**Supplemental Information for**

**Reconsidering the validity of transcriptome-wide association studies**

**Table of Contents**

[Mathematical structure of TWAS 2](#_Toc79957407)

[Relation to joint association testing 3](#_Toc79957408)

[Non-linear TWAS models 4](#_Toc79957409)

[Comparison with CoMM 4](#_Toc79957410)

[Simulation results 5](#_Toc79957411)

[Supplemental figures 8](#_Toc79957412)

[Supplemental Figure 1. Primary simulation type 1 error results, evaluated under a null of no genetic covariance 8](#_Toc79957413)

[Supplemental Figure 2. Primary simulation type 1 error results for CoMM. 9](#_Toc79957414)

[Supplemental Figure 3. Results for empirical type 1 error rate simulations for BMI. 10](#_Toc79957415)

[Supplemental tables 11](#_Toc79957416)

[Supplemental Table 1. Comparison of results for main TWAS analysis with TWAS on gene-tissue pairs with no detectable eQTL signal. 11](#_Toc79957417)

[Supplemental Table 2. Estimated type 1 error rate inflation for gene ENSG00000277639 in analysis of BMI 11](#_Toc79957418)

[Supplemental Table 3. Level of association and effect size in local genetic correlation analysis. 11](#_Toc79957419)

[Supplemental references 12](#_Toc79957420)

# Mathematical structure of TWAS

Given a specific genotype matrix , we can partition the gene expression and outcome phenotype into a local genetic component and a residual such that , and equivalently with as per equations (1) and (3). These local genetic components reflect all the genetic association with and that is (linearly) captured by the SNPs in , either because they are causal themselves or are in LD with other SNPs that are. The terms are thus by definition independent of .

 In the second stage of TWAS, the estimate of obtained from the eQTL data is used to compute , ie. the estimated local genetic component of , in the target GWAS sample. Under the assumption that both samples are drawn from the same population, this constitutes an unbiased estimate of for the individuals in this GWAS sample even though was obtained from an external sample (provided that the estimation procedure used to obtain guarantees an unbiased estimator of to begin with).

 We can express this estimate as , with reflecting the error of the estimate relative to its true value . This error is dependent on the specific eQTL sample used, and drawing a new sample thus generates a new value of . The form of the sampling distribution of , and thus that of given , depends on what estimation procedure is used to obtain . In some cases it may not have an easily expressed parametric form, though when using, for example, singular value decomposition pruning, it is a multivariate normal distribution.

 If we use linear regression to fit the model in equation (2), this specifies a conditional distribution of the outcome given . Crucially, this means that under this model we consider to be fixed, ignoring the uncertainty of as an estimate of . The coefficient can be expressed as , and thus with the null hypothesis is equivalent to . Note that , summing over each SNP . Each can be estimated from marginal SNP association statistics, which is how TWAS can be rewritten to be performed using only GWAS summary statistics for the outcome phenotype.

 From the above it follows that . Since as noted is independent of , and as is purely a linear function of , the term is zero and hence . As such, we can see that the relation tested by TWAS in equation (2) is the covariance between an estimate of on the one hand, and the true value of on the other.

 We can further rewrite , showing that this covariance is essentially offset from the true genetic covariance by the error term (denoted as in the main text). Averaging over the distribution of the expected value of will be zero, and hence will have a statistical distribution centered on under a sampling distribution that treats as random.

However, since the linear regression equation for (2) used in TWAS simply conditions on (treating it as fixed), rather than modeling its full distribution around , this distribution of around will not be accounted for in any test statistic based on . Since this effectively ignores a source of sampling variance, the resulting standard errors will be too small if evaluating the null hypothesis of no genetic covariance, ie. .

Moreover, since and therefore , this null hypothesis implies that . In practice this error term has an unknown but non-zero value specific to the data used, and as such, under the TWAS null hypothesis the true genetic covariance will not equal zero.

Effectively, relative to testing , the TWAS model assumptions treat the eQTL data as if it is the true underlying population of interest, rather than merely a sample drawn from that population. This can be illustrated by an analogy to the t-test. To compare the true population means and of groups and for some variable of interest, we would typically use an independent samples t-test to evaluate the null hypothesis . This would utilize a test statistic based on the sample means and as well as their standard errors and , accounting for the estimation uncertainty of both sample means.

In this context, TWAS is analogous to computing the sample mean for group , then performing a single sample t-test in group with the null hypothesis . The resulting test statistic would be based on , and , but as it does not incorporate it fails to account for the uncertainty in as an estimate of . As such, no inference can be performed on the relation between the two population means and . In the same vein, TWAS only considers the relationship that the *sample-specific* genetic component of the gene expression has with the true *population-level* genetic component of the phenotype. Since the specific sample on which this is based on is unlikely to be of particular scientific interest, this will have little practical use.

# Relation to joint association testing

As shown in the main text, the second stage of the TWAS analysis can be interpreted as a test of joint association between the SNPs in and the phenotype , under the constraint that the joint SNP effect size vector in equation (3) is constrained to . That is, the joint genetic association effect sizes with the outcome phenotype are assumed to be proportional to the estimated joint genetic association effect sizes for the gene expression.

 Since implies and vice versa, the null hypothesis being tested () is the same as in an unconstrained model, and only the power characteristics of the analysis differ. In cases where is (sufficiently close to) proportional to , imposing this proportionality constraint would not substantially reduce the fit of the model while using only a single parameter. Since can potentially contain hundreds or even thousands of parameters, this would yield a large increase in statistical power to detect an association. However, this comes at the expense of reducing power when this proportionality is not present, possibly reducing it to almost zero if the true associations between and are inconsistent with the directions in .

 It would be tempting to assume that because the weights used in TWAS are based on eQTL data, it is still possible to draw *some* sort of conclusions from significant TWAS results regarding the involvement of the expression of that gene and tissue in the genetics of the phenotype. After all, a significant result means that is at least to some degree associated with .

 However, this association can potentially come from chance correlations with the noise term included in , and the results in Supplemental Table 1 bear this out: even when no discernable genetic association with the gene expression is present (ie. ) and the computed is effectively reduced to just that noise term , there is still a considerable number of significant TWAS associations. The noise term will be just as capable of driving TWAS associations when a non-zero component is present, meaning that associations between and simply cannot be relied upon to draw any conclusions about the involvement of the gene expression in the genetic signal of . All that can be concluded from it is what it’s null hypothesis, , suggests: whether or not there is genetic association between and the SNPs in .

The nature of TWAS as a form of joint association testing does also suggests a relation with gene-based analysis1–3, since this performs a test of joint association of a set of SNPs with an outcome as well. However, interpretation of gene-based analysis hinges on the fact that the SNPs used are all located inside, or very close to, the transcription region of the gene being tested. It therefore localizes genetic association to that transcription region. By contrast, TWAS typically includes SNPs from an area orders of magnitude larger than the transcription region of the genes. As such they do not localize genetic association to nearly the same degree, and therefore cannot really be interpreted as a form of gene-based analysis, unless SNP selection is similarly restricted to a more localized area.

# Non-linear TWAS models

Several of the methods listed in Table 1 (‘non-linear models’) more explicitly approach TWAS as a form of joint association testing, and seek to generalize the model to relax this expectation of consistent effect directions, as follows. For a linear regression based on equation (2), the corresponding test statistic can be expressed as , with the correlation between SNP and the phenotype, and the corresponding (signed) weight derived from the eQTL data (and a scaling constant). This follows from the fact that ).

If the signs of and are either all the same or all opposite, each term will have the same sign, and the test statistic as a whole will tend to have a large value. If the signs are not consistent however, the for different SNPs will have different signs, and they (partially) cancel each other out when summed.

To get around this, we can define a generalized test statistic , raising the terms to the power before summing them. This has the dual effect of giving more weight in the test statistic to SNPs with stronger correlations with , as well as removing the expectation of consistent effect directions for even values of . A notable special case of this test statistic is , which yields the test statistic of the weighted SKAT model4.

Different ‘non-linear’ methods in Table 1 use this test statistic in different ways (with all the ‘linear’ methods using only ). Tang (2021)5 uses instead of , Xu (2017)6 uses an adaptive combination of and , and Zhang (2020)7 uses an adaptive combination of (where . Since all three of these methods include , they have better power to detect joint associations when effect sizes are not consistent with (though at the expense of somewhat lower power in scenarios where the linear model is a good fit). However, in terms of interpretation this moves even further away from testing a genetic relation between gene expression and phenotype, since these tests can no longer be defined in terms of , much less .

# Comparison with CoMM

Unique among the TWAS methods in Table 1, CoMM8 explicitly accounts for the uncertainty in the eQTL estimates. It accomplishes this by specifying the linear equations (translating to our notation) and (with the two equations sharing the same parameter ), which, if we substitute , translates to and . The parameters and are estimated simultaneously in a single model using both these equations, thus circumventing the use of a separately estimated and failing to account for the estimation uncertainty in that .

 Although this would therefore seem to resolve the issue discussed in this paper, that is not the case. The problem is that the model has can only account for genetic association between and via the parameter . And since in the CoMM model the two equations are estimated simultaneously, the parameter will be fitted to the relation between and as well as the relation between and . Thus is in effect somewhere intermediate between parameters and from equations (1) and (3).

 The issue with this becomes readily apparent when we inspect the model under the null hypothesis tested by CoMM. Under that null, the first equation is unchanged since does not occur in it. For the second equation however, we obtain . In other words, because under the CoMM model the genetic associations between and are parameterized as , when is set to 0 under the null this forces the assumption that and are completely independent (since is defined in the model as independent of ). In other words, this null is again equivalent to testing (since the CoMM model essentially assumes ), reducing to a joint association test just as the other TWAS models do (see section *Relation to joint association testing* above).

# Simulation results

The primary simulations (see *Methods - Primary simulations*) were performed to evaluate the type 1 error rates of TWAS under different conditions, when it is used to test the null hypothesis of no genetic covariance (ie. ). In this case, simulations were performed using a constant sample size of 10,060 (except for one set of conditions, which used for the eQTL sample size), varying only the local heritability values for the gene expression and outcome phenotype effects.

It should be noted however that simulations at particular heritability and sample size are approximately equivalent to simulations at a different sample size and a heritability . That is, simulations at, for example, 10% heritability and are equivalent to simulating at with a heritability of approximately 1.1%, or at with a heritability of approximately 52.6%. As such, the simulations at this sample size are representative of a range of scenarios at other sample sizes as well.

The main results for these simulations are depicted in Figure 2, and, as these show there is considerable type 1 error rate inflation in many of the conditions, varying as a function of the heritability parameters. Firstly, with larger the type 1 error rate increases (and the same would happen with increasing ). The main reason for this is that, as noted in the main text, under the actual null hypothesis tested by TWAS . As such, relative to the TWAS null, the simulations performed here essentially constitute a power analysis, and naturally the power increases as either the effect size (here determined by ) or the sample size increase.

Secondly, the results also show an opposite effect for , with the type 1 error rate inflation decreasing for larger . This is because the error rate inflation results from the uncertainty in the eQTL estimates that is essentially ignored in the TWAS model, and with larger the estimates improve and hence there is less uncertainty to ignore. The same is true, and for the same reasons, for increases in sample size for the gene expression data, as the difference between the left and right panels in Figure 2 demonstrate.

Finally, as shown in Figure 3, the error rate inflation also gets progressively worse at lower significance thresholds. The reason for this is that the relative difference between the true sampling distribution and the distribution actually used by TWAS gets larger the further we move into the tails of those distributions. This is also illustrated in the figure below, plotting the -log10 p-values as a function of the test statistic value for a sampling distribution, and for a corresponding distribution with underestimated standard errors ( and respectively, computing one-sided p-values). As shown, the discrepancy between the correct p-value (black) and the p-value that results from using the wrong distribution (red) gets larger as the test statistic value increases, and this causes a corresponding discrepancy in type 1 error rates for these distributions as well.



 Additional results for the simulations are given in Supplemental Figure 1, where the “Local rG” values show that the type 1 error rate is well controlled for these simulations when using LAVA local genetic correlation analysis. The “TWAS (TWAS null)” values show the same for the LAVA TWAS implementation when simulating under the TWAS null instead. It should also be noted that type 1 error rate simulations for existing methods have generally been of this latter kind (eg. Su (2018)9, Zhang 2020)7), only simulating the phenotype for given eQTL effects. Or alternatively, simulating both gene expression and phenotype, but under a model of no genetic association for the phenotype (eg. Xu (2017)6, Tang (2021)5, Yang (2019)8). This in part explains why the issues discussed in this paper were not discovered before.

Also shown are results for FUSION (note that these are based on only 1,000 iterations, and therefore somewhat noisier). For the elastic net model, the performance is the same as for as the LAVA TWAS, though for FUSION with the LASSO model the decrease in error rate inflation with increasing appears to be absent.

 Finally, Supplemental Figure 2 shows the simulation results for CoMM, with considerably more severe type 1 error rate inflation than the traditional TWAS models. The reason for this is that CoMM only imposes a soft constraint on , since the parameter it uses (see section *Comparison with CoMM* above) is partially determined by the genetic associations of with . By contrast, the traditional TWAS model effectively imposes a hard constraint of , leaving only the parameter free. This gives the CoMM model more freedom to fit those genetic associations, resulting in greater type 1 error rate inflation. Of note is that although CoMM does fully model the uncertainty in the eQTL estimates, the inverse relation between (or corresponding sample size) and error rate inflation is still present. The likely reason for this is that as the eQTL signal becomes stronger, the parameter is pulled more strongly towards , and therefore away from .

 Following the bivariate analyses, additional simulations were performed using real data estimates of and for each tested gene-tissue pair. This provides an estimate of the type 1 error rate inflation for each individual tested gene-tissue pair in the TWAS, given the levels of genetic signal for the gene expression and outcome found in the data for that gene and tissue. Although this cannot tell us whether specific significant TWAS hits are false positives or not, it does provide a reasonable approximation of the level of type 1 error rate inflation the TWAS for that phenotype.

 As shown in Table 3, the type 1 error rate inflation is extremely variable across genes and tissues. This is in large part due to the fact that the local and values vary considerably across genes and tissues, and, as the primary simulations demonstrated, the type 1 error rate inflation is strongly affected by both those values (and note that the sample size varies across tissues as well). This is also borne out by Supplemental Figure 3, showing the relation between the and estimates and the type 1 error rate inflation (at ). Both the negative relation of the inflation with and the positive relation with that were found in the primary simulations are visible here as well.

 An illustration of what this can look like is given in Supplemental Table 2, showing the gene with the most extreme estimated type 1 error rate inflation (gene ENSG00000277639) from the BMI analysis. The estimated error rates for all five of the tested tissues are orders of magnitude higher than the should be, and for both artery and skin gene expression the TWAS analysis also yields a p-value much lower than the corresponding local genetic correlation analysis, and highly significant in the case of skin.

This demonstrates the big discrepancies in p-value that can arise when the uncertainty in the eQTL estimates is accounted for compared to when it is not, and the repercussions of that for the conclusions drawn from the analysis. Based on the TWAS results, a researcher might conclude with considerable confidence that the expression of this gene, particularly in the skin, plays a role in the genetic aetiology of BMI. Yet the local genetic correlation analysis paints a very different picture, showing no real evidence for any such relation, with none of the five tissues anywhere close to significance.

Moreover, based on the local genetic correlation estimates it is clear that the relationship would be very weak even had it been significant, with only an estimated 2% of the local genetic signal for BMI being explained by the genetic component of the skin expression for gene ENSG00000277639. This makes it unlikely to be a biologically very relevant or informative relationship, regardless of its p-value. For comparison, Supplemental Table 3 shows a summary of significant local genetic correlations and their effect sizes.

# Supplemental figures



*Supplemental Figure 1. Primary simulation type 1 error results, evaluated under a null of no genetic covariance* **().** Type 1 error rates were evaluated for a significance threshold of 0.05; see Figure 2 and 3 and Methods - Primary simulations for details on simulation settings. TWAS analysis was performed with either the LAVA-TWAS implementation or with FUSION. For “TWAS (TWAS null)”, simulations were performed under the TWAS null model instead to validate the LAVA-TWAS implementation. Note that, for ease of comparison, y-axes for all four plots are on the same scale.



*Supplemental Figure 2. Primary simulation type 1 error results for CoMM.*Type 1 error rates were evaluated for a significance threshold of 0.05; see Figure 2 and 3 and Methods - Primary simulations for details on simulation settings. Simulations were performed either under the true null model with no genetic covariance (top row, ), or under the null model of regular TWAS (bottom row, ). Note that for CoMM, only 1,000 SNPs per block were used rather than the 5,000 used for the other simulations (except “TWAS - 1K SNPs in Supplemental Figure 1).



*Supplemental Figure 3. Results for empirical type 1 error rate simulations for BMI.*Results are for a significance threshold of 0.05, for each of the 84,567 gene-tissue pairs that were tested in the bivariate analyses. The type 1 error rate inflation levels found in the simulations are shown as a function of the (left) and (right) estimates found in for that gene-tissue pair in the bivariate analyses. For visual clarity, the horizontal axis for the plot has been put on a log scale.

# Supplemental tables

## Supplemental Table 1. Comparison of results for main TWAS analysis with TWAS on gene-tissue pairs with no detectable eQTL signal.

|  |  |  |
| --- | --- | --- |
|  | **Univariate eQTL signal (main analysis)** | **No univariate eQTL signal (additional analysis)** |
|  | **Gene-tissue pairs** | **Genes** | **Gene-tissue pairs** | **Genes** |
| **Phenotype** | **No. tested** | **No. signif** | **% signif.** | **No. signif** | **% signif.** | **No. tested** | **No. signif** | **% signif.** | **No. signif** | **% signif.** |
| BMI | 84,567 | 1,948 | 2.30 | 1,218 | 4.90 | 95,763 | 1,175 | 1.23 | 922 | 3.71 |
| Blood pressure | 54,622 | 687 | 1.26 | 477 | 1.92 | 37,475 | 188 | 0.50 | 170 | 0.68 |
| Diabetes | 18,967 | 290 | 1.53 | 189 | 0.76 | 7,551 | 137 | 1.81 | 98 | 0.39 |
| Educational attainment | 45,160 | 746 | 1.65 | 441 | 1.78 | 45,094 | 375 | 0.83 | 293 | 1.18 |
| Schizophrenia | 61,137 | 576 | 0.94 | 406 | 1.63 | 43,028 | 232 | 0.54 | 212 | 0.85 |

Inclusion in either category was based on univariate joint association p-values (see *Methods - Local genetic correlation*). For the main TWAS analysis, presence of genetic association with expression was defined as having a univariate p-value below ; for the additional analysis, absence of eQTL signal was defined as having a univariate p-value above 0.05. Significance for the TWAS in both analyses was determined using a significance threshold of for each phenotype , where is the sum of the two ‘no. tested’ columns.. A gene was counted as significant if at least one tissue had a significant TWAS association for that gene.

## Supplemental Table 2. Estimated type 1 error rate inflation for gene ENSG00000277639 in analysis of BMI

|  |  |  |
| --- | --- | --- |
| **Tissue** | **Local genetic correlation analysis** | **TWAS analysis** |
| **P-value** | **Estimate ()** | **Expl. Var.a (%)** | **P-value** | **Expl. Var.b (%)** | **Error rate** | **Inflation** |
| Artery (tibial) |  | -0.087 | 0.76 |  | 0.0028 | 0.0092 | 15615 |
| Esophagus (mucosa) |  | 0.025 | 0.06 |  | 0.0002 | 0.0041 | 6898 |
| Muscle (skeletal) |  | 0.005 | 0.00 |  | 0.0000 | 0.0271 | 45898 |
| Skin (lower leg) |  | 0.142 | 2.00 |  | 0.0078 | 0.0066 | 11205 |
| Thyroid |  | -0.030 | 0.09 | 0.152 | 0.0004 | 0.0047 | 7945 |

The significance threshold and expected error rate per analysis for BMI is , error rate inflation is computed as the observed error rate divided by this significance threshold.

a This reflects the percentage of the local genetic variance of BMI that can be accounted for by the (genetic component of) the gene expression, computed as (ie. the proportion of that can be explained by )

b This reflects the percentage of the total phenotypic variance of BMI (ie. ) that can be accounted for by the estimated expression

## Supplemental Table 3. Level of association and effect size in local genetic correlation analysis.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Phenotype** | **Signif. genes** | **% of univariate** | **Mean**  | **Max.**  | **Quantiles**  |
| **Univariate** | **Bivariate** | **By gene** | **By area** | **5%** | **25%** | **Median** | **75%** | **95%** |
| BMI | 16,493 | 695 | 4.21 | 26.6 | 17.7 | 61.9 | 7.16 | 11.0 | 15.8 | 21.8 | 34.8 |
| Blood pressure | 9,862 | 293 | 2.97 | 19.7 | 24.0 | 77.2 | 10.1 | 16.4 | 21.5 | 28.9 | 46.1 |
| Diabetes | 4,052 | 114 | 2.81 | 12.7 | 18.4 | 53.9 | 4.54 | 10.3 | 16.0 | 25.4 | 37.5 |
| Educational attainment | 9,235 | 292 | 3.16 | 17.2 | 23.5 | 58.2 | 9.92 | 15.6 | 21.8 | 29.3 | 44.6 |
| Schizophrenia | 12,552 | 228 | 1.82 | 11.6 | 21.2 | 50.3 | 10.2 | 14.5 | 19.2 | 26.8 | 37.1 |

Genes were counted as univariate significant if univariate tests were significant () for the phenotype and expression in at least one tissue. Genes were counted as bivariate significant if the local with the phenotype was significant (at thresholds listed in Table 2) for at least one tissue. The % of univariate by area was computed as the total genomic region covered by bivariate significant genes as a percentage of the total genomic region covered by univariate significant genes. The values are expressed as percentages, and quantify how much of the local genetic signal for the phenotype can be explained by the local genetic signal for the gene expression.

# Supplemental references

1. Neale, B. M. & Sham, P. C. The Future of Association Studies: Gene-Based Analysis and Replication. *Am. J. Hum. Genet.* **75**, 353–362 (2004).

2. Kang, G., Jiang, B. & Cui, Y. Gene-based Genomewide Association Analysis: A Comparison Study. *Curr. Genomics* **14**, 250–255 (2013).

3. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput. Biol.* **11**, e1004219 (2015).

4. Wu, M. C. *et al.* Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test. *Am. J. Hum. Genet.* **89**, 82–93 (2011).

5. Tang, S. *et al.* Novel Variance-Component TWAS method for studying complex human diseases with applications to Alzheimer’s dementia. *PLOS Genet.* **17**, e1009482 (2021).

6. Xu, Z., Wu, C., Wei, P. & Pan, W. A Powerful Framework for Integrating eQTL and GWAS Summary Data. *Genetics* **207**, 893–902 (2017).

7. Zhang, J., Xie, S., Gonzales, S., Liu, J. & Wang, X. A fast and powerful eQTL weighted method to detect genes associated with complex trait using GWAS summary data. *Genet. Epidemiol.* **44**, 550–563 (2020).

8. Yang, C. *et al.* CoMM: a collaborative mixed model to dissecting genetic contributions to complex traits by leveraging regulatory information. *Bioinformatics* **35**, 1644–1652 (2019).

9. Su, Y.-R. *et al.* A Mixed-Effects Model for Powerful Association Tests in Integrative Functional Genomics. *Am. J. Hum. Genet.* **102**, 904–919 (2018).