# Genetic T-cell receptor diversity at 1 year following allogeneic hematopoietic stem cell transplantation

Buhler, Stéphane; Bettens, Florence; Dantin, Carole Alice; Ferrari-Lacraz, Sylvie; Ansari Djaberi, Marc Georges; Mamez, Anne-Claire; Masouridi Levrat, Stavroula; Chalandon, Yves; Villard, Jean

**How to cite**

BUHLER, Stéphane et al. Genetic T-cell receptor diversity at 1 year following allogeneic hematopoietic stem cell transplantation. In: Leukemia, 2020, vol. 34, n° 5, p. 1422–1432. doi: 10.1038/s41375-019-0654-y

This publication URL: https://archive-ouverte.unige.ch//unige:143931
Publication DOI: 10.1038/s41375-019-0654-y

Accepted for publication in Leukemia (13 of November 2019)

# Genetic T-cell receptor diversity at one year following allogeneic hematopoietic stem cell transplantation

Stéphane Buhler[1*], Florence Bettens[1], Carole Dantin[2], Sylvie Ferrari-Lacraz[1], Marc Ansari[3], Anne-Claire Mamez[2], Stavroula Masouridi-Levrat[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] Pediatric Department, Onco-Hematology Unit, Geneva University Hospitals and Cansearch research laboratory, Geneva Medical School, Geneva, Switzerland

Corresponding authors *:

Stéphane Buhler and Jean Villard, Transplantation Immunology Unit and National Reference Laboratory for Histocompatibility, Geneva University Hospitals, Gabrielle-Perret-Gentil 4, 1211 Geneva 4, +41 22 372 9401 / +41 79 553 3595, stephane.buhler@hcuge.ch; jean.villard@hcuge.ch

Short title: T-cell reconstitution at one year post-HSCT

Key words: immune reconstitution, TCR, haematopoietic stem cell transplantation, cytomegalovirus, next-generation sequencing

Text word count: 4 000, abstract word count: 183, number of figures: 4, number of tables: 3, number of references: 60.

The authors declare no competing interests

## Abstract

After allogeneic hematopoietic stem cell transplantation (HSCT), immune reconstitution leads to the development of a new T-cell repertoire. Immune reconstitution could be influenced by events such as conditioning, infections and graft versus host disease (GVHD). Factors influencing the TCR diversity are of great interest to fine-tune the strategy for donor selection and to optimize standard of care. In this work, immunosequencing of the TCR CDR3β region was carried out in a large cohort of 116 full chimeric recipients at one year post-HSCT and their respective donors prior to transplantation. The repertoire overlap before and after HSCT was minimal, supporting de novo reconstitution as a primary pathway at any age. Among the parameters investigated, increased patient and/or donor age as well as positive CMV serologic status reinforced by CMV infection/reactivation were the ones significantly associated with a reduced diversity at one year post-HSCT. CMV-specific T-cell clones were shown to influence the clonality of the repertoire alongside the expansion of limited numbers of non-CMV T-cell populations. Interestingly, at the exception of CMV infection/reactivation, TCR diversity was not predictive of GVHD, relapse, death or infections post-HSCT.

## Introduction

Allogeneic hematopoietic stem cell transplantation (alloHSCT) is a standard treatment of hematologic disorders such as leukaemia and primary immunodeficiencies. The polymorphism of human leukocyte antigen (HLA) genes is a major factor for the global outcome, notably for avoiding graft rejection and for minimizing the risk of relapse and development of severe graft versus host disease (GVHD) (1-3). HLA is also instrumental in the processes leading to immune reconstitution, especially the T-cell repertoire diversity, considered a key factor for prognosis and long term survival. Homeostatic proliferation, which controls the size of the T-cell pool during the reconstitution of the T-cell compartment (4), is mediated by the presentation of antigens by HLA molecules to T cells. Direct presentation in case of HLA mismatch, indirect presentation of alloantigens (minor histocompatibility antigens) or pathogen peptides, (e.g. derived from latent viruses present in the recipient) drive a cytokine-mediated expansion of the mature donor T cells infused with the graft (5, 6). This pathway, which is independent of the thymus, result in a skewed repertoire closely associated with infections and GVHD (7-9). By contrast, de novo maturation of naïve T cells derived from lymphoid precursors of the donor and selected in the thymus by the self-HLA molecules presenting self-peptides will restore a broad and fully responsive repertoire. HLA has been suggested to bias the T-cell receptor (TCR) V gene usage of maturing thymocytes (10). The thymus is highly sensitive to conditioning, immunosuppression and GVHD and is significantly influenced by age. Age is associated with thymic involution, impairing the renewed thymopoiesis (11-15). It has been suggested that thymus-dependent reconstitution take months following HSCT, leaving the patients at risk of infections and other complications. However, it is not well understood whether T-cell diversity is directly predictive of these clinical events.

Pre, peri and post-transplant factors influencing the T-cell repertoire have been investigated leading to an improvement of the standard of care (16-18). Recent technological developments have also provided the capacity to analyse more precisely and in-depth the complex processes of reconstitution, offering new perspectives for the post-HSCT follow-up. On the one hand, system-level profiling is now possible with mass cytometry, allowing comprehensive monitoring of immune reconstitution (19, 20). On the other hand, the central role and dynamics of T cells reconstitution can be investigated by high throughput immunosequencing (21-23), providing a more detailed snapshot of the TCR diversity compared to methodologies like spectratyping (24, 25). Current protocols target the most variable complementary-determining region of the somatically rearranged TCR alpha and beta chains (i.e., CDR3 at the V-J or V-D-J junctions, respectively), allowing to track down single clones among thousand T cells sequenced in parallel. A limited number of studies have been published using this technology after alloHSCT. These studies are characterised by the inclusion of a small number of patients and a variety of factors which could influence the T-cell repertoire reconstitution. One study reported a reduction of the TCR diversity following GVHD or disease relapse (26). In another study, a better recovery of the T-cell diversity was reported in recipients of double cord blood units compared to conventional or T depleted peripheral blood stem cells (PBSC) (27). Acute GVHD and steroid treatment were associated with a high diversity in contrast to cytomegalovirus (CMV) and Epstein Barr virus (EBV) infections (27). A role of CMV reactivation generating holes in the underlying repertoire was recently proposed with a massive expansion of effector memory ($T_{EM}$) cells and contraction of naïve subsets post-HSCT (28). Antithymocyte therapy has been reported by some authors to lead to lower TCR diversity in recipients of CMV-positive donors, while other variables (e.g., GVHD and CMV reactivation) were not predictive of the diversity (29). By contrast, a

reduction of diversity was observed with GVHD but not with conditioning by another group (30).

In this work, immunosequencing of the TCR CDR3β region was carried out in a large cohort of 116 full chimeric recipients at one-year post-HSCT and their respective HSC donors. The first aim was to investigate repertoire overlap before and after HSCT. Within the second aim, parameters possibly associated with the reconstitution of the repertoire at one year were analyzed. The third aim was to assess whether T-cell diversity could be used as a marker to predict post-transplant complications.

## Material and Methods

### Patients and donors

Patients receiving an allogeneic HSC graft in Geneva between 2000 and 2016, who had full donor chimerism without sign of relapse at one year were selected. The cohort consisted of 116 donor/recipient pairs with characteristics described in Table 1. Post-HSCT complications were recorded during the first year and also after that period (Table 2). A standard immunosuppressive treatment consisting of methotrexate or mycophenolate combined with cyclosporine A or tacrolimus was provided to all patients. Partial T-cell depletion is sometimes included in the institution protocol and was considered as an explanatory variable in the statistical analyses (31). This study was approved by the ethical committee of the institution (CER 06-208 and 08-208R).

### HLA typing and chimerism

DNA extracted on an automatic system (QIAGEN GmbH, Hilden, Germany) from Ficoll purified peripheral blood mononuclear cells (PBMCs) was obtained from donors shortly

before transplantation (time point 1) and from their full chimeric recipients at one-year post-HSCT (time point 2). HLA typing was performed by reverse PCR-sequence-specific oligonucleotide microbead arrays and high throughput sequencing (One Lambda, Canoga Park, CA, USA) or PCR-sequence-specific primers (Genovision, Milan Analytika AG, Switzerland). Chimerism was performed by STR analysis (AmpFlSTR® Identifiler, Invitrogen-Thermofisher, Waltham MA, USA), the detection sensitivity is <3% (i.e., patients were selected for the study if the donor chimerism was ≥97%).

**Immunosequencing**

High throughput sequencing of the TCR CDR3β region was carried out on Illumina MiSeq and HiSeq systems following a multiplex PCR (Adaptive Biotechnologies ImmunoSEQ© assay) (32, 33). Donor/recipient pairs were analyzed at survey resolution targeting 120 000 T cells. Reproducibility and sampling performance was assessed in five selected pairs using triplicates analyzed at deep resolution (i.e., targeting 400 000 T cells). Productive rearrangements were retrieved from the ImmunoSEQ© analyzer platform and formatted for the analyses to be carried out in R with the help of GNU/Linux scripts. The counts of clones with CDR3 rearrangements sharing synonymous nucleotide substitutions (i.e., an identical amino acid sequence) were pooled.

**Peptide binding predictions**

The FASTA sequences of the 190 canonical proteins of CMV strain AD169 were downloaded from https://www.uniprot.org/. This is one reference proteome for CMV which includes manual annotations and Swiss-Prot reviewing. These data were then submitted to the NetMHCpan 4.0 server available at http://www.cbs.dtu.dk/services/NetMHCpan/ to

perform HLA class I binding predictions (34). The predictions were performed on all possible AD169 nonamers (i.e. 9mer represent the preferred length of peptides bound by HLA class I molecules) and with each of the 91 alleles observed in our patients at high resolution (i.e. second field (35)). The binding predictions performed on individual alleles were combined according to the HLA types in order to estimate the theoretical capacity of each patient to present CMV derived peptides. We considered the total number of strong and weak binders using affinity ranks <0.5% or >0.5% and <2%, respectively. Mean peptide binding affinity in nanoMolar units was also estimated in each patient over retained 9mer binders. Association with CMV infection/reactivation within the first year following transplantation was analysed through logistic regression.

**CMV-specific clones**

T cells were characterized as potentially specific for CMV if their CDR3β sequence matched one of the 164 clones identified by Emerson et al. as CMV-associated (i.e., clones with significant enrichment in CMV positive subjects) and/or one of the 919 clones reported as CMV-reactive (i.e., clones able to recognize CMV antigens) (36). Both sets are partially overlapping, the second one being particularly biased toward reactivity to 65 kDa phosphoprotein (pp65) and 55 kDa immediate-early protein 1 (IE1). HLA restriction at low resolution (i.e., first field level (35)) is proposed for about half of these clones, we thus checked if we could find a concordance with HLA types in patients.

**Statistical analyses**

Clone's frequencies were estimated from the number of sequenced templates and used to describe the commonness/rareness of given T cells or of groups of T cells (e.g. CMV-specific

clones). Overlap of the TCR repertoire before and after HSCT was estimated by Jaccard and Morisita's similarity indices (37, 38). These indices vary between 0 (no overlap) and 1 (complete overlap). The standard Jaccard index gauges similarity using the ratio of shared clones at both time points on the total number of clones, the standardized index additionally weights the similarity using clone's frequencies and the Morisita's index is based on statistical dispersion of clones assuming that diversity increases with sample size. The diversity of the TCR repertoire at both time points (i.e. pre and post-HSCT) was estimated by productive clonality. This index is robust to sampling and accounts both for the richness and eveness of the repertoire and ranges between 0 (polyclonal/diverse) and 1 (monoclonal/invariant). The difference of clonality between both time points was considered as a surrogate for repertoire reconstitution during the first year. Graphical inspection of the data and univariate linear modelling were used to identify independent variables significantly associated with reconstitution. Post hoc tests (Tukey HSD) were applied to determine group(s) with a different mean. Multivariable models were subsequently explored and validated with diagnostic plots of the residuals. We also assessed whether donor parameters were associated with the clonality of the repertoire infused with the graft. Furthermore, logistic regressions were performed considering clonality as a possible explanatory variable for post-transplant complications within or after the first year, based on the repertoire of the donor or recipient, respectively. Additional analyses are detailed in supplementary figure legends where appropriate. All the analyses were performed in R (version 3.5.0) using the packages ggplot2, reshape2, tcR, GGally, scales and party.

## Results

**Repertoire overlap**

Immunosequencing yielded a total of 3 582 584 private and public CDR3β clones with variable frequencies pre and post-HSCT, including 2 894 321 unique rearrangements (i.e., meaning that almost 81% of the rearrangements were private to a given donor/recipient pair). The repertoire overlap at both time points was low according to the three similarity indices, but slightly less in some pairs for Morisita's index (Figures 1a, 1b and S1). Mean (±SD) values obtained were 0.017 (±0.014) for standard Jaccard, 0.026 (±0.025) for generalized Jaccard and 0.13 (±0.16) for Morisita's index, respectively, with significantly correlated distributions of values (Figure 1S). Deep resolution sequencing provided a very good concordance with survey resolution and high reproductibility between replicates (not shown). We also verified that the number of productive templates sequenced before and after alloHSCT was not correlated with the indices of similarity.

**Parameters influencing repertoire reconstitution**

Productive clonality exhibited a skewed pattern with a shift from a mostly polyclonal repertoire pre-HSCT to more predominant oligoclonal profiles post-HSCT (paired t-test $p<2.2e-16$). Several parameters were significantly associated with this change of clonality according to the univariate analyses (Table S1). The reduction of diversity (i.e. equal to an increased clonality), was lower in young recipients (≤ 20 years old, Figure 2a) or when the graft was infused from a young donor (≤ 30 years old, Figure 2b), but only according to these discrete categories (Table S1). The CMV serologic status and CMV infection/reactivation (defined as CMV DNA in plasma above the limit of detection, currently 2.1E+1 UI/ml, in patients with or without clinical symptoms) were both significantly associated with a reduced diversity post-HSCT (Figure 2c). Specifically, the CMV positive donor/recipient

(D+/R+) group differed from the other groups according to Tukey HSD. Regarding CMV infection/reactivation, it significantly reduced the repertoire diversity in all groups (no observation in the D-/R- group). Moreover, the significant association was mainly driven by CMV, although a trend was observed for other viruses (Figure S2). Conditioning, T-cell depletion, source of stem cells, donor lymphocyte infusions (DLI), acute and chronic GVHD, relapse and other infections (e.g., bacterial) were not significant. Age and CMV-related variables were included in multivariable models and were significant taken two-by-two (Table 3), but without interaction (not shown). The strong influence of CMV serologic status and CMV infection/reactivation was also observed using conditional inference framework analysis (Figure S3). Looking at the repertoire infused with the graft, a significant effect of donor's age and CMV status was found. Both variables were interacting with a shift toward oligoclonality detected in the group of CMV positive donors aged >30 years (Table S2 and Figure S7a).

**Impact of cytomegalovirus on repertoire reconstitution**

*CMV-specific clones*

A total of 1 978 CDR3β clones were defined as specific for CMV in different donor/recipient pairs, including 299 with a unique rearrangement. Thus, many of these rearrangements were public, including one observed in 59 pairs. In addition, 10.5% of the CMV-specific clones were shared in the cohort before and after alloHSCT compared to only 1.6% of the non-CMV clones, representing a drastic increase of overlap (Fisher's exact test $p<2.2e-16$). Among the 299 rearrangements, 167 were described with an HLA restriction (36). Cross-tabulating this information with HLA types in patients, we could observe the allele corresponding to the proposed restriction in 43% of the cases supporting the assignation of these clones as

specific for CMV. Indeed, the number of HLA alleles, the flexibility of possible interactions at the TCR-peptide-HLA interface and the randomness of TCR rearrangements in distinct individuals makes this concordance highly improbable just by chance (39-41). The number and cumulated frequency of CMV-specific clones are reported in Figure 3, according to CMV serologic status and infection/reactivation post-HSCT. An increase above a frequency threshold of 1/1000 post-HSCT was only observed in the D+/R+ group when no infection/reactivation occurred. By contrast, all groups (except D-/R-, no observation) exhibited increased frequencies in case of infection/reactivation. We also investigated the relationship between the cumulative frequency of CMV-specific clones and clonality and found a significant correlation (Pearson's r=0.249, p=0.007; Spearman's rho=0.318, p=0.0005, Figure S4a). This was confirmed by a powerful resampling approach (Figure S4b). Finally, we demonstrated that although the frequency of CMV specific clones increases according to CMV status, such clones were never dominant in the whole repertoire of the donors and patients (Figure 4).

*CMV peptide binding predictions and infection/reactivation*

A total of 64 054 9mer were derived from the proteins of AD169 and submitted to NetMHCpan. Among this large number of possible 9mer, close to half of them (29 054, 45.4%) were predicted as binders to one or several HLA class I alleles (5.2 and 6.6 alleles on average for strong and weak 9mer binders, respectively). On average, HLA-A, B, and C alleles were predicted to bind strongly (or weakly) 809 (1 587), 784 (1 693) and 1 358 (3 134) peptides, respectively. Thus, considering HLA class I conjointly, every patient could theoretically cover a broad spectrum of CMV derived peptides (Figures S5). Furthermore, no significant association with CMV infection/reactivation was observed (Figure S6).

**Is T-cell diversity predictive of post-transplant complications?**

Overall, clonality was not associated with an increased risk of clinical events within and after the first year post-alloHSCT (Table S2 and S3), excepted that the risk of CMV infection/reactivation slightly increased with grafts from donors with a lower TCR diversity (Table S2 and Figure S7b).

## Discussion

Using immunosequencing we could describe the reconstitution of the T-cell repertoire diversity at one year after alloHSCT in a cohort of 116 full donor chimeric patients. Although the transplant infused usually includes a large amount of T cells, our data show that the repertoire one year after the procedure is very different, with only a few overlaps. This strongly suggests that a new repertoire can be reconstituted at any age through thymic dependent or independent pathways (9, 17, 18). TCR monitoring by immunosequencing is very powerful, but it only provides a snapshot of the repertoire and the technology can be challenging in terms of analyses (37, 38). However, with three indices of similarity going in the same direction and with a serie of controls performed at deep resolution to exclude sample size issues, we are pretty confident that our data are robust.

The diversity of the repertoire has been correlated to many clinical factors, however in this study, only three of them are significantly associated with clonality: age of the patient, age of the donor and CMV. Interestingly, GVHD which has previously been associated with repertoire diversity in some reports (26, 30) was not significant in our cohort. Yet, a tendency was observed between acute GVHD and CMV infection/reactivation, although not

significant, indicating that a more robust immunosuppressive treatment in patients with higher grade acute GVHD led to a higher prevalence of CMV infection/reactivation (p=0.063). The association between acute GVHD and post-HSCT infections, especially CMV, is well described (8, 42, 43), the risk being dependent on the dose of steroid administration (44) and the GVHD grade (45).

An impact of the age of recipients on clonality was expected. Previous reports have already demonstrated similar findings which reflect the capacity of the thymus of the recipient to generate a more diverse repertoire at a younger age (12, 13, 15, 46, 47). The contribution of young donors is also significant to explain a more diverse repertoire post-HSCT in our cohort. It has been previously proposed that advanced donor age could delay immune recovery (48), possibly because of the decreased frequency of pre-thymic T-cell progenitors within the graft. In addition, the repertoire diversity is higher with lymphoid progenitors from cord blood compared to bone marrow donors (49). Similarly, the proliferation of peripheral T cells from young donors could be more efficient than from older donors. However we would expect a better overlap of the repertoire before and after T-cell reconstitution. We thus speculate that T-cell precursors coming from young donors associated with the thymic independent pathway (50, 51) best explain these results. The importance of donor age is of great interest in the strategy of finding the best donor for alloHSCT patients, e.g., what is the best option between an old haploidentical donor or an HLA matched or mismatched young unrelated donor? T-cell repertoire diversity at one year could be a parameter to take into consideration in this context. The third factor significantly influencing the repertoire reconstitution is the CMV serologic status of the recipient and the donor reinforced by CMV infection/reactivation. In the D-/R- group, no primo infection was detected and the T-cell diversity is higher (less oligoclonal) compared to the other groups. The D+/R- status is also

less associated with oligoclonality. In this later group, the rate of infection/reactivation is low, as expected, with only two reported cases. The reservoir of CMV is coming with the donor cells and could be controlled by the donor-specific T cells infused with the graft (i.e., only 6 of the 116 alloHSCT analyzed in this study were with 100% depleted T cells). The D-/R+ group is more prone to develop oligoclonality in case of infection/reactivation. This could be explained either by a) residual recipient CMV-specific T cells (like tissue resident memory cells) which have escaped the conditioning regimen and proliferate, b) by the primary response of donor specific T cells, c) by other crossreactive donor T cells or d) by the proliferation of non specific T cells induced as a bystander effect of inflammation. The D+/R+ group, with and without CMV infection/reactivation, exhibits the strongest association with oligoclonality. In this group, the reservoir of latent CMV of the recipient and donor can stimulate donor CMV-specific T cells infused with the graft. Recipient CMV positive serostatus has been proposed as the main factor determining CMV infection post-HSCT (52) and CMV-specific cytotoxic T cells were observed more frequently in D+/R+ grafts (53). To validate these explanations, we analysed the presence of CMV-specific T cells and confirmed that their frequencies were mirroring the change of clonality observed in the different groups of our cohort depending on infection/reactivation. Interestingly, all patients have a good capacity to present CMV derived peptides according to their HLA class I restriction. Therefore, the presence of T cells able to respond broadly to CMV is expected in every individual. In agreement with our data, T-cell response to CMV is substantial and directed toward multiple antigens (54, 55). CMV is the largest among known human viruses against which up to 10% of the CD4 and CD8 memory compartments are committed in the blood of seropositive subjects (55-57). By comparing CMV-specific clones and non-CMV clones in our cohort, we could also show that the repertoire overlap is significantly driven by

cytomegalovirus but otherwise remains very limited before and after alloHSCT. Interestingly, in our cohort the combined frequencies of CMV-specific clones were well below 10% and none of them were found among clones described as dominant (Figure 4). Dominant clones were observed more often in recipients with severe (i.e., ≥ grade 2) acute GVHD ($p=0.01$) and were especially overrepresented in the D+/R+ group ($p<2.5e-10$), but were not significantly associated with CMV infection/reactivation. Some of these clones can be truly CMV-specific and are not yet described in the current database. Alternatively these clones could be the signature of concomitant viral infections, indeed we found a trend, although not significant, of the impact of other infections on clonality. Finally, the presence of oligoclonality due to the presence of non-CMV dominant clones could be explained by the homeostatic proliferation of nonspecific T cells triggered by cytokines secreted during the symptomatic or asymptomatic anti-CMV immune response, especially in the D+/R+ group. We speculate that these nonspecific clones could be recent thymic emigrant or mature naïve T cells which proliferate very efficiently (58), much more than memory anti-CMV specific T-cell clones. Moreover, severe GVHD could also compound the reduction of diversity alongside CMV by promoting the expansion of alloreactive clones. A similar observation was made in a recent study based on TCRα diversity where dominant clones not observed before transplantation were found in patients suffering from CMV reactivation or extensive chronic GVHD (29). In our study, in non-infected patients the repertoire is more diverse and a weak overlap is observed before and after transplantation. This suggests that at one-year post-HSCT the repertoire is mainly composed of new T cells emerging from thymic dependent or independent pathways. This is also true but to a lesser degree in patients with CMV infection/reactivation as discussed above. Unfortunately, we do not have information about

the naïve or memory phenotype of these T cells, but previous reports have already established that at one year the T-cell populations are mainly naïve (CD45RA+) (17, 18).

It is important to stress some limitations of our study. One drawback is that we could not differentiate CD4 and CD8 subpopulations because we did not have enough cells to perform cell sorting. For instance, a durable and significant imprint of CMV on T-cell reconstitution specifically in shaping the CD8+ memory T-cell compartment has been described (28, 59, 60).

The diversity of the T-cell repertoire at one year was interestingly not predictive of any subsequent clinical event such as infection (including CMV), GVHD, relapse or mortality. This strongly suggests that clinical management remains the key factor to prevent and treat any event post-transplantation.

The analysis of the donor T-cell repertoire was also instructive as it revealed that diversity was not significantly associated with GVHD, relapse or any infection except CMV infection/reactivation. This suggests that either grafts with reduced T cell diversity may lack CMV protective T cells and thus increase the risk of infection, or more probably that CMV positive donors (i.e. with a more oligoclonal repertoire) are overrepresented in CMV positive recipients, the more at risk of developing an infection/reactivation. Our data also suggest that despite being significantly shifted in some donors and in many recipients, the repertoire infused with the graft, although transient, was diverse enough to afford protection in the early reconstitution phase and was then reconstituted to a sufficient extent at one year.

In conclusion, our study demonstrates the weak repertoire overlap before and at one year after alloHSCT. Age of the patient and the donor play a significant role. Reduced diversity at one year is mainly associated with CMV serostatus and infection/reactivation. It is to note

that although CMV-specific clones are central to the observed shift of clonality they never predominate in the repertoire.

## Acknowledgements

This study was supported by the Swiss National Science Foundation (grant #310030_173237/1), the Academic Society of the University of Geneva, IRGHET (International Research Group on unrelated Hematopoietic stem cell Transplantation), the Dr Henri Dubois-Ferrière Dinu Lippatti foundation and the Philantropy Settlement.

The authors are grateful to Lydie Brunet for her technical expertise and to the technicians of the LNRH for their most efficient support for HLA typing. We would also like to thanks Dr. José Manuel Nunes (Biosc-iGE3) for his valuable suggestions on data analysis using R.

## Disclosure of Conflicts of Interest

None to disclose

## References


1. Petersdorf EW. Genetics of graft-versus-host disease: the major histocompatibility complex. Blood reviews. 2013;27(1):1-12.

2. Fleischhauer K, Beelen DW. HLA mismatching as a strategy to reduce relapse after alternative donor transplantation. Seminars in hematology. 2016;53(2):57-64.

3. Fleischhauer K, Shaw BE. HLA-DP in unrelated hematopoietic cell transplantation revisited: challenges and opportunities. Blood. 2017;130(9):1089-96.

4. Jameson SC. Maintaining the norm: T-cell homeostasis. Nat Rev Immunol. 2002;2(8):547-56.

5. Degauque N, Brouard S, Soulillou JP. Cross-Reactivity of TCR Repertoire: Current Concepts, Challenges, and Implication for Allotransplantation. Frontiers in immunology. 2016;7:89.

6. DeWolf S, Sykes M. Alloimmune T cells in transplantation. The Journal of clinical investigation. 2017;127(7):2473-81.

7. Inman CF, Eldershaw SA, Croudace JE, Davies NJ, Sharma-Oates A, Rai T, et al. Unique features and clinical importance of acute alloreactive immune responses. JCI Insight. 2018;3(10).

8. Phan TL, Pritchett JC, Leifer C, Zerr DM, Koelle DM, Di Luca D, et al. HHV-6B infection, T-cell reconstitution, and graft-vs-host disease after hematopoietic stem cell transplantation. Bone marrow transplantation. 2018;53(12):1508-17.

9. Mackall CL, Bare CV, Granger LA, Sharrow SO, Titus JA, Gress RE. Thymic-independent T cell regeneration occurs via antigen-driven expansion of peripheral T cells resulting in a repertoire that is limited in diversity and prone to skewing. Journal of immunology. 1996;156(12):4609-16.

10. Sharon E, Sibener LV, Battle A, Fraser HB, Garcia KC, Pritchard JK. Genetic variation in MHC proteins is associated with T cell receptor expression biases. Nature genetics. 2016.

11. Chaudhry MS, Velardi E, Malard F, van den Brink MR. Immune Reconstitution after Allogeneic Hematopoietic Stem Cell Transplantation: Time To T Up the Thymus. Journal of immunology. 2017;198(1):40-6.

12. Clave E, Busson M, Douay C, Peffault de Latour R, Berrou J, Rabian C, et al. Acute graft-versus-host disease transiently impairs thymic output in young patients after allogeneic hematopoietic stem cell transplantation. Blood. 2009;113(25):6477-84.

13. Hakim FT, Memon SA, Cepeda R, Jones EC, Chow CK, Kasten-Sportes C, et al. Age-dependent incidence, time course, and consequences of thymic renewal in adults. The Journal of clinical investigation. 2005;115(4):930-9.

14. Lynch HE, Goldberg GL, Chidgey A, Van den Brink MR, Boyd R, Sempowski GD. Thymic involution and immune reconstitution. Trends Immunol. 2009;30(7):366-73.

15. Weinberg K, Blazar BR, Wagner JE, Agura E, Hill BJ, Smogorzewska M, et al. Factors affecting thymic function after allogeneic hematopoietic stem cell transplantation. Blood. 2001;97(5):1458-66.

16. Bosch M, Khan FM, Storek J. Immune reconstitution after hematopoietic cell transplantation. Current opinion in hematology. 2012;19(4):324-35.

17. Ogonek J, Kralj Juric M, Ghimire S, Varanasi PR, Holler E, Greinix H, et al. Immune Reconstitution after Allogeneic Hematopoietic Stem Cell Transplantation. Frontiers in immunology. 2016;7:507.

18. Storek J, Geddes M, Khan F, Huard B, Helg C, Chalandon Y, et al. Reconstitution of the immune system after hematopoietic stem cell transplantation in humans. Semin Immunopathol. 2008;30(4):425-37.

19. Stern L, McGuire H, Avdic S, Rizzetto S, Fazekas de St Groth B, Luciani F, et al. Mass Cytometry for the Assessment of Immune Reconstitution After Hematopoietic Stem Cell Transplantation. Frontiers in immunology. 2018;9:1672.

20. Lakshmikanth T, Olin A, Chen Y, Mikes J, Fredlund E, Remberger M, et al. Mass Cytometry and Topological Data Analysis Reveal Immune Parameters Associated with Complications after Allogeneic Stem Cell Transplantation. Cell Rep. 2017;20(9):2238-50.

21. Heather JM, Ismail M, Oakes T, Chain B. High-throughput sequencing of the T-cell receptor repertoire: pitfalls and opportunities. Brief Bioinform. 2018;19(4):554-65.

22. Robins H. Immunosequencing: applications of immune repertoire deep sequencing. Current opinion in immunology. 2013;25(5):646-52.

23. Six A, Mariotti-Ferrandiz ME, Chaara W, Magadan S, Pham HP, Lefranc MP, et al. The past, present, and future of immune repertoire biology - the rise of next-generation repertoire analysis. Frontiers in immunology. 2013;4:413.

24. Cochet M, Pannetier C, Regnault A, Darche S, Leclerc C, Kourilsky P. Molecular detection and in vivo analysis of the specific T cell response to a protein antigen. European journal of immunology. 1992;22(10):2639-47.

25. Gorski J, Yassai M, Zhu X, Kissela B, Kissella B, Keever C, et al. Circulating T cell repertoire complexity in normal individuals and bone marrow recipients analyzed by CDR3 size spectratyping. Correlation with immune status. Journal of immunology. 1994;152(10):5109-19.

26. Yew PY, Alachkar H, Yamaguchi R, Kiyotani K, Fang H, Yap KL, et al. Quantitative characterization of T-cell repertoire in allogeneic hematopoietic stem cell transplant recipients. Bone marrow transplantation. 2015;50(9):1227-34.

27. van Heijst JW, Ceberio I, Lipuma LB, Samilo DW, Wasilewski GD, Gonzales AM, et al. Quantitative assessment of T cell repertoire recovery after hematopoietic stem cell transplantation. Nat Med. 2013;19(3):372-7.

28. Suessmuth Y, Mukherjee R, Watkins B, Koura DT, Finstermeier K, Desmarais C, et al. CMV reactivation drives posttransplant T-cell reconstitution and results in defects in the underlying TCRbeta repertoire. Blood. 2015;125(25):3835-50.

29. Link-Rachner CS, Eugster A, Rucker-Braun E, Heidenreich F, Oelschlagel U, Dahl A, et al. T cell receptor alpha repertoire of CD8+ T cells following allogeneic stem cell transplantation using next-generation sequencing. Haematologica. 2018.

30. Gkazi AS, Margetts BK, Attenborough T, Mhaldien L, Standing JF, Oakes T, et al. Clinical T Cell Receptor Repertoire Deep Sequencing and Analysis: An Application to Monitor Immune Reconstitution Following Cord Blood Transplantation. Frontiers in immunology. 2018;9:2547.

31. Chalandon Y, Roosnek E, Mermillod B, Waelchli L, Helg C, Chapuis B. Can only partial T-cell depletion of the graft before hematopoietic stem cell transplantation mitigate graft-versus-host disease while preserving a graft-versus-leukemia reaction? A prospective phase II study. Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation. 2006;12(1):102-10.

32. Carlson CS, Emerson RO, Sherwood AM, Desmarais C, Chung MW, Parsons JM, et al. Using synthetic templates to design an unbiased multiplex PCR assay. Nature communications. 2013;4:2680.

33. Robins HS, Campregher PV, Srivastava SK, Wacher A, Turtle CJ, Kahsai O, et al. Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. Blood. 2009;114(19):4099-107.

34. Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. Journal of immunology. 2017;199(9):3360-8.

35. Nunes E, Heslop H, Fernandez-Vina M, Taves C, Wagenknecht DR, Eisenbrey AB, et al. Definitions of histocompatibility typing terms. Blood. 2011;118(23):e180-3.

36. Emerson RO, DeWitt WS, Vignali M, Gravley J, Hu JK, Osborne EJ, et al. Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. Nature genetics. 2017;49(5):659-65.

37. Laydon DJ, Bangham CR, Asquith B. Estimating T-cell repertoire diversity: limitations of classical estimators and a new approach. Philos Trans R Soc Lond B Biol Sci. 2015;370(1675).

38. Rempala GA, Seweryn M. Methods for diversity and overlap analysis in T-cell receptor populations. J Math Biol. 2013;67(6-7):1339-68.

39. Bentzen AK, Such L, Jensen KK, Marquard AM, Jessen LE, Miller NJ, et al. T cell receptor fingerprinting enables in-depth characterization of the interactions governing recognition of peptide-MHC complexes. Nat Biotechnol. 2018.

40. Birnbaum ME, Mendoza JL, Sethi DK, Dong S, Glanville J, Dobbins J, et al. Deconstructing the peptide-MHC specificity of T cell recognition. Cell. 2014;157(5):1073-87.

41. Illing PT, Pymm P, Croft NP, Hilton HG, Jojic V, Han AS, et al. HLA-B57 micropolymorphism defines the sequence and conformational breadth of the immunopeptidome. Nature communications. 2018;9(1):4693.

42. Ljungman P, Perez-Bercoff L, Jonsson J, Avetisyan G, Sparrelid E, Aschan J, et al. Risk factors for the development of cytomegalovirus disease after allogeneic stem cell transplantation. Haematologica. 2006;91(1):78-83.

43. Miller HK, Braun TM, Stillwell T, Harris AC, Choi S, Connelly J, et al. Infectious Risk after Allogeneic Hematopoietic Cell Transplantation Complicated by Acute Graft-versus-Host Disease. Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation. 2017;23(3):522-8.

44. Hakki M, Riddell SR, Storek J, Carter RA, Stevens-Ayers T, Sudour P, et al. Immune reconstitution to cytomegalovirus after allogeneic hematopoietic stem cell transplantation: impact of host factors, drug therapy, and subclinical reactivation. Blood. 2003;102(8):3060-7.

45. Cantoni N, Hirsch HH, Khanna N, Gerull S, Buser A, Bucher C, et al. Evidence for a bidirectional relationship between cytomegalovirus replication and acute graft-versus-host disease. Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation. 2010;16(9):1309-14.

46. Storek J, Joseph A, Dawson MA, Douek DC, Storer B, Maloney DG. Factors influencing T-lymphopoiesis after allogeneic hematopoietic cell transplantation. Transplantation. 2002;73(7):1154-8.

47. Toubert A, Glauzy S, Douay C, Clave E. Thymus and immune reconstitution after allogeneic hematopoietic stem cell transplantation in humans: never say never again. Tissue antigens. 2012;79(2):83-9.

48. Baron F, Storer B, Maris MB, Storek J, Piette F, Metcalf M, et al. Unrelated donor status and high donor age independently affect immunologic recovery after nonmyeloablative conditioning. Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation. 2006;12(11):1176-87.

49. Talvensaari K, Clave E, Douay C, Rabian C, Garderet L, Busson M, et al. A broad T-cell repertoire diversity and an efficient thymic function indicate a favorable long-term immune reconstitution after cord blood stem cell transplantation. Blood. 2002;99(4):1458-64.

50. Fusco A, Panico L, Gorrese M, Bianchino G, Barone MV, Grieco V, et al. Molecular evidence for a thymus-independent partial T cell development in a FOXN1-/- athymic human fetus. PLoS One. 2013;8(12):e81786.

51. McClory S, Hughes T, Freud AG, Briercheck EL, Martin C, Trimboli AJ, et al. Evidence for a stepwise program of extrathymic T cell development within the human tonsil. The Journal of clinical investigation. 2012;122(4):1403-15.

52. Melendez-Munoz R, Marchalik R, Jerussi T, Dimitrova D, Nussenblatt V, Beri A, et al. Cytomegalovirus Infection Incidence and Risk Factors Across Diverse Hematopoietic Cell Transplantation Platforms Using a Standardized Monitoring and Treatment Approach: A Comprehensive Evaluation from a Single Institution. Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation. 2018.

53. Ogonek J, Varanasi P, Luther S, Schweier P, Kuhnau W, Gohring G, et al. Possible Impact of Cytomegalovirus-Specific CD8(+) T Cells on Immune Reconstitution and Conversion to Complete Donor Chimerism after Allogeneic Stem Cell Transplantation. Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation. 2017;23(7):1046-53.

54. Elkington R, Walker S, Crough T, Menzies M, Tellam J, Bharadwaj M, et al. Ex vivo profiling of CD8+-T-cell responses to human cytomegalovirus reveals broad and multispecific reactivities in healthy virus carriers. J Virol. 2003;77(9):5226-40.

55. Sylwester AW, Mitchell BL, Edgar JB, Taormina C, Pelte C, Ruchti F, et al. Broadly targeted human cytomegalovirus-specific CD4+ and CD8+ T cells dominate the memory compartments of exposed subjects. J Exp Med. 2005;202(5):673-85.

56. Camargo JF, Komanduri KV. Emerging concepts in cytomegalovirus infection following hematopoietic stem cell transplantation. Hematol Oncol Stem Cell Ther. 2017;10(4):233-8.

57. Boeckh M, Geballe AP. Cytomegalovirus: pathogen, paradigm, and puzzle. The Journal of clinical investigation. 2011;121(5):1673-80.

58. Fink PJ, Hendricks DW. Post-thymic maturation: young T cells assert their individuality. Nat Rev Immunol. 2011;11(8):544-9.

59. Itzykson R, Robin M, Moins-Teisserenc H, Delord M, Busson M, Xhaard A, et al. Cytomegalovirus shapes long-term immune reconstitution after allogeneic stem cell transplantation. Haematologica. 2015;100(1):114-23.

60. Lugthart G, van Ostaijen-Ten Dam MM, Jol-van der Zijde CM, van Holten TC, Kester MG, Heemskerk MH, et al. Early cytomegalovirus reactivation leaves a specific and dynamic imprint on the reconstituting T cell compartment long-term after hematopoietic stem cell transplantation. Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation. 2014;20(5):655-61.

## Figure legends

**Figure 1**

Overlap of the TCR CDR3β repertoire in donors prior to transplantation and in their full chimeric recipients at one year post-HSCT. (a) Prototypical examples of overlap in two selected donor/recipient pairs. The clone's frequencies pre and post-TX (along x and y axis, respectively) and number of unique clones (dot size) are represented by scatter plots. Clones that are only observed at one time point are colored in red, while clones observed at both time points (i.e. overlapping clones) are colored in green. The values for standard Jaccard, standardized Jaccard and Morisita's index are 0.021, 0.014 and 0.013 for pair #1 plotted on the left and 0.082, 0.117 and 0.264 for pair #2 plotted on the right, respectively. (b) Repertoire overlap according to the indices of similarity among the 116 donor/recipient pairs at several cut-off values. These indices vary between 0 (no overlap) and 1 (complete overlap) and are represented along the x axis. The number of donor/recipient pairs at each cut-off is plotted along the y axis. TX: allogeneic HSCT.

**Figure 2**

Productive clonality of the TCR CDR3β repertoire in donors prior to transplantation and in their full chimeric recipients at one year post-HSCT. Clonality is shown along the y axis according to (a) age of recipients at transplantation (n=13 and 103 for recipients ≤ or > than 20 years old, respectively), (b) age of donors at transplantation (n=27 and 79 for donors ≤ or > than 30 years old, respectively, the age of 10 donors is unknown) and (c) cytomegalovirus (CMV) serologic status and occurrence of CMV infection/reactivation within the first year post-HSCT (see Table 1 for the numbers included in each category). In (a) clonality is plotted separately for both time points (i.e. pre and post-TX) with gray lines connecting

donor/recipient pairs. Clonality varies between 0 (polyclonal/diverse repertoire) and 1 (monoclonal/invariant repertoire). In (b) and (c) the variation of the repertoire diversity is plotted as clonality post-TX minus clonality pre-TX. Thus, a negative/positive value indicates an increased/decreased repertoire diversity at one year post-HSCT, respectively. D: donor CMV negative (-) or positive (+), R: recipient CMV negative (-) or positive (+), TX: allogeneic HSCT.

**Figure 3**

Cumulative frequency (represented along the log scaled y axis) and number (represented by the dot size) of CMV-specific T-cell clones in donors prior to transplantation and in their full chimeric recipients at one year post-HSCT according to CMV serologic status (shown in the four panels along x axis) and CMV infection/reactivation within the first year post-HSCT (shown in the upper and lower panel). A frequency threshold of 1/1000 is indicated by the dotted line. D: donor CMV negative (-) or positive (+), R: recipient CMV negative (-) or positive (+), TX: allogeneic HSCT.

**Figure 4**

Distribution of T-cell clones in donors prior to transplantation (upper panels) and in recipients at one year post-HSCT (lower panels). Clones are categorized according to two parameters, (1) their CMV specificity, non-specific clones being classified as other and (2) their dominance (i.e. clones exhibiting a frequency of 10% or more in a given pair, either in the donor or in the recipient, are considered as dominant). This allowed to classify all sequenced clones into three subclasses shown in different colors on the plot. Of note, a fourth subclass consisting in dominant CMV-specific clones was not observed in any donor or

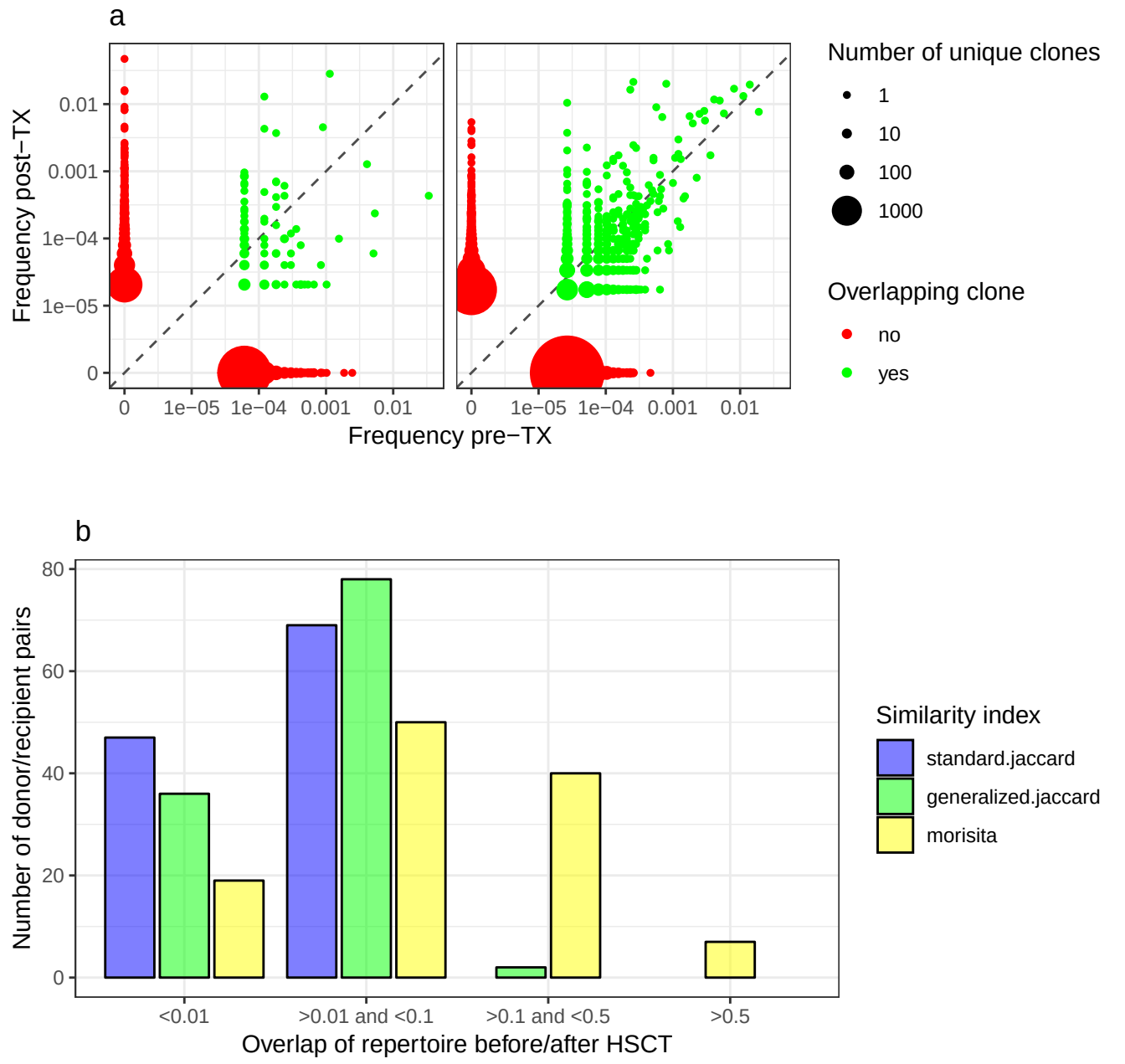
patient. The y axis represents the cumulative frequencies of clones comprised within each subclass while donor/recipient pairs are listed along the x axis according to CMV serologic status (indicated on the top). The sizes of the faceted plots are proportional to the number of pairs comprised within each group.

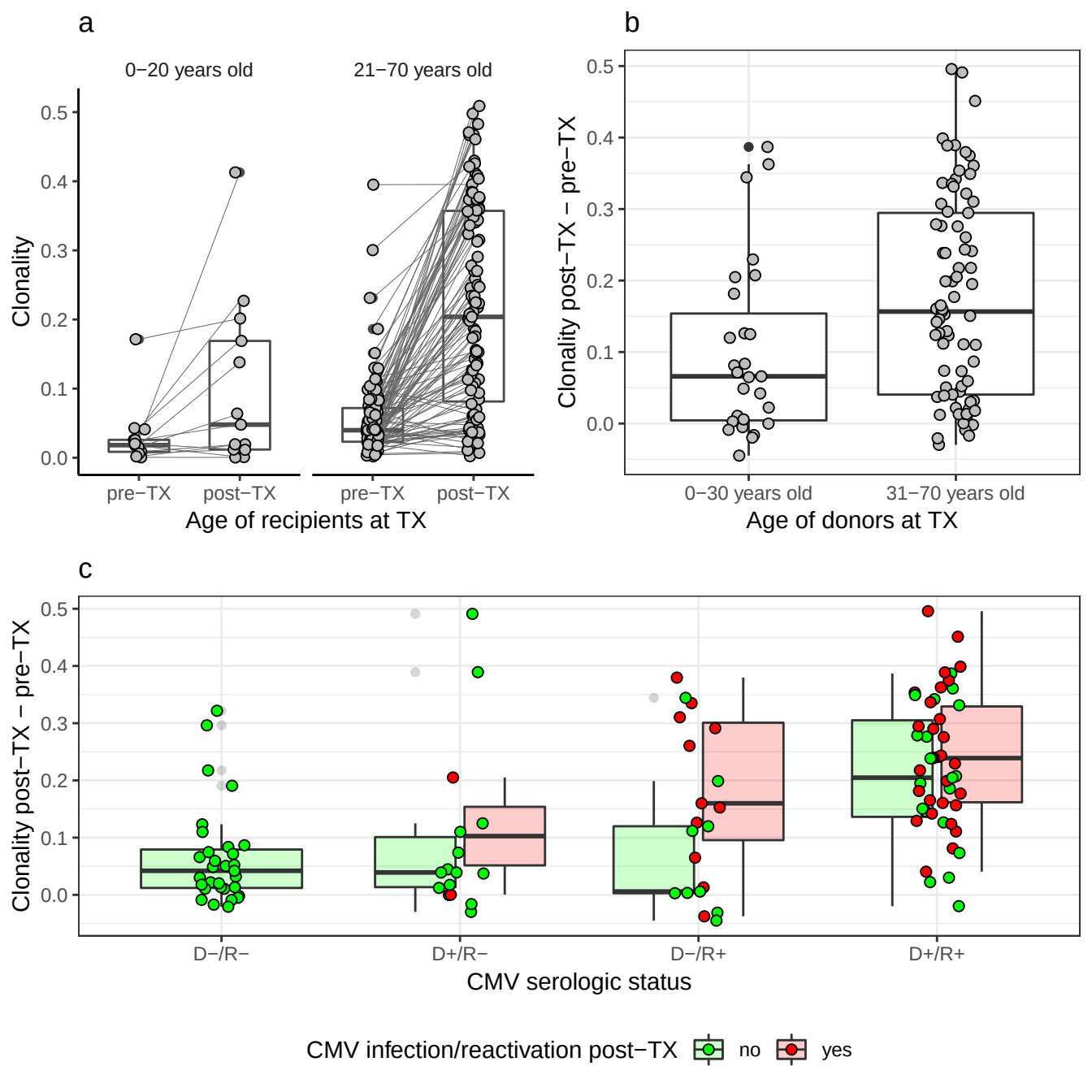
## Table legends

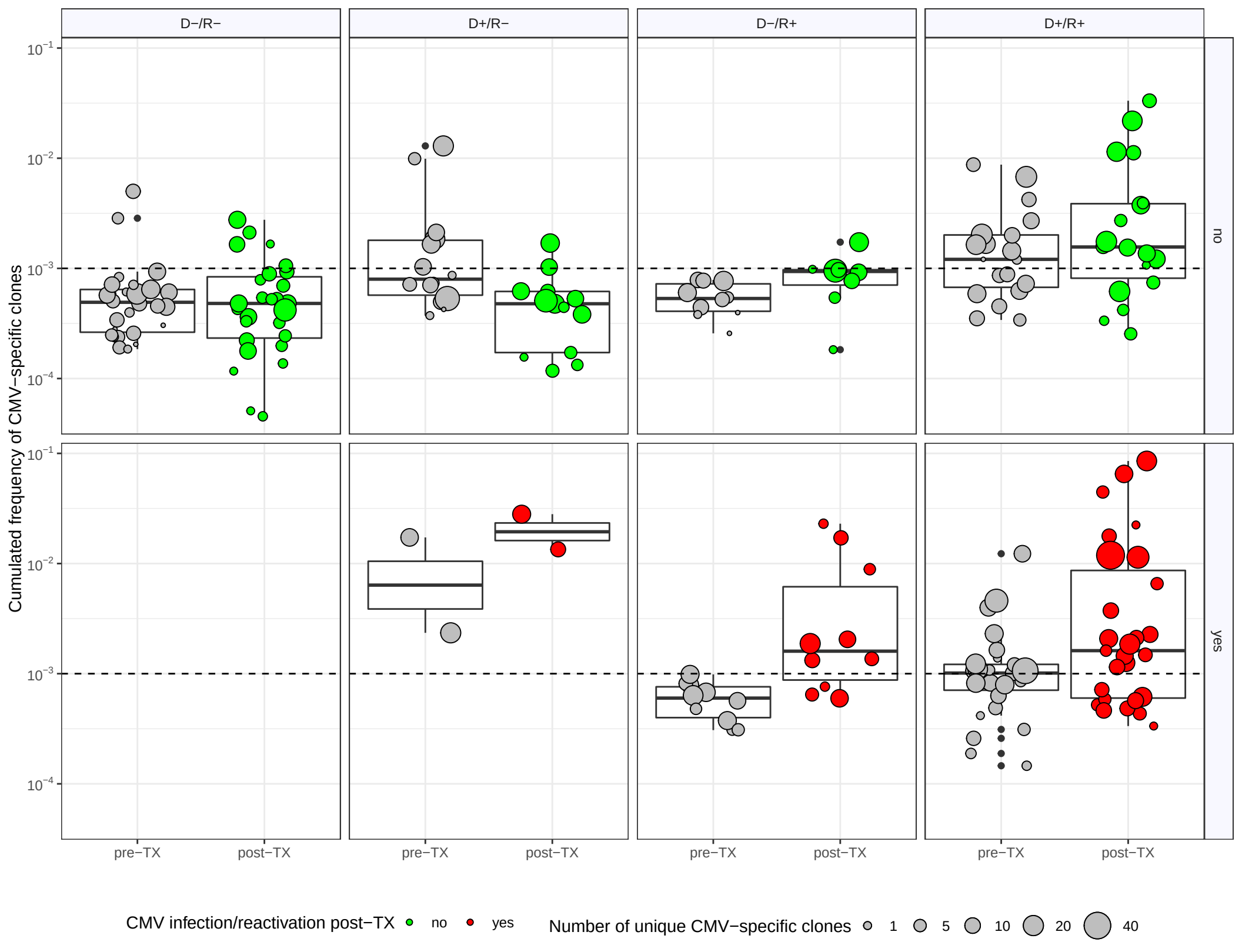
**Table 1** Patient's and transplant's characteristics

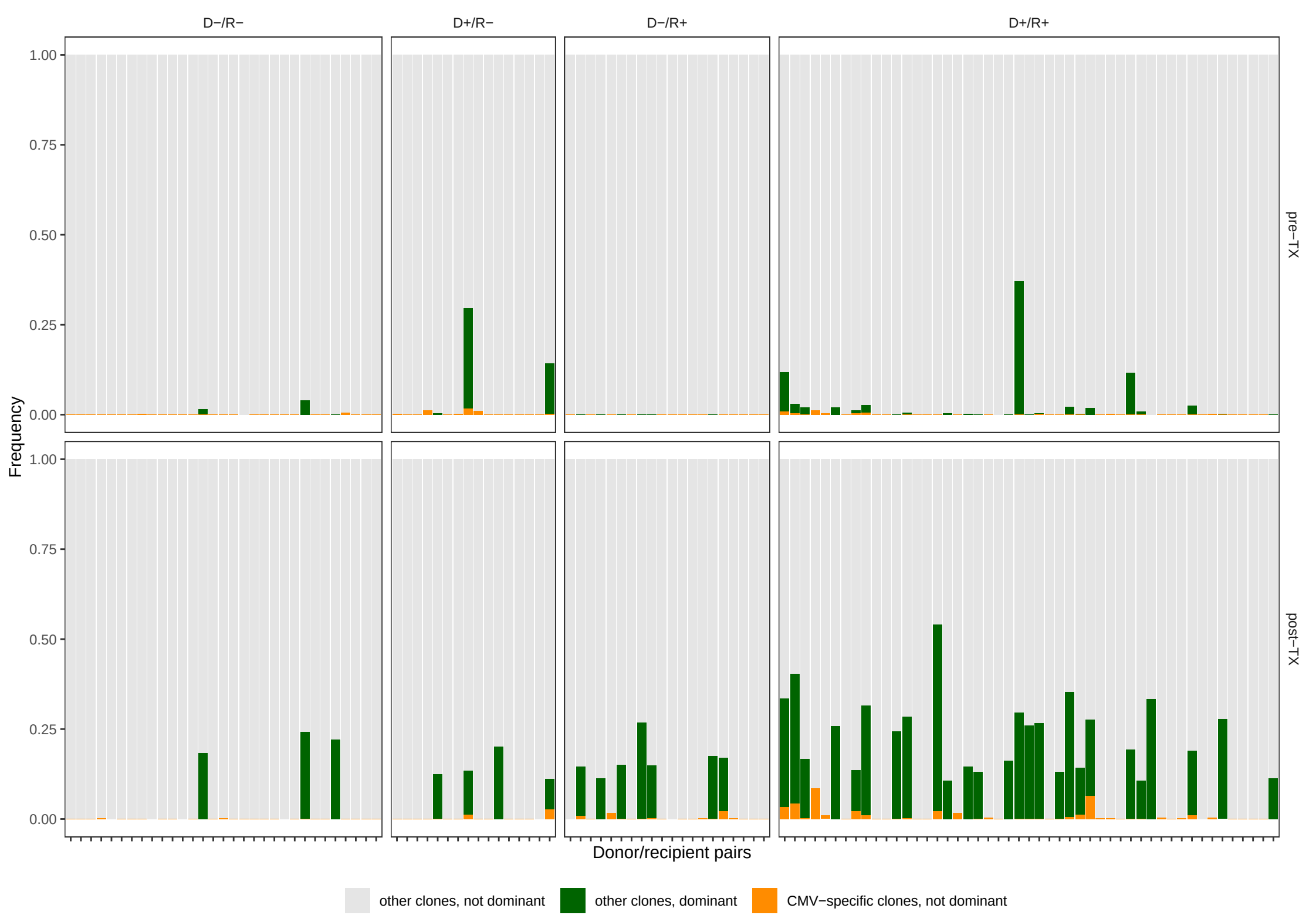
**Table 2** Post-transplantation complications among 116 alloHSCT recipients

**Table 3** Multivariable analyses by linear regression for TCR reconstitution at one year


a
Frequency post−TX
0.01
0.001
1e−04
1e−05
0
0
1e−05
1e−04
0.001
0.01
Frequency pre−TX
Number of unique clones
1
10
100
1000
Overlapping clone
no
yes
b
Number of donor/recipient pairs
80
60
40
20
0
<0.01
>0.01 and <0.1
>0.1 and <0.5
>0.5
Overlap of repertoire before/after HSCT
Similarity index
standard.jaccard
generalized.jaccard
morisita


a
0−20 years old
21−70 years old
Clonality
0.0
0.1
0.2
0.3
0.4
0.5
pre−TX
post−TX
pre−TX
post−TX
Age of recipients at TX
b
Clonality post−TX − pre−TX
0.0
0.1
0.2
0.3
0.4
0.5
0−30 years old
31−70 years old
Age of donors at TX
c
Clonality post−TX − pre−TX
0.0
0.1
0.2
0.3
0.4
0.5
D−/R−
D+/R−
D−/R+
D+/R+
CMV serologic status
CMV infection/reactivation post−TX
no
yes


D−/R−
D+/R−
D−/R+
D+/R+
no
yes
Cumulated frequency of CMV−specific clones
$10^{-1}$
$10^{-2}$
$10^{-3}$
$10^{-4}$
pre−TX
post−TX
CMV infection/reactivation post−TX
no
yes
Number of unique CMV−specific clones
1
5
10
20
40


D−/R−
D+/R−
D−/R+
D+/R+
pre−TX
post−TX
Frequency
1.00
0.75
0.50
0.25
0.00
Donor/recipient pairs
other clones, not dominant
other clones, dominant
CMV−specific clones, not dominant

| Parameter | N=116 donor/recipient pairs |
|---|---|
| Recipient age in years | median: 45.5, minimum: 0 (4,5 months), maximum: 66 |
| Recipient sex | female: 49, male: 67 |
| Primary disease | AA: 6, ALL: 14, AML: 36, CLL: 3, CML: 16, inborn error: 1, lymphoma: 13, MDPS: 5, MDS: 13, MPS: 2, myeloma: 5, solid tumor: 2 |
| Donor age in years | median: 39.5, minimum: 1, maximum 65 |
| Donor sex | female: 50, male: 66 |
| Type of donor | unrelated: 42, related: 70, haploidentical: 4 |
| Source of stem cells | BM: 16, CB: 2, PBSC: 98 |
| Conditioning | MAC: 78 (including 33 with ATG), RIC: 38 (including 31 with ATG) |
| *Ex vivo* T-cell depletion | fraction depleted: 0%: 48, 50%: 62, 100%: 6 |
| Number of DLI | 0: 90, 1: 6, 2: 6, 3: 7, 4: 4, 5: 2, 7: 1 |
| HLA matching | 10/10: 101, <10/10: 15 |
| Sex matching | F/F: 23, F/M: 27, M/F: 26, M/M: 40 |
| CMV serologic status | D-/R-: 31, D-/R+: 20, D+/R-: 16, D+/R+: 49 |

ATG: anti-thymocyte globulin, AA: aplastic anemia, ALL: acute lymphoblastic leukemia, AML: acute myeloid leukemia, BM: bone marrow, CB: cord blood, CLL: chronic lymphocytic leukemia, CML: chronic myeloid leukemia, D: donor, DLI: donor lyphocyte infusion, F: female, M: male, MAC: myeloablative conditioning, MDPS: combined myelodysplastic syndrome / myeloproliferative neoplasm, MDS: myelodysplastic syndrome, MPS: myeloproliferative syndrome, PBSC: peripheral blood stem cells, R:recipient, RIC: reduced intensity conditioning.

**Table 1** Patient's and transplant's characteristics

| Type of complication | Within one year, n (%) | After one year, n (%) |
|---|---|---|
| Acute GVHD | 68 (58.6%) | 16 (13.8%) |
| | Grade 1: 23 (19.8%) | Grade 1: 7 (6%) |
| | Grade ≥2: 43 (37.1%) | Grade ≥2: 9 (7.8%) |
| | Non-available: 2 (1.7%) | - |
| Chronic GVHD | 28 (24.1%) | 20 (17.2%) |
| Relapse | 24 (20.7%) | 19 (16.4%) |
| Death | - | 24 (20.7%) |
| Any infection | 84 (72.4%) | 34 (29.3%) |
| Viral infection (CMV and/or others) | 63 (54.3%) | 22 (20%) |
| Cytomegalovirus infection/reactivation | 43 (37.1%) | 7 (6%) |
| | CMV alone: 23 (19.8%) | CMV alone: 3 (2.6%) |
| | CMV and other virus: 20 (17.2%) | CMV and other virus: 4 (3.4%) |
| Bacterial infection | 48 (41.4%) | 19 (16.4%) |
| Fungal infection | 10 (8.6%) | 6 (5.2%) |
| Parasitic infection | 2 (1.7%) | 2 (1.7%) |

**Table 2** Post-transplantation complications among 116 alloHSCT recipients

| Response variable | Explanatory variables | Categories tested | Baseline | Coefficient estimate | p-value |
|---|---|---|---|---|---|
| Clonality post-TX - pre-TX | (1) recipient age group | 0-20, 21-70 years old | 0-20 years old | 0.08 (21-70 years old) | *0.026* |
| | | | | 0.04 (D+/R-) | 0.287 |
| | (2) CMV serologic status | D-/R-, D-/R+, D+/R-, D+/R+ | D-/R- | 0.08 (D-/R+) | *0.022* |
| | | | | 0.16 (D+/R+) | *1.34E-08* |
| | (1) donor age group | 0-30, 31-70 years old | 0-30 years old | 0.06 (31-70 years old) | *0.023* |
| | | | | 0.04 (D+/R-) | 0.266 |
| | (2) CMV serologic status | D-/R-, D-/R+, D+/R-, D+/R+ | D-/R- | 0.1 (D-/R+) | *0.009* |
| | | | | 0.17 (D+/R+) | *4.61E-08* |
| | (1) CMV infection/reactivation | no, yes | no | 0.06 (yes) | *0.043* |
| | | | | 0.02 (D+/R-) | 0.5 |
| | (2) CMV serologic status | D-/R-, D-/R+, D+/R-, D+/R+ | D-/R- | 0.05 (D-/R+) | 0.209 |
| | | | | 0.13 (D+/R+) | *4.82E-05* |

The intercepts of the regressions are not shown. There was no significant interaction between the above variables (i.e. models testing for interaction are not shown).

CMV: cytomegalovirus, D: donor, R: recipient.

**Table 3** Multivariable analyses by linear regression for TCR reconstitution at one year

| Response variable | Explanatory variable | Categories tested | Baseline | Coefficient estimate | p-value |
|---|---|---|---|---|---|
| Clonality post-TX - pre-TX | Recipient age at TX | continuous variable | - | 0.001 | 0.135 |
| | Recipient age group | 0-20, 21-70 years old | 0-20 years old | 0.09 (21-70 years old) | *0.025* |
| | Recipient sex | male, female | female | 0.02 (male) | 0.447 |
| | Primary disease | AA, ALL, AML, CLL, CML, inborn error, lymphoma, MDPS, MDS, MPS, myeloma, solid tumor | AA | 0.05 (ALL) | 0.454 |
| | | | | 0.03 (AML) | 0.654 |
| | | | | 0.13 (CLL) | 0.193 |
| | | | | 0.11 (CML) | 0.081 |
| | | | | -0.11 (inborn error) | 0.444 |
| | | | | 0.12 (lymphoma) | 0.077 |
| | | | | 0.02 (MDPS) | 0.848 |
| | | | | 0.01 (MDS) | 0.896 |
| | | | | -0.004 (MPS) | 0.974 |
| | | | | 0.03 (myeloma) | 0.683 |
| | | | | -0.06 (solid tumor) | 0.595 |
| | Recipient CMV | positive, negative | negative | 0.13 (positive) | *2.06E-07* |
| | Type of donor | unrelated, related, haploidentical | haploidentical | 0.035 (related) | 0.624 |
| | | | | 0.002 (unrelated) | 0.978 |
| | Donor age | continuous variable | - | 0.002 | 0.063 |
| | Donor age group | 0-30, 31-70 years old | 0-30 years old | 0.075 (31-70 years old) | *0.016* |
| | Donor sex | male, female | female | -0.04 (male) | 0.099 |
| | Donor CMV | positive, negative | negative | 0.1 (positive) | *3.17E-05* |
| | HLA matching | matched (10/10), mismatched (<10/10) | matched 10/10 | 0.005 (mismatched <10/10) | 0.901 |
| | CMV status | D-/R-, D-/R+, D+/R-, D+/R+ | D-/R- | 0.03 (D+/R-) | 0.391 |
| | | | | 0.07 (D-/R+) | *0.033* |
| | | | | 0.17 (D+/R+) | *1.78E-08* |
| | Sex matching | F/F, M/F, F/M, M/M (donor/recipient) | F/F | -0.06 (M/F) | 0.15 |
| | | | | 0.01 (F/M) | 0.796 |
| | | | | -0.02 (M/M) | 0.501 |

| Response variable | Explanatory variable | Categories tested | Baseline | Coefficient estimate | p-value |
|---|---|---|---|---|---|
| Clonality post-TX - pre-TX | Conditioning (model 1) | MAC, RIC | MAC | 0.01 (RIC) | 0.641 |
| | Conditioning (model 2) | MAC-ATG, MAC-noATG, RIC-ATG, RIC-noATG | MAC-ATG | 0.01 (MAC-noATG) | 0.662 |
| | | | | 0.04 (RIC-ATG) | 0.299 |
| | | | | -0.05 (RIC-noATG) | 0.415 |
| | Source of stem cells | BM, CB, PBSC | BM | -0.12 (CB) | 0.261 |
| | | | | 0.04 (PBSC) | 0.263 |
| | T cell depletion | no, yes | no | -0.04 (yes) | 0.109 |
| | T cell fraction depleted | 0%, 50%, 100% | 0% | 0.05 (50%) | 0.079 |
| | | | | -0.01 (100%) | 0.856 |
| | Number of DLI | n=0, 1, 2, 3, 4, 5, 7 | n=0 | 0.03 (n=1) | 0.596 |
| | | | | -0.03 (n=2) | 0.627 |
| | | | | 0.05 (n=3) | 0.398 |
| | | | | 0.06 (n=4) | 0.405 |
| | | | | 0.07 (n=5) | 0.498 |
| | | | | -0.15 (n=7) | 0.297 |
| | Acute GVHD | grades 0, 1, ≥2 | grade 0 | -0.005 (grade 1) | 0.88 |
| | | | | 0.03 (grade ≥2) | 0.255 |
| | Chronic GVHD | no, yes | no | 0.02 (yes) | 0.565 |
| | Relapse within first year | no, yes | no | 0.009 (yes) | 0.776 |
| | Any infection | no, yes | no | 0.04 (yes) | 0.153 |
| | Viral infection (model 1) | no, yes | no | 0.1 (yes) | *7.72E-05* |
| | Viral infection (model 2) | no, yes excluding CMV, yes including CMV | no | 0.04 (yes excluding CMV) | 0.248 |
| | | | | 0.13 (yes including CMV) | *3.21E-06* |
| | Viral infection (model 3) | no, yes excluding CMV, yes CMV only, yes CMV and other virus(es) | no | 0.04 (yes excluding CMV) | 0.248 |
| | | | | 0.11 (yes CMV only) | *6.10E-04* |
| | | | | 0.14 (yes CMV and other virus(es)) | *2.75E-05* |
| | CMV infection/reactivation | no, yes | no | 0.12 (yes) | *5.04E-06* |
| | Bacterial infection | no, yes | no | -0.02 (yes) | 0.536 |

| Response variable | Explanatory variable | Categories tested | Baseline | Coefficient estimate | p-value |
|---|---|---|---|---|---|
| Clonality post-TX - pre-TX | Fungal infection | no, yes | no | -0.03 (yes) | 0.467 |
| | Parasitic infection | no, yes | no | 0.002 (yes) | 0.982 |

The intercepts of the univariate linear regressions are not shown.

ATG: anti-thymocyte globulin, AA: aplastic anemia, ALL: acute lymphoblastic leukemia, AML: acute myeloid leukemia, BM: bone marrow, CB: cord blood, CLL: chronic lymphocytic leukemia, CML: chronic myeloid leukemia, CMV: cytomegalovirus, D: donor, DLI: donor lyphocyte infusion, F: female, GVHD: graft versus host disease, M: male, MAC: myeloablative conditioning, MDPS: combined myelodysplastic syndrome / myeloproliferative neoplasm, MDS: myelodysplastic syndrome, MPS: myeloproliferative syndrome, PBSC: peripheral blood stem cells, R:recipient, RIC: reduced intensity conditioning.

**Table S1** Summary of the univariate analyses on parameters possibly influencing TCR reconstitution as measured by the difference of clonality at one year post-alloHSCT and prior to transplantation

| Response variable | Explanatory variable | Categories tested | Baseline | Coefficient estimate | p-value |
|---|---|---|---|---|---|
| Clonality pre-TX* | Type of donor | unrelated, related, haploidentical | haploidentical | 0.01 (related) | 0.651 |
| | | | | 0 (unrelated) | 0.98 |
| | Donor age | continuous variable | - | 0.0014 | *0.00122* |
| | Donor age group | 0-30, 31-70 years old | 0-30 years old | 0.03 (31-70 years old) | *0.01* |
| | Donor sex | male, female | female | 0.0066 (male) | 0.541 |
| | Donor CMV | positive, negative | negative | 0.037 (positive) | *0.0004* |
| | Source of stem cells | BM, CB, PBSC | BM | -0.04 (CB) | 0.312 |
| | | | | 0.008 (PBSC) | 0.588 |
| | multivariate model 1 | | | | |
| | (1) Donor age group | continuous variable | - | 0.0003 | 0.68 |
| | (2) Donor CMV | positive, negative | negative | -0.027 (positive) | 0.422 |
| | (3) age : CMV (interaction) | - | - | 0.0016 (age : positive) | 0.063 |
| | multivariate model 2 | | | | |
| | (1) Donor age | 0-30, 31-70 years old | 0-30 years old | 0.002 (31-70 years old) | 0.908 |
| | (2) Donor CMV | positive, negative | negative | -0.003 (positive) | 0.867 |
| | (3) age group : CMV (interaction) | - | - | 0.05 (31-70 years old : positive) | *0.038* |
| **Response variable** | **Explanatory variable** | **Categories tested** | **Baseline** | **Coefficient estimate** | **p-value** |
| Acute GVHD** | Clonality pre-TX | continuous variable | - | -5.1 | 0.159 |
| Chronic GVHD** | | | | 1.6 | 0.65 |
| Relapse** | | | | 1.76 | 0.639 |
| Any infection** | | | | 0.5 | 0.9 |
| Viral infection** | | | | 4.1 | 0.273 |
| CMV infection/reactivation** | | | | 8.2 | *0.04* |
| Bacterial infection** | | | | -1.5 | 0.662 |
| Fungal infection** | | | | -13.6 | 0.243 |
| Parasitic infection** | | | | -14.2 | 0.589 |

The intercepts of the regressions (linear * or logistic **) are not shown.

**Table S2** (upper part) donor parameters possibly influencing the clonality of the donor's repertoire infused with the graft; (lower part) predicting clinical events within the first year post-alloHSCT based on the donor's repertoire clonality

| Response variable | Explanatory variable | Categories tested | Baseline | Coefficient estimate | p-value |
|---|---|---|---|---|---|
| Death | Clonality post-TX | continuous variable | - | -0.9 | 0.549 |
| Relapse | | | | 0.5 | 0.744 |
| Acute GVHD | | | | -0.5 | 0.785 |
| Chronic GVHD | | | | 1.1 | 0.497 |
| Any infection | | | | 1.9 | 0.167 |
| Viral infection | | | | 2.1 | 0.178 |
| CMV infection/reactivation | | | | 3.8 | 0.156 |
| Bacterial infection | | | | 2.8 | 0.101 |
| Fungal infection | | | | 2.3 | 0.403 |
| Parasitic infection | | | | 4.9 | 0.331 |

The intercepts of the logistic regressions are not shown.

**Table S3** Predicting clinical events after one year based on the patient's repertoire clonality at one year post-alloHSCT

**Supplementary Figures**

**Figure S1**

Overlap of the TCR CDR3β repertoire in donors prior to transplantation and in their full chimeric recipients at one year post-HSCT according to three indices of similarity (i.e. standard Jaccard, generalized Jaccard and Morisita, see material and methods for the characteristics of each index). The similarity of repertoire before and after HSCT, as measured by the three indices, is provided on pairwise scatter plots with each point representing a given donor/recipient pair. Density distribution of each index across donor/recipient pairs is also plotted alongside their pairwise correlation (the three indices are significantly correlated, values of 0.732, 0.362, 0.599, p-values ranging between 2.2e-16 and 6.58e-05).

**Figure S2**

Impact of different types of viral infection on the TCR CDR3β repertoire within the first year post-HSCT. The variation of diversity before and after alloHSCT is estimated as clonality post-TX minus clonality pre-TX. The plot discloses a gradient with the lowest deviation of diversity observed in non-infected patients and the highest found in patients with a CMV infection/reactivation. The patients with other viral infection(s) (i.e. excluding CMV) exhibit an intermediate profile. A Tukey HSD test confirms that the only significant difference is CMV infection/reactivation versus the baseline consisting of non-infected patients (also see Table S1). TX: allogeneic HSCT.

**Figure S3**

Conditional inference trees on the variation of productive clonality before and after alloHSCT computed as clonality post-TX minus clonality pre-TX. The analysis was performed using the

four significant explanatory variables detected by linear regression (i.e. recipient's age group, donor's age group, donor/recipient CMV serologic status and occurrence of CMV reactivation/infection within the first year post-HSCT). Briefly, the purpose of conditional inference is to test the global null hypothesis of independence between one or several input variables (as listed above) and a response variable (here the variation of productive clonality). The partitioning algorithm works as follows: (1) by keeping the input variable with the strongest association, if any, to the response (and providing a p-value for this association), (2) by implementing a binary split among its subgroups (or categories), (3) by repeating this process recursively until no significant input variable remains. Put simply, the goal is to identify which independent input variables and which subgroups (or categories) among the identified variables are strongly associated with the response variable. The graphical output obtained is a tree with the variation of clonality represented along the y axis in the form of box and whisker plots and nodes representing the identified variables, their p-values and the binary partitioning into subgroups. Detailed information on conditional inference framework is available in the R documentation and the party package (https://www.rdocumentation.org/packages/party/versions/1.3-3).

According to the conditional inference trees, the most significant variable associated to the variation of clonality is CMV serologic status ($p<0.001$), and more specifically the D+/R+ group (i.e. patients with a reservoir of latent and persistent virus who receive a graft from a positive donor which may contain CMV-reactive T cells rapidly expandable upon infection/reactivation post-HSCT). Occurrence of CMV infection/reactivation is also significantly associated to the global shift of clonality but less strongly than serologic status (i.e. it is significant as a second level node in the tree ($p=0.036$)). This suggests that subclinical CMV reactivation was probably sufficient to drive the expansion of CMV-specific T cells and to skew the clonality of the

repertoire, an observation that fits well with the data shown on Figures 2c and 3. Indeed, several D+/R+ pairs were asymptomatic regarding occurrence of CMV infection/reactivation within the first year post-HSCT but they still exhibit increased cumulated frequencies of potential CMV-specific clones and oligoclonal profiles at one year. In addition, age is not associated to clonality when considered alongside the two CMV-related input variables, although both patient's and donor's age groups are found to be slightly significant if considered separately as input variable to the tree (not shown). This is in agreement with the linear regression analyses (see Table 3 and Table S1). TX: allogeneic HSCT.

**Figure S4**

Investigating the linear relationship between productive clonality and the cumulative frequency of potential CMV-specific clones before and after HSCT within our cohort. Both variables are compared by measuring the values post-TX minus the values pre-TX and (a) are significantly correlated but with a modest coefficient (Pearson's r=0.249, p=0.007; Spearman's rho=0.318, p=0.0005). (b) Simulating the correlation between both variables using a permutation procedure. CMV specificity was attributed randomly to T-cell clones under the same conditions than in observed data (i.e. 1,978 clones were randomly assigned as CMV-specific within the total number of clones). This procedure was repeated 10,000 times to generate an empirical distribution of the correlation coefficient. The observed coefficient is extreme among simulated values (i.e. quantile 9 997 of 10 000), meaning that it was not possible to reproduce a similar correlation using the frequencies of randomly chosen clones. This strongly suggests that expanded CMV-specific clones are the main drivers of the reduction of repertoire diversity observed at one year post-HSCT.

**Figure S5**

Heat map of the total number of unique CMV-derived 9mer peptides that are theoretically bound by patients according to their HLA class I molecules considered conjointly (HLA-A, B and C) and according to NetMHCpan binding predictions for the corresponding alleles. The 116 patients are plotted in no particular order along the x axis. The number of theoretically bound 9mer is presented separately for each of the 190 proteins of CMV strain AD169 from which they are derived (along the y axis) and subdivided into peptides of weak or strong binding affinity (left and right panels). The number of theoretically bound 9mer is quite variable between different viral proteins and among patients, ranging from 0 (white on the heat map) to 550 (brown on the heat map). However, the peptide coverage is large across most viral proteins for every patient (i.e. median of 43 9mer bound per protein). The four viral proteins from which the largest number of 9mer are derived and theoretically bound are the major DNA-binding protein, the DNA polymerase catalytic subunit, the large tegument protein deneddylase and the major capsid protein. Noteworthy, the 65 kDa phosphoprotein (pp65) and 55 kDa immediate-early protein 1 (IE1) usually considered among the most immunodominant antigens of CMV are derived into an average of 34/85 and 30/71 strong/weak 9mer, respectively, potentially presented on the patient's HLA class I molecules. Although these numbers suggest a large peptide coverage of both pp65 and IE1 in every patient, they do not stand as specifically elevated in comparison to other viral proteins. However, the notion of immunodominance of a given protein cannot solely be based on the breadth of peptide coverage and should also account for additional and more subtle characteristics not captured by our approach. For instance, the quantity of displayed peptides at the cell surface, the stability and conformation of the peptide-HLA complexes and the

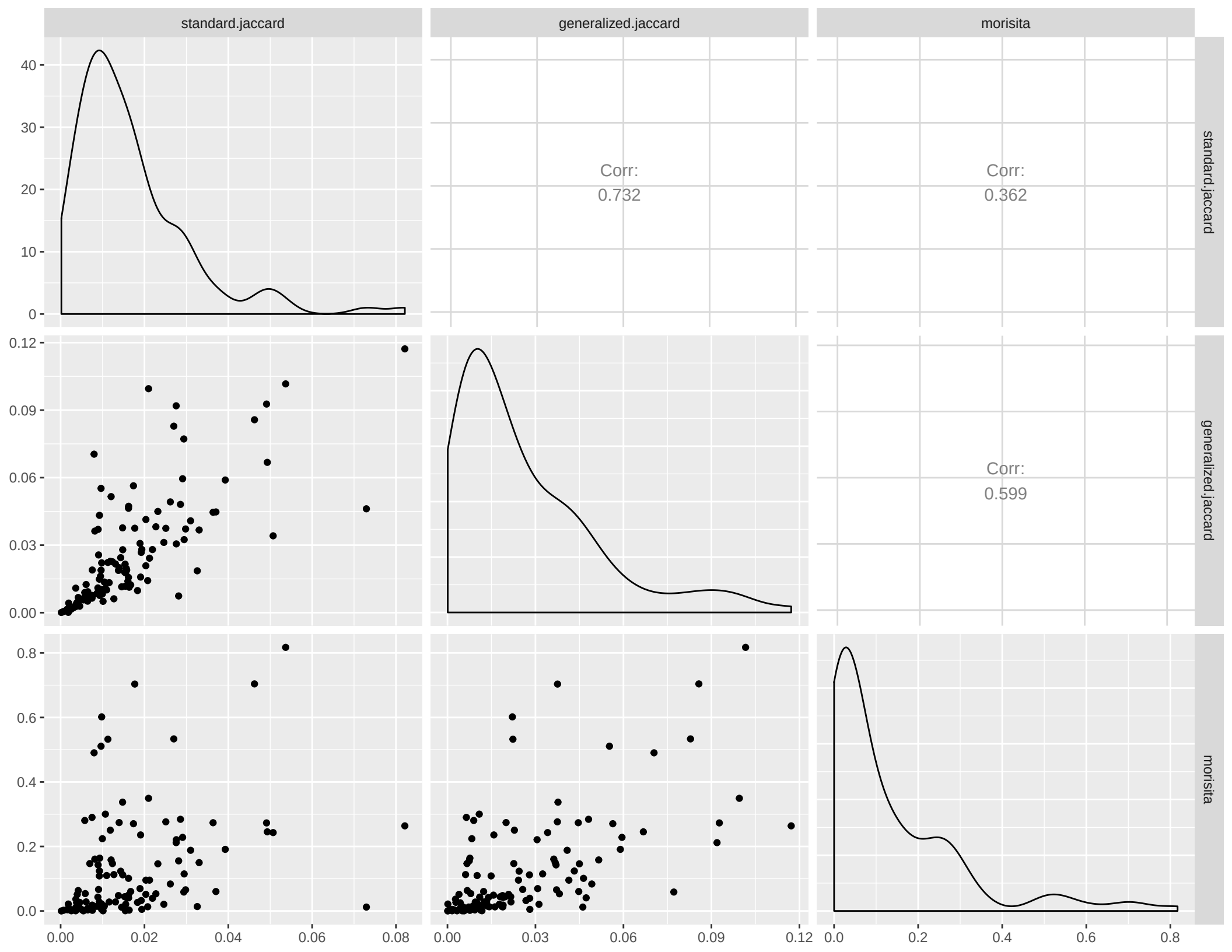
spectrum of T cells that will be able to recognize these complexes with high affinity in different individuals are key components of the adaptive immune response.

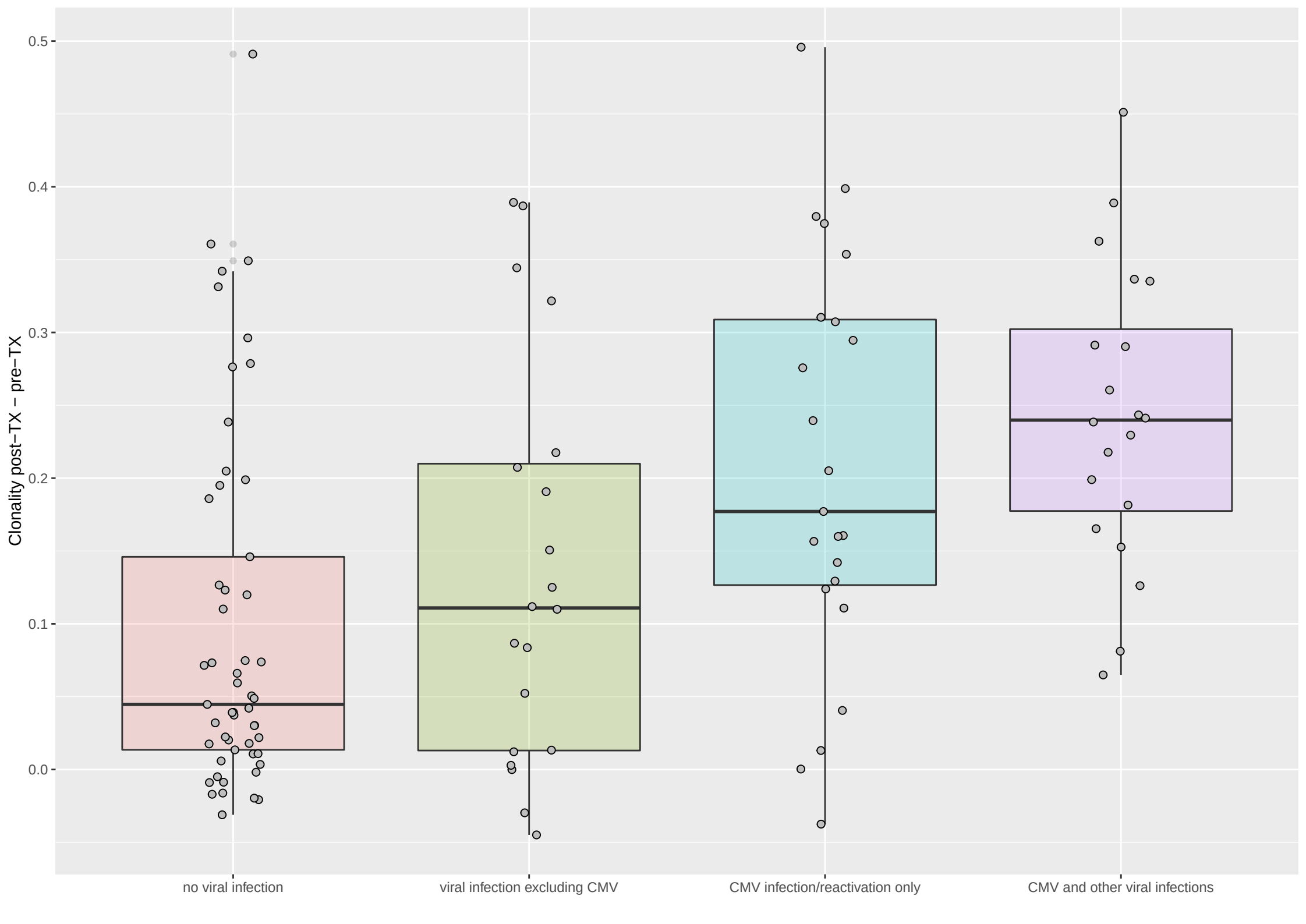
**Figure S6**

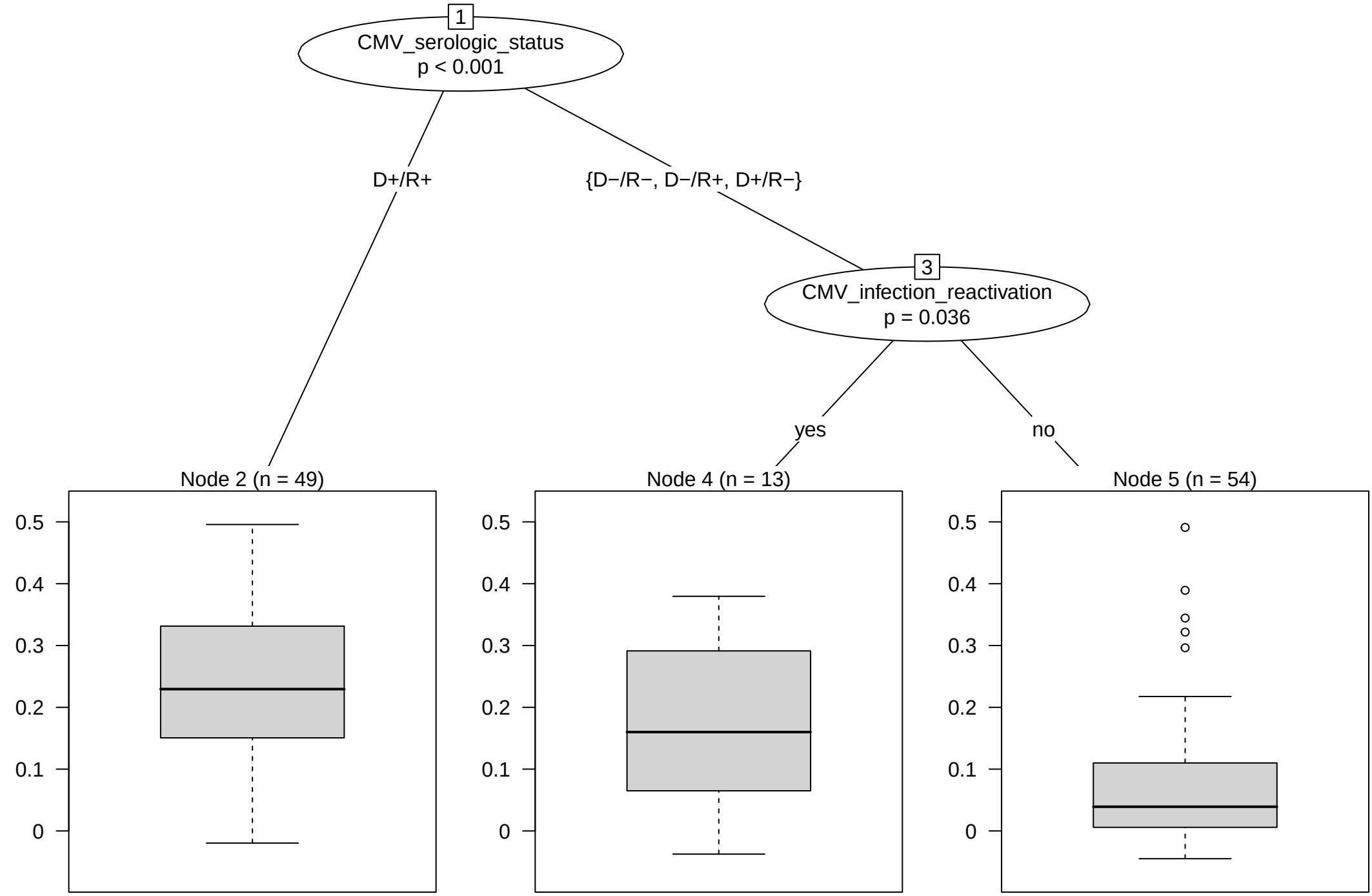
Theoretical capacity of patients to present peptides derived from CMV strain AD169 on their HLA class I molecules as estimated by (a) the mean peptide binding affinity in nanomolar (nM) averaged over all retained potential binders and (b) the total number of potentially bound 9mer (subdivided into strong and weak binders). Patients are grouped according to non-occurrence/occurrence of CMV infection/reactivation within the first year post-transplantation. The visual inspection of boxplots and formal testing by logistic regression does not reveal an association between potential binding coverage and occurrence of CMV infection/reactivation post-HSCT (p-values of 0.58 and 0.50, respectively).

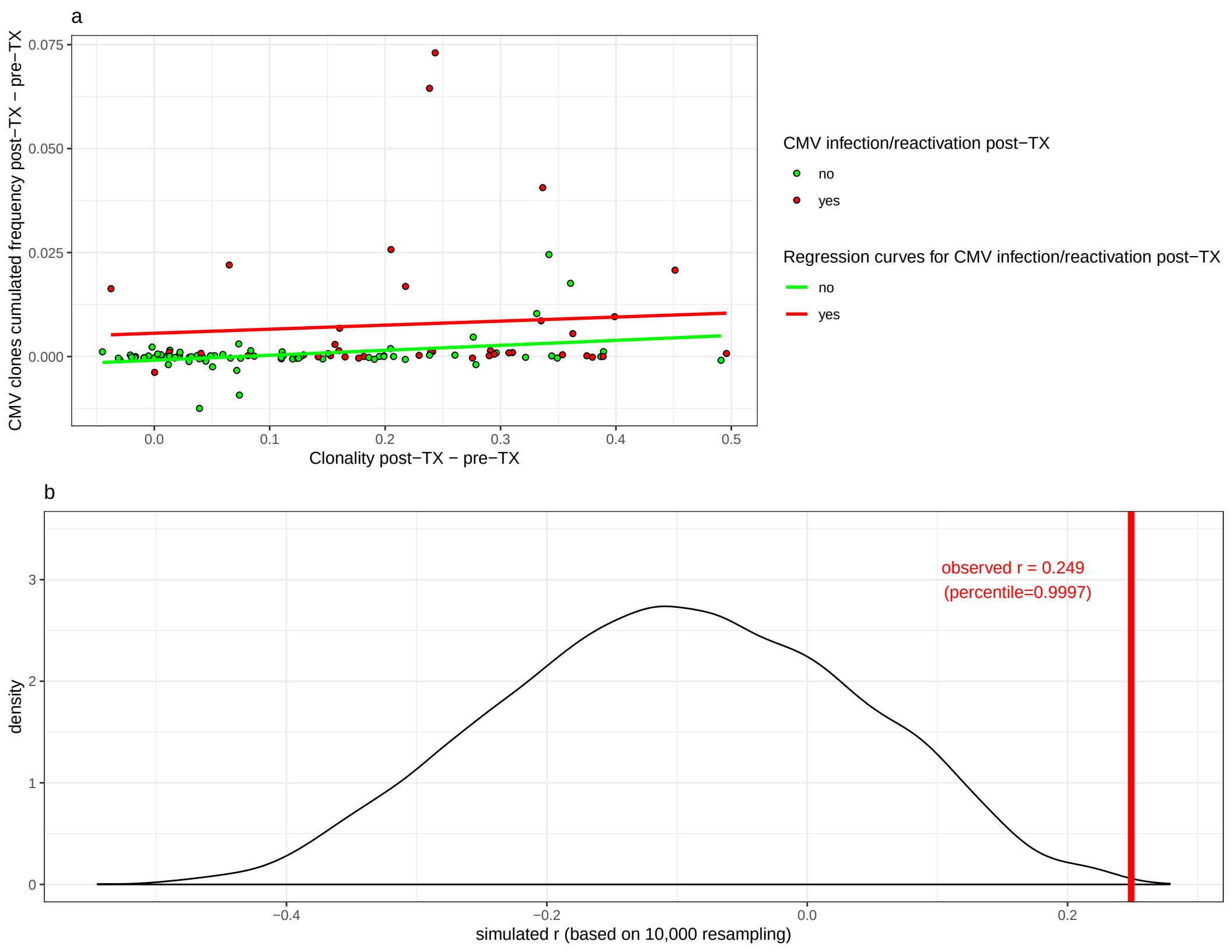
**Figure S7**

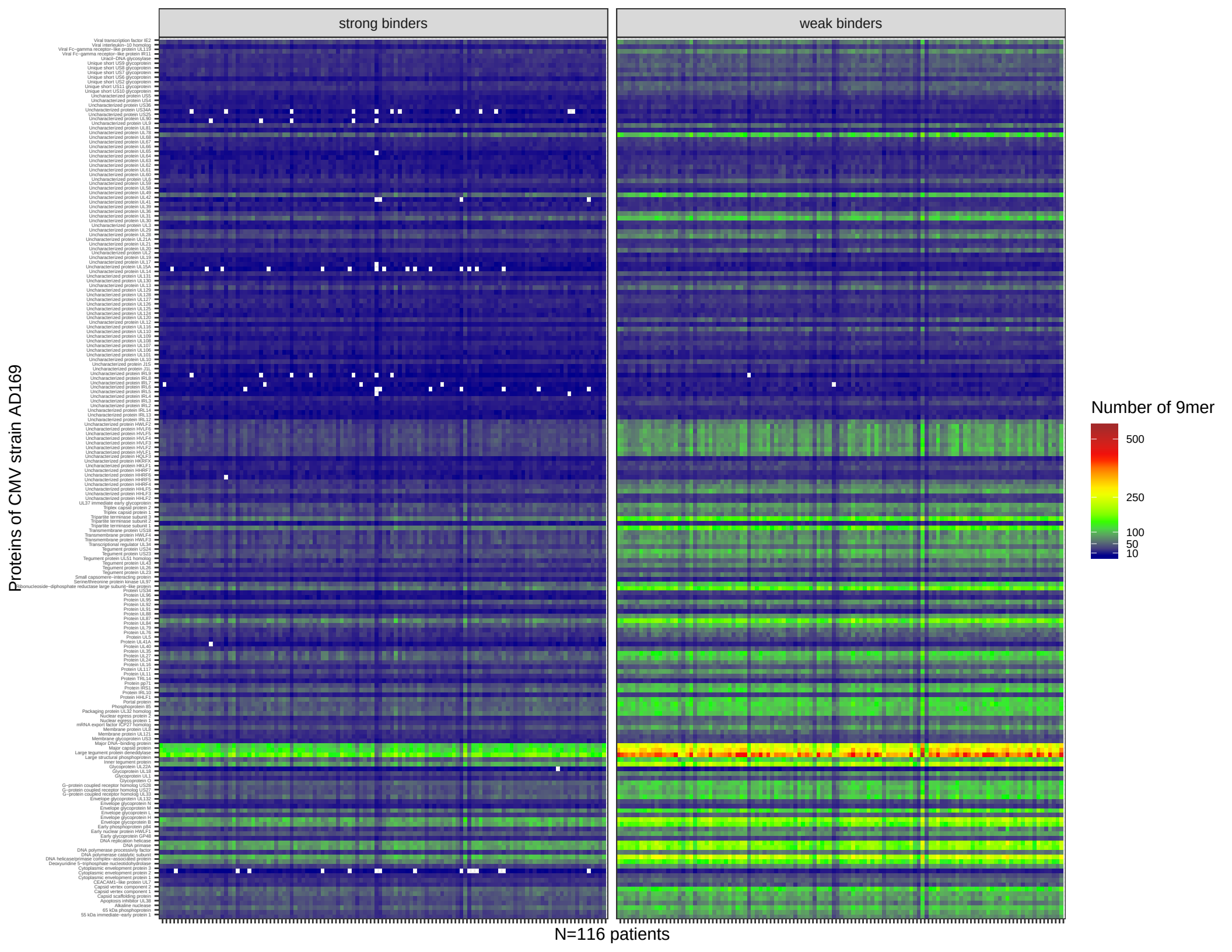
Clonality of the TCR CDR3β repertoire infused with the graft (a) according to age and CMV serologic status of the donor. In (b) the predictive value of clonality on the occurrence of CMV infection/reactivation within the first year post-alloHSCT is assessed. The CMV serologic status of the donor/recipient pair (D-/R-, D-/R+, D+/R- and D+/R+) is indicated by the colored dots.

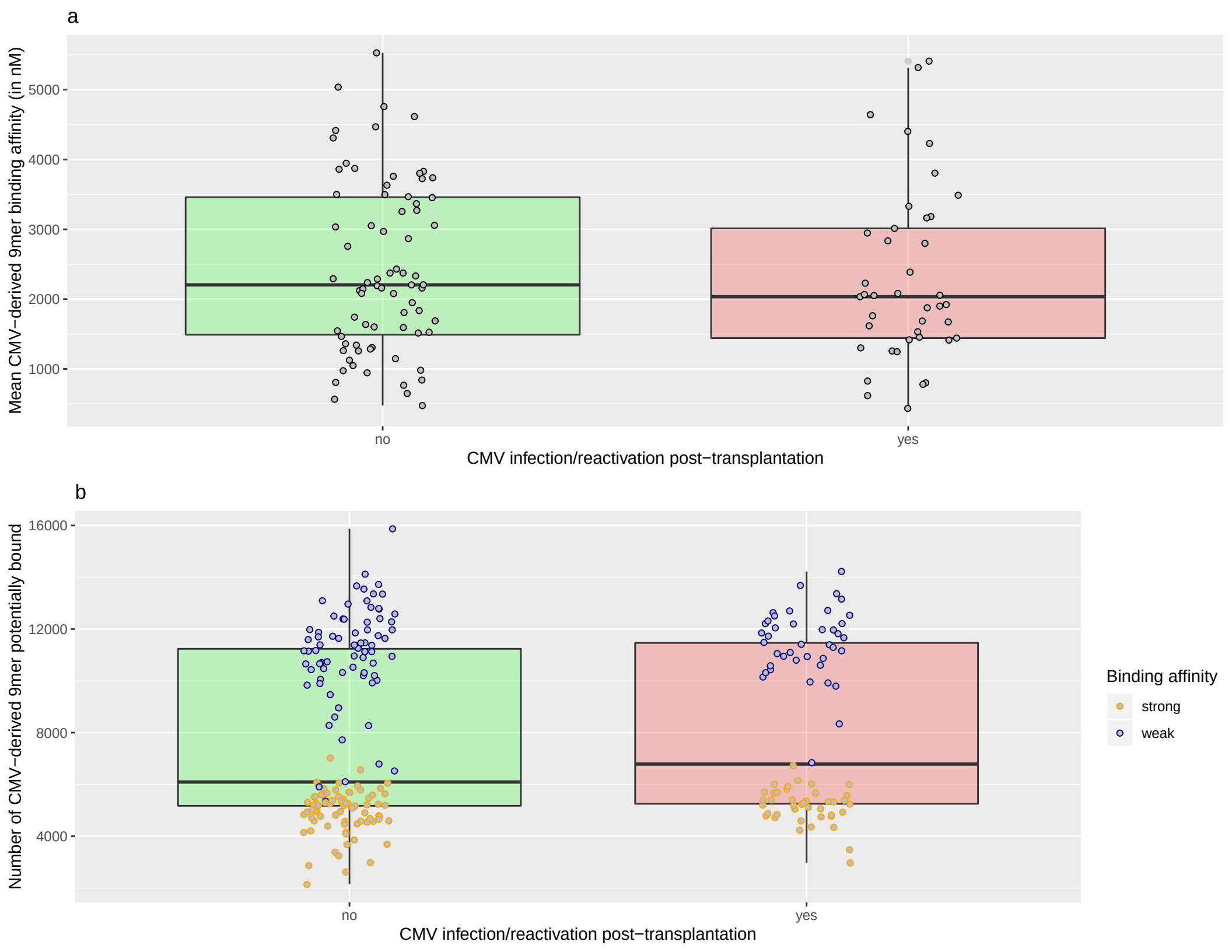

standard.jaccard
generalized.jaccard
morisita
Corr:
0.732
Corr:
0.362
Corr:
0.599
standard.jaccard
generalized.jaccard
morisita
0
10
20
30
40
0.00
0.03
0.06
0.09
0.12
0.0
0.2
0.4
0.6
0.8
0.00
0.02
0.04
0.06
0.08
0.00
0.03
0.06
0.09
0.12
0.0
0.2
0.4
0.6
0.8

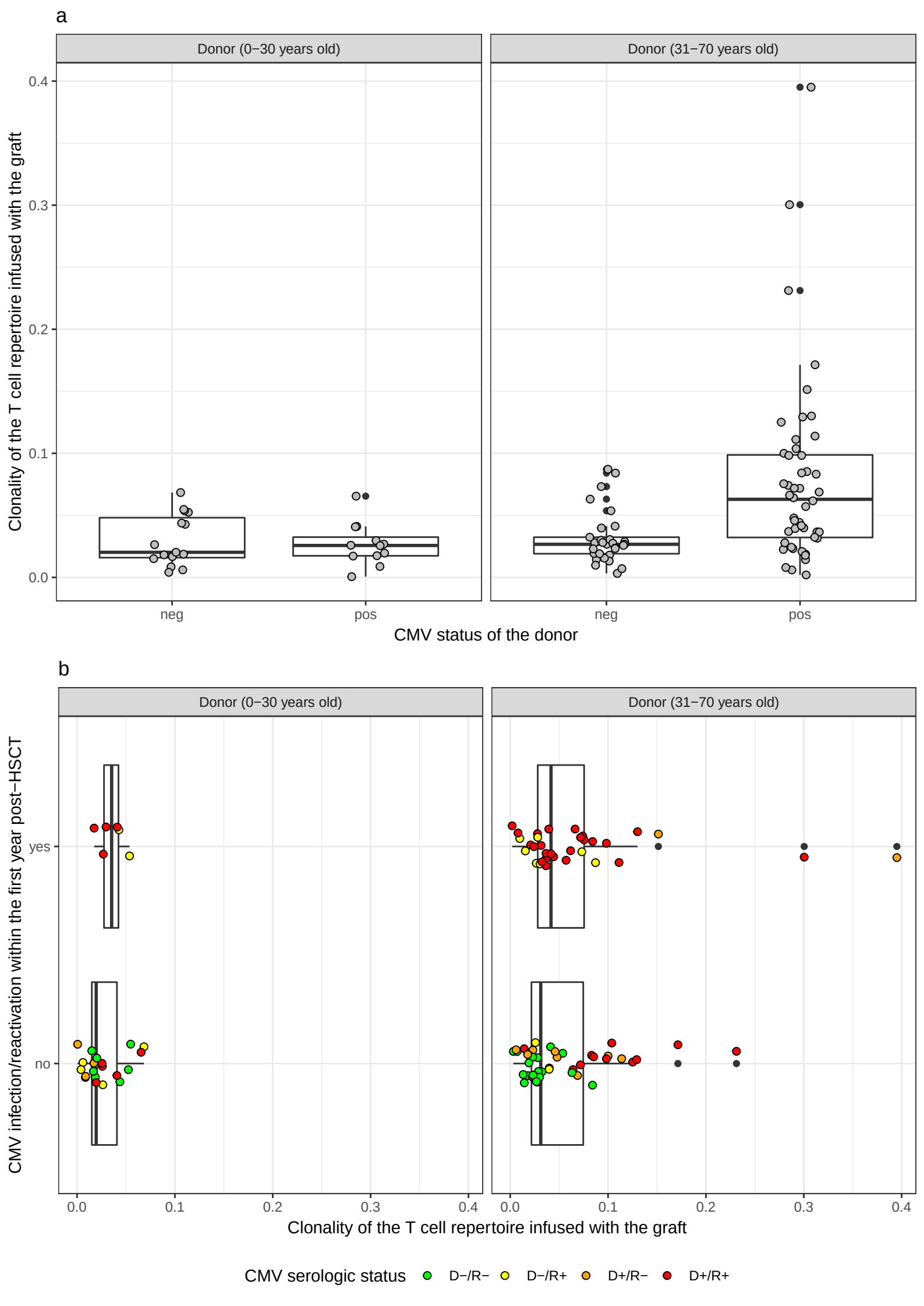

Clonality post−TX − pre−TX
0.5
0.4
0.3
0.2
0.1
0.0
no viral infection
viral infection excluding CMV
CMV infection/reactivation only
CMV and other viral infections


1
CMV_serologic_status
p < 0.001
D+/R+
{D−/R−, D−/R+, D+/R−}
3
CMV_infection_reactivation
p = 0.036
yes
no
Node 2 (n = 49)
Node 4 (n = 13)
Node 5 (n = 54)
0.5
0.4
0.3
0.2
0.1
0


a
CMV clones cumulated frequency post−TX − pre−TX
0.075
0.050
0.025
0.000
0.0
0.1
0.2
0.3
0.4
0.5
Clonality post−TX − pre−TX
CMV infection/reactivation post−TX
no
yes
Regression curves for CMV infection/reactivation post−TX
no
yes
b
density
3
2
1
0
observed r = 0.249
(percentile=0.9997)
−0.4
−0.2
0.0
0.2
simulated r (based on 10,000 resampling)


strong binders
weak binders
Proteins of CMV strain AD169
N=116 patients
Number of 9mer
500
250
100
50
10


a
Mean CMV−derived 9mer binding affinity (in nM)
5000
4000
3000
2000
1000
no
yes
CMV infection/reactivation post−transplantation
b
Number of CMV−derived 9mer potentially bound
16000
12000
8000
4000
no
yes
CMV infection/reactivation post−transplantation
Binding affinity
strong
weak


a
Donor (0−30 years old)
Donor (31−70 years old)
Clonality of the T cell repertoire infused with the graft
0.0
0.1
0.2
0.3
0.4
neg
pos
CMV status of the donor
b
CMV infection/reactivation within the first year post−HSCT
yes
no
Clonality of the T cell repertoire infused with the graft
CMV serologic status
D−/R−
D−/R+
D+/R−
D+/R+