


ORIGINAL ARTICLE

Expression of specific HLA class II alleles is associated with an increased risk for active tuberculosis and a distinct gene expression profile

Leila Y. Chihab^{1,2} | Rebecca Kuan¹ | Elizabeth J. Phillips^{3,4} |
Simon A. Mallal^{3,4} | Virginie Rozot⁵ | Mark M. Davis^{6,7} | Thomas J. Scriba⁵ |
Alessandro Sette^{1,8} | Bjoern Peters^{1,8} | Cecilia S. Lindestam Arlehamn¹  | the
SATVI Study Group

¹Center for Infectious Disease and Vaccine Research, La Jolla Institute for Immunology, La Jolla, California, USA

²Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, California, USA

³Institute for Immunology and Infectious Diseases, Murdoch University, Perth, Western Australia, Australia

⁴Vanderbilt University School of Medicine, Nashville, Tennessee, USA

⁵South African Tuberculosis Vaccine Initiative, Institute of Infectious Disease and Molecular Medicine, Division of Immunology, Department of Pathology, University of Cape Town, Cape Town, South Africa

⁶Institute for Immunity, Transplantation and Infection, Stanford University School of Medicine, Stanford, California, USA

⁷Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California, USA

⁸Department of Medicine, University of California San Diego, La Jolla, California, USA

Correspondence

Cecilia S. Lindestam Arlehamn, Center for Infectious Disease and Vaccine Research, La Jolla Institute for Immunology, La Jolla, CA, 92037, USA
Email: cecilia@lji.org

Funding information

National Institutes of Health; S10 OD016262; S10 RR027366; U19 AI118626

Several HLA allelic variants have been associated with protection from or susceptibility to infectious and autoimmune diseases. Here, we examined whether specific HLA alleles would be associated with different *Mycobacterium tuberculosis* (*Mtb*) infection outcomes. The HLA alleles present at the -A, -B, -C, -DPA1, -DPB1, -DQA1, -DQB1, -DRB1, and -DRB3/4/5 loci were determined in a cohort of 636 individuals with known *Mtb* infection outcomes from South Africa and the United States. Among these individuals, 203 were Quantiferon (QFT) negative, and 433 were QFT positive, indicating *Mtb* exposure. Of these, 99 QFT positive participants either had active tuberculosis (TB) upon enrollment or were diagnosed in the past. We found that *DQA1*03:01*, *DPB1*04:02*, and *DRB4*01:01* were significantly more frequent in individuals with active TB (susceptibility alleles), as judged by Odds Ratios and associated p-values, while *DPB1*105:01* was associated with protection from active TB. Peripheral blood mononuclear cells (PMBCs) from a subset of individuals were stimulated with *Mtb* antigens, revealing individuals who express any of the three susceptibility alleles were associated with lower magnitude of responses. Furthermore, we defined a gene signature associated with individuals expressing the susceptibility alleles that was characterized by lower expression of

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. HLA: Immune Response Genetics published by John Wiley & Sons Ltd.

APC-related genes. In summary, we have identified specific HLA alleles associated with susceptibility to active TB and found that the expression of these alleles was associated with a decreased *Mtb*-specific T cell response and a specific gene expression signature. These results will help understand individual risk factors in progressing to active TB.

KEYWORDS

HLA association, *Mycobacterium tuberculosis*, susceptibility, T-lymphocytes, transcriptome

1 | INTRODUCTION

Mycobacterium tuberculosis (*Mtb*), the cause of tuberculosis (TB) disease, infects about one-quarter of the world's population and causes approximately 1.5 million deaths annually.¹ Although *Mtb* infection is fairly prevalent worldwide, less than 10% of infections lead to active tuberculosis (ATB) disease.^{1,2} In some cases individuals with *Mtb* infection can completely clear all living bacteria, while in the most *Mtb* infection cases the bacteria persist and these individuals have an increased risk of bacterial reactivation leading to ATB.³ Latent infection can last weeks, months, or years with the likelihood of progression to ATB being highest within the first 2 years following infection. Certain conditions can increase the likelihood of progression to ATB disease (e.g., HIV co-infection, immune suppression therapy, malnutrition, and diabetes).⁴ To reduce the incidence of TB the WHO has identified a strategy which includes identification and treatment of individuals with ATB, screening of individuals at high risk, as well as preventative therapy to those at high risk of progression to ATB.⁵ Importantly, the biological correlates that distinguish between different *Mtb* infection outcomes have yet to be fully elucidated and thus there is a need for improved prognostic and diagnostic tests to identify those at risk for ATB.

human leukocyte antigens (HLAs) are heterodimeric proteins that play a critical role in the immune recognition of antigens. HLA molecules present pathogen-derived peptides (epitopes) for recognition by T cells.⁶ HLA molecules are grouped into class I versus class II, each encompassing several highly polymorphic loci with more than 34,000 different possible HLA allelic variants.⁷ In general, different HLA allelic variants present distinct epitope repertoires, and this high level of diversity results in a highly diverse set of peptide epitopes being presented in different individuals in a given population. In this context, genetic analyses and genome wide association studies (GWAS) have demonstrated that the presence of specific HLA allelic variants can exert either a protective

or susceptibility effect for different autoimmune or infectious diseases.⁸⁻¹⁴ Several HLA alleles that are associated with ATB has previously been identified in populations from different geographical locations.¹⁵⁻²⁰

Here, we examined whether any specific HLA alleles are associated with ATB and thus could be used to identify individuals at risk for ATB, and whether expression of these HLA alleles can be linked to differences in the *Mtb*-specific immune response.

2 | MATERIALS AND METHODS

2.1 | ETHICS STATEMENT

This research was performed in accordance with approval from the Human Research Ethics Committees of the University of Cape Town (Cape Town, South Africa) and the La Jolla Institute for Immunology (La Jolla, USA). All participants provided written informed consent before participation. Adolescents provided written informed assent, with a parent or legal guardian providing written informed consent.

2.2 | Study Participants

Study participants ($n = 620$) from the Worcester region of South Africa were investigated. Participants were divided into a healthy cohort of QFT positive ($n = 318$) and negative (203) individuals (QuantIFEROn-TB Gold In-Tube, Cellestis) and into individuals with ongoing (sputum XpertMTB/RIF-positive (Cepheid, Sunnyvale, CA)) TB disease ($n = 28$) or a self-reported history of individuals with previously diagnosed active TB who underwent TB treatment ($n = 71$). We also enrolled QFT+ participants ($n = 16$) at the University of California, San Diego Anti-Viral Research Center (San Diego, USA). QFT status and whether individuals have or have had active TB is listed in Supplementary Table 1.

2.3 | PBMC Isolation

PBMC were purified from whole blood either using CPT tubes (BD), for the adolescents, or layered onto Ficoll (for adults) using 50 ml Leukosep tubes (Greiner) by density-gradient centrifugation on Ficoll, according to the manufacturer's instructions. Cells were suspended in FBS containing 10% (vol/vol) DMSO and cryopreserved in liquid nitrogen. The cryopreserved cells from South Africa were shipped to LJI for analysis.

Cryopreserved PBMC were quickly thawed by incubating each cryovial at 37°C for 2 min, and cells transferred to cold medium (RPMI 1640 with L-glutamin and 25 mM HEPES; Omega Scientific), supplemented with 5% human AB serum (GemCell), 1% penicillin streptomycin (Life Technologies), 1% glutamax (Life Technologies) and 20 U/ml benzonase nuclease (MilliporeSigma). Cells were centrifuged and resuspended in medium to determine cell concentration and viability using trypan blue and a hemacytometer.

2.4 | Peptides and other stimuli

Peptides were synthesized as crude material on a small (1 mg) scale by A&A LLC (San Diego). Peptides were pooled into peptide pools, re-lyophilized and reconstituted at a concentration of 0.7 mg/mL for MTB300 (PMID27409590). *Mtb* whole cell lysate (WCL) from strain H37Rv was obtained from BEI Resources, NIAID, NIH (NR-14822).

2.5 | HLA typing

Participants were HLA-typed at the La Jolla Institute, by the American Society for Histocompatibility and Immunogenetics (ASHI)-accredited laboratory at Murdoch University (Western Australia), or Immucor, Sirona Genomics, CA.

Typing at LJI was performed by next-generation sequencing.²¹ Specifically, amplicons were generated from the appropriate class II locus for exons 2 through 4 by PCR amplification. From these amplicons, sequencing libraries were generated (Illumina Nextera XT) and sequenced with MiSeq Reagent Kit v3 as per the manufacturer's instructions (Illumina). Sequence reads were matched to HLA alleles and participant genotypes were assigned.

Typing in Australia for class I (HLA A, B, and C) and class II (DQA1, DQB1, DRB1, DRB3, DRB4, DRB5, and DPB1) was performed using locus-specific PCR amplification of genomic DNA. The HLA testing process at the

Institute for Immunology and Infectious Diseases in Perth Western Australia is accredited by the American Society for Histocompatibility and Immunogenetics (ASHI) and the National Association of Testing Authorities (NATA). Specific HLA Loci were PCR amplified using sample specific MID-tagged primers that amplify polymorphic exons from Class I (A, B, C Exons 2 and 3) and Class II (DQA1 and DQB1; Exons 2 and 3, DRB and DPB1; Exon 2) major histocompatibility complex (MHC) genes relevant to epitope binding and presentation. Therefore, rare alleles that differs by a single nucleotide in exon 1 cannot be excluded due to missing coverage of exon 1. MID tagged primers were optimized to minimize allele dropouts and primer bias. Amplified DNA products from unique MID tagged products (up to 96 MIDs) were quantitated and pooled in equimolar ratios and subjected to library preparation using NEBNext Ultra II library prep kits (New England Biolabs). Libraries were quantified using the Jetseq library quantitation kit (Meridian Bioscience) and High sensitivity D1000 screentape on an Agilent 2200 TapeStation (Agilent) for concentration and size distribution. Normalized libraries were sequenced on the Illumina MiSeq platform using the MiSeq V3 600-cycle kit (2x300bp reads). Sequences were separated by MID tags, reads were quality-filtered and alleles called using an in house-accredited HLA caller software pipeline minimizing the influence of sequencing errors. Alleles were called using the latest IMGT HLA allele database as the allele reference library. Samples were called using the IMGT HLA allele database version 3190 or 3230. The algorithm was developed by E.J.P. and S.A.M. and relies on periodically updated versions of the freely available international immunogenetics information system (<http://www.imgt.org>) and an ASHI-accredited HLA allele caller software pipeline, IID HLA analysis suite (<http://www.iiid.com.au/laboratory-testing/>). Sample report integrity were tracked and checked using proprietary and accredited Laboratory Information and Management System (LIMS) and HLA analyze reporting software that performs comprehensive allele balance and contamination checks on the final dataset.

Typing at Immucor, Sirona Genomics was performed using the MIA FORA NGS kit and analysis software (Immucor, Inc.) as previously described.²²

The HLA type of each subject is listed in Supplementary Table 1.

Population coverage of HLA alleles were calculated using the population coverage tool from the Immune epitope database (<http://www.iedb.org>). To identify allele associations with TB, odds ratios and p-values between active TB and QFT negative/QFT positive individuals were calculated utilizing the RATE algorithm.²³ A type 1 error rate of 0.05 was accounted for via a Bonferroni

correction on alleles expressed in at least 5% of the total study cohort, where the threshold for a significant p -value (0.0006) was calculated as $0.05/87$ (the number of alleles expressed in at least 5% of the study cohort). In addition, individual corrected p -values for each allele was calculated by multiplying the unadjusted p -value with 87.

2.6 | ELISPOT assay

PBMCs were thawed and antigen-specific cellular responses were measured by IFN γ ELISPOT assay with all antibodies and reagents from Mabtech (Nacka Strand, Sweden). Plates (96-well Immobilion-P; Millipore) were coated overnight at 4°C with 5 $\mu\text{g/ml}$ anti-IFN γ (1-D1K; Mabtech). Briefly, 200,000 cells were plated in each well of the pre-coated plate and were stimulated in triplicate with 10 $\mu\text{g/ml}$ *Mtb* whole cell lysate, 1 $\mu\text{g/ml}$ MTB300, 10 $\mu\text{g/ml}$ PHA as a positive control, or DMSO corresponding to the concentration present in the peptide pool. After 20 h incubation at 37°C, wells were washed with PBS/0.05% Tween 20 and incubated with biotinylated anti-IFN γ (7-B6-1; Mabtech) for 2 h. Spots were developed using Vectastain APC peroxidase (Vector Laboratories) and 3-amino-9-ethylcarbazole (Sigma-Aldrich). Spots were counted by computer-assisted image analysis (AID iSpot; Aid Diagnostica). Responses were considered positive if the net spot-forming cells (SFC) per 10^6 PBMC were ≥ 20 , the stimulation index ≥ 2 , and $p \leq 0.05$ (Student's t -test, mean of triplicate values of the response against relevant stimuli vs. the DMSO control).

2.7 | Cell sorting and RNA purification

PBMCs were thawed and added at a density of 1×10^6 cells per well to a round-bottom 96-well plate. They were stimulated for 24 h at 37°C with MTB300, *Mtb* whole cell lysate, anti-CD28 (1 $\mu\text{g/ml}$, CD28.2; eBioscience) together with anti-CD3 (1 $\mu\text{g/ml}$ pre-coated overnight at 4°C, UCHT1; BioLegend), as well as left unstimulated as a negative control. After 24 hrs, cells were resuspended in PBS with 10% (v/v) FBS and incubated at 4°C for 10 min. Cells were then stained with fixable viability dye eFluor506 (eBioscience) and an antibody mixture containing anti-human CD3-AF700 (UCHT1; ThermoFisher), CD4-APCeFluor780 (RPA-T4; eBioscience), CD8-BV650 (RPA-T8; BioLegend), CCR7-PerCPCy5.5 (G043H7, Biolegend) CD14-V500 (M5E2; BD Bioscience), CD19-V500 (HIB19; BD Bioscience), CD25-FITC (M-A251; BD Bioscience), CD45RA-eFluor450 (HI100, eBiosciences), CD137-APC (4B4-1; BioLegend), OX40-PECy7 (Ber-ACT35; BioLegend), and PD-L1-PE

(29 E.2A3; BioLegend) for 20 min at room temperature. Cells were washed and resuspended in PBS before being transferred into a 5 ml polypropylene FACS tube (BD Bioscience). PBMCs were sorted, based on forward and side scatter excluding debris and doublets, on a FACSaria into TRIzol LS (Thermo Fisher). Acquisition files were analyzed using FlowJo. Total RNA was extracted from $\sim 100,000$ cells in TRIzol LS using the miRNeasy Micro Kit (Qiagen) on a QIACube (Qiagen). Total RNA was amplified according to Smart Seq protocol.² cDNA was purified using AMPure XP beads. cDNA was used to prepare a standard barcoded sequencing library (Illumina). Samples were sequenced using an Illumina HiSeq2500 to obtain 50-bp single end reads. Samples that failed to be sequenced due to limited sample availability or failed the quality control were eliminated from further sequencing and analysis. The full protocol can be found at protocols.io (<https://doi.org/10.17504/protocols.io.bxr6pm9e>).

2.8 | RNA-seq analysis

Paired-end reads that passed Illumina filters were filtered for reads aligning to tRNA, rRNA, adapter sequences, and spike-in controls. The reads were aligned to the GRCh38 reference genome and Gencode v27 annotations using STAR (v2.6.1).²⁴ DUST scores were calculated with PRINSEQ Lite v0.20.3²⁵ and low-complexity reads (DUST >4) were removed from BAM files. The alignment results were parsed via SAMtools²⁶ to generate SAM files. Read counts to each genomic feature were obtained with the featureCounts (v 1.6.5)²⁷ using the default option along with a minimum quality cut off (Phred >10). After removing absent features (zero counts in all samples), the raw counts were imported into R v3.6.1 and genes with an average TPM < 1 were removed. R/Bioconductor package DESeq2 v.1.24.0²⁸ was used to normalize raw counts. Variance stabilizing transformation was applied to normalized counts to obtain log₂ gene expression values. Quality control was performed using boxplots and Principal component analysis (PCA), using the 'prcomp' function in R, on log₂ expression values. Differentially expressed genes were identified using the DESeq2 Wald test, and p -values were adjusted for multiple test correction using the Benjamini Hochberg algorithm.²⁹ Genes with adjusted p values <0.05 and log₂ fold change >1 or < -1 were considered differentially expressed. Cell-type enrichment was performed using DICE.³⁰ The RNAseq data have been submitted to the Gene Expression Omnibus under accession number GSE207214 (<http://www.ncbi.nlm.nih.gov/geo/>).

TABLE 1 HLA allele association with active TB

HLA allele ^a	HLA allele frequency (%)	No. participants with allele		No. participants lacking allele		Odds ratio	<i>P</i> -value ^b	Corrected <i>p</i> -value	Population coverage (%)
		Active TB	QFT+ & QFT-	Active TB	QFT+ & QFT-				
<i>DQA1*03:01</i>	7.1	24	21	75	516	7.9	8.0 E-10	6.9 E-08	26.4
<i>DPB1*04:02</i>	10.3	25	40	74	495	4.2	2.0 E-06	1.7 E-04	19.2
<i>DRB4*01:01</i>	27.0	42	127	57	401	2.3	0.0003	0.03	47.4 ^c
<i>DPB1*105:01</i>	25.6	12	150	87	385	0.4	0.0006	0.05	0.1 ^d
<i>A*23:17</i>	6.8	0	43	99	494	0.0	0.0008	>0.05	0.0 ^e
<i>DQA1*01:05</i>	5.3	0	34	99	503	0.0	0.006	>0.05	2.3
<i>DRB1*11:01</i>	12.9	21	61	78	474	2.1	0.01	>0.05	10.5
<i>DQA1*03:03</i>	21.7	13	125	86	412	0.5	0.02	>0.05	6.9
<i>B*15:03</i>	9.9	16	47	83	489	2.0	0.03	>0.05	1.3
<i>DQB1*02:01</i>	15.4	8	90	90	447	0.4	0.03	>0.05	23.0
<i>DRB1*03:01</i>	17.5	10	101	89	434	0.5	0.04	>0.05	17.8
<i>DRB1*07:01</i>	17.8	25	88	74	447	1.7	0.04	>0.05	18.2

^aWhen no result was obtained for a particular locus those participants were excluded from the calculations, the total number of participants for each allele does therefore not add up to 636, or ATB ($n = 99$), and QFT ± ($n = 537$, where QFT + ($n = 334$), and QFT - ($n = 203$)).

^bBonferroni corrected threshold for a significant *p*-value = 0.0006.

^cCoverage based in linkage with DR4, 7, and 9 alleles.

^dRare worldwide, but 25.4% in South African Black and 20% in Brazilian Black.

^eRare worldwide, but 7.7% in South African Black and 8% in Brazilian Black.

3 | RESULTS

3.1 | *HLA DQA1*03:01, DRB4*01:01, and DPB1*04:02* are associated with susceptibility for active TB and *DPB1*105:01* with protection

To determine whether specific HLA alleles are associated with active TB, samples were obtained from a cohort of 636 participants, of which 203 individuals were QFT negative, the remaining 433 individuals were QFT positive including 99 individuals with a history of active TB (ATB) or who were diagnosed with active TB upon enrollment. Participants were HLA typed by determining the specific alleles expressed at the HLA-A, -B, -C, -DPA1, -DPB1, -DQA1, -DQB1, -DRB1, and -DRB3/4/5 loci. The results of the HLA typing were utilized to investigate whether the expression of particular HLA alleles would be associated with *Mtb* infection outcomes. To identify allele associations with TB, odds ratios and *p*-values between active TB and QFT negative/QFT positive individuals were calculated utilizing the RATE algorithm.²³ A type 1 error rate of 0.05 was accounted for via a Bonferroni correction on alleles expressed in at least 5% of the total study cohort (Table 1).

Of the 324 HLA alleles expressed in one or more individuals of our cohort, we focused on the 87 alleles (37 class I and 50 class II alleles) that were present in at least 5% (i.e. 32 individuals) of the cohort to ensure we had sufficient power to detect differences in the HLA distribution. For these 87 alleles, 12 had a significant *p*-value when comparing their frequency in individuals with ATB compared with QFT negative and positive individuals (Table 1). Most of these variants are commonly expressed in the general population (Table 1). Of these variants, HLA *DQA1*03:01* had the strongest association with ATB, with an odds ratio of 7.9 and *p*-value of 8.0 E-10. A total of 45 participants carried this allele type, 24 of which (53%) had ATB. In addition, the *DPB1*04:02* and *DRB4*01:01* alleles were also associated with ATB, with odds ratios of 4.2 and 2.3 respectively. Finally, HLA *DPB1*105:01* appeared to confer protection against ATB, since it was inversely correlated with ATB (OR = 0.4, $p = 0.0006$). After Bonferroni correction, both the positive association between alleles *DQA1*03:01*, *DPB1*04:02*, *DRB4*01:01* and ATB ($p = 6.9$ E-8, 1.7 E-4, 0.03), and the negative association between HLA *DPB1*105:01* and ATB ($p = 0.05$) were maintained.

TABLE 2 HLA allele association with *Mtb* infection

HLA allele ^a	HLA allele frequency (%)	No. participants with allele		No. participants lacking allele		Odds ratio	P-value ^b	Corrected p-value
		Active TB & QFT+	QFT-	Active TB & QFT+	QFT-			
<i>DQA1*03:01</i>	7.1	43	2	390	201	11.1	1.00 E-05	0.001
<i>DQA1*04:01</i>	13.1	42	41	391	162	0.4	3.80 E-04	0.033
<i>DRB1*03:02</i>	10.9	34	35	398	167	0.4	0.0006	0.050
<i>DQB1*04:02</i>	14.8	50	44	382	159	0.5	0.001	>0.05
<i>DQA1*01:02</i>	45.9	217	75	216	128	1.7	0.002	>0.05
<i>C*17:01</i>	14.2	49	41	384	161	0.5	0.003	>0.05
<i>A*74:01</i>	5.8	33	4	400	199	4.1	0.003	>0.05
<i>B*42:01</i>	8.8	28	28	404	175	0.4	0.004	>0.05
<i>DRB1*15:03</i>	18.0	90	24	342	178	2.0	0.006	>0.05
<i>DQB1*06:02</i>	37.3	177	60	255	143	1.7	0.006	>0.05
<i>DPA1*02:02</i>	37.1	135	61	191	142	1.6	0.010	>0.05
<i>DRB4*01:01</i>	27.0	128	41	298	160	1.7	0.012	>0.05
<i>A*30:01</i>	12.1	43	34	390	169	0.5	0.018	>0.05
<i>DQA1*01:03</i>	20.6	78	53	355	150	0.6	0.021	>0.05
<i>B*15:03</i>	9.9	51	12	381	191	2.1	0.022	>0.05
<i>DRB3*01:01</i>	28.7	110	70	316	131	0.7	0.023	>0.05
<i>DPB1*04:01</i>	24.9	96	62	335	141	0.7	0.030	>0.05
<i>DRB1*07:01</i>	17.8	67	46	365	156	0.6	0.034	>0.05
<i>DRB1*11:01</i>	12.9	64	18	368	184	1.8	0.042	>0.05

^aWhen no result was obtained for a particular locus those participants were excluded from the calculations, the total number of participants for each allele does therefore not add up to 636, or ATB ($n = 99$), and QFT \pm ($n = 537$, where QFT + ($n = 334$), and QFT - ($n = 203$)).

^bBonferroni corrected threshold for a significant p -value = 0.0006.

3.2 | *DQA1*04:01* and *DRB1*03:02* are associated with protection from *Mtb* infection

To protect from active disease, a factor like HLA could either protect from infection, or from progression to active TB. To investigate that we examined the alleles that were associated with susceptibility to active disease, and tested if these were enriched within the QFT positive. We found no alleles that were more frequent in QFT positive individuals compared with QFT negative (not shown).

We next investigated the differences in HLA expression between *Mtb* infected (both QFT positive and ATB) participants and QFT negative participants, using the same approach as above (Table 2). We identified 19 alleles present in at least 5% of the participants with a significant p -value. In this comparison, *DQA1*03:01* was again identified to have the closest association to *Mtb* infection with an odds ratio of 11.1 and a corrected p -value of 0.001. In addition, *DQA1*04:01* and *DRB1*03:02* were associated

with protection from *Mtb* infection, both with odds ratios of 0.4 ($p = 0.033$ and 0.050, respectively). In conclusion, we identified several alleles associated with ATB and *Mtb* infection susceptibility. We focused the rest of the study on the alleles associated with susceptibility for ATB.

3.3 | *DQB1*02:02*, **03:03*, **04:02*, **06:01/02/09* co-expression with *DQA1*03:01* increases susceptibility to ATB

Mature HLA DQ molecules expressed on the cell surface are a dimer of DQ alpha and beta chains, and particular DQA alleles are preferentially co-expressed with particular DQB allelic variants. For this reason, we next sought to determine which DQ beta chain variants are most likely to be co-expressed with HLA *DQA1*03:01* in ATB cases. To do so, we calculated the odds ratios and p -values of HLA DQ alpha and beta pairs to identify which pair is more closely associated to ATB (Table 3). Using the most prevalent DQ beta chains in our cohort,

TABLE 3 DQB1 alleles expressed with *DQA1*03:01*

DQA1	DQB1	No. ATB w allele	No. ATB w/o allele	No. QFT + QFT-w allele	No. QFT + QFT-w/o allele	OR	p-value
<i>DQA1*03:01</i>	-	24	75	21	516	7.9	8.0 E-10
<i>DQA1*03:01</i>	<i>DQB1*02:02</i>	9	89	1	536	54.2	3.1 E-07
-	<i>DQB1*02:02</i>	23	75	85	452	1.6	0.08
<i>DQA1*03:01</i>	<i>DQB1*04:02</i>	6	92	1	536	35.0	7.2 E-05
-	<i>DQB1*04:02</i>	11	87	83	454	0.7	0.35
<i>DQA1*03:01</i>	<i>DQB1*06:02</i>	8	90	5	532	9.5	1.6 E-04
-	<i>DQB1*06:02</i>	44	54	193	344	1.5	0.11
<i>DQA1*03:01</i>	<i>DQB1*03:03</i>	3	95	2	535	8.4	0.03
-	<i>DQB1*03:03</i>	5	93	23	514	1.2	0.79
<i>DQA1*03:01</i>	<i>DQB1*06:01</i>	2	96	0	537	inf	0.02
-	<i>DQB1*06:01</i>	8	90	32	505	1.4	0.37
<i>DQA1*03:01</i>	<i>DQB1*06:09</i>	1	97	0	537	inf	0.15
-	<i>DQB1*06:09</i>	6	92	47	490	0.7	0.55
<i>DQA1*03:01</i>	DQB1 Combined	21	77	9	528	16.0	0.26

we paired various beta chains with *DQA1*03:01* and determined the odds ratios with a confidence interval of 95%. When paired with beta chains *DQB1*02:02*, **04:02*, **03:03*, and **06:01/02*, the odds ratios significantly increased from 7.9 to 54.2 (*DQB1*02:02*), 35.0 (*DQB1*04:02*), 8.4 (*DQB1*03:03*), 9.5 (*DQB1*06:01*), and infinity (*DQB1*06:02*; e.g., no expression in QFT positive or negative individuals), respectively. The odds ratio of association when *DQA1*03:01* is paired to all possible DQB1 chains combined is 16, an increase from the original 7.9. Thus, when *DQA1*03:01* is co-expressed with the beta chains highlighted in Table 2, the association of *DQA1*03:01* with ATB becomes even more pronounced.

3.4 | Expression of susceptibility alleles is associated with lower magnitude of *Mtb*-specific response in an IFN γ ELISPOT assay

We next investigated if the expression of the HLA alleles associated with ATB (Table 1) was associated with differences in the *Mtb*-specific immune response. PBMC samples from a subset of our cohort (all of which were QFT positive and 59 of which were ATB cases) were tested for IFN γ production in response to an *Mtb*-derived peptide pool (MTB300). For all three of the susceptibility alleles (*DQA1*03:01*, *DPB1*04:02*, or *DRB4*01:01*) we found a trend toward a lower magnitude of IFN γ -specific responses, with the strongest trend being for *DQA1*03:01*

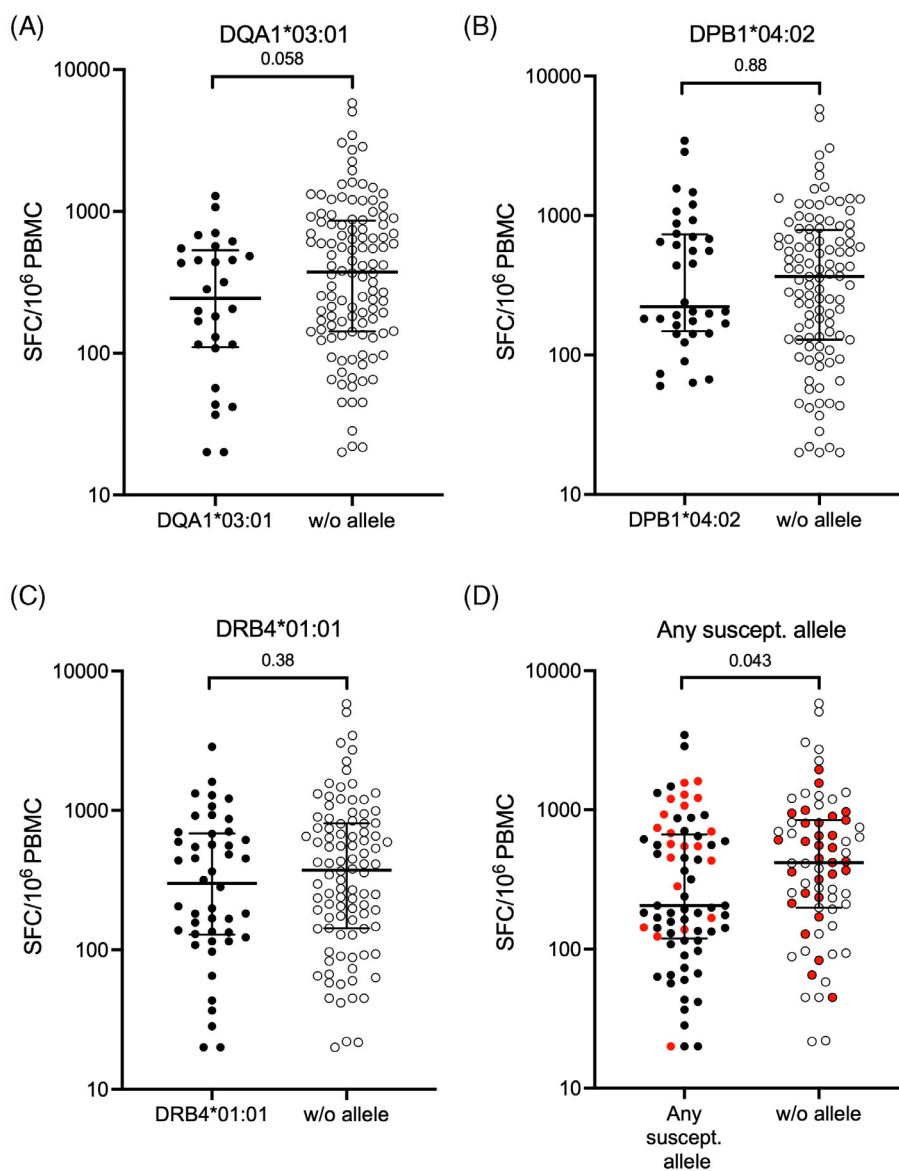
(Figure 1A–C). When the three alleles were considered together, we found a significantly lower magnitude of IFN γ -specific response in individuals with the susceptibility alleles compared with those without (Figure 1d). No significant difference was detected when comparing the responses from the ATB cohort with the QFT positive cohort (not shown). In conclusion, the expression of the three susceptibility alleles influences the *Mtb*-specific IFN γ response.

3.5 | The presence of ATB susceptibility alleles is correlated with specific gene signatures

To further investigate the biological mechanisms of HLA association with ATB susceptibility, PBMCs from a subset of QFT positive healthy individuals (Figure 1D, red color) were stimulated with MTB300 or a whole cell lysate (WCL) of the H37Rv TB strain. Live, singlet cells were sorted and their global gene expression was determined by RNA-Seq. The differential gene expression analysis of individuals who express the susceptibility alleles versus individuals who lack the expression of these alleles in each stimulation condition from the ELISPOT assay (unstimulated, MTB300, and WC), revealed a specific gene signature.

A total of 84 differentially expressed genes (DEGs) were identified across the three different stimulation conditions (Supplementary Table 2; adjusted *p* values <0.05

FIGURE 1 The expression of susceptibility alleles is associated with a lower magnitude of *Mtb*-specific IFN γ response. Magnitude of response against MTB300 as SFC per 10⁶ PBMC measured in an IFN γ ELISPOT assay. Each point represents one individual (total $n = 144$), median \pm interquartile range is shown. Two-tailed Mann–Whitney U test. Individuals with susceptibility alleles are represented by closed circles and individuals without by open circles. (A) *DQA1*03:01* $n = 28$ with and $n = 116$ without, (B) *DPB1*04:02* $n = 36$ with, $n = 108$ without, (C) *DRB4*01:01* $n = 46$ with, $n = 98$ without, (D) any susceptibility allele $n = 73$ with, $n = 71$ without. Red color indicates participants included in the subsequent gene expression experiments.



and log₂ fold change >1 or < -1). For the unstimulated samples, 5 genes were upregulated in individuals with the susceptibility alleles, and 29 genes were upregulated in those without. For the MTB300 stimulation condition, 8 genes were upregulated in with and 10 w/o, and for WCL it was 21 versus 19 genes (Figure 2A). The 84 genes were filtered based on concordant expression across the different stimuli, for example, if a gene was upregulated in individuals with the susceptibility allele in one condition but up in individuals without in another it was removed. This resulted in a gene signature of 63 genes (Figure 2B and Supplementary Table 2). These genes could be grouped into 4 clusters based on their expression in individuals with versus without the susceptibility alleles and across stimuli conditions (Figure 2B). Cluster 1 encompasses genes that were upregulated in the unstimulated and MTB300 condition, but with no difference

between allele expression. The genes corresponding to Cluster 2 were upregulated in individuals with expression of the susceptibility alleles across the three different conditions compared to individuals lacking the alleles. Cluster 3 consisted of genes upregulated in individuals without the susceptibility alleles across different stimulations. Interestingly, cluster 3 included HLA-DQA1 and HLA-DQB1, as well as JUN and TIRAP. Cluster 4 was composed of genes with different patterns of upregulation in response to WCL, but no difference between allele expression.

The 63 genes associated with the signature were enriched in genes expressed in B cells and monocytes (both classical and non-classical; Figure 2C). Finally, to visualize the gene expression changes between cohorts in each stimulus, 6 representative genes were chosen, 2 from cluster 2 (XPOT and CAB39) and 4 from cluster 3

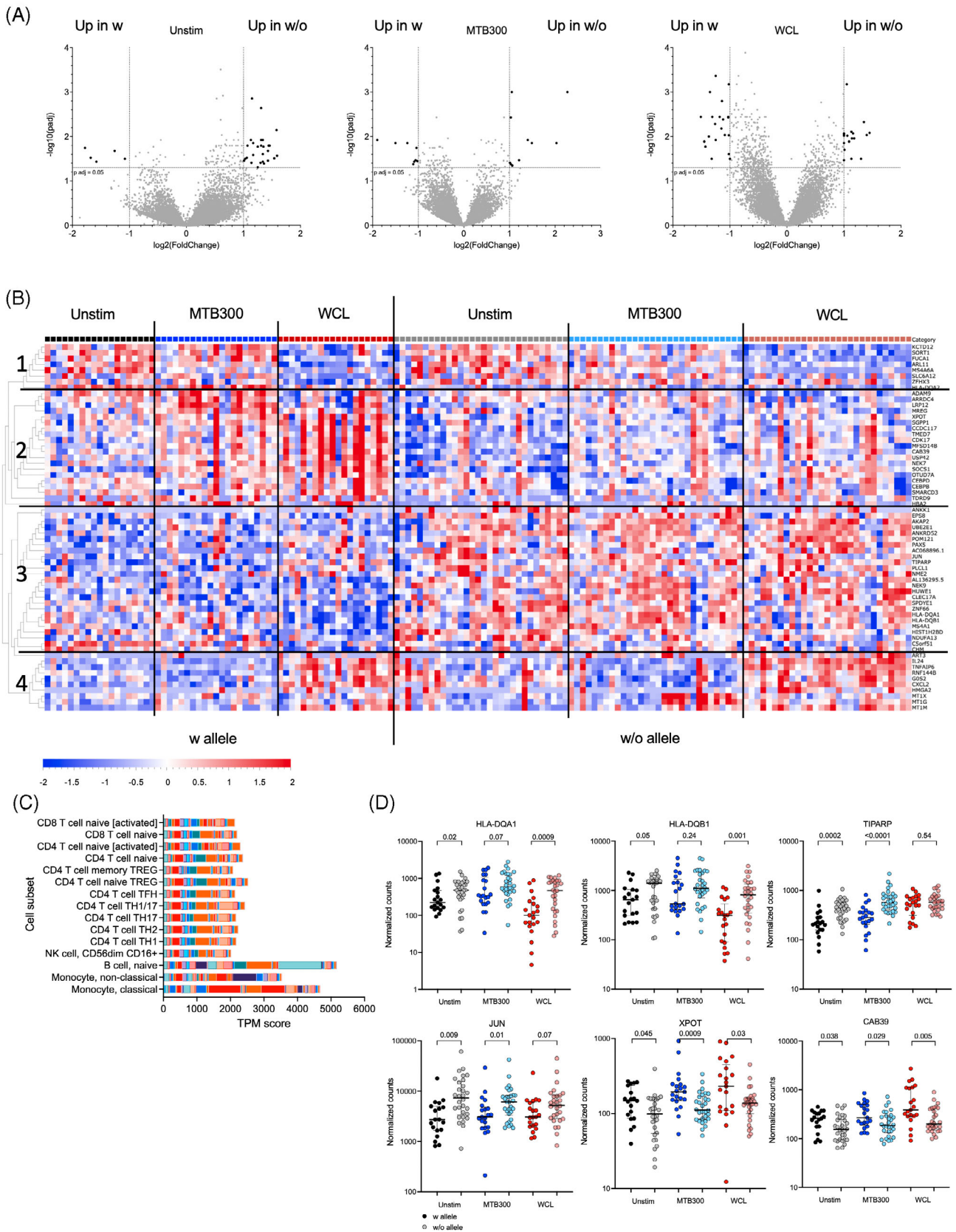


FIGURE 2 Legend on next page.

(JUN, TIPARP, HLA-DQA1, and HLA-DQB1; Figure 2D). The 4 genes from cluster 3 exhibited a trend for lower expression in individuals with the susceptibility alleles in each of the stimuli conditions. Conversely, the 2 genes from cluster 2 had higher expression in each stimuli condition in individuals with the susceptibility alleles.

3.6 | Gene signatures associated with ATB susceptibility alleles and low T cell responsiveness to *Mtb* antigens

Finally, we investigated whether DEGs could be identified in PBMCs from participants who had different magnitudes of an *Mtb*-specific IFN γ response. Specifically, we compared the unstimulated samples from individuals expressing any of the three susceptibility alleles and who had a magnitude of response below the median for individuals included in the RNAseq (<568 SFCs) with individuals lacking those alleles and who had a magnitude of response above the median. Unstimulated samples were used for this analysis to avoid detecting genes differentially expressed solely due to the *Mtb*-specific magnitude, that is, potentially higher levels of T cell activation genes, etc. This analysis revealed 21 DEGs which were all upregulated in individuals without the susceptibility alleles (Figure 3A). This gene signature could distinguish between these two cohorts (with alleles; low expression and without alleles; high expression, Figure 3B, C). Furthermore, 5 of these 21 DEGs overlapped with the 63-gene signature described above (CXCL2, JUN, TNFAIP6, ZNF66, and AC068896.1).

Taken together, individuals expressing any of the susceptibility alleles have a specific gene signature that is enriched in genes related to antigen-presenting cells (APCs), possibly related to the lower magnitude of *Mtb*-specific IFN γ responses detected.

4 | DISCUSSION

In this study, we examined the potential association between the expression of specific HLA alleles and ATB

disease. The identification of risk factors for ATB will help identify individuals at high risk of progression and could inform preventative treatment strategies. Using samples from 636 participants, we identified 3 HLA alleles associated with susceptibility for ATB. The expression of these alleles was associated with a lower magnitude of *Mtb*-specific T cell responses and also correlated with a lower expression of APC-related genes.

We identified that *DQA1*03:01*, *DPB1*04:02*, and *DRB4*01:01* were associated with susceptibility to ATB, and in contrast, *DPB1*105:01* was associated with protection against ATB. Interestingly, *DPB1*04:02* and *DPB1*105:01* have an identical exon 2 which is encoding for the antigen recognition site in the MHC molecule and differ only by a single nucleotide in exon 3.⁷ Single-nucleotide variants within HLA have been associated with various diseases in a number of studies.^{31–34} However, future studies in other populations should investigate whether the observation in this study is a coincidence of the high frequency of expression of *DPB1*105:01* in South Africa or whether this translates to other aspects of the immunity against *Mtb*.

HLA alleles are highly polymorphic in general, and the genetic heterogeneity in HLA alleles also varies across different ethnic groups and geographical populations. However, the alleles identified as associated with differential disease outcomes in the present study are relatively common in the most ethnic groups. While we identified an association of ATB with *DQA1*03:01* in our study cohort from South Africa and the United States, this allele was also found in association with pulmonary tuberculosis in a cohort from Iceland and Russia.¹⁹ We further identified the DQB1 alleles most likely expressed together with *DQA1*03:01*. These DQB1 alleles were *DQB1*02:02*, **04:02*, **03:03*, and **06:01/02*. A meta-analysis of 12 studies found that HLA *DQB1*06:01* was significantly associated with an increased risk of ATB as well.¹⁶

Sveinbjornsson et al. reported the association of the HLA class II allele *DQA1*03* with *Mtb* infection and identified a noncoding variant, rs557011, located between the HLA-DQA1 and HLA-DRB1 alleles that also associated with *Mtb* infection.¹⁹ Many other groups have reported

FIGURE 2 Expression of the susceptibility alleles is associated with a specific gene signature. (A) Volcano plots showing differentially expressed genes comparing individuals with susceptibility alleles ($n = 21$) to those without ($n = 30$) in three different stimuli conditions; unstimulated, MTB300, and WCL. Black indicates DEGs (adjusted p value <0.05 and log₂ fold change >1 and < -1). (B) Heatmap displaying the 63 DEGs from (A). Genes are ordered by hierarchical clustering, and samples are ordered according to allele expression and then according to stimulation condition. Black vertical lines indicate allele and stimuli separation and horizontal lines indicate cluster separation. (C) Cell-type enrichment (<http://dice-database.org>) for the 63 DEGs. Each color represents one gene. (D) Gene expression at the mRNA level. Gene expression values in counts normalized by sequencing depth calculated by the DESeq2 package. Two-tailed Mann–Whitney U test.

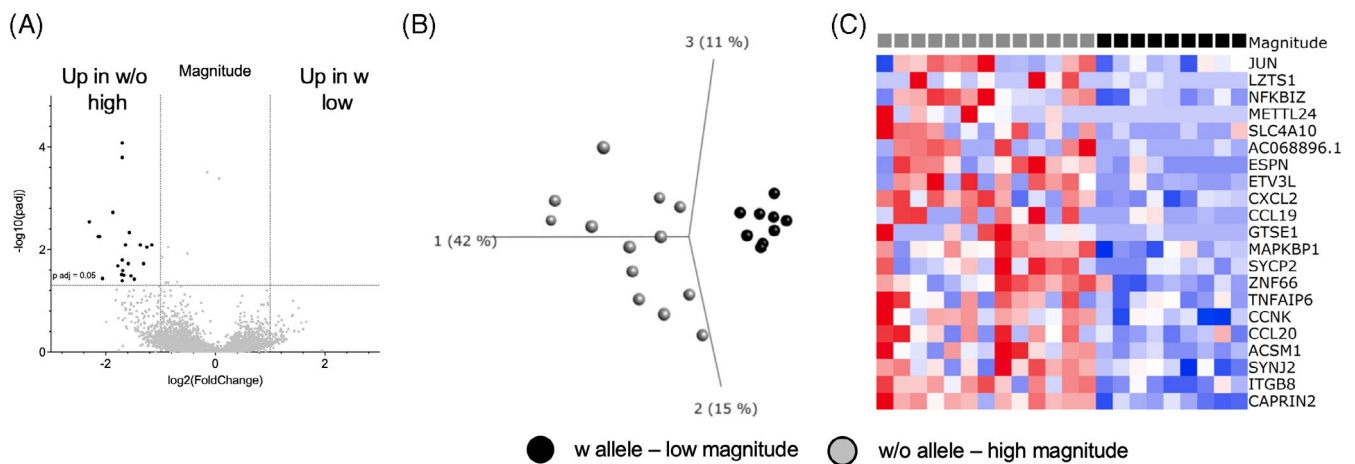


FIGURE 3 Differences in *Mtb*-specific magnitude of response is associated with a specific gene signature. (A) Volcano plot showing differentially expressed genes comparing individuals with susceptibility alleles and low magnitude of expression ($n = 9$) to those without and high magnitude of expression ($n = 13$) in unstimulated samples. Black indicates DEGs (adjusted p -value < 0.05 and \log_2 fold change > 1 and < -1). (B) PCA plot illustrating differences between the two cohorts for the 21 DEGs. (C) Heatmap displaying the 20 DEGs from (A). Genes are ordered by hierarchical clustering, and samples are ordered according to the two cohorts.

similar findings for different class I and class II HLA alleles.^{15,17,18,20} Because of the polymorphic nature and heterogeneity of HLA allele expression in diverse geographical locations, further studies in additional cohorts across different ethnicities may identify or confirm additional HLA alleles associated with susceptibility to ATB. In addition, HLA alleles present a strong linkage disequilibrium in which preferential combinations of alleles are inherited together in the genome more often than expected. The series of alleles at linked loci on a single chromosome is named “haplotype”.³⁵ Specific haplotypes of association with ATB may be identified through more detailed HLA typing in which these linked loci are taken into consideration. In our study we did not determine these linkages.

We further linked expression of the susceptibility alleles to changes in the *Mtb*-specific immune response. Individuals expressing *DQA1*03:01*, *DPB1*04:02*, and *DRB4*01:01* had a lower *Mtb*-specific immune response compared to those lacking these alleles. For the individual alleles the most pronounced effect was observed when comparing *Mtb*-specific immune responses in those expressing *DQA1*03:01* compared to others, which is in agreement with this allele having the strongest association with ATB. However, one limitation is that the *Mtb*-specific immune responses were measured by IFN γ ELISPOT which is not an absolute quantitative test of the immune response. Future studies will investigate additional immune responses in more detail. Other groups have reported impaired immune responses due to the expression of specific HLA alleles in patients with human immunodeficiency virus (HIV) and hepatitis c virus

(HCV) infection.^{36–39} Huang et al. identified an *HLA-B*35* variant that was associated with accelerated HIV infection in a population of patients of European descent. They described that individuals who expressed the *HLA-B*35* susceptibility allele had significantly impaired dendritic cell (DC) function than individuals lacking the susceptibility allele and hypothesized that this impaired DC function was the mechanism by which the susceptibility allele contributed to HIV disease.³⁷ Additionally, HLA class I and class II variants have been found to confer protection against severe HCV infection and these protective alleles have been reported to better present immunodominant HCV epitopes.^{36,38,39} Finally, HLA-DQ variants have been reported to present streptococcal and staphylococcal superantigens at different strengths and influence the overall T cell-mediated response to bacterial superantigens.⁴⁰ Overall, HLA allele variants have been linked previously to altered immune responses in the context of different diseases.

In addition, using transcriptomics we defined gene signatures that could distinguish individuals with versus without expression of the susceptibility alleles. The majority of DEGs were identified in response to WCL, reflecting the complexity of this stimuli compared to unstimulated and MTB300 stimulated samples. The DEGs could be divided into four clusters that provided further insight into the role of the HLA susceptibility alleles and *Mtb* infection. Two clusters corresponded to genes that were differentially expressed in response to the specific stimuli unstim and MTB300 versus WCL stimulation. Cluster 2 was comprised of genes that were upregulated in individuals who express the susceptibility

alleles. These genes included the transcription factors CEBPD and CEBPB, as well as the chromatin remodeling factor SMARCD3, which are highly expressed on monocytes (<http://dice-database.org> and www.proteinatlas.org). CEBPD and CEBPB are involved in the regulation of inflammatory genes such as TNF and IL-12 in monocytes, and these transcription factors have been shown to both upregulate and downregulate inflammatory genes.⁴¹ Similarly, SMARCD3 has been shown to regulate the expression of cytokines such as IL-6.⁴² The upregulation of these genes in those individuals who express the susceptibility alleles suggest that they are involved in downregulating the expression of inflammatory cytokines needed for an effective immune response against *Mtb*. Additional experiments are needed to understand this in detail. Finally, cluster 3 was comprised of genes that were upregulated in individuals who are not carriers of the susceptibility alleles. This cluster included both HLA-DQA1 and HLA-DQB1, as well as PAX5, JUN and CLEC17A which are all highly expressed on B cells (<http://dice-database.org>). HLA-DQA1 and HLA-DQB1 proteins are generally upregulated in response to stress, such as high concentration of inflammatory cytokines,^{43,44} highlighting the potential effect of susceptibility alleles with lower DQA1/DQB1 expression under stress. Moreover, JUN is involved in toll-like receptor activity and antigen-cross presentation,⁴⁵ and downregulation (as in individuals expressing the susceptibility alleles) leads to impairment of those functions. This cluster also contained TIRAP, which encodes for a molecule that is associated with toll-like receptors and also plays an important role in innate immunity.⁴⁶ Thus, individuals expressing the susceptibility alleles have downregulation of genes related to antigen-presentation, which may contribute to an increased risk of ATB.

Lastly, we identified a gene signature related to the magnitude of *Mtb*-specific T cell responses. This signature was upregulated in individuals lacking the susceptibility alleles and 5 genes overlapped with the signature detected when comparing DEGs irrespective of the *Mtb*-specific T cell response detected. This signature included JUN, as mentioned above, as well as, CCL19 which is involved in the homing of lymphocytes to the lung in *Mtb* infection.⁴⁷ The signature also included CXCL2 which is a neutrophil-attracting chemokine.⁴⁸ Taken together, these gene signature results suggest that the susceptibility alleles potentially increase the risk for ATB by impairing antigen loading and presentation to immune cells for proper clearance of *Mtb* infection.

RNAseq is a sensitive method to quantify the gene expression in a sample which requires relatively few cells as input. It can give insight into relative protein levels and can be used to infer cellular function. Proteomics methods, on the other hand, can provide exact protein

levels and a more concrete understanding of the relationship between protein abundance and function, albeit requiring greater numbers of cells. Several studies have shown poor correlation between mRNA and protein levels.^{49–51} This can be explained by the well-known fact that the relative abundance of different proteins in a cell is not solely the result of differential mRNA expression levels, but also due to differences in translation, post-transcriptional regulation, and degradation.⁵² Proteomics can serve as a useful approach in future studies to gain further insight into the functional impact of the gene signatures highlighted in our study.

In conclusion, our study identified three HLA alleles (*HLA DQA1*03:01*, *DPB1*04:02*, and *DRB4*01:01*) associated with susceptibility to ATB. Further investigation revealed that individuals expressing these alleles have a decreased *Mtb*-specific immune response. The identified HLA alleles and associated gene signatures can help identify individuals who are at risk of ATB.

AUTHOR CONTRIBUTIONS

Thomas J. Scriba, Alessandro Sette, Bjoern Peters, and Cecilia S. Lindestam Arlehamn participated in the design and direction of the study. Leila Y. Chihab, Rebecca Kuan, and Cecilia S. Lindestam Arlehamn performed literature search and participated in data collection, analysis and interpretation. Virginie Rozot and Thomas J. Scriba and the SATVI-Ja Jolla Study Group recruited participants. Elizabeth J. Phillips, Simon A. Mallal, and Mark M. Davis coordinated and performed HLA typing. Leila Y. Chihab and Cecilia S. Lindestam Arlehamn wrote the first draft. All authors reviewed, edited and approved the final version of the manuscript before submission.

ACKNOWLEDGMENTS

We thank the sequencing core and flow cytometry core at La Jolla Institute for Immunology for their help in these studies. This study was supported by the National Institutes of Health grants U19 AI118626, S10 RR027366, S10 OD016262. The SATVI study group: Thomas Scriba (PI), Michelle Fisher (Project Manager), Humphrey Mulenga (Data Manager), Ashley Veldsman (Regulatory Manager), Nicole Bilek (Laboratory Manager), Angeliqe Luabeya (Site PI), Jateel Kassim (Site PI [Previous]), Marwou de Kock (Site Manager), Cynthia Ontong (Study Co-ordinator), Elizabeth Filander (Research Nurse), Gail Jacobs (CRW), Nambitha Nqakala (CRW), Yulandi Herselman (Data Administrator), Simba Mabwe (Site Lab Manager), Denis Arendsen (Lab Technologist [study supervisor]), Palesa Dolo (Lab Technologist), Hadn Africa (Lab Technologist [Expert knowledge regarding processing with blood bags]), Marcia Steyn (Lab Technologist), and Lebohang Makhethhe (Lab Technologist).


CONFLICT OF INTEREST

The authors have no conflict of interest to report.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

ORCID

Cecilia S. Lindestam Arlehamn  <https://orcid.org/0000-0001-7302-8002>

REFERENCES

- WHO. (2020). Global Tuberculosis Report 2020.
- Petruccioli E, Scriba TJ, Petrone L, et al. Correlates of tuberculosis risk: predictive biomarkers for progression to active tuberculosis. *Eur Respir J*. 2016;48:1751-1763. doi:10.1183/13993003.01012-2016
- Barry CE 3rd, Boshoff HI, Dartois V, et al. The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nat Rev Micro*. 2009;7:845-855.
- Pai M, Behr MA, Dowdy D, et al. Tuberculosis. *Nature Reviews Disease Primers*. 2016;2:16076. doi:10.1038/nrdp.2016.76
- WHO. The end TB strategy. (2017).
- Cruz-Tapias P, Castiblanco J, JM A. In: Anaya JM, Shoenfeld Y, Rojas-Villarraga A, eds. *Autoimmunity: From Bench to Bedside*. El Rosario University Press; 2013.
- Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, Marsh SGE. IPD-IMGT/HLA Database. *Nucleic Acids Res*. 2020;48:D948-D955. doi:10.1093/nar/gkz950
- Blackwell JM, Jamieson SE, Burgner D. *HLA and Infectious Diseases*. Clinical Microbiology Reviews; Vol 22; 2009:370-385, Table of Contents. doi:10.1128/CMR.00048-08
- Dendrou CA, Petersen J, Rossjohn J, Fugger L. HLA variation and disease. *Nat Rev Immunol*. 2018;18:325-339. doi:10.1038/nri.2017.143
- Sanchez-Mazas A. A review of HLA allele and SNP associations with highly prevalent infectious diseases in human populations. *Swiss Med Wkly*. 2020;150:w20214. doi:10.4414/smw.2020.20214
- Wang L, Zou ZQ, Wang K. Clinical relevance of HLA gene variants in HBV infection. *J Immunol Res*. 2016:9069375. doi:10.1155/2016/9069375
- Gough SC, Simmonds MJ. The HLA region and autoimmune Disease: associations and mechanisms of action. *Curr Genomics*. 2007;8:453-465. doi:10.2174/138920207783591690
- Eriksson D, Royrvik EC, Aranda-Guillen M, et al. GWAS for autoimmune Addison's disease identifies multiple risk loci and highlights AIRE in disease susceptibility. *Nat Commun*. 2021; 12:959. doi:10.1038/s41467-021-21015-8
- Tait BD, Colman PG, Morahan G, et al. HLA genes associated with autoimmunity and progression to disease in type 1 diabetes. *Tissue Antigens*. 2003;61:146-153. doi:10.1034/j.1399-0039.2003.00013.x
- Amirzargar AA, Yalda A, Hajabolbaghi M, et al. The association of HLA-DRB, DQA1, DQB1 alleles and haplotype frequency in Iranian patients with pulmonary tuberculosis. *Int J Tuberc Lung Dis*. 2004;8:1017-1021.
- Li C, Zhou Y, Xiang X, Zhou Y, He M. The relationship of HLA-DQ alleles with tuberculosis risk: a meta-analysis. *Lung*. 2015;193:521-530. doi:10.1007/s00408-015-9747-1
- Salie M, van der Merwe L, Moller M, et al. Associations between human leukocyte antigen class I variants and the *Mycobacterium tuberculosis* subtypes causing disease. *J Infect Dis*. 2014;209:216-223. doi:10.1093/infdis/jit443
- Souza de Lima D, Morishi Ogusku M, Porto Dos Santos M, et al. Alleles of HLA-DRB1*04 associated with pulmonary tuberculosis in Amazon Brazilian population. *PLoS One*. 2016; 11(2):e0147543. doi:10.1371/journal.pone.0147543
- Sveinbjornsson G, Gudbjartsson DF, Halldorsson BV, et al. HLA class II sequence variants influence tuberculosis risk in populations of European ancestry. *Nat Genet*. 2016;48:318-322. doi:10.1038/ng.3498
- Wamala D, Buteme HK, Kirimunda S, Kallenius G, Joloba M. Association between human leukocyte antigen class II and pulmonary tuberculosis due to *Mycobacterium tuberculosis* in Uganda. *BMC Infect Dis*. 2016;16:23. doi:10.1186/s12879-016-1346-0
- McKinney DM, Fu Z, Le L, Greenbaum JA, Peters B, Sette A. Development and validation of a sample sparing strategy for HLA typing utilizing next generation sequencing. *Hum Immunol*. 2015;76:917-922. doi:10.1016/j.humimm.2015.04.007
- Thorstenson YR, Creary LE, Huang H, et al. Allelic resolution NGS HLA typing of class I and class II loci and haplotypes in Cape Town, South Africa. *Hum Immunol*. 2018;79:839-847. doi:10.1016/j.humimm.2018.09.004
- Paul S, Dillon MB, Lindestam Arlehamn CS, et al. A population response analysis approach to assign class II HLA-epitope restrictions. *J Immunol*. 2015;194:6164-6176. doi:10.4049/jimmunol.1403074
- Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15-21. doi:10.1093/bioinformatics/bts635
- Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*. 2011;27:863-864. doi:10.1093/bioinformatics/btr026
- Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25: 2078-2079. doi:10.1093/bioinformatics/btp352
- Liao Y, G S, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30:923-930. doi:10.1093/bioinformatics/btt656
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550. doi:10.1186/s13059-014-0550-8
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Methodol*. 1995;57:289-300.
- Schmiedel BJ, Singh D, Madrigal A, et al. P. Impact of Genetic polymorphisms on human immune cell gene expression. *Cell*. 2018;175:1701-1715, 1716. doi:10.1016/j.cell.2018.10.022
- Chen C, Song N, Dong Q, et al. Association of Single-Nucleotide Variants in the human leukocyte antigen and other loci with childhood Hodgkin lymphoma. *JAMA Netw Open*. 2022;5:e2225647. doi:10.1001/jamanetworkopen.2022.25647
- Hill-Burns EM, Factor SA, Zabetian CP, Thomson G, Payami H. Evidence for more than one Parkinson's disease-

- associated variant within the HLA region. *PLoS One*. 2011; 6(11):e27109. doi:10.1371/journal.pone.0027109
33. Shimizu C, Kim J, Eleftherohorinou H, et al. HLA-C variants associated with amino acid substitutions in the peptide binding groove influence susceptibility to Kawasaki disease. *Hum Immunol*. 2019;80:731-738. doi:10.1016/j.humimm.2019.04.020
34. Chen DP, Chang SW, Jaing TH, Wang WT, Hus FP, Tseng CP. Single nucleotide polymorphisms within HLA region are associated with disease relapse for patients with unrelated cord blood transplantation. *PeerJ*. 2018;6:e5228. doi:10.7717/peerj.5228
35. Shiina T, Hosomichi K, Inoko H, Kulski JK. The HLA genomic loci map: expression, interaction, diversity and disease. *J Hum Genet*. 2009;54:15-39. doi:10.1038/jhg.2008.5
36. Harcourt G, Hellier S, Bunce M, et al. Effect of HLA class II genotype on T helper lymphocyte responses and viral control in hepatitis C virus infection. *J Viral Hepat*. 2001;8:174-179. doi:10.1046/j.1365-2893.2001.00289.x
37. Huang J, Goedert JJ, Sundberg EJ, et al. HLA-B*35-px-mediated acceleration of HIV-1 infection by increased inhibitory immunoregulatory impulses. *J Exp Med*. 2009;206:2959-2966. doi:10.1084/jem.20091386
38. Neumann-Haefelin C, Spangenberg HC, Blum HE, Thimme R. Host and viral factors contributing to CD8+ T cell failure in hepatitis C virus infection. *World J Gastroenterol*. 2007;13:4839-4847. doi:10.3748/wjg.v13.i36.4839
39. Wertheimer AM, Miner C, Lewinsohn DM, Sasaki AW, Kaufman E, Rosen HR. Novel CD4+ and CD8+ T-cell determinants within the NS3 protein in subjects with spontaneously resolved HCV infection. *Hepatology*. 2003;37:577-589. doi:10.1053/jhep.2003.50115
40. Llewelyn M, Srisakandan S, Peakman M, et al. HLA class II polymorphisms determine responses to bacterial superantigens. *J Immunol*. 2004;172:1719-1726. doi:10.4049/jimmunol.172.3.1719
41. Huber R, Pietsch D, Panterodt T, Brand K. Regulation of C/EBPbeta and resulting functions in cells of the monocytic lineage. *Cell Signal*. 2012;24:1287-1296. doi:10.1016/j.cellsig.2012.02.007
42. Tartey S, Matsushita K, Vandenbon A, et al. Akirin2 is critical for inducing inflammatory genes by bridging I kappaB-zeta and the SWI/SNF complex. *EMBO J*. 2014;33:2332-2348. doi:10.15252/embj.201488447
43. Kagnoff MF. Celiac disease: pathogenesis of a model immunogenetic disease. *J Clin Invest*. 2007;117:41-49. doi:10.1172/JCI30253
44. Carey BS, Poulton KV, Poles A. Factors affecting HLA expression: a review. *Int J Immunogenet*. 2019;46:307-320. doi:10.1111/iji.12443
45. Novoszel P, Drobits B, Holcman M, et al. The AP-1 transcription factors c-Jun and JunB are essential for CD8alpha conventional dendritic cell identity. *Cell Death Differ*. 2021;28:2404-2420. doi:10.1038/s41418-021-00765-4
46. Belhaouane I, Hoffmann E, Chamailard M, Brodin P, Machelart A. Paradoxical roles of the MAL/Tirap adaptor in pathologies. *Front Immunol*. 2020;11:569127. doi:10.3389/fimmu.2020.569127
47. Khader SA, Rangel-Moreno J, Fountain JJ, et al. In a murine tuberculosis model, the absence of homeostatic chemokines delays granuloma formation and protective immunity. *J Immunol*. 2009;183:8004-8014. doi:10.4049/jimmunol.0901937
48. Dorhoi A, Iannaccone M, Maertzdorf J, Nouailles G, Weiner J 3rd, Kaufmann SH. Reverse translation in tuberculosis: neutrophils provide clues for understanding development of active disease. *Front Immunol*. 2014;5:36. doi:10.3389/fimmu.2014.00036
49. Liu Y, Beyer A, Aebersold R. On the dependency of cellular protein levels on mRNA abundance. *Cell*. 2016;165:535-550. doi:10.1016/j.cell.2016.03.014
50. Takemon Y, Chick JM, Gerdes Gyuricza I, et al. Proteomic and transcriptomic profiling reveal different aspects of aging in the kidney. *Elife*. 2021;10:e62585. doi:10.7554/eLife.62585
51. Koussounadis A, Langdon SP, Um IH, Harrison DJ, Smith VA. Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. *Sci Rep*. 2015;5:10775. doi:10.1038/srep10775
52. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet*. 2012;13:227-232. doi:10.1038/nrg3185

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Chihab LY, Kuan R, Phillips EJ, et al. Expression of specific HLA class II alleles is associated with an increased risk for active tuberculosis and a distinct gene expression profile. *HLA*. 2023;101(2):124-137. doi:10.1111/tan.14880