

Commentary

doi: 10.1111/j.1469-1809.2011.00640.x

Correcting Away the Hidden Heritability

Scott M. Williams* and Jonathan L. Haines

Center for Human Genetics Research, Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN

Summary

In the age of high-density genome-wide association (GWAS) data, correcting for multiple comparisons is a substantial issue for genetic epidemiological studies. However, the current manuscript review process generally requires both stringent correction and independent replication. The result of this stringency is that studies that are published suffer from inflated Type 2 error rates (false negatives), thereby removing many likely real signals from follow-up. Elimination of these alleles, if they are truly associated, from further study will slow research progress in studies of complex disease. We argue that this method of correction is overly conservative, especially in an age when high-density follow-up experiments are possible and reasonably inexpensive.

Keywords: GWAS, multiple testing, heritability, complex disease

Studies of the genetics of complex disease have begun to move from the realm of genotyping single nucleotide polymorphisms (SNPs) to that of exome or whole genome sequencing. This is at least partially motivated by the inability of GWAS to identify the majority of the heritable fraction of disease (Manolio et al., 2009). Typically, reviewers and editors require strict Bonferroni or a similar correction for multiple comparisons in the first stage of a GWAS based on results from simulation work (Dudbridge & Gusnanto, 2008). A further requirement is to follow such analyses by a second and/or third stage of independent replication. We argue that severe correction for multiple comparisons in the first of a two-stage analysis is overly conservative and serves to reduce the discoverable amount of the heritability that is in fact buried in the mass of GWAS data. In addition, this process does not serve the primary purpose for which this and other “hypothesis-free” approaches were originally intended, that is, the generation of new *biologically* testable hypotheses. Although this approach represents a strong and understandable desire by editors and reviewers to effectively eliminate false-positive results, it likely serves to ignore innumerable true signals. Determining the appropriate balance of false positive to false-negative results is not a new discussion. For decades, scientists have debated whether correction is necessary and if so, how it is best accomplished (Rothman, 1990; Boehringer

et al., 2000; Nyholt, 2001; Krawczak et al., 2004). However, the high density of data in GWAS has caused the concern about multiple comparison correction to be revisited with renewed vigor by the genetics community. Lost in this discussion are significant considerations about the amount of independence among marker genotypes and an explicitly stated preference on balancing Type 1 versus Type 2 errors (Dudbridge et al., 2000; Kimmel et al., 2008). While correction has now become a mantra, we argue it requires a more balanced approach that incorporates both biological context of putative signals in the broadest sense and a better understanding of the role that statistical analyses can play in the detection of genes/variants that affect complex disease.

We caution against the groundswell that mandates indiscriminant use of Bonferroni or comparable correction. Our position rests less on standard statistical arguments, but rather on technological advances that enable adequate and independent replication not just of a few “top hits” but also of all the thousands of nominal associations. The *P* value as a metric is meant to assess the probability to observe an event as extreme or even more extreme than the one observed. Certainly, in GWAS most nominal findings are due to chance alone or to technical artifacts, despite the extensive quality control procedures commonly used. The intent of correction, then, is to minimize those associations that are chance events – that is, to minimize the number of false-positive results. However, it should still be noted that even extremely small *P* values can be consistent with the null hypothesis, and therefore this metric alone may not serve as the best arbiter of true association or more importantly real biological meaning (Zaykin & Zhivotovsky, 2005). We argue that, when examining a large

*Corresponding author: Scott M. Williams, Center for Human Genetics Research, Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN 37232. Tel: 615-322-8036; Fax: 615-343-8619; E-mail: scott.williams@chgr.mc.vanderbilt.edu

array of nominally positive findings, statistical stringency alone does not permit us to determine which findings are by chance and which are not, and therefore, setting too stringent cutoff for Type 1 error criterion for association decreases power to find real associations. Clearly, the preferred conservative solution among editors and the majority of statistical geneticists is that it is better to remove those findings more likely due to chance events from further consideration by applying conservative rules. This approach results in having few reports that can be readily proven wrong due to Type 1 error, but also slows the progress of research by inflating the Type 2 error rate.

An alternative view is that by conservatively correcting for false positives, we are in fact eliminating the vast majority of truly associating SNPs, the very thing we wish to find. Almost all high-impact journals require, in addition to stringent correction, that independent replication is demonstrated in a second or third dataset. By requiring both a very small P value and an independent replication, we agree that the chances of reporting genes that are truly associated with disease are greatly increased. However, we adopt and endorse the view that the “gold” standard of marginal effects is in fact replication (Chanock et al., 2007). Thus, if one requires replication as a criterion for publication, then we question the duplicative need for severe correction for multiple testing in the discovery dataset.

We argue that we are missing many more genes showing genuine associations that do not pass this stringent set of criteria, but are nonetheless real. So, if we reassert the initial intention of GWAS as part of a discovery phase of investigation, these current requirements only allow us to find the low hanging, statistically discovered fruit while ignoring more subtle, yet very real and perhaps even common, biological effects. This rigorous process could very easily explain some and perhaps much of the missing heritability (Manolio et al., 2009) that is currently being chased with alternative technologies and underlying hypotheses (e.g., multiple rare variants/common disease).

Previously, the problem with using replication as the true gold standard was a lack of appropriate technology to assay all putatively true associations. Current technologies now allow all SNPs at $P < 0.05$ to be assayed relatively cheaply on custom chips. Two recent studies exemplify how the current correction paradigm can actually lead to false-negative results that account for some of the missing heritability. Easton et al. (2007) performed a three-stage analysis of breast cancer. Rather than focusing on only genome-wide significant SNPs, they tested the most significant 5% of SNPs from the discovery dataset in a second confirmatory dataset, and from that 30 SNPs were tested in a third dataset. Of particular note, the majority of the final SNPs that replicated did not meet genome-wide significance levels in stage 1 or even stage 2.

McCauley et al. (2010) performed a comprehensive follow-up of all available SNPs with $P < 0.10$ in an initial GWAS of multiple sclerosis (MS) (Hafler et al., 2007). They genotyped ~30,000 SNPs, using a cost-effective custom chip array. None of the SNPs with $P < 0.0001$ in the initial discovery dataset replicated in follow-up studies. In contrast, multiple new confirmed associations were identified, but many with SNPs (and/or genes) that ranked below the top 0.5% of SNPs. The majority of these associations had initial GWAS P values only slightly less than a nominal P of 0.05.

GWAS has been a successful tool for confirming existing and identifying many new associations of common variation to common disease. However, the formulaic and rigid application of severe statistical correction has in fact muted this success, preventing many, and perhaps most, of the true associations from being reported and confirmed. In addition, the underlying assumptions of the standard approach as discussed in this paper, specifically of single gene, marginal effects, may violate many biologically real phenomena such as gene–gene and gene–environment interactions that can fail in a replication phase even if they are real. This raises other important considerations that we have not directly addressed but certainly warrant attention in the genetics community (Greene et al., 2009). As the MS (McCauley et al., 2010) and breast cancer (Easton et al., 2007) data clearly show, a large number of true associations lurk in the midst of nominally significant P values, and the only way to illuminate them is to return to the initial intent of GWAS as a discovery and filtering tool for future studies.

References

- Boehringer, S., Epplen, J. T. & Krawczak, M. (2000) Genetic association studies of bronchial asthma – a need for Bonferroni correction? *Hum Genet* **107**, 197.
- Chanock, S. J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D. J., Thomas, G., Hirschhorn, J. N., Abecasis, G., Altshuler, D., Bailey-Wilson, J. E., Brooks, L. D., Cardon, L. R., Daly, M., Donnelly, P., Fraumeni, J. F., Jr., Freimer, N. B., Gerhard, D. S., Gunter, C., Guttmacher, A. E., Guyer, M. S., Harris, E. L., Hoh, J., Hoover, R., Kong, C. A., Merikangas, K. R., Morton, C. C., Palmer, L. J., Phimister, E. G., Rice, J. P., Roberts, J., Rotimi, C., Tucker, M. A., Vogan, K. J., Wacholder, S., Wijisman, E. M., Winn, D. M. & Collins, F. S. (2007) Replicating genotype-phenotype associations. *Nature* **447**, 655–660.
- Dudbridge, F. & Gusnanto, A. (2008) Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol* **32**, 227–234.
- Dudbridge, F., Koeleman, B. P., Todd, J. A. & Clayton, D. G. (2000) Unbiased application of the transmission/disequilibrium test to multilocus haplotypes. *Am J Hum Genet* **66**, 2009–2012.
- Easton, D. F., Pooley, K. A., Dunning, A. M., Pharoah, P. D., Thompson, D., Ballinger, D. G., Struwing, J. P., Morrison, J., Field, H., Luben, R., Wareham, N., Ahmed, S., Healey, C. S., Bowman, R., Meyer, K. B., Haiman, C. A., Kolonel, L. K., Henderson, B. E., Le, M. L., Brennan, P., Sangrajrang, S., Gaborieau,

- V., Odehrey, F., Shen, C. Y., Wu, P. E., Wang, H. C., Eccles, D., Evans, D. G., Peto, J., Fletcher, O., Johnson, N., Seal, S., Stratton, M. R., Rahman, N., Chenevix-Trench, G., Bojesen, S. E., Nordestgaard, B. G., Axelsson, C. K., Garcia-Closas, M., Brinton, L., Chanock, S., Lissowska, J., Peplonska, B., Nevanlinna, H., Fagerholm, R., Eerola, H., Kang, D., Yoo, K. Y., Noh, D. Y., Ahn, S. H., Hunter, D. J., Hankinson, S. E., Cox, D. G., Hall, P., Wedren, S., Liu, J., Low, Y. L., Bogdanova, N., Schurmann, P., Dork, T., Tollenaar, R. A., Jacobi, C. E., Devilee, P., Klijn, J. G., Sigurdson, A. J., Doody, M. M., Alexander, B. H., Zhang, J., Cox, A., Brock, I. W., MacPherson, G., Reed, M. W., Couch, F. J., Goode, E. L., Olson, J. E., Meijers-Heijboer, H., Van Den Ouweland A., Uitterlinden, A., Rivadeneira, F., Milne, R. L., Ribas, G., Gonzalez-Neira, A., Benitez, J., Hopper, J. L., McCredie, M., Southey, M., Giles, G. G., Schroen, C., Justenhoven, C., Brauch, H., Hamann, U., Ko, Y. D., Spurdle, A. B., Beesley, J., Chen, X., Mannermaa, A., Kosma, V. M., Kataja, V., Hartikainen, J., Day, N. E., Cox, D. R. & Ponder, B. A. (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087–1093.
- Greene, C. S., Penrod, N. M., Williams, S. M. & Moore, J. H. (2009) Failure to replicate a genetic association may provide important clues about genetic architecture. *PLoS One* **4**, e5639.
- Hafler, D. A., Compston, A., Sawcer, S., Lander, E. S., Daly, M. J., De Jager, P. L., de Bakker, P. I., Gabriel, S. B., Mirel, D. B., Ivinson, A. J., Pericak-Vance, M. A., Gregory, S. G., Rioux, J. D., McCauley, J. L., Haines, J. L., Barcellos, L. F., Cree, B., Oksenberg, J. R. & Hauser, S. L. (2007) Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med* **357**, 851–862.
- Kimmel, G., Karp, R. M., Jordan, M. I. & Halperin, E. (2008) Association mapping and significance estimation via the coalescent. *Am J Hum Genet* **83**, 675–683.
- Krawczak, M., Boehringer, S. & Epplen, J. T. (2004) Correcting for multiple testing in genetic association studies: the legend lives on. *Hum Genet* **109**, 566–567.
- Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorf, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., Boehnke, M., Clark, A. G., Eichler, E. E., Gibson, G., Haines, J. L., Mackay, T. F., McCarroll, S. A. & Visscher, P. M. (2009) Finding the missing heritability of complex diseases. *Nature* **461**, 747–753.
- McCauley, J. L., Zuvich, R. L., Beecham, A. H., DeJager, P. L., Konidari, I., Whitehead, P. L., Aubin, C., Ban, M., Pobywajlo, S., 't Hooft, F. M., Briskin, R., Romano, S., Aggarwal, N. T., Piccio, L., McArdle, W. L., Strachan, D. P., Evans, D., Cross, A. H., Rioux, J. D., Barcellos, L. F., Ivinson, A. J., Compston, A., Hafler, D. A., Hauser, S. L., Oksenberg, J. R., Sawcer, S., Pericak-Vance, M. A. & Haines, J. L. (2010) Comprehensive follow-up of the first genome-wide association study of multiple sclerosis identifies KIF21B and TMEM39A as susceptibility loci. *Hum Mol Genet* **19**, 953–962.
- Nyholt, D. R. (2001) Genetic case-control association studies—correcting for multiple testing. *Hum Genet* **109**, 564–567.
- Rothman, K. J. (1990) No adjustments are needed for multiple comparisons. *Epidemiology* **1**, 43–6.
- Zaykin, D. V. & Zhivotovsky, L. A. (2005) Ranks of genuine associations in whole-genome scans. *Genetics* **171**, 813–823.

Received: 26 October 2010

Accepted: 23 December 2010