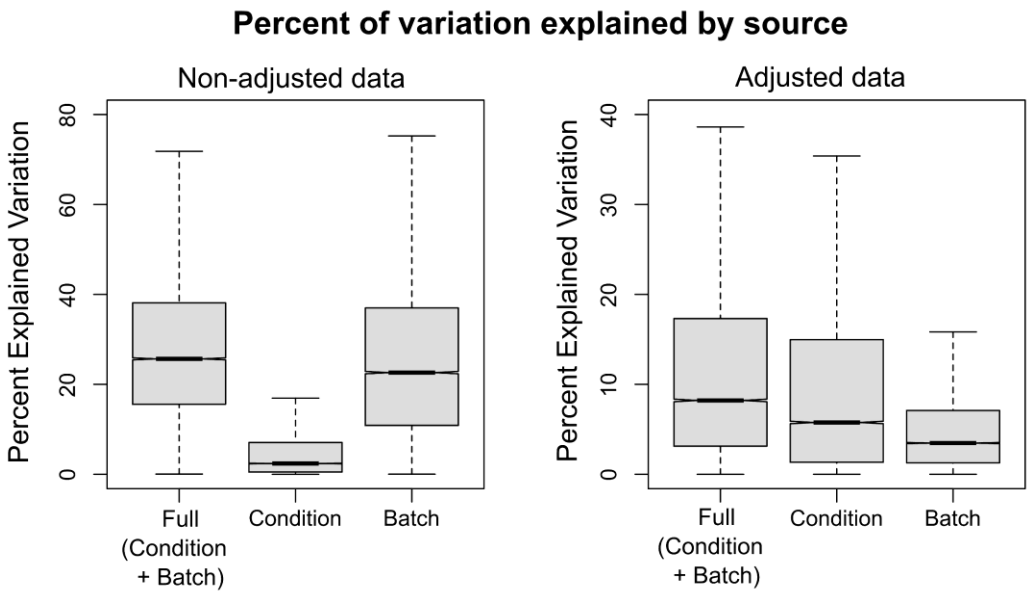
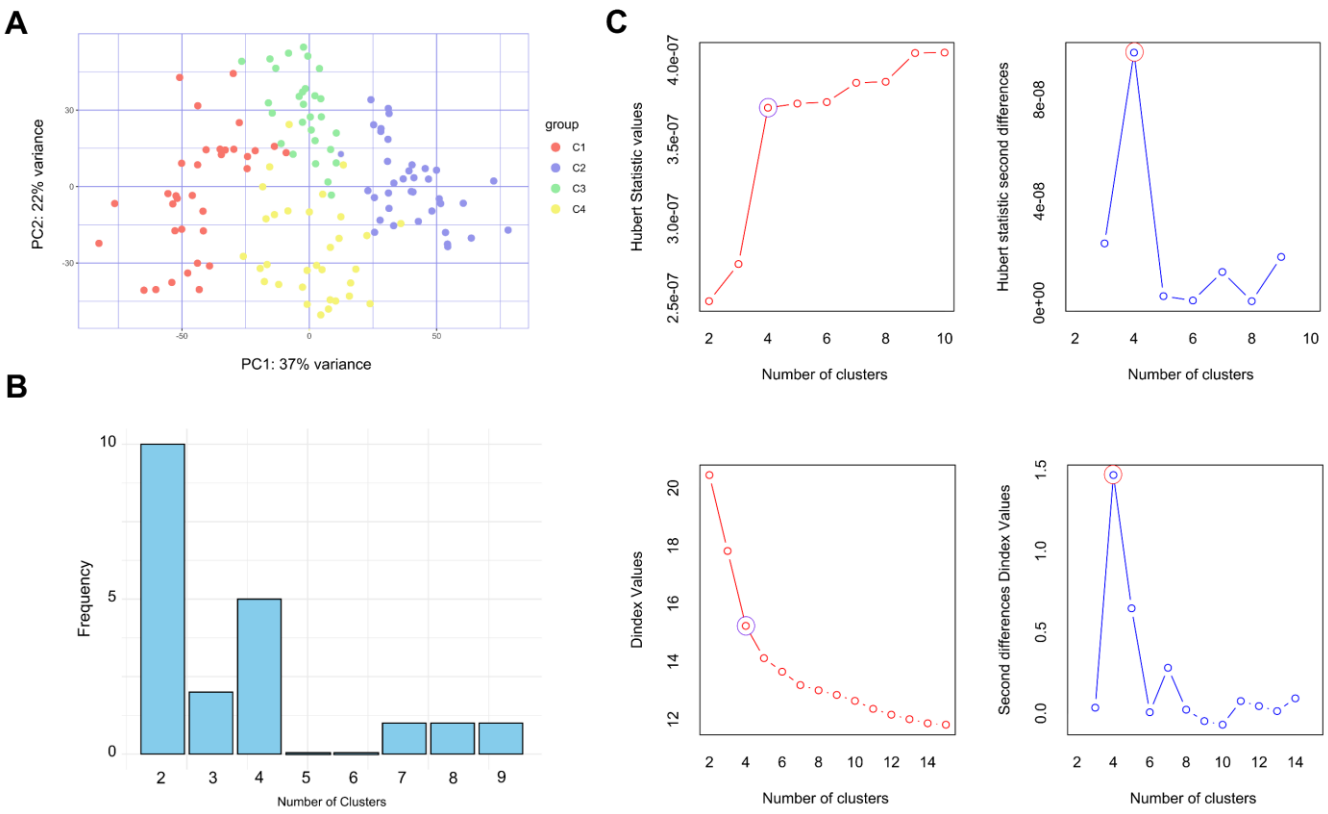
INTEGRATIVE FUNCTIONAL GENOMICS ANALYSIS OF KAPOSI SARCOMA COHORTS

Ezequiel Lacunza [1], Valeria Fink [2], Julian Naipauer [5], María E. Salas [1], Ana M. Gun [2], Mary J. Goldman [3], Jingchun Zhu [3], Sion Williams [4], María I. Figueroa [2], Pedro Cahn [2], Omar Coso [5], Ethel Cesarman [6], Juan C. Ramos [4], Martín C. Abba [1,*].

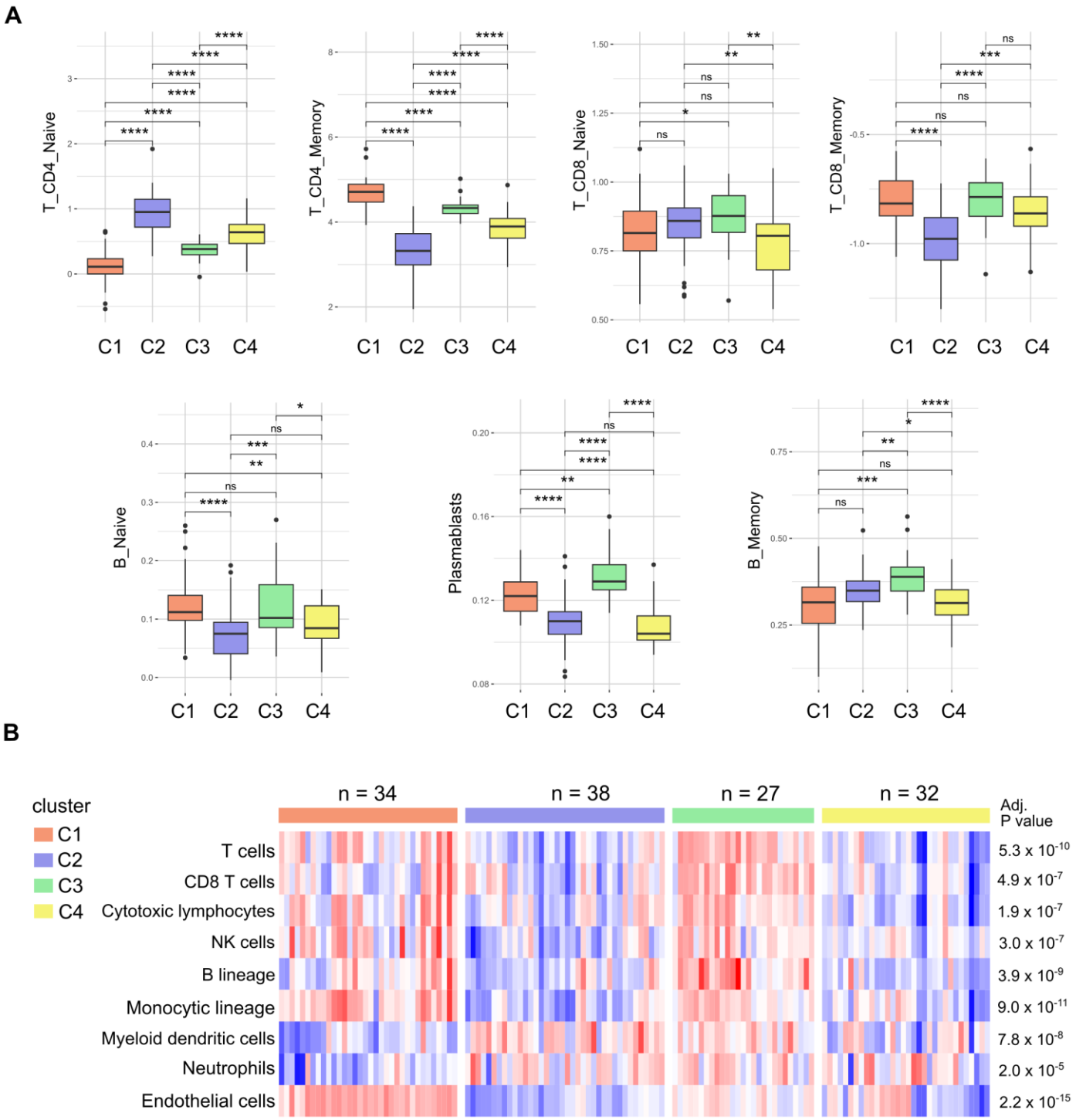
Additional figures



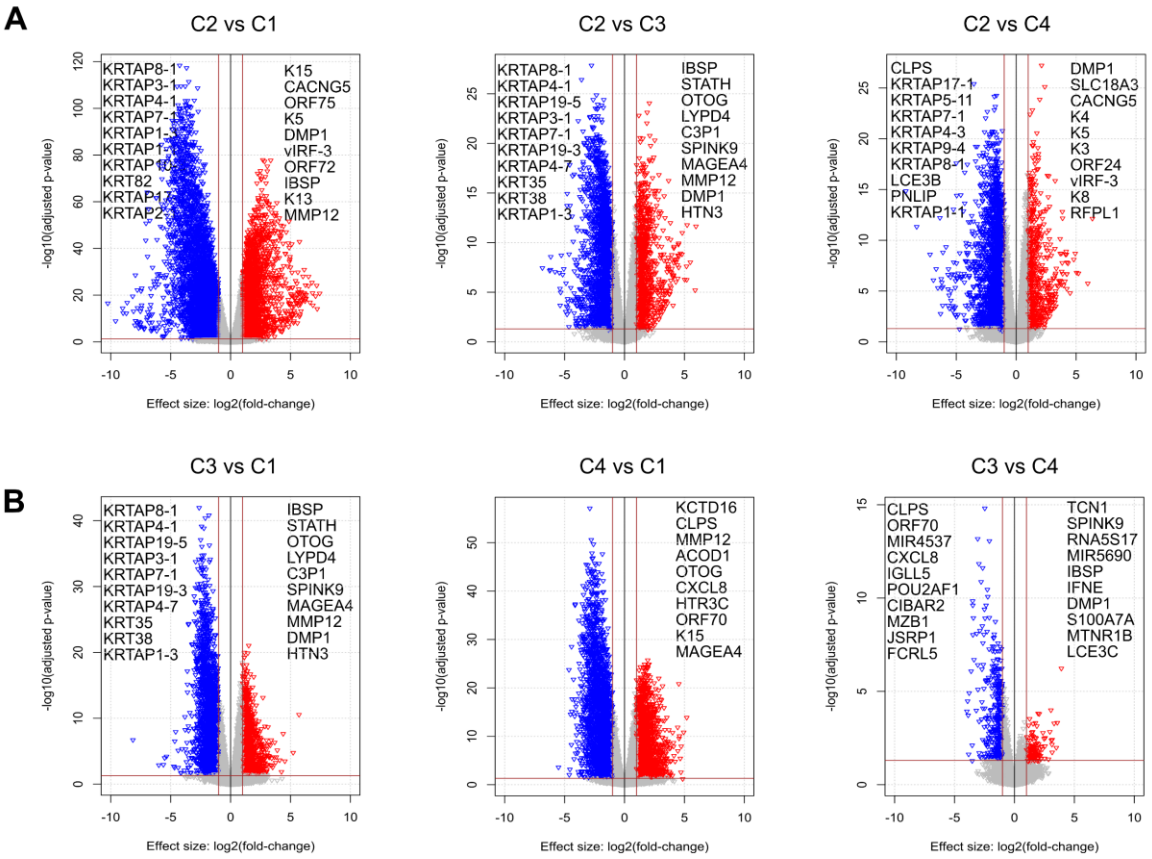
Supplementary Figure 1: Box plots illustrating the percentage of variation explained across the transcriptome by three factors: full model (condition + batch), condition (control or KS), and batch. The effect of batch adjustment can be noted in the right boxplot of the variation analysis compared with the non-adjusted data in the left boxplot. After adjustment, the variation analysis showed decreased evidence of batch effects and most of the biological variation was retained as the condition source.



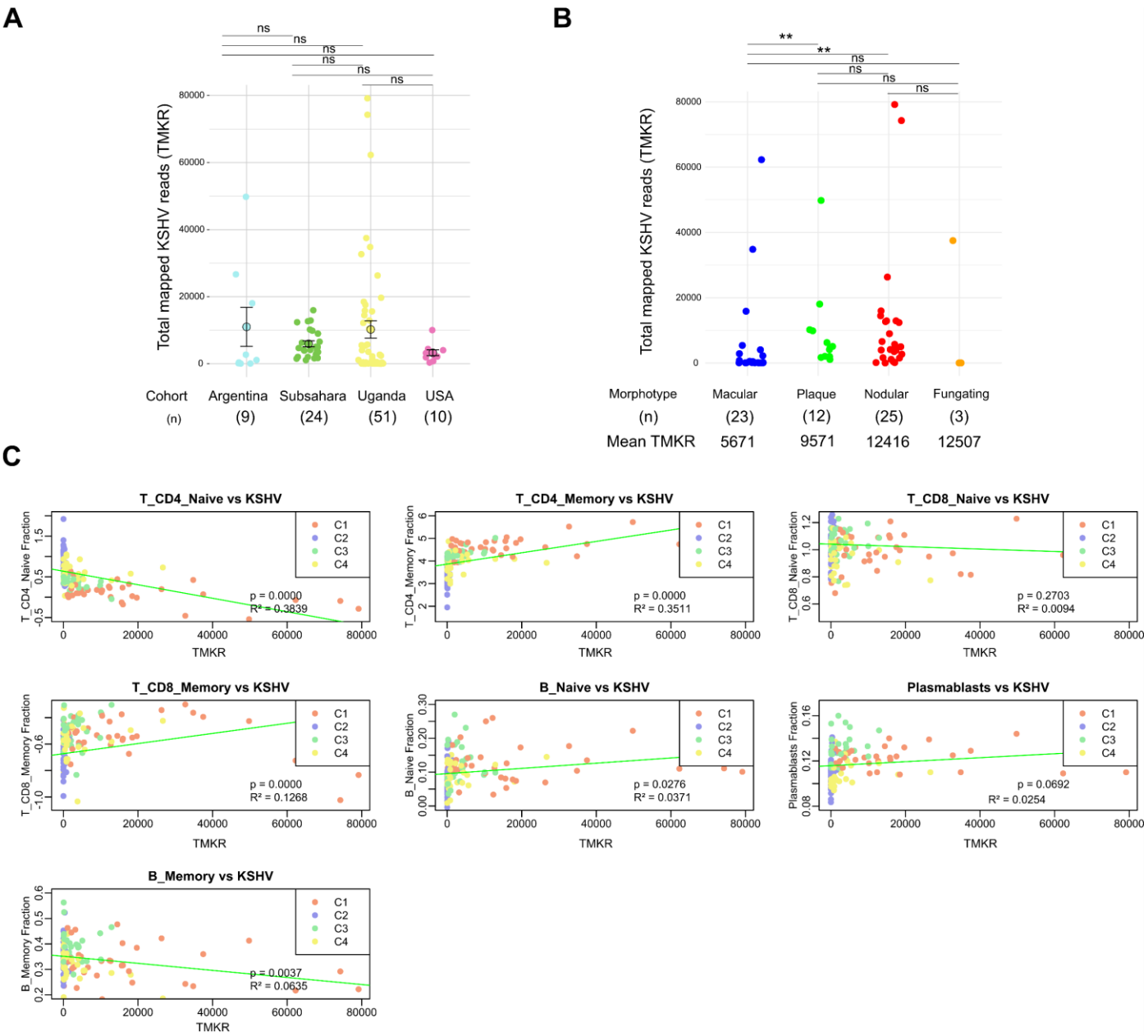
Supplementary Figure 2: Clustering analysis and validation of sample groupings using PCA and clustering indices. A. Principal Component Analysis (PCA) of samples, color-coded according to predefined clusters C1-C4, showing distinct groupings that reflect the underlying clustering structure. **B.** Histogram depicting the frequency distribution of clustering indices and their suggested number of clusters, with the majority of indices indicating either 2 or 4 clusters. **C.** Hubert's Index (upper half) and Dindex (bottom) plots, both assessing the optimal number of clusters, with both methods indicating that 4 clusters (highlighted with a red/blue circle) provide the best partitioning of the data.



Supplementary Figure 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.



Supplementary Figure 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.



Supplementary Figure 5: KSHV loads versus cohorts, morphotype and immune infiltration. A. Scatter plot of Total Mapped KSHV Reads (TMKR) by cohort populations in KS lesions. No significant difference in KSHV expression was observed across populations, assessed using the Kruskal-Wallis H test. **B.** Scatter plot of TMKR by KS morphotype, ordered from early to advanced clinical stages (macular, plaque, nodular, and fungating). The macular group (n = 23, mean = 5671) exhibited significantly lower viral load compared to the plaque (n = 12, mean = 9571) and nodular (n = 25, mean = 12561) groups. **C.** Linear regression plots between immune cell fractions and KSHV expression (TMKR). Cluster labels (C1, C2, C3, C4) and statistical significance (p) for the relationship between immune cell fractions and KSHV expression are shown. R^2 indicates the goodness of fit for the regression model. ** $p < 0.01$; ns, indicates no significant differences.