RESEARCH ARTICLE OPEN ACCESS

Risk HLA Variants Affect the T-Cell Repertoire in Multiple Sclerosis

Melissa Sorosina, PhD, Silvia Santoro, PhD, Laura Ferrè, MD, PhD, Elisabetta Mascia, PhD, Ferdinando Clarelli, MS, Antonino Giordano, MD, Miryam Cannizzaro, MD, Moiola Lucia, MD, Vittorio Martinelli, MD, Massimo Filippi, MD, and Federica Esposito, MD, PhD

Correspondence
Dr. Esposito
esposito.federica@hsr.it

Neurol Neuroimmunol Neuroinflamm 2023;10:e200093. doi:10.1212/NXI.0000000000200093

Abstract

Background and Objectives

The major histocompatibility complex (MHC) locus has a predominant role in the genetic predisposition to multiple sclerosis (MS), with 32 associations found to be involved. We aimed to investigate the impact of MHC MS-risk alleles on T-cell repertoire in patients with MS.

Methods

We studied 161 untreated patients with relapsing-remitting MS for whom Class I and II human leukocyte antigen (HLA) alleles were inferred from whole-genome genotyping data, and T-cell receptor (TCR) CDR3 sequences were obtained through next-generation sequencing. T-cell repertoire features including diversity, public clones, and architecture were evaluated.

Results

We identified 5 MS-risk loci associated with TCR diversity: HLA-DRB1*15:01 (7.65×10^{-3}), rs9271366 (1.96×10^{-3}), rs766848979 A (1.89×10^{-2}), rs9277626 (2.95×10^{-2}), and rs11751659 (1.92×10^{-2}), with evidence of expanded clonotypes in carriers of risk alleles. Moreover, HLA-DRB1*15:01 (4.99×10^{-3}), rs9271366 (6.54×10^{-3}), rs1049079 C (4.37×10^{-2}), AA DQβ1 position −5 L (1.05×10^{-3}), and AA DQβ1 position 221 Q (9.39×10^{-4}) showed an association with the CDR3 aminoacidic sequence architecture, suggesting an impact on the antigen recognition breadth as well. Evaluating the sharing of clones across MS-risk allele carrier individuals revealed the presence of highly shared clonotypes predicted to target viral antigens, including Epstein-Barr virus.

Discussion

Our study supports the association between MHC-risk alleles and macrofeatures of the T-cell repertoire in the context of MS. Further studies are needed to understand the underlying molecular mechanisms.

From the Laboratory of Neurological Complex Disorders (M.S., S.S., L.F., E.M., F.C., A.G., M.C., F.E.), Division of Neuroscience, Institute of Experimental Neurology (INSPE), IRCCS San Raffaele Scientific Institute; Neurology and Neurorehabilitation Unit (L.F., A.G., M.C., M.L., V.M., M.F., F.E.), IRCCS San Raffaele Scientific Institute, Milan, Italy; Vita-Salute San Raffaele University (M.F.); Neurophysiology Unit (M.F.), IRCCS San Raffaele Scientific Institute; and Neuroimaging Research Unit, Division of Neuroscience, Institute of Experimental Neurology (INSPE) (P.F.), IRCCS San Raffaele Scientific Institute, Milan, Italy.

Go to Neurology.org/NN for full disclosures. Funding information is provided at the end of the article.

The Article Processing Charge was funded by the authors.

Glossary

CDR = complementarity-determining region; **CMV** = cytomegalovirus; **EBV** = Epstein-Barr virus; **EDSS** = Expanded Disability Status Scale; **gDNA** = genomic DNA; **HLA** = human leukocyte antigen; **HLAGB** = HLA genetic burden; **LevD** = Levenshtein distance; **MAF** = minor allele frequency; **MHC** = major histocompatibility complex; **MS** = multiple sclerosis; **PCA** = principal component analysis; **SNP** = single nucleotide polymorphism; **RRMS** = relapsing-remitting MS; **TCR** = T-cell receptor.

Multiple sclerosis (MS) is a progressive disorder of the CNS affecting approximately 2.3 million people worldwide.[1] In the past decade, genome-wide association studies significantly contributed to unravel the genetics of MS, highlighting a polygenic architecture,[2] with the major histocompatibility complex (MHC) locus that includes 32 associations accounting for approximately 20% of MS genetic predisposition.[2] Such risk is dominated by the HLA-DRB1*15:01, whose association with MS has been known for decades, with carriers of the HLA-DRB1*15:01 allele having a 3-fold risk to develop MS compared with noncarriers.[2]

Extensive evidence supports a key role for the adaptive immune system in the pathogenesis of MS. Autoreactive T cells are known to migrate from peripheral blood inside the CNS, where they are reactivated by antigen-presenting cells, leading to worsening of the inflammatory condition and tissue injury.[3] Each T-cell clone expresses a unique T-cell receptor (TCR) able to recognize the target peptide presented by the MHC, leading to the immune response. The majority of human TCRs are heterodimers composed of 2 chains: α and β, containing complementary determining regions (CDRs) that drive the ability of a T lymphocyte to recognize its specific target antigen. CDR3 represents the most variable region, the primary responsible for the interaction with the targets.

To ensure the ability of the immune system to recognize a large variety of peptides, each individual has a highly diverse repertoire, and this high grade of diversity, described by several dimensions as clones frequency, convergence, and repertoire richness, is guaranteed by the recombination of gene segments in lymphocytes together with random addition/deletion of nucleotides.[4] This diversity, which is estimated to exceed 10^6 sequences in humans,[5] is profoundly different across individuals and is influenced by both genetics and the immunologic encounters throughout life.

In this context, the role of MHC is essential because the interaction between the peptides bound to the MHC molecule affects the immune cell activation, tolerance, and self-antigen recognition.[4,6] In addition, MHC configuration seems to be able to shape the TCR repertoire,[7,8] potentially affecting the responsiveness of the individual adaptive immune system, including autoimmune reactions.

Recent technological advances in the field of next-generation sequencing have allowed to investigate hundreds of thousands of TCR sequences, leading to the generation of immune repertoire maps that can be seen as a singular entity. Given the importance of MHC in shaping the TCR repertoire and its role in the onset of autoimmune diseases, we aimed to investigate the impact of the 32 MHC MS-risk alleles on T-cell repertoire in patients with MS. The influence on diversity and sequence architecture have been analyzed, together with the presence of specific clonotypes that could be enriched in MS-risk allele carriers.

Methods

Patients

A total of 161 patients with MS were recruited at the IRCCS San Raffaele Hospital (Milan, Italy). Inclusion criteria were as follows: (1) diagnosis of relapsing-remitting MS (RRMS) according to Poser and McDonald criteria[9] and (2) age older than 18 years. Exclusion criteria were as follows: (1) progressive MS course, (2) treatment with disease-modifying drugs potentially affecting the TCR repertoire (interferon β, teriflunomide, dimethyl fumarate, fingolimod, natalizumab, cladribine, and immunosuppressive therapies) in the previous 90 days, (3) previous treatment with alemtuzumab or ocrelizumab to avoid any long-term impact on the TCR repertoire, (4) treatment with corticosteroids in the previous 30 days, (5) a history of infections and antimicrobial or antiviral therapy in the previous 30 days, and (6) pregnancy. Female:male ratio is 99:62, the mean age 37.2 years (SD = 9.3), the mean disease duration 5.8 years (SD = 6.6), and the median Expanded Disability Status Scale (EDSS) at sampling 1.5 (min = 0; max = 4.5). Whole blood was collected from each patient.

Standard Protocol Approvals, Registrations, and Patient Consents

The study was approved by the local ethical committee (study protocol: GR2016-01).

Genetic Quality Controls and HLA Imputation

Single nucleotide polymorphisms (SNPs) genotype data were assayed on the Illumina BeadChip Human OmniExpress-24 platform. Standard quality controls were performed: samples were discarded if discordant with their ascertained sex on X chromosome, if they had a call rate <95% or exceeded mean heterozygosity rate >3 SD, and if showing kinship according to identity by descend analysis. To account for population substructure, principal component analysis (PCA) was conducted with PLINK v1.9,[10] estimating eigenvectors on a pruned subset of SNPs in low linkage disequilibrium (pairwise $r^2 < 0.2$ window of 50 SNPs). Individuals estimated as outliers (eigenvectors values>|mean ± 6*SD|) according to the first 10 principal components were excluded. After quality controls, 10 individuals were excluded. The final dataset was composed of 151

patients with MS. SNPs with minor allele frequency (MAF) <1%, with genotypic call rate <95%, and departing from Hardy-Weinberg equilibrium ($p < 1 \times 10^{-4}$) were excluded.

Classical HLA Class I and II alleles, SNPs, and amino acid sequences were imputed using SNP2HLA,[11] based on a reference panel of 5,225 patients of European ancestry from the type 1 Diabetes Genetics Consortium. After imputation, we retained best-guess genotypes for alleles with a high-quality imputation score ($r^2 > 0.6$).

According to the most recent study from the International MS Genetics Consortium,[2] 32 associations with MS were identified within the HLA locus. In our dataset, 2 of them were excluded because of a lower imputation quality ($r^2 < 0.6$). Rs114071505 was not present in the reference panel; thus, it was retrieved after imputation on the Haplotype Reference Consortium Rel 1.1 2016 European (EUR) Panel using Eagle2.4[12] for prephasing and Minimac4[13,14] for imputation.

A final set of 30 HLA loci was included in the analyses (mean imputation quality score: 0.98; range: 0.78–1). Statistical analyses were referred to the risk alleles, also in the case of protective loci. The HLA genetic burden (HLAGB) was calculated for each individual[15] as the sum of the risk allele burden of each HLA locus calculated as the allele dose (0, 1, and 2) multiplied by the natural log of the effect size (retrieved from[2]).

TCR Sequencing

Genomic DNA (gDNA) extraction was performed with phenol-chloroform standard method or the Maxwell 16 blood DNA purification kit (Promega). Quantity and quality were determined by nanodrop (ThermoFisher) and 1% agarose gel. TCR sequencing was performed using the immunoSEQ assay (hsTCRB kit; Adaptive Biotechnology), targeting the CDR3 of the human T cell receptor beta variable gene sequences. Libraries were prepared in triplicate starting from 3,000 ng of gDNA per sample following the manufacturer's instruction. Sequencing was performed on Illumina NextSeq500 sequencer in paired-end modality (156 + 9 cycles, NextSeq500/550 Mid Output flowcells).

Data Analysis

Sequencing Data

Raw sequencing data have been uploaded and processed by the immunoSEQ Analyzer to retrieve raw demultiplexed data that were next analyzed using the R *immunarch* package.[16] CDR3 sequences representing productive rearrangements (in frame) were analyzed, while those representing nonproductive rearrangements (out of frame or with a stop codon) were filtered out.

Diversity Estimation

Repertoire diversity is a combination of 2 information: richness (the number of unique clones in a repertoire dataset) and evenness (the similarity of frequencies of the different clonotypes), and it was estimated based on the aminoacidic sequence using the inverse Simpson diversity index (INV.S), calculated within to the *immunarch* package.[16] It represents the effective number of clonotypes that is obtained when the weighted arithmetic mean is used to quantify the average proportional abundance of types in the entire dataset. High values indicate a richer and more even distribution of TCR clones, while low values indicate an enrichment of some T-cell clones. Due to the nature of diversity estimation, which is affected by the number of clones and clonotypes in the sample, and directly related to the depth of sequencing, we performed a downsampling step to overcome biases due to the difference at sequencing depth level. We decided to use 65,000 reads as threshold as the best compromise between the inclusion of the largest number of samples (samples with a lower number of reads have been excluded) and the maintenance of a high degree of correlation between the diversity estimation calculated predownsampling and postdownsampling (Spearman correlation: $p < 2.2 \times 10^{-16}$, $r = 0.9998$). This led to the exclusion, after quality controls, of 14 individuals from the analysis and to a final sample set of 137 individuals. Given the stochastic nature of the downsampling, we performed 1,000 independent runs of downsampling followed by diversity estimation. The mean value of the INV.S across the different runs was used for association test.

Public Clones

Public clones were identified on the nonsubsampled repertoire analyzing separately HLA-risk allele carriers and noncarrier individuals. A clone was defined as public if the CDR3 aminoacidic sequence is shared across at least 2 repertoires.

Clonotypes Annotation

Target prediction of relevant clonotypes was performed with TCRMatch (tools.iedb.org/tcrmatch/)[17] querying the identified CDR3 aminoacidic sequence against the Immune Epitope Database (iedb.org, updated on July 12, 2021). The conserved C/F + W motif in input CDR3 aminoacidic sequences were trimmed, and a stringency level of 0.95 was used.

Repertoire Architecture

Networks of CDR3 aminoacidic sequences repertoires were constructed on the nonsubsampled repertoire as previously described[18] based on the top 10,000 aminoacidic CDR3 sequences. Within each network, each node is a CDR3 aminoacidic sequence, and the Levenshtein distance (LevD) was calculated and used to define connections, connecting CDR3 sequences with LevD = 1 (=1 aminoacidic sequence change). LevD was calculated using the *stringdist* R package,[19] while the degree of each node, namely the number of its connections, was calculated with the *igraph* package.[20] Comparisons between networks were performed according to the percentage of connected clonotypes (meaning the percentage of clonotypes showing at least 1 connection) out of the total number of clonotypes for each individual.

Table 1 Association Results for TCR Diversity

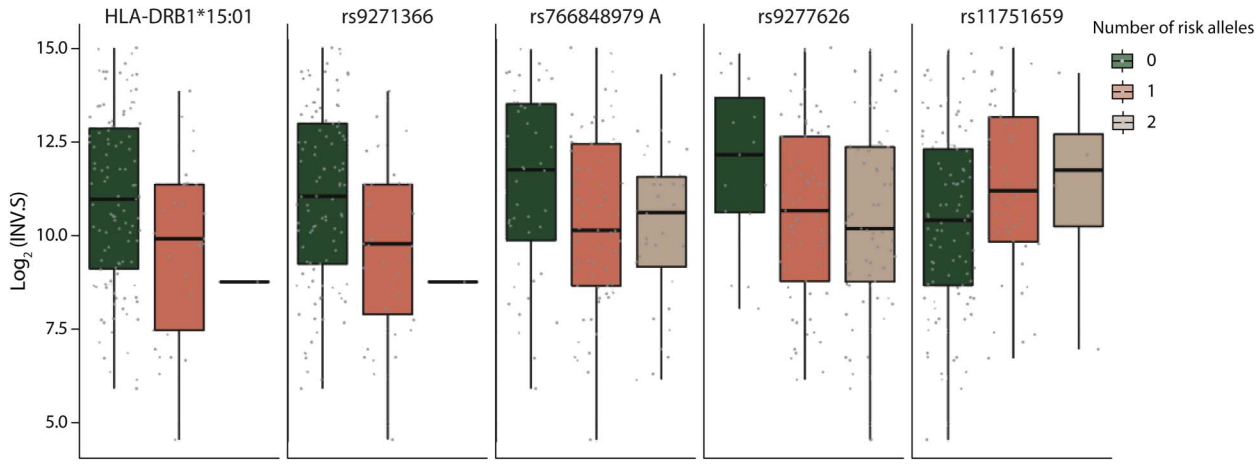
HLA loci	A1 (risk allele)	A2	OR	*p* Value	Beta	
HLA-DRB1*15:01	Present	Absent	2.90	7.65×10^{-3}	−1.11	a
rs1071743 G°	Absent	Present	1.45	0.91	−0.04	
rs3097671	C	G	1.34	0.36	0.29	
rs67476479 CAA	Present	Absent	1.32	0.76	−0.10	
rs2844482	A	G	1.35	0.81	−0.08	
rs9266629°	T	C	1.22	0.85	−0.06	
HLA-B*38:01°	Absent	Present	2.08	0.49	−0.57	
rs4081559	T	C	1.31	0.39	0.28	
HLA-DRB1*13:03	Present	Absent	1.79	0.13	1.37	
rs7454108	C	T	1.25	0.41	−0.48	
AA B position 45 TK	Present	Absent	1.13	0.69	−0.15	
AA DQβ1 position −5 L	Present	Absent	1.24	0.99	0.00	
AA DQα1 position 130 S	Present	Absent	1.48	0.63	−0.25	
rs11751659	G	A	1.17	1.42×10^{-2}	0.84	a
rs114071505°	G	C	1.28	0.40	0.60	
rs9271366	G	A	1.57	1.96×10^{-3}	−1.22	a
HLA-B*52:01°	Absent	Present	2.26	0.34	0.85	
rs3135024	C	T	1.16	0.31	0.31	
rs2523500°	T	C	1.09	0.40	−0.24	
rs1049079 C°	Absent	Present	1.14	0.68	0.31	
rs10093°	C	G	1.14	0.26	0.33	
rs766848979 A°	Absent	Present	1.19	1.89×10^{-2}	−0.58	a
HLA-B*35:03	Present	Absent	1.33	0.47	−0.57	
rs9277626°	A	G	1.09	2.95×10^{-2}	−0.62	a
HLA-DQA1*06	Present	Absent	1.85	NA	NA	
rs2229092	C	A	1.17	0.67	−0.18	
rs3819292	A	C	1.09	0.57	−0.18	
rs17493811°	C	G	1.22	0.27	0.99	
rs3093982	T	C	1.11	0.71	0.15	
AA DQβ1 position 221 Q°	Absent	Present	1.46	0.72	0.09	

Abbreviations: A1 = reference and risk allele; A2 = alternative allele; Beta = regression coefficient; NA = impossible to calculate; OR = odds ratio retrieved from the IMSGC paper[2] referring to the association with MS; *p* = *p* value of association derived from linear regression model including sex, age, and PC1 as covariates.
[a] Significant association ($p < 0.05$, Bonferroni correction threshold = $1.67*10^{-3}$). °: protective alleles according to the most recent study from the IMSGC,[2] where rs1071743 G represents the HLA-A*02:01 effect; for these alleles, statistics (OR, *p*, and Beta) refer to the indicated risk allele.

Statistical Analyses

p Values of comparisons were calculated according to the Wilcoxon test, Student *t* test, or regression analyses as appropriate, based on the type of studied variable and distribution. Allelic association was analyzed under additive model of inheritance, within linear regression framework using PLINK v.1.9.[21] Sex, age, EDSS at sampling, number of combined unique active lesions at sampling, and number of relapses 1 year before sampling were considered as covariates and included in the models when appropriate. To account for

Figure 1 Associations Between MS-Risk HLA Loci and Diversity



The $\log_2$ (INV.S) is plotted against the number of MS-risk alleles in HLA-DRB1*15:01, rs9271366, rs766848979 A, rs9277626, and rs11751659. Representation is in boxplot format (the Tukey method) with each dot representing 1 individual. MS = multiple sclerosis.

population genetic structure, the first eigenvector derived from the PCA was included in the model. According to the exploratory nature of this study, we set significance at nominal level, alpha = 0.05.

Healthy Participants Dataset

TCRβ repertoire sequence data and HLA typing information of 666 healthy individuals derived from the studies conducted by Emerson et al.[22] and De Witt et al.[23] were analyzed. Sequencing data were downloaded from the Adaptive biotechnologies website (clients.adaptivebiotech.com/pub/Emerson-2017-NatGen) while information on HLA alleles was downloaded from Zenodo repository (dx.doi.org/10. 5,281/zenodo.1248193). HLA alleles were indicated as "present" or "not present." Due to the batch effect between the healthy and MS datasets, a direct comparison between the 2 groups was not possible.

Data Availability

Anonymized data not published within this article are available from the corresponding author on reasonable request.

Results

TCR Sequencing Metrics

High-quality sequencing data were produced, with a mean of 88.6% (min-max: 83.3%–92.5%) of bases with a Qscore>30 (inferred base call accuracy >99.9%) and a mean of 89.7% of clusters passing the quality filter (min-max: 84.6%–91.25%) across the sequencing runs. A total of 21,791,911 rearrangements were identified. Among these, 17,727,335 were classified as productive, corresponding to functional TCRs (thus to single clones), with a mean of 110,107 (range: 33,841–289,477). Clonotypes, which are represented by unique sequences, were 12,307,813 (mean = 76,446.04; range: 29,229–144,391).

MS-Associated HLA Risk Alleles Affect the TCR Diversity in Patients With MS

A total of 6,768,090 clonotypes (mean = 49,402.11; range=29,289–63,612) were identified. Due to the skewed distribution of the INV.S index (eFigure 1, links.lww.com/NXI/A804), we used the log-transformed values for the following analyses.

Among all the tested clinical and demographic features, age showed an association with $\log_2$(INV.S) distribution (Wilcoxon signed-rank test, $p = 1.89 \times 10^{-7}$) while sex showed a marginal association (χ^2 test, $p = 0.08$), as already previously reported for healthy individuals,[24-27] and were included as covariates in the model. No association was found for the other clinical baseline features ($p > 0.05$).

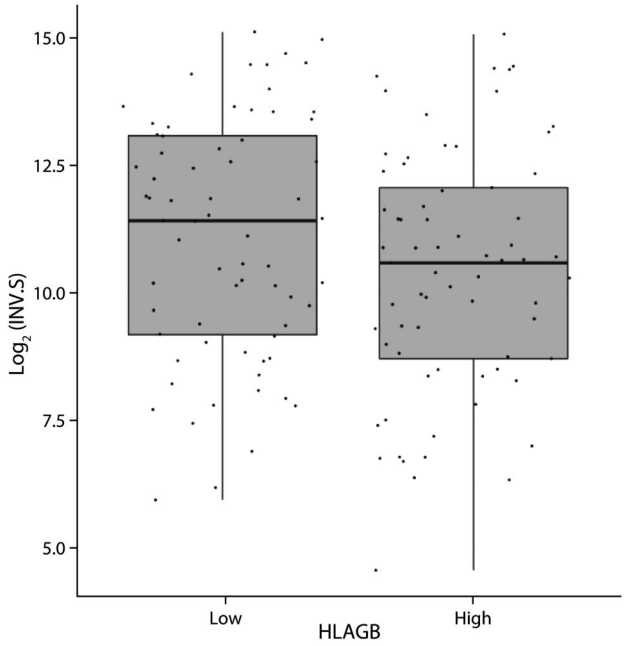
As summarized in Table 1, 5 loci were significantly associated with the $\log_2$ (INV.S), with individuals carrying risk alleles in HLA-DRB1*15:01, rs9271366, rs766848979 A, or rs9277626 showing a lower diversity; conversely, carriers of risk alleles in rs11751659 have shown a higher diversity (Figure 1).

Using the HLAGB as a measure of the cumulative genetic burden of the entire MS-associated HLA locus, after stratification of each individual in "high HLAGB" and "low HLAGB" based on the median value of the entire cohort, we observed a lower diversity in the "high HLAGB" group (linear regression $p = 0.022$, age, sex, and PC1 included in the model as covariates), as shown in Figure 2.

Individuals Carrying MS HLA Risk Variants Show Polarization of Clonotypes

To test whether the difference in diversity was explained by polarization of specific clonotypes, we evaluated the number of clonotypes required to occupy 10% of the entire

Figure 2 Associations Between HLAGB and Diversity



The $\log_2$ (INV.S) is plotted against dichotomized HLAGB. "Low" and "High" were defined based on the median value of the entire analyzed MS dataset. Representation is in boxplot format (the Tukey method) with each dot representing 1 individual. HLAGB = HLA genetic burden; MS = multiple sclerosis.

repertoire-stratifying patients based on their genotype in the 4 associated HLA loci.

As expected, as shown in eFigure 2, links.lww.com/NXI/A804, we observed a trend of lower numbers of clonotypes representing the 10% of the entire repertoire in individuals carrying the risk allele in HLA-DRB1*15:01, rs9271366, rs766848979 A, or rs9277626 ($p = 0.09$, $p = 0.034$, $p = 0.016$, and $p = 0.14$, respectively, according to regression analysis, including age, sex, and PC1 as covariates), while for rs11751659, the trend is opposite ($p = 0.033$ according to regression analysis, including age, sex and PC1 as covariates), in line with what was observed for the diversity measure.

To better explain these findings, we evaluated the clonal space homeostasis, meaning the proportion of repertoire occupied by clonal groups with specific abundances. According to this analysis, when stratifying clonotypes based on their specific frequencies (small: 1×10^{-5}<frequency<1×10^{-4}; medium: 1×10^{-4}<frequency<0.001; large: 0.001 < frequency<0.01; hyperexpanded: frequency >0.01), we observed that the clonotypes with the greatest difference were the large and hyperexpanded clonotypes subsets, with individuals carrying the risk allele in HLA-DRB1*15:01, rs9271366, rs766848979 A, or s9277626 showing a higher number of large and hyperexpanded clonotypes and a minor number of less frequent clonotypes (frequency<1×10^{-4}) (eFigure 3, links.lww.com/NXI/A804). Regarding rs11751659, we did not observe a clear pattern of difference across the different subsets.

Public Clones in Risk-Allele Carriers Show an Enrichment of Species Involved in Viral Infections

Aimed to identify clonotypes that are more frequent in risk-allele carrier individuals, we investigated public clones, thus those clonotypes that are shared across at least 2 individuals carrying the MS-risk allele in HLA-DRB1*15:01, rs9271366, rs766848979 A, rs9277626, or rs11751659. Given the rarity in our cohort of homozygous individuals for the risk allele in HLA-DRB1*15:01, rs9271366, and rs11751659, we decided to jointly analyze as carrier individuals that are homozygous or heterozygous for the risk allele (n = 36, 42, and 47 individuals for HLA-DRB1*15:01, rs9271366, and rs11751659 respectively), while for rs9277626 and rs766848979 A, we analyzed as carriers only homozygous individuals (n = 68 and n = 37 for rs9277626 and rs766848979 A, respectively) vs individuals homozygous for the nonrisk allele (n = 17 and n = 45 for rs9277626 and rs766848979 A, respectively). Among the public clones identified in risk allele carriers that are not present, not even 1 copy, in noncarriers (specifically, 40,255 clonotypes for HLA-DRB1*15:01, 50,907 clonotypes for rs9271366, 78,617 clonotypes for rs766848979 A, 287,214 clonotypes for rs9277626, and 95,117 clonotypes for rs11751659), we selected the top 10 shared clonotypes (eTable 1, links.lww.com/NXI/A805) for each group to evaluate their potential targets according to TCRMatch.[17] Although arbitrary, this threshold allowed us to focus on highly shared clonotypes (15%–36% of the carrier group) that are completely absent in the noncarrier group. Target prediction analysis of these clonotypes reveals the presence of several clonotypes that are potentially able to target viral antigens, including Epstein-Barr virus (EBV, 17% of the analyzed clonotypes), which is a known risk factor of MS[28] (eTable 2, links.lww.com/NXI/A806).

MS-Associated HLA Risk Alleles Affect the Immune Architecture

The similarity landscape of CDR3 amino acid sequences constitutes the clonal architecture of the immune repertoire and reflects its antigen recognition breadth: if all immune receptor sequences are similar, then the recognition breadth would be small, while a higher number of dissimilar sequences means that a larger number of antigens may be recognized. To evaluate whether MS risk alleles affect also this complementary layer of information, we constructed a network of CDR3 aminoacidic sequences for each individual.

Age was found to be associated with the percentage of connected CDR3 aminoacidic sequences ($p = 0.0135$) and was included as a covariate in the statistical model, together with sex, PC1, and the number of productive clones to take into account the different number of clones across samples.

Five loci showed a significant association with the percentage of connected CDR3 aminoacidic sequences (Table 2), with individuals carrying risk alleles in HLA-DRB1*15:01, rs9271366, or rs1049079 C showing a lower degree of connectivity, while carriers of the risk allele in AA DQB1 position −5 L or AA

Table 2 Association Results for the Percentage of Connected CDR3 Sequences

HLA loci	A1 (risk allele)	A2	OR	*p* Value	Beta	
HLA-DRB1*15:01	Present	Absent	2.90	4.99×10^{-3}	-8.32×10^{-3}	a
rs1071743 G°	Absent	Present	1.45	0.43	1.88×10^{-3}	
rs3097671	C	G	1.34	0.14	3.50×10^{-3}	
rs67476479 CAA	Present	Absent	1.32	0.56	-1.46×10^{-3}	
rs2844482	A	G	1.35	0.07	4.22×10^{-3}	
rs9266629°	T	C	1.22	0.32	2.53×10^{-3}	
HLA-B*38:01°	Absent	Present	2.08	0.87	-1.02×10^{-3}	
rs4081559	T	C	1.31	0.73	8.15×10^{-4}	
HLA-DRB1*13:03	Present	Absent	1.79	0.37	-5.72×10^{-3}	
rs7454108	C	T	1.25	0.71	-1.48×10^{-3}	
AA B position 45 TK	Present	Absent	1.13	0.69	1.09×10^{-3}	
AA DQβ1 position −5 L	Present	Absent	1.24	1.05×10^{-3}	6.12×10^{-3}	a
AA DQα1 position 130 S	Present	Absent	1.48	0.72	1.46×10^{-3}	
rs11751659	G	A	1.17	0.85	-4.76×10^{-4}	
rs114071505°	G	C	1.28	0.31	-5.26×10^{-3}	
rs9271366	G	A	1.57	6.54×10^{-3}	-7.77×10^{-3}	a
HLA-B*52:01°	Absent	Present	2.26	0.37	6.05×10^{-3}	
rs3135024	C	T	1.16	0.18	2.92×10^{-3}	
rs2523500°	T	C	1.09	0.38	1.87×10^{-3}	
rs1049079 C°	Absent	Present	1.14	4.37×10^{-2}	−0.0108	a
rs10093°	C	G	1.14	0.90	2.60×10^{-4}	
rs766848979 A°	Absent	Present	1.19	0.07	-3.37×10^{-3}	
HLA-B*35:03	Present	Absent	1.33	0.79	-1.62×10^{-3}	
rs9277626°	A	G	1.09	0.82	4.60×10^{-4}	
HLA-DQA1*06	Present	Absent	1.85	NA	NA	
rs2229092	C	A	1.17	0.92	3.27×10^{-4}	
rs3819292	A	C	1.09	0.42	1.80×10^{-3}	
rs17493811°	C	G	1.22	0.54	3.85×10^{-3}	
rs3093982	T	C	1.11	0.16	-4.03×10^{-3}	
AA DQβ1 position 221 Q°	Absent	Present	1.46	9.39×10^{-4}	6.21×10^{-3}	a

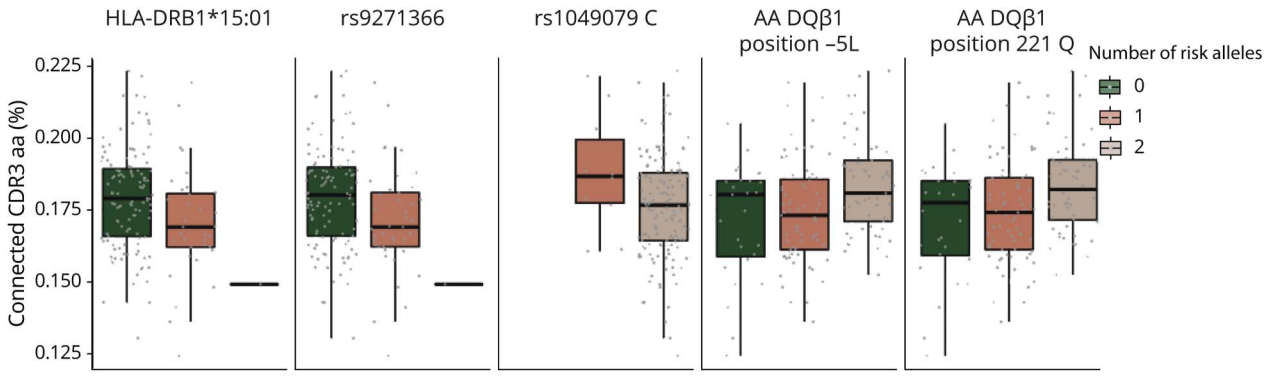
Abbreviations: A1 = reference and risk allele; A2 = alternative allele; Beta = regression coefficient; NA = impossible to calculate; OR = odds ratio retrieved from the IMSGC paper[2] referring to the association with MS; *p* = *p* value of association derived from linear regression model including age, sex, PC1, and the number of productive clones as covariates.
[a] Significant association ($p < 0.05$, Bonferroni correction threshold = $1.67*10^{-3}$). °: protective alleles according to the most recent study from the IMSGC,[2] where rs1071743 G represents the HLA-A*02:01 effect; for these alleles, statistics (OR, *p*, and Beta) refer to the indicated risk allele.

DQB1 position 221 Q showing a higher degree of connectivity (Figure 3). The cumulative HLAGB did not show any association with the degree of connectivity (linear regression $p = 0.98$, including age, sex, and the number of productive clones included as covariates).

Association of HLA-DRB1*15:01 With Immune Repertoire Features Occurs in Patients With MS Only

To answer the question of whether the correlation observed in patients with MS is specific to an MS environment or it is also

Figure 3 Associations Between MS-Risk HLA Loci and the Percentage of Connected CDR3 Aminoacidic Sequences



The percentage of connected CDR3 aminoacidic sequences is plotted against the number of MS-risk alleles in (A) HLA-DRB1*15:01, (B) rs9271366, (C) rs1049079 C, (D) AA DQB1 position −5 L, and (E) AA DQB1 position 221 Q. Representation is in boxplot format (the Tukey method) with each dot representing 1 individual. MS = multiple sclerosis.

present in nondisease conditions, we took advantage of a large dataset of healthy individuals[22] to test the observed association also in a healthy population. Among the identified associated HLA loci, data were available only for HLA-DRB1*15:01 (630 individuals), which was then tested for association with immune repertoire metrics.

For the diversity subanalysis, to ensure consistency with the MS patient dataset, we performed a subsampling to 65,000 clones as performed for patients, resulting in the exclusion of 16 samples. Conversely to what observed for patients with MS, we did not detect any association between genotype in HLA-DRB1*15:01 and $\log_2$ (INV.S) (dominant model, linear regression $p = 0.45$, Beta = 0.16, including sex, age, and cytomegalovirus [CMV] infection status as covariates). Similarly, including only the 367 Caucasian individuals (as for the MS dataset), no association was found (dominant model, linear regression $p = 0.63$, Beta = 0.12, including sex, age, and CMV infection status as covariates).

Regarding the architecture metrics, the entire dataset was included in the subanalysis with no downsampling step, as performed for patients with MS. No differences for degree of connectivity were observed across the 2 genotype groups (linear regression $p = 0.83$, Beta = −0.0006, including age, sex, known CMV infection and the number of productive clones as covariates). Analyzing the Caucasian individuals alone, no association was found (linear regression $p = 0.91$, Beta = −0.0004, including age, known CMV infection, and the number of productive clones as covariates).

Discussion

The importance of the HLA region in MS is well established, with evidence that MHC genetic variants affect the susceptibility and expression of the disease.[2,29-32] The genetic influence exerted by the HLA region on MS risk is complex and intricate. HLA-DRB1*15:01 was the first risk allele associated, while nowadays several risk or protective variants are known. The locus with the highest impact is the HLA-DRB1, but also other loci, including the HLA-A, with the HLA-A*02:01 allele being the first protective allele associated with MS, HLA-B, harboring 6 independent effects, nonclassical HLA, and non-HLA genes in the MHC region are involved.[2]

Several mechanisms could be hypothesized to explain the contribution of HLA to MS. The most obvious is that disease-related HLA Class II molecules could preferentially present self-peptides, stimulating autoreactive $CD4^+$ T lymphocytes and favoring their migration into the CNS. On the contrary, HLA contributes also to the modeling of the immune system, with MHC type II showing self-antigens in the thymus, and through the activation and expansion of peripheral T lymphocytes necessary for the immune response triggered by antigens. In this context, we hypothesized that the genetic association between the HLA region and MS may be related to the influence to the general structure of the immune repertoire, as already shown,[7] creating an environment favoring the individual at risk to develop MS.

In this study, we evaluated the association between MS-risk HLA loci and 2 main features of the immunologic repertoire: diversity and architecture, representing 2 complementary and alternative ways to define the extent of the immune system to recognize antigens and react to immune triggers.[18,33] Our results support our hypothesis because we observed an association between HLA alleles and immune repertoire diversity and with the degree of connectivity of the CDR3 aminoacidic sequences. This association seems to be allele specific, with alleles showing different directions of effect, although a mild relationship with the diversity index was observed analyzing the cumulative genetic risk carried by the entire HLA region (Figure 2, $p = 0.022$), with patients carrying a higher genetic burden showing also a lower diversity. No clear pattern was observed between protective and risk alleles. Given the overall heterogeneity, the interpretation of results could be trivial, and each signal could represent a different unrelated underlying mechanism, similarly to what

has been proposed for the protective effect of HLA Class I alleles that seem to be implicated in the type I interferon signaling instead of the antigen-presenting function.[34]

The lower diversity observed in individuals carrying specific HLA alleles may indicate a polarization of clonotypes, suggesting that their presence may favor the reactivity toward a small number of antigens. Indeed, the analysis of clonal proportion distribution across individuals carrying or not carrying the MS risk alleles supports this (eFigures 2, eFigure3, links.lww.com/NXI/A804), in line with previous reports showing that T-cell expansion of specific clonotypes exist in patients with MS.[35-37] At the same time, this polarization does not seem to be absolute, contrary to what is observed in pathologies driven by viral infections,[38] but mixed. An alternative point of view could be the possibility that our results reflect a different degree of diversity in T-lymphocyte subtypes.[26] Indeed, a limited diversity in specific subtypes of T lymphocytes, such as T-regulatory cells, was suggested to be a risk factor for autoimmune diseases.[39] The analysis of specific T-cell populations would be useful to elucidate this.

Hypothesizing that relevant clonotypes could be shared across risk-allele carriers, we identified shared clonotypes in the carrier group not present in noncarriers. Although this may be a conservative approach, given also the size of the 2 sets of patients, it allowed us to prioritize 15 clonotypes that deserve further investigations (eTable 1, links.lww.com/NXI/A805). Although it has not been possible to identify their molecular target based on currently available databases, it is very interesting that several clonotypes are predicted to potentially target EBV (eTable 2, links.lww.com/NXI/A806), supporting its involvement in MS susceptibility.[28] However, it is also important to underline that this approach is not free of potential intrinsic biases; therefore, experiments specifically aimed at identifying their target will help to clarify this aspect.

Contrary to what has been observed for other autoimmune diseases (such as lupus erythematosus and rheumatoid arthritis[40,41]), previous reports showed that patients with MS have greater diversity than healthy individuals,[38,42] suggesting that having a higher diversity may favor the development of the disease. Although it is not known whether the different immune structure is generated as a consequence of the disease state or it represents an intrinsic factor involved in the pathogenesis, we could not exclude that certain HLA alleles, such as rs11751659, may act in this direction, leading the immune repertoire to be more prone to recognize more antigens that will be polarized after the autoimmune insult. This can also explain the heterogeneity observed across MHC alleles, with some risk alleles pointing to a lower diversity and other alleles being associated with a higher diversity, even if in less extent.

Given that the immune repertoire is extremely dependent on the encounters occurring during the whole life, our results most likely reflect the interplay between genetic and environmental factors contributing to the final repertoire, instead of a genetically driven effect only. Given its importance in MS,[28] the EBV infection might represent the most relevant factor. Moreover, given the fact that we did not observe the association between HLA-DRB1*15:01 and repertoire characteristics in the healthy dataset, we can hypothesize that the impact of genetics becomes evident only in the context of MS risk environment, characterized by the presence of additional genetic factors and environmental ones. It is important to note that the 2 cohorts (MS and healthy individuals) are composed by geographically and ethnically distinct individuals; the missing replication in the healthy dataset could be due to their exposure to a different environment and not to the disease status.

It is also noteworthy to note the presence of the HLA-DRB1*15:01 allele, the primary genetic factor associated with MS,[2] among the alleles found to be associated with the immune repertoire features. Several studies have been conducted to understand, from a functional point of view, the involvement of the DRB1*15:01 allele in MS, showing that it is potentially involved also in the shaping of T-cell repertoire.[43] As an example, autoreactive T cells for myelin basic protein (MBP) were found to be able to recognize epitopes in the context of DR2b (the heterodimer of HLA-DRα with the HLA-DRβ chain of HLA-DRB1*15:01)[44,45]; another evidence shows that, in humanized transgenic mouse models, MBP-specific TCRs restricted by DR2b lead to spontaneous experimental autoimmune encephalomyelitis, the mouse model of MS.[46] It is of interest that data suggest that the DR2b itself could serve as a source of antigens[47] and DR2b could be involved in the autoproliferation of T cells, even in the absence of antigenic stimulus,[48] supporting a role in peripheral homeostasis and maintenance. In this context, our results support the involvement of the HLA-DRB1*15:01 allele in the perturbation of T-cell repertoire dynamics in MS. We could hypothesize that such allele affects the T-cell repertoire toward a narrower one, enriched of T cells with a higher affinity for specific peptides, more prone to be activated in the presence of environmental risk factors. Similar mechanisms can also be hypothesized for the other MS-risk HLA alleles that we found to be associated. It is also correct to point out that given the high linkage disequilibrium existing within the HLA region, it would be of interest to test the observed results in a larger dataset, to be able to discern the true signal between HLA-DRB1*15:01 and rs9271366, which in our dataset show a relatively high linkage disequilibrium ($r^2 = 0.78$), suggesting that they probably reflect the same underlying signal.

This study has some drawbacks that could be improved. First, we analyzed the entire repertoire of T lymphocytes, without focusing on specific cell subtypes. This may have masked specific effects of the different subtypes. Moreover, starting from DNA extracted from limited volume of peripheral blood, only a small fraction of the entire set of T lymphocytes has been studied; thus, we may lose information about clones that were not randomly sampled. However, the study of macrofeatures makes it possible to partially overcome this problem because it represents a general measure of the repertoire that

is maintained even after technical (i.e. blood sample volume) or postanalysis subsampling. It is also important to note that most of the associations were significant at nominal level; thus, a replication in a larger dataset is advised. On the contrary, a main strength of this study comes from the large cohort of patients with MS, enrolled in the same MS center and composed only by untreated individuals, reducing potential biases related to drug's effect on the immune structure.[49,50]

In conclusion, our study supports the association between HLA risk alleles and macrofeatures of the T-cell repertoire in the context of MS. We have identified clones of interest, which are highly shared between carriers of risk alleles. Further studies are needed to better delineate this association and understand the underlying molecular mechanisms.

Study Funding

Italian Ministry of Health (project code: GR-2016-02363997).

Disclosure

The authors report no relevant disclosures. Go to Neurology.org/NN for full disclosures.

Publication History

Received by *Neurology: Neuroimmunology & Neuroinflammation* August 2, 2022. Accepted in final form December 14, 2022. Submitted and externally peer reviewed. The handling editor was Deputy Editor Scott S. Zamvil, MD, PhD, FAAN.

Appendix Authors

Name	Location	Contribution
Melissa Sorosina, PhD	Laboratory of Neurological complex disorders, Division of Neuroscience, Institute of Experimental Neurology (INSPE), IRCCS San Raffaele Scientific Institute, Milan, Italy	Drafting/revision of the article for content, including medical writing for content; major role in the acquisition of data; study concept or design; and analysis or interpretation of data
Silvia Santoro, PhD	Laboratory of Neurological complex disorders, Division of Neuroscience, Institute of Experimental Neurology (INSPE), IRCCS San Raffaele Scientific Institute, Milan, Italy	Major role in the acquisition of data; analysis or interpretation of data
Laura Ferrè, MD, PhD	Laboratory of Neurological complex disorders, Division of Neuroscience, Institute of Experimental Neurology (INSPE), IRCCS San Raffaele Scientific Institute, Milan, Italy; Neurology and Neurorehabilitation Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy	Major role in the acquisition of data; analysis or interpretation of data
Elisabetta Mascia, PhD	Laboratory of Neurological complex disorders, Division of Neuroscience, Institute of Experimental Neurology (INSPE), IRCCS San Raffaele Scientific Institute, Milan, Italy	Major role in the acquisition of data
Ferdinando Clarelli, MS	Laboratory of Neurological complex disorders, Division of Neuroscience, Institute of Experimental Neurology (INSPE), IRCCS San Raffaele Scientific Institute, Milan, Italy	Major role in the acquisition of data; analysis or interpretation of data
Antonino Giordano, MD	Laboratory of Neurological complex disorders, Division of Neuroscience, Institute of Experimental Neurology (INSPE), IRCCS San Raffaele Scientific Institute, Milan, Italy; Neurology and Neurorehabilitation Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy	Major role in the acquisition of data
Miryam Cannizzaro, MD	Laboratory of Neurological complex disorders, Division of Neuroscience, Institute of Experimental Neurology (INSPE), IRCCS San Raffaele Scientific Institute, Milan, Italy; Neurology and Neurorehabilitation Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy	Major role in the acquisition of data
Moiola Lucia, MD	Neurology and Neurorehabilitation Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy	Major role in the acquisition of data
Vittorio Martinelli, MD	Neurology and Neurorehabilitation Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy	Major role in the acquisition of data
Massimo Filippi, Professor, MD	Neurology and Neurorehabilitation Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy; Vita-Salute San Raffaele University, Milan, Italy; Neurophysiology Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy; Neuroimaging Research Unit, Division of Neuroscience, Institute of Experimental Neurology (INSPE), IRCCS San Raffaele Scientific Institute, Milan, Italy	Major role in the acquisition of data
Federica Esposito, MD, PhD	Laboratory of Neurological complex disorders, Division of Neuroscience, Institute of Experimental Neurology (INSPE), IRCCS San Raffaele Scientific Institute, Milan, Italy; Neurology and Neurorehabilitation Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy	Drafting/revision of the article for content, including medical writing for content; major role in the acquisition of data; study concept or design; and analysis or interpretation of data

References

1. Browne P, Chandraratna D, Angood C, et al. Atlas of Multiple Sclerosis 2013: a growing global problem with widespread inequity. *Neurology.* 2014;83(11):1022-1024. doi. 10.1212/WNL.0000000000000768.
2. International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science.* 2019;365(6460):eaav7188. doi. 10.1126/science.aav7188.

3. Dendrou CA, Fugger L, Friese MA. Immunopathology of multiple sclerosis. *Nat Rev Immunol.* 2015;15(9):545-558. doi. 10.1038/nri3871.
4. Rudolph MG, Stanfield RL, Wilson IA. How TCRs bind MHCs, peptides, and coreceptors. *Annu Rev Immunol.* 2006;24(1):419-466. doi. 10.1146/annurev.immunol.23.021704.115658.
5. Robins HS, Campregher PV, Srivastava SK, et al. Comprehensive assessment of T-cell receptor β-chain diversity in αβ T cells. *Blood.* 2009;114(19):4099-4107. doi. 10.1182/blood-2009-04-217604.
6. Riedhammer C, Weissert R. Antigen presentation, autoantigens, and immune regulation in multiple sclerosis and other autoimmune diseases. *Front Immunol.* 2015;6:322. doi. 10.3389/fimmu.2015.00322.
7. Sharon E, Sibener LV, Battle A, Fraser HB, Garcia KC, Pritchard JK. Genetic variation in MHC proteins is associated with T cell receptor expression biases. *Nat Genet.* 2016;48(9):995-1002. doi. 10.1038/ng.3625.
8. Logunova NN, Kriukova VV, Shelyakin PV, et al. MHC-II alleles shape the CDR3 repertoires of conventional and regulatory naïve CD4+ T cells. *Proc Natl Acad Sci U S A.* 2020;117(24):13659-13669. doi. 10.1073/pnas.2003170117.
9. Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol.* 2018;17(2):162-173. doi. 10.1016/S1474-4422(17)30470-2.
10. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559-575. doi. 10.1086/519795.
11. Jia X, Han B, Onengut-Gumuscu S, et al. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One.* 2013;8(6):e64683. doi. 10.1371/journal.pone.0064683.
12. Loh P-R, Danecek P, Palamara PF, et al. Reference-based phasing using the haplotype reference Consortium panel. *Nat Genet.* 2016;48(11):1443-1448. doi. 10.1038/ng.3679.
13. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet.* 2012;44(8):955-959. doi. 10.1038/ng.2354.
14. Fuchsberger C, Abecasis GR, Hinds DA. minimac2: faster genotype imputation. *Bioinformatics.* 2015;31(5):782-784. doi. 10.1093/bioinformatics/btu704.
15. De Jager PL, Chibnik LB, Cui J, et al. Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. *Lancet Neurol.* 2009;8(12):1111-1119. doi. 10.1016/S1474-4422(09)70275-3.
16. ImmunoMind Team. *Immunarch: An R Package for Painless Bioinformatics Analysis of T-Cell and B-Cell Immune Repertoires*. Zenodo; 2019. doi.org/10.5281/zenodo.3367200.
17. Chronister WD, Crinklaw A, Mahajan S, et al. TCRMatch: predicting T-cell receptor specificity based on sequence similarity to previously characterized receptors. *Front Immunol.* 2021;12:640725. doi. 10.3389/fimmu.2021.640725.
18. Madi A, Poran A, Shifrut E, et al. T cell receptor repertoires of mice and humans are clustered in similarity networks around conserved public CDR3 sequences. *Elife.* 2017;6:e22057. doi. 10.7554/eLife.22057.
19. Loo M, Mark PJ. The stringdist package for approximate string matching. *R J.* 2014;6(1):111. doi. 10.32614/RJ-2014-011.
20. Csardi G, Nepusz T. The Igraph software package for complex network research. *InterJournal.*; 2006. Complex Sy:1695. igraph.org.
21. The International HapMap Consortium, Frazer KA, Ballinger DG, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature.* 2007;449(7164):851-861. doi. 10.1038/nature06258.
22. Emerson RO, DeWitt WS, Vignali M, et al. Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nat Genet.* 2017;49(5):659-665. doi. 10.1038/ng.3822.
23. DeWitt WS, Smith A, Schoch G, Hansen JA, Matsen FA, Bradley P. Human T cell receptor occurrence patterns encode immune history, genetic background, and receptor specificity. *Elife.* 2018;7:e38358. doi. 10.7554/eLife.38358.
24. Britanova OV, Putintseva EV, Shugay M, et al. Age-related decrease in TCR repertoire diversity measured with deep and normalized sequence profiling. *J Immunol.* 2014;192(6):2689-2698. doi. 10.4049/jimmunol.1302064.
25. Krishna C, Chowell D, Gönen M, Elhanati Y, Chan TA. Genetic and environmental determinants of human TCR repertoire diversity. *Immun Ageing.* 2020;17(1):26. doi. 10.1186/s12979-020-00195-9.
26. Qi Q, Liu Y, Cheng Y, et al. Diversity and clonal selection in the human T-cell repertoire. *Proc Natl Acad Sci U S A.* 2014;111(36):13139-13144. doi. 10.1073/pnas.1409155111.
27. Gong M, Li X, Zheng A, et al. Age-related changes in the TRB and IGH repertoires in healthy adult males and females. *Immunol Lett.* 2021;240:71-76. doi. 10.1016/j.imlet.2021.10.002.
28. Bjornevik K, Cortese M, Healy BC, et al. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science.* 2022;375(6578):296-301. doi. 10.1126/science.abj8222.
29. Isobe N, Keshavan A, Gourraud P-A, et al. Association of HLA genetic risk burden with disease phenotypes in multiple sclerosis. *JAMA Neurol.* 2016;73(7):795-802. doi. 10.1001/jamaneurol.2016.0980.
30. the International Multiple Sclerosis Genetics Consortium. Class II HLA interactions modulate genetic risk for multiple sclerosis. *Nat Genet.* 2015;47(10):1107-1113. doi. 10.1038/ng.3395.
31. Okuda DT, Srinivasan R, Oksenberg JR, et al. Genotype-Phenotype correlations in multiple sclerosis: HLA genes influence disease severity inferred by 1HMR spectroscopy and MRI measures. *Brain.* 2009;132(1):250-259. doi. 10.1093/brain/awn301.
32. Masterman T, Ligers A, Olsson T, Andersson M, Olerup O, Hillert J. HLA-DR15 is associated with lower age at onset in multiple sclerosis. *Ann Neurol.* 2000;48(2):211-219. doi. 10.1002/1531-8249(200008)48:2<211::aid-ana11>3.0.co;2-r.
33. Aversa I, Malanga D, Fiume G, Palmieri C. Molecular T-cell repertoire analysis as source of prognostic and predictive biomarkers for checkpoint blockade immunotherapy. *Int J Mol Sci.* 2020;21(7):2378. doi. 10.3390/ijms21072378.
34. Lundtoft C, Pucholt P, Imgenberg-Kreuz J, et al. Function of multiple sclerosis-protective HLA class I alleles revealed by genome-wide protein-quantitative trait loci mapping of interferon signalling. *PLoS Genet.* 2020;16(10):e1009199. doi. 10.1371/journal.pgen.1009199.
35. Skulina C, Schmidt S, Dornmair K, et al. Multiple sclerosis: brain-infiltrating CD8+ T cells persist as clonal expansions in the cerebrospinal fluid and blood. *Proc Natl Acad Sci U S A.* 2004;101(8):2428-2433. doi. 10.1073/pnas.0308689100.
36. Jacobsen M, Cepok S, Quak E, et al. Oligoclonal expansion of memory CD8+ T cells in cerebrospinal fluid from multiple sclerosis patients. *Brain.* 2002;125(3):538-550. doi. 10.1093/brain/awf059.
37. Lossius A, Johansen JN, Vartdal F, et al. High-throughput sequencing of TCR repertoires in multiple sclerosis reveals intrathecal enrichment of EBV-reactive CD8+ T cells. *Eur J Immunol.* 2014;44(11):3439-3452. doi. 10.1002/eji.201444662.
38. Alves Sousa AdP, Johnson KR, Ohayon J, Zhu J, Muraro PA, Jacobson S. Comprehensive analysis of TCR-β repertoire in patients with neurological immune-mediated disorders. *Sci Rep.* 2019;9(1):344. doi. 10.1038/s41598-018-36274-7.
39. Wing JB, Sakaguchi S. TCR diversity and Treg cells, sometimes more is more. *Eur J Immunol.* 2011;41(11):3097-3100. doi. 10.1002/eji.201142115.
40. Thapa DR, Tonikian R, Sun C, et al. Longitudinal analysis of peripheral blood T cell receptor diversity in patients with systemic lupus erythematosus by next-generation sequencing. *Arthritis Res Ther.* 2015;17(1):132. doi. 10.1186/s13075-015-0655-9.
41. Liu X, Zhang W, Zhao M, et al. T cell receptor β repertoires as novel diagnostic markers for systemic lupus erythematosus and rheumatoid arthritis. *Ann Rheum Dis.* 2019;78(8):1070-1078. doi. 10.1136/annrheumdis-2019-215442.
42. Hayashi F, Isobe N, Glanville J, et al. A new clustering method identifies multiple sclerosis-specific T-cell receptors. *Ann Clin Transl Neurol.* 2021;8(1):163-176. doi. 10.1002/acn3.51264.
43. Martin R, Sospedra M, Eiermann T, Olsson T. Multiple sclerosis: doubling down on MHC. *Trends Genet.* 2021;37(9):784-797. doi. 10.1016/j.tig.2021.04.012.
44. Martin R, Jaraquemada D, Flerlage M, et al. Fine specificity and HLA restriction of myelin basic protein-specific cytotoxic T cell lines from multiple sclerosis patients and healthy individuals. *J Immunol.* 1990;145(2):540-548. doi. 10.4049/jimmunol.145.2.540.
45. Pette M, Fujita K, Wilkinson D, et al. Myelin autoreactivity in multiple sclerosis: recognition of myelin basic protein in the context of HLA-DR2 products by T lymphocytes of multiple-sclerosis patients and healthy donors. *Proc Natl Acad Sci U S A.* 1990;87(20):7968-7972. doi. 10.1073/pnas.87.20.7968.
46. Vergelli M, Kalbus M, Rojo SC, et al. T cell response to myelin basic protein in the context of the multiple sclerosis-associated HLA-DR15 haplotype: peptide binding, immunodominance and effector functions of T cells. *J Neuroimmunol.* 1997;77(2):195-203. doi. 10.1016/s0165-5728(97)00075-1.
47. Wang J, Jelcic I, Mühlenbruch L, et al. HLA-DR15 molecules jointly shape an autoreactive T cell repertoire in multiple sclerosis. *Cell.* 2020;183(5):1264-1281.e20. doi. 10.1016/j.cell.2020.09.054.
48. Kondo T, Cortese I, Markovic-Plese S, et al. Dendritic cells signal T cells in the absence of exogenous antigen. *Nat Immunol.* 2001;2(10):932-938. doi. 10.1038/ni711.
49. Amoriello R, Greiff V, Aldinucci A, et al. The TCR repertoire reconstitution in multiple sclerosis: comparing one-shot and continuous immunosuppressive therapies. *Front Immunol.* 2020;11:559. doi. 10.3389/fimmu.2020.00559.
50. Chiarini M, Sottini A, Bertoli D, et al. Newly produced T and B lymphocytes and T-cell receptor repertoire diversity are reduced in peripheral blood of fingolimod-treated multiple sclerosis patients. *Mult Scler.* 2015;21(6):726-734. doi. 10.1177/1352458514551456.