Published in final edited form as:
Nat Genet. 2019 January ; 51(1): 180–186. doi:10.1038/s41588-018-0271-0.

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

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

[1]Wellcome Sanger Institute, Wellcome Genome Campus, CB10 1SD Hinxton, Cambridge, UK [2]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, CB10 1SD, Hinxton, Cambridge, UK [3]University of Cambridge, Cambridge, UK [4]Microsoft Research New England, Cambridge, Massachusetts, USA [5]A full list of consortium members and affiliations is presented 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), 69120, Heidelberg, Germany

Abstract

Different exposures, including diet, physical activity, or external conditions can contribute to genotype-environment interactions (GxE). Although high-dimensional environmental data are increasingly available, and multiple exposures have been implicated with GxE at the same loci, multi-environment tests for GxE are not established. Here, we propose the structured linear mixed model (StructLMM), a computationally efficient method to identify and characterize loci that interact with one or more environments. After validating our model using simulations, we apply StructLMM to body mass index in UK Biobank, where our model yields previously known and novel GxE signals. Finally, in an application to a large blood eQTL dataset, we demonstrate that StructLMM can be used to study interactions with hundreds of environmental variables.

[#]Correspondence should be addressed to Inês Barroso (ib1@sanger.ac.uk), Oliver Stegle (oliver.stegle@embl.de).
[*]These authors contributed equally.

URLs

Haplotype Reference panel, http://www.haplotype-reference-consortium.org/site; Phase 3 1,000 Genomes reference panel, http://grch37.rest.ensembl.org

Data availability

The BIOS RNA data can be obtained from the European Genome-phenome Archive (EGA; accession/EGAS00001001077). Genotype data are available from the respective biobanks.

Code availability

StructLMM is available from https://github.com/limix/struct-lmm and is supported within the LIMIX framework51 at https://github.com/limix/limix. For tutorials and illustrations on how to use the model, see http://struct-lmm.readthedocs.io.

Author contributions. R.M., F.P.C., I. B., & O.S. conceived the method. R.M., F.P.C., D.H. implemented the methods. R.M., F.P.C., M.J.B. analysed the data. L.F. provided data resources. R.M., F.P.C., I.B., & O.S. interpreted results and wrote the paper.

Competing interests

F.P.C. was employed at Microsoft while performing the research.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Introduction

Large population cohorts that combine genetic profiling with deep phenotype and environmental data, including diet, physical activity and other lifestyle covariates, have fostered interest to study genotype-environment interactions (GxE). Already, such analyses have identified GxE for different traits in humans, including disease risk1,2, and molecular traits3,4.

Established GxE methods test for interactions between a single environmental variable and individual genetic variants5 (Fig. 1a). Recent extensions enable assessing GxE across sets of genetic variants, either using genetic risk scores6 or variance component tests7–9. Whilst there is evidence that multiple environments can interact with a single genetic locus to influence phenotypes, for example a number of environments have been shown to alter the effect of *FTO* on BMI, including physical activity10–13, diet12–15 and smoking12, there are no robust methods for the joint GxE analysis of multiple environmental variables.

Multivariate GxE tests can have power advantages, in particular to identify interactions that are simultaneously driven by multiple environments, or because combinations of multiple environmental variables act as proxy for unobserved drivers of GxE. Additionally, joint tests reduce the multiple testing burden. Thus, as increasingly high-dimensional environmental data are available in population cohorts, and given the desire to fully understand the impact of multiple environments in complex traits and diseases, there is a growing need for multi-environment GxE tests.

Here, we present the structured linear mixed model (StructLMM), a variance component test to identify and characterize GxE interactions with multiple environments. Our model can handle hundreds of environmental variables and it can be applied to large cohorts of hundreds of thousands of individuals.

Results

Conventional linear mixed models (LMMs) are used to test for associations with constant genetic effect sizes across individuals in the population, also called persistent genetic effects. Covariates and additional random effect components are included to account for population structure, environment, or other additive (confounding) effects. StructLMM extends the LMM framework by modelling heterogeneity in effect sizes due to GxE

$$\boldsymbol{y} = \underbrace{\boldsymbol{X}\boldsymbol{b}}_{\text{covariates}} + \underbrace{\boldsymbol{x}\beta_{\text{G}}}_{\text{persistent G}} + \underbrace{\boldsymbol{x} \odot \boldsymbol{\beta}_{\text{G}\times\text{E}}}_{\text{G}\times\text{E}} + \underbrace{\boldsymbol{e}}_{\text{environment}} + \underbrace{\boldsymbol{\psi}}_{\text{noise}} . \quad (1)$$

Here, β_{G} denotes the effect size of a conventional persistent genetic effect component and $\boldsymbol{\beta}_{\text{G}\times\text{E}} = \left[\beta^1_{\text{G}\times\text{E}}, \ldots, \beta^N_{\text{G}\times\text{E}}\right]^{\text{T}}$ is a vector of per-individual effect sizes to account for heterogeneous genetic effects, which follows a multivariate normal distribution, $\boldsymbol{\beta}_{\text{G}\times\text{E}} \sim N(\mathbf{0}, \sigma^2_{\text{G}\times\text{E}}\,\boldsymbol{\Sigma})$. Depending on the functional form of the environmental covariance $\boldsymbol{\Sigma}$, this model can account for different types of GxE, for example hierarchies of discrete

environmental groups, or as considered here, GxE effects based on a set of continuous and discrete environmental covariates (Fig. 1b, c). The same environmental covariance is also used to account for additive environmental effects, $\boldsymbol{e} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. StructLMM is technically related to existing variance component tests for rare variants16 and epistasis17; see Supplementary Note for a comparison to alternative methods.

Using the multi-environment model defined above (Eq. (1)), we propose a score test to identify loci with significant GxE interaction effects. In addition, the same framework can be used to define a joint association test that accounts for the possibility of heterogeneous effect sizes due to GxE, which generalizes previous two degrees of freedom single-environment association tests5,18. Both tests are computationally efficient, enabling genome-wide analyses using hundreds of environmental variables on cohorts of hundreds of thousands of individuals. The model facilitates different analyses to characterise GxE effects at individual loci, including estimation of the fraction of genetic variance explained by GxE (ρ, Methods), and estimating per-individual allelic effects based on environmental profiles in the population (Fig. 1d), thus identifying individuals at increased/decreased trait risk. Finally, StructLMM can be used to explore which environments are most relevant for GxE, by comparing models that contain all environmental factors and models with environmental variables removed (Fig. 1e). See Methods for a full derivation.

Model validation using simulated data

Initially, we considered simulated data using genotypes from the 1000 Genomes project19 to assess the statistical calibration and power of StructLMM. To mimic environmental distributions as observed in real settings, we simulated GxE based on 60 environmental covariates from UK Biobank, including physical activity, diet, and other lifestyle factors (Methods). We varied the sample size of the simulated population, the magnitude of GxE effects, the number of driving environments for GxE, and other parameters (Supp. Table 1).

First, we confirmed the statistical calibration of the StructLMM interaction test (referred to as StructLMM-int), either considering phenotypes simulated without any genetic effects (Fig. 2a, Supp. Fig. 1a,b) or simulated from a persistent effect model without interactions (i.e. the null model of StructLMM-int; Supp. Fig. 1a,b).

Next, we simulated phenotypes with variable fractions of the genetic variance explained by GxE (ρ, Methods), and assessed power of StructLMM-int. For comparison, we also considered a single-environment one-degree of freedom fixed effect test (SingleEnv-Renv-int, Supp. Table 2; e.g. Gauderman et al.18, Bonferroni adjusted for the number of environments, Methods), using the same random effect component (as for StructLMM) to account for additive environmental effects under the null.

The power of both tests increases as the fraction of the genetic effect explained by GxE (ρ) increases, noting that StructLMM-int is substantially better powered than the SingleEnv-Renv-int test (Fig. 2b). As a second parameter, we varied the number of active environments that contribute to GxE but used all 60 environmental variables during testing. The results of this analysis show that StructLMM-int increasingly outperforms the corresponding

SingleEnv-Renv-int GxE test as the number of active environments increases (Fig. 2c, Supp. Fig. 2b).

We considered a number of additional settings, including varying the total number of observed environments, the fraction of phenotypic variance explained by additive environmental effects and simulating interaction effects using environments that are not included at the testing stage. The latter corresponding to GxE effects driven by environments for which there are no measurements available, a scenario that is likely to occur in practice. We also considered settings where the environments were heritable themselves, varied the extent of distributional skew and considered binary environments with different frequencies. Across all settings, StructLMM-int had consistent power advantages over alternative methods and remained calibrated (Supp. Fig. 2-4).

For the same settings, we also considered the StructLMM joint association test, which accounts for the possibility of heterogeneous effect sizes due to GxE, and compared it to a two degrees of freedom single-environment test using fixed effects (e.g. Kraft et al.5; Bonferroni adjusted for the number of environments, Methods; Supp. Table 2), as well as to conventional association tests that only model persistent effects (Supp. Table 2). In these experiments (Supp. Fig. 2-4), the StructLMM joint association test yielded similar power advantages as StructLMM-int when testing for interactions, indicating that StructLMM can be useful to discover additional associations, in particular for variants with strong GxE (Supp. Fig. 2a).

Finally, we considered alternative implementations of interaction and association tests (Supp. Table 2), using fixed effects to account for additive environment instead of a random effect component, which yielded near-identical results (Supp. Fig. 2). We also note that multi-environment GxE tests can in principle be implemented based on fixed effect tests with as many degrees of freedoms as environments (Supp. Table 2). However, we observed that such tests were not always calibrated (Supp. Fig. 1b), in particular for large numbers of environments, and in addition had lower performance (Supp. Fig. 1c).

Taken together, these results show increased power and robustness of StructLMM compared to existing methods, in particular when large numbers of environments drive GxE interaction effects, as might be expected to occur for the majority of complex traits and diseases.

Application to data from UK Biobank

Initially, we applied StructLMM-int to test for GxE interactions at 97 variants (corresponding genes as annotated by GIANT20) that have previously been linked to BMI using independent data20. We considered 252,188 unrelated individuals of European ancestry, for which BMI and 64 lifestyle covariates, similar to those used in Young *et al.*13 (12 diet-related factors, three factors linked to physical activity and six lifestyle factors, modelled as gender-adjusted and age-adjusted, Methods, Supp. Fig. 5,6), were available in the full release of UK Biobank21. StructLMM-int identified four significant GxE effects ($\alpha < 0.05$, Bonferroni adjusted), whereas a single-environment one-degree of freedom fixed effect test (SingleEnv-Renv-int) identified only two of these interactions (Fig. 3a, Supp. Fig. 7, Supp. Table 3). Among the loci identified by StructLMM-int was the *FTO* locus

(rs1241085, ρ=0.14, Supp. Fig. 8a), which has previously been implicated with GxE for multiple environments10,12–14, *MC4R* (Fig. 3b) for which an interaction with physical activity in females aged 20-40yrs has been previously suggested (P-adj=0.025 reported in12), *SEC16B* (Supp. Fig. 8b), for which secondary analyses provided some evidence for an interaction (P=0.025) with physical activity in Europeans11 and in a separate study in Hispanics22 and *PARK2* (Supp. Fig. 8c), a gene that has been linked to time-dependent variation in BMI23. StructLMM also enhanced the significance of interactions identified by both tests (P StructLMM-int = 4.23×10^{-16} versus P-adj SingleEnv-Renv-int = 6.76×10^{-6} and P StructLMM-int = 1.15×10^{-4} versus P-adj SingleEnv-Renv-int = 4.48×10^{-4} for *FTO* and *SEC16B* respectively). Larger differences in the number of discoveries were observed at more lenient thresholds, e.g. 11 versus six loci with GxE at FDR<5% (Benjamini-Hochberg adjustment, Supp. Table 3).

We also considered additional fixed effect interaction tests, including multi-environment GxE tests based on fixed effects, which identified fewer interactions than StructLMM-int (N=2 versus N=4; $\alpha < 0.05$), as well as alternative implementations of the single environment interaction test, which consistent with the results on simulations, yielded near-identical results to SingleEnv-Renv-int (Supp. Fig. 9,10, Supp. Table 3).

Finally, as an alternative filtering strategy, we applied the same interaction tests to 17,606 variants with significant persistent associations with BMI in UK Biobank ($P < 5 \times 10^{-8}$; LMM-Renv). StructLMM-int identified 23 loci with GxE interactions (FDR<5%; Benjamini-Hochberg adjusted, +/-500kb, $r^2 > 0.1$), including *SEC16B*, *MC4R* and *FTO*, compared to at most 11 loci identified by alternative methods (Supp. Fig 11, Supp. Table 3).

The StructLMM framework can also be used to test for associations while accounting for the possibility of effect size heterogeneity due to GxE. To explore this, we applied the StructLMM joint association test to BMI, using low-frequency and common variants (imputed variants, MAF>1%, 7,515,856 variants in total) and the same set of 64 lifestyle covariates as considered in the interaction analysis. For comparison we also considered an LMM using the same random effect component to account for additive environmental effects as in StructLMM (LMM-Renv) and a linear model without accounting for additive environment (LM). Although the choice of null model can have a large impact on loci discovery ($P < 5 \times 10^{-8}$, +/-500kb, r2>0.1, LMM-Renv 327 loci of which 14.37% were not detected by the LM, LM 379 loci of which 25.59% were not detected by LMM-Renv; Supp. Table 4, Supp. Fig. 12), StructLMM identified 23 loci that were not detected by other methods (351 unique loci in total; Fig. 3c, Supp. Table 4, Supp. Fig. 13-16), indicating that the StructLMM joint association test can be used to identify additional loci with a strong GxE component. One such locus lies in the *ADAMTSL3* gene (*rs4842838*, P StructLMM = 9.35×10^{-10}, P LMM = 3.83×10^{-7}, P LM = 2.37×10^{-5}), which codes for a glycoprotein24. Other variants within this gene have been linked to BMI-related traits, including lean body mass25, waist circumference26 and hip circumference adjusted for BMI27.

Once GxE loci have been identified, StructLMM can be used for the interpretation of these effects, and in particular to estimate per-individual allelic effects based on environmental profiles to identify individuals with increased/decreased trait risk (Fig. 4a). We confirmed

the robustness of these estimates using hold-out validation, providing further evidence for possible opposite directions of effect at *PARK2* (Supp. Fig. 17). To explore which environmental variables are most relevant for individual GxE signal, we calculated Bayes factors (BF) between the full model and models with individual environmental exposures removed (Supp. Fig. 18), identifying between 20 and 25 environments with putative GxE effects (BF>0). Since the environments are not independent of one another, we used backward elimination based on BF between the full model and models with increasing numbers of environments removed. These analyses identified physical activity measures for females (no evidence for males) as contributing to GxE at *MC4R*, in agreement with12, but also yielded a number of additional environments (Fig. 4b, Supp. Fig. 18). For all loci, we consistently observed that multiple environments contribute to GxE but there is evidence of differences in the GxE architecture, with *FTO* being associated with the largest number of environments whilst *SEC16B* and *PARK2* were associated with a smaller number of environments (Supp. Fig. 18). Differences in the environments that contribute to GxE effects were also apparent when correlating per-individual allelic effect size estimates across loci (Supp. Fig. 19).

Identification of eQTL interactions with cellular state

As a second application, we considered a gene expression dataset28 to illustrate how StructLMM can be used to identify context-dependent regulatory effects on gene expression, for example due to external stimuli4 or differences in cell type composition29, using hundreds of environment covariates. Insights into context-dependent genetic regulation of gene expression are important to identify disease-relevant cell types and molecular pathways30–32.

We reanalysed a large whole-blood expression dataset comprising of 2,040 genotyped individuals profiled with RNA-seq28 (Methods) and applied StructLMM-int to test for cell-context interactions at *cis* expression quantitative trait loci (eQTL). Following Zhernakova *et al.*28, we considered gene expression levels both as phenotypes but also as proxy (environmental) variables, which can tag variation in blood cell composition and other factors across individuals. Specifically, we considered a set of 443 highly variable genes as environmental variables in our analysis (Methods).

Initially, we applied a linear model to identify lead *cis* eQTL variants for 23,506 expressed genes (within plus or minus 250 kb from the centre of the gene, Methods). Next, we applied StructLMM-int to test for cell-context interactions at lead variants for each of these genes. The model produced calibrated P values despite the large number of environments (Supp. Fig. 20), identifying 3,483 eQTL with a cell-context interaction (FDR<5%, termed interaction eQTL; Supp. Table 5). Although globally, interactions with cell-context tended to explain small fractions of the *cis* genetic variance on gene expression ($\rho<0.2$, for 68.0% of interaction eQTL, Fig. 5a), GxE explained more variance than persistent genetic effects for 532 genes ($\rho>0.5$, for 15.3% of interaction eQTL). We also compared StructLMM-int to alternative multi-environment interaction tests based on fixed effects, which were markedly less robust and identified fewer interaction eQTL (Supp. Fig. 20). Similarly, we compared the discovered interaction eQTL to results from a stepwise procedure that was used to

identify interaction eQTL in the primary analysis of the same data28 (see Supplementary Note for details), which yielded markedly fewer interactions (3,372 versus 1,841 interaction eQTL, considering StructLMM and the approach in 28; FDR<5%; Supp. Fig. 20; considering 17,952 genes assessed in both studies). Finally, we considered alternative approaches to normalise the expression data (Methods), thereby assessing potential biases due to gene-exposure correlations and distributional skew of counts-based gene expression profiles. These results indicated that StructLMM is robust to both potential sources of bias (Supp. Fig. 21).

Next, we overlapped the interaction eQTL with risk variants from the NHGRI-EBI GWAS catalog V1.0.133, identifying 64 putative colocalisation events (r^2>0.8 between lead eQTL and GWAS variants, Supp. Fig. 22, Methods), including GWAS variants for autoimmune diseases, infectious diseases and blood cell traits (Supp. Table 6, Supp. Dataset 1,2). Notably, 46 of these interaction eQTL were not reported in the primary analysis28. One example is an interaction eQTL for *CTSW* expression (Fig. 5b-d, P StructLMM-int = $2.2x10^{-15}$, ρ=0.12), which is in linkage disequilibrium (LD) with a risk variant for Crohn's disease *rs568617* (r^2=0.98, Supp. Fig. 23). To investigate the molecular pathways that are associated with this interaction, we stratified the population into strata with the smallest and largest allelic effects as estimated using StructLMM (Fig. 5b,c), and tested for pathways that were enriched among differentially expressed genes between these groups (Methods). This identified *T cell selection (GO:0045058), positive T cell selection (GO:0046632)* and *positive regulation of interleukin-17 secretion* (*GO:0032740*) as the top three processes for this interaction eQTL (Fisher exact test; see Supp. Table 6 for genome-wide enrichment results), GO terms that are consistent with known roles of IL-17 producing $CD4^+$ T cells in the pathogenesis of inflammatory bowel disease, including Crohn's disease34.

Taken together, this analysis demonstrates the broad applicability of StructLMM, including in settings with very large numbers of environmental factors.

Discussion

We propose a method based on variance component tests to identify GxE interactions using multiple environments. Conceptually, our approach is related to set tests for groups of variants, but instead of aggregating across multiple genetic variants, StructLMM jointly models multiple environmental variables to identify GxE interactions. Compared to conventional single and multiple degrees of freedom fixed effect GxE tests, this approach enjoys power advantages (Fig. 2, Supp. Fig. 1-4), and yields increased robustness, in particular when analysing large numbers of environmental variables (Supp. Fig. 1, 20).

We applied StructLMM to data from UK Biobank to assess GxE at 97 GIANT variants associated with BMI, confirming established GxE effects at *FTO*, and we identified, for the first time, three additional GxE signals at stringent thresholds (FWER<5%, Fig. 3a), some of which confirm prior evidence11,12,14,15,22,23. More lenient FDR-based significance thresholds, as frequently employed for GxE analyses6,12, yielded 11 GIANT variants with evidence for GxE (FDR<5%; Benjamini-Hochberg adjusted; Supp. Table 3), and a genome-wide analysis based on all variants that are associated with BMI identified 23 loci with

significant GxE effects (Supp. Fig. 11, Supp. Table 3). We also show that the same framework can be used to test for associations, demonstrating that accounting for heterogeneity in effect sizes can identify additional loci, similarly to previously reported benefits of 2-df fixed effect tests5.

In addition to offering power advantages, StructLMM yields per-individual allelic effect size estimates that reflect GxE. We have shown that this allows for different downstream analyses, including the identification of individuals with increased or decreased genetic risk. This would be of particular interest in the complex disease field as it may provide further explanation as to why individuals who share the same set of risk variants may have different outcomes in longitudinal follow-up. In particular, identifying sets of environments that may decrease disease risk for individuals carrying the same genetic burden may provide useful avenues for targeted disease prevention. We also explore which environments are putative drivers of the observed GxE effects. However, such downstream analyses, when using the same dataset for discovery, should be interpreted with caution. Ultimately, independent validation cohorts will be required to confirm such findings.

As a second use case, we applied StructLMM to test for cell-context interactions in a large blood eQTL study, where the same modelling principles enabled the identification of context-specific eQTL. Several of these interaction eQTL colocalised with GWAS variants and the marker genes of the cellular environments that underlie these interaction effects could be connected to plausible molecular pathways (Supp. Table 6).

Although we found that StructLMM is a robust and powerful alternative to conventional linear interaction tests, our approach is not free of limitations. First, there are general challenges when analysing GxE that although not specific to our model need to be taken into consideration. One such challenge are environmental variables that are themselves heritable. Accounting for heritable covariates in association tests can lead to spurious associations due to collider bias35. Our results indicate that interaction tests are more robust to such correlations (Supp. Fig. 3). However, gene exposure associations alter the interpretation of interactions, reflecting epistatic relationships between genetic factors. A second generic challenge is the selection of candidate variants for GxE tests. To reduce the multiple testing burden, we selected variants that have persistent effects on the phenotype. However, the fact that our association test identifies novel loci with strong GxE (ρ) if applied genome-wide indicates that this filter is not optimal.

Among more specific limitations and areas of future work for StructLMM, we note the computational requirements of the model are more demanding than conventional LMMs, despite scaling linearly with the number of individuals. A second potential limitation is that StructLMM does not currently enable accounting for relatedness. Although the model has an additive random effect component, it is currently used to model additive environmental effects. Generalizations to simultaneously account for a relatedness could be considered, for example using suitable low-rank approximations36 or other speed-ups to retain scalability to large sample sizes. Finally, while StructLMM can in principle be used in conjunction with any environmental covariance, we have limited our attention to linear covariances. The model could be extended to account for non-linear interactions, for example using

polynomial covariance functions. Future developments in this direction will be increasingly valuable as larger cohort sizes enable detecting higher order interaction effects.

Online Methods

The Structured Linear Model

A conventional linear mixed model (LMM) to test for associations can be cast as

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{x}\beta + \boldsymbol{u} + \boldsymbol{\psi},$$

where β is the focal variant effect size, $\boldsymbol{X}$ the fixed effect design matrix of k covariates and $\boldsymbol{b}$ the corresponding effect sizes. The variable u denotes additive (confounding) factors, and ψ denotes iid noise. The random effect component u and the noise vector ψ follow multivariate normal distributions, $\boldsymbol{u} \sim N(\mathbf{0}, \sigma_e^2 \boldsymbol{\Sigma_u})$ and $\boldsymbol{\psi} \sim N(\mathbf{0}, \sigma_n^2 \boldsymbol{I})$, where the covariance matrix $\boldsymbol{\Sigma u}$ reflects the covariance of population structure, environment or other (confounding) factors. Association tests for non-zero effects of the focal variant correspond to alternative hypothesis $\beta \neq 0$.

StructLMM generalizes the conventional LMM for association testing by introducing per-individual effect sizes due to GxE

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{x}\beta + \boldsymbol{x} \odot \boldsymbol{\beta}_{\mathrm{G\times E}} + \boldsymbol{u} + \boldsymbol{\psi}, \quad (1)$$

where $\beta_{\mathrm{G\times E}}$ is a per-individual allelic effects vector, which follows a multivariate normal distribution with environment covariance $\boldsymbol{\Sigma}$:

$$\boldsymbol{\beta}_{\mathrm{G\times E}} \sim N(\mathbf{0}, \sigma_{GxE}^2 \boldsymbol{\Sigma}). \quad (2)$$

The covariance $\boldsymbol{\Sigma}$ captures heterogeneity in allelic effects in the population and is estimated using a linear covariance function based on a set of observed environmental variables, where we assume $\boldsymbol{\Sigma}_{\mathrm{u}} = \boldsymbol{\Sigma}$. If collider bias35 is a concern, non-heritable environmental variables should be selected. Non-linear environmental effects can be modelled by combining observed environmental variables (e.g., effects from environments × age or environments × gender); see Supplementary Note.

Statistical Testing

Based on Eq. (1), we define an interaction test ($\sigma^2_{\mathrm{GxE}} > 0$) where persistent genetic and additive environment effects are accounted for in the null model; and an association test $\sigma^2_{\mathrm{GxE}} > 0$ and $\boldsymbol{\beta} \neq 0$), which jointly tests for associations while accounting for the possibility of heterogeneous genetic effects due to GxE. Both tests are implemented as efficient score tests, similar to the approach in SKAT and SKAT-O37,38, with linear complexity in the number of individuals (Supplementary Note).

Estimation of ρ

Estimates of the fraction of the genetic variance explained by GxE (ρ) can be obtained from maximum likelihood estimates of the model in Eq. (3)

$$\rho = \frac{\mathrm{Var}^{\mathrm{GxE}}}{\mathrm{Var}^{\mathrm{G}} + \mathrm{Var}^{\mathrm{GxE}}} \quad (5)$$

with $\mathrm{Var}^{\mathrm{G}}$ denoting the fraction of the variance explained by persistent effects, and $\mathrm{Var}^{\mathrm{GxE}}$ variance due to GxE.

Exploring the most relevant environments for GxE

Bayes factors between the full model and models with individual environments or sets of environments removed from the environmental covariance $\mathbf{\Sigma}$ (Supplementary Note) can be used to assess the relevance of environments.

Estimation of per-individual allelic effects

Per-individual (i.e. for each environment state) allelic effects can be estimated using BLUP39. Additionally, the model yields posterior estimates of the realisation of the unobserved environmental state that explains the GxE effect (see Supplementary Note).

Simulations

Simulation procedure overview—Simulations were based on genotypes of European individuals from the 1000 Genomes project19 (phase 1, 1,092 individuals, 379 Europeans), considering 103,527 variants on chromosome 21 (minor allele frequency >=2%). Following40,41, synthetic genotypes of unrelated individuals were generated for different sample sizes, while preserving the population structure of the seed population (see9). We considered 33 environmental exposures using empirical environmental covariates from 70,282 UK Biobank individuals (based on the Interim release), augmented using element wise interactions with gender and age, resulting in 100 environmental variables. These environmental variables were preprocessed as in the UK Biobank analysis (see below) and randomly assigned to synthetic genotypes. See Supplementary Note for details.

Assessment of statistical calibration—Statistical calibration of different tests was assessed using phenotypes simulated from an empirical null model, considering i) no genetic effect (Fig. 2a, Supp. Fig. 1a,b) and ii) simulated persistent genetic effects (100 persistent genetic effect variants, no GxE interactions, Supp. Fig. 1a,b). Calibration was assessed using QQ plots and genomic control ($\lambda_{\mathrm{GC}} = \frac{log_{10}(m)}{log_{10}(0.5)}$; m is the median P value), based on P values from chromosome 21 pooled across 100 repeat experiments.

Power simulations—Phenotypes with GxE interactions were simulated, varying the fraction of variance explained by GxE, the number of active environments and other parameters (Supp. Table 1, Supplementary Note). We also studied the effect of gene-

exposure correlations (Supp. Fig. 3, Supplementary Note) and considered synthetic environments to assess the effect of (rare) binary environmental variables (Supp. Fig. 4, Supplementary Note). We considered 1,000 repeat experiments for each setting, randomly selecting a segment of approximately 2 Mb from chromosome 21 and simulating GxE effects from one causal variant. Power at 1% FWER (Bonferroni adjusted across variants) was assessed considering variants in linkage disequilibrium with selected true causal variants ($r^2 \geq 0.8$) as true positives, reporting average power across repeat experiments (which is either 1 or 0).

Comparison methods—We compared StructLMM to alternative single- and multi-environment models, as well as standard genetic association tests. For interaction tests, we considered alternative single-environment GxE interaction tests i) using random effect (SingleEnv-Renv-int) or ii) fixed effect (SingleEnv-Fenv-int) components to account for additive environmental effects due to all enviornments, and finally iii) an additive single-environment fixed effect term based on the specific environment considered in the GxE test only (SingleEnv-Senv-int). The same models were considered to test for associations, using a 2 df statistical test5 (SingleEnv-Renv, SingleEnv-Fenv, SingleEnv-Senv, respectively). Additionally, for association tests, we considered linear association tests, again either using a multi-environment random effect for additive environment (LMM-Renv) or a multi-environment fixed effect (LM-Fenv) component to account for additive environmental effects, as well a linear model with no additive environment effect term (LM). All tests were implemented using LRT, considering Bonferroni adjusted minimum P value per variant across environments for single-environment models. Finally, we assessed the performance of fixed-effect multi-environment interaction- and association tests, again considering either random or fixed additive environment components based on all observed environments, considering either an LRT or score test. Performance was assessed using the average area under the curve (AUC) across repeat experiments (using true positive definitions as for power), computed in the range FPR<0.10 and normalised to the 0-1 range such that 0 corresponds to chance performance and 1 is the performance of an ideal model. See Supp. Table 2 for an overview and Supplementary Note for details.

Analysis of BMI in UK Biobank

This research has been conducted using the full release of the UK Biobank Resource (Application 14069)21. The UK Biobank study has approval from the North West Multi-Centre Research Ethics Committee and all participants included in the analyses provided informed consent to UK Biobank.

Data pre-processing—BMI phenotype data is 'Instance 0' of UK Biobank data field 21001. Individuals with missing BMI data were discarded from the analysis and BMI log transformed13,42. Following13, we considered 21 lifestyle covariates as environments, discarding individuals with outlying or missing environmental variables (Supplementary Note). We further discarded individuals of non-British ancestry as well as related individuals. After filtering and QC on the BMI phenotype, genotype and the environmental variables, we obtained a set of 252,188 individuals for analysis. Principal component for

population structure adjustment were calculated using flashpca version 2.043 using 147,604 variants as indicated by the field 'in_PCA' from the released marker QC file.

Genotype data—We used genotypes that were imputed with the HRC panel (build GRCh37). We performed QC of the remaining imputed variants on the fly, using a fast bgen reader, implemented as part of StructLMM, treating genotype-sample pairs with low imputation accuracy (max. probability <0.5) as missing, and discarding variants with missingness>5%, MAF<1%, HWE $P<1x10^{-6}$ and INFO score $r^2<0.4$ (based on the UK Biobank imputation MAF and info file). Genotype dosages were calculated using available probabilities (including genotype-sample pairs with low imputation accuracy) and mean imputation used for any genotype-sample pairs with missing data. 7,515,856 variants passed these filters.

Environmental Covariance and Covariates—To generate the environment matrix **E,** we augmented all 21 environmental variables described above (excluding age) by gender and age, by multiplying the continuous age vector, the binary male indicator vector and the binary female indicator vector with each of the 21 environment variables, which resulted in 63 covariates. The environmental covariance was estimated based on standardised environmental variables (not including zero values due to augmentation when mean adjusting) followed by per-individual standardisation (Supp. Fig. 5, 6; see Supplementary Note for full details). In all analyses, a mean vector, genotype chip, gender, age^2, age^3, gender x age, gender x age^2, gender x age^3, 10 genetic principal components were included as covariates.

Calibration of interaction and association tests—To validate the tested methods and QC procedures, we assessed the empirical calibration using permuted genotype variants (173,297 variants) on chromosome 20 (Supp. Fig. 7).

Interaction testing—We considered 97 GIANT variants previously associated with BMI20 to test for GxE interactions using StructLMM-int as well as single-environment fixed effect interaction tests (SingleEnv-Renv-int and SingleEnv-Senv-int, 1 df, Supplementary Note), and a multi-environment fixed-effect based interaction test (64 df, MultiEnv-Renv-LRT-int, Supplementary Note). Variants with significant GxE were reported at FWER 5% (i.e., $P < 0.05/97$), and alternatively using a more lenient threshold at FDR<5% (Benjamini-Hochberg adjustment 44; Supp. Table 3).

We also selected the 17,606 variants with LMM-Renv P values $<5x10^{-8}$ and compared results using StructLMM-int to those from single-environment fixed effect interaction tests (SingleEnv-Renv-int, Supplementary Note), and a multi-environment fixed-effect based interaction test (64 df, MultiEnv-Renv-LRT-int, Supplementary Note). This filter is valid as LMM-Renv corresponds to the null model of both StructLMM-int and SingleEnv-Renv-int. Variants with significant GxE were reported at FDR<5% (Benjamini-Hochberg44 adjustment), followed by LD clumping to define independent loci: we iteratively (i) selected the most significant variant (using the FDR-adjusted P values) and (ii) removed all variants in LD ($r^2 > 0.1$) within +/-500kb, until no variant was left, resulting in 23, 11 and 9 clumps (loci), respectively (Supp. Table 3).

Association testing—We used StructLMM, LMM-Renv and LM for genome-wide association analyses, reporting significant associations at $P<5x10^{-8}$, for which ρ was estimated using StructLMM (Fig. 3c, Supp. Table 4). LD clumping was used to define independent loci identified by each of the three methods; we iteratively (i) selected the most significant variant and (ii) removed all variants in LD ($r^2 > 0.1$) within +/-500kb, until no variant was left, resulting in 351, 327 and 379 loci, respectively. We compared the methods pairwise, identifying loci found by only one method, by calculating the LD (r^2) between significant variants within a clump identified by one method and significant variants from the other method that lie within +/-500kb, resulting in 32 and 16 loci (StructLMM and LMM-Renv), 65 and 98 (StructLMM and LM) and 47 and 97 (LMM-Renv and LM). We also compared the genome-wide results of StructLMM to those from a multiple degrees of freedom (65 df) fixed effect association test MultiEnv-Renv-LRT, again using genome-wide significance thresholds of $5x10^{-8}$ and estimating ρ for all significant variants (Supp. Fig. 9).

Per-individual allelic effect estimation—We performed in-sample estimation of the allelic effect (see above) for each of 252,188 individuals at each of the four interaction loci (FWER 5%; Fig. 4a). Allelic effects were assessed out of sample by randomly splitting the cohort into training and test fractions, to assess out-of-sample predictions (Supp. Fig. 17; Supplementary Note). To assess whether the same set of individuals are at the extreme ends of the effect size spectrum across multiple interaction variants (5% FDR-adjusted), we computed the squared Spearman's correlation coefficient and then used ward hierarchical clustering (Supp. Fig. 19).

Explorative analysis of driving environments—We explored which environments had putative effects on GxE by comparing the log marginal likelihood of the full model to models with individual or sets of environments excluded. We initially assessed the relevance of individual environments based on the log(Bayes factor) of removing single environments (Supp. Fig. 18). To account for correlations between environments, we also used a backwards elimination procedure (Supplementary Note), greedily removing environments until there is evidence that we have selected a full set of environments that can drive the observed GxE effect (Fig. 4b, Supp. Fig. 18).

Analysis of cell-context eQTL in a large blood cohort

Genotype data pre-processing—We used freeze one from the BIOS consortium (EGA; accession/EGAS00001001077), and analysed 2,040 samples for which genotypes and QC-passing RNA-seq data were available. Processed genotype and expression data were taken from the primary analysis28. Imputed genotypes (from the four biobanks CODAM, LifeLines, the Leiden Longevity Study and the Rotterdam Study) were merged to perform a mega-analysis, as opposed to the meta-analysis in the original paper. After merging, we performed joint QC of the genetic variants, retaining variants that met the following conditions: MACH-R2>0.5, call-rate>0.95, $HWE>10^{-4}$ and MAF>5%, resulting in 5,683,643 variants for analysis.

Ethical approval—The ethical approval for this study lies with the individual participating cohorts (CODAM, LLD, LLS and RS)45–48.

Expression data—The expression data was taken from the original quantifications (after TMM normalisation) and we selected features that were identified in at least 10% of the samples, resulting in 23,506 expressed genes for analysis. Expression values were quantile normalised and we used ENSEMBL 71 as gene annotation.

Environmental Covariance—We used gene expression levels to build the StructLMM covariance capturing cell-type composition and other sources of cell-context heterogeneity. Specifically, we considered a set of highly variable proxy genes, identified through a two-step procedure: (i) we selected the top 25% most variable genes based on the interquartile range of non-quantile normalised data, (ii) we pruned this set, ranking the genes by variability and removing genes with $r^2 \geq 0.2$ with a higher ranked feature. This resulted in a set of 443 proxy genes, which we used to build a linear covariance for StructLMM based on quantile-normalised expression levels.

cis-eQTL map—We identified *cis*-eQTL using a linear association test, considering genetic variants within 250kb from the centre of the gene body. Following the primary analysis28, we considered the following 53 factors as covariates: the first 25 principal components calculated from the full gene expression profiles, the leading ten MDS components on the genotypes (computed using PLINK v1.90b3.32), cell counts of neutrophils, eosinophils, basophils, lymphocytes and monocytes, age, gender, dataset batch and the first eight principal components derived from SAMtools flagstat and Picard tools (Supplementary Note).

Interaction eQTL analysis—For each of the 23,506 genes, we tested for interactions at the lead variant from the *cis*-eQTL map using StructLMM-int. For comparison, we also considered a multivariate fixed effect test, MultiEnv-Renv-LRT-int (Supplementary Note) either using the same environmental variables as in StructLMM or based on a reduced representation using the leading twenty principal components (Supplementary Note). Significant interactions were reported at FDR < 5% (Storey's procedure49). Calibration of all methods was assessed by repeating the analysis with permuted genotypes. We considered analogous analyses using residual gene expression levels as environments, regressing out the *cis* genetic variant tested from all environments (Supp. Figure 21a-c), to rule out potential spurious effects due to strong gene-exposure correlations. As additional control, we considered an alternative normalisation of the expression data, using boxcox normalisation followed by removal of outliers (2.5 standard deviations, Supp. Figure 21d-f).

Overlap with GWAS hits and pathways analysis—We overlapped our set of interaction eQTL with GWAS variants that are part of the NHGRI-EBI GWAS catalogue33 that pass genome-wide significance threshold ($P<5x10^{-8}$). We defined a colocalisation event based on (i) eQTL and GWAS variants are within 10kb and (ii) high linkage disequilibrium between variants ($r^2 \geq 0.8$, estimated from Phase 3 1,000 Genomes reference panel). For the pathway enrichment analysis, we used the following procedure for each analysed interaction eQTL: (i) we used StructLMM to predict per-individual allelic effects (described above); (ii) we defined the groups of samples with the highest/lowest predicted allelic effect, each containing 10% of the total number of samples (N=204); (iii) we computed rank-based

correlation of genome-wide expression levels and the vector of group binary indicators (based on N=408 samples); (iv) we defined the 100 genes with highest positive correlation as differentially expressed; (v) we performed enrichment analysis of GO biological processes in the differentially expressed test using topGO50 (standard Fisher exact test, algorithm=classic, nodeSize=5). In Supp. Table 6, we report both the top-enriched broad biological process and the three top-enriched narrow processes (broad/narrow terms are defined as those with more/less than 100 annotated genes in the background set). In the *CTSW* example in Fig. 5, the aggregate interacting environment was estimated as described above.

Further statistical details and derivates are provided in Supplementary Note.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The others would like to thank Christoph Lippert and Leopold Parts for helpful discussions. This research has been conducted using the UK Biobank Resource (Application Number 14069). R.M. was supported by a PhD fellowship from the Mathematical Genomics and Medicine programme, funded by the Wellcome Trust. F.P.C., D.H. & O.S. received support from core funding of the European Molecular Biology Laboratory and the European Union's Horizon2020 research and innovation programme under grant agreement N635290. IB acknowledges funding from Wellcome (WT098051 and WT206194). M.J.B. was supported by a fellowship from the EMBL Interdisciplinary Postdoc (EI3POD) program under Marie Skłodowska-Curie Actions COFUND (grant number 664726). The Biobank-Based Integrative Omics Studies (BIOS) Consortium is funded by BBMRI-NL, a research infrastructure financed by the Dutch government (NWO 184.021.007).

BIOS consortium banner

Bastiaan T. Heijmans[7], Peter A.C. 't Hoen[8], Joyce van Meurs[9], Aaron Isaacs[10], Rick Jansen[11], Lude Franke[6], Dorret I. Boomsma[12], René Pool[12], Jenny van Dongen[12], Jouke J. Hottenga[12], Marleen M.J. van Greevenbroek[13], Coen D.A. Stehouwer[13], Carla J.H. van der Kallen[13], Casper G. Schalkwijk[13], Cisca Wijmenga[6], Alexandra Zhernakova[6], Ettje F. Tigchelaar[6], P. Eline Slagboom[7], Marian Beekman[7], Joris Deelen[7], Diana van Heemst[14], Jan H. Veldink[10], Leonard H. van den Berg[10], Cornelia M. van Duijn[15], Bert A. Hofman[16], André G. Uitterlinden[9], P. Mila Jhamai[9], Michael Verbiest[8], H. Eka D. Suchiman[7], Marijn Verkerk[9], Ruud van der Breggen[7], Jeroen van Rooij[9], Nico Lakenberg[7], Hailiang Mei[17], Maarten van Iterson[7], Michiel van Galen[8], Jan Bot[18], Peter van 't Hof[17], Patrick Deelen[6], Irene Nooren[18], Matthijs Moed[7], Martijn Vermaat[8], Dasha V. Zhernakova[6], René Luijk[7], Marc Jan Bonder[6], Freerk van Dijk[6,19], Wibowo Arindrarto[20], Szymon M. Kielbasa[17], Morris A. Swertz[6,19], Erik W. van Zwet[20]

[6] Department of Genetics, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands

[7] Molecular Epidemiology Section, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands

[8] Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands

[9] Department of Internal Medicine, ErasmusMC, Rotterdam, The Netherlands

[10] Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands

[11] Department of Psychiatry, VU University Medical Center, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands

[12] Department of Biological Psychology, VU University Amsterdam, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands

[13] Department of Internal Medicine and School for Cardiovascular Diseases (CARIM), Maastricht University Medical Center, Maastricht, The Netherlands

[14] Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands

[15] Department of Genetic Epidemiology, ErasmusMC, Rotterdam, The Netherlands

[16] Department of Epidemiology, ErasmusMC, Rotterdam, The Netherlands

[17] Sequence Analysis Support Core, Leiden University Medical Center, Leiden, The Netherlands

[18] SURFsara, Amsterdam, the Netherlands

[19] Genomics Coordination Center, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

[20] Medical Statistics Section, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands

References

1. Hunter DJ. Gene-environment interactions in human diseases. Nat Rev Genet. 2005; 6:287–98. [PubMed: 15803198]
2. Ritz BR, et al. Lessons Learned From Past Gene-Environment Interaction Successes. Am J Epidemiol. 2017; 186:778–786. [PubMed: 28978190]
3. Brown AA, et al. Genetic interactions affecting human gene expression identified by variance association mapping. Elife. 2014; 3:e01381. [PubMed: 24771767]
4. Fairfax BP, et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science. 2014; 343
5. Kraft P, Yen YC, Stram DO, Morrison J, Gauderman WJ. Exploiting gene-environment interaction to detect genetic associations. Hum Hered. 2007; 63:111–9. [PubMed: 17283440]
6. Rask-Andersen M, Karlsson T, Ek WE, Johansson A. Gene-environment interaction study for BMI reveals interactions between genetic factors and physical activity, alcohol consumption and socioeconomic status. PLoS Genet. 2017; 13:e1006977. [PubMed: 28873402]

Europe PMC Funders Author Manuscripts

7. Lin X, Lee S, Christiani DC, Lin X. Test for interactions between a genetic marker set and environment in generalized linear models. Biostatistics. 2013; 14:667–81. [PubMed: 23462021]
8. Lin X, et al. Test for rare variants by environment interactions in sequencing association studies. Biometrics. 2016; 72:156–64. [PubMed: 26229047]
9. Casale FP, Horta D, Rakitsch B, Stegle O. Joint genetic analysis using variant sets reveals polygenic gene-context interactions. PLoS Genet. 2017; 13:e1006693. [PubMed: 28426829]
10. Kilpelainen TO, et al. Physical activity attenuates the influence of FTO variants on obesity risk: a meta-analysis of 218,166 adults and 19,268 children. PLoS Med. 2011; 8:e1001116. [PubMed: 22069379]
11. Ahmad S, et al. Gene x physical activity interactions in obesity: combined analysis of 111,421 individuals of European ancestry. PLoS Genet. 2013; 9:e1003607. [PubMed: 23935507]
12. Bjornland T, Langaas M, Grill V, Mostad IL. Assessing gene-environment interaction effects of FTO, MC4R and lifestyle factors on obesity using an extreme phenotype sampling design: Results from the HUNT study. PLoS One. 2017; 12:e0175071. [PubMed: 28384342]
13. Young AI, Wauthier F, Donnelly P. Multiple novel gene-by-environment interactions modify the effect of FTO variants on body mass index. Nat Commun. 2016; 7
14. Corella D, et al. Statistical and biological gene-lifestyle interactions of MC4R and FTO with diet and physical activity on obesity: new effects on alcohol consumption. PLoS One. 2012; 7:e52344. [PubMed: 23284998]
15. Qi Q, et al. Fried food consumption, genetic risk, and body mass index: gene-diet interaction analysis in three US cohort studies. BMJ. 2014; 348
16. 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. 2012; 91:224–37. [PubMed: 22863193]
17. Crawford L, Zeng P, Mukherjee S, Zhou X. Detecting epistasis with the marginal epistasis test in genetic mapping studies of quantitative traits. PLoS Genet. 2017; 13:e1006869. [PubMed: 28746338]
18. Gauderman WJ, et al. Update on the State of the Science for Analytical Methods for Gene-Environment Interactions. Am J Epidemiol. 2017; 186:762–770. [PubMed: 28978192]
19. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015; 526:68–74. [PubMed: 26432245]
20. Locke AE, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015; 518:197–206. [PubMed: 25673413]
21. Bycroft C, et al. Genome-wide genetic data on~ 500,000 UK Biobank participants. bioRxiv. 2017
22. Richardson AS, et al. Moderate to vigorous physical activity interactions with genetic variants and body mass index in a large US ethnically diverse cohort. Pediatr Obes. 2014; 9:e35–46. [PubMed: 23529959]
23. Ahmad S, et al. Established BMI-associated genetic variants and their prospective associations with BMI and other cardiometabolic traits: the GLACIER Study. Int J Obes (Lond). 2016; 40:1346–52. [PubMed: 27121252]
24. Hall NG, Klenotic P, Anand-Apte B, Apte SS. ADAMTSL-3/punctin-2, a novel glycoprotein in extracellular matrix related to the ADAMTS family of metalloproteases. Matrix Biol. 2003; 22:501–10. [PubMed: 14667842]
25. Zillikens MC, et al. Large meta-analysis of genome-wide association studies identifies five loci for lean body mass. Nat Commun. 2017; 8:80. [PubMed: 28724990]
26. Wen W, et al. Genome-wide association studies in East Asians identify new loci for waist-hip ratio and waist circumference. Sci Rep. 2016; 6
27. Shungin D, et al. New genetic loci link adipose and insulin biology to body fat distribution. Nature. 2015; 518:187–196. [PubMed: 25673412]
28. Zhernakova DV, et al. Identification of context-dependent expression quantitative trait loci in whole blood. Nat Genet. 2017; 49:139–145. [PubMed: 27918533]
29. Westra HJ, et al. Cell Specific eQTL Analysis without Sorting Cells. PLoS Genet. 2015; 11:e1005223. [PubMed: 25955312]

30. Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M. Mapping complex disease traits with global gene expression. Nat Rev Genet. 2009; 10:184–94. [PubMed: 19223927]
31. Maurano MT, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012; 337:1190–5. [PubMed: 22955828]
32. Emilsson V, et al. Genetics of gene expression and its effect on disease. Nature. 2008; 452:423–8. [PubMed: 18344981]
33. MacArthur J, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 2017; 45:D896–D901. [PubMed: 27899670]
34. Galvez J. Role of Th17 Cells in the Pathogenesis of Human IBD. ISRN Inflamm. 2014; 2014
35. Day FR, Loh P-R, Scott RA, Ong KK, Perry JR. A robust example of collider bias in a genetic association study. The American Journal of Human Genetics. 2016; 98:392–393. [PubMed: 26849114]
36. Listgarten J, Lippert C, Heckerman D. FaST-LMM-Select for addressing confounding from spatial structure and rare variants. Nature Genetics. 2013; 45:470. [PubMed: 23619783]
37. Wu MC, et al. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet. 2011; 89:82–93. [PubMed: 21737059]
38. Lee S, Wu MC, Lin X. Optimal tests for rare variant effects in sequencing association studies. Biostatistics. 2012; 13:762–75. [PubMed: 22699862]
39. Schaeffer L. Application of random regression models in animal breeding. Livestock Production Science. 2004; 86:35–45.
40. Casale FP, Rakitsch B, Lippert C, Stegle O. Efficient set tests for the genetic analysis of correlated traits. Nat Methods. 2015; 12:755–8. [PubMed: 26076425]
41. Loh PR, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet. 2015; 47:284–90. [PubMed: 25642633]
42. Fesinmeyer MD, et al. Genetic risk factors for BMI and obesity in an ethnically diverse population: results from the population architecture using genomics and epidemiology (PAGE) study. Obesity. 2013; 21:835–846. [PubMed: 23712987]
43. Abraham G, Qiu Y, Inouye M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. Bioinformatics. 2017; 33:2776–2778. [PubMed: 28475694]
44. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the royal statistical society. Series B (Methodological). 1995:289–300.
45. Van Greevenbroek MM, et al. The cross-sectional association between insulin resistance and circulating complement C3 is partly explained by plasma alanine aminotransferase, independent of central obesity and general inflammation (the CODAM study). European journal of clinical investigation. 2011; 41:372–379. [PubMed: 21114489]
46. Tigchelaar EF, et al. Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. BMJ open. 2015; 5:e006772.
47. Hofman A, et al. The Rotterdam Study: 2014 objectives and design update. European journal of epidemiology. 2013; 28:889–926. [PubMed: 24258680]
48. Skyler JS. Pulmonary insulin update. Diabetes technology & therapeutics. 2005; 7:834–839. [PubMed: 16241894]
49. Storey JD. A direct approach to false discovery rates. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2002; 64:479–498.
50. Alexa A, Rahnenfuhrer J. topGO: enrichment analysis for gene ontology. R package version. 2010; 2
51. Lippert C, Casale FP, Rakitsch B, Stegle O. LIMIX: genetic analysis of multiple traits. bioRxiv. 2014

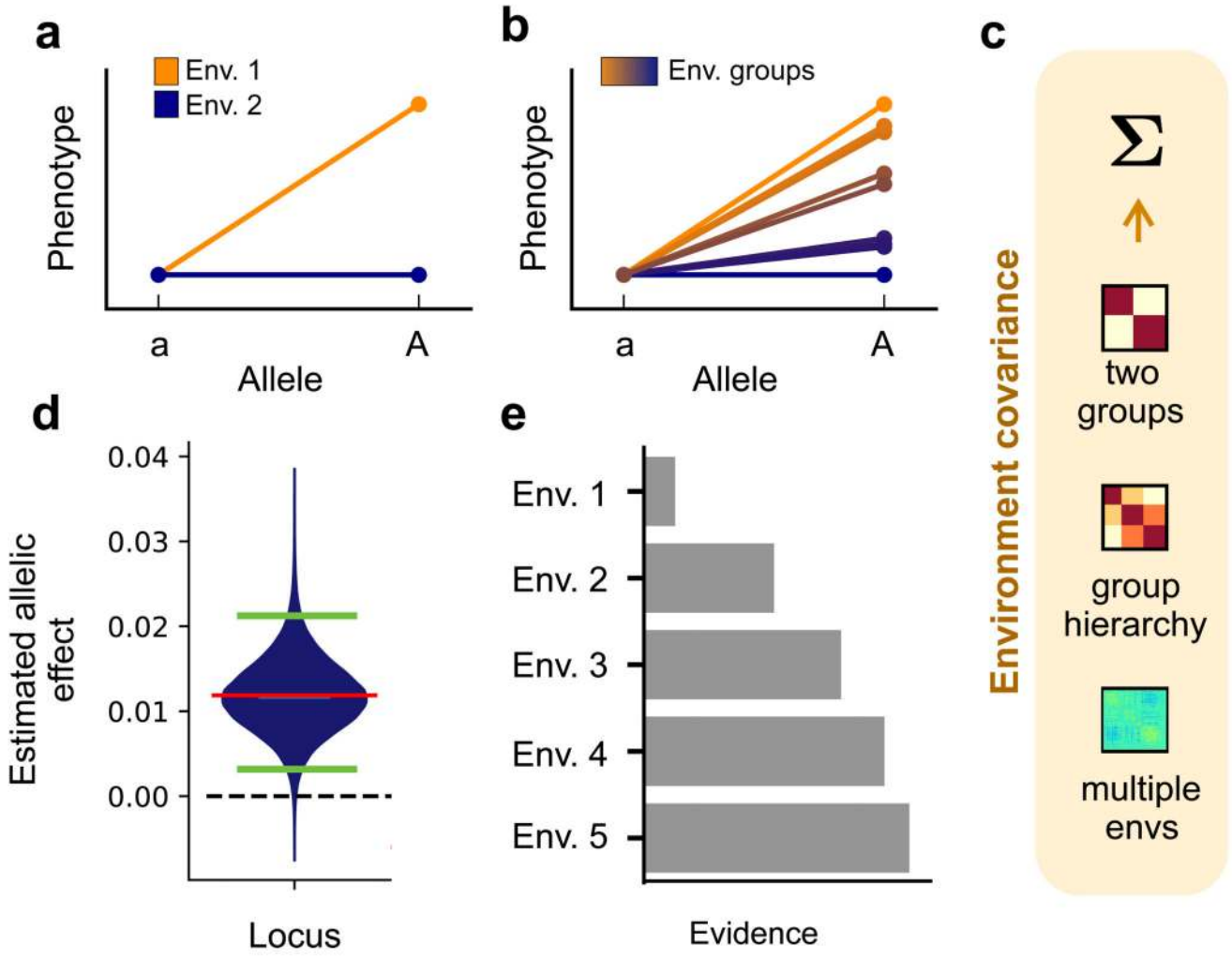

Figure 1. Overview of the StructLMM model.

(**a**) Basic genotype-environment interaction, with a genetic effect that is specific to one of two groups (blue and orange lines correspond to the average phenotypes observed within two environmental groups for two alleles). (**b**) Interaction with multiple environmental groups or bins of continuous environmental states (average phenotypes for groups exerting increasing GxE effects from blue to orange for two alleles). (**c**) StructLMM accounts for possible heterogeneity in effect sizes due to GxE using a multivariate normal prior, where alternative choices of the environmental covariance **Σ** can capture discrete (two groups, group hierarchy; see **a**,**b**) or continuous substructure of environmental exposures in the population (multiple envs). (**d,e**) Different illustrative example analyses using StructLMM. (**d**) Estimation of per-individual allelic effects in the population at individual loci. The violin plot displays the density of estimated allelic effect sizes for individuals in the population. Median and the top and bottom 5% quantiles of the effect size distribution are indicated by the red and green bars, respectively. (**e**) Bayes factors between the full model and models with environmental variables removed, thereby identifying environments that are most relevant for GxE.

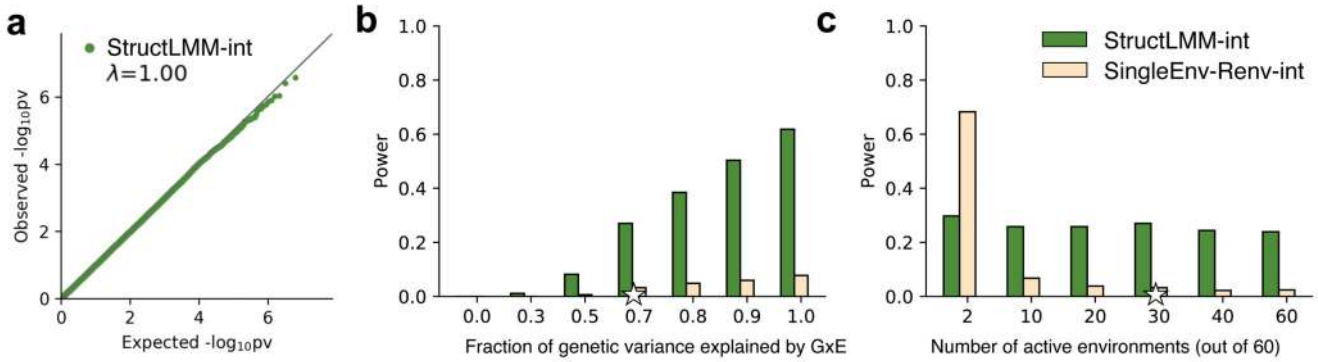

Figure 2. Assessment of statistical calibration and power using simulated data.

(**a**) QQ plots of negative log P values from the StructLMM interaction test (green, StructLMM-int) using phenotypes simulated from the null (no genetic effect) for 103,527 variants on chromosome 21. (**b**) Comparison of power for detecting GxE interactions for increasing fractions of the genetic variance explained by GxE (ρ). Compared are the StructLMM interaction test (StructLMM-int) and a single-environment interaction test (SingleEnv-Renv-int). (**c**) Analogous power analysis, when simulating GxE using increasing numbers of active environments with non-zero GxE effects (out of 60 environments total, considered in all sts; ρ=0.7). All 60 environments contribute to the simulated additive environment effect. Models were assessed in terms of power (at Family Wise Error Rate - FWER<1%) for detecting variants with true GxE effects (Methods). Stars denote default values of genetic parameters, which were retained when varying other parameters (Supp. Table 1). A synthetic CEU population of 5,000 individuals based on the 1000 Genomes Project was used for all experiments.

Europe PMC Funders Author Manuscripts

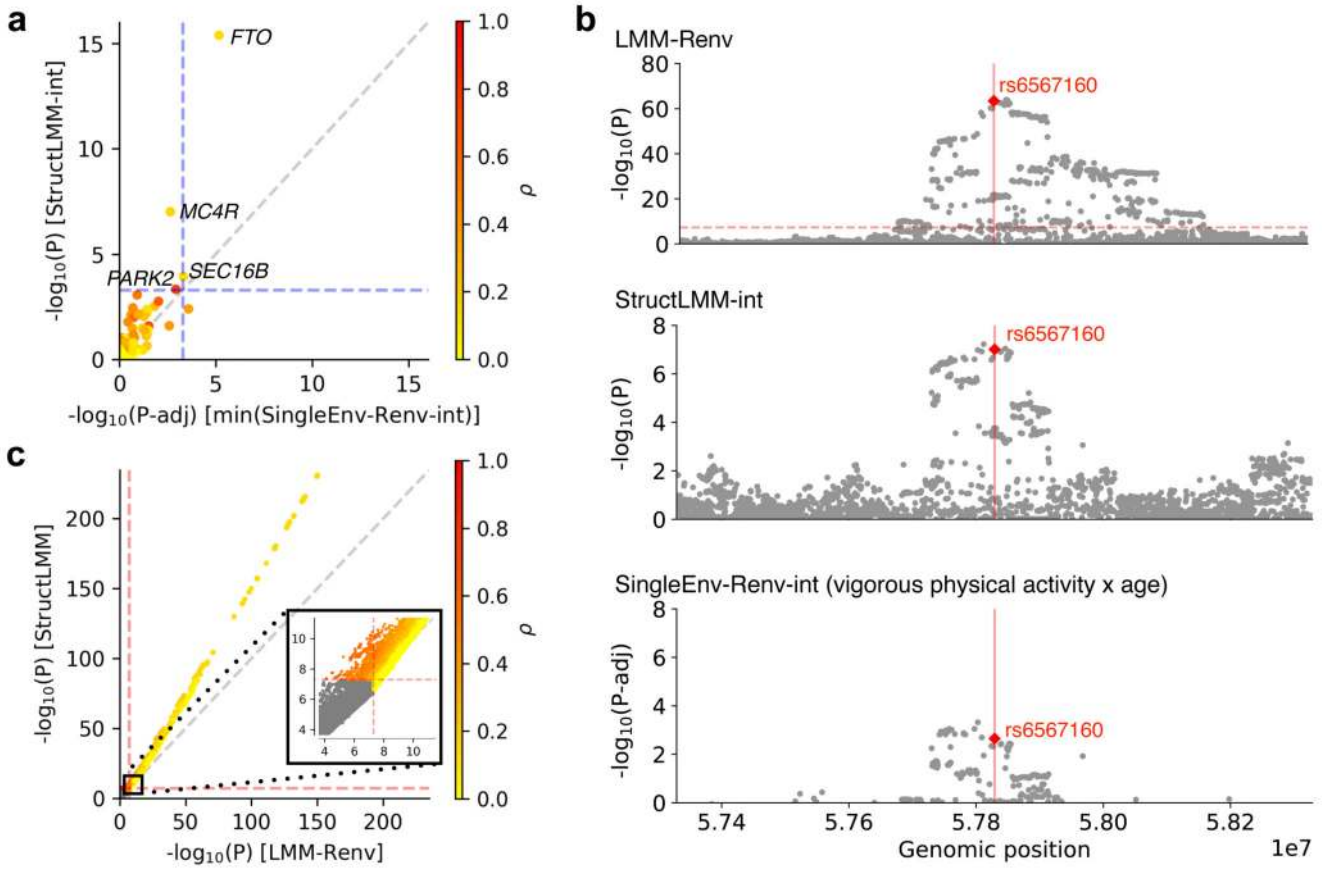

Figure 3. Applications to model GxE on body mass index (BMI) in UK Biobank.

(**a**) Scatter plot of negative log P values from GxE interaction tests at 97 GIANT variants20, considering a single-environment fixed effect GxE tests (SingleEnv-Renv-int, x-axis, P values Bonferroni adjusted for the number of tested environments) versus the StructLMM interaction test (StructLMM-int, y-axis). Dashed lines correspond to $\alpha<0.05$, Bonferroni adjusted for the number of tests. (**b**) Local Manhattan plots of an interaction identified by StructLMM-int at *MC4R*. From top to bottom: LMM association test (LMM-Renv), StructLMM interaction test (StructLMM-int), single-environment LMM interaction test (SingleEnv-Renv-int) for the environment with the strongest GxE effect at the GIANT SNP, age-adjusted vigorous physical activity (vigorous physical activity x age). The red vertical line and diamond symbol indicates the GIANT SNP as in **a**. (**c**) Scatter plot of genome-wide negative log P values from LMM association test (LMM-Renv, x-axis) versus the StructLMM association test (y-axis). Dashed lines indicate genome-wide significance at $P<5x10^{-8}$ and colour denotes the estimated extent of heterogeneity (fitted parameter ρ), where yellow/red corresponds to variants with low/high GxE components. The inset displays a zoom-in view of variants close to genome-wide significance. n = 252,188 unrelated individuals of European ancestry for all experiments.

Europe PMC Funders Author Manuscripts

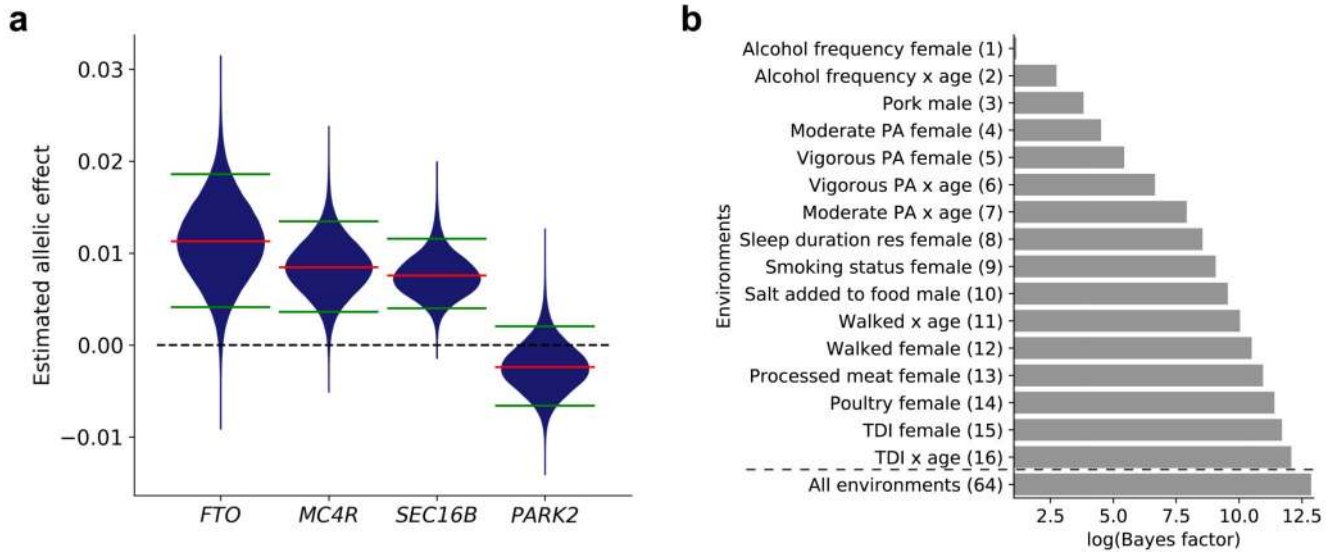

Figure 4. Downstream analysis to explore identified GxE loci.

(**a**) Violin plots showing distributions of the in-sample estimated allelic effect size (effect of heterozygous versus homozygous reference carriers for environmental states realised in the population; n = 252,188 unrelated individuals of European ancestry for all experiments; Methods) on BMI for the four GIANT variants with GxE (α<0.05, Fig. 3a). Estimated persistent genetic effects are shown by the red bar and the green bars indicate top and bottom 5% quantiles of variation in effect sizes due to GxE. (**b**) Cumulative evidence of environmental variables that explain GxE at *MC4R,* showing Bayes factors between the full model and models with increasing numbers of environmental variables removed using backward elimination. For comparison, shown is the evidence for all 64 environmental variables. 'Alcohol frequency female', is selected as the first environmental factor, followed by 'Alcohol frequency x age' and so on.

Europe PMC Funders Author Manuscripts

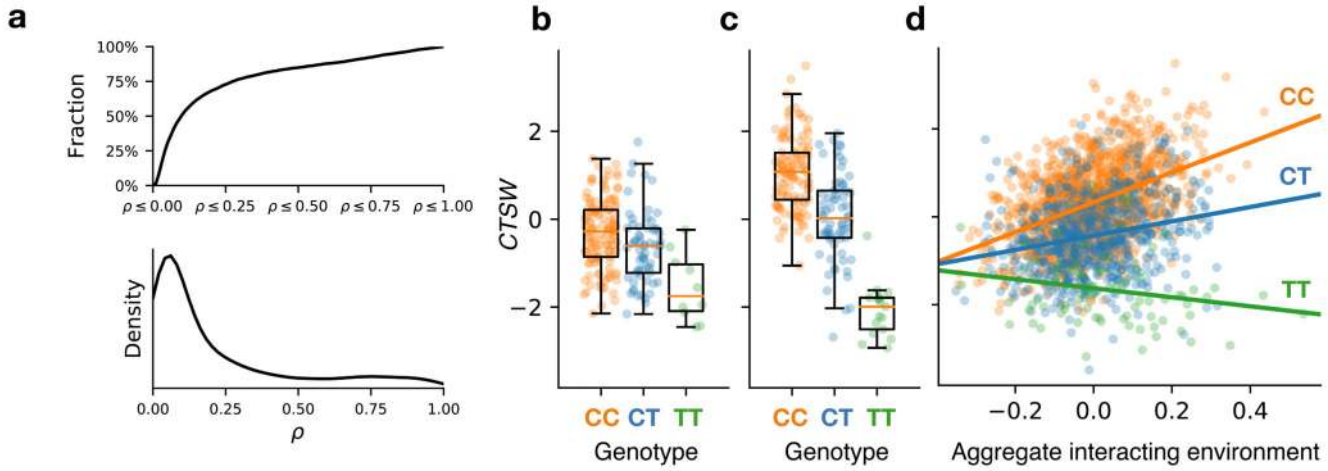

Figure 5. Gene-context interactions in a blood gene expression cohort.

(**a**) Cumulative fraction (top) and density (bottom) of eQTL with interactions (3,483 interaction eQTL; FDR<5%) as a function of the estimated extent of heterogeneity (fitted parameter ρ). (**b**-**d**) Example of an interaction eQTL for *CTSW* at the lead variant *rs568617*, which is in LD with *rs568617* (r^2=0.98, Supp. Fig. 23), a known risk variant for Crohn's disease. (**b**,**c**) Expression level of *CTSW* for different alleles at the lead eQTL variant, considering 10% strata of individuals (n = 204 independent samples) with the smallest (**b**) and largest (**c**) per-individual allelic effects as estimated using StructLMM, displaying the 25th, 50th and 75th percentiles, with whiskers extending to 1.5 times the interquartile range. (**d**) Scatter plot of *CTSW* expression level versus the aggregate environmental signal for the GxE effect at *rs568617* (aggregate interacting environment), estimated using StructLMM (Supplementary Note). Individuals are stratified by the alleles at the eQTL lead variant. Solid lines denote regression lines for each genotype group.

Europe PMC Funders Author Manuscripts