nature biotechnology

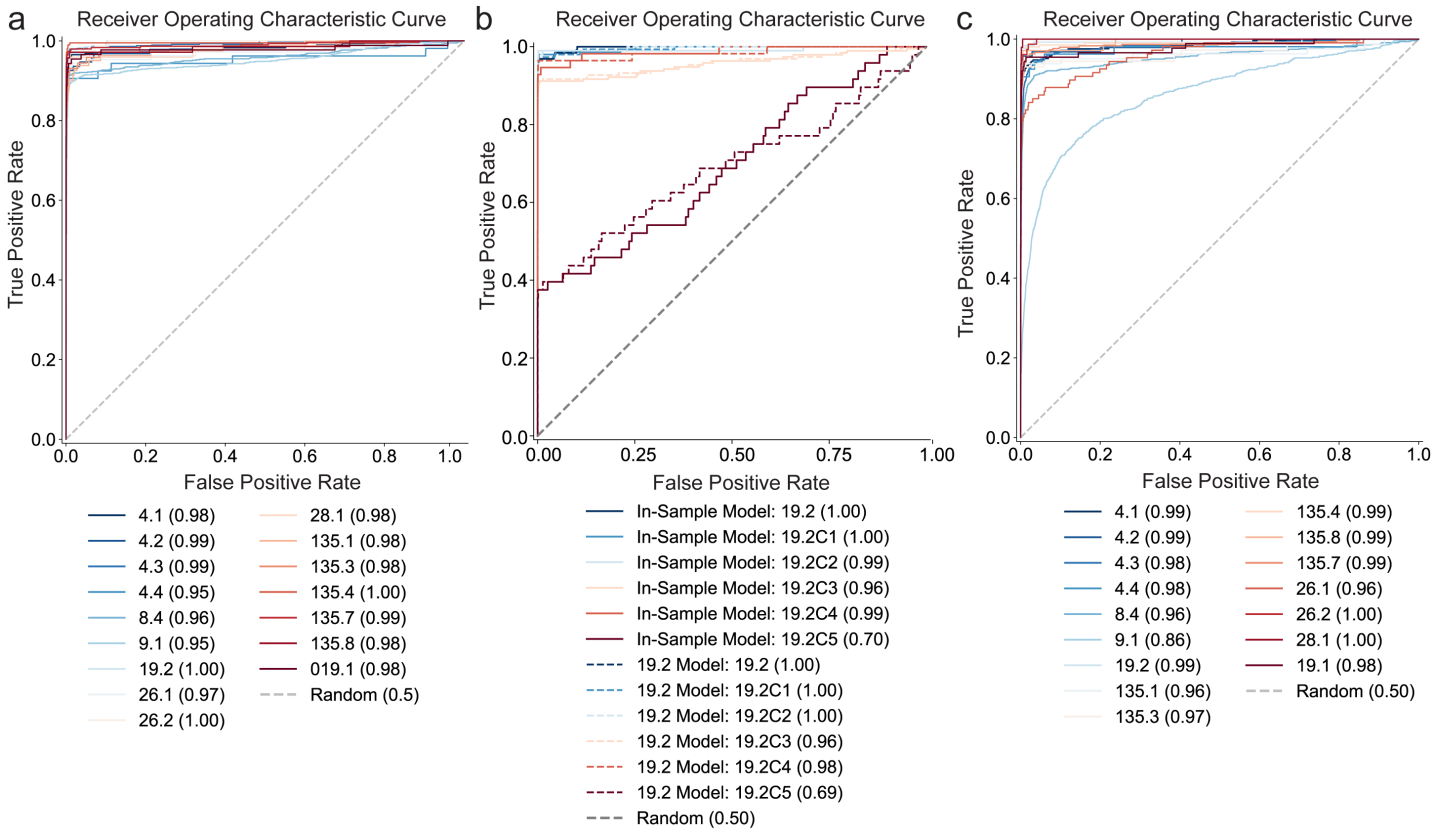
Supplementary information

https://doi.org/10.1038/s41587-026-03128-x

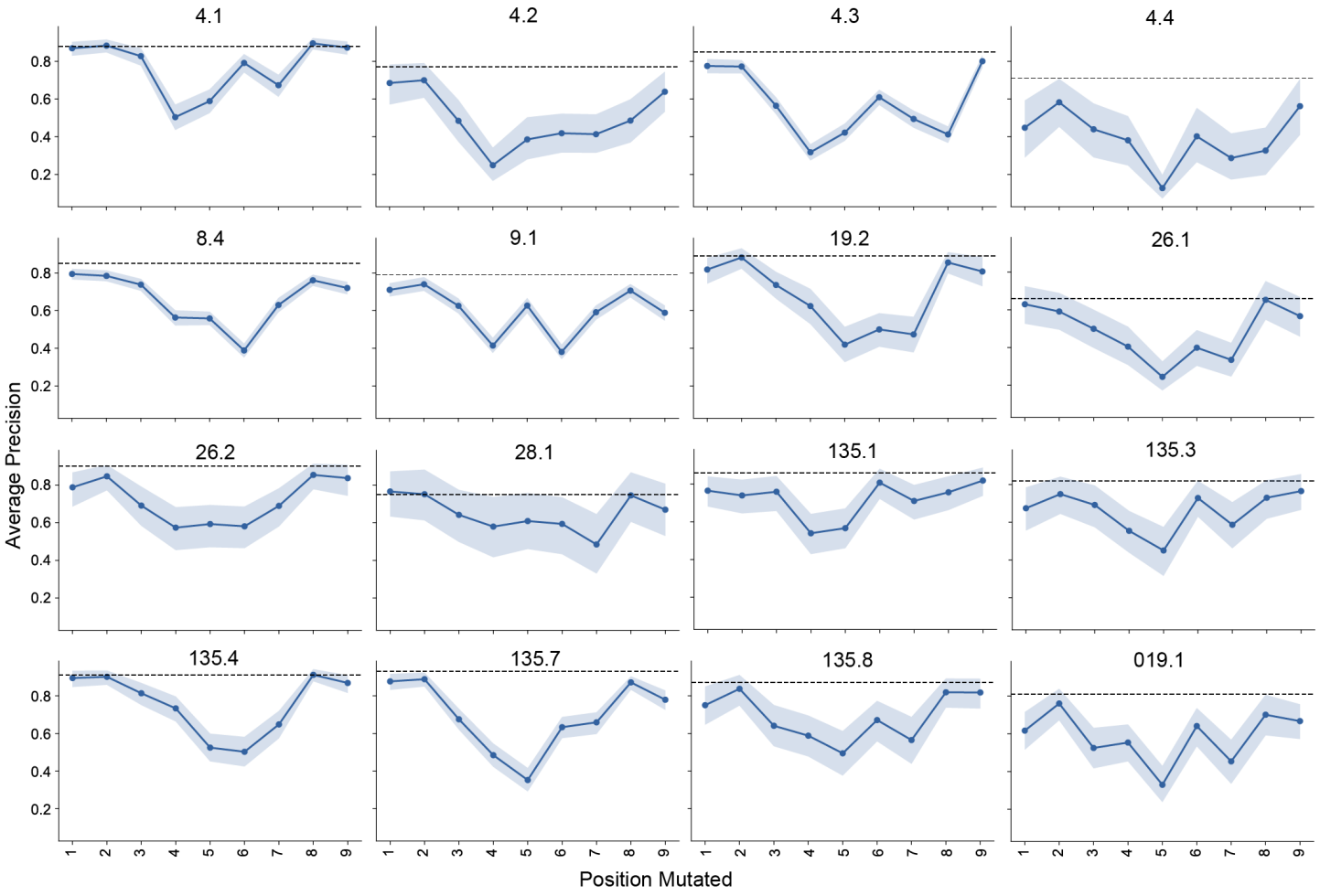
Deep peptide recognition profiling decodes TCR specificity and enables disease-associated antigen discovery

In the format provided by the authors and unedited

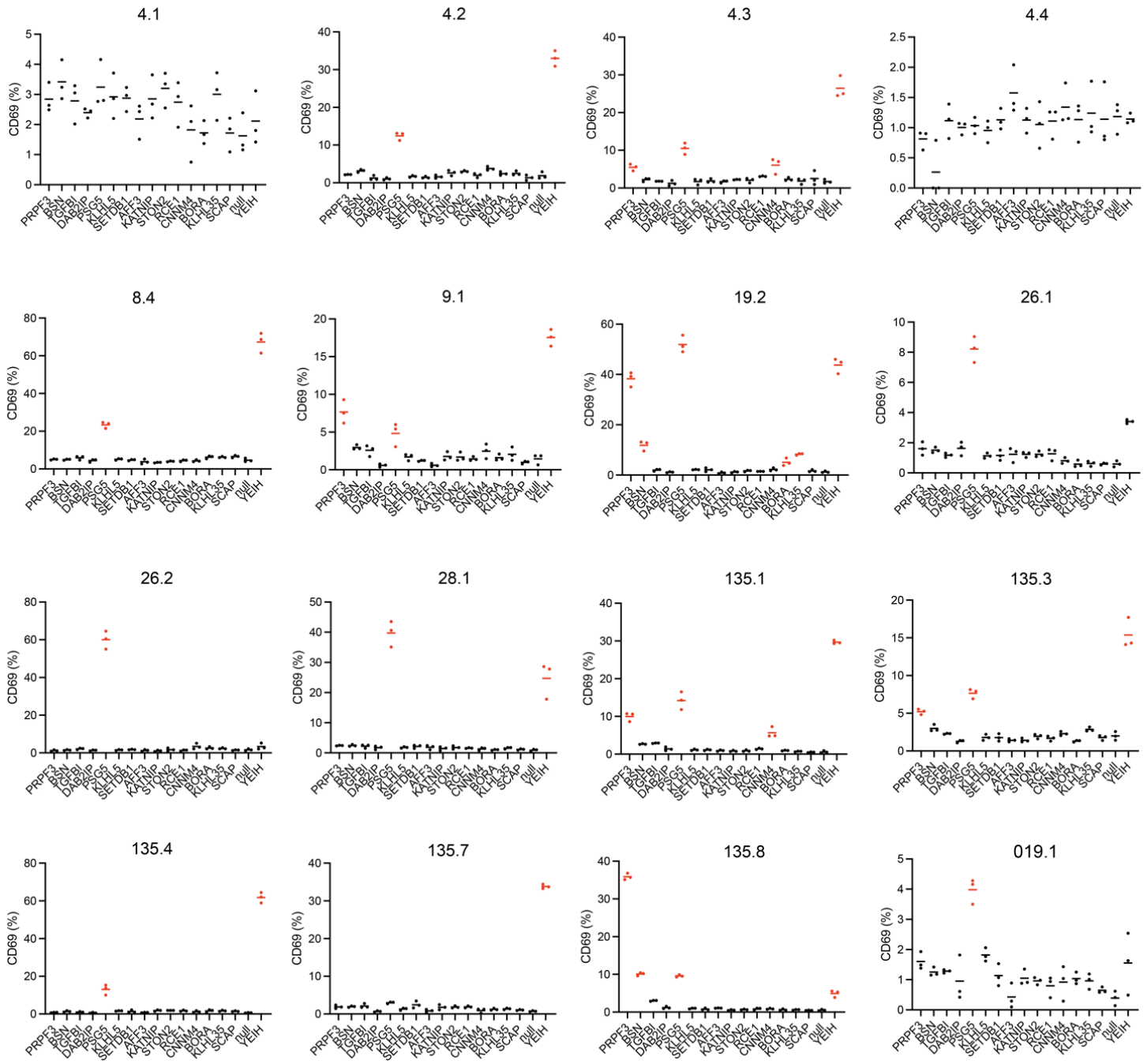
Supporting Information



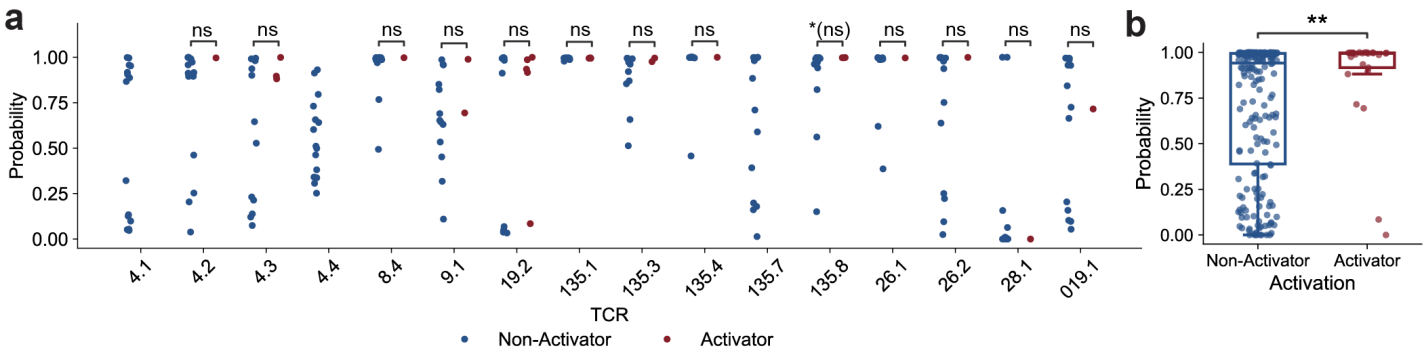
Supplementary Figure 1. Model performance and validation of peptide recognition prediction.
a, Receiver operating characteristic (ROC) curves for individual β-chain-specific models trained on peptide recognition profiling (PRP) data.
b, ROC comparison of models trained individually on each variant (in-sample), while dashed lines show predictions using the model trained on the parental TCR 19.2.
c, Leave-one-TCR-out cross-validation performance across all TCRs, assessed by ROC curves for peptide binding prediction. The dashed diagonal line represents random performance (AUC = 0.5).



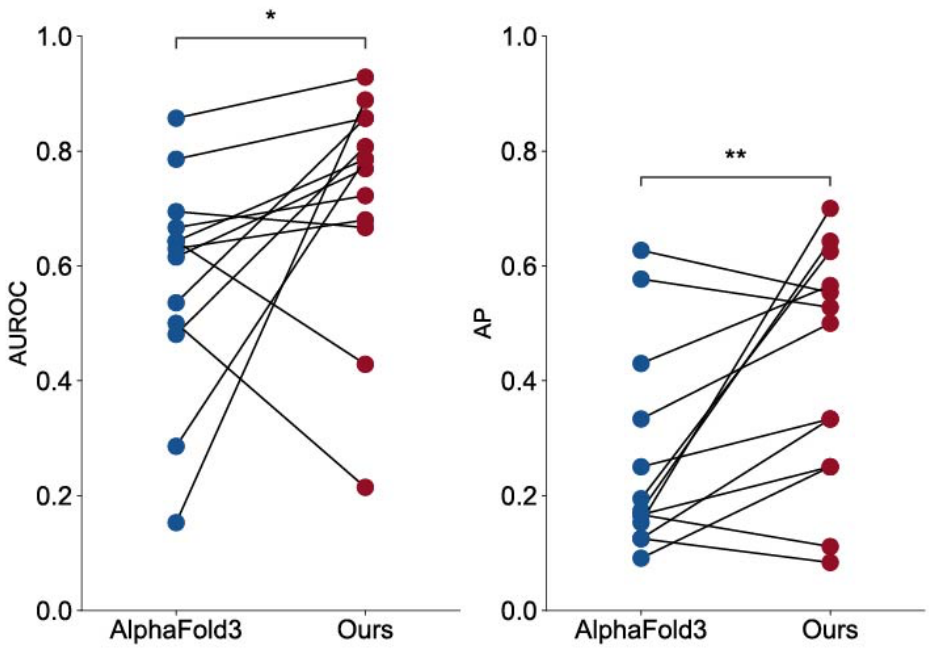
Supplementary Figure 2. Effect of position-wise amino acid substitution on test set performance. Average precision (AP) on the test set following random substitution of amino acids at each peptide position. Each point represents the AP obtained when a given position is perturbed, and shading indicates the 95% confidence interval estimated by bootstrapping. The dashed line denotes the AP of the unperturbed (original) test set.



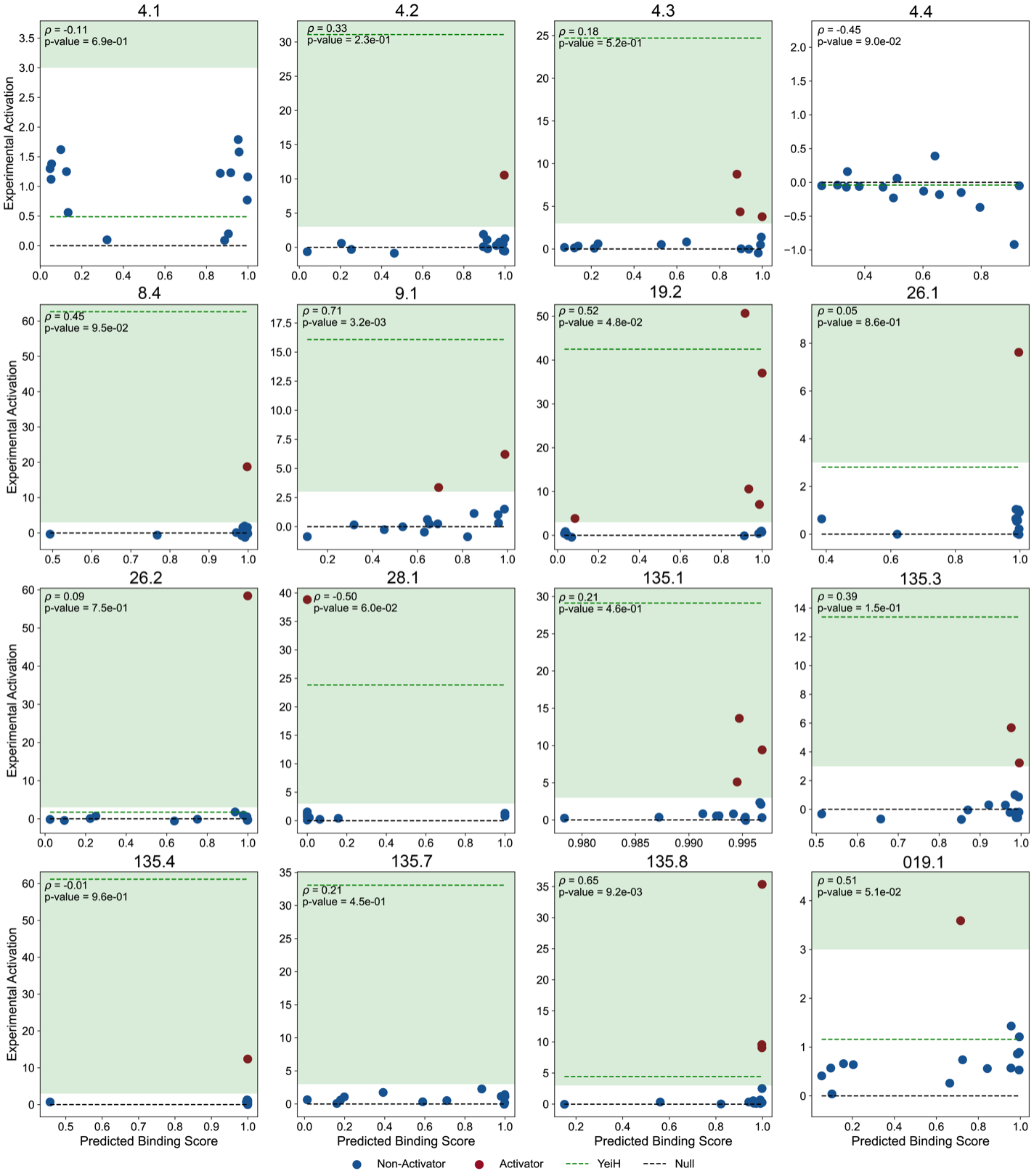
Supplementary Figure 3. Synthetic peptide activation of AS/AAU patient-derived TCRs.
CD8^{+} SKW-3 cells were transduced with individual AS/AAU-associated TCRs. HLA-B*27:05^{+} K562 cells were generated by lentiviral transduction and sorted for surface expression. APCs were pulsed with peptides (100 µM, 2 h), co-cultured with TCR-transduced SKW-3 cells for 16 h, and stained for CD69 expression. CD69-based activation of individual TCRs in response to candidate peptides. Red points indicate peptides classified as activators, defined as CD69% normalized by subtracting the average null peptide response, with a threshold > 3%.



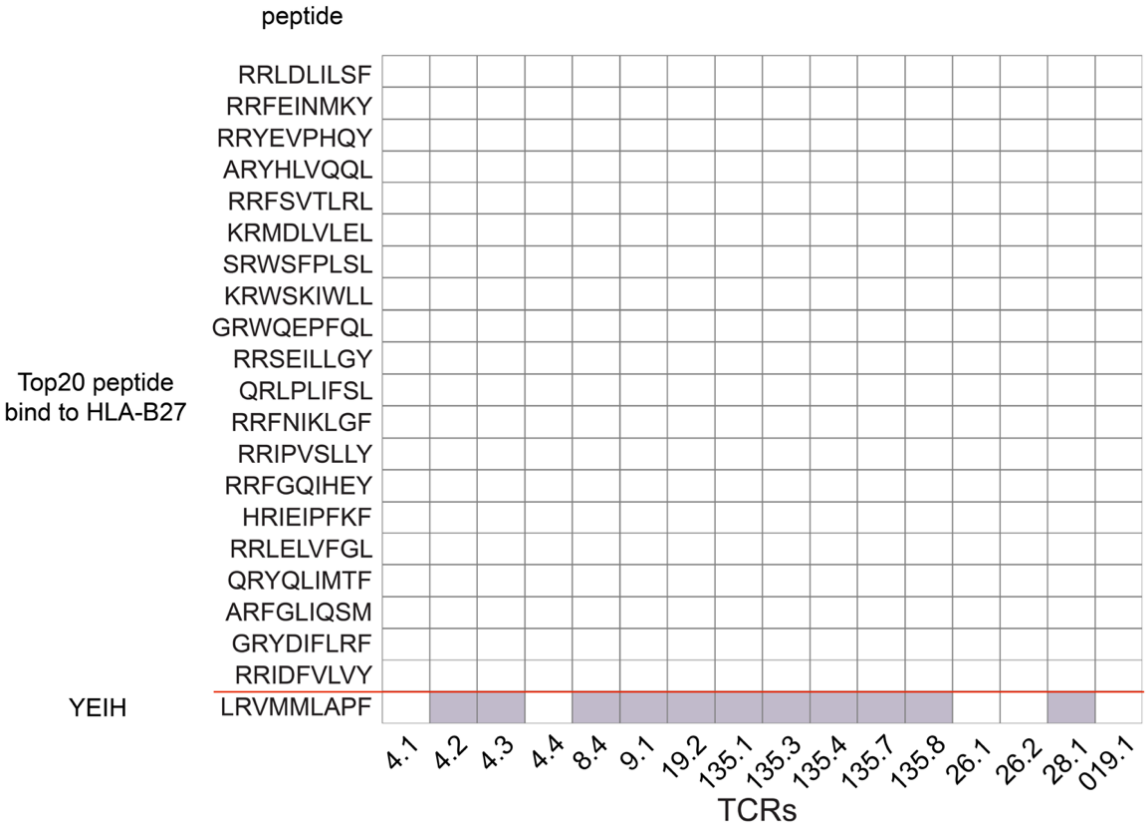
Supplementary Figure 4. Model-predicted binding probability scores for activators and non-activators.
a, Predicted binding probability scores for each individual TCR. Parentheses indicate statistical significance after Benjamini-Hochberg correction.
b, Boxplot showing aggregated predicted binding probability scores across all TCRs for activating and non-activating peptide. Statistical significance was assessed using a one-sided Mann-Whitney U test (**$p < 0.01$, $p=3.7 \times 10^{-3}$).



Supplementary Figure 5. Comparison of AlphaFold3 and model activation predictions.
Paired line plots comparing prediction performance between AlphaFold3 and our model across individual TCRs, assessed by AUROC (left) and average precision (AP; right). Statistical significance was evaluated using a one-sided Wilcoxon signed-rank test (AUROC *p=0.029, AP **p <=0.0052).

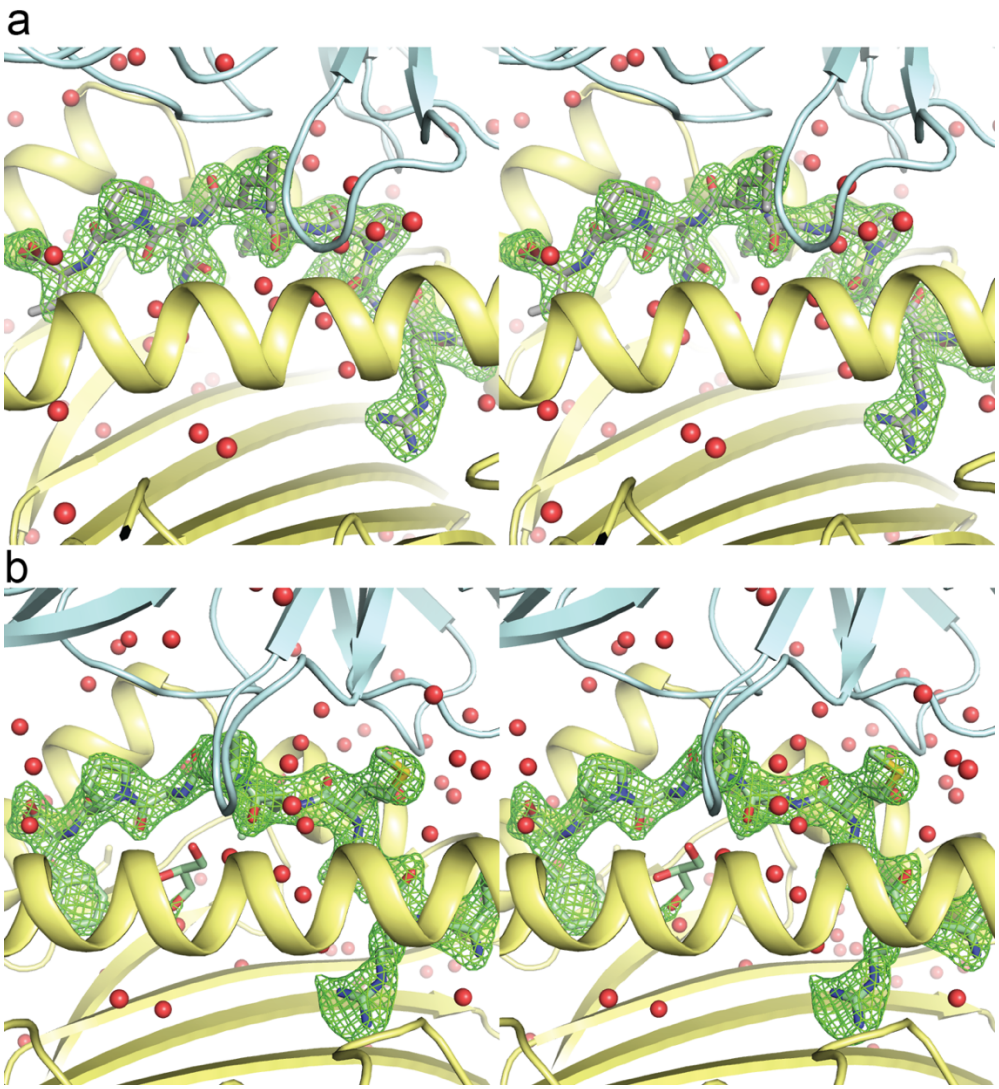


Supplementary Figure 6. Parity plots comparing predicted binding scores and experimental activation. Parity plots for each TCR showing the relationship between predicted binding scores and experimental activation levels. Each point represents a tested peptide (red = activators; blue = non-activators). The dashed black line indicates the response to the null peptide, while the green dashed line marks the response to the known activating peptide YeiH. Green shaded regions denote the activation threshold. Spearman correlation coefficient (ρ) and corresponding *P* values are reported for each TCR.

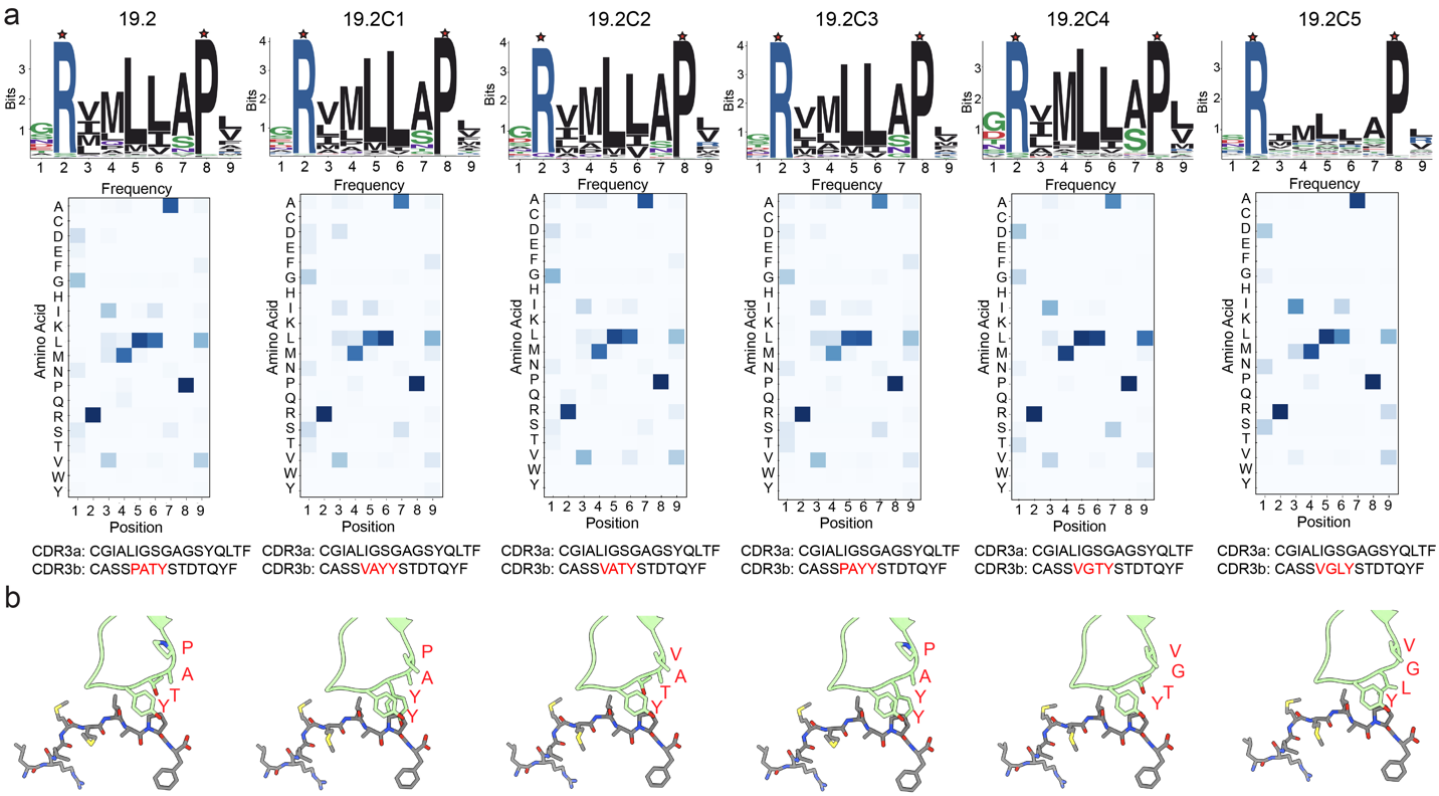


Supplementary Figure 7. Activation of NetMHCpan-predicted HLA-B*27:05-binding peptides from the human proteome.

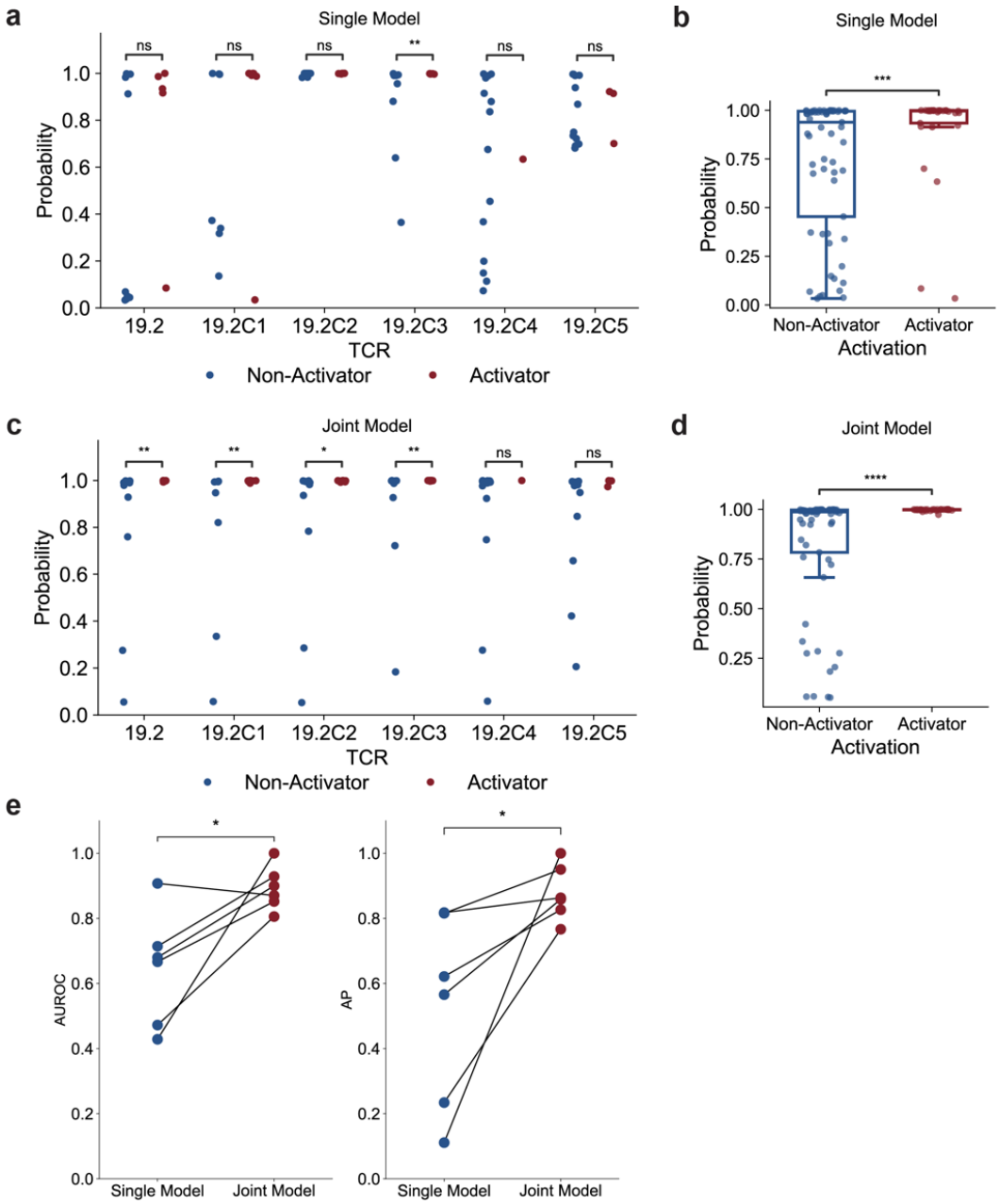
Shown are the top 20 peptides from human proteome with the highest predicted HLA-B*27:05 binding affinity, as determined by NetMHCpan, evaluated across all modeled TCRs. The YEIH peptide is included as a positive control.



Supplementary Figure 8. Electron density maps of crystal structures
Simulated annealing mFo-DFc maps were calculated for **(a)** the PSG5-HLA-B*27:05-TCR19.2 and **(b)** the YEIH-HLA-B*27:05-TCR19.2 structures. Structures are shown as walleye stereo images. TCR is depicted as a cartoon in pale cyan, HLA-B*27:05 in pale yellow, water molecules as red spheres, glycerol molecules as sticks, PSG5 and YEIH peptides as sticks in gray and pale green respectively, and maps are contoured in green at 3.0 sigma.



Supplementary Figure 9. Peptide recognition profiles of 19.2 CDR3β mutants and structural modeling.
a, Sequence logos and amino acid enrichment heatmaps for positive-binding peptides identified for TCR 19.2 and its CDR3β mutants (19.2C1-19.2C5).
b, Structural models showing CDR3β contact residues (mutated positions highlighted) interacting with a bound peptide, illustrating the impact of CDR3β mutations on peptide recognition patterns.



Supplementary Figure 10. Model-predicted binding probability scores for 19.2 variant activators and non-activators in single-variant and joint models.

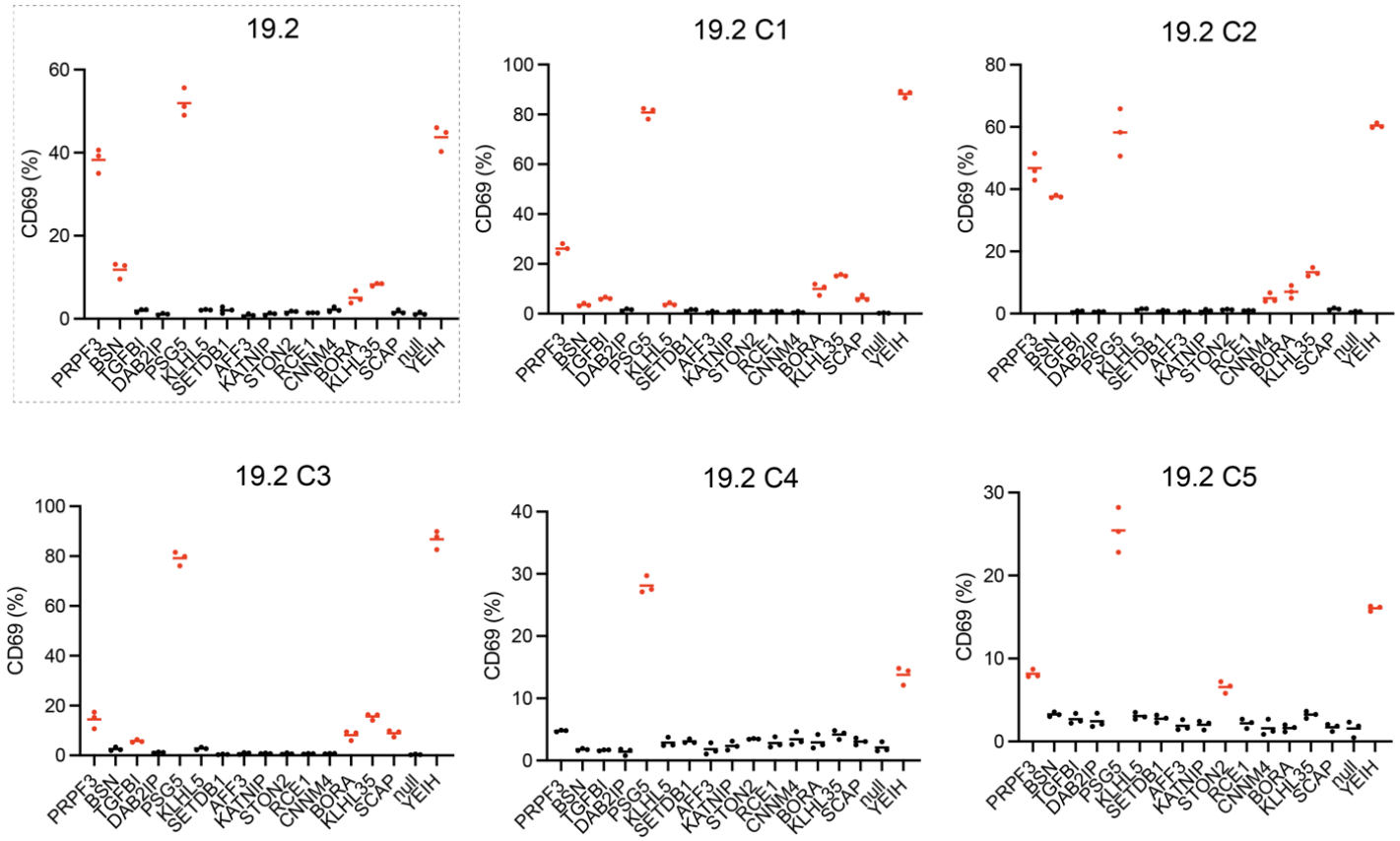
a, Predicted binding probability scores from single TCR models for each 19.2 variant. Parentheses indicate whether statistical significance changed after Benjamini-Hochberg correction.

b, Boxplot showing aggregated predictions from single TCR models across 19.2 variants for activating (1) and non-activating (0) peptides. Statistical significance was assessed using a one-sided Mann-Whitney U test (***$p < 0.001$, $p = 8.1 \times 10^{-4}$).

c, Predicted binding probability scores from the joint model for each 19.2 TCR variant. Parentheses indicate whether statistical significance changed after Benjamini-Hochberg correction.

d, Boxplot showing aggregated joint model predictions across 19.2 variants for activating (1) and non-activating (0) peptides. Statistical significance was assessed using a one-sided Mann-Whitney U test (****$p < 0.0001$, $p = 6.3 \times 10^{-9}$).

e, Paired line plots comparing prediction performance between single and joint models across 19.2 variants, assessed by AUROC (left) and average precision (AP; right). Statistical significance was evaluated using a one-sided Wilcoxon signed-rank test (AUROC: $p = 0.03125$; AP: $p = 0.015625$).



Supplementary Figure 11. Synthetic peptide activation of 19.2 TCR variants.
CDR-mutant variants of the 19.2 TCR tested against selected peptides. Five mutant TCRs (C1-C5) were generated based on 19.2 and evaluated in the same peptide stimulation assay. The 19.2 wild-type data (upper left, gray box) is reproduced from Supplementary Figure 6 and included here as a reference.

Supplementary Table 1. CD69 activation values for each TCR, calculated as the average CD69% minus the null control. Red values indicate peptides classified as activators.

Gene	Peptide	4.1	4.2	4.3	4.4	8.4	9.1	19.2	135.1	135.3	135.4	135.7	135.8	26.1	26.2	28.1	019.1
PRPF3	TRLALIAPK	1.22	0.28	3.78	-0.37	0.41	6.21	37.03	9.41	3.23	0.21	1.17	35.36	1.02	-0.41	1.51	1.21
BSN	TRVPMIAPR	1.79	1.28	0.49	-0.92	0.31	1.50	10.58	2.10	1.01	0.73	1.23	9.57	0.92	-0.14	1.54	0.86
TGFBI	ERLTLLAPL	1.16	-0.63	0.09	-0.07	1.29	1.13	0.74	2.35	0.30	0.34	1.41	2.51	0.58	0.54	1.31	0.89
DAB2IP	GRPQLLAPL	0.77	-0.85	-0.47	-0.18	-0.20	-0.87	-0.11	0.81	-0.70	0.00	-0.04	0.67	1.04	-0.29	0.81	0.56
PSG5	GRLPLLNPI	1.62	10.55	8.76	-0.15	18.73	3.36	50.66	13.64	5.68	12.38	2.27	9.07	7.62	58.42	38.83	3.59
KLHL5	IRLPLLAPQ	1.30	-0.20	0.02	-0.23	0.48	0.24	0.92	0.57	-0.19	1.04	0.59	0.46	0.54	-0.10	0.90	1.43
SETDB1	GKNPLLVPL	1.25	-0.46	0.18	-0.05	0.08	-0.26	0.86	0.55	-0.20	0.72	1.74	0.26	0.57	0.16	1.22	0.74
AFF3	GRNELLSPL	0.56	-0.30	-0.03	0.39	-0.80	-0.86	-0.43	0.37	-0.57	0.17	0.08	0.54	0.65	-0.38	0.94	0.04
KATNIP	GRHMWLAPI	1.23	0.75	0.52	-0.06	-1.26	0.31	-0.08	0.25	-0.57	1.20	1.07	0.02	0.64	-0.55	0.54	0.66
STON2	SRVILFNPL	1.58	1.12	0.34	-0.13	-0.63	0.23	0.42	0.32	-0.04	1.29	1.12	0.15	0.64	0.09	0.88	0.57
RCE1	LRNQVIAPL	1.12	0.09	1.39	-0.07	-0.23	-0.01	0.22	0.83	-0.32	1.07	1.15	0.33	0.68	-0.32	0.68	0.41
CNNM4	GRLLLAAPV	0.20	1.90	4.35	0.16	-0.35	1.02	0.97	5.10	0.31	0.61	0.18	0.34	0.22	1.81	0.42	0.53
BORA	GRIPVLNPF	0.10	0.53	0.60	-0.05	1.60	0.15	3.85	0.38	-0.67	1.00	0.51	0.11	-0.01	0.98	0.11	0.64
KLHL35	VRLPLLAPA	1.38	0.60	0.13	0.06	1.58	0.61	7.05	0.14	0.87	1.05	0.64	0.10	0.03	0.74	0.71	0.57
SCAP	ARLMLLHPS	0.09	-0.54	0.82	-0.04	1.94	-0.47	0.40	-0.06	-0.21	0.84	0.33	-0.03	0.00	-0.16	0.24	0.26
YEIH	LRVMMLAPF	0.49	31.08	24.72	-0.04	62.66	16.08	42.50	29.11	13.38	61.15	33.08	4.42	2.81	1.74	23.83	1.16

Supplementary Table 2. Participant demographics.

	Sex	Age	Diagnosis
HC252	M	40	
HC289	M	34	
HC310	F	40	
HC327	F	41	
HC343	F	24	
HC770	F	37	
AI028	F	63	AAU, nr-axSpA
AI041	M	42	AS
AI065	F	53	AAU, nr-axSpA
AI126	F	42	AS
AI206	M	42	AS
AI246	M	54	AAU

Supplementary Table 3. Antibodies used for flow cytometry

Target	Conjugate	Clone	Vendor	Catalog Number
CD3	BUV395	OKT3	ThermoFisher	363-0037-42
CD4	PerCP	RPA-T4	ThermoFisher	11-0049-42
CD8	cFluor V420	SK1	Cytek	R7-20238
CD19	FITC	HIB19	ThermoFisher	11-0199-42
TCRgd	FITC	B1.1	ThermoFisher	11-9959-42
TCR Va7.2	FITC	3C10	Biolegend	351703
TCR Va24-Ja18	FITC	6B11	Biolegend	342905

Supplementary Table 4. Primers for library generation and sequencing

Name	Sequence
Primer F1	TAATTCTACTTCATACATTTTCAATTAAGATGCAGTTACTTCGCTGTTTT
Primer library	TTTCTGTTATTGCTAGCGTTTTGGCTNNKAGANNKNNKNNKNNKCCANNKGGAGGTGGCGGGAGTGG
Primer F2	GAGGTTCTGGGGGCGGAGGTAGTATTCAAAGAACTCCAAAAATTCAAGTTTACTCAAGAC
Primer F3	CTGCAGAAAACGGTAAATCCAATTTCTTGAACTGTTACGTATCTGGTTTCCACCCATCAG
Primer R1	AGCCAAAACGCTAGCAATAACAGAAAATATTGAAAAACAGCGAAGTAACTGCATCTTAAT
Primer R2	TCCGCCCCCAGAACCTCCACCGCCACTCCCGCCACCTCC
Primer R3	GAAATTGGATTTACCGTTTTCTGCAGGATGTCTTGAGTAAACTTGAATTTTTGGAGTTCT
Primer R4	ATTCTTTCACCATTTTTCAACAAATCGACTTCAACGTCTGATGGGTGGAAACCAGATACG
Secondary F	TAATTCTACTTCATACATTTTCAATTAAGATGC
Secondary R	ATTCTTTCACCATTTTTCAACAAATCG
Illumina sequencing primer F	CTACACGACGCTCTTCCGATCTNNNNNNNNATCGATTAATTCTACTTCATACATTTTCAATTAAGATGCAGTTACTTCGCTGTTTTTCAATATTTT
Illumina sequencing primer R	ATTCCTGCTGAACCGCTCTTCCGATCTNNNNNNNNTGCAGGATGTCTTGAGTAAACTTGAATTTTTGGAGTTCTTTGAATACTACCTCCGCCCCC

Supplementary Table 5. Data collection and refinement statistics (molecular replacement)

	AS19.2 - YEIH	AS19.2 - PSG5
Wavelength (Å)	0.9795	0.9795
Resolution range (Å)	46.75 - 2.0 (2.05 - 2.0)	44.59 - 2.13 (2.21 - 2.13)
Space group	P 1 2_1 1	P 1 21 1
Unit cell (Å, °)	84.84 52.68 102.04 90 96.51 90	84.08 52.72 100.32 90 96.35 90
Total reflections	203565 (13540)	161001 (16548)
Unique reflections	59741 (3971)	47974 (4828)
Multiplicity	3.4 (3.4)	3.4 (3.4)
Completeness (%)	97.48 (96.84)	97.16 (97.75)
Mean I/sigma(I)	7.03 (0.97)	6.32 (1.02)
Wilson B-factor ($Å^2$)	35.8	39.0
R_{merge}	0.109 (1.81)	0.143 (1.40)
R_{meas}	0.129 (2.15)	0.170 (1.66)
R_{pim}	0.0691 (1.14)	0.0910 (0.879)
$CC_{1/2}$	0.995 (0.552)	0.993 (0.374)
Reflections used in refinement	59476 (3918)	47949 (4822)
Reflections used for R_{free}	2142 (138)	1436 (146)
R_{work}	0.220 (0.421)	0.210 (0.326)
R_{free}	0.258 (0.413)	0.247 (0.389)
Number of non-hydrogen atoms	6849	6830
macromolecules	6611	6650
heteroatoms	82	40
solvent	156	140
Protein residues	827	832
RMS(bonds) (Å)	0.002	0.003
RMS(angles) (°)	0.48	0.59
Ramachandran favored (%)	95.8	97.1
Ramachandran outliers (%)	4.04	2.92
Rotamer outliers (%)	0.12	0.00
Clashscore	1.38	1.10
Average B-factor ($Å^2$)	2.52	2.30
macromolecules	72.7	53.9
heteroatoms	73.1	54.1
solvent	77.3	76.4
PDB Accession code	9PBG	9PBH

Statistics for the highest-resolution shell are shown in parentheses.