Additive, Epistatic, and Environmental Effects Through the Lens of Expression Variability QTL in a Twin Cohort

Gang Wang,*,[1] Ence Yang,*,[1] Candice L. Brinkmeyer-Langford,* and James J. Cai*,†,[2]
*Department of Veterinary Integrative Biosciences and †Interdisciplinary Program in Genetics, Texas A&M University, College Station, Texas 77843

ABSTRACT The expression of a gene can vary across individuals in the general population, as well as between monozygotic twins. This variable expression is assumed to be due to the influence of both genetic and nongenetic factors. Yet little evidence supporting this assumption has been obtained from empirical data. In this study, we used expression data from a large twin cohort to investigate the influences of genetic and nongenetic factors on variable gene expression. We focused on a set of expression variability QTL (evQTL)—*i.e.*, genetic loci associated with the variance, as opposed to the mean, of gene expression. We identified evQTL for 99, 56, and 79 genes in lymphoblastoid cell lines, skin, and fat, respectively. The differences in gene expression, measured by the relative mean difference (RMD), tended to be larger between pairs of dizygotic (DZ) twins than between pairs of monozygotic (MZ) twins, showing that genetic background influenced the expression variability. Furthermore, a more profound RMD was observed between pairs of MZ twins whose genotypes were associated with greater expression variability than the RMD found between pairs of MZ twins whose genotypes were associated with smaller expression variability. This suggests that nongenetic (*e.g.*, environmental) factors contribute to the variable expression. Lastly, we demonstrated that the formation of evQTL is likely due to partial linkages between eQTL SNPs that are additively associated with the mean of gene expression; in most cases, no epistatic effect is involved. Our findings have implications for understanding divergent sources of gene expression variability.

VARIATION and variability are central concepts in biology (Hallgrímsson and Hall 2005). Although often used interchangeably in the scientific literature, the two are not synonymous. Variation refers to the differences among individuals, whereas variability refers to the potential of a population to vary (Wagner 1995; Wagner and Altenberg 1996). In many cases, greater phenotypic variability (*e.g.*, transcriptional noise) is disadvantageous (Kemkemer *et al.* 2002; Bahar *et al.* 2006; Wang and Zhang 2011) unless it gives rise to greater organismal plasticity—first at the level of an individual organism and eventually at the population level. Genetic factors resulting in more variable phenotypes become favored when they enable a population to more effectively respond to environmental changes (Hill and Zhang 2004; Kaern *et al.* 2005; Acar *et al.* 2008; Zhang *et al.* 2009). Thus, understanding to what extent and in what ways genotypes influence phenotypic variability is of fundamental importance.

Much effort has been focused on identifying genetic loci such as expression quantitative trait loci, or eQTL (Stranger *et al.* 2005, 2007; Choy *et al.* 2008; Montgomery *et al.* 2010; Pickrell *et al.* 2010; Montgomery and Dermitzakis 2011), that affect the average value of a phenotype, while ignoring those that affect the variance of a phenotype. However, there is increasing evidence across species for genetic loci that affect the variance of phenotype (Queitsch *et al.* 2002; Jimenez-Gomez *et al.* 2011; Ronnegard and Valdar 2011; Perry *et al.* 2012; Shen *et al.* 2012; Yang *et al.* 2012). Recently we introduced the concept of expression variability QTL, or evQTL (Hulse and Cai 2013). By definition, an evQTL is a genetic locus linked to or associated with genetic variation influencing the variance of gene expression in a population. To identify evQTL, we previously adapted the method developed by Ronnegard and Valdar (2011) for detecting vQTL based on the double generalized linear model (dglm) (Verbyla and Smyth

doi: 10.1534/genetics.113.157503
Manuscript received September 13, 2013; accepted for publication November 21, 2013; published Early Online December 2, 2013.
Supporting information is available online at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.157503/-/DC1.
[1]These authors contributed equally to this work.
[2]Corresponding author: Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77843. E-mail: jcai@tamu.edu

1998). The dglm method tests for expression variances and measures the contribution of genetic variants to the expression heteroscedasticity. It compares the fit of a full model, which takes into account the contribution of genotype to both the mean and the variance of gene expression simultaneously, and a mean model, which only takes into account the contribution of genotype to the mean and ignores the contribution to the variance. A significant result of a dglm test shows the nonrandom association between genotypes and gene expression variances. Using this method, we have conducted a genome-wide scan for evQTL in the human genome (Hulse and Cai 2013).

How an evQTL is created in the first place is not clear. We consider two possible scenarios, emphasizing either the role of environmental or genetic factors. The first possibility is that specific genetic variants disrupt the stabilizing genetic architecture that buffers stochastic variation in phenotype. As a result of such an effect of decanalization, along with the sensitization of the stabilizer (*e.g.*, heat-shock protein 90), the phenotype becomes more sensitive to external environment and varies more greatly between individuals (Ronnegard and Valdar 2011; Hulse and Cai 2013).

The other possible scenario concerns the role of genetic interactions in the formation of evQTL. Through either epistatic or nonepistatic (*e.g.*, additive and dominance) effects within or between loci, genetic interactions contribute to genotypic variance. Epistasis may increase the variance of a quantitative trait (Pare *et al.* 2010; Ronnegard and Valdar 2012). However, it is extremely difficult to distinguish the contributions of epistatic or nonepistatic effects to variable expression of genes. Epistasis, in particular, is known to produce predominantly additive and dominance genetic variance when the low-frequency alleles of some SNPs are involved (Cheverud and Routman 1995).

Here we investigated the distribution and formation of evQTL by leveraging the existing dataset (Grundberg *et al.* 2012) derived from a population-based cohort of twin studies (Moayyeri *et al.* 2013). We interrogated this dataset for evQTL and investigated the roles of genetic and nongenetic factors in the formation of the evQTL we identified. The twin cohort offered a unique advantage for studying the relative contributions of various factors that influence expression variability. Importantly, comparing expression data of monozygotic and dizygotic twins allowed us to distinguish between genetic and nongenetic effects. In the following sections, we present the descriptive statistics for expression variability in the twin cohort, describe the detection of evQTL, and finally estimate the relative contributions of genetic and nongenetic factors, as well as epistatic and nonepistatic effects, to the creation of evQTL.

Materials and Methods

The TwinsUK dataset

We obtained the TwinsUK dataset including both genotype and expression data, which had been used in the eQTL study of Grundberg *et al.* (2012). Here, we briefly describe the cohort and data processing performed in this previous study (Grundberg *et al.* 2012). The TwinsUK cohort includes 856 female individuals of European descent recruited from the TwinsUK adult twin registry (Spector and Williams 2006; Moayyeri *et al.* 2013). Subcutaneous adipose tissue, skin tissue, and lymphoblastoid cell lines (LCLs) were collected from each individual. Genotyping was performed with a combination of Illumina HumanHap300, HumanHap610Q, 1-M Duo and 1.2-M Duo 1-M chips. Genotypes were called with the Illuminus calling algorithm (Teo *et al.* 2007), and SNPs were filtered for minor allele frequency (MAF) of >5%. Gene expression levels were measured in LCLs, skin, and adipose (Grundberg *et al.* 2012). Expression profiling of the samples, each with either two or three technical replicates, was performed using Illumina Human HT-12 V3 BeadChips (Illumina). All samples were randomized before array hybridization, and replicates were hybridized on different BeadChips. Raw data were imported to Illumina BeadStudio software, and probes with fewer than three beads present were excluded. Log_2-transformed expression signals were normalized separately per tissue, with quantile normalization of the replicates of each individual followed by quantile normalization across all individuals (Nica *et al.* 2011).

In this study, we used available gene expression data for both individuals of a twin pair. All 48,804 probe sequences were mapped by BLAST to the reference genome (hg18), and probes found to map to more than one location were not used. Polymorphisms in the target mRNA sequence can greatly affect the binding affinity of microarray probe sequences, leading to false-positive and false-negative signals with any other polymorphisms in linkage disequilibrium (LD) (Ramasamy *et al.* 2013). To control for this, we used a comprehensive compendium of SNPs in European (CEU) population of the 1000 Genomes Project Consortium (2012) to remove an additional 13,600 probes found to anneal in regions with SNPs present at a MAF of >5%. Similarly, probes mapping to nonautosomal locations were excluded from further analysis. Finally, 35,078 probes were left for our analysis.

The coefficient of variation (CV) is used as a normalized measure of the dispersion of expression distribution (Maheshri and O'Shea 2007; Ansel *et al.* 2008; Ronnegard and Valdar 2011). The CV was computed as $CV = \sigma/\mu$, where σ and μ are the standard deviation and the mean of gene expression levels, respectively. LD block plots were obtained by using HaploView (Barrett *et al.* 2005).

Identification of evQTL using the dglm method

First we used the F–K test filter to greatly reduce the number of SNPs for computationally intensive model fitting. We then adapted the dglm method (Verbyla and Smyth 1998) to test for inequality in expression variances and measure the contribution of genetic variants to the expression heteroscedasticity. We considered the following model: $y_i = \mu + x_i\beta + g_i\alpha + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma^2\exp(g_i\theta))$, where y_i indicates a gene expression trait of individual i, g_i is the

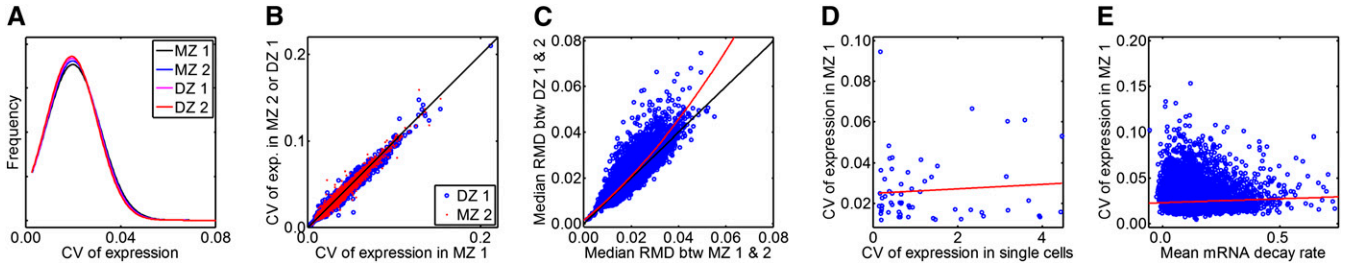

Figure 1 Distributions of expression variability in LCLs. (A) Distribution of CVs of gene expression (probe n = 35,078) measured in MZ and DZ twins. MZ 1 is the set of first pairs of all MZ twins and MZ 2 is the set of second pairs of all MZ twins. Similarly, DZ 1 is the set of first pairs of all DZ twins and DZ 2 is the set of second pairs of all DZ twins. (B) Scatter plot of CVs of gene expression (probe n = 35,078) in MZ 1 against those in MZ 2 (blue) or DZ 1 (red) cohorts. (C) Scatter plot of median RMD between pairs of MZ twins against median RMD between pairs of DZ twins. Each blue dot indicates a single expression probe (or a gene) and the position of the blue dot indicates the median value of RMD of expression between all MZ pairs (MZ 1 − MZ 2) on the x-axis and that between all DZ pairs (DZ 1 − DZ 2) on the y-axis. The red line is based on quadratic regression to show a more pronounced difference between MZ and DZ with greater RMD. (D) Scatter plot of CVs of gene expression (n = 59) in single cells against CVs of gene expression in MZ 1. (E) Scatter plot of mean mRNA decay rate against CVs of gene expression in the MZ 1 cohort. The red line is based on the linear regression.

genotype at the given SNP (encoded as 0, 1, or 2 for homozygous rare, heterozygous, and homozygous common alleles, respectively), ε_i is the residual with variance σ^2, and θ is the corresponding vector of coefficients of genotype g_i on the residual variance. Age of subjects and the batch of data collection were modeled as covariates x_i. With this full model, both mean and variance of expression y_i were controlled by SNP genotype g_i. We coded the fitting procedure using the dglm package in R. A snippet of R code for the dglm analysis is available in the supporting information of Ronnegard and Valdar (2011). We assumed that the input gene expression data were approximately normally distributed, conditional on the evQTL and covariates, and set family = Gaussian in the dglm R code to specify the error distribution and link function used. We tested for each input probe–SNP pair and obtained two P-values: $P_{\text{dispersion}}$ and P_{mean}, for the effects of genotypes on the variance and the mean of expression levels, respectively (Ronnegard and Valdar 2011). Probe–gene pairs that did not make the algorithm converge during computation were discarded. To control for the effect of outlier expression data points, permutation tests (Stranger *et al.* 2005) were conducted for all $P_{\text{dispersion}}$ significant pairs. Specifically, for each probe–SNP pair, we performed 10,000 permutations of expression phenotype relative to SNP genotypes. An association was considered significant if the P-value from the analysis of the observed $P_{\text{dispersion}}$ was lower than the threshold of the 0.001 tail of the distribution of the $P_{\text{dispersion}}$ from the 10,000 permutations ($P_{\text{permutation}} < 0.001$).

Single-cell expression and mRNA decay rate

Expression level of 96 genes was measured in 1440 single lymphoblastoid single cells by qPCR assays in another study (Livak *et al.* 2013). We used this data to compute the CV of expression of the same gene in different cells. The mRNA decay rates of 16,823 genes were estimated in 70 human LCLs (Pai *et al.* 2012). We obtained the mRNA decay rate data to compute the average mRNA decay rate for each gene among these LCL samples.

Identification of interacting SNPs

We used a two-step procedure to identify SNPs that may "interact" with evSNPs. Assuming an additive interaction between the SNP to be identified and an evSNP, we first partitioned individuals into L and S groups according to genotypes of the evSNP, which were associated with large (L) and small (S) variances of gene expression. Next we scanned genome-wide SNPs. For each SNP, we computed heterozygosity of the polymorphic site among individuals in L and S groups as $\text{Het}_{\text{L}} = 1 - \sum_{i=1}^{n} p_{i\_\text{L}}^2$ and $\text{Het}_{\text{S}} = 1 - \sum_{i=1}^{n} p_{i\_\text{S}}^2$, respectively, where P_i are allele frequencies of SNPs and, for diallelic SNPs, $n = 2$. All SNPs were then ranked by the value of $\text{Het}_{\text{L}} - \text{Het}_{\text{S}}$ and the top 100 SNPs with largest values were selected for further analysis. In the next step, a typical eQTL (not evQTL) analysis (Stranger *et al.* 2005) was conducted among individuals of the L group. In other words, for each top SNP with high genotype heterozygosity difference, a simple linear regression was performed between the SNP's genotypes and gene expression. The most significant SNPs were retained after applying a Bonferroni adjusted P-value cutoff = 0.05 and were reported as candidate interacting SNPs. To maintain sample independence, only one group of the twin sets was used in this analysis.

Results

Expression and genotype data

To investigate the genetic influences underlying variable gene expression, we revisited the published expression data (Grundberg *et al.* 2012) of the MuTHER (Multiple Tissue Human Expression Resource) project (Nica *et al.* 2011). In that study, gene expression was measured for LCL, adipose tissue (subcutaneous fat), and skin (tissue biopsies) using Illumina Human HT-12 V3 BeadChips. These tissues were sampled from a cohort of 856 female twins from the TwinsUK adult registry, including 154 monozygotic (MZ) twin pairs, 232 dizygotic (DZ) twin pairs and 84 singletons (Moayyeri *et al.* 2013). After quality control, expression data

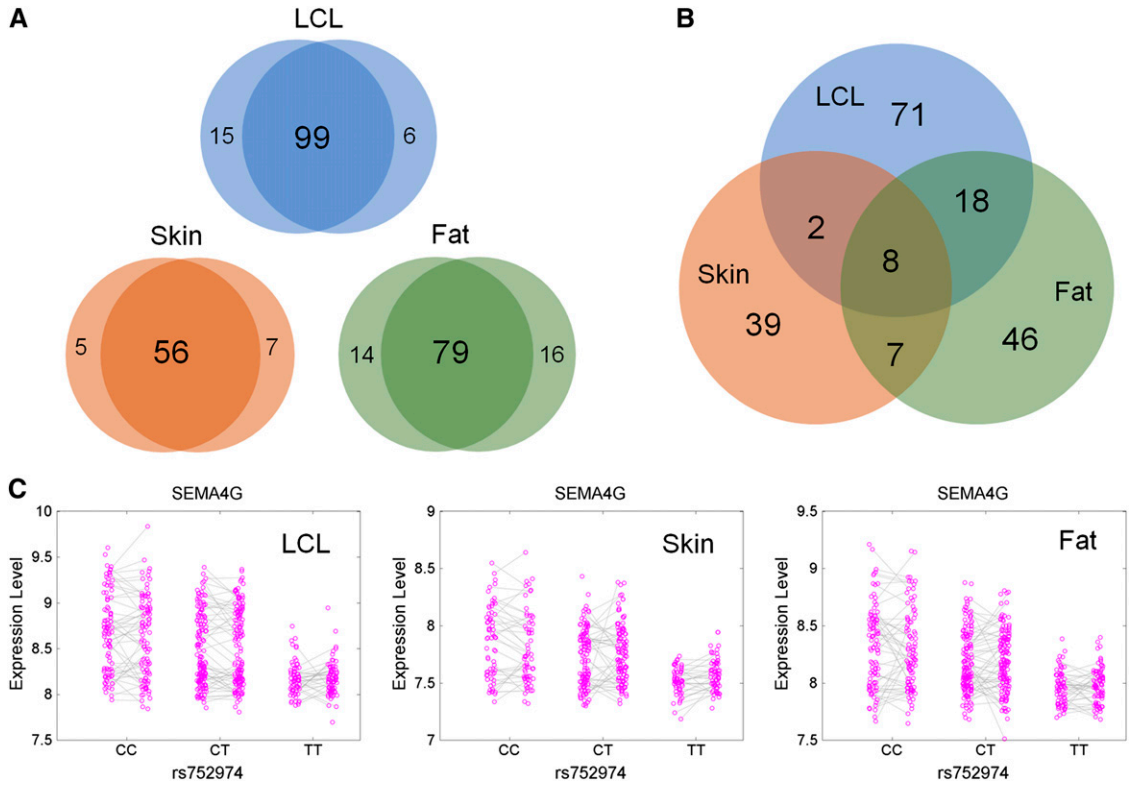

Figure 2 Numbers of evQTL in LCL, skin, and fat. (A) Venn diagrams of evQTL genes detected in two groups of twin sets. Each group of the twin sets is composed of one set of unrelated twin individuals. Overlapping areas of the Venn diagrams contain numbers of validated evQTL genes identified with both sets of twins. (B) Numbers correspond to evQTL genes within a subset of tissues. (C) One example of evQTL shared by all three tissues: evQTL at *SEMA4G*.

for 825 (adipose and LCL) and 705 (skin) individuals were retained (Grundberg *et al.* 2012). For each tissue, we downloaded the processed MuTHER expression data files deposited at ArrayExpression (http://www.ebi.ac.uk/arrayexpress/) using accession E-TABM-1140. The data were the quantile-normalized $\log_2$-transformed expression signals. Quantile normalization was performed first across the replicates of a single individual and then across all individuals as described in Grundberg *et al.* (2012). Along with the expression data, we also obtained the genotype data of this cohort (Grundberg *et al.* 2012). In our analysis, all available twin pairs with complete expression and genotype information were included, corresponding to 134 MZ and 195 DZ pairs with LCL profiles, 139 MZ and 188 DZ pairs with adipose profiles, and 105 MZ and 148 DZ pairs with skin profiles. Members of the TwinsUK cohort have health and lifestyle characteristics that are comparable to those of population singletons (Andrew *et al.* 2001). Because of this, we were able to use this cohort as a representative general population to investigate both genetic and nongenetic factors behind expression variability in this study.

Expression variability in the twin cohort

Here we present basic, descriptive statistics for expression data (independent of genotype information), with particular attention to disparities in gene expression among individuals. We chose to focus on the LCL data for this analysis, due to the availability of additional expression-related statistics (such as single-cell expression data and mRNA decay data).

We used the quantile-normalized $\log_2$-transformed expression data in all analyses throughout the article unless otherwise indicated. From this data, we first determined that expression values for most probes (n = 35,078) approximately fit the normal distribution: 97% of probes were with a skewness between −0.80 and 0.80 and a kurtosis of ∼3.0 (Supporting Information, Figure S1A); <7% of probes were rejected by Shapiro–Wilk test of normality with Bonferroni adjustment to the level of $\alpha = 0.01$. These justified the use of the Gaussian error distribution and link function in our dglm model (*Materials and Methods*). Retrospectively, we showed that the profile distributions for evQTL probes are approximately normal before and after Box–Cox transformation (Figure S1B).

To measure the level of dispersion of gene expression values, we computed the CV for each probe. The CVs ranged from 0.0024 [for ILMN_1765043 (*RPL38*)] to 0.2115 [for ILMN_1715169 (*HLA-DRB1*)], with a median of 0.0154. The distributions of CVs measured in subcohorts are indistinguishable from one another such as when comparing one group of MZ twin sets with the other (*i.e.*, MZ 1 *vs.* MZ 2) or comparing a group of MZ twin sets with a group of DZ twin sets (*e.g.*, MZ 1 *vs.* DZ 1) (Figure 1A). Probe data points are

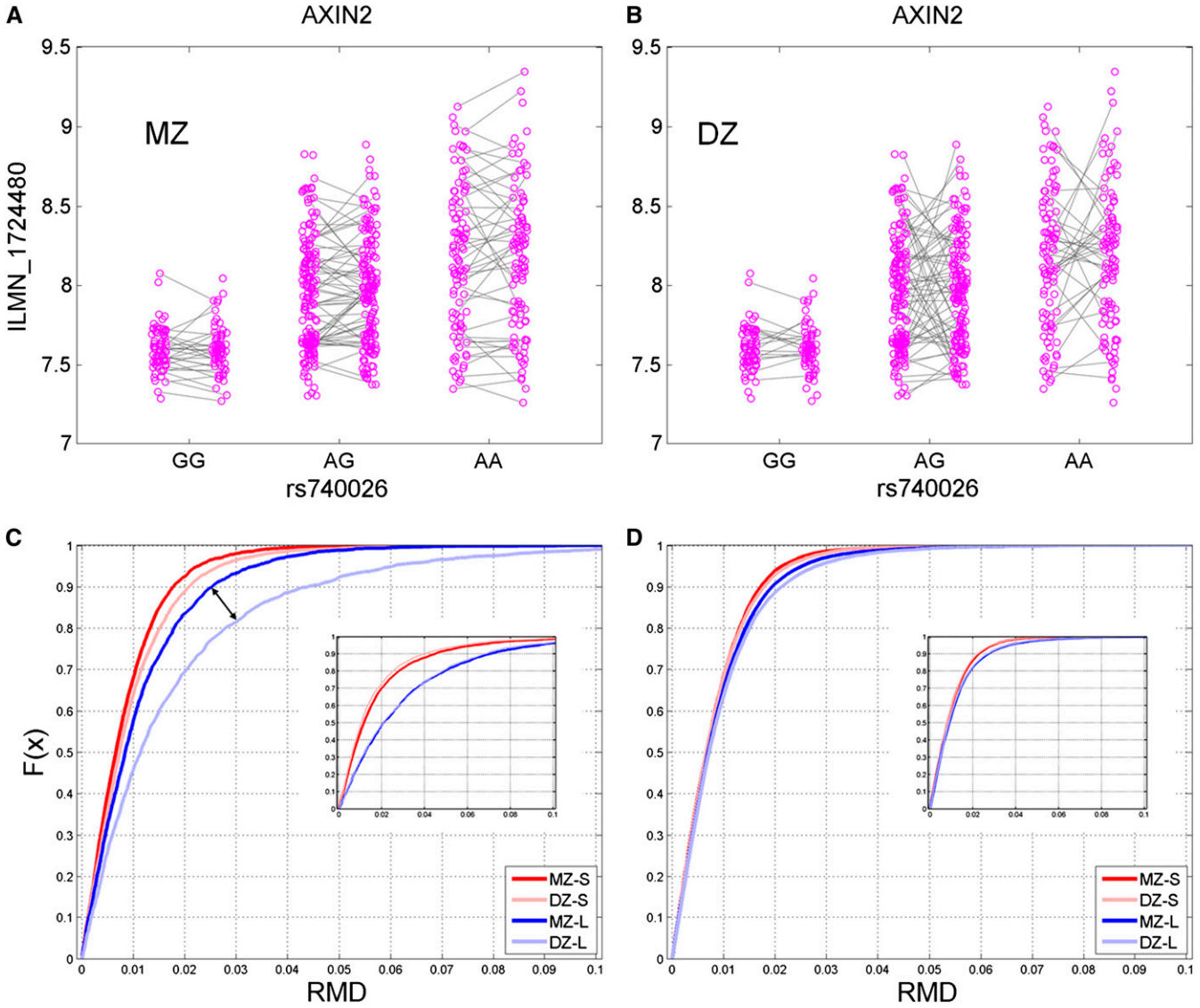

Figure 3 Dissection of genetic and nongenetic effects of evQTL using twins data. (A) The evQTL between *AXIN2* and rs740026. The expression data points from pairs of MZ twins are linked. (B) Same as A except that DZ twins are linked. (C) CDFs of RMD between twins classified into four groups, namely MZ-S, DZ-S, MZ-L, and DZ-L (see main text for definitions). The double arrow highlights the highly significant discrepancy in RMD distribution between MZ-L and DZ-L (K–S test, $P < 0.01$). Insert shows the same CDFs of RMD recomputed after randomly shuffling identities of corresponding MZ and DZ pairs. (D) Same as C except that data are randomly sampled from non-evQTL genes.

located along or close to the 1–1 diagonal line in the CV–CV scatter plot for the majority of probes, regardless of the CV–CV comparison between MZ 1 and MZ 2 or between MZ 1 and DZ 1 (Figure 1B). These results indicate that the extent and overall distribution of expression variability measured between individuals across different MZ and DZ cohorts are highly similar when all genes are taken into account.

Next, we measured expression differences between each pair of twins. For each probe, we computed the relative mean difference (RMD) in expression between MZ twin pairs and DZ twin pairs, separately. For a pair of MZ twins, for example, the RMD was computed using $\text{RMD} = ((1/2) \cdot |y_{\text{MZ1}} - y_{\text{MZ2}}|)/\bar{y}$, where $\bar{y}$ is the arithmetic mean of the expression levels, y_{MZ1} and y_{MZ2}, for the MZ twin pair. For most probes, the median RMD of expression between DZ pairs is larger than it is between MZ pairs, as indicated by the fact that most genes are located above the 1–1 diagonal line in the scatter plot (Figure 1C). That is to say, the normalized difference in gene expression between DZ pairs (DZ 1 and DZ 2) tends to be larger than that between MZ pairs (MZ 1 and MZ 2), suggesting that genetic factors influence expression variability for most of these genes.

To determine the influence of single-cell expression variability on population-level expression variability, we computed the CVs of expression for a selection of genes whose expression levels have been measured in single LCL cells (Livak *et al.* 2013). No correlation between the single-cell CVs and the between-individual CVs measured was

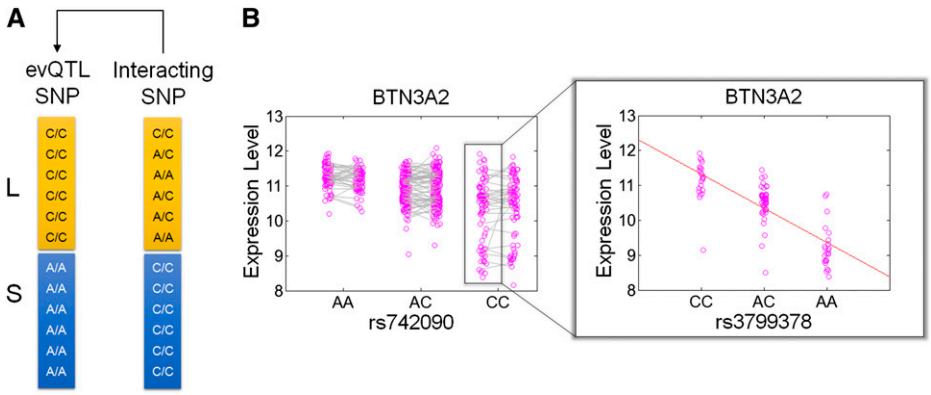

Figure 4 Schematic and example of an interacting SNP that helps the creation of an evQTL. (A) L indicates the group of individuals with evSNP genotype (C/C) associated with large variance in gene expression, while S indicates that with evSNP genotype (A/A) associated with small variance. The interacting SNP shows large genotype heterogeneity in the L group and small or no genotype heterogeneity in the S group. (B) Real example of evSNP rs742090 and interacting SNP rs3799378 at BTN3A2. Individuals with rs742090-CC genotype are further broken down by rs3799378 into three subgenotype groups, which are associated with different means of gene expression levels (shadowed panel).

detected for MZ 1 (Spearman's correlation test, $P = 0.21$, $n = 59$; Figure 1D). This suggests a limited contribution of single-cell expression variability (or transcriptional noise at the single-cell level) to the variability between individuals (or transcriptional noise at the population level).

Finally, we hypothesized that variable gene expression may be due to different mRNA decay rates for different genes. To test this, we used the mRNA decay rate data from the study of Pai *et al.* (2012). The correlation between mean mRNA decay rate and CV of expression among genes is not specific as shown by the opposite signs of two correlation coefficients: Spearman's $\rho = -0.027$ ($P = 0.00498$) and Pearson's $r = 0.044$ ($P = 4e\text{-}6$, $n = 11{,}083$; Figure 1E). Thus, gene expression variability showed no signs of correlation with the mRNA decay rates of genes.

Genetic variants underlying expression variability

To systematically assess the genetic influence on expression variability, we identified genome-wide evQTL using the method we previously established (Hulse and Cai 2013). We focused on *cis*-acting evQTL by limiting our search to those SNPs that flanked probes within 1.0 Mb on either side.

After filtering for quality control (*Materials and Methods*), a total of 35,078 probes were available for analysis. On average, each probe corresponded to 1212 SNPs in the 2-Mb *cis*-region (*i.e.*, 6 SNPs per 10 kb). For each SNP–probe pair, we conducted a three-step test to determine the evQTL relationship as described previously (Hulse and Cai 2013). Briefly, we first tested for the homogeneity of variances in gene expression among different genotype groups using the Fligner–Killeen (F–K) test (Fligner and Killeen 1976). Only those SNPs with a $P < 0.01$ (following Fraser and Schadt 2010) were carried on to the next step of analysis. We then applied the dglm method (Ronnegard and Valdar 2011) to each SNP–probe pair, ultimately computing $P_{\text{dispersion}}$ for a total of 1,251,611 SNP–probe pairs. To account for multiple tests performed between these probe and SNP pairs, we used the threshold of $P_{\text{dispersion}} < 1 \times 10^{-8}$, which is roughly equivalent to Bonferroni-adjusted $P < 0.01$, to assess the genome-wide significance. Finally, we conducted permutation tests for each significant SNP–probe pair to control for the influence of outlier data points on the dglm results (*Materials and Methods*). The detection of evQTL was performed independently for each of the two groups of twin sets. The assignment of individual twins to each group was random and did not influence the overall results. Each evQTL detected with one group of twin sets was then validated with the other group of twin sets to confirm its authenticity. For all three tissues, concordance was prevalent (Figure 2A) and the cases of discordance were mostly due to outliers present in one group of twins but not in the other group. The direction of effect (association with increased or decreased gene expression variability) was the same between the two groups of the twin sets for all evaluated SNPs.

A total of 99, 75, and 59 genes were identified and confirmed to have at least one validated *cis*-evQTL SNP (evSNP) in LCLs, fat, and skin, respectively (Table S1). Eight genes (corresponded to 4.2% of all unique evQTL genes) were shared in all three tissues (Figure 2B). One of these shared evQTL genes, *SEMA4G*, is given as an example to illustrate the consistent influence of genotypes on the variance of gene expression across the three tissues (Figure 2C). All evQTL shared across tissues showed the same directional effect, defined as either increasing or decreasing the variance of gene expression. That is to say, the directionality of some evQTL effects is not tissue- or cell-type specific. However, understanding regulatory variability in many different tissues might yield insights into the basic biological processes that influence tissue differentiation.

Given that many evQTL genes have more than one *cis*-evSNP, we examined the structure of haplotypes of these multiple *cis*-evSNPs. We found that *cis*-evSNPs of the same gene are likely to be located within the same LD block and that typically these blocks contained only a few prominent haplotypes (see Figure S2 for an example involving gene *ALG11*). This suggests that multiple evSNPs are likely to be linked with the same causal variant. We furthermore retrieved the ancestral allele information for SNPs from the 1000 Genomes Project Consortium (2012). The prediction of ancestral alleles was based on the phylogenetic trees constructed

Table 1 SNPs associated with gene expression variability and human complex traits

Gene (evSNP)	Tissue	GWAS complex trait	Risk allele	Reference
PAX8 (rs11123170)	LCL, fat, skin	Renal function-related traits (BUN)	rs11123170-G^{L}	Okada *et al.* (2012)
WDR41 (rs163030)	LCL, fat, skin	Caudate nucleus volume	rs163030-A^{L}	Stein *et al.* (2011)
HCG22 (rs2517532)	LCL	Hypothyroidism	rs2517532-G^{S}	Eriksson *et al.* (2012)
TBKBP1 (rs8070463)	LCL	Multiple sclerosis, ankylosing spondylitis	rs8070463-T^{S}	Patsopoulos *et al.* (2011)
			rs8070463-C^{L}	Evans *et al.* (2011)

^{L} and ^{S} indicate that individuals carrying homozygotic genotype of the risk allele have large and small variance in gene expression, respectively. GWAS, genome-wide association studies; LCL, lymphoblastoid cell line.

with sequences of human, chimpanzee, orangutan, and rhesus macaque. We found that ancestral alleles of evSNPs are more likely to be associated with smaller expression variability than derived alleles (Fisher's exact test: $P = 0.0036$, 0.022, and 0.036 for LCLs, skin, and fat, respectively).

Dissect genetic and nongenetic effects of evQTL

Twin data facilitated the dissection of the contributions of genetic and nongenetic factors. Variability measured between pairs of DZ twins is expected to be larger than that between pairs of MZ twins, as the phenotypic difference between DZ pairs may result from both genetic and environmental (nongenetic) effects while differences between genetically identical MZ pairs are attributable to the environment, assuming that the environments influencing MZ and DZ twin individuals are essentially identical. Figure 3 depicts the difference in expression level of evQTL gene *AXIN2* in three genotypes (GG, AG, and AA) defined by rs740026. Figure 3, A and B illustrate genotypes at rs740026 by linking the two data points for each twin pair by a straight line: Figure 3A shows genotype similarities between MZ twins, while in Figure 3B, similarities between DZ twin pairs are shown. Note that linkers between DZ twin pairs with different genotypes at the SNP site (*i.e.*, DZ 1 $\neq$ DZ 2) are not plotted. The expression difference between a pair of twins can be visually quantified by the slope of the straight line: a steeper line reflects a more dissimilar expression level between the twins. In the case of *AXIN2*, it is apparent that expression differences between DZ pairs tend to be larger than between MZ pairs. This is especially true for the AA genotype group, which shows a larger variance in expression between individuals.

For each evSNP and its associated genes in LCLs, we computed the RMD in gene expression between all pairs of MZ or DZ twins, as long as the genotypes of two individuals of the pair of twins were both identical to each other and homozygous at the SNP site. By definition, each evSNP allele is associated with either larger (L) or smaller (S) variance in gene expression. Thus, the RMD values (for evSNPs and associated genes) were separated according to whether homozygous genotypes defined by evSNPs were associated with L or S variances in gene expression. The cumulative distribution functions (CDFs) of these RMD values were plotted (Figure 3C). The curves were based on the RMD values calculated between all possible twin pairs for all evSNPs and genes and classified into four groups: MZ-S, MZ-L, DZ-S, and DZ-L. The MZ-S and DZ-S groups included pairs whose genotypes showed a small (S) amount of variance, while the MZ-L and DZ-L groups included pairs whose genotypes were associated with large (L) variances. In the end, the four groups—MZ-S, MZ-L, DZ-S, and DZ-L—contained 3629, 2548, 3825, and 2520 RMD values, respectively. Detailed statistics for the distributions of RMD values in each of these four groups are provided (Table S2). We found that CDF curves for the large-variance groups (MZ-L and DZ-L) were shifted toward the right compared to those for small-variance groups (MZ-S and DZ-S) [Kolmogorov–Smirnov (K–S) test, all $P < 10^{-5}$]. This indicated that the distribution of RMD between twin pairs (either MZ or DZ) in the large-variance groups was significantly different from that of the small-variance groups, with a larger RMD median for the large-variance groups. This difference (in RMD distribution between L and S groups) remained even when we randomly assigned the identities of MZ and DZ pairs (see insert in Figure 3C). Together, these results suggested that the increased discrepancy in gene expression between twin pairs (shown as a larger median RMD) contributed to the elevated variability in expression, which is true for both MZ and DZ twins. Because MZ twins are genetically identical, the increased RMD between MZ pairs was likely due to an increased sensitivity of gene expression to environmental factors.

More importantly, we found a significant discrepancy in distribution of RMD between MZ and DZ: DZ groups tended to have larger RMD values than MZ groups. This trend applied to both L and S groups, but was more obvious in the L group (all K–S tests, $P < 0.01$) (Figure 3C). These results suggested that the different genetic backgrounds resulted in a larger difference in gene expression between DZ twin pairs, which is more pronounced than that observed between MZ twin pairs.

For comparison, we randomly selected the same number of genes and *cis*-SNPs and conducted the same analysis of RMD distribution. There was no difference between CDFs of RMD in these non-evQTL genes in regard to either MZ or DZ twins, larger or smaller variance groups, as well as before or after shuffling of the twin identities. CDFs of all groups were more similar to each other (K–S test, all $P > 0.025$, except between MZ-S and DZ-L, $P = 2.9e\text{-}4$; Table S2, Figure 3D). That is to say, the influence of genetic and/or environmental effects on variable expression was not detected at the genomic level for all genes, but was limited to evQTL regions.

Finally, we repeated the CDF analyses using the RMD values computed from the Box–Cox normalized $\log_2$-transformed expression data, as well as using the absolute difference (instead of RMD) in gene expression. In both cases, our findings were highly similar to those obtained above (Figure S3), which supports the robustness of the results presented above.

Possible evQTL replicated by using RNA-seq data and SNPs of the 1000 Genomes Project

We obtained genotype data for fully sequenced samples of European ancestry (CEU) from the phase 1 release of the 1000 Genomes Project (1000 Genomes Project Consortium 2012), along with short reads from RNA-seq experiments in LCLs for these same individuals ($n = 43$) (Montgomery *et al.* 2010). After mapping the short reads, we estimated the expression level in fragments per kilobase of exon per million fragments (FPKM) (Trapnell *et al.* 2013) for all genes. For the same evQTL gene–SNP pairs detected in LCLs (*i.e.*, those included in Table S1), we plotted the relationships between genotype and FPKM value for each. Many evQTL relationships could be recognized by visual inspection (examples are presented in Figure S4), though none were statistically significant due to the small sample size. It is noteworthy that the algorithm for computing FPKM models the dispersion in a transcript's fragment count with a negative binomial distribution (Trapnell *et al.* 2013), which may introduce a relationship between the mean and variance of the count. The relationship should be taken into account in FPKM-based evQTL analyses.

Partially linked SNPs contribute to variable gene expression

Recent theoretical work showed that the within-genotype variance of a quantitative trait varies when a nonadditive genetic interaction or epistasis is present (Pare *et al.* 2010; Struchalin *et al.* 2010). Alternatively, the variance of a quantitative trait may be from the result of the interaction between genetic variants additively associated with the mean of the quantitative trait. To test this, we employed a two-step procedure to identify SNPs partially associated with (or interacting) with evSNPs through an incomplete haplotype structure (*Materials and Methods*). In an ideal scenario (Figure 4A), the genotype heterozygosity of the partially linked SNPs is large among individuals whose evSNP genotype is associated with larger expression variability (L group), while the genotype heterozygosity is small or equals zero among individuals whose evSNP genotype is associated with smaller expression variability (S group). If the interacting SNP is associated with the mean level of gene expression, then the association between the evSNP genotype and greater expression variability is likely due to the partial linkage between the evSNP and the interacting SNP.

Given these considerations, we performed a genome-wide search to identify a set of candidate interacting SNPs for each evQTL SNP and then used simple linear regression analysis to evaluate whether the potential interacting SNPs are significantly associated with gene expression among L-group individuals (*Materials and Methods*). For the 99 evQTL in LCLs, we identified 56 with at least one interacting SNP (Table S3). Among these interacting SNPs, 54 are located within the *cis*-region of the evSNPs with which they interact. Figure 4B presents one such relationship between evSNP rs742090 and interacting SNP rs3799378, both at *BTN3A2*. Individuals with a CC genotype for evSNP rs742090 were further sorted by rs3799378 genotypes. Clearly, the expression level of *BTN3A2* in individuals with the rs742090-CC genotype is significantly influenced by rs3799378 genotypes. The increased variability in gene expression showed in individuals with rs742090-CC genotype is caused by the heterogeneity of rs3799378 genotypes. These results suggest that local haplotype structure between SNPs contributed to the creation of evQTL.

Linking evQTL with complex disease phenotypes

Several studies have utilized eQTL data to interpret the discoveries of association studies of complex traits (Emilsson *et al.* 2008; Nica *et al.* 2010; Nicolae *et al.* 2010). Along this same vein, we identified evQTL associated with complex traits from the catalog of published genome-wide association studies (GWAS) (http://www.genome.gov/gwastudies/). From the results of these GWAS, we identified 61 reported genes as evQTL genes (Table S4). In four cases, the exact same SNP was found to be both an evSNP and a marker SNP associated with risk or susceptibility to the complex trait (Table 1). Intriguingly, the "T" allele of rs8070463, associated with smaller expression variability of *TBKBP1*, is a reported risk allele in multiple sclerosis (Patsopoulos *et al.* 2011), while the "C" allele for this same SNP, associated with larger expression variability, is linked with risk for ankylosing spondylitis (Evans *et al.* 2011).

Discussion

There is empirical evidence across several species that the variance among phenotypes is genotype dependent (Ansel *et al.* 2008; Wolc *et al.* 2009; Hill and Mulder 2010; Jimenez-Gomez *et al.* 2011). Understanding genetic control of phenotypic variability has become increasingly important in evolutionary biology, human medicine, the agricultural industry and other branches of biological science (Gibson 2009; Yang *et al.* 2012). Despite the importance, few research programs focus on genetic variants associated with trait variance, while studies of trait averages abound. Recently, a powerful statistical framework based on the dglm model has been developed for studying phenotypic variability of complex traits (Ronnegard and Valdar 2011). Given that gene expression is a complex trait with highly variable and heritable patterns (Stranger *et al.* 2005; Williams *et al.* 2007; Montgomery and Dermitzakis 2011), we have previously adapted the dglm method to investigate genetic variants controlling expression variability (Hulse and Cai 2013).

In this study, we further investigated the relative contributions of genetic and nongenetic (environmental) factors

to expression variability and the roles of these factors in the formation of evQTL. We started by exploring basic gene expression statistics measured in the TwinsUK cohort. For all genes, expression level dispersions were highly similar in and between both MZ and DZ twins. No correlations with expression variability were detected when compared between individuals, between single cells, or relative to the average mRNA decay rate, highlighting the marked discrepancies in variability measured at population and molecular levels. Further results showed that the discordance in expression between each pair of DZ twins was more pronounced than that between MZ twins, implying that the increased amount of genetic variation between DZ twins influences expression variability. Next, we systematically identified *cis*-acting evQTL in three tissues of the TwinsUK cohort. Twin data greatly facilitated the validation of detected evQTL and revealed overall robust signals that would otherwise not be appreciable in studies of nontwin design. Focusing on the detected evQTL, we showed that the discordance in expression between DZ pairs was larger than that between MZ pairs, and further showed that the discordance in expression between MZ pairs whose genotypes were associated with large expression variability was significantly larger than that between MZ pairs whose genotypes were associated with small expression variability. It is intriguing to find that the phenotypic discordance remained even in the absence of genetic variation between MZ twins. This might be explained by incomplete penetrance of mutations, which is frequent in isogenic model organisms in homogeneous environments (Horvitz and Sulston 1980; Gartner 1990). This might also be epigenetic: for example, DNA methylation, which can be influenced by environmental factors such as diet and lifestyle, is known to affect gene expression (Badano and Katsanis 2002; Baranzini *et al.* 2010). Lastly, much to our surprise, we found that more than half of evQTL could be explained by a conceptually simple scenario in which the evSNP was occasionally associated with a nearby SNP that influenced gene expression both additively and independently. We suspect there should be many different ways of nonepistatic interaction between two or more genetic variants, such as the mode of partial linkage we have described here, giving rise to genotype-dependent expression variances. That is to say, the majority of phenotypic variability across individuals might be explained without invoking epistasis (Hill *et al.* 2008; Powell *et al.* 2013).

In light of our new findings, several related considerations are discussed below.

Methodological considerations for studying phenotypic variability

The procedure we used for identifying evQTL (*Materials and Methods* and Hulse and Cai 2013) consisted of three steps. First, the F–K test was applied to test for the heterogeneity of variances of gene expression between different genotypes and identify corresponding SNPs. Next, the dglm method was applied to the selected SNPs. The significant results of the dglm test were then subjected to permutation tests to reduce the influence of outliers in the data. This procedure is less likely to be susceptible to issues related to multiple testing and outliers in input data, though a formal assessment of its statistical power remains to be done.

Statistical methods, including Levene's test (Pare *et al.* 2010; Struchalin *et al.* 2010), squared residual value linear modeling (SVLM) (Struchalin *et al.* 2012), and the methods by Yang *et al.* (2011), have been applied in studying phenotypic variability (see review in Geiler-Samerotte *et al.* 2013). As a full parametric approach, the dglm method (Ronnegard and Valdar 2011) has several advantages. For example, it accounts for the uncertainty of fitted parameters for both the mean and the variance aspects of the model and also allows fitting of covariates (Ronnegard and Valdar 2012); it is also highly flexible, allowing for any response distribution from the exponential family (Smyth 2002) (such as binomial, Poisson, or gamma) to be modeled (see section 2 in the Supporting Theory of Ronnegard and Valdar 2011 for a sample of modeling the gamma-distributed traits).

Given the flexibility of the dglm method, we acknowledge that the results of our evQTL analysis are likely to be dependent on how the dglm analysis was set up. For this study, we adapted the Gaussian error distribution and link function because no significant departure from normality was found in the expression data. The effect of different methods of normalization on statistical interpretation of gene expression remains subject to careful scrutiny (Bolker *et al.* 2009; Qin *et al.* 2012; Geiler-Samerotte *et al.* 2013). For example, normalizations may perturb the covariance structure of input data or change the scale of the resulting data. Thus, the impacts of different methods of data transformation and normalization should be carefully considered in future studies involving evQTL analysis. Finally, we acknowledge that the dglm analysis described in this article may be influenced by the scale effect (*e.g.*, mean–variance relationship). It is not uncommon for trait variance to change with trait mean, often causing trait skewness. If this occurs, any SNP associated with a large increase in mean expression would also be associated with an increase in variability (Ronnegard and Valdar 2011). Analyses studying a specific phenotype and/or with a more narrowly targeted focus than that of the broad-based study described in this article should employ a more conservative approach in which QTL associated strictly with variance (*i.e.*, those affecting only variability and not the mean) are identified, using the procedure proposed by Ronnegard and Valdar (2011).

Additive vs. epistatic effect of genotypes on phenotypic variation in a population

Quantitative geneticists partition the genetic effect on phenotypic variation between individuals into additive, dominance, and epistatic components. The additive component describes the variance associated with the independent contributions of alleles, while dominance describes the variance contributed by interactions between alleles at the

same locus, and epistasis refers to the contribution of interactions between alleles at different loci. For most complex traits, quantitative genetic theory (Hill *et al.* 2008; Crow 2010) suggests that epistasis is unlikely to contribute substantially to the between-individual variation. That is to say, most of the variation in a population will be due to the additive effects of specific allelic combinations. Yet this assertion is not without controversy. The results of empirical linkage mapping and association studies suggest that epistasis seems to explain considerable variation in complex trait characteristics within natural populations (Carlborg and Haley 2004; Zuk *et al.* 2012).

Our results showed that >50% of evQTL can be explained by a partial linkage (*i.e.*, partial association between haplotypes) of the evQTL SNP and another SNP nearby. Our interacting SNP analysis only considered a simplistic scenario of the association. There are many other possible ways of partial associations in which SNPs interact. For example, consider the genotyped SNP "A/a" and the causative expression QTL "Q/q," with only three haplotypes segregating in the population: AQ, aQ, and aq (as would occur if the novel "q" allele arose on the "a" haplotype). Then the "a" SNP allele would be associated with a changed trait mean and a higher trait variance as the expression QTL segregates within that genotype. If we could take all possible partial associations into account, we would anticipate that even more evQTL could be explained by the effect of partial linkage, rather than epistasis. We therefore conclude that much variance in a quantitative trait may be explained by partial linkage between local genetic variants, each additively associated with the trait. Our view is supported by the results of recent studies. Powell *et al.* (2013) conducted a gene expression study using blood samples from 862 individuals from nuclear families containing MZ or DZ twin pairs, using both pedigree and genotype information. They found that the genetic architecture of gene expression is predominantly additive, with a minority of transcripts displaying nonadditive effects. Hill *et al.* (2008) evaluated the evidence from empirical studies of genetic variance components and found that additive variance typically accounts for over half and often close to 100% of the total genetic variance.

Detecting evQTL as a shortcut for detecting epistasis?

Detection of the variance of a quantitative trait in genetic association studies is thought to increase knowledge about the interaction between genetic variants. More specifically, detecting variability QTL (*e.g.*, evQTL) is considered to be a shortcut for detecting genetic interactions (Ronnegard and Valdar 2011, 2012). So far, many methods for detecting genetic interactions are based on testing for different variances of phenotype between genotypes, with the underlying assumption that the variance of a quantitative trait is likely to differ under the influence of epistasis (Pare *et al.* 2010; Ronnegard and Valdar 2011). However, our new discovery that evQTL are formed due to the partial haplotype association between SNPs refutes this assumption. As stated above, more than half (and probably many more) of the evQTL identified could be explained by partial linkage between SNPs with additive effects. Both additive and epistatic effects can result in increased phenotypic variation (as schematically illustrated in Figure S5). Merely detecting the variance of a quantitative trait cannot in itself distinguish between the additive and epistatic effects; thus, no specific conclusions can be made. The relationship between partially associated SNPs, each additively associated with phenotypic variation, needs to be integrated more carefully in the study of phenotypic variability. Thus, the variance of a quantitative trait should not serve as a hallmark of genetic interaction or epistasis.

Phenotypic variability and implications in complex traits and diseases

High-throughput sequencing and genotyping technologies have spurred an increasing number of studies detecting genotype–phenotype relationships and mapping in complex, polygenic traits and human diseases (Hindorff *et al.* 2009). The remarkable success of GWAS is accompanied by the issue of the "missing heritability" (Manolio *et al.* 2009), namely the fact that the trait-associated SNPs identified through GWAS often account for only a small proportion of the observed correlations in phenotype between relatives. The reason behind this issue has been thought to be that additional genetic factors remain to be found, and that the presence of epistasis is a particular cause for concern (Carlborg and Haley 2004; Moore and Williams 2009; Ueki and Cordell 2012). In reality, if the effect of one locus is altered or masked by effects at another locus, power to detect the first locus is likely to be reduced and elucidation of the joint effects at the two loci will be hindered by their interaction. Consequently, a large amount of research has been devoted to the detection and investigation of epistatic interactions; a number of methods for detecting the interaction between SNPs have been proposed (Pare *et al.* 2010; Struchalin *et al.* 2010; Shang *et al.* 2011; Daye *et al.* 2012; Struchalin *et al.* 2012), yet there has been much confusion in the literature over definitions and interpretations of epistasis (Cordell 2002).

This study, along with other studies (Hill *et al.* 2008; Powell *et al.* 2013), have clearly pointed out that a detailed investigation of local haplotype structure between SNPs at the same locus is necessary to reveal their combined influences on phenotypes of complex traits. For example, we have identified a list of evSNPs that are also associated with human complex traits (see Table 1). Further investigations on partial linkage between SNPs that may influence these traits should be performed. The same should also be done for the *FTO* (fat mass and obesity associated) gene locus whose genotype is associated with phenotypic variability of body mass index (Yang *et al.* 2012).

Finally, we point out that an interaction detected via statistical models is different from the biological interaction (Phillips 1998; Cordell 2002; Wang *et al.* 2010). The lack of direct correspondence between statistical and biological

interactions makes it difficult to make strong inferences concerning biological mechanisms based on interaction terms from a statistical model (Ueki and Cordell 2012). Therefore, detection of statistical interaction merely provides a good starting point for a more focused investigation of the joint involvement of the relevant factors, which can perhaps be better addressed through other types of experimental data. Our findings suggest that there is a lot that can be done at the statistical level to prioritize those loci that are most likely to produce significant experimental results.

Conclusions

In conclusion, we used evQTL as a statistical model system for studying phenotypic variability and dissected the genetic and nongenetic effects by using the twin data. Our findings concerning evQTL offer new insights into relative contribution of genetic and environmental factors in the formation of evQTL. Dissecting the genetic components underlying phenotypic variability into additive and epistatic effects allowed the dominant role of additive effect to be revealed.

Acknowledgments

We thank the TwinUK study for producing the datasets, Ken Livak for sharing the single-cell expression data, and the Whole Systems Genomics Initiative of Texas A&M University for providing computational resources. The TwinUK study was funded by the Wellcome Trust and the European Community's Seventh Framework Programme (FP7/2007–2013). The study also received support from the National Institute for Health Research (NIHR)'s Clinical Research Facility at Guy's and St. Thomas' National Health Service (NHS) Foundation Trust and NIHR's Biomedical Research Centre based at Guy's and St. Thomas' NHS Foundation Trust and King's College London. SNP genotyping was performed by the Wellcome Trust Sanger Institute and the National Eye Institute via National Institutes of Health/Computerised Infectious Disease Reporting (CIDR).

Literature Cited

Acar, M., J. T. Mettetal, and A. van Oudenaarden, 2008 Stochastic switching as a survival strategy in fluctuating environments. Nat. Genet. 40: 471–475.

Andrew, T., D. J. Hart, H. Snieder, M. de Lange, T. D. Spector *et al.*, 2001 Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. Twin Res. 4: 464–477.

Ansel, J., H. Bottin, C. Rodriguez-Beltran, C. Damon, M. Nagarajan *et al.*, 2008 Cell-to-cell stochastic variation in gene expression is a complex genetic trait. PLoS Genet. 4: e1000049.

Badano, J. L., and N. Katsanis, 2002 Beyond Mendel: an evolving view of human genetic disease transmission. Nat. Rev. Genet. 3: 779–789.

Bahar, R., C. H. Hartmann, K. A. Rodriguez, A. D. Denny, R. A. Busuttil *et al.*, 2006 Increased cell-to-cell variation in gene expression in ageing mouse heart. Nature 441: 1011–1014.

Baranzini, S. E., J. Mudge, J. C. van Velkinburgh, P. Khankhanian, I. Khrebtukova *et al.*, 2010 Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. Nature 464: 1351–1356.

Barrett, J. C., B. Fry, J. Maller, and M. J. Daly, 2005 Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21: 263–265.

Bolker, B. M., M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen *et al.*, 2009 Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol. Evol. 24: 127–135.

Carlborg, O., and C. S. Haley, 2004 Epistasis: Too often neglected in complex trait studies? Nat. Rev. Genet. 5: 618–625.

Cheverud, J. M., and E. J. Routman, 1995 Epistasis and its contribution to genetic variance components. Genetics 139: 1455–1461.

Choy, E., R. Yelensky, S. Bonakdar, R. M. Plenge, R. Saxena *et al.*, 2008 Genetic analysis of human traits in vitro: drug response and gene expression in lymphoblastoid cell lines. PLoS Genet. 4: e1000287.

Cordell, H. J., 2002 Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Hum. Mol. Genet. 11: 2463–2468.

Crow, J. F., 2010 On epistasis: why it is unimportant in polygenic directional selection. Philos. Trans. R. Soc. Lond. B Biol. Sci. 365: 1241–1244.

Daye, Z. J., J. Chen, and H. Li, 2012 High-dimensional heteroscedastic regression with an application to eQTL data analysis. Biometrics 68: 316–326.

Emilsson, V., G. Thorleifsson, B. Zhang, A. S. Leonardson, F. Zink *et al.*, 2008 Genetics of gene expression and its effect on disease. Nature 452: 423–428.

Eriksson, N., J. Y. Tung, A. K. Kiefer, D. A. Hinds, U. Francke *et al.*, 2012 Novel associations for hypothyroidism include known autoimmune risk loci. PLoS ONE 7: e34442.

Evans, D. M., C. C. A. Spencer, J. J. Pointon, Z. Su, D. Harvey *et al.*, 2011 Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. Nat. Genet. 43: 761–767.

Fligner, M. A., and T. J. Killeen, 1976 Distribution-free 2-sample tests for scale. J. Am. Stat. Assoc. 71: 210–213.

Fraser, H. B., and E. E. Schadt, 2010 The quantitative genetics of phenotypic robustness. PLoS ONE 5: e8635.

Gartner, K., 1990 A third component causing random variability beside environment and genotype. A reason for the limited success of a 30 year long effort to standardize laboratory animals? Lab. Anim. 24: 71–77.

Geiler-Samerotte, K., C. Bauer, S. Li, N. Ziv, D. Gresham *et al.*, 2013 The details in the distributions: why and how to study phenotypic variability. Curr. Opin. Biotechnol. 24: 752–759.

Gibson, G., 2009 Decanalization and the origin of complex disease. Nat. Rev. Genet. 10: 134–140.

Grundberg, E., K. S. Small, A. K. Hedman, A. C. Nica, A. Buil *et al.*, 2012 Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat. Genet. 44: 1084–1089.

Hallgrímsson, B., and B. K. Hall, 2005 *Variation*. Academic Press, Amsterdam.

Hill, W. G., and X. S. Zhang, 2004 Effects on phenotypic variability of directional selection arising through genetic differences in residual variability. Genet. Res. 83: 121–132.

Hill, W. G., and H. A. Mulder, 2010 Genetic analysis of environmental variation. Genet. Res. 92: 381–395.

Hill, W. G., M. E. Goddard, and P. M. Visscher, 2008 Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet. 4: e1000008.

Hindorff, L. A., P. Sethupathy, H. A. Junkins, E. M. Ramos, J. P. Mehta *et al.*, 2009 Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl. Acad. Sci. USA 106: 9362–9367.

Horvitz, H. R., and J. E. Sulston, 1980 Isolation and genetic characterization of cell-lineage mutants of the nematode Caenorhabditis elegans. Genetics 96: 435–454.
Hulse, A. M., and J. J. Cai, 2013 Genetic variants contribute to gene expression variability in humans. Genetics 193: 95–108.
Jimenez-Gomez, J. M., J. A. Corwin, B. Joseph, J. N. Maloof, and D. J. Kliebenstein, 2011 Genomic analysis of QTLs and genes altering natural variation in stochastic noise. PLoS Genet. 7: e1002295.
Kaern, M., T. C. Elston, W. J. Blake, and J. J. Collins, 2005 Stochasticity in gene expression: from theories to phenotypes. Nat. Rev. Genet. 6: 451–464.
Kemkemer, R., S. Schrank, W. Vogel, H. Gruler, and D. Kaufmann, 2002 Increased noise as an effect of haploinsufficiency of the tumor-suppressor gene neurofibromatosis type 1 in vitro. Proc. Natl. Acad. Sci. USA 99: 13783–13788.
Livak, K. J., Q. F. Wills, A. J. Tipping, K. Datta, R. Mittal *et al.*, 2013 Methods for qPCR gene expression profiling applied to 1440 lymphoblastoid single cells. Methods 59: 71–79.
Maheshri, N., and E. K. O'Shea, 2007 Living with noisy genes: how cells function reliably with inherent variability in gene expression. Annu. Rev. Biophys. Biomol. Struct. 36: 413–434.
Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff *et al.*, 2009 Finding the missing heritability of complex diseases. Nature 461: 747–753.
Moayyeri, A., C. J. Hammond, D. J. Hart, and T. D. Spector, 2013 The UK Adult Twin Registry (TwinsUK Resource). Twin Res. Hum. Genet. 16: 144–149.
Montgomery, S. B., and E. T. Dermitzakis, 2011 From expression QTLs to personalized transcriptomics. Nat. Rev. Genet. 12: 277–282.
Montgomery, S. B., M. Sammeth, M. Gutierrez-Arcelus, R. P. Lach, C. Ingle *et al.*, 2010 Transcriptome genetics using second generation sequencing in a Caucasian population. Nature 464: 773–777.
Moore, J. H., and S. M. Williams, 2009 Epistasis and its implications for personal genetics. Am. J. Hum. Genet. 85: 309–320.
Nica, A. C., S. B. Montgomery, A. S. Dimas, B. E. Stranger, C. Beazley *et al.*, 2010 Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. PLoS Genet. 6: e1000895.
Nica, A. C., L. Parts, D. Glass, J. Nisbet, A. Barrett *et al.*, 2011 The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. PLoS Genet. 7: e1002003.
Nicolae, D. L., E. Gamazon, W. Zhang, S. Duan, M. E. Dolan *et al.*, 2010 Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet. 6: e1000888.
Okada, Y., X. Sim, M. J. Go, J. Y. Wu, D. F. Gu *et al.*, 2012 Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. Nat. Genet. 44: 904.
Pai, A. A., C. E. Cain, O. Mizrahi-Man, S. De Leon, N. Lewellen *et al.*, 2012 The contribution of RNA decay quantitative trait loci to inter-individual variation in steady-state gene expression levels. PLoS Genet. 8: e1003000.
Pare, G., N. R. Cook, P. M. Ridker, and D. I. Chasman, 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.
Patsopoulos, N. A., Bayer Pharma MS Genetics Working Group; Steering Committees of Studies Evaluating IFNβ-1b and a CCR1-Antagonist; ANZgene Consortium; GeneMSA *et al.*, 2011 Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. Ann. Neurol. 70: 897–912.
Perry, G. M. L., K. W. Nehrke, D. A. Bushinsky, R. Reid, K. L. Lewandowski *et al.*, 2012 Sex modifies genetic effects on residual variance in urinary calcium excretion in rat (Rattus norvegicus). Genetics 191: 1003–1013.
Phillips, P. C., 1998 The language of gene interaction. Genetics 149: 1167–1171.
Pickrell, J. K., J. C. Marioni, A. A. Pai, J. F. Degner, B. E. Engelhardt *et al.*, 2010 Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature 464: 768–772.
Powell, J. E., A. K. Henders, A. F. McRae, J. Kim, G. Hemani *et al.*, 2013 Congruence of additive and non-additive effects on gene expression estimated from pedigree and SNP data. PLoS Genet. 9: e1003502.
Qin, S., J. Kim, D. Arafat, and G. Gibson, 2012 Effect of normalization on statistical and biological interpretation of gene expression profiles. Front Genet 3: 160.
Queitsch, C., T. A. Sangster, and S. Lindquist, 2002 Hsp90 as a capacitor of phenotypic variation. Nature 417: 618–624.
Ramasamy, A., D. Trabzuni, J. R. Gibbs, A. Dillman, D. G. Hernandez *et al.*, 2013 Resolving the polymorphism-in-probe problem is critical for correct interpretation of expression QTL studies. Nucleic Acids Res. 41: e88.
Ronnegard, L., and W. Valdar, 2011 Detecting major genetic loci controlling phenotypic variability in experimental crosses. Genetics 188: 435–447.
Ronnegard, L., and W. Valdar, 2012 Recent developments in statistical methods for detecting genetic loci affecting phenotypic variability. BMC Genet. 13: 63.
Shang, J., J. Zhang, Y. Sun, D. Liu, D. Ye *et al.*, 2011 Performance analysis of novel methods for detecting epistasis. BMC Bioinformatics 12: 475.
Shen, X., M. Pettersson, L. Ronnegard, and O. Carlborg, 2012 Inheritance beyond plain heritability: variance-controlling genes in Arabidopsis thaliana. PLoS Genet. 8: e1002839.
Smyth, G. K., 2002 An efficient algorithm for REML in heteroscedastic regression. J. Comput. Graph. Statist. 11: 836–847.
Spector, T. D., and F. M. K. Williams, 2006 The UK Adult Twin Registry (TwinsUK). Twin Res. Hum. Genet. 9: 899–906.
Stein, J. L., D. P. Hibar, S. K. Madsen, M. Khamis, K. L. McMahon *et al.*, 2011 Discovery and replication of dopamine-related gene effects on caudate volume in young and elderly populations (N=1198) using genome-wide search. Mol. Psychiatry 16: 927–937.
Stranger, B. E., M. S. Forrest, A. G. Clark, M. J. Minichiello, S. Deutsch *et al.*, 2005 Genome-wide associations of gene expression variation in humans. PLoS Genet. 1: e78.
Stranger, B. E., A. C. Nica, M. S. Forrest, A. Dimas, C. P. Bird *et al.*, 2007 Population genomics of human gene expression. Nat. Genet. 39: 1217–1224.
Struchalin, M. V., A. Dehghan, J. C. Witteman, C. van Duijn, and Y. S. Aulchenko, 2010 Variance heterogeneity analysis for detection of potentially interacting genetic loci: method and its limitations. BMC Genet. 11: 92.
Struchalin, M. V., N. Amin, P. H. Eilers, C. M. van Duijn, and Y. S. Aulchenko, 2012 An R package "VariABEL" for genome-wide searching of potentially interacting loci by testing genotypic variance heterogeneity. BMC Genet. 13: 4.
Teo, Y. Y., M. Inouye, K. S. Small, R. Gwilliam, P. Deloukas *et al.*, 2007 A genotype calling algorithm for the Illumina BeadArray platform. Bioinformatics 23: 2741–2746.
The 1000 Genomes Project Consortium, 2012 An integrated map of genetic variation from 1,092 human genomes. Nature 491: 56–65.
Trapnell, C., D. G. Hendrickson, M. Sauvageau, L. Goff, J. L. Rinn *et al.*, 2013 Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. 31: 46–53.
Ueki, M., and H. J. Cordell, 2012 Improved statistics for genome-wide interaction analysis. PLoS Genet. 8: e1002625.
Verbyla, A. P., and G. K. Smyth, 1998 Double generalized linear models: approximate residual maximum likelihood and diag-

nostics, pp. 1–15 in *Research Report*. Department of Statistics, University of Adelaide, Adelaide, Australia.
Wagner, G., 1995 Adaptation and the modular design of organisms, pp. 315–328 in *Advances in Artificial Life*, edited by F. Morán, A. Moreno, J. Merelo, and P. Chacón. Springer-Verlag, Berlin.
Wagner, G. P., and L. Altenberg, 1996 Complex adaptations and the evolution of evolvability. Evolution 51: 967–976.
Wang, X., R. C. Elston, and X. Zhu, 2010 The meaning of interaction. Hum. Hered. 70: 269–277.
Wang, Z., and J. Zhang, 2011 Impact of gene expression noise on organismal fitness and the efficacy of natural selection. Proc. Natl. Acad. Sci. USA 108: E67–E76.
Williams, R. B., E. K. Chan, M. J. Cowley, and P. F. Little, 2007 The influence of genetic variation on gene expression. Genome Res. 17: 1707–1716.
Wolc, A., I. M. White, S. Avendano, and W. G. Hill, 2009 Genetic variability in residual variation of body weight and conformation scores in broiler chickens. Poult. Sci. 88: 1156–1161.
Yang, J., R. J. Loos, J. E. Powell, S. E. Medland, E. K. Speliotes *et al.*, 2012 FTO genotype is associated with phenotypic variability of body mass index. Nature 490: 267–272.
Yang, Y., O. F. Christensen, and D. Sorensen, 2011 Use of genomic models to study genetic control of environmental variance. Genet. Res. 11: 1–14.
Zhang, Z., W. Qian, and J. Zhang, 2009 Positive selection for elevated gene expression noise in yeast. Mol. Syst. Biol. 5: 299.
Zuk, O., E. Hechter, S. R. Sunyaev, and E. S. Lander, 2012 The mystery of missing heritability: Genetic interactions create phantom heritability. Proc. Natl. Acad. Sci. USA 109: 1193–1198.

Communicating editor: E. Petretto

GENETICS

Supporting Information
http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.157503/-/DC1

Additive, Epistatic, and Environmental Effects Through the Lens of Expression Variability QTL in a Twin Cohort

Gang Wang, Ence Yang, Candice L. Brinkmeyer-Langford, and James J. Cai

DOI: 10.1534/genetics.113.157503

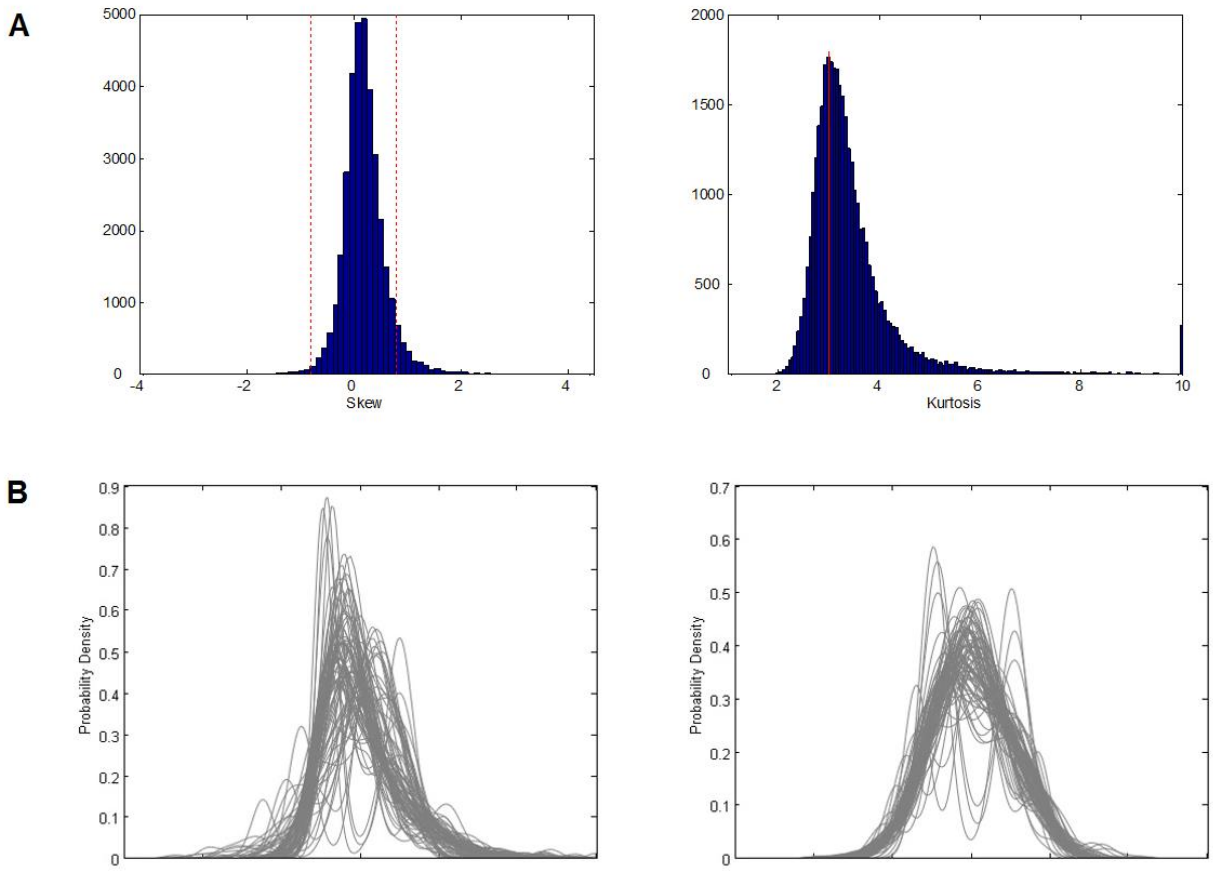

Figure S1 Normality of expression data measured in LCLs. (**A**) Distributions of skewness and kurtosis. Red dashed lines indicate -0.8 and +0.8 skewness; red solid line indicates kurtosis = 3. (**B**) Profile distributions of expression data for selected probes (i.e., probes involved in evQTLs). (Left) Quantile-normalized expression data; (Right) Box-Cox normalized expression data.

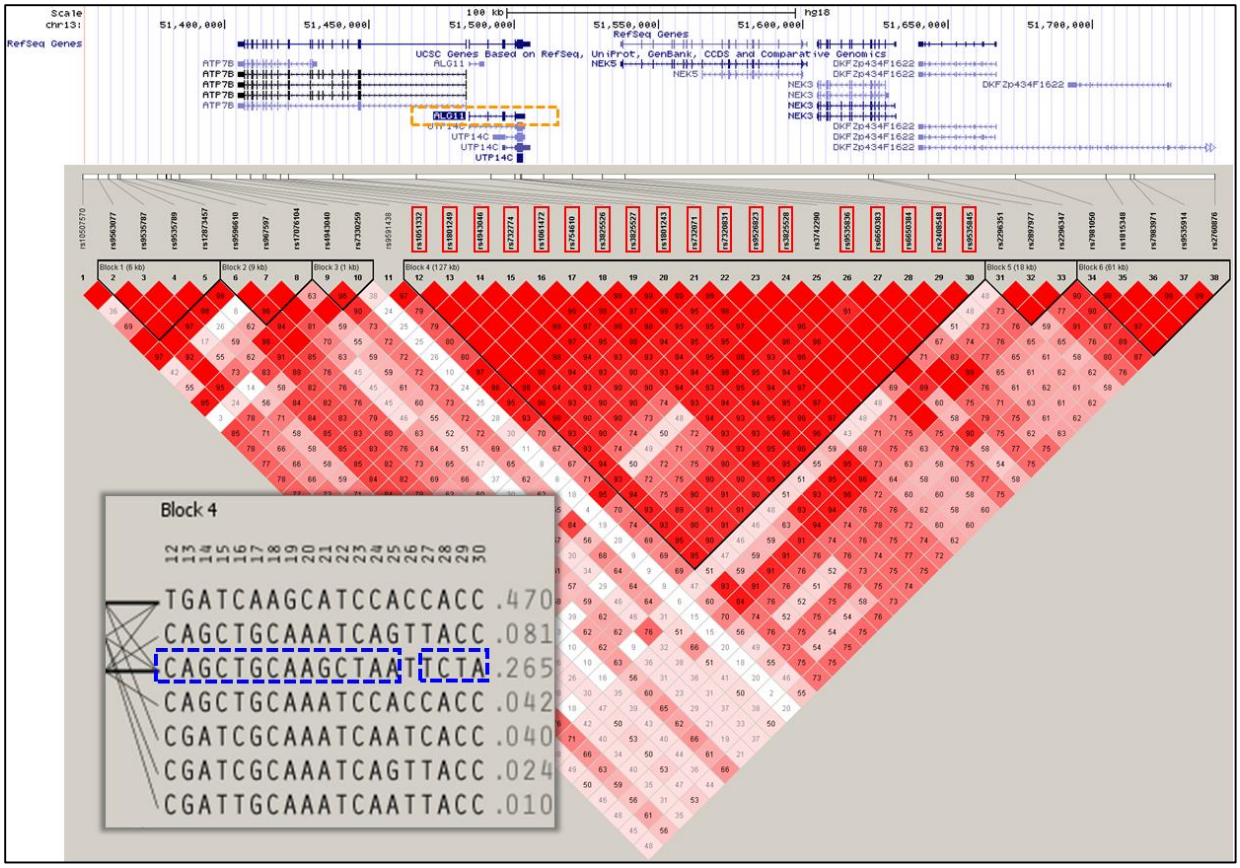

Figure S2 Linkage disequilibrium patterns of the genomic region surrounding evQTL at *ALG11*. The entire region of the analysis included 38 SNPs over a ~400 kb span. The *cis*-evSNPs are indicated with red boxes. The haplotypes in the LD block accommodating evSNPs are displayed in insert, with corresponding haplotype frequencies. Of note, alleles of evSNPs resulting in larger variance of gene expression are allocated in one haplotype highlighted with blue box.

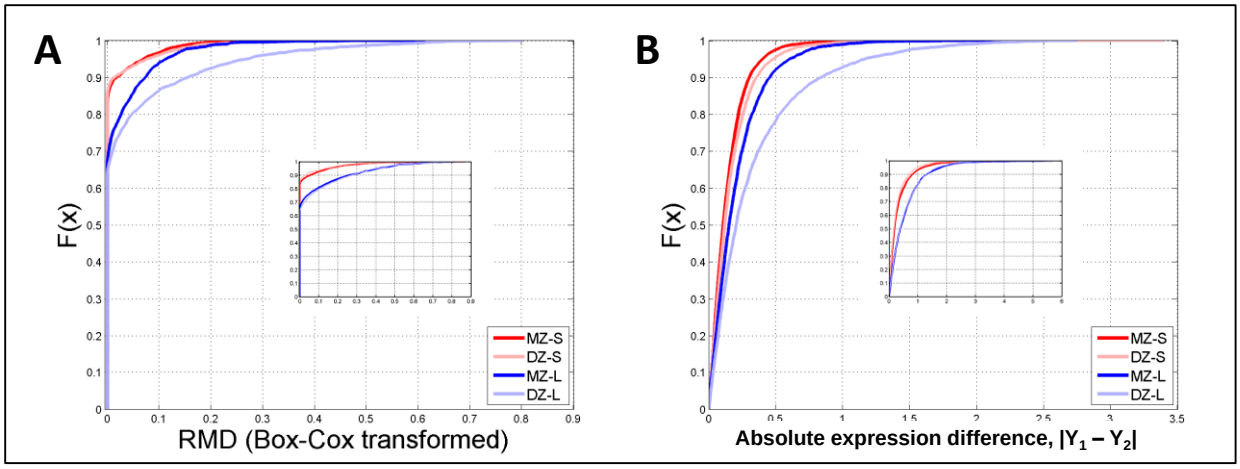

Figure S3 Comparison between results of the CDF analysis for the expression difference between twin pairs in evQTL genes. (**A**) Results obtained using RMD of Box-Cox normalized $\log_2$-transformed data between twin pairs. (**B**) Results obtained using absolute difference of $\log_2$-transformed data between twin pairs.

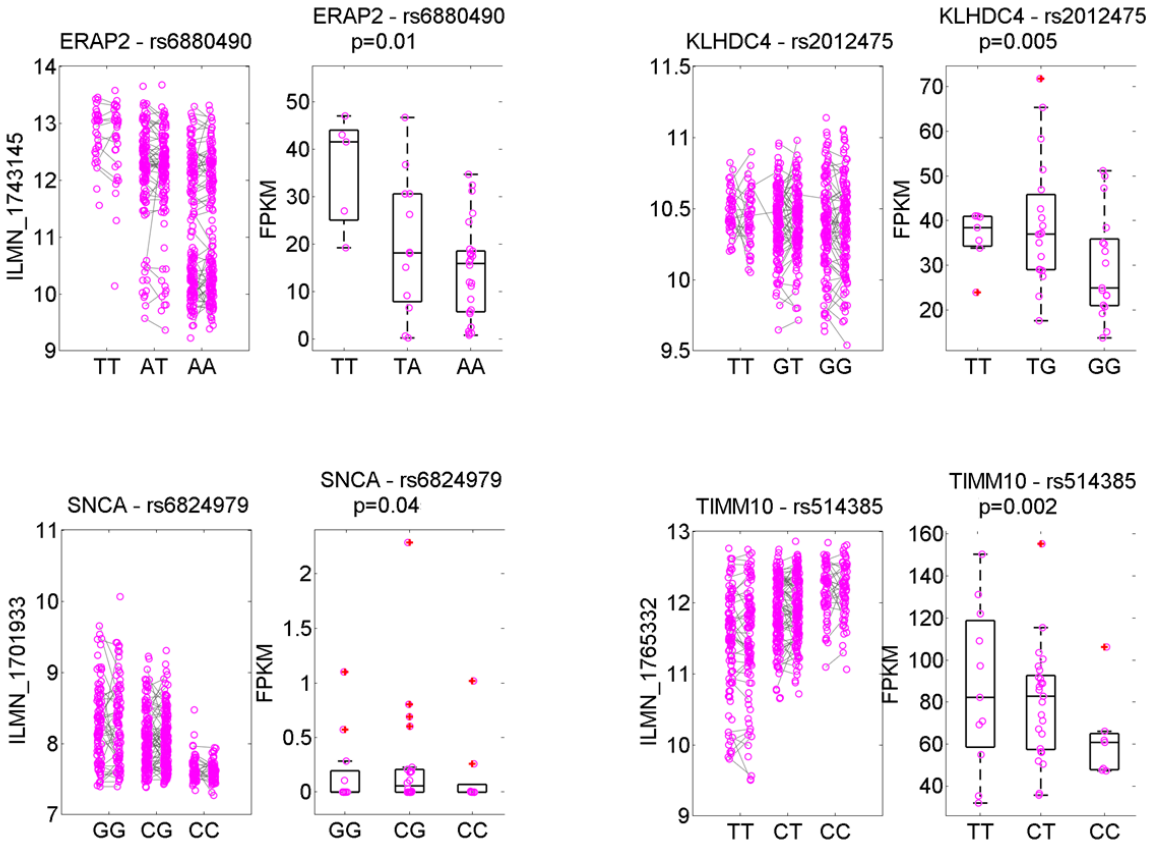

Figure S4 Comparison between evQTLs depicted using TwinsUK samples (left) and those of the same gene-SNP pairs depicted using RNA-seq data (MONTGOMERY et al. 2010) and the 1,000 Genomes Project genotype data (THE-1000-GENOMES-PROJECT-CONSORTIUM et al. 2012) (right). *P*-value of Levene's test for the equality of variances for expression levels calculated for two homozygous genotype groups is shown in the subtitle of the right panel in each figure.

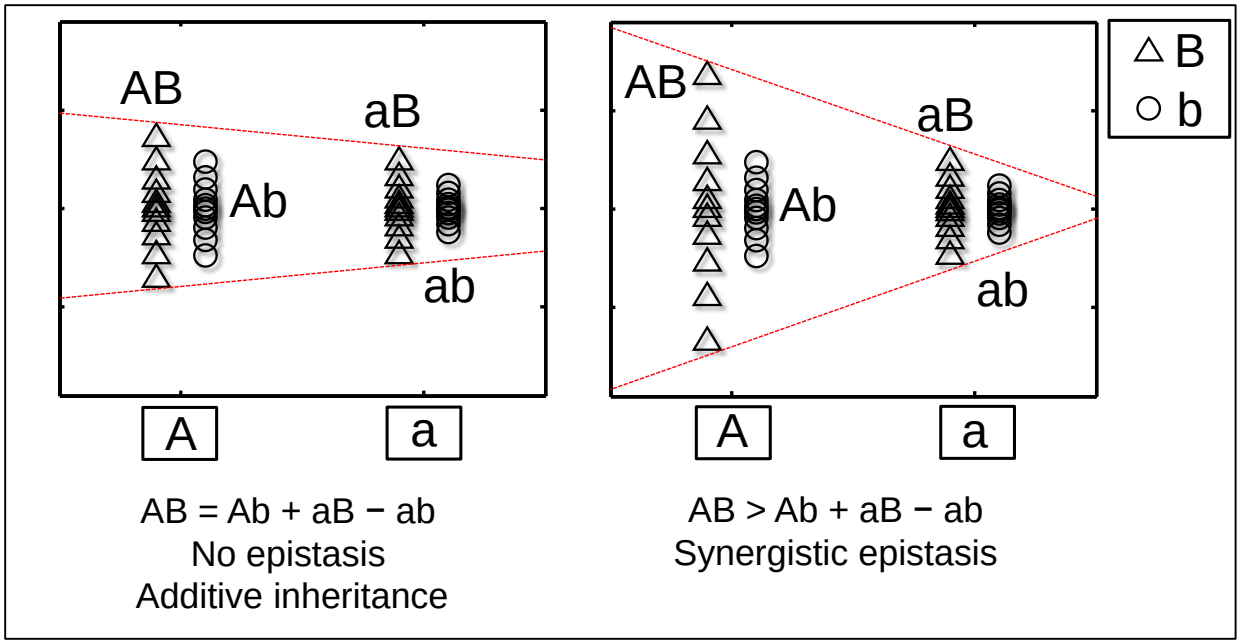

Figure S5 Schematic shows that both additive (left) and epistatic (right) effects reate similar evQTL signals. “A” and “a” are two alleles of evSNP, while “B” and “b” are alleles of interacting SNP.

Table S1 List of evQTLs detected in LCL, skin, and fat tissues.

Tissue		Gene	# of *cis*-evQTL SNPs	Representative evQTL SNP	Alleles	Allele associated with larger expression variance	Ancestral allele
LCL							
	1	*AGTPBP1*	61	rs7847905	T/C	C	T
	2	*ALG11*	171	rs9526819	G/A	G	A
	3	*ATP13A1*	108	rs4808964	C/A	C	A
	4	*AXIN2*	2	rs740026	G/A	A	G
	5	*BBS2*	41	rs14306	C/A	A	G
	6	*BRWD2*	43	rs729704	T/C	T	T
	7	*BTN3A2*	107	rs742090	C/A	A	G
	8	*C12ORF28*	173	rs775439	G/A	A	A
	9	*C12ORF54*	27	rs2731094	T/C	C	T
	10	*C16ORF30*	15	rs2076441	G/C	C	G
	11	*C17ORF97*	2	rs6565724	T/C	T	C
	12	*C7ORF28B*	3	rs10242703	G/A	A	A
	13	*CAPN11*	1	rs6938938	G/A	G	G
	14	*CDCP1*	3	rs6441894	T/C	C	T
	15	*CDO1*	21	rs698366	T/G	T	C
	16	*CHST3*	11	rs4148929	C/A	C	C
	17	*COPG2*	38	rs12706942	G/C	C	G
	18	*CORO2A*	12	rs12351801	T/C	T	C
	19	*COX4NB*	24	rs301164	T/C	C	C
	20	*CPA4*	74	rs6953940	G/A	A	G
	21	*CPNE1*	205	rs224354	G/C	G	C
	22	*DDIT4L*	64	rs2866194	G/A	A	A
	23	*EEF1G*	6	rs1058678	T/C	C	A
	24	*EIF5A*	1	rs7220464	T/G	T	G
	25	*EMR1*	2	rs4807916	T/C	C	C
	26	*ERAP2*	2	rs6880490	T/A	A	T
	27	*FLJ10916*	15	rs4246598	C/A	A	C
	28	*GAD1*	1	rs13007861	C/A	C	C
	29	*GDA*	19	rs7019060	G/A	A?	G
	30	*GOLPH4*	1	rs13100726	G/A	G	A
	31	*HCG22*	100	rs1064191	T/C	C	G
	32	*HCLS1*	2	rs4472078	T/C	T	G
	33	*HEY2*	36	rs628009	T/G	G?	A
	34	*HLA-DPB1*	59	rs9277341	T/C	T	C
	35	*HLA-DRB1*	27	rs9272723	T/C	C	T
	36	*HMHB1*	3	rs17100739	G/A	A	G
	37	*HYAL4*	8	rs12674456	T/C	T	T
	38	*IKZF1*	14	rs7789635	T/C	C	T
	39	*IPO8*	32	rs1371053	T/C	C	C
	40	*IRF5*	3	rs10156169	G/A	A	G
	41	*ITIH4*	26	rs2240915	T/A	T	T
	42	*KANK1*	17	rs2361106	G/A	G	A
	43	*KLHDC4*	2	rs2012475	T/G	G	T
	44	*LDHC*	19	rs3993291	T/C	T	C
	45	*LEMD3*	77	rs11175680	G/C	G	A
	46	*LOC440160*	8	rs12052294	G/A	A	G
	47	*LOC642290*	1	rs7255207	T/C	T	C
	48	*LOC648453*	3	rs17753176	T/A	A	A
	49	*LOC650263*	44	rs726920	G/C	G	G
	50	*LOC650557*	445	rs1980495	C/A	C	G
	51	*LTBR*	3	rs2364484	C/A	A	A
	52	*MED4*	53	rs9534912	T/C	T	C
	53	*MFF*	29	rs7573035	C/A	A	A
	54	*MGAT4A*	4	rs13032507	C/A	C	A
	55	*MMRN1*	3	rs7693616	T/G	G	T

	56	*MOCOS*	5	rs1893269	G/C	C	C
	57	*MTERFD2*	16	rs3815291	C/A	C	A
	58	*MTRR*	32	rs327588	G/C	C	G
	59	*MXRA7*	30	rs11077850	T/C	T	C
	60	*MYEOV*	24	rs7939250	G/A	G	A
	61	*MYH11*	18	rs3851702	T/C	C	G
	62	*N4BP2*	1	rs6824761	C/A	A	A
	63	*NLRP2*	5	rs11667481	G/A	A	G
	64	*NMNAT3*	21	rs4367021	C/A	A	C
	65	*NUDT2*	15	rs1337593	T/C	T	C
	66	*OAS1*	70	rs1293742	T/C	T	A
	67	*PAX8*	4	rs2863243	T/C	C	C
	68	*PGM5*	52	rs12685375	T/C	C	C
	69	*PHACS*	4	rs7928485	T/C	T	T
	70	*PILRB*	2	rs5015756	T/C	T	C
	71	*PJA2*	27	rs12719109	T/C	C	T
	72	*PLXDC2*	3	rs2038912	T/C	T	C
	73	*PPIC*	30	rs30063	G/A	G	T
	74	*PPIL3*	2	rs3731714	T/C	T	G
	75	*PRPF31*	19	rs254259	T/C	C	A
	76	*PTER*	8	rs7913889	G/A	A	C
	77	*RAMP1*	10	rs7578855	T/C	T	T
	78	*RGS17*	19	rs3870364	G/A	A	G
	79	*RPS6KA2*	24	rs9356531	G/A	G	A
	80	*SCD5*	49	rs2125171	T/C	C	G
	81	*SEMA4G*	20	rs752974	T/C	T	G
	82	*SERPINB10*	3	rs967538	T/G	G	T
	83	*SERPINB6*	7	rs7759176	T/C	C	G
	84	*SF1*	1	rs563536	T/C	C	T
	85	*SLC39A8*	80	rs151402	G/A	A	A
	86	*SNCA*	124	rs6824979	G/C	G	G
	87	*SNHG5*	8	rs12190637	G/A	G	A
	88	*STEAP2*	101	rs42617	T/A	T	T
	89	*TBKBP1*	55	rs1912483	G/A	G	A
	90	*TIMM10*	49	rs514385	T/C	T	C
	91	*TIMM22*	3	rs2586306	T/A	T	A
	92	*UBE3C*	8	rs1182392	T/C	T	G
	93	*USP6*	61	rs6502843	G/A	A	A
	94	*WBSCR27*	40	rs13244770	T/G	T	T
	95	*WDR41*	109	rs4704434	G/A	A	A
	96	*XRN2*	157	rs804384	G/A	A	G
	97	*XYLT1*	7	rs4238652	T/C	C	C
	98	*YPEL4*	44	rs633129	G/A	A	C
	99	*ZNF544*	1	rs260462	G/A	A	G
SKIN							
	1	*ACCN1*	2	rs7207800	A/G	G	A
	2	*ACTC1*	4	rs8039278	T/G	G	T
	3	*AGR3*	2	rs17629719	G/T	G	G
	4	*ATP13A1*	102	rs741706	A/G	G	A
	5	*C10ORF53*	1	rs751595	A/G	G	C
	6	*C8ORF42*	1	rs17665859	C/T	T	T
	7	*CACNA1H*	1	rs2745167	T/C	T	C
	8	*CCL23*	28	rs864104	C/T	C	A
	9	*CLLU1OS*	4	rs4760407	T/C	T	C
	10	*COPG2*	1	rs10863	A/G	A	G
	11	*EEF1G*	1	rs11231168	T/C	T	C
	12	*FLJ10916*	7	rs1878809	C/G	G	C
	13	*FLJ35429*	1	rs6457121	G/A	A	G
	14	*FLJ45964*	3	rs6713353	G/C	G	C
	15	*FTHL7*	1	rs7991197	C/T	T	T
	16	*GUCY1B2*	31	rs9568497	C/T	T	T
	17	*HERC6*	7	rs713175	C/T	T	C
	18	*LOC147650*	2	rs8102075	A/G	A	G

	19	*LOC283398*	1	rs10859604	A/G	A	A
	20	*LOC643453*	2	rs4876979	G/T	T	C
	21	*LOC643747*	2	rs12301198	G/A	A	A
	22	*LOC644366*	2	rs3851781	T/C	T	C
	23	*LOC644889*	3	rs11024097	G/A	G	C
	24	*LOC645351*	1	rs7712944	G/A	G	G
	25	*LOC645797*	9	rs7481311	A/G	G	C
	26	*LOC647541*	3	rs2808788	A/G	G	G
	27	*LOC648226*	11	rs7850776	C/T	C	C
	28	*LOC652367*	2	rs860214	C/T	T	G
	29	*LOC653034*	1	rs231674	T/C	C	C
	30	*LOC653486*	1	rs3758653	C/T	C	T
	31	*LOC730684*	2	rs667470	C/T	T	T
	32	*LTBP3*	1	rs2236684	G/T	T	A
	33	*MDGA1*	4	rs7748388	C/T	C	G
	34	*MED4*	69	rs9567986	T/C	C	T
	35	*MUC16*	1	rs12459695	C/G	G	G
	36	*NOX3*	1	rs11756851	C/T	C	T
	37	*NUPL1*	1	rs9507399	A/G	A	G
	38	*PAX8*	5	rs7589901	C/T	T	A
	39	*PECR*	12	rs7570208	C/A	C	G
	40	*PKD1L2*	10	rs9934926	C/T	T	A
	41	*SEMA4G*	18	rs807029	C/T	C	G
	42	*SH2D4A*	7	rs748208	T/C	T	C
	43	*SOX5*	1	rs17468457	C/T	T	T
	44	*SPINK7*	95	rs999741	C\T	T	C
	45	*TFG*	86	rs9878163	A\G	G	A
	46	*TIMM10*	119	rs7943793	A/G	G	T
	47	*TMEM25*	3	rs573971	G/A	A	A
	48	*UNC45A*	8	rs871078	G/C	C	A
	49	*UPK1A*	15	rs4806197	C/T	T	C
	50	*USMG5*	27	rs4918003	C/G	G	T
	51	*WDR41*	39	rs7714170	C/T	T	A
	52	*WFIKKN1*	20	rs7205409	C/T	C	C
	53	*XYLT1*	1	rs1045885	G/C	C	C
	54	*ZNF365*	63	rs7923561	C/T	C	A
	55	*ZNF418*	63	rs9749429	T/C	T	C
	56	*ZNF713*	1	rs4948003	A/T	A	T
FAT							
	1	*ADH4*	30	rs7694844	A/C	A	T
	2	*AGTPBP1*	33	rs918941	C/T	T	G
	3	*ATP13A1*	98	rs968525	A/T	T	C
	4	*B4GALT4*	4	rs4449310	A/G	A	C
	5	*BRWD2*	45	rs9325569	A/T	T	G
	6	*CCL23*	13	rs864104	A/G	G	A
	7	*CDK6*	3	rs982692	C/G	C	T
	8	*CHST13*	5	rs6786437	C/A	C	C
	9	*CLEC12A*	5	rs7313235	T/C	C	C
	10	*CLUAP1*	2	rs6501178	A/C	A	A
	11	*COG6*	1	rs2180960	T/G	G	G
	12	*COPG2*	34	rs992859	A/G	A	C
	13	*DAPL1*	2	rs759033	C/G	C	T
	14	*DLG4*	4	rs8067250	C/G	G	C
	15	*DMKN*	2	rs6510492	A/G	G	T
	16	*DPYSL4*	23	rs902627	G/T	G	G
	17	*DSEL*	8	rs9946666	C/T	T	T
	18	*EIF3H*	1	rs1695714	G/T	T	A
	19	*EMR2*	18	rs8107048	A/G	A	C
	20	*FGA*	1	rs1316990	C/T	T	C
	21	*FLJ10916*	7	rs7425197	C/G	G	C
	22	*FTHL7*	1	rs3945859	G/T	G	T
	23	*FUT4*	6	rs596855	A/G	G	C

	24	*GUCA1C*	1	rs9821923	C/A	C	A
	25	*HLA-DPB1*	29	rs9469485	T/G	T	C
	26	*HLA-DRB1*	54	rs9275141	A/C	A	G
	27	*HMHB1*	2	rs17100749	C/G	G	G
	28	*IKZF1*	4	rs6952409	A/G	G	G
	29	*IL1RN*	2	rs315948	T/C	T	G
	30	*IL8*	18	rs552582	T/G	T	A
	31	*IPO8*	21	rs9300199	C/T	T	T
	32	*IRF5*	41	rs7808907	A/G	G	C
	33	*ITGA4*	1	rs13429032	A/G	G	G
	34	*KCTD10*	42	rs9593	C/T	?	A
	35	*KCTD12*	2	rs9600792	A/G	G	T
	36	*KCTD8*	7	rs7684615	G/C	C	G
	37	*LAIR2*	1	rs4806747	A/G	A	A
	38	*LDHC*	65	rs7946331	C/G	G	G
	39	*LEMD3*	65	rs990609	A/C	A	C
	40	*LILRA2*	3	rs1671176	A/C	C	C
	41	*LOC644366*	3	rs3851781	C/T	C	C
	42	*LOC644889*	2	rs1544506	T/C	T	C
	43	*LOC645460*	1	rs17762258	G/A	G	A
	44	*LOC646625*	7	rs9554219	G/T	T	T
	45	*LOC653034*	1	rs7225168	G/T	T	G
	46	*LRRC25*	6	rs8101804	G/A	A	C
	47	*LZTS1*	1	rs13264395	G/C	G	C
	48	*MAPK7*	3	rs9911451	A/T	A	A
	49	*MARCH8*	81	rs984955	C/T	T	A
	50	*MED4*	101	rs9595825	C/T	T	T
	51	*MEGF9*	50	rs991121	C/T	T	G
	52	*MFF*	14	rs9646881	A/C	C	G
	53	*MGC12965*	5	rs682229	T/C	C	A
	54	*MOCOS*	2	rs604271	G/A	G	A
	55	*MYH1*	50	rs995362	T/C	T	G
	56	*NUDT2*	15	rs4480190	C/G	C	C
	57	*PAX8*	14	rs895417	T/C	T	A
	58	*PLIN*	1	rs16942690	G/A	G	A
	59	*PMP2*	1	rs13261227	A/G	G	C
	60	*PPIL3*	5	rs7582581	A/G	G	T
	61	*PRPF31*	11	rs42318	G/A	A	G
	62	*PTCHD3*	39	rs7342136	A/G	A	T
	63	*PTER*	2	rs2298126	A/C	C	A
	64	*RAB37*	2	rs3850121	T/C	C	G
	65	*RIMS2*	37	rs4734094	T/A	A	G
	66	*RNF126P1*	14	rs9303389	G/C	G	G
	67	*RPL14*	127	rs9968170	T/C	C	T
	68	*SEMA4G*	30	rs807029	C/T	C	G
	69	*SFTPD*	52	rs9421727	C/T	C	T
	70	*SILV*	4	rs3213122	A/G	A	C
	71	*SLC39A8*	51	rs9705	A/G	G	G
	72	*STMN2*	15	rs4305892	A/G	A	T
	73	*TFG*	109	rs989795	A/G	G	T
	74	*TIMM10*	162	rs929934	A/G	G	G
	75	*TRPC4AP*	17	rs6120827	T/G	T	T
	76	*USMG5*	26	rs7913461	C/G	C	A
	77	*WBSCR27*	31	rs8629	G/C	C	C
	78	*WDR41*	5	rs335655	C/A	C	G
	79	*ZNF232*	87	rs9912506	A/G	A	G

Table S2 (**A) *P*-values of Kolmogorov-Smirnov (K-S) test of pairwise comparison between groups in the RMD distribution and basic descriptive statistics for the RMD values.** Upper diagonal *P*-values are from randomly selected non-evQTL genes. Results obtained from the Box-Cox normalized and not normalized data are given. **(B) The per-probe statistics of RMD values for larger variance groups (MZ-L and DZ-L).** Probes with no SNP in either groups are not shown.

A	**Log2-transformed data**				
		MZ-S	DZ-S	MZ-L	DZ-L
	MZ-S	-	0.40	0.031	2.94e-04
	DZ-S	9.34e-06	-	0.033	0.026
	MZ-L	8.82e-18	1.66e-06	-	0.16
	DZ-L	1.40e-78	1.66e-52	7.28e-22	-
	Box-Cox normalized log2-transformed data				
		MZ-S	DZ-S	MZ-L	DZ-L
	MZ-S	-	0.1670	4.49e-08	1.98e-14
	DZ-S	0.0486	-	9.60e-11	8.30e-20
	MZ-L	1.59e-68	1.57e-80	-	0.0333
	DZ-L	2.97e-85	1.69e-99	1.84e-07	-

B	Probe	RMD in MZ-L and DZ-L groups					
		MZ-L			DZ-L		
		# of SNPs	Mean	Median	# of SNPs	Mean	Median
	ILMN_1658247	37	0.0126	0.0103	47	0.0366	0.0150
	ILMN_1660086	75	0.0127	0.0091	84	0.0180	0.0123
	ILMN_1664641	6	0.0062	0.0060	31	0.0131	0.0106
	ILMN_1667229	24	0.0138	0.0098	25	0.0212	0.0129
	ILMN_1669032	62	0.0142	0.0098	44	0.0178	0.0175
	ILMN_1670841	6	0.0048	0.0047	3	0.0135	0.0180
	ILMN_1675616	36	0.0102	0.0094	39	0.0111	0.0101
	ILMN_1676528	41	0.0127	0.0083	45	0.0332	0.0181
	ILMN_1676575	29	0.0118	0.0108	35	0.0255	0.0167
	ILMN_1677124	31	0.0079	0.0042	29	0.0141	0.0144
	ILMN_1678974	50	0.0115	0.0085	28	0.0175	0.0148
	ILMN_1682034	124	0.0100	0.0083	178	0.0112	0.0086
	ILMN_1696537	16	0.0139	0.0101	5	0.0090	0.0047
	ILMN_1701933	27	0.0204	0.0176	31	0.0325	0.0238
	ILMN_1704598	4	0.0114	0.0119	2	0.0145	0.0145
	ILMN_1706959	32	0.0068	0.0051	35	0.0115	0.0083
	ILMN_1707137	9	0.0128	0.0089	7	0.0251	0.0083
	ILMN_1709173	6	0.0093	0.0090	2	0.0041	0.0041
	ILMN_1709590	19	0.0172	0.0114	6	0.0126	0.0114
	ILMN_1712400	7	0.0101	0.0084	10	0.0261	0.0237
	ILMN_1713803	9	0.0109	0.0129	7	0.0205	0.0090
	ILMN_1715169	20	0.0214	0.0162	11	0.0323	0.0264
	ILMN_1715693	75	0.0112	0.0090	69	0.0173	0.0130
	ILMN_1716218	11	0.0119	0.0125	9	0.0148	0.0071
	ILMN_1718932	111	0.0148	0.0106	143	0.0277	0.0151
	ILMN_1719170	31	0.0087	0.0075	28	0.0149	0.0115
	ILMN_1719204	50	0.0067	0.0064	53	0.0130	0.0075
	ILMN_1721727	29	0.0117	0.0079	21	0.0233	0.0131
	ILMN_1723984	41	0.0109	0.0093	43	0.0095	0.0076
	ILMN_1724480	47	0.0108	0.0083	35	0.0206	0.0171
	ILMN_1726624	11	0.0088	0.0101	8	0.0175	0.0116
	ILMN_1743836	2	0.0077	0.0077	20	0.0089	0.0087
	ILMN_1745043	11	0.0064	0.0040	8	0.0120	0.0112
	ILMN_1749070	12	0.0082	0.0081	9	0.0129	0.0111
	ILMN_1751559	14	0.0124	0.0087	11	0.0096	0.0066
	ILMN_1752150	7	0.0176	0.0155	7	0.0120	0.0116
	ILMN_1753164	49	0.0079	0.0062	46	0.0260	0.0085
	ILMN_1753312	26	0.0151	0.0118	29	0.0145	0.0117

	ILMN_1756579	41	0.0079	0.0055	33	0.0090	0.0065
	ILMN_1764754	35	0.0096	0.0070	27	0.0179	0.0128
	ILMN_1765274	24	0.0114	0.0085	17	0.0150	0.0103
	ILMN_1765332	38	0.0088	0.0067	42	0.0199	0.0171
	ILMN_1766165	27	0.0118	0.0077	31	0.0141	0.0119
	ILMN_1778347	21	0.0084	0.0088	10	0.0132	0.0089
	ILMN_1778488	14	0.0069	0.0070	14	0.0141	0.0052
	ILMN_1780601	8	0.0166	0.0132	7	0.0133	0.0117
	ILMN_1781388	19	0.0155	0.0119	6	0.0212	0.0121
	ILMN_1784294	32	0.0149	0.0111	31	0.0218	0.0195
	ILMN_1787199	62	0.0077	0.0059	53	0.0083	0.0065
	ILMN_1795336	52	0.0080	0.0045	54	0.0209	0.0107
	ILMN_1809496	27	0.0108	0.0099	17	0.0153	0.0098
	ILMN_1813746	18	0.0173	0.0118	13	0.0237	0.0177
	ILMN_1835359	9	0.0125	0.0104	3	0.0232	0.0180
	ILMN_1836218	29	0.0090	0.0080	31	0.0147	0.0088
	ILMN_1851610	2	0.0029	0.0029	22	0.0079	0.0056
	ILMN_1854132	13	0.0102	0.0081	10	0.0077	0.0055
	ILMN_1864228	1	0.0141	0.0141	11	0.0123	0.0064
	ILMN_1885273	14	0.0124	0.0095	6	0.0116	0.0088
	ILMN_1897741	12	0.0090	0.0075	8	0.0296	0.0290
	ILMN_1900622	1	0.0047	0.0047	1	0.0024	0.0024
	ILMN_1900994	10	0.0062	0.0033	5	0.0086	0.0075
	ILMN_1904238	53	0.0120	0.0091	46	0.0192	0.0141
	ILMN_2086222	61	0.0140	0.0116	61	0.0119	0.0090
	ILMN_2120982	4	0.0167	0.0157	7	0.0111	0.0080
	ILMN_2134224	15	0.0086	0.0086	7	0.0067	0.0039
	ILMN_2147424	21	0.0124	0.0101	24	0.0239	0.0187
	ILMN_2154287	3	0.0294	0.0370	1	0.0061	0.0061
	ILMN_2173294	45	0.0062	0.0049	28	0.0147	0.0095
	ILMN_2181363	20	0.0141	0.0106	12	0.0074	0.0034
	ILMN_2183938	20	0.0220	0.0172	11	0.0441	0.0346
	ILMN_2196479	14	0.0129	0.0090	13	0.0197	0.0128
	ILMN_2198408	15	0.0196	0.0148	11	0.0237	0.0209
	ILMN_2200659	109	0.0069	0.0058	131	0.0146	0.0091
	ILMN_2223922	4	0.0104	0.0116	1	0.0088	0.0088
	ILMN_2233539	55	0.0192	0.0135	58	0.0266	0.0151
	ILMN_2237428	10	0.0134	0.0115	5	0.0129	0.0075
	ILMN_2246083	53	0.0247	0.0188	40	0.0299	0.0257
	ILMN_2262288	91	0.0079	0.0046	113	0.0153	0.0096

Table S3 LCL evQTLs and interacting SNPs. *P*-value indicates the linear regression significance between interacting SNP genotype and gene expression of evQTL gene (see Methods). Two genes close to interacting SNPs at *trans*-genomic location are given.

LCL evQTL gene	evSNP	evSNP position	Interacting SNP	Interacting SNP position	P-value	*trans*-located gene
PTER	rs7913889	chr10:16559381	rs7909832	chr10:16556710	8.66E-77	
ERAP2	rs6880490	chr5:96185549	rs11135484	chr5:96221889	4.76E-45	
BBS2	rs14306	chr16:56510299	rs13337155	chr16:56504724	5.77E-39	
IPO8	rs1371053	chr12:30845705	rs7968343	chr12:30817004	6.16E-38	
WDR41	rs4704434	chr5:76890568	rs441102	chr5:76738970	4.31E-33	
SEMA4G	rs752974	chr10:102762256	rs12571302	chr10:102742763	4.13E-32	
N4BP2	rs6824761	chr4:40054818	rs4974962	chr4:40115239	4.92E-29	
PPIL3	rs3731714	chr2:202060820	rs7606251	chr2:201736734	7.47E-28	
PILRB	rs5015756	chr7:100013457	rs7341507	chr7:99951315	9.06E-27	
TIMM10	rs514385	chr11:57306011	rs3851118	chr11:57280951	1.36E-23	
LOC440160	rs12052294	chr2:132433268	rs6720375	chr2:132440073	1.15E-21	
C12ORF28	rs775439	chr12:70096374	rs811822	chr12:70107646	1.35E-20	
TIMM22	rs2586306	chr17:909451	rs2241931	chr17:909653	2.05E-20	
SERPINB10	rs967538	chr18:61509075	rs6567399	chr18:61544546	1.60E-19	
BTN3A2	rs742090	chr6:26415637	rs3799378	chr6:26404374	4.29E-19	
SCD5	rs2125171	chr4:83835698	rs6830527	chr4:83798446	7.00E-18	
GOLPH4	rs13100726	chr3:167707035	rs9873288	chr3:167726096	2.77E-17	
NUDT2	rs1337593	chr9:34334015	rs7032924	chr9:34167694	2.59E-16	
LOC642290	rs7255207	chr19:28385923	rs7258333	chr19:28269920	8.94E-16	
NMNAT3	rs4367021	chr3:139420218	rs9822952	chr3:139385257	1.44E-15	
IRF5	rs10156169	chr7:128684571	rs8043	chr7:128607384	9.33E-15	
IKZF1	rs7789635	chr7:50473610	rs10216316	chr7:50462418	5.65E-14	
C12ORF54	rs2731094	chr12:48881661	rs12318285	chr12:48905575	8.20E-14	
LEMD3	rs11175680	chr12:65605539	rs2133323	chr12:65577453	1.16E-13	
MTRR	rs327588	chr5:7908359	rs1801394	chr5:7870973	1.40E-12	
WBSCR27	rs13244770	chr7:73280691	rs4440516	chr7:73224041	3.61E-12	
XYLT1	rs4238652	chr16:17199665	rs7190386	chr16:17183373	4.25E-12	
ITIH4	rs2240915	chr3:52859526	rs17331178	chr3:52847544	5.13E-12	
ATP13A1	rs4808964	chr19:19603692	rs2304128	chr19:19746151	1.18E-11	
KLHDC4	rs2012475	chr16:87760814	rs2290019	chr16:87741662	2.95E-11	
HCLS1	rs4472078	chr3:121477634	rs12493927	chr3:121378927	7.57E-11	
MTERFD2	rs3815291	chr2:242032540	rs7559967	chr2:242006736	8.75E-10	
HCG22	rs1064191	chr6:31075375	rs2523857	chr6:31021504	2.12E-09	
LOC650557	rs1980495	chr6:32346794	rs3129878	chr6:32408735	3.03E-09	
CDCP1	rs6441894	chr3:45206579	rs7633169	chr3:45185519	1.21E-08	
HLA-DPB1	rs9277341	chr6:33039625	rs9501259	chr6:33055551	1.15E-07	
LDHC	rs3993291	chr11:18479739	rs4757662	chr11:18439738	1.19E-07	
AXIN2	rs740026	chr17:63561681	rs757558	chr17:63561592	1.25E-07	
FLJ10916	rs4246598	chr2:88438050	rs2970924	chr2:88433305	2.24E-07	
MED4	rs9534912	chr13:48588564	rs9526455	chr13:48712474	2.62E-07	
RPS6KA2	rs9356531	chr6:167276965	rs9355601	chr6:167269159	3.29E-07	
C16ORF30	rs2076441	chr16:1590576	rs2235643	chr16:1585115	1.76E-06	
PLXDC2	rs2038912	chr10:20207349	rs1326233	chr10:20238864	2.64E-06	
UBE3C	rs1182392	chr7:157036200	rs10271990	chr7:157037521	3.76E-06	
SERPINB6	rs7759176	chr6:2915664	rs7751676	chr6:2931879	6.59E-06	
COPG2	rs12706942	chr7:130149061	rs10128	chr7:130151694	7.90E-06	
EIF5A	rs7220464	chr17:7210836	rs11078672	chr17:7215142	9.29E-06	
ALG11	rs9526819	chr13:52563475	rs3742289	chr13:52603194	2.63E-05	
SLC39A8	rs151402	chr4:103190486	rs151372	chr4:103174196	2.65E-05	
HMHB1	rs17100739	chr5:143112574	rs10498677	chr6:11466011	3.20E-05	*TMEM170B*
COX4NB	rs301164	chr16:85813880	rs9923691	chr16:85824190	6.63E-05	
CPNE1	rs224354	chr20:34054609	rs1118233	chr20:34228349	7.84E-05	
YPEL4	rs633129	chr11:57685196	rs12790660	chr11:57667222	9.68E-05	
PRPF31	rs254259	chr19:54606405	rs16985368	chr19:54614590	1.17E-04	
C7ORF28B	rs10242703	chr7:6870635	rs2792574	chr14:48676427	2.86E-04	*MDGA2*
CPA4	rs6953940	chr7:129947007	rs3807344	chr7:129934219	4.33E-04	

Table S4 The evQTL genes associated with complex traits and reported in the GWAS catalog (accessed in May 2013).

	GWAS gene	Disorder or trait	GWAS reported SNP and risk allele
LCL			
	RAMP1	Obesity-related traits	rs10185142-A
	IKZF1	Systemic lupus erythematosus|Red blood cell traits|Inflammatory bowel disease|Crohn's disease|Systemic lupus erythematosus|Mean corpuscular volume|Acute lymphoblastic leukemia (childhood)|Acute lymphoblastic leukemia (childhood)|Hippocampal atrophy	rs10276619-G|rs12718598-T|rs1456896-T|rs1456896-T|rs4917014-A|rs12718597-A|rs4132601-C|rs11978267-G|rs10276619-?
	HLA-DPB1	Aspirin exacerbated respiratory disease in asthmatics|Hepatitis B (viral clearance)|Hepatitis B (viral clearance)|Hepatitis B vaccine response|Systemic sclerosis|Hepatitis B|Nephropathy|Hepatitis B	rs1042151-G|rs9277535-G|rs9277535-G|rs9277535-?|rs987870-?|rs9277535-G|rs1883414-?|rs9277535-G
	PTER	Obesity	rs10508503-C
	PAX8	Renal function-related traits (BUN)	**rs11123170-G**
	GDA	Suicidal ideation|Suicidal ideation	rs11143230-C|rs11143230-C
	NLRP2	Inflammatory bowel disease	rs11672983-A
	SLC39A8	Diastolic blood pressure|Hypertension|Systolic blood pressure|Blood pressure|Body mass index|HDL cholesterol	rs13107325-T|rs13107325-T|rs13107325-T|rs13107325-T|rs13107325-T|rs13107325-T
	BTN3A2	Schizophrenia	rs13194053-T
	ERAP2	Inflammatory bowel disease|Crohn's disease	rs1363907-A|rs2549794-C
	PLXDC2	Diabetic retinopathy	rs1571942-C
	WDR41	Caudate nucleus volume	**rs163030-A**
	PPIC	Aortic root size	rs17470137-A
	OAS1	Response to fenofibrate (adiponectin levels)	rs2384207-?
	ZNF544	Attention deficit hyperactivity disorder	rs260461-?
	LDHC	Amyloid A Levels	rs2896526-G
	BRWD2	Obesity-related traits|Visceral fat	rs2919009-A|rs7085142-T
	HCG22	Hematology traits|Behcet's disease|Hypothyroidism	rs3130544-A|rs4947296-?|**rs2517532-G**
	PGM5	Bipolar disorder	rs3750552-?
	MTRR	Capecitabine sensitivity|Capecitabine sensitivity	rs4702484-?|rs4702484-?
	SF1	Urate levels	rs606458-T
	MYEOV	Breast cancer	rs614367-T
	RPS6KA2	Dental caries|Inflammatory bowel disease	rs635808-?|rs1819333-T
	SNCA	Parkinson's disease|Parkinson's disease|Parkinson's disease|Parkinson's disease|Parkinson's disease|Parkinson's disease|Parkinson's disease|Parkinson's disease|Parkinson's disease|Parkinson's disease	rs6532194-?|rs356219-?|rs356220-?|rs356220-T|rs356219-G|rs356220-T|rs356220-A|rs356220-T|rs2736990-?|rs11931074-?|rs2736990-C
	MMRN1	Parkinson's disease	rs6532197-G
	MTERFD2	Sex hormone-binding globulin levels	rs6721345-A
	UBE3C	Response to citalopram treatment|Response to citalopram treatment|Quantitative traits	rs6966038-?|rs6966038-?|rs2527866-C
	IRF5	Systemic lupus erythematosus|Ulcerative colitis|Systemic lupus erythematosus|Systemic sclerosis|Systemic sclerosis|Systemic sclerosis|Systemic sclerosis|Systemic sclerosis|Primary biliary cirrhosis|Systemic lupus erythematosus|Systemic lupus erythematosus|Ulcerative colitis|Primary biliary cirrhosis|Rheumatoid arthritis|Systemic sclerosis|Systemic lupus erythematosus|Systemic lupus erythematosus|Systemic lupus erythematosus	rs729302-A|rs4728142-A|rs729302-A|rs10488631-?|rs10488631-?|rs10488631-?|rs10488631-?|rs10488631-?|rs10488631-C|rs12531711-G|rs12531711-G|rs10488631-C|rs4728142-A|rs10488631-C|rs10488631-C|rs10488631-C|rs4728142-A|rs12537284-A|rs10488631-C
	MGAT4A	Bipolar disorder|Bipolar disorder	rs7578035-G|rs12618769-T
	C12ORF28	Attention deficit hyperactivity disorder and conduct disorder	rs789560-G
	TBKBP1	Multiple sclerosis|Ankylosing spondylitis	**rs8070463-T|rs8070463-C**
	HLA-DRB1	Lymphoma|Epstein-Barr virus immune response (EBNA-1)|Leishmaniasis (visceral)|Systemic lupus erythematosus|Crohn's disease|Schizophrenia|Ulcerative colitis|Hepatocellular carcinoma|Immune response to anthrax vaccine |Hypothyroidism|IgA nephropathy|Multiple sclerosis|Nodular sclerosis Hodgkin lymphoma|Nodular sclerosis Hodgkin lymphoma|Systemic sclerosis|Response to interferon beta therapy|Rheumatoid arthritis|Nephropathy|Ulcerative colitis|Rheumatoid arthritis|Immunoglobulin A |Immunoglobulin A |Lumiracoxib-related liver injury|Rheumatoid arthritis|Rheumatoid	rs9268853-C|rs477515-T|rs9271858-G|rs9270984-T|rs7765379-G|rs4530903-?|rs6927022-A|rs9272105-A|rs3104402-A|rs3129720-C|rs660895-G|rs3129889-G|5-SNP haplotype 6|5-SNP haplotype 3|rs3129763-?|rs9272105-A|rs7765379-?|rs9275596-?|rs9268853-T|rs7765379-?|rs9271366-?|rs2187668-A|rs3129900-?|rs13192471-G|rs6910071-G|rs9271100-?|rs9271366-G|rs3135388-A|rs2395185-?|rs3129934-T|rs6457620-?|rs2395148-?|rs660895-?|rs2647044-A|rs615672-?|DRB1*07

		arthritis|Systemic lupus erythematosus|Multiple sclerosis|Multiple sclerosis|Ulcerative colitis|Multiple sclerosis|Rheumatoid arthritis|Arthritis (juvenile idiopathic)|Rheumatoid arthritis|Type 1 diabetes|Rheumatoid arthritis|Response to ximelagatran treatment	
	ITIH4	Ulcerative colitis|Immune reponse to smallpox (secreted IL-2)|Schizophrenia|Bipolar disorder	rs9847710-C|rs17331151-A|rs2239547-?|rs736408-C
FAT			
	CDK6	White blood cell types|White blood cell count|Height|Height|Rheumatoid arthritis|Height|Height|Height	rs445-C|rs445-T|rs42235-T|rs2282978-?|rs42041-G|rs2282978-C|rs2040494-C|rs2282978-C
	FGA	Venous thromboembolism|Fibrinogen|D-dimer levels|Fibrinogen	rs7659024-A|rs6050-C|rs13109457-A|rs6056-A
	IRF5	Systemic lupus erythematosus|Ulcerative colitis|Systemic lupus erythematosus|Systemic sclerosis|Systemic sclerosis|Systemic sclerosis|Systemic sclerosis|Systemic sclerosis|Primary biliary cirrhosis|Systemic lupus erythematosus|Systemic lupus erythematosus|Ulcerative colitis|Primary biliary cirrhosis|Rheumatoid arthritis|Systemic sclerosis|Systemic lupus erythematosus|Systemic lupus erythematosus|Systemic lupus erythematosus	rs729302-A|rs4728142-A|rs729302-A|rs10488631-?|rs10488631-?|rs10488631-?|rs10488631-?|rs10488631-C|rs12531711-G|rs12531711-G|rs10488631-C|rs4728142-A|rs10488631-C|rs10488631-C|rs10488631-C|rs4728142-A|rs12537284-A|rs10488631-C
	IKZF1	Systemic lupus erythematosus|Red blood cell traits|Inflammatory bowel disease|Crohn's disease|Systemic lupus erythematosus|Mean corpuscular volume|Acute lymphoblastic leukemia (childhood)|Acute lymphoblastic leukemia (childhood)|Hippocampal atrophy	rs10276619-G|rs12718598-T|rs1456896-T|rs1456896-T|rs4917014-A|rs12718597-A|rs4132601-C|rs11978267-G|rs10276619-?
	MEGF9	Response to statin therapy	rs16909449-C
	FUT4	Response to angiotensin II receptor blocker therapy|Response to angiotensin II receptor blocker therapy (opposite direction w/ diuretic therapy)	rs11020821-C|rs11020821-C
	KCTD8	Response to amphetamines|Anticoagulant levels	rs17641529-?|rs13130255-?
	PAX8	Renal function-related traits (BUN)	**rs11123170-G**
	COG6	Psoriasis	rs7993214-?
	TRPC4AP	Prothrombin time	rs2295888-A
	BRWD2	Obesity-related traits|Visceral fat	rs2919009-A|rs7085142-T
	IL1RN	Obesity-related traits|C-reactive protein|Protein quantitative trait loci	rs4252023-A|rs6734238-G|rs6761276-?
	DAPL1	Obesity-related traits	rs16843372-G
	PTER	Obesity	rs10508503-C
	ITGA4	Monocyte count|White blood cell types|Celiac disease	rs2124440-G|rs12988934-T|rs13010713-G
	DSEL	Major depressive disorder|Cognitive performance	rs17077540-G|rs2124349-?
	HLA-DRB1	Lymphoma|Epstein-Barr virus immune response (EBNA-1)|Leishmaniasis (visceral)|Systemic lupus erythematosus|Crohn's disease|Schizophrenia|Ulcerative colitis|Hepatocellular carcinoma|Immune response to anthrax vaccine |Hypothyroidism|IgA nephropathy|Multiple sclerosis|Nodular sclerosis Hodgkin lymphoma|Nodular sclerosis Hodgkin lymphoma|Systemic sclerosis|Response to interferon beta therapy|Rheumatoid arthritis|Nephropathy|Ulcerative colitis|Rheumatoid arthritis|Immunoglobulin A |Immunoglobulin A |Lumiracoxib-related liver injury|Rheumatoid arthritis|Rheumatoid arthritis|Systemic lupus erythematosus|Multiple sclerosis|Multiple sclerosis|Ulcerative colitis|Multiple sclerosis|Rheumatoid arthritis|Arthritis (juvenile idiopathic)|Rheumatoid arthritis|Type 1 diabetes|Rheumatoid arthritis|Response to ximelagatran treatment	rs9268853-C|rs477515-T|rs9271858-G|rs9270984-T|rs7765379-G|rs4530903-?|rs6927022-A|rs9272105-A|rs3104402-A|rs3129720-C|rs660895-G|rs3129889-G|5-SNP haplotype 6|5-SNP haplotype 3|rs3129763-?|rs9272105-A|rs7765379-?|rs9275596-?|rs9268853-T|rs7765379-?|rs9271366-?|rs2187668-A|rs3129900-?|rs13192471-G|rs6910071-G|rs9271100-?|rs9271366-G|rs3135388-A|rs2395185-?|rs3129934-T|rs6457620-?|rs2395148-?|rs660895-?|rs2647044-A|rs615672-?|DRB1*07
	DLG4	Liver enzyme levels (alkaline phosphatase)	rs314253-C
	IL8	Inflammatory bowel disease	rs2472649-G
	KCTD10	HDL cholesterol	rs9943753-G
	PTCHD3	Fasting insulin-related traits (interaction with BMI)	rs1334893-?
	ADH4	Esophageal cancer (alcohol interaction)	rs3805322-?
	SLC39A8	Diastolic blood pressure|Hypertension|Systolic blood pressure|Blood pressure|Body mass index|HDL cholesterol	rs13107325-T|rs13107325-T|rs13107325-T|rs13107325-T|rs13107325-T|rs13107325-T
	LZTS1	Dental caries|Dental caries|Major depressive disorder (broad)	rs4922199-?|rs10111661-?|rs1106634-A
	STMN2	Creutzfeldt-Jakob disease	rs1460163-A
	EIF3H	Corneal curvature|Colorectal cancer|Colorectal cancer	rs11987235-A|rs16892766-?|rs16892766-A
	SFTPD	Chronic obstructive pulmonary disease-related biomarkers|Chronic obstructive pulmonary disease-related biomarkers	rs7078012-T|rs3923564-G
	WDR41	Caudate nucleus volume	**rs163030-A**
	HLA-DPB1	Aspirin exacerbated respiratory disease in asthmatics|Hepatitis B (viral clearance)|Hepatitis B (viral	rs1042151-G|rs9277535-G|rs9277535-G|rs9277535-?|rs987870-?|rs9277535-

		clearance)|Hepatitis B vaccine response|Systemic sclerosis|Hepatitis B|Nephropathy|Hepatitis B	G|rs1883414-?|rs9277535-G
	LDHC	Amyloid A Levels	rs2896526-G
SKIN			
	LTBP3	Urate levels|Non-alcoholic fatty liver disease histology (lobular)	rs642803-T|rs6591182-A
	PAX8	Renal function-related traits (BUN)	**rs11123170-G**
	ACTC1	Prostate cancer (gene x gene interaction)|Refractive error	rs543686-?|rs634990-C
	ACCN1	Obesity-related traits|vWF and FVIII levels	rs17808461-A|rs1354492-A
	SOX5	Non-obstructive azoospermia|Systemic sclerosis|Systemic sclerosis|Response to statin therapy|Response to antipsychotic treatment|PR interval|AIDS	rs10842262-?|rs11047102-?|rs11047102-?|rs7979575-C|rs1464500-?|rs11047543-A|rs1522232-C
	HERC6	Metabolite levels	rs1440581-?
	WDR41	Caudate nucleus volume	**rs163030-A**
	PKD1L2	Attention deficit hyperactivity disorder and conduct disorder	rs4889240-T
	ZNF365	Atopic dermatitis|Crohn's disease|Breast size|Crohn's disease|Intelligence|Ewing sarcoma|Breast cancer|Mammographic density|Crohn's disease|Breast cancer|Crohn's disease	rs10995251-C|rs10761659-G|rs7089814-C|rs7076156-G|rs10995170-?|rs224278-C|rs10822013-T|rs10995190-A|rs10761659-G|rs10995190-G|rs10995271-C
	PECR	Alcohol dependence	rs7590720-G