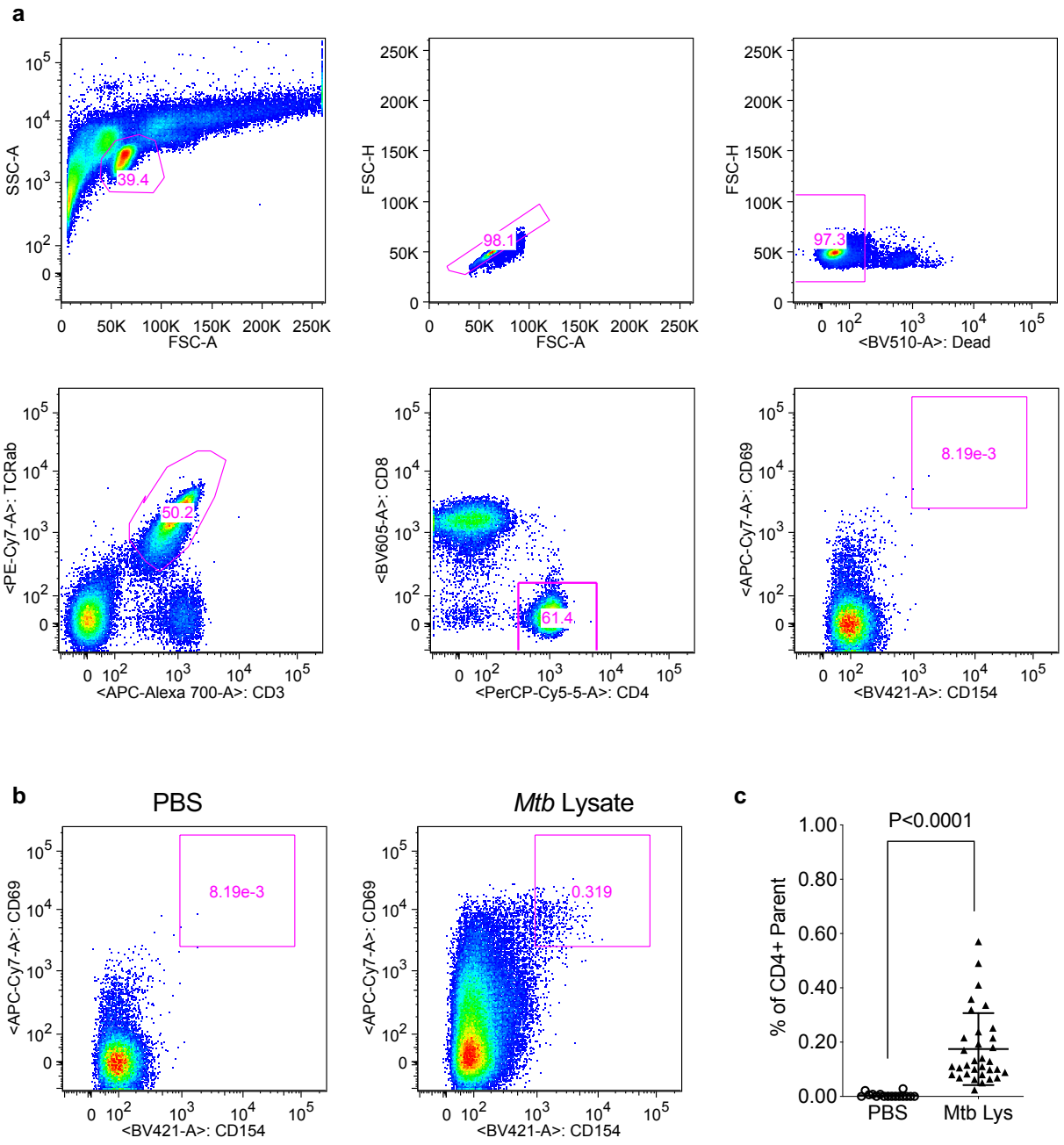
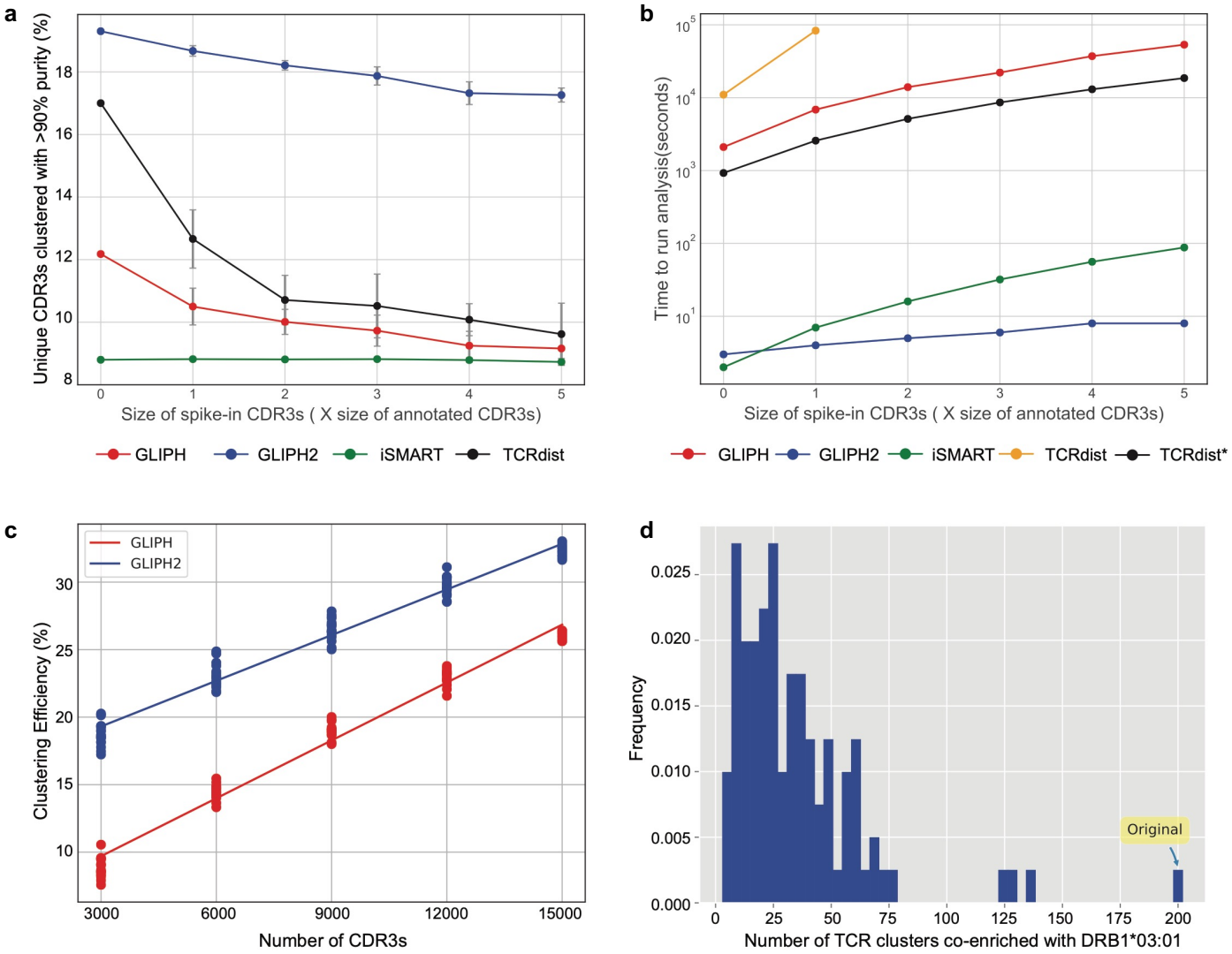
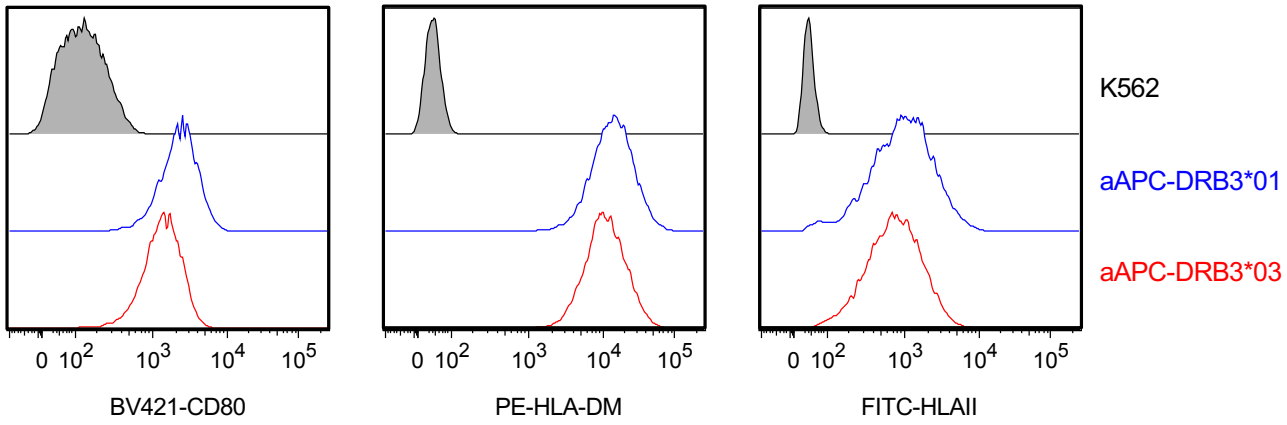


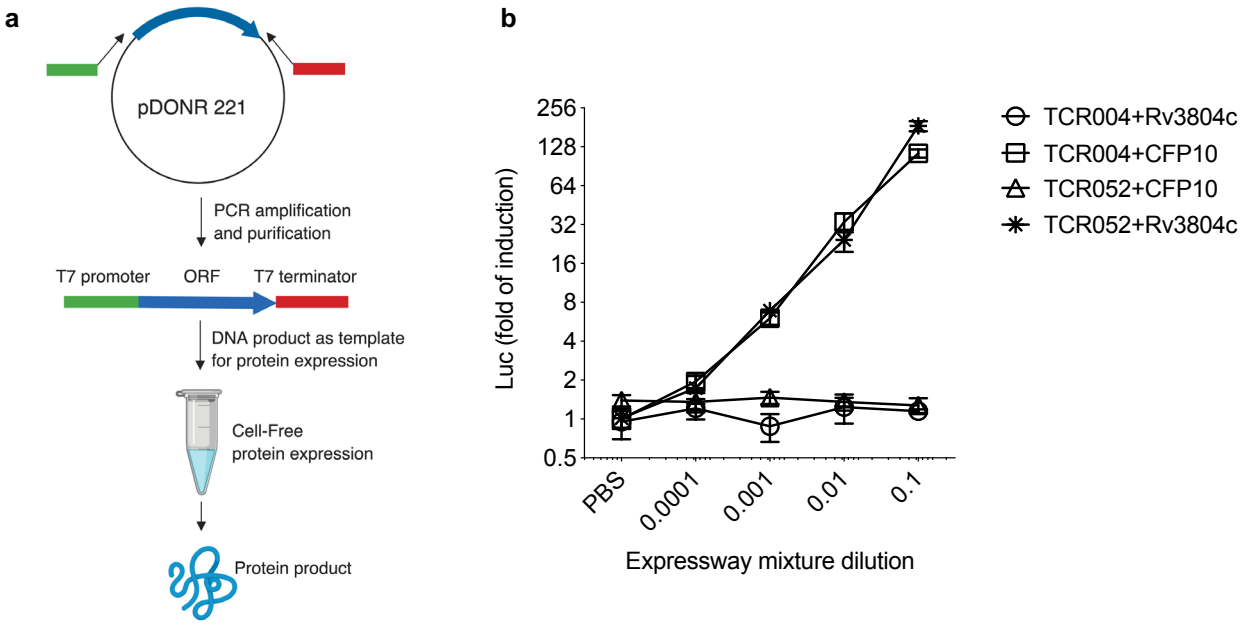
Supplementary Figure 1. Stimulation of PBMC with *Mtb* Lysate. a, Gating strategy used for isolating and sorting *Mtb*-reactive CD4⁺ T cells after 12-hour *Mtb* lysate stimulation. **b**, The frequency increase of CD4⁺CD69⁺CD154⁺ population after *Mtb* lysate stimulation. Plots shown as representative of two independent experiments. **c**, Frequency of CD4⁺CD69⁺CD154⁺ population as a proportion of the CD4⁺ T cells from PBS (n=16) and *Mtb* lysate stimulation (n=24). Mean ± s.d. shown. P value calculated by two-sided paired Wilcoxon test.



Supplementary Figure 2. Benchmark of GLIPH2. a-b, Performance comparison among GLIPH, GLIPH2, TCRdist (star indicates a modified code for clustering analysis only) and iSMART. 3264 unique CDR3β sequences from 15 epitopes were compiled from VDJdb (Supplementary Table 2). X-axis indicates the amount of spike-in data randomly selected from a $CD4^+$ T cell reference database, 0: no spike-in, 1-5: different fold of spike-in data were mixed with annotated CDR3β sequences. **a**, Y-axis indicates the percentage of correctly clustered unique CDR3β sequences. Clusters with more than 90% CDR3β sequences labeled with one correct epitope are considered correct clusters. s.d of 10 repeat random samples of spike-in reported on bars. **b**, Y-axis indicates the amount of time needed to run each algorithm. TCRdist indicates the whole package analysis while TCRdist* indicates the clustering analysis only. **c**, Analyzing the 19,044 *Mtb*-specific CDR3β sequences, GLIPH2 clusters up to 32% of *Mtb* reactive TCRs while the original GLIPH clusters 26% of *Mtb* reactive TCRs. **d**, The histogram shows the distribution of number of TCR clusters co-enriched with DRB1*0301 after 100-time permutations of HLA information among all the individuals. The original number is highlighted with an arrow.



Supplementary Figure 3. Generation of aAPC using K562. Expression analysis of surface molecules on generated aAPC (artificial antigen presenting cell). Filled histograms: K562 cells as baseline control, colored histograms: respective aAPC. Plots shown as representative of two independent experiments.



Supplementary Figure 4. *In vitro* cell-free protein expression as antigen source for TCR activation. a, A schematic depicting the workflow of cell-free protein expression. Briefly, individual ORF in vector pDONR 221 was PCR amplified using a pair of primers containing T7 promoter and terminator on each end. PCR product was then purified and used as template for protein expression using Expressway system in a single tube. The reaction mixture with protein product could be used immediately to activate its corresponding TCR without further purification. **b**, Protein CFP10 and Rv3804c were produced *in vitro* as described in **a**. TCR004 was activated by its target protein CFP10 but not by an unrelated protein Rv3804c. Similarly, TCR052 only activated by its target protein Rv3804c, but not by CFP10. x-axis indicates the titration of protein product. Mean ± s.d. (n=3, biological replicates) shown.

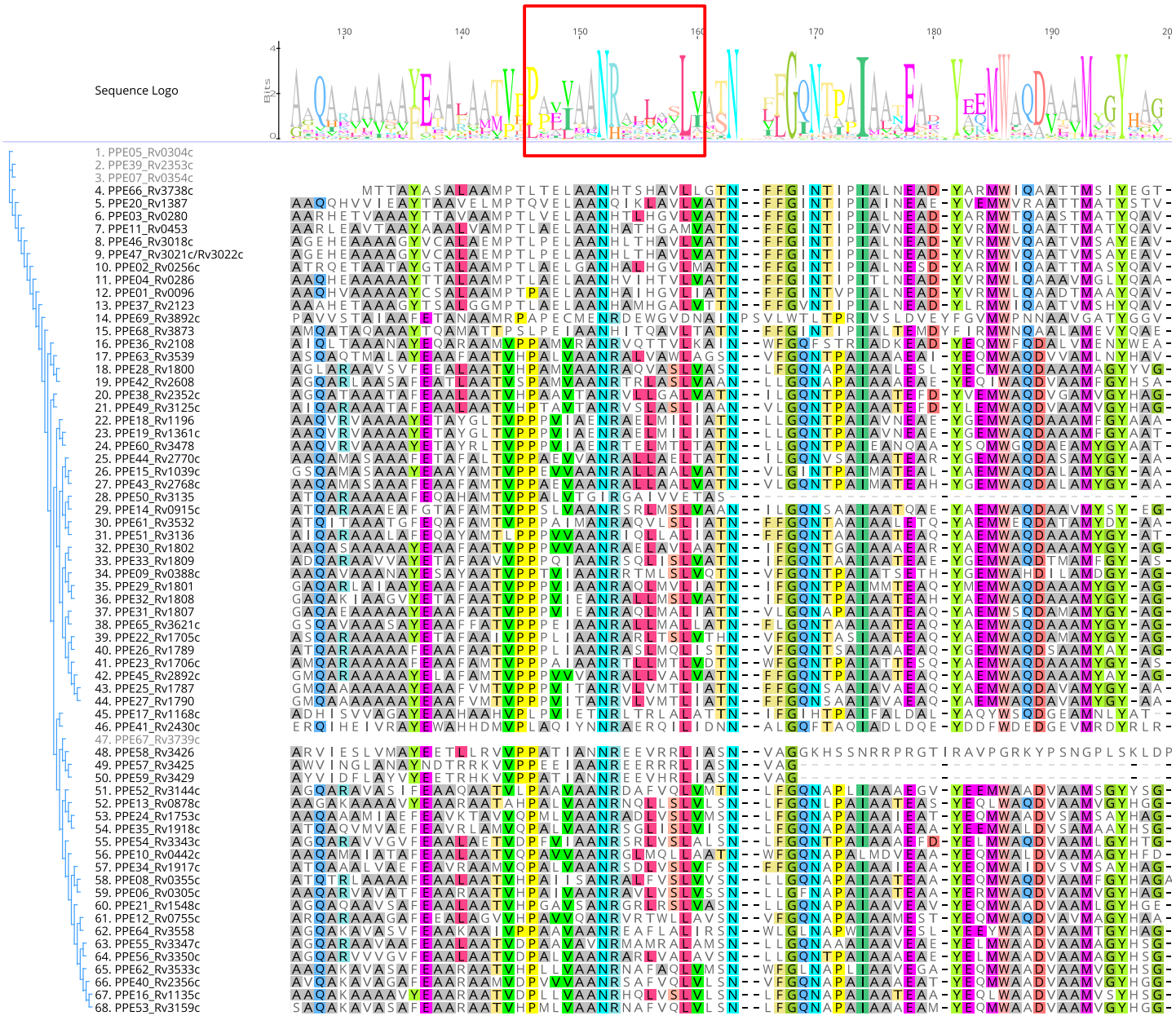
	PL1	PL2	PL3	PL4	PL5	PL6	PL7	PL8	PL9	PL10	PL11	PL12
A	PL1A	PL2A	PL3A	PL4A	PL5A	PL6A	PL7A	PL8A	PL9A	PL10A	PL11A	PL12A
B	PL1B	PL2B	PL3B	PL4B	PL5B	PL6B	PL7B	PL8B	PL9B		PL11B	PL12B
C	PL1C	PL2C	PL3C	PL4C	PL5C	PL6C	PL7C	PL8C	PL9C		PL11C	PL12C
D	PL1D	PL2D	PL3D	PL4D	PL5D	PL6D	PL7D	PL8D	PL9D		PL11D	PL12D
E	PL1E	PL2E	PL3E	PL4E	PL5E	PL6E	PL7E	PL8E	PL9E		PL11E	PL12E
F	PL1F	PL2F	PL3F	PL4F	PL5F	PL6F	PL7F	PL8F	PL9F		PL11F	PL12F
G	PL1G	PL2G	PL3G	PL4G	PL5G	PL6G	PL7G	PL8G	PL9G		PL11G	PL12G
H	PL1H	PL2H	PL3H	PL4H	PL5H	PL6H	PL7H	PL8H	PL9H		PL11H	PL12H

	PL13	PL14	PL15	PL16	PL17	PL18	PL19	PL20	PL21	PL22	PL23	PL24
A	PL13A	PL14A	PL15A	PL16A	PL17A	PL18A	PL19A	PL20A	PL21A	PL22A	PL23A	PL24A
B	PL13B	PL14B	PL15B	PL16B	PL17B	PL18B	PL19B	PL20B	PL21B	PL22B	PL23B	PL24B
C	PL13C	PL14C	PL15C	PL16C	PL17C	PL18C	PL19C	PL20C	PL21C	PL22C	PL23C	PL24C
D	PL13D	PL14D	PL15D	PL16D	PL17D	PL18D	PL19D	PL20D	PL21D	PL22D	PL23D	PL24D
E	PL13E	PL14E	PL15E	PL16E	PL17E	PL18E	PL19E	PL20E	PL21E	PL22E	PL23E	PL24E
F	PL13F	PL14F	PL15F	PL16F	PL17F	PL18F	PL19F	PL20F	PL21F	PL22F	PL23F	PL24F
G	PL13G	PL14G	PL15G	PL16G	PL17G	PL18G	PL19G	PL20G	PL21G	PL22G	PL23G	PL24G
H	PL13H	PL14H	PL15H	PL16H	PL17H	PL18H	PL19H	PL20H	PL21H	PL22H	PL23H	PL24H

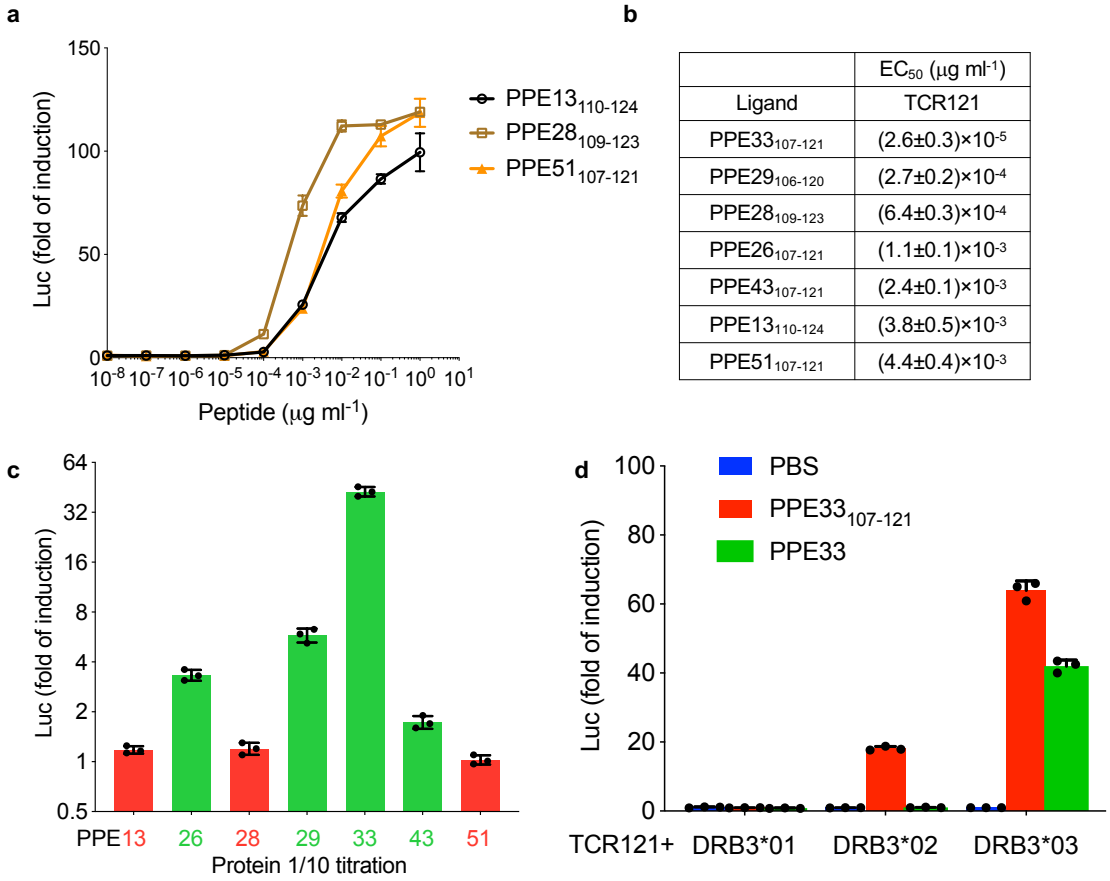
	PL25	PL26	PL27	PL28	PL29	PL30	PL31	PL32	PL33	PL34	PL35	PL36
A	PL25A	PL26A	PL27A	PL28A	PL29A	PL30A	PL31A	PL32A	PL33A	PL34A	PL35A	PL36A
B	PL25B	PL26B	PL27B	PL28B	PL29B	PL30B	PL31B	PL32B	PL33B	PL34B	PL35B	PL36B
C	PL25C	PL26C	PL27C	PL28C	PL29C	PL30C	PL31C	PL32C	PL33C	PL34C	PL35C	PL36C
D	PL25D	PL26D	PL27D	PL28D	PL29D	PL30D	PL31D	PL32D	PL33D	PL34D	PL35D	PL36D
E	PL25E	PL26E	PL27E	PL28E	PL29E	PL30E	PL31E	PL32E	PL33E	PL34E	PL35E	PL36E
F	PL25F	PL26F	PL27F	PL28F	PL29F	PL30F	PL31F	PL32F	PL33F	PL34F	PL35F	PL36F
G	PL25G	PL26G	PL27G	PL28G	PL29G	PL30G	PL31G	PL32G	PL33G	PL34G	PL35G	PL36G
H	PL25H	PL26H	PL27H	PL28H	PL29H	PL30H	PL31H	PL32H	PL33H	PL34H	PL35H	PL36H

	PL37	PL38	PL39	PL40	PL41	PL42						
A	PL37A	PL38A	PL39A	PL40A	PL41A	PL42A						
B	PL37B	PL38B	PL39B	PL40B	PL41B	PL42B						
C	PL37C	PL38C	PL39C	PL40C	PL41C	PL42C						
D	PL37D	PL38D		PL40D	PL41D	PL42D						
E	PL37E	PL38E		PL40E	PL41E	PL42E						
F	PL37F	PL38F		PL40F	PL41F							
G	PL37G	PL38G		PL40G	PL41G							
H	PL37H	PL38H		PL40H	PL41H							

Supplementary Figure 5. 96-well plate diagram with subpool ID. Originally, 3724 individual *Mtb* ORF Clones (Supplementary Table 5) were provided by BEI in 96-well plate format (42 plates in total). In order to facilitate the screening process, every row from an original plate (12 clones) were pooled together as a subpool, indicated above as a plate number combined with a row letter. As an example, “PL1A” subpool stands for the original clone plate 1 row A. Thus, the original 42 plates were minimized to 4 plates (321 subpools) as shown above and used for screening.



Supplementary Figure 6. Sequence conservation of PPE family proteins. Sequences of 68 PPE proteins from mycobacterial strain H37Rv were retrieved from UniProt database and aligned using Muscle alignment. Red outline highlights the conserved peptides containing “AANR” region which is recognized by specificity group I TCR121 and TCR122.



	EC_{50} (μg ml^{-1})
Ligand	TCR121
PPE33$_{107-121}$	(2.6±0.3)×10^{-5}
PPE29$_{106-120}$	(2.7±0.2)×10^{-4}
PPE28$_{109-123}$	(6.4±0.3)×10^{-4}
PPE26$_{107-121}$	(1.1±0.1)×10^{-3}
PPE43$_{107-121}$	(2.4±0.1)×10^{-3}
PPE13$_{110-124}$	(3.8±0.5)×10^{-3}
PPE51$_{107-121}$	(4.4±0.4)×10^{-3}

Supplementary Figure 7. Peptide titration assay for TCR121. a, Dose-dependent response of Group I TCR121 to three PPE peptides containing the "AANR" motif. Mean ± s.d. (n=3, biological replicates) shown. **b**, EC50 values were determined from dose-response curves and ordered from lowest to highest among 7 PPE peptides that are ubiquitously reactive to TCR121. The average EC50 value and S.D. for each ligand were calculated from three different experiments. **c**, From PBMC carrying DRB3*03 allele, monocyte-derived dendritic cells were generated as antigen presenting cell (APC) to activate TCR121 with different PPE family proteins. Green indicating the four positive PPE proteins with known potency and red indicating the three proteins with unknown potency. **d**, Monocyte-derived dendritic cells were generated from three different individuals' PBMC carrying DRB3*01, DRB3*02 and DRB3*03 allele respectively and then used as APC to activate TCR121 with both peptide PPE33$_{107-121}$ and the whole PPE33 protein. Mean ± s.d. (n=3, biological replicates) shown.