BIOINFORMATICS CORRIGENDUM Vol. 29 no. 20 2013, pages 2667–2668
doi:10.1093/bioinformatics/btt444

Genome analysis

Corrigendum of 'High throughput analysis of epistasis in genome-wide association studies with BiForce'

Attila Gyenesei[1], Colin A.M. Semple[2], Chris S. Haley[2] and Wen-Hua Wei[2,*]

[1]Finnish Microarray and Sequencing Centre, Turku Centre for Biotechnology, University of Turku and Åbo Akademi University, 20520, Turku, Finland and [2]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine at the University of Edinburgh, Western General Hospital, Edinburgh, EH4 2XU, UK

Associate Editor: Alfonso Valencia

Contact: Wenhua.Wei@igmm.ed.ac.uk

Received on April 12, 2013; revised on July 19, 2013; accepted on July 29, 2013

Following the publication of our article, describing the use of BiForce (http://bioinfo.utu.fi/biforcetoolbox) for the analysis of epistasis (Gyenesei *et al.*, 2012), we observed that inflated evidence for epistasis may arise under exceptional circumstances when analyzing quantitative traits. This may occur when two neighboring (e.g. <200 kb apart) single nucleotide polymorphisms (SNPs) in an epistatic pair are in linkage disequilibrium (LD) and at least one of them carries strong marginal effects. Similar inflation was discovered recently in other LD- or haplotype-based methods for the analysis of epistasis in disease traits (Ueki and Cordell, 2012). This issue does not affect the analysis of disease traits in our case because BiForce uses logistic regression as the final step to generate the results for such traits (Wan *et al.*, 2010). Thus, Table 3 of the original paper (Gyenesei *et al.*, 2012) is correct. However for quantitative traits, BiForce uses contingency table-based F ratio tests for interactions without the fitting step applied in linear regression. It is known that such tests are not orthogonal, but they are robust when LD between two SNPs is low, allowing the fast screening achieved by BiForce. When LD is high, however, the test for interaction between two correlated SNPs is inflated by the marginal effects of the pair of SNPs, and therefore the inflation is critical when marginal effects are strong but not when marginal effects are weak. Nevertheless, because BiForce uses stringent Bonferroni-adjusted thresholds by default, the chance of inflated epistatic pairs being genome-wide significant should be low in general.

This issue affected the results in Table 2 of the original article (Gyenesei *et al.*, 2012). The correct interaction *P*-values (P_{int}) of each SNP pair are listed in the updated Table 2 later in the text, suggesting that none remained genome-wide significant in C-reactive protein (CRP), glucose (GLU), high-density lipoprotein (HDL), low-density lipoprotein (LDL) and triglycerides (TRI). Results from the analyses of simulated data on quantitative and disease traits are unaffected because in simulation SNPs were randomly drawn from a chromosome assuming they were in Hardy–Weinberg equilibrium, i.e. the chance of high LD coming together with strong marginal effects is very low, which is supported by the results of false positive rate in Figure 3 of the original article. In summary, the issue only affects a small part of

Table 2. Previous genome-wide significant epistatic pairs identified from the NFBC199 cohort[a] (update)

Trait	SNP_1	SNP_2	P_{int}	distance	LD (r^2)	correct P_{int}
CRP	rs1811472[b] (1q23.2; 0.41)	rs2592887[b] (1q23.2; 0.40)	3.0E-12	10 590	0.86	2.1E-01
CRP	rs1811472[b] (1q23.2; 0.41)	rs2794520[b] (1q23.2; 0.36)	3.5E-11	36 467	0.62	1.1E-01
CRP	rs2592887[b] (1q23.2; 0.40)	rs2794520[b] (1q23.2; 0.36)	2.9E-12	25 877	0.70	1.6E-01
CRP	rs2650000[b] (12q24.31; 0.45)	rs7953249[b] (12q24.31; 0.48)	2.6E-09	14 762	0.76	9.1E-01
CRP	rs1169300[b] (12q24.31; 0.32)	rs2464196[b] (12q24.31; 0.32)	3.4E-10	4202	0.99	4.1E-01
GLU	rs560887[b] (2q31.1; 0.30)	rs563694[b] (2q31.1; 0.34)	1.3E-08	10 923	0.81	5.2E-01
HDL	rs3764261[b] (16q13; 0.28)	rs1532624[b] (16q13; 0.41)	2.0E-14	12 155	0.53	6.8E-01
LDL	rs157580[b] (19q13.32; 0.29)	rs405509 (19q13.32; 0.46)	6.9E-10	13 570	0.35	1.6E-04
TRI	rs1260326[b] (2p23.3; 0.36)	rs780094 (2p23.3; 0.36)	5.8E-08	10 297	0.95	9.6E-01

[a]All SNP pairs listed detected as marginal-SNP interactions, with the threshold of 1.5E-08 for CRP, 2.2E-08 for HDL, 3.9E-08 for GLU and LDL, 7.7E-07 for TRI; SNP_1 (SNP_2) – name, genomic location and minor allele frequency (the latter two in bracket) of the first (second) SNP; P_{int} – *P*-value of the interaction test; distance – the distance in base pairs between two SNPs; LD – linkage disequilibrium (in r^2) between a pair of SNPs; the SNP pair in HDL was also detected via the pair-wise genome scan ($P < 9.54E-13$); *correct* P_{int} – the corrected *P*-value of the interaction test. [b]The marginal-SNP.

*To whom correspondence should be addressed.

the results (i.e. Table 2) concerning analyses of quantitative traits in real data. The main biological results and overall conclusions are unaffected. A script to address this issue is available from the dedicated website and will be incorporated in the next BiForce release.

Conflict of Interest: none declared.

REFERENCES

Gyenesei,A. *et al.* (2012) High throughput analysis of epistasis in genome-wide association studies with BiForce. *Bioinformatics*, **28**, 1957–1964.

Ueki,M. and Cordell,H.J. (2012) Improved statistics for genome-wide interaction analysis. *PLoS Genet*, **8**, e1002625.

Wan,X. *et al.* (2010) BOOST: a fast approach to detecting gene-gene interactions in genome-wide case-control studies. *Am. J. Hum. Genet.*, **87**, 325–340.