Author/s:
Erlichster, M;Bedo, J;Skafidas, E;Kwan, P;Kowalczyk, A;Goudey, B

Title:
Letter: improved parsimony of genetic risk scores for coeliac disease through refined HLA modelling

Date:
2021-03-01

Citation:
Erlichster, M., Bedo, J., Skafidas, E., Kwan, P., Kowalczyk, A. & Goudey, B. (2021). Letter: improved parsimony of genetic risk scores for coeliac disease through refined HLA modelling. Alimentary Pharmacology and Therapeutics, 53 (6), pp.759-760. https://doi.org/10.1111/apt.16263.

Persistent Link:
https://hdl.handle.net/11343/298278

DR BENJAMIN GOUDEY (Orcid ID : 0000-0002-2318-985X)

Article type : Letter to the Editors

DR BENJAMIN GOUDEY (Orcid ID : 0000-0002-2318-985X)

Article type : Letter to the Editors

Corresponding author mail id: bwgoudey@gmail.com

Letter: improved parsimony of genetic risk scores for coeliac disease through refined HLA modelling.

Authors: Michael Erlichster[1,2], Justin Bedo[3,4], Efstratios Skafidas[2, 5], Patrick Kwan[2,6], Adam Kowalczyk[4,7,8,9], Benjamin Goudey[4,9,10]

1. MX3 Diagnostics, Melbourne, Victoria, Australia
2. Department of Medicine (Royal Melbourne Hospital), The University of Melbourne, Melbourne, VIC, Australia
3. Bioinformatics Division, Walter and Eliza Hall Institute, Melbourne, VIC, Australia
4. School of Computing and Information Systems, The University of Melbourne, Melbourne, VIC, Australia
5. Department of Electrical and Electronic Engineering, The University of Melbourne, Melbourne, VIC, Australia
6. Department of Neuroscience, Central Clinical School, Monash University, Melbourne, VIC, Australia
7. Diversity Arrays Technology Pty Ltd, Canberra, ACT, Australia
8. Institute for Applied Ecology, University of Canberra, ACT, Australia
9. Centre for Epidemiology and Biostatistics, The University of Melbourne, Melbourne, VIC, Australia
10. IBM Research Australia, Melbourne, VIC, Australia

Editors,

We read with interest the recent article of Sharp et al. describing a single nucleotide polymorphism (SNP) genetic risk score (GRS) for coeliac disease[1]. The authors report a 42 SNP model, combining the 4 established HLA risk alleles for coeliac disease (DQ2.5, DQ2.2, DQ8, DQ7.5) and 38 non-risk variants, which achieves a similar quality of prediction (**AUC = 0.875**) as the best previously reported GRS model with several hundreds of SNPs (**AUC = 0.87–0.89**)[2]. Using data from the UK Biobank, the authors demonstrate that the resulting model provides better discrimination of coeliac disease risk than HLA-DQ stratification alone (**AUC = 0.881 vs 0.815**), where the contribution of the HLA region is represented using three risk categories. The report by Sharp et al. highlights an important direction of research - the development of parsimonious models that can be clinically deployed for low cost. Moreover, the reduced complexity of their model allows for greater biological insight. However, this idea can be extended further when assessing coeliac disease risk.

Building on the same observation – that combinatorial HLA-DQ risk genotypes may be used to more effectively stratify risk – we recently published a GRS constructed on the same European coeliac disease GWAS datasets[3]. In this work we described two HLA-based risk model achieving equivalent quality of prediction to the best-performing genome-wide coeliac disease risk models[2]; HDQ_{15} – a model derived by ordering the 15 HLA-DQ risk genotypes with 4 tagging SNPs (**AUC = 0.871**) and HDQ_{17} – an extension integrating 2 novel HLA risk alleles and 2 additional SNPs (**AUC=0.879**). In particular, our HDQ_{15} is very similar to the HLA stratification presented by Sharp et al. Importantly, we found that the contribution of non-HLA variants beyond the 4 or 6 SNPs used in these models was limited.

To highlight the utility of these even more parsimonious HLA-only models, we applied each model to 409,624 Caucasian Europeans from UK Biobank cohort, defining coeliac disease by hospital admission code and self-reported questionnaires as per Sharp et al. In this population, the 42 SNP Sharp model (AUC = **0. 8674**) still significantly outperforms a 3 category HLA risk model (**AUC = 0.781,** one-side Delong's test p-value=1.3×10^{-127}). However, our 6 SNP HDQ_{17} is almost identical in AUC to the Sharp model (**AUC = 0.8672**, p-value=0.47), while the 4 SNP HDQ_{15} has only marginally lower performance (**AUC: 0. 855**, p-value=4.9×10^{-6}).

These results suggest that most of the improvement in AUC reported by Sharp et al. can be achieved with fewer variants through a more nuanced use of HLA-attributed risk. The HLA-only models, HDQ15 and HDQ17, can be implemented using only 4 or 6 SNPs respectively, which may improve robustness and decrease cost in clinical deployment. While non-HLA genomic risk variants are critical for coeliac disease aetiology, our results also highlight that further investigation is needed to understand the relative contributions of HLA and non-HLA variations for predicting coeliac disease risk.

References

1 Sharp, S. A. *et al.* A single nucleotide polymorphism genetic risk score to aid diagnosis of coeliac disease: a pilot study in clinical care. *Aliment Pharmacol Ther 2020;* **52**, 1165-1173.

2 Abraham, G. *et al.* Accurate and robust genomic prediction of celiac disease using statistical learning. *PLoS Genetics* 2014; **10**, e1004137.

3 Erlichster, M. *et al.* Improved HLA-based prediction of coeliac disease identifies two novel genetic interactions. *European Journal of Human Genetics 2020*, 1-10.

Acknowledgements

This research has been conducted using the UK Biobank Resource (application 51064).

Declaration of personal interests: None.
Declaration of funding interests: None.

Authorship

Guarantor of the article: Dr Benjamin Goudey is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Author contributions: ME and BG designed the study, contributed to analysis and wrote the manuscript. JB, ES, PK, and AK contributed to analysis and reviewed the manuscript. All authors contributed to the discussion and reviewed or edited the manuscript. All authors approved the final version of the manuscript.

Figures

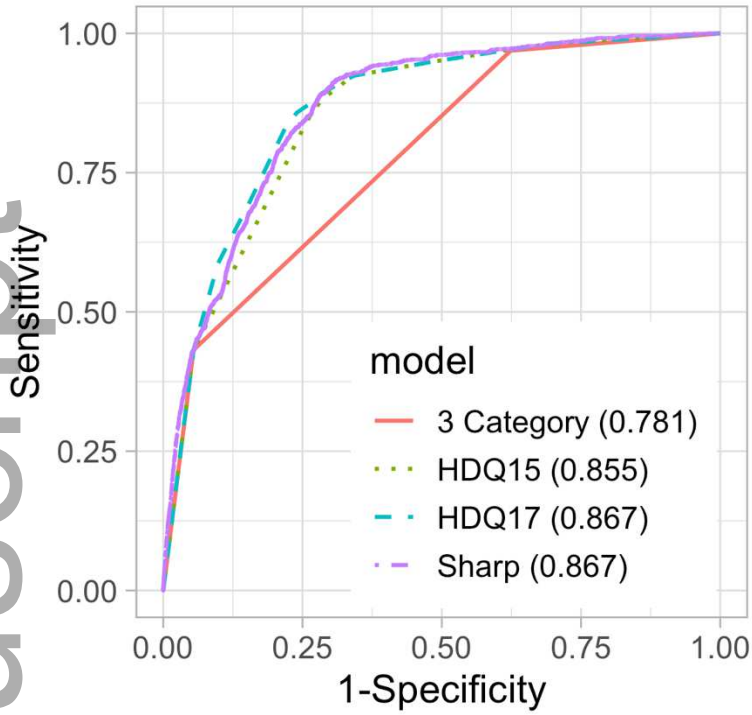
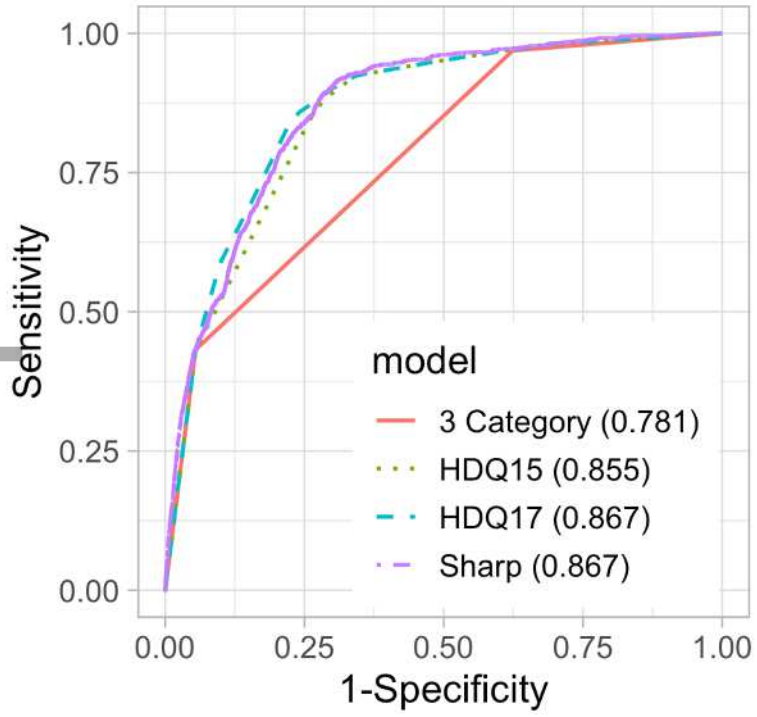


Figure 1: ROC curves of four CD risk models showing the baseline 3-SNP HLA model ("3 Category"), the 4- and 6-SNP HLA based models ("HDQ15" ,"HDQ17") from Erlichster et al. and the 42 SNP model from Sharp et al ("Sharp"). *All model were evaluated on the same European subset of the UK Biobank, with weights for each model derived on external cohorts. The AUCs for each model is shown in parentheses next to their names in the legend.*



apt_16263_f1.png