



OPEN Integrative functional genomics analysis of Kaposi sarcoma cohorts

Ezequiel Lacunza^{1,7}, Valeria Fink^{2,7}, Julián Naipauer^{5,7}, María E. Salas^{1,7}, Ana M. Gun^{2,7}, Mary J. Goldman³, Jingchun Zhu³, Sion Williams^{4,7}, María I. Figueroa^{2,7}, Pedro Cahn^{2,7}, Omar Coso^{5,7}, Ethel Cesarman⁶, Juan C. Ramos^{4,7} & Martín C. Abba^{1,7}✉

Kaposi sarcoma (KS) is an AIDS-defining cancer and a significant global health challenge caused by KS-associated herpesvirus (KSHV). NGS-based approaches have profiled KS lesions in a minimal number of studies compared with other neoplastic diseases. Here we present a compiled and harmonized dataset of 131 KS and non-tumor cutaneous samples in the context of their predicted pathway activities, immune infiltrate, KSHV and HIV gene expression profiles, and their associated clinical data representing patient populations from Argentina, United States (USA), and Sub-Saharan Africa cohorts. RNA-seq data from 9 Argentinian KS lesions were generated and integrated with previously published datasets derived from the USA and sub-Saharan African cohorts from Tanzania, Zambia, and Uganda. An unsupervised analysis of 131 KS-related samples allowed us to identify four KS clusters based on their host and KSHV gene expression profiles, immune infiltrate, and the activity of specific signaling pathways. The compiled RNA-seq profile is shared with the research community through the UCSC Xena browser for further visualization, download, and analysis (<https://kaposi.xenahubs.net/>). These resources will allow biologists without bioinformatics knowledge to explore and correlate the host and viral transcriptome in a curated dataset of different KS RNA-seq-based cohorts, which can lead to novel biological insights and biomarker discovery.

Keywords Kaposi's sarcoma, RNAseq, Transcriptome, Xena browser

Kaposi sarcoma (KS) is a prominent AIDS-defining malignancy, and a critical global health challenge primarily linked to infection with KS-associated herpesvirus (KSHV)^{1–3}. KSHV contains a complex genome comprising over 80 genes, which are regulated to facilitate either latency with minimal viral expression or lytic replication, resulting in the production of new virions. KSHV can infect epithelial cells, endothelial cells, B cells, and, more recently, it has been found to infect neurons and mesenchymal stem cells (MSCs), further highlighting its diverse pathogenic potential^{1–5}.

Among the various forms of KS, the most aggressive variant is observed in individuals with HIV, known as acquired immunodeficiency syndrome-associated KS (AIDS-KS). This form is characterized by widespread dissemination, affecting the skin and visceral organs, including the gastrointestinal tract and lungs⁶. Moreover, in men who have sex with men (MSM), the HIV infection rate is 4.9–10.5%, and in transgender women (TGW), HIV reported prevalence is up to 58%. These two populations have disproportionately higher risks of developing KS caused by KSHV^{7,8}.

NGS-based transcriptomic analyses of KS have revealed significant molecular insights into the pathogenesis of the disease. For example, Tso et al.⁹ reported alterations in glucose and lipid metabolism, with changes in genes associated with metabolic disorder pathways. A recent study further stratified HIV+ KS lesions into two molecular subtypes—one characterized by endothelial and proliferative features and the other enriched in inflammatory transcripts¹⁰. Additionally, profiling KS lesions from different anatomical sites has revealed variations in immune composition, angiogenic factors, and KSHV gene expression, underscoring the role of the tumor microenvironment in disease progression¹¹. Unsupervised clustering of KSHV gene expression has also identified distinct tumor groups with varying latent and lytic viral profiles^{5,9,12–14}. Furthermore, morphologically distinct KS tumors from the same individual exhibit similar KSHV gene expression patterns, suggesting that viral activation is influenced by the local microenvironment and host immune response¹⁴. This molecular

¹CINIBA, Facultad de Ciencias Médicas, Universidad Nacional de La Plata, La Plata, Argentina. ²Dirección de Investigaciones, Fundación Huésped, Buenos Aires, Argentina. ³UC Santa Cruz Genomics Institute, University of California, Santa Cruz, CA, USA. ⁴Sylvester Comprehensive Cancer Center, University of Miami, Miami, FL, USA. ⁵Instituto de Fisiología, Biología Molecular y Neurociencias, Buenos Aires, Argentina. ⁶Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, NY, USA. ⁷UM-CFAR/SCCC Argentina Consortium for Research and Training in Virally Induced AIDS-Malignancies, Miami, USA. ✉email: mcabba@gmail.com

heterogeneity emphasizes the need for further studies to develop predictive biomarkers and optimize therapeutic strategies for KS.

Most of these RNAseq studies come from subjects in sub-Saharan Africa where KSHV seroprevalence is notably high, with over 50% of the population exhibiting detectable anti-KSHV antibodies^{15,16}. This raises the need for a broader integrative analysis including diverse cohorts to capture the global spectrum of KS pathology.

This study aimed to integrate RNAseq-based gene expression profiles from four distinct cohorts of KS covering North America, South American, and sub-Saharan Africa to enhance our understanding of the disease. Additionally, we intend to share these valuable datasets with the research community through the UCSC Xena resource, promoting further exploration and discovery in the field of KS research.

Materials and methods

Sample collection and participants

Nine KS skin lesions were collected within an Argentinian study that included men at birth with or without HIV, recruited at Fundación Huésped and Hospitals network as part of our collaborative U54 consortium (Miami CFAR SCCC—Argentina Consortium for Research and Training in virally induced AIDS-Malignancies). Participants were >18 years old (median: 42 years; range: 34–62 years).

RNA sequencing

Total RNA was isolated from KS samples preserved in RNeasy Lysis Buffer using the miRNeasy Tissue/Cells Advanced Kits (Qiagen) following standard manufacturer's protocol. RNA concentration and integrity were measured on an Agilent 2100 Bioanalyzer (Agilent Technologies). RNA samples with RNA integrity number (RIN) over 5 were considered for RNA sequencing. The RNA samples were processed for directional RNA-seq library construction using the Illumina Total RNA Prep with Ribo-Zero Plus library preparation kit according to the manufacturer's protocol. We performed 101 nt paired-end sequencing using an Illumina Novaseq 6000 platform and obtained ~60 million clusters per sample with 92% >Q30. The RNAseq raw data has been submitted to NCBI GEO database with accession number GSE271303.

For the three public cohorts included in this study, RNA extraction and library preparation methods were obtained from their associated publications or public repositories. In all cases, tissues were preserved in RNeasy Lysis Buffer prior to processing. GSE147704 and SRP486827 used the miRNeasy and RNeasy Mini Kits (Qiagen), respectively, while GSE241095 used the Direct-zol RNA Miniprep Kit (Zymo) following TRIzol lysis. Library preparation was performed using the TruSeq RNA Library Prep Kit v2 (Illumina) for GSE147704 and SRP486827, and the NEBNext Ultra Low Input Total RNA Library Prep Kit with rRNA Depletion (NEB) for GSE241095.

Ethics approval and consent to participate

The protocol and study-related documents were approved by the institutional review board (Comité de Bioética, Fundación Huésped). Informed consent was obtained from all participants prior to their involvement in any study procedures. All methods were conducted in full compliance with the applicable guidelines and regulations.

RNAseq data pre-processing and integration of Kaposi's sarcoma cohorts

RNAseq raw data were retrieved from GEO/SRA using the SRA Toolkit (<https://github.com/ncbi/sra-tools>) from GSE147704 (Tanzania & Zambia cohort: KS = 24, Control = 24, and Normal = 3), GSE241095 (USA cohort: KS = 10 and Control = 10) and SRP486827 (Uganda cohort: KS = 51). The raw short-read sequences generated from the Argentinian KS samples and retrieved from GEO/SRA were quality-checked and trimmed to remove adapters and low-quality bases using the Rfastp R/Bioconductor package. The preprocessed reads were then aligned and mapped to the human genome reference GRCh38 using the Subread aligner algorithm provided by the Rsubread R/Bioconductor package. The remaining non-human reads were aligned to the KSHV reference sequence NC_009333. The KSHV reference genome used in our study for read sequence alignment and mapping belongs to the Human herpesvirus 8 strain GK18. It provides a valuable complete genome for global transcriptome analysis, but it does not capture the full extent of KSHV's genetic diversity, including the highly diverse K1 and K15 subtypes present in sub-Saharan populations. The aligned reads (BAM files) from each sample were used to calculate gene expression abundance at the whole-genome level using the featureCounts function provided by the Rsubread package. To assess the Argentine cohort's impact, we used the RnaSeqSampleSize R package to estimate statistical power based on sample size, effect size, sequencing depth, and data dispersion.

The raw read counts for each cohort were integrated, and batch effects were diagnosed using the BatchQC R/Bioconductor package¹⁷, followed by batch effects adjustment with the ComBat-seq algorithm from the "sva" R/Bioconductor package¹⁸. Briefly, ComBat-seq algorithm implements a negative binomial regression model to estimate batch effects based on the count matrix obtained for each RNA-seq study allowing us to include the condition variable (control or KS) as covariates in the regression model behind this method to account for the impact of batch and retaining biological signal.

KS and KSHV transcriptome analysis of the compiled dataset

The transcriptome profile of the compiled batch-effect adjusted KS dataset consists of 131 KS and non-tumor cutaneous samples and their associated metadata. To identify differentially expressed genes between KS and controls or between KS groups (e.g.: from HIV+ vs HIV- or with or without ART, etc.), we computed fold changes and adjusted p-values using the edgeR R/Bioconductor package based on the normalized log2-based count per million values. Genes showing a log-fold change greater than 1 and an adjusted p-value below 0.05 were considered significantly differentially expressed. Functional enrichment analysis and Gene Set Variation Analysis (GSVA) of differentially expressed genes were performed with the clusterProfiler and GSVA R packages. Tumor immune cell infiltration scores were estimated with the ABIS algorithm and the MCP-counter

tool from the `immunedeconv` (<https://github.com/omnideconv/immunedeconv>) and `MCPcounter` (<https://github.com/ebecht/MCPcounter>) R packages, respectively, on normalized count matrices. ABIS performs absolute deconvolution on RNA-Seq data to estimate the proportions of human immune cell types¹⁹. MCP-counter estimates immune and stromal cell populations based on transcriptomic markers that are strongly, specifically, and stably expressed in unique cell types, with scores proportional to the abundance of each population in the tumor, enabling intersample comparison and large cohort analyses²⁰. Unsupervised hierarchical clustering analysis and heatmaps representations were performed with the MultiExperimentViewer (MeV 4.9.0) software. To determine the optimal number of clusters, we applied Principal Component Analysis (PCA) alongside the `NbClust` package, which evaluates the number of clusters using 30 different indices (<https://github.com/cran/NbClust/blob/master/R/NbClust.R>).

The relationship between a categorical and a quantitative variable was assessed using the Wilcoxon rank-sum test or the Kruskal–Wallis test. Pearson correlation was used for two quantitative variables. P values were corrected for multiple testing using the Bonferroni or Benjamini–Hochberg methods.

Data sharing and visualization through a UCSC Xena hub

UCSC Xena is a high-performance visualization and analysis tool for both large public repositories and private datasets²¹. UCSC Xena has two components: the front-end Xena Browser and the back-end Xena Hubs. We have deployed a public Xena Hub (called KS-omics) to host and share the functional genomics data of the compiled KS dataset with the research community (<https://kaposi.xenahubs.net/>). Briefly, Xena Browser allows biologists without bioinformatics knowledge to explore data with a variety of visualizations and analytic tools.

Results and discussion

Integrative analysis of host and viral gene expression profiles in KS using multi-cohort RNA sequencing

To perform an integrative analysis of host and viral gene expression profiles in KS, we combined 94 KS with 37 non-tumor cutaneous cases obtained from four cohorts, one of them generated by us and three available in a public database (Fig. 1A). Clinical variables considered for the study included age, ART treatment, sex at birth, HIV status, origin (endemic/epidemic), and KS morphotype. To generate a homogeneous, uniformly curated, and preprocessed dataset, the RNAseq raw data was directly retrieved from GEO/SRA using the SRA Toolkit and equally preprocessed by our bioinformatics pipeline. Additionally, raw read count matrices from each cohort were integrated. Variation analysis (Fig. S1) showed that batch effects explained a larger proportion of variation than the condition (control or KS), highlighting the need for batch correction. Batch effects were then adjusted using the ComBat-seq algorithm (Fig. S1). Importantly, the transcriptomic profile of the four independent KS cohorts were generated using Illumina sequencing platforms (Table 1).

Figure 1B shows the scatter plot of cases projected on the first two dimensions in unadjusted data, and in data adjusted by ComBat-seq. We observed a strong batch effect in the unadjusted data, which was well addressed by ComBat-seq. A favorable adjustment would pool control/non-tumor cutaneous samples from the cohorts (batches), while keeping all KS lesions separated from the controls and from each other. In the multidimensional plot of ComBat-seq adjusted data, we observed the expected pattern of data if there were no batch effects, in which the control/normal cases are clustered together, while the KS cases from four cohorts are scattered at different locations. These results suggest a successful adjustment of batch effect from ComBat-seq.

UCSC Xena is a high-performance visualization and analysis tool for both large public repositories and private datasets²¹. In this sense, we have implemented a public cloud-based backend Xena Hubs hosting the human and KSHV batch effect adjusted gene expression, immune and pathway activity profiles as well as their associated phenotypic data obtained from the four KS cohorts (<https://kaposi.xenahubs.net/>). The frontend Xena Browser provides a wide variety of visualizations and analyses including scatter plots, bar graphs, statistical tests, genomic signatures, as well as a unique Visual Spreadsheet view (Fig. 1C). The Xena Visual Spreadsheet was designed to enable and enhance integration across diverse data modalities, providing researchers with a more biologically complete understanding of genomic events and tumor biology. The UCSC Xena browser enables users to explore functional genomic datasets for correlations between genomic and phenotypic variables. It also allows for differential expression analysis between user-defined sample groups (Fig. 1D) and functional enrichment analysis of differentially expressed genes (Fig. 1E).

Unsupervised clustering reveals distinct transcriptomic and immune profiles in KS lesions

To characterize the transcriptomic, immune, and functional profiles of the samples, we employed an unsupervised clustering approach. We defined clusters based on the entire transcriptome (host and KSHV). This approach allows us to dissect the contributions of both host and viral factors to the overall expression landscape, providing a comprehensive understanding of the interplay between the host's immune response and viral activity in KS.

The unsupervised analysis revealed a clear distinction between KS lesions and non-tumor controls (Fig. 2A), underscoring the impact of KSHV on the host transcriptome⁹. To further validate the optimal number of clusters, we applied the `NbClust` package. The majority rule suggested either 2 or 4 clusters as the most likely options. Considering this, along with the structure observed in the dendrogram and PCA analysis, we determined that 4 clusters were the most appropriate choice for our data (Fig. S2). We identified at least three distinct groups of lesions, designated as clusters 1 (C1), 3 (C3), and 4 (C4). Uninvolved skin from KS patients and skin from non-KS subjects were classified within the control cluster (C2), which comprised two subclusters and included two KS cases (Fig. 2A). Cluster C1 included most endemic tumors, along with some epidemic cases, while clusters C3 and C4 were primarily composed of epidemic lesions ($p < 0.001$, Data S1). Importantly, no significant associations were observed between lesion clusters and HIV status or ART treatment (Data S1).

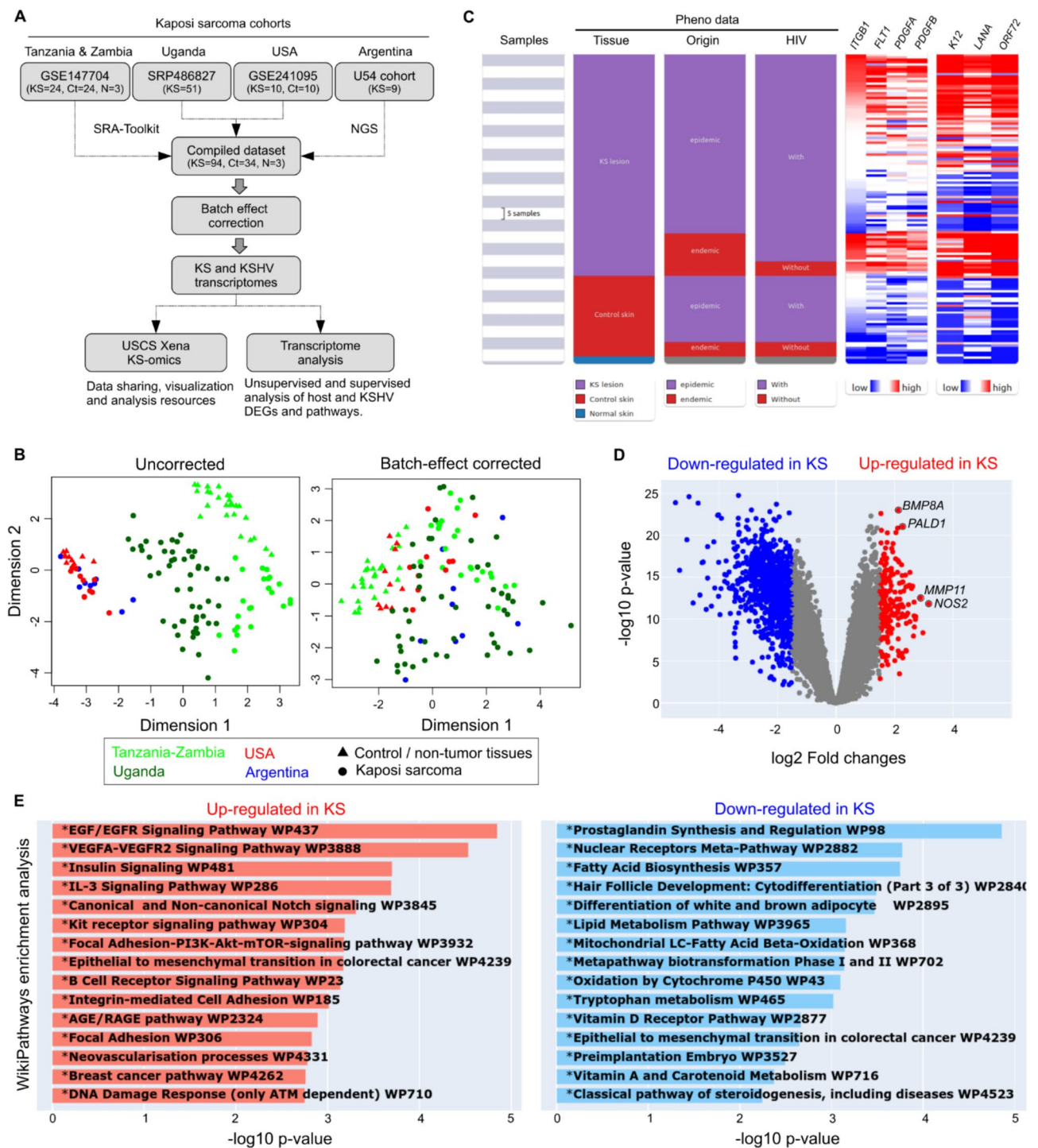


Fig. 1. Compiled KS dataset of 131 KS and matched non-tumor tissues. (A) Diagram of the Strategy followed to obtain the compiled KS dataset. (B) Multidimensional scaling plot of the non-adjusted (left) and batch-effect adjusted (right) gene expression data of KS (circle) and matched non-tumor tissues (triangle) among cohorts. (C) An example Xena Browser Visual Spreadsheet examining the gene expression profiles of selected humans (*ITGB1*, *FLT1*, *PDGFA* and *PDGFB*) and KSHV genes (*K12*, *LANA*, *ORF72*) among the compiled KS dataset in the context of their phenotypic data (<https://kaposi.xenahubs.net/>). Gene expression data is colored red to blue for high to low expression respectively. UCSC Xena browser provides analytic tools to correlate the human and KSHV gene expression levels as well as to identify differentially expressed genes-based groups defined by the user (e.g. KS lesions vs. control skin). (D) Representative volcano plot of differentially expressed genes detected between KS and control samples using UCSC Xena browser. (E) Functional enrichment analysis of differentially expressed genes with Xena.

Characteristics	Tanzania/Zambia GSE147704	Uganda SRP486827	USA GSE241095	Argentina GSE271303
# of Cases	24 KS	51 KS	10 KS	9 KS
	24 Controls		10 Controls	
	3 Normal skin			
Age (median, IQR)	33 (29, 42)	35 (26, 46)	43 (34, 49)	42 (34, 50)
Sex at birth				
Male	16	44	10	9
Female	8	5	0	0
n/a	–	2	–	–
Race				
White	0	0	2 (20%)	0
Black	24 (100%)	51 (100%)	4 (40%)	0
Hispanic	0	0	4 (40%)	9 (100)
HIV characteristics				
HIV co-infection	18 (75%)	51 (100%)	9 (90%)	9 (90%)
CD4 T-cell count (cells/ul)	na	na	39	144
HIV Viral load (cps/mL)	12000	na	74000	125000
On ART at biopsy	18 (75%)	0 (0%)	10 (100%)	3 (33%)

Table 1. Characteristics of the study subjects included in each KS cohort.

However, a notable association with KS morphotype was found, with cluster C4 exhibiting a higher percentage of macular lesions ($p < 0.01$; Data S1).

We defined the immune fraction profile using the ABIS algorithm¹⁹, which demonstrated a significant immune infiltrate enrichment in the lesions compared to the controls (Fig. 2B). Clusters C1 and C3 showed the highest percentage of immune fractions, indicating a greater infiltrate compared to clusters C2 and C4. Cluster C1 was characterized by a significant enrichment of memory CD4 T lymphocytes, whereas naive CD4 T cells were predominant in the control group (C2) (Fig. 3). In contrast, cluster C3 exhibited a notable enrichment of plasmablasts and memory B cells (Fig. 3). Additionally, all three lesion clusters showed enrichment in CD8 memory T cells compared to the control group (Fig. 3).

To strengthen the analysis, we examined the composition of the sample microenvironment using the microenvironment cell populations (MCP)-counter method²⁰. The cell composition varied significantly between clusters (Fig. 3). Cluster C1 displayed elevated expression of genes specific to immune populations, including T cells, CD8+ T cells, natural killer (NK) cells, and cytotoxic lymphocytes. However, it was primarily characterized by a significant upregulation of endothelial cell-related genes (Fig. 3). Cluster C3 showed elevated expression of most immune cell populations, with B lineage signatures and T cells as key determinants. Cluster C4 displayed a generally low immune profile, similar to the control cluster (C2), but differed by a higher expression of endothelial cell markers (Fig. 3).

These results complemented the findings by Tso et al., who reported a significant infiltrate of immune cells in the lesions, primarily of B cells, although T cells were not significantly present. However, they predicted the activation of numerous chemokines that play an important role in the recruitment of T cells⁹. Additionally, Lidenge et al. have shown that, despite the upregulation of chemokines like CxCL-9, immune cells, including CD4+ T cells, CD8+ T cells, and NK cells, are not co-localized with KSHV-infected regions^{12,13}. This discrepancy highlights the differences between transcriptomic-based immune profiling and protein-level detection, suggesting that immune cells may be recruited to the tumor microenvironment but may not be localized to KSHV-infected areas. In addition, the marginal tissues captured in biopsies likely contain immune populations that are present but infiltrate the tumor at low frequencies.

Given the increasing relevance of immune checkpoint inhibitors (ICIs) in the treatment of KS, we evaluated the transcriptomic expression levels of a set of ICIs. A distinct pattern of expression levels was observed, correlating with the infiltrate (Fig. 2C). In this context, clusters C1 and C3 were the most enriched in these genes. These results, while needing further validation, along with the immune profile of the clusters, are relevant in the context of the KS microenvironment and immune checkpoint ICIs immunotherapies. They are also consistent with previous studies. For instance, Petitprez et al.²² established an immune-based classification of sarcoma immune classes (SICs) and demonstrated that the immune-high group, characterized by the presence of B cell-rich lineage, exhibited a strong response to PD1 immunotherapy. The tumor microenvironment in KS is known to be composed of activated B cells and tumor-associated macrophages²³. More recently, it has been shown that KS, along with other sarcoma subtypes, displayed the highest response rates to ICIs and longest survival, with PD-L1 expression of $\geq 1\%$ being associated with increased response to ICIs²⁴.

Pathways activity analysis

The pathway activity analysis based on Gene Ontology (GO) and Hallmarks revealed a reduction in epithelial differentiation, including keratinocyte differentiation and epithelial-mesenchymal transition (EMT), in lesions compared to controls (Fig. 2D). Additionally, decreases in metabolic processes such as lipid metabolism,



Fig. 2. Transcriptomic, Immune and Functional Profiling of Kaposi Sarcoma Lesions. **(A)** Clustering results reveal distinct groups of KS lesions, with endemic lesions primarily in Cluster C1, while epidemic lesions are distributed across the three clusters. **(B)** Immune profiling using the ABIS algorithm demonstrates a higher immune cell infiltrate in KS lesions compared to controls. **(C)** Transcriptomic expression levels of immune checkpoint inhibitors (ICIs) indicate that Clusters C1 and C3 exhibit the highest expression levels. **(D)** Pathway activity analysis using Gene Ontology reveals reduced epithelial differentiation and increased immune activity in KS lesions, particularly in Clusters C1 and C3. Cluster C4 shows lower immune activity but a strong innate antiviral response. **(E)** Pathway activity analysis using Hallmarks reveals reduced metabolic processes in KS lesions compared to controls, along with increased proliferative activity, angiogenesis, and PI3K/Akt/mTOR signaling, particularly in Clusters C1 and C3.

cholesterol homeostasis, and response to estrogen were observed (Fig. 2E). This finding aligns with previous studies demonstrating that KSHV plays a significant role in cellular transformation and metabolic reprogramming, both of which are key factors in the pathogenesis of KS^{9,25}. Moreover, the impaired EMT is particularly important in KS, as it promotes the formation of new blood vessels and supports tumor growth, emphasizing the complex relationship between these processes in the disease^{4,5}.

Furthermore, the analysis revealed higher levels of immune pathways—such as adaptive immune responses, T cell differentiation, and B cell differentiation—in the lesions of clusters C1 and C3 compared to cluster C4. KS lesions in cluster C3 exhibited more pronounced humoral immune related pathways than those in cluster C1, which corresponds with its higher proportion of B cells (Figs. 2D, 3). Since KSHV has a known tropism for B cells and can infect them, leading to lymphoproliferative lesions, cluster C3, enriched in B cell lineages, may reflect this interaction^{26–28}. However, the transcriptomic data does not clarify whether these B cells are directly infected by KSHV or are merely part of the recruited immune response.

Lesions in cluster C4 displayed a pattern more similar to non-tumor controls, showing lower enrichment levels across most pathways compared to clusters C1 and C3 (Fig. 2D). However, these lesions exhibited enrichment in pathways related to innate antiviral responses, whereas tumors in clusters C1 and C3 were associated with broader and more robust immune responses (Fig. 2D).

Cluster C1 exhibited lesions with a molecular pattern of pathway activation characteristic of KS, marked by increased proliferative activity (e.g., cell cycle, mitotic spindle) and activation of the PI3K/Akt/mTOR and IL6/JAK/STAT3 pathways (Fig. 2D,E). Both clusters C1 and C3 also displayed angiogenic activity (Fig. 2D,E). These

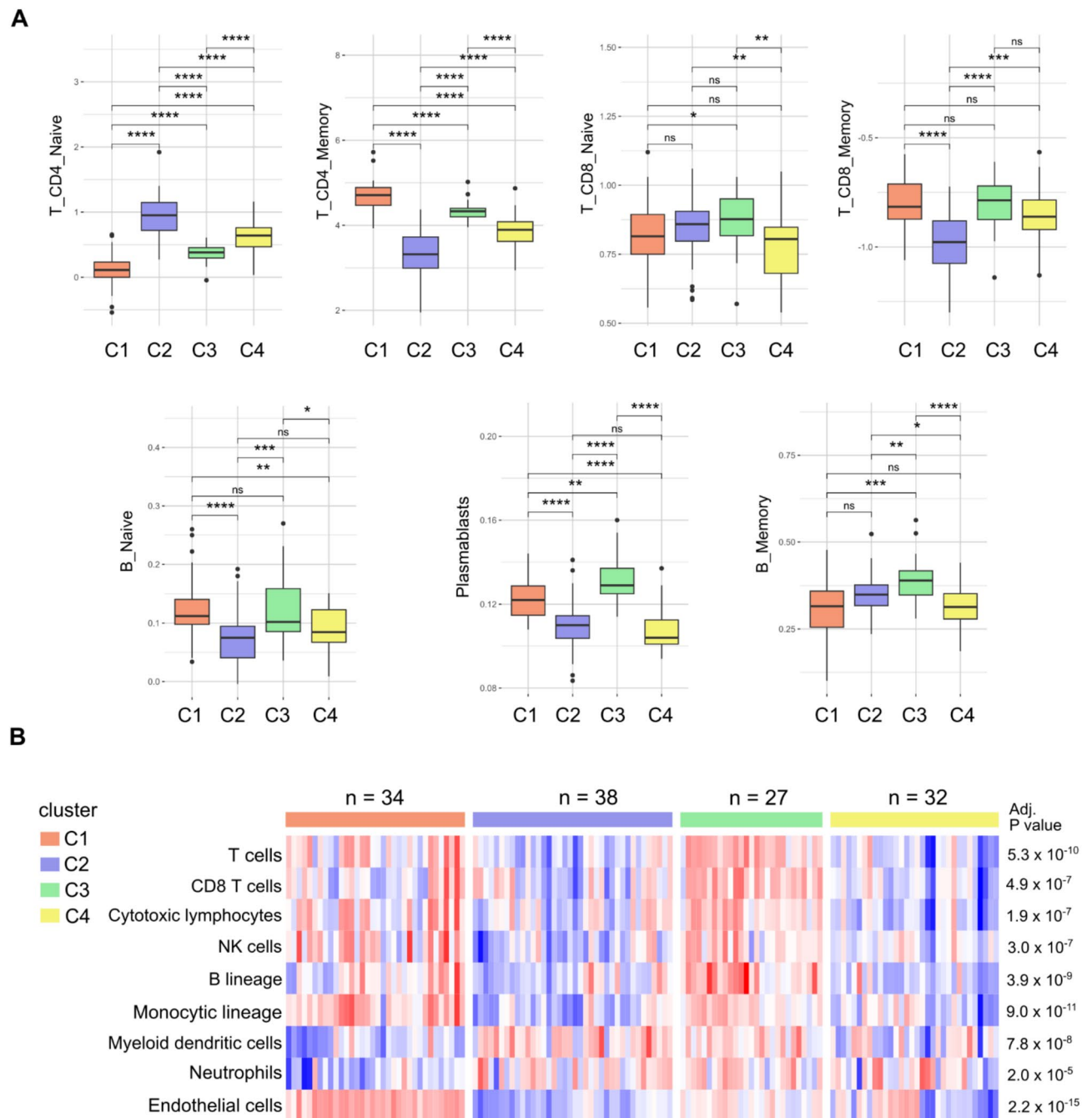


Fig. 3. Immune cell fractions across different clusters. **(A)** Box plots showing the distribution of immune cell fractions within each cluster assessed by ABIS algorithm, highlighting differences in immune composition. Statistical differences were assessed using the Wilcoxon rank test, with p-value thresholds set at *0.05, **0.01, ***0.001, and ****0.0001. **(B)** Composition of the microenvironment by cluster as defined by the MCP-counter scores. Adjusted P values were obtained from Benjamini–Hochberg correction of two-sided Kruskal–Wallis tests P values.

differences in immune and pathways profiling may have important implications for disease progression and patient response.

Gene expression profiling analysis

To identify the representative genes or signatures associated with the functional processes in each lesion cluster, we conducted differential expression analysis, comparing controls (C2) with lesions (C1, C3, and C4), as well as among the lesion groups themselves. Cluster C1 exhibited the highest number of dysregulated genes compared to non-tumor samples, with the majority of KSHV genes expressed (Fig. 4, Table S1). KSHV genes *K15*, *K13*, *ORF72*, *ORF75*, *vIRF-2*, and *vIRF-4* were among the top upregulated genes. Among the numerous affected

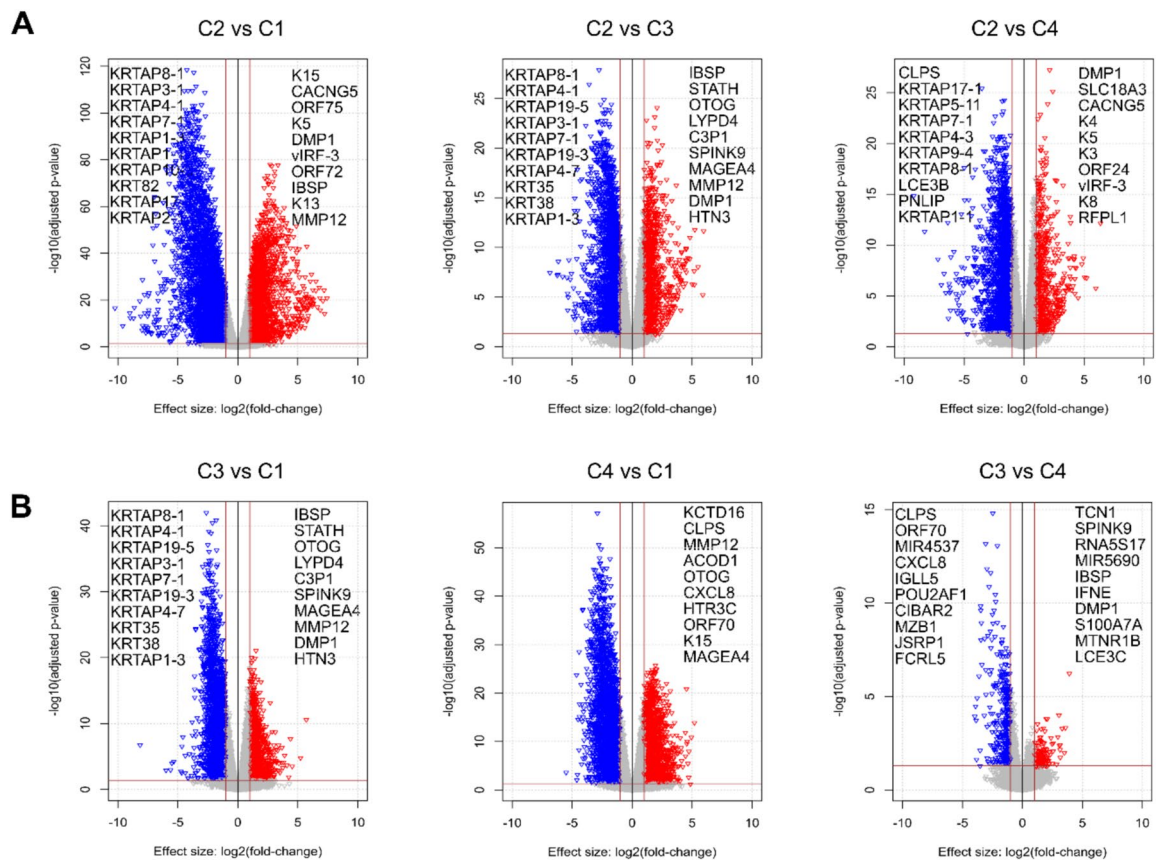


Fig. 4. Volcano plots showing differentially expressed genes (DEGs) across clusters. (A) Comparison of DEGs between the control cluster and the lesion clusters. (B) Comparison of DEGs between different lesion clusters. Each quadrant of the plots indicates the top 10 significantly upregulated or downregulated genes. The volcano plots are interpreted as follows: for a comparison such as C2 vs. C1, genes on the right are upregulated in C1, while those on the left are upregulated in C2 or downregulated in C1.

host genes, those linked to extracellular matrix remodeling stood out, including several metalloproteinases and collagens (e.g. *MMP9-17*, *ADAMTS5-7*, *COL4A1*), along with significant genes such as *ITGB1*, *FLT1*, *FLT2*, *PDGFA*, *PDGFB*, *AKT3*, *ANGPT2*, *CXCL10*, *ROCK2*, *IFNG*, and *CXCR4*, all of which are associated with processes relevant to KS, such as endothelial differentiation, angiogenesis, the PI3K/Akt pathway, and cytokine/chemokine cascades¹. Additionally, several genes related to cell proliferation (e.g. *PLK1*, *BUB1B*, *TTK*) and T cell lymphocytes (e.g. *CD86*, *CD274*, *CCR7*) were also upregulated. There was also a marked decrease in the expression of several genes linked to protein biosynthesis (ribosomal proteins), cellular respiration (e.g. *NDUFA1*, *NDUFA3*, *CYC1*), and lipid metabolism (*ADIPOQ*, *ACOX1*, *ACOX2*, *CYP4B1*) This underscores the substantial impact of KSHV on host cell metabolism, particularly under hypoxic conditions^{29–31}.

When comparing cluster C3 with controls (C2), a similar trend emerged, with a high expression of KSHV genes and a notable impact on host genes, though to a lesser extent than in cluster C1 (Fig. 4, Table S1). In cluster C3, there was a prominent enrichment of genes associated with lymphocyte activation and proliferation (e.g., *CD80*, *ITK*, *CD84*, *CD7*, *CD28*, *CD247*, *TLR9*, *LYN*, *CD22*). Cluster C4 also exhibited expression of most KSHV genes when compared to non-tumor controls; however, the host gene enrichment was primarily linked to the viral infection response (e.g., *IFIT1*, *IFIT2*, *IFIT3*, *IRF1*, *IRF7*, *IF16*) (Table S1).

The comparison between lesion clusters revealed distinct differences. In the cluster C1 vs. cluster C3 analysis, cluster C1 exhibits gene expression profiles more like KS tumors, specifically angiogenic genes (e.g. *TX1*, *ANGPTL3*, *PROX1*, *SCUBE1*, *KDR*, *ROBO4*, *PDGFB*, *SERPINA5*), mitotic genes (e.g. *CENPE*, *CDK1*, *KIF3C*, *BRCA2*, *TOP2A*, etc.), and remodeling matrix components including metalloproteinases⁵. Conversely, cluster C3 was distinguished by greater differentiation (higher keratin expression) and increased metabolic activity, evidenced by a higher expression of lipid metabolism-related genes and xenobiotic metabolism genes with activation of numerous cytochrome P450 genes (Table S1).

The comparison between clusters C1 and C4 reveals distinct metabolic and immune response profiles in the lesions. Cluster C4 shows elevated expression of oxidative and xenobiotic metabolism genes, along with ERBB signaling genes (e.g., *ERBB2*, *ERBB3*, *ERBB4*, *EGFR*, *EREG*), suggesting an adaptive response to environmental stressors and a focus on cellular survival and proliferation (Table S1). This points to potential metabolic reprogramming that could support tumorigenesis. The ERBB family of kinases has recently been implicated in regulating the KSHV latent-to-lytic replication switch, emphasizing their role in KSHV-related pathogenesis³².

In contrast, cluster C1 was marked by a strong activation of immune-related genes, particularly those associated with lymphocyte activation and inflammatory responses (e.g., *CXCL8*, *CXCL13*, *CXCL2*, *IL6*, *IFNG*, *IL10*, *CCR5*, *CXCR4*), indicating a robust immune environment potentially engaged in combating the tumor (Table S1).

Additionally, the comparison of clusters C3 and C4 revealed a lower number of differentially expressed genes (Fig. 4). C3 exhibited higher expression of genes linked to T and B cell activation (e.g., *CD84*, *CD86*, *CD70*, *CD79A*, *CD79B*, *IL10*), while C4 showed increased expression of genes involved in antiviral responses (e.g., *IFI44*, *IFIH1*, *IFI27*, *RSAD2*, *XAF1*, *OAS1*, *OAS2*, *OAS3*) (Table S1).

In summary, the transcriptomic and functional analysis identifies three distinct groups of KS. Cluster C1, primarily composed of endemic cases with some epidemic ones, represents tumors characterized by poorly differentiated, endothelial, and proliferative phenotypes. This cluster also showed enrichment of the PI3K/mTOR pathway and a strong presence of memory CD4 T lymphocytes in the immune fraction. Cluster C3, composed of predominantly epidemic cases, showed greater metabolic activity and was enriched in B cells. In contrast, cluster C4 had the least immune infiltration and highest differentiation, with strong activation of antiviral defense genes and the macular morphotype, typical of early-stage KS¹.

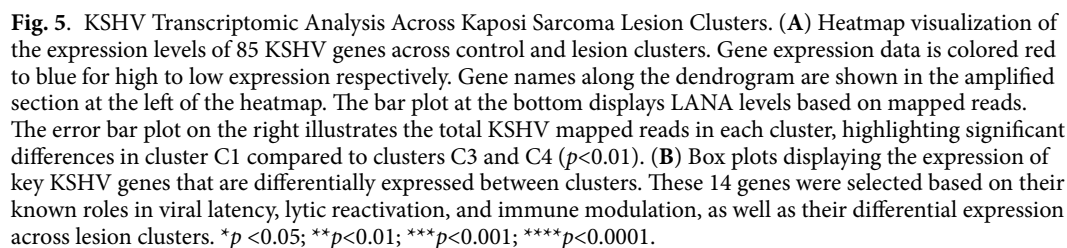
Overall, consistent with the findings of Lidenge et al.^{12,13}, which showed that endemic and epidemic KS transcriptomes are highly similar, with enhanced dysregulation of tumorigenesis and immune functions in endemic KS, we observed a segregation between endemic (C1) and epidemic (C3 and C4) cases. This distinction was driven by the magnitude and the specific subsets of expression changes. C1 (primarily endemic) displayed stronger activation of angiogenesis, endothelial differentiation, immune-related genes, cell proliferation, and tumorigenic pathways such as PI3K/Akt/mTOR. In contrast, C3 (epidemic) exhibited higher metabolic activity, with increased expression of genes related to lipid and xenobiotic metabolism, while C4 showed features typical of an early-stage viral response, including activation of antiviral defense pathways. These differences highlight functional divergence between C1, C3, and C4, supporting the clustering of endemic and epidemic KS lesions based on distinct immune and metabolic signatures.

KSHV transcriptome analysis

To evaluate the relationship between the clusters identified from the whole transcriptome and KSHV expression levels while minimizing variability from host gene expression, we chose to analyze the KSHV transcriptome alone between clusters. This approach clarified the viral contributions to the disease and helped to determine if different viral expressions correlate with the observed clusters. First, it is important to note that there were no significant differences in overall KSHV expression levels (measured by total mapped KSHV reads, TMKR) across the different cohort populations (Argentina, Tanzania/Zambia, and Uganda; $p > 0.05$; Fig. S3). However, when comparing the clusters, cluster C1 contained tumors with the highest levels of KSHV expression, as indicated by the TMKR, compared to clusters C3 and C4 ($p < 0.01$; Fig. 5A). This difference was further supported by the elevated levels of LANA in cluster C1 (Fig. 5A). The analysis of differentially expressed genes among the lesion clusters confirmed that cluster C1 exhibited the highest expression levels of genes such as *ORF72*, *ORF73* (LANA), *ORF74*, *ORF75*, as well as *K12*, *K13*, and *K15*, which were among the most highly expressed transcripts across all clusters. These genes have been implicated in various functions, including immune evasion, angiogenesis, cell proliferation, and apoptosis regulation. Based on their biological relevance and differential expression patterns, a focused panel of 14 representative KSHV genes was selected for visualization in Fig. 5B. LANA is known to stabilize the viral genome within host cells, facilitating persistence and contributing to tumorigenesis^{33,34}. *ORF74* is a viral GPCR that binds various human chemokines (*CXCL1*, *CXCL8*) contributing to immune evasion and consequently viral dissemination^{35,36}. Similarly, *K12* and *K13* play roles in modulating host immune responses³⁷, suggesting that the elevated expression of these genes in cluster C1 may correlate with more aggressive tumor behavior or enhanced viral persistence. Although no significant differences were observed in the total mapped reads between clusters C3 and C4 (Fig. 5A), these clusters were differentiated by the expression of specific genes, including LANA and RTA (*ORF50*). Additionally, lytic genes such as the viral Bcl-2 (*ORF16*), which inhibits apoptosis, the protease (*ORF17*) associated with capsid maturation, and the gene expression modulator (*ORF18*) were all more abundantly expressed in C3 compared to C4 (Fig. 5B, Table S2). This suggests that KSHV is contributing to the intrinsic transcriptomic differences among these clusters of epidemic tumors, potentially influencing their pathogenic behavior and responses to treatment. The distinct expression profiles of these genes highlight the complex interplay between viral replication and host cell dynamics in KS.

We conducted further correlation analyses to examine the relationship between viral tissue load (quantified via TMKR) and KS morphotypes. The macular group exhibited a significantly lower viral load compared to the plaque and nodular groups (Fig. S3). Although no significant differences were found between the plaque, nodular, and fungating groups, an increasing trend in mean TMKR values was observed from macular to fungating lesions (Fig. S3). These findings correlate with the observation that cluster C4, primarily composed of macular lesions, exhibited lower viral load, consistent with the early stage of these lesions. This suggests an association between viral load and lesion stage in KSHV-driven tumorigenesis. We also explored the relationship between immune infiltration patterns and viral loads across clusters (Fig. S3). Linear regression analysis revealed that higher KSHV loads were associated with a decrease in naïve CD4+ T cells and an increase in memory CD4+ T cells, particularly in Cluster C1. A similar trend, though weaker, was observed for CD8+ T cells, with memory CD8+ T cells showing a positive association with KSHV load. No significant associations were found between KSHV load and B cell subsets, suggesting that B cell enrichment in cluster C3 may be an intrinsic feature of these lesions, independent of KSHV load.

While our study provides valuable insights into the molecular diversity of KS, several limitations should be acknowledged. First, RNA-seq data, while powerful for transcriptomic profiling, may not fully capture the presence and spatial distribution of immune cells within the tumor microenvironment. Immune profile algorithms infer immune cell composition based on gene expression rather than direct cellular localization,



KS pathology across different populations. Although the Argentine cohort is smaller in sample size, its inclusion increased the statistical power of the combined dataset from 84 to 91% for detecting differentially expressed genes between normal and KS samples ($\text{FDR} < 0.01$, fold change > 2). This highlights the cohort's meaningful contribution to the sensitivity and robustness of the analysis, underscoring the value of integrating diverse, high-quality datasets. Further validation, including spatial transcriptomics, will be necessary to refine our understanding of KS molecular landscapes and their therapeutic implications.

Conclusion

In summary, this is the first study to integrate transcriptomic profiles of various cohorts of KS. The RNA-seq profile generated from the Argentinian KS samples compiled with previously published datasets as well as their associated metadata and predicted pathway activities, immune infiltrate, host and viral gene expression profiles are shared with the research community through the development of a KS-omics resource leveraging UCSC Xena browser. This resource will allow biologists without bioinformatics knowledge to explore and correlate the host and viral transcriptomic profiles in a compiled and harmonized dataset to develop novel biological insights and discover biomarkers.

Our comprehensive analysis enabled us to define and characterize three distinct lesion groups that reflect both the biology of the virus and the intrinsic molecular characteristics of the lesions. Notably, while endemic lesions were primarily concentrated in cluster C1, epidemic lesions were distributed across all three clusters.

Epidemic lesions in clusters C3 and C4 did not show significant differences in overall KSHV expression but differed in their intrinsic functional molecular characteristics. Cluster C3 exhibited a high level of B cell infiltration and ICIs expression, while cluster C4 was characterized by an enrichment of genes involved in early response against viral infection. Whether these different clusters imply differences in treatment approaches or response, it still needs to be explored.

Thus, we defined three groups of KS: endothelial, proliferative and well-defined KS tumors (C1); B-cell-enriched and metabolic tumors (C3); and less immunogenic, early-stage KS tumors (C4). Overall, our findings provide valuable insights into the molecular diversity of KS and underscore the potential for tailored therapeutic strategies based on lesion characteristics.

Data availability

The raw data have been submitted to NCBI GEO database with accession number GSE271303. The rest of the data are available from the corresponding author upon reasonable request.

Received: 18 March 2025; Accepted: 23 June 2025

Published online: 10 October 2025

References

- Cesarman, E. et al. Kaposi sarcoma. *Nat. Rev. Dis. Primers* **5**(1), 9 (2019).
- Dittmer, D. P. & Damania, B. Kaposi sarcoma-associated herpesvirus: Immunobiology, oncogenesis, and therapy. *J. Clin. Investig.* **126**(9), 3165–3175 (2016).
- Mesri, E. A., Cesarman, E. & Boshoff, C. Kaposi's sarcoma and its associated herpesvirus. *Nat. Rev. Cancer* **10**(10), 707–719 (2010).
- Naipauer, J. & Mesri, E. A. The Kaposi's sarcoma progenitor enigma: KSHV-induced MEndT-EndMT axis. *Trends Mol. Med.* **29**(3), 188–200. <https://doi.org/10.1016/j.molmed.2022.12.003> (2023).
- Lacunza, E. et al. Unveiling the role of KSHV-infected human mesenchymal stem cells in Kaposi's sarcoma initiation. *J. Med. Virol.* **96**(5), e29684. <https://doi.org/10.1002/jmv.29684> (2024).
- Haverkos, H. W. & Drotman, D. P. Prevalence of Kaposi's sarcoma among patients with AIDS. *N. Engl. J. Med.* **312**(23), 1518 (1985).
- Hutchison, L. M., Boscoe, F. P. & Feingold, B. J. Cancers disproportionately affecting the New York state transgender population, 1979–2016. *Am. J. Public Health* **108**(9), 1260–1262 (2018).
- AIDS-defining Cancer Project Working Group for IeDEA and COHERE in EuroCoord. Comparison of Kaposi sarcoma risk in human immunodeficiency virus-positive adults across 5 continents: A multiregional multicohort study. *Clin. Infect. Dis.* **65**(8), 1316–1326. <https://doi.org/10.1093/cid/cix480> (2017).
- Tso, F. Y. et al. RNA-Seq of Kaposi's sarcoma reveals alterations in glucose and lipid metabolism. *PLoS Pathog.* **14**(1), e1006844 (2018).
- Moorad, R. et al. A prospective cohort study identifies two types of HIV+ Kaposi Sarcoma lesions: proliferative and inflammatory. *Int. J. Cancer* **153**(12), 2082–2092. <https://doi.org/10.1002/ijc.34689> (2023).
- Ramaswami, R. et al. Transcriptional landscape of Kaposi sarcoma tumors identifies unique immunologic signatures and key determinants of angiogenesis. *J. Transl. Med.* **21**(1), 653 (2023).
- Lidenge, S. J. et al. Lack of CD8+ T-cell co-localization with Kaposi's sarcoma-associated herpesvirus infected cells in Kaposi's sarcoma tumors. *Oncotarget* **11**(17), 1556–1572 (2020).
- Lidenge, S. J. et al. 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**(7), e1008681 (2020).
- Rose, T. M. et al. Quantitative RNAseq analysis of Ugandan KS tumors reveals KSHV gene expression dominated by transcription from the LTR downstream latency promoter. *PLoS Pathog.* **14**(12), e1007441 (2018).
- Uldrick, T. S. & Whitby, D. Update on KSHV epidemiology, Kaposi Sarcoma pathogenesis, and treatment of Kaposi Sarcoma. *Cancer Lett.* **305**(2), 150–162 (2011).
- Damania, B. & Dittmer, D. P. Today's Kaposi sarcoma is not the same as it was 40 years ago, or is it?. *J. Med. Virol.* **95**(5), e28773. <https://doi.org/10.1002/jmv.28773>. PMID:37212317; PMCID:PMC10266714 (2023).
- Manimaran, S. et al. BatchQC: Interactive software for evaluating sample and batch effects in genomic data. *Bioinformatics* **32**(24), 3836–3838. <https://doi.org/10.1093/bioinformatics/btw538> (2016).
- Zhang, Y., Parmigiani, G. & Johnson, W. E. ComBat-seq: Batch effect adjustment for RNA-seq count data. *NAR Genom. Bioinform.* **2**(3), Iqaa078 (2020).
- Monaco, G. et al. RNA-Seq signatures normalized by mRNA abundance allow absolute deconvolution of human immune cell types. *Cell Rep.* **26**(6), 1627–1640.e7 (2019).

20. Becht, E. et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* **17**, 218 (2016).
21. Goldman, M. J. et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-020-0546-8> (2020).
22. Petitprez, F. et al. Transcriptomic analysis of the tumor microenvironment to guide prognosis and immunotherapies. *Cancer Immunol. Immunother.* **67**(6), 981–988. <https://doi.org/10.1007/s00262-017-2058-z> (2018).
23. Joest, B. et al. Stage-related PD-L1 expression in Kaposi sarcoma tumor microenvironment. *J. Cutan. Pathol.* **47**(10), 888–895. <https://doi.org/10.1111/cup.13716> (2020).
24. Lee, A. Q., Hao, C., Pan, M., Ganjoo, K. N. & Bui, N. Q. Histologic and immunologic factors associated with response to immune checkpoint inhibitors in advanced sarcoma. *Clin. Cancer Res.* <https://doi.org/10.1158/1078-0432.CCR-24-3485> (2024).
25. Zhu, Y., da Silva, S. R. & Gao, S. J. Metabolic reprogramming by KSHV in cellular transformation. *Cancer Metab.* **2**(Suppl 1), P88. <https://doi.org/10.1186/2049-3002-2-S1-P88> (2014).
26. Rappocciolo, G. et al. Human herpesvirus 8 infects and replicates in primary cultures of activated B lymphocytes through DC-SIGN. *J. Virol.* **82**, 4793–4806. <https://doi.org/10.1128/JVI.01587-07> (2008).
27. Ballon, G., Chen, K., Perez, R., Tam, W. & Cesarman, E. Kaposi sarcoma herpesvirus (KSHV) vFLIP oncoprotein induces B cell transdifferentiation and tumorigenesis in mice. *J. Clin. Invest.* **121**, 1141–1153. <https://doi.org/10.1172/JCI44417> (2011).
28. Nicol, S. M. et al. Primary B lymphocytes infected with Kaposi's sarcoma-associated herpesvirus can be expanded in vitro and are recognized by LANA-specific CD4+ T cells. *J. Virol.* **90**(8), 3849–3859. <https://doi.org/10.1128/JVI.02377-15> (2016).
29. Delgado, T. et al. Induction of the Warburg effect by Kaposi's sarcoma herpesvirus is required for the maintenance of latently infected endothelial cells. *Proc. Natl. Acad. Sci. USA* **107**, 10696–10701 (2010).
30. Singh, R. K., Lang, F., Pei, Y., Jha, H. C. & Robertson, E. S. Metabolic reprogramming of Kaposi's sarcoma associated herpes virus infected B-cells in hypoxia. *PLoS Pathog.* **14**(5), e1007062. <https://doi.org/10.1371/journal.ppat.1007062>. PMID:29746587;PMCID:PMC5963815 (2018).
31. Méndez-Solís, O. et al. Kaposi's sarcoma herpesvirus activates the hypoxia response to usurp HIF2 α -dependent translation initiation for replication and oncogenesis. *Cell Rep.* **37**(13), 110144. <https://doi.org/10.1016/j.celrep.2021.110144>. PMID:34965440;PMCID:PMC9121799 (2021).
32. Olson, A. T. et al. Polypharmacology-based kinome screen identifies new regulators of KSHV reactivation. *PLoS Pathog.* **19**(9), e1011169. <https://doi.org/10.1371/journal.ppat.1011169>. PMID:37669313;PMCID:PMC10503724 (2023).
33. Nakajima, K., Inagaki, T., Espere, J. M. & Izumiya, Y. Kaposi's sarcoma-associated herpesvirus (KSHV) LANA prevents KSHV episomes from degradation. *J. Virol.* **98**(2), e0126823. <https://doi.org/10.1128/jvi.01268-23> (2024).
34. Uppal, T., Banerjee, S., Sun, Z., Verma, S. C. & Robertson, E. S. KSHV LANA: The master regulator of KSHV latency. *Viruses* **6**(12), 4961–4998. <https://doi.org/10.3390/v6124961>. PMID:25514370;PMCID:PMC4276939 (2014).
35. de Munnik, S. M. et al. The viral G protein-coupled receptor ORF74 hijacks β -arrestins for endocytic trafficking in response to human chemokines. *PLoS ONE* **10**(4), e0124486. <https://doi.org/10.1371/journal.pone.0124486> (2015).
36. Medina, M. V. et al. KSHV G-protein coupled receptor vGPCR oncogenic signaling upregulation of Cyclooxygenase-2 expression mediates angiogenesis and tumorigenesis in Kaposi's sarcoma. *PLoS Pathog.* **16**(10), e1009006. <https://doi.org/10.1371/journal.ppat.1009006> (2020).
37. Liang, C., Lee, J. S. & Jung, J. U. Immune evasion in Kaposi's sarcoma-associated herpes virus associated oncogenesis. *Semin. Cancer Biol.* **18**(6), 423–436. <https://doi.org/10.1016/j.semcancer.2008.09.003> (2008).

Acknowledgements

The authors want to acknowledge the invaluable contributions from all study participants and from all the research team at Fundación Huésped, where participants were recruited. This work was supported by the National Cancer Institute of the National Institutes of Health under Award Number U54CA221208 (OC and JCR), P30CA240139 (JCR), 5U24CA258407 (UCSC Xena Browser), and the National University of La Plata M250 I+D grant (MCA).

Author contributions

EL: Investigation, formal analysis, writing the article. VF, MIF: Resources and logistics for sample and clinical data collection. MES, AMG, SW: Methodology, research assistance, clinical data of participants. MJG and JZ: UCSC Xena. JN and EC: Advised on the study and revised the manuscript. PC, OC, JCR: Resources. MCA: Conceived the study, supervision, formal analysis, writing the article. All authors read and approved the final manuscript.

Funding

This work was supported by the National Cancer Institute of the National Institutes of Health under Award Number U54CA221208 (OC and JCR), P30CA240139 (JCR), 5U24CA258407 (UCSC Xena Browser), and the National University of La Plata M250 I+D grant (MCA).

Declarations

Ethics declarations

MJG is a consultant for the Fred Hutchinson Cancer Research Center. The other authors declare no competing interests.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-08564-4>.

Correspondence and requests for materials should be addressed to M.C.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025