Supplementary materials

Antonia Schäfer[1], Zuleika Calderin Sollet[1], Marie-Priscille Hervé[1], Stéphane Buhler[1], Sylvie Ferrari-Lacraz[1], Paul J. Norman[3], Katherine M. Kichula[3], Ticiana D. J. Farias[4], Stavroula Masouridi-Levrat[2], Anne-Claire Mamez[2], Amandine Pradier[2], Federico Simonetta[2], Yves Chalandon[2], Jean Villard[1*]

[1] *Transplantation Immunology Unit and National Reference Laboratory for Histocompatibility, Department of Diagnostic, Geneva University Hospitals, Geneva, Switzerland*

[2] *Service of Haematology, Department of Oncology, Geneva University Hospitals and Faculty of Medicine, University of Geneva, Geneva, Switzerland*

[3] *Department of Biomedical Informatics, University of Colorado School of Medicine, Aurora, CO, USA*

[4] *Department of Biological Sciences, The University of North Carolina at Charlotte, Charlotte, NC, USA*

Corresponding author:

*Jean Villard, Transplantation Immunology Unit and National Reference Laboratory for Histocompatibility, Geneva University Hospitals, Gabrielle-Perret-Gentil 4, 1211 Geneva 4, +41 22 372 93 94/ +41 79 553 34 09, jean.villard@hcuge.ch

Description of supplementary material and methods

Supplementary methods

1. Sample collection and processing
2. Bulk TCR CDR3β sequencing
3. High resolution KIR genotyping
4. Spectral flow cytometry immunophenotyping
5. TCR data processing
6. Identification of public and private TCR clonotypes
7. Spectral flow cytometry data preprocessing and gating

Supplementary tables and figures

Table S1

Description: List of antibodies used for the full spectrum flow cytometry panel

Table S2

Description: Demographic and transplant related characteristics of the study cohort stratified according to the CMV groups (seronegative, seropositive and reactivated)

Table S3

Description: Post-transplant clinical events and transplant outcomes of the study cohort stratified according to the CMV groups (seronegative, seropositive and reactivated)

Table S4

Description: HLA restriction and peptide information related to the CMV specific TCR clonotypes

Table S5

Description: Association between TCR/NK cell repertoire reconstitution and different potential modulatory events

Table S6

Description: Absolute counts of NK and $CD4^+$ / $CD8^+$ T cells in the study cohort (*n* = 47)

Figure S1

(A) Description: Study design overview
(B) Description: Swimmer plots detailing the patients' clinical course post-HSCT

Figure S2

(A) Description: Gating strategy for cleaning and identification of viable $CD45^+$ immune cells
(B) Description: Gating hierarchy for identification of major immune cell subsets

Figure S3

Description: Cumulative productive frequency of pathogen-specific TCR clones identified by *in silico* matching with public databases at serial timepoints post-HSCT

Figure S4

(A) Description: Cumulative productive frequency of donor-recipient non-shared and shared anti-CMV specific TCR clones in matched and mismatched CMV reactivated recipients at indicated timepoints post-HSCT
(B) Description: Cumulative productive frequency of donor-recipient non-shared and shared anti-CMV specific TCR clones according to CMV seropositivity/reactivation

serostatus and reactivation.

Figure S5

(A) Description: Proportion of $NKp44^+$, $NKp30^+$, $NKp46^+$, $TRAIL^+$, $FasL^+$, $NKG2D^+$, $PD\text{-}1^+$, $TIGIT^+$ $CD56^{dim}$ cell subsets at indicated timepoints post-HSCT
(B) Description: Proportion of $KIR2DL5^+$, $KIR2DS1^+$, $KIR2DS4^+$, $KIR3DS1^+$ $CD56^{dim}$ cell subsets at indicated timepoints post-HSCT

Figure S6

(A) Description: Representative single-cell plots displaying CD56 and CD107a expression frequency according to stimulation
(B) Description: Frequency of $CD107a^+$ $CD56^+$ NK cells in unstimulated and stimulated cells

Figure S7

(A) Description: CMV peak viral loads in CMV reactivated recipients in relation to the productive Simpson clonality of the TCR repertoire

(B) Description: CMV peak viral loads in CMV reactivated recipients in relation to the frequency of CD57^{+} NKG2C^{+} CD56^{dim} NK cells at 12 months post-alloHSCT

Figure S8

Description: Immune reconstitution of representative CMV seronegative (a.) and CMV reactivated (b.), (c.) recipients.

Supplementary material and methods

1.Sample collection and processing

Peripheral blood mononuclear cell (PBMC) isolation was performed by Ficoll density gradient centrifugation. Briefly, whole blood was first diluted with the equal amount of Phosphate-buffered saline (PBS), layered onto the Ficoll-Paque (Gibco Life Technologies, Oslo, Norway) and centrifuged at 800 *x g* for 20-25 min. Subsequently, PBMCs were isolated from the interface, washed twice with PBS and cryopreserved in fetal bovine serum supplemented with 10 % dimethyl sulfoxide (Merck, Darmstadt, Germany). Samples were stored in liquid nitrogen until use. Genomic DNA was extracted from PBMCs using NucleoSpin kit (Machery-Nagel, Düren, Germany) or directly extracted from the whole blood and purified using the EZ1 Advanced XL machine (QIAGEN, GmbH, Hilden, Germany). All samples were stored at 4°C until usage.

2. Bulk TCR CDR3β sequencing

DNA quality and concentration was assessed using *ds*DNA Qubit kit. The starting DNA input concentration for each sample was between 30 and 44 ng/µl. High-throughput TCR complementary determining region 3 beta (CDR3β) sequencing at survey resolution was performed using the immunoSEQ ® Assay (Adaptive Biotechnologies, Seattle, WA, USA) according to the manufacturer's instructions. Briefly, genomic DNA was amplified by a multiplex PCR reaction using forward and reverse amplification primers targeting all V and J gene segments. TCR pooled libraries were sequenced on an Illumina MiSeq instrument using v3 150 cycle kit chemistry (Illumina, San Diego, CA, USA).

3. High resolution KIR genotyping

DNA purity and concentration was assessed by Nanodrop 200 spectrophotometer. For library preparation, 500 ng of genomic DNA was first fragmented by digestive enzymes (New England Biolabs, Boston, MA, USA) followed by barcode ligation with unique adaptors. After postligation cleanup, dual size selection was performed with AMPure magnetic beads (Beckmann Coulter, Brea, CA, USA) to acquire fragment sizes of 800 to 1200 bp length. In a second step, a pool of oligonucleotide probes specific for KIR and HLA genomic region were used for the targeted capture. Final enriched libraries were normalized to a concentration of 12 pmol/l. Paired-end sequencing was performed using NovaSeq instrument with a sequencing length of 2 x 250 bp (Illumina, San Diego, CA, USA).

4. Spectral flow cytometry immunophenotyping

The antibody panel for the immunophenotyping comprised a total of 35 markers. Antibody clones, concentration, corresponding fluorochromes and supplier are summarized in supplementary Table S1. Antibody master mix was prepared fresh before each staining experiment. Cryopreserved samples were quickly thawed in a water-bath at 37°C and washed twice with warm RPMI-1640 (Gibco, Life Technologies, Oslo Norway) supplemented with 10% fetal bovine serum and – if required – with Pierce Universal Cell Nuclease (Thermo Scientific™, Waltham, MA, USA). Cells were subsequently resuspended in RPMI-1640 with 10% fetal bovine serum and rested overnight in a 37°C, 5 % CO2 incubator to be revitalized. After overnight resting, samples were washed with PBS by centrifugation at 400 *x g* for 5 min. Viability staining was performed by adding 5 µl of a 1:500 diluted ViaDye Red viability staining solution (Cytek® Biosciences, Fremont, CA, USA) to the cells and incubated for 20 min at room temperatur. Cells were subsequently washed with cell staining buffer (CSB) containing 1% BSA, 0.5 M EDTA and 0.02% NaNH3 in PBS. For extracellular staining, Fc receptors were first blocked by adding 5 µl of Fc-receptor blocking (BioLegend, Fell, Germany) solution to each sample followed by 10 min incubation. Afterwards, 80 µl of antibody master mix were added to the cells and incubated for additional 20 min at 4°C protected from the light. Following two washes with 2 ml of CSB, cells were fixed and permeabilized for the intracellular staining: To this end, cell were resuspended in 200 µl of Fix/Perm FoxP3 Solution (Foxp3/Transcription Factor Staining Buffer Set, eBiosciences™, San Diego, CA, USA) and left for 20 min at room temperature. Cells were then washed twice by pelleting with FoxP3 permeabilization buffer at 800 x *g*. After fixation and permeabilization, cells were resuspended in 50 µL of cytoplasmic /intracellular antibody cocktail and incubated for another 20 min at 4°C protected from the light. Cells were then washed twice by pelleting with CSB and resuspended in 400 µl CSB. On the

day of acquisition, cells were finally filtered through a 35 µm nylon mesh filter. Samples were acquired on a 5-Laser Aurora system (Cytek® Biosciences, Fremont, CA, USA) using the SpectroFlo® Software v3.1.0. The instrument QC was done a daily basis using SpectroFlo® QC Beads (Lot 2004) following manufacturers recommendations. Samples were acquired at a high flow rate of ~ 50 µl / sec. For compensation, cells were unmixed using the stored reference controls with autofluorescence extraction option selected.

5. TCR data processing

Raw data processing and sequencing analyses were performed using the immunoSEQ Analyzer 4.0 online platform (http://www.adaptivebiotech.com/immunoseq) and in R environment (version 4.1.3, R Core team, R foundation for statistical computing). At first, raw sequencing datasets were cleaned as follows: (i) According to the criteria of the International ImMunoGeneTics Collaboration, the TCRβ CDR3 region starts at the second conserved cysteine encoded by the 3' position of the Vβ gene segment and ends with the conserved phenylalanine encoded by the 5' position of the Jβ gene segment. Thus, all sequences matching this requirement were kept while all other sequences generating non-productive events were discarded. (ii) We collapsed sequences with the same CDR3 composition (including sequences with different V-D-J rearrangements) into the same clonotype and computed to each clonotype, the sum of the productive frequencies of all the different DNA sequences coding for that same CDR3.

The Simpson clonality was computed as a TCR repertoire diversity marker as follows: $Simpson\ index = 1 - \lambda = 1 - \sum_{i=1}^{N} p2$. Output values range from 0 to 1, where values tending to 1 indicating an uneven with dominant clones distributed repertoire and inversely values tending to 0 a more even polyclonal repertoire. The degree of overlap in terms of TCR clone identity and abundance was computed using the Morisita-Horn index, mathematically defined as follows: $Morisita - Horn\ index = MHI\ (i) = \frac{2\Sigma(xiyi)}{(Dx+Dy)XY}$.

6. Identification of public and private TCR clonotypes

CDR3aa sequences were used to assess TCR clones publicness. Public TCR CDR3β clonotypes were either defined as matching clonotypes present in public databases with antigen-specific validated TCR

(section 2.8.3) or being present in ≥ 2 individuals in our cohort. Identification of identical clonotypes within a D/R pair were considered as private unless the requirement of full chimerism was not set.

7. Spectral flow cytometry data preprocessing and gating

FCS3.0 files were extracted from SpectroFlo® Software v3.1.0 after spectral unmixing and compensation adjustments. For further downstream analysis, non-randomized data were uploaded in FlowJo v10.7.2 software (Tree Star, Ashland, OR, USA): (i) the datasets were checked whether they follow appropriate quality standards using the FlowAI plugin (ii) the datasets were then manually pre-gated on single, alive, CD45+ immune cells (Supplementary Fig. 2A) (iii) Cell populations of interest were further identified via manual gating (Supplementary Fig. 2B).

Supplementary table S1. List of antibodies used for the full spectrum flow cytometry panel

Target	Fluorochrome	Antibody clone	Provider	ug/test	Staining step
CD45	cFluor V547	HI30	Cytek custom	0.2	Surface
CD3	APC/Fire 810	SK7	Biolegend	0.1	Surface
CD4	PE/Fire 640	SK3	Biolegend	0.1	Surface
CD8a	BV570	RPA-T8	Biolegend	0.4	Surface
CD19	BV750	HIB19	Biolegend	0.4	Surface
CD14	BUV805	M5E2	BD biosciences	0.2	Surface
CD56	BV785	5.1H11	Biolegend	0.05	Surface
CD16	BUV496	3G8	BD biosciences	0.05	Surface
CD57	BV605	QA17A04	Biolegend	0.4	Surface
NKG2A	PE-Cy5	S19004C	Biolegend	0.4	Surface
NKG2C	BUV737	134591	BD biosciences	0.2	Surface
NKG2D	cFluor BYG710	1D11	Cytek custom	2.5	Surface
NKp30	BUV563	P30-15	BD biosciences	1	Surface
NKp44	BUV615	P44-8	BD biosciences	1	Surface
NKp46	BV711	9E2	BD biosciences	1	Surface

KIR2DL1	PE-Vio 770	REA284	Miltenyi	1	Surface
KIR2DL1/ S1	APC-Vio 770	REA1010	Miltenyi	1	Surface
KIR2DL2/ L3 /S2	BB700	CH-L	BD biosciences	0.4	Surface
KIR2DL3	APC	REA147	Miltenyi	2.5	Surface
KIR2DL4	PE	181703	R&D Systems	2.5	Surface
KIR2DL5	BV421	UPR1	BD biosciences	1	Surface
KIR2DS4	PE – Vio 615	REA860	Miltenyi	2.5	Surface
KIR3DL1	VioGreen	REA1005	Miltenyi	1	Surface
KIR3DL2	Alexa Fluor 647	539304	R&D Systems	2.5	Surface
KIR3DL3	Alexa Fluor 700	1136B	R&D Systems	2.5	Surface
KIR3DL1 / S1	FITC	REA168	Miltenyi	1	Surface
CD107a	BUV395	H4A3	BD biosciences	0.05	Intra – cellular
TNFa	BV650	Mab11	Biolegend	1	Intra – cellular
PD-1	BUV661	EH12.1	BD biosciences	0.4	Surface
TIGIT	cFluor BYG750	A15153G	Cytek custom	1	Surface
TRAIL	VioBright B515	REA1113	Miltenyi	2.5	Surface
Fas-L	BV480	NOK1	BD biosciences	1	Surface
Granzyme B	Pacific Blue	GB11	Biolegend	0.2	Intra – cellular
Perforin	cFluor R685	DG9	Cytek custom	1	Intra – cellular
Viability	ViaDyeRed		Cytek custom		Surface

Table S2. Demographic and transplant related characteristics of the study cohort stratified according to CMV the groups (seronegative, seropositive and reactivated)

Parameter	All (*n* = 54)	CMV seronegative *n* = 20	CMV seropositive *n* = 20	CMV reactivation (D-/R+) *n* = 9	CMV reactivation (D+/R+) *n* = 4
Recipient age at HSCT in yr (median, IQR)	57 (62)	57 (10)	54 (18)	63 (12)	54.5 (33.5)
Recipient genre (M:F)	37 : 17	15:5	13:8	2:2	7:2
Donor age in yr (median, IQR)	41 (21.5)	38.5 (26.3)	51 (19)	38 (19)	40.5 (5.5)
Donor genre (M:F)	39 : 15	15:5	14:7	7:2	3:1
Underlying diagnosis, *n* (%)					
AML	27 (50)	8 (40)	10 (47.6)	7 (77.8)	2 (50)
ALL	4 (7.4)	3 (15)	1 (4.8)	-	-
AL non specific	2 (3.7)	1 (5)	1 (4.8)	-	-
CML /CLL	2 (3.7)	-	2 (9.5)	-	-
Lymphoma	4 (7.5)	1 (5)	1 (4.8)	2 (22.2)	-
Myeloma	1 (1.9)	1 (5)	-	-	-
MDPS /MDS /MPS	12 (22.2)	6 (30)	4 (19)	-	2 (50)
Hemoglobinopathy	2 (3.7)	-	2 (9.5)	-	-
Donor type, *n* (%)					
MUD	22 (40.7)	11 (55)	9 (42.9)	1 (11.1)	1 (25)
MMUD	2 (3.7)	1 (5)	1 (4.8)	-	-
MRD	9 (16.7)	5 (25)	4 (19)	-	-
Haplo-identical	21 (38.9)	3 (15)	7 (33.3)	8 (88.9)	3 (75)
First transplantation, *n* (%)	49 (90.7)	18 (90)	19 (90.5)	8 (88.9)	4 (100)
Conditioning with ATG, *n* (%)	39 (72.2)	15 (75)	15 (71.4)	7 (77.8)	2 (50)
PTCy, *n* (%)	22 (40.7)	4 (20)	8 (38.1)	8 (88.9)	2 (50)
No T-cell depletion, *n* (%)	53 (98.1)	20 (100)	20 (95.2)	9 (100)	4 (100)
Cryopreservation, *n* (%)	43 (79.6)	19 (95)	14 (66.7)	6 (66.7)	4 (100)
Stem cell source, *n* (%)					
BM	3 (5.5)	-	3 (14.3)	-	-
PBSC	51 (94.4)	20 (100)	18 (85.7)	9 (100)	4 (100=

CMV serostatus, *n* (%)					
D + /R +	22 (40.7)	-	18 (85.7)	-	4 (100)
D - /R +	12 (22.2)	-	3 (14.3)	9 (100)	-
D + /R -	7 (13)	7 (35)	-	-	-
D - /R -	13 (24.1)	13 (65)	-	-	-
Disease risk index, *n* (%)					
Low	4 (7.4)	-	4 (19)	-	-
Intermediate	41 (75.9)	17 (85)	13 (61.9)	7 (77.8)	4 (100)
High / very high	7 (13)	3 (15)	2 (9.5)	2 (22.2)	-
NA	2 (3.7)	-	2 (9.5)	-	-

ALL; acute lymphoblastic leukemic, AML; acute myeloid leukemia, ATG; anti-thymocyte globulin, BM; bone marrow, CLL; chronic lymphoblastic leukemia, CML; chronic myeloid leukemia, CMV; cytomegalovirus, IQR; interquartile range, MDS; myelodysplastic syndrome, MPS; myeloproliferative syndrome, MDPS; myelodysplastic/myeloproliferative syndrome, MMUD; mismatched unrelated donor, MRD; matched related donor, MUD; matched unrelated donor, PBSC; peripheral blood stem cells, PTCy; post-transplant cyclophosphamide

Table S3. Post-transplant clinical events and transplant outcomes of the study cohort stratified according to the CMV groups (seronegative, seropositive and reactivated)

Parameter	All (*n* = 54)	CMV serongative *n* = 20	CMV seropositive *n* = 21	CMV reactivation (D-/R+) *n* = 9	CMV reactivation (D+/R+) *n* = 4
Immunogenic post-transplant complications					
aGvHD, *n* (%)	40 (74.1)	15 (75)	14 (66.7)	8 (88.9)	3 (75)
> 1 episode within 1-year post-HSCT, *n* (%)	8 (14.8)	4 (20)	2 (9.5)	2 (22.2)	-
Severe aGvHD, grade ≥ III, *n* (%)	6 (11.1)	3 (15)	2 (9.5)	-	1 (25)
cGvHD, *n* (%)	13 (24.1)	6 (30)	4 (19)	2 (22.2)	1 (25)
> 1 episode within 1-year post-HSCT, *n* (%)	2 (3.7)	1 (5)	-	-	1 (25)
Moderate to severe cGvHD, *n* (%)	9 (16.7)	4 (20)	3 (14.3)	1 (11.1)	1 (25)
Infectious post-transplant complications					
Chronic viral infection, *n* (%)					
CMV	13 (24)	-	-	9 (100)	4 (100)
EBV	11 (20.4)	2 (10)	6 (28.6)	1 (11.1)	2 (50)
HHV-6	6 (11.1)	2 (10)	3 (14.3)	1 (11.1)	-
HHV-8	1 (1.9)	-	1 (4.7)	-	-
HSV	3 (5.6)	1 (5)	1 (4.7)	-	1 (25)
BKV	5 (9.3)	-	3 (14.3)	2 (22.2)	-
Respiratory viral infections, *n* (%)					
Influenza	5 (9.3)	1 (5)	1 (4.7)	3 (33.3)	-
Sars-Cov2	16 (29.6)	5 (25)	8 (38.1)	1 (11.1)	2 (50)
RSV	4 (7.4)	2 (10)	2 (9.5)	-	-
Rhinovirus	5 (9.3)	2 (10)	2 (9.5)	1 (11.1)	-
Bacterial infections, *n* (%)	26 (48.2)	8 (40)	10 (47.6)	6 (66.7)	2 (50)
Transplant outcomes, *n* (%)					
Disease relapse	11 (20.4)	4 (20)	4 (19)	3 (33.4)	-
Overall survival at 1 year post-HSCT	50 (92.6)	19 (95)	18 (85.7)	9 (100)	4 (100)

Table S4. HLA restriction and peptide information related to the CMV specific TCR clonotypes

HLA	Peptide	*n*	frequency
HLA-A*03:01	KLGGALQAK	12657	0,632280947
HLA-A*02	NLVPMVATV	4048	0,202218004
HLA-A*02:01	NLVPMVATV	748	0,03736637
HLA-C*07:02	CRVLCCYVL	430	0,021480667
HLA-B*07:02	TPRVTGGGAM	421	0,021031072
HLA-DRA1*01	LLQTGIHVRVSQPSL	302	0,015086422
HLA-C*07:02	FRCPRRFCF	263	0,013138176
HLA-B*07:02	RPHERNGFTVL	224	0,011189929
HLA-A*01:01	VTEHDTLLY	194	0,009691278
HLA-B*35:01	IPSINVHHY	93	0,004645819
HLA-A*01:01	YSEHPTFTSQY	80	0,003996403
HLA-B*44	NEGVKAAW	76	0,003796583
HLA-A*02	MLNIPSINV	72	0,003596763
HLA-A1	VTEHDTLLY	69	0,003446898
HLA-A*24:02	QYDPVAALF	43	0,002148067
HLA-B*44:03:08	NEGVKAAW	40	0,001998202
HLA-B*08:01	QIKVRVKMV	35	0,001748426
HLA-A*24:02	AYAQKIFKI	32	0,001598561
HLA-DRA1*01	EHPTFTSQYRIQGKL	29	0,001448696
HLA-A*02:01	YSEHPTFTSQY	22	0,001099011
HLA-B*07	TPRVTGGGAM	19	0,000949146
HLA-B*08	QIKVRVDMV	14	0,000699371
HLA-B*08:01	ELRRKMMYM	13	0,000649416
HLA-A*02:01	VLEETSVML	10	0,00049955
HLA-B*35	IPSINVHHY	10	0,00049955
HLA-B*35:08	FPTKDVAL	10	0,00049955
HLA-B*08:01	QIKVRVDMV	9	0,000449595
HLA-B7	TPRVTGGGAM	9	0,000449595
HLA-A*02	VLEETSVML	6	0,00029973
HLA-A2	NLVPMVATV	5	0,000249775
HLA-B*18	ELKRKMIYM	5	0,000249775
HLA-A*02:01	ARNLVPMVATVQGQN	3	0,000149865
HLA-A*02:01	YILEETSVM	3	0,000149865
HLA-B*08	ELRRKMMYM	3	0,000149865
HLA-A*01	VTEHDTLLY	2	9,99101E-05
HLA-A*02:01	LSEFCRVLCCYVLEE	2	9,99101E-05
HLA-B*07	RPHERNGFTVL	2	9,99101E-05
HLA-B*18	CVETMCNEY	2	9,99101E-05
HLA-B*18	DEEDAIAAY	2	9,99101E-05
HLA-B*35:08	CPSQEPMSIYVY	2	9,99101E-05
HLA-E*01:01:01:03	VMAPRTLIL	2	9,99101E-05
HLA-A*01	VLEETSVML	1	4,9955E-05

HLA-A*01	YSEHPTFTSQY	1	4,9955E-05
HLA-A*02:01	EDVPSGKLFMHVTLG	1	4,9955E-05
HLA-A*02:01:98	NLVPMVATV	1	4,9955E-05
HLA-B*07	RPHERNGFTV	1	4,9955E-05
HLA-B*12	EFFWDANDIY	1	4,9955E-05
HLA-B*35:42:01	IPSINVHHY	1	4,9955E-05

Table S5. Association between TCR/NK cell repertoire reconstitution and different potential modulatory events within the study cohort (*n* = 54)

Parameter	**PTCy**	**aGvHD/cGvHD**	**Sars-Cov-2**
	n = 22 with PTCy *n* = 32 without PTCy	*n* = 43 with GvHD *n* = 11 without GvHD	*n* = 16 with PTCy *n* = 38 without PTCy
TCR clonality	No significant difference between the groups except at T6 with a higher clonality in PTCy treated recipients (p = 0.043). Between 3 and 12 months, stable TCR clonality (p = 0.29).	No significant difference between these different groups (p>0.05). Between 3 and 12 months, significant increase in recipients with GvHD (p = 0.0275).	No significant difference between the groups and between 3 and 12 months (p>0.05).
TCR overlap	Significant reduced TCR overlap in PTCy treated recipients early post-HSCT (t_0-t_3, p = 0.049; t_3-t_6, p = 0.049). At latter timepoints, no significant difference was noted between the groups (t_3-t_{12}, p = 0.39, t_9- t_{12}, p = 0.75).	No significant difference between the groups (p>0.05).	No significant difference between fhe groups (p>0.05).
NK/T cell subsets	Significant higher frequency of $CD8^+$ T cells at 12 months in PTCy treated recipients (p = 0.015). Significant reduced proprtion of $CD56^{bright}$ NK cells at 12 months in PTCy treated recipients (p = 0.02).	Significant reduced proportion of $CD16^+$ $CD56^{dim}$ NK cells at 12 months in GvHD recipients (p = 0.0067). Significant increase in monocytes CD14+ at 6 (p = 0.045) and 12 months in GvHD recipients (p = 0.0093).	Significant reduced frequency of NK cells at 12 months in Sars-Cov-2 infected recipients (p = 0.0485)

aGvHD; acute graft-versus-host disease, cGvHD; chronic graft-versus-host disease, PTCy; post-transplant cyclophosphamide

Table S6. Absolute counts of NK and $CD4^+$ / $CD8^+$ T cells in the study cohort (*n* = 47)

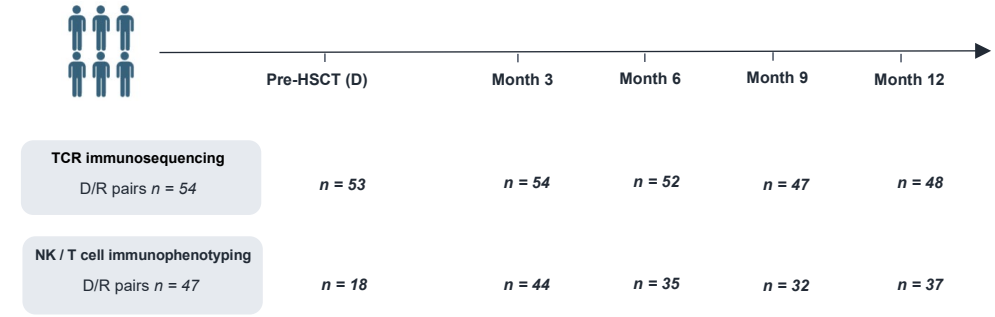
Sample	Timepoint	CD3+ cells	CD4 T - cells	CD8 T - cells	NK cells
p01	T03	263094	62020	109808	227643
p01	T06	324462	86547	210836	215048
p01	T09	95543	17822	74315	48434
p01	T0	422729	309570	95769	54294
p02	T03	91848	65370	20393	103544
p02	T06	89754	68221	19287	88844
p02	T12	113093	59593	32259	28934
p02	T0	807453	542799	233852	48206
p03	T03	122230	21728	43086	25747
p03	T06	90670	16723	64314	62911
p03	T09	107064	21281	77442	70971
p04	T03	75953	39164	24037	33431
p04	T06	287193	60500	193684	62014
p04	T09	383971	107038	191291	40555
p05	T03	5489	3927	1458	4539
p05	T06	5751	2942	2575	5776
p05	T09	7162	3320	3479	3689
p07	T03	165749	159809	4628	52397
p07	T06	157909	103023	13404	179783
p07	T09	261847	180679	40901	101134
p07	T12	144689	94466	22791	58976
p07	T0	400697	302317	67540	23531
p09	T03	76082	18150	47572	117489
p09	T06	89972	24354	47722	132241
p09	T12	75650	19631	48971	83250
p10	T03	4051	811	1486	50527
p10	T06	81425	59793	16424	57699
p10	T12	5857	3871	1031	1703
p11	T03	21597	2383	11106	13695
p11	T06	106398	41918	43788	56447
p11	T09	93065	42748	29285	59268
p11	T12	225686	90311	86755	76215
p12	T03	39022	8129	16105	27926
p12	T06	31096	11441	11726	2727
p12	T12	21575	2784	14589	4186
p13	T03	432997	82727	321251	42994
p13	T06	473462	114716	328086	73480
p13	T09	45564	26968	9828	22047
p13	T12	418132	137397	253077	67192
p13	T0	294424	179721	89258	34084
p14	T03	23477	18902	3333	25880
p14	T06	340769	88474	222208	82304

p14	T12	19383	3893	14547	2219
p17	T03	98094	60693	23446	183950
p17	T06	117668	23547	18185	530672
p17	T09	63646	48651	10420	112683
p17	T12	142149	76990	33095	219829
p18	T06	3138	2605	112	1670
p18	T09	8415	6080	360	9173
p19	T03	71375	29980	35686	34045
p19	T09	252263	37774	183117	61218
p19	T12	182217	39686	128158	20714
p19	T0	234019	178033	44889	24114
p20	T03	210766	69139	87892	40218
p20	T06	396341	72192	189134	45924
p20	T09	248377	44514	110125	11546
p20	T12	427469	75637	235740	62850
p20	T0	352118	194778	101113	63124
p21	T03	62973	61146	908	28168
p21	T09	172986	54559	101058	14875
p21	T12	49047	27824	13881	48400
p21	T0	361461	216965	94604	101416
p22	T03	12862	3114	5008	45964
p22	T12	44505	11274	21994	7564
p23	T03	73563	38268	28304	49289
p23	T06	116616	84230	27922	15440
p23	T09	119487	75777	39339	7251
p23	T12	87892	59013	21815	9472
p23	T0	160169	93437	33874	20612
p24	T03	77860	43099	23167	23918
p24	T09	188904	44935	101407	136110
p24	T12	202856	69889	105677	142659
p25	T03	2679	720	1312	118280
p25	T06	266787	15082	187193	51196
p25	T09	460331	50207	333912	78266
p25	T12	378281	58136	272489	17123
p25	T0	545673	360047	133792	64103
p27	T03	23351	8286	11656	88239
p27	T06	137555	51599	40647	98126
p27	T09	160276	58344	79405	113573
p27	T12	40213	13910	20983	18077
p28	T03	127161	30231	78210	65526
p28	T12	287788	93382	157991	43985
p30	T03	55721	12597	30820	46796
p30	T12	203548	86528	92954	20768
p32	T03	10022	5967	403	4089
p32	T06	1393	680	130	655
p32	T0	63098	39750	18412	3318

p34	T03	19853	12854	6085	214031
p34	T06	18180	8807	6764	28694
p34	T09	101029	38327	47488	64041
p34	T12	225987	81963	118721	175827
p34	T0	202960	134161	43271	37639
p35	T03	327128	84631	212950	20589
p35	T09	514120	93212	382631	26047
p35	T12	462108	97810	339818	36455
p35	T0	332195	194985	61686	75072
p36	T03	4400	3134	618	59987
p36	T09	88182	24086	55111	118589
p36	T12	117559	47844	55994	80544
p36	T0	246898	188586	50885	25662
p37	T03	246845	64689	105435	60467
p37	T06	177777	50295	92768	19616
p37	T09	348585	95503	200902	61437
p37	T12	376810	118873	214527	74913
p37	T0	215739	137749	57752	31961
p38	T03	20162	10160	5793	96279
p38	T06	237302	42008	100008	38563
p38	T09	132948	31441	68944	12770
p39	T03	22108	18701	2024	8310
p39	T09	3216	1960	662	503
p39	T12	159401	114539	38809	13337
p40	T03	4580	2271	962	528
p40	T06	71010	34601	25318	2166
p40	T12	387881	185000	149620	25708
p41	T03	9453	1527	2952	7248
p41	T06	258847	43006	175467	83339
p41	T12	249514	90242	134902	27181
p42	T06	13150	4801	6109	16498
p42	T12	49698	23454	21400	26034
p42	T0	55862	28961	22785	8254
p43	T03	10203	3508	1439	35726
p43	T06	257316	14021	68119	70095
p43	T09	285954	36990	73786	91004
p43	T12	295962	74231	108200	50126
p44	T06	12943	5975	3894	2056
p44	T12	331116	141680	155674	53608
p44	T0	50099	22107	21503	8727
p45	T03	154404	49067	85459	76956
p45	T06	84767	26532	46076	11116
p45	T12	202913	55066	123126	24928
p47	T03	28456	22186	4375	14039
p47	T06	41728	28115	11915	6781
p47	T09	52995	33775	17995	2452

p48	T03	206113	17105	113743	51914
p48	T06	242572	38843	153596	61852
p48	T09	269283	36875	188478	96328
p48	T0	376273	184628	141439	31090
p49	T03	18315	9220	6394	3156
p49	T06	17663	3058	7765	6713
p49	T09	128245	36320	76847	37633
p49	T12	48047	19380	23779	7543
p49	T0	131243	63733	42438	8613
p50	T03	38236	7575	21568	29622
p50	T06	27492	6637	12084	12878
p50	T09	66818	18499	38151	66406
p50	T12	61996	24299	30427	30126
p52	T03	45715	27278	14862	91698
p52	T06	29930	18198	7984	76376
p52	T09	29423	22575	4027	53120
p53	T03	47030	28214	12890	7365
p53	T06	46827	24490	13934	4034
p53	T09	431502	238648	138050	24892
p53	T12	142678	86561	45180	2798
p55	T03	107422	39501	60046	26500
p55	T09	159123	61336	83996	8559
p55	T12	273310	65192	178918	17227
p59	T03	18576	11922	3499	19598
p59	T09	101015	85777	11187	22302
p59	T12	234907	115046	64881	26412
p61	T03	125145	58908	34386	15040
p61	T12	404667	111497	220583	64000
p66	T03	85143	4027	43664	5936
p66	T06	586291	101094	306430	98920
p66	T09	502321	174829	260046	81323
p66	T12	199037	80890	91169	27526

A


Pre-HSCT (D)
Month 3
Month 6
Month 9
Month 12
TCR immunosequencing
D/R pairs n = 54
n = 53
n = 54
n = 52
n = 47
n = 48
NK / T cell immunophenotyping
D/R pairs n = 47
n = 18
n = 44
n = 35
n = 32
n = 37

B

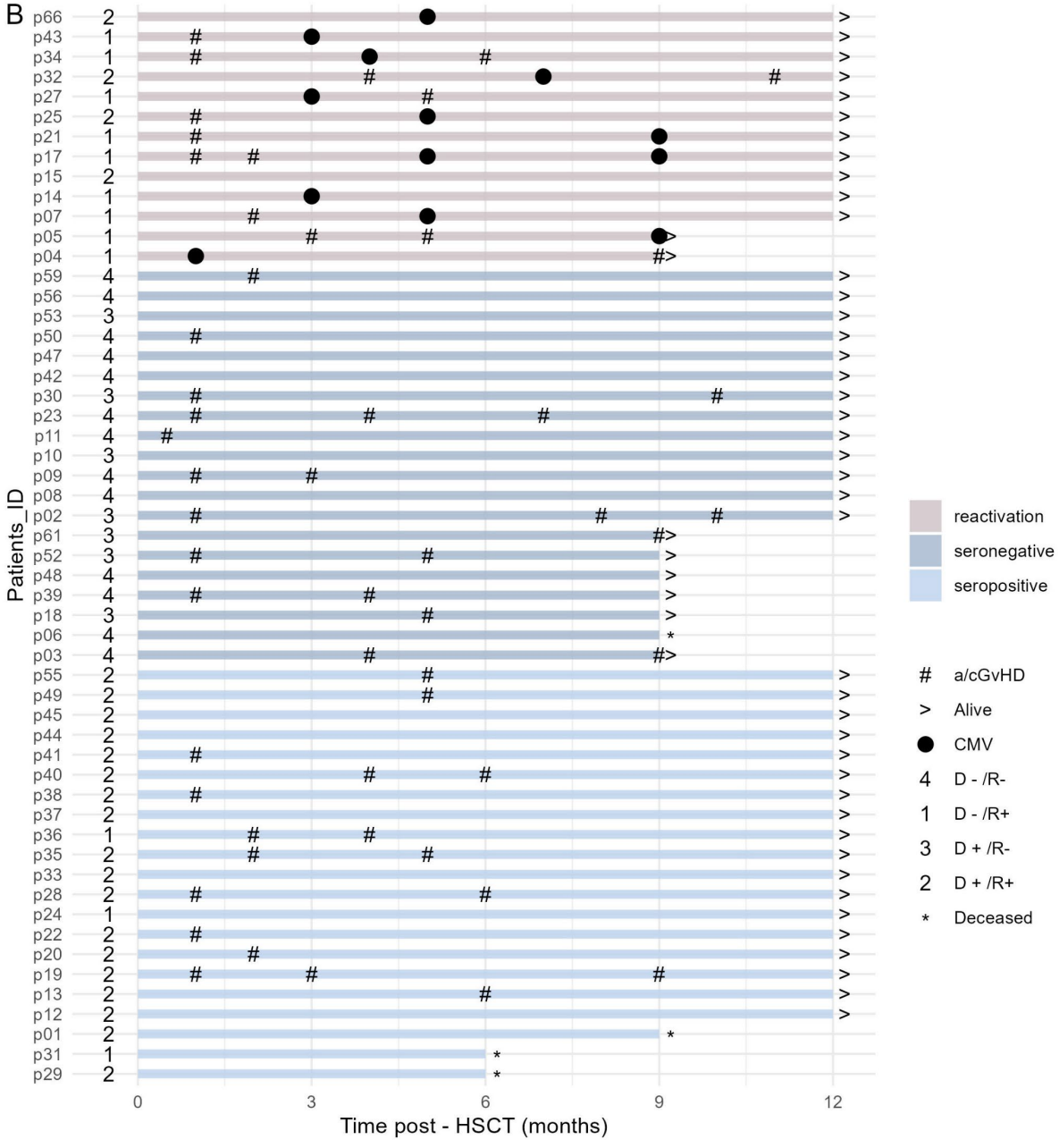
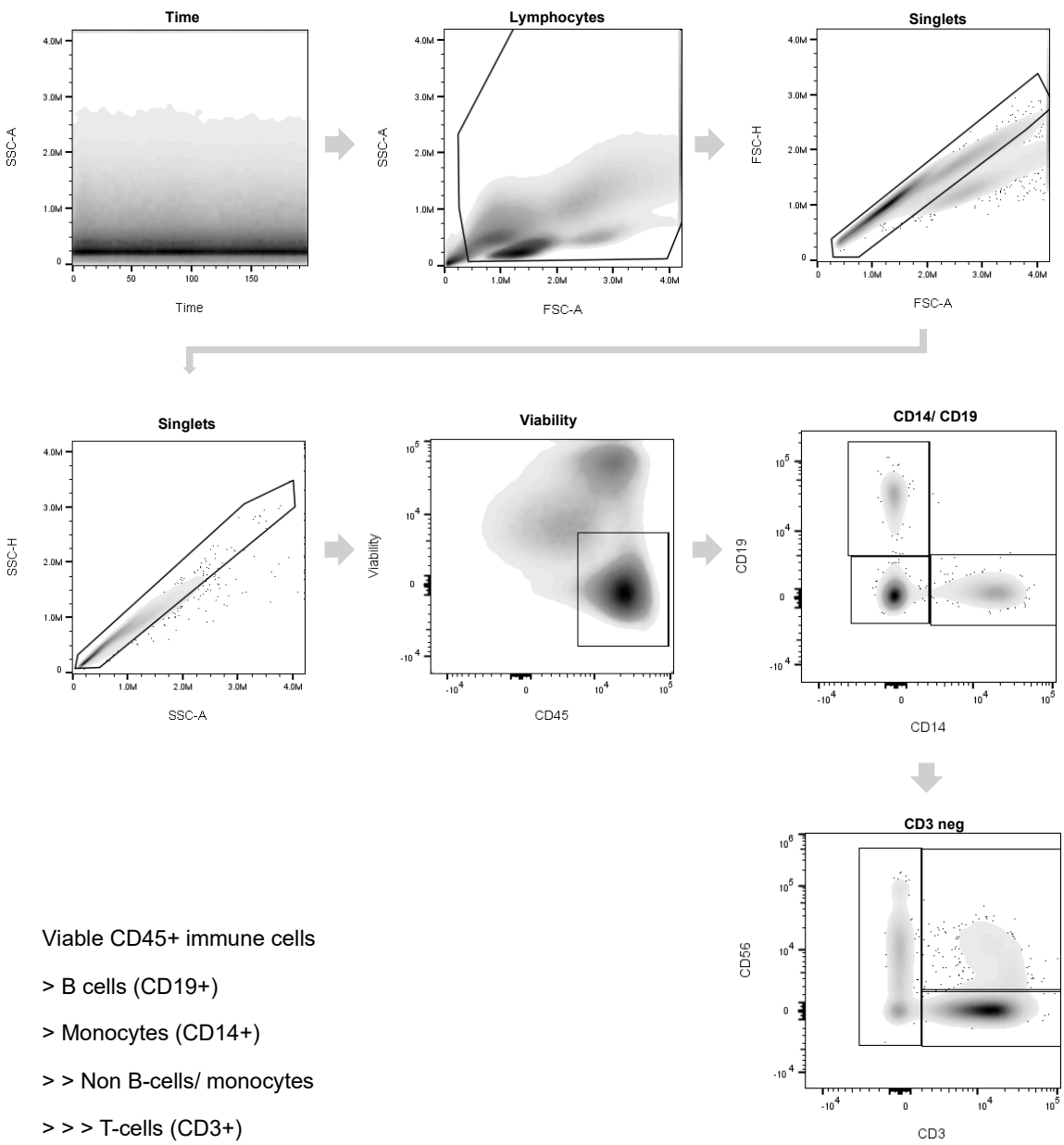

Patients_ID
Time post - HSCT (months)
0
3
6
9
12
reactivation
seronegative
seropositive
a/cGvHD
> Alive
CMV
4 D - /R-
1 D - /R+
3 D + /R-
2 D + /R+
* Deceased

Fig. S1 (A) Peripheral blood samples derived from five different timepoints (pre-HSCT from the donor and 3-, 6-, 9-, 12-months from the recipient) were analyzed by spectral flow cytometry and TCR immunosequencing **(B)** Swimmer plots detailing the patients' clinical course with selective clinical events and stratified according CMV serostatus and infection/reactivation episodes. Bottom horizontal axis display months from HSCT and each line represents an HSCT recipient.

A



B

Viable CD45+ immune cells

> B cells (CD19+)

> Monocytes (CD14+)

> > Non B-cells/ monocytes

> > > T-cells (CD3+)

> > > > > CD4+ T-cells (CD4+ CD8a-)

> > > > > CD8+ T-cells (CD4- CD8a+)

> > > > > CD4+ CD8+ DP T-cells

> > > > > CD8- CD4- DN T-cells

> > > NKT cells (CD3+ CD56+)

> > > Non T cells (CD3-)

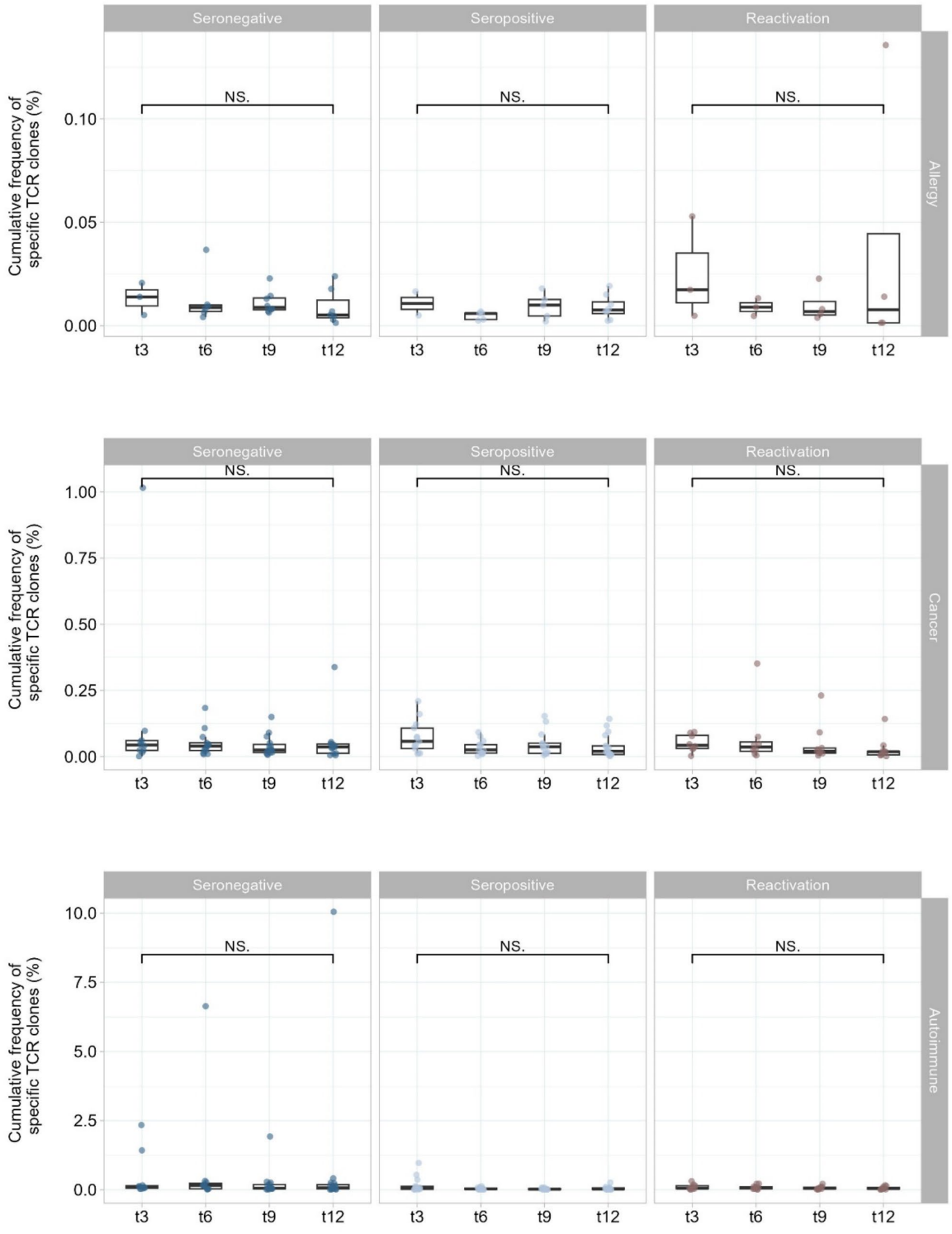
> > > > > NK-cells (CD56+ CD16+)

> > > > > > CD56bright NK-cells (CD56++ CD16+/-)

> > > > > > CD56dim NK-cells (CD56+ CD16+/-)

> > > > > Unconventional NK-cells (CD56- CD16+)

Fig. S2 (A) Gating strategy for cleaning and identification of viable $CD45^+$ immune cells **(B)** Gating hierarchy for identification of major immune cell subsets based on following markers: CD3, CD8a, CD4, CD19, CD14, CD16 and CD56.


Seronegative
Seropositive
Reactivation
Allergy
Cancer
Autoimmune
Cumulative frequency of specific TCR clones (%)
NS.
t3
t6
t9
t12
0.00
0.05
0.10
0.25
0.50
0.75
1.00
0.0
2.5
5.0
7.5
10.0

1

2

3

4

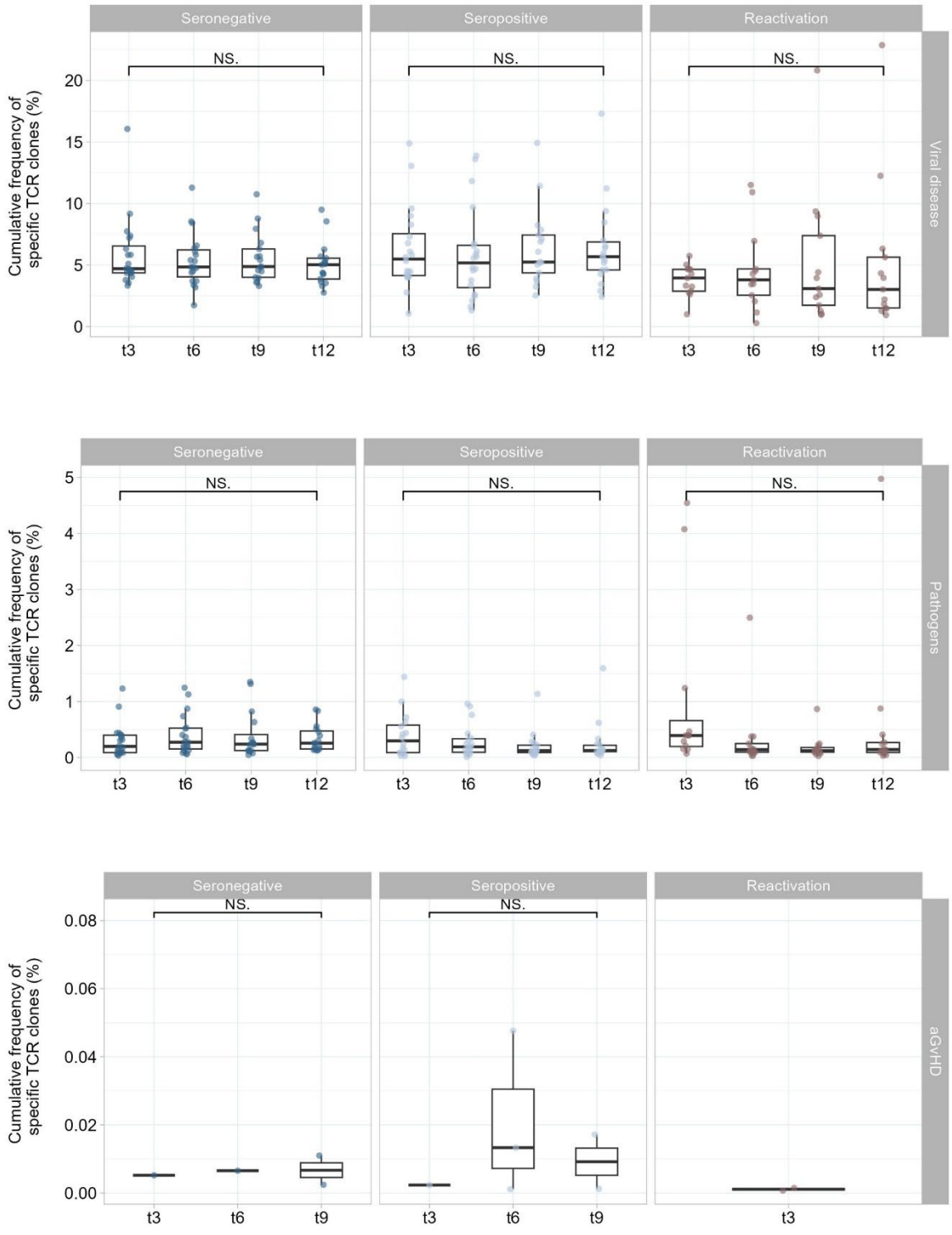


Fig. S3 Cumulative productive frequency of pathogen-specific TCR clones identified by *in silico* matching with public databases at serial timepoints post-HSCT (t0, *n* = 53; t3, *n* = 54; t6, *n* = 52; t9, *n* = 47; t12, *n* = 48). Box plots display medians and interquartile ranges (IQR), with whiskers representing 1.5× IQR. Wilcoxon sum rank test in (A). All *p* values were 2 sided. Statistical thresholds: (NS) not significant

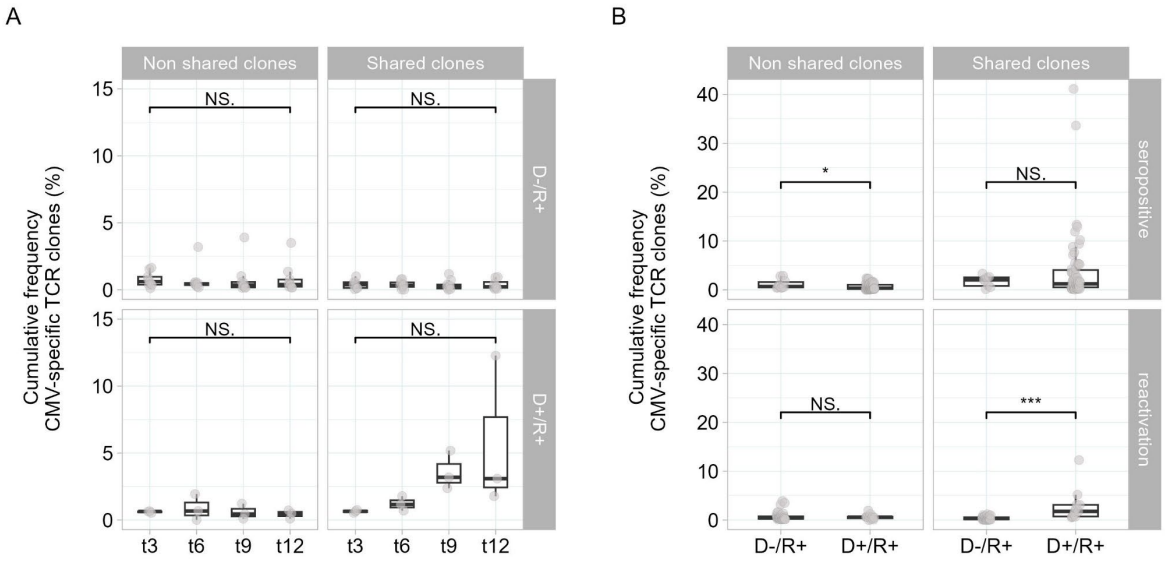
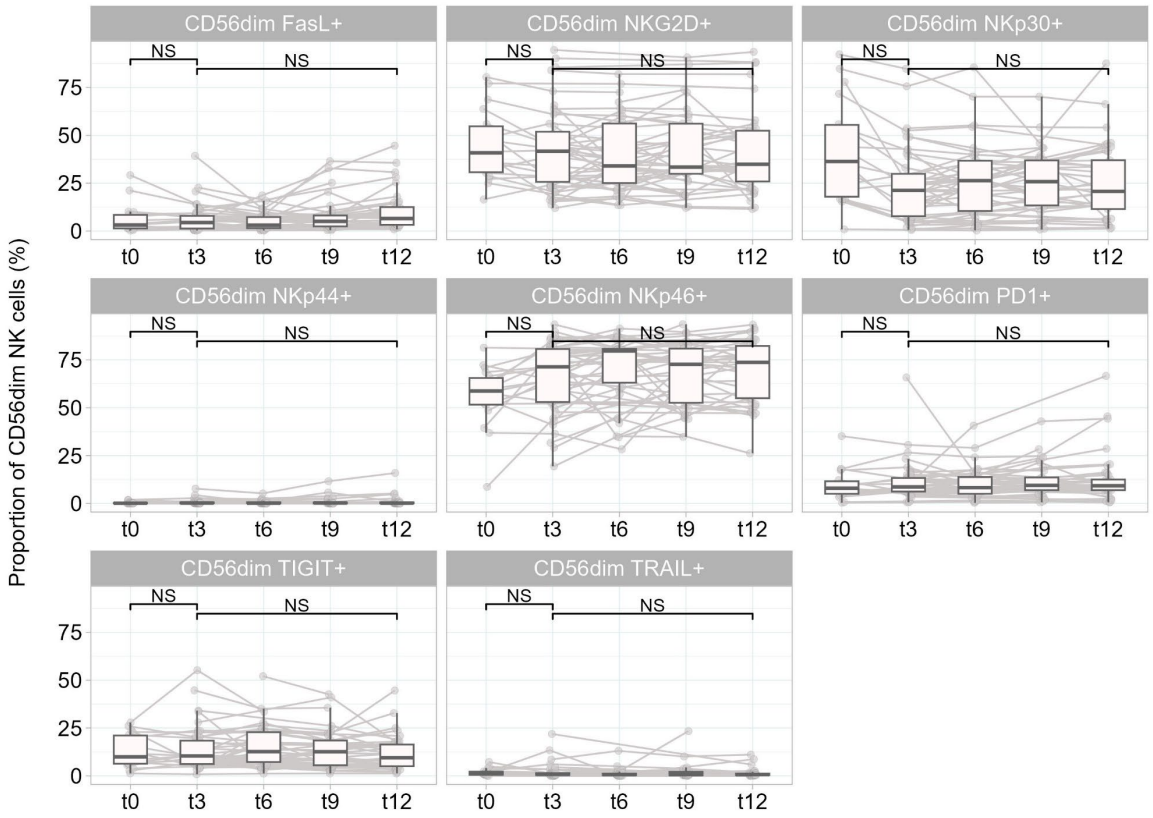
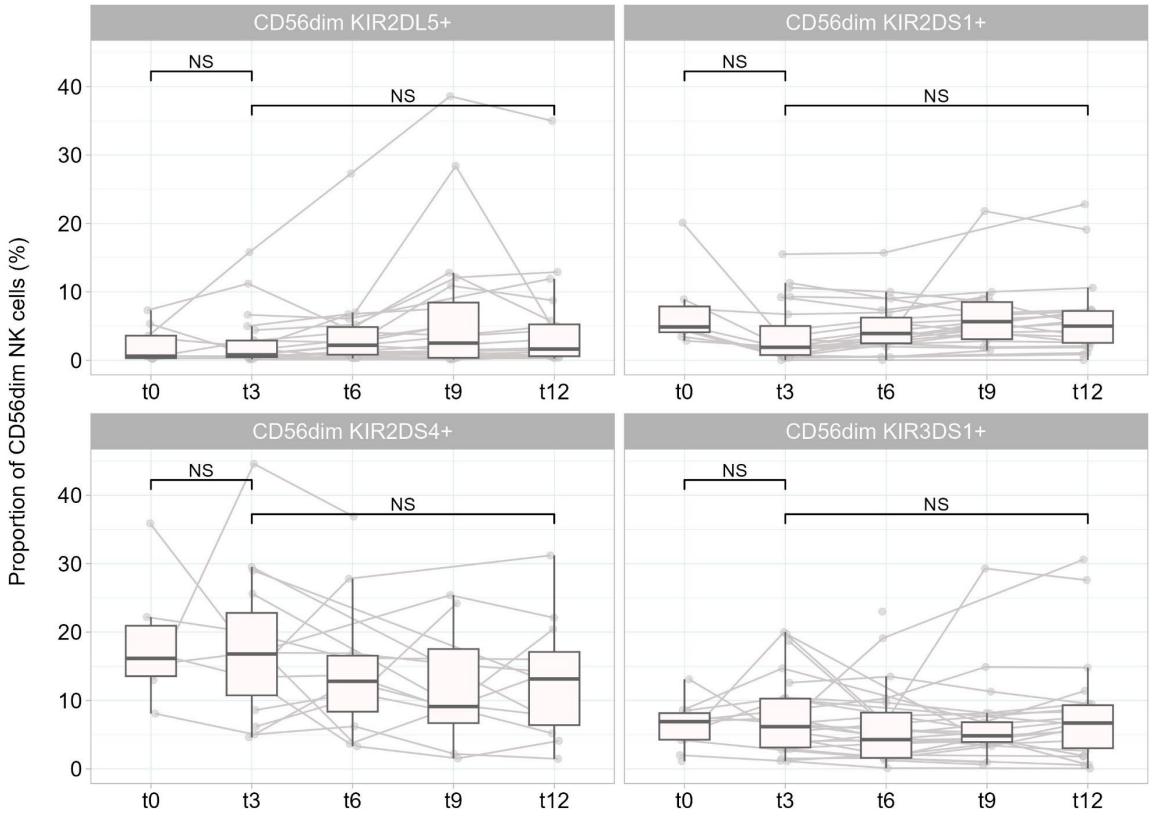


Fig. S4 Stratification according to following groups: seronegative (*n* = 20), seropositive (*n* = 21) and reactivated (*n* = 13) CMV recipients applies to plot (A) and (B). **(A)** Cumulative productive frequency of donor-recipient non-shared and shared CMV- specific TCR clones in D-/R+ (*n* = 9) and D+/R+ (*n* = 4) CMV reactivated recipients at indicated timepoints post-HSCT. **(B)** Cumulative productive frequency of donor-recipient non-shared and shared CMV-specific TCR clones according to CMV serostatus and reactivation. Box plots display medians and interquartile ranges (IQR), with whiskers representing 1.5× IQR. Wilcoxon rank sum test in (A), (B). All *p* values were 2 sided. Statistical thresholds: (*) $p < 0.05$, (***) $p < 0.001$ and (NS) not significant.

A


CD56dim FasL+
CD56dim NKG2D+
CD56dim NKp30+
CD56dim NKp44+
CD56dim NKp46+
CD56dim PD1+
CD56dim TIGIT+
CD56dim TRAIL+
NS
Proportion of CD56dim NK cells (%)
0
25
50
75
t0
t3
t6
t9
t12

B


CD56dim KIR2DL5+
CD56dim KIR2DS1+
CD56dim KIR2DS4+
CD56dim KIR3DS1+
NS
Proportion of CD56dim NK cells (%)
0
10
20
30
40
t0
t3
t6
t9
t12

1

Fig. S5 (A) Proportion of NKp44^{+}, NKp30^{+}, NKp46^{+}, TRAIL^{+}, FasL^{+}, NKG2D^{+}, PD-1^{+}, TIGIT^{+} CD56^{dim} cell subsets at indicated timepoints post-HSCT: t0 (n = 18), t3 (n = 44), t6 (n = 35), t9 (n = 32), t12 (n = 37). **(B)** Proportion of KIR2DL5+, KIR2DS1+, KIR2DS4+, KIR3DS1+ CD56dim cell subsets at indicated timepoints post-HSCT: t0 (n = 18), t3 (n = 44), t6 (n = 35), t9 (n = 32), t12 (n = 37). Lines connect paired samples. Box plots display medians and interquartile ranges (IQR), with whiskers representing 1.5× IQR. Wilcoxon rank sum test with FDR correction in (A), (B). All *p* values were 2 sided. Statistical thresholds: (NS) not significant.

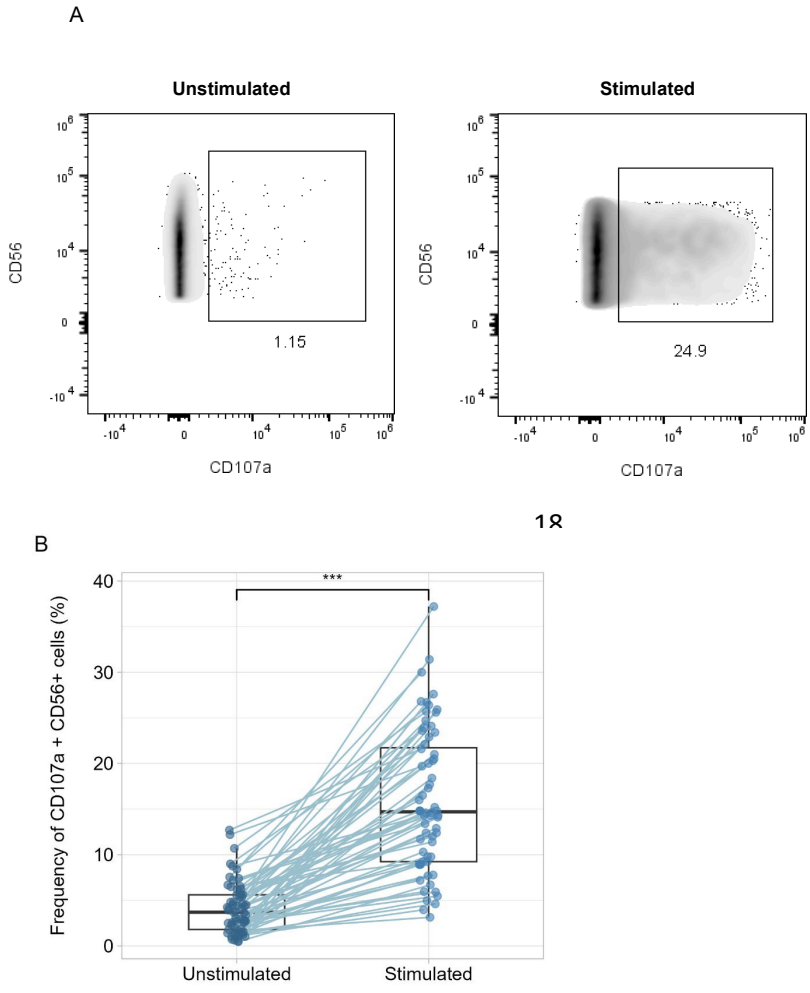


Fig. S6 (A) Representative single-cell plots displaying CD56 and CD107a expression frequency without K562 and post-K562 stimulation. **(B)** Frequency of CD107a^{+} CD56^{+} NK cells in unstimulated and stimulated cells. Lines connect paired samples. Box plots display medians and interquartile ranges (IQR), with whiskers representing 1.5× IQR. Wilcoxon rank sum test in (B). All *p* values were 2 sided. Statistical thresholds: (***) $p < 0.001$

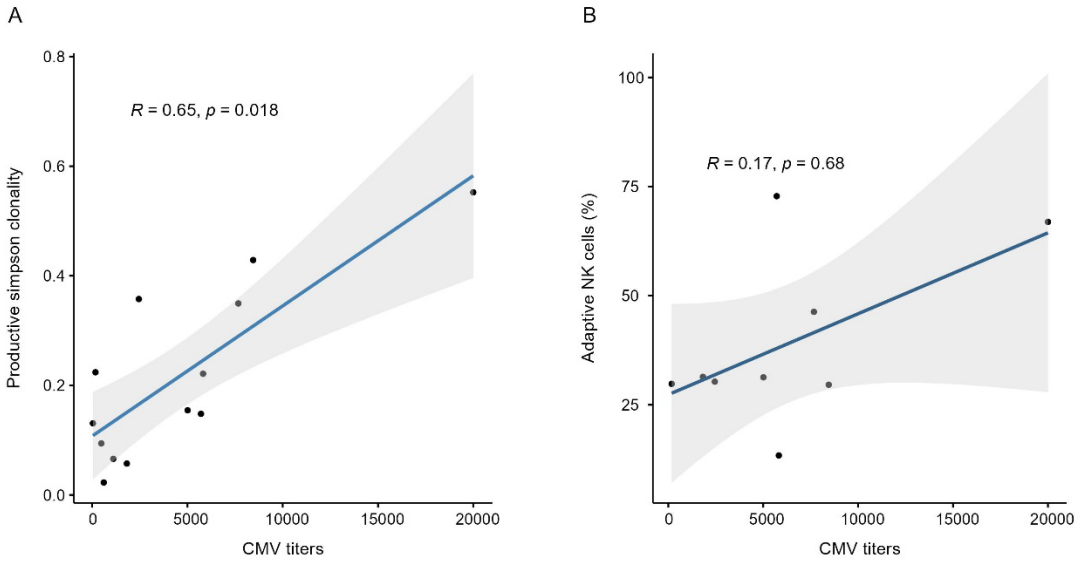
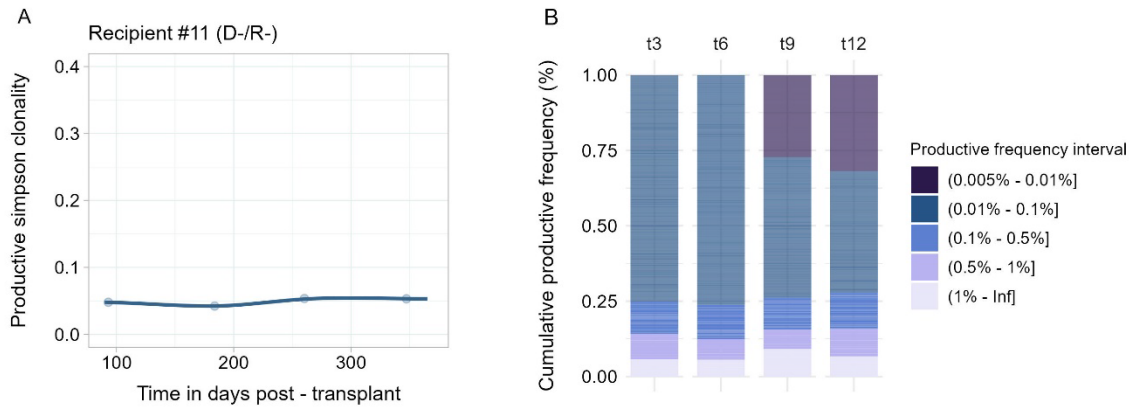
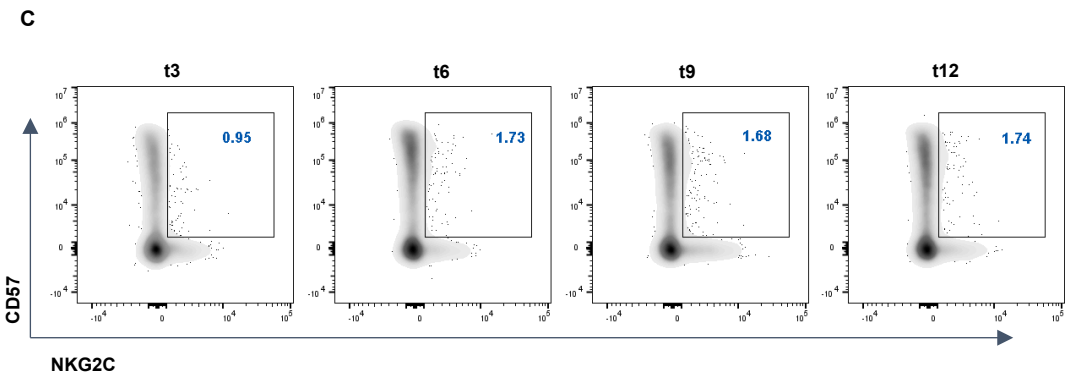
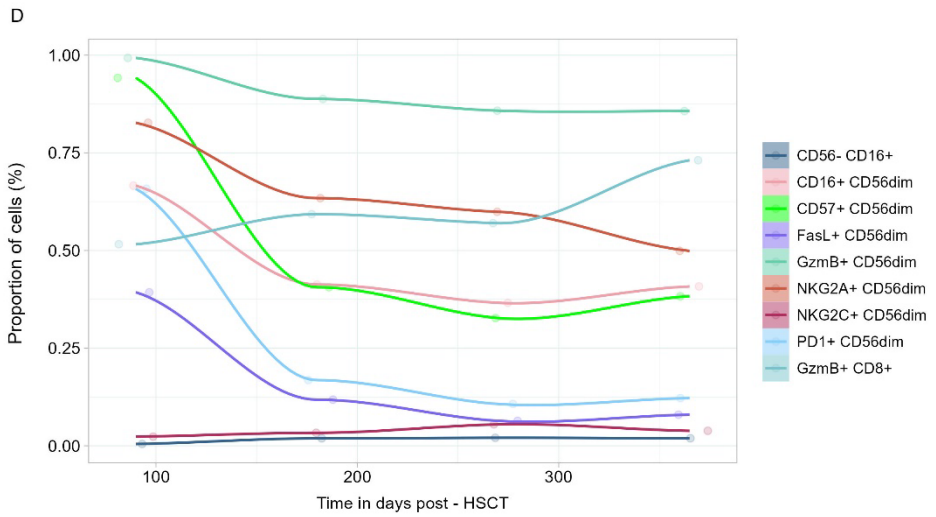


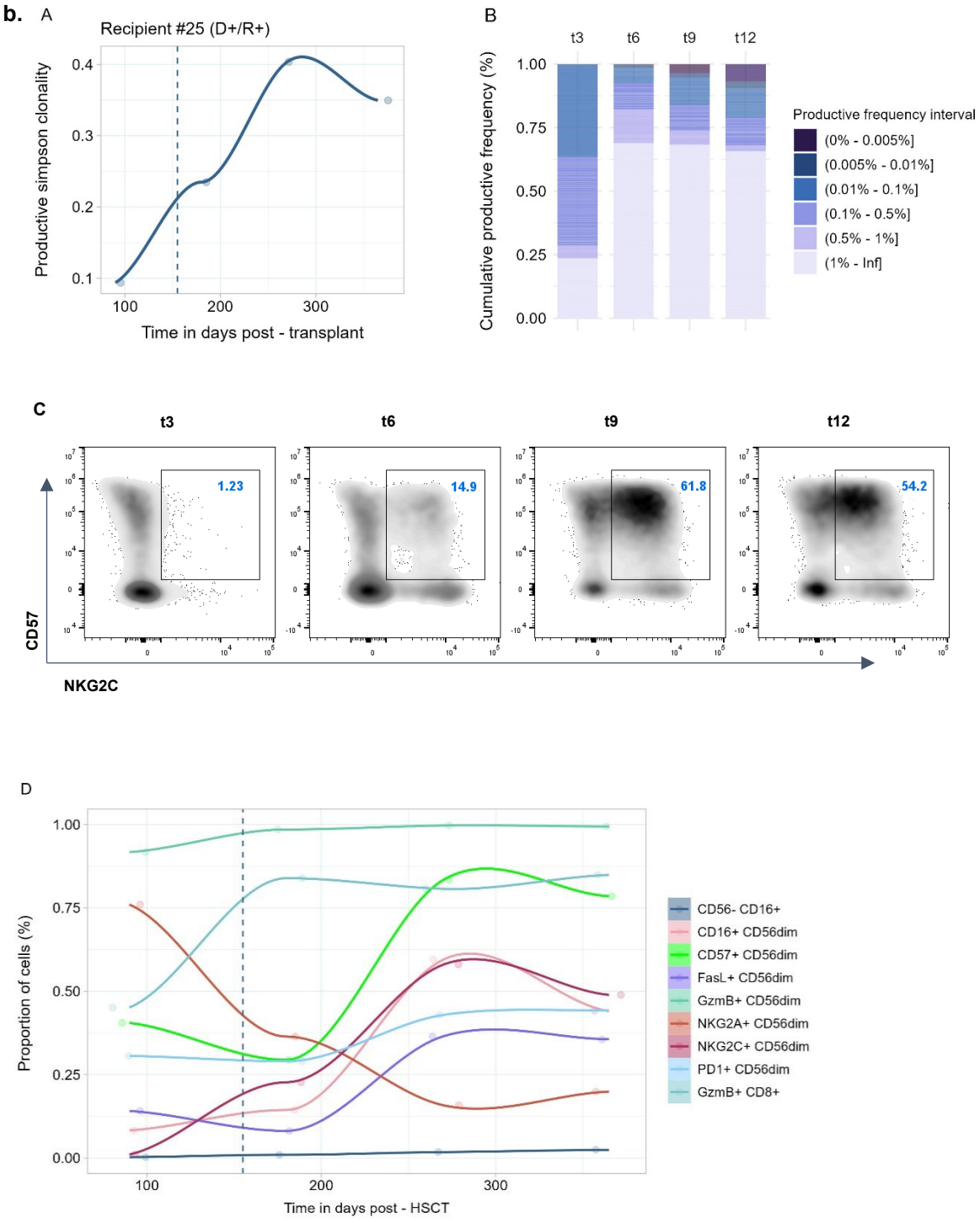
Fig. S7: (A) Scatter plot showing CMV peak viral loads in CMV reactivated recipients (*x* axis) in relation to the productive Simpson clonality of the TCR repertoire (*y* axis) **(B)** Scatter plot showing CMV peak viral loads in CMV reactivated recipients (*x* axis) in relation to the frequency of $CD57^{+}$ $NKG2C^{+}$ $CD56^{dim}$ NK cells at 12 months post-alloHSCT (*y* axis). Spearman rank correlation coefficient is indicated on each plot with the corresponding *p*-value.

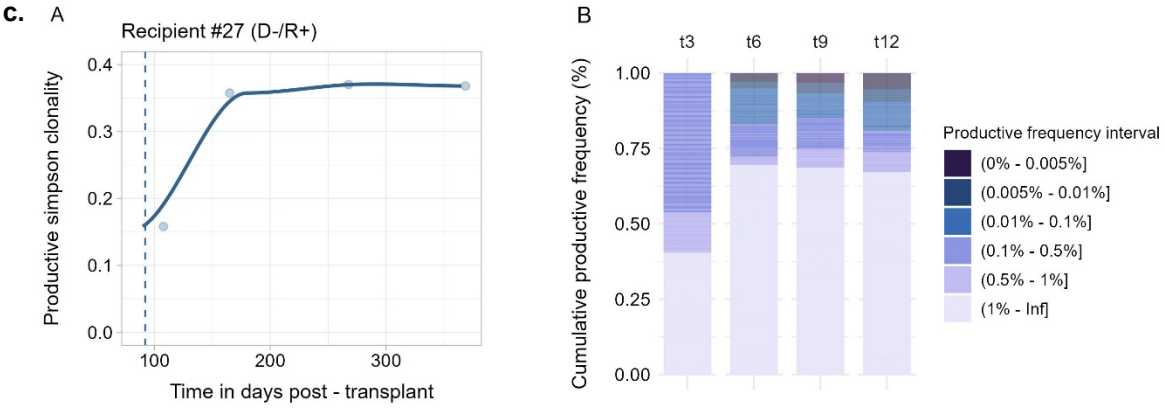
a.

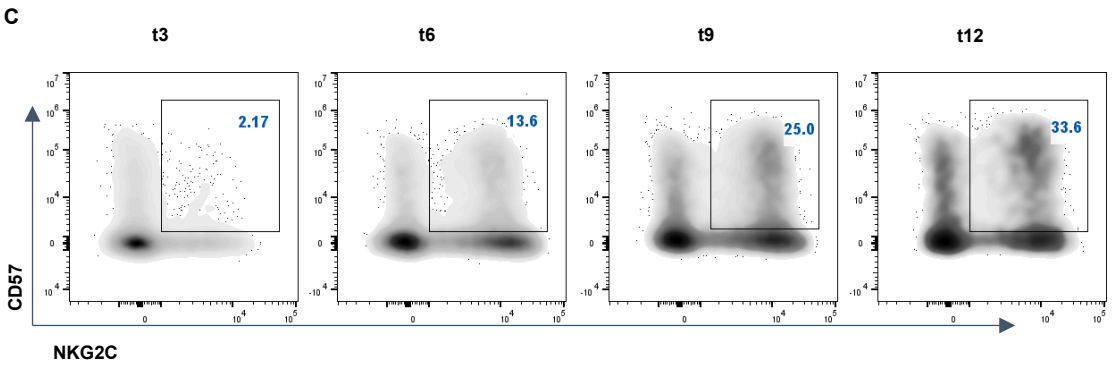

A
Recipient #11 (D-/R-)
Productive simpson clonality
0.4
0.3
0.2
0.1
0.0
100
200
300
Time in days post - transplant
B
t3
t6
t9
t12
Cumulative productive frequency (%)
1.00
0.75
0.50
0.25
0.00
Productive frequency interval
(0.005% - 0.01%]
(0.01% - 0.1%]
(0.1% - 0.5%]
(0.5% - 1%]
(1% - Inf]


C
t3
t6
t9
t12
0.95
1.73
1.68
1.74
CD57
NKG2C


D
Proportion of cells (%)
1.00
0.75
0.50
0.25
0.00
100
200
300
Time in days post - HSCT
CD56- CD16+
CD16+ CD56dim
CD57+ CD56dim
FasL+ CD56dim
GzmB+ CD56dim
NKG2A+ CD56dim
NKG2C+ CD56dim
PD1+ CD56dim
GzmB+ CD8+


b.
A
Recipient #25 (D+/R+)
Productive simpson clonality
0.4
0.3
0.2
0.1
100
200
300
Time in days post - transplant
B
t3
t6
t9
t12
Cumulative productive frequency (%)
1.00
0.75
0.50
0.25
0.00
Productive frequency interval
(0% - 0.005%]
(0.005% - 0.01%]
(0.01% - 0.1%]
(0.1% - 0.5%]
(0.5% - 1%]
(1% - Inf]
C
t3
t6
t9
t12
1.23
14.9
61.8
54.2
CD57
NKG2C
D
Proportion of cells (%)
1.00
0.75
0.50
0.25
0.00
100
200
300
Time in days post - HSCT
CD56- CD16+
CD16+ CD56dim
CD57+ CD56dim
FasL+ CD56dim
GzmB+ CD56dim
NKG2A+ CD56dim
NKG2C+ CD56dim
PD1+ CD56dim
GzmB+ CD8+


C.
A
Recipient #27 (D-/R+)
Productive simpson clonality
0.4
0.3
0.2
0.1
0.0
100
200
300
Time in days post - transplant
B
t3
t6
t9
t12
Cumulative productive frequency (%)
1.00
0.75
0.50
0.25
0.00
Productive frequency interval
(0% - 0.005%]
(0.005% - 0.01%]
(0.01% - 0.1%]
(0.1% - 0.5%]
(0.5% - 1%]
(1% - Inf]


C
t3
t6
t9
t12
2.17
13.6
25.0
33.6
CD57
NKG2C

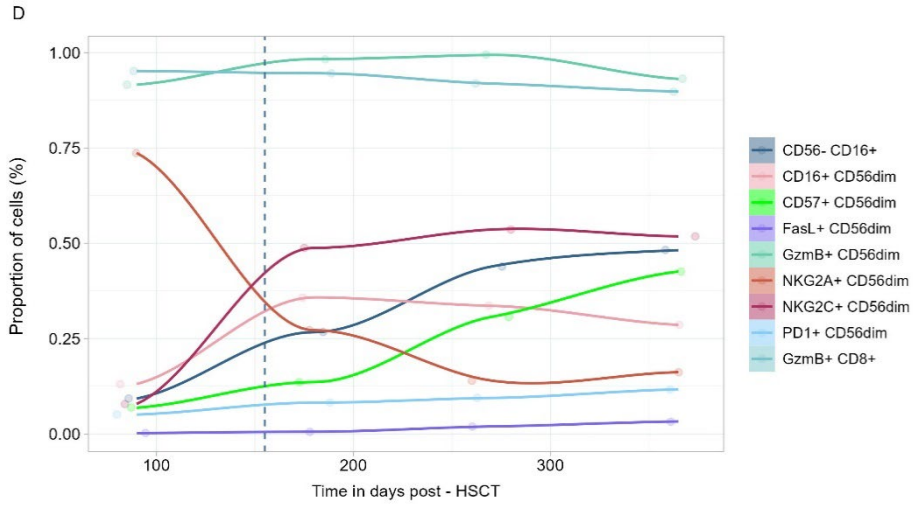

D
Proportion of cells (%)
1.00
0.75
0.50
0.25
0.00
100
200
300
Time in days post - HSCT
CD56- CD16+
CD16+ CD56dim
CD57+ CD56dim
FasL+ CD56dim
GzmB+ CD56dim
NKG2A+ CD56dim
NKG2C+ CD56dim
PD1+ CD56dim
GzmB+ CD8+

Fig. S8: Immune reconstitution of representative CMV seronegative (a.) and CMV reactivated (b.), (c.) recipients. **(A)** Evolution of the TCR repertoire post-alloHSCT based on the productive Simpson clonality (*y* axis) in a representative recipient. The vertical dotted line represents time of CMV reactivation. **(B)** TCR fractal clonal size organization defined by the productive frequency of clones at indicated timepoints. The color-coded legend bar represents the cumulative frequency of clones stratified according to the individual clone's productive frequency. **(C)** Representative flow cytometry plot of $NKG2C^{+}$ $CD57^{+}$ NK cells gated on $CD56^{dim}$ NK cells at different timepoints post-alloHSCT. **(D)** Evolution of the frequency of distinct cell subsets at different timepoints post-alloHSCT. The color-coded legend represents the distinct cell subsets.