Supplementary Data

HTLV-1-Induced Neuroimmunome Correlates with Disease Progression and Severity

Fernando Yuri Nery do Vale[1,2*], Carlota Miranda-Solé[3*], Adriel Leal Nóbile[1,2] , Júlia Nakanishi Usuda[1,2] , Dennyson Leandro M Fonseca[4], Lena F. Schimke[5], Mauro Cesar Cafundó Morais[2,6,7], Débora Gomes de Albuquerque Freitas[1,5], Anny Silva Adri[1,2], Roseane Galdioli Nava[1,4], Yohan Lucas Gonçalves Corrêa[1,2], Hugo Fernando Nery do Vale[1,8,], Adriana Simizo[6], Niels Olsen Camara[5], Gustavo Cabral-Miranda[5], Helder I. Nakaya[2,6,7], Rodrigo Dalmolin[9], Haroldo Dutra Dias[1,10,11], Yuki Saito[12,13], Yasunori Kogure[12], Junji Koya[12,14], Keisuke Kataoka[12,14], Igor Salerno Filgueiras[1,5], Margarita Dominguez-Villar[3*], Otávio Cabral Marques[1,8,12*]

[1] Laboratory of Psychoneuroimmunology, Selye Lab, University of São Paulo School of Medicine, São Paulo, Brazil.
[2] Department of Clinical and Toxicological Analyses, School of Pharmaceutical Sciences, University of São Paulo, São Paulo, Brazil.
[3] Imperial College London, Department of Infectious Diseases, Imperial College, UK.
[4] Interunit Postgraduate Program on Bioinformatics, University of São Paulo, São Paulo, Brazil.
[5] Department of Immunology, Institute of Biomedical Sciences, University of São Paulo, Brazil.
[6] Hospital Israelita Albert Einstein, São Paulo, Brazil.
[7] Institut Pasteur de São Paulo, São Paulo, Brazil.
[8] D'Or Institute for Research and Education, Rio de Janeiro, Brazil.
[9] Bioinformatics Multidisciplinary Environment, Federal University of Rio Grande do Norte, Brazil.
[10] Department of Medicine, Division of Molecular Medicine, Laboratory of Medical Investigation 29, University of São Paulo, Brazil.
[11] Ann Romney Center for Neurologic Diseases, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[12] Division of Molecular Oncology, National Cancer Center Research Institute, Tokyo, Japan.
[13] Division of Gastroenterology and Hepatology, Department of Internal Medicine, Keio University School of Medicine, Tokyo, Japan.
[14] Division of Hematology, Department of Medicine, Keio University School of Medicine, Tokyo, Japan
*Contributed Equally

Corresponding authors

Fernando Yuri Nery do Vale; E-mail: feryurinery@usp.br

Otavio Cabral-Marques, MSc, PhD; Email: otavio.cmarques@usp.br

Supplementary methods

Research Approach

We employed a multi-dimensional systems biology framework to investigate the molecular underpinnings of HTLV-1 pathogenesis (**Supplementary Figure 1**). Building on our single-cell RNA sequencing (scRNA-seq) dataset comprising 233,093 PBMCs, we aimed to validate and further explore the neuroimmunome dysregulation identified in this cohort. To that end, we integrated publicly available bulk RNA-sequencing from PBMCs and microarray datasets (n = 200) derived from independent cohorts of individuals with HTLV-1 infection and healthy controls. This integrative strategy enabled cross-platform validation of neuroimmune gene signatures and reinforced the consistency and generalizability of the observed molecular alterations across distinct biological contexts and transcriptomic resolutions. To strengthen these findings, we also performed flow cytometric analysis to validate key neuroimmune markers at the protein level and assessed their association with HTLV-1 proviral load across clinical subgroups.

All Statistical analyses were performed with R Studio software (The R Foundation for Statistical Computing)[31] (v. 4.3.2). R packages from bioconductor[32] and CRAN repository were used to analyse the transcriptomic data and to perform the data visualization.

Data availability

The tables used to generate and visualize the results of our study are provided as supplementary files. These tables include comprehensive data inputs for the analysis and detailed output results that encapsulate the findings and

insights derived from our research. The supplementary input tables outline the dataset structure, variables, and relevant parameters, while the output tables display processed results, statistical metrics, and visual summaries that support our conclusions. Datasets used in the analysis were publicly available at the GEO datasets under the accession codes GSE29312, GSE38537, GSE233437, and GSE33615, and in the European Genome-phenome Archive (EGA) under the accession number EGAS00001004936. The custom code used for all analyses is available at https://github.com/feryurinv23.

Single cell analysis.

scRNA-seq data from 30 ATL patients, 11 HTLV-1 asymptomatic carriers (ACs), and 4 healthy donors were analyzed. The Seurat object was loaded into RStudio using the Seurat package[1–3] for further processing. Quality control measures were applied to ensure high data integrity. Cells with fewer than 200 or more than 2,500 detected RNA features were excluded to eliminate low-quality cells, such as dead or dying cells with minimal transcriptional activity, and potential doublets with abnormally high RNA content. Additionally, cells in which more than 5% of unique molecular identifiers (UMIs) were derived from mitochondrial genes were removed to avoid artifacts associated with cellular stress or apoptosis. The data were then normalized and scaled to correct for variations between cells, enabling more accurate comparisons.

Dimensionality reduction was performed using principal component analysis (PCA), followed by cell clustering via the Louvain algorithm. The resulting structure was visualized using Uniform Manifold Approximation and Projection (UMAP), as previously described by Koya et al. (2021)[4]. Cell types,

malignant and non-malignant cells were assigned based on previous reports by Koya et al., 2021[4], which used T-cell marker expression, TCR repertoire analysis, and mRNA/ADT levels of canonical lineage markers[4]. To confirm these classifications, we applied the FeaturePlot() function using gene markers listed in **Supplementary Table 01**.

For downstream analyses, differentially expressed genes (DEGs) were identified using the FindClusters() function, comparing gene expression profiles between ACs, ATL subtypes, and the healthy control group.

Functional enrichment, clustering, and interactome

Gene Ontology (GO) enrichment analysis was performed using the clusterProfiler package[5–7] in R (v. 4.3.2) and the EnrichR[8–10] bioinformatics tool to identify enriched biological processes (BPs). Pathway analysis was conducted using the KEGG database[11], and for all databases, a significance threshold of adjusted *p*-value ≤ 0.05 was applied. To evaluate the nervous and immune clusters formed by BPs, the Appyters[12] web tool, integrated with EnrichR[12], was utilized. This approach facilitated the identification of enriched BPs, which were then grouped into clusters and visualized using a scatter plot. The Leiden algorithm was applied for clustering[13], with dimensionality reduction performed using Uniform Manifold Approximation and Projection (UMAP). The results were visualized using the ggplot2 package[14] in R (v. 4.3.2). Additionally, a network analysis of genes associated with enriched BPs was conducted to explore the neuroimmunological interactome. This was achieved using the cnetplot() function from the enrichplot package[15] in R (v. 4.3.2).

Characterization of neuroimmunome genes and terms related to biological processes

To characterize DEGs involved in the neuroimmune response, we utilized the Uberon[16] and BPs database to identify genes annotated as belonging to both the immune and nervous systems. In the subsequent step, we selected BPs terms for dotplot and network visualization. For neurological processes, we used the following keywords: “synapse”, “synaptic”, “neuro”, “glia”, “nervous”, “axon”, “glutamate”, “gliogenesis”, and “nerve.” For immune-related processes, we included: “response to virus”, “immune response”, “B cell”, “T cell”, “cytokine”, “leukocytes”, “humoral immune”, “lymphocytes”, “immunoglobulin”, “interferon”, “innate immune”, and “adaptive immune.” This enrichment analysis provided deeper insight into how these neuroimmune genes may contribute to both systems’ biological functions and their relevance in the context of our study.

Assessing Synaptic-related Processes at Single-Cell Level

We apply a comprehensive workflow to identify pathways and BPs related to synaptic activity and neuro-immune process, aiming to gain new insights into the role of peripheral neuroimmune communication at the single-cell level in the context of HTLV-1. By integrating different bioinformatics approaches, we explore the peripheral synaptic gene expression associated with DEGs in PBMCs from individuals with HTLV-1 infection, contributing to our hypothesis on neuroimmune interactions.

Enrichment analysis of the synaptic neurotransmitter gene set was performed using the ESCAPE

First, we performed gene set enrichment analysis using the ESCAPE R package[17], applying single-sample GSEA (ssGSEA) to calculate enrichment scores at the single-cell level. To construct the gene set, we curated genes from the KEGG 2021 database[11], a comprehensive resource for understanding high-level functions and utilities of biological systems. The gene set included genes associated with synaptic activity and neuro-ligand receptor interactions, particularly those involved in GABAergic and dopaminergic synapses. We then used the runEscape() function, which implements the Gene Set Enrichment Analysis (GSEA) method[18], to compute enrichment scores for each pathway and integrate the results directly into the single-cell object. For visualization, we applied the heatmapEnrichment() function to display enrichment scores across different cell clusters.

Neurotransmission-linked gene networks associated to disease stage

Additionally, neurotransmitter pathways were analyzed to investigate the DEGs identified in individuals with HTLV-1 infection in PBMCs associated with the ACs and ATL subtypes, aiming to understand the role of neurotransmitter-related gene expression and the neurotransmitter families to which they belong. This analysis was conducted using the KEGG database[11], combined with network visualization using ggplot2[14,19] to illustrate functional interactions. We constructed a pathway-specific interaction network highlighting the expression of GABAergic, glutamatergic, cholinergic, dopaminergic, and serotonergic genes across PBMC subtypes and clinical stages using Cytoscape software[20].

Cell-enriched Island using ArchipelaGO

We applied the workflow to identify cell-enriched "islands" associated with specific BPs, using UMAP components as spatial coordinates. This package integrates a gene expression matrix, spatial cell coordinates from UMAP, and user-defined gene sets to identify spatially enriched regions, termed "islands". These islands represent clusters of cells with high enrichment for specific biological processes. We focused on the top five BPs gene sets related to neuroimmune interactions, obtained using the annotationDBi[21] and org.Hs.eg.db[22] packages from Bioconductor.

Initially, enrichment scores for each cell were calculated using over-representation analysis (ORA), with Fisher's exact test employed to determine statistical significance. The p-values were transformed into enrichment scores defined as -log(p). These scores were then used to generate spatial enrichment landscapes via kernel density estimation, where UMAP components and respective scores were integrated to create a continuous topographical map.Islands were identified by applying a user-defined threshold ("water level") to the topographical map, delineating regions of significant enrichment. Each island was further characterized by its cellular composition, determined based on the cell types present within its polygonal contours.

Metanalysis of transcriptomic data.

Eligible studies were identified using the Gene Expression Omnibus (GEO)[23] by searching for peer-reviewed datasets indexed with the terms "HTLV-1," "Homo sapiens," and expression profiles from “high throughput sequencing” (bulk or single-cell RNA-seq) and “expression profiling by array.” Studies had to contain healthy controls and include a minimum of 8 total samples. Initially, 64 studies reporting transcriptional data from human subjects were identified (**Fig. 4**).

Studies were excluded based on the following criteria: (1) therapeutic-focused analyses (n = 12); (2) use of cultured cell lines, tissues, or genetically modified viruses (n = 30); (3) fewer than three infected samples (n = 11); (4) duplicated or replicated samples (n = 3); (5) time-course or longitudinal datasets (n = 3); and (6) super series datasets (n = 1). After applying the exclusion criteria, we incorporated two datasets provided by our collaborators, resulting in a total of five datasets deemed eligible for inclusion in our meta-analysis workflow, along with one additional single-cell dataset.

These datasets included a total of 200 participants: 135 with ATL, 47 ACs, 18 with HAM/TSP, and 45 healthy controls (**Fig. 4**). To comprehensively evaluate the transcriptional overlap between DEGs from scRNA-seq of 233,093 PBMCs and the metaDEGs (obtained through meta-analysis as described below), further analyses were conducted.

The meta-analysis was conducted using the web tool "ExpressAnalyst"[24,25] to quantitatively synthesize results from multiple studies on gene expression. Specifically, the datasets GSE29312, GSE38537, GSE233437,

GSE33615, and EGAD000010007024 (bulk RNAseq) were utilized for the meta-analysis under the conditions asymptomatic, ATL, and HAM/TSP.

Data annotation involved specifying the value type (raw data), data type (microarray or bulk RNAseq), selecting the human organism, and specifying the ID type for all datasets. Missing values were addressed by excluding features with more than 50% missing values and estimating the remaining missing values using KNN imputation (feature-wise). Subsequently, filtering and normalization were performed using a 15% variance filter based on the inter-quantile range and a 5% abundance filter (relative percentile).

Significant genes were identified by combining p-values using Fisher's method. A significance threshold of adjusted p-value < 0.05 and average fold change ≠ 0.0 was applied.

Principal component analysis.

PCA was performed using spectral decomposition with the factoextra[26] and FactoMineR packages[27] in R (v. 4.3.2). Eigenvalues were calculated based on the contributions of neuroimmune genes associated with BPs to assess their influence on PCA. Eigenvalues greater than one were considered critical for determining group segregation[28]. The first two principal components were analyzed to evaluate group separation and the role of neuroimmune genes in distinguishing healthy controls from ATL conditions. The top 20 genes contributing most to the first two dimensions were selected for further analysis.

Relative effect analysis

We evaluated the relative effects of the highest neuroimmune genes identified by PCA (top 20 genes) in ATL compared to a healthy control group using a Multivariate Analysis of Variance (MANOVA) with a bootstrap[29] approach for enhanced reliability, involving 1,000 re-samplings. This method allowed us to account for variability in the data and calculate confidence intervals (CIs) for the relative effects, providing a robust measure of association. The statistical analysis was performed using the R packages npmv[30] and reshape[31].

Gradient boosting machine

We used the Gradient Boosting Machine (GBM) method based on decision tree methodology using gbm R package[32] for feature selection to identify the most relevant genes contributing to the disease. This approach helped reduce the dimensionality of genomic data and focus on the most informative genes[33]. The model was performed using the top 20 high-score neuroimmune genes identified by PCA on 67 samples from GSE36516 (46 samples in the ATL group and 21 samples in the control group). A total of 500 trees were used in the analysis, and the Huberized distribution method was applied. The partial effects of the genes analyzed in the GBM model were examined using all the trees in the model and a grid resolution of 100 points.

Additional data visualization features

To better understand the differences in gene expression between the healthy control group and the ATL condition, we visualized gene expression using violin plots generated with the ggplot2[19] package. Statistical differences between

groups were assessed using the Wilcoxon test, implemented through the stat_compare_means function from the ggpubr package[34].

Ethical Study Approval

Ethical approval for the collection of blood samples from healthy individuals has been granted by the Imperial College Research Ethics Committee (ICREC reference number 6931613). Ethical approval for blood collection from people with HTLV-1 infection has been granted by the Health Research Authority UK (South Central – Oxford C Research Ethics Committee, REC reference: 20/SC/0226). The dataset EGAS00001004936 was generated by Koya *et al,* 2021[36] as part of studies approved by the Institutional Review Board of the National Cancer Center and other participating institutions. The data have been deposited in the European Genome-phenome Archive (EGA) for controlled access and secure sharing.

Flow cytometric analysis

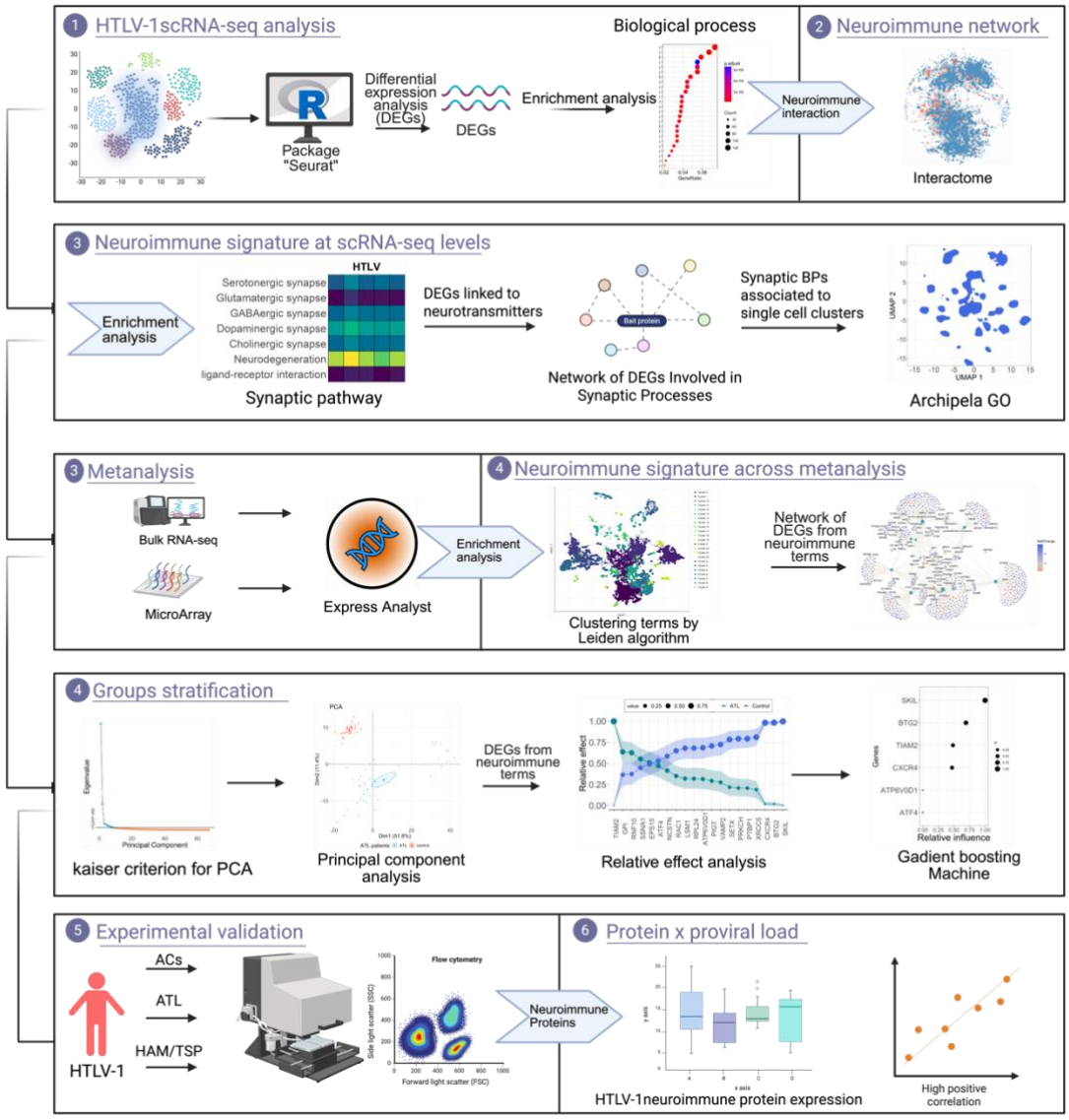
Cryopreserved PBMCs from healthy individuals, ACs, and people with ATL, or HAM/TSP were prepared and stained for analysis. First, the PBMCs were thawed in RPMI 1640 medium, supplemented with human AB serum and Benzonase. After centrifugation and resuspension in RPMI 1640medium supplemented with 2mM L-Glutamine, 5% Human Serum AB and 1x Pen/Strep; the cells were rested at 37°C for 2 hours. The cells were first stained with Fixable Blue viability dye, followed by Fc receptor blocking and a 30 min incubation with a chemokine receptor antibody for CXCR4 (12G5, Biolegend) at 37°C. These were further stained for surface markers: CD3 (UCHT1, Biolegend), CD4 (S3.5,

Thermofisher), CD45RA (HI100, Biolegend), CD45RO (UCHL1, Biolegend), CD16 (3G8 RUO, BD), CD14 (M5E2 RUO, BD) and CD20 (2H7, Biolegend); and fixed using the FoxP3 fixation and staining buffers (Thermo Scientific). After fixation, the cells underwent a permeabilization step with Perm/Wash solution, followed by intracellular blocking and intracellular staining using antibodies targeting SKIL (OTI3E2, Novus Biologicals), ATF4 (S360A-24, Novus Biologicals), V-ATPase D1 (*ATP6V0D1*; D-4, Santa Cruz Biotechnology), Senataxin (*SETX;* 4G1*,* Novus Biologicals), PTBP1 (3D6D8, Thermofisher), VAMP2 (EPR12790, Abcam), and Ku80 (*XRCC5;* 2G5E7, Thermofisher). This staining pipeline had been previously established and validated by our collaborators[35–37].

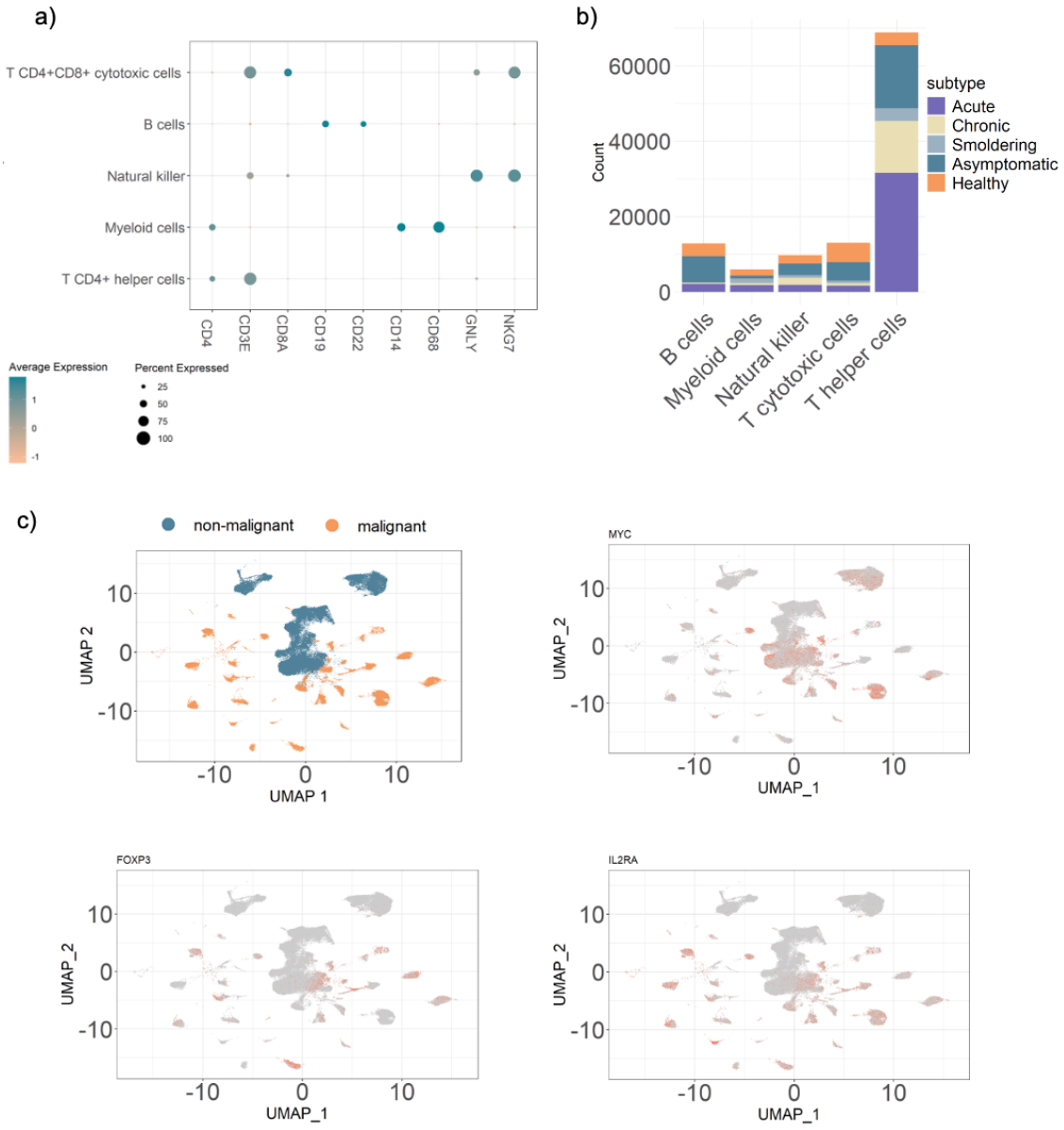
Ethical Study Approval

Ethical approval for the collection of blood samples from healthy individuals has been granted by the Imperial College Research Ethics Committee (ICREC reference number 6931613). Ethical approval for blood collection from people with HTLV-1 infection has been granted by the Health Research Authority UK (South Central – Oxford C Research Ethics Committee, REC reference: 20/SC/0226). The dataset EGAS00001004936 was generated by Koya *et al,* 2021[36] as part of studies approved by the Institutional Review Board of the National Cancer Center and other participating institutions. The data have been deposited in the European Genome-phenome Archive (EGA) for controlled access and secure sharing.

Supplementary Figures

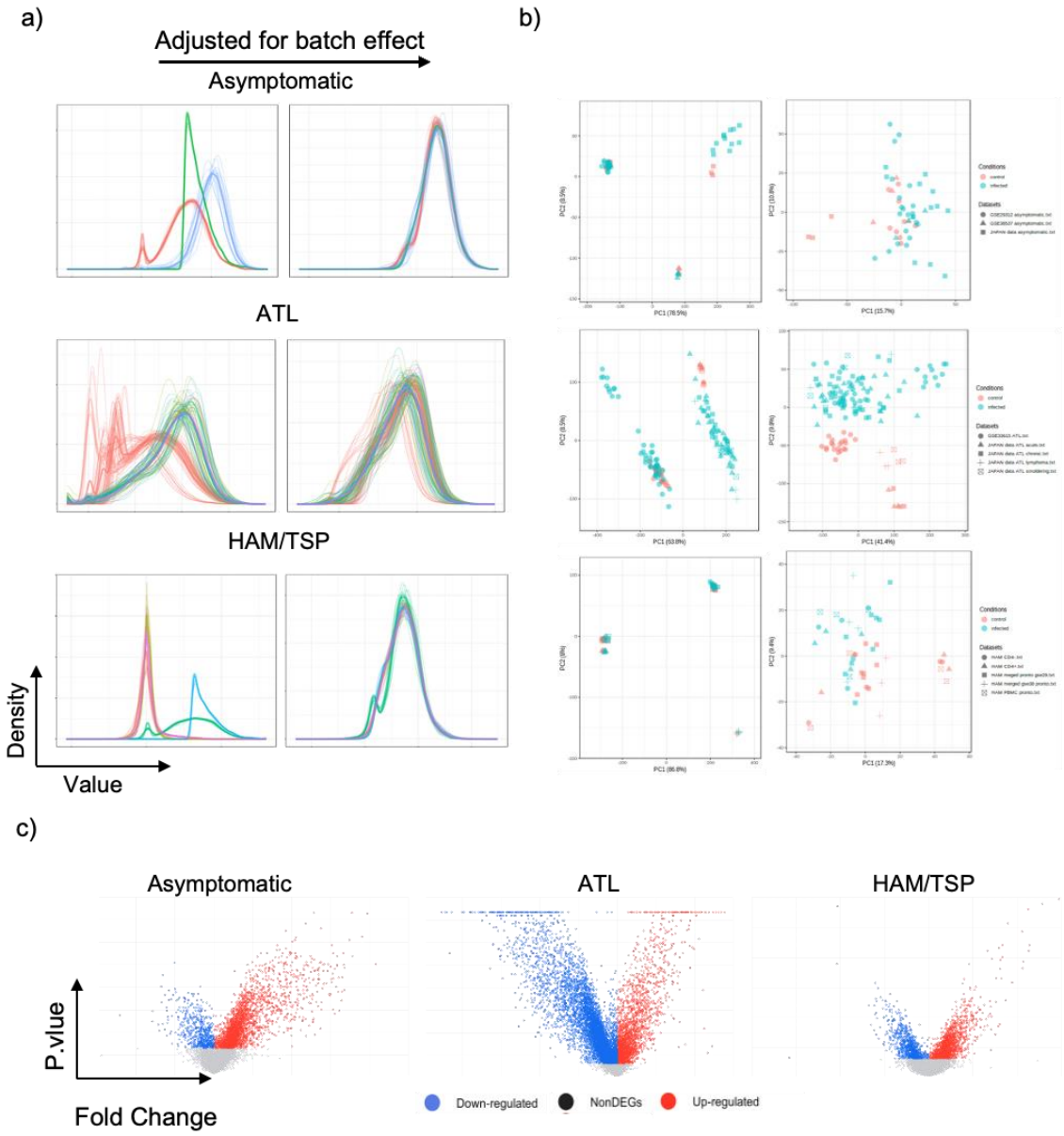


Supplementary Figure 01: Systems biology workflow illustrates the integrative analytical pipeline used in this study as described in the Materials and Methods.



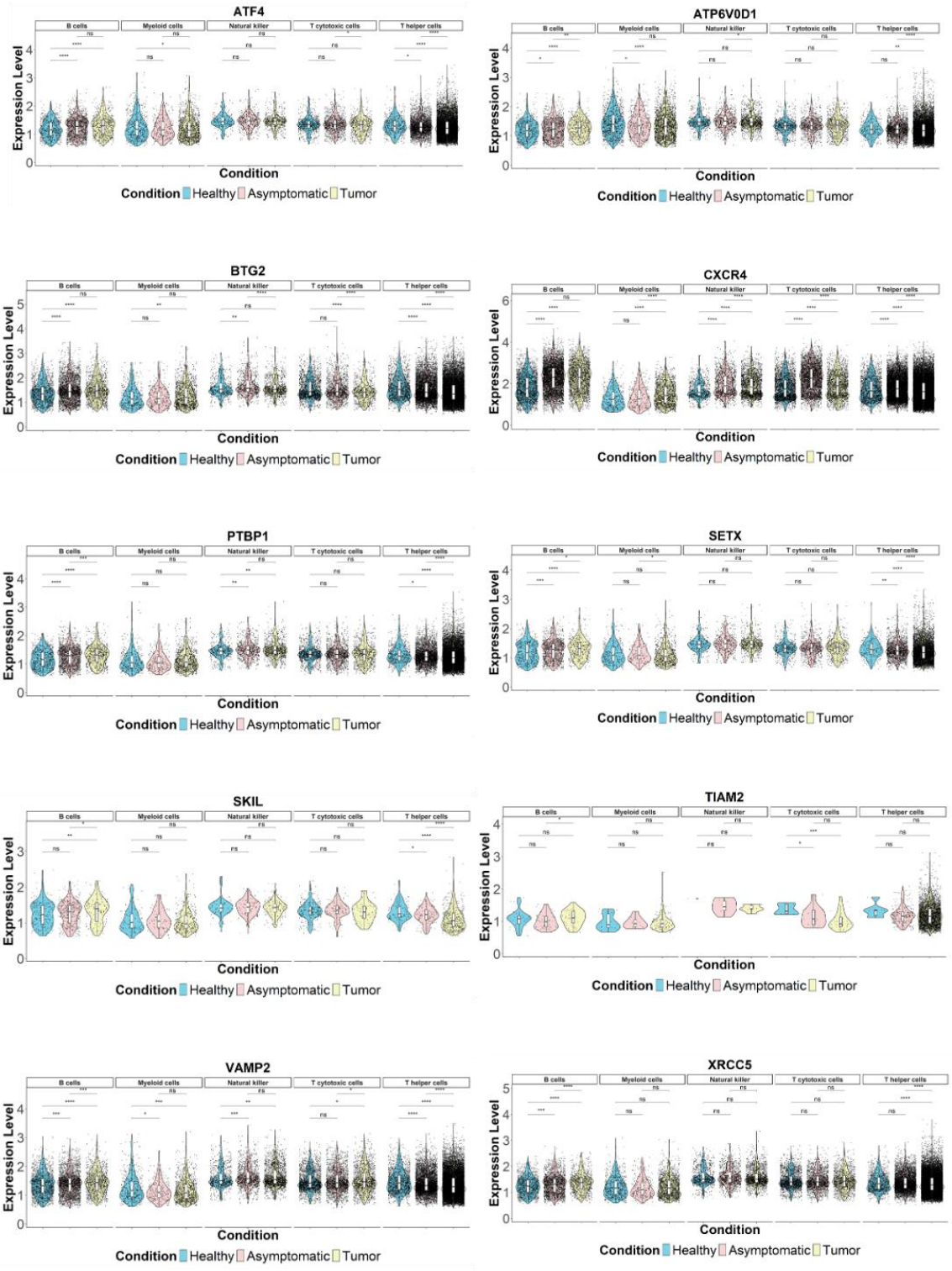
Supplementary Figure 2: Neuro-Immune Gene Expression Profiles Associated with Malignant Cells in Single-Cell Data. (a) Dot plot showing the expression of selected immune markers across major immune cell types, highlighting the specificity of genes such as *CD3E*, *CD4*, *CD14*, *CD19*, *CD68*, and *NKG7*. (**b**) Barplot showing the distribution of major immune cell subsets stratified by clinical condition (Healthy, Asymptomatic, Smoldering, Chronic, Acute). (**c)** UMAP feature plots displaying the expression of selected genes

markers (*MYC*, *FOXP3*, *IL2RA*, *PLCG1*), indicating their association with malignant cell clusters.

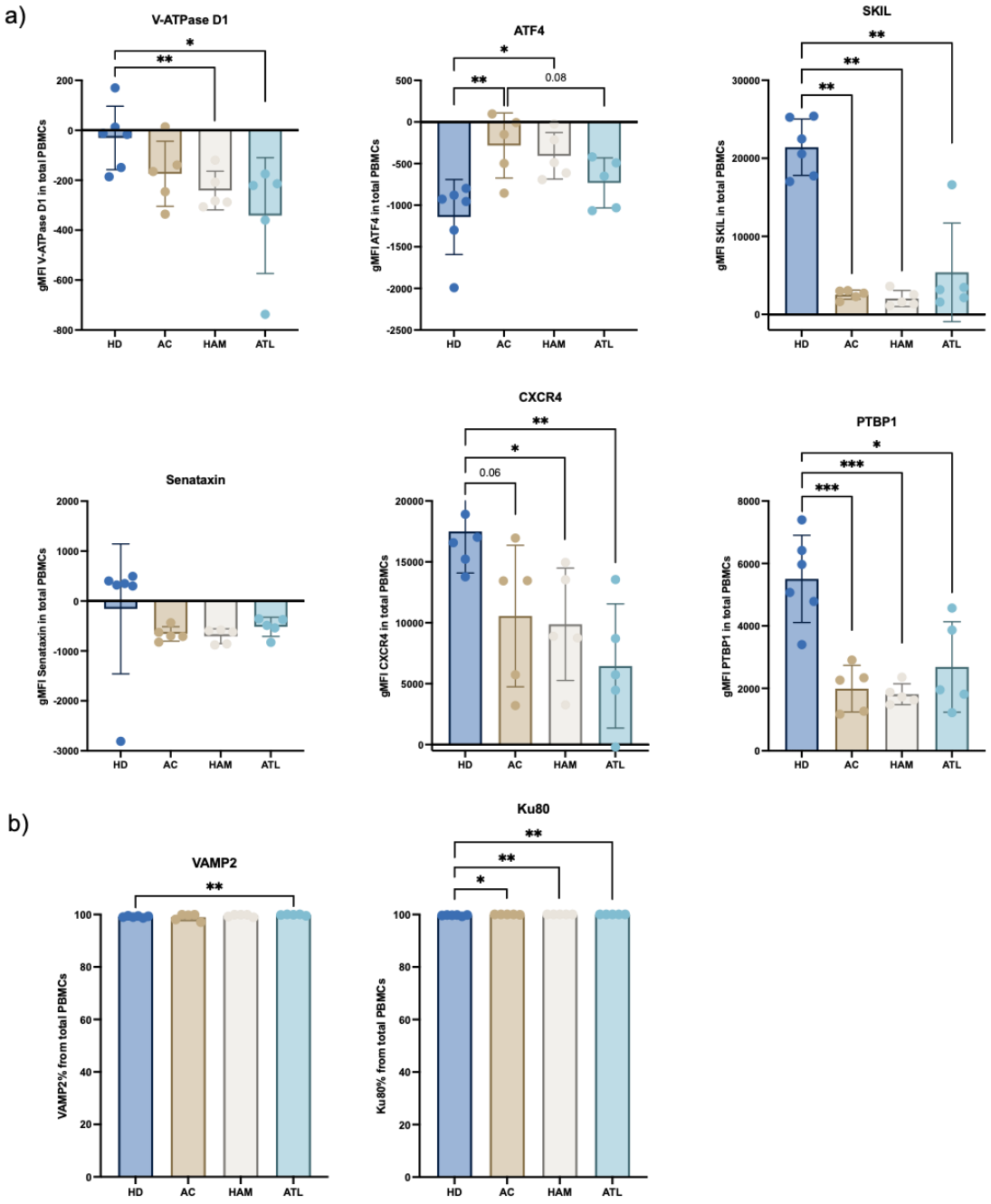


Supplementary Figure 3: Meta-Analysis Reveals Transcriptional Patterns Across Multiple Conditions. (a) Density plots showing distributions before and after batch effect correction, highlighting the harmonization of expression profiles across datasets and conditions**. (b)** Principal Component Analysis (PCA) plots illustrating the clustering of samples before (left) and after (right) batch effect correction, demonstrating improved data integration across different cohorts and

disease stages. **(c)** Volcano plots of differentially expressed genes (DEGs) for each condition analyzed, showing significantly upregulated (red) and downregulated (blue) genes (adjusted $p < 0.05$).

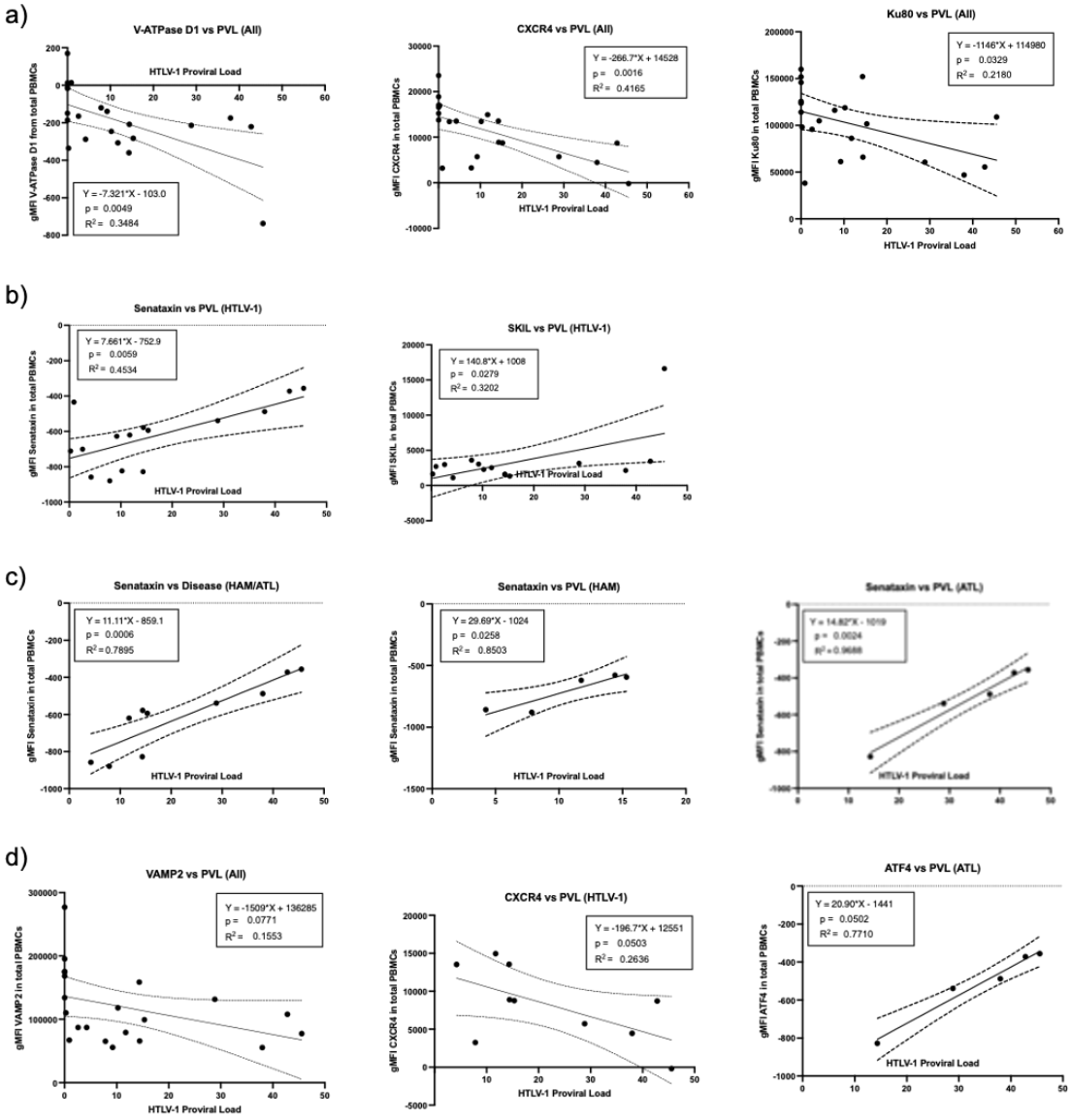


Supplementary Figure 4: Cell-Type-Specific Gene Expression Patterns Reveal Immune Signatures Across HTLV-1 Conditions: Violin plots depict the expression profiles of twelve genes (ATF4, ATP6V0D1, BTG2, CXCR4, PTBP1, SETX, SKIL, TIAM2, VAMP2, and XRCC5) across five major immune cell populations, B cells, myeloid cells, natural killer (NK) cells, T cytotoxic cells, and T helper cells, under three clinical conditions: Asymptomatic (red), Healthy (blue), and Tumor (yellow). The overall distribution of gene expression across all cell types is represented by the black violins. Statistical comparisons between conditions were performed using the Wilcoxon rank-sum test, with significance levels indicated as follows: $p < 0.05$ (), $p < 0.01$ (), $p < 0.001$ (), $p < 0.0001$ (****), and non-significant (ns).



Supplementary Figure 5: Intracellular protein expression in PBMCs from healthy individuals and individuals with HTLV-1 infection. (a) Geometric mean fluorescence intensity (gMFI) of intracellular **V-ATPase D1, ATF4, SKIL, Senataxin (*SETX*), PTBP1** and **CXCR4**, measured by multiparameter flow cytometry in PBMCs from healthy donors (HD), ACs, and patients with HAM/TSP or ATL. **(b)** Frequency of total PBMCs expressing intracellular **VAMP2** and **Ku80 *(XRCC5)***. Both proteins were ubiquitously expressed (>95% of PBMCs). Data

points represent individual donors, with bars showing mean ± SD. Statistical significance was determined using unpaired t-tests with Welch's correction. $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***), $p < 0.0001$ (****).



Supplementary Figure 6: Correlation analysis between HTLV-1 proviral load and intracellular protein expression in PBMCs across disease subgroups. Linear regression plots depicting the correlation between HTLV-1 proviral load (PVL) and the geometric mean fluorescence intensity (gMFI) of selected intracellular proteins in total PBMCs across **(a)** all individuals regardless of HTLV-

1 infection, **(b)** individuals with HTLV-1 infection or **(c)** individuals with symptomatic HTLV-1 infection (HAM/ATL). **(d)** Correlations that did not reach statistical significance but show a trend towards it. Each point represents one subject; regression lines with 95% confidence intervals are shown.

References

1. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.
2. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573-3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.
3. Hao, Y., Stuart, T., Kowalski, M.H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., et al. (2024). Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol *42*, 293–304. https://doi.org/10.1038/s41587-023-01767-y.
4. Koya, J., Saito, Y., Kameda, T., Kogure, Y., Yuasa, M., Nagasaki, J., McClure, M.B., Shingaki, S., Tabata, M., Tahira, Y., et al. (2021). Single-Cell Analysis of the Multicellular Ecosystem in Viral Carcinogenesis by HTLV-1. Blood Cancer Discov *2*, 450–467. https://doi.org/10.1158/2643-3230.BCD-21-0044.
5. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284–287. https://doi.org/10.1089/OMI.2011.0118.
6. Xu, S., Hu, E., Cai, Y., Xie, Z., Luo, X., Zhan, L., Tang, W., Wang, Q., Liu, B., Wang, R., et al. (2024). Using clusterProfiler to characterize multiomics data. Nat Protoc. https://doi.org/10.1038/s41596-024-01020-z.
7. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation *2*, 100141. https://doi.org/10.1016/j.xinn.2021.100141.
8. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 1–14. https://doi.org/10.1186/1471-2105-14-128/FIGURES/3.
9. Kuleshov, M. V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res *44*, W90–W97. https://doi.org/10.1093/nar/gkw377.

10. Xie, Z., Bailey, A., Kuleshov, M. V., Clarke, D.J.B., Evangelista, J.E., Jenkins, S.L., Lachmann, A., Wojciechowicz, M.L., Kropiwnicki, E., Jagodnik, K.M., et al. (2021). Gene Set Knowledge Discovery with Enrichr. Curr Protoc *1*, e90. https://doi.org/10.1002/CPZ1.90.
11. Kanehisa, M. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res *28*, 27–30. https://doi.org/10.1093/nar/28.1.27.
12. Clarke, D.J.B., Jeon, M., Stein, D.J., Moiseyev, N., Kropiwnicki, E., Dai, C., Xie, Z., Wojciechowicz, M.L., Litz, S., Hom, J., et al. (2021). Appyters: Turning Jupyter Notebooks into data-driven web apps. Patterns *2*. https://doi.org/10.1016/j.patter.2021.100213.
13. Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep *9*, 5233. https://doi.org/10.1038/s41598-019-41695-z.
14. Villanueva, R.A.M., and Chen, Z.J. (2019). ggplot2: Elegant Graphics for Data Analysis (2nd ed.). Measurement (Mahwah N J) *17*, 160–167. https://doi.org/10.1080/15366367.2019.1565254.
15. Guangchuang Yu (2024). Package “enrichplot” Title Visualization of Functional Enrichment Result.
16. Mungall, C.J., Torniai, C., Gkoutos, G. V, Lewis, S.E., and Haendel, M.A. (2012). Uberon, an integrative multi-species anatomy ontology. Genome Biol *13*, R5. https://doi.org/10.1186/gb-2012-13-1-r5.
17. Borcherding, N., Vishwakarma, A., Voigt, A.P., Bellizzi, A., Kaplan, J., Nepple, K., Salem, A.K., Jenkins, R.W., Zakharia, Y., and Zhang, W. (2021). Mapping the immune environment in clear cell renal carcinoma by single-cell genomics. Commun Biol *4*, 122. https://doi.org/10.1038/s42003-020-01625-6.
18. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences *102*, 15545–15550. https://doi.org/10.1073/pnas.0506580102.
19. Wickham, H. (2016). ggplot2 (Springer International Publishing) https://doi.org/10.1007/978-3-319-24277-4.
20. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res *13*, 2498–2504. https://doi.org/10.1101/gr.1239303.
21. Pagès, H., Carlson, M., Falcon, S., and Li, N. (2024). AnnotationDbi: Manipulation of SQLite-based annotations in Bioconductor. Preprint.
22. Carlson, M. (2024). org.Hs.eg.db: Genome wide annotation for Human. Preprint.
23. Clough, E., Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., et al. (2024). NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update. Nucleic Acids Res *52*, D138–D144. https://doi.org/10.1093/nar/gkad965.

24. Liu, P., Ewald, J., Pang, Z., Legrand, E., Jeon, Y.S., Sangiovanni, J., Hacariz, O., Zhou, G., Head, J.A., Basu, N., et al. (2023). ExpressAnalyst: A unified platform for RNA-sequencing analysis in non-model species. Nat Commun *14*, 2995. https://doi.org/10.1038/s41467-023-38785-y.
25. Ewald, J., Zhou, G., Lu, Y., and Xia, J. (2023). Using ExpressAnalyst for Comprehensive Gene Expression Analysis in Model and Non-Model Organisms. Curr Protoc *3*. https://doi.org/10.1002/cpz1.922.
26. Kassambara, A., and Mundt, F. (2020). factoextra: Extract and Visualize the Results of Multivariate Data Analyses. Preprint.
27. Lê, S., Josse, J., and Husson, F. (2008). FactoMineR : An *R* Package for Multivariate Analysis. J Stat Softw *25*. https://doi.org/10.18637/jss.v025.i01.
28. Kaiser, H.F. (1960). The Application of Electronic Computers to Factor Analysis. Educ Psychol Meas *20*, 141–151. https://doi.org/10.1177/001316446002000116.
29. Xu, L.-W. (2015). Parametric bootstrap approaches for two-way MANOVA with unequal cell sizes and unequal cell covariance matrices. J Multivar Anal *133*, 291–303. https://doi.org/10.1016/j.jmva.2014.09.015.
30. Burchett, W.W., Ellis, A.R., Harrar, S.W., and Bathke, A.C. (2017). Nonparametric Inference for Multivariate Data: The *R* Package npmv. J Stat Softw *76*. https://doi.org/10.18637/jss.v076.i04.
31. Wickham, H. (2007). Reshaping Data with the reshape Package. J Stat Softw *21*. https://doi.org/10.18637/jss.v021.i12.
32. Ridgeway, G., and Developers, G. (2003). gbm: Generalized Boosted Regression Models. Preprint, https://doi.org/10.32614/CRAN.package.gbm https://doi.org/10.32614/CRAN.package.gbm.
33. Boldini, D., Grisoni, F., Kuhn, D., Friedrich, L., and Sieber, S.A. (2023). Practical guidelines for the use of gradient boosting for molecular property prediction. J Cheminform *15*, 73. https://doi.org/10.1186/s13321-023-00743-7.
34. Kassambara, A. (2016). ggpubr: "ggplot2" Based Publication Ready Plots. Preprint, https://doi.org/10.32614/CRAN.package.ggpubr https://doi.org/10.32614/CRAN.package.ggpubr.
35. Jones, E.M., Sourij, C., Stradner, M., Schlenke, P., Sereban, N., Moser, O., Quinlan, R., Short, C.-E., Harris, B.H.L., Fertleman, M., et al. (2025). Accelerated memory T cell decline and tolerogenic recall responses to SARS-CoV-2 vaccination in diabetes. Preprint, https://doi.org/10.1101/2025.02.08.637262 https://doi.org/10.1101/2025.02.08.637262.
36. Coulombeau, R., Selck, C., Giang, N., Al-Mohammad, A., Ng, N., Maher, A.K., Argüello, R., Scalfari, A., Varley, J., Nicholas, R., et al. (2025). <scp>Sphingosine</scp> -1-Phosphate Signalling Inhibition Suppresses Th1-Like Treg Generation by Reversing Mitochondrial Uncoupling. Immunology *174*, 153–166. https://doi.org/10.1111/imm.13870.
37. Maher, A.K., Aristodemou, A., Giang, N., Tanaka, Y., Bangham, C.R.M., Taylor, G.P., and Dominguez-Villar, M. (2024). HTLV-1 induces an inflammatory CD4+CD8+ T cell population in HTLV-1–associated myelopathy. JCI Insight *9*. https://doi.org/10.1172/jci.insight.173738.