Supplemental Materials

Integrated Genomic Analyses of Cutaneous T Cell Lymphomas Reveal the Molecular Bases for Disease Heterogeneity

Supplemental Methods

Sample collection and sorting strategy details

Recent reports suggesting that leukemic CTCLs harbor heterogeneous immune phenotypes, including naïve (N), central memory (CM), effector memory (EM), or terminally differentiated effector memory cells (EMRA)[1,2]. In healthy CD4+ T cells, these different immune subsets respond differently to TCR stimulation. Therefore, to test these cells functionally, we immunophenotyped our samples to avoid potentially confounding immunophenotype-dependent differences. 4 immune subsets by pairs of CD45RA/CD45RO or CCR7/CD45RA as follows: CD45RA+CD45RO- for RA, CD45RA-CD45RO+ for RO, CD45RA+CCR7+ for naïve (**N**), CD45RA-CCR7+ for central memory (**CM**), CD45RA-CCR7- for effector memory (**EM**), and CD45RA+CCR7- for terminally differentiated effector memory cells (**EMRA**).

For DNA-seq, malignant cells from step 3 above were subject to DNA extraction. For RNA-seq, **Supplemental Table 3** presents available immune subsets for each patient. In detail, for NU114, NU134, NU161, NU16, NU30, NU51, NU55, NU56, NU78, and NU80, RA/RO sorting was done and subject to RNAseq. For NU115, NU201, NU208, and NU8, N/CM/EM/EMRA staining was done and subject to RNAseq. The samples for RA isolation were chosen due to low levels of EMRA population (<10%). For functional assay, N/CM/EM/EMRA sorting was done for all samples, and available subsets for each patient are specified in **Supplemental Table 3**.

Flow cytometry antibodies

FACS antibodies used were as follows: CD3-Pacific blue (Biolegend #317313), CD3-APC (Biolegend #317318), CD8-PerCPCy5.5 (eBioscience #45-0088-42), CD26-PE (Biolegend #302705), TCR Vb2-PE (Miltenyi Biotec #130-110-095), TCR Vb13-FITC (eBioscience #11-5792-41), TCR Vb14-PE (Miltenyi Biotec #130-108- 804), TCR Vb17-PE (Beckman coulter #1M2048), CD45RO-Pacific blue (Biolegend #304216), CD45RA-PECy7 (Biolegend #304126), CCR7-APC (Biolegend #353214), anti-PD-1-FITC (Biolegend #329903), anti-IL-2-FITC (Biolegend #500305), anti-IL-4-FITC (Biolegend #500806), anti-IL-17A-APC-Cy7 (Biolegend #512319), anti-IFNg-APC-Cy7 (Biolegend #502559), and anti-granzyme-B-FITC (Biolegend #515403).

DNA sequencing analysis additional information

We performed DNA-sequencing on samples from 94 patients with CTCL (**Supplemental Table 1**). 75 patients (80%) had their tumor analyzed by WGS, and 19 (20%) by WES. 52 patients (55%) had matched normal DNA available. 44 patients (47%) had samples isolated from blood, in which malignant cells were sorted by flow cytometry (see Methods). 50 samples (53%) had skin tumors analyzed by DNA sequencing, of which 30 tumors (60%) were fresh frozen and 20 tumors (40%) were formalin fixed, paraffin embedded.

To identify SCNVs in WGS samples, we utilized Patchwork with a window size of 10,000 bp[3]. For quality control, we excluded calls with discordant $\log_2$ read ratio and delta B-allele frequency (**BAF**) or present at a high frequency in GnomAD[4]. For WES samples, GATK4CNV was used to call copy number variants[5]. Chromosomal arms >50% deleted or amplified were considered broad events. To identify significant focal deletions, GISTIC2.0 software was utilized[6]. Clipping Reveals Structure (CREST) software was utilized to call breakpoints in matched WGS samples[7]. Soft clipped reads were extracted for each chromosome, germline structural variant events were removed, and breakpoints were identified. Samples successfully analyzed by CREST are indicated in **Supplemental Table 1**. To identify genes affected by translocations more frequently than by chance alone, the probability of the observed number of translocations occurring by chance was calculated using the binomial distribution

based on the rate of translocations, adjusted for gene length. Bonferroni correction was applied for multiple hypothesis testing.

Identification of putative driver genes

To implicate specific genes within each region of significantly recurrent SCNVs, we employed a multi-tiered approach (**Supplemental Fig. 1**). First, we identified recurrent SCNVs that occurred more often than expected by chance using GISTIC2.0[6].

Then to identify the target gene, we first used gene localizing mutations to identify the target gene in each GISTIC confidence interval. We calculated the likelihood ratio of each gene in each interval. We calculated the observed combined burden of SSNVs and focal SCNVs as previously described[8]. In brief, the probability of the observed number of SSNVs and focal deletions occurring within each gene within significant SCNVs was calculated from the binomial distribution, adjusting for gene expression and gene length. The likelihood ratio was calculated as the ratio between the reciprocal of the *P* value of the most significant and next most significant gene, and a likelihood ratio >5 was considered significant.

Next, we identified peaks with single genes within the minimally commonly deleted region (i.e., a lone gene common to the highest number of samples), genes previously implicated in CTCL, consensus cancer genes as determined by COSMIC, and finally genes with known function in T cell biology based on the literature.

MutSigCV[9] was employed as previously described, using RNA-seq data from this study as the gene expression covariate.

Hotspot mutations occurring more often than expected by chance alone and genes with a significant burden of damaging mutations were identified as described previously[10]. For hotspot mutations, the probability of a codon being mutated the observed number of times was calculated with the gene expression adjusted rate of mutations for each codon using the binomial distribution.

For damaging mutations, the probability of a gene having the observed number of damaging mutations (including frameshift, stopgain, and splice site mutations) was calculated with the gene expression and gene length adjusted rate of damaging mutations using the binomial distribution. Then, genes not expressed in normal CD4+ T cells were excluded. For each analysis, Bonferroni correction for multiple hypothesis testing was applied.

Ex vivo proliferation assay and analysis details

Isolated cells were stained with carboxyfluorescein succinimidyl ester Cell Division Tracker Kit (Biolegend) and cultured in a 96 well round bottom plate for 7 days in RPMI-1640 (Gibco) supplemented with 10% heat-inactivated fetal bovine serum (Gemini) and penicillin/streptomycin (Gibco). 5 different stimuli, 1.anti-CD3 (10µg/ml, Biolegend #317325), 2.anti-CD3 (10µg/ml, Biolegend #317325)/ anti-CD2 (10µg/ml, Biolegend #309212), 3.anti-CD3/CD28 (1:1 ratio, Dynabeads, Life Technologies), 4.IL7 (50ng/ml, Biolegend #581902)/ IL15 (200ng/ml, Biolegend #570302), 5.IL4 (100ng/ml, Biolegend #574002)/ IL6 (50ng/ml, Biolegend #570802)/ IL10 (50ng/ml, Biolegend #571002)/ TNFa (20ng/ml, Biolegend #570102) were used to determine subset specific proliferation. Proliferation Tool from FlowJo v9.9.6. (Tree Star) was used to analyze flow cytometry data from CFSE assays. Original population was assigned based on the peak from the anti-CD3 condition, which showed no proliferation. Number of peaks was assigned using the calculation of 1+log2(FITC-A of original population/FITC- A of most divided population). Default settings were used for other options, and model adjustment was used to generate the model fit to the actual flow plot. Division index indicates the average number of cell divisions that a cell in the original population underwent as calculated by this tool.

Intracellular cytokine staining

For intracellular cytokine staining, cells were cultured with phorbol myristate acetate (50ng/ml, Sigma), ionomycin (500ng/ml, Sigma) and Golgi plug (1µg/ml, eBioscience) for 6 hours, then washed, fixed and permeabilized (FIX & PERM kit, Invitrogen).

RNAseq mouse samples

For murine RNAseq experiments, ITK-SYK$^{CD4\text{-}CreERT2}$ and ITK-SYK$^{CD4\text{-}CreERT2}$;PD1$^{-/-}$ mice were induced with 1 mg tamoxifen per mouse. At day 5 after injection, single-cell suspensions from spleens and lymph nodes were generated and T cells were FACS-sorted for eGFP ex vivo.

Total RNA was extracted from approximately 100,000 cells, using RNeasy Micro Plus Kit (Qiagen). Library preparation from 1 ng total RNA was performed with the SMART-Seq v2 Ultra Low Input RNA Kit (Takara Bio Inc), and SE-50 bp sequencing was performed on Illumina HiSeq2000 machines. Read alignment to the mm10 genome and transcriptome assembly were performed using HISAT2 and StringTie. GSEA was performed with BubbleGUM.

Microarray mouse samples

Global gene expression profiling of FACS-sorted ITK-SYK$^{CD4\text{-}Cre}$ and C57BL/6 wildtype CD4+ T cells was performed on SurePrint G3 Mouse Gene Expression 8x60k microarrays (Agilent Technologies) using 60 ng of total RNA according to the manufacturer's protocol (one-color Low Input Quick Amp Labeling Kit, Agilent Technologies). Raw gene expression data were extracted as text files with the Feature Extraction software 11.0.1.1 (Agilent Technologies). All data analysis was conducted using R (v3.2.2). Data quality assessment, filtering, preprocessing, normalization, batch correction based on nucleic acid labeling batches and data analyses were carried out with the R-packages limma, Agi4x44PreProcess and the ComBat function of the sva R-package.

Analysis of published datasets

For analysis of SSNVs, we included our previously described collection of published CTCL and other T cell lymphoma sequencing studies, detailed in Park et al 2017[10] (Supplemental Tables 1, 2 and 8 from that manuscript). For SCNV analysis, published samples from Choi et al, 2015[8] and McGirt et al, 2015[11] were utilized in conjunction with our in-house samples.

For analysis of CRISPR screening data, we utilized data from Shifrut et al, 2018[12]. Log fold change and *P* values were obtained for each putative CTCL driver gene.

Supplemental Tables

Supplemental Table 1. Samples analyzed by DNA-sequencing.

Supplemental Table 2. Clinical and demographic information of patients studied.

Supplemental Table 3. Samples analyzed by transcriptomic, immunophenotypic, and functional analyses.

Supplemental Table 4. Comparison of whole genome and whole exome sequencing in identification of copy number variants in CTCL.

Supplemental Table 5. Identification of significant deletion and amplifications in CTCL using GISTIC2.0.

Supplemental Table 6. Statistically significant translocation events, point mutations, and damaging mutations in CTCL.

Supplemental Table 7. Published role of novel putative tumor suppressors identified by SCNVs in CTCL.

Supplemental Table 8. Mutational signature profile of CTCL.

Supplemental Table 9. Integration of CTCL driver genes with published CRISPR screening data in primary human T cells.

Supplemental Table 10. List of surface markers conjugated with metal ions in CTLC custom and Immuno-Oncology mass cytometry (CyTOF) panels.

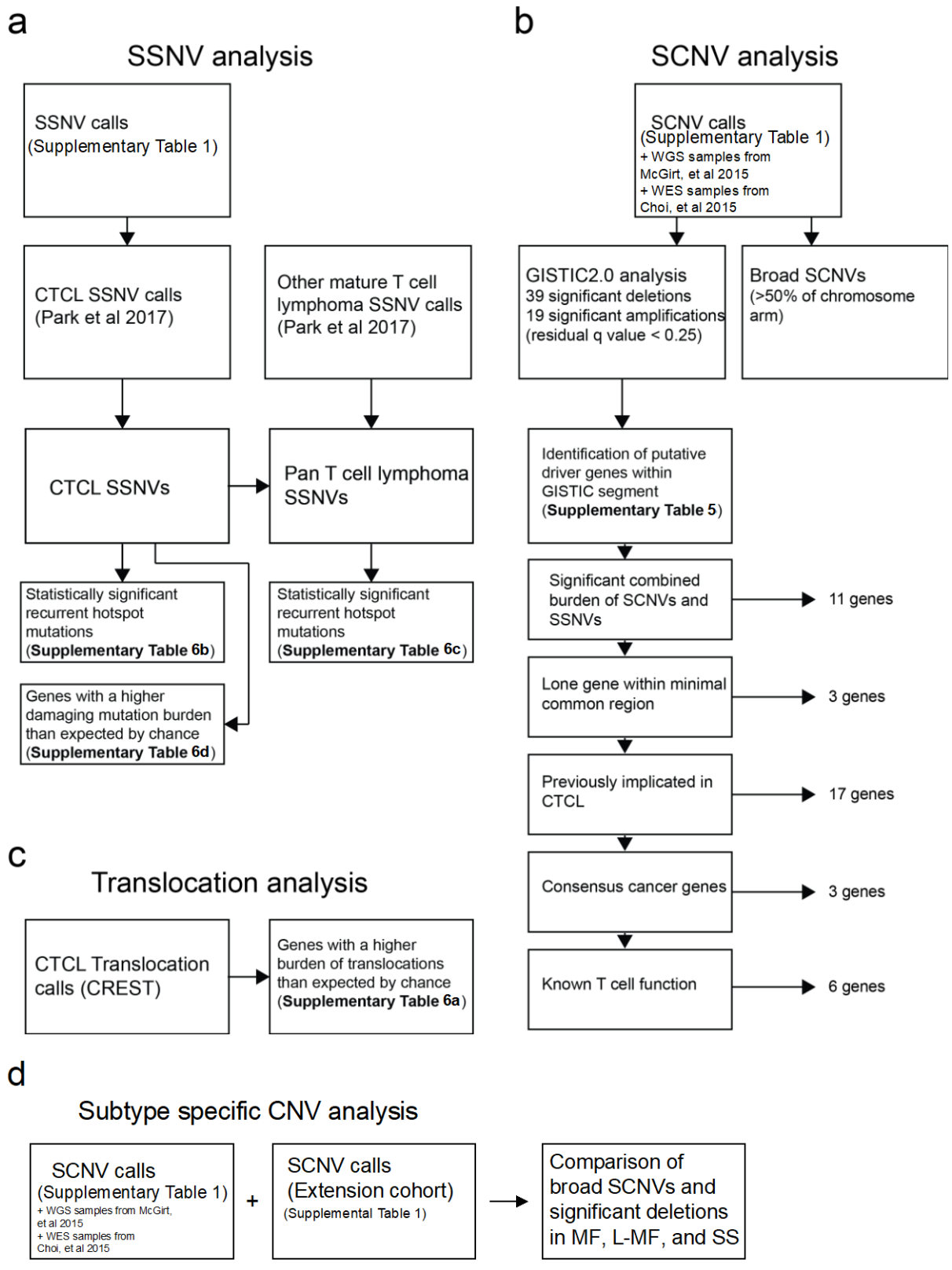
Supplemental Table 11. Differential transcripts between high proliferative CTCL cells and non-proliferative CTCL cells.

Supplemental Table 12. Genetic analysis of high proliferative and non-proliferative CTCLs.

Supplemental Table 13. Differential RNA expression among 3 CTCL groups.

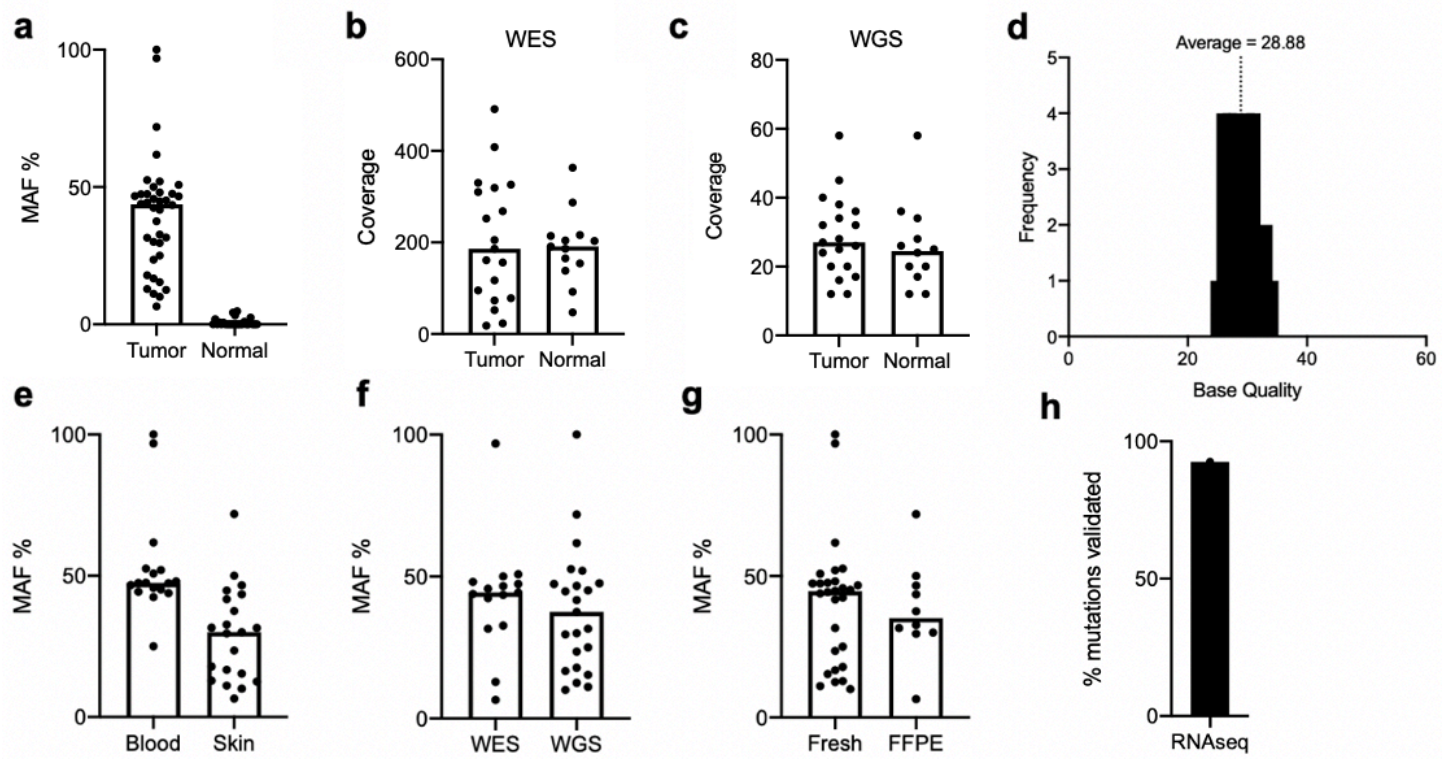
Supplemental Figures

Supplemental Figure 1



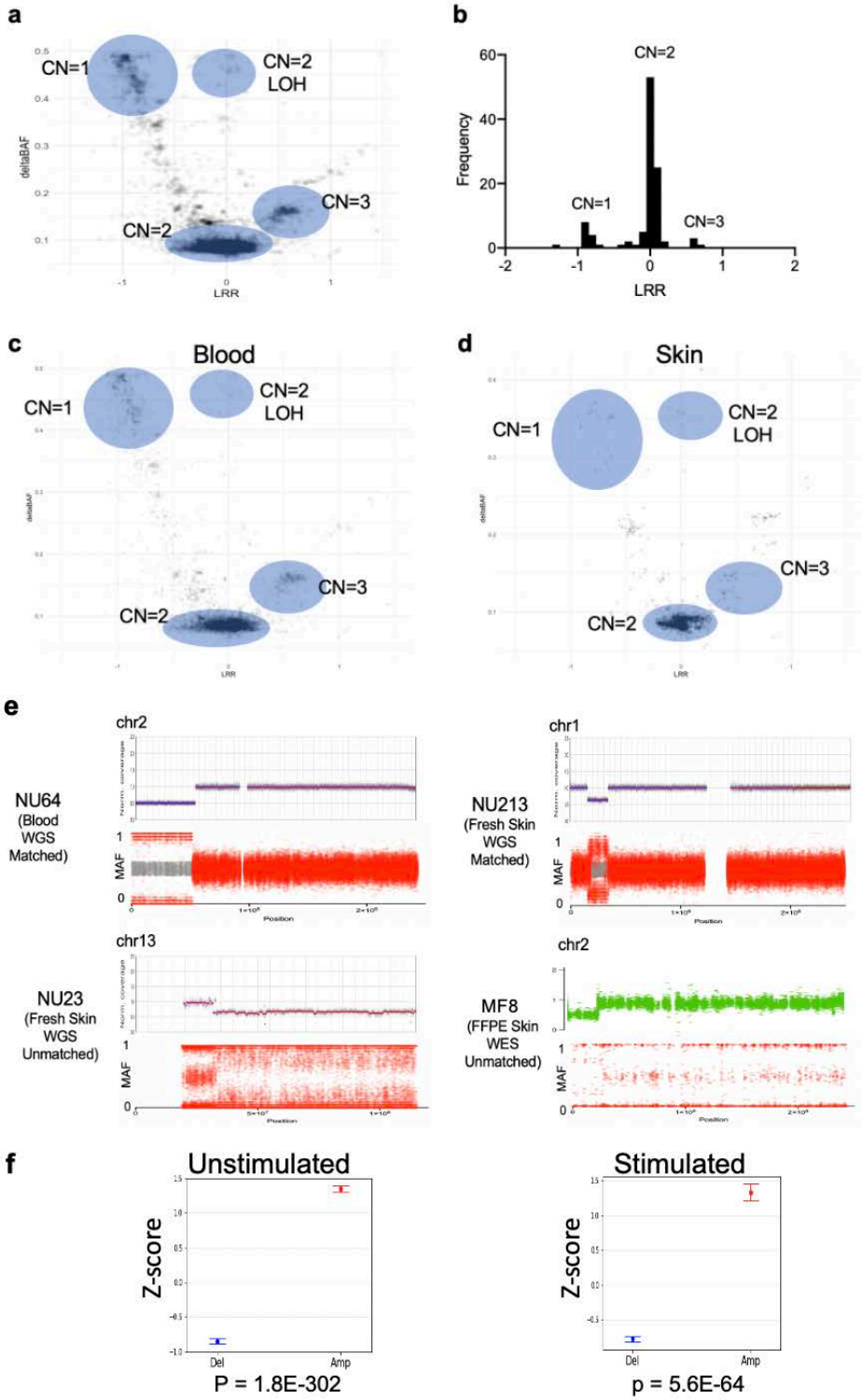
Supplemental Figure 1. Approach to identify putative driver genes and subtype analysis.
a, Approach to identify statistically significant hotspot mutations and genes with a higher burden of damaging mutations than expected by chance in CTCL. Somatic single nucleotide variant (**SSNV**) calls from the present study were combined with SSNV calls from publicly available CTCL datasets (studies described in Park, et al 2017). Then, CTCL SSNVs were assessed for statistically significant hotspot and damaging mutations as previously described using the binomial distribution (see **Supplemental Methods** for further detail). CTCL SSNVs were combined with SSNVs from other mature T cell lymphomas to identify hotspot mutations recurrent across numerous types of T cell lymphoma. **b,** Approach to identify broad somatic copy number variants (**SCNVs**), significant amplifications and deletions as assessed by GISTIC2.0, and putative driver genes within focal significant SCNVs. Broad SCNVs were considered chromosomal arms >50% deleted or >50% amplified. GISTIC2.0 was utilized to identify significant focal deletions and amplifications. A multitiered pipeline was employed to identify putative driver genes residing on significantly deleted or duplicated SCNVs (see **Supplemental Methods** for further detail of statistical methods). Briefly, the probability was calculated that any gene on a GISTIC confidence interval would be subject to an SCNV or SSNV. The P value for the observed vs. expected number of mutations for each gene was then calculated using the binomial distribution, adjusting for gene expression and gene length. Genes were considered significant if they had a >5-fold lower P-value compared to the neighbors. For one peak only a single gene was identified within the peak (*ATXN1*). The minimal common region was defined was the region affected by the highest number of focal SCNVs within each peak. Analysis of the literature was performed to identify genes implicated in CTCL previously, consensus tumor suppressors and oncogenes (COSMIC), and genes with known T cell function. Number of genes indicates how many putative driver genes were identified at each step of the algorithm. **c,** Strategy to identify genes with a higher burden than expected by chance of chromosomal translocations. See Supplemental Methods for further detail of statistical approach. **d,** Approach to analysis of subtype-specific mutations. SCNV calls were combined with SCNV calls from extension cohort (16 FFPE samples, 11 WGS and 5 WES samples with GATK QC score <0.08). Broad SCNVs and significant deletions in MF, L-MF, and SS were then compared. For all analyses, further details of the samples utilized are included in Supplemental Table 1.

Supplemental Figure 2



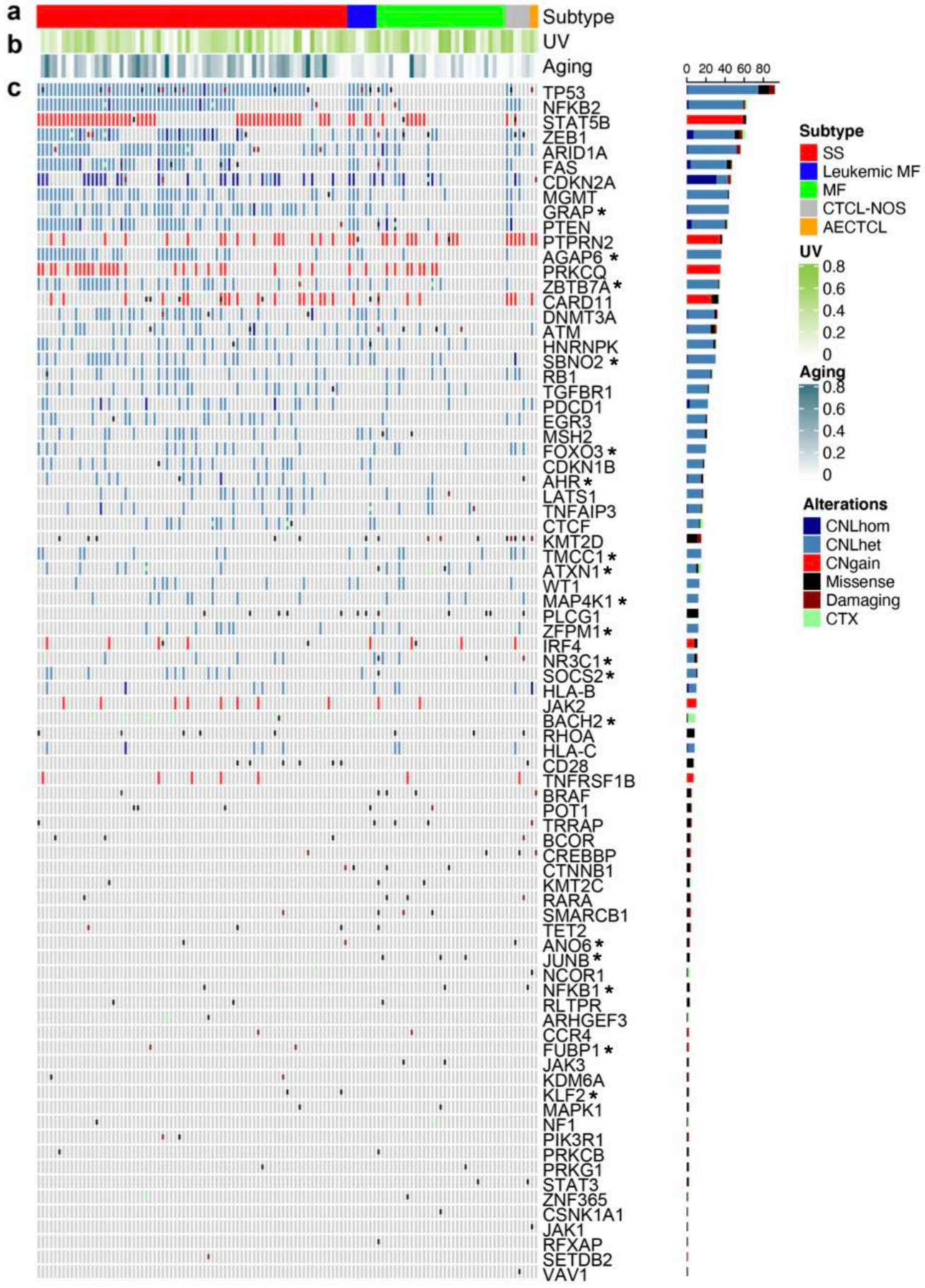
Supplemental Figure 2. Quality control of SSNVs. a, Minor allele frequencies (**MAF**) of putative driver gene mutations identified in CTCL in tumor and normal. **b-c**, Coverage depth of putative driver gene mutations in WES (**b**) and WGS (**c**) in tumor and normal. **d**, Histogram of base quality for CTCL putative driver mutations. **e-g**, MAFs of putative drive mutations across different tissues sampled (**e**), sequencing modalities (**f**), and sample preparations (**g**). **h**, Percentage of mutations identified by DNAseq that were validated by RNAseq in samples with matched DNA/RNAseq available, for mutations covered at least 25x in both DNA and RNAseq.

Supplemental Figure 3



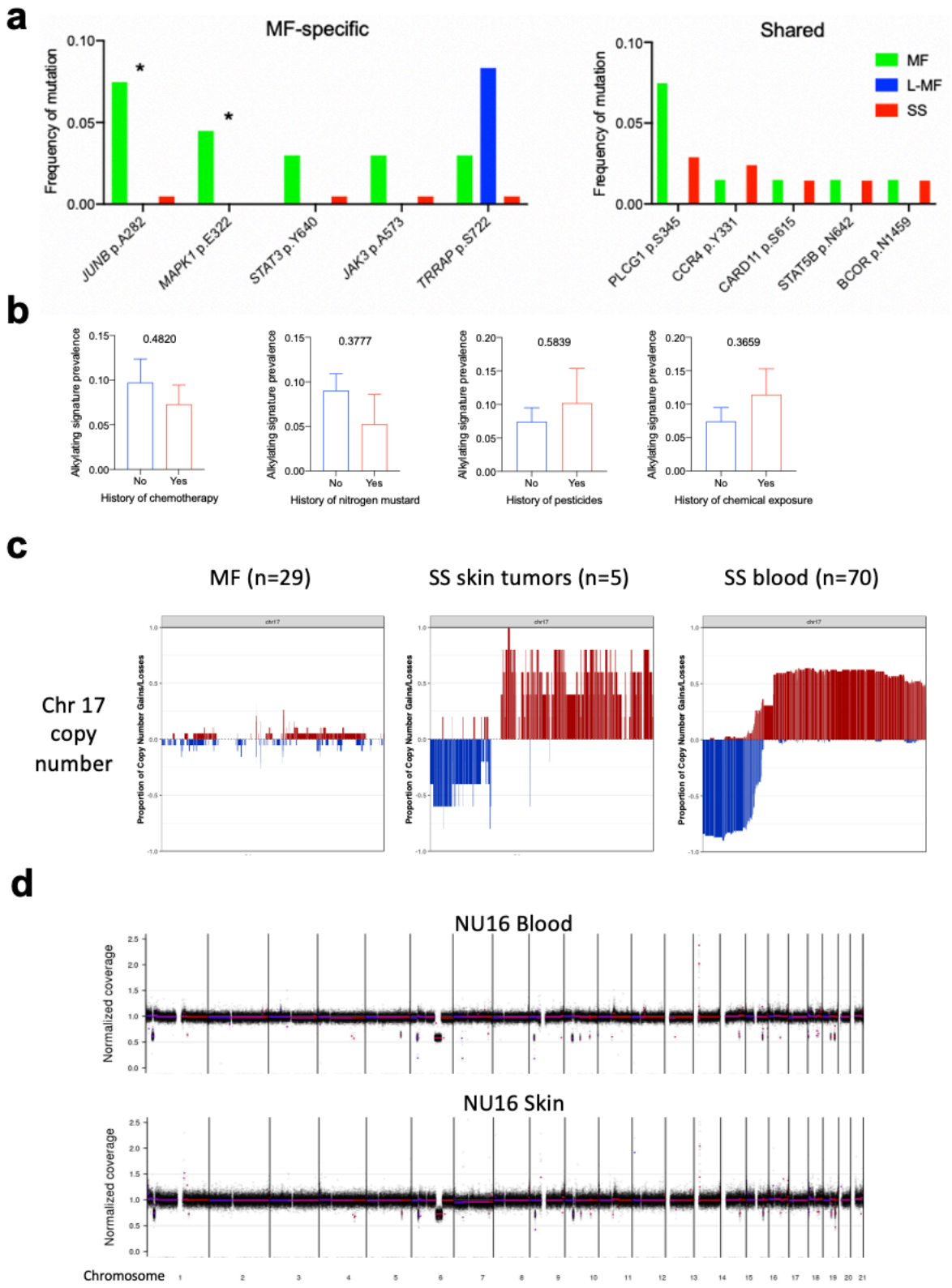
Supplemental Figure 3. Quality control of SCNVs. a, Plot of delta B allele frequency (BAF) versus log2 read ratio (LRR) reveals concordance of allelic imbalance and coverage ratio changes across samples to identify copy number (CN) states. **b**, Histogram of LRR across CTCL samples. **c-d**, Plots of BAF versus LRR in samples isolated from blood and skin. **e**, Representative plots of normalized coverage and MAF across a chromosome positions observed across samples from diverse sources of tissue, sequencing modalities, and with or without matched normal samples. In each case, regions of allelic imbalance correspond to reduced coverage. **f**, Z-scores of RNA-seq expression levels for genes deleted (blue) or amplified (red) in stimulated and unstimulated conditions. As expected, deleted genes were significantly under-expressed and amplified genes significantly over-expressed compared to diploid samples. P values calculated by Welch's one-sided T test.

Supplemental Figure 4



Supplemental Fig. 4. **Mutational landscape of CTCL**. a, Diagnosis for each patient. b, Fraction of enrichment of each COSMIC point mutation signatures in each sample. c, Matrix highlighting copy number and point mutations identified in putative CTCL driver genes. Rows and columns represent genes and samples, respectively. Bar plot on the right represents percent of samples with an alteration in each gene. CNLhom, CNLhet; homozygous and heterozygous copy number loss, respectively. CNgain; copy number gain, CTX; chromosomal translocation, CTCL-NOS; cutaneous T cell lymphoma not otherwise specified, AECTCL; aggressive epidermotropic cytotoxic T cell lymphoma. * indicates novel putative driver gene.

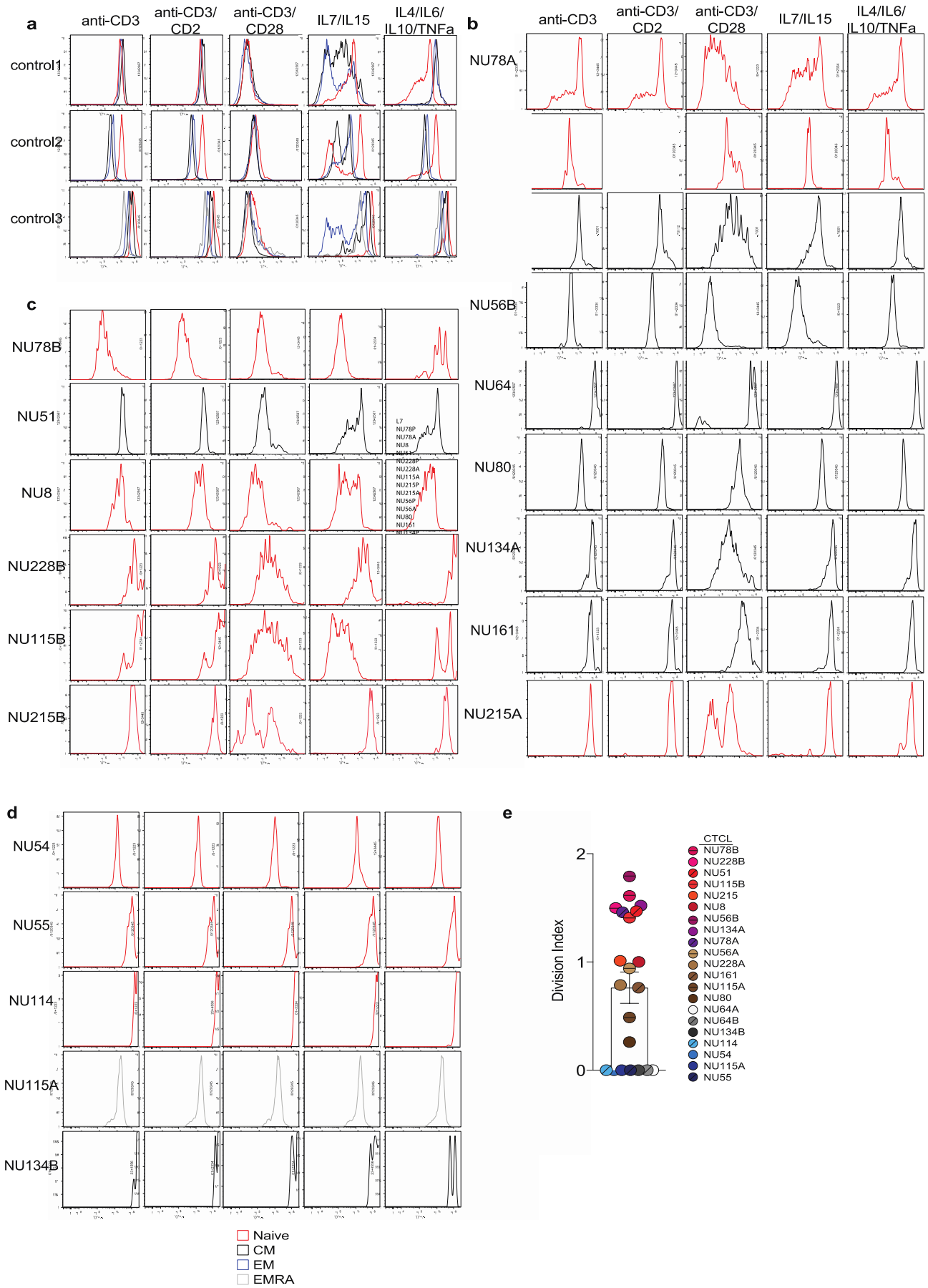
Supplemental Figure 5



Supplemental Figure 5. Genetic alterations in CTCL samples by clinical subtype. a, Frequency of recurrent hotspot mutations. Shared indicates mutations present in >1% of both MF and SS; MF-specific indicates mutations >5 fold enriched in MF. * indicates P value < 0.05, Fisher's exact test comparing MF and SS. **b**, Prevalence of alkylating signature to MF patients with indicated history of exposure. *P* values determined by unpaired t test. **c**, Profile of copy number gains/losses in chromosome 17 demonstrate an absence of deletion or amplification in skin-limited MF samples and frequent 17p deletions and 17q

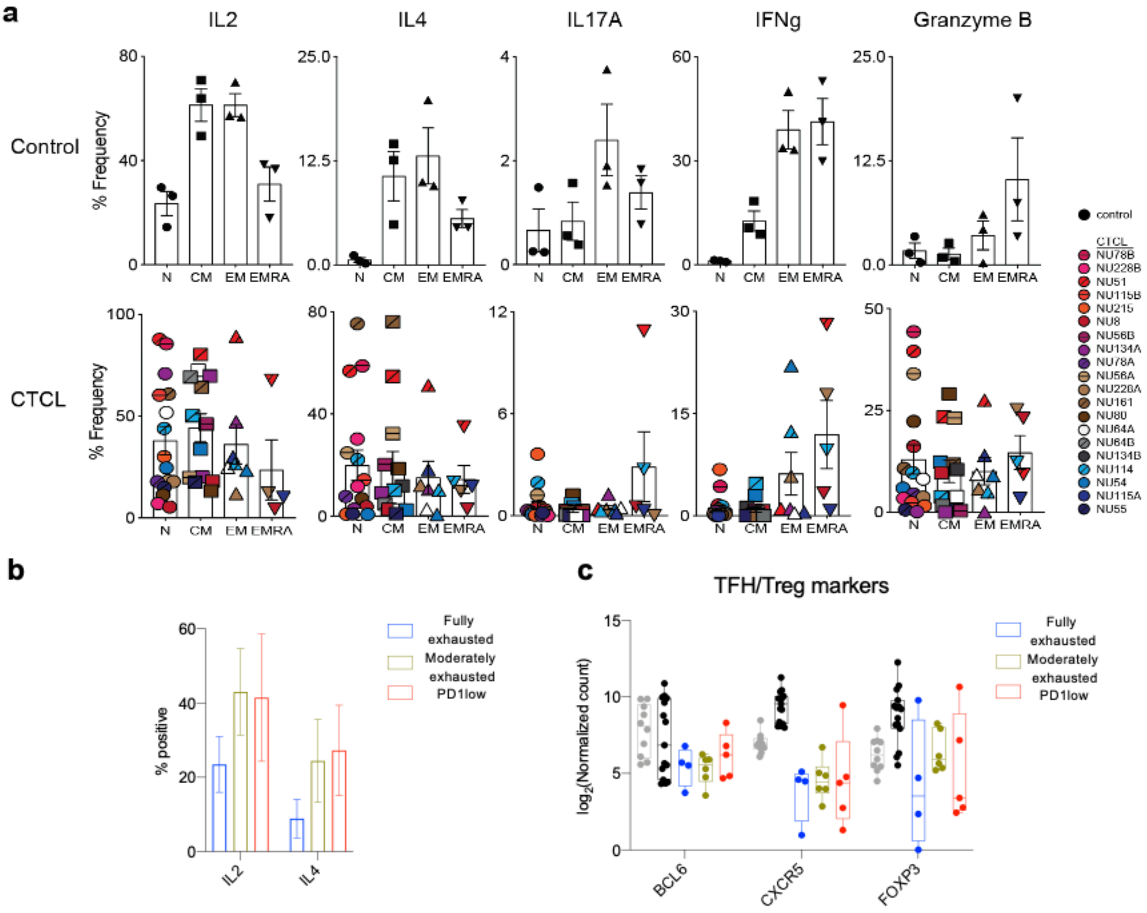
amplifications frequent in patients with leukemic disease regardless of sampling from skin or blood. **d**, Copy number profile of patient NU16 in blood and skin. For this patient with whole-genome sequencing of blood and skin from two different time points, copy number profiles across the genome are similar regardless of site of sampling.

Supplemental Figure 6



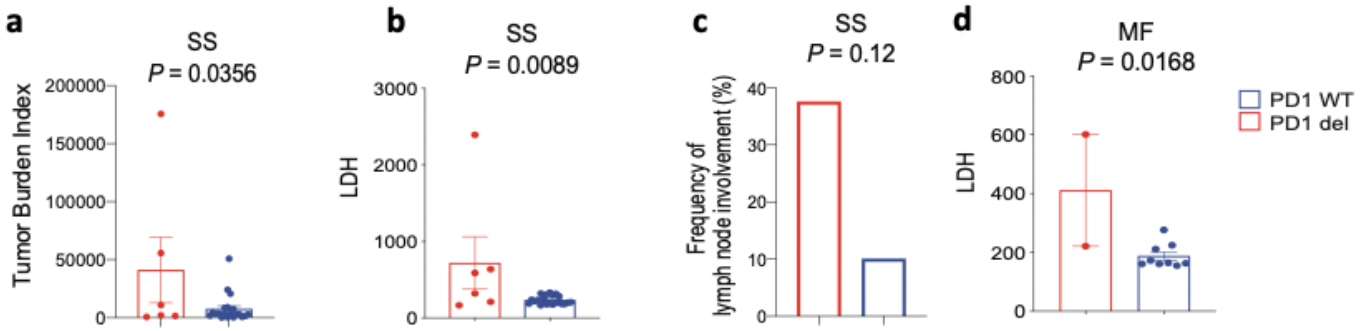
Supplemental Figure 6. Characterization of ex vivo CTCL functions. **a**, Subset specific proliferation of naïve (**N**, CD45RA+CCR7+), central memory (**CM**, CD45RA-CCR7+), effector memory (**EM**, CD45RA-CCR7-) and terminally differentiated effector memory (**EMRA**, CD45RA+CCR7-) of CD3+CD4+ T cells from 3 healthy controls under 5 different stimuli. In parentheses represent the cell types expected to proliferate to each condition. 1.anti-CD3 only (none), 2.anti-CD3/CD2 (memory), 3.anti-CD3/CD28 (both), 4.IL7/IL15 (memory cells only) 5.IL4/IL6/IL10/TNFα (naïve cells only). Concentrations were as follows: 1.anti-CD3 (10μg/ml, Biolegend), 2.anti-CD3 (10μg/ml, Biolegend)/ anti-CD2 (10μg/ml, Biolegend), 3.anti-CD3/CD28 (1:1 ratio, Dynabeads, Life Technologies), 4.IL7 (50ng/ml, Biolegend)/ IL15 (200ng/ml, Biolegend), 5.IL4 (100ng/ml, Biolegend)/ IL6 (50ng/ml, Biolegend)/ IL10 (50ng/ml, Biolegend)/ TNFα (20ng/ml, Biolegend). **b**-**d**, *Ex vivo* proliferation of malignant CTCLs subsets (CD3+CD4+Vβ+ or CD3+CD4+CD26-, see details in **Supplemental Methods**). Patients that have two different time points of collection dates (NU78, NU64, NU115, NU134, NU215, and NU228) were specified as A for first collection date and B for second collection date. For each sample, the predominant immune subset (See **Supplemental Table 3**) is shown. High proliferative CTCL cells (**c**) showed proliferation upon 3 or more stimuli, intermediate proliferative CTCL cells (**b**) showed proliferation upon 1 or 2 stimuli, and no proliferative CTCL cells (**d**) showed no proliferation upon any stimulus. **e**, Quantification of proliferative activity in FlowJo (Tree Star) using division index. Division index of major subsets of each CTCL samples upon anti-CD3/CD28 are presented as bar graph.

Supplemental Figure 7



Supplemental Figure 7. Characterization of T cell phenotypes in CTCL cells. a, Effector cytokine production following 6 hours of stimulation with phorbol myristate acetate (PMA) and ionomycin from CTCL and control cells. Flow cytometry data represent mean % frequency of positive population ± s.e.m. **b**, IL2 and IL4 production according to CTCL subgroup. Data represent mean % frequency of positive population ± s.e.m. **c**, Markers of T follicular helper (TFH) and regulatory T cells (Treg) in CTCL subgroups.

Supplemental Figure 8.



Supplemental Figure 8. PD1 deletions are associated with adverse prognostic factors. a,b Adverse prognostic factors related to disease severity in SS presented according to PD1 genotype. Data represent mean ± s.e.m of tumor burden index at peak (see **Supplemental Methods**, **a**) and lactate dehydrogenase (**LDH**) at diagnosis (**b**). *P* values are from two-tailed unpaired Student's t-test. PD1 WT; PD1-wild-type patients, PD1 del; PD1-deleted patients. **c**, Percentage of patients with non-skin draining lymph node involvement in SS according to PD1 genotype. Data represent mean of frequency and *P* value is from Fisher's exact test. **d**, LDH at diagnosis according to PD1 genotype in MF. Data represent mean ± s.e.m and *P* value is from two-tailed unpaired Student's t-test.

Supplemental References

1. Horna P, Moscinski LC, Sokol L, Shao H. Naive/Memory T-Cell Phenotypes in Leukemic Cutaneous T-Cell Lymphoma: Putative Cell of Origin Overlaps Disease Classification. *Cytometry B Clin Cytom*. 2018.
2. Roelens M, Delord M, Ram-Wolff C, et al. Circulating and skin-derived Sezary cells: clonal but with phenotypic plasticity. *Blood*. 2017;130(12):1468-1471.
3. Mayrhofer M, Dilorenzo S, Isaksson A. Patchwork: allele-specific copy number analysis of whole-genome sequenced tumor tissue. 2013;14(3):R24.
4. Karczewski KJ, Francioli LC, Tiao G, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*. 2019:531210.
5. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*. 2010;20(9):1297-1303.
6. 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.
7. Wang J, Mullighan CG, Easton J, et al. CREST maps somatic structural variation in cancer genomes with base-pair resolution. 2011;8(8):652-654.
8. Choi J, Goh G, Walradt T, et al. Genomic landscape of cutaneous T cell lymphoma. *Nat Genet*. 2015;47(9):1011-1019.
9. Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013;499(7457):214-218.
10. 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.
11. McGirt LY, Jia P, Baerenwald DA, et al. Whole-genome sequencing reveals oncogenic mutations in mycosis fungoides. *Blood*. 2015;126(4):508-519.
12. 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 e1915.