Single-cell analysis of Sézary syndrome reveals novel markers and shifting gene profiles associated with treatment

Tracking no: ADV-2021-005991R1

Nicholas Borcherding (University of Iowa, United States) Kevin Severson (Mayo Clinic, United States) Nicholas Henderson (University of Iowa, United States) Luana Dos Santos Ortolan (Seattle Childrens Research Institute, United States) Allison Rosenthal (Mayo Clinic Health System, United States) Andrew Bellizzi (University of Iowa, United States) Vincent Liu (University of Iowa, United States) Brian Link (University of Iowa, United States) Aaron Mangold (Mayo Clinic Arizona, United States) Ali Jabbari (Seattle Children's Research Institute, United States)

Abstract:
Cutaneous T cell lymphomas (CTCLs) are a spectrum of diseases with varied clinical courses caused by malignant clonal proliferation of skin-tropic T cells. Most patients have an indolent disease course managed with skin-directed therapies. In contrast, others, especially in advanced stages of disease or with specific forms, have aggressive progression and poor median survival. Sézary syndrome (SS), a leukemic variant of CTCL, lacks highly consistent phenotypic and genetic markers that may be leveraged to prevent the delay in diagnosis experienced by most patients with CTCL and could be useful for optimal treatment selection. Using single-cell mRNA and T-cell receptor sequencing of peripheral blood immune cells in SS, we extensively mapped the transcriptomic variations of nearly 50 000 T cells of both malignant and nonmalignant origins. We identified potential diverging SS cell populations, including quiescent and proliferative populations shared across multiple patients. In particular, the expression of *AIRE* was the most highly upregulated gene in our analysis, and *AIRE* protein expression could be observed over a variety of CTCLs. Furthermore, within a single patient, we were able to characterize differences in cell populations comparing malignant T cells over the course of treatment with histone deacetylase inhibition and photopheresis. New cellular clusters after progression on the therapy notably exhibited increased expression of the transcriptional factor *FOXP3*, a master regulator of regulatory T cell function, raising the potential implication of an evolving mechanism of immune evasion.

Conflict of interest: No COI declared

COI notes:

Preprint server: No;

Author contributions and disclosures: NB processed the data, analyzed the data and wrote the manuscript. K.J.S., N.H., L.S.O collected and processed clinical samples. A.C.R., provided clinical insight and assistance in recruiting patients. A.M.B. and V.L. performed IHC staining and quantification, provided clinical insight, and assisted in drafting manuscript. B.K.L provided clinical insight and assisted in drafting manuscript. A.R.M. and A.J. designed research, assisted in drafting manuscript, and supervised the work.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: Raw and processed data are available at GEO accession: GSE124899 and GSE146586

Clinical trial registration information (if any):

Title: Single-cell analysis of Sézary syndrome reveals novel markers and shifting gene profiles associated with treatment

Authors: Nicholas Borcherding[1,2], Kevin J. Severson[3], Nicholas Henderson[4], Luana S. Ortolan[4,5], Allison C. Rosenthal[9], Andrew M. Bellizzi[1], Vincent Liu[1,4,6], Brian K. Link[6,7], Aaron R. Mangold[3], Ali Jabbari[1,4,6,8*]

Affiliations:

[1] Department of Pathology, University of Iowa, Iowa City, IA, 52242

[2] Department of Pathology and Immunology, Washington University, St Louis, MO, 63110

[3] Department of Dermatology, Mayo Clinic, Scottsdale, AZ, 85259

[4] Department of Dermatology, University of Iowa, Iowa City, IA, 52242

[5] Seattle Children's Research Institute, Seattle, WA, 98101

[6] Holden Comprehensive Cancer Center, University of Iowa, Iowa City, IA, 52242

[7] Department of Internal Medicine, University of Iowa, Iowa City, IA, 52242

[8] Iowa City Veterans Affairs Medical Center, Iowa City, IA 52242

[9] Department of Hematology Oncology, Mayo Clinic, Phoenix, AZ, 85054

*Correspondence: ali-jabbari@uiowa.edu (A.J.)

Key points

Integrated clonotype and mRNA expression levels in 6 patients, identifying common and divergent transcriptional states.

Characterization of transcriptional changes with progression on photopheresis and vorinostat.

Abstract

Cutaneous T cell lymphomas (CTCLs) are a spectrum of diseases with varied clinical courses caused by malignant clonal proliferation of skin-tropic T cells. Most patients have an indolent disease course managed with skin-directed therapies. In contrast, others, especially in advanced stages of disease or with specific forms, have aggressive progression and poor median survival. Sézary syndrome (SS), a leukemic variant of CTCL, lacks highly consistent phenotypic and genetic markers that may be leveraged to prevent the delay in diagnosis experienced by most patients with CTCL and could be useful for optimal treatment selection. Using single-cell mRNA and T-cell receptor sequencing of peripheral blood immune cells in SS, we extensively mapped the transcriptomic variations of nearly 50 000 T cells of both malignant and nonmalignant origins. We identified potential diverging SS cell populations, including quiescent and proliferative populations shared across multiple patients. In particular, the expression of *AIRE* was the most highly upregulated gene in our analysis, and *AIRE* protein expression could be observed over a variety of CTCLs. Furthermore, within a single patient, we were able to characterize differences in cell populations comparing malignant T cells over the course of treatment with histone deacetylase inhibition and photopheresis. New cellular clusters after progression on the therapy notably exhibited increased expression of the transcriptional factor *FOXP3,* a master regulator of regulatory T cell function, raising the potential implication of an evolving mechanism of immune evasion.

Introduction

Cutaneous T cell lymphomas (CTCL) are a group of heterogeneous malignancies of T cell origin that traffic to the skin. Mycosis fungoides (MF) and Sézary syndrome (SS) are CTCLs comprised of skin-tropic CD4[+] T cells with varying degrees of blood involvement [1,2]. Despite the clonal nature of both MF and SS, the course of the diseases can be highly variable, with a subset possessing a relatively indolent disease course whereas other patients manifest widespread disease burden beyond the skin; in particular, SS patients exhibit prominent blood involvement and carry a poor prognosis. Advanced stages of CTCL have a 5-year survival as low as 24% [1], with patients often cycling through therapies that do not offer long-term durable responses. Thus, a deeper understanding of the CTCL disease process and determinants of tumor behavior are critical unmet knowledge gaps that could assist in determining optimal therapeutic regimens for patients with SS and MF.

Although CTCL represents a clonal proliferation of malignant T cells, we and others have previously demonstrated single-cell transcriptional heterogeneity of the circulating malignant T cell population of SS patients [3,4]. This heterogeneity has also been reported from single-cell sequencing of CTCL skin lesions, demonstrating marked intertumoral and intratumoral transcriptional heterogeneity [5]. One theory for the differential presentation and course of CTCL for patients is the varied composition of distinct subsets of malignant T cells [2]. Single-cell sequencing enables the investigation of distinct subset analysis for the leukemic SS, which may shed light on niche development and potential vulnerabilities to specific treatments.

The transcriptional heterogeneity of CTCL may explain the extensive treadmilling of therapeutic response and recurrence. Histone deacetylase inhibitors (HDACi) are a mainstay of CTCL treatment and are thought to alter key genetic programs involved in tumor development and progression [6]. However, the trials for which the approval of HDACi was based found only 30-35% of CTCL patients respond to HDACi, with a median duration of response of 13.7-15 months [7,8]. Interestingly, in-depth analysis of HDACi in SS patients found highly individualized

responses in not only T cell number but variable genomic and mRNA expression changes [9,10]. Brentuximab vedotin (BV), another therapy for CTCL, consists of a chimeric antibody directed against CD30, a cell marker variably found in MF/SS, conjugated to monomethyl auristatin E, an inhibitor of microtubule polymerization and hydrolysis. Studies examining the efficacy of BV in advanced-stage CTCL have indicated that there is no correlation between tumoral CD30 expression and time to response, duration of response, progression-free survival, and event-free survival [11,12] supporting that other patients or tumor-specific features play a role. These data underscore the need for a greater understanding of the maintenance of the malignant potential of CTCL and the need for better predictive biomarkers in CTCL patients undergoing treatment.

We use single-cell mRNA and T cell receptor (TCR) sequencing across isolated CD45[+] immune cells in the peripheral blood of six SS patients. With over 50,000 single-cells, our studies revealed extensive common and differential gene expression markers and cellular pathway alterations within malignant SS T cells. In addition to this analysis, our study uncovered alterations at the single-cell level of a SS patient during treatment with HDACi and photopheresis through therapeutic failure and disease progression, offering a novel insight into resistance mechanisms for combinatorial therapy. Therefore, these data indicate that determinations of tumor heterogeneity and composition by transcriptionally distinct malignant cell subpopulations may hold predictive value in treatment response and utility in clinical decision-making and therapeutic choice.

Methods

Patient Recruitment

The current study was approved by the University of Iowa or the Mayo Clinic Institutional Review Board and conducted under the Declaration of Helsinki Principles. The patients were recruited from the Department of Dermatology Cutaneous Lymphoma clinic at the University of

Iowa and the Cutaneous Lymphoma Clinic at the Mayo Clinic in Scottsdale, Arizona. Samples from the Mayo Clinic were shipped overnight at 4° C to the University of Iowa. Informed written consent was received from the participant before inclusion in the study.

Flow Cytometry

Blood draws were performed, and peripheral blood mononuclear cells were isolated using a Ficoll gradient. Cells were labeled for CD45 and flow-sorted to isolate immune cells on a Becton Dickinson Aria II. The Patient 1 sample isolation involved flow-sorted peripheral blood CD4[+] T cells only, and the baseline data were previously described [3].

Single-Cell RNA Sequencing

Sequencing for 5' gene expression and T-cell receptor was performed using the Chromium (10x Genomics, Pleasanton, CA) and Illumina (San Diego, CA) sequencing technologies. Amplified cDNA was used to construct both 5' expression and TCR enrichment libraries. Libraries were pooled and run on separate lanes on an Illumina HiSeq 4000. Each lane consisted of 150 bp paired-end reads. Basecalls were converted into FASTQs using the Illumina bcl2fastq. FASTQ files were aligned to the human genome (GRCh38) using the CellRanger v3.0.1 pipeline as described by the manufacturer. The TCR V(D)J sequences were aligned to the vdj_GRCh38_alts_ensembl genome build provided by the manufacturer.

Single-cell data processing and analysis

Initial processing of immune cells from peripheral blood of six patients used the Seurat R package (v3.0.2). Single cells used in subsequent analysis were comprised of CTCL1 (n=7,828), CTCL2 (n=32,592), CTCL3 (n=2,250), CTCL4 (n=1,479), CTCL5 (n=1,090) and CTCL6 (n=2,677). CTCL1 sequencing data was derived from the previous publication[3] and included as part of the baseline/integrated data. To this sample, we sequenced a second time

point after progression on vorinostat (n=7,127) which has not been previously published. Samples were combined into a single Seurat object using canonical correlational analysis and mutual nearest neighbors (MNN) [13,14]. Dimensional reduction to form the UMAP plots utilized the top 40 calculated dimensions and a resolution of 0.5. Cluster markers and differential gene expression analyses were performed using the Wilcoxon rank-sum test with an unsupervised approach involving no gene filtering. *FOXP3*+ cells were defined as individual cells with *FOXP3* counts greater than or equal to 1. Percent expression graphs were created using the schex R package (v1.1.5, development version), setting nbins equal to 80 and in the *make_hexbin()* call. Cell trajectories were created using the monocle (v2.10.1) R package [15] using the top 4,750 genes for ordering and the DDRtree dimensional reduction strategy. In addition, the individual time points were used as the residual model string during the reduction of dimensions to eliminate differences based on batch effect. Cell cycle scoring was performed on single-cells using the previously described gene set [16] and the CellCycleScoring function in the Seurat package.

Single-cell immune phenotyping utilized the SingleR (v1.0.1) R package [17]; the cell-type based correlations and single-cell expression values were compared to transcriptional profiles from pure cell populations in Human Primary Cell Atlas [18]. Single-sample Gene Set Enrichment Analysis (ssGSEA) utilized the escape R package [19]. Individual gene sets were derived from GSEA Hallmark library [20], Kyoto encyclopedia of genes and genome [21], and Biocarta. Expression data were visualized using the ggplot2 (v3.2.1) and pheatmap (v1.0.12) R packages. TCR sequencing results were processed using the scRepertoire (v1.0.0) [22]. Malignant clones were defined as clones with immune repertoire occupancy greater than 0.1. Genes upregulated at time point 2 (n=1,428) underwent gene set enrichment analysis using the Enrichr software [23] using the same libraries as above. Results with adjusted p-value < 0.25 are

presented in supplemental Table 6. Processed data is available in the GEO accession: GSE124899 and GSE146586.

Bulk-Sequencing analysis.

Raw expression data for GSE113113 [24] were downloaded from the NCBI Sequence Read Archive and converted to FASTQ files using the SRA toolkit. Samples were aligned with the kallisto pseudoalignment [25] and GRCh38 build for the human genome, aggregating the estimated counts for transcripts into gene-level quantifications based on Human Genome Organization Gene Nomenclature Committee gene symbol. Sample expression values were processed using the Sleuth R Package (v0.30.0). ssGSEA for the bulk samples was performed using the GSVA (v1.30.0) R package using the Poisson function for cumulative distribution. Immune receptor quantification for bulk RNA sequencing was performed using MiXCR using the default pipeline [26]. Recovered immune receptor reads were then filtered for the dominant T cell receptor V gene, corresponding to either αβ or γδ T cells, and normalized by the entire recovered immune receptor repertoire.

Immunohistochemistry

Immunohistochemistry was performed on a Dako Autostainer Link 48 after deparaffinization, rehydration, and heat-induced epitope retrieval (HIER) on a Dako PT Link. All antibodies were retrieved in Dako Target Retrieval Solution, HpH (EDTA-based). All primary antibody incubations were of a 30-minute duration. The polymer-based Dako EnVision FLEX kit was used for detection; all detections were 30 minutes. Diaminobenzidine was used as the chromogen. Slides were subsequently lightly counterstained with Harris hematoxylin, dehydrated, and coverslipped for pathologist review. The primary antibodies used in this study included CD3 (Dako; polyclonal; 1:800), CD4 (Novocastra; clone 4B12; 1:100), and AIRE (Invitrogen; clone 5H12; 1:50). A multi-tissue block including tonsil was used as the positive

tissue control for CD3 and CD4, while thymus was used for AIRE. Nuclear AIRE staining was quantified as mean positive cells across a minimum of five high-power fields.

Statistical Analysis

Statistical analyses were performed in R (v3.6.1). Two-sample significance testing utilized Welch's T-test, with significance testing for more than three samples using one-way analysis of variance (ANOVA) with Tukey honest significance determination for correcting multiple comparisons. Unless otherwise noted in the figure legend, significance for gene expression is based on the cut-off of log-fold change ≥ 1 or ≤ -1 and an adjusted p-value < 0.05. Single-cell relative percentages by categorical variables used the total number of cells in the indicated sequencing run as a denominator for normalization. Analysis of the distribution of cells across multiple categorical variables utilized the χ^2 test.

Results

Sequencing results of the peripheral blood of 6 SS patients.

A total of 54,994 cells were sequenced and passed, filtering from the peripheral blood of six separate patients with SS (Figure 1A). Patient information and a summary of the sequencing results are available in Supplemental Table 1. Based on mRNA expression, we observed 16 distinct clusters. Using both percent of cells expressing lineage markers (Figure 1B) and correlations (Figure 1C) based on the human primary cell atlas, we identified 11 T cell clusters (C1, C2, C3, C6, C7, C8, C9, C10, C12, and C15), 1 B cell clusters (C13), and 3 myeloid cell clusters (C5, C11, and C14). The majority of the sequenced peripheral cells were T cells, with a mean percent of T cells per patient equal to 76.9 ±15.2% (Figure 1D). Myeloid and B cells comprised a mean percentage of 19.2±12.7% and 3.9±3.9%, respectively Each patient sample varied in the distribution of cell types (Figure 1E). Notably, the Patient 2 sample consisted of a

relatively high number of T cells in the C2, C3, and C6 T cell clusters (Figure 1E), while the

Patient 5 and 6 samples had a wide distribution of cells across all three cell types (Figure 1E).

Defining malignant T cells in the peripheral blood of SS patients

Across the 11 T cell clusters (Figure 2A), we next explored the expression patterns of

common T cell and CTCL markers (Figure 2B). Although all clusters expressed *CD3D* and *CD5*,

pan-T cell genes, there was delineation in the expression of *CD4* (high clusters: C1, C2, C3,

and C6) and *CD8A* (high clusters C8; moderate clusters C4, C7, C9, C10). A general absence

of *CD7*, a common feature in CTCL, was observed in clusters C1-3 and C6 (Figure 2B). These

clusters possessed prominent expression of *CCR4*, *CXCL13*, and PD-1 (*PDCD1*) (Figure 2B),

genes previously associated with SS [3,5,27,28]. Accompanying the mRNA classification of

malignant and nonmalignant T cells, we also used the V(D)J single-cell sequencing results for

the TCR, recovering the sequences for 43,074 T cells (89.9%). Using the scRepertoire [22] R

package, we defined clonotypes as both the sequence of genes and the nucleotide sequence of

both the TCR A and B chains (Supplemental Figure 1A); a variety of unique clonotypes were

isolated across the six samples (Supplemental Figure 1B). The amount of unique clonotypes

sequenced did not predict the clonal homeostasis or occupied repertoire space for each sample,

with 95.2% of Patient 2 T cells formed from the dominant clonotype for that sample despite

having 1,254 unique clonotypes (Supplemental Figure 1C, D).

T cell clonal frequency was then placed into categories by copies within the patient

sample. TCR clonal categories were organized by proportion of total productive TCR reads per

patient with hyperexpanded ($0.1 < X \leq 1$), large ($0.01 < X \leq 0.1$), medium ($0.001 < X \leq 0.01$),

small ($1e-4 < X \leq 0.001$) and rare clonotypes ($0 < X \leq 1e-4$). The C1, C2, C3, C10, C12, and

C15 clusters possessed T cells with greater than 90% hyperexpanded clonotypes, so-called

malignant (M) clusters (Figure 2C), with each sample having a single, dominant clone. Notably,

nonmalignant (NM) clusters were C4, C7, C8, and C9. To investigate potential novel markers or

therapeutic targets of the peripheral SS T cells, we next performed differential gene expression analysis across T cells with a hyperexpanded clonotype compared to T cells with single, small, medium, or large clonotypes. Unlike bulk sequencing approaches, SC mRNA sequencing allows for the comparison of differential genes in the context of the percent of cells expressing the gene or genes of interest. This allows for not only the difference in terms of log-fold change but also in terms of the difference in the percent of malignant versus nonmalignant T cells expressing the gene(s) (Δ Percent). Non-TCR genes with the highest log-fold change (LFC) and greatest discrimination between malignant versus nonmalignant T cells were *AIRE* (LFC = 2.42, Δ = 49.2%), *NEDD4L* (LFC = 2.39, Δ = 28%), *IGFBP4* (LFC = 2.28, Δ = 22.4%), *TUSC3* (LFC = 2.24, Δ = 16.7%), and *PDLIM1* (LFC = 2.20, Δ 28.1%) (Figure 2D,E). Conversely, *CD2* (LFC = -0.66, Δ=35.2%), CD7 (LFC = -2.21, Δ = 58.3%), and CD49f (*ITGA4*, LFC = 3.14, Δ = 30.3%) were among the most downregulated genes in the malignant SS T cells (Supplemental Figure 2). The complete results for the differential expression analysis are available in Supplemental Table 2. We also compared our list of differentially upregulated genes (n=284) to previously identified upregulated genes in peripheral blood of SS (n=53) [29]. The comparison yielded an overlap coefficient of 49.1% and 26 common genes shared between the analyses (Figure 2F). Using a 50-patient bulk RNA cohort of normal skin, MF, and SS biopsies [24], we first established the relationship of clonality and diversity in the samples (Supplemental Figure 3A-B). In this cohort, 237 of the 284 genes were expressed, and we found 72 of the 237 genes significantly and directly correlated with the proportion of dominant V-gene recovered, a marker of clonality (Supplemental Figure 3C). Of the genes with non-significant increased expression in the single-cell cohort and expressed in the bulk sequencing data (n=2,594), 532 genes were positively correlated with dominant V-gene proportion. Contingency comparison by Fischer's exact test showed a strong association between the single-cell differentially-expressed genes and correlations (p=6.47e-4). We also found that 48 of the 284 genes inversely correlated with the diversity of the immune repertoire (Supplemental Figure 3C). Top genes, ranked by LFC in both

comparisons (Figure 2F, left list) and by LFC in our cohort (Figure 2F, right list), are displayed.

Half of these top 20 genes were significantly correlated with clonality (directly) or diversity

(inversely) in the secondary bulk RNA cohort (Supplemental Figure 3D-E). We chose to survey

skin biopsies of MF (n=17), SS (n=5), and T-cell non-Hodgkin's lymphoma (NHL, n = 13)

patients using IHC for one of these novel markers AIRE (Figure 2G). Nuclear staining of AIRE

ranged across the patients, we found AIRE in 11 of 17 MF and 5 of 6 SS samples (Figure 2H).

Although the mean density of AIRE+ SS was varied, a comparison between CTCL staining

(MF+SS) and NHL produced a p-value of 0.005321.

Transcriptional heterogeneity within malignant and nonmalignant T cells.

We next wanted to understand the differences in the gene expression that produced the

different T cells clusters. Although the malignant clusters C1, C2, and C3 shared a number of

common gene markers described above (Figure 3A), other malignant clusters expressed unique

markers. For example, C12 had high levels of expression of *TOP2A* and *MKI67*, markers of cell

cycle activation, compared to any other T cell cluster (Figure 3A). The nonmalignant cluster, C8,

had high granzyme and perforin expression levels, consistent with cytotoxic functions (Figure

3A) and its high expression of *CD8A.* To better understand the possible implications of the

differential gene expression across the T cell clusters, we performed principal component

analysis (PCA) on scaled mRNA levels of genes involved in T cell development and

differentiation (Figure 3B) with the full list of genes available in Supplemental Table 3. In the

malignant C1, C3, and C12 clusters, the PCA-identified expression consisted of *GATA3*, *CCR7*,

and *TOX* expression (Figure 3B). The nonmalignant C8 cluster was also unique in the high

expression of cytotoxic markers, such as *GZMA*, *GZMB*, and *PRF1,* while detectable, albeit low

levels of expression of these mRNA species were only seen in the malignant C12 cluster

(Figure 3A). We also examined selected canonical gene markers of CTCL transcriptional

profiles. Among the skin trafficking markers, C12 and C15 had the highest expression of

SELPLG (CLA precursor) and the fucosyl-transferase, *FUT7* [30], although they possessed lower

chemokine receptor levels, *CCR4* (Figure 3C). Memory phenotypes have also previously been

described in CTCL [31,32]; high expression of *CCR7* across malignant T cells compared to

nonmalignant T cells is consistent with prior reports of SS cells either emerging from, or at the

least possessing, central memory T cell-like characteristics [33]. Although we have previously

described both clonal central memory and *FOXP3*$^+$ regulatory T cell (Treg) clusters examining a

single SS patient [3], the aggregated single-cell analysis failed to reveal a representative cluster

for Treg or Treg-like cells (Figure 3C), possibly due to the increased number of cells, the

inclusion of additional immune cell types, or drop-out effect of single-cell sequencing. With the

exception of CTCL3 and CTCL6 samples, *FOXP3*+ T cells were clustered across C1, C3, and

C4, forming a range of 2.7-38.9% of the cells in the clusters by patient ID. We also looked at

intra and inter-patient diversity using the gene expression count matrices and found the richness

of malignant gene expression in malignant T cells to generally be higher than non-malignant T

cells and varied within and across patients (Supplemental Figure 4). Beyond gene expression

and representativeness, we also investigated potential functional differences in terms of cell

cycle and signaling pathway enrichments as well as cell subset-specific signatures. Cell cycle

regression demonstrated key differences in cell cycle assignment (p < 2.2e-16, χ^2 test). C12

was highly proliferative, with 89.3% of cells in S or G2M phases (Figure 3D). In contrast, C10

was more quiescent in terms of cell cycle, with 20.6% of cells in S or G2M phases. The

remaining malignant clusters have more intermediate distributions within our cell cycle analysis

(Figure 3D).

To ensure our malignant clustering was representative of CTCL T cells, we isolated

highly expressed genes for each cluster to perform gene set enrichment on previously published

bulk CTCL RNA-seq results (Figure 3E) [24]. With the normal controls, we demonstrated that C3,

C6, and C12 signatures had the highest level of separation (Figure 3E). Conversely, the C10

signature had the opposite trend with lower enrichment in the SS/MF compared to normal

control skin. Using the percent of dominant V-gene recovered, a surrogate marker for clonality,

in the cohort, we found enrichment for C1, C2, C3, and C6 signature enrichment directly and

significantly correlated with clonality (Figure 3F). Both C12 and C15 had insignificant

correlations, with C12 trending towards direct correlation (Figure 3F); this may be a result of the

two clusters representing more rare populations in CTCL. Similarly, nonmalignant T cell

clusters, C4, C7, and C8, were insignificantly correlated with the dominant V gene

(Supplemental Figure 5). Only C10 gene set enrichment negatively correlated with clonality

(Figure 3F).

The C10 cluster was aberrant not only in global position on UMAP compared to other

CTCL clusters but also lacked consistent markers of CTCL (Figure 3C). Closer examination

found that of the hyperexpanded cells in C10, 63.6% of the cells in this cluster were derived

from CTCL4. Next, we performed gene set enrichment across several diverse cellular

processes to compare potentially functional differences of CTCL clusters and nonmalignant

peripheral blood T cells (Supplemental Table 4). Gene set enrichment results with column

clustering based on the Manhattan distance found 3 distinct enrichment patterns. The C2-C6-

C10 patterns had enrichment in FOS signaling and complement activation (Figure 3G). C10

also displayed unique enrichment in amino acid and nitrogen metabolism (Figure 3G).

Nonmalignant and C1 clusters had unremarkable enrichment across pathways. The C3-C12-

C15 pattern consisted of elevated enrichment of downstream signaling from cytokines (Figure

3G). For example, C15 had unique enrichment of ALK, eicosanoid, and cytokine signaling, while

C12 had unique enrichment for DNA replication, epithelial-to-mesenchymal transition, and

resistance to histone-deacetylase (HDAC) inhibition (Figure 3G). The latter is particularly

interesting as HDAC inhibition (HDACi) represents a major treatment paradigm for CTCL.

High-resolution mRNA expression changes from HDAC inhibitor therapy.

At a single-cell resolution, we have previously characterized the T cell expression heterogeneity within a 61-year old male patient with stage 4 (T4N1M0B2) SS Field 3 under treatme a HDACi. The previously published work loont with photopheresis and vorinostatked at flow-sorted peripheral blood malignant (CD3$^+$CD4$^+$CD5^{bright}SSC^{hi}) and nonmalignant (CD3$^+$CD4$^+$CD5^{int}SSC^{int}) T cells at an initial time point during combined HDACi and photopheresis (n=7,954) therapy. Adding to the previous data, we sequenced the same populations of peripheral blood 9 months later following disease progression (n=7,127) and integrated the single-cell sequencing runs into a single UMAP projection (Figure 4A). The result was 11 distinct clusters (Figure 4A) with malignant/nonmalignant flow-based phenotypic classifications (Figure 4B, upper panel) and distinct time points (Figure 4B, lower panel). Most clusters represented mixes of baseline (T1) and progressed (T2) time points (Figure 4B, lower panel); however, Clusters 6 and 9 were solely comprised of T cells from the T2 time point. TCR sequencing also enabled the assignment of clonotype grouping similar to the previous cohort analysis (Figure 4C). Malignant clusters 0, 3, 6, 8, 9, and 10 were defined as possessing greater than 50% of T cells with hyperexpanded clonotypes (Figure 4C).

We also examined the expression distribution of common CTCL markers (Figure 4D), finding a subset of both nonmalignant and malignant T cells with expression of CD5, a feature of our patient previously described both at the mRNA level and protein level [3]. Comparing the T cells with hyperexpanded clonotypes from T2 to T1, we found a total of 278 genes with significantly altered expression of adjusted p-value < 0.05 and 1- > LFC < 1 (Figure 4E). The secondary time point had pronounced decreases across *JUN/FOS* genes, previously implicated in response to HDACi [9] (Figure 4E). Genes that were upregulated after progression on the combined therapy were diverse and included immune mediators and transcription factors (i.e., *ITGA4*, *CXCR1*, *FOXP3*, and *TIGIT*), chromatin modifiers (*KLF10*, *MCRS1*, and *SMARCB1*), and mitochondrial transporters (*SLC25A26*) (Figure 4F). A complete list of differentially regulated genes is available in Supplemental Table 4. When ssGSEA was performed comparing

pathway enrichment for T1 and T2, we found an overall increase in metabolic gene sets, such as fatty acid oxidation, glycolysis, and oxidative phosphorylation (Figure 4G). Conversely, we found decreased levels in gene set enrichment associated with immune ligands, such as CCR5, IL-6, and interferons (Supplemental Table 5).

Defining altered expression dynamics in SS pre- and post-HDACi therapy using single cells.

Focusing on the malignant SS cells, we defined the clusters by time point predominance with clusters 6 and 9 new to T2 and cluster 0 (67.3% T1) and cluster 8 (72.1% T1) enriched at T1 (Figure 5A). We next attempted to characterize the gene expression differences between the clusters, finding a high degree of gene expression overlap between clusters (Figure 5B). Within Cluster 0, increased JUN and LAIR2, recently associated with HDACi resistanceField 10, was observed. With our previous characterization, we noted a single isolated cluster of *FOXP3*-high malignant T cells [3] corresponding to Cluster 3 in our new analysis. T2 had the emergence of 2 additional clusters, 6 and 9, with moderate levels of *FOXP3* expression, and cluster 10, with low levels of expression (Figure 5B). Cluster 8, the only T1-predominant cluster, had high cysteine dioxygenase type I (CDO1) gene levels. In addition to gene expression analysis, we also performed ssGSEA and examined the distribution of the malignant SS T cells and principal component analysis. Notably, the T2-predominant clusters 6 and 9 had greater representation in the lower right quadrant associated with the metabolic pathway enrichment previously mentioned (Figure 5C). In contrast, Cluster 8 had preferential enrichment in growth factor and cytokine signaling (Figure 5C).

In order to better understand the transition of the malignant SS T cells at baseline and following progression on HDACi therapy, we used reverse-graph embedding [15] to construct a cellular trajectory manifold that can identify distinct cellular fates or transcriptional states (Figure 5D). Defining each branch of the manifold as transcriptional states, we found four distinct states with a clear separation in relative percent of cells by time point. The T1 states were defined as

State 2 and State 4, while State 1 and State 3 were formed from the majority of T2 malignant

SS cells (Figure 5D). Differences were also noted in the placement of time point cells along the

manifold with T2 cells of the State 3 and State 4 branches at the distal portions compared to the

more proximal T1 cells. We performed differential gene expression analysis for each

transcriptional state, comparing the state to the rest of the cells (Figure 5E). Most interestingly,

State 1, comprised of the *FOXP3*+ clusters 3, 6, and 9, was a T2-predominant transcriptional

state with multiple clusters with *FOXP3*, potentially implicating a therapy-induced genetic

program. We next wanted to understand how the branches of the manifold contribute to the

resistance to therapy. Using recent gene expression results for patients sensitive and resistant

to HDACi [10], we compared differential gene expression between all four states and found a total

of 33 genes previously reported (Figure 5F). Of the 1,395 genes that were altered that were not

previously reported, the preponderance of the most significant genes was increased in the State

1 cells (Figure 5G). It should be noted that these genes may represent new mechanisms of

resistance, underlying genetic changes induced by HDACi, or a combination of the two.

Towards understanding the functional import of these genes, we performed gene set

enrichment analysis (Supplemental Figure 6). Among the most significantly enriched gene sets

were amino and fatty acid metabolism, DNA repair pathways, and cytokine signaling.

Discussion

Despite being a clonal T cell neoplasm, we and others have reported that a degree of

epigenomic, transcriptional, and protein heterogeneity exists in both SS and MF [3–5,34,35]. Unlike

previous single-cell quantifications of CTCL, we integrated our transcriptomic data across

patients, allowing for the quantification of CTCL heterogeneity across patients. Indeed,

leveraging the single-cell mRNA and TCR sequencing of circulating nonmalignant and SS cells

for assessment of clonality, we find a high degree of inter and intra-patient heterogeneity (Figure

2,3) across over 50,000 cells. This variability may contribute to the discordance in the clinical or

histopathological definition of CTCL, prolonged median time to diagnosis of 3 years, and therapeutic treadmilling observed inpatient management [36,37]. A new report has even shown distinct transcriptional signatures of skin-derived and blood-derived MF cells at the level of single patients [38]. Despite this heterogeneity, we found common SS markers that may warrant further investigation, including several previously identified, such as *NEDD4L*[29,39], *PLS3*[40], and *TOX*[41,42]. Notably, the *AIRE* gene was expressed across samples in 58% of malignant cells versus 8.7% of nonmalignant cells. The largest difference for non-TCR-related genes and only recently *AIRE* expression was incidentally reported in circulating mycosis fungoides cells [43]. The autoimmune regulator, AIRE, functions in central tolerance during T cell maturation and is expressed by thymic epithelial and dendritic cells[44]. Loss of AIRE leads to autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy, a pleiotropic genetic immune dysfunction characterized by the loss of thymus-generated Tregs [45]. The role of AIRE in the ectopic expression of peripheral antigens raises the possibility that its expression by CTCL cells may function to recruit and stimulate a wider array of regulatory T cells, potentially as a mechanism to avoid immune responses.

Within the CTCL literature, there is a lack of consensus regarding the existence of a FOXP3+ regulatory or regulatory-like phenotype of malignant T cells [46–49]. This variation in the literature may be a result of the regulation of FOXP3 expression by environmental factors, such as bacterial toxin [50] or therapeutic regimens. In terms of the latter, the use of HDACi led to the development of *FOXP3*-moderate-expressing clusters (labeled 6 and 9 in Figure 5F). This is underscored by previous findings from a single-cell ATAC-seq study, which reported differential accessibility of FOXP3 in clonal SS cells [35]. Upregulation of HDAC9 has been previously reported in CTCL [9,51], which along with HDAC6, has been shown to regulate the expression of *FOXP3* in regulatory T cells [52]. A subset of HDACi-responsive CTCL patients had an opening of the chromatin associated with the FOXP3 enhancers, suggesting an increase in *FOXP3* expression may be a surrogate marker of response to HDAC inhibition [9].

We previously showed that *FOXP3* was the single strongest predictor of early-stage disease in a cohort of 152 CTCL patients [3,53,54]. With transcriptional State 1, a post-HDACi-predominant state, consisting of multiple populations of *FOXP3*+ malignant SS cells, further research is warranted to distinguish whether therapy selects for diverging populations of FOXP3+ or pushes the transcriptional state to that of earlier-stage disease. The latter has been suggested as HDACi increases chromatin access to the *FOXP3* loci [9]. However, two major caveats to this concept include 1) the shift in the transcriptional state may depend on the specific HDACi agent and 2) further work is needed to understand if this is truly a transcriptional state versus epigenomic shift. To the former, vorinostat inhibits HDAC class I, II (to include HDAC9), and IV, while the potent bicyclic romidepsin is more specific for HDAC class I molecules. Key differences in HDACi-agent-specific effects on genome accessibility, with romidepsin preferentially altering promoters and active enhancer regions compared to vorinostat, likely alters gene expression [9]. This differential effect of HDACi on heterogeneity via flow cytometry has been previously reported with *ex vivo* treatment of SS with vorinostat and romidepsin [4].

Acknowledgments

Authorship

Contribution: NB processed the data, analyzed the data and wrote the manuscript. K.J.S., N.H., L.S.O collected and processed clinical samples. A.C.R., provided clinical insight and assistance in recruiting patients. A.M.B. and V.L. performed IHC staining and quantification, provided clinical insight, and assisted in drafting manuscript. B.K.L provided clinical insight and assisted in drafting manuscript. A.R.M. and A.J. designed research, assisted in drafting manuscript, and supervised the work.

Declaration of Interests: The authors declare no competing interests.

References

1. Willemze R, Jaffe ES, Burg G, et al. WHO-EORTC classification for cutaneous lymphomas. *Blood*. 2005;105(10):3768–3785.

2. Kirsch IR, Watanabe R, O'Malley JT, et al. TCR sequencing facilitates diagnosis and identifies mature T cells as the cell of origin in CTCL. *Sci. Transl. Med.* 2015;7(308):.

3. Borcherding N, Voigt AP, Liu V, et al. Single-cell profiling of cutaneous T-cell lymphoma reveals underlying heterogeneity associated with disease progression. *Clin. Cancer Res.* 2019;25(10):2996–3005.

4. Buus TB, Willerslev-Olsen A, Fredholm S, et al. Single-cell heterogeneity in Sézary syndrome. *Blood Adv.* 2018;2(16):2115–2126.

5. Gaydosik AM, Tabib T, Geskin LJ, et al. Single-cell lymphocyte heterogeneity in

advanced cutaneous T-cell lymphoma skin tumors. *Clin. Cancer Res.* 2019;25(14):4443–4454.

6. Bolden JE, Peart MJ, Johnstone RW. Anticancer activities of histone deacetylase inhibitors. *Nat. Rev. Drug Discov.* 2006;5(9):769–784.

7. Whittaker SJ, Demierre M-F, Kim EJ, et al. Final results from a multicenter, international, pivotal study of romidepsin in refractory cutaneous T-cell lymphoma. *J Clin Oncol*. 2010;28(29):4485–4491.

8. Piekarz RL, Frye R, Turner M, et al. Phase II multi-institutional trial of the histone deacetylase inhibitor romidepsin as monotherapy for patients with cutaneous T-cell lymphoma. *J. Clin. Oncol.* 2009;27(32):5410.

9. Qu K, Zaba LC, Satpathy AT, et al. Chromatin Accessibility Landscape of Cutaneous T Cell Lymphoma and Dynamic Response to HDAC Inhibitors. *Cancer Cell*. 2017;32(1):27–41.

10. Andrews JM, Schmidt JA, Carson KR, et al. Novel cell adhesion/migration pathways are predictive markers of HDAC inhibitor resistance in cutaneous T cell lymphoma. *EBioMedicine*. 2019;46:170–183.

11. Duvic M, Tetzlaff MT, Gangar P, et al. Results of a phase II trial of brentuximab vedotin for CD30+ cutaneous T-cell lymphoma and lymphomatoid papulosis. *J. Clin. Oncol.* 2015;10(33):3759–65.

12. Kim YH, Tavallaee M, Sundram U, et al. Phase II investigator-initiated study of brentuximab vedotin in mycosis fungoides and Sézary syndrome with variable CD30 expression level: A multi-institution collaborative project. *J. Clin. Oncol.* 2015;33(32):3750–8.

13. Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. *Cell*. 2019;177(7):1888–1902.

14. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell

transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 2018;36(5):411–420.

15. Qiu X, Mao Q, Tang Y, et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods*. 2017;14(10):.

16. Nestorowa S, Hamey FK, Pijuan Sala B, et al. A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood*. 2016;128(8):e20–e31.

17. Aran D, Looney AP, Liu L, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* 2019;20(2):163–172.

18. Mabbott NA, Baillie JK, Brown H, Freeman TC, Hume DA. An expression atlas of human primary cells: Inference of gene function from coexpression networks. *BMC Genomics*. 2013;14(1):632.

19. Borcherding N, Vishwakarma A, Voigt AP, et al. Mapping the immune environment in clear cell renal carcinoma by single-cell genomics. *Commun. Biol.* 2021;4(1):1–11.

20. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 2005;102(43):15545–15550.

21. Ogata H, Goto S, Sato K, et al. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 1999;27(1):29–34.

22. Borcherding N, Bormann NL. scRepertoire: An R-based toolkit for single-cell immune receptor analysis. *F1000Research*. 2020;9:47.

23. Kuleshov M V., Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 2016;44(W1):W90–W97.

24. Querfeld C, Leung S, Myskowski PL, et al. Primary t cells from cutaneous t-cell lymphoma skin explants display an exhausted immune checkpoint profile. *Cancer Immunol. Res.* 2018;6(8):900–909.

25.   Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq
        quantification. *Nat. Biotechnol.* 2016;34(5):525–527.

26.   Bolotin DA, Poslavsky S, Mitrophanov I, et al. MiXCR: Software for comprehensive
        adaptive immunity profiling. *Nat. Methods*. 2015;12(5):380–1.

27.   Picchio MC, Scala E, Pomponi D, et al. CXCL13 is highly produced by Sézary cells and
        enhances their migratory ability via a synergistic mechanism involving CCL19 and CCL21
        chemokines. *Cancer Res.* 2008;68(17):7137–7146.

28.   Çetinözman F, Jansen PM, Vermeer MH, Willemze R. Differential expression of
        Programmed Death-1 (PD-1) in Sézary syndrome and mycosis fungoides. *Arch.*
        Dermatol. 2012;148(12):1379–1385.

29.   Booken N, Gratchev A, Utikal J, et al. Sézary syndrome is a unique cutaneous T-cell
        lymphoma as identified by an expanded gene signature including diagnostic marker
        molecules CDO1 and DNM3. *Leukemia*. 2008;22(2):393–399.

30.   Ferenczi K, Fuhlbrigge RC, Pinkus JL, Pinkus GS, Kupper TS. Increased CCR4
        expression in cutaneous T cell lymphoma. *J. Invest. Dermatol.* 2002;119(6):1405–1410.

31.   Campbell JJ, Clark RA, Watanabe R, Kupper TS. Sézary syndrome and mycosis
        fungoides arise from distinct T-cell subsets: A biologic rationale for their distinct clinical
        behaviors. *Blood*. 2010;116(5):767–771.

32.   Clark RA, Watanabe R, Teague JE, et al. Skin effector memory T cells do not recirculate
        and provide immune protection in alemtuzumab-treated CTCL patients. *Sci. Transl. Med.*
        2012;4(117):117ra7-117ra7.

33.   Clark RA, Shackelton JB, Watanabe R, et al. High-scatter T cells: A reliable biomarker for
        malignant T cells in cutaneous T-cell lymphoma. *Blood*. 2011;117(6):1966–1976.

34.   Stoeckius M, Hafemeister C, Stephenson W, et al. Simultaneous epitope and
        transcriptome measurement in single cells. *Nat. Methods*. 2017;14(9):865–868.

35.   Satpathy AT, Saligrama N, Buenrostro JD, et al. Transcript-indexed ATAC-seq for

precision immune profiling. *Nat. Med.* 2018;24(5):580–590.

36. Pimpinelli N, Olsen EA, Santucci M, et al. Defining early mycosis fungoides. *J. Am. Acad. Dermatol.* 2005;53(6):1053–1063.

37. Scarisbrick JJ, Quaglino P, Prince HM, et al. The PROCLIPI international registry of early-stage mycosis fungoides identifies substantial diagnostic delay in most patients. *Br. J. Dermatol.* 2019;181(2):350–357.

38. Herrera A, Cheng A, Mimitou EP, et al. Multimodal single-cell analysis of cutaneous T cell lymphoma reveals distinct sub-clonal tissue-dependent signatures. *Blood*. 2021;

39. Wang Y, Su M, Zhou LL, et al. Deficiency of SATB1 expression in Sézary cells causes apoptosis resistance by regulating FasL/CD95L transcription. *Blood*. 2011;117(14):3826–3835.

40. Su MW, Dorocicz I, Dragowska WH, et al. Aberrant Expression of T-Plastin in Sezary Cells. *Cancer Res.* 2003;63(21):7122–7127.

41. Zhang Y, Wang Y, Yu R, et al. Molecular markers of early-stage mycosis fungoides. *J. Invest. Dermatol.* 2012;132(6):1698–1706.

42. Huang Y, Su MW, Jiang X, Zhou Y. Evidence of an oncogenic role of aberrant TOX activation in cutaneous T-cell lymphoma. *Blood*. 2015;125(9):1435–1443.

43. Rindler K, Bauer WM, Jonak C, et al. Single-Cell RNA Sequencing Reveals Tissue Compartment-Specific Plasticity of Mycosis Fungoides Tumor Cells. *Front. Immunol.* 2021;12:1257.

44. Anderson MS, Su MA. AIRE expands: New roles in immune tolerance and beyond. *Nat. Rev. Immunol.* 2016;16(4):247–258.

45. Laakso SM, Laurinolli TT, Rossi LH, et al. Regulatory T cell defect in APECED patients is associated with loss of naive FOXP3+ precursors and impaired activated population. *J. Autoimmun.* 2010;35(4):351–357.

46. Berger CL, Tigelaar R, Cohen J, et al. Cutaneous T-cell lymphoma: Malignant

proliferation of T-regulatory cells. *Blood*. 2005;105(4):1640–1647.

47.   Tiemessen MM, Mitchell TJ, Hendry L, et al. Lack of suppressive CD4+CD25+FOXP3+ T
        cells in advanced stages of primary cutaneous T-cell lymphoma. *J. Invest. Dermatol.*
        2006;126(10):2217–23.

48.   Krejsgaard T, Gjerdrum LM, Ralfkiaer E, et al. Malignant Tregs express low molecular
        splice forms of FOXP3 in Sézary syndrome. *Leukemia*. 2008;22(12):2230–2239.

49.   Gjerdrum LM, Woetmann a, Odum N, et al. FOXP3+ regulatory T cells in cutaneous T-
        cell lymphomas: association with disease stage and survival. *Leuk. Off. J. Leuk. Soc.*
        Am. Leuk. Res. Fund, U.K. 2007;21(12):2512–2518.

50.   Willerslev-Olsen A, Buus TB, Nastasi C, et al. Staphylococcus aureus enterotoxins
        induce FOXP3 in neoplastic T cells in Sézary syndrome. *Blood Cancer J. 2020 105*.
        2020;10(5):1–11.

51.   Lee CS, Ungewickell a., Bhaduri a., et al. Transcriptome sequencing in Sezary
        syndrome identifies Sezary cell and mycosis fungoides-associated lncRNAs and novel
        transcripts. *Blood*. 2012;120(16):3288–3297.

52.   Beier UH, Wang L, Han R, et al. Histone deacetylases 6 and 9 and sirtuin-1 control
        Foxp3+ regulatory T cell function through shared and isoform-specific mechanisms. *Sci.*
        Signal. 2012;5(229):ra45.

53.   Lefrançois P, Tetzlaff MT, Moreau L, et al. TruSeq-Based Gene Expression Analysis of
        Formalin-Fixed Paraffin-Embedded (FFPE) Cutaneous T-Cell Lymphoma Samples:
        Subgroup Analysis Results and Elucidation of Biases from FFPE Sample Processing on
        the TruSeq Platform. *Front. Med.* 2017;4(153):.

54.   Litvinov I V., Tetzlaff MT, Thibault P, et al. Gene expression analysis in Cutaneous T-Cell
        Lymphomas (CTCL) highlights disease heterogeneity and potential diagnostic and
        prognostic indicators. *Oncoimmunology*. 2017;6(5):e1306618.

Figures

Figure 1: Single-cell map of peripheral blood across 6 patients with Sézary Syndrome. **A**.

UMAP projection of the flow-sorted CD45$^+$ immune cells (n=54,994) across 6 patients: T

cells/CTCL cells (Cluster 1, 2, 3, 4, 6, 7, 8, 9, 10, 12, and 15), myeloid cells (Cluster 5, 11, and

14), and B cells (Cluster 13). **B**. Percent of cells expressing selected lineage markers for T cells

(*CD2* and *CD3G*), myeloid cells (*ITGAM* [CD11b] and *ITGAX* [CD11c]), and B cells (*CD19* and

CD79A). **C.** Normalized correlation values across rows ranging from 0 to 1 for predicted immune

cell phenotypes from the Human Primary Cell Atlas [18] based on the SingleR R package [17] for

each cluster. **D**. Percentage of total cells for each patient by cell type. **E**. Cell distribution by

patient sample both as contour graphs and summarized in the accompanying bar chart by

patient and cluster. Clusters representing greater than 15% of total patient cells are labeled.

Figure 2: Differentiating malignant and nonmalignant T cells. A. UMAP projection of CD3$^+$ T

cell clusters. **B.** Percent of cells expressing selected common markers for CTCL across the

UMAP projection. **C**. UMAP projection with proportion of total reads per sample of TCR V(D)J-

based clonotype attached into the following grouping: Hyperexpanded (0.1< X ≤ 1), Large

(0.01< X ≤ 0.1), Medium (0.001 < X ≤ 0.01), Small (1e-4 < X ≤ 0.001) and rare copies (0 < X ≤

1e-4) using the scRepertoire R package [22] with the percentage of cells in each clonotype

category in the lower bar chart. Malignant T cell clusters (C1, C2, C3, C6, C10, C12, and C15)

had greater than 50% of cells with hyperexpanded clonotypes. **D**. Percent difference (Δ percent)

between malignant and nonmalignant T cells versus log2-fold change with the top ten significant

(Bonferroni p-value < 0.05) upregulated genes by both percentages and fold change labeled. **E.**

Violin plot of relative mRNA expression of the top 12 significant genes by Δ percent difference

between malignant (M) and nonmalignant (NM) T cells. **F**. Significantly upregulated genes in the

single-cell (SC) cohort and a secondary SS cohort [29] with the top ten genes by a log-fold change

in overlapping comparison and unique to the SC cohort. **G.** Representative immunohistochemistry images for SS at 400x magnification. Representative images and data summary (bar graph) of 35 patients with MF, SS, or T-cell NHL. Black arrows indicate nuclear staining of AIRE, and images displayed are at 200x magnification.

Figure 3: Heterogeneity in both normal and malignant SS T cells. A. Heatmap of the top ten genes for each of the single-cell clusters with Bonferroni P-value < 0.05. Bars on the left indicate malignant (red) or nonmalignant (blue) predominance. **B**. Density plots of the principal component analysis (PC1 and PC2) of genes involved in T cell differentiation. **C.** Selection of genes by cluster identity for skin-homing, central memory, and regulatory T-cell phenotypes (*IL7R*-low/negative associated with Tregs). **D**. Percent of cells in each cluster in phases of the cell cycle by regression analysis. **E**. ssGSEA enrichment scores for gene signatures derived from the malignant SS clusters applied to a mixed MF/SS bulk RNA cohort [24]. **F**. The proportion of dominant V read versus ssGSEA enrichment scores for gene signatures derived from the malignant SS clusters applied to a mixed MF/SS bulk RNA cohort [24]. **G**. Z-score scaled ssGSEA enrichment scores for Biocarta, Kegg, and Hallmark libraries by single-cell cluster. The median values across nonmalignant (NM) clusters are reported in a single column.

Figure 4: High-resolution mRNA expression change from HDAC inhibitor therapy. A. UMAP projection of CD3$^+$ T cell clusters. **B.** Relative percentage of flow-sorted malignant (CD3$^+$CD4$^+$CD5^{bright}SSC^{hi}) and nonmalignant (CD3$^+$CD4$^+$CD5^{int}SSC^{int}) T cells by cluster and relative percent of clusters by initial sample collection (T1) and 9 months later on HDACi therapy (T2). **C.** UMAP projection with the frequency of TCR V(D)J-based clonotype attached into the following grouping: Hyperexpanded (200 < X ≤ 15100), Large (20 < X ≤ 200), Medium (5 < X ≤ 20), Small (1 < X ≤ 5) and single copies (X = 1) using the scRepertoire R package [22] with the percentage of cells in each clonotype category in the lower bar chart. **D.** Percent of cells

expressing selected common markers for CTCL across the UMAP projection. **E.** Volcano plot of differential genes based on the comparison of T2 versus T1 malignant T cells. Significant genes, as defined by 1 > log-fold change < -1 and Bonferroni p-value < 0.05, highlighted in colors. **F**. Violin plots relative mRNA expression of selected immune, epigenetic, and resistance factor genes for T1 and T2 malignant cells. **G**. Violin plots of ssGSEA enrichment of selected metabolic gene sets for T1 and T2 malignant cells.

Figure 5: Single-cell analysis of pre and post HDACi therapy shows diverging cell populations. A. UMAP projection of malignant T cells at initial sample collection (T1, n=2,603) and post-HDACi progression (T2, n=2,778) with percent of cells in clusters by time point. **B**. Violin plots of cluster-specific gene markers compared to the entire cohort of malignant T cells across both time points. **C**. Density plots of the principal component analysis of ssGSEA results by malignant T cell clusters. **D**. Trajectory manifold of malignant T cells using the monocle R package with both the Seurat clusters and time point displayed. **E**. Median mRNA expression of the top 20 significantly (adjusted p-value < 0.05) upregulated markers by transition state. **F**. Venn diagram of the overlap in significant genes between the transcriptional states in the cell trajectory and comparing HDACi resistant versus sensitive SS patients in GSE132053 [10] with the selection of genes from 33 gene overlap displayed as violin plots. Significant genes were defined as log-fold change ≥ 0.5 and adjusted p-value < 0.05. **G**. Gene expression patterns of significantly expressed genes by transition state that are previously reported upregulated in a HDACi-resistant SS [10] along the cell trajectory manifold.
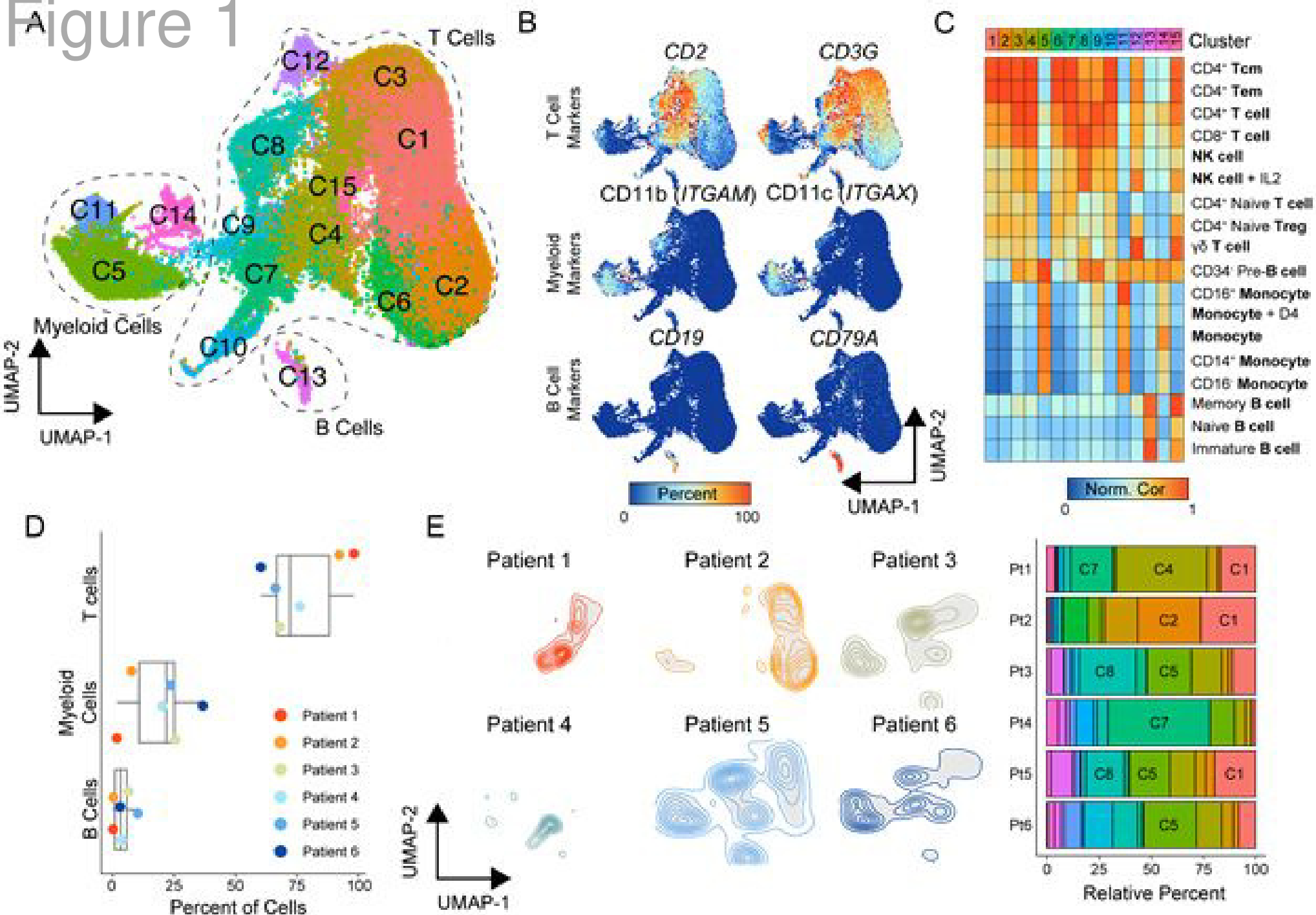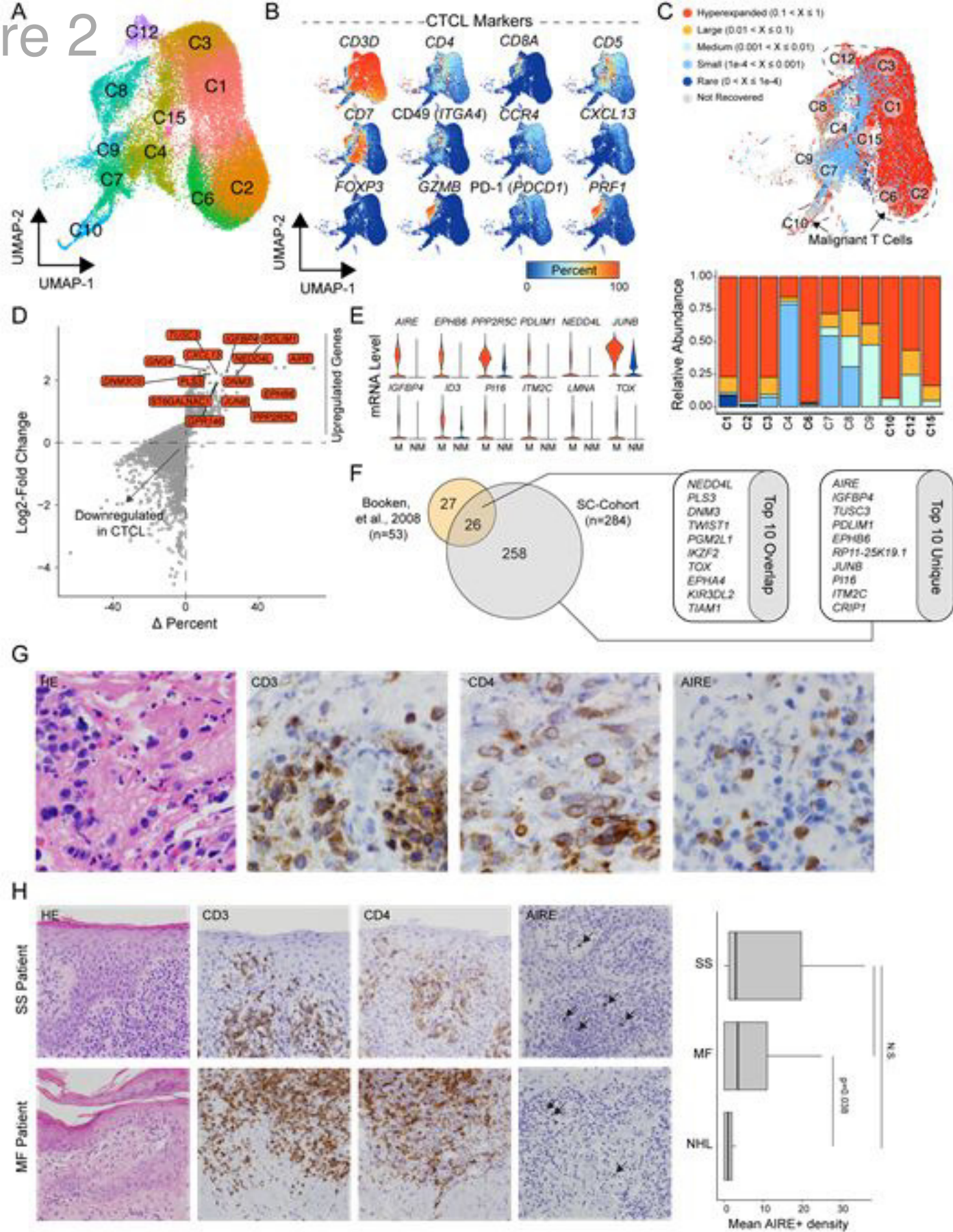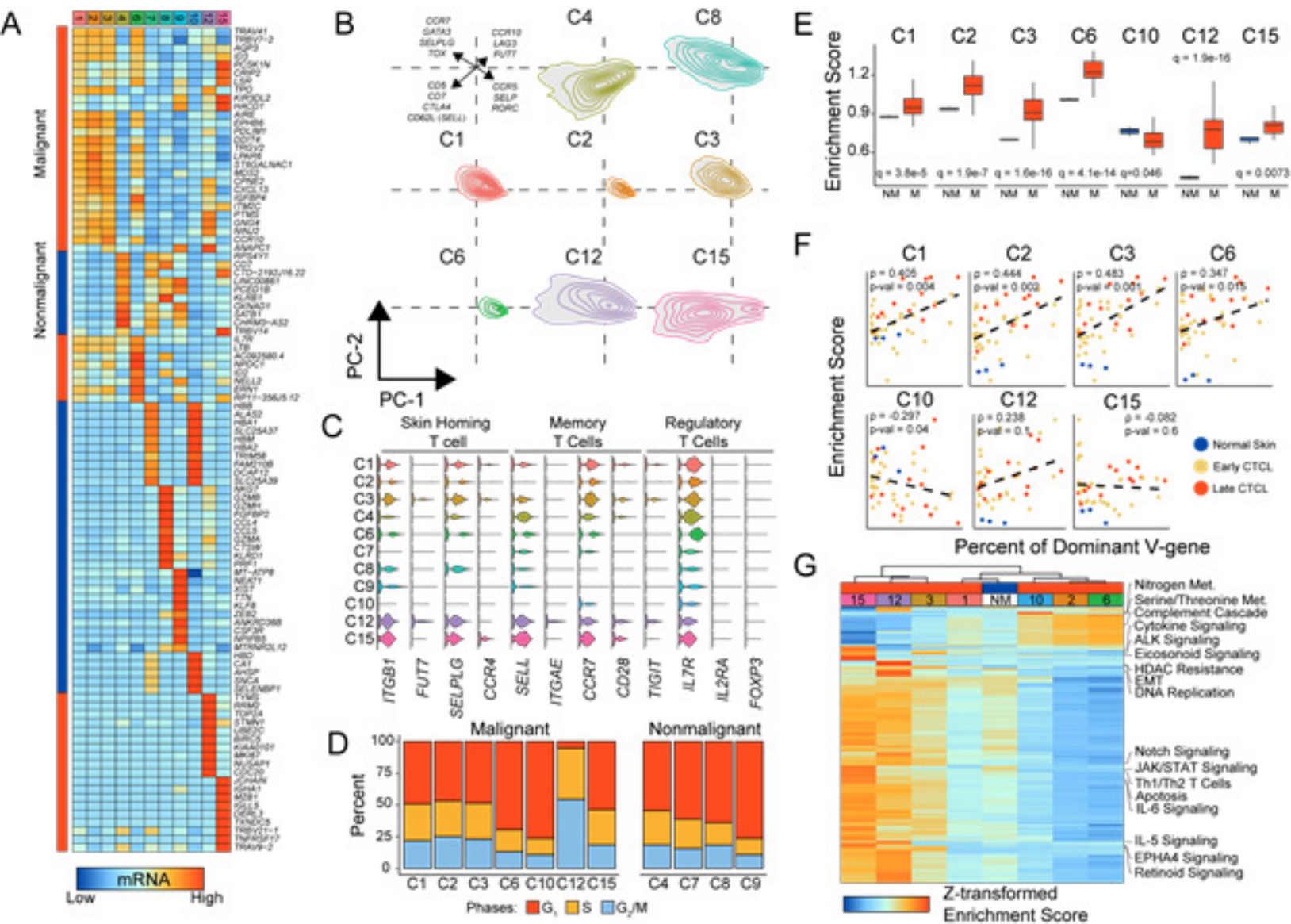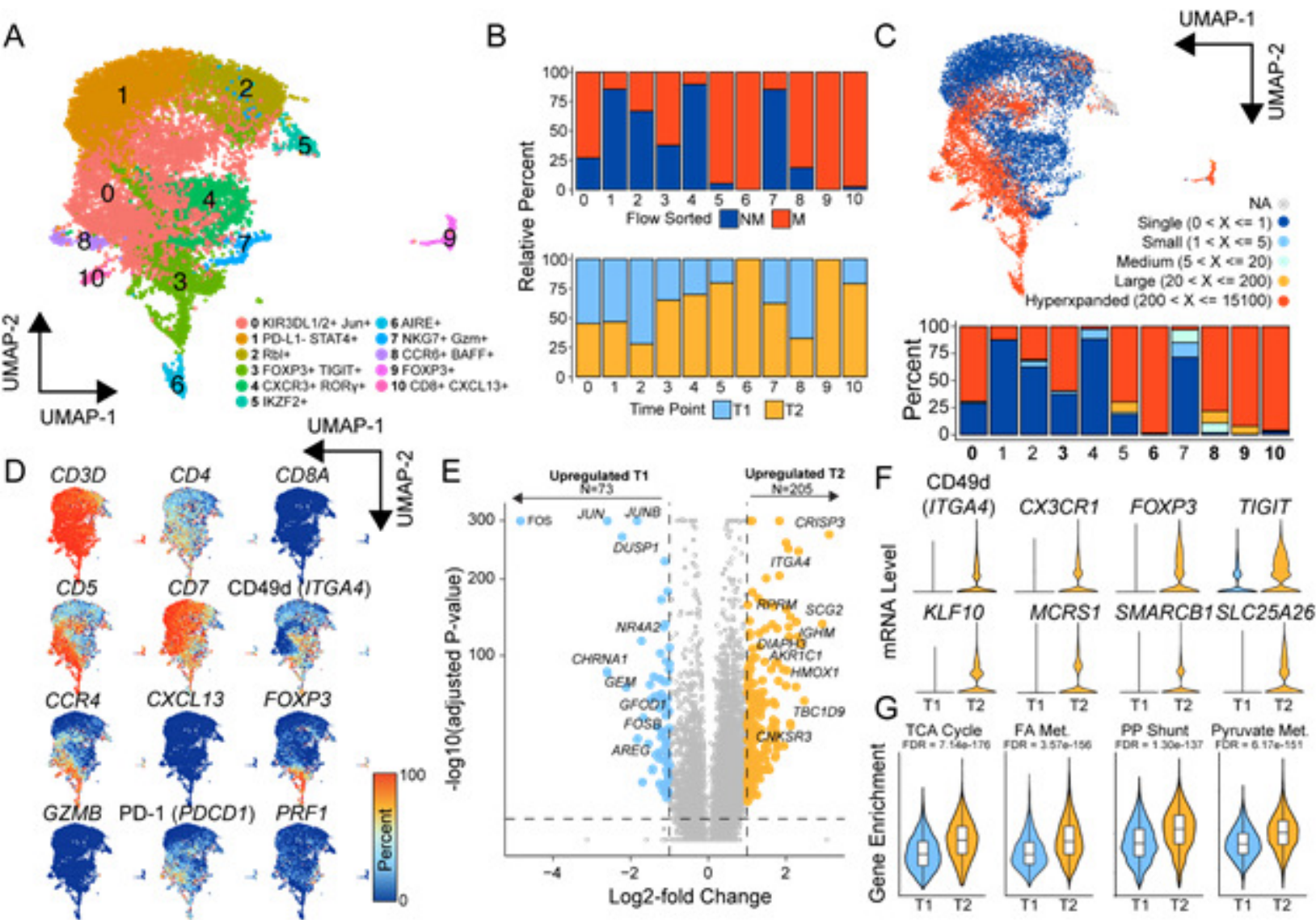
Figure 1

Figure 2

Figure 3

Figure 4

Figure 5