A conserved human $CD4^+$ T cell subset recognizing the mycobacterial adjuvant, trehalose monomycolate

Yuki Sakai[1,2,#], Minori Asa[1,#], Mika Hirose[3], Wakana Kusuhara[1,2], Nagatoshi Fujiwara[4], Hiroto Tamashima[5], Takahiro Ikazaki[5], Shiori Oka[6], Kota Kuraba[6], Kentaro Tanaka[8], Takashi Yoshiyama[9], Masamichi Nagae[1,2], Yoshihiko Hoshino[10], Daisuke Motooka[8], Ildiko Van Rhijn[11,12,13], Xiuyuan Lu[2], Eri Ishikawa[1,2], D. Branch Moody[11], Takayuki Kato[3], Shinsuke Inuki[6,7], Go Hirai[5] and Sho Yamasaki[1,2,14,*]

[1]Department of Molecular Immunology, Research Institute for Microbial Diseases, Osaka University, Suita 565-0871, Japan
[2]Laboratory of Molecular Immunology, Immunology Frontier Research Center (IFReC), Osaka University, Suita 565-0871, Japan
[3]Laboratory of Cryo-EM Structural Biology, Institute for Protein Research, Osaka University, Suita 565-0871, Japan
[4]Department of Food and Nutrition, Faculty of Contemporary Human Life Science, Tezukayama University, Nara, 631-0034, Japan
[5]Graduate School of Pharmaceutical Sciences, Kyushu University, Fukuoka 812-8582, Japan
[6]Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto 606-8501, Japan
[7]Graduate School of Biomedical Sciences, Tokushima University, Tokushima 770-8505, Japan
[8]Genome Information Research Center, Research Institute for Microbial Diseases, Osaka University, Suita 565-0871, Japan
[9]Respiratory Disease Center, Fukujuji Hospital, Japan Anti-Tuberculosis Association, Tokyo 204-8522, Japan
[10]Department of Mycobacteriology, Leprosy Research Center, National Institute of Infectious Diseases, Higashimurayama, Tokyo189-0002, Japan
[11]Division of Rheumatology, Immunity and Inflammation, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA
[12]Faculty of Veterinary Medicine, Department of Infectious Diseases and Immunology, University Utrecht, 3584CL, Utrecht, Netherlands

[13]Department of Medical Biology, Amsterdam University Medical Center, 1105AZ, Amsterdam, Netherlands
[14]Center for Infectious Disease Education and Research (CiDER), Osaka University, Suita 565-0871, Japan

[#]These two authors contributed equally to this manuscript.

*Correspondence: Sho Yamasaki, Department of Molecular Immunology, Research Institute for Microbial Diseases (RIMD), Immunology Frontier Research Center (IFReC), Osaka University3-1 Yamadaoka, Suita, 565-0871, Japan. Phone: +81-6-6879-8306; Email: yamasaki@biken.osaka-u.ac.jp

Abstract

Mycobacterium tuberculosis causes human tuberculosis. As mycobacteria are protected by thick lipid cell wall, humans have developed immune responses against diverse mycobacterial lipids. Most of these immunostimulatory lipids are known as adjuvants acting through innate immune receptors, such as C-type lectin receptors. Although a few mycobacterial lipid antigens activate unconventional T cells, antigenicity of most adjuvantic lipids are unknown. Here, we identified that trehalose monomycolate (TMM), an abundant mycobacterial adjuvant, activates human T cells bearing a unique αβTCR. This recognition was restricted by CD1b, a monomorphic antigen-presenting molecule conserved in primates but not mice. Single-cell TCR-RNA sequencing using newly established CD1b-TMM tetramers revealed that TMM-specific T cells are present as $CD4^+$ effector memory T cells in the periphery of uninfected donors, but express IFNγ, TNF and anti-mycobacterial effectors upon TMM stimulation. TMM-specific T cells are detected in cord blood and PBMCs of non-BCG-vaccinated donors, but are expanded in active tuberculosis patients. A cryo-electron microscopy study of CD1b-TMM-TCR complexes revealed unique antigen recognition by conserved features of TCRs, positively-charged CDR3α and long CDR3β regions. These results indicate that humans have a commonly-shared and pre-formed $CD4^+$ T cell subset recognizing a typical mycobacterial adjuvant as an antigen. Furthermore, the dual role of TMM justifies reconsideration of the mechanism of action of adjuvants.

Introduction

Diseases caused by mycobacteria, including tuberculosis, leprosy, Buruli ulcer and nontuberculous mycobacterial (NTM) lung disease, rank among the top causes of death and disability worldwide. Mycobacteria are distinguished from other bacteria by a thick cell envelope comprised of unique outer membrane of neutral lipids and glycolipids, which forms the primary barrier against the host and activates host immunity (1). The mycobacterial cell envelope, provided as a mixture of compounds, has been broadly administered in vivo in animals as ‘complete Freund’s adjuvant’ (CFA) that can promote a strong vaccine response (2). Yet, limited understanding of the defined immunogenic components of CFA has to date prevented its therapeutic use as a vaccine adjuvant in humans (3). Recently, the receptors for some mycobacterial immunogens have been identified, which involve *N*-acetyl muramyl dipeptide (MDP)–NOD2 and trehalose mono/dimycolate (TMM/TDM)–Mincle axes (4-6). However, there remain many mycobacterial lipids for which receptors and mechanisms of action are not known.

In this study, we sought to identify mycobacterial cell wall components leading to immune activation ex vivo. After demonstrating strong T cell activation, we purified the stimulatory component and established its structure as TMM. T cell activation by TMM was not mediated by Mincle but unique TCRs restricted by CD1b. Using TMM-loaded CD1b tetramers and single-cell TCR-RNA sequencing (sc-TCR-RNA-seq), we identified a naturally occurring memory T cell population, which exists in cord blood and uninfected human subjects but is expanded in humans during *M. tuberculosis* infection. Structural analysis identified unique TCR motifs that are shared across unrelated humans and mediate TMM recognition. Finally, clonotype tracking revealed that TMM-specific T cells produce typical mycobactericidal effectors including granulysin (7), granzyme B (8), IFNγ (9) and TNF (10, 11) when stimulated with TMM.

Results

Identification of mycobacterial lipid-reactive T cells in human PBMCs

As a relatively unbiased and comprehensive way to search for defined immunostimulatory components in the mycobacterial cell envelope, PBMCs from healthy donors were stimulated with total *M. tuberculosis* compounds extracted into chloroform/methanol (2:1, vol/vol) and coated onto tissue culture plates (12). Based on cell trace violet (CTV) dilution, we subjected mycobacterial lipid-responsive T cells to single-cell TCR-RNA sequencing (scTCR-RNA-seq) to identify potentially diverse clonotypes along with their effector function and TCR profiles (Figure 1A)(GSE260931). Among 26,502 detected clonotypes, we selected 52 clonotypes that are expanded by *M. tuberculosis* lipids and reconstituted their TCRαβ pairs in NFAT-GFP reporter cells (Figure 1A and Supplemental Table 1).

After surface expression of the TCR complex was confirmed for 44 clonotypes (Figure 1B), we tested these cells for responses to plate-coated *M. tuberculosis* lipids in the presence of cytokine-differentiated human monocytes as antigen-presenting cells (APCs). One clonotype derived from $CD4^+$ T cells, Y-50, responded strongly to mycobacterial lipids based on GFP and CD69 upregulation (Figure 1C). Analysis of scTCR-RNA-seq data revealed that the Y-50 clonotype was expressed by fourteen individual cells within the $CD4^+$ T cell clusters (Figure 1D). These cells were characterized by the expression of granzyme B (*GZMB*), perforin-1 (*PRF1*), granulysin (*GNLY*), TNF (*TNF*) and IFNγ (*IFNG*), regardless of the expression level of *CD4* (Figure 1E and Supplemental Figure 1). Y-50 also expressed innate-like T cell markers, like

CD161/killer cell lectin like receptor B1 (*KLRB1*) and CCAAT enhancer binding protein delta (*CEBPD*) (Figure 1E). Some Y-50 cells were also detected in a Ki67^{hi} proliferating CD4^{+} cluster (Figure 1D), in agreement with their CTV^{lo} status used for sorting (Figure 1A). These results suggest that mycobacterial lipid-reactive Y-50-expressing CD4^{+} T cells exhibit an innate and cytolytic signature after ex vivo lipid stimulation.

TMM activates the Y-50 clonotype T cells

To identify the lipid stimulus, we separated crude lipids by thin layer chromatography (TLC) and measured responses to each fraction. We collected 16 fractions and found potent antigen activity (Figure 2A). The active peak shifted to lower retention factor (Rf) values (fraction 2) under more hydrophobic solvent conditions (Figure 2B), suggesting that the antigenic component was likely a moderately polar lipid. We thus analyzed the fraction by matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS) (Figure 2C), finding ions that matched and largely overlapped in chain length and saturation patterns with purified trehalose monomycolate (TMM) from *M. tuberculosis* H37 Rv (13) (Figure 2D and Supplemental Figure 2). In addition to purified TMM, Y-50 TCR-expressing reporter cells were also activated by APCs cocultured with *M. tuberculosis* H37 Rv and *M. bovis* BCG (Figure 2E), demonstrating the origin of the stimulus from intact bacteria.

CD1b restricts TMM recognition by the Y-50 TCR

TMM is most well known as a major cell wall glycolipid with adjuvanticity that is also a biosynthetic intermediate to TDM, known as cord factor. As TDM and TMM were both known to potently activate myeloid cells through the innate receptor, Mincle (12, 14), TMM might activate T cells via Mincle on APCs. However, the Y-50 T cells selectively recognized TMM but not TDM (Figure 2F), suggesting that the response was specific to some aspects of the TMM structure and was not Mincle mediated.

The major alternative hypothesis was CD1 presentation of TMM to TCRs, as prior studies reported that another mycobacterial glycolipid, glucose monomycolate (GMM), was a CD1-restricted T cell antigen (15). Yet, GMM was not a Y-50 antigen (Figure 2F), suggesting that TMM might be a new T cell antigen presented by CD1.

As all four types of human CD1 antigen presenting molecules can present lipids (16), we examined the effect of blocking antibodies against human CD1a, CD1b, CD1c and CD1d and found that only anti-CD1b selectively suppressed TMM-induced activation of Y-50 reporter cells in the presence of APCs (Figure 3A). Conversely, ectopic expression of CD1b on HEK293T cells conferred Y-50 TCR reactivity to TMM (Figure 3B). Thus, only CD1b is necessary and sufficient for the presentation of TMM to the Y-50 TCR.

Clone Y-50 broadly recognizes TMM from various mycobacterial species

TMM is produced broadly among mycobacterial species (17). To further characterize the selectivity of antigenic lipids recognized by the Y-50 TCR, we purified TMM possessing different lipid lengths from *M. intracellulare* and *M. smegmatis* (C60-

C88) as well as *Rhodococcus* species with shorter mycolate moieties (C28-C36), finding that all showed antigenic activity (Figure 3C). However, other related mycolyl lipids that varied in the head group moiety, GMM, mannose monomycolate (MMM), glycerol monomycolate (GroMM) and free mycolic acid (MA) (18, 19), lacked antigenicity (Figure 3C). Thus, the T cell reactivity identified here was new, and the head group comprised of the trehalose disaccharide was required for Y-50 TCR recognition.

To exclude the possibility of contaminants or mitogens in natural TMM preparations, we carried out complete synthesis of TMM, starting with hexa-*O*-TMS trehalose 6,6′-diol (Figure 3D). Synthetic TMM also induced T cell activation, formally ruling in this structure as an antigen (Figure 3E). However, other synthetic TMM analogues lacking an α-branched alkyl chain or a β-hydroxy group did not activate (Figure 3E). Thus, while Y-50 broadly recognizes TMM mycolate chains of varying length present across mycobacterial species, it discriminates the chemical features that define mycobacterial TMM.

Cationic residues in Y-50 TCR are critical for TMM recognition

A striking characteristic of Y-50 TCRα is the presence of four positively-charged arginine residues (R107, R108, R113 and R114) within the CDR3α region (Figure 4A). To investigate the contribution of these residues to TMM recognition, we introduced alanine mutations and evaluated their effects using TCR-reconstituted reporter cells. TCRα containing alanine substitutions at R107, R113 or R114 showed impaired reporter activity, whereas R108A had no impact (Figure 4A).

Compared to the average CDR3β (14.4 residues) (20), the Y-50 CDR3β was much longer, encoded by 20 residues (Figure 4B). To assess the contribution of amino acid insertion during VDJ recombination to antigen recognition, we engineered four shorter Y-50 TCRβs lacking residues within the junctional region. None of these TCRβ mutants recognized TMM (Figure 4B), suggesting that certain aspects of this long CDR3β are required for the recognition of TMM by the Y-50 TCR.

Structural characterization of the Y-50 TCR

The cationic and long loop motifs suggested a TCR binding mechanism controlled by electrostatic interactions and a flexible TCR surface. To gain structural insight into the Y-50 TCR, a soluble TCRαβ was constructed for crystallization. We obtained a crystal structure of the TCRαβ complex (PDB: 8XUB) that diffracted to a resolution of 2.5 Å (Figure 4C and Supplemental Table 2). Three CDR3α arginine residues that were found to be critical for TMM recognition were facing toward the TCRβ, whereas R108, which was dispensable, was oriented away from the TCRα-β interface (Supplemental Figure 3A). The electron density of the TCRαβ was clear except for the CDR3β loop region, implying that the extra-long CDR3β loop may be highly flexible, as hypothesized (Figure 4, D and E).

Determination of ternary complex structure of Y-50 TCR-TMM-CD1b

However, solving the definitive recognition mechanism required a ternary structure, so we conducted cryo-electron microscopy (cryo-EM) analysis. Recombinant

CD1b was refolded with synthetic TMM (Supplemental Figure 3B) and incubated with soluble Y-50 TCRαβ. The cryo-EM map of the ternary complex was successfully reconstructed to a resolution of 3.18 Å (PDB: 8ZOX)(Figure 5A, Supplemental Figure 3, C-E, Supplemental Table 3, and Supplemental Video 1). In the area between TCRαβ and CD1b, we observed clear density map that exactly overlapped with the chemical structure of TMM (Supplemental Figure 3F). TMM lipid chains were buried inside CD1b pockets, like GMM (21), with its sugar head group exposed toward TCR (Supplemental Figure 3G). TCRαβ tightly contacted with its protruding bulky trehalose moiety through CDR3 regions (Supplemental Figure 3G).

This observation was supported in detail by the superimposition of the structure of the Y-50 TCR alone (Figure 4C, PDB: 8XUB) with its structure within the ternary complex (Figure 5A, PDB: 8ZOX). The positions of backbone Cα atoms in both structures largely overlapped; however, the location of the long CDR3β region, was noticeably shifted upon TMM-CD1b contact (Figure 5B). Whereas the CDR3β loop hung 'downwards' in the structure without antigen, it was 'drawn up' like a curtain to allow the recognition of TMM presented by CD1b (Figure 5B), providing detailed insight into how the long CDR3β creates the flexible TCR interface with CD1b-TMM complex by avoiding steric hindrance.

We next determined the mode of specific TMM recognition by Y-50 TCR. The experimentally observed functional importance of cationic CDR3α residues (Figure 4A) was explained by the ternary structure. R114 of TCRα chain formed a hydrogen bond with β-hydroxy group of TMM (Figure 5C), likely explaining both the strong effects of alanine mutation (Figure 4A) and altered recognition of TMM lacking the β-hydroxy group (Figure 3E). To further test the significance of this interaction, we synthesized TMM stereoisomers that differed in the stereochemistry of the acyl group (Supplemental Figure 4). Y-50 TCR recognized natural TMM (*R,R*) but not non-natural isomers (Figure 5D). Overall, the Y-50 TCR is highly specific for the natural stereo-configuration of the TMM lipid moiety through interaction with the cationic residue R114.

Trehalose is a diglucose. Another critical cationic residue, TCRα R107, formed a hydrogen bond with hydroxy group at C-2′ atom of the distal glucose in trehalose, while the proximal glucose interacted with TCRβ D114 through hydroxy group at C-4 atom (Figure 5E). Furthermore, α^{R107} or β^{D114} formed a salt bridge (Figure 5E). Thus, Y-50 TCRα and TCRβ may cooperate for TMM recognition by interacting with distinct epitopes. In line with this interpretation, Y-50 TCR recognition of TMM was impaired in α^{R107A} or β^{D114A} single mutation and more severely in double mutation (α^{R107A}-β^{D114A}) (Supplemental Figure 5A).

These structural analyses further revealed the molecular basis by which the characteristic features of Y-50 TCR recognize TMM-CD1b complex. Alanine scanning functionally confirmed this mode of antigen recognition and found additional critical residues that are involved in the interaction with CD1b (Supplemental Figure5, B-D). For example, TCRβ G110 was also important (Supplemental Figure 5D) due to an interaction with E80 of CD1b (Figure 5F). In addition to CDR3β, other regions derived from the TRBV4-1-encoded TCR Vβ chain also bound the surface of CD1b itself. A positively charged R37 in CDR1β formed a salt bridge with a negatively charged D83 of CD1b.

CDR2β also interacted with CD1b α1 helix through a salt bridge (E63–R79) and a hydrogen bond (Y58–E80) (Figure 5, F and G, and Supplemental Video 2). Thus, the TRBV4-1-encoded TCR Vβ loops interact with a triply charged patch on CD1b defined as an ^{79}RExxD^{83} sequence in an antigen-independent manner (Figure 5G). This charged patch can explain the preferential usage of TRBV4-1 in CD1b- and also CD1c-restricted TCRs (22-24), as an identical ^{79}RExxD^{83} motif is also found in CD1c, but not other human CD1 isoforms (Figure 5H).

Characterization of TMM-specific T cells in the periphery using tetramers

Key questions relating to whether pre-primed memory cells recognizing CD1b exist in humans, as well as their potential effector function in the periphery without in vitro expansion were largely unknown. We therefore sought to generate TMM-loaded CD1b tetramers and combine them with single cell analysis. These TMM-CD1b and unloaded control tetramers were validated by binding to cell lines expressing TCRs and TMM-stimulated human PBMCs (Supplemental Figure 6). First, we investigated gene expression profiles of freshly-isolated and TMM-stimulated T cells bearing Y-50 TCR using scTCR-RNA-seq. Y-50 T cells were separated into different clusters before and after stimulation in UMAP plot (Figure 6A), implying that gene expression signatures were altered by antigen stimulation. Before ex vivo stimulation, Y-50 T cells expressed typical effector memory markers, such as CD44 (*CD44*), IL-7 receptor α chain (*IL7R*) and integrin β_1 (*ITGB1*) (Figure 6B, and Supplemental Figure 7, A and B), suggesting that resting TMM-specific T cells exist as naturally occurring memory T cells, which is the characteristic feature of NKT and other innate T cells (25). Upon TMM stimulation, Y-50 T cells moved to the cluster characterized by the expression of cytotoxic and bactericidal effector genes (*GZMB, PRF1, GNLY*) (7, 8) and anti-mycobacterial protective cytokines and chemokines (*IFNG, TNF* and *CCL5*) (26, 27), whereas the expression levels of stemness-related molecules, such as *IL7R*, *TCF7* and *CXCR4*, were downregulated (Figure 6, C-E). These results suggest that Y-50 T cells are innate-like T cells exhibiting anti-mycobacterial potential in the periphery, whose effector signature is markedly enhanced upon antigen stimulation.

TMM-specific T cells with similar TCR motifs are shared across humans

To examine whether TMM-specific T cells are shared across genetically unrelated individuals, we sorted TMM-CD1b-tet$^+$ cells from fresh PBMCs of additional donors and examined their characteristics by scTCR-RNA-seq. TMM-CD1b-tet$^+$ clonotypes from unrelated donors were mainly localized within the CD4$^+$ effector memory clusters (Figure 7, A and B), which expressed CD44, IL-7 receptor or integrin β_1 but not homing receptor CCR7 (Supplemental Figure 7C), similar to unstimulated Y-50 T cells (Supplemental Figure 7B). TCRs expressed by these clonotypes were reconstituted in reporter cells and confirmed their specific reactivities to TMM (Figure 7, C and D). Importantly, sequences of these TCRs showed them not be identical to Y-50 but they possessed similar characteristics, including positively-charged CDR3α sequences, biased TCR Vβ that is encoded by TRBV4-1 and long CDR3β sequences (Figure 7C). Further, these trends could be seen in comprehensive TCR analysis of TMM-

CD1b tetramer-sorted T cells (Figure 7, E and F). Thus, TMM and CD1b-reactive T cells show clear evidence for conserved features across numerous clonotypes from different donors, constituting a new public TCR-antigen linkage that establish a donor-unrestricted T cell subset in humans (28).

Quantification of TMM-specific T cells during Mtb infection

Finally, we used tetramers to examine the frequency of TMM-specific T cells in PBMCs from uninfected and active tuberculosis (TB) donors that were recruited consecutively based on smear positive and *M. tuberculosis* culture positive sputum samples (Supplemental Table 4). TMM-CD1b-tet^{+} T cells were detected in most uninfected donors, consistent with prior experiments of single cell or tetramer-based outcomes (Figure 6 and Figure 7), further suggesting that these cells are preformed innate-type T cells, as these frequencies are similar to those of other unconventional T cells (29). The frequency was significantly increased in active TB patients (Figure 7G), suggesting that these T cells may react to TMM during mycobacterial infection in the host. Since these samples were from Japanese healthy donors who had received BCG vaccination, we also examined PBMCs and cord blood cells of healthy donors from North America where BCG is no longer widely used. The frequency of tetramer^{+} T cells was comparable among all three uninfected groups (Figure 7G), indicating that TMM-specific T cells are present in naïve donors and developed without exposure to BCG or other environmental antigens.

Discussion

A basic paradigm for adaptive MHC-restricted T cells is that naïve cells are primed by antigen to differentiate to memory T cells that persist in elevated numbers with memory markers in the periphery. In contrast, CD1d-restricted iNKT cells express memory markers in the absence of defined antigenic stimulation, and functionally circulate in larger numbers and respond rapidly as a cohort to antigen challenge (25, 30). The extent to which human group 1 CD1 (CD1a, CD1b and CD1c)-restricted T cells behave as naïve or memory T cells in the periphery, remains poorly understood owing mainly to technical challenges in directly addressing these questions in humans and limited tractable infection models that express CD1a, CD1b and CD1c (which are lacking in mice) (31). For example, prior studies have emphasized long term in vitro cultured T cell clones (7, 15, 18, 32) or indirect detection of T cells by activation assays (33, 34) rather than tetramer-based scTCR-RNA-seq analysis. Through the discovery of TMM as a T cell antigen, generation of CD1b-TMM tetramers applied across unrelated donors, identification of new binding motifs and structural dissection of lipid and non-lipid interaction by cryo-EM ternary complex, this study advanced our understanding of human pathogen-specific T cell responses in the CD1b system.

Our key findings were the identification of mycobacteria-specific TCRs in peripheral T cells without infection, and the rapid induction of anti-mycobacterial effectors by stimulation ex vivo. IFNγ and TNF are canonical anti-mycobacterial protective cytokines produced from CD4$^+$ Th1 cells (9-11). Recently, granulysin, granzyme B and perforin secreted from CD8$^+$ T cells are also recognized to be important for protection against mycobacterial infection (26, 35). TMM-specific T cells are a unique cell subset that rapidly upregulated all of these effectors simultaneously upon antigen stimulations. While host protection is difficult to demonstrate directly in human experimental systems as discussed below, this evidence supports that TMM-reactive T cells express a host protective effector function.

Although TMM-reactive T cells expressed CD4, this coreceptor seems dispensable for the recognition of CD1b-TMM complex, as the reporter cells used in our assay do not express human CD4. However, we cannot fully exclude the possibility that, like conventional T cells, CD4 is required for those T cells to be selected by MHC class II. Clinically, the well-known susceptibility of human immunodeficiency virus (HIV)-infected patients to tuberculosis resulting from the reduction of $CD4^+$ T cells (36) might also be partly due to the loss of $CD4^+$ TMM-specific T cells.

Known patterns of TMM biosynthesis and expression support plausible scenarios for TMM antigen function during infection. TMM is expressed by most mycobacterial species and is used for further biosynthesis of other cell wall components, including arabinogalactan and TDM (37). Unlike TDM, which is downregulated in mycobacteria upon infection of the host, the level of TMM is relatively constant (38). Compared with GMM, TMM may be resistant to stresses such as oxidation because the reducing ends of both glucoses are occupied. Thus, given the importance of TMM for multiple stages of the mycobacterial life cycle, the presence of T cells that recognize TMM with various lipid chain lengths plausibly could allow effective induction of responses to a broad spectrum of mycobacteria.

Taking advantage of the direct detection by tetramers and single cell analysis, we provide several lines of evidence that TMM-specific T cells exist before the host is exposed to mycobacteria, as they were detected in random blood donors, as well as non-TB or non-BCG vaccinated donors and even in cord blood cells. However, as contrasted with NKT and MAIT cells, *PLZF* was not highly expressed in TMM-specific T cells, so it is unclear whether they were selected by DP thymocytes like other innate-like T cells (39, 40). Even assuming the involvement of CD1b for selection, the selecting ligand(s) is unclear. It is also possible that the intrinsic affinity of TRBV4-1 to CD1b patch might allow less ligand-dependent positive selection.

Determination of the ternary structure of TMM-specific TCR provides similarities and differences in the mode of glycolipid recognition with previously reported glycolipid-specific T cells (21). CDR3α loop regions of both TCRs interact with β-hydroxy group of GMM and TMM, which is a defining chemical feature of foreign mycolic acids as contrasted with self fatty acids, which allows T cells to discriminate the natural configuration of mycolyl lipids. Compared to the monosaccharide in GMM, two sugar moieties of TMM interacted more extensively with TCR residues, likely determining T cell antigen specificity (Supplementary Video 3). Furthermore, long CDR3β region uniquely found in TMM-specific TCRβ and its demonstrated compression in the ternary structure show how the TCRβ chain moves upwards to accommodate the bulky TMM head group presented by CD1b.

Biased usage of TRBV4-1 has been reported for CD1b-restricted T cells from blood (22, 23) and tuberculous pleural effusions (41), suggesting that such TCRs are clinically relevant to host response in TB disease and could be biomarkers of *M. tuberculosis* infection. However, a detailed mechanism underlying this preference was unclear, as ternary complex structure of TRBV4-1$^+$ TCR with CD1-bound antigens had never been solved. Our cryo-EM structure provides direct evidence for the presence of an antigen-independent 'patch' by which TRBV4-1-encoded residues interact with CD1b. Conservation of this motif among CD1b and even CD1c, but not in other human CD1 molecules, may support the high frequency of TRBV4-1 in CD1b- and CD1c-restricted T cells (24). The first identification of CD1b/c 'patch' implies the presence of any other motifs for known biased TCR Vβ genes, such as TRBV7-9$^+$ T cells restricted by human CD1c (42) or CD1d which are associated with Crohn's disease (43-45).

TMM is recognized by pattern recognition receptor Mincle (12). Thus, this mycobacterial lipid represents a 'dual ligand' that can activate both pattern recognition receptors (PRRs) and TCRs, so our data suggest that TMM can act simultaneously as a PAMP and an T cell antigen, respectively. Freund's adjuvant cannot be used as a human therapeutic due to its bacterial origin and undefined mechanism. However, chemically defined, dual acting molecules like the synthetic TMM studied here could be promising therapeutic options as both adjuvant and antigen, to prevent various diseases caused by mycobacterial species. In addition to tuberculosis, NTM lung disease is one of the most urgent targets as cases are dramatically increasing and current drug treatments are ineffective (46). TMM from NTM species are also – and more strongly – recognized by TMM-specific T cells. Furthermore, the potent activity and higher hydrophilicity of short-chain TMMs could be advantageous in terms of efficacy, formulation and administration.

Detailed analysis on the structure-activity relationships and the protective role of TMM-specific T cells against mycobacterial infection using primate models will contribute to the establishment of treatment and prevention options.

Methods

Sex as a biological variable

Our study examined human PBMCs from both male and female donors. Sex was not considered as a biological variable.

Human subjects

PBMCs from healthy donors were collected after obtaining informed consent. Peripheral and cord blood cells from healthy donors were also obtained from Veritas Corporation (Tokyo, Japan; Batch 210570303C, 220771201C, 220771404C, 220772503C, 220781001C, 2208409001, 2208411000, 220873101C, 220880801C, 220881703C (PBMC); Batch 2211410002, 2211416002, 2211422002, 2211422002, 2211422003, 2211423000, 2211423001, 2212406005, 2212414001, 2212414003, 2212420000 (CBMC)). Active tuberculosis cases (13 cases) are those that were admitted to the hospital as sputum smear and culture positive pulmonary tuberculosis cases that consecutively included in the study from April 2023 to July 2023. All participants were enrolled after giving written informed consent. Blood was taken for active tuberculosis before starting treatment.

Bacteria

M. tuberculosis strain H37Rv was kindly provided by Dr. Ikuya Yano. For inactivation, the bacterium was heated at 65°C for 1 h, followed by incubation at 60°C for overnight. *M. bovis* BCG was purchased from Japan BCG Laboratory.

Lipid extraction and purification for stimulation

M. tuberculosis strain H37Rv lipids were extracted as previously described (47). Briefly, 10 ml of chloroform/methanol (2:1, vol/vol) or acetone were added to 100 mg of bacteria and sonicated at 40°C for 10 min. The organic phase was collected and dried and dissolved in chloroform/methanol (2:1, vol/vol) for storage and aliquoting into various assays as a crude lipid. For lipid fractionation, crude lipids were separated by high performance thin layer chromatography (Merck) followed by charring with copper(II) acetate-phosphoric acid. TMM was purified from *M. tuberculosis* H37Rv, *M. tuberculosis* CDC1551, *M. bovis* BCG, *M. intracellulare*, *M. smegmatis*, *R. equi* and *R. sp* 4306, TDM was purified from *M. tuberculosis* CDC1551, GMM and MMM was purified from *R. ruber*, and GroMM was purified from *M. bovis* BCG as previously described (13, 48-52). Briefly, the heat-killed bacteria were sonicated in chloroform/methanol (2:1, vol/vol) for 15 min on ice, and water was added (1/20 total volume). The organic layer was collected and evaporated completely. The crude lipids were separated by thin-layer chromatography (Merck) and fractions were extracted. Mycolic acid was purified from *M. tuberculosis* H37Rv as described (47). Synthetic GMM was provided by Adriaan Minnaard (Groningen University) (53). For stimulation of cells, lipids dissolved in chloroform/methanol (2:1, vol/vol) were diluted in isopropanol, applied to 96-well plates at 20 μl/well and air-dried prior to adding media.

MALDI-TOF MS analysis

TMM was detected by matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) with an UltrafleXtreme (Bruker Daltonics, Billerica, MA). In brief, purified lipid fractions and TMM standards were dissolved in chloroform/methanol (3:1, vol/vol) at a concentration of 1 mg/m1, and 1 μl of sample was applied directly to the sample plate, followed by addition of 1 μl 2,5-dihydroxybenzoic acid (10 mg/m1 in chloroform/methanol, 1:1, vol/vol) as a matrix. The samples were analyzed in the reflection mode with an accelerating voltage operating in positive mode of 20 kV (13, 54).

Chemical synthesis

Reactions were carried out under a nitrogen atmosphere unless noted otherwise and monitored by thin-layer chromatography using Merck Silica Gel 60 F254 plates. Flash chromatography was performed using flash silica gel 60N (spherical neutral, particle size 40–50 μm, Kanto Chemical Co. Ltd). NMR spectra were recorded using a Bruker Avance III (500 MHz) device with a Prodigy (nitrogen-based) cryoprobe or a JNM-ECZL600R (600 MHz) device with a ROYAL HFX probe. Chemical shifts were reported in the scale relative to $CHCl_3$ (δ 7.26 ppm for ^{1}H NMR, 77.16 ppm for ^{13}C NMR) or pyridine (δ 7.58 ppm for ^{1}H NMR, 135.91 ppm for ^{13}C NMR) as an internal reference. Splitting patterns are designated as s: singlet, d: doublet, t: triplet, q: quartet, br: broadening, and m: multiplet. High-resolution mass spectrometry (HRMS) was obtained with a Bruker MicrOTOF II detector or a Bruker MALDI-TOF MS Autoflex Speed device. Gel permeation chromatography (GPC) was executed using LaboACE LC–5060 equipped with JAIGEL-1HR and JAIGEL-2HR ($CHCl_3$). HPLC purification was performed on the HITACHI HPLC system consisting of the following: pump, L6250; detector, L-3350 RI monitor; column, Senshu-Pak PEGASIL silica SP100; mobile phase, hexane/EtOAc.

TMM (C32, *RR*): $[\alpha]_D^{27}$ +110.01 (c = 0.28, $CHCl_3$/MeOH 4/1); ^{1}H NMR (500 MHz, $CDCl_3$/CD_3OD 4/1) δ 4.97 (d, *J* = 3.7 Hz, 1H), 4.93 (d, *J* = 3.7 Hz, 1H), 4.48 (dd, *J* = 11.9, 1.8 Hz, 1H), 4.05 (ddd, *J* = 10.0, 7.0, 1.8 Hz, 1H), 3.94 (dd, *J* = 11.9, 7.0 Hz, 1H), 3.75 (m, 1H), 3.72 (dd, *J* = 11.0, 1.8 Hz, 1H), 3.69 (dd, *J* = 9.8, 9.2 Hz, 1H), 3.67 (dd, *J* = 9.8, 9.2 Hz, 1H), 3.56–3.50 (m, 2H), 3.42 (dd, *J* = 9.8, 3.7 Hz, 1H), 3.37 (dd, *J* = 9.8, 3.7 Hz, 1H), 3.18 (dd, *J* = 9.6, 9.2 Hz, 1H), 3.14 (dd, *J* = 10.0, 9.2 Hz, 1H), 2.29 (ddd, *J* = 10.1, 7.8, 4.6 Hz, 1H), 1.50–1.39 (m, 2H), 1.38–1.28 (m, 2H), 1.28–1.05 (m, 50H), 0.75 (t, *J* = 7.0 Hz, 6H); ^{13}C NMR (126 MHz, $CDCl_3$/CD_3OD 4/1) δ 175.5, 94.43, 94.37, 72.7, 72.62, 72.58, 72.3, 71.6, 71.4, 71.0, 70.8, 70.0, 64.0, 61.9, 52.5, 34.7, 31.8 (2C), 29.64–29.48 (15C), 29.4, 29.34, 29.26 (2C), 29.2, 27.2, 25.2, 22.6 (2C), 13.9 (2C); HRMS-MALDI (*m/z*): $[M+Na]^+$ calcd for $C_{44}H_{84}NaO_{13}$, 843.5810; found 843.58.

TMM (C32, *SS*): $[\alpha]_D^{27}$ +8.03 (c = 0.22, $CHCl_3$/MeOH 4/1); ^{1}H NMR (500 MHz, $CDCl_3$/CD_3OD 4/1) δ 4.98 (d, *J* = 3.7 Hz, 1H), 4.95 (d, *J* = 3.7 Hz, 1H), 4.29 (dd, *J* = 11.7, 2.0 Hz, 1H), 4.19 (dd, *J* = 12.1, 5.0 Hz, 1H), 3.92–3.87 (m, 1H), 3.72–3.66 (m, 4H), 3.57–3.50 (m, 2H), 3.41 (dd, *J* = 9.8, 3.7 Hz, 1H), 3.37 (dd, *J* = 9.8, 3.7 Hz, 1H), 3.24 (dd, *J* = 9.3, 9.2 Hz, 1H), 3.19 (dd, *J* = 9.3, 9.2 Hz, 1H), 2.29 (ddd, *J* = 10.2, 7.8, 4.6 Hz, 1H), 1.49–1.39 (m, 2H), 1.38–1.28 (m, 2H), 1.28–1.03 (m, 50H), 0.74 (t, *J* = 7.0 Hz, 6H); ^{13}C NMR (126 MHz, $CDCl_3$/CD_3OD 4/1) δ 175.5, 94.0, 93.9, 72.8, 72.7, 72.5, 72.2, 71.6 (2C), 70.7, 70.3, 70.2, 63.1, 61.9, 52.9, 34.7, 31.8 (2C), 29.6–29.5 (15C), 29.4, 29.3, 29.2

(2C), 29.1, 27.2, 25.1, 22.6 (2C), 13.9 (2C); HRMS-MALDI (*m/z*) [M+Na]$^+$ calcd for $C_{44}H_{84}NaO_{13}$, 843.5810; found 843.58.

TMM (C32, *RS* +*SR*): $[\alpha]_D^{27}$ +9.02 (c = 0.23, $CHCl_3$/MeOH 4/1); ^{1}H NMR (500 MHz, $CDCl_3/CD_3OD$ 4/1) δ 4.97 (d, J = 3.7 Hz, 1H), 4.96 (d, J = 3.4 Hz, 1H), 4.94 (d, J = 3.7 Hz, 1H), 4.92 (d, J = 3.7 Hz, 1H), 4.51 (dd, J = 11.8, 1.8 Hz, 1H), 4.23–4.17 (m, 2H), 4.00–3.88 (m, 3H), 3.74–3.63 (m, 10H), 3.540 (dd, J = 12.7, 6.7 Hz, 1H), 3.538 (dd, J = 11.9, 6.1 Hz, 1H), 3.404 (dd, J = 9.8, 3.7 Hz, 1H), 3.399 (dd, J = 9.8, 3.7 Hz, 1H), 3.37 (dd, J = 9.6, 3.7 Hz, 1H), 3.35 (dd, J = 9.8, 3.7 Hz, 1H), 3.25–3.12 (m, 4H), 2.40–2.32 (m, 2H), 1.60–1.50 (m, 2H), 1.38–1.28 (m, 6H), 1.28–1.03 (m, 100H), 0.74 (t, J = 6.9 Hz, 12H); ^{13}C NMR (126 MHz, $CDCl_3/CD_3OD$ 4/1) δ 175.04, 174.98, 94.3, 94.1 (2C), 94.0, 73.0, 72.8 (2C), 72.7, 72.4, 72.3, 72.2 (2C), 71.6 (3C), 71.5, 70.9, 70.8, 70.7, 70.3, 70.0, 69.9, 63.8, 63.3, 61.9 (2C), 51.6, 51.3, 33.8, 33.5, 31.8 (4C), 29.65–29.48 (30C), 29.46 (2C), 29.4 (2C), 29.2 (4C), 27.9, 27.8, 26.6, 26.1 (2C), 25.8, 22.6 (4C), 13.9 (4C); HRMS-MALDI (*m/z*) [M+Na]$^+$ calcd for $C_{44}H_{84}NaO_{13}$, 843.5810; found 843.58.

Other synthetic precursors and TMM analogues were synthesized as described in the supplementary information.

The stereoselective synthesis of **TMM (C32, *RR*)** was carried out by a modified method of Nishizawa et al. (55): $[\alpha]_D$ +28.8 (c 0.49, $CHCl_3$/MeOH = 1:1); ^{1}H NMR (500 MHz, $CDCl_3/CD_3OD$ = 1:1) δ 5.10 (d, J = 4.0 Hz, 2H), 4.49 (dd, J = 12.0, 2.3 Hz, 1H), 4.19 (dd, J = 12.0, 5.7 Hz, 1H), 4.11–4.04 (ddd, J = 9.7, 5.7, 2.3 Hz, 1H), 3.85–3.76 (m, 4H), 3.72–3.66 (m, 2H), 3.55–3.46 (m, 2H), 3.38–3.30 (m, 2H), 2.48–2.40 (m, 1H), 2.21–1.94 (m, 2H), 1.68–1.15 (m, 52H), 0.89 (t, J = 6.9 Hz, 6H); ^{13}C{^{1}H} NMR (150 MHz,

$CDCl_3/CD_3OD$ = 1:1) δ 174.7, 93.4, 93.3, 72.6, 72.5, 71.8, 71.22, 71.17, 70.2, 70.1, 69.5, 62.8, 61.0, 52.2, 33.9, 31.2, 29.0, 28.94, 28.89, 28.86, 28.8, 28.7, 28.6, 28.2, 26.7, 24.7, 21.9, 21.6, 13.0; HRMS (ESI-TOF) *m/z*: $[M + Na]^+$ calcd for $C_{44}H_{84}NaO_{13}$, 843.5804; found, 843.5778.

Antibodies

Human Fc block (Fc1) was purchased from BD Pharmingen. Anti-human CD3 (HIT3a), anti-human CD19 (SJ25C1), TotalSeq-C Hashtags (LNH-94; 2M2), anti-mouse CD3 (2C11, 17A2), anti-mouse CD69 (H1.2F3), and anti-rat CD2 (OX-34), anti-human CD1a (HI149), anti-human CD1b (SN13), anti-human CD1c (L161), anti-human CD1d (51.1), mouse IgG1 κ isotype control (MG1-45) and mouse IgG2b κ isotype control (MPC-11) antibodies were purchased from BioLegend.

In vitro stimulation of PBMCs

Cryopreserved human PBMCs were thawed and labeled by CellTrace Violet (CTV, Thermo Fisher Scientific), then quenched and washed by RPMI 1640 medium (Sigma) supplemented with 5% human AB serum (Gemini Bio), penicillin (Sigma), streptomycin (MP Biomedicals), and 2-mercaptoethanol (Nacalai Tesque). 10^6 CTV labeled PBMCs were stimulated in the same medium with plate-coated 3 μg of Mtb-crude lipids, 3 μg of synthetic GMM or 3 μg of heat-killed Mtb H37Rv for 10 days. Recombinant human IL-2 (1 ng/m1, PeproTech), human IL-7 (5 ng/m1, PeproTech) and human IL-15 (5 ng/m1, PeproTech) were added at day 2, day 5 and day 8. After staining with anti-human CD3 antibody, CTV^{low}CD3^{+} cells were sorted by SH800 Cell Sorter (Sony Biotechnology) and used for single-cell TCR- and RNA-seq analyses.

Single-cell–based transcriptome and TCR repertoire analysis

Single-cell transcriptome and TCR repertoire analysis were performed using Chromium Controller (10x Genomics) according to the manufacturer's instructions as previously described (56). Libraries were sequenced on Illumina NovaSeq 6000 in the paired-end mode. The raw reads were processed by Cell Ranger v6.0.0-7.1.0 (10x Genomics). TCR repertoire analysis was conducted using Scirpy (v0.11.1) and gene expression-based clustering was performed using Scanpy (1.9.1). UMAP plot, heatmap, volcano plot and differential expression analysis were performed by Seurat R package (v5.0.1).

Bulk TCR-sequencing

3×10^5 PBMCs were lysed in QIAzol (QIAGEN). Full-length cDNA was then synthesized using SMARTer technology (Takara Bio), and the variable regions of TCR α and TCRβ genes were amplified using TRAC/TRBC-specific primers. After sequencing of the variable region amplicons, each pair of reads was assigned a clonotype (defined as TR(A/B)V and TR(A/B)J genes and CDR3) using MiXCR software (57).

APCs

For the preparation of cytokine-differentiated human monocytes, CD14^{+} monocytes were sorted from freshly isolated human PBMCs using MACS cell separation column

(Miltenyi Biotec), followed by cultured in RPMI 1640 supplemented with 10% FBS, non-essential amino acids, 10 ng/m1 human GM-CSF, and 10 ng/m1 human IL-4 for 7 days. Human CD1b was cloned into the retroviral vector pMX-IRES-human CD8 (58) using Phoenix packaging cells and PEI MAX (Polysciences). Supernatant containing retroviruses was used for infection into the mouse dendritic cell line DC2.4 (ATCC).

TCR reconstitution and stimulation

TCR α and β chain cDNA sequences were synthesized with eblock (IDT) and cloned into retroviral vectors pMX-IRES-rat CD2. TCRα mutants were constructed by site-directed mutagenesis. The vectors were transduced into mouse T cell hybridoma with an NFAT-GFP reporter gene (56, 59) using retroviruses described above to reconstitute TCRαβ pairs. For antigen stimulation, TCR-reconstituted cells were cocultured with stimulants in the presence of APCs unless indicated otherwise. After 20 h, T cell activation was assessed by GFP and CD69 expression.

NGS-based mutagenesis scanning

For the mutant libraries, synthesized mutant TCRα or β cDNA sequences were pooled, and reconstituted into reporter cell lines with WT TCRβ or TCRα, respectively. Library cells were left unstimulated or stimulated with TMM for 20 h and then sorted by GFP negative/positive populations. Each sorted population was analyzed by bulk TCR-sequencing (GSE261269). The proportion of the read counts of each mutant within GFP+ or GFP^{-} population were shown as percentage.

CD1b tetramers

Unloaded human CD1b monomers (biotinylated) were obtained from the NIH tetramer facility. For TMM loading, 16 μg of Mtb TMM was sonicated at 45°C for 1 h in 45 μl of 0.5% CHAPS 50 mM sodium citrate buffer (pH 4.5), added to 5 μl of CD1b monomers (2 mg/m1) and incubated overnight at 37°C. For the preparation of control CD1b tetramers (endo-CD1b tet), CD1b monomers were treated as described above without TMM loading. Monomers were then neutralized by 5 μl 1 M Tris (pH 8) and tetramerized using streptavidin-PE (Biolegend) or streptavidin-APC (eBioscience).

Tetramer staining and isolation of CD1b tetramer$^+$ T cells

10^7 human PBMCs were incubated with 20 μg/m1 of PE-conjugated TMM-CD1b tetramers and 20 μg/m1 of APC-conjugated endo-CD1b tetramers in 40 μl 1% BSA/PBS at room temperature in the dark for 15 min. Without washing, 2 μl of human Fc block (50 μg/ml), 60 μl of anti-human CD3-FITC (2 μg/ml) and 2 μl of TotalSeq-C Hashtags were added and incubated on ice for 20 min. Before sorting, cells were washed and filtered by nylon mesh, and incubated with propidium iodide. 2492 TMM-tetramer$^+$ endo-tetramer$^-$ cells within the CD3$^+$ gated population were sorted by SH800 Cell Sorter (Sony Biotechnology) and subjected to scTCR-RNA-seq. Unsorted PBMCs from the same donors were also subjected to scTCR-RNA-seq. Among 1737 TMM-tetramer$^+$ cells (1559 clonotypes) obtained, clonotypes that were detected more abundantly in unsorted T cells than in TMM-tetramer$^+$ T cells were excluded.

Characterization of Y-50 before and after TMM stimulation

For unstimulated cells, PBMCs from 3 donors were stained with anti-CD3Ab and PE-conjugated TMM-CD1b tetramer and $CD3^+$TMM-tetramer$^+$ cells were sorted. For TMM-simulated cells, CTV-labeled PBMCs from 5 donors were stimulated with TMM for 8-10 days and $CD3^+CTV^{lo}$ cells were sorted. These samples were subjected to scTCR-RNA-seq analysis and projected on the same UMAP plot. On the basis of scTCR-seq, the cells expressing clonotype identical to Y-50 TCRαβ were indicated as Y-50 clonotype.

Crystal structural analysis

Complementary DNA encoding the ectodomains of Y-50 TCRα (from Ala-1 to S-205) and β (from Asp-1 to Asp-249) with a nidogen signal sequence, a 6 × His-tag and a tobacco etch virus protease cleavage site at the N-terminal were synthesized (Thermo GeneArt) and cloned into pcDNA3.1(+) vector. To improve the efficiency of protein expression, artificial disulfide bond and stabilizing mutations were introduced as described previously (60, 61). The plasmids were transformed into Expi293 cells in the presence of the mannosidase inhibitor kifunensine. The cells were cultured with shaking at 120 rpm 37°C 8% CO_2 for 4 days. After being filtered through a 0.22 μm, the supernatant was applied onto 5 ml nickel-nitrilotriacetic acid agarose (FUJIFILM Wako), and His-tagged TCRαβ were eluted with elution buffer (50 mM Tris-HCl [pH 8.0], 300 mM NaCl, and 250 mM imidazole). After removal of His-tag by tobacco etch virus protease, the eluted protein was concentrated and further applied to Superdex 75 (Cytiva) equilibrated with 20 mM Tris-HCl (pH 8.0) buffer containing 100 mM NaCl. The crystals were formed by the sitting-drop vapor-diffusion method. 0.4 μl of protein solution (5 mg/m1 in 100 mM NaCl, 20 mM Tis-HCl (pH8.0)) was mixed with 0.4 μl of mother liquid containing 0.2 M potassium sulfate 0.1M Bis-Tris pH 5.5, 25% PEG3350 and incubated at 20°C. The diffraction data were collected in a cold nitrogen gas stream on an EIGER X 9M detector (DECTRIS) at a wavelength of 1.0 Å. The resulting datasets were processed, integrated by XDS (62), and scaled by AIMLESS (63). Structures were clarified by molecular replacement with the TCR complex (PDB: 8ZO4 as a search model, by MOLREP) as implemented in CCP4i software (63). The models were refined using REFMAC5 and PHENIX1.20 software (64, 65). The structures were rebuilt using COOT 0.9.8.92 (66) and further modified based on σ-weighted $(2|F_{\mathrm{obs}}|—|F_{\mathrm{calc}}|)$ and $(|F_{\mathrm{obs}}|—|F_{\mathrm{calc}}|)$ electron density maps. Crystallographic images were created using PyMOL software (Schrödinger, LLC.). Data collection and refinement statistics were summarized in Supplementary Table 2.

Cryo-EM structural analysis

TMM loaded CD1b ectodomain was refolded as follows: Denatured proteins of CD1b (24.8 mg) and β2m (9.6 mg) were mixed with 3.28 mg TMM and refolded in the buffer containing 0.1 M Tris-HCl (pH 8.0), 1 M L-arginine (pH 8.0), 5 M urea, 5 mM reduced glutathione, and 0.5 mM oxidized glutathione. The refolded proteins were then dialyzed four times against 0.01 M Tris-HCl (pH 8.0) and applied onto HiTrap Q HP 5 m1 column (Cytiva). Purified TMM-loaded CD1b was mixed with Y-50 TCR at the ratio of

1:1. 2.2 μl of sample (1.0 mg/ml) was applied onto the glow-discharged Quantifoil Au 0.6/1.0 200 mesh grid (Quantifoil Micro Tools GmbH, Germany) and frozen in liquid ethane using a Vitrobot IV (FEI, 4°C and 95% humidity). Cryo-EM data collection was performed on a Titan Krios cryo-TEM equipped with a Cs corrector (Thermo Fisher Scientific, USA) operating at 300 keV in EFTEM nanoprobe mode. Images were acquired as movies using Gatan BioQuantum energy filter (slit width of 20 eV) and K3 direct electron detector camera (Gatan, Inc., USA) in electron counting mode. A total of 8,533 movies were collected at a dose rate of 8.532 e^{-}/pixel/s, a pixel size of 0.675 $Å^2$, and a total dose of 60 $e^{-}/Å^2$. SerialEM software (67) was used for automated data collection using a 3 × 3-hole pattern beam-image shift scheme with a nominal defocus range of −0.6 to −1.8 μm. All of image processes were carried out on cryoSPARC v4.4.1 software (68). After motion correction of movies and CTF parameter estimation, an initial round of particle picking was performed using the blob picker tool (diameter 100 - 140 Å). After four iterations of 2D class and manual selection, 12,424 particles were selected. Classification into three classes using Ab Initio Reconstruction and manual selection was repeated twice, and the resulting 6,961 particles were used as training data for Topaz picks. The 2,266,363 particles were automatically picked using Topaz picking algorithm. After two rounds of 2D classification and 3D classification using Ab Initio Reconstruction and Heterogeneous refinement, 599,402 particles were selected. A subsequent round of 2D classification further narrowed the selection to 232,909 particles. Three maps were reconstructed using Ab-initio Reconstruction with C1 symmetry. Several maps were duplicated and used as the initial model for the heterogeneous refinement. In this process 2,266,363 particles picked by TOPAZ were used, but the resolution was not high enough, so 3,158,103 particles picked by blob picker. As a result, one of these classified particles (599,402 particles) were applied to 2D classification, Ab-initio Reconstruction and Non-uniform Refinement. Finally, 11,404 particles were selected and the density map from the refinement was obtained at 3.31 Å resolution. Each particle was subjected to Reference Based Motion Correction. As the results of Non-uniform Refinement, a map of the complex at 3.18 Å resolution was obtained. Local resolution of the obtained map was estimated by Local resolution estimation job on cryoSPARC. 3D structures of Y-50 (PDB: 8XUB) and CD1b (PDB: 5L2K) were automatically fitted into map with program phenix.dock_in_map in PHENIX program suite (65). Chemical structure of TMM was idealized by phenix.elbow. The atomic model of ternary complex was manually modified using COOT and refined with phenix.real_space_refine of PHENIX suite. Stereochemistry of the refined structure was evaluated with MolProbity (69). Validation of the final model is summarized in Table S3.

Statistics

Data were analyzed using GraphPad Prism v9.1.0 software (GraphPad Software). Statistical differences between two groups were determined by unpaired two-tailed Welch's *t*-test. A *P* value < 0.05 was considered statistically significant.

Study approval

The protocol for collecting human blood samples from healthy donors was approved by Osaka University Institutional Review Board (IRB) (approval number 898-4). Informed consent was obtained from all participants before the first blood sampling. The protocol for collecting human blood samples from active tuberculosis patients was reviewed and approved by the medical research ethics committee of the National Institute of Infectious Diseases for inclusion of human subjects (#1343 and #1491) and of Fukujuji Hospital (#22034). All participants were enrolled after giving written informed consent.

Data availability

All reagents used in this study will be made available upon reasonable request to the corresponding author. All single-cell TCR-RNA sequencing and Bulk TCR-sequencing data were deposited in NCBI's Gene Expression Omnibus (GEO) database (GSE260931 and GSE261269). Values for all data points in graphs are in the Supporting Data Values file.

Author contributions

YS, MA, MH, WK, XL, and EI did investigation; NF, HT, TI, SO, KK, TY, YH, IVR, SI, and GH provided resources; MH, KT, MN, DM, and TK did data curation; SY supervised research; YS, MA, DBM, and SY wrote the manuscript.

Acknowledgements

We thank Y-C. Liu, D. Okuzaki, Y. Nogi, M. Okuda, A. Yamashita, N. Kamata and Y. Takeuchi for experimental support; C. Schutt, M. Ato, Y. Tsujimura, A. Shahine, J. Rossjohn and I. Yano for discussion. This research was supported by AMED (21gm0910010, 223fa627002, 21ak0101070 (SY), 23fk0108608 and 24fk0108701 (YH)), JSPS KAKENHI (JP20H00505, JP22H05183 (SY), JP20K06938, JP20H04773 (SI), JP20K05899, JP23K06554 (NF), JP24KJ0147 and JP24K18464 (MA)) and JPMJSP2138 (YS). This research was also supported by BINDS from AMED (JP23ama121001, support number 4199). We thank the NIH Tetramer Core Facility (contract number 75N93020D00005) for providing biotinylated CD1b monomers.

Declaration of interests

The authors declare no competing financial interests.

Figure legends

Figure 1. Identification of mycobacterial lipid-reactive T cells.

A) Schematic representation of experimental procedure. Human PBMCs were cultured with plate-coated crude lipids extracted from *M. tb*. The expanded T cells were sorted and analyzed by single-cell TCR-RNA sequencing. Highly expanded CTV^{lo} TCR clonotypes were reconstituted into NFAT-GFP reporter cells to examine the reactivity to *M. tb* lipids.

B) Forty-four TCR clonotypes were reconstituted into reporter cells and analyzed for their surface expression using anti-CD3 antibody.

C) NFAT-GFP reporter cells (44 clonotypes) expressing each different TCR were stimulated with *M. tb* crude lipids in the presence of PMBCs or cytokine-differentiated monocytes as APCs, and after 20 h incubation, analyzed for GFP and CD69 expression. Representative results from two independent experiments are shown.

D) UMAP plot of T cells expanded in response to *M. tb* lipids (left panel). T cell clones expressing Y-50 clonotype are highlighted in red dots (right panel). CTL, cytotoxic T lymphocytes.

E) Gene expression signature of Y-50 cells. The expression of characteristic genes in each cell expressing Y-50 TCR clonotype are shown.

Figure 2. Identification of TMM as a T cell antigen.

A-B) *M. tuberculosis* H37Rv crude lipids were fractionated by HPTLC using chloroform/methanol/water (65:25:4; vol/vol/vol (A) and 90:10:1; vol/vol/vol (B)) and stained with copper(II) acetate-phosphoric acid. Y-50 reporter cells were stimulated with each fraction in the presence of APCs and analyzed for GFP and CD69 expression. Open and closed arrowheads denote the origin and the solvent front, respectively.

C) MALDI-TOF MS spectrum of lipid fraction 2 (Fr2).

D) Chemical structure of TMM of α-mycolate is shown, and methoxy-mycolate, and keto-mycolate are the other major subclasses of mycolate found in *M. tuberculosis* TMM.

E) Y-50 reporter cells were co-cultured with cytokine-differentiated human monocytes pre-incubated with whole bacteria (heat killed *M. tb* H37Rv or living BCG) and analyzed for GFP and CD69 expression.

F) Y-50 reporter cells were stimulated with the indicated concentration of TMM, TDM or GMM. GFP and CD69 expression are shown in bar graphs. Schematic ligand structures are shown below.

Data are shown as the means ± SD of triplicate assays (E, F) and representative results from two independent experiments are shown (A, B, E and F).

Figure 3. CD1b restricts TMM recognition by Y-50 T cells.

A) Y-50 reporter cells were co-cultured with cytokine-differentiated human monocytes and TMM (0.3 nmol /well) in the presence of 5 μg/m1 of anti-CD1a, CD1b, CD1c, CD1d or isotype control antibodies (IgG1 and IgG2b) and analyzed for GFP and CD69 expression.

B) The reporter cells expressing Y-50 TCR were stimulated with TMM (1 nmol/well) in the presence of HEK293T cells transfected with human CD1a, CD1b, CD1c or CD1d.

C) Y-50 reporter cells were stimulated with TMM (1 nmol/well) purified from *M. tuberculosis* CDC1551, *M. bovis* BCG, *M. intracellulare, M. smegmatis*, *Rhodococcus equi* and *R. sp* 4306. Also, GMM, mannose-monomycolate (MMM), glycerol monomycolate (GroMM) and mycolic acid (MA) were tested in the presence of human CD1b-expressing DC2.4 cells (CD1b-DC2.4).

D) Synthetic scheme for TMM.

E) Y-50 reporter cells were stimulated with synthetic TMM harboring a β-hydroxy group (hydroxy) and α-branched alkyl chains (branch) or synthetic analogues lacking hydroxy (−, +) or both moieties (−, −) in the presence of CD1b-DC2.4 cells as APCs.

Data are shown as the means ± SD of triplicate assays and representative results from two independent experiments are shown (A, B, C and E).

Figure 4. Mutagenesis and structural analysis of TMM-reactive TCR.

A) The amino acid sequences of Y-50 CDR3α within the arginine mutants shown in red. TMM reactivities of each mutant were shown as percentage of the maximum response induced by plate coated anti-CD3 Ab. The number of amino acids were shown in accordance with the ImMunoGeneTics (IMGT) definition (https://imgt.org/IMGTScientificChart/).

B) Nucleotide and amino acid sequences of the Y-50 TCR CDR3β region and its junction-deletion mutants (Δ). D region and N or P nucleotide sequences that constitute junctional sequences were unshaded. Cells were stimulated as indicated in (A).

C) Crystal structure of the Y-50 TCRαβ heterodimer (PDB: 8XUB). The main chains of TCRα and β are shown in violet and brown, respectively. CDR3αβ regions are boxed. D-E) 2*Fo-Fc* map contoured at 2.0 σ (D) and *B*-factor diagram (E) of CDR3αβ are shown as gray mesh and color gradient, respectively. Junction regions of CDR3β are boxed. Data are shown as the means ± SD of triplicates (A and B) and representative results from two independent experiments are shown.

Figure 5. Ternary complex structure of Y-50 TCR-TMM-CD1b.

A) Overall structure of the Y-50 TCR-TMM-CD1b complex. The main chains of TCRα, TCRβ, and CD1b are shown as indicated. TMM was presented as yellow sphere.

B) Superimposition of the structure of Y-50 TCR alone (PDB: 8XUB)(pink) and Y-50 TCR bound to TMM-CD1b (PDB: 8ZOX)(blue)(upper panel). CDR3β regions (boxed) are magnified (lower panels).

C) Close-up view of TMM (*R,R*) and the side chain of R114 within CDR3α. β-hydroxy group of TMM is shown in red.

D) Y-50 reporter cells were stimulated with natural configuration of synthetic TMM (*R,R*) or non-natural stereoisomers, (*S,S*) or (*S,R*+*R,S*), in the presence of CD1b-DC2.4 and analyzed for GFP and CD69 expression. Structures of stereoisomers are shown below (*R*, red; *S*, black). Data are shown as the means ± SD of triplicates and a representative result from two independent experiments is shown.

E) Close-up view of TMM (*R,R*) and the side chain of R107 (CDR3α) and D114 (CDR3β). Hydroxy groups of TMM that formed hydrogen bonds to the TCR side chains are shown in red.

F) Close-up view of the side chains of R79, E80 and D83 in CD1b that interact with the side chains of R37 (CDR1β), Y58 and E63 (CDR2β) and G110 (CDR3β).

G) Multi-bonded interaction of CD1b RExxD motif and TRBV4-1 residues. Individual interaction was shown as dotted lines.

H) Conservation of RExxD motif in human CD1b and CD1c. Amino acid sequence of CD1a (NP_001307581), CD1b (NP_001755.1), CD1c (NP_001756.2) and CD1d (NP_001306074) are aligned. Numbers indicate the amino acid position of mature peptide (excluding signal peptide).

Figure 6. Functional maturation of TMM-reactive T cells upon TMM stimulation.

A-B) Cluster shift of Y-50 clonotype before and after TMM stimulation. T cells expressing Y-50 clonotype defined by scTCR-RNA-seq were overlayed (A) on UMAP plot of PBMCs from donors including the donor used in Figure 1 (B) as described in Material and Methods. T_{EM}, effector memory T cells. CTL, cytotoxic T lymphocytes.

C-E) Differentially expressed genes in Y-50 T cells upon TMM stimulation. The expression of representative genes encoding cytotoxic effector molecules (C), pro-inflammatory cytokines and chemokines (D) and stemness-related molecules (E) are shown in violin plots.

Figure 7. TMM-specific T cells with similar characteristics are shared among individuals.

A-B) Frequent TMM-specific clonotypes identified by TMM-CD1b-tetramer sorting and scTCR-RNA-seq were overlayed (A) on UMAP plot of all TMM-tetramer-sorted T cells and unsorted CD3$^+$ T cells from 13 healthy donors (B). Three clones were detected from different individual donors; two clones (clone 17 and 439) were from another donor. CD4$^+$ T$_{EM}$, CD4$^+$ effector memory T cells. CD4 T$_{CM}$, CD4$^+$ central memory T cells. Naïve T cells were rare within TMM-tetramer$^+$ cells and were not clustered on UMAP.

C) TCR usages, CDR3 sequences and length of CDR3β region of clonotypes detected in (A).

D) Each clonotype was reconstituted into reporter cells and analyzed for TMM, TDM and GMM reactivity using CD1b-DC2.4 as APCs. Data are shown as the means ± SD of triplicate assays and representative results from two independent experiments are shown. Reporter cells were stained with PE-conjugated endo-CD1b (Cont.) or TMM-CD1b (TMM) tetramers and anti-CD3 antibodies. Y-50 TCR is shown as a control.

E-F) Frequency of TCRVβ usage (D) and length of CDR3β region (E) of unsorted or top 27 TMM-CD1b tetramer$^+$ T cell clonotypes.

G) PBMCs from Japanese donors (uninfected donors, n = 7; TB patients, n = 13) or PBMC (n = 10) and cord blood cells (n = 10) from uninfected US donors were stained with PE-conjugated TMM-loaded CD1b tetramer, APC-conjugated CD1b-endo tetramer and anti-CD3 antibody. The percentages of TMM-CD1b tetramer positive and endo-CD1b tetramer negative population in CD3$^+$ T cells are shown (TMM-CD1b-tet$^+$).

Medians are indicated with horizontal bars. *P* values were calculated using unpaired two-tailed Welch's *t*-test; *$P < 0.05$; N.S., not significant.

Supplemental Figure 1. Characteristic genes in each cluster of UMAP plot in Figure 1D.

Log-normalized expression of the top 10 genes for each T cell cluster in Figure 1D are shown in heatmap. The colors of annotated clusters are consistent with the dots shown in Figure 1D.

Supplemental Figure 2. Identification of antigen recognized by Y-50.

A-B) MALDI-TOF MS spectrum of lipid fraction 2 (Fr2) and the reference spectrum of *M. tuberculosis* H37Rv TMM (lower) (A). The subclass of mycolate and chain length annotated based on the detected *m/z* matching the structural formula were indicated on upper panel of (A) (B). Related to Figure 2C.

C) Structural characteristics of TMM composed of three mycolic acid subclasses.

Supplemental Figure 3. Structural analysis of Y-50 TCR-TMM-CD1b ternary complex.

A) Close-up view of the side chains of CDR3α arginine residues (R107, R108, R113 and R114) in the crystal structure of Y-50 (PDB 8XUB).

B) Refolding of recombinant CD1b (CD1b) with synthetic TMM (TMM). Red arrow indicates refolded CD1b-TMM-β2m complex separated by anion-exchange chromatography (left), which was confirmed by SDS-PAGE followed by CBB staining (right). NR, non-reducing; R, reducing.

C-D) Representative micrograph (C) and 2D classification (D) of cryo-EM analysis.

E) Flow chart of cryo-EM image processing. Gold standard FSC curves are also shown.

F) Cryo-EM density map (gray mesh) and a model structure TMM (yellow).

G) TMM (yellow sphere) and surface representation of the CDR3 (violet), CDR3β (brown) loops and CD1b (gray) are shown. Oxygen atoms of TMM are shown in red.

Supplemental Figure 4. Synthesis of TMM.

A) NMR chart of synthetic TMMs ((*R,R*), (*S,S*) and (*R,S* + *S,R*)).

B) Antigenic activity of synthetic TMM. Y-50 reporter cells were stimulated with stereoselective synthesized (left) or chiral column-separated synthetic (right) TMM (*R,R*) in the presence of CD1b-DC2.4 and analyzed for GFP and CD69 expression. Data are shown as the means ± SD of triplicate assays and representative results from two independent experiments.

Supplemental Figure 5. Mutagenesis analysis of Y-50 TCR.

A) Reporter cells expressing indicated combination of Y-50 TCRαβ mutants were stimulated with TMM on CD1b-DC2.4 and percentages of GFP$^+$CD69$^+$ cells are shown. Data are shown as the means ± SD of triplicates and a representative result from two independent experiments is shown.

B) Schematic procedure of NGS-based mutagenesis scanning.

C-D) Distribution of reporter cells expressing each mutant in GFP$^+$ and GFP$^-$ populations with or without TMM stimulation. The ratio of each CDR3α mutant (C) and CDR3β mutant (D) are shown as percentages. The number of amino acids were shown in accordance with the ImMunoGeneTics (IMGT) definition.

Supplemental Figure 6. Validation of TMM-loaded CD1b tetramers.

A) Reporter cells expressing Y-50 and mutant Y-50 TCR shown in Figure 3B (Δ12) were stained with PE-conjugated TMM-loaded CD1b tetramer and anti-CD3 antibody. As a control tetramer, CD1b biotinylated monomer conjugated by SA-APC (endo-CD1b tet) was prepared. Endo-CD1b tet reactivity was confirmed by Y-50 (negative control) and autoreactive Fee TCR (70) (positive control).

B) HEK 293T cells were transfected with plasmids encoding Fee TCR and human CD3s-IRES-GFP and stained with PE-conjugated TMM-CD1b or APC-conjugated endo-CD1b tetramers.

C) TMM-stimulated PBMCs were co-stained with all combination of four tetramers (APC-labeled CD1-TMM, PE-labeled CD1-endo tetramers, PE-labeled CD1-TMM and APC-labeled CD1-endo tetramers). Numbers in red indicate the frequency of TMM-specific T cells with CD3$^+$ cells for each staining combination.

Supplemental Figure 7. Gene expression profile of TMM-specific T cells with or without TMM stimulation.

A) Volcano plot showing differentially expressed genes of Y-50 T cells in response to TMM stimulation.

B) Heatmap showing log-normalized expression of marker genes in individual Y-50 cells in the presence or absence of TMM stimulation.

C) Heatmap showing log-normalized expression of marker genes for each CD3$^+$ T clusters shown in Figure 7A-B. Typical signature of CD4$^+$ effector memory T cells is boxed.

D) Representative plots of the tetramer staining of PBMCs from uninfected (left) and TB-infected (right) donors. PBMCs were stained with PE-conjugated TMM-loaded CD1b tetramer, APC-conjugated CD1b-endo tetramer and anti-CD3 antibody. Numbers within boxes indicate the percentages of TMM-CD1b tetramer positive and endo-CD1b tetramer negative population in CD3$^+$ T cells are shown.

References

1. Dulberger CL, Rubin EJ, and Boutte CC. The mycobacterial cell envelope - a moving target. *Nat Rev Microbiol.* 2020;18(1):47-59.
2. Freund J, Casals J, and Hosmer EP. Sensitization and antibody formation after injection of tubercle bacilli and paraffin oil. *Proceedings of the Society for Experimental Biology and Medicine.* 1937;37(3):509-13.
3. Dube JY, McIntosh F, Zarruk JG, David S, Nigou J, and Behr MA. Synthetic mycobacterial molecular patterns partially complete Freund's adjuvant. *Sci Rep.* 2020;10(1):5874.
4. Ishikawa E, Mori D, and Yamasaki S. Recognition of Mycobacterial Lipids by Immune Receptors. *Trends Immunol.* 2017;38(1):66-76.
5. Brown GD, Willment JA, and Whitehead L. C-type lectins in immunity and homeostasis. *Nat Rev Immunol.* 2018;18(6):374-89.
6. Coulombe F, Divangahi M, Veyrier F, de Leseleuc L, Gleason JL, Yang Y, et al. Increased NOD2-mediated recognition of N-glycolyl muramyl dipeptide. *J Exp Med.* 2009;206(8):1709-16.
7. Stenger S, Hanson DA, Teitelbaum R, Dewan P, Niazi KR, Froelich CJ, et al. An antimicrobial activity of cytolytic T cells mediated by granulysin. *Science.* 1998;282(5386):121-5.
8. Walch M, Dotiwala F, Mulik S, Thiery J, Kirchhausen T, Clayberger C, et al. Cytotoxic cells kill intracellular bacteria through granulysin-mediated delivery of granzymes. *Cell.* 2014;157(6):1309-23.
9. Flynn JL, Chan J, Triebold KJ, Dalton DK, Stewart TA, and Bloom BR. An essential role for interferon gamma in resistance to *Mycobacterium tuberculosis* infection. *J Exp Med.* 1993;178(6):2249-54.
10. Flynn JL, Goldstein MM, Chan J, Triebold KJ, Pfeffer K, Lowenstein CJ, et al. Tumor necrosis factor-alpha is required in the protective immune response against *Mycobacterium tuberculosis* in mice. *Immunity.* 1995;2(6):561-72.
11. Keane J, Gershon S, Wise RP, Mirabile-Levens E, Kasznica J, Schwieterman WD, et al. Tuberculosis associated with infliximab, a tumor necrosis factor alpha-neutralizing agent. *N Engl J Med.* 2001;345(15):1098-104.
12. Ishikawa E, Ishikawa T, Morita YS, Toyonaga K, Yamada H, Takeuchi O, et al.

Direct recognition of the mycobacterial glycolipid, trehalose dimycolate, by C-type lectin Mincle. *J Exp Med.* 2009;206(13):2879-88.

13. Harris SP, Fujiwara N, Mealey RH, Alperin DC, Naka T, Goda R, et al. Identification of *Rhodococcus equi* lipids recognized by host cytotoxic T lymphocytes. *Microbiology (Reading).* 2010;156(Pt 6):1836-47.
14. Decout A, Silva-Gomes S, Drocourt D, Barbe S, Andre I, Cueto FJ, et al. Rational design of adjuvants targeting the C-type lectin Mincle. *Proc Natl Acad Sci U S A.* 2017;114(10):2675-80.
15. Moody DB, Reinhold BB, Guy MR, Beckman EM, Frederique DE, Furlong ST, et al. Structural requirements for glycolipid antigen recognition by CD1b-restricted T cells. *Science.* 1997;278(5336):283-6.
16. Huang S, Shahine A, Cheng TY, Chen YL, Ng SW, Balaji GR, et al. CD1 lipidomes reveal lipid-binding motifs and size-based antigen-display mechanisms. *Cell.* 2023;186(21):4583-96 e13.
17. Fujita Y, Naka T, Doi T, and Yano I. Direct molecular mass determination of trehalose monomycolate from 11 species of mycobacteria by MALDI-TOF mass spectrometry. *Microbiology (Reading).* 2005;151(Pt 5):1443-52.
18. Beckman EM, Porcelli SA, Morita CT, Behar SM, Furlong ST, and Brenner MB. Recognition of a lipid antigen by CD1-restricted alpha beta+ T cells. *Nature.* 1994;372(6507):691-4.
19. Layre E, Collmann A, Bastian M, Mariotti S, Czaplicki J, Prandi J, et al. Mycolic acids constitute a scaffold for mycobacterial lipid antigens stimulating CD1-restricted T cells. *Chem Biol.* 2009;16(1):82-92.
20. Shugay M, Bagaev DV, Zvyagin IV, Vroomans RM, Crawford JC, Dolton G, et al. VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Res.* 2018;46(D1):D419-D27.
21. Gras S, Van Rhijn I, Shahine A, Cheng TY, Bhati M, Tan LL, et al. T cell receptor recognition of CD1b presenting a mycobacterial glycolipid. *Nat Commun.* 2016;7:13257.
22. Van Rhijn I, Gherardin NA, Kasmar A, de Jager W, Pellicci DG, Kostenko L, et al. TCR bias and affinity define two compartments of the CD1b-glycolipid-specific T Cell repertoire. *J Immunol.* 2014;192(9):4054-60.
23. Reinink P, Shahine A, Gras S, Cheng TY, Farquhar R, Lopez K, et al. A TCR beta-

Chain Motif Biases toward Recognition of Human CD1 Proteins. *J Immunol.* 2019;203(12):3395-406.

24. Guo T, Koo MY, Kagoya Y, Anczurowski M, Wang CH, Saso K, et al. A Subset of Human Autoreactive CD1c-Restricted T Cells Preferentially Expresses TRBV4-1(+) TCRs. *J Immunol.* 2018;200(2):500-11.

25. Bendelac A, Savage PB, and Teyton L. The biology of NKT cells. *Annu Rev Immunol.* 2007;25:297-336.

26. Nunes-Alves C, Booty MG, Carpenter SM, Jayaraman P, Rothchild AC, and Behar SM. In search of a new paradigm for protective immunity to TB. *Nat Rev Microbiol.* 2014;12(4):289-99.

27. Vesosky B, Rottinghaus EK, Stromberg P, Turner J, and Beamer G. CCL5 participates in early protection against *Mycobacterium tuberculosis*. *J Leukoc Biol.* 2010;87(6):1153-65.

28. Van Rhijn I, and Moody DB. Donor Unrestricted T Cells: A Shared Human T Cell Response. *J Immunol.* 2015;195(5):1927-32.

29. Godfrey DI, Uldrich AP, McCluskey J, Rossjohn J, and Moody DB. The burgeoning family of unconventional T cells. *Nat Immunol.* 2015;16(11):1114-23.

30. Engel I, and Kronenberg M. Making memory at birth: understanding the differentiation of natural killer T cells. *Curr Opin Immunol.* 2012;24(2):184-90.

31. Bagchi S, He Y, Zhang H, Cao L, Van Rhijn I, Moody DB, et al. CD1b-autoreactive T cells contribute to hyperlipidemia-induced skin inflammation in mice. *J Clin Invest.* 2017;127(6):2339-52.

32. Sieling PA, Chatterjee D, Porcelli SA, Prigozy TI, Mazzaccaro RJ, Soriano T, et al. CD1-restricted T cell recognition of microbial lipoglycan antigens. *Science.* 1995;269(5221):227-30.

33. Moody DB, Ulrichs T, Muhlecker W, Young DC, Gurcha SS, Grant E, et al. CD1c-mediated T-cell recognition of isoprenoid glycolipids in *Mycobacterium tuberculosis* infection. *Nature.* 2000;404(6780):884-8.

34. Gilleron M, Stenger S, Mazorra Z, Wittke F, Mariotti S, Bohmer G, et al. Diacylated sulfoglycolipids are novel mycobacterial antigens stimulating CD1-restricted T cells during infection with *Mycobacterium tuberculosis*. *J Exp Med.* 2004;199(5):649-59.

35. Winchell CG, Nyquist SK, Chao MC, Maiello P, Myers AJ, Hopkins F, et al.

CD8+ lymphocytes are critical for early control of tuberculosis in macaques. *J Exp Med.* 2023;220(12).

36. Geldmacher C, Ngwenyama N, Schuetz A, Petrovas C, Reither K, Heeregrave EJ, et al. Preferential infection and depletion of *Mycobacterium tuberculosis*-specific CD4 T cells after HIV-1 infection. *J Exp Med.* 2010;207(13):2869-81.

37. Su CC, Klenotic PA, Bolla JR, Purdy GE, Robinson CV, and Yu EW. MmpL3 is a lipid transporter that binds trehalose monomycolate and phosphatidylethanolamine. *Proc Natl Acad Sci U S A.* 2019;116(23):11241-6.

38. Matsunaga I, Naka T, Talekar RS, McConnell MJ, Katoh K, Nakao H, et al. Mycolyltransferase-mediated glycolipid exchange in Mycobacteria. *J Biol Chem.* 2008;283(43):28835-41.

39. Savage AK, Constantinides MG, Han J, Picard D, Martin E, Li B, et al. The transcription factor PLZF directs the effector program of the NKT cell lineage. *Immunity.* 2008;29(3):391-403.

40. Seach N, Guerri L, Le Bourhis L, Mburu Y, Cui Y, Bessoles S, et al. Double-positive thymocytes select mucosal-associated invariant T cells. *J Immunol.* 2013;191(12):6002-9.

41. Shao MM, Yi FS, Huang ZY, Peng P, Wu FY, Shi HZ, et al. T Cell Receptor Repertoire Analysis Reveals Signatures of T Cell Responses to Human *Mycobacterium tuberculosis*. *Front Microbiol.* 2022;13:829694.

42. Roy S, Ly D, Li NS, Altman JD, Piccirilli JA, Moody DB, et al. Molecular basis of mycobacterial lipid antigen presentation by CD1c and its recognition by alphabeta T cells. *Proc Natl Acad Sci U S A.* 2014;111(43):E4648-57.

43. Almeida CF, Smith DGM, Cheng TY, Harpur CM, Batleska E, Nguyen-Robertson CV, et al. Benzofuran sulfonates and small self-lipid antigens activate type II NKT cells via CD1d. *Proc Natl Acad Sci U S A.* 2021;118(34).

44. Rosati E, Rios Martini G, Pogorelyy MV, Minervina AA, Degenhardt F, Wendorff M, et al. A novel unconventional T cell population enriched in Crohn's disease. *Gut.* 2022;71(11):2194-204.

45. Minervina AA, Pogorelyy MV, Paysen S, Luening U, Degenhardt F, Franke A, et al. Crohn's-associated invariant T cells (CAITs) recognise small sulfonate molecules on CD1d. *Gut.* 2023;73(1):205-6.

46. Daley CL, Iaccarino JM, Lange C, Cambau E, Wallace RJ, Jr., Andrejak C, et al.

Treatment of nontuberculous mycobacterial pulmonary disease: an official ATS/ERS/ESCMID/IDSA clinical practice guideline. *Eur Respir J.* 2020;56(1).

47. Nishimura N, Tomiyasu N, Torigoe S, Mizuno S, Fukano H, Ishikawa E, et al. Mycobacterial mycolic acids trigger inhibitory receptor Clec12A to suppress host immune responses. *Tuberculosis (Edinb).* 2023;138:102294.

48. Fujiwara N, Naka T, Ogawa M, Yamamoto R, Ogura H, and Taniguchi H. Characteristics of *Mycobacterium smegmatis* J15cs strain lipids. *Tuberculosis (Edinb).* 2012;92(2):187-92.

49. Fujiwara N, Oka S, Ide M, Kashima K, Honda T, and Yano I. Production and partial characterization of antibody to cord factor (trehalose 6,6'-dimycolate) in mice. *Microbiol Immunol.* 1999;43(8):785-93.

50. Matsunaga I, Oka S, Fujiwara N, and Yano I. Relationship between induction of macrophage chemotactic factors and formation of granulomas caused by mycoloyl glycolipids from *Rhodococcus ruber* (*Nocardia rubra*). *J Biochem.* 1996;120(3):663-70.

51. Naka T, Maeda S, Niki M, Ohara N, Yamamoto S, Yano I, et al. Lipid phenotype of two distinct subpopulations of *Mycobacterium bovis* Bacillus Calmette-Guerin Tokyo 172 substrain. *J Biol Chem.* 2011;286(51):44153-61.

52. Ueda S, Fujiwara N, Naka T, Sakaguchi I, Ozeki Y, Yano I, et al. Structure-activity relationship of mycoloyl glycolipids derived from *Rhodococcus sp*. 4306. *Microb Pathog.* 2001;30(2):91-9.

53. Tahiri N, Fodran P, Jayaraman D, Buter J, Witte MD, Ocampo TA, et al. Total Synthesis of a Mycolic Acid from *Mycobacterium tuberculosis*. *Angew Chem Int Ed Engl.* 2020;59(19):7555-60.

54. Bhatt A, Fujiwara N, Bhatt K, Gurcha SS, Kremer L, Chen B, et al. Deletion of kasB in *Mycobacterium tuberculosis* causes loss of acid-fastness and subclinical latent tuberculosis in immunocompetent mice. *Proc Natl Acad Sci U S A.* 2007;104(12):5157-62.

55. Nishizawa M, Yamamoto H, Imagawa H, Barbier-Chassefiere V, Petit E, Azuma I, et al. Efficient syntheses of a series of trehalose dimycolate (TDM)/trehalose dicorynomycolate (TDCM) analogues and their interleukin-6 level enhancement activity in mice sera. *J Org Chem.* 2007;72(5):1627-33.

56. Lu X, Hosono Y, Nagae M, Ishizuka S, Ishikawa E, Motooka D, et al.

Identification of conserved SARS-CoV-2 spike epitopes that expand public cTfh clonotypes in mild COVID-19 patients. *J Exp Med.* 2021;218(12).

57. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods.* 2015;12(5):380-1.

58. Yamasaki S, Ishikawa E, Sakuma M, Ogata K, Sakata-Sogawa K, Hiroshima M, et al. Mechanistic basis of pre-T cell receptor-mediated autonomous signaling critical for thymocyte development. *Nat Immunol.* 2006;7(1):67-75.

59. Matsumoto Y, Kishida K, Matsumoto M, Matsuoka S, Kohyama M, Suenaga T, et al. A TCR-like antibody against a proinsulin-containing fusion peptide ameliorates type 1 diabetes in NOD mice. *Biochem Biophys Res Commun.* 2021;534:680-6.

60. Boulter JM, Glick M, Todorov PT, Baston E, Sami M, Rizkallah P, et al. Stable, soluble T-cell receptor molecules for crystallization and therapeutics. *Protein Eng.* 2003;16(9):707-11.

61. Froning K, Maguire J, Sereno A, Huang F, Chang S, Weichert K, et al. Computational stabilization of T cell receptors allows pairing with antibodies to form bispecifics. *Nat Commun.* 2020;11(1):2330.

62. Kabsch W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr.* 2010;66(Pt 2):133-44.

63. Potterton E, Briggs P, Turkenburg M, and Dodson E. A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr.* 2003;59(Pt 7):1131-7.

64. Vagin AA, Steiner RA, Lebedev AA, Potterton L, McNicholas S, Long F, et al. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr.* 2004;60(Pt 12 Pt 1):2184-95.

65. Liebschner D, Afonine PV, Baker ML, Bunkoczi G, Chen VB, Croll TI, et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol.* 2019;75(Pt 10):861-77.

66. Emsley P, Lohkamp B, Scott WG, and Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr.* 2010;66(Pt 4):486-501.

67. Schorb M, Haberbosch I, Hagen WJH, Schwab Y, and Mastronarde DN. Software

tools for automated transmission electron microscopy. *Nat Methods.* 2019;16(6):471-7.

68. Punjani A, Rubinstein JL, Fleet DJ, and Brubaker MA. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods.* 2017;14(3):290-6.

69. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* 2018;27(1):293-315.

70. Gherardin NA, Redmond SJ, McWilliam HEG, Almeida CF, Gourley KHA, Seneviratna R, et al. CD36 family members are TCR-independent ligands for CD1 antigen-presenting molecules. *Sci Immunol.* 2021;6(60).

Figure 1

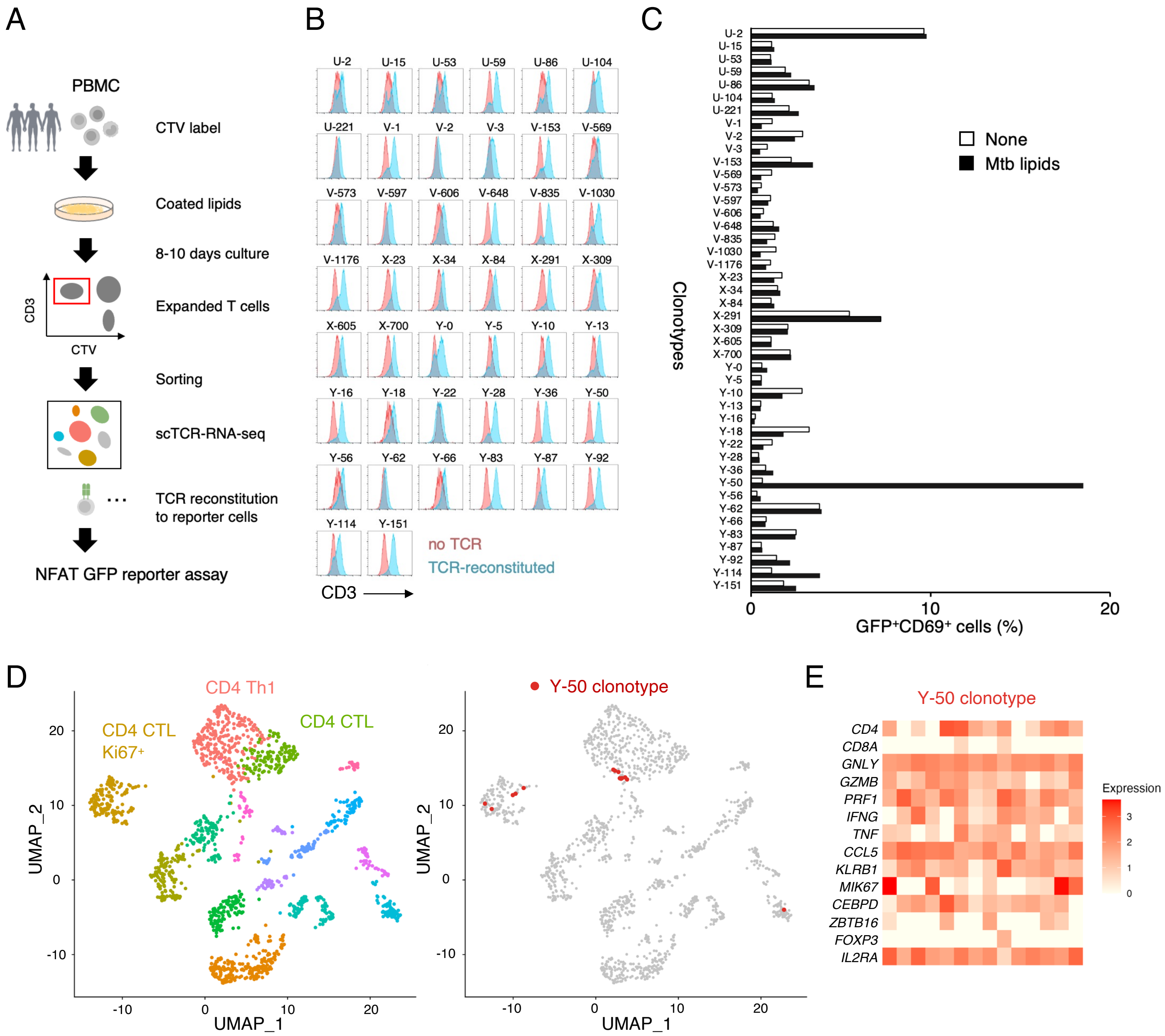

Figure 1. Identification of mycobacterial lipid-reactive T cells.

A) Schematic representation of experimental procedure. Human PBMCs were cultured with plate-coated crude lipids extracted from *M. tb*. The expanded T cells were sorted and analyzed by single-cell TCR-RNA sequencing. Highly expanded CTV^{lo} TCR clonotypes were reconstituted into NFAT-GFP reporter cells to examine the reactivity to *M. tb* lipids.

B) Forty-four TCR clonotypes were reconstituted into reporter cells and analyzed for their surface expression using anti-CD3 antibody.

C) NFAT-GFP reporter cells (44 clonotypes) expressing each different TCR were stimulated with *M. tb* crude lipids in the presence of PMBCs or cytokine-differentiated monocytes as APCs, and after 20 h incubation, analyzed for GFP and CD69 expression. Representative results from two independent experiments are shown.

D) UMAP plot of T cells expanded in response to *M. tb* lipids (left panel). T cell clones expressing Y-50 clonotype are highlighted in red dots (right panel). CTL, cytotoxic T lymphocytes.

E) Gene expression signature of Y-50 cells. The expression of characteristic genes in each cell expressing Y-50 TCR clonotype are shown.

Figure 2

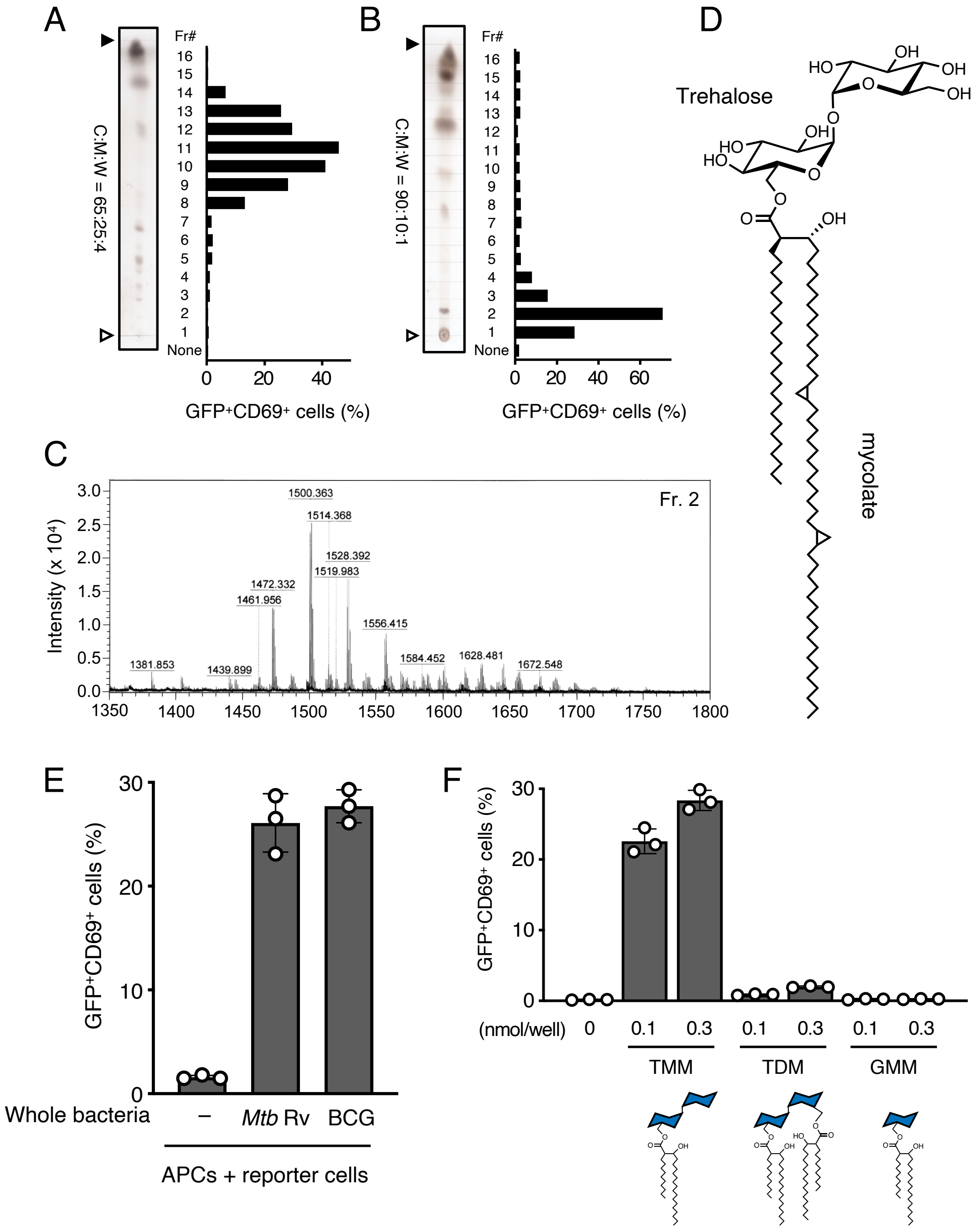

Figure 2. Identification of TMM as a T cell antigen.

A-B) *M. tuberculosis* H37Rv crude lipids were fractionated by HPTLC using chloroform/methanol/water (65:25:4; vol/vol/vol (A) and 90:10:1; vol/vol/vol (B)) and stained with copper(II) acetate-phosphoric acid. Y-50 reporter cells were stimulated with each fraction in the presence of APCs and analyzed for GFP and CD69 expression. Open and closed arrowheads denote the origin and the solvent front, respectively.

C) MALDI-TOF MS spectrum of lipid fraction 2 (Fr2).

D) Chemical structure of TMM of α-mycolate is shown, and methoxy-mycolate, and keto-mycolate are the other major subclasses of mycolate found in *M. tuberculosis* TMM.

E) Y-50 reporter cells were co-cultured with cytokine-differentiated human monocytes pre-incubated with whole bacteria (heat killed *M. tb* H37Rv or living BCG) and analyzed for GFP and CD69 expression.

F) Y-50 reporter cells were stimulated with the indicated concentration of TMM, TDM or GMM. GFP and CD69 expression are shown in bar graphs. Schematic ligand structures are shown below.

Data are shown as the means ± SD of triplicate assays (E, F) and representative results from two independent experiments are shown (A, B, E and F).

Figure 3

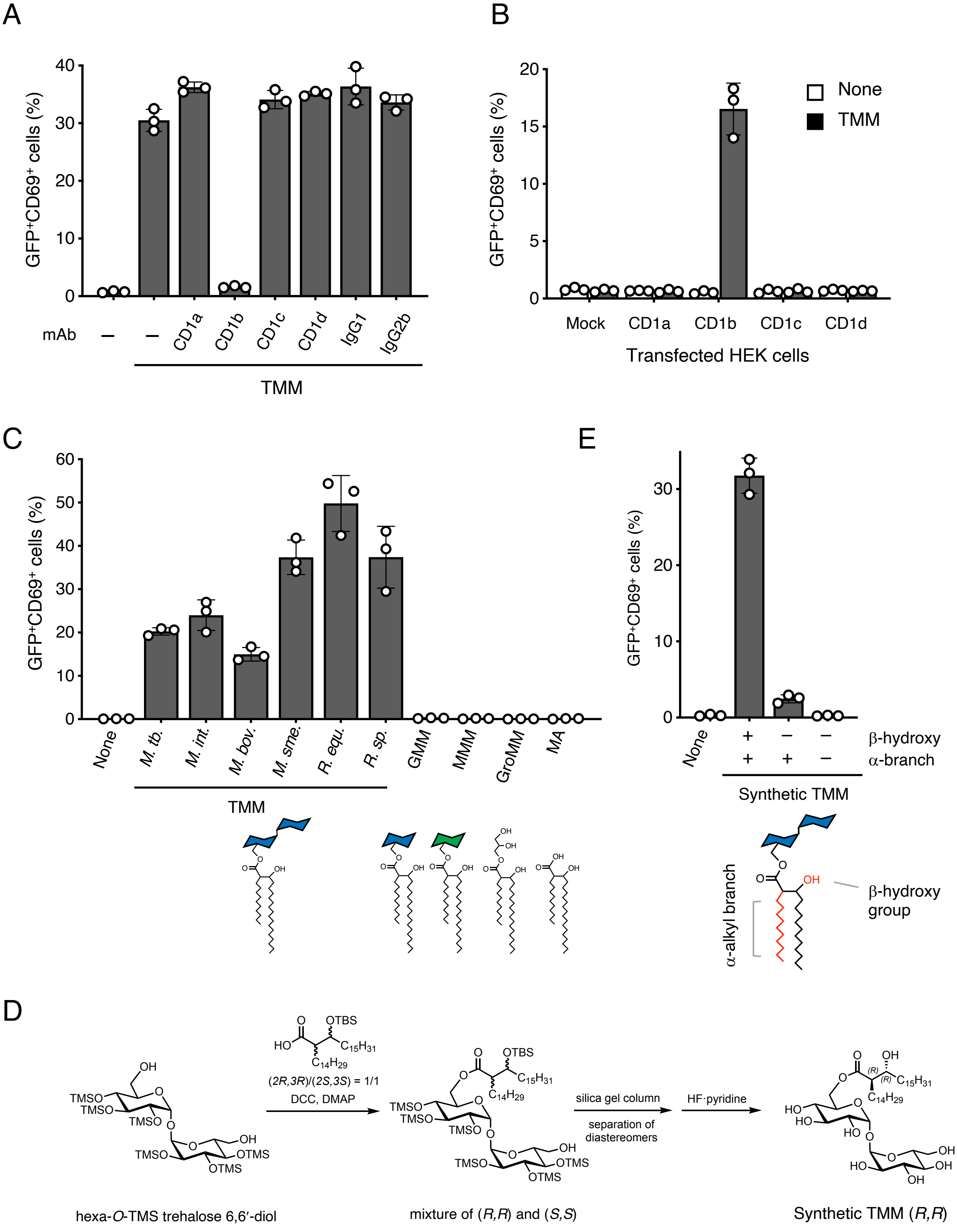

Figure 3. CD1b restricts TMM recognition by Y-50 T cells.

A) Y-50 reporter cells were co-cultured with cytokine-differentiated human monocytes and TMM (0.3 nmol /well) in the presence of 5 µg/ml of anti-CD1a, CD1b, CD1c, CD1d or isotype control antibodies (IgG1 and IgG2b) and analyzed for GFP and CD69 expression.

B) The reporter cells expressing Y-50 TCR were stimulated with TMM (1 nmol/well) in the presence of HEK293T cells transfected with human CD1a, CD1b, CD1c or CD1d.

C) Y-50 reporter cells were stimulated with TMM (1 nmol/well) purified from *M. tuberculosis* CDC1551, *M. bovis* BCG, *M. intracellulare, M. smegmatis*, *Rhodococcus equi* and *R. sp* 4306. Also, GMM, mannose-monomycolate (MMM), glycerol monomycolate (GroMM) and mycolic acid (MA) were tested in the presence of human CD1b-expressing DC2.4 cells (CD1b-DC2.4).

D) Synthetic scheme for TMM.

E) Y-50 reporter cells were stimulated with synthetic TMM harboring a β-hydroxy group (hydroxy) and α-branched alkyl chains (branch) or synthetic analogues lacking hydroxy (–, +) or both moieties (–, –) in the presence of CD1b-DC2.4 cells as APCs.

Data are shown as the means ± SD of triplicate assays and representative results from two independent experiments are shown (A, B, C and E).

Figure 4

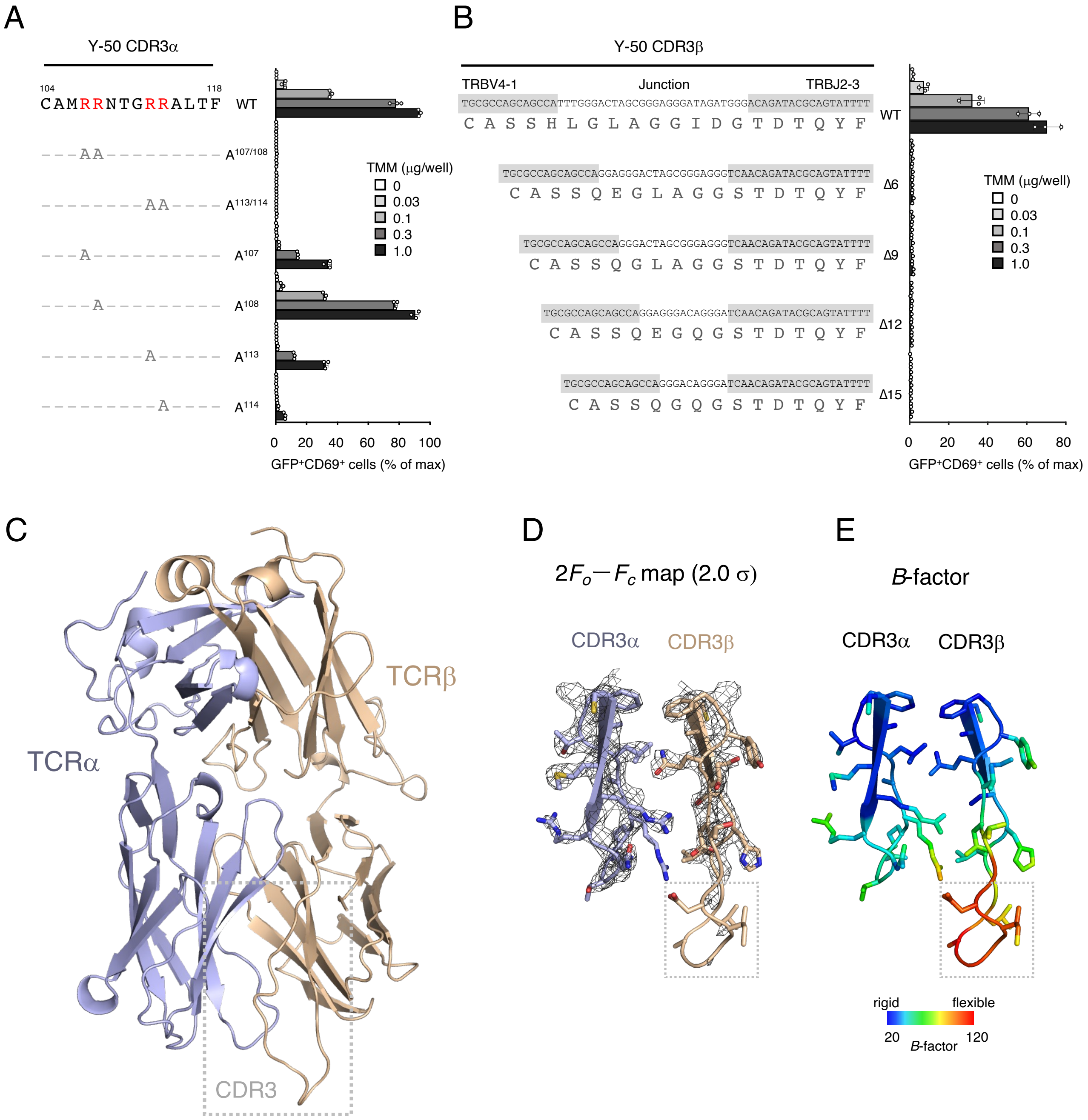

Figure 4. Mutagenesis and structural analysis of TMM-reactive TCR.

A) The amino acid sequences of Y-50 CDR3α within the arginine mutants shown in red. TMM reactivities of each mutant were shown as percentage of the maximum response induced by plate coated anti-CD3 Ab. The number of amino acids were shown in accordance with the ImMunoGeneTics (IMGT) definition (https://imgt.org/IMGTScientificChart/).

B) Nucleotide and amino acid sequences of the Y-50 TCR CDR3β region and its junction-deletion mutants (Δ). D region and N or P nucleotide sequences that constitute junctional sequences were unshaded. Cells were stimulated as indicated in (A).

C) Crystal structure of the Y-50 TCRαβ heterodimer (PDB: 8XUB). The main chains of TCRα and β are shown in violet and brown, respectively. CDR3αβ regions are boxed.

D-E) 2*Fo-Fc* map contoured at 2.0 σ (D) and *B*-factor diagram (E) of CDR3αβ are shown as gray mesh and color gradient, respectively. Junction regions of CDR3β are boxed.

Data are shown as mean ± SD of triplicates (A and B) and representative results from two independent experiments are shown.

Figure 5

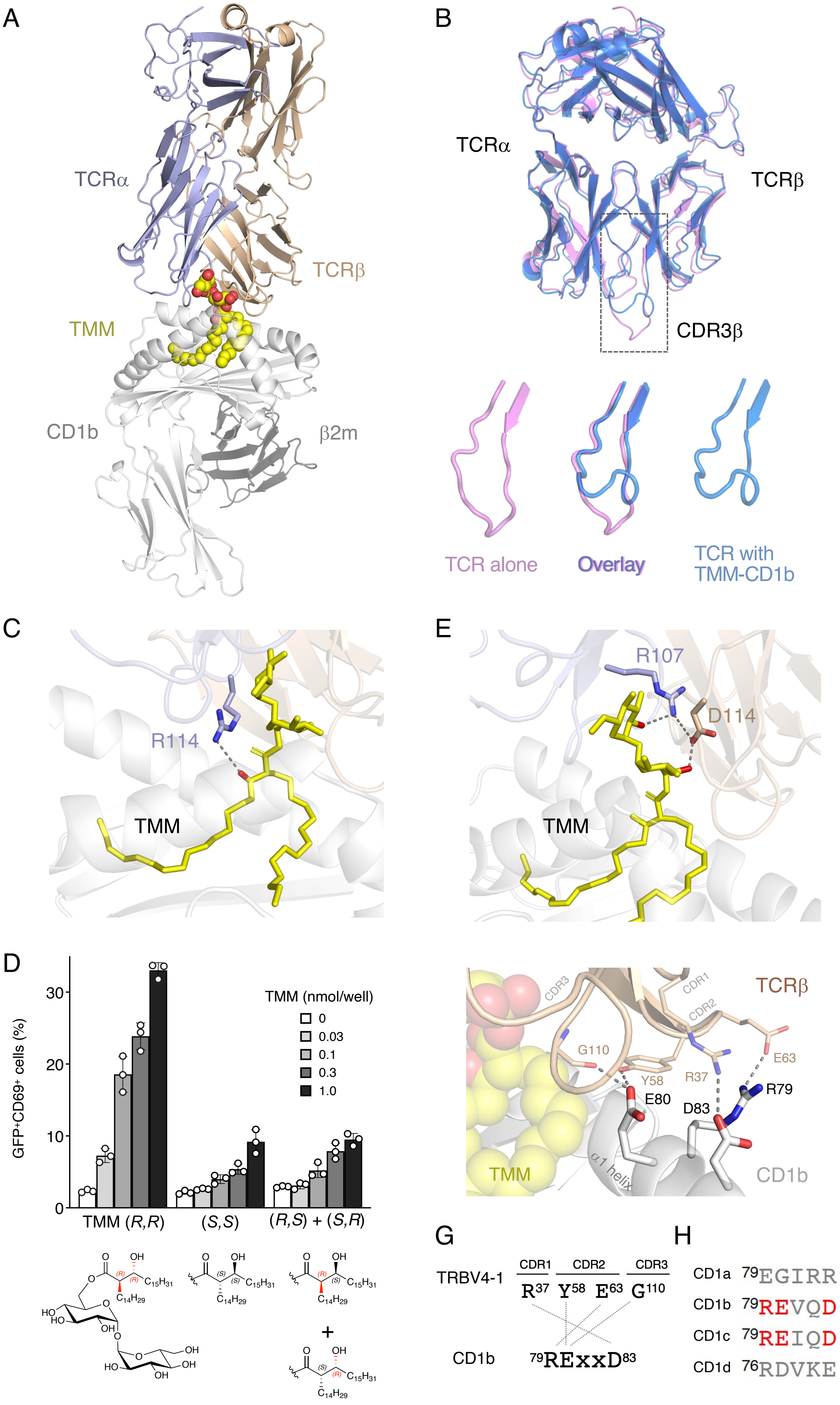
A
TCRα
TCRβ
TMM
CD1b
β2m
B
TCRα
TCRβ
CDR3β
TCR alone
Overlay
TCR with TMM-CD1b
C
R114
TMM
E
R107
D114
TMM
CDR3
CDR1
CDR2
TCRβ
G110
Y58
R37
E63
E80
D83
R79
α1 helix
TMM
CD1b
D
GFP⁺CD69⁺ cells (%)
0
10
20
30
TMM (nmol/well)
☐ 0
☐ 0.03
☐ 0.1
☐ 0.3
■ 1.0
TMM (R,R)
(S,S)
(R,S) + (S,R)
OH
C15H31
C14H29
(R)
(S)
HO
OH
O
+
G
TRBV4-1
CDR1
CDR2
CDR3
R37 Y58 E63 G110
CD1b
79RExxD83
H
CD1a 79EGIRR
CD1b 79REVQD
CD1c 79REIQD
CD1d 76RDVKE

Figure 5. Ternary complex structure of Y-50 TCR-TMM-CD1b.

A) Overall structure of the Y-50 TCR-TMM-CD1b complex. The main chains of TCRα, TCRβ, and CD1b are shown as indicated. TMM was presented as yellow sphere.

B) Superimposition of the structure of Y-50 TCR alone (PDB: 8XUB)(pink) and Y-50 TCR bound to TMM-CD1b (PDB: 8ZOX)(blue)(upper panel). CDR3β regions (boxed) are magnified (lower panels).

C) Close-up view of TMM (*R,R*) and the side chain of R114 within CDR3α. β-hydroxy group of TMM is shown in red.

D) Y-50 reporter cells were stimulated with natural configuration of synthetic TMM (*R,R*) or non-natural stereoisomers, (*S,S*) or (*S,R+R,S*), in the presence of CD1b-DC2.4 and analyzed for GFP and CD69 expression. Structures of stereoisomers are shown below (*R*, red; *S*, black). Data are shown as the means ± SD of triplicates and a representative result from two independent experiments is shown.

E) Close-up view of TMM (*R,R*) and the side chain of R107 (CDR3α) and D114 (CDR3β). Hydroxy groups of TMM that formed hydrogen bonds to the TCR side chains are shown in red.

F) Close-up view of the side chains of R79, E80 and D83 in CD1b that interact with the side chains of R37 (CDR1β), Y58 and E63 (CDR2β) and G110 (CDR3β).

G) Multi-bonded interaction of CD1b RExxD motif and TRBV4-1 residues. Individual interaction was shown as dotted lines.

H) Conservation of RExxD motif in human CD1b and CD1c. Amino acid sequence of CD1a (NP_001307581), CD1b (NP_001755.1), CD1c (NP_001756.2) and CD1d (NP_001306074) are aligned. Numbers indicate the amino acid position of mature peptide (excluding signal peptide).

Figure 6

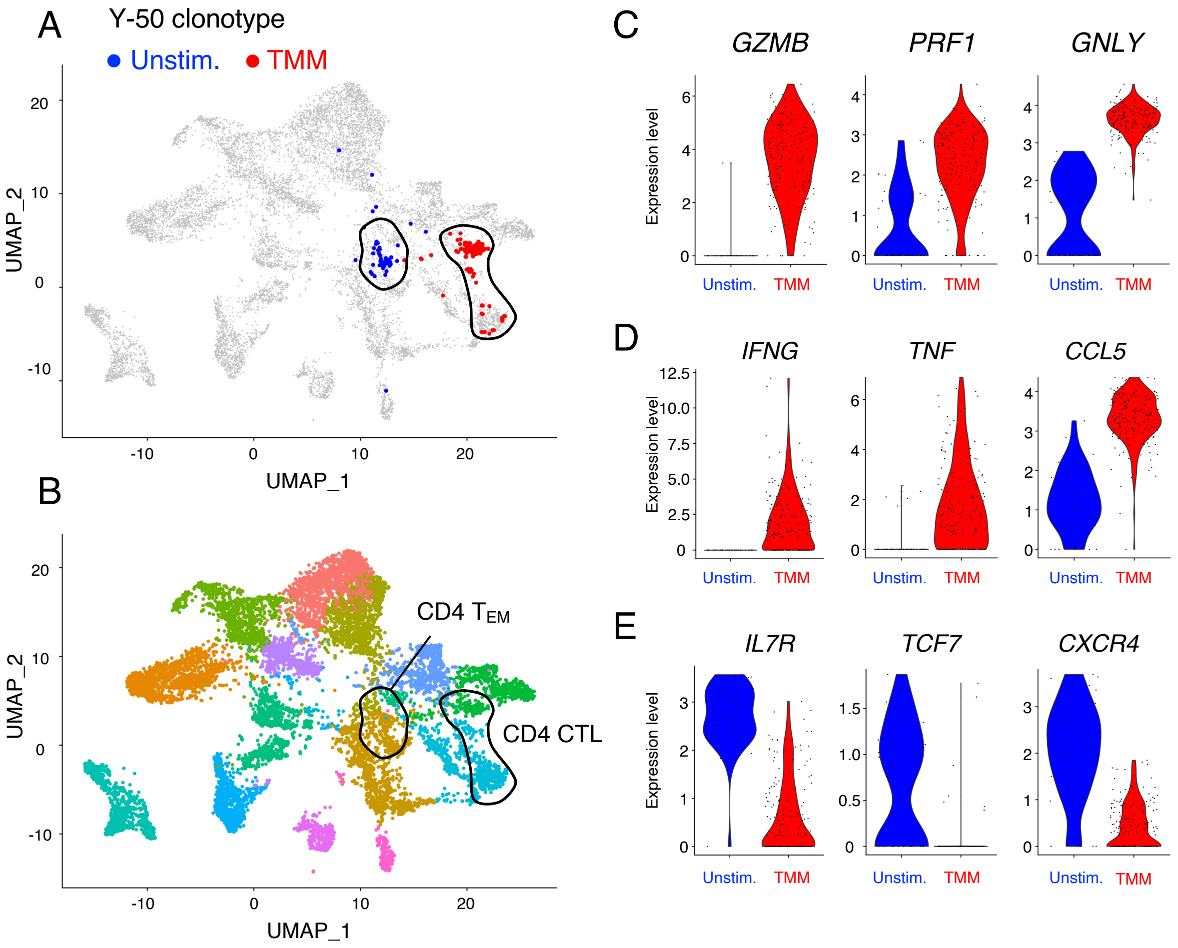

Figure 6. Functional maturation of TMM-reactive T cells upon TMM stimulation.
A-B) Cluster shift of Y-50 clonotype before and after TMM stimulation. T cells expressing Y-50 clonotype defined by scTCR-RNA-seq were overlayed (A) on UMAP plot of PBMCs from donors including the donor used in Figure 1 (B) as described in Material and Methods. T_{EM}, effector memory T cells. CTL, cytotoxic T lymphocytes.
C-E) Differentially expressed genes in Y-50 T cells upon TMM stimulation. The expression of representative genes encoding cytotoxic effector molecules (C), pro-inflammatory cytokines and chemokines (D) and stemness-related molecules (E) are shown in violin plots.

Figure 7

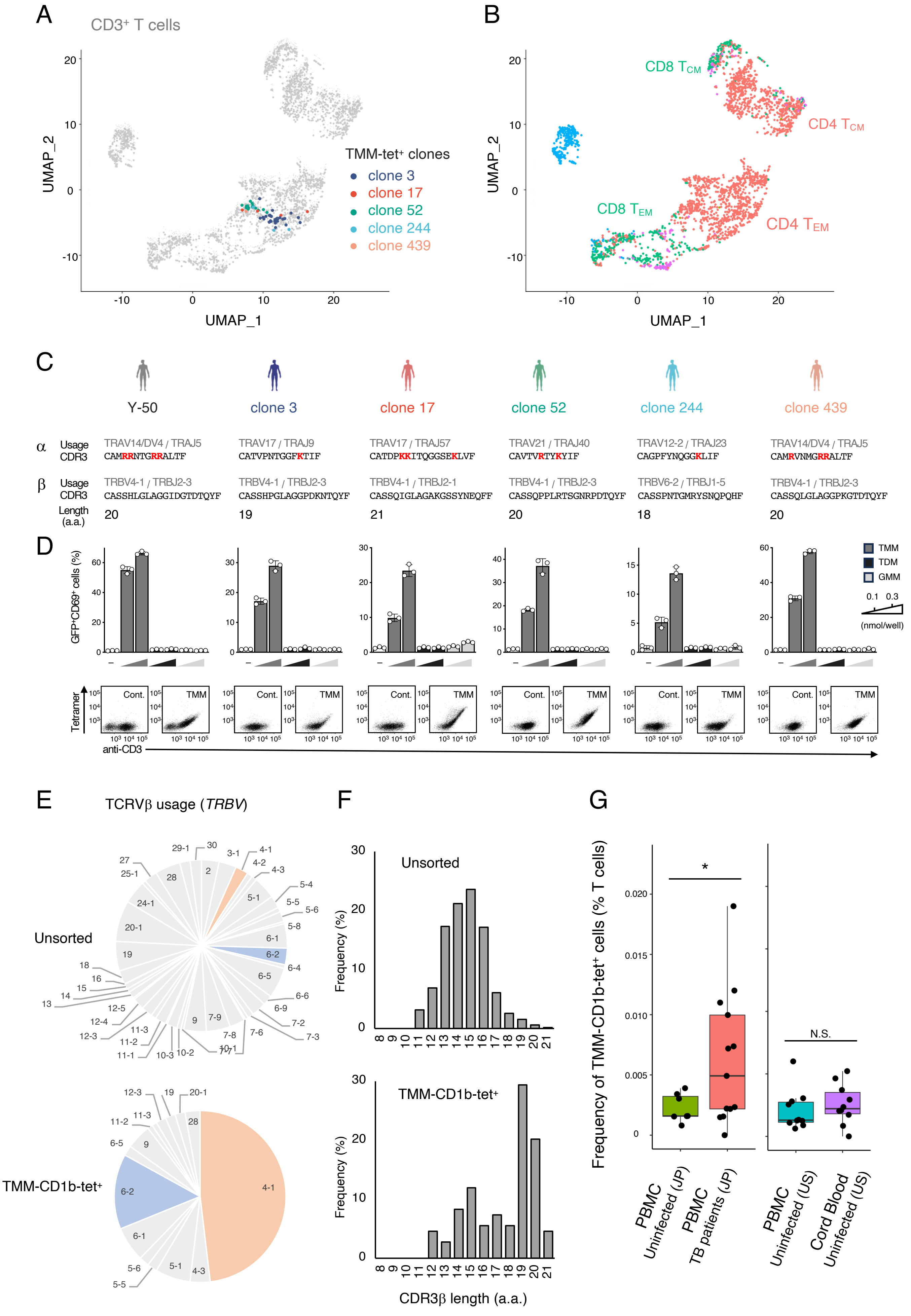
A
CD3+ T cells
UMAP_2
UMAP_1
TMM-tet+ clones
clone 3
clone 17
clone 52
clone 244
clone 439
B
CD8 TCM
CD4 TCM
CD8 TEM
CD4 TEM
C
Y-50
clone 3
clone 17
clone 52
clone 244
clone 439
α Usage CDR3
TRAV14/DV4 / TRAJ5
CAMRRNTGRRALTF
TRAV17 / TRAJ9
CATVPNTGGFKTIF
TRAV17 / TRAJ57
CATDPKKITQGGSEKLVF
TRAV21 / TRAJ40
CAVTVRTYKYIF
TRAV12-2 / TRAJ23
CAGPFYNQGGKLIF
TRAV14/DV4 / TRAJ5
CAMRVNMGRRALTF
β Usage CDR3
TRBV4-1 / TRBJ2-3
CASSHLGLAGGIDGTDTQYF
TRBV4-1 / TRBJ2-3
CASSHPGLAGGPDKNTQYF
TRBV4-1 / TRBJ2-1
CASSQIGLAGAKGSSYNEQFF
TRBV4-1 / TRBJ2-3
CASSQPPLRTSGNRPDTQYF
TRBV6-2 / TRBJ1-5
CASSPNTGMRYSNQPQHF
TRBV4-1 / TRBJ2-3
CASSQLGLAGGPKGTDTQYF
Length (a.a.)
20
19
21
20
18
20
D
GFP+CD69+ cells (%)
TMM
TDM
GMM
0.1 0.3
(nmol/well)
Cont.
TMM
Tetramer
anti-CD3
E
TCRVβ usage (TRBV)
Unsorted
TMM-CD1b-tet+
F
Unsorted
TMM-CD1b-tet+
Frequency (%)
CDR3β length (a.a.)
G
Frequency of TMM-CD1b-tet+ cells (% T cells)
*
N.S.
PBMC Uninfected (JP)
PBMC TB patients (JP)
PBMC Uninfected (US)
Cord Blood Uninfected (US)

Figure 7. TMM-specific T cells with similar characteristics are shared among individuals.

A-B) Frequent TMM-specific clonotypes identified by TMM-CD1b-tetramer sorting and scTCR-RNA-seq were overlayed (A) on UMAP plot of all TMM-tetramer-sorted T cells and unsorted CD3$^+$ T cells from 13 healthy donors (B). Three clones were detected from different individual donors; two clones (clone 17 and 439) were from another donor. CD4$^+$ T_{EM}, CD4$^+$ effector memory T cells. CD4 T_{CM}, CD4$^+$ central memory T cells. Naïve T cells were rare within TMM-tetramer$^+$ cells and were not clustered on UMAP.

C) TCR usages, CDR3 sequences and length of CDR3β region of clonotypes detected in (A).

D) Each clonotype was reconstituted into reporter cells and analyzed for TMM, TDM and GMM reactivity using CD1b-DC2.4 as APCs. Reporter cells were stained with PE-conjugated endo-CD1b (Cont.) or TMM-CD1b tetramers (TMM) and anti-CD3 antibodies. Y-50 TCR is shown as a control. Data are shown as the means ± SD of triplicate assays and representative results from two independent experiments are shown.

E-F) Frequency of TCRVβ usage (D) and length of CDR3β region (E) of unsorted or top 27 TMM-CD1b tetramer$^+$ T cell clonotypes.

G) PBMCs from Japanese donors (uninfected donors, n = 7; TB patients, n = 13) or PBMC (n = 10) and cord blood cells (n = 10) from uninfected US donors were stained with PE-conjugated TMM-loaded CD1b tetramer, APC-conjugated CD1b-endo tetramer and anti-CD3 antibody. The percentages of TMM-CD1b tetramer positive and endo-CD1b tetramer negative population in CD3$^+$ T cells are shown (TMM-CD1b-tet$^+$). Medians are indicated with horizontal bars. *P* values were calculated using unpaired two-tailed Welch's t-test; *$P < 0.05$; N.S., not significant.