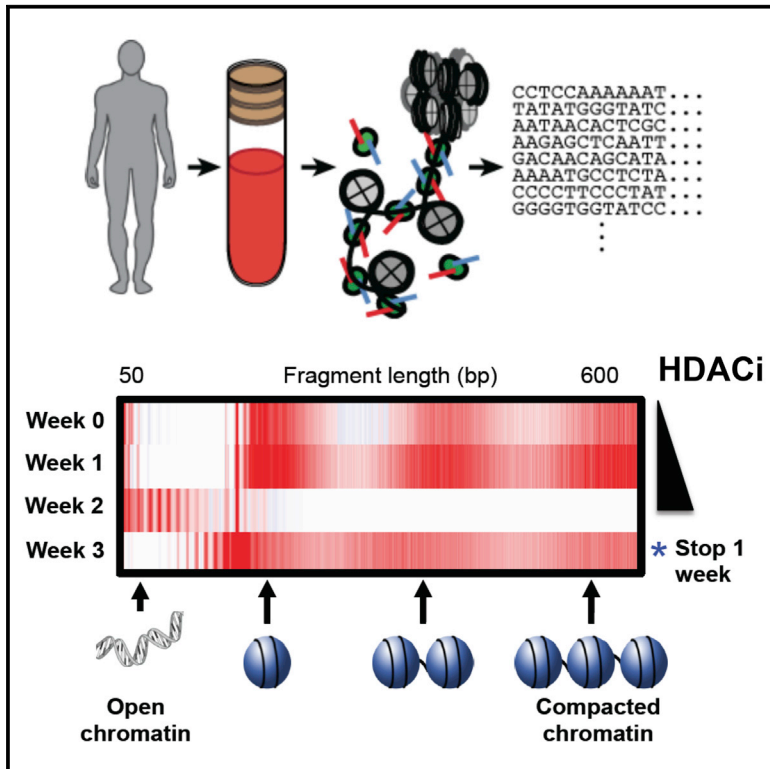


# Chromatin Accessibility Landscape of Cutaneous T Cell Lymphoma and Dynamic Response to HDAC Inhibitors

## Graphical Abstract



## Authors

Kun Qu, Lisa C. Zaba,  
Ansuman T. Satpathy, ...,  
William J. Greenleaf, Youn H. Kim,  
Howard Y. Chang

## Correspondence

younkim@stanford.edu (Y.H.K.),  
howchang@stanford.edu (H.Y.C.)

## In Brief

Qu et al. show that the accessible chromatin landscape distinguishes leukemic from host T cells in cutaneous T cell lymphoma (CTCL) patients as well as T cells from healthy individuals. The clinical response of CTCL to HDAC inhibitors strongly associates with a concurrent gain in chromatin accessibility.

## Highlights

- Regulome signatures distinguish CTCL leukemic, CTCL host, and normal CD4<sup>+</sup> T cells
- CTCL regulomes show three dominant patterns of transcription factor activity
- Specific chromatin dynamics are associated with clinical response to HDACi
- HDACi accentuates pre-existing DNA access in CTCL leukemic and host CD4<sup>+</sup> T cells

# Chromatin Accessibility Landscape of Cutaneous T Cell Lymphoma and Dynamic Response to HDAC Inhibitors

Kun Qu,<sup>1,2,9</sup> Lisa C. Zaba,<sup>2,3,9</sup> Ansuman T. Satpathy,<sup>2,3</sup> Paul G. Giresi,<sup>4</sup> Rui Li,<sup>2</sup> Yonghao Jin,<sup>1</sup> Randall Armstrong,<sup>5</sup> Chen Jin,<sup>1</sup> Nathalie Schmitt,<sup>6</sup> Ziba Rahbar,<sup>3</sup> Hideki Ueno,<sup>7</sup> William J. Greenleaf,<sup>2,8</sup> Youn H. Kim,<sup>3,\*</sup> and Howard Y. Chang<sup>2,3,10,\*</sup>

<sup>1</sup>CAS Key Laboratory of Innate Immunity and Chronic Diseases, School of Life Sciences and Medical Center, University of Science and Technology of China, Hefei 230027, China

<sup>2</sup>Center for Personal Dynamic Regulomes and Program in Epithelial Biology, Stanford University School of Medicine, CCSR 2155c, 269 Campus Drive, Stanford, CA 94305-5168, USA

<sup>3</sup>Department of Dermatology, Stanford University School of Medicine, Stanford, CA 94305, USA

<sup>4</sup>Epinomics, Menlo Park, CA 94035, USA

<sup>5</sup>Stanford Blood and Marrow Transplantation Cellular Therapy Facility, Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA

<sup>6</sup>CNRSUMR 5164, Université de Bordeaux, Bordeaux 33076, France

<sup>7</sup>Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>8</sup>Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA

<sup>9</sup>These authors contributed equally

<sup>10</sup>Lead Contact

\*Correspondence: [younkim@stanford.edu](mailto:younkim@stanford.edu) (Y.H.K.), [howchang@stanford.edu](mailto:howchang@stanford.edu) (H.Y.C.)

<http://dx.doi.org/10.1016/j.ccell.2017.05.008>

## SUMMARY

Here, we define the landscape and dynamics of active regulatory DNA in cutaneous T cell lymphoma (CTCL) by ATAC-seq. Analysis of 111 human CTCL and control samples revealed extensive chromatin signatures that distinguished leukemic, host, and normal CD4<sup>+</sup> T cells. We identify three dominant patterns of transcription factor (TF) activation that drive leukemia regulomes, as well as TF deactivations that alter host T cells in CTCL patients. Clinical response to histone deacetylase inhibitors (HDACi) is strongly associated with a concurrent gain in chromatin accessibility. HDACi causes distinct chromatin responses in leukemic and host CD4<sup>+</sup> T cells, reprogramming host T cells toward normalcy. These results provide a foundational framework to study personal regulomes in human cancer and epigenetic therapy.

## INTRODUCTION

Cutaneous T cell lymphoma (CTCL) is a heterogeneous group of T cell neoplasms with primary involvement of the skin. Mycosis fungoides (MF) and Sézary syndrome (SS) constitute the majority of CTCLs and are believed to originate from skin-tropic mature CD4<sup>+</sup> T cells (Willemze et al., 2005). In the early stages, patients often have skin-restricted disease and, in advanced stages of MF, the malignant T cells can involve the lymph node, viscera, and/or blood compartments. SS is the leukemic subtype of

CTCL where patients present with generalized skin erythema. CTCL is the first clinical indication approved by the US Food and Drug Administration for treatment with histone deacetylase inhibitors (HDACi), such as vorinostat and romidepsin, highlighting the power of therapies that target the epigenome (New et al., 2012; Rodríguez-Paredes and Esteller, 2011). However, only a subset of CTCL patients (30%–35%) respond to HDACi, and molecular and predictive biomarkers of clinical response to HDACi are needed. Despite CTCL being the first disease targeted by HDACi therapy, the landscape of the CTCL epigenome

### Significance

Cutaneous T cell lymphoma (CTCL) is a heterogeneous group of T cell neoplasms, and is the first FDA-approved indication for epigenetic treatment with HDAC inhibitors (HDACi). However, the landscape of the CTCL epigenome and its dynamic response to therapy *in vivo* are not known. Here we characterized the accessible chromatin landscapes in clinical CTCL samples and their dynamic response to HDACi. The pattern of accessible chromatin can predict clinical response to HDACi, while DNA mutations do not. Regulome profiling in cancer may thus provide unique prognostic and mechanistic insights.

in vivo and its response to therapy are not known. Moreover, it is appreciated that CTCL comprises a complex interplay between malignant T cells and the host immune system. The way in which CTCL reprograms host immunity and potential dynamic response of these interacting systems to therapy are unclear. Systematic analysis of the epigenomic landscape from primary clinical samples is needed to address these issues.

Assay of transposase-accessible chromatin with sequencing (ATAC-seq) is a recently introduced and sensitive method to map open chromatin sites, predict transcription factor (TF) binding, and determine nucleosome position from as few as 500 cells (Buenrostro et al., 2013; Lara-Astiaso et al., 2014; Lavin et al., 2014), or even in single cells (Buenrostro et al., 2015; Cusanovich et al., 2015). This technology enables clinicians to track the epigenomic state of patient-derived samples in real time and affords a “personal regulome”; a summary of gene-regulatory events in a snapshot of time within a single individual (Qu et al., 2015). In this study, we developed a systematic approach to characterize chromatin dynamics in CTCL using ATAC-seq, and addressed the regulatory dynamics in leukemic epigenomes from CTCL patients treated with HDACi.

## RESULTS

### Landscape of DNA Accessibility in Normal CD4<sup>+</sup>, CTCL Leukemia, and Host T Cells

We generated and analyzed 111 high-resolution personal regulomes, 81 from 14 patients with CTCL and 30 from 10 healthy donors, of a single cell type (human CD4<sup>+</sup> T cells) that comprised over 6 billion measurements (Figure 1A; Table S1). We interrogated the landscapes of chromatin accessibility in these samples and developed methods to integrate diverse sources of genomic and epigenomic information to address the regulatory dynamics in leukemic epigenomes from CTCL patients treated with HDACi (Figure 1A). Thirteen of 14 patients had S (stage IV, significant leukemic T cells); 1 patient had stage III MF, where the disease was not blood predominant (Table S2). Because MF/SS are typically characterized by a dominant CD4<sup>+</sup> T cell clone bearing a unique T cell receptor, we purified leukemic T cells from patients (defined by CD4<sup>+</sup>, CD26<sup>−</sup>, and T cell receptor V-beta clone<sup>+</sup>) versus non-leukemic host CD4<sup>+</sup> T cell (defined by CD4<sup>+</sup>, V-beta clone<sup>−</sup>) from the same patients by fluorescence-activated cell sorting (FACS) (Figure 1B). Thereafter, we refer to the non-malignant CD4<sup>+</sup> T cells from CTCL patients as “host T cells.” Bulk CD4<sup>+</sup> T cells were also obtained using RosetteSep Human CD4<sup>+</sup> T Cell Enrichment Cocktail. Leukemic, host and bulk T cells were obtained from 9 out of 14 patients who had detectable V-beta clone, and only bulk T cells were obtained for the remaining 5 patients without detectable V-beta clone. Although number and proportion of leukemic and host T cells varies depending on the stage and drug response of each individual, we were able to obtain at least 50,000 CD4<sup>+</sup> T cells per sample (Figure 1B). To provide additional comparative framework, we also analyzed 30 longitudinally collected ATAC-seq profiles of CD4<sup>+</sup> T cells obtained from 10 healthy donors (Qu et al., 2015).

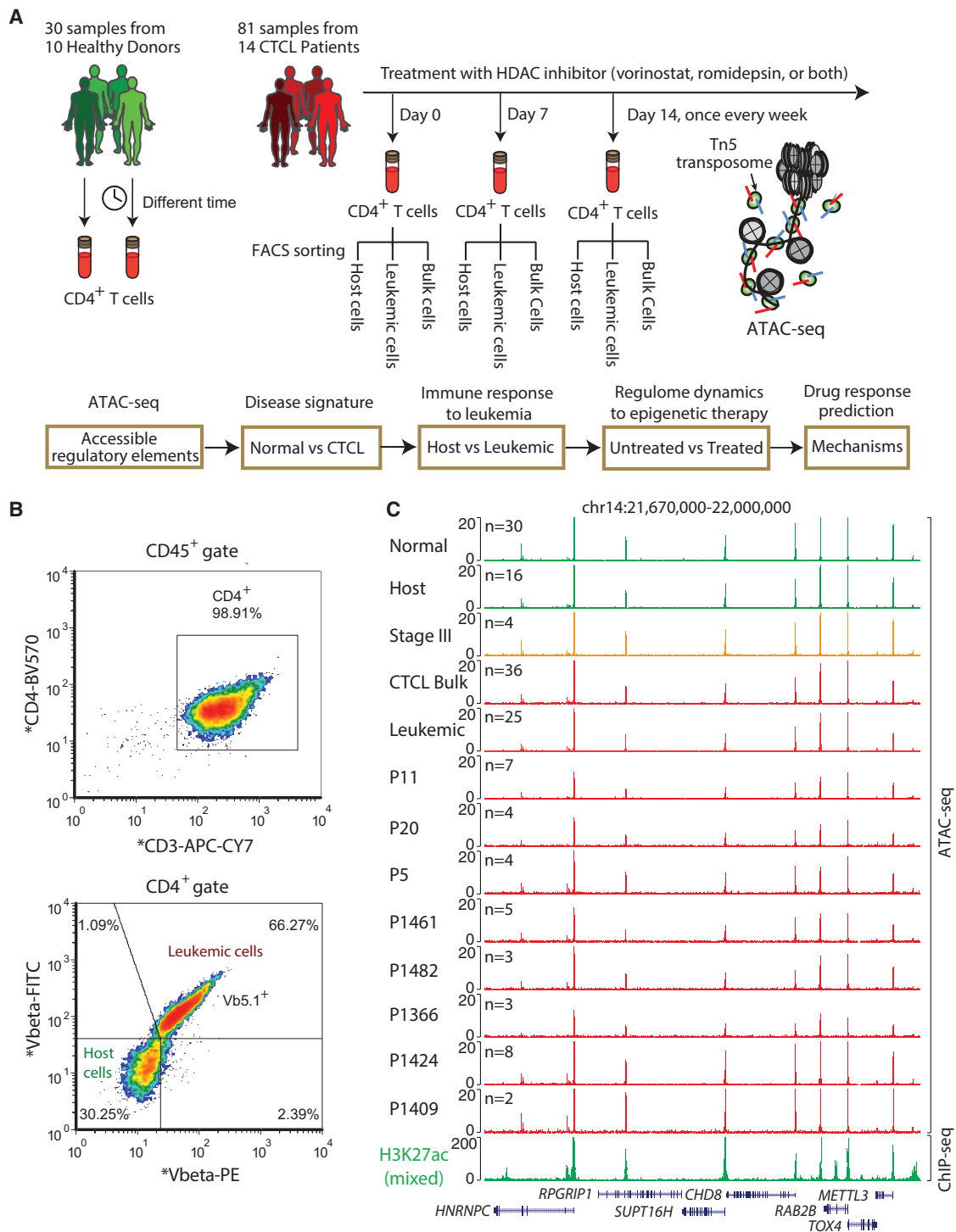
We performed ATAC-seq to map the location and accessibility of regulatory elements genome wide. Each library was sequenced to obtain, on average, more than 55 million paired-

end reads (Table S1). With this dataset we identified a total of 71,464 peaks of DNA accessibility; visual inspection and quantitative analysis indicate that these data are of high quality with a strong signal to background ratio (Figure 1C). ATAC-seq signal in multiple patients aligned very well with the active enhancer mark histone H3K27ac (Spearman correlation = 0.68–0.77,  $p < 10^{-10}$ ) (Limbach et al., 2016), indicating that detected regions of DNA are open and accessible. Correlation and clustering analysis of all the accessible sites recapitulate the group classification of the clinical samples, separating healthy donors, CTCL bulk, and leukemic and host cells (Figure S1). Interestingly, samples from the one patient with stage III MF clustered together with those from normal donors, in agreement with the clinical stage classification where malignant T cells in stage III MF are located predominantly in the skin or lymph node, but not in the peripheral blood. Non-leukemic host CD4<sup>+</sup> T cell samples also clustered together adjacent to normal T cell samples from healthy donors. These results demonstrate the feasibility of obtaining high-quality epigenomic data from primary cancer samples using standard clinical infrastructure.

### Epigenomic Signatures of CTCL Leukemic and Non-leukemic Host Cells

To identify differences in regulatory DNA activity among leukemic, host, bulk, and normal cells, we applied pairwise comparison of the corresponding samples using DESeq. We discovered 7,498 elements of differential DNA accessibility across the genome (Figure 2A). Known sites related to inter-individual variability among healthy donors (Qu et al., 2015) were removed from further consideration. Unsupervised hierarchical clustering of the differential peaks reveals three distinct clusters of regulatory elements, suggesting potential normal and leukemic epigenetic signatures that distinguish healthy, leukemic, and host T cells in CTCL patients.

We used GREAT to assess the genomic features that are enriched in each cluster of peaks (Figure S2). Cluster I is comprised of 1,995 elements that are more accessible in normal and host CD4<sup>+</sup> cells compared with bulk and purified CTCL leukemic cells; these elements may reflect the normal epigenomic signature of T cell homeostasis that is lost in the malignant T cells. GREAT analysis revealed that elements in this cluster are highly enriched in immune-related gene ontology functions, immune system morphology, and hematopoietic system diseases (Figure S2) (false discovery rate [FDR] < 0.05 for each), suggesting that they are critical for proper function of human T cells. Several signaling pathways are enriched in cluster I peaks such as interleukin-23 (IL-23)-mediated ( $p = 6.1 \times 10^{-12}$ ), nuclear factor of activated T cell-dependent ( $p = 5.0E-10$ ) and platelet-derived growth factor ( $p = 6.1E-10$ ) pathways, which have been reported to driving hematopoietic cancers including CTCL (da Silva Almeida et al., 2015). In addition, IL-2 signaling pathway events mediated by phosphatidylinositol 3-kinase (PI3K) ( $p = 2.3 \times 10^{-8}$ ) were also found to be significant, consistent with prior transcriptome analysis of SS, which identified PI3K/AKT as the top dysregulated signaling pathway (Lee et al., 2012). Two intragenic elements in *PIK3R1*, encoding the regulatory subunit of PI3K, are upregulated in healthy donor and non-leukemic host cells, and repressed in CTCL bulk and leukemic samples, exemplifying elements in cluster I (Figure 2B).

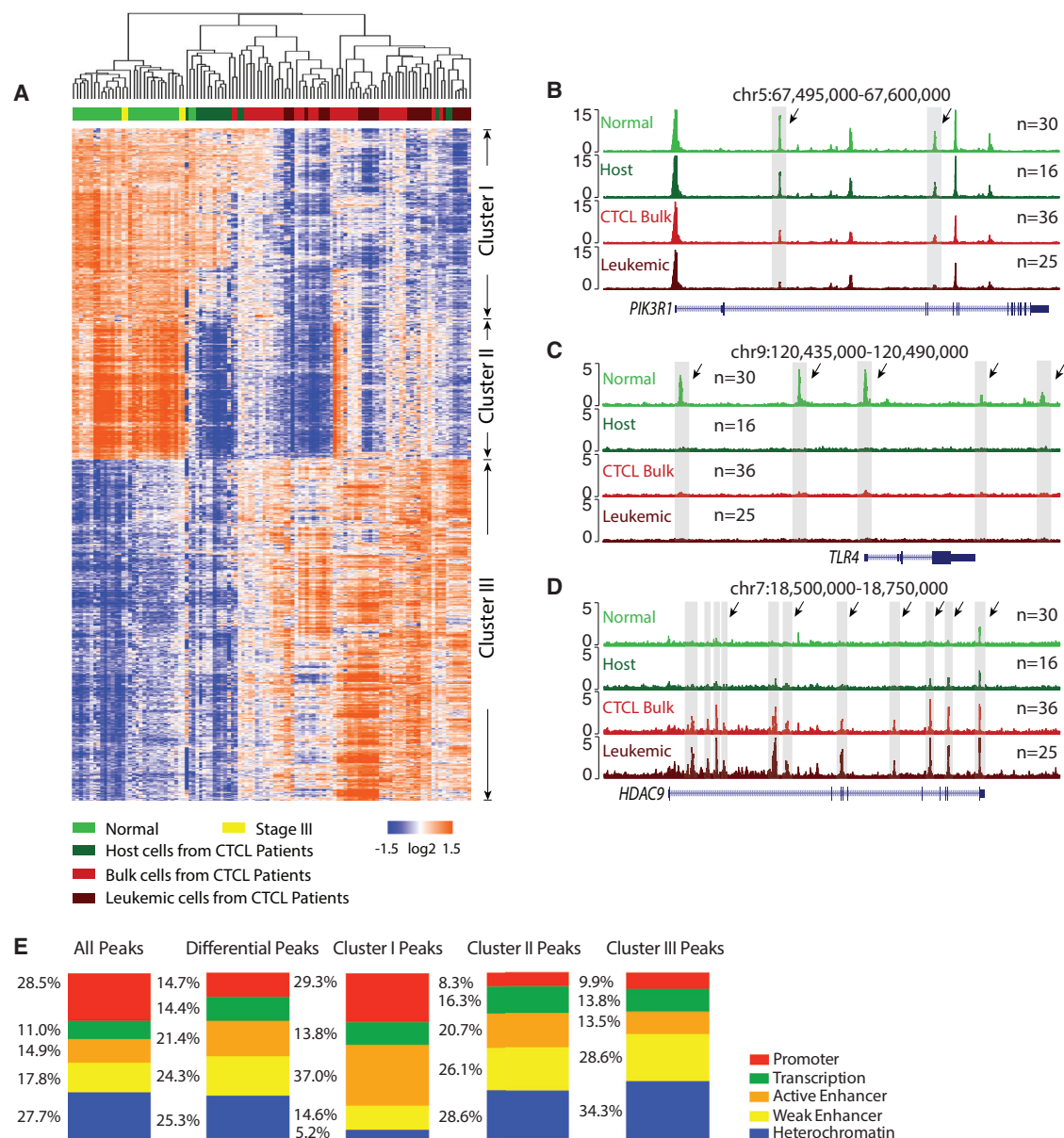


**Figure 1. Landscape of DNA Accessibility in Normal CD4<sup>+</sup>, CTCL, Leukemic, and Host Cells**

(A) Schematic outline of the study design shows 30 samples from 10 healthy donors and 81 samples from 14 CTCL patients under HDACi therapy (top), with a bioinformatics pipeline for data analysis (bottom). Host and leukemic cells were isolated from FACS based on CD3<sup>+</sup>CD4<sup>+</sup>Vβ<sup>-</sup> and CD3<sup>+</sup>CD4<sup>+</sup>Vβ<sup>+</sup>, respectively. ATAC-seq was then performed on normal CD4<sup>+</sup>, CTCL, leukemic, and host cells.

(B) CD3<sup>+</sup>, CD4<sup>+</sup>, leukemic (Vβ<sup>+</sup>), and non-leukemic host (Vβ<sup>-</sup>) T cells sorted from FACS.

(C) Normalized ATAC-seq profiles at a locus in normal, host, stage III, bulk CTCL, and leukemic samples and bulk cells from individual patients, together with normalized H2K27ac chromatin immunoprecipitation sequencing profile. See also [Figure S1](#); [Tables S1](#) and [S2](#).



**Figure 2. Epigenomic Signatures of CTCL Leukemic and Host Cells**

(A) Heatmap of 7,498 regulatory elements with differential activity. Each column is a sample; each row is an element. Samples and elements are organized by 2D unsupervised hierarchical clustering. The color scale shows relative ATAC-seq signals as indicated. Top: samples were categorized into five groups, normal, stage III patient, host cells, bulk cells, and leukemic cells from CTCL patients. Samples from the same group are labeled with the same color.

(B–D) Normalized ATAC-seq profiles at the *PIK3R1* (B), *TLR4* (C), and *HDAC9* (D) locus in normal, host, bulk CTCL, and leukemic cells. Shaded regions are representative peaks in clusters I (B), II (C), and III (D), respectively.

(E) Distribution of genomic features of all, differential, cluster I, II, and III regulatory elements. Five genomic features were studied, promoter, transcription, active enhancer, weak enhancer, and heterochromatin. See also Figures S2 and S3.

Cluster II is comprised of 1,696 elements that are highly accessible in CD4<sup>+</sup> T cells from healthy donors, but are less accessible in both the leukemic and host non-leukemic cells. This behavior suggests that these elements have a disease-specific signature, and identifies elements that are reprogrammed in host immune cells in leukemic patients. Prime examples of elements in cluster II include the promoter and several distal elements of gene *TLR4* (Figure 2C). The protein encoded by this gene is a member of the

Toll-like receptor (TLR) family, which plays a fundamental role in pathogen recognition and activation of innate immunity. Consistently, *TLR4* mRNA is strongly expressed in normal CD4<sup>+</sup> T cells and significantly downregulated in host cells and leukemic cells from CTCL patients (Figure S3A) (Lee et al., 2012; Quinn et al., 2015). Compared with cluster I-enriched features, cluster II elements are enriched for more general immune functions that are not specific to lymphocytes or leukocytes (Figure S2),



suggesting that the loss of accessibilities of peaks in cluster II may cause general immunodeficiency diseases, but not necessarily leukemia, which was regulated primarily by peaks in cluster I. Mouse phenotype and disease ontology analysis show that cluster II peaks are enriched in immune-related diseases, but none of which is specific to leukemic cells in SS (Figure S2). These results further indicate the generalized disease-relevant function of cluster II elements; however, peaks in cluster I maybe the drivers for CTCL specifically.

Cluster III consists of 3,807 elements, which are highly accessible in bulk CTCL and leukemic cells, but not in host CD4<sup>+</sup> T cells and normal cells from healthy donors. Cluster III may thus represent a leukemic signature associated with abnormal growth or differentiation. Gene ontology terms show that peaks in cluster III are highly enriched terms related to cell development and differentiation (Figure S2). Examples of cluster III include multiple elements in the *HDAC9* locus, encoding a histone deacetylase homolog that is a component of co-repressor complexes, which become strongly accessible in bulk and purified CTCL leukemia cells (Figure 2D), suggesting a signature of malignant T cells. HDACs regulate chromatin remodeling and gene expression, as well as the functions of more than 50 TFs and non-histone proteins. Transcriptome analysis shows that, among all the HDACs, *HDAC9* is the only gene with significant differential expression between host and leukemic cells (Figure S3B) (Lee et al., 2012). It has also been reported that *HDAC9* proved particularly important in regulating Foxp3-dependent suppression in T regulatory (Treg) cells, and HDACi therapy in vivo enhanced Treg-mediated suppression of homeostatic proliferation, and decreased inflammatory bowel disease through Treg-dependent effects (Tao et al., 2007). These results suggest the chromatin-accessible elements on *HDAC9* may play a significant role in driving disease progression in this T cell malignancy.

We next checked the genomic distribution of all differential and cluster peaks by overlapping each peak list with regions of features in T cells defined in the Epigenomic Roadmap. Promoter and active enhancers are highly enriched in cluster I peaks, suggesting that these peaks may have a greater influence on gene expression (Figure 2E).

### Change in DNA Accessibility Is Correlated with Differential Messenger RNA Expression in CTCL

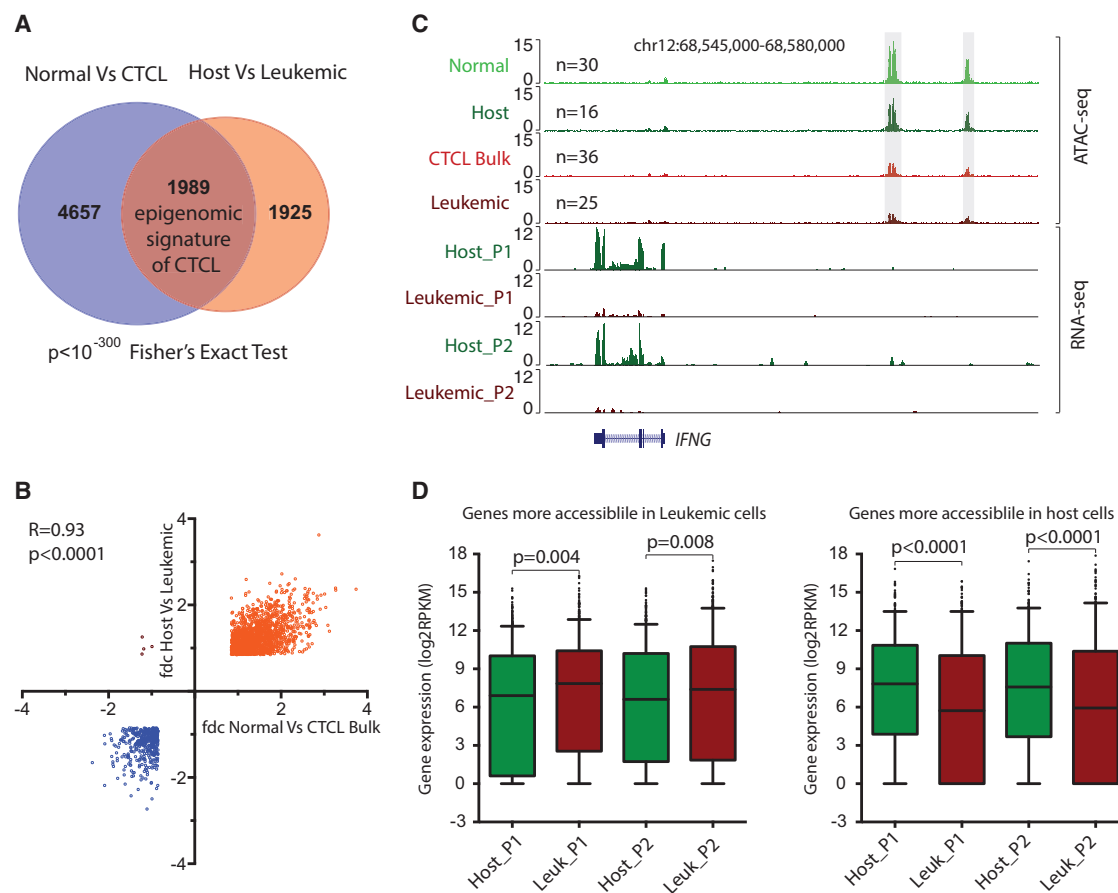
We next examined whether the normal and leukemic chromatin signatures in DNA accessibility correlate with that of gene expression. We purified host and leukemic cells from bulk T cells obtained from the CTCL patients using FACS, and performed chromatin accessibility analysis by ATAC-seq, and compared the results with whole-transcriptome analysis by RNA sequencing (RNA-seq) for both cells. Accessible sites that are (1) differential between total CD4<sup>+</sup> cells of healthy donor and CTCL patients, and (2) also coordinately differential between leukemic and host CD4<sup>+</sup> T cells purified from CTCL patients, were defined as the “CTCL signature.” This analysis identified 1,989 elements that represent the epigenomic signature of CTCL (Figures 3A and S3C); 1,557 elements are more accessible in normal and host cells; 432 elements more accessible in bulk CTCL and leukemic cells (Figure 3B). A quantitative comparison of each pair suggests that the fold changes between normal and leukemic epigenomic signature are highly consistent.

The *IFNG* locus emerged as a prime example of the intersection of predicted regulatory divergence in normal and leukemic cells (Figure 3C). *IFNG* encodes interferon gamma (IFN- $\gamma$ ), a soluble cytokine that is secreted by cells of both the innate and adaptive immune systems, and is a key regulator of immune response and T helper 1 (Th1) cell differentiation (Platanias, 2005). IFN- $\gamma$  is currently being used in combination therapy to treat CTCL patients (Jawed et al., 2014). We then asked whether the epigenomic profiling is generally correlated with gene expression. By comparing the genome-wide RNA-seq data in host versus leukemic cells obtained from an independent study (Lee et al., 2012), we found that, on average, gene loci that gain ATAC-seq signal showed a significant increase of gene expression level ( $p < 0.008$ , Student's *t* test); gene loci that lose ATAC-seq signal also had decreased expression ( $p < 0.0001$ , Student's *t* test) (Figure 3D), indicating a high correlation of epigenetic and RNA profiling.

### Transcription Factor Occupancy Networks in CTCL

One of the main advantages of the ATAC-seq is that this technology can potentially inform the transcriptional regulatory network in the disease. Because TF binding to their cognate DNA sequences, termed motifs, often obligates nucleosome eviction and creation of an accessible DNA site, integration of known TF motifs with DNA accessibility data from ATAC-seq can predict a genome-wide regulatory network in any state of interest (Qu et al., 2015). We applied this analytical technique to identify CTCL-specific differences in the gene-regulatory network from ATAC-seq data. We first obtained a total of 242 vertebrate TF motifs from the Jasper database (Mathelier et al., 2016), identified their genome-wide distribution using HOMER, and overlaid these sites with the differential ATAC-seq peaks shown in Figure 2A. We then used Genomica to select statistically significant motifs that are enriched or depleted in each sample, producing a patient-specific regulatory network (Figure 4A).

Our analysis revealed distinct patterns of TF access of DNA in CTCL and host cells (Figures 4A and 4B). In bulk CTCL and purified leukemia samples, we observed nearly uniform activation of nuclear factor  $\kappa$ B (NF- $\kappa$ B), together with activation of one of three TF motif patterns: (1) Jun-AP1, (2) CTCF, or (3) a motif set that includes EGR, SMAD, MYC, and KLF. Each CTCL sample showed NF- $\kappa$ B activation, but the three companion TF motifs are coordinately enriched in only a subset of CTCL samples and appeared mutually exclusive, indicating that the transcriptional logic of CTCL can be patient specific. These results may highlight the core transcriptional circuitry in CTCL. The broad activation of NF- $\kappa$ B is consistent with the recent discovery in CTCL of activating somatic mutations in TNF receptor 2, downstream signaling components, and *NFKB2* itself, which activates NF- $\kappa$ B activity (Choi et al., 2015; Ungewickell et al., 2015). NF- $\kappa$ B and c-Jun are onco-proteins that regulate many types of cancers including CTCL (Woollard et al., 2016), suggesting that our unbiased results are consistent with previous discoveries. In contrast, the primary role of CTCF is thought to be in regulating the 3D structure of chromatin (Phillips and Corces, 2009); dysregulation of chromosome neighborhoods can also cause cancer through altered gene expression (Flavahan et al., 2016). Regulators with a CTCF-like motif have been reported in regulating leukemia and other cancers (Dolnik et al., 2012;



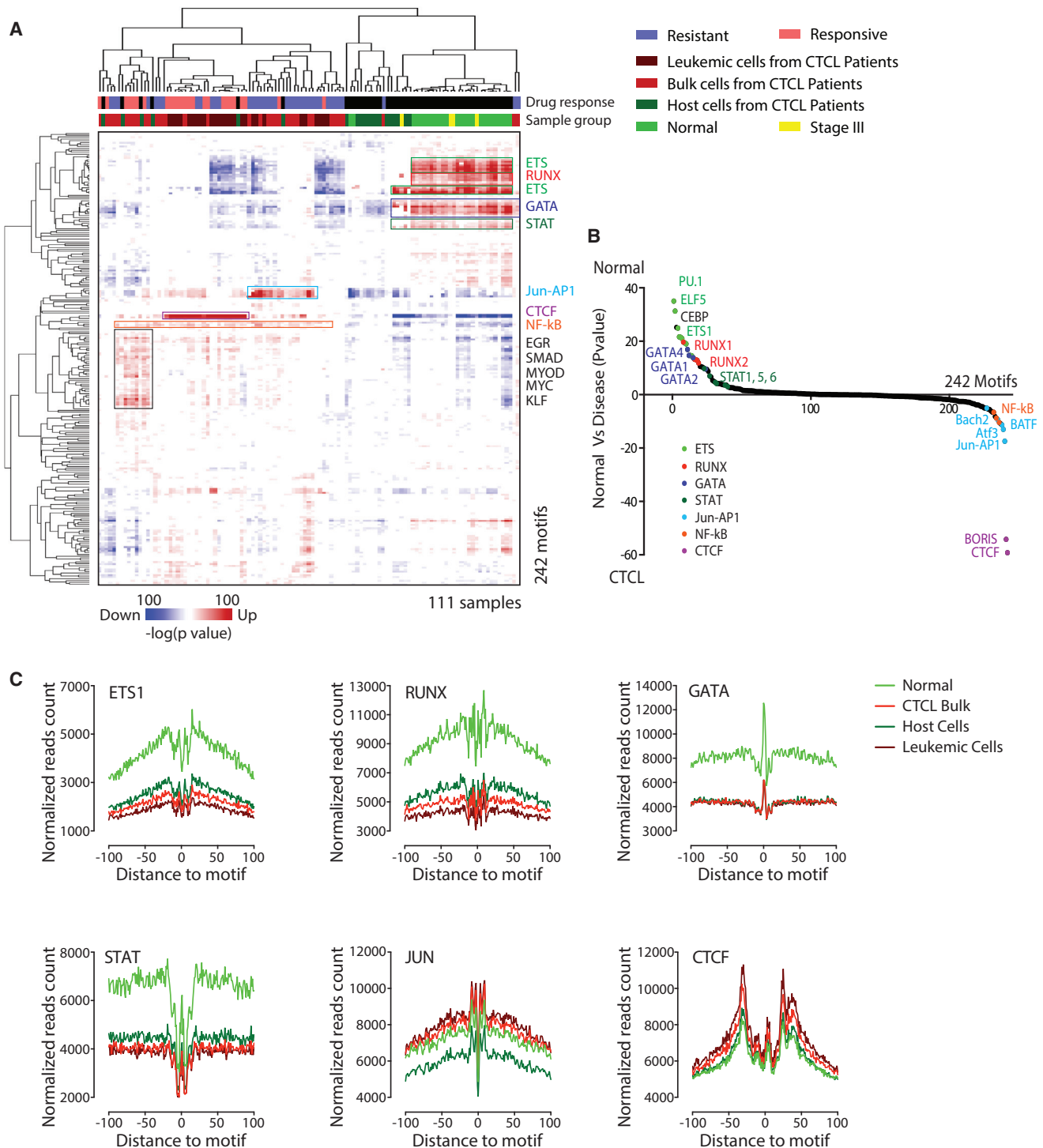
**Figure 3. Change in DNA Accessibility Is Correlated with Differential mRNA Expression in CTCL**

(A) Overlap of differential peaks in normal versus bulk CTCL samples, and differential peaks in host versus leukemic samples define a leukemic signature. (B) Scatterplot of the average fold change of the leukemic signature peaks in host versus leukemic cells versus that of normal versus CTCL bulk cells. Orange indicates peaks that are more accessible in normal and host samples, and blue indicates those that are more accessible in CTCL and leukemic cells. Four outlier dots indicate peaks that are more accessible in host cells and CTCL bulk cells. (C) Normalized ATAC-seq profiles at the *IFNG* locus in normal, host, bulk CTCL, and leukemic cells, and normalized RNA-seq profiles in host and leukemic cells from two individual patients at the same locus. (D) Boxplots (the line in the box shows the median, the upper and lower borders of the box indicate the upper and lower quartile, lines below and above the box indicate the 5th and 95th percentiles, data points beyond the limit of lines mean a minimum to the 5th and a maximum to the 95th percentile) of mRNA expression levels in the leukemic and host cells from two CTCL patients, and of the genes that are more accessible in leukemic cells (left) or host cells (right). The p value was estimated from Student's t test. See also Figure S3.

Yoshida et al., 2013), and CTCF mutations were noted in a subset of CTCLs (Choi et al., 2015). Finally, one of the most interesting discoveries is that motif patterns could be associated with patients' drug responses. We noticed 89.5% of samples (17 out of 19) enriched with CTCF were from patients subsequently responsive to the HDACi romidepsin, and 88.2% of samples (15 out of 17) enriched with Jun-AP1 were from patients resistant to HDACi drugs. While our sample set is not sufficiently powered to address this question, this initial observation suggests patient-specific regulome may inform drug response.

Conversely, DNA access at TF motifs for ETS, RUNX, GATA, and STAT are strongly enriched in normal CD4<sup>+</sup> T cells, but are lost in both leukemic and host non-leukemic CD4<sup>+</sup> T cells. These TFs are involved in a wide variety of functions including the regulation of cellular differentiation, cell-cycle control, cell migration, cell proliferation, apoptosis, and angiogenesis (Sharrocks,

2001). For example, the STAT TF family is well known in mediating many aspects of cancer inflammation and immunity (Yu et al., 2009). STAT3 activation restrains anti-tumor immune responses by antagonizing NF- $\kappa$ B and STAT1-mediated expression of anti-tumor Th1 cytokines such as IL-12 and IFN- $\gamma$ , which are necessary for both innate and T cell-mediated anti-tumor immunity (Yu et al., 2009). It is notable that the depletion of ETS/RUNX/GATA/STAT activity occurred in both purified leukemic cells and host CD4<sup>+</sup> cells from CTCL patients (Figure 4A). This result suggests that there is broad reprogramming of T cell homeostasis in CTCL patients that is shared in the leukemic compartment and host immune response. Quantitative ranking of TF motif enrichments between normal and disease samples largely recapitulated the dominant regulators identified above (Figure 4B). We found that ETS, CEBP, RUNX, GATA, and STAT are the top motifs enriched specifically in normal T cells



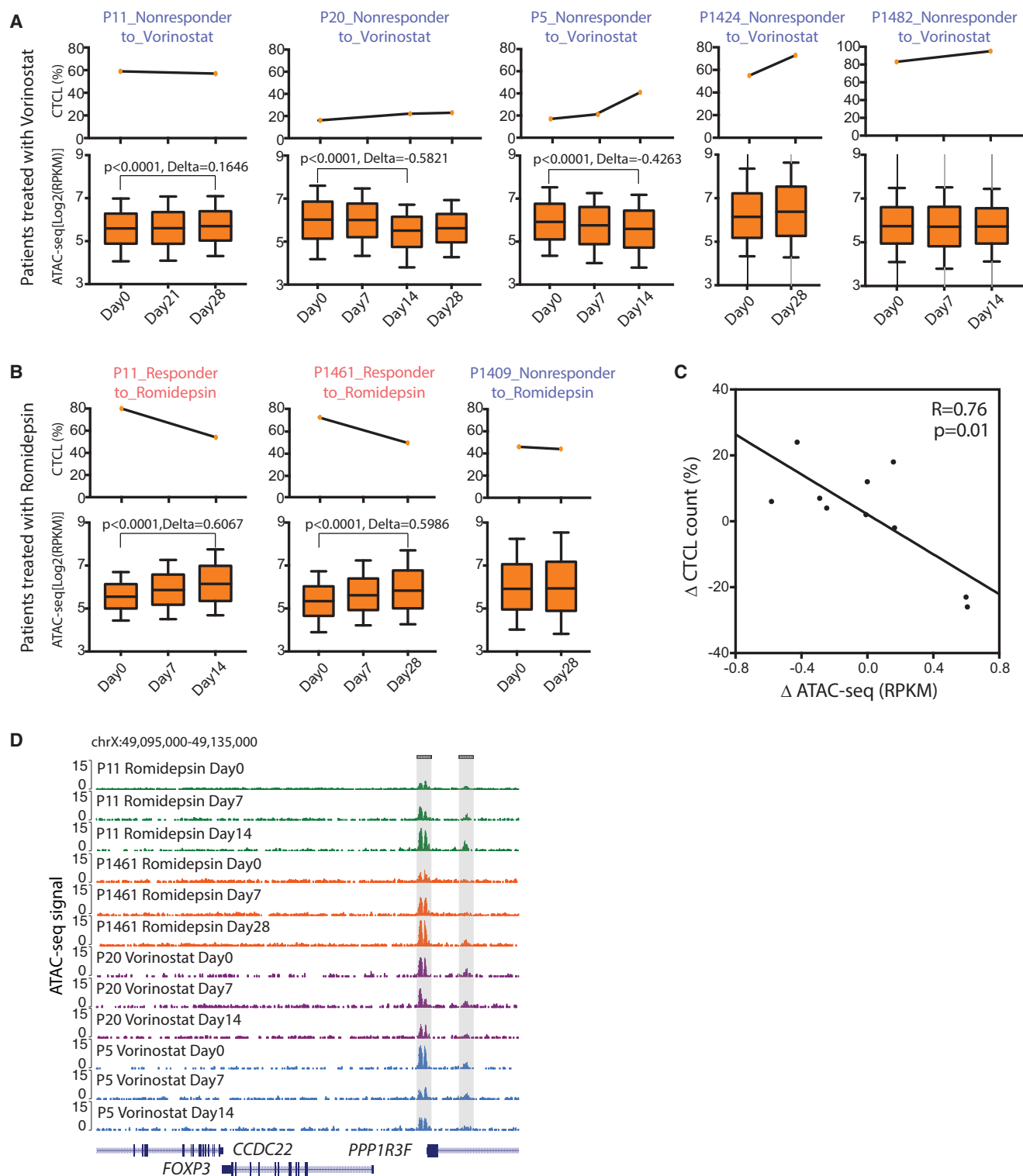
**Figure 4. Transcription Factor Occupancy Networks in CTCL**

(A) Enrichment of known transcription factor motifs in differential accessible elements for all samples. Each row is a motif and each column is a sample. Values in the matrix indicate the significance, in terms of  $-\log(p \text{ value})$ , of the enrichment estimated from Genomica. Top ranked motifs were annotated on the right. Unsupervised hierarchical clustering was performed. Top: the first color bar indicates whether a sample was drug responsive or resistant; the second color bar indicates the category of each sample: normal, host, stage III, bulk CTCL, or leukemic cells.

(B) Ranking of the most differential motifs between normal versus disease samples. The top ones were the most enriched motifs in normal samples and the bottom ones most enriched in disease samples.

(C) Visualization of ATAC-seq footprint for motifs ETS1, RUNX, GATA, STAT, JUN, and CTCF, in normal, host, bulk CTCL, and leukemic cells. The ATAC-seq signal across all the motif binding sites in the genome were aligned on the motif and averaged.





**Figure 5. Personal Chromatin Dynamics with HDACi Treatment**

(A and B) Percentage of CTCL counts (top) and chromatin states indicated by ATAC-seq signal (bottom) of patients during vorinostat (A) or romidepsin (B) treatment from 0 to 5 weeks. Boxplot: the line in the box shows the median, the upper and lower borders of the box indicate the upper and lower quartiles, lines below and above the box indicate the 10th and 90th percentiles. Clinical responders were colored in pink, and non-responders were colored in dark blue. The p values were estimated from Student's t test. Delta means the average reads per kilobase of transcript per million mapped reads (RPKM) fold change between a treated versus an untreated state (day 0).

(legend continued on next page)

but not CTCL, while CTCF, Jun-AP1, NF- $\kappa$ B, and BATF are the top motifs specifically activated in CTCL.

Analysis of our ATAC-seq revealed the “TF footprint” on genomic DNA directly from clinical CTCL samples and primary human T cells (Figure 4C). DNA sequences that are directly occupied by DNA-binding proteins are protected from transposition, analogous to DNase digestion footprints. For the most differential motifs in normal versus disease comparisons, we observed their footprints under all conditions (Figure 4C). We found that TFs enriched in normal T cells, such as ETS1, RUNX, GATA, and STAT, show deeper “footprints” and higher DNA accessibility, flanking their motifs in normal cells compared with leukemic cells. In contrast, the “footprints” of the TFs enriched in leukemic cells, such as JUN and CTCF, show the opposite trend. The degree of TF footprint loss is greater than that of the footprint gain of the respective TFs. Collectively, the results from the orthogonal footprint analysis are consistent with the motif-enrichment analysis using Genomica.

### Personal Chromatin Dynamics with HDACi Treatment

HDACi (here vorinostat and romidepsin) are thought to induce therapeutic effects in cancer by modulation of gene expression (particularly induction of tumor suppressor genes) by increasing histone acetylation and DNA accessibility. However, only a subset, ~43% (6 out of 14 included in this study) of CTCL patients demonstrates clinical response to HDACi, while other patients with similar disease by stage and existing molecular criteria are resistant. The basis for this therapeutic heterogeneity is not known. HDACi treatment is currently given without any ability to visualize its impact on the CTCL epigenome or on gene transcription. We profiled personal regulomes of HDACi treatment in real time, in order to evaluate ahead of time the potentially optimal therapeutic approach, before major changes in cell composition. We focused our analysis on patients who showed eventual clinical response to HDACi rather than on those who did not (“non-responders” thereafter); the latter may have stable disease versus progressive disease, which we do not have sufficient power to distinguish on a molecular level at present.

We found that clinical response to HDACi is correlated with a dynamic change in CTCL DNA accessibility. Patients with clinical response in the blood compartment showed global increase of DNA accessibility during HDACi treatment, while patients resistant to HDACi did not (Figure S4A). We found 62%, 5%, and 33% of cluster I, II, and III peaks (defined in Figure 2A), respectively, were responsive to romidepsin in CTCL patients (Figure S4B). The romidepsin-responsive peaks in cluster I are enriched in immune-relevant gene ontology terms (Figure S4C), suggesting that T cells were activated during drug treatment and turned leukemic cells toward normalcy.

We plotted the CTCL leukemic count versus the DNA accessibility (defined as the reads per kilobase of transcript per million mapped reads normalized ATAC-seq reads count of all peaks) of their bulk CD4<sup>+</sup> T cell, before and during HDACi

therapy for all available patients. We separated patients into two groups, those treated with either vorinostat or romidepsin. Patients who did not respond to HDACi, such as P11 on vorinostat or P1409 on romidepsin, showed negligible changes in DNA accessibility (Figures 5A and 5B). The CTCL counts of patients P20 and P5 actually increased while on vorinostat treatment over 4 weeks, and the chromatin accessibility significantly decreased (Figure 5A). In contrast, when patient P11 was subsequently treated with romidepsin, and had a CTCL count reduction of almost 30%, the ATAC-seq profile showed a significant increase in DNA accessibility. Similar results were also observed in patient P1461 (Figure 5B). These results suggest that the dynamics of chromatin accessibility during HDACi treatment can be patient specific and may predict clinical outcome. Summarizing our entire set of patients, we found that the decrease in CTCL count in response to HDACi is strongly correlated with the quantitative gain of in DNA accessibility as measured by ATAC-seq (Figure 5C). These results suggest that CTCL patients’ clinical responses to HDACi are associated with a specific and dynamic pattern of chromatin decompaction.

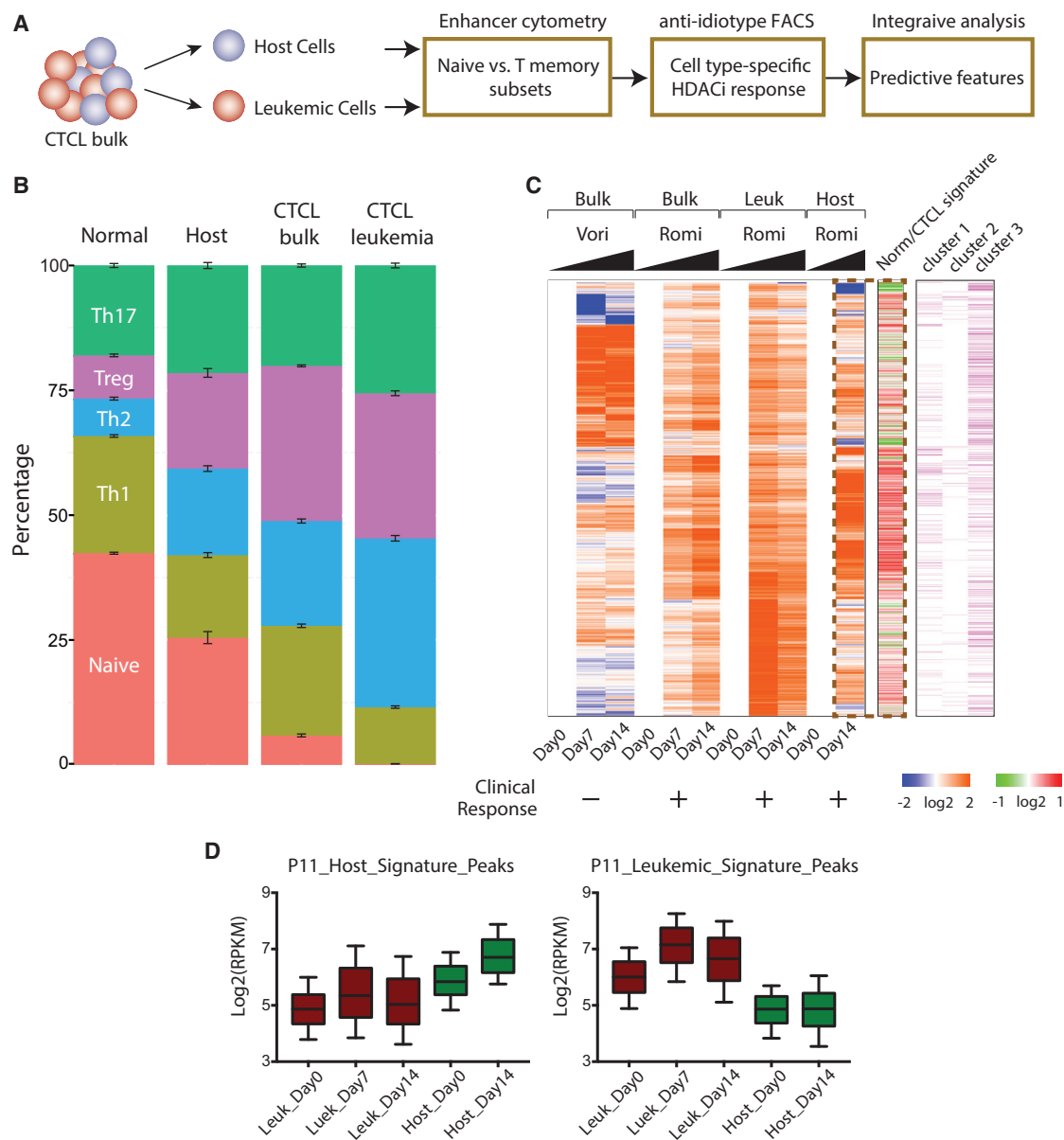
Several notable genes demonstrate HDACi responsiveness in CTCL patients. *FOXP3* is a master regulator of the regulatory pathway in the development and function of Treg cells (Fontenot et al., 2003, 2005; Hori et al., 2003). For responsive patients P11 and P1461, two *FOXP3* enhancers were gradually opened up during the drug treatment, while that of the other two drug-resistant patients, P20 and P5, remained unchanged or even became closed, suggesting that *FOXP3* might be a critical regulator in drug response (Figure 5D). As another example, *IFIT3* is an interferon-induced protein that is upregulated in pancreatic and hypopharynx cancers (Niess et al., 2015; Xu et al., 2013). Figure S4D shows how the enhancer became accessible during drug treatment in clinical responders but remained unchanged or even compacted in clinical non-responders.

We also investigated whether HDACi-induced DNA accessibility around tumor suppressor genes anticipated a HDACi therapeutic effect. In HDACi responders (e.g., P11 and P1461 on romidepsin), many elements flanking tumor suppressor genes became accessible during drug treatment, while in non-responders DNA accessibility of tumor suppressors actually decreased (Figure S5A). The promoter of *TP53*, a well-known tumor suppressor gene, illustrates this dynamic selectively in HDACi responders (Figure S5B).

Furthermore, we adapted a previously described hidden Markov model analysis tool ChromHMM (Ernst and Kellis, 2012) to check which chromatin states in primary T cells were enriched in the top 5,000 altered sites under vorinostat and romidepsin therapy. This analysis revealed that peaks mostly altered under romidepsin treatment were highly enriched in promoters and active enhancers compared with vorinostat (Figure S6), suggesting that peaks altered by romidepsin may have greater influence on gene expression.

(C) Dot plot of delta percentage of CTCL counts versus delta ATAC-seq signal in the form of RPKM. The solid line was fit from linear regression, and the p value and correlation R were estimated from Pearson correlation analysis.

(D) Normalized ATAC-seq profiles at the *FOXP3* locus of patients P11, P1461, P20, and P5 at different time points during HDACi treatment. Shaded regions are peaks identified gradually opened up during the HDACi treatment for responsive patients (P11 and P1461), and remained unchanged or even closed up for resistant patients (P20 and P5). See also Figures S4–S6.



**Figure 6. Enhancer Cytometry and Cell-Type-Specific Response to HDACi Therapy**

(A) Schematic outline of study design of cell-type-specific response to HDACi therapy.

(B) Enhancer cytometry analysis from CIBERSORT using T subtype cell ATAC-seq signatures as eigenvectors to evaluate the T subtype cell composition in a mixture of normal, host, CTCL, and leukemic cells. Error bars represent SEM from all samples in each category. Each T subtype cell was colored differently.

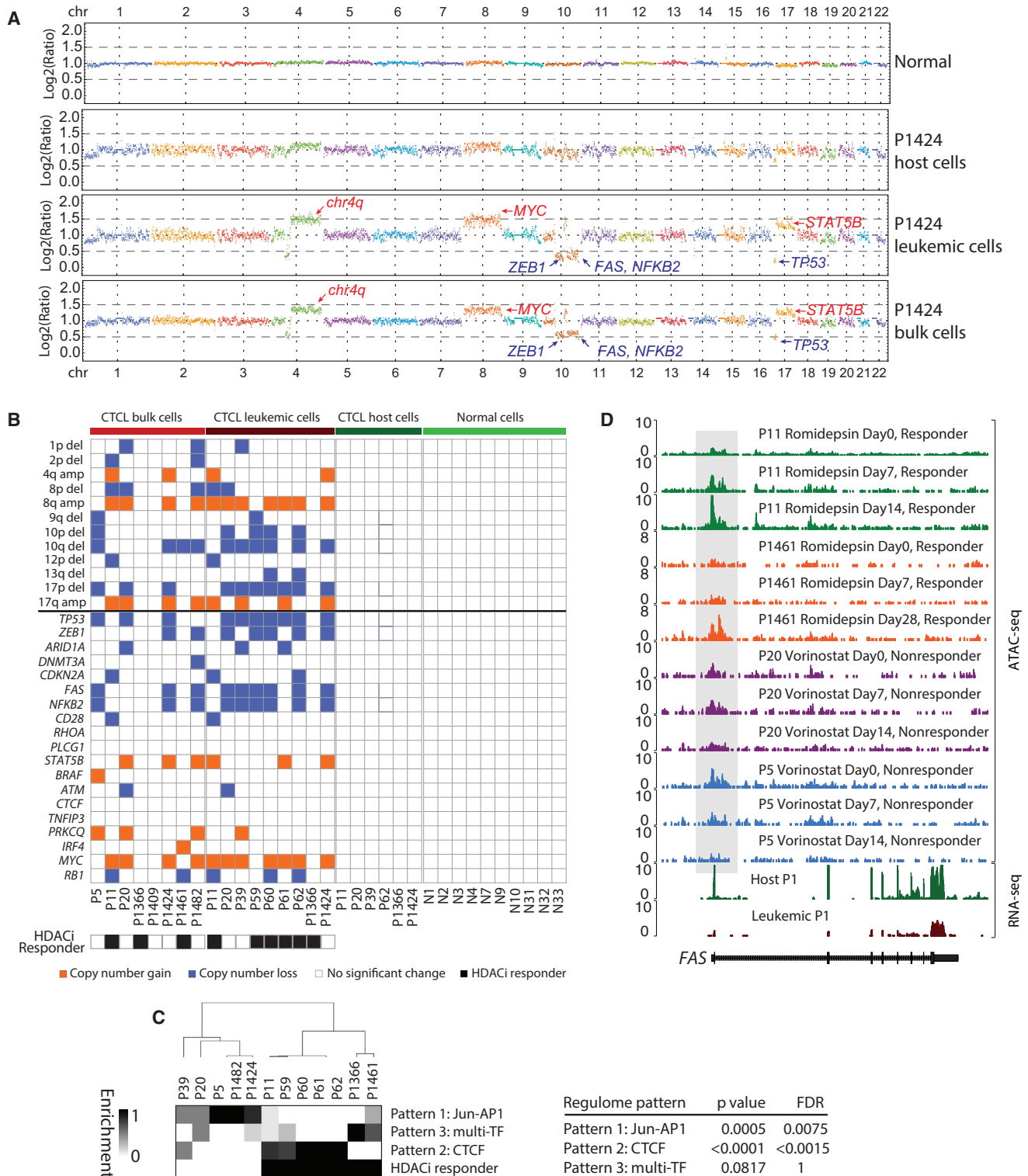
(C) Left: heatmap of relative fold changes of treated states versus their corresponding untreated states (day 0) of the top 5,000 altered peaks in patient P11 treated with vorinostat or romidepsin as indicated. Middle: average fold change of the corresponding peaks in normal versus CTCL samples. Right: whether a corresponding peak belongs to cluster I, II, or III as defined in Figure 2A. Time points at the bottom indicate when samples were subject to ATAC-seq: +, clinical response; –, resistant.

(D) Boxplots (the line in the box shows the median, the upper and lower borders of the box indicate the upper and lower quartiles, lines below and above the box indicate the 10th and 90th percentiles) of host signature peaks (left) and leukemic signature peaks (right) in leukemic (dark red) and host (dark green) cells of patient P11 under treatment with romidepsin. See also Figure S7.

### Enhancer Cytometry and Cell-Type-Specific Response to HDACi Therapy

CD4<sup>+</sup> T cells are heterogeneous, and consist of multiple subtypes, including naive and memory T cell populations. Memory CD4<sup>+</sup> T cells can be further divided based on their differentiation into cytokine-polarized Th1, Th2, Th17, and Treg cells. We

recently developed “enhancer cytometry” wherein we can accurately enumerate the frequency of cell types in complex cellular mixtures based on DNA accessibility data (Corces et al., 2016), and we applied this method to deconvolve the composition of subtypes of T cells in normal and CTCL samples, including purified leukemic and host cells from CTCL patients (Figure 6A).



**Figure 7. Integration of Genomic and Epigenomic Landscape in CTCL**

(A) Human chromosomal diagram showing the areas of DNA copy-number gain and loss identified by ATAC-seq in normal (donor 1), and host, leukemic, and bulk cells from patient P1424. Selected implicated driver genes are shown.

(B) Distribution of recurrent chromosomal (top) and gene (bottom) copy-number variants in CTCL patients (bulk, leukemic, and host cells) and normal donors. Orange boxes indicate copy-number gain, blue boxes copy-number loss, and white boxes no significant change. Patients with black boxes are HDACi drug responders.

(legend continued on next page)

We first isolated human blood-naïve and memory Th1, Th2, Th17, and Treg cells using FACS, and performed ATAC-seq to determine the landscape of chromatin accessibility of each cell type in healthy subjects. We then computed the contribution of each signature to aggregate leukemic cell profiles. Interestingly, bulk CTCL and leukemic samples demonstrated a significantly lower percentage of naïve T cells and a concomitant increase in memory T cells compared with normal or host cells, indicating that CTCL cells are highly similar to memory T cells at the epigenomic level (Figure 6B). Further fractionation of memory T cell profiles showed that CTCL cells demonstrate an increase in Th2 and Treg profiles relative to normal cells. An excess of Treg cells in cancer patients can prevent the immune system from destroying cancer cells (Josefowicz et al., 2012), and the adoption of a Treg-like regulome in CTCL cells may facilitate evasion of anti-tumor immunity. In host T cells of CTCL patients, Th1 frequency is reduced and Th2 frequency is increased (Figure 6B). This result suggests that the balance of T cell homeostasis is disrupted in CTCL patients, either directly or indirectly, with the repression of the Th1 cell signature and enrichment of the Th2 signature. This observation is also consistent with the down-regulation of *IFNG* (Figure 3C), which is a key effector in Th1 cell differentiation and anti-tumor immunity (Platanias, 2005).

We next sought to deconvolve the composition of host and leukemic cells in bulk CTCL samples. As a positive control, enhancer cytometry reported that host cells were predominant in purified host samples, and that leukemic cells were dominant in purified leukemic samples, as expected (Figure S7). Enhancer cytometry of bulk CD4<sup>+</sup> T cells from patients P11 and P1424 was also consistent with their clinical course (Figures 5A and 5B), independently confirming the findings from flow cytometry based on a small number of cell surface markers (Figure S7).

An important question in epigenetic therapy of cancer is the target cell identity. Is it the leukemic cells, the non-malignant host CD4<sup>+</sup> cells, or both, that are altered in response to HDACi? Patient P11 provides a case study. P11 was treated with vorinostat with no clinical response, and then treated with romidepsin with positive response. We were able to obtain ATAC-seq profiles from bulk CD4<sup>+</sup> cells during vorinostat treatment, and bulk, purified leukemic, and host CD4<sup>+</sup> cells during romidepsin treatment. Visualizing the top 5,000 most differentially altered elements in DNA accessibility illustrated three key points (Figure 6C). First, the chromatin accessibility landscape of P11 responded differently to vorinostat versus romidepsin, indicating drug-specific molecular effects. Second, romidepsin induced distinct patterns of DNA accessibility in leukemic versus host CD4<sup>+</sup> T cells (Figure 6C). Intriguingly, the HDACi-induced change in host CD4<sup>+</sup> T cells was much more correlated with the chromatin signature that distinguishes normal versus CTCL cells (Pearson correlation 0.36 for host cells and −0.04 for leukemic cells), suggesting that the host CD4<sup>+</sup> cells may be more relevant

in HDACi's ability to normalize the chromatin state in CTCL patients (Figure 6C, dotted box). Third, and most importantly, we observed that a baseline state of DNA accessibility appears to predict the elements that respond to HDACi. When we compared the romidepsin response with the three clusters of DNA elements that distinguished normal, host CD4<sup>+</sup>, and CTCL cells (defined in Figure 2A), we observed that, in CTCL cells, romidepsin increased DNA accessibility in cluster III elements, which are more accessible in CTCL than normal or host cells. Similarly, in host CD4<sup>+</sup> cells, romidepsin increased accessibility more in cluster I elements, which are accessible in normal and host CD4<sup>+</sup> cells but lost in CTCL. This same point is reinforced by quantitative analysis of DNA elements that are differentially accessible in leukemic cells or host cells (Figure 6D). Host signature elements have a higher accessibility in host cells than CTCL cells at baseline (as expected), and host elements have an increased accessibility in response to romidepsin only in host cells but not leukemic cells. Conversely, leukemic signature elements have a higher accessibility in leukemic cells than in host cells, and romidepsin treatment increases their accessibility in leukemic, but not host cells (Figure 6D). In other words, HDACi accentuate the existing pattern of DNA accessibility in each given cell type, rather than switching on inaccessible sites in either CTCL or host cells.

### Integration of Genomic and Epigenomic Landscapes in CTCL

Genome and exome sequencing of CTCL and SS demonstrated chromosomal copy-number variations (CNV) in patients with the diseases (da Silva Almeida et al., 2015; Choi et al., 2015). We used a method to detect chromosomal CNV from ATAC-seq background reads (Denny et al., 2016), and integrated genomic and epigenomic analyses in CTCL. For each sample, we pinpointed the genomic amplifications and deletions from ATAC-seq data, which recapitulated chromosomal CNVs such as 10q and 17p deletion and 8q and 17q amplification, confirming discoveries from exome and whole-genome DNA sequencing (da Silva Almeida et al., 2015; Choi et al., 2015) (Figures 7A and 7B). Recurrently mutated genes in CTCL, such as *FAS*, *MYC*, *NFKB2*, *STAT5B*, and *TP53* were also discovered (Figures 7A and 7B). In addition, we integrated the chromatin accessibility profile and response to HDACi for each sample with genomic CNV. We found that none of the CNVs can predict the clinical response to HDACi (FDR > 0.05), but, in contrast, the chromatin regulome profile significantly predicted HDACi response ( $p < 0.05$ , FDR < 0.05, chi-square test) (Figure 7C). The presence of the CTCF pattern and the absence of the Jun-AP1 pattern are both predictive of HDACi response. At the level of individual genes, *FAS* encodes a cell surface death receptor, the expression of which in tissue culture experiments has been proposed to explain the efficacy of epigenetic therapy in CTCL (Wu and

(C) Association of CTCL regulome pattern discovered in Figure 4A and HDACi response. Left: 2D unsupervised hierarchical clustering of patients based on their enrichment of each regulome pattern. Right: association of CTCL regulome pattern and HDACi response, the p value was estimated from a chi-square test, and the FDR from Bonferroni correction.

(D) Normalized ATAC-seq profiles at the *FAS* locus of patients P11 (green), P1461 (orange), P20 (purple), and P5 (blue) at different time points during HDACi treatment. Shaded regions are peaks identified as gradually opening up during the HDACi treatment for responsive patients (P11 and P1461), and remaining unchanged or even closed up for resistant patients (P20 and P5). Normalized RNA-seq profiles in host and leukemic cells from patient 1 at the same locus are also shown.



Wood, 2011). However, we found that genomic deletion of the *FAS* locus is not sufficient to predict HDACi clinical response (Figure 7B), because only a subset of patients with intact *FAS* locus de-repressed the gene upon HDACi. We show that HDACi can indeed increase chromatin accessibility at the *FAS* promoter in CTCL from human patients, and that each patient who had increased chromatin accessibility at *FAS* experienced subsequent clinical response to HDACi (Figure 7D). These results highlight the unique prognostic and mechanistic insights that are potentially gained by integration of genomic and epigenomic analysis of CTCL.

## DISCUSSION

Here, we surveyed the landscape of active regulatory DNA in CTCL using the sensitive method of ATAC-seq. Because only ~1% of the human genome is accessible in any given cell type, the identity and pattern of DNA accessibility is highly informative of cell identity, activity state, and regulatory programs. In CTCL, we found distinct patterns of DNA accessibility in both leukemic and host CD4<sup>+</sup> T cells that differ from CD4<sup>+</sup> T cells in healthy individuals. These DNA elements are coordinately associated with genes that intimately control T cell growth, immunity, and homeostasis, adding to the concept that CTCL is both a neoplasm and a systemic disease that alters host immune functions.

In CTCL, the gene-regulatory network is altered with nearly ubiquitous activation of NF- $\kappa$ B, plus one of three mutually exclusive sets of DNA bindings factors Jun-AP1, CTCF, and a set of TFs that includes MYC. Because inhibitors to several of these factors or their upstream regulators have been developed, such as JNK inhibitors (Zhang et al., 2012), Aza-C for CTCF (Flavahan et al., 2016), or BRD4 inhibitors for Myc deactivation (Filippakopoulos et al., 2010), knowledge of the dominant regulators in each patient sample may afford rational matching of patients to targeted therapies. In host CD4<sup>+</sup> T cells, we found systemic deactivation of programs that confer immune functions driven by STAT, GATA, ETS, and RUNX factors. The net effect of these regulatory perturbations is an expansion of memory Treg and Th2 regulomes in CTCL leukemic cells and host cells, at the expense of naive T cells and Th1 subsets.

The nearly complete loss of naive regulome signatures in leukemic cells suggests that CTCL may derive from founder memory T cell clones. Since memory T cells are long lived and share self-renewal properties with hematopoietic stem cells, these data may indicate a model of lymphoma evolution that mirrors the process observed in myeloid leukemia (Corces-Zimmerman et al., 2014; Jan et al., 2012). Namely, memory T cells may serve as a reservoir for the accumulation of cancer-causing mutations, which eventually lead to frank lymphoma. Our findings using DNA accessibility and enhancer cytometry are consistent with the extensive literature on the immune profile of CTCL (Dulmage and Geskin, 2013), and are advantageous in that multiple insights (including TF drivers not available from previous assays) can be learned from a single regulome profile. Personal regulome analysis of acute myeloid leukemia has identified prognostic features and therapeutic targets (Corces et al., 2016; Mazumdar et al., 2015). Our results with CTCL further demonstrate the feasibility and potential insights from the application of cutting-edge epigenomic technology on patient samples for precision medicine.

Despite the fact that CTCL is the first clinical indication that led to the approval of HDACi for human use, the chromatin dynamics of CTCL patients being treated with HDACi are not known. We combined CTCL purification, based on the clonal T cell receptor V-beta idotype, and ATAC-seq profiling, to track CTCL patients throughout treatment with HDACi vorinostat or romidepsin. We were fortunate to identify a patient who had a positive clinical response to romidepsin and whose CTCL had a V-beta clone that we were able to sort and purify. Our results held two surprises that ran contrary to the conventional thinking on the epigenetic treatment of cancer. First, we found that the HDACi romidepsin had profound effects on the DNA accessibility of CTCL and host CD4<sup>+</sup> T cells, and it was the change in host T cells that was more aligned with normalizing the chromatin signature. Thus, the host immune system may be equally important as the cancer cell for epigenetic therapy. Recent studies of EZH2 and DNA methyltransferase inhibitors have shown an important role in de-repressing host immunity against cancer in preclinical models (Chiappinelli et al., 2015; Peng et al., 2015). Our results in CTCL patient are consistent with this concept and suggest HDACi as another possible agent to manipulate host immunity in cancer.

Second, we found that the clinical response to HDACi is associated with a global increase in CTCL DNA accessibility, and that HDACi appears to accentuate the status quo pattern of DNA accessibility rather than evoking new accessible elements. In our patient series, clinical response differentiated vorinostat versus romidepsin. It is unclear whether a higher dose of vorinostat could have induced DNA accessibility or clinical response. Nonetheless, HDACi therapy is currently given in a set regimen without evaluating the effect on CTCL chromatin state. Functional feedback on the therapeutic target, as is currently practiced for measuring blood culture for anti-microbial therapy or clotting time for dosing anti-coagulation therapy, greatly improves the precision and likelihood of reaching the therapeutic goal. Our findings provide a rationale for potentially monitoring the chromatin state of cancer cells and host cells during epigenetic therapy of cancer. Moreover, we found that existing accessible elements, rather than dormant tumor suppressor genes or off-lineage regulators, are preferentially induced by HDACi therapy in CTCL and host cells. This finding may explain the observation that restoration of host immunity occurs with HDACi therapy. Our results are based on a small number of observations, and thus should be evaluated for replication in additional patients in future studies.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- METHOD DETAILS
  - Patients
  - Cell Isolation
  - ATAC-Seq
  - Primary Data Processing and Peak Calling
  - Data Quality Control
  - Significance Analysis
  - Genomic Segmentation Analysis

- Integrative Analysis of Chromatin Accessibility (ATAC-seq) and Gene Expression (RNA-seq) Profiles
- Regulatory Network and TF Footprinting Analysis
- Drug Response Analysis
- Enhancer Cytometry Analysis
- Copy Number Variation Analysis
- **DATA AND SOFTWARE AVAILABILITY**
  - GEO Accession
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

## SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and two tables and can be found with this article online at <http://dx.doi.org/10.1016/j.ccell.2017.05.008>.

## AUTHOR CONTRIBUTIONS

L.C.Z., Y.H.K., and H.Y.C. conceived the project. L.C.Z., R.A., and A.T.S. performed all cell sorting. R.L., L.C.Z., A.T.S., and P.G.G. performed ATAC-seq experiments. N.S., Z.R., and H.U. collected patient samples and clinical information. K.Q., Y.J., and C.J. performed all data analysis. W.J.G. provided analysis methods. K.Q. and H.Y.C. wrote the manuscript with inputs from all authors.

## ACKNOWLEDGMENTS

We thank members of Chang and Greenleaf labs for discussions, Grant Ognibene and Ililsha Rajasansi for coordinating and collecting patient samples, and all the patients for participation. This work was supported by Haas Family Foundation (to Y.H. and H.Y.C.), NIH grant P50-HG007735 (to H.Y.C. and W.J.G.), R35-CA209919 (to H.Y.C.), Stanford Cancer Institute (to Y.H. and H.Y.C.), the National Natural Science Foundation of China grant 91640113 (to K.Q.), the Chinese Government 1000 Youth Talent Program (to K.Q.), the research start-up from the University of Science and Technology of China (to K.Q.). H.Y.C. is a scientific co-founder of Epinomics and a member of its scientific advisory board. W.G.J. is a scientific co-founder of Epinomics and a member of its scientific advisory board. P.G.G. is a co-founder and employee of Epinomics. H.Y.C. is a member researcher of the Parker Institute for Cancer Immunotherapy. We thank the Supercomputing Center of the University of Science and Technology of China for providing supercomputing resources for this project.

Received: October 24, 2016

Revised: March 17, 2017

Accepted: May 15, 2017

Published: June 15, 2017

## REFERENCES

- Adey, A., Morrison, H.G., Asan, X., Kitzman, J.O., Turner, E.H., Stackhouse, B., MacKenzie, A.P., Caruccio, N.C., Zhang, X., et al. (2010). Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol.* **11**, R119.
- Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218.
- Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490.
- Chiappinelli, K.B., Strissel, P.L., Desrichard, A., Li, H., Henke, C., Akman, B., Hein, A., Rote, N.S., Cope, L.M., Snyder, A., et al. (2015). Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell* **162**, 974–986.
- Choi, J., Goh, G., Walrad, T., Hong, B.S., Bunick, C.G., Chen, K., Bjornson, R.D., Maman, Y., Wang, T., Tordoff, J., et al. (2015). Genomic landscape of cutaneous T cell lymphoma. *Nat. Genet.* **47**, 1011–1019.
- Corces-Zimmerman, M.R., Hong, W.J., Weissman, I.L., Medeiros, B.C., and Majeti, R. (2014). Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *Proc. Natl. Acad. Sci. USA* **111**, 2548–2553.
- Corces, M.R., Buenrostro, J.D., Wu, B., Greenside, P.G., Chan, S.M., Koenig, J.L., Snyder, M.P., Pritchard, J.K., Kundaje, A., Greenleaf, W.J., et al. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203.
- Cusanovich, D.A., Daza, R., Adey, A., Pliner, H.A., Christiansen, L., Gunderson, K.L., Steemers, F.J., Trapnell, C., and Shendure, J. (2015). Epigenetics. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914.
- da Silva Almeida, A.C., Abate, F., Khiabani, H., Martinez-Escala, E., Guitart, J., Tensen, C.P., Vermeer, M.H., Rabadan, R., Ferrando, A., and Palomero, T. (2015). The mutational landscape of cutaneous T cell lymphoma and Sezary syndrome. *Nat. Genet.* **47**, 1465–1470.
- de Hoon, M.J., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. *Bioinformatics* **20**, 1453–1454.
- Denny, S.K., Yang, D., Chuang, C.H., Brady, J.J., Lim, J.S., Gruner, B.M., Chiou, S.H., Schep, A.N., Baral, J., Hamard, C., et al. (2016). Nf1b promotes metastasis through a widespread increase in chromatin accessibility. *Cell* **166**, 328–342.
- Dolnik, A., Engelmann, J.C., Scharfenberger-Schmeer, M., Mauch, J., Kelkenberg-Schade, S., Haldemann, B., Fries, T., Kronke, J., Kuhn, M.W., Paschka, P., et al. (2012). Commonly altered genomic regions in acute myeloid leukemia are enriched for somatic mutations involved in chromatin remodeling and splicing. *Blood* **120**, e83–92.
- Dulmage, B.O., and Geskin, L.J. (2013). Lessons learned from gene expression profiling of cutaneous T-cell lymphoma. *Br. J. Dermatol.* **169**, 1188–1197.
- Dummer, R., Heald, P.W., Nestle, F.O., Ludwig, E., Laine, E., Hemmi, S., and Burg, G. (1996). Sezary syndrome T-cell clones display T-helper 2 cytokines and express the accessory factor-1 (interferon-gamma receptor beta-chain). *Blood* **88**, 1383–1389.
- Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216.
- Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W.B., Fedorov, O., Morse, E.M., Keates, T., Hickman, T.T., Felletar, I., et al. (2010). Selective inhibition of BET bromodomains. *Nature* **468**, 1067–1073.
- Flavahan, W.A., Drier, Y., Liao, B.B., Gillespie, S.M., Venteicher, A.S., Stemmer-Rachamimov, A.O., Suva, M.L., and Bernstein, B.E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110–114.
- Fontenot, J.D., Gavin, M.A., and Rudensky, A.Y. (2003). Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. *Nat. Immunol.* **4**, 330–336.
- Fontenot, J.D., Rasmussen, J.P., Williams, L.M., Dooley, J.L., Farr, A.G., and Rudensky, A.Y. (2005). Regulatory T cell lineage specification by the forkhead transcription factor foxp3. *Immunity* **22**, 329–341.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589.
- Hori, S., Nomura, T., and Sakaguchi, S. (2003). Control of regulatory T cell development by the transcription factor Foxp3. *Science* **299**, 1057–1061.
- Jan, M., Snyder, T.M., Corces-Zimmerman, M.R., Vyas, P., Weissman, I.L., Quake, S.R., and Majeti, R. (2012). Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Sci. Transl. Med.* **4**, 149ra118.
- Jawed, S.I., Myskowski, P.L., Horwitz, S., Moskowitz, A., and Querfeld, C. (2014). Primary cutaneous T-cell lymphoma (mycosis fungoides and Sezary syndrome): part II. Prognosis, management, and future directions. *J. Am. Acad. Dermatol.* **70**, 223.e1–17, quiz 240–2.
- Josefowicz, S.Z., Lu, L.F., and Rudensky, A.Y. (2012). Regulatory T cells: mechanisms of differentiation and function. *Annu. Rev. Immunol.* **30**, 531–564.

- Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* 22, 1813–1831.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.
- Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretzky, I., Jaitin, D.A., David, E., Keren-Shaul, H., Mildner, A., Winter, D., Jung, S., et al. (2014). Immunogenetics. Chromatin state dynamics during blood formation. *Science* 345, 943–949.
- Lavin, Y., Winter, D., Blecher-Gonen, R., David, E., Keren-Shaul, H., Merad, M., Jung, S., and Amit, I. (2014). Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* 159, 1312–1326.
- Lee, C.S., Ungewickell, A., Bhaduri, A., Qu, K., Webster, D.E., Armstrong, R., Weng, W.K., Aros, C.J., Mah, A., Chen, R.O., et al. (2012). Transcriptome sequencing in Sezary syndrome identifies Sezary cell and mycosis fungoides-associated lncRNAs and novel transcripts. *Blood* 120, 3288–3297.
- Limbach, M., Saare, M., Tserel, L., Kisand, K., Eglit, T., Sauer, S., Axelsson, T., Syvanen, A.C., Metspalu, A., Milani, L., et al. (2016). Epigenetic profiling in CD4+ and CD8+ T cells from Graves' disease patients reveals changes in genes associated with T cell receptor signaling. *J. Autoimmun.* 67, 46–56.
- Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C.Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., et al. (2016). JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 44, D110–D115.
- Mazumdar, C., Shen, Y., Xavy, S., Zhao, F., Reinisch, A., Li, R., Corces, M.R., Flynn, R.A., Buenrostro, J.D., Chan, S.M., et al. (2015). Leukemia-associated cohesin mutants dominantly enforce stem cell programs and impair human hematopoietic progenitor differentiation. *Cell Stem Cell* 17, 675–688.
- McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* 28, 495–501.
- Morita, R., Schmitt, N., Bentebibel, S.E., Ranganathan, R., Bourdery, L., Zurawski, G., Foucat, E., Dullaers, M., Oh, S., Sabzghabaei, N., et al. (2011). Human blood CXCR5(+)CD4(+) T cells are counterparts of T follicular cells and contain specific subsets that differentially support antibody secretion. *Immunity* 34, 108–121.
- New, M., Olzscha, H., and La Thangue, N.B. (2012). HDAC inhibitor-based therapies: can we interpret the code? *Mol. Oncol.* 6, 637–656.
- Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12, 453–457.
- Niess, H., Camaj, P., Mair, R., Renner, A., Zhao, Y., Jackel, C., Nelson, P.J., Jauch, K.W., and Bruns, C.J. (2015). Overexpression of IFN-induced protein with tetratricopeptide repeats 3 (IFIT3) in pancreatic cancer: cellular "pseudoinflammation" contributing to an aggressive phenotype. *Oncotarget* 6, 3306–3318.
- Olsen, E.A., Whittaker, S., Kim, Y.H., Duvic, M., Prince, H.M., Lessin, S.R., Wood, G.S., Willemze, R., Demierre, M.F., Pimpinelli, N., et al. (2011). Clinical end points and response criteria in mycosis fungoides and Sezary syndrome: a consensus statement of the International Society for Cutaneous Lymphomas, the United States Cutaneous Lymphoma Consortium, and the Cutaneous Lymphoma Task Force of the European Organisation for Research and Treatment of Cancer. *J. Clin. Oncol.* 29, 2598–2607.
- Peng, D., Kryczek, I., Nagarsheth, N., Zhao, L., Wei, S., Wang, W., Sun, Y., Zhao, E., Vatan, L., Szeliga, W., et al. (2015). Epigenetic silencing of TH1-type chemokines shapes tumour immunity and immunotherapy. *Nature* 527, 249–253.
- Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. *Cell* 137, 1194–1211.
- Platanias, L.C. (2005). Mechanisms of type-I- and type-II-interferon-mediated signalling. *Nat. Rev. Immunol.* 5, 375–386.
- Qu, K., Zaba, L.C., Giresi, P.G., Li, R., Longmire, M., Kim, Y.H., Greenleaf, W.J., and Chang, H.Y. (2015). Individuality and variation of personal regulomes in primary human T cells. *Cell Syst.* 1, 51–61.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Quinn, E.M., Coleman, C., Molloy, B., Dominguez Castro, P., Cormican, P., Trimble, V., Mahmud, N., and McManus, R. (2015). Transcriptome analysis of CD4+ T cells in coeliac disease reveals imprint of BACH2 and IFNGamma regulation. *PLoS One* 10, e0140049.
- Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330.
- Rodriguez-Paredes, M., and Esteller, M. (2011). Cancer epigenetics reaches mainstream oncology. *Nat. Med.* 17, 330–339.
- Segal, E., Friedman, N., Koller, D., and Regev, A. (2004). A module map showing conditional activity of expression modules in cancer. *Nat. Genet.* 36, 1090–1098.
- Sharrocks, A.D. (2001). The ETS-domain transcription factor family. *Nat. Rev. Mol. Cell Biol.* 2, 827–837.
- Sherwood, R.I., Hashimoto, T., O'Donnell, C.W., Lewis, S., Barkal, A.A., van Hoff, J.P., Karun, V., Jaakkola, T., and Gifford, D.K. (2014). Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat. Biotechnol.* 32, 171–178.
- Simon Anders, W.H. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11, R106.
- Tao, R., de Zoeten, E.F., Ozkaynak, E., Chen, C., Wang, L., Porrett, P.M., Li, B., Turka, L.A., Olson, E.N., Greene, M.I., et al. (2007). Deacetylase inhibition promotes the generation and function of regulatory T cells. *Nat. Med.* 13, 1299–1307.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578.
- Ungewickell, A., Bhaduri, A., Rios, E., Reuter, J., Lee, C.S., Mah, A., Zehnder, A., Ohgami, R., Kulkarni, S., Armstrong, R., et al. (2015). Genomic analysis of mycosis fungoides and Sezary syndrome identifies recurrent alterations in TNFR2. *Nat. Genet.* 47, 1056–1060.
- Willemze, R., Jaffe, E.S., Burg, G., Cerroni, L., Berti, E., Swerdlow, S.H., Ralfkiaer, E., Chimenti, S., Diaz-Perez, J.L., Duncan, L.M., et al. (2005). WHO-EORTC classification for cutaneous lymphomas. *Blood* 105, 3768–3785.
- Woollard, W.J., Pullabhatla, V., Lorenc, A., Patel, V.M., Butler, R.M., Bayega, A., Begum, N., Bakr, F., Dedhia, K., Fisher, J., et al. (2016). Candidate driver genes involved in genome maintenance and DNA repair in Sezary syndrome. *Blood* 127, 3387–3397.
- Wu, J., and Wood, G.S. (2011). Reduction of Fas/CD95 promoter methylation, upregulation of Fas protein, and enhancement of sensitivity to apoptosis in cutaneous T-cell lymphoma. *Arch. Dermatol.* 147, 443–449.
- Xu, C.Z., Shi, R.J., Chen, D., Sun, Y.Y., Wu, Q.W., Wang, T., and Wang, P.H. (2013). Potential biomarkers for paclitaxel sensitivity in hypopharynx cancer cell. *Int. J. Clin. Exp. Pathol.* 6, 2745–2756.
- Yoshida, K., Toki, T., Okuno, Y., Kanezaki, R., Shiraiishi, Y., Sato-Otsubo, A., Sanada, M., Park, M.J., Terui, K., Suzuki, H., et al. (2013). The landscape of somatic mutations in Down syndrome-related myeloid disorders. *Nat. Genet.* 45, 1293–1299.
- Yu, H., Pardoll, D., and Jove, R. (2009). STATs in cancer inflammation and immunity: a leading role for STAT3. *Nat. Rev. Cancer* 9, 798–809.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-seq (MACS). *Genome Biol.* 9, R137.
- Zhang, T., Inesta-Vaquera, F., Niepel, M., Zhang, J., Ficarro, S.B., Machleidt, T., Xie, T., Marto, J.A., Kim, N., Sim, T., et al. (2012). Discovery of potent and selective covalent inhibitors of JNK. *Chem. Biol.* 19, 140–154.
- Zhao, M., Sun, J., and Zhao, Z. (2013). TSGene: a web resource for tumor suppressor genes. *Nucleic Acids Res.* 41, D970–D976.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT OR RESOURCE	SOURCE	IDENTIFIER
Clinical staging and assessment of CTCL	(Olsen et al., 2011)	<a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3422534/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3422534/</a>
<b>Antibodies</b>		
PerCP/Cy5.5 anti-human CD45RA	Biolegend Inc	BioLegend Cat# 304122; RRID: AB_893357
Brilliant Violet 510™ anti-human CD127 (IL-7R $\alpha$ )	Biolegend Inc	BioLegend Cat# 351331; RRID: AB_2561935
APC/Cy7 anti-human CD4	Biolegend Inc	BioLegend Cat# 344616; RRID: AB_2028483
PE anti-human CD196 (CCR6)	Biolegend Inc	BioLegend Cat# 353410; RRID: AB_10913815
FITC anti-human CD25	Biolegend Inc	BioLegend Cat# 302603; RRID: AB_314273
Brilliant Violet 421™ anti-human CD183 (CXCR3)	Biolegend Inc	BioLegend Cat# 353715; RRID: AB_11124720
<b>Software and Algorithms</b>		
Bowtie2	(Langmead and Salzberg, 2012)	<a href="http://bowtie-bio.sourceforge.net/bowtie2/index.shtml">http://bowtie-bio.sourceforge.net/bowtie2/index.shtml</a>
MACS2	(Zhang et al., 2008)	<a href="https://github.com/taoliu/MACS">https://github.com/taoliu/MACS</a>
DESeq	(Simon Anders, 2010)	<a href="http://bioconductor.org/packages/release/bioc/html/DESeq.html">http://bioconductor.org/packages/release/bioc/html/DESeq.html</a>
BedTools	(Quinlan and Hall, 2010)	<a href="http://bedtools.readthedocs.io/en/latest/">http://bedtools.readthedocs.io/en/latest/</a>
Cluster 3.0	(de Hoon et al., 2004)	<a href="http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm">http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm</a>
Java Treeview		<a href="http://jtreeview.sourceforge.net/">http://jtreeview.sourceforge.net/</a>
IDR	(Landt et al., 2012)	<a href="https://www.encodeproject.org/software/idr/">https://www.encodeproject.org/software/idr/</a>
GREAT	(McLean et al., 2010)	<a href="http://bejerano.stanford.edu/great/public/html/">http://bejerano.stanford.edu/great/public/html/</a>
TopHat	(Trapnell et al., 2012)	<a href="https://ccb.jhu.edu/software/tophat/index.shtml">https://ccb.jhu.edu/software/tophat/index.shtml</a>
Jaspar		<a href="http://jaspar.genereg.net/">http://jaspar.genereg.net/</a>
HOMER	(Heinz et al., 2010)	<a href="http://homer.ucsd.edu/homer/motif/index.html">http://homer.ucsd.edu/homer/motif/index.html</a>
Genomica	(Segal et al., 2004)	<a href="https://genomica.weizmann.ac.il/">https://genomica.weizmann.ac.il/</a>
PIQ	(Sherwood et al., 2014)	<a href="http://piq.csail.mit.edu/">http://piq.csail.mit.edu/</a>
CIBERSORT	(Newman et al., 2015)	<a href="https://cibersort.stanford.edu/">https://cibersort.stanford.edu/</a>
<b>Data used in this paper</b>		
Deposited data (All self-generated data)	GSE85853	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=adolemkolctcbncd&amp;acc=GSE85853">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=adolemkolctcbncd&amp;acc=GSE85853</a>
ChIP-seq of H3K27Ac in Normal CD4 <sup>+</sup> T cell	GSE71952 (Limbach et al., 2016)	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE71952">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE71952</a>
mRNA expression RNA-seq data in Normal CD4 <sup>+</sup> T cells	GSE69549 (Lee et al., 2012; Quinn et al., 2015)	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE69549">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE69549</a>
mRNA expression RNA-seq data in CTCL host and leukemic T cells	SRA046736 (Lee et al., 2012; Quinn et al., 2015)	<a href="http://www.bloodjournal.org/content/bloodjournal/suppl/2012/08/30/blood-2012-04-423061.DC1/blood-2012-04-423061-1.pdf">http://www.bloodjournal.org/content/bloodjournal/suppl/2012/08/30/blood-2012-04-423061.DC1/blood-2012-04-423061-1.pdf</a>



## CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to the Lead Contact Howard Y. Chang at [howchang@stanford.edu](mailto:howchang@stanford.edu)

## METHOD DETAILS

### Patients

Following informed consent per the Declaration of Helsinki, Sézary syndrome (SS) patient samples were collected under a protocol approved by the Institutional Review Board at Stanford University Medical Center. Patient characteristics are described in [Table S2](#). All samples were obtained from patients with either clinical stage III or IV MF or SS. All diagnoses were confirmed by a board-certified dermatopathologist. Clinical staging and assessment of response to treatment are as defined by CTCL international consensus criteria ([Olsen et al., 2011](#)).

### Cell Isolation

Normal donors were recruited under a Stanford University IRB-approved protocol. Informed consent was obtained. Standard blood draws in green-top tube were obtained for each time point. 1-5mL of whole blood was enriched for CD4<sup>+</sup> cells using RosetteSep Human CD4<sup>+</sup> T Cell Enrichment Cocktail (StemCell Technology) as described ([Buenrostro et al., 2013](#)). PBMCs were prepared by Ficoll-Hypaque density-gradient centrifugation. PBMCs were stained with fluorochrome-labeled anti-human monoclonal antibodies (Biolegend Inc) to CD45 (clone HI30), CD4 (clone RPA-T4), and CD3 (clone HIT3a). T-cell receptor (TCR) V $\beta$  clonality was determined with the TCR V $\beta$  Repertoire Kit (Beckman-Coulter). Antibody-stained patient lymphocytes were sorted into CD3<sup>+</sup>CD4<sup>+</sup>V $\beta$ <sup>+</sup> and CD3<sup>+</sup>CD4<sup>+</sup>V $\beta$ <sup>-</sup> fractions with the use of an Influx flow cytometer (Becton Dickinson) ([Dummer et al., 1996](#)). At least 50,000 CD4<sup>+</sup> T cells were enriched by negative selection without ex vivo expansion. For CD4<sup>+</sup> T helper cell subtypes, cells were sorted as previously described ([Morita et al., 2011](#)). Briefly, naive cells were sorted as CD4<sup>+</sup>CD25<sup>-</sup>CD45RA<sup>+</sup>, Th1 cells as CD4<sup>+</sup>CD25<sup>-</sup>CD45RA<sup>+</sup>CXCR3<sup>+</sup>CCR6<sup>-</sup>, Th2 cells as CD4<sup>+</sup>CD25<sup>-</sup>CD45RA<sup>+</sup>CXCR3<sup>+</sup>CCR6<sup>-</sup>, Th17 cells as CD4<sup>+</sup>CD25<sup>-</sup>CD45RA<sup>+</sup>CXCR3<sup>+</sup>CCR6<sup>+</sup>, and Treg cells as CD4<sup>+</sup>CD25<sup>+</sup>CD127<sup>lo</sup>. >95% post-sort purities were confirmed prior to ATAC-seq.

### ATAC-Seq

ATAC-seq was performed as described ([Buenrostro et al., 2013](#)), and 2x50 paired-end sequencing performed on Illumina NextSeq500 to yield on average 55M reads/sample.

### Primary Data Processing and Peak Calling

Adapter sequences trimming, mapping to Hg19 using Bowtie2 ([Langmead and Salzberg, 2012](#)) (with option `-very-sensitive`), and PCR duplicate removal were as described ([Buenrostro et al., 2013](#)). Mapped reads were shifted +4/-5bp depending on the strand of the read, so that the first base of each mapped read representing the Tn5 cleavage position. All mapped reads were then extended to 50bp centered by the cleavage position. Reads mapped to repeated regions and chromosome M were removed. Peak calling was performed using MACS2 ([Zhang et al., 2008](#)) with options `-f BED`, `-g hs`, `-q 0.01`, `-nomodel`, `-shift 0`. Peaks in each sample were called for QC purpose. Samples were then grouped into 5 categories, Normal, Host, Stage3, BulkCTCL and Leukemic. Reads from the same category were concatenated, on which peak calling was performed using the same options. Peaks in each category were further filtered, and high quality peaks with  $FDR < 10E-7$  were obtained. Peaks for all the categories were then merged together into a unique peak list, and number of raw reads mapped to each peak at each condition was quantified using `intersectBed` function in `BedTools` ([Quinlan and Hall, 2010](#)). Peak raw counts were quartile normalized using `DESeq` ([Simon Anders, 2010](#)) package in R. Peak intensity was defined as  $\log_2$  of the normalized counts. After these steps, an  $N \times M$  data matrix was obtained where  $N$  indicates the number of merged peaks,  $M$  indicates the number of samples, and value  $D_{i,j}$  indicates the peak intensity of peak  $i$  ( $i=1$  to  $N$ ) in sample  $j$  ( $j=1$  to  $M$ ). For each pairs of samples, Pearson correlation was calculated based on the  $\log_2$  normalized counts of all the peaks. Unsupervised correlation of the Pearson correlation matrix was performed using `Cluster 3.0` ([de Hoon et al., 2004](#)), and visualized in `Java Treeview`.

### Data Quality Control

ATAC-seq data quality measure was comprehensively studied in our previous work ([Qu et al., 2015](#)). The ability of using ATAC-seq to detect accessible sites was proved by comparing ATAC-seq signal with DNaseI hypersensitivity sequencing (DHSseq), which serves as a gold standard assay and positive control for open chromatin. We found that ATAC-seq was highly reproducible between replicates. From an irreproducibility discovery rate (IDR) analysis ([Landt et al., 2012](#)), which was an ENCODE-specified method to evaluate data reproducibility across replicates, we found that the number of reproducible peaks plateau between 11-12 million mappable reads, irrespective of the IDR cutoff ([Qu et al., 2015](#)). In this study, we have on average 26 million uniquely mapped reads ([Table S1](#)), suggesting the sequencing in this study was deep enough to confidently capture the majority of the regions of interest. We then using several metrics to evaluate the quality of each sample, including the number of raw reads, overall alignment rate, final mapped reads, final mapped rate, percentage of reads mapped to chrM, percentage of reads mapped to repeat regions (black list), percentage of reads filtered out by low MAPQ score, percentage of PCR duplicates, TSS enrichment score (reads that enriched at +/-2kb around TSS versus the background), read length distribution and number of peaks ([Table S1](#)). We then end up with 111 high quality samples.



### Genomic Segmentation Analysis

Genomic location classification was defined by chromHMM in Epigenomic Roadmap Consortium (Roadmap Epigenomics et al., 2015) on Naïve T cells. A chromHMM 25-state classification was then combined to 5 states, promoter, transcription, active enhancer, weak enhancer and heterochromatin. All merged peaks, differential peaks and peaks in each cluster in Figure 2A were then overlaid with these 5 genomic segments, e.g. if the center of a peak resides in promoter, then this peak was assigned to promoter, and same for all the other segments.

### Integrative Analysis of Chromatin Accessibility (ATAC-seq) and Gene Expression (RNA-seq) Profiles

We analyzed public datasets from a transcriptome analysis of Sezary syndrome (Lee et al., 2012) to validate the positive correlation between the ATAC-seq signals with the mRNA expressions of the genes nearby. Nearby genes were defined from GREAT (McLean et al., 2010). RNA-seq reads were aligned to the human reference sequence NCBI (NCBI) build GRCh37/hg19 with TopHat (Trapnell et al., 2012). A combination of RefSeq, UCSC genes, and Gencode databases were used as reference annotations. Multiple isoforms of a same gene were collapsed so that only values of fragments per kilobase per million mapped reads of each gene were calculated with a self-developed script (K.Q.). Boxplot of gene expression difference in Figure 3D was performed using Prism 7, and p values were estimated from Student T-test.

### Regulatory Network and TF Footprinting Analysis

Input motifs set was obtained from jasper (<http://jaspar.genereg.net/>) on vertebrates. We searched for input motifs in each significant differential peak using HOMER (Heinz et al., 2010), and generated a peak versus motif matrix, where each row is a peak and each column is a motif. Value 1 in the matrix represents this motif is found in the peak and value 0 represents NOT. From peak significant analysis, we had a peak by sample matrix where each row is a peak and each column is a sample, and values in the matrix represent the peak intensity in the corresponding sample. We then mean-centered the values in each peak to obtain the relative fold changes of each peak across all the samples. By integrating this two matrixes into Genomica (Segal et al., 2004), using the ModuleMap algorithm, we were able to ask whether and to what extent a motif was enriched in each sample, and obtain a motif by sample matrix, where each row is a motif, each column is a sample, and the values in this matrix represent the significance of enrichment by  $-\log(p \text{ value})$ . If the average fold change of a motif in a sample was positive, then this motif was defined as positive enriched in the sample; and if that of a motif in a sample was negative, then this motif was defined as negative enriched in the sample. The top enriched/depleted between normal and disease was defined as the  $-\log(p \text{ value})$  differences between normal samples and disease samples (here CTCL Bulk and Leukemic samples). The higher this value is, the more enriched this motif is in normal samples, and the lower this value is, the more enriched this motif is in disease samples. The genome-wide motif footprinting analysis was performed using PIQ v1.2 (Sherwood et al., 2014). For footprinting, we adjusted the read start sites to represent the center of the transposon's binding event (see above). Previous descriptions of the Tn5 transposase show that the transposon binds as a dimer and inserts two adaptors separated by 9 bp (Adey et al., 2010). Therefore, we modified the reads' aligned file in sam format by offsetting +4bp for all the reads aligned to the forward strand, and -5bp for all the reads aligned to the reverse strand. We then converted a shifted base sam file to bam format and had the bam file sorted using samtools. We concatenated ATAC-seq reads each category of samples (Normal, Host, CTCL Bulk and Leukemic) and randomly selected 100M reads from each group and made merged bam files. PIQ takes a sorted bam file and a list of motif position weight matrix (PWM) file as inputs. We took default settings and run PIQ as instructed here: <https://bitbucket.org/thashim/piq-single>. PIQ predicted the genomic occupation of 242 motifs with binding affinity estimated by purity scores. We filtered the PIQ predictions using a purity score cutoff at 0.7, and overlaid these predictions with randomly selected 100M reads in each category. We then averaged the overlaid read counts of 200bp genomic region centered by the motif sites for motif footprints.

### Drug Response Analysis

To evaluate how the entire chromatin responds to HDACi anti-cancer drugs, we used an in house generated script (K.Q.) to calculate the RPKMs of all the chromatin accessible sites, and see how they were changed during the drug treatment. The top 5000 most altered peaks in terms of RPKM fold changes, either gain or lose accessibility, were selected and the fold changes of treated states versus their corresponding untreated states (Day0) of the same patient were plotted (Figures 6C and S4A). Samples were ordered according to their drug treatment time points, and unsupervised clustering was performed on the peaks using software Cluster 3.0 and visualized in Treeview. The average fold changes of the corresponding peak in normal versus CTCL samples was also plotted, together with whether a peak belonged to cluster 1, 2, or 3 defined in Figure 2A. An entire list of tumor suppressor genes (1217 genes) was obtained from Tumor Suppressor Gene Database (Zhao et al., 2013), and a list of 1301 peaks -5kb to 2kb around the promoters of these genes were defined as tumor suppressor peaks. Drug response of these tumor suppressor peaks was analyzed and plotted in a similar way (Figure S5A). The clinical response of leukemic (Sézary) cells with HDACi therapy was assessed using flow cytometric analysis conducted as part of clinical care practice (Stanford Hospital Clinical Laboratory, Stanford, CA).

### Enhancer Cytometry Analysis

CIBERSORT (Newman et al., 2015) with "Custom Signature Gene" option was used to deconvolve the composition of subtype T cells in bulk CTCL samples. ATAC-seq on primary Naïve, Th1, Th2, Treg and Th17 cells were performed and data was processed in a same way as above. Reads in each cell type were concatenated and peaks were called using MACS2 as above. Only high quality peaks in

each cell type were kept and merged to a T cell subtype peak list. Peak intensities in subtype T cells as well as Bulk CTCL and other samples were obtained using BedTools. Peaks mapping to chromosome X and promoter/TSS regions were manually removed. Manual removal of peaks improves the predictive capabilities of CIBERSORT (Corces et al., 2016). Peak intensity matrix in subtype T cells was used as the reference sample file, and phenotype class file was manually generated for CIBERSORT to obtain a custom gene signature. Peak intensity matrix in Bulk CTCL and other samples was used as the mixture file in CIBERSORT, which then deconvolved the composition of T subtypes in each sample in the mixture file. Predicted percentages of each T subtype cells in each sample were grouped into Normal, Host, CTCL bulk and CTCL leukemic samples, and histogram was performed in R.

### Copy Number Variation Analysis

Copy number variation analysis was performed according to (Denny et al., 2016). ATAC-seq reads from the same patient or healthy donor at each category (bulk, host or leukemic) were concatenated to address the chromosomal copy number of the patient or donor. Each chromosome was divided into windows size of 1Mbp at sliding step of 500Kbp. Within each window, reads mapped to ATAC-seq peaks were removed, remaining only the background reads. The average coverage for window was defined as follows:

$$\text{Average\_coverage} = \frac{\text{Reads mapped into window} - \text{Reads mapped into peaks}}{\text{Window size} - \text{sum of peak size}}$$

We then performed 3 normalizing steps to estimate the copy number of each sliding window in each sample: (1) by sequencing depth from each concatenated sample; (2) by the median coverage of the normal samples and calculated the Log2 ratio over that median coverage; and (3) by the median coverage of each sample itself. We defined a copy number gain or loss by Z-score greater than 2. Regulome patterns were defined in Figure 4A, and the enrichment of each pattern for each patient was defined as the number of samples in each pattern divided by the total number of samples of the same patient in all the patterns.

### DATA AND SOFTWARE AVAILABILITY

#### GEO Accession

Raw reads and called peaks for all samples are available through the Gene Expression Omnibus (GEO) via accession GSE85853.

### QUANTIFICATION AND STATISTICAL ANALYSIS

Student T-tests were applied in Figures 3D, 5A, 5B, and S3A.

Chi-Square Test and Bonferroni correction were applied in Figure 7C.