The signature of a T-cell response to KSHV persists across space and time in individuals with epidemic and endemic KS from Uganda

Shashidhar Ravishankar[1†], Andrea M.H. Towlerton[1,2†], Peter Mooka[2], James Kafeero[2], David G. Coffey[3], Lauri D. Aicher[1], Kelvin R. Mubiru[2], Lazarus Okoche[2], Prisca Atwinirembabazi[2], Joseph Okonye[2], Warren T. Phipps[4,5], Edus H. Warren[1,4,5,6]

[1]Translational Science and Therapeutics Division, Fred Hutchinson Cancer Center, Seattle, WA, United States
[2]Hutchinson Centre Research Institute – Uganda, Kampala, Uganda
[3]Division of Myeloma, Sylvester Comprehensive Cancer Center, University of Miami, Miami, FL, United States
[4]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Center, Seattle, WA, United States
[5]Department of Medicine, University of Washington, Seattle, WA, United States
[6]Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA, United States

[†] These authors have contributed equally to this work and share first authorship.

*Address for Correspondence:
Edus H. Warren
Translational Science and Therapeutics Division
Fred Hutchinson Cancer Center
1100 Fairview Avenue N, S3-204
Seattle, WA, 98109-1024, USA
ehwarren@fredhutch.org

Summary

Transcriptional profiling and T-cell receptor repertoire analysis of tumor biopsies from 144 Ugandan adults with KS demonstrate that a systemic, persistent, polyspecific KSHV-specific T-cell response exists and is likely inhibited by M2 macrophages within the tumor micro-environment.

Abstract

Inadequate T-cell control of Kaposi sarcoma-associated herpesvirus (KSHV) infection predisposes to development of Kaposi sarcoma (KS), but little is known about the T-cell response to KSHV. Postulating that KS tumors contain abundant KSHV-specific T-cells, we performed transcriptional profiling and T-cell receptor (TCR) repertoire analysis of tumor biopsies from 144 Ugandan adults with KS. We show that $CD8^+$ T-cells and M2-polarized macrophages dominate the tumor micro-environment (TME). The TCR repertoire of KS tumor infiltrating lymphocytes (TIL) is shared across non-contiguous tumors and persists across time. Clusters of T-cells with predicted shared specificity for uncharacterized antigens, potentially encoded by KSHV, comprise ~25% of KS TIL, and are shared across tumors from different time points and individuals. Single-cell RNA-sequencing of blood identifies a non-proliferating effector memory phenotype and captured the TCRs in 14,698 putative KSHV-specific T-cells. These results suggest that a polyspecific KSHV-specific T-cell response inhibited by M2 macrophages exists within the KS TME, and provide a foundation for studies to define its specificity at a large scale.

Introduction

T-lymphocytes play a central role in the control of chronic viral infections, particularly herpesvirus infections. The importance of T-cell immunity to herpesviruses is reflected in the fact that T-cells specific for Epstein-Barr virus (EBV) and cytomegalovirus (CMV), arguably the two most extensively studied human herpesviruses, comprise a significant fraction of the peripheral blood T-cell repertoire throughout life (Lindau et al., 2019; Lanfermeijer et al., 2021). The evolutionary struggle between herpesviruses and T-cells has led to the acquisition by herpesviruses of multiple strategies for evading recognition and elimination by T-cells. Failure to control herpesvirus replication can lead to fulminant disease and, in the case of the oncogenic gamma herpesviruses EBV and Kaposi sarcoma-associated herpesvirus (KSHV), predispose to tumorigenesis.

Although the $CD4^+$ and $CD8^+$ T-cell response to EBV has been well characterized, comparatively little is known about the T-cell response to KSHV (Nalwoga et al., 2021; Nalwoga and Whitby, 2022). One study used monocyte-derived dendritic cells transduced with recombinant lentiviruses encoding 31 different KSHV-encoded open reading frames to stimulate autologous $CD4^+$ and $CD8^+$ T-cells and demonstrated T-cell proliferation in response to several early and late lytic genes (Robey et al., 2009). Subsequent studies using interferon-gamma ELISpot assays in which T-cells from KSHV-exposed donors were stimulated with pools of overlapping peptides collectively tiling specific KSHV-encoded proteins or most of the predicted KSHV proteome concluded that T-cell responses to KSHV-encoded peptides are infrequent, weak, and heterogeneous (Hislop and Sabbah, 2008; Roshan et al., 2017; Nalwoga et al., 2021). T-cell proliferative responses to the critically important, multifunctional LANA-1 (ORF73)

protein, which is required for viral genome maintenance, appear to be the strongest and most frequently observed. Furthermore, IFN-γ-secreting $CD4^+$ T-cell clones recognizing LANA-1-derived peptides in an MHC class II-dependent manner were described in one study (Hislop and Sabbah, 2008). To date, however, few, if any, $CD8^+$ T-cell clones with rigorously documented, specific, high-avidity recognition for KSHV-encoded peptides have been reported.

That a potent and specific T-cell response to KSHV exists is suggested by the observation that uncontrolled KSHV replication predisposing to the development of Kaposi sarcoma (KS) is most often observed in the setting of T-cell deficiency or dysfunction. KSHV replication, for example, is frequently seen in KSHV-seropositive individuals with HIV infection or recipients of solid organ or allogeneic hematopoietic cell transplants. In these settings, moreover, KS can remit following initiation of antiretroviral therapy (ART) or withdrawal of immune suppression. In the present study we hypothesize that, as is the case with other herpesviruses, KSHV can also elicit a strong T-cell response. Furthermore, we hypothesize that KS tumors, which contain high levels of KSHV (Lidenge et al., 2020a; Tso et al., 2018; Rose et al., 2018), should contain significant numbers of $CD4^+$ and $CD8^+$ T-cells that are specific for KSHV-encoded antigens. Finally, we hypothesize that the dysfunctional T-cell response previously reported in individuals with KS (Nalwoga et al., 2021; Lidenge et al., 2020b) could potentially be due to T-cell inhibitory factors within the KS tumor micro-environment (TME).

To identify T-cells specific for KSHV we used adaptive immune receptor repertoire sequencing (AIRR-Seq) to define and enumerate the T-cell receptors (TCRs) expressed in 299 KS tumor biopsies (221 epidemic (HIV^+) KS, 78 endemic (HIV^-) KS) and 64 samples of uninvolved axillary skin (39 epidemic KS, 25 endemic KS) from 106 individuals with epidemic KS and 38 with endemic KS. The mean number of tumors analyzed per individual was 1.8

(range, 1-7). For most individuals, it was possible to analyze multiple spatially distinct tumors obtained at multiple time points over one year of observation. To further characterize the KS TME, we profiled the transcriptome of 51 KS tumors (39 epidemic and 12 endemic) using bulk RNA sequencing (RNA-seq) and performed targeted gene expression analysis on 27 additional tumors and 5 samples of normal uninvolved skin. Finally, to capture TCRs carried in putative KSHV-specific T-cells and profile their transcriptomes, we performed single-cell RNA sequencing of 20 PBMC samples (10 epidemic KS, 10 endemic KS) from nine individuals (5 epidemic KS, 4 endemic KS) collected across multiple time points over one year of observation.

Our results demonstrate that the TME in both epidemic and endemic KS is dominated by $CD8^+$ T-cells and macrophages with M2 polarization. Analysis of the TCR repertoire in KS tumors reveals a polyclonal idiosyncratic T-cell response that is largely private to each individual but shared across spatially and temporally distinct tumors from the same individual. This T-cell response has limited intra-tumoral expansion. Clustering TCR β chain CDR3 amino acid sequences by sequence similarity, we find that ~25% of the TCRs expressed by tumor infiltrating lymphocytes (TIL) in KS tumors fall into antigenic specificity groups with strong MHC association that are predicted to recognize similar or identical peptide/MHC complexes. In contrast to the lack of inter-individual sharing observed at the level of TCR CDR3 sequences, considerable inter-individual sharing of antigenic specificity groups was identified. To capture TCRs expressed by T-cells with predicted specificity for antigens encoded by KSHV, we performed single-cell TCR α/β chain sequencing and transcriptional profiling of peripheral blood cells from 5 patients with epidemic and 4 with endemic KS. We identified 14,698 T-cells bearing 4,283 unique αβ TCRs that belonged to antigenic specificity groups with predicted

specificity for KSHV, laying a foundation for future studies to define the specificity of the KSHV-specific T-cell response at large scale.

Results

We used a multi-omic approach to study the T-cells infiltrating Kaposi sarcoma tumors from HIV-seropositive and HIV-seronegative adults presenting to the Uganda Cancer Institute (Phipps et al., 2023). We compiled a total of 363 T-cell receptor β chain (*TRB*) repertoires from KS tumor biopsies and biopsies of uninvolved axillary skin from 106 HIV^+ and 38 HIV^- adults at study entry. We sequenced the *TRB* repertoire in 103 samples from endemic KS (78 tumors; 25 NATs) and 260 samples from epidemic KS (221 tumors; 39 NATs). We also explored the transcriptional profile of the TME in a subset of samples using bulk RNA-Seq (39 tumors from 29 individuals with epidemic KS and 12 tumors from 10 individuals with endemic KS). In addition, we quantified the expression of select human, CMV, EBV, HIV, and KSHV genes in 27 KS tumor samples and 5 normal skin samples from the cohort using a targeted gene expression analysis platform; 8 of these tumors were among the 51 on which bulk RNA-seq was performed. Finally, we profiled the transcriptome and T-cell repertoire in 20 KS PBMC samples from 9 individuals (5 epidemic KS, 4 endemic KS) using scRNA-seq (Supplementary Table S1).

KSHV and HIV gene expression in KS tumors

RNA-seq of 39 epidemic and 12 endemic KS tumors revealed consistent expression of both latent and lytic cycle KSHV genes within the TME. Kaposin (*K12*) showed the highest expression of all KSHV genes in all but two tumors (Figure 1A) and was also the most strongly

expressed gene in the 22 tumors evaluated with the targeted expression assay (Supplementary Figure 1A). The key latency gene *ORF73* (LANA-1) and the key lytic gene *ORF75* were both highly expressed across all tumor samples, reflecting the dynamic state of KSHV within the KS TME (Figure 1A). Consistent expression of other latency genes (*e.g.*, *ORF71*, *ORF72*) as well as early lytic genes (*e.g.*, *K15*, *K2*, *K7*) was also observed.

HIV gene expression was detected in all 39 tumors obtained from individuals with epidemic KS and 4 of the 12 tumors obtained from individuals with endemic KS, all of whom were ART-naive at study entry (Figure 1B). HIV Gag/Pol was consistently expressed in all 39 samples with epidemic KS, suggesting active replication of HIV within the TME. Targeted gene expression analysis with probes for HIV Env/Rev and Integrase in 22 KS tumors confirmed expression of HIV Env/Rev in 9 (41%) tumors and Integrase in 5 (23%) tumors (Supplementary Figure 1A).

M2 macrophages and $CD8^+$ T-cells are abundant in KS tumors

Macrophages with M2 polarization were predicted to be the most common immune cell type, accounting for a mean of 22% of the immune cells in the TME of epidemic KS and 38% of endemic KS (Figure 1C, 1D). $CD8^+$ T-cells were predicted to be the second largest group of immune cells within the TME, accounting for a mean of 21% and 18% of the cells in epidemic and endemic tumors, respectively (Figure 1C, 1D). Neither epidemic nor endemic KS tumors were predicted to contain any naïve $CD4^+$ T-cells. However, $CD4^+$ resting memory T-cells and $CD4^+$ activated memory T-cells were predicted to make up 5% and 4% of the immune cells in

endemic KS tumors. Epidemic KS tumors, as expected, appeared to contain few, if any, $CD4^+$ memory T-cells (1.3% memory resting; 0.2% memory activated) (Figure 1C, 1D).

Cell type deconvolution of previously published RNA-Seq datasets from 28 KS tumor - NAT tissue pairs and skin from 3 non-KS control subjects similarly showed an abundance of M2-polarized macrophages and $CD8^+$ T-cells. M2 macrophages accounted for a mean of 33% and 25% respectively, of all immune cells in the tumors from Lidenge *et al*. (Lidenge et al., 2020a) and Tso *et al*. (Tso et al., 2018). In the GTEx control skin tissues from both sun-exposed and non-sun-exposed skin, our deconvolution analysis identified high levels of resting mast cells, 38% and 35% respectively, but mast cells were uncommon in KS tumors. Low levels of $CD8^+$ T-cells were observed in the GTEx skin samples, with a mean of 7% observed in the sun-exposed and 6% in non-sun-exposed samples (Figure 2A).

We observed a negative correlation between percent abundance of M2 macrophages and $CD8^+$ T-cells across all 51 KS tumors from this study and 28 public KS tumor datasets. We observed a stronger negative correlation in endemic KS tumors (this study: $R = -0.61$, $p = 0.035$; Lidenge *et al.,*: $R= -0.88$, $p = 0.019$) compared to epidemic KS tumors (This study: $R = -0.23$, $p = 0.162$; Lidenge *et al.*: $R = -0.59$; $p = 0.01$; Tso *et al.*: $R = -0.82$, $p = 0.18$), with the exception of the 4 KS tumors from Tso *et al.* (Tso et al., 2018) (Figure 1E, Figure 2B). A similar negative correlation was observed across NAT samples from individuals with epidemic KS (Lidenge *et al.:* $R = -0.19$; $p = 0.44$; Tso *et al.*: $R = -0.96$, $p = 0.04$). However, in the limited number of NAT samples from individuals with endemic KS, there was a weak positive correlation between abundance of M2 macrophage and $CD8^+$ T-cells (Lidenge *et al.:* $R = 0.17$, $p = 0.74$).

Targeted gene expression analysis of KS tumors, when compared with samples of uninvolved skin, confirmed increased expression of lymphatic/endothelial markers as well as markers of M2-polarized macrophages and $CD8^+$ T-cells (Supplementary Figure 1). Increased expression of *LYVE1*, *NRP1*, *PROX1*, and *VEGFR3* in the tumors confirmed the presence of tumor cells of lymphatic/endothelial origin (Supplementary Figure 1A). Marked upregulation in KS tumors compared with uninvolved skin of the canonical M2 macrophage gene *CD163* as well as *TGFB1* was observed, as was upregulation of *CD8A* and *CD8B* (Supplementary Figure 1B). Increased expression of *CTLA4*, *PDCD1*, and *TIGIT* was also observed, indicating activity in pathways associated with inhibition of T-cell function.

KS TIL repertoires are diverse and poorly correlated with plasma KSHV viral load at study entry and 1-year survival

We assessed the repertoire diversity of the T-cell receptor β chain repertoire within the KS TME from 144 individuals with epidemic and endemic KS. For comparison we also analyzed AIRR-Seq data from tumor biopsies from 49 children and adolescents with Burkitt lymphoma (BL) from Uganda and Ghana (Lombardo et al., 2017). The 106 individuals with epidemic KS had significantly higher plasma KSHV titers (mean of 36,091 copies/μL vs 8,428 copies/μL) and significantly lower $CD4^+$ T-cell counts (mean of 226 cells/mL vs 808 cells/mL) than the 38 individuals with endemic KS at study entry (Supplementary Figure 2A; 2B) (Phipps et al., 2023).

We used the generalized entropy measure, Rényi entropy, calculated at different values of the parameter α (alpha), to assess the diversity of the TCR repertoires at the time of study enrollment

across the different groups. T-cell repertories from individuals with endemic KS showed significantly higher species richness and Shannon entropy when compared to individuals with epidemic KS from the current study and individuals with BL from Ghana (Figure 3A; 3B). However, individuals with endemic KS and BL from Uganda show no significant differences in diversity when looking at the two lower order diversity metrics ($\alpha = 0, 1$).

Simpson's diversity in tumors and NAT from individuals with epidemic KS was significantly lower compared to all other groups (Figure 3C). Nonetheless, the Simpson's diversity was comparable across endemic KS NAT, endemic KS tumors, and BL – Uganda tumors. The Inverse Berger-Parker diversity index showed no significant difference across endemic KS, and BL samples from Uganda (Figure 3D). However, the higher order Rényi diversity metrics ($\alpha = 2, \infty$) were significantly lower in both NAT and tumors from participants with epidemic KS and BL tumors from Ghana (Figure 3C, 3D). The average diversity of expanded clones (Figure 3D) in the KS tumors was comparable, if not greater, to that observed in AIRR-Seq datasets generated from the pediatric BL samples from Uganda and Ghana.

The Rényi curves for T-cell repertoires from endemic KS are consistently higher than those for the repertoires from epidemic KS at all values of the Rényi parameter α, indicating greater overall diversity of endemic repertoires (Figure 3E). The repertoires from endemic KS tumors are more diverse than those from BL tumors from Ghana but appear to be less diverse than the repertoires of BL tumors from Uganda. Nevertheless, repertoires from epidemic KS tumors have consistently lower diversity across all values of α when compared to endemic KS and BL tumors from Uganda.

Generalized Rényi diversity metrics on KS tumors obtained at the time of study entry show negligible negative correlation with $CD4^+$ counts, measured at study entry, at lower α values (Pearson correlation of -0.1 at $\alpha <= 2$) and negligible positive correlation with KSHV plasma titers at study entry at higher α values (Pearson correlation of 0.1 at $\alpha >= 4$) (Supplementary Figure 2C). Repertoire diversity metrics also showed minimal positive correlation with 1-year survival (Point-Biserial correlation between 0.1 – 0.16 at $\alpha = 0,1,2,\infty$) irrespective of HIV status.

TIL repertoires of KS tumors are largely private to each individual

We examined the frequency with which TCR β chain sequences are shared across the repertoires of KS tumors and NAT samples from 144 individuals in our HIPPOS KS cohort. Alluvial plots in (Figure 4A, 4B, 4C) show the frequencies of the 500 most frequent TCRs across KS tumor and NAT repertoires from representative pairs of individuals with epidemic and endemic KS from whom one normal adjacent tissue (NAT) sample along with three non-contiguous tumor biopsies samples, all obtained at study entry, were available for analysis. In individuals with endemic KS, a majority of the most frequent TCR sequences in the tumor are shared across the sets of three tumor biopsies obtained from each of the individuals (Figure 4A; Figure 5A). However, there is little, if any, overlap between the sequences observed in the three tumors and the matched NAT sample (Figure 4A; Figure 5A). We did not observe inter-subject sharing of any of the 500 most frequent TCR sequences in repertoires from endemic KS tumor biopsies or their matched NATs.

In contrast, tumor biopsies and matched NAT samples from individuals with epidemic KS show considerable intra-subject TCR sequence overlap (Figure 4B; Figure 5B). Indeed, a large majority of the 500 most expanded sequences in these representative subjects are shared between the KS tumors and the NAT sample from the same individual, and only a small minority are observed only in the KS tumors and not in the NAT sample. The large difference in the magnitude of TCR sequence sharing across autologous KS tumors and NAT samples observed in individuals with endemic and epidemic KS is highlighted in Figure 4C, which juxtaposes one set of alluvial plots from each of Figures 4A and 4B. The contrasts illustrated in the alluvial plot analysis are confirmed by analysis with UpSetR intersection plots, which illustrate the exact number of shared sequences between the KS tumor and NAT samples from representative pairs of individuals with endemic or epidemic KS (Figure 5, A-C).

The Morisita-Horn repertoire similarity indices for intra-subject tumor-tumor comparisons are significantly higher than those for intra-subject NAT-tumor comparisons, but the magnitude of this difference is far larger for individuals with endemic KS than for epidemic KS. The Morisita-Horn index for inter-subject repertoire comparisons between all possible pairs of individuals with endemic KS/endemic KS, epidemic KS/epidemic KS, and endemic KS/epidemic KS is near zero on average (Figure 4, D-F). Thus, at the level of TCR sequences, there is very limited overlap between the repertoires of KS tumors and NAT samples from different individuals, regardless of HIV status.

KS TIL carry the signature of a polyclonal response to unknown antigens

To determine whether TCR sequences carried in the 299 KS tumor TCR repertoires have previously been associated with a T-cell response to a specific pathogen or tissue antigen or a specific pathological condition, we systematically interrogated two publicly available databases, VDJdb (26, 27), and the McPAS-TCR (28) database, for annotated TCR sequences with high confidence of antigenic specificity (See Methods).

We observed that TCR sequences described in VDJdb and/or the McPAS-TCR database and associated with a T-cell response to a specific pathogen or tissue antigen appeared at very low levels (< 1% cumulative frequency) in tumors and NAT samples from individuals with endemic or epidemic KS (Figure 4G). The vast majority (99%) of TCR sequences observed in our KS TIL repertoires were not listed in VDJdb or the McPAS-TCR database, and thus could not be readily associated with previously defined T-cell responses to specific pathogen or tissue antigens (far right side of Figure 4G).

Of the very small subset (<1%) of KS tumor and NAT sequences that could be found in one or both databases, the most frequently associated pathogen/tissue specificities were CMV and Influenza A (Figure 4G). The TRB CDR3 amino acid sequence CASSSANYGYTF, for example, which has been associated with an HLA-A*02:01-restricted $CD8^+$ T-cell response to the prototypic CMV $pp65_{495-503}$ peptide with sequence NLVPMVATV (Trautmann et al., 2005), comprised 2.3% of the TIL repertoire in the tumors from an HLA-A*02:01^+ individual (008_049) with epidemic KS. Sequences known to be associated with T-cell responses to HIV were identified in samples from individuals with epidemic but not endemic KS. The TRB CDR3 amino acid sequence CASSVDKGGTDTQYF, for example, which has been associated with an HLA-B*42:01-restricted $CD8^+$ T-cell response to the HIV $Pol_{982-990}$ peptide with sequence

IIKDYGKQM (Kløverpris et al., 2015), was observed at low levels (mean frequency, 0.35%) in multiple KS tumors from 13 individuals, all of whom were HIV^+ and HLA-B*42:01$^+$.

Mirroring what is observed at the level of sequences (Figure 4G), specificity groups containing TRB CDR3 amino acid sequences associated in public databases with T-cell responses to defined pathogen or tissue antigens comprise <1% of the T-cell repertoires in KS tumors and NAT samples (Figure 4H). On average, however, 24% of the T-cell repertoire in NATs and tumors from individuals with endemic and epidemic KS – represented by the "Clustered Unknowns" in Figure 4H – is encompassed within antigenic specificity groups with as yet unknown specificity, labeled as such because they do not contain any sequences listed in public databases as associated with a T-cell response to a defined pathogen or tissue antigen. In total 151,771 unique TRB CDR3 amino acids sequences were classified as "clustered unknowns". We hypothesize that a significant fraction of the specificity groups classified as "Clustered Unknowns" are associated with T-cell responses to KSHV or KS. The remaining 74% of TRB CDR3 amino acid sequences in tumor and NAT samples from individuals with epidemic and endemic KS (far right in Figure 4H) were not grouped into any specificity groups defined by GLIPH2, and are termed "Unclustered Unknowns."

"Clustered unknown" specificity groups are shared across individuals

Although the repertoires of TCR sequences observed in TIL from both epidemic and endemic KS tumors are largely private and idiosyncratic, with negligible sharing between individuals, measurable inter-individual sharing of GLIPH2-defined antigenic specificity groups is observed.

We measured the pairwise sharing of specificity groups between KS tumor and NAT samples within and between individuals with epidemic and endemic KS using the Jaccard index. Among individuals with endemic KS, there is a significantly higher intra-subject tumor-tumor sharing of specificity groups when compared to intra-subject NAT-tumor, inter-subject NAT-NAT, and inter-subject tumor-tumor sharing (Figure 4I). Looking first at intra-subject sharing, we observed 26,509 specificity groups that were shared between at least two autologous tumors from spatially distinct, non-contiguous tumors derived from individuals with endemic KS. Looking at inter-subject sharing, we observed 42,747 specificity groups that were shared by two or more tumors from different individuals with endemic KS.

Similarly, in tumors and NAT samples from individuals with epidemic KS, we observe significantly greater intra-subject sharing of antigenic specificity groups between non-contiguous autologous KS tumors when compared to intra-subject NAT-tumor, inter-subject NAT-NAT, and inter-subject tumor-tumor sharing (Figure 4J). A total of 33,469 specificity groups are shared between two or more individuals with epidemic KS and 26,735 specificity groups are shared between two or more individuals with endemic KS. In addition, 35,731 specificity groups are shared between two or more individuals in the combined epidemic and endemic KS cohort, represented by the increased pairwise sequence sharing of specificity groups (Figure 4K).

In both the epidemic and endemic KS case, there is an enrichment of T-cells belonging to clusters of cells with unknown antigenic specificity in the tumors in both groups, with some tumors showing between two- and ten-fold more distinct T-cells compared to NATs. However, sequences belonging to these specificity groups each make up a small fraction of individual repertoires across endemic KS NAT samples (mean of 0.11%), endemic KS tumors (mean of

0.01%), epidemic KS NAT samples (mean of 0.06%), and epidemic KS tumors (mean of 0.04%) (Supplementary Figure 3).

T-cells from clustered unknown antigenic specificity groups persist in KS tumors across time

For 51 of the 144 individuals in the cohort, we collected additional KS tumors at several subsequent time points, most commonly 3 weeks, 2 months, and 6 months after study enrollment. We examined the persistence of T-cells from the clustered unknown specificity groups that we hypothesize could contain KSHV- or KS-specific T-cells. TCR sequences belonging to GLIPH2-defined clustered unknown antigenic specificity groups are consistently present at low levels of expansion in tumors obtained at multiple different time points after study enrollment. In both endemic (Figure 6A) and epidemic KS (Figure 6B) tumors, we observe remarkably stable intratumoral persistence of T-cells from clustered unknown specificity groups (blue and red bands respectively in Figure 6A, 6B) from study entry through the end of the one-year period of observation. On average these T-cells were found at a cumulative frequency of 30%, 26% and 23% at visit 1, 5 and 8, respectively, in tumors from individuals with endemic KS. In individuals with epidemic KS, this group comprised 31%, 27% and 25% of the repertoire in tumors collected at visits 1, 5, and 8, respectively.

T-cells in clustered unknown specificity groups are antigen-experienced effector cells

To define the phenotype of T-cells assigned to clustered unknown specificity groups and to capture their TCR α and β chain sequences at large scale, we performed scRNA-seq with TCR α/β chain sequencing on 20 PBMC samples obtained at multiple time points after study entry from 9 individuals in our cohort (5 epidemic KS, 4 endemic KS). The datasets were normalized, integrated, and mapped to a reference PBMC single-cell atlas, creating a combined dataset comprising 246,521 single cells (Supplementary Figure 4), of which 61,763 cells carry a productive αβ TCR (mean, 3,087 such cells per sample), collectively carrying 31,576 unique αβ TCRs (Figure 7A). Of these 61,763 cells, 38,411 (62%) carrying 9,805 unique productive αβ TCRs could be presumptively mapped back to the KS TIL repertoire by virtue of their TCR β chain sequences, and 14,698 of those cells, carrying 4,283 unique αβ TCRs, belonged to "clustered unknown" antigenic specificity groups defined by GLIPH2 analysis (Figure 7B).

Cell-type classification of the 61,763 cells carrying a productive αβ TCR revealed that 59,411 (96.19%) were classified as T-cells and 1,180 (1.9%) as NK cells. The remaining 1,172 cells were classified as other types of immune cells despite carrying a productive αβ TCR, which is likely due to inadequate coverage of marker genes used for classification. Similarly, 14,094 of the 14,698 (95.8%) cells carrying a productive αβ TCR that mapped back to the KS TIL repertoire and were assigned to a "clustered unknown" antigenic specificity group by GLIPH2 were classified as T-cells, with 306 classified as NK cells. Upon closer inspection of T-cell subtypes, we observed abundant $CD8^+$ effector memory T-cells (mean of 45% in endemic KS and 68% in epidemic KS) and $CD4^+$ central memory T-cells (mean of 12% in endemic KS and

8% in epidemic KS) in both epidemic and endemic KS samples (Figure 7C, 7D). Endemic KS samples contained a prominent population of $CD4^+$ cytotoxic T lymphocytes (mean, 18%; Figure 7C). However, this trend is skewed for 008_241 and 008_255, where we observed very few “clustered unknown” T-cells at time points V01 (008_241: 46 T-cells), V09 (008_241: 5, 008_255: 12) and V10 (008_255: 66).

“Clustered unknown” T-cells that were classified as $CD8^+$ T-cells, NK T-cells, and $CD4^+$ cytotoxic T lymphocytes expressed high levels of markers of effector function, including *GZMB*, *PRF1*, and *IFNG* (Figure 8). Cells from these clusters also expressed T-cell inhibitory markers including *LAG3*, *TIGIT*, and *TOX*. A large majority of the T-cells showed expression of memory markers *IL7R* and *KLRB1*, but only a small subset of these cells showed expression of the proliferation marker *MKI67*. Cells from the $CD4^+$ T-cell, $CD8^+$ T-cell, and NK cell clusters also express *KLRG1*, *BATF*, and *PRDM1*, which are markers associated with proliferative dysfunction in T- and NK cells. These genes are also upregulated in short-lived terminally differentiated $CD4^+$ T-cells (Borys et al., 2022). However, no notable differentially expressed genes between "clustered unknown" T-cells and other T-cells found in the blood were observed.

T-cells from different individuals that were assigned to the same antigenic specificity group frequently demonstrated heterogeneous transcriptional profiles. A total of 105 antigenic specificity groups were shared by at least 3 of the nine individuals whose PBMCs were studied at the single-cell level (Figure 9A). Interestingly, despite $CD8^+$ effector memory T-cells being the prevailing cell type, some specificity groups comprised TCRs from cells displaying both $CD8^+$ and $CD4^+$ T-cell phenotypes (Figure 9B). The most frequent specificity groups, identified by the motifs SLG%GE and SP%NTGE, were observed in 4 and 3 of the 9 individuals, respectively (Figure 9C). Despite their prevalence across samples, these T-cells were found at low overall

frequencies (Mean frequency: 0.02% and Max frequency: 2.8%), consistent with our observations from the KS TME.

Discussion

In this study we investigated the T-cell repertoire and gene expression profile of KS tumors and non-involved skin in 144 individuals with KS (106 epidemic and 38 endemic KS) from Uganda. We sequenced the *TRB* repertoire of 299 KS tumor biopsies (221 epidemic and 78 endemic KS), and 64 samples of uninvolved axillary skin (39 epidemic and 25 endemic KS). To provide context in which to evaluate the TCR repertoire of T-cells infiltrating KS tumors, we included in our analysis analogous T-cell AIRR-Seq datasets, generated using identical wet- and dry-lab workflows, from 49 pediatric Burkitt lymphoma (another gamma herpesvirus-associated malignancy) tumors from Uganda and Ghana. The gene expression profile of a subset of 51 KS tumors was determined using bulk RNA-seq. We obtained a complementary targeted assessment of gene expression from 22 KS tumor and 5 normal skin samples from a subset of individuals with epidemic and endemic KS. We characterized the T-cell receptor repertoire and profiled the phenotype of putative KSHV-specific T-cells of interest using scRNA-seq GEX+VDJ data from 20 PBMC samples collected at multiple time points from a subset of 5 individuals with epidemic KS and 4 with endemic KS. Finally, we included in our analyses RNA-seq data generated from 6 paired epidemic KS tumor - Normal Adjacent Skin (NAT) samples from Tso et al. (Tso et al., 2018) and 18 paired epidemic KS tumor - NAT samples, 6 paired endemic KS tumor - NAT samples, and 3 control skin samples from Lidenge et al. (Lidenge et al., 2020a). To our knowledge, this is the most comprehensive assessment of the KS TME and KS TIL repertoire to date.

RNA-seq analysis of gene expression within the KS TME showed consistent expression of both latent and lytic cycle KSHV genes. We also found significant expression of a diverse set of

HIV transcripts in all 39 epidemic KS tumors studied using RNA-seq. Surprisingly, RNA-seq analysis also suggested expression of a smaller, variable subset of HIV transcripts in 4 of 12 endemic tumors; we believe that this observation is an artefact of the *k*-mer-based strategy used for quantifying gene expression from the human and HIV transcriptomes. Targeted gene expression analysis of a partially overlapping set of KS tumors confirmed expression of both latent and lytic cycle KSHV genes in KS tumors. The presence of high-level KSHV gene expression in KS tumors suggests that T-cells specific for KSHV are very likely to be found there, as well.

The immune cell repertoire of KS TME is dominated by M2-polarized macrophages and $CD8^+$ T-cells. It is likely that any recognition of viral transcripts by these $CD8^+$ T-cells would be targeting both latent and lytic stage genes of the KSHV life cycle, also suggested in prior literature (Robey et al., 2009; Lambert et al., 2006; Guihot et al., 2006; Wilkinson et al., 2002; Wang et al., 2000; Hislop and Sabbah, 2008; Nalwoga et al., 2021; Roshan et al., 2017; Woodberry et al., 2005). Our assessment of the TIL repertoire in KS using AIRR-Seq shows that the diversity of T-cells within the TME is comparable to that seen in Burkitt lymphoma, another gamma herpesvirus-associated malignancy that is common in children in Uganda and other countries in sub-Saharan Africa. The KS TIL repertoire is largely idiosyncratic and polyclonal, with low frequencies of T-cells with known antigenic specificity. This is supported by prior studies which have found few if any public T-cell responses in individuals seropositive for KSHV (Phipps et al., 2023).

Our GLIPH2 analysis predicts that nearly 25% of the KS TIL repertoire is comprised of T-cells that can be assigned to antigenic specificity groups predicted to recognize similar or

identical, but as yet uncharacterized, antigens. These specificity groups do not contain T-cells with defined antigenic specificity as annotated in VDJdb or the McPAS-TCR database. A total of 14,698 T-cells belonging to these specificity groups were also identified in the peripheral blood at multiple time points using scRNA-seq. We hypothesize that at least a subset of these T-cells with yet uncharacterized antigenic specificity are likely specific for antigens encoded by KSHV.

Transcriptional profiling with scRNA-seq revealed that "clustered unknown" T-cells were primarily comprised of $CD8^+$ effector memory cells (74%), and to a lesser extent $CD4^+$ cytotoxic T-lymphocytes (8%) and $CD4^+$ central memory T-cells (5%). These T-cells showed consistent expression of effector genes including *GZMB*, *PRF1*, and *IFNG.* A majority of T-cells expressed markers associated with memory phenotype, including *IL7R* and *KLRB1*, with only a small population expressing proliferative markers such as *MKI67.* We observe variable expression of genes associated with proliferative dysfunction such as *KLRG1*, *BATF*, and *PMRD1*. T-cell inhibitory/exhaustion markers, including *LAG3*, *TIGIT*, and *TOX*, are consistently expressed at low levels across all T-cell subtypes in the peripheral blood. Collectively, this highlights the persistence of a dysfunctional phenotype across this subset of putative KSHV-specific-T-cells. These transcriptional characteristics should be interpreted cautiously, however, given that the phenotypic characterization was performed on T-cells from peripheral blood, and the phenotype of T-cells that carry the same TCRs but are located within the TME could be different. Nonetheless, our targeted gene expression analysis also shows an enrichment of markers associated with T-cell inhibition and exhaustion (*LAG3*, *TIGIT*, *PDCD1*, and *CTLA4)* in KS tumors. Furthermore, while "clustered unknown" T-cells make up 24-25% of the repertoire in both the KS TME and peripheral blood, their limited expansion is suggestive of proliferative dysfunction. The fact that these putative KSHV-specific T-cell neighborhoods are found in

multiple individuals, in spatially distinct, metachronous tumor lesions and circulating in blood across the time span of one year, provide compelling support for the notion that these are not just random occurrences but *prima facie* evidence of T-cell recognition of KSHV-encoded antigens within the KS TME.

M2-polarized macrophages are predicted to be the dominant immune cell type not only in the KS tissue samples from the present study, but also in 56 epidemic and endemic matched KS tumor-NAT pairs and 3 control skin samples from sub-Saharan Africa from prior publications (Tso et al., 2018; Lidenge et al., 2020a). Characterization of fold change in gene expression between KS tumor and NAT from a subset of individuals with epidemic and endemic KS using targeted gene expression assays further confirmed the enrichment of the M2 macrophage marker *CD163* within KS tumors.

We hypothesize that M2 macrophages could play a key role in the limited expansion of the putative KSHV-specific T-cells that we identified within the KS TME. From our RNA-seq analysis, we show that, though present at lower levels than in KS tumors, M2 macrophages are nonetheless also found in normal adjacent skin of individuals with KS. M2 macrophage-associated T-cell inhibitory activity could in part explain difficulties encountered by multiple investigators trying to identify KSHV-specific $CD8^+$ cytotoxic T-cells in individuals who are seropositive for KSHV. *In vitro* studies have shown that KSHV infection of macrophages induces M2 polarization (Montani et al., 2020). It is reasonable to hypothesize that M2 macrophages play an important role in the pathogenesis of KS lesions by inhibiting the proliferation of KSHV-specific T-cells in the TME. This possibility is one of the foci of our current and future studies.

Our study has several limitations. Our RNA-seq data from the 39 epidemic KS tumors and 12 endemic KS were generated from two different sequencing runs, raising the possibility of batch effects confounding any inferred differences between the epidemic and endemic tumors. However, the overall consistency of our findings across all 51 RNA-Seq datasets as well as across 59 comparable RNA-Seq datasets from prior publications (Tso et al., 2018; Lidenge et al., 2020a) indicate the presence of a shared biological mechanism. Nonetheless, we chose not to perform extensive analysis of the differences in the TME between epidemic and endemic KS given the possibility of residual batch effects. Phenotypic characterization of T-cells of interest was limited to that observed within the peripheral blood. The characterization of immune cell-types within the tumor was limited to deconvolution of bulk RNA-Seq of KS tumors. This limits our ability to comment on the phenotype of putative KSHV-specific T-cells within the TME in KS. However, there is a high degree of consensus in the cell types identified in the blood and within the TME. Understanding the changes to T-cell states within the TME is also a major focus of our current and future studies.

Despite these limitations, we have identified a repertoire of T-cells bearing 4,283 unique αβ TCRs that are consistently found in KS tumors across space and time and are strongly associated with the expression of specific MHC alleles. We postulate that at least some of these T-cells are specific for either KSHV or KS tumors. Identification of the 4,283 unique TCRs carried by this repertoire of T cells and of the MHC alleles with which they are closely associated provides a comprehensive foundation for defining the targets of the KSHV-specific T-cell response. Indeed, significant current effort in our group is focused on identifying the targets of putative KSHV-specific T-cells that are closely associated with and therefore likely restricted by MHC alleles that are common in the Ugandan population. Complementary studies are investigating the

transcriptional profile and specific localization of T-cells bearing putative KSHV-specific TCRs within the KS TME, and exploring how their function and proliferation might be inhibited by the prominent population of M2 macrophages that our studies have also identified in the TME. Understanding the factors that limit antiviral T-cell response within the KS TME will hopefully guide the development of new therapeutic strategies based on vaccination or adoptive cell transfer to prevent or treat KS and other KSHV-associated diseases.

Materials and Methods

Subjects and samples

Adults 18 years or older with histologically confirmed Kaposi sarcoma who were naïve to ART and had received no prior chemotherapy were enrolled between October 2012 and May 2021 on a prospective cohort study (termed “HIPPOS”) at the Uganda Cancer Institute (UCI) in Kampala, Uganda, as previously described (Phipps et al., 2023). Participants completed an enrollment visit and up to 10 follow-up visits over ~1 year. At enrollment, participants completed a standardized medical history and physical examination, a blood sample and an oral swab were collected for KSHV testing, and peripheral blood mononuclear cells (PBMC) were isolated. HIV serology was performed, and $CD4^+$ T-cell and plasma HIV RNA testing was completed for HIV-seropositive participants. UCI medical charts were reviewed to record routine blood count data and chest X-ray findings. Full-thickness punch biopsies (4 x 4 mm diameter) of at least three non-contiguous cutaneous KS lesions and a punch biopsy of axillary skin not visibly involved with KS were also obtained and placed in either neutral buffered formalin or RNA*later* (Thermo Fisher Scientific, Waltham, MA, USA).

At follow-up visits, participants completed an interim medical history and detailed physical exam to assess treatment response and provided additional blood samples and oral swabs for KSHV testing. Up to three additional biopsies of non-contiguous KS lesions and a blood sample for PBMC isolation were obtained at follow-up visits at 3-, 36-, and 48-weeks following enrollment, and placed in RNA*later*.

Participants were offered chemotherapy at the UCI per standard guidelines independent of the study. First-line therapy consisted of combination bleomycin and vincristine (BV) or paclitaxel; both regimens were given every 3 weeks for six cycles. HIV-seropositive participants not already in HIV care were referred to local HIV clinics for management and initiation of ART according to Ugandan Ministry of Health guidelines.

Biopsies of Burkitt lymphoma (BL) were acquired from 2 cohorts of African children after obtaining verbal or written informed consent using institutional review board–approved protocols, as previously described (Lombardo et al., 2017). The first cohort comprised 12 children with BL who presented to the Uganda Cancer Institute (UCI) in Kampala, Uganda between August 2013 and March 2014. The second cohort comprised 37 children diagnosed with BL between 1975 and 1992 at Korle Bu Hospital in Accra, Ghana, from whom cryopreserved, archival BL tumor biopsies were available.

Laboratory procedures

Plasma and oral swab samples were evaluated for KSHV DNA by quantitative, real-time PCR at the UCI-Fred Hutch Cancer Centre Laboratory in Kampala, Uganda as previously described (Phipps et al., 2023; Rose et al., 2018). Samples with >150 copies/mL of KSHV DNA were considered positive (Magaret et al., 2007). $CD4^+$ T-cell counts were measured by flow cytometry, and HIV-1 RNA levels were measured using real-time RT-PCR.

RNA Sequencing (RNA-seq)

Total RNA was extracted from cryopreserved KS tumor biopsies using the RNeasy Fibrous Tissue Mini Kit (QIAGEN, Venlo, The Netherlands), quantitated on a Qubit 3.0 Fluorometer

(Thermo Fisher), and assessed on an Agilent 2200 Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA). RNA samples with RNA integrity number (RIN) greater than 7 were used for sequencing library preparation. Sequencing libraries from RNA extracted from cryopreserved KS tumor biopsies from 39 HIV-seropositive HIPPOS participants (epidemic KS) were prepared in the Fred Hutchinson Cancer Center (FHCC) Genomics Core Facility (Seattle, WA) using the TruSeq RNA Sample Prep Kit v2 (Illumina, San Diego, CA, USA), as previously described (Rose et al., 2018). Library size distributions were validated using the Agilent 2200 TapeStation. Library QC, pooling of indexed libraries, and cluster optimization were performed using the Qubit 3.0 Fluorometer. Pooled libraries were sequenced on an Illumina HiSeq 2500 in "High Output" mode with a paired-end, 50-base read configuration. Sequencing libraries from RNA extracted from cryopreserved KS tumor biopsies from an additional 12 HIV-seronegative HIPPOS participants (endemic KS) were prepared and sequenced at Azenta (South Plainfield, NJ) using the NEBNext Ultra II RNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA). The sequencing libraries were validated on the Agilent TapeStation, quantified using the Qubit 3.0 Fluorometer as well as by quantitative PCR (KAPA Biosystems, Wilmington, MA, USA), and sequenced on a partial S4 lane of an Illumina NovaSeq 6000 with a paired-end, 150-base read configuration.

Targeted gene expression analysis

RNA was extracted from formalin-fixed paraffin-embedded (FFPE) slides from 25 epidemic KS and 2 endemic KS tumor biopsies and 5 samples of uninvolved skin using the AllPrep DNA/RNA FFPE Kit (QIAGEN). Gene expression analysis was performed on RNA samples on the Nanostring platform (Nanostring Technologies, Seattle, WA, USA) using the Immunology

V2 Panel supplemented with 30 user-defined viral and host gene probes (Supplementary Table S1). Of the 25 epidemic KS tumors profiled, 8 were also profiled using bulk RNA-Seq. Raw probe counts were normalized using the geometric mean of 15 housekeeping genes followed by analysis using nSolver software (Version 4.0) and the nSolver Advanced Analysis package (version 2.0.115).

T-cell receptor β chain variable region sequencing

Genomic DNA was extracted from 363 cryopreserved biopsies of KS tumor and normal skin (termed normal adjacent tissue, NAT) using the DNeasy Blood & Tissue Kit (QIAGEN) and prepared for T-cell receptor β chain (*TRB*) variable region sequencing at survey level resolution (Robins, 2013) using the ImmunoSEQ hsTCRB v3.0 assay (Adaptive Biotechnologies, Seattle, WA). *TRB* sequencing was performed on pools of libraries from 11 - 15 tumor or skin samples on the Illumina MiSeq platform using a v3 150-cycle kit (Illumina, San Diego, CA) in the FHCC Genomics Core Facility (Seattle, WA) or the UCI-Fred Hutch Cancer Centre Laboratory (Kampala, Uganda).

MHC genotyping

High-resolution genotyping of the MHC class I and class II loci of all participants in the epidemic and endemic KS cohorts and the Uganda BL cohort was performed on genomic DNA via next-generation sequencing as a contract service by Scisco Genetics, Seattle, WA. The full standardized protocol is available online (Genetics, 2019). Sequencing of the PCR amplicons spanning the MHC loci was performed on the Illumina MiSeq platform using a v2 500-cycle kit

(Illumina, San Diego, CA) in the FHCC Genomics Core Facility or the UCI-Fred Hutch Cancer Centre Laboratory.

Single-cell RNA sequencing (scRNA-seq)

Viable cells were thawed from cryopreserved KS PBMC specimens and subjected to scRNA-seq using the 10X Genomics (10X Genomics, Pleasanton, CA) platform. Single index chemistry was performed for all libraries using the following reagents to generate scRNA-seq libraries for V1.0 chemistry: the Chromium™ Single Cell 5' Library Construction Kit, Chromium™ Single Cell A Chip Kit, ChromiumTM Single Cell 5' Library & Gel Bead Kit, Chromium™ Single Cell 5' Feature Barcode Library Kit, and the Chromium™ Single-Cell Enrichment, Human T-Cell kit. For libraries generated using V1.1 chemistry, the Chromium™ Next GEM Single Cell 5' Library and Gel Bead Kit v1.1, and the Chromium™ Next GEM Chip G Single Cell Kit were used.

Libraries were quantified using both the Qubit 3.0 Fluorometer (Thermo Fisher, Waltham, MA), TapeStation using the Agilent TapeStation (Agilent, Santa Clara, CA) and by qPCR using the KAPA Library Quantification Kit for Illumina (Roche, Basel, Switzerland). Libraries were pooled and sequenced on the Illumina NovaSeq 6000 (Illumina, San Diego, CA) on an SP1 flow cell in the FHCC Genomics Core.

Analysis of RNA-seq data

A total of 4,353,735,624 paired-end reads were obtained from 51 RNA-Seq datasets (Mean: 85,367,365 reads per sample) generated from 39 epidemic KS and 12 endemic KS tumor samples. We used BBDuk (Brian, 2014) to remove any adapter contamination using a starting k-mer size of 23 and min-k size of 11, with an allowed Hamming distance of 1 between query k-

mer and reference adapter k-mers. K-mers with average quality less than 6 were trimmed from both ends of the read. Reads with length ≤ 10bp were discarded after trimming. (18) We quantified the transcript and gene level expression against the human (Gencode v42), HHV8 (GCF_000838265.1), and HIV (GCF_000864765.1) transcriptome index using Salmon (v1.9.0) (Patro et al., 2017). K-mer sizes of 31 for the human transcriptome and 21 for the viral transcriptomes (HIV and HHV8) were used to build the reference index and quantify gene expression from the filtered reads.

We generated a gene-expression matrix from the Salmon gene-level expression data with sample name as rows, gene ensemble ID as columns, and expression values in each cell represented in transcripts per million reads (TPM). Using the LM-22 immune cell-type model from CibersortX (Newman et al., 2015), we quantified the cell-type composition of the tumor samples.

We included in our analysis RNA-seq data generated from 6 paired epidemic KS tumor - Normal Adjacent Skin (NAT) samples from Tso *et al.* (Tso et al., 2018) and 18 paired epidemic KS tumor - NAT samples, 6 paired endemic KS tumor - NAT samples, and 3 control skin samples from Lidenge *et al.* (Lidenge et al., 2020a) in the analysis to validate the findings from our cohort. The raw dataset was downloaded from SRA using the NCBI SRA-Toolkit. The 59 single-end RNA-seq datasets from the two cohorts were processed using the same BBMAP-Salmon pipeline modified to accommodate single-end sequencing datasets. Gene expression matrices with TPMs for 30 sun-exposed and 30 non-sun-exposed control skin tissue data from the GTEx v8 data release (https://gtexportal.org/home/) were used to assess the baseline transcriptomic profile of normal skin. These 119 RNA-seq quantification datasets from 6 batches

of sequencing were formatted for analysis with CibersortX and the immune cell-type composition for each dataset was determined using the LM-22 immune cell-type model available through the CibersortX web-interface (https://cibersortx.stanford.edu/).

T-cell receptor repertoire analysis

T-cell repertoire analyses were conducted using the R statistical programming language (Team, 2021). A reproducible workflow of all steps in data QC, data transformation, and analysis was developed using Jupyter notebooks and the Nextflow workflow management system (Tommaso et al., 2017). Repertoire analysis was performed with open-source R packages and custom scripts were implemented in LymphoSeq2, an R package that we developed for the exploration of Adaptive Immune Receptor Repertoire Sequencing (AIRR-Seq) data (Wu et al., 2021).

To investigate intra-tumoral T-cell heterogeneity, we sequenced a total of 474 TCR repertoires from both tumor and adjacent normal tissues collected from 146 individuals enrolled in the study. Repertoires with less than 1,000 total *TRB* sequences were removed from the analysis, leaving a total of 363 TCR repertoires from 299 KS tumor biopsies (221 epidemic KS, 78 endemic KS), 64 samples of uninvolved axillary skin (39 epidemic KS, 25 endemic KS), and 7 samples from single-cell tumor suspensions subjected to a Rapid Expansion Protocol (6 epidemic KS and 1 endemic KS). In addition, we included in the repertoire analyses 49 AIRR-Seq datasets generated from BL tumors from children from Uganda and Ghana using the same wet- and dry-lab platforms as the KS tumor and NAT tissues, which also passed the same sequencing threshold of 1,000 total *TRB* sequences.

To minimize the impact of differences in the effective depth of sequencing achieved on the samples on the assessment of repertoire diversity, repertoires were sampled down to a total count of 1,000 *TRB* nucleotide sequences (Towlerton et al., 2022). The sampled repertoires were aggregated over the unique *TRB* CDR3 nucleotide sequences and repertoire diversity statistics were averaged over 100 iterations.

Repertoire diversity was measured using Rényi generalized entropy at different values of α ($\alpha = 0, 0.25, 0.5, 1, 2, 4, 8, 16, 32, 64, \infty$) to capture different characteristics of the T-cell repertoire diversity. Species richness ($\alpha = 0$) and Shannon entropy ($\alpha = 1$) are entropy measures that are more affected by the presence of distinct T-cells in the repertoire at any degree of abundance. Simpson's diversity ($\alpha = 2$) and Inverse Berger-Parker ($\alpha = \infty$) are entropy measures that weigh the more abundant T-cells in the repertoire to a greater extent. Hence, they give a summary of the diversity among expanded T-cells in the repertoire. We used a custom R wrapper around Vegan's Rényi function to calculate the depth normalized Rényi diversity profile for each dataset (J et al., 2023).

Prediction of T-cell antigenic specificity and MHC restriction

The TRB CDR3 amino acid sequences in each repertoire were compared with the VDJdb (Bagaev et al., 2020; Goncharov et al., 2022; Shugay et al., 2017) and McPAS-TCR (Tickotsky et al., 2017) databases of annotated TRB sequences, with high confidence of antigenic specificity (VDJdb score >1; McPAS-TCR antigen identification method < 3), to identify sequences that have previously been associated with a T-cell response to a specific microbial or tissue antigen or a specific pathological condition. TCRs from public databases were filtered to ensure that the

recorded MHC allele in the database was also found in the individual with the sequence of interest.

Antigenic specificity groups of T-cell receptors carrying specific TRB CDR3 amino acid sequences were computationally predicted using GLIPH2 (Huang et al., 2020). Sequences from T-cell receptor repertoires were labeled with repertoire ID and phenotype (epidemic KS, endemic KS). MHC typing information from the KS cohort was formatted as per GLIPH2 requirements. A reference database for GLIPH2 was generated using 35 control AIRR-Seq PBMC datasets from sub-Saharan Africa. The custom reference catalogued the CDR3 length distribution, V-gene usage, and CDR3 composition of 932,116 unique TRB CDR3 sequences.

TRB CDR3 amino acid sequences with likely antigenic specificity for CMV, EBV, HIV, *Mycobacterium tuberculosis* (*Mtb*), or an identified tissue antigen, were identified by first finding all specificity groups that contained an annotated TRB sequence reported in one or more of the public databases. All TRB CDR3 amino acid sequences belonging to these specificity groups were identified as sequences with likely antigenic specificity for CMV, EBV, HIV, *Mtb*, a specific tissue antigen, or pathological condition (Towlerton et al., 2022). GLIPH2 groups that did not contain any TRB CDR3 amino acid sequences with associated antigenic specificity were categorized as “Clustered Unknowns”. Sequences not grouped into antigenic specificity groups by GLIPH2 and with yet undetermined antigenic specificity were categorized as “Unclustered Unknowns”.

T-cell repertoire similarity

The similarity between two TCR repertoires was calculated with both the Morisita-Horn Index and Jaccard Index using Compairr (Rognes et al., 2022). The Morisita-Horn index was used to assess the overlap between the *TRB* CDR3 amino acid sequences found in two repertoires. The Jaccard index was used to calculate the extent to which GLIPH2-defined specificity groups were shared between two repertoires. Overlapping sequences within repertoires were visualized using alluvial plots generated with LymphoSeq2 and UpSetR (Conway et al., 2017).

Analysis of single-cell RNA-Seq gene expression (GEX) and VDJ libraries

We built a structured analysis pipeline using NextFlow to process scRNA-seq GEX + VDJ sequencing libraries from PBMC samples of 20 individuals with epidemic and endemic KS. We quantified the number of cells across both GEX and VDJ libraries using the Cell Ranger multi (pipeline version: Cell Ranger-7.0.1). The references for Cell Ranger GEX (Human reference (GRCh38) - 2020-A) and VDJ (Human V(D)J reference (GRCh38) ensembl-7.0.0) were obtained from the 10X Cell Ranger website (https://www.10xgenomics.com/support/software/cell-ranger/downloads). For quality control and preprocessing, we utilized the Bioconductor package scuttle to compute per-cell metrics (McCarthy et al., 2016). We used the isOutlier function from the Bioconductor package scater to identify outliers based on per-cell metrics, with a threshold of 5 median average deviations. Cell counts were normalized, and clustering was performed using scuttle and (Lun et al., 2016; McCarthy et al., 2016). Doublet clusters were identified based on gene expression using the scDblFinder (Germain et al., 2021). Seuratv4 (Hao et al., 2020), and scRepertoire (Borcherding

et al., 2020) were employed to merge gene expression and VDJ data for each sample. T-cell sequence data from VDJ libraries were incorporated into the metadata field of the Seurat objects.

Gene expression and VDJ data from the 20 PBMC single-cell datasets were integrated using the Seurat rpca protocol [https://satijalab.org/seurat/articles/integration_rpca.html]. Ensembl IDs were replaced with gene symbols to prepare the integrated object for cell type classification.

Antigen grouping and phenotypic characterization of scRNA-seq GEX + VDJ libraries

We used the SeuratV4 reference mapping protocol (https://satijalab.org/seurat/articles/multimodal_reference_mapping#intro-seurat-v4-reference-mapping) for cell-type classification of cells in the single-cell libraries utilizing a multimodal PBMC atlas comprising 161,764 cells as a reference (Hao et al., 2021). To ensure consistency with the reference atlas normalization, we first normalized cell counts within the integrated Seurat object using the scTransform function from Seurat. We identified anchors between the reference and query through the FindTransferAnchors function, transferring cell type labels using the MapQuery function.

We merged GLIPH2 results from bulk TRB sequencing data with TRB sequences from single-cell datasets to identify antigen groups. The merged table was filtered to include only TRB sequences classified as “Clustered Unknowns” from the previous section. The persistence of these T-cells and their predicted phenotypes across all single-cell datasets was visualized using T-cell pack plots. The outer circles of the T-cell pack plots represent GLIPH2-identified “Clustered

Unknown" specificity groups, while inner circles depict individual TCR sequences belonging to each group. The circles are color-coded by subject ID or predicted cell type, with circle size indicating T-cell count in the respective samples.

Statistical methods

The statistical significance of differences between epidemic (HIV-seropositive) and endemic (HIV-seronegative) KS participants and BL participants from Ghana and Uganda was assessed the Wilcoxon rank-sum test for continuous variables. The Pearson correlation test was used to calculate the correlation between clinical characteristics such as KSHV plasma viral load, HIV viral load, and $CD4^+$ T-cell counts with diversity metrics from AIRR-Seq data. We used a generalized linear model (GLM) to examine the relationship between percentage abundance of $CD8^+$ T-cell and M2 Macrophages from RNA-Seq datasets on KS tumors. The correlation between percent abundance of the two cell types in KS tumors was measured using a two-sided Pearson test. To measure correlation between diversity metrics from AIRR-Seq data at $\alpha = 0, 1, 2, \infty$ and 1-year survival, we first convert response data to a binary classification. CR (Complete response) and PR (Partial response) responses were grouped as responders [assigned a value of 1]. SD (Stable disease) and PD (Progressive disease) responses were grouped as non-responders [assigned a value of 0]. The correlation between the two variables was then assessed using the Point-Biserial correlation test.

Ethical considerations

Approval for the study was obtained from the Makerere University School of Medicine Research and Ethics Committee, Fred Hutchinson Cancer Center IRB, and Uganda National Council for Science and Technology. All participants provided documentation of informed consent.

Supplementary Material

Supplementary Table S1: Summary of all sequencing data from individuals with epidemic and endemic KS included in the present study.

Supplementary Figure 1: Targeted gene expression profiling of epidemic and endemic KS tumors. (**A)** Heatmap of expression of four human endothelial genes and select CMV, EBV, KSHV, and HIV genes in 2 tumors from individuals with endemic KS (top) and 20 tumors from individuals with epidemic KS (bottom) determined with a targeted gene expression platform. (**B)** Heatmap of expression of 21 human immune-related genes defining 10 characteristics of the KS TME determined using a targeted gene expression platform in 27 KS tumors (left) and 5 NAT samples (right), displayed with the Viridis palette. The log2 fold-change between average expression of each gene in the KS tumors vs. NAT samples is displayed in the column at far right using the Plasma palette.

Supplementary Figure 2: Plasma KSHV and HIV viral load and $CD4^+$ T-cell count are not correlated with T-cell repertoire diversity in KS tumors. **(A)** Pooled distribution of all

plasma KSHV titers (copies/mL) and **(B)** $CD4^+$ T-cell counts from individuals with endemic (Blue) and epidemic (Red) KS. Significance between comparisons is indicated by : ns ($p > 0.05$), * ($p <= 0.05$), ** ($p <= 0.01$), *** ($p <= 0.001$), **** ($p <= 0.0001$). **(C)** Pearson correlation matrix of plasma KSHV and HIV viral load and $CD4^+$ T-cell count at study entry with from individuals with epidemic and endemic KS and generalized Rényi diversity of the T-cell repertoire diversity in their respective KS tumors for the indicated values of the Rényi parameter α (0, 0.25, 0.5, 1, 2, 64, ∞). Although the generalized Rényi diversity metrics are strongly correlated with each other (min r = 0.5; max r = 1), negligible correlation is observed between the clinical metrics and repertoire diversity.

Supplementary Figure 3: Frequency of T-cells in GLIPH2-defined clusters with unknown antigenic specificity in tumors and NAT samples from individuals with epidemic and endemic KS. Cumulative frequency (in %) of the ensemble of TCR sequences assigned to GLIPH2-defined specificity groups with unknown antigenic specificity in Epidemic KS – NAT samples (Pink), Epidemic KS – Tumors (Red), Endemic KS – NAT samples (Light blue) and Endemic KS – Tumors (Blue).

Supplementary Figure 4: Single-cell RNA sequencing of PBMC from individuals with epidemic and endemic KS reveals an abundance of $CD8^+$ effector memory cells, $CD4^+$ central memory cells, monocytes, and NK cells. (A) Integrated reference mapped UMAP showing 246,521 cells from 20 single-cell GEX+VDJ libraries generated with PBMCs from 9 individuals with KS (4 endemic KS, 5 epidemic KS). (B, C) Alluvial plots tracking the relative

abundance of 19 cell-types from 20 PBMC samples collected across multiple time points from (B) 4 individuals with endemic KS and (C) 5 with epidemic KS.

Data Availability

AIRR-Seq datasets from the 363 KS tumor and NAT samples and 49 BL tumor samples from Ghana and Uganda are publicly available for access on the Adaptive Biotechnologies immunoSEQ Analyzer portal at (https://clients.adaptivebiotech.com/pub/ravishankar-2024-jem) https://clients.adaptivebiotech.com/pub/. All R scripts, metadata, GLIPH2 references and Nextflow workflows to analyze and visualize the data used in the study are available on GitHub (https://github.com/shashidhar22/ks_manuscript). Data generated by Nextflow pipelines and GLIPH2 used to generate figures for the manuscript are available on Zenodo (https://zenodo.org/records/10594758). RNA-Seq datasets from 51 KS tumors and scRNA-seq GEX+VDJ datasets from 20 KS PBMC libraries are available through the NCBI under BioProject number PRJNA1068629 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1068629).

Acknowledgements

We would like to acknowledge the HIPPOS study team, the Data and Laboratory Teams at the Hutchinson Center Research Institute – Uganda, HIPPOS study participants, and their families for enabling this research. We would also like to thank The Genotype-Tissue Expression (GTEx) Project for data used in this manuscript: the Sun-exposed skin and Non-sun-exposed skin datasets from the v8 data release, obtained through the GTEx Portal on 04/27/23. The Genotype-

Tissue Expression (GTEx) Project is supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

Ethics Statement

The studies involving human participants were reviewed and approved by the Makerere University School of Medicine Research and Ethics Committee, the Fred Hutchinson Cancer Center IRB, and the Uganda National Council for Science and Technology. All participants provided documentation of informed consent.

Author Contributions

EW, WP, AT and SR conceived and designed the study. PM and JK collected the tissue samples. AT, KM, LO, PA and JO generated the AIRR-Seq and RNA-Seq data. SR, DC, AT, and EW analyzed the data. SR, AT, and EW drafted the manuscript. All authors edited the manuscript. All authors contributed to the article review and approved the submitted version.

Funding

This research was supported by awards from the National Cancer Institute (R01 CA217138, R01 CA239287, P30 CA015704-043S3) and from the Immunotherapy and Pathogen-Associated Malignancies Integrated Research Centers at the Fred Hutchinson Cancer Center. This research was also supported by the Specimen Processing and Research Cell Bank and the Genomics and

Bioinformatics Shared Resources of the Fred Hutch/University of Washington Cancer Consortium (P30 CA015704). Bioinformatic analysis was supported by the Scientific Computing Infrastructure at the Fred Hutchinson Cancer Center funded by ORIP grant S10OD028685.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be constructed as a potential conflict of interest.

Figure legends

Figure 1: Epidemic and endemic KS tumors show both lytic and latent KSHV gene expression and contain abundant macrophages with M2 polarization and CD8$^+$ T-cells. (A) KSHV gene expression in 12 endemic (top) and 39 epidemic (bottom) KS tumors. Latent (La), Immediate early (IE), Early lytic stage 1 (E1), Early lytic stage 2 (E2), Early lytic stage 3 (E3), Late lytic stage 4 (L4), and Late lytic stage 5 (L5). (B) Expression of HIV genes across all 51 KS tumors. (C) The relative abundance of 22 immune cell types in the 51 KS tumors estimated using the LM22 model from CibersortX. (D) Average relative abundance of the indicated 8 immune cell types in the 12 endemic (Blue) and 39 epidemic (Red) KS tumors, estimated using CibersortX. Boxes represent the interquartile range (IQR) and the upper/lower whiskers extend from the hinge to the largest/smallest value, respectively, no further than 1.5 * IQR from the hinge. Significance between comparisons is indicated by ns ($p > 0.05$), * ($p <= 0.05$), ** ($p<= 0.01$), *** ($p<= 0.001$), **** ($p <= 0.0001$). (E) CD8$^+$ T-cell density is negatively correlated with M2 macrophage density in both epidemic and endemic KS tumors.

Figure 2: KS tumors show an enrichment of M2 macrophages and CD8$^+$ T-cells in comparison to control skin samples. (A) Heatmap of relative abundance of the indicated immune cell types found in KS tumors and skin samples from individuals with KS as well as control skin samples from the GTEx consortium, determined via deconvolution of RNA-seq data using CibersortX. The RNA-seq datasets analyzed were derived from (left to right) 3 control skin samples, 6 paired tumor and NAT samples from individuals with endemic KS, and 18 paired tumor and NAT samples from individuals with epidemic KS from Lidenge *et al.*; 4 paired tumor and NAT samples from individuals with epidemic KS from Tso *et al.*; and 30 non-sun-exposed control skin samples and 30 sun-exposed control skin samples from the GTEx consortium. **(B)** CD8$^+$ T-cell abundance shows negative correlation with M2 macrophage abundance across samples from individuals with KS, however this trend is not significant in control dataset from the GTEx consortium.

Figure 3: Diversity of T-cells infiltrating endemic KS tumors is comparable to the diversity observed in TIL from EBV-associated Burkitt lymphoma. (A, B, C, D, E) Rényi entropy (RE) comparisons, at different values of α, across tissue samples: epidemic KS – NAT (Pink), epidemic KS – tumor (Red), endemic KS – NAT (Light blue), endemic KS – tumor (Blue), Burkitt lymphoma (BL) tumors from Ghana (Yellow), and BL tumors from Uganda (Brown). (A) Species richness (RE $\alpha = 0$) and (B) Shannon entropy (RE $\alpha = 1$) are metrics influenced by the number of unique T-cells found in the repertoire. (C) Simpson's diversity (RE $\alpha = 2$) and (D) Berger-Parker index ($\alpha = \infty$) are metrics that are influenced by expanded T-cell clones in the repertoire. (E) Rényi diversity profiles for all six cohorts show the variation in repertoire diversity with change in α. The T-cell repertoire of one group is more diverse than another group

only when its Rényi profile is consistently higher at all values of α. Significance between comparisons is indicated by ns ($p > 0.05$), * ($p <= 0.05$), ** ($p<= 0.01$), *** ($p<= 0.001$), **** ($p <= 0.0001$).

Figure 4: T-cell repertoires in epidemic and endemic KS tumors are idiosyncratic and contain few TCRs specific for viral/self-antigens but share T-cells clusters with predicted specificity for uncharacterized antigen groups. (A, B, C) Alluvial plots tracking the 500 most frequent TRB sequences in three spatially distinct tumors and one NAT sample from representative samples acquired at study entry, from (A) endemic KS, (B) epidemic KS, and (C) epidemic KS and endemic KS tumors. (D, E, F) Pairwise similarity between T-cell repertoires in KS tumors and NAT samples from individuals with epidemic and endemic KS was quantified using the Morisita-Horn index: intra-subject tumor-tumor (Purple), intra-subject NAT-tumor (Orange), inter-subject NAT-NAT (Green) and inter-subject tumor-tumor (Magenta). (G, H) Frequency distributions of TRB CDR3 amino acid sequences: (G) in public databases or (H) computationally predicted to be associated with an MHC-restricted T-cell response to a specific pathogen or tissue antigen. Sequences listed in these databases/computationally predicted to be associated with T-cell responses to two or more pathogens are classified as “Multi-Pathogen”. (G) Sequences not found in these databases are classified as having “Unknown” specificity. (H) Antigenic specificity clusters that did not contain any sequences listed in the public databases as being associated with a T-cell response to a known pathogen or tissue were classified as “Clustered Unknowns”. Sequences that were not assigned to an antigenic specificity cluster were classified as “Unclustered Unknowns”. (I, J, K) The Jaccard index quantifying the pairwise similarity of repertoires of GLIPH2-defined antigenic specificity groups between KS tumors and

NAT samples from individuals with (I) endemic KS, (J) epidemic KS, and (K) endemic and epidemic KS colored by the same categories as in (D, E, F). Significance between comparisons is indicated by ns ($p > 0.05$), * ($p <= 0.05$), ** ($p<= 0.01$), *** ($p<= 0.001$), **** ($p <= 0.0001$).

Figure 5: Synchronously acquired, non-contiguous tumors from endemic KS show a greater degree of intra-tumor TCR sharing than epidemic KS, and inter-tumor sharing of TCR sequences is low both within and across epidemic and endemic KS. (A, B, C) UpSetR plots showing the overlap of unique TCR sequences between NAT and synchronously obtained tumor samples from two representative individuals with (A) endemic KS or (B) epidemic KS, or (C) one individual with endemic and one individual with epidemic KS. Patterns of sequence sharing are grouped along the horizontal axis and coded by color: sequences found in a single endemic KS tumor (Blue), sequences found in a single endemic NAT sample (Light Blue), sequences found in a single epidemic NAT sample (Pink), sequences found in a single epidemic KS tumor (Red), shared intra-subject tumor-tumor (Purple), intra-subject NAT-tumor (Orange), inter-subject tumor-tumor (Magenta), and inter-subject NAT-NAT (Green) sequences, sequences with other patterns of sharing (Grey).

Figure 6: T-cells from GLIPH2-defined groups with specificity for unknown antigens persist in the TME across time and space in individuals with epidemic and endemic KS. (A, B) Alluvial plots of the most frequent 500 TRB CDR3 amino acid sequences observed in tumor samples obtained at different time points in representative individuals with (A) Endemic KS and (B) Epidemic KS. Highlighted are persistent TCR β chain CDR3 sequences from GLIPH2-defined “clustered unknown” antigenic specificity groups in both Endemic (Blue) and Epidemic

(Red) KS lesions. Colored dots on the body figures above each alluvial plot denote the sites from which each tissue sample was collected and are reproduced below the corresponding column containing the repertoire data for that tissue sample.

Figure 7: T-cells in "clustered unknown" antigenic specificity groups are effector memory cells. (A) Integrated, reference-mapped UMAP showing cell type classification of 61,763 cells carrying a productive αβ TCR from 20 single-cell GEX+VDJ libraries generated from PBMCs from 9 individuals with KS (4 endemic and 5 epidemic KS). (B) UMAP highlighting 14,698 "clustered unknown" T-cells (red) identified in blood that can be mapped back to the KS TME based on their TCR β chain sequence. (C, D) Alluvial plots tracking the relative abundance over time of cells carrying a productive αβ TCR with the indicated phenotypes from (C) 4 individuals with endemic KS and (D) 5 with epidemic KS.

Figure 8: Gene expression profiles from the integrated single-cell UMAP of "clustered unknown" T-cell from KS PBMC samples show expression of T-cell effector, inhibition, memory, proliferation, and dysfunction markers. Integrated reference mapped UMAPs highlight the expression of T-cell effector function (first row), inhibition (second row), memory (third row), proliferation (third row), and dysfunction (fourth row) markers across all "clustered unknown" T-cells detected in PBMC samples from individuals with epidemic and endemic KS.

Figure 9: "Clustered unknown" antigenic specificity groups are observed across serially acquired PBMC samples and shared between individuals with epidemic and endemic KS. (A, B) T-cell pack plots highlighting T-cells (inner circles) belonging to "clustered unknown"

GLIPH2 specificity groups (outer circles) present in single-cell PBMC datasets from at least 3 individuals. Inner T-cell circles are colored (A) by the type of KS, in shades of red when found in samples with epidemic KS and shades of blue when found in samples with endemic KS, or (B) by the predicted cell phenotype. Numbers within the inner circles in (A) indicate the count of T-cells with that TCR. (C) Amino acid sequence logos for the TCR α and β chain CDR3 regions in the αβ TCRs that comprise the two most frequently shared "clustered unknown" specificity groups across all 20 single-cell PBMC datasets from individuals with epidemic and endemic KS.

References

Bagaev, D.V., R.M.A. Vroomans, J. Samir, U. Stervbo, C. Rius, G. Dolton, A. Greenshields-Watson, M. Attaf, E.S. Egorov, I.V. Zvyagin, N. Babel, D.K. Cole, A.J. Godkin, A.K. Sewell, C. Kesmir, D.M. Chudakov, F. Luciani, and M. Shugay. 2020. VDJdb in 2019: database extension, new analysis infrastructure and a T-cell receptor motif compendium. *Nucleic Acids Res*. 48:D1057–D1062. doi:10.1093/nar/gkz874.

Borcherding, N., N.L. Bormann, and G. Kraus. 2020. scRepertoire: An R-based toolkit for single-cell immune receptor analysis. *F1000Research*. 9:47. doi:10.12688/f1000research.22139.2.

Borys, S.M., A.K. Bag, L. Brossay, and D.O. Adeegbe. 2022. The Yin and Yang of Targeting KLRG1+ Tregs and Effector Cells. *Front. Immunol.* 13:894508. doi:10.3389/fimmu.2022.894508.

Brian, B. 2014. BBMap: A Fast, Accurate, Splice-Aware Aligner. *Lawrence Berkeley National Laboratory*.

Conway, J.R., A. Lex, and N. Gehlenborg. 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*. 33:2938–2940. doi:10.1093/bioinformatics/btx364.

Genetics, S. 2019. HLA-24S-v6 Protocol.

Germain, P.-L., A. Lun, C.G. Meixide, W. Macnair, and M.D. Robinson. 2021. Doublet identification in single-cell sequencing data using scDblFinder. *F1000Research*. 10:979. doi:10.12688/f1000research.73600.2.

Goncharov, M., D. Bagaev, D. Shcherbinin, I. Zvyagin, D. Bolotin, P.G. Thomas, A.A. Minervina, M.V. Pogorelyy, K. Ladell, J.E. McLaren, D.A. Price, T.H.O. Nguyen, L.C. Rowntree, E.B. Clemens, K. Kedzierska, G. Dolton, C.R. Rius, A. Sewell, J. Samir, F. Luciani, K.V. Zornikova, A.A. Khmelevskaya, S.A. Sheetikov, G.A. Efimov, D. Chudakov, and M. Shugay. 2022. VDJdb in the pandemic era: a compendium of T cell receptors specific for SARS-CoV-2. *Nat. Methods*. 19:1017–1019. doi:10.1038/s41592-022-01578-0.

Guihot, A., N. Dupin, A.-G. Marcelin, I. Gorin, A.-S. Bedin, P. Bossi, L. Galicier, E. Oksenhendler, B. Autran, and G. Carcelain. 2006. Low T Cell Responses to Human Herpesvirus 8 in Patients with AIDS-Related and Classic Kaposi Sarcoma. *J. Infect. Dis.* 194:1078–1088. doi:10.1086/507648.

Hao, Y., S. Hao, E. Andersen-Nissen, W.M. Mauck, S. Zheng, A. Butler, M.J. Lee, A.J. Wilk, C. Darby, M. Zagar, P. Hoffman, M. Stoeckius, E. Papalexi, E.P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L.B. Fleming, B. Yeung, A.J. Rogers, J.M. McElrath, C.A. Blish, R. Gottardo, P. Smibert, and R. Satija. 2020. Integrated analysis of multimodal single-cell data. *bioRxiv*. 2020.10.12.335331. doi:10.1101/2020.10.12.335331.

Hao, Y., S. Hao, E. Andersen-Nissen, W.M. Mauck, S. Zheng, A. Butler, M.J. Lee, A.J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E.P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L.M. Fleming, B. Yeung, A.J. Rogers, J.M. McElrath, C.A. Blish, R. Gottardo, P. Smibert, and R. Satija. 2021. Integrated analysis of multimodal single-cell data. *Cell*. 184:3573-3587.e29. doi:10.1016/j.cell.2021.04.048.

Hislop, A.D., and S. Sabbah. 2008. CD8+ T cell immunity to Epstein-Barr virus and Kaposi's sarcoma-associated herpes virus. *Semin. Cancer Biol.* 18:416–422. doi:10.1016/j.semcancer.2008.10.005.

Huang, H., C. Wang, F. Rubelt, T.J. Scriba, and M.M. Davis. 2020. Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. *Nat Biotechnol*. 38:1194–1202. doi:10.1038/s41587-020-0505-4.

J, O., S. G, B. F, K. R, L. P, M. P, O. R, S. P, S. M, S. E, W. H, B. M, B. M, B. B, B. D, C. G, C. M, D.C. M, D. S, E. H, F. R, F. M, F. B, H. G, H. M, L. L, M. D, O. M, R.C. E, S. T, S. A, T.B. C, and W. J. 2023. vegan: Community Ecology Package.

Kløverpris, H.N., R. McGregor, J.E. McLaren, K. Ladell, M. Harndahl, A. Stryhn, J.M. Carlson, C. Koofhethile, B. Gerritsen, C. Keşmir, F. Chen, L. Riddell, G. Luzzi, A. Leslie, B.D. Walker, T. Ndung'u, S. Buus, D.A. Price, and P.J. Goulder. 2015. CD8+ TCR Bias and Immunodominance in HIV-1 Infection. *J. Immunol.* 194:5329–5345. doi:10.4049/jimmunol.1400854.

Lambert, M., M. Gannagé, A. Karras, M. Abel, C. Legendre, D. Kerob, F. Agbalika, P.-M. Girard, C. Lebbe, and S. Caillat-Zucman. 2006. Differences in the frequency and function of HHV8-specific CD8 T cells between asymptomatic HHV8 infection and Kaposi sarcoma. *Blood*. 108:3871–80. doi:10.1182/blood-2006-03-014225.

Lanfermeijer, J., P.C. de Greef, M. Hendriks, M. Vos, J. van Beek, J.A.M. Borghans, and D. van Baarle. 2021. Age and CMV-Infection Jointly Affect the EBV-Specific CD8+ T-Cell Repertoire. *Front. Aging*. 2:665637. doi:10.3389/fragi.2021.665637.

Lidenge, S.J., A.V. Kossenkov, F.Y. Tso, J. Wickramasinghe, S.R. Privatt, O. Ngalamika, J.R. Ngowi, J. Mwaiselage, P.M. Lieberman, J.T. West, and C. Wood. 2020a. Comparative transcriptome analysis of endemic and epidemic Kaposi's sarcoma (KS) lesions and the secondary role of HIV-1 in KS pathogenesis. *Plos Pathog*. 16:e1008681. doi:10.1371/journal.ppat.1008681.

Lidenge, S.J., F.Y. Tso, O. Ngalamika, J. Kolape, J.R. Ngowi, J. Mwaiselage, C. Wood, and J.T. West. 2020b. Lack of CD8+ T-cell co-localization with Kaposi's sarcoma-associated herpesvirus infected cells in Kaposi's sarcoma tumors. *Oncotarget*. 11:1556–1572. doi:10.18632/oncotarget.27569.

Lindau, P., R. Mukherjee, M.V. Gutschow, M. Vignali, E.H. Warren, S.R. Riddell, K.W. Makar, C.J. Turtle, and H.S. Robins. 2019. Cytomegalovirus Exposure in the Elderly Does Not Reduce CD8 T Cell Repertoire Diversity. *J. Immunol. Author Choice*. 202:476–483. doi:10.4049/jimmunol.1800217.

Lombardo, K.A., D.G. Coffey, A.J. Morales, C.S. Carlson, A.M.H. Towlerton, S.E. Gerdts, F.K. Nkrumah, J. Neequaye, R.J. Biggar, J. Orem, C. Casper, S.M. Mbulaiteye, K.G. Bhatia, and E.H. Warren. 2017. High-throughput sequencing of the B-cell receptor in African Burkitt lymphoma reveals clues to pathogenesis. *Blood Adv.* 1:535–544. doi:10.1182/bloodadvances.2016000794.

Lun, A.T.L., D.J. McCarthy, and J.C. Marioni. 2016. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research*. 5:2122. doi:10.12688/f1000research.9501.2.

Magaret, A.S., A. Wald, M.-L. Huang, S. Selke, and L. Corey. 2007. Optimizing PCR Positivity Criterion for Detection of Herpes Simplex Virus DNA on Skin and Mucosa. *J. Clin. Microbiol.* 45:1618–1620. doi:10.1128/jcm.01405-06.

McCarthy, D.J., K.R. Campbell, A.T.L. Lun, and Q.F. Wills. 2016. scater: pre-processing, quality control, normalisation and visualisation of single-cell RNA-seq data in R. *bioRxiv*. 069633. doi:10.1101/069633.

Montani, M.S.G., L. Falcinelli, R. Santarelli, M. Granato, M.A. Romeo, N. Cecere, R. Gonnella, G. D'Orazi, A. Faggioni, and M. Cirone. 2020. KSHV infection skews macrophage polarisation towards M2-like/TAM and activates Ire1 α-XBP1 axis up-regulating pro-

tumorigenic cytokine release and PD-L1 expression. *Brit J Cancer*. 123:298–306. doi:10.1038/s41416-020-0872-0.

Nalwoga, A., R. Roshan, K. Moore, V. Marshall, W. Miley, N. Labo, M. Nakibuule, S. Cose, R. Rochford, R. Newton, and D. Whitby. 2021. Kaposi's sarcoma-associated herpesvirus T cell responses in HIV seronegative individuals from rural Uganda. *Nat. Commun.* 12:7323. doi:10.1038/s41467-021-27623-8.

Nalwoga, A., and D. Whitby. 2022. Adaptive immune responses to Kaposi's sarcoma-associated herpesvirus. *Curr. Opin. Immunol.* 77:102230. doi:10.1016/j.coi.2022.102230.

Newman, A.M., C.L. Liu, M.R. Green, A.J. Gentles, W. Feng, Y. Xu, C.D. Hoang, M. Diehn, and A.A. Alizadeh. 2015. Robust enumeration of cell subsets from tissue expression profiles. *Nat. methods*. 12:453–457. doi:10.1038/nmeth.3337.

Patro, R., G. Duggal, M.I. Love, R.A. Irizarry, and C. Kingsford. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods*. 14:417–419. doi:10.1038/nmeth.4197.

Phipps, W., S.V. Adams, P. Mooka, J. Kafeero, S. Sekitene, D. Mubiru, J. Nankoma, C. Namirembe, L. Okoche, E.B. Namubiru, S. Kayemba, K.K. Baker, M.W. Redman, C. Casper, J. Orem, and E.H. Warren. 2023. A prospective study of clinical outcomes of HIV-associated and HIV-negative Kaposi sarcoma in Uganda. *Aids*. 37:51–59. doi:10.1097/qad.0000000000003376.

Robey, R.C., D. Lagos, F. Gratrix, S. Henderson, N.C. Matthews, R.J. Vart, M. Bower, C. Boshoff, and F.M. Gotch. 2009. The CD8 and CD4 T-Cell Response against Kaposi's Sarcoma-Associated Herpesvirus Is Skewed Towards Early and Late Lytic Antigens. *PLoS ONE*. 4:e5890. doi:10.1371/journal.pone.0005890.

Robins, H. 2013. Immunosequencing: applications of immune repertoire deep sequencing. *Curr. Opin. Immunol.* 25:646–652. doi:10.1016/j.coi.2013.09.017.

Rognes, T., L. Scheffer, V. Greiff, and G.K. Sandve. 2022. CompAIRR: ultra-fast comparison of adaptive immune receptor repertoires by exact and approximate sequence matching. *Bioinformatics*. 38:4230–4232. doi:10.1093/bioinformatics/btac505.

Rose, T.M., A.G. Bruce, S. Barcy, M. Fitzgibbon, L.R. Matsumoto, M. Ikoma, C. Casper, J. Orem, and W. Phipps. 2018. Quantitative RNAseq analysis of Ugandan KS tumors reveals KSHV gene expression dominated by transcription from the LTd downstream latency promoter. *Plos Pathog*. 14:e1007441. doi:10.1371/journal.ppat.1007441.

Roshan, R., N. Labo, M. Trivett, W. Miley, V. Marshall, L. Coren, E.M.C. Castro, H. Perez, B. Holdridge, E. Davis, R. Matus-Nicodemos, V.I. Ayala, R. Sowder, K.M. Wyvill, K. Aleman, C. Fennessey, J. Lifson, M.N. Polizzotto, D. Douek, B. Keele, T.S. Uldrick, R. Yarchoan, C.

Ohlen, D. Ott, and D. Whitby. 2017. T-cell responses to KSHV infection: a systematic approach. *Oncotarget*. 8:109402–109416. doi:10.18632/oncotarget.22683.

Shugay, M., D.V. Bagaev, I.V. Zvyagin, R.M. Vroomans, J.C. Crawford, G. Dolton, E.A. Komech, A.L. Sycheva, A.E. Koneva, E.S. Egorov, A.V. Eliseev, E. Van Dyk, P. Dash, M. Attaf, C. Rius, K. Ladell, J.E. McLaren, K.K. Matthews, E.B. Clemens, D.C. Douek, F. Luciani, D. van Baarle, K. Kedzierska, C. Kesmir, P.G. Thomas, D.A. Price, A.K. Sewell, and D.M. Chudakov. 2017. VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Res.* 46:gkx760-. doi:10.1093/nar/gkx760.

Team, R.C. 2021. R: A Language and Environment for Statistical Computing. Vienna, Austria.

Tickotsky, N., T. Sagiv, J. Prilusky, E. Shifrut, and N. Friedman. 2017. McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. *Bioinformatics*. 33:2924–2929. doi:10.1093/bioinformatics/btx286.

Tommaso, P.D., M. Chatzou, E.W. Floden, P.P. Barja, E. Palumbo, and C. Notredame. 2017. Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* 35:316–319. doi:10.1038/nbt.3820.

Towlerton, A.M.H., S. Ravishankar, D.G. Coffey, C.E. Puronen, and E.H. Warren. 2022. Serial Analysis of the T-Cell Receptor β-Chain Repertoire in People Living With HIV Reveals Incomplete Recovery After Long-Term Antiretroviral Therapy. *Front. Immunol.* 13:879190. doi:10.3389/fimmu.2022.879190.

Trautmann, L., M. Rimbert, K. Echasserieau, X. Saulquin, B. Neveu, J. Dechanet, V. Cerundolo, and M. Bonneville. 2005. Selection of T Cell Clones Expressing High-Affinity Public TCRs within Human Cytomegalovirus-Specific CD8 T Cell Responses. *J. Immunol.* 175:6123–6132. doi:10.4049/jimmunol.175.9.6123.

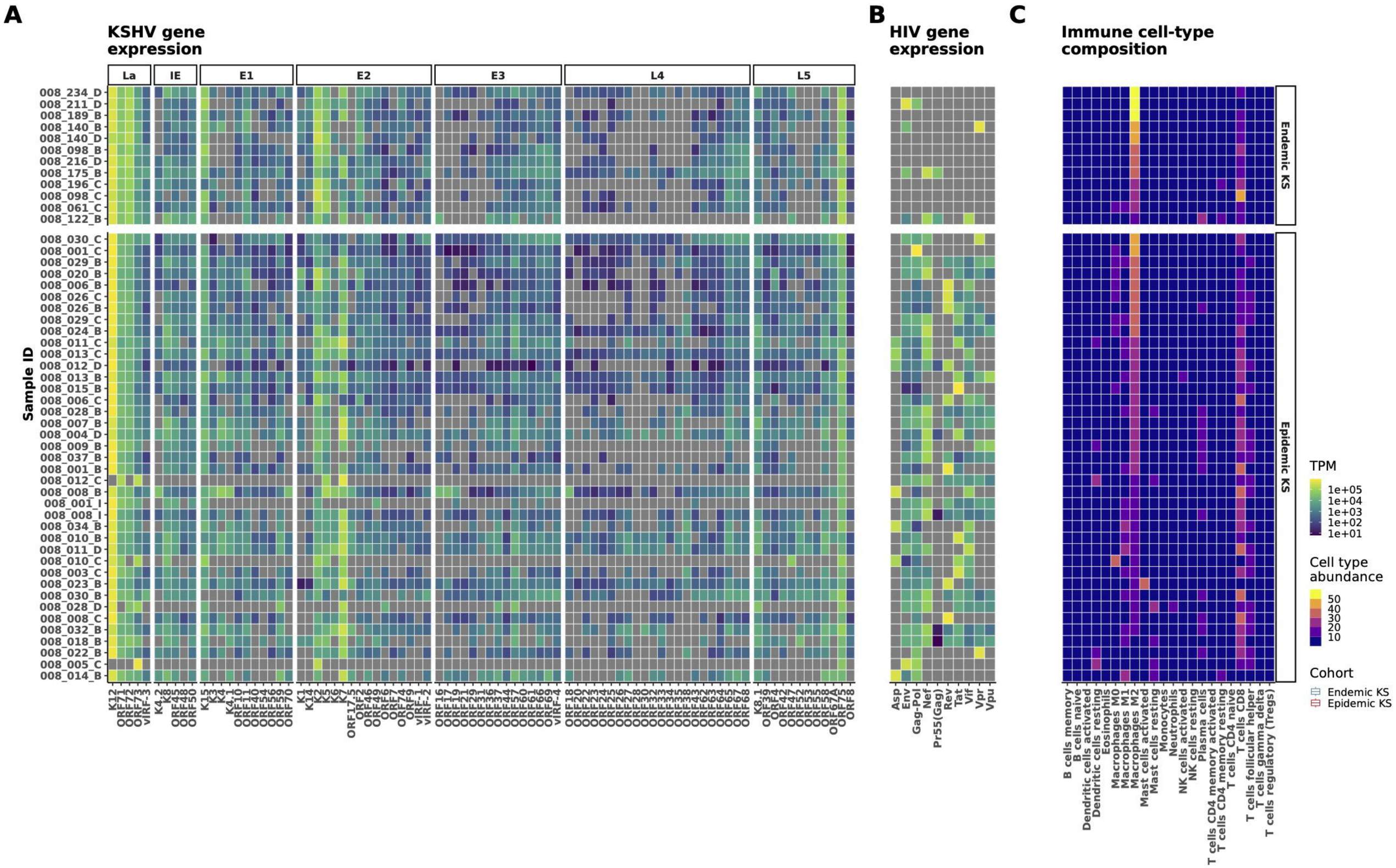
Tso, F.Y., A.V. Kossenkov, S.J. Lidenge, O. Ngalamika, J.R. Ngowi, J. Mwaiselage, J. Wickramasinghe, E.H. Kwon, J.T. West, P.M. Lieberman, and C. Wood. 2018. RNA-Seq of Kaposi's sarcoma reveals alterations in glucose and lipid metabolism. *Plos Pathog*. 14:e1006844. doi:10.1371/journal.ppat.1006844.

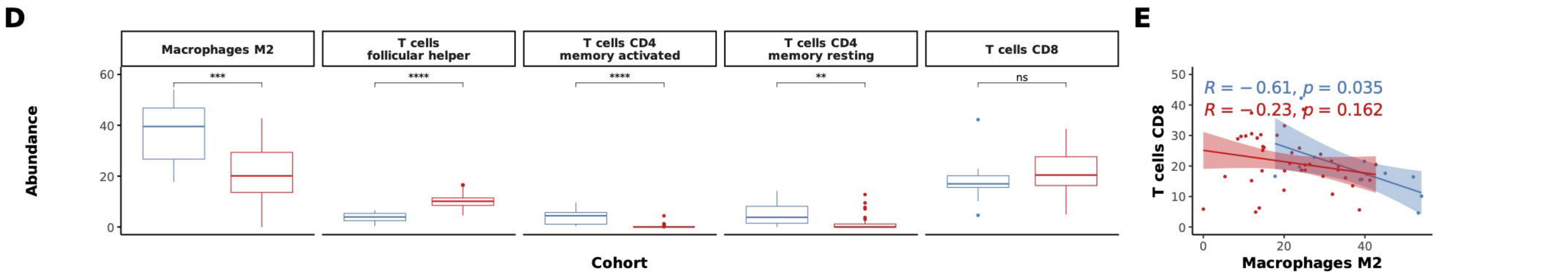
Wang, Q.J., F.J. Jenkins, L.P. Jacobson, Y.-X. Meng, P.E. Pellett, L.A. Kingsley, K.G. Kousoulas, A. Baghian, and C.R. Rinaldo. 2000. CD8+ Cytotoxic T Lymphocyte Responses to Lytic Proteins of Human Herpes Virus 8 in Human Immunodeficiency Virus Type 1—Infected and —Uninfected Individuals. *J. Infect. Dis.* 182:928–932. doi:10.1086/315777.

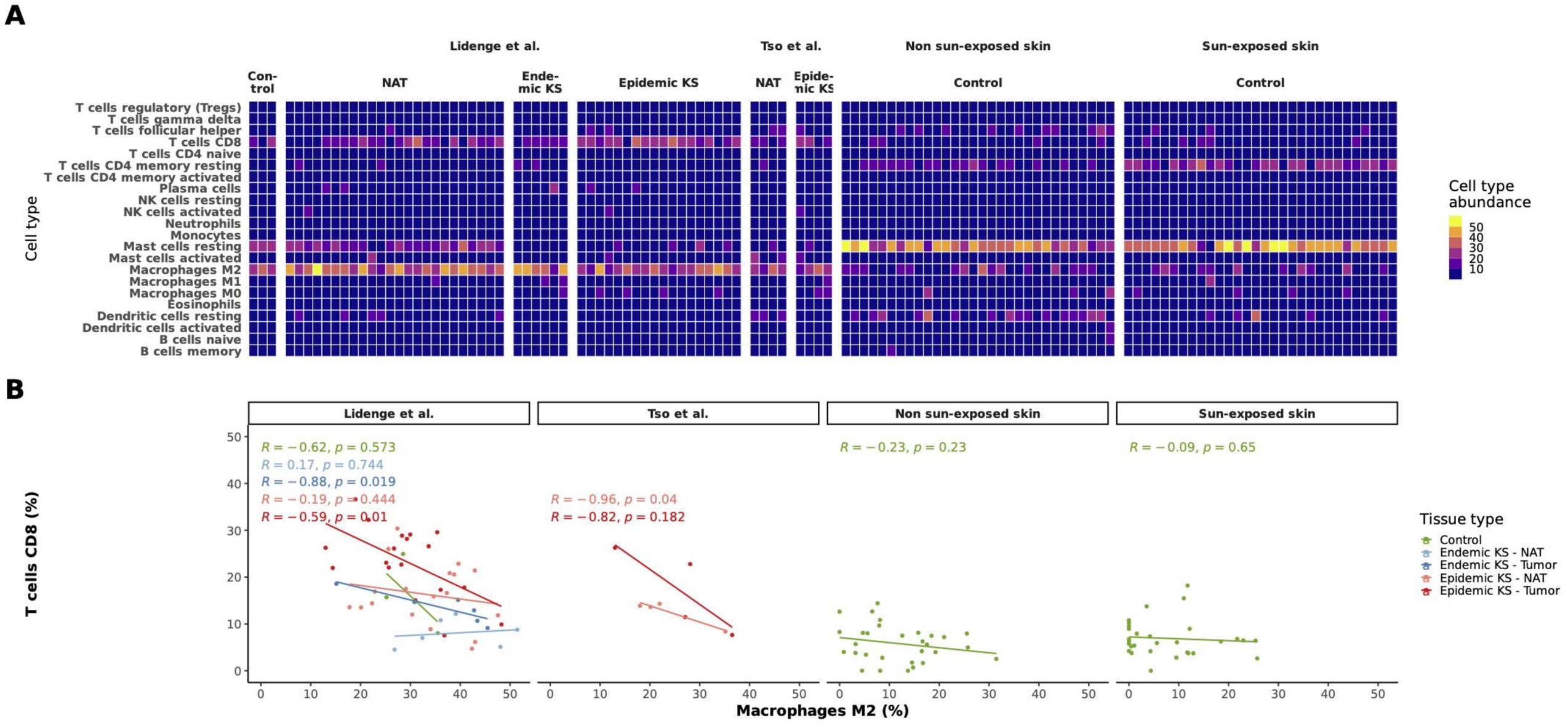
Wilkinson, J., A. Cope, J. Gill, D. Bourboulia, P. Hayes, N. Imami, T. Kubo, A. Marcelin, V. Calvez, R. Weiss, B. Gazzard, C. Boshoff, and F. Gotch. 2002. Identification of Kaposi's Sarcoma-Associated Herpesvirus (KSHV)-Specific Cytotoxic T-Lymphocyte Epitopes and Evaluation of Reconstitution of KSHV-Specific Responses in Human Immunodeficiency Virus Type 1-Infected Patients Receiving Highly Active Antiretroviral Therapy. *J. Virol.* 76:2634–2640. doi:10.1128/jvi.76.6.2634-2640.2002.

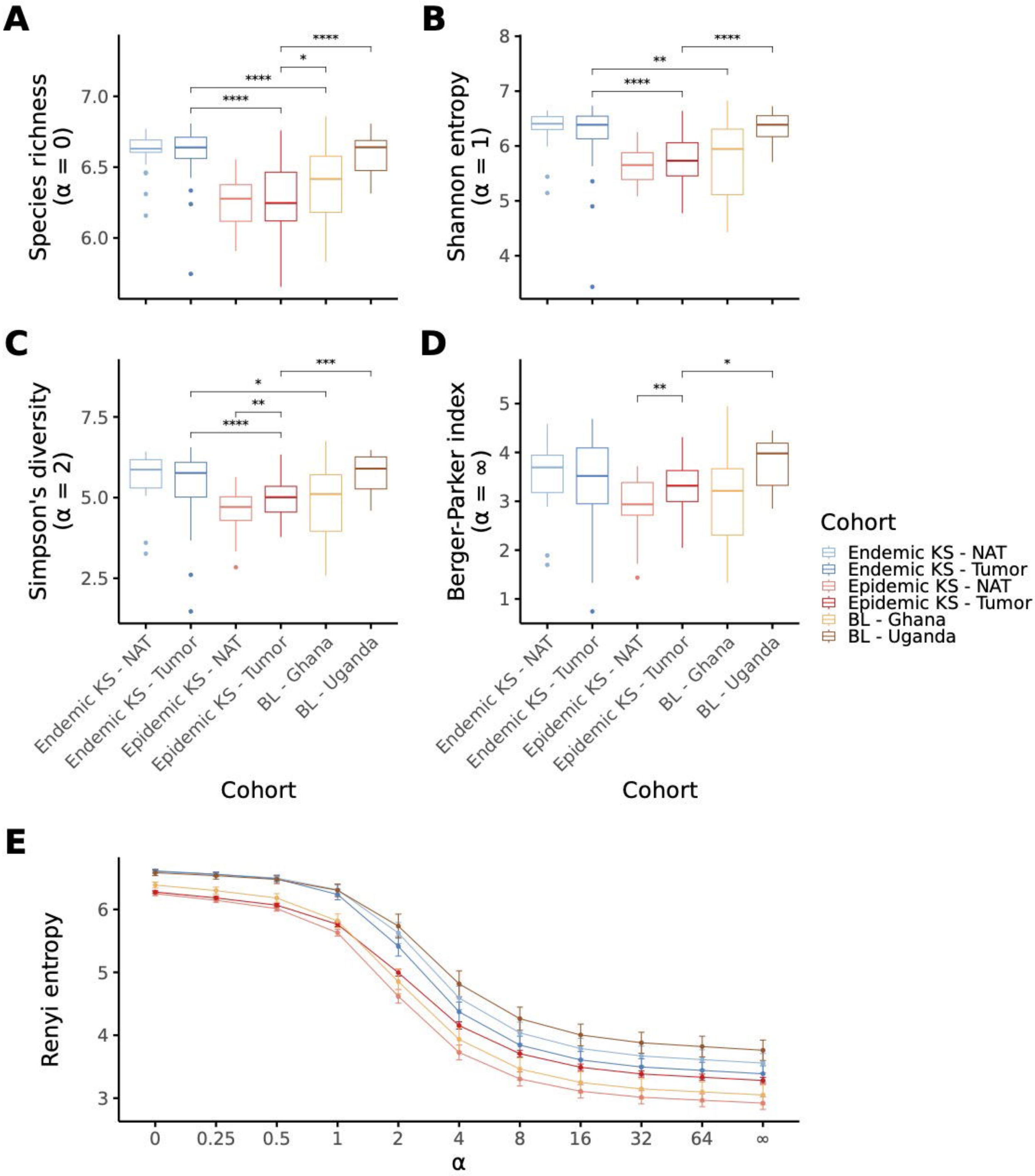
Woodberry, T., T.J. Suscovich, L.M. Henry, J.N. Martin, S. Dollard, P.G. O'Connor, J.K. Davis, D. Osmond, T.-H. Lee, D.H. Kedes, A. Khatri, J. Lee, B.D. Walker, D.T. Scadden, and C. Brander. 2005. Impact of Kaposi Sarcoma–Associated Herpesvirus (KSHV) Burden and HIV Coinfection on the Detection of T Cell Responses to KSHV ORF73 and ORF65 Proteins. *J. Infect. Dis.* 192:622–629. doi:10.1086/432103.

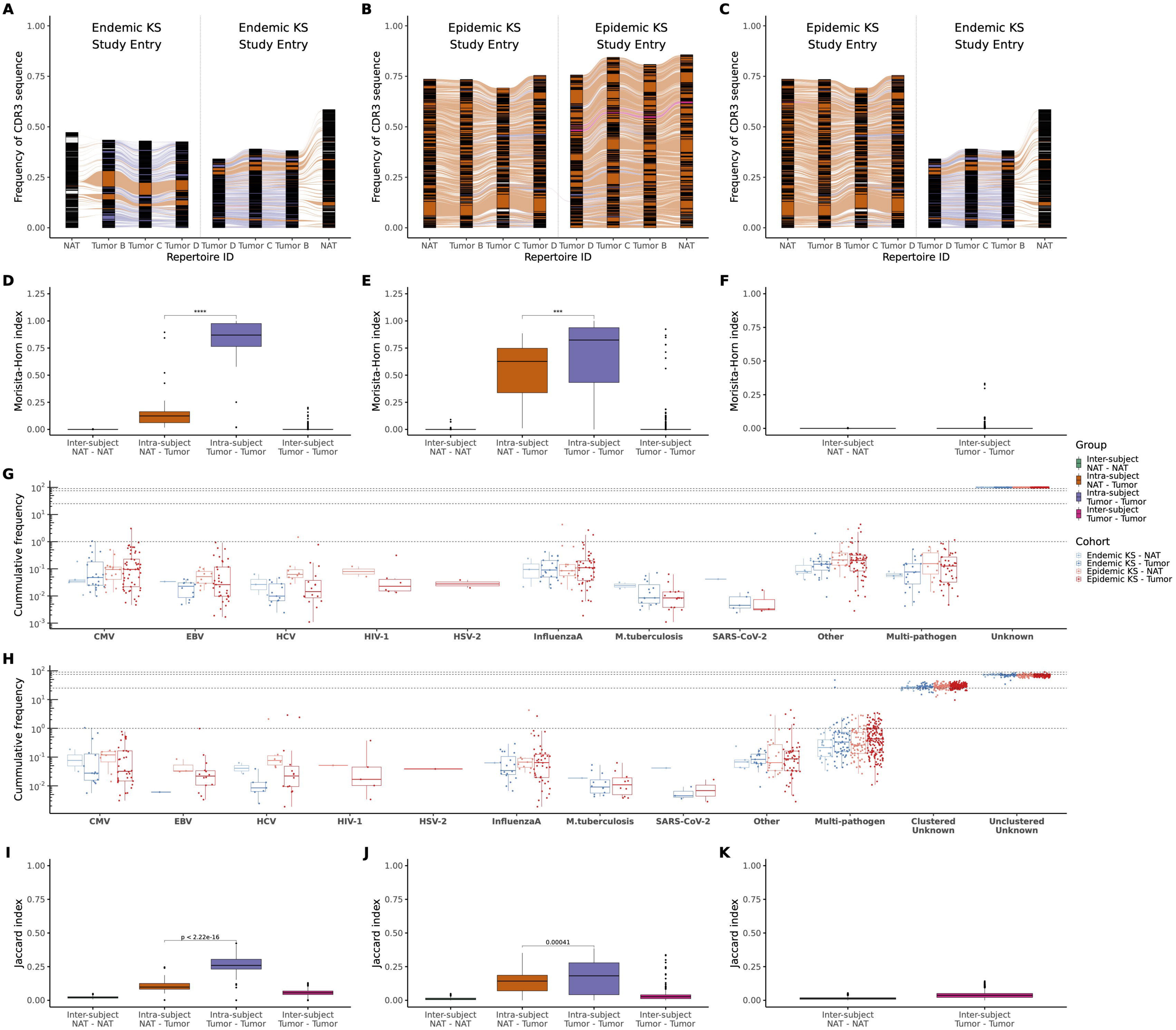
Wu, E., D.G. Coffey, and S. Ravishankar. 2021. LymphoSeq2.

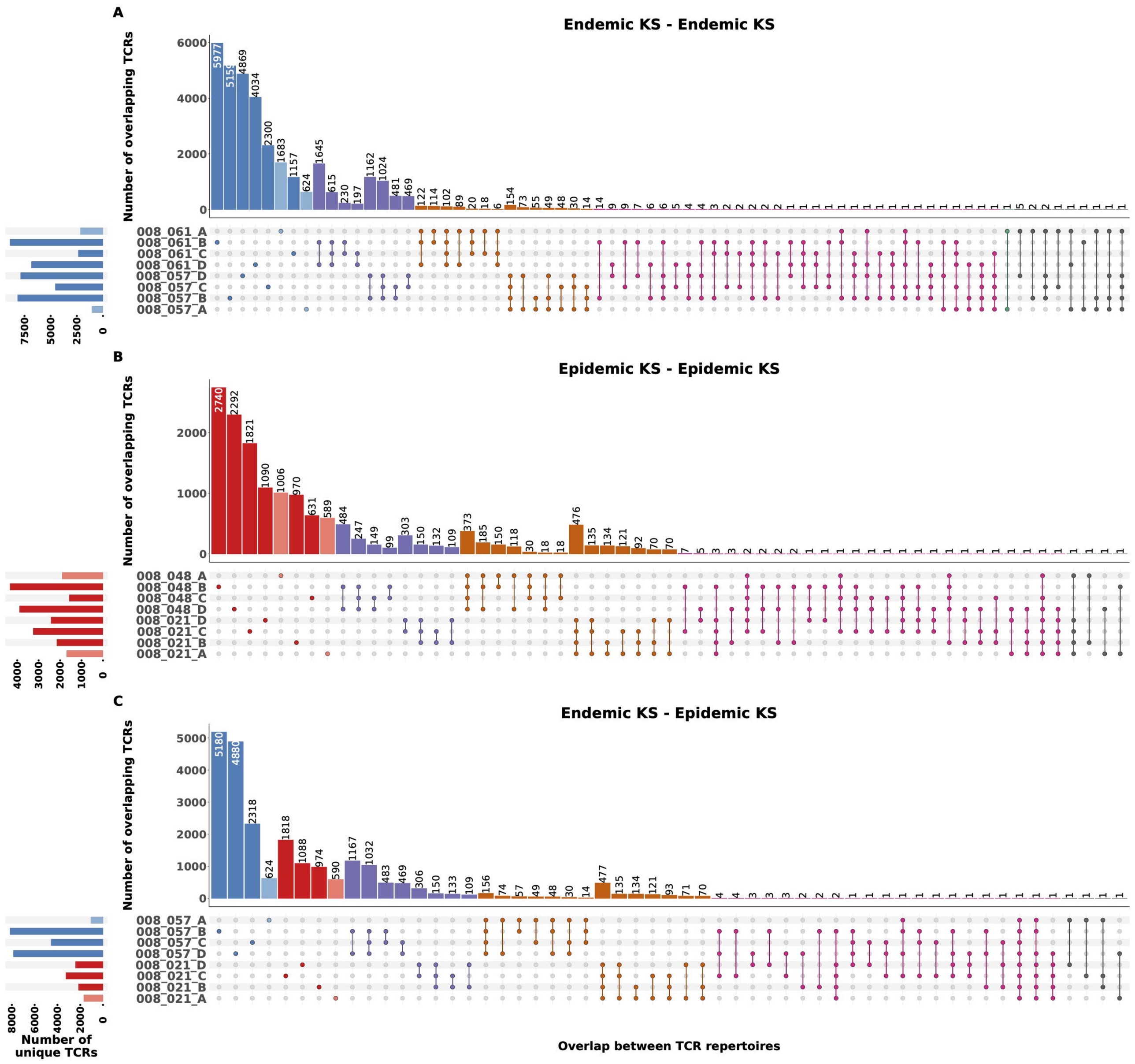

A
KSHV gene expression
La
IE
E1
E2
E3
L4
L5
Sample ID
B
HIV gene expression
Asp
Env
Gag-Pol
Nef
Pr55(Gag)
Rev
Tat
Vif
Vpr
Vpu
C
Immune cell-type composition
Endemic KS
Epidemic KS
B cells memory
B cells naive
Dendritic cells activated
Dendritic cells resting
Eosinophils
Macrophages M0
Macrophages M1
Macrophages M2
Mast cells activated
Mast cells resting
Monocytes
Neutrophils
NK cells activated
NK cells resting
Plasma cells
T cells CD4 memory activated
T cells CD4 memory resting
T cells CD4 naive
T cells CD8
T cells follicular helper
T cells gamma delta
T cells regulatory (Tregs)
TPM
1e+05
1e+04
1e+03
1e+02
1e+01
Cell type abundance
50
40
30
20
10
Cohort
Endemic KS
Epidemic KS

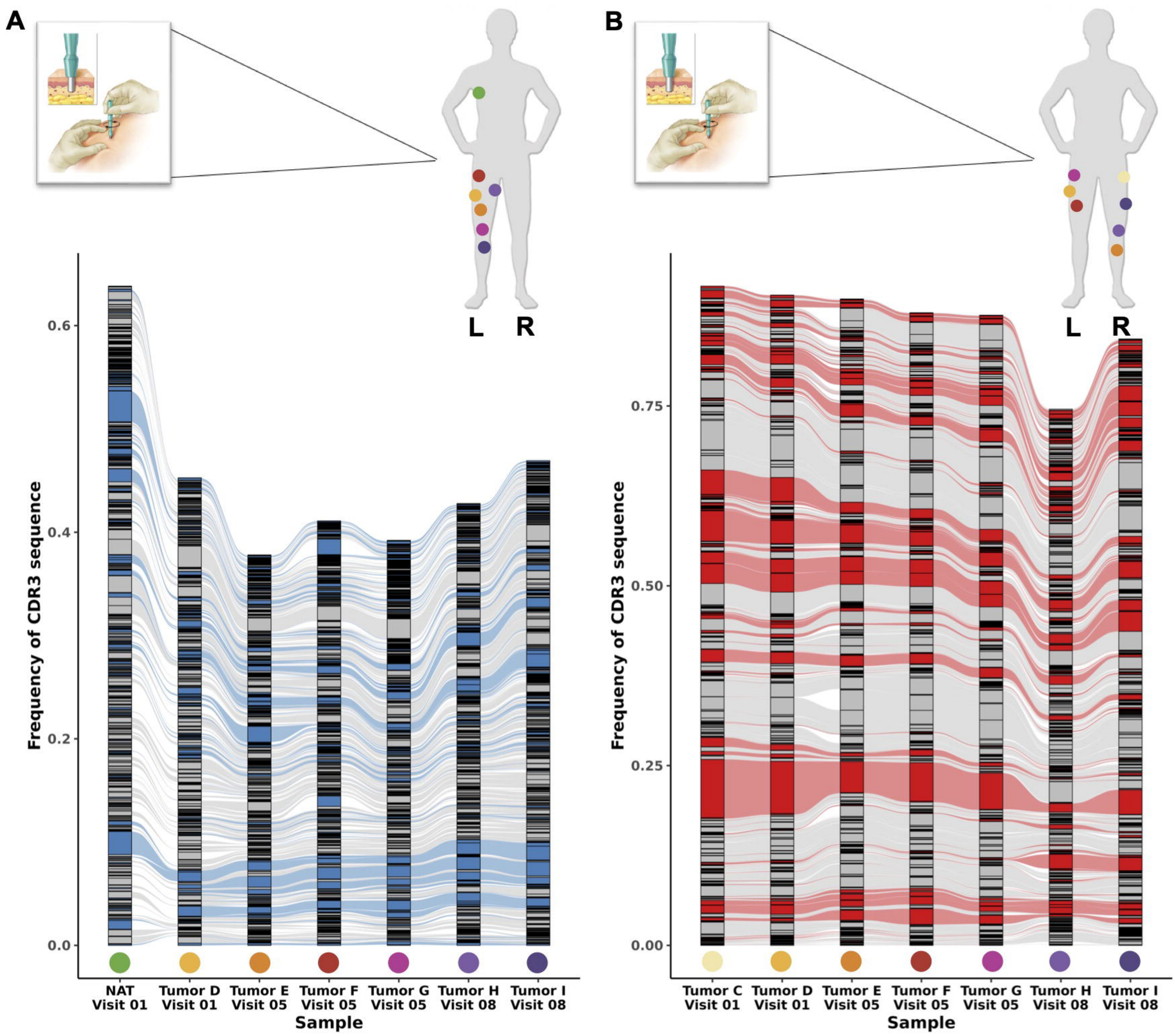

D
Macrophages M2
T cells follicular helper
T cells CD4 memory activated
T cells CD4 memory resting
T cells CD8
***
****
****
**
ns
Abundance
Cohort
E
R = −0.61, p = 0.035
R = −0.23, p = 0.162
T cells CD8
Macrophages M2

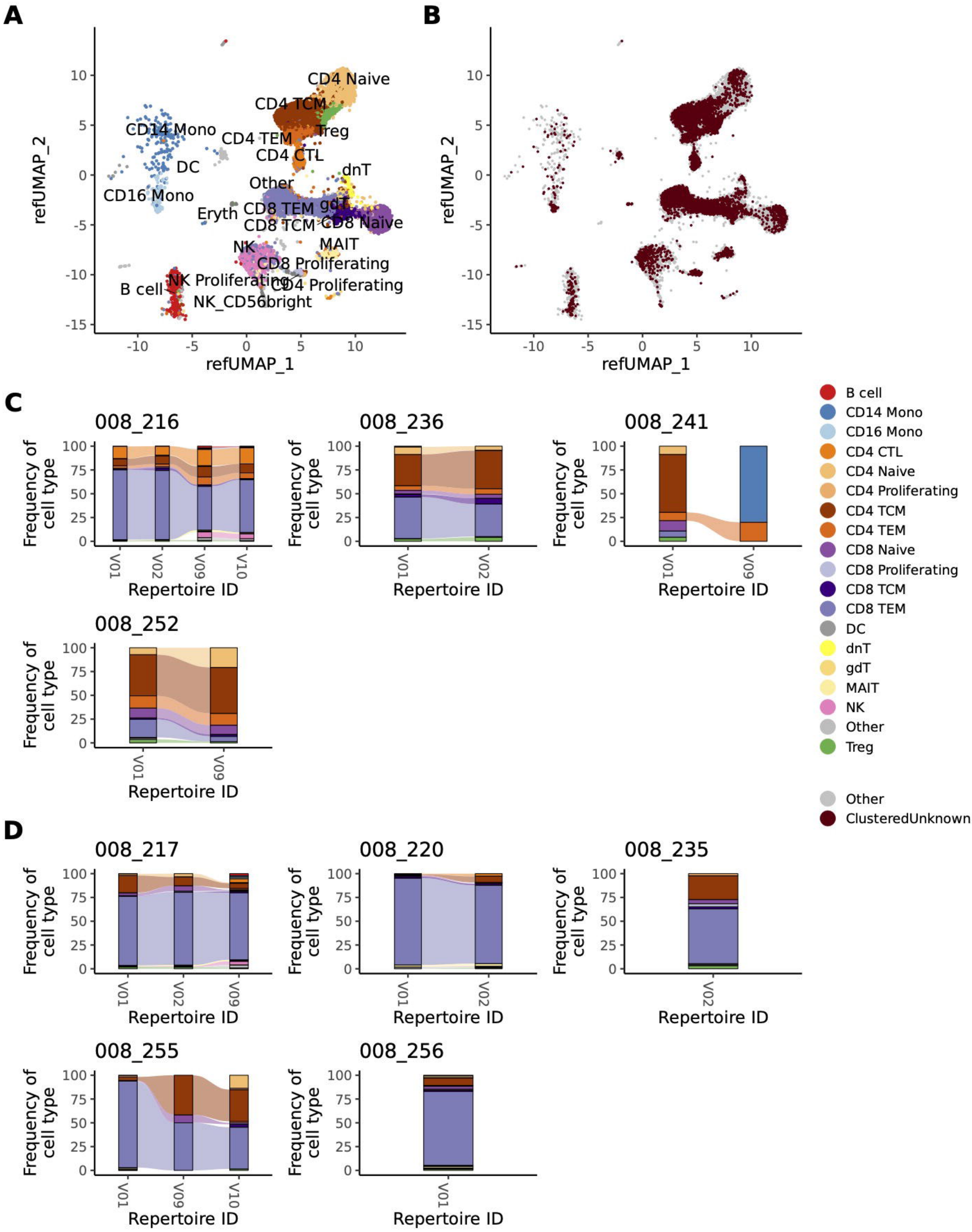

A
Lidenge et al.
Tso et al.
Non sun-exposed skin
Sun-exposed skin
Con-trol
NAT
Ende-mic KS
Epidemic KS
NAT
Epide-mic KS
Control
Control
Cell type
T cells regulatory (Tregs)
T cells gamma delta
T cells follicular helper
T cells CD8
T cells CD4 naive
T cells CD4 memory resting
T cells CD4 memory activated
Plasma cells
NK cells resting
NK cells activated
Neutrophils
Monocytes
Mast cells resting
Mast cells activated
Macrophages M2
Macrophages M1
Macrophages M0
Eosinophils
Dendritic cells resting
Dendritic cells activated
B cells naive
B cells memory
Cell type abundance
50
40
30
20
10
B
Lidenge et al.
Tso et al.
Non sun-exposed skin
Sun-exposed skin
R = −0.62, p = 0.573
R = 0.17, p = 0.744
R = −0.88, p = 0.019
R = −0.19, p = 0.444
R = −0.59, p = 0.01
R = −0.96, p = 0.04
R = −0.82, p = 0.182
R = −0.23, p = 0.23
R = −0.09, p = 0.65
T cells CD8 (%)
Macrophages M2 (%)
Tissue type
Control
Endemic KS - NAT
Endemic KS - Tumor
Epidemic KS - NAT
Epidemic KS - Tumor

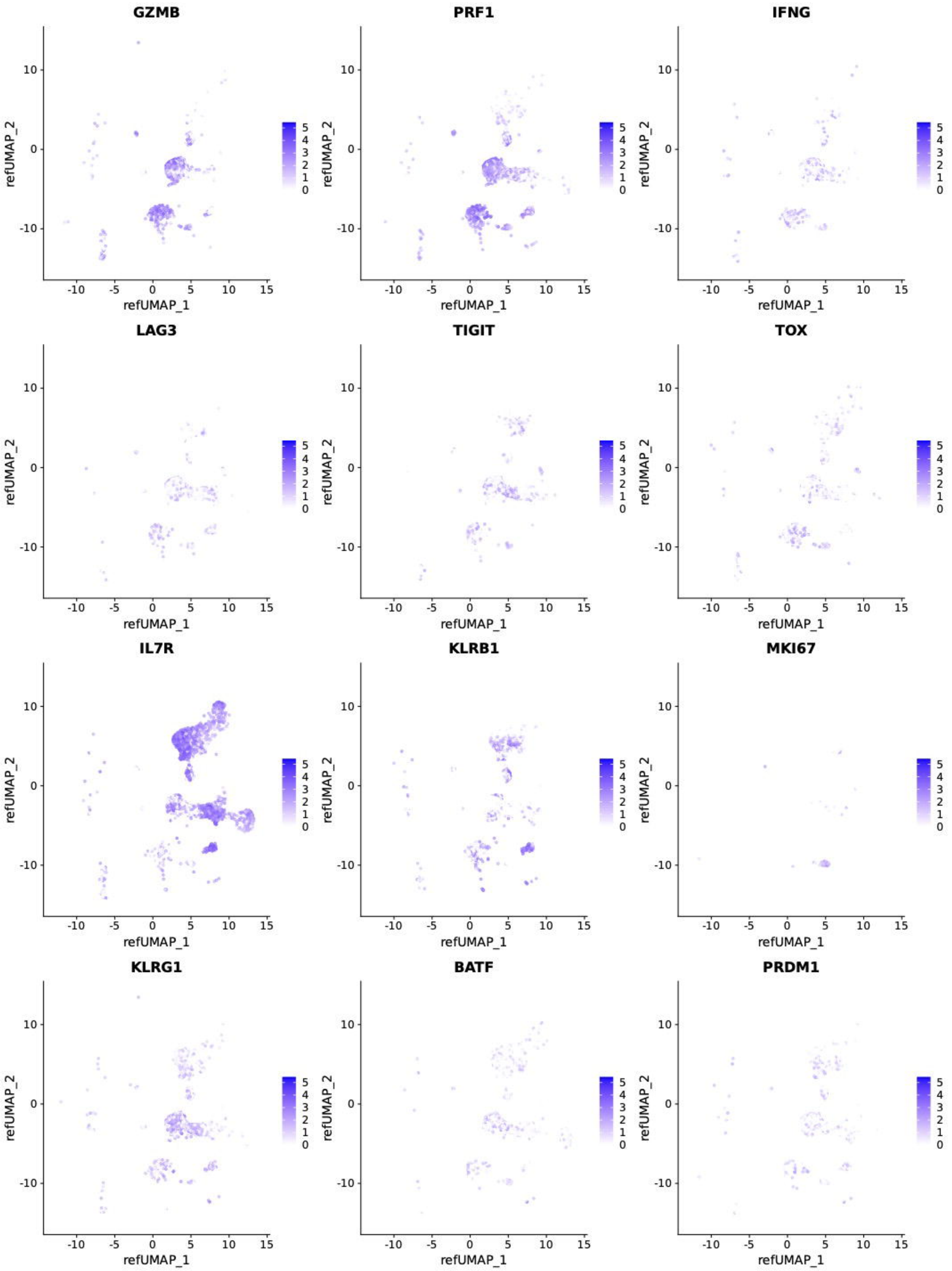

A
Species richness (α = 0)
7.0
6.5
6.0
****
****
*
****
B
Shannon entropy (α = 1)
8
7
6
5
4
****
**
****
C
Simpson's diversity (α = 2)
7.5
5.0
2.5
****
**
*
***
D
Berger-Parker index (α = ∞)
5
4
3
2
1
**
*
Endemic KS - NAT
Endemic KS - Tumor
Epidemic KS - NAT
Epidemic KS - Tumor
BL - Ghana
BL - Uganda
Cohort
Cohort
Cohort
Endemic KS - NAT
Endemic KS - Tumor
Epidemic KS - NAT
Epidemic KS - Tumor
BL - Ghana
BL - Uganda
E
Renyi entropy
6
5
4
3
0
0.25
0.5
1
2
4
8
16
32
64
∞
α

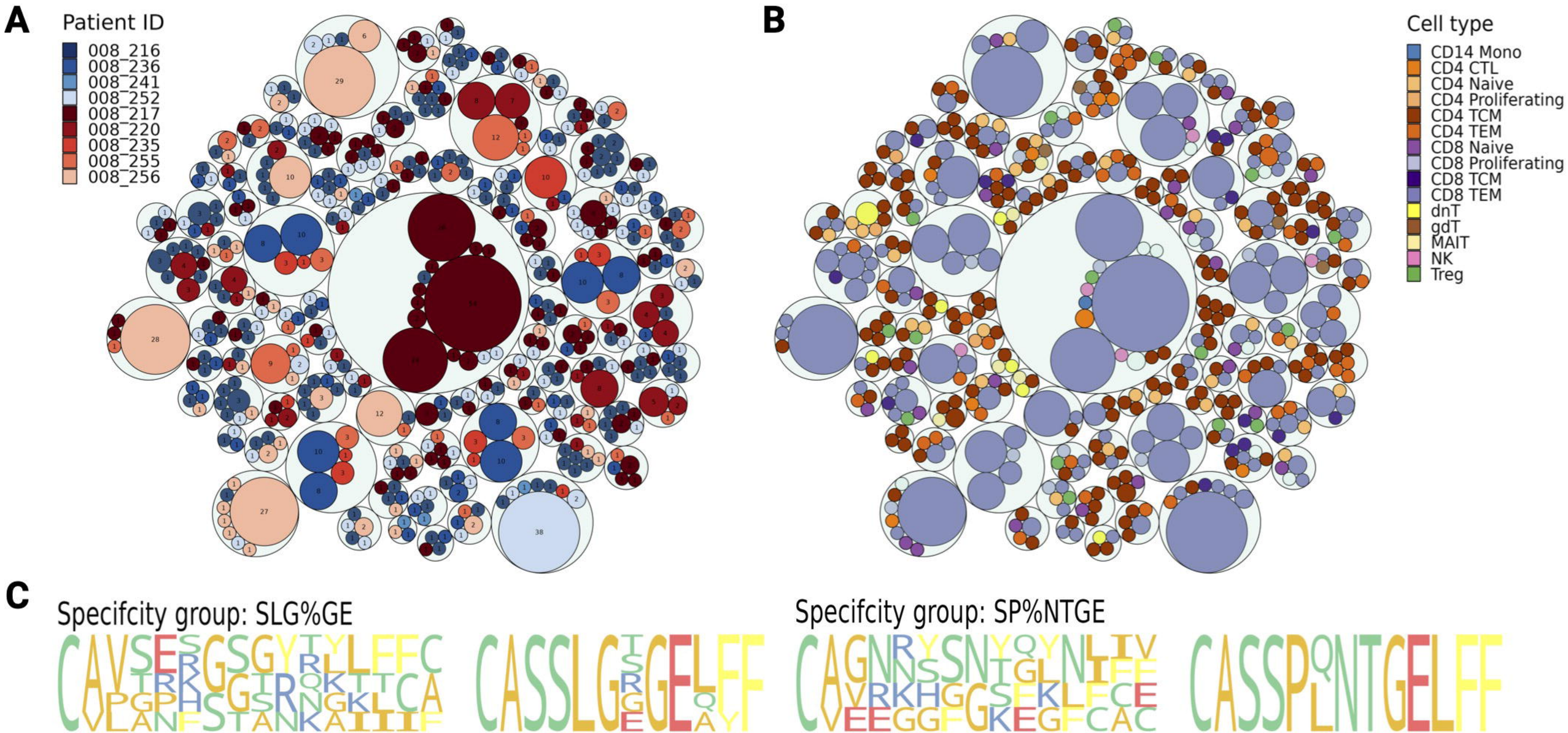

A
Endemic KS Study Entry
Endemic KS Study Entry
Frequency of CDR3 sequence
Repertoire ID
NAT Tumor B Tumor C Tumor D Tumor D Tumor C Tumor B NAT
B
Epidemic KS Study Entry
Epidemic KS Study Entry
C
Epidemic KS Study Entry
Endemic KS Study Entry
D
Morisita-Horn index
****
Inter-subject NAT - NAT
Intra-subject NAT - Tumor
Intra-subject Tumor - Tumor
Inter-subject Tumor - Tumor
E
***
F
G
Cummulative frequency
CMV
EBV
HCV
HIV-1
HSV-2
InfluenzaA
M.tuberculosis
SARS-CoV-2
Other
Multi-pathogen
Unknown
Group
Inter-subject NAT - NAT
Intra-subject NAT - Tumor
Intra-subject Tumor - Tumor
Inter-subject Tumor - Tumor
Cohort
Endemic KS - NAT
Endemic KS - Tumor
Epidemic KS - NAT
Epidemic KS - Tumor
H
Clustered Unknown
Unclustered Unknown
I
Jaccard index
p < 2.22e-16
J
0.00041
K


A
Endemic KS - Endemic KS
Number of overlapping TCRs
008_061_A
008_061_B
008_061_C
008_061_D
008_057_D
008_057_C
008_057_B
008_057_A
B
Epidemic KS - Epidemic KS
Number of overlapping TCRs
008_048_A
008_048_B
008_048_C
008_048_D
008_021_D
008_021_C
008_021_B
008_021_A
C
Endemic KS - Epidemic KS
Number of overlapping TCRs
008_057_A
008_057_B
008_057_C
008_057_D
008_021_D
008_021_C
008_021_B
008_021_A
Number of unique TCRs
Overlap between TCR repertoires


A
L
R
0.6
0.4
0.2
0.0
Frequency of CDR3 sequence
NAT Visit 01
Tumor D Visit 01
Tumor E Visit 05
Tumor F Visit 05
Tumor G Visit 05
Tumor H Visit 08
Tumor I Visit 08
Sample
B
L
R
0.75
0.50
0.25
0.00
Frequency of CDR3 sequence
Tumor C Visit 01
Tumor D Visit 01
Tumor E Visit 05
Tumor F Visit 05
Tumor G Visit 05
Tumor H Visit 08
Tumor I Visit 08
Sample


A
B
C
D
refUMAP_1
refUMAP_2
CD4 Naive
CD4 TCM
CD14 Mono
CD4 TEM
Treg
CD4 CTL
DC
dnT
Other
CD16 Mono
gdT
Eryth
CD8 TEM
CD8 TCM
CD8 Naïve
MAIT
NK
CD8 Proliferating
NK Proliferating
CD4 Proliferating
B cell
NK_CD56bright
008_216
008_236
008_241
008_252
008_217
008_220
008_235
008_255
008_256
Frequency of cell type
Repertoire ID
V01
V02
V09
V10
B cell
CD14 Mono
CD16 Mono
CD4 CTL
CD4 Naive
CD4 Proliferating
CD4 TCM
CD4 TEM
CD8 Naive
CD8 Proliferating
CD8 TCM
CD8 TEM
DC
dnT
gdT
MAIT
NK
Other
Treg
Other
ClusteredUnknown


GZMB
PRF1
IFNG
LAG3
TIGIT
TOX
IL7R
KLRB1
MKI67
KLRG1
BATF
PRDM1
refUMAP_1
refUMAP_2
10
0
-10
-10
-5
0
5
10
15
5
4
3
2
1
0


A
Patient ID
008_216
008_236
008_241
008_252
008_217
008_220
008_235
008_255
008_256
B
Cell type
CD14 Mono
CD4 CTL
CD4 Naive
CD4 Proliferating
CD4 TCM
CD4 TEM
CD8 Naive
CD8 Proliferating
CD8 TCM
CD8 TEM
dnT
gdT
MAIT
NK
Treg
C
Specifcity group: SLG%GE
Specifcity group: SP%NTGE