Online methods

1 Discovery data

The Brisbane Systems Genetics Study (BSGS) comprises 846 individuals of European descent from 274 independent families.[1] DNA samples from each individual were genotyped on the Illumina 610-Quad Beadchip by the Scientific Services Division at deCODE Genetics Iceland. Full details of genotyping procedures are given in Medland et al.[2] Standard quality control (QC) filters were applied and the remaining 528,509 autosomal SNPs were carried forward for further analysis.

Gene expression profiles were generated from whole blood collected with PAXgene TM tubes (QIAGEN, Valencia, CA) using Illumina HT12-v4.0 bead arrays. The Illumina HT-12 v4.0 chip contains 47,323 probes, although some probes are not assigned to RefSeq genes. We removed any probes that did not match the following criteria: contained a SNP within the probe sequence with MAF > 0.05 within 1000 genomes data; did not map to a listed RefSeq gene; were not significantly expressed (based on a detection p-value < 0.05) in at least 90% of samples. After this stringent QC 7339 probes remained for 2D-eQTL mapping.

2 Normalisation

Gene expression profiles were normalised and adjusted for batch and polygenic effects. Profiles were first adjusted for raw background expression in each sample. Expression levels were then adjusted using quantile and log2 transformation to standardise distributions between samples. Batch and polygenic effects were adjusted using the linear model

$$y = \mu + \beta_1 c + \beta_2 p + \beta_3 s + \beta_4 a + g + e \tag{1}$$

where μ is the population mean expression levels, c, p, s and a are vectors of chip, chip position, sex and generation respectively, fitted as fixed effects; and g is a random additive polygenic effect with a variance covariance matrix

$$G_{ijk} = \begin{cases} \sigma_a^2 + \sigma_e^2 & j = k \\ 2\phi_{ijk}\sigma_a^2 & j \neq k \end{cases} \tag{2}$$

The parameters σ_a^2 and σ_e^2 are variance components for additive background genetic and environmental effects respectively. Here, we are using family based pedigree information rather than SNP based IDB to account for relationships between individuals and so ϕ_{ijk} is the kinship coefficient between individuals j and k. The residual, e, from equation 1 is assumed to follow a multivariate normal distribution with a mean of zero. Residuals were normalised by rank transformation and used as the adjusted phenotype for the pairwise epistasis scan. The GenABEL package for R was used to perform the normalisation.[3]

3   Exhaustive 2D-eQTL analysis

We used epiGPU[4] software to perform an exhaustive scan for pairwise interactions, such that each SNP is tested against all other SNPs for statistical association with the expression values for each of the 7339 probes. For each SNP pair there are 9 possible genotype classes. We treat each genotype class as a fixed effect and fit an 8 *d.f.* *F*-test to test the following hypotheses:

$$H_0 : \sum_{i=1}^{3} \sum_{j=1}^{3} (\bar{x}_{ij} - \mu)^2 = 0; \tag{3}$$

$$H_1 : \sum_{i=1}^{3} \sum_{j=1}^{3} (\bar{x}_{ij} - \mu)^2 > 0; \tag{4}$$

where μ is the mean expression level and x_{ij} is the pairwise genotype class mean for genotype i at SNP 1 and genotype j at SNP 2. This type of test does not parameterize for specific types of epistasis, rather it tests for the joint genetic effects at two loci. This has been demonstrated to be statistically more efficient when searching for a wide range of epistatic patterns, although will also include any marginal effects of SNPs which must be dealt with post-hoc.[5]

The complete exhaustive scan for 7339 probes comprises 1.03×10^{15} *F*-tests. We used permutation analysis to estimate an appropriate significance threshold for the study. To do this we performed a further 1600 exhaustive 2D scans on permuted phenotypes to generate a null distribution of the extreme *p*-values expected to be obtained from this number of multiple tests given the correlation structure between the SNPs. We took the most extreme *p*-value from each of the 1600 scans and set the 5% FWER to be the 95% most extreme of these *p*-values, $T_* = 2.13 \times 10^{-12}$. The effective number of tests in one 2D scan being performed is therefore $N_* = 0.05/T_* \approx 2.33 \times 10^{10}$. To correct for the testing of multiple traits we established an experiment wide threshold of $T = 0.05/(N_* \times 7339) = 2.91 \times 10^{-16}$. This is likely to be conservative as it assumes independence between probes.

Following filtering on this threshold only SNP pairs with at least 5 data points in all 9 genotype classes were kept. We then calculated the LD between interacting SNPs and removed any pairs with $r^2 > 0.1$. If multiple SNP pairs were present on the same chromosomes for a particular expression trait then

only the sentinal SNP pair was retained, *i.e.* if a probe had multiple SNP pairs that were on chromosomes one and two then only the SNP pair with the most significant p-value was retained. At this stage 6404 filtered SNP pairs remained. We then performed a second filtering screen that was identical to the first but an additional step was included where any SNPs that had previously been shown to have a significant additive or dominant effect ($p < 1.29 \times 10^{-11}$) was removed, creating a second set of 4751 filtered SNP pairs.

To ensure that interacting SNPs were driven by epistasis and not marginal effects we performed a nested ANOVA on each pair in the filtered set to test if the interaction terms were significant. We did this by contrasting the full genetic model (8 *d.f.*) against the reduced marginal effects model which included the additive and dominance terms at both SNPs (4 *d.f.*). Thus, a 4 *d.f.* F-test was performed on the residual genetic variation, representing the contribution of epistatic variance. Significance of epistasis was determined using a Bonferroni threshold of $0.05/(6404 + 4751) = 4.48 \times 10^{-6}$. This resulted in 432 and 117 SNP pairs with significant interaction terms from filters 1 and 2, respectively.

All variance decompositions were calculated using the NOIA method.[6]

4 Replication

We have attempted replication of the 529 significant interactions from the discovery set using three independent cohorts; Fehrmann, EGCUT, and CHDWB. It was required that LD $r^2 < 0.1$ between interacting SNPs, and all nine genotype classes had at least 5 individuals present in order to proceed with statistical testing for replication in both datasets. We also excluded any putative SNPs that had discordant allele frequencies in any of the datasets. Details of the cohorts are as follows.

Fehrmann: $n = 1240$ The Fehrmann dataset[7] consists of whole peripheral blood samples of 1240 unrelated individuals from the United Kingdom and the Netherlands. Some of these individuals are patients, while others are healthy controls. Individuals were genotyped using the Illumina 610 Quad platform. RNA levels were quantified using the HT12v3 platform.

EGCUT: $n = 891$ The Estonian Genome Center of the University of Tartu (EGCUT) study[8] consists of whole blood samples of 891 unrelated individuals from Estonia. They were genotyped using the Illumina 610 Quad platform. RNA levels were quantified using the HT12v3 platform.

CDHWB: $n = 139$ The Center for Health Discovery and Well Being (CD-HWB) Study[9] is a population based cohort consisting of 139 individuals of European descent collected in Atlanta USA. Gene expression profiles were generated with Illumina HT-12 V3.0 arrays from whole blood collected from Tempus tubes that preserve RNA. Whole genome genotypes were measured using Illumina OmniQuad arrays. Due to the small sample size, most SNP pairs did

not pass filtering in this dataset and so we have excluded it from the rest of the analysis.

4.1 Meta Analysis

The 4 *d.f.* interaction *p*-values for each independent replication dataset were calculated using the same statistical test as was performed in the discovery dataset. We then took the interaction *p*-values from EGCUT and Fehrmann and calculated a joint *p*-value as $-2\log p_1 - 2\log p_2 \sim \chi^2_{4d.f.}$.

5 Additive vs epistatic variance comparison

To compare the relative contribution of genetic variance between additive and epistatic variances one would ideally take all the possible additive effects

6 Functional enrichment analysis

6.1 H3K4me3

6.2 Chromosome interactions

6.3 Transcription factor enrichment

6.4 SNP colocalisation with genic features

7 References

References

[1] Powell, J. E. *et al.* The Brisbane Systems Genetics Study: genetical genomics meets complex trait genetics. *PloS one* **7**, e35430 (2012).

[2] Medland, S. E. *et al.* Common variants in the trichohyalin gene are associated with straight hair in Europeans. *American journal of human genetics* **85**, 750–5 (2009).

[3] Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics (Oxford, England)* **23**, 1294–6 (2007).

[4] Hemani, G., Theocharidis, A., Wei, W. & Haley, C. EpiGPU: exhaustive pairwise epistasis scans parallelized on consumer level graphics cards. *Bioinformatics (Oxford, England)* **27**, 1462–5 (2011).

[5] Hemani, G., Knott, S. & Haley, C. An Evolutionary Perspective on Epistasis and the Missing Heritability. *PLoS Genetics* **9**, e1003295 (2013).

[6] Alvarez-Castro, J., Le Rouzic, A., Carlborg, O., Álvarez Castro, J. M. & Carlborg, O. How to perform meaningful estimates of genetic effects. *PLoS Genetics* **4**, e1000062 (2008).

[7] Fehrmann, R. S. N. *et al.* Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS genetics* **7**, e1002197 (2011).

[8] Metspalu, A. The Estonian Genome Project. *Drug Development Research* **62**, 97–101 (2004).

[9] Preininger, M. *et al.* Blood-informative transcripts define nine common axes of peripheral blood gene expression. *PLoS genetics* **9**, e1003362 (2013).