Addendum

We reported the detection and replication of epistatic interactions for gene expression (Hemani et al. 2014). Subsequently, in a Brief Communications Arising (BCA), Wood et al. (2014) proposed an alternative explanation for the reported interactions, namely the existence of an imperfectly tagged causal variant in *cis* which is in linkage disequilibrium (LD) with both genotyped SNPs. Using an independent data set with gene expression and genome sequences, Wood et al. 2014 reported a substantial reduction of signal for the interaction term by adjusting the expression trait by the top associated genetic variant from the sequence data. We acknowledged that such “haplotype effects” could explain a number of the reported *cis-cis* interactions but it was not clear whether this alternative explanation applied to all reported interactions, in particular the *cis-trans* interactions (Hemani et al. BCA reply). For the alternative explanation involving *cis-cis* interactions, both SNPs in *cis* are correlated with an unknown causal variant, whereas for *cis*-trans interactions, the SNP in trans is not correlated with either the SNP in *cis* or the unknown causal variant.

We have now performed additional empirical analyses on exactly the same data used in the original publication (Hemani et al. 2014) to attempt to resolve the alternative explanations for the results given by Hemani et al. (2014) and Wood et al. (2014 BCA). We used additional genome-wide association testing, simulation studies and empirical derivations to study the behaviour of the test statistic for epistasis when epistatic interactions are not present. These investigations have revealed that the F-test we used to detect epistasis is not robust to large additive effects of causal variants in *cis* that are not perfectly tagged by genotyped SNPs. A large imperfectly tagged additive effect in *cis* violates the assumed null distribution of no epistatic interaction because residuals are not normally distributed (they are a mixture of binomial and normal errors). This violation of assumption can create widespread genome-wide inflation of the test statistic, both for *cis-cis* and *cis-trans* effects, which can cause an inflated false-positive rate if the theoretical F-distribution is used to calculate a p-value (as was done in Hemani 2014).

In Figure 1 below we report the genome-wide inflation factor (the median of equivalent *χ*21 divided by 0.455) of the F-test for epistasis for the 501 SNPs pairs that passed our significance filters (Hemani 2014). The inflation factor was calculated from a genome-wide analysis that held the SNP with the strongest additive effect in *cis* constant and then performing an interaction test with SNPs on other chromosomes. The genome-wide analysis excluded SNPs in a 5 Mb region around the 2nd SNPs of the reported pair and SNPs that were on the same chromosome as the 1st SNP. In Table 1 we report the average inflation of the test statistic per probe, for probes that had at least 5 SNP pairs that contributed to the 501 selected pairs. Figure 1 and Table 1 show that for many probe *cis*-SNP pairs, in particular for the probes for which we reported *cis-trans* effects in Hemani et al. (2014), there is a large inflation of the test statistic. Note that since we removed 5Mb regions around the 2nd SNPs the actual inflation could be even (slightly) larger than those shown in Figure 1 and Table 1. The two probes for which replication was reported in Table 1 of Hemani (2014), TMEM149 and MBLL1, have median genome-wide test statistics that are, respectively, more than 5 and 3 times the value under the null hypothesis of no epistasis.

The consequence of these findings is that in the presence of imperfectly tagged eQTLs with large additive effects, the false positive rate from the test statistic for epistasis we employed is larger than assumed. A larger false positive rate is consistent with finding apparent replication (since the same applies to the replication data) and with having a much lower replication rate than expected given the stringency in the discovery sample. It is also the likely explanation why the statistical evidence for *cis-trans* effects is reduced after an adjustment for the additive effect of the best-tagging sequence variant in the *cis* region (Wood et al. 2014).

Whilst we cannot resolve with certainty alternative explanations for all reported epistatic interactions (either the 501 pairs discovered or the 30 pairs replicated) with these additional analyses, they show that in the presence of imperfectly tagged large *cis*-eQTLs, tests for epistasis such as that employed by us can result in inflated false positive rate. This conclusion is general and alternative explanations may also apply to other studies that have reported pairwise interactions in the presence of large additive effects using SNP-array based genotype data in particular. Unfortunately, there is, to our knowledge, no existing or proposed solution to this problem. The problem could be attenuated using whole genome sequencing data or very well imputed data. Creating an empirical null distribution by correction for genome-wide inflation for any pair of SNPs on different chromosomes in a full genome-wide scan for epistasis is at present computationally too demanding, and would not resolve *cis-cis* interactions.

Figure 1: Distribution of the ratio of the median F-statistic for epistasis and its value under the null hypothesis from a genome-wide empirical pairwise epistasis analysis that held the cis-SNP constant, for each of the 501 SNP pairs that passed the significance filters in Hemani et al. (2014).

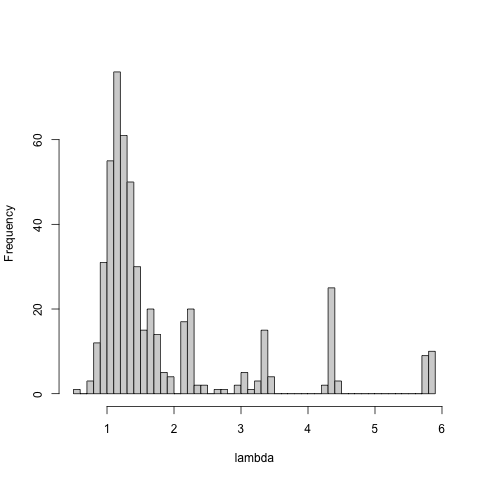


Table 1: Probes with multiple cis-trans interactions have genome-wide inflation of the test statistic for epistasis.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Probe Name | SNP 1 | SNP 2 | Gene | Mean Lambda | N pairs |
| ILMN\_1704730 | rs1884655 | rs10255470 | CD93 | 2.88 | 10 |
| ILMN\_1710752 | rs2123758 | rs2786014 | NAPRT1 | 2.15 | 8 |
| ILMN\_1717234 | rs1157079 | rs7733671 | CAST | 4.31 | 17 |
| ILMN\_1720059 | rs12435486 | rs7837237 | HMBOX1 | 2.29 | 7 |
| ILMN\_1738784 | rs10930170 | rs12120009 | PPP2R5A | 2.24 | 6 |
| ILMN\_1755589 | rs11080134 | rs11169322 | DIP2B | 1.16 | 6 |
| ILMN\_1786426 | rs2839013 | rs8106959 | TMEM149 | 5.65 | 20 |
| ILMN\_1804396 | rs1293455 | rs2655991 | C14ORF4 | 1.38 | 7 |
| ILMN\_2313158 | rs10869600 | rs13069559 | MBNL1 | 3.15 | 15 |
| ILMN\_2372639 | rs17159840 | rs10059004 | TRAPPC5 | 4.17 | 17 |
| ILMN\_3231952 | rs12947580 | rs8079215 | ARL17B | 2.16 | 6 |