Matters arising

Phantom epistasis between unlinked loci

https://doi.org/10.1038/s41586-021-03765-z

Received: 3 June 2016

Accepted: 23 June 2021

Published online: 11 August 2021

Check for updates

Gibran Hemani[1,2,16]✉, Joseph E. Powell[3,4,16], Huanwei Wang[5], Konstantin Shakhbazov[15], Harm-Jan Westra[6], Tonu Esko[7,8,9], Anjali K. Henders[5], Allan F. McRae[5], Nicholas G. Martin[10], Andres Metspalu[7], Lude Franke[6], Grant W. Montgomery[5], Michael E. Goddard[11], Greg Gibson[12], Jian Yang[5,13,14,16] & Peter M. Visscher[5,16]

ARISING FROM G. Hemani et al. *Nature* https://doi.org/10.1038/nature13005 (2014)

Statistical evidence for genetic interactions[1] (epistasis) involving loci that are unlinked can be attenuated by fine-mapping additive effects at one of the interacting loci[2], but the reason for this has remained unclear[3]. Here we show, using theory, simulation and data on the 501 genetic interactions we previously reported[1] to influence gene expression, a previously unrecognized property of the gold-standard statistical test to detect interactions, namely that the presence of imperfectly tagged additive causal variants can lead to phantom epistasis between unlinked markers. Therefore, the false positive rate in studies that use the test may not be sufficiently controlled and, to our knowledge, no current statistical fix exists for this problem.

In Hemani et al.[1] we applied a four 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 variants[4]. 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 theory[5], and has been used routinely in the linkage study era and the genome-wide association study (GWAS) era[6–8]. The level of epistasis can be tested for statistical significance using an F-test with 4, $n-9$ degrees of freedom, in which 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). Our analysis, on 846 individuals, yielded 501 pairwise interactions that surpassed a family-wise significance threshold of $P < 2.31 \times 10^{-16}$ (hereafter referred to as the H2014 interactions). Most of these interactions were long-range '*cis-trans*' associations, in which 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 ($P < 0.05/501$).

Soon after publication, these findings were further statistically replicated in an independent dataset by Wood et al.[2]. However, after 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 covariates[3]. 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.[2] 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 linkage disequilibrium 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 detail[9]. However, this explanation is not plausible for most of the H2014 signals that were *cis–trans* interactions, in which 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 indicates 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 in which we performed a genome-wide 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 genome-wide interaction test statistics (Supplementary Methods). Some loci have no obvious genomic inflation, whereas for many loci the inflation factor is much larger than that expected under the null (Supplementary Fig. 1). 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, although this is unlikely given the discovery sample size[10], and the simplest interpretation here is that the F-statistics depart 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 linkage disequilibrium with a causal additive variant, which mimics the *cis–trans* interactions that form most of the H2014 signals.

[1]MRC Integrative Epidemiology Unit, University of Bristol, Oakfield House, Oakfield Grove, Bristol, UK. [2]Population Health Sciences, University of Bristol, Bristol, UK. [3]UNSW Cellular Genomics Futures Institute, School of Medical Sciences, University of New South Wales, Sydney, New South Wales, Australia. [4]Garvan-Weizmann Centre for Cellular Genomics, Garvan Institute of Medical Research, Sydney, New South Wales, Australia. [5]Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia. [6]Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. [7]Estonian Genome Center, University of Tartu, Tartu, Estonia. [8]Medical and Population Genetics, Broad Institute, Cambridge, MA, USA. [9]Divisions of Endocrinology, Children's Hospital, Boston, MA, USA. [10]Queensland Institute of Medical Research, Brisbane, Queensland, Australia. [11]Faculty of Veterinary and Agricultural Science, University of Melbourne, Melbourne, Victoria, Australia. [12]School of Biology and Centre for Integrative Genomics, Georgia Institute of Technology, Atlanta, GA, USA. [13]School of Life Sciences, Westlake University, Hangzhou, Zhejiang, China. [14]Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang, China. [15]Unaffiliated. [16]These authors contributed equally: Gibran Hemani, Joseph E. Powell, Jian Yang, Peter M. Visscher. ✉e-mail: g.hemani@bristol.ac.uk

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 behaviour of the test statistic in the diploid context with *cis*-acting additive effect sizes that attempt to mimic the H2014 signals (Supplementary Methods). We began by recreating the conditions within the *MBNL1* locus, in which 11 independent *cis*–*trans* associations were originally discovered, 5 of which were 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 Fig. 2), and that as genomic inflation grows, the number of false positive interactions grows (Supplementary Fig. 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 among 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 effect 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 Fig. 4). However, if the discovery had a significant interaction owing to inflation of the test statistic, 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 Fig. 5). We also found that the sign of the most significant interaction term was not likely to be replicated more than chance ($P = 0.83$), in contrast to the H2014 signals originally reported.

The implied solution to avoiding the interaction test statistic inflation is to control for the fine-mapped *cis*-additive expression quantitative trait loci (eQTLs). However, this may not reliably control the type 1 error rate under at least two scenarios. First, we explored the effect 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 Figs. 6, 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 Fig. 8).

There is a long history of problems arising in genetic analysis owing to the interaction between statistical tests and background genetic architecture being poorly understood or experimental design being misaligned[11,12]. In the case of the *F*-statistic used for detecting epistasis, the problem of inflation that we describe here arises owing to two forces. First, when there is imperfect linkage disequilibrium 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 linkage disequilibrium properties that maximize the interaction test statistic inflation.

Going forwards, adjusting for fine-mapped additive effects should be done routinely when testing for interactions, as in many situations it will attenuate the test statistic inflation described here. Per locus permutation testing strategies will be difficult to apply at scale, but could serve as a post-discovery sensitivity analysis, although their interpretation is unlikely to be straightforward (Supplementary Fig. 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.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-03765-z.

Data availability

The gene expression data used to generate the original H2014 paper is available at the Gene Expression Omnibus (GEO) under accession code GSE53195. The ALSPAC genotype data can be accessed via http://www.bristol.ac.uk/alspac/. This study makes use of data from dbGaP (accessions phs000428.v1.p1) and EGA (accessions EGAS00001000108 and EGAS00001000090) (see Supplementary Information for a full list of acknowledgements to these datasets).

Code availability

Code is available at: https://github.com/explodecomputer/eqtl-2d.

1. Hemani, G. et al. Detection and replication of epistasis influencing transcription in humans. *Nature* **508**, 249–253 (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**, 1462–1465 (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.* **15**, 722–733 (2014).
8. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
9. de los Campos, G., Sorensen, D. A. & Toro, M. A. imperfect linkage disequilibrium generates phantom epistasis (& perils of big data). *G3 (Bethesda)* **9**, 1429–1436 (2019).
10. Yang, J. et al. Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807–812 (2011).
11. 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).
12. 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).

Acknowledgements G.H. was funded by the Wellcome Trust and Royal Society (208806/Z/17/Z). J.E.P. is supported by National Health and Medical Research Council Fellowship (APP1175781). J.Y. and P.M.V. 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). J.Y. is supported by the Westlake Education Foundation. 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.

Author contributions G.H., J.E.P., J.Y. and P.M.V. conceived the analyses. M.E.G. and P.M.V. developed the theory. G.H., J.E.P. and H.W. conducted the analyses. G.H., J.E.P., P.M.V., J.Y. and G.G. wrote the paper. All authors provided critical feedback on the interpretation of the results

Competing interests The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-03765-z.

Correspondence and requests for materials should be addressed to G.H.

nature research

Corresponding author(s): *Double-blind peer review submissions: write DBPR and your manuscript number here instead of author names.*

Last updated by author(s): *YYYY-MM-DD*

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	*Provide a description of all commercial, open source and custom code used to collect the data in this study, specifying the version used OR state that no software was used.*
Data analysis	*Provide a description of all commercial, open source and custom code used to analyse the data in this study, specifying the version used OR state that no software was used.*

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The gene expression data used to generate the original H2014 paper is available at the Gene Expression Omnibus under accession code GSE53195.
The ALSPAC genotype data can be accessed via http://www.bristol.ac.uk/alspac/

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Simulation sample sizes are drawn from the original sample sizes used in https://doi.org/10.1038/nature13005
Data exclusions	ALSPAC data was used as a linkage disequilibrium reference and was restricted to unrelated samples of European ancestry
Replication	Replication was achieved by splitting the ALSPAC sample into discovery and replication samples
Randomization	Randomisation was not built into the sampling process of the simulations here and is not relevant to the question
Blinding	Blinding is not relevant to the question here, no experimental data is being collected

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging