Dear Reviewers

We are grateful for your time and efforts on the proceedings related to our 2014 paper "Detection and replication of epistasis influencing transcription in humans". The reason that this matter has not yet been resolved is that it seems to be quite a complex case. We are grateful to the editors for being communicative about the process and allowing us to make our case to you over how we believe the matter should be concluded. In this letter we wish to:

* briefly explain the progress that has been made in understanding the problems that gave rise to the observations made in correspondence with Wood et al 2014
* make a case against retraction, though we are open to guidance from the reviewers and editors on any outcome

## Timeline

To begin, please allow us to demonstrate that we have acted in good faith throughout this. The timeline below shows that we have tried to be proactive in being up front about the problem and trying to understand and communicate the exact nature of the problem. We have also opposed the offer for voluntary retraction after it was requested in mid-2017, which has led to a lengthy editorial process. This is a position that we believe in, but we do regret that it has led to delays.

**Feb 2014:** Original paper (Hemani et al 2014, henceforth H2014) published

**Oct 2014:** Letter from Wood et al, and our response published, showing that cis-additive effects can abrogate almost all the interactions reported in H2014. While haplotype effects is a possible explanation for the cis-cis interactions, there is no forthcoming explanation for why this is happening for the cis-trans interactions.

**October 2015:** We submit a new correspondence (as an Addendum) that briefly describes work detailing an issue with the statistical test for interaction that the entire field uses routinely (and still does), which explains the phenomenon with the cis-trans interactions.

**June 2017:** Nature responds, offering to publish the letter as a Brief Communication Arising (BCA), and requesting that we voluntarily retract H2014

**October 2017:** Following conversation with editors at Nature, we formally request explanation for why H2014 should be retracted

**December 2017:** We are told to voluntarily retract by January 2018 or an editorial retraction will be issued

**January 2018:** We submit text for a voluntary retraction with a revised version of the BCA

**March 2019:** We receive a response from Nature, asking all authors to sign the retraction note. However, the retraction note has been edited in a manner that suggests editorial misunderstanding around the scientific issue with H2014

**July 2019:** Following more discussion, Nature offers that we write a more detailed manuscript describing the problems with the statistical test used in H2014.

This brings us to the present. We have a new manuscript that substantially expands on the original BCA that we sent in July 2015. We have written it as an appraisal of the statistical method for detecting genetic interactions that we used in H2014, which is widely used throughout the literature. We use the H2014 results as examples throughout.

## The nature of the problem

In H2014 we reported genetic interactions between variants associated with gene expression in humans. In most cases, the interacting variants were on different chromosomes, but one of the variants was close to the gene that was being influenced. We called these *cis-trans* interactions. There were also *cis-cis* interactions reported.

These interactions were statistically replicated by us in two independent datasets. They have subsequently been statistically replicated by at least two other groups in independent datasets also.

However, Wood et al 2014 showed that a fine-mapped additive variant can explain away the *cis-cis* and *cis-trans* interaction terms. They attributed the original interactions to 'haplotype effects', however this cannot explain the behaviour of the *cis-trans* interactions, as we pointed out in our Reply to Wood et al. 2014.

In H2014 we used a fundamental statistical test for detecting genetic interactions. This is the predominant test that has been used for detecting genetic interactions throughout the literature. It typically involves performing a 4 d.f. F-test to contrast a model of marginal effects and a model with marginal and interaction terms. We now show in detail the reason that some of the genetic interactions can be explained by fine-mapped cis-effects is that when the cis-interacting locus is in imperfect linkage disequilibrium (LD) with the *cis*-causal variant, the residual error becomes a mixture of binomial and normal distributions, violating the assumptions of the linear model and inflating the interaction term's test statistic. We demonstrate that the problem is likely difficult to solve.

The remainder of this letter will argue against retraction of H2014. We are finding this an uncomfortable thing to be doing because we do not wish to sound like we are eroding high standards in scientific publishing. But here are the arguments, and if it is deemed that retraction is appropriate then we will value that independent judgement.

## Definition of retraction

The Committee on Publication Ethics (COPE) suggests journals issue a retraction if: *"they have clear evidence that the findings are unreliable, either as a result of misconduct (e.g. data fabrication) or honest error (e.g. miscalculation or experimental error)"* (<https://publicationethics.org/retraction-guidelines>). The question here is whether the findings are unreliable. We believe that due to the replication of the findings, and that none of these issues have arisen due to miscalculation or experimental error, H2014 does not fall under this category. Rather, we correctly implemented the standard statistical test for detecting epistasis, and the signals were replicated by us and others in independent datasets. We subsequently learned that **interpretation** of those results has problems that were previously unknown. If this were to form a part of the definition of retraction, a large fraction of the scientific literature would be retracted, in our opinion.

Perhaps it is important to emphasise that this is not a simple miscalculation or experimental error. Highly esteemed statistical geneticists have since published an examination of the statistical test that we used and their conclusion was that it is unreliable for *cis-cis* interactions but should be fine for the *cis-trans* associations as long as the loci are uncorrelated (<https://www.ncbi.nlm.nih.gov/pubmed/30877081>). We now demonstrate that it is actually unreliable for *cis-trans* associations also, but the point is that understanding this problem has proven difficult even with the benefit of hindsight.

The Nature editors provided us another definition of retraction by which they were operating, which we never received in writing but was verbally communicated to us as approximately: “if information about the conclusions in a paper arises subsequent to its publication that would have originally precluded its publication, then we retract the original paper”. In our opinion, a large fraction of papers would be retracted on this basis.

## Important conclusions of the paper stand

One of the major conclusions (spanning the final three paragraphs of H2014) was that epistatic interactions are likely to be few and contribute little to the variance of gene expression, relative to additive effects. This result and interpretation is unaffected, and is the main reason for the paper continuing to be cited. While the presentation of the paper would have been very different had knowledge of the problem with the standard method come to light earlier, we still maintain that given the scale of the analysis, the importance of the conclusion that the instances of epistasis are few compared to additive effects remains. Negative results pertaining to questions about genetic architecture have been seen in Nature previously (e.g. <https://www.nature.com/articles/nature12170>).

We do think it’s important to emphasise that while there is a question mark over the originally reported signals, the important scientific conclusion on the small overall contribution of pairwise interactions to phenotypic variation in gene expression remains valid.

## Precedence

We have surveyed all Nature retractions that have previously been issued, and they only arise due to fraud, errors in calculation, or failure to follow standard protocols correctly (<https://www.nature.com/nature/articles?type=retraction>). We also surveyed Retraction Watch (<https://retractionwatch.com/>) and found no examples of retractions where new understanding of standard statistical methods led to retraction of older papers.

It seems unjust to place H2014 in the same frame of reference as current retractions.

## Consistency

There are ample examples of problematic papers, and indeed entire fields of problematic papers, for which retraction has not been the answer, and organic scientific progress has. A decade of failure to replicate genetic linkage signals for complex traits did not lead to retractions. Another decade of failure to replicate candidate gene studies did not lead to retractions. Our own review of the epistasis literature (<https://pubmed.gov/25200660>), which suggested that no previously published examples were reliable, did not lead to any retractions.

We will not enumerate a long list of problematic studies, but it does seem that the only reason that retraction is being requested for H2014 is that we voluntarily tried to ‘self-correct’ – by elaborating on a problem that was already observed in the Wood et al 2014 correspondence. This could be quite a dangerous message that deters future self-correction, or it implies quite a substantive change in the definition and culture of retraction.

## Nothing has changed since 2014

The biological plausibility of the interactions that we reported in H2014 was weakened following the correspondence with Wood et al (2014). We are not attempting to mask this, in fact we corroborate the findings in the correspondence with Wood et al in 2014, and we tried to publicly and loudly explain why this attenuation is occurring in our 2015 correspondence (first as an Addendum, which became a BCA). That BCA, and this new manuscript, does not change that situation. At best, there is some indication that the set of interactions we reported are a mixture of false and true positives (we observed higher replication rates than expected under null test statistic inflation, detailed in this new manuscript). But our original BCA and this new manuscript simply explains why there can be test statistic inflation with the standard method.

We remain open to guidance on the matter and hope to have the issue resolved quickly.

Yours sincerely,

Gibran Hemani, Konstantin Shakhbazov, Harm-Jan Westra, Tonu Esko, Anjali K. Henders, Allan F. McRae, Jian Yang, Greg Gibson, Nicholas G. Martin, Andres Metspalu, Lude Franke, Grant W. Montgomery, Peter M. Visscher & Joseph E. Powell