Reply to : On powerful GWAS in admixed populations

Atkinson, Elizabeth G.

2021-12

Atkinson , E G , Bloemendal , A , Maihofer , A X , Nievergelt , C M , Daly , M J & Neale , B M 2021 , ' Reply to : On powerful GWAS in admixed populations ' , Nature Genetics , vol. 53 , þÿ n o . 1 2 , p p . 1 6 3 4 1 6 3 5 . h t t p s : / / d o i . o r g / 1 0 . 1 0 3 8 / s 4 1 5 8 8 - 0 2 1 - 0 0 9 7 5 - z

http://hdl.handle.net/10138/353721
https://doi.org/10.1038/s41588-021-00975-z

Reply to Hou et al. comment; *Nature Genetics* 'Matters Arising'

Elizabeth G. Atkinson[1,2,*], Alex Bloemendal[1,2], Adam Maihofer[3], Caroline M. Nievergelt[3], Mark J. Daly[1,2,4], Benjamin M. Neale[1,2]

1 Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA

2 Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

3 Department of Psychiatry, University of California, San Diego, La Jolla, CA, USA

4 Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland

*Corresponding author and contact information:

Elizabeth G Atkinson, PhD

eatkinso@broadinstitute.org

(202) 246-0666

Analytic and Translational Genetics Unit

Richard B. Simches Building, 6th Floor

185 Cambridge Street

Boston, MA 02114

Admixed populations are sorely underrepresented in genomics research[1,2]. To ensure that medical genetic breakthroughs equitably benefit individuals of all ancestries[3], there is a need for the development of tools that facilitate the study of diverse and admixed populations. Our manuscript[4] proposes a novel methodology for the inclusion of admixed individuals in well-calibrated genome-wide association studies (GWAS) through the incorporation of local ancestry.

In their comment, Hou et al. argue that alternative GWAS methods that do not include local ancestry can attain improved power in circumstances where the effect sizes are equivalent across ancestries. We wish to clarify that while we indeed observe a power drop due to the increase in the number of parameters estimated in this edge case (addressed in the Tractor manuscript in Figure 2, Extended Data Figure 3-4 and *Discussion*), in all other scenarios modeled we observe a power boost. Given that minor allele frequencies and patterns of linkage disequilibrium regularly differ between populations genome-wide[5–8], as well as that differences in case ascertainment, epistasis, and gene-environment interactions may differ across ancestries and induce marginal effect size differences, we expect the instance of perfectly identical marginal effect sizes to be the exception, not the rule, even assuming identical causal effects. Tractor is therefore expected to outperform other methods at most GWAS loci.

By *marginal effect*, we mean the estimand (large-sample limit) of GWAS-style single variant regression (including control for stratification). By *causal effect,* we mean the effect of allelic substitution on an isogenic (and iso-environmental) background. The causal effects may of course be unidentifiable from observed data; if all variation were measured and indexed with no perfect LD, and there were no population stratification, then they would represent the estimand or large-sample limit of the full multivariate regression. The vector of marginal effects is related to the vector of causal effects by multiplication by the LD matrix.

The Tractor model parametrizes ancestry-specific *marginal* effects; hence the power loss highlighted in Hou et al.'s comment pertains only to the case where these are all equal, which requires not merely identical causal effects, but also identical patterns of AF and LD across ancestries. Imperfect tagging will induce heterogeneity by ancestry at the tagging variant even if the causal one has no allelic heterogeneity by ancestry, reducing the power of models that assume no heterogeneity such as the 1 degree of freedom tests suggested by Hou et al. The

vast majority of GWAS loci identified are not causal[9–11] but rather represent tagging variants and their estimated marginal effect sizes.

Tractor was specifically developed to function in diverse admixed datasets and has several further advantages beyond leveraging genomic differences across ancestries to gain power. Even when failing to boost power over competing methods, it produces accurate ancestry-specific effect size estimates, which can be vital for efforts utilizing GWAS summary statistics, such as the construction of polygenic scores for understudied populations. Tests considering admixed individuals' component ancestries in aggregate do not produce ancestry-specific results. Tractor additionally improves the resolution of GWAS signal, even before fine-mapping, thanks to its ability to track ancestry breakpoints within admixed genomes, helping streamline the interpretation of significant loci. By iterating between statistical phasing and local ancestry inference, we also improve the recovery of long-range haplotypes in admixed individuals. In sum, the Tractor model allows for the well-calibrated study of admixed individuals while boosting power under most scenarios.

Author Contributions

E.G.A. drafted the primary text with input from A.B., A.M., B.M.N., C.M.N. and M.J.D. All authors reviewed and approved the final draft.

Competing Interests statement

M.J.D. is a founder of Maze Therapeutics. B.M.N. is a member of the Deep Genomics Scientific Advisory Board and serves as a consultant for the Camp4 Therapeutics Corporation, Takeda Pharmaceutical and Biogen. The remaining authors declare no competing interests.

Data and Code Availability

All data referred to in this reply is available as described in the original Tractor publication[4].

References

1. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* vol. 538 161–164 (2016).
2. Sirugo, G., Williams, S. M. & Tishkoff, S. A. The Missing Diversity in Human Genetic Studies. *Cell* **177**, 1080 (2019).
3. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
4. Atkinson, E. G. *et al.* Tractor uses local ancestry to enable the inclusion of admixed individuals in GWAS and to boost power. *Nat. Genet.* **53**, 195–204 (2021).
5. Wang, Y. *et al.* Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations. *Nat. Commun.* **11**, 3865 (2020).
6. Bergström, A. *et al.* Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**, (2020).
7. 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
8. Skotte, L., Jørsboe, E., Korneliussen, T. S., Moltke, I. & Albrechtsen, A. Ancestry-specific association mapping in admixed populations. *Genet. Epidemiol.* **43**, 506–521 (2019).
9. Broekema, R. V., Bakker, O. B. & Jonkers, I. H. A practical view of fine-mapping and gene prioritization in the post-genome-wide association era. *Open Biol.* **10**, 190221 (2020).
10. MacArthur, D. G. *et al.* Guidelines for investigating causality of sequence variants in human disease. *Nature* **508**, 469–476 (2014).
11. Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* **19**, 491–504 (2018).