

**Prediction and characterization of the T cell epitopes for the major soybean protein
allergens using bioinformatics approaches**

Running Title: T cell epitopes of soybean allergens

Fanlin Zhou ^a, Shudong He ^{*,a}, Yi Zhang ^{*,b}, Yongfei Wang ^a, Hanju Sun ^a, Qian Liu ^c

^a Engineering Research Center of Bio-process of Ministry of Education, School of Food and Biological Engineering, Hefei University of Technology, Hefei 230009, Anhui, PR China;

^b IPREM, E2S UPPA, CNRS, Université de Pau et des Pays de l'Adour, 64000 Pau, France

^c College of Food Science, Northeast Agricultural University, Harbin 150030, Heilongjiang, PR China

*** Corresponding Authors:**

Shudong He: Shudong.he@hfut.edu.cn; Tel. (86) 551-62901505; Fax. (86) 551-62901516.

Yi Zhang: yi.zhang@univ-pau.fr

ORCID:

Shudong He: 0000-0002-1155-4928

Yi Zhang: 0000-0003-1780-0963

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the [Version of Record](#). Please cite this article as doi: [10.1002/prot.26233](https://doi.org/10.1002/prot.26233) © 2021 Wiley Periodicals, Inc.

Received: Jul 19, 2021; Revised: Aug 14, 2021; Accepted: Aug 30, 2021

Abstract

Protein allergens is a health risk for consumption of soybeans. To understand allergenicity mechanism, T cell epitopes of 7 soybean allergens were predicted and screened by abilities to induce cytokine interleukin (IL) 4. The relationships among amino acid composition, properties, allergenicity and pepsin hydrolysis sites were analyzed. Among the 138 T cell epitopes identified, YIKDVFRVIPSEVLS, KDVFRVIPSEVLSNS, DVFRVIPSEVLSNSY of Gly m 6.0501 (P04347), and AKADALFKAIEAYLL, ADALFKAIEAYLLAH of Gly m 4.0101 (P26987) were the most possible epitope candidates. In T cell epitopes pattern, the frequencies of amino acids Q, D, E, P and G decreased, while F, I, N, V, K, H, A, L and S increased. Hydrophobic residues at positions p1 and p2 and positively charged residues in positions p13 might contribute to allergenicity. Most of epitopes could be hydrolyzed by pepsin into small polypeptides within 12 residues length, and the anti-digestive epitope regions contained I, V, S, N, and Q residues. T cell epitopes EEQRQQEGVIVELSK from Gly m 5.03 (P25974) showed resistance to pepsin hydrolysis and would cause a higher Th2 cell response. This research provides basis for the development of hypoallergenic soybean products in the soybean industry as well as for the immunotherapy design for protein allergy.

Keywords: T cell epitope, soybean protein, food allergen, bioinformatics

1. Introduction

In recent years, the incidence of food allergies has been on the rise, and about 10% of people in the world are suffering from food allergies [1]. So far, 390 proteins have been identified as food protein allergens by the World Health Organization and the International Union of Immunological Societies (WHO/IUIS) Allergen Nomenclature Sub-Committee, and more than 50% of the food protein allergens belong to eggs, milk, nuts, wheat, crustaceans, beans, fish, and peanuts [2]. Soybean is a common source for legume allergen, and about 25% of food allergies are caused by soybeans [3, 4], not only will induce a variety of pathological reactions such as intestinal injury, stomach discomfort, or allergic dermatitis, anaphylactic shock, [3], but also limit the development and application of soybean products [4]. So far, eight soybean protein allergens have been identified and submitted to the allergen database (<http://www.allergen.org/>), namely hydrophobic protein (Gly m 1), defensin (Gly m 2), profilin (Gly m 3), pathogenesis-related protein (Gly m 4), β -conglycinin (Gly m 5), glycinin (Gly m 6), seed biotinylated protein (Gly m 7) and 2S albumin (Gly m 8). Besides, Gly m Bd 28k, Gly m Bd 30k, trypsin inhibitor, lectin, and Gly 50 kDa have been also confirmed as the common soybean allergens [5]. However, among these soybean proteins, only seven of them, i.e., Gly m 4.0101 (Uniprot ID: P26987), Gly m 5.0201 (Uniprot ID: P11827), Gly m 5.03 (Uniprot ID: P25974), Gly m 6.0101 (Uniprot ID: P04776), Gly m 6.0501 (Uniprot ID: P04347), trypsin inhibitor (Uniprot ID: P01070) and lectin (Uniprot ID: P05046) (Table 1), have been characterized with crystal structures as seen in the Uniprot database (<https://www.uniprot.org/>).

Previous studies had indicated that soybean allergy is usually mediated by type 2 CD4⁺ T cells (Th2) [5]. In detail, the T cell epitopes of soybean allergens combine with major

Accepted Article

histocompatibility complex (MHC) class II proteins, which can be recognized by Th2 cells and induce the release of IL-4, IL-5, and IL-13, and other interleukins that promote B cell proliferation and differentiation to produce antibodies and play an important role in activating mast cells, basophils, and eosinophils [2]. Since antigen recognition by T cells is a critical step, the analyses of T cell epitopes and related cytokines are of vital importance for understanding of the mechanism of anaphylaxis. Generally, T cell epitopes typically consist of 12-20 amino acids, and the T cell epitope chain less than 12 amino acids would not efficiently stimulate CD4 cells [6]. For the peptide fragment that are bound by MHC class II molecules, they are usually longer (13-15 amino acids) where the core sequences have length of 9 amino acids [7]. MHC molecules are known as peptide-binding glycoproteins, a complex that affect the immune response. For major histocompatibility antigen system in human body, the histocompatibility leukocyte antigens (HLAs) encoded by the MHC genes produce the Class I and class II molecules, and HLA class II molecules (HLA-DP, HLA-DQ, HLA-DR) alleles are gene complexes encoding human MHC class II proteins [8]. However, the alleles of HLA class II alleles are not clear, and the epitope peptides fragment bound to HLA class II molecules as well as the binding affinity are rarely described.

The identification of T cell epitopes of food allergens helps with the understanding of the allergy mechanisms and the development of hypoallergenic foods and immunotherapy design. As previously reported, employing appropriate proteases to pre-hydrolyze food allergens are effective to disrupt the allergenicity of food proteins, but the protease enzymes need to be selected based on the anti-digestion area of epitopes [9]. For example, there are many hypoallergenic children's formula milk products that are developed via enzymatic pre-

hydrolysis method, in order to effectively prevent children from milk allergy, and improve taste and nutritional value [10]. Furthermore, T cell epitopes have been considered as safe and effective immunotherapy modulators [2]. For example, the allergic reactions from shrimp tropomyosin and egg ovalbumin studied on mice could be effectively reduced by T cell epitopes oral immunotherapy [11, 12]. The knowledge of allergen-specific T cell epitopes is also useful to develop T cell epitope targeted vaccines for immunotherapy [13].

Allergen T cell epitopes are conventionally identified using peptide scanning technology via synthesizing multiple overlapping peptides that span the length of the protein in the combination with *in vitro* T cell stimulation tests, but this approach was costly in the peptide synthesis and serum preparation [14]. Although the lymphocyte proliferation assay in the T cell test is simple, inexpensive and sensitive, it takes long time and can not identify and quantify the cytokines. Flow cytometry is a novel T cell epitope identification method with high throughput screening to save time, but the equipment is expensive [15]. In this context, bioinformatics has been applied as an initial screening method and widely used in the prediction of food allergen epitopes with advantages including high throughput, fast analysis speed, and low cost, benefiting from database development [16]. In our previous studies, seven different bioinformatics tools, ProPred, SYFPEITHI, NetMHCII, NN-align, SMM-align, NetMHCIIpan and RANKPEP, were employed to predict the T cell epitopes of a black kidney bean, from which two of three potential T cell epitopes were confirmed using cytokine and lymphocyte proliferation experiments [17], suggesting the prediction accuracy of bioinformatics analysis. Besides, thirty-six T cell epitopes of the peanut allergen Ara h 1 were obtained by NetMHCIIpan 2.0 tool, and 14 of predicted T cell epitopes bind to HLA

Accepted Article

molecules in vitro, and 35 of T cell epitopes induced T cell proliferation differentiation and the production of IL13 [14]. By now, only one soybean allergen, Gly m 6, with identified epitopes, is recorded in SDAP database, and most of the relevant studies pay attention to the exploration of B cell epitopes from soybean allergens. For instance, Zeece et al. [18] identified the IgE epitope (aa 192-306) in the region of soybean globulin G1 acidic subunit by western blot. Helm et al. [19] used the peptide scanning technique to identify 11 linear B cell epitopes of soybean globulin G1, and the epitopes AGVALSRCTLN (aa 62-72) had cross-reactivity with a peanut allergen. Saeed et al. [20] identified 9 glycinin epitopes and found that glycinin cross-reacted with peanuts, almonds, and walnuts. However, there is a lack of research on the T cell epitopes of soybean allergens. In addition to epitope prediction, bioinformatics can also be used to analyze epitope amino acids by building machine learning models. For example, random forest is a nonlinear ensemble learning method based on bagging algorithm to measure the significant degree of predictive parameters and produce precise outcomes [21]. Thus, the relationship between amino acid descriptor and epitope allergenicity can be established by mathematical method and random forest algorithm. In this study, in order to make a comprehensive understanding of T cell epitopes of soybean allergens, seven soybean protein allergens characterized with crystal structures were subjected for bioinformatic analyses. T cell epitopes of soybean allergens were predicted by in silico tools, and the binding affinity with different HLA II alleles and the ability to induce IL-4 as well as the allergenicity of epitopes were also evaluated to understand the allergenicity mechanisms. Amino acid composition and the significance as well as the pepsin hydrolysis sites of epitopes were analyzed.

2. Methods

2.1 Allergen amino acid sequence retrieval

The amino acid sequences of P01070 (trypsin inhibitor), P04347 (Gly m 6.0501), P04776 (Gly m 6.0101), P05046 (lectin), P11827 (Gly m 5.0201), P25974 (Gly m 5.03), and P26987 (Gly m 4) were downloaded from the Uniprot database (<https://www.uniprot.org/>), and the detailed information of the seven soy allergens were shown in Table 1.

2.2 T cell epitope prediction

The “MHC-II Binding Predictions” tool in the IEDB (Immune Epitope Database Analysis Resource) database was used to predict the soybean allergen peptides that could bind to HLA class II molecules [22]. The potential sequence was submitted to the software in Fasta format, and the IEDB recommended method, combining with the consensus method and the NetMHCIIpan method, was selected for prediction, and thereinto, the consensus method considers the combination of any three of four methods, including artificial neural network (ANN) alignment method, stabilization matrix (SMM) alignment method, combinatorial library method, and Sturnioli method. A total of 27 HLA molecules (15 HLA-DR molecules, 6 HLA-DQ molecules, and 6 HLA-DP molecules) were used to predict T cell epitopes, and the epitope length parameter was set to 15 amino acids as a suggestion. T cell peptides with percentile ranking value ≤ 4 , stable matrix method $IC_{50} \leq 250$ nM and neural network method $IC_{50} \leq 250$ nM obtained by the consensus method, as well as the peptides with the percentile ranking value ≤ 4 and NetMHCIIpan $IC_{50} \leq 250$ nM obtained by the NetMHCIIpan method were selected for further analyses.

2.3 Evaluation of the ability to induce IL-4

The IL4pred tool (<https://webs.iitd.edu.in/raghava/il4pred/>) was used to calculate the

ability of epitopes to induce IL-4 secretion by Th2 cells, and a hybrid prediction method consisting of support vector machine and motif was selected. The support vector machine threshold was selected as 0.2 [23].

2.4 Allergenicity evaluation

The amino acid sequences of the T cell epitopes were submitted to the online tool AllerTop 2.0 (<https://www.ddg-pharmfac.net/AllerTOP/>), and the amino acid properties, such as hydrophobicity, size, helix-forming propensity, relative abundance, and β -strand forming propensity, were applied for the allergenicity prediction [24, 25].

2.5 Epitope amino acid composition analysis

The amino acid composition of T cell epitopes and the frequency were analyzed according to the methods of He et al. [17] and Zheng et al. [26]. In brief, the amino acid types and numbers of T cell epitopes in each soybean allergen were counted, and the amino acid frequency was calculated by the ratio of the amino acid number of T cell epitope in the whole protein.

2.6 Significance analysis of epitope amino acids

The amino acid descriptor 3z scales (z1 descriptor represents hydrophilicity, z2 represents the bulk, and z3 represents electricity properties) were used as the X variable and the T cell epitope allergenicity (yes/no) was set as the Y variable [27] (shown in supplementary Table S1). Each T cell epitope had 15 aa and the number of variables for each soybean allergen was 45 (3*15), expressed as p1z1-p15z3. The discriminant analysis was performed with the XLSTAT software (Trial version, Addinsoft, New York, NY) to confirm the accuracy of the amino acid descriptor 3z scales, then the significance analysis of T cell epitope amino acids was carried out by using the logistic regression and random forests via the Easy Fit function of

XLSTAT software.

2.7 Simulated pepsin digestion of epitopes

The online tool PeptideCutter (https://web.expasy.org/peptide_cutter/) was used to predict the pepsin digestion sites of the T cell epitopes of soybean allergens [28].

3. Results

3.1 T cell epitope prediction results

In order to compare the binding affinity between different peptides and HLA class II molecules, a unified scale has been generated in the prediction results of the IEDB-percentile ranking value, as the lower percentile ranking value with higher affinity [29]. In addition, the inhibitory concentration 50 (IC₅₀) value was also used to calculate the peptide binding affinity, as the lower, IC₅₀ value indicated the stronger affinity [25]. Therefore, peptides with a strong binding ability to MHC class II molecules would be screened out based on the peptide binding affinity score with IC₅₀ ≤ 250 nM and percentile ranking value ≤ 4 (shown in supplementary Table S2-S15). Subsequently, screened T cell peptides were further selected by the IL-4 inducing ability analysis, and the AllerTop 2.0 tool was finally used to confirm the allergenicity potential.

3.1.1 P01070

Soybean trypsin inhibitor (Uniprot ID: P01070) is an anti-nutritional factor in soybean [30], consisting of 216 amino acids, with a protein molecular weight of 24005 Da. Forty-five peptides with the high binding ability to MHC class II molecules were obtained by setting IC₅₀ ≤ 250 nM and percentile ranking value ≤ 4, and 22 peptides could induce IL-4 secretion by Th2

cells based on the IL4pred tool analysis (shown in supplementary Table S2-S3). Furthermore, a total of 14 peptides were confirmed as the T cell epitopes via the allergenicity analysis using the AllerTOP v. 2.0 tool (Table 2), which were mainly located in three regions of the protein: aa 44-60, aa 86-111, and aa 179-199. The region 44-60 are located inside the protein, whereas segments aa 86-111, and aa 179-199 were exposed to the protein surface (Table 2). Thereinto, the epitopes "YRIRFIAEGHPLSLK" (aa 86-100) and "RIRFIAEGHPLSLKF" (aa 87-101) might have a higher ability to induce Th2 to produce IL4 because of the higher IL-4pred scores. Furthermore, as shown in Table 2, the bindings of the epitopes "PLSLKFDSFAVIMLC" (aa 96-110) and "LSLKFDSEFAVIMLCV" (aa 97-111) with the most diverse HLA class II alleles were observed, indicating these two epitopes can cause more people to be allergic than other epitopes.

3.1.2 P04347

Gly m 6.0101, Gly m 6.0201, Gly m 6.031, Gly m 6.0401, and Gly m 6.0501 are the 5 subunits (G1-G5) of the Glycinin, which belong to the 11S plant seed storage protein [31]. Gly m 6.0501 (Uniprot code: P04347) containing 516 amino acids with a molecular weight of 57956 Da was analyzed in the study, and a total of 22 T cell epitopes were screened (Table 3), which were mainly located in 5 protein regions of aa 183-199, aa 231-246, aa 260-275, aa 390-407, and aa 461-495. Thereinto, protein fragments of aa 231-246, aa 260-275, and aa 471-494 are posited on the surface of the protein (Table 3). Compared with other epitopes, the four epitopes of "GLEYYVVKTHHNAVS" (aa 461-475), "LEYVVKTHHNAVSS" (aa 462-476), "YVVKTHHNAVSSYI" (aa 464-478), and "VVKTHHNAVSSYIK" (aa 465-479) had higher IL-4pred scores.

3.1.3 P04776

Gly m 6.0101 (Uniprot ID: P04776) is the G1 subunit of glycinin, which contains 495 amino acids and has a molecular weight of 55706 Da. As shown in Table 4, a total of 34 T cell epitopes, mainly concentrated in the 7 regions of aa 158-172, aa 217-236, aa 319-333, aa 347-362, aa 366-387, aa 412-442 and aa 468-491, that could induce Th2 cell to produce IL4 and have allergenicity potential. Among them, the regions of aa 217-236, aa 319-333, and aa 468-491 are exposed on the surface (Table 4). T cell epitopes in the aa 217-236 region could bind the largest number of HLA class II alleles, indicating the allergic susceptibility, and the epitopes "ILSGFTLEFLEHAFS" (aa 220-234) and "LSGFTLEFLEHAFSV" (aa 221-235) had higher IL-4pred scores.

3.1.4 P05046

Soybean lectin (Uniprot ID: P05046) is an anti-nutritional factor with 285 amino acids and the molecular weight is approximately 120 kDa [32]. As shown in Table 5, a total of 25 peptides were confirmed as the T cell epitopes that can induce Th2 to produce IL4 and have allergenicity potential. All T cell epitopes are located on the surface of the protein except for the region aa 234-250 (Table 5). Furthermore, the epitopes "LVLLTSKANS AETVS" (aa 23-37), "EWVRIGFSAATGLDI" (aa 234-248), "VRIGFSAATGLDIPG" (aa 236-250), "HDVLSWSFASNLP HA" (aa 253-267), "DVLSWSFASNLP HAS" (aa 254-268) and "VLSWSFASNLP HASS" (aa 255-269) exhibited higher IL-4pred scores. Especially, the epitope "ASFAASFNF TFYAPD" (aa 100-114) was predicted by three methods at the same time, and more HLA class II alleles were observed to be bound by this epitope.

3.1.5 P11827

Soybean β -conglycinin is a 7S seed storage protein containing three subunits (α , α' and β) [33]. Gly m 5.0201 (Uniprot ID: P11827) is the α' subunit of β -conglycinin and has high amino acid sequence homology with the α and β subunits, and the molecular weight is 72,228 Da containing 621 amino acids [34]. Through IEDB tools, IL-4pred, and AllerTOP v. 2.0 prediction, a total of 17 T cell epitopes were screened (Table 6), and the epitopes "PFHFNSKRFQTLFKN" (aa 211-225) and "FHFNSKRFQTLFKNQ" (aa 212-226) combined the most types of alleles, and they also had strong potentials for inducing Th2 cells to produce IL4. Also, these two epitopes are exposed on the surface of the protein (Table 6).

3.1.6 P25974

Soybean β -conglycinin β subunit (Uniprot code: P25974) is composed of 439 amino acids and has a molecular weight of 50476 Da. A total of 15 T cell epitopes were obtained with allergic potential (Table 7). Compared with other epitopes, the epitopes "EEQRQQEGVIVELSK" (aa 201-215), "EQRQQEGVIVELSKE" (aa 202-216), and "RNPIYSNNFGKFFEI" (aa 245-259) had higher IL-4pred scores and were located on the surface of the protein.

3.1.7 P26987

Gly m 4 (Uniprot ID: P26987) is the pathogenesis-related protein that belongs to the PR protein family. It has cross-reactivity with apple allergen Mal d 1, birch pollen allergen Bet v 1, and other allergens [35]. Gly m 4 has a protein length of 158 amino acids and a molecular weight of 16772 Da, while a total of 12 T cell epitopes were screened by the prediction method (Table 8). These 12 T cell epitopes are all posited on the surface of the protein. Compared with other epitopes, the epitopes "AKADALFKAIEAYLL" (aa 138-152) and

"ADALFKAIEAYLLAH" (aa 140-154) bond the most types of HLA class II alleles, 13 and 15 different HLA alleles, respectively.

3.2 Amino acid composition of T cell epitopes

As shown in Fig. 1, the ratio of each amino acid in the entire protein and that in the T cell epitope fragment were counted, respectively. All 20 types of amino acids were observed in four of the seven proteins (P01070, P04347, P04776, and P11827), and the P05046 protein and P25974 protein contained 19 types of different amino acids except for cysteine (C) and tryptophan (W), respectively, while 17 kinds of amino acids were found in the P26987 protein except for arginine (R), tryptophan (W), and cysteine (C). In short, lysine (L), serine (S), valine (V), glutamate (E), and asparagine (N) were found to be rich in the seven protein allergens, and fewer tryptophane (W), methionine (M) and cysteine (C) were observed.

By comparing the frequency of each amino acid in the protein sequence with that in the T cell epitopes, the frequencies of glutamine (Q) and glycine (G) obviously reduced in the T cell epitope region of the six proteins (except for P25974 protein). The appearance frequency of glutamate (E) in the T cell epitope regions of P01070, P04347, P04776, P05046, and P11827 proteins also decreased, and a similar phenomenon was also observed in the T cell epitope region of P04347, P05046, P11827, P25974, and P26987 proteins for proline (P), while aspartate (D) decreased in P01070, P04347, P05046, P25974 and P26987 proteins. On the other hand, the frequency of phenylalanine (F) in T cell epitope regions increased in all seven proteins. Histidine (H) and isoleucine (I) increased in six proteins, and the appearance frequency of asparagine (N) alanine (A), and serine (S) increased in five proteins. In addition, the frequency of leucine (L), valine (V), and lysine (K) in T cell epitope regions also increased

in more than half of the proteins (4/7) (shown in Fig. 1 and supplementary Table S16).

3.3 Significance analysis of epitope amino acids

A discriminant analysis model was used to analyze the relationship between the allergenicity (yes or no, Y- variable) of the epitope peptides and the physical and chemical properties of the amino acids (X- variables).

The present study used 3 descriptors to describe the amino acid physical and chemical properties of the T cell epitopes for building the random forests models with variable importance analysis. The confusion matrix (validation sample) correct (%) of seven soybean allergens were: 40% (P01070), 57.143% (P04347), 41.667% (P04776), 55.556% (P05406), 72.727% (P11827), 60% (P25974), and 75% (P26987). The variable importance of the X-variables were determined by examining the mean decrease accuracy obtained through random forests analysis of the quantitative X- variables and qualitative Y-variable (Fig.2).

As shown in Fig.2, the variable p1z1 significantly contributed to the allergenicity (yes) in the five soybean allergens (P01070, P04776, P05406, P11827, P25974). According to mean decrease accuracy values, variables p2z1, p6z2, p13z3 were beneficial to the allergenicity (yes) in four soybean allergens. Through the calculation, the occurrence of allergenicity (yes), p1, and p6 was the most important position to allergenicity (yes), followed by p2, p4, p5, and p13. The soybean allergens P01070, P11827, and P25974 expressed the bulk of the amino acid at the p1 position, whereas P04776 and P05406 expressed the electronic property of the amino acid at the p1 position. The amino acids at the p6 position can have a good contribution to allergenicity (yes) in the six allergens except for P04776. Especially, for the allergen P26987, the hydrophobicity, bulk, and electronic property of amino acid at position p6 promoted the

allergenicity (yes).

In the soybean allergens including P04347, P11827, P25974, P26987, the most important amino acid property for allergenicity (yes) is z1 (hydrophobicity), followed by z2 (bulk) and z3 (electronic property). Electronic property is the most important amino acid property for the allergen P05046, whereas bulk is the most important amino acid property for the allergens of P01070 and P04776.

Except for the Y- variable allergenicity (yes), the random forest models also provided the relationship between X- variable and Y- variable allergenicity (total) to find important variables. From the Fig.2, we found that the variable p1z1 can affect both allergenicity (no) and allergenicity (yes) in most allergens (P01070, P04776, P05406, and P25974), and the variable p6z1 contributed to both allergenicity (no) and allergenicity (yes) in three allergens (P04347, P11827, and P26987).

3.4 Simulated pepsin digestion of T cell epitopes

Most of the soybean proteins would be broken down by pepsin in the stomach, and only the anti-digestive peptides containing epitopes could enter the gastric mucosa and interact with immune cells to trigger allergic reactions [36]. Thus, the Peptidecutter tool was used to simulate the gastric digestion of seven soybean allergens, respectively, and the pepsin hydrolysis sites of the T cell epitopes were shown in Table 9, which further confirm the potential allergenicity.

As shown in Table 9, except for the T cell epitope "SKDNVISQIPSQVQE" in P11827 protein, all other epitopes contained multiple pepsin hydrolysis sites, and most of the digested peptides lengths were less than 12 amino acids. Since the length of the T cell epitope is generally believed between 12-20 amino acids [15], the peptide sequences of

"DDGTRRLVVSKNKP" (aa 179-192) in P01070 protein, "TPVVAVSIIDTNS" (aa 158-170), "RHNIGQTSSPDI" (aa 322-333), and "SQARQIKNNNP" (aa 468-479) in P04776 protein, "NPHIGINVNSIRSIKTTS" (aa 168-185) in P05046 protein, "SKDNVISQIPSQVQE" (aa 558-572) in P11827 protein, "EEQRQQEGVIVE" (aa 201-212) in P25974 protein and the "LVTADADNVIPKA" (aa 23-34) in P26987 protein would be the digestion resistant regions. Among the 8 anti-digestive regions, isoleucine (I), valine (V), serine (S), asparagine (N), and glutamine (Q) were the amino acids with the highest content, while isoleucine (I), valine (V), and asparagine (N) covered 7 anti-digestive areas and serine (S) covered 6 anti-digestive areas. Interestingly, all 8 anti-digestive regions are exposed on the surface of the proteins except "TPVVAVSIIDTNS" (aa 158-170) in P04776 (shown in Table 2-Table 8), and as shown in Table 9, six peptides remain allergenicity among these 8 peptides. It has been reported that amino acids which had the high surface exposure were likely to bind with antibodies and formed B cell epitopes, however, the relationship between T cell epitopes and surface exposure has not been elaborated yet.

4. Discussion

At present, the food allergy is not exactly explained, which has hindered the development of hypoallergenic foods and immunotherapy [37, 38]. The identification of T cell epitopes can provide important information about the cellular mechanisms involved in the transition from food tolerance to allergy [39], and the analysis of cytokines related to Th2 cells can also improve our understanding of the immune response. Soybean is one of the most widely cultivated edible legumes. In the past 50 years from 1968 to 2018, the world soybean

Accepted Article

production increased 8.4 times, and the global planted area also increased 4.3 times, while global soybean consumption increased from 24 million tons to 57 million tons between 1999 and 2019 [40]. Soybean is rich in carbohydrates, protein, fat, cellulose, and trace elements such as calcium and magnesium [41]. Although soybean is the high-quality protein source for humans and animals, the prevalence of sensitization of the major allergens from soybean has aroused wide attention. After entering the human body, soybean protein allergens will be treated by lysosomes and then combined with MHC class II molecules, and then the MHC class II-T cell epitope complex will interact with CD4⁺ T cells to synthesize and secrete the cytokine such as IL4, which will play a key role in promoting the proliferation and differentiation of antigen-presenting cells and stimulating the production of IgE [23]. In this study, the T cell epitopes of seven soybean protein allergens were comprehensively identified based on the binding affinity between epitopes and HLA II molecular, the inducing ability of IL4 production, and the antigenicity analysis, while the larger fragments released by pepsin digestion were also analyzed. As far as we have known, it might be the first paper to analyze T cell epitopes of soybean allergens.

The combination of consensus method and NetMHCIIpan in IEDB have been used to predict T cell epitopes of allergens that can bind to MHC II molecules, and the NetMHCIIpan method has been recognized as the most accurate epitope prediction method [42]. Ramesh et al. [14] predicted 54 T cell epitopes of peanut allergen Ara h 1 by NetMHCIIpan method and identified 32 T cell epitopes were confirmed by using ProImmune REVEAL assay, indicated the higher accuracy (32 confirmed/37 predicted) of the NetMHCIIpan method. Pascal et al. [43] identified 4 T cell epitope regions of peanut allergen Ara h 2 via T cell proliferation assay and

Accepted Article

cytokine profile analysis, and all of the T cell epitopes were further proved to be MHC II binders by NetMHCIIpan method and NetMHCII method. In order to further improve accuracy on the prediction of T cell epitopes, combining consensus method and the NetMHCIIpan method in the IEDB database with other bioinformatics tools were explored in the previous studies, and all the predicted 6 T cell epitopes exhibited good potentials in T-cell proliferation and cytokines release [44]. Except for previous T cell epitopes prediction methods, the allergenicity and the ability to promote Th2 response-related cytokines IL-4 release of T cell epitopes were also evaluated in the current studies, which would further improve the prediction accuracy.

As shown in Tables 2-8, since the literature on T cell epitopes mapping of soybean has not been available until now, all the T-cell epitopes are first identified. The soybean allergen Gly m 4 belongs to the PR family and has cross-reactivity with the birch pollen allergen Bet v 1. Jahn- Schmid et al. [45] found that the "TLLRAVESYLLAHSD" (aa 142-156) of the Bet v 1 allergen was the major T cell epitope and could cross-react with the peptide segment of soybean Gly m 4 "ALFKAIEAYLLAHPD" (aa 142-156), and this peptide fragment is consistent with the T cell epitopes "ADALFKAIEAYLLAH" (aa 140-154) and "DALFKAIEAYLLAHP" (aa 141-155) of Gly m 4 predicted in this study. In addition, comparing with the conformational epitope of Gly m 4 reported by Husslik et al. [46], many conformational epitope amino acid residues were existed in the Gly m 4 T cell epitope regions "LYKALVTDADNVIPKA" (aa 19-34), "KKITFLEDGETKFVLHKIESI" (aa 54-74), "AGPNGGSAGKLTVKY" (aa 106-120), and "AKADALFKAIEAYLL" (aa 138-152) reported in the current experiment.

Accepted Article

Allergen-specific T cells play an important role in allergic reactions and are obvious targets for immunotherapy intervention in diseases [17]. Recently, Candreva et al. [47] developed a new strategy for preventing milk allergy on the basis of oral administration of a soybean-derived peptide that was cross-reactive with bovine caseins. This peptide contained both T and B cell epitopes of soybean allergen Gly-m-Bd-30K and could stimulate T cells without causing IgE cross-linking on basophils and mast cells. In this study, the T cell epitope region "GGSILSGFTLEFLEHAFSVD" (217-236) in the P04776 allergen reported in this experiment is consistent with the previous results of the B cell epitope "GGSILSGFTLEFLEHAFSV" (217-235) identified by peptide scanning [48]. The T cell epitope regions "EEQRQQEGVIVELSK" (aa 201-215), "NPIYSNNFGKFFEIT" (aa 246-260), and "DIFLSSVDINEGALLLPHFNS" (aa 271-291) in P25974 allergen reported in this experiment are consistent with the epitopes identified by co-immunoprecipitation and mass spectrometry [49]. These T cell epitopes which contained B cell epitope amino acids would be expected for allergy treatment via Synthetic Peptide Immuno-Regulatory Epitopes for related cross-reactive soybean.

Selecting potential T cell epitopes would be a difficult task because of the complexity of HLA alleles. In this study, since the epitopes "YIKDVFRVIPSEVLS" (aa 477-491), "KDVFRVIPSEVLSNS" (aa 479-493), "DVFRVIPSEVLSNSY" (aa 480-494) in P04347 protein could bind with more than 9 HLA molecules, and the epitopes "AKADALFKAIEAYLL" (aa 138-152), "ADALFKAIEAYLLAH" (aa 140-154) in P26987 protein could bind with more than 13 HLA molecules, and these epitopes can be predicted by all three methods (shown in Tables 3 and 8), the five new fragments are considered as the most

possible epitope candidates.

Compared with the entire protein sequence, the frequency of phenylalanine (F), isoleucine (I), asparagine (N), valine (V) lysine (K), alanine (A), and serine (S), leucine (L) and histidine (H) increased in T cell epitopes of most soybean allergens. According to previous studies, the presence of lysine (K) had a significant effect on T cell stimulation and secondary structure alpha-helix could promote the antigenicity [50]. Another study also found that isoleucine (I) and histidine (H) was the key amino acids in the T cell epitope of the dust mite allergen Der p 2, because they might direct contact moieties, and might also indirectly affect peptide binding by changing the conformation of adjacent amino acid side chains [51].

A previous study showed that MHC molecular tended to bind hydrophobic amino acids in most positions except penultimate position, and could bind both hydrophilic and hydrophobic amino acids in positions 4 and 6, which could bind hydrophobic amino acids in positions 1, 2, 3, 5, and 8 [7]. Similarly, in the present paper, random forest models also showed that amino acids in positions p1, p2, p4, p5, p6, and p13 had a good contribution to the allergenicity. For positions p4, p5, and p6, three physicochemical properties (z1, z2, and z3) contributed to allergenicity in a similar way. For positions p1 and p2, the physicochemical properties that contribute to allergenicity were in the order $z1 > z2 > z3$, whereas the order for position p13 was $z3 > z2 > z1$. Furthermore, the hydrophobic (z1) residues at positions p1 and p2 contributed to the allergenicity in most soybean allergen models. In the position p1 of P04776, P05046, P11827, and P25974 models and the position p2 of P04347, P04776, P11827, and P26987, there are many hydrophobic residues including which can promote allergenicity (shown in Table 2-8 and Fig. 2). Positively charged amino acids, such as arginine (R), lysine

(K), and histidine (H) tend to locate at the position p13 and play an important role in the allergenicity in the soybean models P04347, P04776, P25974, and P05046. In this study, the frequency of phenylalanine (F), isoleucine (I), asparagine (N), valine (V) lysine (K), and histidine (H) residues increased in the T cell epitope region, which was consistent with the current analysis, as hydrophobic residues including phenylalanine (F), isoleucine (I), and valine (V) located at the positions p1 and p2 contributed to the allergenicity, whereas positively charged amino acids, such as lysine (K) and histidine (H) located at the positions p6 and p13 and promoted the allergenicity (shown in Table 3-5, Table 7, and Fig. 2), and previous studies also confirmed the positively charged amino acids would provide a net charge to increase the activity of the epitope peptide antigen [51].

In simulated pepsin digestion experiments, most of the T cell epitopes from soybean allergen could be hydrolyzed by pepsin into small peptides (<12 aa), and most of the anti-digestive fragments are located on the surface of the proteins. The digestion-resistant epitope region contained much hydrophobic amino acids including isoleucine (I), valine (V), and non-charged amino acids including serine (S), asparagine (N), and glutamine (Q), and the second position of this peptide with hydrophobic amino acid might contribute to the allergenicity, such as the anti-digestive fragment "LVTADNVIPKA" (aa 23-34) in P26987. The T cell epitopes EEQRQQEGVIVELSK" (aa 201-215), and "EQRQQEGVIVELSKE" (aa 202-216) with higher IL-4pred scores also can be resistant to pepsin hydrolysis and have a great potential to enter into bodies to cause Th2 cell response. In order to development of soybean hypoallergenic products, proteases that can hydrolyze the above five amino acid sites (I, V, S, N, and Q) can be selected. For example, the endopeptidase Glu-C can hydrolyze glutamate (E) and glutamine

(Q) [52, 53, 54] can be used to specifically destroy the digestion-resistant T cell epitopes "SKDNVISQIPSQVQE" (aa 558-572) (P11827) and EEQRQQEGVIVE" (aa 201-212) (P25974).

5. Conclusions

T cell epitopes of seven soybean allergens were predicted by using the "MHC-II Binding Predictions" tools in the IEDB database, and 138 T cell epitopes were obtained by the further evaluation of the ability to induce the production of IL4 and allergenicity by bioinformatics tools, and 5 predicted fragments of "YIKDVFRVIPSEVLS" (aa 477-491), "KDVFRVIPSEVLSNS" (aa 479-493), "DVFRVIPSEVLSNSY" (aa 480-494) in P04347 protein and "AKADALFKAIEAYLL" (aa 138-152), "ADALFKAIEAYLLAH" (aa 140-154) in P26987 protein were considered as the most possible epitope candidates. Based on the amino acid composition analysis and random forest model, hydrophobic residues (such as phenylalanine, isoleucine, and valine) at positions p1 and p2 and positively charged amino acid residues (such as lysine and histidine) in positions p13 would have a good contribution to the allergenicity. Furthermore, most of the epitopes of T cells could be hydrolyzed by pepsin into small molecular peptides (<12 aa), and anti-digestive epitope regions contained more isoleucine (I), valine (V), serine (S), asparagine (N), and glutamine (Q). In the current studies, bioinformatics strategies provide an efficient method for epitope prediction and analysis, and the results may aid to study the mechanism of soybean sensitization. Additionally, epitopes reported here are expected to apply in immunotherapy designing and hypoallergenic soybean products developing in the food industry.

Abbreviation

WHO/IUIS, World Health Organization and the International Union of Immunological Societies; MHC, Major Histocompatibility Complexes; HLA, Human Leukocyte Antigens; IEDB, Immune Epitope Database Analysis Resource; ANN, Artificial Neural Network; SMM, Stabilization Matrix; IL, Interleukin

Access dates of websites

Allergen nomenclature. WHO/IUIS Allergen Nomenclature Sub-Committee. (2020). <http://www.allergen.org/> Accessed 21 July 2020. Uniprot database. (2020). <https://www.uniprot.org/> Accessed 10 September 2020. IL4pred tool. (2020). <https://webs.iiitd.edu.in/raghava/il4pred/> Accessed 23 November 2020. AllerTop 2.0. (2020). <https://www.ddg-pharmfac.net/AllerTOP/> Accessed 2 November 2020. PeptideCutter. (2020). https://web.expasy.org/peptide_cutter/ Accessed 25 December 2020.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgments

The authors thank the hard work and diligence of the predecessors who worked in the food protein allergen area and appreciate the helpful comments from the anonymous reviewers. The authors are also grateful for the financial support from the National Natural Science

Foundation of China (No. 31701524), Major Special Science and Technology Planning Project of Anhui Province, China (No. 202003b06020030), Key Research and Development Program of Anhui Province, China (No. 202104a06020016).

References

- [1] S.H. Sicherer, H.A. Sampson, Food Allergy: A review and update on epidemiology, pathogenesis, diagnosis, prevention and management, *Journal of Allergy & Clinical Immunology*, 141 (2017) 41-58. doi: 10.1016/j.jaci.2017.11.003.
- [2] F. Zhou, S. He, H. Sun, Y. Wang, Y. Zhang, Advances in epitope mapping technologies for food protein allergens: A review, *Trends in Food Science & Technology*, 107 (2020) 226-239. doi: 10.1016/j.tifs.2020.10.035.
- [3] X. Sun, X. Shan, Z. Yan, Y. Zhang, L. Guan, Prediction and characterization of the linear IgE epitopes for the major soybean allergen β -conglycinin using immunoinformatics tools, *Food & Chemical Toxicology*, 56 (2013) 254-260. doi: 10.1016/j.fct.2013.02.014.
- [4] G. Bu, T. Li, T. Zhu, G. Xi, Identification of the linear immunodominant epitopes in the β subunit of β -conglycinin and preparation of epitope antibodies, *International Journal of Biological Macromolecules*, 154 (2020) 724-731. doi: 10.1016/j.ijbiomac.2020.03.159.
- [5] A.K. Verma, S. Kumar, M. Das, P.D. Dwivedi, A comprehensive review of legume allergy, *Clinical Reviews in Allergy & Immunology*, 45 (2013) 30-46. doi: 10.1007/s12016-012-8310-6.
- [6] Malherbe, Laurent, T-cell epitope mapping, *Annals of Allergy Asthma & Immunology*, 103 (2009) 76-79. doi: 10.1016/S1081-1206(10)60147-0

- [7] P. Walden, T-cell epitope determination, *Current Opinion in Immunology*, 8 (1996) 68. doi: 10.1016/S0952-7915(96)80107-5.
- [8] W.-L. Fan, M.-S. Shiao, R.C.-Y. Hui, S.-C. Su, C.-W. Wang, Y.-C. Chang, W.-H. Chung, HLA association with drug-induced adverse reactions, *Journal of Immunology Research*, 2017 (2017) 1-10. doi: 10.1155/2017/3186328.
- [9] T. Rahaman, T. Vasiljevic, L. Ramchandran, Effect of processing on conformational changes of food proteins related to allergenicity, *Trends in Food Science & Technology*, 49 (2016) 24-34. doi: 10.1016/j.tifs.2016.01.001.
- [10] G. Bu, Y. Luo, F. Chen, K. Liu, T. Zhu, Milk processing as a tool to reduce cow's milk allergenicity: a mini-review, *Dairy Science & Technology*, 93 (2013) 211-223. doi: 10.1007/s13594-013-0113-x.
- [11] C.Y.Y. Wai, N.Y.H. Leung, P.S.C. Leung, K.H. Chu, T cell epitope immunotherapy ameliorates allergic responses in a murine model of shrimp allergy, *Clinical & Experimental Allergy*, 46 (2016) 491-503. doi: 10.1111/cea.12684.
- [12] P. Rupa, Y. Mine, Oral immunotherapy with immunodominant T-cell epitope peptides alleviates allergic reactions in a Balb/c mouse model of egg allergy, *Allergy*, 67 (2012) 74-82. doi: 10.1111/j.1398-9995.2011.02724.x.
- [13] R.E. Soria-Guerra, R. Nieto-Gomez, D.O. Govea-Alonso, S. Rosales-Mendoza, An overview of bioinformatics tools for epitope prediction: Implications on vaccine development, *Journal of Biomedical Informatics*, 53 (2015) 405-414. doi: 10.1016/j.jbi.2014.11.003.
- [14] M. Ramesh, A. Yuenyongviwat, G.N. Konstantinou, J. Lieberman, M. Pascal, M. Masilamani, H.A. Sampson, Peanut T-cell epitope discovery: Ara h 1, *Journal of Allergy and*

Clinical Immunology, 137 (2016) 1764-1771. doi: 10.1016/j.jaci.2015.12.1327.

[15] T.A. Ahmad, A.E. Eweida, L.H. El-Sayed, T-cell epitope mapping for the design of powerful vaccines, Vaccine Reports, 6 (2016) 13-22. doi: 10.1016/j.vacrep.2016.07.002.

[16] C. Wang, Y. Wang, G. Liu, L. Fu, Food allergomics based on high-throughput and bioinformatics technologies, Food Research International, 130 (2020) 108942. doi: 10.1016/j.foodres.2019.108942.

[17] S. He, J. Zhao, W. Elfalleh, M. Jemaà, H. Sun, X. Sun, M. Tang, Q. He, Z. Wu, F. Lang, In silico identification and in vitro analysis of B and T-cell epitopes of the black turtle bean (*Phaseolus vulgaris* L.) lectin, Cellular Physiology & Biochemistry, 49 (2018) 1600-1614. doi: 10.1159/000493496.

[18] M.G. Zeece, T.A. Beardslee, J.P. Markwell, G. Sarath, Identification of an IgE-binding region in soybean acidic glycinin G1, Food & Agricultural Immunology, 11 (1999) 83-90. doi: 10.1080/095401099999942.

[19] R.M. Helm, G. Cockrell, C. Connaughton, H.A. Sampson, G.A. Bannon, V. Beilinson, N.C. Nielsen, A.W. Burks, A soybean G2 glycinin allergen, International Archives of Allergy and Immunology, 123 (2000) 205-212. doi: 10.1159/000024446.

[20] H. Saeed, C. Gagnon, E. Cober, S. Gleddie, Using patient serum to epitope map soybean glycinins reveals common epitopes shared with many legumes and tree nuts, Molecular Immunology, 70 (2016) 125-133. doi: 10.1016/j.molimm.2015.12.008.

[21] S. Talukdar, K.U. Eibek, S. Akhter, S. Ziaul, A. Reza, M.T. Islam, J. Mallick, Modeling fragmentation probability of land-use and land-cover using the bagging, random forest and random subspace in the Teesta River Basin, Bangladesh, Ecological Indicators, 126 (2021)

107612. doi: 10.1016/j.ecolind.2021.107612.

[22] R. Vita, J.A. Overton, J.A. Greenbaum, J. Ponomarenko, J.D. Clark, J.R. Cantrell, D.K.

Wheeler, J.L. Gabbard, D. Hix, A. Sette, B. Peters, The immune epitope database (IEDB) 3.0, *Nucleic Acids Research*, 43 (2015) 405-412. doi: 10.1093/nar/gku938.

[23] S.K. Dhanda, S. Gupta, P. Vir, G.P.S. Raghava, Prediction of IL4 inducing peptides, *Clinical & Developmental Immunology*, 2013 (2013) 1-9. doi: 10.1155/2013/263952.

[24] I. Dimitrov, D.R. Flower, I. Doytchinova, AllerTOP v.2--a server for in silico prediction of allergens, *BMC Bioinformatics*, 20 (2014) 1-6. doi: 10.1186/1471-2105-14-S6-S4.

[25] K. Narayanan, In silico prediction of potential vaccine candidates on capsid protein of human bocavirus 1, *Molecular Immunology*, 93 (2018) 193-205. doi: 10.1016/j.molimm.2017.11.024.

[26] L.-N. Zheng, H. Lin, R. Pawar, Z.-X. Li, M.-H. Li, Mapping IgE binding epitopes of major shrimp (*Penaeus monodon*) allergen with immunoinformatics tools, *Food & Chemical Toxicology*, 49 (2011) 2954-2960. doi: 10.1016/j.fct.2011.07.043.

[27] S. Hellberg, M. Sjoestroem, B. Skagerberg, S. Wold, Peptide quantitative structure-activity relationships, a multivariate approach, *Journal of Medicinal Chemistry*, 30 (1987) 1126-1135. doi: 10.1021/jm00390a003.

[28] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein identification and analysis tools on the ExPASy server, (In) John M. Walker (ed): *The Proteomics Protocols Handbook*, Humana Press (2005) pp. 571-607.

[29] Sinu Paul, C.S.L. Arlehamn, T.J. Scriba, M.B.C. Dillon, C. Oseroff, D. Hinz, D.M. McKinney, S.C. Pro, J. Sidney, B. Peters, A. Sette, Development and validation of a broad

scheme for prediction of HLA class II restricted T cell epitopes, *Journal of Immunological Methods*, 422 (2015) 28-34. doi: 10.1016/j.jim.2015.03.022.

[30] V. Kumar, A. Rani, P. Mittal, M. Shuaib, Kunitz trypsin inhibitor in soybean: contribution to total trypsin inhibitor activity as a function of genotype and fate during processing, *Journal of Food Measurement & Characterization*, 13 (2019) 1583-1590. doi: 10.1007/s11694-019-00074-y.

[31] S. Natarajan, C. Xu, H. Bae, B.A. Bailey, P. Cregan, T.J. Caperna, W.M. Garrett, D. Luthria, Proteomic and genetic analysis of glycinin subunits of sixteen soybean genotypes, *Plant Physiology & Biochemistry*, 45 (2007) 436-444. doi: 10.1016/j.plaphy.2007.03.031.

[32] T. Wang, G.-x. Qin, Z.-w. Sun, Y. Zhao, Advances of research on glycinin and β -conglycinin: A review of two major soybean allergenic proteins, *Critical Reviews in Food Science & Nutrition*, 54 (2014) 850-862. doi: 10.1080/10408398.2011.613534.

[33] L. L'Hocine, J.I. Boye, Allergenicity of soybean: New developments in identification of allergenic proteins, cross-reactivities and hypoallergenization technologies, *Critical Reviews in Food Science and Nutrition*, 47 (2007) 127-143. doi: 10.1080/10408390600626487.

[34] H.B. KRISHNAN, W.-S. KIM, S. JANG, M.S. KERLEY, All three subunits of soybean β -conglycinin are potential food allergens, *Journal of Agricultural and Food Chemistry*, 57 (2009) 938-943. doi: 10.1021/jf802451g.

[35] T. Geng, K. Liu, R. Frazier, L. Shi, E. Bell, K. Glenn, J.M. Ward, Development of a sandwich ELISA for quantification of Gly m 4, a soybean allergen, *Journal of Agricultural & Food Chemistry*, 63 (2015) 4947-4953. doi: 10.1021/acs.jafc.5b00792.

[36] J. Huang, C. Liu, Y. Wang, C. Wang, M. Xie, Y. Qian, L. Fu, Application of in vitro and

in vivo models in the study of food allergy, *Food Science & Human Wellness*, 7 (2018) 235-243. doi: 10.1016/j.fshw.2018.10.002.

[37] L.T. Villuendas, J. Cuestaherranz, M. Gonzalezmuñoz, L.F. Pacios, E. Compes, B. Garcíacarrasco, G.S. Duran, A.D. Perales, T-cell epitopes of the major peach allergen, Pru p 3: Identification and differential T-cell response of peach-allergic and non-allergic subjects, *Molecular Immunology*, 46 (2009) 722-728. doi: 10.1016/j.molimm.2008.10.018.

[38] Glaspole, I. N., De Leon, M. P., Rolland, J. M., & O'Hehir, R. E. Characterization of the T-cell epitopes of a major peanut allergen, Ara h 2, *Allergy*, 60 (2005) 35-40. doi: 10.1111/j.1398-9995.2004.00608.x.

[39] E.A. Pastorello, M. Monza, V. Pravettoni, R. Longhi, P. Bonara, J. Scibilia, L. Primavesi, R. Scorza, Characterization of the T-cell epitopes of the major peach allergen Pru p 3, *International Archives of Allergy and Immunology*, 153 (2010) 1-12. doi: 10.1159/000301573.

[40] M. De Maria, E.J. Robinson, J.R. Kangile, R. Kadigi, I. Dreoni, M. Couto, N. Howai, J. Peci, Global soybean trade-the geopolitics of a bean, (2020). doi: 10.34892/7yn1-k494.

[41] N. Güzeler, Ç. Yildirim, The utilization and processing of soybean and soybean products, *Journal of Agricultural Faculty of Uludag University*, 30 (2016) 546-553.

[42] L. Zhang, K. Udaka, H. Mamitsuka, S. Zhu, Toward more accurate pan-specific MHC-peptide binding prediction: a review of current methods and tools, *Briefings in Bioinformatics*, 13 (2012) 350-364. doi: 10.1093/bib/bbr060.

[43] M. Pascal, G. Konstantinou, M. Masilamani, J. Lieberman, H. Sampson, In silico prediction of Ara h 2 T cell epitopes in peanut - allergic children, *Clinical & Experimental Allergy*, 43 (2013) 116-127. doi: 10.1111/cea.12014.

- [44] L.L. Xu, M. Gasset, H. Lin, C. Yu, J.L. Zhao, X.W. Dang, Z.X. Li, Identification of the dominant T-cell epitopes of Lit v 1 shrimp major allergen and their functional overlap with known B-cell epitopes, *Journal of Agricultural and Food Chemistry*, 69 (2021) 7420–7428. doi: 10.1021/acs.jafc.1c02231.
- [45] B. Jahn-Schmid, A. Radakovics, D. Lüttkopf, S. Scheurer, S. Vieths, C. Ebner, B. Bohle, Bet v 1142-156 is the dominant T-cell epitope of the major birch pollen allergen and important for cross-reactivity with Bet v 1-related food allergens, *Journal of Allergy and Clinical Immunology*, 116 (2005) 213-219. doi: 10.1016/j.jaci.2005.04.019.
- [46] F. Husslik, J. Nürnberg, C.S.v. Loetzen, T. Mews, B.K. BallmerWeber, J. Kleine-Tebbe, R. Treudler, J.-C. Simon, S. Randow, E. Völker, A. Reuter, P. Rösch, S. Vieths, T. Holzhauser, D. Schiller, The conformational IgE epitope profile of soya bean allergen Gly m 4, *Clinical & Experimental Allergy*, 46 (2016) 1484-1497. doi: 10.1111/cea.12796.
- [47] N.M. Candreva, P.L. Smaldini, A. Cauerhff, S. Petruccelli, G.H. Docena, A novel approach to ameliorate experimental milk allergy based on the oral administration of a short soy cross-reactive peptide, *Food Chemistry*, 346 (2020) 128926. doi: 10.1016/j.foodchem.2020.128926.
- [48] T.A. Beardslee, M.G. Zeece, G. Sarath, J.P. Markwell, Soybean glycinin G1 acidic chain shares IgE epitopes with peanut allergen Ara h 3, *International Archives of Allergy & Immunology*, 123 (2000) 299-307. doi: 10.1159/000053642.
- [49] Y. Zhao, G. Naren, J. Qiang, G. Qin, N. Bao, M.H. Farouk, Identification of allergic epitopes of soybean β -conglycinin in different animal species, *Frontiers in Veterinary Science*, 7 (2020) 599546. doi: 10.3389/fvets.2020.599546.
- [50] I. Berkower, G.K. Buckenmeyer, J.A. Berzofsky, Molecular mapping of a

histocompatibility-restricted immunodominant T cell epitope with synthetic and natural peptides: implications for T cell antigenic structure, *Journal of Immunology*, 136 (1986) 2498-2503. doi: 10.1084/jem.163.4.1030.

[51] B. Wu, G. Toussaint, L.V. Elst, C. Granier, M.G. Jacquemin, J.-M.R. Saint-Remy, Major T cell epitope-containing peptides can elicit strong antibody responses, *European Journal of Immunology*, 30 (2000) 291-299. doi: 10.1002/1521-4141(200001)30:1<291::AID-IMMU291>3.0.CO;2-T.

[52] M. Weiss, B. Dhayalan, Y.-S. Chen, N. Phillips, M. Swain, N. Rege, A. Mirsalehi, M. Jarosinski, F. Ismail-Beigi, N. Metanis, Reassessment of an innovative insulin analogue excludes protracted action yet highlights the distinction between external and internal diselenide bridges, *Chemistry-A European Journal*, 26 (2020) 4695. doi: 10.1023/A:1007042825783.

[53] N. Priolo, S.M.d. Valle, M.C. Arribére, L. López, a.N. Caffini, Isolation and characterization of a cysteine protease from the latex of *Araujia hortorum* fruits, *Journal of Protein Chemistry*, 19 (2000) 39-49. doi: 10.1023/A:1007042825783.

[54] C.E. Osorio, N. Wen, J.H. Mejías, S. Mitchell, D.v. Wettstein, S. Rustgi, Directed-mutagenesis of *flavobacterium meningosepticum* prolyl-oligopeptidase and a glutamine-specific endopeptidase from barley, *Frontiers in Nutrition*, 7 (2020) 11. doi: 10.3389/fnut.2020.00011.

Table/Figure Captions

Table 1 The names, structures, and related properties of seven soybean allergens.

Table 2 List of putative epitopes of P01070 along with interacting HLA class II alleles and relevant scores.

Table 3 List of putative epitopes of P04347 along with interacting HLA class II alleles and relevant scores.

Table 4 List of putative epitopes of P04776 along with interacting HLA class II alleles and relevant scores.

Table 5 List of putative epitopes of P05046 along with interacting HLA class II alleles and relevant scores.

Table 6 List of putative epitopes of P11827 along with interacting HLA class II alleles and relevant scores.

Table 7 List of putative epitopes of P25974 along with interacting HLA class II alleles and relevant scores.

Table 8 List of putative epitopes of P26987 along with interacting HLA class II alleles and relevant scores.

Table 9 Pepsin cleavage site of T cell epitopes.

Fig. 1 Analysis of the amino acid composition of T cell epitopes of seven soybean allergens.

(P01070: Amino acid composition analysis of P01070 protein; P01070-T: Amino acid composition analysis of T cell epitopes of P01070 protein. P04347: Amino acid composition

analysis of P04347 protein; P04347-T: Amino acid composition analysis of T cell epitopes of P04347 protein. P04776: Amino acid composition analysis of P04776 protein; P04776-T: Amino acid composition analysis of T cell epitopes of P04776 protein. P05046: Amino acid composition analysis of P05046 protein; P05046-T: Amino acid composition analysis of T cell epitopes of P05046 protein. P11827: Amino acid composition analysis of P11827 protein; P11827-T: Amino acid composition analysis of T cell epitopes of P11827 protein. P25974: Amino acid composition analysis of P25974 protein; P25974-T: Amino acid composition analysis of T cell epitopes of P25974 protein. P26987: Amino acid composition analysis of P26987 protein; P26987-T: Amino acid composition analysis of T cell epitopes of P26987 protein.)

Fig. 2 Mean decrease accuracy of the 3-z scale models of the (A) P01070 T cell epitopes, (B) P04347 T cell epitopes, (C) P04776 T cell epitopes, (D) P05046 T cell epitopes, (E) P11827 T cell epitopes, (F) P25974 T cell epitopes, and (G) P26987 T cell epitopes.

Table 1 The names, structures, and related properties of seven soybean allergens.

Protein name	IUIS name	Uniprot name	Number of amino acids	Molecular weight (Da)
Pathogenesis-related protein	Gly m 4.0101	P26987	158	16772
β-conglycinin	Gly m 5.0201	P11827	481	54242
	Gly m 5.0301/Gly m 5.0302	P25974	439	50476
Glycinin	Gly m 6.0101	P04776	495	55706
	Gly m 6.0501	P04347	516	57956
Kunitz trypsin inhibitor	-	P01070	216	24005
Lectin	-	P05046	285	30928

Table 2 List of putative epitopes of P01070 along with interacting HLA class II alleles and relevant scores.

HLA class II alleles	Method	Position	Sequence	IL-4pred	AllerTOP v. 2.0	Spatial position
HLA-DRB1*01:01/ HLA-DQA1*05:01/ HLA-DQB1*03:01	A	44-58	LSDITAFGGIRAAPT	0.89	YES	Inner
HLA-DQA1*05:01/ HLA-DQB1*03:01	A	45-59	SDITAFGGIRAAPTG	0.95	YES	Inner
HLA-DRB1*01:01/ HLA-DQA1*05:01/ HLA-DQB1*03:01	A	46-60	DITAFGGIRAAPTGN	0.84	YES	Inner
HLA-DRB3*02:02	C	86-100	YRIRFIAEGHPLSLK	1.33	YES	Exposed
HLA-DRB3*02:02	C	87-101	RIRFIAEGHPLSLKF	1.5	YES	Exposed
HLA-DPA1*01:03/ HLA-DPB1*04:01/ HLA-DPA1*01:03/ HLA-DPB1*02:01	A,C	93-107	EGHPLSLKFDSFAVI	0.59	YES	Exposed
HLA-DPA1*01:03/ HLA-DPB1*02:01/ HLA-DPA1*03:01/ HLA-DPB1*04:02/ HLA-DPA1*02:01/ HLA-DPB1*01:01/ HLA-DRB1*07:01/HLA-DRB1*09:01/ HLA-DPA1*01:03/ HLA-DPB1*04:01	A,C	96-110	PLSLKFDSFAVIMLC	0.31	YES	Exposed
HLA-DPA1*01:03/ HLA-DPB1*02:01/ HLA-DPA1*03:01/ HLA-DPB1*04:02/ HLA-DPA1*02:01/ HLA-DPB1*01:01/ HLA-DRB1*07:01/ HLA-DPA1*01:03/ HLA-DPB1*04:01	A,C	97-111	LSLKFDSFAVIMLCV	0.3	YES	Exposed
HLA-DRB1*07:01/HLA-DRB1*13:02	A,B	179-193	DDGTRRLVVSKNKPL	0.22	YES	Exposed
HLA-DRB1*13:02/ HLA-DRB1*07:01	A,B,C	180-194	DGTRRLVVSKNKPLV	0.3	YES	Exposed
HLA-DRB1*13:02/ HLA-DRB1*07:01/ HLA-DRB1*15:01	A,B,C	181-195	GTRRLVVSKNKPLVV	0.29	YES	Exposed
HLA-DRB1*13:02/ HLA-DRB1*07:01/ HLA-DRB1*15:01	A,B,C	183-197	RRLVVSKNKPLVVQF	0.29	YES	Exposed
HLA-DRB1*13:02/ HLA-DRB1*07:01/ HLA-DRB3*02:02	A,B,C	184-198	RLVVSKNKPLVVQFQ	0.3	YES	Exposed
HLA-DRB1*13:02/ HLA-DRB3*02:02	B,C	185-199	LVVSKNKPLVVQFQK	0.3	YES	Exposed

Note: A: Combination method of SMM method ANN method, and combinatorial library method; B: Combination method of SMM method, ANN method, and Sturniolo method; C: NetMHCIIpan method.

Table 3 List of putative epitopes of P04347 along with interacting HLA class II alleles and relevant scores.

HLA class II alleles	Method	Position	Sequence	IL-4pred	AllerTOP v. 2.0	Spatial position
HLA-DRB3*02:02	C	183-197	NPRVFYLAGNPDIEH	0.47	YES	Inner
HLA-DRB3*02:02	C	184-198	PRVFYLAGNPDIEHP	0.47	YES	Inner
HLA-DRB3*02:02	C	185-199	RVFYLAGNPDIEHPE	0.51	YES	Inner
HLA-DPA1*01:03/ HLA-DPB1*04:01	C	231-245	GFSKHFLAQSFNTNE	0.37	YES	Exposed
HLA-DPA1*01:03/ HLA-DPB1*04:01	C	232-246	FSKHFLAQSFNTNED	0.34	YES	Exposed
HLA-DQA1*05:01/ HLA-DQB1*03:01	A	260-274	QIVTVEGGLSVISPK	0.46	YES	Exposed
HLA-DQA1*05:01/ HLA-DQB1*03:01	A	261-275	IVTVEGGLSVISPKW	0.37	YES	Exposed
HLA-DRB1*15:01/ HLA-DRB1*11:01	B	390-404	GLSAQYVVLYRNGIY	0.32	YES	Inner
HLA-DRB1*15:01/ HLA-DRB1*11:01	B	391-405	LSAQYVVLYRNGIYS	0.32	YES	Inner
HLA-DRB1*15:01/ HLA-DRB1*11:01	B	392-406	SAQYVVLYRNGIYSP	0.32	YES	Inner
HLA-DRB1*15:01/ HLA-DRB1*11:01	B	393-407	AQYVVLYRNGIYSPD	0.31	YES	Inner
HLA-DRB1*04:01/ HLA-DRB3*02:02	B,C	461-475	GLEYYVFKTHHNAVS	1.22	YES	Inner
HLA-DRB3*02:02/ HLA-DRB1*04:01	B,C	462-476	LEYVVFKTHHNAVSS	1.23	YES	Inner
HLA-DRB1*04:01/ HLA-DRB3*02:02	B,C	464-478	YVVFKTHHNAVSSYI	1.24	YES	Inner
HLA-DRB3*02:02/ HLA-DRB1*04:01	B,C	465-479	VVFKTHHNAVSSYIK	1.26	YES	Inner
HLA-DRB3*01:01	A	471-485	HNAVSSYIKDVFRVI	0.29	YES	Exposed
HLA-DRB3*01:01	A	472-486	NAVSSYIKDVFRVIP	0.29	YES	Exposed
HLA-DRB3*01:01	A	473-487	AVSSYIKDVFRVIPS	0.29	YES	Exposed
HLA-DRB1*09:01/ HLA-DRB1*07:01/ HLA-DRB1*15:01/ HLA-DRB5*01:01/ HLA-DPA1*02:01/ HLA-DPB1*14:01/ HLA-DRB3*02:02/ HLA-DPA1*01:03/ HLA-DPB1*04:01/	A,B,C	477-491	YIKDVFRVIPSEVLS	0.29	YES	Exposed
HLA-DRB1*09:01/ HLA-DRB1*07:01/ HLA-DRB1*15:01/ HLA-DRB1*01:01/ HLA-DRB1*04:01/ HLA-DRB1*04:05/ HLA-DRB5*01:01/ HLA-DPA1*02:01/ HLA-DPB1*14:01 / HLA-DRB3*02:02	A,B,C	479-493	KDVFRVIPSEVLSNS	0.32	YES	Exposed
HLA-DRB1*09:01/ HLA-DRB1*07:01/ HLA-DRB1*15:01/ HLA-DRB1*04:01/ HLA-DRB5*01:01/HLA-DPA1*02:01/ HLA-DPB1*14:01/ HLA-DRB3*02:02	A,B,C	480-494	DVFRVIPSEVLSNSY	0.31	YES	Exposed
HLA-DPA1*02:01/ HLA-DPB1*14:01/ HLA-DRB3*02:02	C	481-495	VFRVIPSEVLSNSYN			

Note: A: Combination method of SMM method ANN method, and combinatorial library method; B: Combination method of SMM method, ANN method, and Sturniolo method; C: NetMHCIIpan method.

Table 4 List of putative epitopes of P04776 along with interacting HLA class II alleles and relevant scores.

HLA class II alleles	Method	Position	Sequence	IL-4pred	AllerTOP v. 2.0	Spatial position
HLA-DRB4*01:01	A	158-172	TPVVAVSIIDTNSLE	0.21	YES	Inner
HLA-DPA1*02:01/ HLA-DPB1*01:01/ HLA-DPA1*01:03/ HLA-DPB1*02:01/ HLA-DPA1*03:01/ HLA-DPB1*04:02/ HLA-DPA1*01:03/ HLA-DPB1*04:01	A,C	217-231	GGSILSGFTLEFLEH	1.46	YES	Exposed
HLA-DPA1*02:01/ HLA-DPB1*01:01/ HLA-DPA1*03:01/ HLA-DPB1*04:02/ HLA-DPA1*01:03/ HLA-DPB1*02:01/ HLA-DPA1*01:03/ HLA-DPB1*04:01	A,C	220-234	ILSGFTLEFLEHAFS	1.80	YES	Exposed
HLA-DPA1*02:01/ HLA-DPB1*01:01/ HLA-DPA1*03:01/ HLA-DPB1*04:02/ HLA-DPA1*01:03/ HLA-DPB1*02:01/ HLA-DPA1*01:03/ HLA-DPB1*04:01	A,C	221-235	LSGFTLEFLEHAFSV	1.83	YES	Exposed
HLA-DPA1*03:01/ HLA-DPB1*04:02/ HLA-DPA1*01:03/ HLA-DPB1*02:01/ HLA-DPA1*01:03/ HLA-DPB1*04:01	A,C	222-236	SGFTLEFLEHAFSVD	0.83	YES	Exposed
HLA-DRB3*02:02	C	319-333	MRLRHNIGQTSSPDI	0.68	YES	Exposed
HLA-DPA1*03:01/ HLA-DPB1*04:02/ HLA-DPA1*01:03/ HLA-DPB1*04:01	A,C	347-361	LDFPALSWLRLSAEF	0.25	YES	Inner
HLA-DPA1*03:01/ HLA-DPB1*04:02	A	348-362	DFPALSWLRLSAEFG	0.26	YES	Inner
HLA-DRB3*02:02	C	366-380	KNAMFVPHYNLNANS	0.26	YES	Inner
HLA-DRB1*04:01/ HLA-DRB1*13:02/ HLA-DRB1*07:01/ HLA-DRB3*02:02	A,B,C	368-382	AMFVPHYNLNANSII	0.27	YES	Inner
HLA-DRB1*13:02/ HLA-DRB3*01:01/ HLA-DRB1*07:01/ HLA-DRB3*02:02	A,B,C	369-383	MFVPHYNLNANSIIY	0.26	YES	Inner
HLA-DRB1*13:02/ HLA-DRB3*01:01/ HLA-DRB3*02:02	A,B,C	370-384	FVPHYNLNANSIIYA	0.26	YES	Inner
HLA-DQA1*01:02/ HLA-DQB1*06:02/ HLA-DRB1*13:02/ HLA-DRB3*01:01/ HLA-DRB3*02:02	A,B,C	371-385	VPHYNLNANSIIYAL	0.23	YES	Inner
HLA-DQA1*01:02/ HLA-DQB1*06:02/ HLA-DRB1*13:02/ HLA-DRB3*01:01/ HLA-DRB3*02:02	A C	372-386	PHYNLNANSIIYALN	0.24	YES	Inner
HLA-DQA1*01:02/ HLA-DQB1*06:02/ HLA-DRB1*13:02/ HLA-DRB3*02:02	A,B,C	373-387	HYNLNANSIIYALNG	0.25	YES	Inner
HLA-DRB1*11:01	B	412-426	LIVPQNFVVAARSQS	0.42	YES	Inner
HLA-DRB1*11:01	B	413-427	IVPQNFVVAARSQSD	0.45	YES	Inner
HLA-DRB1*11:01	B	414-428	VPQNFVVAARSQSDN	0.40	YES	Inner
HLA-DRB1*11:01	B	415-429	PQNFVVAARSQSDNF	0.38	YES	Inner
HLA-DRB1*11:01	B	416-430	QNFVVAARSQSDNFE	0.40	YES	Inner
HLA-DRB1*11:01	B	417-431	NFVVAARSQSDNFEY	1.40	YES	Inner
HLA-DPA1*01:03/ HLA-DPB1*02:01/ HLA-DPA1*02:01/ HLA-DPB1*01:01/ HLA-DPA1*01:03/ HLA-DPB1*04:01	A,C	423-437	RSQSDNFEYVSFKTN	1.3	YES	Inner
HLA-DPA1*01:03/ HLA-DPB1*02:01/ HLA-DPA1*02:01/ HLA-DPB1*01:01	A	424-438	SQSDNFEYVSFKTND	0.3	YES	Inner
HLA-DRB1*04:05/ HLA-DPA1*01:03/ HLA-DPB1*02:01/ HLA-DPA1*02:01/ HLA-DPB1*01:01	A,B	425-439	QSDNFEYVSFKTNDT	0.3	YES	Inner
HLA-DRB1*04:05/ HLA-DPA1*01:03/ HLA-DPB1*02:01/ HLA-DPA1*02:01/ HLA-DPB1*01:01	A,B	426-440	SDNFEYVSFKTNDTP	0.3	YES	Inner
HLA-DRB1*04:05/ HLA-DRB1*04:01/ HLA-DRB3*01:01/ HLA-DRB1*13:02/ HLA-DRB3*02:02	A,B,C	428-442	NFEYVSFKTNDTPMI	0.3	YES	Inner
HLA-DRB1*13:02	B	468-482	SQARQIKNNNPFKF	1.33	YES	Exposed
HLA-DRB1*13:02/ HLA-DRB3*01:01/ HLA-DRB3*02:02	A,B,C	470-484	QARQIKNNNPFKFLV	1.29	YES	Exposed
HLA-DRB1*13:02/ HLA-DRB3*01:01/ HLA-DRB3*02:02	A,B,C	471-485	ARQIKNNNPFKFLVP	1.29	YES	Exposed
HLA-DRB1*13:02/ HLA-DRB3*01:01/ HLA-DRB3*02:02	A,B,C	472-486	RQIKNNNPFKFLVPP	1.28	YES	Exposed

HLA-DRB1*13:02/ HLA-DRB3*02:02	B,C	473-487	QIKNNNPFKFLVPPQ	1.28	YES	Exposed
HLA-DRB1*04:05/ HLA-DRB1*13:02	B	474-488	IKNNNPFKFLVPPQE	1.28	YES	Exposed
HLA-DRB1*04:05	B	475-489	KNNNPFKFLVPPQES	0.28	YES	Exposed
HLA-DRB1*04:05	B	477-491	NNPFKFLVPPQESQK	0.28	YES	Exposed

Note: A: Combination method of SMM method ANN method, and combinatorial library method; B: Combination method of SMM method, ANN method, and Sturniolo method; C: NetMHCIIpan method.

Table 5 List of putative epitopes of P05046 along with interacting HLA class II alleles and relevant scores.

HLA class II alleles	Method	Position	Sequence	IL-4pred	AllerTOP v. 2.0	Spatial position
HLA-DRB3*02:02	C	23-37	LVLLTSKANS AETVS	0.97	YES	Exposed
HLA-DPA1*01:03/ HLA-DPB1*02:01/ HLA-DPA1*01:03/ HLA-DPB1*04:01	A,C	34-48	ETVSFSWNKFVPKQP	0.29	YES	Exposed
HLA-DRB3*02:02	C	38-52	FSWNKFVPKQPNMIL	0.29	YES	Exposed
HLA-DRB3*02:02	C	40-54	WNKFVPKQPNMILQG	0.29	YES	Exposed
HLA-DRB1*13:02	B	78-92	SSLGRALYSTPIHIW	0.24	YES	Exposed
HLA-DQA1*05:01/ HLA-DQB1*03:01	A	92-106	WDKETGSVASFAASF	0.42	YES	Exposed
HLA-DQA1*05:01/ HLA-DQB1*03:01	A	93-107	DKETGSVASFAASFN	0.44	YES	Exposed
HLA-DQA1*05:01/ HLA-DQB1*03:01	A	94-108	KETGSVASFAASFNF	0.52	YES	Exposed
HLA-DQA1*05:01/ HLA-DQB1*03:01	A	95-109	ETGSVASFAASFNFT	0.51	YES	Exposed
HLA-DPA1*01:03/ HLA-DPB1*02:01/ HLA-DRB1*04:05/ HLA-DPA1*01:03/ HLA-DPB1*04:01	A,B,C	100-114	ASFAASFNFTFYAPD	0.22	YES	Exposed
HLA-DRB1*13:02/ HLA-DRB3*02:02	B,C	168-182	NPHIGINVNSIRSIK	0.24	YES	Exposed
HLA-DRB1*13:02/ HLA-DRB3*02:02	B,C	169-183	PHIGINVNSIRSIKT	0.24	YES	Exposed
HLA-DRB1*13:02/ HLA-DRB3*02:02	B,C	170-184	HIGINVNSIRSIKTT	0.37	YES	Exposed
HLA-DRB1*13:02/ HLA-DRB3*02:02	B,C	171-185	IGINVNSIRSIKTTS	0.39	YES	Exposed
HLA-DRB1*04:01/ HLA-DRB1*08:02/ HLA-DRB1*04:05	B	172-186	GINVNSIRSIKTTSW	0.38	YES	Exposed
HLA-DRB1*04:01/ HLA-DRB1*04:05/ HLA-DRB1*08:02	B	173-187	INVNSIRSIKTTSWD	0.35	YES	Exposed
HLA-DRB1*04:01/ HLA-DRB1*04:05	B	174-188	NVNSIRSIKTTSWDL	0.36	YES	Exposed
HLA-DRB1*13:02/ HLA-DRB3*02:02	B,C	184-198	TSWDLANNKVAKVLI	0.23	YES	Exposed
HLA-DRB1*13:02/ HLA-DRB3*02:02	B,C	185-199	SWDLANNKVAKVLIT	0.22	YES	Exposed
HLA-DRB1*13:02/ HLA-DRB3*02:02	B,C	186-200	WDLANNKVAKVLITY	0.25	YES	Exposed
HLA-DRB1*07:01/ HLA-DRB1*09:01/ HLA-DQA1*05:01/ HLA-DQB1*03:01	A	234-248	EWVRIGFSAATGLDI	1.05	YES	Inner
HLA-DRB1*07:01/ HLA-DRB1*09:01	A	236-250	VRIGFSAATGLDIPG	0.98	YES	Inner
HLA-DRB1*09:01	A	253-267	HDVLSWSFASNLPHA	1.14	YES	Exposed
HLA-DRB1*09:01/ HLA-DRB3*02:02	A,C	254-268	DVLSWSFASNLPHAS	1.10	YES	Exposed
HLA-DRB1*09:01/ HLA-DRB3*02:02	A,C	255-269	VLSWSFASNLPHASS	1.08	YES	Exposed

Note: A: Combination method of SMM method ANN method, and combinatorial library method; B: Combination method of SMM method, ANN method, and Sturniolo method; C: NetMHCIIpan method.

Table 6 List of putative epitopes of P11827 along with interacting HLA class II alleles and relevant scores.

HLA class II alleles	Method	Position	Sequence	IL-4pred	AllerTOP v. 2.0	Spatial position
HLA-DPA1*03:01/ HLA-DPB1*04:02/ HLA-DRB1*11:01/ HLA-DPA1*01:03/ HLA-DPB1*04:01	A,B,C	211-225	PFHFNSKRFQTLFKN	1.34	YES	Exposed
HLA-DPA1*03:01/ HLA-DPB1*04:02/ HLA-DRB1*11:01/ HLA-DPA1*01:03/ HLA-DPB1*04:01	A,B,C	212-226	FHFNSKRFQTLFKNQ	1.34	YES	Exposed
HLA-DPA1*01:03/ HLA-DPB1*04:01	C	213-227	HFNSKRFQTLFKNQY	1.32	YES	Exposed
HLA-DRB5*01:01	B	221-235	TLFKNQYGHVRLQR	0.35	YES	Exposed
HLA-DRB5*01:01	B	222-236	LFKNQYGHVRLQRF	0.34	YES	Exposed
HLA-DRB5*01:01	B	223-237	FKNQYGHVRLQRFN	0.33	YES	Exposed
HLA-DRB5*01:01/ HLA-DRB1*11:01	B	225-239	NQYGHVRLQRFNKR	1.48	YES	Exposed
HLA-DRB1*11:01/ HLA-DRB5*01:01	B	226-240	QYGHVRLQRFNKRS	1.48	YES	Exposed
HLA-DRB1*04:01	B	248-262	DYRILEFNSKPNTLL	1.31	YES	Inner
HLA-DRB1*04:01/ HLA-DRB5*01:01/ HLA-DRB3*02:02	B,C	249-263	YRILEFNSKPNTLLL	1.23	YES	Inner
HLA-DRB3*01:01	A	262-276	LLPHHADADYLIVIL	0.22	YES	Inner
HLA-DRB1*04:05	B	275-289	ILNGTAILTLVNND	0.25	YES	Inner
HLA-DRB1*04:05	B	276-290	LNGTAILTLVNND	0.25	YES	Inner
HLA-DRB1*04:05	B	277-291	NGTAILTLVNND	0.26	YES	Inner
HLA-DQA1*05:01/ HLA-DQB1*03:01	A	520-534	DIFVIPAGYPVVVNA	0.26	YES	Inner
HLA-DQA1*05:01/ HLA-DQB1*03:01	A	521-535	IFVIPAGYPVVVNAT	0.28	YES	Inner
HLA-DRB4*01:01	A	558-572	SKDNVISQIPSQVQE	0.40	YES	Exposed

Note: A: Combination method of SMM method ANN method, and combinatorial library method; B: Combination method of SMM method, ANN method, and Sturniolo method; C: NetMHCIIpan method.

Table 7 List of putative epitopes of P25974 along with interacting HLA class II alleles and relevant scores.

HLA class II alleles	Method	Position	Sequence	IL-4pred	AllerTOP v. 2.0	Spatial position
HLA-DRB1*13:02/ HLA-DRB3*02:02	B,C	41-55	SFQTLFENQNGRIRL	0.43	YES	Exposed
HLA-DRB1*13:02/ HLA-DRB3*02:02	B,C	42-56	FQTLFENQNGRIRLL	0.22	YES	Exposed
HLA-DRB3*02:02	C	43-57	QTLFENQNGRIRLLQ	0.21	YES	Exposed
HLA-DRB1*15:01	B	71-85	DYRIVQFQSKPNTIL	0.34	YES	Inner
HLA-DQA1*01:02/ HLA-DQB1*06:02	A	201-215	EEQRQQEGVIVELSK	1.31	YES	Exposed
HLA-DQA1*01:02/ HLA-DQB1*06:02	A	202-216	EQRQQEGVIVELSKE	1.32	YES	Exposed
HLA-DRB1*11:01	B	212-226	ELSKEQIRQLSRRAK	0.21	YES	Exposed
HLA-DRB3*01:01	A	244-258	SRNPIYSNNFGKFFE	0.42	YES	Exposed
HLA-DRB3*01:01	A	245-259	RNPIYSNNFGKFFEI	1.37	YES	Exposed
HLA-DRB3*01:01	A	246-260	NPIYSNNFGKFFEIT	0.41	YES	Exposed
HLA-DRB1*13:02/ HLA-DRB1*04:05	B	271-285	DIFLSSVDINEGALL	0.87	YES	Inner
HLA-DRB1*13:02/ HLA-DQA1*01:02/ HLA-DQB1*06:02	A, B	275-289	SSVDINEGALLLPHF	0.27	YES	Inner
HLA-DQA1*01:02/ HLA-DQB1*06:02	A	276-290	SVDINEGALLLPHFN	0.27	YES	Inner
HLA-DQA1*01:02/ HLA-DQB1*06:02	A	277-291	VDINEGALLLPHFNS	0.27	YES	Inner
HLA-DRB3*01:01	A	335-349	SEDDVVFVIPAAYPFV	0.27	YES	Inner

Note: A: Combination method of SMM method ANN method, and combinatorial library method; B: Combination method of SMM method, ANN method, and Sturniolo method; C: NetMHCIIpan method.

Table 8 List of putative epitopes of P26987 along with interacting HLA class II alleles and relevant scores.

HLA class II alleles	Method	Position	Sequence	IL-4pred	AllerTOP v. 2.0	Spatial position
HLA-DRB3*01:01	A	19-33	LYKALVTDADNVIPK	1.09	YES	Exposed
HLA-DRB3*01:01/ HLA-DRB1*13:02	A,B	20-34	YKALVTDADNVIPKA	1.02	YES	Exposed
HLA-DPA1*02:01/ HLA-DPB1*01:01	A	49-63	GPGTIKKITFLEDGE	1.48	YES	Exposed
HLA-DPA1*02:01/ HLA-DPB1*01:01	A	50-64	PGTIKKITFLEDGET	1.40	YES	Exposed
HLA-DRB3*01:01/ HLA-DRB1*03:01	A,B	54-68	KKITFLEDGETKFVL	1.27	YES	Exposed
HLA-DRB3*01:01/ HLA-DRB1*03:01	A,B	55-69	KITFLEDGETKFVLH	1.27	YES	Exposed
HLA-DRB3*01:01/ HLA-DRB1*03:01	A,B	56-70	ITFLEDGETKFVLHK	1.31	YES	Exposed
HLA-DRB1*07:01	A	60-74	EDGETKFVLHKIESI	0.28	YES	Exposed
HLA-DQA1*05:01/ HLA-DQB1*03:01	A	106-120	AGPNGGSAGKLTVKY	0.21	YES	Exposed
HLA-DPA1*02:01/ HLA-DPB1*05:01/ HLA-DRB1*09:01/ HLA-DRB1*15:01/ HLA-DPA1*02:01/ HLA-DPB1*01:01/ HLA-DRB5*01:01/ HLA-DRB1*07:01/ HLA-DRB1*01:01/ HLA-DPA1*01:03/ HLA-DPB1*04:01	A,B,C	138-152	AKADALFKAIEAYLL	1.23	YES	Exposed
HLA-DPA1*02:01/ HLA-DPB1*05:01/ HLA-DRB1*01:01/ HLA-DRB1*09:01/ HLA-DRB1*15:01/ HLA-DPA1*02:01/ HLA-DPB1*01:01/ HLA-DRB5*01:01/ HLA-DRB1*04:01/ HLA-DRB1*07:01/ HLA-DPA1*02:01/ HLA-DPB1*14:01/ HLA-DPA1*01:03/ HLA-DPB1*04:01/ HLA-DRB3*02:02	A,B,C	140-154	ADALFKAIEAYLLAH	1.22	YES	Exposed
HLA-DPA1*02:01/ HLA-DPB1*05:01	A	141-155	DALFKAIEAYLLAHP	1.22	YES	Exposed

Note: A: Combination method of SMM method ANN method, and combinatorial library method; B: Combination method of SMM method, ANN method, and Sturniolo method; C: NetMHCIIpan method.

Table 9 Pepsin cleavage site of T cell epitopes.

Protein name	Position and sequence	Pepsin digestion sites	Anti-digestive peptides	Allergenicity of anti-digestive peptides
P01070	↓ ↓↓ 44-60 L SDITA F GGIRAAPTGN ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ 86-111 YRIRF IAEGHP LS L K F DS F AVIM L CV ↓ ↓ ↓ 179-199 DDGTRRLVVSKNKP LVVQ F QK	44, 49, 50, 90, 96, 98, 99, 100, 101, 103, 104, 108, 109, 192, 196, 197	DDGTRRLV VSKNKP	Yes
P04347	↓ ↓ ↓ 183-199 NPRV FY L AGNPDIEHPE ↓ ↓ ↓ ↓ ↓ ↓ 231-245 G F SKH FLAQS F NTNED ↓ ↓ 260-275 QIVTVEGG L SVISPKW ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ 390-407 G L SAQ Y VV L Y RNGI YSPD ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ 461-495 G L E Y VV F KTHHNAVSS Y IKDVF RVIPSEV L SNS Y N	186, 188, 189, 231, 232, 235, 240, 241, , 267, 268, 390, 391, 394, 395, 397, 398, 399, 403, 461, 462, 463, 464, 466, 467, 476, 477, 482, 489, 490, 493, 494	-	-
P04776	↓ ↓ 158-172 TPVVAVSIIDTNS L E ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ 217-236 GGSIL SG F T L E F L EHA FSVD ↓ 319-333 MRL RHNIGQTSSPDI ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ 347-362 L DF PAL S W L RL SAE F G ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ 366-387 KNAM FVPHY NL NANSII Y A L NG ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ 412-442 LIVPQN F VVAARSQSDN F E Y VS F KTNDTPMI ↓ ↓ ↓ 468-491 SQQARQIKNNNP FK F LVPPQESQK	170, 171, 220, 221, 223, 224, 225, 226, 227, 228, 229, 232, 321, 347, 349, 352, 353, 354, 355, 357, 360, 361, 369, 374, 376, 382, 383, 384, 385, 417, 418, 428, 429, 430, 431, 433, 434, 479, 481, 482	TPVVAVSII DTNS,	Yes
			RHNIGQTS SPDI,	No
			SQQARQIK NNNP	Yes
P05046	↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ 23-54 L V L L TSKANSAETVS F S W NK FVPKQPNMI L QG ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ 78-114 SS L GRA LY STPIHI WDKETGSVA S F AAS F N F T F ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ 168-200 NPHIGINVNSIRSIKTTS W D L ANNKVAKV LIT Y ↓ ↓ ↓ ↓ ↓ 234-250 EW VRIGF SAATG L DIPG ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ 253-269 HDVL S W S F ASNL PHASS	23, 24, 25, 26, 37, 38, 39, 40, 42, 51, 52, 79, 80, 83, 85, 91, 101, 102, 105, 106, 107, 108, 109, 110, 185, 186, 187, 188, 196, 199, 235, 240, 245, 246, 256, 257, 258, 259, 260, 264	NPHIGINVN SIRSIKTTS	Yes
P11827	↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ 211-240 PFH F NSKRFQT L F KNQY GHV RV LQRF NKRS ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ 248-291 D YRI LE F NSKPNT L LL PHHADAD Y L IVI L NGTAI L T L VNNDDRD ↓ ↓ ↓ ↓ ↓ 520-535 DI F VIPAGY PVVVNAT 558-572 SKDNVISQIPSQVQE	213, 214, 221, 222, 223, 227, 232, 236, 248, 251, 253, 254, 260, 261, 263, 270, 271, 272, 275, 276, 281, 282, 283, 284, 521, 522, 528	SKDNVISQI PSQVQE	Yes
P25974	↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ 41-57 S F QT L F ENQNGRIRL LQ ↓ ↓ ↓ ↓ ↓ 71-85 D YRIVQ F QSKPNTI L ↓ ↓ ↓ ↓ ↓ 201-226 EEQRQQEGVIVE L SKEQIRQ LSRRAK ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ 244-260 SRNPIY SNN F GK F FEIT ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ 271-291 DI F L SSVDINEGA L LL PHF NS ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ 335-349 SEDDV F VIPAAY P FV	41, 42, 44, 45, 46, 55, 71, 76, 77, 84, 212, 213, 220, 249, 252, 253, 255, 256, 272, 273, 274, 283, 284, 286, 289, 339, 340, 346, 347	EEQRQQEG VIVE	No
P26987	↓ ↓ ↓	19, 20, 22, 58, 59, 65,	LVTADANV	Yes

19-34 L Y KA LVTDADNVIPKA ↓ ↓ ↓ ↓ ↓ 49-74 GPGTIKKITF L EDGETK F VL HKIESI ↓ ↓ ↓ 106-120 AGPNGGSAGK L TVK Y ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ 138-155 AKADA L F KAIEA Y L L AHP	66, 68, 115, 116, 119, 142, 143, 144, 149, 150, 151, 152	IPKA	
--	--	------	--

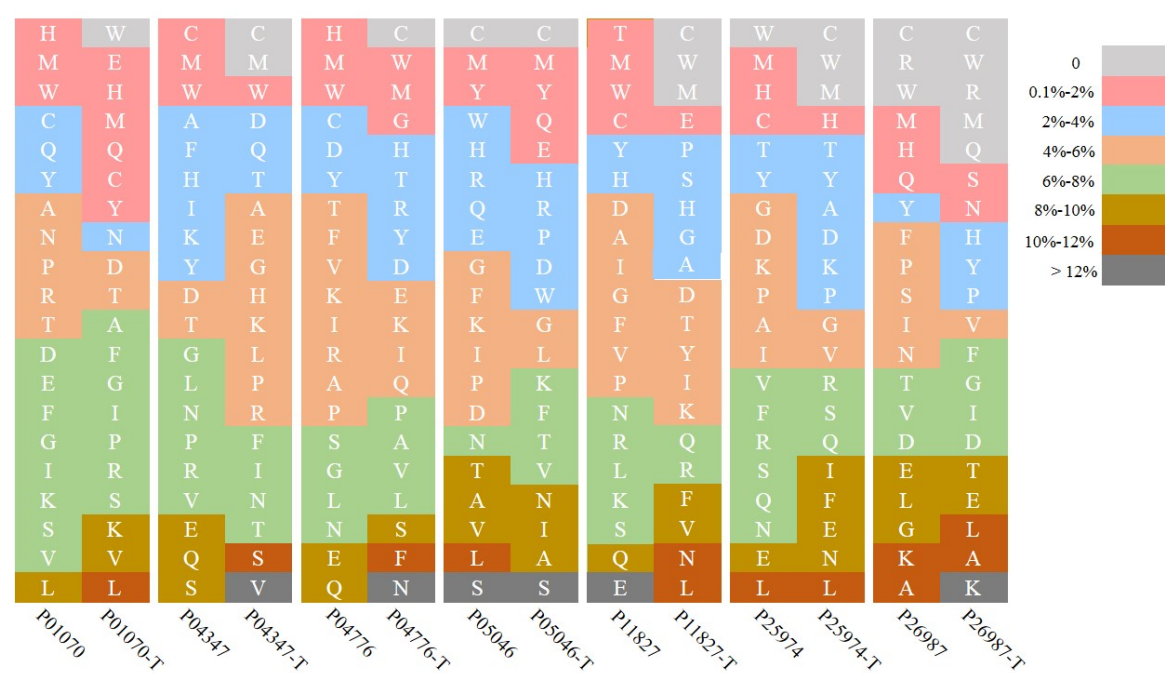
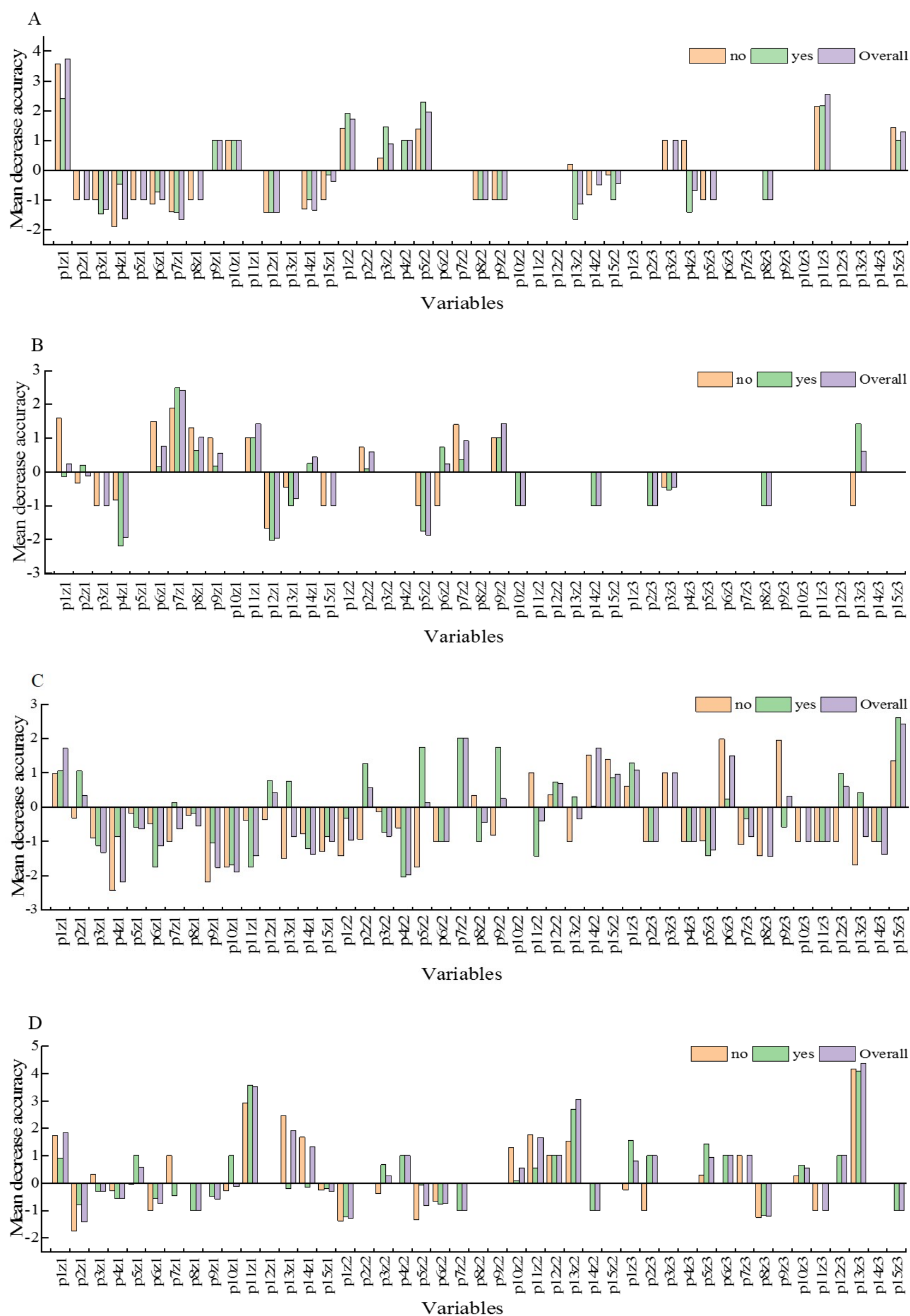


Fig. 1 Analysis of the amino acid composition of T cell epitopes of seven soybean allergens.

(P01070: Amino acid composition analysis of P01070 protein; P01070-T: Amino acid composition analysis of T cell epitopes of P01070 protein.
P04347: Amino acid composition analysis of P04347 protein; P04347-T: Amino acid composition analysis of T cell epitopes of P04347 protein.
P04776: Amino acid composition analysis of P04776 protein; P04776-T: Amino acid composition analysis of T cell epitopes of P04776 protein.
P05046: Amino acid composition analysis of P05046 protein; P05046-T: Amino acid composition analysis of T cell epitopes of P05046 protein.
P11827: Amino acid composition analysis of P11827 protein; P11827-T: Amino acid composition analysis of T cell epitopes of P11827 protein.
P25974: Amino acid composition analysis of P25974 protein; P25974-T: Amino acid composition analysis of T cell epitopes of P25974 protein.
P26987: Amino acid composition analysis of P26987 protein; P26987-T: Amino acid composition analysis of T cell epitopes of P26987 protein.)



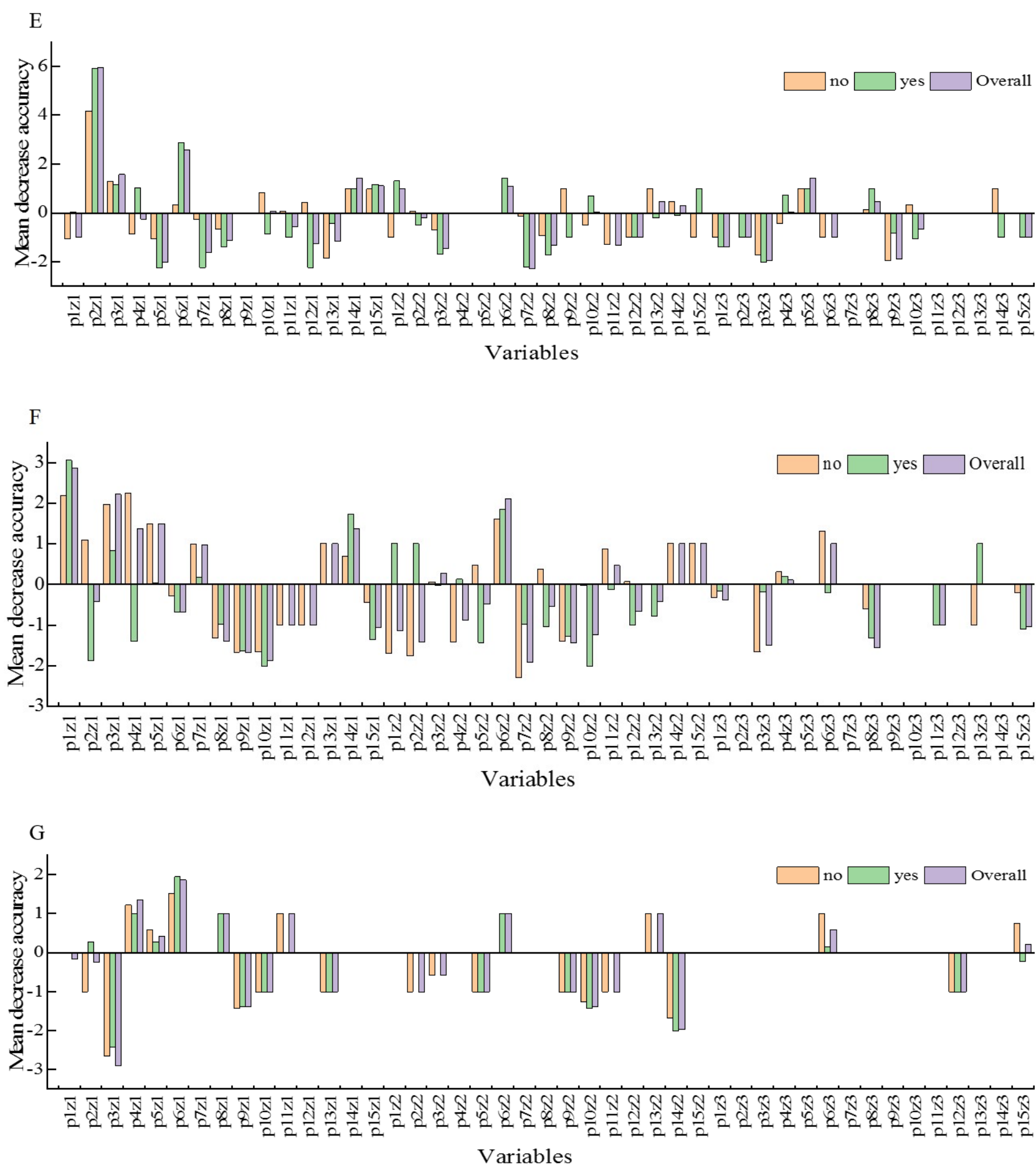


Fig. 2 Mean decrease accuracy of the 3-z scale models of the (A) P01070 T cell epitopes, (B) P04347 T cell epitopes, (C) P04776 T cell epitopes, (D) P05046 T cell epitopes, (E) P11827 T cell epitopes, (F) P25974 T cell epitopes, and (G) P26987 T cell epitopes.

Table S1. Descriptor scales for characterized amino acids

Abbreviation	Name	z_1	z_2	z_3
Ala	alanine	0.07	-1.73	0.09
Arg	arginine	2.88	2.52	-3.44
Asn	asparagine	3.22	1.45	0.84
Asp	aspartic acid	3.64	1.13	2.36
Cys	cysteine	0.71	-0.97	4.13
Gln	glutamine	2.18	0.53	-1.14
Glu	glutamic acid	3.08	0.39	-0.07
Gly	glycine	2.23	-5.36	0.30
His	histidine	2.41	1.74	1.11
Ile	isoleucine	-4.44	-1.68	-1.03
Leu	leucine	-4.19	-1.03	-0.98
Lys	lysine	2.84	1.41	-3.14
Met	methionine	-2.49	-0.27	-0.41
Phe	phenylalanine	-4.92	1.30	0.45
Pro	phenylalanine	-1.22	0.88	2.23
Ser	serine	1.96	-1.63	0.57
Thr	threonine	0.92	-2.09	-1.40
Trp	tryptophan	-4.75	3.65	0.85
Tyr	tyrosine	-1.39	2.32	0.01
Val	valine	-2.69	-2.53	-1.29

Reference:

S. Hellberg, M. Sjoestroem, B. Skagerberg, S. Wold, Peptide quantitative structure-activity relationships, a multivariate approach, Journal of Medicinal Chemistry, 30 (1987) 1126-1135. doi: 10.1021/jm00390a003.

Table S2. List of epitopes of P01070 predicted by consensus method along with interacting HLA class II alleles and relevant scores.

Allele	Star t	End	Method	Peptide	Percentile rank	Sm m IC50	Nn IC50	IL-4 pred	AllerTOP v. 2.0
HLA-DPA1*01:03/DPB1*02:01	1	15	Consensus (comb.lib./smm/nn)	MKSTIFFLFLFCAFT	0.32	18	34.5	0.05	
HLA-DPA1*01:03/DPB1*02:01	2	16	Consensus (comb.lib./smm/nn)	KSTIFFLFLFCAFTT	0.32	18	36.2	0.04	
HLA-DPA1*01:03/DPB1*02:01	3	17	Consensus (comb.lib./smm/nn)	STIFFLFLFCAFTTS	0.3	17	36	-0.12	
HLA-DPA1*01:03/DPB1*02:01	4	18	Consensus (comb.lib./smm/nn)	TIFFLFLFCAFTTSY	1.5	19	43.5	0.04	
HLA-DPA1*01:03/DPB1*02:01	5	19	Consensus (comb.lib./smm/nn)	IFFLFLFCAFTTSYL	1.1	37	33.1	0.04	
HLA-DRB1*15:01	5	19	Consensus (smm/nn/sturniolo)	IFFLFLFCAFTTSYL	2.3	94	181. 5	0.04	
HLA-DRB1*07:01	5	19	Consensus (comb.lib./smm/nn)	IFFLFLFCAFTTSYL	2.6	46	22.3	0.04	
HLA-DRB1*07:01	6	20	Consensus (comb.lib./smm/nn)	FFLFLFCAFTTSYLP	0.73	48	10	0.04	
HLA-DPA1*01:03/DPB1*02:01	6	20	Consensus (comb.lib./smm/nn)	FFLFLFCAFTTSYLP	1.1	40	34.1	0.04	
HLA-DPA1*02:01/DPB1*01:01	6	20	Consensus (comb.lib./smm/nn)	FFLFLFCAFTTSYLP	2.6	89	126. 2	0.04	
HLA-DRB1*15:01	6	20	Consensus (smm/nn/sturniolo)	FFLFLFCAFTTSYLP	3.2	118	119. 1	0.04	
HLA-DRB1*07:01	7	21	Consensus (comb.lib./smm/nn)	FLFLFCAFTTSYLPS	0.79	48	10.6	0.06	
HLA-DPA1*01:03/DPB1*02:01	7	21	Consensus (comb.lib./smm/nn)	FLFLFCAFTTSYLPS	1.6	65	33.1	0.06	
HLA-DRB1*04:01	7	21	Consensus (smm/nn/sturniolo)	FLFLFCAFTTSYLPS	3.3	131	82.1	0.06	
HLA-DRB1*07:01	8	22	Consensus (comb.lib./smm/nn)	LFLFCAFTTSYLP	0.99	48	11.9	0.09	
HLA-DRB1*04:01	8	22	Consensus (smm/nn/sturniolo)	LFLFCAFTTSYLP	2.6	134	44.1	0.09	
HLA-DPA1*01:03/DPB1*02:01	8	22	Consensus (comb.lib./smm/nn)	LFLFCAFTTSYLP	3.3	128	29.4	0.09	
HLA-DRB1*07:01	9	23	Consensus (comb.lib./smm/nn)	FLFCAFTTSYLPSAI	0.94	52	11.6	0.14	
HLA-DPA1*01:03/DPB1*02:01	9	23	Consensus (comb.lib./smm/nn)	FLFCAFTTSYLPSAI	3.3	128	34.4	0.14	
HLA-DRB1*09:01	9	23	Consensus (comb.lib./smm/nn)	FLFCAFTTSYLPSAI	3.7	84	43.4	0.14	
HLA-DRB1*04:01	9	23	Consensus (smm/nn/sturniolo)	FLFCAFTTSYLPSAI	3.8	164	36.3	0.14	
HLA-DRB1*07:01	10	24	Consensus (comb.lib./smm/nn)	LFCAFTTSYLPSAIA	1.3	147	13.7	0.16	
HLA-DRB1*09:01	10	24	Consensus (comb.lib./smm/nn)	LFCAFTTSYLPSAIA	2.6	144	34.9	0.16	
HLA-DRB1*07:01	11	25	Consensus (comb.lib./smm/nn)	FCAFTTSYLPSAIAD	2	168	18.5	0.17	
HLA-DRB1*09:01	11	25	Consensus (comb.lib./smm/nn)	FCAFTTSYLPSAIAD	3.4	145	41.4	0.17	
HLA-DRB1*09:01	12	26	Consensus (comb.lib./smm/nn)	CAFTTSYLPSAIADF	2.9	175	37.2	0.33/YES	
HLA-DRB1*13:02	22	36	Consensus (smm/nn/sturniolo)	AIADFLVDNEGNPLE	3.8	30	28.2	0.43/YES	
HLA-DRB1*13:02	23	37	Consensus (smm/nn/sturniolo)	IADFLVDNEGNPLEN	3.7	30	26.9	0.42/YES	
HLA-DRB3*01:01	36	50	Consensus (comb.lib./smm/nn)	ENGTTYILSDITAF	1.3	170	30	-0.08	YES
HLA-DRB3*01:01	37	51	Consensus (comb.lib./smm/nn)	NGGTTYILSDITAFG	0.85	174	20.9	-0.09	YES
HLA-DRB1*04:01	37	51	Consensus (smm/nn/sturniolo)	NGGTTYILSDITAFG	1.2	53	41.4	-0.09	YES
HLA-DRB1*03:01	37	51	Consensus (smm/nn/sturniolo)	NGGTTYILSDITAFG	3.2	226	101. 4	-0.09	YES
HLA-DRB1*04:01	38	52	Consensus (smm/nn/sturniolo)	GGTTYILSDITAFGG	0.47	52	24.8	-0.10	YES
HLA-DRB3*01:01	38	52	Consensus (comb.lib./smm/nn)	GGTTYILSDITAFGG	1.1	178	26.3	-0.10	YES
HLA-DRB1*03:01	38	52	Consensus (smm/nn/sturniolo)	GGTTYILSDITAFGG	3	238	63	-0.10	YES
HLA-DRB1*04:01	39	53	Consensus (smm/nn/sturniolo)	GTTYILSDITAFGGI	0.36	51	22	0.85/YES	
HLA-DRB3*01:01	39	53	Consensus (comb.lib./smm/nn)	GTTYILSDITAFGGI	0.94	173	23	0.85/YES	
HLA-DRB1*03:01	39	53	Consensus (smm/nn/sturniolo)	GTTYILSDITAFGGI	3	227	44.7	0.85/YES	
HLA-DRB1*04:01	40	54	Consensus (smm/nn/sturniolo)	TTYILSDITAFGGIR	0.36	57	21.6	0.92/YES	
HLA-DRB3*01:01	40	54	Consensus (comb.lib./smm/nn)	TTYILSDITAFGGIR	1.5	184	33.3	0.92/YES	
HLA-DRB1*03:01	40	54	Consensus (smm/nn/sturniolo)	TTYILSDITAFGGIR	3	223	40.3	0.92/YES	
HLA-DRB1*04:01	41	55	Consensus (smm/nn/sturniolo)	YYILSDITAFGGIRA	0.65	59	29.6	0.86/YES	
HLA-DRB1*03:01	41	55	Consensus (smm/nn/sturniolo)	YYILSDITAFGGIRA	3	236	57.8	0.86/YES	
HLA-DRB1*01:01	44	58	Consensus (comb.lib./smm/nn)	LSDITAFGGIRAAPT	1.8	13	7.7	0.89/YES	YES
HLA-DQA1*05:01/DQB1*03:01	44	58	Consensus (comb.lib./smm/nn)	LSDITAFGGIRAAPT	2.3	14	42.8	0.89/YES	YES
HLA-DRB1*01:01	45	59	Consensus (comb.lib./smm/nn)	SDITAFGGIRAAPTG	1.8	13	4.7	0.95/YES	YES
HLA-DQA1*05:01/DQB1*03:01	45	59	Consensus (comb.lib./smm/nn)	SDITAFGGIRAAPTG	2.1	14	40.5	0.95/YES	YES
HLA-DRB1*01:01	46	60	Consensus (comb.lib./smm/nn)	DITAFGGIRAAPTGN	1.8	13	4.1	0.84/YES	YES
HLA-DQA1*05:01/DQB1*03:01	46	60	Consensus (comb.lib./smm/nn)	DITAFGGIRAAPTGN	2.1	14	40.4	0.84/YES	YES
HLA-DQA1*05:01/DQB1*03:01	47	61	Consensus (comb.lib./smm/nn)	ITAFGGIRAAPTGNE	1.7	16	37.2	-0.02	
HLA-DRB1*01:01	47	61	Consensus (comb.lib./smm/nn)	ITAFGGIRAAPTGNE	1.8	13	4.3	-0.02	
HLA-DRB1*01:01	48	62	Consensus (comb.lib./smm/nn)	TAFGGIRAAPTGNER	2	14	5.2	0.02	YES
HLA-DQA1*05:01/DQB1*03:01	48	62	Consensus (comb.lib./smm/nn)	TAFGGIRAAPTGNER	2.3	23	43.8	0.02	YES
HLA-DPA1*01:03/DPB1*02:01	93	107	Consensus (comb.lib./smm/nn)	EGHPLSLKFDSFAVI	1.3	48	38.6	0.59/YES	YES
HLA-DPA1*01:03/DPB1*02:01	94	108	Consensus (comb.lib./smm/nn)	GHPLSLKFDSFAVIM	1.1	47	33.8	0.51/YES	

HLA-DPA1*01:03/DPB1*02:01	95	109	Consensus (comb.lib./smm/nn)	HPLSLKFDSFAVIML	0.95	45	27.5	0.4/YES	
HLA-DPA1*03:01/DPB1*04:02	95	109	Consensus (comb.lib./smm/nn)	HPLSLKFDSFAVIML	2.3	106	67.3	0.4/YES	
HLA-DPA1*02:01/DPB1*01:01	95	109	Consensus (comb.lib./smm/nn)	HPLSLKFDSFAVIML	2.4	105	106.2	0.4/YES	
HLA-DRB1*07:01	95	109	Consensus (comb.lib./smm/nn)	HPLSLKFDSFAVIML	3.4	50	73.4	0.4/YES	
HLA-DRB1*09:01	95	109	Consensus (comb.lib./smm/nn)	HPLSLKFDSFAVIML	3.4	134	93.3	0.4/YES	
HLA-DPA1*01:03/DPB1*02:01	96	110	Consensus (comb.lib./smm/nn)	PLSLKFDSFAVIMLC	0.97	46	28.9	0.31/YES	YES
HLA-DPA1*03:01/DPB1*04:02	96	110	Consensus (comb.lib./smm/nn)	PLSLKFDSFAVIMLC	2.3	107	63.9	0.31/YES	YES
HLA-DPA1*02:01/DPB1*01:01	96	110	Consensus (comb.lib./smm/nn)	PLSLKFDSFAVIMLC	2.4	108	102.1	0.31/YES	YES
HLA-DRB1*07:01	96	110	Consensus (comb.lib./smm/nn)	PLSLKFDSFAVIMLC	3.3	49	95	0.31/YES	YES
HLA-DRB1*09:01	96	110	Consensus (comb.lib./smm/nn)	PLSLKFDSFAVIMLC	3.5	136	121.4	0.31/YES	YES
HLA-DPA1*01:03/DPB1*02:01	97	111	Consensus (comb.lib./smm/nn)	LSLKFDSDFAVIMLCV	0.92	44	29.8	0.3/YES	YES
HLA-DPA1*03:01/DPB1*04:02	97	111	Consensus (comb.lib./smm/nn)	LSLKFDSDFAVIMLCV	2.3	102	63.7	0.3/YES	YES
HLA-DPA1*02:01/DPB1*01:01	97	111	Consensus (comb.lib./smm/nn)	LSLKFDSDFAVIMLCV	2.4	93	103.3	0.3/YES	YES
HLA-DRB1*07:01	97	111	Consensus (comb.lib./smm/nn)	LSLKFDSDFAVIMLCV	3.2	47	89.7	0.3/YES	YES
HLA-DPA1*01:03/DPB1*02:01	98	112	Consensus (comb.lib./smm/nn)	SLKFDSFAVIMLCVG	3	116	40.5	0.3/YES	
HLA-DPA1*03:01/DPB1*04:02	98	112	Consensus (comb.lib./smm/nn)	SLKFDSFAVIMLCVG	3.2	139	104.9	0.3/YES	
HLA-DRB1*07:01	98	112	Consensus (comb.lib./smm/nn)	SLKFDSFAVIMLCVG	3.3	49	141	0.3/YES	
HLA-DPA1*02:01/DPB1*01:01	98	112	Consensus (comb.lib./smm/nn)	SLKFDSFAVIMLCVG	3.8	133	175.5	0.3/YES	
HLA-DRB1*07:01	179	193	Consensus (comb.lib./smm/nn)	DDGTRRLVVSKNKPL	1.2	16	13.5	0.22/YES	YES
HLA-DRB1*13:02	179	193	Consensus (smm/nn/sturniolo)	DDGTRRLVVSKNKPL	2.1	63	15.5	0.22/YES	YES
HLA-DRB1*13:02	180	194	Consensus (smm/nn/sturniolo)	DGTRRLVVSKNKPLV	0.97	60	8.1	0.3/YES	YES
HLA-DRB1*07:01	180	194	Consensus (comb.lib./smm/nn)	DGTRRLVVSKNKPLV	1.1	15	12.3	0.3/YES	YES
HLA-DRB1*13:02	181	195	Consensus (smm/nn/sturniolo)	GTRRLVVSKNKPLVV	0.14	17	3.3	0.29/YES	YES
HLA-DRB1*07:01	181	195	Consensus (comb.lib./smm/nn)	GTRRLVVSKNKPLVV	0.66	14	9.4	0.29/YES	YES
HLA-DRB1*15:01	181	195	Consensus (smm/nn/sturniolo)	GTRRLVVSKNKPLVV	3.7	134	41.7	0.29/YES	YES
HLA-DRB1*13:02	182	196	Consensus (smm/nn/sturniolo)	TRRLVVSKNKPLVVQ	0.1	17	3.1	0.29/YES	
HLA-DRB1*07:01	182	196	Consensus (comb.lib./smm/nn)	TRRLVVSKNKPLVVQ	0.88	14	11.2	0.29/YES	
HLA-DRB1*15:01	182	196	Consensus (smm/nn/sturniolo)	TRRLVVSKNKPLVVQ	3.7	135	33.1	0.29/YES	
HLA-DRB1*13:02	183	197	Consensus (smm/nn/sturniolo)	RRLVVSKNKPLVVQF	0.1	17	3	0.29/YES	YES
HLA-DRB1*07:01	183	197	Consensus (comb.lib./smm/nn)	RRLVVSKNKPLVVQF	1.1	14	12.4	0.29/YES	YES
HLA-DRB1*15:01	183	197	Consensus (smm/nn/sturniolo)	RRLVVSKNKPLVVQF	3.8	137	35.2	0.29/YES	YES
HLA-DRB1*13:02	184	198	Consensus (smm/nn/sturniolo)	RLVVSKNKPLVVQFQ	0.14	19	3.3	0.3/YES	YES
HLA-DRB1*07:01	184	198	Consensus (comb.lib./smm/nn)	RLVVSKNKPLVVQFQ	2.8	41	18.1	0.3/YES	YES
HLA-DRB1*13:02	185	199	Consensus (smm/nn/sturniolo)	LVVSKNKPLVVQFQK	0.33	19	4.9	0.3/YES	YES
HLA-DRB1*13:02	186	200	Consensus (smm/nn/sturniolo)	VVSKNKPLVVQFQKL	1.4	59	10.9	0.13	YES
HLA-DRB1*11:01	192	206	Consensus (smm/nn/sturniolo)	PLVVQFQKLDKESLA	3.7	75	183.2	-0.01	
HLA-DRB1*11:01	193	207	Consensus (smm/nn/sturniolo)	LVVQFQKLDKESLAK	3.9	78	118	-0.11	
HLA-DRB1*11:01	194	208	Consensus (smm/nn/sturniolo)	VVQFQKLDKESLAKK	3.7	75	91.2	-0.20	YES

Table S3. List of epitopes of P01070 predicted by NetMHCIIpan method along with interacting HLA class II alleles and relevant scores.

Allele	Star t	End	Method	Peptide	Percentile rank	NetMHCIIpan IC50	IL-4pred	AllerTO P v. 2.0
HLA-DPA1*01:03/DPB1*04:01	5	19	NetMHCIIpan	IFFLFLCAFTTSYL	3.3	183.5	0.04	
HLA-DPA1*01:03/DPB1*04:01	6	20	NetMHCIIpan	FFLFLCAFTTSYLP	2.7	156.98	0.04	
HLA-DPA1*01:03/DPB1*04:01	7	21	NetMHCIIpan	FLFLCAFTTSYLP	1.6	116.23	0.06	
HLA-DPA1*01:03/DPB1*04:01	8	22	NetMHCIIpan	LFLCAFTTSYLPSA	0.81	82.02	0.09	
HLA-DPA1*01:03/DPB1*04:01	9	23	NetMHCIIpan	FLCAFTTSYLPSAI	0.63	73.76	0.14	
HLA-DPA1*01:03/DPB1*04:01	10	24	NetMHCIIpan	LCAFTTSYLPSAIA	0.95	91.02	0.16	
HLA-DPA1*01:03/DPB1*04:01	11	25	NetMHCIIpan	FCAFTTSYLPSAIAD	2.1	133.53	0.17	
HLA-DRB3*02:02	38	52	NetMHCIIpan	GGTYIILSDITAFGG	4	232.01	-0.10	YES
HLA-DRB3*02:02	39	53	NetMHCIIpan	GTYIILSDITAFGGI	2.9	164.99	0.85/ YES	
HLA-DRB3*02:02	40	54	NetMHCIIpan	TYIILSDITAFGGIR	3.1	176.58	0.92/ YES	
HLA-DRB3*02:02	86	100	NetMHCIIpan	YRIRFIAEGHPLSLK	3.4	200.18	1.33/	YES

							YES	
HLA-DRB3*02:02	87	101	NetMHCIIpan	RIRFIAEGHPLSLKF	3.3	191.58	1.5/YES	YES
HLA-DPA1*01:03/DPB1*04:01	93	107	NetMHCIIpan	EGHPLSLKFDSFAVI	2.3	139.57	0.59/ YES	
HLA-DPA1*01:03/DPB1*04:01	94	108	NetMHCIIpan	GHPLSLKFDSFAVIM	0.99	93.9	0.51/ YES	
HLA-DPA1*01:03/DPB1*04:01	95	109	NetMHCIIpan	HPLSLKFDSFAVIML	0.41	59.25	0.4/YES	
HLA-DPA1*01:03/DPB1*04:01	96	110	NetMHCIIpan	PLSLKFDSFAVIMLC	0.56	67.28	0.31/ YES	YES
HLA-DPA1*01:03/DPB1*04:01	97	111	NetMHCIIpan	LSLKFDSDFAVIMLCV	0.61	72.5	0.3/YES	YES
HLA-DPA1*01:03/DPB1*04:01	98	112	NetMHCIIpan	SLKFDSFAVIMLCVG	1.3	105.88	0.3/YES	
HLA-DPA1*01:03/DPB1*04:01	99	113	NetMHCIIpan	LKFDSFAVIMLCVGI	3.2	179.09	0.3/YES	
HLA-DRB3*02:02	180	194	NetMHCIIpan	DGTRRLVVSKNKPL V	1.7	97	0.30/ YES	YES
HLA-DRB3*02:02	181	195	NetMHCIIpan	GTRRLVVSKNKPLV V	0.72	44.5	0.29/ YES	YES
HLA-DRB3*02:02	182	196	NetMHCIIpan	TRRLVVSKNKPLVV Q	0.49	36.08	0.29/ YES	
HLA-DRB3*02:02	183	197	NetMHCIIpan	RRLVVSKNKPLVVQ F	0.42	33.43	0.29/ YES	YES
HLA-DRB3*02:02	184	198	NetMHCIIpan	RLVVSKNKPLVVQF Q	0.8	48.68	0.3/YES	YES
HLA-DRB3*02:02	185	199	NetMHCIIpan	LVVSKNKPLVVQFQ K	1.9	105.19	0.3/YES	YES

Table S4. List of epitopes of P04347 predicted by consensus method along with interacting HLA class II alleles and relevant scores.

Allele	Start	End	Method	Peptide	Percentile rank	Sm m IC50	Nn IC50	IL-4 pred	AllerTOP v. 2.0
HLA-DPA1*02:01/DPB1*01:01	1	15	Consensus (comb.lib./smm/nn)	MGKPFFTLSSLCL	0.42	25	35.5	-0.16	
HLA-DPA1*03:01/DPB1*04:02	1	15	Consensus (comb.lib./smm/nn)	MGKPFFTLSSLCL	0.81	46	24.5	-0.16	
HLA-DRB1*04:05	1	15	Consensus (smm/nn/sturniolo)	MGKPFFTLSSLCL	1.7	71	47.3	-0.16	
HLA-DRB1*01:01	1	15	Consensus (comb.lib./smm/nn)	MGKPFFTLSSLCL	2	14	13.1	-0.16	
HLA-DRB1*07:01	1	15	Consensus (comb.lib./smm/nn)	MGKPFFTLSSLCL	3.6	54	15.2	-0.16	
HLA-DPA1*01:03/DPB1*02:01	1	15	Consensus (comb.lib./smm/nn)	MGKPFFTLSSLCL	4	153	47.8	-0.16	
HLA-DPA1*02:01/DPB1*01:01	2	16	Consensus (comb.lib./smm/nn)	GKPFFTLSSLCLL	0.35	22	36.8	-1.43	
HLA-DPA1*03:01/DPB1*04:02	2	16	Consensus (comb.lib./smm/nn)	GKPFFTLSSLCLL	0.58	35	18.2	-1.43	
HLA-DRB1*01:01	2	16	Consensus (comb.lib./smm/nn)	GKPFFTLSSLCLL	1.6	12	9.6	-1.43	
HLA-DRB1*04:05	2	16	Consensus (smm/nn/sturniolo)	GKPFFTLSSLCLL	1.6	67	45.3	-1.43	
HLA-DPA1*01:03/DPB1*02:01	2	16	Consensus (comb.lib./smm/nn)	GKPFFTLSSLCLL	2.1	88	33.4	-1.43	
HLA-DRB1*07:01	2	16	Consensus (comb.lib./smm/nn)	GKPFFTLSSLCLL	3.3	49	13.2	-1.43	
HLA-DPA1*03:01/DPB1*04:02	3	17	Consensus (comb.lib./smm/nn)	KPFFTLSSLCLLL	0.54	34	17.7	-1.2	
HLA-DPA1*02:01/DPB1*01:01	3	17	Consensus (comb.lib./smm/nn)	KPFFTLSSLCLLL	0.56	21	53.8	-1.2	
HLA-DRB1*01:01	3	17	Consensus (comb.lib./smm/nn)	KPFFTLSSLCLLL	1.6	12	8.9	-1.2	
HLA-DRB1*04:05	3	17	Consensus (smm/nn/sturniolo)	KPFFTLSSLCLLL	1.7	63	46.8	-1.2	
HLA-DPA1*01:03/DPB1*02:01	3	17	Consensus (comb.lib./smm/nn)	KPFFTLSSLCLLL	2.1	89	33.4	-1.2	
HLA-DRB1*07:01	3	17	Consensus (comb.lib./smm/nn)	KPFFTLSSLCLLL	3.1	46	12.7	-1.2	
HLA-DPA1*03:01/DPB1*04:02	4	18	Consensus (comb.lib./smm/nn)	PFFTLSSLCLLLL	0.36	24	21.1	-1.02	
HLA-DPA1*02:01/DPB1*01:01	4	18	Consensus (comb.lib./smm/nn)	PFFTLSSLCLLLL	1.5	39	88.4	-1.02	
HLA-DRB1*01:01	4	18	Consensus (comb.lib./smm/nn)	PFFTLSSLCLLLL	2.6	18	11.3	-1.02	
HLA-DPA1*01:03/DPB1*02:01	4	18	Consensus (comb.lib./smm/nn)	PFFTLSSLCLLLL	2.8	111	38.2	-1.02	
HLA-DRB1*07:01	4	18	Consensus (comb.lib./smm/nn)	PFFTLSSLCLLLL	3	44	15.9	-1.02	
HLA-DRB1*04:05	4	18	Consensus (smm/nn/sturniolo)	PFFTLSSLCLLLL	3.5	50	74.5	-1.02	
HLA-DPA1*03:01/DPB1*04:02	5	19	Consensus (comb.lib./smm/nn)	FFTLSSLCLLLLS	0.52	24	27.8	-1.02	
HLA-DPA1*01:03/DPB1*02:01	5	19	Consensus (comb.lib./smm/nn)	FFTLSSLCLLLLS	2.8	111	54.2	-1.02	
HLA-DRB1*01:01	5	19	Consensus (comb.lib./smm/nn)	FFTLSSLCLLLLS	3.3	21	17	-1.02	
HLA-DRB1*07:01	5	19	Consensus (comb.lib./smm/nn)	FFTLSSLCLLLLS	3.6	75	18.6	-1.02	
HLA-DPA1*02:01/DPB1*01:01	5	19	Consensus (comb.lib./smm/nn)	FFTLSSLCLLLLS	3.7	39	159. 8	-1.02	
HLA-DPA1*03:01/DPB1*04:02	6	20	Consensus (comb.lib./smm/nn)	FTLSSLCLLLSS	2.2	25	58.1	-1.02	YES
HLA-DPA1*03:01/DPB1*04:02	7	21	Consensus (comb.lib./smm/nn)	TLSSLCLLLSSA	3.4	28	80.7	-0.91	YES
HLA-DQA1*05:01/DQB1*03:01	93	107	Consensus (comb.lib./smm/nn)	IIVVQGKGAIGFAP	4	45	43.9	0.22/ YES	
HLA-DQA1*05:01/DQB1*03:01	95	109	Consensus (comb.lib./smm/nn)	VVQGKGAIGFAFPGC	3.6	62	40.6	0.17	
HLA-DQA1*05:01/DQB1*03:01	96	110	Consensus (comb.lib./smm/nn)	VQGKGAIGFAFPGCP	3.6	73	40.7	0.19	
HLA-DQA1*05:01/DQB1*03:01	97	111	Consensus (comb.lib./smm/nn)	QGKGAIGFAFPGCPE	3.6	110	44.4	0.18	
HLA-DQA1*05:01/DQB1*03:01	98	112	Consensus (comb.lib./smm/nn)	GKGAIGFAFPGCPET	3.6	132	48.9	0.29/ YES	
HLA-DQA1*05:01/DQB1*03:01	260	274	Consensus (comb.lib./smm/nn)	QIVTVEGGLSVISPK	3.6	27	59.3	0.46/ YES	YES
HLA-DQA1*05:01/DQB1*03:01	261	275	Consensus (comb.lib./smm/nn)	IVTVEGGLSVISPKW	3.7	30	60.8	0.37/ YES	YES
HLA-DRB1*04:01	372	386	Consensus (smm/nn/sturniolo)	AGRISTLNSLTLPAL	3.4	70	84.3	-0.42	YES
HLA-DRB1*04:01	373	387	Consensus (smm/nn/sturniolo)	GRISTLNSLTLPALR	2.9	73	73.9	-0.73	
HLA-DRB1*15:01	390	404	Consensus (smm/nn/sturniolo)	GLSAQYVVLYRNGIY	1.6	70	27.6	0.32/ YES	YES
HLA-DRB1*11:01	390	404	Consensus (smm/nn/sturniolo)	GLSAQYVVLYRNGIY	2.5	24	238	0.32/ YES	YES
HLA-DRB1*15:01	391	405	Consensus (smm/nn/sturniolo)	LSAQYVVLYRNGIYS	1.1	66	16.8	0.32/ YES	YES
HLA-DRB1*11:01	391	405	Consensus (smm/nn/sturniolo)	LSAQYVVLYRNGIYS	2.5	24	173. 2	0.32/ YES	YES
HLA-DRB1*15:01	392	406	Consensus (smm/nn/sturniolo)	SAQYVVLYRNGIYSP	1.2	67	15	0.32/ YES	YES
HLA-DRB1*11:01	392	406	Consensus (smm/nn/sturniolo)	SAQYVVLYRNGIYSP	2.5	24	151. 9	0.32/ YES	YES
HLA-DRB1*15:01	393	407	Consensus (smm/nn/sturniolo)	AQYVVLYRNGIYSPD	1.2	68	16.6	0.31/ YES	YES

HLA-DRB1*11:01	393	407	Consensus (smm/nn/sturniolo)	AQYVVLYRNGIYSPD	2.5	25	161.5	0.31/ YES	YES
HLA-DRB1*15:01	394	408	Consensus (smm/nn/sturniolo)	QYVVLYRNGIYSPD W	1.6	79	19.3	0.28/ YES	
HLA-DRB1*15:01	395	409	Consensus (smm/nn/sturniolo)	YVVLYRNGIYSPDW N	1.9	156	32	0.24/ YES	
HLA-DRB1*04:01	400	414	Consensus (smm/nn/sturniolo)	RNGIYSPDWNLNANS	3.2	241	80.4	0.08	YES
HLA-DRB1*04:01	401	415	Consensus (smm/nn/sturniolo)	NGIYSPDWNLNANSV	3.2	234	75.8	0.03	YES
HLA-DRB1*04:01	402	416	Consensus (smm/nn/sturniolo)	GIYSPDWNLNANSVT	3.2	230	61.5	0.12	
HLA-DRB1*04:01	461	475	Consensus (smm/nn/sturniolo)	GLEYYVFKTHHNAV S	1.1	117	33.4	1.22/ YES	YES
HLA-DRB1*04:01	462	476	Consensus (smm/nn/sturniolo)	LEYVVFCKTHHNAVSS	1.1	118	25.6	1.23/ YES	YES
HLA-DRB1*04:01	463	477	Consensus (smm/nn/sturniolo)	EYVVFCKTHHNAVSSY	1.1	156	24.3	1.25/ YES	
HLA-DRB1*04:01	464	478	Consensus (smm/nn/sturniolo)	YVVFCKTHHNAVSSYI	1.1	153	29.6	1.24/ YES	YES
HLA-DRB1*04:01	465	479	Consensus (smm/nn/sturniolo)	VVFCKTHHNAVSSYIK	1.2	177	39.7	1.26/ YES	YES
HLA-DRB3*01:01	471	485	Consensus (comb.lib./smm/nn)	HNAVSSYIKDVFRVI	1.2	122	130.9	0.29/ YES	YES
HLA-DRB3*01:01	472	486	Consensus (comb.lib./smm/nn)	NAVSSYIKDVFRVIP	1.2	125	145.6	0.29/ YES	YES
HLA-DRB3*01:01	473	487	Consensus (comb.lib./smm/nn)	AVSSYIKDVFRVIPS	1.2	126	167.4	0.29/ YES	YES
HLA-DRB3*01:01	474	488	Consensus (comb.lib./smm/nn)	VSSYIKDVFRVIPSE	1.2	126	195.1	0.3/YES	
HLA-DRB1*09:01	476	490	Consensus (comb.lib./smm/nn)	SYIKDVFRVIPSEVL	0.94	68	10.5	0.33/ YES	
HLA-DRB1*07:01	476	490	Consensus (comb.lib./smm/nn)	SYIKDVFRVIPSEVL	1.6	28	4	0.33/ YES	
HLA-DRB1*15:01	476	490	Consensus (smm/nn/sturniolo)	SYIKDVFRVIPSEVL	3	84	44.7	0.33/ YES	
HLA-DRB5*01:01	476	490	Consensus (smm/nn/sturniolo)	SYIKDVFRVIPSEVL	3.9	40	75.7	0.33/ YES	
HLA-DRB1*09:01	477	491	Consensus (comb.lib./smm/nn)	YIKDVFRVIPSEVLS	0.94	68	9	0.29/ YES	YES
HLA-DRB1*07:01	477	491	Consensus (comb.lib./smm/nn)	YIKDVFRVIPSEVLS	1.6	28	4.5	0.29/ YES	YES
HLA-DRB1*15:01	477	491	Consensus (smm/nn/sturniolo)	YIKDVFRVIPSEVLS	1.9	84	24.2	0.29/ YES	YES
HLA-DRB5*01:01	477	491	Consensus (smm/nn/sturniolo)	YIKDVFRVIPSEVLS	3.9	40	39.6	0.29/ YES	YES
HLA-DRB1*09:01	478	492	Consensus (comb.lib./smm/nn)	IKDVFRVIPSEVLSN	0.97	69	9	0.29/ YES	
HLA-DRB1*07:01	478	492	Consensus (comb.lib./smm/nn)	IKDVFRVIPSEVLSN	1.6	28	5	0.29/ YES	
HLA-DRB1*15:01	478	492	Consensus (smm/nn/sturniolo)	IKDVFRVIPSEVLSN	1.9	85	20.1	0.29/ YES	
HLA-DRB1*01:01	478	492	Consensus (comb.lib./smm/nn)	IKDVFRVIPSEVLSN	3.6	9	11.9	0.29/ YES	
HLA-DRB1*04:01	478	492	Consensus (smm/nn/sturniolo)	IKDVFRVIPSEVLSN	3.9	203	93.5	0.29/ YES	
HLA-DRB5*01:01	478	492	Consensus (smm/nn/sturniolo)	IKDVFRVIPSEVLSN	3.9	40	33	0.29/ YES	
HLA-DRB1*09:01	479	493	Consensus (comb.lib./smm/nn)	KDVFRVIPSEVLSNS	0.94	68	8.9	0.29/ YES	YES
HLA-DRB1*07:01	479	493	Consensus (comb.lib./smm/nn)	KDVFRVIPSEVLSNS	1.7	29	5.5	0.29/ YES	YES
HLA-DRB1*15:01	479	493	Consensus (smm/nn/sturniolo)	KDVFRVIPSEVLSNS	1.9	85	18.8	0.29/ YES	YES
HLA-DRB1*01:01	479	493	Consensus (comb.lib./smm/nn)	KDVFRVIPSEVLSNS	3.2	9	11	0.29/ YES	YES

HLA-DRB1*04:01	479	493	Consensus (simm/nn/sturniolo)	KDVFRVIPSEVLSNS	3.3	195	81.8	0.29/ YES	YES
HLA-DRB1*04:05	479	493	Consensus (simm/nn/sturniolo)	KDVFRVIPSEVLSNS	3.9	183	83	0.29/ YES	YES
HLA-DRB5*01:01	479	493	Consensus (simm/nn/sturniolo)	KDVFRVIPSEVLSNS	3.9	40	30.3	0.29/ YES	YES
HLA-DRB1*09:01	480	494	Consensus (comb.lib./simm/nn)	DVFRVIPSEVLSNSY	1.3	77	11.2	0.32/ YES	YES
HLA-DRB1*07:01	480	494	Consensus (comb.lib./simm/nn)	DVFRVIPSEVLSNSY	1.8	31	6.7	0.32/ YES	YES
HLA-DRB1*15:01	480	494	Consensus (simm/nn/sturniolo)	DVFRVIPSEVLSNSY	2.3	94	27.3	0.32/ YES	YES
HLA-DRB1*04:01	480	494	Consensus (simm/nn/sturniolo)	DVFRVIPSEVLSNSY	3.7	210	90.4	0.32/ YES	YES
HLA-DRB5*01:01	480	494	Consensus (simm/nn/sturniolo)	DVFRVIPSEVLSNSY	3.9	41	47.6	0.32/ YES	YES

Table S5. List of epitopes of P04347 predicted by NetMHCIIpan method along with interacting HLA class II alleles and relevant scores.

Allele	Star t	End	Method	Peptide	Percentile rank	NetMHCIIpan IC50	IL-4 pred	AllerTOP v. 2.0
HLA-DPA1*01:03/DPB1*04:01	1	15	NetMHCIIpan	MGKPFFTLSSLCL	0.82	83.39	-0.16	
HLA-DPA1*01:03/DPB1*04:01	2	16	NetMHCIIpan	GKPFFTLSSLCLL	0.61	69.45	-1.43	
HLA-DRB3*02:02	2	16	NetMHCIIpan	GKPFFTLSSLCLL	3.6	213.13	-1.43	
HLA-DPA1*01:03/DPB1*04:01	3	17	NetMHCIIpan	KPFFTLSSLCLLL	0.52	63.83	-1.2	
HLA-DRB3*02:02	3	17	NetMHCIIpan	KPFFTLSSLCLLL	3.7	220.41	-1.2	
HLA-DPA1*01:03/DPB1*04:01	4	18	NetMHCIIpan	PFFTLSSLCLLLL	0.73	79.86	-1.02	
HLA-DPA1*01:03/DPB1*04:01	5	19	NetMHCIIpan	FFTLSSLCLLLLS	1.2	102.65	-1.02	
HLA-DPA1*01:03/DPB1*04:01	6	20	NetMHCIIpan	FTLSSLCLLLLS	3.2	177.4	-1.02	YES
HLA-DRB3*02:02	72	86	NetMHCIIpan	SKRTLNRNGSHLPSY	2.7	158.7	0.06	
HLA-DRB3*02:02	73	87	NetMHCIIpan	KRTLNRNGSHLPSYL	2.1	118.4	-0.03	
HLA-DRB3*02:02	74	88	NetMHCIIpan	RTLNRNGSHLPSYLP	3.3	198.25	-0.02	
HLA-DRB3*02:02	182	196	NetMHCIIpan	QNPRVFYLAGNPDIE	3.2	188.16	0.48/ YES	
HLA-DRB3*02:02	183	197	NetMHCIIpan	NPRVFYLAGNPDIEH	2.3	133.08	0.47/ YES	YES
HLA-DRB3*02:02	184	198	NetMHCIIpan	PRVFYLAGNPDIEHP	2.2	131.08	0.47/ YES	YES
HLA-DRB3*02:02	185	199	NetMHCIIpan	RVFYLAGNPDIEHPE	2.5	143.42	0.51/ YES	YES
HLA-DPA1*01:03/DPB1*04:01	230	244	NetMHCIIpan	SGFSKHFLAQSFNTN	1.2	100.51	0.38/ YES	
HLA-DPA1*01:03/DPB1*04:01	231	245	NetMHCIIpan	GFSKHFLAQSFNTNE	1.1	96.12	0.37/ YES	YES
HLA-DPA1*01:03/DPB1*04:01	232	246	NetMHCIIpan	FSKHFLAQSFNTNED	1.4	110.02	0.34/ YES	YES
HLA-DPA1*01:03/DPB1*04:01	233	247	NetMHCIIpan	SKHFLAQSFNTNEDT	3	171.87	0.34/ YES	
HLA-DRB3*02:02	350	364	NetMHCIIpan	ICTMKLHENIARPSR	3.9	228.89	0.06	YES
HLA-DRB3*02:02	351	365	NetMHCIIpan	CTMKLHENIARPSRA	3	172.8	0.02	
HLA-DRB3*02:02	352	366	NetMHCIIpan	TMKLHENIARPSRAD	3.6	213.39	-0.04	
HLA-DRB3*02:02	402	416	NetMHCIIpan	GIYSPDWNLNANSVT	3.7	220.12	0.12	
HLA-DRB3*02:02	403	417	NetMHCIIpan	IYSPDWNLNANSVTM	1.9	104.54	0.15	
HLA-DRB3*02:02	404	418	NetMHCIIpan	YSPDWNLNANSVTMT	1.4	77.6	0.14	YES
HLA-DRB3*02:02	405	419	NetMHCIIpan	SPDWNLNANSVTMTR	0.87	53.35	0.12	
HLA-DRB3*02:02	406	420	NetMHCIIpan	PDWNLNANSVTMTRG	0.99	59.86	0.15	YES
HLA-DRB3*02:02	407	421	NetMHCIIpan	DWNLNANSVTMTRG K	1.1	66.19	0.15	YES
HLA-DRB3*02:02	408	422	NetMHCIIpan	WNLNANSVTMTRGK G	1.8	99.3	0.09	YES
HLA-DRB3*02:02	461	475	NetMHCIIpan	GLEYYVFKTHHNAVS	2.4	142.43	1.22/ YES	YES
HLA-DRB3*02:02	462	476	NetMHCIIpan	LEYVVFVKTHHNAVSS	1.4	81.23	1.23/ YES	YES

HLA-DRB3*02:02	463	477	NetMHCIIpan	EYVVFKTHHNAVSSY	0.93	55.33	1.25/ YES	
HLA-DRB3*02:02	464	478	NetMHCIIpan	YVVFKTHHNAVSSYI	0.78	47.54	1.24/ YES	YES
HLA-DRB3*02:02	465	479	NetMHCIIpan	VVFKTHHNAVSSYIK	1.2	70.83	1.26/ YES	YES
HLA-DRB3*02:02	466	480	NetMHCIIpan	VFKTHHNAVSSYIKD	2.2	129.52	1.26/ YES	
HLA-DPA1*01:03/DPB1*04:01	476	490	NetMHCIIpan	SYIKDVFRVIPSEVL	3.1	175.28	0.33/ YES	
HLA-DRB3*02:02	476	490	NetMHCIIpan	SYIKDVFRVIPSEVL	3.8	221.73	0.33/ YES	
HLA-DPA1*02:01/DPB1*14:01	477	491	NetMHCIIpan	YIKDVFRVIPSEVLS	0.12	208.39	0.29/ YES	YES
HLA-DRB3*02:02	477	491	NetMHCIIpan	YIKDVFRVIPSEVLS	2.1	115.93	0.29/ YES	YES
HLA-DPA1*01:03/DPB1*04:01	477	491	NetMHCIIpan	YIKDVFRVIPSEVLS	3.6	194.22	0.29/ YES	YES
HLA-DPA1*02:01/DPB1*14:01	478	492	NetMHCIIpan	IKDVFRVIPSEVLSN	0.11	190.14	0.29/ YES	
HLA-DRB3*02:02	478	492	NetMHCIIpan	IKDVFRVIPSEVLSN	1.7	93.4	0.29/ YES	
HLA-DPA1*02:01/DPB1*14:01	479	493	NetMHCIIpan	KDVFRVIPSEVLSNS	0.09	173.29	0.29/ YES	YES
HLA-DRB3*02:02	479	493	NetMHCIIpan	KDVFRVIPSEVLSNS	1.5	83.96	0.29/ YES	YES
HLA-DPA1*02:01/DPB1*14:01	480	494	NetMHCIIpan	DVFRVIPSEVLSNSY	0.2	232.16	0.32/ YES	YES
HLA-DRB3*02:02	480	494	NetMHCIIpan	DVFRVIPSEVLSNSY	2.3	131.97	0.32/ YES	YES
HLA-DPA1*02:01/DPB1*14:01	481	495	NetMHCIIpan	VFRVIPSEVLSNSYN	0.2	229.3	0.31/ YES	YES
HLA-DRB3*02:02	481	495	NetMHCIIpan	VFRVIPSEVLSNSYN	3	170.25	0.31/ YES	YES
HLA-DRB3*02:02	500	514	NetMHCIIpan	QVRQLKYQGNSGPLV	2.1	116.97	0.07	YES
HLA-DRB3*02:02	501	515	NetMHCIIpan	VRQLKYQGNSGPLVN	0.99	58.24	0.01	YES
HLA-DRB3*02:02	502	516	NetMHCIIpan	RQLKYQGNSGPLVNP	0.95	55.81	0	

Table S6. List of epitopes of P04776 predicted by consensus method along with interacting HLA class II alleles and relevant scores.

Allele	Start	End	Method	Peptide	Percentile rank	Smm IC50	Nn IC50	IL-4 pred	AllerTOP v. 2.0
HLA-DPA1*01:03/DPB1*02:01	1	15	Consensus (comb.lib./smm/nn)	MAKLVFSLCFLLFSG	1.1	49	24.4	-0.95	
HLA-DPA1*02:01/DPB1*01:01	1	15	Consensus (comb.lib./smm/nn)	MAKLVFSLCFLLFSG	1.2	18	155.8	-0.95	
HLA-DPA1*01:03/DPB1*02:01	2	16	Consensus (comb.lib./smm/nn)	AKLVFSLCFLLFSGC	1.1	50	26.4	-0.83	
HLA-DPA1*02:01/DPB1*01:01	2	16	Consensus (comb.lib./smm/nn)	AKLVFSLCFLLFSGC	1.2	19	209.2	-0.83	
HLA-DPA1*01:03/DPB1*02:01	3	17	Consensus (comb.lib./smm/nn)	KLVFSLCFLLFSGCC	1.4	57	32.3	-0.78	
HLA-DPA1*01:03/DPB1*02:01	4	18	Consensus (comb.lib./smm/nn)	LVFSLCFLLFSGCCF	2.6	104	39.9	-0.73	YES
HLA-DPA1*01:03/DPB1*02:01	5	19	Consensus (comb.lib./smm/nn)	VFSLCFLLFSGCCFA	3.4	160	50.9	-0.73	YES
HLA-DRB4*01:01	158	172	Consensus (comb.lib./smm/nn)	TPVVAVSIIDTNSLE	3.9	174	69	0.21/ YES	YES
HLA-DRB4*01:01	189	203	Consensus (comb.lib./smm/nn)	EQEFLKYQQEQGGHQ	3.2	92	59.8	0.31/ YES	
HLA-DRB4*01:01	190	204	Consensus (comb.lib./smm/nn)	QEFLKYQQEQGGHQS	2.8	91	53.1	0.28/ YES	
HLA-DPA1*02:01/DPB1*01:01	215	229	Consensus (comb.lib./smm/nn)	NEGGSILSGFTLEFL	0.72	37	48.3	0.32/ YES	
HLA-DPA1*01:03/DPB1*02:01	215	229	Consensus (comb.lib./smm/nn)	NEGGSILSGFTLEFL	2.5	78	60.4	0.32/ YES	
HLA-DPA1*02:01/DPB1*01:01	216	230	Consensus (comb.lib./smm/nn)	EGGSILSGFTLEFLE	0.68	36	37.8	0.47/ YES	
HLA-DPA1*01:03/DPB1*02:01	216	230	Consensus (comb.lib./smm/nn)	EGGSILSGFTLEFLE	1.9	76	47	0.47/ YES	
HLA-DPA1*03:01/DPB1*04:02	216	230	Consensus (comb.lib./smm/nn)	EGGSILSGFTLEFLE	3	100	70.8	0.47/ YES	
HLA-DPA1*02:01/DPB1*01:01	217	231	Consensus (comb.lib./smm/nn)	GGSilSGFTLEFLEH	0.47	27	26.6	1.46/ YES	YES
HLA-DPA1*01:03/DPB1*02:01	217	231	Consensus (comb.lib./smm/nn)	GGSilSGFTLEFLEH	1.5	55	41.8	1.46/ YES	YES
HLA-DPA1*03:01/DPB1*04:02	217	231	Consensus (comb.lib./smm/nn)	GGSilSGFTLEFLEH	2.2	100	39.1	1.46/ YES	YES
HLA-DPA1*02:01/DPB1*01:01	218	232	Consensus (comb.lib./smm/nn)	GSILSGFTLEFLEHA	0.37	23	21.9	1.46/ YES	
HLA-DPA1*03:01/DPB1*04:02	218	232	Consensus (comb.lib./smm/nn)	GSILSGFTLEFLEHA	0.74	43	13.9	1.46/ YES	
HLA-DPA1*01:03/DPB1*02:01	218	232	Consensus (comb.lib./smm/nn)	GSILSGFTLEFLEHA	1.2	45	36.5	1.46/ YES	
HLA-DPA1*02:01/DPB1*01:01	219	233	Consensus (comb.lib./smm/nn)	SILSGFTLEFLEHAF	0.37	23	21.5	1.75/ YES	
HLA-DPA1*03:01/DPB1*04:02	219	233	Consensus (comb.lib./smm/nn)	SILSGFTLEFLEHAF	0.74	43	13	1.75/ YES	
HLA-DPA1*01:03/DPB1*02:01	219	233	Consensus (comb.lib./smm/nn)	SILSGFTLEFLEHAF	1.3	44	38.7	1.75/ YES	
HLA-DPA1*02:01/DPB1*01:01	220	234	Consensus (comb.lib./smm/nn)	ILSGFTLEFLEHAFS	0.87	41	29.7	1.80/ YES	YES
HLA-DPA1*03:01/DPB1*04:02	220	234	Consensus (comb.lib./smm/nn)	ILSGFTLEFLEHAFS	1.1	55	14.9	1.80/ YES	YES
HLA-DPA1*01:03/DPB1*02:01	220	234	Consensus (comb.lib./smm/nn)	ILSGFTLEFLEHAFS	1.9	65	48.7	1.80/ YES	YES
HLA-DPA1*02:01/DPB1*01:01	221	235	Consensus (comb.lib./smm/nn)	LSGFTLEFLEHAFSV	0.87	41	37.2	1.83/ YES	YES
HLA-DPA1*03:01/DPB1*04:02	221	235	Consensus (comb.lib./smm/nn)	LSGFTLEFLEHAFSV	1.1	54	14.6	1.83/ YES	YES
HLA-DPA1*01:03/DPB1*02:01	221	235	Consensus (comb.lib./smm/nn)	LSGFTLEFLEHAFSV	1.5	63	43.6	1.83/ YES	YES
HLA-DPA1*03:01/DPB1*04:02	222	236	Consensus (comb.lib./smm/nn)	SGFTLEFLEHAFSVD	1.2	62	18.3	0.83/ YES	YES
HLA-DPA1*01:03/DPB1*02:01	222	236	Consensus (comb.lib./smm/nn)	SGFTLEFLEHAFSVD	2.3	77	57.3	0.83/ YES	YES
HLA-DPA1*01:03/DPB1*02:01	223	237	Consensus (comb.lib./smm/nn)	GFTLEFLEHAFSVDK	3.4	125	79.8	0.62/ YES	
HLA-DRB1*08:02	252	266	Consensus (smm/nn/sturniolo)	KGAIIVTVKGGLSVIK	2.8	180	118.5	1.24/	

									YES	
HLA-DRB1*08:02	253	267	Consensus (smm/nn/sturniolo)	GAIVTVKGGLSVIKP	3.2	183	130	0.25/ YES		
HLA-DQA1*05:01/DQB1*03:01	331	345	Consensus (comb.lib./smm/nn)	PDIYNPQAGSVTTAT	3.7	41	11.3	0.09	YES	
HLA-DQA1*05:01/DQB1*03:01	332	346	Consensus (comb.lib./smm/nn)	DIYNPQAGSVTTATS	3.5	39	10.3	0.12	YES	
HLA-DQA1*05:01/DQB1*03:01	333	347	Consensus (comb.lib./smm/nn)	IYNPQAGSVTTATSL	3.7	41	10.1	0.14	YES	
HLA-DQA1*05:01/DQB1*03:01	334	348	Consensus (comb.lib./smm/nn)	YNPQAGSVTTATSLD	4	44	11.4	0.05	YES	
HLA-DPA1*01:03/DPB1*02:01	343	357	Consensus (comb.lib./smm/nn)	TATSLDFPALS WLRL	2.5	100	48.3	1.21/ YES		
HLA-DPA1*01:03/DPB1*02:01	344	358	Consensus (comb.lib./smm/nn)	ATSLDFPALS WLRLS	2.5	99	55.9	1.21/ YES		
HLA-DPA1*03:01/DPB1*04:02	345	359	Consensus (comb.lib./smm/nn)	TSLDFPALS WLRLSA	2.8	56	68.1	0.21/ YES		
HLA-DPA1*01:03/DPB1*02:01	345	359	Consensus (comb.lib./smm/nn)	TSLDFPALS WLRLSA	2.9	97	69.7	0.21/ YES		
HLA-DPA1*03:01/DPB1*04:02	346	360	Consensus (comb.lib./smm/nn)	SLDFPALS WLRLSAE	1.8	57	51.8	0.21/ YES		
HLA-DPA1*01:03/DPB1*02:01	346	360	Consensus (comb.lib./smm/nn)	SLDFPALS WLRLSAE	3.6	100	85	0.21/ YES		
HLA-DPA1*03:01/DPB1*04:02	347	361	Consensus (comb.lib./smm/nn)	LDFPALS WLRLSAEF	1.8	58	49.8	0.25/ YES	YES	
HLA-DPA1*03:01/DPB1*04:02	348	362	Consensus (comb.lib./smm/nn)	DFPALS WLRLSAEFG	3.8	165	61.2	0.26/ YES	YES	
HLA-DPA1*03:01/DPB1*04:02	349	363	Consensus (comb.lib./smm/nn)	FPALS WLRLSAEFGS	3.8	169	69.1	0.11	YES	
HLA-DRB1*04:01	366	380	Consensus (smm/nn/sturniolo)	KNAMFVPHYNLNAN S	2.8	240	71.7	0.26	YES	
HLA-DRB1*04:01	367	381	Consensus (smm/nn/sturniolo)	NAMFVPHYNLNANSI	2.5	216	63.8	0.27		
HLA-DRB1*04:01	368	382	Consensus (smm/nn/sturniolo)	AMFVPHYNLNANSII	1.2	197	39.5	0.27/ YES	YES	
HLA-DRB1*13:02	368	382	Consensus (smm/nn/sturniolo)	AMFVPHYNLNANSII	3.1	42	41.1	0.27/ YES	YES	
HLA-DRB1*07:01	368	382	Consensus (comb.lib./smm/nn)	AMFVPHYNLNANSII	3.7	55	22.1	0.27/ YES	YES	
HLA-DRB1*13:02	369	383	Consensus (smm/nn/sturniolo)	MFVPHYNLNANSIIY	2.7	42	19.6	0.26/ YES	YES	
HLA-DRB3*01:01	369	383	Consensus (comb.lib./smm/nn)	MFVPHYNLNANSIIY	3.2	240	79.4	0.26/ YES	YES	
HLA-DRB1*07:01	369	383	Consensus (comb.lib./smm/nn)	MFVPHYNLNANSIIY	3.6	54	26.9	0.26/ YES	YES	
HLA-DRB1*13:02	370	384	Consensus (smm/nn/sturniolo)	FVPHYNLNANSIIYA	1.3	41	9.8	0.26/ YES	YES	
HLA-DRB3*01:01	370	384	Consensus (comb.lib./smm/nn)	FVPHYNLNANSIIYA	2.9	241	68.8	0.26/ YES	YES	
HLA-DQA1*01:02/DQB1*06:02	371	385	Consensus (comb.lib./smm/nn)	VPHYNLNANSIIYAL	0.39	40	29.7	0.23/ YES	YES	
HLA-DRB1*13:02	371	385	Consensus (smm/nn/sturniolo)	VPHYNLNANSIIYAL	0.88	40	7.8	0.23/ YES	YES	
HLA-DRB3*01:01	371	385	Consensus (comb.lib./smm/nn)	VPHYNLNANSIIYAL	2.6	235	56.2	0.23/ YES	YES	
HLA-DQA1*01:02/DQB1*06:02	372	386	Consensus (comb.lib./smm/nn)	PHYNLNANSIIYALN	0.26	40	21.8	0.24/ YES	YES	
HLA-DRB1*13:02	372	386	Consensus (smm/nn/sturniolo)	PHYNLNANSIIYALN	1.2	41	8.7	0.24/ YES	YES	
HLA-DRB3*01:01	372	386	Consensus (comb.lib./smm/nn)	PHYNLNANSIIYALN	3.3	242	81.5	0.24/ YES	YES	
HLA-DQA1*01:02/DQB1*06:02	373	387	Consensus (comb.lib./smm/nn)	HYNLNANSIIYALNG	0.22	38	19.3	0.25/ YES	YES	
HLA-DRB1*13:02	373	387	Consensus (smm/nn/sturniolo)	HYNLNANSIIYALNG	2.8	119	10.8	0.25/ YES	YES	
HLA-DQA1*01:02/DQB1*06:02	374	388	Consensus (comb.lib./smm/nn)	YNLNANSIIYALNGR	0.22	38	23.2	0.03	YES	
HLA-DRB1*13:02	374	388	Consensus (smm/nn/sturniolo)	YNLNANSIIYALNGR	3.5	119	25.5	0.03	YES	
HLA-DQA1*01:02/DQB1*06:02	375	389	Consensus (comb.lib./smm/nn)	NLNANSIIYALNGRA	0.79	69	33.5	-0,02	YES	
HLA-DRB1*09:01	376	390	Consensus (comb.lib./smm/nn)	LNANSIIYALNGRAL	0.64	169	15	-0.92	YES	
HLA-DQA1*01:02/DQB1*06:02	376	390	Consensus (comb.lib./smm/nn)	LNANSIIYALNGRAL	1.7	104	62.4	-0.92	YES	

HLA-DRB1*01:01	376	390	Consensus (comb.lib./smm/nn)	LNANSIIYALNGRAL	2.2	17	9	-0.92	YES
HLA-DRB1*09:01	377	391	Consensus (comb.lib./smm/nn)	NANSIIYALNGRALI	0.3	155	11.4	-0.96	YES
HLA-DRB1*01:01	377	391	Consensus (comb.lib./smm/nn)	NANSIIYALNGRALI	0.54	13	5	-0.96	YES
HLA-DRB1*13:02	377	391	Consensus (smm/nn/sturniolo)	NANSIIYALNGRALI	3.9	163	12.3	-0.96	YES
HLA-DRB1*01:01	378	392	Consensus (comb.lib./smm/nn)	ANSIIYALNGRALIQ	0.27	13	4.2	-0.97	YES
HLA-DRB1*09:01	378	392	Consensus (comb.lib./smm/nn)	ANSIIYALNGRALIQ	0.3	154	11.1	-0.97	YES
HLA-DRB1*13:02	378	392	Consensus (smm/nn/sturniolo)	ANSIIYALNGRALIQ	3.9	163	11.8	-0.97	YES
HLA-DRB1*01:01	379	393	Consensus (comb.lib./smm/nn)	NSIIYALNGRALIQV	0.25	12	4.1	-0.90	YES
HLA-DRB1*09:01	379	393	Consensus (comb.lib./smm/nn)	NSIIYALNGRALIQV	0.33	151	11.6	-0.90	YES
HLA-DRB1*13:02	379	393	Consensus (smm/nn/sturniolo)	NSIIYALNGRALIQV	4	165	11.5	-0.90	YES
HLA-DRB1*09:01	380	394	Consensus (comb.lib./smm/nn)	SIIYALNGRALIQVV	0.49	180	13.9	-0.91	
HLA-DRB1*01:01	380	394	Consensus (comb.lib./smm/nn)	SIIYALNGRALIQVV	0.56	13	5.1	-0.91	
HLA-DRB1*01:01	381	395	Consensus (comb.lib./smm/nn)	IYYALNGRALIQVVN	2.2	29	9	-0.91	YES
HLA-DRB1*11:01	412	426	Consensus (smm/nn/sturniolo)	LIVPQNFVVAARSQS	4	33	61.2	0.42/ YES	YES
HLA-DRB1*11:01	413	427	Consensus (smm/nn/sturniolo)	IVPQNFVVAARSQSD	3.5	34	34.3	0.45/ YES	YES
HLA-DRB1*11:01	414	428	Consensus (smm/nn/sturniolo)	VPQNFVVAARSQSDN	2.5	34	25.5	0.40/ YES	YES
HLA-DRB1*11:01	415	429	Consensus (smm/nn/sturniolo)	PQNFVVAARSQSDNF	2.2	33	21.6	0.38/ YES	YES
HLA-DRB1*11:01	416	430	Consensus (smm/nn/sturniolo)	QNFVVAARSQSDNFE	2.1	33	20.7	0.40/ YES	YES
HLA-DRB1*11:01	417	431	Consensus (smm/nn/sturniolo)	NFVVAARSQSDNFEY	4	97	35.6	1.40/ YES	YES
HLA-DPA1*01:03/DPB1*02:01	423	437	Consensus (comb.lib./smm/nn)	RSQSDNFEYVSFKTN	1.8	73	44.5	1.3/YES	YES
HLA-DPA1*02:01/DPB1*01:01	423	437	Consensus (comb.lib./smm/nn)	RSQSDNFEYVSFKTN	3.8	134	62.2	1.3/YES	YES
HLA-DPA1*01:03/DPB1*02:01	424	438	Consensus (comb.lib./smm/nn)	SQSDNFEYVSFKTND	1.9	73	48.4	0.3/YES	YES
HLA-DPA1*02:01/DPB1*01:01	424	438	Consensus (comb.lib./smm/nn)	SQSDNFEYVSFKTND	3.8	133	56.7	0.3/YES	YES
HLA-DRB1*04:05	425	439	Consensus (smm/nn/sturniolo)	QSDNFEYVSFKTNDT	0.47	40	127.7	0.3/YES	YES
HLA-DPA1*01:03/DPB1*02:01	425	439	Consensus (comb.lib./smm/nn)	QSDNFEYVSFKTNDT	2.1	75	52	0.3/YES	YES
HLA-DPA1*02:01/DPB1*01:01	425	439	Consensus (comb.lib./smm/nn)	QSDNFEYVSFKTNDT	3.8	133	53.7	0.3/YES	YES
HLA-DRB1*04:05	426	440	Consensus (smm/nn/sturniolo)	SDNFEYVSFKTNDTP	0.49	41	87	0.3/YES	YES
HLA-DPA1*01:03/DPB1*02:01	426	440	Consensus (comb.lib./smm/nn)	SDNFEYVSFKTNDTP	2.5	75	60.7	0.3/YES	YES
HLA-DPA1*02:01/DPB1*01:01	426	440	Consensus (comb.lib./smm/nn)	SDNFEYVSFKTNDTP	4	139	57	0.3/YES	YES
HLA-DRB1*04:05	427	441	Consensus (smm/nn/sturniolo)	DNFEYVSFKTNDTPM	0.47	40	79.1	0.3/YES	
HLA-DRB1*04:01	427	441	Consensus (smm/nn/sturniolo)	DNFEYVSFKTNDTPM	3.8	70	91.9	0.3/YES	
HLA-DRB1*04:05	428	442	Consensus (smm/nn/sturniolo)	NFEYVSFKTNDTPMI	0.42	38	55.6	0.3/YES	YES
HLA-DRB1*04:01	428	442	Consensus (smm/nn/sturniolo)	NFEYVSFKTNDTPMI	1.6	59	47.6	0.3/YES	YES
HLA-DRB3*01:01	428	442	Consensus (comb.lib./smm/nn)	NFEYVSFKTNDTPMI	3.1	117	76.9	0.3/YES	YES
HLA-DRB1*13:02	428	442	Consensus (smm/nn/sturniolo)	NFEYVSFKTNDTPMI	4	113	32.8	0.3/YES	YES
HLA-DRB1*04:05	429	443	Consensus (smm/nn/sturniolo)	FEYVSFKTNDTPMIG	0.44	39	64	0.28/ YES	
HLA-DRB1*04:01	429	443	Consensus (smm/nn/sturniolo)	FEYVSFKTNDTPMIG	1.2	59	41.4	0.28/ YES	
HLA-DRB3*01:01	429	443	Consensus (comb.lib./smm/nn)	FEYVSFKTNDTPMIG	3.1	120	75.9	0.28/ YES	
HLA-DRB1*13:02	429	443	Consensus (smm/nn/sturniolo)	FEYVSFKTNDTPMIG	3.8	115	27.9	0.28/ YES	
HLA-DRB1*04:01	430	444	Consensus (smm/nn/sturniolo)	EYVSFKTNDTPMIGT	2.2	120	43.3	0.17	YES
HLA-DRB3*01:01	430	444	Consensus (comb.lib./smm/nn)	EYVSFKTNDTPMIGT	3.1	120	75.7	0.17	YES
HLA-DRB1*04:05	430	444	Consensus (smm/nn/sturniolo)	EYVSFKTNDTPMIGT	3.2	115	118.9	0.17	YES
HLA-DRB1*13:02	430	444	Consensus (smm/nn/sturniolo)	EYVSFKTNDTPMIGT	3.5	115	25	0.17	YES
HLA-DRB1*04:01	431	445	Consensus (smm/nn/sturniolo)	YVSFKTNDTPMIGTL	2.7	120	68	0.15	YES
HLA-DRB3*01:01	431	445	Consensus (comb.lib./smm/nn)	YVSFKTNDTPMIGTL	2.7	117	64.7	0.15	YES
HLA-DRB1*13:02	431	445	Consensus (smm/nn/sturniolo)	YVSFKTNDTPMIGTL	3	111	21.4	0.15	YES
HLA-DRB1*04:05	431	445	Consensus (smm/nn/sturniolo)	YVSFKTNDTPMIGTL	3.2	116	185.5	0.15	YES
HLA-DRB3*01:01	432	446	Consensus (comb.lib./smm/nn)	VSFKTNDTPMIGTLA	3.6	116	89.8	0.15	YES
HLA-DQA1*05:01/DQB1*03:01	436	450	Consensus (comb.lib./smm/nn)	TNDTPMIGTLAGANS	4	37	149.5	-0.03	
HLA-DQA1*05:01/DQB1*03:01	437	451	Consensus (comb.lib./smm/nn)	NDTPMIGTLAGANSL	4	34	117.2	0.07	
HLA-DQA1*05:01/DQB1*03:01	438	452	Consensus (comb.lib./smm/nn)	DTPMIGTLAGANSLL	4	32	110.2	-0.09	
HLA-DQA1*05:01/DQB1*03:01	439	453	Consensus (comb.lib./smm/nn)	TPMIGTLAGANSLLN	4	37	97	-0.09	YES
HLA-DRB1*09:01	447	461	Consensus (comb.lib./smm/nn)	GANSLLNALPEEVIQ	3.4	121	41.8	-0.52	
HLA-DRB1*09:01	448	462	Consensus (comb.lib./smm/nn)	ANSLLNALPEEVIQH	4	121	46	-0.55	

HLA-DRB1*04:01	456	470	Consensus (smm/nn/sturniolo)	PEEVIQHTFNLKSQQ	3.6	240	86.5	0.21/ YES	
HLA-DRB1*04:01	457	471	Consensus (smm/nn/sturniolo)	EEVIQHTFNLKSQQA	3	229	75.9	0.20/ YES	
HLA-DRB1*04:01	458	472	Consensus (smm/nn/sturniolo)	EVIQHTFNLKSQQAR	3.2	242	79.6	0.25/ YES	
HLA-DRB1*13:02	468	482	Consensus (smm/nn/sturniolo)	SQQARQIKNNNPFKF	0.18	14	3.5	1.33/ YES	YES
HLA-DRB1*13:02	469	483	Consensus (smm/nn/sturniolo)	QQARQIKNNNPFKFL	0.09	14	2.6	1.33/ YES	
HLA-DRB3*01:01	469	483	Consensus (comb.lib./smm/nn)	QQARQIKNNNPFKFL	3.3	177	81.1	1.33/ YES	
HLA-DRB1*13:02	470	484	Consensus (smm/nn/sturniolo)	QARQIKNNNPFKFLV	0.08	13	2.4	1.29/ YES	YES
HLA-DRB3*01:01	470	484	Consensus (comb.lib./smm/nn)	QARQIKNNNPFKFLV	2.4	173	55.6	1.29/ YES	YES
HLA-DRB1*13:02	471	485	Consensus (smm/nn/sturniolo)	ARQIKNNNPFKFLVP	0.09	14	2.4	1.29/ YES	YES
HLA-DRB3*01:01	471	485	Consensus (comb.lib./smm/nn)	ARQIKNNNPFKFLVP	2.7	177	63.3	1.29/ YES	YES
HLA-DRB1*13:02	472	486	Consensus (smm/nn/sturniolo)	RQIKNNNPFKFLVPP	0.09	15	2.7	1.28/ YES	YES
HLA-DRB3*01:01	472	486	Consensus (comb.lib./smm/nn)	RQIKNNNPFKFLVPP	3.9	185	98.1	1.28/ YES	YES
HLA-DRB1*13:02	473	487	Consensus (smm/nn/sturniolo)	QIKNNNPFKFLVPPQ	0.59	41	5.8	1.28/ YES	YES
HLA-DRB1*04:05	474	488	Consensus (smm/nn/sturniolo)	IKNNNPFKFLVPPQE	3.1	113	55.9	1.28/ YES	YES
HLA-DRB1*13:02	474	488	Consensus (smm/nn/sturniolo)	IKNNNPFKFLVPPQE	3.4	45	24.2	1.28/ YES	YES
HLA-DRB1*04:05	475	489	Consensus (smm/nn/sturniolo)	KNNNPFKFLVPPQES	3	112	34.2	0.28/ YES	YES
HLA-DRB1*04:05	476	490	Consensus (smm/nn/sturniolo)	NNNPFKFLVPPQESQ	3	111	27.4	0.28/ YES	
HLA-DRB1*04:05	477	491	Consensus (smm/nn/sturniolo)	NNPFKFLVPPQESQK	2.9	110	26.6	0.28/ YES	YES
HLA-DRB1*04:05	478	492	Consensus (smm/nn/sturniolo)	NPFKFLVPPQESQKR	3.4	119	32.2	0.28/ YES	

Table S7. List of epitopes of P04776 predicted by NetMHCIIpan method along with interacting HLA class II alleles and relevant scores.

Allele	Start	End	Method	Peptide	Percentile rank	NetMHCIIpan IC50	IL-4 pred	AllerTOP v. 2.0
HLA-DRB3*02:02	67	81	NetMHCIIpan	VALSRCTLNRNALRR	2	113.55	0.2	
HLA-DRB3*02:02	68	82	NetMHCIIpan	ALSRCTLNRNALRRP	1.9	103.12	0.2	YES
HLA-DRB3*02:02	69	83	NetMHCIIpan	LSRCTLNRNALRRPS	1.5	87.3	0.2	YES
HLA-DRB3*02:02	70	84	NetMHCIIpan	SRCTLNRNALRRPSY	1.9	107.36	0.24/YES	
HLA-DRB3*02:02	71	85	NetMHCIIpan	RCTLNRNALRRPSYT	2.6	149.92	0.24/YES	
HLA-DPA1*01:03/DPB1*04:01	215	229	NetMHCIIpan	NEGGSILSGFTLEFL	2.4	143.96	0.32/YES	
HLA-DPA1*01:03/DPB1*04:01	216	230	NetMHCIIpan	EGGSILSGFTLEFLE	2	132.16	0.47/YES	
HLA-DPA1*01:03/DPB1*04:01	217	231	NetMHCIIpan	GGSILSGFTLEFLEH	1.4	109.82	1.46/YES	YES
HLA-DPA1*01:03/DPB1*04:01	218	232	NetMHCIIpan	GSILSGFTLEFLEHA	0.61	72.68	1.57/YES	
HLA-DPA1*01:03/DPB1*04:01	219	233	NetMHCIIpan	SILSGFTLEFLEHAF	0.38	58.42	1.75/YES	
HLA-DPA1*01:03/DPB1*04:01	220	234	NetMHCIIpan	ILSGFTLEFLEHAFS	0.52	64.19	1.80/YES	YES
HLA-DPA1*01:03/DPB1*04:01	221	235	NetMHCIIpan	LSGFTLEFLEHAFSV	0.43	60.27	1.83/YES	YES
HLA-DPA1*01:03/DPB1*04:01	222	236	NetMHCIIpan	SGFTLEFLEHAFSVD	0.61	71.22	0.83/YES	YES
HLA-DPA1*01:03/DPB1*04:01	223	237	NetMHCIIpan	GFTLEFLEHAFSVDK	0.82	83.62	0.62/YES	
HLA-DPA1*01:03/DPB1*04:01	224	238	NetMHCIIpan	FTLEFLEHAFSVDKQ	1.6	116.5	0.49/YES	
HLA-DRB3*02:02	315	329	NetMHCIIpan	TICTMRLRHNIGQTS	3.3	194.22	-1.09	
HLA-DRB3*02:02	316	330	NetMHCIIpan	ICTMRLRHNIGQTSS	1.6	91.42	-1.08	YES
HLA-DRB3*02:02	317	331	NetMHCIIpan	CTMRLRHNIGQTSSP	1.6	89.92	-1.13	YES
HLA-DRB3*02:02	318	332	NetMHCIIpan	TMRLRHNIGQTSSPD	1.9	103.36	-0.34	YES
HLA-DRB3*02:02	319	333	NetMHCIIpan	MRLRHNIGQTSSPDI	2.2	127.51	0.68/YES	YES
HLA-DPA1*01:03/DPB1*04:01	344	358	NetMHCIIpan	ATSLDFPALSWLRLS	3.2	178.66	1.21/YES	

HLA-DPA1*01:03/DPB1*04:01	345	359	NetMHCIIpan	TSLDFPALSWLRLSA	2.8	159.22	0.21/YES	
HLA-DPA1*01:03/DPB1*04:01	346	360	NetMHCIIpan	SLDFPALSWLRLSAE	2.9	166.18	0.21/YES	
HLA-DPA1*01:03/DPB1*04:01	347	361	NetMHCIIpan	LDFPALSWLRLSAEF	3.1	174.96	0.25/YES	YES
HLA-DRB3*02:02	366	380	NetMHCIIpan	KNAMFVPHYNLNAN S	4	235.42	0.26/YES	YES
HLA-DRB3*02:02	367	381	NetMHCIIpan	NAMFVPHYNLNANSI	2.5	143.19	0.27/YES	
HLA-DRB3*02:02	368	382	NetMHCIIpan	AMFVPHYNLNANSII	0.87	53.11	0.27/YES	YES
HLA-DRB3*02:02	369	383	NetMHCIIpan	MFVPHYNLNANSIIY	0.27	28.26	0.26/YES	YES
HLA-DRB3*02:02	370	384	NetMHCIIpan	FVPHYNLNANSIIYA	0.11	19.48	0.26/YES	YES
HLA-DRB3*02:02	371	385	NetMHCIIpan	VPHYNLNANSIIYAL	0.1	17.89	0.23/YES	YES
HLA-DRB3*02:02	372	386	NetMHCIIpan	PHYNLNANSIIYALN	0.12	20.16	0.24/YES	YES
HLA-DRB3*02:02	373	387	NetMHCIIpan	HYNLNANSIIYALNG	0.2	26.12	0.25/YES	YES
HLA-DRB3*02:02	374	388	NetMHCIIpan	YNLNANSIIYALNGR	1.2	69.67	0.03	YES
HLA-DRB3*02:02	376	390	NetMHCIIpan	LNANSIIYALNGRAL	2.8	160.35	-0.92	YES
HLA-DRB3*02:02	377	391	NetMHCIIpan	NANSIIYALNGRALI	0.71	43.38	-0.96	YES
HLA-DRB3*02:02	378	392	NetMHCIIpan	ANSIIYALNGRALIQ	0.38	31.92	-0.97	YES
HLA-DRB3*02:02	379	393	NetMHCIIpan	NSIIYALNGRALIQV	0.21	26.59	-0.90	YES
HLA-DRB3*02:02	380	394	NetMHCIIpan	SIYALNGRALIQVV	0.21	26.58	-0.91	
HLA-DRB3*02:02	381	395	NetMHCIIpan	IYALNGRALIQVVN	0.44	33.64	-0.91	YES
HLA-DRB3*02:02	382	396	NetMHCIIpan	IYALNGRALIQVVNC	0.9	54.74	-0.74	YES
HLA-DPA1*01:03/DPB1*04:01	423	437	NetMHCIIpan	RSQSDNFEYVSFKTN	3.6	195.31	1.30/YES	YES
HLA-DRB3*02:02	428	442	NetMHCIIpan	NFEYVSFKTNDTPMI	0.77	46.87	0.30/YES	YES
HLA-DRB3*02:02	429	443	NetMHCIIpan	FEYVSFKTNDTPMIG	0.22	26.96	0.28/YES	
HLA-DRB3*02:02	430	444	NetMHCIIpan	EYVSFKTNDTPMIGT	0.14	21.01	0.17	YES
HLA-DRB3*02:02	431	445	NetMHCIIpan	YVSFKTNDTPMIGTL	0.06	16.42	0.15	YES
HLA-DRB3*02:02	432	446	NetMHCIIpan	VSFKTNDTPMIGTLA	0.17	23.88	0.15	YES
HLA-DRB3*02:02	433	447	NetMHCIIpan	SFKTNDTPMIGTLAG	0.68	41.86	0.20	
HLA-DRB3*02:02	469	483	NetMHCIIpan	QQARQIKNNNPFKFL	1.8	101.29	1.33/YES	
HLA-DRB3*02:02	470	484	NetMHCIIpan	QARQIKNNNPFKFLV	1.1	66.99	1.29/YES	YES
HLA-DRB3*02:02	471	485	NetMHCIIpan	ARQIKNNNPFKFLVP	1.2	70.65	1.29/YES	YES
HLA-DRB3*02:02	472	486	NetMHCIIpan	RQIKNNNPFKFLVPP	1.6	88.32	1.28/YES	YES
HLA-DRB3*02:02	473	487	NetMHCIIpan	QIKNNNPFKFLVPPQ	3.5	205.67	1.28/YES	YES

Table S8. List of epitopes of P05046 predicted by consensus method along with interacting HLA class II alleles and relevant scores.

Allele	Start	End	Method	Peptide	Percentile rank	Smm IC50	Nn IC50	IL-4 pred	AllerTOP v. 2.0
HLA-DRB1*04:01	20	34	Consensus (smm/nn/sturniolo)	TLVLVLLTSKANSAE	3.8	64	92.6	-0.86	YES
HLA-DRB1*04:01	21	35	Consensus (smm/nn/sturniolo)	LVLVLLTSKANSAET	3.9	70	93.5	-0.79	YES
HLA-DPA1*01:03/DPB1*02:01	32	46	Consensus (comb.lib./smm/nn)	SAETVSFSWNKFVVPK	1.7	69	42.1	0.29/ YES	YES
HLA-DPA1*01:03/DPB1*02:01	33	47	Consensus (comb.lib./smm/nn)	AETVSFSWNKFVVPKQ	1.7	69	45.6	0.29/ YES	
HLA-DPA1*01:03/DPB1*02:01	34	48	Consensus (comb.lib./smm/nn)	ETVSFSWNKFVVPKQP	2.1	69	52.5	0.29/ YES	
HLA-DPA1*01:03/DPB1*02:01	35	49	Consensus (comb.lib./smm/nn)	TVSFSWNKFVVPKQPN	3.2	70	77.9	0.3/YES	
HLA-DRB1*13:02	77	91	Consensus (smm/nn/sturniolo)	PSSLGRALYSTPIHI	3.7	97	26.9	0.13	
HLA-DRB1*13:02	78	92	Consensus (smm/nn/sturniolo)	SSLGRALYSTPIHIW	3.8	98	28.2	0.24/ YES	YES
HLA-DQA1*05:01/DQB1*03:01	92	106	Consensus (comb.lib./smm/nn)	WDKETGSVASFAASF	3.5	39	85.1	0.42/ YES	YES
HLA-DQA1*05:01/DQB1*03:01	93	107	Consensus (comb.lib./smm/nn)	DKETGSVASFAASFN	3.3	37	82.6	0.44/ YES	YES
HLA-DQA1*05:01/DQB1*03:01	94	108	Consensus (comb.lib./smm/nn)	KETGSVASFAASFNF	3.7	41	73.9	0.52/ YES	YES
HLA-DQA1*05:01/DQB1*03:01	95	109	Consensus (comb.lib./smm/nn)	ETGSVASFAASFNFT	3.7	45	60.8	0.51/ YES	YES
HLA-DQA1*05:01/DQB1*03:01	96	110	Consensus (comb.lib./smm/nn)	TGSVASFAASFNFTF	3	63	51.5	0.08	YES
HLA-DRB1*04:01	96	110	Consensus (smm/nn/sturniolo)	TGSVASFAASFNFTF	3.9	166	74.2	0.08	
HLA-DRB1*04:01	97	111	Consensus (smm/nn/sturniolo)	GSVASFAASFNFTFY	3.9	168	42.5	0.06	
HLA-DQA1*05:01/DQB1*03:01	97	111	Consensus (comb.lib./smm/nn)	GSVASFAASFNFTFY	4	105	65.3	0.06	
HLA-DPA1*01:03/DPB1*02:01	100	114	Consensus (comb.lib./smm/nn)	ASFAASFNFTFYAPD	2	82	42	0.22/ YES	
HLA-DRB1*04:05	100	114	Consensus (smm/nn/sturniolo)	ASFAASFNFTFYAPD	3.1	91	70	0.22/ YES	YES
HLA-DPA1*01:03/DPB1*02:01	101	115	Consensus (comb.lib./smm/nn)	SFAASFNFTFYAPDT	2	82	42	0.21/ YES	YES
HLA-DPA1*01:03/DPB1*02:01	102	116	Consensus (comb.lib./smm/nn)	FAASFNFTFYAPDTK	2.2	85	54.4	0.21/ YES	
HLA-DPA1*01:03/DPB1*02:01	103	117	Consensus (comb.lib./smm/nn)	AASFNFTFYAPDTKR	3.8	85	90.2	-0.22	
HLA-DQA1*01:01/DQB1*05:01	118	132	Consensus (comb.lib./smm/nn)	LADGLAFFLAPIDTK	2.1	250	109.8	-0.03	
HLA-DRB1*04:05	133	147	Consensus (smm/nn/sturniolo)	PQTHAGYLGLFNENE	2.1	82	60.7	0.11	
HLA-DRB1*04:05	134	148	Consensus (smm/nn/sturniolo)	QTHAGYLGLFNENES	1.8	81	32.9	0.24/ YES	YES
HLA-DRB1*04:05	135	149	Consensus (smm/nn/sturniolo)	THAGYLGLFNENESG	1.8	82	33.7	0.24/ YES	
HLA-DRB1*04:05	136	150	Consensus (smm/nn/sturniolo)	HAGYLGLFNENESGD	1.8	80	32.8	0.27/ YES	
HLA-DRB1*04:05	137	151	Consensus (smm/nn/sturniolo)	AGYLGLFNENESGDQ	1.8	80	41.4	0.19	
HLA-DRB1*04:05	154	168	Consensus (smm/nn/sturniolo)	AVEFDTRNSWDPPN	4	104	84	0.25/ YES	
HLA-DRB1*13:02	167	181	Consensus (smm/nn/sturniolo)	PNPHIGINVNSIRSI	2.1	99	5.4	0.17	YES
HLA-DRB1*13:02	168	182	Consensus (smm/nn/sturniolo)	NPHIGINVNSIRSIK	2.1	99	5.1	0.24/ YES	YES
HLA-DRB1*13:02	169	183	Consensus (smm/nn/sturniolo)	PHIGINVNSIRSIKT	2.2	103	5.1	0.24/ YES	YES
HLA-DRB1*13:02	170	184	Consensus (smm/nn/sturniolo)	HIGINVNSIRSIKTT	2.4	108	5.5	0.37/ YES	YES
HLA-DRB1*13:02	171	185	Consensus (smm/nn/sturniolo)	IGINVNSIRSIKTTS	2.7	118	8.5	0.39/ YES	YES
HLA-DRB1*04:01	172	186	Consensus (smm/nn/sturniolo)	GINVNSIRSIKTTSW	1.6	95	47.7	0.38/ YES	YES
HLA-DRB1*08:02	172	186	Consensus (smm/nn/sturniolo)	GINVNSIRSIKTTSW	2.6	165	113.2	0.38/ YES	YES
HLA-DRB1*04:05	172	186	Consensus (smm/nn/sturniolo)	GINVNSIRSIKTTSW	4	183	84	0.38/ YES	YES

HLA-DRB1*04:01	173	187	Consensus (smm/nn/sturniolo)	INVNSIRSIKTTSWD	1.4	96	36.4	0.35/ YES	YES
HLA-DRB1*04:05	173	187	Consensus (smm/nn/sturniolo)	INVNSIRSIKTTSWD	2.3	181	58.1	0.35/ YES	YES
HLA-DRB1*08:02	173	187	Consensus (smm/nn/sturniolo)	INVNSIRSIKTTSWD	3.2	166	130. 3	0.35/ YES	YES
HLA-DRB1*04:01	174	188	Consensus (smm/nn/sturniolo)	NVNSIRSIKTTSWDL	1.8	108	33.2	0.36/ YES	YES
HLA-DRB1*04:05	174	188	Consensus (smm/nn/sturniolo)	NVNSIRSIKTTSWDL	3.5	202	75.8	0.36/ YES	YES
HLA-DRB1*04:01	175	189	Consensus (smm/nn/sturniolo)	VNSIRSIKTTSWDLA	1.6	103	34.8	0.33/ YES	
HLA-DRB1*08:02	175	189	Consensus (smm/nn/sturniolo)	VNSIRSIKTTSWDLA	3.2	161	128. 4	0.33/ YES	
HLA-DRB1*04:01	176	190	Consensus (smm/nn/sturniolo)	NSIRSIKTTSWDLAN	1.8	109	38.5	0.32/ YES	
HLA-DRB1*08:02	176	190	Consensus (smm/nn/sturniolo)	NSIRSIKTTSWDLAN	2.4	177	108. 4	0.32/ YES	
HLA-DRB1*13:02	182	196	Consensus (smm/nn/sturniolo)	KTTSWDLANNKVAK V	2.3	83	16.7	0.25/ YES	
HLA-DRB1*13:02	183	197	Consensus (smm/nn/sturniolo)	TTSWDLANNKVAKVL	1.6	80	11.9	0.25/ YES	
HLA-DRB1*13:02	184	198	Consensus (smm/nn/sturniolo)	TSWDLANNKVAKVLI	1.5	77	9.7	0.23/ YES	YES
HLA-DRB1*13:02	185	199	Consensus (smm/nn/sturniolo)	SWDLANNKVAKVLIT	1.4	76	10.4	0.22/ YES	YES
HLA-DRB1*13:02	186	200	Consensus (smm/nn/sturniolo)	WDLANNKVAKVLITY	1.8	79	13.8	0.25/ YES	YES
HLA-DRB1*07:01	194	208	Consensus (comb.lib./smm/nn)	AKVLITYDASTSLLV	0.78	8	10.5	-0.59	
HLA-DRB1*09:01	194	208	Consensus (comb.lib./smm/nn)	AKVLITYDASTSLLV	2.8	17	36.6	-0.59	
HLA-DRB3*01:01	194	208	Consensus (comb.lib./smm/nn)	AKVLITYDASTSLLV	2.8	247	7.3	-0.59	
HLA-DRB1*04:01	194	208	Consensus (smm/nn/sturniolo)	AKVLITYDASTSLLV	3	58	75.9	-0.59	
HLA-DRB1*03:01	194	208	Consensus (smm/nn/sturniolo)	AKVLITYDASTSLLV	3.6	161	82.6	-0.59	
HLA-DRB1*07:01	195	209	Consensus (comb.lib./smm/nn)	KVLITYDASTSLLVA	0.94	8	11.6	-0.8	
HLA-DRB1*04:01	195	209	Consensus (smm/nn/sturniolo)	KVLITYDASTSLLVA	2	57	55.7	-0.8	
HLA-DRB3*01:01	195	209	Consensus (comb.lib./smm/nn)	KVLITYDASTSLLVA	2.8	247	7.7	-0.8	
HLA-DRB1*09:01	195	209	Consensus (comb.lib./smm/nn)	KVLITYDASTSLLVA	2.9	17	37.6	-0.8	
HLA-DRB1*03:01	195	209	Consensus (smm/nn/sturniolo)	KVLITYDASTSLLVA	3.2	161	71.9	-0.8	
HLA-DRB1*07:01	196	210	Consensus (comb.lib./smm/nn)	VLITYDASTSLLVAS	1.3	8	14	-0.87	
HLA-DQA1*01:02/DQB1*06:02	196	210	Consensus (comb.lib./smm/nn)	VLITYDASTSLLVAS	1.9	183	78.6	-0.87	
HLA-DRB1*04:01	196	210	Consensus (smm/nn/sturniolo)	VLITYDASTSLLVAS	2.3	72	61.5	-0.87	
HLA-DRB1*09:01	196	210	Consensus (comb.lib./smm/nn)	VLITYDASTSLLVAS	3.3	17	40.1	-0.87	
HLA-DQA1*01:02/DQB1*06:02	197	211	Consensus (comb.lib./smm/nn)	LITYDASTSLLVASL	0.86	102	46.5	-0.83	
HLA-DRB1*07:01	197	211	Consensus (comb.lib./smm/nn)	LITYDASTSLLVASL	1.4	8	14.3	-0.83	
HLA-DRB1*04:01	197	211	Consensus (smm/nn/sturniolo)	LITYDASTSLLVASL	3	92	75.1	-0.83	
HLA-DRB1*01:01	197	211	Consensus (comb.lib./smm/nn)	LITYDASTSLLVASL	3.7	23	57.6	-0.83	
HLA-DRB1*09:01	197	211	Consensus (comb.lib./smm/nn)	LITYDASTSLLVASL	3.9	16	45.4	-0.83	
HLA-DQA1*01:02/DQB1*06:02	198	212	Consensus (comb.lib./smm/nn)	ITYDASTSLLVASLV	0.66	100	41.5	0.08	YES
HLA-DRB1*07:01	198	212	Consensus (comb.lib./smm/nn)	ITYDASTSLLVASLV	1.6	8	15.8	0.08	YES
HLA-DRB1*01:01	198	212	Consensus (comb.lib./smm/nn)	ITYDASTSLLVASLV	2.5	17	69.4	0.08	YES
HLA-DRB1*09:01	198	212	Consensus (comb.lib./smm/nn)	ITYDASTSLLVASLV	3.5	16	42.6	0.08	YES
HLA-DQA1*01:02/DQB1*06:02	199	213	Consensus (comb.lib./smm/nn)	TYDASTSLLVASLVY	0.69	103	42.3	0.1	
HLA-DRB1*07:01	199	213	Consensus (comb.lib./smm/nn)	TYDASTSLLVASLVY	1.9	25	18.1	0.1	
HLA-DQA1*01:02/DQB1*06:02	200	214	Consensus (comb.lib./smm/nn)	YDASTSLLVASLVYP	1.1	134	53.1	0.09	
HLA-DQA1*01:02/DQB1*06:02	201	215	Consensus (comb.lib./smm/nn)	DASTSLLVASLVYPS	1.3	115	62.1	0.06	YES
HLA-DRB1*03:01	215	229	Consensus (smm/nn/sturniolo)	SQRTSNILSDVVDLK	1.8	33	45.6	0.25/ YES	
HLA-DRB1*03:01	216	230	Consensus (smm/nn/sturniolo)	QRTSNILSDVVDLKT	0.65	32	21	0.18	
HLA-DRB1*03:01	217	231	Consensus (smm/nn/sturniolo)	RTSNILSDVVDLKTS	0.38	32	15.8	0.22/ YES	
HLA-DRB1*03:01	218	232	Consensus (smm/nn/sturniolo)	TSNILSDVVDLKTSL	0.33	30	14.8	0.22/ YES	
HLA-DRB1*03:01	219	233	Consensus (smm/nn/sturniolo)	SNILSDVVDLKTSLP	0.5	31	17.4	0.16	
HLA-DRB1*03:01	220	234	Consensus (smm/nn/sturniolo)	NILSDVVDLKTSLPE	1	90	29.2	0.17	YES

HLA-DQA1*05:01/DQB1*03:01	231	245	Consensus (comb.lib./smm/nn)	SLPEWVRIGFSAATG	4	45	47.8	0.19	YES
HLA-DQA1*05:01/DQB1*03:01	232	246	Consensus (comb.lib./smm/nn)	LPEWVRIGFSAATGL	3	35	46.6	0.19	
HLA-DRB1*07:01	232	246	Consensus (comb.lib./smm/nn)	LPEWVRIGFSAATGL	3	45	11.3	0.19	
HLA-DRB1*09:01	232	246	Consensus (comb.lib./smm/nn)	LPEWVRIGFSAATGL	3.6	138	35.4	0.19	
HLA-DQA1*05:01/DQB1*03:01	233	247	Consensus (comb.lib./smm/nn)	PEWVRIGFSAATGLD	3	34	51.5	0.10	
HLA-DRB1*07:01	233	247	Consensus (comb.lib./smm/nn)	PEWVRIGFSAATGLD	3.1	46	13	0.10	
HLA-DRB1*09:01	233	247	Consensus (comb.lib./smm/nn)	PEWVRIGFSAATGLD	3.5	137	31.5	0.10	
HLA-DRB1*07:01	234	248	Consensus (comb.lib./smm/nn)	EWVRIGFSAATGLDI	0.64	16	4.7	1.05/ YES	YES
HLA-DRB1*09:01	234	248	Consensus (comb.lib./smm/nn)	EWVRIGFSAATGLDI	1.5	83	8.5	1.05/ YES	YES
HLA-DQA1*05:01/DQB1*03:01	234	248	Consensus (comb.lib./smm/nn)	EWVRIGFSAATGLDI	3.8	33	62.9	1.05/ YES	YES
HLA-DRB1*07:01	235	249	Consensus (comb.lib./smm/nn)	WVRIGFSAATGLDIP	0.64	16	5.1	1.05/ YES	
HLA-DRB1*09:01	235	249	Consensus (comb.lib./smm/nn)	WVRIGFSAATGLDIP	1.6	87	6.8	1.05/ YES	
HLA-DRB1*07:01	236	250	Consensus (comb.lib./smm/nn)	VRIGFSAATGLDIPG	0.64	16	5.9	0.98/ YES	YES
HLA-DRB1*09:01	236	250	Consensus (comb.lib./smm/nn)	VRIGFSAATGLDIPG	1.6	87	6.6	0.98/ YES	YES
HLA-DRB1*07:01	237	251	Consensus (comb.lib./smm/nn)	RIGFSAATGLDIPGE	1.1	21	6.8	0.97/ YES	
HLA-DRB1*09:01	237	251	Consensus (comb.lib./smm/nn)	RIGFSAATGLDIPGE	2.8	118	6.9	0.97/ YES	
HLA-DRB1*07:01	238	252	Consensus (comb.lib./smm/nn)	IGFSAATGLDIPGES	1.2	22	11.2	-0.07	
HLA-DRB1*09:01	238	252	Consensus (comb.lib./smm/nn)	IGFSAATGLDIPGES	2.9	122	9.3	-0.07	
HLA-DRB1*09:01	250	264	Consensus (comb.lib./smm/nn)	GESHDVLSWSFASNL	3.7	142	26.4	0.20	YES
HLA-DRB1*09:01	251	265	Consensus (comb.lib./smm/nn)	ESHDVLSWSFASNLP	4	147	28.2	0.20	YES
HLA-DRB1*09:01	252	266	Consensus (comb.lib./smm/nn)	SHDVLSWSFASNLP	3.9	146	26.9	0.14	YES
HLA-DRB1*09:01	253	267	Consensus (comb.lib./smm/nn)	HDVLSWSFASNLP	1.9	145	26.9	1.14/ YES	YES
HLA-DRB1*09:01	254	268	Consensus (comb.lib./smm/nn)	DVLSWSFASNLP	2	110	27.6	1.10/ YES	YES
HLA-DRB1*09:01	255	269	Consensus (comb.lib./smm/nn)	VLSWSFASNLP	2.5	227	34.1	1.08/ YES	YES

Table S9. List of epitopes of P05046 predicted by NetMHCIIpan method along with interacting HLA class II alleles and relevant scores.

Allele	Start	End	Method	Peptide	Percentile rank	NetMHCIIpan IC50	IL-4 pred	AllerTOP v. 2.0
HLA-DRB3*02:02	20	34	NetMHCIIpan	TLVLVLLTSKANS	3.2	182.8	-0.86	YES
HLA-DRB3*02:02	21	35	NetMHCIIpan	LVLVLLTSKANS	2.3	137.06	-0.79	YES
HLA-DRB3*02:02	22	36	NetMHCIIpan	VLVLLTSKANS	2	113.73	-1.00	YES
HLA-DRB3*02:02	23	37	NetMHCIIpan	LVLLTSKANS	2.3	138.44	0.97/YES	YES
HLA-DPA1*01:03/DPB1*04:01	32	46	NetMHCIIpan	SAETVSFSWNKFVPK	3.1	174.95	0.29/YES	
HLA-DPA1*01:03/DPB1*04:01	33	47	NetMHCIIpan	AETVSFSWNKFVPKQ	2.9	168.5	0.29/YES	
HLA-DPA1*01:03/DPB1*04:01	34	48	NetMHCIIpan	ETVSFSWNKFVPKQP	3.3	183.6	0.29/YES	YES
HLA-DRB3*02:02	38	52	NetMHCIIpan	FSWNKFVPKQPNMIL	3.6	213.29	0.29/YES	YES
HLA-DRB3*02:02	39	53	NetMHCIIpan	SWNKFVPKQPNMILQ	3.2	183.3	0.29/YES	
HLA-DRB3*02:02	40	54	NetMHCIIpan	WNKFVPKQPNMILQG	3.3	196.69	0.29/YES	YES
HLA-DPA1*01:03/DPB1*04:01	96	110	NetMHCIIpan	TGSVASFAASFNFTF	3.7	197.33	0.08	
HLA-DPA1*01:03/DPB1*04:01	97	111	NetMHCIIpan	GSVASFAASFNFTFY	1.3	104.27	0.06	
HLA-DPA1*01:03/DPB1*04:01	98	112	NetMHCIIpan	SVASFAASFNFTFYA	0.68	76.1	0.27/YES	
HLA-DPA1*01:03/DPB1*04:01	99	113	NetMHCIIpan	VASFAASFNFTFYAP	0.43	59.9	0.18	YES
HLA-DPA1*01:03/DPB1*04:01	100	114	NetMHCIIpan	ASFAASFNFTFYAPD	0.32	52.88	0.22/YES	YES
HLA-DPA1*01:03/DPB1*04:01	101	115	NetMHCIIpan	SFAASFNFTFYAPDT	0.37	58.04	0.21/YES	
HLA-DPA1*01:03/DPB1*04:01	102	116	NetMHCIIpan	FAASFNFTFYAPDTK	0.65	74.89	0.21/YES	
HLA-DPA1*01:03/DPB1*04:01	103	117	NetMHCIIpan	AASFNFTFYAPDTKR	1.6	116.51	-0.22	YES
HLA-DPA1*01:03/DPB1*04:01	104	118	NetMHCIIpan	ASFNFTFYAPDTKRL	2.9	170.12	0.75/YES	
HLA-DRB3*02:02	167	181	NetMHCIIpan	PNPHIGINVNSIRSI	1.9	105.32	0.17	YES
HLA-DRB3*02:02	168	182	NetMHCIIpan	NPHIGINVNSIRSIK	1.4	77.04	0.24/YES	YES
HLA-DRB3*02:02	169	183	NetMHCIIpan	PHIGINVNSIRSIKT	1.4	77.75	0.24/YES	YES
HLA-DRB3*02:02	170	184	NetMHCIIpan	HIGINVNSIRSIKTT	1.6	90.51	0.37/YES	YES

HLA-DRB3*02:02	171	185	NetMHCIIpan	IGINVNSIRSIKTTTS	3.1	178.55	0.39/YES	YES
HLA-DRB3*02:02	182	196	NetMHCIIpan	KTTSWDLANNKVAKV	3.5	207.39	0.25/YES	
HLA-DRB3*02:02	183	197	NetMHCIIpan	TTSWDLANNKVAKVL	1.6	88.06	0.25/YES	
HLA-DRB3*02:02	184	198	NetMHCIIpan	TSWDLANNKVAKVLI	1.1	63.31	0.23/YES	YES
HLA-DRB3*02:02	185	199	NetMHCIIpan	SWDLANNKVAKVLIT	1.1	63.74	0.22/YES	YES
HLA-DRB3*02:02	186	200	NetMHCIIpan	WDLANNKVAKVLITY	1.5	85.64	0.25/YES	YES
HLA-DRB3*02:02	187	201	NetMHCIIpan	DLANNKVAKVLITYD	3.1	177.28	0.11	
HLA-DRB3*02:02	194	208	NetMHCIIpan	AKVLITYDASTSLLV	4	235.49	-0.59	
HLA-DRB3*02:02	195	209	NetMHCIIpan	KVLITYDASTSLLVA	2.6	151.11	-0.8	
HLA-DRB3*02:02	196	210	NetMHCIIpan	VLITYDASTSLLVAS	2.9	166.05	-0.87	
HLA-DRB3*02:02	197	211	NetMHCIIpan	LITYDASTSLLVASL	3.3	193.73	-0.83	
HLA-DRB3*02:02	254	268	NetMHCIIpan	DVLSWSFASNLPHAS	1.1	61.45	1.1/YES	YES
HLA-DRB3*02:02	255	269	NetMHCIIpan	VLSWSFASNLPHASS	0.42	33.15	1.08/YES	YES
HLA-DRB3*02:02	256	270	NetMHCIIpan	LSWSFASNLPHASSN	0.35	30.8	1.13/YES	
HLA-DRB3*02:02	257	271	NetMHCIIpan	SWSFASNLPHASSNI	0.32	29.12	0.13	
HLA-DRB3*02:02	258	272	NetMHCIIpan	WSFASNLPHASSNID	0.66	40.59	0.17	
HLA-DRB3*02:02	259	273	NetMHCIIpan	SFASNLPHASSNIDP	1.6	88.37	0.10	

Table S10. List of epitopes of P11827 predicted by consensus method along with interacting HLA class II alleles and relevant scores.

Allele	Start	End	Method	Peptide	Percentile rank	Smm IC50	Nn IC50	IL-4 pred	AllerTOP v. 2.0
HLA-DPA1*01:03/DPB1*02:01	5	19	Consensus (comb.lib./smm/nn)	RFPLLLLGVVFLASV	3.7	47	86.9	-1	
HLA-DRB1*07:01	9	23	Consensus (comb.lib./smm/nn)	LLLGVVFLASVSVSF	2.7	9	23.8	-1.21	
HLA-DRB3*01:01	9	23	Consensus (comb.lib./smm/nn)	LLLGVVFLASVSVSF	2.8	154	66.2	-1.21	
HLA-DQA1*05:01/DQB1*03:01	9	23	Consensus (comb.lib./smm/nn)	LLLGVVFLASVSVSF	3.3	32	77.6	-1.21	
HLA-DRB1*07:01	10	24	Consensus (comb.lib./smm/nn)	LLGVVFLASVSVSFG	2.7	9	35.1	-0.51	YES
HLA-DRB3*01:01	10	24	Consensus (comb.lib./smm/nn)	LLGVVFLASVSVSFG	3.1	158	75.7	-0.51	YES
HLA-DRB1*07:01	11	25	Consensus (comb.lib./smm/nn)	LGVVFLASVSVSFGI	1.5	9	15.2	-0.22	YES
HLA-DRB3*01:01	11	25	Consensus (comb.lib./smm/nn)	LGVVFLASVSVSFGI	2.5	153	57.1	-0.22	YES
HLA-DRB1*04:01	11	25	Consensus (smm/nn/sturniolo)	LGVVFLASVSVSFGI	3.1	49	77.7	-0.22	YES
HLA-DRB1*07:01	12	26	Consensus (comb.lib./smm/nn)	GVVFLASVSVSFGIA	1.8	9	17.8	-0.29	
HLA-DRB1*04:01	12	26	Consensus (smm/nn/sturniolo)	GVVFLASVSVSFGIA	2.2	49	59.4	-0.29	
HLA-DRB3*01:01	12	26	Consensus (comb.lib./smm/nn)	GVVFLASVSVSFGIA	2.6	154	61.3	-0.29	
HLA-DRB1*07:01	13	27	Consensus (comb.lib./smm/nn)	VVFLASVSVSFGIAY	2	9	18.6	-0.3	
HLA-DRB1*09:01	13	27	Consensus (comb.lib./smm/nn)	VVFLASVSVSFGIAY	2.7	91	36	-0.3	
HLA-DRB1*04:01	13	27	Consensus (smm/nn/sturniolo)	VVFLASVSVSFGIAY	2.8	52	71.7	-0.3	
HLA-DRB3*01:01	13	27	Consensus (comb.lib./smm/nn)	VVFLASVSVSFGIAY	3.8	164	95.4	-0.3	
HLA-DRB1*07:01	14	28	Consensus (comb.lib./smm/nn)	VFLASVSVSFGIAYW	2.2	27	19.9	-0.21	
HLA-DRB1*04:01	14	28	Consensus (smm/nn/sturniolo)	VFLASVSVSFGIAYW	3.3	140	82.4	-0.21	
HLA-DRB1*07:01	15	29	Consensus (comb.lib./smm/nn)	FLASVSVSFGIAYWE	2.7	28	24.5	0	
HLA-DRB1*11:01	206	220	Consensus (smm/nn/sturniolo)	HKNKNPFHFNSKRFAQ	3.7	12	78.8	1.38/ YES	
HLA-DRB1*11:01	207	221	Consensus (smm/nn/sturniolo)	KNKNPFHFNSKRFAQT	3.7	12	37.3	0.38/ YES	
HLA-DRB1*11:01	208	222	Consensus (smm/nn/sturniolo)	NKNPFHFNSKRFAQTL	2.9	11	27.9	0.4/YES	
HLA-DRB1*11:01	209	223	Consensus (smm/nn/sturniolo)	KNPFHFNSKRFAQTLF	2.3	11	22.7	0.46/ YES	
HLA-DRB1*11:01	210	224	Consensus (smm/nn/sturniolo)	NPFHFNSKRFAQTLFK	2.2	11	21.8	0.33/ YES	
HLA-DPA1*03:01/DPB1*04:02	210	224	Consensus (comb.lib./smm/nn)	NPFHFNSKRFAQTLFK	3	112	72.9	0.33/ YES	
HLA-DPA1*03:01/DPB1*04:02	211	225	Consensus (comb.lib./smm/nn)	PFHFNSKRFAQTLFKN	2.4	112	57.3	1.34/ YES	YES
HLA-DRB1*11:01	211	225	Consensus (smm/nn/sturniolo)	PFHFNSKRFAQTLFKN	2.5	34	25.3	1.34/ YES	YES
HLA-DPA1*03:01/DPB1*04:02	212	226	Consensus (comb.lib./smm/nn)	FHFNSKRFAQTLFKNQ	3	114	73.2	1.34/ YES	YES
HLA-DRB1*11:01	212	226	Consensus (smm/nn/sturniolo)	FHFNSKRFAQTLFKNQ	3.7	34	43.9	1.34/ YES	YES
HLA-DRB5*01:01	217	231	Consensus (smm/nn/sturniolo)	KRFQTLFKNQYGHVR	1.8	25	13.2	0.33/ YES	
HLA-DRB5*01:01	218	232	Consensus (smm/nn/sturniolo)	RFQTLFKNQYGHVRV	0.88	24	8.6	0.35/ YES	
HLA-DRB5*01:01	219	233	Consensus (smm/nn/sturniolo)	FQTLFKNQYGHVRVL	0.81	22	7.3	0.36/ YES	
HLA-DRB5*01:01	220	234	Consensus (smm/nn/sturniolo)	QTLFKNQYGHVRVLQ	0.81	22	6.8	0.36/ YES	
HLA-DRB5*01:01	221	235	Consensus (smm/nn/sturniolo)	TLFKNQYGHVRVLQR	0.6	19	5.9	0.35/ YES	YES
HLA-DRB5*01:01	222	236	Consensus (smm/nn/sturniolo)	LFKNQYGHVRVLQRF	1.7	40	6.6	0.34/ YES	YES
HLA-DRB5*01:01	223	237	Consensus (smm/nn/sturniolo)	FKNQYGHVRVLQRFN	1.8	42	12.1	0.33/ YES	YES
HLA-DRB5*01:01	224	238	Consensus (smm/nn/sturniolo)	KNQYGHVRVLQRFNK	3.5	99	22.4	1.47/ YES	
HLA-DRB5*01:01	225	239	Consensus (smm/nn/sturniolo)	NQYGHVRVLQRFNKR	3.2	100	20.6	1.48/ YES	YES
HLA-DRB1*11:01	225	239	Consensus (smm/nn/sturniolo)	NQYGHVRVLQRFNKR	3.4	126	23.1	1.48/ YES	YES
HLA-DRB1*11:01	226	240	Consensus (smm/nn/sturniolo)	QYGHVRVLQRFNKRS	3.4	132	16	1.48/ YES	YES

									YES	
HLA-DRB5*01:01	226	240	Consensus (smm/nn/sturniolo)	QYGHVRVLQRFNKRS	3.4	230	21.9	1.48/ YES	YES	
HLA-DRB1*11:01	227	241	Consensus (smm/nn/sturniolo)	YGHVRVLQRFNKRSQ	2.1	44	12.6	1.55/ YES		
HLA-DRB1*11:01	228	242	Consensus (smm/nn/sturniolo)	GHVRVLQRFNKRSQQ	2.1	44	11.9	1.51/ YES		
HLA-DRB1*11:01	229	243	Consensus (smm/nn/sturniolo)	HVRVLQRFNKRSQQL	2.1	45	18	1.52/ YES		
HLA-DRB1*11:01	230	244	Consensus (smm/nn/sturniolo)	VRVLQRFNKRSQQLQ	3.4	45	39.4	0.52/ YES		
HLA-DRB1*04:01	248	262	Consensus (smm/nn/sturniolo)	DYRILEFN SKPNTLL	3.9	167	138.8	1.31/ YES	YES	
HLA-DRB1*04:01	249	263	Consensus (smm/nn/sturniolo)	YRILEFN SKPNTLLL	3.4	163	84.3	1.23/ YES	YES	
HLA-DRB5*01:01	249	263	Consensus (smm/nn/sturniolo)	YRILEFN SKPNTLLL	3.9	71	137.3	1.23/ YES	YES	
HLA-DRB3*01:01	261	275	Consensus (comb.lib./smm/nn)	LLPHHADADYLIVI	2.8	250	125	0.26/ YES	YES	
HLA-DRB3*01:01	262	276	Consensus (comb.lib./smm/nn)	LLPHHADADYLIVIL	2.7	243	90.6	0.22/ YES		
HLA-DRB1*01:01	267	281	Consensus (comb.lib./smm/nn)	ADADYLIVILNGTAI	2	14	75.2	-1.39		
HLA-DRB1*13:02	267	281	Consensus (smm/nn/sturniolo)	ADADYLIVILNGTAI	2.2	101	12.1	-1.39		
HLA-DRB1*13:02	268	282	Consensus (smm/nn/sturniolo)	DADYLIVILNGTAIL	0.39	33	4.2	-1.41		
HLA-DRB1*01:01	268	282	Consensus (comb.lib./smm/nn)	DADYLIVILNGTAIL	1.6	12	43.3	-1.41		
HLA-DRB1*15:01	268	282	Consensus (smm/nn/sturniolo)	DADYLIVILNGTAIL	1.6	78	137.5	-1.41		
HLA-DRB1*13:02	269	283	Consensus (smm/nn/sturniolo)	ADYLIVILNGTAILT	0.39	33	4.3	-1.35		
HLA-DRB1*01:01	269	283	Consensus (comb.lib./smm/nn)	ADYLIVILNGTAILT	1.1	10	25.5	-1.35		
HLA-DRB1*15:01	269	283	Consensus (smm/nn/sturniolo)	ADYLIVILNGTAILT	1.5	77	97.6	-1.35		
HLA-DRB1*13:02	270	284	Consensus (smm/nn/sturniolo)	DYLIVILNGTAILTL	0.35	31	4.2	-1.38		
HLA-DRB1*01:01	270	284	Consensus (comb.lib./smm/nn)	DYLIVILNGTAILTL	0.91	9	20.2	-1.38		
HLA-DRB1*15:01	270	284	Consensus (smm/nn/sturniolo)	DYLIVILNGTAILTL	1.5	76	81.5	-1.38		
HLA-DPA1*03:01/DPB1*04:02	270	284	Consensus (comb.lib./smm/nn)	DYLIVILNGTAILTL	3.3	41	78.1	-1.38		
HLA-DRB1*13:02	271	285	Consensus (smm/nn/sturniolo)	YLIVILNGTAILTLV	0.35	31	4.2	-1.25		
HLA-DRB1*01:01	271	285	Consensus (comb.lib./smm/nn)	YLIVILNGTAILTLV	0.91	9	20.2	-1.25		
HLA-DPA1*03:01/DPB1*04:02	271	285	Consensus (comb.lib./smm/nn)	YLIVILNGTAILTLV	2.2	40	59	-1.25		
HLA-DRB1*15:01	271	285	Consensus (smm/nn/sturniolo)	YLIVILNGTAILTLV	2.5	97	82.8	-1.25		
HLA-DRB1*13:02	272	286	Consensus (smm/nn/sturniolo)	LIVILNGTAILTLVN	0.46	34	5.5	-1.19	YES	
HLA-DPA1*03:01/DPB1*04:02	272	286	Consensus (comb.lib./smm/nn)	LIVILNGTAILTLVN	2.3	40	60.2	-1.19	YES	
HLA-DRB1*01:01	272	286	Consensus (comb.lib./smm/nn)	LIVILNGTAILTLVN	3.3	21	24.6	-1.19	YES	
HLA-DRB1*13:02	273	287	Consensus (smm/nn/sturniolo)	IVILNGTAILTLVNN	1.7	86	8.9	-0.14	YES	
HLA-DPA1*03:01/DPB1*04:02	273	287	Consensus (comb.lib./smm/nn)	IVILNGTAILTLVNN	2.6	40	66.2	-0.14	YES	
HLA-DQA1*01:02/DQB1*06:02	273	287	Consensus (comb.lib./smm/nn)	IVILNGTAILTLVNN	2.8	67	215.3	-0.14	YES	
HLA-DRB1*01:01	273	287	Consensus (comb.lib./smm/nn)	IVILNGTAILTLVNN	3.7	23	37	-0.14	YES	
HLA-DQA1*01:02/DQB1*06:02	274	288	Consensus (comb.lib./smm/nn)	VILNGTAILTLVNND	2.8	70	211	-0.12	YES	
HLA-DPA1*03:01/DPB1*04:02	274	288	Consensus (comb.lib./smm/nn)	VILNGTAILTLVNND	3.6	42	122	-0.12	YES	
HLA-DRB1*04:05	275	289	Consensus (smm/nn/sturniolo)	ILNGTAILTLVNDD	2.1	63	54.5	0.25/ YES	YES	
HLA-DRB1*04:05	276	290	Consensus (smm/nn/sturniolo)	LNGTAILTLVNDDDR	1.1	63	26.5	0.25/ YES	YES	
HLA-DRB1*04:05	277	291	Consensus (smm/nn/sturniolo)	NGTAILTLVNDDDRD	1.1	62	18.2	0.26/ YES	YES	
HLA-DRB1*04:05	278	292	Consensus (smm/nn/sturniolo)	GTAILTLVNDDDRDS	1.1	62	17.9	0.1		
HLA-DRB1*04:05	279	293	Consensus (smm/nn/sturniolo)	TAILTLVNDDDRDSY	1.1	64	22.1	0.19		
HLA-DRB1*04:05	280	294	Consensus (smm/nn/sturniolo)	AILTLVNDDDRDSYN	2.3	189	48.3	0.21/ YES		
HLA-DPA1*01:03/DPB1*02:01	328	342	Consensus (comb.lib./smm/nn)	PVNKPGRFESFFLSS	2.9	113	87.3	0.39/ YES		
HLA-DPA1*01:03/DPB1*02:01	329	343	Consensus (comb.lib./smm/nn)	VNKPGRFESFFLSST	2.2	110	54.4	0.43/ YES		
HLA-DPA1*02:01/DPB1*01:01	329	343	Consensus (comb.lib./smm/nn)	VNKPGRFESFFLSST	3.4	123	67.7	0.43/ YES		
HLA-DPA1*01:03/DPB1*02:01	330	344	Consensus (comb.lib./smm/nn)	NKPGRFESFFLSSTQ	2.1	110	52.7	0.39/ YES		

HLA-DPA1*02:01/DPB1*01:01	330	344	Consensus (comb.lib./smm/nn)	NKPGRFESFFLSSTQ	3.4	122	53.1	0.39/ YES	
HLA-DPA1*01:03/DPB1*02:01	331	345	Consensus (comb.lib./smm/nn)	KPGRFESFFLSSTQA	2.1	109	52.8	0.38/ YES	
HLA-DPA1*02:01/DPB1*01:01	331	345	Consensus (comb.lib./smm/nn)	KPGRFESFFLSSTQA	3.4	120	47.5	0.38/ YES	
HLA-DRB1*04:05	331	345	Consensus (smm/nn/sturniolo)	KPGRFESFFLSSTQA	3.8	99	81.2	0.38/ YES	
HLA-DPA1*01:03/DPB1*02:01	332	346	Consensus (comb.lib./smm/nn)	PGRFESFFLSSTQAAQ	2	177	50.7	0.36/ YES	
HLA-DRB1*04:05	332	346	Consensus (smm/nn/sturniolo)	PGRFESFFLSSTQAAQ	2.1	94	37.9	0.36/ YES	
HLA-DRB1*04:01	332	346	Consensus (smm/nn/sturniolo)	PGRFESFFLSSTQAAQ	2.6	146	67.2	0.36/ YES	
HLA-DRB1*04:01	333	347	Consensus (smm/nn/sturniolo)	GRFESFFLSSTQAQQ	1.4	110	44.7	0.33/ YES	
HLA-DRB1*04:05	333	347	Consensus (smm/nn/sturniolo)	GRFESFFLSSTQAQQ	2.1	91	23.1	0.33/ YES	
HLA-DPA1*01:03/DPB1*02:01	333	347	Consensus (comb.lib./smm/nn)	GRFESFFLSSTQAQQ	3.6	177	84	0.33/ YES	
HLA-DPA1*02:01/DPB1*01:01	333	347	Consensus (comb.lib./smm/nn)	GRFESFFLSSTQAQQ	3.8	133	63.8	0.33/ YES	
HLA-DRB1*04:01	334	348	Consensus (smm/nn/sturniolo)	RFESFFLSSTQAQQS	1.1	117	37.6	0.30/ YES	
HLA-DRB1*04:05	334	348	Consensus (smm/nn/sturniolo)	RFESFFLSSTQAQQS	2.1	184	17.9	0.30/ YES	
HLA-DRB1*04:01	335	349	Consensus (smm/nn/sturniolo)	FESFFLSSTQAQQSY	1.1	123	37.1	0.28/ YES	
HLA-DRB1*04:05	335	349	Consensus (smm/nn/sturniolo)	FESFFLSSTQAQQSY	2.1	193	16.6	0.28/ YES	
HLA-DRB1*04:01	336	350	Consensus (smm/nn/sturniolo)	ESFFLSSTQAQQSYL	1.2	122	40	-0.16	
HLA-DRB1*04:01	337	351	Consensus (smm/nn/sturniolo)	SFFLSSTQAQQSYLQ	1.6	199	48	-0.4	
HLA-DRB1*04:01	461	475	Consensus (smm/nn/sturniolo)	MNEGALFLPHFNSKA	2.9	244	38.9	0.28/ YES	
HLA-DRB1*04:01	462	476	Consensus (smm/nn/sturniolo)	NEGALFLPHFNSKAI	2.9	213	23.3	0.28/ YES	
HLA-DRB1*04:01	463	477	Consensus (smm/nn/sturniolo)	EGALFLPHFNSKAIV	2.9	206	19.4	0.29/ YES	
HLA-DRB1*04:01	464	478	Consensus (smm/nn/sturniolo)	GALFLPHFNSKAIVV	2.9	199	18.2	0.29/ YES	
HLA-DRB1*04:01	465	479	Consensus (smm/nn/sturniolo)	ALFLPHFNSKAIVVL	2.9	201	24.9	0.28/ YES	
HLA-DQA1*05:01/DQB1*03:01	519	533	Consensus (comb.lib./smm/nn)	QDIFVIPAGYPVVVN	3.2	33	54.7	0.21/ YES	
HLA-DQA1*05:01/DQB1*03:01	520	534	Consensus (comb.lib./smm/nn)	DIFVIPAGYPVVVNA	3.1	31	53.2	0.26/ YES	YES
HLA-DQA1*05:01/DQB1*03:01	521	535	Consensus (comb.lib./smm/nn)	IFVIPAGYPVVVNAT	2.8	33	40.6	0.28/ YES	YES
HLA-DQA1*05:01/DQB1*03:01	522	536	Consensus (comb.lib./smm/nn)	FVIPAGYPVVVNATS	3.8	36	62	0.3/YES	
HLA-DRB1*13:02	524	538	Consensus (smm/nn/sturniolo)	IPAGYPVVVNATSDL	3.9	161	9.7	0.25/ YES	
HLA-DRB1*13:02	527	541	Consensus (smm/nn/sturniolo)	GYPVVVNATSDLNFF	4	165	6.1	0.10	YES
HLA-DRB4*01:01	558	572	Consensus (comb.lib./smm/nn)	SKDNVISQIPSQVQE	3.3	130	62.2	0.40/ YES	YES
HLA-DRB4*01:01	559	573	Consensus (comb.lib./smm/nn)	KDNVISQIPSQVQEL	2.8	141	49.2	0.46/ YES	
HLA-DRB4*01:01	560	574	Consensus (comb.lib./smm/nn)	DNVISQIPSQVQELA	3.1	109	58.4	0.42/ YES	
HLA-DRB4*01:01	561	575	Consensus (comb.lib./smm/nn)	NVISQIPSQVQELAF	3.3	149	61.4	0.49/ YES	

Table S11. List of epitopes of P11827 predicted by NetMHCIIpan method along with interacting HLA class II alleles and relevant scores.

Allele	Star	End	Method	Peptide	Percentile rank	NetMHCIIpan	IL-4 pred	AllerTOP
--------	------	-----	--------	---------	-----------------	-------------	-----------	----------

t			IC50			v. 2.0	
HLA-DPA1*01:03/DPB1*04:01	1	15	NetMHCIIpan	MMRARFPLLLLGVVVF	3.5	192.15	-0.94
HLA-DRB3*02:02	10	24	NetMHCIIpan	LLGVVFLASVSVSFG	2.8	161.97	-0.51 YES
HLA-DRB3*02:02	11	25	NetMHCIIpan	LGVVFLASVSVSFGI	2.2	126.63	-0.22 YES
HLA-DRB3*02:02	12	26	NetMHCIIpan	GVVFLASVSVSFGIA	2	109.89	-0.29
HLA-DRB3*02:02	13	27	NetMHCIIpan	VVFLASVSVSFGIAY	2.9	164.58	-0.3
HLA-DPA1*01:03/DPB1*04:01	209	223	NetMHCIIpan	KNPFHFNSKRFQTLF	2.7	157.1	0.46/YES
HLA-DPA1*01:03/DPB1*04:01	210	224	NetMHCIIpan	NPFHFNSKRFQTLFK	2.3	139.9	0.33/YES
HLA-DPA1*01:03/DPB1*04:01	211	225	NetMHCIIpan	PFHFNSKRFQTLFKN	2.1	135.19	1.34/YES YES
HLA-DPA1*01:03/DPB1*04:01	212	226	NetMHCIIpan	FHFNSKRFQTLFKNQ	2.4	142.48	1.34/YES YES
HLA-DPA1*01:03/DPB1*04:01	213	227	NetMHCIIpan	HFNSKRFQTLFKNQY	3	171.1	1.32/YES YES
HLA-DRB3*02:02	249	263	NetMHCIIpan	YRILEFN SKPNTLLL	2.9	165.03	1.23/YES YES
HLA-DRB3*02:02	250	264	NetMHCIIpan	RILEFN SKPNTLLLP	3.7	216.1	0.19
HLA-DRB3*02:02	268	282	NetMHCIIpan	DADYLIVILNGTAIL	1.5	84.63	-1.41
HLA-DRB3*02:02	269	283	NetMHCIIpan	ADYLIVILNGTAILT	0.77	47.08	-1.35
HLA-DRB3*02:02	270	284	NetMHCIIpan	DYLIVILNGTAILTL	0.66	40.29	-1.38
HLA-DRB3*02:02	271	285	NetMHCIIpan	YLIVILNGTAILTLV	0.65	39.85	-1.25
HLA-DRB3*02:02	272	286	NetMHCIIpan	LIVILNGTAILTLVN	0.85	51.11	-1.19 YES
HLA-DRB3*02:02	273	287	NetMHCIIpan	IVILNGTAILTLVNN	1.2	72.53	-0.14 YES
HLA-DPA1*01:03/DPB1*04:01	328	342	NetMHCIIpan	PVNKPGRFESFFLSS	2.7	155.67	0.39/YES
HLA-DPA1*01:03/DPB1*04:01	329	343	NetMHCIIpan	VNKPGRFESFFLSST	0.95	91.11	0.43/YES
HLA-DPA1*01:03/DPB1*04:01	330	344	NetMHCIIpan	NKPGRFESFFLSSTQ	0.91	87.81	0.39/YES
HLA-DPA1*01:03/DPB1*04:01	331	345	NetMHCIIpan	KPGRFESFFLSSTQA	0.75	80.36	0.38/YES
HLA-DPA1*01:03/DPB1*04:01	332	346	NetMHCIIpan	PGRFESFFLSSTQAQ	1.2	99.42	0.36/YES
HLA-DPA1*01:03/DPB1*04:01	333	347	NetMHCIIpan	GRFESFFLSSTQAQQ	1.7	122.42	0.33/YES
HLA-DRB3*02:02	333	347	NetMHCIIpan	GRFESFFLSSTQAQQ	3.2	189.26	0.33/YES
HLA-DRB3*02:02	334	348	NetMHCIIpan	RFESFFLSSTQAQQS	2.2	131.49	0.3/YES
HLA-DPA1*01:03/DPB1*04:01	334	348	NetMHCIIpan	RFESFFLSSTQAQQS	3	171.27	0.3/YES
HLA-DRB3*02:02	335	349	NetMHCIIpan	FESFFLSSTQAQQSY	2.1	114.64	0.28/YES
HLA-DRB3*02:02	336	350	NetMHCIIpan	ESFFLSSTQAQQSYL	2.4	142	-0.16
HLA-DRB3*02:02	337	351	NetMHCIIpan	SFFLSSTQAQQSYLQ	3.2	189.73	-0.40
HLA-DPA1*01:03/DPB1*04:01	445	459	NetMHCIIpan	KNPQLRDL DVFLSVV	3.4	188.52	-0.38

HLA-DPA1*01:03/DPB1*04:01	447	461	NetMHCIIpan	PQLRDLDVFLSVVDM	3.8	204.8	-0.06	
HLA-DRB3*02:02	463	477	NetMHCIIpan	EGALFLPHFNSKAIV	2.5	146.05	0.29/YES	
HLA-DRB3*02:02	464	478	NetMHCIIpan	GALFLPHFNSKAIVV	1.9	107.24	0.29/YES	
HLA-DRB3*02:02	465	479	NetMHCIIpan	ALFLPHFNSKAIVVL	2.1	118.53	0.28/YES	
HLA-DRB3*02:02	466	480	NetMHCIIpan	LFLPHFNSKAIVVLV	2.7	154.24	0.28/YES	
HLA-DRB3*02:02	474	488	NetMHCIIpan	KAIVVLVINEGEANI	2.7	158.84	-1.35	
HLA-DRB3*02:02	475	489	NetMHCIIpan	AIVVLVINEGEANIE	2.5	143.29	-0.63	
HLA-DRB3*02:02	476	490	NetMHCIIpan	IVVLVINEGEANIEL	2	113.84	-0.65	YES
HLA-DRB3*02:02	477	491	NetMHCIIpan	VVLVINEGEANIELV	2.5	147.26	-0.68	YES
HLA-DRB3*02:02	527	541	NetMHCIIpan	GYPVVVNATSDLNFF	3.6	215.04	0.10	YES
HLA-DRB3*02:02	540	554	NetMHCIIpan	FFAFGINAENNQRNF	3.3	199.33	0.14	YES

Table S12. List of epitopes of P25974 predicted by consensus method along with interacting HLA class II alleles and relevant scores.

Allele	Start	End	Method	Peptide	Percentile rank	Smm IC50	Nn IC50	IL-4 pred	AllerTOP v. 2.0
HLA-DPA1*03:01/DPB1*04:02	5	19	Consensus (comb.lib./smm/nn)	RFPLLVLGTVFLAS	2.3	49	59.4	-1.14	YES
HLA-DPA1*01:03/DPB1*02:01	5	19	Consensus (comb.lib./smm/nn)	RFPLLVLGTVFLAS	3.6	24	85.3	-1.14	YES
HLA-DPA1*03:01/DPB1*04:02	6	20	Consensus (comb.lib./smm/nn)	FPLLVLGTVFLASV	0.77	45	32.5	-1.13	
HLA-DPA1*01:03/DPB1*02:01	6	20	Consensus (comb.lib./smm/nn)	FPLLVLGTVFLASV	2.6	23	62.2	-1.13	
HLA-DPA1*03:01/DPB1*04:02	7	21	Consensus (comb.lib./smm/nn)	PLLVLGTVFLASVC	0.81	46	33.5	-0.81	YES
HLA-DPA1*01:03/DPB1*02:01	7	21	Consensus (comb.lib./smm/nn)	PLLVLGTVFLASVC	2.9	24	68.9	-0.81	YES
HLA-DPA1*03:01/DPB1*04:02	8	22	Consensus (comb.lib./smm/nn)	LLVLGTVFLASVCV	0.77	45	30.2	-0.79	YES
HLA-DPA1*01:03/DPB1*02:01	8	22	Consensus (comb.lib./smm/nn)	LLVLGTVFLASVCV	3.5	23	82	-0.79	YES
HLA-DPA1*03:01/DPB1*04:02	9	23	Consensus (comb.lib./smm/nn)	LVLLGTVFLASVCVS	0.87	46	35.1	-0.80	YES
HLA-DPA1*03:01/DPB1*04:02	10	24	Consensus (comb.lib./smm/nn)	VLLGTVFLASVCVSL	3.4	114	81	0.14	
HLA-DRB3*01:01	22	36	Consensus (comb.lib./smm/nn)	VSLKVREDENNPFYL	3.7	37	92.8	-0.74	YES
HLA-DRB3*01:01	23	37	Consensus (comb.lib./smm/nn)	SLKVREDENNPFYLR	3.5	37	84.3	-0.63	YES
HLA-DRB1*04:01	29	43	Consensus (smm/nn/sturniolo)	DENNPFYLRSSNSFQ	1.2	86	64.9	-0.55	
HLA-DRB1*04:05	29	43	Consensus (smm/nn/sturniolo)	DENNPFYLRSSNSFQ	2.1	172	27	-0.55	
HLA-DRB1*04:01	30	44	Consensus (smm/nn/sturniolo)	ENNPYLRSSNSFQT	0.9	81	32.2	-0.53	
HLA-DRB1*04:05	30	44	Consensus (smm/nn/sturniolo)	ENNPYLRSSNSFQT	2.1	164	16.2	-0.53	
HLA-DRB1*04:01	31	45	Consensus (smm/nn/sturniolo)	NNPYLRSSNSFQTL	0.81	77	23.7	-0.52	
HLA-DRB1*04:05	31	45	Consensus (smm/nn/sturniolo)	NNPYLRSSNSFQTL	2.1	155	13.2	-0.52	
HLA-DRB1*07:01	31	45	Consensus (comb.lib./smm/nn)	NNPYLRSSNSFQTL	3.6	65	29.8	-0.52	
HLA-DRB1*04:01	32	46	Consensus (smm/nn/sturniolo)	NPFYLRSSNSFQTLF	0.82	78	21.8	-0.57	
HLA-DRB1*04:05	32	46	Consensus (smm/nn/sturniolo)	NPFYLRSSNSFQTLF	2.1	157	12.8	-0.57	
HLA-DRB1*07:01	32	46	Consensus (comb.lib./smm/nn)	NPFYLRSSNSFQTLF	3.3	65	28.1	-0.57	
HLA-DRB1*04:01	33	47	Consensus (smm/nn/sturniolo)	PFYLRSSNSFQTLFE	1.1	87	26.9	0.26/ YES	
HLA-DRB1*04:05	33	47	Consensus (smm/nn/sturniolo)	PFYLRSSNSFQTLFE	2.1	146	13.7	0.26/ YES	
HLA-DRB1*04:01	34	48	Consensus (smm/nn/sturniolo)	FYLRSSNSFQTLFEN	1.7	192	49.9	1.25/ YES	
HLA-DRB1*13:02	41	55	Consensus (smm/nn/sturniolo)	SFQTLFENQNGRIRL	3.7	8	27.1	0.43/ YES	YES
HLA-DRB1*13:02	42	56	Consensus (smm/nn/sturniolo)	FQTLFENQNGRIRLL	3.4	8	24	0.22/ YES	YES
HLA-DRB1*11:01	47	61	Consensus (smm/nn/sturniolo)	ENQNGRIRLLQRFNK	1.8	56	16.9	0.09	
HLA-DRB1*11:01	48	62	Consensus (smm/nn/sturniolo)	NQNGRIRLLQRFNKR	1.8	54	8.2	0.11	
HLA-DRB1*11:01	49	63	Consensus (smm/nn/sturniolo)	QNGRIRLLQRFNKRS	1.8	53	6	0.11	
HLA-DRB1*11:01	50	64	Consensus (smm/nn/sturniolo)	NGRIRLLQRFNKRS	1.8	44	5.7	0.08	
HLA-DRB1*11:01	51	65	Consensus (smm/nn/sturniolo)	GRIRLLQRFNKRS	1.8	46	5.6	0.22/ YES	
HLA-DRB1*11:01	52	66	Consensus (smm/nn/sturniolo)	RIRLLQRFNKRS	1.8	80	9	0.16	
HLA-DRB1*11:01	53	67	Consensus (smm/nn/sturniolo)	IRLLQRFNKRS	3	80	29	0.34/ YES	
HLA-DRB1*15:01	71	85	Consensus (smm/nn/sturniolo)	DYRIVQFQSKPNTIL	4	90	56.3	0.34/ YES	YES
HLA-DRB1*15:01	72	86	Consensus (smm/nn/sturniolo)	YRIVQFQSKPNTILL	3.4	92	49	0.04	YES
HLA-DRB1*01:01	90	104	Consensus (comb.lib./smm/nn)	ADADFLLFVLSGRAI	2.9	19	46.7	0.37/ YES	
HLA-DRB1*15:01	90	104	Consensus (smm/nn/sturniolo)	ADADFLLFVLSGRAI	3.5	39	49.6	0.37/ YES	
HLA-DRB1*15:01	91	105	Consensus (smm/nn/sturniolo)	DADFLLFVLSGRAIL	1.6	35	28.3	0.35/ YES	
HLA-DRB1*01:01	91	105	Consensus (comb.lib./smm/nn)	DADFLLFVLSGRAIL	2.2	15	18.6	0.35/ YES	
HLA-DRB1*07:01	91	105	Consensus (comb.lib./smm/nn)	DADFLLFVLSGRAIL	3.6	56	29.9	0.35/ YES	
HLA-DRB1*01:01	92	106	Consensus (comb.lib./smm/nn)	ADFLLFVLSGRAILT	1.3	11	9.4	0.36/ YES	
HLA-DRB1*15:01	92	106	Consensus (smm/nn/sturniolo)	ADFLLFVLSGRAILT	1.3	34	23.1	0.36/ YES	
HLA-DRB1*07:01	92	106	Consensus (comb.lib./smm/nn)	ADFLLFVLSGRAILT	3.7	55	32.5	0.36/ YES	

HLA-DRB1*01:01	93	107	Consensus (comb.lib./smm/nn)	DFLLFVLSGRAILTL	1.1	10	7.7	-0.64	
HLA-DRB1*15:01	93	107	Consensus (smm/nn/sturniolo)	DFLLFVLSGRAILTL	1.3	37	22.7	-0.64	
HLA-DRB1*07:01	93	107	Consensus (comb.lib./smm/nn)	DFLLFVLSGRAILTL	3.5	51	32.3	-0.64	
HLA-DPA1*03:01/DPB1*04:02	93	107	Consensus (comb.lib./smm/nn)	DFLLFVLSGRAILTL	3.8	165	94.1	-0.64	
HLA-DRB1*01:01	94	108	Consensus (comb.lib./smm/nn)	FLLFVLSGRAILTLV	1.3	11	7.6	-1.07	
HLA-DRB1*15:01	94	108	Consensus (smm/nn/sturniolo)	FLLFVLSGRAILTLV	1.6	79	25.9	-1.07	
HLA-DPA1*03:01/DPB1*04:02	94	108	Consensus (comb.lib./smm/nn)	FLLFVLSGRAILTLV	3.2	159	76	-1.07	
HLA-DRB1*07:01	94	108	Consensus (comb.lib./smm/nn)	FLLFVLSGRAILTLV	3.5	52	30.2	-1.07	
HLA-DRB1*01:01	95	109	Consensus (comb.lib./smm/nn)	LLFVLSGRAILTLVN	2.1	18	8.8	-1.03	
HLA-DPA1*03:01/DPB1*04:02	95	109	Consensus (comb.lib./smm/nn)	LLFVLSGRAILTLVN	3.8	165	98	-1.03	
HLA-DRB1*15:01	95	109	Consensus (smm/nn/sturniolo)	LLFVLSGRAILTLVN	4	96	57.1	-1.03	
HLA-DRB1*01:01	96	110	Consensus (comb.lib./smm/nn)	LFVLSGRAILTLVNN	3.7	23	13.1	-0.93	YES
HLA-DRB1*04:05	98	112	Consensus (smm/nn/sturniolo)	VLSGRAILTLVNDD	1.5	59	43.9	0.09	YES
HLA-DRB1*04:05	99	113	Consensus (smm/nn/sturniolo)	LSGRAILTLVNDDR	0.94	59	21.8	0.11	YES
HLA-DRB1*04:05	100	114	Consensus (smm/nn/sturniolo)	SGRAILTLVNDDR	0.98	61	15.3	0.11	YES
HLA-DRB1*04:05	101	115	Consensus (smm/nn/sturniolo)	GRAILTLVNDDRDS	0.97	60	15.4	0.12	YES
HLA-DRB1*04:05	102	116	Consensus (smm/nn/sturniolo)	RAILTLVNDDRDSY	1.1	63	18.6	0.20	
HLA-DRB1*04:05	103	117	Consensus (smm/nn/sturniolo)	AILTLVNDDRDSY N	2.3	189	48.3	0.21/ YES	
HLA-DRB4*01:01	137	151	Consensus (comb.lib./smm/nn)	PHDHQNLKIIKLAIP	1.3	89	17.4	0.28/ YES	
HLA-DRB4*01:01	138	152	Consensus (comb.lib./smm/nn)	HDHQNLKIIKLAIPV	1.1	85	11.2	0.28/ YES	
HLA-DRB4*01:01	139	153	Consensus (comb.lib./smm/nn)	DHQNLKIIKLAIPVN	0.1	102	9.4	0.29/ YES	
HLA-DRB4*01:01	140	154	Consensus (comb.lib./smm/nn)	HQNLKIIKLAIPVNK	0.08	96	8.5	1.44/ YES	
HLA-DRB1*12:01	140	154	Consensus (smm/nn)	HQNLKIIKLAIPVNK	1.44	204	82.1	1.44/ YES	
HLA-DRB5*01:01	140	154	Consensus (smm/nn/sturniolo)	HQNLKIIKLAIPVNK	3.1	45	19.9	1.44/ YES	
HLA-DRB4*01:01	141	155	Consensus (comb.lib./smm/nn)	QNLKIIKLAIPVNKP	0.13	104	9.6	1.49/ YES	
HLA-DRB1*12:01	141	155	Consensus (smm/nn)	QNLKIIKLAIPVNKP	1.91	197	111.2	1.49/ YES	
HLA-DRB5*01:01	141	155	Consensus (smm/nn/sturniolo)	QNLKIIKLAIPVNKP	2.5	47	16.4	1.49/ YES	
HLA-DRB4*01:01	142	156	Consensus (comb.lib./smm/nn)	NLKIIKLAIPVNKPG	0.23	189	13.2	1.49/ YES	
HLA-DRB5*01:01	142	156	Consensus (smm/nn/sturniolo)	NLKIIKLAIPVNKPG	2.7	49	17.4	1.49/ YES	
HLA-DRB1*12:01	142	156	Consensus (smm/nn)	NLKIIKLAIPVNKPG	3.36	186	199.6	1.49/ YES	
HLA-DRB4*01:01	143	157	Consensus (comb.lib./smm/nn)	LKIIKLAIPVNKPGR	0.81	205	24.1	1.49/ YES	
HLA-DRB5*01:01	143	157	Consensus (smm/nn/sturniolo)	LKIIKLAIPVNKPGR	2.3	47	15.5	1.49/ YES	
HLA-DRB5*01:01	144	158	Consensus (smm/nn/sturniolo)	KIIKLAIPVNKPGRY	3.4	49	21.5	1.49/ YES	
HLA-DRB1*04:01	155	169	Consensus (smm/nn/sturniolo)	PGRYDDFFLSSTQAQ	3.6	191	88.7	0.17	
HLA-DRB1*04:01	156	170	Consensus (smm/nn/sturniolo)	GRYDDFFLSSTQAQ Q	2.1	130	56.6	0.17	YES
HLA-DRB1*04:01	157	171	Consensus (smm/nn/sturniolo)	RYDDFFLSSTQAQQS	1.5	127	46.7	0.13	
HLA-DRB1*04:01	158	172	Consensus (smm/nn/sturniolo)	YDDFFLSSTQAQQSY	1.2	129	41.2	0.19	YES
HLA-DRB1*04:01	159	173	Consensus (smm/nn/sturniolo)	DDFFLSSTQAQQSYL	1.3	127	43.9	0.19	
HLA-DRB1*04:01	160	174	Consensus (smm/nn/sturniolo)	DFFLSSTQAQQSYLQ	1.9	204	52.8	0.18	
HLA-DQA1*01:02/DQB1*06:02	201	215	Consensus (comb.lib./smm/nn)	EEQRQQEGVIVELSK	3.5	176	93.1	1.31/ YES	YES
HLA-DQA1*01:02/DQB1*06:02	202	216	Consensus (comb.lib./smm/nn)	EQRQQEGVIVELSKE	3.5	174	68.7	1.32/ YES	
HLA-DQA1*01:02/DQB1*06:02	203	217	Consensus (comb.lib./smm/nn)	QRQQEGVIVELSKEQ	3.5	173	72.1	1.32/ YES	
HLA-DRB1*11:01	212	226	Consensus (smm/nn/sturniolo)	ELSKEQIRQLSRRAK	3.2	8	31.9	0.21/ YES	YES

HLA-DRB1*11:01	213	227	Consensus (smm/nn/sturniolo)	LSKEQIRQLSRRAKS	1.3	7	13.4	0.29/ YES	
HLA-DRB1*11:01	214	228	Consensus (smm/nn/sturniolo)	SKEQIRQLSRRAKSS	0.71	7	9.5	0.13	
HLA-DRB5*01:01	214	228	Consensus (smm/nn/sturniolo)	SKEQIRQLSRRAKSS	3.9	112	24.7	0.13	
HLA-DRB1*11:01	215	229	Consensus (smm/nn/sturniolo)	KEQIRQLSRRAKSSS	0.5	7	7.9	0.02	
HLA-DRB5*01:01	215	229	Consensus (smm/nn/sturniolo)	KEQIRQLSRRAKSSS	3.6	112	22.8	0.02	
HLA-DRB1*11:01	216	230	Consensus (smm/nn/sturniolo)	EQIRQLSRRAKSSSR	0.53	7	8.2	0.12	YES
HLA-DRB5*01:01	216	230	Consensus (smm/nn/sturniolo)	EQIRQLSRRAKSSSR	4	80	25.9	0.12	YES
HLA-DRB1*11:01	217	231	Consensus (smm/nn/sturniolo)	QIRQLSRRAKSSSRK	1.4	22	15.2	-0.02	
HLA-DRB5*01:01	217	231	Consensus (smm/nn/sturniolo)	QIRQLSRRAKSSSRK	2	118	12.2	-0.02	
HLA-DRB5*01:01	218	232	Consensus (smm/nn/sturniolo)	IRQLSRRAKSSSRKT	2	115	11.9	-0.02	
HLA-DRB1*11:01	218	232	Consensus (smm/nn/sturniolo)	IRQLSRRAKSSSRKT	3.7	22	36.6	-0.02	
HLA-DRB3*01:01	243	257	Consensus (comb.lib./smm/nn)	RSRNPIYSNNFGKFF	2.9	104	69.5	0.46/ YES	
HLA-DRB3*01:01	244	258	Consensus (comb.lib./smm/nn)	SRNPIYSNNFGKFFE	2.3	104	52	0.42/ YES	YES
HLA-DRB3*01:01	245	259	Consensus (comb.lib./smm/nn)	RNPIYSNNFGKFFEI	1.6	101	36.4	1.37/ YES	YES
HLA-DRB3*01:01	246	260	Consensus (comb.lib./smm/nn)	NPIYSNNFGKFFEIT	1.9	102	41.1	0.41/ YES	YES
HLA-DRB3*01:01	247	261	Consensus (comb.lib./smm/nn)	PIYSNNFGKFFEITP	2.6	103	62.9	0.41/ YES	
HLA-DRB1*04:05	267	281	Consensus (smm/nn/sturniolo)	LRDLDFLSSVDINE	1.4	55	43.2	0.92/ YES	
HLA-DRB1*04:05	268	282	Consensus (smm/nn/sturniolo)	RDLDFLSSVDINEG	0.91	56	36	0.88/ YES	
HLA-DRB1*04:05	269	283	Consensus (smm/nn/sturniolo)	DLDIFLSSVDINEGA	2.3	93	37.5	0.89/ YES	
HLA-DRB1*04:05	270	284	Consensus (smm/nn/sturniolo)	LDIFLSSVDINEGAL	2.1	90	43.3	0.98/ YES	
HLA-DRB1*13:02	271	285	Consensus (smm/nn/sturniolo)	DIFLSSVDINEGALL	3.2	82	22.7	0.87/ YES	YES
HLA-DRB1*04:05	271	285	Consensus (smm/nn/sturniolo)	DIFLSSVDINEGALL	3.7	126	66.1	0.87/ YES	YES
HLA-DRB1*13:02	272	286	Consensus (smm/nn/sturniolo)	IFLSSVDINEGALLL	2	78	14.9	-1.18	YES
HLA-DRB1*13:02	273	287	Consensus (smm/nn/sturniolo)	FLSSVDINEGALLLP	1.9	81	14.6	-1.20	
HLA-DRB1*13:02	274	288	Consensus (smm/nn/sturniolo)	LSSVDINEGALLLPH	1.9	81	14.7	-0.84	
HLA-DRB1*13:02	275	289	Consensus (smm/nn/sturniolo)	SSVDINEGALLLPHF	2.4	83	17.5	0.27/ YES	YES
HLA-DQA1*01:02/DQB1*06:02	275	289	Consensus (comb.lib./smm/nn)	SSVDINEGALLLPHF	3.3	169	91.1	0.27/ YES	YES
HLA-DQA1*01:02/DQB1*06:02	276	290	Consensus (comb.lib./smm/nn)	SVDINEGALLLPHFN	3.3	169	82.6	0.27/ YES	YES
HLA-DQA1*01:02/DQB1*06:02	277	291	Consensus (comb.lib./smm/nn)	VDINEGALLLPHFNS	3.2	166	77.7	0.27/ YES	YES
HLA-DRB1*07:01	285	299	Consensus (comb.lib./smm/nn)	LLPHFNSKAIVILVI	3.6	172	30.5	0.28/ YES	
HLA-DRB3*01:01	335	349	Consensus (comb.lib./smm/nn)	SEDDVFVIPAAYPFV	2.7	241	168.2	0.27/ YES	YES
HLA-DRB3*01:01	336	350	Consensus (comb.lib./smm/nn)	EDDVVFVIPAAYPFVV	2.6	236	123.7	0.27/ YES	
HLA-DRB5*01:01	336	350	Consensus (smm/nn/sturniolo)	EDDVVFVIPAAYPFVV	2.9	57	173.5	0.27/ YES	
HLA-DRB1*07:01	336	350	Consensus (comb.lib./smm/nn)	EDDVVFVIPAAYPFVV	4	58	31.9	0.27/ YES	
HLA-DRB3*01:01	337	351	Consensus (comb.lib./smm/nn)	DDVFVIPAAYPFVVN	2.7	237	127.4	0.28/ YES	
HLA-DRB5*01:01	337	351	Consensus (smm/nn/sturniolo)	DDVFVIPAAYPFVVN	2.9	57	160.1	0.28/ YES	
HLA-DRB3*01:01	338	352	Consensus (comb.lib./smm/nn)	DVFVIPAAYPFVVNA	2.7	237	148.7	0.28/ YES	
HLA-DRB5*01:01	338	352	Consensus (smm/nn/sturniolo)	DVFVIPAAYPFVVNA	2.9	56	167.1	0.28/ YES	
HLA-DRB1*13:02	342	356	Consensus (smm/nn/sturniolo)	IPAAYPFVVNATSNL	2.8	122	11	0.4/YES	

HLA-DRB1*04:01	343	357	Consensus (smm/nn/sturniolo)	PAAYPFVVNATSNL N	1.9	112	90.4	0.43/ YES	
HLA-DRB1*13:02	343	357	Consensus (smm/nn/sturniolo)	PAAYPFVVNATSNL N	2.4	111	7.8	0.43/ YES	
HLA-DRB1*04:01	344	358	Consensus (smm/nn/sturniolo)	AAYPFVVNATSNLN F	1.9	110	56.8	0.36/ YES	
HLA-DRB1*13:02	344	358	Consensus (smm/nn/sturniolo)	AAYPFVVNATSNLN F	2.3	107	6.6	0.36/ YES	
HLA-DRB1*07:01	344	358	Consensus (comb.lib./smm/nn)	AAYPFVVNATSNLN F	2.8	41	23.3	0.36/ YES	
HLA-DRB1*04:01	345	359	Consensus (smm/nn/sturniolo)	AYPFVVNATSNLNFL	1.8	128	51.7	0.34/ YES	
HLA-DRB1*13:02	345	359	Consensus (smm/nn/sturniolo)	AYPFVVNATSNLNFL	2.2	104	5.6	0.34/ YES	
HLA-DRB1*07:01	345	359	Consensus (comb.lib./smm/nn)	AYPFVVNATSNLNFL	2.5	38	19.1	0.34/ YES	
HLA-DRB1*04:01	346	360	Consensus (smm/nn/sturniolo)	YPFVVNATSNLNFLA	2.1	128	56.6	0.24/ YES	
HLA-DRB1*13:02	346	360	Consensus (smm/nn/sturniolo)	YPFVVNATSNLNFLA	2.3	105	6.5	0.24/ YES	
HLA-DRB1*07:01	346	360	Consensus (comb.lib./smm/nn)	YPFVVNATSNLNFLA	2.6	39	22.3	0.24/ YES	
HLA-DRB1*04:01	347	361	Consensus (smm/nn/sturniolo)	PFVVNATSNLNFLAF	3.1	179	77.3	0.31/ YES	
HLA-DRB1*07:01	347	361	Consensus (comb.lib./smm/nn)	PFVVNATSNLNFLAF	4	59	29	0.31/ YES	
HLA-DRB1*04:05	354	368	Consensus (smm/nn/sturniolo)	SNLNFLAFGINAENN	3.7	193	80.1	-0.50	
HLA-DRB1*04:05	355	369	Consensus (smm/nn/sturniolo)	NLNFLAFGINAENNQ	3.5	191	75.5	-0.59	YES
HLA-DRB1*04:05	356	370	Consensus (smm/nn/sturniolo)	LNFLAFGINAENNQR	3.9	194	82.5	-0.48	YES

Table S13. List of epitopes of P25974 predicted by NetMHCIIpan method along with interacting HLA class II alleles and relevant scores.

Allele	Start	End	Method	Peptide	Percentile rank	NetMHCIIpan IC50	IL-4 pred	AllerTOP v. 2.0
HLA-DPA1*01:03/DPB1*04:01	5	19	NetMHCIIpan	RFPLLVLGTVFLAS	2.6	149.77	-1.14	YES
HLA-DPA1*01:03/DPB1*04:01	6	20	NetMHCIIpan	FPLLVLGTVFLASV	2.4	143.92	-1.13	
HLA-DPA1*01:03/DPB1*04:01	7	21	NetMHCIIpan	PLLVLGTVFLASVC	3.2	178.43	-0.81	
HLA-DRB3*02:02	30	44	NetMHCIIpan	ENNPFYLRSSNSFQT	2.3	137	-0.53	
HLA-DRB3*02:02	31	45	NetMHCIIpan	NNPFYLRSSNSFQTL	1.2	71.04	-0.52	
HLA-DRB3*02:02	32	46	NetMHCIIpan	NPFYLRSSNSFQTLF	0.95	55.74	-0.57	
HLA-DRB3*02:02	33	47	NetMHCIIpan	PFYLRSSNSFQTLFE	1.1	66.16	0.26/ YES	
HLA-DRB3*02:02	34	48	NetMHCIIpan	FYLRSSNSFQTLFEN	1.7	92.65	1.25/ YES	
HLA-DRB3*02:02	41	55	NetMHCIIpan	SFQTLFENQNGRIRL	2.3	135.06	0.43/ YES	YES
HLA-DRB3*02:02	42	56	NetMHCIIpan	FQTLFENQNGRIRLL	1.9	105.94	0.22/ YES	YES
HLA-DRB3*02:02	43	57	NetMHCIIpan	QTLFENQNGRIRLLQ	2.5	147.58	0.21/ YES	
HLA-DRB3*02:02	44	58	NetMHCIIpan	TLFENQNGRIRLLQR	3.8	220.55	0.23YES	
HLA-DRB3*02:02	93	107	NetMHCIIpan	DFLLFVLSGRAILTL	3.3	198.22	-0.64	
HLA-DRB3*02:02	94	108	NetMHCIIpan	FLLFVLSGRAILTLV	3	175.63	-1.07	
HLA-DRB3*02:02	95	109	NetMHCIIpan	LLFVLSGRAILTLVN	3.5	203.85	-1.03	

HLA-DRB3*02:02	157	171	NetMHCIIPa n	RYDDFFLSSTQAQQS	3.5	204.25	0.13	
HLA-DRB3*02:02	158	172	NetMHCIIPa n	YDDFFLSSTQAQQSY	3	175.43	0.19	YES
HLA-DRB3*02:02	294	308	NetMHCIIPa n	IVILVINEGDANIEL	3.5	209.6	-0.60	
HLA-DRB3*02:02	342	356	NetMHCIIPa n	IPAAYPFVVNATSNL	0.12	20.46	0.4/YES	
HLA-DRB3*02:02	343	357	NetMHCIIPa n	PAAYPFVVNATSNL N	0.01	11.91	0.43/ YES	
HLA-DRB3*02:02	344	358	NetMHCIIPa n	AAYPFVVNATSNLN F	0.01	8.61	0.36/ YES	
HLA-DRB3*02:02	345	359	NetMHCIIPa n	AYPFVVNATSNLNFL	0.01	8.08	0.34/ YES	
HLA-DRB3*02:02	346	360	NetMHCIIPa n	YPFVVNATSNLNFLA	0.01	8.75	0.24/ YES	
HLA-DRB3*02:02	347	361	NetMHCIIPa n	PFVVNATSNLNFLAF	0.04	14.17	0.31/ YES	
HLA-DRB3*02:02	348	362	NetMHCIIPa n	FVVNATSNLNFLAFG	0.74	46.19	0.31/ YES	
HLA-DRB3*02:02	358	372	NetMHCIIPa n	FLAFGINAENNQRNF	3.9	229.08	-0.22	YES

Table S14. List of epitopes of P26987 predicted by consensus method along with interacting HLA class II alleles and relevant scores.

Allele	Start	End	Method	Peptide	Percentile rank	Smm IC50	Nn IC50	IL-4 pred	AllerTOP v. 2.0
HLA-DRB1*04:05	14	28	Consensus (smm/nn/sturniolo)	VAPATLYKALVTDAD	1.7	63	46.7	-0.38	
HLA-DRB1*04:05	15	29	Consensus (smm/nn/sturniolo)	APATLYKALVTDADN	1.1	63	28.8	-0.11	
HLA-DRB1*04:05	16	30	Consensus (smm/nn/sturniolo)	PATLYKALVTDADNV	0.97	60	28	-0.11	
HLA-DRB1*04:05	17	31	Consensus (smm/nn/sturniolo)	ATLYKALVTDADNVI	0.92	58	32.6	-0.01	YES
HLA-DRB3*01:01	17	31	Consensus (comb.lib./smm/nn)	ATLYKALVTDADNVI	1.5	150	87.9	-0.01	YES
HLA-DRB1*04:05	18	32	Consensus (smm/nn/sturniolo)	TLYKALVTDADNVIP	1.3	64	41.9	0.03	YES
HLA-DRB3*01:01	18	32	Consensus (comb.lib./smm/nn)	TLYKALVTDADNVIP	1.6	154	83	0.03	YES
HLA-DRB3*01:01	19	33	Consensus (comb.lib./smm/nn)	LYKALVTDADNVIPK	1.6	155	100.8	1.09/YES	YES
HLA-DRB3*01:01	20	34	Consensus (comb.lib./smm/nn)	YKALVTDADNVIPKA	1.6	156	107.3	1.02/YES	YES
HLA-DRB1*13:02	20	34	Consensus (smm/nn/sturniolo)	YKALVTDADNVIPKA	4	159	29.7	1.02/YES	YES
HLA-DRB3*01:01	21	35	Consensus (comb.lib./smm/nn)	KALVTDADNVIPKAL	1.6	155	142.9	0.13	
HLA-DPA1*02:01/DPB1*01:01	49	63	Consensus (comb.lib./smm/nn)	GPGTIKKITFLEDGE	3.3	84	150.5	1.48/YES	YES
HLA-DPA1*02:01/DPB1*01:01	50	64	Consensus (comb.lib./smm/nn)	PGTIKKITFLEDGET	3	82	139.6	1.40/YES	YES
HLA-DRB3*01:01	52	66	Consensus (comb.lib./smm/nn)	TIKKITFLEDGETKF	3	86	71.9	1.32/YES	
HLA-DRB3*01:01	53	67	Consensus (comb.lib./smm/nn)	IKKITFLEDGETKFFV	1.5	84	32.9	1.29/YES	
HLA-DRB3*01:01	54	68	Consensus (comb.lib./smm/nn)	KKITFLEDGETKFFVL	0.93	82	22.6	1.27/YES	YES
HLA-DRB1*03:01	54	68	Consensus (smm/nn/sturniolo)	KKITFLEDGETKFFVL	3.1	128	70.1	1.27/YES	YES
HLA-DRB3*01:01	55	69	Consensus (comb.lib./smm/nn)	KITFLEDGETKFFVLH	0.96	83	23.2	1.27/YES	YES
HLA-DRB1*03:01	55	69	Consensus (smm/nn/sturniolo)	KITFLEDGETKFFVLH	2.8	128	62.4	1.27/YES	YES
HLA-DRB3*01:01	56	70	Consensus (comb.lib./smm/nn)	ITFLEDGETKFFVLHK	1.7	85	38	1.31/YES	YES
HLA-DRB1*03:01	56	70	Consensus (smm/nn/sturniolo)	ITFLEDGETKFFVLHK	3.7	130	86.5	1.31/YES	YES
HLA-DRB1*07:01	60	74	Consensus (comb.lib./smm/nn)	EDGETKFFVLHKIESI	3.3	127	28	0.28/YES	YES
HLA-DRB1*01:01	77	91	Consensus (comb.lib./smm/nn)	ANLGYSYSVVGGAA L	2	14	56.4	-0.82	YES
HLA-DRB1*09:01	77	91	Consensus (comb.lib./smm/nn)	ANLGYSYSVVGGAA L	2.6	112	26.9	-0.82	YES
HLA-DRB1*01:01	78	92	Consensus (comb.lib./smm/nn)	NLGYSYSVVGGGAALP	2.2	15	51.2	-0.82	YES
HLA-DRB1*09:01	78	92	Consensus (comb.lib./smm/nn)	NLGYSYSVVGGGAALP	3	124	33.4	-0.82	YES
HLA-DQA1*05:01/DQB1*03:01	78	92	Consensus (comb.lib./smm/nn)	NLGYSYSVVGGGAALP	3.1	37	36.2	-0.82	YES
HLA-DQA1*05:01/DQB1*03:01	79	93	Consensus (comb.lib./smm/nn)	LGYSYSVVGGGAALPD	2.2	28	28.5	-0.82	
HLA-DRB1*01:01	79	93	Consensus (comb.lib./smm/nn)	LGYSYSVVGGGAALPD	2.2	15	55.6	-0.82	
HLA-DRB1*09:01	79	93	Consensus (comb.lib./smm/nn)	LGYSYSVVGGGAALPD	3.1	125	39.1	-0.82	
HLA-DQA1*05:01/DQB1*03:01	80	94	Consensus (comb.lib./smm/nn)	GYSYSVVGGGAALPDT	1.9	26	21.1	-0.82	YES
HLA-DRB1*01:01	80	94	Consensus (comb.lib./smm/nn)	GYSYSVVGGGAALPDT	2.2	15	63.9	-0.82	YES
HLA-DQA1*05:01/DQB1*03:01	81	95	Consensus (comb.lib./smm/nn)	YSYSVVGGGAALPDTA	1.7	23	15	-0.87	
HLA-DRB1*01:01	81	95	Consensus (comb.lib./smm/nn)	YSYSVVGGGAALPDTA	2.2	15	88.7	-0.87	
HLA-DQA1*05:01/DQB1*03:01	82	96	Consensus (comb.lib./smm/nn)	SYSVVGGGAALPDTAE	1.8	24	14.9	-0.90	
HLA-DQA1*05:01/DQB1*03:01	83	97	Consensus (comb.lib./smm/nn)	YSVVGGGAALPDTAEK	2.5	30	17.5	-1.01	YES
HLA-DQA1*05:01/DQB1*03:01	84	98	Consensus (comb.lib./smm/nn)	SVVGGGAALPDTAEKI	3.1	38	27.9	-1.17	YES
HLA-DRB1*03:01	92	106	Consensus (smm/nn/sturniolo)	PDTAEKITFDSKLV A	2.4	184	16.2	0.05	YES
HLA-DRB1*03:01	93	107	Consensus (smm/nn/sturniolo)	DTAEKITFDSKLVAG	2.6	194	11.6	0.08	YES
HLA-DRB1*03:01	94	108	Consensus (smm/nn/sturniolo)	TAEKITFDSKLVAGP	2.7	203	13.1	0.09	
HLA-DRB1*03:01	95	109	Consensus (smm/nn/sturniolo)	AEKITFDSKLVAGPN	2.7	203	12.9	0.10	
HLA-DRB1*03:01	96	110	Consensus (smm/nn/sturniolo)	EKITFDSKLVAGPNG	2.7	203	16.5	0.15	YES
HLA-DQA1*05:01/DQB1*03:01	101	115	Consensus (comb.lib./smm/nn)	DSKLVAGPNGGSAGK	3	35	120.8	0.10	
HLA-DQA1*05:01/DQB1*03:01	102	116	Consensus (comb.lib./smm/nn)	SKLVAGPNGGSAGKL	2.5	25	84	0.09	
HLA-DQA1*05:01/DQB1*03:01	103	117	Consensus (comb.lib./smm/nn)	KL VAGPNGGSAGKLT	2.5	27	50.7	0.02	
HLA-DQA1*05:01/DQB1*03:01	104	118	Consensus (comb.lib./smm/nn)	LVAGPNGGSAGKLT V	2.5	28	49.3	0.07	
HLA-DQA1*05:01/DQB1*03:01	105	119	Consensus (comb.lib./smm/nn)	VAGPNGGSAGKLT V K	2.7	32	48.9	0.06	
HLA-DQA1*05:01/DQB1*03:01	106	120	Consensus (comb.lib./smm/nn)	AGPNGGSAGKLT V K Y	3	35	67.3	0.21/YES	YES
HLA-DPA1*02:01/DPB1*05:01	138	152	Consensus (comb.lib./smm/nn)	AKADALFKAIEAYLL	0.24	140	43.6	1.23/YES	YES
HLA-DRB1*09:01	138	152	Consensus (comb.lib./smm/nn)	AKADALFKAIEAYLL	0.49	32	13.9	1.23/YES	YES

HLA-DRB1*15:01	138	152	Consensus (smm/nn/sturniolo)	AKADALFKAIEAYLL	0.87	57	13.6	1.23/YES	YES
HLA-DPA1*02:01/DPB1*01:01	138	152	Consensus (comb.lib./smm/nn)	AKADALFKAIEAYLL	0.95	73	48	1.23/YES	YES
HLA-DRB5*01:01	138	152	Consensus (smm/nn/sturniolo)	AKADALFKAIEAYLL	1.3	10	80.1	1.23/YES	YES
HLA-DRB1*07:01	138	152	Consensus (comb.lib./smm/nn)	AKADALFKAIEAYLL	1.6	12	15.8	1.23/YES	YES
HLA-DRB1*01:01	138	152	Consensus (comb.lib./smm/nn)	AKADALFKAIEAYLL	1.8	3	7.8	1.23/YES	YES
HLA-DPA1*02:01/DPB1*05:01	139	153	Consensus (comb.lib./smm/nn)	KADALFKAIEAYLLA	0.22	136	15.4	1.22/YES	
HLA-DRB1*09:01	139	153	Consensus (comb.lib./smm/nn)	KADALFKAIEAYLLA	0.48	32	13.6	1.22/YES	
HLA-DRB1*01:01	139	153	Consensus (comb.lib./smm/nn)	KADALFKAIEAYLLA	0.51	3	4.8	1.22/YES	
HLA-DRB1*15:01	139	153	Consensus (smm/nn/sturniolo)	KADALFKAIEAYLLA	0.86	56	9.3	1.22/YES	
HLA-DPA1*02:01/DPB1*01:01	139	153	Consensus (comb.lib./smm/nn)	KADALFKAIEAYLLA	0.95	68	15.2	1.22/YES	
HLA-DRB5*01:01	139	153	Consensus (smm/nn/sturniolo)	KADALFKAIEAYLLA	1.3	10	31.8	1.22/YES	
HLA-DRB1*07:01	139	153	Consensus (comb.lib./smm/nn)	KADALFKAIEAYLLA	1.7	12	16.6	1.22/YES	
HLA-DRB1*04:01	139	153	Consensus (smm/nn/sturniolo)	KADALFKAIEAYLLA	3.1	89	79.1	1.22/YES	
HLA-DPA1*02:01/DPB1*05:01	140	154	Consensus (comb.lib./smm/nn)	ADALFKAIEAYLLAH	0.24	138	11.8	1.22/YES	YES
HLA-DRB1*01:01	140	154	Consensus (comb.lib./smm/nn)	ADALFKAIEAYLLAH	0.27	3	4.2	1.22/YES	YES
HLA-DRB1*09:01	140	154	Consensus (comb.lib./smm/nn)	ADALFKAIEAYLLAH	0.68	32	15.7	1.22/YES	YES
HLA-DRB1*15:01	140	154	Consensus (smm/nn/sturniolo)	ADALFKAIEAYLLAH	0.86	56	7.8	1.22/YES	YES
HLA-DPA1*02:01/DPB1*01:01	140	154	Consensus (comb.lib./smm/nn)	ADALFKAIEAYLLAH	0.95	69	8.4	1.22/YES	YES
HLA-DRB5*01:01	140	154	Consensus (smm/nn/sturniolo)	ADALFKAIEAYLLAH	1.3	10	23.8	1.22/YES	YES
HLA-DRB1*04:01	140	154	Consensus (smm/nn/sturniolo)	ADALFKAIEAYLLAH	2	90	55.7	1.22/YES	YES
HLA-DRB1*07:01	140	154	Consensus (comb.lib./smm/nn)	ADALFKAIEAYLLAH	2.2	12	19.7	1.22/YES	YES
HLA-DPA1*02:01/DPB1*05:01	141	155	Consensus (comb.lib./smm/nn)	DALFKAIEAYLLAHP	0.24	141	11.1	1.22/YES	
HLA-DRB1*01:01	141	155	Consensus (comb.lib./smm/nn)	DALFKAIEAYLLAHP	0.44	3	4.6	1.22/YES	
HLA-DRB1*15:01	141	155	Consensus (smm/nn/sturniolo)	DALFKAIEAYLLAHP	0.84	55	7.2	1.22/YES	
HLA-DRB1*09:01	141	155	Consensus (comb.lib./smm/nn)	DALFKAIEAYLLAHP	0.9	32	18.5	1.22/YES	
HLA-DPA1*02:01/DPB1*01:01	141	155	Consensus (comb.lib./smm/nn)	DALFKAIEAYLLAHP	0.95	59	8.3	1.22/YES	
HLA-DRB5*01:01	141	155	Consensus (smm/nn/sturniolo)	DALFKAIEAYLLAHP	1.3	10	25.8	1.22/YES	
HLA-DRB1*04:01	141	155	Consensus (smm/nn/sturniolo)	DALFKAIEAYLLAHP	1.7	91	49.8	1.22/YES	
HLA-DRB1*07:01	141	155	Consensus (comb.lib./smm/nn)	DALFKAIEAYLLAHP	2.9	13	25	1.22/YES	
HLA-DPA1*02:01/DPB1*05:01	142	156	Consensus (comb.lib./smm/nn)	ALFKAIEAYLLAHPD	0.52	224	10.2	1.22/YES	
HLA-DRB1*01:01	142	156	Consensus (comb.lib./smm/nn)	ALFKAIEAYLLAHPD	0.58	3	5.2	1.22/YES	
HLA-DRB1*15:01	142	156	Consensus (smm/nn/sturniolo)	ALFKAIEAYLLAHPD	0.87	57	7.5	1.22/YES	
HLA-DPA1*02:01/DPB1*01:01	142	156	Consensus (comb.lib./smm/nn)	ALFKAIEAYLLAHPD	0.95	79	8.7	1.22/YES	
HLA-DRB1*09:01	142	156	Consensus (comb.lib./smm/nn)	ALFKAIEAYLLAHPD	1.2	36	21.6	1.22/YES	
HLA-DRB5*01:01	142	156	Consensus (smm/nn/sturniolo)	ALFKAIEAYLLAHPD	1.3	10	26	1.22/YES	
HLA-DRB1*04:01	142	156	Consensus (smm/nn/sturniolo)	ALFKAIEAYLLAHPD	2	95	55.5	1.22/YES	
HLA-DRB1*07:01	142	156	Consensus (comb.lib./smm/nn)	ALFKAIEAYLLAHPD	3	13	26	1.22/YES	
HLA-DPA1*02:01/DPB1*01:01	143	157	Consensus (comb.lib./smm/nn)	LFKAIEAYLLAHPDY	0.95	132	10.8	1.24/YES	
HLA-DRB5*01:01	143	157	Consensus (smm/nn/sturniolo)	LFKAIEAYLLAHPDY	1.3	30	42.2	1.24/YES	
HLA-DRB1*01:01	143	157	Consensus (comb.lib./smm/nn)	LFKAIEAYLLAHPDY	1.4	10	7.1	1.24/YES	
HLA-DRB1*09:01	143	157	Consensus (comb.lib./smm/nn)	LFKAIEAYLLAHPDY	3	98	38.1	1.24/YES	
HLA-DRB5*01:01	144	158	Consensus (smm/nn/sturniolo)	FKAIEAYLLAHPDYN	1.3	31	92.8	1.29/YES	
HLA-DPA1*02:01/DPB1*01:01	144	158	Consensus (comb.lib./smm/nn)	FKAIEAYLLAHPDYN	3.9	191	21.8	1.29/YES	

Table S15. List of epitopes of P26987 predicted by NetMHCIIpan method along with interacting HLA class II alleles and relevant scores.

Allele	Start	End	Method	Peptide	Percentile rank	NetMHCIIpan IC50	IL-4 pred	AllerTOP v. 2.0
HLA-DPA1*01:03/DPB1*04:01	138	152	NetMHCIIpan	AKADALFKAIEAYLL	1.1	96.61	1.23/YES	YES
HLA-DPA1*02:01/DPB1*14:01	139	153	NetMHCIIpan	KADALFKAIEAYLLA	0.07	150.04	1.22/YES	
HLA-DPA1*01:03/DPB1*04:01	139	153	NetMHCIIpan	KADALFKAIEAYLLA	0.35	55.82	1.22/YES	
HLA-DPA1*02:01/DPB1*14:01	140	154	NetMHCIIpan	ADALFKAIEAYLLAH	0.03	121.3	1.22/YES	YES
HLA-DPA1*01:03/DPB1*04:01	140	154	NetMHCIIpan	ADALFKAIEAYLLAH	0.32	51.92	1.22/YES	YES
HLA-DRB3*02:02	140	154	NetMHCIIpan	ADALFKAIEAYLLAH	4	234.39	1.22/YES	YES
HLA-DPA1*02:01/DPB1*14:01	141	155	NetMHCIIpan	DALFKAIEAYLLAHP	0.03	112.21	1.22/YES	
HLA-DPA1*01:03/DPB1*04:01	141	155	NetMHCIIpan	DALFKAIEAYLLAHP	0.32	52.06	1.22/YES	

HLA-DRB3*02:02	141	155	NetMHCIIPa n	DALFKAIEAYLLAHP	3.9	230.46	1.22/YES
HLA-DPA1*02:01/DPB1*14:01	142	156	NetMHCIIPa n	ALFKAIEAYLLAHPD	0.03	106.95	1.22/YES
HLA-DPA1*01:03/DPB1*04:01	142	156	NetMHCIIPa n	ALFKAIEAYLLAHPD	0.33	54.24	1.22/YES
HLA-DPA1*02:01/DPB1*14:01	143	157	NetMHCIIPa n	LFKAIEAYLLAHPDY	0.07	149.94	1.24/YES
HLA-DPA1*01:03/DPB1*04:01	143	157	NetMHCIIPa n	LFKAIEAYLLAHPDY	0.59	68.33	1.24/YES
HLA-DPA1*01:03/DPB1*04:01	144	158	NetMHCIIPa n	FKAIEAYLLAHPDYN	2.8	161.65	1.29/YES

Table S16. Amino acids frequency of seven soybean allergens and their T cell epitopes.

Name Amino acids	P01070	P01070-T	P04347	P04347-T	P04776	P04776-T	P05046	P05046-T	P11827	P11827-T	P25974	P25974-T	P26987	P26987-T
A	5.56%	6.67%	3.88%	4.85%	5.86%	7.06%	8.07%	9.33%	4.35%	3.53%	5.24%	3.11%	11.39%	11.67%
R	4.63%	7.62%	6.40%	4.24%	5.46%	2.94%	2.11%	2.67%	6.60%	7.45%	7.06%	6.22%	0.00%	0.00%
N	4.63%	3.33%	6.40%	6.67%	7.48%	13.53%	6.32%	9.07%	6.28%	11.37%	7.52%	9.78%	5.06%	1.67%
D	7.87%	4.29%	4.65%	3.33%	3.43%	3.53%	5.97%	3.20%	4.35%	4.71%	4.78%	3.56%	7.60%	7.22%
C	2.32%	0.95%	1.55%	0.00%	2.22%	0.00%	0.00%	0.00%	0.81%	0.00%	0.23%	0.00%	0.00%	0.00%
Q	2.32%	1.91%	8.72%	2.12%	9.70%	5.49%	2.46%	1.07%	8.54%	7.06%	7.52%	7.56%	0.63%	0.00%
E	6.48%	1.91%	8.14%	4.24%	8.28%	4.12%	3.16%	1.87%	13.04%	1.18%	8.43%	9.33%	8.23%	8.89%
G	7.87%	6.67%	7.95%	4.24%	7.27%	1.77%	4.56%	4.27%	4.83%	3.92%	4.56%	5.33%	9.49%	7.22%
H	1.39%	1.43%	2.91%	4.24%	1.62%	2.35%	2.11%	2.13%	3.54%	3.92%	1.82%	1.33%	1.27%	2.78%
I	7.41%	6.19%	3.30%	7.27%	5.25%	5.29%	5.26%	9.33%	4.51%	5.49%	5.92%	8.44%	5.06%	7.78%
L	8.80%	11.91%	7.95%	5.76%	7.48%	7.45%	11.23%	5.87%	7.73%	11.37%	10.93%	11.11%	8.23%	12.22%
K	6.02%	8.57%	3.68%	4.24%	5.05%	4.31%	5.26%	6.13%	6.92%	5.49%	4.78%	3.56%	10.13%	12.78%
M	1.39%	0.95%	0.97%	0.00%	1.41%	0.98%	0.70%	0.53%	0.64%	0.00%	0.46%	0.00%	0.63%	0.00%
F	6.48%	6.67%	3.49%	6.06%	4.85%	10.59%	4.91%	6.40%	4.83%	8.63%	6.61%	9.33%	4.43%	6.67%
P	5.09%	6.19%	7.36%	5.15%	5.86%	6.86%	5.61%	3.73%	5.64%	3.53%	5.01%	4.00%	5.06%	3.33%
S	7.41%	7.62%	8.33%	11.82%	6.87%	9.61%	12.28%	15.20%	7.57%	3.53%	7.52%	7.56%	5.06%	1.11%
T	4.63%	4.76%	4.07%	2.42%	4.04%	2.94%	8.07%	6.93%	1.93%	5.10%	2.51%	2.22%	6.96%	8.33%
W	0.93%	0.00%	0.78%	0.30%	0.81%	0.39%	2.11%	4.00%	0.48%	0.00%	0.00%	0.00%	0.00%	0.00%
Y	2.32%	0.48%	2.91%	7.88%	2.22%	3.53%	1.75%	0.80%	2.42%	4.31%	2.73%	2.22%	3.80%	3.33%
V	6.48%	11.91%	6.59%	15.15%	4.85%	7.26%	8.07%	7.47%	4.99%	9.41%	6.38%	5.33%	6.96%	5.00%

Notes: P01070-T: T cell epitopes of P01070 protein; P04347-T: T cell epitopes of P04347 protein.; P04776-T: T cell epitopes of P04776 protein; P05046-T: T cell epitopes of P05046 protein; P11827-T: T cell epitopes of P11827 protein. P25974-T: T cell epitopes of P25974 protein. P26987-T: T cell epitopes of P26987 protein.