

1 Comprehensive epitope mutational scan database 2 enables accurate T cell receptor cross-reactivity 3 prediction

4 Amitava Banerjee , David J Pattinson , Cornelia L. Wincek , Paul Bunk , Armend
5 Axhemi , Sarah R. Chapin , Saket Navlakha , and Hannah V. Meyer

6 ¹Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

7 ²Department of Pathobiological Sciences, School of Veterinary Medicine, University of Wisconsin-Madison,
8 Madison, WI 53711, USA

9 ³Medical Research Center and Clinic for Medical Oncology and Hematology, Cantonal Hospital St. Gallen,
10 9007 St. Gallen, Switzerland

11 ⁴School of Biological Sciences, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

12 ⁵W.M. Keck Structural Biology Laboratory, Howard Hughes Medical Institute, New York, NY, USA

13 ⁶Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

14 To whom correspondence should be addressed. E-mail: hmeyer@cshl.edu, navlakha@cshl.edu

15 ABSTRACT

Predicting T cell receptor (TCR) activation is challenging due to the lack of both unbiased benchmarking datasets and computational methods that are sensitive to small mutations to a peptide. To address these challenges, we curated a comprehensive database, called BATCAVE, encompassing complete single amino acid mutational assays of more than 22,000 TCR-peptide pairs, centered around 25 immunogenic human and mouse epitopes, across both major histocompatibility complex classes, against 151 TCRs. We then present an interpretable Bayesian model, called BATMAN, that can predict the set of peptides that activates a TCR. We also developed an active learning version of BATMAN, which can efficiently learn the binding profile of a novel TCR by selecting an informative yet small number of peptides to assay. When validated on our database, BATMAN outperforms existing methods and reveals important biochemical predictors of TCR-peptide interactions. Finally, we demonstrate the broad applicability of BATMAN, including for predicting off-target effects for TCR-based therapies and polyclonal T cell responses.

17 Introduction

A single T cell receptor (TCR) can recognize a variety of peptides, a property known as TCR cross-reactivity [1, 2]. Predicting which peptides a TCR cross-reacts to is critical for numerous applications, including predicting viral escape [3], cancer neoantigen immunogenicity [4], autoimmunity [2, 5], and off-target toxicity of T cell-based therapies [6, 7]. However, predicting interactions among TCRs, peptides, and major histocompatibility complexes (TCR-pMHCs) remains challenging [8–11] due to: (a) limited TCR cross-reactivity assay data, (b) few experimentally validated negative examples [12], which are important for model discrimination (Figure 1A), and (c) limited number of available ground-truth TCR-pMHC structures [13]. Most existing computational methods are designed to cluster different TCRs that bind the same peptide [8, 14]. But the opposite task — predicting peptides that bind a given TCR — remains outstanding [9, 11, 15, 16]. This is largely due to the sensitivity required to discriminate among single amino acid (AA) mutants [11, 17] of a TCR’s known index peptide, i.e., the peptide to which the TCR was identified to strongly bind. To address this challenge, we offer both a comprehensive experimental mutational scan database of TCR-pMHC binding, and a method that can predict how peptide mutations affect TCR activation. Additionally, we also developed an active learning method to minimize the number of TCR-pMHC experiments that need to be performed to learn the peptide cross-reactivity of a novel TCR, thereby reducing costs associated with performing TCR-pMHC assays.

32 BATCAVE: a database of TCR-specific peptide mutational scans reveals TCR-pMHC 33 interaction rules

34 To address the lack of TCR cross-reactivity training sets with balanced positive and negative experimentally-validated TCR-
35 pMHC pairs, we curated a database of continuous-valued TCR activation data by pMHCs from mutational scan assays (hereafter
36 referred to as TCR activation data; [Figure 1B,C](#)). This database, termed BATCAVE (“Benchmark for Activation of T-cells
37 with Cross-reactive Avidity for Epitopes), includes (1) 35 fully-sequenced CD8⁺ and 43 (7 fully-sequenced) CD4⁺ mouse
38 TCR clones, and (2) 69 (59 fully-sequenced) CD8⁺ and 4 fully-sequenced CD4⁺ human TCR clones ([Extended Data Fig 1](#)).
39 Together, TCRs in the BATCAVE recognize a total of 25 unique index peptides that are of length 8 to 20 ($L \in [8, 20]$) AAs
40 presented by 11 class-I MHCs and 4 class-II MHCs molecules, and which are involved in cancer, viral infection or autoimmunity.
41 For 101 of 151 TCRs, BATCAVE records the raw activation of all possible $L \times 19$ single-AA mutant peptides, and for the rest
42 of TCRs, the corresponding data for all mutant peptides with mutations only in non-MHC-anchor positions ([Figure 1C](#)). To
43 provide a consistent activation scale across diverse TCR-pMHC avidities and experimentally measured variables (e.g., IFN γ
44 concentration and NFAT-GFP fluorescence intensity), we normalized the raw activation data to the maximum activation for
45 each TCR. In addition to the TCR activation data, we collected matched TCR-pMHC dissociation constants of all single-AA
46 mutant peptides for 3 MHC-I-restricted TCRs (868Z11, c259, TIL1383I), and 5 MHC-II-restricted TCRs (all specific for the
47 H2-IAb-bound ‘3K’ antigen FEAQKAKANKAVD). Overall, we achieved (1) full coverage of the single-AA mutant antigenic
48 space for over 100 TCRs, and (2) balanced numbers of high-confidence, true positive (TCR activated) and true negative (TCR
49 inactive) data for each TCR, both unprecedented among existing datasets ([Figure 1A](#)).

50 BATCAVE showed that single AA changes of the index peptide can result in both loss and gain of TCR activation over
51 orders of magnitude ([Figure 1C, Extended Data Fig 2](#)), and it recorded how different mutants of the same index peptide activate
52 different TCRs [18–20], allowing us to distill broad, interpretable biochemical and cellular features of TCR-pMHC interactions.
53 First, we analyzed how pMHC binding affects TCR activation. Since different MHC alleles have different peptide-binding
54 anchor residues, we first focused on the most prevalent MHC allele in our database, HLA-A*02:01 (43 TCRs, 11 index peptides).
55 We selected the 41 TCRs recognizing nine distinct 9-AA-long index peptides with all peptide positions mutagenized, and we
56 used pMHC binding level prediction from NetMHCPan4.0 [21] to classify pMHC interaction as ‘no’ [> 2% rank], ‘weak’
57 [0.5–2% rank], or ‘strong’ [< 0.5% rank] binders. For ease of interpretation, we also discretized continuous TCR activation
58 values into three levels ([Figure 1C](#)): ‘strong’ [$\geq 50\%$ of maximum], ‘weak’ [10–50%] or ‘no’ [<10%] activation. We found
59 that the fraction of TCR strong-activating mutants increased from 0.18 for MHC non-binders, to 0.23 (weak-binders) and 0.29
60 (strong-binders) with predicted MHC binding. However, mutant peptides were distributed over all $3 \times 3 = 9$ possible paired
61 (MHC-binding level \times TCR activation level) categories ([Figure 1D](#)), with similar observations for other MHCIs ([Extended
62 Data Fig 3](#)). This suggests that MHC binding alone cannot be used to predict TCR activation and demonstrates the need for
63 TCR activation and peptide immunogenicity prediction models beyond pMHC affinity prediction.

64 Second, we utilized BATCAVE to deduce the relationship between TCR-pMHC binding (measured as K_d) and TCR
65 activation. This is a classic question in immunology with many features contributing to the relationship between binding and
66 activation, such as pMHC concentration [22], TCR structural changes [23, 24], catch bonds [25, 26], and co-receptors [27–29].
67 We found a general positive association between TCR-pMHC binding and TCR activation across both MHC I and MHC II TCRs
68 ([Figure 1E](#)). However, there are exceptions; for example, for TCR 868Z11, which binds the disulfide-stabilized HLA-A02:01
69 molecule [30], we found that strong TCR-pMHC binding of many mutants is not sufficient for TCR activation.

70 Third, we asked how TCR cross-reactivity depends on the wildtype (WT) AAs of the index peptide. Since previous studies
71 have associated hydrophobic peptide residues with increased immunogenicity [31–33], we asked whether WT AA residue
72 hydrophobicity dictates the hydrophobicity of allowed mutations. To quantify this, we calculated the difference between
73 average hydrophobicity (as measured by the Wimley-White interfacial hydrophobicity scale [34]) of TCR non-activating and
74 strong-activating mutated AAs for each WT AA across all TCRs and peptide positions ([Figure 1F](#)). Positive differences in
75 average hydrophobicity between strong-activating and non-activating mutated AAs were more frequent when WT AAs were
76 hydrophobic (89%) than for any other class of WT AA (36%), with the mean differences in hydrophobic WT AA and all other
77 AAs being 0.18 and 0.00, respectively. This indicated that peptide positions with non-hydrophobic WT AAs do not have any
78 preference for the mutant AA for TCR activation, while hydrophobic WT AA residues preferentially allowed hydrophobic
79 mutant AAs, likely since the latter are crucial TCR binding residues.

80 Finally, we studied how the TCR CDR3 chain hydrophobicity dictates the hydrophobicity and number of allowed mutations.
81 For TCRs with available $CDR3\alpha$ and $CDR3\beta$ sequences, we again calculated the difference between average hydrophobicity
82 of TCR non-activating and TCR-activating mutated peptide AAs. Next, we computed the hydrophobicity of the $CDR3\alpha\beta$
83 chains by summing across the hydrophobicity of their individual AA residues. We found that hydrophobic (hydrophilic)
84 $CDR3\beta$ chains prefer hydrophobic (hydrophilic) mutant AAs for activation ([Figure 1G](#)), confirming a trend previously seen for
85 TCR-pMHC structural analyses [32, 35, 36].

86 Overall, BATCAVE provides the largest to-date collection of TCR-pMHC interaction data across multiple MHC classes,

87 index peptides, and TCRs and reveals broad as well as TCR-specific insights into properties of TCR-pMHC binding. Together,
88 this database serves as a strong foundation for benchmarking TCR-pMHC prediction methods.

89 Existing TCR-pMHC models poorly predict TCR activation by mutant peptides

90 We first asked how well existing TCR-pMHC models predict TCR activation by mutant peptides, a question we could answer
91 using BATCAVE as a benchmarking dataset. Since many index peptides appear multiple times for different TCRs in BATCAVE,
92 we selected a subset of BATCAVE with each index peptide appearing only once. This selection yielded a diverse and balanced
93 dataset, spanning 18 MHCI-restricted human TCRs specific for unique 9- and 10- AA-long index peptides and 4 MHCII-
94 restricted TCRs specific for index peptides of various lengths (Figure 2A). For TCRs with the same index peptide, we included
95 the one with the least class imbalance.

96 We benchmarked 15 pre-trained TCR-pMHC models with different machine learning architectures and input data requirements
97 (Figure 2A), selected based on the availability of web-servers or source code. These models are trained on large publicly
98 available databases (e.g., VDJdb [37], McPAS-TCR [38], and single-cell immune repertoire profiling data [39]). While these
99 methods are designed for either classifying TCRs binding to selected immunogenic peptides [8, 14, 40] or predicting peptide
100 immunogenicity [32, 41], our benchmarking involved a distinct problem of classifying which set of peptides activate a given
101 TCR. Similar to previous results on smaller and less diverse datasets [11, 17], these models predicted only marginally better
102 than random on our benchmarking dataset, with the best classification AUCs being from TITAN [42] (mean AUC=0.56 for
103 MHCI-restricted and 0.52 for MHCII-restricted peptides) and ImRex [9] (mean AUC=0.55 for MHCI-restricted peptides).
104 General immunogenicity prediction tools, which do not use TCR sequence information for prediction, overall performed
105 worse, with none included among the top 6 performers. Overall, predicted TCR-pMHC interaction scores from most tested
106 TCR-pMHC models were uncorrelated with true TCR activation values for the mutant peptides (Extended Data Fig 4). Thus,
107 these results demanded development of new TCR-pMHC models for predicting how peptide mutations affect TCR activation.

108 BATMAN: a Bayesian inference model provides state-of-the-art prediction of TCR activation 109 by mutant peptides

110 We present BATMAN — "Bayesian Inference of Activation of TCR by Mutant Antigens" — a hierarchical Bayesian model
111 that can predict TCR activation by single-AA mutant peptides based on their distances to the TCR's index peptide. Our
112 peptide-to-index distance is a product of (a) a learned positional weight profile, specific for individual TCRs, corresponding to
113 the weighted effects of mutated residues at different positions in the sequence, (b) a learned AA substitution distance matrix
114 from the index AA to the mutant AA, and (c) an optional scalar weight, also specific for individual TCRs, corresponding to the
115 pMHC binding category (Figure 2B, Methods). BATMAN can be used for classification and continuous regression tasks for
116 both TCR-specific and cross-TCR activation datasets.

117 We validated BATMAN over our benchmarking dataset in two modes (Figure 2C):

- 118 1. *Within-TCR* mode, where we performed 5-fold cross-validation for each TCR, with the folds stratified by TCR activation
119 category and mutation position. The inferred positional weight profiles and pMHC binding level weights were TCR-
120 specific, while the inferred AA matrix was TCR independent. Even though we inferred TCR-specific parameters, the
121 hierarchical Bayesian framework of BATMAN allowed pooling of training data across TCRs to infer TCR-independent
122 model hyperparameters and the AA matrix (see Methods).
- 123 2. *Leave-one-TCR-out (LOO-TCR)* mode, where we trained on data from all TCRs but one, and tested on the held-out TCR.
124 The inferred positional weight profiles, pMHC binding level weights, and inferred AA matrix were all TCR-independent,
125 and learned by training across all TCRs except the held-out one (see Methods).

126 Since all of the pretrained TCR-pMHC models (Figure 2A) performed close to random in our benchmarking study, we
127 compared BATMAN's performance to that of *pTEAM* [17], an existing method designed to predict the effect of peptide mutation
128 effects on TCR activation. For a fair comparison, we re-trained *pTEAM* using the same data as BATMAN.

129 BATMAN outperformed *pTEAM* in both within-TCR (mean AUC=0.836 versus *pTEAM* AUC=0.795) and LOO-TCR
130 (mean AUC=0.687 versus *pTEAM* AUC=0.613) classification (Figure 2D). Critical to achieving BATMAN's performance
131 was learning the AA distance matrices by pooling training data across TCRs. For example, applying BATMAN with the
132 BLOSUM100 AA matrix, the best-performing conventionally-used distance matrix (Extended Data Figs 5 to 7), dropped the
133 within-TCR AUC to 0.811 (Figure 2D). Extended Data Figs 5 to 7 further highlight the superior performance of BATMAN
134 over 68 conventional AA substitution distance matrices when tested on all 151 TCRs, both in classification and continuous
135 regression tasks. BATMAN performance also slightly improved with the addition of pMHC binding level information, with the
136 within-TCR AUC dropping to 0.833 without this input (Figure 2D). *pTEAM* can also incorporate TCR sequence information

137 into its model, but we found no gain in performance with this additional information ([Figure 2D](#)), showing its limitations when
138 the training data is sufficiently diverse.

139 How do BATMAN and *pTEAM* compare when training data becomes limited? We calculated the within-TCR AUCs from
140 both models over random subsamples of previous training folds, maintaining stratification over TCR activation classes and
141 mutation positions. In each case, we also trained within-TCR BATMAN in two ways: (1) pooled across all TCRs like before,
142 and, to investigate effects of pooling across TCRs and the overall training data size, (2) trained and tested individually for each
143 TCR. In this limited-data regime, BATMAN pooled across TCRs consistently performs better than *pTEAM*, with BATMAN
144 achieving close to its 5-fold cross-validation performance with only a random 40% subsample of training data in each fold (7 per
145 mutation position per TCR) ([Figure 2E](#)). Pooled BATMAN performed better than unpooled BATMAN, for which MHCI AUC
146 dropped to 0.775 for the above 40% subsampling, though the difference narrowed with more training data (MHCI AUC=0.830
147 and 0.817 respectively for unpooled and pooled with full within-TCR data; see also [Extended Data Fig 7](#)). Unpooled BATMAN
148 also performed comparably to *pTEAM* in the very low training data regime ([Figure 2E](#)). Overall, these results highlight the
149 unique ability of BATMAN to improve TCR-pMHC interaction prediction by pooling across TCRs compared to *pTEAM*, which
150 must be trained separately for every TCR.

151 Finally, we compared BATMAN and *pTEAM* for LOO-TCR tasks where the training data is restricted to TCRs with the
152 same index peptide. For this benchmark, we selected 8 MHCI and 2 MHCII index peptides for which mutational scan data
153 against at least 3 TCRs were found in BATCAVE. BATMAN performed better than *pTEAM* for 9 out of 10 index peptides
154 tested ([Figure 2F](#)), with their performances becoming comparable only for index peptides with a large number of TCRs tested
155 (e.g., TPQDLNTML (n=8), NLVPMVATV (n=25) and SIINFEKL (n=16)). These benchmarking results collectively establish
156 BATMAN as the state-of-the-art predictor of TCR activation by mutant peptides across diverse TCR-pMHC training data
157 contexts.

158 Inferred BATMAN parameters capture interpretable features of TCR-pMHC interactions

159 BATMAN inferred positional weights were consistent across different AA matrices ([Extended Data Fig 8](#)), indicating that they
160 might correspond to key TCR-pMHC interaction features. Thus, we asked what the inferred positional weight profiles and AA
161 substitution matrices of BATMAN can reveal about TCR-pMHC interaction rules.

162 First, we inferred a pan-TCR positional weight profile and AA matrix by training BATMAN over all TCRs binding to
163 9-AA-long peptides in our benchmarking dataset of [Figure 2A](#). The learned pan-TCR positional weights ([Figure 3A](#)) peaked
164 near the middle of the peptide chain, reflecting the fact that central AA residues more directly affect TCR binding compared to
165 flanking MHC-anchor residues [19, 41, 43–46] (see also [Extended Data Fig 2](#)).

166 Second, we pooled across TCRs but inferred TCR-specific weight profiles, which showed that while central residues are
167 indeed often highly weighted, there can be significant variability of binding motifs for individual TCRs ([Figure 3A](#), [Extended](#)
168 [Data Fig 8](#)), suggesting that TCR cross-reactivity constitutes a diverse spectrum. Some TCRs (e.g., a3a, T1) are activated when
169 very specific AAs exist at certain positions (e.g., P2-P9 for T1, P1-P5 for a3a), and by almost all AAs at other positions (P1 and
170 P6-P8, respectively). These TCRs have almost binary positional weight profiles and resulted in better within-TCR classification
171 performance ([Figure 2D](#)). Similarly, positional weights could reveal how different TCRs that bind the same index peptide
172 use different binding motifs. We trained BATMAN by pooling over all 25 BATCAVE TCRs specific for the NLV peptide
173 and indeed found heterogeneity in positional weights across TCRs ([Figure 3B](#)). For example, the M5 residue is important for
174 binding to most TCRs, while N1 and A7 residues show high variation in their importance across TCRs.

175 Third, we asked whether BATMAN can identify the exact TCR epitopes from mutational scans of longer MHCII-bound
176 peptides, which would be important for T cell epitope discovery from peptidome-wide tiling scans (e.g. [47]). To test this,
177 we investigated BATMAN weights for the CD4⁺ TCRs TCR-F5 and TCR-3598-2 in the TScan-II dataset [47] performing
178 mutational scans across all positions in 20-AA-long peptides. The inferred positional weights ([Figure 3A](#)) peaked at peptide
179 positions included in the experimentally-verified central TCR-binding epitopes, FRDYVDRFYKTLRAEQASQE [48] and
180 LPVPGVLLKEFTVSGNILTI [47]. This suggests BATMAN weights can reveal the exact TCR epitopes from longer peptide
181 tiles with many TCR-irrelevant positions mutagenized.

182 Fourth, to study the correspondence between BATMAN positional weights and TCR-pMHC 3D structures quantitatively,
183 we assessed the interactions between peptide residues and TCR chains for 6 MHCI and 2 MHCII TCRs in BATCAVE with
184 available ground-truth crystallographic structures of their corresponding TCR-pMHC complexes in the Protein Data Bank
185 (PDB). BATMAN positional weights correlated strongly (r^2 between 0.61-0.81) with the number of interactions between
186 peptide residues and TCR chains ([Figure 3C](#)). For example, for the highly cross-reactive [2] autoimmune TCR 1E6 ([49], PDB
187 ID: 3UTT) binding the insulin peptide ALWGPDPAAA, positional weights are large for the solvent exposed GPD bulge of the
188 peptide, over which the TCR binds [49]. Similar correspondence between large inferred positional weights and important TCR
189 recognition motifs on peptides, characterized by a large number of TCR interactions, were seen for the other TCRs ([Figure 3C](#)).

190 This suggests that BATMAN weights encapsulate key structural features of TCR-pMHC complexes that affect TCR-peptide
191 binding.

192 Fifth, we investigated additional features of TCR-pMHC interactions revealed by the inferred positional weight profiles.
193 As a test case, we used a dataset spanning a wide range of cross-reactivity, consisting of both low-avidity ‘naive’ (broadly
194 cross-reactive) and antigen-expanded ‘educated’ (specific) TCRs binding to the SIINFEKL peptide [18]. We trained BATMAN
195 by pooling over all TCRs and calculated the width of the posterior distribution of the inferred positional weights for each
196 peptide residue which contains 90% of the density. To eliminate differences among TCRs due to their different pMHC binding
197 requirements, we focused only on the non-MHC-anchor positions (P1, P4, P6, P7) of the SIINFEKL peptide. We found that
198 naive TCRs with broader cross-reactivity profile have broader posterior distributions of inferred weights (Figure 3D), suggesting
199 that the weight profiles capture essential features of cross-reactivity.

200 Finally, the BATMAN AA distance matrix (Figure 3E) revealed three biochemical features of TCR-pMHC interactions.
201 First, large changes in TCR activation correspond to non-aromatic to aromatic AA substitutions (e.g., valine and glycine to
202 phenylalanine; methionine to tryptophan and tyrosine) affecting side-chain interactions [31, 32, 41]. Second, swapping in
203 hydrophobic isoleucine and leucine residues for non-hydrophobic residues overall increases TCR activation, in line with these
204 residues considered to increase immunogenicity [31–33] and our earlier observation in BATCAVE data (Figure 1F) that strong-
205 activating mutant AA residues are more hydrophobic than non-activating residues for highly hydrophobic WT AAs. Third,
206 while previous works have reported both gain [50] and loss [51] of TCR activation with peptide cysteine residue substitutions,
207 we found an increase in TCR activation for valine, glutamine and serine to cysteine substitution while there was a decrease in
208 activation for phenylalanine and tyrosine to cysteine substitutions (Figure 3E), reflecting the high context-dependence of such
209 substitutions.

210 Overall, the interpretable parameters learned by BATMAN capture a host of biological features that reveal the nature of
211 TCR-pMHC interactions.

212 Active learning with BATMAN boosts performance for novel TCRs

213 Designing cross-reactivity assays for a novel TCR is challenging because the number of possible peptides to test increases
214 exponentially with the number of mutations. Can we identify a small set of informative peptides within this space, whose
215 binding information can be used to predict TCR activation to the many other peptides not assayed?

216 We developed a BATMAN-guided active learning framework (called ‘BATMAN-AL’), which iteratively selects a small
217 number of mutant peptides that provide the maximum information about the cross-reactivity of a given TCR. Intuitively, these
218 peptides should (a) be diverse, so as to adequately cover the peptide space, and (b) lie near current decision boundaries, thus
219 helping classifiers better delineate between activation classes. Specifically, in each experimental round, the TCR binding of
220 a set of mutant peptides is determined experimentally, and then the binding results are fed-back to update the model, which
221 subsequently picks the next set of maximally informative peptides to test in the next round, until predictive performance
222 converges. Computationally, it may be advantageous to select only one peptide per round, so that the model can use binding
223 information from this peptide to better select the next peptide; however, this may be experimentally prohibitive because of
224 constraints on the total number of rounds a TCR can be assayed, thus requiring the selection of multiple peptides per round. To
225 address these requirements, we developed AL strategies to maximize classification performance boost of each AL round by
226 sampling the most informative and diverse set of peptides in each round (Extended Data Fig 9).

227 BATMAN-AL achieved a superior classification performance with mean AUC of 0.786, only slightly lower than the AUC
228 for within-TCR task (which used 144 training peptides), using only ~ 45 total peptides per TCR selected over 5 rounds
229 (Figure 4A). Even if constrained to only select 9 total peptides in a single round, BATMAN-AL achieved a mean AUC of
230 0.730, outperforming the leave-one-TCR-out mean AUC of 0.687 (Figure 4A), which suggests that even a minimal amount of
231 active learning can boost performance. BATMAN-AL also outperformed random active learning (BATMAN-RL; Figure 4A),
232 where training peptides are chosen randomly. Furthermore, BATMAN-AL demonstrated much less variance across TCRs
233 compared to BATMAN under the LOO-TCR mode, and the variance was comparable to within-TCR levels after only 5 rounds,
234 demonstrating how small, carefully chosen training set sizes can boost robustness across TCRs.

235 We next explored the performance of BATMAN-AL under various constraints on the total number of peptides to test and
236 number of experimental rounds. For example, if a fixed number of peptides are tested (say, N_p), is it best to select all N_p of them
237 in a single round, or might it be better to select $N_p/4$ peptides per round over 4 rounds? In general, we found that performing
238 more rounds of experiments for a fixed total number of peptides improves predictive performance (Figure 4B; e.g., for $N_p = 72$,
239 mean AUC increases from 0.787 in single round to 0.798 for 4 rounds), although for fewer total peptides (e.g., $N_p = 36$), a
240 single active learning round is sufficient. Testing more peptides, unsurprisingly, improves performance, although if a large
241 number of peptides can be tested (e.g., $N_p = 108$), a single active learning round is again sufficient to achieve within-TCR-level
242 performance.

What determines how much BATMAN will benefit from active learning for a novel TCR? We found that larger median peptide-to-index distances of all single-AA mutants of its index peptide (calculated using the pan-TCR weight profile and AA substitution matrix) correlated with performance boost with one AL round with 9 peptides (Figure 4C). One reason for this could be that larger median distances indicate a broad range of effects for single-AA mutants (Extended Data Fig 9A), and thus better separability between classes that can be learned by AL. Furthermore, TCRs with smaller median peptide-to-index distances often corresponded to narrower within-TCR-activation-class distance distributions (e.g., among strong-activating mutants of TCRs A23 and 18A2 in Extended Data Fig 9A), and so already better LOO-TCR classification performance than other TCRs, gaining only a slight additional performance boost by AL (Figure 4A). For example, TCR a3a, having the largest median peptide-to-index distance, demonstrated the most improvement with AL, with its AUC approaching within-TCR levels after one AL round using 9 peptides (Figure 4A). Comparison of a3a positional weights profiles (Figure 4D) showed that after the single AL round, the weight profile indeed closely matched the within-TCR profile.

In summary, BATMAN-AL can efficiently learn the binding profile of novel TCRs from only a handful of example and can provide custom testing strategies based on experimental constraints.

BATMAN TCR activation predictions generalize to multiple peptide mutations

While BATCAVE records effects of single-AA mutations on activation of individual TCRs, we asked whether BATMAN predictions can be generalized beyond single-AA mutants. For this benchmark, we collected multi-AA mutant TCR activation datasets for the BATCAVE TCRs for which single-AA mutants were evaluated by the same assay in the same source publication (n=6 TCRs). While BATMAN benchmarked on single-AA mutants of unique index peptides showed significant improvement with pooling across TCRs (Figure 2E), here we used unpooled BATMAN because (1) the training set is biased in the number of TCRs per index peptide (3 MAGE-A3-specific TCRs, but only one specific TCR for the 3 other index peptides), and (2) the training set consists of full single-AA data while the test set consists of all available multi-AA mutant data. Thus, for each TCR, we trained BATMAN in unpooled within-TCR mode over the full single-AA data of individual TCRs, and inferred the TCR-specific positional weights and AA matrix as before. We defined the peptide-to-index distance of multi-AA mutants by summing over the product of positional weight and AA substitution distance for all mutational positions. We then used the peptide-to-index distances to rank multi-AA mutants and investigated whether this ranking predicts TCR activation.

We found that BATMAN could identify strong TCR activating mutants at all Hamming distances (Figure 5A). For example, among 32 Hamming distance 8 mutants for TCR c259 (index peptide SLLMWITQC), the sole binder NVSLWLSAV [52] was correctly ranked as top. As another motivating example, we considered the a3a TCR, which was designed to target the cancer-testis antigen MAGE-A3 (EVDPIGHLY). However, it resulted in fatal off-target cross-reactivity to muscle antigen Titin (ESDPIVAQY), differing from MAGE-A3 by 4 AA mutations [53, 54]. We found that BATMAN, trained on only single-AA mutants, ranked Titin 6th among 29 4-AA mutants and 8th among all 111 multi-AA mutants in our a3a dataset. These results indicate that a complete single-AA mutational scan can generalize to some extent to multi-AA mutated cross-reactive targets using BATMAN.

Can BATMAN-AL better identify TCR-activating multi-AA mutants from only single-AA mutant training? We tested this using the a3a TCR and found that the peptide-to-index distances of BATMAN-AL, using only 9 single-AA mutants, performed better than pan-TCR distances in classifying multi-AA mutant activation, with the rank of Titin improving from 41st to 16th among 111 multi-AA mutants in our a3a dataset (Figure 5B). These results suggest that BATMAN and BATMAN-AL predictions can identify cross-reactive targets for a novel TCR beyond those in the 1-Hamming space.

BATMAN predicts polyclonal T cell response

While we have thus focused on predicting activation of individual TCR clones against index peptide mutations, in some situations (e.g., predicting neoantigen immunogenicity, viral escape mutants), we are interested in the collective response of T cell clones present in a polyclonal repertoire with different frequencies. For example, if a viral peptide elicits strong response from the peptide-specific polyclonal T cell repertoire, can BATMAN predict which mutants of that peptide could fail to elicit response from the repertoire, and thus may lead to viral escape? For the Human Cytomegalovirus peptide (NLVPMVATV), BATCAVE contained full mutational scan training data for 25 individual TCRs. For testing BATMAN predictions on polyclonal T cell response, we collected the TScan assay dataset measuring continuous-valued, NLV-expanded polyclonal TCR repertoire response towards all single-AA NLV mutants individually [55].

How well can individual NLV-specific TCR clone activation data in BATCAVE predict polyclonal NLV-specific TCR response? To test this, we trained BATMAN in within-TCR mode, over the full single-AA mutant data for individual TCR clones and inferred the AA matrix and TCR-specific positional weights as before. We then used the resulting weights to calculate TCR-specific peptide-to-index distances for all NLV mutants, and calculated their correlations with the polyclonal T cell response. Different NLV-specific TCRs recognize different NLV mutants [17, 55] and thus, are representative of the

295 full polyclonal response to various degrees. This was reflected in the variable correlations of the peptide-to-index distances
296 from individual TCRs with the polyclonal response (e.g., $r=0.63$ for NLV3 but $r=0.47$ for TCR2 and $r=0.25$ for TCR52-10;
297 **Figure 5C**), indicating that individual NLV-specific TCR clone activation data to NLV mutants cannot be reliably used to predict
298 polyclonal peptide-specific T cell responses.

299 In contrast, when training BATMAN in pan-TCR mode on all NLV-specific TCR data in BATCAVE, pan-TCR predictions
300 correlated strongly ($r=0.61$; **Figure 5C**) to the polyclonal response. This indicates that pooling across TCRs is again critical,
301 as the pan-TCR prediction found from pooling across individual TCRs statistically resembles the polyclonal response more
302 accurately than individual TCRs. These results suggest that BATMAN provides useful predictions about polyclonal T cell
303 response, peptide immunogenicity, and potential viral escape mutants.

304 Discussion

305 We curated BATCAVE, a comprehensive database of experimentally validated positive and negative TCR-pMHC interactions
306 from mutational scan and TCR-pMHC binding affinity assays. BATCAVE contains raw and normalized TCR activation data
307 for both single- and multi- AA peptide mutations, and for both MHCI-restricted and MHCII-restricted TCRs. In addition,
308 BATCAVE contains full TCR sequence information and pMHC binding affinities when available. Using this database, we
309 unraveled the nature of TCR cross-reactivity, including (1) the relationship between pMHC binding and TCR activation; (2)
310 the relationship between TCR-pMHC binding and TCR activation; and (3) insights about how biochemical features (e.g.,
311 hydrophobicity) of the TCR's wildtype AA affects TCR binding.

312 Despite these advances, there are some biases in BATCAVE that require careful interpretation. First, many datasets included
313 used *in vitro* TCR-pMHC assays with, e.g., UV-mediated peptide exchange for MHC loading (see supplementary notes), and
314 thus the TCR activation level by such pMHCs (**Figure 1D**) might be different than *in vivo* settings. Second, most BATCAVE
315 TCRs belong to peripheral cytotoxic T cells, which survived thymic selection [56]. This thymic selection bias on TCR repertoire
316 [57–62] may affect, for example, the relationship between pMHC binding and TCR activation, or between TCR activation
317 and TCR-pMHC binding affinity. Third, recorded in BATCAVE as a single number, T cell activation for a given TCR-pMHC
318 pair can be heterogeneous, depending on, e.g., expression of co-signaling molecules [63, 64] (e.g., CD5, CD8, and CD4).
319 Notwithstanding these caveats, we believe this database will provide a strong benchmark for evaluating future methods and for
320 uncovering further principles of TCR cross-reactivity.

321 Using this database, we then developed a computational method, called BATMAN, that predicts TCR activation of
322 peptides based on their distances to the TCR's index peptide. BATMAN achieved state-of-the-art performance for predicting
323 cross-reactivity across diverse TCRs and index peptides present in the BATCAVE database. Crucial towards BATMAN's
324 performance were model-inferred positional weights and AA substitution matrices. In comparison, simple distance functions
325 based on mismatch rules (e.g., Hamming distance) with equal weights assigned to each position proved far from sufficient to
326 accurately characterize cross-reactivity. Furthermore, using the learned weights and substitution matrices allowed us to project
327 cross-reactivity outside the single-AA-mutational space, which we demonstrated by accurately ranking immunogenic peptides
328 of clinically relevant TCRs. In addition, the parameters learned by BATMAN revealed striking variation in weight profiles for
329 different TCRs, which was also supported by structural analyses. We then developed an active learning version of BATMAN,
330 which helps BATMAN generalize to novel TCRs. This version provides an efficient way to sample from the prohibitively large
331 antigenic space by iteratively selecting peptides to assay that provide the best improvement of novel TCR activation prediction
332 accuracy. Overall, BATMAN fills a hitherto unoccupied niche of TCR-pMHC prediction methods by accurately discriminating
333 between small differences in peptide sequences, which we show existing methods fail to predict, but which are essential for
334 understanding neoantigen immunogenicity and off-target effects of TCR-based therapies.

335 Finally, there are several directions of future work and corresponding applications enabled by this study. First, BATMAN
336 could be improved by incorporating TCR sequence information into the model and by training on datasets from other types of
337 experimental TCR cross-reactivity assays [65] (e.g., yeast display library enrichment [5, 66], T-Scan [55], and SABR [67]),
338 which sample more comprehensively outside the one AA-mutational scan space. Second, insights from our database and method
339 could be used to predict how TCR sequences may determine the relationship between TCR binding affinity and cross-reactivity,
340 and to design high-affinity TCRs with limited off-target cross-reactivity, which has numerous clinical applications.

341 Author contributions

342 AB, DJP, SN and HVM conceptualized the work; AB developed the software; AB and DJP designed the model; AB and CW
343 implemented the user interface with help from SRC; AB and PB curated the database; AB conducted all formal analyses except
344 Protein Data Bank structure analyses (AA and HVM); AB, SN, and HVM wrote the original draft; all authors reviewed and
345 edited the final draft; SN and HVM supervised the work.

346 Competing interests

347 The authors declare no competing interests.

348 Acknowledgments

349 We thank our lab members and the anonymous referees for discussion and feedback on method development and figure design,
350 Paul G. Thomas and Zachary Sethna for discussion on TCR datasets, Anastasia Troshina and Vasilisa A. Kovaleva for designing
351 the BATMAN logo, and all authors who shared their datasets with us. TCR-pMHC schematics in [Figure 1B](#), [Figure 2B,C](#) were
352 created with BioRender (<https://www.biorender.com/>).

353 Funding

354 The research was supported by the Simons Center for Quantitative Biology at Cold Spring Harbor Laboratory; US National
355 Institutes of Health Grants S10OD028632 and 1R01AI167862; and the Simons Pivot Fellowship. This work was discussed in
356 part at the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-2210452. The funders had
357 no role in the template design or decision to publish.

358 References

- 359 ¹A. K. Sewell, “Why must t cells be cross-reactive?”, *Nature Reviews Immunology* **12**, 669–677 (2012).
- 360 ²L. Wooldridge, J. Ekeruche-Makinde, H. a. Van Den Berg, A. Skowera, J. J. Miles, M. P. Tan, G. Dolton, M. Clement, S.
361 Llewellyn-Lacey, D. A. Price, et al., “A single autoimmune t cell receptor recognizes more than a million different peptides”,
362 *Journal of Biological Chemistry* **287**, 1168–1177 (2012).
- 363 ³R. Keeton, M. B. Tincho, A. Ngomti, R. Baguma, N. Benede, A. Suzuki, K. Khan, S. Cele, M. Bernstein, F. Karim, et al., “T
364 cell responses to sars-cov-2 spike cross-recognize omicron”, *Nature* **603**, 488–492 (2022).
- 365 ⁴M. Łuksza, Z. M. Sethna, L. A. Rojas, J. Lihm, B. Bravi, Y. Elhanati, K. Soares, M. Amisaki, A. Dobrin, D. Hoyos, et al.,
366 “Neoantigen quality predicts immunoediting in survivors of pancreatic cancer”, *Nature* **606**, 389–395 (2022).
- 367 ⁵X. Yang, L. I. Garner, I. V. Zvyagin, M. A. Paley, E. A. Komech, K. M. Jude, X. Zhao, R. A. Fernandes, L. M. Hassman,
368 G. L. Paley, et al., “Autoimmunity-associated t cell receptors recognize hla-b* 27-bound peptides”, *Nature* **612**, 771–777
369 (2022).
- 370 ⁶K. Ishii, J. S. Davies, A. L. Sinkoe, K. A. Nguyen, S. M. Norberg, C. P. McIntosh, T. Kadakia, C. Serna, Z. Rae, M. C. Kelly,
371 et al., “Multi-tiered approach to detect autoimmune cross-reactivity of therapeutic t cell receptors”, *Science Advances* **9**,
372 eadg9845 (2023).
- 373 ⁷N. S. Abdelfattah, T. Kula, and S. J. Elledge, “T-switch: a specificity-based engineering platform for developing safe and
374 effective t cell therapeutics”, *Immunity* **57**, 2945–2958 (2024).
- 375 ⁸D. Hudson, R. A. Fernandes, M. Basham, G. Ogg, and H. Koohy, “Can we predict t cell specificity with digital biology and
376 machine learning?”, *Nature Reviews Immunology*, 1–11 (2023).
- 377 ⁹P. Moris, J. De Pauw, A. Postovskaya, S. Gielis, N. De Neuter, W. Bittremieux, B. Ogunjimi, K. Laukens, and P. Meysman,
378 “Current challenges for unseen-epitope tcr interaction prediction and a new perspective derived from image classification”,
379 *Briefings in Bioinformatics* **22**, bbaa318 (2021).
- 380 ¹⁰D. M. Mason and S. T. Reddy, “Predicting adaptive immune receptor specificities by machine learning is a data generation
381 problem”, *Cell Systems* **15**, 1190–1197 (2024).
- 382 ¹¹F. Drost, A. Chernysheva, M. Albahah, K. Kocher, K. Schober, and B. Schubert, “Benchmarking of t-cell receptor-epitope
383 predictors with eptope-tcr”, *bioRxiv*, 2024–11 (2024).
- 384 ¹²C. Dens, K. Laukens, W. Bittremieux, and P. Meysman, “The pitfalls of negative data bias for the t-cell epitope specificity
385 challenge”, *Nature Machine Intelligence* **5**, 1060–1062 (2023).
- 386 ¹³B. McMaster, C. Thorpe, G. Ogg, C. M. Deane, and H. Koohy, “Can alphafold’s breakthrough in protein structure help
387 decode the fundamental principles of adaptive cellular immunity?”, *Nature Methods*, 1–11 (2024).
- 388 ¹⁴H. Koohy, D. Hudson, A. Lubbock, and M. Basham, “A comparison of clustering models for inference of t cell receptor
389 antigen specificity”, *bioRxiv*, 2023–08 (2023).
- 390 ¹⁵F. Grazioli, A. Mösch, P. Machart, K. Li, I. Alqassem, T. J. O’Donnell, and M. R. Min, “On tcr binding predictors failing to
391 generalize to unseen peptides”, *Frontiers in Immunology* **13**, 1014256 (2022).

- 392 ¹⁶L. Deng, C. Ly, S. Abdollahi, Y. Zhao, I. Prinz, and S. Bonn, “Performance comparison of tcr-pmhc prediction tools reveals a
393 strong data dependency”, *Frontiers in Immunology* **14**, 1128326 (2023).
- 394 ¹⁷E. Dorigatti, F. Drost, A. Straub, P. Hilgendorf, K. I. Wagner, B. Bischl, D. Busch, K. Schober, and B. Schubert, “Predicting t
395 cell receptor functionality against mutant epitopes”, *bioRxiv*, 2023–05 (2023).
- 396 ¹⁸A. Straub, S. Grassmann, S. Jarosch, L. Richter, P. Hilgendorf, M. Hammel, K. I. Wagner, V. R. Buchholz, K. Schober, and
397 D. H. Busch, “Recruitment of epitope-specific t cell clones with a low-avidity threshold supports efficacy against mutational
398 escape upon re-infection”, *Immunity* **56**, 1269–1284 (2023).
- 399 ¹⁹A. K. Bentzen, L. Such, K. K. Jensen, A. M. Marquard, L. E. Jessen, N. J. Miller, C. D. Church, R. Lyngaa, D. M. Koelle,
400 J. C. Becker, et al., “T cell receptor fingerprinting enables in-depth characterization of the interactions governing recognition
401 of peptide–mhc complexes”, *Nature biotechnology* **36**, 1191–1196 (2018).
- 402 ²⁰C. H. Coles, R. M. Mulvaney, S. Malla, A. Walker, K. J. Smith, A. Lloyd, K. L. Lowe, M. L. McCully, R. Martinez Hague,
403 M. Aleksic, et al., “Tcrs with distinct specificity profiles use different binding modes to engage an identical peptide–hla
404 complex”, *The Journal of Immunology* **204**, 1943–1953 (2020).
- 405 ²¹B. Reynisson, B. Alvarez, S. Paul, B. Peters, and M. Nielsen, “Netmhcpant-4.1 and netmhciipan-4.0: improved predictions of
406 mhc antigen presentation by concurrent motif deconvolution and integration of ms mhc eluted ligand data”, *Nucleic acids
407 research* **48**, W449–W454 (2020).
- 408 ²²M. Lever, P. K. Maini, P. A. Van Der Merwe, and O. Dushek, “Phenotypic models of t cell activation”, *Nature Reviews
409 Immunology* **14**, 619–629 (2014).
- 410 ²³R. A. Mariuzza, P. Agnihotri, and J. Orban, “The structural basis of t-cell receptor (tcr) activation: an enduring enigma”,
411 *Journal of Biological Chemistry* **295**, 914–925 (2020).
- 412 ²⁴L. Sušac, M. T. Vuong, C. Thomas, S. von Bülow, C. O’Brien-Ball, A. M. Santos, R. A. Fernandes, G. Hummer, R. Tampe,
413 and S. J. Davis, “Structure of a fully assembled tumor-specific t cell receptor ligated by pmhc”, *Cell* **185**, 3201–3213 (2022).
- 414 ²⁵L. V. Sibener, R. A. Fernandes, E. M. Kolawole, C. B. Carbone, F. Liu, D. McAfee, M. E. Birnbaum, X. Yang, L. F. Su,
415 W. Yu, et al., “Isolation of a structural mechanism for uncoupling t cell receptor signaling from peptide-mhc binding”, *Cell*
416 **174**, 672–687 (2018).
- 417 ²⁶H.-K. Choi, P. Cong, C. Ge, A. Natarajan, B. Liu, Y. Zhang, K. Li, M. N. Rushdi, W. Chen, J. Lou, et al., “Catch bond models
418 may explain how force amplifies tcr signaling and antigen discrimination”, *Nature communications* **14**, 2616 (2023).
- 419 ²⁷S. J. Davis and P. A. van der Merwe, “Tcr triggering: co-receptor-dependent or-independent?”, *Trends in immunology* **24**,
420 624–626 (2003).
- 421 ²⁸A. M. Mørch, Š. Bálint, A. M. Santos, S. J. Davis, and M. L. Dustin, “Coreceptors and tcr signaling—the strong and the weak
422 of it”, *Frontiers in cell and developmental biology* **8**, 597627 (2020).
- 423 ²⁹B. Laugel, H. A. van den Berg, E. Gostick, D. K. Cole, L. Wooldridge, J. Boulter, A. Milicic, D. A. Price, and A. K. Sewell,
424 “Different t cell receptor affinity thresholds and cd8 coreceptor dependence govern cytotoxic t lymphocyte activation and
425 tetramer binding properties”, *Journal of Biological Chemistry* **282**, 23799–23810 (2007).
- 426 ³⁰A. Moritz, R. Anjanappa, C. Wagner, S. Bunk, M. Hofmann, G. Pszolla, A. Saikia, M. Garcia-Alai, R. Meijers, H.-G.
427 Rammensee, et al., “High-throughput peptide-mhc complex generation and kinetic screenings of tcrs with peptide-receptive
428 hla-a* 02: 01 molecules”, *Science Immunology* **4**, eaav0860 (2019).
- 429 ³¹F. Duan, J. Duitama, S. Al Seesi, C. M. Ayres, S. A. Corcelli, A. P. Pawashe, T. Blanchard, D. McMahon, J. Sidney,
430 A. Sette, et al., “Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer
431 immunogenicity”, *Journal of Experimental Medicine* **211**, 2231–2248 (2014).
- 432 ³²J. Schmidt, A. R. Smith, M. Magnin, J. Racine, J. R. Devlin, S. Bobisse, J. Cesbron, V. Bonnet, S. J. Carmona, F. Huber, et al.,
433 “Prediction of neo-epitope immunogenicity reveals tcr recognition determinants and provides insight into immunoediting”,
434 *Cell Reports Medicine* **2** (2021).
- 435 ³³D. Chowell, S. Krishna, P. D. Becker, C. Cocita, J. Shu, X. Tan, P. D. Greenberg, L. S. Klavinskis, J. N. Blattman, and
436 K. S. Anderson, “Tcr contact residue hydrophobicity is a hallmark of immunogenic cd8+ t cell epitopes”, *Proceedings of the
437 National Academy of Sciences* **112**, E1754–E1762 (2015).
- 438 ³⁴W. C. Wimley and S. H. White, “Experimentally determined hydrophobicity scale for proteins at membrane interfaces”,
439 *Nature structural biology* **3**, 842–848 (1996).

- 440 ³⁵V. K. Karnaikhov, D. S. Shcherbinin, A. O. Chugunov, D. M. Chudakov, R. G. Efremov, I. V. Zvyagin, and M. Shugay,
441 “Structure-based prediction of t cell receptor recognition of unseen epitopes using tcen”, *Nature Computational Science* **4**,
442 510–521 (2024).
- 443 ³⁶S. Dai, E. S. Huseby, K. Rubtsova, J. Scott-Browne, F. Crawford, W. A. Macdonald, P. Marrack, and J. W. Kappler,
444 “Crossreactive t cells spotlight the germline rules for $\alpha\beta$ t cell-receptor interactions with mhc molecules”, *Immunity* **28**,
445 324–334 (2008).
- 446 ³⁷D. V. Bagaev, R. M. Vroomans, J. Samir, U. Stervbo, C. Rius, G. Dolton, A. Greenshields-Watson, M. Attaf, E. S. Egorov,
447 I. V. Zvyagin, et al., “Vdjdb in 2019: database extension, new analysis infrastructure and a t-cell receptor motif compendium”,
448 *Nucleic Acids Research* **48**, D1057–D1062 (2020).
- 449 ³⁸N. Tickotsky, T. Sagiv, J. Prilusky, E. Shifrut, and N. Friedman, “Mcpas-tcr: a manually curated catalogue of pathology-
450 associated t cell receptor sequences”, *Bioinformatics* **33**, 2924–2929 (2017).
- 451 ³⁹10X Genomics, “LIT047_a New Way of Exploring Immunity - Linking Highly Multiplexed Antigen Recognition to Immune
452 Repertoire and Phenotype”, *10xGenomics*, 1–13 (2019).
- 453 ⁴⁰T. J. O’Donnell, C. Kanduri, G. Isacchini, J. P. Limenitakis, R. A. Brachman, R. A. Alvarez, I. H. Haff, G. K. Sandve, and
454 V. Greiff, “Reading the repertoire: progress in adaptive immune receptor analysis using machine learning”, *Cell Systems* **15**,
455 1168–1189 (2024).
- 456 ⁴¹J. J. Calis, M. Maybeno, J. A. Greenbaum, D. Weiskopf, A. D. De Silva, A. Sette, C. Keşmir, and B. Peters, “Properties of
457 mhc class i presented peptides that enhance immunogenicity”, *PLoS computational biology* **9**, e1003266 (2013).
- 458 ⁴²A. Weber, J. Born, and M. Rodriguez Martínez, “Titan: t-cell receptor specificity prediction with bimodal attention networks”,
459 *Bioinformatics* **37**, i237–i244 (2021).
- 460 ⁴³S. Frankild, R. J. De Boer, O. Lund, M. Nielsen, and C. Kesmir, “Amino acid similarity accounts for t cell cross-reactivity
461 and for “holes” in the t cell repertoire”, *PloS one* **3**, e1831 (2008).
- 462 ⁴⁴H. Xia, J. McMichael, M. Becker-Hapak, O. C. Onyeador, R. Buchli, E. McClain, P. Pence, S. Supaphol, M. M. Richters,
463 A. Basu, et al., “Computational prediction of mhc anchor locations guides neoantigen identification and prioritization”,
464 *Science immunology* **8**, eabg2200 (2023).
- 465 ⁴⁵M. E. Birnbaum, J. L. Mendoza, D. K. Sethi, S. Dong, J. Glanville, J. Dobbins, E. Özkan, M. M. Davis, K. W. Wucherpfennig,
466 and K. C. Garcia, “Deconstructing the peptide-mhc specificity of t cell recognition”, *Cell* **157**, 1073–1087 (2014).
- 467 ⁴⁶J. J. Adams, S. Narayanan, M. E. Birnbaum, S. S. Sidhu, S. J. Blevins, M. H. Gee, L. V. Sibener, B. M. Baker, D. M. Kranz,
468 and K. C. Garcia, “Structural interplay between germline interactions and adaptive recognition determines the bandwidth of
469 tcr-peptide-mhc cross-reactivity”, *Nature immunology* **17**, 87–94 (2016).
- 470 ⁴⁷M. H. Dezfulian, T. Kula, T. Pranzatelli, N. Kamitaki, Q. Meng, B. Khatri, P. Perez, Q. Xu, A. Chang, A. C. Kohlgruber,
471 et al., “Tscan-ii: a genome-scale platform for the de novo identification of cd4+ t cell epitopes”, *Cell* (2023).
- 472 ⁴⁸S. Ranasinghe, P. A. Lamothe, D. Z. Soghoian, S. W. Kazer, M. B. Cole, A. K. Shalek, N. Yosef, R. B. Jones, F. Donaghey,
473 C. Nwonu, et al., “Antiviral cd8+ t cells restricted by human leukocyte antigen class ii exist during natural hiv infection and
474 exhibit clonal expansion”, *Immunity* **45**, 917–930 (2016).
- 475 ⁴⁹A. M. Bulek, D. K. Cole, A. Skowera, G. Dolton, S. Gras, F. Madura, A. Fuller, J. J. Miles, E. Gostick, D. A. Price, et al.,
476 “Structural basis for the killing of human beta cells by cd8+ t cells in type 1 diabetes”, *Nature immunology* **13**, 283–289
477 (2012).
- 478 ⁵⁰C. Szeto, P. Zareie, R. C. Wirasinha, J. B. Zhang, A. T. Nguyen, A. Riboldi-Tunnicliffe, N. L. La Gruta, S. Gras, and
479 S. R. Daley, “Covalent tcr-peptide-mhc interactions induce t cell activation and redirect t cell fate in the thymus”, *Nature
480 Communications* **13**, 4951 (2022).
- 481 ⁵¹A. Sachs, E. Moore, Z. Kosaloglu-Yalcin, B. Peters, J. Sidney, S. A. Rosenberg, P. F. Robbins, and A. Sette, “Impact of
482 cysteine residues on mhc binding predictions and recognition by tumor-reactive t cells”, *The Journal of Immunology* **205**,
483 539–549 (2020).
- 484 ⁵²A. R. Karapetyan, C. Chaipan, K. Winkelbach, S. Wimberger, J. S. Jeong, B. Joshi, R. B. Stein, D. Underwood, J. C. Castle,
485 M. van Dijk, et al., “Tcr fingerprinting and off-target peptide identification”, *Frontiers in Immunology* **10**, 2501 (2019).
- 486 ⁵³G. P. Linette, E. A. Stadtmauer, M. V. Maus, A. P. Rapoport, B. L. Levine, L. Emery, L. Litzky, A. Bagg, B. M. Carreno,
487 P. J. Cimino, et al., “Cardiovascular toxicity and titin cross-reactivity of affinity-enhanced t cells in myeloma and melanoma”,
488 *Blood, The Journal of the American Society of Hematology* **122**, 863–871 (2013).

- 489 ⁵⁴B. J. Cameron, A. B. Gerry, J. Dukes, J. V. Harper, V. Kannan, F. C. Bianchi, F. Grand, J. E. Brewer, M. Gupta, G. Plesa,
490 et al., “Identification of a titin-derived hla-a1–presented peptide as a cross-reactive target for engineered mage a3-directed t
491 cells”, *Science translational medicine* **5**, 197ra103–197ra103 (2013).
- 492 ⁵⁵T. Kula, M. H. Dezfulian, C. I. Wang, N. S. Abdelfattah, Z. C. Hartman, K. W. Wucherpfennig, H. K. Lyerly, and S. J. Elledge,
493 “T-scan: a genome-wide method for the systematic discovery of t cell epitopes”, *Cell* **178**, 1016–1028 (2019).
- 494 ⁵⁶K. M. Ashby and K. A. Hogquist, “A guide to thymic selection of t cells”, *Nature Reviews Immunology* **24**, 103–117 (2024).
- 495 ⁵⁷E. S. Huseby, J. White, F. Crawford, T. Vass, D. Becker, C. Pinilla, P. Marrack, and J. W. Kappler, “How the t cell repertoire
496 becomes peptide and mhc specific”, *Cell* **122**, 247–260 (2005).
- 497 ⁵⁸B. D. Stadinski, K. Shekhar, I. Gómez-Touriño, J. Jung, K. Sasaki, A. K. Sewell, M. Peakman, A. K. Chakraborty, and
498 E. S. Huseby, “Hydrophobic cdr3 residues promote the development of self-reactive t cells”, *Nature immunology* **17**, 946–955
499 (2016).
- 500 ⁵⁹J. Textor, F. Buytenhuijs, D. Rogers, È. M. Gauthier, S. Sultan, I. M. Wortel, K. Kalies, A. Fähnrich, R. Pagel, H. J. Melichar,
501 et al., “Machine learning analysis of the t cell receptor repertoire identifies sequence features of self-reactivity”, *Cell Systems*
502 **14**, 1059–1073 (2023).
- 503 ⁶⁰K. A. Lagattuta, J. B. Kang, A. Nathan, K. E. Pauken, A. H. Jonsson, D. A. Rao, A. H. Sharpe, K. Ishigaki, and S.
504 Raychaudhuri, “Repertoire analyses reveal t cell antigen receptor sequence features that influence t cell fate”, *Nature
505 immunology* **23**, 446–457 (2022).
- 506 ⁶¹D. Luppov, E. Vlasova, D. Chudakov, and M. Shugay, “Comprehensive analysis of $\alpha\beta$ -t-cell receptor repertoires reveals
507 signatures of thymic selection”, *bioRxiv* (2025).
- 508 ⁶²J. Lu, F. Van Laethem, A. Bhattacharya, M. Craveiro, I. Saba, J. Chu, N. C. Love, A. Tikhonova, S. Radaev, X. Sun, et al.,
509 “Molecular constraints on cdr3 for thymic selection of mhc-restricted tcgs from a random pre-selection repertoire”, *Nature
510 communications* **10**, 1019 (2019).
- 511 ⁶³J. Cabezas Caballero, A. Huhn, M. A. Kutuzov, V. Andre, A. Shomuradova, P. A. van der Merwe, and O. Dushek, “Generation
512 of t cells with reduced off-target cross-reactivities by engineering co-signalling receptors”, *bioRxiv*, 2024–10 (2024).
- 513 ⁶⁴S. Singhaviranon, J. P. Dempsey, A. T. Hagymasi, I. I. Mandoiu, and P. K. Srivastava, “Low-avidity t cells drive endogenous
514 tumor immunity in mice and humans”, *Nature Immunology*, 1–12 (2025).
- 515 ⁶⁵A. V. Joglekar and G. Li, “T cell antigen discovery”, *Nature methods* **18**, 873–880 (2021).
- 516 ⁶⁶M. H. Gee, A. Han, S. M. Lofgren, J. F. Beausang, J. L. Mendoza, M. E. Birnbaum, M. T. Bethune, S. Fischer, X. Yang,
517 R. Gomez-Eerland, et al., “Antigen identification for orphan t cell receptors expressed on tumor-infiltrating lymphocytes”,
518 *Cell* **172**, 549–563 (2018).
- 519 ⁶⁷A. V. Joglekar, M. T. Leonard, J. D. Jeppson, M. Swift, G. Li, S. Wong, S. Peng, J. M. Zaretsky, J. R. Heath, A. Ribas, et al.,
520 “T cell antigen discovery via signaling and antigen-presenting bifunctional receptors”, *Nature methods* **16**, 191–198 (2019).
- 521 ⁶⁸D. Chandler, “Interfaces and the driving force of hydrophobic assembly”, *Nature* **437**, 640–647 (2005).
- 522 ⁶⁹A. Kucukelbir, D. Tran, R. Ranganath, A. Gelman, and D. M. Blei, “Automatic differentiation variational inference”, *Journal
523 of machine learning research* **18**, 1–45 (2017).
- 524 ⁷⁰E. C. Meng, T. D. Goddard, E. F. Petersen, G. S. Couch, Z. J. Pearson, J. H. Morris, and T. E. Ferrin, “Ucsf chimeraX: tools
525 for structure building and analysis”, *Protein Science* **32**, e4792 (2023).
- 526 ⁷¹J. E. J. Mills and P. M. Dean, “Three-dimensional hydrogen-bond geometry and probability information from a crystal
527 survey”, en, *Journal of Computer-Aided Molecular Design* **10**, 607–622 (1996).

528 Methods

529 TCR activation dataset collection and processing

530 We collected continuous TCR-pMHC datasets for complete single-AA mutational scans from all publications containing raw
531 datasets ($n=21$). To normalize datasets across publications, we scaled TCR activation values by the maximum activation value
532 over all recorded peptides tested against that TCR. The only exceptions to this normalization scheme were for experiments
533 where the TCR activation measurements were on a logarithmic scale (e.g., EC50 values), in which case we used the logarithm
534 of the TCR activation values and linearly transformed them to map to the [0,1] interval. Following previous works [17],
535 we discretized the normalized TCR activation values to 3 ordered levels for downstream classification tasks: no activation
536 ($a_{no} \in [0, 0.1]$), weak activation ($a_{weak} \in [0.1, 0.5]$), and strong activation ($a_{strong} \in [0.5, 1]$). For regression tasks, we directly
537 used the normalized TCR activation values. More technical TCR-specific notes on data collection and processing, as well as
538 links to source publications, can be found in the Supplementary Notes. A number of publications (see Supplementary Materials
539 for citations) contained further mutational scan experiments relevant for our database, but the associated raw datasets were not
540 readily available to us.

To select the benchmark subset of TCRs (e.g., for Figure 2A,D) among all TCRs specific for a common index peptide, we selected the one with the highest balance of TCR activation classes. For this purpose, for each TCR, we calculated the entropy of the frequency distribution of the TCR activation classes, as follows

$$S_{TCR} = -f_S \log f_S - f_W \log f_W - f_N \log f_N, \quad (1)$$

541 with f_S, f_W, f_N being the fraction of strong-, weak-, and non-activating single-AA mutant peptides respectively. For each
542 unique index peptide, we selected the TCR with the highest S_{TCR} . The only exception to this rule was for the index peptide
543 FEAQKAKANKAVD; while all other index peptides had TCR sequences available for either all or none of their specific
544 TCRs, only a subset of FEAQKAKANKAVD-specific TCRs had available sequence. So, for this peptide, we selected the TCR
545 B3K508, with the second highest S_{TCR} among FEAQKAKANKAVD-specific TCRs, since the corresponding TCR sequence,
546 necessary for many TCR-pMHC model predictions in Figure 2A, was available.

547 Web application for visualizing TCR-pMHC interactions from our database

548 TCR-pMHC interactions from our database (Figure 1B) are visualized via the web application at <https://batman.cshl.edu/>. All
549 interactive plots are deployed as a RShiny application, using *ShinyDashboard* (v. 0.7.2). The scatter plot displaying peptide
550 clustering based on index-to-mutant distance was generated via *ggplot2* (v. 2_3.4.4) and rendered using *plotly* (v. 4.10.3). The
551 heatmap presenting normalized peptide activation per index peptide was generated with *InteractiveComplexHeatmap* (v. 1.8.0).
552 The Alluvium plot visualizing the binding of index and mutated peptides to TCRs was generated with *ggplot2* and *ggalluvial* (v.
553 0.12.5). The code for the application is available at https://github.com/meyer-lab-cshl/batman_shiny.

554 Amino acid hydrophobicity calculation

555 To calculate AA hydrophobicity, we used the Wimley-White interfacial hydrophobicity scale [34], which is known to best
556 describe TCR-pMHC AA binding characteristics [58]. This is likely due to the fact that the binding sites of unligated TCRs
557 and pMHC are solvated, and hydrophobic residues at protein-protein interfaces often constitute binding hotspots due to the
558 hydrophobic effect [36, 68]. For AA residues histidine, glutamic acid, and aspartic acid, the above scale provides hydrophobicity
559 of both neutral and charged residues, where we have used the latter. The resulting AA hydrophobicity ordering can be found,
560 e.g., in the horizontal axis of Figure 1F.

561 MHC binding predictions

562 For all MHC-peptide pairs in BATCAVE, we predicted MHC binding using the command-line interface of netMHCpan4.0 [21].
563 We restricted binding predictions to the full length of the peptide sequence and classified binding based on ‘Rank_BA’ into ‘no’
564 [$> 2\%$ rank], ‘weak’ [$0.5\text{--}2\%$ rank], or ‘strong’ [$< 0.5\%$ rank] binders.

565 Training and validation of BATMAN

566 Bayesian hierarchical classifier for TCR activation

We first describe how BATMAN works for a given TCR in within-TCR mode. For classification tasks, BATMAN (Figure 2B) performs Bayesian logistic regression to predict the ordered categorical activation level for the given TCR and peptide, $a(\text{peptide}) \in \{a_{no}, a_{weak}, a_{strong}\}$, using the peptide-to-index distance $d(\text{peptide}, \text{index})$ corresponding to the index peptide of

the TCR, using this link function:

$$\text{Prob}[a(\text{peptide}) | d(\text{peptide}, \text{index})] = \begin{cases} 1 - \text{logit}^{-1}(d_0 - d(\text{peptide}, \text{index}) - c_1), & \text{if } a(\text{peptide}) = a_{\text{non}} \\ \text{logit}^{-1}(d_0 - d(\text{peptide}, \text{index}) - c_1) - \text{logit}^{-1}(d_0 - d(\text{peptide}, \text{index}) - c_2), & \text{if } a(\text{peptide}) = a_{\text{weak}} \\ \text{logit}^{-1}(d_0 - d(\text{peptide}, \text{index}) - c_2), & \text{if } a(\text{peptide}) = a_{\text{strong}} \end{cases} \quad (2)$$

where the inverse logit function is defined as $\text{logit}^{-1}(x) = \frac{1}{1+e^{-x}}$, d_0 is a constant intercept and c_1 and c_2 are two constant cutpoints with the constraint $c_1 < c_2$, with the following hyperprior distributions:

$$d_0 \sim \text{Normal}(\mu_0, \sigma_0), \quad (3)$$

$$c_1, c_2 \stackrel{iid}{\sim} \text{Normal}(0, 2), \quad (4)$$

$$\mu_0 \sim \text{Normal}(0, 2), \quad (5)$$

$$\sigma_0 \sim \text{HalfNormal}(2). \quad (6)$$

For any peptide-index sequence pair, the peptide-to-index distance $d(\text{peptide}, \text{index})$ is computed based on position-dependent weights $w(\text{position})$ and a 20x20 AA substitution distance matrix M :

$$d(\text{peptide}, \text{index}) = \sum_{\substack{\text{position} \\ \in \{1, 2, \dots, L\}}} w(\text{position}) M[\text{aa}(\text{index}, \text{position}), \text{aa}(\text{peptide}, \text{position})], \quad (7)$$

where each element in $M[\text{aa}(\text{index}, \text{position}), \text{aa}(\text{peptide}, \text{position})]$ corresponds to the substitution of amino acid residue $\text{aa}(\text{index}, \text{position})$ to $\text{aa}(\text{peptide}, \text{position})$ at a given position in the index and peptide sequences, respectively. The diagonal elements of M are all zero, such that the distance from the index peptide to itself is zero. BATMAN infers the weights $w(\text{position})$ and AA distance matrix elements of $M[\text{aa}_1, \text{aa}_2]$ with $\text{aa}_1, \text{aa}_2 \in \{\text{A,C,D,...,W,Y}\}$.

Position-dependent weights $w(\text{position}) \in [0, 1]$ with position $\in \{1, 2, \dots, L\}$, where L is the length of the TCR's index peptide, have the prior:

$$w(\text{position}) \stackrel{iid}{\sim} \text{Beta}(\alpha, \beta). \quad (8)$$

Elements of M follow:

$$M[\text{aa}_1, \text{aa}_2] = D[\text{aa}_1, \text{aa}_2](1 + \delta[\text{aa}_1, \text{aa}_2]), \quad (9)$$

$$\delta[\text{aa}_1, \text{aa}_2] \stackrel{iid}{\sim} \text{Normal}(\mu, \sigma) \quad (10)$$

where D is a pre-defined AA distance matrix (e.g., BLOSUM100) used for constructing the prior for the inferred AA matrix M . The hyperparameters of $d(\text{peptide}, \text{index})$ have the following, weakly informative hyperprior distributions:

$$\alpha, \beta \stackrel{iid}{\sim} \text{Gamma}(2, 2), \quad (11)$$

$$\mu \sim \text{Normal}(0, 0.5), \quad (12)$$

$$\sigma \sim \text{Exponential}(1). \quad (13)$$

BATMAN is also designed to optionally use pMHC binding information. In such cases, [Equation \(7\)](#) is modified as following:

$$d(\text{peptide}, \text{index}) = \sum_{\substack{\text{position} \\ \in \{1, 2, \dots, L\}}} w(\text{position}) M[\text{aa}(\text{index}, \text{position}), \text{aa}(\text{peptide}, \text{position})] + d_{\text{MHC}}, \quad (14)$$

with the scalar d_{MHC} assuming one of 3 possible values based on NetMHCPan-predicted pMHC binding level,

$$d_{\text{MHC}} = \begin{cases} d_{\text{SB}}, & \text{if pMHC pair is strong binder} \\ d_{\text{WB}}, & \text{if pMHC pair is weak binder} \\ d_{\text{NB}}, & \text{if pMHC pair is non binder,} \end{cases} \quad (15)$$

with the prior distribution

$$d_{\text{SB}}, d_{\text{WB}}, d_{\text{NB}} \stackrel{iid}{\sim} \text{Beta}(\alpha_M, \beta_M) \quad (16)$$

with the hyperparameter prior being

$$\alpha_M, \beta_M \stackrel{iid}{\sim} \text{Gamma}(2, 2). \quad (17)$$

572 We verified via prior predictive sampling that these assumptions can yield all anticipated outcomes i.e. activation levels.

573 Regression tasks with BATMAN

To use BATMAN for regression tasks of predicting continuous-valued normalized TCR activation $a(\text{peptide}) \in [0, 1]$, we modified [Equation \(2\)](#) to

$$\text{Prob}[a(\text{peptide}) | d(\text{peptide}, \text{index})] = \text{Normal}(d_0 - d(\text{peptide}, \text{index}), \sigma), \sigma \sim \text{Exponential}(1), \quad (18)$$

574 with all other steps being identical as described above for classification tasks. The performance of such an application is shown
575 in [Extended Data Fig 5A](#).

576 Pooling across TCRs for training BATMAN

The hierarchical Bayesian inference set-up allows BATMAN to integrate datasets from multiple TCRs having the same index peptide length ('pooling across TCRs'). In such cases, the positional weight profiles $w(\text{position}, \text{TCR})$ and the intercepts $d_0(\text{TCR})$ are TCR-specific, but have the same prior distributions as specified above, i.e.,

$$w(\text{position}, \text{TCR}) \stackrel{iid}{\sim} \text{Beta}(\alpha, \beta), \quad (19)$$

and

$$d_0(\text{TCR}) \stackrel{iid}{\sim} \text{Normal}(\mu_0, \sigma_0), \quad (20)$$

with the hyperparameters α, β, μ_0 and σ_0 having hyperpriors as above. These TCR-specific weight profiles are used to calculate TCR-specific peptide-to-index distances $d(\text{peptide}, \text{index}, \text{TCR})$ similarly as above,

$$d(\text{peptide}, \text{index}, \text{TCR}) = \sum_{\substack{\text{position} \\ \in \{1, 2, \dots, L\}}} w(\text{position}, \text{TCR}) M[\text{aa}(\text{index}, \text{position}), \text{aa}(\text{peptide}, \text{position})]. \quad (21)$$

When we used pMHC binding information, we used the following TCR-specific version of [Equation \(14\)](#),

$$d(\text{peptide}, \text{index}, \text{TCR}) = \sum_{\substack{\text{position} \\ \in \{1, 2, \dots, L\}}} w(\text{position}, \text{TCR}) M[\text{aa}(\text{index}, \text{position}), \text{aa}(\text{peptide}, \text{position})] + d_{\text{MHC}}(\text{TCR}), \quad (22)$$

577 with the TCR-specific quantities $d_{\text{MHC}}(\text{TCR})$ having the same priors as d_{MHC} before.

TCR-specific peptide-to-index distances are consequently used, similar to [Equation \(2\)](#), to construct TCR-specific activation probabilities $a(\text{peptide}, \text{TCR})$,

$$\begin{aligned} \text{Prob}[a(\text{peptide}, \text{TCR}) | d(\text{peptide}, \text{index}, \text{TCR})] = & \\ \begin{cases} 1 - \text{logit}^{-1}(d_0(\text{TCR}) - d(\text{peptide}, \text{index}, \text{TCR}) - c_1), & \text{if } a_1 \\ \text{logit}^{-1}(d_0(\text{TCR}) - d(\text{peptide}, \text{index}, \text{TCR}) - c_1) - \text{logit}^{-1}(d_0(\text{TCR}) - d(\text{peptide}, \text{index}, \text{TCR}) - c_2), & \text{if } a_2 \\ \text{logit}^{-1}(d_0(\text{TCR}) - d(\text{peptide}, \text{index}, \text{TCR}) - c_2). & \text{if } a_3 \end{cases} \end{aligned} \quad (23)$$

where

$$a_1 : a(\text{peptide}, \text{TCR}) = a_{\text{non}}$$

$$a_2 : a(\text{peptide}, \text{TCR}) = a_{\text{weak}}$$

$$a_3 : a(\text{peptide}, \text{TCR}) = a_{\text{strong}}$$

578 In both within-TCR and cross-TCR cases, pooling was performed over different positions in the peptide sequence, and different
579 elements of the matrix M , corresponding to different AA substitutions. Pooling across AA substitutions allowed us to assign
580 $M[\text{aa}_1, \text{aa}_2] = D[\text{aa}_1, \text{aa}_2](1 + \mu)$ for AA substitutions absent in the training set but present in the test set.

581 **Unpooled BATMAN training**

582 We tested different parameter inference and pooling schemes for BATMAN. For all MHCII TCRs, [Figure 2E](#) ‘BATMAN
583 unpooled’ results, all [Figure 5A,C](#) (except ‘Pan-TCR’ results) TCRs, [Extended Data Figs 5](#) and [6](#), and ‘Full_unpooled’ results
584 in [Extended Data Fig 7](#), we did not pool across TCRs, i.e., BATMAN was trained individually for each TCR separately. In
585 these cases, the unpooled inferred weights had a Beta(2, 2) distribution as the prior. We first used conventional AA substitution
586 distance matrices, and performed 5-fold cross validation separately for each TCR, e.g., for 9-mer-binding TCRs using about
587 144 random peptides from the set of single-AA mutants for the TCR for training, and the remaining 36 for testing.

588 **Pooling schemes for BATMAN**

589 In [Figure 2D](#) ‘BATMAN (BLOSUM100)’ results, [Extended Data Figs 5](#) and [6](#), we did not infer the AA matrix, i.e., M was set to
590 the indicated AA distance matrix. In other cases where we inferred the matrix M , the pre-defined matrix D was always chosen
591 to be BLOSUM100, since it performed the best in classification tasks among all the conventional AA distance functions in
592 unpooled training ([Extended Data Fig 5B](#),[Extended Data Fig 6](#)). For [Extended Data Fig 7](#) *Symmetric_** results, we constrained
593 M to be symmetric, which showed slightly worse performance than when we did not impose this constraint. Thus, in all
594 other cases, we inferred the full AA matrix. Indeed, the asymmetric part of the inferred full AA matrix was prominent for
595 hydrophobic AA residues ([Figure 3E](#)). For plotting the inferred AA matrix in [Figure 3E](#), we divided the inferred matrix by
596 the corresponding values of $1 + \mu$, to make the ratios of inferred matrix elements to BLOSUM100 matrix elements more
597 interpretable.

598 When pooling across TCRs, for [Figure 2D](#) and [Figure 2E](#) ‘BATMAN pooled’ results, we pooled across the selected
599 14 9-mer-binding TCRs and 4 10-mer-binding human TCRs separately. In [Figure 2F](#), [Figure 3B](#), and [Extended Data Fig 7](#)
600 **_within* results, we pooled within TCRs specific for an index peptide. In [Figure 2D](#), and [Figure 2E](#) ‘BATMAN pooled’ results,
601 BATMAN is pooled across TCRs specific for all index peptides of same length in the benchmarking TCR set of [Figure 2A](#),
602 and across all TCRs specific for all index peptides of same length in [Extended Data Fig 7](#) **_across*. BATMAN performance
603 improved by pooling the training data across TCRs ([Figure 2E](#), [Extended Data Fig 7](#)), even when inferring TCR-specific
604 weights.

605 We used pMHC binding information in BATMAN prediction for the results in [Figure 2D](#) ‘BATMAN’, [Figure 2E,F](#), and
606 [Figure 3](#). We did not use pMHC binding information in BATMAN-AL, since LOO-TCR performance using pMHC binding had
607 a broader range across TCRs with using pMHC binding ([Figure 2D](#)). For multi-AA mutant results in [Figure 5A-B](#), since the
608 experimentally tested multi-AA mutants are biased towards MHC-binders, we did not use pMHC binding information input.

609 Finally, while in most cases we inferred TCR-specific positional weight profiles, for leave-one-TCR-out tasks (e.g.,
610 [Figure 2D,F](#), [Figure 4A,B](#), [Extended Data Fig 9](#) etc) and ‘Pan-TCR’ results of [Figure 3A](#) and [Figure 5C](#), we inferred a common
611 weight profile for all TCRs in the training set.

612 **Training schemes for BATMAN**

613 For within-TCR validation tasks, we performed 5-fold cross-validation of BATMAN. The folds were stratified by TCR activation
614 levels for classification tasks and TCR activation deciles for regression tasks, and kept identical among all methods (averaged
615 over folds) for comparison.

616 For TCRs with a sufficient number of peptide examples (≥ 5) of all 3 activation levels to perform 5-fold cross validation,
617 BATMAN classification performance was quantified in terms of 3 pairwise AUCs based on the peptide-to-index distance
618 d (peptide, index) of each mutant peptide, calculated using TCR-specific or cross-TCR positional weight profile and AA
619 distance matrix inferred by BATMAN. In such cases (e.g., [Figure 2D-F](#),[Figure 4A,B](#), [Extended Data Fig 5](#) etc) an average of
620 the 3 AUCs are plotted, whereas [Extended Data Figs 6](#) and [7](#) show individual AUCs. For some TCRs, and test folds, only 2
621 activation classes were available, where the single AUCs corresponding to the only available activation class pair are plotted.
622 For LOO-TCR tests in [Figure 2F](#) restricted among SIINFEKL-specific TCRs, we only used educated TCRs (OT1 and Ed TCRs
623 in [Extended Data Fig 2](#)) and discarded low-avidity naive TCRs, for which *pTEAM* performed worse [[17](#)].

624 **AA distance matrices in prior distribution**

To convert conventional AA substitution matrices (D' set to BLOSUM_*, PAM_*, Dayhoff, or Gonnet) into distance matrices
625 D suitable to be used in priors for BATMAN, we performed the transformation

$$D[aa_1, aa_2] = \left(1 - \frac{D'[aa_1, aa_2]}{D'[aa_1, aa_1]}\right) \left(1 - \frac{D'[aa_2, aa_1]}{D'[aa_2, aa_2]}\right), \quad (24)$$

626 so that the AA distance matrix D was always symmetric, with diagonal elements equal to zero. The Hamming matrix had all
off-diagonal elements equal to 1.

627 **Parameter inference from posterior sampling for BATMAN**

628 BATMAN is implemented in python (v. 3.11.5), using *pymc* (v. 5.10.1) packages. We sampled inferred parameters from
629 approximated posteriors using the "Automatic Differentiation Variational Inference" (ADVI) method [69], with the convergence
630 criterion being that the evidence lower bound (ELBO) loss function did not change by more than 0.1% if the number of
631 iterations was doubled. For unpooled BATMAN training on limited subsamples in [Figure 2E](#), we found that we required a
632 minimum of 30% of within-TCR training fold data for the convergence of estimated BATMAN parameters.

633 **Other TCR-pMHC interaction prediction methods**

634 **Training dataset summary of different TCR-pMHC methods**

635 We compared our benchmarking dataset with the training datasets of existing TCR-pMHC interaction prediction methods
636 ([Figure 1A](#)). We estimated (1) the total number of TCRs and pMHCS considered by each, and (2) the statistics of all
637 experimentally validated examples of TCR-pMHC interactions spanning their respective full training datasets ([Figure 1A](#)).
638 We discarded any subsampling and artificial generation of training dataset (e.g., by random pairing of pMHCS and TCRs,
639 commonly used to generate artificial negative examples). Further method-specific notes on acquisition of training dataset
640 statistics can be found in the Supplementary Notes.

641 **Implementation of different TCR-pMHC methods**

642 We tested a subset of pre-trained TCR-pMHC methods on our database. The selection was based on availability of web servers,
643 pretrained models, and ease of installing and running models locally. We trained *pTEAM* in both within-TCR and leave-
644 one-TCR-out modes. For the rest of the methods, we used available pre-trained models on our dataset. Each tested method
645 yielded a continuous-valued TCR-pMHC interaction score for each mutant-TCR pair, which was used to calculate 3 AUCs for
646 classification tasks that were subsequently averaged in the final results. The Supplementary Notes section contains links and
647 summaries of different methods tested, and more technical details on their applications on our database.

648 **Implementing *pTEAM***

649 A recent method, *pTEAM*, was specifically developed to predict TCR activation by mutants. We implemented *pTEAM* following
650 the description in its source preprint [17]. Briefly, we used Atchley embeddings for index and mutant peptides, and, for
651 leave-one-TCR-out tasks, aligned TCR sequences. These embeddings were used as inputs to random forests with 250 trees for
652 classification and regression tasks, with same folds as BATMAN. Each pairwise AUC was calculated by averaging over two
653 AUCs corresponding to 3 activation level probabilities output from the random forests. We used *R* to align TCR sequences
654 with the *muscle* (v 3.40.0) package and implement the random forests with the *randomForest* (v 4.7-1.1) package. AUCs were
655 calculated using the *multiclass.roc* function from the *pROC* (v. 1.18.4) package in *R*.

656 **Active Learning**

657 **Active Learning with BATMAN**

658 To initialize parameters, we first trained BATMAN-AL in the leave-one-TCR-out mode, and learned pan-TCR positional
659 weights ($w_{\text{pan-TCR}}$) and the AA substitution distance matrix (M). When applying active learning to a novel TCR, we only modify
660 the TCR-specific positional weights. For all candidate peptides, we used Eqn. (7) to calculate peptide-to-index distances.

661 Our strategy to select peptides for training was based on a combination of two criteria: (a) picking peptides near decision
662 boundaries so that classifiers can finely delineate classes, and (b) picking diverse peptides so as to cover different regions of
663 peptide space. For (a), if all peptide candidates (single-AA mutants) were sorted based on their peptide-to-index distance, then
664 assuming balanced data, we reasoned that class boundaries would fall at roughly the median distance (for 2-class data) or the
665 two tertile distances (1/3, 2/3 for 3-class data). For (b), we explored diversity in the positions of the selected peptides, as well as
666 diversity in the distances themselves.

667 Using these intuitions, our BATMAN-AL strategy was as follows. For each position, we first selected one mutant peptide.
668 For roughly half (5 out of 9) the positions — those with the smallest pan-TCR positions weights — we selected the peptide
669 having the peptide-to-index distance closest to the median peptide-to-index distance. For the remaining (4 out of 9) positions,
670 we selected the peptide closest to the first tertile of the peptide-to-index distances ([Extended Data Fig 9A](#)). To avoid sampling
671 redundant peptides, when selecting more than one peptide per mutation position, we selected the additional peptides randomly.

To infer positional weights for an unseen TCR, we trained BATMAN in within-TCR mode with the sampled peptides and
the index peptide as before, except now the priors for the positional weights (w_1 , the positional weight vector for the 1st AL
round) are informed by the pan-TCR weights as follows,

$$w_1(\text{position}) \sim \text{Normal}(w_{\text{pan-TCR}}(\text{position}), \sigma(\text{position})), \quad (25)$$

with

$$\sigma(\text{position}) \stackrel{iid}{\sim} \text{Exponential}(1). \quad (26)$$

For the next AL round, we sample from the remaining peptides, calculating peptide-to-index distances using the pan-TCR AA matrix and the learned positional weights from the last AL round as a prior:

$$w_i(\text{position}) \sim \text{Normal}(w_{i-1}(\text{position}), \sigma(\text{position})). \quad (27)$$

After each AL round, we normalized the learned weights to [0,1], and calculated AUC for the remaining unsampled peptides.

As mentioned above, when selecting more than one peptide per position, our strategy selects the additional peptides randomly. We compared the performance of 4 strategies for selecting random peptides:

1. Peptides are not selected randomly; i.e., peptides are chosen according to their distance from the median or the first tertile of the peptide-to-index distances, depending on the positional weights (i.e., the AL strategy introduced above).
2. All peptides are chosen randomly in each round, while still sampling one peptide per position (“random learning”)
3. All peptides are chosen randomly in even rounds, and in odd rounds, peptides are chosen as per strategy (1) above.
4. All peptides are chosen randomly after the first round, with peptides chosen according to strategy (1) in the first round

For benchmarking performance of AL strategies, among our benchmarking TCR set of [Figure 2A](#), we selected 9-mer-binding TCRs with all positions mutagenized. In all cases involving random peptide selection, we performed AL for 100 realizations of random peptide choice and parameter inference, and averaged the results. The results ([Extended Data Fig 9B](#)) showed that out of the 4 strategies, the 3rd strategy exhibited a monotonic increase in median AUC with the number of peptides sampled, and the smallest variance of AUCs, across different TCRs ([Extended Data Fig 9B](#)). Thus, we plotted the results corresponding to this strategy in the main text ([Figure 4A,B](#)). These results also showed the importance of random peptide choice in AL algorithms. For example, AL without any random peptide sampling did not exhibit monotonic performance increase with more experimental rounds ([Extended Data Fig 9B](#)).

Assessing the relationship between inferred peptide weights with TCR-pMHC structure

For 6 TCRs in the BATCAVE dataset with full peptide mutational scan, there were corresponding PDB entries with the TCRs in complex with either (1) their respective index peptides in the BATCAVE database (TCRs 47BE7, 1E6, and TIL1383I), (2) a single-AA mutant of their index peptide (TCRs NYE-S with BATCAVE index peptide being SLLMWITQC and structure present for SLLMWITQV), (3) the TCR-binding epitope which is embedded in the longer BATCAVE index peptide (TCRs F5 and F24 with BATCAVE index peptide being FRDYVDRFYKTLRAEQASQE and structure present for RFYKTLRAEQASQ). Since these were the only experimental TCR-pMHC PDB structures available for BATCAVE TCRs, notwithstanding the differences between BATCAVE index peptides and corresponding PDB peptides, we correlated BATMAN positional weights of these TCRs with corresponding structural features with TCR-pMHC interactions.

To compare weight profiles across TCRs, we normalized the positional weights to their TCR-specific maximum. For TCRs F5 and F24, while BATMAN inferred positional weights for all 20 positions of FRDYVDRFYKTLRAEQASQE mutagenized in BATCAVE, PDB structures contained only peptide residues R7-Q19, and thus we restricted our analyses to those for structural analysis.

To correlate the BATMAN inferred positional weights with TCR-pMHC interaction, we analyzed the polar and nonpolar interactions of each peptide residue in the complex with TCR for TCRs with available structural data in the Protein Data Bank (PDB), listed here by TCR name: PDB identifier : NYE-S1:6rbp, NYE-S2: 6rpa, NYE-S3: 6rp9, 47BE7: 7na5, TCR-1E6: 3utt, TIL1383I: 7rk7, TCR-F5: 6cq, and F24: 6cql. All analyses were performed with UCSF ChimeraX [[70](#)] (version 1.8). To identify hydrogen bonding interactions between residues of the peptide to neighboring residues in either TCR, MHC or peptide itself, we used the ‘hbonds’ function which uses atom types and geometric criteria to identify hydrogen bonds. For this, we allowed a distance tolerance of 0.4 Å and an angle tolerance of 20°. To identify all polar and nonpolar interactions involving each of the residues of the peptide to neighboring residues in either TCR, MHC or peptide itself we used the ‘contacts’ function. For this, we allowed a center-center distance of 4 Å. Threshold parameters for interactions represent default cut-offs for the range of the respective interaction [[71](#)]. For each pMHC-TCR structure, we then summed the number of polar and non-polar interactions at each peptide position and fit a linear model with interactions as the independent and BATMAN-inferred peptide positional weight as dependent variable.

Data availability

The fully curated database of TCR-pMHC interactions can be downloaded from https://github.com/meyer-lab-cshl/BATMAN/tree/main/TCR_epitope_database.

716 **Code availability**

717 Custom analysis code was written in python (version $\geq 3.10.11$) or R (version $\geq 3.4.0$). All analyses and code to re-produce
718 figures in this manuscript are available at: <https://github.com/meyer-lab-cshl/BATMAN-paper>. The python implementation of
719 BATMAN ('pyBATMAN') can be installed from <https://pypi.org/project/pybatman/> and run locally. pyBATMAN installation
720 instructions and input file specifications can be found at <https://github.com/meyer-lab-cshl/BATMAN/>. Example TCR-pMHC
721 input dataset and python script for running pyBATMAN can be found at https://github.com/meyer-lab-cshl/BATMAN/tree/main/run_batman. Interactive Jupyter notebook tutorials on pyBATMAN and BATMAN-AL usage can be downloaded
722 from https://github.com/meyer-lab-cshl/BATMAN/blob/main/run_batman/pyBATMAN_Tutorial.ipynb and https://github.com/meyer-lab-cshl/BATMAN/blob/main/run_batman/BATMAN_AL.ipynb respectively.
723
724

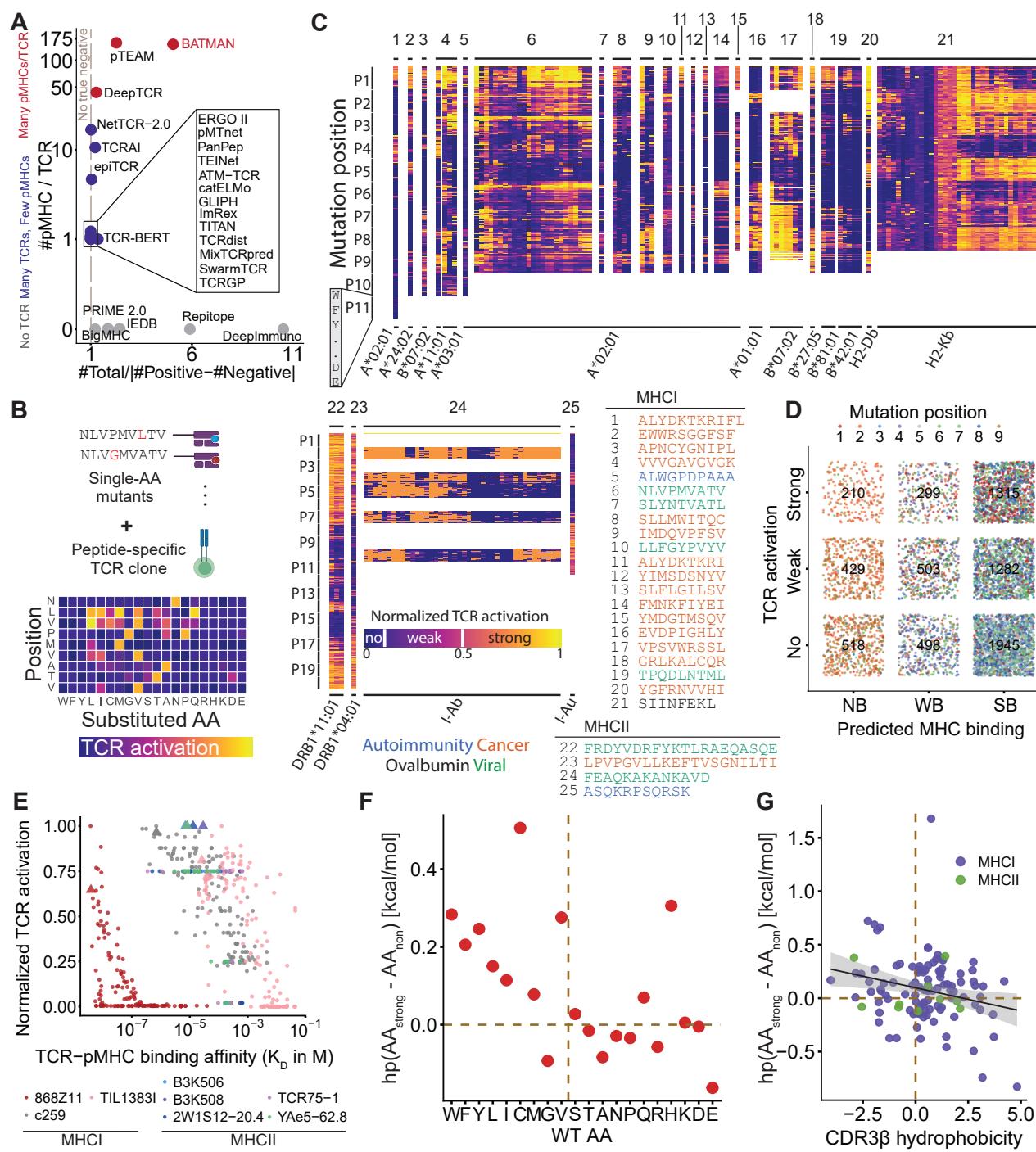


Figure 1. The BATCAVE database reveals important insights into TCR-pMHC interactions **A.** Training dataset summary metrics for the number of pMHCs tested per TCR and the class balance of experimentally verified TCR-pMHC interactions, shown for selected TCR-pMHC interaction prediction methods. **B.** Schematic of a mutational scan assay, reporting activation of a TCR clone against all single-AA mutants of its index (here, NLVPMVATV). **C.** Curated BATCAVE mutational scan database for TCR activation, with each column corresponding to a TCR clone, grouped by their index peptide (indicated above each column) and recognized MHC (below), and each row corresponding to the substituted AA at a specific position, ordered according to their interfacial hydrophobicity. **D.** Dependency of mutant TCR activation on NetMHCPan-predicted pMHC binding, shown for all HLA-A*02:01-restricted 9-mer binding TCRs in BATCAVE, with the number of mutants (single dot) in each of the nine sub categories (squares) indicated. **E.** Dependency of TCR activation on TCR-pMHC binding, with dots representing TCR-pMHC pairs and TCRs respectively (index peptides indicated by triangles). Hydrophobicity difference between allowed and disallowed AA substitutions depend on WT peptide AA (**F**, dots representing average over all occurrences of a WT AA) and CDR3 β hydrophobicity (**G**, dots representing TCRs).

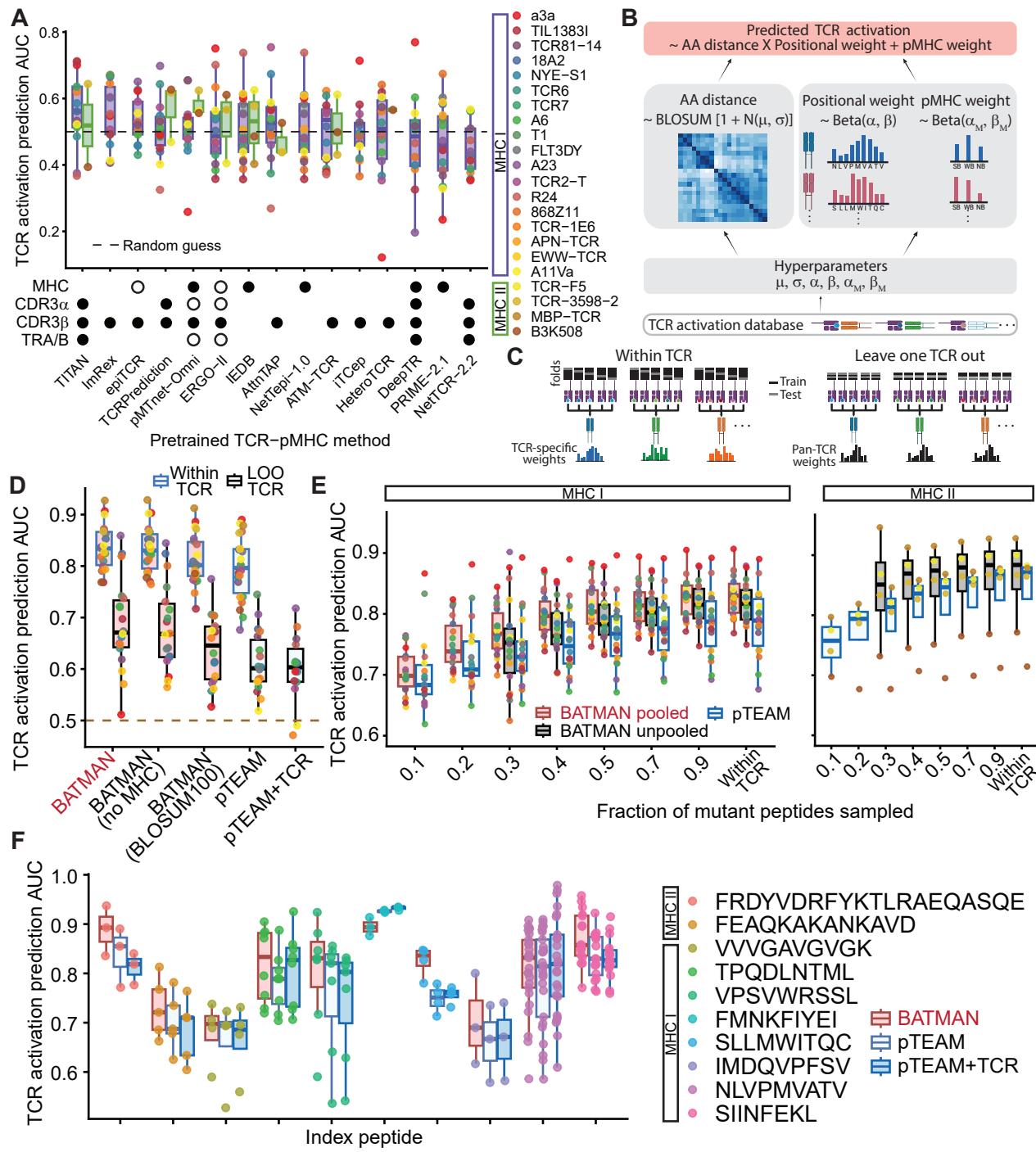


Figure 2. BATMAN outperforms existing TCR-pMHC interaction prediction methods. A TCR activation classification area under the curve (AUC) scores for different pre-trained methods, with their respective requirements indicated by dot matrix (filled: mandatory, open: optional, used if available in database); MHC: name of peptide-binding MHC allele; CDR3 α/β : sequence of the TCR CDR3 α/β chain; TRA/B: gene name or sequence of at least one of the TRAB, TRAJ, TRBV, and TRBJ gene name or sequence. Schematics for **B** BATMAN TCR activation prediction method and **C** two validation modes, within-TCR and LOO-TCR. **D**. BATMAN outperforms pTEAM in both within-TCR and LOO-TCR modes. **E**. BATMAN outperforms pTEAM for different numbers of training samples, showing improvement with pooling across TCRs (TCR color scheme same as in A). **F** BATMAN and pTEAM LOO-TCR performance comparison among TCRs specific for a common index peptide.

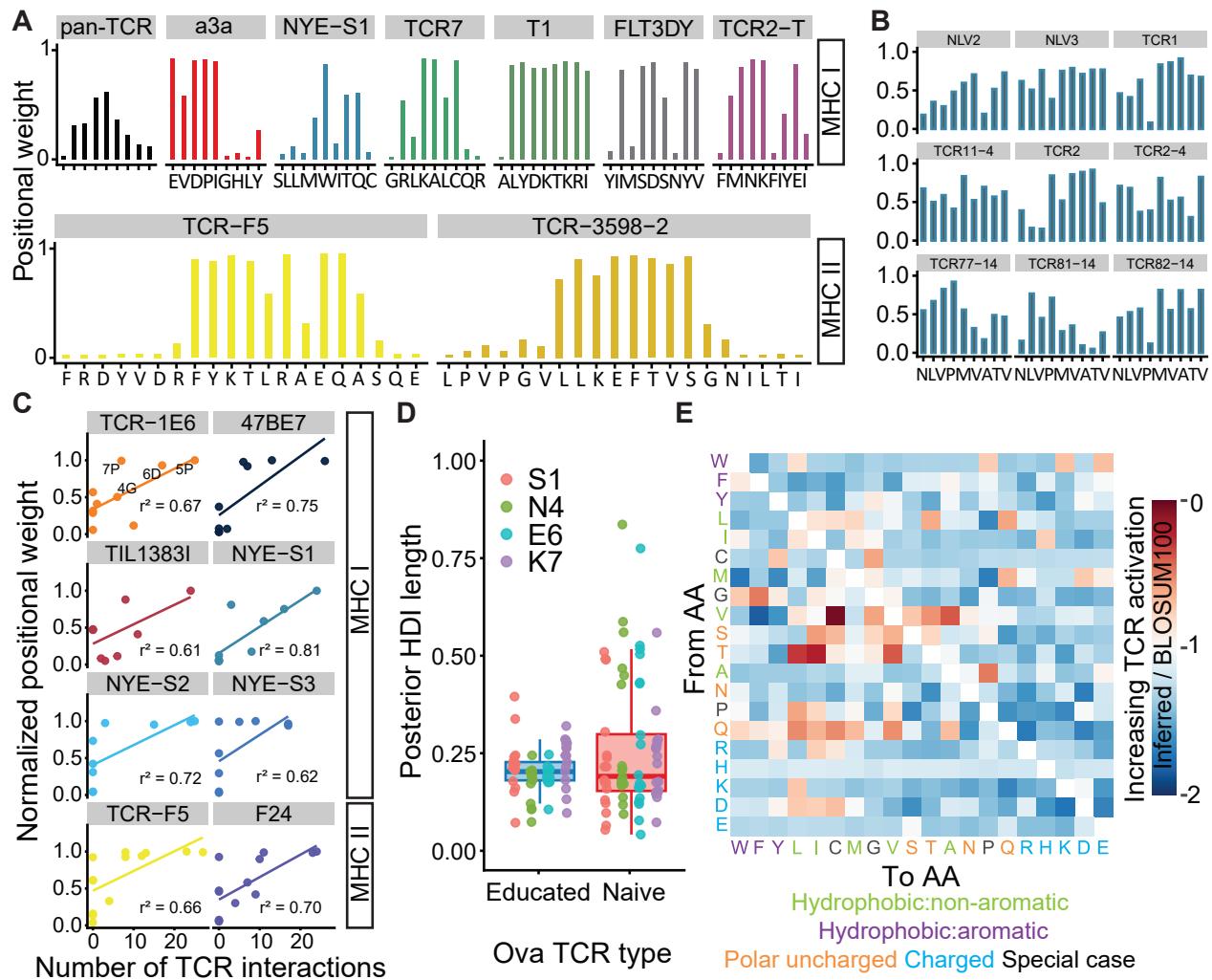


Figure 3. Inferred BATMAN parameters capture interpretable biochemical features of TCR-pMHC interactions. **A.** pan-TCR positional weights inferred from all 9-mer-binding TCRs in benchmark dataset (first panel), and examples of TCR-specific positional weights of selected TCRs with different index peptides (remainder of panels). **B.** TCR-specific positional weights of nine different NLV-specific TCRs, showing important peptide residues for TCR binding. **C.** Inferred positional weights correlate the number of interactions of peptide with TCR chains (derived from TCR-pMHC structure). To make different weight profiles comparable across TCR panels, we normalized positional weights by their maximum over all positions for each weight profile. 4 TCR-binding peptide residues are highlighted for TCR-1E6. **D.** 90% high-density intervals (HDIs) of positional weight posterior distributions for non-MHC-anchor positions are wider for low-avidity ‘naive’ SIINFEKL-specific TCRs than SIINFEKL-expanded ‘educated’ ones. **E.** Example ratio of inferred matrix elements to BLOSUM100, the best performing conventional AA distance matrix, for within-TCR classification.

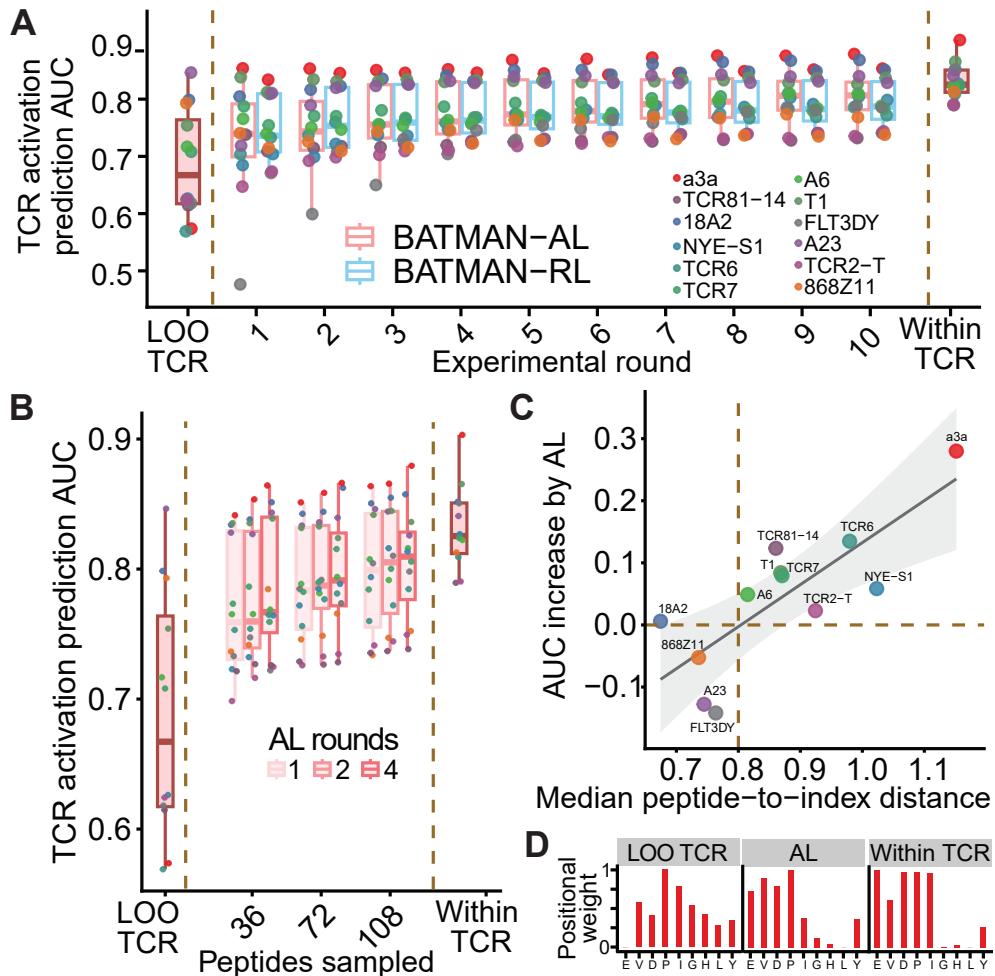


Figure 4. Active learning (AL) with BATMAN predicts TCR-specific activation with limited training samples. **A.** Average AUCs for BATMAN with active (AL) and random learning (RL). At each learning step (Experimental round, x-axis), 9 peptides are sampled. **B.** BATMAN-AL performance when the same total number of peptides are sampled over different numbers of AL rounds. For performance comparison in panels **A** and **B**, leave-one-TCR-out and within-TCR AUCs are shown. **C.** Larger median peptide-to-index distances are associated with higher AL performance boost with BATMAN. **D.** After only one AL round with 9 peptides, the positional weight profile of the a3a TCR approximates the within-TCR weight profile.

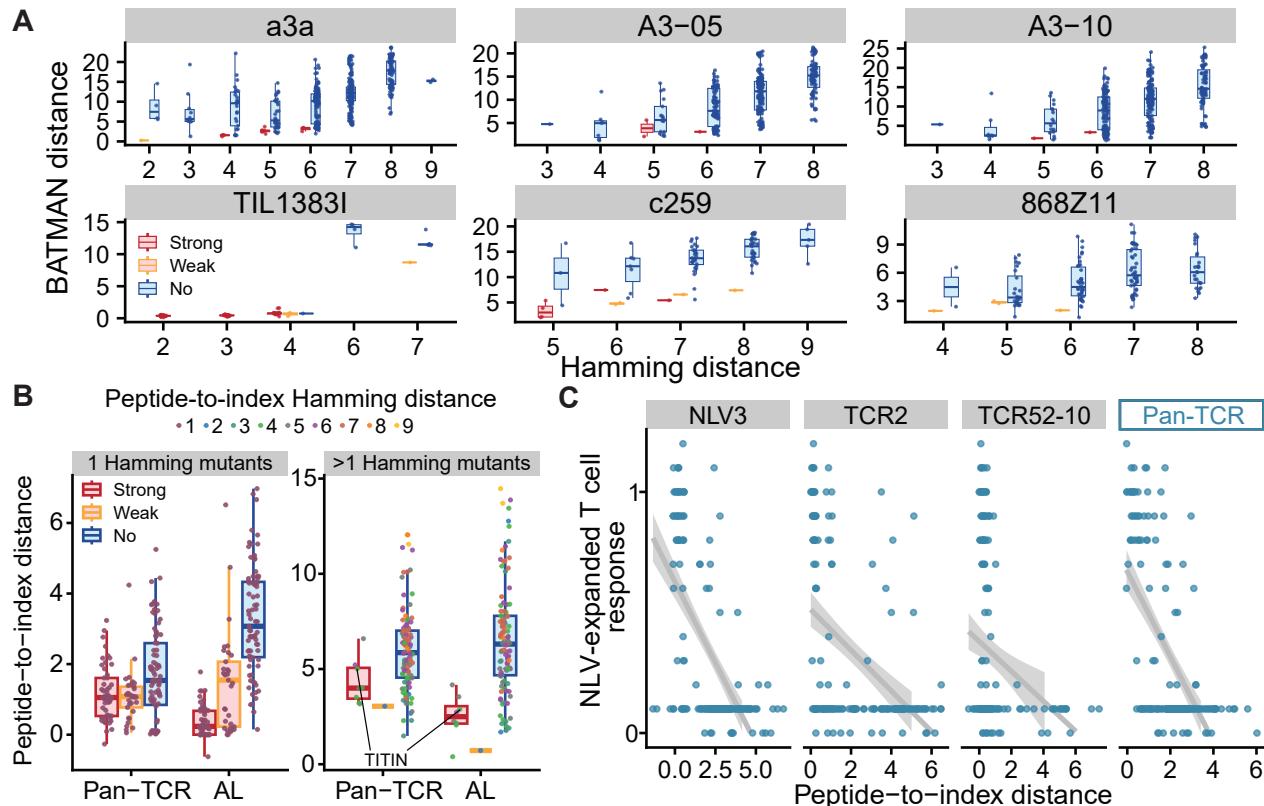
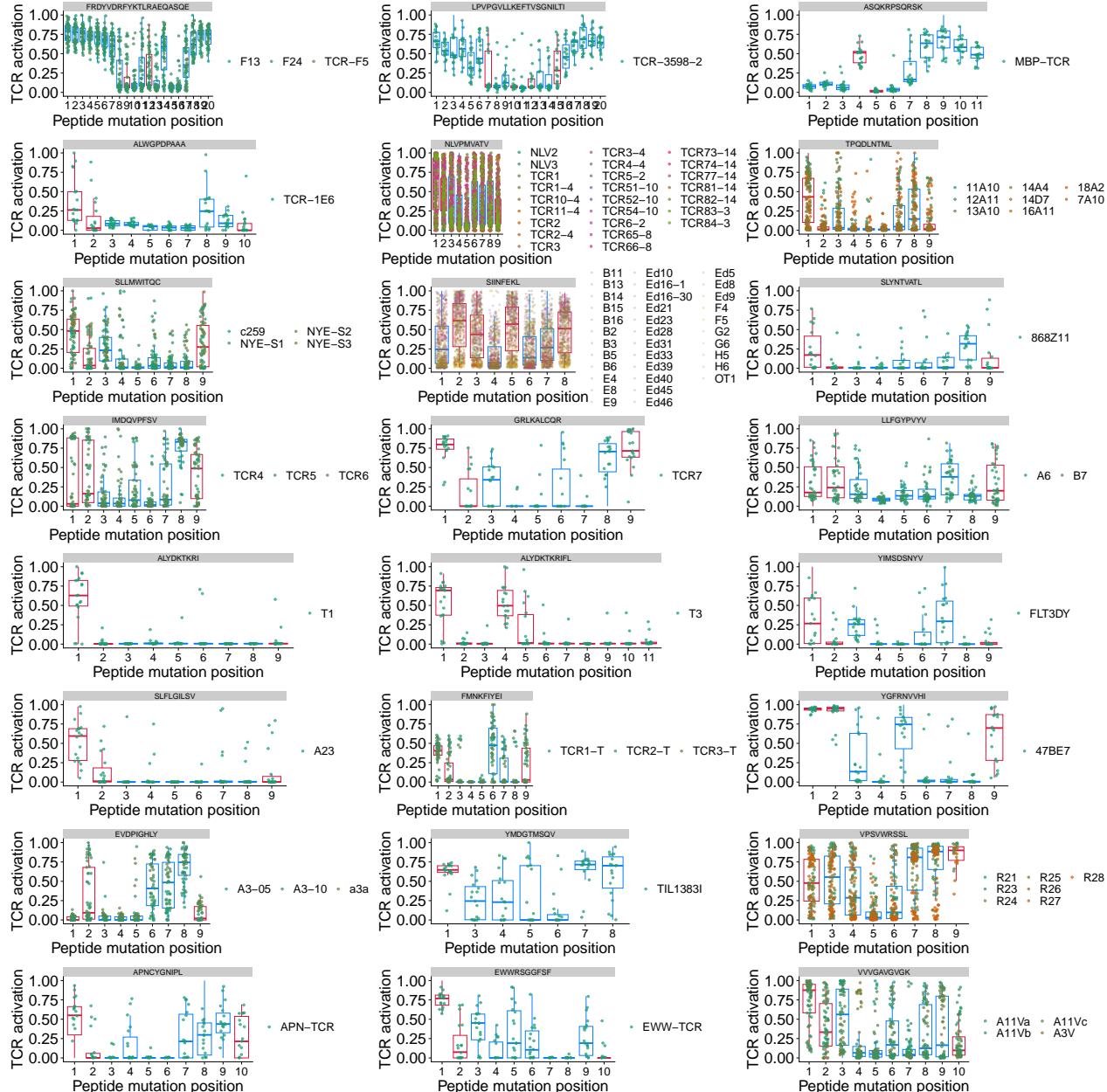


Figure 5. BATMAN predictions generalize to multiple-AA mutations and polyclonal T cell response. A. BATMAN peptide-to-index distances from single-AA mutant training separates multi-AA mutants according to their TCR activation. **B.** BATMAN-AL with only 9 peptides for the a3a TCR allows improved TCR activation prediction of both 1 Hamming and >1 Hamming mutant peptides. **C.** NLV-expanded polyclonal T cell repertoire response against NLV mutant peptides correlates with the response of individual NLV-specific TCRs to various degrees (shown are results from 3 random example TCRs), whereas Pan-TCR response, predicted from pooling across all 25 NLV-specific BATCAVE TCRs, correlating more strongly.

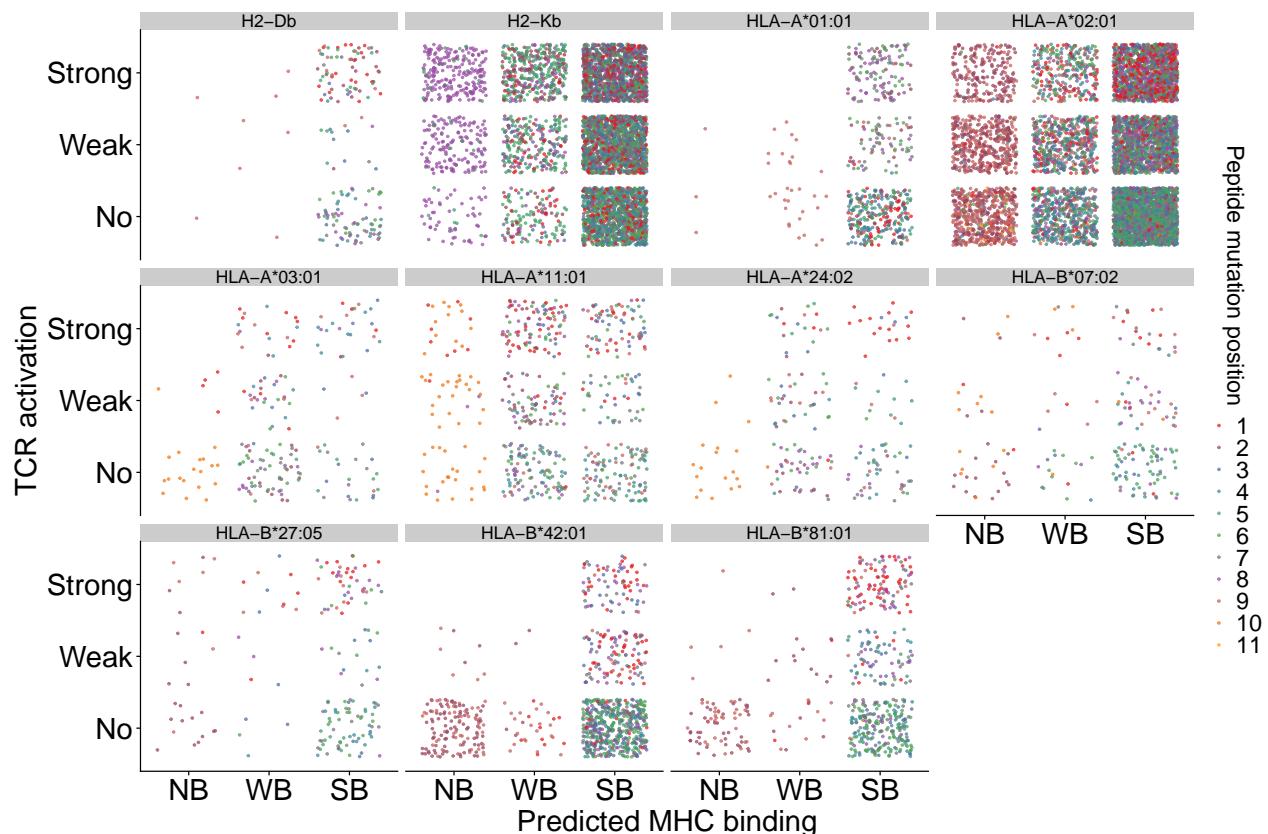


■ charged ■ hydrophobic aromatic ■ hydrophobic non-aromatic ■ polar uncharged ■ special case

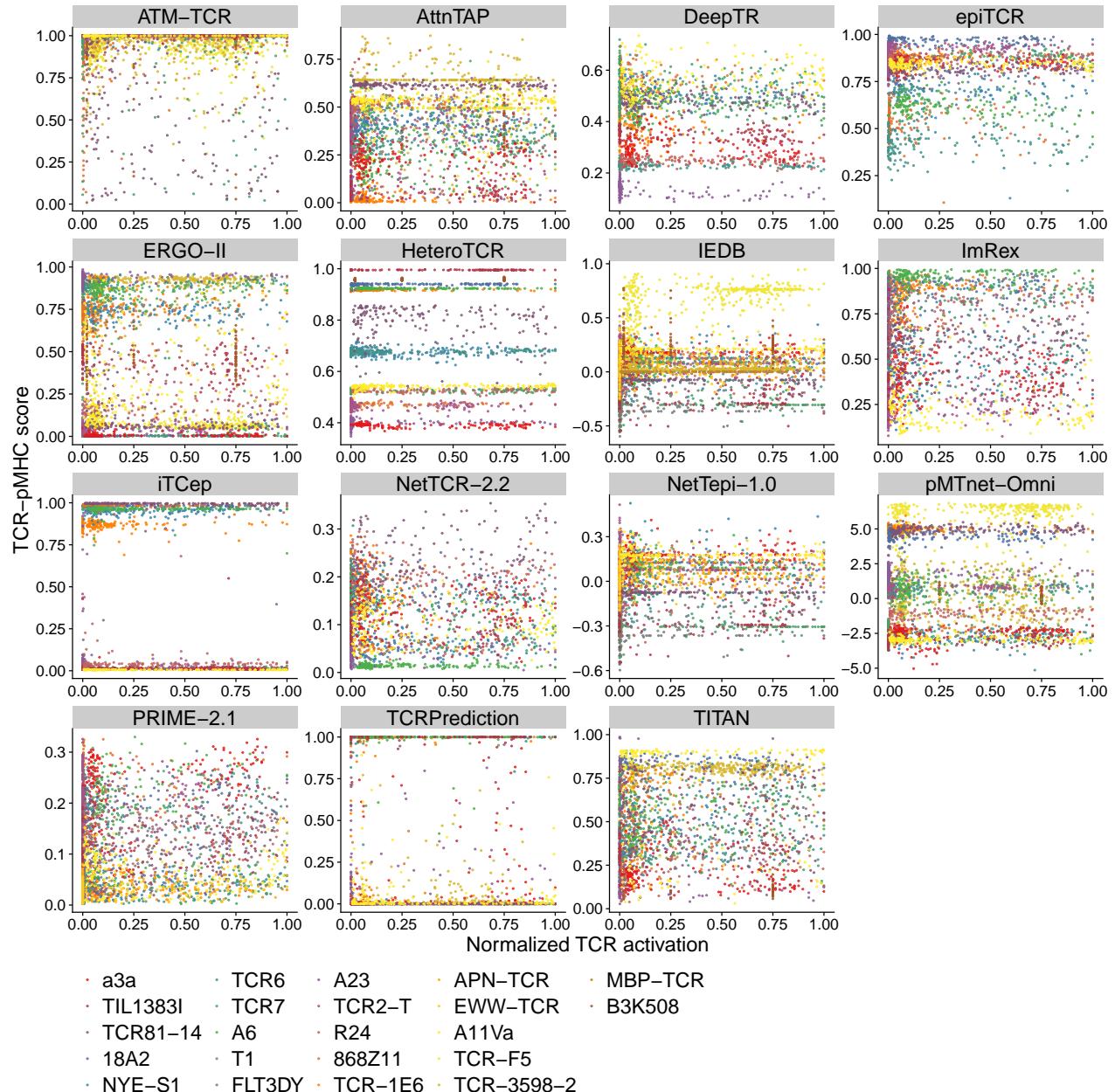
Extended Data Figure 1. BATCAVE TCR CDR3 α and CDR3 β sequence diversity, grouped by index peptides.



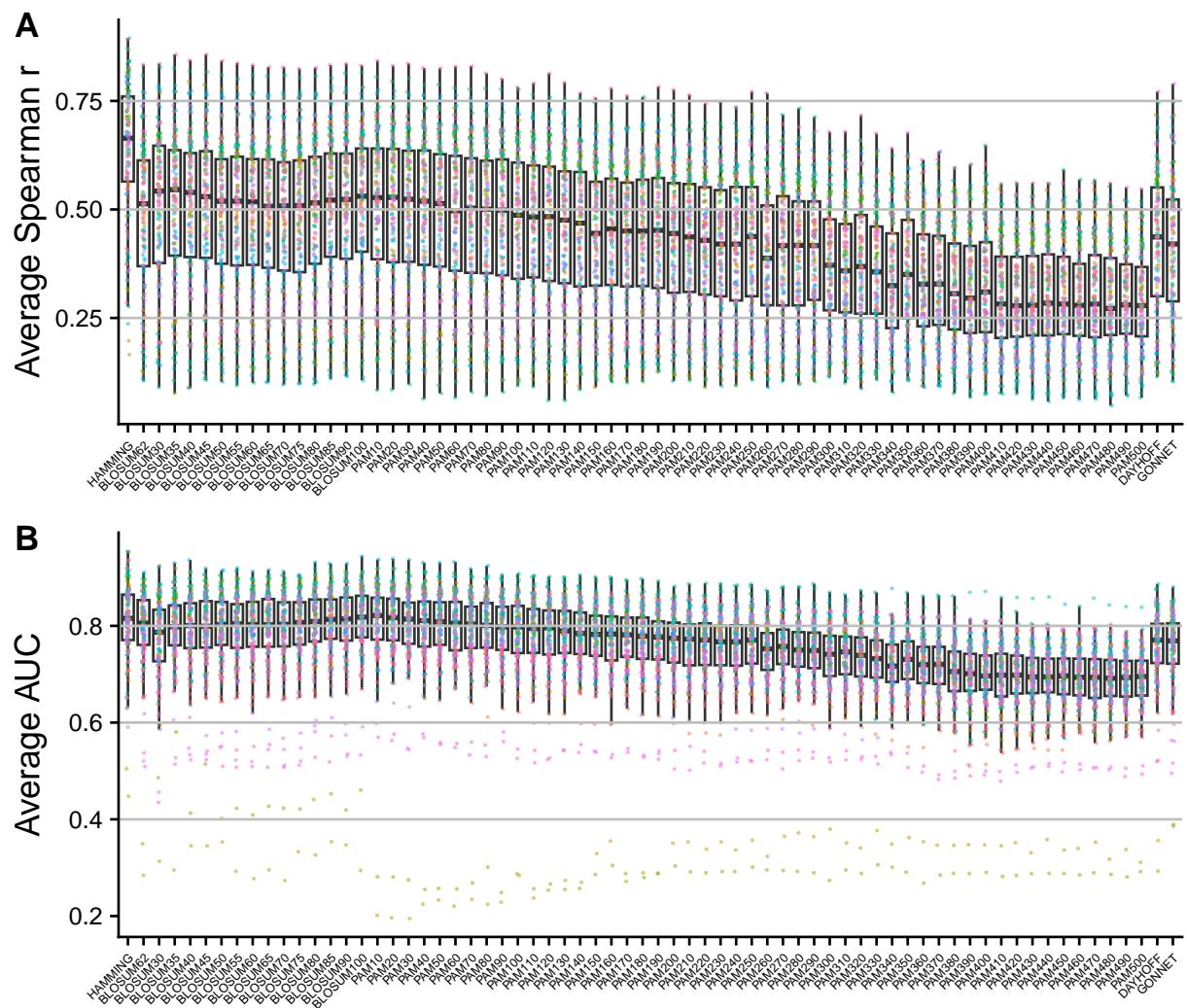
Extended Data Figure 2. Normalized TCR activation by mutant peptides, grouped by index peptides and mutation position (MHC anchor residues marked red).



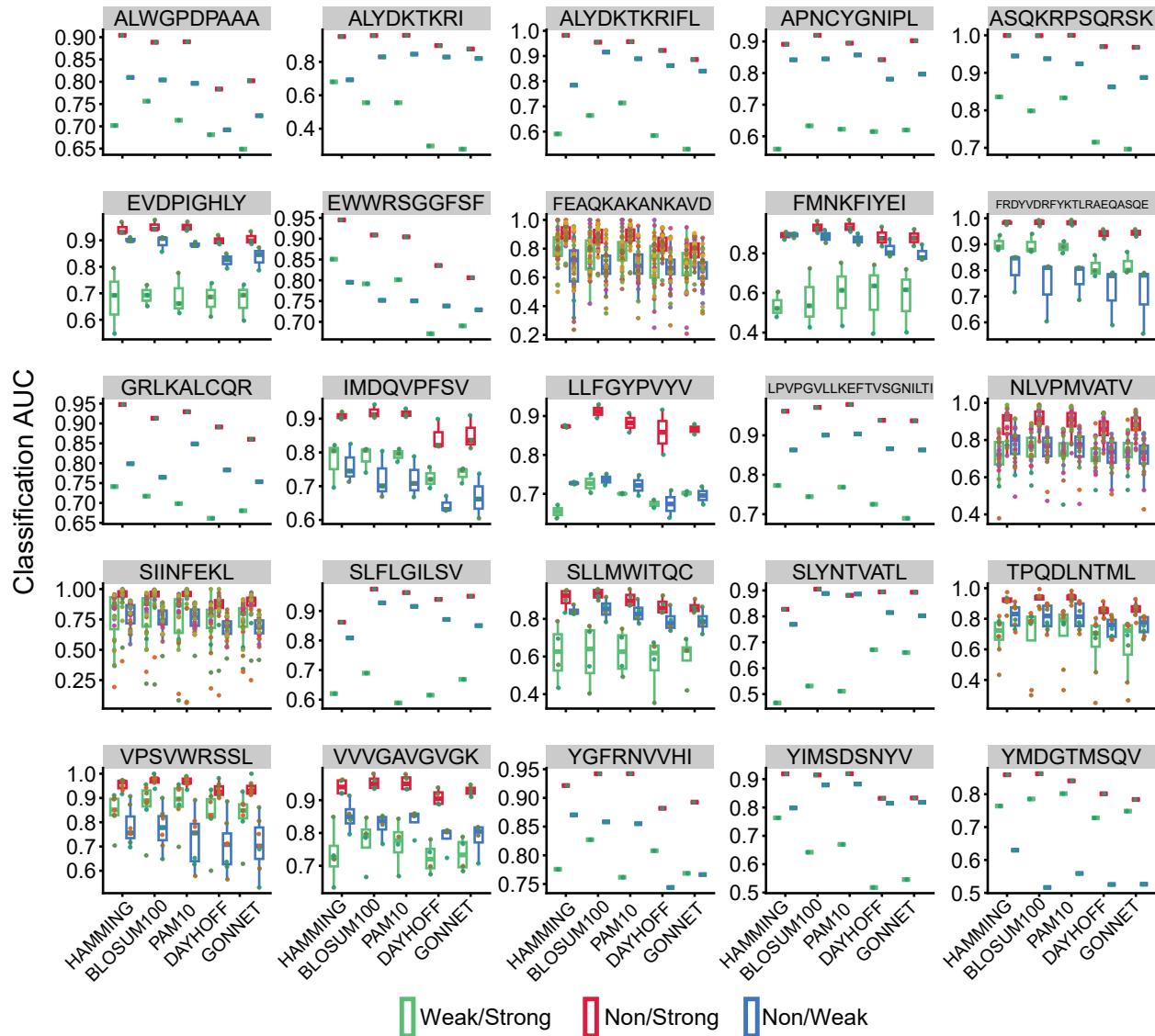
Extended Data Figure 3. Relationship between pMHC binding and TCR activation in BATCAVE database. Dependency of BATCAVE mutant peptide TCR activation category on NetMHCPan-predicted pMHC binding, grouped by MHCs, for MHCI peptides for which all positions were mutagenized.



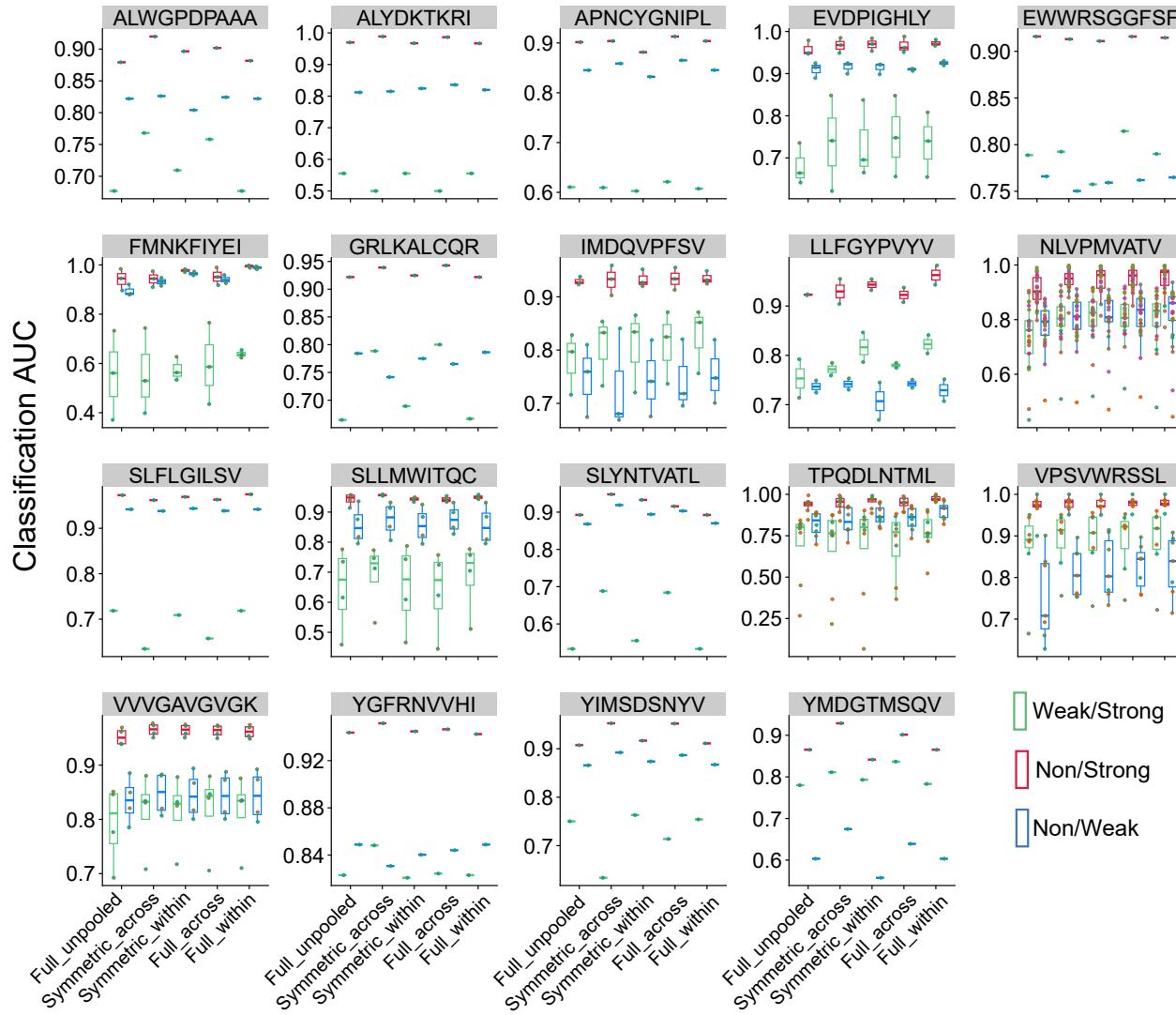
Extended Data Figure 4. TCR-pMHC scores from different methods do not correlate with TCR activation by mutant peptides
TCR-pMHC interaction scores and normalized TCR activation of mutant peptides for the TCRs selected in Figure 2a.



Extended Data Figure 5. Extended, unpooled performance analyses for BATMAN. (A) Classification and (B) regression performances in within-TCR tests without cross-TCR pooling using different amino acid distance matrices (points colored by TCRs)



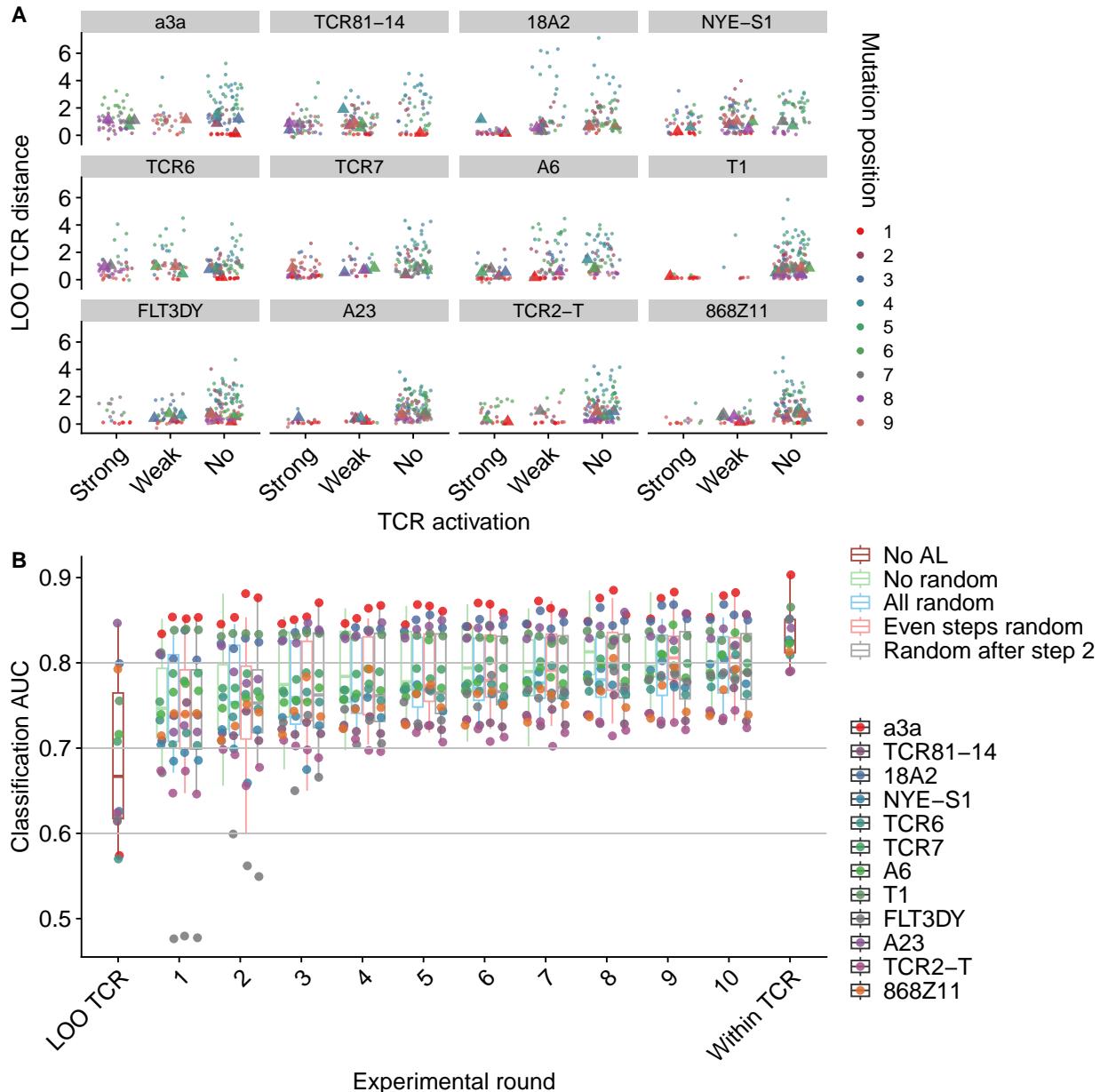
Extended Data Figure 6. Extended, unpooled performance analyses for BATMAN with selected AA substitution matrices. Pairwise classification AUC for selected amino acid distances for results plotted in [Extended Data Fig 5B](#) (points colored by TCRs with the same color scheme as [Extended Data Fig 2](#), grouped by index peptides).



Extended Data Figure 7. Pooling across TCRs improves within-TCR classification performance. Pairwise classification area under the curve (AUC) for BATCAVE TCRs with 9 and 10 AA long index peptides, with different inferred AA matrices (*Symmetric_**, and *Full_**), and pooling modes (**_within* TCRs specific for a index peptide and **_across* TCRs specific for all index peptides of same length). Unpooled results with inferred Full AA matrices shown for comparison (points colored by TCRs with the same color scheme as [Extended Data Fig 2](#), grouped by index peptides).



Extended Data Figure 8. Analyses of inferred positional weights. Positional weights are consistent across different conventional and BATMAN-inferred full and symmetric amino acid distance matrices.



Extended Data Figure 9. Peptide-to-index distances for mutant peptides are associated with active learning performance boost with BATMAN. A BATMAN peptide-to-index distances for single-AA mutant peptides corresponding to individual left out TCRs, using pan-TCR positional weight profile and AA matrix inferred from all other TCRs. The points are separated by TCR activation levels of the peptides, and colored by mutation position. Triangles correspond to the 9 mutant peptides chosen in the first AL round. **B** BATMAN performance for different AL strategies, each sampling 9 peptides per experimental round. For performance comparison, LOO TCR and Within TCR AUCs from Figure 4a are shown.