nature portfolio

Peer Review File

Deep peptide recognition profiling decodes TCR specificity and enables disease-associated antigen discovery

Corresponding Author: Professor K. Christopher Garcia

Parts of this Peer Review File have been redacted as indicated as we could not obtain permission to publish the reports from one of the peer reviewers.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The manuscript describes an experimental pipeline to screen very large libraries of peptides displayed on MHC molecules for recognition by specific TCRs. Using a set of highly similar TCRs (same Va, same Vb, almost same Jb, almost same CDR3b length, restricted to the same allele (HLA-B27:05) with many if not all binding to the same YeiH epitope), the authors show that one can build Peptide Recognition Profiles which are powerful to screen peptides from the human proteome that can elicit TCR recognition. The authors further explore how small changes in CDR3b loops impact peptide recognition profiles and how conservation of peptide recognition profiles relates to CDR3 sequence similarity.

While I have several comments (and I hope they can contribute to improving the manuscript), I strongly believe that these types of studies (including the powerful yeast-display pipeline) are extremely useful to better understand TCR-epitope recognition and cross-reactivity.

Main comments

- It should be better emphasized that R at P2 and P at P8 are part of the design of the library (was not clear to me when reading the results). While I understand that this is useful to have higher yields of peptides recognized by the TCRs, this represents an important restriction for the Peptide Recognition Profile and exploration of cross-reactive peptides from the human proteome. This is especially the case for Pro at P8, which is not part of the motif of HLA-B*27:05. By fixing it, one excludes almost 95% of the human peptides (Pro frequency in the human proteome is ~6.3%), and there is not much evidence that none of these peptides could be recognized by the input TCRs.

- I found it difficult to understand where the different TCRs came from. First, there is almost no mention of the selection criteria in the main text. Then in the method, for several TCRs, like TCR 4.1, it just says that they were found by scTCR-Seq of expanded TRBV9+ cells. I doubt that only one TCR was found in these experiments, so the question is how this specific one was selected (based on Va, Jb, CDR3b length?). The second underlying question is whether all of them bind the same epitope (YeiH). This is important for the interpretation of the results.

- This leads to my main issue, namely that all TCRs used in this study are actually extremely similar, with the same Va, the same Vb, the same Jb (with two exceptions) and same CDR3b length (with one exception). As a result most PRPs show high similarity (which is somehow reassuring). The authors make several claims that TCR sequence similarity or sequence-based clustering does not correlate/predict PRP similarity (especially in Figure 6). However, this applies only within this set of pre-selected highly similar TCRs. In a real-life scenario, one is faced with very diverse TCRs with completely different sequences and completely different PRPs. I very strongly suspect that TCR sequence similarity (e.g., CDR1-CDR2-CDR3) is actually a very good predictor of PRP similarity. To address this point, what needs to be done is to take existing data (or generate data with the pMHC yeast-display) for other TCRs binding to unrelated epitopes (e.g., NY-ESO-1, GIL,…), derive PRPs and compare TCR sequence similarity with PRP similarity. This is almost 100% guaranteed to show that TCR sequence similarity actually correlates strongly with PRP similarity. For this reason, all claims that TCR sequence similarity does not correlate with PRP similarity should be removed, since they only apply within a biased set of pre-selected TCRs

that show extremely high sequence similarity. Overall, this is a rather good news, because it means that we can use global TCR sequence similarity (likely with a strict threshold) to map PRPs with a reasonable accuracy, as demonstrated by the data generated in this work. This does not exclude that one can do even slightly better with the approach proposed here for TCRs that are extremely similar, but overall, the differences between the predicted motifs and a simple average motif across all TCRs from this study appear to be modest, and this simple 'averaging' approach may fare equally well in several benchmarks.

- In Figure 2c, the authors observe clusters. However, apart from the manually drawn circle, I do not see much structure in these data, and I strongly suspect that the layout could considerably vary if other approaches were used (PCA,UMAP, tSNE) or other choices of parameters (Number of PCA components, centering of the data, normalization, random seed, etc). I'm also a bit puzzled that TCR 4.2, 4.3 and 4.4 are so far from each other, since they share this specificity for R/K at P9 (which I know is compatible with the HLA-B27:05 motif, even if it differs from the hydrophobic residues found in other cases).

- Can the authors come up with some explanations regarding the peculiar pattern in the PRP of TCR4.2/3/4, where the specificity at P9 differs from other cases? Looks like an interesting structural question where different TCRs recognize peptides with differences mostly in the P9 anchor.

- In Figure 4, the authors claim that their scores correlate with functional activation. However, first, it's not fully clear to me that this is really the case, as demonstrated in Figure S9 where only 2/16 cases pass statistical significance. I also think that the authors address an artificially much too challenging benchmarking scenario. In real life applications, one has to find a few peptides recognized by a specific TCRs out of a sea of self-peptides. I would strongly encourage the authors to consider this scenario in their work and include some (say 20) randomly selected peptides from the human proteome (maybe including the binding to B27 filtering criteria), and test them. This would provide much more convincing evidences that their tool actually works remarkably well at identifying some 'cross-reactive' peptides (and most likely the majority of TCRs would pass statistical significance in this more realistic scenario).

- I'm extremely puzzled by the results in Figure 5e. By looking at the motifs (Figure 5b), they all show very high similarity. I have extensive experience working with motifs, and whenever motifs show this level of similarity, training whatever predictors on one dataset or multiple ones makes very little difference (and if differences are found it's often due to the larger number of peptides when considering the multiple datasets). I strongly suspect that there may be some hidden overfitting or other biases (or just chance, or a bug somewhere in a code, or some specific choices in the architecture of the predictor,…) that resulted in the important increase in AUC in Figure 5e.

- In many parts, the authors claim that they work addresses the "TCR specificity" question. While I have no doubt that this study contributes to it, it's still quite far from solving it. For instance, all the observations (i.e. PRPs models, including the ability to predict them for new TCRs) likely apply only to TCRs showing the same Va, same Vb, same Jb, same CDRb length. A quick calculation based on frequency of AV21, BV9, (BJ2-7+BJ2-3+BJ2-2) and CDR3b_L15 in baseline TCR repertoires, shows that this represents less than 0.02% of a standard TCR repertoire from donors/patients. While this does not undermine the value of the work (and that the same limitations apply to many other studies), this should be clearly acknowledged, and some claims should be toned down. Along these lines, I feel that the title ("decodes TCR specificity") is a bit exaggerated, since the "decoding" only applies to a very very tiny fraction of all existing TCRs.

- I could not find the actual data in Supp, and therefore could not review this part.

Minor comments:

- In Figure 1c, the 26.2 motifs seems highly similar to 135.3 and 135.8. Does not sound like the best example, or maybe there were some confusions in the making of the Figure. This further underlines the fact that actually all PRPs found in this study show high similarity, likely because the input TCRs are very similar in their sequences.
- Line 52-56: While I agree that many examples (including work by the author of this manuscript) support this claim, I'm still wondering whether this is the exception or the rule. For instance, from a statistical point of view, TCRs binding to the same epitope do show more sequence similarity compared to random TCRs. Not a major point for this paper, but one should always be careful that looking too much at exceptions can obscure some of the statistical rules that are valid and are useful to make (still imperfect, of course) predictions.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
The manuscript by Wang et al. describes a novel and exciting, if relatively focused, approach to predict TCR recognition. The authors start by leveraging yeast display, using a system they recently focused on to identify TCRs and TCR ligands associated with B2705, relevant in ankylosing spondylitis. After profiling several TCRs with yeast display, they train a machine learning approach on this data, incorporating advances such as a fine tuned protein language model, but notably only focusing on CDRbeta chains of the TCR. From this data they generate a impressive tool for characterizing and

predicting molecular recognition in the system. Several conceptual advances are described that will be very useful for the field, such as the descriptions of PRPs, neighborhoods, validation through clever engineering, etc.

The prediction pipeline utilizes an innovative deep learning architecture, pairing transformer encoder layers with convolution to build highly sophisticated CDRbeta-peptide representations. These are then passed to a multilayer perceptron (MLP) for the final binding probability determination. Further, the authors subject the model to rigorous validation strategies, including a leave-one-TCR-out method of cross-validation, a gold standard approach in assessing generalizability. Taken together with protein language models being fine tuned on their PRP data, this work represents a potential major improvement in the predictive power and accuracy of sequence-based prediction models.

Despite this enthusiasm however there are significant limitations to the study. Although the machine learning performance is impressive, the study relies entirely on TCR CDR3beta chains. The authors claim that the TCRs show significant CDR loop diversity (line 112), but the CDR3beta chains of the TCRs studied are almost all identical (Fig. 2a). Thus the overall performance and generalization across TCR datasets are oversold: if a model that relies on CDR3beta is focused on TCRs that have essentially the same CDR3b, then strong performance across these receptors is to be expected. It is true that the TCRs have different CDR3alpha chains, and thus the claimed performance with TCRs of *overall* sequence is technically correct, but as the authors even point out, this system relies heavily on CDR3beta (line 104).

Overall then the predictive model from their statistics indeed appears strong but it will be highly focused on this select set of CDR3beta-similar TCRs. A much stronger approach would be to incorporate both chains and ask how a broader model reflects divergences between the TCRs and then wholly novel TCRs that actually have diversity in both chains.

Some more detailed concerns:

On lines 162-174, the authors describe the ability of the model to predict T cell activation. They report an AUC = 0.72 and an AUPRC = 0.42. Given the class imbalance (Fig. 4B), the PR baseline is only around 0.1 so the model does predict better than randomly. However, the authors must indicate baseline performance on the plots. Additionally, the AUC is not particularly helpful given class imbalance from the yeast display data.
The authors claim that the predictions are improved over structural approaches, using AlphaFold3 iPTM scores. This is a bit of a bait and switch – there is no indication that AF3 iPTM is useful at all for predicting T cell activation. There are reports showing that while AF3 *can* accurately predict some TCR-peptide:MHC complexes, in others it fails poorly, without clear indication from the confidence metrics. Additionally, later on in the Discussion the others talk about divergence between binding and function – given all this, even if AF3 iPTM scores were reliable, should we expect AF3 iPTM scores to predict T cell activation?

On this latter point, in the Discussion the authors mention the “often elusive gap between in vitro affinity and cellular response” (line 285). However, the training data is yeast display, which relies on TCR binding via tetramer as opposed to functional data. While they do indeed show power in predicting function, this statement is misleading given their methodology.

Thus, despite introducing new conceptual advances in how we think about specificity and its relationship to sequence, etc., the use of very powerful yeast display technology to identify new ligands to train on, introduction of concepts such as PRMs and neighborhoods, incorporation of structural biology and validation via mutagenesis, and a strong performing model, the approach is still very limited, and there are stretches in the interpretation of the data and results.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
The manuscript describes the collection of a large set (~10^6 peptides) of peptide recognition data for 21 closely related TCRs derived from AS and AAU patients. These data are then used to generate a machine learning based model relating TCR sequence to the peptide recognition space. The authors go on to show that the model generated by this method outperforms sequence-based predictions, and can be used to predict novel TCR recognition profiles (in a leave one out validation scenario). They also use the generated peptide profiles to identify a novel candidate self-antigen, which is then validated functionally.
There is much to commend in the paper, in that it provides a large set of peptide recognition data that is generally hard to obtain and will be of use to the field. The broader claim that the authors have advanced a new understanding of the TCR recognition problem is not supported by the analysis, as the authors have created a strawman of sorts that they utilize throughout the manuscript.
In addition to the revisions to the framing and clarity, the experiment in item 4 I think is essential.
1) The authors claim that their model refutes the utility of sequence based analysis for TCR prediction. First, their model explicitly utilizes the TCR sequence so this claim is quite baffling throughout. When they predict “unseen” TCRs (leave one out) their machine learning model is of course relying on some distance in abstract TCR space to predict its peptide recognition profile. It may be doing using a much more sophisticated model that projects novel features of that TCR sequence into some peptide recognition space, but it still starts with the sequence. It might seem an obvious point, but the

opacity of the writing in the paper does obscure the fact that they are just learning a better mapping of sequence to function.
2) The next issue is that the authors have addressed an important, but relatively “niche” problem and are comparing it to a much broader classification problem. This divergence in the scale of comparison leads to basically all their top line conclusions as a trivial consequence of the framing of the problem. The authors are examining 21 (or 16?) highly related TCRs. They say it is 21 TCRs, but I only see 16 in Figure 2. These TCRs are extremely close in TCR space, however you define it. All 16 use exactly the same TRAV. 12/15 use exactly the same TRBV, and the other 4 use another TRBV (so TRBVs total). The admittedly coarse grained nature of pure sequence based TCR analyses (GLIPH, TCRDist) will not have great resolution in these TCRs that are remarkably similar, particularly in the CDR3b which is the comparator in the manuscript. It is certainly interesting and even fascinating that they still observe such divergence in recognition space between such closely related TCRs, but the uses of sequence based measures have never been applied to such fine distinctions. This is just a point about the framing of the manuscript—they aren’t really testing the problems that other repertoire classification papers tackle in terms of the variation in TCRs. The appropriate framing would be that you need much more resolution in peptide target space once you are narrowed in on a such a small region of TCR space.
3) One of the more baffling choices in the paper is the focus on the CDR3b exclusively. This only works of course because they never vary the TCRa. They say this is because structural models suggest that the CDR3b is doing most of the work for at least one of the peptides examined (but who knows for all six million!). They also do acknowledge this as a limitation, but it’s actually fatal to the application of this model to anything beyond TCRs where the alpha is already known to maintain specificity—so totally useless for de novo acquired data. Clearly the alpha has some restrictions on it, which is why they are all TRAV21, and further it’s known that swapping alphas even in a beta dominated epitope-response can abolish specificity or radically modify it.
4) One does wonder how much the variation in positive peptides is alpha mediated. The fact that the models work without including alpha suggests (and leads the authors to conclude) that the alpha doesn’t matter, but it’s also possible the model is just overfit on a small number of TCRs. One experiment they should do is to start swapping the alphas among the betas and show that unique positive peptide hits and the PRP tracks the beta and not the alpha and that the model keeps working with such high accuracy. They would in fact predict that swapping the alphas does not matter at all, which is very unlikely to be the case.
5) A last point on the framing—the authors suggest this approach should replace sequence-based classification, which is entirely impractical. They analyzed 21 (or 16?) TCRs when one experiment regularly generates data—even some limited specificity data—on hundreds to thousands. There’s no acknowledgment that this approach doesn’t scale at all, and no assessment of how many TCRs minimally you have to do to build a valid model even in this narrow corner of TCR space. There are clearly practical applications of this approach—the discovery of a novel self-antigen is a great one—but it’s not how they’ve framed the entire manuscript. I suggest they remove the idea that this is the short cut to a generalizable TCR specificity model and focus instead on how this approach “blows up” narrow regions of TCR space with incredible specificity. This has immediate applications in cancer and autoimmunity (as they demonstrate).
6) Minor point: the intro is basically the abstract repeated, including copied sentences. There is a lot of repetition in the discussion as well.
7) Minor point: The figure 6 description both in the results and legends could be more reader friendly—ideally this would be a broad immunological audience. The Mahalanobis distance isn’t right at everyone’s fingertips. A little wayfinding would be helpful.

(Remarks on code availability)
I briefly looked at the github repository. They appear to provide code to download their learned models from a google drive, but they have not provided code to generate your own models from your own data. The training single TCR model section is empty with a "todo".

Reviewer #4

(Remarks to the Author)
The authors combined high throughput yeast display library screening of HLA-B27-bound peptides recognized by selected TCRs, with fine-tuned protein language models to generate Peptide Recognition Profiles (PRPs) for these TCRs. The main finding is that TCRs cluster by functional rather than sequence similarity. Importantly, the PRPs predicted T cell activation for select peptides with greater accuracy than existing modeling approaches. In addition to the general technical and conceptual advance, this led to the discovery of a novel candidate autoantigen derived from PSG5 (GRLPLLNPI) for which there was greater autoreactivity in HLA-B27+ patients compared to HLA-B27+ control subjects. Taken together, the work is novel and represents a significant advance in discovering potential autoreactive self-peptides.
Despite the advances shown here, I have some questions pertaining to the biological application and generalizability of this methodology.
1. How were the 21 HLA-B27-restricted TCRs selected? In the results (line 92-95) the authors state that ‘multiple HLA-B*27:05-restricted TCRs implicated in AS and AAU’ were used. What is the evidence that these 21 TCRs are implicated in AS/AAU? Have all of these TCRs been shown to be expanded in HLA-B27+ individuals compared to HLA-B27+ healthy controls? Please clarify and cite the studies where appropriate.
2. Irrespective of the proposed paradigm using PRPs over sequence similarity, which seems to be warranted based on the results, the generalizability of these findings is limited by the focus on only CDR3β modelling since this ignores the alpha-chain and CDR1/2 regions. For HLA-B27 specifically, several TCRs are beta-centric, but alpha-chain and CDR1/2 contacts with the alpha-helices can still modulate recognition thresholds and explain why some peptides are seen and others aren’t. Even in Fig. 4g, for example, there is a CDR1 -peptide contact. The fact that germline-encoded CDR1/2 loops could impact peptide contact is also noted in the previous manuscript (PMID 36477533). Since CDR1/2 often contact HLA helices and can influence docking geometry and peptide sensitivity, please justify the CDR3β-only assumption made in this model and

manuscript. This could be addressed experimentally by an ablation experiment addressing whether alpha/beta/CDR1/2 features alter peptide rankings for a representative subset. If feasible, a mutational or alpha-chain swap (between two related TCRs) would strengthen the conclusion that the observed peptide preferences are not contingent on germline-encoded contacts.
3. The authors should better justify and/or discuss the limitations of a peptide library where the HLA heavy chain has been mutated in positions that are likely to affect peptide binding, the P2 and P8 positions of the HLA-B27-bound peptide are fixed, all peptides are 9-mers, and the N- and C-termini of the peptide can't be in their normal orientation in the peptide binding groove since the single chain construct precludes this arrangement? This must have some structural consequences for the way peptide-MHC complex are seen by TCRs. Moreover, are the HLA-B27 constructs used in this study expressing bona fide HLA-B*27:05, or HLA-B27 with the Y84A, L5M, H114Y, and A153D mutations like the yeast display library? (Is Cys67 mutated as well?) If so, they should not be called HLA-B*27:05. Does the PSG5 peptide bind to HLA-B*27:05 or HLA-B27 with the four substitutions? If both, are there affinity differences?
4. In Fig. 4f, the expansion of CD8+ T cells recognizing the flu peptide is nearly as strong (~3.8 fold) as for the PSG5 peptide (~4.9 fold) in the individual shown. In Supplemental Fig. 10c, the lack of statistical significance for expanded B27/flu-specific CD8s is most likely due to insufficient power with only 6 patients and 6 controls. Unless there are differences in influenza exposure/infection, this calls into question the specificity of the 'autoreactive' response to the PSG5 peptide. More data are needed to substantiate this claim.
5. In the antigen presentation assays, peptide concentrations are 100 μM, which is very high, and could lead to non-specific binding. Please justify the use of this high concentration.
6. Do the 21 TCRs studied here recognize other HLA-B alleles?
7. Although more of a comment than a question, it seems to me that these results challenge the molecular mimicry hypothesis in its simplest form; i.e. a foreign antigenic peptide derived from a microbe and presented by HLA-B27 leads to an immune response that cross-reacts with an HLA-B27 self-peptide because of sequence similarity between the foreign and self-peptides. This has been suggested to occur for HLA-B27 because it theoretically presents self-peptides that resemble certain microbial peptides because of its unique peptide binding specificity. While sequence-based molecular mimicry could still occur, it seems the situation is far more complicated, raising again the question of why HLA-B27 is so strongly linked to this disease.

(Remarks on code availability)

Author Rebuttal letter:

General responses:
We thank the reviewers for their insightful feedback. A key concern raised by multiple reviewers was the generalizability of our findings, given the focus on a specific, HLA-B*27:05-restricted TCR family. We were remiss in not clarifying this important point of distinction of our paper to the readers. In revisions, we now clarify that this system was deliberately chosen because of the exclusive peptide recognition by CDR3-beta, which allows us to visualize and rationalize the cross-reactivity data and predictions with extremely high focus and precision and without interference of confounding factors like the role of other CDRs in peptide specificity. Thus the structure-activity relationships we can infer in our work are extremely granular. This enables high-resolution functional mapping within a disease-associated TCR neighborhood, serving as a biologically grounded benchmark that complements rather than replaces repertoire-scale analyses.

The key revisions include:
1. New α-chain swap experiments and complementary structural analyses confirming that β-chain features predominantly drive peptide specificity in this clonotype family.
2. Revised the Abstract, Results, and Discussion to emphasize that our platform provides fine-grained resolution within local TCR neighborhoods and is designed to complement repertoire-scale sequence-based approaches.
3. Text revisions throughout to specify TCR provenance, emphasize conclusions within this defined family, and outline extensions to diverse repertoires.

Point-to-point responses:
Reviewer #1
The manuscript describes an experimental pipeline to screen very large libraries of peptides displayed on MHC molecules for recognition by specific TCRs. Using a set of highly similar TCRs (same Va, same Vb, almost same Jb, almost same CDR3b length, restricted to the same allele (HLA-B27:05) with many if not all binding to the same YeiH

epitope), the authors show that one can build Peptide Recognition Profiles which are powerful to screen peptides from the human proteome that can elicit TCR recognition. The authors further explore how small changes in CDR3b loops impact peptide recognition profiles and how conservation of peptide recognition profiles relates to CDR3 sequence similarity.

While I have several comments (and I hope they can contribute to improving the manuscript), I strongly believe that these types of studies (including the powerful yeast-display pipeline) are extremely useful to better understand TCR-epitope recognition and cross-reactivity.
Main points:
1. It should be better emphasized that R at P2 and P at P8 are part of the design of the library (was not clear to me when reading the results). While I understand that this is useful to have higher yields of peptides recognized by the TCRs, this represents an important restriction for the Peptide Recognition Profile and exploration of cross-reactive peptides from the human proteome. This is especially the case for Pro at P8, which is not part of the motif of HLA-B*27:05. By fixing it, one excludes almost 95% of the human peptides (Pro frequency in the human proteome is ~6.3%), and there is not much evidence that none of these peptides could be recognized by the input TCRs.

We agree that fixing P2 (Arg) and P8 (Pro) narrows the peptide space relative to the full HLA-B*27:05 repertoire. In prior work (Yang et al., Nature 2022), an unconstrained P8 yielded Pro convergence across TCR selections (e.g., below - round 4 enrichments for TCRs 19.2 and 135.4; see figure below). Thus, we fixed P8 as Pro to allocate diversity to more informative positions (e.g., P9) while capturing family-specific features. This was a necessary trade-off to maximize library stability and functional diversity at the central TCR-contact residues within the constraints of yeast display transformation efficiency.

Revision: We have revised the Methods page 15 line #515-516, Results page 4 line #96-102 and Discussion page 12 line #390-397 to explicitly acknowledge this constraint. We clarify that while this design limits the discovery of non-canonical binders, the PRP framework itself is modular and can be adapted to other anchor motifs in future studies.
2. I found it difficult to understand where the different TCRs came from. First, there is almost no mention of the selection criteria in the main text. Then in the method, for several TCRs, like TCR 4.1, it just says that they were found by scTCR-Seq of expanded TRBV9+ cells. I doubt that only one TCR was found in these experiments, so the question is how this specific one was selected (based on Va, Jb, CDR3b length?). The second underlying question is whether all of them bind the same epitope (YeiH). This is important for the interpretation of the results.

We apologize for the lack of clarity. We analyzed 21 TCRs total: 16 disease-derived for training/analysis (Figs. 1b, 2a) and 5 engineered 19.2 variants for mutagenesis validation only (Fig. 5). The 16 derive from three sources (detailed in revised Methods and Fig. 1a legend):
1) BV9–Y/FSTDTQ–BJ2.3/TRAV21 public family (Yang et al., Nature 2022): TCRs 4.1–4.4, 8.4, 9.1 from scTCR-seq of TRBV9-enriched AS/AAU samples, selected for the disease-enriched motif.
2) Non-BV9 example (TCR019.1; Paley & Yang et al., JCI Insight 2024): From AAU aqueous humor scTCR-seq; functional assays confirmed YeiH recognition despite alternate β germline.
3) YeiH-tetramer–sorted clonotypes (Paley & Yang et al., JCI Insight 2024): From expanded PBMCs of HLA-B*27:05+ AAU donors, yielding TRBV9/TRBV5 users beyond the public family.

Regarding whether all TCRs in our panel bind the same YeiH epitope: a subset of receptors were isolated directly by YeiH–HLA-B*27:05 tetramer staining, and others were obtained independently from synovial or ocular fluid based on their enrichment for the disease-associated CDR3β motif that has been functionally linked to YeiH. Together, the panel therefore spans both tetramer-sorted YeiH-specific TCRs and motif-guided receptors enriched in AS/AAU contexts, rather than being restricted to a single tetramer-defined clonotype.

Revision: We have added a dedicated section in the Methods page 12 line #413-431 detailing the source of each TCR (synovial fluid, ocular fluid, or tetramer sort) and updated Figure 1a to visually distinguish between the training set and the engineered validation set.

3. This leads to my main issue, namely that all TCRs used in this study are actually extremely similar, with the same Va, the same Vb, the same Jb (with two exceptions) and same CDR3b length (with one exception). As a result most PRPs show high similarity (which is somehow reassuring). The authors make several claims that TCR sequence similarity or sequence-based clustering does not correlate/predict PRP similarity (especially in Figure 6). However, this applies only within this set of pre-selected highly similar TCRs. In a real-life scenario, one is faced with very diverse TCRs with completely different sequences and completely different PRPs. I very strongly suspect that TCR sequence similarity (e.g., CDR1-CDR2-CDR3) is actually a very good predictor of PRP similarity. To address this point, what needs to be done is to take existing data (or generate data with the pMHC yeast-display) for other TCRs binding to unrelated epitopes (e.g., NY-ESO-1, GIL,…), derive PRPs and compare TCR sequence similarity with PRP similarity. This is almost 100% guaranteed to show that TCR sequence similarity actually correlates strongly with PRP similarity. For this reason, all claims that TCR sequence similarity does not correlate with PRP similarity should be removed, since they only apply within a biased set of pre-selected TCRs that show extremely high sequence similarity. Overall, this is a rather good news, because it means that we can use global TCR sequence similarity (likely with a strict threshold) to map PRPs with a reasonable accuracy, as demonstrated by the data generated in this work. This does not exclude that one can do even slightly better with the approach proposed here for TCRs that are extremely similar, but overall, the differences between the predicted motifs and a simple average motif across all TCRs from this study appear to be modest, and this simple 'averaging' approach may fare equally well in several benchmarks.

We appreciate this perspective. We agree that across the global repertoire, sequence similarity is a strong predictor of specificity (as leveraged by GLIPH/TCRdist). However, our study focuses on a related but different hard problem that is frequently encountered: distinguishing functional divergence within a "neighborhood" of highly similar TCRs where standard clustering methods often merge distinct specificities. Thus, we introduce both the platform and data resource to probe fine-scale divergence within a near-clonal, disease-linked family (conserved TRAV21; TRBV9/5; similar CDR3β). Our cross-reactivity screens show how fine-specificity is shaped by even conservative point mutations in CDR3β that would not be apparent from sequence similarity approaches.

Revision: We have softened the language that implied sequence similarity "fails" broadly. Instead, we now frame our results to show that within high-identity TCR families, PRPs provide the necessary resolution that global sequence metrics cannot. We emphasize that our approach complements repertoire-scale clustering by resolving these dense local neighborhoods in Introduction page 3 line #55-57 and Discussion page11 line #381-389.

4. In Figure 2c, the authors observe clusters. However, apart from the manually drawn circle, I do not see much structure in these data, and I strongly suspect that the layout could considerably vary if other approaches were used (PCA,UMAP, tSNE) or other choices of parameters (Number of PCA components, centering of the data, normalization, random seed, etc). I'm also a bit puzzled that TCR 4.2, 4.3 and 4.4 are so far from each other, since they share this specificity for R/K at P9 (which I know is compatible with the HLA-B27:05 motif, even if it differs from the hydrophobic residues found in other cases).

To confirm robustness, we re-analyzed PRPs via PCA, UMAP, and t-SNE with varying PCs, normalization, seeds, etc. We now provide representative PCA and UMAP embeddings (Supplementary Fig. 3c, d) to demonstrate robustness and clarify that our conclusions rely on quantitative JS distance (Fig. 2), not on any single 2D projection.

Regarding TCR 4.2/4.3/4.4, we clarify in the Results that their distinct P9 preference is partly driven by the scaffold constraints of the library used for those specific receptors (from Yang et al., 2022), and we mask these positions during quantitative distance calculations to prevent artifacts. We have revised the Methods page 15 line #515-516 and Supplementary Fig. 2 legend to make these distinctions explicit.

Revision: We have performed sensitivity analyses included in Supplementary Fig. 3. The separation of functional clusters remains robust across methods.

5. Can the authors come up with some explanations regarding the peculiar pattern in the PRP of TCR4.2/3/4, where the specificity at P9 differs from other cases? Looks like

an interesting structural question where different TCRs recognize peptides with differences mostly in the P9 anchor.

We now clarify that the P9 pattern for TCR 4.2/4.3/4.4 is primarily a scaffold artifact rather than a distinct biological preference. For these three TCRs (Yang et al., Nature 2022), P2 and P9 were fixed (to Lys/Arg), while P8 was unconstrained and converged to Pro. For the remaining TCRs in this study, we fixed P2 and P8 (Pro) and allowed P9 to vary. Thus, for 4.2/4.3/4.4, P9 carries little biological information; the apparent divergence at P9 reflects library design rather than differential recognition of variable P9 residues.

From a computational analysis standpoint, we handle this by masking constrained positions when computing distance and similarity metrics to avoid driving similarity based on library composition.

Revision: Our Methods highlight this in sections "Motif analysis and sequence logos" in page 18 and "Calculating TCR distances" in page 20. In sum, the P9 peculiarity for TCR4.2/4.3/4.4 reflects scaffold constraints rather than a biologically distinct P9 preference, and our conclusions are robust to analyses that control for this factor. We have revised the Methods page 15 line #515-516.

6. In Figure 4, the authors claim that their scores correlate with functional activation. However, first, it's not fully clear to me that this is really the case, as demonstrated in Figure S9 where only 2/16 cases pass statistical significance. I also think that the authors address an artificially much too challenging benchmarking scenario. In real life applications, one has to find a few peptides recognized by a specific TCRs out of a sea of self-peptides. I would strongly encourage the authors to consider this scenario in their work and include some (say 20) randomly selected peptides from the human proteome (maybe including the binding to B27 filtering criteria), and test them. This would provide much more convincing evidences that their tool actually works remarkably well at identifying some 'cross-reactive' peptides (and most likely the majority of TCRs would pass statistical significance in this more realistic scenario).

This is an excellent suggestion. To rigorously test the model's ability to discriminate true antigens from high-affinity non-stimulatory peptides, we selected the top 20 strongest predicted HLA-B*27:05 binders from the human proteome (using NetMHCpan) and tested them experimentally.

Revision: As shown in Supplementary Fig. 12, none of these high-affinity binders induced activation in the AS/AAU TCRs. This highlights that our PRP model captures fine-grained TCR recognition features that standard MHC-binding predictions miss.

7. I'm extremely puzzled by the results in Figure 5e. By looking at the motifs (Figure 5b), they all show very high similarity. I have extensive experience working with motifs, and whenever motifs show this level of similarity, training whatever predictors on one dataset or multiple ones makes very little difference (and if differences are found it's often due to the larger number of peptides when considering the multiple datasets). I strongly suspect that there may be some hidden overfitting or other biases (or just chance, or a bug somewhere in a code, or some specific choices in the architecture of the predictor,…) that resulted in the important increase in AUC in Figure 5e.

We took strict measures to prevent data leakage. We appreciate this concern and have explicitly checked for data leakage between training and activation datasets:

1) The yeast-display data used for training and the T cell activation data used for evaluation represent distinct experimental readouts.
2) We systematically compared all training peptides to those in the activation assays and found only one overlapping peptide (PRPF3).
3) Removing this peptide caused a slight decrease in AUC and average precision, indicating that the observed performance is not driven by overlap.
4) We did not evaluate joint-training performance on the deep-sequencing data themselves, precisely to avoid potential leakage when pooling datasets.

Revision: We performed an overlap analysis and found only a single peptide (PRPF3) shared between training and activation sets. Removing this peptide resulted in negligible changes to AUC/AUPRC, confirming that performance is driven by learned features, not memorization. We have added this analysis to the Methods in sections " Engineered TCR Variants and α-Chain Swaps" in page 13. Finally, we have clarified in

the figure legends which specific datasets were used for the analyses presented in Figure 5b and 5d.

8. In many parts, the authors claim that their work addresses the “TCR specificity” question. While I have no doubt that this study contributes to it, it’s still quite far from solving it. For instance, all the observations (i.e. PRPs models, including the ability to predict them for new TCRs) likely apply only to TCRs showing the same Va, same Vb, same Jb, same CDRb length. A quick calculation based on frequency of AV21, BV9, (BJ2-7+BJ2-3+BJ2-2) and CDR3b_L15 in baseline TCR repertoires, shows that this represents less than 0.02% of a standard TCR repertoire from donors/patients. While this does not undermine the value of the work (and that the same limitations apply to many other studies), this should be clearly acknowledged, and some claims should be toned down. Along these lines, I feel that the title (“decodes TCR specificity”) is a bit exaggerated, since the “decoding” only applies to a very very tiny fraction of all existing TCRs.

We recognize that the present study addresses TCR specificity within a restricted subset of clonotypes. However, our dataset does not rely on a single Vβ germline: in addition to the well-characterized TRBV9 family, we also include TCRs utilizing TRBV5, thereby extending the analysis beyond a single β chain. We now clarify this point in the Results page 5 line #128 and Methods page 13 line #430 to avoid the impression that all TCRs shared identical V gene usage.

Our intent with the term "decodes" is to reflect the depth of our analysis rather than its breadth. Our platform successfully models the complex, non-linear rules of TCR-peptide recognition within a reductionist, ground-truth model system. To address the reviewer’s concern about generalizability, we have extensively revised the Introduction and Discussion to explicitly frame this work as a high-resolution analysis of a specific TCR family. We believe these textual clarifications ensure the title is read in its intended, focused context.
Revision: We have revised the main text to clearly delineate the scope of our conclusions as decoding peptide specificity for this AS/AAU-associated TCR family, rather than for the entire TCR repertoire, while retaining the title to reflect this focused but comprehensive analysis.

9. I could not find the actual data in Supp, and therefore could not review this part.

All processed data have now been uploaded and are available in the Supplementary Tables and the associated data repository in page 21. To make this more transparent, we have revised the manuscript to explicitly direct readers to these files.

We hope this resolves the reviewer’s concern and will ensure that all data can be easily accessed and reviewed.

Minor points:
1. In Figure 1c, the 26.2 motifs seem highly similar to 135.3 and 135.8. Does not sound like the best example, or maybe there were some confusions in the making of the Figure. This further underlines the fact that actually all PRPs found in this study show high similarity, likely because the input TCRs are very similar in their sequences.

We have carefully re-checked the motif alignments and updated Figure 1c to avoid potential confusion. Specifically, we replaced the motif for TCR 26.2 with TCR 019.1 that is less similar to 135.3 and 135.8, so as to better illustrate the diversity captured by PRPs.

2. Line 52-56: While I agree that many examples (including work by the author of this manuscript) support this claim, I’m still wondering whether this is the exception or the rule. For instance, from a statistical point of view, TCRs binding to the same epitope do show more sequence similarity compared to random TCRs. Not a major point for this paper, but one should always be careful that looking too much at exceptions can obscure some of the statistical rules that are valid and are useful to make (still imperfect, of course) predictions.

We agree that, from a statistical perspective, TCRs recognizing the same epitope tend to be more similar to each other than to random TCRs, and that this property has been leveraged by methods such as GLIPH and TCRdist. Our statement in previous lines 52–56 was intended to highlight well-documented exceptions to this trend, which motivate the need for higher-resolution approaches such as PRPs. In response to the reviewer’s

comment, we have revised the Introduction in page 3 line #55-57 and Discussion in page 12 line #381-389 to explicitly acknowledge both aspects: (i) that broad statistical rules of sequence similarity remain valid and useful for prediction, and (ii) that exceptions exist, and these exceptions are precisely where our framework provides added value.

Reviewer #2:

The manuscript by Wang et al. describes a novel and exciting, if relatively focused, approach to predict TCR recognition. The authors start by leveraging yeast display, using a system they recently focused on to identify TCRs and TCR ligands associated with B2705, relevant in ankylosing spondylitis. After profiling several TCRs with yeast display, they train a machine learning approach on this data, incorporating advances such as a fine tuned protein language model, but notably only focusing on CDRbeta chains of the TCR. From this data they generate an impressive tool for characterizing and predicting molecular recognition in the system. Several conceptual advances are described that will be very useful for the field, such as the descriptions of PRPs, neighborhoods, validation through clever engineering, etc.

The prediction pipeline utilizes an innovative deep learning architecture, pairing transformer encoder layers with convolution to build highly sophisticated CDRbeta-peptide representations. These are then passed to a multilayer perceptron (MLP) for the final binding probability determination. Further, the authors subject the model to rigorous validation strategies, including a leave-one-TCR-out method of cross-validation, a gold standard approach in assessing generalizability. Taken together with protein language models being fine tuned on their PRP data, this work represents a potential major improvement in the predictive power and accuracy of sequence-based prediction models.

Despite this enthusiasm however there are significant limitations to the study. Although the machine learning performance is impressive, the study relies entirely on TCR CDR3beta chains. The authors claim that the TCRs show significant CDR loop diversity (line 112), but the CDR3beta chains of the TCRs studied are almost all identical (Fig. 2a). Thus the overall performance and generalization across TCR datasets are oversold: if a model that relies on CDR3beta is focused on TCRs that have essentially the same CDR3b, then strong performance across these receptors is to be expected. It is true that the TCRs have different CDR3alpha chains, and thus the claimed performance with TCRs of *overall* sequence is technically correct, but as the authors even point out, this system relies heavily on CDR3beta (line 104).

Overall then the predictive model from their statistics indeed appears strong but it will be highly focused on this select set of CDR3beta-similar TCRs. A much stronger approach would be to incorporate both chains and ask how a broader model reflects divergences between the TCRs and then wholly novel TCRs that actually have diversity in both chains.

Major points:

1. On lines 162-174, the authors describe the ability of the model to predict T cell activation. They report an AUC = 0.72 and an AUPRC = 0.42. Given the class imbalance (Fig. 4B), the PR baseline is only around 0.1 so the model does predict better than randomly. However, the authors must indicate baseline performance on the plots. Additionally, the AUC is not particularly helpful given class imbalance from the yeast display data.

We believe the comment refers to Fig. 4d. In response, we have added a baseline to the PR curve in the original Fig. 4d as well as to all other PR plots (Figs. 3b, 5b, 5d, and 6b).

Revision: We have updated Figure 4d and all Precision-Recall (PR) plots to explicitly show the random baseline (fraction of positive samples). We moved the AUC plots for the highly imbalanced yeast datasets to Supplementary Fig. 4 as suggested, retaining PR curves in the main text to better visualize performance on imbalanced data.

2. The authors claim that the predictions are improved over structural approaches, using AlphaFold3 iPTM scores. This is a bit of a bait and switch – there is no indication that AF3 iPTM is useful at all for predicting T cell activation. There are reports showing that while AF3 *can* accurately predict some TCR-peptide:MHC complexes, in others it fails poorly, without clear indication from the confidence metrics. Additionally, later on in the Discussion the others talk about divergence between binding and function – given all this, even if AF3 iPTM scores were reliable, should we expect AF3 iPTM scores to predict T cell activation?

We agree that AF3 and similar methods optimize for structural stability rather than functional activation. However, because these tools are frequently used by the community, we felt it was critical to empirically benchmark whether structural confidence scores correlate with actual T cell triggering in this system.

Revision: We have added “AlphaFold3 and tFold predictions” in Method page 21 and clarified in the Results line #215 that AF3 and tFold-TCR serve as structural baselines, not direct competitors. We further emphasized that our model’s superior performance arises from learning peptide-TCR recognition rules directly from experimental data, rather than from structural confidence.

3. On this latter point, in the Discussion the authors mention the “often elusive gap between in vitro affinity and cellular response” (line 285). However, the training data is yeast display, which relies on TCR binding via tetramer as opposed to functional data. While they do indeed show power in predicting function, this statement is misleading given their methodology.
The “gap” we refer to is the conceptual gap between binding and activation; our work shows that models trained on enriched binding data can still capture aspects of functional specificity, but we do not claim to train directly on functional measurements. We agree that our training data are derived from yeast-display selections, which measure binding via peptide-MHC multimer enrichment rather than direct T cell functional readouts. The superior performance of our model (new Figure 4d) confirms that learning from high-throughput interaction data captures functional rules that current structural models do not.

Revision: We revised the Discussion in page 11 line #344-348 to clearly distinguish between the binding data used for training and the functional data used for validation. Our intention in the Discussion was not to imply that we trained directly on functional data, but rather to emphasize that the relationship between binding and activation is complex, and that predictive models must account for this gap.

4. Thus, despite introducing new conceptual advances in how we think about specificity and its relationship to sequence, etc., the use of very powerful yeast display technology to identify new ligands to train on, introduction of concepts such as PRMs and neighborhoods, incorporation of structural biology and validation via mutagenesis, and a strong performing model, the approach is still very limited, and there are stretches in the interpretation of the data and results.

We recognize that the present work focuses on a restricted TCR family and is trained on yeast-display data and thus is not intended as a general solution to TCR specificity. Our goal is to introduce PRPs, local TCR neighborhoods, and their integration with machine learning as a framework for resolving fine-grained functional heterogeneity within a disease-associated clonotype family. This β-centric design allows us to isolate the dominant CDR3β-driven interactions and define interpretable peptide-recognition rules at high resolution, complementing rather than replacing broader repertoire-scale analyses.

Revision: We have revised the Discussion in page 11 line #360-364 to more carefully frame our contributions as complementary and system-level in scope. Please also see our detailed response to Reviewer #3 (Point 4) regarding the new α-chain swap experiments, setting a baseline for dual-chain extensions.
Reviewer #3:
The manuscript describes the collection of a large set (~10^6 peptides) of peptide recognition data for 21 closely related TCRs derived from AS and AAU patients. These data are then used to generate a machine learning based model relating TCR sequence to the peptide recognition space. The authors go on to show that the model generated by this method outperforms sequence-based predictions, and can be used to predict novel TCR recognition profiles (in a leave-one-out validation scenario). They also use the generated peptide profiles to identify a novel candidate self-antigen, which is then validated functionally.

There is much to commend in the paper, in that it provides a large set of peptide recognition data that is generally hard to obtain and will be of use to the field. The broader claim that the authors have advanced a new understanding of the TCR recognition problem is not supported by the analysis, as the authors have created a strawman of sorts that they utilize throughout the manuscript.

In addition to the revisions to the framing and clarity, the experiment in item 4 I think is

essential.

This reviewer raised several specific points that are addressed below:
Major points:
1. The authors claim that their model refutes the utility of sequence based analysis for TCR prediction. First, their model explicitly utilizes the TCR sequence so this claim is quite baffling throughout. When they predict “unseen” TCRs (leave one out) their machine learning model is of course relying on some distance in abstract TCR space to predict its peptide recognition profile. It may be doing using a much more sophisticated model that projects novel features of that TCR sequence into some peptide recognition space, but it still starts with the sequence. It might seem an obvious point, but the opacity of the writing in the paper does obscure the fact that they are just learning a better mapping of sequence to function.

We agree. Our goal is to show that simple similarity measures on CDR3β sequence (e.g., edit distance, motif clustering) are insufficient to capture fine-grained differences in PRPs within this narrow TCR family, not that sequence is uninformative. Our argument is not that sequence is uninformative, but that simple edit-distance metrics are insufficient for resolving determining specificity within this dense TCR cluster.

Revision: We have revised the Introduction in page 3 line #55-57 and #84-87 to remove "refutation" language. We now frame the work as using ML to learn a non-linear mapping from sequence to function, which outperforms linear distance metrics in this context.

2. The next issue is that the authors have addressed an important, but relatively “niche” problem and are comparing it to a much broader classification problem. This divergence in the scale of comparison leads to basically all their top line conclusions as a trivial consequence of the framing of the problem. The authors are examining 21 (or 16?) highly related TCRs. They say it is 21 TCRs, but I only see 16 in Figure 2. These TCRs are extremely close in TCR space, however you define it. All 16 use exactly the same TRAV. 12/15 use exactly the same TRBV, and the other 4 use another TRBV (so TRBVs total). The admittedly coarse grained nature of pure sequence based TCR analyses (GLIPH, TCRDist) will not have great resolution in these TCRs that are remarkably similar, particularly in the CDR3b which is the comparator in the manuscript. It is certainly interesting and even fascinating that they still observe such divergence in recognition space between such closely related TCRs, but the uses of sequence based measures have never been applied to such fine distinctions. This is just a point about the framing of the manuscript—they aren’t really testing the problems that other repertoire classification papers tackle in terms of the variation in TCRs. The appropriate framing would be that you need much more resolution in peptide target space once you are narrowed in on a such a small region of TCR space.

We recognize that the present study addresses TCR specificity within a restricted subset of disease-associated public clonotypes.

Revision: We revise the Discussion in page 11 line #360-364 to better emphasize that repertoire-scale methods and our local, high-resolution approach address different scales of the specificity problem. Our contribution is to show how much functional diversity can exist even within near-clonal families and to provide both the platform and resources for dissecting it.

Regarding the number of receptors, our dataset includes 21 TCRs in total, but only the 16 disease-derived receptors were used for model training and clustering (Fig. 2). The remaining 5 receptors are engineered variants of clonotype 19.2 used solely for mutagenesis validation (Fig. 5). We have updated the Method page13 “Engineered TCR Variants and α-Chain Swaps” and Figure 5a legends to make this breakdown explicit.

3. One of the more baffling choices in the paper is the focus on the CDR3b exclusively. This only works of course because they never vary the TCRa. They say this is because structural models suggest that the CDR3b is doing most of the work for at least one of the peptides examined (but who knows for all six million!). They also do acknowledge this as a limitation, but it’s actually fatal to the application of this model to anything beyond TCRs where the alpha is already known to maintain specificity—so totally useless for de novo acquired data. Clearly the alpha has some restrictions on it, which is why they are all TRAV21, and further it’s known that swapping alphas even in a beta dominated epitope-response can abolish specificity or radically modify it.

We recognize that focusing exclusively on CDR3β limits the general applicability of the model. We were remiss in not emphasizing that this choice was deliberate and reflects a unique opportunity presented by this disease-associated TCR family. High resolution structural analyses (see figure below) have shown that in these HLA-B*27:05-restricted clonotypes, CDR3β loops dominate peptide contacts, with germline-encoded CDR1α playing a supporting role, whereas CDR3α sits farther from the peptide with minimal contact, and its sequences, which are very diverse, appear largely shaped by neutral drift, and contributes minimally to peptide discrimination.

This unusual β-centric architecture provides a clean, reductionist system in which sequence–function coupling can be mapped with high precision, without the confounding variability typically introduced by CDR3α. This system offers a valuable "ground truth" for defining the limits of predictive modeling when sequence–function relationships can be studied in isolation.

Revision: We have revised the Result in page 5 #131-139 describing the HLA-B*27:05 restricted TCR panel to explicitly explain the β-centric architecture and conserved CDR1α contacts, and have added a new structural panel (Fig. 2b) highlighting these features.

4. One does wonder how much the variation in positive peptides is alpha mediated. The fact that the models work without including alpha suggests (and leads the authors to conclude) that the alpha doesn't matter, but it's also possible the model is just overfit on a small number of TCRs. One experiment they should do is to start swapping the alphas among the betas and show that unique positive peptide hits and the PRP tracks the beta and not the alpha and that the model keeps working with such high accuracy. They would in fact predict that swapping the alphas does not matter at all, which is very unlikely to be the case.

In our system α-chain variation can modulate TCR recognition indirectly and that α-β pairing contributes to directly to fine-tuning specificity in many systems. In this study, the α-chain was held constant deliberately - not because it is unimportant, but to control for its variability and test how well functional peptide recognition can be captured from β-chain information alone. The strong correlation between model predictions and experimentally measured activation indicates that, within this well-defined HLA-B*27:05-restricted clonotype family, the β-chain encapsulates the dominant sequence features governing peptide specificity.

To directly assess whether α-chain variation influences peptide recognition in this system, we employed a three-tiered validation strategy combining structural analysis, computational modeling, and the suggested experimental chain-swapping.

(1) Structural Evidence (CDR3 β Dominance)
As we illustrate in Point #3, we analyzed crystal structures of TCRs within this family complexed with peptide-HLA-B*27:05. As shown in Point #3, peptide recognition is structurally dominated by the CDR3β loop, which makes extensive contacts with the peptide backbone and side chains. In contrast, the CDR3α loop is positioned away from the peptide-binding groove. Consequently, the observed variation in CDR3α sequences across this family likely reflects neutral drift rather than antigen-driven selection. This structural reality provides the biological justification for our β-centric modeling approach.

(2) Computational Verification (Feature Importance)
To test this hypothesis quantitatively, we retrained our machine learning models to include CDR3α sequences as input features (additive and stacked encoding). As shown in Supplementary Fig. 6a-b, the inclusion of CDR3α data did not yield a statistically significant improvement in predictive performance (Wilcoxon $p = 0.39$). This confirms that, for this specific biological system, the information content required to predict peptide specificity is contained almost entirely within the β -chain.

(3) Experimental Validation (α -Chain Swaps)
Following the reviewer's suggestion, we performed α-chain swap assays among representative TCRs. Because J-region pairing can affect α–β compatibility, we selected three α-chains (19.2a, 8.4a, and 135.4a) sharing the AJ28 segment, fixed the highly cross-reactive β-chain (19.2b), and substituted α-chains accordingly. As shown in the activation data (new Supplementary Fig. 6c), α-swaps led to only minor quantitative differences in CD69 induction, with overall activation patterns closely resembling that of

the parental 19.2 TCR and remaining distinct from those of the 8.4 or 135.4 TCRs. Consistent with structural analyses, these results indicate that β-chain identity dictates the dominant peptide-recognition pattern, while α-swaps preserving TRAV21 pairing exert minimal influence on specificity.

We further evaluated α-chain contributions computationally by incorporating CDR3α sequences into the model via additive or stacked encoding. As shown in the ROC and PR curves (new Supplementary Fig. 6a,b), including CDR3α features did not improve performance (Wilcoxon $p = 0.39$), indicating that β-derived features alone are sufficient to recapitulate activation patterns in this system.

Importantly, these results should not be interpreted to mean that α-chains are generally irrelevant for TCR specificity. Rather, they underscore that within this clonotypically constrained, β-dominated family, peptide specificity can be predicted from β-chain information with high fidelity. This provides a rigorous baseline for extending the framework to more heterogeneous α-β combinations in future studies, where both chains and germline segments can be incorporated to dissect their additive and modulatory roles.

Revision: As shown in the new Supplementary Fig. 6, the α-swaps resulted in minimal changes to the peptide activation profile, which continued to track the parental β-chain identity. Furthermore, computationally adding CDR3α features to the model did not improve performance (Supplementary Fig. 6a,b). This confirms that in this specific disease-associated family, the β-chain is the dominant determinant of specificity, validating our modeling choice. We have revised the Results in page 6 line #175-186

5. A last point on the framing—the authors suggest this approach should replace sequence-based classification, which is entirely impractical. They analyzed 21 (or 16?) TCRs when one experiment regularly generates data—even some limited specificity data—on hundreds to thousands. There's no acknowledgment that this approach doesn't scale at all, and no assessment of how many TCRs minimally you have to do to build a valid model even in this narrow corner of TCR space. There are clearly practical applications of this approach—the discovery of a novel self-antigen is a great one—but it's not how they've framed the entire manuscript. I suggest they remove the idea that this is the shortcut to a generalizable TCR specificity model and focus instead on how this approach "blows up" narrow regions of TCR space with incredible specificity. This has immediate applications in cancer and autoimmunity (as they demonstrate).

We agree that our approach is not a replacement for high-throughput repertoire classification. Our dataset of 16 disease-derived clonotypes plus 5 engineered variants was deliberately designed as a reductionist system to probe how peptide recognition can diverge even among highly similar TCRs. The central message is that, within such near-clonal families, sequence similarity alone can fail to predict functional specificity, underscoring the need for empirical peptide-level mapping in representative systems.

Revision: We have revised the Discussion in page 11 line #325-331 and #340-343 to frame our platform as a "magnifying glass" for deep interrogation of specific, high-value clonotypes (e.g., in autoimmunity or neoantigen discovery), rather than a tool for broad repertoire binning.

Minor points:
1. the intro is basically the abstract repeated, including copied sentences. There is a lot of repetition in the discussion as well.

We have streamlined the Introduction to avoid repetition of the Abstract and removed redundant paragraphs in the Discussion in page 11 line #325-397, focusing on key conceptual advances and limitations.

2. The figure 6 description both in the results and legends could be more reader friendly—ideally this would be a broad immunological audience. The Mahalanobis distance isn't right at everyone's fingertips. A little wayfinding would be helpful.

We have revised the description of Figure 6 in both the Results in page 10 line #305-309 and #315-320 and the Figure legend to make it more accessible for a broad immunology audience. Specifically, we now provide a plain-language explanation of the Mahalanobis distance metric and added guiding text to help readers interpret the plots.

3.I briefly looked at the github repository. They appear to provide code to download their learned models from a google drive, but they have not provided code to generate your

own models from your own data. The training single TCR model section is empty with a "todo". (code availability)
We have updated the GitHub repository to include training code and documentation for building models from new datasets, and we now specify this in the Code Availability section.
Reviewer #4:
The authors combined high throughput yeast display library screening of HLA-B27-bound peptides recognized by selected TCRs, with fine-tuned protein language models to generate Peptide Recognition Profiles (PRPs) for these TCRs. The main finding is that TCRs cluster by functional rather than sequence similarity. Importantly, the PRPs predicted T cell activation for select peptides with greater accuracy than existing modeling approaches. In addition to the general technical and conceptual advance, this led to the discovery of a novel candidate autoantigen derived from PSG5 (GRLPLLNPI) for which there was greater autoreactivity in HLA-B27+ patients compared to HLA-B27+ control subjects. Taken together, the work is novel and represents a significant advance in discovering potential autoreactive self-peptides.

Despite the advances shown here, I have some questions pertaining to the biological application and generalizability of this methodology.

1. How were the 21 HLA-B27-restricted TCRs selected? In the results (line 92-95) the authors state that 'multiple HLA-B*27:05-restricted TCRs implicated in AS and AAU' were used. What is the evidence that these 21 TCRs are implicated in AS/AAU? Have all of these TCRs been shown to be expanded in HLA-B27+ individuals compared to HLA-B27+ healthy controls? Please clarify and cite the studies where appropriate.

In total, we examined 21 TCRs (Fig. 1a). Of these, 16 disease-derived TCRs from AS/AAU samples were used for all model training and evaluation (Figs. 1b and 2a). The remaining 5 TCRs are engineered variants of clonotype 19.2 (19.2C1–19.2C5), which were used exclusively for the mutagenesis validation (Fig. 5b) and were not included in model training or clustering. We have clarified this distinction in the Fig. 1a legend and Methods to avoid confusion.

For the 16 disease-derived TCRs, we now summarize their provenance in the Results and provide full details in the Methods. Briefly, TCR 4.1 was isolated from TRBV9$^+$ CD8$^+$ T cells expanded from the blood of an HLA-B27$^+$ AS patient; TCRs 8.4 and 9.1 from TRBV9$^+$ CD8$^+$ T cells in AS synovial fluid; TCR 019.1 from the aqueous humor of an HLA-B27$^+$ AAU patient; and the remaining receptors from YeiH–HLA-B*27:05 tetramer$^+$ CD8$^+$ T cells sorted from the blood of HLA-B27$^+$ AAU participants. Thus, all 16 TCRs used for modeling were obtained from AS or AAU disease samples.

With respect to disease linkage, prior work has shown that the BV9–Y/FSTDTQ–BJ2.3 motif is strongly enriched in AS synovial fluid relative to blood and clonally expanded in AAU ocular fluid (Yang et al., Nature 2022), and that YeiH-specific CD8$^+$ T cells are expanded in HLA-B27$^+$ AS/AAU patients compared with HLA-B27$^+$ controls, whereas influenza-specific responses are equivalent across groups (Paley & Yang et al., JCI Insight 2024). While not every individual clonotype in our 16-TCR panel was independently quantified against controls, these motif-level and antigen-specific enrichment data support that the modeled TCRs are representative of disease-associated populations.

Revision: We have updated the Methods in page 12 line #412-431 to explicitly list the provenance of all 16 training TCRs (AS synovial fluid, AAU ocular fluid, or Tetramer+ sort from patients). All are directly linked to the disease state.

2. Irrespective of the proposed paradigm using PRPs over sequence similarity, which seems to be warranted based on the results, the generalizability of these findings is limited by the focus on only CDR3β modelling since this ignores the alpha-chain and CDR1/2 regions. For HLA-B27 specifically, several TCRs are beta-centric, but alpha-chain and CDR1/2 contacts with the alpha-helices can still modulate recognition thresholds and explain why some peptides are seen and others aren't. Even in Fig. 4g, for example, there is a CDR1a-peptide contact. The fact that germline-encoded CDR1/2 loops could impact peptide contact is also noted in the previous manuscript (PMID 36477533). Since CDR1/2 often contact HLA helices and can influence docking geometry and peptide sensitivity, please justify the CDR3β-only assumption made in this model and manuscript. This could be addressed experimentally by an ablation experiment addressing whether alpha/beta/CDR1/2 features alter peptide rankings for a representative subset. If feasible, a mutational or alpha-chain swap (between two related TCRs) would strengthen the conclusion that the observed peptide preferences

are not contingent on germline-encoded contacts.

As we illustrate in Reviewer #2, Point #3, we analyzed crystal structures of TCRs within this family complexed with peptide-HLA-B*27:05. As shown in Point #3, peptide recognition is structurally dominated by the CDR3β loop, which makes extensive contacts with the peptide backbone and side chains. In contrast, the CDR3α loop is positioned away from the peptide-binding groove. Consequently, the observed variation in CDR3α sequences across this family likely reflects neutral drift rather than antigen-driven selection. This structural reality provides the biological justification for our β-centric modeling approach.
Our decision to model CDR3β alone was not based on the assumption that α- or germline contacts are irrelevant, but on structural and experimental evidence showing that, within this specific public clonotype family, peptide specificity is dominated by β-chain interactions (Yang et al., Nature 2022). Our new α-swap data (see response to Reviewer #3, Point #4) confirms that experimentally, the CDR3β is the primary variable driving differential peptide recognition in this system. To directly test α-chain effects, we performed α-swap experiments among clonotypes and observed only modest shifts in activation that did not alter the overall recognition patterns dictated by β identity (new Supplementary Fig. 6c). Including CDR3α in the model did not improve predictive performance (Supplementary Fig. 6a,b).

Revision: We have revised the Result in page 5 #131-139 describing the HLA-B*27:05 restricted TCR panel to explicitly explain the β-centric architecture and conserved CDR1α contacts, and have added a new structural panel (Fig. 2b) highlighting these features. Furthermore, we have revised the Discussion in page 11 #336-343 to note that applying this framework to other TCR systems would likely require incorporating α-chain features.

3. The authors should better justify and/or discuss the limitations of a peptide library where the HLA heavy chain has been mutated in positions that are likely to affect peptide binding, the P2 and P8 positions of the HLA-B27-bound peptide are fixed, all peptides are 9-mers, and the N- and C-termini of the peptide can't be in their normal orientation in the peptide binding groove since the single chain construct precludes this arrangement? This must have some structural consequences for the way peptide-MHC complex are seen by TCRs. Moreover, are the HLA-B27 constructs used in this study expressing bona fide HLA-B*27:05, or HLA-B27 with the Y84A, L5M, H114Y, and A153D mutations like the yeast display library? (Is Cys67 mutated as well?) If so, they should not be called HLA-B*27:05. Does the PSG5 peptide bind to HLA-B*27:05 or HLA-B27 with the four substitutions? If both, are there affinity differences?

The yeast display construct used in this study follows established approaches for stabilizing peptide-MHC complexes on the yeast surface, and therefore includes mutations (Y84A, L5M, H114Y, and A153D, with C67S to prevent aberrant disulfide formation) that improve folding and surface expression. These substitutions are a technical requirement of the yeast display system and are not unique to HLA-B*27:05, as similar strategies have been widely employed across multiple HLA alleles. We agree that such modifications may alter fine aspects of peptide binding and orientation, and we now explicitly acknowledge this limitation in the revised Discussion in page 12 line #390-397.

Importantly, all downstream validation was performed using wild-type HLA-B*27:05, not the mutated yeast construct. PSG5 recognition was confirmed with WT HLA-B27:05 tetramers, in functional activation assays, and by structural analysis of the PSG5–B27:05–TCR complex. These results demonstrate that PSG5 is stably bound and presented in the context of the native allele, and that recognition is not an artifact of the yeast-display mutations.

4. In Fig. 4f, the expansion of CD8+ T cells recognizing the flu peptide is nearly as strong (~3.8 fold) as for the PSG5 peptide (~4.9 fold) in the individual shown. In Supplemental Fig. 10c, the lack of statistical significance for expanded B27/flu-specific CD8s is most likely due to insufficient power with only 6 patients and 6 controls. Unless there are differences in influenza exposure/infection, this calls into question the specificity of the 'autoreactive' response to the PSG5 peptide. More data is needed to substantiate this claim.

We agree that comparison to influenza-specific responses is important. As shown in our previous work (Paley & Yang et al., JCI Insight 2024) and confirmed by new data added to Supplementary Fig. 10, Influenza-specific responses are equivalent between patients and controls. In contrast, PSG5 reactivity is specifically enriched in AS/AAU patients.

This differential enrichment supports the classification of PSG5 as a disease-associated autoantigen candidate, whereas Flu serves as a non-pathogenic control.

In particular, our prior study (Paley & Yang et al., JCI Insight 2024, Fig. 4C) demonstrated no significant difference in influenza epitope-specific CD8+ T cell frequencies between AS/AAU patients and healthy HLA-B27+ controls. To further illustrate this point, we provide here an expanded unpublished dataset (see figure below) in which influenza-specific CD8+ T cell responses were assessed across a larger cohort, again showing equivalent responses in patients and controls. Together, these data support the conclusion that influenza reactivity is a shared control response and not disease-associated, whereas PSG5 reactivity is specifically enriched in patients. We have also cited the JCI Insight study in the revised manuscript in line #247 to make this point clear to readers.

5. In the antigen presentation assays, peptide concentrations are 100 µM, which is very high, and could lead to non-specific binding. Please justify the use of this high concentration.

We used 100 µM peptide concentrations in our antigen presentation assays to ensure robust peptide loading onto HLA-B*27:05 in vitro, which is consistent with prior studies (Yang et al., Nature 2022). While this concentration is higher than the estimated physiological range, it is commonly used to compensate for variability in peptide stability, cell uptake, and presentation efficiency. Importantly, we observed clear peptide specific activation patterns across TCRs, and control peptides included in these assays did not elicit non-specific responses at the same concentration, indicating that activation was sequence dependent rather than due to non-specific binding.

6. Do the 21 TCRs studied here recognize other HLA-B alleles?
We did not test cross-reactivity to other HLA-B alleles in the present study. All TCRs were isolated from HLA-B*27:05-positive AS/AAU patients, and our experiments were designed around B*27:05. Investigating potential cross-reactivity across HLA-B alleles will therefore be an important direction for future study.

7. Although more of a comment than a question, it seems to me that these results challenge the molecular mimicry hypothesis in its simplest form; i.e. a foreign antigenic peptide derived from a microbe and presented by HLA-B27 leads to an immune response that cross-reacts with an HLA-B27 self-peptide because of sequence similarity between the foreign and self-peptides. This has been suggested to occur for HLA-B27 because it theoretically presents self-peptides that resemble certain microbial peptides because of its unique peptide binding specificity. While sequence-based molecular mimicry could still occur, it seems the situation is far more complicated, raising again the question of why HLA-B27 is so strongly linked to this disease.

We agree. Our results do not negate peptide-level mimicry but add a layer of complexity: TCRs with similar sequences can have divergent functional profiles.

Previous work (Yang et al., Nature 2022) has elegantly demonstrated sequence- and structure-based molecular mimicry between microbial and self-peptides bound by HLA-B*27, supporting the arthritogenic peptide hypothesis. Our findings are not intended to contradict this important evidence, but rather to expand the framework of mimicry. Specifically, our manuscript directly demonstrates that highly similar TCR CDR3β sequences can diverge in their peptide recognition profiles, indicating that TCR sequence similarity does not necessarily predict functional similarity. In contrast, peptides are structurally simpler, and the present study does not directly test peptide sequence versus function. Thus, our conclusions should not be interpreted as negating molecular mimicry at the peptide level, but instead highlight that TCR cross-reactivity cannot always be inferred from sequence alone. We have revised the Discussion page 11 line #375-380 to clarify this point.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Most of my comments have been addressed, although a few remaining points need clarification.

1) The authors mention TRBV5. However, there are multiple TRBV5-x genes, with some very important differences in CDR1/CDR2 sequences. Would be important to clarify which one was used there.
2) I appreciate the improved explanations for the design of the library, and the fact that P2 and P8 were fixed. However, I think this could be made even more clear by explicitly showing it in the Figures. For instance, having a little red star in each motif below or above positions that were fixed in the peptide libraries would really help a lot. This would also clarify that K/R specificity at P9 for some motifs results from a different design of the libraries, and not some complex allosteric mechanism. I presume that results from this study may be used by many people in slides, and not showing this library design information in the Figure could lead to many misunderstandings or confusion.
3) Wrt to my comment 3), the authors have clarified nicely the scope of their work. However, I do think it would add a lot to show a PRP for a distant TCR, as I had requested. This could be done in Figure 6a, based on the conceptual schematic of TCR and peptide recognition space, where a PRP could be displayed for one the “Screened TCRs” (e.g. 135.1), a quite similar PRP for a “nearby” TCRs (e.g., 4.1, with some subtle differences examined in this work) and then a completely different PRP for a distant TCR (e.g. NY-ESO-1 specific TCRs where PRPs are publicly available). This would help a lot providing the broader perspective and highlighting the focus of this study. I would also encourage to make a smaller circle for the screened TCRs, since they correspond to a very tiny fraction of the TCR space.
Finally, I bring to the attention of the authors a similar ‘mirror image’ model proposed in PMID: 38615042 (Figure 3E) for how TCR motifs are evolving across epitope space – but as this is a paper from my lab, I leave it up to the authors’ choice to discuss this or not.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
The authors have addressed most of my points thoroughly. My one remaining concern comes from the alpha-swap experiments and their discussion. I appreciate the addition of new data, and the lack of gain in predictive capacity when including the alpha chain sequence in the modeling framework. I think some clarification for the reader and a re-framing of the b-domination claims are still warranted.

In their response, the authors say “Importantly, these results should not be interpreted to mean that α-chains are generally irrelevant for TCR specificity. Rather, they underscore that within this clonotypically constrained, β-dominated family, peptide specificity can be predicted from β-chain information with high fidelity.” This is only the case though when the alpha chain is one selected (potentially from a relatively large group) specifically to maintain specificity. The alpha chain may not make direct contact with the peptide, but it is still extraordinarily constrained, so you can’t predict peptide specificity from b-chain information alone. This needs to be made clear. Something like “Conditional on a suitable alpha chain that permits the proper orientation of the beta chain to the peptide MHC”. It is likely a relatively small portion of alpha chains that are indeed suitable—even if we say “all TRAJ28 TCRs will work “ (which is very unlikely) that’s still about 2.5% of the repertoire. From the data provided we can’t really say what the number of complementary alphas is. In all likelihood, this means that the vast majority of TCRs containing exactly the TCRbetas profiles here will not be specific for these antigens.
One way of demonstrating this would be to do some alpha swaps with more distant TCRa chains and show that specificity is indeed broken (or not! I might be wrong, but the authors thought they needed to keep the TRAJ28). Alternatively just clarifying the language that this beta prediction works only in the presence of a relatively small subset of alphas that they haven’t defined here would be helpful. Otherwise, the way the broader field will interpret this paper is “as long as a TCR has this beta it is autoreactive and has precisely this response profile” and that should not be the conclusion.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
The authors have provided a robust response to reviewers' critiques including additional experiments and significant changes to the manuscript. The manuscript now presents the data, the authors' interpretation, and conclusions in a much more appropriate context given the limitations of the model they are studying. I have one more minor point that the results shown in supplemental figure 6c (CD69 activation with various alpha chain swaps) should show the values (with statistical variation) rather than just a heat map which essentially can only be read as 'off' or 'on'. How are negatives and positivies differentiated? The figure legend should also contain more details.

(Remarks on code availability)

Author Rebuttal letter:

Point-to-point responses:

Reviewer #1:
Most of my comments have been addressed, although a few remaining points need clarification.

1) The authors mention TRBV5. However, there are multiple TRBV5-x genes, with some very important differences in CDR1/CDR2 sequences. Would be important to clarify which one was used there.
We agree that the TRBV5 designation encompasses multiple germline genes with distinct CDR1 and CDR2 sequences. We have now explicitly specified the exact TRBV5 gene usage for the relevant TCRs in the manuscript Fig. 2a and Methods (page 13, line 436), to avoid ambiguity and ensure clarity for readers.

2) I appreciate the improved explanations for the design of the library, and the fact that P2 and P8 were fixed. However, I think this could be made even more clear by explicitly showing it in the Figures. For instance, having a little red star in each motif below or above positions that were fixed in the peptide libraries would really help a lot. This would also clarify that K/R specificity at P9 for some motifs results from a different design of the libraries, and not some complex allosteric mechanism. I presume that results from this study may be used by many people in slides, and not showing this library design information in the Figure could lead to many misunderstandings or confusion.
We have updated Fig. 3c, Fig. 6g, and Supplementary Fig. 2 to explicitly annotate the fixed anchor positions (P2 and P8, or P2 and P9 where applicable) directly on the peptide motifs by little red star, with corresponding legend updates. This clarifies the library design and avoids misinterpretation of motif features as intrinsic biological effects.

3) Wrt to my comment 3), the authors have clarified nicely the scope of their work. However, I do think it would add a lot to show a PRP for a distant TCR, as I had requested. This could be done in Figure 6a, based on the conceptual schematic of TCR and peptide recognition space, where a PRP could be displayed for one the “Screened TCRs” (e.g. 135.1), a quite similar PRP for a “nearby” TCRs (e.g., 4.1, with some subtle differences examined in this work) and then a completely different PRP for a distant TCR (e.g. NY-ESO-1 specific TCRs where PRPs are publicly available). This would help a lot providing the broader perspective and highlighting the focus of this study. I would also encourage to make a smaller circle for the screened TCRs, since they correspond to a very tiny fraction of the TCR space.
To address this point, we updated Fig. 6a to include a representative peptide recognition profile (PRP) from a functionally distant TCR (Gee et al., 2018), shown alongside PRPs for screened and nearby TCRs. We also reduced the visual footprint of the screened TCRs to reflect their limited coverage of the overall TCR space. Together, these changes highlight that the present study interrogates a narrow local neighborhood within TCR recognition space, rather than the global TCR repertoire.

4) Finally, I bring to the attention of the authors a similar ‘mirror image’ model proposed in PMID: 38615042 (Figure 3E) for how TCR motifs are evolving across epitope space – but as this is a paper from my lab, I leave it up to the authors’ choice to discuss this or not.
We thank the reviewer for highlighting this conceptual convergence. The intuition that TCR-peptide relationships can be understood through mapping of recognition landscapes aligns well with our framework. We have added a brief acknowledgment in the Discussion (page 10, line 337): "Our framework shares conceptual parallels with prior models describing how TCR motifs vary systematically across epitope space(27), reinforcing the broader machine learning intuition that TCR-peptide relationships can be productively viewed through mapping of recognition landscapes."

Reviewer #2 :
The revised manuscript addressed most of the comments raised before.

1) However, the alpha chain swaps were still severely constrained. The broader point would be to include an alpha chain that 1) maintained specificity to some epitopes but did not use the major motif features (i.e. not using the same J region) and 2) showing that most alpha chains would actually break specificity. So in a fundamental sense, the alpha can only be excluded from the model because it is assumed to be from an extremely narrow portion of the repertoire.

We agree that α-chain variation is biologically important and that our β-centric predictions are conditional on compatible α-chain pairing. The α-chain is not excluded because it is unimportant, but because the TCRs analyzed represent a narrowly selected, disease-associated family in which α-β pairing has already been strongly constrained by biological selection.

We further note that the TCRα chain comprises TRAV, CDR3α, and TRAJ segments, and both TRAV and TRAJ are known to play important roles in α-β pairing and epitope specificity. Altering TRAJ usage (e.g., moving away from TRAJ28) would be expected to impact receptor assembly and peptide recognition. As α and β chains are co-selected as structurally compatible pairs, extensive α-chain swapping beyond closely related, compatibility-preserving contexts would primarily reflect pairing incompatibility rather than informative specificity rules. Systematically mapping the space of compatible α-β combinations would require a fundamentally different experimental design and is beyond the scope of the present study.

We have clarified these assumptions in the Results (page 6, line 184) and revised the Discussion to make these constraints explicit (page 10, line 343): "However, these findings should not be interpreted to mean that α-chains are dispensable for TCR specificity more broadly. Our β-chain-based predictions hold only conditional on the presence of a structurally compatible α-chain that permits productive pMHC engagement." The limitations section further notes (page 12, line 393): "Additionally, our β-centric modeling applies specifically to this clonotypically constrained family in which α-β pairing has already been subject to strong biological selection; generalization to diverse α-β pairings will require expanded profiling."

Reviewer #3:
The authors have addressed most of my points thoroughly.

1) My one remaining concern comes from the alpha-swap experiments and their discussion. I appreciate the addition of new data, and the lack of gain in predictive capacity when including the alpha chain sequence in the modeling framework. I think some clarification for the reader and a re-framing of the b-domination claims are still warranted.

In their response, the authors say “Importantly, these results should not be interpreted to mean that α-chains are generally irrelevant for TCR specificity. Rather, they underscore that within this clonotypically constrained, β-dominated family, peptide specificity can be predicted from β-chain information with high fidelity.” This is only the case though when the alpha chain is one selected (potentially from a relatively large group) specifically to maintain specificity. The alpha chain may not make direct contact with the peptide, but it is still extraordinarily constrained, so you can’t predict peptide specificity from b-chain information alone. This needs to be made clear. Something like “Conditional on a suitable alpha chain that permits the proper orientation of the beta chain to the peptide MHC”. It is likely a relatively small portion of alpha chains that are indeed suitable—even if we say “all TRAJ28 TCRs will work “ (which is very unlikely) that’s still about 2.5% of the repertoire. From the data provided we can’t really say what the number of complementary alphas is. In all likelihood, this means that the vast majority of TCRs containing exactly the TCRbetas profiles here will not be specific for these antigens.

One way of demonstrating this would be to do some alpha swaps with more distant TCRa chains and show that specificity is indeed broken (or not! I might be wrong, but the authors thought they needed to keep the TRAJ28). Alternatively, just clarifying the language that this beta prediction works only in the presence of a relatively small subset of alphas that they haven’t defined here would be helpful. Otherwise, the way the broader field will interpret this paper is “as long as a TCR has this beta it is autoreactive and has precisely this response profile” and that should not be the conclusion.

We thank the reviewer for this important clarification and fully agree that our β-dominance claims must be interpreted conditionally. As noted in our response to Reviewer 2, β-chain-based prediction is not intended to imply that α-chains are generally dispensable. Rather, our conclusions apply within a clonotypically constrained system in which α-β pairing has already been subject to strong biological selection.

Regarding the suggested distant α-swap experiments: α and β chains are co-selected as structurally compatible pairs during thymic selection, and productive TCR assembly requires compatible germline pairing, proper folding, and appropriate docking geometry

onto peptide-MHC. Swapping in distant or motif-incompatible α-chains will likely result in TCRs that fail to express, fold, or engage pMHC productively. If such chimeric receptors showed loss of specificity, we could not distinguish whether this reflected a genuine change in peptide recognition rules or simply structural incompatibility at the level of pairing or docking. Conversely, if by chance a distant α-chain happened to preserve function, it would not change our core conclusion that β-chain features dominate peptide specificity within this constrained family. For these reasons, we focused on compatibility-preserving swaps, which isolate the contribution of α-chain variation to fine-tuning within a functional receptor architecture.

We have revised the Results (page 6, line 184) and Discussion (page 10, line 343; page 12, line 393) to explicitly state that β-chain-based prediction holds only conditional on the presence of a suitable α-chain, and that compatible α-chains likely represent a restricted subset of the repertoire.

Reviewer #4:
The authors have provided a robust response to reviewers' critiques including additional experiments and significant changes to the manuscript. The manuscript now presents the data, the authors' interpretation, and conclusions in a much more appropriate context given the limitations of the model they are studying.

1) I have one more minor point that the results shown in supplemental figure 6c (CD69 activation with various alpha chain swaps) should show the values (with statistical variation) rather than just a heat map which essentially can only be read as 'off' or 'on'. How are negatives and positives differentiated? The figure legend should also contain more details.

We have revised Supplementary Fig. 6c to display the underlying quantitative values (mean of three independent experiments) and expanded the figure legend to clarify how positive responses are defined.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
My comments have been addressed

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
all concerns have been addressed--no new comments.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
I'm satisifed that the authors have addressed the reviewers' concerns adeuqately.

(Remarks on code availability)

Author Rebuttal letter:

Response to Reviewers

We thank the reviewers for their thoughtful evaluation of our manuscript and for their constructive feedback during the revision process. We are pleased that all reviewers are satisfied with the current version of the manuscript.

Reviewer #1:

Remarks to the Author: My comments have been addressed
Response: We thank the reviewer for their time and for the positive feedback.

Reviewer #3:

Remarks to the Author: all concerns have been addressed--no new comments.
Response: We appreciate the reviewer's thorough assessment and are glad that all concerns have been resolved.

Reviewer #4:

Remarks to the Author: I'm satisfied that the authors have addressed the reviewers' concerns adequately.
Response: We thank the reviewer for their positive comments and are glad that all concerns have been addressed.