Autoantibody Repertoires in Healthy Individuals Vary According to HLA Class II Genotype

Rayo Suseno[1,*], Juliano A. Boquett[1,*], Ravi Dandekar[1], Thomas Ituarte[1], Bonny Alvarenga[1], Cynthia Vierra-Green[2], Stephen Spellman[2], Martin Maiers[2], Michael Wilson[1], Jill A. Hollenbach[1,3,✉]

[1] Department of Neurology, University of California, San Francisco, United States;

[2] Center for International Blood and Transplant Research (CIBMTR) and NMDP, Minneapolis, MN, United States;

[3] Department of Epidemiology and Biostatistics, University of California, San Francisco, United States

[*] Equal contributors

✉ Corresponding author: Jill A. Hollenbach (jill.hollenbach@ucsf.edu)

Abstract

The link between *HLA* genotype and the presence of pathogenic autoantibodies has previously been established across different autoimmune disorders. However, the functional process linking specific antibodies to specific *HLA* remains unclear. Additionally, autoantibodies – usually associated with autoimmune disease – are also present in healthy individuals. To this end, we sought to determine the spectrum of self-antigen antibody specificity in healthy individuals, stratified by *HLA* genotypes. We utilized Phage ImmunoPrecipitation Sequencing (PhIP-Seq), a programmable phage display for interrogation of antibody specificity, which encompasses over 700,000 peptides tiled across the entire human proteome. Serum from 741 donors without diagnosed autoimmune disease were grouped by their five homozygous *HLA-DRB1* genotypes and analyzed for differences in autoantibody reactivity. We applied a custom filter to the PhIP-Seq normalized data and obtained a set of enriched peptides for each *HLA-DRB1* genotype. We found that autoantibody profiles in healthy individuals with different *HLA-DRB1* genotypes are generally distinct. Binary logistic regression successfully identified whether a sample belongs to a specific *HLA-DRB1* genotype (*HLA-DRB1*01, HLA-DRB1*03, HLA-DRB1*04, HLA-DRB1*07, and HLA-DRB1*15*) with the following accuracies: 96%, 92%, 90%, 94%, and 90%, and multinomial models predicted *HLA-DRB1* genotype with up to 90% accuracy. Finally, gene-level analysis suggests that individuals with specific *HLA* autoimmune risk alleles may harbor potentially pathogenic autoantibodies in the absence of, or prior to the establishment of, overt disease. Our analysis demonstrates that autoantibody repertoires in healthy people vary according to *HLA* class II genotype, and may provide insight into the pathological processes associated with development of autoimmune and other immune-mediated diseases.

Introduction

Located on human chromosome 6, Human Leukocyte Antigen (*HLA*) genes encode cell surface receptors that present antigens to T cells, playing a pivotal role in the adaptive immune response[1]. *HLA* loci are the most polymorphic of the human genome, and this remarkable variation leads to distinctive peptide-binding repertoires, mediating key differences in antigen presentation and differentially eliciting immune responses[2]. Accordingly, *HLA* variation has been associated to a multitude of diseases, ranging from autoimmune, infection, neurological, to cancer phenotypes[3–8].

Patterns of antibody response and the specificity of antibody epitope recognition in both infectious and autoimmune disease have driven the search for antigen bound by *HLA* and presented to T cells, in the interest of exploration of disease triggers, therapeutics, and vaccine development. B cells function as professional antigen presenting cells and are capable of accessing antigen for phagocytosis through the B cell receptor (BCR), a bound immunoglobulin (Ig). Evidence suggests that the BCR itself may be presented by *HLA* molecules and therefore be recognized as epitopes by T cells[9]. Data also suggests that there exists operative T-cell-mediated immunosurveillance with respect to follicular lymphoma cells, and that this control is mediated by the strength of binding of BCR-derived peptides to *HLA*[9]. Thus**,** the peripheral BCR repertoire may be shaped by *HLA* alleles in healthy individuals and controlled by T-cell mediated recognition of BCR peptides.

It has been shown that *HLA* class II modulates antibody response to both viral and bacterial antigens in vaccination[10–13]. Furthermore, there is evidence for a link between *HLA* class II genotype and the presence of specific autoantibodies across multiple autoimmune disorders[2]. The presence of autoantibodies may precipitate initiation of autoimmune disease prior to the formation of immune complexes and subsequent pathology[14]. While autoantibodies are known to be present in individuals with autoimmune disease, they are also found in healthy individuals[15]. However, there is a lack of understanding of the relationship between *HLA*

variation and the presence of specific autoantibodies. Here, we examined the extent to which *HLA* genotype influences the autoantibody repertoire in healthy individuals, and show that distinct autoantibody profiles are predictive of HLA class II genotype.

Methods

Subjects

Our analysis included serum samples from 741 healthy individuals who volunteered as hematopoietic stem cell donors stored in the NMDP (formerly National Marrow Donor Program) Research Sample Repository (ClinicalTrials.gov protocol NCT04920474). All individuals consented to research and publication of research results and had been previously genotyped for *HLA* class I and class II. These were convenience samples selected based on homozygosity for *HLA-DRB1*, and where possible across the class II region, in order to reduce confounding in the interpretation of the results. Detailed donor demographic information is shown in Supplemental Tables 1 and 2.

The donors included in the study presented a total of 14 *HLA-DRB1* high-resolution (two-field) genotypes. In order to increase statistical power, individuals that shared *HLA* genotype at the first field were merged as a single group (Figure 1). For instance, *HLA-DRB1*15:01* and *HLA-DRB1*15:03* were grouped into a single *HLA-DRB1*15* group. In this manner, five common *HLA-DRB1* allelic groups were chosen as core analysis groups based on sufficient sample size. With a count of 192 subjects, *HLA-DRB1*15* represents the largest core group. The four other groups were *HLA-DRB1*07:01*, *HLA-DRB1*04* (*HLA-DRB1*04:01*, *HLA-DRB1*04:04*, and *HLA-DRB1*04:07*), *HLA-DRB1*03:01*, and *HLA-DRB1*01* (*HLA-DRB1*01:01* and *HLA-DRB1*01:02*), with sample sizes of 123, 69, 148, and 96 respectively. The remaining 113 samples were included in the analysis as additional samples to compare against the core groups, which include *HLA-DRB1*11:01*, *HLA-DRB1*11:04*, *HLA-DRB1*13:01*, *HLA-DRB1*13:02*, and *HLA-DRB1*08:02*.

Phage Immunoprecipitation Sequencing (PhIP-Seq)

We employed a programmable phage display assay comprised of 731,000 peptides tiled across the human proteome, representing the entire human peptidome, termed PhIP-Seq[17,18]. To generate the human peptidome, we computationally divided the entire human proteome, including all known structural and splice variants, into oligonucleotide sequences encoding 49 amino acid peptides with 24-residue overlap. The phage library was generated using classic molecular cloning such that individual phage particles contain 5-15 copies of a single human peptide expressed on the surface capsid protein[19,20].

PhIP-Seq was performed using a high-throughput protocol, which starts with deep-well polypropylene 96-well plates, where the phage library was exposed to 1 µl of human sera in 1:1 storage buffer. The deep-well plates with library and sample were incubated overnight at 4°C to facilitate antibody-phage binding. Beads were added to each reaction well and a sequence of washes were performed. The immunoprecipitated solution was resuspended and amplified. Filtered solution was transferred to a new pre-blocked deep-well plate where individuals' sera was added and subjected to another round of immunoprecipitation and amplification, followed by next-generation sequencing (NGS) library prep. Phage DNA from each sample was barcoded and amplified (Phusion PCR, 18 rounds) and subjected to NGS on an Illumina NovaSeq instrument at an average read depth of 1 million reads per sample. Fully detailed PhIP-Seq protocol is described elsewhere[19,20].

Binding quantification and peptide linear ranking

Raw sequencing reads from the PhIP-Seq assay were subsequently aligned to a reference library using Bowtie2 to obtain the read count for each sample to each peptide[20,21]. Normalized read counts were then calculated by multiplying the read proportion of each peptide

in a sample with 100,000. This operation transformed the raw read count to number of reads per kilo (RPK), meaning that the normalized reads would sum up to 100,000 across the human peptidome. To normalize samples against background, fold change (FC) over mock-IP was calculated by dividing the normalized read by the mean percentage of all bead-only controls in the same peptide. Once the RPK and FC have been calculated, we applied a set of filters (MIN_RPK = 0, FC_THRESH1 = 10, FC_THRESH2 = 100, SUM_RPK_THRESH = 50) intended to reduce noise for downstream analysis, yielding us with 102,716 peptides. The data were then transformed into binary hits, using another set of filters to ensure that the gene in which the peptide reside needs to have 1) a summed RPK of 50 or above, and 2) a FC of 100 or above.

Autoantibody reactivity analysis

To determine which autoantibodies were statistically significantly different in their reactivity between one genotype group against the others, we used the Kruskal-Wallis rank sums test[22]. To adjust for multiple comparisons, Bonferroni correction was applied such that a p-value threshold of 0.05/6 is enforced. There are a total of six in our analysis: the five primary genotype groups (*HLA-DRB1*15*, *HLA-DRB1*07*, *HLA-DRB1*04*, *HLA-DRB1*03*, and *HLA-DRB1*01*), and one additional group representing all other genotypes.

HLA genotype prediction from autoantibody profiles

To test whether *HLA* genotype can be predicted from autoantibody profiles, we employed multinomial and binary logistic regression. The multinomial model predicts whether a sample belongs to one of five genotype groups: *HLA-DRB1*15*, *HLA-DRB1*07*, *HLA-DRB1*04*, *HLA-DRB1*03*, or *HLA-DRB1*0*1. The binary logistic regression model predicts whether a sample has a specific *HLA-DRB1* genotype. Thus, a total of five binary logistic models were employed: *HLA-DRB1*15* vs rest, *HLA-DRB1*07* vs rest, *HLA-DRB1*04* vs rest, *HLA-DRB1*03* vs rest, and *HLA-DRB1*01* vs rest.

All regression models were performed using only peptides that were significantly different according to the Kruskal-Wallis rank sums test and are exclusive to each genotype group. Due to sample size imbalance between each group against the rest of the samples, we balanced the model during the training process by under-sampling groups that are overrepresented. This was done by using the *class_weight = "balanced"* argument in Scikit-learn's LogisticRegression function[23].

For each binary logistic regression, we ran five-fold cross validation to avoid overfitting. Three metrics that we reported to assess the model are accuracy, AUC-ROC (Area Under the Receiver Operating Characteristic Curve), and F1 score. The accuracy can be calculated by dividing the true positive (TP) and true negative (TN) with the total samples ($\frac{TP+TN}{TP+TN+FP+FN}$). The AUC-ROC was computed by plotting the true positive rate (TPR) against the false positive rate (FPR) across different thresholds and then calculating the area under this curve to assess the overall model discrimination. The F1 score is particularly informative because it reveals if class imbalance is causing the model to disproportionately predict one class. The F1 score for a particular class was calculated using Precision ($\frac{TP}{TP+FP}$) and Recall ($\frac{TP}{TP+FN}$) with the following formula: $F1_{class} = \frac{2*Precision_{class}*Recall_{class}}{Precision_{class}+Recall_{class}}$, where a class is either *HLA-DRB1*01, HLA-DRB1*03, HLA-DRB1*04, HLA-DRB1*07*, or *HLA-DRB1*15*. All three metrics were calculated using Scikit-learn[23].

To understand what drives the model's decision, we ranked the binary logistic regression coefficients to find features (here, specific autoantibodies) that are the most influential. We ranked the absolute value of the coefficients to better understand the magnitude of the features, while keeping track of its direction. Since the model ran through five-fold cross-validation, we rank the coefficients based on the average coefficients across five iterations.

Multidimensionality reduction analysis

We performed the Uniform Manifold Approximation and Projection (UMAP)[24] on the five main *HLA-DRB1* genotype groups. A total of 9,358 peptides were used in this UMAP, consisting of peptides that passed through the Kruskal-Wallis rank sums test and were exclusive to each genotype group. The binary table is first scaled using the StandardScaler() function so that each feature has zero mean and unit variance. Subsequently, the following arguments were used within Python's UMAP function: n_neighbors = 15, min_dist = 0.0, n_components = 2, metric = cosine.

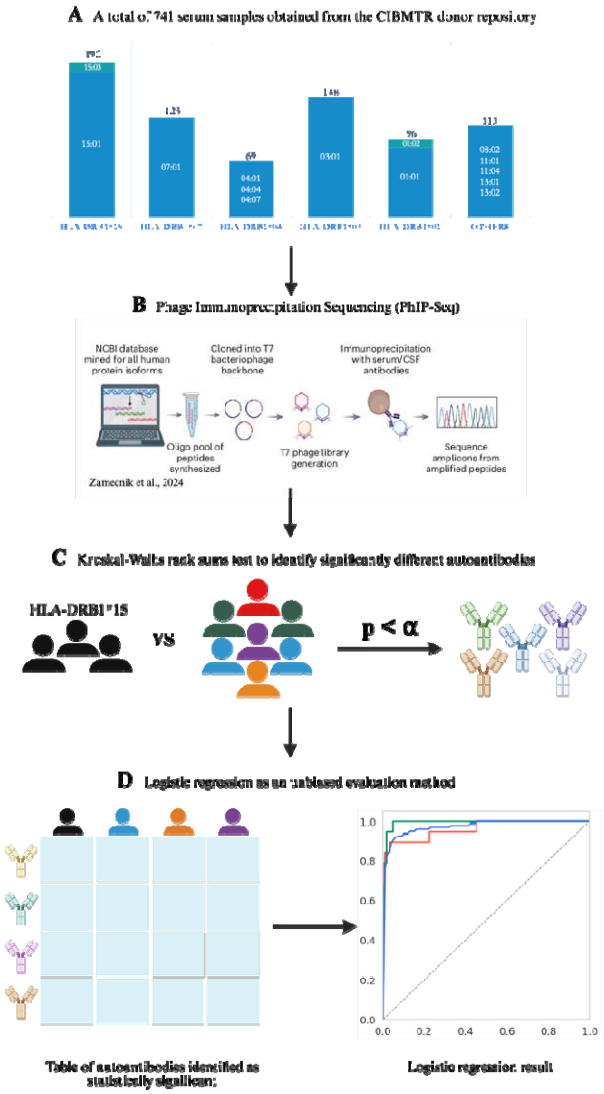

Figure 1: An outline of this study's design. **A)** serum retrieval of healthy individuals from the CIBMTR donor repository, **B)** Phage Immunoprecipitation Sequencing (PhIP-Seq) performed on the CIBMTR healthy cohort, **C)** Kruskal-Wallis rank sums test to identify autoantibodies that have different reactivity among different genotypes, **D)** logistic regression as means of evaluation.

Results

Autoantibody profiles vary by HLA-DRB1 genotype

Using the Kruskal-Wallis rank-sum test to compare autoantibody reactivity between each *HLA-DRB1* group versus all others (*HLA-DRB1*01*, *HLA-DRB1*03*, *HLA-DRB1*04*, *HLA-DRB1*07*, and *HLA-DRB1*15* vs. the rest), we identified 1,029, 1,297, 5,691, 909 and 453 self-peptides, respectively, that were significantly bound by donor serum antibodies in each group relative to the others. Most autoantibodies identified in specific *HLA-DRB1* groups are private, with very few autoantibodies identified in two or more *HLA-DRB1* groups (Figure 2A). This demonstrates that the autoantibody profile in healthy individuals with different *HLA-DRB1* genotypes are generally distinct.

Considering that the peptidome library was built with tiled peptides with 24-mer overlap, it is expected that individuals sharing an *HLA-DRB1* genotype with autoantibodies to a specific peptide may also have antibodies specific to adjacent, overlapping peptides from the same protein. As an example, *ESPN_526263* and *ESPN_526264*, both originating from the *ESPN* gene, are adjacent to one another and both react to *HLA-DRB1*15* samples. To capture protein-level information from our analysis, we merged all peptides for a given protein bound by autoantibodies in each genotype group. When considering protein-level autoantibody specificity, we find substantially more sharing between genotype groups (Figure 2B). This demonstrates that individuals with different *HLA-DRB1* may have autoantibodies to the same protein (Figure 2B), but through distinct peptides (Figure 2A). Nevertheless, we highlight that the majority of protein-level specificity remains exclusive to each *HLA-DRB1* group (196 for *HLA-DRB1*01*,

310 for *HLA-DRB1*03*, 1507 for *HLA-DRB1*04*, 178 for *HLA-DRB1*07* and 269 for *HLA-DRB1*15*). The list of significantly different autoantibodies (peptides and gene/protein) alongside their hit counts can be found on Supplemental Table 3.1 to 3.5.

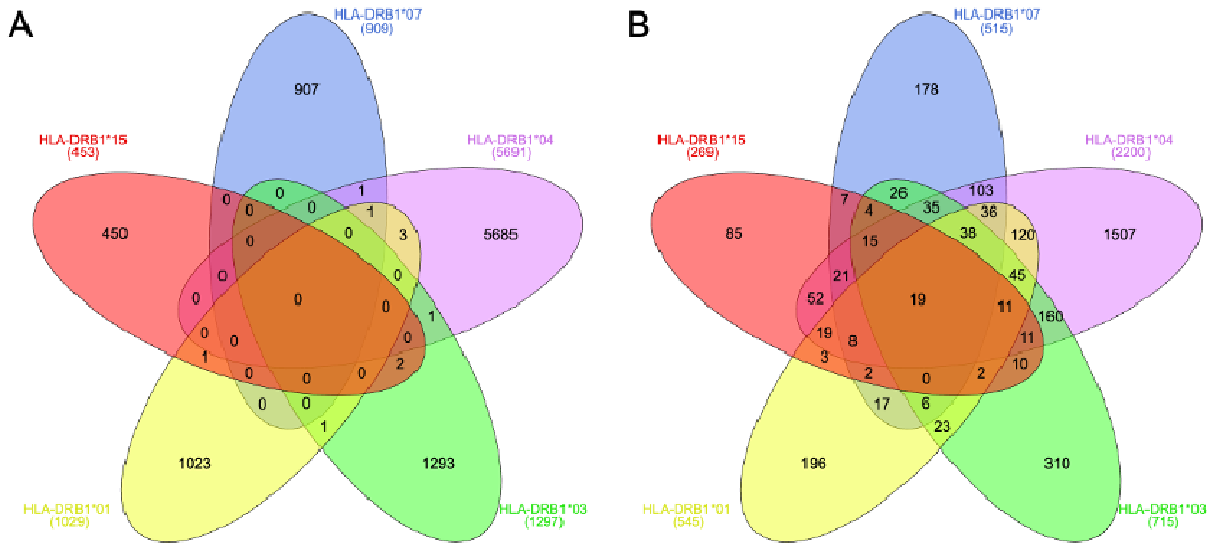

Figure 2: Venn diagrams showing sharing of autoantibody specificity across genotype groups. **A)** Number of peptides bound by autoantibodies significantly expressed in each *HLA-DRB1* group. **B)** Number of gene/proteins with peptides that bound to autoantibodies in each *HLA-DRB1* group.

Autoantibody repertoire is predictive of HLA class II genotype

As an unbiased measure of whether the autoantibody repertoire can be used to predict *HLA* class II genotype, we employed binary and multinomial logistic regression using autoantibody peptide specificity as predictors. In the binary model, autoantibody specificities with the following criteria served as predictors: 1) they must be statistically significantly different relative to other groups, and 2) they must be exclusive to the group of interest. As an example, in the *HLA-DRB1*15 vs*. rest model, the matrix contained 1,023 peptides that satisfied these criteria across 628 samples, including *HLA-DRB1*15* (192), *HLA-DRB1*07* (123), *HLA-*

*DRB1*04* (69), *HLA-DRB1*03* (148), and *HLA-DRB1*01* (96). While the model's accuracy and AUC-ROC score are informative metrics to report, we also present the F1 score to ensure that our models are not biased to one class due to data imbalance.

For the binary model, all groups presented an AUC ranging from 0.96 to 0.98 after undergoing 5-fold cross validation (Figure 3). The *HLA-DRB1*01* group specificities performed best in predicting whether a sample has an *HLA-DRB1*01* genotype ($F1_{HLA\text{-}DRB1*01}$ = 0.85) from the autoantibody repertoire. In contrast, *HLA-DRB1*04* specificities performed poorly in predicting whether a sample has an *HLA-DRB1*04* genotype ($F1_{HLA\text{-}DRB1*04}$ = 0.13). This effect is more apparent from the confusion matrix, shown in Supplemental Figure 1, where the model only correctly predicted one out of fourteen *HLA-DRB1*04* sample, albeit with satisfactory AUC and accuracy. We hypothesized that the poor predictive performance was likely due to the small sample size of the HLA-DRB1*04 group (n = 69), which may have caused the Kruskal–Wallis rank-sum test to identify a disproportionately large number of significantly different autoantibodies (5,691 in this case—several orders of magnitude higher than the average of 922 in other groups).

To understand whether low sample count for the *HLA-DRB1*04* was underlying poor performance in that group, we conducted a sensitivity analysis based on sample size, randomly removed samples from *HLA-DRB1*15* to match the number *HLA-DRB1*04* samples. Running Kruskal-Wallis rank sums test between 69 *HLA-DRB1*15* samples and the 549 remaining yielded 6,372 autoantibodies that are statistically significantly different, similar in magnitude to *HLA-DRB1*04*. We also performed binary logistic regression with these 618 samples and 6,372 autoantibodies, which obtained an accuracy and AUC of 0.89 and 0.99. However, much like *HLA-DRB1*04*, we observed an $F1_{HLA\text{-}DRB1*15}$ value of 0.13, indicating that the model did not perform optimally due to the class imbalance.

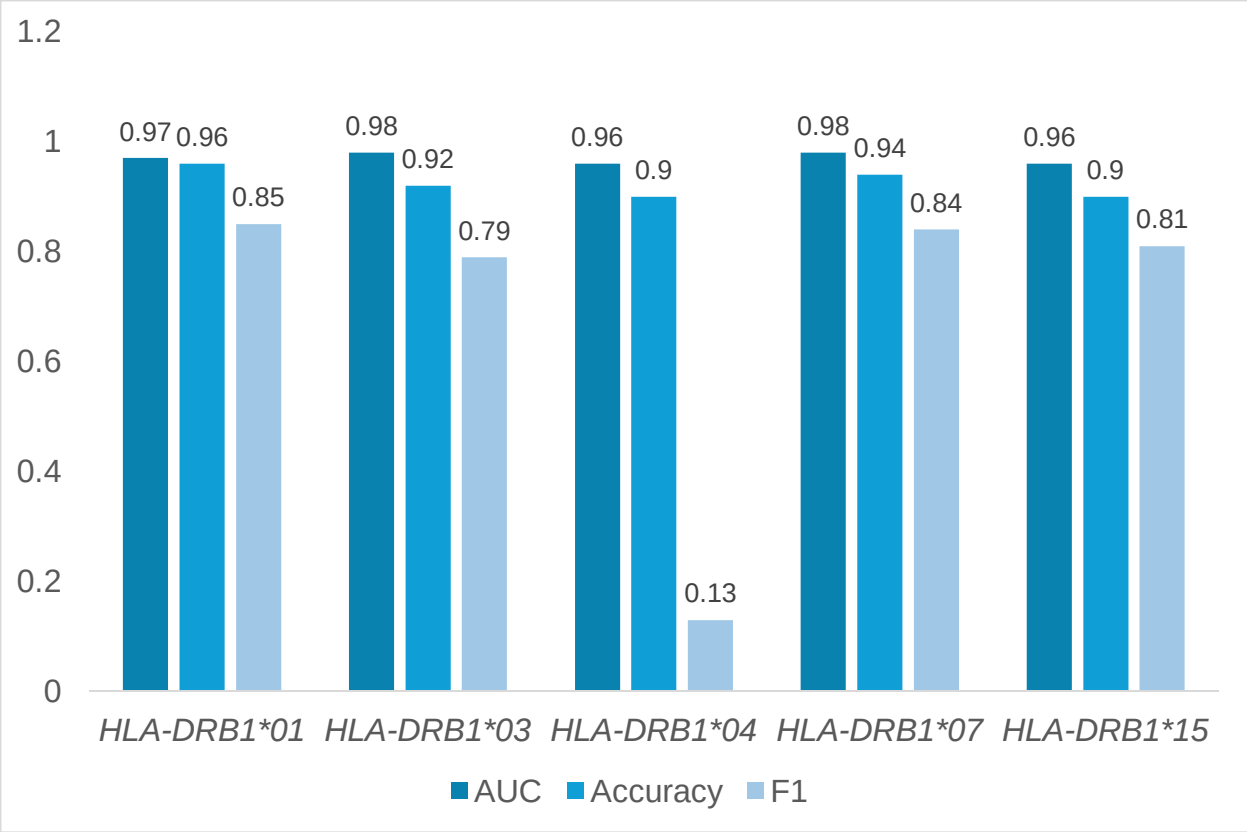

Figure 3: Binary logistic regression performance metrics, which includes the AUC-ROC, accuracy, and F1 score. The models successfully determined a sample's genotype according to autoantibody reactivity, as shown by an average AUC of 0.97 and an average accuracy of 0.92 across the groups. The F1 score for *HLA-DRB1*04* is particularly low due to its sample size being smaller than the rest of the groups. This caused the model to preferentially predict the majority group (non-*HLA-DRB1*04*), even after balancing the training data.

To understand which features are driving decisions in the binary model, we ranked the coefficients that represent the importance of each autoantibody specificity. These coefficients can either be a positive predictor (shown in blue, Figure 4), or a negative predictor (shown in red, Figure 4). A positive predictor is a peptide that shifts the model's decision towards categorizing a sample to the genotype of interest, whereas a negative predictor pushes that decision in the opposite direction. For example, in the *HLA-DRB1*03* logistic regression model,

autoantibody specificity for *NHSL2_57958* functions as a positive predictor (blue box), while *MAML3_311538* serves as a negative predictor (red box). Thus, higher levels of *NHSL2_57958* increase the likelihood that a sample will be classified as *HLA-DRB1*03*, whereas higher levels of *MAML3_311538* push the classification toward non-*HLA-DRB1*03*.

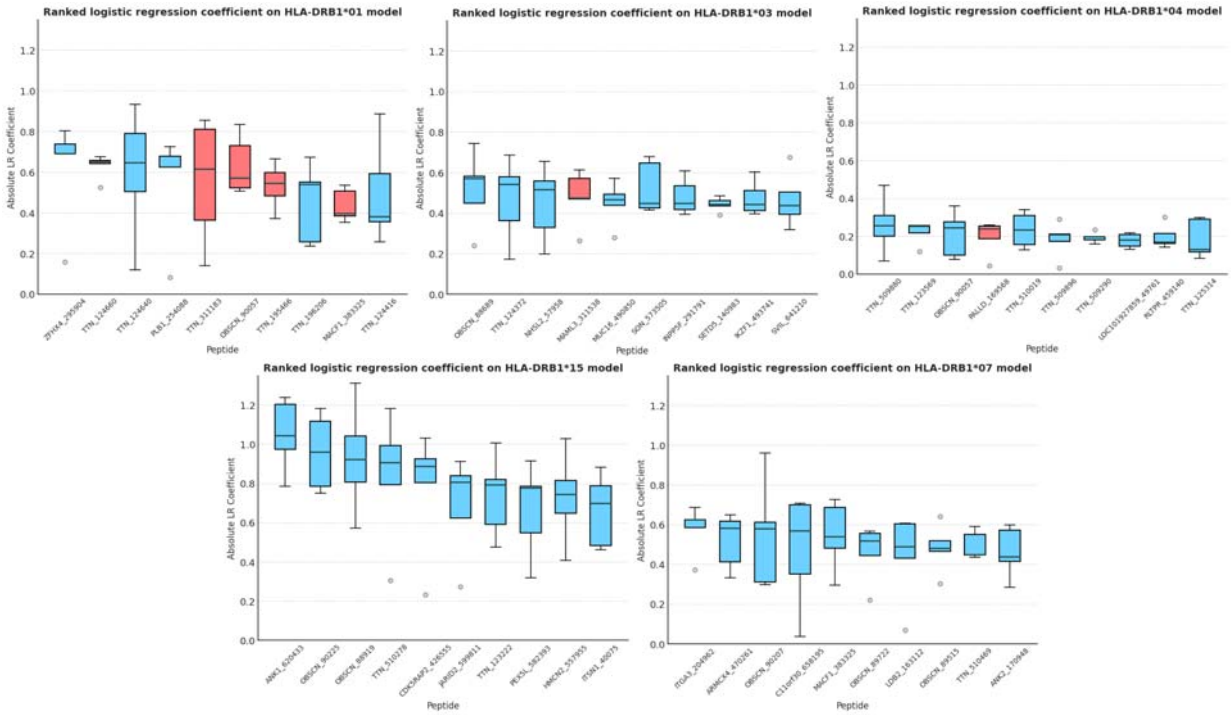

Figure 4: Ranked binary logistic regression coefficients, based on five iterations of cross-validation. Each box in the boxplot represents the feature's coefficient from five iterations of five-fold cross-validation. The colors of the boxes indicate whether a peptide acts as a positive or negative predictor. In this context, a positive predictor is a feature (a peptide) that shifts the model's decision toward a positive outcome.

Multinomial logistic regression returned an accuracy of 76% when the five *HLA-DRB1* groups were tested (Figure 5A). The *HLA-DRB1*04* group showed an F1 score of 0.55 while the F1 score ranged from 0.75 to 0.81 for the remaining four groups. When we performed a multinomial logistic regression excluding *HLA-DRB1*04* group, we observed an accuracy improvement to 90% (Figure 5B). The F1 score ranged from 0.89 (*HLA-DRB1*01* and *HLA-DRB1*03*) to 0.93 (*HLA-DRB1*07*). These results support the notion that a given *HLA-DRB1*

genotype can be predicted from a given autoantibody repertoire, owing to distinct specificity profile.

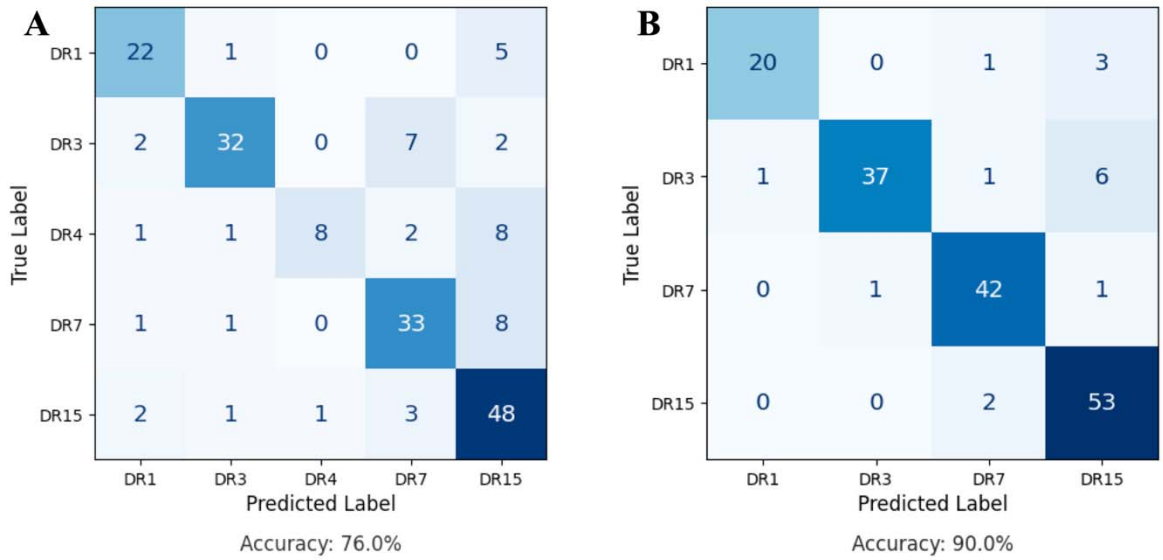

Figure 5: Confusion matrix from the resulting multinomial logistic regression. **A)** Multinomial logistic regression including all five *HLA-DRB1* groups. **B)** Multinomial logistic regression excluding *HLA-DRB1*04* group.

Autoantibody profiles cluster donors according to HLA-DRB1 genotype

To examine whether *HLA* class II genotype groups cluster according to autoantibody repertoire, we applied Uniform Manifold Approximation and Projection (UMAP) to reduce the dimensionality of our data (628 samples × 9,358 peptides) into a 2D representation, as shown in Figure 6. By color-coding each point according to the sample's *HLA-DRB1* genotype, we observed a clear tendency for samples with the same genotype to cluster together, indicating genotype-specific patterns in autoantibody reactivity. Unsurprisingly, *HLA-DRB1*04* exhibits weaker clustering performance compared to the remaining groups, which is likely a consequence of its small sample size.

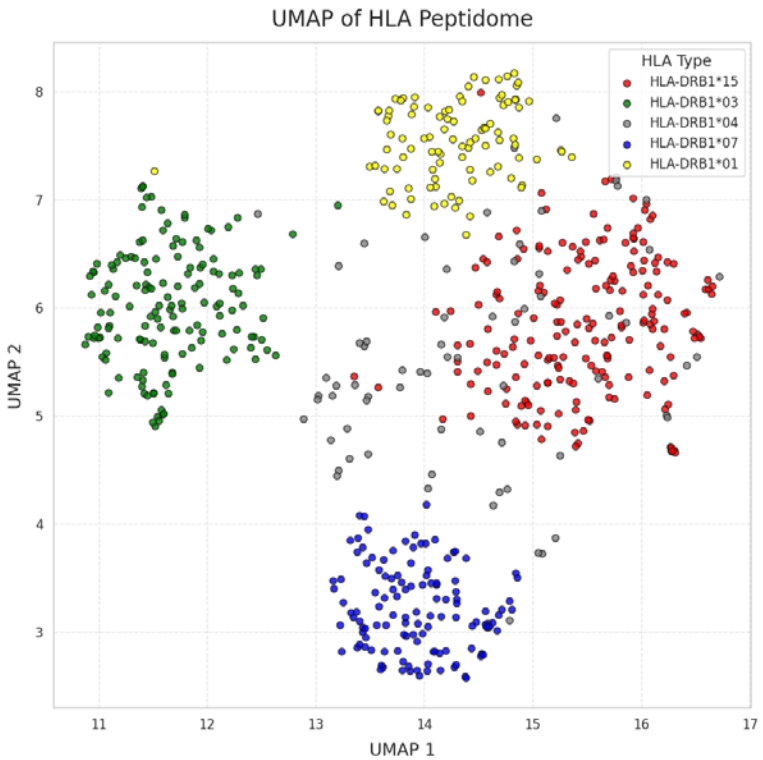

Figure 6: UMAP of autoantigen enrichments of *HLA-DRB*1 groups. Each point in this plot is a sample with a known genotype. A total of 9,358 peptides were represented in this UMAP, consisting of peptides that passed through the Kruskal-Wallis rank sums test and were exclusive to each genotype group.

Elevated autoantibodies specific to independent risk loci in individuals with autoimmune HLA risk haplotypes

We observed several instances of antibodies specific for disease-associated proteins showing elevated levels in an *HLA* genotype group that is a known risk factor for the same disease. For example, out of 450 peptides that were used in the binary logistic regression model for *HLA-DRB1*15,* Figure 4 highlights 10 peptides that are most important in the decision-making process. Among those 10, a peptide from the cyclin-dependent kinase 5 regulatory subunit-associated protein 2 (*CDK5RAP2*) is of interest since it houses rs10984186 (g.118708097G>A) risk variant for Alzheimer's Disease (AD)[25]; we have previously shown that *HLA-DRB1*15* is a part of the risk haplotype that confers risk for AD[26].

A similar case can be observed with Ikaros family zinc finger 1 (*IKZF1*), a gene that plays a role in regulating the lymphocyte differentiation[27], where its peptide appears to be an important predictor for *HLA-DRB1*03*, as seen on Figure 4. This particular gene is of interest because a single nucleotide polymorphism (rs4132601, g.50402906T>G) within the *IKZF1* region has been shown as a susceptible locus for Systemic lupus erythematosus (SLE)[28]; *HLA-DRB1*03:01* has repeatedly been shown as a risk factor for SLE[29–31].

Discussion

Previous studies have demonstrated the presence of autoantibodies in disease-free individuals[15,32,33]. Further, it has been shown that individual IgG autoantibody profiles in the blood are stable over long periods of time[15,34], there is no gender bias[33], and healthy individuals share common autoantibodies[32,33]. Haller-Kikkatalo *et al*., (2017) conducted autoantibody profiling in disease-free individuals and concluded that presence of tissue non-specific autoantibodies might serve as prognostic markers for future disease. Given the role of autoantibodies and the known association of *HLA* class II variation with autoimmune disease risk, in the present study we sought to understand whether *HLA* class II genotype influences the autoantibody repertoire in the absence of overt autoimmune disease. We examined 741 serum samples from healthy *HLA-DRB1* homozygous donors for autoantibody specificity against the entire human proteome, consisting of over 731,000 peptides, demonstrating that autoantibodies that are differentially enriched between the five *HLA-DRB1* groups analyzed.

The PhIP-Seq platform described here has previously been used to identify specific autoantibody specificities in a disease setting. For instance, in a case-control study, Zamecnik *et al*., (2024) found an autoantibody signature that is predictive to multiple sclerosis (MS)[20], while Bodansky *et al*., (2024) carried out a similar comparison with multisystem inflammatory syndrome in children (MIS-C)[35]. Unlike these case-control comparisons, our model emphasizes

variation within the healthy population to show the influence of the *HLA* class II towards the autoantibody repertoire.

While only considering donors without known autoimmune disease, we observed that autoantibody profile might provide information on autoimmune disease risk. The association between *HLA* allele and autoimmune disease such as Alzheimer's and SLE has been established[26,29]. In separate studies, the link between genes in which the autoantibody peptides originate and autoimmune disease has also been found[25,28]. The rs10984186 SNP in *CDK5RAP2* (a predictor for *HLA-DRB1*15* in our model) may regulate AD pathology by cleavage of p35 to p25, which causes *CDK5* overactivation that promotes tau phosphorylation, thus increasing the production of neurofibrillary tangles, a trademark in AD patient's brain.[25]

Despite our finding that *HLA* class II shapes the autoantibody repertoire, the mechanism of this process remains unknown. *HLA* class II alleles determine which peptides are presented to T cells, leading to their activation and the provision of T cell help. This process triggers cytokine release, which stimulates B cells to produce antibodies—or, in cases of self-peptide presentation, autoantibodies. To prevent self-peptide presentation, autoreactive T cells are typically eliminated via clonal deletion in the thymus during its development.[36] This process, however, is imperfect and autoreactive T cells often escape clonal deletion.[36]

To the best of our knowledge, this is the first study to showcase that the autoantibody repertoire in healthy individuals is shaped by the *HLA* class II genotype. Further, through a supervised machine learning regression model we were able to demonstrate that the autoantibody repertoire can be used to predict *HLA-DRB1* genotype, reaching a performance accuracy of up to 90% for a multinomial model and 96% for a binomial model (*HLA-DRB1*01*). Future studies in more diverse cohorts and including other *HLA* genotypes should provide more evidence regarding how *HLA* class II variation influences the landscape of our autoantibody repertoire. Our findings also bring bolster the notion that the autoantibody repertoire may be used to predict the onset of autoimmune disease before its manifestation.

References

1. Markov, P. V. & Pybus, O. G. Evolution and diversity of the human leukocyte antigen(HLA). *Evol. Med. Public Health* **2015**, 1 (2015).
2. Lenz, T. L., Spirin, V., Jordan, D. M. & Sunyaev, S. R. Excess of Deleterious Mutations around HLA Genes Reveals Evolutionary Cost of Balancing Selection. *Mol. Biol. Evol.* **33**, 2555–2564 (2016).
3. Chapman, S. J. & Hill, A. V. S. Human genetic susceptibility to infectious disease. *Nature Reviews Genetics* vol. 13 175–188 Preprint at https://doi.org/10.1038/nrg3114 (2012).
4. Kirino, Y. *et al.* Genome-wide association analysis identifies new susceptibility loci for Behçet's disease and epistasis between HLA-B*51 and ERAP1. *Nat. Genet.* **45**, 202–207 (2013).
5. Fakiola, M. *et al.* Common variants in the HLA-DRB1-HLA-DQA1 HLA class II region are associated with susceptibility to visceral leishmaniasis. *Nat. Genet.* **45**, 208–213 (2013).
6. Parham, P. MHC class I molecules and KIRS in human history, health and survival. *Nature Reviews Immunology* vol. 5 201–214 Preprint at https://doi.org/10.1038/nri1570 (2005).
7. Misra, M. K., Damotte, V. & Hollenbach, J. A. The immunogenetics of neurological disease. *Immunology* vol. 153 399–414 Preprint at https://doi.org/10.1111/imm.12869 (2018).
8. Hollenbach, J. A. & Oksenberg, J. R. The immunogenetics of multiple sclerosis: A comprehensive review. *Journal of Autoimmunity* vol. 64 13–25 Preprint at https://doi.org/10.1016/j.jaut.2015.06.010 (2015).
9. Koning, M. T. *et al.* Massive Parallel Sequencing of Full-Length B-Cell Receptor Sequences Reveals HLA-Dependent Shaping of the B-Cell Immune Repertoire. *Blood* **124**, 4143–4143 (2014).
10. Ovsyannikova, I. G. *et al.* Associations between human leukocyte antigen (HLA) alleles and very high levels of measles antibody following vaccination. in *Vaccine* vol. 22 1914–1920 (2004).
11. Stern, L. J. & Calvo-Calle, J. M. *HLA-DR: Molecular Insights and Vaccine Design*. http://hla.alleles.org/.
12. Pajewski, N. M. *et al.* The role of HLA-DR-DQ haplotypes in variable antibody responses to Anthrax Vaccine Adsorbed. *Genes Immun.* **12**, 457–465 (2011).
13. Prentice, H. A. *et al.* HLA class II genes modulate vaccine-induced antibody responses to affect HIV-1 acquisition. *Sci. Transl. Med.* **7**, (2015).
14. Suurmond, J. & Diamond, B. Autoantibodies in systemic autoimmune diseases: Specificity and pathogenicity. *Journal of Clinical Investigation* **125**, 2194–2202 (2015).
15. Nagele, E. P. *et al.* Natural IgG Autoantibodies Are Abundant and Ubiquitous in Human Sera, and Their Number Is Influenced By Age, Gender, and Disease. *PLoS One* **8**, (2013).
16. Madbouly, A. *et al.* Validation of statistical imputation of allele-level multilocus phased genotypes from ambiguous HLA assignments. *Tissue Antigens* **84**, 285–292 (2014).
17. Larman, H. B. *et al. PhIP-Seq Characterization of Autoantibodies from Patients with Multiple Sclerosis, Type 1 Diabetes and Rheumatoid Arthritis*. *J Autoimmun* vol. 43 (2013).
18. Mohan, D. *et al.* PhIP-Seq characterization of serum antibodies using oligonucleotide-encoded peptidomes. *Nat. Protoc.* **13**, 1958–1978 (2018).
19. Zamecnik, C. R. *et al.* ReScan, a Multiplex Diagnostic Pipeline, Pans Human Sera for SARS-CoV-2 Antigens. *Cell Rep. Med.* **1**, (2020).
20. Zamecnik, C. R. *et al.* An autoantibody signature predictive for multiple sclerosis. *Nat. Med.* **30**, 1300–1308 (2024).

21. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
22. DiCillo, E. B. *et al.* Patterns of autoantibody expression in multiple sclerosis identified through development of an autoantigen discovery technology. *Journal of Clinical Investigation* **135**, (2025).
23. Pedregosa FABIANPEDREGOSA, F. *et al. Scikit-Learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot*. *Journal of Machine Learning Research* vol. 12 http://scikit-learn.sourceforge.net. (2011).
24. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. http://arxiv.org/abs/1802.03426 (2020).
25. Miron, J. *et al.* CDK5RAP2 gene and tau pathophysiology in late-onset sporadic Alzheimer's disease. *Alzheimer's and Dementia* **14**, 787–796 (2018).
26. Steele, N. Z. R. *et al.* Fine-mapping of the human leukocyte antigen locus as a risk factor for Alzheimer disease: A case–control study. *PLoS Med.* **14**, (2017).
27. Dieudonné, Y., Guffroy, A., Vollmer, O., Carapito, R. & Korganow, A. S. IKZF1 Loss-of-Function Variant Causes Autoimmunity and Severe Familial Antiphospholipid Syndrome. *Journal of Clinical Immunology* vol. 39 353–357 Preprint at https://doi.org/10.1007/s10875-019-00643-2 (2019).
28. Chen, L. *et al.* IKZF1 polymorphisms are associated with susceptibility, cytokine levels, and clinical features in systemic lupus erythematosus. *Medicine (United States)* **99**, E22607 (2020).
29. Morris, D. L. *et al.* Unraveling multiple MHC gene associations with systemic lupus erythematosus: Model choice indicates a role for HLA alleles and non-HLA genes in europeans. *Am. J. Hum. Genet.* **91**, 778–793 (2012).
30. Miglioranza Scavuzzi, B. *et al.* The lupus susceptibility allele DRB1*03:01 encodes a disease-driving epitope. *Commun. Biol.* **5**, (2022).
31. Kim, K. *et al.* The HLA-DRβ1 amino acid positions 11-13-26 explain the majority of SLE-MHC associations. *Nat. Commun.* **5**, 5902 (2014).
32. Haller-Kikkatalo, K. *et al.* Demographic associations for autoantibodies in disease-free individuals of a European population. *Sci. Rep.* **7**, (2017).
33. Shome, M. *et al.* Serum autoantibodyome reveals that healthy individuals share common autoantibodies. *Cell Rep.* **39**, (2022).
34. Lacroix-Desmazes, S., Mouthon, L., Kaveri, S. V, Kazatchkine, M. D. & Weksler, M. E. Stability of natural self-reactive antibody repertoires during aging. *J Clin Immunology* (1999).
35. Bodansky, A. *et al.* Molecular mimicry in multisystem inflammatory syndrome in children. *Nature* **632**, 622–629 (2024).
36. Owen, D. L., Sjaastad, L. E. & Farrar, M. A. Regulatory T Cell Development in the Thymus. *The Journal of Immunology* **203**, 2031–2041 (2019).