nature portfolio

Peer Review File

Mapping the immunopeptidome of seven SARS-CoV-2 antigens across common HLA haplotypes

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
Major comments:
• In general, the description of the methods and protocols used are lacking the necessary details in the main text and in the material and methods section. Simply referring to other papers is not enough. The authors therefore need to provide, at least, a detailed Supplementary Methods section.
• The authors should argue why two different peptide purification strategies were used and give pros and cons of both methods.
• The authors have chosen to set a 5% FDR threshold to validate the identified peptides which needs to be justified. In the mat&met section, they mention “For SARS-CoV-2 peptides additional high confidence re-testing of detection threshold at 1% FDR was performed.” but no results are provided nor discussed here on the 1%-5% output comparison. Are many of the identifications close to the 5% threshold?
• The authors say they have identified 128 class I and 158 unique class II SARS-CoV-2 peptides, while the total number of unique peptides is only 248. Overlaps should be clearly indicated.
• The authors state they selected 56 peptides out of the total 248 based on the predicted HLA binding. Which cutoff was used? Do the authors have an explanation as to why only half of the peptides are predicted to be binders? Could it be that because of the cutoff potentially interesting peptides have been left out of further analysis.
• The authors have performed an experiment with synthetic peptides to validate the found hits, which is not convincing. The visualisations used, in particular the retention times and ion mobility graphs, are not fit for purpose. They should be replaced by distribution plots of the difference between the synthetic versus endogenous peptides. The authors should clearly state whether or not some peptides have been excluded based on this validation experiment. Have different gradients been used to explain such big retention time shifts? The pearson correlation coefficients distribution looks bad for synthetic peptides, these days, even PCC with predictions look better.
• Recently rescoring peptide spectrum matches have been introduced and shown to have a substantial impact on the identification rate of immunopeptides, have the reviewers considered using one of these implementations such as Prosit rescoring, MS2Rescore or inSPIRE? Potentially more xeno-epitopes could be found and the authors could lower the FDR threshold to the typical 1% FDR.
• The authors always performed an experiment without a viral protein in the cells, have they searched those with the same database containing the viral protein(s)? Were there any viral protein hits there that were also found in the experiment with the viral protein?
• Have the authors performed multiple immunopeptidomics experiments for the same samples? Were the xeno-epitopes identified in multiple runs or only in single runs?
• On lines 191-192 the authors state that the two methods used can be complementary in a way, however for the N protein it seems the direct delivery method does not provide any extra information since the hotspots were already identified with the stable transfected method. The same seems to be true for the nsp9 protein. Therefore, it seems a bold statement to say both methods allow for broader coverage, especially since this method does not seem to work well for

the N protein. The authors should go more into detail here as to why they think it is more valuable to use both methods and add the pro's and con's of both methods.
• Regarding the post-translational modifications, were the peptides with a modification also identified without the modifications, or are these modifications specific for the peptides? Also on line 266 the authors claim that the modifications are in line with what is to be expected. This argument must be referenced.
• The github link for the cross reactivity check does not work.
• The authors have done a substantial amount of work to validate antigenicity of 56 SARS-CoV-2 peptides that were identified in the immunopeptidomics experiment. However, more in-depth information and discussion is required on this part of the manuscript.
o Figure 5 shows a lot of information however these all give a very precise information where certain pools activate for certain donors, however, a more general donor to donor overview is missing. Are there donors who show response for multiple peptide pools, or for none?
o Some of the DMSO stimulated cells also show a lot of IFNγ+ TNF+ expression, is this usual? The comparison to the blank controls should also be more analysed in depth as here there seems to be a lot of deviation in DMSO experiments in between the pool or peptide experiments, which you would hope is not the case for a blank control.
o Have the authors considered other metrics such as MFI (median/mean fluorescence intensity) to compare activation.
o Were these experiments performed multiple times for the same donor, and do they show similar expression rates for both DMSO as well as the peptide pools?
o Figure 5A for instance shows percentages that are very close to each other and show similar activation for pool 1 DMSO and pool 1. The authors should spend a bit more time on this in the discussion or results section.

Minor comments:
• Figure 1 should be revised by clearly mentioning the immunoaffinity purification step and the large scale/small scale elution notion introduced in the mat&met section could already be added in figure 1.
• Overall the figure legends are vague and lack minimal information to be self-understandable.

Reviewer #2:
Remarks to the Author:
In this study, Braun and colleagues analyzed the immunopeptidome of SARS-CoV-2, focusing on non-spike structural and non-structural antigens. The novelty of this study resides in the mass-spectrometry-based approach and HLA-processing that has only been scarcely reported for T cell epitope mapping. Although this methodology offers many significant advantages over the traditional epitope discovery techniques using predicted HLA-binders or overlapping peptides with similar peptide lengths, results are somewhat unexpected and challenging to conciliate with the vast amount of literature in the SARS-CoV-2 T cell field. Nevertheless, I believe the manuscript could be improved by addressing the suggestions below:

General comments

• My primary concern is the fact that the findings are not universal and somewhat dissimilar to the well-characterized repertoire of SARS-CoV-2 antigen targets and immunodominant responses. For example, the authors identified a large number of naturally processed peptides from the nsp8 and nsp9 antigens, which are typically associated with low T cell memory reactivity in COVID-19 convalescent individuals. Indeed, the testing of newly identified nsp9-specific peptides later in the manuscript confirmed the overall low nsp9 reactivity.

• The testing of the identified peptides and the result description in Figure 5 is confusing and could benefit from a more stringent statistical analysis. Also, as the outcome of these experiments is dissociated from the rest of the manuscript findings, the authors should have made an attempt to connect the two parts.

Specific comments

Introduction

Well written. I would suggest that more updated literature be used, particularly when referring to the durability of T cell responses and how it can be affected by further vaccine boosters and breakthrough infections or when referring to mutations and consequences of novel variants in shaping the repertoire of T cell responses.

Results

When providing a rationale for focusing on non-spike antigens, nothing is mentioned regarding why Nsp1, Nsp4, and Nsp5 were chosen. Also, the fact that the other proteins were chosen based on the short kinetics of infection in vitro, might not translate to the kinetics or individual protein abundance of in vivo host replication. For example, the RNA-dependent RNA polymerase (NSP12) is known to elicit high T cell reactivity and to play a significant role in the viral infection process. Please comment.

In Figures 2B and 3B, it Is unclear why the “fraction” of peptides to the most abundant peptide length is being plotted instead of the overall abundance. Please clarify

When comparing both approaches, the authors classify the N protein as one of the less presentable canonical proteins of SARS-CoV-2. However, Nucleocapsid is one of the most common CD4+ and CD8+ T cell targets that dominate the response magnitude, breadth, and frequency after SARS-CoV-2 infection. Indeed, the most immunodominant SARS-CoV-2 CD8+ T cell epitope reported to date is a HLA-B*07:02 restricted N epitope. Can the authors elaborate on the discrepancies?

The authors attempt to make a straight comparison between direct delivery and stable transfection

approaches, but only 2 antigens are common to both approaches, and only a partial overlap is observed for nsp9. Why were different antigens considered for the different approaches?

A description or a table with clinical data and other relevant information about the cohort of 9 individuals should be presented. Likewise, more details (e.g. how many times were cells fed with IL-2?) should be added to the protocol for expanding PBMCs with peptide pools. Why were cells chosen to be expanded for 10 days or 12-13 days? I am concerned that de novo responses could also be developed besides a memory recall response. The authors should also show FACS plots of the phenotype of the expanding cells, mainly CD4+ and CD8+ T cell populations and memory sub-populations.

In Figure 5A at least for CD4+ T cell responses, the levels of DMSO are notably high, suggesting bystander activation. If a stimulation index>2 (i.e. specific signal over DMSO control), which translates to a higher confidence in identifying bone-fide-specific responses, were to be considered, none would be considered positive responders. In this regard, in the graphic with the summary of net responses (background DMSO subtracted), the value of the representative donor in the dot plot on the left is not represented. For example, Pool 2 net value of 0.21 (0.049-0.028) is not depicted.

In Figure 5B, why is the sum of the individual peptides (nested and non-nested) much greater than the pool's response with the peptides combined? Are they tested individually or pooled at the same concentration? Also, the dot plot from Nsp9_46-54 stimulation depicts a very different population shape, typically associated with FMO controls. Is this the correct visualization of that particular peptide stimulation?

For Figure 5C I have the same criticism as for Figure 5A. Indeed, the only case where peptide stimulation has a seemingly higher signal over DMSO control is shown for the representative donor of Figure 5B. However, the authors did not show a response summary across several donors. The fact that none of the figures in Figure 5 have any statistical analysis makes it impossible to draw any conclusion. I would suggest making this section more scientifically accurate by taking my suggestions above. Also, to call these responses “robust” in the abstract seems markedly exaggerated.

Discussion

Too short. A much bigger emphasis should be placed on the overall conclusions and the reconciliation of the findings within the manuscript and the vast amount of literature. Additional considerations regarding the capability of the identified epitopes to cross-recognize common cold coronaviruses or different SARS-CoV-2 variants at the global population level should be discussed.

Some additional limitations of the study, pointed out above, should also be added.

Reviewer #3:
Remarks to the Author:
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

We thank both reviewers for their positive assessment of our study and their constructive criticism and suggestions. We have now addressed all of the reviewers' comments and queries, as outlined below. We believe that the modifications to our manuscript, in particular the addition of new figures, have significantly improved our study and our overall conclusions.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Major comments:

• In general, the description of the methods and protocols used are lacking the necessary details in the main text and in the material and methods section. Simply referring to other papers is not enough. The authors therefore need to provide, at least, a detailed Supplementary Methods section.

Response:

For the main immunopeptidomics approach, both of the papers cited are protocol papers from our lab describing either the large scale protocol (citation 16; Nat. Prot. 2019) or the small scale protocol (citation 31; Curr. Prot. 2021). Both of these protocol papers have been specifically written to provide a highly detailed account of how we process immunopeptidomics samples. We feel that a detailed supplementary methods section would merely duplicate our published work. Instead, we have added some more details to the materials and methods immunopeptidomics section.

In general, the methods section follows the same format of similar papers from our lab and others when performing mass spectrometry and subsequent immunogenicity screening using T cell activation assays. If more specific details of the kind of missing information are provided, we would be happy to extend any particular sections that require further attention.

• The authors should argue why two different peptide purification strategies were used and give pros and cons of both methods.

Response:

An additional section has been added to address this important point, lines 216-226.

"Overall, the exogenous introduction of protein in the direct delivery approach directs antigen presentation predominantly to the HLA class II pathway, while the transfection approach allows to detect HLA class I and class II peptides in a more balanced ratio, at least when B cells are studied which express high levels of HLA class II and may be less efficient than other antigen presenting cells to cross-present antigen. Where time is critical and HLA class II presentation is under investigation, the direct delivery approach provides a clear time advantage but if protein expressed in bacteria is used, this approach might not capture mammalian post-translational modifications. The transfection approach, though slower and more labour-intensive, provides a more complete gamut of HLA class I (and class II) antigens derived from the transfected gene of interest and is critical for approaches where the depth of antigen discovery is key."

• The authors have chosen to set a 5% FDR threshold to validate the identified peptides which needs to be justified. In the mat&met section, they mention "For SARS-CoV-2 peptides additional high

confidence re-testing of detection threshold at 1% FDR was performed." but no results are provided nor discussed here on the 1%-5% output comparison. Are many of the identifications close to the 5% threshold?

Response:

Whilst the consensus in bottom-up proteomics analysis of high abundance peptides is that a more stringent 1% FDR cut-off is appropriate, for immunopeptidomics analysis of low abundance MHC peptides, our experience is that a less stringent 5% FDR threshold is more appropriate and many hits at 1-5% FDR can be successfully validated. Indeed, the variable performance of FDR cut-offs in immunopeptidomics has recently been discussed in detail using scan fate tracing in datasets using synthetic peptide libraries using different database sizes by Parker et al (MCP 2021)*. The vast majority of SARS-CoV-2 peptides presented in this work and their modifications can be detected at both, the 1% and 5% FDR thresholds (229/302, 76%; Table S1 column H), this additional information has been added in line 505. The 1% retesting values in supplementary files allow the reader to filter any detected peptides in a more stringent manner if desired.

*Parker R, et al. The Choice of Search Engine Affects Sequencing Depth and HLA Class I Allele-Specific Peptide Repertoires. Mol Cell Proteomics. 2021;20:100124.

• The authors say they have identified 128 class I and 158 unique class II SARS-CoV-2 peptides, while the total number of unique peptides is only 248. Overlaps should be clearly indicated.

Response:

All peptides and their source antigens can be found in Supplementary Table S1. This table includes each peptide version, including peptides with/without PTM or identical peptides found across different experiments and in class I and class II immunoprecipitation steps.

• The authors state they selected 56 peptides out of the total 248 based on the predicted HLA binding. Which cutoff was used? Do the authors have an explanation as to why only half of the peptides are predicted to be binders? Could it be that because of the cutoff potentially interesting peptides have been left out of further analysis.

Response:

The cut-off used is mentioned in material and methods, lines 515-516: "NetMHCpan and NetMHCIIpan binding prediction was used with %Rank cut-off 2 and 5 respectively to include strong and weak binders in the analysis."

Whilst binding predictors like NetMHC provide additional validation, the prediction algorithms have been trained on experimental data which can be skewed and partially limited, especially for the HLA-C alleles that are not as highly expressed as HLA-A/HLA-B. One further limitation for binding prediction is that it performs well for the canonical 9-mer length, but MHC I molecules can present slightly longer and shorter versions of the same peptide, e.g. the detected SARS-CoV-2 peptide "ALLSDLQDL" is a 9-mer and a predicted strong binder of HLA-A*02:01, but the detected 10-mer "ALLSDLQDLK" is not predicted to bind by the algorithm (S1 table). Additionally, binding predictions for the longer MHC II peptides are inherently difficult and less evolved because different amino acid positions can be accommodated in MHC binding pockets. Furthermore, most historical data that was

used to train the NetMHC algorithm came from older instruments with less sensitivity and Orbitrap data, whereas here we present data from the highly sensitive TimsTOF Pro2 instrument.

• The authors have performed an experiment with synthetic peptides to validate the found hits, which is not convincing. The visualisations used, in particular the retention times and ion mobility graphs, are not fit for purpose. They should be replaced by distribution plots of the difference between the synthetic versus endogenous peptides. The authors should clearly state whether or not some peptides have been excluded based on this validation experiment. Have different gradients been used to explain such big retention time shifts? The Pearson correlation coefficients distribution looks bad for synthetic peptides, these days, even PCC with predictions look better.

Response:

To address this query, we now include mirror plots of the experimental and synthetic peptide spectra with associated PCC values as Supplemental Fig. 3C. For three peptides where a natural to synthetic spectrum match was not possible (e.g. different charges/PTM), we conducted a follow-on analysis of the natural to predicted spectrum (Suppl. Fig. 3D). Upon reviewing and introducing the raw data mirror plots to Suppl. Fig 3 and a distinction drawn between synthetic or predicted matches, we have now adjusted the PCC, IM and RT plots (formerly Suppl. Fig 3 C-E, now panel B) as requested in this review to rather show distribution plots for IM and RT. The mentioned RT shift is predominantly unidirectional across the peptides indicating a systematic change, together with the fact that experiments spanned two years and vendors/manufacturers around this time were introducing slightly upgraded products (e.g. Aurora columns used in this study), so this not entirely unexpected.

Synthetic peptides were ordered for combined MS-based validation of the targets and immunogenicity testing. This validation was conducted in parallel and as such all peptides progressed to T cell assays, which could be considered the ultimate validation.

• Recently rescoring peptide spectrum matches have been introduced and shown to have a substantial impact on the identification rate of immunopeptides, have the reviewers considered using one of these implementations such as Prosit rescoring, MS2Rescore or inSPIRE? Potentially more xeno-epitopes could be found and the authors could lower the FDR threshold to the typical 1% FDR.

Response:

The rescoring of peptides using spectral intensity prediction and sequence features is one method to reduce FDR rates but is arguably an artificial way to prioritise xeno-peptides. We decided to use less stringent FDR cut-offs without rescoring the data and screen for peptides prioritised based on their source antigen and, importantly, by their absence of detection in untreated cell pellets. Moreover, the data has been deposited and open to anyone to be reanalysed with their currently preferred rescoring algorithm should this be the reader's preference. The goal of this work was to find high confidence candidates for immunogenicity testing, the T cell assays have been completed and due to the scarcity of HLA-typed PBMCs from convalescent donors, it was not be possible for us to extend them to all peptides.

• The authors always performed an experiment without a viral protein in the cells, have they

searched those with the same database containing the viral protein(s)? Were there any viral protein hits there that were also found in the experiment with the viral protein?

Response:

Yes, as mentioned above, we performed a total SARS-CoV-2 proteome search of the parental cells and control cells without protein during direct delivery as a strategy to identify *bona fide* viral antigen derived peptides. None of the reported viral peptides were present in the negative controls. A sentence to clarify this has been added (lines 196-198).

• Have the authors performed multiple immunopeptidomics experiments for the same samples? Were the xeno-epitopes identified in multiple runs or only in single runs?

Response:

Immunopeptidomics for the purpose of antigen discovery was performed on as much material as possible. For this reason, we combined all collected cell pellets of a sample into one immunopeptidomics run. Each sample was run as a single injection (direct delivery) or single injections of individual fractions (transfectants). Additional information in regards to this has been added in line 487.

• On lines 191-192 the authors state that the two methods used can be complementary in a way, however for the N protein it seems the direct delivery method does not provide any extra information since the hotspots were already identified with the stable transfected method. The same seems to be true for the nsp9 protein. Therefore, it seems a bold statement to say both methods allow for broader coverage, especially since this method does not seem to work well for the N protein. The authors should go more into detail here as to why they think it is more valuable to use both methods and add the pro's and con's of both methods.

Response:

We have revised our statements and re-written the section comparing transfection to direct delivery (lines 205-208 and 216-226). We have provided an additional paragraph contrasting both methods with what we believe are their pros and cons and note with interest that there is not always strong correlation between the two antigen delivery methods.

• Regarding the post-translational modifications, were the peptides with a modification also identified without the modifications, or are these modifications specific for the peptides? Also on line 266 the authors claim that the modifications are in line with what is to be expected. This argument must be referenced.

Response:

Both cases are present, i.e. for some peptides we see both native and PTM bearing peptides whilst for others only the PTM peptide was observed. The individual information for each peptide is accessible to the reader via Table S1. Additional references have been included highlighting the common nature of the observed PTMs in a range of immunopeptidomics studies (line 254).

• The github link for the cross reactivity check does not work.

Response:

This has now been corrected, the link is active and confirmed at

https://github.com/PurcellLab/agrep_for_crossreactivity

• The authors have done a substantial amount of work to validate antigenicity of 56 SARS-CoV-2 peptides that were identified in the immunopeptidomics experiment. However, more in-depth information and discussion is required on this part of the manuscript.
Figure 5 shows a lot of information however these all give a very precise information where certain pools activate for certain donors, however, a more general donor to donor overview is missing. Are there donors who show response for multiple peptide pools, or for none?

Response:

Thank you for this suggestion. We have now expanded the discussion (lines 361-373).

In Figure 6 (formerly **Figure 5**), we have now included heat maps of results from each pool for each donor to make this information more readily accessible. While some donors show more reactivity than others, no donor responded to all three pools. Additional donor information including prior disease severity and HLA typing can be found in Supplementary Table S5.

o Some of the DMSO stimulated cells also show a lot of IFNγ+ TNF+ expression, is this usual? The comparison to the blank controls should also be more analysed in depth as here there seems to be a lot of deviation in DMSO experiments in between the pool or peptide experiments, which you would hope is not the case for a blank control.

Response:

We have not included blank samples as they would not be a correct control in this experiment. DMSO used at the same dilution as peptides is a much more appropriate control. After 10-13 days in culture, it is not unexpected to have some level of bystander activation in T cell assays which can vary between donors, hence DMSO controls were strictly included every time.

o Have the authors considered other metrics such as MFI (median/mean fluorescence intensity) to compare activation.

Response:

There are two reasons to analyse the response as a % of IFN+ cells rather than via MFI analysis. Biologically, T cells responding with a lower or higher level of IFNγ production still indicate a valid T cell response. Analytically, a low percentage of positive cells within a major population of non-responders will not result in an MFI shift in flow cytometry, hence looking for % changes is appropriate in this assay. Overall, the expected range of response in these assays was in line to our previous work, particularly for epitopes following a primary SARS-CoV-2 infection.

- Nguyen THO, et al.. CD8+ T cells specific for an immunodominant SARS-CoV-2 nucleocapsid epitope display high naive precursor frequency and TCR promiscuity. Immunity. 2021; 54(5):1066-1082.e5.

- Habel JR, et al.. Suboptimal SARS-CoV-2-specific CD8+ T cell response associated with the prominent HLA-A*02:01 phenotype. Proc Natl Acad Sci U S A. 2020; 117(39):24384-24391.

o Were these experiments performed multiple times for the same donor, and do they show similar expression rates for both DMSO as well as the peptide pools?

Response:

Due to the scarcity of clinical samples, the experiments were performed only once per donor. The DMSO control vs peptide response is shown in Fig. 5 for all such cases.

o Figure 5A for instance shows percentages that are very close to each other and show similar activation for pool 1 DMSO and pool 1. The authors should spend a bit more time on this in the discussion or results section.

Response:

Figure 5A (now 6A): Similar percentages across DMSO and pool 1/3 responses in Donor 1 indicate that there was no response against these pools. A higher response to pool 2 compared to DMSO in Donor 1 (nearly 2-fold) indicates that there was a response against pool 2. It is expected that any donor will only respond to a limited number of peptides given that each donor has a different HLA haplotype.

Minor comments:
• Figure 1 should be revised by clearly mentioning the immunoaffinity purification step and the large scale/small scale elution notion introduced in the mat&met section could already be added in figure 1.

Response:

The figure has been updated to clearly include an immunoaffinity purification step. The figure now also introduces the “Transfection” vs “Direct delivery” expressions used throughout the manuscript. The expressions “large scale” and “small scale” are now included in the figure legend.

• Overall the figure legends are vague and lack minimal information to be self-understandable.

Response:

As indicated in the previous response, we have altered the figure legend to make it more informative. If further changes are required to more figure legends, please indicate the specific information that is lacking and we would be happy to add it.

Reviewer #2 (Remarks to the Author):

In this study, Braun and colleagues analyzed the immunopeptidome of SARS-CoV-2, focusing on non-spike structural and non-structural antigens. The novelty of this study resides in the mass-spectrometry-based approach and HLA-processing that has only been scarcely reported for T cell epitope mapping. Although this methodology offers many significant advantages over the traditional

epitope discovery techniques using predicted HLA-binders or overlapping peptides with similar peptide lengths, results are somewhat unexpected and challenging to conciliate with the vast amount of literature in the SARS-CoV-2 T cell field. Nevertheless, I believe the manuscript could be improved by addressing the suggestions below:

General comments

• My primary concern is the fact that the findings are not universal and somewhat dissimilar to the well-characterized repertoire of SARS-CoV-2 antigen targets and immunodominant responses. For example, the authors identified a large number of naturally processed peptides from the nsp8 and nsp9 antigens, which are typically associated with low T cell memory reactivity in COVID-19 convalescent individuals. Indeed, the testing of newly identified nsp9-specific peptides later in the manuscript confirmed the overall low nsp9 reactivity.

Response:

One of the goals of our study was to focus on antigens less well studied than say the spike glycoprotein. However, to alleviate these concerns, we have performed additional bioinformatic analysis of the ligand and epitope data available in the IEDB database, now in Figure 5, and compared this directly with our own findings. Here, it becomes evident that a significant number of NSP8 and NSP9 epitopes have been confirmed in patient samples. We do not regard the detection of responses against two nsp9 peptides as overall low reactivity. Overall, 14% of all tested peptides were confirmed as T cell epitopes, somewhat lower than the report in IEDB for identified ligands (any assay) with a corresponding positive ICS epitope response (~48%), this is however with perfect HLA match, which was not possible in the current study. It is furthermore evident from the newly performed analysis of IEDB ligands and epitopes, that the manuscript adds significantly to the somewhat sparse ligand information for HLA-A*01:01, HLA-B*08:01/B*27:05 and HLA-C*01:02/C*07:01 and to some degree for HLA-DQA*05:01/DQB1*02:01 and HLA-DRA1*01/HLA-DRB1*01:01 /DRB1*03:01

It was furthermore not the goal of this work to find antigens with the highest memory T cell response, we wanted to validate new antigens found through our discovery workflow. This approach is known to be successful from previous work on influenza and other viral pathogens.

• The testing of the identified peptides and the result description in Figure 5 is confusing and could benefit from a more stringent statistical analysis. Also, as the outcome of these experiments is dissociated from the rest of the manuscript findings, the authors should have made an attempt to connect the two parts.

Response:

This was a discovery project and the T cell assays were performed to find whether any COVID convalescent patients would respond to any of the viral peptides discovered via immunopeptidomics. An additional transition sentence has been added to link antigen discovery to immunogenicity testing (lines287-289).

Given that each donor has a different HLA haplotype and, for the most part, we have found only one donor responding to a peptide, it is not possible to do further statistical analysis. Furthermore, for the analysis DMSO was subtracted to account for bystander activation, therefore there is no additional negative control to run statistics against. Generally, there is no lower threshold limit for

responses in T cell activation assays. Epitope-specific T cells are generally rare and the results are in line with expectations.

Additionally, our access to patient material is limited and does not allow for multiple repeats. The COVID disease burden in Australia was rather low, meaning that a limited number of PBMC samples was used across a number of projects. Other global groups are better positioned to perform broader validation screens in the future.

Specific comments

Introduction

- Well written. I would suggest that more updated literature be used, particularly when referring to the durability of T cell responses and how it can be affected by further vaccine boosters and breakthrough infections or when referring to mutations and consequences of novel variants in shaping the repertoire of T cell responses.

Response:
We have updated details in the introduction and introduce a thorough analysis of annotated T cell and ligand assays in the IEDB (Figure 5) putting our results into a more contemporary context. Vaccination is currently only focussed on the Spike glycoprotein so we cannot comment on how these more novel epitopes would respond to boosters etc.

Results

- When providing a rationale for focusing on non-spike antigens, nothing is mentioned regarding why Nsp1, Nsp4, and Nsp5 were chosen. Also, the fact that the other proteins were chosen based on the short kinetics of infection in vitro, might not translate to the kinetics or individual protein abundance of in vivo host replication. For example, the RNA-dependent RNA polymerase (NSP12) is known to elicit high T cell reactivity and to play a significant role in the viral infection process. Please comment.

Response:

When the work was started, nothing was known about nsp12 or most SARS-CoV-2 proteins for that matter. We took a pragmatic approach to cover as many viral proteins as possible subject to access to reagents and their lack of then current coverage in publicly accessible databases

- In Figures 2B and 3B, it Is unclear why the “fraction” of peptides to the most abundant peptide length is being plotted instead of the overall abundance. Please clarify

Response:

These figures display the length distribution across a high number of experiments from 2A/3A. The total peptide numbers detected in individual experiments vary and hence, to normalise the length distribution across all experiments, the number of 9/15-mers was set to 1. The caveat in setting the overall abundance to 1 as suggested is that different immunopeptidomics groups choose to include different lengths, e.g. 8-12-mers or 8-14-mers in their analysis. This makes it hard to compare relative abundance of shorter/longer peptides across papers. If 9/15-mers are set to 1, the datasets become more readily comparable.

- When comparing both approaches, the authors classify the N protein as one of the less presentable canonical proteins of SARS-CoV-2. However, Nucleocapsid is one of the most common CD4+ and CD8+ T cell targets that dominate the response magnitude, breadth, and frequency after SARS-CoV-2 infection. Indeed, the most immunodominant SARS-CoV-2 CD8+ T cell epitope reported to date is a HLA-B*07:02 restricted N epitope. Can the authors elaborate on the discrepancies?

Response:

This statement has been amended to reflect the reviewer's comment, and the data presented below in tabular comparing data from the IEDB and this study, on the number of ligands/epitopes identified for the N protein. It would have been interesting whether we see the mentioned immunodominant N epitope, however B*07:02 was not expressed by the BLCLs used in this study.

HLA	**Ligand/Epitope count (IEDB)**	**Ligands (This Study)**
HLA-A*01:01	1	4
HLA-A*02:01	10	3
HLA-A*02:03	1	-
HLA-A*03:01	4	-
HLA-A*11:01	3	-
HLA-A*33:03	2	-
HLA-B*07:02	4	-
HLA-B*08:01	1	3
HLA-B*15:01	2	-
HLA-B*27:05	1	1
HLA-B*35:01	3	-
HLA-B*40:01	1	-
HLA-DRB1*01:01	2	-
HLA-DRB1*07:01	2	-
HLA-DQA1*05:01/DQB1*02:01	-	14

HLA-DRB1*03:01	-	10
HLA-DRB3*01:01	-	10

- The authors attempt to make a straight comparison between direct delivery and stable transfection approaches, but only 2 antigens are common to both approaches, and only a partial overlap is observed for nsp9. Why were different antigens considered for the different approaches?

Response:

The direct antigen loading approach was dictated by access to recombinant proteins of sufficient solubility for the assay conditions whilst the transfection was more broadly applicable hence the limited overlap.

- A description or a table with clinical data and other relevant information about the cohort of 9 individuals should be presented.

Response:

Information on the age, gender, HLA type and Days post COVID can be found in Supplementary Table 5, we have now also added a description of disease severity to the dataset.

- Likewise, more details (e.g. how many times were cells fed with IL-2?) should be added to the protocol for expanding PBMCs with peptide pools. Why were cells chosen to be expanded for 10 days or 12-13 days?

Response:

IL-2 was added from Day 4 and as required depending on expansion, at most every other day. Some further information has been added in lines 552-555, 560-561. All cells were tested on Day 10 against the same pool as Day 0. Where a response was detected, T cell cultures were further tested on day 12/13 with individual peptides as is now detailed in lines 555-556, 560-561.

- I am concerned that de novo responses could also be developed besides a memory recall response. The authors should also show FACS plots of the phenotype of the expanding cells, mainly CD4+ and CD8+ T cell populations and memory sub-populations.

Response:

In our experience with this assay, the timeframe is too short and the concentration of peptide too limiting to elicit substantial *de novo* responses. Moreover, the expanded T cell population is usually low in absolute numbers, making it impossible to perform further analysis of T cell subsets with any degree of confidence. Irrespective of whether these responses are memory or *de novo* the immunogenicity of the peptides is not in doubt and we do not attempt to argue that the responses detected contributed to disease outcome in the donors.

- In Figure 5A at least for CD4+ T cell responses, the levels of DMSO are notably high, suggesting bystander activation. If a stimulation index>2 (i.e. specific signal over DMSO control), which translates to a higher confidence in identifying bone-fide-specific responses, were to be considered, none would be considered positive responders. In this regard, in the graphic with the summary of net responses (background DMSO subtracted), the value of the representative donor in the dot plot on the left is not represented. For example, Pool 2 net value of 0.21 (0.049-0.028) is not depicted.

Response:

We thank the reviewer for highlighting the mismatch between the representative FACS plot and graph. We have now corrected the numbers on the representative plot to correspond to the graphed data. We have extensive experience (GS link) with T cell assays and consider several factors carefully, such as cytokine production as a readily measurable read-out, but also factors like proliferation observed during the culture. Here we show the antigen testing data in a fully transparent manner for all individual responses. More work will be needed to determine whether any of the detected antigens are able to elicit immunodominant responses.

- In Figure 6B, why is the sum of the individual peptides (nested and non-nested) much greater than the pool's response with the peptides combined? Are they tested individually or pooled at the same concentration?

Response:

We have combined nested peptides in the same pool to be able to pool peptides that likely bind to the same HLA. Hence, responses against individual peptides (10 µM) will likely have a degree of cross-reactive T cell responses resulting in a higher additive value than the testing of a pool (10 µM total concentration).

The total amount of peptides is always the same, i.e. less of each peptide in a pool versus in individual testing. This also might additionally explain higher responses against individually tested peptides. A clarification has been added in lines 552-553.

- Also, the dot plot from Nsp9_46-54 stimulation depicts a very different population shape, typically associated with FMO controls. Is this the correct visualization of that particular peptide stimulation?

Response:

No FMO controls were used in this experiment. The plot was used to show that although this individual responded to several peptides, it was not due to nonspecific activation as the same cells did not respond to other peptides, e.g. Nsp9_46-54.

- For Figure 5C I have the same criticism as for Figure 5A. Indeed, the only case where peptide stimulation has a seemingly higher signal over DMSO control is shown for the representative donor of Figure 5B. However, the authors did not show a response summary across several donors. The fact that none of the figures in Figure 5 have any statistical analysis makes it impossible to draw any conclusion. I would suggest making this section more scientifically

accurate by taking my suggestions above. Also, to call these responses “robust” in the abstract seems markedly exaggerated.

Response:

The heatmap which has now been added to 5A shows CD4 responses against pool 2, which are equally relevant for 5A and 5C. While some donors show a strong response, other donors respond at a lower level. Here, we use a very transparent approach to present all raw data and derivative bar graphs for the readers to be able to assess the results independently. As mentioned above, statistical analysis is not possible for this assay. We have changed the wording from robust to long-lived.

Discussion

- Too short. A much bigger emphasis should be placed on the overall conclusions and the reconciliation of the findings within the manuscript and the vast amount of literature. Additional considerations regarding the capability of the identified epitopes to cross-recognize common cold coronaviruses or different SARS-CoV-2 variants at the global population level should be discussed.
  Some additional limitations of the study, pointed out above, should also be added.

Response:

As outlined in a previous comment we have added a new section to the results, covering all reported ligands and epitopes from the IEDB to date, to have a defined basis for comparison. The main points from this analysis have been added to the now extended discussion, together with key considerations on the reconciliation with this data and what has been described in literature. We have furthermore assessed the conservation of the five identified peptides across all variants of concern, interest and being monitored as defined by the CDC. Here it is evident that the peptides are highly conserved, with only one single point mutation for E:58-73 identified in the Beta variants (B.1.351). This has been added to the discussion, and an overview of mutations and variants added as supplementary file S6_SARS-CoV-2_variants_coverage.xlsx.

Reviewer #3 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Response:

Noted.

Additional note:

During the review process we noticed a typo in Figure 3 and have changed it from 11457 to the correct number of 11357 mammalian 8-12-mers in the last line of Fig 3A. This does not change any of the of the main findings or messages of the paper.

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
The authors have provided satisfying responses to all of my recommendations.

Reviewer #2:
Remarks to the Author:
I would like to express my gratitude to the authors for their efforts in addressing my comments. However, I believe that many comments were overlooked and not adequately addressed from my perspective. As a result, some of my initial concerns, which I consider important for the interpretation and relevance of the findings, remain unsubstantiated. I still believe it is a very interesting study potentially limited by the sample availability.