Science Advances
AAAS

Supplementary Materials for

Somatic mutations associated with clonal expansion of CD8$^+$ T cells

Sofie Lundgren *et al.*

Corresponding author: Satu Mustjoki, satu.mustjoki@helsinki.fi; Sofie Lundgren, sofie.lundgren@helsinki.fi

Sci. Adv. **10**, eadj0787 (2024)
DOI: 10.1126/sciadv.adj0787

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:

Supplementary Methods

Quality control of samples sequenced with immunogene panel

Sequencing depth for targeted genomic region was calculated with Samtools (*64*). For each sample, average sequencing depth per gene was calculated by averaging the depth across the targeted regions of the gene. Samples with low sequencing coverage in immunogene panel sequencing (less than 80% of target area had minimum coverage of 50x) and their matched counterpart samples were excluded from variant analysis. Lineage-specific variants in samples with low purity of CD4+ or CD8+ T cells in flow cytometry analysis (≤60% of live cells) were also excluded from analysis. Purity analysis of sorted CD4+ and CD8+ T cell samples are shown in **Supplementary Fig. 13A-B**.

Short variant analysis

Pre-processing of short read data was done with Trimmomatic software (*53*). Adapter sequences, bases with low quality and less than 36bp log reads were removed from sequence data before alignment to human reference genome build 38 (EnsEMBL v82) with BWA-MEM (*65*) with default parameters. PCR duplicates were removed using the SortSAM and MarkDuplicate modules of Picard Toolkit (*66*). Raw sequencing data from the previously published cohorts (*20,30*) had been processed with the same pipeline, and hence commensurable with that released in this study.

Single nucleotide and indel variant analysis was performed with Genome Analysis Toolkit (*67*) (GATK), as previously described (*30*). Variant calling from previously published cohorts was newly performed for this study, to allow integrated analysis. Analysis involved variant discovery on individual samples and a subsequent genotyping of all variants across all samples. MuTect2 was used for variant calling. Variant discovery was done in tumor-only mode without a paired normal as well as by pairing samples with their matched counterpart (i.e. CD4+ datasets paired with CD8+ datasets and vice versa). For the genotyping analysis, variant calls passing filters and supported by ten or more reads were aggregated across samples and supplemented by a set known STAT3 lost-of-function variants and hotspot variants (detected ≥30 samples and constituting ≥1% of the listed mutations of the each gene in COSMIC (*68*)) in genes recurrently

mutated in T cell neoplasms, as previously described (*30*). These variants were then genotyped across samples in tumor-only mode using GATK-4.1.3.0 Mutect2 program.

Following variant calling, several filtering procedures were applied to remove false positive and germline variants from the analysis. Recurrent variant calling artifacts were filtered using a panel of normals (PON) created from healthy T cell (n = 42), skin (n = 1) and fibroblast (n = 1) samples analyzed with immunogene panel. Variants detected in three or more PON samples were excluded from the analysis, with exception of previously reported recurrent mutations in lymphoid or myeloid malignancies (reported in ≥5 hematopoietic or lymphoid tissue samples in COSMIC database (*68*)). Remaining variant calls obtained from genotyping analyses were filtered for vector contamination, RNA or pseudogene artefacts to distinguish variants with a low variant allele frequency from technical or biological artefacts as described previously (*69*). Variant annotation and filtering was performed with Annovar tool against the RefGene database. At first, variant calls were normalized using bcftools (*70*). Variants other than those passing all MuTect2 filters and located in intronic and intergenic regions were then filtered as well as variants with a total depth ≤30, a strand-specific variant read depth <1 in both directions (hence 2 variant reads were required at minimum), <1 variant read in the F1R2 configuration, <1 variant read in the F2R1 configuration, quality value ≤40, variant allele frequency ≤2% or ≥30%, strand odd ratio for SNVs ≥3.00, strand odd ratio for indels ≥11.00, minor allele frequency ≥1% in the 1KG database and EPS database, minor allele frequency ≥0.1% in general, American, African, Finnish, East Asian, and Non-Finnish European ExAC, gnomAD v2 exome, or gnomAD v2 genome databases, and minor allele frequency ≥0.1% in general exome, Finnish exome, general genome, Finnish genome gnomAD v2 and v3 databases. Variants with a variant allele frequency 1-2% were accepted, if supported by five or more COSMIC samples. Finally, all variants were manually curated by visual inspection using Integrative Genomics Viewer 2.4.10 (IGV, Broad Institute). During this step, indel variants falling within simple repeat regions were removed.

CNV identification from immunogene panel sequencing data

CNVs were called from sequencing data by reprocessing small variant analysis alignment files with DRAGEN Bio-IT Platform version 4.2 (*57*) (Illumina) with Illumina Multigenome Graph Reference hg38 for DRAGEN v4.2 (Illumina). Background reference set was constructed from

healthy donor T cell (n=42), skin (n=1) and fibroblast (n=1) samples sequenced with same gene panel. Variant calls were annotated with annotSV (*58*). The DRAGEN Bio-IT Platform command and annotSV commands used in CNV analysis are provided in the **Supplementary Table 10**. Low quality CNVs and CNVs with population allele frequency of ≥1% in any of the reported reference databases were removed. Only CNVs overlapping with ≥2 exons and length of ≥10 000 bp were included in the analysis. Recurrent CNV calling artifacts were further filtered by omitting those CNVs with ≥10% overlap in 3 or more PON samples. Next, ancestral CNVs called in both lineages of a matched sample pair (≥80% overlap) and samples with an average CNV correlation (**Supplementary Table 11**) less than 0.75 against reference set samples were discarded. CNV correlation values were extracted from DRAGEN outputs. Finally, all remaining CNVs were manually curated by visual inspection in IGV.

Filtering of healthy T donor short variants

In the analysis of T cell somatic mutations from hematological patients, we used the healthy samples in the Panel of Normals (PON, see above). For this reason, these results were biased towards alterations that are absent in healthy donor samples. To enable comparison of healthy donor T cell mutations to those of hematological patients, we re-analyzed both patient and healthy variant data with a PON formed from all samples (biasing towards non-recurrent mutations but doing that similarly to healthy controls and patients). We filtered out all somatic variants called in ≥10 individuals across all samples before subsequent filtering with identical criteria to original somatic variant analysis (see above). In total, 856/923 (92.7%) lineage-specific and 32/34 (94.1%) shared variants from the original analysis were identified with the alternative PON, with 45 additional lineage-specific mutations identified in the patient samples (no additional ancestral mutations were identified). The healthy T cell variants from this analysis are listed in **Supplementary Table 5**. The patient variants from alternative PON filtering were only used in comparisons to healthy control samples.

Single-cell sequencing

9 cGVHD patient samples and 6 healthy controls were subjected to the analysis. Frozen MNC from peripheral blood were sorted with BD Influx Cell sorter and the gene and V(D)J transcript profiles were studied with 10x Genomics Chromium Single Cell Immune Profiling platform. The

Chromium Single Cell 5'RNAseq run and library preparation were done using the Chromium Next GEM Single Cell Immune Profiling version 1.1 chemistry.

When thawing, the cryo-preserved samples were resuspended to 13ml of +37°C plain RPMI (Roswell Park Memorial Institute) and after centrifugation of 5 minutes at 300g, washed with 10ml of PBS + 2mM EDTA buffer. After 2nd 5 minute centrifugation at 300g, cells were resuspended to PBS + 0.05% BSA in concentration of 10,000 cells/ul. 2ul of CD45-APC-H7 (2D1, BD, cat. no 641417) antibody per 1 million cells was added and samples were incubated with the antibody for 15 minutes in room temperature. Cells were then washed with 2ml of PBS-BSA buffer (5 minute centrifugation at 300g) and resuspended to pre-chilled RPMI at concentration of 5 million cells/ml. From this point, samples were kept on ice (before and after sorting). Sorting of CD45+ MNCs was performed with BD Influx Cell sorter, sorting strategy as presented in **Supplementary Fig. 13C**. 300,000 target cells were collected to Protein Lo-Bind tubes (Eppendorf, cat. no 0030108).

After sorting, single-cell samples were partitioned using a Chromium Controller (10X Genomics) and scRNA-seq and TCRαβ-libraries were prepared using Chromium Single Cell 5' Library & Gel Bead Kit (10X Genomics), according to manufacturer's instructions (CG000086 Rev D). 12,000 cells from each sample, suspended in 0.04% BSA in PBS, were loaded on the Chromium Single Cell A Chip. During the run, single-cell barcoded cDNA is generated in nanodroplet partitions. The droplets were subsequently reversed and the remaining steps were performed in bulk. Full length cDNA was amplified using 14 cycles of PCR (Veriti, Applied Biosystems). TCR cDNA was further amplified in a hemi-nested PCR reaction using Chromium Single Cell Human T Cell V(D)J Enrichment Kit (10X Genomics). Finally, the total cDNA and the TCR-enriched cDNA were subjected to fragmentation, end repair and A-tailing, adaptor ligation, and sample index PCR (14 and 9 cycles, respectively). All libraries were sequenced using NovaSeq 6000 system (Illumina) using read lengths: 26bp (Read 1), 8bp (i7 Index), 0 bp (i5 Index) and 91bp (Read 2). Length configurations used for TCR enriched libraries: Read1=150, i7=8, i5=0, Read2=150. The raw data was processed using Cell Ranger v3.1 pipelines. Cell Ranger pipeline "cellranger mkfastq" was used to produce FASTQ (raw sequence data) files, "cellranger count" to perform alignment, filtering and UMI counting for the 5' gene

expression data and "cellranger vdj" to perform V(D)J sequence assembly and paired clonotype calling for the V(D)J data. “Cellranger mkfastq” was run using the Illumina bcl2fastq v2.2.0 and alignment was done against human genome GRCh38.

scRNA+TCRαβ-seq data analysis

Quality control was done before subsequent analysis. Cells with high amount of mitochondrial or ribosomal transcripts (>15% or >50% of all UMI counts, respectively), cells with less than 500 or more than 4000 genes expressed, and cells with low (<1000) or high (>15000) UMI counts were removed.

To remove batch-effect and analyse all samples integratively, we used deep generative modeling tool scVI as described previously (*62*). The latent embeddings created by scVI were used for graph-based clustering and UMAP-dimensionality reduction with default parameters in RunUMAP function in Seurat package (v4) (*60*) . With iterative clustering and subsetting of cells, we removed clusters consisting of doublets or low-quality cells based on UMI and gene counts and divided cells into main immune cell classes: monocytes, T cells (and subsequently CD4+ and CD8+ T cells), NK cells, B cells, dendritic cells and progenitor cells. Differential expression analyses were performed based on t-test, where Bonferroni adjusted p-values below 0.05 were denoted as significant. Clusters were annotated using differentially expressed genes, comparison to bulk-RNA-seq profiles of sorted immune subsets and canonical markers (**Supplementary Fig. 8C** for CD8+ **and 9C** for CD4+ T cells) and cytolytic score. List of differentially expressed genes for each main cell class as well as subpopulations (comparing to other cells in same main cell class) are provided in **Supplementary Tables 6 and 7** (for CD8+ and for CD4+ T cells, respectively). The V(D)J sequences of each cell were integrated into the Seurat object as metadata for gene expression and clonotype analysis. Clonotypes were identified based on the total nucleotide level TCRα and TCRβ. Cytolytic score **(Supplementary Fig. 8D and S9D**) was calculated with the AddModuleScore function, including genes defined by Dufva and Pölönen *et al* (*71*): GZMA, GZMH, GZMM, PRF1, and GNLY.

Mutation identification from scRNA+TCRαβ-seq data

All mutations that had been identified with immunogene panel sequencing from the 9 cGVHD patients (included in the scRNA+TCRαβ-seq analysis) were genotyped from single-cell transcriptome data by Vartrix (*42*) with default parameters. We successfully identified three somatic SNVs in three genes: *TYW1* in GVHD-11, *PTPRE* in GVHD-14 and *TNFRSF1B* in GVHD-21. For these three variants, we identified the reference (REF) genotype from significant number of cells from all 9 patients (in total 2385, 15 472 and 1046 cells with REF genotype for *TYW1*, *PTPRE*, and *TNFRSF1B*, respectively, **Supplementary Fig. 10A**). For *TYW1* and *TNFRSF1B* variants, all cells with variant (ALT) genotype (9 and 7 cells for *TYW1* and *TNFRSF1B*, respectively) were T cells from the same patient from whom the same variant had been identified with immunogene panel sequencing. For *PTPRE* variant, one cell with ALT genotype was identified in GVHD-7, possibly due to index hopping or sequencing error. All other cells with ALT or ALT/REF genotype (88 cells) were from GVHD-14, from whom we identified the same mutation with immunogene panel sequencing. The extremely low number of false positive variants in scRNA+TCRαβ-seq data suggested a high accuracy of identified genotypes. After identification of mutated cells, we imputed the mutated clonotype based on TCRβ as in Huuhtanen & Bhattacharya *et al* (*72*): 5/5, 44/45 and 2/2 mutated T cells with mapped TCRβ had exactly the same TCRβ sequence within GVHD-11 (*TYW1* mutation), GVHD-14 (*PTPRE*) and GVHD-21 (*TNFRSF1B*), respectively. Odds ratios were calculated for mutated clonotypes against all other clonotypes within each patient by using "oddsratio" function in R.

CNV identification from scRNA+TCRαβ-seq data

To identify CNVs in scRNA+TCRαβ-seq data, we executed Numbat (v1.3.0) (*61*) with its default settings. We used the cluster annotations from Seurat analysis (see above) to segregate the CD4+ and CD8+ T cells, running Numbat separately for each sample and the two T cell subtypes.

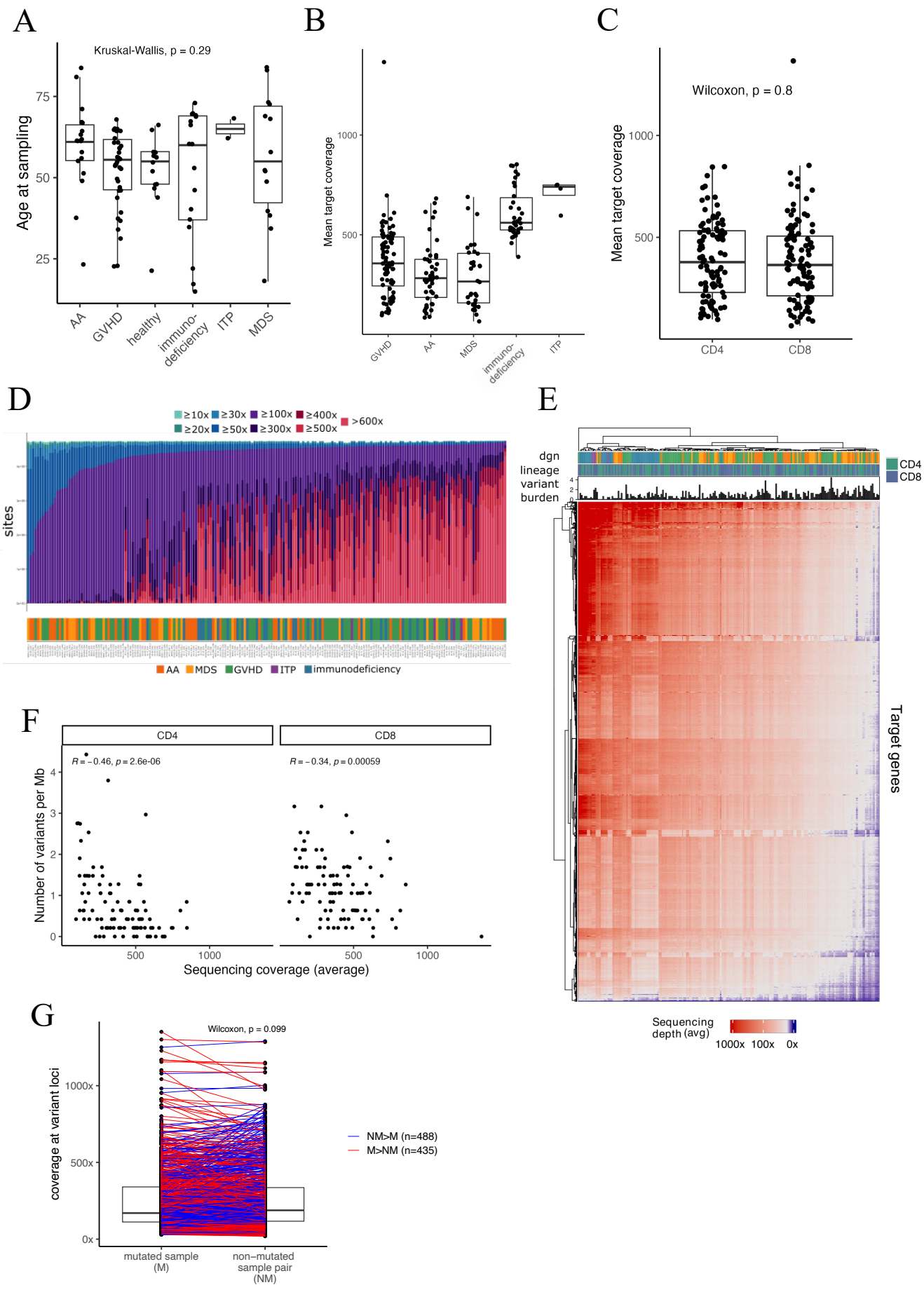
Differential expression and cell-cell interaction analysis of mutated clonotypes

Differential expression between mutated and non-mutated clonotypes was performed with Seurat package (*60*), by using t-test for comparisons and adjusting for multiple comparisons with

Bonferroni method. Only cells from each patient and each dominating phenotypic cluster (for instance, 0 CD8 Temra cluster) were included in the comparisons.

Cell-cell communication analysis were conducted using CellChat (v1.6.1) (*63*). As input for CellChat, we used normalized gene expression data (LogNormalize method) and a cell grouping based on cluster annotation from Seurat analysis. We modified the cell grouping by dividing the CD8+ and CD4+ Temra clusters (only the dominating phenotypic cluster for each patient) with detected mutations into two subsets based on the presence of the mutation.

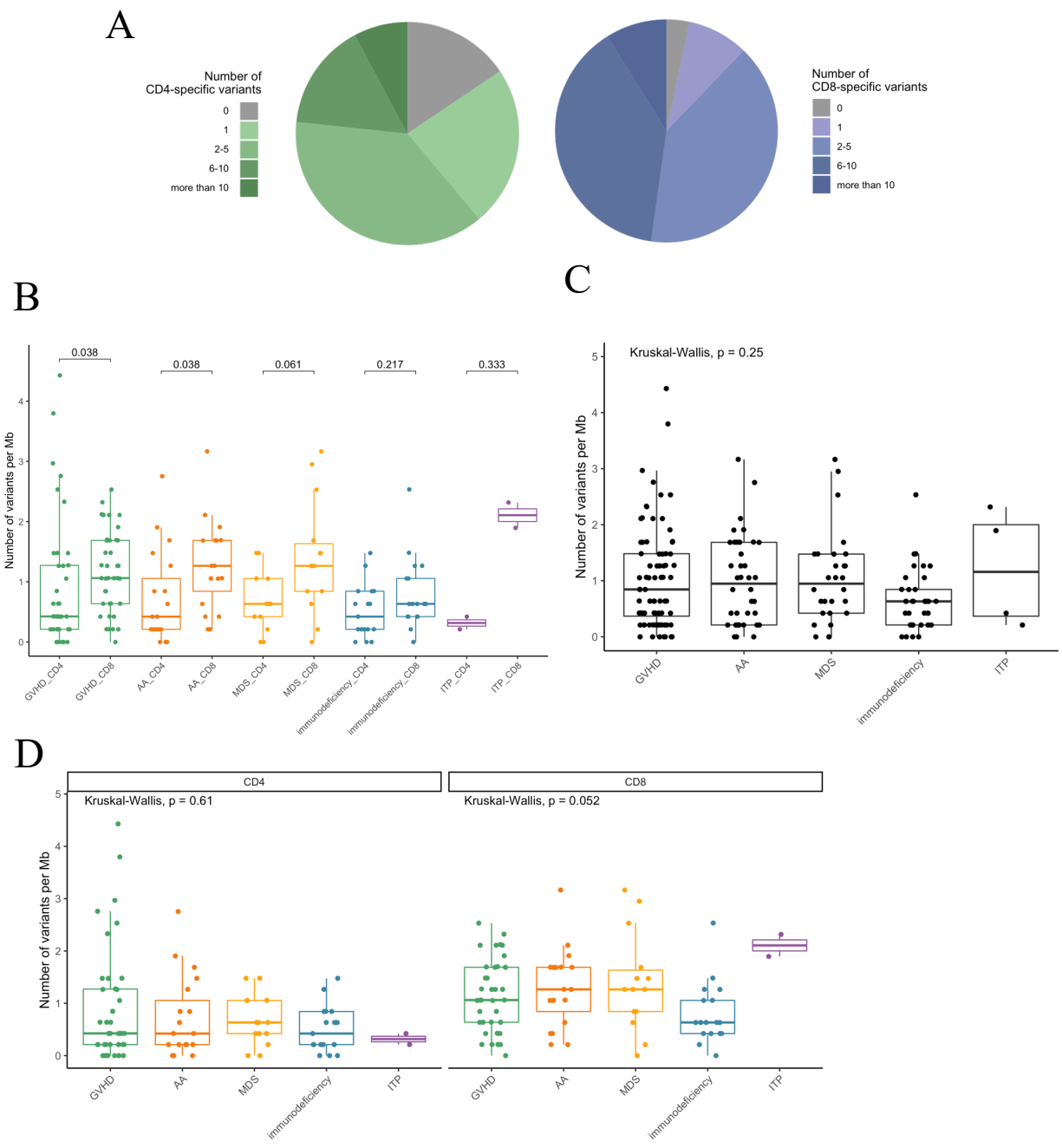
Supplementary Figure 1



Supplementary Fig. 1. Age distribution and sequencing coverage in immunogene panel analysis.

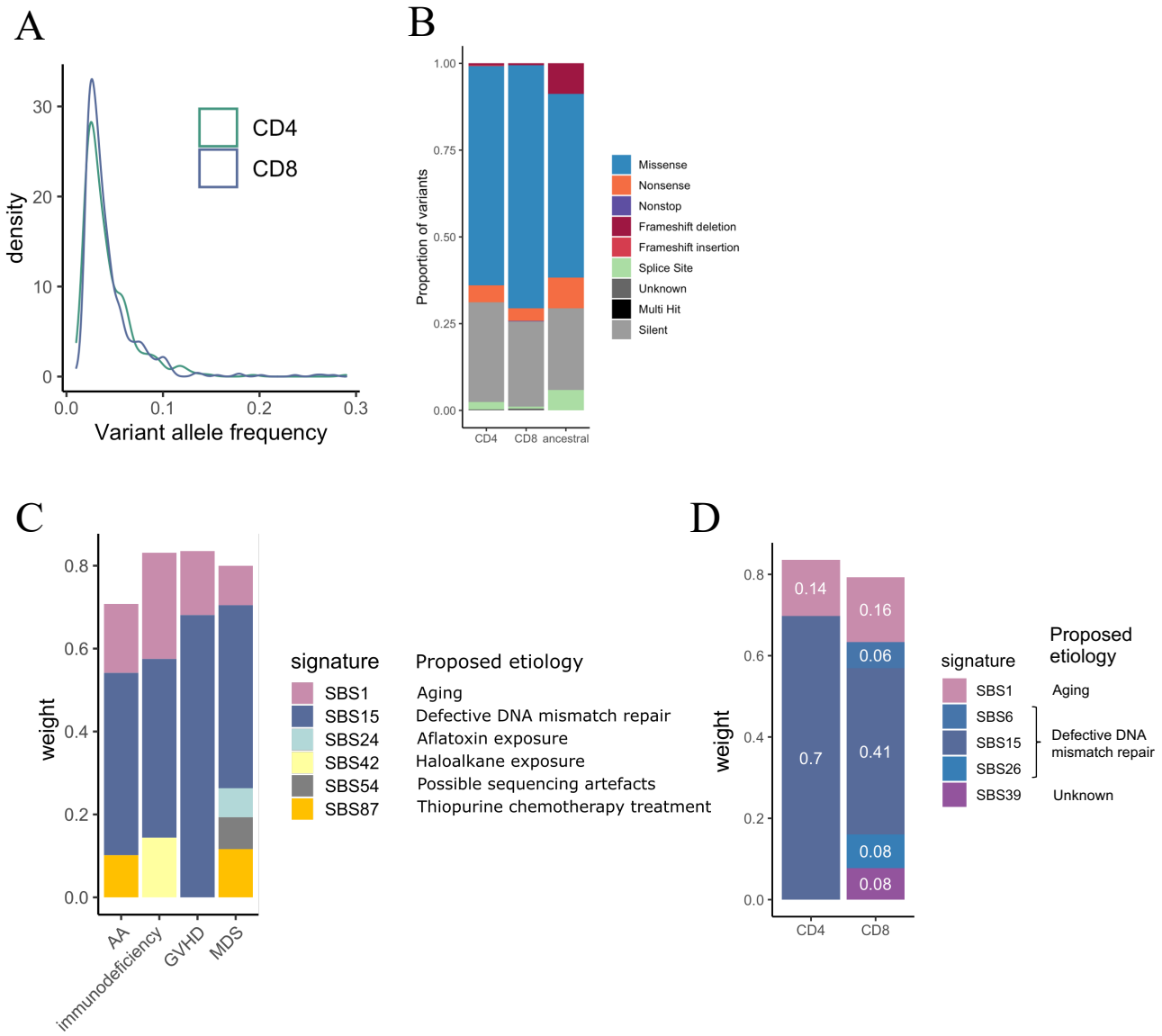
(**A**) Age at sampling in all analyzed cohorts. P values have been calculated with Kruskal-Wallis test. (**B**) Average sequencing coverage in immunogene panel sequencing in different patient cohorts. (**C**) Average sequencing coverage in CD4+ and CD8+ samples of hematological patients. P value has been calculated with Mann-Whitney test. (**D**) Target coverage in all patient samples, presented as sites covered with depth of 10-19x, 20-29x, 30-49x, 50-99x, 100-299x, 300-399x, 400-499x, 500-599x, and >600x. Samples are ordered based on the area covered at minimum depth of 100x. Samples with less than 80% of target area covered with a depth of less than 50x were excluded from the analysis. (**E**) Average sequencing depth across targeted genes (y axis) in CD4 and CD8 samples of hematological patients (x axis). Patient group and sample type (CD4 or CD8) are shown in the top panel. Rows and columns are clustered using Euclidean distance. (**F**) Correlation of sequencing coverage and mutation burden in CD4+ and CD8+ samples. P values and correlation coefficients were calculated with Pearson correlation test. (**G**) Sequencing coverage in patients' T cell samples at mutated loci (M) and each paired CD4 or CD8 sample with no mutation at the same loci (NM). P value has been calculated with Mann-Whitney test.

Supplementary Figure 2



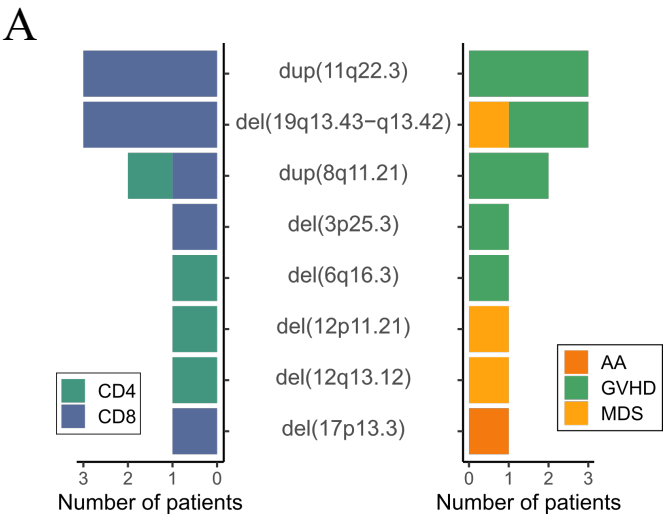
Supplementary Fig. 2. Lineage-specific mutations in CD4+ and CD8+ T cells. (**A**) Distribution of CD4+ and CD8+ specific mutations. We identified lineage-specific mutations in 84% (76/90) and 97% (87/90) of patients, respectively. (**B**) Lineage-specific mutation burden in CD8+ T cells and CD4+ T cells in each cohort separately. P values have been calculated for comparisons within one cohort (Mann-Whitney test) and adjusted for multiple comparisons with Benjamini-Hochberg method. (**C**) Lineage-specific mutation in different patient cohorts (both CD4+ and CD8+ mutations). P value has been calculated with Kruskal-Wallis test. (**D**) CD4-specific or CD8-specific mutation burden was compared separately between patient cohorts with Kruskal-Wallis test.

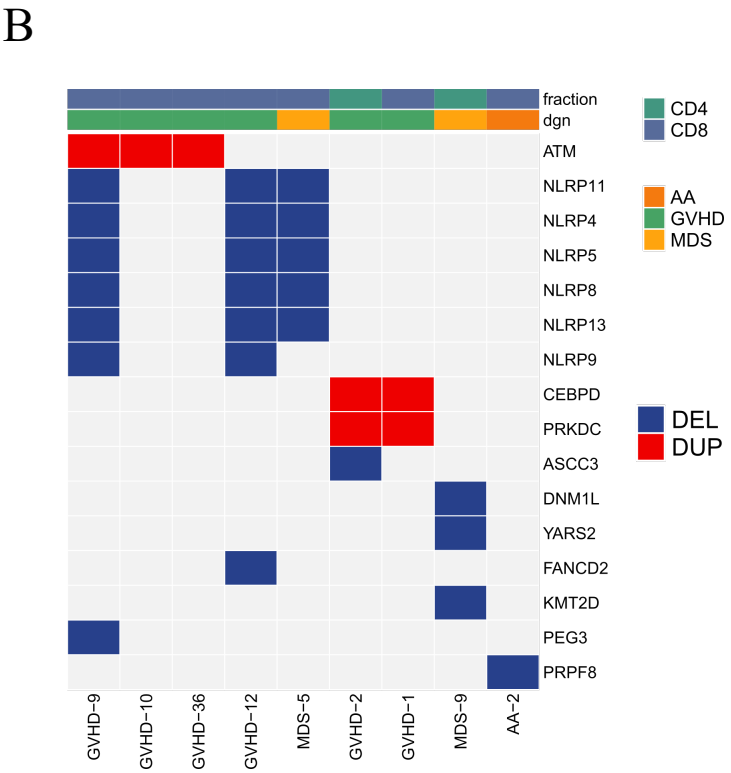
Supplementary Figure 3

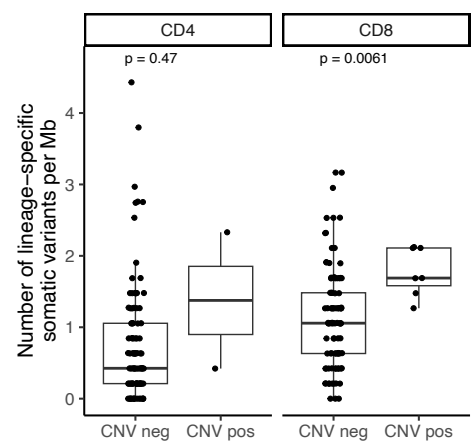


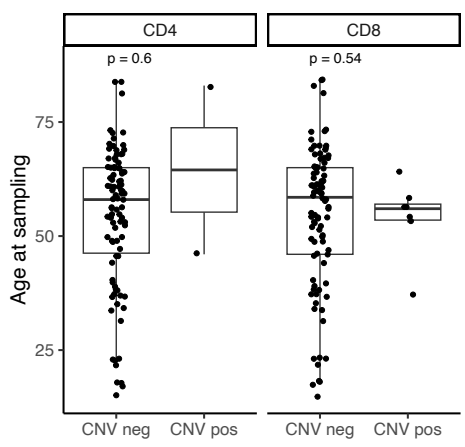
Supplementary Fig. 3. Variant types and mutational signatures in T cells of hematological patients. (**A**) Density plot of variant allele frequencies (VAFs) in CD4+ and CD8+ T cells. There was no significant difference between two lineages ($p = 0.48$ with Mann-Whitney test). (**B**) Variant types of lineage-specific and ancestral (detected in both CD4+ and CD8+ compartments in same patient) mutations. (**C**) Mutational signatures weights for each identified signature in different patient groups. All variants (including lineage-specific and ancestral variants) were included in the analysis. (**D**) Mutational signature weights for each identified signature in CD4+ or CD8+ T cells (only lineage-specific mutations).

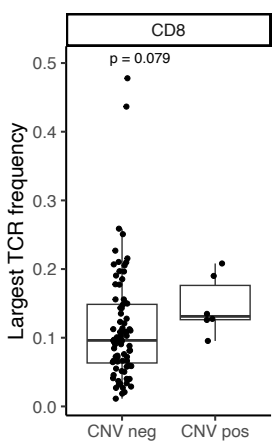
Supplementary Figure 4





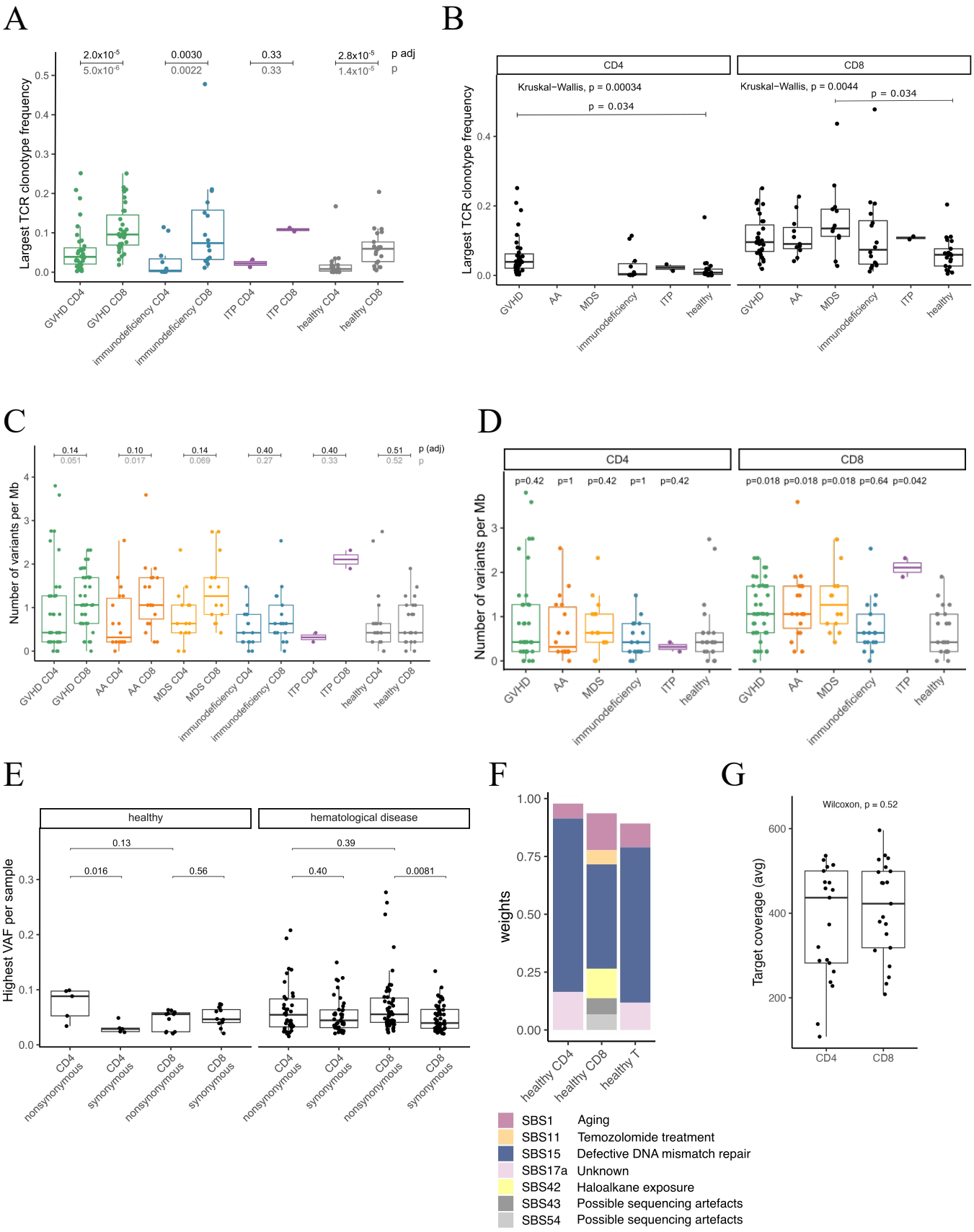






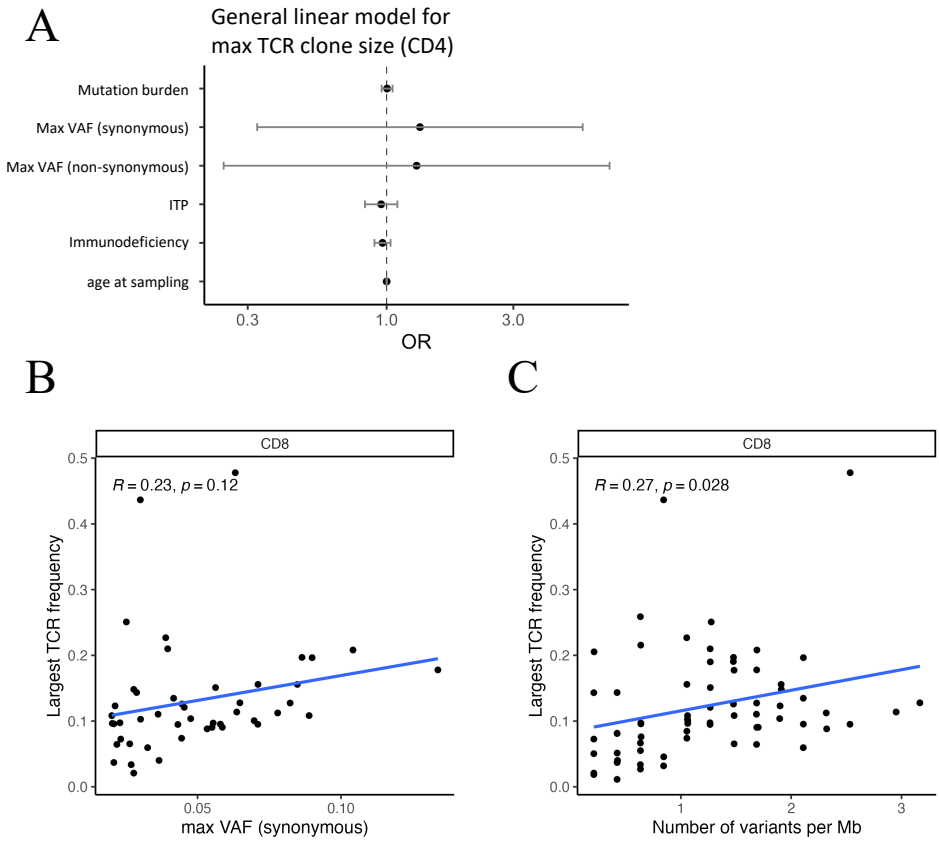
Supplementary Fig. 4. Lineage-specific CNVs in CD4+ and CD8+ T cells. (**A**) Number of patients with different CNVs. Patient groups for each alteration are shown on the right panel and fractions where each alteration was detected are shown in the left panel. (**B**) Immunogene panel target genes affected by lineage-specific CNVs. Duplications (DUP) are marked with red and deletions (DEL) with blue color in the oncoplot. Fraction and patient group (dgn) are shown in the top panel. (**C**) Lineage-specific somatic mutation burden in CD4 and CD8 samples with and without CNVs. P values have been calculated with Mann-Whitney test. (**D**) Age in patients with and without CNVs in CD4 or CD8 samples. P values have been calculated with Mann-Whitney test. (**E**) Largest TCR frequency in CD8 samples with and without CNVs. TCR data was not available for CD4 samples with CNVs. P value has been calculated with Mann-Whitney test**.**

Supplementary Figure 5



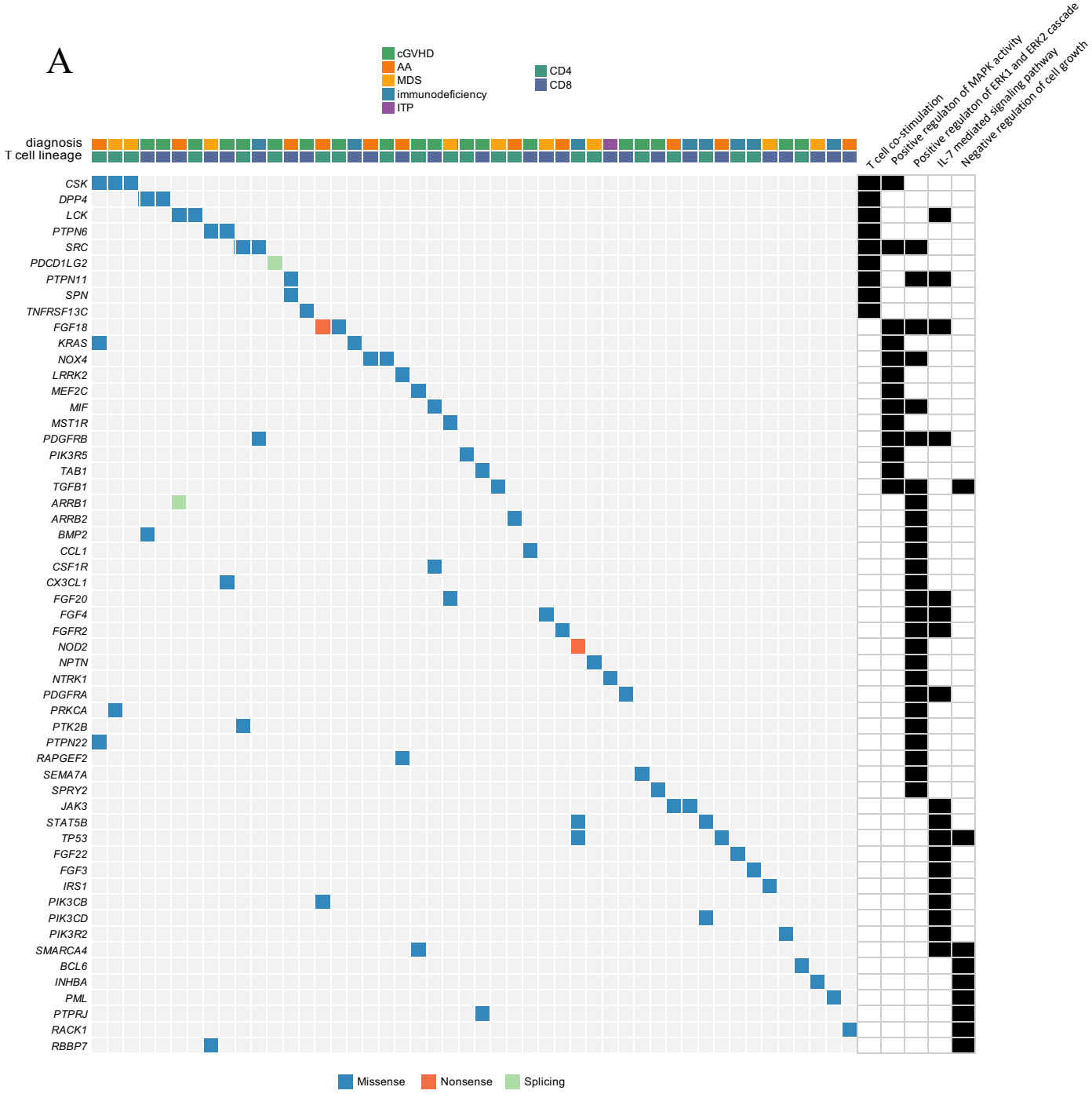
Supplementary Fig. 5. Somatic mutation analysis of healthy T cells and comparisons to hematological patients. (**A**) Largest TCR clonotype frequency in CD4 and CD8 samples in each cohort. P values (comparisons within each cohort) have been calculated with Mann-Whitney test and adjusted for multiple comparisons with Benjamini-Hochberg method. Both adjusted (p adj) and unadjusted (p) are shown. (**B**) Largest TCR clonotype frequency in different cohorts. Because there was significant difference between CD4 and CD8 samples in all patient groups (panel **A**), CD4 and CD8 samples are shown separately. P values (comparisons within each cell type) have been calculated with Kruskal-Wallis test, post-hoc analysis was performed with Dunn's test and p values for pair-wise comparisons were adjusted for multiple comparisons with Benjamini-Hochberg method. Only significant p values are shown for clarity. (**C**) Lineage-specific mutation burden in CD4 and CD8 cells in each cohort. P values have been calculated for comparisons within each cohort with Mann-Whitney test and adjusted for multiple comparisons with Benjamin-Hochberg method. Both adjusted (p adj) and unadjusted p values are shown. (**D**) Lineage-specific mutation burden in CD4 cells (left) and CD8 cells (right) in different patient cohorts, in comparison with healthy T cells. P value for each patient group (comparison to healthy) has been calculated with Mann-Whitney test and adjusted for multiple comparisons with Benjamini-Hochberg method. (**E**) The highest synonymous and non-synonymous VAF per sample in CD4 and CD8 samples in healthy donors (left) and patients (right). P value for all shown comparisons have been calculated with Mann-Whitney test and adjusted for multiple comparisons with Benjamini-Hochberg method.**) (F)** Mutational signature analysis of healthy donor T cell mutations. (**G**) Average target coverage of healthy donor CD4 and CD8 samples. P value has been calculated with Mann-Whitney test.

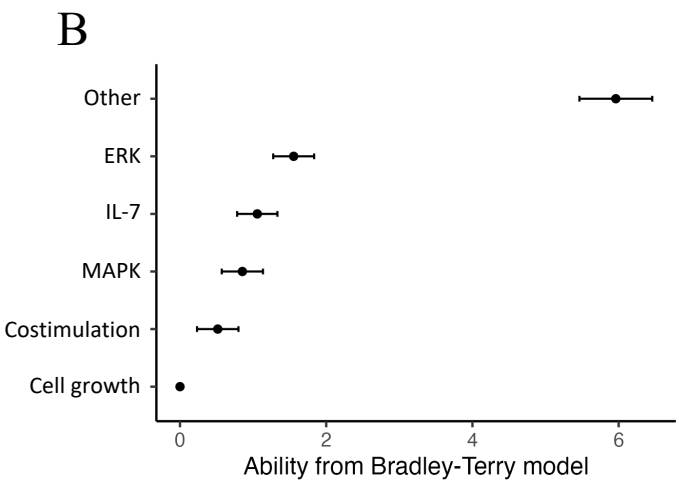
Supplementary Figure 6

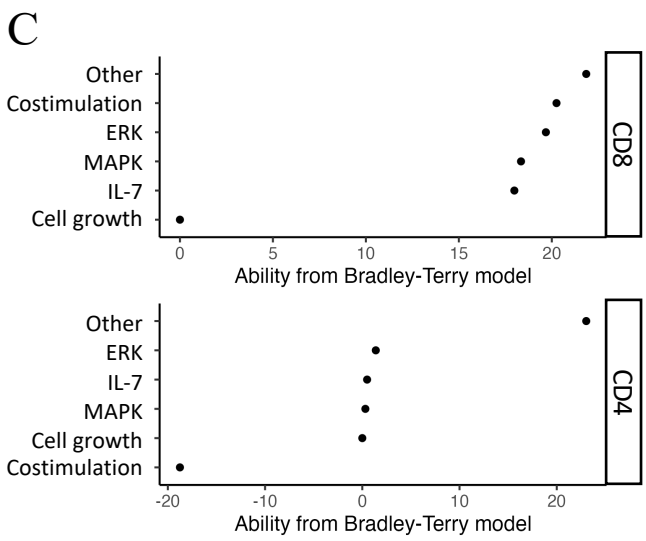


Supplementary Fig. 6. Variant and TCR associations. (**A**) Multivariate general linear model for the largest CD4+ T cell expansion clone size, defined by TCRβ sequencing. Odds ratios (OR) with their associated 95% confidence intervals are shown for each variable included in the model (mutation burden (Number of variants per Mb), the highest non-synonymous and synonymous VAF, diagnosis (GVHD, AA, MDS, ITP, immunodeficiency), and age at sampling). None of the tested variables were significantly associated with size of the largest CD4+ T cell expansion, and hence CD4+ clone size was not included in the next figures. For CD8+ T cells, Pearson correlation analysis of individual factors included in the multivariate model of the biggest TCR clone (**Fig. 2A**) size are shown in (**B**) (the highest nonsynonymous VAF) and (**C**) (mutation burden).
TCR, T cell receptor; VAF, variant allele frequency; OR, odds ratio; Mb, megabase

Supplementary Figure 7

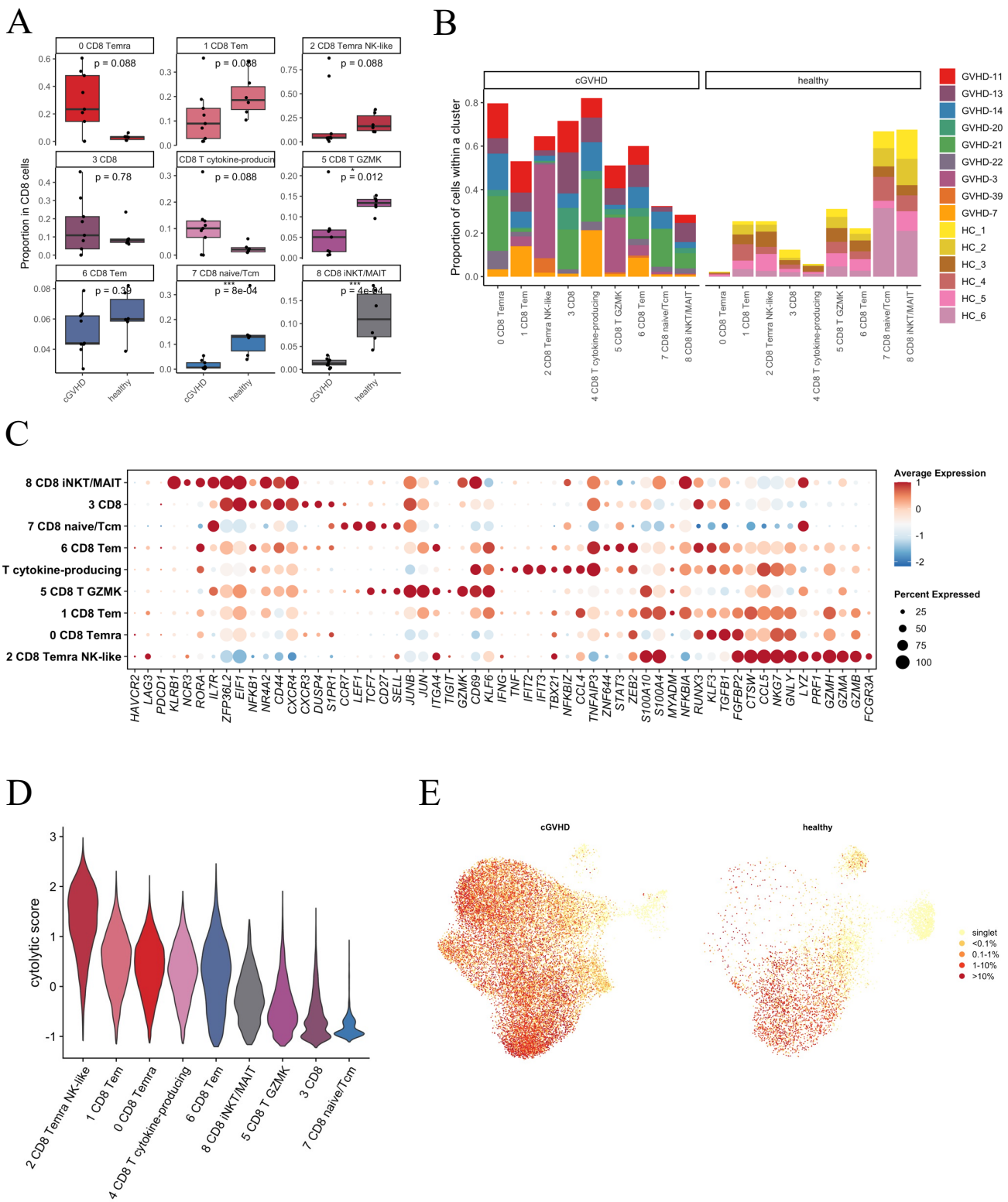






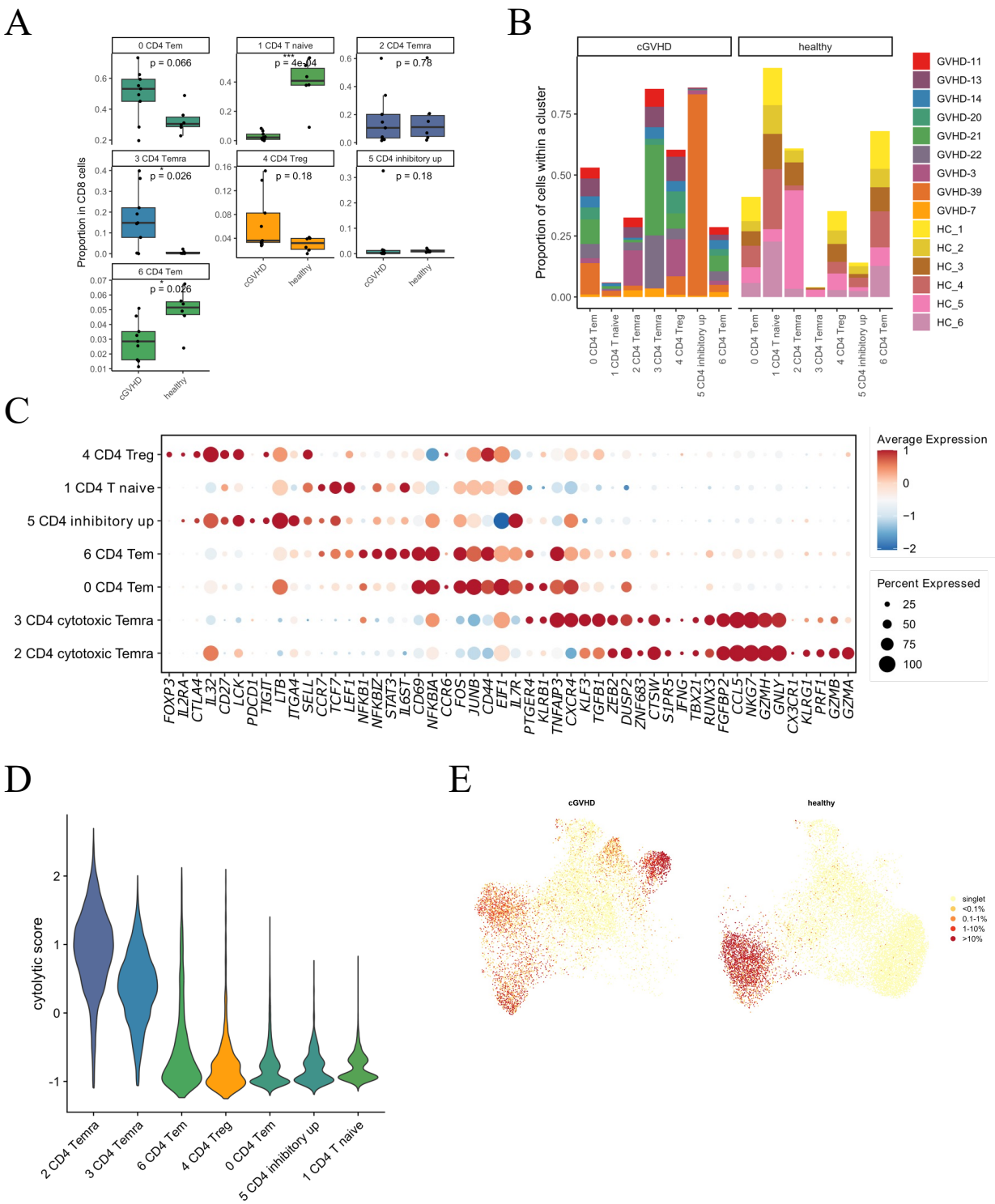
Supplementary Fig. 7. Significantly mutated pathways in CD4+ and CD8+ T cells. (**A**) Co-mutation plot of lineage-specific mutations in the most significantly mutated biologically relevant pathways, based on OncodriveFM analysis. Each column represents one sample and each row represents one gene. Diagnosis and T cell lineage (CD4+ or CD8+) for each sample are shown in the top panel and pathway annotation is shown in the right panel. (**B**) Bradley-Terry model was used to study the relative order of mutation acquisition from pairwise relationships. When a pathway was not mutated in a sample, its VAF was set to 0. Points and error bars represent the Bradley-Terry model results for the point estimate and 95% confidence interval for relative pathway ordering. (**C**) Modified Bradley Terry model analysis (using only co-occurring mutations) was performed for variants in CD4+ and CD8+ samples separately. Results were not statistically significant, due to the low number of samples with co-occurring mutations.

Supplementary Figure 8



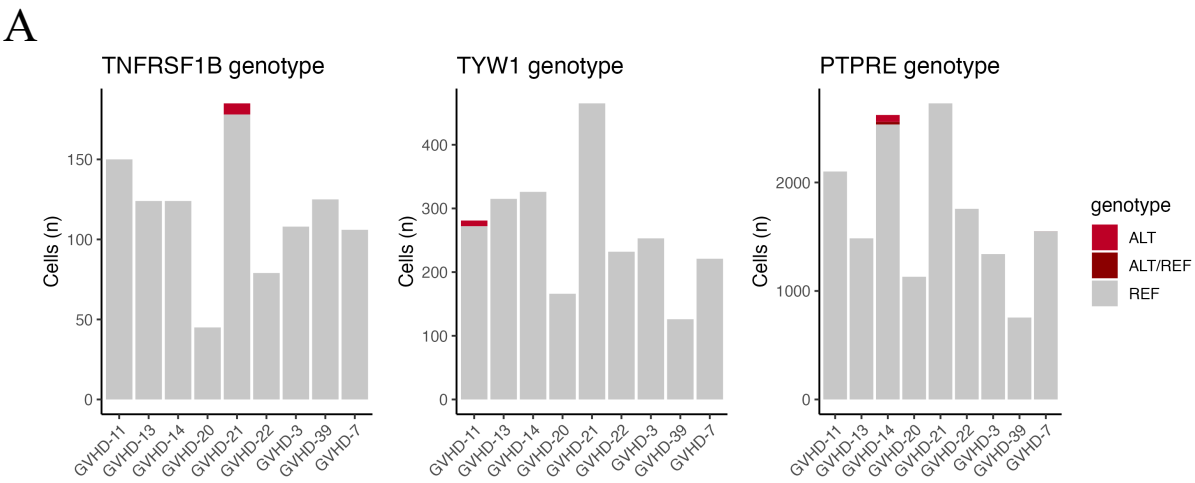
Supplementary Fig. 8. Single-cell transcriptome analysis of 9 patients with cGVHD and 6 age-matched healthy controls (CD8). (**A**) Proportions of each phenotype cluster in cGVHD patients and healthy controls' CD8+ T cells. P values have been calculated with Mann-Whitney test and significantly different proportions between cGVHD and healthy individuals ($p \leq 0.05$) have been marked with an asterisk (*). (**B**) Proportion of cells from each patient within each CD8+ phenotype cluster. (**C**) Differentially expressed genes used in the annotation of the CD8+ T cell phenotype clusters. Colour shows the average expression in each cluster, while the size of the dot shows the percentage of cells expressing each marker. (**D**) Cytolytic score (see Supplementary methods) for CD8+ T cell phenotype clusters. (**E**) UMAP for cGVHD patients' and healthy controls' CD8+ T cells with TCRβ. Colors show the size of the TCR clone (singlet, <0.1%, 0.1-1%, or 1-10% frequency within CD8+ T cells, analyzed separately for each sample).

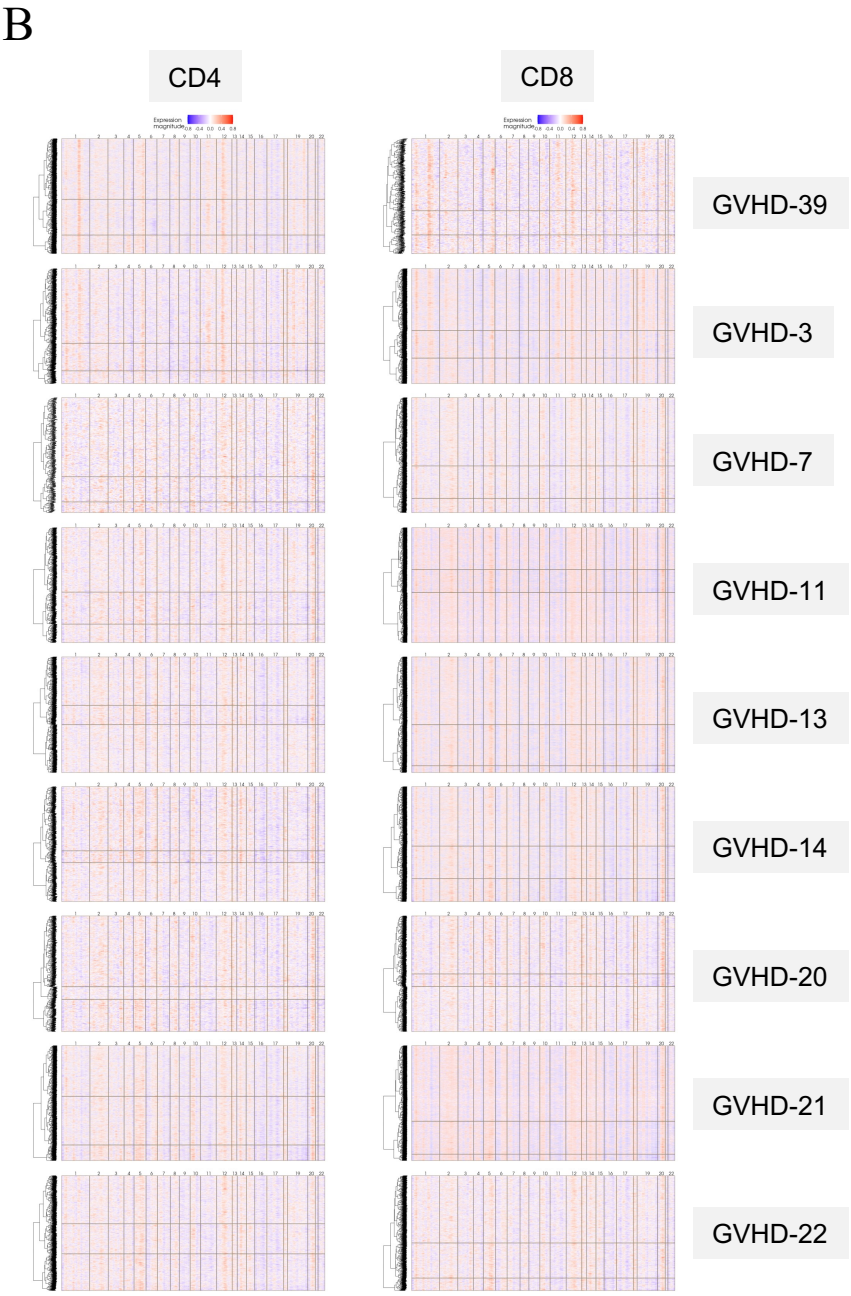
Supplementary Figure 9



Supplementary Fig. 9. Single-cell transcriptome analysis of 9 patients with cGVHD and 6 age-matched healthy controls (CD4). (**A**) Proportions of each phenotype cluster in cGVHD patients and healthy controls' CD4+ T cells. Significantly different proportions between cGVHD and healthy individuals ($p \leq 0.05$ with Mann-Whitney test) have been marked with an asterisk (*). (**B**) Proportion of cells from each patient within each CD4+ phenotype cluster. (**C**) Differentially expressed genes used in the annotation of the CD4+ T cell phenotype clusters. Colour shows the average expression in each cluster, while the size of the dot shows the percentage of cells expressing each marker. (**D**) Cytolytic score (see Supplementary methods) for CD4+ T cell phenotype clusters. (**E**) UMAP for cGVHD patients' and healthy controls' CD4+ T cells with TCRβ. Colors show the size of the TCR clone (singlet, <0.1%, 0.1-1%, or 1-10% frequency within CD4+ T cells, analyzed separately for each sample).

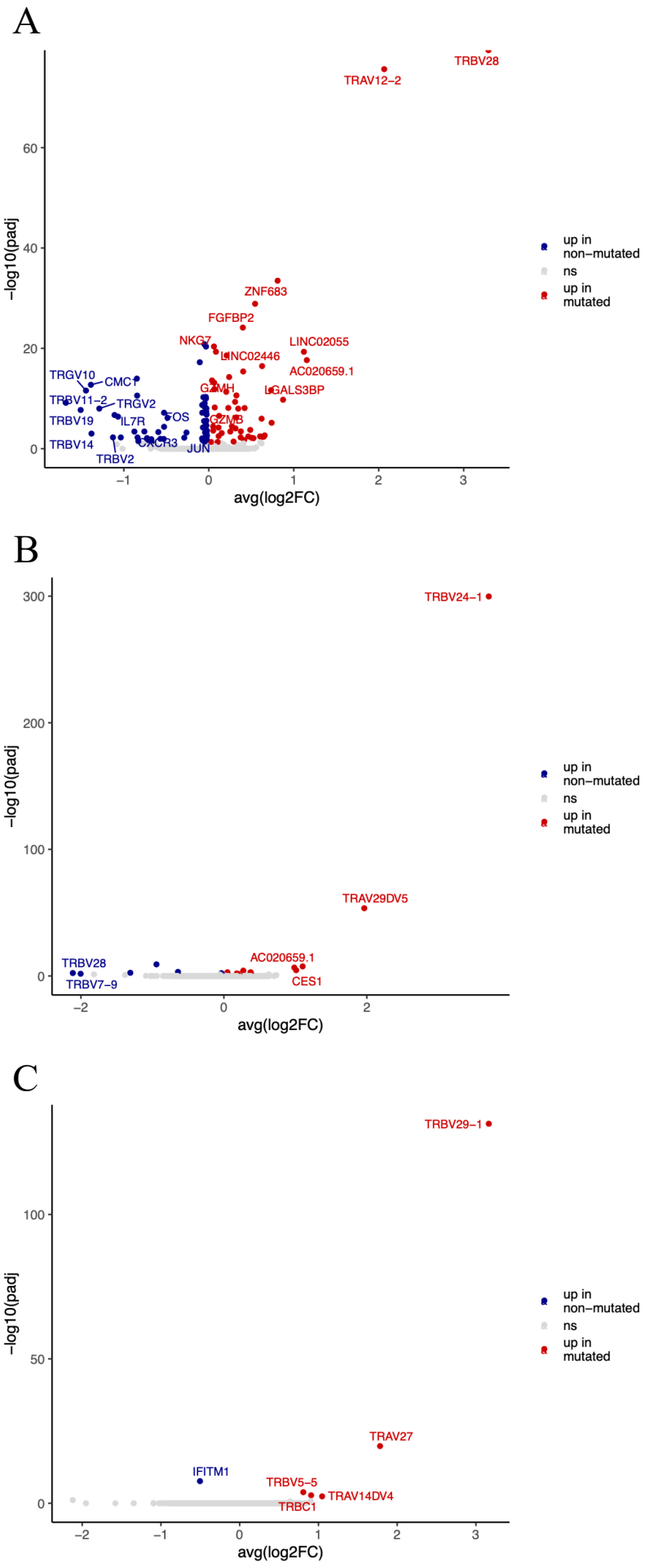
Supplementary Figure 10





Supplementary Fig. 10. Mutation and CNV identification from scRNA+TCRαβ-seq data. **(A)** Mutation identification from scRNA+TCRαβ-seq data with Vatrix from 9 cGVHD patient samples. Red denotes cells with variant genotype (ALT or ALT/REF), while grey denotes cells with reference genotype (REF). Results are shown from three immunogene panel variants that were successfully genotyped from scRNA+TCRαβ-seq data. (**B**) CNV analysis with Numbat *(61)* from CD4 and CD8 cells of 9 cGVHD patients. Each panel shows the window-smoothed normalized expression profiles of CD4 (left) and CD8 (right) cells. The dendrogram shows the initial hierarchical clustering result of single cells based on smoothed expression.

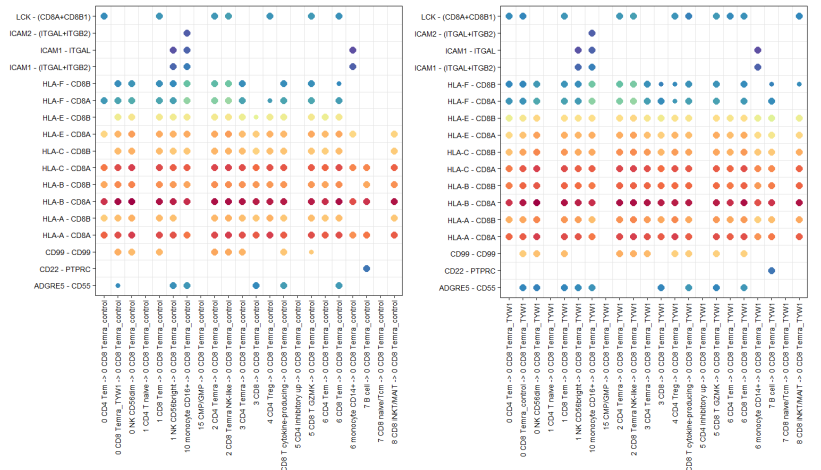
Supplementary Figure 11



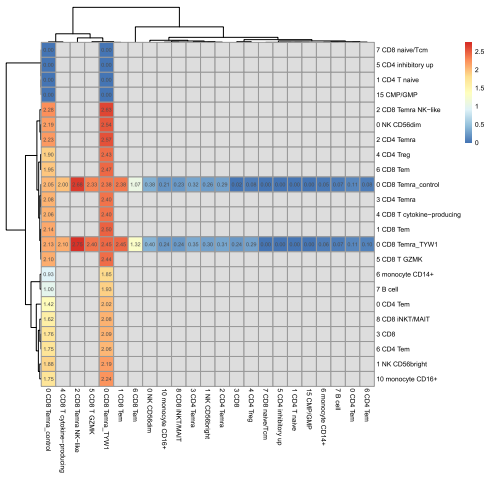
Supplementary Fig. 11. Differential gene expression analysis of mutated vs non-mutated cells of 3 patients with cGVHD. Analysis was performed within cells belonging to "0 CD8 Temra" cluster for clonotypes with *TYW1* and *PTPRE* mutations (panels **A** and **B**) and within cells belonging to "3 CD4 Temra" cluster for clonotype with *TNFRSF1B* mutation (panel **C**). Differential expression of genes was calculated with t test and adjusted for multiple comparisons with Bonferroni method. In each panel X axis shows the average of log2 fold change between mutated and non-mutated cells and y axis shows the -log10 adjusted p value. Significantly up- and downregulated genes have been marked with red (upregulated in mutated clonotype) and blue (upregulated in cells not belonging to mutated clonotype).

Supplementary Figure 12

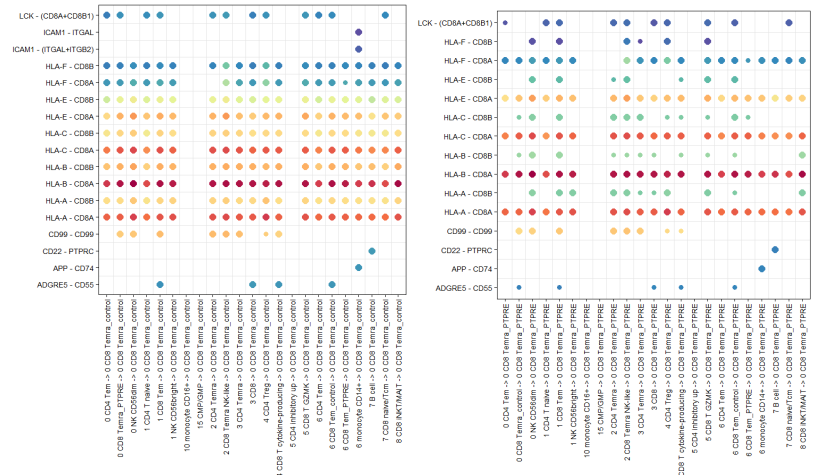
A



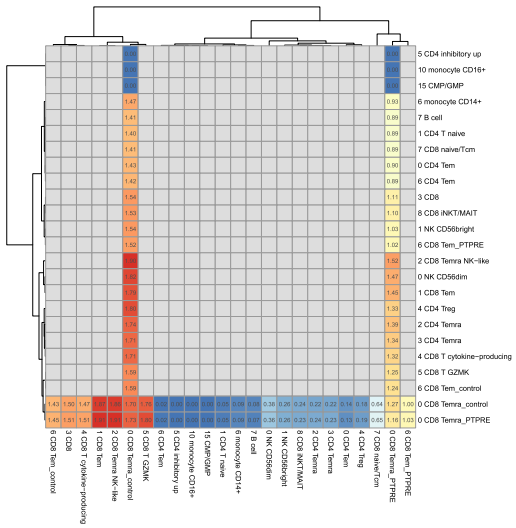
B



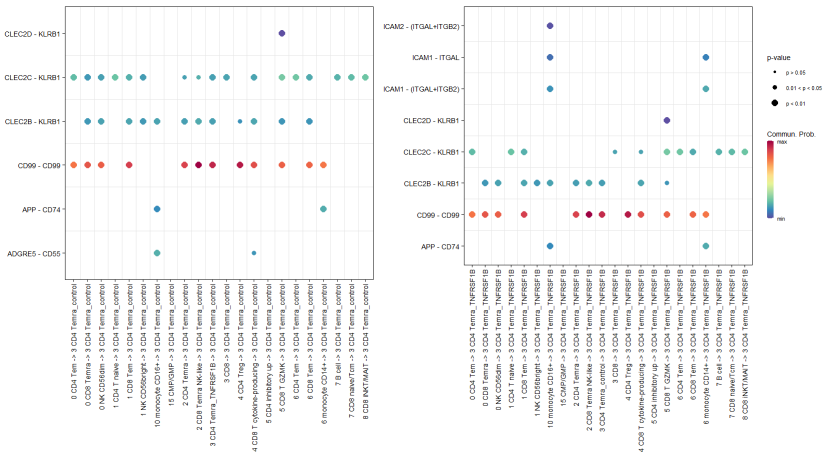
C



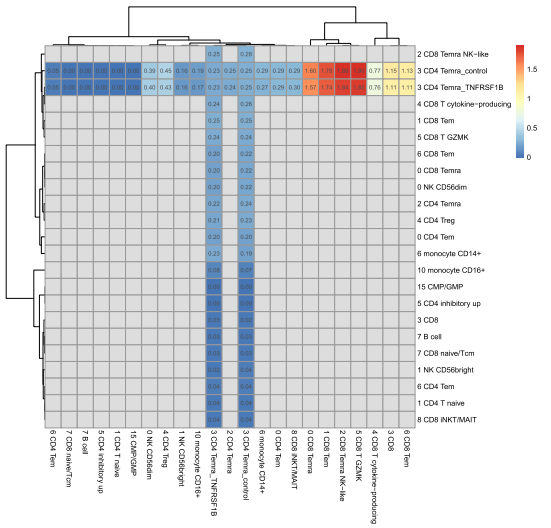
D



E



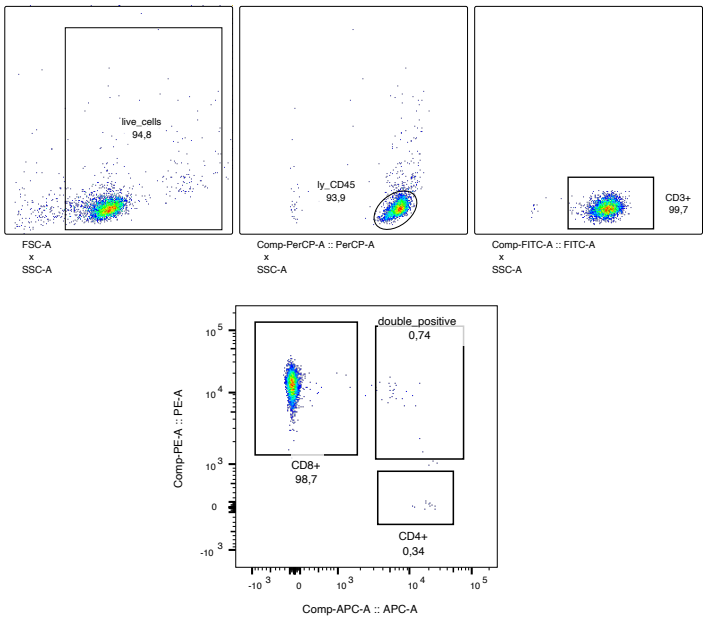
F



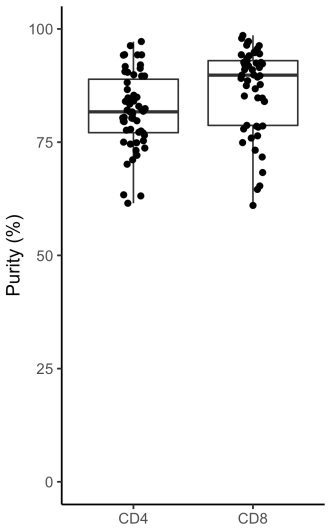
Supplementary Fig. 12. Cell-cell communication analysis. Interaction analysis of T cells, comparing cell-cell communication patterns of mutated and non-mutated T cells with CellChat (*63*). Only cell-cell interactions for the dominating T cell phenotype for each mutated clonotype are shown in the plots. The dotplots (panels **A**, **C**, and **E**) show the statistically significant incoming interactions with respect to the mutated (left) and non-mutated (control, right) clonotypes of the three samples. The heatmaps (panels **B**, **D**, and **F**) show the total interaction strengths (weights) between the cell groups. Each row indicates the communication probability originating from the "sender" (ligand-expressing) cell group to other cell groups (indicated by the columns). (**A-B**) Predicted cell-cell interactions of *TYW1* mutated and *TYW1* wild-type CD8+ Temra cells (CD8 cluster 0). (**C-D**) Predicted cell-cell interactions of *PTPRE* mutated vs *PTPRE* non-mutated CD8+ temra cells (CD8 cluster 0) (**E-F**) Predicted cell-cell interactions of *TNFRSF1B* mutated vs *TNFRSF1B* non-mutated CD4+ Temra cells (CD4 cluster 3).

Supplementary Figure 13

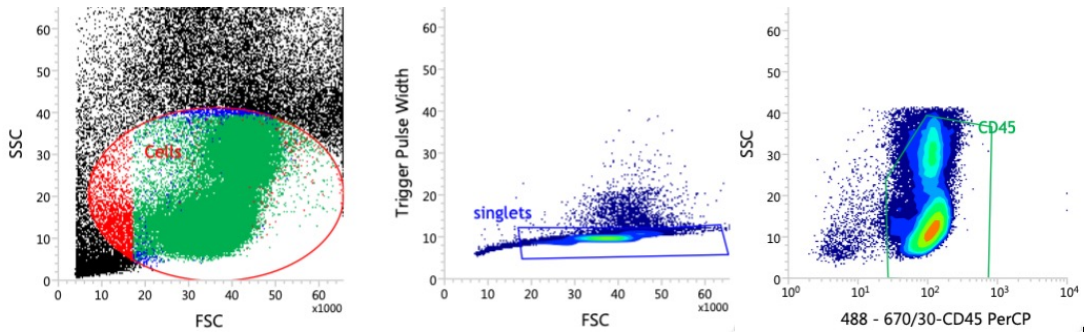
A



B



C



Supplementary Fig. 13. (**A**) Gating strategy for flow-assisted cell sorting and purity analysis of bead-sorted CD3+ or CD8+ samples. (**B**) Purity analysis of bead-sorted samples included in lineage-specific variant analysis. Samples with low purity (less than 60% of live cells were CD4+ or CD8+ T cells) were excluded from the analysis. (**C**) Gating strategy for flow-assisted cell sorting of scRNA+TCRαβ-seq samples.

Supplementary Tables 1-11 (separate excel file).

Supplementary Table 1. Clinical characteristics of patients analyzed with immunogene panel sequencing.
Supplementary Table 2. Lineage-specific somatic mutations in CD4+ or CD8+ compartments.
Supplementary Table 3. Ancestral somatic mutations in CD4+ and CD8+ compartments.
Supplementary Table 4. Lineage-specific CNVs in CD4+ and CD8+ compartments
Supplementary Table 5. Somatic mutations used in comparisons with healthy control T cells
Supplementary Table 6. Size of 5 largest TCR clonotypes in CD4+ and CD8+ samples with TCRβ sequencing.
Supplementary Table 7. Viral epitope analysis for TCRs matching with the highest non-synonymous VAF.
Supplementary Table 8. Differentially expressed genes in CD8+ T cells with scRNA sequencing.
Supplementary Table 9. Differentially expressed genes in CD4+ T cells with scRNA sequencing.
Supplementary Table 10. DRAGEN Bio-IT Platform and annotSV commands used in CNV analysis.
Supplementary Table 11. CNV correlation values for patient samples analyzed with immunogene panel sequencing.

REFERENCES AND NOTES

1. S. Mustjoki, N. S. Young, Somatic mutations in “benign” disease. *N. Engl. J. Med.* **384**, 2039–2052 (2021).

2. N. Kakiuchi, S. Ogawa, Clonal expansion in non-cancer tissues. *Nat. Rev. Cancer* **21**, 239–256 (2021).

3. L. B. Alexandrov, P. H. Jones, D. C. Wedge, J. E. Sale, P. J. Campbell, S. Nik-Zainal, M. R. Stratton, Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47**, 1402–1407 (2015).

4. F. Blokzijl, J. de Ligt, M. Jager, V. Sasselli, S. Roerink, N. Sasaki, M. Huch, S. Boymans, E. Kuijk, P. Prins, I. J. Nijman, I. Martincorena, M. Mokry, C. L. Wiegerinck, S. Middendorp, T. Sato, G. Schwank, E. E. S. Nieuwenhuis, M. M. A. Verstegen, L. J. W. van der Laan, J. de Jonge, J. N. M. IJzermans, R. G. Vries, M. van de Wetering, M. R. Stratton, H. Clevers, E. Cuppen, R. van Boxtel, Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* **538**, 260–264 (2016).

5. L. Moore, A. Cagan, T. H. H. Coorens, M. D. C. Neville, R. Sanghvi, M. A. Sanders, T. R. W. Oliver, D. Leongamornlert, P. Ellis, A. Noorani, T. J. Mitchell, T. M. Butler, Y. Hooks, A. Y. Warren, M. Jorgensen, K. J. Dawson, A. Menzies, L. O’Neill, C. Latimer, M. Teng, R. van Boxtel, C. A. Iacobuzio-Donahue, I. Martincorena, R. Heer, P. J. Campbell, R. C. Fitzgerald, M. R. Stratton, R. Rahbari, The mutational landscape of human somatic and germline cells. *Nature* **597**, 381–386 (2021).

6. G. Genovese, A. K. Kähler, R. E. Handsaker, J. Lindberg, S. A. Rose, S. F. Bakhoum, K. Chambert, E. Mick, B. M. Neale, M. Fromer, S. M. Purcell, O. Svantesson, M. Landén, M. Höglund, S. Lehmann, S. B. Gabriel, J. L. Moran, E. S. Lander, P. F. Sullivan, P. Sklar, H.

Grönberg, C. M. Hultman, S. A. McCarroll, Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* **371**, 2477–2487 (2014).

7. S. Jaiswal, P. Fontanillas, J. Flannick, A. Manning, P. V. Grauman, B. G. Mar, R. C. Lindsley, C. H. Mermel, N. Burtt, A. Chavez, J. M. Higgins, V. Moltchanov, F. C. Kuo, M. J. Kluk, B. Henderson, L. Kinnunen, H. A. Koistinen, C. Ladenvall, G. Getz, A. Correa, B. F. Banahan, S. Gabriel, S. Kathiresan, H. M. Stringham, M. I. McCarthy, M. Boehnke, J. Tuomilehto, C. Haiman, L. Groop, G. Atzmon, J. G. Wilson, D. Neuberg, D. Altshuler, B. L. Ebert, Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* **371**, 2488–2498 (2014).

8. M. Xie, C. Lu, J. Wang, M. D. McLellan, K. J. Johnson, M. C. Wendl, J. F. McMichael, H. K. Schmidt, V. Yellapantula, C. A. Miller, B. A. Ozenberger, J. S. Welch, D. C. Link, M. J. Walter, E. R. Mardis, J. F. Dipersio, F. Chen, R. K. Wilson, T. J. Ley, L. Ding, Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* **20**, 1472–1478 (2014).

9. L. I. Shlush, S. Zandi, A. Mitchell, W. C. Chen, J. M. Brandwein, V. Gupta, J. A. Kennedy, A. D. Schimmer, A. C. Schuh, K. W. Yee, J. L. McLeod, M. Doedens, J. J. F. Medeiros, R. Marke, H. J. Kim, K. Lee, J. D. McPherson, T. J. Hudson; HALT Pan-Leukemia Gene Panel Consortium, A. M. K. Brown, F. Yousif, Q. M. Trinh, L. D. Stein, M. D. Minden, J. C. Y. Wang, J. E. Dick, Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* **506**, 328–333 (2014).

10. K. von Beck, T. von Beck, P. B. Ferrell Jr., A. G. Bick, A. Kishtagari, Lymphoid clonal hematopoiesis: Implications for malignancy, immunity, and treatment. *Blood Cancer J.* **13**, 5 (2023).

11. A. El-Kenawi, B. Ruffell, Inflammation, ROS, and Mutagenesis. *Cancer Cell* **32**, 727–729 (2017).

12. D. Hormaechea-Agulla, K. A. Matatall, D. T. Le, B. Kain, X. Long, P. Kus, R. Jaksik, G. A. Challen, M. Kimmel, K. Y. King, Chronic infection drives Dnmt3a-loss-of-function clonal hematopoiesis via IFNγ signaling. *Cell Stem Cell* **28**, 1428–1442.e6 (2021).

13. S. Avagyan, J. E. Henninger, W. P. Mannherz, M. Mistry, J. Yoon, S. Yang, M. C. Weber, J. L. Moore, L. I. Zon, Resistance to inflammation underlies enhanced fitness in clonal hematopoiesis. *Science* **374**, 768–772 (2021).

14. F. Caiado, L. V. Kovtonyuk, N. G. Gonullu, J. Fullin, S. Boettcher, M. G. Manz, Aging drives *Tet2*+/- clonal hematopoiesis via IL-1 signaling, *Blood* **141**, 886–903 (2023).

15. J. A. Fraietta, C. L. Nobles, M. A. Sammons, S. Lundh, S. A. Carty, T. J. Reich, A. P. Cogdill, J. J. D. Morrissette, J. E. DeNizio, S. Reddy, Y. Hwang, M. Gohil, I. Kulikovskaya, F. Nazimuddin, M. Gupta, F. Chen, J. K. Everett, K. A. Alexander, E. Lin-Shiao, M. H. Gee, X. Liu, R. M. Young, D. Ambrose, Y. Wang, J. Xu, M. S. Jordan, K. T. Marcucci, B. L. Levine, K. C. Garcia, Y. Zhao, M. Kalos, D. L. Porter, R. M. Kohli, S. F. Lacey, S. L. Berger, F. D. Bushman, C. H. June, J. J. Melenhorst, Disruption of TET2 promotes the therapeutic efficacy of CD19-targeted T cells. *Nature* **558**, 307–312 (2018).

16. N. N. Shah, H. Qin, B. Yates, L. Su, H. Shalabi, M. Raffeld, M. A. Ahlman, M. Stetler-Stevenson, C. Yuan, S. Guo, S. Liu, S. H. Hughes, T. J. Fry, X. Wu, Clonal expansion of CAR T cells harboring lentivector integration in the CBL gene following anti-CD22 CAR T-cell therapy. *Blood Adv.* **3**, 2317–2322 (2019).

17. C. L. Nobles, S. Sherrill-Mix, J. K. Everett, S. Reddy, J. A. Fraietta, D. L. Porter, N. Frey, S. I. Gill, S. A. Grupp, S. L. Maude, D. L. Siegel, B. L. Levine, C. H. June, S. F. Lacey, J. J.

Melenhorst, F. D. Bushman, CD19-targeting CAR T cell immunotherapy outcomes correlate with genomic modification by vector integration. *J. Clin. Invest.* **130**, 673–685 (2020).

18. B. Prinzing, C. C. Zebley, C. T. Petersen, Y. Fan, A. A. Anido, Z. Yi, P. Nguyen, H. Houke, M. Bell, D. Haydar, C. Brown, S. K. Boi, S. Alli, J. C. Crawford, J. M. Riberdy, J. J. Park, S. Zhou, M. P. Velasquez, C. DeRenzo, C. R. Lazzarotto, S. Q. Tsai, P. Vogel, S. M. Pruett-Miller, D. M. Langfitt, S. Gottschalk, B. Youngblood, G. Krenciute, Deleting DNMT3A in CAR T cells prevents exhaustion and enhances antitumor activity. *Sci. Transl. Med.* **13**, eabh0272 (2021).

19. H. L. M. Koskela, S. Eldfors, P. Ellonen, A. J. van Adrichem, H. Kuusanmäki, E. I. Andersson, S. Lagström, M. J. Clemente, T. Olson, S. E. Jalkanen, M. M. Majumder, H. Almusa, H. Edgren, M. Lepistö, P. Mattila, K. Guinta, P. Koistinen, T. Kuittinen, K. Penttinen, A. Parsons, J. Knowles, J. Saarela, K. Wennerberg, O. Kallioniemi, K. Porkka, T. P. Loughran Jr., C. A. Heckman, J. P. Maciejewski, S. Mustjoki, Somatic STAT3 mutations in large granular lymphocytic leukemia. *N. Engl. J. Med.* **366**, 1905–1913 (2012).

20. P. Savola, T. Martelius, M. Kankainen, J. Huuhtanen, S. Lundgren, Y. Koski, S. Eldfors, T. Kelkka, M. A. I. Keränen, P. Ellonen, P. E. Kovanen, S. Kytölä, J. Saarela, H. Lähdesmäki, M. R. J. Seppänen, S. Mustjoki, Somatic mutations and T-cell clonality in patients with immunodeficiency. *Haematologica* **105**, 2757–2768 (2020).

21. D. Kim, G. Park, J. Huuhtanen, S. Lundgren, R. K. Khajuria, A. M. Hurtado, C. Muñoz-Calleja, L. Cardeñoso, V. Gómez-García de Soria, T. H. Chen-Liang, S. Eldfors, P. Ellonen, S. Hannula, M. Kankainen, O. Brück, A. Kreutzman, U. Salmenniemi, T. Lönnberg, A. Jerez, M. Itälä-Remes, M. Myllymäki, M. A. I. Keränen, S. Mustjoki, Somatic mTOR mutation in clonally expanded T lymphocytes associated with chronic graft versus host disease. *Nat. Commun.* **11**, 2246 (2020).

22. P. Savola, O. Brück, T. Olson, T. Kelkka, M. J. Kauppi, P. E. Kovanen, S. Kytölä, T. Sokka-Isler, T. P. Loughran, M. Leirisalo-Repo, S. Mustjoki, Somatic STAT3 mutations in Felty syndrome: An implication for a common pathogenesis with large granular lymphocyte leukemia. *Haematologica* **103**, 304–312 (2018).

23. P. Savola, T. Kelkka, H. L. Rajala, A. Kuuliala, K. Kuuliala, S. Eldfors, P. Ellonen, S. Lagström, M. Lepistö, T. Hannunen, E. I. Andersson, R. K. Khajuria, T. Jaatinen, R. Koivuniemi, H. Repo, J. Saarela, K. Porkka, M. Leirisalo-Repo, S. Mustjoki, Somatic mutations in clonally expanded cytotoxic T lymphocytes in patients with newly diagnosed rheumatoid arthritis. *Nat. Commun.* **8**, 15869 (2017).

24. M. Valori, L. Jansson, A. Kiviharju, P. Ellonen, H. Rajala, S. A. Awad, S. Mustjoki, P. J. Tienari, A novel class of somatic mutations in blood detected preferentially in CD8+ cells. *Clin. Immunol.* **175**, 75–81 (2017).

25. M. Valori, J. Lehikoinen, L. Jansson, J. Clancy, S. A. Lundgren, S. Mustjoki, P. Tienari, High prevalence of low-allele-fraction somatic mutations in *STAT3* in peripheral blood CD8+ cells in multiple sclerosis patients and controls. *PLOS ONE* **17**, e0278245 (2022).

26. L. Van Horebeek, N. Dedoncker, B. Dubois, A. Goris, Frequent somatic mosaicism in T lymphocyte subsets in individuals with and without multiple sclerosis. *Front. Immunol.* **13**, 993178 (2022).

27. Z.-Y. Qiu, L. Fan, L. Wang, C. Qiao, Y.-J. Wu, J.-F. Zhou, W. Xu, J.-Y. Li, STAT3 mutations are frequent in T-cell large granular lymphocytic leukemia with pure red cell aplasia. *J. Hematol. Oncol.* **6**, 82 (2013).

28. T. Kawakami, N. Sekiguchi, J. Kobayashi, T. Imi, K. Matsuda, T. Yamane, S. Nishina, Y. Senoo, H. Sakai, T. Ito, T. Koizumi, M. Hirokawa, S. Nakao, H. Nakazawa, F. Ishida, Frequent

STAT3 mutations in CD8+ T cells from patients with pure red cell aplasia. *Blood Adv.* **2**, 2704–2712 (2018).

29. A. Jerez, M. J. Clemente, H. Makishima, H. Rajala, I. Gómez-Seguí, T. Olson, K. McGraw, B. Przychodzen, A. Kulasekararaj, M. Afable, H. D. Husseinzadeh, N. Hosono, F. LeBlanc, S. Lagström, D. Zhang, P. Ellonen, A. Tichelli, C. Nissen, A. E. Lichtin, A. Wodnar-Filipowicz, G. J. Mufti, A. F. List, S. Mustjoki, T. P. Loughran Jr., J. P. Maciejewski, STAT3 mutations indicate the presence of subclinical T-cell clones in a subset of aplastic anemia and myelodysplastic syndrome patients. *Blood* **122**, 2453–2459 (2013).

30. S. Lundgren, M. A. I. Keränen, M. Kankainen, J. Huuhtanen, G. Walldin, C. M. Kerr, M. Clemente, F. Ebeling, H. Rajala, O. Brück, H. Lähdesmäki, S. Hannula, T. Hannunen, P. Ellonen, N. S. Young, S. Ogawa, J. P. Maciejewski, E. Hellström-Lindberg, S. Mustjoki, Somatic mutations in lymphocytes in patients with immune-mediated aplastic anemia. *Leukemia* **35**, 1365–1379 (2021).

31. S. Nik-Zainal, L. B. Alexandrov, D. C. Wedge, P. Van Loo, C. D. Greenman, K. Raine, D. Jones, J. Hinton, J. Marshall, L. A. Stebbings, A. Menzies, S. Martin, K. Leung, L. Chen, C. Leroy, M. Ramakrishna, R. Rance, K. W. Lau, L. J. Mudie, I. Varela, D. J. McBride, G. R. Bignell, S. L. Cooke, A. Shlien, J. Gamble, I. Whitmore, M. Maddison, P. S. Tarpey, H. R. Davies, E. Papaemmanuil, P. J. Stephens, S. McLaren, A. P. Butler, J. W. Teague, G. Jönsson, J. E. Garber, D. Silver, P. Miron, A. Fatima, S. Boyault, A. Langerød, A. Tutt, J. W. M. Martens, S. A. J. R. Aparicio, Å. Borg, A. V. Salomon, G. Thomas, A.-L. Børresen-Dale, A. L. Richardson, M. S. Neuberger, P. A. Futreal, P. J. Campbell, M. R. Stratton; Breast Cancer Working Group of the International Cancer Genome Consortium, Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).

32. E. Jokinen, J. Huuhtanen, S. Mustjoki, M. Heinonen, H. Lähdesmäki, Predicting recognition between T cell receptors and epitopes with TCRGP. *PLoS Comput. Biol.* **17**, e1008814 (2021).

33. M. Goncharov, D. Bagaev, D. Shcherbinin, I. Zvyagin, D. Bolotin, P. G. Thomas, A. A. Minervina, M. V. Pogorelyy, K. Ladell, J. E. McLaren, D. A. Price, T. H. O. Nguyen, L. C. Rowntree, E. B. Clemens, K. Kedzierska, G. Dolton, C. R. Rius, A. Sewell, J. Samir, F. Luciani, K. V. Zornikova, A. A. Khmelevskaya, S. A. Sheetikov, G. A. Efimov, D. Chudakov, M. Shugay, VDJdb in the pandemic era: A compendium of T cell receptors specific for SARS-CoV-2. *Nat. Methods* **19**, 1017–1019 (2022).

34. A. Gonzalez-Perez, N. López-Bigas, Functional impact bias reveals cancer drivers. *Nucleic Acids Res.* **40**, e169 (2012).

35. A. Niroula, A. Sekar, M. A. Murakami, M. Trinder, M. Agrawal, W. J. Wong, A. G. Bick, M. M. Uddin, C. J. Gibson, G. K. Griffin, M. C. Honigberg, S. M. Zekavat, K. Paruchuri, P. Natarajan, B. L. Ebert, Distinction of lymphoid and myeloid clonal hematopoiesis. *Nat. Med.* **27** (2021), 1921, 1927.

36. S. W. Brady, K. G. Roberts, Z. Gu, L. Shi, S. Pounds, D. Pei, C. Cheng, Y. Dai, M. Devidas, C. Qu, A. N. Hill, D. Payne-Turner, X. Ma, I. Iacobucci, P. Baviskar, L. Wei, S. Arunachalam, K. Hagiwara, Y. Liu, D. A. Flasch, Y. Liu, M. Parker, X. Chen, A. H. Elsayed, O. Pathak, Y. Li, Y. Fan, J. R. Michael, M. Rusch, M. R. Wilkinson, S. Foy, D. J. Hedges, S. Newman, X. Zhou, J. Wang, C. Reilly, E. Sioson, S. V. Rice, V. P. Loyola, G. Wu, E. Rampersaud, S. C. Reshmi, J. Gastier-Foster, J. M. G. Auvil, P. Gesuwan, M. A. Smith, N. Winick, A. J. Carroll, N. A. Heerema, R. C. Harvey, C. L. Willman, E. Larsen, E. A. Raetz, M. J. Borowitz, B. L. Wood, W. L. Carroll, P. A. Zweidler-McKay, K. R. Rabin, L. A. Mattano, K. W. Maloney, S. S. Winter, M. J. Burke, W. Salzer, K. P. Dunsmore, A. L. Angiolillo, K. R. Crews, J. R. Downing, S. Jeha, C.-

H. Pui, W. E. Evans, J. J. Yang, M. V. Relling, D. S. Gerhard, M. L. Loh, S. P. Hunger, J. Zhang, C. G. Mullighan, The genomic landscape of pediatric acute lymphoblastic leukemia. *Nat. Genet.* **54**, 1376–1389 (2022).

37. T. Girardi, C. Vicente, J. Cools, K. De Keersmaecker, The genetics and molecular biology of T-ALL. *Blood* **129**, 1113–1123 (2017).

38. A. Stengel, W. Kern, M. Zenger, K. Perglerová, S. Schnittger, T. Haferlach, C. Haferlach, Genetic characterization of T-PLL reveals two major biologic subgroups and JAK3 mutations as prognostic marker. *Gene. Chromosomes Canc.* **55**, 82–94 (2016).

39. C. M. Arends, J. Galan-Sousa, K. Hoyer, W. Chan, M. Jäger, K. Yoshida, R. Seemann, D. Noerenberg, N. Waldhueter, H. Fleischer-Notter, F. Christen, C. A. Schmitt, B. Dörken, U. Pelzer, M. Sinn, T. Zemojtel, S. Ogawa, S. Märdian, A. Schreiber, A. Kunitz, U. Krüger, L. Bullinger, E. Mylonas, M. Frick, F. Damm, Hematopoietic lineage distribution and evolutionary dynamics of clonal hematopoiesis. *Leukemia* **32**, 1908–1919 (2018).

40. A. G. Bick, J. S. Weinstock, S. K. Nandakumar, C. P. Fulco, E. L. Bao, S. M. Zekavat, M. D. Szeto, X. Liao, M. J. Leventhal, J. Nasser, K. Chang, C. Laurie, B. B. Burugula, C. J. Gibson, A. E. Lin, M. A. Taub, F. Aguet, K. Ardlie, B. D. Mitchell, K. C. Barnes, A. Moscati, M. Fornage, S. Redline, B. M. Psaty, E. K. Silverman, S. T. Weiss, N. D. Palmer, R. S. Vasan, E. G. Burchard, S. L. R. Kardia, J. He, R. C. Kaplan, N. L. Smith, D. K. Arnett, D. A. Schwartz, A. Correa, M. de Andrade, X. Guo, B. A. Konkle, B. Custer, J. M. Peralta, H. Gui, D. A. Meyers, S. T. McGarvey, I. Y.-D. Chen, M. B. Shoemaker, P. A. Peyser, J. G. Broome, S. M. Gogarten, F. F. Wang, Q. Wong, M. E. Montasser, M. Daya, E. E. Kenny, K. E. North, L. J. Launer, B. E. Cade, J. C. Bis, M. H. Cho, J. Lasky-Su, D. W. Bowden, L. A. Cupples, A. C. Y. Mak, L. C. Becker, J. A. Smith, T. N. Kelly, S. Aslibekyan, S. R. Heckbert, H. K. Tiwari, I. V. Yang, J. A.

Heit, S. A. Lubitz, J. M. Johnsen, J. E. Curran, S. E. Wenzel, D. E. Weeks, D. C. Rao, D. Darbar, J.-Y. Moon, R. P. Tracy, E. J. Buth, N. Rafaels, R. J. F. Loos, P. Durda, Y. Liu, L. Hou, J. Lee, P. Kachroo, B. I. Freedman, D. Levy, L. F. Bielak, J. E. Hixson, J. S. Floyd, E. A. Whitsel, P. T. Ellinor, M. R. Irvin, T. E. Fingerlin, L. M. Raffield, S. M. Armasu, M. M. Wheeler, E. C. Sabino, J. Blangero, L. K. Williams, B. D. Levy, W. H.-H. Sheu, D. M. Roden, E. Boerwinkle, J. E. Manson, R. A. Mathias, P. Desai, K. D. Taylor, A. D. Johnson; NHLBI Trans-Omics for Precision Medicine Consortium, P. L. Auer, C. Kooperberg, C. C. Laurie, T. W. Blackwell, A. V. Smith, H. Zhao, E. Lange, L. Lange, S. S. Rich, J. I. Rotter, J. G. Wilson, P. Scheet, J. O. Kitzman, E. S. Lander, J. M. Engreitz, B. L. Ebert, A. P. Reiner, S. Jaiswal, G. Abecasis, V. G. Sankaran, S. Kathiresan, P. Natarajan, Inherited causes of clonal haematopoiesis in 97,691 whole genomes. *Nature* **586**, 763–768 (2020).

41. M. Kato, M. Sanada, I. Kato, Y. Sato, J. Takita, K. Takeuchi, A. Niwa, Y. Chen, K. Nakazaki, J. Nomoto, Y. Asakura, S. Muto, A. Tamura, M. Iio, Y. Akatsuka, Y. Hayashi, H. Mori, T. Igarashi, M. Kurokawa, S. Chiba, S. Mori, Y. Ishikawa, K. Okamoto, K. Tobinai, H. Nakagama, T. Nakahata, T. Yoshino, Y. Kobayashi, S. Ogawa, Frequent inactivation of A20 in B-cell lymphomas. *Nature* **459**, 712–716 (2009).

42. A. A. Petti, S. R. Williams, C. A. Miller, I. T. Fiddes, S. N. Srivatsan, D. Y. Chen, C. C. Fronick, R. S. Fulton, D. M. Church, T. J. Ley, A general approach for detecting expressed mutations in AML cells using single cell RNA-sequencing. *Nat. Commun.* **10**, 3660 (2019).

43. G. X. Y. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, M. T. Gregory, J. Shuga, L. Montesclaros, J. G. Underwood, D. A. Masquelier, S. Y. Nishimura, M. Schnall-Levin, P. W. Wyatt, C. M. Hindson, R. Bharadwaj, A. Wong, K. D. Ness, L. W. Beppu, H. J. Deeg, C. McFarland, K. R. Loeb, W. J.

Valente, N. G. Ericson, E. A. Stevens, J. P. Radich, T. S. Mikkelsen, B. J. Hindson, J. H. Bielas, Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).

44. I. Martincorena, A. Roshan, M. Gerstung, P. Ellis, P. Van Loo, S. McLaren, D. C. Wedge, A. Fullam, L. B. Alexandrov, J. M. Tubio, L. Stebbings, A. Menzies, S. Widaa, M. R. Stratton, P. H. Jones, P. J. Campbell, Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880–886 (2015).

45. H. E. Machado, E. Mitchell, N. F. Øbro, K. Kübler, M. Davies, D. Leongamornlert, A. Cull, F. Maura, M. A. Sanders, A. T. J. Cagan, C. McDonald, M. Belmonte, M. S. Shepherd, F. A. Vieira Braga, R. J. Osborne, K. Mahbubani, I. Martincorena, E. Laurenti, A. R. Green, G. Getz, P. Polak, K. Saeb-Parsy, D. J. Hodson, D. G. Kent, P. J. Campbell, Diverse mutational landscapes in human lymphocytes. *Nature* **608**, 724–732 (2022).

46. E. Masle-Farquhar, K. J. L. Jackson, T. J. Peters, G. Al-Eryani, M. Singh, K. J. Payne, G. Rao, D. T. Avery, G. Apps, J. Kingham, C. J. Jara, K. Skvortsova, A. Swarbrick, C. S. Ma, D. Suan, G. Uzel, I. Chua, J. W. Leiding, K. Heiskanen, K. Preece, L. Kainulainen, M. O'Sullivan, M. A. Cooper, M. R. J. Seppänen, S. Mustjoki, S. Brothers, T. P. Vogel, R. Brink, S. G. Tangye, J. H. Reed, C. C. Goodnow, STAT3 gain-of-function mutations connect leukemia with autoimmune disease by pathological NKG2Dhi CD8+ T cell dysregulation and accumulation. *Immunity* **55**, 2386–2404.e8 (2022).

47. E. Holzelova, C. Vonarbourg, M.-C. Stolzenberg, P. D. Arkwright, F. Selz, A.-M. Prieur, S. Blanche, J. Bartunkova, E. Vilmer, A. Fischer, F. Le Deist, F. Rieux-Laucat, Autoimmune lymphoproliferative syndrome with somatic Fas mutations. *N. Engl. J. Med.* **351**, 1409–1418 (2004).

48. K. C. Dowdell, J. E. Niemela, S. Price, J. Davis, R. L. Hornung, J. B. Oliveira, J. M. Puck, E. S. Jaffe, S. Pittaluga, J. I. Cohen, T. A. Fleisher, V. K. Rao, Somatic FAS mutations are common in patients with genetically undefined autoimmune lymphoproliferative syndrome. *Blood* **115**, 5164–5169 (2010).

49. J. E. Niemela, L. Lu, T. A. Fleisher, J. Davis, I. Caminha, M. Natter, L. A. Beer, K. C. Dowdell, S. Pittaluga, M. Raffeld, V. K. Rao, J. B. Oliveira, Somatic KRAS mutations associated with a human nonmalignant syndrome of autoimmunity and abnormal leukocyte homeostasis. *Blood* **117**, 2883–2886 (2011).

50. L. Mularoni, R. Sabarinathan, J. Deu-Pons, A. Gonzalez-Perez, N. López-Bigas, OncodriveFML: A general framework to identify coding and non-coding regions with cancer driver mutations. *Genome Biol.* **17**, 128 (2016).

51. B. Fattizzo, J. Rosa, J. A. Giannotta, L. Baldini, N. S. Fracchiolla, The physiopathology of T-cell acute lymphoblastic leukemia: Focus on molecular aspects. *Front. Oncol.* **10**, 273 (2020).

52. C. A. Taylor, R. A. Watson, O. Tong, W. Ye, I. Nassiri, J. J. Gilchrist, A. V. de los Aires, P. K. Sharma, S. Koturan, R. A. Cooper, V. K. Woodcock, E. Jungkurth, B. Shine, N. Coupe, M. J. Payne, D. N. Church, V. Naranbhai, S. Groha, P. Emery, K. Mankia, M. L. Freedman, T. K. Choueiri, M. R. Middleton, A. Gusev, B. P. Fairfax, IL7 genetic variation and toxicity to immune checkpoint blockade in patients with melanoma. *Nat. Med.* **28**, 2592–2600 (2022).

53. S. Groha, S. A. Alaiwi, W. Xu, V. Naranbhai, A. H. Nassar, Z. Bakouny, T. El Zarif, R. M. Saliby, G. Wan, A. Rajeh, E. Adib, P. V. Nuzzo, A. L. Schmidt, C. Labaki, B. Ricciuti, J. V. Alessi, D. A. Braun, S. A. Shukla, T. E. Keenan, E. Van Allen, M. M. Awad, M. Manos, O. Rahma, L. Zubiri, A.-C. Villani, B. Fairfax, C. Hammer, Z. Khan, K. Reynolds, Y. Semenov, D.

Schrag, K. L. Kehl, M. L. Freedman, T. K. Choueiri, A. Gusev, Germline variants associated with toxicity to immune checkpoint blockade. *Nat. Med.* **28**, 2584–2591 (2022).

54. M. Frick, W. Chan, C. M. Arends, R. Hablesreiter, A. Halik, M. Heuser, D. Michonneau, O. Blau, K. Hoyer, F. Christen, J. Galan-Sousa, D. Noerenberg, V. Wais, M. Stadler, K. Yoshida, J. Schetelig, E. Schuler, F. Thol, E. Clappier, M. Christopeit, F. Ayuk, M. Bornhäuser, I. W. Blau, S. Ogawa, T. Zemojtel, A. Gerbitz, E. M. Wagner, B. M. Spriewald, H. Schrezenmeier, F. Kuchenbauer, G. Kobbe, M. Wiesneth, M. Koldehoff, G. Socié, N. Kroeger, L. Bullinger, C. Thiede, F. Damm, Role of donor clonal hematopoiesis in allogeneic hematopoietic stem-cell transplantation. *J. Clin. Oncol.* **37**, 375–385 (2019).

55. I. Martincorena, K. M. Raine, M. Gerstung, K. J. Dawson, K. Haase, P. Van Loo, H. Davies, M. R. Stratton, P. J. Campbell, Universal patterns of selection in cancer and somatic tissues. *Cell* **171**, 1029–1041.e21 (2017).

56. R. Rosenthal, N. McGranahan, J. Herrero, B. S. Taylor, C. Swanton, deconstructSigs: Delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* **17**, 31 (2016).

57. N. A. Miller, E. G. Farrow, M. Gibson, L. K. Willig, G. Twist, B. Yoo, T. Marrs, S. Corder, L. Krivohlavek, A. Walter, J. E. Petrikin, C. J. Saunders, I. Thiffault, S. E. Soden, L. D. Smith, D. L. Dinwiddie, S. Herd, J. A. Cakici, S. Catreux, M. Ruehle, S. F. Kingsmore, A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. *Genome Med.* **7**, 100 (2015).

58. AnnotSV: An integrated tool for structural variations annotation, *Bioinformatics | Oxford Academic*; https://academic.oup.com/bioinformatics/article/34/20/3572/4970516.

59. M. Shugay, D. V. Bagaev, M. A. Turchaninova, D. A. Bolotin, O. V. Britanova, E. V. Putintseva, M. V. Pogorelyy, V. I. Nazarov, I. V. Zvyagin, V. I. Kirgizova, K. I. Kirgizov, E. V. Skorobogatova, D. M. Chudakov, VDJtools: Unifying post-analysis of T cell receptor repertoires. *PLoS Comput. Biol.* **11**, e1004503 (2015).

60. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).

61. T. Gao, R. Soldatov, H. Sarkar, A. Kurkiewicz, E. Biederstedt, P.-R. Loh, P. V. Kharchenko, Haplotype-aware analysis of somatic copy number variations from single-cell transcriptomes. *Nat. Biotechnol.* **41**, 417–426 (2023).

62. R. Lopez, J. Regier, M. B. Cole, M. I. Jordan, N. Yosef, Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).

63. S. Jin, C. F. Guerrero-Juarez, L. Zhang, I. Chang, R. Ramos, C.-H. Kuan, P. Myung, M. V. Plikus, Q. Nie, Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).

64 H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

65. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 (2013).

66. *Picard Tools - By Broad Institute*; https://broadinstitute.github.io/picard/.

67. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

68. J. G. Tate, S. Bamford, H. C. Jubb, Z. Sondka, D. M. Beare, N. Bindal, H. Boutselakis, C. G. Cole, C. Creatore, E. Dawson, P. Fish, B. Harsha, C. Hathaway, S. C. Jupe, C. Y. Kok, K. Noble, L. Ponting, C. C. Ramshaw, C. E. Rye, H. E. Speedy, R. Stefancsik, S. L. Thompson, S. Wang, S. Ward, P. J. Campbell, S. A. Forbes, COSMIC: The catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019).

69. S. Adnan Awad, M. Kankainen, T. Ojala, P. Koskenvesa, S. Eldfors, B. Ghimire, A. Kumar, S. Kytölä, M. M. Kamel, C. A. Heckman, K. Porkka, S. Mustjoki, Mutation accumulation in cancer genes relates to nonoptimal outcome in chronic myeloid leukemia. *Blood Adv.* **4**, 546–559 (2020).

70. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).

71. O. Dufva, P. Pölönen, O. Brück, M. A. I. Keränen, J. Klievink, J. Mehtonen, J. Huuhtanen, A. Kumar, D. Malani, S. Siitonen, M. Kankainen, B. Ghimire, J. Lahtela, P. Mattila, M. Vähä-Koskela, K. Wennerberg, K. Granberg, S.-K. Leivonen, L. Meriranta, C. Heckman, S. Leppä, M. Nykter, O. Lohi, M. Heinäniemi, S. Mustjoki, Immunogenomic landscape of hematological malignancies. *Cancer Cell* **38**, 380–399.e13 (2020).

72. J. Huuhtanen, D. Bhattacharya, T. Lönnberg, M. Kankainen, C. Kerr, J. Theodoropoulos, H. Rajala, C. Gurnari, T. Kasanen, T. Braun, A. Teramo, R. Zambello, M. Herling, F. Ishida, T. Kawakami, M. Salmi,
T. Loughran, J. P. Maciejewski, H. Lähdesmäki, T. Kelkka, S. Mustjoki, Single-cell characterization of leukemic and non-leukemic immune repertoires in CD8+ T-cell large granular lymphocytic leukemia. *Nat. Commun.* **13**, 1981 (2022).