Supplementary Note 1: Mathematical rationale

We assume that gene expression is log-linear in environmental effects, as in common approaches to modeling expression,

$$\log t_i^a = \gamma_e^a e_i + \beta_h \eta_i^a + \beta_{h\times e}\eta_i^a e_i + \mu^a + \epsilon_i^a \tag{1}$$

$$\epsilon_i^a \sim N(0, v^a) \tag{2}$$

where all parameters should additionally be indexed by the particular locus s, but we omit this for clarity. Here t_i^a is the "true" expression of the alternative allele in individual i, η_i^a is the haplotype of a candidate regulatory SNP (note we only consider the top eSNP), γ_e^a is the allele specific effect of the environment e, β_h is the main effect of the SNP in phase with a, and $\beta_{h\times e}$ is an interaction term. Note that including γ_e^a means that we model allele specific effects of environment without explicitly needing to test a candidate SNP. ϵ_i^a is noise resulting from both technical factors such as PCR amplification, as well as other unexplained inter-individual variation, e.g. resulting from unmeasured covariates. We do not need to assume the variance v is equal for both alleles. Given these latent expression levels we model the observed read counts as Poisson distributed,

$$\begin{aligned} y_i^a &\sim \text{Pois}(t_i^a) \\ y_i^r &\sim \text{Pois}(t_i^r) \end{aligned} \tag{3}$$

where y_i^a and y_i^r are the alternative and reference read counts respectively for individual i. Since our intention is to use only the allele specific signal, we condition on the total read count $n_i = y_i^a + y_i^r$, which results in the alternative read count being binomially distributed, i.e.

$$y_i^a | t, n_i \sim \text{Bin}(n_i, t_i^a/(t_i^a + t_i^r)). \tag{4}$$

From Equation 1 we can obtain an expression for the p parameter of the binomial in Equation 4

$$t_i^a/(t_i^a + t_i^r) = \sigma((\gamma_e^a - \gamma_e^r)e_i + \beta_h(\eta_i^a - \eta_i^r) + \beta_{h\times e}(\eta_i^a - \eta_i^r)e_i + (\mu^a - \mu^r) + \epsilon_i) \tag{5}$$

$$\epsilon_i = \epsilon_i^a + \epsilon_i^r \sim N(0, v^a + v^r) \tag{6}$$

assuming independence between ϵ_i^a and ϵ_i^r, and the logistic function $\sigma(\xi) = 1/(1 + e^{-\xi})$. Noting that $\eta_i^a - \eta_i^r = \pm h_i$ is the heterozygosity of the candidate SNP, defining $\beta_e := \gamma_e^a - \gamma_e^r$ and $\mu := \mu^a - \mu^r$, and considering the symmetrized likelihood for $\min(y_i^a, y_i^r)$ we obtain our likelihood model

$$\min(y_i^a, y_i^r) | \beta, \mu, \epsilon_i \sim \text{Binomial}\left[n_i, \sigma(\beta_e e_i + \beta^h h_i + \beta^{g\times e} e_i h_i + \mu + \epsilon_i)\right] \tag{7}$$

By evaluating absolute allelic imbalance by taking the minimum between y_i^a and y_i^r we are able to address individuals with opposite directions of allelic response, handling both unphased data along with the possibility of distinct causal variants between individuals.

1.1 Overdispersion model

The noise ϵ_{is} in Equation 7 is distributed as $N(0, v_s)$ where $v_s = v_s^a + v_s^r$. Note we have reintroduced the indexing with respect to the locus s. Since v_s^a and v_s^r are clearly not identifiable under our likelihood model, we work directly with the overdispersion variance v_s instead.

We explored several prior specifications for v_s. The first was to assume constant $v_s = v$ for all loci s. This still resulted in significant inflation of p-values under permutations, even if not to the same extent as the binomial GLM. The second was to allow v_s to be different for each locus and use the conjugate prior, which is an inverse-gamma on v_s, or equivalently gamma on the precision $1/v_s \sim G(a, b)$, where $\{a, b\}$ are the shape and rate respectively. This was the setting we chose because p-values under permutations are not inflated (see Supplementary Figure 6). Thirdly, observing that MCMC estimates of v_s showed a relationship with log average read depth t_s at locus s, we tested a model $\log v_s \sim N(mt_s + c, r)$. While this approach finds a somewhat larger number of associations, nominal p-values are still slightly overinflated under the null. While this can be resolved by running permutations for the top nominal associations, we chose to avoid the additional computational burden. None-theless, this option is included in the EAGLE software.

Supplementary Note 2: Simulation study

In order to test EAGLE vs. naive methods for associating environmental factors and allelic imbalance we performed an extensive simulation study. In order to render our simulations as realistic as possible we use real data from the DGN cohort as much as possible. In particular, we test the same exonic SNPs, with the same number of heterozygous individuals, the same total expression read counts and simulate overdispersion of alternative allele counts using parameters estimated from the true data. Our simulation procedure, repeated for each exonic SNP s, is as follows:

1. Fit the EAGLE model with no covariates and no min:

$$\begin{aligned} y_i^a | \mu, \epsilon_i &\sim \text{Binomial}\left[n_i, \sigma(\mu + \epsilon_i)\right] \\ \epsilon_i | v_s &\sim N(0, v_s) \\ v_s | a, b &\sim G(a, b) \end{aligned} \tag{8}$$

   where $a = 1.1, b = 3 \times 10^{-3}$ are the hyperparameters learned across all exonic SNPs. EAGLE provides an estimate of μ, denoted $\hat{mu}$, and the variational posterior distribution $q_s(1/v_s) = G(1/v_s, a_s, b_s)$, with corresponding expected posterior random effect variance $\hat{v}_s := \mathbb{E}_q[v_s] = b_s/a_s$.

2. For each heterozygous individual we simulate allelic read counts y^{sim} as

$$\begin{aligned} y_i^{\text{sim}} | \hat{mu}, \epsilon_i &\sim \text{Binomial}\left[n_i, \sigma(\hat{\mu} + \beta_x x_i + \epsilon_i)\right] \\ \epsilon_i | \hat{v}_s &\sim N(0, \hat{v}_s) \end{aligned} \tag{9}$$

where β_x is a "true" effect size that we choose (or set to zero for no effect) and x_i is the environment (time of day in all simulations presented herein).

Having simulated new alternative allele count we test for association using EAGLE, a binomial GLM, or Spearman correlation. The binomial GLM is equivalent to EAGLE with no overdispersion, i.e. $v_s = 0$, and is implemented using the R `glm` function. Spearman correlation is calculated between allelic imbalance (defined as $|\frac{1}{2} - y_i/n_i|$) and the environment, x_i.

We did not attempt to use simulations to compare EAGLE to standard interaction QTL testing since these methods condition on different data, and such a simulation would therefore require making significant assumptions about unknown effects on NGS read counts. In particular the relative effect sizes and noise levels for the simulated allelic ratios and total expression would determine which approach performs best.

Supplementary Note 3: Pathway enrichment

We sought to characterize the properties of the genes whose genetic regulation is modulated by each environment. Since the number of genome-wide significant associations remains relatively modest even with the improved power from EAGLE, we performed enrichment analysis using the top 50 associations for each environment (the results are relatively robust to the number of top associations used, Supplementary Figure S9b). We first tested these associations against a curated set of pathways taken from GO, KEGG and BioCarta (restricted to those with fewer than 100 genes), using a standard hypergeometric test with the entire set of genes tested by EAGLE as the background. The strongest enrichment is for smoking and the BioCarta CCR5 pathway. CCR5 itself has been implicated in smoking induced emphysema (Ma et al., 2005).

Supplementary Note 4: Transcription factor enrichment

Since our hypothesis is that GxE interactions for gene expression are often driven by allele specific binding of environmentally-responsive transcription factors, we tested for enrichment of transcription factor binding sites (TFBS) proximal to environment-associated genes. We inferred transcription factor binding sites in seven cell type were inferred from DNase-seq measurements using a modified version of CENTIPEDE (Pique-Regi et al., 2011). CENTIPEDE relies on two observations: (1) chromatin around motif instances bound by transcription factors typically has higher DNase I sensitivity than chromatin around unbound motif instances, and (2) each transcription factor has a characteristic DNase I cleavage profile around bound motif instances. Based on these observations, given a putative bound motif instance, CENTIPEDE models the number of reads mapped to each base pair along a window around the motif site as a

mixture of two components (bound vs unbound), and infers the probability that each site is bound. Specifically, conditional on being bound (or unbound), CENTIPEDE models (1) the total number of DNase-seq reads using a negative binomial distribution, and (2) DNase-seq read counts in a window around the putative binding site using a multinomial distribution conditioned on the total number of reads, with independent sets of parameters for bound and unbound sites.

However, when multiple replicate DNase-seq measurements are available for the same cell type, CENTIPEDE has often been applied after pooling replicates. If there is substantial heterogeneity in DNase-seq data between replicates, then pooling replicates tends to introduce more variation in the total read counts and read count profiles, limiting the ability of CENTIPEDE to accurately identify TF binding sites. We used a modified version of CENTIPEDE by modeling the data from each replicate separately. Specifically, we model the total number of DNase-seq reads at bound and unbound sites using separate negative binomial distributions for each replicate. Since we expect the DNase cleavage profile at bound sites to remain the same across replicates, we use the same multinomial distribution to model the DNase-seq read count profiles for all replicates.

We combined predicted TFBS from the following seven cell types (ENCODE Project Consortium, 2004):

1. CD20+_RO01778 B-cells
2. GM12878 lymphoblastoid cell line (CEPH/Utah)
3. Hmveclbl lung-derived blood microvascular endothelial cells
4. Hpaf pulmonary artery fibroblasts (blood)
5. Primary Th1 T-cells
6. Primary Th2 T-cells
7. Primary Th17 T-cells
8. Regulatory T cells

Since we only expect to see GxE when there is corresponding genetic variation, we filtered for TFBS within 5kb of each gene that also contained at least one variant previously identified in the 1000 Genomes Project, resulting in an average of 7.7 TFBS per gene across 282 TF motifs. We again used a hypergeometric test for enrichment for each environment (Supplementary Figure S9c). While no associations were significant at 10% FDR using Benjamini-Hochberg, it is notable that the strongest association (nominal $p = 10^{-4}$) is for smoking-associated genes and the transcription factor TBX4. TBX4 is known to be regulated by SOX9, variants in which influence lung function specifically in smokers (Melén and Bottai, 2012). Additionally, genes showing a GxE interaction for blood pressure medication are enriched in binding of SP1, which is known to respond to antihypertensive drugs (Negoro et al., 1995) and regulates angiotensin receptor transcription (Kubo et al., 2003; Rohrwasser et al., 2002).

Supplementary Note 5: *trans*-eQTL enrichment

Further, we investigated additional evidence for co-regulation of EAGLE hits of each environment based on trans-eQTLs. DGN's relatively large sample size enables the detection of inter-chromosomal trans-eQTLs (138 unique trans-eQTL genes at 5% FDR, Battle et al. (2014)). Applying a relaxed threshold of $p < 10^{-5}$ yielded a trans-network with 55,313 edges involving 48,163 SNPs and 7473 genes. We investigated whether the top 50 genes associated by EAGLE for each environment tend to share distal regulatory SNPs in this network. Against an empirical null distribution generated by randomly sampling sets of 50 genes from those tested for each environment, we found the number of SNPs regulating more than one of the 50 genes is significantly increased ($p < 0.05$) for age, exercise, family history of depression and opiate use. The trans-network involving SNPs regulating more than one gene in the top 50 list for exercise is shown in Supplementary Figure S9d. Interestingly all five of the genes (IFIT2, MX2, IFI44L, ADAR, RSAD2) implicated in this network are interferon inducible, highlighting the impact of exercise on immune response (Walsh et al., 2011).

Supplementary Note 6: Confounding and standard interaction QTL testing

Total gene expression suffers from high confounding. Results from EAGLE or standard methods could represent interactions with (potentially unmeasured) factors that are correlated with the tested environmental variables. EAGLE however should be less susceptible to false positives from some technical confounders (Supplementary Figure 1). We here use a second simulation study to test the hypothesis that this is a key reason we fail to detect many GxE effects on the transcriptome using standard interaction QTL testing. It is extremely challenging to remove the influence of hidden technical confounding without subtracting true biological signal.

1. Gene expression was quantified as $\log_2(2+(5\times10^7)c_{id}/T_i)$ where T_i is the total number of mapped reads for sample i, c_{id} are the counts for gene d. These values are quantile normalized to a Gaussian (for each gene) to give expression values y_{id}.
2. Principal component analysis (PCA) is performed on the matrix Y.
3. For each gene d we fit a Bayesian regression model

$$\begin{aligned} y_{id} &= \beta_x x_i + \sum_s \beta_s g_{is} + \sum_{k=1}^{K} \beta_k u_{ik} + \epsilon_i \\ \beta_x, \beta_s, \beta_k &\sim N(0, \sigma_\beta^2) \\ \epsilon_i &\sim N(0, \sigma_n^2) \end{aligned} \tag{10}$$

where x is the environment, g_{is} is the genotype of cis-SNP s for this gene (taking SNPs within 100kb), u_{ik} is the principal component k for individual i and K denotes the number of confounders factors (PCs) that contribute highly to expression (we take $K = 20$). We analytically integrate over the β's and optimize σ_β^2 and σ_n^2 (using L-BFGS) to obtain MLEs $\hat{\sigma}_\beta^2$ and $\hat{\sigma}_n^2$ and posterior means of the β's, denoted by $\hat{\beta}$.

4. We generate simulation expression values as

$$\begin{aligned} y_{id}^{\text{sim}} &= \gamma g_{is'} x_i + \hat{\beta}_x x_i + \sum_s \hat{\beta}_s g_{is} + \sum_{k=1}^{K'} \hat{\beta}_k u_{ik} + \epsilon_i \\ \beta_x, \beta_s, \beta_k &\sim N(0, \hat{\sigma}_\beta^2) \\ \epsilon_i &\sim N(0, \hat{\sigma}_n^2) \end{aligned} \tag{11}$$

where γ is the true GxE effect size (set to 0 or 0.1 with equal probability in our simulation), s' indexes the lead eSNP and $K' \leq K$ is the number of confounding PCs we include in the simulated data.

5. We then use standard interaction QTL testing, combined with Bonferroni correction across cis-SNPs, to detect the "true" simulated GxE associations.

In particular we vary the number of confounding PCs included in the simulation, K', finding, as expected, that performance is highly sensitive to the amount of confounding included in the simulation (Supplementary Figure 11).

Supplementary Note 7: Obtaining interaction p-values from Fairfax et al.

The Fairfax et al. (2014) study does not provide test statistics for interaction terms, and the raw data is not available. However, we can approximate interaction p-values using the t-statistics provided for eQTL testing in the naive and stimulated conditions respectively. Let t_0 and t_1 be the t-statistics for a particular SNP-gene pair in the naive and stimulated state respectively. Under the additive linear model used by Fairfax et al., t_0 and t_1 both have non-central student-t distributions with degrees of freedom (d.o.f.) $\nu = n - 2$, where n is the number of samples (228 in this case), and $\mu_i = \frac{\beta_i}{\sigma c}$ where σ is the true noise standard deviation (assumed equal for both perturbed and unperturbed state, as with standard interaction testing), β_i is the true effect size (with $i \in \{0, 1\}$ denoting the two states) and $c = \sqrt{[(X^T X)^{-1}]_{11}}$ comes from the Fisher infor-

mation matrix (with X the design matrix), i.e.

$$t_i \stackrel{d}{=} \frac{\mu_i + z}{\sqrt{v/\nu}}, \tag{12}$$

$$z \sim N(0,1), \tag{13}$$

$$v \sim \chi^2_\nu \tag{14}$$

To test the null hypothesis that $\beta_0 = \beta_1$ we consider the test statistic $t_\Delta = t_1 - t_0$. Unfortunately the difference of two independent non-central t-variates does not have a standard form, so we resort to approximating t_0, t_1 by Gaussians with means and variances

$$\mathrm{E}\left[t_i\right] = \mu_i \sqrt{\frac{\nu}{2}} \frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}, \tag{15}$$

$$\mathrm{Var}\left[t_i\right] = \frac{\nu(1+\mu_i^2)}{\nu-2} - \frac{\mu_i^2 \nu}{2}\left(\frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}\right)^2. \tag{16}$$

Under the null $\mu_0 = \mu_1$ so $E[t_\Delta] = E[t_1] - E[t_0] = 0$. To calculate $\mathrm{Var}\left[t_\Delta\right] = \mathrm{Var}\left[t_0\right] + \mathrm{Var}\left[t_1\right]$ exactly we would need μ_0, μ_1. Since we don't know these we approximate it by the observed t-statistic, i.e. using $\mu_i \approx \hat{t}_i$ in Equation 16. Using the resulting approximation $\hat{\mathrm{Var}}[t_\Delta]$ we obtain p-values as

$$F_{N-2}(-|t_\Delta| / \sqrt{\hat{\mathrm{Var}}[t_\Delta]}) \tag{17}$$

where $F_\rho(\cdot)$ is the student-t cdf with ρ degrees of freedom. To confirm that $\hat{\mathrm{Var}}[t_\Delta]$ is good approximation we simulated a SNP with minor allele frequency (MAF) 0.20 in $N = 228$ individuals in two conditions, with varying shared effect sizes β and noise standard deviation 0.3. We calculated t-statistics using an additive linear model for both conditions and investigated the distribution of t_Δ under these simulations. Using 1000 simulations we first confirmed that t_Δ is approximately normally distributed (Shapiro-Wilk test $p = 0.32$). We then compared the standard deviation of the sampling distribution for t_Δ under simulations vs. that calculated using Equation 16, finding excellent agreement across a range of effect sizes (Supplementary Figure 13a). For shared $\beta = 5$ we confirmed that the corresponding p-values calculated using Equation 17 are well calibrated (Supplementary Figure 13b).

Supplementary Note 8: The regulation of Ces1f by PPARα

PPARα had an allele-specific effect on Ces1f expression, suggesting different genetic backgrounds may explain variability in reports of the strength of PPARα's regulatory effect (Zhang et al., 2012; Jones et al., 2013). This result highlights the importance of considering different genetic backgrounds when mapping the regulatory targets of transcription factors.

Supplementary Note 9: PPARα TFBS enrichment

In order to test whether the genes EAGLE detected as having allele-specific responses to PPARα agonists had an enrichment of appropriate binding motifs we first scanned the *Rattus norvegicus* rn5 genome using FIMO (Grant et al., 2011) Version 4.11.1, part of the MEME online suite (Bailey et al., 2009), for the motifs for PPARγ::RXR (MA0065.2) and PPARγ (MA0066.1) obtained from the JASPAR core motif database (2016 server). PPARα and PPARγ have similar binding behavior, with both typically binding as a heterodimer with RXR. The default p-value threshold of 10^{-4} was used in FIMO. We fit a Poisson GLM where the response was the number of occurrences within a window w of the TSS of each gene tested by EAGLE, and the single covariate was whether the gene was associated by EAGLE (at an FDR of 10%). The window size w was varied as shown in Figure 3c of the main paper, and we report the Wald test statistic (z-score).

Supplementary Note 10: Filtering criteria

We use the following criteria to choose which loci (exonic SNPs) to test:

- The SNP must not show significant mapping bias, based on simulated reads (see Battle et al. (2014) for simulation details)
- We require at least 20 individuals to be heterozygous at this locus (since the homozygotes do not contribute to the likelihood).
- At least 10 of these heterozygotes must not have the same environmental factor (this covers both continuous and discrete, or even mixed, environmental covariates)
- Fewer than half the heterozygotes have monoallelic expression, with the later defined as either fewer than three reads, or fewer than 1%, mapping to one allele. The rationale for this filter is that extensive monoallelic expression is typically an indicator of more complex events than simple *cis*-regulation. For example, we have observed imprinted genes or genes with common deletions of an exon manifest as monoallelic expression present in many individuals. Additionally, mapping errors that escape standard filters may also appear as monoallelic expression. Even the strongest *cis*-eQTLs we find in DGN do not completely suppress expression of one allele in heterozygotes.

We do not use an explicit filter on the total number of reads since our count based likelihood will naturally put less weight on individuals with few total reads.

Having chosen which loci to test we chose what covariates to include in the model. In all cases the null model H_0 will include an intercept term and the alternative H_1 will include an intercept term and the environmental covariate, i.e.

$$\begin{aligned} H_0 &: \min(y_i^a, y_i^r)|\mu, \epsilon_i \sim \text{Binomial}\left[n_i, \sigma(\mu + \epsilon_i)\right] \\ H_1 &: \min(y_i^a, y_i^r)|\beta, \mu, \epsilon_i \sim \text{Binomial}\left[n_i, \sigma(\beta_e e_i + \mu + \epsilon_i)\right] \end{aligned} \tag{18}$$

For the 70% of testable genes where we have a cis-eQTL from total expression analysis with a Bonferroni corrected p-value less than 10^{-3}, we consider including h_{is}, a binary variable indicating whether this lead eSNP is heterozygous in individual i. This term is included in both hypotheses if, out of the heterozygotes at s, there are at least 5 homozygotes and 5 heterozygotes at the eSNP, i.e. $\min(\sum_{n\in\mathcal{H}_s} h_{is}, \sum_{n\in\mathcal{H}_s}(1-h_{is})) > 5$ where $\mathcal{H}_s$ is the set of individual heterozygous at s. We then have

$$\begin{aligned} H_0 &: \min(y_i^a, y_i^r)|\mu, \beta, \epsilon_i \sim \text{Binomial}\left[n_i, \sigma(\beta^h h_i + \mu + \epsilon_i)\right] \\ H_1 &: \min(y_i^a, y_i^r)|\beta, \mu, \epsilon_i \sim \text{Binomial}\left[n_i, \sigma(\beta_e e_i + \beta^h h_i + \mu + \epsilon_i)\right] \end{aligned} \tag{19}$$

Finally, we add an interaction term between the lead eSNP and the environmental factor to the alternative model if no group is created with fewer than 5 individuals, so that

$$\begin{aligned} H_0 &: \min(y_i^a, y_i^r)|\mu, \beta, \epsilon_i \sim \text{Binomial}\left[n_i, \sigma(\beta^h h_i + \mu + \epsilon_i)\right] \\ H_1 &: \min(y_i^a, y_i^r)|\beta, \mu, \epsilon_i \sim \text{Binomial}\left[n_i, \sigma(\beta_e e_i + \beta^h h_i + \beta^{g\times e} e_i h_i + \mu + \epsilon_i)\right] \end{aligned} \tag{20}$$

Supplementary Note 11: Parameter estimation and inference

Here we describe how we estimate parameters under our model and calculate the required marginal likelihoods to perform likelihood ratio testing. The overall testing strategy is

1. Jointly fit the null model for all loci and the noise shape and rate hyperparameters $\{a, b\}$
2. Holding $\{a, b\}$ fixed, fit the alternative model for all loci
3. Perform likelihood ratio testing using the approximate marginal likelihoods obtained from Step 1 and 2

Our strategy for steps 1 and 2 is to approximately integrate over the overdispersion noise ϵ and the per locus overdispersion parameter v_s, while optimizing the fixed effect regression coefficients (the intercept, environment coefficient, and possibly eSNP heterozygosity and interaction terms), and noise hyperparameters $\{a, b\}$.

11.1 Variational EM algorithm

We use variational expectation-maximization (EM, Jordan et al., 1999) to optimize β_s, μ_s, a, b in the M-step while integrating out the overdispersion noise ϵ_{is} and random effect variance v_s using Nonconjugate Variational Message Passing (NCVMP, Knowles and Minka, 2011) in the E-step.

Likelihood. For notational clarity it will be convenient to introduce the auxiliary variable $g_{is} = \beta_s x_i + \mu_s + \epsilon_{is} \sim N(\beta_s x_i + \mu_s, v_s)$, where x_i is the vector of fixed effect covariates (potentially including environment, heterozygosity of the lead eSNP, and an interaction term, see Section 10) so that the data likelihood is

$$\prod_{i,s} \sigma(g_{is})^{y_{is}} [1 - \sigma(g_{is})]^{n_{is} - y_{is}} =: f_{is}(g_{is}). \tag{21}$$

The log likelihood is then

$$\sum_{is} g_{is} y_{is} - n_{is} \log(1 + e^{g_{is}}) \tag{22}$$

Variational approximation. We assume a fully factorized variational approximation $\prod_s q_s(1/v_s) \prod_i q_{is}(g_{is})$ where we use a Gaussian variational distribution $q_{is}(g_{is}) = N(m_{is}, v_{is})$ on each g_{is}, and a gamma distribution $q_s(1/v_s) = G(1/v_s, a_s, b_s)$ on the random effect precisions $1/v_s$, where a_s, b_s are shape and rate parameters respectively.

E-step: random effects. The updates for the variational posterior on g_{is}, i.e. $q_{is}(g_{is}) = N(m_{is}, v_{is})$ are

$$\begin{aligned} \frac{1}{v_{is}} &\leftarrow \frac{1}{v_f^{is}} + \frac{1}{\hat{v}_s} \\ \frac{m_{is}}{v_{is}} &\leftarrow \frac{m_f^{is}}{v_f^{is}} + \frac{\beta_s x_i + \mu_s}{\hat{v}_s} \end{aligned} \tag{23}$$

where the first term in both updates come from the Gaussian approximation to f_{is} provided by NCVMP, i.e. $N(m_f^{is}, v_f^{is})$, the second term comes from the linear regression, and $1/\hat{v_s} := \langle 1/v_s \rangle_q$ is the expected random effect precision (reciprocal variance) under the current $q_s(v_s)$. The NCVMP message (Gaussian approximation) to g_{is} from f_{is} is $N(m_f^{is}, v_f^{is})$, calculated as follows (suppressing the indexing over i and s),

$$\begin{aligned} \frac{1}{v_f} &= -2\frac{dS}{dv} \\ \frac{m_f}{v_f} &= \frac{m}{v_f} + \frac{dS}{dm} \end{aligned} \tag{24}$$

where $S := \int q(g) \log f(g) dg = \mathbb{E}_q[\log f(g)]$. See Knowles and Minka (2011) for details. The term $\log f(g)$ is

$$gy - n \log(1 + e^g) \tag{25}$$

Taking the expectation w.r.t. $q(g)$ for the first term is straightforward but the second we lower bound as follows (Saul and Jordan, 1998):

$$\mathbb{E}_q[\log(1 + e^g)] \leq \frac{1}{2}\alpha^2 v + \log(1 + e^{m+(1-2\alpha)v/2}) \tag{26}$$

where $\alpha \in [0, 1]$ is an additional variational parameter we optimize using the fixed point update $\alpha \leftarrow \sigma(m + (1-2\alpha)v/2)$. It is then easy to show the updates are

$$\begin{aligned} \frac{1}{v_f} &\leftarrow n\sigma(m + (1-2\alpha)v/2)[1 - \sigma(m + (1-2\alpha)v/2)] \\ \frac{m_f}{v_f} &\leftarrow \frac{m}{v_f} + y - n\sigma(m + (1-2\alpha)v/2) \end{aligned} \tag{27}$$

E-step: random effect variance. The updates for the random effect variance posterior distribution $q_s(1/v_s) = G(1/v_s, a_s, b_s)$ are

$$a_s \leftarrow a + \frac{1}{2}N_s \tag{28}$$

$$\begin{aligned} b_s &\leftarrow b + \frac{1}{2}\sum_{i=1}^{N_s} \mathbb{E}_q\left[(g_{is} - (\beta_s x_i + \mu_s))^2\right] \\ &= b + \frac{1}{2}\sum_{i=1}^{N_s} \left[(m_{is} - (\beta_s x_i + \mu_s))^2 + v_{is}\right] \end{aligned} \tag{29}$$

where N_s is the number of individuals testable at exonic SNP s.

M-step: fixed effects. The updates for β, μ are straightforward: letting $\tilde{\mathbf{x}}_s = [\mathbf{x}_s, \mathbf{1}]$ then

$$[\beta_s, \mu_s]^T \leftarrow (\tilde{\mathbf{x}}_s^T \tilde{\mathbf{x}}_s)^{-1} \tilde{\mathbf{x}}_s^T m_{:s}$$

We precompute Cholesky factors of $\tilde{\mathbf{x}}_s^T \tilde{\mathbf{x}}_s$ for efficiency.

M-step: overdispersion hyperparameters. The updates for a, b are equivalent to fitting the parameters of a gamma distribution, expect that we have uncertainty in the data. Define

$$\xi := \log\left(\frac{1}{S}\sum_{s=1}^{S} a_s/b_s\right) - \frac{1}{S}\sum_{i=1}^{S}[\psi(a_s) - \log b_s] \tag{30}$$

where ψ is the digamma function and S is the total number of tested sites (exonic SNPs). Then the maximum likelihood shape a is approximately

$$a \approx \frac{3 - \xi + \sqrt{(\xi - 3)^2 + 24\xi}}{12\xi} \tag{31}$$

(see Minka (2002) for details). Using Equation 31 as an initialization we proceed with a Newton-Raphson update

$$a \leftarrow a - \frac{\log(a) - \psi(a) - \xi}{\frac{1}{a} - \psi'(a)} \tag{32}$$

which we run to convergence (defined as absolute change in a less than 10^{-6}, which typically takes fewer than 20 iterations). Given the MLE of a the MLE of b is simply

$$b \leftarrow a \left(\frac{1}{S} \sum_{s=1}^{S} a_s / b_s \right)^{-1} \tag{33}$$

Calculating the variational lower bound. Variational EM provides an approximation to the marginal likelihood $P(y|\beta, \mu, a, b)$ through the lower bound

$$\log P(y|\beta, \mu, a, b) \geq \mathbb{E}_q[\log P(y, g, v|\beta, \mu, a, b)] + H[q] \tag{34}$$

where the entropy $H[q] = -\mathbb{E}_q[\log q(g, v)]$. The right hand side of Equation 34 is optimized (maximized) w.r.t. to q (in the E-step) and β, μ, a, b (in the M-step) by the variational EM procedure. Since q is a product of univariate Gaussian and gamma distributions calculating $H[q]$ is straightforward:

$$H[q] = \sum_s H_G(a_s, b_s) + \sum_{is} H_N(m_{is}, v_{is}) \tag{35}$$

$$H_G(a, b) := -\log b + \log[\Gamma(a)] + (1 - a)\psi(a) \tag{36}$$

$$H_N(m, v) := \frac{1}{2}[\log(2\pi v) + 1] \tag{37}$$

The data term $\mathbb{E}_q[\log P(y, g, v|\beta, \mu, a, b)]$ is

$$\begin{aligned} &\sum_s \Bigg(a \log b - \log \Gamma(a) + (a - 1)(\psi(a_s) - \log b_s) - b\frac{a_s}{b_s} \\ &+ \sum_i \Big\{ -\frac{1}{2}\log 2\pi + \frac{1}{2}(\psi(a_s) - \log b_s) - \frac{1}{2}\frac{a_s}{b_s}[(m_{is} - \beta_s^T \tilde{x}_s)^2 + v_{is}] \\ &+ m_{is} y_{is} - n_{is}\left[\frac{1}{2}\alpha_{is}^2 v_{is} + \log\left(1 + \exp\left(m_{is} + (1 - 2\alpha_{is})v_{is}/2\right)\right)\right] \Big\} \Bigg) \end{aligned} \tag{38}$$

where the three lines correspond to the gamma prior on $1/v_s$, the likelihood for g_{is} and the data likelihood f_{is} respectively. Adding Equation 35 and 38 gives

the global lower bound which we use to assess convergence of the variational EM algorithm (we require a change less than 0.1). Extracting only the terms for a specific exonic SNP s provides an approximation of $\log P(y_{:s}|\beta_s, a, b)$, with $g_{:s}$ and v_s integrated out, which we use as the likelihood in our likelihood ratio test.

References

Aich, J., Mabalirajan, U., Ahmad, T., Agrawal, A., and Ghosh, B. (2012). Loss-of-function of inositol polyphosphate-4-phosphatase reversibly increases the severity of allergic airway inflammation. *Nature communications*, 3:877.

Al-Daghri, N. M., Clerici, M., Al-Attas, O., Forni, D., Alokail, M. S., Alkharfy, K. M., Sabico, S., Mohammed, A. K., Cagliani, R., and Sironi, M. (2013). A nonsense polymorphism (R392X) in TLR5 protects from obesity but predisposes to diabetes. *The Journal of Immunology*, 190(7):3716–3720.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., and Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic acids research*, 37(Web Server issue):W202–8.

Battle, A., Mostafavi, S., Zhu, X., Potash, J. B., Weissman, M. M., McCormick, C., Haudenschild, C. D., Beckman, K. B., Shi, J., Mei, R., et al. (2014). Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome research*, 24(1):14–24.

Biondi, O., Villemeur, M., Marchand, A., Chretien, F., Bourg, N., Gherardi, R. K., Richard, I., and Authier, F.-J. (2013). Dual effects of exercise in dysferlinopathy. *The American journal of pathology*, 182(6):2298–2309.

Boer, R. A., Voors, A. A., Muntendam, P., Gilst, W. H., and Veldhuisen, D. J. (2009). Galectin-3: a novel mediator of heart failure development and progression. *European journal of heart failure*, 11(9):811–817.

Chen, Z., Tang, H., Qayyum, R., Schick, U. M., Nalls, M. A., Handsaker, R., Li, J., Lu, Y., Yanek, L. R., Keating, B., et al. (2013). Genome-wide association analysis of red blood cell traits in African Americans: the COGENT network. *Human molecular genetics*, 22(12):2529–2538.

ENCODE Project Consortium (2004). The ENCODE (ENCyclopedia of DNA elements) project. *Science*, 306(5696):636–640.

Fairfax, B. P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., Jostins, L., Plant, K., Andrews, R., McGee, C., et al. (2014). Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science*, 343(6175).

Grant, C. E., Bailey, T. L., and Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics (Oxford, England)*, 27(7):1017–8.

Jacobo-Albavera, L., Aguayo-de la Rosa, P. I., Villarreal-Molina, T., Villamil-Ramírez, H., León-Mimila, P., Romero-Hidalgo, S., López-Contreras, B. E., Sánchez-Muñoz, F., Bojalil, R., González-Barrios, J. A., et al. (2012). VNN1 gene expression levels and the G-137T polymorphism are associated with HDL-C levels in Mexican prepubertal children. *PLOS ONE*.

Jones, R. D., Taylor, A. M., Tong, E. Y., and Repa, J. J. (2013). Carboxylesterases are uniquely expressed among tissues and regulated by nuclear hormone receptors in the mouse. *Drug Metabolism and Disposition*, 41(1):40–49.

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., and Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine learning*, 37(2):183–233.

Knowles, D. A. and Minka, T. (2011). Non-conjugate variational message passing for multinomial and binary regression. In *Advances in Neural Information Processing Systems*, pages 1701–1709.

Kubo, T., Kinjyo, N., Ikezawa, A., Kambe, T., and Fukumori, R. (2003). Sp1 decoy oligodeoxynucleotide decreases angiotensin receptor expression and blood pressure in spontaneously hypertensive rats. *Brain research*, 992(1):1–8.

Liu, X., Niu, Y., Yuan, H., Huang, J., and Fu, L. (2015). AMPK binds to sestrins and mediates the effect of exercise to increase insulin-sensitivity through autophagy. *Metabolism*, 64(6):658–665.

Lumeng, C. N., DeYoung, S. M., Bodzin, J. L., and Saltiel, A. R. (2007). Increased inflammatory properties of adipose tissue macrophages recruited during diet-induced obesity. *Diabetes*, 56(1):16–23.

Ma, B., Kang, M.-J., Lee, C. G., Chapoval, S., Liu, W., Chen, Q., Coyle, A. J., Lora, J. M., Picarella, D., and Homer, R. J. (2005). Role of ccr5 in ifn-γ–induced and cigarette smoke–induced emphysema. *The Journal of Clinical Investigation*, 115(12):3460–3472.

Melén, E. and Bottai, M. (2012). On lung function and interactions using genome-wide data. *PLoS Genetics*, 8(12):e1003174.

Michou, L., Cornélis, F., Levesque, J.-M., Bombardieri, S., Balsa, A., Westhovens, R., Barrera, P., Alves, H., van de Putte, L., Migliorini, P., et al. (2012). A genetic association study of the CLEC12A gene in rheumatoid arthritis. *Joint Bone Spine*, 79(5):451–456.

Minka, T. P. (2002). Estimating a gamma distribution. Technical report, Microsoft Research, Cambridge, UK.

Mollan, T. L., Banerjee, S., Wu, G., Siburt, C. J. P., Tsai, A.-L., Olson, J. S., Weiss, M. J., Crumbliss, A. L., and Alayash, A. I. (2013). α-Hemoglobin stabilizing protein (AHSP) markedly decreases the redox potential and reactivity of α-subunits of human HbA with hydrogen peroxide. *Journal of Biological Chemistry*, 288(6):4288–4298.

Negoro, N., Kanayama, Y., Haraguchi, M., Umetani, N., Nishimura, M., Konishi, Y., Iwai, J., Okamura, M., Inoue, T., and Takeda, T. (1995). Blood pressure regulates platelet-derived growth factor A-chain gene expression in vascular smooth muscle cells in vivo. An autocrine mechanism promoting hypertensive vascular hypertrophy. *The Journal of clinical investigation*, 95(3):1140–50.

Pastore, A. and Piemonte, F. (2012). S-Glutathionylation signaling in cell biology: progress and prospects. *European Journal of Pharmaceutical Sciences*, 46(5):279–292.

Pejnovic, N. N., Pantic, J. M., Jovanovic, I. P., Radosavljevic, G. D., Milovanovic, M. Z., Nikolic, I. G., Zdravkovic, N. S., Djukic, A. L., Arsenijevic, N. N., and Lukic, M. L. (2013). Galectin-3 deficiency accelerates high-fat diet–induced obesity and amplifies inflammation in adipose tissue and pancreatic islets. *Diabetes*, 62(6):1932–1944.

Philip, V., Duvvuru, S., Gomero, B., Ansah, T., Blaha, C., Cook, M., Hamre, K., Lariviere, W., Matthews, D., Mittleman, G., et al. (2010). High-throughput behavioral phenotyping in the expanded panel of BXD recombinant inbred strains. *Genes, Brain and Behavior*, 9(2):129–159.

Pique-Regi, R., Degner, J. F., Pai, A. A., Gaffney, D. J., Gilad, Y., and Pritchard, J. K. (2011). Accurate inference of transcription factor binding from dna sequence and chromatin accessibility data. *Genome Research*, 21(3):447–455.

Pollin, T. I., Hsueh, W.-C., Steinle, N. I., Snitker, S., Shuldiner, A. R., and Mitchell, B. D. (2004). A genome-wide scan of serum lipid levels in the Old Order Amish. *Atherosclerosis*, 173(1):89–96.

Rice, T., Rankinen, T., Province, M. A., Chagnon, Y. C., Pérusse, L., Borecki, I. B., Bouchard, C., and Rao, D. (2000). Genome-wide linkage analysis of systolic and diastolic blood pressure The Québec Family Study. *Circulation*, 102(16):1956–1963.

Rohrwasser, A., Zhang, S., Dillon, H. F., Inoue, I., Callaway, C. W., Hillas, E., and Lalouel, J.-M. (2002). Contribution of Sp1 to initiation of transcription of angiotensinogen. *Journal of human genetics*, 47(5):249–56.

Saul, L. and Jordan, M. (1998). A mean field learning algorithm for unsupervised neural networks. In *Learning in graphical models*, pages 541–554. Springer.

Sharma, M., Batra, J., Mabalirajan, U., Sharma, S., Nagarkatti, R., Aich, J., Sharma, S. K., Niphadkar, P. V., and Ghosh, B. (2008). A genetic variation in inositol polyphosphate 4 phosphatase a enhances susceptibility to asthma. *American journal of respiratory and critical care medicine*, 177(7):712–719.

Steinberg, S., Stefansson, H., Jonsson, T., Johannsdottir, H., Ingason, A., Helgason, H., Sulem, P., Magnusson, O. T., Gudjonsson, S. A., Unnsteinsdottir, U., et al. (2015). Loss-of-function variants in ABCA7 confer risk of alzheimer's disease. *Nature genetics*.

van der Harst, P., Zhang, W., Leach, I. M., Rendon, A., Verweij, N., Sehmi, J., Paul, D. S., Elling, U., Allayee, H., Li, X., et al. (2012). Seventy-five genetic loci influencing the human red blood cell. *Nature*, 492(7429):369–375.

Walsh, N. P., Gleeson, M., Shephard, R. J., Gleeson, M., Woods, J. A., Bishop, N. C., Fleshner, M., Green, C., Pedersen, B. K., Hoffman-Goetz, L., Rogers, C. J., Northoff, H., Abbasi, A., and Simon, P. (2011). Position statement part one: Immune function and exercise.

Zhang, Y., Cheng, X., Aleksunes, L., and Klaassen, C. D. (2012). Transcription factor-mediated regulation of carboxylesterase enzymes in livers of mice. *Drug Metabolism and Disposition*, 40(6):1191–1197.

Zhao, L.-J., Xiao, P., Liu, Y.-J., Xiong, D.-H., Shen, H., Recker, R. R., and Deng, H.-W. (2007). A genome-wide linkage scan for quantitative trait loci underlying obesity related phenotypes in 434 Caucasian families. *Human genetics*, 121(1):145–148.

Environment	Non-zero counts	Description
Sex	274 (males)	
Age	922 (continuous)	
Number cigarettes per day	118	
Smoked same day	101	
Time of day	922 (continuous)	
Blood pressure medication	183	
Exercise	166	
OBCP	84	Over-the-counter birth control medication
Antidepressants	219	
Benzodiazepine	63	
Ace inhibitors	66	
Cholesterol lowering	114	
Anti histamine	99	
Thyroid medication	88	
PPI	84	Protein pump inhibitors
NSAIDs	65	Nonsteroidal anti-inflammatory drug
Beta blockers	58	
MDD	459	Major depressive disorder
Family history depression	452	
Diuretic medication	54	
Opiates	55	
Cardiac medication	49	
Steroid medication	53	
Decongestant medication	51	
Bupropion	53	
Oral hypoglyemic	46	
SNRI	46	Serotonin-norepinephrine reuptake inhibitor
BMI	922 (continuous)	Body mass index
Alcohol use	36	
Cannabis use	32	

Table 1: Environmental factors tested across the $N = 922$ individuals.

Environment	Gene	EAGLE p	QTL p	Notes
Sex	TLR5	3.9×10^{-6}	0.993	Innate pathogen recognition. LoF variant has sex specific influence on BMI (Al-Daghri et al., 2013)
Age	THEMIS	8.4×10^{-7}	1.000	Regulation of T-cell maturation.
Smoke before	INPP4A	2.9×10^{-7}	0.049	Phosphatidylinositol signaling pathway enzyme. Influences asthma susceptibility: (Sharma et al., 2008; Aich et al., 2012)
Time of day	NCAPH2,SCO2	1.5×10^{-5}	1.000	SCO2: cytochrome c oxidase. Circadb: p=1e-18
Time of day	NCR3	8.9×10^{-7}	1.000	Natural killer cell cytotoxicity receptor.
Time of day	HLA-DOB	1.8×10^{-7}	1.000	Major histocompatibility complex. Immune activity varies with circadian rhythym.
Time of day	OLFM4	1.2×10^{-9}	1.000	Stem cell marker.
Blood pressure meds	NPRL3	2.1×10^{-6}	1.000	Homeostasis of fluid volume. Expressed in the vasculature. Associated with mean corpuscular hemoglobin concentration (MCHC). (Chen et al., 2013; van der Harst et al., 2012)
Exercise	IFI44L	5.2×10^{-7}	1.000	Interferon-induced protein 44-like
Exercise	SESN3,RP11-712B9.2	1.1×10^{-6}	1.000	Sestrin stress-induced protein. Reduces intracellular oxygen, regulates blood glucose. Influenced by exercise (Liu et al., 2015)
Exercise	DYSF	1.5×10^{-6}	1.000	Linked with skeletal muscle repair. (Biondi et al., 2013)
Exercise	CLEC12A	4.2×10^{-6}	1.000	Suppresses monocyte immune cells. Associated with RA (Michou et al., 2012)
Antidepressants	ABCA7	3.4×10^{-8}	0.786	ATP-binding cassette sub-family A member 7. LoF variants double Alzheimer's risk (Steinberg et al., 2015)
Antidepressants	NKG7	8.8×10^{-6}	0.014	Natural Killer Cell Granule Protein.
Ace inhibitors	STEAP4,AC003991.3	7.2×10^{-6}	1.000	Metalloreductase
Thyroid meds	PHGDH	1.9×10^{-9}	1.000	Serine biosynthesis
Fam. hist. depression	FCGR2B	1.5×10^{-6}	1.000	Inhibits antibody production in B-cells.
Fam. hist. depression	CHIT1	4.5×10^{-7}	1.000	Secreted by macrophages
Fam. hist. depression	ABCA7	1.5×10^{-6}	1.000	See above
Fam. hist. depression	ABCA7	3.8×10^{-14}	1.000	See above
Opiates	ARHGEF18	4.4×10^{-7}	0.134	Rho/Rac guanine nucleotide exchange factor (GEF) 18
Opiates	ZFAT,ZFAT-AS1	1.4×10^{-7}	0.734	Correlates with activity with morphine in mice (Philip et al., 2010)
Decongestant meds	CDH23,C10orf54	4.9×10^{-6}	0.296	Expressed in the neurosensory epithelium.
Decongestant meds	FES,AC068831.1	4.4×10^{-6}	1.000	Tyrosine-protein kinase with various roles including mast cell signalling
Decongestant meds	ZMAT2	3.6×10^{-6}	0.536	Zinc finger, matrin-type 2
SNRI	FADS1,FADS2	1.8×10^{-5}	0.248	Fatty acid desaturase 1/2
SNRI	EIF4E3	7.0×10^{-6}	0.103	Eukaryotic translation initiation factor 4E
BMI	GSTO1	5.6×10^{-8}	1.000	glutathione S-transferase omega 1. Cellular redox homeostasis, metabolism (Pastore and Piemonte, 2012). Upregulated in macrophages recruited to adipose during obesity (Lumeng et al., 2007)
BMI	LGALS3	4.7×10^{-5}	0.359	Lectin, Galactoside-Binding, Soluble, 3. Cardiovascular disease role (Boer et al., 2009). -/- mice are obese (Pejnovic et al., 2013)s
BMI	CD93	1.6×10^{-5}	0.506	C-type lectin transmembrane receptor. Weak (1e-4) association with total fat (Zhao et al., 2007)
BMI	KIAA0930	2.5×10^{-5}	1.000	Weak BP association (4e-3) (Rice et al., 2000)
BMI	NUP210	3.0×10^{-5}	0.517	Nucleoporin 210kDa. Upregulated in macrophages recruited to adipose during obesity (Lumeng et al., 2007).
BMI	C9orf78	1.1×10^{-5}	0.517	Weak (p=0.003) assocation with total cholesterol (Pollin et al., 2004)
BMI	C9orf78,USP20	4.3×10^{-5}	1.000	
BMI	AHSP	1.5×10^{-5}	0.062	Alpha hemoglobin stabilizing protein. Protects hemoglobin from oxidation (Mollan et al., 2013)
BMI	ITK	1.0×10^{-5}	1.000	Interleukin-2-inducible T-cell kinase
BMI	VNN1	3.8×10^{-5}	0.914	Regulates cholesterol levels (Jacobo-Albavera et al., 2012)

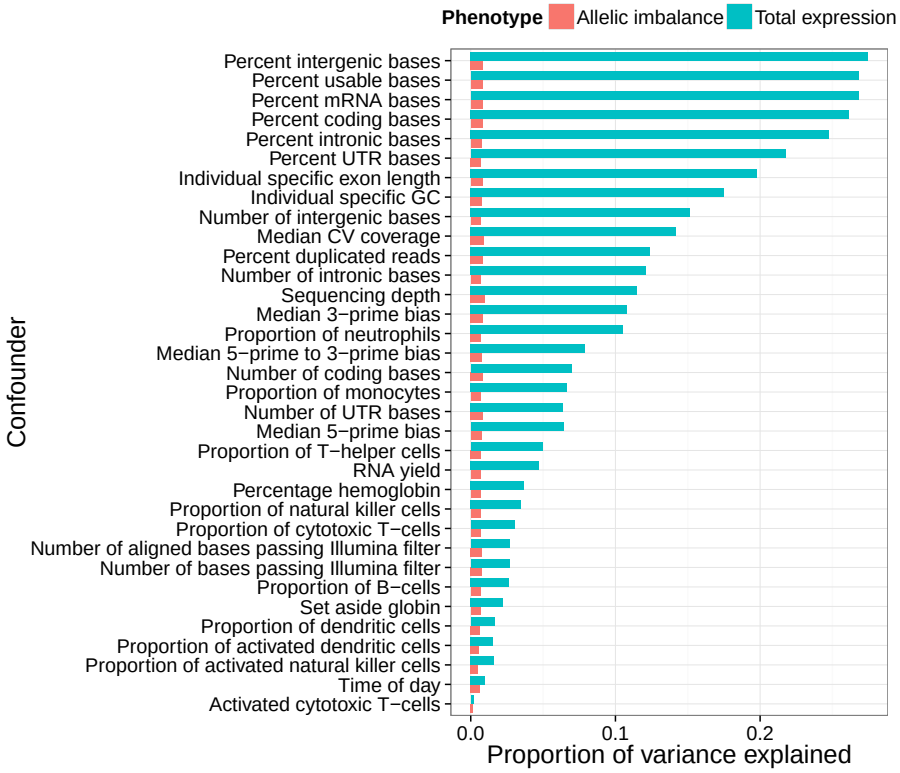
Table 2: Putative GxE interactions detected by EAGLE, at an FDR of 0.1. If the tested SNP lies in an exon of two different overlapping genes then both genes are listed. QTL p is the Bonferroni corrected p-value from standard interaction QTL testing.

Environment	Gene	p-value	Chromosome	ASE Position	Candidate position
Number cig per day	IL10RA	4.8×10^{-2}	11	117864113	117709028
Smoked same day	INPP4A	7.8×10^{-3}	2	99149946	98761589
Blood pressure medication	FAAH	4.4×10^{-2}	1	46870761	46927103
Blood pressure medication	CPT1B	1.9×10^{-2}	22	51015838	51057923
Exercise	IFI44L	2.3×10^{-3}	1	79095581	79167440
Exercise	DYSF	6.7×10^{-3}	2	71688249	71703945
Antidepressants	NKG7	4.7×10^{-2}	19	51875946	51823015
NSAIDs	TAPBPL	9.8×10^{-3}	12	6567907	6585610
NSAIDs	SSNA1	3.6×10^{-3}	9	140084376	140356374
Opiates	ARHGEF18	3.8×10^{-7}	19	7537107	7538117
Opiates	ZFAT	8.9×10^{-3}	8	135612745	136017930
Decongestant medication	CDH23	8.6×10^{-4}	10	73510469	73517225
Decongestant medication	ZMAT2	4.3×10^{-3}	5	140086062	140115433
BMI	GSTO1	1.6×10^{-3}	10	106022789	106033903
BMI	VNN1	2.3×10^{-4}	6	133035098	132996866

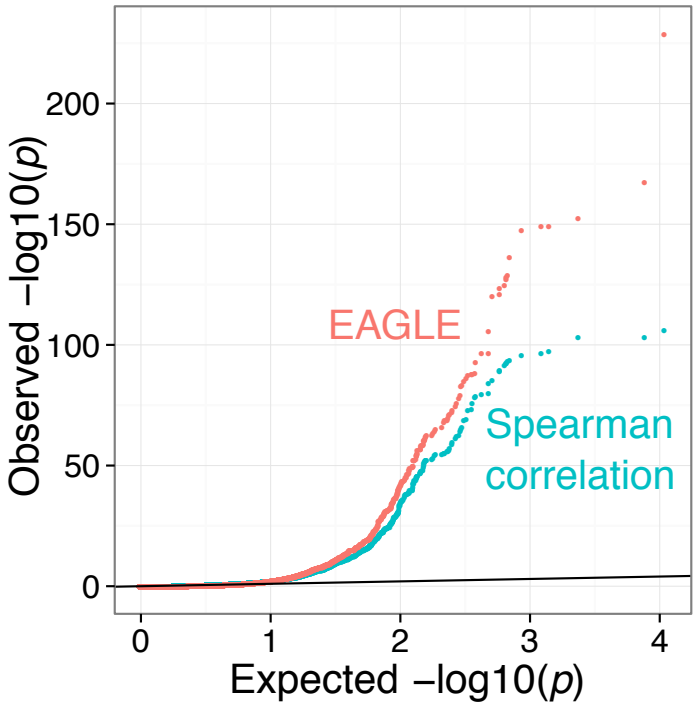
Table 3: Candidate variants detected using two stage analysis. We first detected potential GxE interactions using EAGLE with a lenient FDR of 0.2, resulting in a shortlist of 57 gene-environment pairs. We then tested SNPs within 1Mb of each gene's TSS using both standard interaction testing and EAGLE including heterozygosity of the SNP. We combined p-values from these two tests using Fisher's combined test. The resulting significant p-values are reported here, after Bonferroni correction across the tested SNPs.

Environment	Gene	Chr	Candidate position	GWAS trait	GWAS p-value	GWAS SNP position	LD
Blood pressure meds	TACC3	4	1721654	Height	2.0×10^{-11}	1701317	0.4
Age	THEMIS	6	128269592	Multiple sclerosis	6.0×10^{-9}	128278798	0.42
Age	THEMIS	6	128269592	Celiac disease	4.3×10^{-15}	128282758	0.43
Time of day	NCR3	6	31461132	Hodgkin's lymphoma	7.0×10^{-16}	31446796	0.52
Time of day	NCR3	6	31461132	Autoimmune thyroiditis	7.5×10^{-18}	31448625	0.54
Time of day	NCR3	6	31461132	Celiac disease	1.5×10^{-34}	31448625	0.54
Time of day	NCR3	6	31461132	Multiple schlerosis	6.0×10^{-45}	31448625	0.54
Time of day	NCR3	6	31461132	Psoriasis	1.7×10^{-74}	31448625	0.54
Time of day	NCR3	6	31461132	Rheumatoid arthritis	2.4×10^{-10}	31480272	0.4
BMI	LGALS3	14	55879277	Protein biomarker for LGALS3	2.0×10^{-188}	55614636	0.55
BMI	AHSP	16	31276811	Systemic lupus erythematosus/ systemic sclerosis	1.0×10^{-10}	31326706	0.9
BMI	AHSP	16	31276811	Systemic lupus erythematosus	2.0×10^{-23}	31313253	0.87
Antidepressants	ABCA7	19	1127615	Crohn's disease	8.0×10^{-22}	1124031	0.82
Decongestant meds	FES	15	91543761	Breast cancer	4.0×10^{-8}	91512067	0.43
Decongestant meds	FES	15	91543761	Type 2 diabetes	2.0×10^{-10}	91521337	0.72
Decongestant meds	C20orf3	20	25178119	Allergic rhinitis	1.0×10^{-6}	25206654	0.67
Decongestant meds	C20orf3	20	25178119	Liver enzyme levels (alkaline phosphatase)	7.0×10^{-10}	25298087	0.56

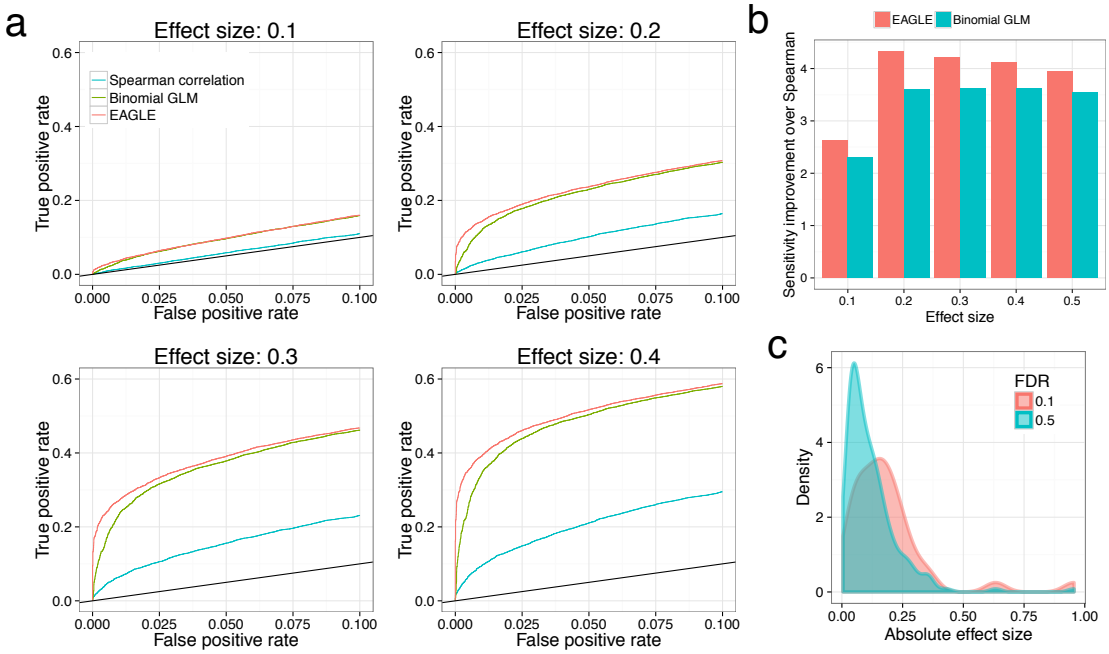
Table 4: GWAS variants in LD with candidate variants from EAGLE.



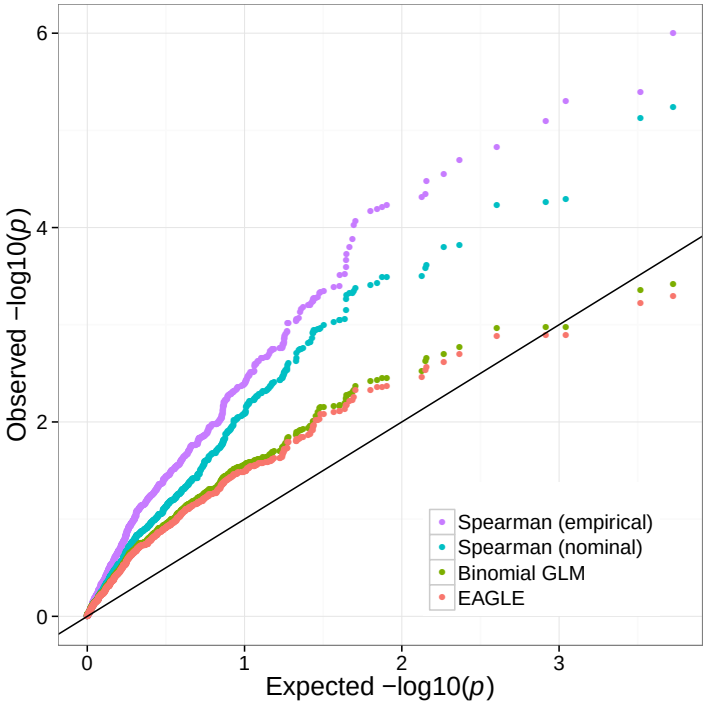
Supplementary Figure 1: Proportion variance explained (PVE) by technical covariates and confounders for allelic imbalance and total expression, across all genes. We define allelic imbalance as $|\frac{1}{2} - y/n|$ where y is the alternative read count and n is the total read count, where we only consider individuals with at total read count $n \geq 30$, and we subsample down to $n = 30$. For total expression we calculated PVE for each gene for the same individuals we could test for allelic imbalance (i.e. individuals with a heterozygous exonic SNP and $n \geq 30$), in order to match power between the two phenotypes. The confounders include sequencing depth, PICARD quality measures, and estimates of cell type proportions (estimated as previously described in Battle et al. (2014)).



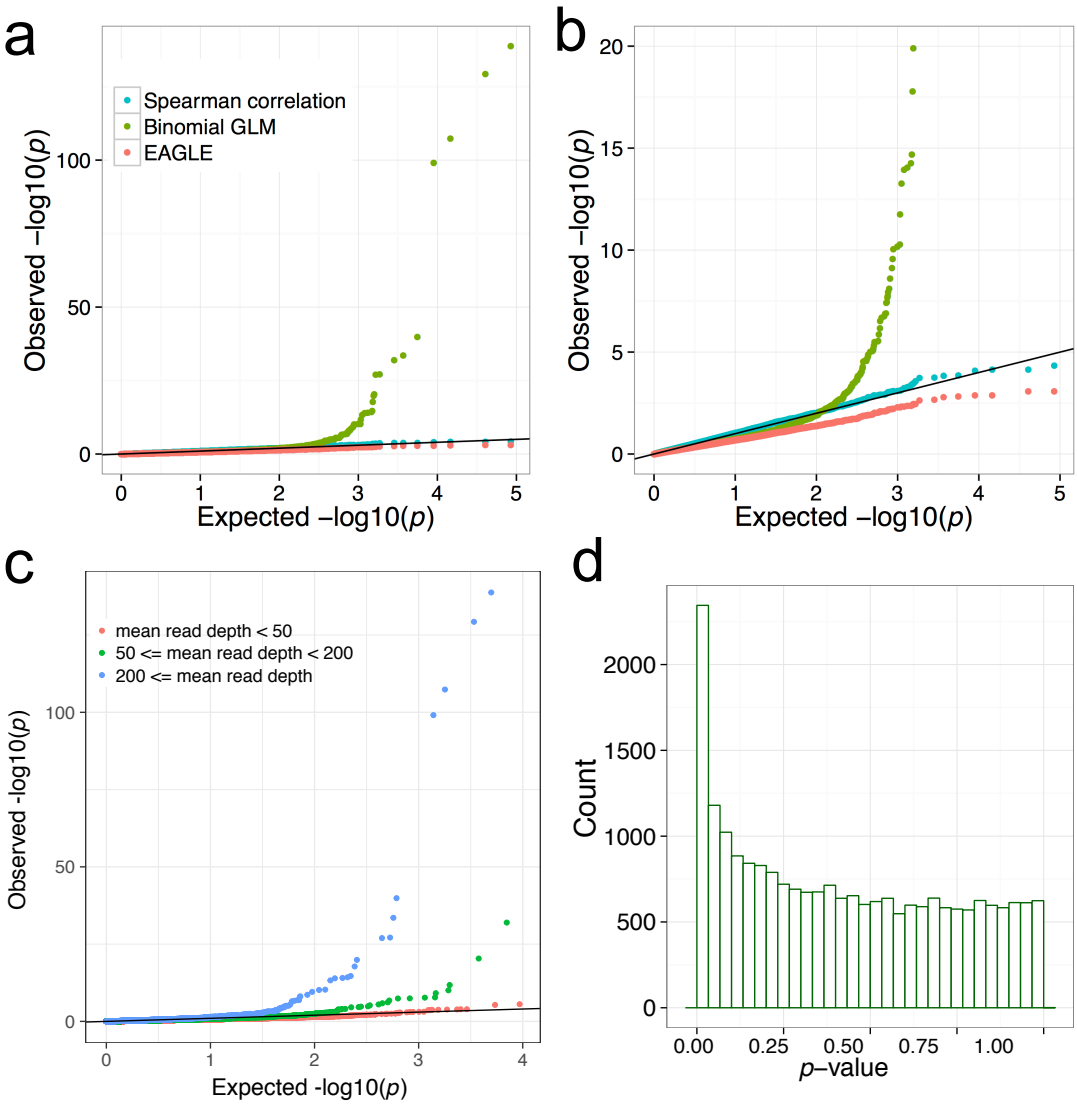
Supplementary Figure 2: EAGLE improves power to detect aseQTLs. Here we tested whether the lead eSNP from total expression analysis was also an allele specific expression QTL (aseQTLs), using both EAGLE and Spearman correlation of allelic imbalance, defined as $|\frac{1}{2} - y/n|$, vs. a binary variable indicating whether the eSNP is heterozygous. By explicitly modeling the count nature of the data EAGLE improves power over Spearman correlation.



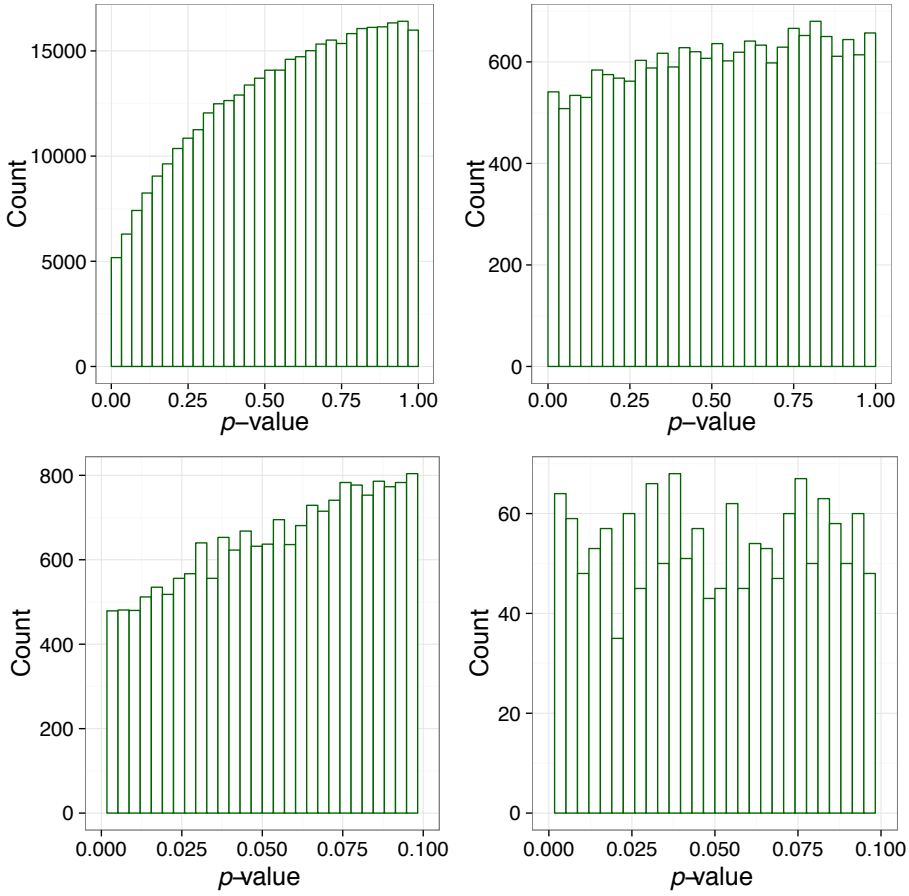
Supplementary Figure 3: EAGLE has better sensitivity and specificity than using a binomial GLM or Spearman correlation between allelic imbalance and environment. We simulated allelic reads at all exonic SNPs tested in the DGN cohort, maintaining the true total counts but sampling the alternative allele count using a maximum likelihood estimate of the overdispersion and varying the true effect size for the time of day environment (see Supplementary Methods Section 2 for simulation details). **a**. ROC curves are shown for Spearman correlation with allelic imbalance defined as $|\frac{1}{2} - y/n|$, a binomial GLM, and EAGLE. Across all effect size settings EAGLE outperforms the binomial GLM and Spearman correlation. The binomial GLM has performance here close to that of EAGLE, but is extremely poorly calibrated (see Supplementary Figure S5). **b**. Relative increase in sensitivity (true positive rate) of EAGLE and the binomial GLM relative to Spearman correlation at a fixed FPR of 0.01 (it is not straightforward to estimate the corresponding FDR without knowing the true underlying proportion of non-null associations). The difference in sensitivity initially increases with effect size, and then slowly decreases. **c**. Distribution of absolute effect sizes for detected associations at FDR 0.1 and 0.5 in DGN. The bulk of the absolute effect sizes fall in the range 0 to 0.4, confirming that the effect sizes used in the simulation are realistic.



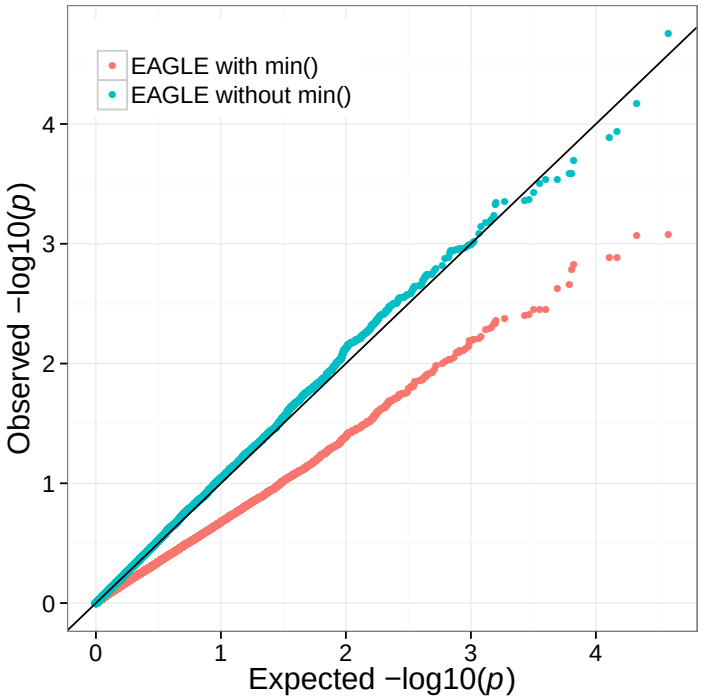
Supplementary Figure 4: Using Spearman correlation with the continuous measurement of allelic imbalance gives false positives in the presence of main effects. We simulated alternative allele counts, with no true effect of environment on allelic imbalance (see Supplementary Methods Section 2), 1000 times for an exonic SNP in the DDIT4 gene, whose total counts have moderate ($R^2 = 0.072$) but significant ($p = 2 \times 10^{-8}$) correlation with time of day. Due to not explicitly modeling the binomial sampling process, the Spearman method is highly anticonservative, and this is not alleviated using permutations to obtain empirical p-values ("Spearman (empirical)"). This is because permuting the allelic imbalance values does not break the relationship between the environment and total count, and due to binomial sampling samples with higher total counts will have allelic ratios closer to 0.5. In contrast, the binomial GLM and EAGLE, which directly model the binomial sampling, are relatively robust to this form of confounding. Even EAGLE is not completely calibrated in this simulated setting where there is a very strong main effect of the environment on total expression. However, there is only overinflation for moderate p-values, which are not critical for Benjamini-Hochberg FDR control, and b) as shown in Supplementary Figure S6, the EAGLE p-value distribution for true data is conservative. While for this particular exonic SNP the binomial GLM is reasonably calibrated, it is not in general (see Supplementary Figure S5).



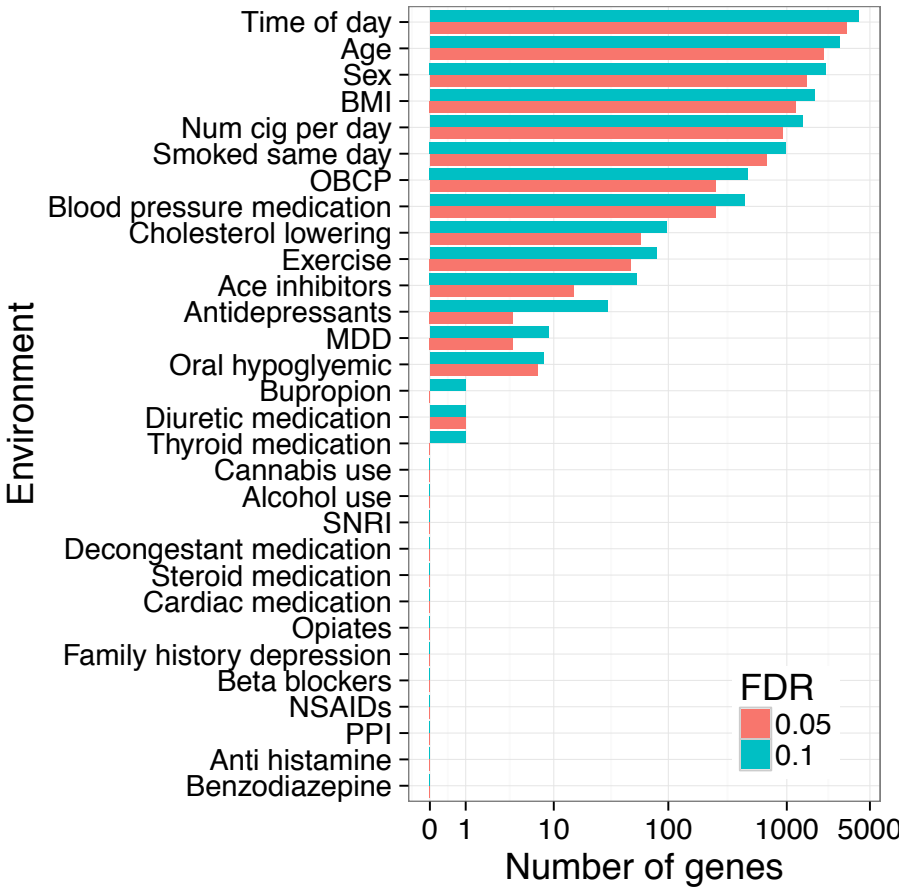
Supplementary Figure 5: Binomial GLM p-values are highly over-inflated in both simulations and permutation analyses. **a,b.** We simulated allelic reads at all exonic SNPs tested in the DGN cohort (Supplementary Note 2) with no true effect of the environment (time of day) on allelic imbalance. The binomial GLM has an excess of extremely small p-values, Spearman is well-calibrated in the absence of main effects (although EAGLE is has better power, see Supplementary Figure S3), and EAGLE is slightly conservative. **c.** The inflation of p-values under the binomial GLM is particularly problematic for exonic SNPs with higher average read depth. **d.** Permutation analysis also shows the inflation of p-values under a binomial GLM.



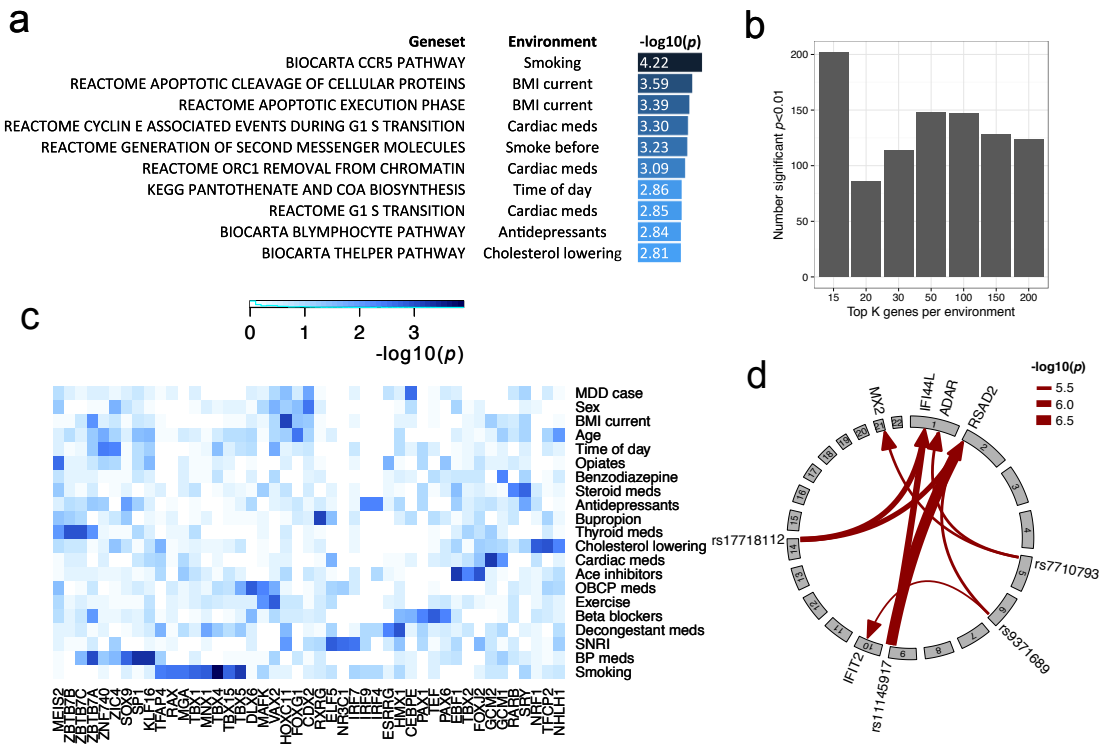
Supplementary Figure 6: EAGLE is conservative under permutations. We permuted all 30 environmental factors independently at each tested exonic SNP. The resulting p-value distribution is somewhat conservative, confirming our findings from the simulation study (Supplementary Figure 5). Based on this analysis we estimate EAGLE's nominal FDR of 10% corresponds to a true FDR of 4.9% and 8.7% for exonic SNPs with read depth below and above 300 respectively. **Left:** All exonic SNPs. **Right:** Exonic SNPs with average read depth > 300. **Top/bottom:** different zoom levels.



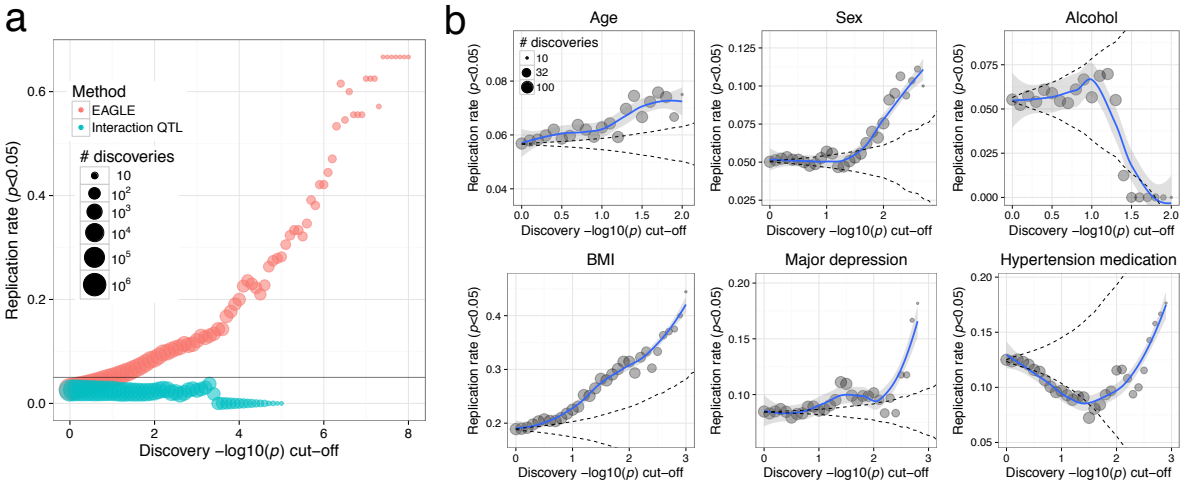
Supplementary Figure 7: Our simulation study (with no true effect) shows that EAGLE is conservative. EAGLE tests *absolute* allelic imbalance by modeling $\min(y, n - y)$. This accounts for the possibility of the causal variant(s) being in different phase with the tested exonic SNP across different individuals. An alternative, which we do not employ in this study, would be to model the *direction* of allelic imbalance by not applying the min transformation, which we show here results in a well-calibrated rather than conservative test.



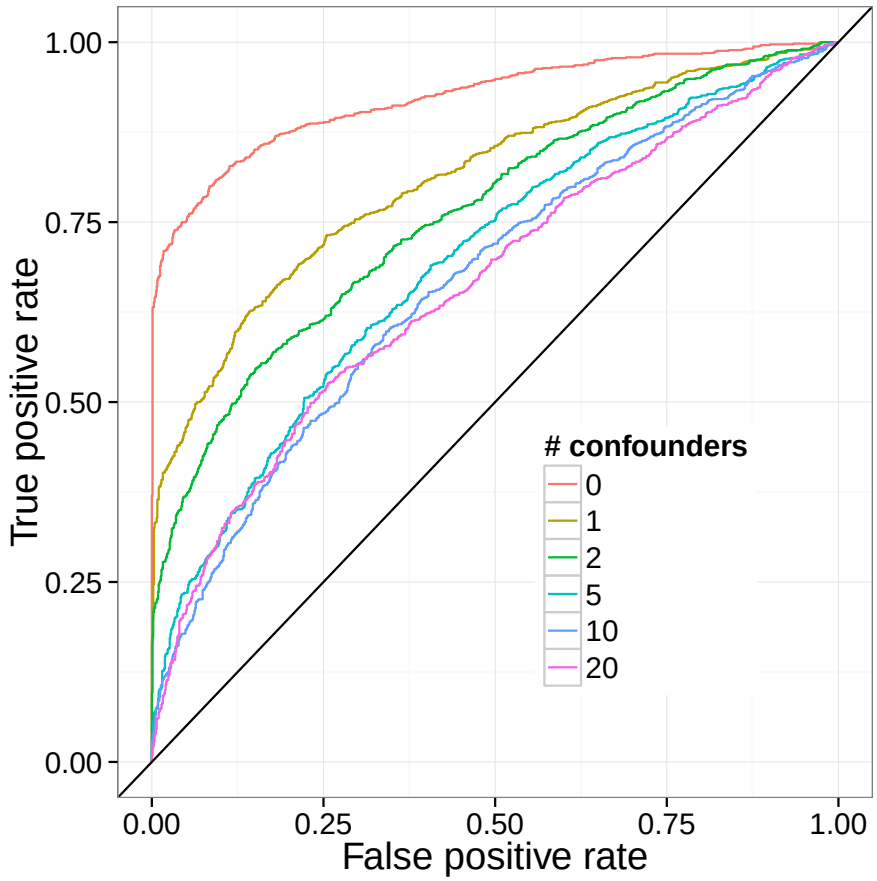
Supplementary Figure 8: Thousands of genes are environmentally responsive. We used Spearman correlation to test the association of 30 environmental factors vs total expression, followed by controlling the FDR at 0.05 or 0.1.



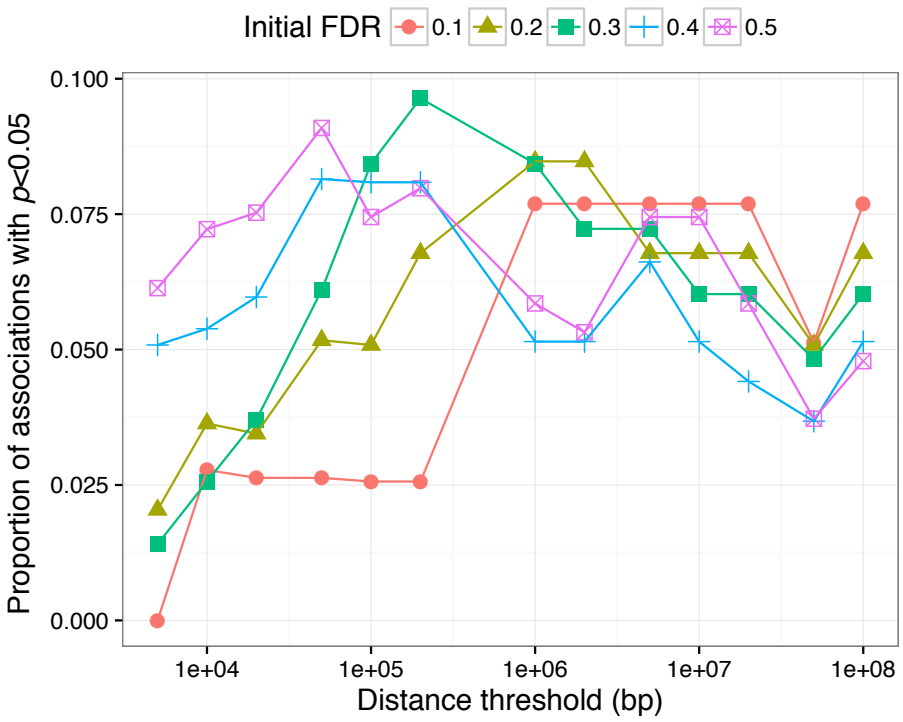
Supplementary Figure 9: Pathway, transcription factor binding site and trans-eQTL network enrichment of genes with GxE interactions for each environment reveals shared regulation. Uncorrected p-values are shown. **a.** The strongest associations between the top $K = 50$ genes for each environment and GO, KEGG and BioCarta pathways with fewer than 100 genes. **b.** The number of significant pathways is relatively robust to K, the number of top associated genes used. **c.** Enrichment of CENTIPEDE predicted TFBS within 5kb of the TSS of the top 50 genes associated with each environment. **d.** Co-regulation of genes modulated by exercise. We tested the top 50 genes associated with each environmental factor for co-regulation in a trans-eQTL network learnt from total expression analysis (taking all trans-eQTLs with $p < 10^{-5}$). Shown are trans-eQTLs regulating at least two genes in the top 50 list for exercise.



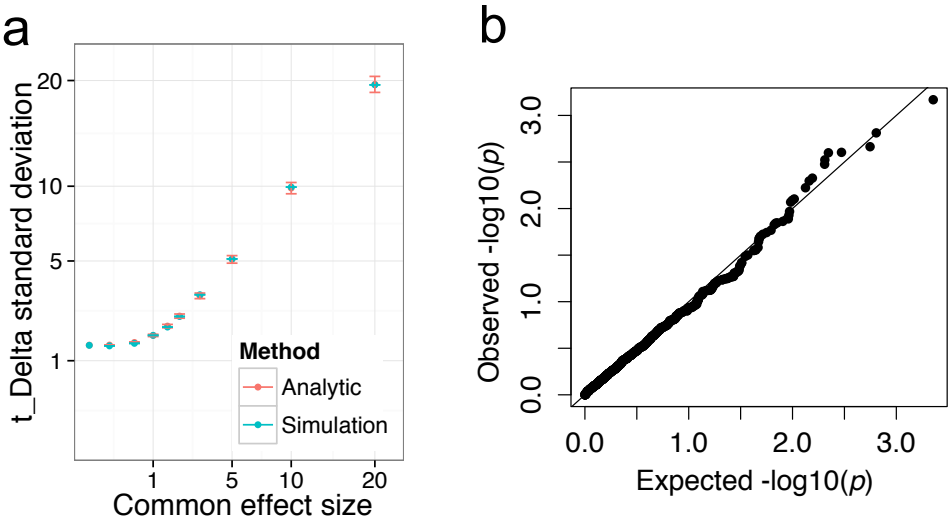
Supplementary Figure 10: **a**. We initially assessed replication by splitting the DGN cohort into two equal halves (approximately matching gender and age), calling GxE associations in one half, and checking for replication in the other half. The proportion of associations replicating at nominal p-value threshold of 0.05 is shown as a function of the p-value threshold used in the discovery set. Despite halving the sample size, using EAGLE we see striking replication as the discovery threshold is increased. In comparison, using standard interaction QTL testing no replication is observed. **b**. EAGLE associations discovered in DGN show replication in CARTaGENE, a cohort of 724 French-Canadian individuals. Plots show the proportion of GxE associations which replicate in CARTaGENE at nominal $p < 0.05$, as a function of the p-value threshold in DGN. Dashed lines denote two standard deviations of the null distribution where the discoveries are randomly chosen (calculated using the hypergeometric distribution). Out of the six environmental factors recorded in both DGN and CARTaGENE, four show significant replication with improving replication for increasingly stringent discovery thresholds. Associations for alcohol do not to replicate well, but alcohol in particular may be problematic since in DGN this variable represents whether the individual consumed alcohol before the blood draw, whereas in CARTaGENE this represents average alcohol consumption. Associations for hypertension medication do not show significant replication, although there is an upturn in the replication rate for the most statistically significant discoveries.



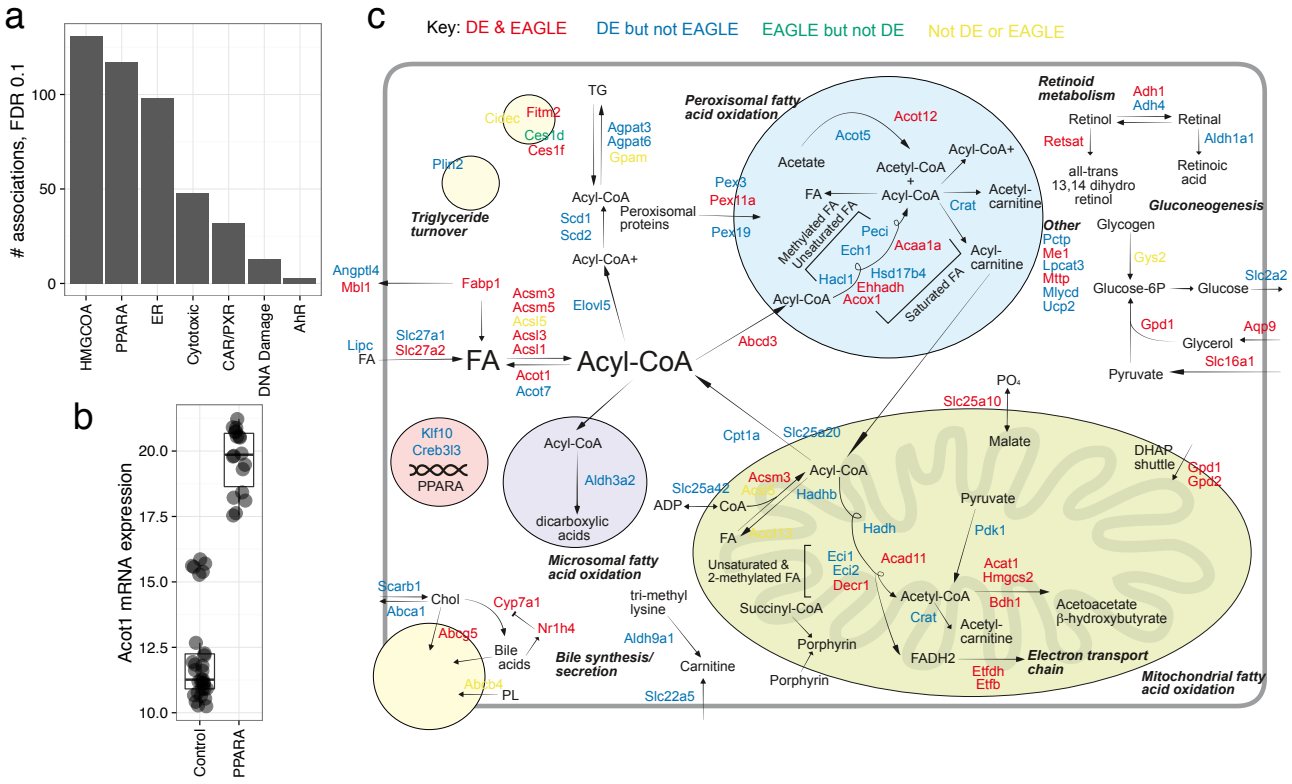
Supplementary Figure 11: Standard interaction QTL testing is sensitive to confounding. We performed a second simulation study using total expression to assess the influence of confounding on standard interaction QTL testing. We use principal components of the gene expression matrix as our potential confounders. In order to simulate data with structured noise matching that in the true data, the effect on the true gene expression values of local cis-SNPs, the environment (time of day) and confounding was assessed using a Bayesian regression model. Gene expression values were then generated from this model, including a varying number of confounders (PCs), and with or without a true GxE effect. The ROC curves here show that as more confounding is included in the simulations, the sensitivity and specificity of standard interaction QTL testing falls dramatically.



Supplementary Figure 12: The two stage analysis, where we first use EAGLE to identify genes whose allelic expression is associated with an environmental factor, and second test for candidate variants, is robust to both the initial FDR threshold used and the cis-window size used in step 2. We show the proportion of nominally significant associations after Bonferroni correction across cis-SNPs.



Supplementary Figure 13: Interaction p-values can be approximated using the t-statistics reported by Fairfax et al. **a.** The sampling distribution for t_Δ. To approximate p-values for a significant interaction term in two conditions from t-statistics we use the test statistic $t_\Delta = t_1 - t_0$. Under the null distribution that the effect size and noise variance in each condition is equal, we can analytically approximate the sampling variance of t_Δ. To confirm the validity of this analytical calculation we compare to empirical simulations, and find excellent agreement even for large shared effect sizes. **b.** Testing t_Δ is well calibrated. We simulated a SNP with a MAF of 0.2 in $N = 228$ individuals, noise s.d. of 0.3 and a shared effect size of 5 in two conditions. We calculated p-values according to our procedure for the Fairfax et al. data. The qqplot here confirms that the resulting p-values are well calibrated.



Supplementary Figure 14: EAGLE detects allele-specific responses to treatment of rat livers with various toxicants. **a.** EAGLE detects 442 associations (10% FDR) between allelic imbalance and treatment, across seven toxicant classes, despite moderate sample sizes. **c**. Shown here are EAGLE-testable orthologous rat genes to mouse genes known to respond to PPARα either from knock-out experiments or agonist treatment, positioned according to their functional roles. EAGLE associations and differential expression (using Spearman correlation) were called at 10% FDR.