Distinctive genomic features of human T-lymphotropic virus type 1-related adult T-cell leukemia-lymphoma in Western populations

Caroline S. Myers,[1] Eli Williams,[2] Carlos Barrionuevo Cornejo,[3] Georgios Pongas,[4] Ngoc L. Toomey,[4] Jose A. Sanches,[5] Maxime Battistella,[6] Samuel Mo,[1] Melissa Pulitzer,[7,8] Cristopher A. Moskaluk,[2] Govind Bhagat,[8,9] Kenneth Ofori,[9] Jonathan J. Davick,[10] Octavio Servitje,[11] Denis Miyashiro,[5] Fina Climent,[11] Kimberley Ringbloom,[1] Daniela Duenas,[3] Calvin Law,[1] Sandro Casavilca Zambrano,[3] Luis Malpica,[12] Brady E. Beltran,[13] Denisse Castro,[13] Luciana Barreto,[4,14] Carlos Brites,[15] Jennifer R. Chapman,[4] Jaehyuk Choi,[1#] Alejandro A. Gru[16#] and Juan C. Ramos[4#]

[1]Department of Dermatology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA; [2]University of Virginia School of Medicine, Charlottesville, VA, USA; [3]Instituto Nacional de Enfermedades Neoplasicas, Lima, Peru; [4]University of Miami, Miami, FL, USA; [5]Universidade de São Paolo, São Paolo, Brazil; [6]Université de Paris, Paris, France; [7]Memorial Sloan-Kettering Cancer Center, New York, NY, USA; [8]New York-Presbyterian Hospital, New York, NY, USA; [9]Columbia University School of Medicine, New York, NY, USA; [10]University of Iowa, Iowa City, IA, USA; [11]Hospital Universitari de Bellvitge, Barcelona, Spain; [12]University of Texas MD Anderson Cancer Center, Houston, TX, USA; [13]Hospital Nacional Edgardo Rebagliati Martins, Lima, Peru; [14]Instituto Nacional de Câncer José Alencar Gomes da Silva, Rio de Janeiro, Brazil; [15]Federal University of Bahia, Salvador, Brazil and [16]Columbia University Irving Medical Center/New York-Presbyterian, New York, NY, USA

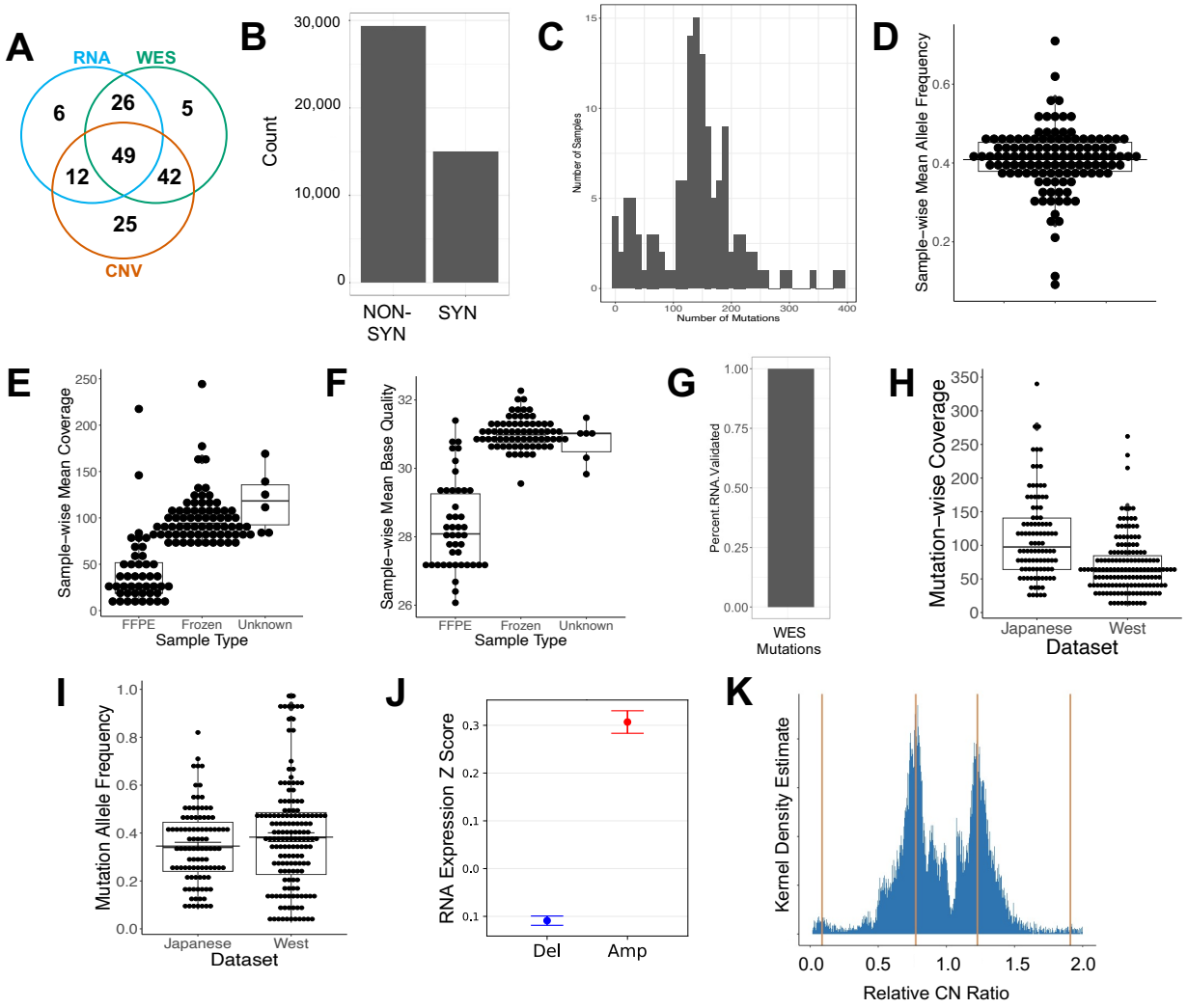
[#]JC, AAG and JCR contributed equally as senior authors.

Correspondence: A.A. Gru
aag2222@cumc.columbia.edu

Received: February 8, 2024.
Accepted: July 29, 2024.
Early view: August 8, 2024.

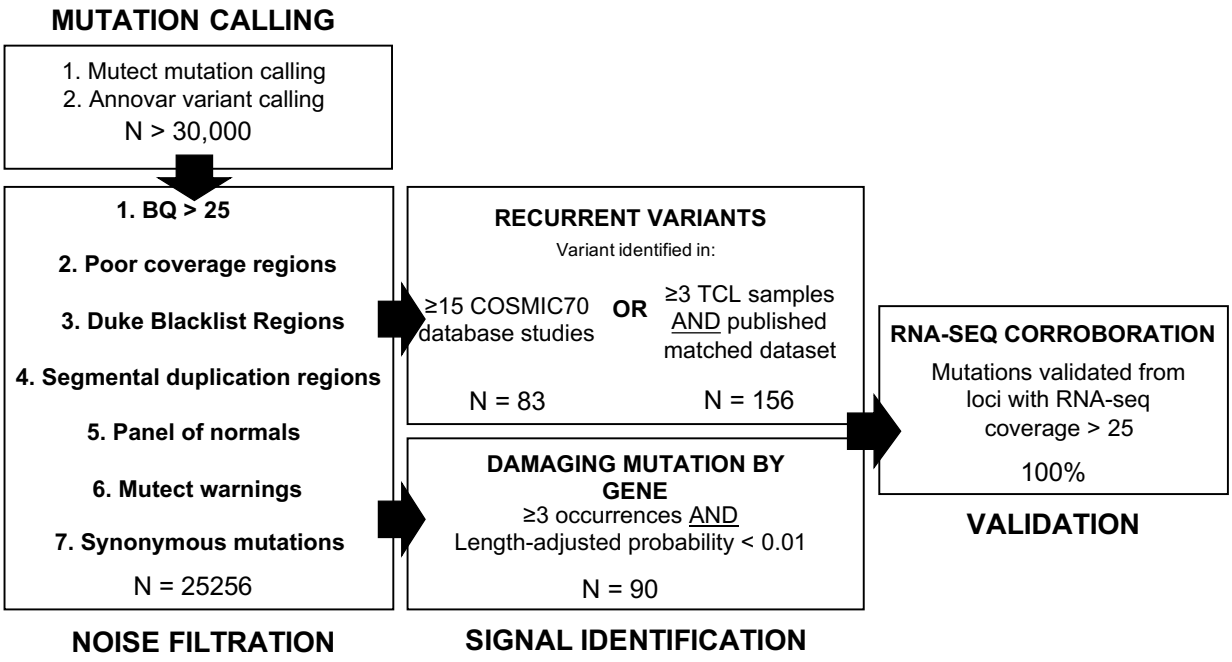
https://doi.org/10.3324/haematol.2024.285233

SUPPLEMENTAL MATERIALS

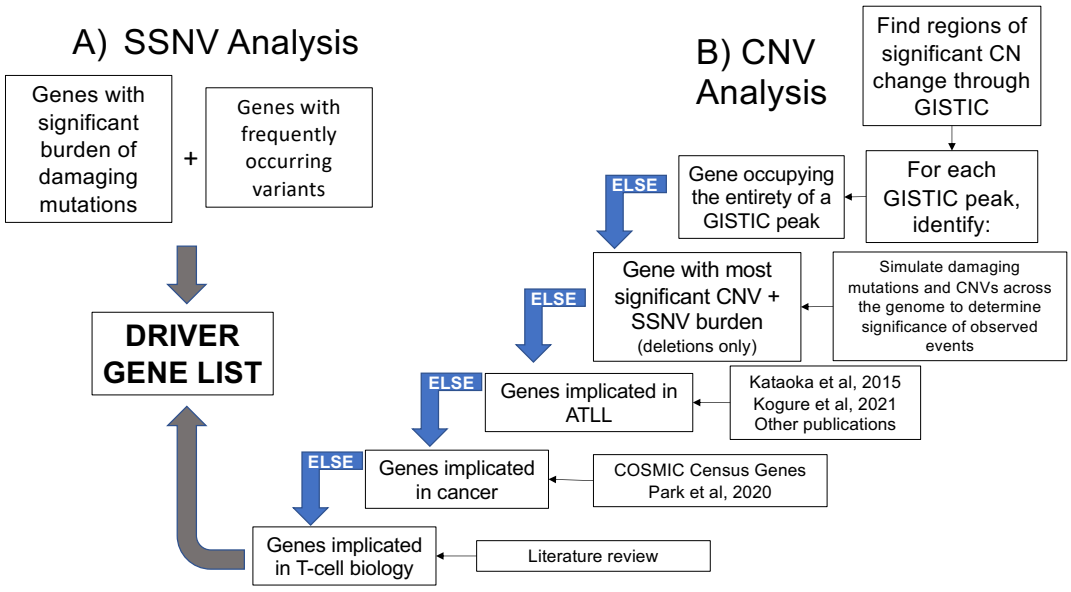


Supplemental Figure S1 | QC Metrics for whole exome and CNV sequencing data. **a**, Venn diagram of Western patients with samples analyzed by RNA-seq (blue) whole exome sequencing (green), copy number variation analysis (orange), or multiple methods. **b,** Ratio of nonsynonymous to synonymous point mutations called across all samples. **c**, Histogram of numbers of point mutations called per sample. **d**, Sample-wise averages of mean allele frequencies. **e-f**, Sample-wise **e**, mean coverage and **f**, mean base quality by sample type. Dots represent individual values and boxplot includes, mean, first and third quartiles. Values not adjacent or connected to the boxplot by a central line represent outliers**. g,** Percentage of putative

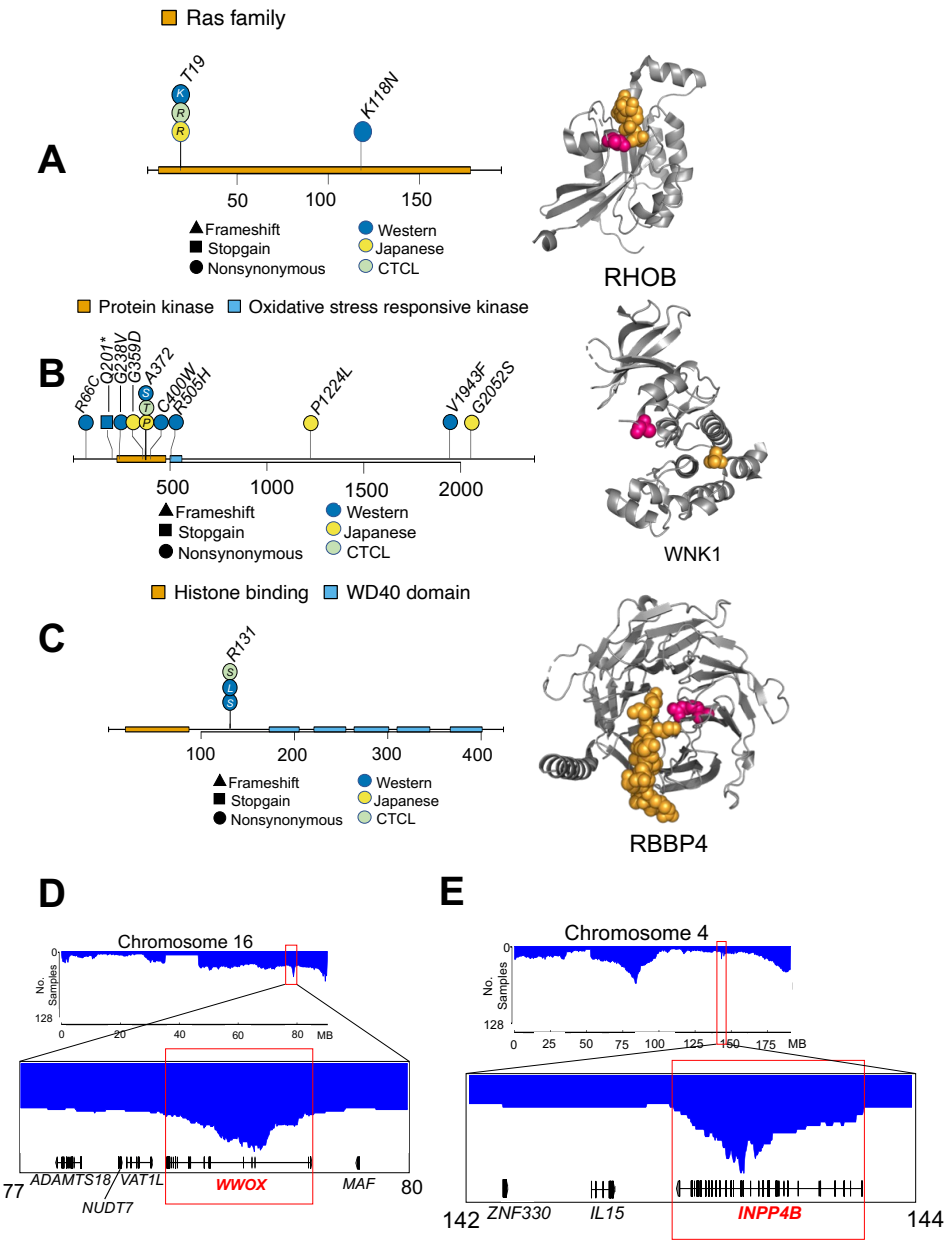
driver mutations with DNA- and RNA-seq coverage > 25 validated by RNA-seq data**. h,** Sequencing coverage and **i,** mutation allele frequency of variants of interest identified in Japanese (left) and Western (right) patients. **j,** RNA-expression of genes within significant copy-number variation (CNV) regions of deletion ("del", blue, left) versus amplification ("amp", red, right). Dots represent mean values and error bars represent standard error**. k,** Kernel density estimate (KDE) of CNV fold change values across all samples. Only CNVs different from wildtype (i.e. fold change not equal to 1) were included. CNV fold change is displayed across the x axis. Vertical lines represent KDE-represented modes.



Supplemental Figure S2 | Schematic of quality control methodology used to identify putative driver genes from point mutation data and eliminate potential sources of bias, noise and ambiguity. Mutations and variants were called by Mutect and Annovar, respectively. They were filtered out based on the criteria in the bottom left, then filtered in by the criteria in the center. Variants of interest were validated by analysis of orthogonal RNA-seq data (far right). A detailed textual explanation can be found in the supplemental methods.

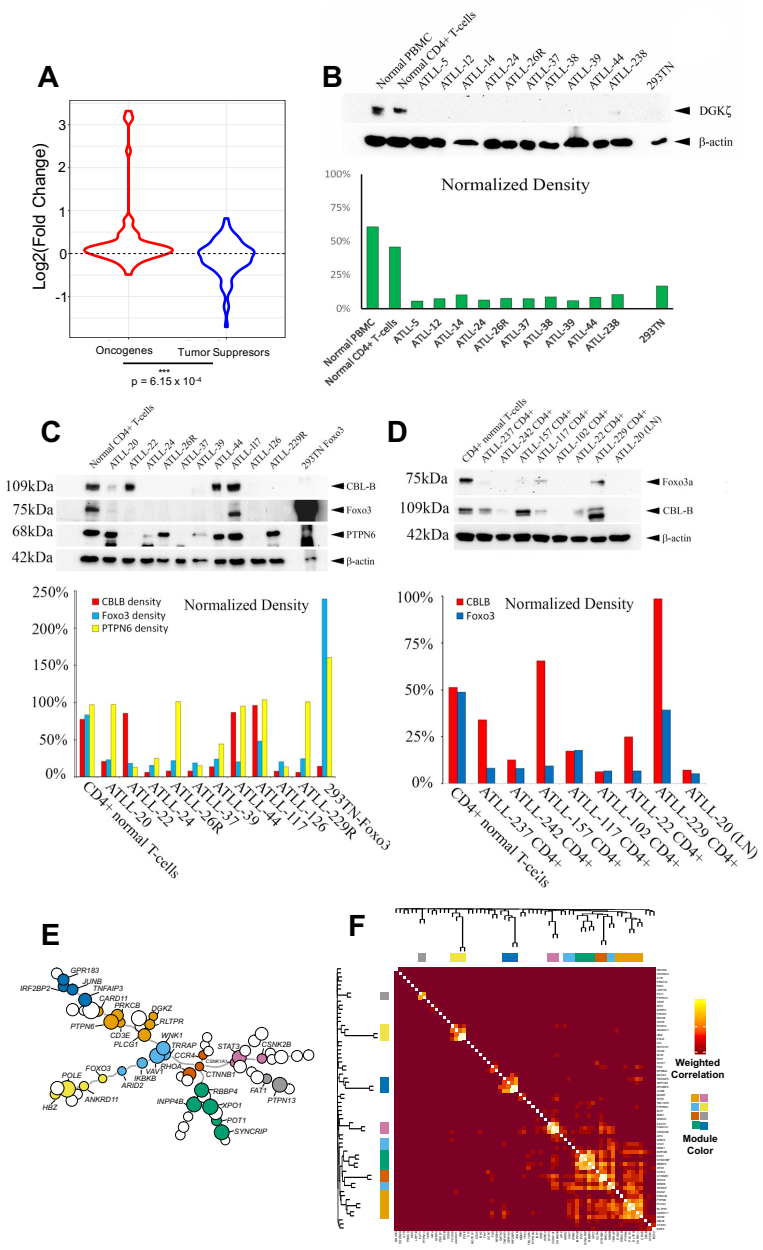


Supplemental Figure S3| Schematic of the hierarchical method used to call driver genes in peaks determined to be significant by GISTIC2.0. Regions of significant amplification/deletion by GISTIC analysis were first examined for putative driver genes satisfying the criterium in the top right. If that criterium were satisfied by a single gene, that gene was called as the putative driver gene and the search process was stopped. If no gene satisfied that criterium, we proceeded to the next criterium, and so forth. If no putative drivers were found by this stepwise search, the search expanded to encompass three neighboring genes on either side of the peak. CNV: copy number variation; SSNV: somatic single nucleotide variation (point mutation). A detailed textual explanation can be found in the Supplemental Methods.



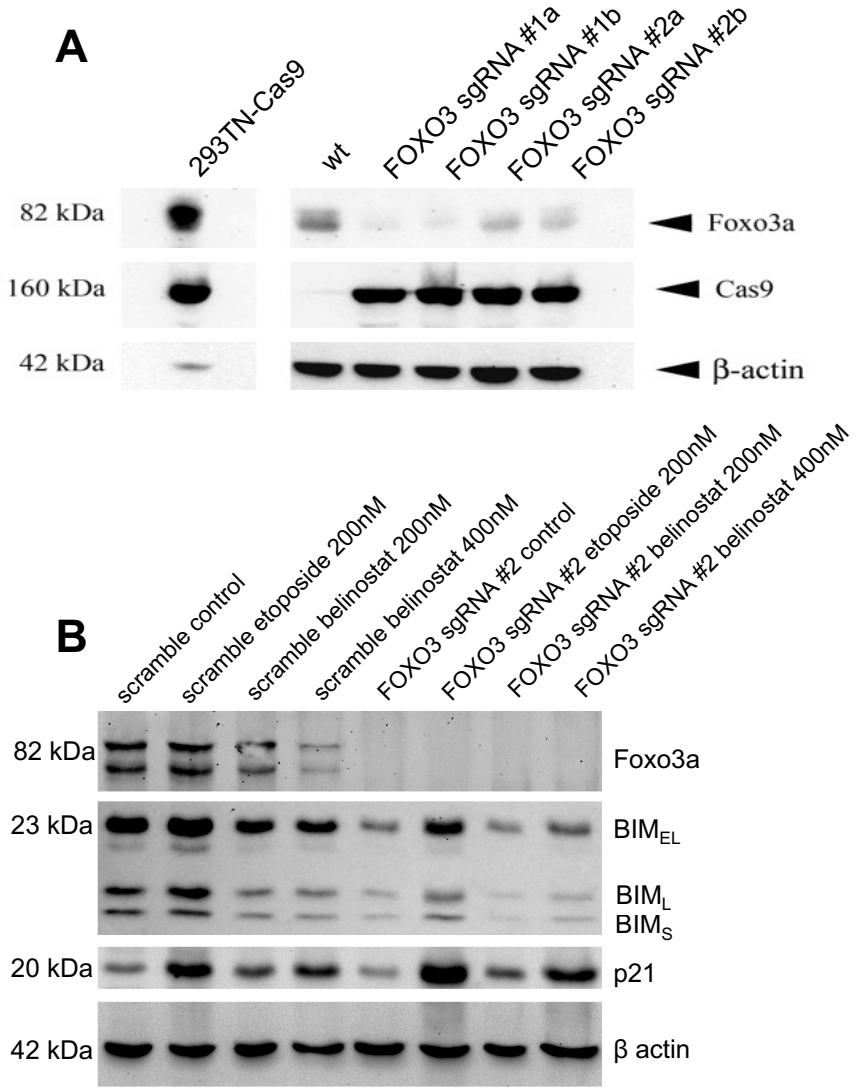
Supplemental Figure S4 | Putative driver genes identified by point mutation and copy number mutation analysis. **a-c** Novel ATLL putative driver genes identified by point analysis. Lollipop plots on the left show the location of point mutations, with mutation types and cohorts indicated by the legend. Right panels show recurrently mutated residues in magenta within the context of their respective protein structures; ligands are shown in yellow or orange. **a**, Activated (phosphorylated) WNK1, with phosphate group shown in orange. **b**, RBBP4 is bound to ZNF827, shown in orange. **c**, RHOB is bound to GDP, shown in orange **d-e** Histograms indicating

segments of copy number loss showing overlapping deletion in **d**, *WWOX* and **e**, *INPP4B.* Surrounding genes are indicated at the bottom of the figure.



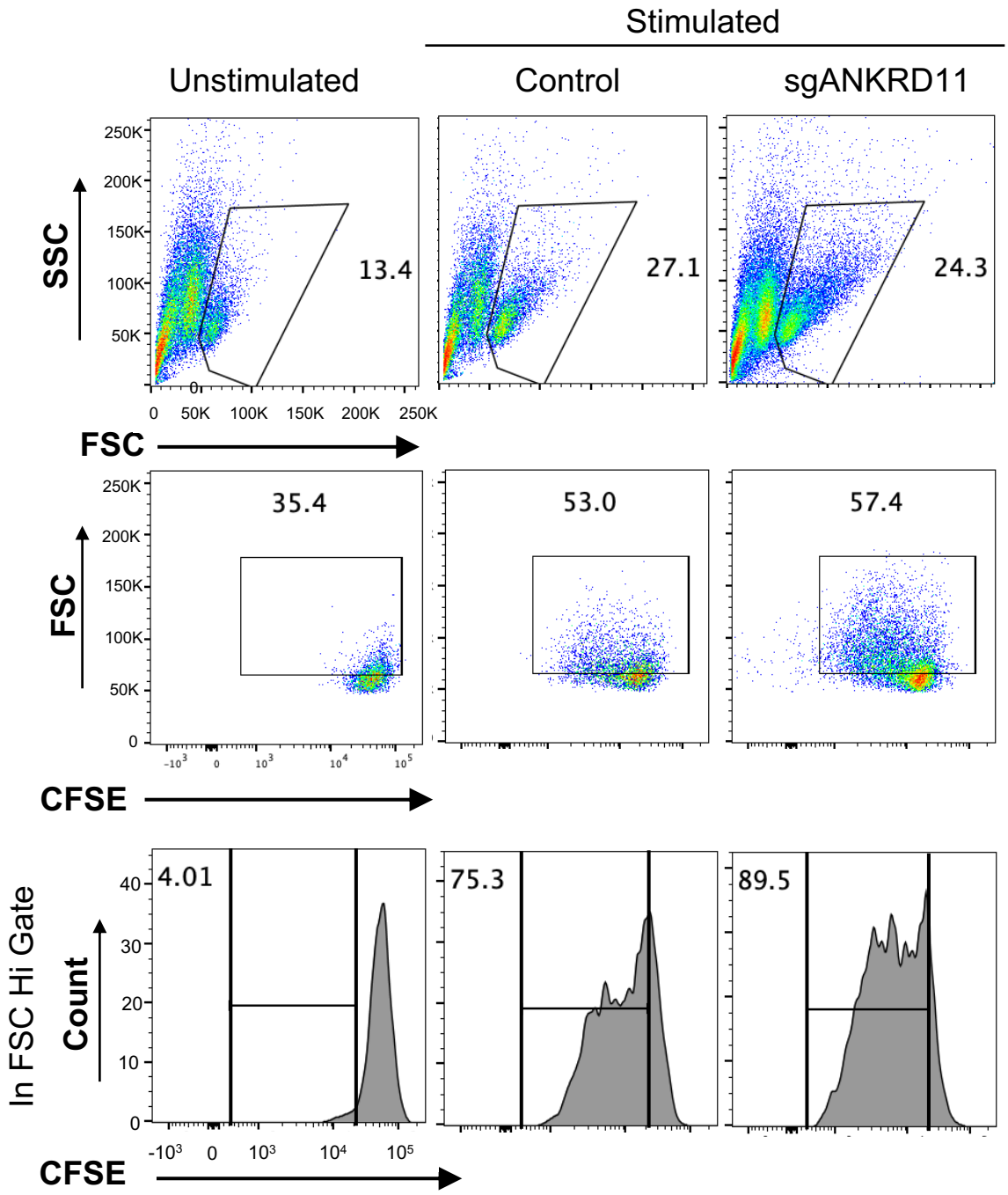
Supplemental Figure S5 | Functional analyses of putative ATLL driver genes. **a,** Violin plots showing log fold change (**LFC**) of sgRNAs for putative oncogenes and putative tumor suppressor in genome-wide CRISPR amplification screens for T-cell receptor dependent proliferation and cytokine production. sgRNA changes for putative oncogenes are shown on the left; those for

putative tumor suppressors are shown on the right. P-value represents the significance of the difference between these distributions (student's two-tailed t-test). **b-d,** Western blot analysis of protein levels in ATLL patient samples. Band density was normalized against β-actin, shown as loading control. **b**, DGKZ protein levels. **c**, FOXO3, CBLB, and PTPN6 protein levels. **d**, FOXO3 and CBLB protein levels. **e-f,** Weighted gene correlation network analysis of RNA expression data for putative oncogenes/tumor suppressors as well as for the viral protein HBZ. Colors represent correlated modules of genes. **e,** Network diagram of RNA expression data, with edges representing maximal correlation between genes (nodes). Node size represents relative levels of gene expression, with larger nodes representing more highly expressed genes. Colors represent closely correlated modules of genes, as determined by network analysis of pairwise weighted correlation shown in **f,** a heatmap of pairwise weighted correlation. Dark red indicates low correlation; yellow-white represents high correlation. Gene members of correlated modules are indicated with labels above and to the left of the graph.



Supplemental Figure S6 Protein levels of Foxo3a and target gene proteins in patient-derived acute ATLL cell lines. β-actin was used as a loading control. **a,** Foxo3a levels in early CRISPR-Cas9 FOXO3 knockouts in ATLL97-c. Two FOXO3 knockout constructs with distinct single guide (sg) RNA sequences were established. Each expressed a distinct FOXO3-specific sgRNA (numbered 1 and 2). Protein levels for each construct were measured in duplicate. 293TN-Cas9 expressing (far left) and unmodified wild-type (wt) ATLL97-c cells were used as a positive control for Cas9 expression and Foxo3a, respectively. **b,** Foxo3 and target gene proteins (BIM and p21) in scramble sg vs. FOXO3 sg #2 CRISPR knockout before and after

anti-neoplastic agents (etoposide and belinostat) at the specified drug concentrations. Control cells were treated with DMSO. The different variants of BIM are shown as: BIM_{EL} = extra-long BIM, BIM_L = long BIM, BIM_S = short BIM.



Supplemental Figure S7 | CFSE proliferation assay gating strategy. Human CD4+ T cells isolated from three different donors were labelled with CFSE 10 days after targeted gene knockdown. After CFSE staining, cells were cultured with or without stimulation. Stringent gating excluded debris and dying cells to capture only live and activated lymphocytes (Top row: Side Scatter, SSC vs Forward Scatter, FSC). Secondary gating for larger proliferating activated cells (Second row: FSC vs CFSE). Proliferation (% divided) of CD4+ T cells with ANKRD11 knockouts (right) were analyzed and compared to non-targeting control (second from left) (Bottom row). Gating was based on diluted CFSE signal in unstimulated controls (far left).

Supplemental Table S1 | Significant Variants determined by COSMIC70 database. This table includes variants that are seen at least once in the ATLL West cohort and in COSMIC70 at least 15 times. Variants with all reads occurring on one strand, all reads starting at the same base pair or with all mutant allele frequencies (MAFs) less than 0.05 were removed as artefacts. “ATLL + TCL" represents the frequency of the specified mutation in the ATLL dataset combined with published T-cell lymphoma” datasets. Samples with DNA sequencing and RNA sequencing available, and with coverage greater than 25 by both DNA- and RNA-sequencing in the location of interest, were considered eligible for RNA-seq mutation validation. Column H represents the number of samples examined in RNA-seq mutation validation as determined by the eligibility criteria above. Column I represents the number of samples with the DNA-seq determined damaging mutation validated by RNA-seq. The "ATLL Expression Quartile" represents how highly that gene is expressed by RNA-seq, with 4 being the most highly expressed quartile and 1 being the least. AA = amino acid, chr = chromosome, pos = position, ref = reference amino acid, alt = mutant amino acid, TCL = T-cell lymphoma. ATLL Western Cohort N = 122; ATLL All N = 122 (West) + 83 (Japanese) = 205; ATLL + TCL N = 1145.

Supplemental Table S2 | Significant variants as determined by frequency of recurrence in ATLL samples. Variants with all reads occurring on one strand, all reads starting at the same base pair or with all MAFs less than 0.05 were removed as artefacts. Counts of recurrences by amino acid positions includes samples analyzed by targeted sequencing. "ATLL + TCL" represents the frequency of the specified mutation in the ATLL dataset combined with published T-cell lymphoma datasets. Samples with DNA sequencing and RNA sequencing available, and with coverage greater than 25 by both DNA- and RNA-sequencing in the location of interest, were considered eligible for RNA-seq mutation validation. Column I represents the number of samples examined in RNA-seq mutation validation as determined by the eligibility criteria above. Column J represents the number of samples with the DNA-seq determined damaging mutation validated by RNA-seq. The "ATLL Expression Quartile" represents how highly that gene is expressed by RNA-seq, with 4 being the most highly expressed quartile and 1 being the least. AA = amino acid, chr = chromosome, pos = position, ref = reference amino acid, alt = mutant amino acid, TCL = T-cell lymphoma. ATLL Western Cohort N = 122; ATLL All N = 122 (West) + 83 (Japanese) = 205; ATLL + TCL N = 1145.

Supplemental Table S3 | Driver genes as determined by frequency of occurrence. Damaging mutations included in this analysis are splicing, startloss, stopgain, stoploss, frameshift insertion, frameshift deletion, frameshift substitution, nonframeshift deletion, nonframeshift insertion, nonframeshift substitution. Genes satisfying the criteria that all damaging mutations were at the same positiona and all mutant allele frequencies (MAF) < 0.05 were considered artifacts and removed. There were 1024 genes with damaging mutations before evaluating genes for artifact as described above. After filtering, there were 977 genes remaining. Driver genes were determined by first examining all genes with n > 3 damaging mutations in the ATLL West cohort. Only mutations with MAF > 0.05 were considered in this count. Possible driver genes were then examined for significance by calculating the probability of observing n damaging mutations in that

gene, adjusting for gene length and expression quartile. All damaging mutations were considered, regardless of MAF. Probability was calculated based on the cumulative binomial distribution. P-values of 0 represent values less than 1×10^{-14}. Genes likely to have the observed number of mutations by chance ($p > 0.01$) were excluded. “ATLL + TCL" represents the frequency of the specified mutation in the ATLL dataset combined with published T-cell lymphoma datasets. Samples with DNA sequencing and RNA sequencing available, and with coverage greater than 25 by both DNA- and RNA-sequencing in the location of interest, were considered eligible for RNA-seq mutation validation. Column K represents the number of samples examined in RNA-seq mutation validation as determined by the eligibility criteria above. Column L represents the number of samples with the DNA-seq determined damaging mutation validated by RNA-seq. ATLL Western Cohort N = 167; ATLL All N = 167 (West) + 83 (Japanese) = 250; ATLL + TCL N = 1190.

Supplemental Table S4 | Driver Genes based off of GISTIC amplifications. Areas of significant amplification were determined from Oncoscan data using the GISTIC2 algorithm. Regions of frequent deletion/amplification (e.g. T-cell receptor-encoding regions) were excluded. Genes were then called within significant GISTIC peaks ($q < 0.001$) using their ENSEMBL database-specified genetic locations. A gene was considered deleted within a region if at least one coding exon was within a GISTIC deletion peak. From each peak, a driver gene was inferred by examining peaks by the following hierarchical criteria, from most important to least: 1) Sole gene within the peak expressed in ATLL T-cells. 2) Having a SNV + CNV burden significantly greater than that observed in a simulation of randomly distributed amplifications and mutations. Because not all nonsynonymous mutations will be activating mutations, only genes deemed significant by a separate recurrent variant analysis (see tables S1 and S2) were kept as driver genes in this filtration step. 3) Implicated in CTCL in previously published studies. 4) Tier-1 evidence for the gene as an oncogene in COSMIC. 5) A gene within 3 genes of the peak had tier-

1 evidence for the gene as an oncogene in COSMIC. 6) Documented in the literature to be involved in T-cell biology or 7) tumorigenesis. AFR = Afro-caribbean descent, AMR = Indigenous American descent, SAS = Southeast Asian descent, chr = chromosome, SNV = single nucleotide variant.

Supplemental Table S5 | Driver Genes based off of GISTIC deletions. Areas of significant deletion were determined from Oncoscan data using the GISTIC2 algorithm. Regions of frequenct deletion/amplification (e.g. T-cell receptor-encoding regions) were excluded. Genes were then called within significant GISTIC peaks ($q < 0.001$) using their ENSEMBL database-specified genetic locations. A gene was considered deleted within a region if at least one coding exon was within a GISTIC deletion peak. From each peak, a driver gene was inferred by examining peaks by the following hierarchical criteria, from most important to least. 1) Sole gene within the peak expressed in ATLL T-cells. 2) Having a SNV + CNV burden significantly greater than that observed in a simulation of random probability based upon the binomial distribution. Genes were required to have at least one damaging position and to have a likelihood ratio > 5 (5x more significant than next most significant gene). 3) Implicated in CTCL in previously published studies. 4) Tier-1 evidence for the gene as an oncogene in COSMIC. 5) A gene within 3 genes of the peak had tier-1 evidence for the gene as an oncogene in COSMIC. 6) Documented in the literature to be involved in T-cell biology or 7) tumorigenesis. AFR = Afro-caribbean descent, AMR = Indigenous American descent, SAS = Southeast Asian descent, chr = chromosome, SNV = single nucleotide variant.

Supplemental Table S6 | Putative driver genes. Putative roles were determined by corroboration of mutational patterns with oncological roles as described in the literature. Column E describes the previous implication of the indicated gene in ATLL, cancer, or neither ("novel"), based upon our review of the literature.

Supplemental Table S7 | Frequency of gene mutations by subtypes. Displayed frequencies represent the percentage of samples with either a copy number variation (CNV) or point mutation in the given gene. The significance of mutational differences between acute and lymphomatous subtypes is shown in column F, as calculated using Fisher's exact test. Consideration of mutational differences between acute and lymphomatous subtypes defined gene "mutation" as amplifications + all point mutations for putative oncogenes and deletions + damaging mutations for putative tumor suppressors.

Supplemental Table S8 | IRF4 Mutation validation by Sanger Sequencing

Supplemental Table S9 | Sample Metadata. Data types include whole exome sequencing (WES), RNA-sequencing, Oncoscan or Japanese whole-exome sequencing. Sample types include frozen versus formalin-fixed and paraffin-embedded (FFPE). Starred data types represent samples that were eliminated during quality control. Age represents the age at first diagnosis. Survival Wks represents the weeks from first patient encounter to patient death (status: deceased) or loss to follow-up (status: lost). Interferon-alpha therapy response (IFN Response) coding: 1 = complete response, 2 = partial response, 3 = stable disease, 4 = progressive disease or no response, 5 = not evaluable. Chemotherapy response (Chemo response) coding: 1 = complete response, 2 = partial response, 3 = stable disease, 4 = progressive disease or no response, 5 = not evaluable. Ethnicity was determined by SNP analysis using Eth-Seq, as described in the text. AFR = Afro-Caribbean, AMR = Indigenous American, SAS = Southeast Asian, EUR = European.

SUPPLEMENTARY METHODS

Inclusion Criteria and Clinical Classification. 165 patients with a confirmed diagnosis of ATLL from centers in the United States (primarily Miami, Florida), South America (primarily Peru and Brazil), and Europe (France and Spain) were included in this study. The diagnosis of ATLL was made for all cases after meeting the following criteria: serologic evidence of HTLV-1 by enzyme-linked immunosorbent assay confirmed by reflex western blot and identification of clonal CD4+CD7−CD25+/− T cells in peripheral blood or tissues as determined by histology, immunophenotyping, and gene rearrangement studies. HTLV-1 PCR validation was used in skin biopsies of limited quantities. Patients were classified according to the Shimoyama criteria into acute, lymphomatous, chronic, and smoldering ATLL[1]. Chronic ATLL with LDH elevation < 2 times (2N) the upper normal limit value was classified as unfavorable chronic[2]. Patients with lymphoma features (e.g. presenting with large or bulky lymphadenopathy) and absolute lymphocyte count $<4 \times 10^9$/L were classified as lymphomatous type regardless presence or absence of blood-circulating ATLL cells. In equivocal cases resembling cutaneous T-cell lymphoma, HTLV-1 DNA was detected by PCR in diagnostic biopsies. Before genomic analyses, the diagnoses in all cases were confirmed by at least 2 independent hematopathologists. A minimum purity of 20% neoplastic lymphocytes was necessary for inclusion in the study.

All patient samples were collected under protocols approved by the local Internal Review Boards from participating institutions in accordance with the Declaration of Helsinki. ATLL specimens used for molecular studies were obtained from PBMCs of patients with leukemic presentation or from residual formalin-fixed, paraffin-embedded (FFPE) tumor tissue after informed consent was obtained. This study also included cryopreserved or FFPE tumor samples from deceased patients.

Demographic Data Collection When available, patients' age, sex, country of origin, biopsy site, total white blood cell count, absolute lymphocyte count, serum calcium level, LDH level, overall survival time, treatment administered, treatment response according to the International Consensus Meeting proposal,[3] and comprehensive immunophenotyping on peripheral blood mononuclear cells (PBMCs) and tissue biopsies were collected. Due to the retrospective nature of this study, which included patients from resource-poor areas with limited clinical data, "complete" datasets included patient age, sex, geographic region, and response to chemotherapy and/or AZT-interferon therapy. Patient ethnicity was determined based upon single nucleotide polymorphisms (SNPs) using EthSeq.[4]

DNA and RNA Isolation and Sequencing. Genomic DNA was isolated from cell pellets using the E.Z.N.A.® Tissue DNA Kit (Omega Bio-Tek, inc., Norcross, GA) following manufacturer instruction. RNA was isolated from cells using the E.Z.N.A Total RNA Kit (Omega Bio-Tek, inc., Norcross, GA) following manufacturer protocol. Library preparation and sequencing was completed by Admera Health as previously described.[5] Whole Exome Sequencing (WES), FFPE RNA-Seq, and fresh RNA-seq libraries were prepared using KAPA Hyper Prep Kit (Illumina), SMARTer Stranded Total RNA-Seq Kit, and SMART-Seqv4 Ultra Low Input RNA Kit (Takara Bio) followed by NexteraXT DNA Library Prep Kit (Illumina), respectively. WES library pools were loaded onto an Illumina Hiseq in 2 x 150 bp format. RNA-Seq samples were sequenced on an Illumina Hiseq with a read length configuration of 150 PE.

Somatic Variant Calling. The tissue samples collected in this study came from rural regions with limited availability of health care resources. Where possible, we collected fresh/frozen samples for genomic analyses. However, formalin fixing and paraffin embedding (FFPE) was often the only feasible preservation method that could withstand the transportation necessary to get samples to a sequencing facility. Due to this variability in preservation, as well as variability in

sequencing quality that we observed (see Supplemental Fig. S1), we employed the following filtration methods to identify only high-confidence variants of interest that were orthogonally validated in other datasets.

We first utilized a "rule-out" methodology to filter out low-quality or high-uncertainty mutation calls. Somatic variants were called using Mutect1 for unmatched samples and Mutect2 for matched samples.[6] Calls were filtered by base quality (>25), mapping quality (>29), strand orientation bias[7] (<0.8), alternate allele read depth (>3) and alignment score. Variants falling within the Duke Blacklisted regions,[8] regions of frequent segmental duplication or regions with low exome coverage were removed. SNPs frequently seen (n > 3) in publicly available panels of normal samples including Kaviar, ExAC, Rockefeller, dbSNP and 1000 genomes were also removed.[9-13] Variants with mutation allele frequencies less than 5% were removed. Variants with all reads occurring on one strand or all reads beginning at the same base pair were considered artefacts and were removed. Variants were validated by orthogonal RNA-seq data where possible (Supplemental Fig. S1G).

Exonic function and amino acid alterations were determined using Annovar.[14] Variants with amino acid change classified as "none," "missing," or "UTR3" by Annovar as well as mutations occurring outside of a gene coding region were removed. Damaging mutations were defined as splicing, start-loss, stop-gain, stop-loss, frameshift insertion/deletion/substitution and non-frameshift insertion/deletion/substitution mutations. Genes with recurrent damaging mutations were defined as those with greater than three damaging mutations in the Western ATLL cohort. Gene significance ($p < 0.001$) was determined by using a cumulative binomial distribution to calculate the probability of observing a given number of mutations in that gene, adjusting for gene length and expression quartile.

We then employed a "rule-in" methodology to orthogonally validate variants against published datasets. Variants of interest were identified by two frequency-based criteria: first, variants seen at least 15 times in the COSMIC70 database[15]; second, variants seen in at least

one matched T-cell lymphoma (TCL) sample and at least three samples in the combined ATLL + publicly available TCL dataset.[15-17] Recurrent tumor suppressors (n > 3 damaging mutations with allele frequency > 0.05) were examined for significance ($p < 0.01$) using a negative binomial distribution adjusting for gene length. An illustration of this pipeline can be seen Supplemental Fig. S2.

CNV Putative Driver Gene Calling Initial peaks were called using the Thermo-Fisher Oncoscan CNV assay, a clinically-utilized and industry-validated commercial copy number variation (CNV) assay. Copy number calls were filtered to remove peaks within the Duke Blacklisted Regions, regions of frequent segmental duplication, or regions with low exome coverage[8]. CNVs >500 kbp and LOH >3 Mbp were manually reviewed by a molecular pathologist for quality. Purity was estimated by applying a kernel density estimate to evaluate the modes of CNV relative ratios (Supplemental Fig. S1K).

These data were then combined with a published dataset of called copy number variation peaks[18] kindly provided by the Ogawa lab to form a combined Japanese and Western dataset. Dataset-wide significance was determined using GISTIC2.0 with a residual q-value cutoff of 0.0001.[19] If multiple peaks of the same directionality were called in the same cytoband, only the more significant peaks were kept. Peaks falling within T-cell receptor gene regions were discarded because these were presumed to have occurred during thymic development. Putative driver genes within a peak were identified in a hierarchical fashion as depicted in Supplemental Figure S3. First, the significance of a gene's mutational burden was determined through simulating the distribution of dataset mutations across the genome 10,000 times. Deleted genes with more damaging mutations than expected by chance and with five times greater significance than the next most significant gene were determined to be driver genes. If no driver gene was identified through this method, our algorithm then searched hierarchically for known ATLL driver mutations.[18, 20, 21] If none were found, the peak was then searched for known tier-one

oncogenes/tumor suppressors in the COSMIC70 Cancer Census database and in-house databases of cancer-associated mutations.[15, 16, 22, 23].If none were found, the search range was expanded to the ten neighboring genes in either direction and another iteration underwent. If no driver gene was identified within the peak by this analysis, a literature search was conducted to manually identify any putative oncogenes/tumor suppressors.

RNA-Seq Analysis. RNA-seq data collected from fresh CD4+ tumor cells and FFPE-preserved tissue samples were each analyzed separately as previously described.[5, 16, 17] Sequencing data was aligned using STAR. Counts were tallied with HT-Seq and normalized using DESeq2. Genes of interest were examined for differential expression between mutant and wildtype samples. Genes of interest were also examined for modules of correlated expression using weighted gene correlation network analysis.[24] Due to known differences between FFPE and fresh/frozen sample RNA sequencing quality,[25] FFPE and fresh/frozen samples in our cohort were each analyzed independently.

Isolation and culture of primary human CD4+ T-cells. Primary human T-cells were isolated from enriched leukapheresis products (Leukopaks, AllCells). PBMCs were isolated from Leukopaks by Ficoll-Hypaque gradient centrifugation. We used Dynabeads CD4 Positive Isolation Kit (Invitrogen #11331D) to isolate CD4+ T-cells from these PBMCs by magnetic positive selection. Isolated CD4+ T-cells were frozen in Fetal Bovine Serum (FBS) with 10% DMSO for later use. Upon thawing, cells were cultured in complete RPMI consisting of RPMI-1640 medium (Gibco #21875034), 10% FBS, 1% pen/strep, 1mM Sodium Pyruvate and 10mM HEPES. Thawed cells were stimulated with plate-bound anti-human CD3 (OKT3, Biolegend #317326) at 10μg/mL and soluble anti-human CD28 (CD28.2, Biolegend #302943) at 5μg/mL with IL-2 at 50U/mL at 1x 10^6 cells/mL.

CRISPR-based gene knockdown in human T-cells. Transfection of human CD4+ T-cells was performed as previously described.[26] Briefly, guide RNAs were designed using the CRISPick online design tool by Broad Institute [27] and purchased from Integrated DNA technologies (IDT, Alt-R CRISPR Cas9 crRNA). Following manufacturer protocol, crRNAs were duplexed with tracrRNA (IDT #1072534) for 40min at 37°C in 5% CO_2 incubator. They were then complexed with Cas9 protein (Macrolab, Berkeley, 40μM stock) at 1:1 molar ratio for 15min at 37°C. After 72 hours of stimulation, cells were collected, pelleted and resuspended in Lonza electroporation buffer P3 (Lonza #V4XP-3032) at 0.5 x 10^6 cells/20μL. Cells were electroporated at 0.5 x 10^6 per well in 16-well cuvettes using pulse code EH115 (Lonza #AAF-1002X). The total number of cells for electroporation was scaled as required. Immediately after electroporation, 80μL of pre-warmed complete RPMI culture media were added to each well of the cuvettes. Cuvettes were placed in incubator at 37°C and 5% CO_2 for 15 minutes. Cells were then transferred to 96-well plates in complete RPMI media containing 50U/mL IL-2 at 2.5 x 10^6 cells/mL and rested for 4 days, adding fresh media with IL-2 at 50U/mL on day 2.

T-cell proliferation assay pipeline After CD4+ T-cell isolation (Day 0), cells were stimulated with plate-bound anti-human CD3/CD28 and IL-2 at 50U/mL as described above. Following 72 hours of stimulation (Day 3), cells were electroporated with crRNP complexes and cultured in complete RPMI media with 50U/mL IL-2 for 4 days with no stimulation. After 4 days of culture (Day 7) cells were split to make protein lysate pellets and genomic DNA pellets for CRISPR knockout verification via western blot and TIDE sequencing, respectively. Remaining cells were restimulated with ImmunoCult Human CD3/CD28/CD2 T-cell Activator (STEMCELL #10970) in complete RPMI media with IL-2 at 50U/mL for 6 days. ImmunoCult was used at 1/8 of manufacturers recommended dose of 25μL/mL of cell culture. On day 13, cells were stained with CFSE as described above and cultured for 4 days in complete RPMI media without IL-2 and with or without ImmunoCult stimulation. Stimulation with ImmunoCult after CFSE staining was used at

1/16th of manufacturers recommended dose of 25µL/mL of cell culture. After 4 days, cell proliferation was assessed by Fluorescent-activated Cell Sorting (FACS). Gating strategy, including positive controls, is illustrated in Supplemental Fig. 7.

CFSE Staining Carboxyfluorescein succinimidyl ester (CFSE) staining of arrayed cells that were edited with crRNPs was done in 96-well deep well 2.0mL microplates (VWR #75870-796). CFSE (Invitrogen #C34554) was prepared per manufacturer's protocol to make a 5mM stock solution in DMSO. At time of use, this stock was diluted 1:1000 in PBS to make a 5µM working solution that was used to resuspend cells washed in PBS. Cells were transferred from 96-well culture plates into deep well 96-well plates using a manual multichannel pipette. They were washed with 1mL of PBS per well using a 10mL serological pipette. After spinning and decanting, cells were resuspended in 0.2mL of 5µM CFSE and incubated in 37°C tissue culture incubator with 5% CO_2 for 20 minutes. Then, 1mL of RPMI with 10% FBS was added to each well to quench and incubate for 5 minutes in tissue culture incubator. Finally, cells were pelleted and resuspended in pre-warmed complete RPMI and incubated for an additional 10 minutes in tissue culture incubator prior to stimulation.

Pseudo viral transductions. For Lenti viral transduction, 1ug of psPAX2, 1ug of pMD2.G, and 1ug of pCDH over-expressing vectors (1ug of pHIT60, 1ug of pHIT123, and 1ug of pMSCV-U6sgRNA(BbsI)-PGKpuro2A-Foxo3 for retro viral transduction) were complexed with 9ug of EcoTransfect reagent (OzBiosciences, San Diego, CA 92126) and transfected to 70% confluence 293TN in 12 wells plate. Pseudo viral particles were harvested at 48 hours and 72 hours, combined and filtered through a 0.45um syringe filter. 1.0 mL of pseudo viral particles was used to infect 250,000 cells of ATLL-84c or ATTL-97c in 12 wells plate for 8 hours in 37C 5% CO2 incubator. After infection, 1.0 mL of additional 1640 + 10% fetal bovine serum was added to wells

and continue to incubate for 48 hours. Puromycin was added to the cells at a concentration of 1ug/mL for selection of positive clones.

Western Blot Analysis. Whole cells protein lysates (25-50µg) from available PBMCs or frozen solid tumor specimens were fractionated on 8% or 12% SDS-PAGE and transferred by electroblotting onto nitrocellulose membranes (Bio-Rad Laboratories, Hercules, CA). Immunoblottings were performed using the following primary antibodies: DGKZ (Novus Biologicals, LLC, Centennial, CO), Foxo3 (BioVision Inc., Milpitas, CA), β-actin (8H10D10), CBLB (D3C12), PTPN6 (C14H6) BIM (C34C5) and p21 (12D1) (Cell Signaling Technology, Danvers, MA). Protein blots were visualized with either SuperSignal West Pico PLUS or SuperSignal West Femto Maximum Sensitivity Substrate (Life Technologies Corporation, Grand Island, NY).

FOXO3 gene overexpression and CRISPR-based knockdown constructs in patient-derived ATLL cell lines. ATLL-84c and ATLL-97c are clonally-proven ATLL cell lines derived from tumor cells carrying the typical CD4+CD25+ ATLL phenotype (established at Ramos lab)[71]. Lenti pseudo viral particles were packaged with psPAX2 (Addgene plasmid #12260) and pMD2.G (Addgene plasmid #12259). mCAT-1 ATLL lines for retroviral transductions were generated by transduction of Lenti pseudo-viral particles prepared with mCAT-1 ORF in a modified pCDH-CuO-MCS-EF1α +Puro plasmid (System Biosciences). *FOXO3* CRISPR knockdown ATLL cell lines were established by the transduction of pseudo viral particles from all-in-one vector pLentiCRISPR v2 -Foxo3 sgRNA target 5'-GACAGAGTGAGCCGTTTGTC-3' or 5'-AGAGAGGCGCATCATCGTCC-3' (GenScript) and control non-silencing sgRNA sequence 5'-GTATTACTGATATTGGTGGG-3' (BRDN0001149198). shRNA-inducible ATLL constructs were established with pseudo viral particles generated from Tet-pLKO-puro (Addgene plasmid# 21915) targeting 5'-GCTCTTGGTGGATCATCAA-3'. Retroviral pseudo particles for over-expression of *FOXO3* were generated with pMSCV-U6sgRNA(BbsI)-PGKpuro2ABFP (Addgene plasmid #

102796) where the blue fluorescent protein (BFP) ORF was replaced with *FOXO3* ORF and packaged with plasmids pHIT60 and pHIT123 (kindly provided by Müschen Lab.) *FOXO3* R177W and D199N mutated constructs were genera rated in ATLL-84c cells using lentiviral transduction. The nucleotide composition of these vectors was verified by DNA PCR and sequencing.

References

1. 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):428-37.
2. Takatsuki K. Adult T-cell Leukaemia. Oxford: Oxford University Press; 1994.
3. Cook LB, Fuji S, Hermine O, Bazarbachi A, Ramos JC, Ratner L, et al. Revised Adult T-Cell Leukemia-Lymphoma International Consensus Meeting Report. J Clin Oncol. 2019;37(8):677-87.
4. Romanel A, Zhang T, Elemento O, Demichelis F. EthSEQ: ethnicity annotation from whole exome sequencing data. Bioinformatics. 2017;33(15):2402-4.
5. Daniels J, Doukas PG, Escala MEM, Ringbloom KG, Shih DJH, Yang J, et al. Cellular origins and genetic landscape of cutaneous gamma delta T cell lymphomas. Nat Commun. 2020;11(1):1806.
6. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013;31(3):213-9.
7. Diossy M, Sztupinszki Z, Krzystanek M, Borcsok J, Eklund AC, Csabai I, et al. Strand Orientation Bias Detector to determine the probability of FFPE sequencing artifacts. Brief Bioinform. 2021;22(6).
8. Amemiya HM, Kundaje A, Boyle AP. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Sci Rep. 2019;9(1):9354.
9. Glusman G, Caballero J, Mauldin DE, Hood L, Roach JC. Kaviar: an accessible system for testing SNV novelty. Bioinformatics. 2011;27(22):3216-7.
10. Karczewski KJ, Weisburd B, Thomas B, Solomonson M, Ruderfer DM, Kavanagh D, et al. The ExAC browser: displaying reference data information from over 60 000 exomes. Nucleic Acids Res. 2017;45(D1):D840-D5.
11. Maffucci P, Bigio B, Rapaport F, Cobat A, Borghesi A, Lopez M, et al. Blacklisting variants common in private cohorts but not in public databases optimizes human exome analysis. Proc Natl Acad Sci U S A. 2019;116(3):950-9.
12. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29(1):308-11.
13. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.

14. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
15. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. 2019;47(D1):D941-D7.
16. Park J, Daniels J, Wartewig T, Ringbloom KG, Martinez-Escala ME, Choi S, et al. Integrated Genomic Analyses of Cutaneous T Cell Lymphomas Reveal the Molecular Bases for Disease Heterogeneity. Blood. 2021.
17. Park J, Yang J, Wenzel AT, Ramachandran A, Lee WJ, Daniels JC, et al. Genomic analysis of 220 CTCLs identifies a novel recurrent gain-of-function alteration in RLTPR (p.Q575E). Blood. 2017;130(12):1430-40.
18. Kataoka K, Nagata Y, Kitanaka A, Shiraishi Y, Shimamura T, Yasunaga J, et al. Integrated molecular analysis of adult T cell leukemia/lymphoma. Nat Genet. 2015;47(11):1304-15.
19. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. 2011;12(4):R41.
20. Boons E, Nogueira TC, Dierckx T, Menezes SM, Jacquemyn M, Tamir S, et al. XPO1 inhibitors represent a novel therapeutic option in Adult T-cell Leukemia, triggering p53-mediated caspase-dependent apoptosis. Blood Cancer J. 2021;11(2):27.
21. Kogure Y, Kameda T, Koya J, Yoshimitsu M, Nosaka K, Yasunaga JI, et al. Whole-genome landscape of adult T-cell leukemia/lymphoma. Blood. 2021.
22. Choi J, Goh G, Walradt T, Hong BS, Bunick CG, Chen K, et al. Genomic landscape of cutaneous T cell lymphoma. Nat Genet. 2015;47(9):1011-9.
23. Pinzaru AM, Hom RA, Beal A, Phillips AF, Ni E, Cardozo T, et al. Telomere Replication Stress Induced by POT1 Inactivation Accelerates Tumorigenesis. Cell Rep. 2016;15(10):2170-84.
24. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
25. Jacobsen SB, Tfelt-Hansen J, Smerup MH, Andersen JD, Morling N. Comparison of whole transcriptome sequencing of fresh, frozen, and formalin-fixed, paraffin-embedded cardiac tissue. PLoS One. 2023;18(3):e0283159.
26. Hultquist JF, Hiatt J, Schumann K, McGregor MJ, Roth TL, Haas P, et al. CRISPR-Cas9 genome engineering of primary CD4(+) T cells for the interrogation of HIV-host factor interactions. Nat Protoc. 2019;14(1):1-27.
27. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nature Biotechnology. 2016;34(2):184-91.