Epistasis and decanalization shape gene expression variability in humans via distinct modes of action

Gang Wang[1,§], Ence Yang[1,2,§], Jizhou Yang[1], Beiyan Zhou[3], Yanan Tian[3], James J. Cai[1,4,*]

[1]Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77843, USA.

[2]Institute for Systems Biomedicine, School of Basic Medical Sciences, Peking University Health Science Center, Beijing 100191, China.

[3]Department of Veterinary Physiology and Pharmacology, Texas A&M University, College Station, TX 77843, USA.

[4]Interdisciplinary Program of Genetics, Texas A&M University, College Station, Texas, USA

[§]These authors contributed equally to this paper

Running title: GxG and GxE effects underlying expression variability QTLs

Keywords: gene expression, phenotypic variability, evQTL, epistasis, decanalization

Abstract

Increasing evidence shows that phenotypic variance is genetically controlled, and the variance itself is a quantitative trait. The precise mechanism of genetic control over the variance, however, remains to be determined. Here, using complex trait analysis of gene expression, we show that common genetic variation contributes to increasing gene expression variability via distinct modes of action—e.g., epistasis and decanalization. We focused on expression variability QTLs (evQTLs), i.e., genetic loci associated with gene express variance, in the human genome. We found that a quarter of evQTLs could be attributed to the presence of "third-party" eQTLs. These SNPs are associated with gene expression in a fraction, rather than the entire set, of samples. Many additional evQTLs do not interact with other SNPs and are thus unexplained by the epistasis model; these are attributable to the decanalizing effect of evQTL variants. Here we present the decanalization model, which predicts that evQTLs influence gene expression variability through modulating the sensitivity of transcriptional machinery to environmental perturbation. To validate the model we measured the discordant gene expression between monozygotic twins, and also estimated the amplitude of stochastic gene expression noise using repeated RT-qPCR assays on single samples. Both measures were found to be associated with genotypes of evQTLs explained by the decanalization model. Together, our results suggest that genetic variants work interactively or independently to influence gene expression variability. We anticipate our analysis to be a starting point for more sophisticated mechanistic analyses and opens a new, variability-centered research avenue for mapping complex traits.

Author Summary

It is increasingly appreciated that phenotypic variance is genetically controlled. The effects of genotypes on phenotypic variance represent a critical source of phenotypic differences among individuals. Here we study the mechanisms of genetic control of phenotypic variance through the lens of evQTLs in humans. We show that two distinct modes of action, namely epistasis and decanalization, contribute to the formation of evQTLs. Combining computational and experimental analyses, we show that gene expression variability in populations is correlated with the stochastic noise of gene expression in individuals. Such a pattern is more likely to be detected with evQTLs resulted from decanalizing variants, rather than epistatic interactions between variants. We demonstrate the decanalizing function conferred by evQTL variants by showing the increased discordant gene expression between monozygotic twins, as well as the noisier gene expression in cell lines. Our findings have implications for complex trait studies, calling for a paradigm shift in the methodology of disease risk loci mapping to take into consideration the impact of genetic variants on phenotypic variability.

Introduction

Phenotypic variability refers to the likelihood of the phenotypic variation being observed in a population. Quantitative genetics assumes that phenotypic variation, i.e., the difference in phenotypic mean between individuals, is genetically controlled [1]. Under such an assumption, phenotypic variation is explained solely by differences in phenotypic mean among genotypes. This deterministic view, however, has come under challenge. New studies show that phenotypic variance is genetically controlled, and the variance itself is a quantitative trait [2-13]. Increasing evidence of genetic control over the variance calls for a paradigm shift in quantitative genetics. Understanding the mechanism of how phenotypic variance is controlled is of great importance for evolutionary biology, agriculture or animal sciences, and medicine [5, 11, 14, 15]. In evolutionary biology, for example, variability offers an adaptive solution to environmental changes [16-18]. Genetic factors resulting in more variable phenotypes become favored when they enable a population to respond more effectively to environmental changes [19-22]. In medicine, diseased states emerge when the relevant phenotype of affected individuals goes beyond a threshold. As such, high variability genotypes will produce a larger proportion of individuals exceeding that threshold than will low variability genotypes, even if these genotypes have the same mean. By ignoring the effect of genotypes on phenotypic variance, an important axis of genetic variation contributing to phenotypic differences among individuals has been overlooked [1, 23]. The lack of empirical studies in this regard has hindered the discovery of variance associated mutations that contribute to modulating disease susceptibility and the phenotypic variability of other human health-related traits.

Several studies have been conducted to reveal gene expression variability, i.e., the differences in variance of gene expression between groups, in various systems [24-27]. Nevertheless, our understanding of how genetic diversity can control or influence gene expression variability remains limited. Promising new developments along this line have come from our findings in complex trait analysis of gene expression. Using variance-association mapping, we and others identified genetic loci associated with gene expression variance, called evQTLs [11, 12] or v-eQTL [9]. How evQTLs are created in the first place is not completely known. While the epistasis has been widely accepted as a mechanism introducing phenotypic variability, here we seek a straightforward explanation, that is, evQTL variants disrupt or stabilize the genetic architecture that buffers stochastic variation in gene expression. As a result of the decanalization, phenotypic expression becomes more sensitive to the external environment and varies more greatly [5, 11]. We reveal evQTLs with epistasis and decanalization, two distinct modes of action, on gene expression variability and lay down the foundation for a new analytical framework that accounts for the genetic contribution to phenotypic variability. We anticipate that methods derived from the new framework will allow us to identify novel causal loci, which would otherwise be missed by traditional mean-focused methods, in complex disease mapping.

Results

Widespread evQTLs in the human genome

We obtained the expression data of 15,124 protein-coding genes measured in 462 lymphoblastoid cell lines (LCLs) by the Geuvadis Project [28]. We also obtained the genotype data at 2,885,326 polymorphic sites determined in the same cell lines by the 1,000 Genomes Project [29]. After data processing, 326 LCL samples from unrelated individuals of European descent (EUR) were retained for this study (Materials and Methods). To identify evQTLs, we first applied the method based on the double generalized linear model (DGLM) [30]. The method has been previously adopted by us [11, 12] and others [5]. Owing to the computational complexity, we restricted the use of this method in the identification of *cis*-acting evQTLs. On average ~1800 SNPs that lay within 1-Mb radii of the transcription start site were tested per

gene. Using a conservative Bonferroni correction cutoff $P = 1.75\times10^{-9}$ (= 0.05 / 28,494,473), we identified a total of 17,949 *cis*-evQTLs in 1,304 unique genes, i.e., 8.6% of all genes tested (**Figure 1A, Table S1**). Next, to identify both *cis*- and *trans*-evQTLs genome-wide, we adopted the method based on the squared residual value linear model (SVLM)[9, 31]. It is a computationally efficient, two-stage method. The effect of variants on gene expression mean (i.e., eQLT effect) is firstly removed by regression, and the residuals are squared to give a measure of expression dispersion. Then the correlation between squared residuals and genotypes is tested. We applied SVLM to test all SNPs against all genes, without pre-filtering SNPs by their locational relationship with tested genes. Such an all-against-all strategy allowed a systematic survey of *cis*- and *trans*-evQTLs across the entire genome. We used the Benjamini-Hochberg procedure [32] to determine the *P*-value cutoff of 3×10^{-9} that gave the false-discovery rate (FDR) of 0.1. At this level, we identified 505 *cis*-evQTLs in 33 unique genes, and 1,008 *trans*-evQTLs in 235 unique genes (**Figure 1B, Table S2**). Two genes *AXIN2* and *FAM86B1* were found to have both *cis*- and *trans*-evQTLs. Applying the same FDR cutoff to detect both *cis*- and *trans*-evQTL resulted in an unbiased picture of the distribution of all evQTLs across autosomes (**Figure 1C**). Comparing the positions of genes and their evQTLs, we did not observe a strong enrichment of data points along the diagonal of the graph, suggesting *cis*-evQTLs not be particularly enriched compared to *trans*-evQTLs. We noticed the pronounced discrepancy in the number of *cis*-evQTLs detected using DGLM and SVLM. This discrepancy may be because that SVLM and DGLM have different detecting powerful. Computer simulations showed that, when the sample size was set to 300, SVLM method had only half of the power of DGLM (**Supplementary Fig. S1**). Furthermore, the huge multiple testing burden associated with the application of SVLM in the all-against-all tests may also contribute to the discrepancy.

Fig. 1. Overview of evQTL detections and the distribution of *cis*- and *trans*-evQTLs in autosomes. (**A**) Flowchart of *cis*-evQTLs identification using DGLM method. (**B**) Flowchart of *cis*- and *trans*-evQTL identification using SVLM method. (**C**) Distribution of SVLM-identified *cis*- and *trans*-evQTLs in autosomes.

Epistatic interactions contribute to increasing gene expression variability

Epistasis, i.e., the interaction between loci (GxG), may increase the phenotypic variability of a population [10, 33]. The evQTLs provided source materials for studying the epistatic effect on gene expression variability [12]. More specifically, we sought to identify “third-party” SNPs that interact with evQTL SNPs. Such interactions result in more variable gene expression of the evQTL genes. In particular, for each evQTL SNP identified by using SVLM, we applied a two-step procedure to identify the third-party SNPs, also known as *partial eQTL SNPs* (see below). These third-party SNPs interact or are partially associated with evQTL SNPs, resulting in the increased gene expression variance [9, 12]. The process of partial eQTL SNP identification is illustrated in **Supplementary Fig. S2**. Briefly, for a given evQTL (for example, the evQTL between gene *X* and SNP *Y*), we extracted samples with a homozygous genotype associated with large expression variance. We called these L group samples. Accordingly, those related to small expression variance was called S group samples. Then, we conducted a genome-wide scan among the extracted L group samples to identify eQTL SNPs (e.g., SNP *Z*) that control the expression of the corresponding evQTL gene (i.e., gene *X*). The identified eQTL SNPs are called *partial* because they are detected in the sub-sampled discovery panel, and their effect on gene expression is restricted to L group samples. The evQTL SNP *Y* and its partial eQTL SNP *Z* may be co-localized proximately on the same chromosome and partially associated as we showed previously [12]. They may also be unlinked, for instance, located on different chromosomes, and interact with each other epistatically [9]. Here, we focused on the 268

evQTLs (33 *cis*- and 235 *trans*-acting ones) identified by using SVLM. In 73 out of 268 evQTL genes, we identified at least one significant interacting SNP, i.e., partial eQTL SNP with simple linear regression test $P < 10^{-8}$ in the L group samples (**Table S3**). These results suggest that more than one-fourth of evQTLs are attributable to partial eQTL SNPs interacting with evQTL SNPs.

Decanalization contributes to increasing gene expression variability without genetic interactions

Here we put forward the decanalization model to explain the formation of evQTLs. The model emphasizes the interaction between gene (or genotype) and environment (GxE). Unlike the GxG model that concerns the epistatic interactions or associations between variants [9, 12], the decanalization (GxE) model concerns a single variant that perturbs stable genetic systems through decanalizing effect. We hypothesized that some evQTL SNPs are associated with gene expression variability because one of their two alleles confers the decanalization function, causing more variable gene expression. In other words, decanalizing SNPs increase gene expression variability via the single-locus effect, without interacting with any other SNPs. Thus, these decanalizing (GxE) evQTLs have a different formation mechanism in contrast to that of epistatic (GxG) evQTLs.

To show the decanalizing effect, by further controlling the diversity of samples' genetic backgrounds, we re-visited the genotype and expression data used in our previous study [12]. The data was derived from LCLs of a cohort of twin pairs [34]. In the previous study, we used a single set of the twin pairs for evQTL analysis and identified *cis*-evQTLs in 99 unique genes [12]. Here, we first classified the 99 evQTLs (between each gene and the most significant SNP) into 56 GxG and 43 GxE evQTLs. The classification was based on whether or not an interacting SNP (i.e., partial eQTL SNP) could be identified using the two-step procedure described above. The idea was that: if no interacting SNP can be detected for an evQTL, then the evQTL cannot be explained by the GxG model. Thus, the evQTL is likely to be a GxE evQTL, explained by the GxE model—the increased gene expression variability is be driven by the allele of evQTL SNPs with decanalizing function. Next, we extracted expression data of the 139 pairs of monozygotic (MZ) twins. We classified MZ twin pairs whose genotypes were homozygous at evQTL SNP sites into MZ-L or MZ-S group, according to whether their evQTL SNPs were associated with large or small variance. For all MZ twin pairs in the same group (either MZ-L or MZ-S), we estimated the discordant gene expression between two individuals of the same pairs. The discordant gene expression was calculated as the relative mean difference (RMD) in gene expression, which is the difference between two individual's gene expression values normalized by the mean (Materials and Methods).

To illustrate the difference in discordant gene expression between groups, we showed two example evQTLs. One is a GxE evQTL between *TBKBP1* and rs1912483 (**Fig. 2A**, left), and the other is a GxG evQTL between *PTER* and rs7913889 (**Fig. 2B**, left). The data points of gene expression levels were grouped by the genotype. Within each genotype category, data points from the same twin pairs are displayed side-by-side. Every two individuals of the same MZ pairs are linked by a line. The slope of the lines is an indicator of discordant gene expression between twin pairs. In the GxE evQTL example, the slopes between MZ twins with genotypes associated with large expression variance (i.e., MZ-L group) tend to be steeper than those with small expression variance (i.e., MZ-S group)(**Fig. 2A**, left). In contrast, in the GxG evQTL example, the difference in slope skewness between MZ-L and MZ-S groups is less pronounced (**Fig. 2B**, left). We pooled RMD values from different twin pairs together by MZ-L or MZ-S group and compared the distributions of RMD values between the two groups. For GxE evQTLs, the distributions of RMD values between L and S groups were significantly different ($P = 1.3\times10^{-5}$,

Fig. 2A, right), with larger RMD values for L group. In contrast, for GxG evQTLs, this difference in RMD distribution was not detected between L and S groups (P = 0.052, **Fig. 2B**, right).

Fig. 2. Dissection of GxE and GxG effects of evQTLs using twins data. (**A**) An example of GxE evQTL, *TBKBP1*-rs1912483. The expression data points for each of two individuals from the same pairs of MZ twins are linked. Twin pairs are grouped as MZ-L and MZ-S based on whether the homozygous genotype at rs1912483 is associated with large or small gene expression variance. The right panel shows the CDF of normalized discordant gene expression (measured using RMD) for MZ-S and MZ-L groups. (**B**) Same as (**A**) but showing an example of GxG evQTL, *PTER*-rs7913889.

Decanalizing evQTL SNPs are associated with gene expression noise

Our decanalization model works by the action of a single genetic variant conferring the decanalizing effect on gene expression. One of the underlying sources of the gene expression variability is stochastic noise in gene expression [35]. We hypothesized that different alleles of a GxE evQTL SNP might be associated with different levels of expression noise of the corresponding evQTL gene. To test this hypothesis, we set out to estimate the expression noise using RT-qPCR by repeatedly measuring gene expression level in the same cell line multiple times. If our hypothesis is true, then the expression variance of individual with an evQTL genotype associated with larger variance should be more pronounced than the expression variance in individual with genotype with smaller variance.

We selected two GxE evQTLs: *ATMIN*-rs1018804 and *BEND4*-rs7659929, for testing. *ATMIN* is an essential cofactor for checkpoint kinase ATM, which transduces genomic stress signals to halt cell cycle progression and promote DNA repair [36]. We picked two LCLs, HG00097 and HG00364, which have the similar *ATMIN* expression level. Both were derived from female individuals of European descendant. The difference is that HG00097's genotype CC at rs1018804 is associated with larger variance, while HG00364's genotype AA at rs1018804 is associated with smaller variance. Thus, HG00097 and HG00364 belonged to L- and S-groups, respectively. We measured the evQTL gene expression level using RT-qPCR with three technical replicates each at five different sampling time points. The same assay was repeated three times independently. Our results showed that, under the same controlled experimental condition, the variance of gene expression (i.e., the variance in ΔCt values) in HG00097 was greater than HG00364. The same trend was observed from all three biological replicates (**Fig. 3A**). In two replicates, the difference was statistically significant (F-test $P < 0.05$).

We repeated the experiment with two biological replicates on the same evQTL *ATMIN*-rs1018804 using a different pair of LCLs (NA12144 and NA12736 from L- and S-group, respectively) to replace HG00097 and HG00364. We obtained the similar results showing a consistent pattern, that is, the gene expression in the cell line of L-group is more variable than that of S-group (**Supplementary Fig. S3**). Furthermore, we repeated the experiment on a different GxE evQTL (*BEND4*-rs7659929) with another pair of LCLs (NA12889 and NA18858). Again, we obtained the consistent pattern that supports the correlation between gene expression variability and stochastic noise (**Supplementary Fig. S3**).

Fig. 3. The correlation between gene expression variability and noise presents in the GxE evQTL, *ATMIN*-rs1018804, but not in the GxG evQTL, *ZNF10*-rs7972363, in the same cell line pair (HG00097 and HG00364). (**A**) The most left panel shows the distribution of gene expression levels of *ATMIN* among three different genotypes defined by two alleles of

rs1018804. Red arrows indicate the genotype and expression level of HG00097 and HG00364. Right panels show the results of three biological replicates of repeated RT-qPCR analysis for *ATMIN* at five different time points ranging from 12 to 60 h after incubation. At each time point of each biological replicate, three technical replicates are performed to obtain ΔCt values. Red circles indicate the average ΔCt values. The difference in variance of ΔCt between two cell lines was tested using *F*-test for equal variances (P = 0.429, 0.012, and 0.049, respectively, for the three replicates). (**B**) Same as (**A**) but showing the results of evQTL *ZNF10*-rs7972363. *P* values of *F*-test for the three replicates are 0.981, 0.195 and 0.066, respectively.

We hypothesized that the correlation between gene expression variability and noise exists exclusively in GxE evQTLs. We did not expect such a correlation could be recapitulated in GxG evQTLs. This is because the two kinds of evQTLs work through different modes of action. To test this, we repeated the same RT-qPCR experiment with a GxG evQTL ZNF10-rs7972363 using the same cell lines HG00097 and HG00364 (**Fig. 3B**). The genotype AA of HG00097 at rs7972363 is associated with larger variance while the genotype GG of HG00364 is associated with smaller variance. As a GxG evQTL, the interacting SNP rs1567910, which interacts with rs7972363 and helps the creation of the evQTL, has been identified. Samples with AA genotype at rs7972363 can be further broken down by rs1567910 into three subgenotype groups associated with different levels of gene expression mean. Consistent with our expectation, the gene expression variance in ΔCt values was similar between HG00097 and HG00364, samples from the L and S group, respectively (**Fig. 3B**). Together, our results suggest that the level of gene expression noise—the random fluctuation of gene expression—is associated with GxE evQTL, but not GxG, SNPs.

Differences in cell cycle status and alternative splicing do not account for the decanalizing function conferred by GxE evQTL SNPs

Finally, we controlled for two additional confounding factors that might account for the increased gene expression variability associated with evQTLs. The first one is the cell cycle status of cell lines. At the same sampling time, cell lines may differ in the percentage or number of cells in different cell cycle phases. Could the difference in cell cycle status explain the difference in gene expression variability or noise between cell lines? To test this, we performed the cell cycle analysis by flow cytometry with HG00097 and HG00364 at 36 h after incubation (Materials and Methods). The results showed no difference in the percentage of cells in G0/G1, S and G2/M phases between the two cell lines (**Supplementary Fig. S4**). The second confounding factor we considered is the alternative splicing pattern. Different splicing patterns between cell lines might result in different gene-level expression measurements. We used the Integrative Genomics Viewer [37] to visualize the alternatively spliced mRNA of *ATMIN* and compared the pattern of splicing between HG00097 and HG00364, as well as that of *BEND4* between NA12889 and NA18858. In either case, we observed no difference in splicing patterns (**Supplementary Fig. S5**).

Discussion

Variability, which refers to the potential of a population to vary, is a central concept in biology [38]. Emerging experimental and statistical techniques have allowed the variability regarding various phenotypes to be rigorously analyzed [14]. Focusing on the variability QTLs of gene expression, we found that evQTLs are abundant and widespread across the human genome [11, 12]. In the light of evQTLs, the present study reveals two distinct modes of action: epistasis and decanalization, through which common genetic variation control or influence gene expression variability. The epistasis model concerns two or more variants, which interact in non-

additive fashion [9, 39] or link to each other through incomplete linkage disequilibrium [12, 40]. In line with this model, a number of methods for identifying epistasis have been proposed, based on detecting the increased variability [10, 33, 41]. The decanalization model is simpler and more direct, concerning single variants that work alone to destabilize the phenotypic expression and pushing a proportion of individuals away from the robust optimum.

Dissecting the GxG and GxE effects, respectively underlying the epistatic and decanalizing modes of action, in the context of variability QTLs is technically challenging. Here we have taken advantage of the identical genetic background of MZ twins and showed that different genotypes are associated with varying degrees of responsiveness to environmental perturbation. We also detected the unexpected link between the population-level gene expression variability and the stochastic gene expression noise measured in single individuals. It suggests that variable gene expression in each sample may be synthesized and aggregated together and eventually contribute to the gene expression variability of the population as a whole. In other words, the same underlying force destabilizing gene expression might be proposed to be a unified explanation for gene expression variability at different scales (i.e., from the population level to the individual level). To the other end of the spectrum, the cell-to-cell variability in gene expression could be examined, thanks to the rapid development of single-cell based technologies [42]. For example, the genetic control of the variability in burst size and frequency of single-cell transcription may not be too different from that of the population and individual levels.

We were unable to provide the precise mechanism of decanalizing function conferred by evQTL variants. However, we were able to utilize bioinformatics analysis to provide a rationale to substantiate the link between the variants, the possible genetic mechanisms, and the phenotype, i.e., expression variability of the corresponding gene. By synthesizing different sources of information, we attempted to build working models for evQTLs, explaining how evQTL variants can influence gene expression variability. Here we use GxE evQTL *ATMIN*-rs1018804 as an example to illustrate one of the tentative models. Rs1018804 is associated with *WDR24* that encodes WD repeat-containing protein 24, a key component of Rag-interacting complex essential for the activation of mTORC1 [43]. The intronic rs1018804, 43-bp downstream from the nearest exon-intron boundary, may play a role in regulating the splicing of *WDR24* mRNA. The WDR24 protein is predicted [44] to interact with Hsp70 and DNAJ proteins [45]. The latter two interact with the dynein light chain DYNLL1 [46]. Finally, *DYNLL1* and *ATMIN* form a feedback regulation loop—one of few known examples of negative auto-regulation of gene expression where a gene product directly inhibits the main transcriptional activator while bound at its own promoter [47]. Taken together, the working model can be represented as rs1018804 → WDR24 → Hsp70/DNAJ → *DYNLL1* ↔ *ATMIN*. This working model offers a workable blueprint for functional dissection of all components involved. The information flow provides new insights into the potential mechanisms of evQTL variants influencing gene expression variability.

We anticipate that our results have implications for studying human diseases, in which the regulatory variation plays critical roles [48]. Decanalization effect has been proposed to influence brain development and contribute to the risk of psychological disorders like schizophrenia [49, 50] and other complex diseases [15]. Increased gene expression variability was found to be associated with the aging in a mouse model [51], and the aggressiveness of lymphocytic leukemia [52]. Understanding how genetic variation contributes to increasing gene expression variability or variability of other phenotypic traits will facilitate the identification of causal variants. This is especially true when gene expression heterogeneity characterizes the disease under consideration. Indeed, many human diseases are characterized by etiological and phenotypic heterogeneity, echoing the so-called “Anna Karenina principle,” that is, each

sick person is sick in his or her own way. Even merely a small fraction of the increased phenotypic variability among patients is due to the variability-controlling mutations (such as evQTL variants), understanding how these mutations influence the variability is still of importance. The better understanding of the control may bring us close to causal mutations underlying individual's predisposition to disease. This strategy, if combined with other methods for estimating the impact of rare mutations, such as aberrant gene expression analysis for private mutations [53], would be more powerful for personalized medicine. Furthermore, we suggest that variability-controlling mutations are potential targets for genomic editing or drug development. Drug targeting these mutations might bring the dysregulated and dysfunctional gene expression in patients back to normal.

Materials and Methods

Gene expression and genotype data for evQTL analysis

The gene expression data generated by the Geuvadis project RNA-seq study [28] was downloaded from the website of EBI ArrayExpress using accession E-GEUV-1. The downloaded data matrix contained the expression values of Gencode (v12)-annotated genes measured in 462 unique LCL samples. The data was normalized by using the method of probabilistic estimation of expression residuals (PEER)[54]. From the data matrix, we extracted the expression values of autosomal protein-coding genes of 345 EUR samples, whose genotype data is available from the website of the 1,000 Genomes Project [29]. Based on the result of a principal component analysis, we excluded 19 samples whose global expression profile apparently deviated from those of the rest of samples. The final data matrix used for the evQTL analysis contained gene expression values of 15,124 protein-coding genes and 326 EUR samples. Also, we obtained the genotype and expression data from a cohort of female twin pairs [34] from the TwinsUK adult twin registry [55]. The data of gene expression in LCLs of 139 pairs of MZ twins was extracted and used in this study.

Identification of evQTLs

We adopted the DGLM method [30] to test for inequality in expression variances and measure the contribution of genetic variants to the expression heteroscedasticity. As did before, 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 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 also used the SVLM procedure [56] to detect evQTLs. The SVLM method consists of two steps. First, a regression analysis is applied where the trait is adjusted for a possible SNP effect and other covariates. Second, regression analysis is applied to the squared values of residuals obtained from the first stage, using the SNP as the predictor.

Identification of partial eQTL SNPs that interact with evQTL SNPs

We used a two-step procedure to identify SNPs that interact with evQTLs. We first partitioned individuals into L and S groups according to whether genotypes of the evQTL SNP are associated with large (L) and small (S) variances of gene expression. Then we scanned genome-wide SNPs. For each SNP, the eQTL analysis by linear regression model was conducted among individuals of the L group. For each top SNP with high genotype heterozygosity difference, a linear regression [57] was performed on the SNP's genotypes and gene expression. The most significant SNPs were retained after applying an arbitrary *P*-value = 0.0005 as cutoff and were reported as candidate interacting SNPs.

Estimation of gene expression noise using repeated RT-qPCR assay

LCLs were purchased from the Coriell Institute (https://catalog.coriell.org/). The cells were maintained in Roswell Park Memorial Institute Medium 1640 with 2mM L-glutamine and 15% FBS (Seradigm) at 37°C in a humidified atmosphere containing 5% CO_2 (v/v). For the time course experiment, cell lines were seeded at 1×10^6 cells per 10 cm dish and then incubated in the culture medium. Cell lines were screened to ensure to be mycoplasma free by using the MycoFluor mycoplasma detection kit (Invitrogen). Cells were collected at 24, 36, 48, 60, and 72 h after growth. Total RNA was extracted using Trizol reagent (Invitrogen). RNase-free DNase (Ambion) was used to remove potential contaminating DNA from RNA samples. RNA purity and concentration were determined using Nanodrop ND-100 Spectrophotometer. The concentrations of total RNA were adjusted to 100 µg/ml. Real-time RT-PCR assays were performed using iTaq Universal SYBR Green One-Step Kit (Bio-Rad Laboratories) with primers shown in **Table S4**. Template total RNA was reverse transcribed and amplified in a Bio-Rad CFX96 Real-Time PCR Detection System (Bio-Rad Laboratories) in 20-µl reaction mixtures containing 10 µl of iTaq universal SYBR Green reaction mix (2×), 0.25 µl of iScript reverse transcriptase, 2 µl of 100 nM of forward and reverse primers mix, 1 µl of total RNA template, and 6.75 µl of nuclease-free water, at 50°C for 10 min, 95°C for 1 min, followed by 30 cycles of 95°C for 10 s and 58°C for 30 seconds. Melting curves were measured from 65°C to 95°C with 0.5°C of increment. The average expression of two housekeeping genes (*CHMP2A* and *C1orf43*) was used for normalization. The choice of using these two genes as reference was based on recent scrutiny of human genes with a constant level of expression using RNA-seq data [58].

Flow cytometric analysis of cells in different phases of the cell cycle

Cell cycle distribution was evaluated by using flow cytometry. This determination was based on the measurement of the DNA content of nuclei labeled with propidium iodide [59]. Cells were harvested at 24, 36, 48, 60, and 72 h after treatment. The cells were resuspended at a concentration of 1×10^6/ml in cold PBS. After 1ml of ice-cold 100% ethanol had been added dropwise, the cells were fixed at 4°C for at least 16 hours. The fixed cells were pelleted, resuspended in 1ml of propidium iodide (PI) staining solution (50 mg/ml propidium iodide, 100 units/ml RNase A in PBS) for at least 1 hour at room temperature and analyzed on a FACS flow cytometer (BD). By using red propidium-DNA fluorescence, 30,000 events were acquired. The percentage of cells in G0/G1, S and G2/M phases of the cell cycle was calculated using the Flowjo software v10 (Tree Star).

Acknowledgements

We thank Dr. Guan Zhu for valuable discussion and technical assistance. We thank Jinting Guan for help with computer simulations, Srikanth Kanameni for help with flow cytometry, and Dr. Vijayanagaram Venkatraj for help with cell line preparation. We acknowledge the Texas A&M Institute for Genome Sciences and Society (TIGSS) 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/Computerized Infectious Disease Reporting (CIDR).

References

1. Lynch M, Walsh B. Genetics and analysis of quantitative traits. Sunderland, Mass.: Sinauer; 1998. xvi, 980 p. p.
2. Yang J, Loos RJ, Powell JE, Medland SE, Speliotes EK, Chasman DI, et al. FTO genotype is associated with phenotypic variability of body mass index. Nature. 2012;490(7419):267-72. doi: 10.1038/nature11401. PubMed PMID: 22982992; PubMed Central PMCID: PMC3564953.
3. Jimenez-Gomez JM, Corwin JA, Joseph B, Maloof JN, Kliebenstein DJ. Genomic analysis of QTLs and genes altering natural variation in stochastic noise. PLoS genetics. 2011;7(9):e1002295. doi: 10.1371/journal.pgen.1002295. PubMed PMID: 21980300; PubMed Central PMCID: PMC3183082.
4. Ansel J, Bottin H, Rodriguez-Beltran C, Damon C, Nagarajan M, Fehrmann S, et al. Cell-to-cell stochastic variation in gene expression is a complex genetic trait. PLoS genetics. 2008;4(4):e1000049. doi: 10.1371/journal.pgen.1000049. PubMed PMID: 18404214; PubMed Central PMCID: PMC2289839.
5. Ronnegard L, Valdar W. Detecting major genetic loci controlling phenotypic variability in experimental crosses. Genetics. 2011;188(2):435-47. doi: 10.1534/genetics.111.127068. PubMed PMID: 21467569; PubMed Central PMCID: PMC3122324.
6. Metzger BP, Yuan DC, Gruber JD, Duveau F, Wittkopp PJ. Selection on noise constrains variation in a eukaryotic promoter. Nature. 2015. doi: 10.1038/nature14244. PubMed PMID: 25778704.
7. Shen X, Pettersson M, Ronnegard L, Carlborg O. Inheritance beyond plain heritability: variance-controlling genes in Arabidopsis thaliana. PLoS genetics. 2012;8(8):e1002839. doi: 10.1371/journal.pgen.1002839. PubMed PMID: 22876191; PubMed Central PMCID: PMC3410891.
8. Ayroles JF, Buchanan SM, O'Leary C, Skutt-Kakaria K, Grenier JK, Clark AG, et al. Behavioral idiosyncrasy reveals genetic control of phenotypic variability. Proc Natl Acad Sci U S A. 2015;112(21):6706-11. doi: 10.1073/pnas.1503830112. PubMed PMID: 25953335; PubMed Central PMCID: PMC4450409.
9. Brown AA, Buil A, Vinuela A, Lappalainen T, Zheng HF, Richards JB, et al. Genetic interactions affecting human gene expression identified by variance association mapping. Elife. 2014;3:e01381. doi: 10.7554/eLife.01381. PubMed PMID: 24771767; PubMed Central PMCID: PMC4017648.
10. Pare G, Cook NR, Ridker PM, Chasman DI. 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 genetics. 2010;6(6):e1000981. doi: 10.1371/journal.pgen.1000981. PubMed PMID: 20585554; PubMed Central PMCID: PMC2887471.
11. Hulse AM, Cai JJ. Genetic variants contribute to gene expression variability in humans. Genetics. 2013;193(1):95-108. doi: 10.1534/genetics.112.146779. PubMed PMID: 23150607; PubMed Central PMCID: PMC3527258.
12. Wang G, Yang E, Brinkmeyer-Langford CL, Cai JJ. Additive, epistatic, and environmental effects through the lens of expression variability QTL in a twin cohort. Genetics. 2014;196(2):413-25. doi: 10.1534/genetics.113.157503. PubMed PMID: 24298061; PubMed Central PMCID: PMC3914615.
13. Fraser HB, Schadt EE. The quantitative genetics of phenotypic robustness. PloS one. 2010;5(1):e8635. doi: 10.1371/journal.pone.0008635. PubMed PMID: 20072615; PubMed Central PMCID: PMCPMC2799522.
14. Geiler-Samerotte KA, Bauer CR, Li S, Ziv N, Gresham D, Siegal ML. The details in the distributions: why and how to study phenotypic variability. Curr Opin Biotechnol.

2013;24(4):752-9. doi: 10.1016/j.copbio.2013.03.010. PubMed PMID: 23566377; PubMed Central PMCID: PMC3732567.
15. Gibson G. Decanalization and the origin of complex disease. Nature reviews Genetics. 2009;10(2):134-40. doi: 10.1038/nrg2502. PubMed PMID: 19119265.
16. Queitsch C, Sangster TA, Lindquist S. Hsp90 as a capacitor of phenotypic variation. Nature. 2002;417(6889):618-24. doi: 10.1038/nature749. PubMed PMID: 12050657.
17. Gibson G, Wagner G. Canalization in evolutionary genetics: a stabilizing theory? Bioessays. 2000;22(4):372-80. doi: 10.1002/(SICI)1521-1878(200004)22:4<372::AID-BIES7>3.0.CO;2-J. PubMed PMID: 10723034.
18. Wolf L, Silander OK, van Nimwegen E. Expression noise facilitates the evolution of gene regulation. Elife. 2015;4. doi: 10.7554/eLife.05856. PubMed PMID: 26080931; PubMed Central PMCID: PMC4468965.
19. Zhang Z, Qian W, Zhang J. Positive selection for elevated gene expression noise in yeast. Mol Syst Biol. 2009;5:299. Epub 2009/08/20. doi: 10.1038/msb.2009.58. PubMed PMID: 19690568; PubMed Central PMCID: PMC2736655.
20. Acar M, Mettetal JT, van Oudenaarden A. Stochastic switching as a survival strategy in fluctuating environments. Nat Genet. 2008;40(4):471-5. Epub 2008/03/26. doi: 10.1038/ng.110. PubMed PMID: 18362885.
21. Kaern M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: from theories to phenotypes. Nat Rev Genet. 2005;6(6):451-64. Epub 2005/05/11. doi: 10.1038/nrg1615. PubMed PMID: 15883588.
22. Hill WG, Zhang XS. Effects on phenotypic variability of directional selection arising through genetic differences in residual variability. Genet Res. 2004;83(2):121-32. Epub 2004/06/29. PubMed PMID: 15219157.
23. Hill WG, Mulder HA. Genetic analysis of environmental variation. Genet Res (Camb). 2010;92(5-6):381-95. doi: 10.1017/S0016672310000546. PubMed PMID: 21429270.
24. Mar JC, Matigian NA, Mackay-Sim A, Mellick GD, Sue CM, Silburn PA, et al. Variance of gene expression identifies altered network constraints in neurological disease. PLoS genetics. 2011;7(8):e1002207. doi: 10.1371/journal.pgen.1002207. PubMed PMID: 21852951; PubMed Central PMCID: PMC3154954.
25. Ho JW, Stefani M, dos Remedios CG, Charleston MA. Differential variability analysis of gene expression and its application to human diseases. Bioinformatics. 2008;24(13):i390-8. doi: 10.1093/bioinformatics/btn142. PubMed PMID: 18586739; PubMed Central PMCID: PMC2718620.
26. Campbell MG, Kohane IS, Kong SW. Pathway-based outlier method reveals heterogeneous genomic structure of autism in blood transcriptome. BMC Med Genomics. 2013;6:34. doi: 10.1186/1755-8794-6-34. PubMed PMID: 24063311; PubMed Central PMCID: PMC3849321.
27. Hasegawa Y, Taylor D, Ovchinnikov DA, Wolvetang EJ, de Torrente L, Mar JC. Variability of Gene Expression Identifies Transcriptional Regulators of Early Human Embryonic Development. PLoS genetics. 2015;11(8):e1005428. doi: 10.1371/journal.pgen.1005428. PubMed PMID: 26288249.
28. Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature. 2013;501(7468):506-11. doi: 10.1038/nature12531. PubMed PMID: 24037378; PubMed Central PMCID: PMC3918453.
29. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491(7422):56-65. doi: 10.1038/nature11632. PubMed PMID: 23128226; PubMed Central PMCID: PMC3498066.
30. Verbyla AP, Smyth GK. Double generalized linear models: approximate residual maximum likelihood and diagnostics. University of Adelaide: Department of Statistics, 1998.

31. Struchalin MV, Amin N, Eilers PH, van Duijn CM, Aulchenko YS. An R package "VariABEL" for genome-wide searching of potentially interacting loci by testing genotypic variance heterogeneity. BMC Genet. 2012;13:4. doi: 10.1186/1471-2156-13-4. PubMed PMID: 22272569; PubMed Central PMCID: PMC3398297.
32. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):289-300. doi: 10.2307/2346101.
33. Struchalin MV, Dehghan A, Witteman JC, van Duijn C, Aulchenko YS. Variance heterogeneity analysis for detection of potentially interacting genetic loci: method and its limitations. BMC Genet. 2010;11:92. doi: 10.1186/1471-2156-11-92. PubMed PMID: 20942902; PubMed Central PMCID: PMC2973850.
34. Grundberg E, Small KS, Hedman AK, Nica AC, Buil A, Keildson S, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat Genet. 2012;44(10):1084-9. doi: 10.1038/ng.2394. PubMed PMID: 22941192; PubMed Central PMCID: PMC3784328.
35. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science. 2002;297(5584):1183-6. doi: 10.1126/science.1070919. PubMed PMID: 12183631.
36. Kanu N, Behrens A. ATMINistrating ATM signalling: regulation of ATM by ATMIN. Cell Cycle. 2008;7(22):3483-6. PubMed PMID: 19001856.
37. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178-92. doi: 10.1093/bib/bbs017. PubMed PMID: 22517427; PubMed Central PMCID: PMC3603213.
38. Hallgrímsson B, Hall BK. Variation. Amsterdam ; Boston: Elsevier Academic Press; 2005. xxi, 568 p. p.
39. Hemani G, Shakhbazov K, Westra HJ, Esko T, Henders AK, McRae AF, et al. Detection and replication of epistasis influencing transcription in humans. Nature. 2014;508(7495):249-53. doi: 10.1038/nature13005. PubMed PMID: 24572353; PubMed Central PMCID: PMC3984375.
40. Wood AR, Tuke MA, Nalls MA, Hernandez DG, Bandinelli S, Singleton AB, et al. Another explanation for apparent epistasis. Nature. 2014;514(7520):E3-5. doi: 10.1038/nature13691. PubMed PMID: 25279928.
41. Daye ZJ, Chen J, Li H. High-Dimensional Heteroscedastic Regression with an Application to eQTL Data Analysis. Biometrics. 2012;68(1):316-26. doi: 10.1111/j.1541-0420.2011.01652.x. PubMed PMID: 22547833; PubMed Central PMCID: PMC3218221.
42. Dey SS, Foley JE, Limsirichai P, Schaffer DV, Arkin AP. Orthogonal control of expression mean and variance by epigenetic features at different genomic loci. Mol Syst Biol. 2015;11(5):806. doi: 10.15252/msb.20145704. PubMed PMID: 25943345.
43. Bar-Peled L, Chantranupong L, Cherniack AD, Chen WW, Ottina KA, Grabiner BC, et al. A Tumor suppressor complex with GAP activity for the Rag GTPases that signal amino acid sufficiency to mTORC1. Science. 2013;340(6136):1100-6. doi: 10.1126/science.1232044. PubMed PMID: 23723238; PubMed Central PMCID: PMCPMC3728654.
44. Wiles AM, Doderer M, Ruan J, Gu TT, Ravi D, Blackman B, et al. Building and analyzing protein interactome networks by cross-species comparisons. BMC Syst Biol. 2010;4:36. doi: 10.1186/1752-0509-4-36. PubMed PMID: 20353594; PubMed Central PMCID: PMCPMC2859380.
45. Kampinga HH. Molecular biology: It takes two to untangle. Nature. 2015;524(7564):169-70. doi: 10.1038/nature14640. PubMed PMID: 26245378.
46. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, et al. A human protein-protein interaction network: a resource for annotating the proteome. Cell. 2005;122(6):957-68. doi: 10.1016/j.cell.2005.08.029. PubMed PMID: 16169070.
47. Jurado S, Conlan LA, Baker EK, Ng JL, Tenis N, Hoch NC, et al. ATM substrate Chk2-interacting Zn2+ finger (ASCIZ) Is a bi-functional transcriptional activator and feedback sensor

in the regulation of dynein light chain (DYNLL1) expression. J Biol Chem. 2012;287(5):3156-64. doi: 10.1074/jbc.M111.306019. PubMed PMID: 22167198; PubMed Central PMCID: PMC3270970.
48. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. Nature reviews Genetics. 2015;16(4):197-212. doi: 10.1038/nrg3891. PubMed PMID: 25707927.
49. McGrath JJ, Hannan AJ, Gibson G. Decanalization, brain development and risk of schizophrenia. Transl Psychiatry. 2011;1:e14. doi: 10.1038/tp.2011.16. PubMed PMID: 22832430; PubMed Central PMCID: PMC3309463.
50. Burrows EL, Hannan AJ. Decanalization mediating gene-environment interactions in schizophrenia and other psychiatric disorders with neurodevelopmental etiology. Front Behav Neurosci. 2013;7:157. doi: 10.3389/fnbeh.2013.00157. PubMed PMID: 24312026; PubMed Central PMCID: PMC3826253.
51. Bahar R, Hartmann CH, Rodriguez KA, Denny AD, Busuttil RA, Dolle ME, et al. Increased cell-to-cell variation in gene expression in ageing mouse heart. Nature. 2006;441(7096):1011-4. doi: 10.1038/nature04844. PubMed PMID: 16791200.
52. Ecker S, Pancaldi V, Rico D, Valencia A. Higher gene expression variability in the more aggressive subtype of chronic lymphocytic leukemia. Genome Med. 2015;7(1):8. doi: 10.1186/s13073-014-0125-z. PubMed PMID: 25632304; PubMed Central PMCID: PMC4308895.
53. Zeng Y, Wang G, Yang E, Ji G, Brinkmeyer-Langford CL, Cai JJ. Aberrant gene expression in humans. PLoS genetics. 2015;11(1):e1004942. doi: 10.1371/journal.pgen.1004942. PubMed PMID: 25617623; PubMed Central PMCID: PMC4305293.
54. Stegle O, Parts L, Durbin R, Winn J. A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. PLoS Comput Biol. 2010;6(5):e1000770. doi: 10.1371/journal.pcbi.1000770. PubMed PMID: 20463871; PubMed Central PMCID: PMC2865505.
55. Moayyeri A, Hammond CJ, Hart DJ, Spector TD. The UK Adult Twin Registry (TwinsUK Resource). Twin Res Hum Genet. 2013;16(1):144-9. doi: 10.1017/thg.2012.89. PubMed PMID: 23088889; PubMed Central PMCID: PMC3927054.
56. Struchalin MV, Amin N, Eilers PHC, van Duijn CM, Aulchenko YS. An R package "VariABEL" for genome-wide searching of potentially interacting loci by testing genotypic variance heterogeneity. BMC Genet. 2012;13. PubMed PMID: WOS:000301045200001.
57. Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, Lyle R, et al. Genome-wide associations of gene expression variation in humans. Plos Genet. 2005;1(6):e78. doi: 10.1371/journal.pgen.0010078. PubMed PMID: 16362079; PubMed Central PMCID: PMC1315281.
58. Eisenberg E, Levanon EY. Human housekeeping genes, revisited. Trends Genet. 2013;29(10):569-74. doi: DOI 10.1016/j.tig.2013.05.010. PubMed PMID: WOS:000325300800005.
59. Vindelov LL, Christensen IJ. A Review of Techniques and Results Obtained in One Laboratory by an Integrated System of Methods Designed for Routine Clinical Flow Cytometric DNA Analysis. Cytometry. 1990;11(7):753-70. PubMed PMID: WOS:A1990DZ69000001.

Supplementary Figure Legends

Fig. S1. Comparison of statistical power of two evQTL detection methods: DGLM and SVLM, using computer simulations with different sample sizes. For simulations, a population of 10,000 individuals were generated, and the MAF of an evQTL SNP was set to 0.4. The genotypes of SNP were encoded to 0, 1, 2 for homozygous minor, heterozygous, and homozygous major alleles, respectively. The gene expression of each genotype was generated from a normal distribution with the same mean but different variances, 1.0, 2.0, and 4.0, respectively. Before testing a method, the population was subsampled to the designated sample size, ranging from 300 to 1,000. For each sample size, the tested method was applied to the subsamples. The whole procedure was repeated 1,000 times, and the power was computed as the ratio of the times of *P*-value being smaller than 5×10^{-5} (i.e., 0.05/1000).

Fig. S2. Schematic illustration of the method for identifying partial eQTLs. After the identification of evQTL, the partial eQTL method involves two steps: (1) extraction of homozygous individuals whose genotype of the evQTL variant is associated with increased expression variability, and (2) identification of the eQTL between the gene and third-party variant among extracted individuals.

Fig. S3. The correlation between gene expression variability and noise presents in two additional GxE evQTLs. (**A**) GxE evQTL ATMIN-rs1018804 gene expression in the cell line pair NA12144 and NA12763. The most left panel shows the distribution of gene expression levels of ATMIN among three different genotypes defined by two alleles of rs1018804. Red arrows indicate the expression levels of NA12144 and NA12736 and their genotypes. Right panels show the results of two biological replicates of repeated RT-qPCR analysis for ATMIN at five different time points at 24, 36, 48, 60, and 72 h after incubation. At each time point of each biological replicate, three technical replicates were performed to obtain ΔCt values, and the average is presented by the red circle. (**B**) Same as (**A**) but showing the results of evQTL BEND4-rs7659929 using cell line pair NA12889 and NA18858.

Fig. S4. Cell cycle analysis to determine the relative abundance of cells in different phases. (**A**) Representative flow cytometric dot plots. (**B**) Representative histograms obtained using TUNEL assay. (**C**) Relative frequencies of cells in G1, S, and G2 phases. (**D**) PCA of cell cycle profiles. (**E**) Relative frequencies of cells in different phases of HG00097 (red) and HG00364 (blue).

Fig. S5. IGV view of RNA-seq read alignments and sashimi plot of mRNA splicing patterns of evQTL genes in different cell lines. (**A**) IGV view of RNA-seq read alignment of *ATMIN* in HG00097 and HG00364. (**B**) Sashimi plots of *ATMIN* mRNA in HG00097 and HG00364. (**C**) IGV view of RNA-seq read alignment of *BEND4* in NA12889 and NA18858. (**D**) Sashimi plots of *BEND4* mRNA in NA12889 and NA18858.

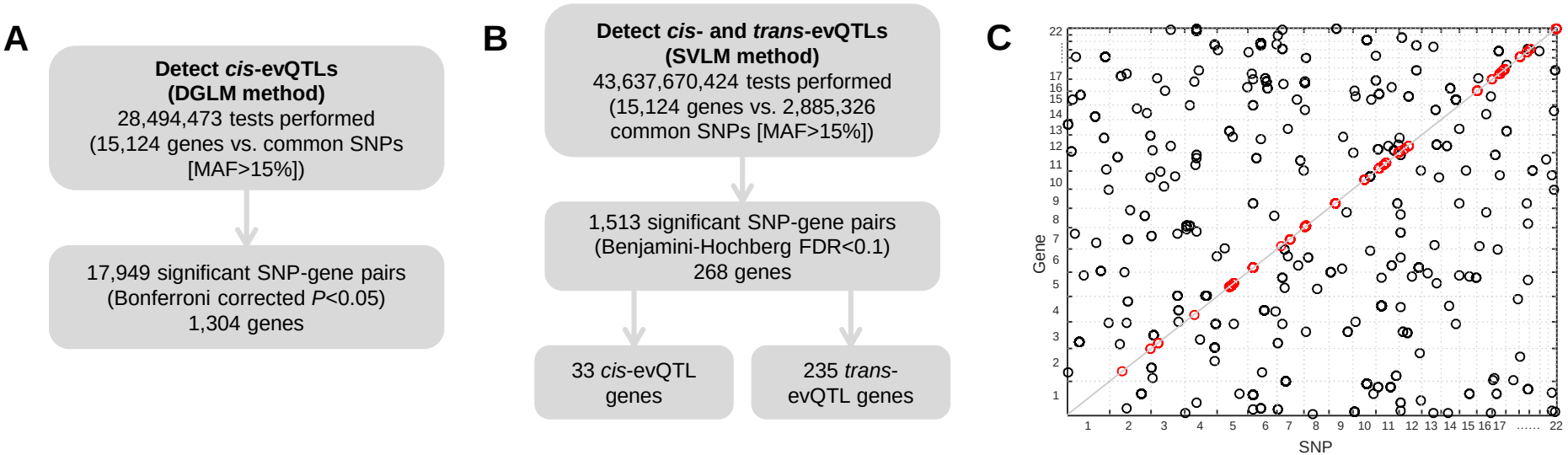
Supplementary Table Legends.

Table S1. The list of *cis*-evQTLs identified using the DGLM method.

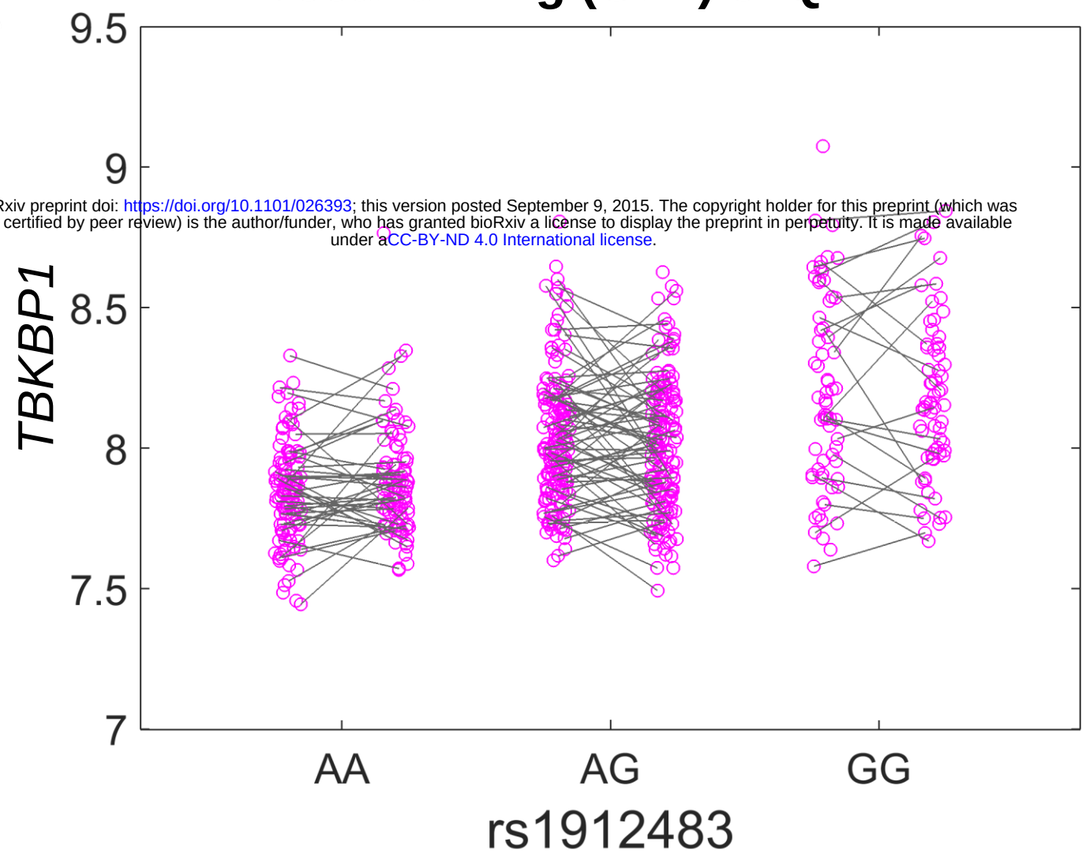
Table S2. The list of *cis*- and *trans*-evQTLs identified using the SVLM method.

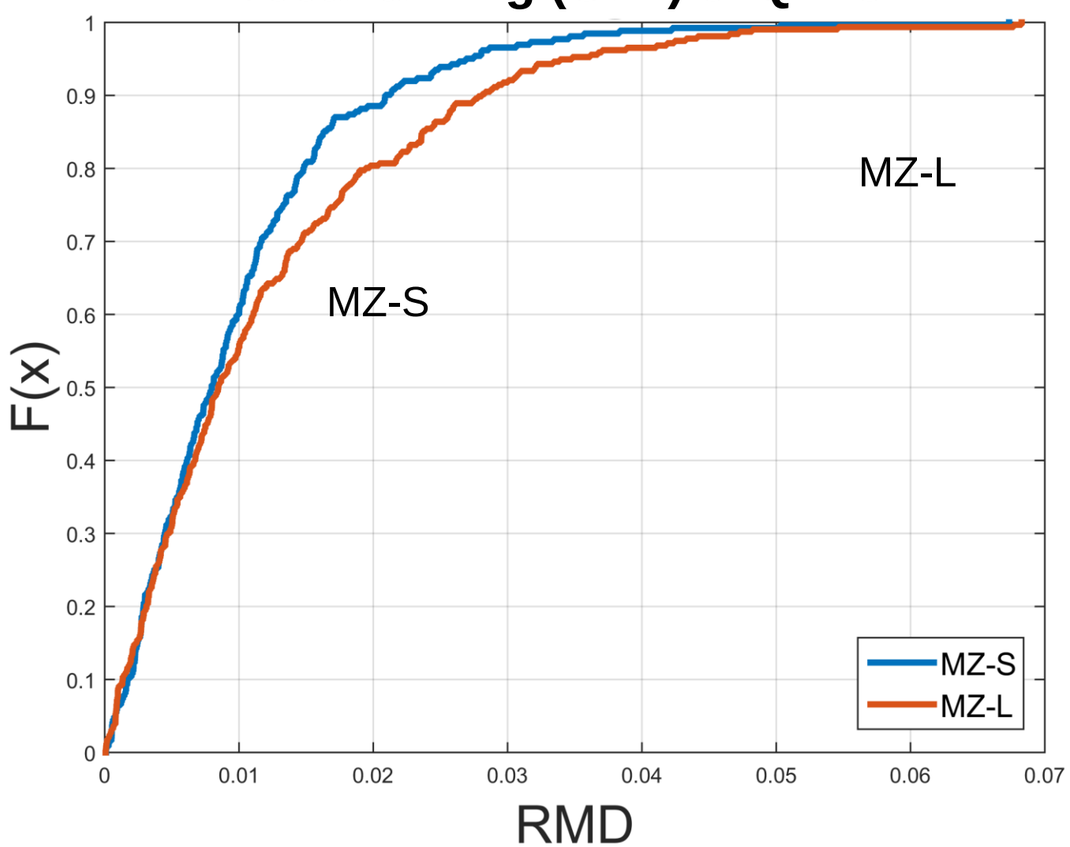
Table S3. The list of partial eQTLs and the corresponding evQTLs.

Table S4. Sequences for primers using in the qRT-PCR assay.

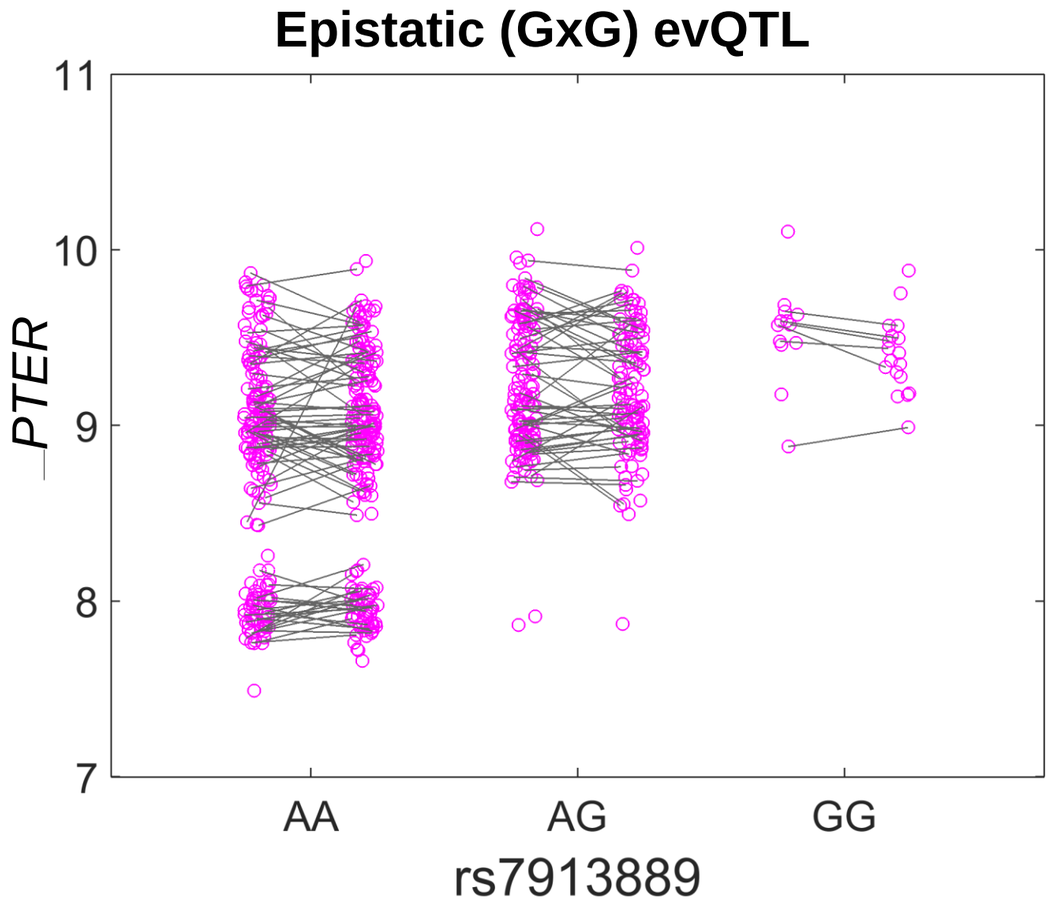

A
Detect cis-evQTLs
(DGLM method)
28,494,473 tests performed
(15,124 genes vs. common SNPs
[MAF>15%])
17,949 significant SNP-gene pairs
(Bonferroni corrected P<0.05)
1,304 genes
B
Detect cis- and trans-evQTLs
(SVLM method)
43,637,670,424 tests performed
(15,124 genes vs. 2,885,326
common SNPs [MAF>15%])
1,513 significant SNP-gene pairs
(Benjamini-Hochberg FDR<0.1)
268 genes
33 cis-evQTL
genes
235 trans-
evQTL genes
C
Gene
SNP

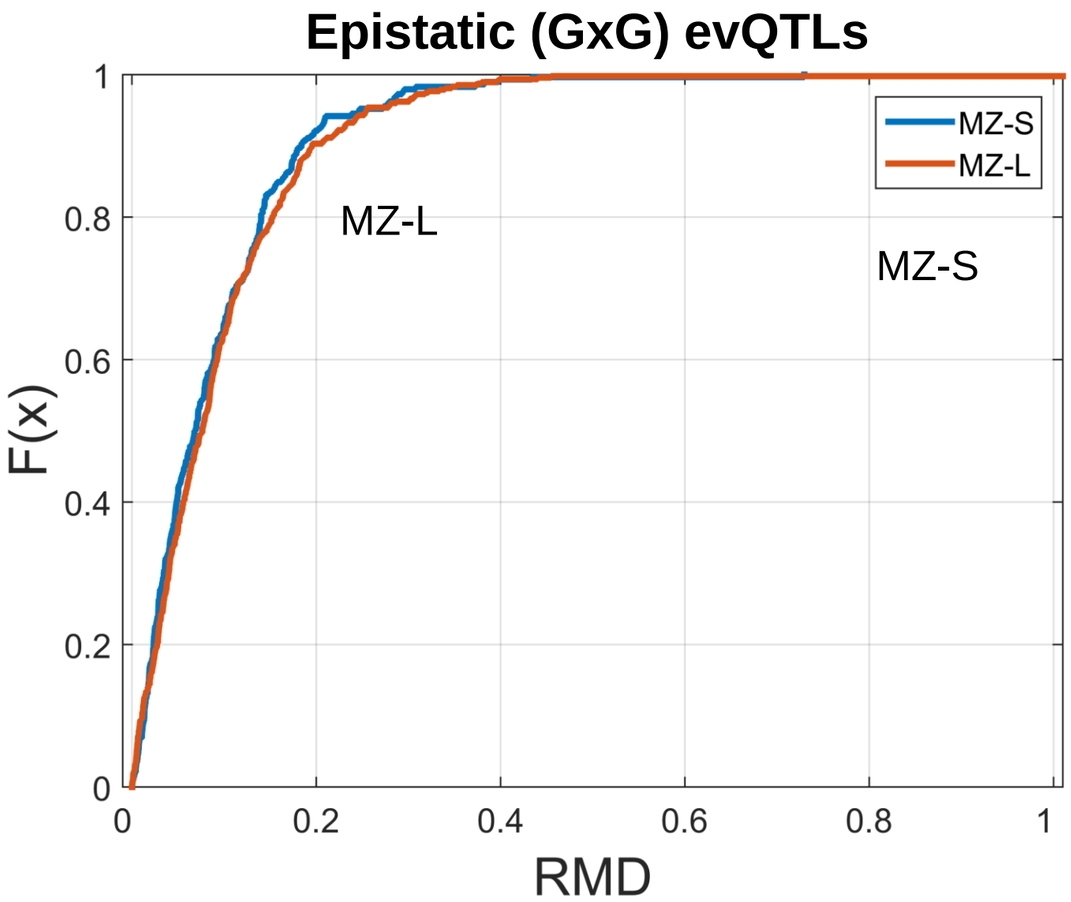
A

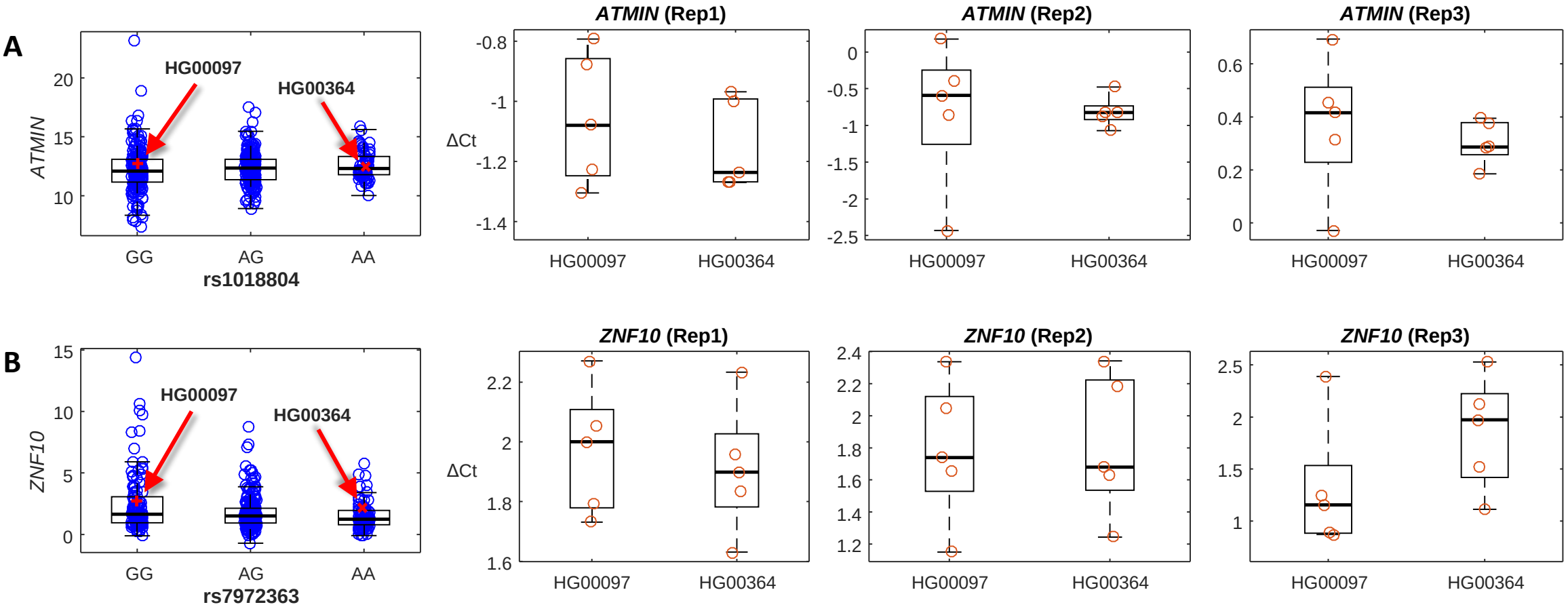




B






A
ATMIN
HG00097
HG00364
GG
AG
AA
rs1018804
ATMIN (Rep1)
ATMIN (Rep2)
ATMIN (Rep3)
ΔCt
HG00097
HG00364
B
ZNF10
rs7972363
ZNF10 (Rep1)
ZNF10 (Rep2)
ZNF10 (Rep3)