medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20217059; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Fine mapping of the *HLA* locus in Parkinson's disease in Europeans

Running head: HLA in Parkinson's disease

Eric Yu, BSc,[1,2] Aditya Ambati, PhD,[3] Maren Stolp Andersen, MD,[4,5] Lynne Krohn, MSc,[1,2] Mehrdad A. Estiar, MSc,[1,2] Prabhjyot Saini, MSc,[1,2] Konstantin Senkevich, MD, PhD,[2,6] Yuri L. Sosero, MD,[1,2] Ashwin Ashok Kumar Sreelatha, MSc, MTech,[7] Jennifer A. Ruskey, MSc,[2,6] Farnaz Asayesh, MSc,[2,6] Dan Spiegelman, MSc,[2,6] Mathias Toft, MD, PhD,[4,5] Marte K. Viken, PhD,[8,9] the International Parkinson's Disease Genomics Consortium (IPDGC), Manu Sharma, PhD,[7] Cornelis Blauwendraat, PhD,[10] Lasse Pihlstrøm, MD, PhD,[4] Emmanuel Mignot, MD, PhD*,[3] Ziv Gan-Or, MD, PhD*.[1,2,6]

*Equal contribution

Affiliations:

1. Department of Human Genetics, McGill University, Montréal, QC, Canada.

2. Montreal Neurological Institute and Hospital, McGill University, Montréal, QC, Canada.

3. Stanford Center For Sleep Sciences and Medicine, Dept of Psychiatry and Behavioral Sciences, Stanford University, Palo Alto, California, USA.

4. Department of Neurology, Oslo University Hospital, Oslo, Norway.

5. Institute of Clinical Medicine, University of Oslo, Oslo, Norway.

6. Department of Neurology and Neurosurgery, McGill University, Montréal, QC, Canada.

7. Centre for Genetic Epidemiology, Institute for Clinical Epidemiology and Applied Biometry, University of Tübingen, Tübingen, Germany

8. Department of Medical Genetics, University of Oslo and Oslo University Hospital, Oslo, Norway.

9. Department of Immunology, Oslo University Hospital, Oslo, Norway.

10. Molecular Genetics Section, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD.

Corresponding authors:

Emmanuel Mignot
Stanford University Center For Sleep Sciences,
3165 Porter Drive, #2178,
Palo Alto CA 94304, USA
Phone: (650) 725-6517
Email: mignot@stanford.edu

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20217059; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Ziv Gan-Or
Montreal Neurological Institute,
McGill University
1033 Pine Avenue, West,
Ludmer Pavilion, room 312
Montreal, QC, H3A 1A1,
Phone: +1-514-398-5845
Email: ziv.gan-or@mcgill.ca

Character Counts
Title: 55
Running Head: 23

Word Counts
Abstract: 248
Introduction: 446
Discussion: 1213
Body: 1912

Attachments
Figures: 2
Tables: 3
Supplementary Figures and Tables:
Supplementary Figures – 1
Supplementary Tables - 7

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20217059; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Abstract

Objective: Previous studies suggested a role for the immune system in Parkinson's disease (PD), and one of the hits in recent genome-wide association studies (GWASs) of PD is within the human leukocyte antigen (*HLA)* locus. Several associations of different *HLA* genes have been suggested, yet it is not clear which associations are relevant for PD.

Methods: We performed a thorough analysis of the *HLA* locus in 13,770 PD patients, 20,214 proxy-cases and 490,861 controls of European origin. We used GWAS data to impute *HLA* types and performed multiple regression models to examine the association of specific *HLA* types, different haplotypes and specific amino acid changes. We further performed conditional analyses to identify specific alleles or genetic variants that drive the association with PD.

Results: Four *HLA* types were associated with PD after correction for multiple comparisons, *HLA-DQA1**03:01, *HLA-DQB1**03:02, *HLA-DRB1**04:01 and *HLA-DRB1**04:04. Haplotype analyses followed by amino-acid analysis and conditional analyses suggested that the association is protective and primarily driven by three specific amino acid polymorphisms present in most *HLA-DRB1**04 subtypes - 11V, 13H and 33H (OR=0.87 95%CI=0.83-0.90, $p<8.23\times10^{-9}$ for all three variants). No other effects were present after adjustment for these amino acids.

Interpretation: Our results suggest that specific variants in the *HLA-DRB1* gene are associated with reduced risk of PD, providing additional evidence for the role of the immune system in PD. Although effect size is small and has no diagnostic significance, understanding the mechanism underlying this association may lead to identification of new targets for therapeutics development.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20217059; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Introduction

Parkinson's disease (PD), a common neurodegenerative movement disorder, is pathologically characterized by degeneration of dopaminergic neurons in the substantia nigra and by accumulation of α□synuclein in Lewy bodies or Lewy neurites in different brain regions. [1] Although PD is primarily neurodegenerative, an potential role of the immune system in the pathophysiology of PD is increasingly recognised based on animal and human studies.[2-4] These data suggest that the immune system can be involved in the initiation of PD, as well as in its progression, and that this involvement can be peripheral and central.[4, 5]

Neuropathological studies have shown evidence for microglia activation in brains of PD patients, although it was initially unclear whether this activation was a part of the disease process, a consequence, or an epiphenomenon.[6] However, recent evidence suggest that microglial activation and a central immune response may have an important role in the pathogenesis of PD. Evidence for peripheral inflammation in the gastrointestinal system and blood of PD patients may suggest that early peripheral immune response may also have a role in the pathogenesis of PD, although the cause of this response is still unknown.[4, 5]Genetic evidence also links the immune system with PD, since genes such as *LRRK2*, the *HLA* locus and possibly *BST1*, all associated with PD[7] and have a role in the immune system.[8-10]

The *HLA* genomic region on chromosome 6 includes genes that encode components of the major histocompatibility complex (MHC).[9] The MHC are divided into three subclasses of genes: class I (including the classical *HLA□A*, *HLA□B* and *HLA□C* genes and the nonclassical, nonpolymorphic *HLA□E*, *HLA□F* and *HLA□G* genes), class II (including *HLA□DPA1*, *HLA□DPB1*, *HLA□DQA1*, *HLA□DQA2*, *HLA□DQB1*, *HLA□DQB2*, *HLA□DRA*, *HLA□DRB1*, *HLA□DRB2*, *HLA□DRB3*, *HLA□DRB4* and *HLA□DRB5*) and class III (including other

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20217059; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

inflammatory response, complement and white blood cell-maturation genes)).[9] Several genome-wide association studies (GWASs) of PD have shown an association between the *HLA* genomic region and risk of PD. In the most recent GWAS meta-analysis, an association with *HLA-DRB5* has been reported, with a potential effect of the rs112485576 single nucleotide polymorphism (SNP) on the expression of *HLA-DRB5*.[7] Previous studies have suggested different associations with *HLA-DQA2*, *HLA-DQB1*, *HLA-DRA*, *HLA-DRB1*, *HLA-DRB5* and with different haplotypes within the *HLA* region in Europeans.[11-16]

In order to establish the association of the *HLA* locus with PD and identify the specific *HLA* type or types associated with PD, we performed the largest HLA alleles, haplotypes and amino acid analyses to date in PD, in a total of 12,137 patients, 14,422 proxy patients and 351,953 controls. We further performed haplotype and amino acid analyses, as well as conditional analyses, to fine map the *HLA* region and identify specific drivers of the association with PD in this region.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20217059; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Methods

Study population

This study was designed as a meta-analysis of multiple cohorts, including a total of 13,770 PD patients, 20,214 proxy-patients and 490,861 controls, as detailed in Supplementary Table 1. In brief, we included cohorts and datasets from eight independent sources: International Parkinson's Disease Genomics Consortium (IPDGC) NeuroX dataset (dbGap phs000918.v1.p1, including datasets from multiple independent cohorts as previously described),[17] McGill University (McGill), National Institute of Neurological Disorders and Stroke (NINDS) Genome-Wide genotyping in Parkinson's Disease (dbGap phs000089.v4.p2),[18] NeuroGenetics Research Consortium (NGRC) (dbGap phs000196.v3.p1),[12] Oslo Parkinson's Disease Study (Oslo), Parkinson's Progression Markers Initiative (PPMI), Vance (dbGap phs000394) and PD cases and proxy-cases from the UK Biobank (UKB). Proxy-cases are first degree relatives of PD patients, thus sharing ~50% of the patients' genetic background and eligible to serve as proxies, as previously described.[19] All cohorts were previously included in the most recent PD GWAS.[7] Study protocols were approved by the relevant Institutional Review Boards and all patients signed informed consent before participating in the study.

Pre-imputation genotype quality control

In order to include only high-quality samples and SNPs, standard quality control (QC) was performed on all datasets individually using PLINK v1.9.[20] Standard GWAS QC was done to filter out samples and SNPs with low call rate, heterozygote outliers along with gender mismatch as previously described.[7] Only samples of European ancestry clustering with HapMap v3 using principal component analysis were included as shown in Supplementary Figure 1. In order to

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20217059; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

exclude related individuals, we examined relatedness in each dataset separately, followed by relatedness test across all datasets combined, to exclude individuals who were included in more than one dataset. All individuals with pi_hat >0.125 were excluded using GCTA v1.26.0.[21]

UK Biobank quality control

For the analysis of the UKB data, unrelated participants of European ancestry (field 22006), with low missingness rate (field 220027) were included after exclusion of heterozygosity outliers as previously described.[7]. PD patients from the UK Biobank were included based on self-report (field 20002) or based on their International Classification of disease diagnosis code (ICD-10, code G20, field 41270). From the remaining participants, proxy-cases were defined as first degree relatives (parents or siblings, field 20112-20114) of patients with PD. Principal components were calculated using flashpca[22] after excluding related individuals as described above. The control group was divided randomly to two groups of controls: one was included in the GWAS comparing PD patients from UKB and controls, and the second was included in the GWAS comparing proxy-cases from UKB and controls. This division was done proportionally to the size of each GWAS.

Imputation

For SNP imputation of each dataset, we used the Michigan Imputation Server on the 1,000 Genomes Project panel (Phase 3, Version 5) using Minimac3 and SHAPEIT v2.r790. Imputed UK Biobank genotyped data v3 were downloaded in July 2019. All variants with an imputation quality (r^2) of >0.30 were labelled as soft calls and >0.80 were labelled as hard calls. Soft calls were only used together with the hard calls for polygenic risk score calculation (see below); hard calls were used for all other analyses.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20217059; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Association analysis of common variants on chromosome 6

Prior to determining HLA types, we performed a simple association test of all SNPs located on chromosome 6, to verify that we identified the same hit in the HLA region as previously described.[7] For this purpose, we generated summary statistics of chromosome 6 for each dataset, and used logistic regression with an additive model adjusting for age at onset for patients and age at enrollment of controls, sex and population stratification (first 10 principal components) with PLINK v2.00a2LM (25 Oct 2019).[20] The UK Biobank data was analyzed similarly using logistic regression adjusting for age, sex, the first 10 principal components and Townsend index to account for additional potential population stratification confounders. Finally, to harmonize effects in cases and proxy-cases, summary statistics for proxy-cases were rescaled based on genome-wide association study by proxy (GWAX) as previously described.[19] To meta-analyze the different datasets, we performed a fixed-effect meta-analysis using METAL with an inverse-variance-based model.[23]

HLA locus analysis

HLA imputation

To impute specific HLA types for each individual, we inferred two field resolution HLA alleles using HIBAG v1.22.0, a statistical method for HLA type imputation in R.[24] HIBAG was shown to be as accurate or more accurate in Europeans compared to other types of HLA imputation tools.[25] HIBAG provided a reference panel for Europeans (n = 2,572) with high imputation accuracy for *HLA-A*, *HLA-B*, *HLA-C*, class I genes, and *HLA-DPB1*, *HLA-DQA1*, *HLA-DQB1*, and *HLA-DRB1*, class II genes. *HLA-DRB3*, *HLA-DRB4* and *HLA-DRB5* imputation models were trained using HIBAG[24] on European origin sample training set (n = 3,267) genotyped on

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20217059; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

the Illumina Infinium PsychArray-24 chip and fully sequenced at 8-digit resolution for HLA loci. These models were validated in a test set (n=886) with high accuracy (Supplementary Table 2). Imputation accuracy for European *DRB1**04 alleles was determined for *DRB1**04:01 *DRB1**04:02 *DRB1**04:03, *DRB1**04:04, *DRB1**04:05, *DRB1**04:07, *DRB1**04:08. Alleles with an imputation probability of <0.5 were defined as undetermined and individuals with two or more undetermined alleles were excluded from the analysis (Supplementary Table 1 details the numbers included for each allele in each cohort after all quality control steps). To further examine imputation accuracy, the results of the *DRB1* imputation were compared against high throughput HLA sequencing in 380 PD samples from Oslo. The combined frequency of seven different *DRB1**04 alleles detected in sequence data was 0.15 with the 04:01 and 04:04 alleles being the most common (Supplementary Table 3). Imputation accuracy for DRB1*04 alleles was very high at 2-digit resolution (Supplementary Table 4).

Statistical analysis

To examine the association of HLA alleles with PD, we used R to perform logistic regression, adjusting for age at onset, sex and the first 10 principal components. The UK Biobank dataset was also adjusted for Townsend index. Haplotype analyses were performed using haplo.stats in R with logistic regression as stated above. Only haplotypes with posterior probability >0.2 and carrier frequency of >1% were included in the analysis. Amino acid association analyses were performed using HIBAG after converting P-coded alleles to amino acid sequences for exon 2, 3 of HLA class I genes and exon 2 of class II genes. Amino acid associations were tested using logistic regression as described above. A polygenic risk score (PRS) was calculated using PRSice v 2.2.11 without LD clumping or P thresholding.[26] The beta weights from the summary statistics of the 90 genome-wide significant variants in the latest PD GWAS[7] were used in the

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20217059; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

PRS. To make sure that all possible variants were included in the PRS analysis, we also performed imputation using the Haplotype Reference Consortium panel (HRC) (Version r1.1 2016) with Minimac4 and Eagle v2.4. Ambiguous variant (rs6658353) and rs112485576 from the HLA region were excluded from the PRS calculation. To examine whether secondary hits exist in the HLA region, we adjusted for significant HLA variants, HLA alleles, HLA amino acid changes and PRS, by introducing significant findings from the first analyses as covariates in the regression models. Statistical analyses were only performed on alleles, haplotypes and amino acid changes with more than 1% carrier frequency. *P* value significance levels were adjusted using Bonferroni correction. Meta-analysis was performed as described above.

Code availability

The scripts used in this analysis is available at https://github.com/gan-orlab/HLA_HIBAG/.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20217059; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Results

Meta-analysis of HLA types in Parkinson's disease suggests a single association

After standard QC, a total of 12,137 patients, 14,422 proxy patients and 351,953 controls were included in the analysis (Supplementary Table 1). As shown in Figure 1A, our SNP-level meta-analysis validated the previous association for SNP, rs112485576, in the *HLA* locus (In the current analysis: OR=0.87, 95% CI= 0.83-0.90, $p=5.00x10^{-13}$, in the previous meta analysis: OR=0.85, 95% CI= 0.82-0.87, $p=6.96x10^{-28}$).[7] No residual HLA effects were found after adjustment of this SNP (Figure 1B, C), indicating that the association of this locus was primarily driven by a single genetic risk factor.

We next performed a meta-analysis association study of all *HLA* types with carrier frequency above 1%. After *HLA* imputation, a total of 141 different *HLA* types across 10 *HLA* loci were included (setting the Bonferroni corrected threshold for statistical significance on $\alpha=3.55x10^{-4}$; 0.05/141). Following these analyses, we found four *HLA* alleles that were associated with PD (Table 1, results for other *HLA* alleles are detailed in Supplementary Table 5): *HLA-DQA1**03:01, *HLA-DQB1**03:02, *HLA-DRB1**04:01, *HLA-DRB1**04:04. These four alleles are all located within a small genomic segment and have similar odds ratio ranging between 0.84-0.89. Three of the four alleles have similar carrier frequencies, indicating that they could be part of the same haplotypes, with the fourth potentially representing a sub-haplotype (Table 1).

HLA haplotype analysis

For haplotype analysis, we allowed for up to three genes to be included in each haplotype, since including more than three genes generated multiple haplotypes with low allele frequency that

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20217059; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

could not be analyzed at the current sample size. A total of 84 different HLA haplotypes (Supplementary Table 6) with allele frequency >1% were identified, setting the cut-off Bonferroni corrected p value for statistical significance at $\alpha=5.95\times10^{-4}$. Three different HLA haplotypes were associated with PD after correction for multiple comparisons: *DQA1**03:01~*DQB1**03:02, *DRB1**04:01~*DQA1**03:03 and *DRB1**04:04~*DQA1**03:01. Upon further examination, this association was found to be driven by several well-known sub-haplotypes, *DRB1**04:04~*DQA1**03:01~*DQB1**03:02 and *DRB1**04:01~*DQA1**03:01/3~*DQB1**03:01/2 (Supplementary Table 6). Because both *DQA1**03:01 and *DQA1**03:03 as well as *DQB1**03:01 and *DQB1**03:03 are present within the extended *DRB1**04:01 haplotype, it is likely that these associations are driven by *DRB1*.

Meta-analysis of the association of HLA amino acid changes with Parkinson's disease

To further identify the specific source of the association in the *HLA* locus, we performed an analysis of 636 amino acid changes in the HLA genes, setting the cut-off Bonferroni corrected p value for statistical significance at $\alpha=7.86\times10^{-5}$. Ten amino acid changes were significantly associated with reduced risk of PD (Supplementary Table 7). The top three associated variants are linked amino-acids 11V, 13H and 33H (Table 3) present in all *DRB1**04 subtypes, complementing the HLA haplotype analysis. Four other variants, 26S, 47Q, 56R and 76V, in the *DQA1* gene, are in perfect LD with each other ($r^2=1$, D'=1, Supplementary Table 7) and in partial LD ($r^2=0.38$, D'=0.85) with the *DRB1* variants. The association of these *DQA1* variants is weaker than the *DRB1* variants in terms of both effect size and statistical association (Supplementary Table 7). Three other variants, 71T, 74E, and 75L, are in the *DQB1* gene, are also in perfect LD with each other ($r^2=1$, D'=1, Supplementary Table 7) and in partial LD ($r^2=0.16$, D'=0.99) with *DRB1* 13H and 33H.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20217059; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Conditional analyses confirm that *DRB104 amino acid variants likely drive the association of the *HLA* locus with PD**

To further determine the specific genes or variants that drive these associations, we performed a set of conditional analyses, and re-analyzed the allele types, haplotypes and amino acid associations with PD. We conditioned the HLA type regression model on the following: rs112485576 (the top GWAS hit in the *HLA* locus), *DQA1**03:01, *DQA1**03:03 and the *DRB1* variant 13H. We have also adjusted for the PD PRS, to examine a potential polygenic effect (Supplementary Tables 5, 7). While the adjustment for the *DRB1* variant 13H completely eliminated the associations in the *DQA1* gene, adjustment for *DQA1**03:01 and *DQA1**03:03 did not completely eliminate the association of the *DRB1* gene (Supplementary Table 5), again supporting this gene and these specific amino acids (11V, 13H and 33H, Table 3) as the drivers of the association in the *HLA* locus. Adjustment for PRS did not change the results.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20217059; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Discussion

In the current study, we performed a thorough analysis of the *HLA* region and examined its association with PD in the European population using a total of 12,137 patients, 14,422 proxy patients and 351,953 controls. Following a series of regression models and conditional analyses, our results indicate that the main, possibly sole, driver of the association in the *HLA* region are three amino acid changes specific of *HLA-DRB1**04 subtypes, 11V, 13H and 33H (Figure 2). Two of these amino acid changes, 13H and 33H are in perfect LD, and 11V is in very strong LD with the other two variants. This study agrees with a smaller HLA sequencing study[13] in 1,597 PD cases and 1,606 controls which also observed a protective effect of *DRB1**04 and the same amino acids, although it also reported additional associations with *DRB1**01:01 and *DRB1*10:01* which were not confirmed in the current study. Interestingly, the V-H-H motif at position 11V, 13H and 33H are central to the DRB1*04:01 heterodimer and contribute to peptide binding, notably through pocket P6[27] (Figure 2).

Previous studies on the *HLA* genomic region in PD have reported associations of different genes and HLA types with PD, including *HLA-DQA2*, *HLA-DQB1*, *HLA-DRA*, *HLA-DRB1* and *HLA-DRB5*.[11-16] The suggested association with *HLA-DRB5* reported in the most recent PD GWAS is based on an expression quantitative trait locus (eQTL) analysis, as the top associated SNP in this region, rs112485576, was also associated with differential expression of *HLA-DRB*.[7] A previous study of 2,000 PD patients and 1,986 controls has implicated a non-coding variant (rs3129882) within *HLA-DRA* as driving the association with PD and suggested that this variant affects the expression of *HLA-DR* and *HLA-DQ* genes.[12] Similarly, another study suggested that the same variant in *HLA-DRA* (rs3129882) is associated with differential expression of MHC-II on immune cells.[28] While our study does not rule out this possibility, since

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20217059; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

the main variants driving the association are amino acid changes in *DRB1**04 that will affect epitope binding ability, it is likely that the effect on PD risk is through these variants and not due to modified expression. Additional functional studies will be required to study this hypothesis.

The current study adds further support to the hypothesis suggesting an involvement of the peripheral and central immune system in PD. On top of the *HLA* locus, several other genes with potential roles in the immune system, including *LRRK2* and potentially *BST1*,[8, 10] have been implicated in PD.[7] In the periphery, there are notable changes in the immune system of PD patients compared to controls, as peripheral monocytes have differential expression of immune related proteins and markers.[4] Whether these changes are drivers of the disease or a result of the disease is still undetermined, but accumulating evidence suggest that they can be part of the pathogenic process of PD. In the central nervous system, pathological studies suggest that microglia cells may have a central role in PD.[29] Microgliosis is a prominent pathological finding in post-mortem brains of PD patients, and evidence suggest that microglia activation occurs early in the disease process and may be involved in the pathogenesis of PD.[4] The specific contribution of *HLA* to these processes is still unclear and needs to be further studied.

One intriguing possibility that may directly involve *HLA* with PD is the potential interaction of HLA-DRB1*04 with α-synuclein, notably an epitope surrounding p.S129. Recent data has shown that α-synuclein fragments can bind MHC and increase T cell reactivity.[30] This activity is proinflammatory, involves both CD4 and CD8 cells and may occur before the onset of motor symptoms,[30, 31] suggesting an involvement of inflammation in early PD pathogenesis. Studying specific α-synuclein fragments has suggested that two major regions of α-synuclein may be associated with increased T cell reactivity in PD, with preferential CD4 activity: an N-terminal region involving amino acid p.Y39 and a C-terminal region surrounding amino acid

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20217059; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

p.S129, two important residues undergoing phosphorylation.[30] The phosphorylation of p.S129 is particularly interesting as it is well known in promoting aggregation.[32, 33] Further analysis focused on the p.Y39 region suggested an association with α-synuclein-specific p.Y39 T cell responses and HLA DRB1*15:01 and DRB5*01:01 presentation. This association was abolished by phosphorylation, which reduced binding of p.Y39-phosphorylated α-synuclein.[30, 31] Other experiments by these authors have also suggested CD8+ T cells responses mediated by HLA-A11*01 presentation of epitopes in the same N-terminal region of α-synuclein. However, in the current study we could not confirm an association of these HLA types with PD.

More interestingly in the context of our work, increased CD4+ T cell response to both p.S129 phosphorylated and unphosphorylated α-synuclein was also demonstrated, suggesting involvement of DQB1*05:01 and DQB1*04:02, as these alleles strongly bound these α-synuclein epitopes.[30] These HLA alleles, however, are not associated with risk of PD in the current study. However, the authors reported in the supplementary data that DRB1*04:01 was also a selective and strong binder of the same α-synuclein epitope with p.S129, but only when the epitope was unphosphorylated.[30] Notably, no other DRB1 alleles that were assayed in this study[30] had increased binding affinity to α-synuclein epitope with p.S129, except for the DRB1*04:01 allele that was a strong binder only when unphosphorylated. Binding register analysis using Immune Epitope Database (IEDB) MHC-II Binding Prediction (http://tools.iedb.org/mhcii/) suggests that this epitope binds a 9 amino acid AYEMPSEEG core, with p.S129 at P6 position, a position postulated to be important based on our HLA-DR amino acid analysis presented above. As CD4+ T cell responses are generally stronger when epitopes are presented by HLA-DR versus HLA-DQ,[34] HLA-DRB1*04 responses to the p.S129 unphosphorylated form of α-synuclein could be dominant in individuals with HLA-DRB1*04,

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20217059; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

explaining the protective effect of this HLA subtype in PD. Additional experiments will be needed to further explore this hypothesis.

Our study has several limitations. First, this study was performed on European populations, and the results may be limited to this population only. Additional studies in other populations are required. Several studies on *HLA* types and PD have been performed in Asian populations,[35-38] and the GWAS risk variant rs112485576 has a similar OR (0.85) in the largest Asian GWAS to date,[39] yet larger studies are required, as well as studies in other populations. An additional potential limitation of our study is its use of imputation rather than fully sequenced HLA types. Given the very high performance of the imputation tool when compared to full sequencing (Supplementary Table 2), the potential effect of imputation inaccuracies is likely small and should be diluted in our large sample size. In addition, we cannot rule out that other, rarer HLA types that were not included in the current analysis may also have a role in PD. An additional limitation of the current study is that by adjusting for sex we eliminate potential sex-specific effects. It is possible that specific HLA types are relevant in one sex more or less than the other, and this should be studied in larger, sex-stratified cohorts.

To conclude, our results suggest a role for the *HLA-DRB1* gene in susceptibility for PD, and provide further evidence for the importance of the immune system in PD. Since the effect is small, it does not merit routine HLA typing in PD, but understanding the mechanism underlying this association may lead to better understanding of PD in general and offer new targets for future immune-related treatment.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20217059; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Acknowledgements

We would like to thank the participants in the different cohorts. This work was financially supported by grants from the Michael J. Fox Foundation, the Canadian Consortium on Neurodegeneration in Aging (CCNA), the Canada First Research Excellence Fund (CFREF), awarded to McGill University for the Healthy Brains for Healthy Lives initiative (HBHL), and Parkinson Canada. We would like to thank all members of the International Parkinson Disease Genomics Consortium (IPDGC). For a complete overview of members, acknowledgements and funding, please see http://pdgenetics.org/partners. This research was also supported in part by the Intramural Research Program of the NIH, National institute on Aging. HLA sequencing of samples from the Oslo Parkinson's disease study was supported by the Norwegian Parkinson Research Fund. PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including Abbvie, Allergan, Amathus Therapeutics, Avid Radiopharmaceuticals, Biogen, BioLegend, Bristol-Myers Squibb, Celgene, Denali Therapeutics, GE Healthcare, Genentech, GlaxoSmithKline, Golub Capital, Handl Therapeutics, Insitro, Janssen, Lilly, Lundbeck, Merck, Meso Scale Discovery, Neurocrine Biosciences, Pfizer, Piramal, Prevail Therapeutics, Roche, Sanofi Genzyme, Servier, Takeda, Teva, UCB, Verily and Voyager Therapeutics. This research has been conducted using the UK Biobank Resource under Application Number 35605.

LP is supported by a grant from the National Health Association, Norway. MSA is supported by a grant from the South-Eastern Regional Health Authority, Norway. KS is supported by a postdoctoral fellowship from the Canada First Research Excellence Fund (CFREF), awarded to McGill University for the Healthy Brains for Healthy Lives initiative (HBHL). ZGO is supported by the Fonds de recherche du Québec - Santé (FRQS) Chercheurs-boursiers award, in

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20217059; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

collaboration with Parkinson Quebec, and by the Young Investigator Award by Parkinson Canada.

Author Contributions

1) conception and design of the study – EY, AA, MT, LP, EM, ZGO

2) acquisition and analysis of data – EY, AA, MSA, LK, MAE, PS, KS, YLS, AAKS, JAR, FA, DS, MT, MKV, MN, CB, LP, EM, ZGO

3) drafting a significant portion of the manuscript or figures – EY, EM, ZGO

Conflicts of interest

The authors have no conflicts to report.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20217059; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

References

1. Poewe W, Seppi K, Tanner CM, et al. Parkinson disease. Nat Rev Dis Primers. 2017 Mar 23;3:17013.

2. Sanchez-Guajardo V, Barnum CJ, Tansey MG, Romero-Ramos M. Neuroimmunological processes in Parkinson's disease and their relation to α-synuclein: microglia as the referee between neuronal processes and peripheral immunity. ASN neuro. 2013;5(2):AN20120066.

3. Wang Q, Liu Y, Zhou J. Neuroinflammation in Parkinson's disease and its potential as therapeutic target. Translational Neurodegeneration. 2015 2015/10/12;4(1):19.

4. Tansey MG, Romero-Ramos M. Immune system responses in Parkinson's disease: Early and dynamic. Eur J Neurosci. 2019 Feb;49(3):364-83.

5. Devos D, Lebouvier T, Lardeux B, et al. Colonic inflammation in Parkinson's disease. Neurobiology of disease. 2013;50:42-8.

6. Zhang W, Wang T, Pei Z, et al. Aggregated α-synuclein activates microglia: a process leading to disease progression in Parkinson's disease. The FASEB Journal. 2005;19(6):533-42.

7. Nalls MA, Blauwendraat C, Vallerga CL, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet Neurol. 2019 Dec;18(12):1091-102.

8. Malavasi F, Deaglio S, Ferrero E, et al. CD38 and CD157 as receptors of the immune system: a bridge between innate and adaptive immunity. Mol Med. 2006 Nov-Dec;12(11-12):334-41.

9. Shiina T, Hosomichi K, Inoko H, Kulski JK. The HLA genomic loci map: expression, interaction, diversity and disease. J Hum Genet. 2009 Jan;54(1):15-39.

10. Wallings RL, Herrick MK, Tansey MG. LRRK2 at the Interface Between Peripheral and Central Immune Function in Parkinson's. Front Neurosci. 2020;14:443.

11. Ahmed I, Tamouza R, Delord M, et al. Association between Parkinson's disease and the HLA-DRB1 locus. Mov Disord. 2012 Aug;27(9):1104-10.

12. Hamza TH, Zabetian CP, Tenesa A, et al. Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. Nat Genet. 2010 Sep;42(9):781-5.

13. Hollenbach JA, Norman PJ, Creary LE, et al. A specific amino acid motif of HLA-DRB1 mediates risk and interacts with smoking history in Parkinson's disease. Proc Natl Acad Sci U S A. 2019 Apr 9;116(15):7419-24.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20217059; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

14. Wissemann WT, Hill-Burns EM, Zabetian CP, et al. Association of Parkinson disease with structural and regulatory variants in the HLA region. The American Journal of Human Genetics. 2013;93(5):984-93.

15. Bandres-Ciga S, Ahmed S, Sabir MS, et al. The Genetic Architecture of Parkinson Disease in Spain: Characterizing Population-Specific Risk, Differential Haplotype Structures, and Providing Etiologic Insight. Mov Disord. 2019 Dec;34(12):1851-63.

16. Hill-Burns EM, Factor SA, Zabetian CP, Thomson G, Payami H. Evidence for more than one Parkinson's disease-associated variant within the HLA region. PLoS One. 2011;6(11):e27109.

17. Nalls MA, Pankratz N, Lill CM, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nat Genet. 2014 Sep;46(9):989-93.

18. Simón-Sánchez J, Schulte C, Bras JM, et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nat Genet. 2009 Dec;41(12):1308-12.

19. Liu JZ, Erlich Y, Pickrell JK. Case-control association mapping by proxy using family history of disease. Nat Genet. 2017 Mar;49(3):325-31.

20. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4(1).

21. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011 Jan 7;88(1):76-82.

22. Abraham G, Inouye M. Fast principal component analysis of large-scale genome-wide data. PLoS One. 2014;9(4):e93766.

23. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190-1.

24. Zheng X, Shen J, Cox C, et al. HIBAG—HLA genotype imputation with attribute bagging. The pharmacogenomics journal. 2014;14(2):192-200.

25. Karnes JH, Shaffer CM, Bastarache L, et al. Comparison of HLA allelic imputation programs. PLoS One. 2017;12(2):e0172444.

26. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. Gigascience. 2019;8(7):giz082.

27. Kamoun M, McCullough KP, Maiers M, et al. HLA Amino Acid Polymorphisms and Kidney Allograft Survival. Transplantation. 2017 May;101(5):e170-e7.

28. Kannarkat GT, Cook DA, Lee JK, et al. Common Genetic Variant Association with Altered HLA Expression, Synergy with Pyrethroid Exposure, and Risk for Parkinson's Disease: An Observational and Case-Control Study. NPJ Parkinsons Dis. 2015;1.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20217059; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

29. Levesque S, Wilson B, Gregoria V, et al. Reactive microgliosis: extracellular micro-calpain and microglia-mediated dopaminergic neurotoxicity. Brain. 2010 Mar;133(Pt 3):808-21.

30. Sulzer D, Alcalay RN, Garretti F, et al. T cells from patients with Parkinson's disease recognize α-synuclein peptides. Nature. 2017 Jun 29;546(7660):656-61.

31. Lindestam Arlehamn CS, Dhanwani R, Pham J, et al. α-Synuclein-specific T cell reactivity is associated with preclinical and early Parkinson's disease. Nat Commun. 2020 Apr 20;11(1):1875.

32. Fayyad M, Erskine D, Majbour NK, et al. Investigating the presence of doubly phosphorylated α-synuclein at tyrosine 125 and serine 129 in idiopathic Lewy body diseases. Brain Pathol. 2020 Jul;30(4):831-43.

33. Li XY, Yang W, Li X, et al. Phosphorylated Alpha-Synuclein in Red Blood Cells as a Potential Diagnostic Biomarker for Multiple System Atrophy: A Pilot Study. Parkinsons Dis. 2020;2020:8740419.

34. Grifoni A, Moore E, Voic H, et al. Characterization of Magnitude and Antigen Specificity of HLA-DP, DQ, and DRB3/4/5 Restricted DENV-Specific CD4+ T Cell Responses. Front Immunol. 2019;10:1568.

35. Chang KH, Wu YR, Chen YC, Fung HC, Chen CM. Association of genetic variants within HLA-DR region with Parkinson's disease in Taiwan. Neurobiol Aging. 2020 Mar;87:140 e13- e18.

36. Ma ZG, Liu TW, Bo YL. HLA-DRA rs3129882 A/G polymorphism was not a risk factor for Parkinson's disease in Chinese-based populations: a meta-analysis. Int J Neurosci. 2015 Apr;125(4):241-6.

37. Sun C, Wei L, Luo F, et al. HLA-DRB1 alleles are associated with the susceptibility to sporadic Parkinson's disease in Chinese Han population. PLoS One. 2012;7(11):e48594.

38. Zhao Y, Gopalai AA, Ahmad-Annuar A, et al. Association of HLA locus variant in Parkinson's disease. Clin Genet. 2013 Nov;84(5):501-4.

39. Foo JN, Chew EGY, Chung SJ, et al. Identification of Risk Loci for Parkinson Disease in Asians and Comparison of Risk Between Asians and Europeans: A Genome-Wide Association Study. JAMA Neurol. 2020 Jun 1;77(6):746-54.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20217059; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Figure legends

Figure 1. Validation of previously associated top *HLA* locus SNP (rs112485576) in our cohort. **A)** Forest plot describing the effect size and 95% confidence interval of rs112485576 for each cohort and fixed-effect meta-analysis. **B-C)** Two LocusZoom plots highlighting the significant variants before (**B**) and after (**C**) the conditional analysis on rs112485576. Dashed lines correspond to the significance threshold. Linkage disequilibrium values are shown with respect to the most significant SNP in the locus.

Figure 2. Association of the *HLA-DRB1* alleles and location of associated amino acids. A) The location of the HLA locus, alleles and amino acids associated with Parkinson's disease in the current study. **B)** 3D model of HLA-DRB1 – HLA-DRA and the location of the 11V, 13H and 33H amino acids associated with PD (highlighted by arrows). The model was generated with PyMol v. 2.4.1 (pdb 4is6).

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20217059; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Tables

Table 1. Meta-analyses of HLA alleles association

Allele	Freq Cases	Freq Controls	OR (95% CI)	P-value[a]	Direction	HetPVal
*DQA1**03:01	0.168	0.185	0.86 (0.81-0.90)	3.54e-08	---------	0.85
*DQB1**03:02	0.182	0.199	0.87 (0.83-0.92)	7.62e-07	---------	0.58
*DRB1**04:01	0.187	0.221	0.89 (0.84-0.94)	2.82e-05	---------	0.87
*DRB1**04:04	0.068	0.082	0.84 (0.77-0.91)	8.21e-05	---------	0.85

Abbreviations. Freq Cases: Frequency of allele in patients; Freq Controls: Frequency of allele in controls; OR (95% CI): Odds ratio and 95% confidence interval; Direction: Direction of beta for each cohort; HetPVal: P-value of heterogeneity.

[a] Bonferroni correction for multiple comparisons set the threshold for statistical significance to $\alpha=3.55x10^{-4}$.

Table 2. Meta-analyses of HLA haplotype association

Haplotype	Freq Cases	Freq Controls	OR (95% CI)	P-value[a]	Direction	HetPVal
*DQA1**03:01-*DQB1**03:02	0.157	0.173	0.87 (0.82-0.93)	7.21e-05	+----+---	0.62
*DRB1**04:04-*DQA1**03:01	0.067	0.081	0.83 (0.76-0.91)	9.82e-05	+--------	0.34
*DRB1**04:01-*DQA1**03:03	0.115	0.143	0.87 (0.81-0.94)	4.96e-04	+--------	0.71

Abbreviations. Freq Cases: Frequency of allele in patients; Freq Controls: Frequency of allele in controls; OR (95% CI): Odds ratio and 95% confidence interval; Direction: Direction of beta for each cohort; HetPVal: P-value of heterogeneity.

[a] Bonferroni correction for multiple comparisons set the threshold for statistical significance to $\alpha=5.95x10^{-4}$.

Table 3. Meta-analyses of HLA amino acid changes association

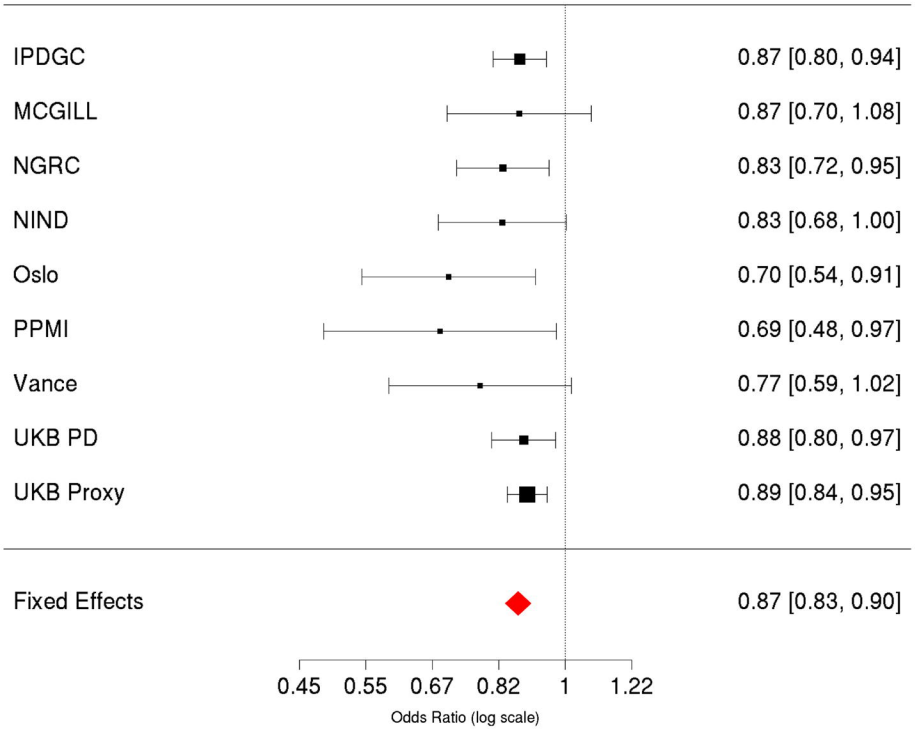
Amino Acid	Freq Cases	Freq Controls	OR (95% CI)	P-value[a]	Direction	HetPVal
DRB1 13H	0.289	0.331	0.87 (0.83-0.91)	4.32e-09	---------	0.60
DRB1 33H	0.289	0.331	0.87 (0.83-0.91)	4.32e-09	---------	0.60
DRB1 11V	0.302	0.342	0.87 (0.83-0.91)	8.22e-09	---------	0.45

Abbreviations. Freq Cases: Frequency of allele in patients; Freq Controls: Frequency of allele in controls; OR (95% CI): Odds ratio and 95% confidence interval; Direction: Direction of beta for each cohort; HetPVal: P-value of heterogeneity. P-value of heterogeneity.

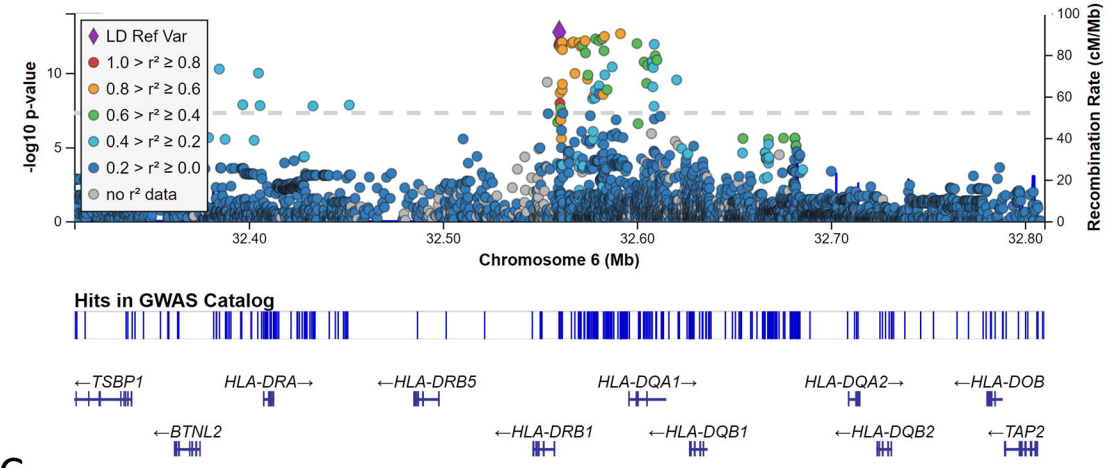
[a] Bonferroni correction for multiple comparisons set the threshold for statistical significance to $\alpha=7.86x10^{-5}$.

(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuit
It is made available under a CC-BY 4.0 International license .

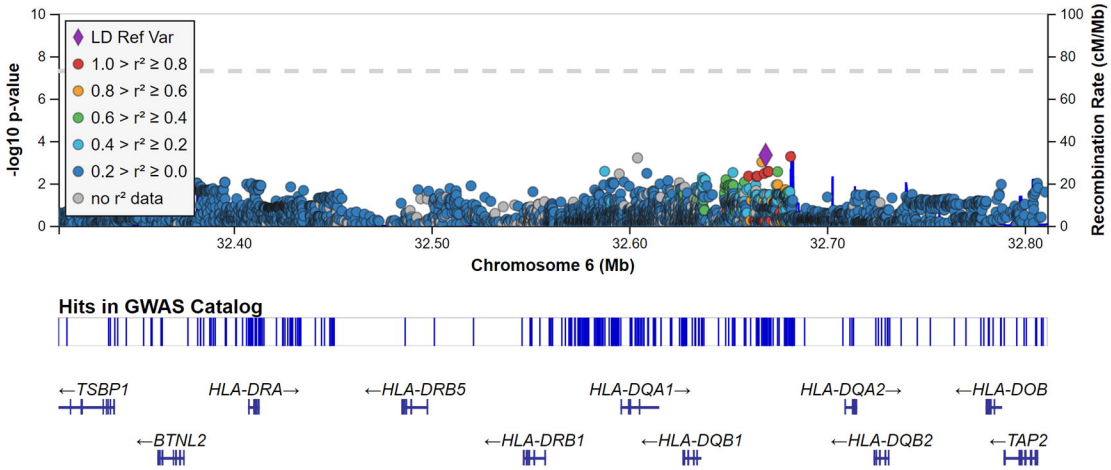
A



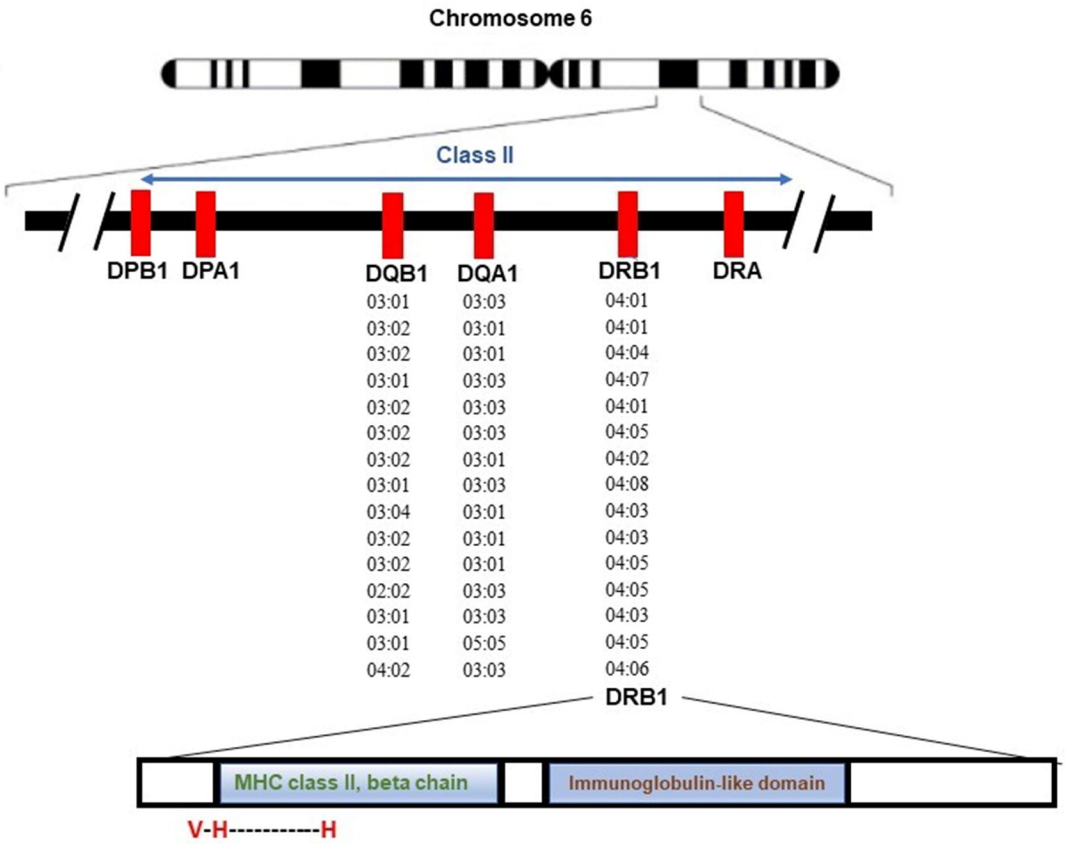
B



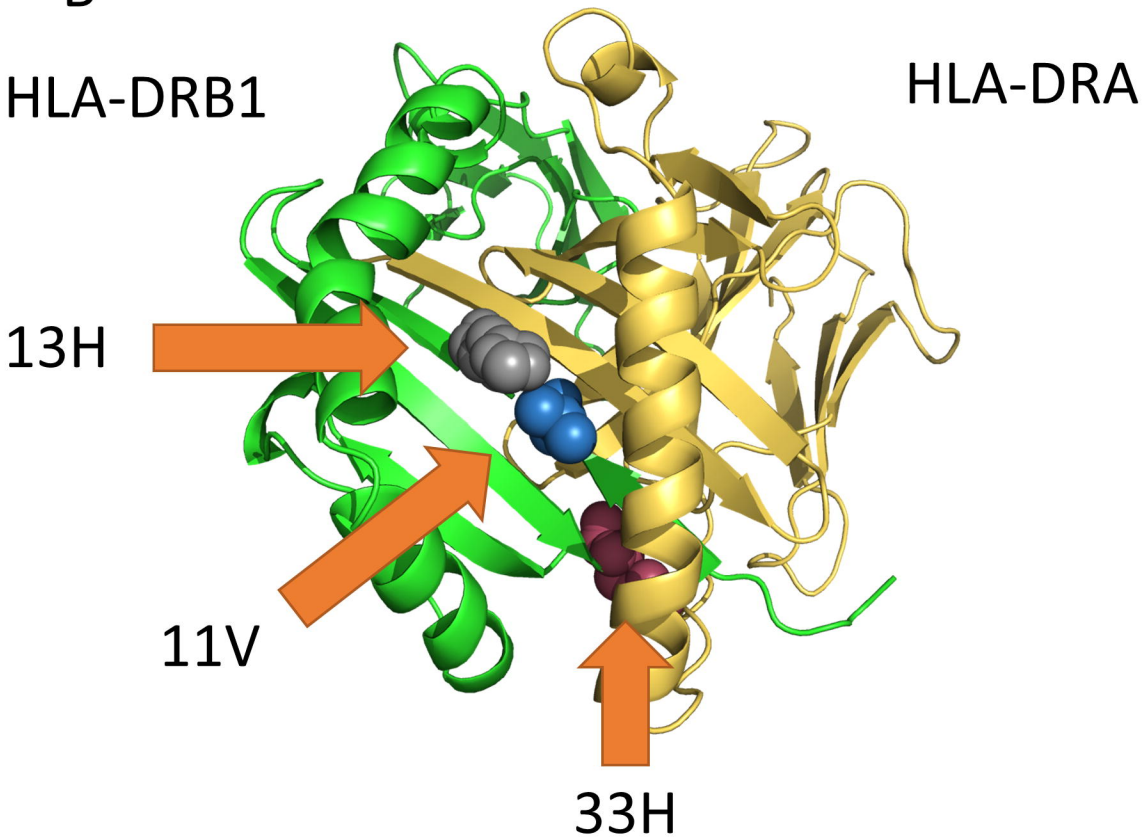
C



A


Chromosome 6
Class II
DPB1
DPA1
DQB1
DQA1
DRB1
DRA
03:01 03:03 04:01
03:02 03:01 04:01
03:02 03:01 04:04
03:01 03:03 04:07
03:02 03:03 04:01
03:02 03:03 04:05
03:02 03:01 04:02
03:01 03:03 04:08
03:04 03:01 04:03
03:02 03:01 04:03
03:02 03:01 04:05
02:02 03:03 04:05
03:01 03:03 04:03
03:01 05:05 04:05
04:02 03:03 04:06
DRB1
MHC class II, beta chain
Immunoglobulin-like domain
V-H----------H

B


HLA-DRB1
HLA-DRA
13H
11V
33H