HTLV-1 infection promotes excessive T cell activation and transformation into adult T cell leukemia/lymphoma

Benjy J.Y. Tan,[1,2,3] Kenji Sugata,[1] Omnia Reda,[1,2,3,4] Misaki Matsuo,[1,2] Kyosuke Uchiyama,[5] Paola Miyazato,[2] Vincent Hahaut,[6,7] Makoto Yamagishi,[8] Kaoru Uchimaru,[8] Yutaka Suzuki,[9] Takamasa Ueno,[10] Hitoshi Suzushima,[11] Hiroo Katsuya,[2,12] Masahito Tokunaga,[13] Yoshikazu Uchiyama,[14] Hideaki Nakamura,[15] Eisaburo Sueoka,[16] Atae Utsunomiya,[13,17] Masahiro Ono,[2,18] and Yorifumi Satou[1,2]

[1]Division of Genomics and Transcriptomics, Joint Research Center for Human Retrovirus Infection, [2]International Research Center for Medical Sciences (IRCMS), and [3]Graduate School of Medical Sciences, Kumamoto University, Kumamoto, Japan. [4]Department of Microbiology, High Institute of Public Health, Alexandria University, Alexandria, Egypt. [5]School of Medicine, Kumamoto University, Kumamoto, Japan. [6]Institute of Molecular and Clinical Ophthalmology Basel, Basel, Switzerland. [7]Department of Ophthalmology, University of Basel, Basel, Switzerland. [8]Laboratory of Tumor Cell Biology, Department of Computational Biology and Medical Sciences and [9]Laboratory of Systems Genomics, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. [10]Division of Infection and Immunity, Joint Research Center for Human Retrovirus Infection, Kumamoto University, Kumamoto, Japan. [11]Department of Hematology, Kumamoto Shinto General Hospital, Kumamoto, Japan. [12]Division of Hematology, Respiratory Medicine and Oncology, Saga University, Saga, Japan. [13]Department of Hematology, Imamura General Hospital, Kagoshima, Japan. [14]Division of Informative Clinical Sciences, Faculty of Life Sciences, Kumamoto University, Kumamoto, Japan. [15]Department of Transfusion Medicine and [16]Department of Clinical Laboratory Medicine, Faculty of Medicine, Saga University, Saga, Japan. [17]Graduate School of Medical and Dental Sciences, Kagoshima University, Kagoshima, Japan. [18]Department of Life Sciences, Imperial College London, London, United Kingdom.

Human T cell leukemia virus type 1 (HTLV-1) mainly infects CD4[+] T cells and induces chronic, persistent infection in infected individuals, with some developing adult T cell leukemia/lymphoma (ATL). HTLV-1 alters cellular differentiation, activation, and survival; however, it is unknown whether and how these changes contribute to the malignant transformation of infected cells. In this study, we used single-cell RNA-sequencing and T cell receptor–sequencing to investigate the differentiation and HTLV-1–mediated transformation of T cells. We analyzed 87,742 PBMCs from 12 infected and 3 uninfected individuals. Using multiple independent bioinformatics methods, we demonstrated the seamless transition of naive T cells into activated T cells, whereby HTLV-1–infected cells in an activated state further transformed into ATL cells, which are characterized as clonally expanded, highly activated T cells. Notably, the greater the activation state of ATL cells, the more they acquire Treg signatures. Intriguingly, the expression of HLA class II genes in HTLV-1–infected cells was uniquely induced by the viral protein Tax and further upregulated in ATL cells. Functional assays revealed that HTLV-1–infected cells upregulated HLA class II molecules and acted as tolerogenic antigen-presenting cells to induce anergy of antigen-specific T cells. In conclusion, our study revealed the in vivo mechanisms of HTLV-1–mediated transformation and immune escape at the single-cell level.

Introduction

Human T cell leukemia virus type 1 (HTLV-1) was the first human retrovirus to be reported (1) and infects at least 5–10 million people worldwide. HTLV-1 inserts its genome into CD4[+] T cells and persists as a provirus in the host cells. After integration, the provirus mainly remains latent and rarely produces viral particles. Consequently, a majority of the infected individuals remain asymptomatic carriers (ACs) throughout their lifetime (2). However, decades after infection, approximately 5%–10% of the infected individuals develop a CD4[+] T cell leukemia/lymphoma, termed adult T cell leukemia/lymphoma (ATL). The underlying pathogenic mechanisms of ATL remain elusive. ATL has 2 major subtypes: (a) aggressive-type ATL, characterized by fast-growing leukemic cells, including acute ATL and lymphoma-type ATL; and (b) indolent-type ATL, with slow disease progression, including chronic ATL and smoldering ATL (3). Both ATL subtypes have long periods of incubation, which hampers experimental investigations.

Classical studies have defined CD25 (IL-2 receptor α chain) as a marker for ATL cells (4). *CD25* expression is induced in normal T cells upon antigenic recognition and subsequent T cell receptor (TCR) signaling. In addition, Tregs, a subset of CD4[+] T cells, constitutively express *CD25*, and this expression is further upregulated when they are activated. Recent studies have identified FOXP3 (5, 6), which is also expressed in Tregs and some activated T cells (7–9), as a marker for ATL cells. These results suggest that ATL transformation involves pathways associated with either activated T cells, Tregs, or both. However, the expression of these markers varies among and within patients (5, 10). In addition, these molecules are dynamically regulated at a single–T cell level in normal T cells during both the steady state and T cell responses.

Meanwhile, the viral genes *tax* and *hbz* have been identified to play key roles in leukemic transformation (11). Tax activates multiple cellular signaling pathways including NF-κB and enhances the survival of host cells (12). HTLV-1 bZIP factor (HBZ) promotes

CD4+ T cell hyperproliferation and differentiation toward the effector/memory phenotype (13, 14). Intriguingly, activating mutations in the ATL genome preferentially accumulate in genes associated with the NF-κB signaling pathway, such as *PLCG1*, *PRKCB*, and *CARD11*, which are a part of TCR signaling (15). However, it is not known how these mutations and viral proteins synergistically and/or cooperatively lead to ATL over the course of the latency period.

These 2 lines of evidence support the idea that HTLV-1 infection and subsequent leukemogenesis are closely associated with the TCR signaling pathway and CD4+ T cell activation and differentiation. Since these processes are dynamically regulated in uninfected and infected T cell populations, single-cell analysis is necessary to fully understand how HTLV-1 infection controls the physiological pathways in T cells and transforms them into ATL cells. In this study, we cross-analyzed healthy donors (HDs), ACs, and patients with ATL using single-cell RNA-Seq (scRNA-Seq) and TCR-Seq to identify the pathways responsible for the in vivo transformation of infected T cells into leukemic cells at the single-cell level.

Results

Single-cell transcriptomic analysis of PBMCs from HTLV-1–infected individuals. In this study, we focused on indolent-type ATL, which has a relatively slow disease progression. In addition, we included healthy individuals to concurrently analyze physiological T cell activation and differentiation with the disease progression and oncogenic transformation in patients with ATL. We generated droplet-based 5′ scRNA-Seq and TCR-Seq libraries from PBMCs obtained from 16 individuals, 4 of whom were ACs (AC1–AC4); 9 were patients with ATL ($n = 3$ with smoldering ATL, SML1–SML3; $n = 5$ with chronic ATL, ATL1–ATL7; and $n = 1$ with lymphoma-type ATL, ATL4); and 3 were HDs (HD1–HD3; Figure 1A and Supplemental Table 1; supplemental material available online with this article; https://doi.org/10.1172/JCI150472DS1). Collectively, the scRNA-Seq profiles from 87,742 cells passed quality control and consisted of 19,903 cells from ACs; 16,357 from patients with SML (SMLs); 40,119 cells from patients with ATL (ATLs); and 11,363 cells from HDs (Supplemental Figure 1A). We identified 25 single-cell clusters on the basis of scRNA-Seq profiles, including T cells, B cells, NK cells, myeloid cells, megakaryocytes, and erythrocytes (Supplemental Figure 1, B–D).

Next, we performed in silico sorting and clustering analysis of T cells to understand the dynamics of T cell differentiation and ATL transformation at the single-cell level. In total, we obtained 46,078 T cells, which included 10,576 from ACs; 7514 from SMLs; 23,845 from ATLs; and 4143 from HDs (Figure 1B). We identified 15 distinct T cell clusters (Figure 1C). Interestingly, cells from HDs, ACs, and SMLs were mixed together in different clusters (H1–H2 and C1–C3), whereas cells from ATLs formed distinct clusters on their own (Supplemental Figure 2, A and B). We identified 3 CD8+ cytotoxic T cells clusters (C1–C3), while the remaining clusters were classified as CD4+ helper T cells on the basis of their key gene expression (Supplemental Figure 2C). We further confirmed our annotations by mapping our data to a reference data set (Supplemental Figure 3, A and B). Cells contributing to clusters I1–I4 and A1–A6 were identified as HTLV-1–infected cells (*HTLV*-sense+*hbz*+*CADM1*+*CCR4*+*CD40LG*lo/−*CD7*lo/−*DPP4*lo/−; Supplemental Figure 2D and refs. 16, 17). To confirm our annotation for HTLV-1–infected

cells, we measured and compared the proviral load (PVL) between CADM1−CD7+ cells and CADM1+CD7−/+ cells and confirmed that CADM1+CD7−/+ cells were HTLV-1 infected as evidenced by the high PVLs (>60%; Supplemental Figure 4C). Since ATL cells are known to be highly expanded T cell clones, we analyzed paired single-cell TCR sequences to identify clonally expanded cells. We noted that among all clusters derived from ATLs, several of the clusters primarily consisted of cells with a single, highly expanded clone (clusters A1–A6; Supplemental Figure 4A). Intriguingly, some cells from HDs clustered together with HTLV-1–infected cells from multiple ATL patients (Supplemental Figure 2A and Supplemental Figure 4B). This suggests that some T cell clones from patients with ATL had a transcriptional profile similar to that of normal T cells in healthy individuals.

Next, we performed differential expression analysis to identify the characteristics of cells belonging to the infected non-ATL clusters and infected ATL clusters. Pathway analysis of upregulated genes showed that these clusters were significantly enriched in genes responsible for TCR- and activation-related pathways, such as those for TCR signaling, antigen presentation, and immune checkpoint signaling (Figure 2A). To identify the key genes accounting for the TCR-related pathways, we assessed the expression profile of all genes in the 10 shared pathways and found a strong enrichment of HLA class II genes among infected/ATL clusters (Figure 2B). A similar analysis of the downregulated genes showed neither common pathways nor enriched genes among these clusters (Supplemental Figure 5, A and B).

ATL cells highly express costimulatory receptors and genes downstream of TCR signaling. Although most ATL cells highly express *IL2RA* (CD25), a key feature for activated T cells and Tregs, they may show variable expression of genes used in activated T cells and Treg markers. We hypothesized that ATL cells in each patient are heterogenous, showing a broad spectrum of the gene expression profiles of activated T cells and Tregs. To test this hypothesis, we applied canonical correspondence analysis (CCA), which is a multidimensional method to quantitatively analyze gene expression profiles associated with the differentiation or activation status of single cells in a data-oriented manner, to the CD4+ T cell data set (18). We first analyzed the CD4+ T cell data using the reference RNA-Seq data for activated T cells to evaluate the "T cell activation score" of individual ATL and non-ATL cells. The analysis showed that the majority of CD4+ T cells from HDs had low T cell activation scores, which is compatible with their naive status, whereas T cells from HTLV-1–infected individuals had higher scores, with T cells from patients with ATL having the highest scores (Figure 3A, left panel, and Supplemental Figure 6A, top row). We next used reference RNA-Seq data for Tregs to analyze the "Treg score" for individual ATL and non-ATL cells and found a similar trend; T cells from HDs showed low Treg scores, whereas cells from patients with SML or ATL had remarkably high Treg scores (Figure 3A, right panel, and Supplemental Figure 6A, bottom row). A 2D plot of the 2 CCA scores showed that, while HDs harbored T cells with an activated phenotype, infected T cells and ATL cells spontaneously became activated, acquired a Treg phenotype, and subsequently progressed to a state of extreme activation, which was maintained throughout the ATL phase (Figure 3B and Supplemental Figure 6B). The dynamics of transcriptional activities for Tregs and activated T cells
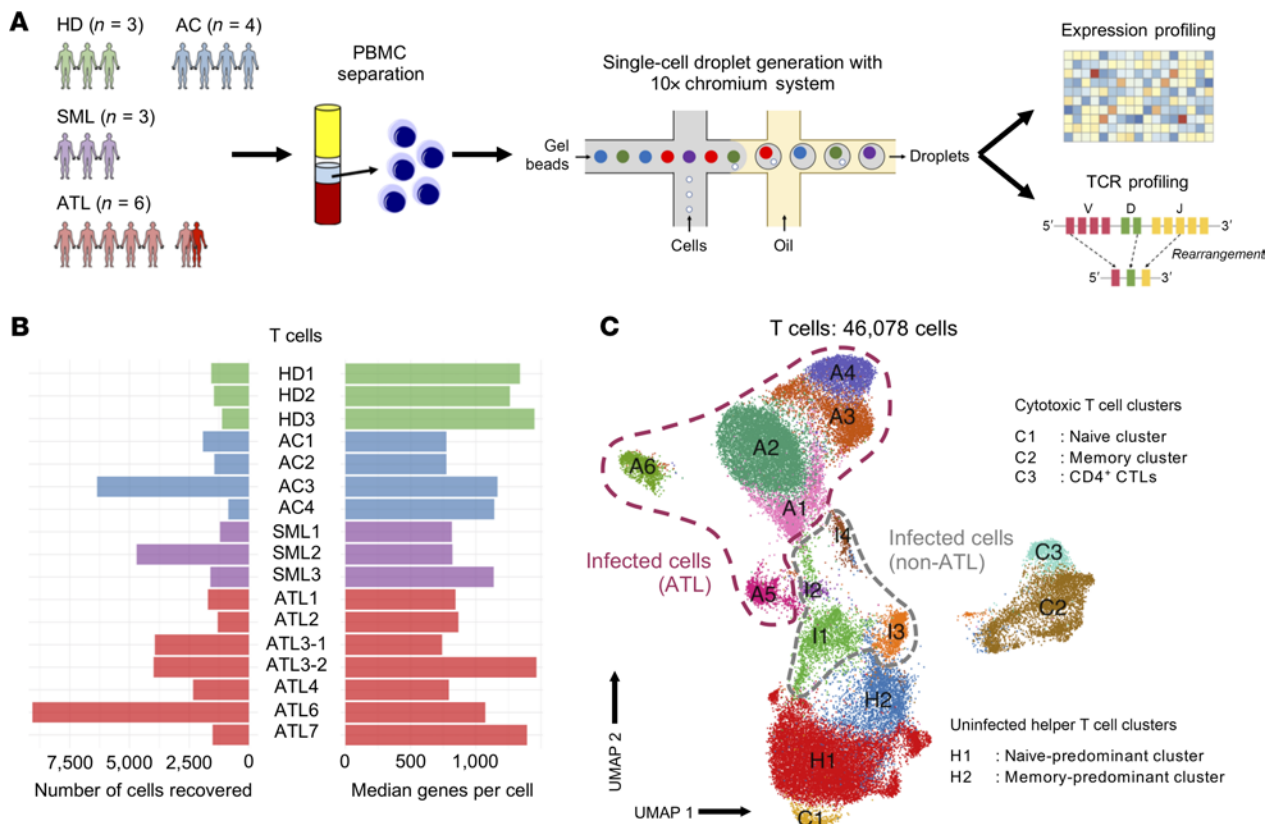
Figure 1. Single-cell transcriptional profiling of HTLV-1–infected individuals. (**A**) Schematic representation of the experimental workflow. (**B**) The number of T cells recovered that passed quality control (left) and their median number of genes per cell (right) for each of the individuals (*n* = 3 HDs, HD1–HD3; *n* = 4 ACs, AC1–AC4; *n* = 3 SMLs, SML1–SML3; and *n* = 7 ATLs, ATL1–ATL7). (**C**) 2D UMAP visualization showing the 4 different groups of T cells we identified.

was further confirmed by 2D-CCA analysis, in which we used both of the variables for T cell activation and Tregs. We found most of the ATL cells in the quadrant between the T cell activation and Treg axes, indicating that they were both highly activated and acquired a Treg phenotype (Supplemental Figure 6C).

Additionally, we analyzed the expression of costimulatory and coinhibitory molecules (i.e., immune checkpoints), defining the activation and exhaustion signatures, respectively. In general, HTLV-1–infected cells had higher activation and exhaustion signatures compared with noninfected and normal T cells (Supplemental Figure 5C). Among the infected cells, the 2 signatures were most pronounced in ATL cells (Supplemental Figure 5C). Interestingly, the activation signature was highly variable between HTLV-1–infected and ATL clusters, and none of the genes were consistently expressed by all the infected and ATL clusters. On the other hand, most of the HTLV-1–infected and ATL clusters expressed similar exhaustion signature genes including *TIGIT*, *CTLA4*, and *LAG3* (Supplemental Figure 5D). This indicates that infected/ATL cells have a broad spectrum of activation states while harboring a phenotype of chronically stimulated, exhausted T cells.

We earlier postulated that by using cells from patients diagnosed with indolent ATL, we would be able to capture cells from different stages of oncogenic transformation and disease progression. To address this hypothesis, we applied a pseudotime analysis to the entire CD4+ T cell population, regardless of the sample identity. Pseudotime analysis showed that uninfected helper

T cells clustered in early pseudotime, while infected/ATL cells clustered around the middle and end of the pseudotime trajectory (Figure 3, C and D, and Supplemental Figure 7, A and B). Interestingly, the trajectory of the cells bifurcated into 2 different branches, suggesting that there were 2 phenotypically different ATL cells. We then assessed how gene expression is regulated across the pseudotime axis using the CCA scores for T cell activation and Tregs. We found that both of the CCA scores gradually and progressively increased along the trajectory, indicating that the transformation of healthy T cells into ATL cells involves activation of infected cells and acquisition of the Treg phenotype (Figure 3E). Next, we analyzed changes in gene expression across the pseudotime. As expected, T cells upregulated genes related to HTLV-1 infection, such as *CADM1*, in the latter half of the pseudotime axis. Intriguingly, T cells from healthy individuals upregulated activation- and Treg-related markers, such as *IL2RA* and *FOXP3*, along the pseudotime axis, which was seamlessly followed by HTLV-1–infected cells and ATL cells toward the end of the pseudotime axis. On the contrary, markers that are associated with naive cells, such as *CCR7* and *IL7R*, were downregulated along the pseudotime trajectory in cells from healthy and infected individuals and from patients with ATL (Figure 3F). These results collectively indicate that our inferred pseudotime trajectory successfully captured the normal T cell development/differentiation and HTLV-1–mediated T cell transformation processes as a continuous physiological and pathological pathway.
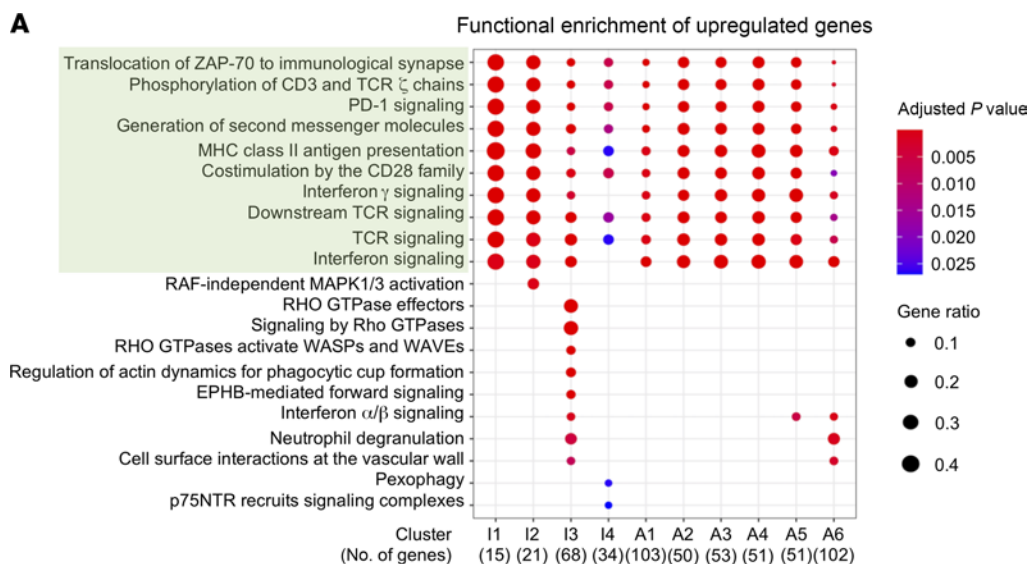
Figure 2. HTLV-1–infected and ATL cells upregulate HLA class II gene expression. (**A**) Dot plot shows the shared and distinct reactome pathways of the upregulated genes for each infected cluster. Common pathways between all clusters are highlighted in green. (**B**) Heatmap shows the expression of the genes participating in the common pathways (highlighted in green in **A**) for all clusters.
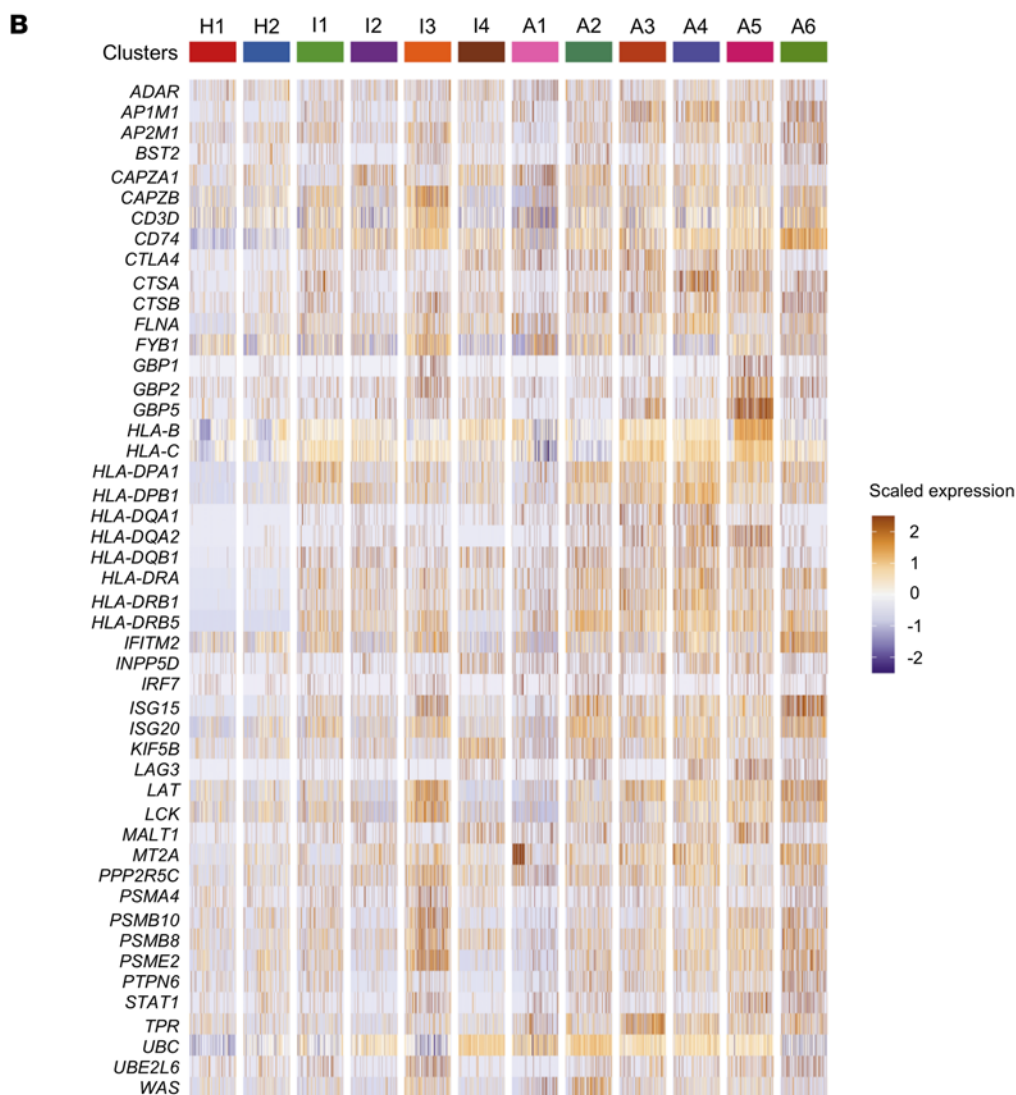
Pseudotime analysis reveals that HLA class II– and Treg-associated genes are upregulated along the ATL trajectory. After performing a collective analysis of all CD4+ T cells, we observed that there were at least 2 phenotypically different ATL cells (Figure 3C). We performed differential expression analysis to evaluate the difference between these 2 branches. Pathway analysis showed that cells in the top branch upregulated genes related to transcription and endoplasmic reticulum stress, while genes upregulated in cells of the bottom branch were involved in metabolic process and TNF signaling (Figure 3G). However, as we showed in Figure 1C, ATL cells mainly formed separate clusters with some overlaps, indicating the large interindividual variability between ATL cells. This large interindividual variation of ATL cells might obscure the unique dynamics of ATL cell transformation in each individual. Therefore, we performed pseudotime analysis by combining cells from HDs (*n* = 3) and individual ATL patients, thereby producing 5 pseudotime data sets (Figure 4A). All 5 data sets successfully identified the trajectory of ATL cells: cells from HDs mainly clustered at the early pseudotime, whereas cells from patients with ATL clustered in the middle and end (Figure 4B). The expression kinetics of T cell–related genes showed that, along the pseudotime trajectory, the expression of genes related to a naive state, such as *CCR7*, was downregulated, while genes related to HTLV-1 infection (*CADM1*) and T cell activation and the Treg phenotype (*IL2RA*, *FOXP3*, and *CTLA4*) were upregulated (Figure 4C and Supplemental Figure 8). This shows that the earlier segment of the pseudotime recapitulates the developmental/differentiation pathway of normal CD4+ T cells as they transition from a naive to an effector state. This was seamlessly followed by the later segment, which revealed the progression of T cells after HTLV-1 infection, during which infected polyclonal cells transition into highly expanded monoclonal cells (Figure 4, B and C).

In order to identify the common factors contributing to ATL progression, we examined the top 200 genes varying as a function of pseudotime for each data set. We found that the top 200 genes shared some common enriched pathways such as those for the involvement of IFN and TCR signaling and antigen presentation by HLA class II. Among the top 200 genes, 36 were commonly expressed in all patients with ATL and consisted of several HLA class II genes (Figure 5, A and B). The expression of these HLA class II genes increased along the pseudotime axis, indicating that as cells transformed into ATL cells, they upregulated HLA class II gene expression (Figure 5C and Supplemental Figure 8).

HLA class II molecules are mainly found on professional antigen presenting cells (APCs), such as monocytes or B cells, but their expression can also be upregulated in T cells upon stimulation. To investigate the significance of this upregulation, we compared the expression levels of HLA class II genes between professional APCs and the different T cell clusters. We observed that very few uninfected T cells expressed HLA class II genes, and the percentage of HLA class II+ cells increased in HTLV-1–infected cells and was the highest in ATL cells (Figure 5D). We also found an increase in the expression of genes related to HLA class II signaling, such as *CIITA*, *RFX5*, and *CD74*, in infected/ATL cells compared with uninfected cells (Supplemental Figure 10A and refs. 19, 20).

To determine whether the pseudotime analysis above recapitulated the real-time course of disease progression, we next

used a pair of longitudinal samples to perform trajectory inference together with samples from healthy individuals (Figure 6A). Pseudotime analysis showed that the cells from the first time point were distributed in the middle of the pseudotime axis, whereas those from the second time point were mainly clustered toward the end of the pseudotime trajectory (Figure 6B), indicating that the ATL cells at the second time point had a more extreme activated phenotype (Supplemental Figure 9A). RNA velocity analysis also showed a similar pattern, with cells transitioning from the naive and memory state to infected cells, further supporting the results of our pseudotime analysis (Figure 6C). Notably, cells in both branches exhibited a gene expression profile that was similar to that observed for the single time-point samples in Figure 4C, with *CCR7* downregulation and *CADM1* and *FOXP3* upregulation across the pseudotime axis (Supplemental Figure 9D). The expression of HLA class II and its related genes also showed a remarkable increase across the pseudotime axis (Figure 6D), confirming our findings that the upregulation of HLA class II molecules was a key feature of ATL progression.

Interestingly, the trajectory of the cells bifurcated into 2 different branches toward the end of the pseudotime axis, indicating that ATL cells in these 2 branches in the patient were considerably different (Figure 6B). Although the ATL cells in both branches showed the key activation features described above, pathway analysis highlighted the intriguing differences between these 2 branches. To elaborate, T cells in the top branch primarily used genes for pathways related to TNF-mediated signaling (*TMSB4X*) and cellular adhesion (*TNFSF9*), whereas T cells in the bottom branch primarily used genes related to T cell activation (*IL6ST*) and immune signaling (*NFKBIZ*) (Figure 6, E and F).

Next, we evaluated whether the difference in the transcriptomes between the 2 branches was due to genomic changes, resulting in 2 distinct malignant clones in the patient. First, we examined the distribution of the most expanded clone by TCR-Seq for each time point and observed that the clones were similarly distributed between the 2 branches (Figure 6B, right panel, and Supplemental Figure 4A), indicating that the ATL cells in the 2 branches originated from the same TCR clone. To explore the origin of this branching, we first analyzed the genomic landscape of these samples by sorting and analyzing HTLV-1–infected cells (CADM1+CD7−/+) from the samples. We found that for both time points, there was only 1 significantly expanded clone, based on HTLV-1 integration site distribution, indicating that the ATL cells originated from 1 parent clone (Figure 6G, pie charts). Next, we performed targeted exome-sequencing on the sorted cells and identified the somatic mutations present in ATL cells and delineated those that differed between the 2 time points (Figure 6G). Subsequently, we analyzed the frequency of these variant alleles in the scRNA-Seq data. However, no mutations were found between the 2 branches in our analysis (Figure 6H), suggesting that the difference between the cells in the 2 branches was due to epigenetic modifications, metabolic status, and/or their reactivity to environmental signals (e.g., cytokine milieu; refs. 21, 22). It remains to be elucidated in future studies whether the 2 major bifurcated fates have any biological and/or clinical significance.

Expression of HLA class II genes is associated with HTLV-1 infection and the viral protein Tax. We next addressed whether HLA
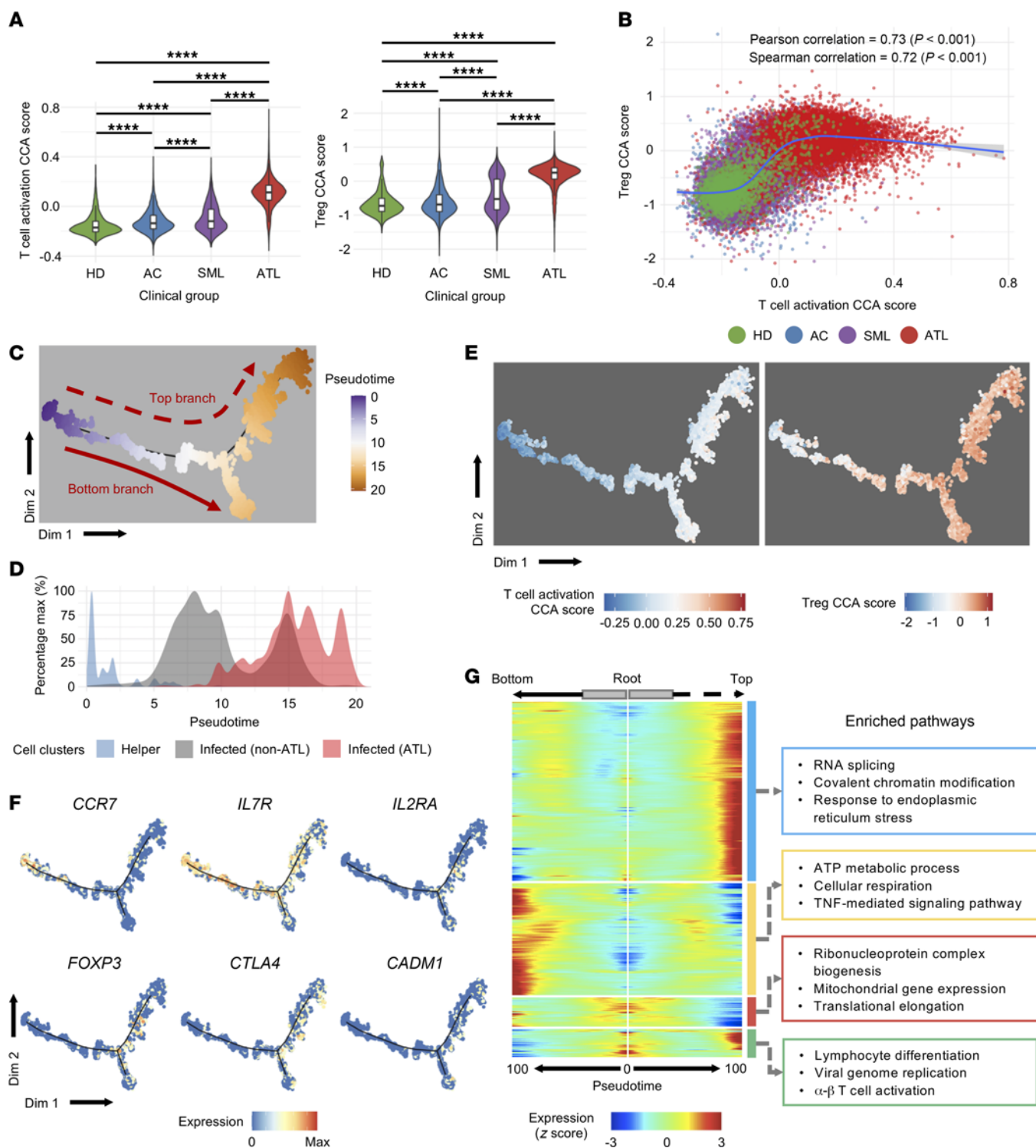
Figure 3. CCA analysis shows that ATL cells acquire the Treg phenotype and are highly activated. (A) Violin plots show the 1D CCA scores for T cell activation and the Treg phenotype, grouped by clinical diagnosis. Box plots in each violin summarize the median (midline) and IQRs. (B) Scatter plot shows the correlation between the 1D CCA score for T cell activation and the Treg phenotype. Blue line shows the regression model of the CCA scores for all cells. (C) Plot shows the pseudotime trajectory for the entire CD4+ T cell population colored by pseudotime. (D) Plot shows the distribution of uninfected helper cells, infected non-ATL cells, and infected ATL cells along the pseudotime axis. (E) Plot shows the distribution of 1D CCA scores for T cell activation and the Treg phenotype in the pseudotime space (for cells from C). (F) Expression of T cell–related marker genes along the pseudotime axis from C. (G) Split heatmap shows the expression profile of genes that vary as a function of pseudotime and are branch dependent. The pseudotime trajectory begins from the middle of the heatmap (gray box) and moves to the left for the bottom branch and to the right for the top branch. The start of the arrow indicates the bifurcation point for the trajectory shown in C. Hierarchical clustering grouped the genes into 4 clusters, indicated by outlined colored boxes on the right, along with the top 3 enriched pathways. ****$P < 0.0001$, by 1-way ANOVA with post hoc Tukey's test (A).
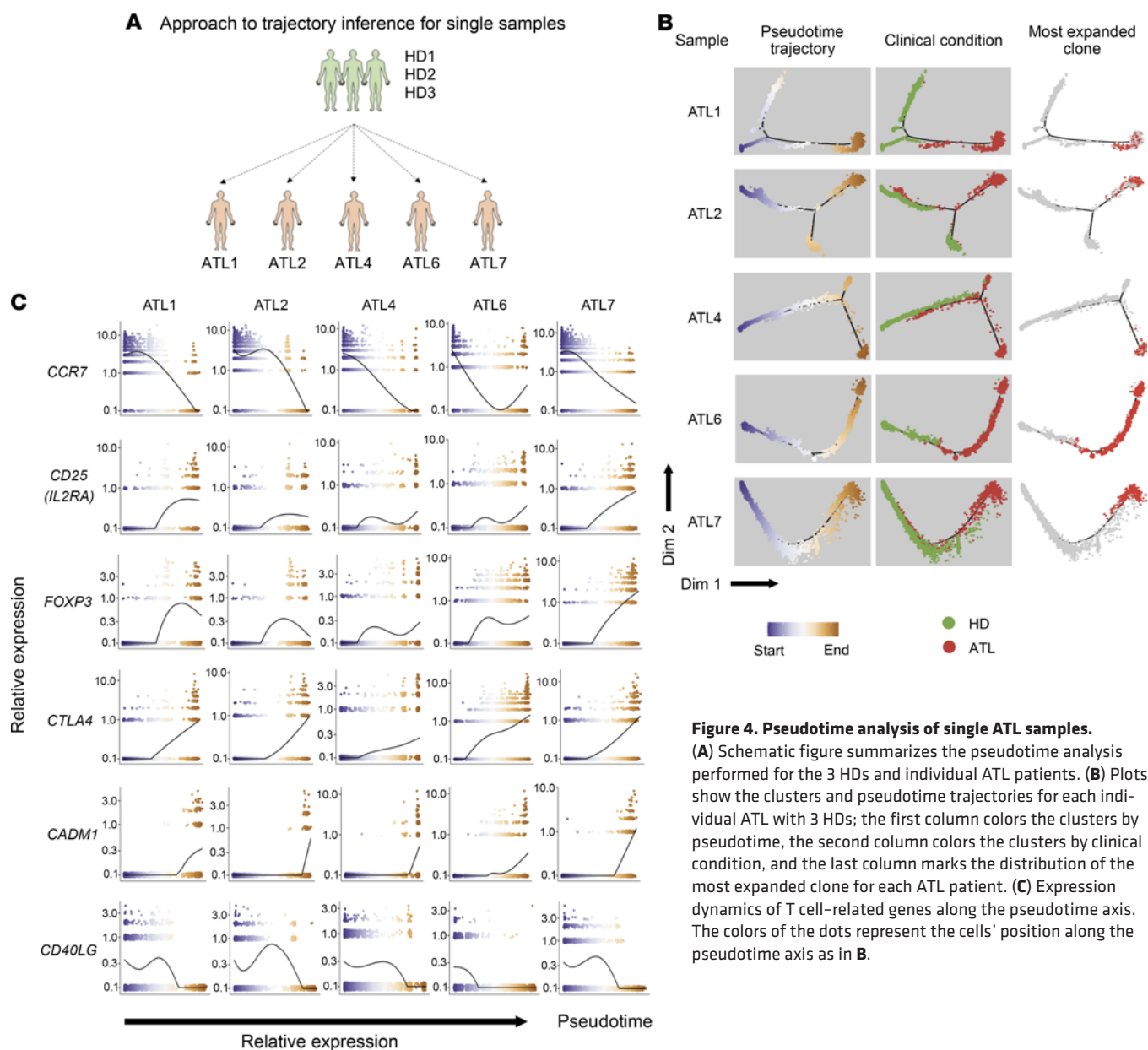
Figure 4. Pseudotime analysis of single ATL samples.
(**A**) Schematic figure summarizes the pseudotime analysis performed for the 3 HDs and individual ATL patients. (**B**) Plots show the clusters and pseudotime trajectories for each individual ATL with 3 HDs; the first column colors the clusters by pseudotime, the second column colors the clusters by clinical condition, and the last column marks the distribution of the most expanded clone for each ATL patient. (**C**) Expression dynamics of T cell–related genes along the pseudotime axis. The colors of the dots represent the cells' position along the pseudotime axis as in **B**.

class II expression was associated with HTLV-1 infection. We first examined the expression of HLA class II in HTLV-1–associated cell lines using flow cytometry. We observed that HLA class II expression was upregulated in all of the HTLV-1–associated cell lines we analyzed, but not in the HTLV-1⁻ ones with the exception of Kit225 cells (Figure 7A). Next, we analyzed freshly isolated CD4⁺ T cells from PBMCs of HTLV-1–infected individuals by comparing HLA class II expression between the uninfected cell fraction (CADM1⁻CD7⁺) and the HTLV-1–infected/ATL cell fraction (CADM1⁺CD7⁻/⁺). We found that the HTLV-1–infected/ATL cell fractions showed significantly higher levels of HLA class II expression compared with the uninfected cell fraction. However, when compared with professional APCs (CD14⁺ monocytes), the upregulation in HTLV-1–infected and ATL cells, was moderate (Figure 7, B and C). It is well established that the HTLV-1 viral genes *tax* and *hbz* have oncogenic functions. Therefore, to determine which of these 2 genes could contribute to the upregulation of HLA class II,

we transfected Kit225 cells with expression plasmids containing either HTLV-1 *tax*, HTLV-1 *hbz*, or HIV-1 *nef* (control). Cells transfected with Tax upregulated HLA class II expression, whereas those transfected with HBZ or Nef did not (Figure 7D). The level of *tax* expression in fresh PBMCs is low but can be induced in bursts. Furthermore, this expression can be induced by ex vivo cultivation (23–25). Therefore, to investigate the dynamics of Tax and HLA class II expression, we performed scRNA-Seq on PBMCs from 3 infected individuals before and after ex vivo cultivation. For all 3 samples, clustering analysis of CD4⁺ T cells revealed 3 clusters consisting of uninfected helper cells, infected cells, and cells with high expression of the HTLV sense strand. Pseudotime analysis inferred 2 trajectories for all 3 samples. The first trajectory transitioned from uninfected cells to infected cells, whereas the second trajectory progressed from uninfected to infected cells and to cells displaying a strong induction of HTLV sense strand transcription (Figure 7E). Consequently, we focused on the second trajectory to
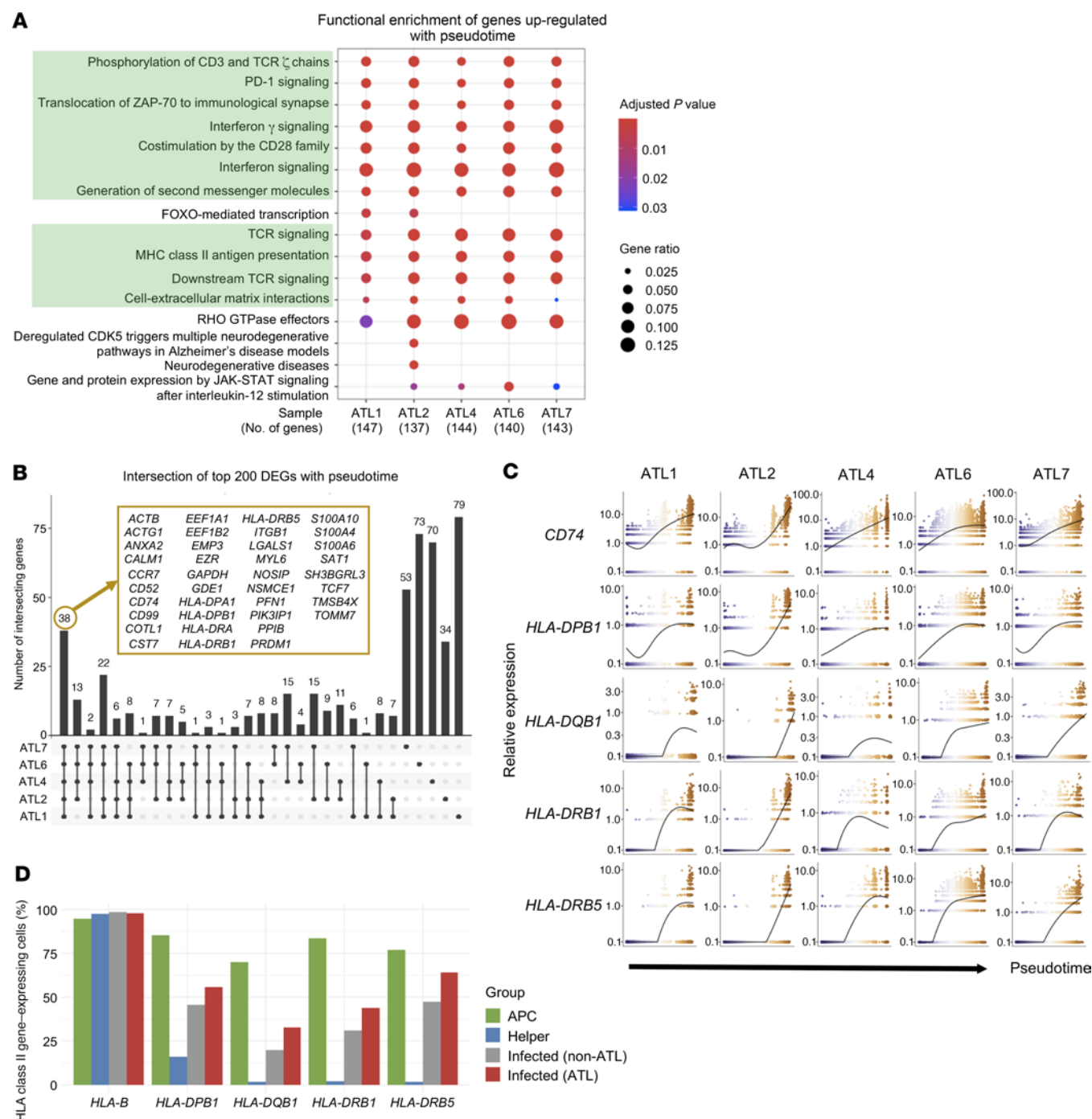
Figure 5. HLA class II genes are upregulated along the pseudotime axis. (**A**) Dot plot shows the shared and distinct reactome pathways of the top 200 genes varying as a function of pseudotime for each ATL sample. (**B**) Upset plot shows the intersection of the top 200 genes varying as a function of pseudotime for each ATL sample. Genes common to all ATL cases are shown in the yellow box. (**C**) Expression dynamics of HLA class II and related genes along the pseudotime axis. The colors of the dots represent the cells' position along the pseudotime axis as in Figure 3B. (**D**) Bar graph shows the percentage of cells in APCs (B cells and monocytes), uninfected helper T cells, infected non-ATL cells, and infected ATL cells that expressed HLA class II genes. *HLA-B* is shown here as control.

investigate the dynamics of HTLV sense strand transcription and HLA class II expression. We observed that along the trajectory, Tax expression was remarkably induced, while the expression of *HBZ* was variable. The expression of HLA class II also increased along the trajectory, and this upregulation occurred concurrently with the increase in Tax expression (Figure 7F). We then aimed to

ascertain the mechanism by which Tax mediated the upregulation of HLA class II. We showed earlier that the expression of genes related to HLA class II signaling was increased in ATL (Supplemental Figure 10A), and one of the upregulated genes, *CIITA*, is known to be a master regulator of HLA class II gene expression. We performed a single-cell assay for transposase-accessible

chromatin using sequencing (scATAC-Seq) on 2 representative samples (SML2 and ATL6) to examine the promoter accessibility of *CIITA*. The results showed that the *CIITA* promoter III was more open in infected cells than in uninfected cells (Supplemental Figure 10D). We then performed a luciferase assay to assess the effect of different concentrations of Tax on *CIITA* promoter III activity and observed that increasing concentrations of Tax caused a subsequent increase in the *trans*-activation of *CIITA* promoter III activity (Supplemental Figure 10E). These results suggest that the viral protein Tax plays a role in the upregulation of HLA class II genes during HTLV-1 infection.

HTLV-1–infected cells can present antigen to responder CD4+ T cells and induce anergy-related genes. Antigen presentation by professional APCs (B cells and monocytes) via HLA class II leads to the activation of CD4+ T cells and induces an immune response. We aimed to investigate whether HTLV-1–infected cells expressing HLA class II could present antigens to T cells. It has been reported that only certain alleles of HLA class II genes allow the presentation of endogenous peptides to T cells and that HLA-DP molecules such as DP2 and DP4 with β-chains encoding Gly84 (DP84Gly) can present antigen to T cells (26). Therefore, we conducted our studies with an in vitro model of TCR stimulation through HLA class II, using an HLA-DP4–restricted TCR (clone 9) that was specific to the cancer testis antigen Wilms tumor protein 1 (WT1) (27) to assess the antigen presentation capability of HTLV-1–infected CD4+ T cells. In fact, HLA-DP4–expressing and WT1 peptide–pulsed HTLV-1–infected cell lines induced cytokine production in responder T cells carrying the clone 9 TCR, while the HTLV-1⁻ T cell lines did not (Figure 8A). Furthermore, peptide-pulsed HTLV-1–infected CD4+ T cells successfully stimulated and induced IFN-γ production in clone 9–expressing CD4+ T cells, albeit at a low level, whereas WT1 peptide–pulsed normal CD4+ T cells and CLIP-pulsed (irrelevant peptide) APCs did not induce any IFN-γ production at all (Figure 8, B and C). These results confirmed that HTLV-1–infected CD4+ T cells could present antigens to CD4+ T cells. However, notably, HTLV-1–infected CD4+ T cells induced a significantly lower number of IFN-γ–producing cells in responder T cells than non–T cell PBMCs, which are enriched with B cells and monocytes and are commonly used as APCs. This indicates that, although HLA class II–expressing HTLV-1–infected CD4+ T cells can present antigen to T cells, they are inefficient APCs and are unable to fully activate CD4+ T cells (Figure 8C and Supplemental Figure 10, B and C). Next, we aimed to evaluate the effects of antigen-specific stimulation on the expression of key genes involved in anergy (Figure 8, B and D). We observed that TCR stimulation with HTLV-1–infected CD4+ T cells caused an increase in the expression of anergy-related molecules compared with the expression detected with non–T cells containing B cells and monocytes (Figure 8D and Supplemental Figure 10F). These results suggest that HTLV-1–infected cells can upregulate HLA class II to present antigens and cause T cell anergy to escape immunosurveillance in vivo.

Discussion

Leukemia, a type of cancer originating from blood or bone marrow cells, is characterized by a large increase in the numbers of abnormal white blood cells. Leukemia can be subdivided into several groups, but, generally, the 5 main categories are acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), chronic lymphocytic leukemia (CLL), chronic myeloid leukemia (CML), and other less common types of leukemia. Inhibition of cell differentiation is a key event during leukemogenesis in ALL and AML. ATL is a leukemia derived from peripheral CD4+ T cells. According to the data from the United States Surveillance, the incidence of peripheral T cell lymphomas is less than 1 case per 100,000 people (28); however, its incidence in the HTLV-1 endemic areas of Japan is approximately 80 per 100,000 infected individuals (29). HTLV-1 infection disrupts the differentiation, activation, proliferation, and apoptosis of the host CD4+ T cells, which leads to malignant transformation during its long latency period (10, 13, 30). Thus, analysis of the mechanism via which activated CD4+ T cells transform into ATL cells has been challenging because of the long latency period of the virus.

Under normal conditions, activated T cells induce the expression of FOXP3 and CTLA4, which can terminate T cell activation and restore homeostasis (31). This homeostatic regulation of T cells is important for preventing excessive inflammation in normal conditions. Our single-cell analyses showed that these proteins were actively induced and sustained in HTLV-1–infected and ATL cells, which may have contributed to immune evasion in the infection. This implies that persistent activation is key for understanding HTLV-1–mediated transformation of T cells. There are at least 3 possibilities for how HTLV-1–infected T cells and ATL cells sustain their highly activated T cell phenotype with negative regulatory mechanisms operating. First, the viral proteins Tax and HBZ may directly activate mechanisms downstream of TCR signals (e.g., NF-κB complexes) and negative feedback mechanisms for T cell activation such as transcription factor complexes associated with FOXP3. Tax enhances NF-κB and AP-1 signaling and contributes to the antiapoptotic and hyperproliferative phenotype of infected cells (12, 32). In contrast, HBZ competes with c-Fos for c-Jun binding, inhibiting AP-1 formation downstream of TCR signaling (33). HBZ also induces Treg differentiation by activating the TGF-β signaling pathway, thereby inducing FOXP3 expression (34). Second, mutations in TCR signaling molecules may sustain the activities of genes downstream of TCR signaling, as described above (15). Last, but not least, HTLV-1–infected T cells and ATL cells may be enriched with T cells with self-reactive TCRs, such as Tregs. It is known that approximately 20% of CD4+ T cells carry self-reactive TCRs and spontaneously receive TCR signals in vivo (18). These 3 mechanisms are not mutually exclusive and might act synergistically in HTLV-1–infected T cells, probably operating during leukemic transformation as well. Frequent T cell overactivation will lead to sustained activation of genes involved in cell-cycle regulation and apoptosis, which might render infected T cells more susceptible to genotoxic stress.

Intriguingly, our pathway analysis suggested that the IFN signaling pathway was actively induced in HTLV-1–infected T cells and ATL cells, supporting the hypothesis that antiviral innate immunity was induced in these T cells. However, anti–HTLV-1 immunity is actively suppressed during ATL. For example, Tax-specific cytotoxic T cells are reduced in ATL patients compared with that in HTLV-1 AC (35). Here, we propose that the unique dynamics of key surface molecules on ATL cells may contribute to the specific suppression of antiviral immunity in HTLV-1 infection and anti-leukemia immunity in ATL (immune evasion). Importantly, our trajectory analysis
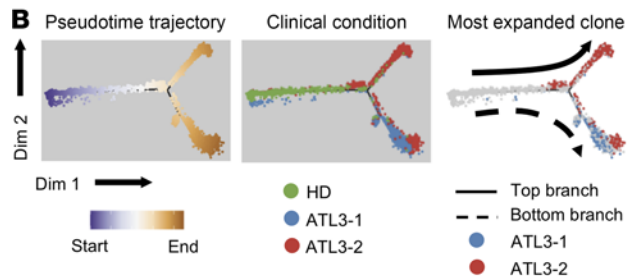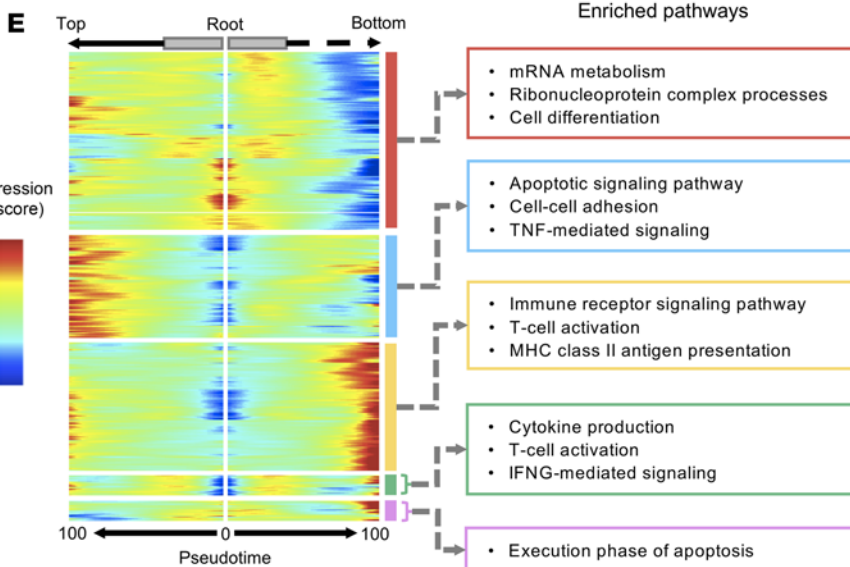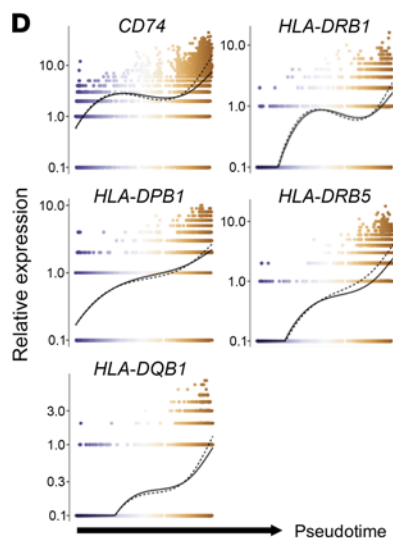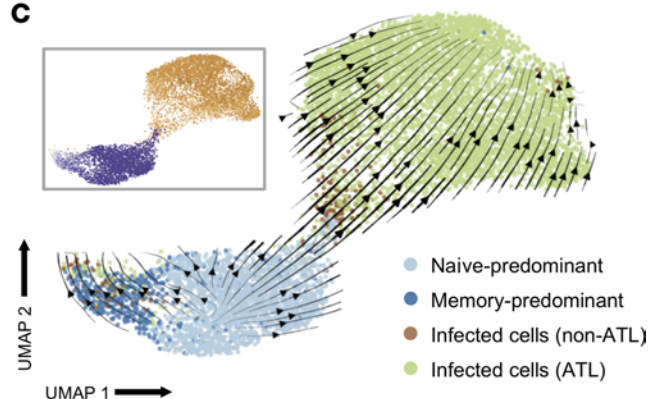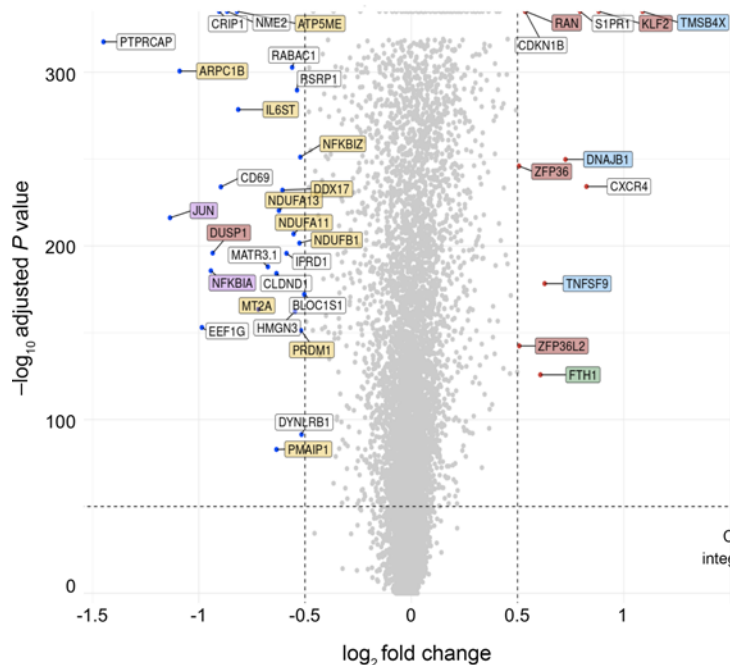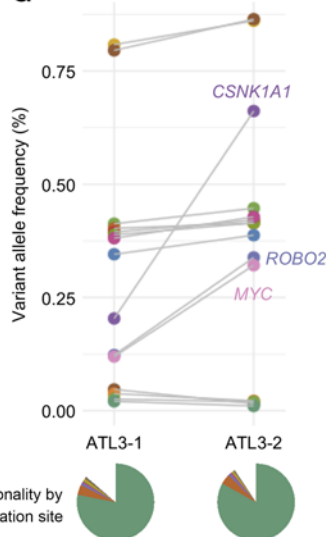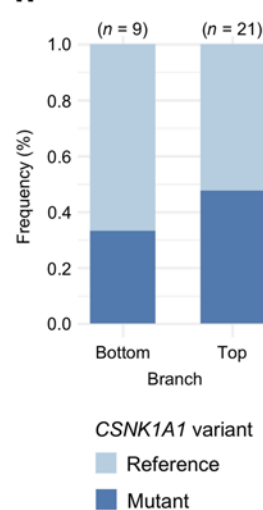
Figure 6. Pseudotime analysis of a paired ATL sample. (**A**) Schematic shows pseudotime analysis of the 3 HDs and the 1 paired ATL sample. (**B**) Plot showing the clusters and pseudotime trajectory; from left to right, clusters are colored by pseudotime, clinical condition, and most expanded clone for each ATL. (**C**) Plot shows the RNA velocity in Seurat-identified clusters. The inset box is colored by pseudotime as shown in **B**. (**D**) Expression dynamics of HLA class II and related genes along the pseudotime axis. The colors of the dots represent the cells' position along the pseudotime axis as in **B**. (**E**) Split heatmap shows the expression profile of genes that vary as a function of pseudotime and are branch dependent. The pseudotime trajectory begins from the middle of the heatmap (gray boxes) and moves to the left for the top branch and to the right for the bottom branch. The start of the arrow indicates the bifurcation point in the trajectory in **B**. Hierarchical clustering grouped the genes into 5 clusters, which are indicated by outlined colored boxes on the right, along with their enriched pathways. (**F**) Volcano plot shows the DEGs between the top and bottom branches. The genes are colored according to the clusters shown in the split heatmap in **E**. (**G**) The dot and line graph shows the change in frequencies of variant alleles across 2 time points, as detected using targeted exome sequencing, for the CADM1⁺CD7⁻/⁺ cell population from the paired ATL case. Genes showing huge changes in frequency are labeled. The pie chart at the bottom shows the ATL cells' clonality identified by the HTLV-1 integration site. (**H**) Bar graph shows the frequency of the *CSNK1A1* variant by trajectory branch detected in the scRNA-Seq data.

showed that T cells consistently upregulated *CTLA4* throughout the ATL trajectory. CTLA4 efficiently binds to CD80/CD86 and removes these CD28 ligands from APCs, thus inhibiting the activities of T cells and APCs (36). Therefore, high CTLA4 expression in ATL cells may inhibit T cell responses in their local microenvironments. In addition, the increased expression of HLA class II in ATL cells may also contribute to the inhibition of T cell responses. T cells express HLA class II molecules following activation (37, 38), although the functional significance is unknown. Activated Tregs (effector Tregs) that express high levels of HLA class II molecules show enhanced suppressive activity (39, 40). Intriguingly, our trajectory analysis showed that HLA class II expression was strongly induced during HTLV-1 infection and further enhanced in ATL. Importantly, to our knowledge, our study provides the first evidence for the ability of HTLV-1–infected cells to present antigen to T cells. However, HTLV-1–infected cells may not be efficient APCs, because the expression of HLA class II and related molecules in HTLV-1–infected and ATL cells is lower than that in professional APCs (Figure 5D and Supplemental Figure 10A). Furthermore, HTLV-1–infected cells did not express the critical ligands for costimulation, such as CD80/CD86, which convey the key costimulatory CD28 signaling in T cells (Supplemental Figure 10, B and C). These results collectively suggest that HTLV-1–infected cells facilitated the escape of infected cells from host immune surveillance (Figure 9), and like tolerogenic DCs, induced anergy and immunosuppressive function in T cells (41, 42). Our new model of HTLV-1–infected cells and HLA-DP4–restricted WT1-specific TCRs will be useful for future studies addressing the underlying molecular mechanism (Figure 8 and ref. 26).

In summary, on the basis of the single-cell data obtained in this study, we propose 2 models to explain how HTLV-1–infected cells deregulate and evade the host immune system (Figure 9). We found that HTLV-1 infection exploited the physiological mechanisms for T cell activation and homeostasis. Our single-cell analysis provides

a model for the transformation of HTLV-1–infected CD4⁺ T cells into leukemic cells during the initial phase of ATL leukemogenesis and demonstrates how an infected clone undergoes malignant transformation in vivo.

Methods

Further details on the methods are provided in the Supplemental Methods.

Patient blood samples. All participants in this study were seen at Imamura General Hospital in Kagoshima City except for the single patient with lymphoma-type ATL, who was seen at Kumamoto Shinto General Hospital in Kumamoto City. Fresh whole blood was obtained from HDs ($n = 3$) and HTLV-1–infected patients with the following clinical diagnoses: AC ($n = 4$), SML ($n = 3$), chronic ATL ($n = 6$), and lymphoma-type ATL ($n = 1$). PBMCs were isolated from whole blood within 24 hours of sample collection using Ficoll-Paque (GE Healthcare) according to the manufacturer's instructions. Briefly, each blood sample was overlaid on top of Ficoll-Paque at a ratio of 2:1 and centrifuged at 500g for 25 minutes at room temperature. Enriched mononuclear cells were washed with PBS and twice centrifuged at 550g for 10 minutes. Cell numbers and viability were checked using a hemocytometer and trypan blue staining.

Cell suspension preparation. Freshly isolated PBMCs were resuspended in PBS to achieve a cell concentration of 5×10^6 cells/mL. One microliter of cell suspension was used for single-cell library preparation, and the remaining PBMCs were resuspended in BAMBANKER cryopreservation medium (Nippon Genetics) and stored at –80°C. Prior to single-cell library preparation, dead cells were removed from the 1 mL cell suspension using a dead cell removal kit (Miltenyi Biotec) according to the manufacturer's instructions. After the final wash, cell numbers were determined using a hemocytometer and centrifuged at 700g for 5 minutes. The resulting cell pellet was then resuspended in PBS plus 0.04% BSA to a concentration of 1000 cells/μL.

PVL measurement of clinical samples by droplet digital PCR. Genomic DNA was extracted from 5×10^6 cells using the DNeasy Blood and Tissue Kit (QIAGEN) following the manufacturer's instructions. Genomic DNA (50 or 100 ng) was used for each droplet digital PCR (ddPCR) reaction. ddPCR was performed using primers and a probe targeting a conserved region in the HTLV-1 pX region and the albumin (*ALB*) gene as previously reported (43). Droplet quantification was performed on a QX200 droplet reader (Bio-Rad), and data were analyzed using QuantaSoft software (version 1.7.4, Bio-Rad). For an objective cutoff with maximum sensitivity, a no-template-control (NTC) sample was used to determine the threshold values in which the highest level of droplet fluorescence in the NTC was designated as the threshold line for negative signal. The PVL was thus calculated as follows: PVL (%) = 100 (HTLV copy number × 2)/ALB copy number.

Single-cell library preparation. Live cell–enriched cell suspensions were loaded onto a Chromium Single Cell Chip (10x Genomics) for coencapsulation with barcoded gel beads from the Single Cell 5′ Gel Bead Kit (10x Genomics) at a target capture rate of 10,000 individual cells per sample. Barcoded cDNAs were pooled, PCR amplified, and then used to generate 2 different libraries. A single-cell transcriptome library was prepared with 50 ng cDNA of amplified product using the Chromium Single Cell 5′ Library Kit (10x Genomics). Enrichment of V(D)J segments and construction of the single-cell TCR library was performed using the Chromium Single Cell V(D)J Enrichment Kit, Human T Cell;
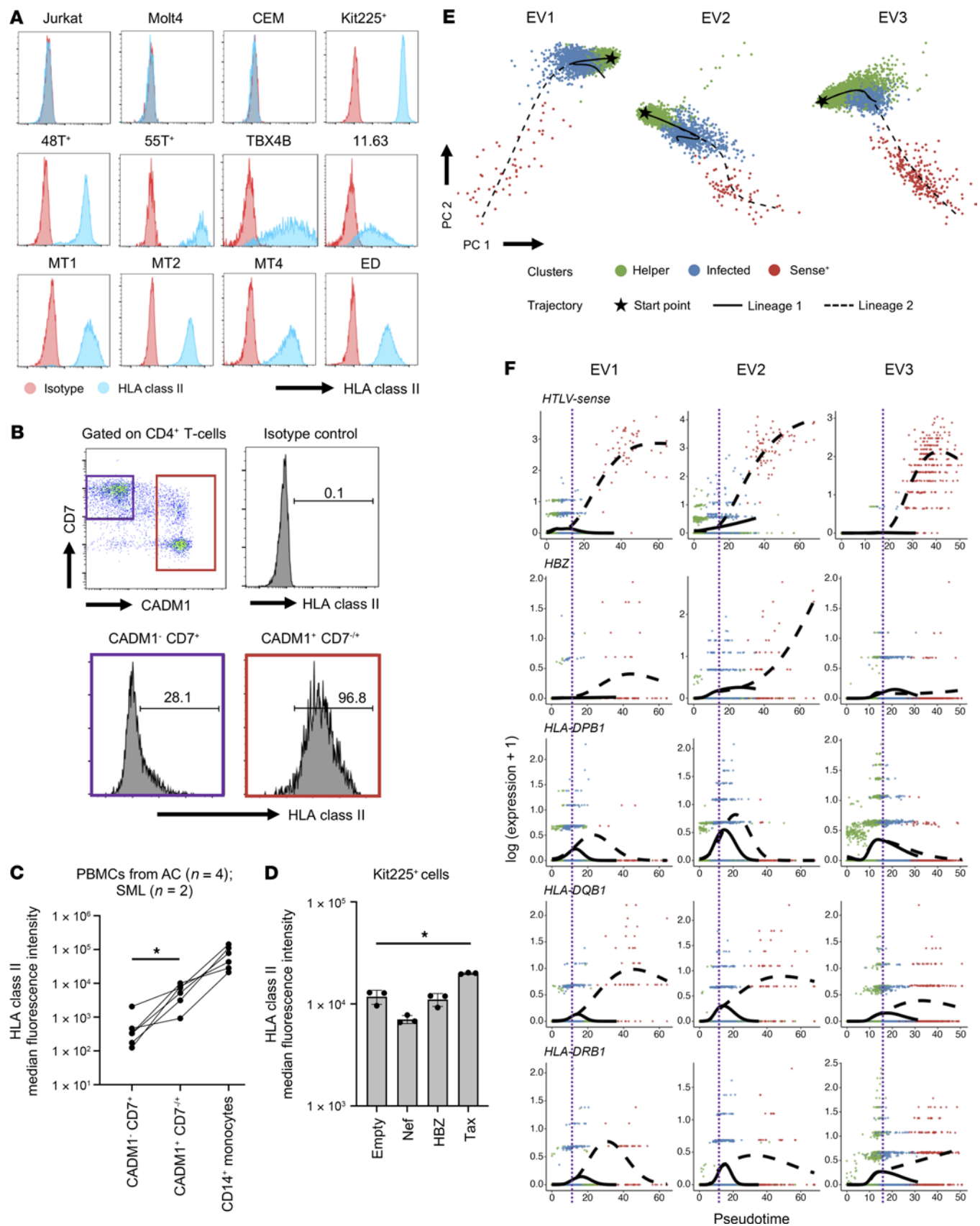
Figure 7. HLA class II expression is associated with HTLV-1 infection and the viral protein Tax. (**A**) Plots showing the detection of HLA class II molecules in HTLV-1–infected cell lines. (**B**) Plots show a representative result of the distribution of freshly isolated CD4$^+$ T cells from HTLV-1–infected individuals based on CADM1 and CD7 expression. ATL cells were identified as CADM1$^+$CD7$^{-/+}$ and highly expressed HLA class II molecules. (**C**) Line graph shows the difference in HLA class II expression between non-ATL cells (CADM1$^-$CD7$^+$), ATL cells (CADM1$^+$CD7$^{-/+}$), and monocytes (CD14$^+$) in 6 individuals infected with HTLV-1. (**D**) Bar graph shows HLA class II expression in Kit225$^+$ cells transfected with either an empty vector, HIV-1 Nef, HTLV-1 HBZ, or Tax. $n = 3$. (**E**) 2D principal component analysis (PCA) plots show the distribution of helper, infected, and HTLV-sense strand–expressing cells in 3 different ex vivo–cultivated T cells from HTLV-1–infected individuals. Also shown are the pseudotime trajectories analyzed by Slingshot. (**F**) Expression dynamics of viral and HLA class II genes along the pseudotime trajectories. The colors of the dots represent the clusters (as in **E**). The purple line indicates the bifurcation point of the pseudotime trajectories. Data indicate the mean ± SD. *$P < 0.05$, by 1-way ANOVA with post hoc Tukey's test (**C** and **D**).

and the Chromium Single Cell 5′ Library Construction Kit, respectively (10x Genomics). The libraries were sequenced with paired-end, single indexing at 150 bp plus 150 bp reads (TCR library) or 26 bp plus 91 bp reads (transcriptome library) on the Illumina HiSeq or NextSeq platform.

CCA. Explanatory variables for CCA were prepared as follows. The T cell activation explanatory variable was defined by the differentially expressed genes (DEGs), obtained using the Bioconductor package DESeq2 (version 1.30.1), between TCR-stimulated-only T cells (Th0) and resting T cells from a published scRNA-Seq data set of CD4$^+$ T cells (https://www.opentargets.org/projects/effectorness; ref. 44). The Treg explanatory variable was defined by the DEGs, obtained using GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/), between resting Tregs and resting conventional T cells (Tconv) from the NCBI's Gene Expression Omnibus (GEO) accession data GSE15390 (45). For both explanatory variables, genes with a FDR of less than 0.05 and a log$_2$ fold change (>0.5 or ≤0.5) were selected. For the 1D CCA, the expression data for CD4$^+$ T cells only were regressed onto each of the explanatory variables above, and correspondence analysis was performed using the Bioconductor package vegan (version 2.5-7). For the 2D CCA, the expression data for CD4$^+$ T cells only were regressed onto both explanatory variables followed by correspondence analysis. The regression model of the CCA scores was obtained by applying a generalized additive model to the data set using the Bioconductor package ggplot2 (version 3.3.5).

Pseudotime analysis. Pseudotime trajectories were identified using the monocle package (version 2.18.0) (46). We prepared the Seurat object for pseudotime analysis by first excluding all cytotoxic T cells (C1–C3) from the Seurat object. Next, we removed cells that expressed CD8A/CD8B and also cells with unidentified clonotype in TCR-Seq. For the analysis involving the entire T cell population, we converted the resulting Seurat object into a CellDataSet (CDS) object and passed it to the monocle R package for pseudotime analysis. For the individual/paired ATL analysis, we performed another sub setting, in which we kept only cells from the 3 healthy individuals and the ATL patient of interest before converting it to a CDS object. Clusters that were enriched with CCR7$^+$ cells were assumed to be the origin of the trajectory. Genes that varied as a function of pseudotime were identified using monocle's "differentialGeneTest" function.

RNA velocity analysis. RNA velocity analysis was performed on the data set for 3 HDs and a paired ATL sample. Using the aligned BAM files

from Cell Ranger (10x Genomics), the number of spliced and unspliced reads was recounted using the Python package velocyto (version 0.17.17) (47) to generate the loom files. To generate the uniform manifold approximation and projection (UMAP) and clusters for these data sets, we first subset these 5 data sets from the Seurat object of T cells. Next, the top 3000 variable features were detected followed by removal of mitochondrial, ribosomal, and TCR genes. Data integration was performed with Harmony, and UMAP was used for visualization of unsupervised clustering. Dynamical model velocities were computed from the loom files and visualized on UMAP using the Python package scVelo (version 0.2.3) (48).

TCR-Seq analysis. TCR sequences were assembled with the Cell Ranger vdj pipeline (version 3.1.0), leading to an output file that contained the identification of the CDR3 sequence and the rearranged TCR gene. Cells with an identified clonotype (i.e., not NA or "none") were kept, and the others were discarded. Clonality was then designated as expanded for cells with 2 or more similar TCRs, whereas the remaining ones were marked as unexpanded.

Quantification of the HTLV-1 PVL in sorted cells. Measurement of the HTLV-1 PVL of sorted cells (CADM1$^-$CD7$^+$ and CADM1$^+$CD7$^{-/+}$ cells) was conducted as described previously (21). Briefly, quantitative multiplex real-time PCR was performed with 2 sets of primers specific for the HTLV-1 provirus and the human gene encoding the RNase P enzyme. PVL was expressed as HTLV-1 copy numbers per 100 PBMCs, based on the assumption that infected cells harbor 1 copy of integrated HTLV-1 provirus per cell.

Targeted exome sequencing and analysis. Target capture was performed using an HTLV-1/ATL panel and the SureSelect Target Enrichment System (Agilent Technologies) as reported previously (49). The resulting libraries were sequenced with 100 bp paired-end reads on the Illumina HiSeq platform. Sequenced data were aligned to the human reference genome hg38 using Burrow-Wheeler Aligner (BWA, version 0.7.15) software. PCR duplicates were removed using Picard (version 2.92) and SAMtools (version 1.2) software. Uninfected T cells (CD4$^+$CADM1$^-$CD7$^+$) were used as matched normal controls to call somatic mutations. Somatic mutation candidates were called using MuTect2 from GATK (version 4.0.12) software and annotated with ANNOVAR (version 20191024). Candidate mutations with (a) 5 or more variant reads in tumor samples; (b) a variant allele frequency (VAF) in tumor samples of 0.01 or higher; (c) a read depth of 200 or higher; and (d) a tumor/normal variant ratio of 2 or higher were adopted and further filtered by excluding synonymous single nucleotide variants (SNVs).

Preparation of PBMCs and cell lines for flow cytometric analysis. PBMCs were obtained via density gradient centrifugation using Ficoll-Paque PLUS (GE Healthcare Life Sciences). The leukemic T cell lines Jurkat, Molt4, and Kit225$^+$ were obtained from American Type Culture Collection (ATCC; TIB-152, CRL-1582, and CRL-1990, respectively). CEM, a leukemic T cell line, was obtained from Hiroaki Takeuchi (Tokyo Medical and Dental University, Tokyo, Japan). The ATL cell lines ATL-48T$^+$, ATL-55T$^+$, MT1, and ED were obtained from Michiyuki Maeda (50). The HTLV-1–infected cell lines MT-2, MT-4, TBX4B, and 11.63 were obtained from Charles R.M. Bangham (51). IL-2–dependent cell lines were cultured in the presence of recombinant IL-2 (100 U/mL). All T cell lines were cultured in RPMI 1640 supplemented with 10% FBS and penicillin/streptomycin.

Analysis of ex vivo–cultivated primary cells. Fresh PBMCs were cultivated in RPMI 1640 supplemented with 10% FBS and penicillin/streptomycin for 18 hours to induce viral sense strand transcription in
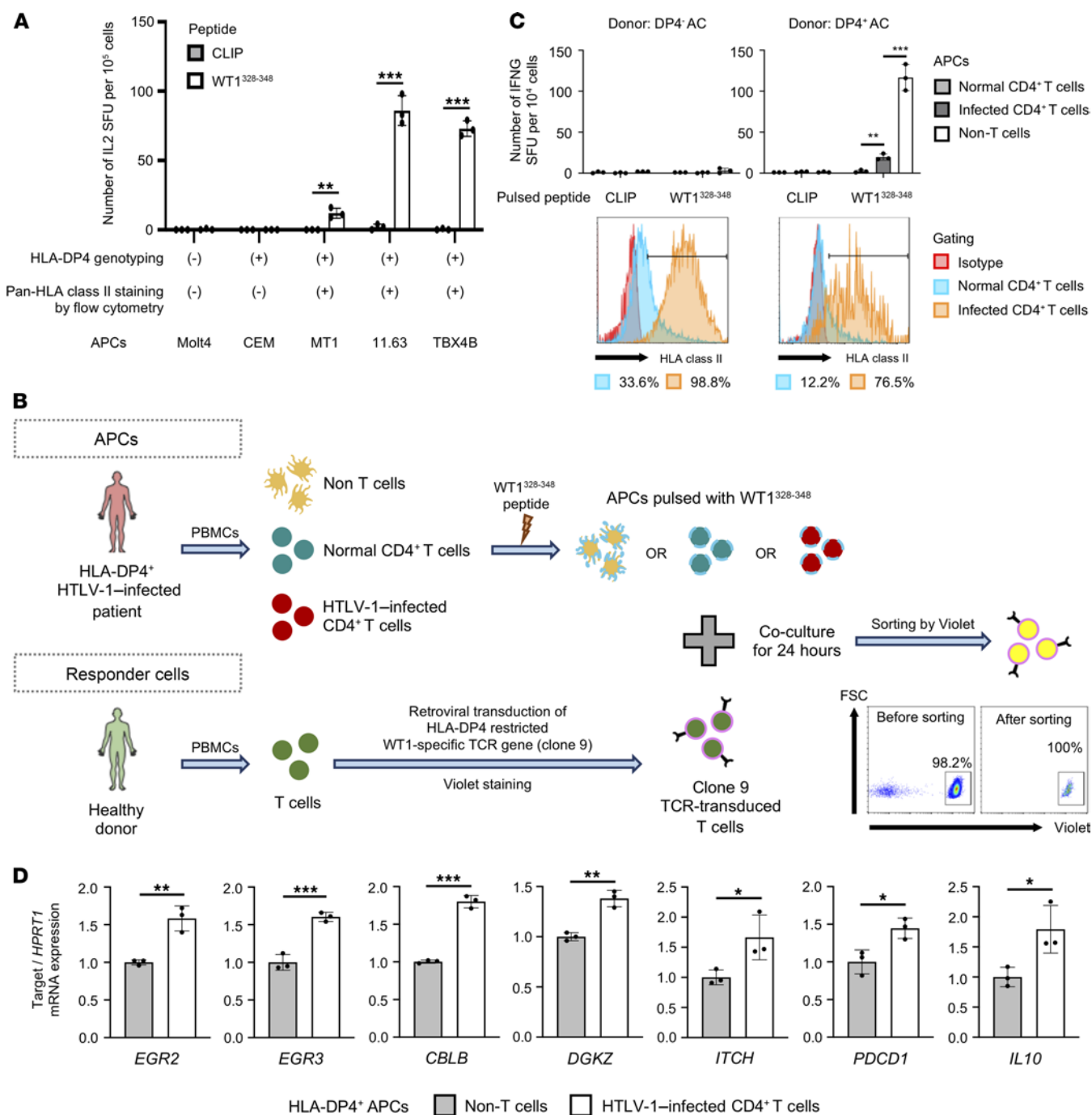
Figure 8. HTLV-1–infected cells can present antigen to responder CD4+ T cells and induce anergy-related genes. (**A**) Bar graph shows the antigen-presenting capabilities of the different HTLV-1–infected T cell lines. Jurkat cells expressing the HLA-DP4–restricted TCR (clone 9) specific for cancer testis antigen WT1 were stimulated with peptide-pulsed cell lines. CLIP peptide was used as a negative control. (**B**) Experimental scheme to evaluate T cell status after antigen-specific stimulation by HTLV-1–infected CD4+ T cells. Similar results were obtained from at least 2 independent donor samples. (**C**) Bar graphs show the antigen-presenting capabilities of different APCs from an HLA-DP4− donor and HLA-DP4+ donor. Normal CD4+ T cells, infected CD4+ T cells, and non–T cells from ACs were prepared as CADM1−CD4+, CADM1+CD4+, and CD4/CD8-depleted PBMCs, respectively. Clone 9–transduced primary T cells were stimulated with APCs pulsed with peptides. Histograms show the expression of HLA class II in normal (CADM1−CD4+) and infected CD4+ T cells (CADM1+CD4+). SFU, spot-forming units. (**D**) Bar graphs show the expression of T cell anergy–related genes in responder T cells after stimulation with non–T cells (conventional APCs) or HTLV-1–infected CD4+ T cells expressing HLA class II molecules. Non–T cells and infected CD4+ T cells were prepared from HLA-DP4+ ACs as CD3-depleted and CADM1+CD4+ PBMCs, respectively. For each APC, the relative expression of each gene was calculated by normalization against the expression level of *HPRT1* in the APCs stimulated with a negative control peptide (CLIP). $n = 3$ (**A**, **C**, and **D**). Data indicate the mean ± SD. *$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$, by unpaired, 2-tailed Student's *t* test.
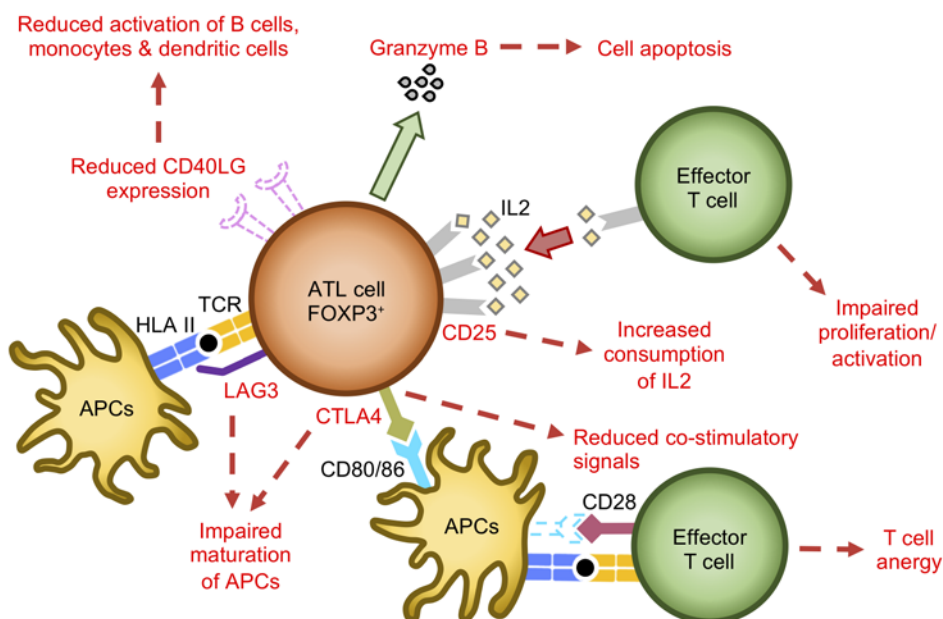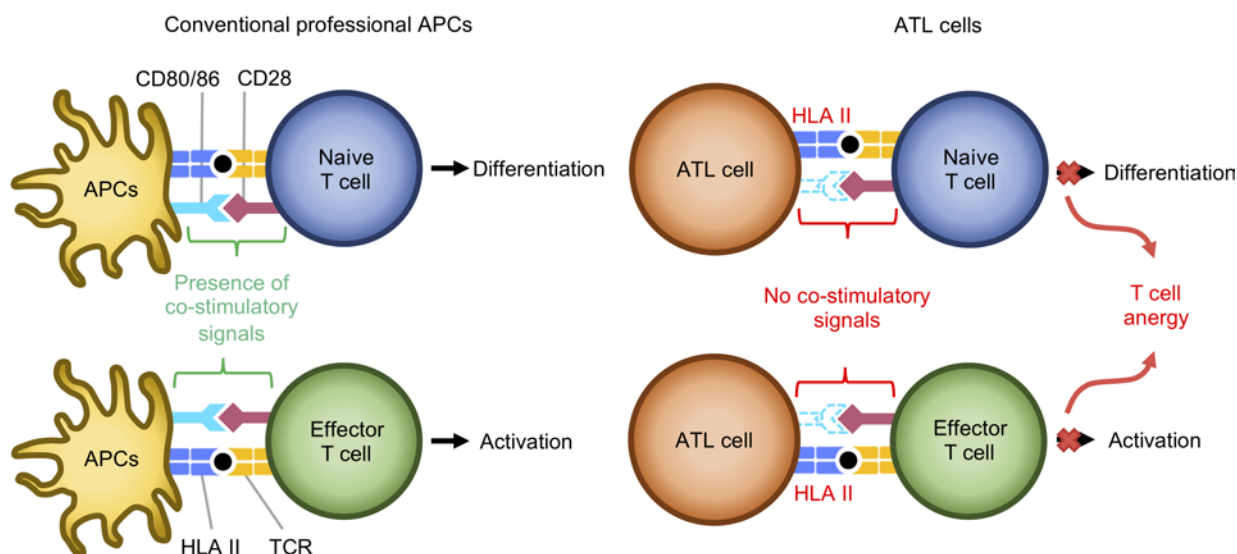
Figure 9. Schematic showing our proposed model of the immunosuppressive properties of ATL cells. (**A** and **B**) ATL cells exert their immunosuppressive function by 2 mechanisms. (**A**) ATL cells acquire the Treg phenotype. ATL cells upregulate coinhibitory molecules (CTLA4, LAG3) to impair the maturation of APCs and reduce costimulatory signals required for complete T cell activation. They also upregulate CD25, leading to an increased consumption of IL-2, which reduces the amount of IL-2 available for effector T cell proliferation and activation. (**B**) ATL cells act as tolerogenic APCs. They upregulate HLA class II for antigen presentation, but since these ATL cells lack costimulatory signals, the target T cells are not activated, leading to T cell unresponsiveness and anergy.

naturally infected cells. Harvested cells were subjected to single-cell library preparation, and the resulting scRNA-Seq data were analyzed as described above. Pseudotime trajectories were identified using the Bioconductor package slingshot (version 1.8.0) (52) and projected onto 2D PC space. Clusters that were enriched with CCR7+ cells were

assumed to be the origin. A negative binomial generalized additive model (GAM) for each gene was fitted using the Bioconductor package tradeSeq (version 1.4.0) (53).

Flow cytometric analysis. The following reagents were used for flow cytometric analysis: BV510-conjugated anti–HLA-II (Tu39, BioLegend);

BV421-conjugated anti-CD7 (M-T701, BD Biosciences); anti-CADM1 (3E1, Medical & Biological Laboratories [MBL]); FITC-conjugated anti-CD80 (2D10, BioLegend); PE-conjugated anti-CD86 (IT2.2, BioLegend); polyclonal Alexa Fluor 647–conjugated goat anti–chicken IgY (Abcam); PE-conjugated streptavidin (BioLegend); PerCP/cyanine5.5-conjugated anti-CD14 (HCD14, BioLegend); and PerCP/cyanine5.5- or biotin-conjugated anti-CD4 (OKT4, BioLegend). Dead cells were distinguished with the LIVE/DEAD Fixable Near-IR Cell Stain Kit (Thermo Fisher Scientific). Primary ATL cells and cell lines were pretreated with goat serum for 20 minutes at room temperature. After washing, cell-surface molecules were stained with specific antibodies for 20 minutes at 4°C. Stained cells were analyzed with FACSVerse (BD Biosciences). Data analysis was performed using FlowJo software (version 10.7.1, Tree Star).

Transfection of tax and hbz. HTLV-1 viral gene expression vectors such as *tax* and *hbz* were generated based on pIRES-EGFP (6029-1, Clontech). As a negative control, HIV-1 *nef* gene expression vectors were also generated based on pIRES-EGFP. Kit225+ cells were transfected with each viral expression vector by electroporation using NEPA21 (Nepa Gene).

Assessment of antigen presentation by upregulated HLA class II molecules in HTLV-1–infected cells. For the ELISPOT assay, all HTLV-1–infected cell lines and donors were typed for HLA-DP at the HLA Foundation Laboratory. The TCR-deficient Jurkat cell line was established previously (54). Synthetic peptides (CLIP: LPKPPKPVSKMRMATPLL-MQALPM and WT1$_{328-348}$: PGCNKRYFKLSHLQMHSRKHT) were purchased from GenScript and dissolved at 50 mg/mL in DMSO. Primary T cells were purified using the Pan T cell Isolation Kit (Miltenyi Biotec). HLA-DP4–restricted, WT1-specific TCRs (clone 9; ref. 27) were transduced into TCR-deficient Jurkat cells or primary T cells using the Plate-GP and PG13 cell–based retrovirus system and Retronectin (Takara Bio) according to the manufacturer's instructions. HLA-DP4/WT1 TCR-transduced Jurkat cells were stimulated with HTLV-1–infected cell lines pulsed with WT1$_{328-348}$ (10 μg/mL) at an effector/target ratio of 5:1 for 24 hours followed by measurement of specific cytokine production using the human IL-2 ELISPOT Kit (Mabtech) and the ELIPHOTO Counter (Minerva Tech).

For stimulation of TCR-transduced primary T cells, CD4+ T cells were first isolated from HLA-DP4+ ACs using the Pan T cell Isolation Kit (Miltenyi Biotec) and biotin-labeled anti-CD8 mAb (RPA-T8, BioLegend). Next, HTLV-1–infected cells were sorted as CD4+CADM1+ T cells using FITC-labeled anti-CADM1 mAb (3E1, MBL) and anti-FITC MicroBeads (Miltenyi Biotec). The CADM1– fractions were used as normal CD4+ T cells. Non–T cells containing monocytes and B cells were prepared by T cell depletion using anti-biotin MicroBeads (Miltenyi Biotec), biotin-labeled anti-CD4 mAb (OKT4, BioLegend), and biotin-labeled anti-CD8 mAb (RPA-T8, BioLegend). Primary T cells (1 × 10^4 cells) expressing clone 9 were stimulated for 24 hours with HLA-DP4+ non–T cells, normal CD4+ T cells, or HTLV-1–infected CD4+ T cells (2 × 10^4 cells) pulsed with WT1$_{328-348}$ (10 μg/mL) followed by measurement of specific cytokine production using the human IFN-γ ELISPOT Kit and the ELIPHOTO Counter.

Assessment of T cell anergy status upon antigen-specific stimulation by HTLV-1–infected cells. Responder T cells were generated by transducing HLA-DP4/WT1 TCRs (clone 9) into primary T cells as shown in Figure 6C and labeled with a tracer dye (CellTrace Violet, Life Technologies, Thermo Fisher Scientific). HTLV-1–infected CD4+ T cells from HLA-DP4+ ACs were as purified as shown in Figure 6C. Non–T cells were prepared by T cell depletion using anti-FITC MicroBeads (Miltenyi Biotec) and FITC-labeled anti-CD3 mAb (SK7, BioLegend).

Both stimulators were pulsed with WT1$_{328-348}$ (10 μg/mL) for 2 hours and then cocultured with responder T cells for 24 hours. Cells were then stained using the LIVE/DEAD Fixable Near-IR Cell Stain Kit (Thermo Fisher Scientific), and Violet+ cells (responder T cells) were sorted using the SH800S Cell Sorter (Sony). Quantification of genes associated with T cell anergy was performed by real-time PCR. cDNA was synthesized using ReverTra Ace qPCR RT Master Mix (Takara Bio) followed by quantification using THUNDERBIRD SYBR qPCR Mix (TOYOBO) and the StepOnePlus Real-Time PCR System (Applied Biosystems). Primer sequences for the genes used are listed in Supplemental Table 2. The relative expression level for each sample was calculated using the ΔΔCt method, with the expression level of *HPRT1* in CLIP peptide–stimulated T cells as the reference sample.

Data availability. The data for this study were deposited in the European Nucleotide Archive (ENA) at the European Bioinformatics Institute (EMBL-EBI) under accession number PRJEB47382 (https://www.ebi.ac.uk/ena/browser/view/PRJEB47382). The source code to reproduce our analyses can be accessed at https://github.com/satou-lab/SingleCell_HTLV_JCI (commit ID 4285bbb62a8024e7ddd7b-708c870e7c1c8f344fe).

Statistics. Statistical analyses for CCA scores were performed with R (version 4.0.3). Statistical analyses for flow cytometric, ELISPOT, and real-time PCR data were performed using GraphPad Prism 7 software (GraphPad Software). A 2-tailed, unpaired Student's *t* test or 1-way ANOVA followed by post hoc Tukey's test was used to determine significance, which was defined as a *P* value of less than 0.05.

Study approval. Study participants provided written informed consent prior to blood collection. The study was approved by the ethics review committee of Kumamoto University (Genome no. 263).

Author contributions

BJYT, MO, and Y Satou conceptualized and designed the study. BJYT, KS, OR, MM, K Uchiyama, PM, MY, Y Suzuki, and HK performed the experiments. BJYT, KS, VH, MY, K Uchimaru, YU, MO, and Y Satou analyzed the data. TU provided materials. HS, MT, HN, ES, and AU took care of patients, procured blood samples, and provided clinical information. BJYT, MO, and Y Satou wrote and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgments

1. Poiesz BJ, et al. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T cell lymphoma. *Proc Natl Acad Sci U S A*. 1980;77(12):7415–7419.

2. Bangham CRM. Human T cell leukemia virus type 1: persistence and pathogenesis. *Annu Rev Immunol*. 2018;36:43–71.

3. Shimoyama M. Diagnostic criteria and classification of clinical subtypes of adult T-cell leukaemia-lymphoma. A report from the Lymphoma Study Group (1984-87). *Br J Haematol*. 1991;79(3):428–437.

4. Uchiyama T, et al. A monoclonal antibody (anti-Tac) reactive with activated and functionally mature human T cells. I. Production of anti-Tac monoclonal antibody and distribution of Tac (+) cells. *J Immunol*. 1981;126(4):1393–1397.

5. Karube K, et al. Expression of FoxP3, a key molecule in CD4CD25 regulatory T cells, in adult T-cell leukaemia/lymphoma cells. *Br J Haematol*. 2004;126(1):81–84.

6. Chen S, et al. Regulatory T cell-like activity of Foxp3+ adult T cell leukemia cells. *Int Immunol*. 2006;18(2):269–277.

7. Fontenot JD, et al. Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. *Nat Immunol*. 2003;4(4):330–336.

8. Hori S, et al. Control of regulatory T cell development by the transcription factor Foxp3. *Science*. 2003;299(5609):1057–1061.

9. Khattri R, et al. An essential role for Scurfin in CD4+CD25+ T regulatory cells. *Nat Immunol*. 2003;4(4):337–342.

10. Satou Y, et al. HTLV-1 modulates the frequency and phenotype of FoxP3+CD4+ T cells in virus-infected individuals. *Retrovirology*. 2012;9:46.

11. Matsuoka M, Jeang KT. Human T-cell leukaemia virus type 1 (HTLV-1) infectivity and cellular transformation. *Nat Rev Cancer*. 2007;7(4):270–280.

12. Giam CZ, Semmes OJ. HTLV-1 infection and adult T-cell leukemia/lymphoma-a tale of two proteins: tax and HBZ. *Viruses*. 2016;8(6):161.

13. Satou Y, et al. HTLV-1 bZIP factor induces T-cell lymphoma and systemic inflammation in vivo. *PLoS Pathog*. 2011;7(2):e1001274.

14. Satou Y, et al. HTLV-I basic leucine zipper factor gene mRNA supports proliferation of adult T cell leukemia cells. *Proc Natl Acad Sci U S A*. 2006;103(3):720–725.

15. Kataoka K, et al. Integrated molecular analysis of adult T cell leukemia/lymphoma. *Nat Genet*. 2015;47(11):1304–1315.

16. Kobayashi S, et al. CADM1 expression and stepwise downregulation of CD7 are closely associated with clonal expansion of HTLV-I-infected cells in adult T-cell leukemia/lymphoma. *Clin Cancer Res*. 2014;20(11):2851–2861.

17. Kagdi HH, et al. Risk stratification of adult T-cell leukemia/lymphoma using immunophenotyping. *Cancer Med*. 2017;6(1):298–309.

18. Bradley A, et al. Elucidating T cell activation-dependent mechanisms for bifurcation of regulatory and effector T cell differentiation by multidimensional and single-cell analysis. *Front Immunol*. 2018;9:1444.

19. Kern I, et al. The two novel MHC class II transactivators RFX5 and CIITA both control expression of HLA-DM genes. *Int Immunol*. 1995;7(8):1295–1299.

20. Schroder B. The multifaceted roles of the invariant chain CD74--More than just a chaperone. *Biochim Biophys Acta*. 2016;1863(6 pt a):1269–1281.

21. Yamagishi M, et al. Polycomb-mediated loss of miR-31 activates NIK-dependent NF-κB pathway in adult T cell leukemia and other cancers. *Cancer Cell*. 2012;21(1):121–135.

22. Wong RWJ, et al. Enhancer profiling identifies critical cancer genes and characterizes cell identity in adult T-cell leukemia. *Blood*. 2017;130(21):2326–2338.

23. Hanon E, et al. Abundant tax protein expression in CD4+ T cells infected with human T-cell lymphotropic virus type I (HTLV-I) is prevented by cytotoxic T lymphocytes. *Blood*. 2000;95(4):1386–1392.

24. Billman MR, et al. Single-cell heterogeneity and cell-cycle-related viral gene bursts in the human leukaemia virus HTLV-1. *Wellcome Open Res*. 2017;2:87.

25. Mahgoub M, et al. Sporadic on/off switching of HTLV-1 Tax expression is crucial to maintain the whole population of virus-induced leukemic cells. *Proc Natl Acad Sci U S A*. 2018;115(6):E1269–E1278.

26. Yamashita Y, et al. HLA-DP(84Gly) constitutively presents endogenous peptides generated by the class I antigen processing pathway. *Nat Commun*. 2017;8:15244.

27. Sugata K, et al. Affinity-matured HLA class II dimers for robust staining of antigen-specific CD4+ T cells. *Nat Biotechnol*. 2021;39(8):958–967.

28. Morton LM, et al. Lymphoma incidence patterns by WHO subtype in the United States, 1992-2001. *Blood*. 2006;107(1):265–276.

29. Satake M, et al. The incidence of adult T-cell leukemia/lymphoma among human T-lymphotropic virus type 1 carriers in Japan. *Leuk Lymphoma*. 2015;56(6):1806–1812.

30. Yasunaga J, et al. Impaired production of naive T lymphocytes in human T-cell leukemia virus type I-infected individuals: its implications in the immunodeficient state. *Blood*. 2001;97(10):3177–3183.

31. Bending D, Ono M. From stability to dynamics: understanding molecular mechanisms of regulatory T cells through Foxp3 transcriptional dynamics. *Clin Exp Immunol*. 2019;197(1):14–23.

32. Yoshida M. Multiple viral strategies of HTLV-1 for dysregulation of cell growth control. *Annu Rev Immunol*. 2001;19:475–496.

33. Basbous J, et al. The HBZ factor of human T-cell leukemia virus type I dimerizes with transcription factors JunB and c-Jun and modulates their transcriptional activity. *J Biol Chem*. 2003;278(44):43620–43627.

34. Zhao T, et al. HTLV-1 bZIP factor enhances TGF-β signaling through p300 coactivator. *Blood*. 2011;118(7):1865–1876.

35. Kozako T, et al. Reduced frequency, diversity, and function of human T cell leukemia virus type 1-specific CD8+ T cell in adult T cell leukemia patients. *J Immunol*. 2006;177(8):5718–5726.

36. Ovcinnikovs V, et al. CTLA-4-mediated transendocytosis of costimulatory molecules primarily targets migratory dendritic cells. *Sci Immunol*. 2019;4(35):eaaw0902.

37. Ko HS, et al. Ia determinants on stimulated human T lymphocytes. Occurrence on mitogen- and antigen-activated T cells. *J Exp Med*. 1979;150(2):246–255.

38. Broeren CP, et al. Activated rat T cells synthesize and express functional major histocompatibility class II antigens. *Immunology*. 1995;84(2):193–201.

39. Baecher-Allan C, et al. MHC class II expression identifies functionally distinct human regulatory T cells. *J Immunol*. 2006;176(8):4622–4631.

40. Miyara M, et al. Functional delineation and differentiation dynamics of human CD4+ T cells expressing the FoxP3 transcription factor. *Immunity*. 2009;30(6):899–911.

41. Iberg CA, Hawiger D. Natural and induced tolerogenic dendritic cells. *J Immunol*. 2020;204(4):733–744.

42. Tsang JY, et al. Antigen presentation by mouse CD4+ T cells involving acquired MHC class II:peptide complexes: another mechanism to limit clonal expansion? *Blood*. 2003;101(7):2704–2710.

43. Katsuya H, et al. The nature of the HTLV-1 provirus in naturally infected individuals analyzed by the viral DNA-capture-seq approach. *Cell Rep*. 2019;29(3):724–735.

44. Cano-Gamez E, et al. Single-cell transcriptomics identifies an effectorness gradient shaping the response of CD4+ T cells to cytokines. *Nat Commun*. 2020;11(1):1801.

45. Beyer M, et al. Repression of the genome organizer SATB1 in regulatory T cells is required for suppressive function and inhibition of effector differentiation. *Nat Immunol*. 2011;12(9):898–907.

46. Qiu X, et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods*. 2017;14(10):979–982.

47. La Manno G, et al. RNA velocity of single cells. *Nature*. 2018;560(7719):494–498.

48. Bergen V, et al. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol*. 2020;38(12):1408–1414.

49. Yamagishi M, et al. Chronological genome and single-cell transcriptome integration characterizes the evolutionary process of adult T cell leukemia-lymphoma. *Nat Commun*. 2021;12(1):4821.

50. Maeda M, et al. Origin of human T-lymphotrophic virus I-positive T cell lines in adult T cell leukemia. Analysis of T cell receptor gene rearrangement. *J Exp Med*. 1985;162(6):2169–2174.

51. Cook LB, et al. The role of HTLV-1 clonality, proviral structure, and genomic integration site in adult T-cell leukemia/lymphoma. *Blood*. 2014;123(25):3925–3931.

52. Street K, et al. Slingshot: cell lineage and pseudo-time inference for single-cell transcriptomics. *BMC Genomics*. 2018;19(1):477.

53. Van den Berge K, et al. Trajectory-based differential expression analysis for single-cell sequencing data. *Nat Commun*. 2020;11(1):1201.

54. Ueno T, et al. Reconstitution of anti-HIV effector functions of primary human CD8 T lymphocytes by transfer of HIV-specific alphabeta TCR genes. *Eur J Immunol*. 2004;34(12):3379–3388.