Testing for genetic interactions with imperfect information about additive causal effects

Gibran Hemani1,\*, Joseph E. Powell2,3,\*, Huanwei Wang4, Konstantin Shakhbazov, Harm-Jan Westra5, Tonu Esko6,7,8, Anjali K. Henders4, Allan F. McRae4, Nicholas G. Martin9, Andres Metspalu6, Lude Franke5, Grant W. Montgomery4, Michael E. Goddard10, Greg Gibson11, Jian Yang4,\*, Peter M. Visscher4,\*

Affiliations:

1. MRC Integrative Epidemiology Unit at the University of Bristol, Oakfield House, Oakfield Grove, Bristol, UK. BS8 2BN.
2. UNSW Cellular Genomics Futures Institute, School of Medical Sciences, University of New South Wales, Sydney, NSW, Australia
3. Garvan-Weizmann Centre for Cellular Genomics, Garvan Institute of Medical Research, Sydney, NSW, Australia
4. Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia
5. Department of Genetics, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9700 RB Groningen, The Netherlands.
6. Estonian Genome Center, University of Tartu, Tartu 51010, Estonia
7. Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts 02142, USA
8. Divisions of Endocrinology, Children's Hospital, Boston, Massachusetts 02115, USA.
9. Queensland Institute of Medical Research, Brisbane, Queensland 4006, Australia.
10. Faculty of Veterinary and Agricultural Science, University of Melbourne, Melbourne 3052 VIC, Australia
11. School of Biology and Centre for Integrative Genomics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA.

\* These authors contributed equally

## Summary

Statistical and computational barriers have made it difficult to gauge the importance of genetic interactions (epistasis) in complex traits. In a previous study1 we attempted to overcome these using a computational strategy that enabled a brute-force pairwise search, which yielded a modest number of instances of epistasis influencing whole blood gene expression levels in humans, and which replicated in two independent datasets. After publication they were further replicated by Wood et al (2014)2, but they also reported that inclusion of fine-mapped sequenced additive effects as covariates in the interaction models attenuated most of the interaction terms. An adequate explanation for the test’s behaviour has remained elusive since this correspondence3, and continued to be unsolved even after receiving attention from others. Using results of 501 genetic interactions that we previously reported to influence gene expression, alongside theory and simulations, this paper seeks to explain why the statistical method that we used, and which has been used in the field for decades, is unreliable, and that without sequence level data and assumptions of homogenous additive effects across individuals being met, the problems are difficult to guard against.

## Background

In Hemani et al (2014) we applied a 4 degree of freedom linear model test for each pairwise combination of 528,509 genotyped autosomal single nucleotide polymorphisms (SNPs), for each of 7,339 gene expression levels in whole blood. The statistical test attempted to capture any joint effect of two independent variants that was not explained by the marginal additive or dominance effect of either of the variants4. Here the additive by additive, additive by dominance, dominance by additive and dominance by dominance terms are jointly assessed in the interaction term. This effect decomposition is fundamental to basic quantitative genetic theory5, and has been used routinely in the linkage study era and the GWAS era6,7. The level of epistasis can be tested for statistical significance using an *F*-test with degrees of freedom, where *n* is the experimental sample size, assuming individuals are present in all pairwise genotype classes. A simpler variation is to parameterise the interaction term to include only the additive by additive term, and what follows in this paper applies to that approach also. **Supplementary Note 1** provides more background on the method. Our analysis, on 846 individuals, yielded 501 pairwise interactions that surpassed a family-wise significance threshold of (henceforth the H2014 interactions). The majority of these interactions were long-range ‘*cis-trans*’ associations, where one interacting variant was close to the gene whose expression level was influenced, and the other interacting variant was on a different chromosome. In two independent datasets, together comprising 2,131 individuals, 30 of these interactions replicated at a Bonferroni multiple testing correction ().

Soon after publication, these findings were further statistically replicated in an independent dataset by Wood *et al*. (2014)2. However, upon including fine-mapped sequenced additive effects as covariates in the interaction models, they found that most of the interaction effects substantially attenuated. We subsequently found a similar attenuation of effects in the original data by using fine-mapped imputed additive effects as covariates3. This exchange raised the question of why a standard method of analysis was giving rise to changeable results, which we explore here.

Wood *et al*. (2014) interpreted the original discovery interactions as so-called haplotype effects, a well-understood mechanism by which two loci can appear epistatic but be due to a simple additive effect. That is, the observed loci flank a causal variant and are in incomplete LD with each other and the causal variant; a statistical interaction between the observed loci can capture more of the additive variance of the causal variant than the marginal additive effects of both the observed loci combined. The haplotype effect model has subsequently been explored in more detail8. However, this explanation is not plausible for the majority of the H2014 signals which were *cis-trans* interactions, where the two interacting loci are on different chromosomes.

## Analysis

If the test statistic for a long-range interaction term can be attenuated with the inclusion of a single additive term, this implies that the interaction test statistic is inflated under the null hypothesis of no epistasis. To explore this assumption we began by estimating the genomic inflation factor for each of the 501 H2014 signals, which is a measure of the extent to which a family of test statistics departs from its distribution. In each case, we ran a genome-wide analysis where we performed an interaction test of the detected *cis*-SNP against every other SNP excluding those on the *cis* chromosome. The genomic inflation factor was then calculated for the interaction test statistics across the set of genome-wide tests (**Supplementary Methods**). **Supplementary Figure 1** shows that some loci have no obvious genomic inflation, while for many loci the inflation factor is much larger than that expected under the null. This is consistent with the idea that for many of the loci the test statistics are inflated. There are other possible explanations that could give rise to high genomic inflation factors, such as an epistatic polygenic component, though this is unlikely given the discovery sample size9, and the simplest interpretation here is that the *F*-statistics are departing from the null distribution in a way that signifies a problem with the data context.

We explored the theoretical mechanism by which the classic interaction test statistic can be inflated when only one of the interaction variants is in LD with a causal additive variant, which mimics the *cis-trans* interactions that form the majority of the H2014 signals. Reducing the problem to a simplified scenario in which individuals are haploid and the additive genetic effects explain all the phenotypic variance, we find that the residuals from a linear model are a mixture of normal and binomial distributions (**Supplementary Note 2**). This leads to systematic inflation or deflation of the F-statistic. We also show that as the effect of the unobserved additive variant gets larger, a larger proportion of variance of the residuals arises from the binomial distribution. Under this model we show that both the mean and the variance of the expected F value from the classical interaction model are increased. This mechanism is entirely separate from the sources of test statistic inflation that have been previously suggested.

Following this finding, we used simulations to explore the behavior of the test statistic in the diploid context with *cis-*acting additive effect sizes attempting to mimic the H2014 signals (**Supplementary Methods**). We began by recreating the conditions within the *MBNL1* locus, where 11 independent *cis-trans* associations were originally discovered, of which five replicated at the Bonferroni level (**Supplementary Note 3**). These simulations show that the genomic inflation factor relates strongly to the variance explained by the additive causal effect (**Supplementary Figure 2**), and that as genomic inflation grows, the number of false positive interactions grows (**Supplementary Figure 3**). We also observe that it is possible to obtain several false discovery signals per simulation even when the genomic inflation factor is low. This is consistent with the variance of the test statistic being inflated as predicted from our theory (**Supplementary Note 2**). Extending these simulations to other loci amongst the H2014 signals resulted in less inflation and lower false discovery rates because we are no longer ascertaining for a locus that is known to have high inflation and high replication rates.

We extended the simulations to evaluate the impact of the test statistic inflation on replication rates of type 1 errors from the discovery sample (**Supplementary Note 4**). We observed that the genomic inflation factor between independent discovery and replication datasets tends to be strongly correlated (**Supplementary Figure 4**). However, if the discovery had a significant interaction due to test statistic inflation, they were seldom independently replicated at the Bonferroni threshold, even at the *MBNL1* locus which showed a relatively high replication rate in the original study (**Supplementary Figure 5**).

The implied solution to avoiding the interaction test statistic inflation is to control for the fine-mapped cis-additive eQTLs. However, this may not reliably control the type 1 error rate under at least two scenarios. First, we explored the impact of measurement error in the *cis* additive causal variant (**Supplementary note 5**). We found that imperfectly adjusting for the additive effect due to realistic levels of imputation error at the *cis* additive causal variant led to poor control of the genomic inflation factor (**Supplementary Figures 6** and **7**).

Second, we evaluated the influence of additive effect heterogeneity on the interaction test statistic inflation (**Supplementary Note 6**). Here, the additive causal variant is simulated to have varying effects across individuals, and when estimating its average effect in the population its variance is only partially captured. The test statistic inflation will not be fully controlled by fitting the additive effect as a covariate, even if the additive variant is sequenced without error (**Supplementary Figure 8**).

There is a long history of problems arising in genetic analysis due to the interplay between statistical tests and background genetic architecture being poorly understood or experimental design being misaligned10,11. In the case of the *F*-statistic used for detecting epistasis, the problem of inflation that we describe here arises due to two forces. First, when there is imperfect LD between causal variant with large additive effect size and a tagging locus nearby, the mean and the variance of the test statistic for interaction terms of the tagging locus will be inflated. Second, a broad search for epistasis, in which strict significance thresholds are applied, is liable to ascertain for loci with large additive effects and specific LD properties that maximises the interaction test statistic inflation.

Going forwards, adjusting for fine-mapped additive effects should be done routinely when testing for interactions as it is likely to reduce test statistic inflation at the least. Per locus permutation testing strategies will be difficult to apply at scale, but could serve as a post-discovery sensitivity analysis, though their interpretation is unlikely to be straightforward (**Supplementary Figure 9**). If there is no large additive effect, as is the case with most complex traits and for most trans regions of ’omic variables, then the problem of the residual being a mixture of binomial and normal distributions is unlikely to be substantial.

## Acknowledgements

GH was funded by the Wellcome Trust and Royal Society [208806/Z/17/Z]. JEP is supported by National Health and Medical Research Council Fellowship (APP1175781). JY and PMV are supported by the Australian Research Council (DP160101343, DP160101056, FT180100186 and FL180100072) and the Australian National Health and Medical Research Council (1078037, 1078901, 1113400, and 1107258). We are extremely grateful to all the families who took part in the ALSPAC study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

## References

1. Hemani, G. *et al.* Detection and replication of epistasis influencing transcription in humans. *Nature* **10**, 249–53 (2014).

2. Wood, A. R. *et al.* Another explanation for apparent epistasis. *Nature* **514**, E3–E5 (2014).

3. Hemani, G. *et al.* Another Explanation for Apparent Epistasis. *Nature* **514**, E5 (2014).

4. Hemani, G., Theocharidis, A., Wei, W. & Haley, C. EpiGPU: exhaustive pairwise epistasis scans parallelized on consumer level graphics cards. *Bioinformatics* **27**, 1462–5 (2011).

5. Cockerham, C. C. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* **39**, 859–882 (1954).

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

7. Wei, W.-H., Hemani, G. & Haley, C. S. Detecting epistasis in human complex traits. *Nat. Rev. Genet.* (2014). doi:10.1038/nrg3747

8. de los Campos, G., Sorensen, D. A. & Toro, M. A. Imperfect Linkage Disequilibrium Generates Phantom Epistasis (&amp; Perils of Big Data). *G3&amp;#58; Genes|Genomes|Genetics* **9**, 1429–1436 (2019).

9. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807–12 (2011).

10. Hemani, G. *et al.* Inference of the Genetic Architecture Underlying BMI and Height with the Use of 20,240 Sibling Pairs. *Am. J. Hum. Genet.* **93**, 865–875 (2013).

11. Gaunt, T. R. & Davey Smith, G. ENOS and coronary artery disease: Publication bias and the eclipse of hypothesis-driven meta-analysis in genetic association studies. *Gene* **556**, 257–258 (2015).