



OPEN

## Cross-reactive MHC class I T cell epitopes may dictate heterologous immune responses between respiratory viruses and food allergens

Kathrin Balz<sup>1,5</sup>, Abhinav Kaushik<sup>2,3,5</sup>, Franz Cemic<sup>4</sup>, Vanitha Sampath<sup>2</sup>, Vanessa Heger<sup>4</sup>, Harald Renz<sup>1</sup>, Kari Nadeau<sup>3</sup> & Chrysanthi Skevaki<sup>1</sup>✉

Respiratory virus infections play a major role in asthma, while there is a close correlation between asthma and food allergy. We hypothesized that T cell-mediated heterologous immunity may induce asthma symptoms among sensitized individuals and used two independent *in silico* pipelines for the identification of cross-reactive virus- and food allergen- derived T cell epitopes, considering individual peptide sequence similarity, MHC binding affinity and immunogenicity. We assessed the proteomes of human rhinovirus (RV1b), respiratory syncytial virus (RSVA2) and influenza-strains contained in the seasonal quadrivalent influenza vaccine 2019/2020 (QIV 2019/2020), as well as SARS-CoV-2 for human HLA alleles, in addition to more than 200 most common food allergen protein sequences. All resulting allergen-derived peptide candidates were subjected to an elaborate scoring system considering multiple criteria, including clinical relevance. In both bioinformatics approaches, we found that shortlisted peptide pairs that are potentially binding to MHC class II molecules scored up to 10× lower compared to MHC class I candidate epitopes. For MHC class I food allergen epitopes, several potentially cross-reactive peptides from shrimp, kiwi, apple, soybean and chicken were identified. The shortlisted set of peptide pairs may be implicated in heterologous immune responses and translated to peptide immunization strategies with immunomodulatory properties.

During the last few decades, the prevalence of allergic diseases has dramatically increased in developed countries. The incidence of asthma has increased fourfold since the 1950's<sup>1</sup> and food allergy prevalence among children has increased to 3.5–8%<sup>2</sup>. Food allergy is often classified into either IgE-mediated, non-IgE-mediated, or mixed IgE/non-IgE-mediated allergic disease. Non-IgE-mediated food allergy is thought to be initiated by T cells, although the pathophysiological mechanisms underlying these reactions are not yet fully understood<sup>3,4</sup>. Studies have observed that food allergy is associated with subsequent increases in the development of allergic rhinitis and asthma<sup>5</sup>. Both consumption and inhalation of food allergens can cause allergic reactions in sensitized individuals. Inhalation of aerosolized wheat, lupin, and other food allergens are thought to stimulate mast cells in the lung and have been associated with respiratory symptoms, including wheeze<sup>6–8</sup>. As wheezing in childhood is mostly associated with viral infections, an indirect correlation between viral infections and food allergy has been suggested. Gastrointestinal viral infections are also more relevant in the context of food allergy<sup>9</sup>. Some mouse studies provide evidence for the subsequent development of food allergen specific IgE on exposure to food allergens after gastrointestinal infection with certain RNA viruses, including murine norovirus type 1 or reovirus<sup>10,11</sup>.

We have previously shown an influenza virus-mediated protective effect over development of experimental asthma in models of ovalbumin and house dust mite-induced asthma. This effect was mediated by cross-reactive

<sup>1</sup>Institute of Laboratory Medicine, Universities of Giessen and Marburg Lung Center (UGMLC), Philipps University Marburg, German Center for Lung Research (DZL), 35043 Marburg, Germany. <sup>2</sup>Division of Pulmonary, Allergy and Critical Care Medicine, Sean N. Parker Center for Allergy and Asthma Research at Stanford University, Stanford, CA 94040, USA. <sup>3</sup>Department of Environmental Health, Harvard T. H. Chan School of Public Health, Boston, MA 02115, USA. <sup>4</sup>Department of Computer Science, TH Mittelhessen, University of Applied Sciences Gießen, 35390 Giessen, Germany. <sup>5</sup>These authors contributed equally: Kathrin Balz and Abhinav Kaushik. ✉email: Chrysanthi.Skevaki@uk-gm.de

T effector memory cells<sup>12</sup>. We hypothesized that cross-reactive T cell epitopes present in respiratory viruses and food allergens may provide the missing link between allergy and wheeze.

Therefore, our aim was to identify potentially cross-reactive T cell epitope pairs among food allergens and clinically relevant respiratory viruses. Such data can serve as the basis for investigating heterologous immune responses among patients with a specific food allergy and a viral infection or even for following outcomes (beneficial or harmful) of antiviral vaccination.

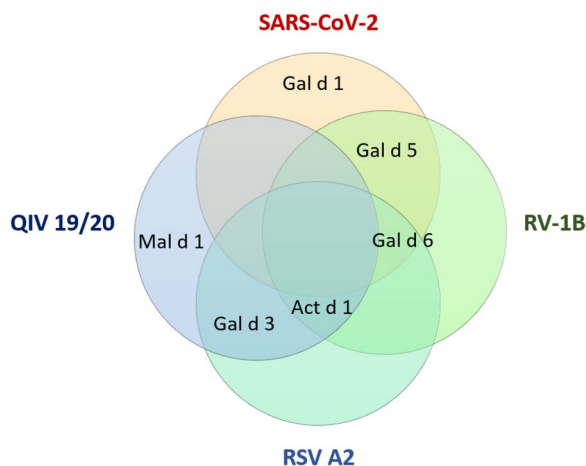
Recently, computational advances have significantly improved our understanding of cross-reactive sites across different allergens. These methods screen for antigenic peptides with similar epitope profile to bind a particular T cell receptor (TCR)<sup>13</sup>. However, most of these methods have their input dataset requirements (e.g., TCRseq or expression data or 3D structure of proteins) and with limited or no options to customize the source of training dataset (e.g., training with experimentally validated T-cell epitopes only) to achieve case-specific objectives<sup>14–16</sup>. Therefore, in this study, we applied two independent approaches developed in-house to predict MHC binding cross-reactive peptide sequences across viral and food allergen sequences that can potentially cause similar T-cell response. These tools also include prediction of MHC binding affinity, which is highly dependent on the so called “anchor residues” of the peptides. Such residues represent at least two amino acids, which are exactly fitting into the groove of the MHC molecule. Thus, anchor residues play an important role in the definition of antigenic peptides<sup>17</sup>.

## Results

We used partially modified versions of our previously published in-silico pipelines for prediction of potentially cross-reactive T cell epitope pairs between allergens and viruses, focusing on food allergens and clinically relevant respiratory viruses. The pipeline-1 uses an agnostic approach to predict the antigenic peptides by means of T cell epitope prediction for both viruses and allergens as well as for the sequence homology of the predicted epitopes, whereas pipeline-2 uses a supervised approach, wherein IEDB T-cell epitope repertoire was used to predict the cross-reactive antigenic regions across viruses and allergens.

Pipeline-1 predicted multiple potentially-cross-reactive T cell epitope pairs for each virus, which were further ranked based on the calculated pair combined score and subsequently scored for the top 5 candidate pairs based on additional criteria. Among the top 5 results for all viruses, we identified T cell epitope pairs with the allergenic epitopes deriving from kiwi (*Actinidia deliciosa*, *Act d*), chicken (*Gallus gallus*, *Gal d*) and apple (*Malus domestica*, *Mal d*) (Fig. 1). The top 5 candidates for RSVA2, RV1b and influenza strains of the seasonal quadrivalent influenza vaccine 2019/2020 (QIV 19/20) on the background of the most frequent human HLA class I alleles are depicted in Table 1. The corresponding top 5 candidate pairs for SARS-CoV-2 for pipeline-1 were published previously<sup>18</sup>.

Interestingly, sequences from apple only appeared in the QIV/19/20 top 5 candidate epitope sequences. Sequences from Gal d 6 and Act d 1 were predicted in 3 out of the 4 virus analyses, suggesting an important role in cross-reactivity between respiratory viruses and food allergens. With regards to the viral proteins, the predicted top 5 sequences for RSVA2 were derived exclusively from the RNA-directed polymerase, whereas the viral protein source for the other viruses were more diverse. Of note, when applying the pipeline for human HLA class II prediction, no candidate epitope pairs could be identified for RV1b and QIV19/20. Epitope pairs for RSVA2 achieved similar scores in both HLA class I and II analyses (Table 2). Importantly, such epitopes derived from a variety of viral proteins and different allergenic sources compared with candidates for HLA class I. Allergenic sources include buckwheat (*Fagopyrum esculentum*, *Fag e*), sesame (*Sesamum indicum*, *Ses i*), potato (*Solanum tuberosum*, *Sola t*), and hazelnut (*Corylus avellana*, *Cor a*), with almost all belonging to the protein family of cupins.



**Figure 1.** Venn diagram depicting the molecular allergen components containing the allergen counterpart of the top 5 predicted epitope pairs for HLA class I prediction with pipeline-1. The top 5 candidates were predicted with pipeline-1 and ranked based on the described scoring system. Molecular allergen components contained in the predicted top 5 allergen epitopes for each virus are depicted in the Venn diagram.

	Allergen epitope	Allergen	Allergen protein family	Viral epitope	Viral protein	MHC allele	Final score	
A: RV1b								
Nr. 1	ASDVICQEY	Gal d 5	Serum albumin	MHDSILVSY	P2-A	HLA-A*01:01	1258	
Nr. 2	VSDDGLNIY	Gal d 6	Lipoprotein	LSCKFLPLY	P2-C	HLA-A*01:01	1032	
Nr. 3	FADLTNEEY	Act d 1	Papain-like cysteine protease	YIPETEDDY	Protease	HLA-A*01:01	610	
Nr. 4	KVFRFSMFK	Gal d 6	Lipoprotein	GIFGENMY	VP2	HLA-A*11:01	508	
Nr. 5	FLGDVIPPPI	Gal d 6	Lipoprotein	LLLAYTPPGI	VP3	HLA-A*02:01	484	
B: RSVA2								
Nr. 1	FLGDKFYTV	Gal d 3	Transferrin	FLPKISLT	RNA-directed RNA polymerase L	HLA-A*02:01	1166	
Nr. 2	VYMDLPHGI	Mal d	Unknown	LYMNLPLMF	RNA-directed RNA polymerase L	HLA-A*24:02	1148	
Nr. 3	TTYKEFLGDK	Gal d 3	Transferrin	TTYNQFLTWK	RNA-directed RNA polymerase L	HLA-A*11:01	642	
Nr. 4	FADLTNEEY	Act d 1	Papain-like cysteine protease	KSNRYNDNY	RNA-directed RNA polymerase L	HLA-A*01:01	586	
Nr. 5	YLLDLLPAA	Gal d 6	Lipoprotein	YLSSELLNSL	RNA-directed RNA polymerase L	HLA-A*02:01	516	
	Allergen epitope	Allergen	Allergen protein family	Viral epitope	Virus protein	Virus strain	MHC allele	Final score
C: Influenza vaccine 2019/2020								
Nr. 1	SPARLYNAL	Mal d 1	Bet v 1 family	NPAMLYNKM	PB2	A/Kansas/2017	HLA-B*07:02	1784
Nr. 2	YLLQNPEAYV	Mal d 1	Bet v 1 family	SLYQNADAYV	HA	A/Brisbane/2018	HLA-A*02:01	636
Nr. 3	FLGDKFYTV	Gal d 3	Transferrin	FMYSDFHFI	PA	A/Brisbane/2018, A/Kansas/2017	HLA-A*02:01	604
Nr. 4	KPTSKRMAI	Act d 1	Papain-like cysteine protease	EPEKRMSL	NS1	B/Colorado/2017, B/Phuket/2013	HLA-B*07:02	556
Nr. 5	YTQTYGVVDY	Mal d	Unknown	YTDTYHSYA	NA	B/Colorado/2017, B/Phuket/2013	HLA-A*01:01	432

**Table 1.** Top 5 candidate epitope pairs for RV1b, RSVA2 and Influenza vaccine 2019/2020 on the background of the most frequent human HLA alleles (Supplementary Table 3), based on pipeline-1 and calculation of the final score as described (see “Materials and methods”).

	Allergen epitope	Allergen	Allergen protein family	Viral epitope	Viral protein	MHC allele	Final score
1	Fag e 1	LPILEFLQLSAQHVV	Cupin	KGAFKYIKPQSQFIV	Matrix protein	HLA-DRB1_01_01	1292
2	Ses i 3	VLFALLASAVVASE	Cupin	MELLILKANAITTIL	Fusion glycoprotein F0	HLA-DRB1_01_01	666
3	Cor a 9	ARRLYNRQETTLAR	Cupin	LRWLTYKLNTPSL	RNA-directed RNA polymerase L	HLA-DRB1_04_01	638
4	Sola t 1	FAKLLSDRKKLRANK	Patatin family	TELNSDDIKLRDNE	Matrix M2-1	HLA-DRB1_03_01	350
5	Fag e 1	LPILEFIQLSAQHVV	Cupin	KGAFKYIKPQSQFIV	Matrix protein	HLA-DRB1_01_01	168

**Table 2.** Top 5 candidate pairs for RSVA2 on the background of the most frequent human HLA class II alleles (Supplementary Table 3) based on pipeline-1 and calculation of the final score as described (see “Materials and methods”).

In order to validate our in-silico results we applied an independent pipeline that uses known T cell epitope features as they appear on IEDB tools. The pipeline is an extension of a computational framework we published previously<sup>18</sup>, wherein similarity between identical *k*-mers with experimentally validated T cell epitopes was used to predict cross-reactive peptides. In this analysis, we observed a limited number of cross-reactivity sites between food and viral antigens, wherein, the largest number of cross-reactive peptides were predicted between *Glycine max* (soybean) and SARS-CoV-2 replicase polyprotein. Besides the polyprotein, SARS-CoV-2 spike glycoprotein also shares potential cross-reactive amino acids with the sarcoplasmic ca-binding proteins of *Crangon crangon* (shrimp). In addition to SARS-CoV-2, we also observed that human respiratory syncytial virus A proteins also share potential cross-reactive sites with food allergen sequences from *Crangon crangon* and *Malus domestica* (apple). Overall, six different protein sequences from two viral species were found to share cross-reactive amino acids with three food allergens. Wherein, we predicted 44 redundant (16 non-redundant) cross-reactive peptide pairs between those six viral sequences with 16 different food allergen sequences (Table 3; Supplementary Table 3). A total of 22 unique peptides from food and viral allergens were finally shortlisted that were predicted to bind 27 unique class-I HLA alleles, that includes both HLA-A and HLA-B genes.

Comparing the T cell epitope pairs predicted with both pipelines, no exact sequences were predicted commonly, due to the differences in peptide length and algorithms used by the developed pipelines. However, predicted allergenic sequences from *malus domestica* and *gallus gallus* were identified with both pipelines and

Virus name	Protein	Viral sequence	Allergen	Protein	Allergen sequence
SARS-CoV2	Replicase_polyprotein	RLQAGNATEVPANST	<i>Glycine max</i>	Beta-conglycinin protein	LQSGDALRVPAGTTF
SARS-CoV2	Spike_glycoprotein	GINITRFQTLALHR	<i>Glycine max</i>	Beta-conglycinin protein	LQSGDALRVPAGTTY
SARS-CoV2	Spike_glycoprotein	LPIGINITRFQTLA	<i>Glycine max</i>	Beta-conglycinin protein	LQSGDALRVPAGTTY
SARS-CoV2	Replicase_polyprotein	DEGNCDTLKEILVTY	<i>Crangon crangon</i>	Iosephosphate isomerase	PCIGEKLDRESNRT
SARS-CoV2	Replicase_polyprotein	LQAGNATEVPANSTV	<i>Crangon crangon</i>	Iosephosphate isomerase	PCIGEKLDRESNRT
HRSV A (strain A2)	RNA-directed RNA polymerase L	RLMEGQTHAQADYLL	<i>Malus domestica</i>	Predicted protein (E4Z8N9)	RLFARTRQVESLTAE
SARS-CoV2	Replicase_polyprotein	RLQAGNATEVPANST	<i>Malus domestica</i>	Predicted protein (E4Z8N9)	RLFARTRQVESLTAE
SARS-CoV2	Replicase_polyprotein	RLQAGNATEVPANST	<i>Malus domestica</i>	Putative COBL7 (COBRA-LIKE 7)	RLFARTRQVESLAAE
HRSV A (strain A2)	Non-structural protein 2	CIVRKLDERQATFTF	<i>Crangon crangon</i>	Sarcoplasmic calcium-binding protein	AGGINIARYQELYAQ
SARS-CoV2	Replicase_polyprotein	DEGNCDTLKEILVTY	<i>Crangon crangon</i>	Sarcoplasmic calcium-binding protein	GINIARYQELYAQFI
HRSV A (strain A2)	Non-structural protein 2	CIVRKLDERQATFTF	<i>Crangon crangon</i>	Sarcoplasmic calcium-binding protein	KAGGINIARYQELYA
SARS-CoV2	Membrane protein	RLFARTRSMWSFNPE	<i>Crangon crangon</i>	Sarcoplasmic calcium-binding protein	AGGINIARYQELYAQ
SARS-CoV2	Spike_glycoprotein	PIGINITRFQTLAL	<i>Crangon crangon</i>	Sarcoplasmic calcium-binding protein	GGINIARYQELYAQF
SARS-CoV2	Membrane protein	RLFARTRSMWSFNPE	<i>Crangon crangon</i>	Sarcoplasmic calcium-binding protein	GINIARYQELYAQFI
SARS-CoV2	Replicase_polyprotein	DEGNCDTLKEILVTY	<i>Gallus gallus</i>	Serum albumin	DHGEADFLKSILIRY
SARS-CoV2	Spike_glycoprotein	GINITRFQTLALHR	<i>Gallus gallus</i>	Serum albumin	DHGEADFLKSILIRY
SARS-CoV2	Spike_glycoprotein	GINITRFQTLALHR	<i>Malus domestica</i>	Two-component response regulator	MKGVTHGACDYLIKIP

**Table 3.** Cross-reactive peptide peptides predicted between viral and allergen sequences using pipeline-2.

sequences from *Gallus gallus* were even predicted to potentially cross-react with viral epitopes from SARS-CoV-2 by both approaches.

## Discussion

We identified several potentially cross-reactive T cell epitope pairs between food allergens and epidemiologically relevant respiratory viruses, using two independent in-silico pipelines. To our knowledge, this is the first study investigating heterologous immunity between T cell epitopes among several clinically relevant RNA viruses and food allergens. Our study revealed allergenic sequences from *Malus domestica* to be important for cross-reactivity. Further, we found that the epitope sequences from SARS-CoV-2 and RSVA2 were those predicted to most likely cross-react with food allergens, as T cell epitope pairs were predicted by both pipelines. Finally, we observed more candidate epitope pairs for HLA class I compared to HLA class II.

In pipeline-1, we made use of several T cell epitope prediction tools followed by alignments based on sequence homology and amino acid properties. The latter is particularly important for detection of cross-reactivity, which is known to also exist between structurally unrelated antigens with little sequence homology<sup>19</sup>. In addition, it is known that the position of specific amino acids in their non-anchor region is critical for T cell epitope immunogenicity<sup>20</sup>. A limitation of pipeline-1 is the lack of such tools for prediction of immunogenicity, which is compensated by pipeline-2. Indeed, the latter utilized the IEDB immunogenicity prediction tool for this purpose<sup>20</sup>.

We applied pipeline-2 to identify only known T cell epitopes in combination with alignment against respiratory bacteria in order to avoid prediction contaminations. Hence, this restriction led to the lack of any identified epitope pairs for RV1b and influenza strains containing the seasonal quadrivalent influenza vaccine. The allergen sequences among the short-listed epitope pairs for SARS-CoV-2 and RSVA2 were derived from different allergen sources, including apple and chicken. Importantly, using pipeline-1, several candidate epitope pairs involving SARS-CoV-2 contained chicken allergens as published previously<sup>18</sup>.

In our study, we did not observe identical peptide sequence pairs, commonly predicted by both applied pipelines. The reason for non-overlapping sequence pairs is the use of independent sequence features by the applied pipelines such as associated immunogenicity and peptide lengths. Moreover, several predicted peptide pairs have not been previously reported and may represent novel sets of epitopes. Overall, these two independent algorithms predicted an exhaustive set of non-redundant epitope pairs that require further experimental investigation.

Of note, we also predicted epitope pairs for RSVA2 on the background of human HLA class II alleles. Interestingly, no candidate pairs could be identified using RV1b or influenza proteins. In addition, we were able to identify candidate pairs for SARS-CoV-2 but with a 10× lower score compared to HLA class I-restricted epitopes<sup>18</sup>. However, no HLA class II candidates were predicted with pipeline-2.

Aerosolized food allergens may trigger asthma symptoms via inhalation and subsequent inflammatory response in the lung<sup>7</sup>. Such cases have been shown for fish allergens (33) and soy proteins<sup>21</sup>. Importantly, we

predicted T cell epitope sequences from RSVA2 and SARS-CoV-2 which appeared to be cross-reactive with sequences from *Crangon crangon*, a shrimp, as well as sequences from RSVA2 which showed cross-reactivity with sequences from soybean.

Our study provides *in-silico* data, which support a yet unexplored pathogenic mechanism for the connection between food allergy and virus-associated asthma. T cell-mediated heterologous immune responses are an important determinant for the final outcome of an infection or allergen exposure and can lead to protective responses or immunopathology, also depending on the balance between antigen load and efficiency of effector T cells. Immunomodulatory effects can thus be elicited by cross-reactive antigens, which may lead to an expansion of T memory cells and in turn to a modified T cell memory pool, changes in patterns of immunodominance and an altered hierarchy of T cell responses<sup>22</sup>. The final outcome depends not only on private specificities of TCR repertoires but also on the host's HLA background.

Future studies using PBMCs from individuals with relevant food allergies may be the first step to validate cross-reactivity involving the predicted T cell epitope pairs. In order to address the diversity of prevalent HLA alleles in distinct ethnic populations, ideally experiments should involve pools of most frequently identified potentially cross-reactive peptides as well as large groups of individuals with distinct HLA haplotypes. Results of such studies may advise peptide immunization strategies for a favorable outcome in the context of allergy and infection.

## Materials and methods

We used our previously published *in-silico* analysis for prediction of potentially cross-reactive T cell epitope pairs between viral and allergenic proteins, applying two independent pipelines. We downloaded all available food allergens from Allergen Online (10.09.2017)<sup>23–25</sup>, as well as protein sequences from the most clinically relevant respiratory viruses SARS-CoV-2, RSVA2, RV1b and influenza strains of the seasonal quadrivalent influenza vaccine 2019/2020 (QIV19/20) (Uniprot and GISAID). Protein sequences are provided in Supplementary Tables 1 and 2.

Pipeline-1 was performed as described previously<sup>18</sup>, using highly prevalent human HLA alleles for Caucasian populations (Supplementary Table 3). Briefly, viral T cell epitopes were predicted using *smm*<sup>26</sup>, *ann*<sup>27</sup> and consensus<sup>28</sup> algorithm tools for MHCI (IC50 threshold  $\leq 5000$  nm), and *netMHCII*<sup>29</sup> for MHCII. Allergenic proteins were aligned against predicted viral epitopes (NCBI protein blast platform)<sup>30</sup> and subsequently used for T cell epitope prediction by *netMHC*<sup>31</sup> and *netMHCpan*<sup>32</sup> for MHCI, and *netMHCII* and *netMHCIIpan*<sup>33</sup> for MHC class II prediction. Viral and allergen epitopes predicted by all methods were pairwise aligned with a Biopython module pairwise<sup>24</sup>. A final pair combined score was calculated, taking the binding affinity of predicted viral and allergen epitopes to MHC molecules into consideration, as well as the score from the pairwise alignment and cross-entropy ( $Pair\ Combined\ Score = \frac{1}{binding\ affinity(nM)(Virus)} \times score\ PwA \times \frac{1}{binding\ affinity(nM)(Allergen)}$ ).

Allergenic epitopes of the top 30 candidate pairs further underwent a comprehensive scoring system taking clinical relevance and conservation criteria into consideration. Epitope pairs were finally ranked for the top 5 candidates for each virus based on the final score. A schematic overview of the pipeline is depicted in Supplementary Fig. 1A. The scoring system is shown in Supplementary Fig. 2.

In addition to the above pipeline-1, we also performed an analysis using a modified version of the independent pipeline we published previously (Supplementary Fig. 1B). We used IEDB database, which hosts known epitope peptide sequences, to predict peptides that are known to bind MHC molecules. To identify cross-reactive antigenic peptides between the given food allergenic protein sequences and viral proteomes, we split each of a given protein sequence into a set of sequential *k-mers* or peptides (length = 15). To avoid contaminations in the downstream analysis, we filtered out the sequential *k-mers* that mapped with bacterial protein sequences, using *blastp* (e-value < 1 and identity > 70% and coverage > 70%). Here, eight bacterial species (*Moraxella catarrhalis*, *Chlamydomphila pneumoniae*, *Mycoplasma pneumoniae*, *Coxiella burnetii*, *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Streptococcus pyogenes* and *Legionella pneumophila*) were selected that are known to cause common respiratory infections. Subsequently, using the IEDB immunogenicity prediction tool<sup>20</sup>, we identified from the remaining sequential *k-mers* those that can potentially form peptide MHC (pMHC) complexes. This tool uses the properties of amino acids within a given peptide to predict its potential to form pMHC complexes. The above steps filtered out a large proportion of sequential *k-mers* to retain only 134,171 high confidence peptide sequences. However, it is likely that some of the peptides may not have a strong MHC-I binding affinity despite the homology, and therefore may be less likely to be presented as antigens by HLA molecules. Therefore, using IEDB tools<sup>35</sup>, the homologous peptides were further evaluated for their binding affinity with human MHC-I molecules for a broad range of alleles ( $n = 54$ ) (Supplementary Table 4a,b). The peptides with MHC binding affinity rank > 50 (arbitrarily selected threshold to retain at most only 50% of the total number of sequences) were further selected for homology analysis. Wherein, using the *blastp* (e-value < 1 and identity > 50% and coverage > 70%), we identified food allergen- and virus- peptides that share homologous sequences (Supplementary Table 4c) and which were considered as potentially cross-reactive T cell epitopes. All utilized algorithms/prediction tools in this study were accessed online rather from a local server.

We evaluated the statistical significance of predicted peptides (from method-2) by comparing the cross-reactive peptides with 1000 randomly generated sequences using BLASTp program. The random peptide sequences were generated using RSAT webserver ([http://rsat.sb-roscoff.fr/random-seq\\_form.cgi](http://rsat.sb-roscoff.fr/random-seq_form.cgi)). We measured the statistical significance, by calculating the p-value, as the number of times each of the predicted cross-reactive peptide matches with randomly generated sequences, using following equation.

$$p = n/1000.$$



Here,  $n$  = number of times a cross-reactive peptide shares sequence identity > 30% with random peptide sequences and  $e$ -value < 1.0 with the set of 1000 random peptide sequences.  $P$ -value < 0.05 was considered as statistically significant, to support alternative hypothesis that cross-reactive peptide sequence is significantly different from random sequences. Here, the low sequence identity threshold was used to filter out any cross-reactive peptide that matches with randomly generated peptide sequences.

## Data availability

All data generated or analysed during this study are included in this published article and its supplementary information files.

Received: 6 March 2023; Accepted: 23 August 2023

Published online: 08 September 2023

## References

- Lambrecht, B. N. & Hammad, H. The immunology of asthma. *Nat. Immunol.* **16**, 45–56 (2015).
- Foong, R.-X., Du Toit, G. & Fox, A. T. Asthma, food allergy, and how they relate to each other. *Front. Pediatr.* **5**, 89 (2017).
- Cianferoni, A. Non-IgE mediated food allergy. *Curr. Pediatr. Rev.* **16**, 95–105 (2020).
- Labrosse, R., Graham, F. & Caubet, J.-C. Non-IgE-mediated gastrointestinal food allergies in children: An update. *Nutrients* **12**, 2086 (2020).
- Hill, D. A. *et al.* The epidemiologic characteristics of healthcare provider-diagnosed eczema, asthma, allergic rhinitis, and food allergy in children: A retrospective cohort study. *BMC Pediatr.* **16**, 133 (2016).
- Emons, J. A. M. & van Gerth, W. R. Food allergy and asthma: Is there a link?. *Curr. Treat. Opt. Allergy* **5**, 436–444 (2018).
- Foong, R.-X., Du Toit, G. & Fox, A. T. Mini review—Asthma and food allergy. *Curr. Pediatr. Rev.* **14**, 164–170 (2018).
- Heymann, P. W. *et al.* Viral infections in relation to age, atopy, and season of admission among children hospitalized for wheezing. *J. Allergy Clin. Immunol.* **114**, 239–247 (2004).
- Cheung, D. S. & Grayson, M. H. Role of viruses in the development of atopic disease in pediatric patients. *Curr. Allergy Asthma Rep.* **12**, 613–620 (2012).
- Fecek, R. J. *et al.* Enteric reovirus infection stimulates peanut-specific IgG2a responses in a mouse food allergy model. *Immunobiology* **215**, 941–948 (2010).
- Chen, X. *et al.* Characterization of intestinal dendritic cells in murine norovirus infection. *Open Immunol. J.* **4**, 22–30 (2011).
- Skevakis, C. *et al.* Influenza-derived peptides cross-react with allergens and provide asthma protection. *J. Allergy Clin. Immunol.* **142**, 804–814 (2018).
- Montemurro, A. *et al.* NetTCR-2.0 enables accurate prediction of TCR-peptide binding by using paired TCR $\alpha$  and  $\beta$  sequence data. *Commun. Biol.* **4**, 1060 (2021).
- Maurer-Stroh, S. *et al.* AllerCatPro-prediction of protein allergenicity potential from the protein sequence. *Bioinformatics* **35**, 3020–3027 (2019).
- Jaravine, V. *et al.* Expitope 2.0: A tool to assess immunotherapeutic antigens for their potential cross-reactivity against naturally expressed proteins in human tissues. *BMC Cancer* **17**, 892 (2017).
- Qiu, T. *et al.* CE-BLAST makes it possible to compute antigenic similarity for newly emerging pathogens. *Nat. Commun.* **9**, 1772 (2018).
- Hobohm, U. & Meyerhans, A. A pattern search method for putative anchor residues in T cell epitopes. *Eur. J. Immunol.* **23**, 1271–1276 (1993).
- Balz, K. *et al.* Homologies between SARS-CoV-2 and allergen proteins may direct T cell-mediated heterologous immune responses. *Sci. Rep.* **11**, 4792 (2021).
- Welsh, R. M. *et al.* Heterologous immunity between viruses. *Immunol. Rev.* **235**, 244–266 (2010).
- Calis, J. J. A. *et al.* Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput. Biol.* **9**, e1003266 (2013).
- Antó, J. M. *et al.* Community outbreaks of asthma associated with inhalation of soybean dust. Toxicoepidemiological Committee. *N. Engl. J. Med.* **320**, 1097–1102 (1989).
- Selin, L. K. *et al.* CD8 memory T cells: Cross-reactivity and heterologous immunity. *Semin. Immunol.* **16**, 335–347 (2004).
- Mari, A., Scala, E. Allergome: A unifying platform. *Arb. Paul. Ehrlich. Inst. Bundesamt. Sera Impfstoffe Frankf. A. M.* **29–39** (2006) (discussion 39–40).
- Mari, A. *et al.* Allergen databases: Current status and perspectives. *Curr. Allergy Asthma Rep.* **9**, 376–383 (2009).
- Mari, A. *et al.* Bioinformatics applied to allergy: Allergen databases, from collecting sequence information to data integration. The Allergome platform as a model. *Cell Immunol.* **244**, 97–100 (2006).
- Peters, B. & Sette, A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinform.* **6**, 132 (2005).
- Buus, S. *et al.* Sensitive quantitative predictions of peptide-MHC binding by a “Query by Committee” artificial neural network approach. *Tissue Antigens* **62**, 378–384 (2003).
- Moutafsi, M. *et al.* A consensus epitope prediction approach identifies the breadth of murine T(CD8+)-cell responses to vaccinia virus. *Nat. Biotechnol.* **24**, 817–819 (2006).
- Nielsen, M. & Lund, O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinform.* **10**, 296 (2009).
- Camacho, C. *et al.* BLAST+: Architecture and applications. *BMC Bioinform.* **10**, 421 (2009).
- Nielsen, M. *et al.* Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci.* **12**, 1007–1017 (2003).
- Hoof, I. *et al.* NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* **61**, 1–13 (2009).
- Andreatta, M. *et al.* Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. *Immunogenetics* **67**, 641–650 (2015).
- Cock, P. J. A. *et al.* Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
- Reynisson, B. *et al.* NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* **48**, W449–W454 (2020).

## Acknowledgements

CS is supported by the Universities Giessen and Marburg Lung Center (UGMLC), the German Center for Lung Research (DZL), University Hospital Giessen and Marburg (UKGM) research funding according to

article 2, section 3 cooperation agreement, and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)-Project-ID 197785619—SFB 1021, KFO 309 (P10), and SK 317/1-1 (Project number 428518790) as well as by the Foundation for Pathobiochemistry and Molecular Diagnostics. This work was also supported by the NIAID AADCRC U19AI057229 (PI Davis), U19AI104209 (PI Galli), U01AI140498 (PI Nadeau), NIEHS R21 ES033049 (PI Nadeau), Sean N. Parker Center for Allergy and Asthma Research at Stanford University and Crown Foundation. We would like to thank L. England, P. Nelson, M. Bruhn for their help with the in-silico pipeline-1 and associated scoring system. We would also like to thank A. Hensel for his administrative support with the submission and revision of the manuscript.

### Author contributions

F.C. and V.H. performed the in-silico analysis of pipeline-1, K.B. further processed and evaluated the data. A.K. performed the in-silico analysis of pipeline-2. C.S. planned the study. All authors wrote or carefully reviewed sections of the manuscript.

### Funding

Open Access funding enabled and organized by Projekt DEAL.

### Competing interests

For CS: Consultancy and research funding, Hycor Biomedical, Bencard Allergie and Thermo Fisher Scientific; Research Funding, Mead Johnson Nutrition (MJN). For KC: Dr. Nadeau reports grants from National Institute of Allergy and Infectious Diseases (NIAID), National Heart, Lung, and Blood Institute (NHLBI), National Institute of Environmental Health Sciences (NIEHS), and Food Allergy Research & Education (FARE); stock options from IgGenix, Seed Health, ClostraBio, and ImmuneID; is Director of the World Allergy Organization Center of Excellence for Stanford, Advisor at Cour Pharma, Consultant for Excellergy, Red tree ventures, Eli Lilly, and Phylaxis, Co-founder of Before Brands, Alladapt, Latitude, and IgGenix; and National Scientific Committee member at Immune Tolerance Network (ITN), and National Institutes of Health (NIH) clinical research centers, outside the submitted work; patents include, “Mixed allergen composition and methods for using the same,” “Granulocyte-based methods for detecting and monitoring immune system disorders,” and “Methods and Assays for Detecting and Quantifying Pure Subpopulations of White Blood Cells in Immune System Disorders.”

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-41187-1>.

**Correspondence** and requests for materials should be addressed to C.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023