Response to: 'Questions on 'Sequencing of the MHC region defines *HLA-DQA*1 as the major genetic risk for seropositive rheumatoid arthritis in Han Chinese population' by Guo *et al*' by Regueiro and Gonzalez

Jianping Guo ,[1] **Tao Zhang,**[2,3] **Xiaowei Li,**[3,4] **Hongzhi Cao,**[5,6,7] **Zhanguo Li**[1]

[1]Department of Rheumatology and Immunology, Peking University People's Hospital, Beijing, China
[2]Institute of Precision Health Research, Beijing genomics institute (BGI)-Shenzhen, Shenzhen, China
[3]Institute of Precision Health Research, China National GeneBank-Shenzhen, BGI-Shenzhen, Shenzhen, China
[4]Technology Service Center, Beijing genomics institute (BGI)-shenzhen, shenzhen, China
[5]Division of Immunology, Shenzhen Digital Life Institute, Shenzhen, China
[6]Division of Artificial Intelligence and Bioinformatics, iCarbonX, ShenZhen, China
[7]Division of Birth Cohort Study, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, China

Correspondence to Professor Jianping Guo, Department of Rheumatology and Immunology, Peking University People's Hospital, Beijing 100044, China; jianping.guo@bjmu.edu.cn

Handling editor Josef S Smolen

We appreciate Dr Gonzalez's interest and comments on our recent publication 'Sequencing of the major histocompatibility complex (MHC) region defines human leukocyte antigen (HLA)-DQA1 as the major genetic risk for seropositive rheumatoid arthritis in Han Chinese population'.[1 2] Dr Gonzalez's comments provide us with an opportunity to clarify and discuss the frequencies of amino-acids at position DQα1:160 and the protective association of DRβ1:37N in rheumatoid arthritis (RA), and to improve our study.

One of the concerns Dr Gonzalez expressed is the frequencies of three amino-acids (Asp (D), Ala (A) and Ser (S)) at DQα1:160 do not sum up to 1.0, that is, 0.20, 0.22 and ≈0.02 in controls; 0.36, 0.37 and ≈0.01 in RA patients. The explanation is that the 'minor frequencies' was set as default for all variants in PLINK. The original frequencies of three amino-acids (D, A and S) were 0.20, *0.78* (1–0.22) and ≈0.02 (the sum is 1.00) in controls and 0.36, *0.63* (1–0.37) and ≈0.01 (the sum is 1.00) in cases, respectively. In our original paper the frequencies of DQα1:160A inhealthy controls were similar to those reported in other Asian studies.[3] These results do not affect the calculation of p value, but do affect the odd ratio (OR) calculation. Indeed, by the omnibus test DQα1:160A showed a protective effect (OR=0.46, p=2.72x 10^{-35}, online supplementary table 10 in Guo *et al*).[2] We appreciate Dr Gonzalez *et al* for this important point and have made a correction for our publication, in which all variants have been presented according to original frequencies instead of minor frequencies.[4]

Regarding the protective effect of DRβ1:37N, although the identified amino-acid DRβ1:37N did not show any significant association in univariate regression analysis, it reached second strong statistical significance after conditioning on DQα1:160D in both discovery and validation stages, indicating an independent association. This phenomenon could be potentially explained by the Simpson's paradox, a striking observation that an association between two variables at the population-level might increase or decrease in quantity, or even change direction within the subgroups, depending on the set of variables being controlled,[5 6] and has been reported in several genetic association studies.[7 8] Notably, the DRβ1:11D also showed an independent protective effect and was in high linkage disequilibrium (LD) with DRβ1:37N (r2=0.62; online supplementary tables 8 and 9 in Guo *et al*).

Regarding other DRB1 variants, as the author indicated, by omnibus test we replicated the findings reported in previous studies,[9–12] including the position 11 and 13 at DRβ1, and the allele *DRB1*04:05*. However, our study focused on single nucleotide polymorphisms (SNPs), classical HLA alleles and the individual amino-acid variants rather than amino-acid positions, because a particular amino-acid(s) may have potential biological function(s). Furthermore, different amino-acids at same position may insert different functions.[13] Taking this into consideration, DQα1:160D remained the top association in omnibus test (OR=2.30, p=1.82x 10^{-38}) (online supplementary table 10 in Guo *et al*). Furthermore, consistent with our findings, Hirata *et al*[14] have also reported that one of DQα1:160D encoding allele *DQA1*0303* was a strong risk for susceptibility to RA in Japanese population (OR=2.65, p=2.0 × 10^{-173}, shown in table 1 in Hirata *et al*).

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Provenance and peer review Commissioned; internally peer reviewed.

To cite Guo J, Zhang T, Li X, *et al*. *Ann Rheum Dis* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/annrheumdis-2020-217064

Received 25 February 2020
Revised 22 April 2020
Accepted 22 April 2020

Linked

► http://dx.doi.org/10.1136/annrheumdis-2020-217031

Ann Rheum Dis 2020;**0**:1–2. doi:10.1136/annrheumdis-2020-217064

ORCID iD
Jianping Guo http://orcid.org/0000-0002-5031-3510

REFERENCES

1 Regueiro C, Gonzalez A. Questions on 'Sequencing of the MHC region defines *HLA-DQA*1 as the major genetic risk for seropositive rheumatoid arthritis in Han Chinese population' by Guo *et al*. *Ann Rheum Dis* 2020:annrheumdis-2020-217031.
2 Guo J, Zhang T, Cao H, *et al*. Sequencing of the MHC region defines *HLA-DQA1* as the major genetic risk for seropositive rheumatoid arthritis in Han Chinese population. *Ann Rheum Dis* 2019;78:773–80.
3 González-Galarza FF, Takeshita LYC, Santos EJM, *et al*. Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic Acids Res* 2015;43:D784–8.
4 Correction: Sequencing of the MHC region defines HLA-DQA1 as the major genetic risk for seropositive rheumatoid arthritis in Han Chinese population. *Ann Rheum Dis* 2020. doi:10.1136/annrheumdis-2018-214725corr1. [Epub ahead of print: 20 Apr 2020].
5 Simpson EH. The interpretation of interaction in contingency tables. *J R Stat Soc Series B* 1951;13:238–41.
6 Wagner CH. Simpson's paradox in real life. *The American Statistician* 1982;36:46–8.
7 Patsopoulos NA, Barcellos LF, Hintzen RQ, *et al*. Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. *PLoS Genet* 2013;9:e1003926.
8 Zhou F, Cao H, Zuo X, *et al*. Deep sequencing of the MHC region in the Chinese population contributes to studies of complex disease. *Nat Genet* 2016;48:740–6.

9 Raychaudhuri S, Sandor C, Stahl EA, *et al*. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet* 2012;44:291–6.
10 Okada Y, Suzuki A, Ikari K, *et al*. Contribution of a non-classical HLA gene, HLA-DOA, to the risk of rheumatoid arthritis. *Am J Hum Genet* 2016;99:366–74.
11 Liu X, Guo J, Jia Y, *et al*. HLA-DRB1 shared epitope-dependent DR-DQ haplotypes are associated with both anti-CCP-positive and -negative rheumatoid arthritis in Chinese Han. *PLoS One* 2013;8:e71373.
12 Okada Y, Kim K, Han B, *et al*. Risk for ACPA-positive rheumatoid arthritis is driven by shared HLA amino acid polymorphisms in Asian and European populations. *Hum Mol Genet* 2014;23:6916–26.
13 Scally SW, Law S-C, Ting YT, *et al*. Molecular basis for increased susceptibility of Indigenous North Americans to seropositive rheumatoid arthritis. *Ann Rheum Dis* 2017;76:1915–23.
14 Hirata J, Hosomichi K, Sakaue S, *et al*. Genetic and phenotypic landscape of the major histocompatibilty complex region in the Japanese population. *Nat Genet* 2019;51:470–80.