bioRxiv preprint doi: https://doi.org/10.1101/026393; this version posted December 14, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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: Genetic mechanisms underlying evQTLs

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

bioRxiv preprint doi: https://doi.org/10.1101/026393; this version posted December 14, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Abstract

Increasing evidence shows that phenotypic variance is genetically controlled; the precise mechanisms of genetic control over the variance remain to be determined. Here, using variance association mapping analysis of gene expression, we show that common genetic variation contributes to gene expression variability via distinct modes of action—e.g., epistasis and decanalization. We focused on the full set of genetic loci associated with gene express variance, i.e., genome-wide expression variability QTLs (or evQTLs), in humans. We found that a quarter of evQTLs, explained by the epistasis model, could be attributed to the presence of “third-party” partial eQTLs that influence gene expression in a fraction, rather than the entire set, of samples. The other three-quarters of evQTLs, explained by the decanalization model, were attributable to the disruptive effect conferred by their own SNPs on transcriptional robustness—that is, these SNPs are responsible for modulating the stability of transcriptional machinery. To validate the model, we measured the discordant expression between monozygotic twins, as well as the level of transcriptional noise in individual cell lines. We showed that decanalizing evQTL SNPs indeed affect the level of transcriptional noise in individuals and contribute to gene expression variability at the population level. Together, our results suggest that common genetic variation works either interactively or independently to influence gene expression variability. These findings may have implications for methodology development toward a new variability-centered research paradigm for mapping quantitative traits.

bioRxiv preprint doi: https://doi.org/10.1101/026393; this version posted December 14, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Introduction

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 also genetically controlled, and the variance itself is a quantitative trait (2-12). 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, 13, 14). In evolutionary biology, for example, variability offers an adaptive solution to environmental changes (15-17). Genetic factors resulting in more variable phenotypes become favored when they enable a population to respond more effectively to environmental changes (18-21). In medicine, disease 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, 22). The lack of empirical studies in this regard has hindered the discovery of variance associated mutations that may contribute to modulating disease susceptibility and the phenotypic variability of other human health-related traits.

Previous studies have shown the existence of substantial gene expression variability (i.e., the difference in gene expression variance between groups) in various systems (23-25). Nevertheless, our understanding of how genetic diversity control or influence gene expression variability remains limited. Promising new developments along this line have come from recent 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 an alternative, more straightforward, explanation—that is, evQTL SNPs disrupt or destabilize the genetic architecture that buffers stochastic variation in gene expression. To emphasize the disruptive effect on the genetic buffering system, we borrow the concept of canalization (i.e., the ability of a population to produce the same phenotype, regardless of variability in its environment or genotype) (26) and regard the disruptive effect as

bioRxiv preprint doi: https://doi.org/10.1101/026393; this version posted December 14, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

"decanalization" that pushes gene expression trait out of its homeostasis or equilibrium to become less robust. We reveal that the formation of evQTLs can be explained by using either the epistasis model or the decanalization model, laying down the foundation for developing a new analytical framework that focuses on the contributions of genetic variation to phenotypic variability. We anticipate that new methods derived from such a framework will allow us to identify novel causal loci, which would otherwise be missed by traditional mean-focused methods, in complex trait 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 (27). We also obtained the genotype data at 2,885,326 polymorphic sites determined in the same cell lines by the 1,000 Genomes Project (28). 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) (29) 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, 30). 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 (31) 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

bioRxiv preprint doi: https://doi.org/10.1101/026393; this version posted December 14, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 marked 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.

Epistatic interactions contribute to increasing gene expression variability

Epistasis, i.e., the interaction between loci, may increase the phenotypic variability of a population (10, 32). 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 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

bioRxiv preprint doi: https://doi.org/10.1101/026393; this version posted December 14, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

We put forward the decanalization model to explain the formation of evQTLs. Unlike the epistasis model that concerns the epistatic interactions or associations between variants (9, 12), the decanalization model concerns a single SNP that perturbs stable genetic systems. We hypothesized that some evQTL SNPs are associated with gene expression variability because one of their two alleles confers the destabilizing function, causing more variable gene expression. These alleles cause the increased gene expression variability without interacting with any other SNPs. Thus, decanalizing evQTLs have such a formation mechanism involving a single-locus, which is different from that of epistatic evQTLs.

To show the decanalizing effect of SNPs, 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 (33). 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 epistatic and 43 decanalizing 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 epistasis model but is likely to be by the decanalization model. Next, we extracted expression data of the 139 pairs of monozygotic (MZ) twins. We classified MZ twin pairs whose genotypes at evQTL SNP sites are homozygous into either MZ-L or MZ-S group based on whether their evQTL SNPs were associated with large or small variance of gene expresion. For MZ twin pairs in the same group (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).

We use two evQTL examples, one decanalizing evQTL and one epistatic evQTL, to illustrate the difference between two types in term of the discordant gene expression between L and S groups. The decanalizing evQTL is between *TBKBP1* and rs1912483 (**Fig. 2A**, right), and the

bioRxiv preprint doi: https://doi.org/10.1101/026393; this version posted December 14, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

epistatic evQTL is between *PTER* and rs7913889 (**Fig. 2B**, right). 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 gray line. The slope of the lines is an indicator of discordant gene expression between twin pairs. In the decanalizing 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**, right). In contrast, in the epistatic evQTL example, the difference in slope skewness between MZ-L and MZ-S groups is less pronounced (**Fig. 2B**, right). 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 decanalizing evQTLs, the distributions of RMD values between L and S groups were significantly different ($P = 1.3\times10^{-5}$, **Fig. 2A**, left), with larger RMD values for L group. In contrast, for epistatic evQTLs, this difference in RMD distribution was not detected between L and S groups ($P = 0.052$, **Fig. 2B**, left).

Decanalizing evQTL SNPs are associated with gene expression noise

Our decanalization model works by the action of a single genetic variant that confers the decanalizing effect on gene expression. One of the underlying sources of the gene expression variability is stochastic noise in gene expression (34). We hypothesized that different alleles of a decanalizing evQTL SNP might be associated with different levels of expression noise for 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 valid, 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 decanalizing 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 (35). 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 ten different sampling time points. The same assay was repeated

bioRxiv preprint doi: https://doi.org/10.1101/026393; this version posted December 14, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 all three replicates, the difference was statistically significant (Brown-Forsythe test, P = 0.034, 0.019, and 0.0096, respectively).

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 decanalizing 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**).

We hypothesized that the correlation between gene expression variability and noise exists exclusively in decanalizing evQTLs. We did not expect such a relationship could be recapitulated in epistatic 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 an epistatic 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 an epistatic evQTL, the interacting SNP rs1567910, which interacts with rs7972363 and helps the creation of the evQTL, has been identified by using the two-step partial eQTL detection method. 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 (**Fig. 3B,** Brown-Forsythe test, P = 0.96, 0.83, and 0.73, for the three replicates, respectively). Together, our results suggest that the level of gene expression noise—the random fluctuation of gene expression—is associated with decanalizing evQTL, but not epistatic, SNPs.

bioRxiv preprint doi: https://doi.org/10.1101/026393; this version posted December 14, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Differences in cell cycle status and alternative splicing do not account for the decanalizing function conferred by decanalizing 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 (36) 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 (37). Emerging experimental and statistical techniques have allowed the variability regarding various phenotypes to be rigorously analyzed (13). 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, 38) or link to each other through incomplete linkage disequilibrium (12, 39). In line with this model, a number of methods for identifying epistasis have been proposed, based on detecting the increased variability (10, 32, 40). 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 epistatic and decanalizing 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

bioRxiv preprint doi: https://doi.org/10.1101/026393; this version posted December 14, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

different genotypes are associated with varying degrees of transcriptional stability. We also detected the 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 individual level to the population 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 (41). 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 the decanalizing function conferred by evQTL SNPs. 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 SNPs can influence gene expression variability. Here we use the decanalizing 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 (42). 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 (43) to interact with Hsp70 and DNAJ proteins (44). The latter two interact with the dynein light chain DYNLL1 (45). Finally, *DYNLL1* and *ATMIN* form a loop of regulatory feedback—one of few known examples of negative autoregulation of gene expression where a gene product directly inhibits the main transcriptional activator while bound at its own promoter (46). 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 SNPs influencing gene expression variability.

We anticipate that our findings have implications for studying human diseases, in which the regulatory variation plays critical roles (47). Decanalization effect has been proposed to influence brain development and contribute to the risk of psychological disorders like

bioRxiv preprint doi: https://doi.org/10.1101/026393; this version posted December 14, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

schizophrenia (48, 49) and other complex diseases (14). Increased gene expression variability was found to be associated with the aging in a mouse model (50), and the aggressiveness of lymphocytic leukemia (51). 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 SNPs), 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 (52), 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 (27) 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 had been quantile normalized and further processed by using the method of probabilistic estimation of expression residuals (PEER) (53). From the downloaded 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 (28). Based on the result of principal component analysis, we excluded 19 samples whose global expression profiles deviated most from those of the rest of samples. The final expression matrix contained the data of 15,124 protein-coding genes for 326 EUR samples. We also obtained the genotype and LCL expression data from a cohort of female twin pairs (33) from the TwinsUK adult twin registry (54). The expression data from 139 pairs of MZ twins was extracted and used in this study.

bioRxiv preprint doi: https://doi.org/10.1101/026393; this version posted December 14, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Identification of evQTLs

We adopted the DGLM method 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: $y_i = \mu + x_i\beta + g_i\alpha + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2 \exp(g_i\theta))$, 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 (30) to detect evQTLs. The SVLM method consists of two steps. First, a regression analysis is applied where the gene expression trait is adjusted for a possible SNP effect and the effect of other covariates is also regressed out. Second, another regression analysis is applied to the squared values of residuals obtained from the first stage, using the SNP as the predictor to capture the effect of the SNP on the expression residuals.

Identification of partial eQTL SNPs that interact with evQTL SNPs

We used a two-step procedure to identify partial eQTL 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) variance 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 (55) 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 ten different time points from 12 to 72 h after growth. Total RNA was extracted using Trizol reagent (Invitrogen).

bioRxiv preprint doi: https://doi.org/10.1101/026393; this version posted December 14, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 a recent RNA-seq study for constantly expressed human genes (56). The expression stability of the two genes was further confirmed by using geNorm and NormFinder programs (57, 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 (PI). 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 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 an 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

bioRxiv preprint doi: https://doi.org/10.1101/026393; this version posted December 14, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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. and Walsh, B. (1998) *Genetics and analysis of quantitative traits*. Sinauer, Sunderland, Mass.
2 Yang, J., Loos, R.J., Powell, J.E., Medland, S.E., Speliotes, E.K., Chasman, D.I., Rose, L.M., Thorleifsson, G., Steinthorsdottir, V., Magi, R. *et al.* (2012) FTO genotype is associated with phenotypic variability of body mass index. *Nature*, **490**, 267-272.
3 Jimenez-Gomez, J.M., Corwin, J.A., Joseph, B., Maloof, J.N. and Kliebenstein, D.J. (2011) Genomic analysis of QTLs and genes altering natural variation in stochastic noise. *PLoS genetics*, **7**, e1002295.
4 Ansel, J., Bottin, H., Rodriguez-Beltran, C., Damon, C., Nagarajan, M., Fehrmann, S., Francois, J. and Yvert, G. (2008) Cell-to-cell stochastic variation in gene expression is a complex genetic trait. *PLoS genetics*, **4**, e1000049.
5 Ronnegard, L. and Valdar, W. (2011) Detecting major genetic loci controlling phenotypic variability in experimental crosses. *Genetics*, **188**, 435-447.
6 Metzger, B.P., Yuan, D.C., Gruber, J.D., Duveau, F. and Wittkopp, P.J. (2015) Selection on noise constrains variation in a eukaryotic promoter. *Nature*, **521**, 344-347.
7 Shen, X., Pettersson, M., Ronnegard, L. and Carlborg, O. (2012) Inheritance beyond plain heritability: variance-controlling genes in Arabidopsis thaliana. *PLoS genetics*, **8**, e1002839.
8 Ayroles, J.F., Buchanan, S.M., O'Leary, C., Skutt-Kakaria, K., Grenier, J.K., Clark, A.G., Hartl, D.L. and de Bivort, B.L. (2015) Behavioral idiosyncrasy reveals genetic control of phenotypic variability. *Proc Natl Acad Sci U S A*, **112**, 6706-6711.
9 Brown, A.A., Buil, A., Vinuela, A., Lappalainen, T., Zheng, H.F., Richards, J.B., Small, K.S., Spector, T.D., Dermitzakis, E.T. and Durbin, R. (2014) Genetic interactions affecting human gene expression identified by variance association mapping. *Elife*, **3**, e01381.
10 Pare, G., Cook, N.R., Ridker, P.M. and Chasman, D.I. (2010) On the use of variance per genotype as a tool to identify quantitative trait interaction effects: a report from the Women's Genome Health Study. *PLoS genetics*, **6**, e1000981.
11 Hulse, A.M. and Cai, J.J. (2013) Genetic variants contribute to gene expression variability in humans. *Genetics*, **193**, 95-108.
12 Wang, G., Yang, E., Brinkmeyer-Langford, C.L. and Cai, J.J. (2014) Additive, epistatic, and environmental effects through the lens of expression variability QTL in a twin cohort. *Genetics*, **196**, 413-425.
13 Geiler-Samerotte, K.A., Bauer, C.R., Li, S., Ziv, N., Gresham, D. and Siegal, M.L. (2013) The details in the distributions: why and how to study phenotypic variability. *Curr Opin Biotechnol*, **24**, 752-759.
14 Gibson, G. (2009) Decanalization and the origin of complex disease. *Nature reviews. Genetics*, **10**, 134-140.

bioRxiv preprint doi: https://doi.org/10.1101/026393; this version posted December 14, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

15 Queitsch, C., Sangster, T.A. and Lindquist, S. (2002) Hsp90 as a capacitor of phenotypic variation. *Nature*, **417**, 618-624.
16 Gibson, G. and Wagner, G. (2000) Canalization in evolutionary genetics: a stabilizing theory? *Bioessays*, **22**, 372-380.
17 Wolf, L., Silander, O.K. and van Nimwegen, E. (2015) Expression noise facilitates the evolution of gene regulation. *Elife*, **4**.
18 Zhang, Z., Qian, W. and Zhang, J. (2009) Positive selection for elevated gene expression noise in yeast. *Mol Syst Biol*, **5**, 299.
19 Acar, M., Mettetal, J.T. and van Oudenaarden, A. (2008) Stochastic switching as a survival strategy in fluctuating environments. *Nat Genet*, **40**, 471-475.
20 Kaern, M., Elston, T.C., Blake, W.J. and Collins, J.J. (2005) Stochasticity in gene expression: from theories to phenotypes. *Nature reviews. Genetics*, **6**, 451-464.
21 Hill, W.G. and Zhang, X.S. (2004) Effects on phenotypic variability of directional selection arising through genetic differences in residual variability. *Genet Res*, **83**, 121-132.
22 Hill, W.G. and Mulder, H.A. (2010) Genetic analysis of environmental variation. *Genet Res (Camb)*, **92**, 381-395.
23 Mar, J.C., Matigian, N.A., Mackay-Sim, A., Mellick, G.D., Sue, C.M., Silburn, P.A., McGrath, J.J., Quackenbush, J. and Wells, C.A. (2011) Variance of gene expression identifies altered network constraints in neurological disease. *PLoS genetics*, **7**, e1002207.
24 Ho, J.W., Stefani, M., dos Remedios, C.G. and Charleston, M.A. (2008) Differential variability analysis of gene expression and its application to human diseases. *Bioinformatics*, **24**, i390-398.
25 Campbell, M.G., Kohane, I.S. and Kong, S.W. (2013) Pathway-based outlier method reveals heterogeneous genomic structure of autism in blood transcriptome. *BMC Med Genomics*, **6**, 34.
26 Waddington, C.H. (1942) Canalization of development and the inheritance of acquired characters. *Nature*, **150**, 563-565.
27 Lappalainen, T., Sammeth, M., Friedlander, M.R., t Hoen, P.A., Monlong, J., Rivas, M.A., Gonzalez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G. *et al.* (2013) Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*, **501**, 506-511.
28 Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T. and McVean, G.A. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56-65.
29 Verbyla, A.P. and Smyth, G.K. (1998) Double generalized linear models: approximate residual maximum likelihood and diagnostics. *Research Report*, 1-15.
30 Struchalin, M.V., Amin, N., Eilers, P.H., van Duijn, C.M. and Aulchenko, Y.S. (2012) An R package "VariABEL" for genome-wide searching of potentially interacting loci by testing genotypic variance heterogeneity. *BMC Genet*, **13**, 4.
31 Benjamini, Y. and Hochberg, Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289-300.
32 Struchalin, M.V., Dehghan, A., Witteman, J.C., van Duijn, C. and Aulchenko, Y.S. (2010) Variance heterogeneity analysis for detection of potentially interacting genetic loci: method and its limitations. *BMC Genet*, **11**, 92.
33 Grundberg, E., Small, K.S., Hedman, A.K., Nica, A.C., Buil, A., Keildson, S., Bell, J.T., Yang, T.P., Meduri, E., Barrett, A. *et al.* (2012) Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet*, **44**, 1084-1089.
34 Elowitz, M.B., Levine, A.J., Siggia, E.D. and Swain, P.S. (2002) Stochastic gene expression in a single cell. *Science*, **297**, 1183-1186.
35 Kanu, N. and Behrens, A. (2008) ATMINistrating ATM signalling: regulation of ATM by ATMIN. *Cell Cycle*, **7**, 3483-3486.

bioRxiv preprint doi: https://doi.org/10.1101/026393; this version posted December 14, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

36 Thorvaldsdottir, H., Robinson, J.T. and Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*, **14**, 178-192.
37 Hallgrímsson, B. and Hall, B.K. (2005) *Variation*. Elsevier Academic Press, Amsterdam ; Boston.
38 Hemani, G., Shakhbazov, K., Westra, H.J., Esko, T., Henders, A.K., McRae, A.F., Yang, J., Gibson, G., Martin, N.G., Metspalu, A. *et al.* (2014) Detection and replication of epistasis influencing transcription in humans. *Nature*, **508**, 249-253.
39 Wood, A.R., Tuke, M.A., Nalls, M.A., Hernandez, D.G., Bandinelli, S., Singleton, A.B., Melzer, D., Ferrucci, L., Frayling, T.M. and Weedon, M.N. (2014) Another explanation for apparent epistasis. *Nature*, **514**, E3-5.
40 Daye, Z.J., Chen, J. and Li, H. (2012) High-Dimensional Heteroscedastic Regression with an Application to eQTL Data Analysis. *Biometrics*, **68**, 316-326.
41 Dey, S.S., Foley, J.E., Limsirichai, P., Schaffer, D.V. and Arkin, A.P. (2015) Orthogonal control of expression mean and variance by epigenetic features at different genomic loci. *Mol Syst Biol*, **11**, 806.
42 Bar-Peled, L., Chantranupong, L., Cherniack, A.D., Chen, W.W., Ottina, K.A., Grabiner, B.C., Spear, E.D., Carter, S.L., Meyerson, M. and Sabatini, D.M. (2013) A Tumor suppressor complex with GAP activity for the Rag GTPases that signal amino acid sufficiency to mTORC1. *Science*, **340**, 1100-1106.
43 Wiles, A.M., Doderer, M., Ruan, J., Gu, T.T., Ravi, D., Blackman, B. and Bishop, A.J. (2010) Building and analyzing protein interactome networks by cross-species comparisons. *BMC Syst Biol*, **4**, 36.
44 Kampinga, H.H. (2015) Molecular biology: It takes two to untangle. *Nature*, **524**, 169-170.
45 Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S. *et al.* (2005) A human protein-protein interaction network: a resource for annotating the proteome. *Cell*, **122**, 957-968.
46 Jurado, S., Conlan, L.A., Baker, E.K., Ng, J.L., Tenis, N., Hoch, N.C., Gleeson, K., Smeets, M., Izon, D. and Heierhorst, J. (2012) 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*, **287**, 3156-3164.
47 Albert, F.W. and Kruglyak, L. (2015) The role of regulatory variation in complex traits and disease. *Nature reviews. Genetics*, **16**, 197-212.
48 McGrath, J.J., Hannan, A.J. and Gibson, G. (2011) Decanalization, brain development and risk of schizophrenia. *Transl Psychiatry*, **1**, e14.
49 Burrows, E.L. and Hannan, A.J. (2013) Decanalization mediating gene-environment interactions in schizophrenia and other psychiatric disorders with neurodevelopmental etiology. *Front Behav Neurosci*, **7**, 157.
50 Bahar, R., Hartmann, C.H., Rodriguez, K.A., Denny, A.D., Busuttil, R.A., Dolle, M.E., Calder, R.B., Chisholm, G.B., Pollock, B.H., Klein, C.A. *et al.* (2006) Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature*, **441**, 1011-1014.
51 Ecker, S., Pancaldi, V., Rico, D. and Valencia, A. (2015) Higher gene expression variability in the more aggressive subtype of chronic lymphocytic leukemia. *Genome Med*, **7**, 8.
52 Zeng, Y., Wang, G., Yang, E., Ji, G., Brinkmeyer-Langford, C.L. and Cai, J.J. (2015) Aberrant gene expression in humans. *PLoS genetics*, **11**, e1004942.
53 Stegle, O., Parts, L., Durbin, R. and Winn, J. (2010) A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLoS Comput Biol*, **6**, e1000770.
54 Moayyeri, A., Hammond, C.J., Hart, D.J. and Spector, T.D. (2013) The UK Adult Twin Registry (TwinsUK Resource). *Twin Res Hum Genet*, **16**, 144-149.

bioRxiv preprint doi: https://doi.org/10.1101/026393; this version posted December 14, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

55 Stranger, B.E., Forrest, M.S., Clark, A.G., Minichiello, M.J., Deutsch, S., Lyle, R., Hunt, S., Kahl, B., Antonarakis, S.E., Tavare, S. *et al.* (2005) Genome-wide associations of gene expression variation in humans. *PLoS genetics*, **1**, e78.
56 Eisenberg, E. and Levanon, E.Y. (2013) Human housekeeping genes, revisited. *Trends Genet*, **29**, 569-574.
57 Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. and Speleman, F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*, **3**, RESEARCH0034.
58 Andersen, C.L., Jensen, J.L. and Orntoft, T.F. (2004) Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res*, **64**, 5245-5250.

Figure Legends

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.

Fig. 2. Dissection of decanalizing and epistatic effects of evQTLs using twins data. (**A**) The left panel shows the cumulative distribution function (CDF) curves of normalized discordant gene expression (measured using RMD) between MZ pairs in S and L groups. The right panel shows an example of decanalizing 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. (**B**) Same as (**A**) but showing an example of epistatic evQTL, *PTER*-rs7913889.

Fig. 3. The correlation between gene expression variability and transcriptional noise presents in the decanalizing evQTL, *ATMIN*-rs1018804, but not in the epistatic 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 ten different time points 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. (**B**) Same as (**A**) but showing the results of epistatic evQTL *ZNF10*-rs7972363.

bioRxiv preprint doi: https://doi.org/10.1101/026393; this version posted December 14, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 SNP 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 decanalizing evQTLs. (**A**) Decanalizing 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**) Principal component analysis 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

bioRxiv preprint doi: https://doi.org/10.1101/026393; this version posted December 14, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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.

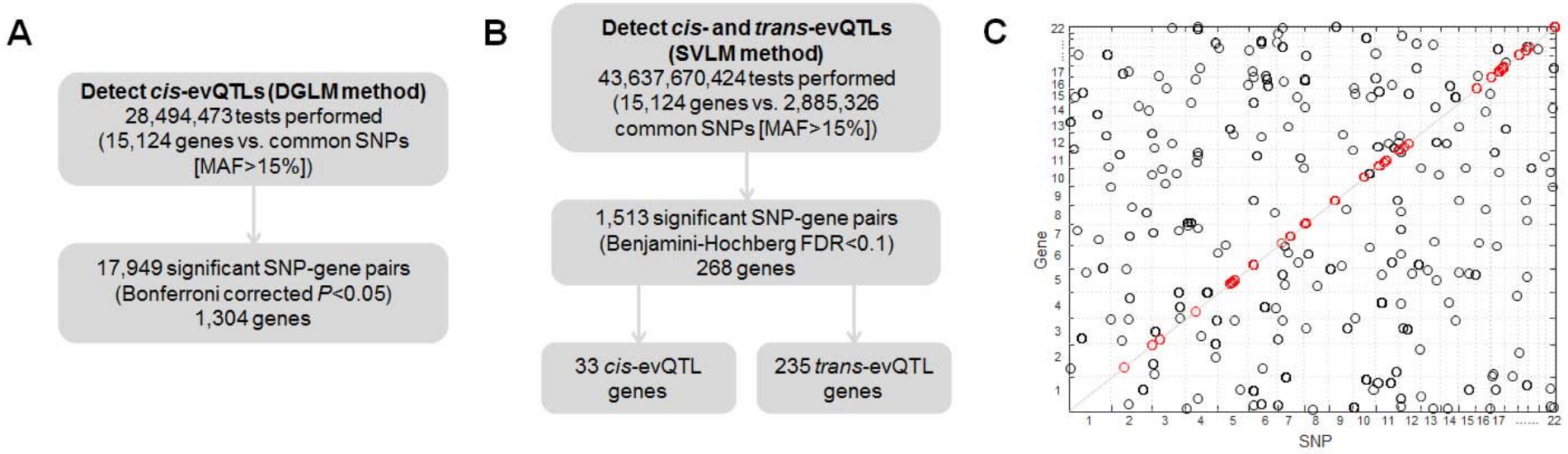
Table S1. The list of *cis*-evQTLs identified using the DGLM method.

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.

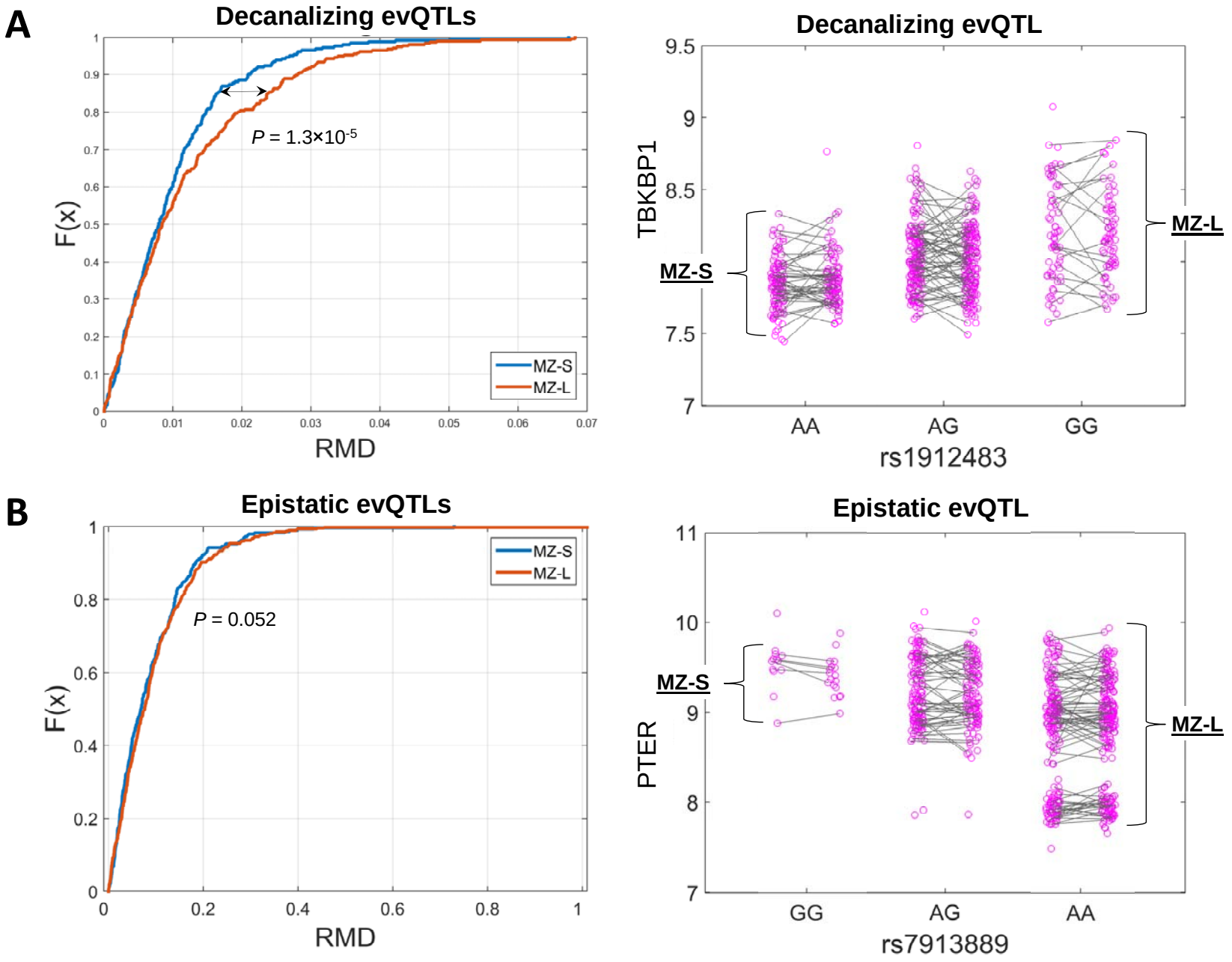
Table S4. Sequences for primers using in the qRT-PCR assay.

Fig. 1.



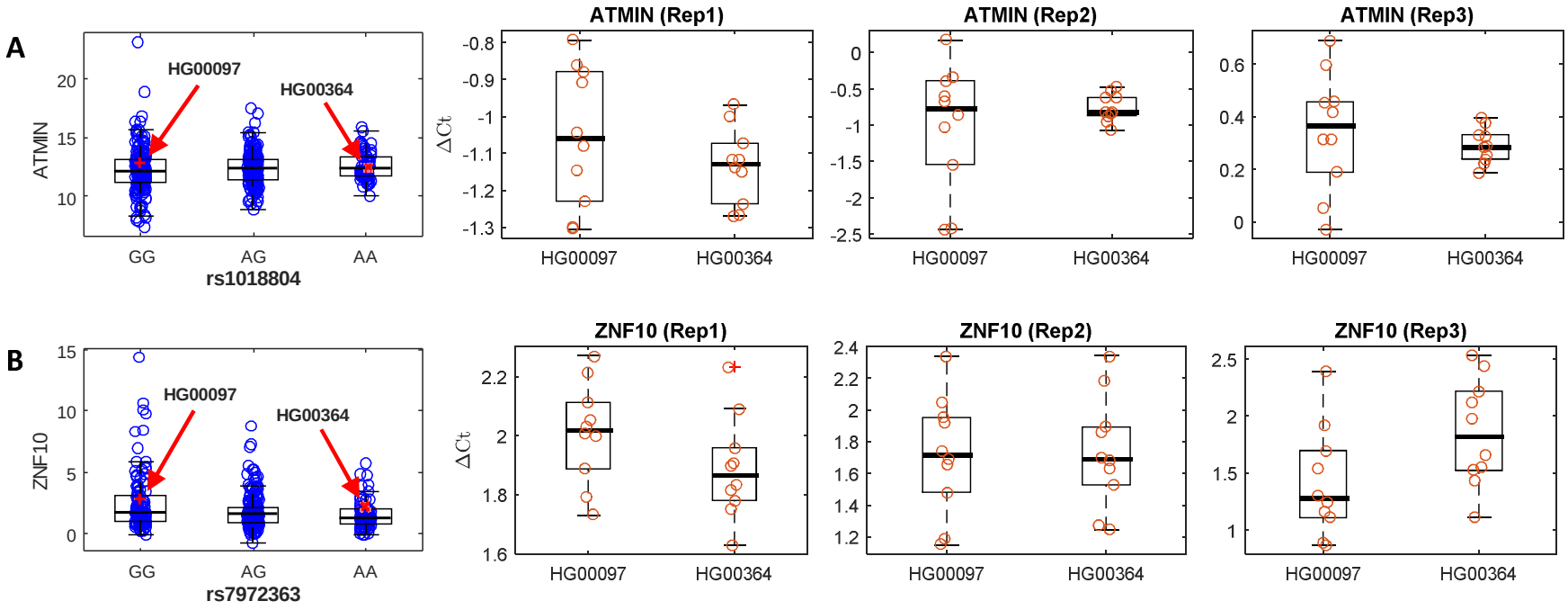
bioRxiv preprint doi: https://doi.org/10.1101/026393; this version posted December 14, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Fig. 2.

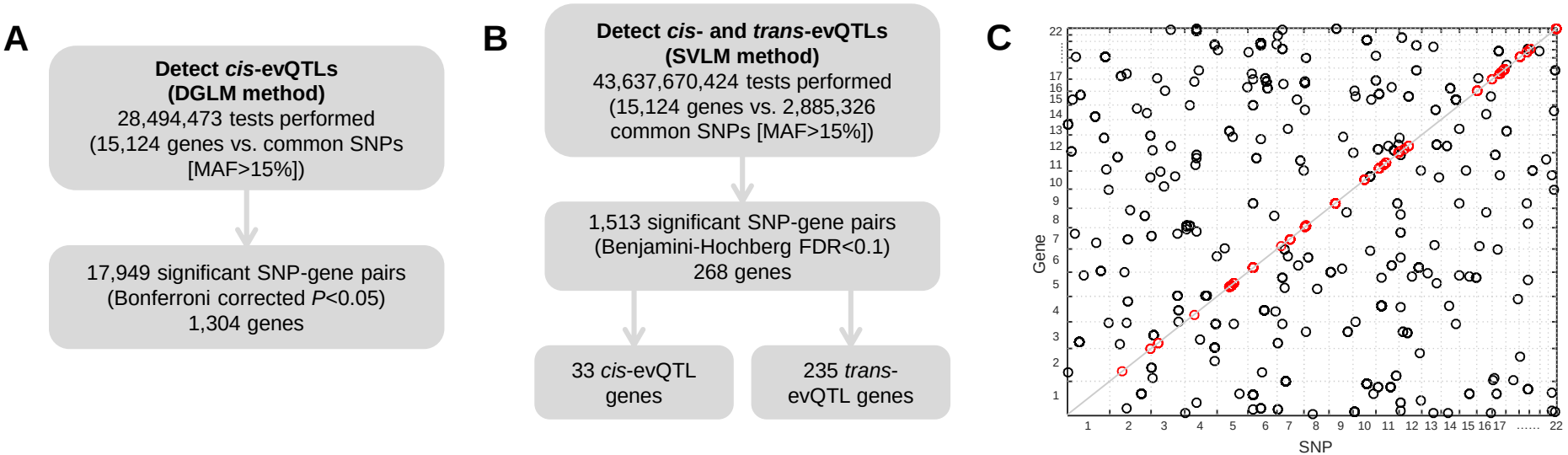


bioRxiv preprint doi: https://doi.org/10.1101/026393; this version posted December 14, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Fig. 3.

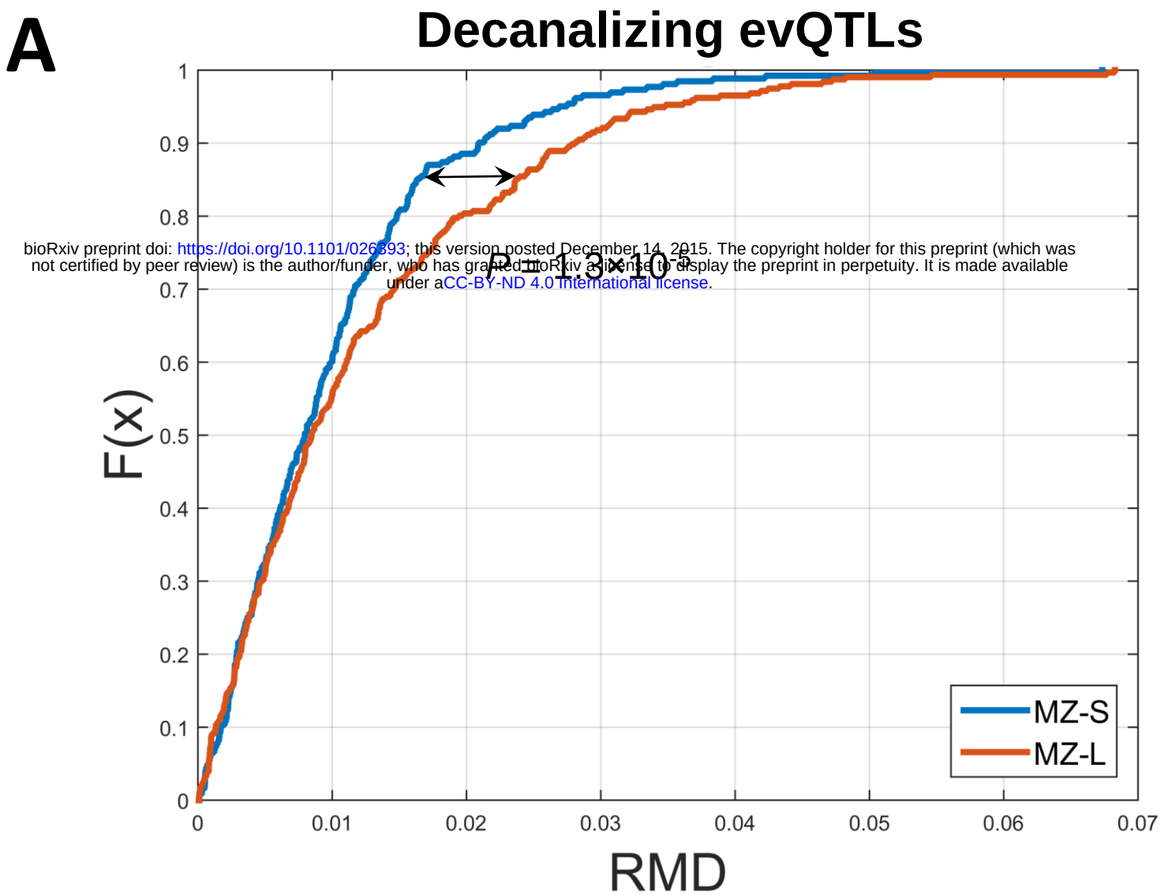


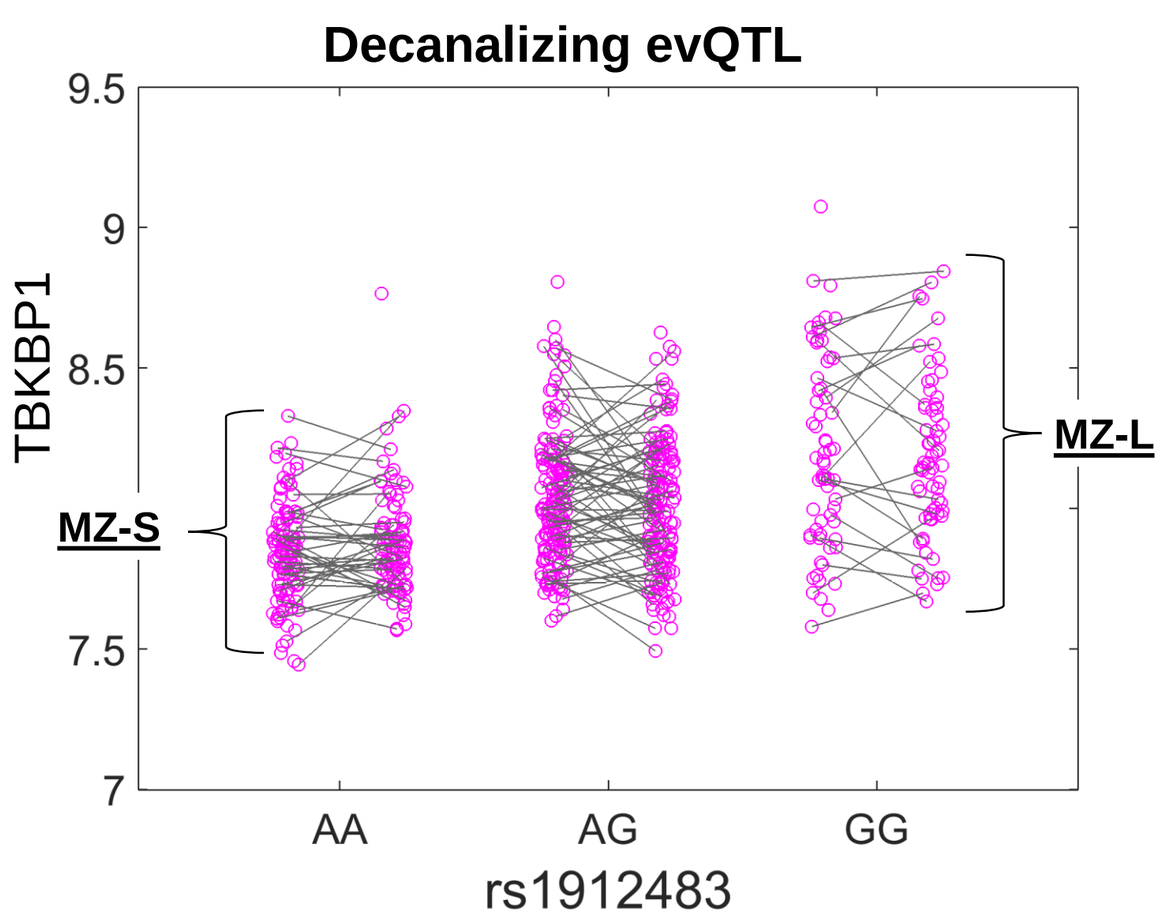
bioRxiv preprint doi: https://doi.org/10.1101/026393; this version posted December 14, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.


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

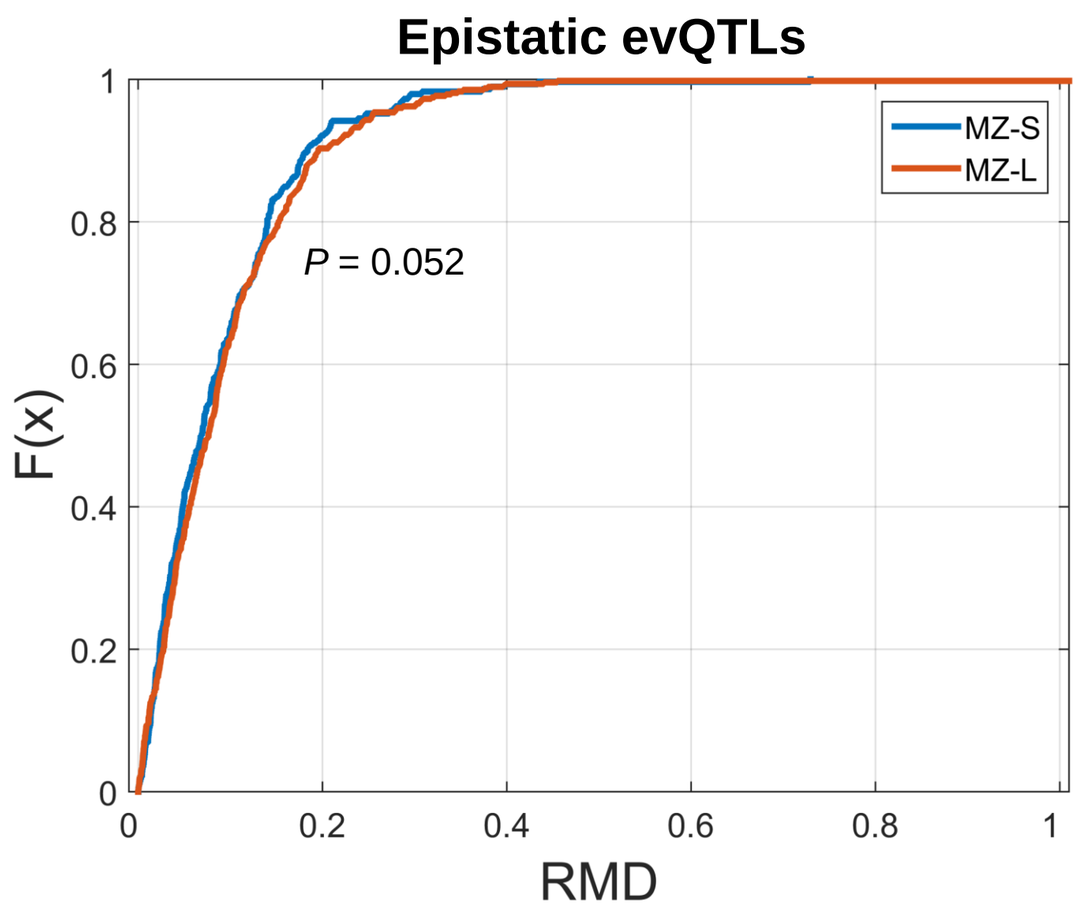
bioRxiv preprint doi: https://doi.org/10.1101/026393; this version posted December 14, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

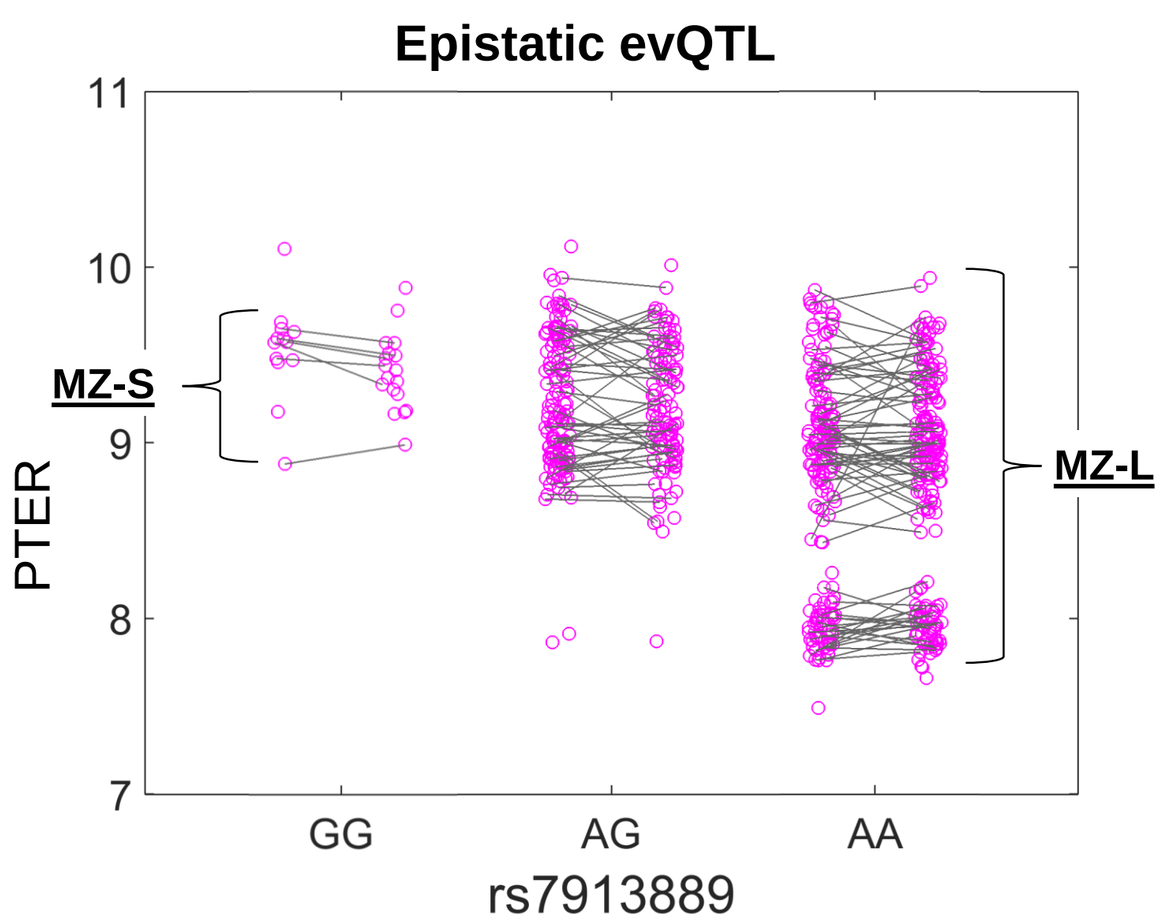
A

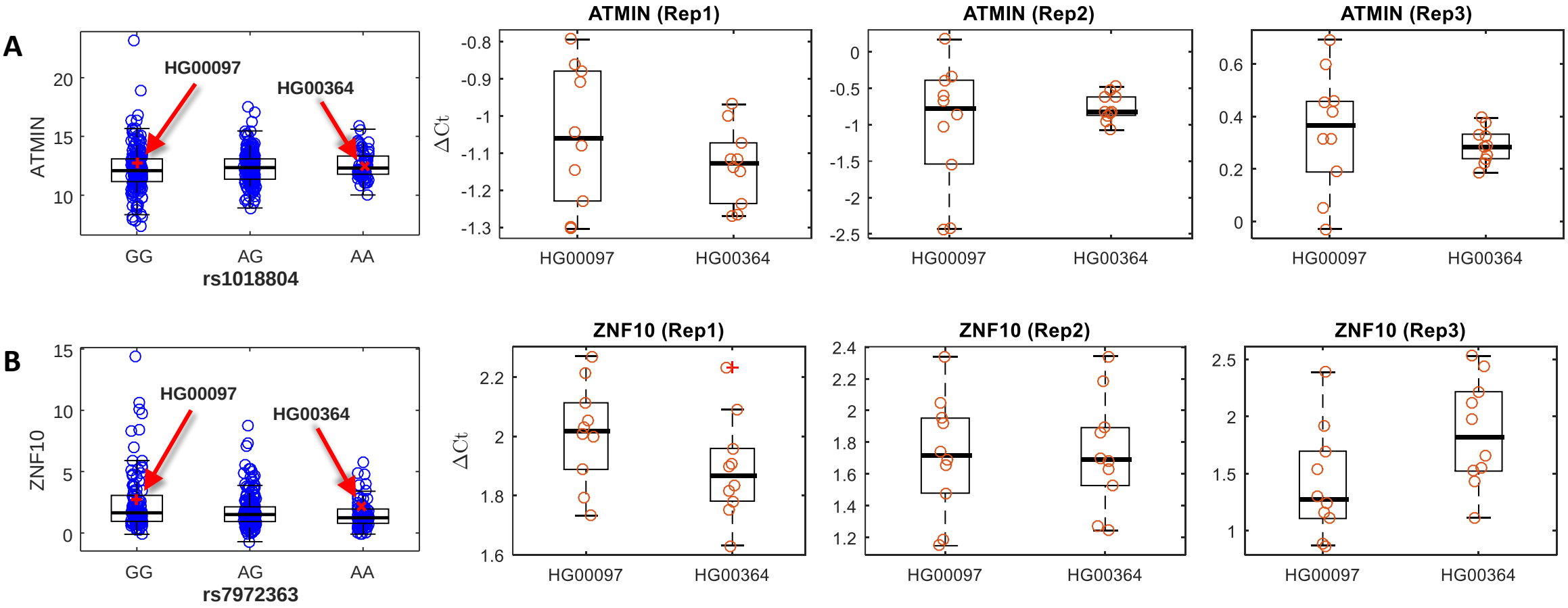




B






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