*Viewpoint*

# Commentary: lessons from molecular genetic studies on reporting false-positive results

*Grant W. Montgomery* [iD]

Institute for Molecular Bioscience, The University of Queensland, 306 Carmody Road, St Lucia,
Qld 4072, Australia. Email: g.montgomery1@uq.edu.au

**Abstract.** Poor replication of published research results is the subject of debate. A common problem is the failure to adequately account for multiple testing issues. In this regard, the evolution of mapping studies to identify genetic risk factors for common diseases has been instructive. Large genome-wide association studies (GWAS) reliably detect the genetic factors with small effects that contribute to risk for many common diseases. GWAS superseded candidate gene studies from the previous decade and looking back, almost no genetic risk factors reported from earlier candidate gene studies replicate in the GWAS results. Candidate gene studies often used small samples and failed to appreciate and adequately account for the multiple testing issues. The failure to replicate results from most candidate gene studies highlights the importance of study power and appropriate statistical analysis to prevent publication of false-positive results.

**Keywords:** gene expression, genetics, genotyping, reproduction.

## Introduction

Poor replication of published research results is a common problem. These issues were highlighted in the publication 'Why Most Published Research Findings Are False' by John Ioannidis (2005), leading to considerable debate. Major contributors to the problem included pressure to publish, inadequate study design and bias towards publishing positive results. Small sample sizes, inappropriate data analysis and poor statistical inference also contributed.

Failure to adequately control for multiple testing is a common problem of statistical inference leading to publication of false-positive results. In this regard, the evolution of mapping studies to identify genetic risk factors for common diseases has been instructive. In the past 10 years, genomic locations for thousands of genetic risk factors for common diseases have been mapped through genome-wide association studies (GWAS; Visscher *et al.* 2017). Empirical results for well-replicated genetic risk factors identified by GWAS demonstrate that effect sizes are small and the significant genetic variants are mostly located within introns and intergenic regions, not in coding regions of genes. Large GWAS studies reliably detect these small effects and the results replicate well across studies because, at the outset, the field set a genome-wide threshold for significance ($P < 5 \times 10^{-8}$) to account for testing association with variants from across the genome (The International HapMap Consortium 2005; Pe'er *et al.* 2008; Fadista *et al.* 2016).

GWAS superseded candidate gene studies from the previous decade. How well do those candidate gene studies replicate in GWAS results? Not well: across multiple diseases and with only a handful of exceptions, GWAS results have shown that published candidate gene associations were mostly false-positive results (Montgomery *et al.* 2008; Duncan *et al.* 2019). The candidate gene studies generally used small samples and were underpowered for the detection of the small size effects now reliably detected in the much larger samples used for GWAS. One reason for the failure is that common variants with the large genetic effects that may have been reliably detected with the small samples used in candidate gene studies are rare for complex diseases.

In addition, there was a failure to appreciate and adequately account for the multiple testing issues. Choosing known variants in coding regions of biologically plausible genes turned out not to increase the chances of finding true results, suggesting we knew less about the biology than we thought (Chabris *et al.* 2012; Rahmioglu *et al.* 2012; Duncan *et al.* 2019). Genetic variants in coding regions of candidate genes were no more likely to be associated with risk for common complex diseases than variants elsewhere in the genome. Consequently, the appropriate distribution to apply in multiple testing correction should have been the genome-wide level of significance used as the standard for reporting results from GWAS. Worse still, many studies failed to correct within their own studies for testing more than one variant or testing multiple genetic models, and results were declared significant at nominal values of $P < 0.05$. Thresholds of $P < 0.05$ or $P < 5 \times 10^{-8}$ give very different estimates of the number of 'significant' hits. For example, a GWAS meta-analysis on endometriosis with 17 045 endometriosis cases and 191 596 controls identified 14 independent risk variants at a genome-wide level of significance (Sapkota *et al.* 2017). In contrast, 24 431 variants passed the nominal significance threshold of $P < 0.05$.

Underpowered candidate gene association studies continue to be published. The clue these results are false positives is that the 'significant' results from these small studies generally report effect sizes well outside the range reported for replicated genome-wide significant results from GWAS (Duncan *et al.* 2019). The lesson is that candidate gene association studies should use genome-wide levels of significance to correct for multiple testing, even when only a small number of variants is genotyped. This would prevent continued publication of false-positive association results. Results should be replicated and, in many cases, this could be easily achieved by looking up the latest relevant GWAS results. Publication of results from small, underpowered candidate gene studies without appropriate statistical analysis and independent replication is no longer justified.

Corrections for multiple testing should also be considered for gene expression studies. The appropriate threshold for microarray or RNA sequence analysis is the subject of some debate. In our studies, we report significance for group comparisons only after multiple testing correction (Fung *et al.* 2018; Mortlock *et al.* 2020). Determining an appropriate *P*-value threshold for statistical significance is critical to differentiate true positives from false positives and false negatives. A key question in expression studies is what number of tests to use for multiple testing correction? In line with experience from genetic studies, we prefer to correct for all genes expressed in the relevant tissue.

There may be a case for choosing a different threshold for some functional studies testing strong hypotheses with good prior evidence. However, we should be wary of relaxed thresholds because many differentially expressed genes from small studies reported to be related to disease pathogenesis do not replicate well in studies with more adequate power. One possible explanation for some false-positive results is the large genetic effects on the expression of many genes (Westra *et al.* 2013; GTEx Consortium 2017). Unequal genotype frequencies between small samples may explain some differences otherwise attributed to treatment comparisons. Large and better-powered gene expression studies including meta-analyses of results from comparable studies at different centres are likely to provide more reliable estimates of disease-related changes in gene expression, as we have seen for well-powered multicentre GWAS results.

We need to understand the limitations of the methods we use. Too many papers in the reproduction literature report on underpowered studies with misleading conclusions based on poor statistical inference. There remains considerable bias favouring publication of 'significant' results over negative results. None of this is helped by the rise in predatory journals with poor editorial standards, but predatory journals are not the only problem. Papers with invalid conclusions based on inadequate study design and poor statistical inference are still published in the top journals in our field. The net result is misinformation or fake news that can easily become dogma, where articles are cited based on their conclusions without scrutiny of the supporting data. Genetic studies on disease risk demonstrate the importance of large studies with adequate power (often requiring collaboration of multiple research groups), combined with rigorous statistical analysis and interpretation. As scientists, reviewers and journal editors, we should reflect on the widespread failure to replicate many published results and insist on higher standards if we want continued support and funding for the vital role research can play in better health care and a productive economy.

## Conflicts of interest

The author declares no conflicts of interest.

## Acknowledgement

## References

Chabris, C. F., Hebert, B. M., Benjamin, D. J., Beauchamp, J., Cesarini, D., van der Loos, M., Johannesson, M., Magnusson, P. K., Lichtenstein, P., Atwood, C. S., Freese, J., Hauser, T. S., Hauser, R. M., Christakis, N., and Laibson, D. (2012). Most reported genetic associations with general intelligence are probably false positives. *Psychol. Sci.* **23**, 1314–1323. doi:10.1177/0956797611435528

Duncan, L. E., Ostacher, M., and Ballon, J. (2019). How genome-wide association studies (GWAS) made traditional candidate gene studies obsolete. *Neuropsychopharmacology* **44**, 1518–1523. doi:10.1038/S41386-019-0389-5

Fadista, J., Manning, A. K., Florez, J. C., and Groop, L. (2016). The (in)famous GWAS *P*-value threshold revisited and updated for low-frequency variants. *Eur. J. Hum. Genet.* **24**, 1202–1205. doi:10.1038/EJHG.2015.269

Fung, J. N., Mortlock, S., Girling, J. E., Holdsworth-Carson, S. J., Teh, W. T., Zhu, Z., Lukowski, S. W., McKinnon, B. D., McRae, A., Yang, J., Healey, M., Powell, J. E., Rogers, P. A. W., and Montgomery, G. W. (2018). Genetic regulation of disease risk and endometrial gene expression highlights potential target genes for endometriosis and polycystic ovarian syndrome. *Sci. Rep.* **8**, 11424. doi:10.1038/S41598-018-29462-Y

GTEx Consortium (2017). Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213. doi:10.1038/NATURE24277

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Med.* **2**, e124. doi:10.1371/JOURNAL.PMED.0020124

Montgomery, G. W., Nyholt, D. R., Zhao, Z. Z., Treloar, S. A., Painter, J. N., Missmer, S. A., Kennedy, S. H., and Zondervan, K. T. (2008). The search for genes contributing to endometriosis risk. *Hum. Reprod. Update* **14**, 447–457. doi:10.1093/HUMUPD/DMN016

Mortlock, S., Kendarsari, R. I., Fung, J. N., Gibson, G., Yang, F., Restuadi, R., Girling, J. E., Holdsworth-Carson, S. J., Teh, W. T., Lukowski, S. W., Healey, M., Qi, T., Rogers, P. A. W., Yang, J., McKinnon, B., and Montgomery, G. W. (2020). Tissue specific regulation of transcription in endometrium and association with disease. *Hum. Reprod.* **35**, 377–393. doi:10.1093/HUMREP/DEZ279

Pe'er, I., Yelensky, R., Altshuler, D., and Daly, M. J. (2008). Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.* **32**, 381–385. doi:10.1002/GEPI.20303

Rahmioglu, N., Missmer, S. A., Montgomery, G. W., and Zondervan, K. T. (2012). Insights into assessing the genetics of endometriosis. *Curr. Obstet. Gynecol. Rep.* **1**, 124–137. doi:10.1007/S13669-012-0016-5

Sapkota, Y., Steinthorsdottir, V., Morris, A. P., Fassbender, A., Rahmioglu, N., De Vivo, I., Buring, J. E., Zhang, F., Edwards, T. L., Jones, S., O, D., Peterse, D., Rexrode, K. M., Ridker, P. M., Schork, A. J., MacGregor, S., Martin, N. G., Becker, C. M., Adachi, S., Yoshihara, K., Enomoto, T., Takahashi, A., Kamatani, Y., Matsuda, K., Kubo, M., Thorleifsson, G.,

Geirsson, R. T., Thorsteinsdottir, U., Wallace, L. M., iPSYCH-SSI-Broad GroupYang, J., Velez Edwards, D. R., Nyegaard, M., Low, S. K., Zondervan, K. T., Missmer, S. A., D'Hooghe, T., Montgomery, G. W., Chasman, D. I., Stefansson, K., Tung, J. Y., and Nyholt, D. R. (2017). Meta-analysis identifies five novel loci associated with endometriosis highlighting key genes involved in hormone metabolism. *Nat. Commun.* **8**, 15539. doi:10.1038/NCOMMS15539

The International HapMap Consortium (2005). A haplotype map of the human genome. *Nature* **437**, 1299–1320. doi:10.1038/NATURE04226

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., and Yang, J. (2017). 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* **101**, 5–22. doi:10.1016/J.AJHG.2017.06.005

Westra, H. J., Peters, M. J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M. W., Fairfax, B. P., Schramm, K., Powell, J. E., Zhernakova, A., Zhernakova, D. V., Veldink, J. H., Van den Berg, L. H., Karjalainen, J., Withoff, S., Uitterlinden, A. G., Hofman, A., Rivadeneira, F., 't Hoen, P. A. C., Reinmaa, E., Fischer, K., Nelis, M., Milani, L., Melzer, D., Ferrucci, L., Singleton, A. B., Hernandez, D. G., Nalls, M. A., Homuth, G., Nauck, M., Radke, D., Volker, U., Perola, M., Salomaa, V., Brody, J., Suchy-Dicey, A., Gharib, S. A., Enquobahrie, D. A., Lumley, T., Montgomery, G. W., Makino, S., Prokisch, H., Herder, C., Roden, M., Grallert, H., Meitinger, T., Strauch, K., Li, Y., Jansen, R. C., Visscher, P. M., Knight, J. C., Psaty, B. M., Ripatti, S., Teumer, A., Frayling, T. M., Metspalu, A., van Meurs, J. B., and Franke, L. (2013). Systematic identification of *trans* eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243. doi:10.1038/NG.2756

Handling Editor: Graeme Martin