Transcriptomic Analysis of CD4+ T Cells Reveals Novel Immune Signatures of Latent Tuberculosis

Julie G. Burel, Cecilia S. Lindestam Arlehamn, Nabeela Khan, Grégory Seumois, Jason A. Greenbaum, Randy Taplitz, Robert H. Gilman, Mayuko Saito, Pandurangan Vijayanand, Alessandro Sette and Bjoern Peters

*J Immunol* 2018; 200:3283-3290; Prepublished online 30 March 2018;
doi: 10.4049/jimmunol.1800118
http://www.jimmunol.org/content/200/9/3283

Supplementary Material

http://www.jimmunol.org/content/suppl/2018/03/30/jimmunol.1800118.DCSupplemental

References

This article cites 46 articles, 10 of which you can access for free at:
http://www.jimmunol.org/content/200/9/3283.full#ref-list-1

Why *The JI*? Submit online.

- **Rapid Reviews!** 30 days* from submission to initial decision
- **No Triage!** Every submission reviewed by practicing scientists
- **Fast Publication!** 4 weeks from acceptance to publication

*average

Subscription

Information about subscribing to *The Journal of Immunology* is online at:
http://jimmunol.org/subscription

Permissions

Submit copyright permission requests at:
http://www.aai.org/About/Publications/JI/copyright.html

Email Alerts

Receive free email-alerts when new articles cite this article. Sign up at:
http://jimmunol.org/alerts
Transcriptomic Analysis of CD4+ T Cells Reveals Novel Immune Signatures of Latent Tuberculosis

Julie G. Burel,* Cecilia S. Lindestam Arlehamn,* Nabeela Khan,* Grégory Seumois,* Jason A. Greenbaum,* Randy Taplitz,† Robert H. Gilman,*‡ Mayuko Saito,‡§¶ Pandurangan Vijayanand,* Alessandro Sette,*¶ Il and Bjorn Peters*¶ Il

In the context of infectious diseases, cell population transcriptomics are useful to gain mechanistic insight into protective immune responses, which is not possible using traditional whole-blood approaches. In this study, we applied a cell population transcriptomics strategy to sorted memory CD4 T cells to define novel immune signatures of latent tuberculosis infection (LTBI) and gain insight into the phenotype of tuberculosis (TB)-specific CD4 T cells. We found a 74-gene signature that could discriminate between memory CD4 T cells from healthy latently Mycobacterium tuberculosis–infected subjects and noninfected controls. The gene signature presented a significant overlap with the gene signature of the Th1* (CCR6 CXCR3 CCR4*) subset of CD4 T cells, which contains the majority of TB-specific reactivity and is expanded in LTBI. In particular, three Th1* genes (ABCB1, c-KIT, and GPA33) were differentially expressed at the RNA and protein levels in memory CD4 T cells of LTBI subjects compared with controls. The 74-gene signature also highlighted novel phenotypic markers that further defined the CD4 T cell subset containing TB specificity. We found the majority of TB-specific epitope reactivity in the CD62L - GPA33 Th1* subset. Thus, by combining cell population transcriptomics and single-cell protein-profiling techniques, we identified a CD4 T cell immune signature of LTBI that provided novel insights into the phenotype of TB-specific CD4 T cells. The Journal of Immunology, 2018, 200: 3283–3290.

T
uberculosis (TB) is the ninth leading cause of death worldwide and the leading cause of death from a single infectious agent in humans (1). Upon infection with Mycobacterium tuberculosis, the majority of individuals develop asymptomatic latent TB infection (LTBI). Active disease (pulmonary or extrapulmonary) can develop directly after primary infection or as a progression from latent infection. Defining the immune responses associated with LTBI is crucial to understand how most individuals can control the infection at the stage of LTBI, whereas others develop active disease.

Whole-blood transcriptomic studies in TB have proven useful to discriminate individuals with active TB from other cohorts, such as LTBI or uninfected individuals (2–6). However, these studies have been less successful at gaining mechanistic insights into what types of immune responses are associated with control of TB, as in the case of LTBI. Gene signatures derived from whole blood are most suited to detect changes in the most abundant cell populations, which might not necessarily be of high relevance in the context of immune mechanisms. For instance, higher neutrophil (7) and monocyte (8) counts in peripheral blood have been described in active TB patients compared with LTBI or uninfected individuals, and these two cell subsets account for the majority of WBCs. Accordingly, a type I IFN gene signature of active TB was identified by Berry et al. (2) in whole blood that is likely driven by neutrophils, and a recently described 16-gene signature that can predict active TB progression among LTBI subjects primarily contains monocyte-associated genes (9). These signatures and associated cell subsets are proinflammatory signatures that are more likely to reflect secondary, rather than primary, immune responses. An alternative strategy is to perform transcriptomic analysis of isolated immune cell subsets of interest. Such an approach has been shown to increase the power of identifying diagnostically, prognostic, and mechanistic gene signatures for a wide range of diseases, including systemic lupus erythematosus (10), chronic inflammatory bowel disease (11), HIV infection (12), and asthma (13). In the case of TB, transcriptomic profiling of macrophages after in vitro infection with M. tuberculosis highlighted the previously unsuspected functional role of IL-32 in host defense and susceptibility to infection (14).

Several lines of evidence highlight the mechanistic role of CD4 T cells in the control of TB infection (15, 16). In particular, the
phenotype and magnitude of TB-specific CD4 T cell responses have been shown to be important for protection in the context of natural infection (17, 18) and immunization (19, 20). Accordingly, we performed transcriptomic studies on sorted memory CD4 T cells to discover novel immune signatures of LTBI and refine the phenotype of TB-specific CD4 T cells. Although previous studies have found very few differentially expressed genes between LTBI subjects and uninfected controls in whole blood (2, 21), we hypothesized that the comparison of the gene-expression profile of CD4 memory T cells between these two groups would have a higher resolution power and that differences will be directly associated with the presence of TB-specific CD4 T cells in the LTBI cohort. Consistent with our hypothesis, we found that a 74-gene signature was differentially expressed in memory CD4 T cells of LTBI subjects compared with controls. This gene signature presented a significant overlap with the gene signature of Th1+, which is the main CD4 T cell subset containing TB-specific peptide reactivity and expanded in LTBI. By combining the transcriptomic data with single-cell protein profiling, we further defined the phenotype of the Th1+ subset displaying TB-specific reactivity and identified novel proteins as promising biomarkers for TB-specific CD4 T cells in the context of LTBI.

Materials and Methods

Ethics statement

Blood samples were obtained from the University of California, San Diego Anti-Viral Research Center clinic and the Universidad Peruana Cayetano Heredia. All samples were obtained for specific use in this study. Ethical approval to carry out this work is maintained through the La Jolla Institute for Allergy and Immunology Institutional Review Board and through Johns Hopkins School of Public Health Institutional Review Board (R.H.G. holds a dual appointment at Universidad Peruana Cayetano Heredia and Johns Hopkins University). All clinical investigations have been conducted according to the principles expressed in the Declaration of Helsinki, and all participants provided written informed consent prior to participation in the study.

Subjects and samples

LTBI status was confirmed in subjects by a positive IFN-γ-release assay (QuantiFERON-TB Gold In-Tube; Cellestis or T-SPOT.TB; Oxford Immunotec) and the absence of clinical and radiographic signs of active TB. TB+ control subjects were also negative for the IFN-γ-release assay. PBMCs were obtained by density gradient centrifugation (Ficoll-Hypaque, Amersham Biosciences) from 100-ml leukapheresis or whole-blood samples, according to the manufacturer’s instructions. Cells were resuspended at 50–100 million cells per milliliter in PBS (Gemini Bio-Products) containing 10% DMSO (Sigma) and cryopreserved in liquid nitrogen.

TB-specific peptide pool

A peptide pool containing 300 M. tuberculosis-derived peptides (MTB300) was prepared as previously described (22). Briefly, individual peptides were resuspended in DMSO, and equal amounts of each peptide were pooled to construct the peptide pool. After lyophilization, the peptide pool was resuspended in DMSO (0.7 mg/ml per peptide) and stored at −80°C.

Bcl2 nuclear expression by flow cytometry

Two million thawed PBMCs were stained with fixable viability dye eFluor 506 (eBioscience) and with anti-human CD3, CD4, CD8, CD19, and CD45RO (see Supplemental Table I for Ab details) for 30 min at 4°C, protected from light. After two washes in PBS, cells were resuspended in 100 μl of PBS containing 0.5% BSA and 2 mM EDTA (pH 8; FACS Buffer), transferred into a 5-ml polystyrene FACS tube (BD Biosciences), and stored at 4°C protected until acquisition. FACS acquisition was performed on a BD LSR II cell analyzer (BD Biosciences), and gating was performed with FlowJo (version 10.1).

ABC1 activity assay

ABC1 activity assay was performed as described previously (23). Briefly, 3 million PBMCs were stained with 0.2 μg/ml Rhodamine 123 (Sigma) for 30 min at 4°C. Subsequently, cells were divided into two samples and incubated with medium alone or with medium containing 0.1 μM clodaric hydrochloride (ABCB1 inhibitor; Tocris Biosciences) for 1 h at 37°C. Following the eflux step, cells were stained with fixable viability dye eFluor 506 (eBioscience) and with anti-human CD3, CD4, CD8, CD19, CD14, CD45RA, and CCR7 for 30 min at 4°C. After two washes in PBS, cells were resuspended in 100 μl of PBS containing 0.5% BSA and 2 mM EDTA (pH 8; FACS Buffer), transferred into a 5-ml polystyrene FACS tube (BD Biosciences), and stored at 4°C protected from light for up to 1 h until FACS acquisition. FACS acquisition was performed on a BD LSR II cell analyzer (BD Biosciences), and gating was performed with FlowJo (version 10.1). Rhodamine 123 fluorescence was detected in the FITC channel.

Mass cytometry

Mass cytometry was performed as described previously (24). Briefly, 2 million PBMCs were stained with viability marker Cell-Id Cisplatin (Fluidigm) for 5 min at room temperature under continuous rotation. After two washes in PBS, cells were stained with anti-human CXCR3 and CCR6 for 30 min at 37°C, followed by staining with the remaining surface Abs for 30 min at room temperature, under continuous rotation (see Supplemental Table II for Ab details). Subsequently, cells were fixed in PBS with 2% paraformaldehyde overnight at 4°C. The following day, cells were permeabilized with Permeabilization Buffer (eBioscience). Just before sample acquisition, cellular DNA was labeled with Cell-Id Intercalar-Ir (Fluidigm). Samples were acquired with a Helios-2 mass cytometer (Fluidigm). At least 190,000 events were recorded per sample. Each fcs file was normalized using EQ Four Element Calibration Beads (Fluidigm) and analyzed with FlowJo (version 10.1).

Memory CD4 T cell sorting

Ten million PBMCs were stained with fixable viability dye eFluor 506 (eBioscience) with anti-human CD3, CD4, CD8, CD45RA, and CCR7 (see Supplemental Table I for Ab details), as described in the flow cytometry sections above. Cell sorting was performed on a BD FACSAria III/Fusion cell sorter (Becton Dickinson). A total of 100,000 memory CD4 T cells (see Supplemental Fig. 1A for gating strategy) was sorted into TRIZol LS reagent (Invitrogen).

RNA sequencing

RNA sequencing was performed as described previously (24). Briefly, total RNA was purified using an miRNeasy Micro Kit (QIAGEN) and quantified by quantitative PCR, as described previously (25). Purified total RNA (1–5 ng) was amplified following the Smart-Seq2 protocol (16 cycles of cDNA amplification) (26). cDNA was purified using Ampure XP beads (Beckman Coulter). From this step, 1 ng of cDNA was used to prepare a standard Nextera XT sequencing library (Nextera XT DNA sample preparation kit and index kit, Illumina). Whole-transcriptome amplification and sequencing-library preparations were performed in a 96-well format to reduce assay-to-assay variability. Quality-control steps were included to determine total RNA quality and quantity, the optimal number of PCR preamplification cycles, and fragment size selection. Samples that failed quality control were eliminated from further downstream steps. Barcoded Illumina sequencing libraries (Nextera; Illumina) were generated using the associated platform (Biomere FXp). Libraries were sequenced on a HiSeq 2500 Illumina platform to obtain 50-bp single-end reads (TruSeq Rapid kit; Illumina).

RNA sequencing analysis

RNA sequencing analysis was performed as previously described (24). Briefly, the single-end reads that passed Illumina filters were filtered for reads aligning to rRNA, rRNA, adapter sequences, and spike-in controls. The reads were then aligned to UCSC hg19 reference genome using TopHat (v 1.4.1) (27). DUST scores were calculated with PRINSEQ Lite (v 0.20.3) (28), and low-complexity reads (DUST > 2) were removed from the BAM files. The alignment results were parsed via SAMtools (29) to generate SAM files. Read counts for each genomic feature were obtained with the htseq-count program (v 0.6.0) (30) using the “union” option. After removing absent features (zero counts in all samples), the raw counts were imported to R/Bioconductor package DESeq2 (31) to identify differentially expressed genes among samples. Genes were considered differentially expressed for adjusted p values < 0.05 and absolute log2 fold change > 0.5. Principal component analysis was performed and heat maps were created using Qlucore on raw counts transformed with the rlog function in R. Pathway analysis enrichment was assessed using the Core Analysis function in Ingenuity Pathway Analysis (QIAGEN). The sequencing data

Downloaded from http://www.jimmunol.org/ by guest on June 4, 2021
presented in this study were submitted to the Gene Expression Omnibus under accession numbers GSE84445 and GSE99373 (https://www.ncbi.nlm.nih.gov/geo) and to ImmPort under study number SDY820 (http://www.immport.org).

TB-specific reactivity of sorted CD4 T cell subsets

A total of 300 million PBMCs was used to negatively isolate CD4 T cells using the CD4+ T Cell Isolation Kit (Miltenyi Biotec), according to the manufacturer’s instructions. Subsequently, CD4 T cells were stained with anti-human CXCR3 and CCR6 for 30 min at 37˚C, followed by staining with fixable viability dye eFluor 506 (eBioscience) and with anti-human CD3, CD4, CD8, CD19, CD14, CCR4, CD62L, GPA33, and CD45RO, as described in the flow cytometry sections above. Cell sorting was performed on a BD FACSAria III cell sorter (Becton Dickinson). CD4 T cell subsets, Th1* subsets, and APCs (see Supplemental Fig. 1B for gating strategy) were sorted in FACS tubes in FACS Buffer. After sorting, CD4 T cell subsets were resuspended in medium and plated in a 96-well U-bottom plate at 5 × 10^6 cells per milliliter overnight at 37˚C. APCs were added at 5 × 10^6 cells per milliliter. The following day, cells were stimulated with the TB-specific MTB300 peptide pool at 2 μg/ml (22) or with plate-bound anti-human CD3 and soluble anti-human CD28 at 1 μg/ml for 24 h at 37˚C. After stimulation, cells were washed and stained with anti-human OX40, PD-L1, and CD4 (see Supplemental Table I for Ab details), as described in the flow cytometry sections above. Acquisition was performed on a BD LSR II analyzer (BD Biosciences).

Results

A 74-gene signature can differentiate memory CD4 T cells of LTBI subjects from those of controls

To define the CD4 T cell immune signature of LTBI, we compared the genome-wide expression profile of memory CD4 T cells (excluding CD45RA+CCR7+ cells) isolated from 29 control subjects and 30 individuals with LTBI. We found a total of 74 differentially expressed genes (adjusted p value < 0.05 and absolute log2 fold change > 0.5, Fig. 1A, Supplemental Table III) between the two cohorts. To estimate the predictive value of this signature, we performed a 5-fold validation. Specifically, the dataset was randomly split into five sets. A gene signature was identified for each combination of four sets (adjusted p value < 0.05, absolute log2 fold change > 0.5 from the DESeq2 analysis), and the PC1 component of the signature was used to predict the LTBI status within the remaining fifth set. Across all dataset combinations, using a PC1 threshold of zero, the gene signature identified was able to correctly classify LTBI subjects with an average specificity of 83% and an average sensitivity of 93% (Fig. 1B). Eighteen of the 74 genes (24%) were consistently classified as differentially expressed in any dataset combination (Supplemental Table IV). Thus, the transcriptomic profile of memory CD4 T cells has robust predictive value to discriminate between LTBI and uninfected subjects.

Of the 74 genes, 50 (68%) were upregulated and 24 (32%) were downregulated in the LTBI cohort compared with controls (Fig. 1C). The 74-gene signature was positively associated with cell proliferation but negatively associated with cell apoptosis (Fig. 1D). Bcl2, an antiapoptotic molecule, was also upregulated in memory CD4 T cells of LTBI subjects compared with controls (Fig. 1E). Thus, memory CD4 T cells in LTBI subjects have a highly specific immune signature that is associated with a proliferative and antiapoptotic state.
The 74-gene signature reflects expansion of the Th1* subset that contains TB-specific reactivity

The memory CD4 T cell compartment from LTBI subjects is expected to contain TB-specific CD4 T cells that might be responsible for the transcriptional differences observed compared with uninfected controls. Comparison of the composition of the CD4 T cell compartment between the two cohorts identified significant differences in the CCR6+ compartment, with CCR6+ CXCR3+CCR4+ cells [also called Th1* (32)] and CCR6+ CXCR3+CCR4− cells (CCR6+ double negative [DNEG]) found at a higher frequency in LTBI subjects compared with controls (Fig. 2A). We (33, 34) and other investigators (35) have previously shown that the majority of TB-specific CD4 T cells are contained within the Th1* subset. We further assessed the TB-specific peptide reactivity of each memory CD4 T cell subset using MTB300 recognized by CD4 T cells (22). A recent genome-wide screen has shown that different cytokine profiles are being produced in response to distinct M. tuberculosis proteins, which do not always include IFN-γ (36). Thus, to capture M. tuberculosis–reactive CD4 T cells in the most unbiased fashion, we elected as a readout the expression of activation-induced markers on the cell surface rather than the production of cytokines. The upregulation of OX40 and PDL1 has been reported as a good surrogate marker for Ag specificity in nonregulatory CD4 T cells (37), and OX40+PDL1+ cells were consistently induced at a higher frequency in our LTBI cohort compared with TB* subjects after M. tuberculosis–specific stimulation, but they were absent in unstimulated samples. Based on this readout and similar to what was reported in previous studies, we found the highest reactivity within the Th1* compartment (Fig. 2B). Thus, Th1* is the CD4 T cell subset containing the majority of TB-specific peptide reactivity, and this subset is expanded in the context of LTBI.

Previous work from our group has elucidated the transcriptional profile of Th1*, with 525 genes differentially expressed compared with other Th subsets (34). To investigate whether the gene-expression profile of memory CD4 T cells in LTBI reflected the expansion of Th1*, we compared the 74-gene signature identified in this study with the Th1* gene signature. Both gene signatures presented a significant overlap, with nine genes in common (representation factor 5.1, p value < 0.00007 for overlap, hypergeometric distribution test, Fig. 2C, 2D). Furthermore, the combined expression of the nine Th1* genes correlated positively with the frequency of Th1* within each subject (Fig. 2E). Thus, the transcriptional signature of memory CD4 T cells in LTBI partially reflects the gene-expression profile of TB-specific CD4 T cells, whose frequency is increased in infected subjects compared with uninfected controls.

Validation of the Th1* gene signature in memory CD4 T cells at the protein level

Next, we thought to validate the overlap between Th1* and the transcriptional signature of memory CD4 T cells in LTBI at the protein level. Of the nine overlapping Th1* genes, ABCB1, c-KIT, and GPA33 were the only three candidates with known protein expression on the cell surface and commercially available Abs for protein profiling. All three genes had robustly adjusted p values in the transcriptomic comparison of memory CD4 T cells in LTBI and controls (Supplemental Table III). In particular, GPA33 was the third most dysregulated gene (adjusted p value = 6.7 × 10−5), and it was the only Th1* protein-coding gene differentially expressed across all dataset combinations (Supplemental Table IV). Using

**FIGURE 2.** The transcriptomic signature of LTBI in memory CD4 T cells has a significant overlap with the TB-specific Th1* subset. (A) Memory CD4 T cell subset composition in LTBI subjects compared with uninfected controls, as determined by flow cytometry. (B) TB-specific peptide reactivity (OX40* PDL1+) among memory CD4 T cell subsets, as determined by flow cytometry, after PBMC stimulation for 24 h with MTB300 (22). Plots are from one representative subject (left panel). Line graph shows combined data from three LTBI subjects (right panel). (C) Overlap between the 74-gene signature identified in Fig. 1 and the previously described Th1* gene signature (34), based on the hypergeometric distribution test (considering the 21,992 transcripts detected within memory CD4 T cells as the total number of genes). (D) Heat map displaying rlog-transformed raw counts of the nine genes overlapping between the 74-gene signature and the Th1* gene signature, with genes ordered with hierarchical clustering and subjects ordered based on PC1 component. (E) Correlation between the PC1 component of the combined expression of the nine Th1* overlapping genes, identified in (C), in memory CD4 T cells and the frequency of Th1* in corresponding subjects, as determined by linear regression analysis. Data were derived from 20 LTBI subjects and 20 controls (A), (C), and (E) or from 3 LTBI subjects (B). **p < 0.01, ***p < 0.001, nonparametric Wilcoxon test.
CyTOF, we successfully validated that the differential gene expression of ABCB1, c-KIT, and GPA33 (Fig. 3A, upper panels) in memory CD4 T cells was also apparent at the protein level (Fig. 3A, lower panels). ABCB1 is an ATP-dependent transporter that is expressed on the cell surface; in addition to its protein expression, its activity can be assessed by measuring its capacity to efflux a fluorescent substrate, such as Rhodamine 123. Using a flow cytometry–based assay (23), we found that memory CD4 T cells of LTBI subjects displayed a higher ABCB1 activity compared with controls (Fig. 3B). Finally, to validate that Th1* was the cell subset responsible for the differential protein expression of the three Th1* genes observed in bulk memory CD4 T cells, we assessed the frequency of cells positive for each marker within each memory CD4 T cell subset. For c-KIT and GPA33, Th1* was the only subset responsible for the differential expression observed in LTBI (Fig. 3C). ABCB1 was differentially expressed between the two cohorts in the Th1* and CCR6+DNEG subset (Fig. 3C). Taken together, our results indicate that the three Th1* genes (ABCB1, c-KIT, and GPA33) are differentially expressed at the RNA and protein levels in memory CD4 T cells of LTBI subjects compared with uninfected controls and that Th1* is the main subset responsible for these dysregulations.

**FIGURE 3.** Validation of the presence of the Th1* gene signature in memory CD4 T cells of LTBI subjects at the protein level. (A) ABCB1, c-KIT, and GPA33 expression at the mRNA (upper panels) and protein (lower panels) levels in memory CD4 T cells of LTBI subjects compared with uninfected controls. Gene-expression data were derived from 30 LTBI subjects and 29 uninfected controls using an Illumina sequencing platform. Protein-expression data were derived from 13 LTBI subjects and 15 uninfected controls using CyTOF. (B) ABCB1 activity in memory CD4 T cells of LTBI subjects (n = 12) compared with controls (n = 14), defined as the ratio of Rhodamine 123 mean fluorescence intensity between inhibitor and noninhibitor conditions (lower panel). Representative line graph (upper panel). (C) Frequency of ABCB1+, c-KIT+, and GPA33− cells among each memory CD4 T cell subset in LTBI subjects (n = 13) compared with controls (n = 15), as determined by CyTOF. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001, nonparametric Wilcoxon test.

TB-specific cells within the Th1* subset have the GPA33−CD62L− phenotype

Finally, we investigated whether the remaining 65 genes from the 74-gene signature identified in Fig. 1 (not overlapping with Th1*) could provide novel protein markers to readily identify TB-specific CD4 T cells. The above data show that TB-specific epitope reactivity is contained within Th1* (Fig. 2B), as well as that Th1* of LTBI subjects upregulated ABCB1 and c-KIT and downregulated GPA33 at the RNA and protein levels compared with controls (Fig. 3A, 3C). Of these three Th1* markers, GPA33 was the only one whose protein expression pattern could reliably divide Th1* into positive and negative cell populations (Fig. 4A); thus, it was selected as the best candidate to further delineate TB-specific CD4 T cells at the protein level. Gene expression correlation between GPA33 and the 65 genes not overlapping with Th1* identified CCR7 and CD62L (alias SELL) as the genes with the highest positive correlation with GPA33 (Fig. 4B). CCR7 and CD62L have been previously reported to be downregulated in TB-specific CD4 T cells (38). At the protein level, CCR7 was unchanged and CD62L was decreased in memory CD4 T cells of LTBI subjects compared with controls (Fig. 4C). Boolean gating
identified that the GPA33$^-$CD62L$^-$ Th1$^*$ subset was the only Th1$^*$ subset with increased frequency in LTBI subjects compared with uninfected controls (Fig. 4D). Subsequently, we sorted Th1$^*$ subsets based on GPA33 and CD62L expression and assessed their reactivity to our TB-specific peptide pool (see Supplemental Fig. 1B for gating strategy). The TB reactivity was higher within the GPA33$^-$CD62L$^-$ subset compared with the GPA33$^+$CD62L$^+$ or the GPA33$^+$ subset (Fig. 4E). Thus, we further refined that the Th1$^*$ subset with TB-specific peptide reactivity that was expanded in LTBI subjects has a GPA33$^-$CD62L$^-$ phenotype.

**Discussion**

In this study, we used a purified cell population transcriptomics strategy to identify novel disease signatures and further define the phenotype of Ag-specific CD4 T cells in the context of LTBI. Transcriptomics on memory CD4 T cells of LTBI subjects identified a gene signature that was directly reflecting the expansion of CD4 T cells with TB-specific reactivity within the *M. tuberculosis*-infected cohort compared with uninfected subjects. In concordance with previous findings (33–35), we found the greatest TB reactivity and expansion within the Th1$^*$ subset and a significant overlap between the memory CD4 T cell gene signature of LTBI and the Th1$^*$ gene signature. Subsequently, by combining the transcriptomic data with single-cell protein profiling techniques and in vitro stimulation with Ag-specific peptide pools, we were able to further refine that TB-specific CD4 T cells were contained within the CD62L$^-$GPA33$^-$ subset of Th1$^*$. Thus, these markers are an avenue to directly isolate the subset of CD4 T cells containing the majority of TB-specific reactivity, without the need for TB-specific stimulation.

It is known that the development of active disease significantly reshapes the CD4 memory T cell population by inducing functional and phenotypic changes in TB-specific CD4 T cells (39, 40), which might be responsible for the loss of immune control and disease development. Our transcriptomic profiling has focused on the comparison of LTBI subjects and uninfected controls, with no inclusion of individuals with active TB. Thus, although it is clear that the 74-gene signature reported in this article and all associated findings are related to *M. tuberculosis* exposure, we cannot ascertain whether they are also responsible for long-term protection. Cell population transcriptomic studies of active TB cohorts are currently being investigated within the Department of Vaccine Discovery at La Jolla Institute for Allergy and Immunology and hopefully will answer these questions.

In parallel with Th1$^*$ expansion, LTBI subjects also had a higher frequency of CCR6$^+$ DNEG cells compared with controls; however, in contrast to Th1$^*$, this cell subset was not associated with TB-specific peptide reactivity. The differentiation pathways leading to Th1$^*$ are yet to be defined (41); hence, it is possible that CCR6$^+$ DNEG cells are precursors of the Th1$^*$ subset. Overall, the CCR6$^+$ compartment of memory CD4 T cells was greatly expanded in LTBI subjects compared with uninfected controls. Upregulation of CCR6 was reported previously in memory T cells from active TB subjects compared with uninfected controls (42). CCR6 is the receptor for CCL20, a chemokine that is highly expressed by monocytes upon infection with *M. tuberculosis* (43) and is upregulated in PBMCs from individuals with active TB after Ag-specific stimulation (42). In our dataset, CCL20 was also upregulated in memory CD4 T cells of LTBI subjects. Taken together, these results suggest that CCL20/CCR6 signaling in CD4 T cells might be an important component of TB immune responses.
The overlap between the 74-gene signature in memory CD4 T cells of LTBI subjects and the Th1* gene signature was modest, with only nine genes in common; however, it was highly statistically significant considering that the total number of transcripts expressed within memory CD4 T cells is close to 22,000. The combined expression of these nine genes correlated, to a great extent, with the frequency of Th1* in each individual, suggesting a robust affiliation with Th1*. To further refine the subset of TB-specific CD4 T cells in LTBI, we selected three genes (ABCB1, c-KIT, and GPA33) from the overlap between the 74-gene signature and the Th1* gene signature. We validated that all three genes were also differentially expressed in memory CD4 T cells and Th1* of LTBI subjects at the protein level. These three genes have been previously shown to be differentially expressed in Th1* compared with other CD4 Th cell subsets (34). In this article, we show that these molecules are also dysregulated within Th1* in LTBI, further suggesting an association with TB-specific CD4 T cells. We found no previous association for these three markers with TB from the literature; more generally, their function in the context of CD4 T cell biology remains largely unknown. ABCB1 is an ATP-dependent membrane efflux transporter that is expressed by a broad range of cell types and has a major role in drug resistance, particularly tumor resistance to chemotherapy (44). The receptor tyrosine kinase c-KIT is the ligand for stem cell factor, is highly expressed on hematopoietic cells, and has been associated with an important role in cell differentiation and maturation (45). In the context of immune responses, c-KIT signaling in dendritic cells is critical for providing help for CD4 T cells (46), but no direct role within CD4 T cells has been reported. GPA33 is a glycoprotein member of the Ig superfamily that is expressed on the surface of lymphocytes and is implicated in cell adhesion. CD62L expression in memory T cells is important for homing to secondary lymphoid organs, with high expression in naive and central memory cells, as opposed to recirculating effector memory cells (48). TB-specific CD4 T cells have been shown to express low levels of CD62L (38), but its expression has not been associated with GPA33 or Th1*. Taken together, our results identified a set of novel CD4 T cell markers that is significantly codyregulated in LTBI and associated with TB-specific peptide reactivity: ABCB1, c-KIT, GPA33, and CD62L. They constitute promising biomarkers for TB-specific CD4 T cell subsets, and the study of their interactions and associated signaling pathways could further our understanding of TB immune responses in the context of LTBI.

Acknowledgments
We thank the Flow Cytometry Core Facility and the Bioinformatics Core Facility of La Jolla Institute for technical assistance.

Disclosures
The authors have no financial conflicts of interest.

References

The Journal of Immunology 3289


