Open access • Posted Content • DOI:10.1101/580126

Prioritizing putatively etiological T cell epitopes across autoimmune diseases

Masato Ogishi

Institutions: University of Tokyo

Published on: 16 Mar 2019 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Human leukocyte antigen, Epitope and Histocompatibility

Prioritizing putatively etiological T cell epitopes across autoimmune diseases

Masato Ogishi [1*]

[1] Division of Infectious Diseases and Applied Immunology, The Institute of Medical Sciences Research Hospital, The University of Tokyo, Tokyo, Japan

***Correspondence:** Masato Ogishi (oogishi-tky@umin.ac.jp)

Keywords: self-tolerance, autoimmunity, genetic predisposition and protection, HLA, T cell epitope

Abstract

Autoimmune diseases remain a leading cause of mortality among adolescents and young adults worldwide. Despite their clinical impact, there are still significant knowledge gaps in our understanding of immunological tolerance and its breach that characterizes the onset of autoimmune diseases. Genetic associations between the histocompatibility leukocyte antigen (HLA) loci and various autoimmune diseases have been well established. The HLA class I and class II molecules present epitopes to T cells, and T cells play indispensable roles both in the maintenance of tolerance and the pathogenesis of autoimmune diseases. Although a vast number of epitopes and reactive T cell clones have been identified from animal model studies and observational studies, however, only a few have been proven to be causally relevant to disease pathogenesis. Here, we propose a computational framework to prioritize etiologically relevant epitopes by integrating the putatively causal associations between HLA alleles and disease risk identified from population genetics; we define a metric, termed "differential presentation index (DPI)," which principally reflects the relative difference of epitope abundance presented onto HLA molecules whose alleles are genetically predisposing to or protective against the specific disease. We systematically examined publicly available epitope sequence data previously studied in the context of autoimmune diseases. Self-epitopes were generally more stably presented on disease-protective HLAs than non-self epitopes, and hence had a negative DPI. Conversely, proteome-wide sequence alignment revealed that epitopes with highly positive DPI were less similar to self. As a case study, we performed a focused analysis of multiple sclerosis (MS), and identified epitopes from myelin basic protein (MBP), a well-established MS autoantigen, based on DPI-guided prioritization. Moreover, we found several non-MBP-derived self-epitopes with high DPI that are potentially involved in the pathogenesis of MS. Our framework facilitates the identification of etiologically relevant epitopes across autoimmune diseases with known HLA allele association, which in turn expedites the development of epitope-specific disease monitoring and intervention strategies.

1 Introduction

Autoimmune diseases have been shown to be more common than previously thought from several epidemiological studies. The estimated total incidence and prevalence are 90 per person-years and 3.2%, respectively, when summed across diseases (1,2). Graves' disease (GD), thyroiditis, and rheumatoid arthritis (RA) are among the most common diseases (>10 per 100,000 person-years) whereas other diseases such as systemic lupus erythematosus (SLE) and multiple sclerosis (MS) are relatively rare (1~10 per 100,000 person-years). Although clinical manifestations considerably vary between diseases, most of the tissue/organ damage is thought to arise from the dysregulated response of the adaptive immunity, both cytotoxic and humoral immunity, to self-antigens. Although significant advances have been made both in the diagnosis and clinical management in the past decade, in most of the diseases there remains no successful therapeutics leading to complete remission and cure. Patients tend to suffer from the chronic, relapsing and refractory nature of the diseases and the adverse effects of treatments, leading to poor quality of life and significant economic burden. Indeed, autoimmune diseases are still among the leading causes of mortality among young and middle-aged adults, particularly among females, and the mortality rate remains relatively constant (3,4).

Very little is known about the disease-initiating processes of autoimmune diseases despite the advances in our understanding of their steady-state pathophysiology. Mechanistic studies have been hampered, at least in part, because of the following reasons: (i) patients with an early stage disease lacking typical clinical manifestations tend not to seek a medical examination or to receive a correct diagnosis, and therefore are rarely studied; (ii) it is often difficult to elucidate the mechanisms essential for disease prevention and/or initiation from observational studies of patients with chronic inflammation and extensive systemic involvements due to several secondary changes; (iii) etiologies identified from studies utilizing genetically predisposing animal models do not necessarily reflect the pathophysiology in humans. Undoubtedly, identification of epitopes relevant to the disease initiation process would have significant translational implications, since recent studies have indicated the possibility of antigen/epitope-specific immunosuppression via various strategies such as transplantation of receptor-engineered T cells with immunomodulatory functions, and more interestingly, systemic administration of peptide-MHC-based nanomedicine to reconstruct physiological regulatory networks (5–9).

Genetics, likewise external environmental factors, is apparently one of the most upstream factors in the cascade of disease initiation. In other words, the genome alteration can be causal to disease, but not *vice versa* (except tumorigenesis). Notably, a number of major histocompatibility complex [MHC; also known human leukocyte antigen (HLA) in humans] loci, as well as hundreds of non-MHC loci, have been associated with predisposition and protection of autoimmune diseases (10–12). Since experimental genetic manipulation is not ethically acceptable, population genetics is a valuable tool to assess the contribution of particular HLA alleles to the onset of autoimmune diseases in humans. In other words, the enrichment or paucity of specific HLA alleles in patients with specific autoimmune diseases could reflect their etiological roles, although one should keep in mind the caveat that other non-HLA loci in strong linkage disequilibrium are of *bona fide* etiological significance. That being said, considering the biological function of MHC molecules to present a short peptide fragment (called epitope) derived from various self- and non-self-antigens to T cells, and the multifaceted roles of T cells as the master regulator of adaptive immunity and self-tolerance, the contribution of T cell epitope to autoimmune diseases has been extensively studied, and numerous self- and non-self-derived epitopes and cognate cross-reactive T cell clones have been identified (13–15). Nevertheless, only a minority of them have been causally linked to either disease

initiation by pathogenic T cells or maintenance of self-tolerance by clonal deletion at thymus and/or induction of regulatory T cells (Tregs) *in vivo*, and consequently, there remain virtually no epitope-specific diagnostic/therapeutic strategies clinically available to date. Therefore, an additional "filter" complementary to experimental validation of T cell recognition *in vitro* is needed to expedite the exploration of the most etiologically responsible epitopes.

In this context, here we propose a simple approach to prioritize epitopes with high pathophysiological relevance by integrating population genetics and MHC binding prediction. We hypothesized that epitopes more stably presented on the HLA molecules encoded by disease-predisposing alleles are more likely to be pathogenic, whereas those preferentially presented on HLA molecules encoded by disease-protective alleles are more likely to contribute to tolerance. HLA loci associated with specific diseases were extracted from the previous phenome-wide association study (PheWAS) data (11), and MHC binding prediction was conducted bioinformatically using NetMHCpan and NetMHCIIpan (16,17). We introduced a metric termed "differential presentation index (DPI)" based on the predicted binding strength among predisposing and protective HLA molecules. As a case study, we screened epitopes studied in the context of MS, and, as expected, found that several epitopes derived from myelin basic protein (MBP), a well-characterized autoantigen, ranked highly based on MS-specific DPI (18,19). Moreover, we found candidates of MS-relevant non-MBP epitopes derived from self-antigens including SIK1, GRK2, IFNB, and EPO, all of which are present and play critical roles in the central nervous system (CNS). Notably, some of those newly identified epitopes had even higher MS-specific DPI than MBP epitopes. Finally, examination of an experimentally verified molecular mimicry epitope dataset revealed two putatively MS-predisposing mycobacterial epitopes homologous to an MBP-derived self-epitope and one putatively RA-protective mycobacterial epitope homologous to mammalian 60kDa heat shock protein (HSP60), a well-known Treg-inducing self-antigen (20,21). Collectively, these findings illustrate the utility of DPI-based epitope prioritization strategy in search of etiologically relevant epitopes. Further characterization in different disease contexts and experimental validation are warranted. The datasets and codes necessary to reproduce the analytical pipeline are made publicly available as the R package *DPA* on GitHub (https://github.com/masato-ogishi/DPA/) to expedite future research.

2 Results

2.1 Datasets

We extracted disease-HLA associations across different autoimmune diseases from the previous study (11). This study involved two populations of European ancestry individuals (N=28,839 and 8,431) and tested the association of HLA variation with 1,368 phenotypes. We identified the following autoimmune disease phenotypes: ankylosing spondylitis (AS), celiac disease (CD), dermatomyositis (DM), giant cell arteritis (GCA), Graves' disease (GD), Juvenile rheumatoid arthritis (JRA), localized lupus and systemic lupus erythematosus (SLE), multiple sclerosis (MS), polymyalgia rheumatica (PMR), polymyositis (PM), primary biliary cirrhosis (PBC), psoriasis and psoriatic arthropathy (PSO), rheumatoid arthritis (RA), systemic sclerosis (SS), type 1 diabetes (T1D), ulcerative colitis (UC), and Wegener's granulomatosis (GPA). We classified the associated HLA alleles into two categories, namely, "predisposing" and "protective," based on the odds ratios. Then we screened diseases that have both predisposing and protective HLA alleles. The following diseases met the criteria: AS, CD, DM, GD, JRA, SLE, MS, PMR, PBC, PSO, RA, SS, T1D, and UC (Table S1). Next, we downloaded the linear T cell epitope data annotated with at least one cell-based functional assay results from the Immune Epitope Database (IEDB) (22). Inclusion/exclusion criteria

in terms of assay annotations are provided in Table S2. Epitopes were considered "immunogenic" if a positive T cell response was recorded from at least one functional assay. We use this definition to capture as many potentially T-cell-activating epitopes. We identified epitopes studied in the context of the following autoimmune diseases: AS, CD, GD, GPA, MS, PBC, PSO, RA, SLE, SS, and T1D (Table S3). It should be noted that these disease contexts do not necessarily guarantee the pathophysiological relevance of the epitopes studied. We screened exact matches to human proteome (UniProt ID: UP000005640) to identify self-derived (S) and non-self (NS) epitopes. We also aligned each epitope sequence against the entire human proteome using the Smith-Waterman local alignment algorithm, with a substitution matrix and gap opening/extension penalty parameters identical to those utilized in the blastp-short tool (see Materials and Methods). We took three representative metrics, namely, mean, maximum, and minimum, of the alignment score distribution as indicators of similarity-to-self for each of the epitopes.

2.2 Differential presentation analysis

Our goal was to identify epitopes the most differentially presented among predisposing and protective HLA molecules in a disease-specific context (Figure 1). To this end, we first merged the disease-HLA and disease-epitope association data to obtain a set of epitopes coupled with disease-specific predisposing/protective HLA allele information. Then, we computed the percentile rank values using either NetMHCpan or NetMHCIIpan. Only four-digit HLA alleles were considered, and two-digit HLA alleles were ignored in the subsequent analysis. Computation for HLA-DP/DQ alleles was problematic because a DP-DQ pair must be provided for the binding prediction, despite the single allelic format of disease association in the PheWAS source data. We ended up computing every single possible combination of DP-DQ alleles that contain at least one disease-associated allele and summarizing the predicted values by taking their medians. We then summarized the sign-inverted log-transformed percentile rank values among protective and predisposing HLAs by taking their mean, maximum and minimum. Note that, after this conversion, the larger value reflects more stable binding. We defined "differential presentation index (DPI)" as the difference between the maximum values among predisposing and protective HLAs. We utilized the maximum value among HLAs tested because any epitope does not necessarily bind strongly to all disease-associated HLAs; strong binding to at least one disease-associated HLA is adequate as the initial inclusion criteria. DPI is an indicator unique to each epitope defined in a disease-specific context. A positive DPI means that the epitope is predicted to be bound more strongly to at least one of the predisposing HLAs than any of the protective HLAs. We then categorized epitopes with DPI > 0.5 and DPI < -0.5 as "predisposing" and "protective," respectively. Note that these epitope-level categories are putative. We excluded epitopes if the minimum percentile rank among all disease-associated HLAs was higher than the recommended threshold (2% and 10% for HLA-I and HLA-II, respectively). Figure 2 shows the distributions of the lowest percentiles among predisposing and protective HLAs. Both the number of available epitope data and their distribution considerably varied between diseases. There was a relatively large number of epitopes associated with MS, RA, and T1D (Figure 2B). Full epitope data is summarized in Table S4.

Next, we explored the common characteristics among the differentially presented epitopes. Interestingly, predisposing epitopes were significantly more likely to be derived from non-self antigens than protective epitopes (P=3.2×10^{-6} and 4.7×10^{-7} in HLA-I and HLA-II, respectively, by chi-square test) (Figure 3A). Meanwhile, we noted that 31/55 (56%) of HLA-I predisposing epitopes were classified as non-immunogenic, meaning that there was not even a single positive T cell assay result (Figure 3B). The difference in terms of immunogenicity between predisposing and protective epitopes was also statistically significant in HLA-I but not in HLA-II epitopes (P=2.5×10^{-7} and 0.35,

respectively, by chi-square test). To quantitatively assess similarity-to-self, we chose the highest sequence alignment scores, representing the best-match sequence-level homology against the human proteome. As expected, epitopes exactly matched to human proteome had apparently higher scores (Figure 3C). Notably, the predisposing epitopes were less similar to self (P=1.6×10^{-6} and 1.1×10^{-6} in HLA-I and HLA-II, respectively, by Wilcoxon rank sum test) (Figure 3D). Consistently, the overall correlation between DPI and similarity-to-self was evident, although we noted considerable variation between diseases (Figure S1). Of note, MS-associated predisposing epitopes with high DPI were more non-self, an observation not contradictory to the molecular mimicry hypothesis (13,23,24). In contrast, no correlation was observed among T1D-associated epitopes, possibly indicating an indispensable role of aberrantly formed self-epitopes (25,26). It is difficult to interpret epitopes associated with other diseases due to the paucity of data. Type 2 ANOVA revealed that neither peptide category or T cell reactivity significantly contributed to the similarity-to-self variation after stratified by peptide origin (S vs. NS) in HLA-I, whereas mutual interaction between peptide category and origin existed (P=2.7×10^{-9}) in HLA-II (Figures 3E and F). It was unexpected to us that T cell recognition was not associated with the similarity-to-self score. We did not either observe any difference among immunogenic epitopes and non-immunogenic MHC binders in the datasets we have previously compiled (N=21,162 and 31,693 for HLA-I and HLA-II epitopes, respectively) (Figure S2) (27) (manuscript under review in *Frontiers in Immunology*).

In summary, (i) predisposing epitopes are generally less similar to self, primarily because peptides derived from non-human antigens tend to be more stably presented on predisposing HLAs; (ii) predisposing HLA-I but not HLA-II epitopes are less likely to be recognized by T cells; and (iii) propensity of T cell recognition is orthogonal to peptide similarity-to-self.

2.3 Case study: multiple sclerosis

We next asked whether the proposed DPI-based framework was able to effectively prioritize known antigens/epitopes. To examine this, we decided to do a focused analysis of MS-associated epitopes because of the abundance of available epitope data. DPIs were calculated based on the binding prediction to four and six HLA-I and HLA-II alleles associated with MS, among which two and three were MS-predisposing, respectively. More than 75% of both HLA-I and HLA-II MS-predisposing epitopes did not match to the human proteome (P=2.3×10^{-4} and 1.1×10^{-5} in HLA-I and HLA-II, respectively, by Fisher's exact test) (Figure 4A). Meanwhile, there was a striking dissociation in terms of T cell recognition between HLA-I and HLA-II MS-predisposing epitopes (P=6.0×10^{-8} and 1.0×10^{-5} in HLA-I and HLA-II, respectively, by Fisher's exact test) (Figure 4B). Interestingly, most of the non-self, non-immunogenic, yet predisposing HLA-I epitopes were derived from human endogenous retrovirus W (HERV-W). Predisposing epitopes had lower similarity-to-self compared to protective epitopes in HLA-I but not in HLA-II data (P=4×10^{-5} and 0.07 in HLA-I and HLA-II, respectively, by Wilcoxon rank sum test) (Figure 4C). We speculated that this is owing to a set of predisposing HLA-II epitopes highly homologous to MBP-derived epitopes. As expected, a subset of non-MBP-derived predisposing HLA-II epitopes shared high sequence homology to MBP (P=0.3 and 1×10^{-4} in HLA-I and HLA-II, respectively, by Wilcoxon rank sum test) (Figure 4D). Finally, we found that DPI-guided prioritization enriched MBP-derived epitopes among self-epitopes (P=0.01 and 1×10^{-4} in HLA-I and HLA-II, respectively, by Wilcoxon rank sum test) (Figure 4E).

Encouraged by these observations, we next sought to screen epitopes potentially relevant to MS etiology. We applied our DPI calculation framework to all self-epitopes identified from the previously compiled epitope datasets (Table S5) (27). The top five putatively MS-predisposing epitopes for both HLA-I and HLA-II are shown in Table 1. Interestingly, the top-ranked HLA-I

epitope RPRPVSPSSL is derived from salt-inducible kinase1 (SIK1). The fourth HLA-I epitope KPRSPVVEL is derived from G-protein coupled receptor kinases 2 (GRK2). Notably, the top-ranked HLA-I epitopes had even higher MS-specific DPI than MBP-derived epitopes. Likewise, among HLA-II epitopes, two interferon beta (IFN-β)- and two erythropoietin (EPO)-derived epitopes were identified (Table 1). MBP-derived HLA-II epitopes were excluded from Table 1 but can be found in Table S5. IFN-β has traditionally been used as an immunomodulatory drug for MS, and is thought to act as a suppressor for T helper type 1 (Th1) cells and T helper 17 (Th17) cells (28–30). EPO, although initially identified as an essential factor for hematopoiesis secreted from kidney, has been implicated as a potent neuroprotective agent (31). In addition to MS-predisposing epitopes, we also explored the putatively MS-protective epitopes. The top five putatively MS-protective HLA-II epitopes are shown in Table 2. Of note, two glutamic acid decarboxylase (GAD) epitopes were identified. Note that we labeled epitopes as "disease-protective" based on genetic association, and hence it does not necessarily guarantee their role in the maintenance of self-tolerance (*e.g.*, induction of Tregs). Nevertheless, it is possible that at least some epitopes with considerably low DPIs act as Treg-inducing epitopes. We will discuss this point in the next chapter.

2.4 Case study: molecular mimicry

Molecular mimicry has been proposed as one of the potential mechanisms underlying the breach of self-tolerance and autoimmunity (23,32–34). However, the pathophysiological significance of the T cell clones recognizing both self- and pathogen-derived epitopes *in vivo* remains undetermined or inconclusive in most cases. Moreover, recognition of a specific epitope by effector T cells *in vitro* does not preclude the possibility of recognition of the same epitope by Tregs *in vivo*, which may overall result in the protection against disease.

Self- and pathogen-derived epitopes associated with diseases in the same direction are likely to share biological roles as well as sequence-level homology. Thus, we searched pairs of human-pathogen epitopes that have the differential presentation category in common. We utilized the miPepBase, a database of experimentally verified self-epitopes and mimicking pathogen-derived epitopes (35). Surprisingly, among the forty-three epitope pairs identified, only four had the same differential presentation category in common (Table 3). Of note, there were two mycobacterium epitopes associated with MS. The role of mycobacteria in the pathogenesis of MS remains controversial (36), and therefore the two mycobacterial epitopes and the corresponding self-epitope appear to be worth being further investigated. Meanwhile, there was one epitope pair associated with putative RA protection. Interestingly, both epitopes are derived from heat shock protein (HSP). It is known that cell stress-induced up-regulation of HSPs and the resultant presentation of HSP epitopes induce Tregs (37). Indeed, T cell clones cross-reactive to epitopes derived from mammalian HSP60 and *M. bovis* HSP65 have been shown to protect against experimental arthritis, strongly suggesting the involvement of Tregs (20,38). The comprehensive set of epitope pairs tested are summarized in Table S6.

3 Discussion

We systematically mined several T cell epitope candidates likely to be etiologically relevant across autoimmune diseases by differential presentation onto genetically inferred disease-predisposing and -protective HLA molecules. T cells have been thought to play multifaceted roles both in the maintenance of physiological self-tolerance and in autoimmune diseases. To date, T cells epitopes presented on HLA molecules whose alleles are genetically associated with disease predisposition have been extensively studied. However, to our knowledge, there are yet few studies

systematically comparing epitope presentation on both predisposing and protective HLAs. Probably, one reason is that testing multiple HLA alleles experimentally is labor-intensive and cost-prohibitive. For example, there are 38 HLA alleles significantly associated with T1D based on the PheWAS data we utilized in this study. Instead of experimental testing, we conducted a comparative analysis through a bioinformatic approach, which can be adequate for a screening purpose. The DPI-based classification system of putatively predisposing and protective epitopes proposed in this study is a novel approach complementary to the current standard approach, *i.e.*, detection of cognate T cell clones. We used percentile rank for differential presentation analysis because this metric is not affected by inherent bias of specific HLA molecules towards higher or lower mean predicted affinities and thus allows a direct comparison between different HLA molecules (16). We used the lowest percentile rank values among predisposing and protective HLA molecules for determining the degree of differential presentation. This definition is conceptually equivalent to the comparison between two hypothetical populations having either all disease-specific predisposing or protective HLA alleles (instead of six HLA alleles per individual in the real-world). We did not consider the odds ratio of different HLA alleles for a specific disease. For example, the odds ratios among T1D- and MS-predisposing HLA-II alleles ranges from 4.420 to 1.713 and 2.672 to 3.475, respectively. Whether odds-ratio-weighted metric contributes to better prioritization of etiologically relevant epitope may be a topic of further investigation.

One of the advantages of computational screening is its unbiasedness; experimental approach often focuses on either known epitopes or epitopes from known antigens, although this type of investigation is likely to overlook epitopes of *bona fide* etiological significance derived from unknown antigens. We tackled this issue by performing an unbiased screen of epitope candidates differentially presented onto MS-predisposing and -protective HLA molecules as a case study. As expected, our analysis revealed several MBP-derived epitopes. The current consensus is that MS is a T cell-mediated autoimmunity against myelinated self-antigens including MBP, and the etiological significance of MBP in MS has been well documented (18,39). Therefore, this enrichment of MBP-derived epitopes can be viewed as an internal positive control for the analysis. Furthermore, we identified several self-antigens, including SIK1, GRK2, IFNB, and EPO, as a potential source of etiologically relevant epitopes. First, SIK1 mutations have been documented as a cause of severe developmental epilepsy (40,41). Second, a previous study of experimental autoimmune encephalomyelitis, a mouse model for MS, has shown that $GRK2^{+/-}$ mice expressing 50% of the GRK2 protein did not suffer from relapses unlike the wild-type animals, and the absence of relapse was associated with a marked reduction of infiltrating inflammatory cells in the CNS (42). Modulation of the Toll-like receptor signaling via GRK2 in microglia was also reported (43). Third, IFN-β is a naturally occurring cytokine mediating a wide range of anti-inflammatory responses in the CNS, and has been used as a therapeutic agent for MS (28,30). The formation of neutralizing autoantibody against therapeutically administered IFN-β strongly suggests a pre-existing humoral immune response against this self-antigen (44). Fourth, EPO is an endogenous neuroprotective protein, and its efficacy has been shown in a pre-clinical model and a small clinical trial, although the subsequent study failed to show its superiority (31,45,46). Besides the putatively MS-predisposing self-antigens, we also identified putatively MS-protective candidate self-antigens including GAD. GAD catalyzes the synthesis of gamma-aminobutyric acid (GABA), an inhibitory neurotransmitter, from glutamate, and hence has an indispensable role in the physiology of GABAergic neurons (47). Notably, anti-GAD autoantibodies are present not only in type 1 and 2 diabetes mellitus, but also in various neurological diseases including stiff-person syndrome, Miller Fisher syndrome, limbic encephalopathy, cerebellar ataxia, eye movement disorders, and epilepsy, but usually not in MS (48). Collectively, aberrant T cell-mediated autoimmunity against these endogenous self-antigens may disrupt the physiological integrity of the CNS microenvironment and thereby contribute to the

pathogenesis of MS. Moreover, these findings underscore the potential utility of an unbiased computational screen in search of etiologically relevant epitopes contributing to the onset and/or aggravation of various autoimmune diseases.

The general principles of Treg-inducing epitopes have remained largely undefined. Treg-inducing epitopes identified from IgG are well-known examples (49,50), which has been thought to at least partially explain the beneficial immunomodulatory effects of high-dose intravenous immunoglobulin (IVIg) therapy for some autoimmune and autoinflammatory diseases (51). However, the immunomodulatory effect of IgG-derived epitopes is apparently not disease-specific. Identification of disease-specific, etiologically relevant novel Treg-inducing epitopes would provide valuable insights into the pathophysiological mechanisms of various autoimmune diseases, and may also pave the way toward epitope-specific immunomodulatory therapeutics (5,9,52). Notably, we found that the majority of the HLA-I predisposing epitopes, although apparently derived from non-self proteome, do not have any positive T cell assay annotation in IEDB. This may be explained as a reflection of tolerance mechanisms leading to clonal anergy or deletion, although overinterpretation should be avoided due to the retrospective nature of this study. In a case study of MS-associated epitopes, we noted that the vast majority of HLA-I predisposing epitopes were derived from HERV-W. HERV-W has also been called MS-associated retrovirus (MSRV) and extensively studied as a potential etiology of MS (53). A recent meta-analysis showed a strong association between detectable HERV-W mRNA and MS (54). Since HERV had been integrated into the human genome in 70 to 30 million years ago, representing almost 8% of the entire human genome, it is not surprising that our adaptive immunity tolerates HERV proteins. Indeed, recent studies highlighted the indispensable role of $CD8^+$ Tregs, as well as conventional $CD4^+Foxp3^+$ Tregs, both in the maintenance of self-tolerance and suppression of antiviral immunity (55,56). Likewise, our analysis of molecular mimicry epitopes revealed human and mycobacteria HSPs as a source of RA-protective HLA-II epitopes, which is consistent with previous observations that HSP-derived epitopes induce $CD4^+$ Tregs and provide protection against adjuvant-induced arthritis in mice (20,38). Therefore, a set of HLA-I non-self epitopes with high DPI but without evidence of T cell recognition, and a set of HLA-II self-epitopes with low DPI, may be considered a good starting point in the exploration of Treg-inducing epitopes in a disease-specific context. Further research is warranted to test the generalizability of this concept.

Molecular mimicry has long been suspected as a general principle of the etiology of various autoimmune diseases. A critical obstacle in the research of molecular mimicry is, in our opinion, that whether any pair of self- and pathogen-derived epitopes have etiological relevance in the same direction, *i.e.*, predisposition to or protection against disease, cannot be argued either based on sequence homology or recognition by the same T cell clones alone. Our framework thus provides one additional layer of criteria for molecular mimicry. It is notable that only four among 43 pairs of self- and pathogen-derived epitopes met the new criteria, which indeed implies that sequence-homologous human and pathogen-derived epitopes could have a distinct impact in terms of disease initiation. This larger-than-expected dissociation between self-epitopes and homologous pathogen-derived epitopes needs to be tested using experimental animal models. In particular, it may be interesting to investigate the putatively disease-predisposing pathogen-derived epitopes homologous to putatively disease-protective self-epitopes. This is because the adaptive immunity in individuals with disease-predisposing HLA alleles may be naïve to the self-epitopes that are likely to be abundantly presented on disease-protective HLA molecules (*i.e.*, potentially tolerance-inducing), and therefore may not be able to tolerate the homologous pathogen-derived epitopes suddenly beginning to be presented upon infection/reactivation of the pathogen(s). Therefore, we would like to propose that etiological roles of pathogen-derived mimicking epitopes should be investigated in a fashion that integrates genetic association, HLA binding, and T cell recognition both *in vitro* and *in vivo*.

The caveats of our concept of differential presentation-based epitope prioritization are as follows. First, the association between disease status and HLA loci does not guarantee an etiological involvement of the encoded HLA molecules, due to the possibility of linkage disequilibrium. Second, affinity to HLA is not the sole parameter explaining pathogenicity (12). For example, the expression level and/or stability of HLA molecules can matter; for example, the low stability of HLA-DQ6 is thought to confer protection against various autoimmune diseases. Moreover, entirely different immunological consequences (*i.e.*, induction of cytopathic vs. regulatory T cells) resulting from alternative TCR docking due to different peptide register have been reported (15). Third, sequence-based affinity prediction may not always reflect the *bona fide* affinity *in vivo*. A good example is post-translationally modified epitopes such as citrullinated epitopes bound to RA-predisposing HLA-DR4 (57). Therefore, further research including experimental validation both *in vitro* and *in vivo* is necessary to comprehensively characterize the HLA-peptidome landscape across autoimmune diseases.

4 Materials and Methods

Computational analysis

All computational analyses were conducted using R ver. 3.5.2 (https://www.r-project.org/) (58). The latest versions of R packages were consistently used. Compiled datasets and essential in-house functions are available as the R package *DPA* on GitHub (https://github.com/masato-ogishi/DPA). Full analytical scripts are available upon request.

Disease-HLA allele association

Associations between autoimmune diseases and HLA alleles were extracted from a previous PheWAS study (11). This study involved two populations of European ancestry individuals (N=28,839 and 8,431) and tested the association of HLA variation with 1,368 phenotypes. The following autoimmune disease phenotypes were identified: ankylosing spondylitis (AS), celiac disease (CD), dermatomyositis (DM), giant cell arteritis (GCA), Graves' disease (GD), Juvenile rheumatoid arthritis (JRA), localized lupus and systemic lupus erythematosus (SLE), multiple sclerosis (MS), polymyalgia rheumatica (PMR), polymyositis (PM), primary biliary cirrhosis (PBC), psoriasis (including psoriasis and related disorders, psoriasis vulgaris, and psoriatic arthropathy) (PSO), rheumatoid arthritis (RA), systemic sclerosis (SS), type 1 diabetes (T1D) (including T1D with ketoacidosis and with neurological/ophthalmic/renal manifestations), ulcerative colitis (UC), and Wegener's granulomatosis (GPA). Only HLA alleles with *P* values of less than 0.01 were included in the analysis. HLA alleles with odds ratios (ORs) of higher and lower than 1 were considered disease-predisposing and disease-protective, respectively. Only autoimmune diseases having both predisposing and protective alleles were included in the subsequent analysis. Disease-HLA associations are summarized in Table S1.

Epitope sequence datasets

HLA-I-restricted epitope sequences of 8-aa to 14-aa lengths and HLA-II-restricted epitope sequences of 9-aa to 32-aa lengths previously studied in the context of autoimmunity with annotations of functional T cell assay results were collected from Immune Epitope Database (IEDB, as of November 19th, 2018) (22). Inclusion/exclusion criteria in terms of T cell assay annotation were provided in Table S2. Epitopes studied in non-human hosts were excluded. Post-translational modifications of epitope sequences were not considered. Epitopes studied in the context of the following autoimmune diseases were identified: AS, CD, GD, GPA, MS, PBC, PSO, RA, SLE, SS, and T1D. Disease-epitope associations and accompanying annotations are summarized in Table S3.

HLA binding prediction

All four-digit HLA alleles significantly associated with a specific autoimmune disease *X* were used for HLA binding prediction of an epitope *Y* that has been studied in the context of *X*. NetMHCpan 4.0 and NetMHCIIpan 3.2 were utilized for HLA binding prediction with default parameter sets (16,17). For HLA-DP and DQ alleles, the prediction was conducted for all combinations between the disease-associated target allele and all available counterparts. For example, HLA-DQA1*0102 is associated with predisposition to MS. In this case, binding was predicted against all possible combinations of HLA-DQA1*0102 and HLA-DQB alleles. Then, medians of both predicted affinities and percentile rank values were taken as representative values for the HLA-DQA1*0102 allele.

Differential presentation analysis

Predicted percentile rank was chosen as a metric for the strength of epitope binding because this metric is not affected by inherent bias of specific HLA molecules towards higher or lower mean predicted affinities and thus allows a direct comparison between different HLA molecules (16). The highest values of sign-inverted log10-transformed percentile ranks, corresponding to the lowest percentile ranks, among predisposing and protective HLA molecules were adopted for differential presentation analysis. Differential presentation index (DPI) was defined as the transformed value of predisposing alleles subtracted by that of protective alleles. DPI is disease-dependent because the sets of predisposing and protective alleles vary between diseases. Epitopes were then categorized in a binary fashion; epitopes with DPI of higher than 0.5 and lower than -0.5 were considered putatively disease-predisposing and putatively disease-protective. Note that epitopes predicted not to bind to any of the disease-associated alleles, with the thresholds being 2% and 10% for HLA-I and HLA-II binding prediction, respectively, were excluded from this binary categorization.

Categorical associations were tested between putative disease association, self/non-self origin, and T cell recognition. Origin was determined by aligning the epitope sequence to the human proteome (UniProt ID: UP000005640). Epitopes with at least one exact match against the human proteome were defined as self-epitopes. T cell recognition (*i.e.*, immunogenicity) was determined from the IEDB annotation. The existence of at least one qualitatively positive functional T cell assay result was considered evidence of immunogenicity. Note that non-functional assays such as binding to peptide-MHC tetramer were not considered for determining immunogenicity. Inclusion/exclusion criteria in terms of T cell assay annotation were provided in Table S2.

Similarity-to-self was measured by aligning the peptide sequence against the entire human proteome (UniProt ID: UP000005640). Pairwise sequence alignment was performed using the *pairwiseAlignment* function implemented in the *Biostrings* package (59). Smith-Waterman local alignment algorithm was employed, with the substitution matrix, gap-opening cost, and gap-extension cost being PAM30, 9, and 1, respectively. These parameters are identical to those utilized in the blastp-short program (see also the BLAST Command Line Applications User Manual: https://www.ncbi.nlm.nih.gov/books/NBK279684/). The highest alignment score among the human protein sequences was used as a metric of similarity-to-self. Likewise, similarity to MBP was defined as the alignment score against human MBP (UniProt ID: P02686) using the same alignment strategy.

5 Acknowledgments

The author thanks Dr. Mai Yamakawa and Dr. Wataru Otsu for helpful discussions.

6 Conflict of Interest

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

7 Author Contributions

M.O. conceived the concept; M.O. performed computational analyses; M.O. wrote the manuscript.

8 Funding

This work is not funded by any external or internal funding sources.

9 Data Availability Statement

The datasets and in-house codes necessary to reproduce the work are available as the R package *DPA* on GitHub (https://github.com/masato-ogishi/DPA/). Full analytical scripts are available upon request.

10 References

1. Cooper GS, Stroehla BC. The epidemiology of autoimmune diseases. *Autoimmun Rev* (2003) **2**:119–125. doi:10.1016/S1568-9972(03)00006-5

2. Wang L, Wang FS, Gershwin ME. Human autoimmune diseases: A comprehensive update. *J Intern Med* (2015) **278**:369–395. doi:10.1111/joim.12395

3. Thomas SL, Griffiths C, Smeeth L, Rooney C, Hall AJ. Burden of mortality associated with autoimmune diseases among females in the United Kingdom. *Am J Public Health* (2010) **100**:2279–2287. doi:10.2105/AJPH.2009.180273

4. Walsh SJ, Rau LM. Autoimmune diseases: A leading cause of death among young and middle-aged women in the United States. *Am J Public Health* (2000) **90**:1463–1466. doi:10.2105/AJPH.90.9.1463

5. Clemente-Casares X, Blanco J, Ambalavanan P, Yamanouchi J, Singha S, Fandos C, Tsai S, Wang J, Garabatos N, Izquierdo C, et al. Expanding antigen-specific regulatory networks to treat autoimmunity. *Nature* (2016) **530**:434–40. doi:10.1038/nature16962

6. Fedorov VD, Themeli M, Sadelain M. PD-1- and CTLA-4-Based Inhibitory Chimeric Antigen Receptors (iCARs) Divert Off-Target Immunotherapy Responses. *Sci Transl Med* (2013) **5**:215ra172-215ra172. doi:10.1126/scitranslmed.3006597

7. Kim YC, Zhang A-H, Su Y, Rieder SA, Rossi RJ, Ettinger RA, Pratt KP, Shevach EM, Scott DW. Engineered antigen-specific human regulatory T cells: immunosuppression of FVIII-specific T- and B-cell responses. *Blood* (2015) **125**:1107–15. doi:10.1182/blood-2014-04-566786

8. Chae W-J, Bothwell ALM. Therapeutic Potential of Gene-Modified Regulatory T Cells: From Bench to Bedside. *Front Immunol* (2018) **9**:303. doi:10.3389/fimmu.2018.00303

9. McGovern JL, Wright GP, Stauss HJ. Engineering Specificity and Function of Therapeutic Regulatory T Cells. *Front Immunol* (2017) **8**:1517. doi:10.3389/fimmu.2017.01517

10. Hirata J, Hosomichi K, Sakaue S, Kanai M, Nakaoka H, Ishigaki K, Suzuki K, Akiyama M, Kishikawa T, Ogawa K, et al. Genetic and phenotypic landscape of the major histocompatibilty complex region in the Japanese population. *Nat Genet* (2019) doi:10.1038/s41588-018-0336-0

11. Karnes JH, Bastarache L, Shaffer CM, Gaudieri S, Xu Y, Glazer AM, Mosley JD, Zhao S, Raychaudhuri S, Mallal S, et al. Phenome-wide scanning identifies multiple diseases and disease severity phenotypes associated with HLA variants. *Sci Transl Med* (2017) **9**:1–14. doi:10.1126/scitranslmed.aai8708

12. Dendrou CA, Petersen J, Rossjohn J, Fugger L. HLA variation and disease. *Nat Rev Immunol* (2018) **18**:325–339. doi:10.1038/nri.2017.143

13. van der Merwe PA, Fugger L, Harkiolaki M, Friese MA, Kranc K, Harlos K, Jensen LT, van Boxel G, Palace J, Holmes SL, et al. T Cell-Mediated Autoimmune Disease Due to Low-

Affinity Crossreactivity to Common Microbial Peptides. *Immunity* (2009) **30**:348–357. doi:10.1016/j.immuni.2009.01.009

14. Pilli D, Zou A, Tea F, Dale RC, Brilot F. Expanding role of T cells in human autoimmune diseases of the central nervous system. *Front Immunol* (2017) **8**:1–17. doi:10.3389/fimmu.2017.00652

15. Ooi JD, Petersen J, Tan YH, Huynh M, Willett ZJ, Ramarathinam SH, Eggenhuizen PJ, Loh KL, Watson KA, Gan PY, et al. Dominant protection from HLA-linked autoimmunity by antigen-specific regulatory T cells. *Nature* (2017) **545**:243–247. doi:10.1038/nature22329

16. Peters B, Paul S, Andreatta M, Jurtz V, Marcatili P, Nielsen M. NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J Immunol* (2017) **199**:3360–3368. doi:10.4049/jimmunol.1700893

17. Jensen KK, Andreatta M, Marcatili P, Buus S, Greenbaum JA, Yan Z, Sette A, Peters B, Nielsen M. Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology* (2018) **154**:394–406. doi:10.1111/imm.12889

18. Krogsgaard M, Wucherpfennig KW, Cannella B, Hansen BE, Svejgaard A, Pyrdol J, Ditzel H, Raine C, Engberg J, Fugger L, et al. Visualization of myelin basic protein (MBP) T cell epitopes in multiple sclerosis lesions using a monoclonal antibody specific for the human histocompatibility leukocyte antigen (HLA)-DR2-MBP 85-99 complex. *J Exp Med* (2000) **191**:1395–412.

19. Ota K, Matsui M, Milford EL, Mackin GA, Weiner HL, Hafler DA. T-cell recognition of an immuno-dominant myelin basic protein epitope in multiple sclerosis. *Nature* (1990) **346**:183–187. doi:10.1038/346183a0

20. Noordzij A, van der Zee R, van Eden W, Anderton SM, Prakken B. Activation of T cells recognizing self 60-kD heat shock protein can protect against experimental arthritis. *J Exp Med* (2004) **181**:943–952. doi:10.1084/jem.181.3.943

21. de Kleer I, Vercoulen Y, Klein M, Meerding J, Albani S, van der Zee R, Sawitzki B, Hamann A, Kuis W, Prakken B. CD30 Discriminates Heat Shock Protein 60-Induced FOXP3+CD4+ T Cells with a Regulatory Phenotype. *J Immunol* (2010) **185**:2071–2079. doi:10.4049/jimmunol.0901901

22. Fleri W, Paul S, Dhanda SK, Mahajan S, Xu X, Peters B, Sette A. The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. *Front Immunol* (2017) **8**:1–16. doi:10.3389/fimmu.2017.00278

23. Wucherpfennig KW, Strominger JL. Molecular mimicry in T cell-mediated autoimmunity: viral peptides activate human T cell clones specific for myelin basic protein. *Cell* (1995) **80**:695–705. doi:10.1016/0092-8674(95)90348-8

24. Vaughan K, Peters B, O'Connor KC, Martin R, Sette A. A molecular view of multiple sclerosis and experimental autoimmune encephalitis: What can we learn from the epitope data? *J Neuroimmunol* (2014) **267**:73–85. doi:10.1016/j.jneuroim.2013.12.009

25. Delong T, Wiles TA, Baker RL, Bradley B, Barbour G, Reisdorph R, Armstrong M, Powell RL, Reisdorph N, Kumar N, et al. Pathogenic CD4 T cells in type 1 diabetes recognize epitopes formed by peptide fusion. *Science (80-)* (2016) **351**:711–714. doi:10.1126/science.aad2791

26. Congia M, Patel S, Cope AP, De Virgiliis S, Sønderstrup G. T cell epitopes of insulin defined in HLA-DR4 transgenic mice are derived from preproinsulin and proinsulin. *Proc Natl Acad Sci U S A* (1998) **95**:3833–8.

27. Ogishi M, Yotsuyanagi H. The landscapes of T cell epitope immunogenicity in sequence space. *bioRxiv* (2018)155317. doi:10.1101/155317

28. Axtell RC, de Jong BA, Boniface K, van der Voort LF, Bhat R, De Sarno P, Naves R, Han M, Zhong F, Castellanos JG, et al. T helper type 1 and 17 cells determine efficacy of interferon-beta in multiple sclerosis and experimental encephalomyelitis. *Nat Med* (2010) **16**:406–12. doi:10.1038/nm.2110

29. Jacobs L, O'Malley J, Freeman A, Ekes R. Intrathecal interferon reduces exacerbations of multiple sclerosis. *Science* (1981) **214**:1026–8.

30. Kieseier BC. The Mechanism of Action of Interferon-β in Relapsing Multiple Sclerosis. *CNS Drugs* (2011) **25**:491–502. doi:10.2165/11591110-000000000-00000

31. Ehrenreich H, Fischer B, Norra C, Schellenberger F, Stender N, Stiefel M, Sirén A-L, Paulus W, Nave K-A, Gold R, et al. Exploring recombinant human erythropoietin in chronic progressive multiple sclerosis. *Brain* (2007) **130**:2577–88. doi:10.1093/brain/awm203

32. Wucherpfennig KW. Structural basis of molecular mimicry. *J Autoimmun* (2001) **16**:293–302. doi:10.1006/jaut.2000.0499

33. McClain MT, Heinlen LD, Dennis GJ, Roebuck J, Harley JB, James J a. Early events in lupus humoral autoimmunity suggest initiation through molecular mimicry. *Nat Med* (2005) **11**:85–89. doi:10.1038/nm1167

34. Fujinami RS, von Herrath MG, Christen U, Whitton JL. Molecular mimicry, bystander activation, or viral persistence: infections and autoimmune disease. *Clin Microbiol Rev* (2006) **19**:80–94. doi:10.1128/CMR.19.1.80

35. Garg A, Kumari B, Kumar R, Kumar M. miPepBase: A database of experimentally verified peptides involved in molecular mimicry. *Front Microbiol* (2017) **8**:1–9. doi:10.3389/fmicb.2017.02053

36. Quayle AJ, Wilson KB, Li SG, Kjeldsen-Kragh J, Oftung F, Shinnick T, Sioud M, Førre Ø, Capra JD, Natvig JB. Peptide recognition, T cell receptor usage and HLA restriction elements of human heat-shock protein (hsp) 60 and mycobacterial 65-kDa hsp-reactive T cell clones from rheumatoid synovial fluid. *Eur J Immunol* (1992) **22**:1315–1322. doi:10.1002/eji.1830220529

37. Wieten L, Broere F, van der Zee R, Koerkamp EK, Wagenaar J, van Eden W. Cell stress induced HSP are targets of regulatory T cells: A role for HSP inducing compounds as anti-

inflammatory immuno-modulators? *FEBS Lett* (2007) **581**:3716–3722. doi:10.1016/j.febslet.2007.04.082

38. van Eden W, Thole JE, van der Zee R, Noordzij A, van Embden JD, Hensen EJ, Cohen IR. Cloning of the mycobacterial epitope recognized by T lymphocytes in adjuvant arthritis. *Nature* (1988) **331**:171–3. doi:10.1038/331171a0

39. Dendrou CA, Fugger L, Friese MA. Immunopathology of multiple sclerosis. *Nat Rev Immunol* (2015) **15**:545–558. doi:10.1038/nri3871

40. Pröschel C, Hansen JN, Ali A, Tuttle E, Lacagnina M, Buscaglia G, Halterman MW, Paciorkowski AR. Epilepsy-causing sequence variations in SIK1 disrupt synaptic activity response gene expression and affect neuronal morphology. *Eur J Hum Genet* (2017) **25**:216–221. doi:10.1038/ejhg.2016.145

41. Hansen J, Snow C, Tuttle E, Ghoneim DH, Yang C-S, Spencer A, Gunter SA, Smyser CD, Gurnett CA, Shinawi M, et al. De Novo Mutations in SIK1 Cause a Spectrum of Developmental Epilepsies. *Am J Hum Genet* (2015) **96**:682–690. doi:10.1016/j.ajhg.2015.02.013

42. Vroon A, Kavelaars A, Limmroth V, Lombardi MS, Goebel MU, Van Dam A-M, Caron MG, Schedlowski M, Heijnen CJ. G protein-coupled receptor kinase 2 in multiple sclerosis and experimental autoimmune encephalomyelitis. *J Immunol* (2005) **174**:4400–6.

43. Palikhe S, Ohashi W, Sakamoto T, Hattori K, Kawakami M, Andoh T, Yamazaki H, Hattori Y. Regulatory Role of GRK2 in the TLR Signaling-Mediated iNOS Induction Pathway in Microglial Cells. *Front Pharmacol* (2019) **10**:59. doi:10.3389/fphar.2019.00059

44. Hoffmann S, Cepok S, Grummel V, Lehmann-Horn K, Hackermüller J, Hackermueller J, Stadler PF, Hartung H-P, Berthele A, Deisenhammer F, et al. HLA-DRB1*0401 and HLA-DRB1*0408 are strongly associated with the development of antibodies against interferon-beta therapy in multiple sclerosis. *Am J Hum Genet* (2008) **83**:219–27. doi:10.1016/j.ajhg.2008.07.006

45. Moransard M, Bednar M, Frei K, Gassmann M, Ogunshola OO. Erythropoietin reduces experimental autoimmune encephalomyelitis severity via neuroprotective mechanisms. *J Neuroinflammation* (2017) **14**:202. doi:10.1186/s12974-017-0976-5

46. Schreiber K, Magyari M, Sellebjerg F, Iversen P, Garde E, Madsen CG, Börnsen L, Romme Christensen J, Ratzer R, Siebner HR, et al. High-dose erythropoietin in patients with progressive multiple sclerosis: A randomized, placebo-controlled, phase 2 trial. *Mult Scler J* (2017) **23**:675–685. doi:10.1177/1352458516661048

47. Monnerie H, Le Roux PD. Glutamate alteration of glutamic acid decarboxylase (GAD) in GABAergic neurons: The role of cysteine proteases. *Exp Neurol* (2008) **213**:145–153. doi:10.1016/j.expneurol.2008.05.013

48. Tohid H. Anti-glutamic acid decarboxylase antibody positive neurological syndromes. *Neurosciences (Riyadh)* (2016) **21**:215–22. doi:10.17712/nsj.2016.3.20150596

49. Cousens LP, Najafian N, Mingozzi F, Elyaman W, Mazer B, Moise L, Messitt TJ, Su Y, Sayegh M, High K, et al. In vitro and in vivo studies of IgG-derived Treg epitopes (Tregitopes): a promising new tool for tolerance induction and treatment of autoimmunity. *J Clin Immunol* (2013) **33 Suppl 1**:S43-9. doi:10.1007/s10875-012-9762-4

50. De Groot AS, Moise L, McMurry JA, Wambre E, Van Overtvelt L, Moingeon P, Scott DW, Martin W. Activation of natural regulatory T cells by IgG Fc-derived peptide "Tregitopes." *Blood* (2008) **112**:3303–3311. doi:10.1182/blood-2008-02-138073

51. Gelfand EW. Intravenous Immune Globulin in Autoimmune and Inflammatory Diseases. *N Engl J Med* (2012) **367**:2015–2025. doi:10.1056/NEJMra1009433

52. Faire MB, Konopacki C, Levine AG, Baptista AP, Germain RN, Treuting PM, Schmidt-Supprian M, Moltedo B, Hemmers S, Rudensky AY, et al. Suppression of lethal autoimmunity by regulatory T cells with a single TCR specificity. *J Exp Med* (2017)jem.20161318. doi:10.1084/jem.20161318

53. Arru G, Mameli G, Astone V, Serra C, Huang Y-M, Link H, Fainardi E, Castellazzi M, Granieri E, Fernandez M, et al. Multiple Sclerosis and HERV-W/MSRV: A Multicentric Study. *Int J Biomed Sci* (2007) **3**:292–7.

54. Morandi E, Tanasescu R, Tarlinton RE, Constantinescu CS, Zhang W, Tench C, Gran B. The association between human endogenous retroviruses and multiple sclerosis: A systematic review and meta-analysis. *PLoS One* (2017) **12**:1–18. doi:10.1371/journal.pone.0172415

55. Holderried TAW, Lang PA, Kim H-J, Cantor H. Genetic disruption of CD8+ Treg activity enhances the immune response to viral infection. *Proc Natl Acad Sci* (2013) **110**:21089–21094. doi:10.1073/pnas.1320999110

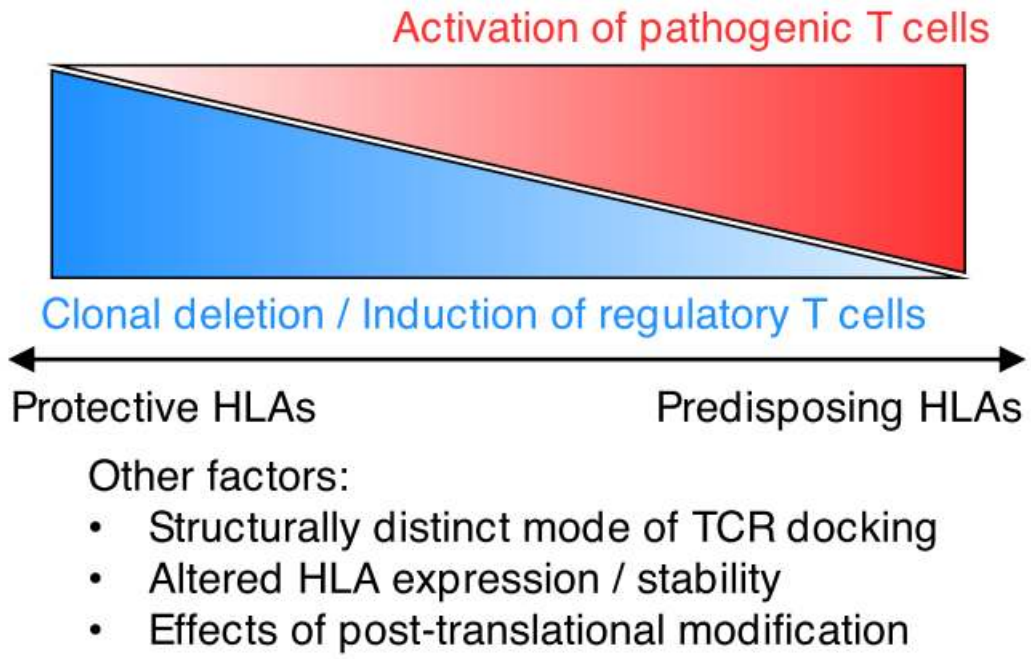
56. Smith TRF, Kumar V. Revival of CD8+ Treg–mediated suppression. *Trends Immunol* (2008) **29**:337–342. doi:10.1016/j.it.2008.04.002

57. Scally SW, Petersen J, Law SC, Dudek NL, Nel HJ, Loh KL, Wijeyewickrema LC, Eckle SBG, van Heemst J, Pike RN, et al. A molecular basis for the association of the HLA-DRB1 locus, citrullination, and rheumatoid arthritis. *J Exp Med* (2013) **210**:2569–82. doi:10.1084/jem.20131241

58. R Core Team. R: A Language and Environment for Statistical Computing. (2018)

59. Pagès H, Aboyoun P, Gentleman R DS. Biostrings: Efficient manipulation of biological strings. (2019)

642 # 11 Figures



643

644 **Figure 1. A diagram of differential peptide presentation in association with autoimmunity.**
645 Several genetic associations between HLA alleles and various autoimmune diseases have been
646 identified. Given the biological function of HLA molecules, we hypothesized that HLA molecules
647 whose alleles are genetically associated with disease predisposition presumably present either
648 pathogenic epitopes more or protective epitopes less, and likewise, HLA molecules whose alleles are
649 genetically associated with disease protection presumably present either pathogenic epitopes less or
650 protective epitopes more. Note that this scheme is oversimplified from the following viewpoints;
651 first, the genetic association with HLA loci could result from other etiologically responsible loci in
652 linkage disequilibrium; second, the different contribution to disease pathogenesis between HLA
653 alleles could be explained in multiple ways other than differential epitope presentation (12); lastly,
654 although in this study we use the percentile rank values predicted by either NetMHCpan or
655 NetMHCIIpan as a surrogate of the stability of epitope presentation, neither affinity or affinity-based
656 percentile rank is the only parameter representing the HLA-epitope thermodynamic interaction.

657

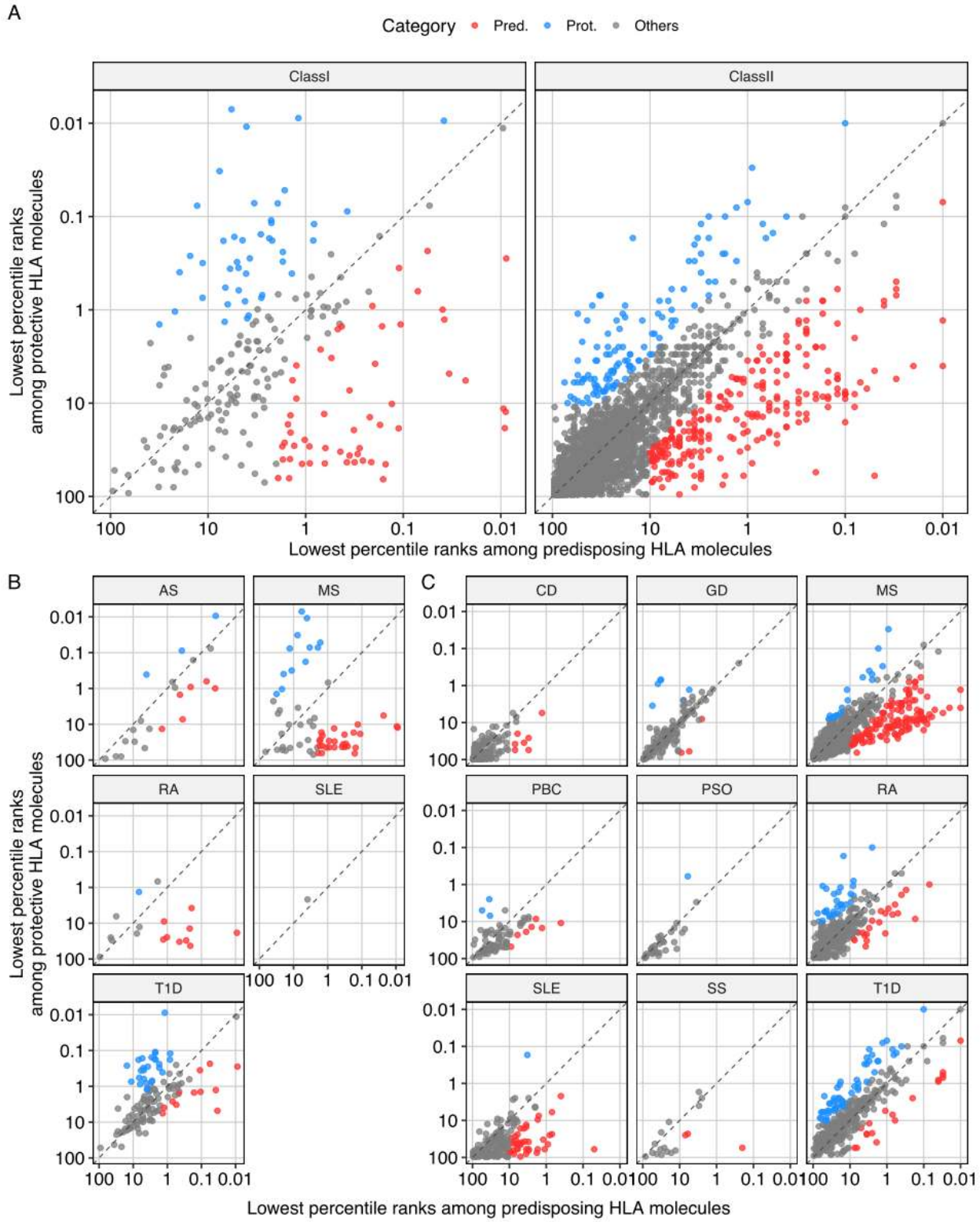


Figure 2. Differentially presented T cell epitopes studied in the context of autoimmune diseases. (A) Differential presentation analysis of autoimmunity-associated HLA-I and HLA-II epitopes. The axes represent the lowest percentile rank values (*i.e.*, the most stable binding) among disease-

predisposing and disease-protective HLA alleles. Epitopes were categorized into putatively disease-predisposing (Pred.), putatively disease-protective (Prot.), or others based on the following criteria: (i) the lowest percentile rank among HLA alleles tested was below the HLA class-specific threshold (*i.e.*, 2% and 10% in HLA-I and HLA-II epitopes, respectively); DPI was either higher than 0.5 or lower than -0.5. The DPI threshold of 0.5 was arbitrarily determined, which roughly corresponds to a three-fold change in the percentile rank. We utilized the lowest percentile rank among disease-associated HLA alleles as a representative metric to capture any epitope bound stably to at least one of the disease-associated HLAs as a potentially etiologically relevant epitope. (B and C) Differential presentation analysis of (B) HLA-I and (C) HLA-II epitopes, stratified by the autoimmune diseases in which the epitopes have been studied. AS, ankylosing spondylitis. CD, celiac disease. GD, Graves' disease. MS, multiple sclerosis. PBC, primary biliary cirrhosis. PSO, psoriasis. RA, rheumatoid arthritis. SLE, systemic lupus erythematosus. T1D, type 1 diabetes.

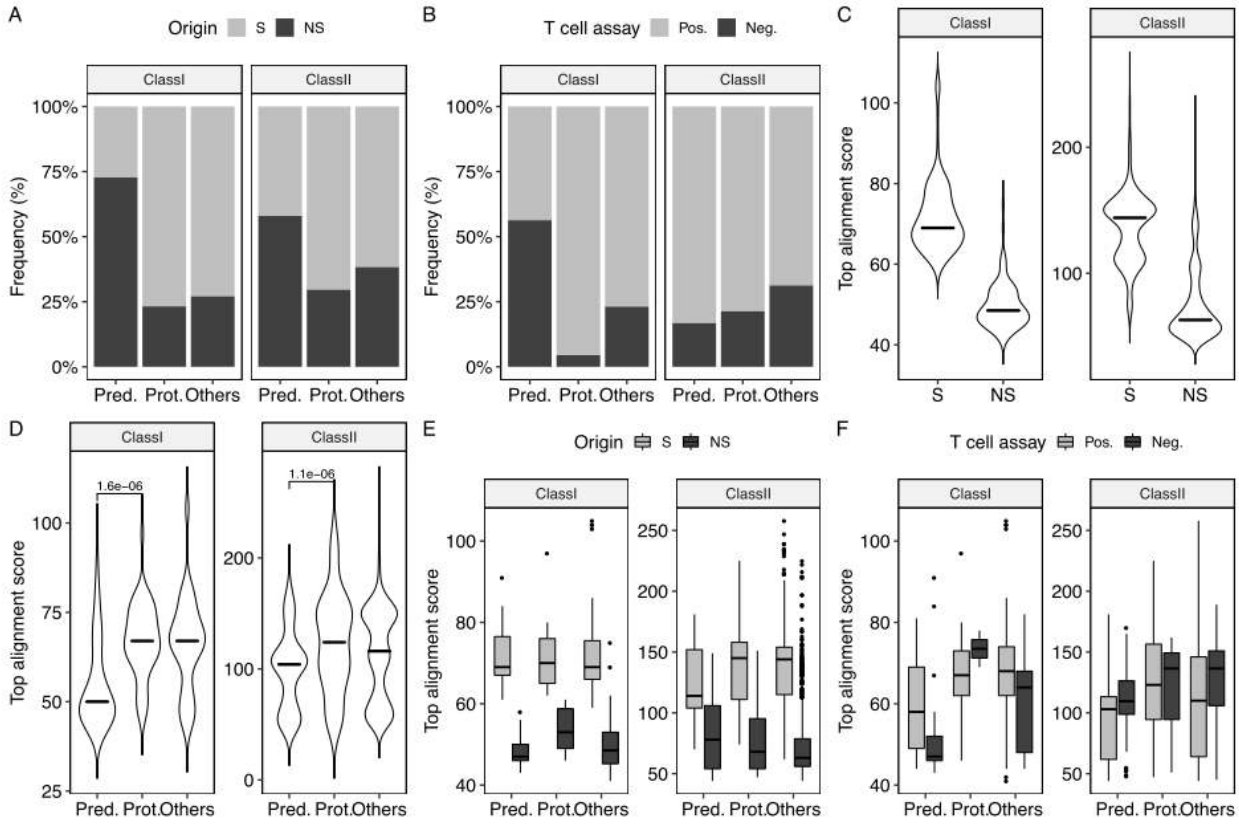
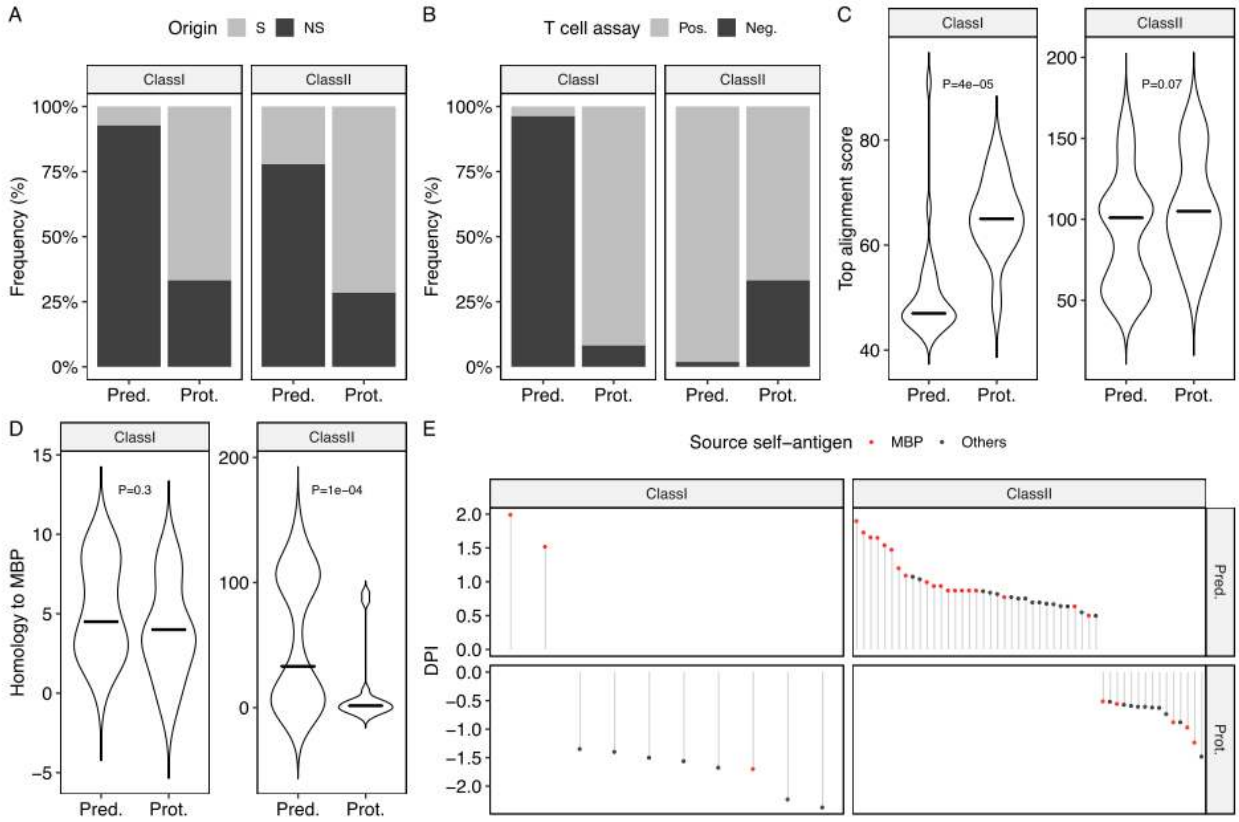


Figure 3. Analysis of differentially presented epitopes across autoimmune diseases. (A) Associations between the differential presentation categories and origins of the epitopes (N=220 and 2520 for HLA-I and HLA-II, respectively). Epitopes with and without at least one exact sequence match in the human proteome (UniProt ID: UP000005640) were considered self (S) and non-self (NS), respectively. (B) Associations between the differential presentation categories and the epitope immunogenicity determined by IEDB-derived annotations of functional T cell assay results. Epitopes with at least one positive assay result were considered positive (Pos.), and those only having negative assay results were considered negative (Neg.). (C-F) Distributions of the top sequence alignment scores against the entire human proteome as a surrogate indicator of similarity to self. Local sequence alignment was conducted by employing the Smith-Waterman algorithm with the substitution matrix and gap-opening/extension costs identical to those used in the blastp-short program. Statistical significance was determined by Wilcoxon's signed rank test.



689

690 **Figure 4. Analysis of differentially presented epitopes associated with multiple sclerosis.** (A-B)
691 (A) Origins and (B) T cell assay results of the MS-associated epitopes (N=62 and 907 for HLA-I and
692 HLA-II, respectively). (C and D) Distributions of the top sequence alignment scores against (C) the
693 entire human proteome and (D) a human MBP protein (UniProt ID: P02686). Statistical significance
694 was determined by Wilcoxon's signed rank test. (E) Prioritization of epitopes with putative
695 etiological relevance based on their DPIs. Epitopes with exact sequence matches to human MBP are
696 shown in red.

697

12 Tables

Table 1. Putatively MS-predisposing non-MBP-derived self-epitopes. DPI was calculated based on the predicted binding to MS-associated HLAs. Top five epitopes were shown. A complete list of epitopes examined can be found in Table S5.

HLA	Peptide	T cell reactivity	DPI	UniProt ID	Gene
HLA-I	RPRPVSPSSL	Negative	3.86	P57059	SIK1
	LPRKPVAGAL	Negative	3.78	Q99627	COPS8
	HPRQEQIAL	Positive	3.75	Q9NZ08	ERAP1
	KPRSPVVEL	Negative	3.74	P25098	GRK2
	RPRHQGVMV	Negative	3.73	P63267	ACTH
HLA-II	DTFRKLFRVYSNFLR	Positive	1.80	P01588	EPO
	IVRVEILRNFYFINR	Negative	1.38	P01574	IFNB
	GWISLWKGFSF	Positive	1.34	Q01955	COL4A3
	HLKRYYGRILHYLKA	Negative	1.33	P01574	IFNB
	LFRVYSNFLRGKLKL	Positive	1.33	P01588	EPO

703 **Table 2. Putatively MS-protective non-MBP-derived self-epitopes.** DPI was calculated based on
704 the predicted binding to MS-associated HLAs. Top five epitopes were shown. A complete list of
705 epitopes examined can be found in Table S5.

HLA	Peptide	T cell reactivity	DPI	UniProt ID	Gene
HLA-II	GWYTYMLVPAALTGL	Negative	-1.80	A1A5B4	ANO9
	NFFRMVISNPA	Positive	-1.60	Q05329	GAD2
	QMTFLRLLSASAHQN	Positive	-1.60	P20908	COL5A1
	FFRMVISNPAATHQD	Positive	-1.54	Q05329	GAD2
	FLKKFHFLKGATLC	Positive	-1.48	Q4ZG55	GREB1

706

Table 3. Human- and pathogen-derived HLA-II epitope pairs having the same differential presentation category in common. A complete list of epitopes examined can be found in Table S6.

Disease	Pathogen	Protein		Peptide		DPI		Category	Ref
		Human	Pathogen	Human	Pathogen	Human	Pathogen		
MS	*M. avium*	MBP	Transposase	ENPVVHFFKNIVTPR	QRCRVHFLRNVLAQV	1.90	0.64	Predisposing	(24)
	M. tuberculosis		Transposase		QRCRVHFMRNLYTAV		1.07		
	B. subtilis		YqeE		ALAVLHFYPDKGAKN		0.61		
RA	*M. bovis*	HSP60	HSP65	HRKPLVIIAEDVDGE	AGKPLLIIAEDVEGE	-1.03	-1.00	Protective	(36)

Supplementary Material

Supplementary Figures

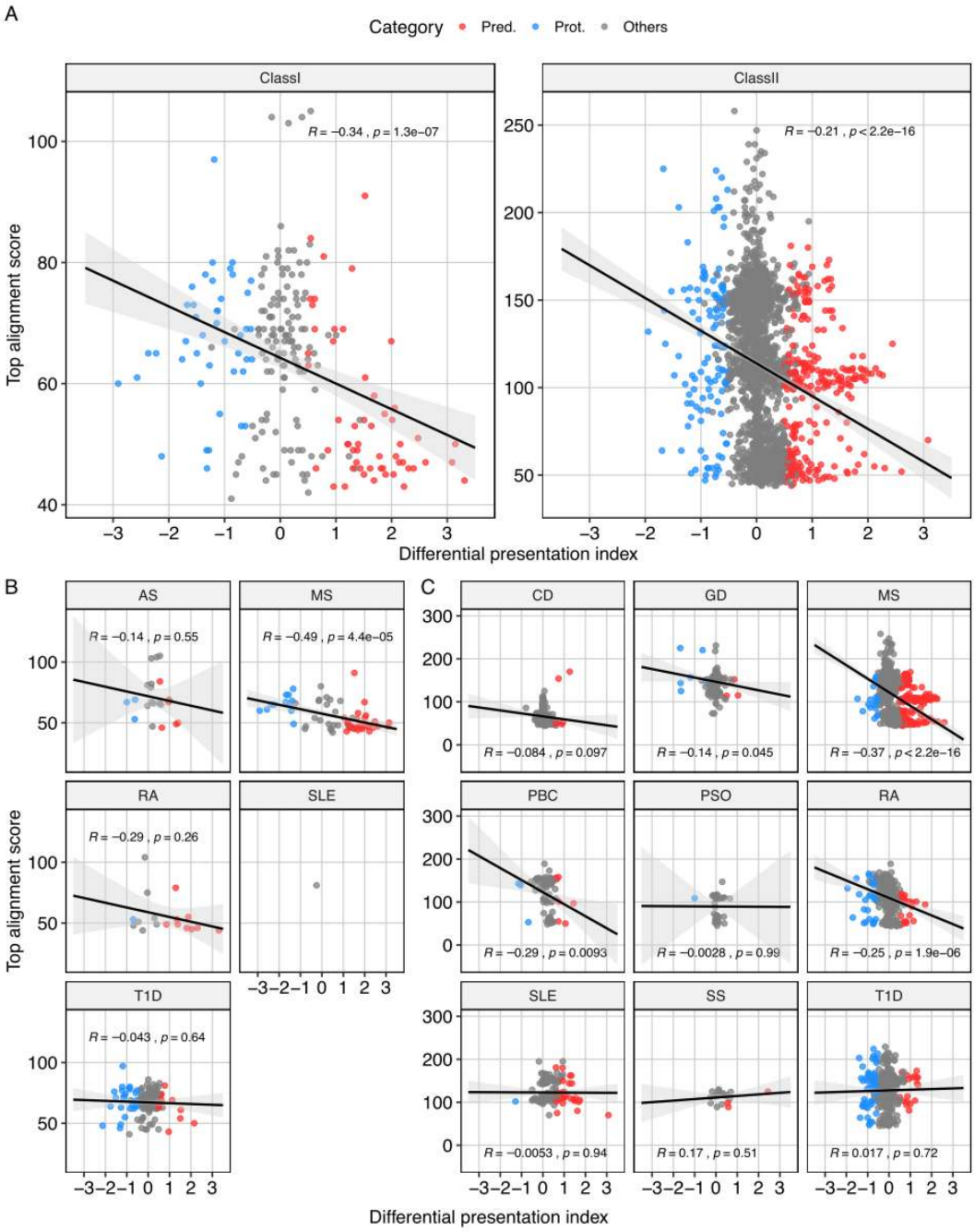


Figure S1. Correlations between DPI and similarity-to-self. (A) DPI-selfness two-dimensional plots of autoimmunity-associated HLA-I and HLA-II epitopes. Autoimmunity-associated epitopes were analyzed, and disease-specific DPI scores were calculated. The highest alignment score against the human proteome (UniProt ID: UP000005640) was also computed for each of the epitopes. (B and C) Two-dimensional plots of (B) HLA-I and (C) HLA-II epitopes facetted by the associated diseases. *R* indicates Pearson's correlational coefficient. For disease abbreviations, see Figure 2.

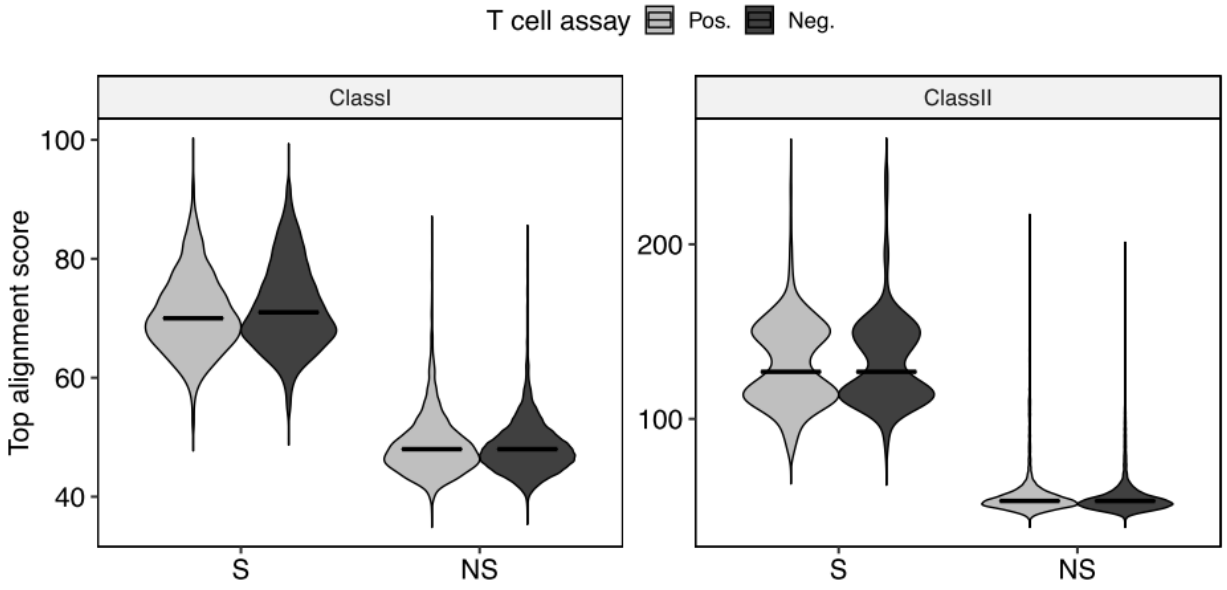


Figure S2. Minimal differences of similarity-to-self among immunogenic and non-immunogenic epitopes. Epitopes studied in various contexts were compiled previously. Epitopes of self (S) and non-self (NS) origins were identified based on the presence or absence of at least one exact sequence match in the human proteome. Immunogenicity was determined based on the presence or absence of at least one positive T cell assay annotation.

Supplementary Tables (Separate files)

Table S1. A summary of disease-HLA allele associations.

Table S2. Inclusion/exclusion criteria for T cell assay annotations.

Table S3. A summary of epitopes previously studied in the context of autoimmune diseases.

Table S4. The results of differential presentation analysis in autoimmunity-associated epitopes.

Table S5. The results of MS-specific differential presentation analysis in self-epitopes previously studied in various contexts.

Table S6. The results of differential presentation analysis in self-epitopes and corresponding pathogen-derived epitopes with evidence of molecular mimicry.