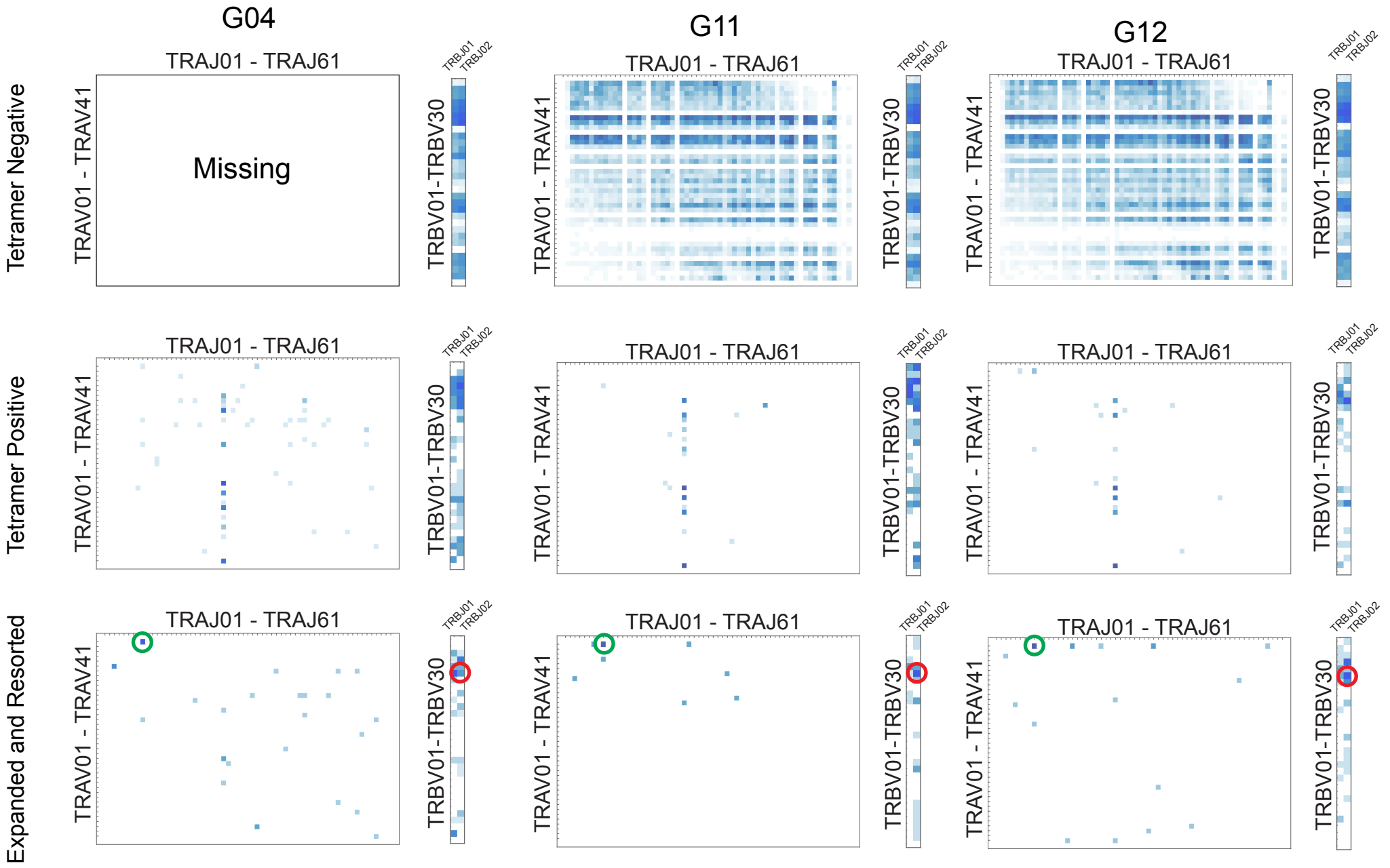
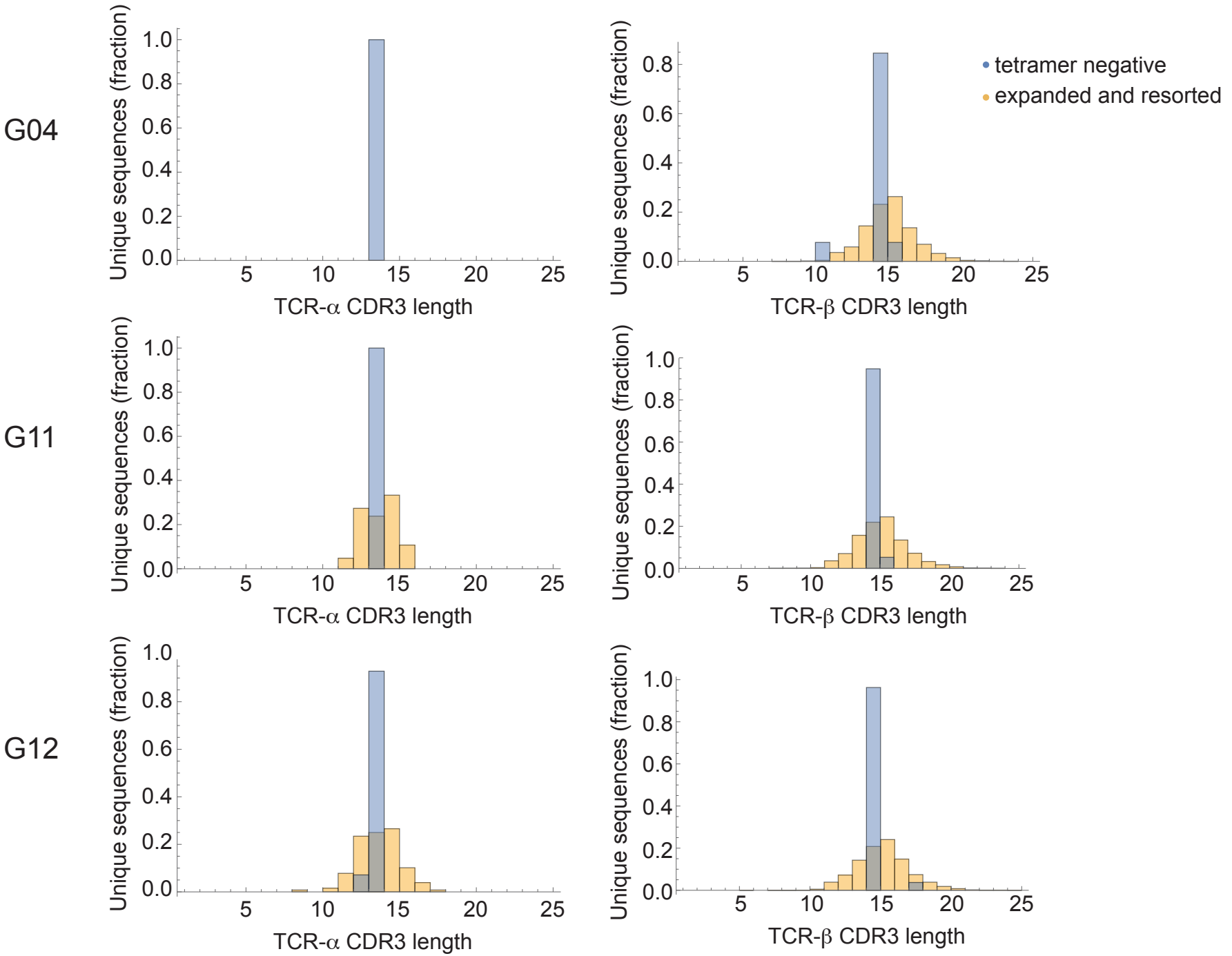
Dewitt et al. Supplementary Figure 1



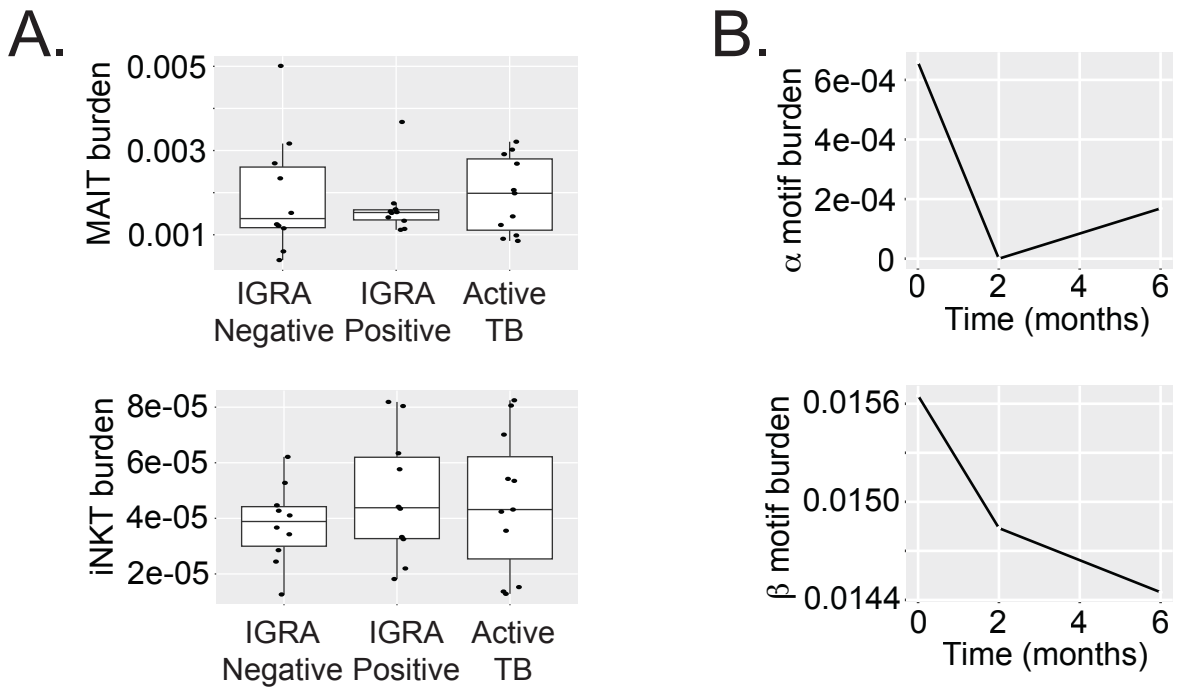
Supplementary Figure 1. GMM-specific T-cell receptor diversity. (A) Heatmaps showing the diversity of TRAV and TRAJ recombination events found among tetramer-negative, tetramer-positive, or expanded and resorted T cells obtained from three additional donors (G04, G11, and G12) with latent tuberculosis infection. Color scale indicates the relative abundance of each V/J combination. The green circle represents the conserved TRAV1-2 and TRAJ9 rearrangement, and the red circle represents the conserved TRBV6 and TRBJ2 rearrangement. The data for the remaining donor is shown in Figure 3 in the main text.

Dewitt et al. Supplemental Figure 2



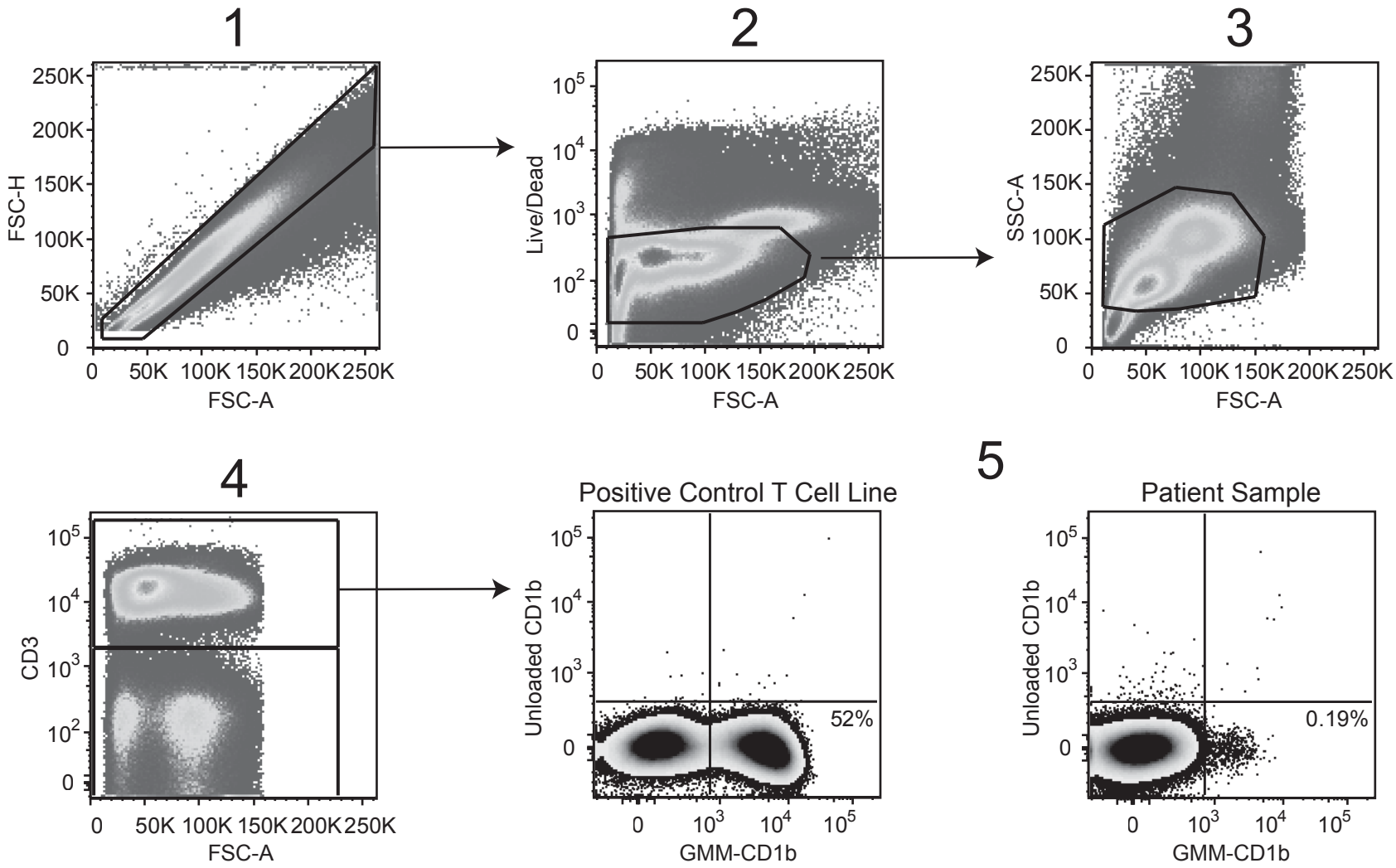
Supplementary Figure 2. CDR3 length distribution among tetramer-negative or expanded and resorted T cells. (A) Unique TCR-α sequences that included a TRAV1 and TRAJ9 rearrangement or (B) unique TCR-β sequences that included a TRBV-6 and TRBJ2 rearrangement in three subjects with latent tuberculosis. Data for the remaining donor is shown in Figure 3C.

Dewitt et al. Supplementary Figure 3



Supplementary Figure 3. TCR-α motif burdens during tuberculosis infection, disease, and treatment. (A) Burden of canonical MAIT or iNKT-specific TCR-α chain rearrangements in IGRA-negative, IGRA-positive, or active TB patients. MAIT seqeuences were identified based on usage of TRAV1-2 and TRAJ33 gene segments and a fixed CDR3 length. For iNKT, we restricted sequences based on usage of TRAV10 and TRAJ18 with 0 or 1 codon deletions and no insertions (Dellabona et al. *J. Exp Med* 1994). There was no difference in either MAIT or iNKT burden when comparing active TB patients to IGRA-negative subjects (p=0.81 and p=0.39, respectively by ANOVA). (B) GMM-specific TCR-α and TCR-β motif burden in PBMCs sampled from one patient with active tuberculosis at diagnosis (0 months), during treatment (2 months), and at the end of therapy (6 months).

Dewitt et al. Supplementary Figure 4



Supplementary Figure 4. Ex vivo quantification of T cells that bind GMM-CD1B tetramer. GMM-loaded CD1B tetramers were used to identify GMM-specific T cells from cryopreserved. Gating strategy proceeds from (1) single cell events to (2) viable cells (AViD Live/Dead) to (3)lymphocytes by size gating, and finally to (4) T cells. The frequency of (5) GMM-specific cells were identified within the CD3+ gate by co-staining with GMM-loaded and unloaded CD1B tetramers as a negative control. Shown here is a representative patient sample and a positive control T cell line.

ID	TCR-α motifs			incidence (fraction of unique TCRs)		p-value
	V	J	CDR3 length	tetramer-negative	expanded/resorted	
G10	TCRAV01	TCRAJ09	13	3/24613	17/76	3.29E-41
	TCRAV08	TCRAJ20	11	4/24613	8/76	2.70E-18
	TCRAV10	TCRAJ26	3	0/24613	2/76	9.35E-06
	TCRAV17	TCRAJ09	13	7/24613	2/76	3.32E-04
	TCRAV25	TCRAJ26	1	2/24613	6/76	1.94E-14
	TCRAV25	TCRAJ26	2	1/24613	3/76	1.12E-07
	TCRAV38	TCRAJ54	2	0/24613	2/76	9.35E-06
	TCRAV41	TCRAJ26	1	1/24613	3/76	1.12E-07
G11	TCRAV01	TCRAJ27	2	0/55066	1/20	3.63E-04
	TCRAV01	TCRAJ09	13	20/55066	7/20	2.25E-19
	TCRAV07	TCRAJ35	3	0/55066	1/20	3.63E-04
	TCRDV03	TCRAJ58	3	0/55066	1/20	3.63E-04
G12	TCRAV01	TCRAJ34	11	25/94785	2/31	3.61E-05
	TCRAV01	TCRAJ09	13	32/94785	13/31	1.86E-36
	TCRAV01	TCRAJ17	14	16/94785	2/31	1.58E-05
	TCRAV03	TCRAJ03	12	2/94785	1/31	9.81E-04
	TCRAV12	TCRAJ26	4	0/94785	1/31	3.27E-04
	TCRAV30	TCRAJ35	15	1/94785	1/31	6.54E-04
	TCRAV41	TCRAJ16	3	0/94785	1/31	3.27E-04

ID	TCR-β motifs			incidence (fraction of unique TCRs)		p-value
	V	J	CDR3 length	tetramer negative	resorted/expanded	
G04	TCRBV04	TCRBJ02	14	697/96689	6/134	4.69E-04
	TCRBV04	TCRBJ02	19	112/96689	9/134	1.40E-13
	TCRBV06	TCRBJ01	14	1413/96689	43/134	1.48E-44
	TCRBV06	TCRBJ02	14	1780/96689	11/134	4.34E-05
	TCRBV30	TCRBJ01	18	13/96689	17/134	1.03E-41
G10	TCRBV01	TCRBJ02	16	18/130339	2/83	7.55E-05
	TCRBV04	TCRBJ02	16	721/130339	13/83	1.89E-15
	TCRBV04	TCRBJ02	20	51/130339	3/83	6.02E-06
	TCRBV06	TCRBJ02	14	1878/130339	46/83	1.06E-61
G11	TCRBV06	TCRBJ02	14	1549/96810	36/67	2.46E-46
	TCRBV26	TCRBJ02	18	0/96810	1/67	6.92E-04
G12	TCRBV04	TCRBJ02	19	115/146215	19/90	6.73E-40
	TCRBV04	TCRBJ02	20	68/146215	2/90	8.79E-04
	TCRBV06	TCRBJ02	14	2406/146215	26/90	5.02E-25
	TCRBV07	TCRBJ02	19	241/146215	6/90	1.20E-08
	TCRBV17	TCRBJ02	17	0/146215	1/90	6.15E-04

Supplementary Table 1. Summary of our motif discovery framework. At the subject level all putative motifs (combinations of V gene family, J gene family and CDR3 length) significant at p-value > 1e-3 are shown. The single motif enriched that was found to be enriched in all subjects (highlighted in red) was chosen for further analysis Note: because of a missing sample, we used three subjects for TCRα, whereas we use all four for TCRβ.

Clone	Variable	CDR3	Joining
A12	TRAV17*01	CANAAIGGSRLTF	TRAJ58*01
	TRBV4-1*01	CASSQEGLRDRGQRGDGYTF	TRBJ1-2*01
C6	TRAV17*01	CANAAIGGSRLTF	TRAJ58*01
	TRBV4-1*01	CASSQEGLRDRGQRGDGYTF	TRBJ1-2*01
G2	TRAV1-2*01	CAVRNTGGFKTIF	TRAJ9*01
	TRBV6-2*01	CASTHRVGGETQYF	TRBJ2-5*01
G7	TRAV12-2*02	CAVKGFGNEKLTF	TRAJ48*01
	TRBV4-1*01	CASSHMGLAGSYSYNEQFF	TRBJ2-1*01
B10	TRAV38-1*01	CAFMKLGTYNQGGKLIF	TRAJ23*01
	TRBV12-4*01	CASSFPRRGVHSEAFF	TRBJ1-1*01
C8	TRAV38-1*01	CAFMKLGTYNQGGKLIF	TRAJ23*01
	TRBV12-4*01	CASSFPRRGVHSEAFF	TRBJ1-1*01
H6	TRAV1-2*01	CAVRNTGGFKTIF	TRAJ9*01
	TRBV6-2*01	CASTHRVGGETQYF	TRBJ2-5*01

Supplementary Table 2. Analysis of GMM-specific T-cell clones. GMM-specific T cells were isolated from a fifth South African donor with latent tuberculosis and plated at limiting dilution to facilitate the isolation of T cell clones. After 28 days of in-vitro expansion with irradiated feeder cells, anti-CD3, and IL-2, we extracted RNA and transcribed RNA to cDNA. We used template switched PCR with gene specific primers targeting the constant region of TCR-α or TCR-β to determine the full length TCR sequence (Jing L et al. J Immunology 2016). These sequences were queried against the IMGT online database determine TCR segment usage and CDR3 sequence (Yousfi Monod et al. Bioinformatics 2004). Shown are the TCR sequences for seven clones with those matching our defined motifs highlighted in yellow.

CDR3 sequences of TCRs matching TCRα motif			
tetramer-negative		expanded/resorted	
CAVIDTGGFKTIF	CAVRGTGGFKTIF	CAARNTGGFKTIF	CAVRGPGRFKTIF
CAVPRGGGFKTIF	CAVRAPGGFKTIF	CAVRKTGGFKTIF	CAVRGTGGFKTIF
CAVRGRGGFKTIF	CAVPNTGGFKTIF	CAVRSLGGFKTIF	CAVRGTGGFKTIF
CAVRDTGGFKTIF	CAVRSTGGFKTIF	CAVRSLGGFKTIF	CAVRGTGGFKTIF
CAVKHTGGFKTIF	CVLMGAGGFKTIF	CAVRKTGGLKTIF	CAVRGTGGFKTIF
CAVQNTGGFKTIF	CAVRSPTGFKTIF	CAVRKTGGLKTIF	CAVRGTGGFKTIF
CAASESGGFKTIF	CAVRDTGGFKTIF	CAARNTGGLKTIF	CAVRGTGGFKTIF
CAVRGTGGFKTIF	CAVTGAGGFKTIF	CAARNTGGFKTIF	CAVRLTGGFKTIF
CAASESGGFKTIF	CAVRTTGGFKTIF	CAVRKPGRFKTIF	CAVRNTGGFKTIF
CAVIGTGGFKTIF	CAVGSTGGFKTIF	CAVRLTGGFKTIF	CAVRGTGGFKTIF
CAVKETGGFKTIF	CAVRGTGGFKTIF	CAVRNTGGFKTIF	CAVRGTGGFKTIF
CAVLNTGGFKTIF	CAVKGTGGFKTIF	CAVRNTGGFKTIF	CAVRGTGGFKTIF
CAVPHTGGFKTIF	CAVSYTGGFKTIF	CAVRLTGGLKTIF	CAVRGTGGFKTFF
CAVWAPGGFKTIF	CAVAGAGGFKTIF	CAVRLTGGFKTIF	CAVRGTGGFKTIF
CAPRNTGGFKTIF	CAVRGTGGFKTIF	CAERLTGGFKTIF	CAVRGTGGFKTIF
CAVAYTGGFKTIF	CAVRNTGGFKTIF	CAVRLTGGFKTIF	CAVRGTGGFKTIF
CAVIITGGFKTIF	CAVRSPGGFKTIF	CAVRLTGGFKTIF	CAVRGTGGFKTIF
CAAKGTGGFKTIF	CAVSNTGGFKTIF	CAVRNTGGFKTIF	CAARGTGGFKTIF
CAAINTGGFKTIF	CALGNTGGFKTIF	CAVRNIGGFKTIF	
CAVINTGGFKTIF	CAVRHTGGFKTIF	CAVRLTGGFKTIF	
CAVRNTGGFKTIF	CAVASTGGFKTIF	CAVRLTGGFKTIF	
CAVSHTGGFKTIF	CAVGNTGGFKTIF	CAVRNTGGFKTIF	
CAVRDTGGFKTIF	CAVRGGGGFKTIF	CAVRLTGGFKAIL	
CAVRDTGGFKTIF	CARGGTGGFKTIF	CAVRLTGGFTTIF	
CAVGDTGGFKTIF	CAVSYTGGFKTIF	CAVRLTGGFKTIY	
CAVAGTGGFKTIF	CAVLNTGGFKTIF	CAARNTGGFKTIF	
CAPRHTGGFKTIF	CAPINTGGFKTIF	CAVRGTGGFKTIF	
CAVLNIGGFKTIF		CAVRGTGGFKTIF	

Supplementary Table 3. Enrichment of residues within the GMM-specific TCR-α motif. This table summarizes unique CDR3 sequences that contain a TRAV1-2 and TRAJ09 rearrangement identified within tetramer-negative or expanded/resorted T cells from one donor with latent tuberculosis. Highlighted in red and blue are the arginine at position 4 and leucine at position 5, respectively.

	CDR3 position	Amino Acid	Fraction of unique TCRs with this amino acid at this position		p-value	enriched(+) suppressed(-)
			tetramer-negative	expanded/resorted		
TCR-α	4	R	20/55	46/46	3.90E-13	+
	5	L	0/55	12/46	3.30E-05	+
TCR-β	1	C	7613/7613	115/119	5.30E-08	-
	1	S	0/7613	4/119	5.30E-08	+
	3	S	7321/7613	85/119	8.70E-20	-
	3	T	158/7613	33/119	2.40E-26	+
	4	Q	51/7613	35/119	1.30E-42	+
	4	S	5195/7613	26/119	7.90E-25	-
	4	R	1025/7613	0/119	6.60E-08	-
	4	T	442/7613	55/119	1.20E-34	+
	5	E	783/7613	0/119	4.40E-06	-
	5	G	592/7613	0/119	1.50E-04	-
	5	L	525/7613	0/119	3.50E-04	-
	5	P	716/7613	103/119	5.20E-85	+
	5	S	560/7613	0/119	2.30E-04	-
	5	Y	1562/7613	8/119	7.20E-05	-
	6	G	1913/7613	0/119	3.20E-15	-
	6	R	772/7613	106/119	3.90E-87	+
	6	T	683/7613	0/119	2.60E-05	-
	7	A	570/7613	0/119	2.40E-04	-
	7	G	2621/7613	0/119	4.40E-22	-
	7	L	433/7613	64/119	1.40E-45	+
	7	S	607/7613	0/119	9.60E-05	-
	7	R	831/7613	35/119	4.10E-08	+
	7	T	666/7613	0/119	4.10E-05	-
	7	V	243/7613	14/119	3.70E-05	+
	8	G	2518/7613	98/119	6.30E-28	+
	8	S	867/7613	1/119	2.30E-05	-
	8	R	549/7613	0/119	0.00023	-
	8	W	54/7613	7/119	3.80E-05	+
	9	A	734/7613	0/119	1.10E-05	-
	9	G	943/7613	111/119	1.30E-87	+
	9	T	1600/7613	0/119	1.20E-12	-
	9	Y	517/7613	0/119	0.00056	-
	10	E	825/7613	27/119	0.00028	+
	10	D	1354/7613	84/119	1.10E-35	+
	10	G	1243/7613	7/119	0.00099	-
	10	N	1323/7613	0/119	2.70E-10	-
	10	Y	2216/7613	0/119	4.20E-18	-
	11	E	5109/7613	34/119	1.90E-17	-
	11	T	2058/7613	77/119	3.50E-17	+
	13	F	2286/7613	9/119	3.60E-09	-
	13	Y	5172/7613	110/119	2.50E-10	+

Supplementary Table 4. Position-wise enrichment of TCR motifs. For each position within the CDR3 of GMM-specific TCR-α and TCR-β motifs, we tabulated the fraction of unique sequences with a particular amino acid at that position among tetramer-negative and expanded/resorted T cells. Statistical significance was computed using a Fisher's Exact test. The table summarizes those residues that were either enriched or suppressed among expanded and resorted T cells as compared to tetramer-negative T cells. These results are summarized graphically in Figure 4C and 4D of the main text.

Adolescent With and Without Latent TB					
PTID	Age	Sex	Ethnicity	IGRA	TST
09-0157	13	male	Coloured	Negative	0
09-0292	15	female	Coloured	Negative	0
11-0083	16	female	White	Negative	0
02-0249	14	male	Coloured	Negative	0
02-0319	13	male	Coloured	Negative	0
02-0320	13	male	Coloured	Negative	0
03-0324	12	male	Black	Negative	0
03-0703	17	female	Black	Negative	0
01-0345	14	male	Coloured	Negative	0
01-0381	14	female	Coloured	Negative	0
01-0457	16	male	Coloured	Positive	14.9
03-0539	12	male	Black	Positive	16.5
03-0558	13	male	Black	Positive	13
01-0667	15	female	Coloured	Positive	15.5
09-0092	13	male	Black	Positive	17
09-0306	15	male	Coloured	Positive	19
01-0872	13	male	Coloured	Positive	12
03-0709	14	female	Black	Positive	16
01-0959	18	female	Coloured	Positive	15
01-0272	14	male	Coloured	Positive	12.4

Adult With Active TB					
PTID	Age	Sex	Ethnicity	Sputum smear	TB Treatment (days)
TB-1052	32	F	Coloured	2+	3
TB-1100	36	F	Coloured	3+	6
TB-1103	31	M	Coloured	3+	3
TB-1104	30	M	Coloured	1+	5
TB-1107	37	F	Coloured	2+	6
TB-1108	21	F	Coloured	2+	6
TB-1112	31	M	Coloured	2+	6
TB-1117	31	F	Coloured	3+	6
TB-1119	19	M	Coloured	1+	6
TB-1124	41	M	Coloured	1+	5
TB-1126	20	M	Coloured	2+	3

Adults with Latent TB					
PTID	Age	Sex	Ethnicity	IGRA	Cell line
DN020-11	27	F	Coloured	Positive	G04
PS 22-2002	24	F	Coloured	Unknown	G10
PS 22-2003	42	F	Coloured	Positive	G11
PS 22-2004	45	F	White	Negative	G12
DN005-10	34	M	Black	Positive	Clones

Supplementary Table 5. Clinical and Demographic Characteristics of Study Subjects. The results reported in this manuscript made use of archived samples from three South African cohorts. Immunosequencing and flow cytometry were performed on adolescents with (n=10) and without(n=10) latent tuberculosis and adults with active tuberculosis (n=10). T-Cell lines and clones were derived from five healthy adult donors with variable infection status. TST = Tuberculin Skin Test and IGRA = Interferon Gamma Release Assay, which are the two clinical assays used to define latent tuberculosis infection. Sputum smear is a semi-quantitative measure of bacterial burden in the sputum of patients with active tuberculosis. Ethnicity was self-described and 'coloured' refers to subjects of mixed-race ancestry in South Africa.