haematologica
Journal of the Ferrata Storti Foundation

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

by Caroline S. Myers, Eli Williams, Carlos Barrionuevo Cornejo, Georgios Pongas, Ngoc L. Toomey, Jose A. Sanches, Maxime Battistella, Samuel Mo, Melissa Pulitzer, Cristopher A. Moskaluk, Govind Bhagat, Kenneth Ofori, Jonathan J. Davick, Octavio Servitje, Denis Miyashiro, Fina Climent, Kimberley Ringbloom, Daniela Duenas, Calvin Law, Sandro Casavilca Zambrano, Luis Malpica, Brady E. Beltran, Denisse Castro, Luciana Barreto, Carlos Brites, Jennifer R. Chapman, Jaehyuk Choi, Alejandro A. Gru, and Juan C. Ramos

Received: Feb 8, 2024.
Accepted: July 29, 2024.

Citation: Caroline S. Myers, Eli Williams, Carlos Barrionuevo Cornejo, Georgios Pongas, Ngoc L. Toomey, Jose A. Sanches, Maxime Battistella, Samuel Mo, Melissa Pulitzer, Cristopher A. Moskaluk, Govind Bhagat, Kenneth Ofori, Jonathan J. Davick, Octavio Servitje, Denis Miyashiro, Fina Climent, Kimberley Ringbloom, Daniela Duenas, Calvin Law, Sandro Casavilca Zambrano, Luis Malpica, Brady E. Beltran, Denisse Castro, Luciana Barreto, Carlos Brites, Jennifer R. Chapman, Jaehyuk Choi, Alejandro A. Gru, and Juan C. Ramos. Distinctive genomic features of human T-lymphotropic virus type 1-related adult T-cell leukemia-lymphoma in Western populations. Haematologica. 2024 Aug 8. doi: 10.3324/haematol.2024.285233 [Epub ahead of print]

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[2#], Juan C. Ramos[4#]

1. Department of Dermatology, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

2. University of Virginia School of Medicine, Charlottesville, Virginia, USA

3. Instituto Nacional de Enfermedades Neoplasicas, Lima, Peru

4. University of Miami, Miami, Florida, USA

5. Universidade de São Paolo, São Paolo, Brazil

6. Université de Paris, Paris, France

7. Memorial Sloan-Kettering Cancer Center, New York, New York, USA

8. New York-Presbyterian Hospital, New York, New York, USA

9. Columbia University School of Medicine, New York, New York, USA

10. University of Iowa, Iowa City, Iowa, USA

11. Hospital Universitari de Bellvitge, Barcelona, Spain

12. University of Texas MD Anderson Cancer Center, Houston, Texas, 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

Author contributions

Authors designated # are senior authors with equal contributions in alphabetical order. A.A.G, J.C, and J.C.R. designed and supervised the study, developed methodology, performed experiments, analyzed and interpreted data, and prepared the manuscript. C.S.M. developed methodology, analyzed and interpreted genomic and clinical data, and prepared the manuscript. E.W. performed Oncoscan CNV analysis and interpreted data. K.R. and S.M. analyzed and interpreted WES data. GP, N.L.T designed CRISPR constructs for FOXO3. N.L.T., L.B., provided technical support, processed tumor samples, performed standard gene sequencing, immunoblotting, and interpreted data. C.L. performed and interpreted targeted CRISPR screens. A.A.G, C.C.B, G.B., J.D., S.C., C.S.M., M.P.P, K.O. and J.R.C performed immunohistochemistry and immunophenotyping studies and analyzed and interpreted data. C.C.B., M.B., D.M., S.C., J.A.S., B.B., D.C., L.M., C.B., O.S., F.C and J.C.R. collected clinical data and developed ATLL patient databases at local institutions. All authors reviewed the manuscript.

Running head

Distinctive Genomic Features of Western Hemisphere ATL

Financial Support

J.C. was supported in part by the National Institutes of Health (R01CA260064-01A1) and the Leukemia and Lymphoma Society grant 1377-21. J.C.R was supported in part by the National Institutes of Health/National Cancer Institute (R01CA223232) and the University of Miami-

Sylvester Comprehensive Cancer Center (P30CA240139). C.S.M. was supported by the Monticello College Foundation Olin Fellowship.

Corresponding Author:

Alejandro A. Gru

622 West 168th Street

VC-15-207New York, NY 10032

aag2222@cumc.columbia.edu

Acknowledgements

We acknowledge the patients who consented to inclusion in this study, the University of Virginia Comprehensive Cancer Center, the University of Miami Sylvester Comprehensive Cancer Center and its Flow Cytometry Shared Resource and data analysis by Patricia Guevara, and the Northwestern University Flow Cytometry Core and Research Computing Services for their invaluable contributions. J.C. was supported in part by the National Institutes of Health (R01CA260064-01A1) and the Leukemia and Lymphoma Society grant 1377-21. J.C.R was supported in by the National Institutes of Health/National Cancer Institute (R01CA223232) and

the University of Miami-Sylvester Comprehensive Cancer Center (P30CA240139). C.S.M. was supported by the Monticello College Foundation Olin Fellowship. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability Statement

Demographic data, mutational counts and CNV peaks may be found in a data supplement available with the online version of this article. For original data, please contact Dr. Jaehyuk Choi at jaehyuk.choi@northwestern.edu

Disclosures

No conflicts of interest to disclose

Abstract

Adult T-cell leukemia-lymphoma (ATLL) is an aggressive Human T-cell Leukemia Virus Type 1 (HTLV-1)-driven malignancy. Although Western hemisphere (Afro-Caribbean and South American) patients face worse prognoses, our understanding of ATLL molecular drivers derives mostly from Japanese studies. We performed multi-omic analyses to elucidate the genomic landscape of ATLL in Western cohorts. Recurrent deletion and/or damaging mutations involving *FOXO3, ANKRD11, DGKZ,* and *PTPN6* implicate these genes as potential tumor suppressors. RNA-seq, published functional data and *in vitro* assays support the roles of *ANKRD11* and *FOXO3* as regulators of T-cell proliferation and apoptosis in ATLL, respectively. Survival data suggest *ANKRD11* mutation may confer a worse prognosis. Japanese and Western cohorts, in addition to acute and lymphomatous subtypes, demonstrated distinct molecular patterns. *GATA3* deletion was associated with unfavorable chronic cases. *IRF4* and *CARD11* mutations frequently emerged in relapses after interferon therapy. Our findings reveal novel putative ATLL driver genes and clinically relevant differences between Japanese and Western ATLL patients.

Introduction

Adult T-Cell Leukemia/Lymphoma (ATLL) is an aggressive hematologic malignancy caused by the human T-cell leukemia virus type I (HTLV-1), which is endemic in South America, the Caribbean, western Africa, and southern Japan.[1, 2] Clinically, ATLL is often characterized by lymphadenopathy with or without lymphocytosis, organomegaly, multi-organ involvement (more commonly skin), and immunosuppression.[3, 4] It can be classified into at least four clinical subtypes. Acute and lymphomatous are by far the most common and lethal variants, while chronic and smoldering forms tend to behave indolently until they ultimately progress to more aggressive subtypes. A chronic variant with unfavorable features ("unfavorable chronic") presents with lymphocytosis and elevated LDH, has a worse prognosis, and progresses to acute subtypes in a shorter time period.[5]

The prognosis of ATLL is dismal. The 4-year survival rates for lymphomatous and acute forms are less than 20%, with median survival less than 11 months.[6, 7] First-line treatment options include multi-agent chemotherapy, biologics such as the anti-CCR4 antibody mogamulizumab,[8] zidovudine (AZT), and interferon-alfa (IFN).[4] However, disease relapse occurs in nearly all patients, and even those who undergo allogeneic stem cell transplantation have median survival of less than 6 months.[6]

Afro-Caribbean ATLL patients present with distinct, more severe clinical features than their Japanese counterparts in published cohorts.[9, 10] These include a younger age at diagnosis by >10 years and a worse overall survival. Despite the clinical impact of these discrepancies, the molecular features of ATLL have not been well characterized among patients in the Western hemisphere. Comprehensive studies of ATLL using genome-wide approaches have come primarily from Japan.[11-13] Western hemisphere studies have been limited in number, size, and breadth.[14, 15] These limitations are due in part to the broad geographic spread of ATLL in the Western hemisphere, encompassing HTLV-1 endemic areas that have not traditionally

participated in research programs.[16] More generally, these features reflect a widespread underrepresentation of Afro-Caribbean and Hispanic patients in genomic research.[16, 17] In the International Cancer Genome Consortium, less than 1% of donors hail from the Caribbean or South American countries.[18] The National Cancer Institute's Genomic Data Commons includes >10% of donors who report African ancestry and >5% who report Hispanic ethnicity.[19] Underrepresentation of these populations in genomic research affects both patients and researchers: it impedes patient inclusion in the growing benefits of personalized medicine and withholds valuable information from our collective knowledge of cancer biology.

Our group sought to overcome these challenges by assembling the largest cohort of Western hemisphere ("Western") ATLL patients to date. We acquired samples from underserved and indigenous populations in South America, the Caribbean, and immigrant communities in the United States. We undertook genome-wide characterization of these patients' molecular features to gain an unparalleled view into the mechanistic basis of ATLL in Western patients. Our multimodal genomic study employed whole-exome sequencing (WES), copy number variation (CNV) data and RNA-seq corroboration to seek novel driver genes. We then looked for molecular drivers of clinical phenotypes including variation in geographic regions, clinical subtypes, and response to therapies.

Our approach is the first to use exome-wide analysis to identify population-based differences between the molecular landscape of Japanese and Western ATLL. Furthermore, the inclusion of underrepresented populations in our genome-wide analysis uncovered novel driver gene candidates that affected apoptosis and T-cell proliferation *in vitro*. Finally, we examined the relationships between molecular features and clinical outcomes in ATLL patients and newly elucidated distinct molecular features characterizing acute and lymphomatous subtypes. Together, these foundational analyses illustrate global patterns of ATLL molecular features. They also yield novel genetic perspective on Western ATLL, thus providing a basis for future pre-clinical and clinical investigation.

Methods

Patient Characteristics and Sample Collection. 165 patients with a confirmed diagnosis of ATLL were included in this study after quality control (Supplemental Table S9). Twelve patients contributed both pre- and post-relapse samples to this study. Specimens for molecular study were obtained from the blood of patients with leukemic presentation or from formalin-fixed, paraffin-embedded (FFPE) or frozen tissue of patients who presented with solid tumors. Immunohistochemistry was performed on representative sections of 32 patient samples, with scoring performed by two independent blinded dermatopathologists. 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 diagnosis criteria, subtype classification, and methodologies for immunohistochemistry, nucleic acid extraction and sequencing can be found in Supplemental methods.

Statistical Analysis. Where possible, statistical analyses used have been indicated in the text. Survival data was analyzed using the R package Survival.[20] Mutual exclusivity analyses used the R Package Discover.[21] Numerical values (e.g. log fold change, RNA-seq counts) were compared between two groups using a student's two-tailed T-test. Frequencies of mutations or clinical outcomes between groups were compared using Fisher's exact test. Specific methodologies used for mutational, CNV and RNA-seq analyses can be found in the Supplemental Methods.

T-cell proliferation assay pipeline. Human T-cells were isolated from enriched leukapheresis products and transfected with Caspase 9-sgRNA ribonucleoprotein (crRNP) complex as previously described (Supplemental Methods). After CRISPR knockdown, cells were stimulated with plate-bound anti-human CD3/CD28 and IL-2 as described in the Supplemental Methods. On day 13, cells were stained with CFSE (Supplemental Methods) and cultured for 4 days in

complete RPMI media without IL-2 and with or without stimulation. After 4 days, cell proliferation was assessed by Fluorescent-activated Cell Sorting (FACS). Gating strategy, including positive controls, is illustrated in Supplemental Fig. S6

FOXO3 gene overexpression and 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). ATLL lines were transduced as described in the Supplemental Methods. Protein levels were verified by Western blot, and the nucleotide composition of *FOXO3* mutant vectors was verified by DNA PCR and sequencing. Transduced cells were exposed to etoposide, belinostat and/or DMSO vehicle. Annexin V staining as an indicator of apoptosis was then evaluated with flow cytometry (Supplemental Methods).

Results

Multimodal analysis of the ATLL genomic landscape.

We compiled a multimodal dataset from 165 Western ATLL patients (Fig. S1A). ATLL diagnoses were confirmed by histopathologic findings, HTLV-1 serologic assays, clonal T-cell populations as determined by immunophenotyping and gene rearrangement studies, and HTLV-1 PCR validation in skin biopsy cases of limited quantities. The geographic coverage of this cohort was broad: countries-of-origin were largely South American (Brazil, Peru, Panama, and Ecuador; N = 76) and the Caribbean (Haiti, Jamaica, Trinidad, Dominican Republic, Bahamas, Antigua, Martinique, St. Vincent, Tortola, West Indies and US Virgin Islands; N = 76). We also included patients of reported African descent from France (N = 6) and Miami, Florida, USA (N = 7). For subsequent data analysis, patient ethnicity was categorized based upon single nucleotide polymorphisms (SNPs) using EthSeq[22] (Table 1).

Data modalities used in this study included Oncoscan Copy Number Variation (CNV), whole-exome sequencing (WES) and RNA-sequencing (RNA-seq). We combined these data

with WES and CNV data from 83 and 426 Japanese patients, respectively.[12] Consistent with the literature,[9] Western patients presented with disease at a younger age (51 vs. 65, $p = 6.62 \times 10^{-14}$, student's two-tailed t-test) and with a greater proportion of aggressive (acute and lymphomatous) subtypes (79% vs 63%, $p = 1.8 \times 10^{-5}$, Fisher's exact test).

We first sought to identify novel genomic drivers in our Western cohort using previously validated methods.[23-26] Our methods included multiple orthogonal analyses. First, we analyzed WES data from 122 patients. We approached mutation calling conservatively, employing previously validated approaches to remove possible sources of sequencing or alignment errors or ambiguity[23-26] (Supplemental Methods; Supplemental Figs. S1 - 2). To further validate these data against independent datasets as well as to identify low-prevalence driver mutations shared with other cancers, we integrated our data with Japanese ATLL data (83 patients),[12] a cohort of published T-cell lymphoma data (699 patients),[24, 25] and publicly available libraries of mutation data across cancer types (>1.4 million tumor samples).[27]

The distribution of mutations within cancer driver genes is not random; we have leveraged these patterns to identify novel cancer-promoting mutations in other T-cell lymphomas.[23-26, 28] Thus, we prioritized mutations that demonstrated patterns characteristic of oncogenes and tumor suppressors (Supplemental Fig. S2). To identify putative oncogenes, we looked for characteristic recurrent "gain-of-function" mutations at amino acid hotspots. Twenty-seven genes harbored recurrent non-synonymous amino acid alterations (Supplemental Tables S1, S2). Tumor suppressors generally contain recurrent damaging mutations, i.e. mutations having a high likelihood of inducing loss-of-function (e.g. stop-gain, start-loss, or splice-site mutations). Thirteen genes had a statistically significant burden of damaging mutations (Supplemental Table S3).

Next, we analyzed exome-wide CNV data from 128 Western ATLL patients in combination with previously published data from 426 Japanese patients.[12, 13] We chose Oncoscan, a clinically utilized and industry-validated commercial CNV assay, for analysis of our

Western datasets.[29] In the combined Japanese and Western dataset, we identified 17 and 35 chromosomal regions subject to statistically significant rates of duplication or deletion, respectively, by methods previously described[24, 25, 30] (Fig. 1A). Identifying potential gene targets of CNVs can be difficult, in part because these CNVs are by nature polygenic. To identify putative tumor suppressors and oncogenes in these recurrent CNV regions, we employed a previously validated stepwise hierarchical algorithm[23-26] (Supplemental Fig. S3). This analysis identified a putative target gene in 39 chromosomal regions (Supplemental Tables S4-5).

In total, we identified 64 putative ATLL driver genes distributed across 15 biologically relevant pathways (Fig. 2A) (Supplemental Table S6). 13 genes were significant by both orthogonal WES and CNV analyses (*GATA3, CCR7, TP53, ARID2, CSNK2B, NOTCH1, CBLB, CD58, ANKRD11, IRF4, CARD11, CTNNB1, CD28)* (Fig. 1B-E). Twenty-five genes were implicated by point mutation analysis alone (Supplemental Tables S1-3) and 26 by copy number mutation analysis alone (Supplemental Tables S4-5). Our results confirmed 42 driver genes reported in Japanese ATLL populations.[12, 13, 31, 32]. Twenty-two genes were newly implicated in ATLL by this analysis (Supplemental Table S6). Four novel putative driver genes were mutated at frequencies of 10% or greater (*FOXO3, APC, WNK1, ANKRD11*). Collectively, mutations in the 22 novel putative driver genes are found in 80% of samples.

Of the 22 genes newly implicated in ATLL, 14 have mutations that have been functionally validated as driver genes in cell or animal models of other cancer types. For an additional 3 genes, we identified new cancer-associated hotspots: in *WNK1* (p.Ala372)*, RBBP4* (p.Arg131) and *FOXO3* (p.Asp199) (Fig. 1G; Supplemental Fig. S4A-B). Five genes were not previously recognized as recurrently mutated in cancer (*CD3E, ANKRD11, DGKZ, PTPN6, RHOB*) (Fig. 1B-F; Supplemental Fig. S4C). Collectively, 44% of samples have at least one mutation in a gene not previously reported in cancer.

The pathways most commonly affected by driver gene mutations include CD28/PI3K-AKT signaling (*CD28, VAV1, PLCG1, PRKCQ, FOXO3,* and negative regulators *CBLB,*

INPP4B, PTPN6), T-Cell Receptor (TCR)/NF-κB signaling (*CD3E, PLCG1, VAV1, PRKCQ, PRKCB, PTPRN2, CARD11, IKBKB, RLTPR, CSNK2B*, *IRF4*, and negative regulators *CBLB, PTPN6*, *DGKZ, NFKBIA, TNFAIP3, TRAF3)*, and cell migration (*CCR7, CCR4, RHOA, RHOB, GPR183, NRXN3, VAV1, WNK1)*. The TCR and PI3K mediated CD28 co-receptor pathways intersect at the activation of PKCθ (*PRKCQ*) via PCLγ1 (*PLCG1)* and PDK1, respectively, which connect proximal TCR and CD28 co-receptor signaling events, ultimately leading to NF-κB activation (Fig. 2B).

Genomic and functional validation of novel putative driver genes.

Our mutational analyses allowed us to identify both known and novel putative driver genes; however, their biological roles in ATLL are not implicit. For example, despite its recurrent damaging mutations (often characteristic of tumor suppressors[28]), *CCR4* has been shown to be an oncogene with functionally validated gain-of-function truncated variants.[33] For this reason, we classified putative driver genes as suspected oncogenes or tumor suppressors based upon their pattern of point mutations leveraged against CNV mutational patterns and published functional studies (Supplemental Table S6). We then sought to validate the functional consequences of these putative oncogenes or tumor suppressors *in vitro*.

To do this, we first utilized previously published, publicly available genome-wide CRISPR screens for T-cell activation. These included two CRISPR interference screens for genes mediating TCR-independent cytokine production, two amplification screens for TCR-independent cytokine production, and an interference screen for TCR-dependent cell proliferation.[34, 35] We predicted that single-guide RNAs (sgRNAs) for oncogenes should show patterns of alteration in CRISPR screens consistent with the promotion of T-cell proliferation or cytokine production. sgRNAs for tumor suppressors should show the opposite patterns.

97% (N = 62) of putative driver genes showed a pattern of sgRNA alteration in one or more CRISPR screens consistent with their predicted tumor suppressor/oncogene roles. The only exceptions were *TP53* and *PDCD1*. Presumably the effects of these genes cannot be easily modeled in short-term cultures (*TP53*) or without ligands (*PDCD1*). 86% (N = 55) displayed the predicted patterns across multiple orthogonal CRISPR screens. 39% (N = 25) showed the predicted sgRNA up- or down-regulation at magnitudes significantly greater than chance (False Discovery Rate, FDR < 0.05) (Fig. 3A-B, S5A).

We next examined novel putative driver genes individually, including novel putative driver genes *ANKRD11*, *DGKZ*, *PTPN6* and *CD3E*, and novel hotspots in cancer-associated gene *FOXO3.* We examined the function of each of these genes in the previously published CRISPR screens described above and performed orthogonal validation assays where possible.

Implicated by both point mutation and copy number mutation analysis, *ANKRD11* encodes a chromatin scaffolding protein binding histone deacetylases involved in the differentiation of neural cells. It has also been postulated to interact with p53 in breast cancer models.[36, 37] Its role in T-cells is unknown. In the Western dataset, 2.5% of samples were affected by recurrent *ANKRD11* damaging mutations (p.Arg838*,* p.Arg1007*,* p.Ser2208*; $p = 4.31 \times 10^{-4}$) (Supplemental Table S3). By orthogonal CNV analysis, we found that *ANKRD11* falls within a 2.5 MB region of significant deletion on chromosome 16 ($q = 2.4 \times 10^{-13}$, GISTIC2.0) deleted in 15% of patients (N = 84) (Fig. 1A-C; Supplemental Table S5). Among all the genes in this region, *ANKRD11* has an outsized proportion (likelihood ratio > 5) of gene-localizing mutations, suggesting it is the target gene of this recurrently deleted chromosomal segment. In 3.5% (N = 6) of samples with both CNV and WES data, both *ANKRD11* alleles were mutated either via biallelic deletion or mutation plus loss of heterozygosity. By RNA-seq, samples with *ANKRD11* deletions had significantly decreased expression of *ANKRD11* transcripts (p = 0.004, student's two-tailed T-test) (Fig. 3C). Finally, we decided to examine the role of *ANKRD11 in vitro.* We performed CRISPR knockout of *ANKRD11* in human T-cells.

Carboxyfluorescein succinimidyl ester (CFSE) proliferation assays showed significant increases in cell division in *ANKRD11* knockout cells upon TCR stimulation ($p = 0.0045$; student's two-tailed T-test) (Fig. 3D-E).

DGKZ is a protein kinase responsible for dampening the effects of TCR stimulation by catalyzing breakdown of the downstream signaling molecule phosphatidic acid.[38] Consistent with its putative tumor suppressor function, *DGKZ* significantly inhibited TCR-mediated proliferation in previously published CRISPR screens (FDR = 5.5×10^{-4}). In tumor samples from Western ATLL patients, we saw reduced or absent DGKZ protein in 10/10 (100%) of patients regardless of *DGKZ* mutation status, suggesting that it could be transcriptionally downregulated in ATLL (Supplemental Fig. S5B).

PTPN6 and *CD3E* are two novel putative driver genes implicated by recurrent point mutations. *CD3E* encodes a subunit of the CD3/TCR complex.[39] Samples in Western, Japanese, and publicly available TCL cohorts were recurrently mutated at p.Ser41Cys (N = 1 Western, 1 Japanese, 1 TCL). Additional samples contained mutations at nearby p.Ser39Cys (N = 1 Western, 1 Japanese, 1 TCL) (Fig. 1F). In genome-wide CRISPR screens, *CD3E* was a significant positive regulator of TCR-mediated cell proliferation (FDR = 3.5×10^{-5}) (Fig. 3B). *PTPN6* encodes for the protein SHP-1, a negative regulator of T-cell activation and PI3K signaling.[40] 2.5% of samples (N = 3) contained damaging mutations in *PTPN6.* By genome-wide CRISPR screen, *PTPN6* was a significant inhibitor of TCR-independent IL-2 production (FDR = 7.2×10^{-5}). By western blot (WB) analysis, 50% of Western ATLL samples showed no significant PTPN6 expression. (Supplemental Fig. S5C).

We noted recurrent mutations at a novel hotspot in cancer-associated gene *FOXO3* (Fig. 1G) (Supplemental Table S2). *FOXO3* encodes a transcription factor regulating T-cell differentiation.[41] It has been functionally validated as a tumor suppressor in several solid-organ malignancies.[42] Mechanistic studies have shown that HTLV-1 viral proteins HBZ and Tax suppress FOXO3 protein function;[43, 44] however, *FOXO3* genetic mutations have not been

directly implicated in ATLL by published genomic studies. 10% of Western cohort patients had *FOXO3* hotspot mutations (Supplemental Tables S1-2). While 31% of these samples contained mutations at cancer hotspot p.Asp199Asn, 54% (N = 7) of Western *FOXO3* mutations occurred in a previously unidentified hotspot, p.Arg177Trp. Both hotspots fall within the *FOXO3* DNA-binding forkhead domain in proximity to the validated dominant negative variant p.Ser256Ala, suggesting they could act as dominant negative mutations (Fig. 1G). By orthogonal CNV analysis, *FOXO3* also falls within a broad region (51MB) of significant deletion ($q = 3.35 \times 10^{-40}$, GISTIC2.0). It is deleted in 9.4% (N = 52) of samples. By WB analysis of a random selection of ATLL patients, FOXO3 protein expression was reduced or absent in 11/13 patients (85%) (Supplemental Fig. S5C-D), including 5/6 (83%) of patients with deletions or mutations. Weighted gene correlation analysis[45] of RNA-seq data suggests that *HBZ* expression relates to expression of *FOXO3* as well as *ANKRD11* (Fig. S5E-F). Consistent with prior large-scale studies on ATLL tumors, we did not find significant expression of *Tax* in the samples analyzed.[12]

We performed CRISPR and shRNA-mediated knockdown of *FOXO3* in our patient-derived ATLL cell lines. We subjected these cells to anti-neoplastic chemotherapy (etoposide) and biological (belinostat) agents used to treat T-cell lymphomas. Knockdown of Foxo3 protein expression in ATLL cell lines by *FOXO3*-specific sgRNAs (sgRNA1 and sgRNA2) conferred resistance to drug-induced apoptosis, as compared to cells transfected with scrambled-(SCR)-sgRNA (Fig. 4A). Similarly, knockdown of *FOXO3* via tetracycline-inducible shRNA protected ATLL cells from dose dependent drug-induced apoptosis (Fig. 4C). In contrast, mCAT-1+ ATLL cells transduced with *FOXO3*-overexpressing pseudoviral particles showed higher rates of drug-induced apoptosis in comparison to cells transduced with empty controls (Fig. 4B). Finally, we examined cells transfected with mutant *FOXO3* p.Asp199Asn and p.Arg177Trp-expressing lentiviral vectors. These cells resembled *FOXO3* knockdown in their phenotype: compared to cells transfected with empty vectors, they showed resistance to drug-induced apoptosis (Fig. 4D). Because the p.Arg177Trp and p.Asp199Asn mutations both map to the Foxo3a DNA-

binding domain, we hypothesized that the effect of *FOXO3* mutation could affect T-cell apoptosis by affecting transcription of pro-apoptotic target genes. However, knockdown of Foxo3a did not appear to affect protein levels of BIM and p21, both encoded by known pro-apoptotic Foxo3a target genes (*BCL2L11* and *CDKN1A,* respectively)(Supplemental Fig. S6B). Collectively, these data support the pro-apoptotic role of *FOXO3* in ATLL, although the mechanisms by which it exerts these effects remains yet to be identified.

The ATLL genomic landscape differs among Western and Japanese patients.

Previous exome-wide studies of ATLL have been conducted in Japanese populations.[11, 12] We sought to determine whether there exist molecular differences between Western and Japanese cohorts that could be associated with the known differences in clinical presentation.[9, 10] We observed broad similarities in commonly mutated genes, but several were mutated at significantly different frequencies. For example, Japanese patients had a significantly greater burden of point mutations in *CCR4,* the most commonly mutated putative driver gene overall (34% versus 16% of patients in the Western populations, p = 0.006, Fisher's exact test) (Fig. 5). *PTPRN2* (34% vs 13%, p = 0.002) and *TRRAP* (43% vs 14%, $p = 3.2 \times 10^{-5}$) were also both more commonly mutated in Japanese than Western samples (Fig. 5C) (Fisher's exact test).

Several putative driver genes were deleted or mutated with significantly greater frequency in the Western cohort. These include *INPP4B* (21% vs 6%, p = 0.006), *ANKRD11* (34% vs 11%, $p = 4.10 \times 10^{-4}$) and *CBLB* (27% vs 13%, p = 0.035) (Fig. 5D)(Fisher's exact test). Of Western patients, 8/13 (62%) had clearly reduced CBLB expression; 50% of these had mutation or deletions (Supplemental Fig. S5C-D). *FOXO3* point mutations were eight times more common in Western patients (10% versus 1.2%, p = 0.017, Fisher's exact test) (Fig. 5A). Mutations at the novel hotspot *FOXO3* p.Arg177Trp occurred only in Western patients of African descent with aggressive leukemic ATLL subtypes (acute N=7, unfavorable chronic N = 1).

Different ATLL clinical subtypes are characterized by distinct mutational patterns.

We next sought to identify molecular drivers of clinical behavior in different ATLL subtypes. Consistent with observations from the Japanese cohort, we identified in the Western cohort enrichment of *STAT3* mutation in "indolent" ATLLs (chronic and smoldering) (37% vs 11%, $p = 1.2x10^{-4}$), enrichment of total number of mutations in "aggressive" ATLLs (acute and lymphomatous) (119 vs 47 mutations, $p = 4.3x10^{-7}$), and enrichment of *TP53, CDKN2A* and *IRF4* mutations in aggressive ATLLs (31% vs 12%, p = 0.01; 40% vs 16%, p = 0.002; 38% vs 18%, p = 0.02, respectively) (Fisher's exact test).[11]

The enrichment of aggressive clinical subtypes in our Western dataset allowed us to distinguish molecular features of acute versus lymphomatous cases in the Western cohort. *NRXN3* and *CCR4* were commonly mutated (>30% of samples) in both subtypes. *CDKN2A* mutations (46% vs 25%, p = 0.009) and *PLCG1* amplification/mutation (37% vs 14%, p = 0.002) were mutated significantly more often in acute cases compared to lymphomatous cases. In contrast, *TP53, WWOX, CD3E, TBL1XR1* and *NFKBIA* were genetically altered significantly more often in lymphomatous cases (Supplemental Table S7). Consistent with their enrichment in distinct disease subtypes, *TP53* and *PLCG1* mutations showed significant mutual exclusivity (q = 0.004).[21]

There is little molecular information available about chronic ATLL subtype with unfavorable features.[9] In our dataset, unfavorable chronic cases (N=17) resembled aggressive cases in their mutational and CNV burden (Fig. 6A-B). They were characterized by an increased frequency of heterozygous *GATA3* deletions. Of the 11 unfavorable chronic cases with CNV data, 45% (N = 5) had *GATA3* deletions (3.8-fold enrichment compared to other cases, p < 0.001, Fisher's exact test). Immunohistochemistry analysis of *GATA3* expression in ATLL patient samples confirms that GATA3 protein levels are lower in unfavorable chronic cases than in other aggressive subtypes (p = 0.05, student's two-tailed T-test) (Fig. 6C).

ATLL molecular features are associated with survival and response to therapy.

In addition to clinical subtype and mutational status, we annotated our dataset with patient survival and response to chemo- and/or AZT-interferon therapy. We used a Cox multivariate regression analysis to test the effect of gene mutation on survival. We limited our analysis to patients with well-annotated clinical data who were seen for at least one follow-up visit (Supplemental Table S9). As expected, indolent subtypes were associated with significantly longer survival times than aggressive subtypes (HR 0.15, 0.26 and 0.26 for smoldering, unfavorable chronic and chronic subtypes, respectively). Treatment type (chemotherapy, zidovudine-interferon, both or neither) did not influence survival. Consistent with previous studies, *STAT3* mutation trended towards significantly decreased mortality in dataset-wide analyses. However, in analyses controlled for clinical subtype and treatment modality and corrected for multiple comparisons, only *ANKRD11* and *TP53* mutation were significantly associated with increased mortality (HR 2.70 and 2.67 respectively, Fig. 6D-E). Pairwise analysis did not show significantly different outcomes for patients with both mutations than patients with either driver mutation alone.

One mutation was associated with response to therapy in the Western cohort: Patients with *CDKN2A* loss were significantly more likely to experience a complete response (CR) to chemotherapy (24% complete response vs 0%, $p = 0.025$) (Fig. 6F). Unfortunately, this mutation was not associated with decreased mortality.

Newly acquired IRF4 and CARD11 mutations in relapsed ATLL.

Consistent with the literature, disease relapse was common in our cohort.[6] Our dataset included 10 patients who experienced disease relapse after receiving AZT-interferon treatment as initial therapy. Two of these patients experienced multiple relapses. We investigated whether the variability in mortality present within our dataset could be related to genomic patterns conferring a high likelihood of relapse.

We first examined patterns of malignant clonality in relapsed samples. As expected, relapse samples shared the same HLA genotype as initial samples, except for one patient who underwent loss of heterozygosity in HLA-A upon relapse. Comparison of transcription data-derived TCR clonotypes in initial versus relapsed samples demonstrated that all relapses harbored the same TCR as the initial clone (Fig. 7A). Interestingly, relapsed samples generally shared few point mutations with initial tumor samples but did share most CNV mutations. Patients generally accumulated additional CNV mutations with relapse (Fig. 7B). We then sought to determine if there were any specific genomic change associated with disease relapse. We found recurrent mutations in *CARD11* (3 samples) and *IRF4* (4 samples) that were acquired in disease relapse samples after AZT-interferon therapy. Although uncommon in the general cohort, *IRF4* mutations were significantly more common in the disease relapse (33% vs 10% of samples, $p = 0.038$, Fisher's Exact Test) (Fig. 7C). Of 12 samples taken after disease relapse, four had new *IRF4* mutations (3/9 patients). One patient who experienced relapse twice developed two different *IRF4* mutations in each relapse instance. Manual review of sequencing data confirmed that none of the *IRF4* mutations in relapsed samples were present even at a subclonal level in parent samples. We further analyzed 7 additional relapsed samples for p.Leu70Val or p.Lys59Arg, the most common *IRF4* mutations in our cohort, by standard Sanger sequencing (Supplemental Table S8)(Fig. 7D). P.Lys59Arg mutations were found in 2 samples from relapsed patients for whom no baseline samples were available. Within the entire cohort, no patient with an *IRF4* mutation experienced complete response to AZT-interferon therapy, suggesting it may be associated with primary or acquired resistance (Supplemental Table S9).

Discussion

Through multimodal exome-wide analysis, we have identified novel genomic features of ATLL that highlight differences between the genomic landscapes of Japanese and Western cohorts.

Newly identified putative driver genes include *FOXO3* and *ANKRD11*. *FOXO3* encodes a transcription factor that has been postulated to be a tumor suppressor and mediator of apoptosis via the PI3K pathway.[46] Our *in vitro* experiments using patient-derived ATLL cell lines demonstrated the acquisition of antineoplastic drug resistance after *FOXO3* knockdown and the augmentation of drug-induced apoptosis with *FOXO3* overexpression, suggesting that FOXO3 may have a pro-apoptotic role in ATLL. Similar to the p.Asp199Asn variant previously observed in solid malignancies,[40] expression of the novel p.Arg177Trp variant observed in our cohort also conferred resistance to antineoplastic drugs *in vitro*. *ANKRD11* is a novel driver gene identified as significant through both mutational and CNV analysis and validated through *in vitro* proliferation assays. *ANKRD11* is especially relevant due to its association with increased mortality. While Japanese studies have noted recurrent deletion of *ANKRD11* in ATLL, it was previously dismissed as breakage at a putative fragile site.[13] Our analysis was the first to distinguish *ANKRD11* and *FOXO3* as putative driver genes through mutational analysis and functional assays. Because both *ANKRD11* and *FOXO3* mutations were significantly enriched in Western patient cohorts, their elucidation as putative driver genes was only possible after the inclusion of underrepresented Western patient cohorts in our analysis.

Previous studies have demonstrated that Western ATLL patients have more severe outcomes than Japanese patients.[9, 10] While differences in medical infrastructure likely contribute to these discrepancies, health care inequalities alone appear inadequate to explain all observed clinical disparities (e.g. the earlier diagnosis of ATLL by >10 years earlier in patients from medical resource-poor regions of the Caribbean).[9] It is plausible that differences in the distribution of ATLL driver mutations may also contribute to the disparate clinical outcomes in Western vs. Japanese ATLL cases.[9, 15]

Our data supported the presence of molecular differences in ATLL that correspond to geography and clinical outcome. For example, mutations in *INPP4B* and *ANKRD11* are more common in Western patients and predict worse survival, revealing potential tumor-associated

causes of disparate clinical outcomes in the West. In contrast, *CCR4* mutations were more common in Japanese patients. Such a finding might account for the better therapeutic responses to mogamulizumab seen in Japanese patients compared to the rates seen in the US.[47]

These data also suggest new clinical targets for ATLL. The collective enrichment of *FOXO3, INPP4B, CBLB, DGKZ, PTPN6* mutations in Western ATLL patients suggest a potentially mounting role of the PI3K-AKT pathway in ATLL oncogenesis. PI3K inhibitors are FDA- and National Comprehensive Cancer Network-approved therapies for chronic lymphocytic leukemia[48] and non-Hodgkin's lymphomas.[49, 50] In recent trials, they have shown to be promising agents in PTCL patients[51] and in preclinical studies of ATLL.[52] These results support investigating further the potential of the PI3K-AKT pathway as a therapeutic target in ATLL, particularly in Western patients.

Our combined Western and Japanese dataset was the largest cohort of exome-wide sequencing data to date. Because of this, we were able to make novel observations regarding the patterns of genomic alterations in different clinical subtypes. *TP53* and *CDKN2A* alterations have previously been shown to be increased in aggressive cases (acute and lymphomatous combined). Our analysis examined acute and lymphomatous cases separately and found that *TP53* loss is more characteristic of lymphomatous cases while *CDKN2A* loss is more characteristic of acute cases. These molecular differences could account for the distinct clinical presentations of acute and lymphomatous ATLL. They might also guide treatment in the future as more targeted therapies become available for ATLL. We also identified *GATA3* deletion as a putative tumor suppressor associated with chronic ATLL with unfavorable features, as it is characterized by recurrent deletion and damaging mutation in our cohort. Consistent with this, it has been implicated in constraining regulatory T-cell proliferation.[53]

In our dataset, disease relapse after AZT-interferon therapy was associated with the acquisition of *IRF4* mutations affecting DNA binding domain. This mutational pattern could be

the result of AZT-interferon treatment-driven selection. IRF4, also known as MUM1, is a transcription factor that regulates the expression of interferon-inducible genes in leukocytes.[54] Lack of IRF4 protein expression has been previously associated with response to AZT-interferon therapy in patients with ATLL.[55]

It has become increasingly evident that different ethnic groups have disparate prognoses for the same diseases. However, our understanding of diversity in disease pathogenesis is limited by the underrepresentation of certain patient populations in genomic research. To address these challenges in the context of ATLL, we generated the largest dataset of untargeted genomic information in Western ATLL patients, including diverse African descendants from North and South America, the Caribbean, and South American indigenous populations, to date. Through the analysis of this dataset, we identified novel molecular ATLL features, some of which are associated with more aggressive disease. Several of these features, including *ANKRD11, INPP4B* and *FOXO3*, were mutated significantly more often in Western patients. Our functional assays support the roles of these genes as tumor suppressors *in vitro.* Other mutations associated with treatment outcomes (e.g. *CCR4*) were more common in Japanese cohorts. Our present findings augment current knowledge and reflect the diversity of ATLL's molecular landscape. Furthermore, they emphasize the need for the inclusion of underrepresented populations in genomic research to better understand and address disparate factors in patient outcomes.

References

1. Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC. 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. Gessain A, Cassar O. Epidemiological Aspects and World Distribution of HTLV-1 Infection. Front Microbiol. 2012;3:388.
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. Cook LB, Fuji S, Hermine O, et al. Revised Adult T-Cell Leukemia-Lymphoma International Consensus Meeting Report. J Clin Oncol. 2019;37(8):677-687.
5. Ishitsuka K, Tamura K. Human T-cell leukaemia virus type I and adult T-cell leukaemia-lymphoma. Lancet Oncol. 2014;15(11):e517-526.
6. Katsuya H, Ishitsuka K, Utsunomiya A, et al. Treatment and survival among 1594 patients with ATL. Blood. 2015;126(24):2570-2577.
7. Imaizumi Y, Iwanaga M, Nosaka K, et al. Prognosis of patients with adult T-cell leukemia/lymphoma in Japan: A nationwide hospital-based study. Cancer Sci. 2020;111(12):4567-4580.
8. Subramaniam JM, Whiteside G, McKeage K, Croxtall JC. Mogamulizumab: first global approval. Drugs. 2012;72(9):1293-1298.
9. Malpica L, Pimentel A, Reis IM, et al. Epidemiology, clinical features, and outcome of HTLV-1-related ATLL in an area of prevalence in the United States. Blood Adv. 2018;2(6):607-620.
10. Shah UA, Shah N, Qiao B, et al. Epidemiology and survival trend of adult T-cell leukemia/lymphoma in the United States. Cancer. 2020;126(3):567-574.
11. Kataoka K, Iwanaga M, Yasunaga JI, et al. Prognostic relevance of integrated genetic profiling in adult T-cell leukemia/lymphoma. Blood. 2018;131(2):215-225.
12. Kataoka K, Nagata Y, Kitanaka A, et al. Integrated molecular analysis of adult T cell leukemia/lymphoma. Nat Genet. 2015;47(11):1304-1315.
13. Kogure Y, Kameda T, Koya J, et al. Whole-genome landscape of adult T-cell leukemia/lymphoma. Blood. 2022;139(7):967-982
14. Marcais A, Lhermitte L, Artesi M, et al. Targeted deep sequencing reveals clonal and subclonal mutational signatures in Adult T-cell leukemia/lymphoma and defines an unfavorable indolent subtype. Leukemia. 2021;35(3):764-776.
15. Shah UA, Chung EY, Giricz O, et al. North American ATLL has a distinct mutational and transcriptional profile and responds to epigenetic therapies. Blood. 2018;132(14):1507-1518.
16. Alvarez-Gomez RM, De la Fuente-Hernandez MA, Herrera-Montalvo L, Hidalgo-Miranda A. Challenges of diagnostic genomics in Latin America. Curr Opin Genet Dev. 2021;66:101-109.
17. Spratt DE, Chan T, Waldron L, et al. Racial/Ethnic Disparities in Genomic Sequencing. JAMA Oncol. 2016;2(8):1070-1074.
18. Zhang J, Bajari R, Andric D, et al. The International Cancer Genome Consortium Data Portal. Nat Biotechnol. 2019;37(4):367-369.
19. Grossman RL, Heath AP, Ferretti V, et al. Toward a Shared Vision for Cancer Genomic Data. N Engl J Med. 2016;375(12):1109-1112.
20. Therneau T. A Package for Survival Analysis in R 2022 [R package version 3.3-1]. Available from: https://CRAN.R-project.org/package=survival. Accessed April 11, 2023.
21. Canisius S, Martens JW, Wessels LF. A novel independence test for somatic alterations in cancer shows that biology drives mutual exclusivity but chance explains most co-occurrence. Genome Biol. 2016;17(1):261.
22. Romanel A, Zhang T, Elemento O, Demichelis F. EthSEQ: ethnicity annotation from whole exome sequencing data. Bioinformatics. 2017;33(15):2402-2404.

23. Daniels J, Doukas PG, Escala MEM, et al. Cellular origins and genetic landscape of cutaneous gamma delta T cell lymphomas. Nat Commun. 2020;11(1):1806.
24. Park J, Daniels J, Wartewig T, et al. Integrated Genomic Analyses of Cutaneous T Cell Lymphomas Reveal the Molecular Bases for Disease Heterogeneity. Blood. 2021;138(14):1225-1236.
25. Park J, Yang J, Wenzel AT, et al. Genomic analysis of 220 CTCLs identifies a novel recurrent gain-of-function alteration in RLTPR (p.Q575E). Blood. 2017;130(12):1430-1440.
26. Choi J, Goh G, Walradt T, et al. Genomic landscape of cutaneous T cell lymphoma. Nat Genet. 2015;47(9):1011-1019.
27. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. 2019;47(D1):D941-D947.
28. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr., Kinzler KW. Cancer genome landscapes. Science. 2013;339(6127):1546-1558.
29. Foster JM, Oumie A, Togneri FS, et al. Cross-laboratory validation of the OncoScan(R) FFPE Assay, a multiplex tool for whole genome tumour profiling. BMC Med Genomics. 2015;8:5.
30. 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.
31. Boons E, Nogueira TC, Dierckx T, 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.
32. Uchida Y, Yoshimitsu M, Hachiman M, et al. RLTPR Q575E: A novel recurrent gain-of-function mutation in patients with adult T-cell leukemia/lymphoma. Eur J Haematol. 2021;106(2):221-229.
33. Nakagawa M, Schmitz R, Xiao W, et al. Gain-of-function CCR4 mutations in adult T cell leukemia/lymphoma. J Exp Med. 2014;211(13):2497-2505.
34. Schmidt R, Steinhart Z, Layeghi M, et al. CRISPR activation and interference screens decode stimulation responses in primary human T cells. Science. 2022;375(6580):eabj4008.
35. Shifrut E, Carnevale J, Tobin V, et al. Genome-wide CRISPR Screens in Primary Human T Cells Reveal Key Regulators of Immune Function. Cell. 2018;175(7):1958-1971.
36. Noll JE, Jeffery J, Al-Ejeh F, et al. Mutant p53 drives multinucleation and invasion through a process that is suppressed by ANKRD11. Oncogene. 2012;31(23):2836-2848.
37. Sirmaci A, Spiliopoulos M, Brancati F, et al. Mutations in ANKRD11 cause KBG syndrome, characterized by intellectual disability, skeletal malformations, and macrodontia. Am J Hum Genet. 2011;89(2):289-294.
38. Chen SS, Hu Z, Zhong XP. Diacylglycerol Kinases in T Cell Tolerance and Effector Function. Front Cell Dev Biol. 2016;4:130.
39. Birnbaum ME, Berry R, Hsiao YS, et al. Molecular architecture of the alphabeta T cell receptor-CD3 complex. Proc Natl Acad Sci U S A. 2014;111(49):17576-17581.
40. Plas DR, Johnson R, Pingel JT, et al. Direct regulation of ZAP-70 by SHP-1 in T cell antigen receptor signaling. Science. 1996;272(5265):1173-1176.
41. Ouyang W, Beckett O, Ma Q, Paik JH, DePinho RA, Li MO. Foxo proteins cooperatively control the differentiation of Foxp3+ regulatory T cells. Nat Immunol. 2010;11(7):618-627.
42. Liu Y, Ao X, Ding W, et al. Critical role of FOXO3a in carcinogenesis. Mol Cancer. 2018;17(1):104.
43. Olagnier D, Sze A, Bel Hadj S, et al. HTLV-1 Tax-mediated inhibition of FOXO3a activity is critical for the persistence of terminally differentiated CD4+ T cells. PLoS Pathog. 2014;10(12):e1004575.

44. Tanaka-Nakanishi A, Yasunaga J, Takai K, Matsuoka M. HTLV-1 bZIP factor suppresses apoptosis by attenuating the function of FoxO3a and altering its localization. Cancer Res. 2014;74(1):188-200.
45. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
46. Li Z, Bridges B, Olson J, Weinman SA. The interaction between acetylation and serine-574 phosphorylation regulates the apoptotic function of FOXO3. Oncogene. 2017;36(13):1887-1898.
47. Phillips AA, Fields PA, Hermine O, et al. Mogamulizumab versus investigator's choice of chemotherapy regimen in relapsed/refractory adult T-cell leukemia/lymphoma. Haematologica. 2019;104(5):993-1003.
48. NCCN. NCCN Clinical Practice Guidelines in Oncology. Chronic Lymphocytic Leukemia/Small Lymphocitic Lymphoma Version 2.2022. Available from: https://www.nccn.org/professionals/physician_gls/pdf/cll.pdf. Accessed on January 18, 2022.
49. NCCN. NCCN Clinical Practice Guidelines in Oncology. B-Cell Lymphomas 2.2022. Available from: https://www.nccn.org/professionals/physician_gls/pdf/b-cell.pdf. Accessed on March 21, 2022.
50. Richardson NC, Kasamon Y, Pazdur R, Gormley N. The saga of PI3K inhibitors in haematological malignancies: survival is the ultimate safety endpoint. Lancet Oncol. 2022;23(5):563-566.
51. Horwitz SM, Koch R, Porcu P, et al. Activity of the PI3K-delta,gamma inhibitor duvelisib in a phase 1 trial and preclinical models of T-cell lymphoma. Blood. 2018;131(8):888-898.
52. Katsuya H, Cook LBM, Rowan AG, Satou Y, Taylor GP, Bangham CRM. Phosphatidylinositol 3-kinase-delta (PI3K-delta) is a potential therapeutic target in adult T-cell leukemia-lymphoma. Biomark Res. 2018;6:24.
53. Yu F, Sharma S, Edwards J, Feigenbaum L, Zhu J. Dynamic expression of transcription factors T-bet and GATA-3 by regulatory T cells maintains immunotolerance. Nat Immunol. 2015;16(2):197-206.
54. Yamagata T, Nishida J, Tanaka S, et al. A novel interferon regulatory factor family transcription factor, ICSAT/Pip/LSIRF, that negatively regulates the activity of interferon-regulated genes. Mol Cell Biol. 1996;16(4):1283-1294.
55. Ramos JC, Ruiz P Jr., Ratner L, et al. IRF-4 and c-Rel expression in antiviral-resistant adult T-cell leukemia/lymphoma. Blood. 2007;109(7):3060-3068.

Table 1 | Demographic characteristics of patient cohorts. Patient demographics are summarized ("total Western") as well as broken down by data type. Numbers represent distinct patients, some of whom may have contributed both pre- and post-relapse samples. Japanese cohort demographics are also summarized (right two columns). Demographics of Japanese CNV data were obtained from published descriptions[12] rather than in-house review.

	TOTAL WESTERN	WES	Oncoscan	RNA	Japanese WES	Japanese CNV
N	165	122	128	92	83	426
Disease Subtype:						
Acute	84	63	70	39	39	194
Chronic with:						95
-Favorable features	9	5	7	2	26	
-Unfavorable features	17	15	11	13	0	
Lymphomatous	45	31	31	31	13	97
Smoldering	9	7	8	7	5	24
Unclassified	1	1	1			
Sex:						
Female	82	64	68	48	43	
Male	83	58	60	44	40	
Age (Avg)	51.48	51.73	51.42	52.18	65.23	
Region:					N/A	
Caribbean	76	69	61	38		
Europe	6	4	5	5		
North America	7	7	7	3		
South America	76	42	55	46		
Ethnicity:					N/A	
African	78	78	63	42		
Native South American	29	29	17	21		
South American/Asian	11	9	9	8		
European	2	2	1			
European/South American	3	3	2	1		
South Asian Islander	42	1	36	3		
Unclassified				17		
Survival Time (Weeks)	83.74	93.03	83.42	97.15	N/A	

Figure 1 | Putative ATLL driver genes discovered through mutation analysis. a, Plot of GISTIC amplification/deletion peaks labeled with their corresponding driver genes. Amplifications are depicted in red and deletions in blue. Peaks are displayed along the genome represented lengthwise. Peak heights indicate GISTIC-calculated significance (G-score). Genes newly implicated in ATLL and/or cancer are highlighted in magenta. Patients included in this analysis include both Western and Japanese cohorts (N = 554). **b, d**, **f, g** Point mutation burden among Western ATLL patients (N = 122), Japanese ATLL patients (N = 83), and published T-cell lymphoma patients (N = 800) across **b,** *ANKRD11*, **d**, *GATA3*. **f,** *CD3E* **g,** *FOXO3.* Shapes and colors indicate mutation type and cohort, respectively, as indicated in the legend. **f-g** Right panels show recurrently mutated residues in magenta color within the context of their respective protein structures; ligands are shown in yellow. **f**, CD3E is shown in dark gray. CD3-zeta and CD3-gamma are shown in light gray. Bound T-cell receptor is shown in yellow. **g**. A novel FOXO3 variant is shown with mutant tryptophan in hot pink and nearby isoleucine shown in orange. Zoomed-in panel shows predicted steric clashing between these residues as red discs, suggesting that a tryptophan mutation may be energetically unfavorable. The other subunit in the FOXO3 homodimer is shown in light gray. DNA is shown in yellow. **c**, **e**, Histograms indicating segments of overlapping deletion in **c**, *ANKRD11* and **e**, *GATA3.* Depicted cohort includes both Japanese and Western patients (N = 554). Surrounding genes are indicated at the bottom of the figure.

Figure 2 | Key pathways involving recurrent mutations identified in Western and Japanese samples. a, Frequencies and types of copy number variation (CNV) and point mutations in Western and Japanese samples. Only samples with both CNV and WES data available are portrayed in this plot (N = 168). Mutation type is indicated by square color. Bars on the bottom indicate gross number of mutations. Colored bars to the left indicate SNP-determined ethnic classification of patients and categorization of their disease subtype. Significance of GISTIC peak in which the mutation is found, if applicable, is indicated by the heatmap on top (values indicate negative logarithm of the q-value). Novel mutations are indicated by gene names bolded in blue (newly implicated in ATLL) or red (newly implicated in cancer). Genes with mutational frequencies differing between subtypes or population cohorts are indicated with a triangle and a star, respectively. **b**, An illustration of CD28- and T-cell receptor (TCR)-initiated signaling

pathways and their associated molecules based on current literature review. Cytoplasmic and nuclear compartments are separated by dashed red nuclear membrane. Downstream DNA-bound transcription factors (FOXO3a, AP-1, NFAT) and pathway- inducible genes (*BIM, FOXP3, IFNG, IL2, IRF4*) are shown. Percentage of Western ATLL samples affected by mutation (Mut), deletion (Del), or amplification (Amp) are shown for each molecule. Stars mark putative driver mutations. HBZ (HTLV-1 bZIP factor) viral protein is shown as dashed oval shape.

Figure 3 | Functional validation of putative driver genes. **a**, Violin plots showing log fold change (**LFC**) of sgRNAs (N = 19114) for putative oncogenes and putative tumor suppressors in genome-wide CRISPR interference 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). A similar analysis of CRISPR amplification screens is included in Supplemental Figure S5. **b**, Patterns of sgRNA alteration for putative oncogenes/tumor suppressors in a genome-wide CRISPR interference screen of T-cell receptor dependent proliferation. X-axis represents the logarithm of the fold change; y-axis represents the negative logarithm of the false discovery rate. Putative oncogenes are shown in red and putative tumor suppressors in blue. CRISPR interference and amplification screens for TCR-independent cytokine production were similarly analyzed. **c**, *ANKRD11* RNA transcript levels in *ANKRD11* mutant (blue) and wildtype (red) samples. Dots represent individual values, central horizontal bar represents mean, and error bars represent standard error. Comparison by student's two-tailed t-test. **d**, Targeted CRISPR knockout screens in normal T-cells for *ANKRD11*. CFSE dilution progresses (CFSE dye diminishes) from right to left. Stimulated cells are shown in red and unstimulated in blue. **e**, Division index in targeted CRISPR knockout screens in normal T-cells for *ANKRD11* versus control. Error bars represent standard error. Statistical comparison by paired-ratio T-test.

Figure 4 | Functional validation of FOXO3 pro-apoptotic role in ATLL cells. Percent apoptotic cells, as represented by the percent of cells staining positively for annexin V by fluorescence-activated cell sorting (FACS) analysis. Brackets represent comparisons made after subtracting dimethyl sulfoxide

vehicle (DMSO) control values. One, two and three stars represent statistical significance by student's two-tailed t-test ($p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively). Error bars represent standard error from separately treated triplicate samples. Panels on right show representative gated data from FACS analysis, with anti-annexin V on the x-axis and propidium iodine (PI-A) on the y-axis. Western blot panels validating protein levels in knockdown/overexpression constructs are shown in on the top left. β-actin was used as protein loading control. **a,** Percent apoptotic cells in ATLL-97c *FOXO3* knockdown constructs exposed to DMSO vehicle, etoposide 200 nM (Eto), or belinostat 200 nM (Bel). Two *FOXO3* knockdown constructs were established using distinct *FOXO3*-specific single guide RNAs (sgRNA1 and sgRNA2). A construct established from a nonspecific scrambled single guide RNA (SCR) was used as a control. **b,** Percent apoptotic cells in ATLL-97c *FOXO3* overexpression constructs exposed to belinostat (Bel) at concentrations of 200, 400 or 800 nM. m-CAT-1 expressing ATLL-97c cells were transduced with pseudoviral particles containing one of two overexpression constructs (#1 or #5). Cells transduced with empty vectors (EV) were used as a control. **c,** Percent apoptotic cells in ATLL-84c doxycycline (Dox)-inducible constructs exposed to DMSO vehicle, belinostat 200 nM (Bel 200) or belinostat 400 nM (Bel 400). Dox (+) cells were treated with doxycycline 1ug/mL at least 72h before and at the start of drug treatment experiments. Dox (-) cells were not exposed to doxycycline **d,** Percent apoptotic cells after exposure to belinostat 400 nM (Bel 400) in ATLL-84c cells transduced with lentiviruses containing either p.Arg177Trp (R177W) or p.Asp199Asn (D199N) *FOXO3* mutant constructs. Cells transfected with an empty lentivirus (EV) were used as controls. Vector nucleotide sequences were verified by sequencing.

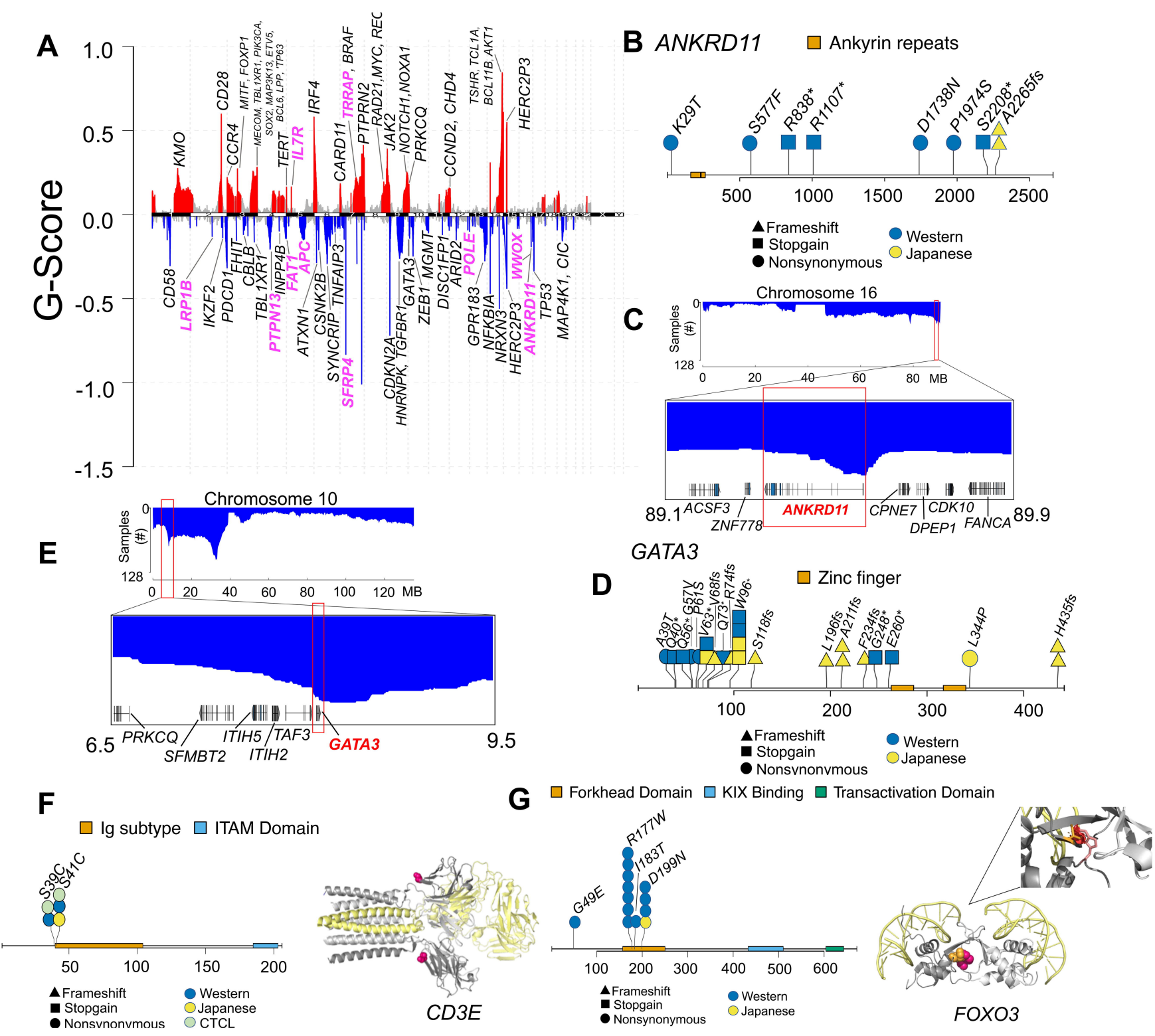
Figure 5 | Differential patterns in mutational frequency between Japanese and Western ATLL cohorts. **a**, **b**, Comparison of point mutation frequency between Japanese/Western and Afro-Caribbean/South American populations. Bar color represents mutation type. Bar height represents the frequency with which a gene is mutated in the specified population. Stars indicate statistical significance: one star indicates $p < 0.05$, two stars indicate $p < 0.01$. Statistical comparison with Fisher's exact test. **a**, Comparison in acute cases (N = 62 Western/39 Japanese, 48 Afro-Caribbean/14 South American). **b**, Comparison in lymphomatous cases (N = 29 Western/13 Japanese, 9 Afro-Caribbean/20 South American). **c-d**, Histograms of **c,** amplifications and **d**, deletions across the genome in Japanese (top)

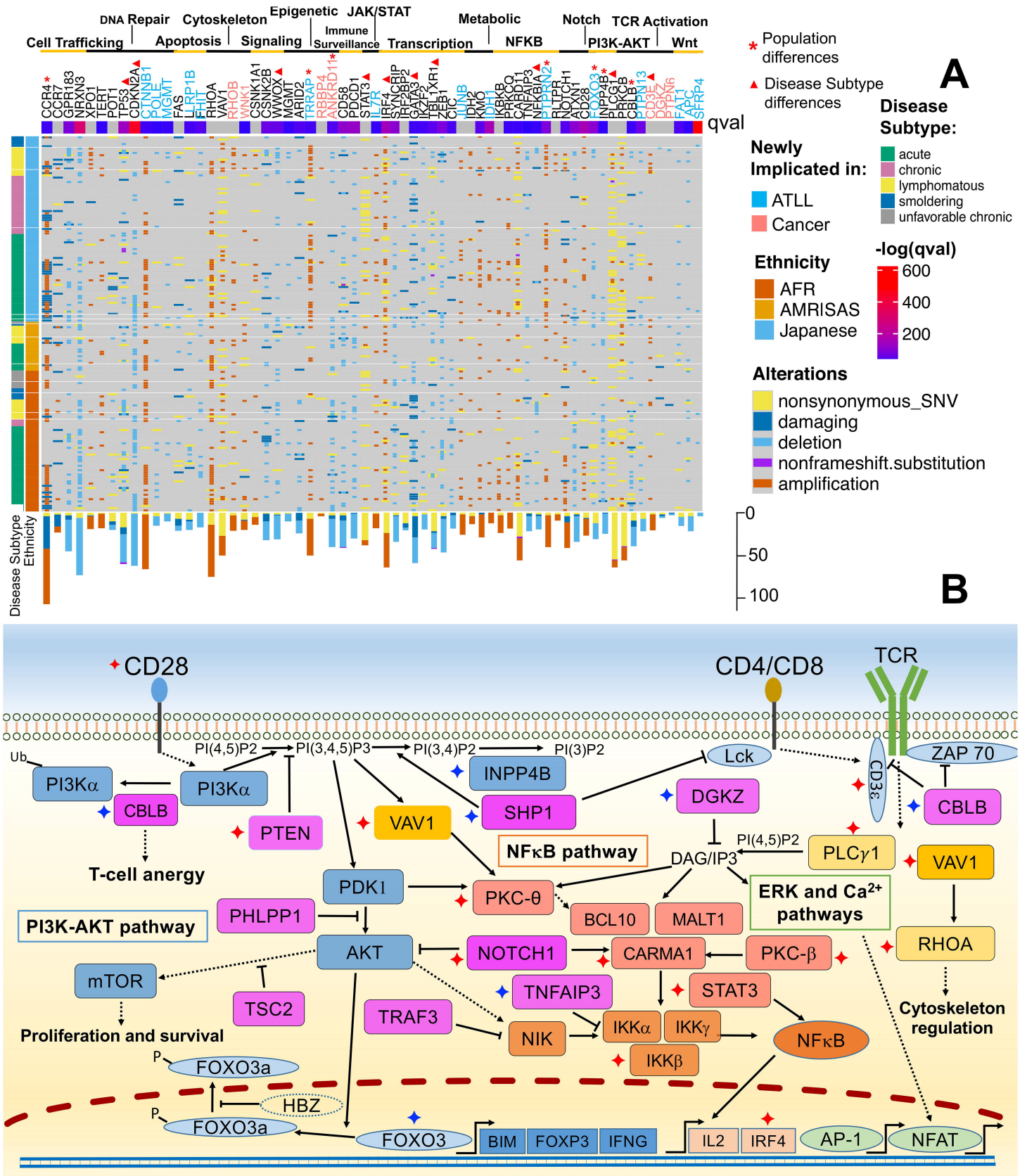
and Western (bottom) cohorts. Y-axis represents the proportion of cohorts with amplification/deletion in a given region. Genes significantly different between cohorts as discussed in the text are highlighted with arrows. Genes newly implicated in ATLL and/or cancer are highlighted in magenta.

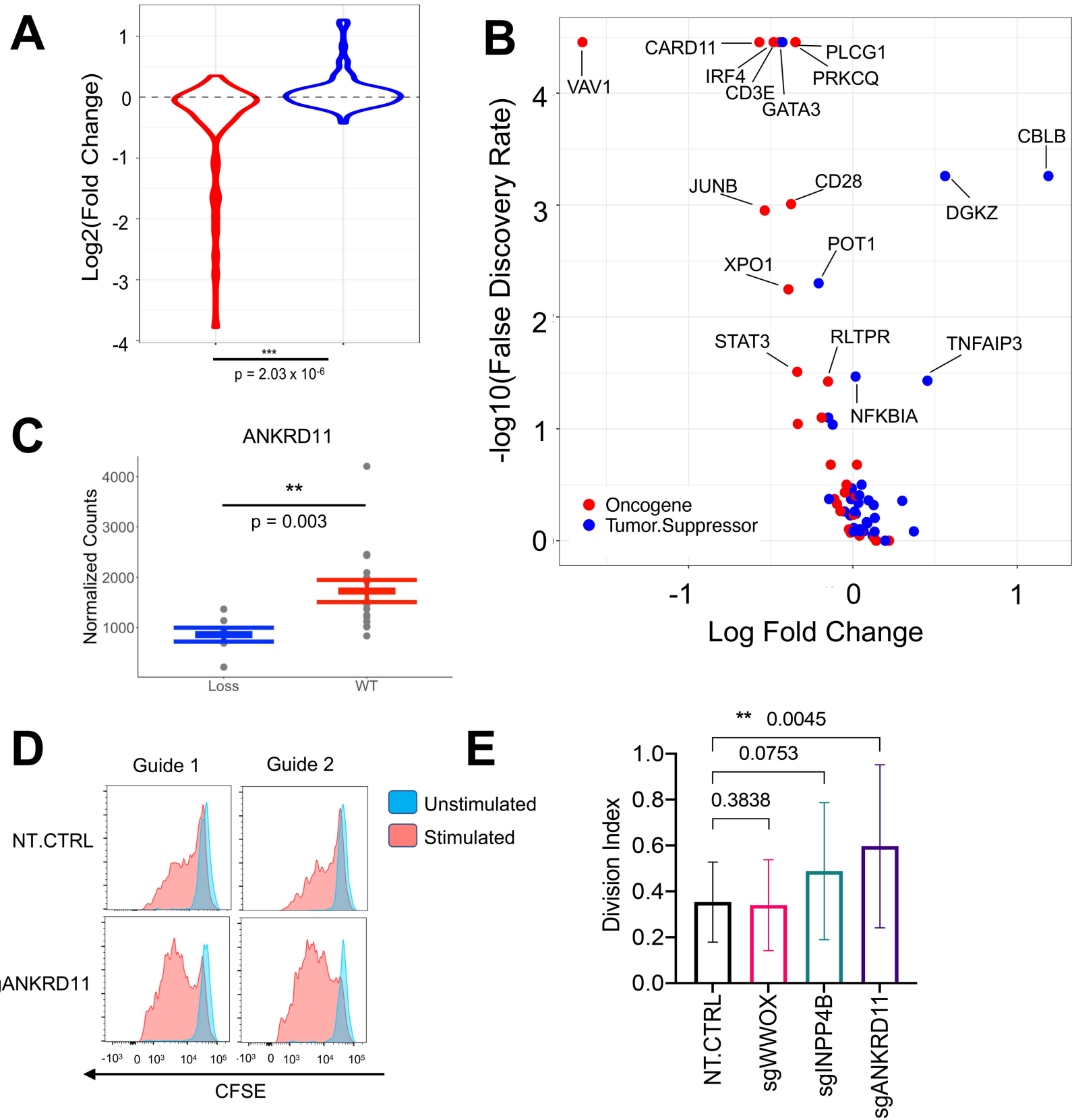
Figure 6 | Genomic alterations associated with clinical subtype and mortality. a, b, Violin plots demonstrating the similarity of unfavorable chronic subtypes (N = 15) to acute (N = 63) and lymphomatous (N = 31) subtypes regarding **a,** the number of genes deleted or amplified per sample and **b,** the number of point mutations per sample. Bars represent mean (central bar) and standard error. One, two and three stars represent $p < 0.05$, 0.01, and 0.001 by student's two-tailed T-test. **c**, Comparison of GATA3 protein levels by immunohistochemistry in aggressive (acute and lymphomatous) compared to unfavorable chronic cases, student's two-tailed t-test. Dots represent individual values, central horizontal bar represents mean, and error bars represent standard error**. d-e**, Kaplan-Meier curves indicating the effect of **d,** *TP53* and **e,** *ANKRD11* mutation on overall survival. Red lines indicate overall mortality for patients with mutation in the specified gene. Blue lines indicate overall mortality for patients without mutation in that gene. Significance was determined using a Cox multivariate analysis of the overall cohort controlling for disease subtype and treatment modality.[45] P_{adj} = p-value after Bonferroni correction for multiple comparisons. WT = wildtype. **f**, Frequencies of chemotherapy responses in patients with *CDKN2A* deletion/damaging mutation (left) versus *CDKN2A* wildtype (right). Star indicates statistical significance, Fisher's exact test. CR = Complete Response, PR = Partial Response.

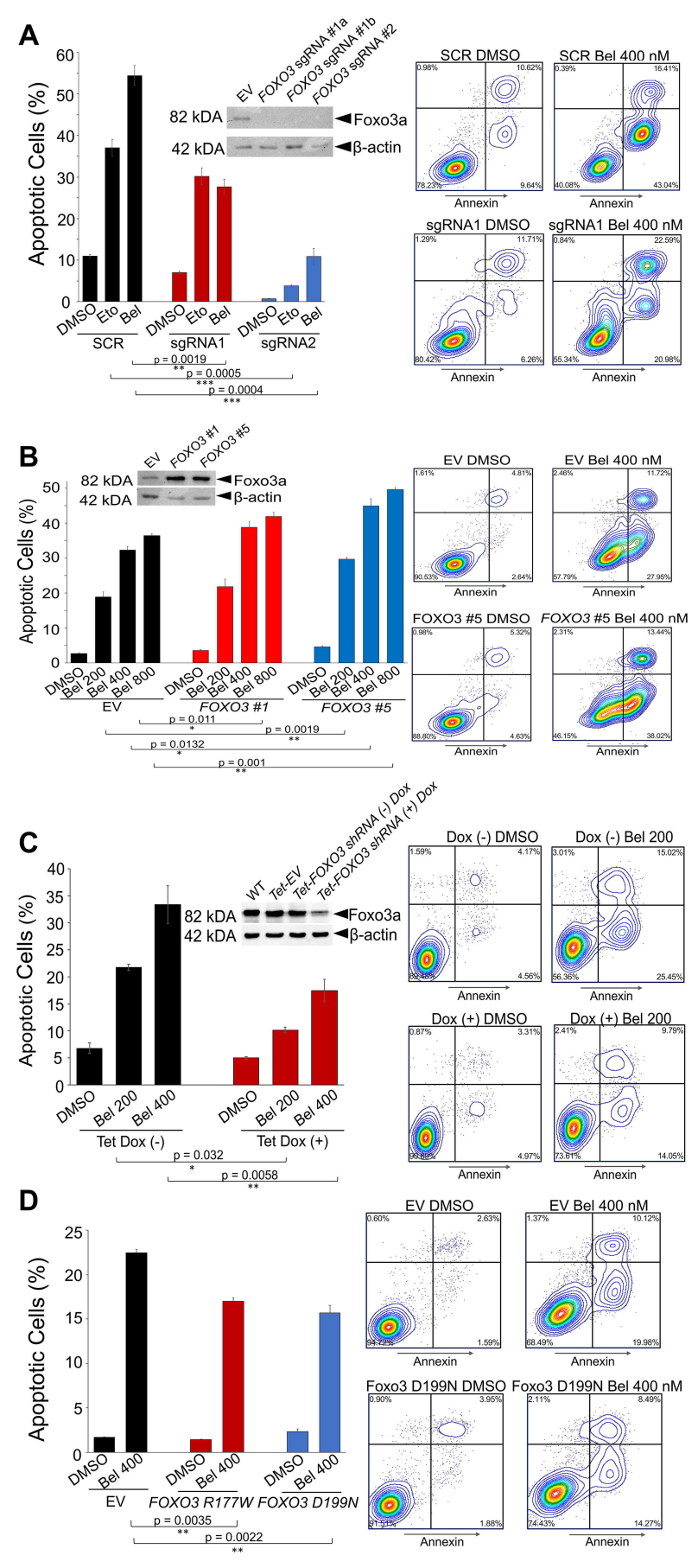
Figure 7 | Association of IRF4 mutation with ATLL relapse. a, Clonality of tumor populations as determined by T-cell receptor sequence identification from RNA-seq data. Patients shown here contributed samples from initial disease as well as subsequent relapse. Clonal populations were defined as cells with identical T-cell receptors. Each blue shape represents the dominant T-cell clone, as determined by α and β T-cell receptor subunit identity. The vertical axis represents the size of clonal populations as determined by the frequencies of the dominant α and β subunit. Relapses are indicated by vertical white lines. Multiple relapses are shown in sequential order, from left to right. **b**, Comparison of point and CNV mutations in initial (I) versus relapsed (R) samples. Pairs of initial/relapsed samples from

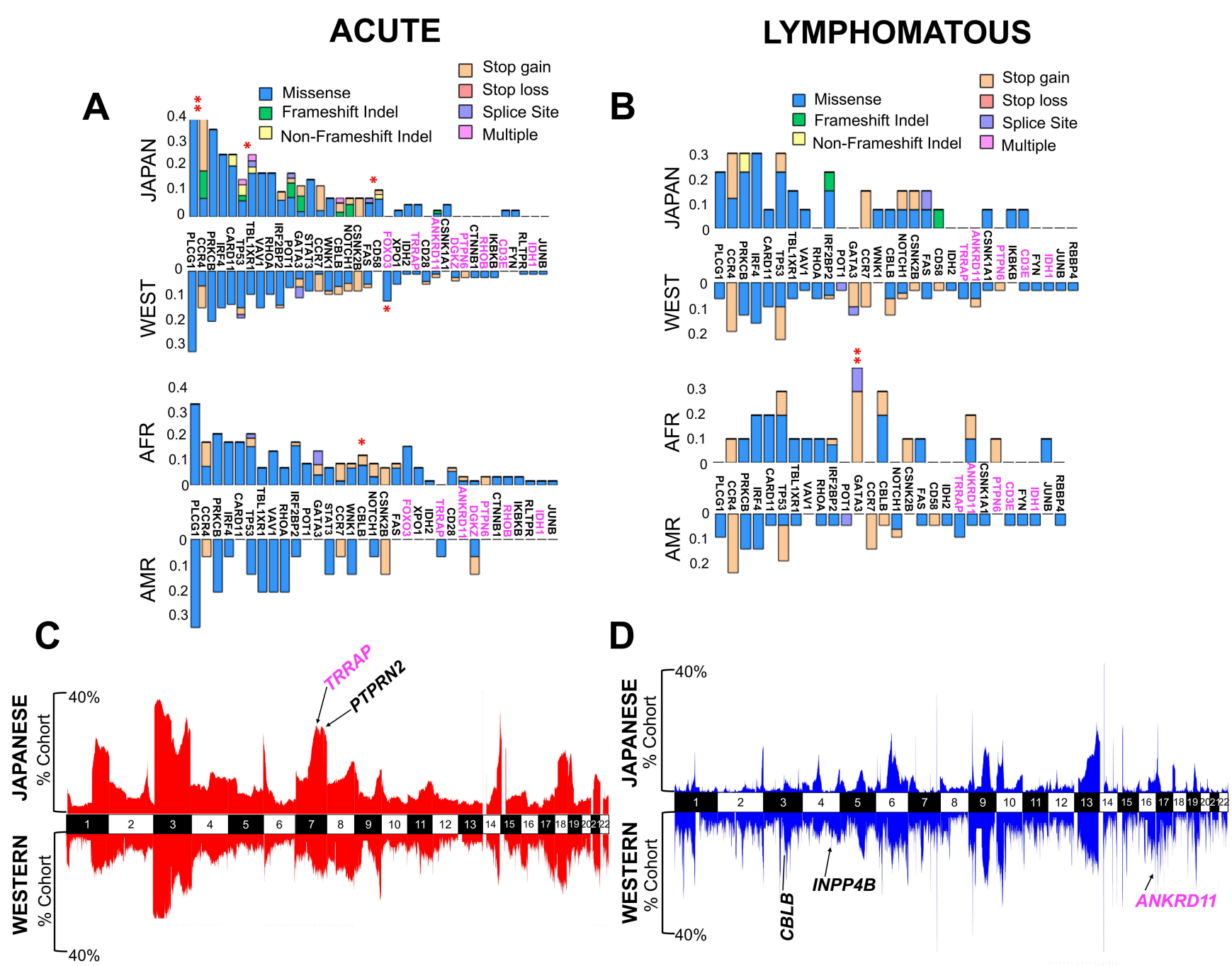
the same patient are indicated in light and dark shades, respectively, of alternating blue and green. Mutation types include deletions, amplifications, damaging mutations and nonsynonymous (missense) mutations, as shown in the legend. Putative driver genes are listed on the left. **c,** Frequency of new *IRF4* mutation in first-time disease (N = 121) compared to relapsed samples (N = 12). Star represents statistical significance, Fisher's exact test. **c**, Distribution of mutations across *IRF4*. Mutations shown in blue occurred in Western samples (N = 121), yellow in Japanese (N = 83). Especially frequent mutations are represented as pie markers indicating the proportion of variants contained in either the Western (blue) or Japanese (yellow) cohort. Marker stem heights represent the number of cases as depicted on the y-axis.

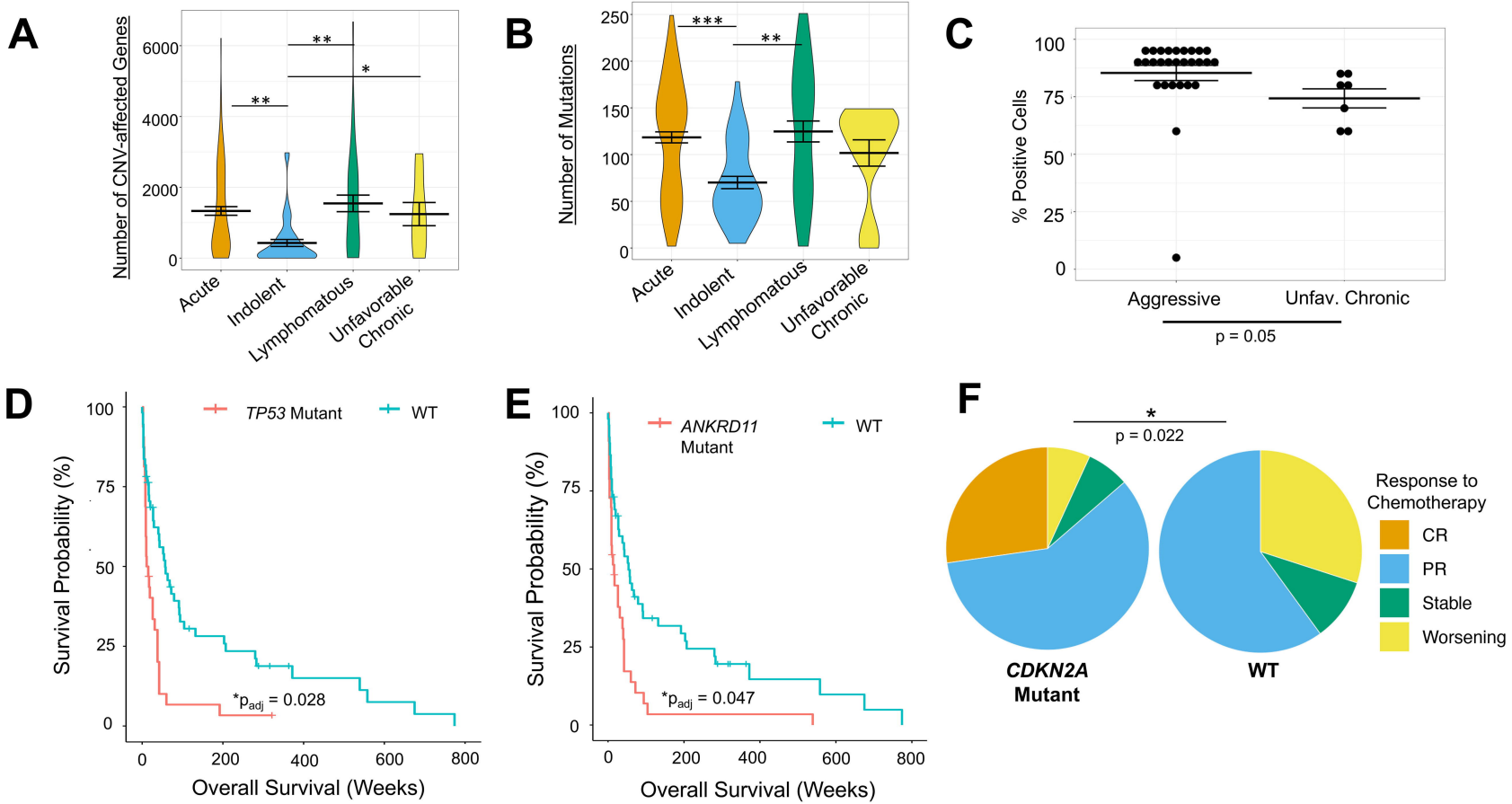

A
G-Score
1.0
0.5
0.0
-0.5
-1.0
-1.5
KMO
CD28
CCR4
MITF, FOXP1
MECOM, TBL1XR1, PIK3CA, SOX2, MAP3K13, ETV5, BCL6, LPP, TP63
TERT
IL7R
IRF4
CARD11
TRRAP, BRAF
PTPRN2
RAD21,MYC, REC
JAK2
NOTCH1,NOXA1
PRKCQ
CCND2, CHD4
TSHR, TCL1A, BCL11B, AKT1
HERC2P3
CD58
LRP1B
IKZF2
PDCD1
FHIT
CBLB
TBL1XR1
PTPN13
INPP4B
FAT1
APC
ATXN1
CSNK2B
SYNCRIP
TNFAIP3
SFRP4
CDKN2A
HNRNPK, TGFBR1
GATA3
ZEB1
MGMT
DISC1FP1
ARID2
POLE
GPR183
NFKBIA
NRXN3
HERC2P3
ANKRD11
WWOX
TP53
MAP4K1, CIC
B
ANKRD11
Ankyrin repeats
K29T
S577F
R838*
R1107*
D1738N
P1974S
S2208*
A2265fs
500
1000
1500
2000
2500
Frameshift
Stopgain
Nonsynonymous
Western
Japanese
C
Chromosome 16
Samples (#)
0
128
0
20
40
60
80
MB
89.1
ACSF3
ZNF778
ANKRD11
CPNE7
CDK10
DPEP1
FANCA
89.9
D
GATA3
Zinc finger
A39T
Q40*
Q56*
G57V
P61S
V63*
V68fs
Q73*
R74fs
W96*
S118fs
L196fs
A211fs
F234fs
G248*
E260*
L344P
H435fs
100
200
300
400
Frameshift
Stopgain
Nonsynonymous
Western
Japanese
E
Chromosome 10
Samples (#)
0
128
0
20
40
60
80
100
120
MB
6.5
PRKCQ
SFMBT2
ITIH5
ITIH2
TAF3
GATA3
9.5
F
Ig subtype
ITAM Domain
S39C
S41C
50
100
150
200
Frameshift
Stopgain
Nonsynonymous
Western
Japanese
CTCL
CD3E
G
Forkhead Domain
KIX Binding
Transactivation Domain
R177W
I183T
D199N
G49E
100
200
300
400
500
600
Frameshift
Stopgain
Nonsynonymous
Western
Japanese
FOXO3

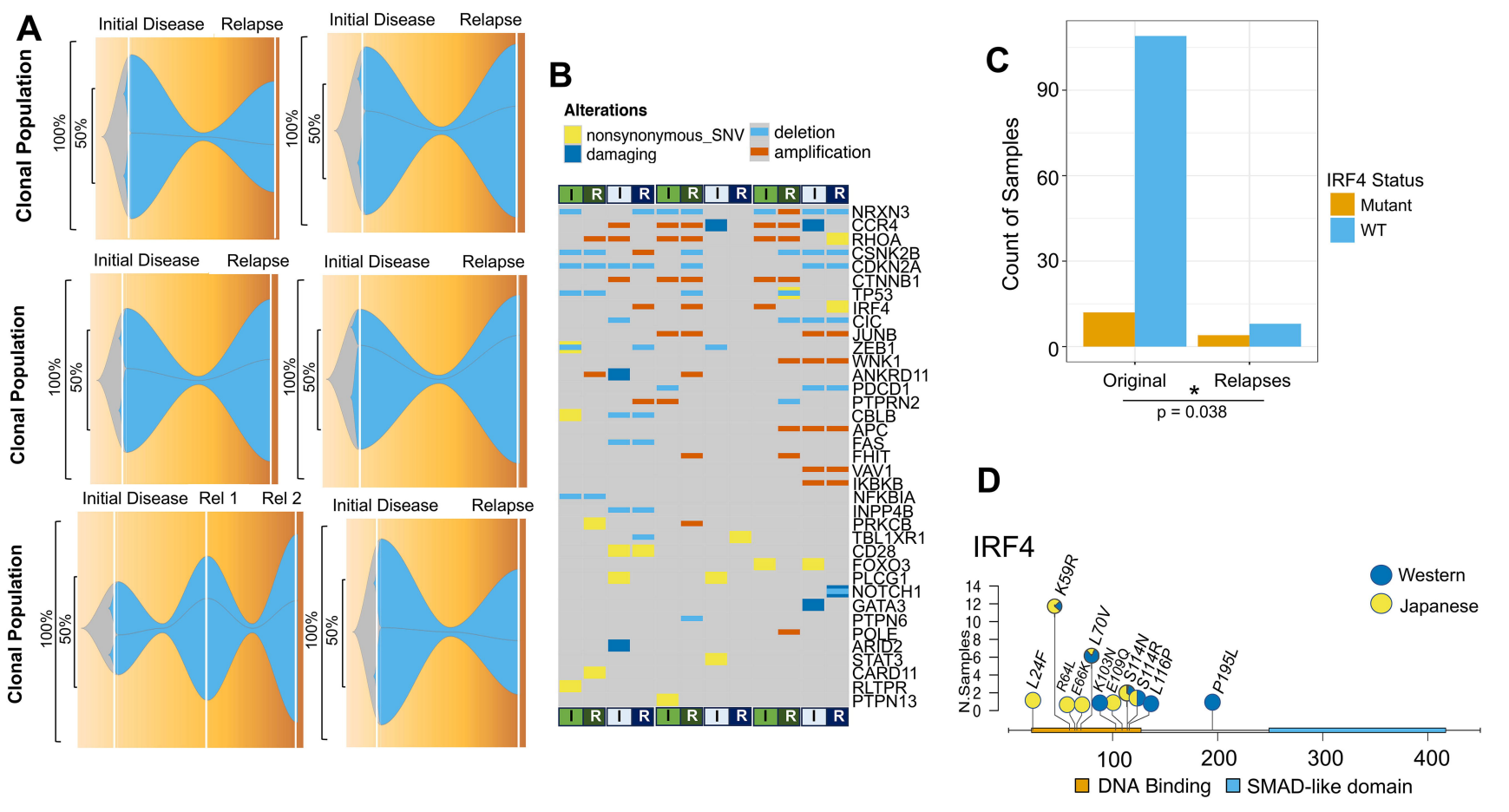

Cell Trafficking
DNA Repair
Apoptosis
Cytoskeleton
Signaling
Epigenetic
Immune Surveillance
JAK/STAT
Transcription
Metabolic
NFKB
Notch
PI3K-AKT
TCR Activation
Wnt
Population differences
Disease Subtype differences
A
Disease Subtype:
acute
chronic
lymphomatous
smoldering
unfavorable chronic
Newly Implicated in:
ATLL
Cancer
Ethnicity
AFR
AMRISAS
Japanese
-log(qval)
600
400
200
Alterations
nonsynonymous_SNV
damaging
deletion
nonframeshift.substitution
amplification
qval
Disease Subtype
Ethnicity
0
50
100
B
CD28
CD4/CD8
TCR
PI(4,5)P2
PI(3,4,5)P3
PI(3,4)P2
PI(3)P2
Ub
PI3Kα
CBLB
T-cell anergy
PTEN
INPP4B
SHP1
VAV1
NFκB pathway
Lck
ZAP 70
CD3ζ
DGKZ
PLCγ1
DAG/IP3
ERK and Ca2+ pathways
PDK1
PKC-θ
PHLPP1
PI3K-AKT pathway
BCL10
MALT1
AKT
NOTCH1
CARMA1
PKC-β
RHOA
mTOR
TSC2
Proliferation and survival
TNFAIP3
STAT3
TRAF3
NIK
IKKα
IKKγ
IKKβ
NFκB
Cytoskeleton regulation
P
FOXO3a
HBZ
FOXO3
BIM
FOXP3
IFNG
IL2
IRF4
AP-1
NFAT


A
Log2(Fold Change)
1
0
-1
-2
-3
-4
***
p = 2.03 x 10-6
B
-log10(False Discovery Rate)
4
3
2
1
0
VAV1
CARD11
IRF4
CD3E
GATA3
PLCG1
PRKCQ
CBLB
JUNB
CD28
DGKZ
POT1
XPO1
STAT3
RLTPR
TNFAIP3
NFKBIA
Oncogene
Tumor.Suppressor
-1
0
1
Log Fold Change
C
ANKRD11
Normalized Counts
4000
3000
2000
1000
**
p = 0.003
Loss
WT
D
Guide 1
Guide 2
NT.CTRL
sgANKRD11
Unstimulated
Stimulated
-10³ 0 10³ 10⁴ 10⁵
CFSE
E
** 0.0045
0.0753
0.3838
Division Index
1.0
0.8
0.6
0.4
0.2
0.0
NT.CTRL
sgWWOX
sgINPP4B
sgANKRD11

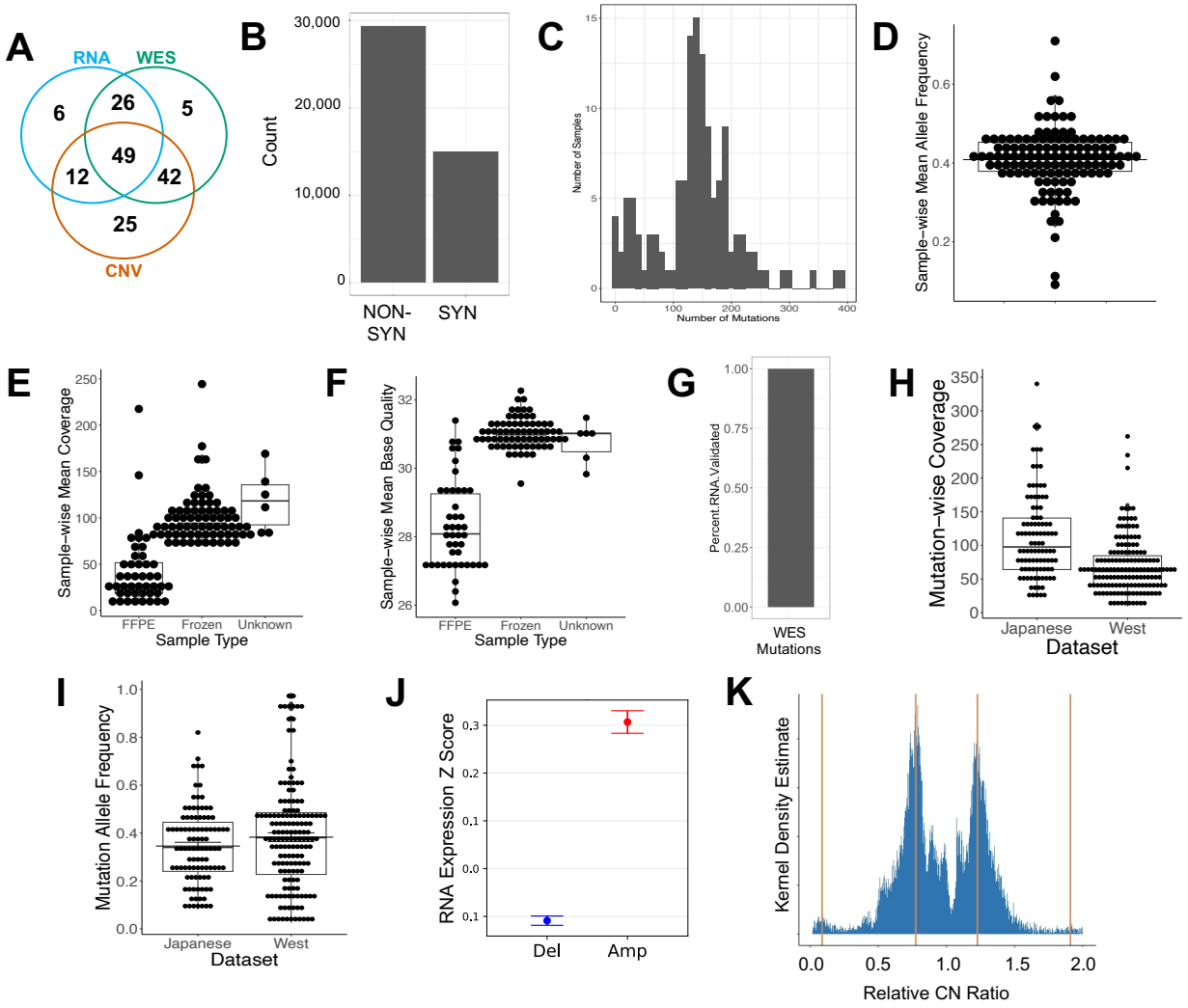

A
Apoptotic Cells (%)
EV
FOXO3 sgRNA #1a
FOXO3 sgRNA #1b
FOXO3 sgRNA #2
82 kDA
42 kDA
Foxo3a
β-actin
DMSO
Eto
Bel
SCR
sgRNA1
sgRNA2
p = 0.0019
**
p = 0.0005
***
p = 0.0004
***
SCR DMSO
SCR Bel 400 nM
sgRNA1 DMSO
sgRNA1 Bel 400 nM
Annexin
B
EV
FOXO3 #1
FOXO3 #5
Bel 200
Bel 400
Bel 800
p = 0.011
*
p = 0.0019
**
p = 0.0132
*
p = 0.001
**
EV DMSO
EV Bel 400 nM
FOXO3 #5 DMSO
FOXO3 #5 Bel 400 nM
C
WT
Tet-EV
Tet-FOXO3 shRNA (-) Dox
Tet-FOXO3 shRNA (+) Dox
Tet Dox (-)
Tet Dox (+)
p = 0.032
*
p = 0.0058
**
Dox (-) DMSO
Dox (-) Bel 200
Dox (+) DMSO
Dox (+) Bel 200
D
FOXO3 R177W
FOXO3 D199N
p = 0.0035
**
p = 0.0022
**
EV DMSO
EV Bel 400 nM
Foxo3 D199N DMSO
Foxo3 D199N Bel 400 nM


ACUTE
LYMPHOMATOUS
A
B
Missense
Frameshift Indel
Non-Frameshift Indel
Stop gain
Stop loss
Splice Site
Multiple
JAPAN
WEST
AFR
AMR
PLCG1 CCR4 PRKCB IRF4 CARD11 TP53 TBL1XR1 VAV1 RHOA IRF2BP2 POT1 GATA3 STAT3 CCR7 WNK1 CBLB NOTCH1 CSNK2B FAS CD58 FOXO3 XPO1 IDH2 TRRAP CD28 ANKRD11 CSNK1A1 DGK2 PTPN6 CTNNB1 RHOB IKBKB CD3E FYN RLTPR IDH1 JUNB RBBP4
C
D
TRRAP
PTPRN2
JAPANESE
% Cohort
WESTERN
% Cohort
40%
CBLB
INPP4B
ANKRD11


A
Number of CNV-affected Genes
6000
4000
2000
0
**
**
*
Acute
Indolent
Lymphomatous
Unfavorable
Chronic
B
Number of Mutations
250
200
150
100
50
0
***
**
Acute
Indolent
Lymphomatous
Unfavorable
Chronic
C
% Positive Cells
100
75
50
25
0
Aggressive
Unfav. Chronic
p = 0.05
D
Survival Probability (%)
100
75
50
25
0
TP53 Mutant
WT
*padj = 0.028
0
200
400
600
800
Overall Survival (Weeks)
E
Survival Probability (%)
100
75
50
25
0
ANKRD11
Mutant
WT
*padj = 0.047
0
200
400
600
800
Overall Survival (Weeks)
F
*
p = 0.022
CDKN2A
Mutant
WT
Response to
Chemotherapy
CR
PR
Stable
Worsening

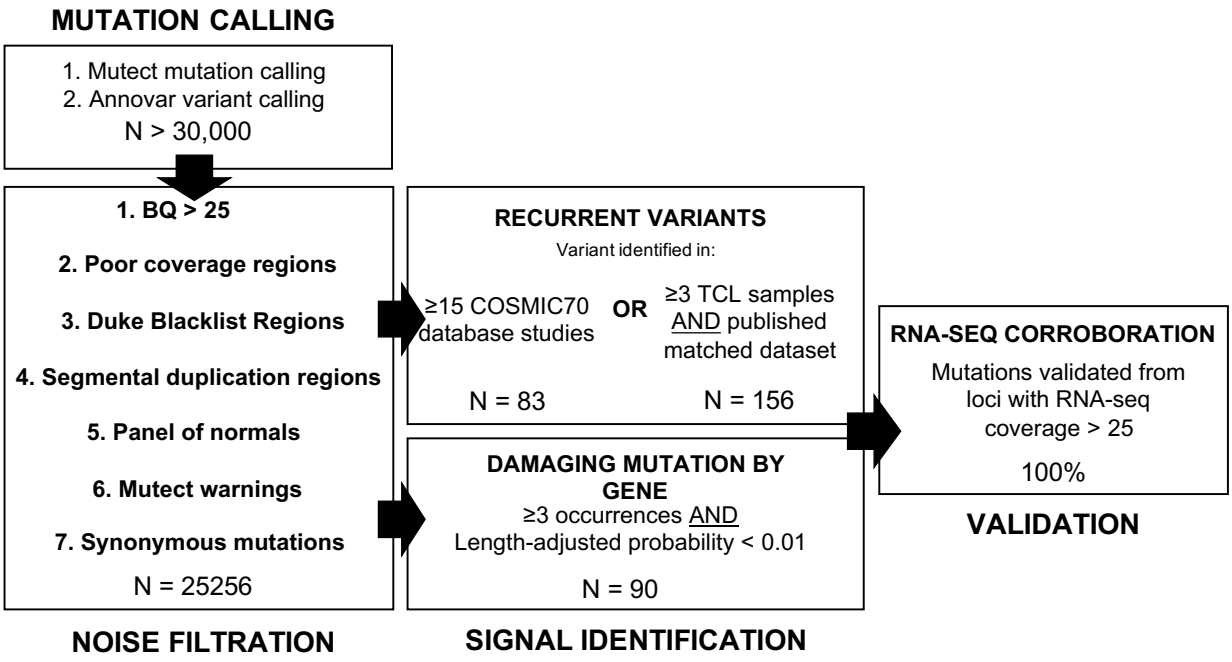

A
Initial Disease
Relapse
Clonal Population
100%
50%
Rel 1
Rel 2
B
Alterations
nonsynonymous_SNV
damaging
deletion
amplification
I R I R I R I R I R I R
NRXN3
CCR4
RHOA
CSNK2B
CDKN2A
CTNNB1
TP53
IRF4
CIC
JUNB
ZEB1
WNK1
ANKRD11
PDCD1
PTPRN2
CBLB
APC
FAS
FHIT
VAV1
IKBKB
NFKBIA
INPP4B
PRKCB
TBL1XR1
CD28
FOXO3
PLCG1
NOTCH1
GATA3
PTPN6
POLE
ARID2
STAT3
CARD11
RLTPR
PTPN13
C
Count of Samples
90
60
30
0
IRF4 Status
Mutant
WT
Original
Relapses
*
p = 0.038
D
IRF4
N.Samples
14
12
10
8
6
4
2
0
K59R
L70V
L24F
R64L
E66K
K103N
E109Q
S114N
S114R
L116P
P195L
Western
Japanese
100
200
300
400
DNA Binding
SMAD-like domain

SUPPLEMENTAL MATERIALS

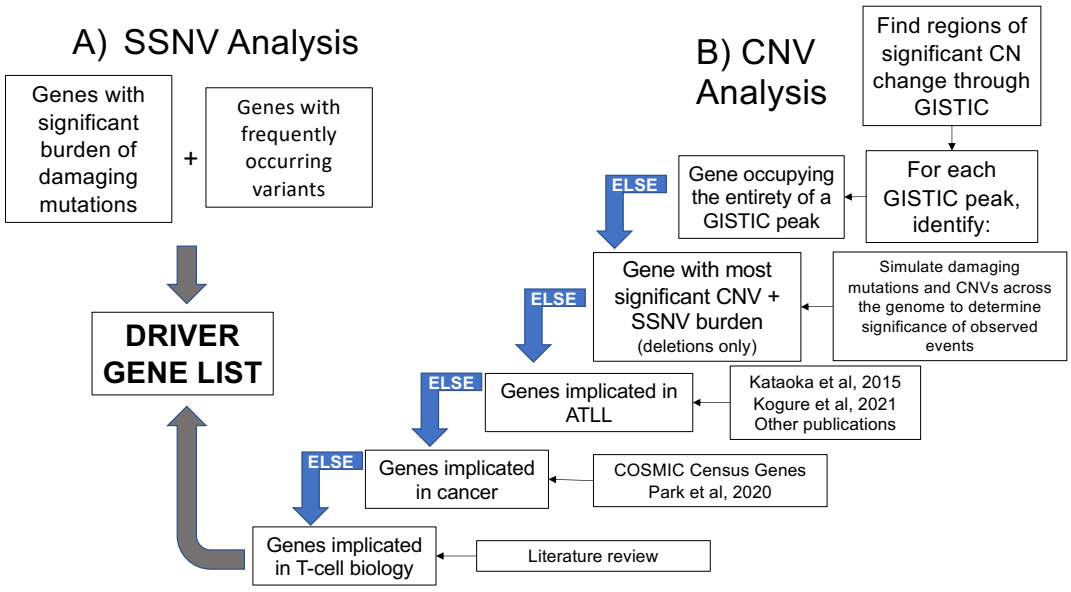


Supplemental Figure S1 | QC Metrics for whole exome and CNV sequencing data. **a**, Venn diagram of Western patients with samples analyzed by RNA-seq (blue) whole exome sequencing (green), copy number variation analysis (orange), or multiple methods. **b,** Ratio of nonsynonymous to synonymous point 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

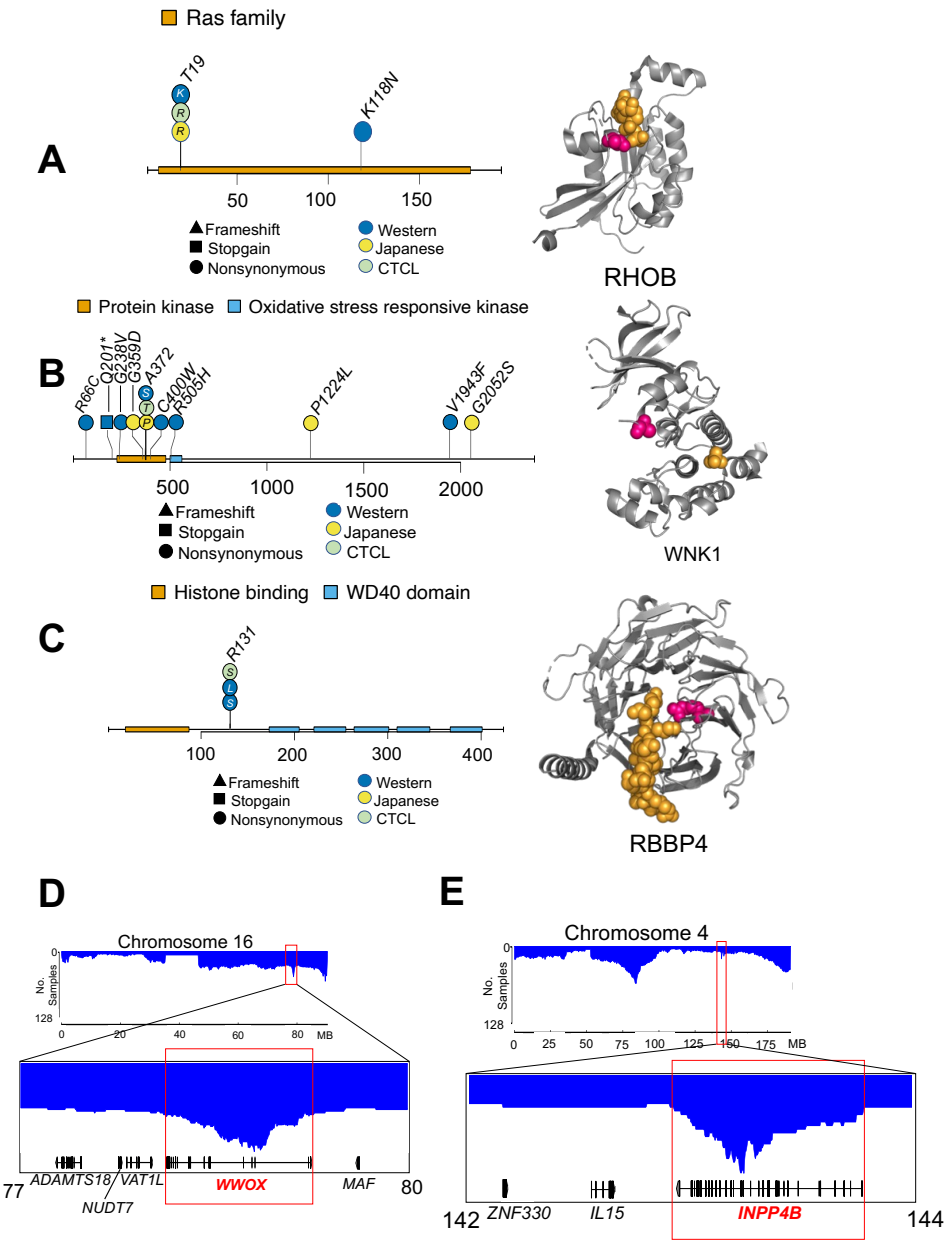
driver mutations with DNA- and RNA-seq coverage > 25 validated by RNA-seq data**. h,** Sequencing coverage and **i,** mutation allele frequency of variants of interest identified in Japanese (left) and Western (right) patients. **j,** RNA-expression of genes within significant copy-number variation (CNV) 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.

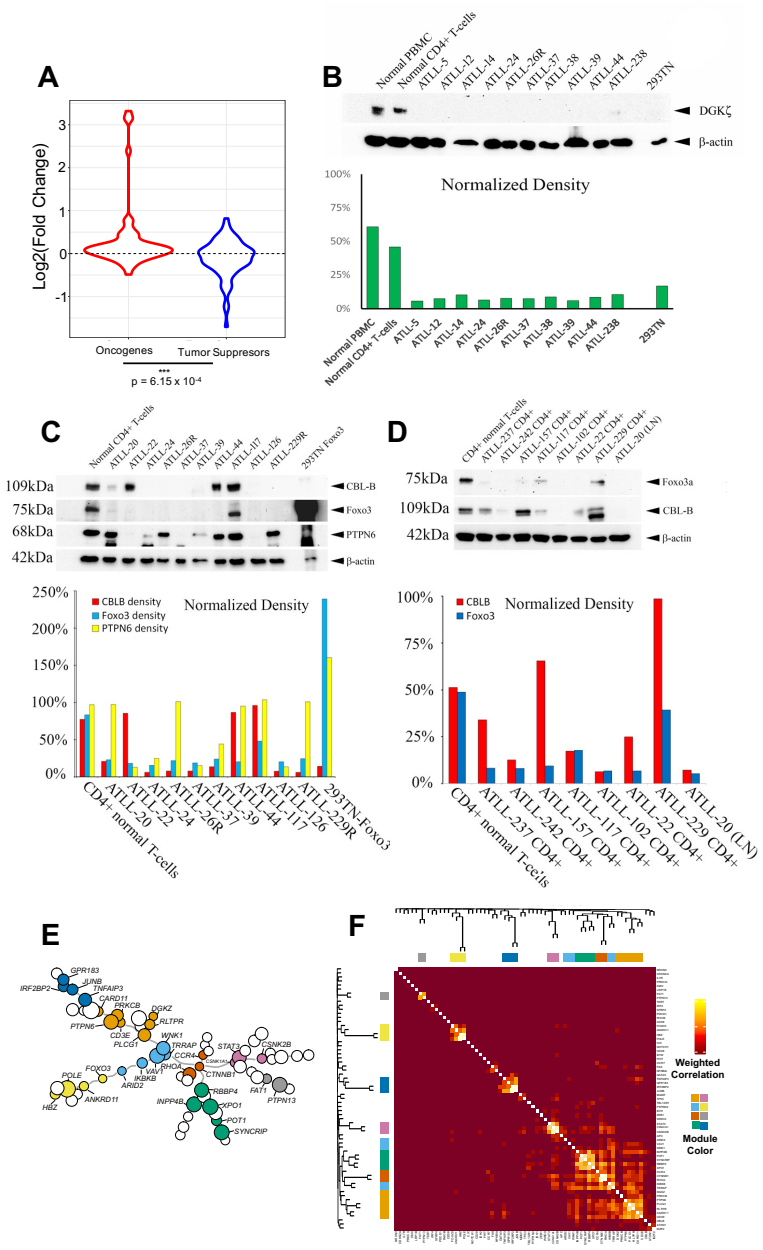


Supplemental Figure S3| Schematic of the hierarchical method used to call driver genes in peaks determined to be significant by GISTIC2.0. Regions of significant amplification/deletion by GISTIC analysis were first examined for putative driver genes satisfying the criterium in the top right. If that 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.



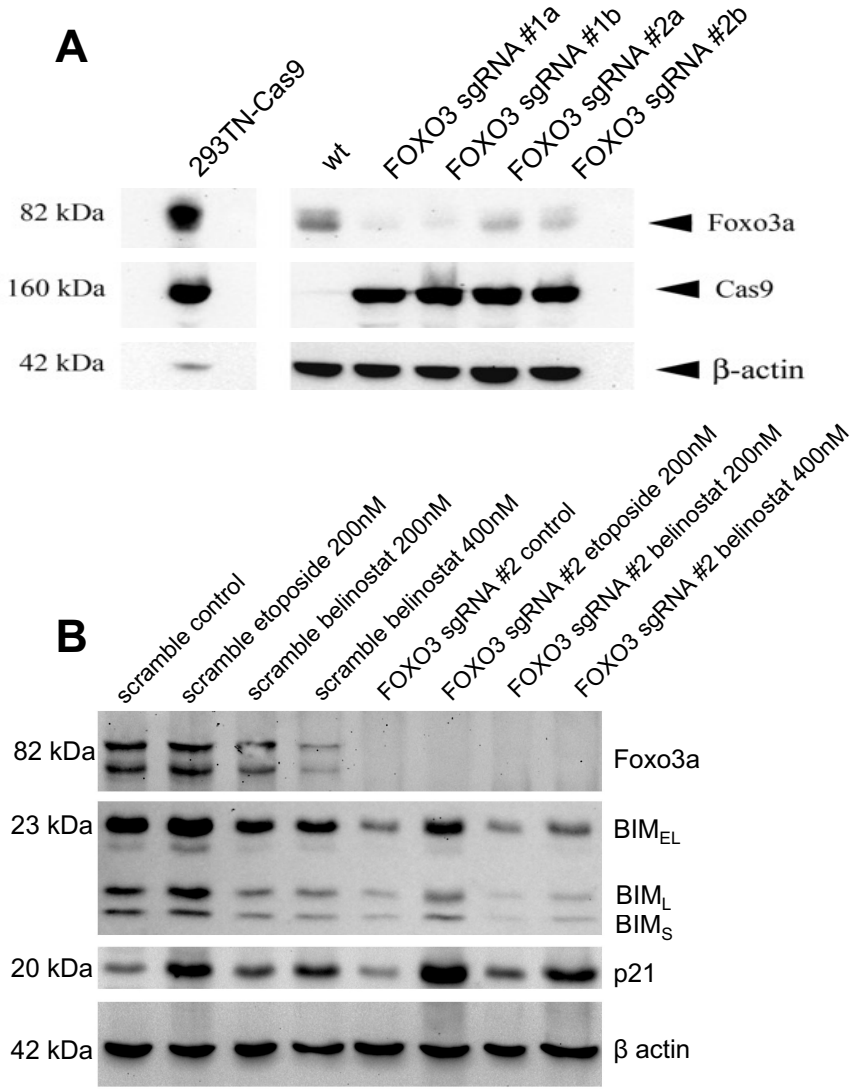
Supplemental Figure S4 | Putative driver genes identified by point mutation and copy number mutation analysis. **a-c** Novel ATLL putative driver genes identified by point analysis. Lollipop plots on the left show the location of point mutations, with mutation types and cohorts indicated by the legend. 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.



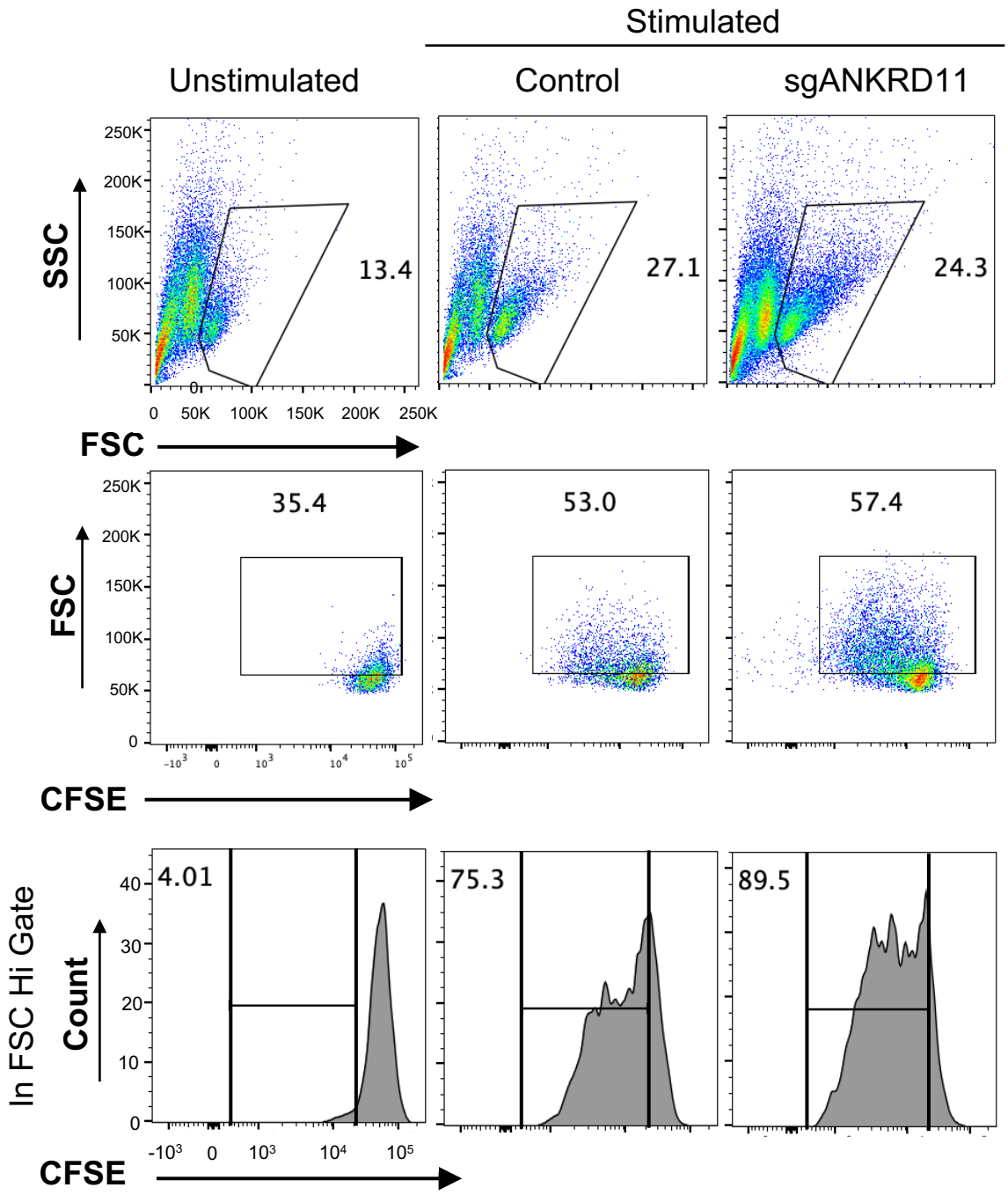
Supplemental Figure S5 | Functional analyses of putative ATLL driver genes. **a,** Violin plots showing log fold change (**LFC**) of sgRNAs for putative oncogenes and putative tumor suppressor in genome-wide CRISPR amplification screens for T-cell receptor dependent proliferation and cytokine production.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

TABLES 1-9 PROVIDES AS EXCEL FILES

Online Supplementary 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.

Online Supplementary 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.

Online Supplementary 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.

Online Supplementary 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.

Online Supplementary 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.

Online Supplementary 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.

Online Supplementary 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.

Online Supplementary Table S8. IRF4 Mutation validation by Sanger Sequencing.

Online Supplementary 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.