

HLA associations differ by ethnicity and aquaporin-4 antibody status in neuromyelitis optica spectrum disorders

Moritz Niederschweiberer, Georgia Panagiotaropoulou, Marcelo A. Fernández-Viña, Gonzalo Montero Martin, Kazutoyo Osoegawa, Megan K. Behne, Andre Franke, Wolfgang Lieb, Sven Jarius, Brigitte Wildemann, Patrick Schindler, Carolin Otto, Friedemann Paul, Eoin P. Flanagan, Lawrence J. Cook, Jorge R. Oksenberg, Michael R. Yeaman, Stephan Ripke, Klemens Ruprecht, with the Guthy-Jackson Charitable Foundation CIRCLES Study Group

This document contains eMethods.

HLA genotyping, sequence data analysis and genotype assignment

EDTA-treated blood samples were collected at CIRCLES sites or at patient education events. These were transported within 24 hours by express courier to a commercial laboratory where genomic DNA was extracted using standardized protocols, including salt-precipitation or phenol/chloroform methods. The resulting DNA was isolated, authenticated for purity and stored at -80°C within 24 hours. HLA genotyping on genomic DNA of patients with NMOSD as well as of White, Hispanic, and Black healthy controls from the University of California San Francisco (UCSF) Multiple Sclerosis Biorepository of the Immunogenetics of Neurological Diseases working Group (<http://indigo.ucsf.edu/> [INDIGO], see also below) was performed as previously described.¹ High-resolution next-generation sequencing (NGS) was applied to define non-ambiguous 3-field alleles for HLA class I (*HLA-A*, *HLA-B*, *HLA-C*) and HLA class II (*HLA-DRB1*, *HLA-DRB3*, *HLA-DRB4*, *HLA-DRB5*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DPA1*, and *HLA-DPB1*) loci using the MIA FORA NGS

high-throughput semi-automated typing protocol (Immucor, Inc., Norcross, GA, USA). Subsequent analyses and alignment of HLA genotypes were likewise performed as previously described.¹ NGS sequencing data were uploaded into the MIA FORA FLEX v3.0 HLA genotype analysis software (Immucor, Inc., Norcross, GA, USA), which analyzed these raw sequencing data based on HLA reference sequences from the IPD-IMGT/HLA database version 3.25.0.² Final HLA genotypes were confirmed after manual review of initial software genotype calls.¹ All HLA allele name data generated during this study was additionally reviewed according to the official updates of the IPD-IMGT/HLA database.³

Imputation and Genotyping of the PopGen control cohort

We imputed HLA alleles (2-field resolution) of participants of the PopGen control cohort from raw genome wide sequence data. Briefly, we first performed genome wide quality control of PopGen data using Ricopili.⁴ The parameters used for retaining subjects and SNPs were: SNP missingness <0.05 (before sample removal); subject missingness <0.02; autosomal heterozygosity deviation ($|F_{het}| < 0.2$); SNP missingness <0.02 (after sample removal); difference in SNP missingness between cases and controls <0.02; and SNP Hardy-Weinberg equilibrium ($p > 10^{-6}$ in controls or $p > 10^{-10}$ in cases). 493,560 SNP (8,198 in the 25-35 Mb HLA region) were retained for further processing, while no subjects were excluded as ethnical outliers. HLA-wide genotype imputation was performed using the pre-phasing/imputation stepwise approach implemented in SHAPEIT/IMPUTE2. The imputation reference set consisted of 1,006 phased haplotypes from the European subset of the 1000 Genomes Project dataset (IMGT 3.28 release with HLA Class II variants imputed from whole-genome sequencing using PolyPheMe⁵). The panel was obtained from

CookHLA⁶ in hg18 alignment, converted to hg19 using liftOver⁷ and then to Ricopili format. The imputation process yielded data on *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DQB1*, and *HLA-DRB1* alleles of participants of the PopGen control cohort.

References

1. Creary LE, Gangavarapu S, Mallempati KC, et al. Next-generation sequencing reveals new information about HLA allele and haplotype diversity in a large European American population. *Hum Immunol* 2019; 80: 807-822. 20190722. DOI: 10.1016/j.humimm.2019.07.275.
2. Robinson J, Halliwell JA, Hayhurst JD, et al. The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res* 2015; 43: D423-431. 20141120. DOI: 10.1093/nar/gku1161.
3. Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools, <http://www.allelefrequencies.net/hla6006a.asp> (2022, accessed 03/13/2023 2023).
4. Lam M, Awasthi S, Watson HJ, et al. RICOPILI: Rapid Imputation for COnsortias PIpeLIne. *Bioinformatics* 2020; 36: 930-933. DOI: 10.1093/bioinformatics/btz633.
5. Abi-Rached L, Gouret P, Yeh JH, et al. Immune diversity sheds light on missing variation in worldwide genetic diversity panels. *PLoS One* 2018; 13: e0206512. 20181026. DOI: 10.1371/journal.pone.0206512.
6. Cook S, Choi W, Lim H, et al. Accurate imputation of human leukocyte antigens with CookHLA. *Nat Commun* 2021; 12: 1264. 20210224. DOI: 10.1038/s41467-021-21541-5.
7. Hinrichs AS, Karolchik D, Baertsch R, et al. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res* 2006; 34: D590-598. DOI: 10.1093/nar/gkj144.