Reply to

nature genetics

Matters arising

https://doi.org/10.1038/s41588-022-01251-4

Reply to: Genotype by sex interactions in ankylosing spondylitis

Received: 14 February 2022

Accepted: 26 October 2022

Published online: 9 January 2023

Check for updates

Elena Bernabeu[1], Konrad Rawlik[2], Oriol Canela-Xandri[1], Andrea Talenti[2], James Prendergast[2] & Albert Tenesa[1,2,3]

ARISING FROM: Bernabeu et al. *Nature Genetics* https://doi.org/10.1038/s41588-021-00912-0 (2021)

Differences between males and females have been described for a broad range of human complex traits, yet their underlying mechanisms are still poorly understood. In this context, our article[1] provides an extensive and agnostic look into the role of the human genome, across 530 complex traits, making use of data originating from around half a million individuals from the UK Biobank. Our article constitutes an important first step, with further studies needed to, among others, assess causal associations and fine-map the genetic variants and regions reported. In their comment, Li et al.[2] have begun this important task, performing a more detailed analysis of the locus linked to ankylosing spondylitis and, more specifically, the roles of the human leukocyte antigen (HLA) region genes *HLA-B* and *MICA*, which have been subject to previous debate[3,4].

Li et al.[2] replicated our finding of a genotype by sex (G × S) interaction for ankylosing spondylitis at the HLA region, and performed a series of conditional logistic regression analyses, making use of the UK Biobank data pertaining to unrelated white individuals of British ancestry (n = 334,996). They found that a variant suggested to tag the *HLA-B27* allele (rs116666910), which was filtered by the allele frequency cutoff in our original analysis, presents the largest evidence of G × S in the region, as opposed to the lead single-nucleotide polymorphism (SNP) at the locus we reported (rs9266267, for which the closest gene is *HLA-B27*, not *MICA*). This observation is not untypical in fine-mapping efforts since when SNP density is increased a better supported marker may arise. We believe these fine-mapping efforts are incredibly valuable when dissecting genetic architecture heterogeneity between the sexes and strongly welcome further such work into potential drivers of the G × S interactions we observed. Linear and logistic models have technical limitations in the presence of large case-control imbalance[5–7] (the number of cases of ankylosing spondylitis in the UK Biobank, considering all white individuals of European ancestry, being 788 in males and 466 in females), which give rise to inflated type I error rates[5,8] (discussed further in the Methods and Supplementary Methods of our article). In linear mixed models, these issues are ameliorated by applying a more stringent minor allele frequency cutoff as we did[8].

After their conditional regression analyses, Li et al.[2] looked to replicate our finding of a sex-biased expression quantitative trait locus (eQTL) at the *MICA* locus. In our work, we searched for potential links between G × S interactions in genetic architecture and G × S interactions in gene expression regulation. We did this through an agnostic search of sex biases in genes near SNPs presenting G × S interactions in any of the 530 traits considered, not just ankylosing spondylitis. This analysis, which made use of the Genotype-Tissue Expression (GTEx) v.6p resource, returned several sex-biased eQTLs, including the eQTL rs56705452 for *MICA* in skeletal muscle tissue. This variant is not in linkage disequilibrium with the rs9266267 variant Li et al.[2] focused on in the first part of their report (r^2 = 0.07 in the 1000 Genomes Project[9]). Making use of a more recent GTEx version (v.8), which contains a larger number of individuals not available to us at the time our study was conducted, Li et al.[2] found the same, albeit less significant (P = 0.0013), rs56705452-*MICA* sex-biased eQTL in the same tissue (Extended Data Fig. 8 in the original article and Fig. 2 in the response by Li et al.). Furthermore, in their study Li et al. did not carry out an interaction analysis between *HLA-B27* tagging variants and sex for the expression of *HLA-B* in any tissue. Hence, while their fine-mapping work is consistent with the hypothesis that *HLA-B27* is linked to ankylosing spondylitis, their functional analysis falls short of that. To correct this, we tested for a G × S interaction in *HLA-B* gene expression across all tissues making use of the GTEx v.8 resource, considering the SNP with the lowest G × S P value in the logistic analysis by Li et al. (rs116666910; Fig. 1b in the response by Li et al.[2]), suggested to tag *HLA-B27*. We found a significant rs116666910 by sex interaction for *HLA-B* expression in skeletal muscle tissue (P = 0.008; Supplementary Table 1) but this is less significant than both the association we originally reported between rs56705452 and *MICA* and the updated result for the same SNP and gene reported by Li et al.[2]. We found no evidence of an interaction of rs116666910 and sex with expression levels in blood of *HLA-B27* (P = 0.95), the target tissue suggested by Li et al.[2]. This reinforces our call for larger sample sizes to truly understand the link between G × S interactions in genetic

[1]Institute of Genetics and Cancer, The University of Edinburgh, Edinburgh, UK. [2]The Roslin Institute, Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Edinburgh, UK. [3]Medical Research Council Human Genetics Unit, The University of Edinburgh, Edinburgh, UK. e-mail: elena.bernabeu@ed.ac.uk; albert.tenesa@ed.ac.uk

architecture and in gene expression regulation, in concordance with previous reports[10,11].

Finally, Li et al.[2] speculated that estimates of heritability may be biased by differences in disease prevalence recorded in the UK Biobank. They presented data suggesting there is a lower 'true' prevalence of the disease in the UK Biobank compared to data obtained from clinical settings. Li et al.[2] based this conclusion on imputed genotypes at the *HLA-B27* allele, the prevalence of which is lower among ankylosing spondylitis cases than in clinical settings. That is, they consider the *HLA-B27* genotype as the gold standard of phenotypic definition for ankylosing spondylitis, an approach we consider as having important limitations. First, there is a circularity element, since it assumes that *HLA-B27* status is the causal allele for ankylosing spondylitis. Second, it assumes that imputation and typing of *HLA-B27* have similar diagnostic accuracy. Even if we were to accept this as a reasonable definition of ankylosing spondylitis, Li et al.[2] presented no evidence that the purported difference in prevalence between the general UK population and the UK Biobank significantly affects the heritability estimates in males or females. We conducted a sensibility analysis with our data using the observed heritability estimated in the UK Biobank for males and females and transformed it to the liability scale using the prevalences reported by Li et al.[2] (Methods and Supplementary Table 2). We found that a difference in heritability between the sexes was still detected ($P = 2.8 \times 10^{-3}$). We further transformed the observed heritability estimates making use of the National Institute for Health and Care Excellence (NICE) prevalence of ankylosing spondylitis, considering both its lower (0.05%) and upper (0.23%) bounds. Significant differences in heritability on the liability scale between males and females were found for both ($P_{lower} = 4.24 \times 10^{-15}$, $P_{upper} = 4.02 \times 10^{-15}$), albeit the estimates were outside the parameter space.

To conclude, we accept that this is a complicated locus, a detailed study of which was outside the scope of our manuscript[1], but we disagree that the evidence presented by Li et al.[2] unequivocally demonstrates the exclusive involvement of *HLA-B27* in the causation of ankylosing spondylitis nor that, as they imply, we made the claim that *MICA*, but not *HLA-B27*, was the only causal gene.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-022-01251-4.

References

1. Bernabeu, E. et al. Sex differences in genetic architecture in the UK Biobank. *Nat. Genet.* **53**, 1283–1289 (2021).
2. Li, Z. et al. Genotype by sex interactions in ankylosing spondylitis. *Nat. Genet*. https://doi.org/10.1038/s41588-022-01250-5 (2023).
3. Zhou, X. & Reveille, J. D. Imputation-based analysis of *MICA* alleles in the susceptibility to ankylosing spondylitis. *Ann. Rheum. Dis.* **79**, e1 (2020).
4. Cortes, A. et al. Imputation-based analysis of *MICA* alleles in the susceptibility to ankylosing spondylitis. *Ann. Rheum. Dis.* **77**, 1691–1692 (2018).
5. Zhou, W. et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).
6. Mbatchou, J. et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.* **53**, 1097–1103 (2021).
7. Jiang, L., Zheng, Z., Fang, H. & Yang, J. A generalized linear mixed model association tool for biobank-scale data. *Nat. Genet.* **53**, 1616–1621 (2021).
8. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat. Genet.* **50**, 906–908 (2018).
9. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
10. Porcu, E. et al. Limited evidence for blood eQTLs in human sexual dimorphism. *Genome Med.* **14**, 89 (2022).
11. Oliva, M. et al. The impact of sex on gene expression across human tissues. *Science* **369**, eaba3066 (2020).

Methods

To assess the existence of an interaction between rs116666910, sex and expression levels of *HLA-B* (ENSG00000234745.10) across all non-sex-specific tissues in the v.8 release of GTEx, linear regression models were fitted using PLINK v.1.9 according to the GTEx Consortium eQTL discovery pipeline. In addition, a sensitivity analysis was carried out to assess the effect of different prevalence estimates of ankylosing spondylitis, including those reported by NICE and by Li et al.[2] on heritability on the liability scale. Methods are further described in the Supplementary Note.

Human research participants

The UK Biobank Project has favorable ethical opinion from the National Health Service North West – Haydock Research Ethics Committee (ref. 21/NW/0157, previously ref. 16/NW/0274).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data used pertain to the GTEx project v.8. Gene expression data are freely available at https://gtexportal.org/home/. GTEx genotype data are available on application through the database of Genotypes and Phenotypes. This research was conducted using the UK Biobank resource under project no. 788. The GTEx regulatory and consent protocols can be found at https://biospecimens.cancer.gov/resources/sops/library.asp.

Code availability

We used PLINK v.1.9 to run our eQTL analysis, which is freely available online at https://www.cog-genomics.org/plink2/. The custom Python code is openly available at *Zenodo* https://doi.org/10.5281/zenodo.7093777.

Author contributions

E.B. drafted the primary text with input from K.R., J.P., A. Tenesa, O.C.-X. and A. Talenti. All authors reviewed and approved the final draft.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-022-01251-4.

Correspondence and requests for materials should be addressed to Elena Bernabeu or Albert Tenesa.

Peer review information *Nature Genetics* thanks Seunggeun Lee and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Albert Tenesa

Last updated by author(s): Oct 26, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	We used PLINK to run our eQTL analysis (v1.9, freely available online at https://www.cog-genomics.org/plink2/). Custom python code used is openly available (https://zenodo.org/record/7093777) with doi: 10.5281/zenodo.7093777.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Data from the Genotype Tissue Expression (GTEx) project V8 release was employed, processed gene expression data downloaded from the GTEx portal, which is openly available: https://gtexportal.org. GTEx genotype data are available upon application through dbGaP.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	The GTEx project possesses genotype and gene expression data for 469 males and 237 females.
Population characteristics	The GTEx project participants were aged ranging between 20-70 at post-mortem recruitment, with ancestry being about 84% Europeans, 15% African American and 1% Asian or other.
Recruitment	The GTEx sample collection, performed on post-mortem adult donors, as described by them: "Data collection is organized under the Cancer Human Biobank (caHUB). Donors of any racial and ethnic group and sex who are age 21–70 in whom biospecimen collection can start within 24 hours of death are eligible. There are few medical exclusionary criteria: human immunodeficiency virus (HIV) infection or high-risk behaviors, viral hepatitis, metastatic cancer, chemotherapy or radiation therapy for any condition within the past 2 years, whole blood transfusion in past 48 hours, or body mass index ≥ 35 or ≤18.5. GTEx donors are identified through low-post-mortem-interval (PMI) autopsy or organ and tissue transplant settings. Currently, there is no way to directly sign up to be a GTEx donor. GTEx has agreements with organizations that are involved in surgery departments in various parts of the United States."
Ethics oversight	The authors did not participate in the recruitment of GTEx.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The GTEx v8 cohort possesses data for 706 individuals.
Data exclusions	No data was excluded.
Replication	No replication was conducted/was not relevant to this study.
Randomization	Not relevant to this study.
Blinding	The authors did not participate in the recruitment of GTEx.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging