scientific reports

Check for updates

OPEN

Homologies between SARS-CoV-2 and allergen proteins may direct T cell-mediated heterologous immune responses

Kathrin Balz[1,4], Abhinav Kaushik[2,4], Meng Chen[2], Franz Cemic[3], Vanessa Heger[3], Harald Renz[1], Kari Nadeau[2,4] & Chrysanthi Skevaki[1,4]✉

The outbreak of the new severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) is a public health emergency. Asthma does not represent a risk factor for COVID-19 in several published cohorts. We hypothesized that the SARS-CoV-2 proteome contains T cell epitopes, which are potentially cross-reactive to allergen epitopes. We aimed at identifying homologous peptide sequences by means of two distinct complementary bioinformatics approaches. Pipeline 1 included prediction of MHC Class I and Class II epitopes contained in the SARS-CoV-2 proteome and allergens along with alignment and elaborate ranking approaches. Pipeline 2 involved alignment of SARS-CoV-2 overlapping peptides with known allergen-derived T cell epitopes. Our results indicate a large number of MHC Class I epitope pairs including known as well as de novo predicted allergen T cell epitopes with high probability for cross-reactivity. Allergen sources, such as *Aspergillus fumigatus*, *Phleum pratense* and *Dermatophagoides* species are of particular interest due to their association with multiple cross-reactive candidate peptides, independently of the applied bioinformatic approach. In contrast, peptides derived from food allergens, as well as MHC class II epitopes did not achieve high in silico ranking and were therefore not further investigated. Our findings warrant further experimental confirmation along with examination of the functional importance of such cross-reactive responses.

Abbreviations

ACE2	Angiotensin converting enzyme 2
ARDS	Acute respiratory distress syndrome
BLAST	Basic local alignment search tool
HLA	Human leukocyte antigen
IEDB	Immune epitope database
MHC	Major histocompatibility complex
NCBI	National Center for Biotechnology Information
RNA	Ribonucleic acid
+ssRNA	Positive-sense single-stranded RNA
Th 1	T-helper 1
Th 2	T-helper 2

The World Health Organization (WHO) has declared the outbreak of the new Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2, ssRNA virus, associated with COVID-19) as a public health emergency. As per the WHO report of 20 September 2020, more than 30 million cases and over 950,000 deaths have been reported worldwide[1]. Human coronaviruses are positive-sense single-stranded RNA (+ssRNA) viruses, with SARS-CoV-2 and SARS-CoV belonging to the B-lineage of the Betacoronavirus genera and MERS-CoV to the C-lineage of the

[1]Institute of Laboratory Medicine, Universities of Giessen and Marburg Lung Center (UGMLC), Philipps University Marburg, German Center for Lung Research (DZL), Marburg, Germany. [2]Division of Pulmonary, Allergy and Critical Care Medicine, Sean N. Parker Center for Allergy and Asthma Research at Stanford University, Stanford, CA, USA. [3]TH Mittelhessen, Department of Computer Science, University of Applied Sciences Gießen, Gießen, Hessen, Germany. [4]These authors contributed equally: Kathrin Balz, Abhinav Kaushik, Kari Nadeau and Chrysanthi Skevaki. ✉email: chrysanthi.skevaki@uk-gm.de

same genera[2,3]. The clinical features in patients affected with these respiratory viruses ranges from asymptomatic carriers to severe respiratory illness with pneumonia and acute respiratory distress syndrome (ARDS). In addition, a number of interesting vascular and inflammatory presentations have been noted, including a multisystem inflammatory syndrome in children.

We have previously reported on heterologous immune responses induced by influenza, another respiratory RNA virus, against allergens, which mediated protection from experimental allergic asthma[4]. Indeed, virus-induced T cell mediated heterologous immunity has been widely described in a variety of settings, which can confer protection or drive immunopathology against other antigens[5,6]. Given that the host immune response to SARS-CoV-2 and associated disease course can be so varied from patient to patient, this spectrum of presentations raises the question of what drives the differential host immune response. There is still little known about asthma phenotypes and severity of COVID-19. In general, asthma has not been shown to be a risk factor for COVID-19 in several published cohorts[7,8]. However, recent studies from the UK and the USA indicated higher numbers of asthmatics in COVID-19 patients[9].

Interestingly, the UK Biobank recently reported that non-allergic patients had a higher risk of severe COVID-19, compared to patients with allergic asthma[10]. Moreover, evidence of T cell activation, as indicated by generation of effector ($CD45RA^{-}CD62L^{-}$) and central memory ($CD45RA^{-}$ $CD62L^{+}$) $\alpha\beta CD4_{conv}$ and $CD8^{+}$ cells was reported among COVID-19 patients with mild or severe disease, suggesting that activation of T cells is inversely associated with severity of SARS-CoV-2 infection[11]. Other investigators reported that T cells in peripheral blood of COVID-19 patients with allergy were increased as compared to patients without allergies[12]. These preliminary clinical and laboratory observations along with our prior experimental evidence involving RNA viruses led us hypothesize that SARS-CoV-2 may share a degree of protein sequence homology to allergens, which may lead to the generation of cross-reactive T cell epitopes. Pre-existing T cells specific for such cross-reactive allergen-derived epitopes may have an impact on COVID-19 outcome via aberrant cytokine responses to the virus peptides. Indeed, these cytokines could prevent an overshooting T1 inflammatory reaction, both locally (as in the case of preexisting pulmonary $CD4^{+}$ T cells specific to inhalant allergens) and/or systemically. Therefore, we sought to predict potentially cross-reactive allergen- and SARS-CoV-2-derived MHC Class I and Class II T cell epitopes, which can be presented by the most prevalent HLA alleles.

Methods and results

In order to examine our working hypothesis, we applied two distinct independent, complementary and systematic bioinformatics approaches (Fig. 1): (a) Pipeline 1-prediction of MHC Class I and Class II epitopes contained in the SARS-CoV-2 proteome and a comprehensive set of allergen protein sequences combined with alignment strategies and ranking of results based on clinical and sequence conservation criteria and (b) Pipeline 2- alignment of SARS-CoV-2 overlapping peptides with known allergen-derived T cell epitopes[13].

Pipeline 1. More than > 2500 allergen protein sequences were downloaded (dates of access 10.09.2017) from Allergome[14–17] (Supplementary Table S1), and protein sequences for SARS-CoV-2 from UniProt[18] (Supplementary Table S2). Viral T cell epitope prediction was performed using smm[19], ann[20] and consensus[21] for MHC Class I (IC50 threshold < = 5000), and netMHCII[22] for MHC Class II (affinity score threshold for strong binders: 0.500; for weak binders: 2.000) (Supplementary Table S3). Epitopes predicted by all methods were aligned against all allergen proteins using a local version of the NCBI protein blast program[23]. Allergen proteins associated with an alignment e-value < 10 were further processed for T cell epitope prediction using netMHC[24] and netMHCpan[25] for MHC Class I, and netMHCII and netMHCIIpan[26] for MHC Class II prediction (affinity score threshold for strong binders: 0.500; for weak binders: 2.000). Viral and allergen epitopes were pairwise aligned with Biopython module pairwise 2[27] and for pairs with a score > 8, a final pair combined score (pcs) was calculated (Supplementary Methods). A higher pair combined score indicates a higher chance for a peptide pair to be cross-reactive and binding to MHC molecules. Duplicates among the resulting candidate epitope pairs were removed before further processing. Therefore, possible sequence repetition due to isoforms and isoallergens (Supplementary Table S1) do not influence further analyses. In total, we obtained more than 5000 candidate pairs for each, MHC Class I and Class II. The top 30 candidate epitope pairs, as per pair combined score, are listed for aero- and food allergens, MHC Class I and Class II presentation background in Supplementary Tables S4–S7, respectively. The top 30 MHC Class II restricted predicted virus-allergen pairs achieved relatively low pcs (24–657) as compared to Class I epitope pairs (1036–10,816). Among our top 30 MHC Class I potentially cross-reactive allergen derived epitopes, we identified more than 20 distinct protein families (Allfam database). In addition to MHC binding affinity and homology between peptide sequences, also other factors (e.g. conservation, association with clinical reactions) are important for the clinical relevance of peptides predicted to be cross-reactive at the T cell level. In order to capture this information level in our ranking, all allergen peptides and associated sources listed among the top 30 candidate epitope pairs were evaluated further with a scoring system (Supplementary Fig. S1 and Supplementary Methods). We found that the top 5 Class I aeroallergens were on average associated with higher pcs as compared to the top 5 potentially cross-reactive food allergens (Table 1 for MHC Class I and Table 2 for MHC Class II peptide pairs).

Pipeline 2. We obtained all known allergen-derived linear T cell epitope peptides from the IEDB, containing peptides known to bind MHC molecules with at least one published experimental evidence (e.g. based on the results of a T cell assay) (Supplementary Table S8). A total of 8207 antigenic peptides from 142 antigens were selected for evaluation, among which, peptides with ambiguous amino acids (e.g. with unknown amino acid 'X' or any special character) were removed from the subsequent analysis. Therefore, all included peptides could be defined in full. Next, SARS-CoV-2 protein sequences were analyzed for the potential antigenic regions by split-

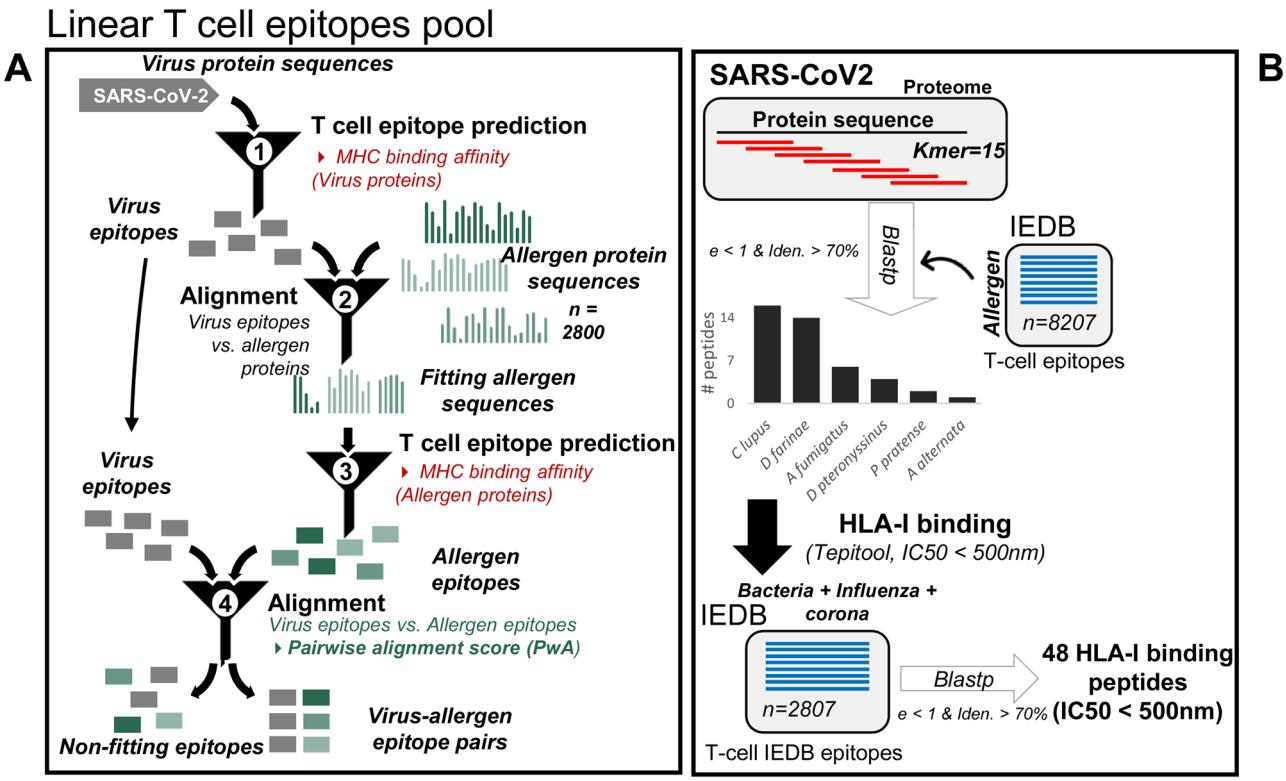


Figure 1. Schematic overview of the bioinformatics approaches. **A** Pipeline 1; SARS-CoV-2 proteins were aligned against > 2500 allergen protein sequences (see methods) and MHC class I-and II- restricted potentially cross-reactive T cell epitope pairs were identified for the most frequent human HLA alleles. **B** Pipeline 2; In an independent framework, we performed the comparative analysis of sequential *kmers* from SARS-CoV-2 protein sequences with known IEDB allergen peptides to predict the cross-reactive viral peptide pool.

MHC I		Top 30 pair combined score					
		Candidate	Allergen epitope	Protein family	Viral epitope	MHC allele	pcs
AERO	Nr.1	Mus m 1	GSNTFTILK	Lipocalin	TSNSFDVLK	HLA-A_11_01	8188
	Nr. 2	Asp f 5	MLYEVLWNL	Fungalysin metalloprotease	YLYALVYFL	HLA-A_02_01	5974
	Nr. 3	Aln g 1	SGVSPVSYQK	Bet v 1 family	ATSRTLSYYK	HLA-A_11_01	2413
	Nr.4	Phl p 5	KYKTFVATF	Group 5/6 grass pollen allergen	MFDAYVNTF	HLA-A_24_02	1910
	Nr. 5	Mus m 1	GSNTFTILK	Lipocalin	VTNNTFTLK	HLA-A_11_01	3286
FOOD	Nr. 1	Gal d 5	FLGHFIYSV	Serum albumin	TMADLVYAL	HLA-A_02_01	1408
	Nr. 2	Gal d 6	YLLDLLPAA	Lipoprotein	TLMNVLTLV	HLA-A_02_01	4185
	Nr. 3	Gal d 6	RPAYRRYLL	Lipoprotein	RPPLNRNYV	HLA-B_07_02	2274
	Nr. 4	Cor a 1	APHGGGSIL	Bet v 1 family	VPGLPGTIL	HLA-B_07_02	2571
	Nr. 5	Gal d 6	KVFRFSMFK	Lipoprotein	LVASIKNFK	HLA-A_11_01	1194

Table 1. The Top 5 candidate human HLA class I T cell potentially cross-reactive epitope pairs between SARS-CoV-2 and aero-and food-allergens based on pair combined score and application of additional clinical and conservation related criteria (see Suppl. Fig. 1) (pipeline 1). Pcs = pair combined score.

ting each of the sequence into sequential *k-mers* (length = 15), and homology with allergen antigenic peptides was then profiled. Within a given threshold range, we found 43 unique SARS-CoV-2 peptides that belong to replicase poly protein and spike glycoprotein (Supplementary Table S9). These peptides demonstrate homology with antigenic peptides of 6 different allergen sources, including *Canis lupus*, *Dermatophagoides farinae*, *Dermatophagoides pteronyssinus*, *Aspergillus fumigatus*, *Alternaria alternata* and *Phleum pratense*, all of which are known to be respiratory allergens and, in the majority, clinically highly relevant (e.g. aeroallergens; Fig. 1).

However, despite the homology, it is likely that some of the peptides may not have strong MHC Class I binding affinity, and thus be less likely to be presented as antigens by HLA molecules. Therefore, we assessed the binding affinity of these peptides with human MHC Class I molecules, across a broad range of alleles that are known to bind viral proteins (52 most common HLA-A and HLA-B alleles). We observed that some of these peptides

MHC II		Top 30 pair combined score					
		Candidate	Allergen epitope	Protein family	Viral epitope	MHC allele	pcs
AERO	Nr.1	Phl p 5	FVATFGAAS	Group 5/6 grass pollen allergen	FSSTFNVPM	HLA-DRB1_04_01	87
	Nr. 2	Asp f 4	LTALAAGSA	Unclassified	VTALRANSA	HLA-DRB1_01_01	479
	Nr. 3	Phl p 5	FVATFGAAS	Group 5/6 grass pollen allergen	FSSTFNVPM	HLA-DRB1_04_01	53
	Nr.4	Phl p 5	FVATFGPAS	Group 5/6 grass pollen allergen	FSSTFNVPM	HLA-DRB1_04_01	29
	Nr. 5	Phl p 5	FKVAATAAN	Group 5/6 grass pollen allergen	FSSTFNVPM	HLA-DRB1_04_01	38
FOOD	Nr. 1	Gal d 5	FLYAPAILS	Serum albumin	FYILPSIIS	HLA-DRB1_01_01	299
	Nr. 2	Gal d 6	ILVDAVLKE	Lipoprotein	VVADAVIKT	HLA-DRB1_03_01	113
	Nr. 3	Gal d 6	VYSDVPIEK	Lipoprotein	VVADAVIKT	HLA-DRB1_03_01	29
	Nr. 4	Ara h 1	FIMPAAHPV	Cupin	FVMMSAPPA	HLA-DRB1_01_01	258
	Nr. 5	Gal d 5	FLYAPAILS	Serum albumin	FLYENAFLP	HLA-DRB1_01_01	81

Table 2. The Top 5 candidate human HLA class II T cell potentially cross-reactive epitope pairs between SARS-CoV-2 and aero-and food-allergens based on pair combined score and application of additional clinical and conservation related criteria (see Fig. 1) (pipeline 1) Pcs = pair combined score.

Allele	HLA-I-Binding Peptide	IC50	SARS-CoV-2 protein name
HLA-A*68:01	NIFGTVYEK	6	R1AB_SARS2_Replicase_polyprotein
HLA-A*02:06	YTVELGTEV	9.4	R1A_SARS2_Replicase_polyprotein
HLA-A*68:02	YTVELGTEV	10.8	R1A_SARS2_Replicase_polyprotein
HLA-B*15:03	LASHMYCSF	10.8	R1A_SARS2_Replicase_polyprotein
HLA-B*40:02	HEGKTFYVL	11	SPIKE_SARS2_Spike_glycoprotein
HLA-B*40:01	GETLPTEVL	11.9	R1AB_SARS2_Replicase_polyprotein
HLA-A*02:06	TVYEKLKPV	13.4	R1AB_SARS2_Replicase_polyprotein
HLA-A*30:02	ASHMYCSFY	13.9	R1A_SARS2_Replicase_polyprotein
HLA-B*40:01	HEGKTFYVL	13.9	SPIKE_SARS2_Spike_glycoprotein
HLA-A*11:01	NIFGTVYEK	24	R1AB_SARS2_Replicase_polyprotein
HLA-B*35:01	LASHMYCSF	24.5	R1A_SARS2_Replicase_polyprotein
HLA-A*68:02	TVYEKLKPV	26	R1AB_SARS2_Replicase_polyprotein
HLA-A*02:01	WLTNIFGTV	34.1	R1AB_SARS2_Replicase_polyprotein
HLA-A*02:06	WLTNIFGTV	34.7	R1AB_SARS2_Replicase_polyprotein
HLA-B*15:03	LTNIFGTVY	35.7	R1AB_SARS2_Replicase_polyprotein
HLA-B*15:25	LASHMYCSF	39.1	R1A_SARS2_Replicase_polyprotein
HLA-B*15:25	LTNIFGTVY	39.8	R1AB_SARS2_Replicase_polyprotein
HLA-A*02:01	TVYEKLKPV	47.8	R1AB_SARS2_Replicase_polyprotein
HLA-B*15:01	LASHMYCSF	48.1	R1A_SARS2_Replicase_polyprotein
HLA-B*15:03	ASHMYCSFY	49	R1A_SARS2_Replicase_polyprotein

Table 3. HLA-I binding high confidence (IC50 < 50 nm) SARS-CoV-2 antigenic peptides (pipeline 2).

(n = 79) were predicted to have MHC Class I binding epitope regions associated with at least one of the Class I HLA alleles with IC50 < 500 nm (Supplementary Table S10). These antigenic peptides were predicted to bind with 20 most frequently occurring HLA Class I alleles, in which HLA*02:03 and HLA*02:06 were predicted to present the highest number of epitope residues. To further investigate if these peptides are specific to the coronavirus family, we performed the BLAST comparison with 2807 known viral antigenic peptides of bacteria, influenza- and corona- virus family (non-SARS CoV-2) from IEDB (with at least one T cell assay evidence) and filtered out matching peptides (Blast e-value < 1 and identity > 70%). Finally, we present 48 high-affinity HLA-binding peptides which are unique to the SARS-CoV-2 proteome, not common to bacteria, influenza and corona virus family antigenic peptides within a given threshold range (Supplementary Table S11) with 14 high confidence HLA Class I binding peptides with IC50 < 50 nm (Table 3).

Application of both complementary pipelines aimed at identifying T cell epitope pairs, which are highly likely to be cross-reactive. Pipeline 1 includes a broader approach by means of considering as many allergen protein sequences are available and subsequently predicting MHC binding affinity and performing alignment of the candidate epitopes. The elaborate scoring system which followed, prioritized these candidates based on clinical aspects and thus relevance. This pipeline has already been used and experimentally validated by our group for similar analyses (Balz K. et al., unpublished data). Nevertheless, immunogenicity and cross-reactivity of the peptides identified by pipeline 1 in our index work remains to be shown. Pipeline 2 takes only known

immunogenic peptides into consideration, leading to more robust results. Further, alignment of these epitopes against the proteome of other organisms, allows identification of epitopes, which are unique for SARS-CoV-2. The approach of pipeline 2, however, does not allow identification of newly described T cell epitopes and less studied allergens are not considered.

Discussion

We have applied two independent, complementary and systematic bioinformatic approaches in order to identify potentially cross-reactive allergen- and SARS-CoV-2- T cell epitopes. Our in silico analysis revealed numerous candidate epitope pairs, including previously published and predicted peptides, while both applied pipelines highlighted an important role of MHC class I inhalant allergens. Epitope pairs including peptides from food allergens appeared to be of lower importance. Our finding may indicate that patients with respiratory allergies including asthma may be more affected by heterologous immune response against SARS-CoV-2. It is of high relevance, that both pipelines highlighted candidate epitopes from *Dermatophagoides* species, as well as *Aspergillus fumigatus* and *Phleum pratense*, suggesting an important role for these allergens. Although the frequency of allergen-specific $CD8^+$ T cells is likely to be low, rare cell subsets have been quite often shown to play an important pathophysiological role[28], and new technologies and bioinformatic approaches for identification of such populations are steadily emerging[29]. Quite importantly, the SARS-CoV-2 $Nsp6_{141-149}$, which was identified among our top potentially cross-reactive epitope pairs, has been recently described by an independent group[30]. To our knowledge, this is the first report on in silico predicted T cell epitope cross-reactivity between SARS-CoV-2 and allergens. While a limitation of our study is the in silico nature of the work, the sequence homology between SARS-CoV-2 and clinically relevant respiratory allergens is along the lines of previously reported cross-reactivity between RNA virus- and allergen-derived peptides at the level of T memory cells[4]. Moreover, our current findings generate further hypotheses in how the adaptive immune system responds differentially with respect to the atopy status of the host. Our present study warrants an immediate investigation of these predicted T cell epitopes to link their possible role in driving the immune response against the SARS-CoV-2 and eventually shape COVID-19 outcome.

There are several different avenues through which the similarities may influence the host immune response. For instance, in hosts sensitized to one of the predicted aeroallergens, the identified similarities with the SARS-CoV-2 proteome may be protective if they prevent an overwhelming Th1 response and the accompanying cytokine storm. Furthermore, allergen-specific T cells may develop a memory response against heterologous SARS-CoV-2 epitopes, which is faster and more efficient. Conversely, such heterologous immune responses could have an adverse outcome by attenuating the antiviral response. T2 immune bias could potentially lead to inadequate virus clearance due to attenuated $CD8^+$ responses. Indeed, there is evidence of a reciprocal relationship between atopy and production of type I and III Interferons in response to viral infections[31]. Given that underlying atopic conditions have not been identified as a significant risk factor for severe clinical courses in those infected with SARS-CoV-2, the epitope homology most likely plays a protective role[7,8]. Interestingly, Jackson et al[32] recently reported that nasal epithelial cells from children with atopic asthma express significantly lower levels of ACE2 receptor as compared to cells from children without asthma or with non-atopic asthma. Similarly, another study using adult bronchial brush samples showed an inverse correlation between ACE2 gene expression and a Th2 dependent gene expression signature[33]. Differential expression of ACE2 receptors among atopic individuals could represent a distinct and unrelated mechanism of action in this context. Our in silico data provide ground to investigate the role of cellular immune responses in regards to the interaction between atopy/asthma and COVID-19. Indeed, the role of SARS-CoV-2-specific T cells in exposed and non-exposed individuals, thereby underlining the importance of heterologous immunity, has been very recently described[34,35]. Further experimental studies are needed to explore the involved pathogenetic mechanisms and potential clinical implications of underlying aeroallergen sensitization on the immune response to SARS-CoV-2.

Data availability

The data used and analyzed in the present study are available from the corresponding author on reasonable request.

Received: 23 September 2020; Accepted: 15 February 2021
Published online: 26 February 2021

References

1. Coronavirus Disease (COVID-19) Situation Reports; 22.09.2020. [Cited 2020 September 22.] Available from https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/.
2. Nextstrain / groups / blab / sars-like-cov; 21.09.2020. [Cited 2020 September 22.] Available from https://nextstrain.org/groups/blab/sars-like-cov.
3. Cascella, M., Rajnik, M., Cuomo, A., Dulebohn, S. C. & Di Napoli, R. StatPearls. Features, Evaluation, and Treatment of Coronavirus (COVID-19). Treasure Island (FL) (2020).
4. Skevaki, C. *et al.* Influenza-derived peptides cross-react with allergens and provide asthma protection. *J. Allergy Clin. Immunol.* **142**, 804–814 (2018).
5. Balz, K., Trassl, L., Härtel, V., Nelson, P. P. & Skevaki, C. Virus-induced T cell-mediated heterologous immunity and vaccine development. *Front. Immunol.* **11**, 513 (2020).
6. Pusch, E., Renz, H. & Skevaki, C. Respiratory virus-induced heterologous immunity: Part of the problem or part of the solution?. *Allergo J. Interdiszipl. Z. Allergol. Umweltmed. Organ Dtsch. Ges. Allerg. Immunitatsforschung* **27**, 28–45 (2018).
7. Liu, S., Zhi, Y. & Ying, S. COVID-19 and asthma: Reflection during the pandemic. *Clin. Rev. Allergy Immunol.* **59**, 78–88 (2020).
8. Carli, G., Cecchi, L., Stebbing, J., Parronchi, P., Farsi, A. Is asthma protective against COVID-19? *Allergy*. https://doi.org/10.1111/all.14426 (2020).

9. Asthma Prevalence; 22.09.2020. [Cited 2020 September 22.] Available from https://www.cdc.gov/asthma/data-visualizations/prevalence.htm.
10. Zhu, Z. *et al.* Association of asthma and its genetic predisposition with the risk of severe COVID-19. *J. Allergy Clin. Immunol.* **146**(327–329), e4 (2020).
11. Odak, I. *et al.* Reappearance of effector T cells is associated with recovery from COVID-19. *EBioMedicine* **57**, 102885 (2020).
12. Shi, W. *et al.* Clinical characteristics of COVID-19 patients combined with allergy. *Allergy* **75**, 2405–2408 (2020).
13. Vita, R. *et al.* The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res.* **47**, D339–D343 (2019).
14. Mari, A., Rasi, C., Palazzo, P. & Scala, E. Allergen databases: current status and perspectives. *Curr. Allergy Asthma Rep.* **9**, 376–383 (2009).
15. Radauer, C., Bublin, M., Wagner, S., Mari, A. & Breiteneder, H. Allergens are distributed into few protein families and possess a restricted number of biochemical functions. *J. Allergy Clin. Immunol.* **121**(847–52), e7 (2008).
16. Cui, J. *et al.* Computer prediction of allergen proteins from sequence-derived protein structural and physicochemical properties. *Mol. Immunol.* **44**, 514–520 (2007).
17. 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).
18. The universal protein resource (UniProt). *Nucleic Acids Res*. 36: D190–D195 (2008).
19. Peters, B. & Sette, A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinform.* **6**, 132 (2005).
20. 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).
21. Moutaftsi, 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).
22. 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).
23. Camacho, C. *et al.* BLAST+: Architecture and applications. *BMC Bioinform.* **10**, 421 (2009).
24. Nielsen, M. *et al.* Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci. Publ. Protein Soc.* **12**, 1007–1017 (2003).
25. Hoof, I. *et al.* NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* **61**, 1–13 (2009).
26. 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).
27. Cock, P. J. A. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics (Oxford, England)* **25**, 1422–1423 (2009).
28. Murray, S. E., Toren, K. G. & Parker, D. C. Peripheral CD4(+) T-cell tolerance is induced in vivo by rare antigen-bearing B cells in follicular, marginal zone, and B-1 subsets. *Eur. J. Immunol.* **43**, 1818–1827 (2013).
29. Arvaniti, E. & Claassen, M. Sensitive detection of rare disease-associated cell subsets via representation learning. *Nat. Commun.* **8**, 14825 (2017).
30. Grifoni, A. *et al.* A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. *Cell Host Microbe* **27**(671–680), e2 (2020).
31. Edwards, M. R. *et al.* Viral infections in allergy and immunology: How allergic inflammation influences viral infections and illness. *J. Allergy Clin. Immunol.* **140**, 909–920 (2017).
32. Jackson, D. J. *et al.* Association of respiratory allergy, asthma, and expression of the SARS-CoV-2 receptor ACE2. *J. Allergy Clin. Immunol.* **146**(203–206), e3 (2020).
33. Bradding, P. *et al.* ACE2, TMPRSS2, and furin gene expression in the airways of people with asthma-implications for COVID-19. *J. Allergy Clin. Immunol.* **146**, 208–211 (2020).
34. Grifoni, A. *et al.* Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell* **181**(1489–1501), e15 (2020).
35. Braun, J. *et al.* SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19. *Nature* **587**, 270–274 (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)-funded SFB 1021 (C04), KFO 309 (P10), and SK 317/1-1 (Project No. 428518790) as well as by the Foundation for Pathobiochemistry and Molecular Diagnostics. KCN is supported by NIH Grants U01AI140498-03 and R01AI140134-01, The Sunshine Foundation, and the Sean N. Paker Center for Allergy and Asthma Research at Stanford University.

Author contributions

Conception or design of the work: C.S., K.C.N.; Data collection: K.B., A.K., F.C., V.H.; Data analysis and interpretation: C.S., K.C.N., K.B., A.K., P.T., N.T., K.E., P.N.; Drafting the article: C.S., K.C.N., K.B., M.C., A.K.; Critical revision of the article: C.S., K.C.N., K.B., M.C., A.K., H.R.; Final approval of the version to be published: All.

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 KCN: Dr. Nadeau reports Grants from National Institute of Allergy and Infectious Diseases (NIAID), Food Allergy Research & Education (FARE), End Allergies Together (EAT), Allergenis, and Ukko Pharma; Grant awardee at NIAID, National Institute of Environmental Health Sciences (NIEHS), National Heart, Lung, and Blood Institute (NHLBI), and the Environmental Protection Agency (EPA); is involved in Clinical trials with Regeneron, Genentech, AImmune Therapeutics, DBV Technologies, AnaptysBio, Adare Pharmaceuticals, and Stallergenes-Greer; Research Sponsorship by Novartis, Sanofi, Astellas, Nestle; Data and Safety Monitoring Board member at Novartis and NHLBI; Cofounded Before Brands, Alladapt, ForTra, and Iggenix; Chief Intellectual Office at FARE, Director of the World Allergy Organization (WAO) Center of Excellence at Stanford, Personal fees from Regeneron, Astrazeneca, ImmuneWorks, and Cour Pharmaceuticals;

Consultant and Advisory Board Member at European Academy of Allergy and Clinical Immunology (EAACI) Research and Outreach Committee, Ukko, Before Brands, Alladapt, IgGenix, Probio, Vedanta, Centecor, Seed, Novartis, NHBLI, EPA, National Scientific Committee of Immune Tolerance Network (ITN) and NIH Programs; US patents for basophil testing, multifood immunotherapy and prevention, monoclonal antibody from plasmoblasts, and device for diagnostics. All other authors have no competing interests to declare.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-84320-8.

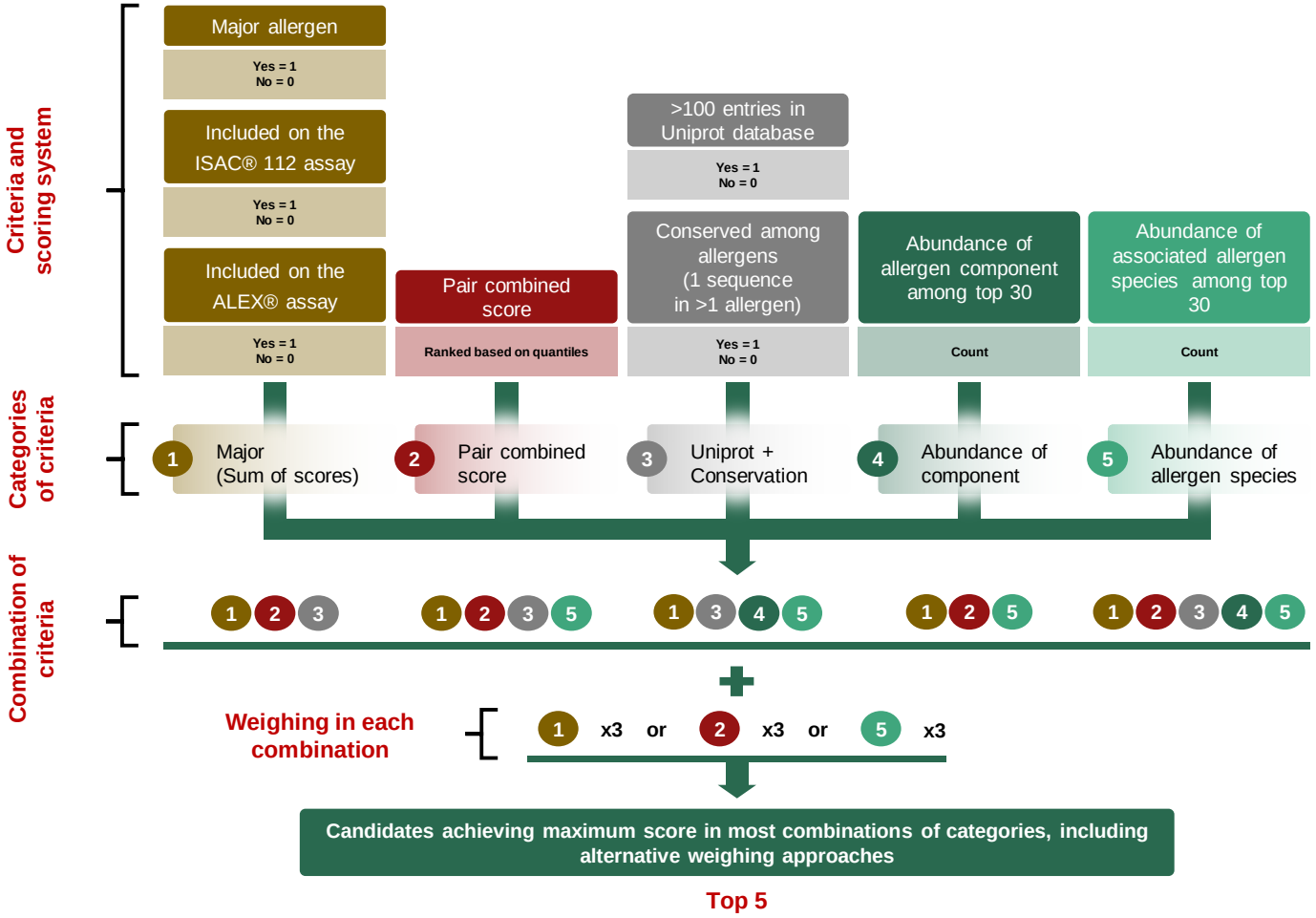
Correspondence and requests for materials should be addressed to C.S.

Reprints and permissions information is available at 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) 2021



Supplementary Figure S1. Criteria and workflow for further scoring of the top 30 potentially cross-reactive predicted allergen-derived T cell epitopes for pipeline 1. Individual molecular allergen components or their associated peptides among the top 30 potentially cross-reactive epitope pairs were scored based on (1) whether they evoked IgE production in >50 % of patients with associated clinical allergy as per WHO/IUIS Allergen Nomenclature Sub-committee (http://allergen.org/index.php) and whether they are included in the commercially available ImmunoCAP ISAC®112 or ALEX® diagnostics assays; (2) pair combined score, defined in 4 ranges, with the help of percentiles : ≤ 10 % = 0; [20-50 %] = 1; [50-90 %] = 2; [90-100 %] = 3; (3) whether they achieve more than 100 entries when entering the epitope sequence in Uniprot and whether the specific epitope sequence is also contained in proteins of other allergen sources and (4) how abundant the specific allergen component or (5) associated allergen source is within the top 30 candidates. Five different combinations of the aforementioned categories were defined and scored, and the cumulative score was calculated for each allergen epitope. Additionally, the cumulative score was calculated three more times, each time multiplying another category by a factor of 3 in order to critically compare alternative weighing of the associated criteria. Allergens were ultimately prioritized for further testing based on the frequency of achieving the maximum score in each of the separately weighed scoring approaches as described above

Supplementary Methods

Calculation of the pair combined score:

The pair combined score takes 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 (cut-off 0,8): Pair combined score= 1/binding affinity (nM) (Virus) * 1/binding affinity (nM) (Allergen) * score PwA. Therefore, a higher score is associated with an increasing probability for MHC binding and a higher degree of similarity between the virus and allergen epitope.

Scoring system:

A scoring system was developed and five categories of criteria were formed, summing up the scores of individual criteria in one group (**Figure E1**). As a next step, five different combinations of the aforementioned categories were defined and the cumulative score was calculated for each allergen epitope. Additionally, the cumulative score was calculated three more times, each time multiplying another category by a factor of 3 in order to critically compare alternative weighing of the associated criteria. The new Top 5 allergens and associated virus epitopes were subsequently ranked based on the frequency of achieving the maximum score in each of the separately weighed scoring approaches as described above.

Supplementary Table S3: The most prevalent human HLA alleles, which were targeted for the in silico epitope prediction in pipeline 1

Human MHC Class I[1]	Human MHC Class II[2]
HLA-A*01:01	DRB1*01:01
HLA-A*02:01	DRB1*03:01
HLA-A*11:01	DRB1*04:01
HLA-A*24:02	
HLA-B*07:02	
HLA-B*40:02	

1=epitopes with a length of 9 or 10 amino acids

2=epitopes with a length of 15 amino acids