

T-cell receptor architecture and clonal tiding provide insight into the transformation trajectory of peripheral T-cell lymphomas

Edith Willscher,^{1*} Christoph Schultheiss,^{2,3*} Lisa Paschold,¹ Franziska Lea Schümann,¹ Paul Schmidt-Barbo,² Benjamin Thiele,^{2,3} Marcus Bauer,⁴ Claudia Wickenhauser,⁴ Thomas Weber¹ and Mascha Binder^{2,3}


¹Internal Medicine IV, Oncology/Hematology, Martin-Luther-University Halle-Wittenberg, Halle, Germany; ²Department of Biomedicine, Translational Immuno-Oncology, University of Basel, Basel, Switzerland; ³Division of Medical Oncology, University Hospital Basel, Basel, Switzerland and ⁴Department of Pathology, Martin-Luther-University Halle-Wittenberg, Halle, Germany

**EW and CS contributed equally as first authors.*

Correspondence: M. Binder
mascha.binder@usb.ch

Received: March 1, 2024.
Accepted: August 22, 2024.
Early view: August 29, 2024.

<https://doi.org/10.3324/haematol.2024.285395>

©2025 Ferrata Storti Foundation
Published under a CC BY-NC license 

Supplementary Methods:

T cell receptor (TCR) immune repertoire sequencing and data analysis

The ReliaPrep™ FFPE gDNA Miniprep System (Promega, Madison, United States) was used for isolation of genomic DNA from FFPE tissue. The V(D)J rearranged TRBV loci were amplified in a multiplex PCR using the Phusion™ High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, USA) and the BIOMED2-FR2/FR3 -TRB-B primer pools. The primers were purchased from Metabion International AG (Martinsried, Germany). PCR amplicons were purified using the Agencourt AMPure XP paramagnetic beads (Beckman Coulter) and subjected to a second PCR for the addition of 7-nucleotide single indices and Illumina adapter sequences. After bead-based purification, PCR amplicons were quantified using the Qubit system (Thermo Fisher) and pooled to a final concentration of six nM. Pools were quality controlled on an Agilent 2100 Bioanalyzer (Agilent Technologies, Böblingen, Germany). Sequencing and demultiplexing was performed on an Illumina MiSeq sequencer (Illumina, San Diego, USA) with a 601-cycle paired-end run and V3-chemistry. The MiXCR framework¹ v.3.0.12 was used for sequence alignment and clonotype assignment. Non-productive reads and clonotypes with less than two reads were discarded. Each unique nucleotide CDR3 sequence was considered a clone. For analysis of repertoire metrics, healthy immune repertoires were proportionally normalized to 30,000 productive reads. Clonality of TRB repertoires was calculated according to the formula $1 - H'/\log_2(S)$ with H' being the Shannon diversity index and S the total number of TRB clonotypes (=richness) in an individual repertoire. Principal component analysis (PCA) of V Gene usage and their contributions were calculated using R package ade4. Data analysis and plotting was performed with Rstudio (version 2023.03.1+446). Monoclonality was assumed if the frequency of the most abundant clonotype in the repertoire matched the respective tumor cell fraction of the sample (+/-20%). Moreover, monoclonality was assumed if the frequency

of the first two clonotypes of the repertoire showed a frequency that matched the tumor cell fraction ($\pm 20\%$), but only if these two clonotypes had a similar frequency ($\pm 20\%$) to account for biallelic rearrangements. All other cases were classified as oligoclonal. Tumor cell fraction was determined by a reference pathologist.

Dynamics of TCR clones

Longitudinal dynamics of lymphoma subclones from patient 003 were calculated according to the approach by Minervina et al^{2,3}. For this, a Euclidean distance matrix of the normalized frequencies of the top 1000 TCR clonotypes for each sampling time point was generated. Clusters were then identified using hierarchical clustering and visualized over time using PCA. Mean trajectories for all four patterns were shown at normalized frequencies.

Tissue microarray staining for TRBV20-1 usage

Available FFPE tissue from our TCR NGS cohort as well as additional T cell lymphoma samples were analyzed using a tissue microarray as reported by Schümann et al⁴. Antigen retrieval was performed using a TRIS-EDTA Buffer pH 9.0 antigen retrieval solution (ZUC029-500, Zytomed, Germany) followed by a blocking step (ZUC007-100, Zytomed, Germany). The TCR V β 2-PE antibody (diluted 1:100, IM2213, Beckmann Coulter, CA, USA) was incubated at room temperature for 60 minutes and a counterstaining with DAPI (Akoya Biosciences, MA, USA) was performed as described in Bauer et al⁵. After a washing step, all slides were mounted with Vectashield Antifade Mounting Medium (H-1000-10, Vector Laboratories, CA; USA) and visualized with a PhenolImager HT (Akoya Biosciences, MA, USA). Tissue microarray (TMA) spot intensity was determined using ImageJ (National Institutes of Health) according to the online protocol provided by the Keith R. Porter Imaging Facility.⁶

Gene panel profiling

Profiling of hotspots mutations associated with T cell lymphomas was performed using a targeted DNA Custom Panels from Qiagen (Hilden, Germany). Genes with recurrent mutations in T cell lymphoma were selected from cbiportal (Supplementary Table S2). Sequencing libraries were constructed using Qiaseq Targeted DNA Custom Panels (Qiagen, Hilden, Germany). The Qubit high-sensitivity double-strand DNA assay kit (Thermo Fisher) was used for quantification of libraries and the Agilent 2100 Bioanalyzer (Agilent) for final quality control. Libraries were sequenced on a Illumina NextSeq 500 platform with 2 x 150 cycles at an average coverage of 52,700 reads per target region. Variant calling of unique molecular identifier (UMI) was performed using CLC Workbench (Quiagen). Mutations were considered as positive if they were found with a variant allele frequency (VAF) exceeding 10% at a read depth of more than 70 reads. To filter for disease relevant mutations, common single nucleotide polymorphisms (SNPs) stated by dbSNP were discarded as well as synonymous variants.

Cell sorting and single-cell transcriptomic profiling

Sorting of lymphoma cells was performed based on their aberrant CD4⁺/CD8⁺ immunophenotype (T-PLL case) or on aberrant CD3 surface expression (AITL case) from cryopreserved peripheral blood mononuclear cells (PBMCs) of these two T cell lymphoma patients. For that, the anti-CD3-APC-H7 (clone SK7, BD Biosciences), anti-CD4-PacificBlue (clone RPA-T4, Biolegend) and the anti-CD8-FITC (clone SK1, BD Biosciences) antibodies were used on a BD FACSAriaIII cell sorter with a 100 µm nozzle. Cells were processed on a 10X Chromium Controller (10X Genomics, Pleasanton, CA, USA) within 1h after collection. Single-cell libraries were generated using the Next GEM Single Cell 5' Kit v2 and Chromium Single Cell Human TCR Amplification kits according to the manufacturer instructions to

detect coupled TCR beta and alpha chains. For the integrated dataset we used two healthy samples provided in Herrera et al⁷ (sample HC1) and one 10X resource dataset.⁸ The two lymphoma samples and the two healthy samples were merged together using package Seurat (v 5.0.0).⁹ Cells with high mitochondrial content (>10%) and more than 2500 RNA features were excluded. Normalization and detection of the top 2000 variable features was done individually for each dataset. To merge data sets, integration anchors were calculated using function *FindIntegrationAnchors* and datasets were integrated with *IntegrateData* to one object. After scaling, PCA and t-distributed stochastic neighbor embedding (tSNE) calculation were performed on 15 dimensions. T cell clusters were assigned according to subset markers found with function *FindAllMarkers* and selected T cell population marker genes. Module scores for feature expression programs were calculated with function *AddModuleScore*. The modules contained following genes: the MYC module contains targets of MYC (CCT3, DUT, RPL34, RPS5, RPL22, RPLP0, RPS6, RPL6, EEF1B2, FBL, SNRPD2, RACK1, RPS3, RPL14, RPS2, RPL18, RPS10, PPIA, EIF3D); the p53 module comprises genes which regulate signal transduction by p53 positively (RPS15, RPS7, UBB, RPL23, RPS20, RPL37); the NF-kB module contains genes, which are expressed when NFkB is activated (UBB,UBC,RACK1,RPS27A,UBA52); TGFb module encloses genes of the TGFb pathway (APP, MAP2K4, MAP2K1, CREBBP, SMAD4, SMURF2, SMURF1, ITGA2, LIMK2, NEDD4L, NEDD9, NUP153, MAPK14, RUNX2, PIAS2, PIAS1, TGFBR3, SKI, MAPK8, RBL1, BTRC, MAP2K6); the STAT5 module contains genes involved in IL2/STAT5 signaling (PHTF2, ABCB1, AHNAK, PTGER2, CTSZ, FURIN, CST7, NDRG1, SOCS2, NFKBIZ, PIM1, LRIG1, CTLA4, SNX9, MAP3K8, ITGAV, TNFRSF4, GABARAPL1, IL4R, GADD45B, CISH, NCOA3, IL10RA, TNFRSF18, FLT3LG, RHOH, TNFRSF1B, IGF2R, HOPX, SERPINB6, IL2RA, BHLHE40, BCL2, LTB, IL18R1).

Fastq files from V(D)J libraries were analysed with cellranger vdj pipeline, and the filtered results were integrated with gene expression data with package scRepertoire (v 1.4.0).

1. Bolotin DA, Poslavsky S, Mitrophanov I, et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods*. 2015;12(5):380-381.
2. Minervina AA, Pogorelyy MV, Komech EA, et al. Primary and secondary anti-viral response captured by the dynamics and phenotype of individual T cell clones. *Elife*. 2020;9.
3. Minervina AA, Komech EA, Titov A, et al. Longitudinal high-throughput TCR repertoire profiling reveals the dynamics of T-cell memory formation after mild COVID-19 infection. *Elife*. 2021;10.
4. Schumann FL, Gross E, Bauer M, et al. Divergent Effects of EZH1 and EZH2 Protein Expression on the Prognosis of Patients with T-Cell Lymphomas. *Biomedicines*. 2021;9(12).
5. Bauer M, Vaxevanis C, Bethmann D, et al. Multiplex immunohistochemistry as a novel tool for the topographic assessment of the bone marrow stem cell niche. *Methods Enzymol*. 2020;635:67-79.
6. <https://kpif.umbc.edu/image-processing-resources/imagej-fiji/determining-fluorescence-intensity-and-positive-signal/>.
7. Herrera A, Cheng A, Mimitou EP, et al. Multimodal single-cell analysis of cutaneous T-cell lymphoma reveals distinct subclonal tissue-dependent signatures. *Blood*. 2021;138(16):1456-1464.
8. Integrated GEX and VDJ analysis of Connect generated library from human PBMCs. <https://www.10xgenomics.com/resources/datasets/>.
9. Hao Y, Stuart T, Kowalski MH, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol*. 2024;42(2):293-304.

Supplementary Table 1: Characteristics of individual patients and samples.*

Pat. ID	Diagnosis	DOSC	Tissue	Age decade at diagnosis	Sex
001_1_1	AITL	ID	3	7	f
001_1_2	AITL	ID	1	7	f
002_1	T-LGLL	ID	1	8	m
003_1	AITL	ID	3	8	m
003_2	AITL	1st relapse/progression	1	8	m
003_3	AITL	2nd relapse/progression	1	8	m
003_4	AITL	3rd relapse/progression	1	8	m
004_1	AITL	ID	3	8	m
005_1	AITL	1st relapse/progression	3	7	m
006_1	AITL	ID	3	5	m
007_1	AITL	ID	3	8	m
008_1	AITL	ID	3	10	m
009_1	AITL	ID	3	8	m
010_1	AITL	ID	3	8	m
010_2	AITL	suspected relapse	3	8	m

010_3	AITL	relapse/progression	1	8	m
010_4	AITL	suspected relapse	1	8	m
011_1	AITL	ID	1	6	f
012_1	PTCL, NOS	ID	2	6	m
013_1	PTCL, NOS	ID	3	7	m
013_2	PTCL, NOS	1st relapse/progression	1	7	m
014_1	AITL	ID	3	6	f
014_2	AITL	1st relapse/progression	3	6	f
015_1	PTCL, NOS	1st relapse/progression	2	4	m
016_1	PTCL, NOS	ID	3	8	f
017_1	PTCL, NOS	1st relapse/progression	4	7	m
017_2	PTCL, NOS	2nd relapse/progression	4	7	m
018_1	PTCL, NOS	ID	3	6	f
019_1	ALCL	ID	10	6	m
020_1	T-PLL	ID	4	7	m
021_1	NKTCL	ID	5	4	m
022_1	NKTCL	ID	1	6	m
023_1	ALCL	ID	3	6	m
024_1	PTCL-TFH	ID	3	6	m
025_1	PTCL, NOS	ID	2	7	m
026_1	AITL	ID	3	7	m
027_1	PTCL, NOS	ID	3	7	m
028_1	SS	3rd relapse/progression	1	7	f
029_1	PTCL, NOS	ID	3	7	f
029_2_1	PTCL, NOS	1st relapse/progression	3	7	f
029_2_2	PTCL, NOS	1st relapse/progression	3	7	f
030_1	MEITL	ID	8	7	f
030_2	MEITL	1st relapse/progression	1	7	f
030_3	MEITL	1st relapse/progression	9	7	f
030_4_1	MEITL	1st relapse/progression	8	7	f
030_4_2	MEITL	1st relapse/progression	8	7	f
031_1_1	PTCL, NOS	ID	5	7	m
031_1_2	PTCL, NOS	ID	1	7	m
032_1	ALCL	ID	1	8	m
033_1	PTLD	ID	3	8	m
034_1	MF	ID	4	8	m
035_1	AITL	ID	3	8	m
035_2	AITL	1st relapse/progression	1	8	m

036_1_1	T-PLL	ID	3	8	f
036_1_2	T-PLL	ID	1	8	f
037_1	T-LGLL	ID	6	8	m
037_2	T-LGLL	1st relapse/progression	1	8	m
037_3	T-LGLL	2nd relapse/progression	1	8	m
038_1	AITL	ID	3	9	f
039_1	SS	1st relapse/progression	1	7	f
040_1	ALCL	ID	3	6	m
041_1	PTCL, NOS	ID	1	9	m
042_1	SPTCL	ID	2	6	m
043_1	ALCL	ID	3	6	f
044_1	PTCL, NOS	ID	3	6	f
045_1	AITL	ID	3	7	f
046_1	EATL	ID	8	5	m
047_1	ALCL	ID	1	6	m
048_1	AITL	ID	3	6	f
049_1	PTCL-TFH	ID	3	9	m
050_1	ALCL	ID	5	6	f
051_1	AITL	ID	1	6	f
052_1	PTCL, NOS	ID	1	5	m
053_1	PTCL, NOS	1st relapse/progression	4	6	f
054_1	PTCL-TFH	ID	3	6	m
055_1	T-PLL	ID	10	7	m
056_1	AITL	1st relapse/progression	10	4	m

*ID = initial diagnosis, DOSC = date of sample collection

Tissue code: 1 = bone marrow, 2 = connective tissue, 3 = lymph node, 4 = skin, 5 = nasopharynx, 6 = spleen, 8 = colon, 9 = peritoneum, 10 = Peripheral blood mononuclear cells (PBMC)

PTCL, NOS = peripheral T cell lymphoma, not otherwise specified

AITL = Angioimmunoblastic T cell lymphoma

ALCL = Anaplastic large cell lymphoma

NKTCL = Extranodal NK-/T-cell lymphoma, nasal type

PTCL-TFH = Peripheral T cell lymphoma, T follicular helper phenotype

MF = Mycosis fungoides

SS = Sézary Syndrom

T-PLL = T cell prolymphocytic leukemia

T-LGLL = T cell large granular lymphocytic leukemia

EATL = Enteropathy-associated T cell lymphoms

MEITL = Monomorphic epitheliotropic intestinal T cell lymphoma

PTLD = Post-transplant lymphoproliferative disorder

SPTCL = Subcutaneos panniculitis-like T cell lymphoma

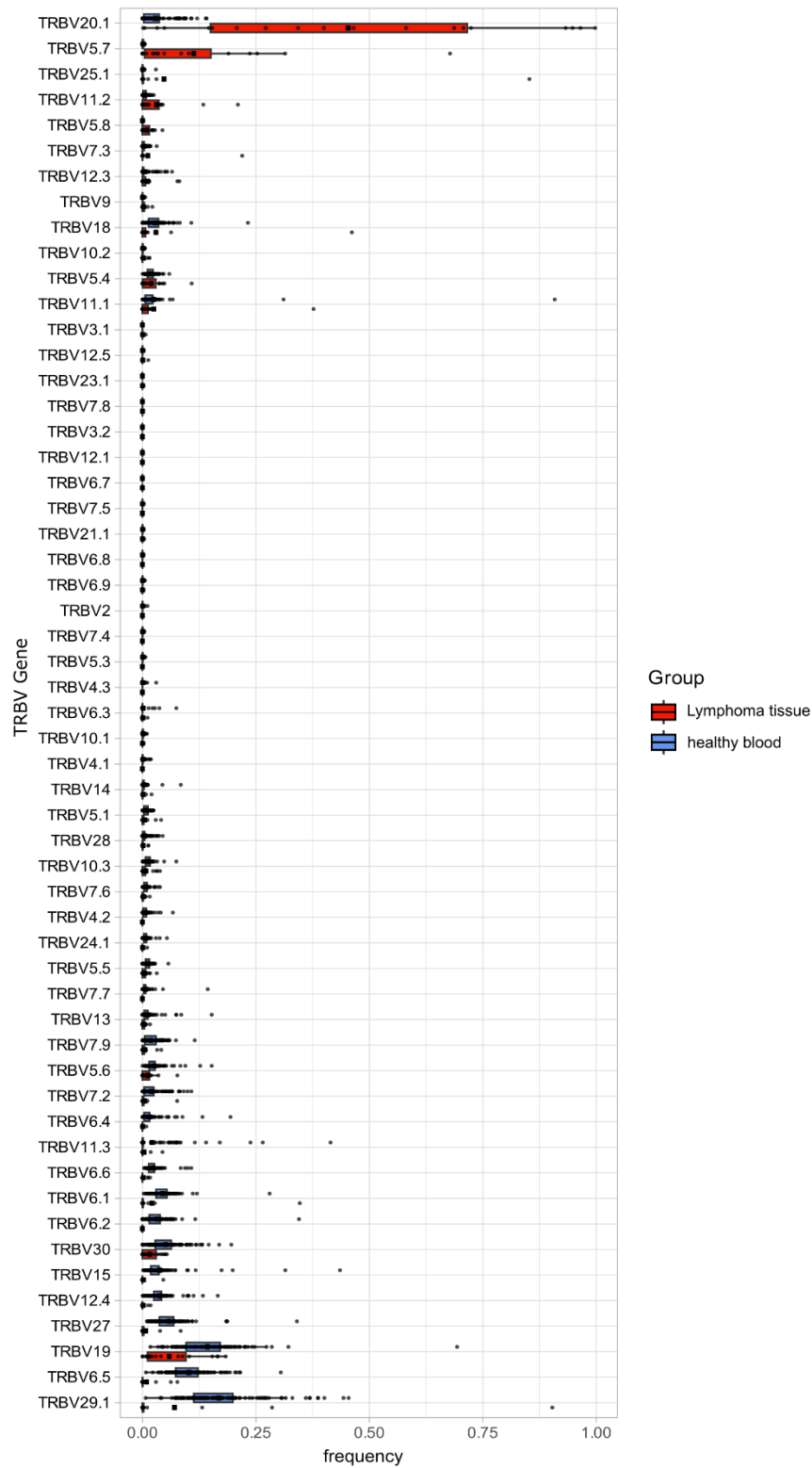
Supplementary Table 2: QIAseq custom DNA panel, covered genes and regions.

chromosome	start	end	covered gene
chr1	120457925	120459275	NOTCH2
chr15	90631832	90631844	IDH2
chr15	90631928	90631940	IDH2
chr16	24135286	24135310	PRKCB
chr16	24166000	24166180	PRKCB
chr16	24183585	24183685	PRKCB
chr16	24185835	24185905	PRKCB
chr16	24192105	24192255	PRKCB
chr16	24196425	24196515	PRKCB
chr16	24196775	24196895	PRKCB
chr16	24202405	24202555	PRKCB
chr16	24231275	24231325	PRKCB
chr17	7572921	7573013	TP53
chr17	7573921	7574038	TP53
chr17	7576531	7576589	TP53
chr17	7576619	7576662	TP53
chr17	7576847	7576931	TP53
chr17	7577013	7577160	TP53
chr17	7577493	7577613	TP53
chr17	7578171	7578294	TP53
chr17	7578365	7578559	TP53
chr17	7579306	7579595	TP53
chr17	7579694	7579726	TP53
chr17	7579833	7579917	TP53
chr17	40354353	40354470	STAT5B
chr17	40354774	40354831	STAT5B
chr17	40359571	40359751	STAT5B
chr17	40362184	40362365	STAT5B
chr17	40474374	40475165	STAT3
chr19	6822246	6822327	VAV1
chr19	6828836	6828916	VAV1
chr19	6829624	6829802	VAV1
chr19	6832097	6832216	VAV1
chr19	6833190	6833301	VAV1
chr19	6833590	6833641	VAV1
chr19	6833718	6833749	VAV1

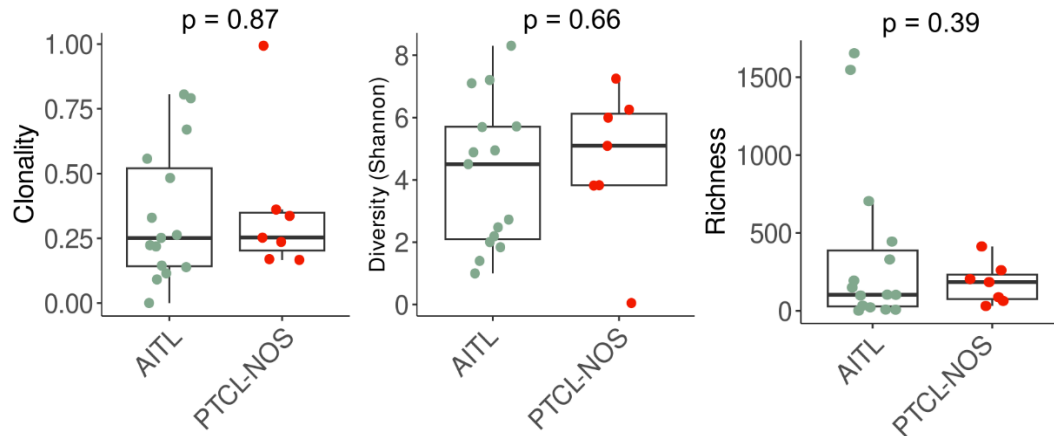
chr19	6833914	6833969	VAV1
chr19	6836438	6836584	VAV1
chr19	6853958	6854030	VAV1
chr19	17942030	17942215	JAK3
chr19	17942480	17942610	JAK3
chr19	17943325	17943520	JAK3
chr19	17943595	17943735	JAK3
chr19	17945375	17945535	JAK3
chr19	17945655	17945815	JAK3
chr19	17945885	17946030	JAK3
chr19	17946730	17946865	JAK3
chr19	17947935	17948025	JAK3
chr19	17948735	17948875	JAK3
chr19	17949065	17949200	JAK3
chr2	25457236	25457248	DNMT3A
chr2	74213525	74213838	TET3
chr2	74230231	74230298	TET3
chr2	74273399	74275543	TET3
chr2	74300670	74300771	TET3
chr2	74307619	74307723	TET3
chr2	74314951	74315170	TET3
chr2	74317018	74317179	TET3
chr2	74320023	74320123	TET3
chr2	74320650	74320798	TET3
chr2	74326397	74326744	TET3
chr2	74327514	74329308	TET3
chr20	39766418	39766430	LOC101927117;PLCG1
chr20	39792578	39792590	PLCG1
chr20	39794133	39794145	PLCG1
chr20	39802379	39802397	PLCG1
chr3	47058577	47058749	SETD2
chr3	47059122	47059234	SETD2
chr3	47061244	47061335	SETD2
chr3	47079150	47079272	SETD2
chr3	47084045	47084195	SETD2
chr3	47087971	47088116	SETD2
chr3	47098305	47098985	SETD2
chr3	47103647	47103841	SETD2
chr3	47108554	47108613	SETD2
chr3	47122451	47122578	SETD2
chr3	47125204	47125877	SETD2
chr3	47127679	47127809	SETD2
chr3	47129597	47129742	SETD2
chr3	47139439	47139576	SETD2
chr3	47142942	47143050	SETD2
chr3	47144830	47144918	SETD2

chr3	47147481	47147615	SETD2
chr3	47155360	47155499	SETD2
chr3	47158107	47158249	SETD2
chr3	47161666	47166043	SETD2
chr3	47168132	47168158	SETD2
chr3	47205338	47205419	SETD2
chr3	49412967	49412982	RHOA
chr3	176743280	176743317	TBL1XR1
chr3	176744155	176744267	TBL1XR1
chr3	176750753	176750929	TBL1XR1
chr3	176751980	176752118	TBL1XR1
chr3	176755880	176755965	TBL1XR1
chr3	176756095	176756227	TBL1XR1
chr3	176763911	176763982	TBL1XR1
chr3	176765082	176765190	TBL1XR1
chr3	176765268	176765342	TBL1XR1
chr3	176767779	176767931	TBL1XR1
chr3	176768260	176768403	TBL1XR1
chr3	176769286	176769519	TBL1XR1
chr3	176771555	176771711	TBL1XR1
chr3	176782702	176782770	TBL1XR1
chr3	176816248	176816334	TBL1XR1
chr3	176849145	176849167	TBL1XR1
chr4	106155094	106158602	TET2
chr4	106162490	106162591	TET2
chr4	106163985	106164089	TET2
chr4	106164721	106164940	TET2
chr4	106180770	106180931	TET2
chr4	106182910	106183010	TET2
chr4	106190761	106190909	TET2
chr4	106193715	106194080	TET2
chr4	106196199	106197681	TET2
chr6	138192359	138192664	TNFAIP3
chr6	138195976	138196177	TNFAIP3
chr6	138196819	138196977	TNFAIP3
chr6	138197127	138197308	TNFAIP3
chr6	138198207	138198398	TNFAIP3
chr6	138199563	138200493	TNFAIP3
chr6	138201202	138201394	TNFAIP3
chr6	138202166	138202461	TNFAIP3
chr9	21968222	21968246	C9orf53;CDKN2A
chr9	21968718	21968775	CDKN2A
chr9	21970895	21971212	CDKN2A
chr9	21974470	21974831	CDKN2A
chr9	21992445	21992484	CDKN2A
chr9	21994132	21994458	CDKN2A

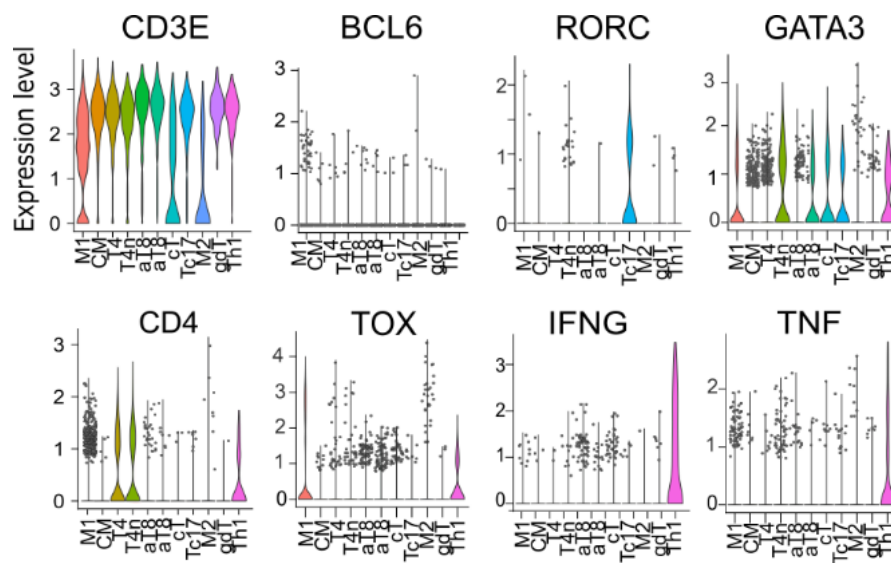
chrX	41193500	41193555	DDX3X
chrX	41193702	41194027	DDX3X
chrX	41196655	41196723	DDX3X
chrX	41198283	41198341	DDX3X
chrX	41200731	41200874	DDX3X
chrX	41201742	41201911	DDX3X
chrX	41201984	41202094	DDX3X
chrX	41202463	41202609	DDX3X
chrX	41202984	41203080	DDX3X
chrX	41203277	41203386	DDX3X
chrX	41203486	41203657	DDX3X
chrX	41204427	41204582	DDX3X
chrX	41204651	41204806	DDX3X
chrX	41205476	41205668	DDX3X
chrX	41205752	41205880	DDX3X
chrX	41206106	41206270	DDX3X
chrX	41206559	41206709	DDX3X
chrX	41206887	41206977	DDX3X



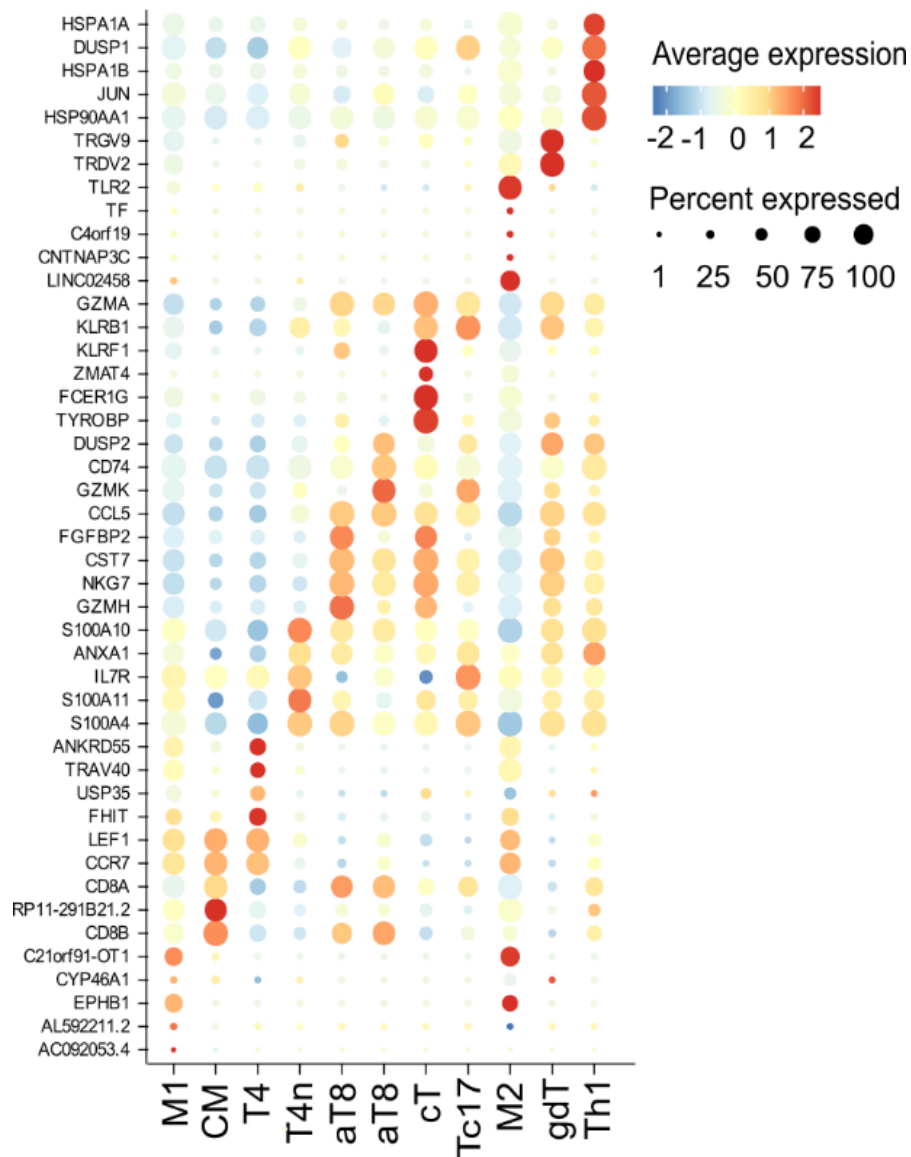
Supplementary Figure 1: TRBV gene usage and frequencies in all lymphoma samples (red, n=19) and healthy controls (blue, n=121). TRBV genes on the axis are sorted for their median group difference.



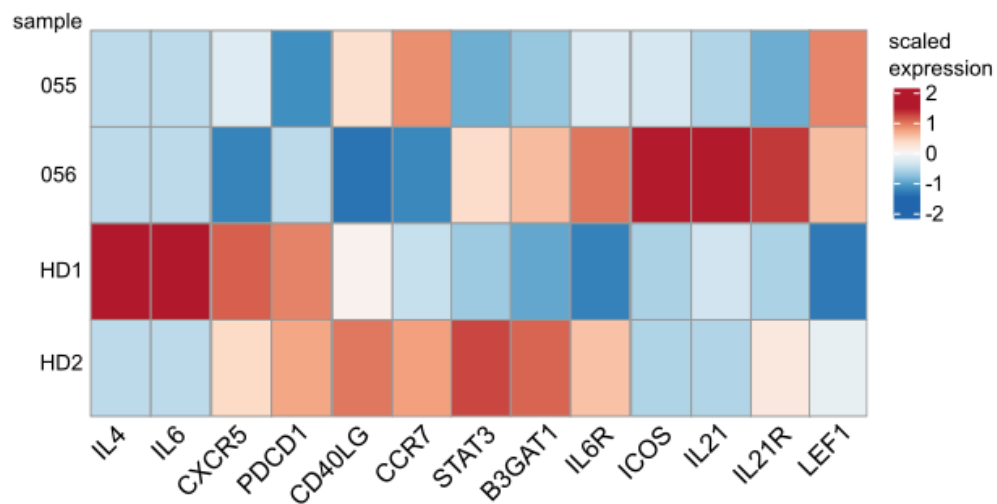
Supplementary Figure 2: TCR metrics in AITL (n=15) and PTCL NOS (n=7) lymphoma samples. Available samples at all time points are shown. Statistics: ANOVA



Supplementary Figure 3: Expression of canonical markers of T cell differentiation and function per cluster. Expression profiles in the integrated single cell dataset of T cells from two healthy and two lymphoma samples.



Supplementary Figure 4: Top five differentiating marker genes for each cluster from the integrated single cell analysis of T cells from two healthy and two lymphoma samples.



Supplementary Figure 5: Expression of markers associated with T follicular helper differentiation in single cell data, separated by samples.