in main text results are highlighted in bold.

Supplementary Table 3 | Interactions identified by StructLMM for BMI in UK Biobank. Provided as supplementary data file Supplementary_table_3.xlsx.

Supplementary Table 4 | Associations identified by StructLMM and LMM in the association analysis of BMI using data from UK Biobank. Provided as supplementary data file Supplementary_table_4.xlsx.

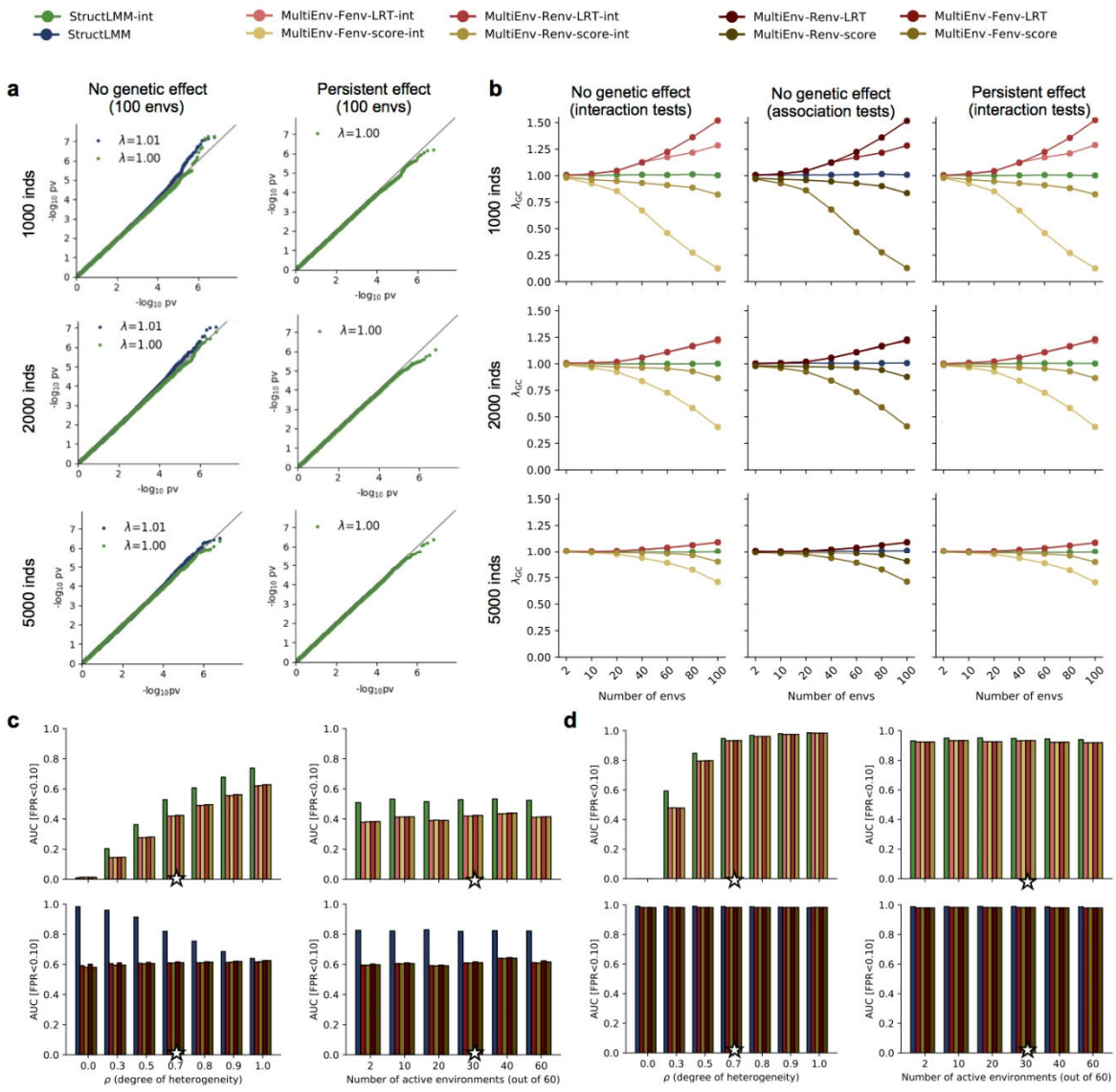
Supplementary Table 5 | Summary table of interaction eQTL analysis in blood cohort. Provided as supplementary data file Supplementary_table_5.xlsx.

Supplementary Table 6 | Pathway enrichment analysis for interactions eQTL that are in linkage with GWAS loci. Provided as supplementary data file Supplementary_table_6.xlsx.

Supplementary Dataset 1 | eQTL Manhattan plots for interaction eQTL that colocalise with disease variants. Manhattan plots provided as supplementary data file Supplementary_data_1.zip.

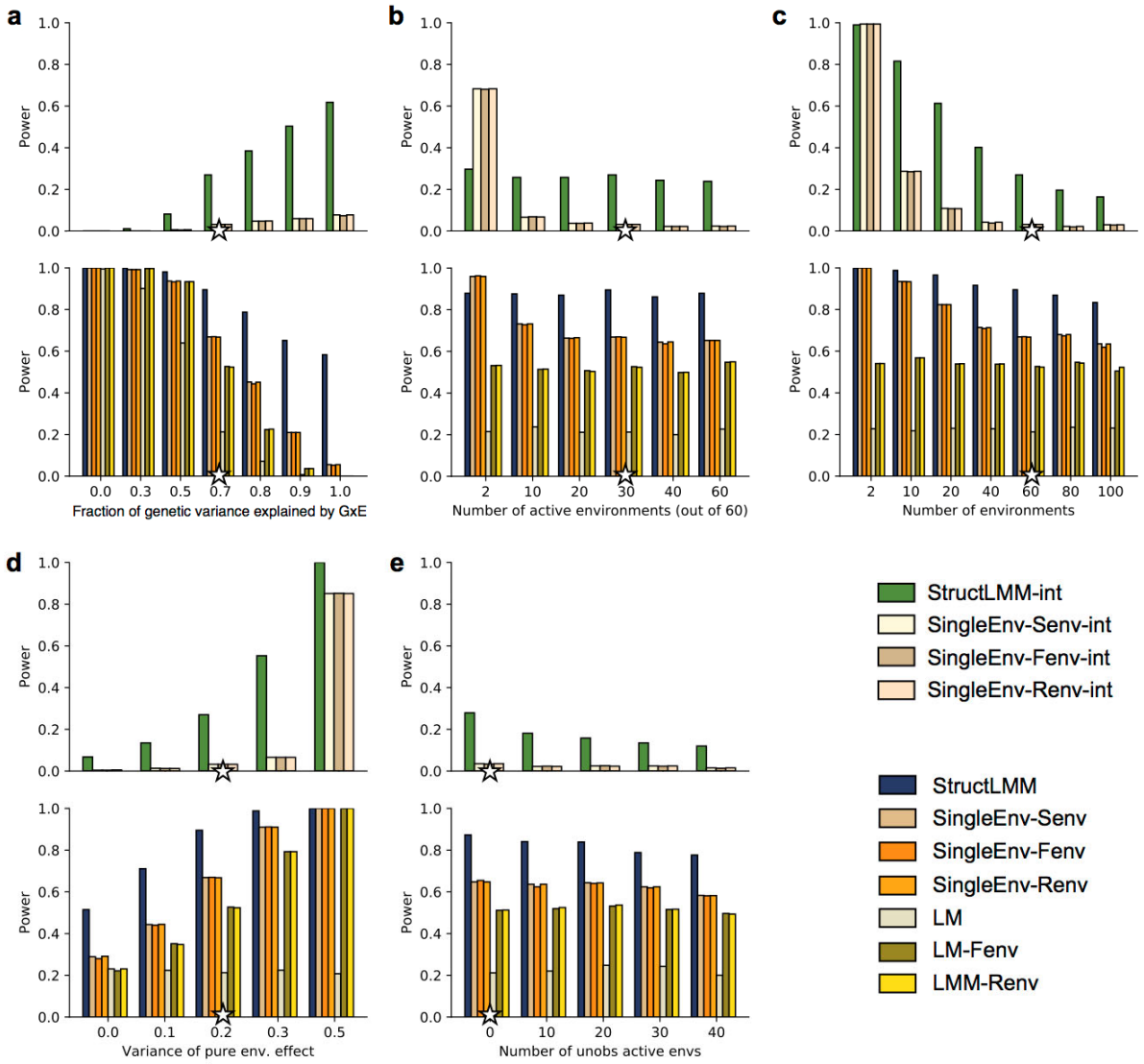
Supplementary Dataset 2 | Interaction eQTL colocalising with disease variants. Figures analogous to **Fig. 5b-d** for 64 interaction eQTL with putative colocalisation with disease variants, provided as supplementary data file Supplementary_data_2.zip.

Supplementary figures

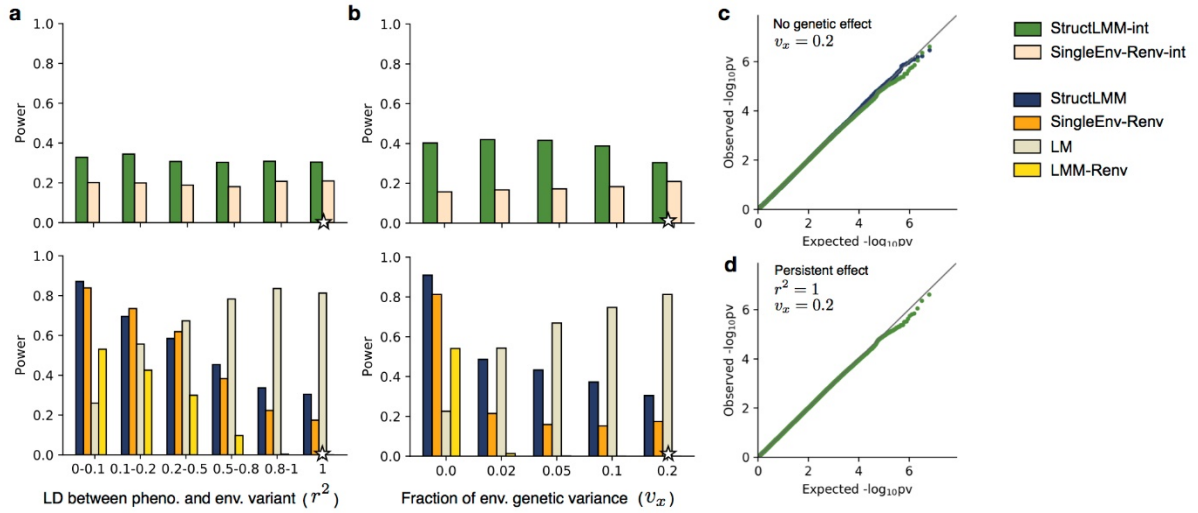


Supplementary Figure 1 | Calibration of StructLMM and comparison with alternative multi-environment tests. (**a**) QQ plots of negative log P values from StructLMM-int (green) and StructLMM (blue) for 103,527 variants on chromosome 21, either simulating no genetic effects (no G, no GxE, $v_g = 0$) or simulating persistent genetic effects (no GxE, $\rho = 0$; only StructLMM-int). Simulations are based on 100 environments ($L = 100$, $\pi = 100\%$). From top to bottom: increasing sample sizes of a synthetic population based on the European population from the 1000 Genomes project: 1,000 individuals, 2,000 individuals and 5,000 individuals. Default parameters were used for all other simulation settings; see **Supp. Table 1.**

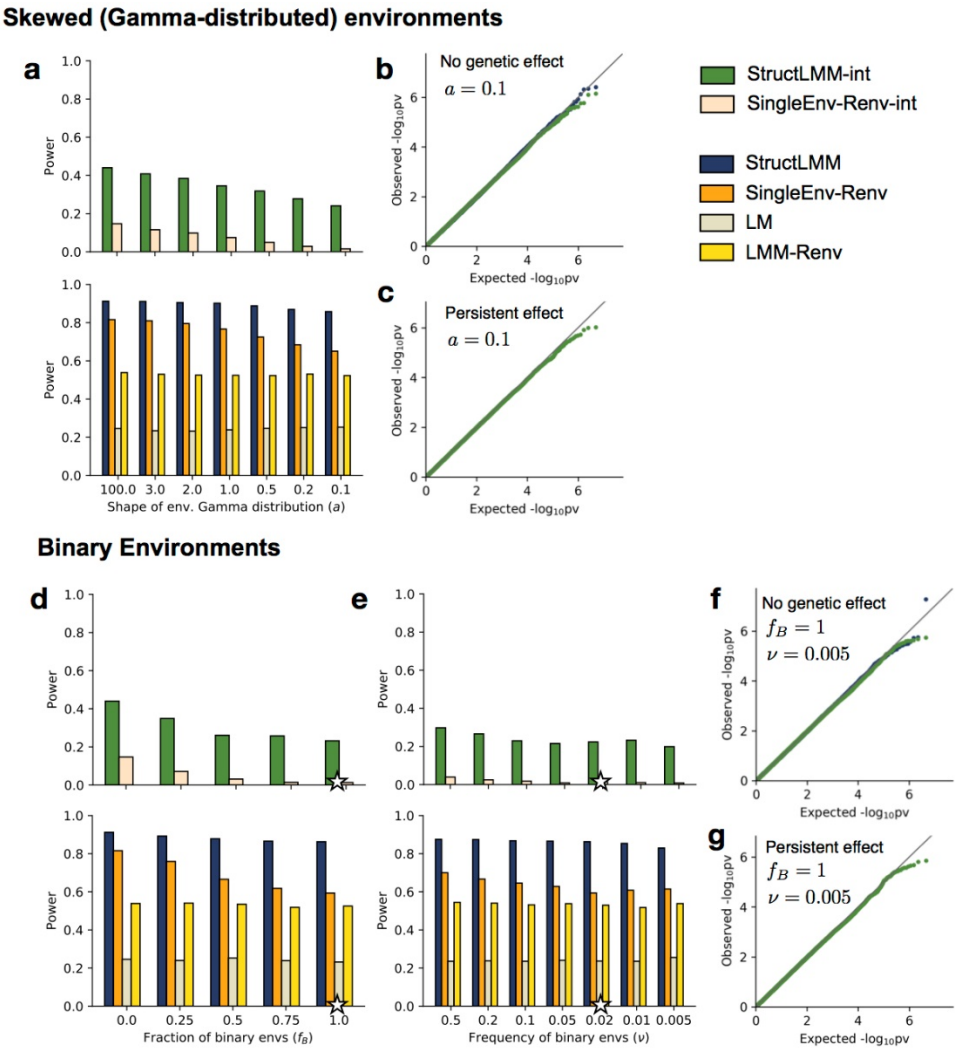
(**b**) Genomic inflation factor λ_{GC} of P values from StructLMM and alternative multi-environment tests based on fixed effects (**Supp. Table 2**), for different numbers of environmental variables (L, $\pi = 100\%$; x-axis) and for increasing sample size (N, top to bottom). Shown are results from StructLMM-int and multi-environment fixed effect interaction tests (column 1), and equivalent association tests (column 2) when no genetic effects are simulated ($v_g = 0$). Column 3 depicts results from StructLMM-int and multi-environment fixed effect interaction tests for simulated persistent genetic effects ($\rho = 0$). Multi-environment fixed effect models using LR tests yielded inflated test statistics (inflation factors λ_{GC}>1) for large numbers of environmental factors in relation to the sample size, whilst score tests yielded deflated statistics (inflation factors λ_{GC}<1) for the corresponding settings. StructLMM was calibrated in all settings. (**c**,**d**) Performance assessment of alternative methods for detecting interactions (top panels) and associations (bottom panels) based on simulated data, using the settings as in **Fig. 2b**,**c** and **Supp. Fig. 2a**,**b** for two sample sizes: $N = 2{,}000$ (**c**) and $N = 5{,}000$ (**d**). Compared were StructLMM-int and alternative multi-environment interaction tests based on fixed effects (**Supp. Table 2**, top panels), StructLMM and additional multi-environment association tests (**Supp. Table 2**, bottom panels). As the fixed effect tests are not always calibrated (see panel **b**), shown are model performance values as assessed by the area under the curve (relative AUC, in the range 0<FPR<0.10, normalised such that 0 corresponds to chance performance and 1 to an ideal model).



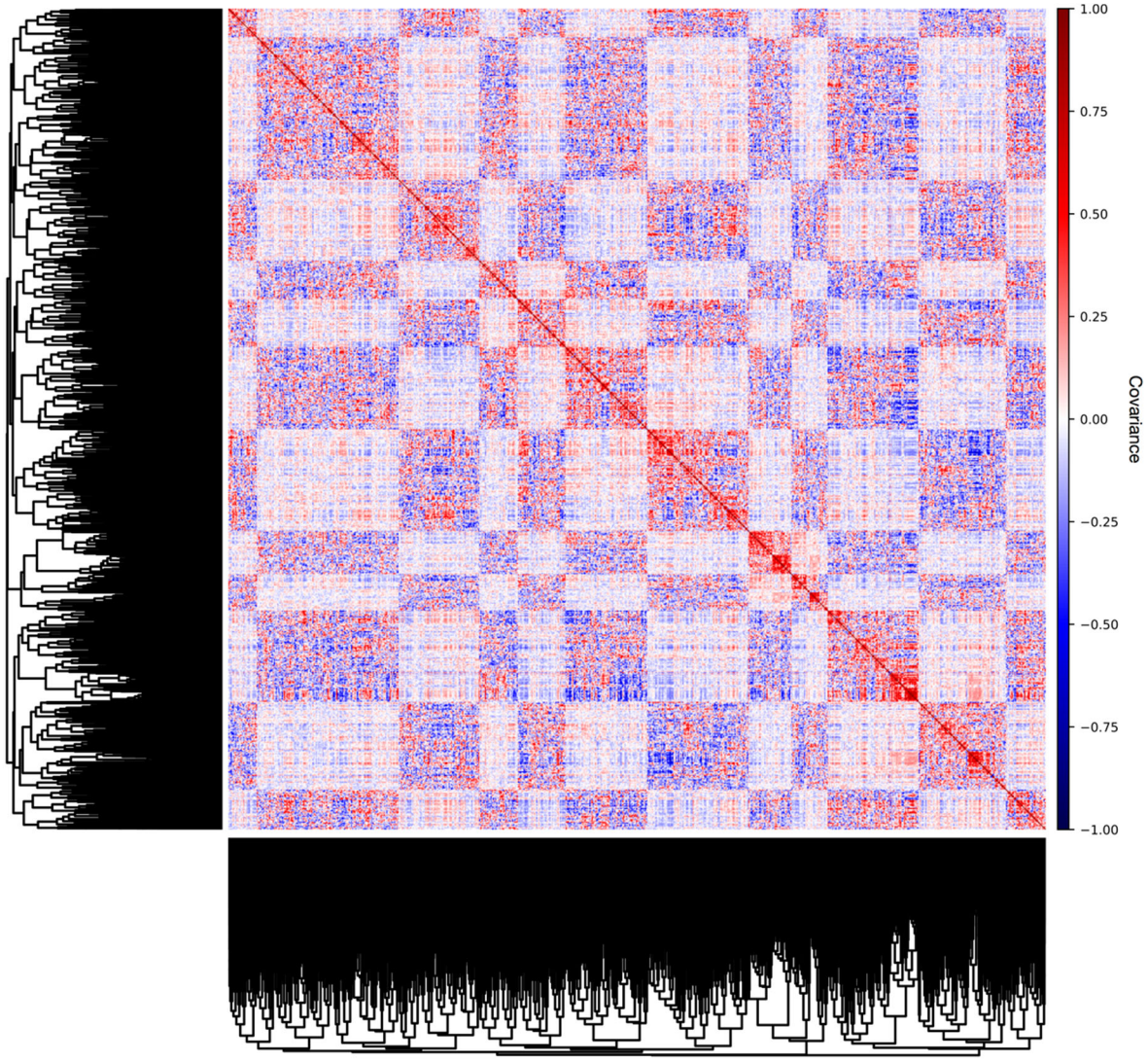
Supplementary Figure 2 | Assessment of power using simulated data. Power comparison of alternative methods for detecting interactions (top panels) and associations (bottom panels) based on simulated data, extending the results as shown in **Fig. 2**, varying (**a**) the fraction of the genetic variance explained by GxE (ρ), (**b**) the number of environments with non-zero GxE effects (π), (**c**) the total number of environments (L, 50% contributing to GxE effects), (**d**) the fraction of variance explained by additive environment effects (v_e) and (**e**) the number of environments that contribute to GxE but are not used (observed, L_{unobs}) for the respective tests. Considered were top panels: the StructLMM interaction test (StructLMM-int) and alternative implementations of single-environment interaction tests (**Supp. Table 2**); bottom panels: StructLMM association test and alternative implementations of 2-df fixed effect tests that jointly tests for persistent associations and interactions with a single environment (**Supp. Table 2**), as well as alternative implementations of linear (mixed) models to test for persistent effects (**Supp. Table 2**). Methods were assessed in terms of power (at FWER<1%) for detecting simulated causal variants (**Methods**). Stars denote default values of genetic parameters, which were retained when varying other parameters (see **Supp. Table 1** & **Methods** for details on the simulation strategy). A synthetic European population of 5,000 individuals based on 1000 Genomes Project genotypes was used for all experiments (**Methods**).



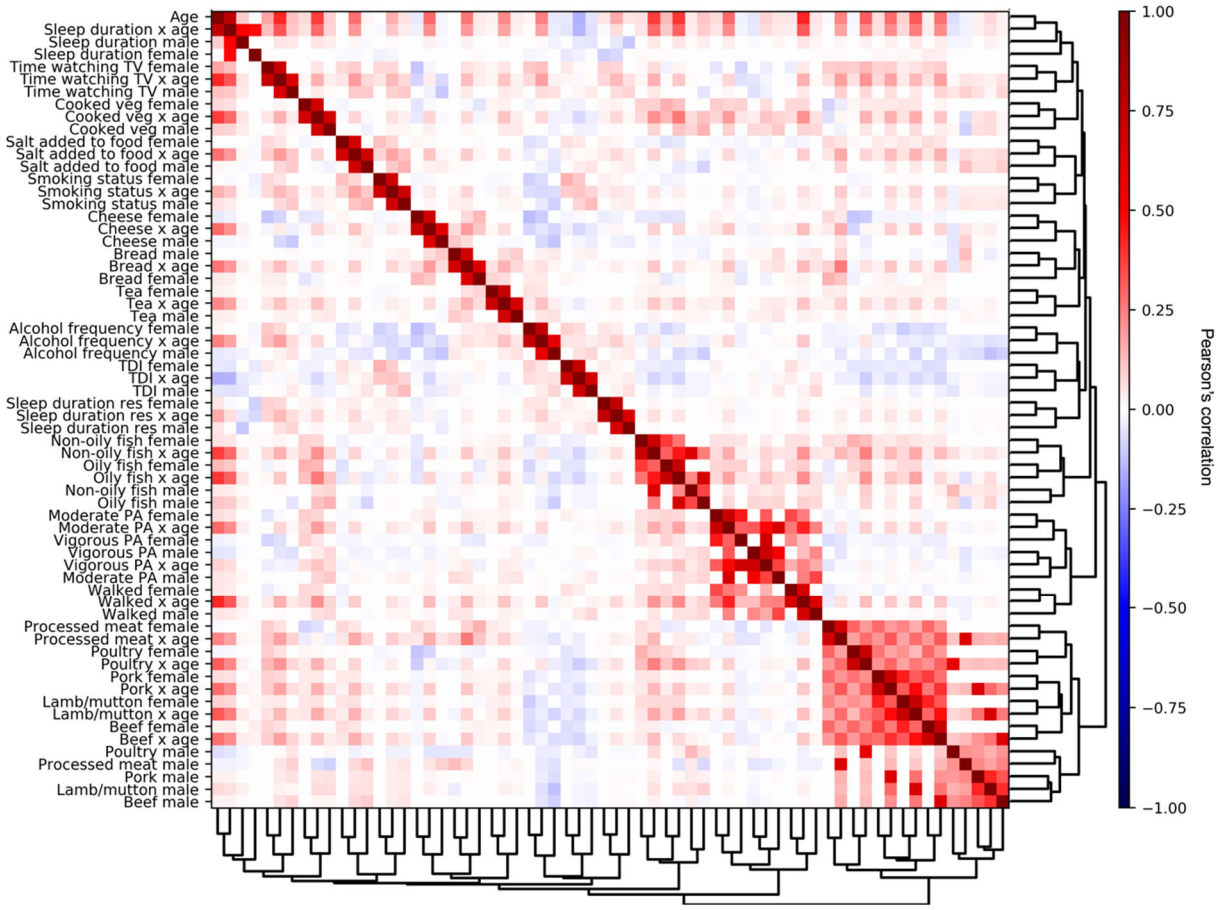
Supplementary Figure 3 | Assessment of power and calibration in the presence of heritable exposures. Power comparison of alternative methods for detecting interactions (top panels) and associations (bottom panels) with simulated genetic effects both on environmental variables and phenotypes. Considered were (**a**) increasing the LD (r^2) between causal variants associated with the environment and those with additive G and GxE effects on the phenotype and (**b**) the average fraction of the exposures variance explained by genetic effects. Stars denote the default value that was retained when varying the other parameter (see **Supp. Table 1**). Top panel: power to detect interactions, considering the StructLMM interaction test (StructLMM-int) and the default implementation of the single-environment interaction tests (SingleEnv-Renv-int, **Supp. Table 2**). Bottom panel: power to detect association, considering StructLMM, the default 2-df fixed effect test that jointly tests for persistent associations and interactions with a single environment (SingleEnv-Renv, **Supp. Table 2**), and linear models to test for persistent effects (LM, LMM-Renv, **Supp. Table 2**). Models were assessed in terms of power (FWER<1%) for detecting simulated causal variants (**Methods**). (**c**) QQ plot of negative log P values obtained from StructLMM-int (green) and StructLMM (blue) for pronounced gene-exposure correlations ($v_x = 0.2$), when no genetic effects are simulated ($v_g = 0$) for 103,527 variants on chromosome 21. (**d**) QQ plot of negative log P values obtained from StructLMM-int (green) when persistent genetic effects were simulated ($\rho = 0$), assuming strong LD between gene-exposure effects and variants associated with phenotype ($v_x = 0.2$, $r^2 = 1$) for 103,527 variants on chromosome 21. Stars denote default values of genetic parameters, which were retained when varying other parameters (see **Supp. Table 1** & **Methods** for details on the simulation strategy). A synthetic European population of 5,000 individuals based on 1000 Genomes Project genotypes was used for all experiments (**Methods**).



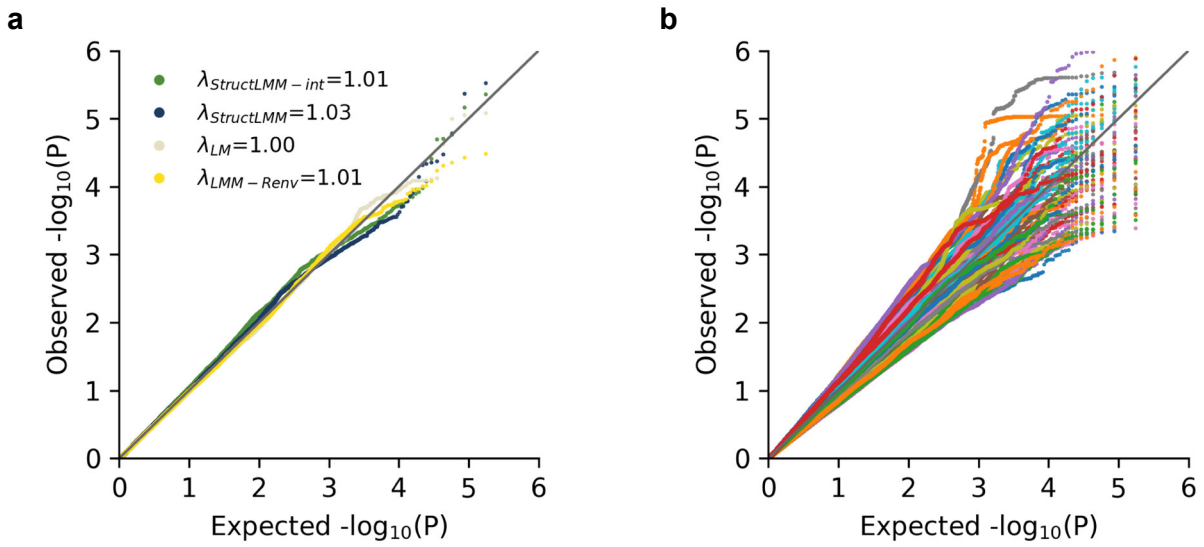
Supplementary Figure 4 | Assessment of power and calibration using simulated data for skewed and binary environments. (**a-c**) Power comparison and calibration when simulating skewed environments drawn from a Gamma distribution. (**a**) Power comparison for varying shape of the gamma distribution (low values correspond to skewed environments, shape 100 correspond to approximately Gaussian distributed environments). (**b**,**c**) QQ plots of negative log P values obtained from StructLMM and StructLMM-int on data with skewed environments for 103,527 variants on chromosome 21, when simulating no genetic effect ($v_g = 0$) (**b**) or when simulating persistent genetic effects (**c,** StructLMM-int only, $\rho = 0$). (**d-g**) Power comparison and calibration for simulated binary environments. (**d**) Power comparison when varying the fraction of binary environments f_B (for constant event frequency ν). (**e**) Power comparison when varying the event frequency of binary environments. Stars denote the default value that was retained when varying the other parameter (see **Supp. Table 1**). (**f**,**g**) QQ plots of negative log P values obtained from StructLMM and StructLMM-int for rare binary environments for 103,527 variants on chromosome 21, either simulating no genetic effect ($v_g = 0$) (**f**) or for simulated persistent genetic effects (**g**, StructLMM-int only, $\rho = 0$). Fixed parameter settings are indicated in the top left corner of the corresponding panels. A synthetic European population of 5,000 individuals based on 1000 Genomes Project genotypes was used for all experiments (**Methods**).



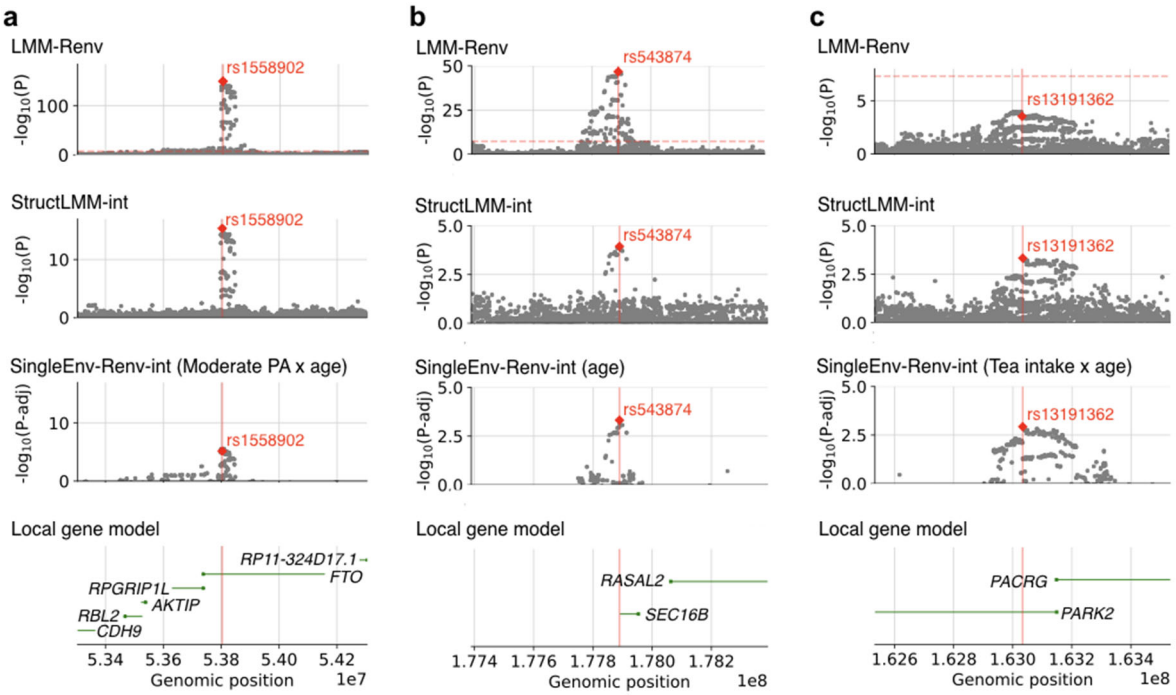
Supplementary Figure 5 | Covariance structure of UK Biobank individuals based on 64 environmental variables. Sample covariance matrix for 5,000 randomly selected individuals, calculated based on 64 environmental variables considered for UK Biobank analyses: 12 diet-related factors, three factors linked to physical activity and a six lifestyle factors, modelled as gender-adjusted and age-adjusted (**Methods**). Dark red denotes pairs of individuals with stronger environmental similarity, whilst blue corresponds to negative covariance of environmental similarity (anti correlation). Blocks of strong correlation/anti correlation, correspond to groups of individuals of the same gender.



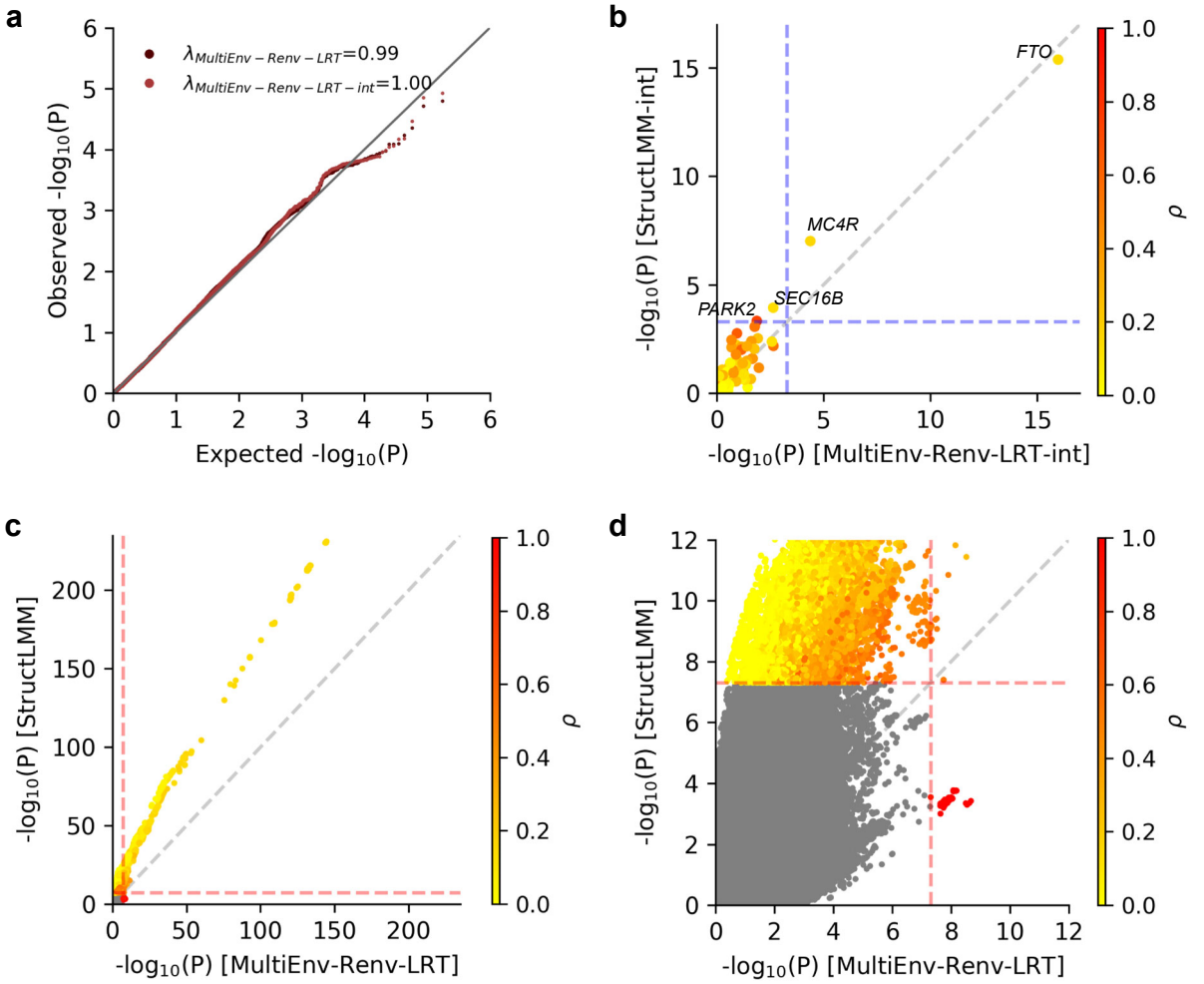
Supplementary Figure 6 | Structure of 64 environmental variables considered for UK Biobank analyses. Shown are correlation coefficients between pairs of environmental variables, considering 12 diet-related factors, three factors linked to physical activity and a six lifestyle factors, modelled as gender-adjusted and age-adjusted (**Methods**). Environments are ordered using hierarchical clustering.



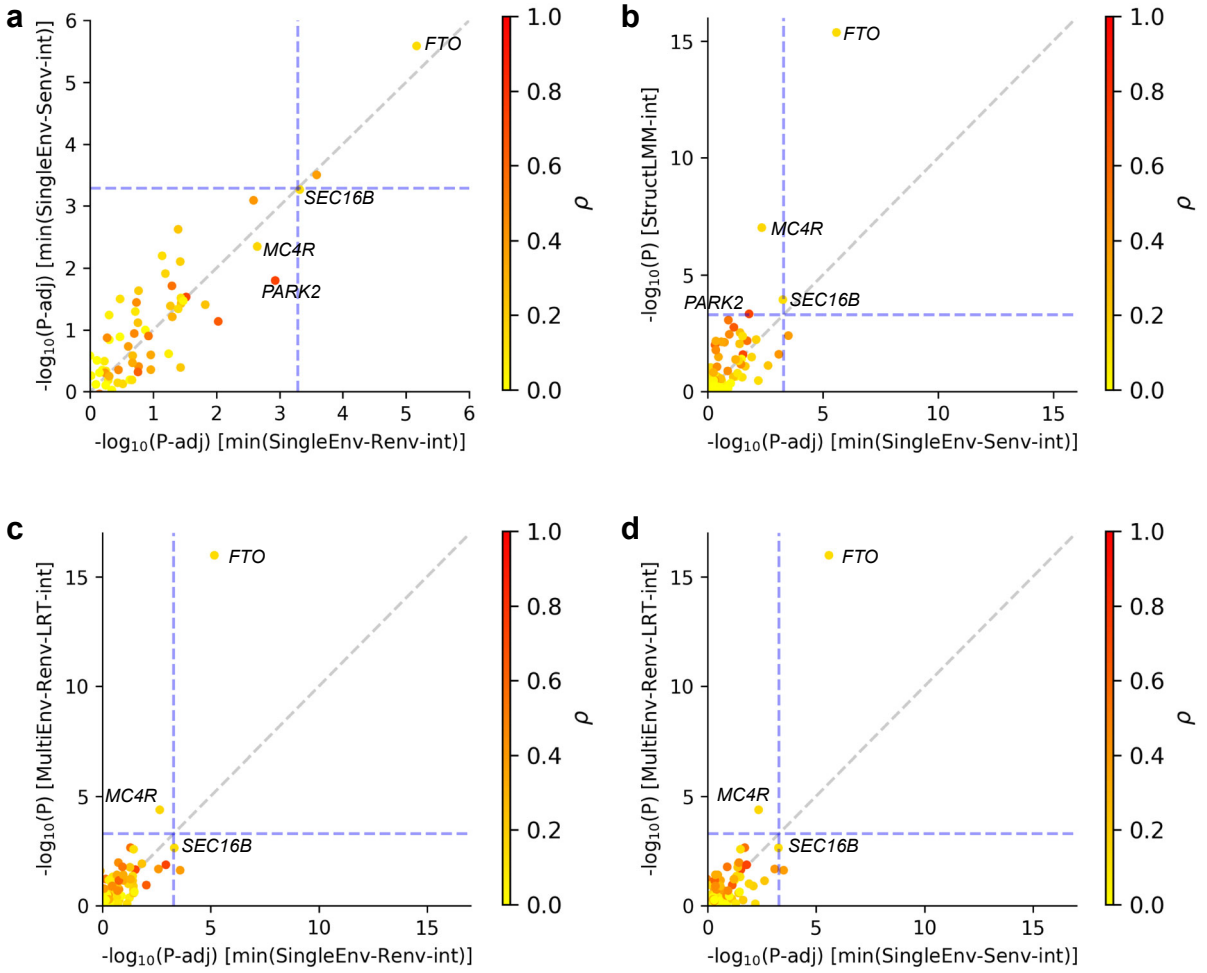
Supplementary Figure 7 | Calibration of interaction and association tests on UK Biobank data. QQ plots of negative log P values from different interaction and association tests applied to UK Biobank BMI phenotype data (n = 252,188 unrelated individuals of European ancestry) based on permuted genetic variants (chromosome 20, 173,297 variants). (**a**) QQ plots of negative log P values for StructLMM-int (green), StructLMM (blue), LM (light grey) and LMM-Renv (yellow). (**b**) QQ plots of negative log P values for SingleEnv-Renv-int, for each of 64 considered environmental variables. LM, LMM-Renv and StructLMM tests were calibrated, whereas the fixed effect interaction tests show variable levels of statistical calibration. See **Supp. Table 2** for an overview of considered methods.



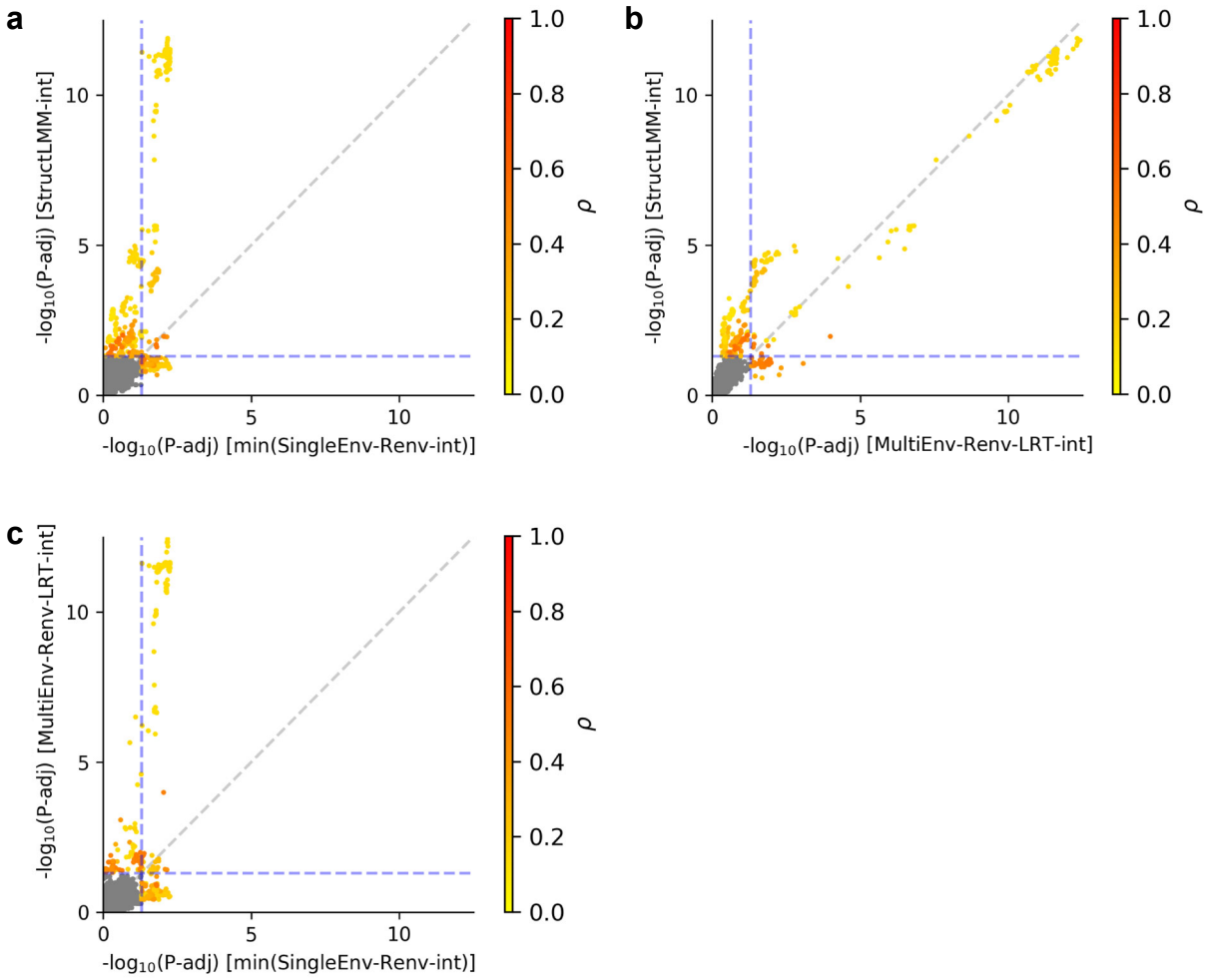
Supplementary Figure 8 | Supplementary results for interactions identified by StructLMM. (**a**-**c**) Local Manhattan plots of interactions identified by StructLMM at *FTO*, *SEC16B* and *PARK2* respectively. From top to bottom: LMM-Renv association test, StructLMM interaction test, SingleEnv-Renv interaction test for the environment with the most significant GxE effect at the respective GIANT variant, local gene models. The red vertical line indicates the position of the GIANT variant as in **Fig. 3a**.



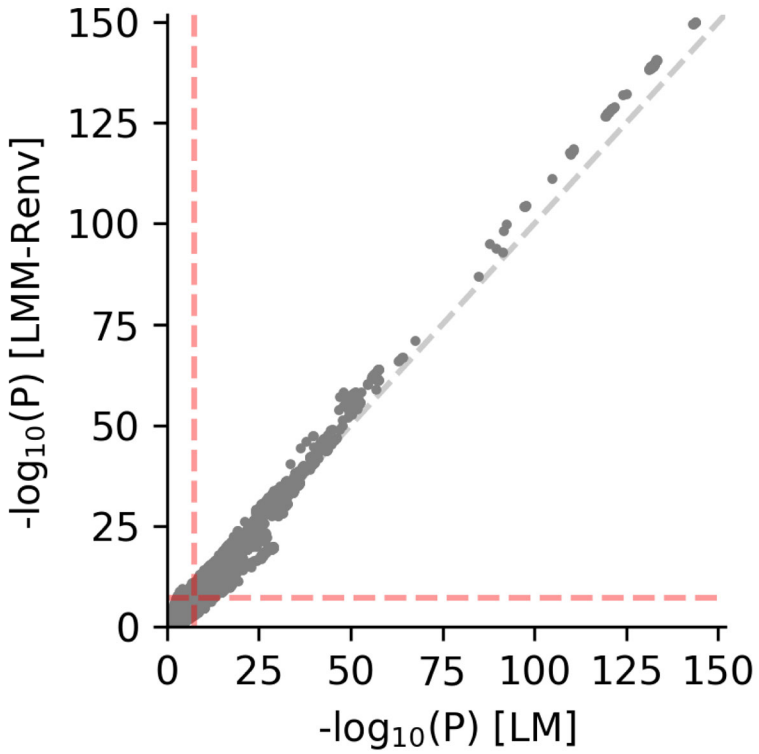
Supplementary Figure 9 | Comparison of multi-environment GxE tests based on fixed effects on UK biobank data. (**a**) QQ plots of negative log P values from a multi-environment fixed effect model to test for associations while accounting for heterogeneity in effect sizes due to GxE (MultiEnv-Renv-LRT) and an interaction test (MultiEnv-Renv-LRT-int) applied to UK Biobank BMI phenotype data and permuted genetic variants (chromosome 20, 173,297 variants), which are calibrated for this sample size (n = 252,188). (**b**) Scatter plot of negative log P values from GxE interaction tests at 97 GIANT variants (Locke *et al*., 2015), considering the MultiEnv-Renv-LRT-int interaction test (x-axis) versus the StructLMM-int test (y-axis). Dashed lines correspond to α<0.05, Bonferroni adjusted for the number of tested GIANT variants and colour denotes the estimated fraction of genetic variance due to GxE (fitted parameter ρ). StructLMM-int and MultiEnv-Renv-LRT-int identified four and two significant loci respectively. (**c**) Scatter plot of negative log P values from the MultiEnv-Renv-LRT association test (x-axis) versus the StructLMM association test (y-axis). Dashed lines indicate genome-wide significance at P<5x10^{-8} and colour denotes the estimated fraction of genetic variance due to GxE (fitted parameter ρ), with (**d**) displaying a zoom-in view of variants close to genome-wide significance. StructLMM and MultiEnv-Renv-LRT identify 17,630 and 2,037 significant variants respectively. All displayed results are generated using 252,188 unrelated individuals of European ancestry.



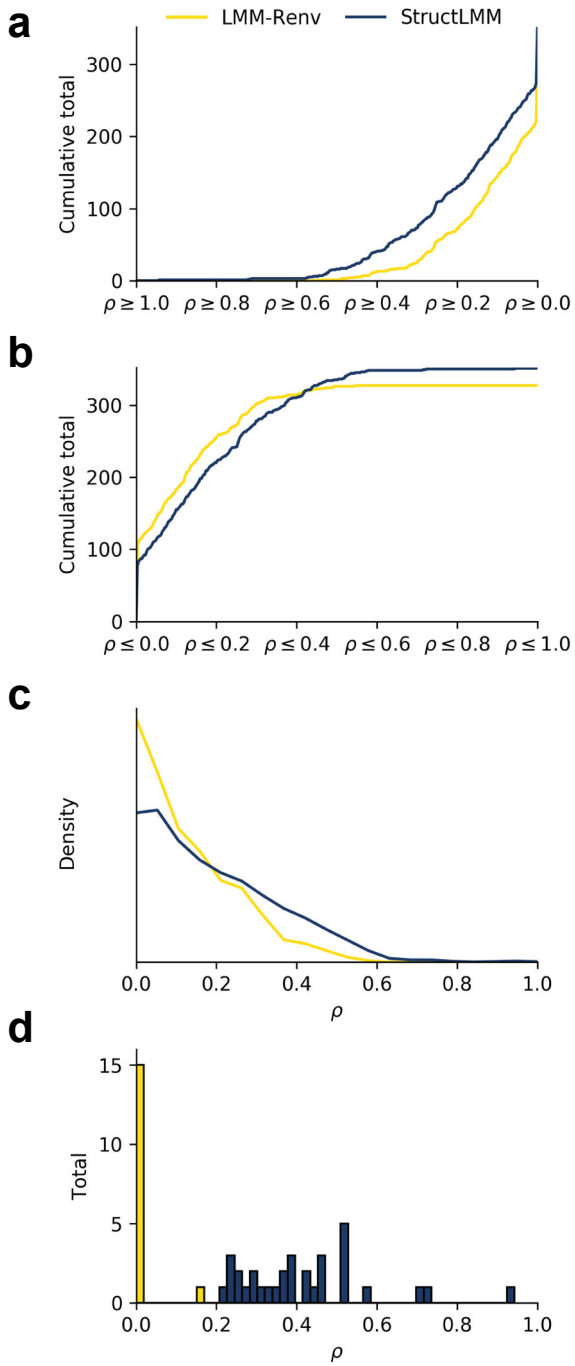
Supplementary Figure 10 | Comparison of different single and multi-environment interaction test based on fixed effects on UK Biobank data. Compared are two alternative single-environment tests (SingleEnv-Renv-int, SingleEnv-Senv-int), as well as a multi-environment test (MultiEnv-Renv-LRT-int) based on fixed effects and StructLMM-int at 97 GIANT variants (Locke *et al.*, 2015). Scatter plot of negative log P values of (**a**) fixed effect tests, either accounting for additive environment effects of all environments (SingleEnv-Renv-int, using random effects, x-axis) or of the single environment that is tested (SingleEnv-Senv-int, using fixed effects, y-axis), (**b**) fixed effect tests, that account for additive environmental effects of the single environment tested (SingleEnv-Senv-int, x-axis) versus StructLMM-int (y-axis), (**c**) single-environment fixed effect interaction test with an additive environmental random effect component (SingleEnv-Renv-int, x-axis) versus the multi-environment fixed effect interaction test (MultiEnv-Renv-LRT-int, y-axis) and (**d**) single-environment fixed effect interaction test with a fixed effect additive environment term accounting for the single environment that is tested for GxE effects under the alternative (SingleEnv-Senv-int, x-axis) versus the multi-environment fixed effect interaction test (MultiEnv-Renv-LRT-int, y-axis). Dashed lines correspond to $\alpha<0.05$ and colour denotes the estimated fraction of genetic variance due to GxE (fitted parameter ρ). For single-environment models (SingleEnv-Renv-int, SingleEnv-Senv-int), shown is the Bonferroni adjusted minimum P value across the set of tested environments (P-adj). SingleEnv-Renv-int and SingleEnv-Senv-int identify the same significant loci, but different significant loci to those identified by MultiEnv-Renv-LRT-int. All displayed results are generated using 252,188 unrelated individuals of European ancestry.



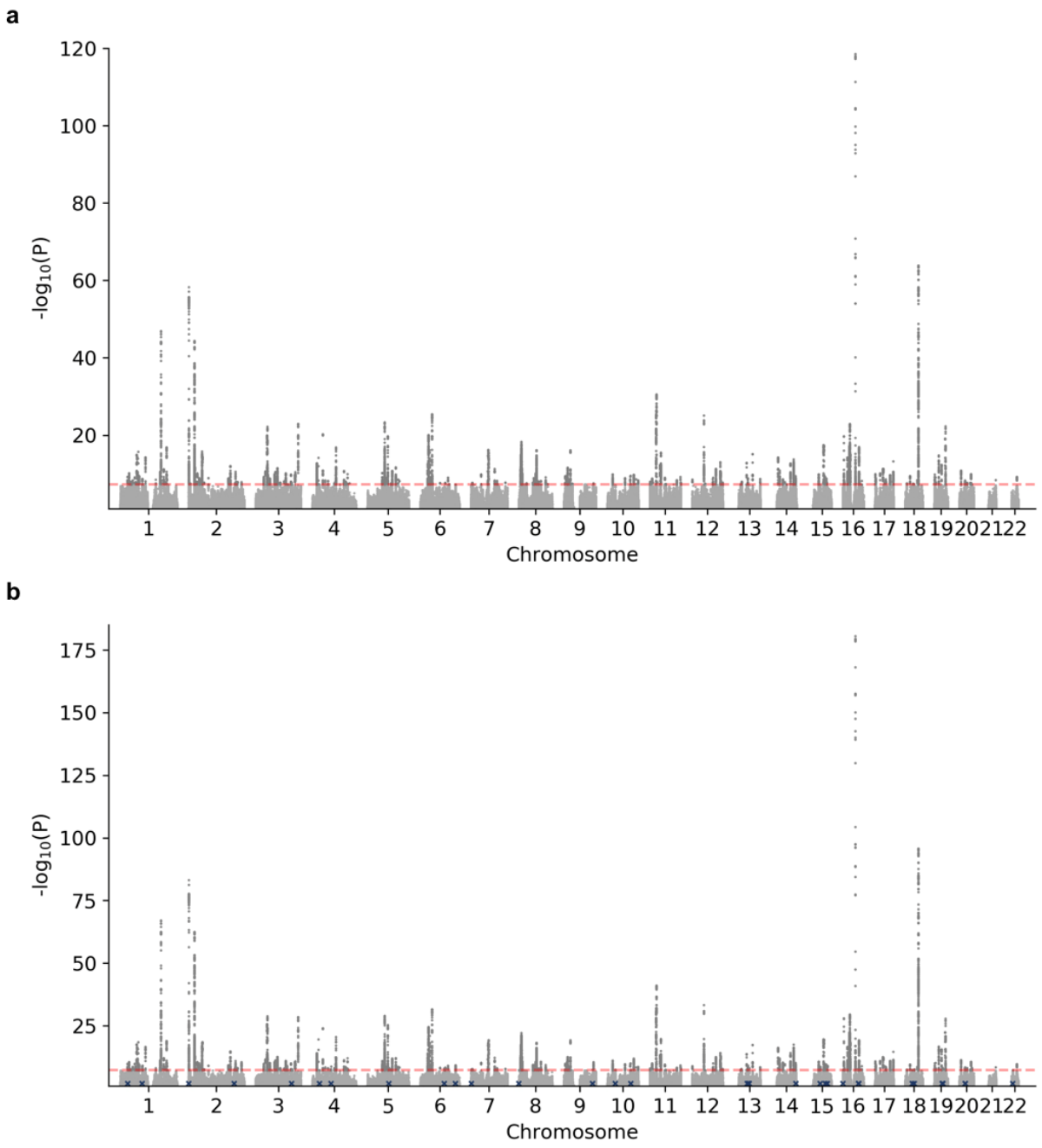
Supplementary Figure 11 | Comparison of interaction test results for genome-wide variants associated with BMI on UK Biobank data. Comparison of the StructLMM interaction test (StructLMM-int), a single-environment interaction test (SingleEnv-Renv-int) as well as multi-environment fixed effect interaction test (MultiEnv-Renv-LRT-int), at 17,606 variants that were significantly associated with BMI based on an LMM (LMM-Renv, $P<5x10^{-8}$). Scatter plots of negative log Benjamini-Hochberg adjusted P values (P-adj) between (**a**) SingleEnv-Renv-int (x-axis) versus StructLMM-int (y-axis), (**b**) MultiEnv-Renv-LRT-int (x-axis) versus StructLMM-int (y-axis) and (**c**) SingleEnv-Renv-int (x-axis) versus MultiEnv-Renv-LRT-int (y-axis). Dashed lines correspond to P-adj<0.05 and colour denotes the estimated fraction of genetic variance due to GxE (fitted parameter ρ based on StructLMM). For the single-environment test (SingleEnv-Renv-int), shown is the Bonferroni adjusted minimum P value across the tested environments (P-adj). StructLMM-int identified 23 loci with GxE, followed by SingleEnv-Renv-int (11 loci) and MultiEnv-Renv-LRT-int (9 loci; FDR<5%, Benjamini-Hochberg adjusted, LD clumped loci, $r^2<0.1$ within 500kb, **Methods**; see **Supp. Table 3** for full results). All displayed results are generated using 252,188 unrelated individuals of European ancestry.



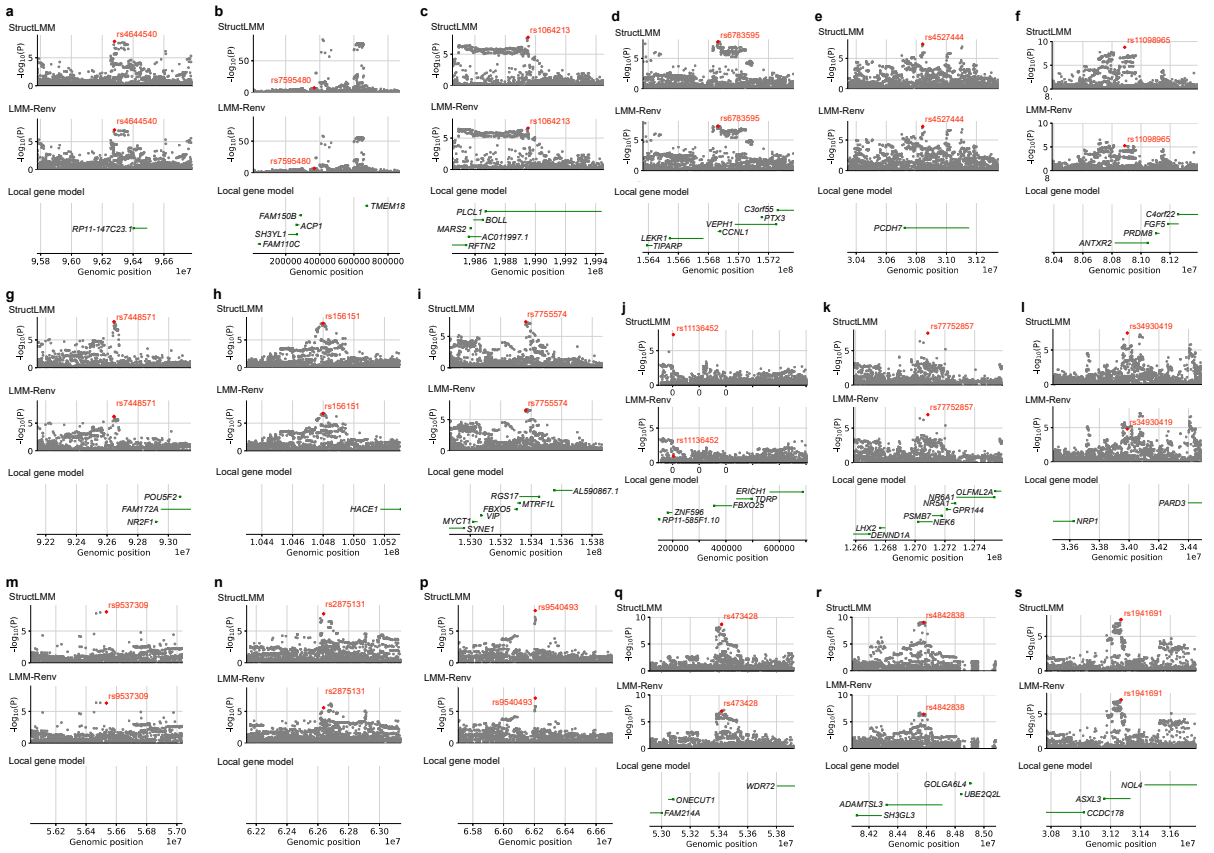
Supplementary Figure 12 | **Comparison of LMM and LM results on UK Biobank data.** Scatter plot of genome-wide negative log P values from the LM association test, without accounting for additive environmental effects (x-axis), versus an LMM association test that accounts for additive environmental effects using the same random effect component as used in StructLMM (LMM-Renv, y-axis). Dashed lines indicate genome-wide significance at $P<5x10^{-8}$. All displayed results are generated using 252,188 unrelated individuals of European ancestry.



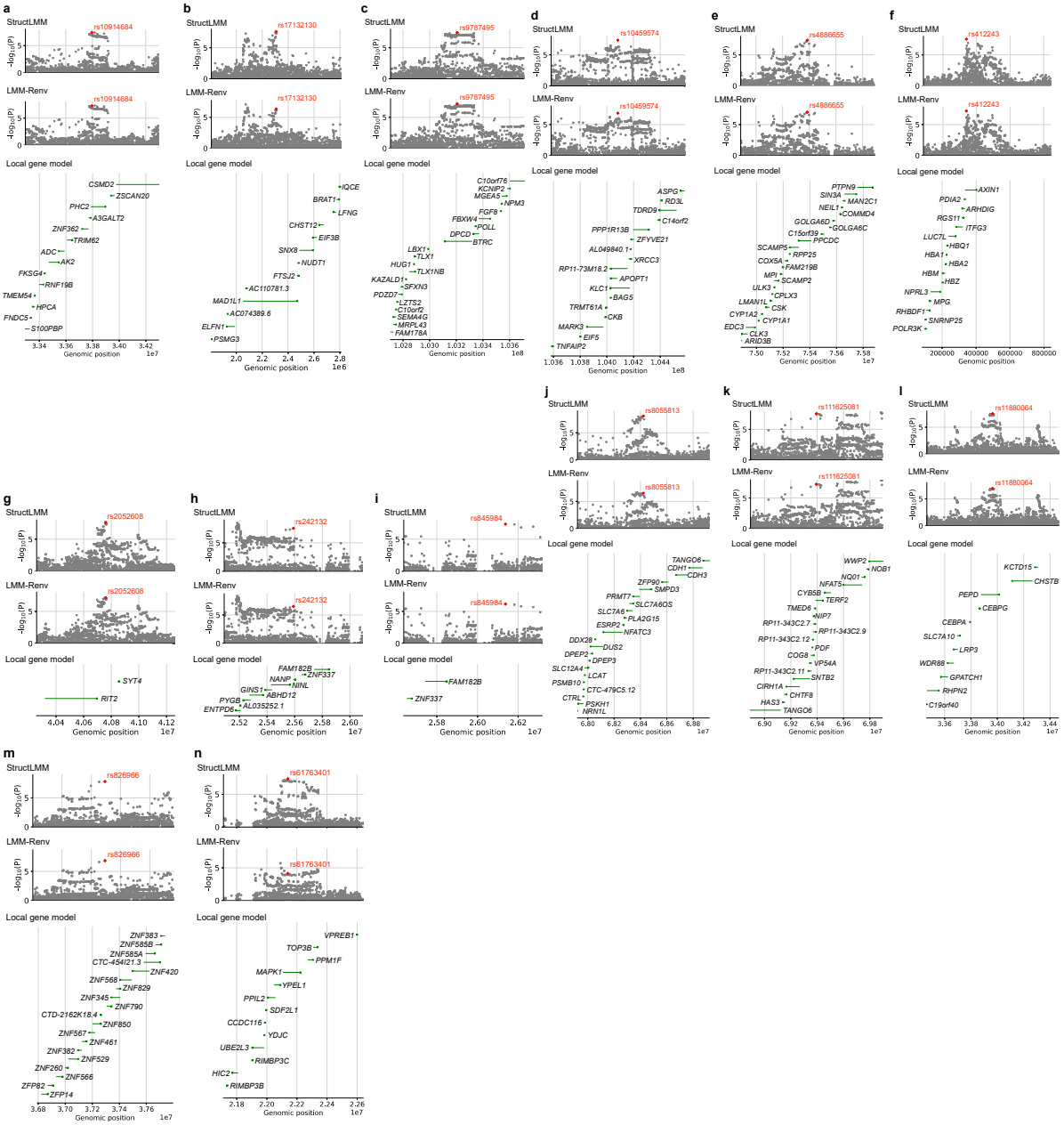
Supplementary Figure 13 | Distribution of the estimated extent of GxE for significant loci identified by the StructLMM association test and LMM-Renv on UK Biobank data. Cumulative number of significant associations ($P<5x10^{-8}$, LD clumped loci, $r^2<0.1$ within 500kb, **Methods**) identified by LMM-Renv (yellow, N=327 loci) and StructLMM (blue, N=351 loci) in decreasing (**a**) and increasing (**b**) order of the estimated extent of GxE (fitted parameter ρ). (**c**) Estimated density of the fraction of genetic variance due to GxE (ρ) for loci identified by LMM-Renv (yellow) and StructLMM (blue). (**d**) Histogram of the fraction of genetic variance due to GxE (ρ), considering the subset of loci exclusively identified by either approach (**Methods**): LMM-Renv (yellow, total 16), StructLMM (blue, total 32). Loci identified by StructLMM tended to have a greater extent of GxE, whereas loci that were LMM-Renv specific tended to have no or little evidence for effect size heterogeneity due to GxE. The latter can be explained by the fact that StructLMM will be slightly less powered than the LMM-Renv when no or very little GxE is present since StructLMM is penalised for testing multiple values ρ (see also **Supp. Fig. 2**).



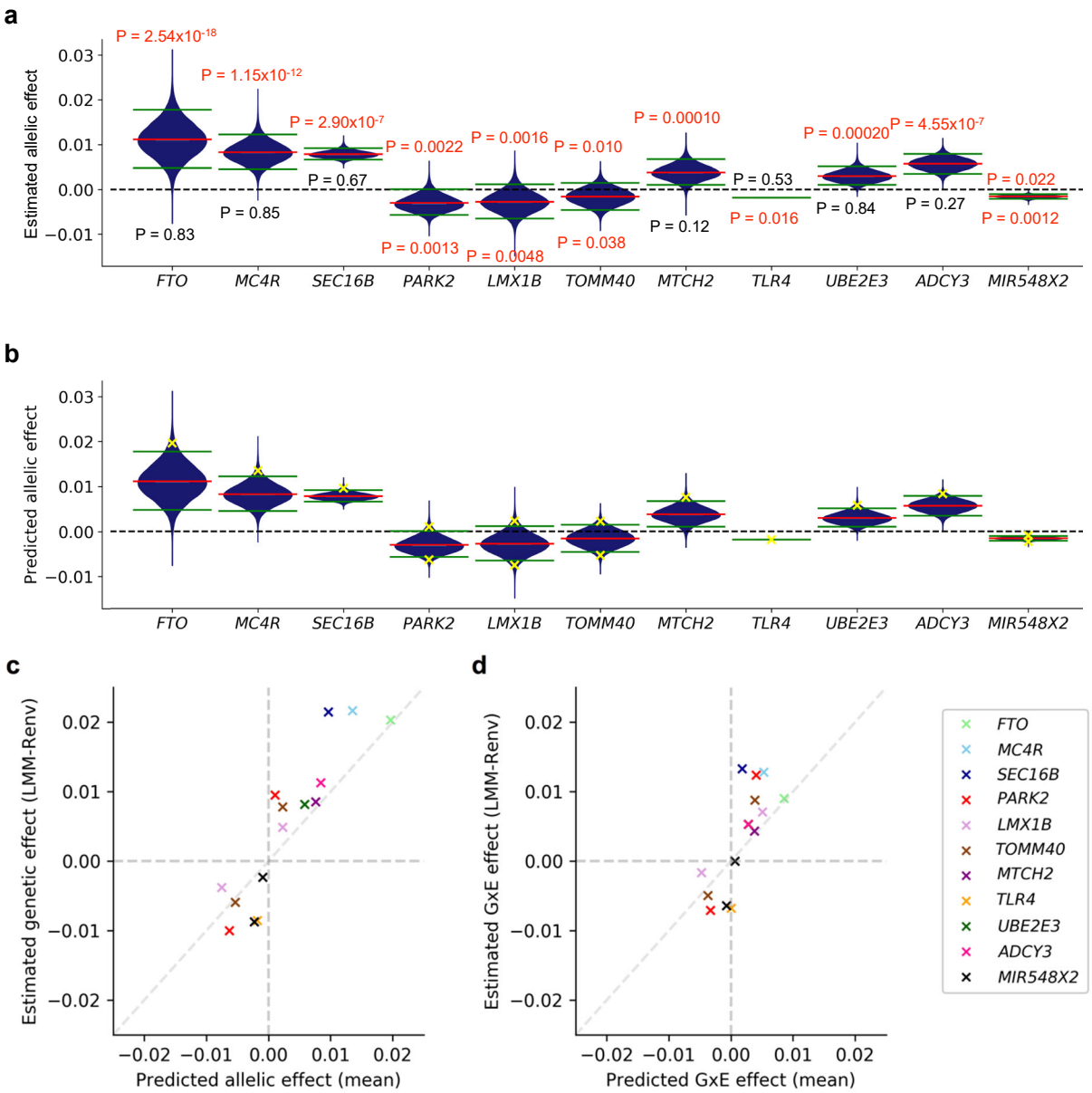
Supplementary Figure 14 | Results from association tests applied to UK Biobank data. Manhattan plots, showing negative log P values obtained from (**a**) LMM-Renv and (**b**) the StructLMM association test. Additional loci identified by StructLMM ($P<5x10^{-8}$) as in **Supp. Table 4** are highlighted with a blue cross. The dashed red line denotes the genome-wide significance threshold ($P<5x10^{-8}$). All displayed results are generated using 252,188 unrelated individuals of European ancestry.



Supplementary Figure 15 | Local Manhattan plots for additional association loci identified by StructLMM versus LMM-Renv on UK Biobank data. Shown are additional associations as in **Supp. Table 4**, with the red vertical line indicating the position of the StructLMM lead variant. From top to bottom: Manhattan plot of negative log P values from StructLMM, LMM-Renv and the local gene models for (**a**-**s**) *RP11-147C23.1*, *FAM150B*, *PLCL1*, *CCNL1*, *PCDH7*, *ANTXR2*, *NR2F1*, *HACE1*, *RGS17*, *ZNF596*, *NEK6*, *NRP1*, rs9537309 (no gene within +/-500 kb), rs2875131 (no gene within +/-500 kb), rs9540493 (no gene within +/-500 kb), *ONECUT1*, *ADAMTSL3* and *ASXL3* respectively. Associations were assigned to genes based on the nearest protein coding gene within +/-500 kb respectively. The maximum LD (r^2) between significant StructLMM variants in the loci exclusive to StructLMM under consideration and significant LMM-Renv variants in other loci within the +/-500 kb region is 0.052, 0.00049 for panels (**b** and **d**) respectively. All displayed results are generated using 252,188 unrelated individuals of European ancestry.

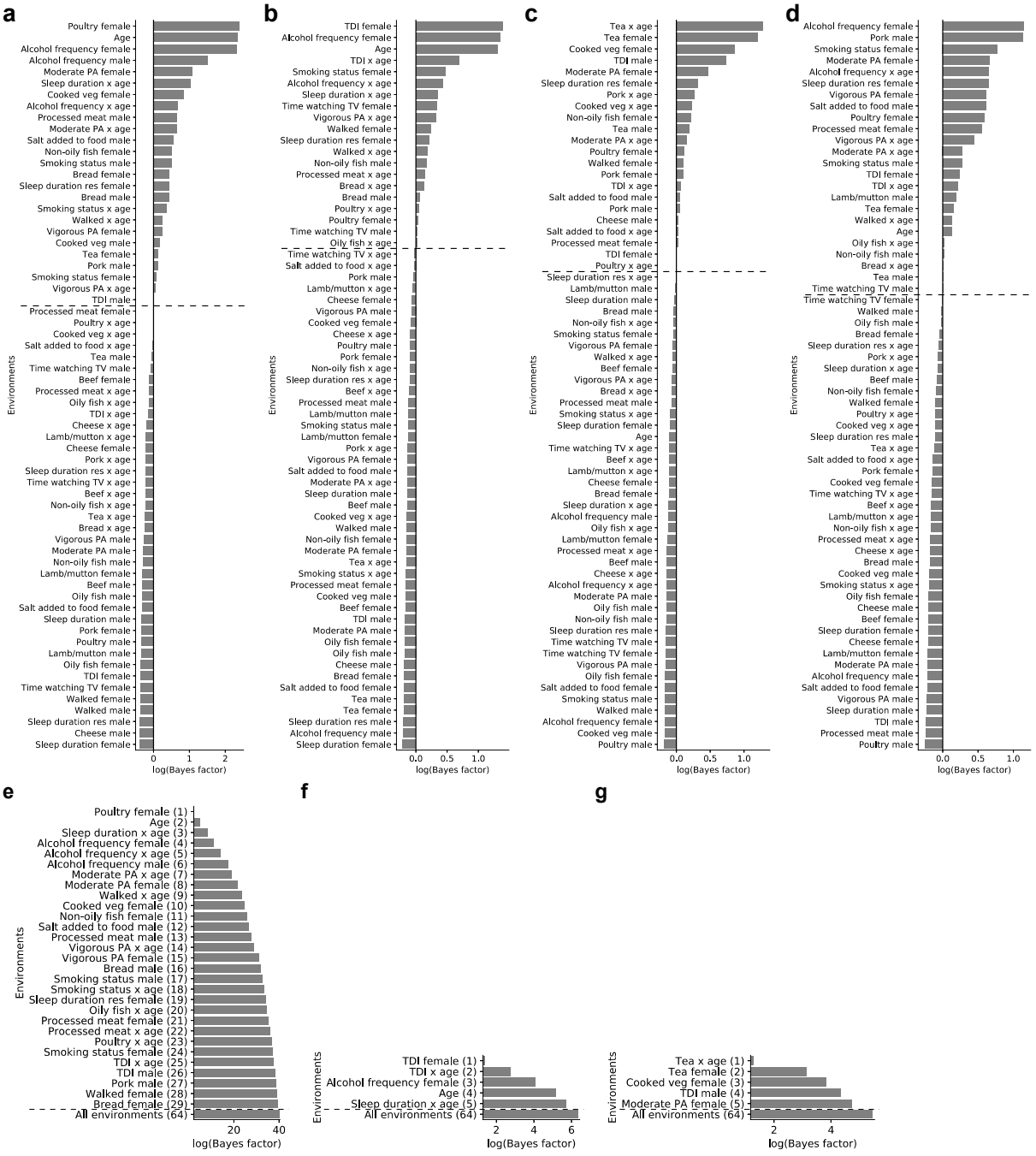


Supplementary Figure 16 | Local Manhattan plots for additional association loci identified by StructLMM versus LMM-Renv on UK Biobank data. Shown are additional associations as in **Supp. Table 4**, with the red vertical line indicating the position of the StructLMM lead variant. From top to bottom: Manhattan plot of negative log P values from StructLMM, LMM-Renv and the local gene models. (**a**-**n**) *PHC2*, *MAD1L1*, *BTRC*, *RP11-73M18.2*, *PPCDC*, *AXIN1*, *RIT2*, *NANP*, *FAM182B*, *SMPD3*, *TERF2*, *PEPD*, *ZNF790* and *MAPK1* respectively. Loci were assigned to genes based on the nearest protein coding gene within +/-500 kb respectively. The maximum LD (r^2) between significant StructLMM variants in the loci exclusive to StructLMM under consideration and significant LMM variants in other loci within the +/-500 kb region is 0.054 for panel (**h**). All displayed results are generated using 252,188 unrelated individuals of European ancestry.

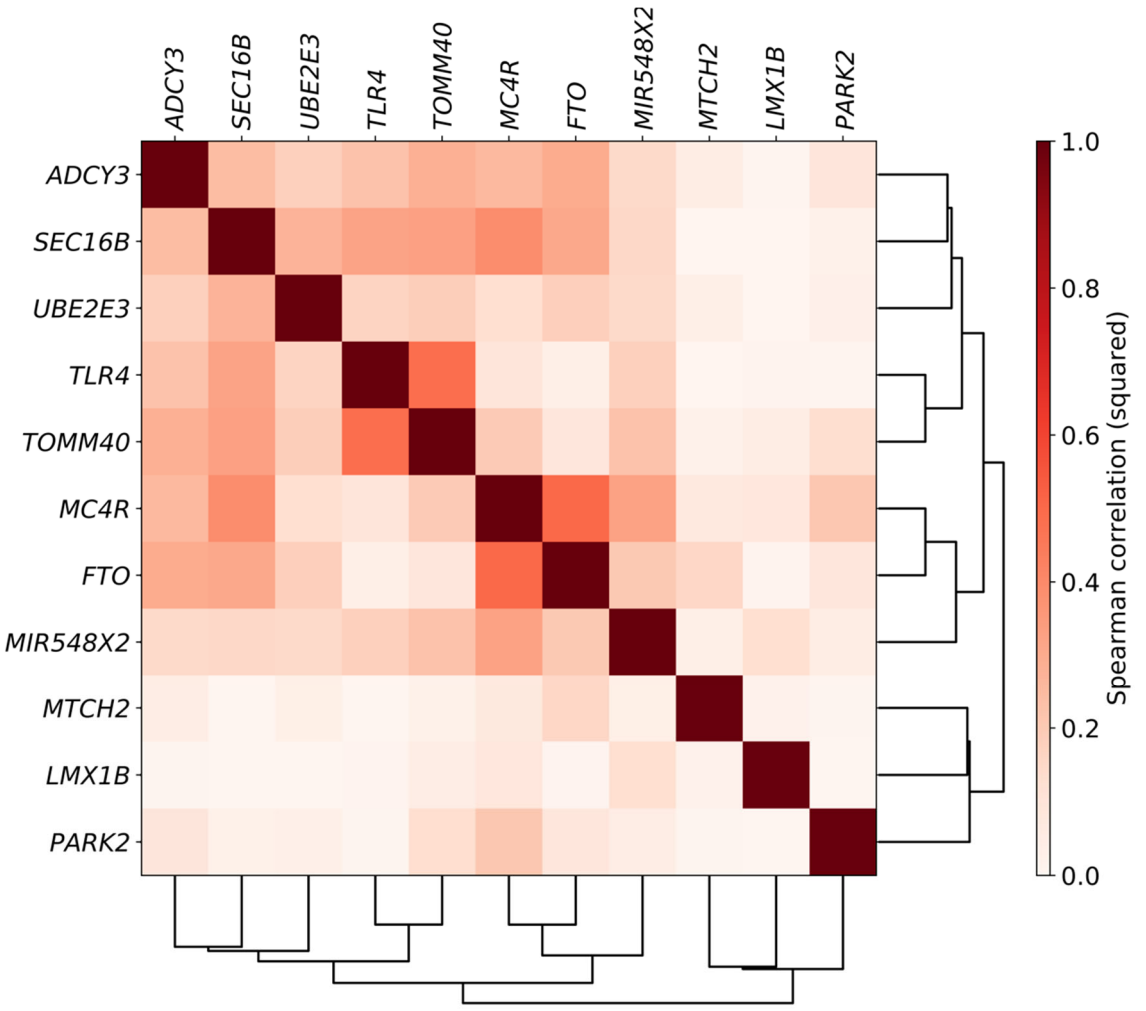


Supplementary Figure 17 | Out of sample prediction of per-individual allelic effect sizes. Assessment of per-individual allelic effects for 11 GIANT variants with evidence for GxE (lenient Benjamini-Hochberg adjustment, FDR<0.05, **Supp. Table. 3**), considering a 50:50 split of the cohort into training and test fractions. (**a**) Violin plots, displaying the estimated density of individuals in the training fraction (n = 126,094) that exert an allelic effect of a particular size given the distribution of the environmental factors within the population (in sample estimates as in **Fig. 4a**; see **Methods**). Mean and the top and bottom 5% strata of the effect size distribution are indicated by the red and green bars, respectively. P values denote the significance of genetic effects assessed using LMM-Renv (not accounting for GxE) using individuals within the respective strata. Nominally significant associations (P<0.05, two-sided LR test) highlighted in red. (**b**) Analogous allelic effect size distribution as in **a**, displaying predictions in the test fraction (out-of-sample predictions, based on environmental states of individuals in the test fraction (n = 126,094) but without using BMI phenotypes; **Methods**). Yellow crosses denote the mean predicted genetic effect within the top and bottom 5% strata (considering strata with nominally significant associations in training fractions, i.e. P values that are highlighted in red in **a**). (**c**) Scatter plot of allelic effect sizes, predicted out of sample

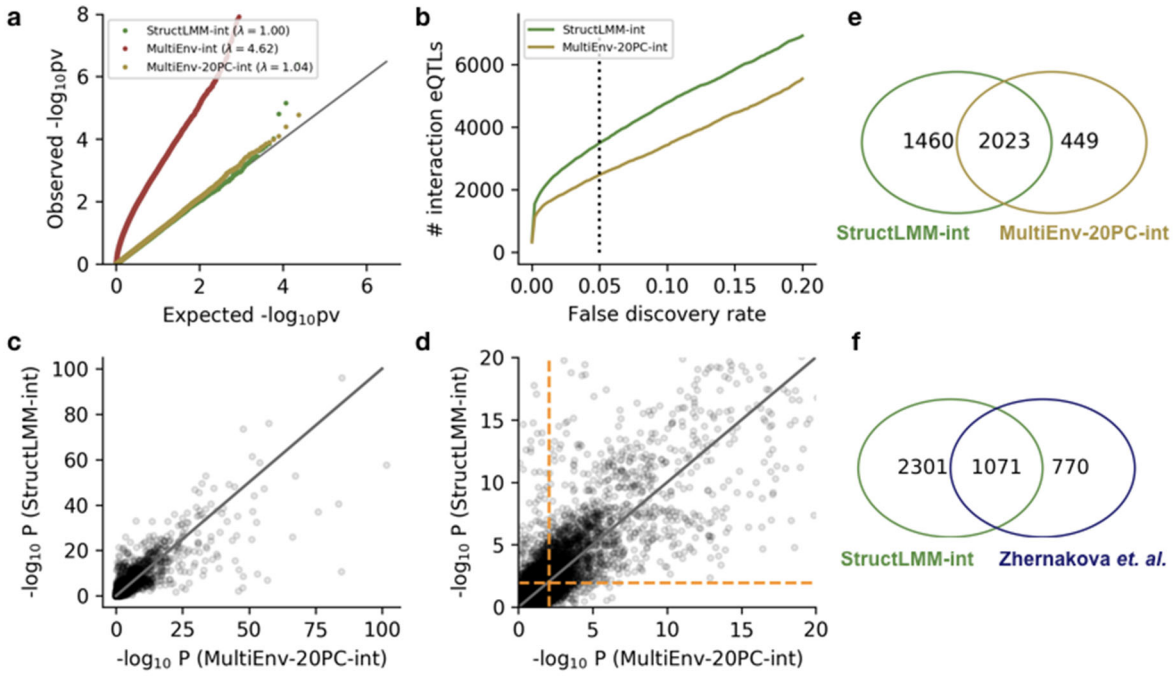
for test set individuals (yellow crosses in **b**, x-axis) versus within-sample estimates obtained from LMM-Renv based on test set individuals in the top and bottom 5% strata (not accounting for GxE, using the genotypes and phenotypes of the test set individuals in the strata). Different GIANT variants are coded in colour. (**d**) Analogous scatter plot as in **c**, however displaying the differences between genetic effects in the 5% strata and population estimates of persistent effects (effect sizes due to GxE).



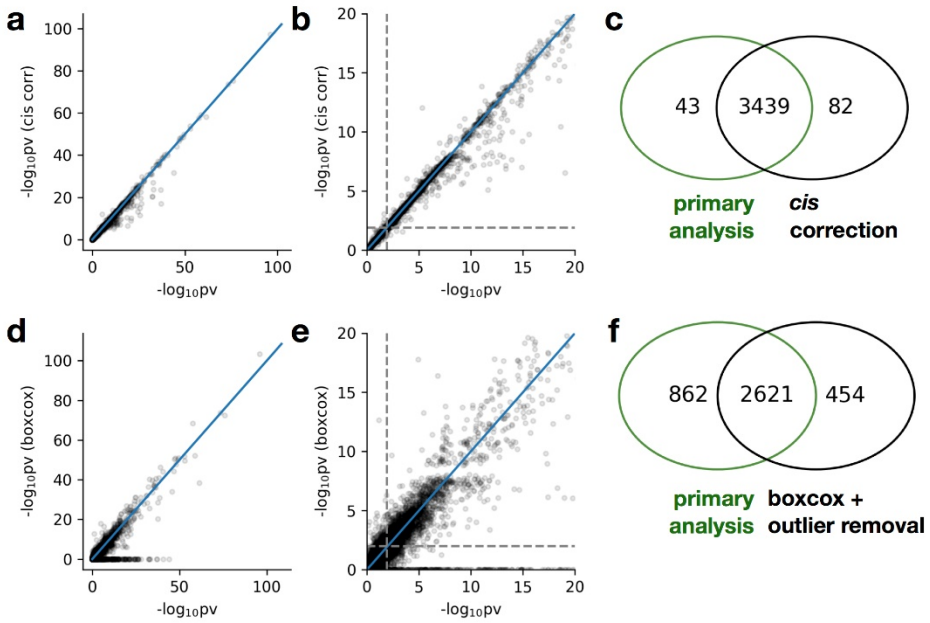
Supplementary Figure 18 | Exploration of putative driving environments at the identified GxE interaction effects UK Biobank data. Relevance of individual environmental variable for GxE effects at four loci, showing Bayes factors (BF) between models containing all 64 environments and models with a single environmental variable removed. Shown are results for (**a**) *FTO*, (**b**) *SEC16B*, (**c**) *PARK2* and (**d**) *MC4R*, ordered by Bayes factor, with positive evidence above the dashed lines whilst those with negative evidence lie below. (**e**-**g**) Cumulative evidence of individual environmental variables contributing to GxE at *FTO*, *SEC16B* and *PARK2,* showing Bayes factors between the full model and models with increasing numbers of environmental variables removed using (greedy) backward elimination (**Methods**). For comparison, shown is the total evidence of all environmental variables. The additional environment that is removed at each elimination step is labelled on the y-axis with the total number of environmental variables removed in considered models shown in brackets.



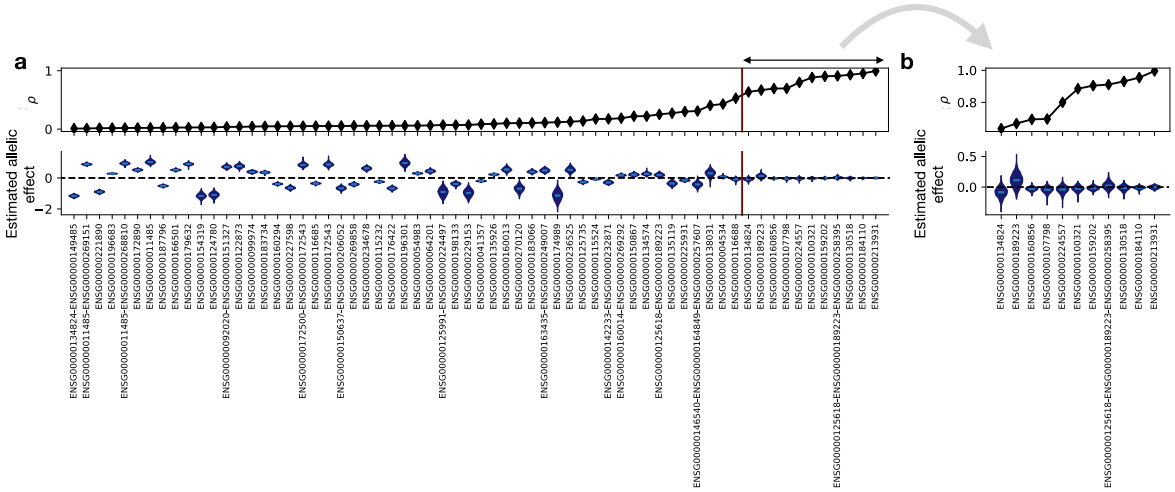
Supplementary Figure 19 | Rank correlation of per-individual genetic effect sizes across loci for UK Biobank data. Shown are squared Spearman correlation coefficients of per-individual (n = 252,188) allelic effects estimates for 11 GIANT variants with evidence for GxE (more lenient Benjamini-Hochberg adjustment, FDR<0.05, **Supp. Table 3**).



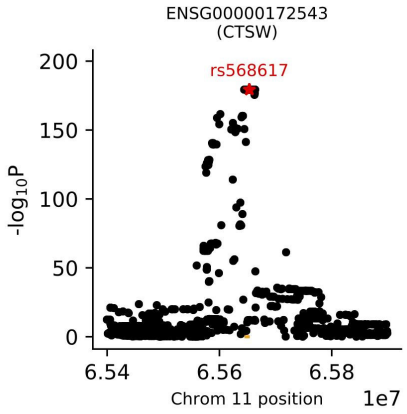
Supplementary Figure 20 | Statistical calibration and power of alternative interaction tests applied to the blood eQTL dataset. Compared were the StructLMM interaction test (StructLMM-int) and multi-environment fixed effect tests with as many degrees of freedom as environmental variables (MultiEnv-Renv-LRT-int, labelled MultiEnv-int in the figure), as well as a multi-environment fixed effect test based on 20 principal components of the environmental variables (MultiEnv-20PC-Renv-LRT-int, labelled MultiEnv-20PC-int). Shown are results from applying all methods to test for cell-context interactions at 23,506 lead eQTL variants (n = 2,040; **Methods**). (**a**) QQ plot of negative log P values obtained on permuted genotype data. Whereas P values from StructLMM-int and MultiEnv-20PC-Renv-LRT-int were calibrated, MultiEnv-Renv-LRT-int yielded inflated P values, most likely owing to the large number of degrees of the freedom, and hence this model was not considered further. (**b**) Number of interaction eQTL discovered by StructLMM-int and MultiEnv-20PC-Renv-LRT-int as a function of the FDR threshold (Benjamini-Hochberg adjustment). (**c**,**d**) Scatter plot of negative log P values from StructLMM-int versus MultiEnv-20PC-Renv-LRT-int with a zoom in-view shown in panel **d**. Dashed orange lines correspond to the 5% FDR thresholds. (**e**,**f**) Overlap of the sets of interaction eQTL identified by StructLMM-int and MultiEnv-20PC-Renv-LRT-int (**e,** FDR<5%; considering all 23,506 tested variants) and interaction eQTL identified in the primary analysis of the data (**f,** Zhernakova *et al.*, FDR<5%; considering 17,952 shared lead eQTL between studies).



Supplementary Figure 21 | Assessment of the robustness of interaction effects on the blood eQTL data to data pre-processing. Compared are the StructLMM-int results from the primary analysis and analogous results obtained using alternative data processing strategies (n = 2,040). First, to assess the potential of spurious associations due to strong gene-exposure associations, we considered a GxE analysis where for each analysed gene, the additive effect of the lead genetic variant was regressed out from all proxy genes used as environments prior to interaction testing (environments were rank-inverse transformed a second time; **Methods**). (**a**) Scatter plot of the negative log P values from the original (x-axis) and the analysis with environments that were adjusted for *cis* genetic effects (cis corr, y-axis). (**b**) Zoom-in view and (**c**) venn diagram of significant interaction effects (FDR<5%). (**d**-**f**) Analogous results considering boxcox normalisation followed by removal of outlying samples (gene expression levels that exceeded 2.5 standard deviations) as an alternative pre-processing strategy. (**d**-**f**) Analogous scatter plots and analysis of overlap as shown in **a**-**c**. Both control analyses recovered a large fraction of the primary interaction eQTL (98.8% and 75.2% respectively), indicating that the interactions identified are sufficiently robust.



Supplementary Figure 22 | Distribution of estimated per-individual allelic effect sizes at interaction eQTL that colocalise with GWAS variants. (**a**) Upper panel: estimated fraction of genetic variance due to GxE effects (ρ). Bottom panel: violin plots displaying the distribution of allelic effects for each locus across the environmental profiles in the population (n = 2,040). Blue lines in the violin plots show the median values of the distribution. (**b**) Zoom-in of **a**, depicting data from the eleven associations with the largest fraction of genetic variance attributable to GxE effects.



Supplementary Figure 23 | **eQTL association for *CTSW*.** Manhattan plot of negative log P values obtained from *cis* eQTL mapping of *CTSW* (n = 2,040). The red star indicates the lead eQTL variant, annotated with the identifier of the colocalising GWAS variant (*rs568617*, r^2=1.00 associated to Crohn's disease); Manhattan plots for other putative colocalisation events are provided in **Supp. Dataset 1.**

Supplementary Note

Contents

1. The StructLMM model

1.1. Model definition

A conventional linear mixed model used for association testing can be cast as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \underbrace{\boldsymbol{x}\beta_{\mathrm{G}}}_{\mathrm{G}} + \boldsymbol{u} + \boldsymbol{\psi}, \tag{1}$$

where $\boldsymbol{y}$ denotes the $N \times 1$ phenotype vector for N individuals, $\boldsymbol{x}$ denotes the $N \times 1$ genotype vector of the focal variant, $\boldsymbol{X} \in \mathbb{R}^{N \times K}$ the fixed effect design matrix of K covariates and $\boldsymbol{b} \in \mathbb{R}^{K \times 1}$ their effect sizes. The variable $\boldsymbol{u}$ is a random effect and can be used to account for additional additive effects, such as population structure, environment or other additive (confounding) factors and $\boldsymbol{\psi}$ denotes iid noise. The random effect component $\boldsymbol{u}$ and the noise vector $\boldsymbol{\psi}$ are assumed to follow multivariate normal distributions

$$\begin{aligned} \boldsymbol{u} &\sim \mathcal{N}\left(\boldsymbol{0}, \sigma_e^2 \boldsymbol{\Sigma}_{\boldsymbol{u}}\right) & (2) \\ \boldsymbol{\psi} &\sim \mathcal{N}\left(\boldsymbol{0}, \sigma_n^2 \boldsymbol{I}_N\right), & (3) \end{aligned}$$

where the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{u}} \in \mathbb{R}^{N \times N}$ is the covariance sample structure of the random effect. Under this conventional linear model, the focal variant $\boldsymbol{x}$ is assumed to have a persistent genetic effect on all samples included in the analysis.

StructLMM extends the conventional linear mixed model by including an additional per-individual effect term that accounts for G×E, which can be represented as an $N \times 1$ vector, $\boldsymbol{\beta}_{\mathrm{G}\times\mathrm{E}}$. The model can be cast as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \underbrace{\boldsymbol{x}\beta_{\mathrm{G}}}_{\mathrm{G}} + \underbrace{\boldsymbol{x} \odot \boldsymbol{\beta}_{\mathrm{G}\times\mathrm{E}}}_{\mathrm{G}\times\mathrm{E}} + \underbrace{\boldsymbol{u}}_{\mathrm{E}} + \boldsymbol{\psi}, \tag{4}$$

where $\odot$ denotes the element-wise (Hadamard) product. The per-individual effect size vector $\boldsymbol{\beta}_{\mathrm{G}\times\mathrm{E}}$ is modelled as random effect, following the multivariate normal distribution

$$\boldsymbol{\beta}_{\mathrm{G}\times\mathrm{E}} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma^2_{\mathrm{G}\times\mathrm{E}} \boldsymbol{\Sigma}\right), \tag{5}$$

where the covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{N \times N}$ parameterises how per-individual effects covary across individuals and is calculated as a function of observed environmental variables $\boldsymbol{\Sigma} \equiv \boldsymbol{\Sigma}(\boldsymbol{E}) \in \mathbb{R}^{N \times N}$, where $\boldsymbol{E}$ is the $N \times L$ matrix of L observed environments.

Note that for the special case $\sigma^2_{\mathrm{G}\times\mathrm{E}} = 0$, this model reduces to a standard linear mixed model for genetic association testing. In StructLMM, the random effect component $\boldsymbol{u}$ is used to account for additive environmental effects. While in general different covariance functions could be considered for additive environmental effects and interactions, we assume $\boldsymbol{\Sigma}_{\boldsymbol{u}} \equiv \boldsymbol{\Sigma}$ for simplicity. Specific choices for the environmental covariance $\boldsymbol{\Sigma}(\boldsymbol{E})$ are discussed in Section 1.2.

Marginal Likelihood For parameter inference, we consider the marginalised form of the model in Eq (4), which is obtained by integrating over the G×E effects $\boldsymbol{\beta}_{\mathrm{G\times E}}$ and the random effect component $\boldsymbol{u}$

$$\boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{X}\boldsymbol{b} + \underbrace{\boldsymbol{x}\beta_{\mathrm{G}}}_{\mathrm{G}}, \underbrace{\sigma^2_{\mathrm{G\times E}}\mathrm{diag}(\boldsymbol{x})\boldsymbol{\Sigma}\mathrm{diag}(\boldsymbol{x})}_{\mathrm{G\times E}} + \underbrace{\sigma^2_e\boldsymbol{\Sigma}}_{\mathrm{E}} + \underbrace{\sigma^2_n\boldsymbol{I}}_{\mathrm{noise}}\right). \tag{6}$$

Here, we used the identity $\boldsymbol{x} \odot \boldsymbol{\beta}_{\mathrm{G\times E}} = \mathrm{diag}(\boldsymbol{x})\boldsymbol{\beta}_{\mathrm{G\times E}}$ in Eq (4), with $\mathrm{diag}(\boldsymbol{x})$ denoting the $N \times N$ diagonal matrix whose diagonal is $\boldsymbol{x}$.

1.2. Definition of the environment covariance

In principle, StructLMM can be applied with any valid covariance function [1] of the observed environments. In this work, we only consider linear covariance functions, defined on a potentially large number of continuous and discrete environmental variables

$$\boldsymbol{\Sigma}(\boldsymbol{E}) = \boldsymbol{E}\boldsymbol{E}^T. \tag{7}$$

The use of a linear covariance function was primarily motivated by two appealing properties. First, as the number of samples typically exceeds the number of environments in larger cohorts ($L << N$), the linear covariance will be low-rank, enabling parameter inference with a computational complexity that scales linearly with the cohort size (Section 1.4.3). Second, a linear covariance is directly interpretable as there is a one-to-one correspondence between StructLMM and multivariate linear regression using L covariates to account for the interaction term (see Section 1.3). Note that, depending on the choice of $\boldsymbol{E}$, the linear covariance can be used to (i) model group-specific effects, which correspond to the setting where $\boldsymbol{\Sigma}$ is a block diagonal matrix with blocks corresponding to the different groups, or per-individual effect sizes (see **Fig. 1b-c**) and (ii) account for non-linear relationships between the observed environmental variables by combining simple observed environments to create more complex ones (similar to the use of basis functions; see below for more details).

Normalisation of environmental variables A standard strategy to normalise the variables for defining a linear covariance is to consider standardised features rescaled by $\frac{1}{\sqrt{L}}$. This normalisation ensures that the resulting random effect has sample mean 0 and sample variance 1 [2]. Other normalisation procedures, including row standardisation, such that the resulting random effect has per-individual variance 1 or is a correlation matrix, are also valid [1]. Different normalisation procedures correspond to slightly different assumptions regarding the variance explained by different environments within the interaction and additive environment term, similar to building linear covariances using genetic factors [3].

Accounting for non-linear relationships between environmental variables Non-linear effects can be accounted for by introducing additional transformed environmental variables, similar to using basis functions for non-linear regression. For example,

to model possible dependencies of G×E and additive environmental effects of L environments $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_L$ with a categorical variable $\boldsymbol{c} \in \{0,1\}^{N\times 1}$ (e.g. gender), one can define $\boldsymbol{\Sigma}$ using the extended set of environments $\boldsymbol{E} = [\boldsymbol{c} \otimes \boldsymbol{e}_1, \ldots, \boldsymbol{c} \otimes \boldsymbol{e}_L, (\boldsymbol{1}_N - \boldsymbol{c}) \otimes \boldsymbol{e}_1, \ldots, (\boldsymbol{1}_N - \boldsymbol{c}) \otimes \boldsymbol{e}_L] \in \mathbb{R}^{N\times 2L}$. This approach is used to define a gender-adjusted and age-adjusted environmental covariance.

1.3. Derivation as marginalised linear G × E interaction model

When using a linear environmental covariance, StructLMM can be interpreted as a random-effect implementation of the multi-environment linear interaction model where persistent genetic and G×E effects are modelled as fixed effects. This relationship is similar to the well-known equivalence between multivariate Bayesian linear regression and a linear mixed model (e.g. [4]). Briefly, denoting with $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_L$ the set of observed $N \times 1$ vectors for L environmental variables, the linear model underlying StructLMM can be cast as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \underbrace{\boldsymbol{x}\beta_G}_{\text{G}} + \underbrace{\sum_{l=1}^{L}(\boldsymbol{x} \odot \boldsymbol{e}_l)\beta_l}_{\text{G×E}} + \underbrace{\sum_{l=1}^{L}\boldsymbol{e}_l\alpha_l}_{\text{E}} + \boldsymbol{\psi}, \tag{8}$$

where $\boldsymbol{\psi} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma_n^2 \boldsymbol{I}_N\right)$. Defining prior variances on G×E and additive environmental effects

$$\beta_l \sim \mathcal{N}(0, \sigma^2_{\text{G×E}}), \tag{9}$$
$$\alpha_l \sim \mathcal{N}(0, \sigma_e^2), \tag{10}$$

and marginalising over β_l and α_l results in the marginalised StructLMM model in Eq. (6). It is of note that both the association and interaction tests in StructLMM can also be directly implemented in the fixed effect framework, without marginalization. These alternative tests are considered in **Supp. Fig. 1b**, where we observe that fixed effect tests are not always calibrated and are underpowered in some settings. See Section 3.2.1 for full details on the implementation of alternative methods that are compared to StructLMM.

1.4. Interaction and joint association tests

We define two variance component score tests based on the StructLMM model, (i) an interaction test to identify loci with G×E and (ii) an association test that accounts for heterogeneity in genetic effect sizes due to G × E (jointly testing the G and G×E effects).

1.4.1. Interaction test

Using the marginalised model in Eq. 6, a test for G×E interactions corresponds to the alternative hypothesis $\sigma^2_{\text{G×E}} > 0$. We define an efficient score-based test that enables the calculation of P values with a complexity that scales linearly in the number of

individuals, provided that the environment covariance $\boldsymbol{\Sigma}$ is low rank. We note that the null of the interaction test reduces to a standard linear mixed model with a low-rank covariance matrix for additive random effects, for which existing efficient inference strategies can be reused [5].

Score test for interaction test In the model in Eq (4-5), the score-test statistics can be computed analogously to the procedure described in [6]

$$\begin{aligned} Q &= \frac{1}{2}\boldsymbol{y}^T\boldsymbol{P}\boldsymbol{K}_1\boldsymbol{P}\boldsymbol{y} \\ &= \frac{1}{2}\boldsymbol{y}^T\boldsymbol{P}\left(\mathrm{diag}(\boldsymbol{x})\boldsymbol{\Sigma}\mathrm{diag}(\boldsymbol{x})\right)\boldsymbol{P}\boldsymbol{y} \\ &= \frac{1}{2}\boldsymbol{y}^T\boldsymbol{P}\underbrace{\left[\mathrm{diag}(\boldsymbol{x})\boldsymbol{E}\right]}_{\boldsymbol{W}}\left[\mathrm{diag}(\boldsymbol{x})\boldsymbol{E}\right]^T\boldsymbol{P}\boldsymbol{y} \\ &= \frac{1}{2}\left\|\boldsymbol{W}^T\boldsymbol{P}\boldsymbol{y}\right\|^2. \end{aligned} \tag{11}$$

where we have defined

$$\begin{aligned} &\boldsymbol{K}_1 = \mathrm{diag}(\boldsymbol{x})\boldsymbol{\Sigma}\mathrm{diag}(\boldsymbol{x}), \\ &\boldsymbol{W} = \mathrm{diag}(\boldsymbol{x})\boldsymbol{E}, \\ &\boldsymbol{P} = \boldsymbol{H}_0^{-1} - \boldsymbol{H}_0^{-1}[\boldsymbol{X},\boldsymbol{x}]\left([\boldsymbol{X},\boldsymbol{x}]^T\boldsymbol{H}_0^{-1}[\boldsymbol{X},\boldsymbol{x}]\right)^{-1}[\boldsymbol{X},\boldsymbol{x}]^T\boldsymbol{H}_0^{-1}. \end{aligned} \tag{12}$$

The matrix $\boldsymbol{H}_0$ denotes the total covariance matrix estimated under the null model

$$\boldsymbol{H}_0 = \hat{\sigma}_e^2\boldsymbol{\Sigma} + \hat{\sigma}_n^2\boldsymbol{I}, \tag{13}$$

where $\hat{\sigma}_e^2$ and $\hat{\sigma}_n^2$ correspond to the (null model) maximum likelihood estimates (MLE) of σ_e^2 and σ_n^2. It can be shown that Q follows a mixture of χ^2 distributions [6, 7]

$$Q \sim \sum_k a_k\chi_1^2, \tag{14}$$

where the vector of the coefficients $\boldsymbol{a} = [a_k]_k$ can be computed as the eigenvalues of $\boldsymbol{P}^{\frac{T}{2}}\boldsymbol{K}_1\boldsymbol{P}^{\frac{1}{2}}$

$$\boldsymbol{a} = \mathrm{eigh}\left(\boldsymbol{P}^{\frac{T}{2}}\boldsymbol{K}_1\boldsymbol{P}^{\frac{1}{2}}\right) \tag{15}$$

$$= \mathrm{eigh}\left(\boldsymbol{W}^T\boldsymbol{P}\boldsymbol{W}\right). \tag{16}$$

As the distribution of the score test statistics is a mixture of χ^2, following the procedure in SKAT [6], P values are computed using Davies method [8] (an exact method which directly inverts the characteristic function), switching to the Liu method [9] (an approximation method based on moment matching) when the Davies method fails to converge (see [10] for full details).

1.4.2. Association test

To define a joint association test that accounts for the possibility of effect size heterogeneity, we consider a fully marginalised form of the model described in Eq. 4, where both the G×E and the persistent effect effect sizes, following distributions $\boldsymbol{\beta}_{\mathrm{G\times E}} \sim \mathcal{N}\left(\mathbf{0}\,,\,\sigma^2_{\mathrm{G\times E}}\boldsymbol{\Sigma}\right)$ and $\beta_{\mathrm{G}} \sim \mathcal{N}\left(0\,,\,\sigma^2_{\mathrm{G}}\right)$, respectively, are integrated out. This fully marginalised model can be cast as

$$\boldsymbol{y} \;\sim\; \mathcal{N}\left(\boldsymbol{X}\boldsymbol{b}\,,\,\underbrace{\sigma^2_{\mathrm{G_{tot}}}\boldsymbol{K}_\rho}_{\mathrm{G+G\times E}} + \underbrace{\sigma^2_e\boldsymbol{\Sigma}}_{\mathrm{E}} + \underbrace{\sigma^2_n\boldsymbol{I}}_{\mathrm{noise}}\right) \tag{17}$$

where $\sigma^2_{\mathrm{G_{tot}}} = \sigma^2_{\mathrm{G}} + \sigma^2_{\mathrm{G\times E}}$ denotes the total variance of the genetic effect (including both persistent and G×E components). Additionally, we have introduced

$$\boldsymbol{K}_\rho \;=\; \underbrace{(1-\rho)\,\boldsymbol{x}\boldsymbol{x}^T}_{\mathrm{G}} + \underbrace{\rho\mathrm{diag}(\boldsymbol{x})\boldsymbol{\Sigma}\mathrm{diag}(\boldsymbol{x})}_{\mathrm{G\times E}}, \tag{18}$$

where the parameter ρ is defined as the fraction of the total genetic variance explained by G×E effects, $\rho = \sigma^2_{\mathrm{G\times E}}/\sigma^2_{\mathrm{G_{tot}}}$. We note that this fully marginalised form of the model can also be derived from the linear model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{x}\odot\boldsymbol{\beta} + \boldsymbol{u} + \boldsymbol{\psi}, \tag{19}$$

by marginalizing over $\boldsymbol{\beta}$, assuming the following multivariate normal prior

$$\boldsymbol{\beta} \;\sim\; \mathcal{N}\left(\mathbf{0}\,,\,\sigma^2_{\mathrm{G_{tot}}}\left[\underbrace{(1-\rho)\mathbf{1}\mathbf{1}^T}_{\mathrm{G}} + \underbrace{\rho\boldsymbol{\Sigma}}_{\mathrm{GxE}}\right]\right). \tag{20}$$

Note that in this representation $\boldsymbol{\beta}$ accounts for G and G×E effects, with the first covariance term corresponding to a persistent genetic effect (G, full correlation of the genetic effect sizes across individuals) whereas the second covariance term accounts for heterogeneity of genetic effect sizes due to G×E.

In the marginalised form of the model given by Eq (17-18), an association test corresponds to assessing the alternative hypothesis $\sigma^2_{\mathrm{G_{tot}}} > 0$. The form of both the model and the test allow us to implement an efficient optimal score test procedure, building on the work in [11]. This score-based test again has complexity that scales linear in the number of individuals, provided that the environment covariance $\boldsymbol{\Sigma}$ is low rank. Below we present details of how the test is implemented, first considering the case of known ρ followed by general case of unknown ρ.

Score test for given ρ. The score test for given ρ is analogous to the one described in Section 1.4.1 but with $\boldsymbol{K}_1$ replaced by $\boldsymbol{K}_\rho = \left((1-\rho)\,\boldsymbol{x}\boldsymbol{x}^T + \rho\mathrm{diag}(\boldsymbol{x})\boldsymbol{\Sigma}\mathrm{diag}(\boldsymbol{x})\right)$ such that $\boldsymbol{W} = \left[\sqrt{1-\rho}\boldsymbol{x} \quad \sqrt{\rho}\mathrm{diag}(\boldsymbol{x})\boldsymbol{E}\right]$ and $[\boldsymbol{X}, \boldsymbol{x}]$ replaced by $\boldsymbol{X}$.

Score test for unknown ρ. As in the test $\sigma^2_{\mathrm{G_{tot}}} > 0$ in the marginalised model in Eq (17-18), the fraction of genetic variance due to G×E (ρ) is in general unknown, one would like to select ρ to maximise detection power. Following [11], we define the optimal score test as follows:

1. Define a grid of R values $\rho_1 < \rho_2 < \cdots < \rho_R$ for ρ_r (by default, we consider $[0, 0.5, 0.75, 0.84, 0.91, 0.96, 0.99, 1]$);
2. For each value ρ_r, compute the score test statistic Q_{ρ_r} using Eq (11) and corresponding P values p_{ρ_r} using the modified Liu method (an approximation method);
3. Compute a P value for the test statistic $T = \min\{p_{\rho_1}, \ldots, p_{\rho_R}\}$.

By definition, this P value can be computed from $Q_{\rho_1}, \ldots, Q_{\rho_R}$ and T as

$$P_T = p(t < T) = 1 - P(Q_{\rho_1} < q_{\rho 1}(T), Q_{\rho_2} < q_{\rho 1}(T), ..., Q_{\rho_\nu} < q_{\rho 1}(T)) \tag{21}$$

where $q_{\rho_r}(T)$ denotes the $(1-T)^{\text{th}}$ percentile of Q_{ρ_r}. It can be shown that Q_ρ can be written as follows (see derivation in Appendix A)

$$Q_\rho = \frac{1}{2}\rho\kappa + \frac{1}{2}\tau_\rho\eta_0, \tag{22}$$

where

$$\begin{aligned}
\kappa &= \sum_{k=1}^{m} \lambda_k \eta_k + \xi && (23)\\
\boldsymbol{\lambda} &= \mathrm{eigh}\,(\boldsymbol{\Lambda}) && (24)\\
\boldsymbol{\Lambda} &= \boldsymbol{\Lambda}_0 - \frac{1}{c}\boldsymbol{\alpha}\boldsymbol{\alpha}^T && (25)\\
\mathrm{Var}\,(\xi) &= \frac{4}{c}\boldsymbol{\alpha}^T\boldsymbol{\Lambda}_0\boldsymbol{\alpha} - \frac{4}{c^2}\left(\boldsymbol{\alpha}^T\boldsymbol{\alpha}\right)^2 && (26)\\
\tau_\rho &= c(1-\rho) + \frac{\rho}{c}\boldsymbol{\alpha}^T\boldsymbol{\alpha} && (27)\\
\eta_k &\overset{\text{iid}}{\sim} \chi^2_1 && (28)\\
\boldsymbol{\Lambda}_0 &= \boldsymbol{E}^T\mathrm{diag}(\boldsymbol{x})\boldsymbol{P}\mathrm{diag}(\boldsymbol{x})\boldsymbol{E} && (29)\\
\boldsymbol{\alpha} &= \boldsymbol{E}^T\mathrm{diag}(\boldsymbol{x})\boldsymbol{P}\boldsymbol{x} && (30)\\
c &= \boldsymbol{x}\boldsymbol{P}\boldsymbol{x}^T. && (31)
\end{aligned}$$

Replacing Eq (22) in Eq (21) results in

$$P_T = p(t < T) \;=\; 1 - \mathbb{E}\left[P(\kappa < \min\left\{\left(2q_{\rho_r}(T) - \tau_{\rho_r}\eta_0\right)/\rho_r \middle|\, \eta_0\right\}_{r=1}^{R}\right], \tag{32}$$

This P value can computed using one-dimensional numerical integration in the same manner as for the interaction test; using Davies method [8] (an exact method which directly inverts the characteristic function), switching to the Liu method [9] (an approximation method based on moment matching) when convergence fails (see [11, 10] for details).

1.4.3. Computational complexity

To evaluate the interaction and joint association tests, we need to first fit the null model and then compute the corresponding score test statistics. A naive implementation of both of these operations would result in computations that scale cubically with the number of individuals, i.e. $O(N^3)$. To reduce the complexity of both operations, we exploit that, when considering a linear covariance of L environments, the total covariance matrix is

$$\sigma_e^2 \boldsymbol{E}\boldsymbol{E}^T + \sigma_n^2 \boldsymbol{I}, \tag{33}$$

where the first term is low rank ($L << N$). In this setting, the null model can be fit with computational complexity $O(NR^2 + R^3)$, where R is the rank of the first covariance term term (in Eq (33), in this case, $R = L$). We refer to [7, 12] for further details.

Concerning the computation of the score test statistics for interaction test, Q can be calculated with computational complexity $O(NL^2 + NK^2 + NLK + L^2K + K^3 + L^3)$ where K corresponds to the number of covariates, while the coefficient $\boldsymbol{a}$ of the χ_1^2-mixture (eigenvalues of $\boldsymbol{W}^T\boldsymbol{P}\boldsymbol{W}$, see Eq (16)), can be computed in $O(NL^2 + NK^2 + NLK + L^2K + K^3 + L^3)$. The computation of P values either using the Davies or Liu method does not depend on the number of individuals.

For the association test for unknown ρ, the standard score-based procedure is repeated multiple times, corresponding to the number of values in the grid search over ρ. The additional variables that need to be computed (see Eqs 23-31) have linear computational complexity with the number of individuals. Finally, the one-dimensional numerical integration step (using Davies or Liu method) has a computational complexity that is independent of the number of individuals.

1.5. Characterisation of G×E loci

The StructLMM framework facilitates different analyses for characterising G×E effects. Although all the computations of these analyses remain linear in the number of individuals, the underlying computations are less efficient than those required for score tests, as maximum likelihood model parameters need to be determined explicitly for each variant that is considered. For this reason, we recommend using these analyses exclusively at the suggestive loci identified by StructLMM association and interaction tests.

1.5.1. Estimation of variance components and the fraction of genetic variance explained by GxE

StructLMM allows for decomposing the phenotypic variance into genetic (G), G×E and additive environment (E) variance components. Based on the marginalised model

in Eq (6), the three variance components can be estimated as

$$\begin{aligned} \mathrm{var}^{(\mathrm{G})} &= \mathrm{var}_S\left(\boldsymbol{x}\hat{\beta}_{\mathrm{G}}\right) = \hat{\beta}^2_{\mathrm{G}}\mathrm{var}_S\left(\boldsymbol{x}\right), & (34)\\ \mathrm{var}^{(\mathrm{G}\times\mathrm{E})} &= \mathrm{var}_{\mathrm{M}}\left(\hat{\sigma}^2_{\mathrm{G}\times\mathrm{E}}(\boldsymbol{x}\odot\boldsymbol{E})(\boldsymbol{x}\odot\boldsymbol{E})^T\right), & (35)\\ \mathrm{var}^{(\mathrm{E})} &= \mathrm{var}_{\mathrm{M}}\left(\hat{\sigma}^2_{E}\boldsymbol{E}\boldsymbol{E}^T\right). & (36) \end{aligned}$$

Here, var_S denotes the sample variance and $\mathrm{var}_{\mathrm{M}}(\boldsymbol{K})$ denotes the expected value of the sample variance of $\boldsymbol{z}$ with $\boldsymbol{z}\sim\mathcal{N}\left(\boldsymbol{0}\,,\,\boldsymbol{K}\right)$. We also use the following two identities for terms that involve the linear covariance matrix, $\boldsymbol{\Sigma}=\boldsymbol{E}\boldsymbol{E}^T$, $\mathrm{diag}(\boldsymbol{x})\boldsymbol{\Sigma}\mathrm{diag}(\boldsymbol{x}) = (\boldsymbol{x}\odot\boldsymbol{E})(\boldsymbol{x}\odot\boldsymbol{E})^{T*}$ and denote the MLE estimates of β_{G}, $\sigma^2_{\mathrm{G}\times\mathrm{E}}$ and σ^2_{E} as $\hat{\beta}_{\mathrm{G}}$, $\hat{\sigma}^2_{\mathrm{G}\times\mathrm{E}}$ and $\hat{\sigma}^2_{\mathrm{E}}$, respectively.

Using $\mathrm{var}_{\mathrm{M}}(\boldsymbol{K}) = \frac{1}{N-1}\mathrm{tr}(\boldsymbol{P}\boldsymbol{K})$, where $P=\boldsymbol{I}-\frac{1}{\dim(\boldsymbol{K})}\boldsymbol{1}\boldsymbol{1}^T$[2, 13], we obtain

$$\begin{aligned} \mathrm{var}^{(\mathrm{G}\times\mathrm{E})} &= \frac{1}{N-1}\hat{\sigma}^2_{G\times E}\left\|\boldsymbol{P}(\boldsymbol{x}\odot\boldsymbol{E})\right\|^2 & (37)\\ \mathrm{var}^{(\mathrm{E})} &= \frac{1}{N-1}\hat{\sigma}^2_{\mathrm{E}}\left\|\boldsymbol{P}\boldsymbol{E}\right\|^2. & (38) \end{aligned}$$

Finally, the estimated fraction of genetic variance explained by G×E effects (ρ) follows as

$$\hat{\rho} = \frac{\mathrm{var}^{(\mathrm{G}\times\mathrm{E})}}{\mathrm{var}^{(\mathrm{G})}+\mathrm{var}^{(\mathrm{G}\times\mathrm{E})}}. \quad (39)$$

It can be seen that this is equivalent to ρ described in Section 1.4.2 (optimal score test procedure), such that $\mathrm{var}^{(\mathrm{G}\times\mathrm{E})}\equiv\sigma^2_{\mathrm{G}\times\mathrm{E}}$ and $\mathrm{var}^{(\mathrm{G})}+\mathrm{var}^{(\mathrm{G}\times\mathrm{E})}\equiv\sigma^2_{\mathrm{tot}}$

Note that the estimate $\hat{\rho}$ is based on a MLE whilst the estimate obtained from the optimal score test procedure corresponds to the minimum P value obtained from searching over a grid of predefined values, which has a different objective.

1.5.2. Exploring the environments that drive the observed interaction effects

The evidence of individual environmental variables or sets of environments for driving the observed G×E effects can be assessed by comparing the log marginal likelihood of models with and without these environments included. The resulting Bayes factors from such comparisons are directly calibrated because the number of parameters fit using maximum likelihood is independent of the number of environmental variables.

Given a variant and a set of L environments $\mathcal{E}=\{\boldsymbol{e}_1,\ldots,\boldsymbol{e}_L\}$, the evidence for a subset of environments, $\mathcal{E}_i\subseteq\mathcal{E}$, $L_i<L$ driving the observed G×E effect can be estimated as

$$\log(\text{Bayes factor})(\mathcal{E}_i) = \mathrm{LML}(\mathrm{M}_{\mathcal{E}}) - \mathrm{LML}(\mathrm{M}_{\mathcal{E}\setminus\mathcal{E}_i}) \quad (40)$$

*We use the $\odot$ operator between vectors and matrices to denote the matrix obtained by the element-wise multiplication of matrix columns by the input vector.

where LML($M_{\mathcal{E}}$) and LML($M_{\mathcal{E}\setminus\mathcal{E}_i}$) denotes the marginal log-likelihood of the model in Eq (6), either considering the full or reduced sets of environments to define the G×E environment covariance respectively. Specifically, while in $M_{\mathcal{E}}$ all environments are used in the G×E covariance, in $M_{\mathcal{E}\setminus\mathcal{E}_i}$ only the environments not in $\mathcal{E}_i$ are considered (the operation $\setminus$ denotes the set difference). Note that the additive environment covariance remains the same in both models and is always estimated using all environments. To identify a putative causal set of driving environments, we use a greedy backward elimination procedure. Initially, we calculate the log(Bayes factor) for each environmental variable, to identify the marginal environment with the most evidence: $i_{\max} = \text{argmax}_i\,(\log(\text{Bayes factor})(\mathcal{E}_i))$. Subsequently, this environment is removed and the process iterated, stopping when there is positive evidence based on the log Kass and Raftery scale [14] that we have selected a full set of environments that can drive the observed G×E effect (i.e. when the delta log(Bayes Factor) between the model under consideration and the model that removes all environmental variables is <1).

1.5.3. Estimation of per-individual allelic effect sizes due to G × E

StructLMM allows for estimating per-individual allelic effect sizes of selected variants based on the distribution of environmental profiles attained in a population. For a given locus, the effect size of the non-reference allele in environment profile $\boldsymbol{e}^\star$ can be estimated using a procedure based on the best linear unbiased predictor (BLUP, [15]):

1. Estimate the G + G × E effect for a homozygous reference individual in environmental profile $\boldsymbol{e}^\star$ using BLUP and the model in Eq (6)

$$\text{ref}(\boldsymbol{e}^\star) = \underbrace{x_r\hat{\beta}_{\text{G}}}_{\text{G}} + \underbrace{\hat{\sigma}^2_{\text{G}\times\text{E}} x_r \boldsymbol{e}^{\star T}(\boldsymbol{x}\odot\boldsymbol{E})^T\hat{\boldsymbol{H}}^{-1}\left(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{b}}-\boldsymbol{x}\hat{\beta}_{\text{G}}\right)}_{\text{G}\times\text{E}}. \tag{41}$$

   Here, $\hat{\boldsymbol{H}}$, $\hat{\beta}$ and $\hat{\boldsymbol{b}}$ are MLE of the total covariance matrix, the genetic effect size β and the covariate effect sizes $\boldsymbol{b}$ in the model in Eq (6), while x_r is the value corresponding to the encoding of the homozygous reference genotype.

2. Compute the BLUP of the G + G × E for a heterozygous individual in environmental profile $\boldsymbol{e}^\star$ using the same model

$$\text{alt}(\boldsymbol{e}^\star) = \underbrace{x_a\hat{\beta}_G}_{\text{G}} + \underbrace{\hat{\sigma}^2_{\text{G}\times\text{E}} x_a \boldsymbol{e}^{\star T}(\boldsymbol{x}\odot\boldsymbol{E})^T\hat{\boldsymbol{H}}^{-1}\left(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{b}}-\boldsymbol{x}\hat{\beta}_G\right)}_{\text{G}\times\text{E}}, \tag{42}$$

   where x_a corresponds to the encoding of the heterozygous genotype.

3. The allelic effect in environment profile $\boldsymbol{e}^\star$ is then defined as the difference be-

tween alt($\boldsymbol{e}^\star$) and ref($\boldsymbol{e}^\star$)

$$\beta^{(\text{tot})}(\boldsymbol{e}^\star) = \text{alt}(\boldsymbol{e}^\star) - \text{ref}(\boldsymbol{e}^\star) \tag{43}$$

$$= (x_a - x_r)\left(\underbrace{\hat{\beta}_G}_{\text{G}} + \underbrace{\hat{\sigma}^2_{\text{G}\times\text{E}}\boldsymbol{e}^{\star T}(\boldsymbol{x}\odot\boldsymbol{E})^T\hat{\boldsymbol{H}}^{-1}\left(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{b}}-\boldsymbol{x}\hat{\beta}_{\text{G}}\right)}_{\text{G}\times\text{E}}\right) \tag{44}$$

$$= (x_a - x_r)\left(\underbrace{\hat{\beta}_{\text{G}}}_{\text{G}} + \underbrace{\beta_{\text{G}\times\text{E}}(\boldsymbol{e}^\star)}_{\text{G}\times\text{E}}\right) \tag{45}$$

In order to compute allelic effects for $N^\star$ different environmental profiles $\{\boldsymbol{e}^\star_1,\ldots,\boldsymbol{e}^\star_{N^\star}\}$, we use Eq (44) in vectorial form as

$$\boldsymbol{\beta}^{(\text{tot})}(\boldsymbol{E}^\star) = (x_a - x_r)\left(\underbrace{\mathbf{1}_N\hat{\beta}_{\text{G}}}_{\text{G}} + \underbrace{\hat{\sigma}^2_{\text{G}\times\text{E}}\boldsymbol{E}^\star(\boldsymbol{x}\odot\boldsymbol{E})^T\hat{\boldsymbol{H}}^{-1}\left(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{b}}-\boldsymbol{x}\hat{\beta}_{\text{G}}\right)}_{\text{G}\times\text{E}}\right), \tag{46}$$

where $\boldsymbol{\beta}^{(\text{tot})}$ is $N^\star\times 1$ vector of genetic effects in the corresponding $N^\star$ profiles and $\boldsymbol{E}^\star$ is the $N^\star\times L$ environment matrix

$$\boldsymbol{E}^\star = \begin{bmatrix}\boldsymbol{e}^\star_1\\ \ldots\\ \boldsymbol{e}^\star_N\end{bmatrix}. \tag{47}$$

This formula can be applied for the whole set of N observed environmental profiles ($\boldsymbol{E}^\star = \boldsymbol{E}$) (in-sample estimations), for environmental profiles from test/validation samples (out-of-sample predictions).
Finally, we note that if the genetic vector $\boldsymbol{x}$ is standardised, then $x_r = \frac{-2p}{\sqrt{2p(1-p)}}$ and $x_a = \frac{1-2p}{\sqrt{2p(1-p)}}$, where p is the observed minor allele frequency (MAF). It follows that $x_a - x_r = \frac{1}{\sqrt{2p(1-p)}}$.

Aggregate environment driving G×E. StructLMM can also be used to estimate the aggregate environment to explain G×E at single loci, which can be interpreted as the MAP posterior estimate of $\boldsymbol{\beta}_{\text{G}\times\text{E}}$. An estimate can be derived based form the linear model equivalence underlying StructLMM in Eq (8) (see Section 1.3), which we use to rewrite the G×E term as

$$\sum_{l=1}^{L}(\boldsymbol{x}\odot\boldsymbol{e}_l)\beta_l = \boldsymbol{x}\odot\left(\sum_{l=1}^{L}\boldsymbol{e}_l\beta_l\right) = \boldsymbol{x}\odot(\boldsymbol{E}\boldsymbol{\beta}'_{\text{G}\times\text{E}}), \tag{48}$$

where $\boldsymbol{\beta}'_{\text{G}\times\text{E}} = [\beta_1,\beta_2,...,\beta_L]^T$ and $\boldsymbol{E}\boldsymbol{\beta}'_{\text{G}\times\text{E}}$ denotes the aggregate environment driving G×E. Using the StructLMM model in Eq (6), a maximum a posteriori estimate of the

aggregate environment is $\boldsymbol{E}\hat{\boldsymbol{\beta}}'_{\mathrm{G\times E}}$, where

$$\hat{\boldsymbol{\beta}}'_{\mathrm{G\times E}} = \hat{\sigma}^2_{\mathrm{G\times E}}(\boldsymbol{x}\odot\boldsymbol{E})^T\hat{\boldsymbol{H}}^{-1}\left(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{b}}-\boldsymbol{x}\hat{\beta}_{\mathrm{G}}\right). \tag{49}$$

1.5.4. Computational Complexity

The post-processing analyses for characterising G×E effects build on the marginal model in Eq. (6), which is amenable to fast inference schemes. Specifically, introducing $\boldsymbol{Z} = [\sigma_{\mathrm{G\times E}}\mathrm{diag}(\boldsymbol{x})\boldsymbol{E},\ \ \sigma_{\mathrm{E}}\boldsymbol{E}]$, the model in Eq (6) can be written as

$$\boldsymbol{y} \sim \mathcal{N}\left(\begin{bmatrix}\boldsymbol{X} & \boldsymbol{x}\end{bmatrix}\begin{bmatrix}\boldsymbol{b}\\ \beta\end{bmatrix},\ \boldsymbol{Z}\boldsymbol{Z}^T+\sigma_n^2\boldsymbol{I}\right). \tag{50}$$

Note, that this model has the same form as the null model, where the first covariance term has rank $2L$. Maximum likelihood can thus be achieved with computational complexity $O(N(2L)^2+(2L)^3)$ (see [7, 12] for details).

2. Relationship with prior work and comparison methods

In this section, we review existing methods for G×E and we describe LMMs that are otherwise related to StructLMM. In Section 2.1 we review related existing G×E testing strategies; in Section 2.2 we discuss technical similarities between StructLMM and existing LMM-based approaches. Finally, in Section 3 we describe alternative methods that are compared with StructLMM.

2.1. Relationship to other G×E tests

Single-environment G×E tests A standard linear model to test for G×E using quantitative traits and a single environmental variable can be cast as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b}+\boldsymbol{x}\beta_{\mathrm{G}}+(\boldsymbol{x}\odot\boldsymbol{e})\beta_{\mathrm{G\times E}}+\boldsymbol{e}\beta_{\mathrm{E}}+\boldsymbol{\psi}, \tag{51}$$

where $\boldsymbol{y}$ is a $N\times 1$ vector of quantitative trait measurements for N individuals, $\boldsymbol{X}$ is the $N\times K$ fixed effect design matrix for K covariates, $\boldsymbol{b}$ is the $K\times 1$ vector of their effect sizes, $\boldsymbol{x}$ is an $N\times 1$ genotype vector for the variant under consideration, β_{G} the corresponding persistent genetic effect, $\boldsymbol{e}$ an $N\times 1$ environment vector for the studied environment, $\beta_{\mathrm{G\times E}}$ the genotype-environment interaction effect and $\boldsymbol{\psi}$ a noise term, typically assumed to be iid normal $\boldsymbol{\psi}\overset{\text{iid}}{\sim}\mathcal{N}\left(0,\sigma_n^2\right)$ [16]. The model in Eq (51) can be used both for testing G×E ($\beta_{\mathrm{G\times E}}\neq 0$) or to test for association while accounting for G×E ($[\beta_{\mathrm{G}},\ \beta_{\mathrm{G\times E}}]\neq\mathbf{0}$) [17, 18].

G×E variant-set tests The model in Eq (51) can be extended to aggregate effects across multiple variants, enabling G×E variant-set tests for single environments. For

a set of R genetic variants $\{\boldsymbol{g}_1, \dots \boldsymbol{g}_R\}$, a variant-set model can be cast as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \sum_{s=1}^{R} \boldsymbol{g}_s \beta_s^{\mathrm{G}} + \sum_{s=1}^{R} (\boldsymbol{g}_s \odot \boldsymbol{e})\, \beta_s^{\mathrm{G}\times E} + \boldsymbol{e}\beta^{\mathrm{E}} + \boldsymbol{\psi}. \tag{52}$$

Tukey's one degree-of-freedom (df) is one of the first G×E variant-set tests [19], which assumes proportionality between interaction and marginal genetic effect sizes (i.e., $\beta_s^{\mathrm{G}\times E} = \theta\beta_s^{\mathrm{G}},\ s = 1, \dots, R$). This assumption can be restrictive, but comes with the advantage that variant-sets can be tested via $\theta \neq 0$ (1 df). Other strategies to re-weight the interaction effects and limit the number of testing parameters have been proposed (for example, Jiao et al [20] proposed using gene-environment correlations).

Another reweighting scheme that is commonly used for G×E analysis is known as the genetic risk score (GRS), where one assumes $\beta_s^{\mathrm{G}} = w_s\beta^{\mathrm{G}}$ and $\beta_s^{\mathrm{G}\times E} = w_s\beta^{\mathrm{G}\times E}$. In GRS analyses, the SNP weight w_s is either i) set to $w_s = 1$ for all s (unweighted GRS, [21]) or ii) to $w_s = \beta_s^{\mathrm{P}}$, where β_s^{P} is the effect size of variant s identified from a previous study that tested only for main genetic effects (weighted GRS, [22]). GRS analyses have been used in G×E interaction analyses of BMI in UK Biobank data [22, 23] using a weighted score based on the effect sizes estimated from the 2015 GIANT association analysis for BMI [24].

A class of G×E variant-set tests that is technically related to StructLMM is based on random effects for $\boldsymbol{\beta}^{\mathrm{G}\times E}$, which build on linear mixed models, similarly to StructLMM. In the following we describe the interaction sequence kernel association test (GESAT) [25], as representative of a family of related methods [25, 26, 27, 28]. Starting from equation Eq (52), GESAT can be derived by defining a multivariate normal prior on the genetic effect sizes, $\boldsymbol{\beta}^{\mathrm{G}\times E} \sim \mathcal{N}(\boldsymbol{0}, \tau\boldsymbol{I}_R)$, where interactions can be tested by assessing $\tau \neq 0$ using a score test (see Section 1.4). GESAT models the additive genetic effects in the null model using fixed effects, estimated using ridge regression to mitigate the large number of covariates in the test [29]. The score-based test statistics of the interaction test is

$$\boldsymbol{Q} = (\boldsymbol{y} - \hat{\boldsymbol{\mu}})^\top \boldsymbol{S}\boldsymbol{S}^\top (\boldsymbol{y} - \hat{\boldsymbol{\mu}}), \tag{53}$$

where $\boldsymbol{S} = [\boldsymbol{g}_1 \odot \boldsymbol{e}, \dots, \boldsymbol{g}_R \odot \boldsymbol{e}]$ and $\hat{\boldsymbol{\mu}}$ is the optimised mean under the null model. $\boldsymbol{Q}$ follows a mixture of χ^2 distributions with 1 df [6] (see Section 1.4). Among the generalisations of GESAT are methods tailored towards rare variants (iSKAT, [26]), with a prior on $\boldsymbol{\beta}^{\mathrm{G}\times E}$ that accounts both for scenarios where the $\mathrm{G} \times E$ effects have consistent effect directions or have independent effects ($\boldsymbol{\beta}^{\mathrm{G}\times E} \sim \mathcal{N}\left(\boldsymbol{0}, \tau(\pi\boldsymbol{1}_R\boldsymbol{1}_R^T + (1-\pi)\boldsymbol{I}_R)\right)$). An optimal score-test is considered to assess statistical significance (see Section 1.4). The testing procedure in iSKAT is related to the association test in StructLMM, but there are important differences. First, the prior used in iSKAT is designed to empower the detection of interaction effects between multiple rare variants and a single environmental variable. Conversely, StructLMM considers a prior on the effect of a single variants due to G×E with multiple environments. Additionally, a technical difference is in how the two methods deal with the high number of covariates: while StructLMM

uses a random effect, iSKAT uses ridge regression. Finally, StructLMM can also be used to efficiently fit the full model with G+E+G×E effects, which is critical to enable for the characterisation of selected G×E loci (see Section 1.5).

StructLMM is also related to a recently proposed random effect set test, iSet [30], which builds on a framework for multi-trait set tests [12]. Critically, the model is limited to analyses of moderately sized cohorts (at most tens of thousands of samples), it only considers categorical environments and it cannot be used for the analysis of multiple environments.

Existing multi-environment G×E anlayses The only multivariate G×E analysis with multiple environments we are aware of has been proposed in [31], in which the authors study G×E effects on BMI at the *FTO* locus. Briefly, the authors employ a two-step approach, first selecting relevant environmental variables based on their additive effect on the phenotype (using a cross validation scheme), and then assess G×E effects using a z-score test statistics using a multivariate fixed effect model (as in Eq. (8)). Thus, the method is similar to the fixed effect model we implement for comparison (Section 1.5.2). At present, there is no software that implements this procedure [31], which also does not scale to genome-wide applications due to the costly cross-validation step.

2.2. Relationship to other LMM implementations

Although the models follow different aims, StructLMM is technically related to the optimal score-based test originally implemented in SKAT-O [11], a rare variant association test. The SKAT-O model can be cast as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{G}\boldsymbol{\beta}^{\mathrm{G}} + \boldsymbol{\psi}, \tag{54}$$

where where $\boldsymbol{G} = [\boldsymbol{g}_1, \dots \boldsymbol{g}_R]$ denote the $N \times R$ genetic design matrix for R rare variants and $\boldsymbol{\beta}^{\mathrm{G}}$ is set to follow the multivariate normal distribution

$$\boldsymbol{\beta}^{\mathrm{G}} \sim \mathcal{N}\left(\boldsymbol{0}, \tau(\pi \boldsymbol{1}_R \boldsymbol{1}_R^T + (1-\pi)\boldsymbol{I}_R\right). \tag{55}$$

This formulation interpolates between fully correlated genetic effects ($\boldsymbol{1}_R\boldsymbol{1}_R^T$) and independent genetic effects ($\boldsymbol{I}_R$). The relation to StructLMM becomes apparent from Eqs (19-20), considering that $\boldsymbol{x} \odot \boldsymbol{\beta} = \mathrm{diag}(\boldsymbol{x})\boldsymbol{\beta}$. In essence, SKAT-O interpolates between models that consider different correlation structures of genetic effects across variants, whereas StructLMM interpolates between different covariance models for per-individual genetic effects.

A second related model is an LMM-based method for gene-gene (G×G) interaction test introduced in [32], between a single variant and multiple others. While conceptually related, the proposed tests scales quadratically with the number of samples (StructLMM is linear), and hence this method cannot be applied to larger datasets.

Additionally, the method does not consider a joint association test but instead is limited to testing for interactions at this point.

Finally, we note that the approaches we employ for efficient fitting of the null and alternative models in StructLMM borrows ideas from previous efficient implementations of linear mixed models using low-rank random effect, including [5, 33, 7, 12, 30].

3. Comparison methods

We compare StructLMM to alternative implementations of single and multi environments GxE models. Additionally, for association tests, we present comparisons to alternative implementations of linear persistent effect tests. Here we provide full details on the models these tests build on. For a compact tabular summary, we refer to **Supp. Table 2** with the methods (SingleEnv-Renv, LMM-Renv and MultiEnv-Renv-LRT) that are used in the main text results displayed in bold. These methods were chosen as the default comparison partners as they use the same random effect term to model the additive environment under the null as StructLMM and hence are most directly comparable.

3.1. Single environment models

Single Environment model with Single environment additive effect (SingleEnv-Senv). Standard interaction tests and joint association tests consider the following model [16, 17] for each environment $\boldsymbol{e}_l$ in isolation:

$$\boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{X}\boldsymbol{b} + \underbrace{\boldsymbol{x}\beta_G}_{\text{G}} + \underbrace{(\boldsymbol{x}\odot\boldsymbol{e}_l)\beta_{\text{G}\times\text{E}}}_{\text{G}\times\text{E}} + \underbrace{(\boldsymbol{e}_l)\gamma}_{\text{E}}, \underbrace{\sigma_n^2\boldsymbol{I}}_{\text{noise}}\right), \tag{56}$$

where $\boldsymbol{y}$ is an $N\times 1$ phenotype vector for N individuals, $\boldsymbol{X}$ is the $N\times K$ fixed effect design matrix for K covariates, $\boldsymbol{b}$ is the $K\times 1$ vector of their effect sizes, $\boldsymbol{x}$ is an $N\times 1$ genotype vector for the variant being tested, β_G the corresponding genetic effect and $\boldsymbol{\psi}$ the residuals. This model can be used to assess the presence of G×E interaction with environment $\boldsymbol{e}_l$ by testing $\beta_{\text{G}\times\text{E}} \neq 0$ (1 df). A joint P value that corresponds to the alternative hypothesis that at least one of L environments is participating in G×E effects, can then be constructed by performing L tests followed by appropriate adjustment for multiple testing (we consider Bonferroni). Similarly, a joint association test that accounts for G×E effects due to single environments $\boldsymbol{e}_l$ can be derived by testing $[\beta_{\text{G}}, \beta_{\text{G}\times\text{E}}] \neq \boldsymbol{0}$ (2 df), where again multiple environments and their corresponding tests can be combined using Bonferroni adjustment.

Single Environment model with multi-environment additive effect as Fixed effect (SingleEnv-Fenv). One can extend the standard approach presented above by mod-

elling additive environment effects from multiple environments:

$$\boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{X}\boldsymbol{b} + \underbrace{\boldsymbol{x}\beta_G}_{\text{G}} + \underbrace{(\boldsymbol{x}\odot\boldsymbol{e}_l)\beta_{\text{G}\times\text{E}}}_{\text{G}\times\text{E}} + \underbrace{\sum_{l=1}^{L}(\boldsymbol{e}_l)\gamma_l}_{\text{E}}, \underbrace{\sigma_n^2\boldsymbol{I}}_{\text{noise}}\right). \tag{57}$$

Again, the presence of interaction and association can be assessed by testing $\beta_{\text{G}\times\text{E}} \neq 0$ and $[\beta_{\text{G}}, \beta_{\text{G}\times\text{E}}] \neq \boldsymbol{0}$ for each environment respectively, where again multiple environments can be combined using Bonferroni adjustment. This model provides a more accurate null model for both tests, modelling additive effects of other relevant environments.

Single Environment model with multi-environment additive effect as Random effect (SingleEnv-Renv). Alternatively, one can use a random effect to model multivariate additive environments, which is analogous to the approach taken in StructLMM. Specifically, one can define an environmental covariance $\boldsymbol{\Sigma}$ based on the observed environments as described in Section 1.2 and consider the model

$$\boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{X}\boldsymbol{b} + \underbrace{\boldsymbol{x}\beta_G}_{\text{G}} + \underbrace{(\boldsymbol{x}\odot\boldsymbol{e}_l)\beta_l}_{\text{G}\times\text{E}}, \underbrace{\sigma_e^2\boldsymbol{\Sigma}}_{\text{E}} + \underbrace{\sigma_n^2\boldsymbol{I}}_{\text{noise}}\right), \tag{58}$$

where again interaction and association tests can be implemented as described above. In the main paper, we consider this model as it has the same background model as StructLMM but perform comparisons with alternative implementations in **Supp. Fig. 2**.

3.2. Association tests

Linear model (LM). For comparison, we consider a conventional linear model (LM), assuming genetic effect sizes that are constant in the population

$$\boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{X}\boldsymbol{b} + \underbrace{\boldsymbol{x}\beta_G}_{\text{G}}, \underbrace{\sigma_n^2\boldsymbol{I}}_{\text{noise}}\right). \tag{59}$$

Association with genetic variant $\boldsymbol{x}$ can be assessed through the 1dof test $\beta_G \neq 0$.

Linear model with multivariate additive environment effects. In a linear model, multivariate additive effects from environment can be accounted for by the means of a fixed effect term, which results in the model (LM-Fenv):

$$\boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{X}\boldsymbol{b} + \underbrace{\boldsymbol{x}\beta_G}_{\text{G}} + \underbrace{\sum_{l=1}^{L}(\boldsymbol{e}_l)\gamma_l}_{\text{E}}, \underbrace{\sigma_n^2\boldsymbol{I}}_{\text{noise}}\right). \tag{60}$$

Alternatively, similarly to StructLMM, one can use a random effect with environment covariance $\boldsymbol{\Sigma}$, obtaining the following linear mixed model (LMM-Renv)

$$\boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{X}\boldsymbol{b} + \underbrace{\boldsymbol{x}\beta_G}_{\text{G}}, \underbrace{\sigma_e^2\boldsymbol{\Sigma}}_{\text{E}} + \underbrace{\sigma_n^2\boldsymbol{I}}_{\text{noise}}\right). \tag{61}$$

In the main paper, we consider the random-effect model as it has the same background model as StructLMM but perform comparisons with alternative implementations in **Supp. Fig. 2**.

3.2.1. Alternative multivariate G×E tests based on fixed effects

Multi Environment model with multi-environment additive effect as Fixed effect (MultiEnv-Fenv). Whilst we implement StructLMM using a random effect term to model GxE, an alternative implementation using multiple fixed effects can be considered. Explicitly, denoting with $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_L$ the $N \times 1$ vectors for L environments, a fixed-effect-based multi-environment model can be cast as

$$\boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{X}\boldsymbol{b} + \underbrace{\boldsymbol{x}\beta_G}_{\text{G}} + \underbrace{\sum_{l=1}^{L}(\boldsymbol{x}\odot\boldsymbol{e}_l)\beta_l}_{\text{G}\times\text{E}} + \underbrace{\sum_{l=1}^{L}(\boldsymbol{e}_l)\gamma_l}_{\text{E}}, \underbrace{\sigma_n^2\boldsymbol{I}}_{\text{noise}}\right), \tag{62}$$

where both interactions due to L environmental variables and their additive effects are modelled as fixed effects. Within this model, the presence of interactions and associations can be assessed by testing $[\beta_1, \ldots, \beta_L] \neq \mathbf{0}$ (L df test) and $[\beta, \beta_1, \ldots, \beta_L] \neq \mathbf{0}$ ($L+1$ df test), respectively.

We consider both an LR-based and a score-based implementations of the tests (in **Supp. Fig. 1b**), which we respectively refer to as MultiEnv-Fenv-LRT and MultiEnv-Fenv-Score. Briefly, for the score test, we employ the Rao's score test statistic [34]

$$\text{RS} = \boldsymbol{U}_0^T\boldsymbol{I}_0^{-1}\boldsymbol{U}_0, \tag{63}$$

where $\boldsymbol{U}_0$ and $\boldsymbol{I}_0$ are respectively the gradient and the Fisher Information matrix (FIM) with respect to the tested parameters, computed use MLE under the null[†]. Rao's score test statistic has an asymptotic chi-square distribution with the number

[†]Specifically, for the test $\boldsymbol{\beta} \neq \mathbf{0}$ in the Gaussian model

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{F}\boldsymbol{\alpha} + \boldsymbol{W}\boldsymbol{\beta}, \boldsymbol{H}), \tag{64}$$

we have

$$U(\beta) = (\boldsymbol{y} - \boldsymbol{F}\boldsymbol{\alpha}_0)^T\boldsymbol{H}_0^{-1}\boldsymbol{W}, \tag{65}$$

$$I(\beta) = \boldsymbol{W}^T\boldsymbol{H}_0^{-1}\boldsymbol{W}, \tag{66}$$

where $\boldsymbol{\alpha}_0$ and $\boldsymbol{H}_0$ are MLE of $\boldsymbol{\alpha}$ and $\boldsymbol{H}$ under the null model.

of tested parameters as degrees of freedom.

Whilst the LR test is inflated, the score test is deflated when the number of environments is large relative to the sample size and again highlighting the benefits of the StructLMM tests that have constant number of parameters independent of the number of environments.

Multi Environment model with multi-environment additive effect as Random effect (MultiEnv-Renv). As with the single environment and linear model tests, the multivariate additive environment effect can be modelled as a random effect with environment covariance $\mathbf{\Sigma}$, analogous to the StructLMM model, resulting in the following model

$$\boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{X}\boldsymbol{b} + \underbrace{\boldsymbol{x}\beta_G}_{\text{G}} + \underbrace{\sum_{l=1}^{L}(\boldsymbol{x} \odot \boldsymbol{e}_l)\beta_l}_{\text{G}\times\text{E}} , \underbrace{\sigma_e^2\mathbf{\Sigma}}_{\text{E}} + \underbrace{\sigma_n^2\boldsymbol{I}}_{\text{noise}} \right), \tag{67}$$

where again interaction and association tests can be performed as described above. Again as described above both LR tests (MultiEnv-Renv-LRT) and score tests (MultiEnv-Renv-Score) can be employed.

PC-based fixed effect G×E test Motivated by the observation that multi-environment fixed effect models for G×E are in general not calibrated, in particular when analysing larger numbers of environments (**Supp. Fig. 1b**), one can in principle considered PCA to reduce the effect number of environmental variables in the test. The model is identical to the approach described in Eq (67), but is based on the leading M principal components calculated based on the full set of L environments, where $M < L$. We consider this approach when analysing the eQTL data, where the proxy for environment is high-dimensional **Supp. Fig. 20**.

3.3. Original analysis of blood eQTL data

We here provide a brief description of the method used in [35] to identify expression quantitative trait loci whose effect depends on cellular context, Briefly, the authors considered the single-environment model in Eq (51) where $\boldsymbol{y}$ is the gene expression of the analysed eQTL gene, $\boldsymbol{g}$ its cis eQTL and $\boldsymbol{e}$ the environment-gene. Then the following procedure was considered:

- Fit model for each eQTL gene as outcome ($\boldsymbol{y}$) and each genome-wide gene as environment ($\boldsymbol{e}$);
- Compute the test statistic $\zeta_l = \sum_i z_{il}^2$ for each environment-gene l, where z_{il} is the z-score of the interaction G×E term from the model with gene eQTL i as outcome ($\boldsymbol{y}$) and gene l as environment ($\boldsymbol{e}$);

- Define the environment-gene with maximal test statistic as proxy gene and regress out its expression from all gene expression levels.

This procedure was repeated ten times, which led to the identification of ten proxy genes. Each proxy gene was then linked to a specific cell type or molecular pathway called module through various analyses (see [35] for more details). Once the proxy genes and the corresponding modules had been identified, Zhernakova *et al.* considered again the single-environment model in Eq (51) to test for interactions with specific modules at individual eQTL (each eQTL gene as outcome and each proxy gene as environment). The significance threshold for this second stage was assessed through permutations and results reported at 5% FDR.

4. Simulations

4.1. Environment covariates

In order to mimic realistic distributions in the real data analysis, we considered environmental exposures from UK Biobank data. Specifically, we extracted 33 environmental factors, which include 9 ordinal dietary variables ('Oily fish intake', 'Non-oily fish intake', 'Processed meat intake', 'Poultry intake', 'Beef intake', 'Lamb/mutton intake, 'Pork intake', 'Cheese intake' and 'Salt added to food'), three continuous dietary variables ('Cooked vegetable intake', 'Bread intake', 'Tea intake'), three physical activity variables ('Number of days/week walked 10+ minutes', 'Number of days/week of moderate physical activity 10+ minutes', 'Number of days/week of vigorous physical activity 10+ minutes'), 'Alcohol intake frequency', 'Sleep duration', 'Sleep duration residuals squared', 'Townsend deprivation index', 'Smoking status', 'Time spent watching television', 'Usual walking pace', 'Frequency of friend/family visits', 'Time spend outdoors in summer', 'Time spent outdoors in winter', 'Time spent using computer', 'Nap during day', 'Overall health rating', 'Nitrogen dioxide air pollution 2010', 'Nitrogen oxides air pollution 2010', 'Traffic intensity on the nearest major road', 'Average daytime sound level of noise pollution', 'Average evening sound level of noise pollution'[‡]. For the three continuous dietary and five pollutant variables, we removed values exceeding the 99th percentile. For 'Sleep duration', we removed the top and bottom percentiles and for each individual, calculated the squared deviations from the mean sleep duration, creating environmental variable, 'Squared sleep duration res.' (33$^{\text{rd}}$ environmental variable). For the four variables ('Time spent watching television' , 'Time spent using computer', 'Time spend outdoors in summer', 'Time spent outdoors in winter'), less than 0.5 hours of was encoded as 0.5 and we excluded individuals in the upper and lower percentile. These variables were further augmented by element wise interactions with gender and age (and the inclusion of age itself), resulting in a total of $L = 100$ environmental variables. We randomly assigned environmental profiles from the 70,282 individuals for which all environmental factors were available (based on the

[‡]Environments in blue were also considered in the BMI analysis and were preprocessed as described in the corresponding section.

Interim release) to the genotypes. The environmental covariance was built as a linear covariance from standardised values.

4.2. Phenotype simulation strategy

Phenotype data were generated as the sum of a persistent genetic effect from a genetic variant ($\boldsymbol{g}$), additive environment effects from a set of E environments ($\boldsymbol{e}$), G×E effects from a subset of $L_1 \leq L$ environments ($\boldsymbol{i}$), a contribution from population structure ($\boldsymbol{u}$) and iid gaussian noise ($\boldsymbol{\psi}$)

$$\boldsymbol{y} = \boldsymbol{g} + \boldsymbol{e} + \boldsymbol{i} + \boldsymbol{u} + \boldsymbol{\psi}. \tag{68}$$

The individual contributions from each term as in Eq (68) were simulated as follows:

- **Persistent genetic effect** from a randomly selected genetic variant is rescaled to have sample variance $(1-\rho) \cdot v_g$;
- **Additive environment effects** are generated as $\boldsymbol{e} = \boldsymbol{E}_{\mathrm{E}}\boldsymbol{\beta}_{\mathrm{E}}$, where $\boldsymbol{E}_{\mathrm{E}}$ is the $N \times L$ design matrix for L environments in N individuals and β_{E} is generated as $\beta_{\mathrm{E}} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$. Subsequently, $\boldsymbol{e}$ is rescaled to have sample variance v_e;
- **G×E effects** are generated as $\boldsymbol{i} = (\boldsymbol{E}_{\mathrm{G\times E}}\boldsymbol{\beta}_{\mathrm{G\times E}}) \otimes \boldsymbol{x}$, where $\boldsymbol{E}_{\mathrm{G\times E}}$ is the $N \times L_1$ design matrix for $L_1 < L$ environments in N individuals and $\beta_{\mathrm{G\times E}}$ is generated as $\beta_{\mathrm{G\times E}} \overset{\text{iid}}{\sim} \{+1, -1\}$. Subsequently, $\boldsymbol{i}$ is rescaled to have sample variance $\rho \cdot v_g$;
- **Population structure** is simulated as $\boldsymbol{u} = \boldsymbol{F}\boldsymbol{\beta}_{\mathrm{pop}}$, where $\boldsymbol{F}$ is the $N \times 10$ matrix of the leading 10 principal components from the realised relatedness matrix and β_{pop} is generated as $\beta_{\mathrm{pop}} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$. Subsequently, $\boldsymbol{u}$ is rescaled to have sample variance v_{pop};
- **Noise** is generated as iid Gaussian and rescaled to have sample variance $v_n = 1 - v_g - v_e - v_{\mathrm{pop}}$.

In a first set of simulation experiments, we vary the sample size (N), the number of environments (L), the percentage of them with G×E effects ($\pi = \frac{L_1}{L}$), the variance explained by the total genetic effect (G+G×E effect, v_g), the proportion of the genetic variance due to G×E effects (ρ) and the variance explained by additive environmental effects (v_e) while we fix the variance explained by population structure ($v_{\mathrm{pop}} = 40\%$). The specific parameter settings and the considered ranges of settings are provided in **Supp. Table 1**. Additionally to these extensive simulations, we also simulated scenarios with gene-exposure correlation, skewed and binary environments. Full details on these additional simulations are given in the next sections.

4.3. Gene-exposure correlations

We simulated environments that are subject to a genetic effect from a variant in LD with the variant affecting the phenotype. Specifically, we generated new environments

as

$$\text{vec}\left(\boldsymbol{E}\right) \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{C} \otimes \boldsymbol{R}\right), \tag{69}$$

where $\boldsymbol{C}$ is the covariance between environments and $\boldsymbol{R}$ is the covariance between individuals. We set $\boldsymbol{C} = \hat{\boldsymbol{C}}$ and $\boldsymbol{R} = v_x \boldsymbol{x}_e \boldsymbol{x}_e^T + (1 - v_x)\hat{\boldsymbol{R}}$, where $\hat{\boldsymbol{C}}$ and $\hat{\boldsymbol{R}}$ are column and row covariances estimated from environments in real data repectively, $\boldsymbol{x}_e$ the genotype of the variant correlated with the environments and v_x the variance explained by this variant on environments. After generation, the environments were standardised and then the phenotype simulation procedure described in Section 4.2 was applied.

In our simulations we varied (i) the LD between the variant correlated with the environments and the one associated with the phenotype and (ii) the average heritability of environments (v_x) when the phenotype and environments are affected by the the same variant. Calibration was assessed in the extreme scenario of same variant and $v_x = 0.20\%$.

4.4. Skewed and binary environments

First, we generated environments as

$$\text{vec}\left(\boldsymbol{E}\right) \sim \mathcal{N}\left(\boldsymbol{0}, \hat{\boldsymbol{C}} \otimes \hat{\boldsymbol{R}}\right), \tag{70}$$

where $\hat{\boldsymbol{C}}$ and $\hat{\boldsymbol{R}}$ are environment and individuals covariances estimated from environments in real data. After generation, the environments were transformed to have non-Gaussian distributions and then the phenotype simulation procedure described in Section 4.2 was applied. To generate skewed distributions, we rank-inverse transformed the generated environments to a Gamma distribution with shape a and scale 1. We varied the shape a of all simulated environments. Calibration was assessed in the extreme scenario $a = 0.1$. For binary environment scenarios, we binarised a fraction f_{B} of the generated environments to have a specific sample frequency ν. We varied the fraction f_{B} of binary environments (fixing $\nu = 2\%$) and the frequency ν (fixing f_{B}=1). Calibration was assessed in the extreme scenario $f_{\text{B}} = 1$ and $\nu = 0.5\%$.

5. Analysis of BMI in UK Biobank

5.1. Hold-out validation of per-individual genetic effect sizes

For out-of-sample assessment of the allelic predictions, we randomly split the 252,188 individuals into training and testing groups (126,094 individuals). We used the training group to fit the model and estimated allelic effects at the 11 interaction loci (5% FDR-adjusted) within sample, similar to the approach taken on the full dataset. We then identified the 5% (6,305) of individuals with the highest and lowest estimated allelic effects. For these strata of individuals, we use their genotype and phenotype information and test for associations using the LMM-Renv from which we identify a subset of strata that have nominally significant P values ($P < 0.05$, **Supp. Fig.**

17a). Next, we predicted allelic effects in the test set (**Supp. Fig. 17b**) based on all data in the training fraction and the environmental profiles in the test fraction, again identifying the 5% strata of individuals with the highest and lowest estimated allelic effects. We averaged over the predicted allelic effects of test individuals in the 5% strata (only for significant strata in the training set estimations (**Supp. Fig. 17a**)) to produce a mean predicted allelic effect (yellow crosses, **Supp. Fig. 17b**). We then used the genotypes and phenotypes of test individuals in in the top and bottom strata to estimate the genetic effect based on their phenotype data (using LMM-Renv). We compared mean predicted allelic effect with the estimated genetic effect (**Supp. Fig. 17c**) and we compared the mean predicted and estimated effects when considering the GxE component only. For the latter comparison, we subtracted the persistent SNP effect (using the estimate from the StructLMM fit on the training fraction) from the predicted mean effect of the strata. For the LMM-Renv estimated effect, this was achieved by the subtracting the variant effect identified from the LMM-Renv using all individuals in the testing set from the variant effect identified from the LMM-Renv using only individuals in the top and bottom 5% strata (**Supp. Fig. 17d**).

6. Analysis of cell-context eQTLs in a large blood cohort

6.1. Generation of principal components from SAMtools flagstat and Picard tools

The first eight principal components derived from SAMtools flagstat and Picard tools were calculated from the following fields: GC, Bam.genome total, Bam.exon total, Counts.number detected genes, PF BASES, PF ALIGNED BASES, CODING BASES, UTR BASES, INTRONIC BASES, INTERGENIC BASES, PCT CODING BASES, PCT UTR BASES, PCT INTRONIC BASES, PCT INTERGENIC BASES, PCT MRNA BASES, PCT USABLE BASES, MEDIAN CV COVERAGE, MEDIAN 5PRIME BIAS, MEDIAN 3PRIME BIAS, MEDIAN 5PRIME TO 3PRIME BIAS.

A. Proof for form of Q

First, let us rewrite Q_ρ

$$
\begin{aligned}
Q_\rho &= \frac{1}{2}\boldsymbol{y}^T\boldsymbol{K}_\rho\boldsymbol{y} && (71)\\
&= \frac{1}{2}(1-\rho)\boldsymbol{y}^T\boldsymbol{P}\mathrm{diag}(\boldsymbol{x})\mathbf{1}\mathbf{1}^T\mathrm{diag}(\boldsymbol{x})\boldsymbol{P}\boldsymbol{y} + \frac{1}{2}\rho\boldsymbol{y}^T\boldsymbol{P}\mathrm{diag}(\boldsymbol{x})\boldsymbol{E}\boldsymbol{E}^T\mathrm{diag}(\boldsymbol{x})\boldsymbol{P}\boldsymbol{y} && (72)\\
&= \frac{1}{2}c(1-\rho)\boldsymbol{v}^T\boldsymbol{M}\boldsymbol{v} + \frac{1}{2}\rho\boldsymbol{v}^T\boldsymbol{Z}\boldsymbol{E}\boldsymbol{E}^T\boldsymbol{Z}^T\boldsymbol{v} && (73)
\end{aligned}
$$

where

$$\begin{aligned}
\boldsymbol{v} &= \boldsymbol{P}^{\frac{T}{2}}\boldsymbol{y} && (74)\\
\boldsymbol{Z} &= \boldsymbol{P}^{\frac{T}{2}}\mathrm{diag}(\boldsymbol{x}) && (75)\\
\boldsymbol{M} &= \frac{1}{c}\boldsymbol{Z}\mathbf{1}\mathbf{1}^T\boldsymbol{Z}^T && (76)\\
c &= \mathbf{1}^T\boldsymbol{Z}^T\boldsymbol{Z}\mathbf{1} && (77)
\end{aligned}$$

Eq (73) can be equivalently written as

$$Q_\rho = \frac{1}{2}\rho\kappa + \frac{1}{2}\tau_\rho\eta_0 \quad (78)$$

where

$$\begin{aligned}
\kappa &= \phi + \xi && (79)\\
\phi &= \boldsymbol{v}^T(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{Z}\boldsymbol{E}\boldsymbol{E}^T\boldsymbol{Z}^T(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{v} && (80)\\
\xi &= 2\boldsymbol{v}^T(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{Z}\boldsymbol{E}\boldsymbol{E}^T\boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{v} && (81)\\
\tau_\rho &= c(1-\rho) + \frac{\rho}{c}\mathbf{1}^T\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{E}\boldsymbol{E}^T\boldsymbol{Z}^T\boldsymbol{Z}\mathbf{1} && (82)\\
\eta_0 &= \boldsymbol{v}^T\boldsymbol{M}\boldsymbol{v} && (83)
\end{aligned}$$

Indeed, replacing Eqs (83) in (78)

$$\begin{aligned}
Q_\rho &= \frac{1}{2}\rho\boldsymbol{v}^T(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{Z}\boldsymbol{E}\boldsymbol{E}^T\boldsymbol{Z}^T(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{v} + && (84)\\
&\quad \rho\boldsymbol{v}^T(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{Z}\boldsymbol{E}\boldsymbol{E}^T\boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{v} + && (85)\\
&\quad \left(c(1-\rho) + \frac{\rho}{c}\mathbf{1}^T\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{E}\boldsymbol{E}^T\boldsymbol{Z}^T\boldsymbol{Z}\mathbf{1}\right)\boldsymbol{v}^T\boldsymbol{M}\boldsymbol{v} && (86)\\
&= \frac{1}{2}\rho\boldsymbol{v}^T\boldsymbol{Z}\boldsymbol{E}\boldsymbol{E}^T\boldsymbol{Z}^T\boldsymbol{v} - \rho\boldsymbol{v}^T\boldsymbol{Z}\boldsymbol{E}\boldsymbol{E}^T\boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{v} + \frac{1}{2}\rho\boldsymbol{v}^T\boldsymbol{M}\boldsymbol{Z}\boldsymbol{E}\boldsymbol{E}^T\boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{v} + && (87)\\
&\quad +\rho\boldsymbol{v}^T\boldsymbol{Z}\boldsymbol{E}\boldsymbol{E}^T\boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{v} - \rho\boldsymbol{v}^T\boldsymbol{M}\boldsymbol{Z}\boldsymbol{E}\boldsymbol{E}^T\boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{v} + && (88)\\
&\quad \frac{1}{2}c(1-\rho)\boldsymbol{v}^T\boldsymbol{M}\boldsymbol{v} + \frac{1}{2c}\rho\mathbf{1}^T\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{E}\boldsymbol{E}^T\boldsymbol{Z}^T\boldsymbol{Z}\mathbf{1}\boldsymbol{v}^T\underbrace{\frac{1}{c}\boldsymbol{Z}\mathbf{1}\mathbf{1}^T\boldsymbol{Z}^T}_{\boldsymbol{M}}\boldsymbol{v} && (89)\\
&= \frac{1}{2}c(1-\rho)\boldsymbol{v}^T\boldsymbol{M}\boldsymbol{v} + \frac{1}{2}\rho\boldsymbol{v}^T\boldsymbol{Z}\boldsymbol{E}\boldsymbol{E}^T\boldsymbol{Z}^T\boldsymbol{v} + && (90)\\
&\quad -\frac{1}{2}\rho\boldsymbol{v}^T\boldsymbol{M}\boldsymbol{Z}\boldsymbol{E}\boldsymbol{E}^T\boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{v} + \frac{1}{2}\rho\boldsymbol{v}^T\underbrace{\frac{1}{c}\boldsymbol{Z}\mathbf{1}\mathbf{1}^T\boldsymbol{Z}^T}_{\boldsymbol{M}}\boldsymbol{Z}\boldsymbol{E}\boldsymbol{E}^T\boldsymbol{Z}^T\underbrace{\frac{1}{c}\boldsymbol{Z}\mathbf{1}\mathbf{1}^T\boldsymbol{Z}^T}_{\boldsymbol{M}}\boldsymbol{v} && (91)\\
&= \frac{1}{2}c(1-\rho)\boldsymbol{v}^T\boldsymbol{M}\boldsymbol{v} + \frac{1}{2}\rho\boldsymbol{v}^T\boldsymbol{Z}\boldsymbol{E}\boldsymbol{E}^T\boldsymbol{Z}^T\boldsymbol{v} && (92)
\end{aligned}$$

Denoting with $\boldsymbol{C}$ a symmetric matrix, with $\boldsymbol{A}$ and $\boldsymbol{B}$ two general matrices and suppose $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, then we have

$$\boldsymbol{v}^T \boldsymbol{C} \boldsymbol{v} \sim \sum_k w_k \chi_1^2, \text{ where } \boldsymbol{w} = \text{eigh}(\boldsymbol{C}) \tag{93}$$

and

$$\begin{aligned} \mathbb{E}\left(\boldsymbol{v}^T \boldsymbol{A} \boldsymbol{v}\right) &= \text{tr}(\boldsymbol{A}) & (94) \\ \text{Cov}\left(\boldsymbol{v}^T \boldsymbol{A} \boldsymbol{v}, \boldsymbol{v}^T \boldsymbol{B} \boldsymbol{v}\right) &= \text{tr}\left(\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{B}^T)\right) & (95) \end{aligned}$$

where we used $\mathbb{E}\left(\boldsymbol{v}^T \boldsymbol{A} \boldsymbol{v} \boldsymbol{v}^T \boldsymbol{B} \boldsymbol{v}\right) = \text{tr}\left(\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{B}^T)\right) + \text{tr}(\boldsymbol{A})\,\text{tr}(\boldsymbol{B})$ from the matrix cookbook section 8.2.4. Using these properties and that $\boldsymbol{M}\boldsymbol{M} = \boldsymbol{M}$[§], we have

- $\phi \sim \sum_{k=1}^m \lambda_k \chi_1^2$ where $\boldsymbol{\lambda} = \text{eigh}(\boldsymbol{\Lambda})$
- $\boldsymbol{\Lambda} = \boldsymbol{E}^T \boldsymbol{Z}^T (\mathbf{I} - \boldsymbol{M}) \mathbf{Z} \boldsymbol{E}$
- $\mathbb{E}(\xi) = 2\text{tr}\left((\boldsymbol{I} - \boldsymbol{M}) \boldsymbol{Z} \boldsymbol{E} \boldsymbol{E}^T \boldsymbol{Z}^T \boldsymbol{M}\right) = 0$
- $\text{Var}(\xi) = 4\text{tr}\left((\boldsymbol{I} - \boldsymbol{M}) \boldsymbol{Z} \boldsymbol{E} \boldsymbol{E}^T \boldsymbol{Z}^T \boldsymbol{M} \boldsymbol{Z} \boldsymbol{E} \boldsymbol{E}^T \boldsymbol{Z}^T\right)$
- $\text{Cov}(\xi, \eta_0) = 4\text{tr}\left((\boldsymbol{I} - \boldsymbol{M}) \boldsymbol{Z} \boldsymbol{E} \boldsymbol{E}^T \boldsymbol{Z}^T \boldsymbol{M}\right) = 0$
- $\text{Cov}(\xi, \phi) = 4\text{tr}\left((\boldsymbol{I} - \boldsymbol{M}) \boldsymbol{Z} \boldsymbol{E} \boldsymbol{E}^T \boldsymbol{Z}^T \boldsymbol{M} (\boldsymbol{I} - \boldsymbol{M}) \boldsymbol{Z} \boldsymbol{E} \boldsymbol{E}^T \boldsymbol{Z}^T (\boldsymbol{I} - \boldsymbol{M})\right) = 0$
- $\text{Cov}(\phi, \eta_0) = 4\text{tr}\left((\boldsymbol{I} - \boldsymbol{M}) \boldsymbol{Z} \boldsymbol{E} \boldsymbol{E}^T \boldsymbol{Z}^T (\boldsymbol{I} - \boldsymbol{M}) \boldsymbol{M}\right) = 0$
- $\eta_0 = \boldsymbol{v}^T \boldsymbol{M} \boldsymbol{v} = \left(\hat{\boldsymbol{z}}^T \boldsymbol{v}\right)^T \left(\hat{\boldsymbol{z}}^T \boldsymbol{v}\right) \sim \chi_1^2$, as $\hat{\boldsymbol{z}} = \frac{\boldsymbol{Z}\boldsymbol{1}}{\|\boldsymbol{Z}\boldsymbol{1}\|}$ is a direction.

Using Eqs (74-77) and introducing

$$\begin{aligned} \boldsymbol{\Lambda}_0 &= \boldsymbol{E}^T \text{diag}(\boldsymbol{x}) \boldsymbol{P} \text{diag}(\boldsymbol{x}) \boldsymbol{E} & (96) \\ \boldsymbol{\alpha} &= \boldsymbol{E}^T \text{diag}(\boldsymbol{x}) \boldsymbol{P} \boldsymbol{x} & (97) \\ c &= \boldsymbol{x} \boldsymbol{P} \boldsymbol{x}^T & (98) \end{aligned}$$

we can rewrite $\boldsymbol{\Lambda}$, $\text{Var}(\xi)$ and τ_ρ as

$$\begin{aligned} \boldsymbol{\Lambda} &= \boldsymbol{\Lambda}_0 - \frac{1}{c} \boldsymbol{\alpha} \boldsymbol{\alpha}^T & (99) \\ \text{Var}(\xi) &= \frac{4}{c} \boldsymbol{\alpha}^T \boldsymbol{\Lambda}_0 \boldsymbol{\alpha} - \frac{4}{c^2} \left(\boldsymbol{\alpha}^T \boldsymbol{\alpha}\right)^2 & (100) \\ \tau_\rho &= c(1 - \rho) + \frac{\rho}{c} \boldsymbol{\alpha}^T \boldsymbol{\alpha} & (101) \end{aligned}$$

[§]from which follows $(\boldsymbol{I} - \boldsymbol{M})(\boldsymbol{I} - \boldsymbol{M}) = (\boldsymbol{I} - \boldsymbol{M})$ and $\boldsymbol{M}(\boldsymbol{I} - \boldsymbol{M}) = 0$

References

[1] Rasmussen, C. E. Gaussian processes in machine learning. In *Advanced lectures on machine learning*, 63–71 (Springer, 2004).

[2] Searle, S. R. & Khuri, A. I. *Matrix algebra useful for statistics* (John Wiley & Sons, 2017).

[3] Hayes, B. J., Visscher, P. M. & Goddard, M. E. Increased accuracy of artificial selection by using the realized relationship matrix. *Genet Res (Camb)* **91**, 47–60 (2009).

[4] Lippert, C. *et al.* FaST linear mixed models for genome-wide association studies. *Nature Methods* **8**, 833–835 (2011).

[5] Lippert, C. *et al.* Fast linear mixed models for genome-wide association studies. *Nat Methods* **8**, 833–5 (2011).

[6] Wu, M. C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *The American Journal of Human Genetics* **89**, 82–93 (2011).

[7] Lippert, C. *et al.* Greater power and computational efficiency for kernel-based association testing of sets of genetic variants. *Bioinformatics* **30**, 3206–14 (2014).

[8] Davies, R. B. The distribution of a linear combination of χ^2 random variables. *Applied Statistics* **29**, 323–333 (1980).

[9] Liu, H., Tang, Y. & Zhang, H. H. A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables. *Computational Statistics & Data Analysis* **53**, 853–856 (2009).

[10] Wu, B., Guan, W. & Pankow, J. S. On Efficient and Accurate Calculation of Significance P-Values for Sequence Kernel Association Testing of Variant Set. *Ann. Hum. Genet.* **80**, 123–35 (2016).

[11] Lee, S. *et al.* Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet* **91**, 224–37 (2012).

[12] Casale, F. P., Rakitsch, B., Lippert, C. & Stegle, O. Efficient set tests for the genetic analysis of correlated traits. *Nat Methods* **12**, 755–8 (2015).

[13] Kostem, E. & Eskin, E. Improving the accuracy and efficiency of partitioning heritability into the contributions of genomic regions. *The American Journal of Human Genetics* **92**, 558–564 (2013).

[14] Kass, R. E. & Raftery, A. E. Bayes Factors. *J. Am. Stat. Assoc.* **90**, 773–795 (1995).

[15] Henderson, C. *Applications of Linear Models in Animal Breeding* (University of Guelph, 1984).

[16] Gauderman, W. J. *et al.* Update on the State of the Science for Analytical Methods for Gene-Environment Interactions. *Am. J. Epidemiol.* **186**, 762–770 (2017).

[17] Kraft, P., Yen, Y. C., Stram, D. O., Morrison, J. & Gauderman, W. J. Exploiting gene-environment interaction to detect genetic associations. *Hum. Hered.* **63**, 111–9 (2007).

[18] Dai, J. Y., Kooperberg, C., Leblanc, M. & Prentice, R. L. Two-stage testing procedures with independent filtering for genome-wide gene-environment interaction. *Biometrika* **99**, 929–944 (2012).

[19] Chatterjee, N., Kalaylioglu, Z., Moslehi, R., Peters, U. & Wacholder, S. Powerful multilocus tests of genetic association in the presence of gene-gene and gene-environment interactions. *Am J Hum Genet* **79**, 1002–16 (2006).

[20] Jiao, S. *et al.* Sberia: set-based gene-environment interaction test for rare and common variants in complex diseases. *Genet Epidemiol* **37**, 452–64 (2013).

[21] Li, S. *et al.* Physical Activity Attenuates the Genetic Predisposition to Obesity in 20,000 Men and Women from EPIC-Norfolk Prospective Population Study. *PLoS Med.* **7**, e1000332 (2010).

[22] Tyrrell, J. *et al.* Gene-obesogenic environment interactions in the UK Biobank study. *Int. J. Epidemiol.* **46**, 559–575 (2017).

[23] 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).

[24] Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).

[25] Lin, X., Lee, S., Christiani, D. C. & Lin, X. Test for interactions between a genetic marker set and environment in generalized linear models. *Biostatistics* **14**, 667–81 (2013).

[26] Lin, X. *et al.* Test for rare variants by environment interactions in sequencing association studies. *Biometrics* **72**, 156–64 (2016).

[27] Tzeng, J.-Y. *et al.* Studying gene and gene-environment effects of uncommon and common variants on continuous traits: a marker-set approach using gene-trait similarity regression. *Am J Hum Genet* **89**, 277–88 (2011).

[28] Zhao, G., Marceau, R., Zhang, D. & Tzeng, J.-Y. Assessing gene-environment interactions for common and rare variants with binary traits using gene-trait similarity regression. *Genetics* **199**, 695–710 (2015).

[29] Hoerl, A. E. & Kennard, R. W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12**, 55–67 (1970).

[30] Casale, F. P., Horta, D., Rakitsch, B. & Stegle, O. Joint genetic analysis using variant sets reveals polygenic gene-context interactions. *PLoS genetics* **13**, e1006693 (2017).

[31] 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).

[32] Crawford, L., Zeng, P., Mukherjee, S. & Zhou, X. Detecting epistasis with the marginal epistasis test in genetic mapping studies of quantitative traits. *PLOS Genet.* **13**, e1006869 (2017).

[33] Listgarten, J. *et al.* A powerful and efficient set test for genetic markers that handles confounding. *Bioinformatics* (2013).

[34] Rao, C. R. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. In *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 44, 50–57 (Cambridge University Press, 1948).

[35] Zhernakova, D. V. *et al.* Identification of context-dependent expression quantitative trait loci in whole blood. *Nature genetics* **49**, 139–145 (2017).