The similarity of class II HLA genotypes defines patterns of autoreactivity in idiopathic bone marrow failure disorders

Pagliuca, Simona

2021-11-09

Pagliuca , S , Gurnari , C , Awada , H , Kishtagari , A , Kongkiatkamon , S , Terkawi , L , Zawit , M , Guan , Y , LaFramboise , T , Jha , B K , Patel , B J , Hamilton , B K , Majhail , N S , Lundgren , S , Mustjoki , S , Sauntharajah , Y , Visconte , V , Chan , T A , Yang , C-Y , Lenz , T L & Maciejewski , J P 2021 , ' The similarity of class II HLA genotypes defines patterns of autoreactivity in idiopathic bone marrow failure disorders ' , Blood , vol. 138 , no. 26 , pp. 2781-2798 . https://doi.org/10.1182/blood.2021012900

http://hdl.handle.net/10138/352380
https://doi.org/10.1182/blood.2021012900

American Society of Hematology
2021 L Street NW, Suite 900,
Washington, DC 20036
Phone: 202-776-0544 | Fax 202-776-0545
editorial@hematology.org

The Similarity of Class II HLA Genotypes Defines Patterns of Autoreactivity in Idiopathic Bone Marrow Failure Disorders

Tracking no: BLD-2021-012900R1

Simona Pagliuca (Cleveland Clinic Foundation, United States) Carmelo Gurnari (Taussig Cancer Institute, Cleveland Clinic, United States) Hassan Awada (Department of Translational Hematology and Oncology Research, Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH, USA, United States) Ashwin Kishtagari (Cleveland Clinic, United States) Sunisa Kongkiatkamon (Cleveland Clinic, United States) Laila Terkawi (Taussig Cancer Institute, Cleveland Clinic, United States) Misam Zawit (University of Cincinnati Medical Center, United States) Yihong Guan (Cleveland Clinic, United States) Thomas LaFramboise (Case Western Reserve University, United States) Babal Jha (Taussig Cancer Institute, Cleveland Clinic, United States) Bhumika Patel (Cleveland Clinic Foundation, United States) Betty Hamilton (Cleveland Clinic, United States) Navneet Majhail (Cleveland Clinic, United States) Sofie Lundgren (Hematology Research Unit Helsinki, University of Helsinki and Helsinki University Hospital Comprehensive Cancer Center, Finland) Satu Mustjoki (University of Helsinki, Finland) Yogen Saunthararajah (Cleveland Clinic, United States) Valeria Visconte (Cleveland Clinic Taussig Cancer Institute, United States) Timothy Chan (Cleveland Clinic Foundation, United States) Chao-Yie Yang (University of Tennessee Health Science Center, United States) Tobias Lenz (University of Hamburg, Germany) Jaroslaw Maciejewski (Cleveland Clinic, United States)

Abstract:
Idiopathic aplastic anemia (IAA) is a rare autoimmune bone marrow failure disorder initiated by a human leukocyte antigen (HLA)-restricted T cell response to unknown antigens. As for other autoimmune disorders, the predilection for certain HLA profiles seems to represent an etiologic factor, however, the structure-function patterns involved in the self-presentation in this disease remain unclear. Herein we analyzed the molecular landscape of HLA complexes of a cohort of 300 IAA patients and almost 3000 healthy and disease controls, by deeply dissecting their genotypic configurations, functional divergence, self-antigen binding capabilities and T cell receptor (TCR) repertoire specificities. Specifically, analysis of the evolutionary divergence of HLA genotypes (HED) showed that IAA patients carried class II HLA molecules whose antigen binding sites were characterized by a high level of structural homology, only partially explained by specific risk allele profiles. This pattern implies reduced HLA binding capabilities, confirmed by binding analysis of hematopoietic stem cell derived self-peptides. IAA phenotype was associated with the enrichment in a few amino acids at specific positions within the peptide binding groove of DRB1 molecules, affecting the interface HLA-antigen-TCR β and potentially constituting the basis of T-cell dysfunction and autoreactivity. When analyzing associations with clinical outcomes, low HED was associated with risk of malignant progression and worse survival, underlying reduced tumor surveillance in clearing potential neoantigens derived from mechanisms of clonal hematopoiesis. Our data shed light on the immunogenetic risk associated with IAA etiology and clonal evolution, and on general pathophysiological mechanisms potentially involved also in other autoimmune disorders.

Conflict of interest: No COI declared

COI notes:

Preprint server: Yes; medRxiv 10.1101/2021.05.28.21258028

Author contributions and disclosures: SP designed the study, collected, analyzed and interpreted the data, performed the bioinformatic and statistical analyses and wrote the manuscript. CG performed NGS experiments, clinical data collection and participated in the analysis interpretation. SK, LT, MZ, AK, HA, YG, BJK helped in sample and data collection. TLa and VV edited the manuscript, helped in data interpretation and gave helpful intellectual insights during the study. YS, BJP, MN, BH actively participated in patient recruitment, management, and follow-up. SL and SM provided genotypic data for the Finnish cohort and critically revised the manuscript. CY performed the molecular modeling, analyzed the crystallographic structures helped in drafting the methods inherent to this part. TC and TLe helped in data interpretation and analytical method development. JM designed and conceptualized the study, interpreted the data analysis and edited the manuscript. First and last authors took responsibility for the integrity and the accuracy of the data presented. All authors reviewed and approved the final version of this manuscript.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: All the data that support the findings of this study are available within the Article and Supplementary Files. Genotyping and clinical data of IAA and HC cohorts have been deposited in the following repository: https://github.com/SMNPAG/HED-IAA. Access can be requested to maciejj@ccf.org and smnpag@gmail.com. TCR sequencing data of the IAA cohort are available through the ImmuneACCESS platform: https://doi.org/10.21417/SP2021B. TCR sequencing data of the HC cohort are provided with the original study and through the ImmuneACCESS platform (https://doi.org/10.21417/B7001Z). Genotypic raw data of the disease-control cohorts can be requested to the authors (MN cohort) and to the National Institute of Diabetes and Digestive and Kidney diseases (T1D cohort).

Clinical trial registration information (if any):

The Similarity of Class II HLA Genotypes Defines Patterns of Autoreactivity in Idiopathic Bone Marrow Failure Disorders

Simona Pagliuca[1,2], Carmelo Gurnari[1,3], Hassan Awada[1], Ashwin Kishtagari[1], Sunisa Kongkiatkamon[1], Laila Terkawi[1], Misam Zawit[1], Yihong Guan[1], Thomas LaFramboise[4], Babal K. Jha[1], Bhumika J. Patel[5], Betty K. Hamilton[6], Navneet S. Majhail[6], Sofie Lundgren[7,8], Satu Mustjoki[7,8,9], Yogen Saunthararajah[1], Valeria Visconte[1], Timothy Chan[10], Chao-Yie Yang[11], Tobias L. Lenz[12,13], and Jaroslaw P. Maciejewski[1].

[1]Translational Hematology and Oncology Research Department, Cleveland Clinic, Cleveland, OH, USA
[2]University of Paris, Paris, France
[3]Department of Biomedicine and Prevention, University of Rome Tor Vergata, Rome, Italy
[4]Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH, USA
[5]Leukemia program, Department of Hematology and Oncology, Cleveland Clinic, Cleveland, OH, USA
[6]Blood and Marrow Transplant Program, Department of Hematology and Oncology, Cleveland Clinic, Cleveland, OH, USA [7]Hematology Research Unit Helsinki, University of Helsinki and Helsinki University Hospital Comprehensive Cancer Center, Helsinki, Finland
[8]Translational Immunology Research program and Department of Clinical Chemistry and Hematology, University of Helsinki, Helsinki, Finland
[9]ICAN Digital Precision Cancer Medicine Flagship, Helsinki, Finland
[10]Center for Immunotherapy and Precision Immuno-Oncology, Cleveland Clinic, Cleveland, OH, USA
[11]Department of Pharmaceutical Sciences, University of Tennessee Health Science Center, Memphis, TN, USA
[12]Research Group for Evolutionary Immunogenomics, Max Planck Institute for Evolutionary Biology, Plön, Germany
[13]Research Unit for Evolutionary Immunogenomics, Department of Biology, University of Hamburg, Hamburg, Germany

Key points:

- **Class II human leukocyte antigen (HLA) loci in idiopathic bone marrow failure (BMF) disorders are characterized by low structural divergence**
- **This immunogenetic pattern contributes to decreasing T-cell receptor repertoire diversity, favoring cross-reactivity and autoimmunity**

Running title: Immunogenetic basis of bone marrow failure disorders

Word count:

Abstract: 250
Main text: 4554
References: 76
Tables: 1
Figures: 7

Corresponding Author:
Jaroslaw P. Maciejewski, MD, PhD
Lerner Research Institute NE6
Cleveland Clinic
9620 Carnegie Ave n building, NE6-314, Cleveland, OH
E-mail: maciejj@ccf.org

Abstract

Idiopathic aplastic anemia (IAA) is a rare autoimmune bone marrow failure disorder initiated by a human leukocyte antigen (HLA)-restricted T cell response to unknown antigens. As for other autoimmune disorders, the predilection for certain HLA profiles seems to represent an etiologic factor, however, the structure-function patterns involved in the self-presentation in this disease remain unclear. Herein we analyzed the molecular landscape of HLA complexes of a cohort of 300 IAA patients and almost 3000 healthy and disease controls, by deeply dissecting their genotypic configurations, functional divergence, self-antigen binding capabilities and T cell receptor (TCR) repertoire specificities. Specifically, analysis of the evolutionary divergence of HLA genotypes (HED) showed that IAA patients carried class II HLA molecules whose antigen binding sites were characterized by a high level of structural homology, only partially explained by specific risk allele profiles. This pattern implies reduced HLA binding capabilities, confirmed by binding analysis of hematopoietic stem cell derived self-peptides. IAA phenotype was associated with the enrichment in a few amino acids at specific positions within the peptide binding groove of DRB1 molecules, affecting the interface HLA-antigen-TCR β and potentially constituting the basis of T-cell dysfunction and autoreactivity. When analyzing associations with clinical outcomes, low HED was associated with risk of malignant progression and worse survival, underlying reduced tumor surveillance in clearing potential neoantigens derived from mechanisms of clonal hematopoiesis. Our data shed light on the immunogenetic risk associated with IAA etiology and clonal evolution, and on general pathophysiological mechanisms potentially involved also in other autoimmune disorders.

Introduction

Among bone marrow failure syndromes (BMF), idiopathic aplastic anemia (IAA) is a hematopoietic stem cell (HSC) disorder mediated by autoimmune T cells. Despite the progress in our understanding of basic disease mechanisms, the strongest evidence for the immune pathogenesis of this disease stems from the successes associated with immune suppressive therapies (IST).[1] In addition, clinical observations such as evolution of paroxysmal nocturnal hemoglobinuria (PNH) clones escaping from autoimmune selection pressures, as well as

somatic loss of human leukocyte antigen (HLA) alleles due to deletion or mutations, further support the immune nature of this disorder.[2,3,4,5,6]

From a pathophysiological point of view, the *primum movens* of bone marrow destruction is thought to be a class I HLA-restricted process, characterized by cytotoxic T lymphocytes (CTLs) recognition of still unknown HSCs antigens.[7,8, 9,10] Experimental data demonstrating activation of CTLs with oligoclonal expansion and skewing of $CD8^+$ T-cell receptor (TCR) repertoire together with an interferon gamma (IFN-γ)-driven FAS-mediated apoptosis of HSCs support this hypothesis.[8,11,12,13,14,15] Furthermore, the imbalance of $CD4^+$ subsets with dysfunctional T helper (Th) 1, Th2 and Th17 responses and consequential impairment of regulatory T cells activities have been demonstrated as additional factors contributing to the aberrant auto-reactivity.[16,17,18,19] However, because the identity of the eliciting antigens has not been ascertained, the laboratory evidence for T cell-mediated pathogenesis, albeit compelling, remains only indirect.

In the context of cellular autoimmune reactions, the pivotal role of HLA molecules in mediating the $CD8^+$ and $CD4^+$ processes is generally established. Specifically, in IAA a mechanistic involvement of HLA is supported by HLA allele predilection and the modes of immune escape *via* somatic reshuffles of HLA locus (e.g., loss or uniparental disomy of chromosome 6p or somatic mutations in class I alleles). [2, 3,5,20,21, 22,23,24]

While external triggers seem essential, genetic disease susceptibility factors (*e.g.*, immunogenetic polymorphisms) appear to be operative. Unlike in other autoimmune disorders, the enrichment of some class I alleles in IAA remains limited to ethnicity-restricted series,[2] whereas the impact of class II loci on disease susceptibility (*e.g.,* DRB1*15:01) has been historically well documented in multiple populations.[25,26,27,28,29,30,31,32,33,34,35,36,37,38]

Conceptually, the predilection of certain HLA alleles could be explained by their structural "suitability" to present specific immunodominant peptides. However, the relatively small contribution of individual risk alleles in defining the etiologic fraction of IAA, scarcely adapts the structure/function relationships into a disease-specific autoantigenic profile.

Recent studies have shown that HLA evolutionary divergence (HED), a metric capturing the pairwise Grantham distance[39] between the peptide binding sites of two homologous HLA molecules encoded by an individual's genotype, correlates with the size of the immunopeptidomic spectrum.[40,41,42] In this virtue, more structurally divergent alleles may

confer a proclivity for more efficient T-cell responses. However, such a postulate, albeit well documented in the context of anti-tumor and anti-infectious immune surveillance,[40,41,43,44] to date has not been explored in the context of an autoimmune disorder such as IAA. A perhaps simplistic assumption would be that higher allelic divergence may increase the probability of self-antigenic presentation, eliciting autoreactive T-cell responses and explaining the association with autoimmune disease phenotypes. **Nevertheless, other hypotheses can be as well contemplated: for instance, the enrichment in less divergent HLA structures may instead contribute to the presentation of a less extended immune-peptidomes accounting for a decreased TCR diversity potentially at basis of cross-reactivity and autoimmune phenomena.**

Herein, we sought to understand the role of HLA functional variability in defining the predisposition and the phenotypic traits of IAA and related disorders. To that end, we performed a large case-control study, in which we analyzed the structural divergence of HLA molecules, its association with risk allele profiles, and the presentation of HSC-related immunopeptidomic specificities. We first performed an allele frequency estimation that helped in the identification and confirmation of the alleles more likely associated with the disease. We then examined the impact of the global genotypic and molecular HLA diversity for both class I and II loci on either disease predisposition or characteristics at diagnosis and clinical outcomes. Finally, we studied how risk and divergent allelic profiles influenced the presentation of HSC immunopeptidome and the TCR repertoire characteristics.

Methods

Study population

We used for this study genotyping information from 263 IAA/PNH patients, 960 healthy subjects 510 patients with myeloid neoplasia, and 1340 subjects with type 1 diabetes. An additional dataset of HLA genotypes for 37 IAA/PNH patients and 128 healthy controls from a Finnish (University of Helsinki) population was built for validation purposes. All patients gave their informed consent to participate to translational research protocols according to local institutional review boards. All procedures were carried out in accordance with guidelines set forth by the Declaration of Helsinki (See supplementary appendix for details of patient cohorts).

HED computation

High quality 4-digit HLA data, in patient and control cohorts were used for the phenotypic association study and the risk allele analysis (see supplementary appendix).

HED scores were computed for all the subjects and all the genotypes in study, using the algorithm published by Pierini and Lenz, applying a customized perl script (https://sourceforge.net/projects/granthamdist/) for the calculation of the amino acid sequence divergence.[41,7] Briefly, starting from a dictionary including all the protein sequences of exons 2 and 3 for class I alleles and exon 2 for class II alleles, assembled from the IPD-IMGT/HLA database v.3.40,[45] we calculated HED for 6 class I (A, B, C) and II HLA loci (DRB1, DQB1, DPB1).

Immunopeptidomic analysis

A HSC specific proteomic reference, as published by Henrich et al,[46] was used to predict peptides binding to alleles enriched in the BMF cohort. Amino acid sequences based on this reference were selected from the human peptidome reference, downloaded from Ensembl[47] (ftp://ftp.ensembl.org/pub/grch37/update/fasta/homo_sapiens/pep//Homo_sapiens.GRCh37.pep.all.fa). The resulting FASTA file was submitted to NetMHCIIpan 4.0[48] within a high performant computational environment. All peptides with a percentile rank of eluted ligand prediction score <2% for strong binders and <5% for weak binders were retained for the downstream analyses.

Crystallographic structures and molecular dynamics

Crystallographic structures of DR molecules were prepared with the PyMOL (www.pymol.org) program. Homology models of DRB1 molecules were constructed using the software I-TASSER.[49,50] The protocol of molecular dynamic simulations was conducted according to a previous study.[51] Details of modeling, simulation procedures and structural motifs used in this study are reported as supplementary materials.

TCRβ chain sequencing and analysis

Immunosequencing of the CDR3 regions of human TCRβ chains was performed using the ImmunoSEQ Assay (Adaptive Biotechnologies, Seattle, WA), as previously described. [52,53,54] ImmunoSeq Analizer 3.0 suite was used for sample export and preliminary statistics and quality control steps while R bioconductor[55] environment and Immunarch R[56] suite were used for all the downstream analyses. **The Immune Epitope Database (IEDB) analysis resources were used to analyze the patterns of cross-reactivity and epitope similarities (see supplementary appendix).**[57]

Statistical analysis

The statistical framework used for the HLA association study and the clinical correlations is described in the supplementary appendix. All of the analyses and data visualization were performed using the statistical computing environment R (4.0.0 R Core Team, R Foundation for Statistical Computing, Vienna, Austria) and excel Microsoft 365.

Data Sharing

All the data that support the findings of this study are available within the Article and Supplementary Files. Genotyping and clinical data of IAA and HC cohorts have been deposited in the following repository: https://github.com/SMNPAG/HED-IAA. Access can be requested to maciejj@ccf.org and smnpag@gmail.com. TCR sequencing data of the IAA cohort are available through the ImmuneACCESS platform: https://doi.org/10.21417/SP2021B. TCR sequencing data of the HC cohort are provided with the original study and through the ImmuneACCESS platform (https://doi.org/10.21417/B7001Z).
Genotypic raw data of the disease-control cohorts can be requested to the authors (MN cohort) and to the National Institute of Diabetes and Digestive and Kidney diseases (T1D cohort).

Results

Association analysis and risk allele imputation

Out of **a large cohort of** adult patients with IAA and hemolytic PNH followed at our institution, 263 patients with completed outcomes and precisely asserted diagnosis had DNA for next generation sequencing (NGS)-based HLA typing (**Table 1, Table S1, Fig.1**). For comparison, we built two different control cohorts: i) 960 healthy subjects (HC) from a prevalently Caucasian North-American population, and ii) 510 patients with myeloid neoplasms (MN), with known HLA genotypes (see methods and supplementary appendix). For the analysis of HLA

associations, the frequency of each allele was evaluated according to a dominant genetic model (see box of definitions in supplementary appendix), assessing the association strength with class assignment.[58] Among 8 loci, 9 class I and II alleles were identified as differentially distributed between HC and IAA/PNH cohort **(Fig. 2A, Table S2**), with 4 alleles significantly enriched in IAA/PNH patients, according to a dominant genetic model: DRB1*15:01 DQB1*06:02, B*07:02 and DQA1*01:02. Analysis of the additive effect confirmed a strong association with the disease phenotype (**Fig.2B, Table S3**). The distribution of the significant 4 risk alleles did not differ among IAA and hemolytic PNH groups (**Fig.S2A**). To determine whether the presence of each IAA risk allele may influence the course of the disease, we performed a logistic regression univariate analysis, which did not reveal any association with malignant progression to myeloid neoplasia (MN) and/or PNH evolution (**Fig.S2B-D**). However, the presence of DQB1*06:02 showed lower odds in terms of response to IST (OR: 0.48 [95%CI 0.26-0.90], p=0.028, **Fig. S2B-D**). By comparison, none of the risk alleles in study were enriched in patients with MN (**Table S4**). Of note is that those 4 alleles together with A*03:01 and C*07:02 belonged to the ancestral haplotype 7.1,[59] which was equally distributed across patient and control groups (4% vs. 5%, respectively, p=0.753). Carriers of at least one risk allele were enriched in the IAA vs. HC with 55% of the cases harboring $\geq$1 risk allele (vs. 39% in HC, OR: 1.94 [95%CI 1.47-2.55], p<0.0001, **Fig.2B**). Furthermore, DRB1*15:01 was associated with DQB1*06:02 in 84% of cases (patients and HC), as a result of the strong linkage disequilibrium. The frequency of homozygous for class II risk alleles in IAA group was higher than in HC, underscoring the role of the "allelic dose" in disease predisposition (**Fig.2C**). Risk allele associations, assessed with binomial regression models, were stronger when considering subgroups of patients with increased autoimmune propensity including i) responders to immunosuppression, ii) age >20yrs.; iii) IAA with PNH clone; iv) all of the above characteristics (**Fig. 2D**, **Table S5)**.

Low class II HLA divergence as an immunogenetic determinant in IAA patients

We applied HED concept to explore the immunogenetic configuration of IAA and PNH patients, taking into account the risk allele background and computing HED metrics for all classical I and II loci (A, B, C, DRB1, DQB1, DPB1). Based on the previously validated locus-specific HED metrics, for class II divergence computation, we accounted for β-chains only, because of the greater variability of α-chains in terms of peptide binding sites.[41] Genotypic differences did not

impact the class-related homozygosity configurations of the three cohorts, with subjects with at least one homozygous locus per class being equally distributed in the groups (**Fig.S2A,B**). When we investigated the organization of global class I and II HED (mean of HED scores for each class) in patients and control groups, no differences were observed for class I (IAA vs. HC: p=0.411 or IAA vs MN p=0.189, **Fig.S2C**). Conversely, a lower mean class II HED was found in IAA/PNH cohort vs. HC (p=0.033; **Fig.S2D**), involving specifically DRB1 (adj. p=0.028) and DQB1 (adj. p=0.028) loci **(Fig.S2E,F**). Associations with lower class II and locus-specific divergence were more evident when we applied generalized linear regression models to predict the risk of immune-enriched disease phenotypes (**Fig.3A**, **Table S6**). This pattern was confirmed also when considering only Caucasian subjects (see supplementary appendix and Table S7). No differences compared to controls instead were found when analyzing mean class I and class II HED metrics in MN cohort (**Fig. 2A-B,** of note is that MN were characterized by a higher HED in locus B compared to HC).

When the additive effect of the presence of risk alleles on class II HED configurations was analyzed, we observed that lower locus specific HED was independently associated with IAA/PNH phenotype (**Fig.3B**). Interestingly, HC carrying DRB1*15:01 had lower class II and locus-specific HED compared to non-carriers **(Fig.S3A, B)**, whereas in IAA both class II and DRB1 divergences were globally reduced either in presence or in absence of this risk allele (**Fig.S3C,D**). Nevertheless, patients without DRB1*15:01 had a significantly lower mean class II (p=0.0005) and DRB1 HED (p<0.0001) compared to HC (**Fig.3C,D**). A similar pattern was seen in DQB1 HED among non-carriers of DQB1*06:02 (**Fig. 3E,F**). When for comparison purposes we examined another autoimmune disease cohort (type 1 diabetes [T1D]), we also found a lower divergence in class II loci (p= 0.00058) and in DRB1 locus compared to HC (p= 0.00104; **Table S7; Fig.S4 A, B, C**). Binomial regression analysis confirmed lower HED in DRB1 locus as a predictor of T1D (OR: 0.97 [95%CI 0.95-0.99], p= 0.00235, **Fig S4D**). Indeed, it is noteworthy the established role of DRB1*15:01 as a protective allele in T1D (**Fig. S4D**), as well as the different risk allele profile dominated by DRB1*03:01 and DRB1*04:01.[60],[61],[62]

To identify alleles structurally similar to those identified as risk alleles in class II, we simulated the range of HED scores between DRB1*15:01/DQB1*06:02 and the pool of alleles present in DRB1 and DQB1 loci in IAA/PNH patients and HC (**Fig.4A-C**). In DRB1 locus, a lower divergence with DRB1*15:01 was obtained for alleles within DRB1*15, DRB1*16, DRB1*04 and DRB1*01 supertypes (**Fig.4A; Table S8**). Combined genotypic frequencies of those DRB1*15:01-like

alleles were higher in IAA/PNH population (OR: 1.89 [95%CI: 1.38-2.61], p=7.22e-05) and in cases with immune-related phenotypes compared to HC (**Fig.4B**). Analogous results were observed for DQB1 locus and DQB1*06 supertype (**Fig.4C, D; Table S9**).

When we studied whether HLA functional divergence may influence characteristics and outcomes of IAA/PNH patients, we found that class II HED scores correlated directly with the size of PNH clone at diagnosis (**Fig.S5A-B;** p= 0.00286, r^2=0.036) and indirectly with age at disease onset (**Fig.S5C-D;** p=0.0013, r^2=0.021). Univariable cox regression models (based on binomial categorization of mean class I and II HED according to the 50th percentile in HC, **Fig.S5E-K**), demonstrated lower probability of survival (p=0.011), higher risk of progression to MDS/AML (p=0.043) and a lower probability of PNH evolution (p=0.0004, **Fig.S5J,L**) in patients with lower mean class II HED, while no impact was seen for class I HED (**Fig.S5G,I,J**).

Recursive analysis of the antigen binding site in DRB1 and DQB1 loci

To investigate whether low divergent patterns could rely on a specific amino acid composition of the peptide binding site of class II HLA molecules, we analyzed the amino acid structure in the antigen binding site (encoded by exon 2) of DRB1 and DQB1 loci. In DRB1 peptide binding groove, 30 out of 89 amino acid positions were variable. By applying a recursive approach we found strongly correlated with the IAA/PNH phenotype 7 amino acids (**Fig. 5A,B; Table S10**), enriched in DRB1*15 group and in all the alleles structurally similar to this supertype. Of note is that most of those residues were non-polar, possibly affecting the physicochemical configuration of the antigen-binding site of DRB1*15-like complexes. When the same analysis was performed on DQB1 locus (**Fig.S6A,B; Table S11**), the majority of variable amino acids in exon 2 were enriched in HC, with the exception of 2 residues of phenylalanine (Phe/F), both belonging to the DQB1*06 supertype and significantly overrepresented in IAA (**Fig.S6A,B**). **Interestingly, none of the antigen binding sites extrapolated from classical class I loci presented specific amino acid signatures enriched in IAA/PNH cohort (Tables S12).**

Because of this lower contribution of DQB1 **and class I** binding sites variability in conferring disease phenotype, our next analysis was focused on modeling antigen interactions within DRB1 locus. When we analyzed the crystallographic structure of the complex DRB1*15:01/DRA*01:01 (PDB: 1BX2), the identified amino acids clustered within the right part of the antigen binding pocket (**Fig.5C**). Only two of 7 amino acids were different in DRB1*16:01

(indicating higher structural similarity to DRB1*15:01; **Fig.5D**), while 5/7 residues differed in DRB1*12:01 (more structurally divergent from DRB1*15:01; **Fig.5E**). We also aligned the crystal structures of DRB1*15:01/DRA*01:01 with three peptides known to have affinity for DRB1 molecules (EBV DNA polymerase, vimentin and myelin binding protein [MBP], see methods) and found two different patterns of interaction within the binding groove: while the peptide portions binding in the left part of the groove tended to assume the same backbone conformation (underlying conservation of the physicochemical characteristics among the structures at this interface), peptide portions allocated within the right side of HLA groove assumed more variable conformations (**Fig.6A**). To investigate how this structural configuration could affect the TCR binding, we then constructed a model structure of TCR α and β chains, DRA*01:01-DRB1*15:01 and CD4 on the reported ternary crystal structure of HLA-peptide-TCR-CD4 (PDB: 3T0E).[63] Based on the alignment of these three antigenic models, we found that the variable peptide segments (right portion) interacted mainly with the TCR Vβ chain (**Fig.6B,C**). These findings provided a proof-of-concept for the importance of this consensus structure in accommodating the interactions with antigen and T-cell specificities within DRB1 locus.

Quantitative thresholds for self-antigenic presentation

To explore the binding capacity of self-generated peptides possibly involved in IAA pathogenesis, we built an *in silico* HSC-specific proteomic reference and we analyzed all the different DRB1 and DQB1 molecules present in IAA and HC cohorts. We generated 15-mer peptides from 40,614 transcripts assembled from 7724 previously identified HSC proteins.[46] For each sequence, the mean number of strong and weak binders (see methods) was determined across all the DRB1 alleles, along with their gene source (**Table S13**). Overall, about 10% of this proteomic reference was found capable of generating self-peptides suitable for the binding of DRB1 or DQB1 molecules (all peptides with a percentile rank of eluted ligand prediction score <5% -corresponding to strong and weak binders- were considered for this analysis, **Fig.S7A,B**). HLA molecules belonging to the same locus had similar origin of derived binders. However, quantitative differences in the number of predicted binders were observed across all the alleles (**Table S13 & S14**). In particular DRB1*15:01 along with other alleles was identified as structurally similar to the DRB1*15 supertype and was characterized by lower binding capacities compared to other alleles (**Fig.S7B, C**). Analysis of the distribution of the number of binders in HC and IAA/PNH groups showed significantly lower binding capacities in DRB1 locus

in patients vs. controls (**Fig.S7D**). This pattern was observed also in the subgroup of homozygous individuals (HED score 0; **Fig.S7E**). For DQB1 locus immune-peptidomic analysis we accounted for the genotypic associations with DQA1 locus and thus we considered only individuals with known DQA1 allele. DQB1*0602-DQA1*01:02 was predicted as one of the haplotypes with the lowest binding capacities (**Fig. S8A-D**) and its propensity to bind HSC self-peptides was decreased compared to other DQB1*06:02/DQA1 combinations (**Fig.S8B**).

To determine whether the findings described above were reproducible across different ethnic groups and populations with unique HLA distributions, we investigated the genotypic and the structural patterns of the DRB1 locus in of IAA patients (N=37, among which N=30 genotyped at 4-digit level) and healthy controls (N=128) from a Finnish cohort. Consistently with our results, DRB1*15:01 was significantly associated with BMF phenotype (**Fig.S9A,B**; OR:5.18; p=0.0001) and divergence in DRB1 locus was lower for IAA patients compared to controls (this configuration was found both for the whole cohort and for non-carriers of DRB1*15:01, **Fig.S9C,D**). Importantly, as for the main cohort, also in the Finnish one, DRB1 allele distribution was characterized by lower binding capacities in patients vs. corresponding HC.

Insights in TCR repertoires

To investigate how risk allele profiles and HED configurations may dictate the patterns of T-cell responses, we performed deep TCR Vβ complementary determining region (CDR3) sequencing for 25 patients with IAA through a multiple sampling **(T**a**ble S15a)**. Patients' Vβ CDR3 spectra were compared to those of 122 HC included in our study (**Table S15b**). After a down-sampling procedure (see Methods), TCR diversity metrics were calculated. As expected, TCR repertoires were characterized by lower diversity compared to HC (IAA vs. HC: p=5.4e-08; number of unique clonotypes: p=0.012; mean size clonal expansion: p=0.512; **Fig.7A-C**) in both DRB1*15:01 carriers and non-carriers (**Fig. 7D-E**). No correlation was found between Vβ TCR diversity metrics and class I or II mean HED (**Fig. 7F-J**). To evaluate how HED could impact on the autoreactive disease-associated spectra, we first built a comprehensive compendium of all CDR3 sequences with known specificity identified in literature (see Methods) and blasted the CDR3 sequences identified in our cohorts against this dictionary. Overall, within the identified clonotypic portion of the repertoire (<2% of the total), the mean proportion of identifiable autoreactive clonotypes was 14% in IAA vs. 6% in HC (**Fig.7L**), while their mean frequency was respectively 0.016% and 0.001% (**Fig.7M;** p<2e-16). In IAA patients, but not in HC, the

productive frequency of those clones inversely correlated with mean class II HED (p=1.332E-19; r^2= 0.041; **Fig.7N**), underscoring the clonal expansion of those specificities in patients with lower class II HLA divergence. Almost all of the autoimmunity-associated clonotypic groups were found hyperexpanded in IAA compared to HC (**Fig.7 O**). **When analyzing the specificity of potential autoreactive clonotypes towards known epitopes we were able to identify (at a mean productive frequency comprised between 0.0005% and 0.09%) CDR3 sequences reactive against autoantigens recognized in other autoimmune diseases (Tables S16-17). These sequences were explored for their similarity with pathogen-associated TCRs and about 16% of them were predicted to be able to bind also infectious epitopes (Tables S16). This feature emphasizes the idea that a background of cross-reactivity, potentially imprinted by HLA, immunepeptidome and TCR interactions may exist in patients with IAA and potentially explain a generalized mechanism of autoreactivity.**

Discussion

In IAA, the autoimmune destruction of hematopoietic progenitor and stem cells is an HLA class I- and II-restricted T cell-mediated process. Here, with a comprehensive immunogenetic approach, encompassing deep NGS of HLA region, TCR sequencing and HSC specific immunopeptidome binding analysis, we intensively assessed HLA structures involved in disease susceptibility and potentially associated with autoimmune propensity. To that end, we used not only a comparative population of HC but also two large disease-control datasets.

The quantitative concept of HED relies on divergent allele advantage, stipulating that structural heterogeneity of HLA alleles allows for a wider spectrum of peptides to be presented and thus a higher probability to mount efficient anti-tumor and anti-infectious responses.[64,65,66,40] Accordingly, a similar principle should apply to autoimmune diseases, with a higher HED reflecting an increased propensity to T cell-mediated autoimmunity. However, we did not found support for this hypothesis and instead observed that HLA molecules in IAA patients were characterized by a high structural similarity, especially in class II, in part due to enrichment for risk alleles and/or alleles structurally similar to risk alleles. These associations were particularly strong for DRB1 and DQB1 loci. Consistently with previous studies, DRB1*15:01 together with DQB1*06:02 were identified as alleles enriched in IAA and PNH.[26,67,27] A special mention deserves the fact that HED metrics were found to be low,

independently of the presence of risk alleles in IAA/PNH setting. This pattern is explained by the global low divergence of class II HLA molecules seen in BMF cohort. Indeed, in IAA patients the non-risk alleles were more structurally similar to each other and to the risk alleles and thus may have analogous (albeit not completely overlapping) peptide recognition spectra. This may contribute to decrease the diversity and increase the clonal expansion of TCR specificities in IAA repertoire. Further, a lower locus specific HED was independently associated with disease phenotype also when performing generalized linear regression models tracking the additive effect of class II HLA risk alleles. Reinforcing the idea that this pattern may be present across different ethnicities and ancestry groups, we showed reproducibility of such findings in an independent cohort of Finnish IAA/PNH patients. Also, we found the same low HLA divergence pattern in an autoimmune disease characterized by different risk allele associations and in which, of note, DRB1*15:01 is a protective allele.[60]

Since our genetic models were built on a mixed population and did not allow to confirm the class I risk alleles previously identified in a prevalently pediatric Caucasian cohort,[2] we performed also an ethnicity stratified subanalysis. Hence, among alleles formerly recognized, we could show only a slightly significant enrichment in B*14:02 in Caucasian patients from the North American cohort, and again confirmed the strong association with the class II alleles identified in the main analysis, as well as the low divergence characterizing class II loci. It is possible that the above dissimilarities with the previously reported class I associations[2] rely on the different age composition of our study cohort, prevalently composed by adult patients (see Supplementary considerations).

In order to deeper analyze the superstructures involved in the decreased divergence within the antigen binding sites, we studied the amino acid specificities found in the variable portion of the antigenic groove of DRB1 and DQB1 molecules with a recursive approach. This analysis enabled the identification of few residues significantly associated with IAA/PNH phenotype, mainly located in pockets involved in antigen and TCR β interactions. This analysis, by modeling analogous antigens known to be binders of DRB*15:01 and alleles with structural similarity, allowed us to identify antigen components involved in the preferential self-presentation of BMF patient, potentially associated with impaired T-cell activation.

Consistent with the lower divergence seen in IAA/PNH population, when we analyzed DRB1 and DQB1 binding predictions covering the HSC specific immunopeptidome, we found that for both loci, HLA genotypes had lower binding capacities in patients compared to healthy individuals,

both in homozygous and heterozygous settings. If, per divergent allele advantage, immunocompetence is supposed to be enhanced in case of higher HLA divergence, it is plausible that less divergent loci and, per extension, HLA molecules with lower binding capacities (such as DRB1*15:01 and DQB1*06:02-DQA1*01:02) may increase the risk of immunological cross-reactivity and molecular mimicry with possible pathogen-associated antigens, by that triggering autoimmune diseases. As an example of autoimmune disorder with known self-antigenic specificities, in multiple sclerosis the link among DRB1*15:01, EBV and central nervous system antigens' molecular mimicry is well established, with indirect evidence that impaired CD4+ presentation may elicit aberrant CD8+ responses and auto-antibody production.[68,69,70,71] **It is also noteworthy to mention recent data demonstrating in private clonotypic responses of IAA patients the presence of a cluster of converging CDR3 sequences, which were potentially epitope-specific, but not predicted to enriched for viral specificities responses.[72,73] We attempted to predict cross-reactivity across the auto-immune repertoire of the IAA samples using an approach of TCR matching based on homologies among CDR3 sequences. This particular analysis helped in identifying similarities between autoreactive clonotypes and TCR known to recognize pathogen-associated epitopes, underscoring the concrete possibility of cross-reactivity". The possible production of autoantibody seems also to fit with the idea that certain self-epitopes of targeted gene-products may be presented by DRB1*15:01 and DQB1*06:02 and associated structures.**

If risk allele profiles did not impact on clinical outcomes, low divergence in class II was associated with an increased probability of malignant progression in IAA/PNH patients. This finding is in line with the idea that anti-tumor surveillance (previously shown as more efficient in highly divergent class I allele pairs)[40] could encompass also HLA class II-restricted T-cell responses.[74,75,76] In this setting, less divergent class II β chains may reduce the neoantigen presentation capabilities configuring an immune escape scenario.

The *in silico* nature of some of our explorations may represent a drawback of our study. However immunopeptidomic studies for BMFs are far from being straightforward because of the limited number of residual HSCs precluding direct application of proteomic tools for the study of potential antigenic structures.

Despite these limitations, our data demonstrate that in BMFs and potentially in other autoimmune disorders, HLA allele configurations with more specific structural patterns and globally lower functional divergence may contribute to decrease the binding capabilities

especially in class II alleles, potentially enhancing antigenic cross-reactivity and, hence, autoimmune propensity.

Analysis of HLA divergence and identification of DRB1 superstructure may be easily translated in clinical practice for better diagnostic and prognostic orientations paving the way for new therapeutical approaches potentially able to modulate the self-antigenic binding capabilities of class II HLA molecules.

AUTHORSHIP AND DISCLOSURES

Authorship contributions:

SP designed the study, collected, analyzed and interpreted the data, performed the bioinformatic and statistical analyses and wrote the manuscript. CG performed NGS experiments, clinical data collection and participated in the analysis interpretation. SK, LT, MZ, AK, HA, YG, BJK helped in sample and data collection. TLa and VV edited the manuscript, helped in data interpretation and gave helpful intellectual insights during the study. YS, BJP, MN, BH actively participated in patient recruitment, management, and follow-up. SL and SM provided genotypic data for the Finnish cohort and critically revised the manuscript. CY performed the molecular modeling, analyzed the crystallographic structures helped in drafting the methods inherent to this part. TC and TLe helped in data interpretation and analytical method development. JM designed and conceptualized the study, interpreted the data analysis and edited the manuscript. First and last authors took responsibility for the integrity and the accuracy of the data presented. All authors reviewed and approved the final version of this manuscript.

Conflict-of-interest disclosure:

This research was conducted in absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data sharing

All the data that support the findings of this study are available within the Article and Supplementary Files. Genotypic raw data of the disease-control cohorts of patients with myeloid neoplasia and T1D are not included in this manuscript but can be requested to the authors (MN) and to the National Institute of Diabetes and Digestive and Kidney diseases (T1DGC) respectively. TCR sequencing data are available through the ImmuneACCESS platform (Adaptive Biotechnology).

Acknowledgements

This work was supported by US National Institute of Health (NIH) grants R35 HL135795, R01HL123904, R01 380HL118281, R01 HL128425, R01 HL132071, Edward P.Evans Foundation (to J.M), Italian Society of Hematology, Fondation ARC pour la Recherche sur le Cancer, Philippe Foundation, Association HPN France – Aplasie medullaire/ Foundation For Rare Diseases (FFRD) (to S.P.), The American-Italian Cancer Foundation (to C.G.); VeloSano Pilot Award, and Vera and Joseph Dresner Foundation–MDS (to V.V.). European Research Council (M-IMM and STRATIFY projects), Academy of Finland, Sigrid Juselius Foundation, and Cancer Foundation Finland (to S.L. and S.M).

We thank Diego Chowell for his helpful insights and critical revision of the manuscript.

We thank the National Institute of Diabetes and Digestive and Kidney diseases and the Type 1 Diabetes Genetic Consortium that provided the T1DGC Immunochip/HLA Reference Panel used in this study and Dr. John Sidney and Prof. Alessandro Sette who kindly provided the ethnicity information for the subjects from San Diego population.

References

1. Young NS, Maciejewski J. The pathophysiology of acquired aplastic anemia. *N Engl J Med*. 1997;336(19):1365–1372.
2. Babushok DV, Duke JL, Xie HM, et al. Somatic HLA Mutations Expose the Role of Class I-Mediated Autoimmunity in Aplastic Anemia and its Clonal Complications. *Blood Adv*. 2017;1(22):1900–1910.
3. Zaimoku Y, Takamatsu H, Hosomichi K, et al. Identification of an HLA class I allele closely involved in the autoantigen presentation in acquired aplastic anemia. *Blood*. 2017;129(21):2908–2916.
4. Katagiri T, Sato-Otsubo A, Kashiwase K, et al. Frequent loss of HLA alleles associated with copy number-neutral 6pLOH in acquired aplastic anemia. *Blood*. 2011;118(25):6601–6609.
5. Afable MG, Wlodarski M, Makishima H, et al. SNP array–based karyotyping: differences and similarities between aplastic anemia and hypocellular myelodysplastic syndromes. *Blood*. 2011;117(25):6876–6884.

6. Young NS, Maciejewski JP. Genetic and environmental effects in paroxysmal nocturnal hemoglobinuria: this little PIG-A goes “Why? Why? Why?” *J Clin Invest*. 2000;106(5):637–641.
7. Young NS. Current concepts in the pathophysiology and treatment of aplastic anemia. *Hematology*. 2013;2013(1):76–81.
8. Risitano AM. Oligoclonal and polyclonal CD4 and CD8 lymphocytes in aplastic anemia and paroxysmal nocturnal hemoglobinuria measured by Vbeta CDR3 spectratyping and flow cytometry. *Blood*. 2002;100(1):178–183.
9. Wlodarski MW, Gondek LP, Nearman ZP, et al. Molecular strategies for detection and quantitation of clonal cytotoxic T-cell responses in aplastic anemia and myelodysplastic syndrome. *Blood*. 2006;108(8):2632–2641.
10. Zeng W, Maciejewski JP, Chen G, Young NS. Limited heterogeneity of T cell receptor BV usage in aplastic anemia. *J Clin Invest*. 2001;108(5):765–773.
11. Risitano AM, Maciejewski JP, Green S, et al. In-vivo dominant immune responses in aplastic anaemia: molecular tracking of putatively pathogenetic T-cell clones by TCR β-CDR3 sequencing. *The Lancet*. 2004;364(9431):355–364.
12. Selleri C, Maciejewski JP, Sato T, Young NS. Interferon-gamma constitutively expressed in the stromal microenvironment of human marrow cultures mediates potent hematopoietic inhibition. *Blood*. 1996;87(10):4149–4157.
13. Sloand E. Intracellular interferon-gamma in circulating and marrow T cells detected by flow cytometry and the response to immunosuppressive therapy in patients with aplastic anemia. *Blood*. 2002;100(4):1185–1191.
14. Maciejewski J, Selleri C, Anderson S, Young NS. Fas antigen expression on CD34+ human marrow cells is induced by interferon gamma and tumor necrosis factor alpha and potentiates cytokine-mediated hematopoietic suppression in vitro. *Blood*. 1995;85(11):3183–3190.
15. Nisticò A, Young NS. gamma-Interferon gene expression in the bone marrow of patients with aplastic anemia. *Ann Intern Med*. 1994;120(6):463–469.
16. Zoumbos NC, Gascón P, Djeu JY, Trost SR, Young NS. Circulating Activated Suppressor T Lymphocytes in Aplastic Anemia. *New England Journal of Medicine*. 1985;312(5):257–265.
17. de Latour RP, Visconte V, Takaku T, et al. Th17 immune responses contribute to the pathophysiology of aplastic anemia. *Blood*. 2010;116(20):4175–4184.
18. Kordasti S, Marsh J, Al-Khan S, et al. Functional characterization of CD4+ T cells in aplastic anemia. *Blood*. 2012;119(9):2033–2043.
19. Kordasti S, Costantini B, Seidl T, et al. Deep phenotyping of Tregs identifies an immune signature for idiopathic aplastic anemia and predicts response to treatment. *Blood*. 2016;128(9):1193–1205.
20. Luzzatto L, Bessler M, Rotoli B. Somatic Mutations in Paroxysmal Nocturnal Hemoglobinuria: A Blessing in Disguise? *Cell*. 1997;88(1):1–4.
21. Katagiri T, Sato-Otsubo A, Kashiwase K, et al. Frequent loss of HLA alleles associated with copy number-neutral 6pLOH in acquired aplastic anemia. *Blood*. 2011;118(25):6601–6609.
22. Kikkawa E, Shiina T, Shigenari A, et al. Detection of 6pLOH in an aplastic anemia patient by in phase HLA genotyping. *HLA*. 2020;95(5):465–469.
23. Imi T, Katagiri T, Hosomichi K, et al. Sustained clonal hematopoiesis by HLA-lacking hematopoietic stem cells without driver mutations in aplastic anemia. *Blood Adv*. 2018;2(9):1000–1012.
24. Mizumaki H, Hosomichi K, Hosokawa K, et al. A frequent nonsense mutation in exon 1 across certain HLA-A and -B alleles in leukocytes of patients with acquired aplastic anemia. *Haematologica*. 2020;
25. Savage SA, Viard M, O’hUigin C, et al. Genome-wide Association Study Identifies HLA-DPB1 as a Significant Risk Factor for Severe Aplastic Anemia. *Am J Hum Genet*. 2020;106(2):264–271.

584 26. Saunthararajah Y, Nakamura R, Nam J-M, et al. HLA-DR15 (DR2) is overrepresented in
585 myelodysplastic syndrome and aplastic anemia and predicts a response to immunosuppression in
586 myelodysplastic syndrome. *Blood*. 2002;100(5):1570–1574.
587 27. Maciejewski JP, Follmann D, Nakamura R, et al. Increased frequency of HLA-DR2 in patients
588 with paroxysmal nocturnal hemoglobinuria and the PNH/aplastic anemia syndrome. *Blood*.
589 2001;98(13):3513–3519.
590 28. Dhaliwal JS, Wong L, Kamaluddin MA, Yin LY, Murad S. Susceptibility to aplastic anemia is
591 associated with HLA-DRB1*1501 in an aboriginal population in Sabah, Malaysia. *Hum Immunol*.
592 2011;72(10):889–892.
593 29. Oguz FS, Yalman N, Diler AS, et al. HLA-DRB1*15 and pediatric aplastic anemia.
594 *Haematologica*. 2002;87(7):772–774.
595 30. Song EY, Park S, Lee DS, Cho HI, Park MH. Association of human leukocyte antigen-DRB1
596 alleles with disease susceptibility and severity of aplastic anemia in Korean patients. *Hum*
597 *Immunol*. 2008;69(6):354–359.
598 31. Yari F, Sobhani M, Vaziri MZ, et al. Association of aplastic anaemia and Fanconi's disease
599 with HLA-DRB1 alleles. *Int J Immunogenet*. 2008;35(6):453–456.
600 32. Nakao S, Takami A, Sugimori N, et al. Response to immunosuppressive therapy and an HLA-
601 DRB1 allele in patients with aplastic anaemia: HLA-DRB1*1501 does not predict response to
602 antithymocyte globulin. *Br J Haematol*. 1996;92(1):155–158.
603 33. Nakao S, Takamatsu H, Chuhjo T, et al. Identification of a specific HLA class II haplotype
604 strongly associated with susceptibility to cyclosporine-dependent aplastic anemia. *Blood*.
605 1994;84(12):4257–4261.
606 34. Moutsianas L, Jostins L, Beecham AH, et al. Class II HLA interactions modulate genetic risk for
607 multiple sclerosis. *Nat Genet*. 2015;47(10):1107–1113.
608 35. Varney MD, Valdes AM, Carlson JA, et al. HLA DPA1, DPB1 alleles and haplotypes contribute
609 to the risk associated with type 1 diabetes: analysis of the type 1 diabetes genetics consortium
610 families. *Diabetes*. 2010;59(8):2055–2062.
611 36. Chinniah R, Rajendran MS, Sivanadham R, et al. Association of HLA class II alleles/haplotypes
612 and amino acid variations in the peptide binding pockets with rheumatoid arthritis. *Int J Rheum*
613 *Dis*. 2019;22(8):1553–1562.
614 37. Wang C, Zheng X, Jiang P, et al. Genome-wide Association Studies of Specific Antinuclear
615 Autoantibody Subphenotypes in Primary Biliary Cholangitis. *Hepatology*. 2019;70(1):294–307.
616 38. Kular L, Liu Y, Ruhrmann S, et al. DNA methylation as a mediator of HLA-DRB1*15:01 and a
617 protective variant in multiple sclerosis. *Nat Commun*. 2018;9(1):2397.
618 39. Grantham R. Amino Acid Difference Formula to Help Explain Protein Evolution. *Science*.
619 1974;185(4154):862–864.
620 40. Chowell D, Krishna C, Pierini F, et al. Evolutionary divergence of HLA class I genotype impacts
621 efficacy of cancer immunotherapy. *Nature Medicine*. 2019;25(11):1715–1720.
622 41. Pierini F, Lenz TL. Divergent Allele Advantage at Human MHC Genes: Signatures of Past and
623 Ongoing Selection. *Molecular Biology and Evolution*. 2018;35(9):2145–2158.
624 42. Krishna C, Chowell D, Gönen M, Elhanati Y, Chan TA. Genetic and environmental
625 determinants of human TCR repertoire diversity. *Immun Ageing*. 2020;17(1):26.
626 43. Arora J, McLaren PJ, Chaturvedi N, et al. HIV peptidome-wide association study reveals
627 patient-specific epitope repertoires associated with HIV control. *Proc Natl Acad Sci U S A*.
628 2019;116(3):944–949.
629 44. Arora J, Pierini F, McLaren PJ, et al. HLA Heterozygote Advantage against HIV-1 Is Driven by
630 Quantitative and Qualitative Differences in HLA Allele-Specific Peptide Presentation. *Mol Biol*
631 *Evol*. 2020;37(3):639–650.
632 45. Robinson J, Barker DJ, Georgiou X, et al. IPD-IMGT/HLA Database. *Nucleic Acids Research*.

2019;gkz950.
46. Hennrich ML, Romanov N, Horn P, et al. Cell-specific proteome analyses of human bone marrow reveal molecular features of age-dependent functional decline. *Nat Commun*. 2018;9(1):4004.
47. Zerbino DR, Achuthan P, Akanni W, et al. Ensembl 2018. *Nucleic Acids Research*. 2018;46(D1):D754–D761.
48. Reynisson B, Barra C, Kaabinejadian S, et al. Improved Prediction of MHC II Antigen Presentation through Integration and Motif Deconvolution of Mass Spectrometry MHC Eluted Ligand Data. *J. Proteome Res.* 2020;19(6):2304–2315.
49. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*. 2010;5(4):725–738.
50. Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res*. 2015;43(W1):W174-181.
51. Yang C-Y. Comparative Analyses of the Conformational Dynamics Between the Soluble and Membrane-Bound Cytokine Receptors. *Sci Rep*. 2020;10(1):7399.
52. Robins HS, Campregher PV, Srivastava SK, et al. Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. *Blood*. 2009;114(19):4099–4107.
53. Robins H, Desmarais C, Matthis J, et al. Ultra-sensitive detection of rare T cell clones. *Journal of Immunological Methods*. 2012;375(1–2):14–19.
54. Carlson CS, Emerson RO, Sherwood AM, et al. Using synthetic templates to design an unbiased multiplex PCR assay. *Nature Communications*. 2013;4(1):.
55. Gentleman R. Bioinformatics and computational biology solutions using R and Bioconductor. New York: Springer Science+Business Media; 2005.
56. Nazarov V, Immunarch.Bot, Rumynskiy E. immunomind/immunarch: 0.6.5: Basic single-cell support. Zenodo; 2020.
57. Vita R, Mahajan S, Overton JA, et al. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res*. 2019;47(D1):D339–D343.
58. Benjamini, Yoav; Hochberg, Yosef. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B 57 (1995), no. 1, 289–300.* .
59. Dorak MT, Shao W, Machulla HKG, et al. Conserved extended haplotypes of the major histocompatibility complex: further characterization. *Genes Immun*. 2006;7(6):450–467.
60. Erlich H, Valdes AM, Noble J, et al. HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk: analysis of the type 1 diabetes genetics consortium families. *Diabetes*. 2008;57(4):1084–1092.
61. Lenz TL, Deutsch AJ, Han B, et al. Widespread non-additive and interaction effects within HLA loci modulate the risk of autoimmune diseases. *Nat Genet*. 2015;47(9):1085–1090.
62. Hu X, Deutsch AJ, Lenz TL, et al. Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. *Nat Genet*. 2015;47(8):898–905.
63. Yin Y, Wang XX, Mariuzza RA. Crystal structure of a complete ternary complex of T-cell receptor, peptide-MHC, and CD4. *Proc Natl Acad Sci U S A*. 2012;109(14):5405–5410.
64. Potts WK, Wakeland EK. Evolution of diversity at the major histocompatibility complex. *Trends in Ecology & Evolution*. 1990;5(6):181–187.
65. Wakeland EK, Boehme S, She JX, et al. Ancestral polymorphisms of MHC class II genes: Divergent allele advantage. *Immunol Res*. 1990;9(2):115–122.
66. Pierini F, Lenz TL. Divergent Allele Advantage at Human MHC Genes: Signatures of Past and Ongoing Selection. *Mol Biol Evol*. 2018;35(9):2145–2158.
67. Sugimori C, Yamazaki H, Feng X, et al. Roles of DRB1 *1501 and DRB1 *1502 in the pathogenesis of aplastic anemia. *Exp Hematol*. 2007;35(1):13–20.

682 68. Zdimerova H, Murer A, Engelmann C, et al. Attenuated immune control of Epstein–Barr virus
683 in humanized mice is associated with the multiple sclerosis risk factor HLA-DR15. *Eur. J. Immunol.*
684 2021;51(1):64–75.
685 69. Angelini DF, Serafini B, Piras E, et al. Increased CD8+ T cell response to Epstein-Barr virus lytic
686 antigens in the active phase of multiple sclerosis. *PLoS Pathog*. 2013;9(4):e1003220.
687 70. Olsson T. Epstein Barr virus infection and immune defense related to HLA-DR15:
688 consequences for multiple sclerosis. *Eur. J. Immunol.* 2021;51(1):56–59.
689 71. Tengvall K, Huang J, Hellström C, et al. Molecular mimicry between Anoctamin 2 and Epstein-
690 Barr virus nuclear antigen 1 associates with multiple sclerosis risk. *Proc Natl Acad Sci U S A*.
691 2019;116(34):16955–16960.
692 72. Huuhtanen J, Lundgren S, Keränen MA, et al. T Cell Landscape of Immune Aplastic Anemia
693 Reveals a Convergent Antigen-Specific Signature. *Blood*. 2019;134(Supplement_1):108–108.
694 73. Glanville J, Huang H, Nau A, et al. Identifying specificity groups in the T cell receptor
695 repertoire. *Nature*. 2017;547(7661):94–98.
696 74. Brightman SE, Naradikian MS, Miller AM, Schoenberger SP. Harnessing neoantigen specific
697 CD4 T cells for cancer immunotherapy. *J Leukoc Biol*. 2020;107(4):625–633.
698 75. Marty Pyke R, Thompson WK, Salem RM, et al. Evolutionary Pressure against MHC Class II
699 Binding Cancer Mutations. *Cell*. 2018;175(2):416-428.e13.
700 76. Haabeth OAW, Tveita AA, Fauskanger M, et al. How Do CD4(+) T Cells Detect and Eliminate
701 Tumor Cells That Either Lack or Express MHC Class II Molecules? *Front Immunol*. 2014;5:174.
702
703
704
705
706

Figure 1

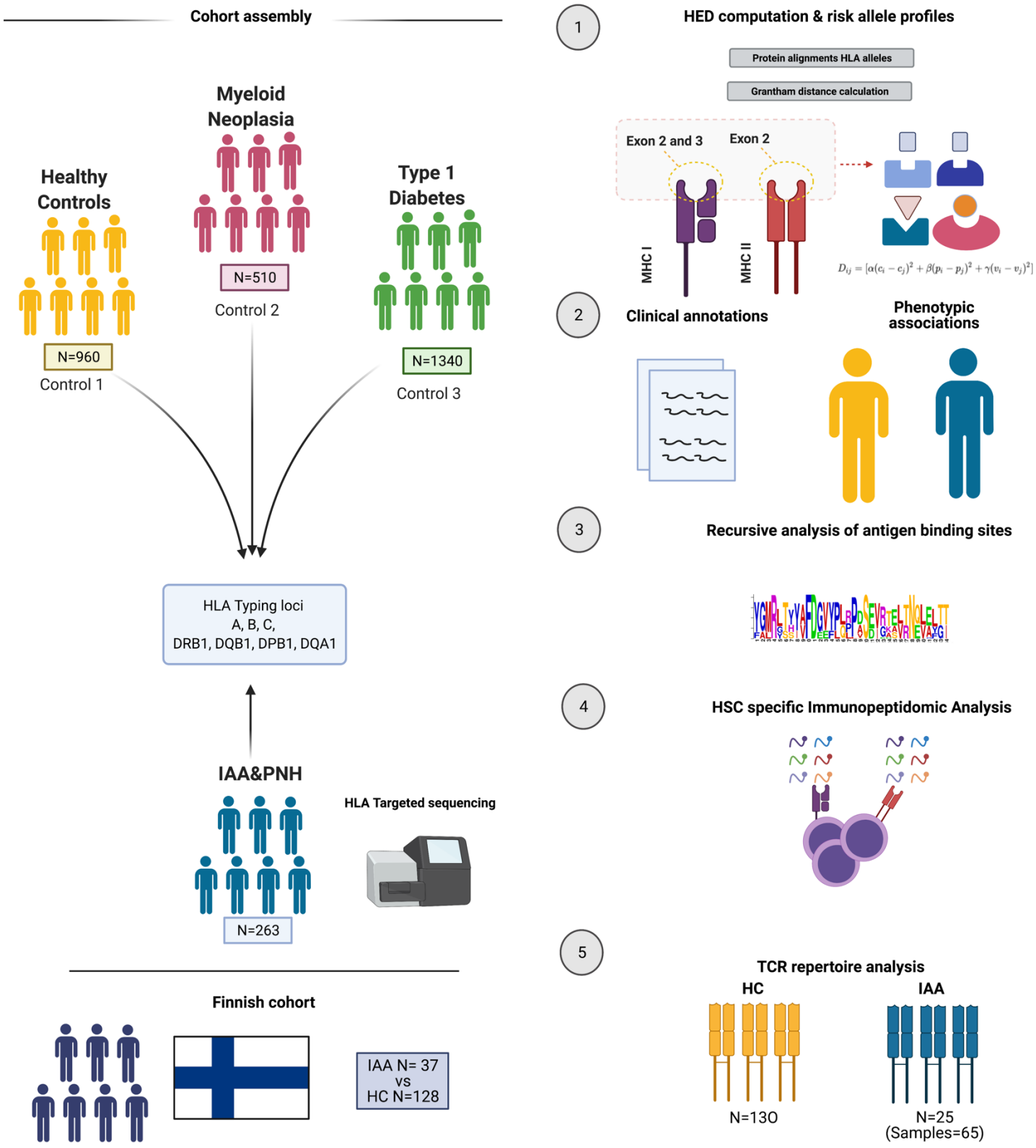
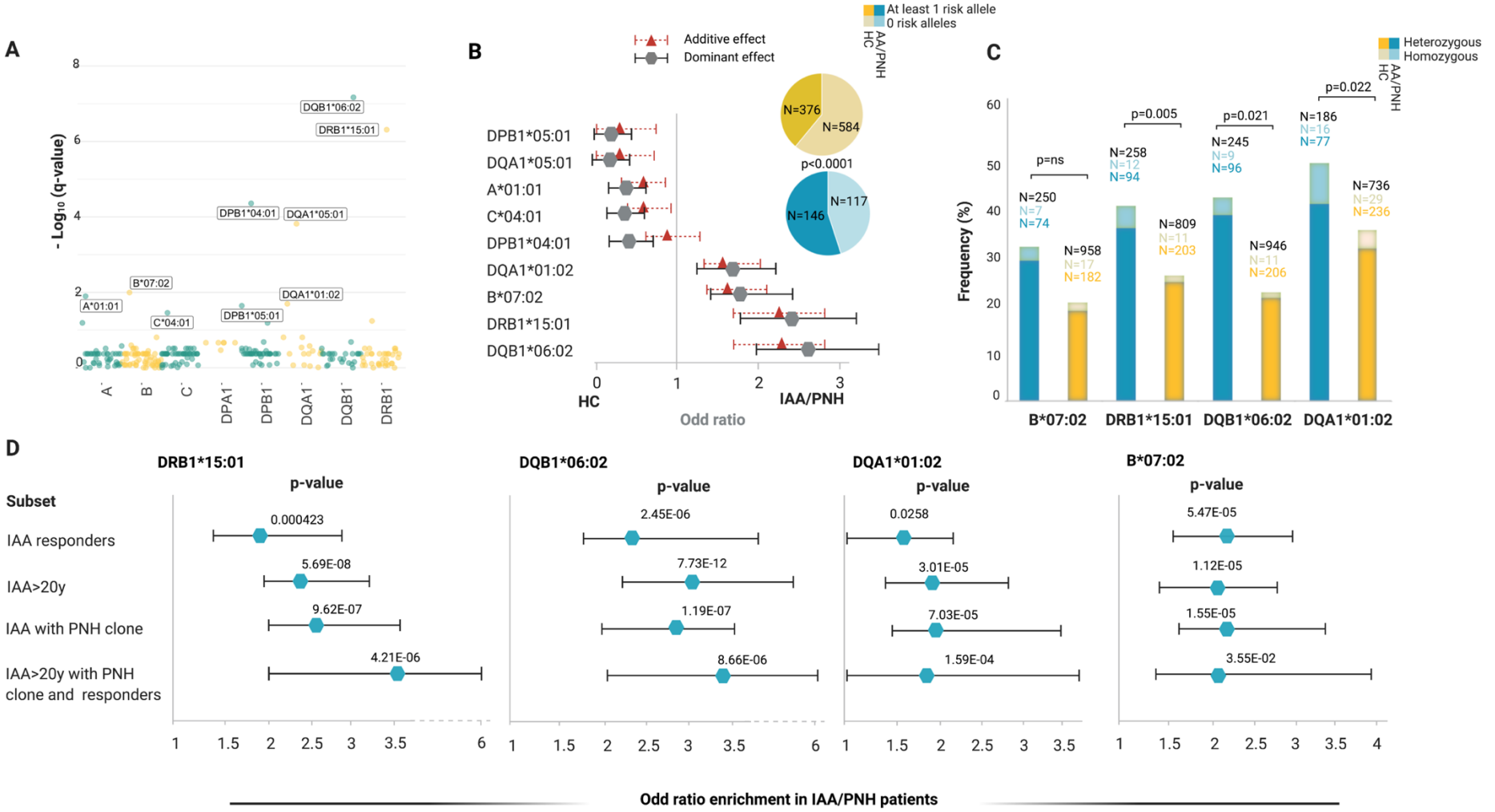


Figure 1: Study design. Sc describing the design and the principal steps of the study.

Figure 2


A
- Log10 (q-value)
DQB1*06:02
DRB1*15:01
DPB1*04:01
DQA1*05:01
B*07:02
A*01:01
DQA1*01:02
C*04:01
DPB1*05:01
A
B
C
DPA1
DPB1
DQA1
DQB1
DRB1
B
Additive effect
Dominant effect
At least 1 risk allele
0 risk alleles
HC
AA/PNH
N=376
N=584
p<0.0001
N=146
N=117
DPB1*05:01
DQA1*05:01
A*01:01
C*04:01
DPB1*04:01
DQA1*01:02
B*07:02
DRB1*15:01
DQB1*06:02
HC
IAA/PNH
Odd ratio
C
Heterozygous
Homozygous
HC
AA/PNH
p=ns
p=0.005
p=0.021
p=0.022
N=250
N=7
N=74
N=958
N=17
N=182
N=258
N=12
N=94
N=809
N=11
N=203
N=245
N=9
N=96
N=946
N=11
N=206
N=186
N=16
N=77
N=736
N=29
N=236
Frequency (%)
B*07:02
DRB1*15:01
DQB1*06:02
DQA1*01:02
D
DRB1*15:01
DQB1*06:02
DQA1*01:02
B*07:02
p-value
Subset
IAA responders
IAA>20y
IAA with PNH clone
IAA>20y with PNH clone and responders
0.000423
5.69E-08
9.62E-07
4.21E-06
2.45E-06
7.73E-12
1.19E-07
8.66E-06
0.0258
3.01E-05
7.03E-05
1.59E-04
5.47E-05
1.12E-05
1.55E-05
3.55E-02
Odd ratio enrichment in IAA/PNH patients

710

Figure 2: Risk allele profile analysis in idiopathic aplastic anemia and paroxysmal nocturnal hemoglobinuria patients

Abbreviations: HED: HLA Evolutionary divergence; MN: myeloid neoplasia; IAA: Idiopathic aplastic anemia; PNH: paroxysmal nocturnal hemoglobinuria; ns: non-significant.

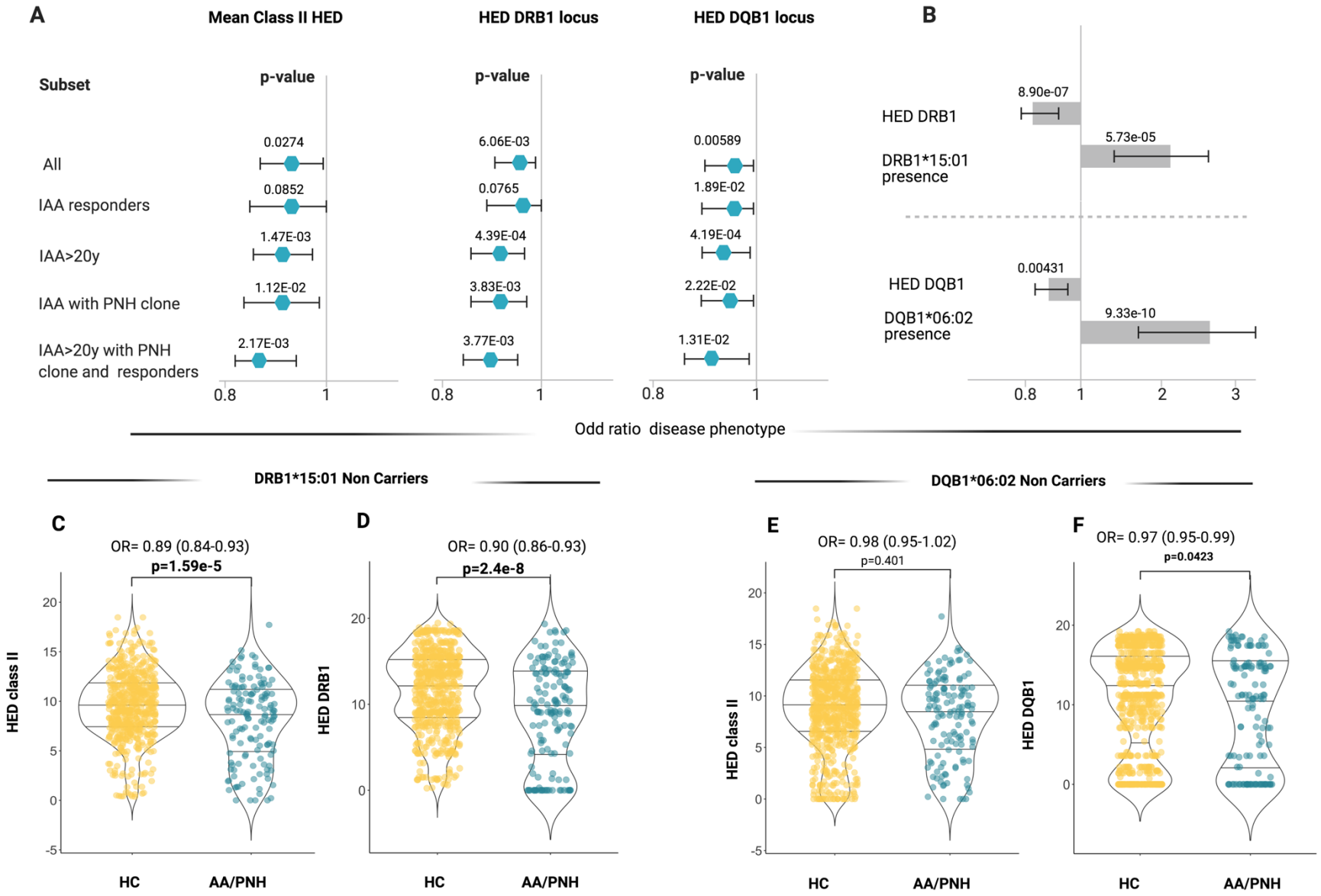
A) Scatterplot representing the negative logarithm base 10 of the adjusted p-values (q-value, Benjamini and Hochberg correction) resulting from the allele association analysis (see methods). Alleles with significantly different genotypic distributions are labeled according to a dominant genetic model.

B) Forest plot reporting the odd ratios (OR) defining estimated effect size of alleles enriched in healthy controls (protective) or in patients (risk). Gray markers describe OR resulting from the analysis of genotypic frequencies (dominant model); red triangles depict the OR deriving from analysis of allelic frequencies (additive model). The pie charts illustrate the distributions of subjects with at least 1 risk allele (darkest colors) or without any risk allele (brighter colors).

C) Barplot depicting the distribution of heterozygous (darkest) and homozygous (brightest) for the 4 risk alleles in controls and patients. Black numbers indicate the total of patients genotyped for the given locus. Colored numbers indicate the number of patients (blue) and controls (yellow) carriers of the risk alleles in heterozygous (darkest) or homozygous (brightest) configurations. Two-sided Fisher test is applied to test the significance of associations with phenotype.

D) Forest plots showing the results of the binomial logistic regression analysis predicting the likelihood of each risk allele association with an aplastic anemia "immune-enriched" phenotype. HC cohort (N=960) was used as comparator group. All (N=263); IAA Responders (N=141); IAA>20 y (N=216); IAA with PNH clones (N=135); IAA>20 y with PNH clone and Responders (N=59).

Figure 3


A
Mean Class II HED
HED DRB1 locus
HED DQB1 locus
Subset
p-value
All
IAA responders
IAA>20y
IAA with PNH clone
IAA>20y with PNH clone and responders
0.0274
0.0852
1.47E-03
1.12E-02
2.17E-03
6.06E-03
0.0765
4.39E-04
3.83E-03
3.77E-03
0.00589
1.89E-02
4.19E-04
2.22E-02
1.31E-02
0.8
1
B
HED DRB1
DRB1*15:01 presence
8.90e-07
5.73e-05
HED DQB1
DQB1*06:02 presence
0.00431
9.33e-10
2
3
Odd ratio disease phenotype
DRB1*15:01 Non Carriers
DQB1*06:02 Non Carriers
C
D
E
F
OR= 0.89 (0.84-0.93)
p=1.59e-5
OR= 0.90 (0.86-0.93)
p=2.4e-8
OR= 0.98 (0.95-1.02)
p=0.401
OR= 0.97 (0.95-0.99)
p=0.0423
HED class II
HED DRB1
HED DQB1
HC
AA/PNH

744

Figure 3: Binomial logistic regression analysis predicting the association between class II HED scores and aplastic anemia phenotypes

Abbreviations: HED: HLA Evolutionary divergence; IAA: Idiopathic aplastic anemia; PNH: paroxysmal nocturnal hemoglobinuria; HC: healthy controls

A) Forest plots showing the results of the binomial logistic regression analysis predicting the likelihood of class II and locus specific HED scores of being associated with an aplastic anemia "immune-enriched" phenotype. All (N=263); IAA Responders (N=141); IAA>20 y (N=216); IAA with PNH clones (N=135); IAA>20 y with PNH clone and Responders (N=59).

B) Multivariable logistic regression analysis testing the independent effect of HED and risk alleles on idiopathic bone marrow failure phenotype. The length of the bars indicates the odd ratio, the error bars show the 95% confident intervals, the numbers on the bars depict the p-values resulting from the likelihood ratio test. Two distinct models are built for DRB1 and DQB1 locus.

C) Violin plots representing the mean class II HED distribution across healthy controls and aplastic anemia patients not carrying DRB1*15:01. Wilcoxon signed rank test was used to calculate the p-value.

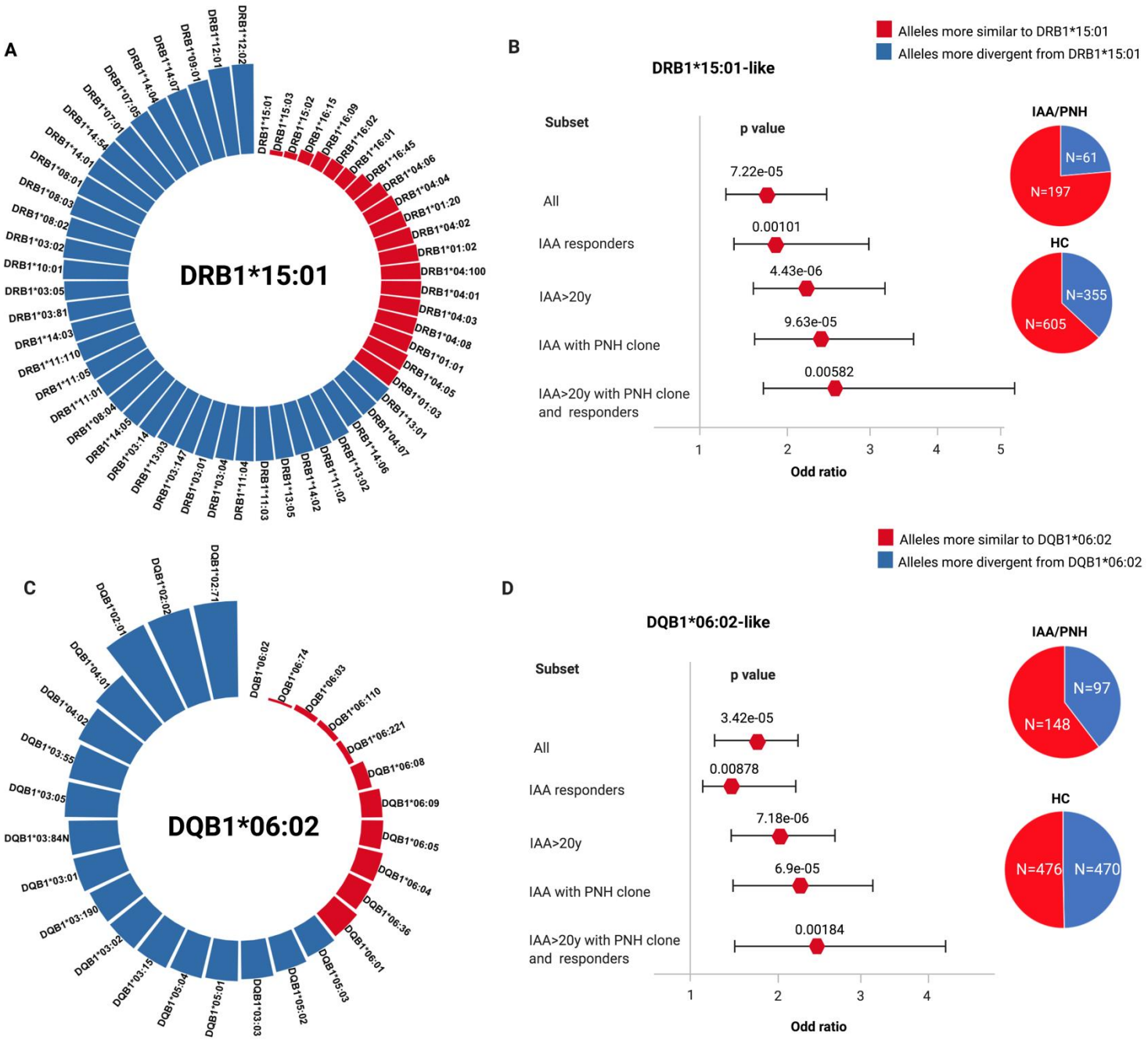
D) Violin plots representing the DRB1 HED distribution across healthy controls and aplastic anemia patients not carrying DRB1*15:01. Wilcoxon signed rank test was used to calculate the p-value.

E) Violin plots representing the mean class II HED distribution across healthy controls and aplastic anemia patients not carrying DQB1*06:02. Wilcoxon signed rank test was used to calculate the p-value.

F) Violin plots representing the DQB1 HED distribution across healthy controls and aplastic anemia patients not carrying DQB1*06:02. Wilcoxon signed rank test was used to calculate the p-value.

Comment on Figure 3C-F: Here our intention was to clarify whether the pattern of lower class II HED seen in IAA/PNH patients was related only to the enrichment in certain risk alleles (such as DRB1*15:01 or DQB1*06:02) or could instead uncover a broader immunogenetic aspect encompassing risk allele profiles. We therefore analyzed HED class II configurations in patients vs controls non carrying those risk alleles and showed that their scores were lower than controls, as a reflex of higher structural similarity between two DRB1 and DQB1 alleles in patients, regardless of the presence of risk genotypes.

Figure 4


A
DRB1*15:01
DRB1*15:01
DRB1*15:03
DRB1*15:02
DRB1*16:15
DRB1*16:09
DRB1*16:02
DRB1*16:01
DRB1*16:45
DRB1*04:06
DRB1*04:04
DRB1*01:20
DRB1*04:02
DRB1*01:02
DRB1*04:100
DRB1*04:01
DRB1*04:03
DRB1*04:08
DRB1*01:01
DRB1*04:05
DRB1*01:03
DRB1*13:01
DRB1*04:07
DRB1*14:06
DRB1*13:02
DRB1*11:02
DRB1*14:02
DRB1*13:05
DRB1*11:03
DRB1*11:04
DRB1*03:04
DRB1*03:01
DRB1*03:147
DRB1*13:03
DRB1*03:14
DRB1*14:05
DRB1*08:04
DRB1*11:01
DRB1*11:05
DRB1*11:110
DRB1*14:03
DRB1*03:81
DRB1*03:05
DRB1*10:01
DRB1*03:02
DRB1*08:02
DRB1*08:03
DRB1*08:01
DRB1*14:01
DRB1*14:54
DRB1*07:01
DRB1*07:05
DRB1*14:04
DRB1*14:07
DRB1*09:01
DRB1*12:01
DRB1*12:02
B
Alleles more similar to DRB1*15:01
Alleles more divergent from DRB1*15:01
DRB1*15:01-like
Subset
p value
All
7.22e-05
IAA responders
0.00101
IAA>20y
4.43e-06
IAA with PNH clone
9.63e-05
IAA>20y with PNH clone and responders
0.00582
1
2
3
4
5
Odd ratio
IAA/PNH
N=61
N=197
HC
N=355
N=605
C
DQB1*06:02
DQB1*06:02
DQB1*06:74
DQB1*06:03
DQB1*06:110
DQB1*06:221
DQB1*06:08
DQB1*06:09
DQB1*06:05
DQB1*06:04
DQB1*06:36
DQB1*06:01
DQB1*05:03
DQB1*05:02
DQB1*03:03
DQB1*05:01
DQB1*05:04
DQB1*03:15
DQB1*03:02
DQB1*03:190
DQB1*03:01
DQB1*03:84N
DQB1*03:05
DQB1*03:55
DQB1*04:02
DQB1*04:01
DQB1*02:01
DQB1*02:02
DQB1*02:71
D
Alleles more similar to DQB1*06:02
Alleles more divergent from DQB1*06:02
DQB1*06:02-like
Subset
p value
All
3.42e-05
IAA responders
0.00878
IAA>20y
7.18e-06
IAA with PNH clone
6.9e-05
IAA>20y with PNH clone and responders
0.00184
1
2
3
4
Odd ratio
IAA/PNH
N=97
N=148
HC
N=476
N=470

790

Figure 4: Simulated structural divergence between each class II risk allele and the pool of alleles present in DRB1 and DQB1 loci in patients and controls.

Abbreviations: HED: HLA Evolutionary divergence; IAA: Idiopathic aplastic anemia; PNH: paroxysmal nocturnal hemoglobinuria; HC: healthy controls

A) Circle graph representing the simulated divergences between DRB1*15:01 and each allele present in DRB1 locus of IAA/PNH patients and HC. Red bars illustrate the alleles more similar to DRB1*15:01 (divergent pairs located under the 25th percentile cutoff of the simulated distribution).

B) Forest plots showing the results of the binomial logistic regression analysis predicting the association between the presence of DRB1*15:01-like alleles and aplastic anemia “immune-enriched” phenotypes. HCs (N=960) were used as comparator group. Pie charts describe the distribution of alleles more similar and more divergent from DRB1*15:01 in patients and controls. All (N=263); IAA Responders (N=140); IAA>20 y (N=216); IAA with PNH clones (N=135); IAA>20 y with PNH clone and Responders (N=59).

C) Circle graph representing the simulated divergences between DQB1*06:02 and each allele present in DQB1 locus of IAA/PNH patients and HC. Red bars illustrate the alleles more similar to DQB1*06:02 (divergent pairs located under the 25th percentile cutoff of the simulated distribution). IAA>20 y (N=216); IAA with PNH clones (N=135); IAA>20 y with PNH clone and Responders (N=59).

D) Forest plots showing the results of the binomial logistic regression analysis predicting the association between the presence of DQB1*06:02-like alleles and an aplastic anemia “immune-enriched” phenotype. HCs (N=960) were used as comparator group. Pie charts describe the distribution of alleles more similar and more divergent from DQB1*06:02 in patients and controls.

Figure 5

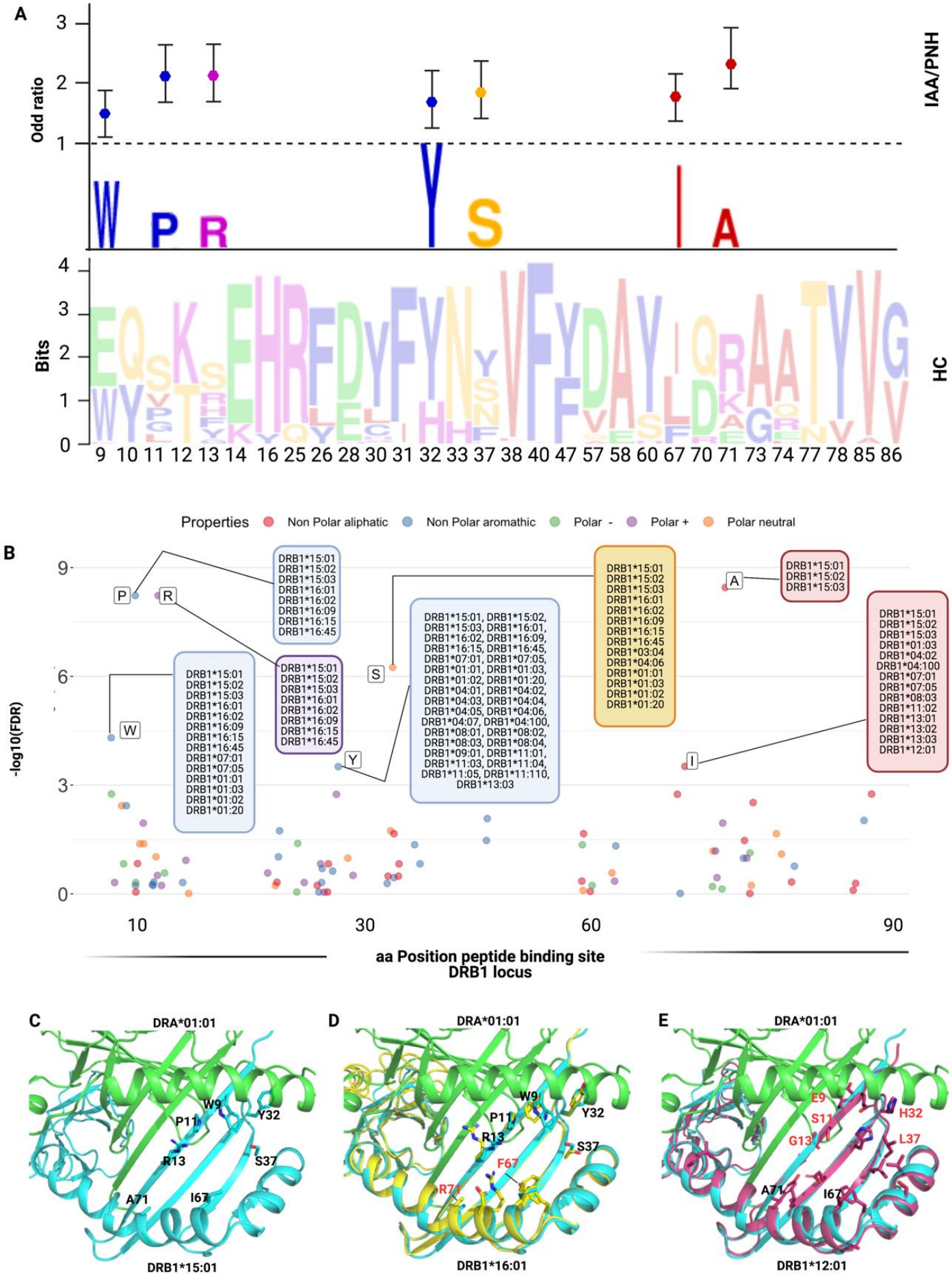

A
Odd ratio
IAA/PNH
Bits
HC
9 10 11 12 13 14 16 25 26 28 30 31 32 33 37 38 40 47 57 58 60 67 70 71 73 74 77 78 85 86
B
Properties
Non Polar aliphatic
Non Polar aromathic
Polar -
Polar +
Polar neutral
-log10(FDR)
aa Position peptide binding site
DRB1 locus
C
DRA*01:01
W9
P11
Y32
R13
S37
A71
I67
DRB1*15:01
D
DRA*01:01
W9
P11
Y32
R13
S37
F67
R71
DRB1*16:01
E
DRA*01:01
E9
S11
H32
G13
L37
A71
I67
DRB1*12:01

Figure 5: Recursive analysis of the amino acid sequence within the peptide binding site of DRB1 locus

Abbreviations: IAA: Idiopathic aplastic anemia; PNH: paroxysmal nocturnal hemoglobinuria; HC: healthy controls; aa: amino acid

A) Lower panel: WebLogo visualization representing the contribution of single aminoacids within the variable portion of the peptide binding site of DRB1 locus. The x-axis indicates each variable position (as per IPD-IMGT-HLA reference). Letters represent each possible amino acid at each given position; Letters' height illustrates the frequency of each amino acid in healthy control population. Colors indicate the chemico-physical properties as per legend in B.
Upper panel: stylized visualization of amino acids differentially distributed between HC and IAA/PNH cohorts. Letters' height illustrates the frequency of each amino acid in disease population. Markers indicate the odd ratios resulting from the logistic regression analysis studying each aminoacidic contribution in determining the phenotype (see methods and Table S7).

B) Scatter plot showing the significance level of each variable amino acid in the peptide binding site of DRB1 locus found enriched in IAA/PNH population compared to HC. Each dot represents the negative logarithm base 10 for the adjusted p-value (q-value) referring to each amino acid. The position on x-axis indicate the position within the peptide binding site according to IPD-IMGT-HLA reference). Only the amino acids presenting a q-value<10-e4 are considered significant for this analysis and labelled in the figure. Alleles presenting the indicated amino acid at the given position are indicated in the boxes. Colors represent the chemico-physical properties as per legend.

C) Crystallographic structure showing the position of the 7 amino acids significantly enriched in the peptide binding groove of DR molecules of IAA/PNH patients. This 3D structure has been visualized with the PyMOL program based on the structure of DRB1*15:01-DRA*01:01-myelin binding protein (PDB:1BX2) and HLA sequences retrieved from IPD-IMGT/HLA database v. 3.40. Only the 7 amino acids identified in the previous analysis have been highlighted.

D) Binding site of DRB1*16:01-DRA*01:01 based on the homology model of DRB1*16:01. The structure of DRB1*16:01 has been superimposed to DRB1*15:01. Residues differing from the risk pattern seen in DRB1*15:01 are colored in red.

E) Binding site of DRB1*12:01-DRA*01:01 based on the homology model of DRB1*12:01. The structure of DRB1*12:01 has been superimposed to the structure of DRB1*15:01. Residues differing from the risk pattern seen in DRB1*15:01 are colored in red.

Figure 6

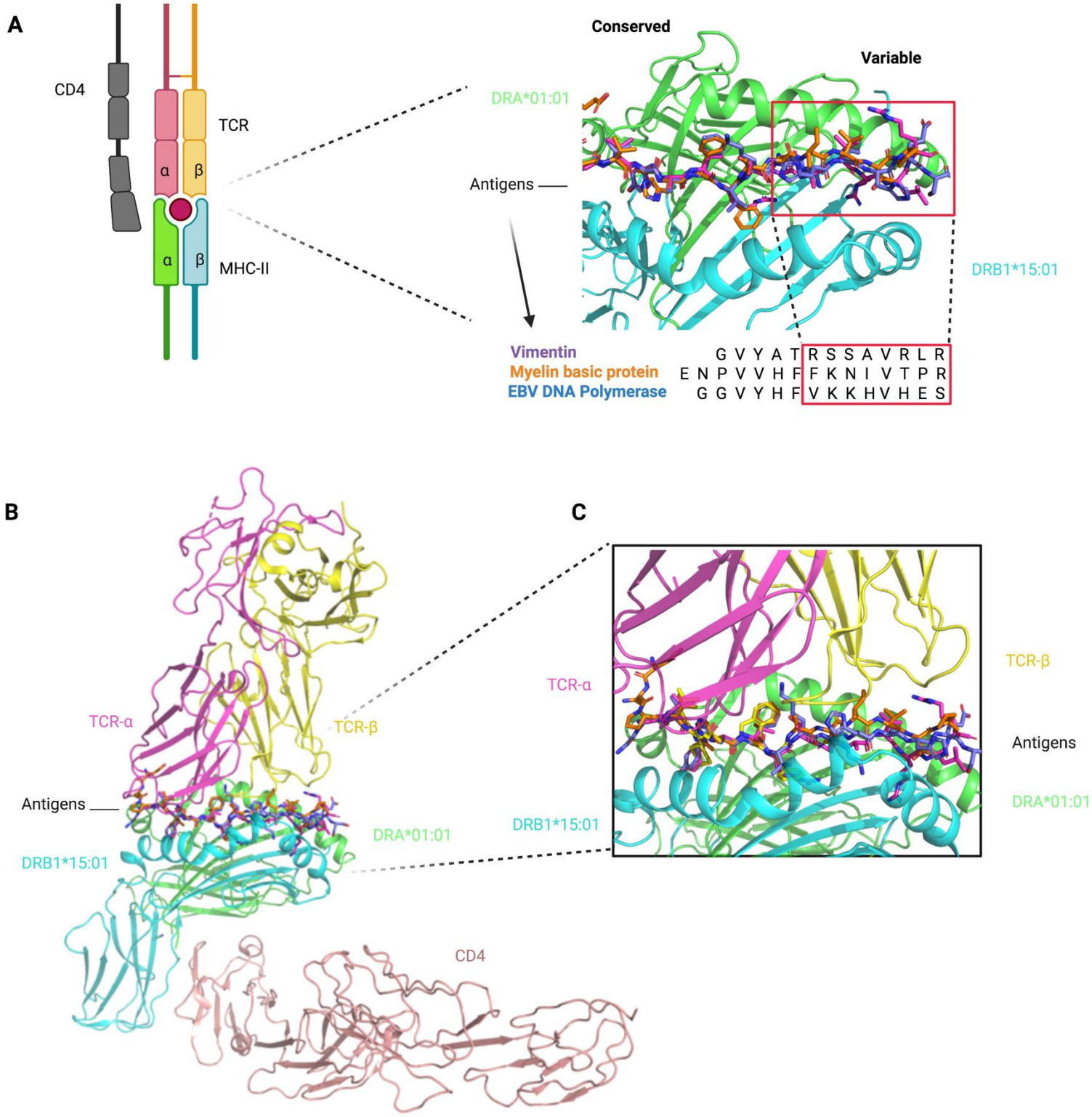

A
CD4
TCR
α
β
α
β
MHC-II
DRA*01:01
Antigens
Conserved
Variable
DRB1*15:01
Vimentin
Myelin basic protein
EBV DNA Polymerase
G V Y A T R S S A V R L R
E N P V V H F F K N I V T P R
G G V Y H F V K K H V H E S
B
TCR-α
TCR-β
Antigens
DRA*01:01
DRB1*15:01
CD4
C
TCR-α
TCR-β
Antigens
DRB1*15:01
DRA*01:01

Figure 6: Structural insight into DRB1-antigen-TCR interactions

Abbreviations: TCR: T-cell receptor; MHC: major histocompatibility complex, MBP: myelin binding protein; EBV: Epstein barr virus

A) Peptides of vimentin (UniProt: P08670, VIME_HUMAN, 59-71), EBV DNA polymerase (UniProt: P03198, DPOL_EBVB9, 628-641), MBP (UniProt: P02686, MBP_HUMAN, 217-231) at the HLA binding site based on the alignment of crystal structures of DRB1*15:01-DRA*01:01/MBP (217-231) (PDB: 1BX2), DRB1*14:02-DRA*01:01/ vimentin (amino acid positions: 59-71, PDB:1H15) and DRB5*01:01-DRA*01:01/EBV DNA polymerase (628-641) (PDB:6ATF). The red squares indicate the peptide portions presenting with more conformational variability (in interaction with the right site of the HLA binding groove). This 3D structure has been prepared with PyMOL using the crystal structure of DRB1*15:01-DRA*01:01/MBP (PDB: 1BX2).

B) Modeled ternary structure of HLA-Antigen-TCR-CD4.

C) Detail of the interaction interface in HLA-antigen-TCR. The three antigenic structures are aligned as shown above. The risk amino acid pattern within the right side of the binding groove interacts with a more variable antigenic portion that contacts directly with the TCR beta chain (software PyMOL).

Figure 7

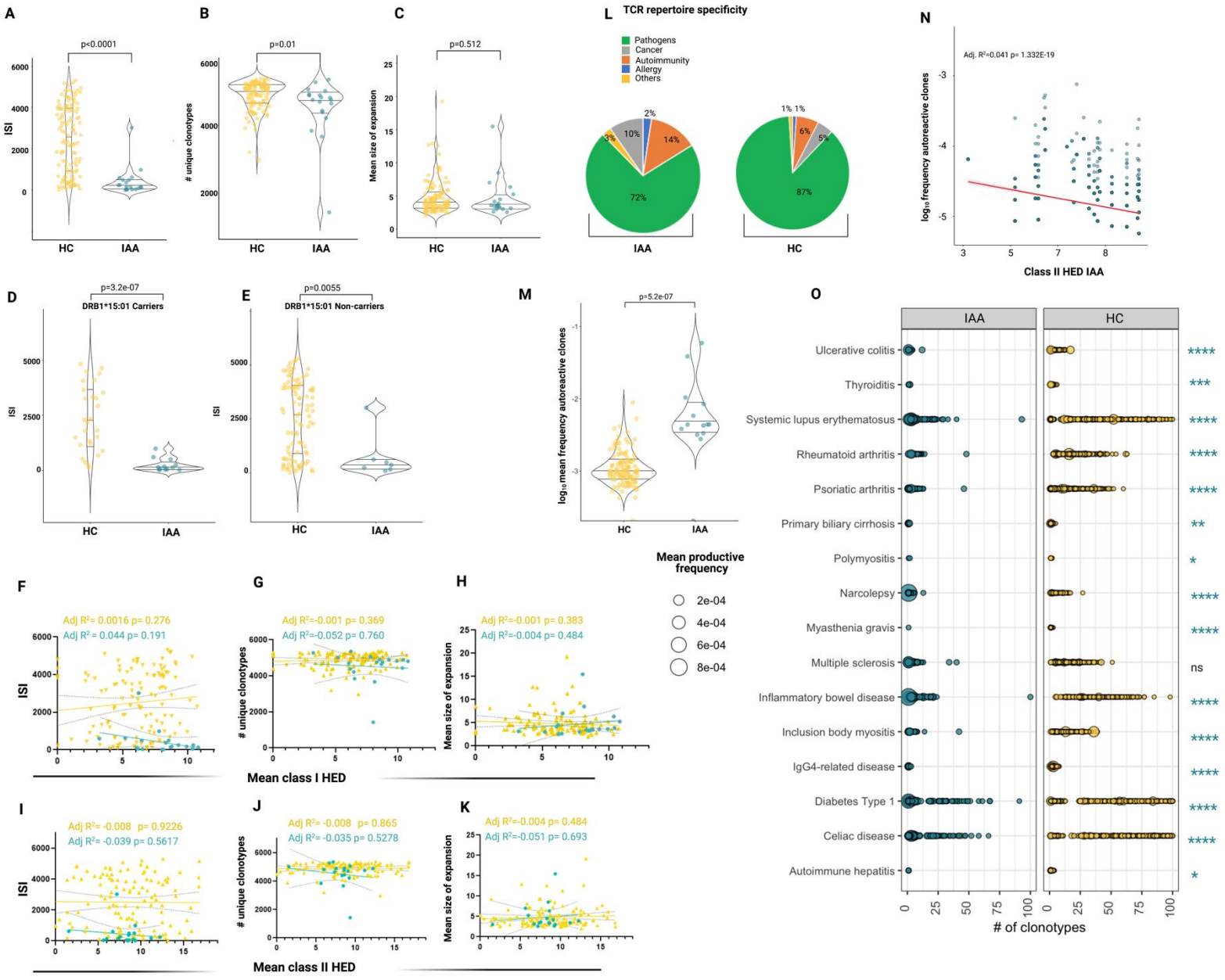

A
p<0.0001
ISI
HC
IAA
B
p=0.01
# unique clonotypes
C
p=0.512
Mean size of expansion
L
TCR repertoire specificity
Pathogens
Cancer
Autoimmunity
Allergy
Others
2%
10%
3%
14%
72%
1% 1%
6%
5%
87%
N
Adj. R²=0.041 p= 1.332E-19
log10 frequency autoreactive clones
Class II HED IAA
D
p=3.2e-07
DRB1*15:01 Carriers
E
p=0.0055
DRB1*15:01 Non-carriers
M
p=5.2e-07
log10 mean frequency autoreactive clones
Mean productive frequency
2e-04
4e-04
6e-04
8e-04
O
IAA
HC
Ulcerative colitis
Thyroiditis
Systemic lupus erythematosus
Rheumatoid arthritis
Psoriatic arthritis
Primary biliary cirrhosis
Polymyositis
Narcolepsy
Myasthenia gravis
Multiple sclerosis
Inflammatory bowel disease
Inclusion body myositis
IgG4-related disease
Diabetes Type 1
Celiac disease
Autoimmune hepatitis
****
***
****
****
****
**
*
****
****
ns
****
****
****
****
****
*
# of clonotypes
F
Adj R²= 0.0016 p= 0.276
Adj R² = 0.044 p= 0.191
G
Adj R²=-0.001 p= 0.369
Adj R²=-0.052 p= 0.760
H
Adj R²=-0.001 p= 0.383
Adj R²=-0.004 p= 0.484
Mean class I HED
I
Adj R²= -0.008 p= 0.9226
Adj R²= -0.039 p= 0.5617
J
Adj R²= -0.008 p= 0.865
Adj R²= -0.035 p= 0.5278
K
Adj R²=-0.004 p= 0.484
Adj R²=-0.051 p= 0.693
Mean class II HED

Figure 7: T cell receptor analysis in IAA patients

Abbreviations: IAA: Idiopathic aplastic anemia; HC: healthy controls; aa: amino acid; ISI: Inverse simpson index

A) ISI distribution in HC and IAA patients (downsampled dataset). Violin plots showing median and interquartile ranges. Wilcoxon signed rank test.
B) Number of unique clonotypes in HC and IAA patients (downsampled dataset). Violin plots showing median and interquartile ranges. Wilcoxon signed rank test.
C) Mean size of expansion of each clonotype of size $\geq$2 templates in HC and IAA patients (down-sampled dataset). Violin plots showing median and interquartile ranges. Wilcoxon signed rank test.
D) ISI distribution in HC and IAA patients in DRB1*15:01 carriers (down-sampled dataset). Violin plots showing median and interquartile ranges. Wilcoxon signed rank test.
E) ISI distribution in HC and IAA patients in non-DRB1*15:01 carriers (down-sampled dataset). Violin plots showing median and interquartile ranges. Wilcoxon signed rank test.
F) Linear regression analysis between ISI and mean class I HED.
G) Linear regression analysis between number of unique clonotypes and mean class I HED
H) Linear regression analysis between mean size of clonotype expansion and mean class I HED.
I) Linear regression analysis between ISI and mean class II HED.
J) Linear regression analysis between number of unique clonotypes and mean class II HED.
K) Linear regression analysis between mean size of clonotype expansion and mean class II HED. R-squared goodness-of-fit are reported along with the p-value in each box (from F to K).
L) Proportion of known complementary determining region 3 (CDR3) specificities in IAA and HC groups (this distribution has been computed in the downsampled dataset).
M) Negative logarithm of mean frequency of autoreactive clonotypes present in HC and IAA patients. Violin plots showing median and interquartile ranges. Each dot represent the mean frequency/per subject (All the values refer to the non-downsampled dataset in order to capture all the possible recognizable CDR3 sequences). Wilcoxon signed rank test
N) Linear regression analysis between frequency of known autoreactive clonotypes in IAA patients and mean class II HED. Each dot represents one autoreactive clonotype. Clonotypes with overlapping frequencies are represented by darkest dots. (All the values refer to the non-downsampled dataset). Wilcoxon signed rank test. R-squared goodness-of-fit are reported along the p-value.
O) Bubble matrix showing the mean frequency of each autoimmune-disease associated clonotype present in the non-downsampled repertoires of IAA and HCs. Each bubble represents the number of clonotypes with known autoimmune specificity (X-axis). The size of each bubble indicates the mean frequency. Wilcoxon signed rank test is used to compare the mean frequencies of each specificity between IAA and HC groups.

Table 1: Patient characteristics at diagnosis[$]		
Category	**Sub-category**	**N (%)/median (IQR)**
All		263
Age (years)		44 (27-62)
Ethnicity	**Caucasian**	214 (84%)
	African American	22 (8%)
	Asian	9 (4%)
	Hispanic or latino	7 (3%)
	Other	2 (2%)
	Unknown	9 (3%)
Gender	**Female**	132 (50.2%)
	Male	131 (49.8%)
Disease phenotype	**IAA +/- non-hemolytic PNH clone**	216 (82%)
	IAA + hemolytic PNH clone	13 (5%)
	Primary hemolytic PNH	34 (13%)
PNH granulocytic clonal size		5.6 (1-37.2)
Severity IAA	**Severe**	166 (71%)
	Moderate	66 (28%)
Treatment characteristics		
Category	**Sub-category**	**N (%)**
First Line Treatment (except allo-HCT)*	ATG/CSA	129 (53%)
	ATG/CSA/Eltrombopag	14 (6%)
	CSA	22 (9%)
	CSA/Eltrombopag	8 (3%)
	CSA/other	5 (2%)
	Androgens	10 (4%)
	Anti-complement	26 (11%)
	Eltrombopag	3 (1%)
	None	14 (6%)
	Other	11 (5%)
Allo-HCT (all lines)		46 (17%)
N° Lines**	0	17 (8%)
	1	96 (46%)
	2	53 (25%)
	3	26 (13%)
	4	13 (6%)
	≥5	3 (14%)

Abbreviations: IQR: Interquartile range, IAA: Idiopathic aplastic anemia, PNH: Paroxysmal Nocturnal Hemoglobinuria; IST: Immunosuppressive treatment, Allo-HCT: allogeneic hematopoietic stem cell transplant, CSA: Cyclosporine, ATG: Anti-thymocyte globuline.

* Data on 242 patients. ** Data on 208 patients.

[$]Data reported for the Cleveland Clinic cohort only. For patient characteristics of the Finnish cohort see Table S23.

953
954

Table 2: Outcome description[$]		
Response to IST* (first line)	**CR/PR**	141 (69%)
	NR	62 (30%)
Secondary PNH from IAA	**N (%)**	28 (10%)
Time to secondary PNH	**median (IQR)**	42.2 (23.4-80.4)
CIF of secondary PNH at 10 years	**Probability (95%CI)**	12.3 (7.8 - 17.8)
Progression to AML/MDS	**N (%)**	32 (12%)
CIF of progression to AML/MDS at 10 years	**Probability (95%CI)**	14.7 (10.5 - 19.6)
Time to progression AML/MDS (months)	**median (IQR)**	44.6 (24.3-77.8)
Follow-up (months)	**median (IQR)**	85 (40-153)

955
956 **Abbreviations:** IQR: Interquartile range, IAA: Idiopathic aplastic anemia, PNH: Paroxysmal Nocturnal
957 Hemoglobinuria; AML: Acute Myeloid Leukemia; MDS: Myelodysplastic syndrome; IST: Immunosuppressive
958 treatment, CIF: Cumulative incidence function, CI: confident interval * Data on 203 patients.
959 [$]Data reported for the Cleveland Clinic cohort only.
960