bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470599; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Single-cell Multiomics Reveals Clonal T-cell Expansions and Exhaustion in Blastic Plasmacytoid Dendritic Cell Neoplasm

Erica A. K. DePasquale[1-4], Daniel Ssozi[1,3], Marina Ainciburu[5], Jonathan Good[1,6], Jenny Noel[1,3], Martin Villanueva[3,7], Charles P. Couturier[3,10,12], Alex K. Shalek[3,8-12], Sary F. Aranki[13], Hari R. Mallidi[13], Gabriel K. Griffin[3,15,16], Andrew A. Lane[2,3,4,17], Peter van Galen[1-4*]

[1]Division of Hematology, Brigham and Women's Hospital, Boston, MA, USA

[2]Department of Medicine, Harvard Medical School, Boston, MA, USA

[3]Broad Institute of MIT and Harvard, Cambridge, MA, USA

[4]Ludwig Center at Harvard, Harvard Medical School, Boston, MA, USA

[5]Hemato-Oncology program, Centro de Investigación Médica Aplicada (CIMA), Universidad de Navarra, IDISNA, Pamplona, Spain

[6]Department of Human Biology, Sattler College, Boston, MA, USA

[7]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA

[8]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA

[9]Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, MA, USA

[10]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, USA

[11]Ragon Institute, Harvard University, Massachusetts Institute of Technology, and Massachusetts General Hospital, Cambridge, MA, USA

[12]Division of Health Science & Technology, Harvard Medical School, Cambridge, MA, USA

[13]Division of Thoracic and Cardiac Surgery, Brigham and Women's Hospital, Boston, MA, USA

[14]The International Mesothelioma Program, Brigham and Women's Hospital, Harvard Medical School Boston, MA, USA

[15]Department of Oncologic Pathology, Dana-Farber Cancer Institute, Boston, MA, USA

[16]Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA

[17]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA

*** Correspondence:**
Peter van Galen
pvangalen@bwh.harvard.edu

Keywords: BPDCN, cancer, single-cell, bioinformatics, multiomics

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470599; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

The immune system represents a major barrier to cancer progression, driving the evolution of immunoregulatory interactions between malignant cells and T-cells in the tumor environment. Blastic plasmacytoid dendritic cell neoplasms (BPDCN), a rare acute leukemia with plasmacytoid dendritic cell (pDC) differentiation, provides a unique opportunity to study these interactions. pDCs are key producers of interferon alpha (IFNA) that play an important role in T-cell activation at the interface between the innate and adaptive immune system. To assess how uncontrolled proliferation of malignant BPDCN cells affects the tumor environment, we catalog immune cell heterogeneity in the bone marrow (BM) of five healthy controls and five BPDCN patients by analyzing 52,803 single-cell transcriptomes, including 18,779 T-cells. We test computational techniques for robust cell type classification and find that T-cells in BPDCN patients consistently upregulate interferon alpha (IFNA) response and downregulate tumor necrosis factor alpha (TNFA) pathways. Integrating transcriptional data with T-cell receptor sequencing via shared barcodes reveals significant T-cell exhaustion in BPDCN that is positively correlated with T-cell clonotype expansion. By highlighting new mechanisms of T-cell exhaustion and immune evasion in BPDCN, our results demonstrate the value of single-cell multiomics to understand immune cell interactions in the tumor environment.

1 Introduction

Innovations in immuno-oncology, such as immune checkpoint blockade (ICB) therapy, have transformed cancer medicine. ICB and CAR-T cells have led to improved outcomes in various solid tumors and B-cell malignancies, respectively (1), however these methods have been less successful in myeloid leukemias than in other cancers. Various mechanisms have been proposed to explain the relative inefficacy of immunotherapies in myeloid leukemias, including expression of immunoregulatory molecules by malignant cells (2,3). The complexity of immune regulation in cancer is well illustrated by expression of interferon (IFN) related genes that influence the effectiveness of immunotherapy through immunostimulatory and immunosuppressive effects (4–6). Chronic interferon signaling has been associated with T-cell exhaustion in the setting of viral infections, but its role in cancer is controversial (7–9). Blastic plasmacytoid dendritic cell neoplasm (BPDCN) is an aggressive form of acute leukemia with few effective therapies that provides a unique opportunity to study IFN dysregulation in cancer. BPDCN is characterized by uncontrolled proliferation of transformed plasmacytoid dendritic cells (pDCs), specialized immune cells that link the innate and adaptive immune systems through the secretion of Type I interferons, including IFNA, particularly during viral infection (10). Studies of BPDCN up to this point have largely focused on the malignant pDC-like tumor cells, but few have focused on the T-cell response. An effective immune response relies on the interaction between healthy innate and adaptive immune systems, which is critical for immunotherapy and may be impacted by IFNA-producing pDCs.

Single-cell RNA-sequencing (scRNA-seq) has provided granular insights into the dynamics and phenotypes of immune cells. Specifically, scRNA-seq has been utilized to define subsets of exhausted T-cells in viral infection and cancer, including those that drive responses to ICB (11–13). Advances in other single-cell technologies — such as single-cell methylation, chromatin accessibility, and mutation status — have allowed for multimodal data collection all from the same cell (14–17). This type of analysis allows us to understand relationships between biological processes and heterogeneous cell types that cannot be studied using unimodal measurements alone. To study T-cell biology, scRNA-seq can be paired with sequencing of the T-cell receptor (TCR) α- and β-chain variable regions. These regions can be enriched from single-cell transcriptomes while maintaining cell barcodes to integrate TCR sequencing data with gene expression profiles (18–20). Recently, TCR sequencing has been applied to solid and blood cancers to tie cell phenotypes to TCR

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470599; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

properties, which has allowed for a deeper examination of T-cell subsets via multiple modalities (20–22). These studies also investigate the phenomenon of T-cell exhaustion, which occurs when cells enter a dysfunctional state in which they lose effector functions, such as proliferation and cytotoxicity, and gain immunoregulatory functions (7,23,24). Known drivers of T-cell exhaustion include continuous antigenic stimulation and chronic inflammation (7). Understanding the mechanism by which cells become exhausted is particularly important in the context of immunotherapy, as exhausted T-cells can regain effector functions via ICB therapy (24,25).

Here, we investigate the immune environment in five BPDCN bone marrow samples and five healthy controls using scRNA-seq and TCR sequencing. We used computational integration tools to create a unified healthy reference and applied multiple classification algorithms to annotate cell types in the BPDCN samples by consensus. These data show variable cell type expansions and a wide range of T-cell proportions between patients. Though we found heterogeneity in cell type proportions, we identified common signatures of upregulated IFNA response, downregulated TNFA signaling, and significantly increased CD8+ T-cell exhaustion. We applied a TCR sequencing approach termed T-cell Receptor Enrichment to linK clonotypes (TREK-seq) that we recently developed to identify expanded T-cell clonotypes in BPDCN samples and showed a correlation between CD8+ T-cell clone size and exhaustion scores. By providing a comprehensive map of cellular heterogeneity and T-cell transcriptomes/clonotypes in BPDCN patients, we lay the foundation for future development and evaluation of immunotherapies in this devastating cancer.

2 Materials and Methods

2.1 Patient samples

Bone marrow aspirate was collected from 5 patients with blastic plasmacytoid dendritic cell neoplasm (BPDCN) who consented to an excess sample banking and sequencing protocol that covered all study procedures and was approved by the Institutional Review Board (IRB) of the Dana-Farber/Harvard Cancer Consortium. Additionally, 5 samples from healthy donors were collected for use as a control in this study, following the same approved protocol or a protocol approved by the IRB of Mass General Brigham. Details for each of these samples are located in **Supplemental Table 1**.

2.2 Single-cell RNA-sequencing

Single-cell RNA-sequencing was performed on cryopreserved iliac crest bone marrow aspirates for BPDCN samples and BM 1-3 controls, and cryopreserved sternum bone marrow for BM 4 and 5 controls. To isolate mononuclear cells, BPDCN samples and BM 1-4 were processed using Ficoll or lymphoprep, whereas BM 5 was processed using Acrodiscs (Pall AP-4952). Cells were stored in liquid nitrogen, thawed using standard procedures, and viable (DAPI negative) cells were sorted on a Sony SH800 flow cytometer. Next, 10,000-15,000 cells were loaded onto a 10x Genomics chip. Further processing was done using the recommended procedures for the 10x Genomics 3' v3.0 or v3.1 chemistry (26). 10x libraries were sequenced on the NovaSeq SP 100 cycle with the following parameters (Read 1: 28 + Read 2: 75 + Index 1 (i7): 10 + Index 2 (i7): 10).

2.3 Dataset processing

Raw sequencing data were processed using CellRanger software (27) to generate FASTQ files and count matrices. Ambient RNAs were estimated and removed from the datasets using SoupX with default parameters (28). Each dataset was filtered to retain cells with >= 500 UMIs, >=1000 genes expressed, and <30% of the reads mapping to the mitochondrial genome. Three of the healthy control

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470599; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bone marrow samples were originally integrated using the IntegrateData() function in Seurat v4.0.3 and clustered using the same software at the default resolution of 0.5 (29). Cell cycle genes ("cc.genes" within the Seurat software) were removed from the integration anchors to combine smaller cell cycle-driven clusters and to eliminate redundancy, though these genes were retained in the final dataset for further analyses. Further, a combination of high resolution sub-clustering with Seurat and analysis of key T-cell gene expression were used to improve granularity in the naive T-cell compartment. Two additional healthy samples from older individuals were further integrated into the original controls using Seurat's TransferData() function. The BPDCN samples were processed and clustered individually using Louvain clustering within Seurat with default options and a resolution of 0.5, though these clusterings were not used in downstream analyses.

2.4 Cell classification

Cells in each BPDCN sample were classified with four classification methods using the integrated healthy control samples as a reference: random forest (30,31), cellHarmony (32), Seurat's TransferData() function (33), and scPred with default parameters (34). Reference input for three of the algorithms was the integrated Seurat object and feature selection was performed separately as a part of each method. cellHarmony instead used a gene expression matrix of the cells by top 50 marker genes for each cluster (as defined by Seurat), the full expression matrix, and a table to cell type classifications for the reference. Given highly consistent classifications between the four methods, we selected cellHarmony-defined cell type labels for all downstream analyses. Classified cells from BPDCN patients were projected into the UMAP space of the integrated controls using the MapQuery() function in Seurat v4.0.3.

2.5 Gene Set Enrichment Analysis (GSEA)

Gene set enrichment was performed separately on cells from each cell type and BPDCN sample relative to the healthy controls. Log fold change values for every gene were extracted from the cellHarmony output (~/input/cellHarmony/OtherFiles), sorted in decreasing order, and used as input for a custom GSEA function (35). This function uses two GSEA implementations in R, gage (36) and fgsea (37) and only reports the results that are significantly enriched with both methods. Hallmark gene sets from the GSEA website were used for this analysis (https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp). Pathways that were significantly enriched in at least one comparison were plotted using the Pheatmap R package.

2.6 Pathway score quantification

To score individual cells for gene signatures, we combined all 52,803 cells from the dataset into one Seurat object, read in a curated list of signatures (IFNA response from the GSEA Hallmark gene set, TNFA signaling by NFKB from the Hallmark gene set, T-cell exhaustion from Penter et al. (21)) and applied the function AddModuleScore to the Seurat object using signature genes as features. Statistical significance for each cell type was calculated by comparing the median scores of normal samples (n = 5) to the median scores of BPDCN samples (n = 5) using the R function wilcox.test and default parameters. The number of biological replicates was less than five if a cell type was not detected in one of the donors (for example, n = 4 for BPDCN CD8+ Memory T-cells).

2.7 T-cell receptor sequencing

T-cell receptor sequencing was performed on both control and BPDCN bone marrow samples using a modified protocol originally developed for Seq-Well (18). This protocol adaptation termed TREK-

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470599; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

seq is described in an independent manuscript (Miller et al. Nat Biotechnol., under review). The modifications to the original protocol are as follows: in the TCR enrichment master mix, we added PartialRead1 and PartialTSO primers at a final concentration of 1.25 µM each. For amplification of TCR transcripts following enrichment, we used the same primers at a final concentration of 0.4 µM each. For the final PCR, we used UPS2-N70x and 10X_SI-PCR_P5 primers at a final concentration of 0.2 µM each to add the Illumina P5 and P7 sequences. The libraries were sequenced using a 150 cycle kit on the Illumina MiSeq and loaded at a final concentration of 10 pM. 28 cycles were used for Read 1, which read the cell barcode and UMI. 150 cycles were used for Index 1, which read the TCR region. TCRα and TCRβ-specific custom sequencing primers were used for index 1 at a final concentration of 2.5 µM. We aimed for a MiSeq cluster density of roughly 450k/mm2. Primer sequences can be found in **Supplemental Table 2**.

2.8 TREK-Seq computational pipeline

Raw sequencing data were demultiplexed with bcl2fastq (v2.20.0), and the resulting fastq files were reformatted to join the cell barcode, UMI and corresponding TCR sequence in the same file. For this analysis, we developed and applied the bioinformatics pipeline WARPT, or Workflow for Association of Receptor Pairs from TREK-seq (https://github.com/mainciburu/WARPT). Briefly, we first corrected cellular barcodes allowing one mismatch with the 10x Single Cell 3' v3 whitelist. UMI correction was also performed by clustering together UMIs with one mismatch. Every corrected barcode and UMI sequence was then added to the corresponding fastq read header. Next, we applied a quality filter to remove every read with an average score < 25. To account for barcode swapping, TCR sequences with identical barcode and UMI were subjected to clustering using an identity threshold of 0.9. The subsequent analysis was carried out exclusively with clusters representing at least 50% of the reads with identical barcodes and UMI and doubling the number of reads from the second most abundant cluster. Next, a consensus sequence was built to summarize the TCR sequences in each of those clusters. We required a minimum of 3 sequences and allowed for a maximum error rate of 0.5 and gap frequency of 0.5 per position. Consensus sequences were aligned to the VDJ segments reference available at the IMGT database. IgBLAST with the default parameters was used for the alignment. Finally, the V, D and J calls and CDR3 sequence with higher UMI counts, for both *TRA* and *TRB* genes, were assigned to each cell barcode. For downstream analysis, only *TRB* variable regions were used.

2.9 Statistics

P-values less than or equal to 0.05 were considered statistically significant. Cell population proportions between younger and older healthy donors were assessed for significant differences using a chi-squared test. For the gene set heatmaps, P-values were calculated via t-test for each comparison of reference to BPDCN sample per cell type and adjusted using the Bonferroni method. Genes that were significant following correction and had an expression value of 0.5 or greater (log normalized expression) in at least 1 sample were retained for the heatmap. A Wilcoxon signed-rank test was used for pathway score quantification, a Kruskal-Wallis rank sum test was used to compare normalized T-cell clonotype sizes, and a Spearman ranked correlation analysis was used to test the relationship between clonotype size and exhaustion scores.

2.10 Data and software availability

All data used in this paper will be made publicly available in the Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/). The scripts used to generate the results in this paper can be found at the project GitHub repository: https://github.com/EDePasquale/BPDCN.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470599; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3 Results

3.1 Identification of Cell Populations in Healthy Bone Marrow

To map the cellular diversity in healthy bone marrow samples, we performed scRNA-seq using the 10x Genomics platform (27). We profiled cells from three younger healthy donors (age 31-45) and two older donors (age 74 and 75) (**Supplemental Table 1**). The two older donors were included as age-appropriate controls for BPDCN, which has been reported to have a median age of ~65 (38). Unique cellular barcodes were used to assign transcripts to cells and individual mRNA molecules were quantified using Unique Molecular Identifiers (UMI). Across all five healthy donors, we retained 25,726 cells following quality control and filtering.

Cells from these healthy donors (BM 1-5) were integrated using the Seurat package (IntegrateData and TransferData functions) to remove batch effects and generate a unified UMAP projection of the data (**Figure 1A and Supplemental Figure 1**). We initially identified 16 unique clusters based on gene expression that were interrogated for expression of canonical marker genes and gene signatures of blood cell types (**Figure 1B**). Additional sub clustering was performed in the lymphocytic compartment, and new T-cell classifications were made based on expression of *CD8A, CD4, ITGB1,* and *CCR7* gene expression. Specifically, CD4+ T-cells and CD8+ T-cells could both be split into Naive (*ITGB1- CCR7+*) and Memory (*ITGB1+ CCR7-/+*) subsets. The final reference was comprised of 17 clusters representing known hematopoietic cell types, including HSCs and progenitors, myeloid, erythroid, and lymphoid cells (**Figure 1C**). All cell types were represented in each donor at broadly comparable levels, though variation exists as expected (**Figure 1D**) (39). We observed no statistically significant differences in population proportions between the donors as an effect of age or collection procedure ($P = 0.2351$).

3.2 Single-Cell Profiling of BPDCN Bone Marrow

Approximately half of BPDCN patients present with skin tumors only, whereas the other half present with bone marrow involvement, i.e. detection of bone marrow tumor blasts by conventional clinical assays such as morphologic examination or flow cytometry (40). To understand the proportional and transcriptional changes that occur in the bone marrow in the presence of BPDCN tumor cells, we selected five patients with 1.5-97% disease involvement at diagnosis (**Supplemental Table 1**). We performed scRNA-seq on BPDCN bone marrow samples using the same protocol as in the healthy samples. From this sequencing, we acquired 27,077 cells across all BPDCN donors following quality control and filtering.

We next wanted to classify cell types in the five BPDCN samples (BPDCN 1-5) using the healthy controls cells as a reference. To this end, we tested four computational classification methods that utilize different underlying algorithms to assign cell type labels to cells: random forest, cellHarmony, Seurat TransferData, and scPred. The resultant cell type predictions were concordant across all four methods in most cell types, particularly in the erythroid and lymphoid lineages, lending confidence to the classification methods (**Figure 2A**). Some variability was noted in the proportion of cells labeled as pDC, cDC, and ProB in BPDCN 4 and 5; we hypothesized that these could be malignant cells that align poorly to any healthy population, which would be consistent with the clinical annotation of high malignant cell content in these bone marrows (**Supplemental Table 1**). To visualize shifts in gene expression programs and differential cell type proportions, we projected the cells from each BPDCN patient onto the same UMAP space as the healthy reference, colored by cellHarmony labels (**Figure 2B**). Indeed, we observed that the pDC and cDC labeled cells in BPDCN 4 and 5 were shifted toward the HSC/Prog and ProB populations, suggesting that the malignant tumor cells in these samples were transcriptionally different from their healthy

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470599; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

counterparts. For this study, we were primarily interested in the role of T-cells in response to and as a potential treatment for BPDCN, therefore we next focused on alterations in cells of the lymphoid lineage.

3.3 T-Cell Populations Show Differential Regulation of IFNA and TNFA Pathways

We used Gene Set Enrichment Analysis (GSEA) to identify pathways that were significantly enriched in the BPDCN samples. Using the Hallmark gene sets and gene expression changes derived via cellHarmony, we generated a heatmap containing all up- or down-regulated pathways enriched in any BPDCN sample compared to the healthy controls (**Figure 3A**). In several T and NK cell populations from BPDCN patients, we found upregulation of “interferon alpha response”, "allograft rejection", "MYC targets", and "oxidative phosphorylation". Interferon alpha (IFNA) response genes were significantly upregulated in 4/4 BPDCN samples in CD8+ Memory T-cells, 3/5 in CD4+ Memory T-cells, 3/3 in CD8+ Naive T-cells, and 3/3 in CD4+ Naive T-cells, along with 3/5 samples in NK cells. IFNA is a marker of immune activation and pDCs produce IFNA, furthering our interest in this pathway. Using a different statistical framework, we next scored all individual CD8+ Memory T-cells for their expression of genes in the IFNA response gene set. We found a significant increase in BPDCN 1-4 relative to control samples, confirming increased IFNA response gene expression in BPDCN T-cells (P = 0.0159, **Figure 3B**). BPDCN 5 did not have cells classified as CD8+ Memory T-cells and was not tested. Increased IFNA response gene scores were also observed in CD4+ Memory T-cells and NK cells (data not shown). We found that tumor necrosis factor alpha (TNFA) signaling was downregulated in many of the T-cell and NK populations: 4/4 CD8+ Memory T-cell, 3/5 CD4+ Memory T-cell, and 2/3 CD4+ Naive T-cell, and 3/5 of the NK cell samples with the exception of CD8+ Naive T-cells (0/3) (**Figure 3A**). Scoring individual CD8+ Memory T-cells for the TNFA gene signature also showed significant reductions in BPDCN (P = 0.0159, **Figure 3C**). These results demonstrate that T/NK-cells in BPDCN exhibit consistent gene expression changes compared to healthy control cells, including increased response to the pDC-related cytokine IFNA and decreased TNFA signaling.

We next investigated which genes in the IFNA response gene set were driving the enrichment in BPDCN T-cells. We generated a heatmap of genes that were significantly different from control expression levels in at least one sample/cell type pair and further filtered to those genes with normalized expression values of at least 0.5 in one pair (p-value ≤ 0.05, Bonferroni adjusted, **Figure 4A**). We also generated heatmaps with the full list of genes in each gene set (**Supplemental Figure 2**). Related to the IFNA response, genes that were upregulated in BPDCN compared to control T-cells in the majority of samples included: *IRF1, PSMB9, PSME2, CD74, ISG15, ISG20, CD47, LY6E, PSMB8,* and *SELL* (**Figure 4A**). The increased expression of *IRF1, PSMB9,* and *PSME2* were notable in all 5 BPDCN samples across lymphoid cell types, with the exception of BPDCN 4’s CD4+ Naive T-cells due to low cell numbers. In contrast, *CD74, ISG15, ISG20, CD47, LY6E, PSMB8,* and *SELL* were most notably increased from control levels in CD8+ Memory T-cells and NK cells, and in BPDCN 1-3 specifically. These genes are involved in T-cell proliferation/differentiation, antigen presentation, and inflammation, consistent with the view that IFNA signaling plays a critical role in immune activation.

Related to TNFA signaling, genes that were downregulated in BPDCN T and NK cells compared to control samples are as follows: *NFKBIA*, *TNFAIP3*, *DUSP1*, *DUSP2*, *FOS*, *JUN*, *ZFP36*, *BTG2*, and *CD69* (**Figure 4B**). These genes were downregulated relative to controls in all T-cell and NK cell subsets in BPDCN 1-4, with the largest expression differences being isolated to the CD4+ and CD8+ Memory T-cell and NK cell clusters. These genes are involved in immune regulation via inflammatory response, proliferation and T-cell activation. Decreased expression of TNFA genes is indicative of a lowered immune response, particularly in relation to T-cell activation.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470599; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Overall, our results suggest an altered balance between inflammatory pathways in T-cells of BPDCN patients, shifting towards IFNA at the expense of TNFA signaling (41). The changes we observe in IFNA and TNFA target genes are likely to affect the function of the adaptive immune system in BPDCN.

3.4 Expanded T-Cell Receptor Clonotypes in BPDCN Correlate with Signatures of Exhaustion

T-cell antigen specificity is determined by the TCR sequence, which consists of the α- and β-chains encoded by *TRAV* and *TRBV* genes that are recombined during T-cell development. We reasoned that sequencing the variable region of the TCR may help to elucidate the biology of T-cells in BPDCN. We performed TCR sequencing on healthy control and BPDCN samples using TREK-seq, a protocol for paired transcript and TCR capture in 10x Genomics data (see **Methods**). We developed and applied the Workflow for Association of Receptor Pairs from TREK-seq (WARPT) computational pipeline to detect groups of cells expressing the same TCR sequence, i.e. T-cell clonotypes. We identified 10,870 T-cell clonotypes in our dataset, each supported by multiple sequencing reads. Next, we ranked clonotypes in each sample by their normalized size (number of cells with the same TCR over all cells in which a TCR was detected). For BM 1-5, we found that none of the control samples were dominated by a single clone or handful of clones, but that the distribution of clonotypes was largely uniform (**Figure 5A-E**). BPDCN samples showed heterogeneity in the distribution of clonotypes: while BPDCN 2-4 do not exhibit a single prevalent clone, similar to the healthy controls, we detected a clone that comprises 30.9% of all TCRs in BPDCN 1 (**Figure 5F-I**). In the healthy controls, TCR sequences were nearly exclusively detected in T-cells (94.7%, **Figure 6A**), and we mainly observed clonal expansions in the CD8+ Memory T-cell compartment, which was expected (**Figure 6B**). These results support that the TREK-seq protocol and analysis pipeline we developed is highly accurate with minimal spurious TCR detection. However, some BPDCN samples showed unexpected cell types containing T-cell receptors (**Figure 6A**). In BPDCN1, one-third of the cells within the largest clone are labeled as NK cells (**Figure 5F**). This sample has a large number of cells classified as NK cells, but a limited tumor cell involvement of 1.5% (**Supplemental Table 1**), leaving open the possibility that the NK-labeled population with TCR results are NKT cells. In BPDCN 4, we unexpectedly captured a minor fraction of rearranged TCR sequences in cDCs (**Figure 5I**). While a rearranged TCR has been detected in tumor cells of some CD4+CD56+ hematologic malignancies (42), this has not been observed in BPDCN, illustrating the potential of single-cell multiomics to reveal unexpected features in the tumor environment.

We next evaluated T-cell exhaustion by scoring each T-cell for the expression of exhaustion-associated genes (21). We overlaid cell type label, exhaustion scores, and normalized clone size parameters onto UMAP plots for each BPDCN sample containing T-cells (**Figure 7**). CD8+ Memory T-cell clusters were enriched for exhausted cells and expanded clonotypes in all patients, which was confirmed by quantification of signature scores and clone sizes in cell types (**Figure 8A**). Overall, exhaustion scores in CD8+ Memory T-cells in BPDCN 1-4 were significantly higher than controls ($P = 0.0159$), though the range of scores among individual cells within each sample was high (**Figure 8B**). These results are consistent with the pathway analysis above, as IFNA-mediated immune stimulation and low levels of TNFA signaling were previously associated with T-cell exhaustion (7,8,43).

To explore the relationship between exhaustion and clonotype sizes, we correlated these two parameters at the single-cell level. We found that CD8+ Memory T-cell clonotypes of various sizes can express T-cell exhaustion signature genes, including some clonotypes that make up a small proportion of the total T-cell population (**Figure 8C-F**). Spearman ranked correlation analysis revealed a significant positive relationship between clonotype size and exhaustion signature

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470599; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

expression: rho of 0.67, 0.61, 0.63, and 0.4 for BPDCN 1-4, respectively (**Supplemental Figure 3**). These results suggest that clonal expansion and exhaustion of CD8+ Memory T-cell clonotypes may be functionally linked in the BPDCN tumor environment.

4 Discussion

To study the immune environment in BPDCN, we utilized paired single-cell sequencing of the transcriptome and T-cell receptors in healthy controls and patient samples. We identified 17 transcriptionally unique cell clusters in an integrated dataset of bone marrow from five younger and older healthy donors. We used this dataset to label cell types in five BPDCN patient bone marrow samples using four distinct classification algorithms, revealing heterogeneity in cell type proportions that correspond to clinical assessment of bone marrow involvement. An analysis of differentially regulated gene sets revealed that IFNA response genes were significantly increased in T-cells of the BPDCN samples and TNFA signaling genes were decreased. A deeper examination of expression within each gene set identified a subset of genes driving this enrichment, with the associated functions of T-cell proliferation, activation, and antigen presentation. TREK-seq showed T-cell specific TCR expression and uniformly distributed clone sizes in healthy controls; both of these results were more variable in BPDCN patient samples, suggesting possible TCR expression by tumor cells and antitumor reactivity by expanded T-cells in BPDCN. Further, we saw increased T-cell exhaustion in CD8+ Memory T-cells that positively correlated with increasing clone sizes. These results demonstrate the utility of single-cell multiomics by establishing a resource of the BPDCN tumor environment that provides new insights relevant to immuno-oncology.

Our results contribute to the body of research surrounding the role of the adaptive immune response in cancer and the relationship between T-cell receptor clonality and exhaustion. In a healthy immune system, activated pDCs produce large quantities of IFNA to stimulate activation of T-cells (10). In this study, we observe increased expression of IFNA response genes in T-cells relative to T-cells in the healthy controls, initially suggesting higher levels of IFNA production by pDC-like tumor cells. However, previous research on BPDCN has shown that pDC-like tumor cells produce lower levels of IFNA relative to pDCs from healthy individuals (44,45). Further, little research has been reported describing the role of TNFA signaling in BPDCN, though studies have shown increased TNFA signaling in the plasma of patients with acute myeloid leukemia (46,47). While the directionality of differential expression in these gene sets in T-cells is in contrast to what has been observed in the adjacent BPDCN cells, it is consistent with our findings of increased T-cell exhaustion. We hypothesize that, despite the tumor cells exhibiting lower IFNA production at the individual cell level, abnormal accumulation of pDC-like tumor cells in BPDCN may lead to increased IFNA production and chronic T-cell activation, eventually leading to T-cell exhaustion and consequent TNFA downregulation.

Exhaustion of cytotoxic T-cells has been a major hurdle in establishing effective immunotherapy treatments for blood cancers (24,25,48). Reactivation of T-cell effector function in exhausted T-cells is the goal of ICB, but aberrant IFNA and IFNG signaling have been shown to inhibit this process (49,50). While the prevalence of exhaustion has been studied in other cancers including acute myeloid leukemia, its role in BPDCN remains unknown. In this study, we establish that BPDCN patients have heightened T-cell exhaustion relative to controls. We also find a positive association between T-cell exhaustion and T-cell clonal expansion, consistent with other cancers (21,51). A lack of TCR expression by malignant cells is one feature that distinguishes BPDCN from mature T-cell malignancies, such as cutaneous T-cell lymphoma (52,53). However, little research has been done on T-cell biology in BPDCN, and none at the single-cell level. Our findings highlight a potential mechanism of T-cell exhaustion via persistent IFNA signaling that might be targeted to restore anti-tumor immunity.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470599; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Multimodal single-cell technologies enable investigation of gene regulation at a high resolution through multiple modalities representing interacting processes within a cell. This abundance of information allows us to deeply examine a small number of samples, however application of these methods to a larger patient population is prohibitively expensive and labor-intensive. For this study, we selected five BPDCN patients with varying levels of clinical bone marrow involvement, though more samples will be needed to ensure that the full spectrum of patient heterogeneity is captured. We observe interesting interactions between T-cell exhaustion and TCR expansion using parallel measurement of gene expression and TCR sequences, but other multimodal technologies could elucidate additional mechanisms of regulation in BPDCN. For example, the application of single-cell genotyping methods in future work would help to classify and characterize the transcriptomes of malignant tumor cells (2,16). Inclusion of additional patient samples and complementary single-cell measurements would strengthen our initial findings and uncover new results that further illuminate T-cell biology in BPDCN.

In summary, we apply scRNA-seq and TCR sequencing with computational techniques to catalog the cellular heterogeneity in BPDCN patient samples. Multiomics technology allows us to gain a deeper understanding of immune cell dynamics by assessing the diversity of immune cell states via scRNA-seq and the expansion of T-cell clonotypes through TREK-seq. Our results suggest that the balance between IFNA and TNFA signaling is disrupted in BPDCN, potentially leading to T-cell clonotype expansion and exhaustion. The discovery of mechanisms by which BPDCN cells evade immune destruction will lead to the development of new cancer therapies that leverage tumor-reactive T-cells.

5 Conflict of Interest

A.K.S. reports compensation for consulting and/or SAB membership from Merck, Honeycomb Biotechnologies, Cellarity, Repertoire Immune Medicines, Ochre Bio, Third Rock Ventures, Hovione, Relation Therapeutics, FL82, and Dahlia Biosciences.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

6 Author Contributions

E.A.K.D., D.S., M.A., J.G., J.N., M.V., C.P.C., A.K.S. and P.v.G. conducted experiments and analyzed the data. E.A.K.D., A.A.L., and P.v.G designed the study and interpreted the data. G.K.G., A.A.L., H.R.M., and S.F.A. provided patient specimens and clinical perspectives. E.A.K.D. and P.v.G. wrote the manuscript. All authors edited the manuscript.

7 Funding

P.v.G. and A.A.L. are supported by the Ludwig Center at Harvard and the Bertarelli Rare Cancers Fund. A.A.L. is supported by the NCI (CA225191), the Mark Foundation for Cancer Research, and is a Leukemia and Lymphoma Society Scholar. P.v.G. is supported by National Institutes of Health (NIH) R00 Award (CA218832), Gilead Sciences, the Harvard Medical School Epigenetics & Gene Dynamics Initiative and is a Glenn Foundation for Medical Research and AFAR Grant for Junior Faculty awardee. G.K.G. is supported by a Physician-Scientist award from the Damon-Runyon Cancer Research Foundation.

8 Acknowledgments

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470599; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

We thank patients for donating cells. We thank Phillip Dexheimer for helpful discussions and Patricia Rogers and the Broad Institute Flow Facility for technical support.

9 References

1. Postow MA, Callahan MK, Wolchok JD. Immune checkpoint blockade in cancer therapy. J Clin Oncol. 2015 Jun 10;33(17):1974–82.

2. van Galen P, Hovestadt V, Wadsworth MH Ii, Hughes TK, Griffin GK, Battaglia S, et al. Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. Cell. 2019 Mar 7;176(6):1265–81.e24.

3. Lichtenegger FS, Krupka C, Haubner S, Köhnke T, Subklewe M. Recent developments in immunotherapy of acute myeloid leukemia. J Hematol Oncol. 2017 Jul 25;10(1):142.

4. Sceneay J, Goreczny GJ, Wilson K, Morrow S, DeCristo MJ, Ubellacker JM, et al. Interferon Signaling Is Diminished with Age and Is Associated with Immune Checkpoint Blockade Efficacy in Triple-Negative Breast Cancer. Cancer Discov. 2019 Sep;9(9):1208–27.

5. Ayers M, Lunceford J, Nebozhyn M, Murphy E, Loboda A, Kaufman DR, et al. IFN-γ-related mRNA profile predicts clinical response to PD-1 blockade. J Clin Invest. 2017 Aug 1;127(8):2930–40.

6. Vadakekolathu J, Minden MD, Hood T, Church SE, Reeder S, Altmann H, et al. Immune landscapes predict chemotherapy resistance and immunotherapy response in acute myeloid leukemia [Internet]. Available from: http://dx.doi.org/10.1101/702001

7. Wherry EJ, Kurachi M. Molecular and cellular insights into T cell exhaustion. Nat Rev Immunol. 2015 Jul 24;15(8):486–99.

8. Dagenais-Lussier X, Loucif H, Murira A, Laulhé X, Stäger S, Lamarre A, et al. Sustained IFN-I Expression during Established Persistent Viral Infection: A "Bad Seed" for Protective Immunity. Viruses [Internet]. 2017 Dec 30;10(1). Available from: http://dx.doi.org/10.3390/v10010012

9. Virgin HW, Wherry EJ, Ahmed R. Redefining chronic viral infection. Cell. 2009 Jul 10;138(1):30–50.

10. Yu H, Zhang P, Yin X, Yin Z, Shi Q, Cui Y, et al. Human BDCA2+CD123+CD56+ dendritic cells (DCs) related to blastic plasmacytoid dendritic cell neoplasm represent a unique myeloid DC subset. Protein Cell. 2015 Apr;6(4):297–306.

11. Miller BC, Sen DR, Al Abosy R, Bi K, Virkud YV, LaFleur MW, et al. Subsets of exhausted CD8+ T cells differentially mediate tumor control and respond to checkpoint blockade. Nat Immunol. 2019 Feb 18;20(3):326–36.

12. Jadhav RR, Im SJ, Hu B, Hashimoto M, Li P, Lin J-X, et al. Epigenetic signature of PD-1+ TCF1+ CD8 T cells that act as resource cells during chronic viral infection and respond to PD-1 blockade. Proc Natl Acad Sci U S A. 2019 Jul 9;116(28):14113–8.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470599; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

13. Efremova M, Vento-Tormo R, Park J-E, Teichmann SA, James KR. Immunology in the Era of Single-Cell Technologies. Annu Rev Immunol. 2020 Apr 26;38:727–57.

14. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. Nature. 2015 Jul 23;523(7561):486–90.

15. Guo H, Zhu P, Guo F, Li X, Wu X, Fan X, et al. Profiling DNA methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite sequencing. Nat Protoc. 2015 May;10(5):645–59.

16. Nam AS, Kim K-T, Chaligne R, Izzo F, Ang C, Taylor J, et al. Somatic mutations and cell identity linked by Genotyping of Transcriptomes. Nature. 2019 Jul;571(7765):355–60.

17. Chaligne R, Gaiti F, Silverbush D, Schiffman JS, Weisman HR, Kluegel L, et al. Epigenetic encoding, heritability and plasticity of glioma transcriptional cell states. Nat Genet. 2021 Oct;53(10):1469–79.

18. Tu AA, Gierahn TM, Monian B, Morgan DM, Mehta NK, Ruiter B, et al. TCR sequencing paired with massively parallel 3′ RNA-seq reveals clonotypic T cell signatures. Nat Immunol. 2019 Nov 19;20(12):1692–9.

19. Singh M, Al-Eryani G, Carswell S, Ferguson JM, Blackburn J, Barton K, et al. High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. Nat Commun. 2019 Jul 16;10(1):3120.

20. Oliveira G, Stromhaug K, Klaeger S, Kula T, Frederick DT, Le PM, et al. Phenotype, specificity and avidity of antitumour CD8+ T cells in melanoma. Nature. 2021 Aug;596(7870):119–25.

21. Penter L, Gohil SH, Huang T, Thrash EM, Schmidt D, Li S, et al. Coevolving JAK2V617F+ relapsed AML and donor T cells with PD-1 blockade after stem cell transplantation: an index case. Blood Adv [Internet]. 2021 Aug 25; Available from: http://dx.doi.org/10.1182/bloodadvances.2021004335

22. Wang X, Chen Y, Li Z, Huang B, Xu L, Lai J, et al. Single-cell RNA-seq of T cells in B-ALL patients reveals an exhausted subset with remarkable heterogeneity. Adv Sci. 2021 Oct;8(19):e2101447.

23. Blank CU, Haining WN, Held W, Hogan PG, Kallies A, Lugli E, et al. Defining “T cell exhaustion.” Nat Rev Immunol. 2019 Nov;19(11):665–74.

24. Zhang Z, Liu S, Zhang B, Qiao L, Zhang Y, Zhang Y. T Cell Dysfunction and Exhaustion in Cancer. Front Cell Dev Biol. 2020 Feb 11;8:17.

25. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat Med. 2018 Oct;24(10):1550–8.

26. Genomics 10x. Chromium Next GEM Single Cell 3′ Reagent Kits v3.1 (Dual Index) [Internet]. 10xGenomics.com. 2020 [cited 2021]. Available from:

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470599; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

https://assets.ctfassets.net/an68im79xiti/D4H8F6Nt7tv3U30Bu9vAY/357f84ca07ab959009ead2b397e9e2e4/CG000315_ChromiumNextGEMSingleCell3__v3.1_Dual_Index_RevA.pdf

27. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017 Jan 16;8:14049.

28. Young MD, Behjati S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. Gigascience [Internet]. 2020 Dec 26 [cited 2021 Sep 17];9(12). Available from: https://academic.oup.com/gigascience/article/9/12/giaa151/6049831

29. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021 Jun 24;184(13):3573–87.e29.

30. Breiman L. Random Forests. Mach Learn. 2001 Oct 1;45(1):5–32.

31. Liaw A, Wiener M. randomForest [Internet]. cran.r-project.org. 2018. Available from: https://cran.r-project.org/web/packages/randomForest/randomForest.pdf; https://www.stat.berkeley.edu/~breiman/RandomForests/

32. DePasquale EAK, Schnell D, Dexheimer P, Ferchen K, Hay S, Chetal K, et al. cellHarmony: cell-level matching and holistic comparison of single-cell transcriptomes. Nucleic Acids Res. 2019 Dec 2;47(21):e138.

33. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, et al. Comprehensive Integration of Single-Cell Data. Cell. 2019 Jun 13;177(7):1888–902.e21.

34. Alquicira-Hernandez J, Sathe A, Ji HP, Nguyen Q, Powell JE. scPred: accurate supervised method for cell-type classification from single-cell RNA-seq data. Genome Biol. 2019 Dec 12;20(1):264.

35. Gudenas B. Gene Set Enrichment Analysis in R [Internet]. Bioinformatics Breakdown. 2021 [cited 2021 Sep 17]. Available from: https://bioinformaticsbreakdown.com/how-to-gsea/

36. Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. GAGE: generally applicable gene set enrichment for pathway analysis. BMC Bioinformatics. 2009 May 27;10:161.

37. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis [Internet]. bioRxiv. 2021 [cited 2021 Sep 17]. p. 060012. Available from: https://www.biorxiv.org/content/10.1101/060012v3.full-text

38. Leukemia and Lymphoma Society. Facts About Blastic Plasmacytoid Dendritic Cell Neoplasm (BPDCN) [Internet]. Leukemia and Lymphoma Society. 2019 [cited 2021]. Available from: https://www.lls.org/sites/default/files/2021-05/FSHP2_BPDCN_FINAL_2019.pdf

39. Burel JG, Qian Y, Arlehamn CL, Weiskopf D, Zapardiel-Gonzalo J, Taplitz R, et al. An Integrated Workflow To Assess Technical and Biological Variability of Cell Population Frequencies in Human Peripheral Blood by Flow Cytometry. The Journal of Immunology. 2017 Feb 15;198(4):1748–58.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470599; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

40. Pemmaraju N, Lane AA, Sweet KL, Stein AS, Vasu S, Blum W, et al. Tagraxofusp in Blastic Plasmacytoid Dendritic-Cell Neoplasm. N Engl J Med. 2019 Apr 25;380(17):1628–37.

41. Cantaert T, Baeten D, Tak PP, van Baarsen LGM. Type I IFN and TNFα cross-regulation in immune-mediated inflammatory disease: basic concepts and clinical relevance. Arthritis Res Ther. 2010 Oct 28;12(5):219.

42. Magro CM, Porcu P, Schaefer J, Erter JW, Furman RR, Shitabata PK, et al. Cutaneous CD4+ CD56+ hematologic malignancies. J Am Acad Dermatol. 2010 Aug;63(2):292–308.

43. Beyer M, Abdullah Z, Chemnitz JM, Maisel D, Sander J, Lehmann C, et al. Tumor-necrosis factor impairs CD4+ T cell–mediated immunological control in chronic viral infection. Nat Immunol. 2016 Mar 7;17(5):593–603.

44. Renosi F, Roggy A, Giguelay A, Soret L, Viailly P-J, Cheok M, et al. Transcriptomic and genomic heterogeneity in blastic plasmacytoid dendritic cell neoplasms: from ontogeny to oncogenesis. Blood Adv. 2021 Mar 9;5(5):1540–51.

45. Togami K, Chung SS, Madan V, Booth CAG, Kenyon CM, Cabal-Hierro L, et al. Sex-biased ZRSR2 mutations in myeloid malignancies impair plasmacytoid dendritic cell activation and apoptosis. Cancer Discov [Internet]. 2021 Oct 6; Available from: http://dx.doi.org/10.1158/2159-8290.CD-20-1513

46. Sanchez-Correa B, Bergua JM, Campos C, Gayoso I, Arcos MJ, Bañas H, et al. Cytokine profiles in acute myeloid leukemia patients at diagnosis: survival is inversely correlated with IL-6 and directly correlated with IL-10 levels. Cytokine. 2013 Mar;61(3):885–91.

47. Tian T, Wang M, Ma D. TNF-α, a good or bad factor in hematological diseases? Stem Cell Investig. 2014 Jun 1;1:12.

48. Noh J-Y, Seo H, Lee J, Jung H. Immunotherapy in Hematologic Malignancies: Emerging Therapies and Novel Approaches. Int J Mol Sci [Internet]. 2020 Oct 27;21(21). Available from: http://dx.doi.org/10.3390/ijms21218000

49. Benci JL, Xu B, Qiu Y, Wu TJ, Dada H, Twyman-Saint Victor C, et al. Tumor Interferon Signaling Regulates a Multigenic Resistance Program to Immune Checkpoint Blockade. Cell. 2016 Dec 1;167(6):1540–54.e12.

50. Ma H, Yang W, Zhang L, Liu S, Zhao M, Zhou G, et al. Interferon-alpha promotes immunosuppression through IFNAR1/STAT1 signalling in head and neck squamous cell carcinoma. Br J Cancer. 2019 Feb;120(3):317–30.

51. Yost KE, Satpathy AT, Wells DK, Qi Y, Wang C, Kageyama R, et al. Clonal replacement of tumor-specific T cells following PD-1 blockade. Nat Med. 2019 Aug;25(8):1251–9.

52. Trottier AM, Cerquozzi S, Owen CJ. Blastic plasmacytoid dendritic cell neoplasm: challenges and future prospects. Blood Lymphat Cancer. 2017 Dec 11;7:85–93.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470599; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

53. Pagano L, Valentini CG, Grammatico S, Pulsoni A. Blastic plasmacytoid dendritic cell neoplasm: diagnostic criteria and therapeutical approaches. Br J Haematol. 2016 Jul;174(2):188–202.

10 Data Availability Statement

Raw and processed data generated for this study will be available in the Gene Expression Omnibus (GEO) at https://www.ncbi.nlm.nih.gov/geo/. The scripts used to generate the results in this paper can be found at the project GitHub repository: https://github.com/EDePasquale/BPDCN.

11 Figure Legends

Figure 1: Integration and labeling of five healthy bone marrow control samples.

(A) UMAP visualization of Seurat-integrated scRNA-seq data for 25,726 hematopoietic cells from five normal BM aspirates.
(B) Expression scores for lineage signatures overlaid on the UMAP of healthy BM in (A).
(C) UMAP shows 17 clusters of cells with similar transcriptional states, identified by Seurat clustering and sub-clustering of healthy BM.
(D) Stacked barplots show the frequencies of cell types in five normal BMs.

Figure 2: Classification of BPDCN samples using healthy references.

(A) Stacked barplots show the frequencies of cell types in five BPDCN samples as classified by four algorithms: RF = random forest, CH = cellHarmony, TD = Seurat TransferData, and SP = scPred.
(B) UMAP visualization of cells from five BPDCN samples in the same UMAP space as the integrated reference. Colors in the cell-type proportions and UMAP visualizations are coded by the legend. For comparison with healthy controls, reference UMAP and proportions for each healthy sample are provided in the black box.

Figure 3: Gene set enrichment shows enrichment of IFNA and depletion of TNFA related genes in BPDCN.

(A) Heatmap shows Gene Set Enrichment Analysis enrichment scores for each cell type and sample (columns) and significantly enriched Hallmark pathway (rows), determined by a P-value ≤ 0.05 in both GSEA tests. Red indicates high enrichment scores and blue indicates low enrichment scores.
(B) Violin plot of IFN Alpha Response gene set scores (y-axis) for CD8+ Memory T-cells from each control and BPDCN sample (x-axis). Each dot within the violin plot represents a cell, with the box in the middle of the violin representing the median and interquartile range of the data. Healthy BM samples are colored in dark blue, BPDCN samples are colored using the sample color scheme from (A).
(C) Violin plot of TNFA Signaling via NFKB gene set scores (y-axis) for CD8+ Memory T-cells from each control and BPDCN sample (x-axis). P-values were calculated by comparing the medians of n = 5 healthy controls to n = 4 BPDCN samples.

Figure 4: Expression of IFNA and TNFA associated gene sets.

(A) Heatmap shows log expression values for genes in the IFN Alpha Response gene set (rows) for each sample and cell type (columns), clustered by row. Genes were filtered to those significant ($p <$

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470599; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

0.05) after multiple testing correction with an expression value for one sample above 0.5. Red indicates higher expression and blue indicates lower expression.
(B) Heatmap shows log expression values for genes in the TNFA Signaling via NFKB gene set.

Figure 5: T-cell clonotype sizes in healthy controls and BPDCN patients.

(A-E) Bar plots of the top 50 ranked normalized clone sizes for BM 1-5 healthy controls. The normalized clone size is the percentage of cells expressing the same TCR of all cells in which a TCR was detected. For each clone (stacked bar), the colors indicate the proportion comprised of that cell type.
(F-I) Bar plots of ranked clones for BPDCN 1-4 patient samples. Axis break in (F) is indicated by dashed line and broken bar.

Figure 6. TCR sequences are detected in T-cells and clonally expanded in CD8+ Memory T-cells.

(A) Pie charts show cell type assignments of cells in which a TCR sequence was detected.
(B) Dot plot shows normalized clone size of all the T-cells in which a TCR sequence was detected. The distribution of values differs between T-cell subsets ($P < 2.2E-16$).

Figure 7: CD8+ Memory T-cells exhibit high expression of T-cell exhaustion-associated genes and larger TCR clones.

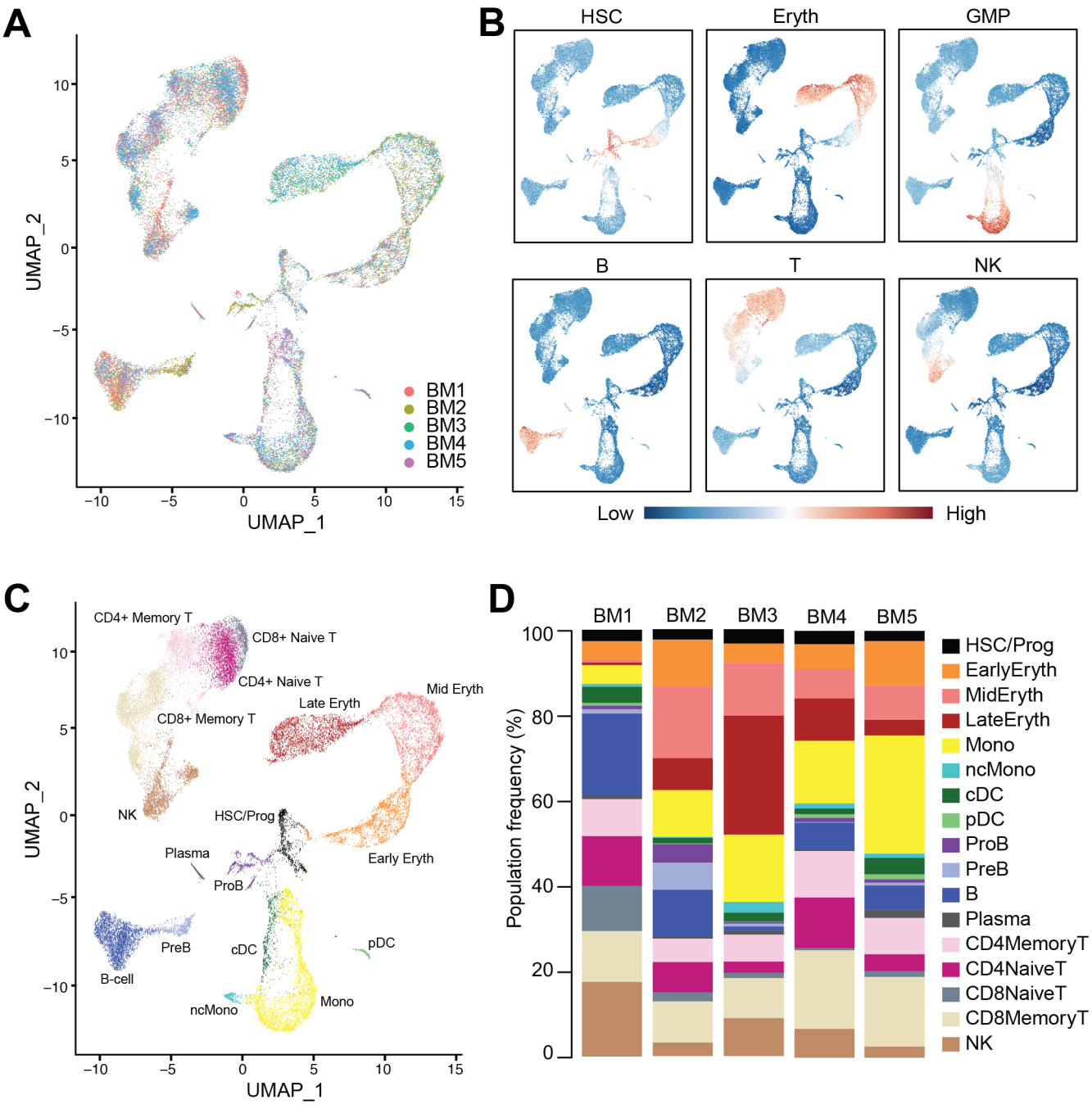
UMAP plots of T-cell populations in each BPDCN sample that contains T-cells (BPDCN 1-4), separated by column. UMAPs are colored by cluster (first row); T-cell exhaustion score, with red indicating high exhaustion and gray indicating low exhaustion (second row); and normalized clone size, percentage of all cells in the dataset that share the same TCR sequence, with blue indicating high clone size and gray indicating low clone size (third row).

Figure 8: Exhaustion scores positively correlate with clone size in BPDCN samples.

(A) Scatter plots show mean exhaustion score per cluster (x-axis) versus mean clone size (y-axis). Dots are scaled by the number of cells of each cluster in the dataset.
(B) Violin plot shows T-cell exhaustion signature scores in CD8+ Memory T-cells in healthy controls (BM 1-5) and BPDCN samples with T-cells (BPDCN 1-4).
(C-F) Box plots of cells binned by TCR clone size (x-axis) versus exhaustion score (y-axis) for T-cells in each BPDCN 1-4 sample as C-F. Parenthetical values on the x-axis labels indicate the normalized clone size.

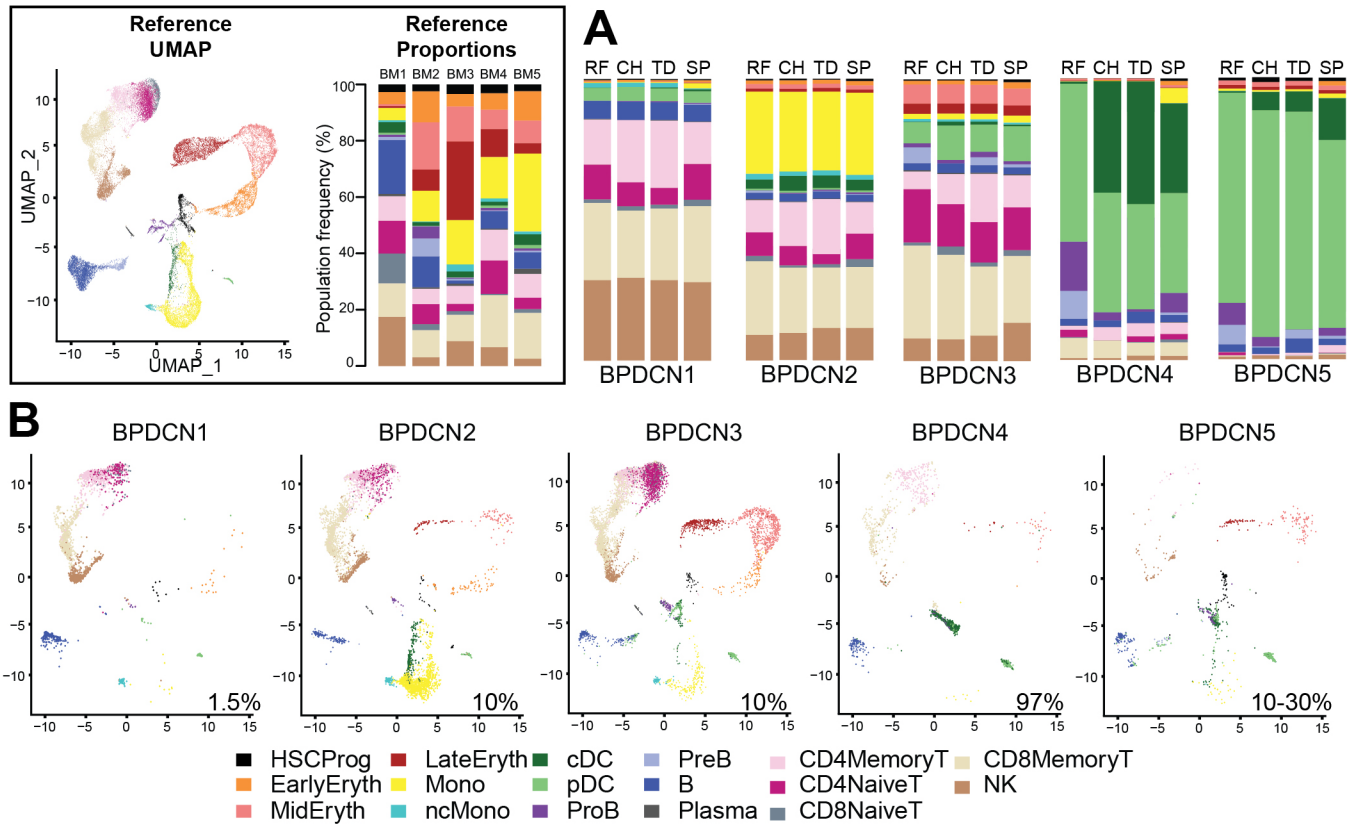
bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470599; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 1



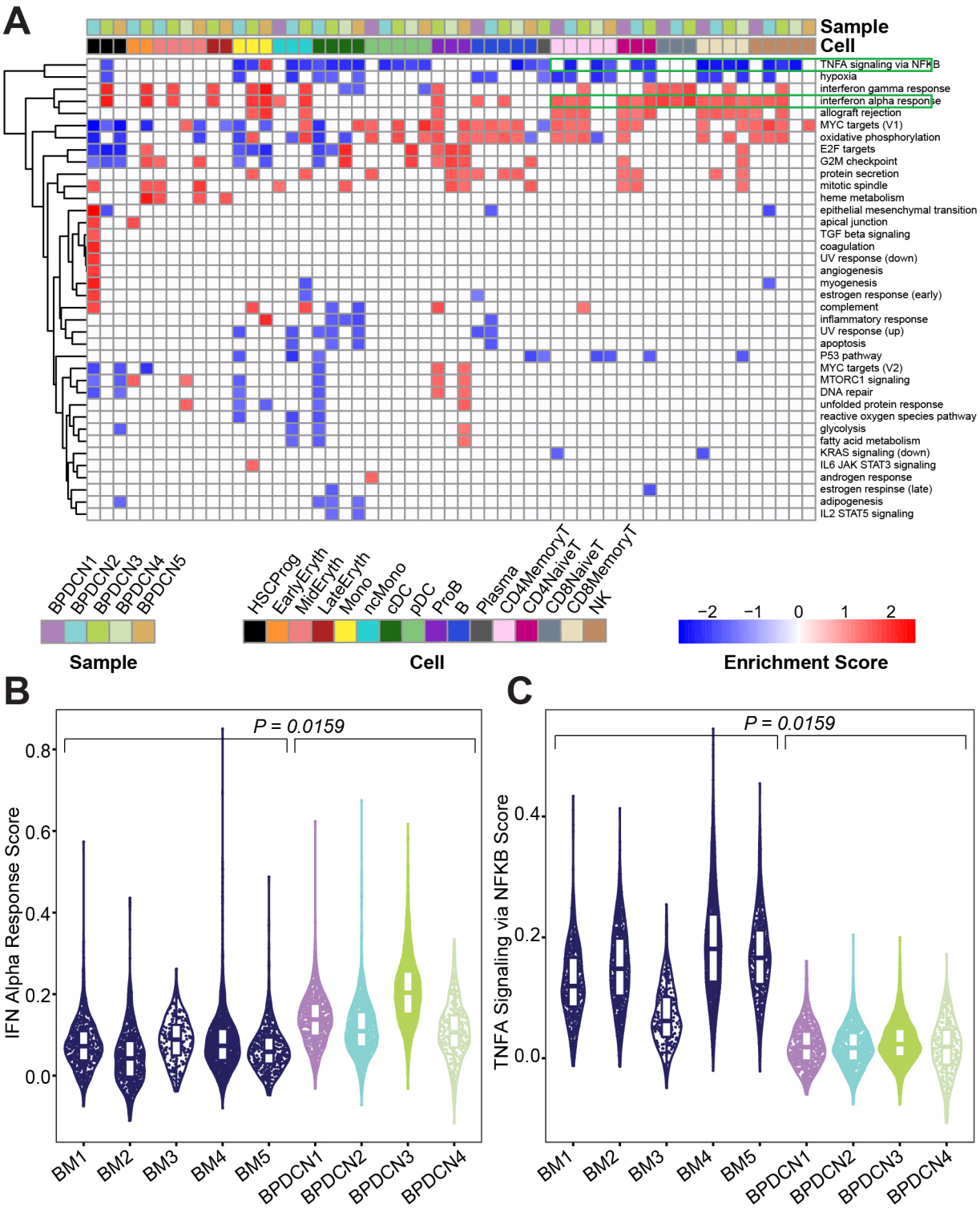
bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470599; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 2



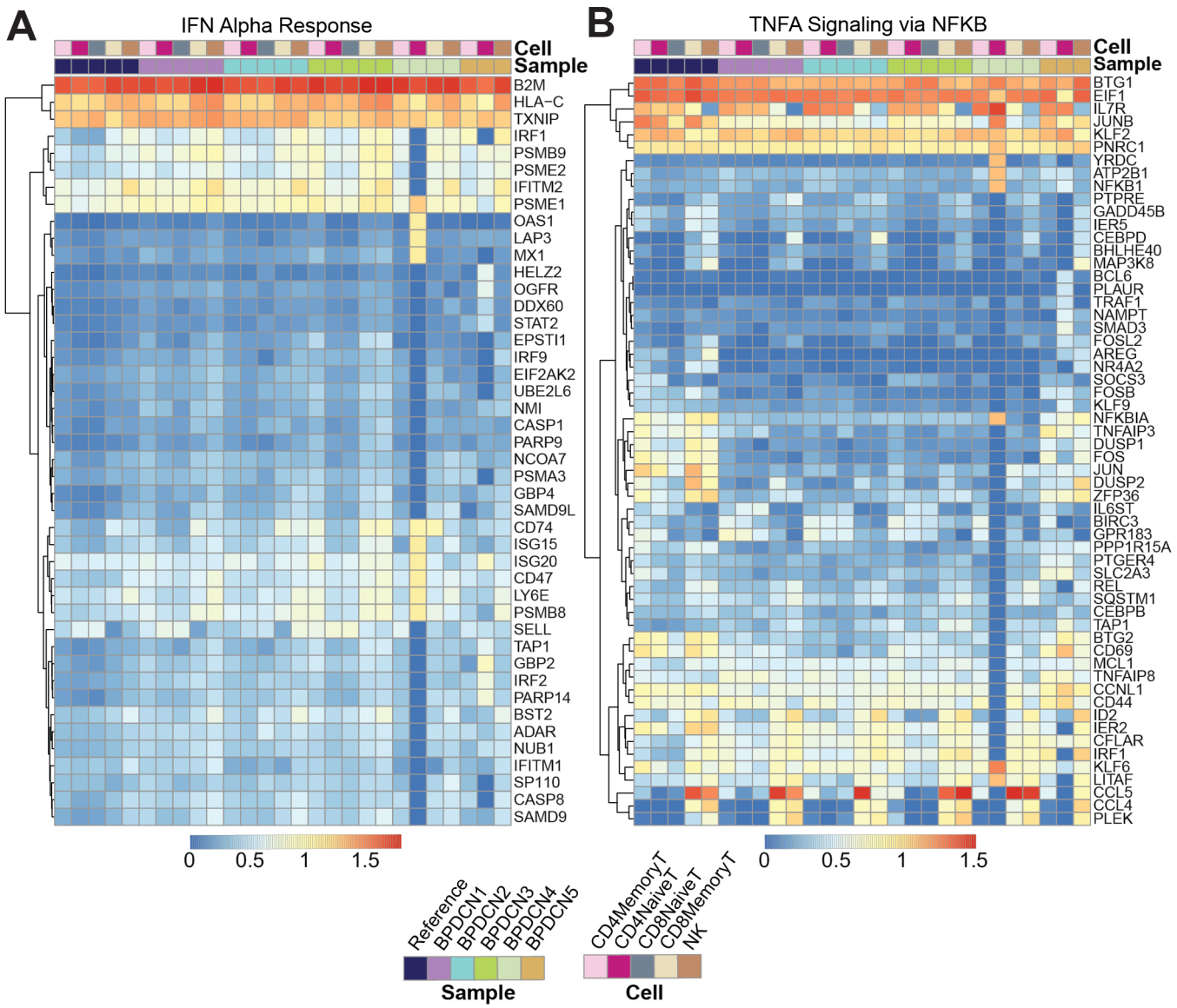
bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470599; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 3



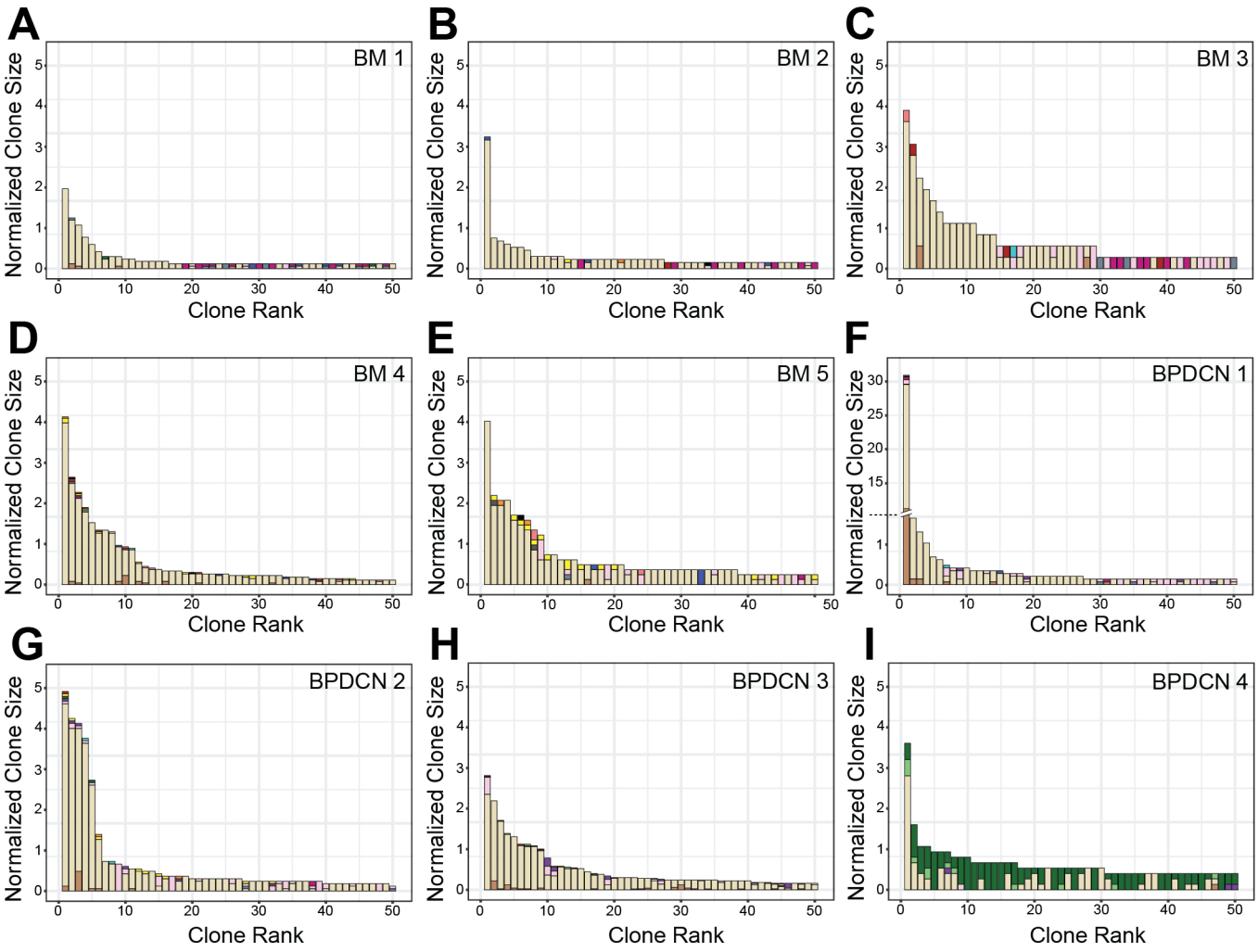
bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470599; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 4



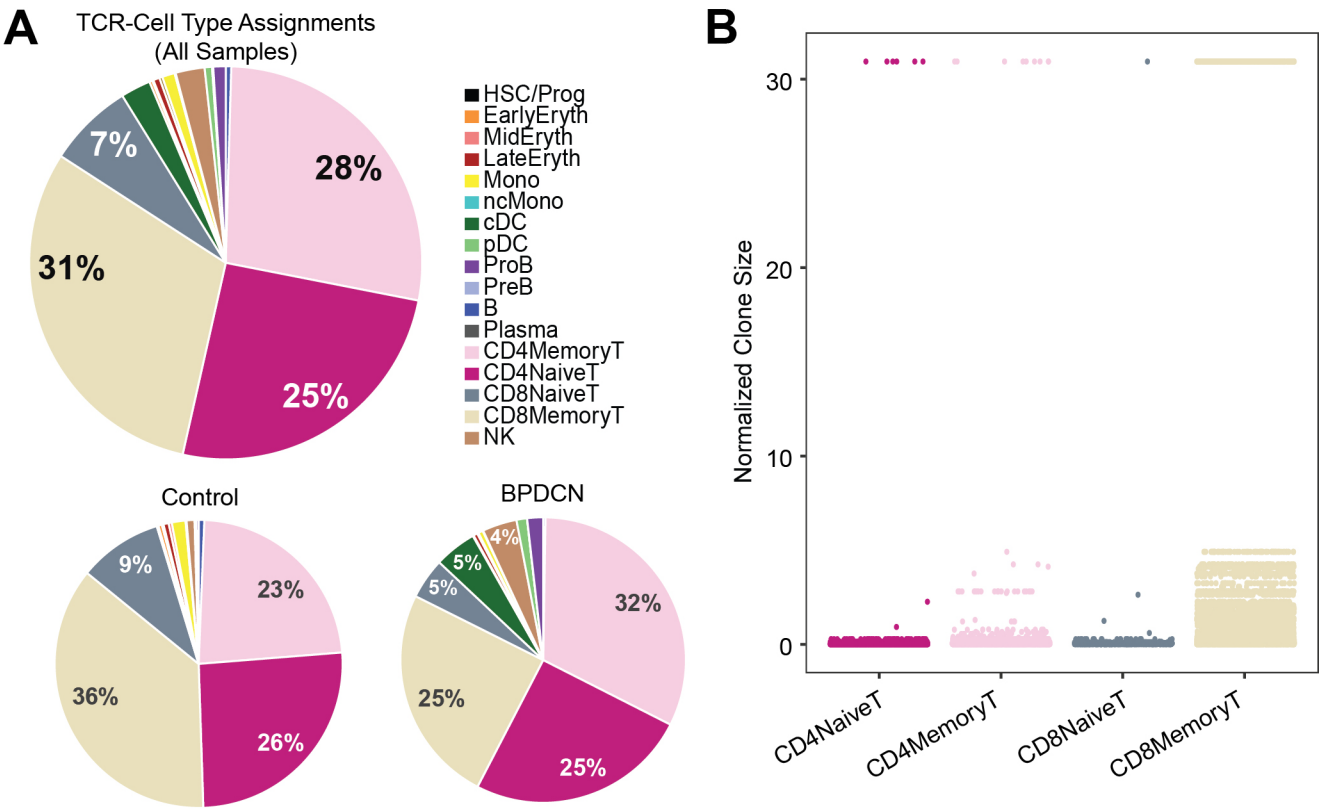
bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470599; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 5



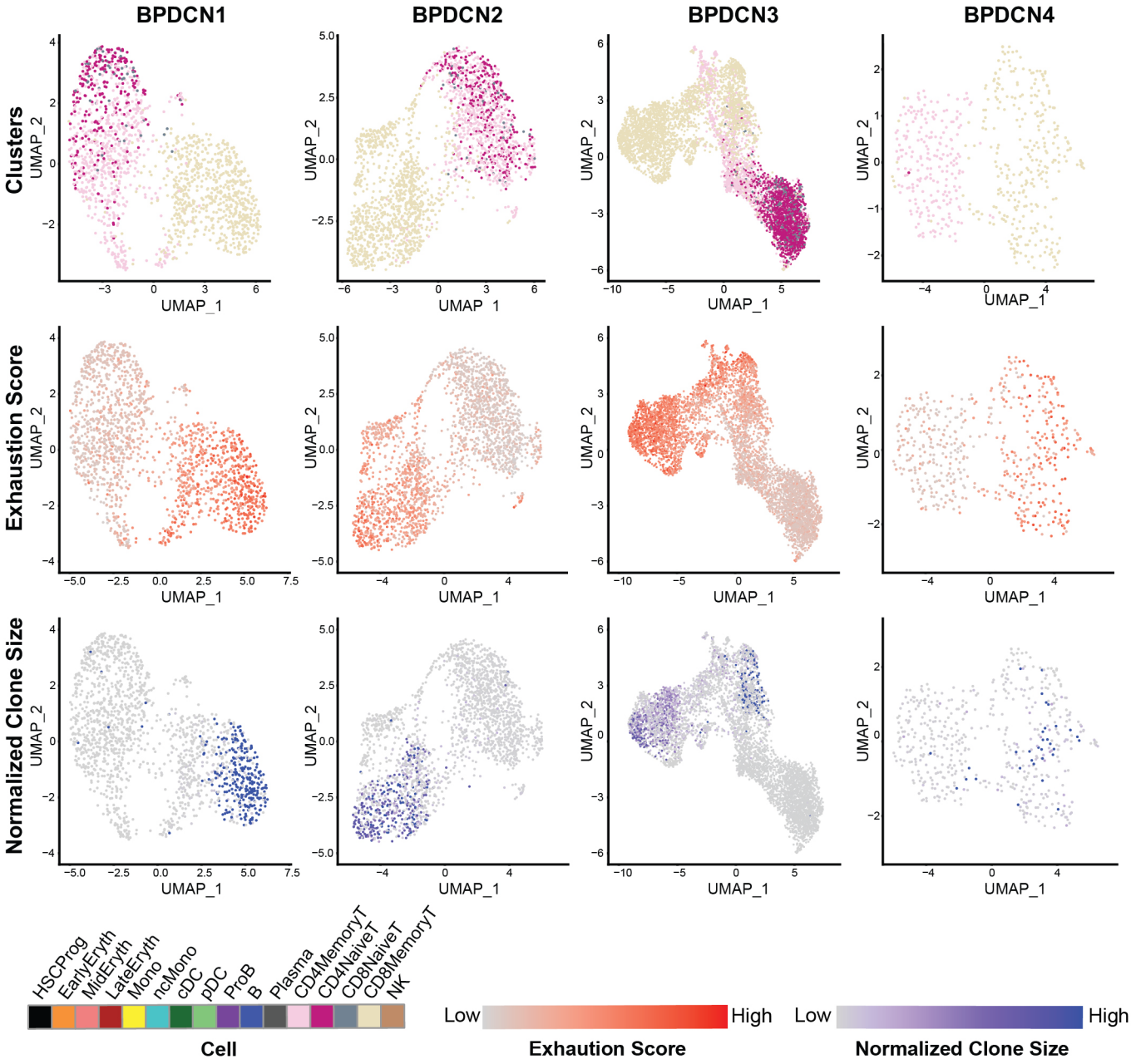
bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470599; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 6



bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470599; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 7



bioRxiv preprint doi: https://doi.org/10.1101/2021.12.01.470599; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

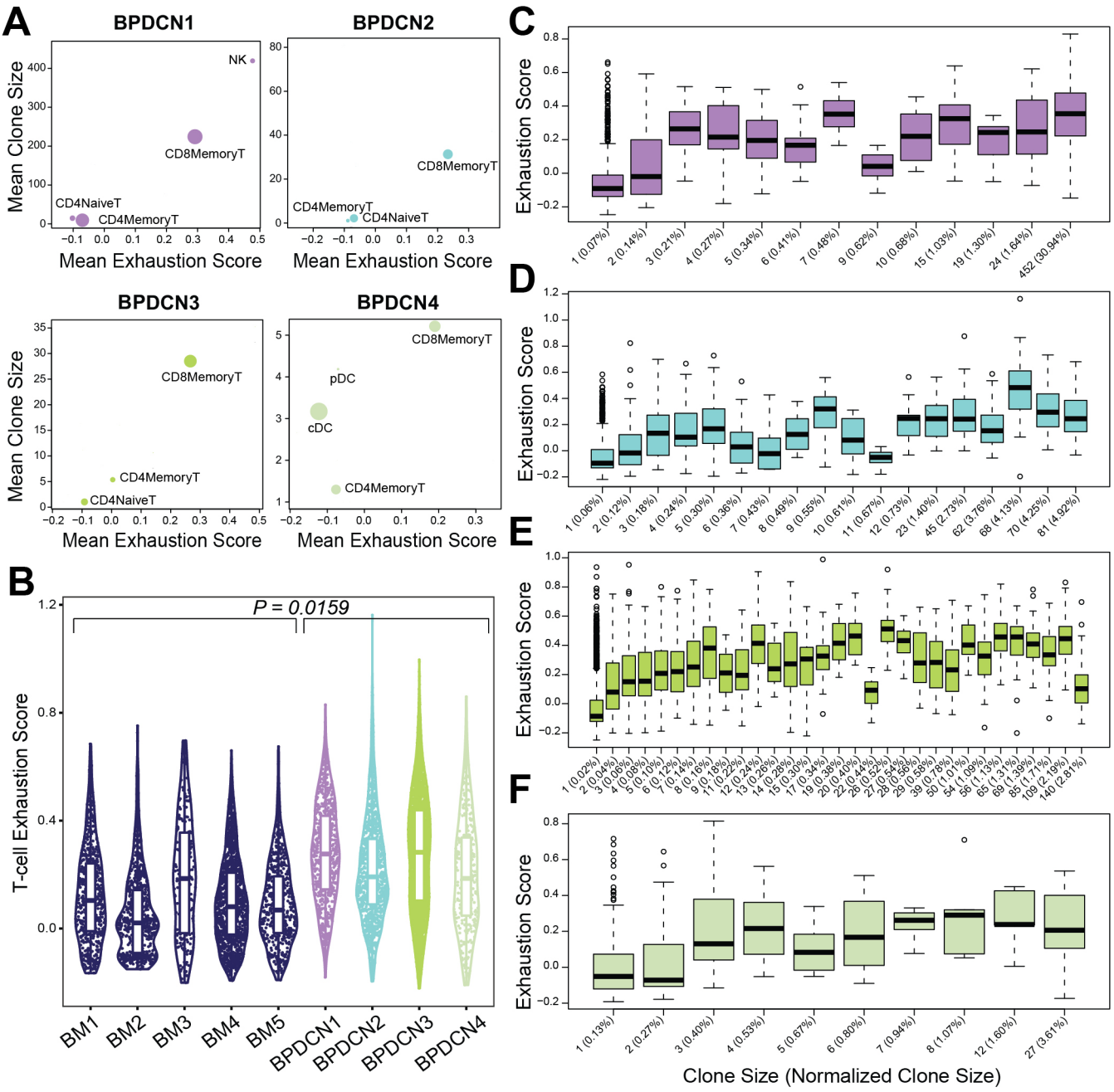



Figure 8