PLOS GENETICS

METHODS

ADELLE: A global testing method for trans-eQTL mapping

Takintayo Akinbiyi[1†], Mary Sara McPeek[1,2]*, Mark Abney[2]*

1 Department of Statistics, The University of Chicago, Chicago, Illinois, United States of America,
2 Department of Human Genetics, The University of Chicago, Chicago, Illinois, United States of America

† Deceased.
* mcpeek@g.uchicago.edu (MSMP); abney@uchicago.edu (MA)

Abstract

Understanding the genetic regulatory mechanisms of gene expression is an ongoing challenge. Genetic variants that are associated with expression levels are readily identified when they are proximal to the gene (i.e., cis-eQTLs), but SNPs distant from the gene whose expression levels they are associated with (i.e., trans-eQTLs) have been much more difficult to discover, even though they account for a majority of the heritability in gene expression levels. A major impediment to the identification of more trans-eQTLs is the lack of statistical methods that are powerful enough to overcome the obstacles of small effect sizes and large multiple testing burden of trans-eQTL mapping. Here, we propose ADELLE, a powerful statistical testing framework that requires only summary statistics and is designed to be most sensitive to SNPs that are associated with multiple gene expression levels, a characteristic of many trans-eQTLs. In simulations, we show that for detecting SNPs that are associated with 0.1%–2% of 10,000 traits, among the 8 methods we consider ADELLE is clearly the most powerful overall, with either the highest power or power not significantly different from the highest for all settings in that range. We apply ADELLE to a mouse advanced intercross line data set and show its ability to find trans-eQTLs that were not significant under a standard analysis. We also apply ADELLE to trans-eQTL mapping in the eQTLGen data, and for 1,451 previously identified trans-eQTLs, we discover trans association with additional expression traits beyond those previously identified. This demonstrates that ADELLE is a powerful tool at uncovering trans regulators of genetic expression.

Author summary

Identification of trans-eQTLs, i.e., genetic variants that regulate expression of genes that are not proximal, has proved challenging, even though previous studies suggest that they may account for a large proportion of complex trait variance. Compared to cis-eQTLs, i.e., variants that regulate expression of proximal genes, trans-eQTLs are much harder to detect because their effect sizes tend to be smaller, and the space of possible genes whose expression they might be associated with is much bigger, leading to a higher burden of multiple comparisons. We developed ADELLE, a powerful statistical method that requires

only summary statistics and is designed to be most sensitive to SNPs that are associated with multiple gene expression levels, a characteristic of many trans-eQTLs. In simulations, we show that for detecting SNPs that are associated with 0.1%–2% of the expression traits, ADELLE is clearly the most powerful overall among the 8 methods we compare. We apply ADELLE to eQTLGen data and also to a mouse advanced intercross line data set and show its ability to detect trans-eQTL signal that was not significant under a standard analysis. This demonstrates that ADELLE is a powerful tool at uncovering trans regulators of genetic expression.

Introduction

eQTL mapping, in which association is tested between gene expression levels and genetic variants, is a useful approach toward understanding mechanisms of genetic regulation. Cis-eQTLs, genetic variants that influence expression of proximal genes, are often readily detected because their effect sizes are commonly large, and the local nature of their effects limits the number of tests and, hence, the multiple testing burden. Because of this, many studies have focused on investigating the role of cis-regulatory effects on gene expression. Recent work, however, has estimated that cis-genetic effects account for a minority of human complex trait variance, perhaps as little as 11%, while trans-genetic effects, i.e. causes that are distant from the gene being regulated, may account for 70% or more of complex trait variance in humans [1, 2]. Unfortunately, even though trans-eQTL effects may dominate the genetic variability of gene expression and of complex traits, the identification of trans-eQTLs has been impeded by two significant hurdles. Compared to cis-eQTLs, trans-eQTLs are much harder to detect because their effect sizes tend to be smaller [2], and the space of possible genes whose expression they might be associated with is much bigger, leading to a higher burden of multiple comparisons.

A basic approach in both model organisms and humans to detect trans-eQTLs is to perform, for each SNP, a test of association against every trans-gene [1, 3–5]. To account for multiple testing, either a Bonferroni correction is applied or a false discovery rate (FDR) procedure is used. Because of the very high number of tests performed, only the strongest of signals achieve statistical significance. This has led to recent efforts to develop methods that will be more effective at detecting trans-eQTLs. Broadly, many of the methods seek to increase the number of discoveries by applying at least one of the following strategies (1) reducing the multiple testing burden by either reducing the number of variants tested [6–10] or reducing the number of genes tested [11–13], or (2) leveraging the expectation that a trans-eQTL will influence the regulation of multiple genes [12–15]. Although incorporating biological or other external information to effectively make the number of tests smaller has the potential to increase power by eliminating either variants or traits where the null hypothesis is true, it also has the potential to miss important signals. On the other hand, even though a trans-eQTL may affect the expression levels of multiple genes the number of these genes will typically be a very small fraction of the total number of genes. Together, these qualities have made the development of effective tools for the discovery of trans-eQTLs very challenging.

We address the problem of developing a powerful statistical method for trans-eQTL detection. In particular, we frame the problem as one where we seek to reject the global null hypothesis that for a candidate trans-eQTL (e.g., a single SNP) none of the expression traits are associated with the SNP. We develop a method that requires only summary statistics of individual tests of association between a SNP and an expression trait. Advantages of only requiring

summary statistics include their ease of being shared and savings in the person and computational effort to generate them.

For the general statistical problem of aggregating a collection of Z scores or p-values into a single test of the global null hypothesis, various methods have been proposed. Examples include Simes's method [16], Cauchy p-value combination [17], higher criticism [18, 19], the Berk-Jones statistic [20], and methods based on equal local levels (ELL) [20–25]. Both the higher criticism and Berk-Jones statistics have generalizations to the case where the tests are dependent, generalized higher criticism [26] and generalized Berk-Jones [27]. These methods were used to test association between a SNP-set and an outcome. Another class of global tests commonly used in genetics corresponds to the sum of χ^2 statistics from different tests [28], which we call Sum-χ^2. Variations and generalizations of this approach underlie methods for rare variant and haplotype association analysis such as SKAT [29] and other variance component tests [30]. The CPMA [14, 31] method has been proposed for combining test statistics for multi-trait mapping. The most commonly-used p-value combination approach is what we call Min-P, which is simply based on the smallest p-value in the collection, with significance assessed by Bonferroni correction or another approach such as Monte Carlo.

In general, there is no uniformly most powerful test of the global null hypothesis. Instead, different tests will be optimal in different alternative model regimes. For instance, the Min-P test, with a multiple testing correction, should do well when there is at least one extremely strong signal among the p-values. On the other hand, sums of χ^2 types of tests are likely to do well when weak signals are spread over a relatively large proportion of the p-values [18, 26]. Here, we propose ADELLE, which is an extension of ELL to the case of dependent tests. Because ADELLE is an ELL-based test, we expect it to show strong performance when the signal is both relatively weak and sparse within a collection of p-values, which is the situation we expect when searching for trans-eQTLs. We assess the performance of ADELLE relative to other methods through simulation studies and application to trans-eQTL detection in (i) mouse data from an advanced intercross line [4] and (ii) eQTLGen data [6].

Description of the methods

We first briefly consider the simplified case in which the expression traits are assumed to be independent and describe how the ELL global testing method could be applied. Then we describe ADELLE, our extension of ELL to the case of dependent traits, which we apply to trans-eQTL mapping.

Global trans-eQTL testing with ELL

In an eQTL mapping study in which $\tilde{D}$ expression traits and M genome-wide SNPs are observed on each of N individuals, suppose each expression trait is tested for association with each genome-wide SNP in the sample leading to a $\tilde{D} \times M$ summary statistic matrix Π of p-values having (d, m)th entry π_{dm} equal to the p-value for testing association between expression trait d and SNP m in the N individuals. In this subsection we make the simplifying assumption that the $\tilde{D}$ traits are independent. We extend to the case of dependent traits in the following subsection.

For a given SNP m, define $\mathcal{D}_m$ to be the subset of expression traits that are considered trans to it, from among the larger set of $\tilde{D}$ traits measured. To detect trans-eQTLs, we propose to perform M global hypothesis tests, one for each SNP, in which the mth global hypothesis test

has null and alternative hypotheses

$$H_0^{(m)}: \text{SNP } m \text{ is not associated with any trait in } \mathcal{D}_m \tag{1}$$

$$H_A^{(m)}: \text{SNP } m \text{ is associated with at least one trait in } \mathcal{D}_m. \tag{2}$$

We now fix a SNP m and describe the ELL method for performing the mth global hypothesis test, where the test statistic is constructed from the p-values in column m of Π. Specifically, we consider a vector of p-values π of length $D_m = |\mathcal{D}_m|$, consisting of the subset of p-values in the mth column of Π that correspond to the traits in $\mathcal{D}_m$. For simplicity of exposition, we drop the subscript m in the remainder of this subsection, so we consider $\mathcal{D}$ to be the set of traits that are trans to the SNP and consider π to be of length D. Under the null hypothesis that the given SNP is not associated with any of its trans traits and the further assumption of independence of traits (and assuming that the method for calculating p-values is well-calibrated), the entries of π would be D independent and identically distributed (i.i.d.) Uniform(0,1) random variables.

ELL is a general global testing method that models the entries of π as i.i.d. from a distribution having cumulative distribution function (cdf) $F_\pi(x)$ for $x \in (0, 1)$. The null hypothesis would be

$$H_0: F_\pi(x) = x \text{ for all } x \in (0, 1), \tag{3}$$

i.e., the p-values are Uniform(0,1), and the one-sided alternative hypothesis would be

$$H_A: F_\pi(x) > x \text{ for at least one } x \in (0, 1), \tag{4}$$

i.e., the p-values tend to be smaller under the alternative than would expected under the null. We use the notation $\pi = (\pi_1, \ldots, \pi_D)$ and for $1 \leq d \leq D$, we define $\pi_{(d)}$ to be the dth order statistic of π, i.e., we sort the entries of π in ascending order and let $\pi_{(d)}$ be the dth component of the sorted vector, so $\pi_{(1)} \leq \pi_{(2)} \leq \ldots \leq \pi_{(D)}$. Under the null hypothesis that the unsorted p-values $\pi_1, \ldots, \pi_D$ are i.i.d. uniform, the entries of $(\pi_{(1)}, \ldots, \pi_{(D)})$ are dependent with a known joint distribution, and marginally each $\pi_{(d)}$ has the Beta($d, D - d + 1$) distribution for $1 \leq d \leq D$.

The ELL test starts by comparing each order statistic to its corresponding beta null distribution and deciding whether it is smaller than expected. Then the ELL test statistic is based on the order statistic that shows the most significant deviation from its corresponding null distribution. On the one hand, if trans-eQTL signals are only of moderate or weak size, then, e.g., $\pi_{(1)}$ and $\pi_{(2)}$ might actually represent null tests, and the true alternatives could be represented by smaller than expected $\pi_{(d)}$ for values of d that are perhaps of small to moderate size. On the other hand, finding that $\pi_{(d)}$ is smaller than expected only for larger values of d, e.g., d close to D, would be difficult to interpret and might not seem compelling evidence for the SNP being a trans-eQTL. Therefore, we propose to base the ELL test statistic on only the smallest fraction q of the p-values, i.e., on order statistics $\pi_{(d)}$ for $1 \leq d \leq qD$, where $q \in (0, 1)$. In the original formulation of ELL, Berk and Jones [20] used $q = 0.5$. In the eQTL mapping context, we take $q = 0.2$ in the simulations and data analysis, i.e., we only the consider the smallest 20% of the p-values for a given SNP. In simulations, we assess the impact of the choice of q on power (see the **Power comparison** subsection of the **Verification and comparison** section). For simplicity of notation, in what follows we assume that qD turns out to be an integer (otherwise it could be replaced by floor(qD)).

To construct the ELL test statistic, we first calculate qD "l-values", one for each $\pi_{(d)}$, $1 \leq d \leq qD$, where the l-value l_d for $\pi_{(d)}$ is the p-value for testing the null hypothesis that $\pi_{(d)}$ is drawn

from a Beta$(d, D - d + 1)$ distribution vs. a one-sided alternative for which we reject the null hypothesis if $\pi_{(d)}$ is sufficiently small. Thus, $l_d =$ pbeta$(\pi_{(d)}, d, D - d + 1)$ where pbeta(x, a, b) is the cdf of the Beta(a, b) distribution evaluated at x. Then the ELL test statistic is

$$T_{ELL} = \min_{1 \le d \le qD} l_d. \tag{5}$$

To assess whether the SNP is a trans-eQTL, we perform a one-sided hypothesis test at level α based on T_{ELL}, where we reject the null hypothesis in Eq 1 if $T_{ELL} < \eta$, where η (the "local level") is a function of α. We refer to this as an equal local level test because the local level η at which we reject H_0 is equal for all l_d. That is, if any of the l-values are less than η we reject H_0. Previous work [24] shows that the ELL test is asymptotically optimal for detecting deviations from a Gaussian distribution for a wide class of rare-weak contamination models.

For the case when the traits are independent, there are existing algorithms [25, 32, 33] to calculate the global level α of the test as a function of the local level η, where we call this function $\alpha(\eta)$. These algorithms are specifically for the case $q = 1$. However, we have adapted the algorithm of Weine et al. 2023 [25] to more general q. To do this, we let $\xi =$ floor(qD), and then obtain $\alpha(\eta)$ as $1 - \sum_{j=0}^{\xi-1} c_j^{(\xi)}$, where $c_j^{(\xi)}$ is a quantity calculated recursively in Algorithm 1 of Appendix B.2 of Weine et al. [25] To invert the function $\alpha(\eta)$ and determine the local level η corresponding to a chosen global level α for the ELL test, we conduct a binary search to find the needed η.

ADELLE: Extension of the ELL method to dependent traits

The ELL approach described in the previous subsection assumes independence of traits, but in practice there is typically correlation among gene transcript levels. Our goal is still to perform, for each SNP, a global test based on the null and alternative hypotheses in Eqs 3 and 4. However, dependence among traits leads to dependence among the elements of the p-value vector π. In that case, it is no longer true that, e.g., $\pi_{(d)}$ is beta distributed under the null as it is in the independence case. Therefore, the methods we describe above for calculation of the ELL test statistic and its null distribution are no longer applicable.

The ADELLE method we propose generalizes the ELL approach to allow for dependent traits. For $1 \le d \le D$, define $F_{(d)}$ to be the cdf of the distribution for $\pi_{(d)}$ under the null hypothesis in the case when the traits are dependent. The basic idea behind ADELLE is that we find an approximation to $F_{(d)}$ and use it to calculate the qD l-values $l_1, \ldots, l_{qD}$ in the case when the traits are dependent. Then we define the ADELLE test statistic T_{ADELLE} to be the minimum of $\{l_1, \ldots, l_{qD}\}$. Finally, we calculate the p-value for the ADELLE test using a Monte Carlo approximation method given in subsection **Monte Carlo p-value calculation**.

First we describe how dependence is incorporated into the model. Rather than directly modeling the dependence on the p-value scale, we instead consider a set of association test statistics $Z_1, \ldots Z_D$, where Z_d tests association between the given SNP and its dth trans trait, $1 \le d \le D$. We assume that under the null hypothesis, each $Z_d \sim N(0, 1)$, where they can be correlated with each other, and we assume that π_d is a two-sided p-value based on Z_d, i.e., $\pi_d = 2\Phi(-|Z_d|)$, where Φ is the standard normal cdf.

Let G denote the genotype vector of the SNP and Y_d the phenotype vector of its dth trans trait. Typical examples of Z_d would be the t-statistic for testing significance of G in a linear model for Y_d or the Wald t-statistic for testing significance of G in a linear mixed model (LMM) for Y_d. In large samples, such a t-statistic will be approximately standard normal under the null hypothesis or, if necessary, could be transformed to be approximately standard normal under the null hypothesis by applying the transformation $\Phi^{-1}(\text{pt}(Z_d))$ where pt is the cdf of the

t-distribution with degrees of freedom $= N - k - 1$ where k is the number of predictors in addition to the intercept in the linear model or LMM. A likelihood ratio χ^2 test statistic for testing significance of G in a LMM for Y_d could also be converted to such a Z_d value by taking a square root of the χ^2_1 test statistic and applying the sign of the estimated coefficient of G in the LMM for Y_d.

We let $Z = (Z_1, \ldots, Z_D)^T$ and, under the global null hypothesis that the SNP is unassociated with any of its trans traits, we model Z as multivariate normal:

$$Z \sim N_D(\mathbf{0}, \mathbf{\Omega}), \tag{6}$$

where N_D denotes the multivariate normal distribution of dimension D, $\mathbf{0}$ is a vector of 0's of length D and $\mathbf{\Omega}$ is the $D \times D$ trait correlation matrix. (See S1 Text for a derivation of this model.) To estimate $\mathbf{\Omega}$, we first form the sample correlation matrix $\hat{C}_{trait}$ for the $\tilde{D}$ traits. However, in the case when $N \ll \tilde{D}$, which is common, the estimate $\hat{C}_{trait}$ would be low rank, so we could regularize it by using the shrinkage estimator [34] $\hat{\Omega} = w\hat{C}_{trait} + (1 - w)I$. (See S1 Text for details on choice of the regularization parameter w.).

To calculate $F_{(d)}(h)$ for $h \in (0, 1)$, where $F_{(d)}$ is the cdf of $\pi_{(d)}$ under the null hypothesis, we first point out the key identity that the two events $E_1 = \{\pi_{(d)} \leq h\}$ and $E_2 = \{\sum_{k=1}^{D} I\{\pi_k \leq h\} \geq d\}$ are the same, where $I\{\cdot\}$ is the indicator function that equals 1 if the event inside the brackets occurs and 0 otherwise, and where E_2 is saying that at least d of the p-values are $\leq h$. By the defined relationship between π_k and Z_k, we have that the events $\{\pi_k \leq h\}$ and $\{|Z_k| \geq -\Phi^{-1}(h/2)\}$ are the same, so $E_2 = \{\sum_{k=1}^{D} I\{|Z_k| \geq -\Phi^{-1}(h/2)\} \geq d\}$. Next, define $S(c) = \sum_{d=1}^{D} I\{|Z_d| \geq c\}$ for $c \geq 0$, where $S(c)$ counts the number of $|Z_d|$ that are greater than or equal to c, and note that $E_2 = \{S(-\Phi^{-1}(h/2)) \geq d\}$. Therefore, the following two events are the same

$$\{\pi_{(d)} \leq h\} = \{S(-\Phi^{-1}(h/2)) \geq d\}. \tag{7}$$

Finally, we have for the l-value

$$l_d(h) \equiv F_{(d)}(h) \equiv P_0(\pi_{(d)} \leq h) = P_0(S(-\Phi^{-1}(h/2)) \geq d), \tag{8}$$

where $P_0(\cdot)$ represents probability under the null hypothesis that the SNP is not associated with any of its D trans traits.

As a consequence, we can obtain needed values of $F_{(d)}$ by considering the distribution of $S(c)$ under the null hypothesis. If $\mathbf{\Omega} = I$, then for $c \geq 0$, $S(c)$ has the null distribution of a Binomial$(D, 2\Phi(-c))$ random variable. When $\mathbf{\Omega} \neq I$, $S(c)$ has the same null mean as a Binomial$(D, 2\Phi(-c))$, but the null variance of $S(c)$ is strictly greater than that for Binomial$(D, 2\Phi(-c))$, i.e., the distribution of $S(c)$ is over-dispersed relative to binomial. The beta-binomial distribution is a standard choice for modeling binomial-like data when there is over-dispersion. Therefore we approximate the distribution of $S(c)$ with a beta-binomial distribution $BB(D, \lambda, \gamma)$ where λ and γ are chosen so that the first and second moments match those of $S(c)$, using techniques of a previous work [26] (see also [27]). The details are given in S1 Text. From the resulting approximation to the distribution of $S(c)$, we obtain an approximation to $F_{(d)}$, which we call $\hat{F}_{(d)}$, based on Eq 7.

To obtain the ADELLE test statistic, we first obtain the qD l-values $l_1, \ldots, l_{qD}$, where l_d is defined to be $\hat{F}_{(d)}$ evaluated at the observed value of $\pi_{(d)}$. Then the ADELLE test statistic is given by $T_{ADELLE} = \min_{1 \leq d \leq qD} l_d$. In the special case when $\mathbf{\Omega} = I$, we get back the same ELL l-values and ELL test statistic used for the independence case in the previous subsection.

Our ADELLE method lends itself to a pre-computation step to reduce computation time when applied to a large number of traits and SNPs. This involves defining a dense grid of points $\mathcal{H} \subset (0, 1)$, and evaluating $\hat{F}_{(1)}(h), \ldots, \hat{F}_{(qD)}(h)$ for all $h \in \mathcal{H}$, which can be efficiently carried out as described in detail in S1 Text.

Connection between ELL and higher criticism

ELL (and its extension to ADELLE to allow for dependence) has a theoretical connection to higher criticism [18, 19] (and its extension to generalized higher criticism [26] to allow for dependence). Specifically, ELL and higher criticism have the same asymptotic behavior. However, higher criticism has been shown [22] to under-perform in finite samples, with ELL generally having higher power (in some cases substantially higher) than higher criticism in finite samples for a sparse normal mixture setting, which is an appropriate setting for trans-eQTL mapping. The higher power for ELL over higher criticism occurs even though both methods are asymptotically optimal for this setting. This is attributed to the extremely slow rate of asymptotic convergence, e.g., not until D is of the order of 10^{69} do the asymptotic results seem to hold for higher criticism. [22]

Monte Carlo p-value calculation

We use a Monte Carlo approach to assess significance of the ADELLE global test statistic. Specifically, we simulate R i.i.d. vectors $\tilde{Z}^{(r)} \sim N_D(\mathbf{0}, \hat{\mathbf{\Omega}})$, $1 \leq r \leq R$, where R is very large (e.g., 2×10^7 in the eQTLGen data analysis), and for each $\tilde{Z}^{(r)}$, we calculate the ADELLE statistic, call it $T^{(r)}$. For any observed ADELLE statistic, T, we calculate its p-value as $(N(T) + 1)/(R + 1)$, where $N(T) = \sum_{r=1}^{R} I\{T^{(r)} \leq T\}$ counts the number of $T^{(r)}$ values that are less than or equal to T. We use the same Monte Carlo method to assess significance of the G-Null, CPMA, Sum-χ^2 and ARCHIE test statistics in our simulation studies, where these are described below in subsection **Additional test statistics included in the comparison**. In the simulations, we verify the empirical type 1 error of our Monte Carlo p-value calculation for the ADELLE, G-Null, CPMA, Sum-χ^2 and ARCHIE methods.

Simulation methods

In the simulations, we consider a setting in which we have summary statistics from association tests of a SNP with each of $D = 10^4$ expression traits, and we want to combine the summary statistics into a global test of the null hypothesis that the SNP is not associated with any of the traits. We use the ADELLE method and each of the 7 different methods described below in subsection **Additional test statistics included in the comparison** to perform the test. To assess type 1 error at level α, for $\alpha = 0.05, 0.01$ and 0.001, we generate 10^5 simulation replicates in which the SNP is not associated with any of the D traits and calculate each of the test statistics on each replicate. To assess type 1 error at smaller α levels of 10^{-4}, 10^{-5} and 2.5×10^{-6}, we instead generate 2×10^7 simulation replicates. For each α level and each testing method, we estimate type 1 error by the proportion of replicates in which the given testing method produced a p-value $< \alpha$. To compare power across methods, we generate 10^3 simulation replicates in which the SNP is associated with exactly A of the D traits, where we perform studies for each of several choices of A from 5 to 200, assuming a sample size of 10^3. For the case $A = 5$, the SNP explains 1.5% of the variance of each associated trait; for $A = 10$, 1% of the variance; for $A = 20$, 0.8% of the variance; for $A = 50$, 0.5% of the variance; for $A = 100$, 0.4% of the variance; and for $A = 200$, 0.2% of the variance of each associated trait. We compare the power of the different methods based on the proportion of replicates in which each method rejects the

null hypothesis. To simulate the data, we start by randomly choosing a true trait correlation matrix Ω (see S1 Text for details). To perform the Monte Carlo p-value calculation described in the previous subsection, we simulate trait values for the 10^4 traits for 10^3 individuals, from which we estimate $\hat{\Omega}$ as described in subsection **ADELLE: extension of the ELL method to dependent traits** above. The estimated $\hat{\Omega}$ is then used in the Monte Carlo p-value calculation. In each simulation replicate, we simulate a vector of Z scores of length D from a multivariate normal distribution with mean vector $\mu = 0$ under the null hypothesis and with correlation matrix Ω. Under the alternative hypothesis, we simulate the Z scores from the same distribution as under the null hypothesis but where the mean vector has exactly A of the D entries equal to c_A (where c_A is chosen so that the SNP explains the specified proportion of variance listed above for each A). The remaining $D - A$ entries of the mean vector are equal to 0.

Additional test statistics included in the comparison

We assessed the type 1 error and power of ADELLE as well as the following methods for testing the global null hypothesis that a given SNP is not associated with any expression trait. For each replicate a vector of (dependent) Z scores was generated as described above and given as input to each method.

Min-P. For each Z score vector, to obtain its p-value using the Min-P method, we calculate $\pi_{(1)}D$, the Bonferroni-corrected minimum p-value among all the test statistics in the vector Z.

Simes. For each Z score vector, to obtain its p-value using the Simes method [16], we calculate $\min_{1 \leq d \leq D} \frac{\pi_{(d)}D}{d}$. The Simes p-value is closely related to the Benjamini-Hochberg procedure [35] for controlling FDR.

Cauchy. For each Z score vector, to obtain its p-value using the Cauchy method [17], we calculate $F_C(D^{-1} \sum_{d=1}^{D} F_C^{-1}(\pi_d))$, where F_C is the Cauchy cdf.

G-Null. The G-Null method is a simpler variation on the ADELLE method. In the ADELLE method, the estimated trait correlation matrix $\hat{\Omega}$ is used both in (1) calculating the l-values used to construct the test statistic and in (2) the Monte Carlo p-value calculation. In contrast, in the G-Null method, the l-values are calculated assuming independence, and $\hat{\Omega}$ is used only for the Monte Carlo p-value calculation. As a result, both methods would be expected to have correct type 1 error (assuming that Ω is well-estimated by $\hat{\Omega}$), and ADELLE would be expected to have higher power when there is dependence among the traits. In simulations, we can investigate to what extent using $\hat{\Omega}$ to calculate the l-values allows ADELLE to improve power over G-Null.

For each Z score vector, to obtain its p-value using the G-Null method, we first calculate the ELL test statistic given in Eq 5. If the elements of Z were independent, we could calculate a p-value by the method given in subsection **Global trans-eQTL testing with ELL**. However, because they are dependent, we instead obtain a Monte Carlo p-value using the method described in subsection **Monte Carlo p-value calculation** above.

Sum-χ^2. The test statistic is $\sum_{d=1}^{D} Z_d^2$ the sum of the squares of the Z scores in the vector. If the Z scores were independent under the null hypothesis, this test statistic would be χ_D^2 distributed. However, in this setting they are dependent, and we instead obtain a Monte Carlo p-value using the method described in subsection **Monte Carlo p-value calculation** above. The Sum-χ^2 test is equivalent to SKAT [29] with equal weights, where the roles of SNPs and traits are reversed, i.e., one SNP is tested with many traits rather than one trait with many SNPs.

CPMA. We used our own implementation of the method described in [14] to compute the CPMA statistic. The CPMA test models the elements of the vector $(-\log(\pi_1), .., \log(\pi_D))$ as

i.i.d. draws from an Exponential(λ) distribution and tests the null hypothesis $\lambda = 1$ vs. the alternative $\lambda \neq 1$. We compute the likelihood ratio statistic for this test. Because the chi-squared null distribution does not hold when the *p*-values are correlated, we instead used the Monte Carlo p-value calculation described above.

ARCHIE. We used the ARCHIE software to obtain the *q* score [13] of the component containing the given SNP. The ARCHIE method [13] requires Monte Carlo to assess significance, so we used the Monte Carlo p-value calculation described above.

In addition, we considered both the generalized higher criticism [26] and generalized Berk-Jones [27] methods but were unable to successfully run the available software on the scale of problems we consider here.

Detection of trans eQTLs in an advanced intercross line

Gonzales et al. [4] described an advanced intercross line (AIL) of mice and undertook genome-wide association studies (GWAS) and eQTL mapping studies in this population. They report finding thousands of cis and trans eQTLs across three brain regions. Here, we focus on trans eQTL associations in the hippocampus region and use summary statistics to test for trans eQTL associations that were not significant in the original study. Gonzales et al. [4] define "trans" to mean that the SNP and gene are on different chromosomes, and we follow their definition in our analysis. Details of the data set and original analysis can be found in Gonzales et al. [4].

For expression traits in the hippocampus, Gonzales et al. determined that in their dataset a p-value threshold of 9.01×10^{-6} corresponded to genome-wide significance of 0.05 when correcting for SNP-wise multiple testing, based on a permutation analysis. This value of 9.01×10^{-6} would thus be an appropriate significance threshold for testing a single expression trait with all SNPs in the genome, and it would also be an appropriate threshold for a global testing method such as ADELLE or any of the other 7 methods described above, in which the p-values for a given SNP with each possible expression trait are combined into a single test statistic, resulting in one test performed for each SNP in the genome. However, if one instead takes a non-global-testing strategy of considering all the p-values for every possible pairing of a SNP and one of its trans traits, then in order to identify a SNP as a trans eQTL with a type 1 error rate of 0.05, it is necessary to correct for both the number of SNPs and the number of traits tested. For any SNP in this study there are approximately 14,000 trans genes against which it is tested. After doing a Bonferroni correction, we, therefore, consider a single SNP-trans gene association to be statistically significant if its p-value is less than 6.4×10^{-10}.

ADELLE only requires summary statistics and a trait correlation matrix, but the available results for this data set only include summary statistics for associations that had *p*-value less than 9.01×10^{-6}. To allow larger p-values to potentially contribute to the global test, we decided to regenerate the complete set of SNP-gene expression *Z* scores. We downloaded the G50–56 LGxSM AIL GWAS data set available at https://palmerlab.org/protocols-data/, filtered the genotype dosage file to include only those mice that had gene expression data in the hippocampus, and pruned SNPs that were in complete LD using Plink [36], leaving 9671 SNPs across the genome. We used the downloaded gene expression matrix *Y* for the hippocampus that had all covariates regressed out and was quantile normalized. We computed the sample trait correlation matrix based on the 15,071 autosomal, expressed genes in *Y* and applied our regularization method to obtain $\hat{\Omega}$. Following the code provided in the supplementary information of Gonzales et al. [4], we used the software package Gemma [37] to construct LOCO GRMs and to do association analysis between each SNP-gene expression pair, which is the equivalent of performing $\sim 15,000$ different GWASs. Using the Monte Carlo assessment of

significance based on 10^7 replicates, we determined an empiric ADELLE p-value for every SNP and used the genomewide significance cutoff of 9.01×10^{-6} that is needed to correct for SNP-wise multiple testing in this dataset.

Detection of trans-eQTLs in the eQTLGen data

Võsa et al. [6] performed cis- and trans-eQTL analyses based on whole-blood expression levels for over 30,000 individuals through the eQTLGen Consortium. For their trans-eQTL analysis, they considered 10,317 SNPs that are each significantly associated with a complex trait, and they defined a SNP-gene pair to be "trans" if the SNP is more than 5 Mb from the gene. They discovered 37% of the trait-associated SNPs as being associated with a distal expression trait, at an FDR of 0.05. However, they flagged 8,984 (12.2%) of the significant trans associations as being potentially caused by cross-mapping of the gene within the cis region of the SNP.

We downloaded the summary statistics for association for the 10,317 SNPs and 19,942 genes from the trans-eQTL analysis of Võsa et al. [6]. For our analysis, we defined "trans" to mean that the SNP is on a different chromosome from the gene. We removed from further consideration all the genes flagged by Võsa et al. as potentially cross-mapping to more than one chromosome, resulting in 18,403 genes remaining. Of these, 15,753 were also in the GTEx dataset, so we restricted to this subset of genes. We used the GTEx data (v10) to estimate the correlation matrix for the expression data of these 15,753 genes, and we then regularized the correlation matrix according to the procedure described in S1 Text. Of the 10,317 SNPs, we only considered the 9,918 SNPs for which association tests were available with all of the genes not on the same chromosome as the SNP.

To address the question of whether there was evidence of additional trans-eQTL signal beyond that discovered by Võsa et al. [6], we removed the most extreme Z-scores from the dataset until there were no discoveries made among the remaining Z-scores at an FDR of 0.05. With the significant Z-scores removed, we then applied both ADELLE and Min-P to the remaining dataset. We used 2×10^7 Monte Carlo replicates to assess p-values for ADELLE. We then applied an FDR of 0.05 to discover SNPs based on their ADELLE or Min-P p-values.

Verification and comparison

Type 1 error

We undertook two type 1 error studies. In the first, we tested type 1 error at significance levels $\alpha = 0.05$, 0.01, and 0.001 for eight methods (ADELLE, G-Null, Min-P, Simes, Cauchy, Sum-χ^2, CPMA and ARCHIE). We performed 10^5 simulation replicates, and for the methods that use empirical p-values (ADELLE, G-Null, Sum-χ^2, CPMA and ARCHIE), we perform an additional 10^5 Monte Carlo replicates. (In S1 Text we show that having equal numbers of simulation and Monte Carlo replicates minimizes the variance of the type 1 error estimates for a fixed budget of total replicates.) As seen in Table 1, all methods control the type 1 error rate, with none of the estimated type 1 error rates significantly different from the nominal level. We undertook a larger simulation study to assess type 1 error at nominal levels 10^{-4}, 10^{-5} and 2.5×10^{-6} for ADELLE, G-Null, Min-P, Simes Cauchy, Sum-χ^2 and CPMA. We performed 20 million simulation replicates, and for the methods that use empirical p-values, we performed an additional 20 million Monte Carlo replicates. The ARCHIE software was much slower to run than the other methods, so it was not feasible to use it in a simulation study of this size. In Table 2 we can see that the type 1 error is well-controlled in all cases, although Sum-χ^2 seems to be slightly conservative in at least one case. Fig 1 depicts the QQ-plot of the 20 million ADELLE p-values in this larger simulation study. The p-values are well within the 95% simultaneous acceptance region for the uniform distribution, verifying that the ADELLE p-values

Table 1. Type 1 error rates of different global testing methods.

Method	Empirical type 1 error at nominal level		
	0.05	**0.01**	**0.001**
ADELLE	0.05101	0.01001	0.00102
G-Null	0.05026	0.00982	0.00096
Min-P	0.04959	0.01014	0.00089
Simes	0.05104	0.01018	0.00089
Cauchy	0.05057	0.01016	0.00089
Sum-χ^2	0.05004	0.00930	0.00091
CPMA	0.04922	0.00963	0.00083
ARCHIE	0.05037	0.00989	0.00090

For ADELLE, G-Null, Sum-χ^2, CPMA and ARCHIE, type 1 error is based on 10^5 simulation replicates and 10^5 Monte Carlo replicates for the empirical p-values, and the acceptance region for a test (at level 0.05) of whether the type 1 error differs from the nominal level is (0.04809, 0.05191) for nominal level 0.05, (0.00913, 0.01087) for nominal level 0.01, and (0.00072, 0.00128) for nominal level 0.001. See S1 Text for details on calculation of acceptance regions for empirical p-values. For Min-P, Simes and Cauchy, type 1 error is based on 10^5 simulation replicates, and the corresponding acceptance regions are (0.04865, 0.05135) for nominal level 0.05, (0.00938, 0.01062) for nominal level 0.01, and (0.00080, 0.00120) for nominal level 0.001.

https://doi.org/10.1371/journal.pgen.1011563.t001

are correctly calibrated. Similar plots for the other methods are given in S1 and S2 Figs. This validates the Monte Carlo p-value calculation approach that we use for ADELLE, G-Null, Sum-χ^2, CPMA and ARCHIE.

Power comparison

The results of the power simulations for all 8 methods considered are given in Tables A-F of S1 Text. Fig 2 shows the results for 6 of the 8 methods. (In order to make the plots less cluttered, the power of the Simes method was not included in Fig 2, as it was approximately equal to that

Table 2. Additional Type 1 studies at smaller α levels.

Method	Empirical type 1 error at nominal level		
	1e-04	**1e-05**	**2.5e-06**
ADELLE	9.725e-05	9.550e-06	3.20e-06
G-Null	9.510e-05	9.400e-06	2.30e-06
Min-P	9.920e-05	9.750e-06	2.30e-06
Simes	9.920e-05	9.750e-06	2.30e-06
Cauchy	9.920e-05	9.750e-06	2.30e-06
Sum-χ^2	8.275e-05*	8.800e-06	2.15e-06
CPMA	9.550e-05	1.085e-05	3.20e-06

A star denotes type 1 error that is significantly different from the nominal (based on a test at level .05). For ADELLE, G-Null, Sum-χ^2 and CPMA, type 1 error is based on 20 million simulation replicates and 20 million Monte Carlo replicates for the empirical p-values, and the acceptance region for a test (at level .05) of whether the type 1 error differs from the nominal level is (9.380e-05, 1.062e-04) for nominal level 1e-04, (8.040e-06, 1.196e-05) for nominal level 1e-05, and (1.520e-06, 3.480e-06) for nominal level 2.5e-06. See S1 Text for details on calculation of acceptance regions for empirical p-values. For Min-P, Simes and Cauchy, type 1 error is based on 20 million simulation replicates, and the corresponding acceptance regions are (9.562e-05, 1.044e-04) for nominal level 1e-04, (8.614e-05, 1.139e-04) for nominal level 1e-05, and (1.807e-06, 3.193e-06) for nominal level 2.5e-06.

https://doi.org/10.1371/journal.pgen.1011563.t002
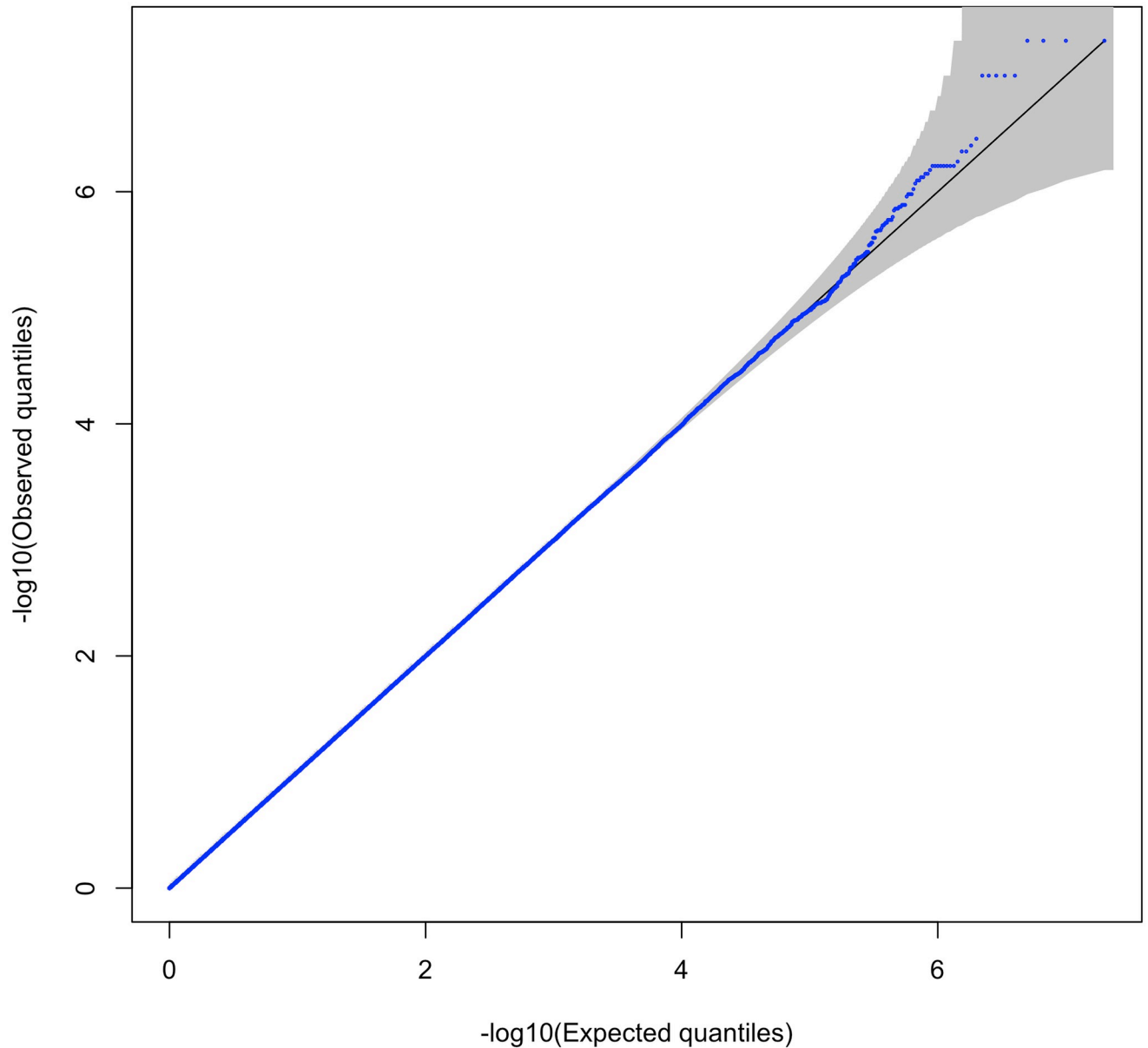
Fig 1. QQ-plot of ADELLE p-values from type 1 error study based on 2×10^7 simulation replicates and 2×10^7 Monte Carlo replicates. The shaded region depicts a 95% simultaneous acceptance region based on ELL [25]; see S1 Text for details.

https://doi.org/10.1371/journal.pgen.1011563.g001

of Min-P in all cases, and similarly, the power of the ARCHIE method was not plotted, as it was approximately equal to that of Sum-χ^2 in all cases.) It is particularly illuminating to examine the relative power of the methods across different numbers of associated expression traits for the tested trans e-QTL, where this is shown in Fig 3.

For each choice of the number of associated expression traits, we plot the power of each method divided by the maximum power observed across all the methods for that setting. We can see that the Min-P, Simes, and Cauchy methods all behave similarly. As expected, they perform best with a small handful of associated traits, e.g., in our simulations, these methods perform better than the other methods when 5 out of 10^4 of the tested traits are associated.
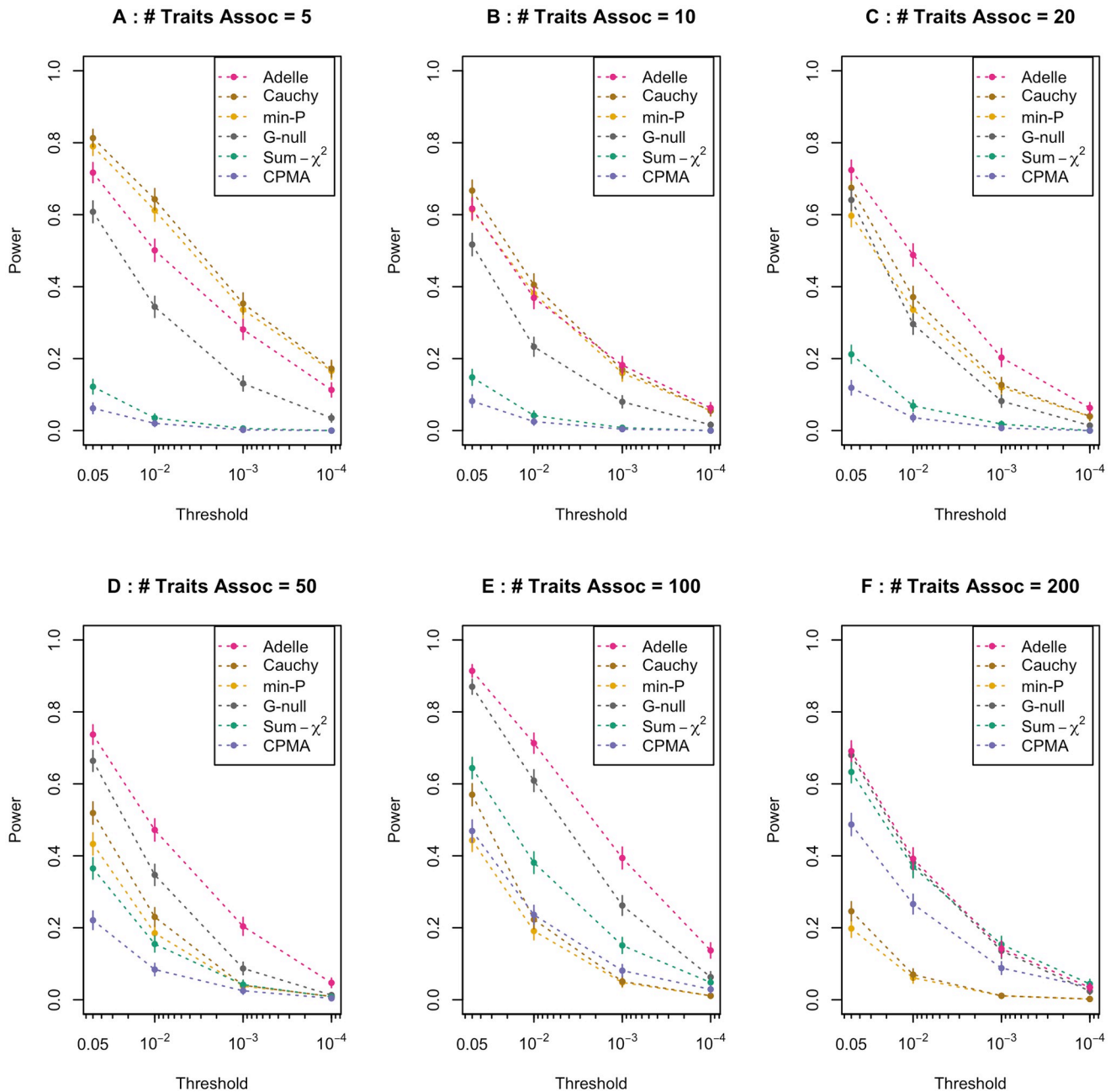
Fig 2. Power curves comparing different global testing methods for detecting a trans-eQTL. Each panel shows power for detecting a trans-eQTL plotted against the significance threshold of the association test, for 6 of the 8 methods considered. In order to make the plots less cluttered, the power of the Simes method was not plotted, as it was approximately equal to that of Min-P in all cases, and similarly, the power of the ARCHIE method was not plotted, as it was approximately equal to that of Sum-χ^2 in all cases (see Tables A—F of S1 Text). For ADELLE, $q = 0.2$ is used. In each panel, power is based on 10^3 simulated replicates. Each panel shows the plot for a setting in which a given number of expression traits are associated with the tested trans-eQTL. For each point of the plot, the corresponding vertical bar represents the 95% confidence interval for power. In Panel A, the number of associated expression traits is 5. In Panels B, C, D, E, and F, the numbers of associated expression traits are 10, 20, 50, 100, and 200, respectively.

https://doi.org/10.1371/journal.pgen.1011563.g002

However, their power is significantly below that of ADELLE with 20 or more associated traits, and they are the worst performing methods with 100 or 200 associated traits. At the opposite end of the spectrum are the Sum-χ^2, ARCHIE, and CPMA methods, which perform the worst with a small handful of associated traits, but outperform Min-P, Simes and Cauchy with a very
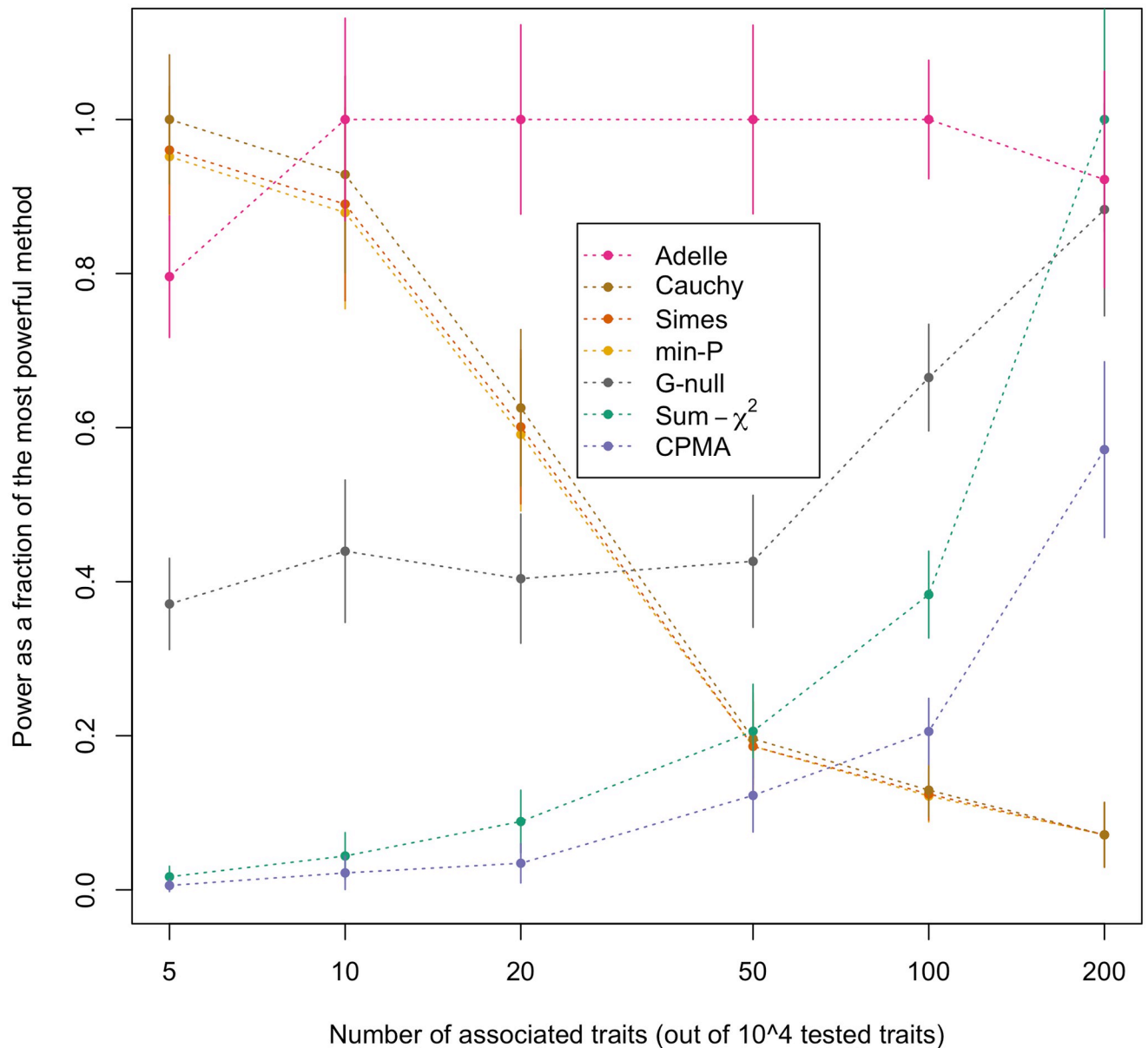
Fig 3. Relative power vs. number of associated traits for different global testing methods. Relative power at significance level 0.001, based on 10^3 simulated replicates, is plotted against the number of associated traits out of 10^4 total traits tested, for 7 of the 8 global testing methods considered. The curve for ARCHIE is visually indistinguishable from that for Sum-χ^2, so it is not plotted separately. For a given number of associated traits, relative power for a method is defined as its power divided by the maximum power achieved by any of the 8 methods for that setting. For each point of the plot, the corresponding vertical bar represents the 95% confidence interval.

https://doi.org/10.1371/journal.pgen.1011563.g003

large number of associated traits. Recall that the Sum-χ^2 method is equivalent to SKAT with equal weights (where the roles of SNPs and traits are reversed, i.e., one SNP is tested with many traits rather than one trait with many SNPs). Our results for the Sum-χ^2 method are consistent with previous observations about the performance of this class of methods [17, 18, 26], which tend to perform well with dense signals. In contrast, the ADELLE method emerges as the most powerful method when there are a moderate number of associated traits. When the number of associated traits is in the range of 10 to 200, ADELLE's power is either the

highest or not significantly different from the highest, and it clearly dominates all the other methods in terms of power when the number of associated traits is in the range of 20 to 100. Overall, ADELLE is the only method that consistently maintained high relative power across the entire range of scenarios we tested (5 to 200 associated traits out of 10^4).

Inclusion of the G-Null method allows us to examine the effect of using the estimated trait correlation matrix $\hat{\Omega}$ in calculating the l-values of the ADELLE statistic. The difference between G-Null and ADELLE is that G-Null uses an identity matrix in place of $\hat{\Omega}$ in calculating the l-values, which leads to a simple closed-form expression. However, from Figs 2 and 3, we can see that ADELLE has significantly greater power than G-Null for most scenarios. From this, we can conclude that it is important to use $\hat{\Omega}$ in calculating the l-values.

We also considered the impact on power of the choice of q, the proportion of order statistics considered by ADELLE. It is reasonable to ask whether choosing q larger than necessary could reduce power. We compared power for q = 5%, 10% and 20%, across the same settings as before, with the number of associated traits equal to 5, 10, 20, 50, 100 or 200 (out of 10^4 total traits), and these results are in Tables A-F and Fig A of S1 Text We found that there was little difference in power across these choices of q, with the average power difference being less than 2 percentage points between q = 5% and q = 20% in our results, and with this holding across the range of number of associated traits considered. Thus, the power of ADELLE seems to be quite robust to the choice of a larger than needed q. In the simulation results in Figs 1–3 and in the data analyses, we use q = 20%.

Computational benchmarking

The computational time needed to apply ADELLE can be divided into 3 parts: (1) obtaining the precompute grid; (2) data analysis; and (3) performing Monte Carlo replicates to obtain genomewide significance. Steps (1) and (2) are quite fast. For example, for the eQTL-Gen application, obtaining the precompute grid took $\sim$1.5 minutes and analyzing the data took $\sim$10 seconds on a 2020 iMac desktop (3.6 GHz 10-Core Intel Core i9 with 128 GB RAM). For the mouse AIL application, obtaining the precompute grid took < 1 minute and analyzing the data took $\sim$10 seconds. To obtain genomewide significance by Monte Carlo, we benchmarked 10^5 replicates at 6 minutes 8 seconds (on the same machine as above), where the compute time for Monte Carlo is linear in the number of replicates. For the Monte Carlo replicates, the most time-consuming step is simulation of multivariate normal random variables.

ADELLE is implemented in a freely downloadable software package that will be made available at https://www.stat.uchicago.edu/~mcpeek/software/index.html.

Applications

Detection of trans eQTLs in an advanced intercross line

In the supplementary information to their article, Gonzales et al. [4] list all trans associations (where a "trans association" is defined to be any association signal that is detected between a SNP and an expression trait for a gene where the SNP and the gene are located on different chromosomes) in the hippocampus that had p-value less than 9.01×10^{-6}, which corresponds to the threshold when correcting for all SNPs in the genome. Thus, many of the listed potential trans eQTLs do not meet the more stringent significance level of 6.4×10^{-10} required when correcting for both SNP-wise and trait-wise multiple testing.

Across the genome we replicated the trans eQTLs discovered by Gonzales et al. [4]. With the exception of one locus on chromosome 12, all trans eQTLs discovered with ADELLE also reached the significance threshold of 6.4×10^{-10} in the previous [4] analysis. However, in the
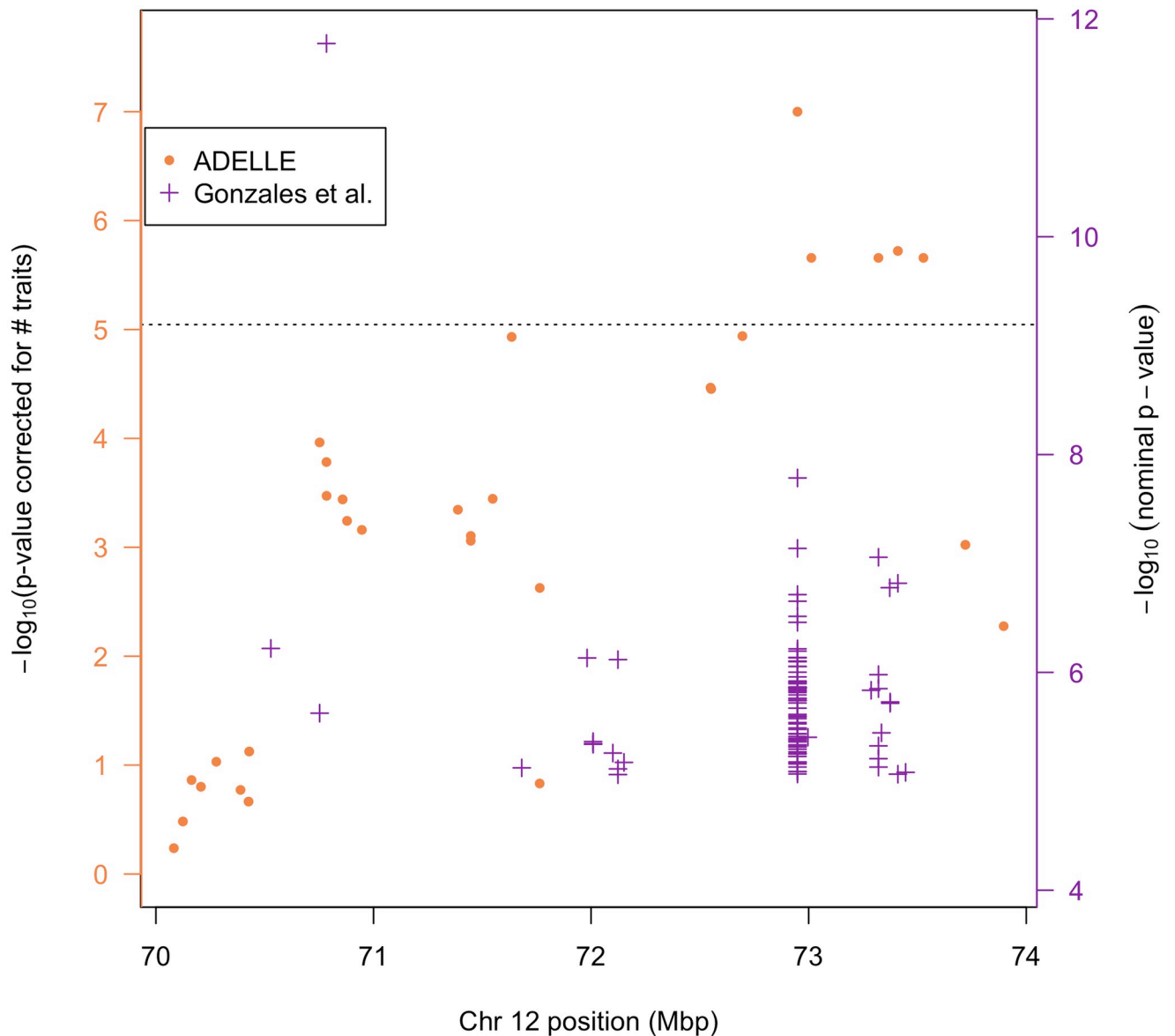
Fig 4. Trans eQTL associations in a region of chromosome 12. The purple "+" symbols in the figure represent single SNP-trait associations in the Gonzales et al. [4] analysis that had p-value less than 9.01×10^{-6}. Among the purple crosses, a single SNP may appear multiple times with different p-values in the figure, representing tests of the same SNP with different traits. The $-\log_{10}$ of these p-values are displayed on the right-hand axis. The left-hand axis represents a Bonferroni-corrected version of the right-hand axis, in which a correction for testing 14,078 traits is made. The ADELLE global testing result for each SNP in the region is shown as an orange dot whose p-value on the $-\log_{10}$ scale is shown on the left-hand axis, because the ADELLE p-value already accounts for testing multiple traits. The dotted line represents the genomewide significance threshold, correcting for both multiple SNPs and multiple traits. Note that this dotted line is more stringent than the one used in Gonzales et al. because we have applied a Bonferroni correction for the number of traits (i.e. gene expressions) tested at each SNP.

https://doi.org/10.1371/journal.pgen.1011563.g004

region of chromosome 12 from 70–74 Mbp, shown in Fig 4, there are several SNPs that are detected as significant by ADELLE but are not detected as significant trans eQTLs by the previous analysis [4] when multiple testing is accounted for. In the previous analysis, these SNPs each show a sub-significant level of association across multiple expression traits. The five SNPs in this chromosome 12 region that were detected as significant by ADELLE are listed in Table G in S1 Text.

In Fig 4, we can see both the ADELLE results and the previously reported results. Among the purple crosses (previous results), a single SNP may appear multiple times, with different *p*-values representing tests of the same SNP with different traits. The Min-P global testing result for a given SNP would be represented by the highest purple cross at a given location, with corresponding $-\log_{10}$ *p*-value given by the left-hand axis. The only SNP in this region that surpasses the threshold to be a trans eQTL in the previous analysis is at approximately 70.8 Mbp. This SNP is strongly associated with only a single trait, while its p-values for association with the remaining traits fit well to the uniform null distribution. In such a case, Min-P is expected to have high power. From Fig 4, we can see that the most significant result in the region by the ADELLE method (SNP rs262318378 at approximately 72.9 Mbp) had many small but sub-significant *p*-values for association in the previous analysis, and so is not significant by Min-P. This is a setting in which ADELLE is expected to have higher power. The other 4 significant ADELLE results in the region correspond to SNPs in high LD with rs262318378, so this set of 5 results may correspond to a single trans eQTL signal. Interestingly, for 2 of these 4 SNPs, the ADELLE test is significant, but none of the individual trait *p*-values for these SNPs was small enough to be reported by Gonzales et al. (i.e., none pass the nominal 9.01e-6 threshhold). In other words, none of the individual SNP-trait *p*-values for these SNPs even meets the significance threshold when correcting only for SNP-wise multiple testing, much less the more stringent standard of correcting for both SNP-wise and trait-wise multiple testing. This occurs because of an enrichment of many small p-values at that SNP, but where none of these p-values by iself is smaller than 9.01e-6. There is also a SNP at approximately 71.6 Mbp (which is not in strong LD with the 5 SNPs having significant ADELLE p-values) that is nearly significant using ADELLE, but for which there is also no single expression trait whose p-value is even as small as 9.01e-6. The data analysis results are consistent with the simulation results that showed ADELLE can gain power for trans eQTL mapping in a setting in which there are a moderate number of associated traits with relatively weak effects.

This region on chromosome 12 with many small but not statistically significant p-values was previously noted [4] and referred to as a "master" eQTL. Trans eQTLs acting as master regulators, that is affecting the expression levels of many genes, have been observed previously [2, 38, 39] and may often be located in trans eQTL hot spots [39, 40]. One possible mechanism for a trans eQTL acting as a master regulator is for it to be a cis eQTL for a transcription factor [39]. In fact, it has been found that a substantial fraction of trans eQTL effects are mediated through a target cis gene [41]. Among the five chromosome 12 SNPs in strong LD that are detected as significant by ADELLE, one was previously shown to be a cis eQTL [4] (see Table G in S1 Text). Other mechanisms by which a SNP may act on trans genes have been discussed [2] and may be relevant for these SNPs.

Detection of trans-eQTLs in the eQTLGen data

The eQTLGen dataset contains many highly significant trans-eQTL results. For example, Võsa et al. [6] prioritize 26 trans-eQTL detections (in their Supplemental Fig 14B) for having high replication rates and low cross-mapping with the cis-region of the the associated SNP. All of these trans-eQTL associations are detected by ADELLE as well. To identify additional trans-eQTL signal not detected by the FDR approach of Võsa et al., we remove the significant *Z* scores from the analysis (as described in subsection **Detection of Trans-eQTLs in the eQTL-Gen Data** of the **Description of the methods** section) and reanalyze the dataset with both ADELLE and Min-P. With the significant *Z*-scores removed from the data, Min-P discovers 0 SNPs, while ADELLE discovers 1,451 SNPs at FDR 0.05. This represents additional trans-eQTL signal beyond that discovered previously, showing that ADELLE is able to combine

multiple sub-significant signals to identify additional trans-eQTL signal in the data. In this case, all 1,451 SNPs were previously identified as trans-eQTLs, with the additional detections made by ADELLE representing additional gene expression traits that were not previously identified as being associated with these SNPs.

Discussion

For trans-eQTL mapping, in order to meet rigorous standards of genomewide significance, the common strategy of considering the entire set of p-values for testing each SNP against each trans trait requires a severe multiple testing correction, because both SNP-wise and trait-wise correction is required. The resulting threshold is too strict for anything other than extremely strong associations to pass. Since a trans-eQTL association signal is not expected to be particularly large, this strategy does not seem well-suited to detecting trans-eQTLs. A global testing strategy in which association test statistics for a single SNP are combined across multiple expression traits into a single test statistic for each SNP has the potential help alleviate this problem because the resulting global test p-values need only be corrected for the number of SNPs. Whether a global test actually represents an improvement can depend entirely on the form of the global test. For example, the global test based on Min-P which is one of the methods considered in our simulations is essentially the same as the common strategy.

We have developed a global testing method ADELLE that is tailored for trans-eQTL mapping. ADELLE is designed to have high power when a trans-eQTL is associated with multiple expression traits, where the proportion of associated traits is small as a subset of all traits tested, and where the individual effect sizes may be relatively weak. We have shown through simulation studies and reanalyses of (i) eQTLGen data and (ii) a mouse AIL data set that our method, ADELLE, is able to detect significant trans eQTL signal that would otherwise not be detected when only individual SNP-trait p-values are considered.

In our simulations, ADELLE was the only method that consistently maintained high relative power when the number of associated expression traits represented.1%–2% of the total number of traits tested, and it had significantly higher power than the other methods when the number of associated expression traits represented around 0.2%–1% of the total number of traits tested. These are particularly relevant ranges for trans eQTLs because it is expected that they will often be associated with many, rather than just a single, gene. In fact, as seen in our analyses of both the eQTLGen and the mouse AIL datasets, ADELLE is able to reject the global null hypothesis even when none of the individual trait p-values for a SNP are particularly small (i.e., they do not meet the significance threshold when correcting only for SNP-wise multiple testing, much less the more stringent standard of correcting for both SNP-wise and trait-wise multiple testing). This shows the ability of ADELLE to effectively combine multiple sub-significant association signals for a given SNP to enable genome-wide significant trans-eQTL detection.

ADELLE needs only summary statistics (consisting of (1) either Z scores or else p-values and the signs of the estimated effect sizes and (2) a sample correlation matrix for the traits) to perform its analysis. A distinct advantage of a method that only requires summary statistics is the ease with which they can be shared. This is especially relevant in human data where concerns regarding privacy and the risk of re-identification can make the sharing of original, individual level data problematic. In addition, sharing of summary statistics avoids the duplication of computation and effort that results when the original data must go through the process of quality control, normalization, testing, etc. multiple times. Sharing of the summary statistics is not without burden, however. The storage and sharing of summary statistics can be demanding, particularly in trans-eQTL studies where pairwise combinations of SNPs and genes result

in a very large number of tests. In practice, even the complete set of Z scores may not be available. An advantage of ADELLE is that it is only based on the qD most significant results for each SNP, where q is set by the user, so a full set of summary statistics is not required. In addition, ADELLE could in principle be modified to use only the summary statistics for tests that meet a certain pre-specified significance level, rather than using a fixed number of top results for each SNP.

Methods for trans-eQTL mapping commonly rely on a Monte Carlo assessment of significance, e.g. [6, 9, 10, 12–14, 42], and ADELLE does as well. With any Monte Carlo-based test, the smallest significance level at which the null hypothesis can be rejected depends on the number of replicates. Specifically, with R Monte Carlo replicates, the smallest significance level at which the null hypothesis can be rejected is $(R + 1)^{-1}$ [43]. With ADELLE, the testing of multiple traits per SNP is already fully accounted for within the test statistic, so the only multiple testing that needs to be considered after p-value calculation is testing across different SNPs, just as in ordinary GWAS. For example, if all SNPs in the human genome were tested against all trans-traits, the standard GWAS genome-wide significance level of 5×10^{-8} would be appropriate for the ADELLE tests, which would require an R of $\sim 2 \times 10^7$ [43]. This number of replicates is computationally feasible in ADELLE (see subsection **Computational benchmarking**) and is what we used in our analysis of the eQTLGen dataset. In fact, in our type 1 error simulation study, we perform a total of 4×10^7 replicates (2×10^7 simulation replicates and an additional 2×10^7 Monte Carlo replicates), which is approximately double the total number of replicates that would be needed to establish genome-wide significance even in a study that included every SNP in the genome. Within the genome-wide significant results, they can be prioritized by their ADELLE statistic, which varies continuously. Therefore, we do not see the use of Monte Carlo to establish genome-wide significance as being a major limitation. A more efficient approach to determine statistical significance is an area for future work.

Understanding the underlying biological mechanisms of trans acting effects on gene expression is a challenging task that will involve combining evidence from various lines of investigation. Here we focused on the statistical problem of identifying SNPs that affect variation in gene expression of distant genes. The combination of relatively weak effects with a very large number of tests make this a particularly difficult problem. The statistical methodology we developed for this problem, however, is general and can easily be applied to a larger set of common problems in genomics. Most any problem that involves an aggregating, or a set-based, test may benefit from our approach. For instance, tests of gene sets, SNP sets, and pathways fall into this category as do phenome wide association tests and tests which involve potential interactions when there are many possibly interacting variables, such as epistasis. In fact, as technology in the field of genomics progresses, and the number of variables, conditions and contexts grows with the size of data sets, we expect highly sensitive methods such as ADELLE to be a valuable tool in the process of developing deeper insights from the data.

Supporting information

S1 Text. Detailed methods and additional results. Detailed description of the methods, including a model for Z, regularization of the sample covariance matrix, the beta-binomial approximation, pre-computation for the ADELLE test, assessment of type 1 error with Monte Carlo p-values, and generation of the correlation matrix for simulations. Additional results consist of numeric power results corresponding to Figs 2 and 3, power results for ADELLE with different choices of q, and significant trans eQTL detections by ADELLE in a region of Chrom 12 in the mouse AIL dataset.
(PDF)

S1 Fig. QQ-plot of Min-P, Cauchy and Simes p-values from type 1 error study. P-values from 20 million simulation replicates under the null hypothesis are shown for each method. The p-values from Min-P, Cauchy and Simes are in blue, red and green, respectively. Because the values are so similar, the 3 curves lie almost perfectly on top of one another, except for the large p-values where the Bonferroni correction used for Min-P is conservative.
(TIF)

S2 Fig. QQ-plot of sum-χ^2, G-Null and CPMA p-values from type 1 error study. Empirical p-values based on 20 million simulation replicates and 20 million Monte Carlo replicates are shown for each method. The p-values from sum-χ^2, G-Null and CPMA are in green, red, and blue, respectively.
(TIF)

Acknowledgments

References

1. Yao C, Joehanes R, Johnson AD, Huan T, Liu C, Freedman JE, et al. Dynamic role of trans regulation of gene expression in relation to complex traits. Am J Hum Genet. 2017; 100(4):571–580. https://doi.org/10.1016/j.ajhg.2017.02.003 PMID: 28285768

2. Liu X, Li YI, Pritchard JK. Trans effects on gene expression can drive omnigenic inheritance. Cell. 2019; 177(4):1022–1034.e6. https://doi.org/10.1016/j.cell.2019.04.014 PMID: 31051098

3. Carlborg O, De Koning DJ, Manly KF, Chesler E, Williams RW, Haley CS. Methodological aspects of the genetic dissection of gene expression. Bioinformatics. 2005; 21(10):2383–2393. https://doi.org/10.1093/bioinformatics/bti241 PMID: 15613385

4. Gonzales NM, Seo J, Hernandez Cordero AI, St Pierre CL, Gregory JS, Distler MG, et al. Genome wide association analysis in a mouse advanced intercross line. Nat Commun. 2018; 9:5162. https://doi.org/10.1038/s41467-018-07642-8 PMID: 30514929

5. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020; 369(6509):1318–1330. https://doi.org/10.1126/science.aaz1776

6. Võsa U, Claringbould A, Westra HJ, Bonder MJ, Deelen P, Zeng B, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat Genet. 2021; 53(9):1300–1310. https://doi.org/10.1038/s41588-021-00913-z PMID: 34475573

7. Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, et al. Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. Nat Genet. 2003; 35(1):57–64. https://doi.org/10.1038/ng1222 PMID: 12897782

8. Liu X, Mefford JA, Dahl A, He Y, Subramaniam M, Battle A, et al. GBAT: a gene-based association test for robust detection of trans-gene regulation. Genome Biol. 2020; 21(1):211. https://doi.org/10.1186/s13059-020-02120-1 PMID: 32831138

9. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. 2013; 45(10):1238–1243. https://doi.org/10.1038/ng.2756 PMID: 24013639

10. Dutta D, VandeHaar P, Fritsche LG, Zöllner S, Boehnke M, Scott LJ, et al. A powerful subset-based method identifies gene set associations and improves interpretation in UK Biobank. Am J Hum Genet. 2021; 108(4):669–681. https://doi.org/10.1016/j.ajhg.2021.02.016 PMID: 33730541

11. Lan H, Stoehr JP, Nadler ST, Schueler KL, Yandell BS, Attie AD. Dimension reduction for mapping mRNA abundance as quantitative traits. Genetics. 2003; 164(4):1607–1604. https://doi.org/10.1093/genetics/164.4.1607 PMID: 12930764

12. Wang L, Babushkin N, Liu Z, Liu X. Trans-eQTL mapping in gene sets identifies network effects of genetic variants. Cell Genom. 2024; 4(4):100538. https://doi.org/10.1016/j.xgen.2024.100538 PMID: 38565144

13. Dutta D, He Y, Saha A, Arvanitis M, Battle A, Chatterjee N. Aggregative trans-eQTL analysis detects trait-specific target gene sets in whole blood. Nat Commun. 2022; 13:4323. https://doi.org/10.1038/s41467-022-31845-9 PMID: 35882830

14. Brynedal B, Choi J, Raj T, Bjornson R, Stranger BE, Neale BM, et al. Large-scale trans-eQTLs affect hundreds of transcripts and mediate patterns of transcriptional co-regulation. Am J Hum Genet. 2017; 100(4):581–591. https://doi.org/10.1016/j.ajhg.2017.02.004 PMID: 28285767

15. Banerjee S, Simonetti FL, Detrois KE, Kaphle A, Mitra R, Nagial R, et al. Tejaas: reverse regression increases power for detecting trans-eQTLs. Genome Biol. 2021; 22(1):142. https://doi.org/10.1186/s13059-021-02361-8 PMID: 33957961

16. Simes RJ. An improved Bonferroni procedure for multiple tests of significance. Biometrika. 1986; 73 (3):751–754. https://doi.org/10.1093/biomet/73.3.751

17. Liu Y, Chen S, Li Z, Morrison AC, Boerwinkle E, Lin X. ACAT: A fast and powerful p value combination method for rare-variant analysis in sequencing studies. Am J Hum Genet. 2019; 104:410–421. https://doi.org/10.1016/j.ajhg.2019.01.002 PMID: 30849328

18. Donoho D, Jin J. Higher criticism for detecting sparse heterogeneous mixtures. Ann Stat. 2004; 32 (3):962–994. https://doi.org/10.1214/009053604000000265

19. Donoho D, Jin J. Higher criticism for large-scale inference, especially for rare and weak effects. Stat Sci. 2015; 30(1):1–25. https://doi.org/10.1214/14-STS506

20. Berk RH, Jones DH. Goodness-of-fit test statistics that dominate the Kolmogorov statistics. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete. 1979; 47(1):47–59. https://doi.org/10.1007/BF00533250

21. Mary D, Ferrari A. A non-asymptotic standardization of binomial counts in higher criticism. Proc IEEE Int Symp Info Theory. 2014; p. 561–565.

22. Gontscharuk V, Landwehr S, Finner H. The intermediates take it all: asymptotics of higher criticism statistics and a powerful alternative based on equal local levels. Biom J. 2015; 57(1):159–180. https://doi.org/10.1002/bimj.201300255 PMID: 24914007

23. Gontscharuk V, Landwehr S, Finner H. Goodness of fit tests in terms of local levels with special emphasis on higher criticism tests. Bernoulli. 2016; 22(3):1331–1363. https://doi.org/10.3150/14-BEJ694

24. Moscovitch A, Nadler B, Spiegelman C. On the exact Berk-Jones statistics and their p-value calculation. Electron J Stat. 2016; 10:2329–2354.

25. Weine E, McPeek MS, Abney M. Application of equal local levels to improve Q-Q plot testing bands with R package qqconf. J Stat Softw. 2023; 106(10):1–31. https://doi.org/10.18637/jss.v106.i10

26. Barnett I, Mukherjee R, Lin X. The generalized higher criticism for testing SNP-set effects in genetic association studies. J Am Stat Assoc. 2017; 112(517):64–76. https://doi.org/10.1080/01621459.2016.1192039 PMID: 28736464

27. Sun R, Lin X. Genetic variant set-based tests using the generalized Berk–Jones statistic with application to a genome-wide association study of breast cancer. J Am Stat Assoc. 2020; 115(531):1079–1091. https://doi.org/10.1080/01621459.2019.1660170 PMID: 33041403

28. Goeman JJ, van de Geer SA, van Houwelingen HC. Testing against a high dimensional alternative. J Roy Stat Soc Series B Stat Methodol. 2006; 68(3):477–493. https://doi.org/10.1111/j.1467-9868.2006.00551.x

29. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet. 2011; 89(1):82–93. https://doi.org/10.1016/j.ajhg.2011.05.029 PMID: 21737059

30. Tzeng JY, Zhang D. Haplotype-based association analysis via variance components score test. Am J Hum Genet. 2007; 81(5):927–938. https://doi.org/10.1086/521558 PMID: 17924336

31. Cotsapas C, Voight BF, Rossin E, Lage K, Neale BM, Wallace C, et al. Pervasive sharing of genetic effects in autoimmune disease. PLOS Genet. 2011; 7(8):e1002254. https://doi.org/10.1371/journal.pgen.1002254 PMID: 21852963

32. Shorack GR, Wellner JA. Empirical processes with applications to statistics. Philadelphia: Society for Industrial and Applied Mathematics; 2009.

33. Moscovich A. Fast calculation of p-values for one-sided Kolmogorov-Smirnov type statistics. Comput Stat Data Anal. 2023; 185:107769. https://doi.org/10.1016/j.csda.2023.107769

34. Warton DI. Penalized normal likelihood and ridge regularization of correlation and covariance matrices. J Am Stat Assoc. 2008; 103(481):340–349. https://doi.org/10.1198/016214508000000021

35. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J Roy Stat Soc Series B Stat Methodol. 1995; 57(1):289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

36. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep, 2007-09; 81(3):559–575. https://doi.org/10.1086/519795 PMID: 17701901

37. Zhou X, Stephens M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat Meth. 2014; 11(4):407–409. https://doi.org/10.1038/nmeth.2848 PMID: 24531419

38. Small KS, Hedman ÅK, Grundberg E, Nica AC, Thorleifsson G, Kong A, et al. Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. Nat Genet. 2011; 43(6):561–564. https://doi.org/10.1038/ng.833 PMID: 21572415

39. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. Nat Rev Genet. 2015; 16(4):197–212. https://doi.org/10.1038/nrg3891 PMID: 25707927

40. Hasin-Brumshtein Y, Khan AH, Hormozdiari F, Pan C, Parks BW, Petyuk VA, et al. Hypothalamic transcriptomes of 99 mouse strains reveal trans eQTL hotspots, splicing QTLs and novel non-coding genes. eLife. 2016; 5:e15614. https://doi.org/10.7554/eLife.15614 PMID: 27623010

41. Pierce BL, Tong L, Chen LS, Rahaman R, Argos M, Jasmine F, et al. Mediation analysis demonstrates that trans-eQTLs are often explained by cis-mediation: a genome-wide analysis among 1,800 South Asians. PLoS Genet. 2014; 10(12):e1004818. https://doi.org/10.1371/journal.pgen.1004818 PMID: 25474530

42. Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, et al. Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. Nat Genet. 2003; 35:57–64. https://doi.org/10.1038/ng1222 PMID: 12897782

43. Hope ACA. A Simplified Monte Carlo Significance Test Procedure. J Roy Stat Soc Series B (Methodol). 1968; 30(3):582–598. https://doi.org/10.1111/j.2517-6161.1968.tb00759.x