Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Mapping of the M72 Vaccine Identifies Conserved Immunodominant PPE18 T Cell Recognition Sequences Across Mycobacterium tuberculosis Lineages

Paul Ogongo
University of California, San Francisco
Julia Huffaker
University of California, San Francisco
Jason D. Limberis
University of California, San Francisco
Drienna D. Holman
Fred Hutchinson Cancer Center
Andrew J. Fiore-Gartland
Fred Hutchinson Cancer Center
Joel D. Ernst
joel.ernst@ucsf.edu

University of California, San Francisco

Article

Keywords:

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11246059/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Although M72/AS01$_E$ induces durable, multifunctional CD4 T cell responses, the specific epitopes driving vaccine-induced immunity remain undefined. Defining these targets is essential for understanding protective mechanisms and identifying immune correlates of protection. We mapped T cell recognition across the M72 fusion protein using 178 overlapping 15-mer peptides. We stimulated PBMC collected before vaccination (Day 0) and after a two-dose regimen (Day 37) from Phase 2b trial participants with peptide pools and individual peptides, and quantified IFNg and IL-17 responses by ELISpot. Vaccine-responsive peptides were identified by comparing paired responses in vaccine and placebo recipients. Epitope conservation was assessed using whole-genome sequences from approximately 34,000 clinical *Mycobacterium tuberculosis* isolates, and HLA class II binding was predicted using NetMHCIIpan 4.1. Peptide deconvolution revealed that immunogenicity was concentrated within PPE18, with five PPE18-derived peptides demonstrating vaccine-associated boosting exclusively in vaccine recipients. These sequences were highly conserved across major *M. tuberculosis* lineages and shared substantial homology with PPE19 and PPE60, suggesting potential cross-reactivity. Multiple HLA class II alleles were predicted to present the five vaccine-responsive peptides. Our findings identify conserved PPE18 epitopes targeted by M72/AS01$_E$-induced T cell responses and provide a foundation for investigating their contribution to vaccine-induced immunity and developing epitope-specific correlates of protection against TB.

Introduction

M72/AS01$_E$ is a recombinant protein subunit tuberculosis (TB) vaccine consisting of two *M. tuberculosis* antigens, PPE18 (Mtb39A, Rv1196) and PepA, also known as MTB32A (Rv0125), formulated with the AS01$_E$ adjuvant. In a Phase 2b clinical trial, a two-dose regimen administered one month apart provided approximately 50% protection against pulmonary TB disease in QuantiFERON-positive adults[1,2]. This unprecedented efficacy for a subunit TB vaccine has advanced M72/AS01$_E$ into an ongoing Phase 3 trial in Africa and Asia, raising the possibility of a new vaccine to reduce the global TB burden. The vaccine induces durable humoral and cellular immune responses across diverse populations, including individuals with previously treated TB[3] and people living with HIV, regardless of antiretroviral therapy status[4,5]. M72 vaccination elicits multifunctional CD4 T cell responses characterized by IFNg, IL-2, TNF-α, IL-13, IL-17, and CD40L production[6,7], with responses maintained for up to three years and enhanced following two-dose compared with single-dose vaccination schedules[1,2,8–10].

Despite extensive characterization of M72/AS01$_E$ immunogenicity, the specific T cell epitopes responsible for these responses remain undefined. This gap is important because the recombinant fusion protein contains potential neoepitopes generated at the PepA–PPE18 junctions that are absent from native *M. tuberculosis* proteins[11] and at the N- and C-termini due to the architecture of the fusion protein. Furthermore, studies in other infections, including HIV and hepatitis B virus, demonstrate that the epitope specificity of T cell responses can distinguish protective from non-protective immunity[12–14]. Therefore, defining epitope-level T cell responses, rather than antigen-level responses alone, is critical for identifying immune correlates of protection since the signature of T cell responses can be determined by the antigens they recognize[15–18].

Here, we do not refer to the experimentally defined T cell recognition domains in the vaccine as epitopes, since HLA typing data was not available. To define the vaccine recognition domains driving M72-induced immunity, we generated overlapping 15-mer peptides (11-amino acid overlap) spanning the entire M72 antigen and assessed antigen-specific responses in PBMC from placebo and vaccinated participants. Peptide pool screening identified

regions inducing IFNg and IL-17 responses, followed by individual peptide deconvolution to resolve immunodominant domains. The strongest responses were concentrated within the PPE18 component, whereas peptides spanning the fusion junctions failed to elicit responses above background. Unblinded analysis of matched Day 0 and Day 37 samples identified five PPE18-derived peptides that demonstrated vaccine-induced effect exclusively in vaccine recipients. Despite reported sequence variability within PPE18, all five peptides that demonstrated a vaccine effect were conserved across major *M. tuberculosis* lineages without lineage-associated non-synonymous mutations. Furthermore, these five PPE18-derived peptides showed extensive sequence homology with PPE19 and PPE60, with more than 51 sequences (54%) exhibiting either complete identity or a single amino acid mismatch. Together, these findings reveal that M72/AS01$_E$-induced T cell immunity is focused on PPE18 peptides that are broadly conserved across *M. tuberculosis* strains and shared among related PPE family members.

Results

Enhanced immune responses to peptide pools in post-vaccination compared to the pre-vaccination time point

To maximize the number of peptides that could be evaluated while conserving the limited number of available PBMC, we combined the 178 individual peptides into pools using a matrix pooling strategy as previously described[19]. This generated 14 peptide pools containing 24–28 peptides each, except for pool 14, which contained 10 peptides (**Supplementary Fig. 1**). We stimulated PBMC with each peptide pool and quantified antigen-specific IFNg and IL-17 responses by ELISpot after subtracting average background responses from unstimulated controls. Both IFNg and IL-17 responses varied considerably among participants and across peptide pools. Comparison of pre-vaccination (Day 0) and post-vaccination (Day 37) samples showed an overall increase in response magnitude following vaccination, with the greatest increases observed for IFNg (Fig. 1).

Peptide pools 1, 3, and 10 consistently elicited the strongest Day 37 responses across participants. Overall, IFNg responses were approximately threefold higher than IL-17 responses to the same peptide pools. Although Day 37 responses were generally higher than Day 0 responses, unblinding analysis revealed no significant vaccine-specific enhancement at the peptide pool level when vaccine recipients were compared with placebo controls (**Supplementary Fig. 2**). Together, these findings indicate an overall increase in immune responses following the two-dose vaccination schedule, but suggest that pooling may have obscured responses to individual immunogenic peptides. The marked variability in responses across peptide pools therefore prompted subsequent deconvolution analyses to identify the specific peptides responsible for the observed and previously reported immune responses induced by M72/AS01$_E$ vaccine.

Immunogenic T cell targets are predominantly derived from the PPE18 component of the M72 vaccine

To identify the individual peptides responsible for the responses observed with peptide pools, we performed peptide deconvolution experiments. Individual peptides were selected from peptide pools that elicited detectable responses above background, irrespective of whether responses increased after vaccination. Thus, Day 0 and Day 37 samples were analyzed independently during the deconvolution phase. Individual peptides were then mapped to the M72 fusion protein, including peptides spanning the junctions between the PPE18 and PepA components

(Fig. 2A). To ensure robust and reproducible analyses, only peptides tested in at least five participants (at either time point) were included.

Individual peptide stimulation revealed substantial heterogeneity in both IFNg and IL-17 responses, with IFNg responses consistently exceeding IL-17 responses in magnitude (Fig. 2B; **Supplementary Fig. 3A**). Mapping responses along the M72 sequence identified discrete regions of high and low immunogenicity. Several clusters of highly immunogenic peptides ("antigenic islands"), including peptides 57–64 and 78–84, elicited strong responses, whereas other regions, such as peptides 49–56, generated little or no detectable response ("antigenic deserts").

Strikingly, the strongest responses were almost exclusively directed against peptides derived from the PPE18 component of the fusion protein. In contrast, peptides originating from the PepA component elicited substantially weaker responses overall. We next assessed the prevalence of the response by determining the number of participants responding to each peptide (Fig. 2C; **Supplementary Fig. 3B**). Consistent with the response magnitude data, PPE18-derived peptides were recognized by a larger proportion of participants, whereas most PepA-derived peptides were recognized by only a few individuals.

Together, these findings demonstrate that the immunogenicity of the M72/AS01$_E$ vaccine is dominated by domains within the PPE18 component, which drive both IFNg and IL-17 responses. However, immunogenicity is not uniformly distributed across PPE18, with distinct peptide clusters eliciting substantially stronger responses than others.

Identification of peptide-specific vaccine effect after two-dose M72 vaccine administration

To identify peptide-specific vaccine-induced responses, we performed an unblinded analysis comparing placebo recipients and vaccinees. Vaccine-induced effects were assessed by comparing matched Day 0 and Day 37 IFNg and IL-17 responses for individual peptides, restricting analyses to peptides tested in at least five participants. Peptides with a paired comparison $p < 0.1$ and assayed in ≥ 5 participants at matched time points were considered as candidate vaccine-boosted immune responses.

Using a paired Wilcoxon signed-rank test and paired AUROC analysis (AUROC = 0 indicates a consistent decrease at Day 37; AUROC = 0.5 indicates no change; AUROC = 1 indicates a consistent increase in responses at Day 37), we identified five peptides, all derived from the PPE18 component of the M72 vaccine, that demonstrated a vaccine effect as defined by IFNg and IL-17 response magnitudes. Of the five, Peptide #61 (VPPPVIAENRAELMI; assayed in $n = 33$) passed the minimum response threshold of at least 35% of the samples > 25 SFC/M and a significant IFNg-specific response (AUROC = 0.72, $p = 0.032$, FDRq = 0.36) (Fig. 3A). Receiver operator curve (ROC) analysis demonstrated that responses to peptide #61 could discriminate pre- and post-vaccination samples (Fig. 3B). Additional peptides (#50, #64, #67, and #77) with p-values < 0.1 and measured in 33 out of 36 vaccine recipients showed increased responses after vaccination and are considered potential candidates with vaccine-induced effect (**Supplementary Table 1**). Since HLA data were not available for these donors, these peptides whose responses are potentially boosted by vaccination require validation with the right HLA genotype background.

Together, these findings describe five PPE18-derived peptides that account for much of the vaccine-induced T cell immunogenicity following the two-dose M72 regimen. The results highlight the heterogeneity of T cell responses

across the M72 antigen and pinpoint discrete PPE18 domains that likely contribute to the vaccine's previously reported immunogenicity.

PPE18-derived peptides that demonstrate vaccine-induced effect are conserved across M. tuberculosis *lineages*

PPE18 is known to be polymorphic[20–23], and having identified five PPE18-derived peptides that exhibited vaccine-induced responses, we next assessed whether these epitopes are conserved across *M. tuberculosis* strains. Using whole-genome sequencing data from approximately 34,000 clinical *M. tuberculosis* isolates[24], we analyzed non-synonymous amino acid variation within the PPE18 gene across major *M. tuberculosis* lineages. Predicted HLA class II core epitopes were mapped to the MTBC reference proteome and evaluated by pairwise sequence alignment to determine epitope conservation.

The five M72 vaccine-responsive peptides had no lineage-associated non-synonymous mutations, indicating that these epitopes are highly conserved across major *M. tuberculosis* lineages (Fig. 4A). Given that PE/PPE proteins comprise approximately 10% of the *M. tuberculosis* coding genome[25] and are implicated in immune modulation and virulence[26,27], we next examined sequence conservation between PPE18 and other PPE family members in the H37Rv proteome. PPE18-derived peptides displayed extensive homology with PPE19 (Rv1361c) and PPE60 (Rv3478) (Fig. 4B). Among the 94 peptides spanning PPE18, 37 were identical to sequences in PPE19, whereas 14 differed by one amino acid. Similarly, 16 peptides were identical to PPE60 sequences, and 24 differed by one amino acid. Across the two pairwise comparisons, this corresponded to 53 perfect-match instances and 38 single-amino-acid-mismatch instances; these counts may include the same PPE18 peptides that are shared with both PPE19 and PPE60 (Fig. 4B). Analysis of perfect and single-amino-acid mismatch alignments further demonstrated that these conserved epitopes are shared predominantly among PPE18, PPE19, and PPE60 (Fig. 4C). Notably, peptides that demonstrated vaccine effect were identical to corresponding sequences in PPE19, PPE60, or both.

Together, these findings demonstrate that the M72 vaccine induces T cell responses against PPE18 domains that are conserved across diverse *M. tuberculosis* lineages and share substantial sequence homology with PPE19 and PPE60, raising the possibility of immune cross-reactivity within this PPE protein subfamily.

Predicted HLA class II presentation of vaccine-responsive PPE18 epitopes

To identify HLA class II molecules that may present the vaccine-responsive domains in the absence of determined HLA types, we performed in silico binding predictions for the five PPE18-derived peptides that exhibited vaccine-induced responses. Because the M72 vaccine is intended for global use, we evaluated peptide binding against the 27-reference human HLA class II allele panel available in the IEDB MHC II Binding Prediction tool (NetMHCIIpan 4.1 EL) using default parameters. Peptides were analyzed individually, and predicted binders were defined using a stringent median percentile rank cutoff of ≤ 5.0, a more conservative threshold than the commonly used ≤ 10 percentile rank for class II binders.

Three of the five peptides (#50, #61, and #67) met this criterion, whereas the strongest predicted binders for peptides #64 and #77 had percentile ranks > 5.0 and were therefore not included (Table 1). We found evidence for presentation promiscuity, as peptides #61 and #67 were predicted to bind *HLA-DRB3*02:02, HLA-DRB1*03:01*, and *HLA-DRB1*13:02*.

Table 1
Predicted HLA binding of the peptides with vaccine-induced boosted responses

Peptide#	Peptide	Allele	median binding percentile	netmhciipan_ el core	netmhciipan_ el score
61	VPPPVIAENRAELMI	HLA-DRB3*02:02	1.2	IAENRAELM	0.416892
		HLA-DRB1*03:01	1.4	VIAENRAEL	0.699264
		HLA-DRB1*13:02	2.5	VIAENRAEL	0.462448
		HLA-DPA1*02:01/DPB1*05:01	3.4	VIAENRAEL	0.084594
		HLA-DRB1*11:01	4.7	VIAENRAEL	0.405934
		HLA-DRB3*01:01	4.8	IAENRAELM	0.109613
67	NTPAIAVNEAEYGEM	HLA-DRB3*02:02	0.13	IAVNEAEYG	0.828831
		HLA-DQA1*03:01/DQB1*03:02	1.6	AIAVNEAEY	0.018174
		HLA-DRB4*01:01	2.2	IAVNEAEYG	0.291368
		HLA-DRB1*13:02	3.6	IAVNEAEYG	0.385124
		HLA-DRB1*03:01	4.4	IAVNEAEYG	0.333719
		HLA-DQA1*05:01/DQB1*02:01	4.9	AIAVNEAEY	0.274658
		HLA-DRB1*04:01	5	IAVNEAEYG	0.366685
50	SAGLMVAAASPYVAW	HLA-DRB1*12:01	2.3	LMVAAASPY	0.452735
		HLA-DRB1*09:01	2.4	MVAAASPYV	0.540128
		HLA-DQA1*05:01/DQB1*03:01	2.5	VAAASPYVA	0.470538

Five peptide sequences that showed a vaccine effect were analyzed for HLA binding predictions using MHC II binding predictions on IEDB (NetMHCIIpan 4.1 EL) with 27 Allele Panel for humans with default settings. The predicted HLA allele cut off was set at median binding percentile = 5.0 and ranked in increasing binding percentile for each peptide sequence. Predicted HLA with greater than 5.0 median binding percentile are not shown.

Overall, these in silico predictions suggest that the vaccine-responsive PPE18 epitopes can be presented by multiple common HLA class II molecules, supporting the potential for broad population coverage of M72-induced CD4 T cell responses across genetically diverse populations.

Discussion

In this study, we defined the peptide-level landscape of T cell responses induced by the M72/AS01$_E$ TB vaccine and identified conserved immunodominant PPE18-derived peptides associated with vaccine-induced boosting. Although previous studies have extensively characterized the magnitude, durability, and multifunctionality of M72-induced immune responses[1,2,6–10], the specific regions of the fusion protein responsible for these responses

remained unknown. By comprehensively mapping 178 overlapping peptides spanning the entire M72 construct, we demonstrate that vaccine immunogenicity is not evenly distributed across the fusion protein but is instead concentrated within discrete regions of the PPE18 component. These findings highlight the importance of epitope-level analyses for resolving vaccine-induced immunity and provide a more refined framework for identifying cellular immune correlates of protection.

A major finding of our study is that the strongest and most frequently recognized domains were derived from PPE18, whereas PepA-derived peptides and sequences spanning the artificial fusion junctions generated limited responses. The absence of responses against junctional peptides suggests that the immune system predominantly recognizes naturally processed *M. tuberculosis*-derived sequences rather than vaccine-specific neoepitopes created during recombinant antigen construction. Furthermore, the identification of five PPE18-derived peptides with vaccine-induced boosting exclusively among M72 recipients demonstrates that only a subset of immunogenic regions contributes to vaccine responsiveness. The level of evidence for any single peptide is limited: low prevalence of responses in the population, unadjusted p-values < 0.05, and FDR-adjusted q-values > 20%. This should be expected since CD4 + T cell responses will be restricted to a small group of participants having the required HLA allele. Still, despite the low power to identify population-level immunodominant responses, there were several peptides with evidence of a response. Notably, the most reactive peptide identified in our study, VPPPVIAENRAELMI, was also reported by Arlehamn et al.[28] as one of the most frequently recognized epitopes in *M. tuberculosis*-infected individuals with no symptoms of TB disease. Although both studies used 15-mer peptides, their libraries differed in overlap length: 11 amino acids in our study versus 10 in Arlehamn et al. This one-residue offset shifted the peptide boundaries, resulting in only partial sequence overlap for three peptides identified in our study: EINSARMYAGPGSAS, NTPAIAVNEAEYGEM, and EASDTAAANQLMNNV. Our results are consistent with observations from other pathogens showing that protective T cell responses are often focused on a limited number of immunodominant epitopes driven by shared HLA alleles[29] rather than broadly distributed across the entire antigen[30–34]. Defining these dominant targets may therefore enable more precise evaluation of vaccine-induced immunity and facilitate discovery of biomarkers associated with protection.

An important implication of our findings is the remarkable conservation of vaccine-responsive PPE18 domains across diverse *M. tuberculosis* lineages. This is consistent with previous reports showing that most *M. tuberculosis* T cell epitopes are hyperconserved[35,36]. Such conservation may reflect functional constraints that limit tolerated sequence variation, limited selective advantage of immune escape, or, potentially, a net benefit to *M. tuberculosis* from certain host T cell responses. The lack of variation could be a result of functional constraints or lack of effective T cell selection pressure. Although PPE proteins exhibit substantial sequence variability[22,27,37,38], none of the five vaccine-responsive epitopes contained lineage-associated non-synonymous mutations in approximately 34,000 clinical isolates. This conservation suggests that M72-induced T cell responses target functionally constrained regions of PPE18 that are preserved despite potential evolutionary pressure exerted by T cell recognition. Moreover, the extensive sequence homology between PPE18-derived epitopes and PPE19/PPE60 raises the possibility of cross-reactive T cell recognition among PPE family members. PE/PPE proteins comprise nearly 10% of the *M. tuberculosis* genome and are among its most immunogenic antigens, yet many exhibit substantial sequence diversity that is thought to contribute to immune evasion[27,39,40]. The identification of conserved immunodominant PPE18 domains recognized following M72 vaccination suggests that protective T cell responses can be directed against regions that are preserved across diverse *M. tuberculosis* lineages despite the overall variability of the PPE protein family. The predicted binding of several vaccine-responsive epitopes to multiple HLA class II molecules further supports the potential for broad population coverage of these responses.

Our study has a number of limitations. First, our analysis was performed using PBMC samples from a Phase 2b clinical trial cohort and focused primarily on IFNg and IL-17 ELISpot responses; additional studies using multiparameter flow cytometry, TCR sequencing, and functional assays will be needed to define the phenotype, clonality, and protective capacity of these epitope-specific T cells. Second, HLA binding predictions were performed in silico because participant HLA genotypes could not be obtained, and experimental validation of peptide–MHC presentation will be required. Finally, while conserved domains identified here represent strong candidates for immune correlates, their association with protection against TB disease requires evaluation in efficacy cohorts. Nonetheless, this study provides the first comprehensive map of immunogenic domains in the M72 vaccine and identifies conserved PPE18 targets that may contribute to protective immunity, offering a foundation for mechanistic studies of M72 efficacy and future TB vaccine development.

Methods

Peptide Synthesis and Reconstitution

15–mer peptides overlapping by 11 amino acids and spanning the full M72 construct (178 total; Fig. 2A) were commercially synthesized at a 20 mg scale using Fmoc chemistry (Biosynthesis, Lewisville, TX) and provided by the Gates Medical Research Institute. Peptides were > 95% pure and supplied lyophilized at 5 mg per vial. Each peptide was reconstituted in DMSO at 5 mg/mL. For ELISPOT assays, peptides were pooled following established methods[19] into 14 pools (24–28 peptides each) at 1 mg/mL or used individually. Pools or single peptides were diluted in RPMI 1640 containing 2 mM L-glutamine, 100 U/mL penicillin, and 100 µg/mL streptomycin to a final working concentration of 2 µg/mL for stimulation assays. Staphylococcal enterotoxin B (SEB; Millipore-Sigma) 1µg/mL was used as a positive control.

Peripheral blood mononuclear cells (PBMC)

PBMC samples were obtained in accordance with the Declaration of Helsinki from a subset of participants who were co-enrolled into the TB018 phase 2b trial of M72/AS01$_E$ and C-041 sub-study, under which samples were collected for the purpose of identifying immune correlates of protection[1,2]. All participants provided written consent, as stipulated in the parent study protocol (Clinical Trial: NCT02097095). Participants were HIV-negative adults, aged 18 to 50 years, with evidence of *M. tuberculosis* infection, as determined by a positive result on the interferon-gamma release assay (QuantiFERON-TB Gold In-Tube assay [QFT, Qiagen]), who had no signs or symptoms of TB disease and who had a negative sputum specimen for *M. tuberculosis* on a polymerase chain reaction (PCR) assay (GeneXpert MTB/RIF, Cepheid) at baseline. The participants received two vaccine doses, 30 days apart. Blood samples were obtained before administration of the first dose of M72/AS01$_E$ (M72: 10 µg; sucrose; polysorbate; TRIS with AS01$_E$: 5µg MPL, 25 µg QS21 in a liposome-based formulation) or placebo (5% Sucrose in phosphate buffer mixed with 23.8 µg α-Tocopherol, 21.4 µg squalene and 9.7 µg Tween 80) (Day 0) and 7 days after the second dose (Day 37) for isolation of PBMC. The parent study protocol was approved by the ethics committees and regulatory authorities in each participating country (Clinical Trial: NCT02097095, registered 2017-02-01).

ELISpot assays

Cryopreserved PBMC were thawed in a 37°C water bath and transferred into a 15ml conical tube with 10mL pre-warmed R10 (RPMI 1640 supplemented with 1% PenStrep, 1% HEPES, 1% L-Glutamine [100x], and 10% FBS). The cells were centrifuged at 900*g* for 5 minutes at room temperature, supernatant discarded, cell pellet dislodged, transferred to a 6-well plate with 5ml R10, and rested for 3–4 hours at 37°C/5% CO_2. Cells were counted based on Trypan blue exclusion criteria. Cells were resuspended in CTL-Test Medium (ImmunoSPOT- supplemented with 1% fresh L-Glutamine) and transferred to a 96-well plate at 200,000 cells/well in 100 µl volume. M72 vaccine peptide pools or individual peptides were added at a concentration of 2 µg/mL. Stimulated cells were cultured for 2 days at 37°C/5% CO_2. For the negative control, triplicate wells containing the same number of cells without antigen stimulation were included for each participant. ELISPOT plates were activated according to the manufacturer's instructions (ImmunoSPOT), and human IFNg /IL-17 capture solution was added and rested overnight at 4°C. The ELISPOT plate was then washed according to the manufacturer's instructions; pre-stimulated cells were transferred from the 96-well culture plate to the ELISPOT plate. Cells were cultured for an additional 24 hours (for a total of 72-hour stimulation). The ELISPOT plate was developed according to the manufacturer's protocol. The developed plate was left to dry and read on the ELISpot machine (S6 Ultra M2 – ImmunoSPOT). Data was cleaned, manually checked, and quality-controlled as needed. Response magnitudes were determined after subtracting background, which was calculated as the mean spot-forming colonies (SFC) of three no-stimulation wells, and responses ≥ 5 SFC/200,000 cells were considered true positive.

Statistical methods for ELISpot peptide response analysis

Biomarkers were evaluated for their ability to detect and quantify a vaccine-induced response. We used rank-based statistics to compare Day 37 and Day 0 responses for a variety of peptide-specific ELISpot responses. A Wilcoxon signed-rank test was used to compute a p-value for the comparison of the Day 37 and Day 0 samples from paired participant samples. To adjust for multiple hypothesis tests across all peptides, we computed false discovery-adjusted q-values (FDR q, Benjamini and Hochberg). To conserve power, we pre-specified that the adjustment would only be computed across peptides that reached a minimum prevalence threshold of > 35% of samples having a detectable response (across participants in both arms and across both time points); as a result, only 11 peptides passed this criteria. Peptides presented in Supp Table 1 were considered to have hypothesis-generating evidence of a vaccine-induced effect with $p < 0.05$, with results requiring further experimental evidence for validation and generalization. Receiver operator curves (ROC) were constructed to assess the performance of a biomarker for classification of the Day 37 and Day 0 samples based on its continuous readout. The paired-AUROC (area under the ROC) is an estimate of the consistency of the sign of the paired differences (e.g., D37 – D0; e.g., all positive or all negative for all participants). Briefly, the differences are computed in one direction (D37 – D0) and assigned an expected result of 1 (positive direction), and then computed in the opposite direction (D0 – D37) and assigned an expected result of 0. Then a standard unpaired AUROC statistic is used on those values (2 x N participants), using the expected result as the outcome and the paired differences as the new predictor. The paired-AUC ranges from 0 to 1, with 0 indicating a consistent decrease at Day 37, 0.5 indicating no separation, and 1 indicating a consistent increase.

PPE18 gene single-nucleotide polymorphism among M. tuberculosis *strains*

We investigated the presence of single-nucleotide polymorphisms (SNPs) in the PPE18 gene in a database of whole-genome sequencing of ~ 34,000 clinical strains[24]. We matched core epitope sequences from MHC Class II binding predictions (https://nextgen-tools.iedb.org/pipeline?tool=tc2) against *M. tuberculosis* complex (MTBC) protein sequences (txid77643). For each match, pairwise local alignment was done between the full predicted

peptide and the matched protein region to assess sequence conservation (https://gist.github.com/SemiQuant/796820dc42075018e5943cbd77aa0a9b).

Epitope mapping to the M. tuberculosis *H37Rv proteome*

PPE18-associated T cell recognition domains (from peptide sequences with a vaccine effect) were aligned to the *M. tuberculosis* H37Rv protein set (Mycobrowser annotated release v4). Each peptide was scanned across the full proteome with Biostrings vmatchPattern. Perfect matches were defined as substring hits with zero mismatches. Single-mismatch hits were defined as hits allowing one substitution, counted only on proteins that lacked a perfect match for that sequence (so one-mismatch tallies do not overlap perfect-match loci). An “overall” summary for loci other than PPE18 reported unique sequences with at least one hit on Rv1361c (PPE19) and/or Rv3478 (PPE60), together with stacked summaries of hit counts.

HLA Prediction

The HLA-binders for the five peptides that demonstrated a vaccine effect were predicted using the Immune Epitope Database (IEDB) [41]. We used the IEDB MHC binding Prediction Tool for class II (https://nextgen-tools.iedb.org/pipeline?tool=tc2), *accessed 03282026*. Peptide sequences were fed into the algorithm with “Human: 27 Allele Panel” and Prediction method: ‘NetMHCIIpan 4.1 EL’ (recommended epitope predictor-2023.09) using default settings. We used the median binding percentile metric to select the HLA-binders, with 5.0 as the cut-off for likely binders for CD4 T cells.

Declarations

Acknowledgements

We thank Nicole Frahm, Thomas Scriba, Peter B. Gilbert, and Margherita Coccia for reviewing the initial draft of the manuscript and for their constructive feedback.

Author Contributions

PO = Study design, Data Analysis, Manuscript Original Draft; JH = Experiment and Data Acquisition, Data Curation, Review and Editing; JDL = Bacterial sequence Data Analysis; DDH, Data Curation and Data Analysis; AJF-G: Data Curation, Data Analysis, Review and Editing; JDE: Funding Acquisition, Supervision, Review and Editing

Funding

This work was funded by the Gates Foundation (INV-039106) grant to J.D.E. *The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Data availability

All data supporting the results of this research are presented in the primary manuscript and its accompanying supplemental files. Further original data and data sources can be obtained from the corresponding author upon justified request.”

References

1. Van Der Meeren, O. *et al.* Phase 2b Controlled Trial of M72/AS01E Vaccine to Prevent Tuberculosis. *N. Engl. J. Med.* 379, 1621–1634 (2018).
2. Tait, D. R. *et al.* Final Analysis of a Trial of M72/AS01E Vaccine to Prevent Tuberculosis. *N. Engl. J. Med.* https://doi.org/10.1056/NEJMoa1909953 (2019) doi:10.1056/NEJMoa1909953.
3. Gillard, P. *et al.* Safety and immunogenicity of the M72/AS01E candidate tuberculosis vaccine in adults with tuberculosis: A phase II randomised study. *Tuberculosis (Edinb)* 100, 118–127 (2016).
4. Kumarasamy, N. *et al.* Long-term safety and immunogenicity of the M72/AS01E candidate tuberculosis vaccine in HIV-positive and -negative Indian adults: Results from a phase II randomized controlled trial. *Medicine (Baltimore)* 97, e13120 (2018).
5. Dagnew, A. F. *et al.* Safety and immunogenicity of investigational tuberculosis vaccine M72/AS01E-4 in people living with HIV in South Africa: an observer-blinded, randomised, controlled, phase 2 trial. *Lancet HIV* 12, e546–e555 (2025).
6. Ji, Z. *et al.* Immunogenicity and Safety of the M72/AS01E Candidate Vaccine Against Tuberculosis: A Meta-Analysis. *Front Immunol* 10, 2089 (2019).
7. Ullah, I. *et al.* The Systematic Review and Meta-Analysis on the Immunogenicity and Safety of the Tuberculosis Subunit Vaccines M72/AS01E and MVA85A. *Front Immunol* 11, 1806 (2020).
8. Leroux-Roels, I. *et al.* Improved $CD4^+$ T cell responses to Mycobacterium tuberculosis in PPD-negative adults by M72/AS01 as compared to the M72/AS02 and Mtb72F/AS02 tuberculosis candidate vaccine formulations: a randomized trial. *Vaccine* 31, 2196–2206 (2013).
9. Spertini, F. *et al.* The candidate tuberculosis vaccine Mtb72F/AS02 in PPD positive adults: a randomized controlled phase I/II study. *Tuberculosis (Edinb)* 93, 179–188 (2013).
10. Thacher, E. G. *et al.* Safety and immunogenicity of the M72/AS01 candidate tuberculosis vaccine in HIV-infected adults on combination antiretroviral therapy: a phase I/II, randomized trial. *AIDS* 28, 1769–1781 (2014).
11. Brennan, M. J. The Enigmatic PE/PPE Multigene Family of Mycobacteria and Tuberculosis Vaccination. *Infect Immun* 85, e00969-16 (2017).
12. Dinges, W. L. *et al.* Virus-specific CD8 + T-cell responses better define HIV disease progression than HLA genotype. *J Virol* 84, 4461–4468 (2010).
13. Gaiha, G. D. *et al.* Structural topology defines protective CD8 + T cell epitopes in the HIV proteome. *Science* 364, 480–484 (2019).
14. de Beijer, M. T. A. *et al.* Discovery and Selection of Hepatitis B Virus-Derived T Cell Epitopes for Global Immunotherapy Based on Viral Indispensability, Conservation, and HLA-Binding Strength. *J Virol* 94, e01663-19 (2020).
15. Moutaftsi, M. *et al.* A consensus epitope prediction approach identifies the breadth of murine T(CD8+)-cell responses to vaccinia virus. *Nat Biotechnol* 24, 817–819 (2006).
16. Lindestam Arlehamn, C. S. *et al.* Memory T cells in latent Mycobacterium tuberculosis infection are directed against three antigenic islands and largely contained in a CXCR3 + CCR6+ Th1 subset. *PLoS Pathog.* 9, e1003130 (2013).
17. Ogongo, P. & Ernst, J. D. Finding antigens for TB vaccines: the good, the bad and the useless. *Nat Med* 29, 35–36 (2023).

18. Ogongo, P. *et al.* Rare variable M. tuberculosis antigens induce predominant Th17 responses in human infection. *JCI Insight* 11, e202134 (2026).
19. Roederer, M. & Koup, R. A. Optimized determination of T cell epitope responses. *J Immunol Methods* 274, 221–228 (2003).
20. Hebert, A. M. *et al.* DNA polymorphisms in the pepA and PPE18 genes among clinical strains of Mycobacterium tuberculosis: implications for vaccine efficacy. *Infect Immun* 75, 5798–5805 (2007).
21. Homolka, S., Ubben, T. & Niemann, S. High Sequence Variability of the ppE18 Gene of Clinical Mycobacterium tuberculosis Complex Strains Potentially Impacts Effectivity of Vaccine Candidate M72/AS01E. *PLoS One* 11, e0152200 (2016).
22. Hakim, J. M. C. & Yang, Z. Predicted Structural Variability of Mycobacterium tuberculosis PPE18 Protein With Immunological Implications Among Clinical Strains. *Front Microbiol* 11, 595312 (2020).
23. Tantivitayakul, P. *et al.* Homoplastic single nucleotide polymorphisms contributed to phenotypic diversity in Mycobacterium tuberculosis. *Sci Rep* 10, 8024 (2020).
24. Phelan, J. *et al.* An open-access dashboard to interrogate the genetic diversity of Mycobacterium tuberculosis clinical isolates. *Sci Rep* 14, 24792 (2024).
25. Zhang, Z., Dong, L., Li, X., Deng, T. & Wang, Q. The PE/PPE family proteins of Mycobacterium tuberculosis: evolution, function, and prospects for tuberculosis control. *Front Immunol* 16, 1606311 (2025).
26. Resstel, C., Madduri, B. T. S. A. & Bell, S. L. Mycobacterial PE/PPE proteins function as 'personal protective equipment' against host defenses. *Front Tuberc* 2, 1458105 (2024).
27. Gan, M., Wang, D., Li, S., Wang, Q. & Liu, Q. Ongoing evolution of PE/PPE genes in Mycobacterium tuberculosis associated with drug resistance and host immune response. *mSystems* 10, e0089825 (2025).
28. Lindestam Arlehamn, C. S. *et al.* A Quantitative Analysis of Complexity of Human Pathogen-Specific CD4 T Cell Responses in Healthy M. tuberculosis Infected South Africans. *PLoS Pathog* 12, e1005760 (2016).
29. Cohen, K. W. *et al.* Trivalent mosaic or consensus HIV immunogens prime humoral and broader cellular immune responses in adults. *J Clin Invest* 133, e163338 (2023).
30. Ramduth, D. *et al.* Immunodominant HIV-1 Cd4 + T cell epitopes in chronic untreated clade C HIV-1 infection. *PLoS One* 4, e5013 (2009).
31. Stenger, R. M. *et al.* Bordetella pertussis proteins dominating the major histocompatibility complex class II-presented epitope repertoire in human monocyte-derived dendritic cells. *Clin Vaccine Immunol* 21, 641–650 (2014).
32. Laher, F. *et al.* HIV Controllers Exhibit Enhanced Frequencies of Major Histocompatibility Complex Class II Tetramer + Gag-Specific CD4 + T Cells in Chronic Clade C HIV-1 Infection. *J Virol* 91, e02477-16 (2017).
33. Tarke, A. *et al.* Comprehensive analysis of T cell immunodominance and immunoprevalence of SARS-CoV-2 epitopes in COVID-19 cases. *Cell Rep Med* 2, 100204 (2021).
34. Wan, X.-F. *et al.* Epitope-spanning antigenic variation reprograms immunodominance and broadens immunity in sequential influenza vaccination. *Nat Commun* 17, 3340 (2026).
35. Comas, I. *et al.* Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved. *Nat Genet* 42, 498–503 (2010).
36. Coscolla, M. *et al.* M. tuberculosis T Cell Epitope Analysis Reveals Paucity of Antigenic Variation and Identifies Rare Variable TB Antigens. *Cell Host Microbe* 18, 538–548 (2015).

37. Karboul, A. *et al.* Frequent homologous recombination events in Mycobacterium tuberculosis PE/PPE multigene families: potential role in antigenic variability. *J Bacteriol* 190, 7838–7846 (2008).
38. McEvoy, C. R. E. *et al.* Comparative analysis of Mycobacterium tuberculosis pe and ppe genes reveals high sequence variation and an apparent absence of selective constraints. *PLoS One* 7, e30593 (2012).
39. Sampson, S. L. Mycobacterial PE/PPE proteins at the host-pathogen interface. *Clin Dev Immunol* 2011, 497203 (2011).
40. D'Souza, C., Kishore, U. & Tsolaki, A. G. The PE-PPE Family of Mycobacterium tuberculosis: Proteins in Disguise. *Immunobiology* 228, 152321 (2023).
41. Kaabinejadian, S. *et al.* Accurate MHC Motif Deconvolution of Immunopeptidomics Data Reveals a Significant Contribution of DRB3, 4 and 5 to the Total DR Immunopeptidome. *Front Immunol* 13, 835454 (2022).

Figures

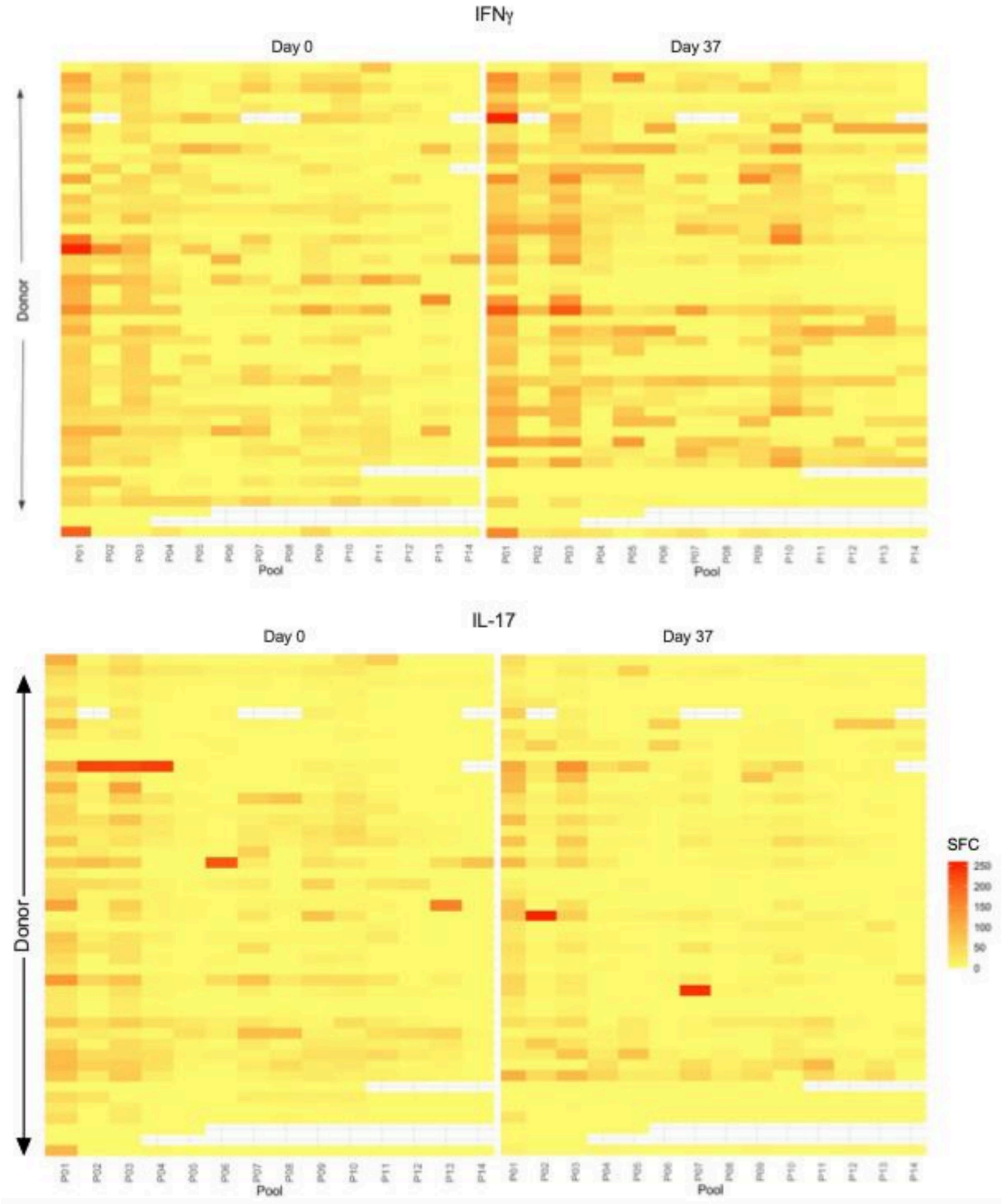


Figure 1

Peripheral blood responses to M72 peptide pools indicate boosted reactivity to specific peptide subsets. A total of 178 15-mer peptides (overlapping by 11 amino acids), 2 mg/mL, covering the entire M72 protein, were combined into 14 pools (24 – 28 peptides per pool) and used to stimulate peripheral blood mononuclear cells (PBMC), followed by detection of IFNg and IL-17 by ELISpot assay. Stimulated cells (200,000 per well) were cultured for 2 days at 37°C/5% CO_2 in a 96-well plate. For control, triplicate wells containing the same number of cells without antigen stimulation were included for each participant. ELISPOT plates were activated according to the

manufacturer's instructions (ImmunoSPOT), and human IFNg /IL-17 capture solution was added and rested overnight at 4°C. The ELISPOT plate was then washed according to the manufacturer's instructions, and pre-stimulated cells were transferred from the 96-well culture plate to the ELISPOT plate. Cells were cultured for an additional 24 hours (for a total of 72-hour stimulation). The heatmaps (IFNg [top] and IL-17 [bottom]) show Spot-Forming Colonies/1,000,000 cells (SFC/M) with donors on the y-axis and the peptide pools on the x-axis, Day 0 (left) and Day 37 (right). White grid areas show participants for whom cells were not available to test all the pools, or for whom the data did not pass quality control.

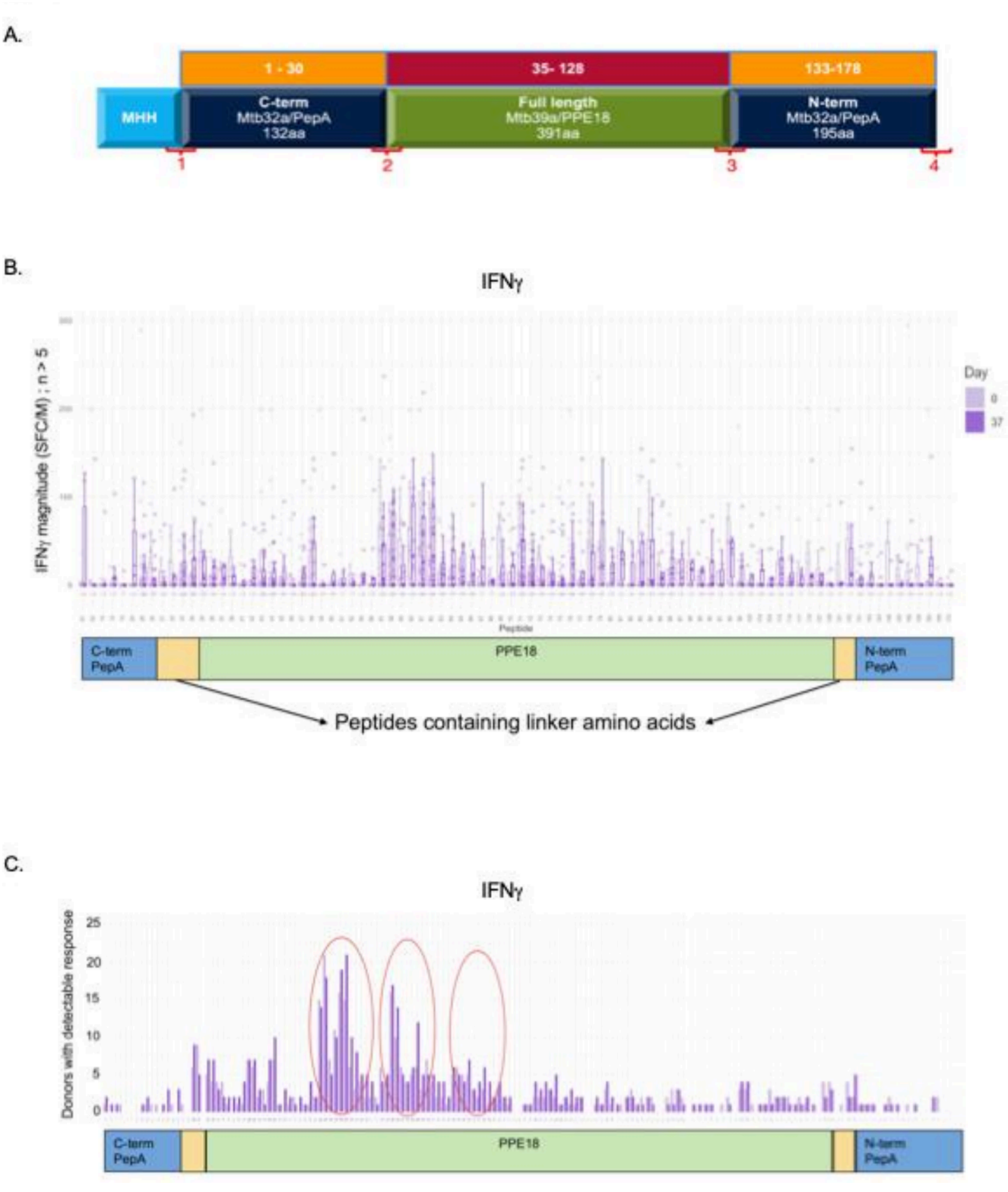


Figure 2

Peptides in the PPE18 (Mtb39A) region of the M72 vaccine construct induce the highest magnitude and prevalence responses. (**A**) Schematic representation of the M72 vaccine construct. Linker regions are highlighted in red, with linkers 2 and 3 joining the two protein subunits of the vaccine (PepA and PPE18). 200,000 PBMC were stimulated with individual peptides (2mg/mL) selected based on peptide pool responses in Figure 1, and IFNg and IL-17 responses were quantified. The individual peptides were then mapped to their position in the M72 vaccine construct. (**B**) IFNg magnitude for peptides assayed in >5 participants (**B**) and the number of participants for whom the peptides elicited a detectable response (**C**). The oval symbols highlight peptide reactivity hotspots (islands).

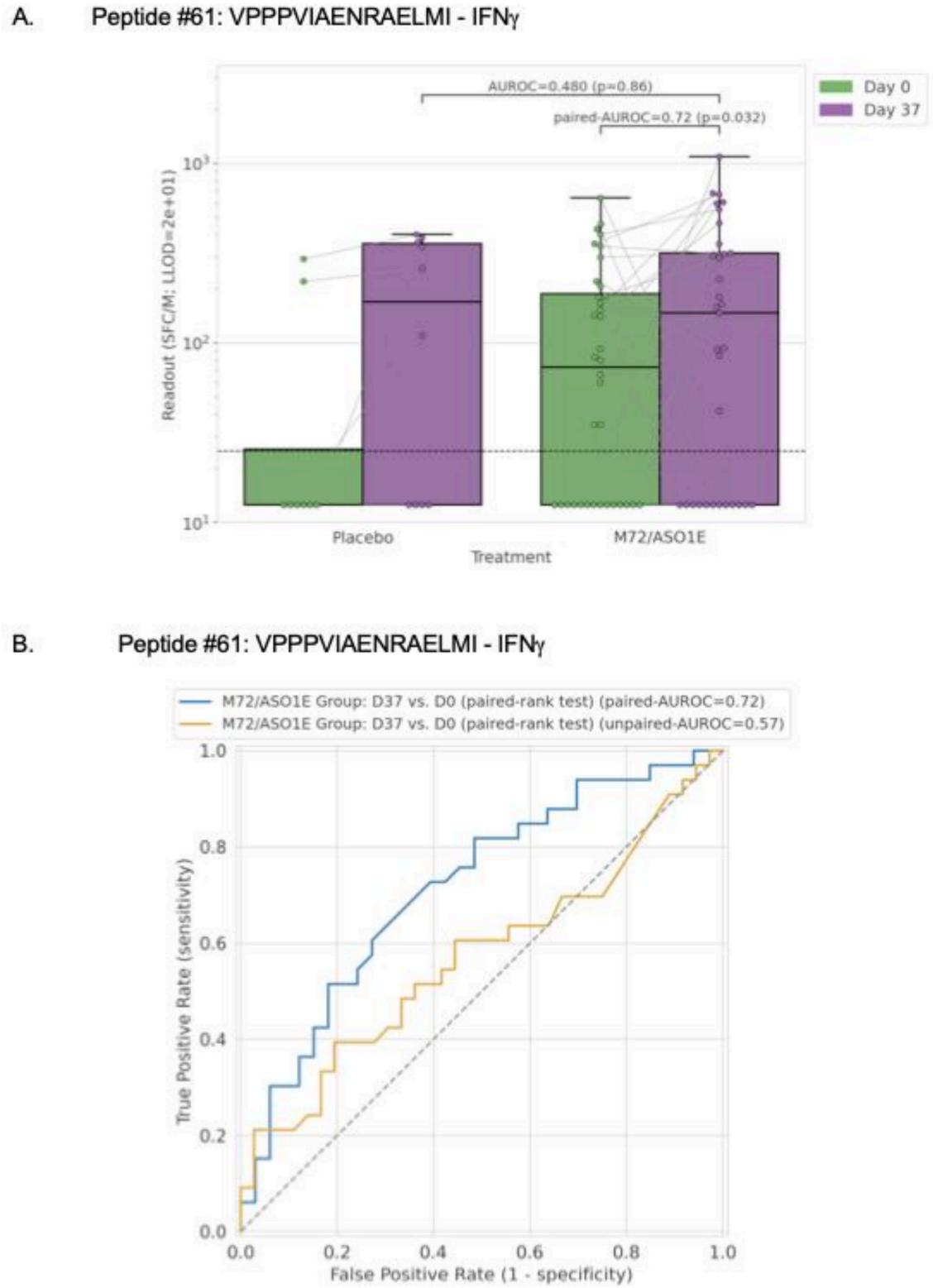


Figure 3

Vaccine-boosted peptide-specific responses. Individual peptide Day 0 and Day 37 responses were compared for the participants with a complete data set in the vaccine arm. The vaccine effect was defined as the difference between Day 0 and Day 37 in the vaccine recipients, and five peptides (all from PPE18) were identified in which the responses between the time points were different, with a p-value <0.1. (**A**) IFNg responses to peptide VPPPVIAENRAELMI (peptide #61), measured in 33 donors. The extent of each boxplot shows the 25th, 50th

(median), and 75th quantiles of the observed responses (SFC/M); whiskers indicate the most extreme observation above or below the median or 1.5 times the interquartile range (whichever is closer to the median). Data are overlaid on top of the boxplots with lines connecting observations from the same participant before and after vaccination. A Wilcoxon signed-rank test was used to compare the pairs of values from each individual, with the paired-AUROC indicating the degree of separation of observations from the two time points (AUROC=0 indicates a consistent decrease at Day 37, AUROC=0.5 indicates no separation, and AUROC=1indicates a consistent increase). (**B**) Receiver operator curves (ROC) show the ability to classify Day 0 from Day 37 samples based on T cell responses to peptide #61. The area under the ROC curve for the paired-sample data (paired-AUC) ranges from 0 to 1, with 0 indicating a consistent decrease at Day 37, 0.5 indicating no separation, and 1indicating a consistent increase. Points along the ROC curve indicate the sensitivity (y-axis) and specificity (one minus the x-axis) for classifying Day 37 from Day 0 samples using various thresholds of biomarker responses, where the biomarker is the SFC/M for an individual peptide.

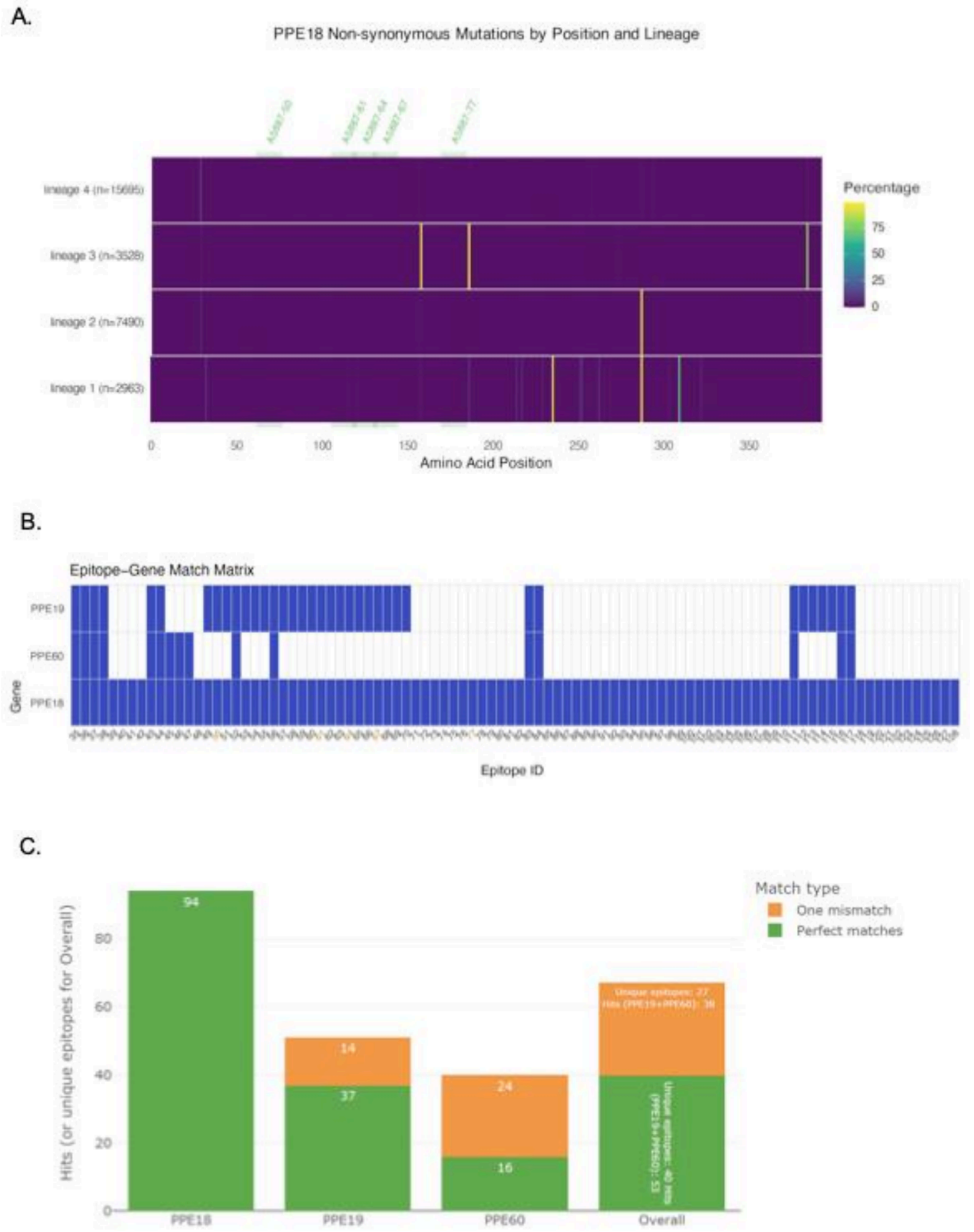


Figure 4

PPE18 peptide responses boosted by vaccination are conserved among *M. tuberculosis* strains and are shared with other PPE proteins. (**A**) PPE18 non-synonymous variation by lineage and amino acid position, in a global collection of ~34,000 clinical isolates, shows conservation in peptides that had boosted IFNg and/or IL-17 responses. (**B**) In silico PPE18-derived epitope matches on proteins in *M. tuberculosis* H37Rv occur in other genes. Rows are H37Rv genes with at least one perfect epitope match; columns are the epitope identifiers. Cells indicate the presence (filled) or absence (empty) of a perfect amino acid match to that gene's protein for the given epitope.

Peptides are shown on the x-axis, with the five peptides that demonstrated vaccine effect color-coded. (**C**) In silico PPE18-derived epitope matches on proteins in *M. tuberculosis* H37Rv occur in other genes. Epitopes were derived from the M72/PPE18 immunogenicity dataset (PPE18 gene only). Each amino acid epitope sequence was searched against the full H37Rv protein set. Bars show summed hit counts on PPE18 (Rv1196), PPE19 (Rv1361c), and PPE60 (Rv3478). The Overall segment height is the number of distinct epitopes with at least one hit on PPE19 and/or PPE60 (union of epitopes mapping to Rv1361c or Rv3478, excluding PPE18-only cases).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial.docx