Genetic interactions affecting human gene expression identified by variance association mapping

Andrew A Brown, Alfonso Buil, Ana Viñuela, Tuuli Lappalainen, Hou-Feng Zheng, John B Richards, Kerrin S Small, Timothy D Spector, Emmanouil T Dermitzakis, Richard Durbin

DOI: http://dx.doi.org/10.7554/eLife.01381

Cite as: eLife 2014;10.7554/eLife.01381

Received: 21 August 2013
Accepted: 13 March 2014
Published: 25 April 2014

Genetic interactions affecting human gene expression identified by variance association mapping

Andrew Anand Brown[1,2], Alfonso Buil[3,4,5], Ana Viñuela[6], Tuuli Lappalainen[3,4,5], Hou-Feng Zheng[7], John B. Richards[6,7], Kerrin S. Small[6], Timothy D. Spector[6], Emmanouil T. Dermitzakis[3,4,5], Richard Durbin[1*]

1. Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK
2. NORMENT, KG Jebsen Centre for Psychosis Research, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway
3. Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland.
4. Institute of Genetics and Genomics in Geneva, University of Geneva, Geneva, Switzerland.
5. Swiss Institute of Bioinformatics, Switzerland
6 Department of Twin Research and Genetic Epidemiology, King's College London, UK
7. Department of Medicine, Human Genetics, Epidemiology and Biostatistics McGill University, Canada.

*Corresponding author: Richard Durbin, rd@sanger.ac.uk

Abstract

Non-additive interaction between genetic variants, or epistasis, is a possible explanation for the gap between heritability of complex traits and the variation explained by identified genetic loci. Interactions give rise to genotype dependent variance, and therefore the identification of variance quantitative trait loci can be an intermediate step to discover both epistasis and gene by environment effects (GxE). Using RNA-sequence data from lymphoblastoid cell lines (LCLs) from the TwinsUK cohort, we identify a candidate set of 508 variance associated SNPs. Exploiting the twin design we show that GxE plays a role in ~70% of these associations. Further investigation of these loci reveals 57 epistatic interactions that replicated in a smaller dataset, explaining on average 4.3% of phenotypic variance. In 24 cases, more variance is explained by the interaction than their additive contributions. Using molecular phenotypes in this way may provide a route to uncovering genetic interactions underlying more complex traits.

Introduction

The discrepancy between the contribution of known genetic factors to variation of a trait and the estimated total contribution of all genetic variants has become known as "missing heritability" (Manolio et al., 2009). Some of the explanations for this discrepancy are: many common variants with small effects; many rare variants with larger effects; and interactions between genetic variants (epistasis) or between variants and environment (GxE). Here, we focus on the discovery and characterization of epistasis, by which we mean that the effect of a genetic variant on a trait depends on the genotype at one or more other locations in the genome. Statistically we define this as a joint effect of two loci on a trait, significant beyond the sum of additive effects.

On long time frames, epistasis plays an important role in evolution (Breen et al., 2012), and has been used to explain the persistence of deleterious mutations under selection (Hemani et al., 2013). Epistasis has frequently been seen in crosses between model organism strains. Huang et al. (2012) looked at mapping variants associated with three traits in two distinct *Drosophila* populations and found very little concordance between the results. They postulated that this could be because the effect of genetic variants was dependent on the genetic background, and found frequent evidence of genetic interactions between two or more variants and the originally associated SNPs. Annotating these interacting SNPs to genes revealed common networks of highly connected genes across both populations. In a study of sources of variation in yeast crosses, Bloom et al. (2013) carried out a scan for epistasis which discovered 78 pairs of loci where the effect of one was dependent on the genotype of the other, affecting 24 traits. In most cases these interactions explained little of the genetic variation in trait, the median was 3%, but in one case 14% of this variance was explained. Significant interactions between variants have also been seen to affect rice yields (Huang et al., 2014) and metabolic traits in yeast (Wentzell et al., 2007). An extended recent review of study designs appropriate to detect epistasis in model organisms, and the evidence thus far collected, can be found in Mackay (2014).

However, epistasis has proved harder to identify in human genome-wide association studies. In particular, with classical complex traits there has not been evidence of epistasis observed on the scale seen in model organisms. This may be in part because of the large number of possible interactions to test in the human genome, and possibly because the genetic architecture is different in a homogeneous outbred population from that of a cross between inbred lines.

Paré et al. (2010) have described how an interaction, either genetic or environmental, can induce genotype dependent variance in phenotypes. This effect can be observed without directly modeling the interacting factor. They suggested that SNPs which showed such effects on variance could be prioritized in the search for interactions. We see an example of why this could be true in Figure 1a: carriers of C allele of SNP rs230273 show reduced expression when also carriers of the G allele of SNP rs3131691. For carriers of this G allele, this induces a bimodality in expression which appears as a large variance in expression. For those with AA genotype at rs3131691, expression appears independent of rs230273 genotype; in the absence of the induced bimodality, the variance within this group is much reduced. The interactions causing genotype dependent variance could be with another genetic variant (epistasis, as in our example and the focus of this paper) or an environmental factor.

We therefore adopt the following two step strategy for uncovering epistasis affecting gene expression. We search for: 1) SNPs affecting the variance of expression (v-eQTL) within the 2Mbp region around the transcription start site (TSS) of the gene, and then 2) SNPs in epistasis with these v-eQTL. Previous work that looked for variance QTL for height and BMI in ~150,000 samples identified one replicated locus (Yang et al., 2012). Wang et al. (2014) also looked at v-eQTL in gene expression in the same cohort as presented here, where expression was quantified using microarrays rather than sequence based technology (Grundberg et al., 2012). They concluded that v-eQTL can often be induced by partial linkage disequilibrium with eQTL. They also discovered differences in expression between monozygotic twins which were dependent on genotype of the twin pair, such differences cannot be induced by these partial linkages and thus point to a gene-environment interaction. The haplotype effect explanation for v-eQTL, combined with a literature which has concluded in many cases epistasis does not contribute to variation in complex traits (Hill et al., 2008), led them to conclude epistasis is not a cause of v-eQTL. However, they do not search for examples of epistasis; we do so in this paper, explicitly ruling out haplotype effects. We note that microarray data are also less suitable than RNA-seq for the purpose of detecting v-eQTL, because saturation of signal limits discrimination at extremes (Wang et al., 2009). In neither Yang et al. (2012) nor Wang et al. (2014) were variance QTL directly used to identify epistatic or GxE interactions.

Two papers have also looked at producing a phenotype related to variance, in both cases using the coefficient of variance (CV) within inbred lines, to map variants which control the stochastic influence in phenotypic variation (Ansel et al. (2008) and Jimenez-Gomez et al. (2011)). In single cell work, and animal models where the environment can be strictly controlled, variance within inbred lines could be seen as stochastic. But we focus our work on where genotype dependent variance is the consequence of a hidden factor, in our case the presence of an interaction between genetic variants, rather than examples where the observations are due to differences in random processes.

There are two other mechanisms by which genotype dependent variance can be induced. Firstly, as Sun et al. (2013) have described, standard eQTL working on mean gene expression levels can be mistaken for having variance effects in the presence of a mean – variance relationship. With RNA-seq data, the relationship between mean and variance is clear; as RNA-seq reads are sampled from a Poisson distribution, a square root transformation breaks this link. Secondly, as discussed by the Wang et al. (2014) paper described above, haplotype effects can appear as v-eQTL. For example, the situation where a recent strong eQTL co-segregates with a more common SNP (i.e. the SNP is in low R^2 with the eQTL, but high D') could be observed as variance effects of a single SNP. This could also by mistaken for epistasis between two variants which jointly tag the eQTL. We control for this possibility by explicitly considering all possible explanatory eQTL in the full sequence data available for our replication sample.

Results

We searched for v-eQTL in a dataset of 765 LCL samples from female Caucasian adult twins in the TwinsUK cohort, including 134 monozygotic (MZ) twin pairs and 192 dizygotic (DZ) pairs. The same samples from this cohort have previously been used for eQTL analysis, with expression quantified using microarrays (Grundberg et al., 2012). The level of expression of 13,660 genes was determined using whole transcriptome sequencing (RNA-seq). Using a non-parametric association test between SNPs within a cis window of ±1Mbp around the TSS and the square of the residuals (see Methods for details), we identified 497 SNPs as peak v-eQTL for 508 genes (False Discovery Rate (FDR) < 0.05, Figure 1-figure supplement 1 and Supplementary Table ST1), 23 reaching Bonferroni significance (nominal *p*-value$<8.9\text{x}10^{-10}$). Many of the FDR defined v-eQTL cluster close to the TSS (9.3% are within 10kb) but they are found at all positions in the window (Figure 1b). One hundred and eighty one v-eQTL are also significant eQTL at a false discovery rate (FDR) of 0.05 (Figure 1-figure supplement 2).

To search for epistasis, we scanned the cis windows for a second variant statistically interacting with each of the peak v-eQTL. A forward stepwise analysis identified independent examples of epistasis, not induced by linkage disequilibrium; a statistical test was applied to remove signals related to dominance (see Methods for details). This identified 256 independent SNPs in apparent epistasis with the peak v-eQTL for 173 genes (Bonferroni, *p*-value < 1.98×10^{-8}; Supplementary Table, ST2). To call these signals as genuine genetic interactions we required two further criteria: i) significant replication in an independent dataset, and ii) that the interaction could not be explained by the effect of a third, possibly rare, variant effecting expression as discussed above.

We replicated our scan for v-eQTL and epistatic interactions in 462 samples with LCL RNA-seq data from 1000 Genomes samples collected by the GEUVADIS consortium (Lappalainen et al., 2013). Table 1 reports the results of replication for v-eQTL and epistasis using both FDR and Bonferroni correction for threshold determination. For the 23 v-eQTL that are significant using the Bonferroni threshold, 16 are significant in the GEUVADIS cohort (FDR<0.05), 15 with same direction of effect. Of the 508 v-eQTL, 28 replicated with an FDR<0.05, 26 with same direction of effect. The ten most significant v-eQTL in the GEUVADIS cohort, with matching direction of effect across the two cohorts, are shown in Figure 1, figure-supplements 3-12.

Of the 256 epistasis associations, information on both the SNP and the gene was available for 246 in the GEUVADIS data. We found that 137 replicated with FDR<0.05, 131 of which had the same direction of effect (Supplementary Table ST2). P-value enrichment analysis (Storey, 2002) indicated that there was replication evidence for 71% of the 246. Moreover, we observed a correlation of 0.58 between the effect sizes of the interactions in both datasets (*p*-value=5.9×10^{-24}), with 202 of the 246 interactions sharing the same direction of effect (*p*-value=2.2×10^{-25}) (Figure 2-figure supplements 1-2).

As discussed in the introduction, it is possible that an observed statistical interaction between two SNPs can be caused by a single true eQTL in linkage disequilibrium with them. For example, a particular combination of alleles across the pair of SNPs could tag a rare causative eQTL. To rule out this possibility, we took advantage of the full sequence for the GEUVADIS replication samples obtained by the 1000 Genomes Project (The 1000 Genomes Project Consortium, 2012). For the 131 replicated examples of epistasis we identified all eQTL for the relevant genes amongst all sequenced cis SNPs or indels (a forward stepwise scan identified all eQTL significant with $p<10^{-5}$, see methods for details). The aim was for good characterisation of eQTL down to low frequency variants, though this is complicated by power and poorer imputation accuracy at such frequencies. We then tested whether the epistatic interaction was still significant in models incorporating each eQTL individually at the same threshold as previously applied. Fifty seven epistasis signals remain significant. Figure 2a shows the effect of the epistasis SNP broken down by genotype group on expression of *TRIT1*, Table 2 and Figure 2, figure supplements 3-12 report the 10 most significant examples of epistasis in the GEUVADIS cohort, a full list is in Supplementary Table ST2. For all plotted interactions, the direction of effect was consistent within v-eQTL genotype groups across cohorts. In at least two instances we see sign epistasis, the effect of one SNP reverses direction conditional on the other SNP (Figure 2-figure supplements 7 and 9).

We estimated the proportion of variance explained by the interaction in the GEUVADIS cohort to avoid over-estimating effects because of winner's curse. As a result, we were able to determine that up to 16% of the variance in gene expression was explained by considering the interaction between the variants, with an average additional variance explained of 4.3% (Table 2, Supplementary Table 2, Figure 3). For the eight genes for which we replicated independent interactions with the v-eQTL, we found that in total up to 10.4% of the variance was explained by these multiple interactions, with an average of 5.1%. For 24 out of 57 the replicated examples of epistasis, the interaction explains more variance than the additive effects of the SNPs. We show as an example the gene *TRIT1* (Figure 2). The v-eQTL (rs3131691) for *TRIT1* lies on the boundary of an ENCODE defined LCL weak enhancer (Dunham et al., 2012, Rosenbloom et al., 2013) upstream of the gene, while the SNP in epistasis (rs230273) lies on the boundary of a downstream LCL enhancer region (Figure 2b). The v-eQTL is also 28bp upstream of a strong eQTL signal (rs34387655). This eQTL has minor allele frequency (MAF) 0.08, and is in high D' with the v-eQTL (MAF=0.30), suggesting that the eQTL could be a recent mutation co-segregating with one allele of the v-eQTL. But this eQTL cannot explain the observed interaction, which was still significant when analyzing only major allele homozygotes for the eQTL (*p*-value=0.0095). Therefore, we conclude that two causal loci act on the weak enhancer in two different ways; rs34387655 has a direct effect on the enhancer while rs3131691 acts in conjunction with the epistasis variant rs230273 (or variants in linkage disequilibrium with these SNPs act in these ways).

The discussion up to this point concerns SNPs in cis with the expressed gene. Looking for examples of trans SNPs (>5Mbp from the TSS) in epistasis with the v-eQTL yielded no hits that replicated in the GEUVADIS cohort. However, using the twin design we were able to address the contribution of long range epistasis by a heritability analysis. Assuming no recombination in the cis region, the proportion of the cis window that dizygotic twins (DZ) inherited identically by descent is either 0, 0.5 or 1 and this allows us to perform a linkage analysis to estimate the proportion of variance explained by variants in the cis region, the trans region (5Mbp away from the TSS) and interactions between the two. We had information about the IBD sharing around 273 of the 508 v-eQTL genes. In 15 of these, interactions between the cis and trans regions explain more than 10% of the variance in expression. For all of these there is greater evidence of cis-trans epistasis affecting expression than an influence of common environment, and for 9 of the 15 the interaction effect was more than the estimated combined direct genetic contribution of both cis and trans variants (Supplementary Table ST3).

The presence of v-eQTL can be induced by gene-environment interactions, as well as epistasis or haplotype effects. Because our data come from a twin cohort, which includes monozygotic (MZ) twin pairs, we have another measure of variability within the dataset: the discordance in expression between MZ twins. Genotype dependent differences in expression within MZ pairs cannot be induced by epistasis or haplotype effects, as both twins share the same genetic background. Therefore, evidence that v-eQTL are also discordant eQTL (d-eQTL) would suggest that v-eQTL could also have a GxE explanation, including possibly interactions between the genome and the epigenome (Martin et al., 1983, Reynolds et al., 2007) (Figure 4a). Using our MZ data, we have tested our 508 v-eQTL for evidence that they are also d-eQTL; using the methods from Storey (2002) we estimate that 70% of the v-eQTL act in this manner. This suggests that GxE interactions are common amongst these variants (Methods, Figure 4b and Supplementary Table ST1). In total, 176 of the 508 v-eQTL show significant effects on discordance (FDR<0.05). Of these 176, we estimate the proportion that are also eQTL as 40.3%, less than the proportion of all v-eQTL which act as eQTL.

By looking at variance between individuals and discordance between monozygotic twins, we mirror an approach which looked at robustness of phenotypes to genetic and environmental influences (Fraser and Schadt, 2010). In this study of gene expression traits, differences between inbred mouse strains were called “genetic robustness QTL” (GR-QTL). These correspond to our definition of v-eQTL, and the paper discusses how they can be induced by epistatic interactions. The paper also looks at QTL for within strain variance, analogous to our d-eQTL and referred to as “environmental robustness QTL” (ER-QTL), and describe them as induced by gene-environment interactions. They reported finding both GR-QTL and ER-QTL in mice, *Arabidopsis* and *S. cerevisiae.*

Discussion

The importance of non-additive variation to explaining missing heritability has been much debated (Hill et al., 2008, Zuk et al., 2012). Here, we were able to report specific examples of interactions explaining noticeable fractions of variation in human gene expression, with in many cases the interaction contributing more than the marginal effects to overall variance. Estimating variance components from pedigrees and twin model studies has concentrated on additive variance, to estimate the narrow sense heritability. The assumption has been that resemblance between related individuals is determined chiefly by additive variation (Falconer and Mackay, 1996). An overview of analyses of many phenotypes in many organisms concluded that there was little evidence for non-additive variation playing a large role in phenotypic variation (Hill et al., 2008). Indeed, the authors provided a theoretical argument that the total contribution of all interacting loci to variance is well approximated by their additive contribution, when the allele frequencies are as predicted by the neutral model. The analysis presented here is powered chiefly to discover common interacting variants, however the result on the neutral model implies there may be many more examples of epistasis which are not statistically detectable without very large samples.

Specifically in gene expression, progress has recently been made to move beyond a solely additive view of variation. Becker et al. (2012) produced evidence for the existence of cis-trans epistasis, though they do not report individual examples which were significant when controlling for all tests and did not consider the contribution of these interactions to phenotypic variation. Further work from Powell et al. (2013) looked to dissect the phenotypes into dominant and additive components. As with our dissection of cis-trans epistasis, additive genetic variation was most consistently observed, though 960 probes had a dominant component to variation; for a subset of these a non-additive eQTL was proposed. All in all, these global results together with the replicated epistatic interactions presented here suggest a moderate influence of non-additive genetic effects on gene transcription variation.

The majority of the interactions are close to each other and to the TSS (Figure 2, figure supplement 13), consistent with a direct molecular interaction. However, despite physical proximity they are, because of the statistical discovery strategy, in low linkage disequilibrium. There has been discussion in the literature about how interactions between variants affecting fitness can change the linkage disequilibrium structure of a region, by bringing variants which alter the local recombination rate under indirect selection (Otto and Feldman, 1997). In the case of positive epistasis, where the combined effect on fitness of the deleterious alleles is mitigated by their joint contribution, selection would favour a decrease in the recombination rate between the loci. This was seen in Lappalainen et al. (2011): non-synonymous, possibly deleterious, coding mutations together with an eQTL which adjusts expression would be an example of positive epistasis. In support of the theoretical result, such variants were frequently observed in high linkage disequilibrium in their results. In contrast, the approach we take here requires linkage disequilibrium to have broken down between variants in order to distinguish an interaction between two variants from a dominant effect of a single locus. As a consequence, we are powered more to detect epistasis which amplifies the effect of deleterious mutations, rather than positive epistasis as described by Lappalainen et al. (2011). Therefore, examples of epistasis of the type they describe would be missed by our methodology (indeed, the five non-synonymous SNPs we discover to be involved in interactions in the TwinsUK dataset are all predicted by PolyPhen score to be benign with the exception of a one (rs150369207) which is classed as possibly damaging for only one out of nine coding transcripts).

A recent paper has also looked for evidence of epistasis affecting transcription in humans (Hemani et al., 2014), using array expression from whole blood and searching the entire space of all possible pairwise interactions. They discover 501 interactions, affecting expression of 238 genes in 846 samples, and replicate 30 examples in an independent dataset at Bonferroni significance level. The interactions discovered are chiefly cis-trans; of the 501 there are 26 cis-cis interactions and 13 trans-trans. The apparent lower replication rate compared to our study may reflect the greater success that has been seen replicating cis effects than trans effects for standard eQTL (Grundberg et al., 2012). Grundberg et al. (2012) also reported that LCLs (the tissue used in our study) showed stronger genetic effects compared to environmental contribution than seen in primary tissues. Finally, RNA-seq has been shown as a more reliable phenotype than array based measures (Marioni et al., 2008). We believe all these factors contribute to our success rate in replicating epistatic interactions.

In conclusion, we report 26 replicated variance eQTL and 57 replicated cis epistatic interactions, which explain up to 16% of the variance of our phenotypes. In almost a half of cases, more variance is explained by the interaction than by single additive effects. Furthermore, we have also shown substantial evidence for gene by environment interactions. We have shown that a proportion of variation of molecular phenotypes can be ascribed to genetic interactions, and that v-eQTL are a valid way of discovering them. Densely phenotyped cohorts are now commonly collecting such molecular data, and therefore there is considerable scope to look both for more of this type of interactions, and for the particular environments involved in GxE. The ability to find genetic interactions affecting molecular phenotypes also suggests a hypothesis driven path by which genetic interactions underlying more complex traits may be identified.

Materials and methods

Genotying and imputation: Samples were genotyped on a combination of the HumanHap300, HumanHap610Q, 1M□Duo and 1.2MDuo 1M Illumnia arrays. Samples were pre-phased using IMPUTE2 (Howie et al., 2009) with no reference panel, then imputed into the 1000 Genomes Phase 1 reference panel (interim, data freeze, 10 November 2010, The 1000 Genomes Project Consortium (2012)). Post imputation, SNPs were removed if MAF < 0.01 or IMPUTE info value < 0.8.

RNA processing: Samples were prepared for sequencing with the Illumina TruSeq sample preparation kit (Illumina, San Diego, CA) according to manufacturer's instructions and were sequenced on a HiSeq2000 machine. Afterwards, the 49-bp sequenced paired-end reads were mapped to the GRCh37 reference genome (The International Human Genome Sequencing Consortium, 2001) with BWA v0.5.9 (Li and Durbin, 2009). We use genes defined as protein coding in the GENCODE 10 annotation (Harrow et al., 2012), removing genes with more than 10% zero read count. RPKM values were root mean transformed. PEER software (Parts et al., 2011) was used to remove 50 latent factors; age and body mass index were included when factors were constructed, to prevent removal of important environmental factors. Data were then quantile normalised.

v-eQTL: GRAMMAR (Aulchenko et al., 2007) was used to remove correlations between related individuals. Expression of each gene was tested against every SNP within 1Mbp of the TSS. First, any eQTL effects were removed by regressing expression on the posterior probability of being a heterozygote and the posterior probability of being a minor allele homozygote. The residuals were squared, giving a measure of distance from the mean expression of that genotype class for all individuals. A Spearman rank correlation test between this "distance" and genotype dosage was used to assess evidence of variance effects. A set of 5 permutations, consistent across all tests to consider linkage disequilibrium structure between SNPs, was applied to the distance residuals and the spearman correlation test was applied as before to estimate the distribution of the test statistic under the complete null hypothesis of no variance effects. An FDR was calculated as the proportion of permuted statistics more significant, divided by 5. This two stage procedure where relatedness was regressed out separately from v-eQTL mapping was adopted to make the full scan for v-eQTL computationally feasible.

Epistasis: The R package lme4 (Bolker, 2013) was used to fit linear mixed models using maximum likelihood to model expression as a function of genetic interactions. The models, with a full description of how the twin structure is captured, are presented in the section "Equations". A forward stepwise scheme, as used in Lappalainen et al. (2013) to map standard eQTL, was used to discover independent examples of epistasis. Assuming the K-1 significant examples of epistasis had been discovered, a complete scan of every SNP in the cis window tested for evidence of epistasis with the v-eQTL (using a likelihood ratio test of equation (2) nested into equation (1), testing the hypothesis c_K=0), conditioned on all previously discovered interactions. If the most significant SNP was Bonferroni significant ($p<1.98\times10^{-8}$), the SNP was added to the list and the process continued, otherwise the list was considered complete. This revealed 275 examples of epistasis, affecting expression of 178 genes. To exclude the possibility that significant interactions could be explained by a non-additive genetic effect of the original v-eQTL appearing as epistasis between the v-eQTL and another variant in tight linkage disequilibrium, a further conditional analysis tested the epistasis term conditional on the model it was discovered in and a non-additive effect of the v-eQTL (testing nested models, equation (3) and equation (4) for $c_K=0$). SNPs which were not Bonferroni significant at the same threshold ($p<1.98\times10^{-8}$) were removed, leaving 256 epistatic interactions affecting 173 genes. Proportion of variance for linear mixed models was calculated as described in Nakagawa and Schielzeth (2012). Scripts to analyse the data are provided in Supplementary material.

Equations:

Denoting individual i, expression by yi, dosage of v-eQTL by S_{iv}, dosage of the kth discovered epistatic SNPs by S_{ik}, probability that the v-eQTL is a heterozygote by S_{iv}^{het}, and the probability that the v-eQTL is a minor allele homozygote by S_{iv}^{hom}, we have modelled expression in the following ways:

$$y_i = \mu + aS_{iv} + \sum_{k=1}^{K-1}(b_kS_{ik} + c_kS_{iv}S_{ik}) + b_KS_{iK} \qquad + \beta_i + \gamma_i + \epsilon_i \quad (1)$$

$$y_i = \mu + aS_{iv} + \sum_{k=1}^{K-1}(b_kS_{ik} + c_kS_{iv}S_{ik}) + b_KS_{iK} + c_KS_{iv}S_{iK} \qquad + \beta_i + \gamma_i + \epsilon_i \quad (2)$$

$$y_i = \mu + a^{het}S_{iv}^{het} + a^{hom}S_{iv}^{hom} + \sum_{k=1}^{K-1}(b_kS_{ik} + c_kS_{iv}S_{ik}) + b_KS_{iK} \qquad + \beta_i + \gamma_i + \epsilon_i \quad (3)$$

$$y_i = \mu + a^{het}S_{iv}^{het} + a^{hom}S_{iv}^{hom} + \sum_{k=1}^{K-1}(b_kS_{ik} + c_kS_{iv}S_{ik}) + b_KS_{iK} + c_KS_{iv}S_{iK} \qquad + \beta_i + \gamma_i + \epsilon_i \quad (4)$$

where

$$\begin{aligned}\beta_i &\sim N(0, \sigma_{FAM}^2)\\ \gamma_i &\sim N(0, \sigma_{MZ}^2)\\ \epsilon_i &\sim N(0, \sigma^2)\end{aligned}$$

To correctly model the twin structure we require that $\beta_i = \beta_j$ when i and j are twins, and $\gamma_i = \gamma_j$ when i and j are MZ twins (capturing the increased genetic correlation of MZ twins).

Heritability: A variance components model was fitted in the program solar (Almasy and Blangero, 1998) where the covariance matrix for the trait is written:

$$\mathbf{\Omega} = \mathbf{\Pi}_{cis}\sigma_{cis}^2 + \mathbf{\Pi}_{trans}\sigma_{trans}^2 + \mathbf{\Pi}_{cis-trans}\sigma_{cis-trans}^2 + \mathbf{I}\sigma_e^2$$

$\mathbf{\Pi}_{cis}$ and $\mathbf{\Pi}_{trans}$ are the proportion of cis and trans alleles that twins share inherited identically by descent and $\mathbf{\Pi}_{cis\text{-}trans}$ is the Hadamard product of these matrices. Parameters were estimated by maximum likelihood and proportion of variance explained by cis-trans interactions was estimated as:

$$\frac{\sigma_{cis-trans}^2}{\sigma_{cis}^2 + \sigma_{trans}^2 + \sigma_{cis-trans}^2 + \sigma_e^2}$$

For comparison, the model without cis-trans interactions but with a common environment term was fitted, and the two models compared using likelihood.

Discordant QTL: Maximum expression of the two twins was regressed on minimum expression of the twin pair and genotype of the twin pair to detect whether the relationship between max and min expression was conditional on genotype.

GEUVADIS replication: Raw RPKM values were root transformed, 20 principal component factors were removed and then the data were quantile normalised. Evidence for v-eQTL and epistasis was calculated as before, with indicator variables for study population (CEU, YRI,TSI, GBR, FIN) to control for population effects. Epistasis was assessed for each SNP individually, as LD induced multiple signals and dominance effects had been removed in the TwinsUK sample. To ensure that our results are not caused by heteroskedasticity, we have considered various transformations to remove this issue and found the results to be robust. In particular, of the 131 statistically significant interactions in the GEUVADIS cohort, 126 are also significant when log transformed data is analysed (a typical way of accounting for heteroskedasticity). To eliminate confounding with eQTL variants, an identical forward stepwise cis eQTL scan to that used in Lappalainen et al. (2013) reported all eQTL significant at $p<10^{-5}$ in the GEUVADIS dataset. A t-test for each reported eQTL assessed significance of the interaction conditional on the v-eQTL, epistasis SNP and the eQTL. If the greatest p value, over all possible eQTL, did not meet the FDR cut-off the SNP was removed from the list of interactions. FDR was calculated using the qvalue package (qvalue 1.34.0) in R (R Development Core Team, 2008) using the default settings with the exception that lambda was restricted to lie within the range of the p values to prevent overly lenient correction. The replication dataset together with functions to reproduce the results are provided in Supplementary file.

ENCODE segmentation: Segmentation analysis for LCL cell line GM12878 was downloaded from the UCSC website on 11/6/2013, url: http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeBroadHmm/wgEncodeBroadHmmGm12878HMM.bed.gz.

Acknowledgements

The TwinsUK study was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013). The study also receives support from the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. SNP Genotyping was performed by The Wellcome Trust Sanger Institute and National Eye Institute via NIH/CIDR**.** Some computation was performed at the Vital-IT (http://www.vital-it.ch) Center for high-performance computing of the SIB Swiss Institute of Bioinformatics.

Competing interests

Emmanouil T. Dermitzakis is a reviewing editor for eLife.

References

ALAN DABNEY AND JOHN D. STOREY AND WITH ASSISTANCE FROM GREGORY R. WARNES qvalue: Q-value estimation for false discovery rate control. R package version 1.34.0 ed.

ALMASY, L. & BLANGERO, J. 1998. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet,* 62**,** 1198-211.

ANSEL, J., BOTTIN, H., RODRIGUEZ-BELTRAN, C., DAMON, C., NAGARAJAN, M., FEHRMANN, S., FRANÇOIS, J. & YVERT, G. 2008. Cell-to-cell stochastic variation in gene expression is a complex genetic trait. *PLoS genetics,* 4**,** e1000049.

AULCHENKO, Y. S., DE KONING, D.-J. & HALEY, C. 2007. Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics,* 177**,** 577-585.

BECKER, J., WENDLAND, J. R., HAENISCH, B., NOTHEN, M. M. & SCHUMACHER, J. 2012. A systematic eQTL study of cis-trans epistasis in 210 HapMap individuals. *Eur J Hum Genet,* 20**,** 97-101.

BLOOM, J. S., EHRENREICH, I. M., LOO, W. T., LITE, T. L. & KRUGLYAK, L. 2013. Finding the sources of missing heritability in a yeast cross. *Nature,* 494**,** 234-7.

BOLKER, D. B. A. M. M. A. B. 2013. lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-2 ed.

BREEN, M. S., KEMENA, C., VLASOV, P. K., NOTREDAME, C. & KONDRASHOV, F. A. 2012. Epistasis as the primary factor in molecular evolution. *Nature,* 490**,** 535-538.

DUNHAM, I., BIRNEY, E., LAJOIE, B. R., SANYAL, A., DONG, X., GREVEN, M., LIN, X., WANG, J., WHITFIELD, T. W. & ZHUANG, J. 2012. An integrated encyclopedia of DNA elements in the human genome.

FALCONER, D. & MACKAY, T. 1996. *Introduction to quantitative genetics*, Longman.

FRASER, H. B. & SCHADT, E. E. 2010. The quantitative genetics of phenotypic robustness. *PLoS One,* 5**,** e8635.

GRUNDBERG, E., SMALL, K. S., HEDMAN, A. K., NICA, A. C., BUIL, A., KEILDSON, S., BELL, J. T., YANG, T. P., MEDURI, E., BARRETT, A., NISBETT, J., SEKOWSKA, M., WILK, A., SHIN, S. Y., GLASS, D., TRAVERS, M., MIN, J. L., RING, S., HO, K., THORLEIFSSON, G., KONG, A., THORSTEINDOTTIR, U., AINALI, C., DIMAS, A. S., HASSANALI, N., INGLE, C., KNOWLES, D., KRESTYANINOVA, M., LOWE, C. E., DI MEGLIO, P., MONTGOMERY, S. B., PARTS, L., POTTER, S., SURDULESCU, G., TSAPROUNI, L., TSOKA, S., BATAILLE, V., DURBIN, R., NESTLE, F. O., O'RAHILLY, S., SORANZO, N., LINDGREN, C. M., ZONDERVAN, K. T., AHMADI, K. R., SCHADT, E. E., STEFANSSON, K., SMITH, G. D., MCCARTHY, M. I., DELOUKAS, P., DERMITZAKIS, E. T., SPECTOR, T. D. & MULTIPLE TISSUE HUMAN EXPRESSION RESOURCE, C. 2012. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet,* 44**,** 1084-9.

HARROW, J., FRANKISH, A., GONZALEZ, J. M., TAPANARI, E., DIEKHANS, M., KOKOCINSKI, F., AKEN, B. L., BARRELL, D., ZADISSA, A., SEARLE, S., BARNES, I., BIGNELL, A., BOYCHENKO, V., HUNT, T., KAY, M., MUKHERJEE, G., RAJAN, J., DESPACIO-REYES, G., SAUNDERS, G., STEWARD, C., HARTE, R., LIN, M., HOWALD, C., TANZER, A., DERRIEN, T., CHRAST, J., WALTERS, N., BALASUBRAMANIAN, S., PEI, B., TRESS, M., RODRIGUEZ, J. M., EZKURDIA, I., VAN BAREN, J., BRENT, M., HAUSSLER, D., KELLIS, M., VALENCIA, A., REYMOND, A., GERSTEIN, M., GUIGO, R. & HUBBARD, T. J. 2012. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res,* 22**,** 1760-74.

HEMANI, G., KNOTT, S. & HALEY, C. 2013. An Evolutionary Perspective on Epistasis and the Missing Heritability. *PLoS Genet,* 9**,** e1003295.

HEMANI, G., SHAKHBAZOV, K., WESTRA, H.-J., ESKO, T., HENDERS, A. K., MCRAE, A. F., YANG, J., GIBSON, G., MARTIN, N. G., METSPALU, A., FRANKE, L., MONTGOMERY, G. W., VISSCHER, P. M. & POWELL, J. E. 2014. Detection and replication of epistasis influencing transcription in humans. *Nature*.

HILL, W. G., GODDARD, M. E. & VISSCHER, P. M. 2008. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genetics,* 4**,** e1000008.

HOWIE, B. N., DONNELLY, P. & MARCHINI, J. 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS genetics,* 5**,** e1000529.

HUANG, A., XU, S. & CAI, X. 2014. Whole-genome quantitative trait locus mapping reveals major role of epistasis on yield of rice. *PLoS One,* 9**,** e87330.

HUANG, W., RICHARDS, S., CARBONE, M. A., ZHU, D., ANHOLT, R. R., AYROLES, J. F., DUNCAN, L., JORDAN, K. W., LAWRENCE, F., MAGWIRE, M. M., WARNER, C. B., BLANKENBURG, K., HAN, Y., JAVAID, M., JAYASEELAN, J., JHANGIANI, S. N., MUZNY, D., ONGERI, F., PERALES, L., WU, Y. Q., ZHANG, Y., ZOU, X., STONE, E. A., GIBBS, R. A. & MACKAY, T. F. 2012. Epistasis dominates the genetic architecture of Drosophila quantitative traits. *Proc Natl Acad Sci U S A,* 109**,** 15553-9.

JIMENEZ-GOMEZ, J. M., CORWIN, J. A., JOSEPH, B., MALOOF, J. N. & KLIEBENSTEIN, D. J. 2011. Genomic analysis of QTLs and genes altering natural variation in stochastic noise. *PLoS Genet,* 7**,** e1002295.

LAPPALAINEN, T., MONTGOMERY, S. B., NICA, A. C. & DERMITZAKIS, E. T. 2011. Epistatic Selection between Coding and Regulatory Variation in Human Evolution and Disease. *American journal of human genetics,* 89**,** 459-463.

LAPPALAINEN, T., SAMMETH, M., FRIEDLANDER, M. R., T HOEN, P. A., MONLONG, J., RIVAS, M. A., GONZALEZ-PORTA, M., KURBATOVA, N., GRIEBEL, T., FERREIRA, P. G., BARANN, M., WIELAND, T., GREGER, L., VAN ITERSON, M., ALMLOF, J., RIBECA, P., PULYAKHINA, I., ESSER, D., GIGER, T., TIKHONOV, A., SULTAN, M., BERTIER, G., MACARTHUR, D. G., LEK, M., LIZANO, E., BUERMANS, H. P., PADIOLEAU, I., SCHWARZMAYR, T., KARLBERG, O., ONGEN, H., KILPINEN, H., BELTRAN, S., GUT, M., KAHLEM, K., AMSTISLAVSKIY, V., STEGLE, O., PIRINEN, M., MONTGOMERY, S. B., DONNELLY, P., MCCARTHY, M. I., FLICEK, P., STROM, T. M., LEHRACH, H., SCHREIBER, S., SUDBRAK, R., CARRACEDO, A., ANTONARAKIS, S. E., HASLER, R., SYVANEN, A. C., VAN OMMEN, G. J., BRAZMA, A., MEITINGER, T., ROSENSTIEL, P., GUIGO, R., GUT, I. G., ESTIVILL, X., DERMITZAKIS, E. T. & THE GEUVADIS CONSORTIUM 2013. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature,* 501**,** 506-11.

LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25**,** 1754-60.

MACKAY, T. F. 2014. Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nat Rev Genet,* 15**,** 22-33.

MANOLIO, T. A., COLLINS, F. S., COX, N. J., GOLDSTEIN, D. B., HINDORFF, L. A., HUNTER, D. J., MCCARTHY, M. I., RAMOS, E. M., CARDON, L. R. & CHAKRAVARTI, A. 2009. Finding the missing heritability of complex diseases. *Nature,* 461**,** 747-753.

MARIONI, J. C., MASON, C. E., MANE, S. M., STEPHENS, M. & GILAD, Y. 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome research,* 18**,** 1509-1517.

MARTIN, N., ROWELL, D. & WHITFIELD, J. 1983. Do the MN and Jk systems influence environmental variability in serum lipid levels? *Clinical Genetics,* 24**,** 1-14.

NAKAGAWA, S. & SCHIELZETH, H. 2012. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*.

OTTO, S. P. & FELDMAN, M. W. 1997. Deleterious mutations, variable epistatic interactions, and the evolution of recombination. *Theor Popul Biol,* 51**,** 134-47.

PARÉ, G., COOK, N. R., RIDKER, P. M. & CHASMAN, D. I. 2010. On the Use of Variance per Genotype as a Tool to Identify Quantitative Trait Interaction Effects: A Report from the Women's Genome Health Study. *PLoS Genet,* 6**,** e1000981.

PARTS, L., STEGLE, O., WINN, J. & DURBIN, R. 2011. Joint genetic analysis of gene expression data with inferred cellular phenotypes. *PLoS Genet,* 7**,** e1001276.

POWELL, J. E., HENDERS, A. K., MCRAE, A. F., KIM, J., HEMANI, G., MARTIN, N. G., DERMITZAKIS, E. T., GIBSON, G., MONTGOMERY, G. W. & VISSCHER, P. M. 2013. Congruence of Additive and Non-Additive Effects on Gene Expression Estimated from Pedigree and SNP Data. *PLoS genetics,* 9**,** e1003502.

R DEVELOPMENT CORE TEAM 2008. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

REYNOLDS, C. A., GATZ, M., BERG, S. & PEDERSEN, N. L. 2007. Genotype–environment interactions: cognitive aging and social factors. *Twin Research and Human Genetics,* 10**,** 241-254.

ROSENBLOOM, K. R., SLOAN, C. A., MALLADI, V. S., DRESZER, T. R., LEARNED, K., KIRKUP, V. M., WONG, M. C., MADDREN, M., FANG, R. & HEITNER, S. G. 2013. ENCODE Data in the UCSC Genome Browser: year 5 update. *Nucleic acids research,* 41**,** D56-D63.

STOREY, J. D. 2002. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 64**,** 479-498.

SUN, X., ELSTON, R., MORRIS, N. & ZHU, X. 2013. What Is the Significance of Difference in Phenotypic Variability across SNP Genotypes? *Am J Hum Genet,* 93**,** 390-7.

THE 1000 GENOMES PROJECT CONSORTIUM 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature,* 491**,** 56-65.

THE INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM 2001. Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860-921.

WANG, G., YANG, E., BRINKMEYER-LANGFORD, C. L. & CAI, J. J. 2014. Additive, Epistatic, and Environmental Effects Through the Lens of Expression Variability QTL in a Twin Cohort. *Genetics,* 196**,** 413-25.

WANG, Z., GERSTEIN, M. & SNYDER, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics,* 10**,** 57-63.

WENTZELL, A. M., ROWE, H. C., HANSEN, B. G., TICCONI, C., HALKIER, B. A. & KLIEBENSTEIN, D. J. 2007. Linking metabolic QTLs with network and cis-eQTLs controlling biosynthetic pathways. *PLoS Genet,* 3**,** 1687-701.

YANG, J., LOOS, R. J. F., POWELL, J. E., MEDLAND, S. E., SPELIOTES, E. K., CHASMAN, D. I., ROSE, L. M., THORLEIFSSON, G., STEINTHORSDOTTIR, V., MAGI, R., WAITE, L., VERNON SMITH, A., YERGES-ARMSTRONG, L. M., MONDA, K. L., HADLEY, D., MAHAJAN, A., LI, G., KAPUR, K., VITART, V., HUFFMAN, J. E., WANG, S. R., PALMER, C., ESKO, T., FISCHER, K., HUA ZHAO, J., DEMIRKAN, A., ISAACS, A., FEITOSA, M. F., LUAN, J. A., HEARD-COSTA, N. L., WHITE, C., JACKSON, A. U., PREUSS, M., ZIEGLER, A., ERIKSSON, J., KUTALIK, Z., FRAU, F., NOLTE, I. M., VAN VLIET-OSTAPTCHOUK, J. V., HOTTENGA, J.-J., JACOBS, K. B., VERWEIJ, N., GOEL, A., MEDINA-GOMEZ, C., ESTRADA, K., LYNN BRAGG-GRESHAM, J., SANNA, S., SIDORE, C., TYRER, J., TEUMER, A., PROKOPENKO, I., MANGINO, M., LINDGREN, C. M., ASSIMES, T. L., SHULDINER, A. R., HUI, J., BEILBY, J. P., MCARDLE, W. L., HALL, P., HARITUNIANS, T., ZGAGA, L., KOLCIC, I., POLASEK, O., ZEMUNIK, T., OOSTRA, B. A., JUHANI JUNTTILA, M., GRONBERG, H., SCHREIBER, S., PETERS, A., HICKS, A. A., STEPHENS, J., FOAD, N. S., LAITINEN, J., POUTA, A., KAAKINEN, M., WILLEMSEN, G., VINK, J. M., WILD, S. H., NAVIS, G., ASSELBERGS, F. W., HOMUTH, G., JOHN, U., IRIBARREN, C., HARRIS, T., LAUNER, L., GUDNASON, V., O/'CONNELL, J. R., BOERWINKLE, E., CADBY, G., PALMER, L. J., JAMES, A. L., MUSK, A. W., INGELSSON, E., PSATY, B. M., BECKMANN, J. S., WAEBER, G., VOLLENWEIDER, P., HAYWARD, C., WRIGHT, A. F., RUDAN, I., et al. 2012. FTO genotype is associated with phenotypic variability of body mass index. *Nature,* 490**,** 267-272.

ZUK, O., HECHTER, E., SUNYAEV, S. R. & LANDER, E. S. 2012. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A,* 109**,** 1193-8.

Test	Threshold	Associations (Available for testing in GEUVADIS)	Replicate, FDR < 0.05 (% success)	Same direction of effect (% success)	π1
v-eQTL	FDR<0.05	508 (485)	28 (5.8%)	26 (93%)	0.30
v-eQTL	Bonf<0.05	23 (23)	16 (70%)	15 (94%)	0.72
Epistasis	Bonf<0.05	256 (246)	137 (56%)	131 (96%)	0.71

Table 1: Replication analysis. Significant associations (at FDR and Bonferroni thresholds) from the TwinsUK sample were replicated in GEUVADIS samples. The number of overlapping SNPs and genes in both datasets per analysis is shown, as well as the percentage of replicated associations. π_1 is an estimate of the proportion of replicating loci in the GEUVADIS cohort (Storey, 2002)

Gene	Chr	v-eQTL	Interacting epistasis SNP	Interaction variance in TwinsUK	Interaction variance in GEUVADIS	Additive variation in GEUVADIS	*P*-value in TwinsUK	*P*-value in GEUVADIS
NUDT2	9	rs10972055	rs10814083	-0.328	-0.128	0.310	1.88×10^{-53}	$5.43\times10\text{-}^{22}$
HLA-DQB2	6	rs114183935	rs1049130	-0.337	-0.161	0.099	1.83×10^{-62}	2.91×10^{-21}
HLA-DQB2	6	rs114183935	rs9274666	-0.368	-0.119	0.158	3.45×10^{-18}	1.04×10^{-16}
SPATA20	17	rs12943759	rs1122634	0.301	0.078	0.404	3.12×10^{-69}	1.42×10^{-15}
POU5F1	6	rs116627368	rs115631087	0.311	0.116	0.008	6.95×10^{-34}	6.63×10^{-14}
SERPINB1	6	rs318452	rs6940344	-0.227	-0.102	0.117	2.40×10^{-36}	7.66×10^{-14}
ANXA5	4	rs6857766	rs12511956	-0.411	-0.104	0.056	3.09×10^{-37}	3.81×10^{-13}
TCF19	6	rs115523621	rs115921994	-0.585	-0.076	0.201	2.59×10^{-36}	1.48×10^{-11}
HLA-C	6	rs114916097	rs116012228	0.160	0.077	0.183	3.35×10^{-18}	2.17×10^{-11}
PHLDB3	19	rs10409591	rs2682547	-0.270	-0.0858	0.0569	1.67×10^{-14}	4.83×10^{-11}

Table 2: Effect size estimates and significance for the ten most significant replicated interactions in TwinsUK and GEUVADIS. Effect sizes are reported as the proportion of variance explained by the interaction. Sign of effect size reflects direction of interaction effect: positive implies combined effect of the alternate alleles is an increase in expression greater than predicted by separate additive effects, and negative that it is less.

Figures:

Figure 1 (a) Genotype dependent variance analysis identifies candidate SNPs for interactions. The plot shows expression of the gene *TRIT1* broken down by v-eQTL genotype (rs3131691), to illustrate how an interaction can be observed as an increase in variance. The genotype at rs3131691 interacts with the genotype of rs230273. Orange individuals are carriers of the C allele at rs230273, that decreases expression only in the AG and GG genotype groups of rs3131691. Observing only expression conditioned on rs3131691, this induced bimodality increases the variance of the observations within these groups. Jitter has been introduced in the x axis to reduce overplotting. **(b) Histogram of distance from transcription start site in kilobases for the 508 peak v-eQTL hits.** Figure shows the clustering of the 508 v-eQTL discovered in the TwinsUK cohort around the transcription start site, with downstream of the TSS counted as positive. The orange triangles below mark the positions of the 26 v-eQTL which replicated in the GEUVADIS cohort.

Figure 2 (a) *TRIT1* expression is affected by an interaction between two SNPs in both TwinsUK and GEUVADIS cohorts. Expression of *TRIT1* is shown, with a separate panel for each v-eQTL (rs3131691) genotype group. Relationship between expression and imputed genotype dosage of the epistasis SNP (rs230273) is shown to be conditional on v-eQTL genotype. Expression from TwinsUK individuals is shown in the upper panels, GEUVADIS individuals in the lower panels. Best fit lines show different SNP effects for the epistatic SNPs in different v-eQTL genotype groups, these lines are constructed ignoring twin structure in the case of the TwinsUK sample and population in the GEUVADIS cohort. **(b) SNPs affecting *TRIT1* expression are near regulatory elements.** Position of v-eQTL (rs3131691), interacting epistasis SNP (rs230273) and a nearby eQTL (rs34387655) affecting *TRIT1* expression are shown. ENCODE segmentation analysis shows regulatory elements around *TRIT1* (reverse strand gene). Colours indicating regions are: yellow = weak enhancer, orange = strong enhancer, red = strong promoter, light red = weak promoter, purple = poised promoter, dark green = transcriptional transition/elongation, light green = weakly transcribed, blue = insulator, and light grey = heterochromatin or repetitive/copy number variation.

Figure 3: Variance explained by additive and interacting variants for 57 replicated examples of epistasis in the GEUVADIS cohort. We show the variation explained by the interaction of two SNPs on phenotype, compared to the additive contribution of the SNPs.

Figure 4 (a) Increased discordance within MZ twin pairs identifies G×E interactions. We show discordance in expression between MZ twin pairs for the gene *BAMBI* broken down by v-eQTL genotype (rs10826519). Discordance is greatest in the GG genotype group (mean difference between MZ twins is 1.12), decreasing with each additional copy of the A allele (mean discordance is 0.85 for GA genotype group, 0.60 for AA). Since MZ twins are genetically identical, genotype dependent discordance in expression must be a consequence of environment, pointing to GxE. We observe that the SNP also has an effect on the mean level of expression ($p=5.42\times10^{-19}$). (**b) -log10 p values for genotype dependent discordance in MZ twins against –log10 p values for peak v-eQTL.** The blue dots represent points where there is a significant epistasis hit with the v-eQTL, orange where no such interaction was detected. For many of the strong v-eQTL with little evidence of discordance we can identify an epistatic interaction which explains the increase in variance. However, for some loci with strong evidence of genotype dependent MZ discordance we also detect an epistatic interaction, suggesting both epistasis and GxE acts on these genes.

Figure 1-figure supplement 1: Peak v-eQTL signals for 13,660 genes. *P*-values for SNPs associated with variance in gene expression (v-eQTL) are plotted against their genomic position. Horizontal line indicates FDR=0.05 cut off. Only the most significant v-eQTL for each gene is plotted, explaining isolated signals and there being few signals with *p*-value > 0.01.

Figure 1-figure supplement 2: -log10 p value for v-eQTL against –log10 p value for eQTL for 508 v-eQTL hits estimated in the TwinsUK cohort.

Figure 1-figure supplement 3: Variance of expression of ENSG00000164978 (*NUDT2*) is dependent on genotype dosage of rs10972055.

Figure 1-figure supplement 4: Variance of expression of ENSG00000105499 (*PLA2GC4*) is dependent on genotype dosage of rs8109684.

Figure 1-figure supplement 5: Variance of expression of ENSG00000043514 (*TRIT1*) is dependent on genotype dosage of rs3131691.

Figure 1-figure supplement 6: Variance of expression of ENSG00000075234 (*TTC38*) is dependent on genotype dosage of rs6008743.

Figure 1-figure supplement 7: Variance of expression of ENSG00000164111 (*ANXA5*) is dependent on genotype dosage of rs6857766.

Figure 1-figure supplement 8: Variance of expression of ENSG00000137054 (*POLR1E*) is dependent on genotype dosage of rs7033474.

Figure 1-figure supplement 9: Variance of expression of ENSG00000168765 (*GSTM4*) is dependent on genotype dosage of rs542338.

Figure 1-figure supplement 10: Variance of expression of ENSG00000232629 (*HLA-DQB2*) is dependent on genotype dosage of rs114183935.

Figure 1-figure supplement 11: Variance of expression of ENSG00000196735 (*HLA-DQA1*) is dependent on genotype dosage of rs9276807.

Figure 1-figure supplement 12: Variance of expression of ENSG00000160284 (*C21orf56*) is dependent on genotype dosage of rs16978976.

Figure 2-figure supplement 1: Evidence for epistasis in twins against evidence for epistasis in 1000 Genomes for the 246 significant hits. The 57 replicated associations after removing possible haplotype effects are shown in blue.

Figure 2-figure supplement 2: Estimate of interaction effect size in 1000 Genomes and twins cohorts. Effect size is reported as proportion of variance explained by the interaction, where sign is positive if when both variants have the alternate allele, the combined effect is a greater increase in expression than predicted by the separate additive effects, negative if expression is decreased comparatively. The 57 replicated associations are shown in blue.

Figure 2-figure supplement 3: ENSG00000164978 (*NUDT2)* expression is affected by an interaction between two SNPs in both TwinsUK and GEUVADIS cohorts. Expression of *NUDT2* is shown, with a separate panel for each v-eQTL (rs10972055) genotype group. Relationship between expression and imputed genotype dosage of the epistasis SNP (rs10814083) is shown to be conditional on v-eQTL genotype. Expression from TwinsUK individuals is shown in the upper panels, GEUVADIS individuals in the lower panels. Best fit lines indicate the different epistatic SNP effects in the different v-eQTL genotype groups and are illustrative only. These lines are constructed ignoring twin structure in the case of the TwinsUK sample and population in the GEUVADIS cohort and do not represent model fit for the analysis performed.

Figure 2-figure supplement 4: ENSG00000232629 (*HLA-DQB2*) expression is affected by an interaction between two SNPs in both TwinsUK and GEUVADIS cohorts. Expression of *HLA-DQB2* is shown, with a separate panel for each v-eQTL (rs114183935) genotype group. Relationship between expression and imputed genotype dosage of the epistasis SNP (rs1049130) is shown to be conditional on v-eQTL genotype. Expression from TwinsUK individuals is shown in the upper panels, GEUVADIS individuals in the lower panels. Best fit lines indicate the different epistatic SNP effects in the different v-eQTL genotype groups and are illustrative only. These lines are constructed ignoring twin structure in the case of the TwinsUK sample and population in the GEUVADIS cohort and do not represent model fit for the analysis performed.

Figure 2-figure supplement 5: ENSG00000232629 (*HLA-DQB2*) expression is affected by an interaction between two SNPs in both TwinsUK and GEUVADIS cohorts. Expression of *HLA-DQB2* is shown, with a separate panel for each v-eQTL (rs114183935) genotype group. Relationship between expression and imputed genotype dosage of the epistasis SNP (rs9274666) is shown to be conditional on v-eQTL genotype. Expression from TwinsUK individuals is shown in the upper panels, GEUVADIS individuals in the lower panels. Best fit lines indicate the different epistatic SNP effects in the different v-eQTL genotype groups and are illustrative only. These lines are constructed ignoring twin structure in the case of the TwinsUK sample and population in the GEUVADIS cohort and do not represent model fit for the analysis performed.

Figure 2-figure supplement 6: ENSG00000006282 (*SPATA20*) expression is affected by an interaction between two SNPs in both TwinsUK and GEUVADIS cohorts. Expression of *SPATA20* is shown, with a separate panel for each v-eQTL (rs12943759) genotype group. Relationship between expression and imputed genotype dosage of the epistasis SNP (rs1122634) is shown to be conditional on v-eQTL genotype. Expression from TwinsUK individuals is shown in the upper panels, GEUVADIS individuals in the lower panels. Best fit lines indicate the different epistatic SNP effects in the different v-eQTL genotype groups and are illustrative only. These lines are constructed ignoring twin structure in the case of the TwinsUK sample and population in the GEUVADIS cohort and do not represent model fit for the analysis performed.

Figure 2-figure supplement 7: ENSG00000204531 (*POU5F1*) expression is affected by an interaction between two SNPs in both TwinsUK and GEUVADIS cohorts. Expression of *POU5F1* is shown, with a separate panel for each v-eQTL (rs116627368) genotype group. Relationship between expression and imputed genotype dosage of the epistasis SNP (rs115631087) is shown to be conditional on v-eQTL genotype. Expression from TwinsUK individuals is shown in the upper panels, GEUVADIS individuals in the lower panels. Best fit lines indicate the different epistatic SNP effects in the different v-eQTL genotype groups and are illustrative only. These lines are constructed ignoring twin structure in the case of the TwinsUK sample and population in the GEUVADIS cohort and do not represent model fit for the analysis performed.

Figure 2-figure supplement 8: ENSG00000021355 (*SERPINB1*) expression is affected by an interaction between two SNPs in both TwinsUK and GEUVADIS cohorts. Expression of *SERPINB1* is shown, with a separate panel for each v-eQTL (rs318452) genotype group. Relationship between expression and imputed genotype dosage of the epistasis SNP (rs6940344) is shown to be conditional on v-eQTL genotype. Expression from TwinsUK individuals is shown in the upper panels, GEUVADIS individuals in the lower panels. Best fit lines indicate the different epistatic SNP effects in the different v-eQTL genotype groups and are illustrative only. These lines are constructed ignoring twin structure in the case of the TwinsUK sample and population in the GEUVADIS cohort and do not represent model fit for the analysis performed.

Figure 2-figure supplement 9: ENSG00000164111 (*ANXA5*) expression is affected by an interaction between two SNPs in both TwinsUK and GEUVADIS cohorts. Expression of *ANXA5* is shown, with a separate panel for each v-eQTL (rs6857766) genotype group. Relationship between expression and imputed genotype dosage of the epistasis SNP (rs12511956) is shown to be conditional on v-eQTL genotype. Expression from TwinsUK individuals is shown in the upper panels, GEUVADIS individuals in the lower panels. Best fit lines indicate the different epistatic SNP effects in the different v-eQTL genotype groups and are illustrative only. These lines are constructed ignoring twin structure in the case of the TwinsUK sample and population in the GEUVADIS cohort and do not represent model fit for the analysis performed.

Figure 2-figure supplement 10: ENSG00000137310 (*TCF19)* expression is affected by an interaction between two SNPs in both TwinsUK and GEUVADIS cohorts. Expression of *TCF19* is shown, with a separate panel for each v-eQTL (rs115523621) genotype group. Relationship between expression and imputed genotype dosage of the epistasis SNP (rs115921994) is shown to be conditional on v-eQTL genotype. Expression from TwinsUK individuals is shown in the upper panels, GEUVADIS individuals in the lower panels. Best fit lines indicate the different epistatic SNP effects in the different v-eQTL genotype groups and are illustrative only. These lines are constructed ignoring twin structure in the case of the TwinsUK sample and population in the GEUVADIS cohort and do not represent model fit for the analysis performed.

Figure 2-figure supplement 11: ENSG00000204525 (*HLA-C*) expression is affected by an interaction between two SNPs in both TwinsUK and GEUVADIS cohorts. Expression of *HLA-C* is shown, with a separate panel for each v-eQTL (rs114916097) genotype group. Relationship between expression and imputed genotype dosage of the epistasis SNP (rs116012228) is shown to be conditional on v-eQTL genotype. Expression from TwinsUK individuals is shown in the upper panels, GEUVADIS individuals in the lower panels. Best fit lines indicate the different epistatic SNP effects in the different v-eQTL genotype groups and are illustrative only. These lines are constructed ignoring twin structure in the case of the TwinsUK sample and population in the GEUVADIS cohort and do not represent model fit for the analysis performed.

Figure 2-figure supplement 12: ENSG00000176531 (*PHLDB3*) expression is affected by an interaction between two SNPs in both TwinsUK and GEUVADIS cohorts. Expression of *PHLDB3* is shown, with a separate panel for each v-eQTL (rs10409591) genotype group. Relationship between expression and imputed genotype dosage of the epistasis SNP (rs2682547) is shown to be conditional on v-eQTL genotype. Expression from TwinsUK individuals is shown in the upper panels, GEUVADIS individuals in the lower panels. Best fit lines indicate the different epistatic SNP effects in the different v-eQTL genotype groups and are illustrative only. These lines are constructed ignoring twin structure in the case of the TwinsUK sample and population in the GEUVADIS cohort and do not represent model fit for the analysis performed.

Figure 2-figure supplement 13: The distance in kilobases from the 246 variants in epistasis to the v-eQTL plotted against the –log10 p value in 1000 Genomes sample. Using the p value in the replication sample avoids inflation by winners curse. The blue dots are the 57 replicated associations after removing haplotype effects.

Supplementary Files:

Supplementary File 1A: Peak vQTL hits in twins sample with evidence of eQTL and discordant QTL.

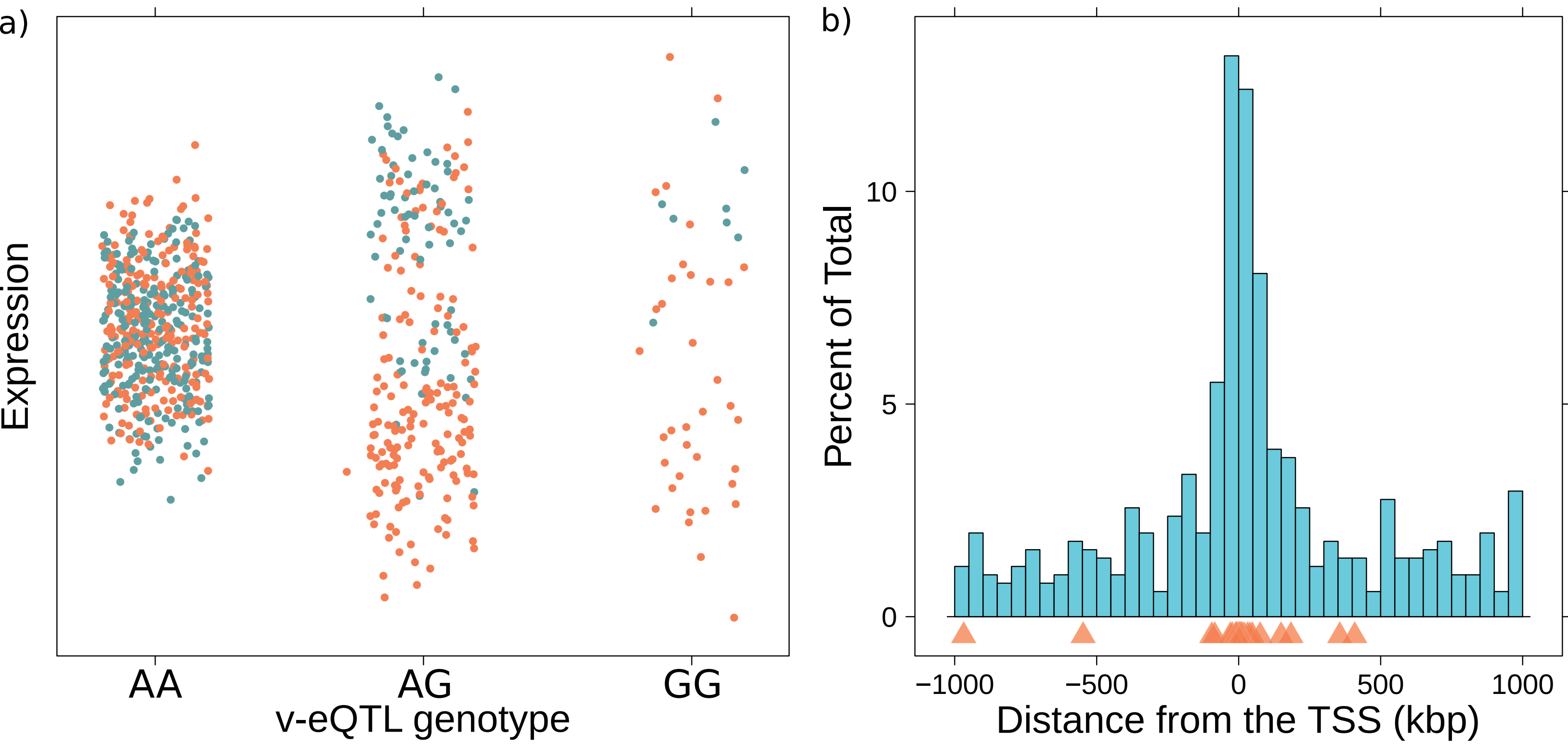
Supplementary File 1B: Significant epistasis hits in twins sample with p values and effect size estimates in 1000 genomes cohort.

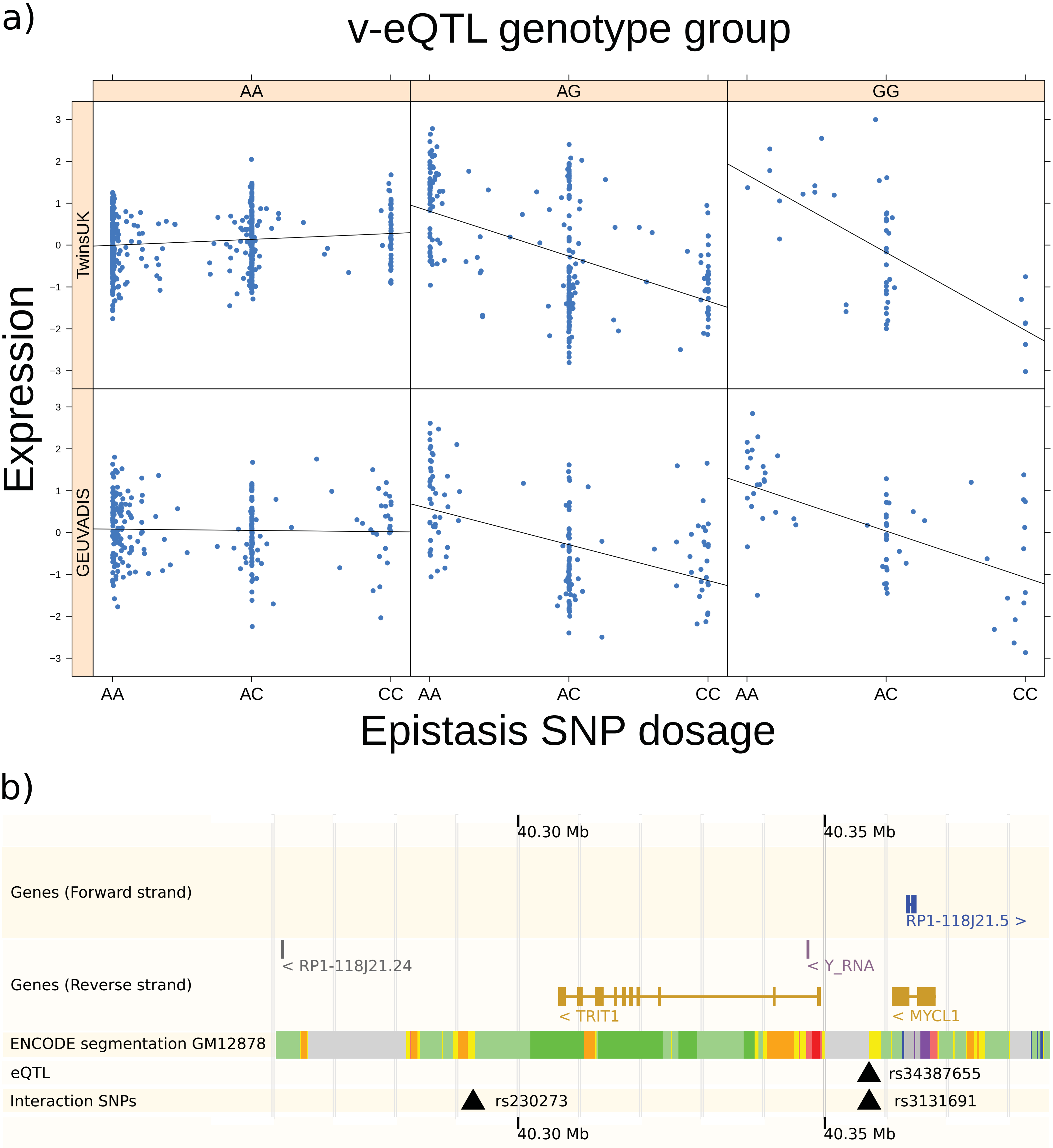
Supplementary File 1C: Contribution of cis variants, trans variants, interactions between the two and unique environment to variation in gene expression.

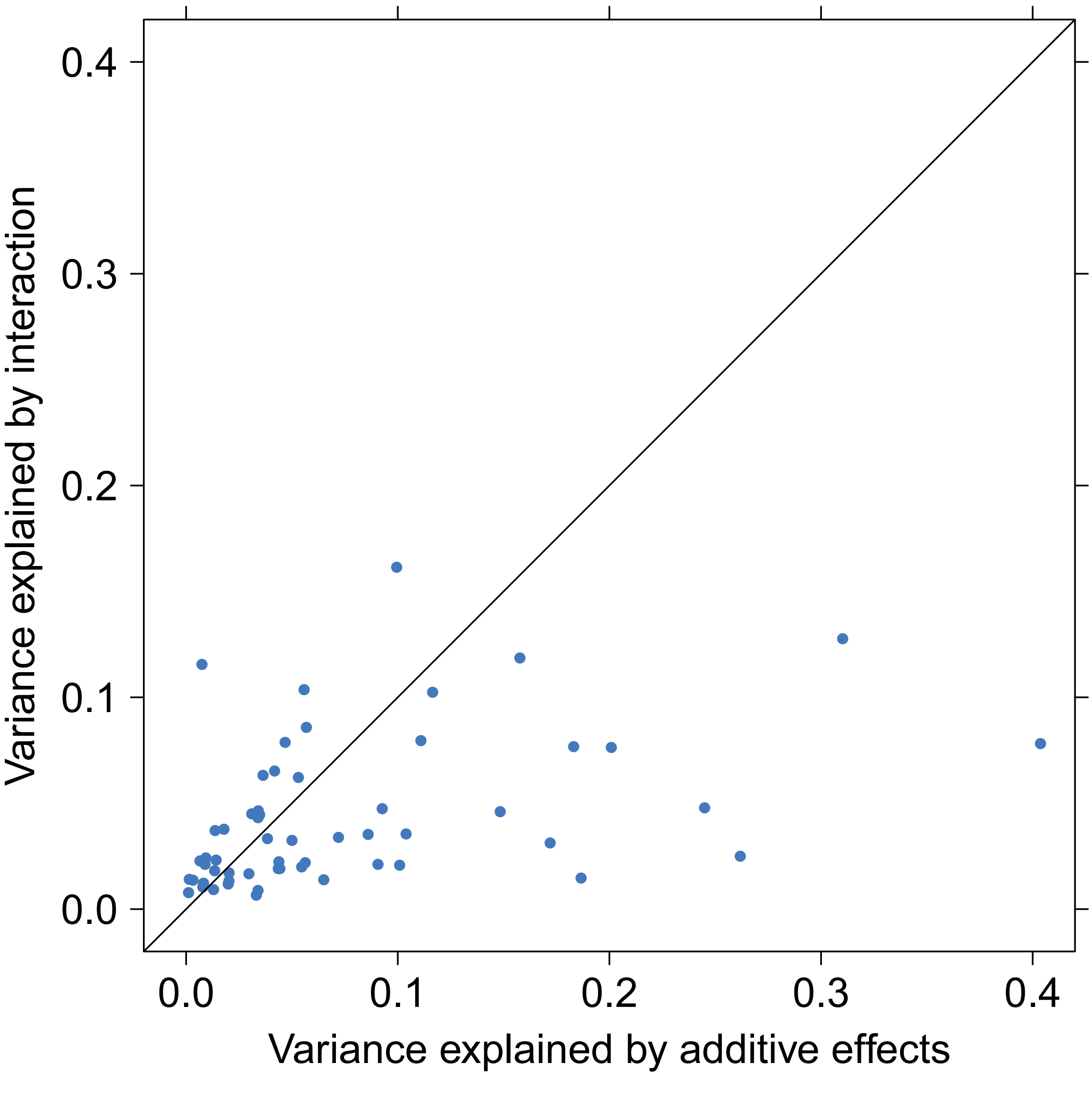
Supplementary File 2: R functions applied to data from the TwinsUK cohort to test individual SNPs for variance effects, to map all independent epistatic interactions with the v-eQTL in the cis window and to eliminate dominance effects from list of epistatic interactions.

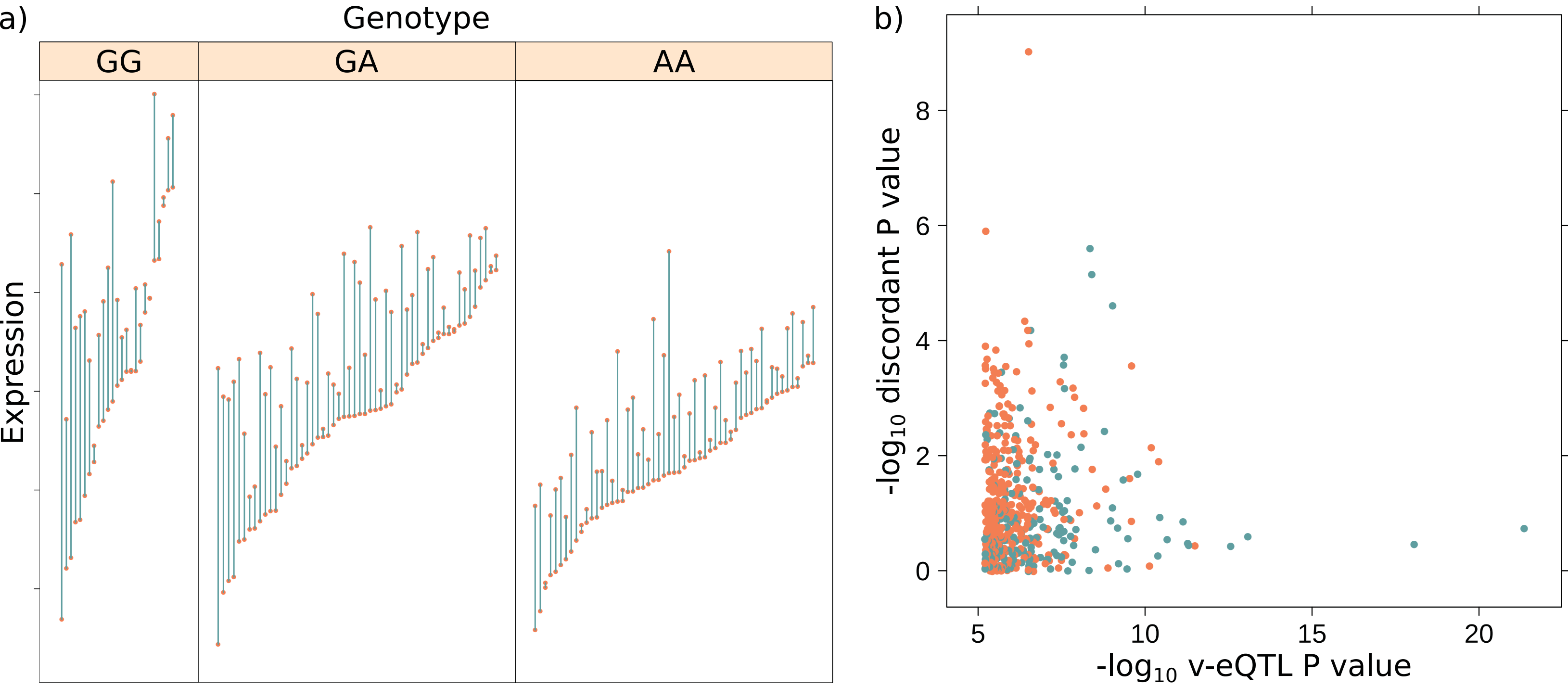
Supplementary File 3: R workspace containing replication data from the GEUVADIS cohort (Lappalainen et al., 2013) together with functions to repeat the replication analysis.

Supplementary File 4: Read me file explaining objects present in SM2.


a)
Expression
AA
AG
GG
v-eQTL genotype
b)
Percent of Total
0
5
10
−1000
−500
0
500
1000
Distance from the TSS (kbp)


a)
v-eQTL genotype group
AA
AG
GG
TwinsUK
GEUVADIS
Expression
3
2
1
0
−1
−2
−3
AA
AC
CC
Epistasis SNP dosage
b)
40.30 Mb
40.35 Mb
Genes (Forward strand)
RP1-118J21.5 >
Genes (Reverse strand)
< RP1-118J21.24
< Y_RNA
< TRIT1
< MYCL1
ENCODE segmentation GM12878
eQTL
rs34387655
Interaction SNPs
rs230273
rs3131691


0.4
0.3
0.2
0.1
0.0
0.0
0.1
0.2
0.3
0.4
Variance explained by interaction
Variance explained by additive effects


a)
Genotype
GG
GA
AA
Expression
b)
8
6
4
2
0
-log$_{10}$ discordant P value
5
10
15
20
-log$_{10}$ v-eQTL P value