Biol Blood Marrow Transplant 25 (2019) 868–882

Biology of Blood and Marrow Transplantation

journal homepage: www.bbmt.org

T Cell Repertoire Evolution after Allogeneic Bone Marrow Transplantation: An Organizational Perspective

Jeremy A. Meier[1], Mahdee Haque[1], Mohamed Fawaz[1], Hamdi Abdeen[1], David Coffey[2], Andrea Towlerton[2], Ahmed Abdeen[1], Abdullah Toor[1], Edus Warren[2], Jason Reed[3], Christopher G. Kanakry[4], Armand Keating[5], Leo Luznik[6], Amir A. Toor[1,*]

[1] *Bone Marrow Transplant Program, Massey Cancer Center, Virginia Commonwealth University, Richmond, Virginia*
[2] *Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington*
[3] *Department of Physics, Virginia Commonwealth University, Richmond, Virginia*
[4] *Experimental Transplantation and Immunology Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland*
[5] *Princess Margaret Cancer Center, Toronto, Ontario, Canada*
[6] *Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, Maryland*

Article history:
Received 4 September 2018
Accepted 11 January 2019

Key Words:
Bone marrow transplantation
Graft-versus-host disease
T cell repertoire
T cell receptor recombination
Translational symmetry
Euclidean distance

A B S T R A C T

High-throughput sequencing (HTS) of human T cell receptors has revealed a high level of complexity in the T cell repertoire, which makes it difficult to correlate T cell reconstitution with clinical outcomes. The associations identified thus far are of a broadly statistical nature, precluding precise modeling of outcomes based on T cell repertoire development following bone marrow transplantation (BMT). Previous work has demonstrated an inherent, mathematically definable order observed in the T cells from a diverse group of donors, which is perturbed in recipients following BMT. In this study, T cell receptor (TCR)-β sequences from HLA-matched related donor and recipient pairs are analyzed to further develop this methodology. TCR-β sequencing from unsorted and sorted T cell subsets isolated from the peripheral blood samples of BMT donors and recipients show conservation and symmetry of VJ segment usage in the clonal frequencies, linked to the organization of the gene segments along the TCR locus. This TCR-β VJ segment translational symmetry is preserved post-transplantation and even in cases of acute graft-versus-host disease (aGVHD), suggesting that GVHD occurrence represents a polyclonal donor T cell response to recipient antigens. The complexity of the repertoire is significantly diminished after BMT, and the T cell clonal hierarchy is altered post-transplantation. Low-frequency donor clones tended to take on a higher rank in the recipients following BMT, especially in patients with aGVHD. Over time, the repertoire evolves to a more donor-like state in the recipients who did not develop GVHD as opposed to those who did. The results presented here support new methods of quantifying and characterizing post-transplantation T cell repertoire reconstitution.

INTRODUCTION

Bone marrow transplantation (BMT) from an HLA-matched or HLA-haploidentical donor provides a potentially curative modality in disorders of hematopoiesis, as well as hematologic malignancies [1]. Despite substantial advances in HLA matching and transplantation procedures, mortality and morbidity following BMT remain unacceptable, owing in large part to variable immune reconstitution contributing to infections and graft-versus-host disease (GVHD) [2]. Immune reconstitution in general, and more specifically T cell recovery following BMT, are crucial factors in determining clinical outcomes; however, the complexities inherent in this process remain elusive [3]. High-throughput sequencing (HTS) has provided insight into how the T cell repertoire might evolve following BMT [4–8]. The T cell repertoire comprises millions of T cell clones, each bearing unique T cell receptors (TCRs) composed of distinct α and β subunits. These unique TCR subunits are generated by the recombination of gene segments on α and β TCR loci, termed TRA and TRB, and located on chromosomes 14q and 7q, respectively [9,10]. Each TCR locus has variable (V), joining (J), and constant (C) gene segments, and TRB has diversity (D) segments as well. One each of the V and J (and if present, D) gene segments on these loci are recombined during T cell ontogeny to yield unique VDJ rearranged, TCR-α and -β subunits (Supplementary Figure 1). The unique receptor subunits together confer an extraordinarily wide diversity to the human

Financial disclosure: See Acknowledgments on page 879.
* Correspondence and reprint requests: Amir A. Toor, MD, 1300 East Marshall Avenue, Richmond, VA, 23298.
E-mail addresses: amir.toor@vcuhealth.org, atoor@vcu.edu, amirtoor@gmail.com (A.A. Toor).

https://doi.org/10.1016/j.bbmt.2019.01.021
1083-8791/

T cell repertoire enabling recognition of millions of potential antigens, including minor histocompatibility antigens or pathogen-derived antigens, presented on either matched or mismatched HLA molecules in the context of allogeneic BMT. Recognition of antigens by TCRs on T cells leads to corresponding antigen-specific T cell clonal expansion driving immune responses.

In the case of GVHD, HTS has been used to identify and track alloreactive T cells [11,12]. Interpretation of the resulting data has been challenging given the unique antigens present in each individual, as well as a diverse pool of potentially alloreactive T cell clones. However, some trends have emerged; for example, acute GVHD (aGVHD) has been associated with restricted clonal diversity of the T cell population [7,13,14]. This has been evident when assessing tissue infiltrating T cells in cases of aGVHD [15,16], which often are distinct among patients and may or may not correlate with the T cells in circulation [8,16,17]. However, other studies have suggested that aGVHD may lead to a more diverse T cell phenotype [6], possibly reflective of transplant type dependency of the observations. Along with GVHD, other factors, such as cytomegalovirus (CMV) reactivation, also have been associated with impairment of immune recovery and TCR clonality post-transplantation [8,18].

A quantitative understanding of the biology of T cell reconstitution will be valuable to optimize clinical outcomes. This may be possible by interpreting TCR HTS data in terms of the VDJ recombination process. This is necessary because it allows for an intuitive organization of the otherwise complex T cell repertoire to understand the quantitative underpinnings of alloreactivity following BMT. Previous work has demonstrated that lymphocyte recovery post-transplantation follows the rules of population growth as in a dynamical system [19–21]. Whole-exome sequencing studies have revealed numerous sequence differences in the exomes of HLA-matched donors and recipients. The resulting large arrays of potential minor histocompatibility antigens in the setting of different HLA types may yield significant diversity in antigen-driven donor T cell clonal responses [20]. In such a system, antigen-driven T cell clones may grow in proportion to the antigen affinity of the relevant TCR. Thus, examining how the T cell repertoire evolves over time after BMT may provide valuable insight into the relationship between immune reconstitution kinetics and clinical outcomes following BMT with donors with a potentially variable minor histocompatibility antigen burden.

It was previously noted that the T cell repertoire has a self-similar fractal organization in BMT donors and tends to be restored to this state with time after BMT. This is suggestive of an inherent order to determining TCR gene segment use in generating the repertoire [22]. This order appears to derive from the organization of the gene segments along the DNA molecules [9,23]. Thus, understanding T cell reconstitution post-BMT in terms of TRB locus organization may yield novel insight into immune recovery and its relationship to both alloreactivity and infection control.

In this study, TCR sequences from BMT donors and recipients uniformly treated with post-transplantation cyclophosphamide [8] were studied to determine the organization of the T cell repertoire and its evolution over time after transplantation. Methods for comprehensively analyzing the T cell repertoire HTS data were evaluated, and the mathematical properties of the T cell repertoire as they relate to the TRB recombination process were analyzed. The possible impact of maladaptive processes, such as aGVHD, on the underlying T cell repertoire was measured, and the magnitude of transformation of the donor T cell repertoire in the recipient was quantified.

METHODS

Patients and Sequencing Data

The TCR-β sequencing data analyzed in this study was a subset derived from patients included in the original Institutional Review Board-approved multi-institutional prospective clinical trial (ClinicalTrials.gov identifier NCT00809276) as described previously [8,24]. Sequencing results are accessible through the publicly available repository through Adaptive Biotechnologies (Seattle, WA; https://clients.adaptivebiotech.com/pub/Kanakry-2016-JCIInsight). All 16 patients for whom there were sequencing data on both donor (n = 16) and corresponding recipient at 1 month (n = 9), 2 to 3 months (n = 9), 1 year (n = 15), and/or 3+ years (n = 3) post-transplantation were included. According to the original study design [24], patients underwent myeloablative conditioning, consisting of once-daily i.v. busulfan and fludarabine. The subset of patients included in this analysis all underwent HLA-matched related allogeneic BMT and received post-transplantation cyclophosphamide (on days +3 and +4) as single-agent GVHD prophylaxis. aGVHD was scored as described previously [24] using the modified Keystone criteria.

Samples and TCR Sequencing

Genomic DNA from peripheral blood samples was obtained as described in the original study [8], with sequencing of TRB loci at survey-level resolution through Adaptive Biotechnologies using the immunoSEQ platform. Data were reported as copy number for specific clones. The immunoSEQ platform combines multiplex PCR with HTS and a sophisticated bioinformatics pipeline (including data normalization) for analysis [4,25]. For T cell subsets ($CD3^+CD4^+$ and $CD3^+CD8^+$), fluorescence-assisted cell sorting (BD FACSAria II; BD Biosciences, San Jose, CA) was performed before external sequencing as described previously [8]. TRB sequencing analyses of those previously published and publicly accessible data are reported herein.

Self-Similarity and Symmetry Analysis

Self-similarity is a mathematical property of many natural systems in which the structure or geometry of an object appears similar at different levels of magnification, following the same organizational rules (Appendix 1). Common examples of this are the appearance of a tree and of the pulmonary airways, where branching patterns are sustained over several orders of division from the trunk to the leaves in the former and from the trachea to the alveoli in the latter. These relationships are quantitatively described by logarithmic scaling and are characterized by a fractal dimension that remains similar irrespective of the level of magnification at which the structure is evaluated (albeit within the limits of that natural system).

For visualizing the self-similarity in the T cell repertoire at different levels of organization, such as the frequency of unique TRB-J segment- and TRB-VJ segment-containing T cell clones, relative proportional distribution (RPD) graphs were generated as described previously [22]. In brief, total counts for unique VJ recombinations were analyzed with V and J segments organized according to their position along the genome at the TCR-β locus relative to the 5′ centromeric end of the TRB gene [9]. Pseudogenes were included in the analysis, bringing the total number of possible V segments to 65 with the potential to recombine with any of the 13 J segments found at the TCR-β locus. Data were visualized using Excel (Microsoft, Redmond, WA) and represent all TCR-β recombinations for that particular sample. Each ring in the graph consists of a different J segment (J1.1 to J2.7), broken up by the proportion with which each V segment is recombined with each J segment for a total of 100% use. For clarity, in this article we use the term “clonal frequency” to refer to TCR genomic DNA copy number.

Self-similar structures are usually symmetric. Symmetry refers to the property in which an object appears the same when viewed from different perspectives or retains its properties after different operations have been performed on it (Appendix 1). For example, a circle appears similar when rotated through different angles, and an equilateral triangle appears the same when rotated by 120°. Human anatomy demonstrates several examples of symmetry; for instance, the morphology of such structures as the extremities, brain, and lungs exhibits reflection symmetry. The TCR recombination process yields VJ recombined receptor arrays that have similar frequencies of specific J segments recombining with specific V segments, with the relative recombination frequencies maintained between individuals [9]. This represents an example of a specific type of symmetry called translational symmetry (Appendix 1). This means that when viewed from the frame of reference of a unique J segment, the recombination frequencies of that J segment with each of the V segments across the TRB locus would always remain the same (Figure 1). This symmetry will also hold true when unique V segment recombination frequencies with each of the J segments are considered.

To establish the presence of translational symmetry of VDJ recombination, radar plots were generated by taking the clonal frequency of individual VJ recombinations plotted on the logarithmic scale. Each line in the radar plot represents a distinct J segment with unique V segment recombinations

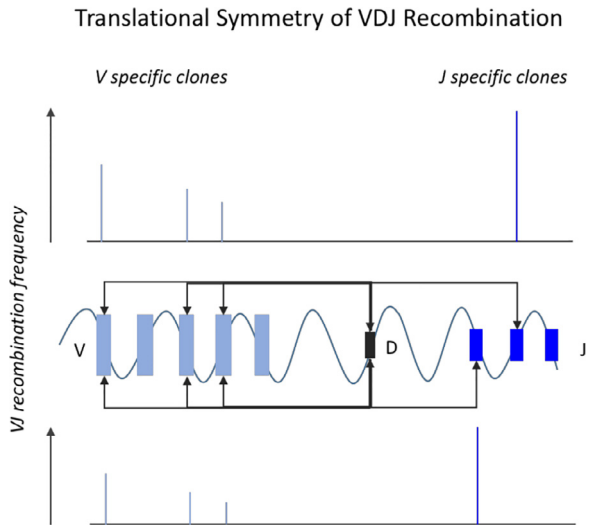


Figure 1. Translational symmetry of TRB VDJ recombination. J segments recombine with specific V segments in similar proportions and vice versa. Bar graphs depict the clonal frequency of specific rearrangements involving the V segments, which vary in absolute magnitude but retain relative proportions when recombining with specific J segments, with D segments acting as an intermediate step.

plotted as the graph progresses across the TRB locus. The corresponding rings in the radar plots denote changing magnitude as it progresses.

Recombination Potentials of TRB V and J Segments

The recombination potential is a method of quantifying the translational symmetry of the TRB J and V segments, while accounting for complexity and depth of the T cell repertoire as a whole. A challenge in accounting for symmetry in the representation of TRB gene segments is the periodic nature of the gene loci across a very long stretch of DNA. To calculate this, the effect of scaling bias (genomic distance between gene segments) must be eliminated. This may be accomplished by accounting for the helical nature of DNA molecules and calculating the angular coordinates of each TRB segment as described previously [9]. In brief, DNA is considered as a propagating wave, with each turn of the DNA helix corresponding to 2π radians in terms of angular distance across the TRB locus. Given that there are 10.4 nucleotides per turn of DNA [26], the angular coordinates of successive TCR gene segments can be determined by the following equation:

$$\theta V_k = \frac{2\pi x_i}{10.4} - \frac{2\pi x_f}{10.4},$$

where θV is the angular distance of the kth TRB V (or J segment) from TRB D2 segments, N represents the total number of segments, x_i represents the 5' initial nucleotide of the TRB-D2 segment, and x_f represents the 3' final nucleotide of the respective V or J segments. The angular distance to the D segments was used in this instance because of the sequence of TRB recombination, which follows J→D and DJ→V sequences.

The recombination potential of each gene segment can then be derived by taking the average of the recombination potential of each J or V segment between donors and recipients,

$$\text{TRB J RecombinationPotential} = \frac{\sum_{V1}^{V30} (\text{Log Clonal Frequency VJ}_i * \cos^2\theta\text{V})}{N}$$

and

$$\text{TRB V Recombination Potential} = \frac{\sum_{J1.1}^{J2.7} (\text{Log Clonal Frequency VJ}_i * \cos^2\theta\text{V})}{N},$$

Where N is the total number of gene segments (V or J).The $\cos^2\theta$V periodic function eliminates the bias introduced by the genomic position of the TRB gene segment on the locus and also, by taking the logarithm of the clonal frequency scales, the variability of clonal frequency across the gene segments being studied. These transformations allow an unbiased view of the contribution of each locus to the repertoire complexity. These analyses were stratified based on aGVHD or recipient CMV serostatus where indicated.

Log-Rank Plots

Log-log plots of assigned rank versus relative clonal frequency of VJ clones were created as described previously [22]. In brief, the relative clonal frequency was calculated by taking the count total of each unique TCR-β VJ recombination and dividing by the sum of all sequence counts for that particular pool to determine the relative use of that clone. Assigned ranks were based on the percent frequency (f) of that clone in the population according to the following scheme:

$$R_{VDJ} = \begin{cases} 1 & \text{if } f \geq 1\% \\ 2 & \text{if } 0.50\% \leq f < 1\% \\ 3 & \text{if } 0.25\% \leq f < 0.50\% \\ 4 & \text{if } 0.15\% \leq f < 0.25\% \\ 5 & \text{if } 0.10\% \leq f < 0.15\% \\ 6 & \text{if } 0.075\% \leq f < 0.10\% \\ 7 & \text{if } 0.06\% \leq f < 0.075\% \\ 8 & \text{if } 0.05\% \leq f < 0.06\% \\ \text{NA otherwise} \end{cases}$$

The absolute value of the slope (a) of the resulting linear regression lines ($y = a \log x + \log k$) reflects the self-similarity of the TCR repertoire and is considered equivalent to the fractal dimension.

Clonal Tracking

Individual TRB rearrangements for BMT donors or BMT recipients at 1 month, 2 to 3 months, or 1 year post-transplantation were arranged according to clonal frequency and sorted based on the donor repertoire to identify those clones that compose the top donor ranks (ie, the top 200 most frequent donor clones); samples from recipients at 3+ years post-transplantation were not included in the analysis. The percentages at which those rearrangements were found in the corresponding recipient rank (ie, the top 200 recipient clones) were determined for all times post-transplantation. Conversely, those clones that populated the top recipient ranks post-transplantation were analyzed for their respective positions in the donor pool of clones. The top 5 recipient clones that were also identified in the donor samples were used for this frequency analysis. These included T cell clones observed in the recipient and ranged in order from clone 1 to clone 68 in the recipient, depending on donor-recipient pair. Clonal tracking results were stratified based on aGVHD history or recipient CMV serostatus.

Clonality, Gini Coefficient, and Diversity Index

Clonality and the Gini coefficient for certain samples were determined using the LymphoSeq package of software tools, available for download at https://bioconductor.org (R package version 1.6.0) as described previously [8]. These measures are expressed on a scale of 0 to 1, with 1 representing a purely monoclonal population. The inverse of Simpson's Diversity Index was used to calculate the diversity of TCR sequencing samples as described previously [6]. It is represented by the following equation: $Diversity = \frac{1}{D_s}$, where $D_s = \frac{\sum n(n-1)}{N(N-1)}$, where n represents unique VJ recombination counts and N represents the sum of all VJ recombination counts in that sample.

Recombination Matrices and Heat Maps

VJ recombination matrices were generated by taking the percent change of total unique sequence counts for each VJ recombination event when comparing donor and recipient at indicated times post-transplantation or by tracking intrarecipient changes over time after transplantation. The resulting percent change values were plotted as a heat map with VJ recombination events of lower frequency represented in green and increasing VJ recombinations depicted in red. Scale was dependent on the particular donor-recipient pair or intrarecipient comparison.

Tracking the Entire Repertoire, Euclidean Distance

To quantify the total change in the clonal frequencies of all the T cell clones in each donor-recipient pair, the entire T cell repertoire (combined donor + recipient at specific time points post-BMT) was considered a multicomponent vector. In this vector, each T cell clone constitutes a vector component, and the change in the clonal frequency (number of reads) from donor to recipient was quantified by measuring the Euclidean distance (ED), using the following equation (see Appendix III for derivation):

$$EDT_n = \sqrt{\sum_{1}^{n} (T_i - T'_i)^2},$$

where EDT_n is the ED between the 2 T cell repertoires, comprising n T cell clones altogether, where the ith T cell clone T_i, changes (transforms) to T_i' over time. EDT_n was divided by the n to yield an average EDT_n. Vectors for measuring the ED were created by combining the donor and recipient T cell repertoires; clones missing in either repertoire were given a value of 0 to allow calculation.

Statistics

Statistical analyses were performed in GraphPad Prism 7 (GraphPad Software, La Jolla, CA) using either multiple t tests (with significance determined

Table 1
Self-Similarity Scores Derived from TRB Clone Copy Numbers Determined from Genomic DNA

	SSS J	SSS V
D011	1.86	1.87
D016	1.83	1.83
D023	1.93	1.95
D031	1.95	1.87
D019	1.63	1.66
Mean ± SD	1.84 ± .13	1.84 ± .11

SSS indicates self-similarity score.

using the Holm-Sidak method with α = .05) or Student's t test (2-sided, α = .05).

RESULTS

Evolution of the T Cell Repertoire after BMT: Self-Similarity

Using TCR-β sequencing data from a group of BMT donor and recipient pairs [8] (Table 1 and Supplementary Table 1), the self-similar organization of the T cell repertoire was evaluated as described previously [22]. TCR-β VJ recombination clonal frequencies (ie, copy numbers) from each donor and recipient time points were arranged in a $n \times m$ matrix in the order of V and J segment arrangement along the TCR-β genomic locus. The clonal frequency of the unique gene segment containing T cell clones consistently demonstrated conservation of the V segment representation across all J segments when the RPD graphs within the donor pool were examined (Figure 2A, top panel, column 1). The self-similarity of the VJ recombination process within the T cell population is evident by the degree to which each ring in the graph aligns with the others, producing a repetitive pattern of TCR-β J recombinations with each respective V segment. Although not identical, the appearance of these rings is similar across the J segment recombinations.

When visualized in the recipients early post-transplantation, the architecture observed in the donors appears to be distorted. The VJ recombination proportions recover with time to assume a more donor-like state 1 year out from transplantation (Figure 2A, top panel, columns 2-4). This suggests that each of the TRB gene segments contributes in a proportional manner to the T cell repertoire as a whole, generating a diverse framework of TCRs.

Translational Symmetry of TRB VJ Recombination

To better identify any perturbations in the T cell repertoire after BMT, radar plots were used to study the clonal frequency of individual VJ recombinations and assess for the presence of translational symmetry in this process. Natural log-transformed clonal frequencies of TCRs bearing unique VJ recombinations were plotted in the radar plots, allowing for direct comparison across multiple recombinations. Translational

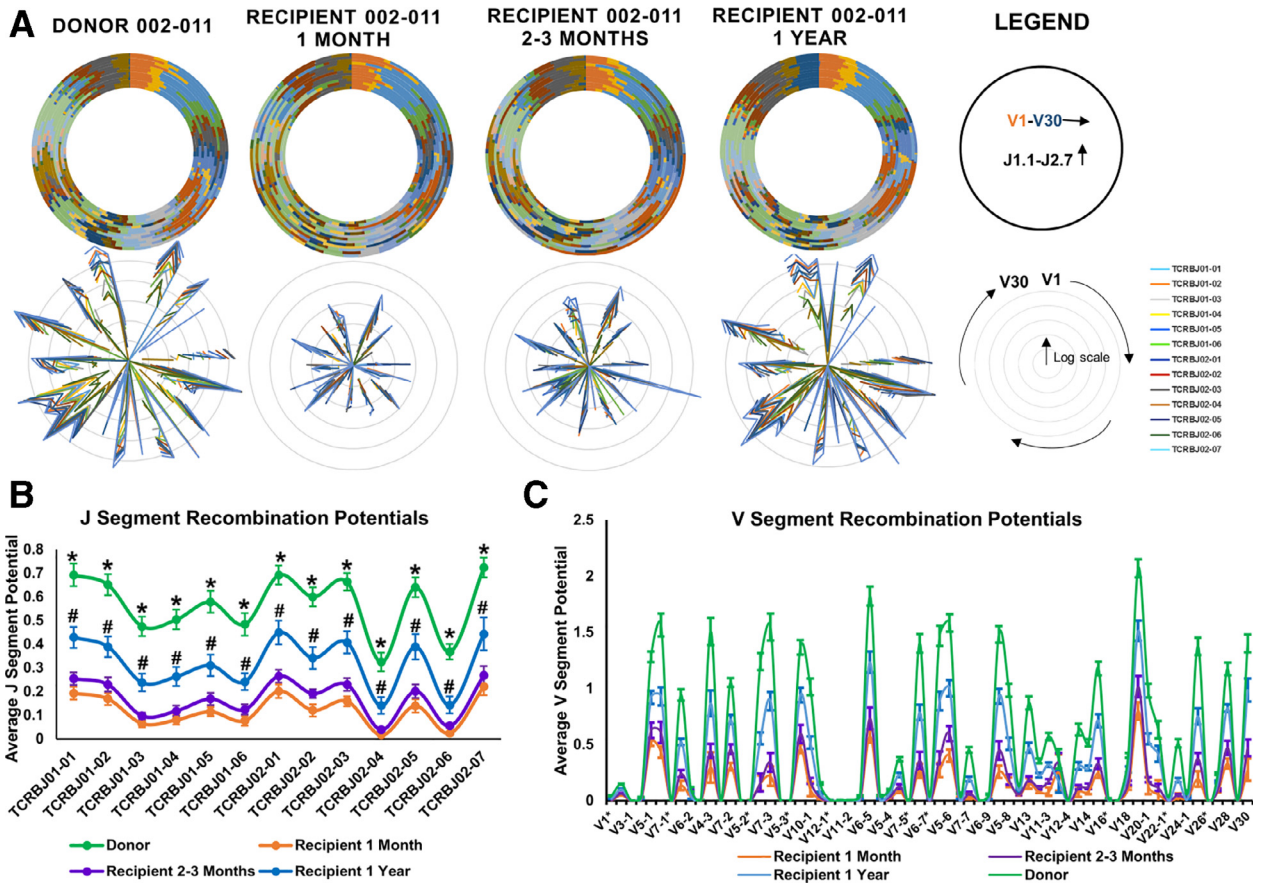


Figure 2. Evolution of the T cell repertoire after BMT. (A) RPD graphs (top) and radar plots (bottom) of a representative BMT donor-recipient pair at the indicated days post-transplantation. Legends for respective figures are on the far right. (B) J segment recombination potential values for donors and recipients at 1 month, 2 to 3 months, and 1-year post-transplantation. *P < .001, donor versus all recipient time points; #P < .05, recipient 1 year versus recipient 1 month and 2 to 3 months (n = 6). (C) Average V segment recombination potentials for donor (n = 13) and recipients at 1 month (n = 7), 2 to 3 months (n = 7), and 1 year (n = 15) post-transplantation.

symmetry was evident across the repertoire when the recombination of each V segment with each J segment is examined in the donor (Figure 2A, bottom panel, column 1). In other words, the frequency of each V segment's recombination with all the J segments is similar relative to its neighboring segments and vice versa. Thus, the recombination frequency of each of the TCR gene segments remains the same across the locus, which manifests as the rings in the radar plots having the same configuration when, for example, all the V segments recombined with unique J segments are analyzed. This suggests that the VDJ recombination, rather than being a stochastic process, may be governed by configuration of the TCR locus.

Further examination of these recombination frequencies reveals a periodicity to how V segments recombine with each J segment across the locus and vice versa. This periodic frequency distribution and the translational symmetry of VJ recombination are still evident following BMT and result in preservation of the "frame" of the TCR repertoire after transplantation (Figure 2A, bottom panel, columns 2-4). However, the overall complexity (or depth) of the repertoire diminishes post-BMT, owing to a reduction in the magnitude of the clonal frequency of each VJ recombination; however, this recovers with time. This suggests that the TCR repertoire has a framework of TRB VJ recombinations present both pre-BMT and post-BMT, with the varying abundance of these recombinations reflecting that seen in the normal setting. This property was observed in all donor-recipient samples analyzed (n = 9). This is comparable to the difference between a tree in the summer when it has its full complement of leaves and in the winter when it is barren. Regardless of season, the basic structure (ie, the trunk and branches) remains intact, akin to the TCR repertoire after transplantation.

Unique Recombination Potential of Each TRB V and J Segment Across the Locus

The foregoing data suggest that the position of the gene segments on the TRB locus appears to determine the probability of each TCR β V segment to recombine with each of the J segments, to form unique TCR-bearing clones [9]. To investigate this, the average recombination potentials were calculated for individual J or V segments (Figure 2B and C). These calculations encompass both the clonal frequency of VJ recombinations and the distance of the segments under study from the *D* segments of the TRB locus. When these recombination potentials were calculated at various times post-BMT, both J (Figure 2B) and V (Figure 2C) segment recombination potentials demonstrated a periodic, oscillatory patterning of TCR segment usage from one end of the locus to the other. The symmetry and periodicity with which TCR-VJ recombinations occur here suggest a predetermined rearrangement scheme dictated by gene segment ordering at the genomic level. As shown in the radar plots, translational symmetry in terms of the frequency of VJ recombinations is maintained after BMT (ie, a similar pattern of gene segment usage) with complexity (ie, clonal frequency reflected by the magnitude of recombination potential) increasing over time (Figure 2B and C). Although recipient J and V segment recombination potentials increase steadily (Figure 2B and C; Supplementary Figure 2A), the complexity of the repertoire still does not recover to that of the donor even 3 years out from transplantation (Supplementary Figure 2B and C). Thus, both the immunodeficient state following BMT and immune recovery over time appear to be relatively broad-based.

TCR Repertoire Architecture Is Preserved within Individual T Cell Subsets

Next, the order observed in the total $CD3^+$ population was evaluated within T cell subsets. $CD3^+CD4^+$- and $CD3^+CD8^+$-sorted cell populations subjected to TCR β locus sequencing demonstrated both conserved organizational principles with evidence of self-similarity when looking at RPD graphs (Figure 3A) and symmetry when looking at the radar plots (Figure 3B). This was evident in donor $CD4^+$ and $CD8^+$ T cell pools, as well as in recipients post-transplantation (Figure 3A and B; Supplementary Figure 3A). This supports the hypothesis that the property of self-similar repertoire organization, with its implications for repertoire diversity if true for T cells as a whole, may extend to all T cell subsets.

The preservation of translational symmetry in terms of TCR usage was also appreciable when evaluating the recombination potential for a representative patient (002-019) at day +242 after transplantation (Supplementary Figure 3B), because the waveforms were unaltered for both $CD3^+CD4^+$ and $CD3^+CD8^+$ cells compared with the total $CD3^+$ population. As suggested in the radar plots (Figure 3B; Supplementary Figure 3A), complexity of the $CD8^+$ repertoire was diminished compared with that of the $CD4^+$ pool (Figure 3C). This was reflected in both the J (Figure 3D) and V (Figure 3E) segment recombination potential values with $CD8^+$ T cells consistently lagging behind the $CD4^+$ T cells regardless of gene segment studied. This methodology recapitulates previous reports demonstrating greater diversity in the $CD4^+$ T cell population [6,27].

Organization of Repertoire Is Maintained in Patients Who Develop GVHD

The impact of the organization of the TCR repertoire and its evolution over time on aGVHD was investigated. Patients who developed aGVHD did not exhibit any obvious distortion of the previously observed symmetry or the time-dependent increase in repertoire complexity when TCR β sequencing data were evaluated with RPD graphs or radar plots (Figure 4A and B). Although it is possible that aGVHD, as well as its accompanying immunosuppressive treatment, may limit the complexity of the developing repertoire, there was no obvious difference in J segment recombination potentials between patients who did and those who did not develop aGVHD (Figure 4C and D). Furthermore, aGVHD did not affect the translational symmetry of the repertoire observed in the recombination potential plots at either 2 to 3 months (Figure 4C, middle panel) or 1 year (Figure 4C, lower panel) post-transplantation. This suggests that in these post-BMT cyclophosphamide-treated patients, aGVHD may either be mediated by a polyclonal expansion of pathogenic T cells or that aGVHD is not associated with disproportionate expansion of a few pathogenic clones in circulation. If such clones are present, they may be distributed in the lymph nodes and tissues.

TRB J (Figure 4C and D) segment recombination potentials at 1-year post-BMT demonstrate a trend toward lower complexity (ie, clonal frequency/magnitude) in patients who developed aGVHD. Examination of the V segment recombination potentials at 1 year post-BMT shows statistically significant differences at some of the loci, with reduced complexity in those patients who develop aGVHD (Figure 4E; Supplementary Table 2). This may be a consequence of both the immunosuppressive regimens administered to treat the GVHD and altered T cell clonal expansion patterns due to GVHD itself. This difference is less evident when repertoires are compared between patients with aGVHD and patients not affected by aGVHD via RPD graphs or radar plots (Supplementary Figure 4A and B).

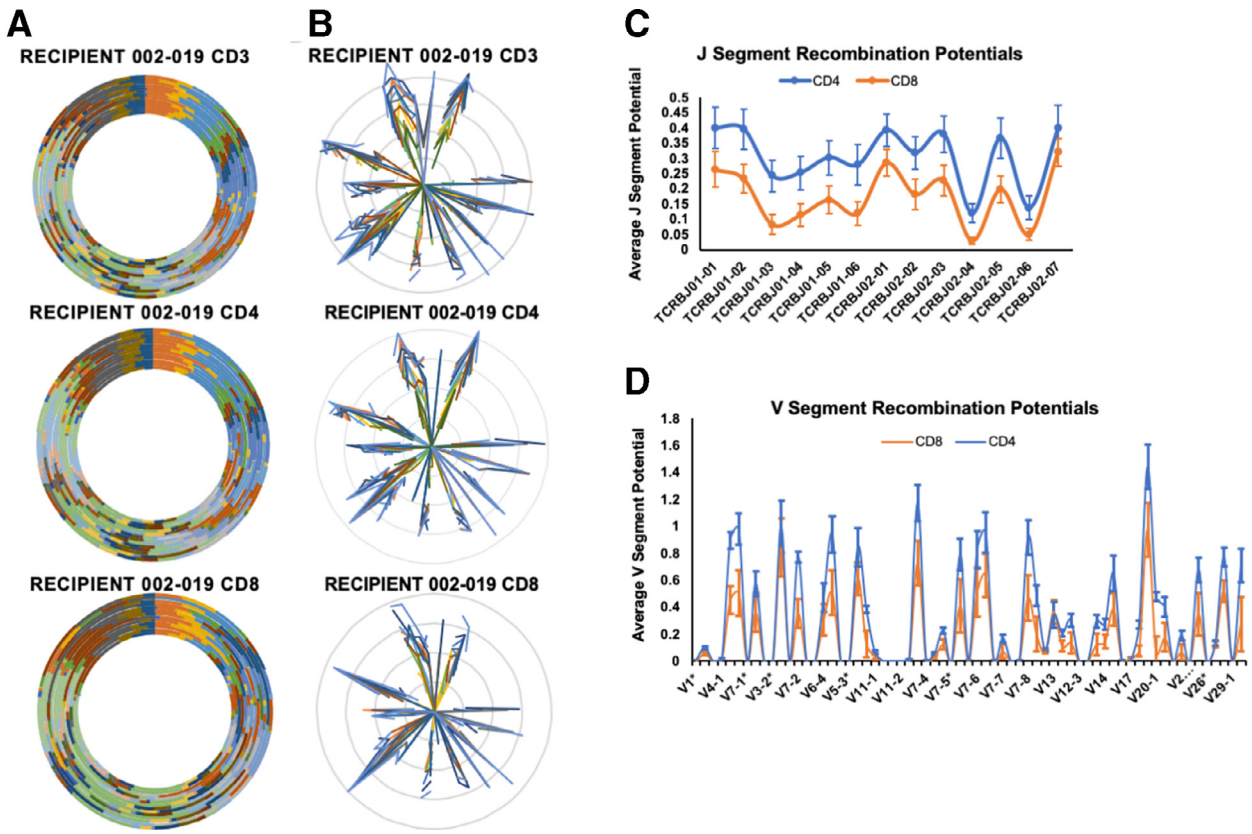


Figure 3. TCR repertoire architecture is preserved within T cell subsets. (A and B) RPD graphs (A) and radar plots (B) of unfractionated $CD3^+$, $CD4^+$, and $CD8^+$ T cell pools in a BMT recipient at day +242 post-transplantation. (C and D) Average J (C) and V (D) segment recombination potentials for $CD4^+$ (n = 5) and $CD8^+$ (n = 5) T cells.

Furthermore, there was no appreciable difference in repertoire complexity when J or V segment recombination potentials were calculated at early time points post-transplantation in patients with aGVHD (Figure 4C, middle panel; Supplementary Figure 4E and F). This paradigm also holds for patients regardless of CMV serostatus (Appendix IV).

Evaluation of differences in donor repertoires revealed no obvious differences in overall repertoire organization or complexity that may be predictive of recipients' tendency to develop aGVHD (Figure 4C, top panel; Supplementary Figure 4G). There was also no difference in the total CD3 counts or absolute lymphocyte counts in those patients who developed aGVHD (Supplementary Figure 5A and B) that might account for the differences observed in recombination potential values at 1 year. Other measures for assessing repertoire complexity did not show any differences between patients with and those without aGVHD (Supplementary Figure 5C and D), even at 1 year post-transplantation. Thus, in some cases, scaling down the repertoire using these measures may obscure relevant differences, making the methodologies reported here an important additional tool in measuring T cell diversity and complexity.

When interpreting these analyses, it is important to recognize the impact that these calculations on scaling of T cell clonal frequencies. Log transformation of the T cell clonal frequencies may obscure relative clonal expansion related to a specific clinical event, such as CMV reactivation or GVHD. However, these findings suggest that the underlying T cell clonal framework is maintained and may provide the basis for eventual recovery of the T cell repertoire.

Self-similar distribution and organization of the T cell repertoire may be evaluated through log-transformed data [22,28]. Consistent with the results from the analysis of the RPD graphs, radar plots, and symmetry calculations, there was no significant effect on the clonal hierarchy in patients who developed aGVHD as assessed by log-rank plots (Figure 5A). Slopes of these plots showed near uniformity, reflective of the preserved self-similar nature, with Power Law clonal frequency distribution of the T cell pool in patients with GVHD and those without GVHD, regardless of post-transplantation time (Supplementary Table 3), again suggesting that aGVHD may be a polyclonal process.

Low-Frequency Donor Clones Expand in Recipients Post-BMT

To determine whether potentially alloreactive T cell clones likely to be present in the donor at a low frequency proliferate and become dominant mediators of alloreactivity in the recipient, the T cell clones constituting the top recipient ranks (ie, highest-frequency clones post-BMT) were evaluated. Patients who developed aGVHD tended to have a greater proportion of their top-ranked T cell clones present in the donor at a much lower frequency or rank compared with the patients who did not develop aGVHD (Figure 5B). Reciprocally, significantly fewer of the infused highest-frequency donor T cell clones were found in recipients who developed aGVHD (Figure 5C), particularly in the early post-transplantation period, coinciding with the time frame of aGVHD pathogenesis. This supports the idea that lower-ranked donor T cell clones may proliferate to a greater extent in recipients with aGVHD, likely on encountering alloreactive antigens not present in the donor. This is consistent with the idea of T cell "vector transformation" post-transplantation, as in a dynamical system [20,29].

Regardless of GVHD status, there is a change in the recipient repertoire compared with the donor repertoire when

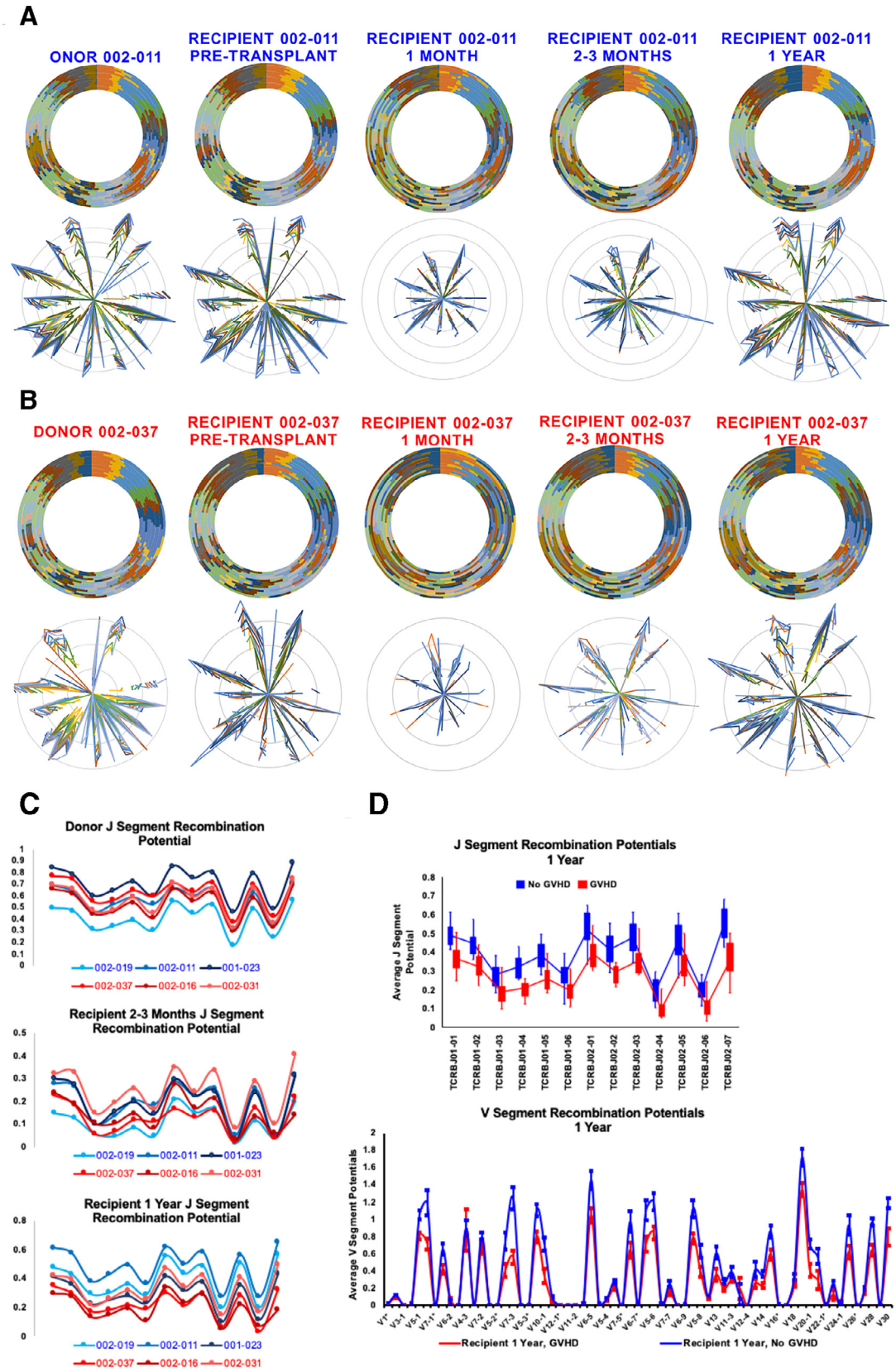


Figure 4. Repertoire organization is maintained in patients who develop aGVHD. (A and B) BMT donor-recipient pairs at indicated time points with RPD graphs (top) and radar plots (bottom) in representative patients without GVHD (blue) (A) and with aGVHD (red) (B), demonstrating comparable reconstitution of the T cell repertoire. (C) J segment recombination potential values for donors (top) and BMT recipients at 2 to 3 months (middle) or 1 year (bottom) post-transplantation in patients with aGVHD and patients without aGVHD. (D) Box-and-whisker representation of recipient 1-year J segment recombination potential values for patients with aGVHD and patients without aGVHD across all J segments (n = 7 for GVHD and n = 8 for GVHD absent). (E) Average 1-year V segment recombination potential values for patients with aGVHD (n = 7) and patients without aGVHD (n = 8). *P* values (aGVHD versus no aGVHD) are presented in Supplementary Table 2.

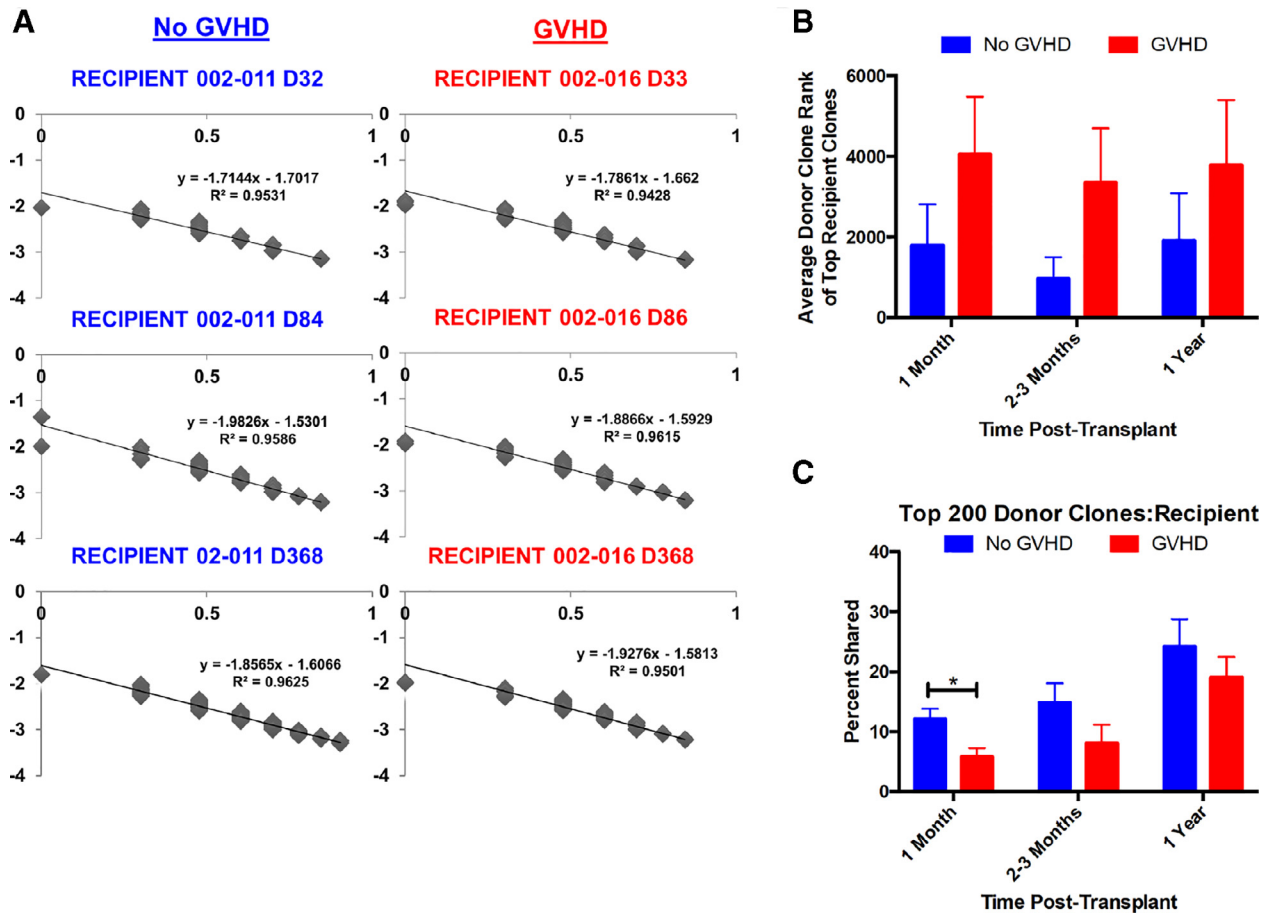


Figure 5. Self-similarity and clonal hierarchy in aGVHD. (A) Log-log rank plots of representative BMT recipients without (left, blue) or with (right, red) aGVHD at the indicated days post-transplantation. Linear regression analysis is included in each plot, where the *x*-axis represents log (relative clonal frequency) and the *y*-axis represents log (clone rank). (B) Average donor rank of the top 5 recipient recombination events (based on productive frequency and shared with donor) in patients with aGVHD (n = 4) and patients without aGVHD (n = 4) ($P = .13$, 2 to 3 months no GVHD versus GVHD). (C) Percentage of top 200 donor clones expressed in the corresponding recipient rank at the indicated time points post-transplantation in patients with aGVHD (n = 4) and patients without aGVHD (n = 4). $^*P < .04$.

comparing donor repertoires with recipient (Figure 6A) or intrarecipient (Figure 6B) repertoires over time using V × J matrices to depict changes in T cell clonal frequency. In both cases, there was an obvious shift in the repertoire hierarchy relative to the donor, with a significant decline in T cell populations early on, as well as later broad-based T cell clonal recovery seen in recipients after BMT.

To quantify the transformation in the overall donor T cell repertoire on transplantation into the recipient, the average EDs of the donor and recipient repertoires were calculated at each time point for which T cell repertoire sequencing data were available. Considering that these patients were all fully donor chimeric in blood or T cells [24], the high ED between donor and recipient T cell repertoire represents a polyclonal proliferation of low frequency and rare donor T cell clones in the recipient and a decline in the dominant clones. A significantly lower ED was measured at 3 months in the patients who developed GVHD when the entire repertoire and the top-ranked T cell clones were analyzed (Figure 7A), possibly consistent with a limited number of alloreactive donor memory T cell clones expanding on encountering recipient antigens and, conversely, a broader-based, more rapid polyclonal T cell reconstitution in patients without GVHD. Given the Power Law distribution of the T cell repertoire, this analysis was restricted to the 10% T cell clones with the highest frequency in the recipient at any given time point. In this analysis, by 1-year post BMT, the average ED had increased significantly in the patients with GVHD, despite an overall increase in the number of T cell clones in all patients (Figure 7B and C; Supplementary Tables 4 and 5). This suggests that patients who do not develop GVHD eventually tend to assume a more donor-like T cell repertoire when their highest-frequency T cell clones are examined and vice versa. In terms of vector transformation, this represents a greater magnitude of vector transformation in patients with GVHD, likely driven by alloreactive antigens.

Using conventional analytic techniques, there was no increase observed in repertoire clonality (based on entropy and Gini coefficient calculations) in patients who developed aGVHD (Supplementary Figure 6A and B). Furthermore, there also were no differences in donor repertoires when stratified by recipient GVHD status (Supplementary Figure 6C and D).

DISCUSSION

HTS of the TCR in BMT donors and recipients demonstrates the incredible complexity of the human T cell repertoire in terms of the T cell clonal makeup and the relative clonal frequencies. When analyzed by conventional statistical methodologies, T cell repertoire analysis fails to yield significant insight into the pathophysiology of GVHD. In this work, these data were analyzed with respect to the TCR VDJ segment positions on the TRB locus. This analysis demonstrates that the T cell clonal hierarchy is defined mathematically, showing

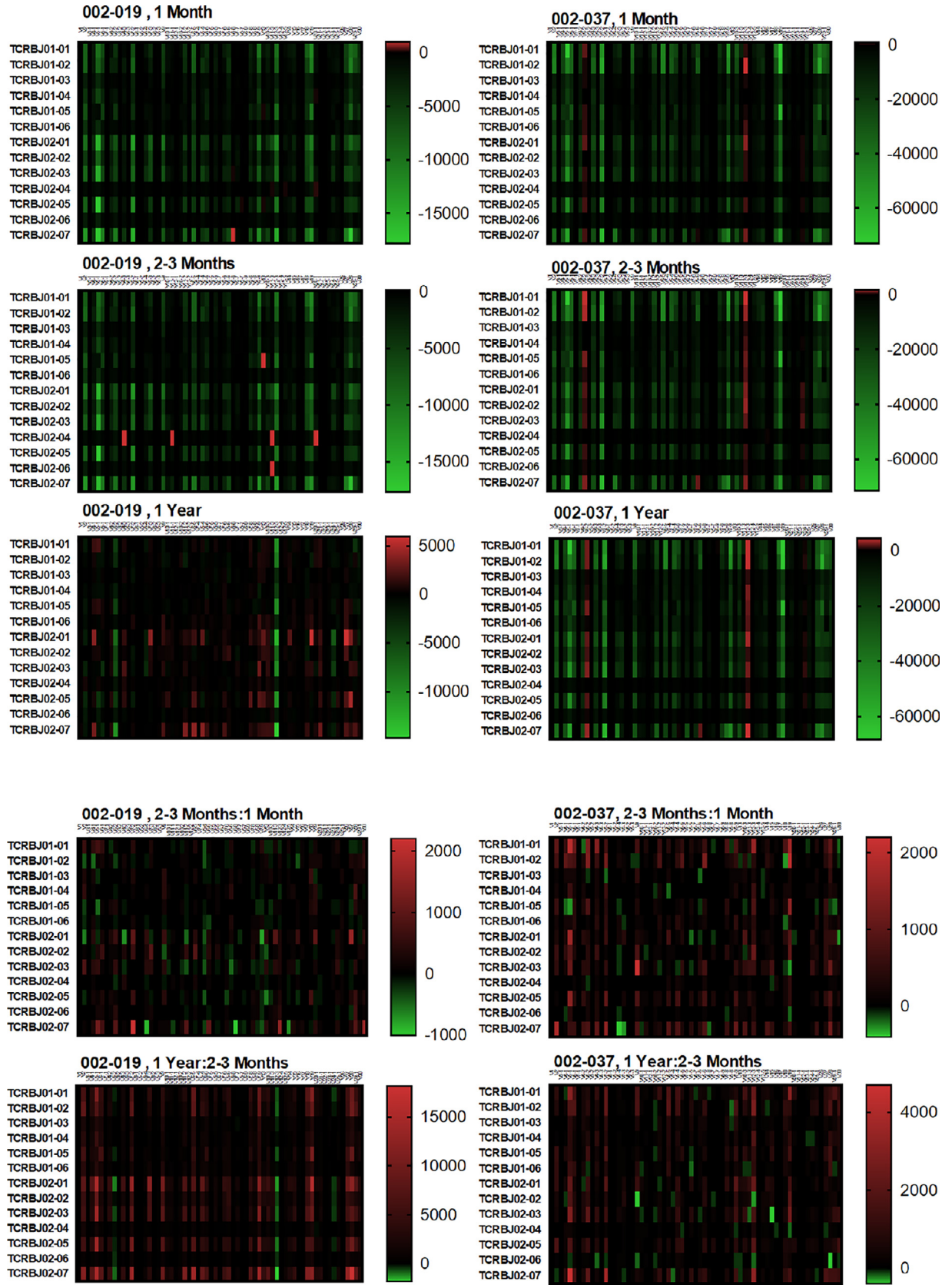


Figure 6. Donor-recipient T cell clonotype differences and dynamic changes post-transplantation. (A) Representative heat maps of VJ recombination matrices demonstrating percentage change between donor and recipient at 1 month, 2 to 3 months, and 1 year post-transplantation across all potential VJ recombination events. Recipients with GVHD are depicted in the right panels (n = 5). Green, decreased; red, increased. (B) Intrarecipient VJ recombination matrices demonstrating percentage change in VJ recombination events within the indicated recipients over time (n = 5).

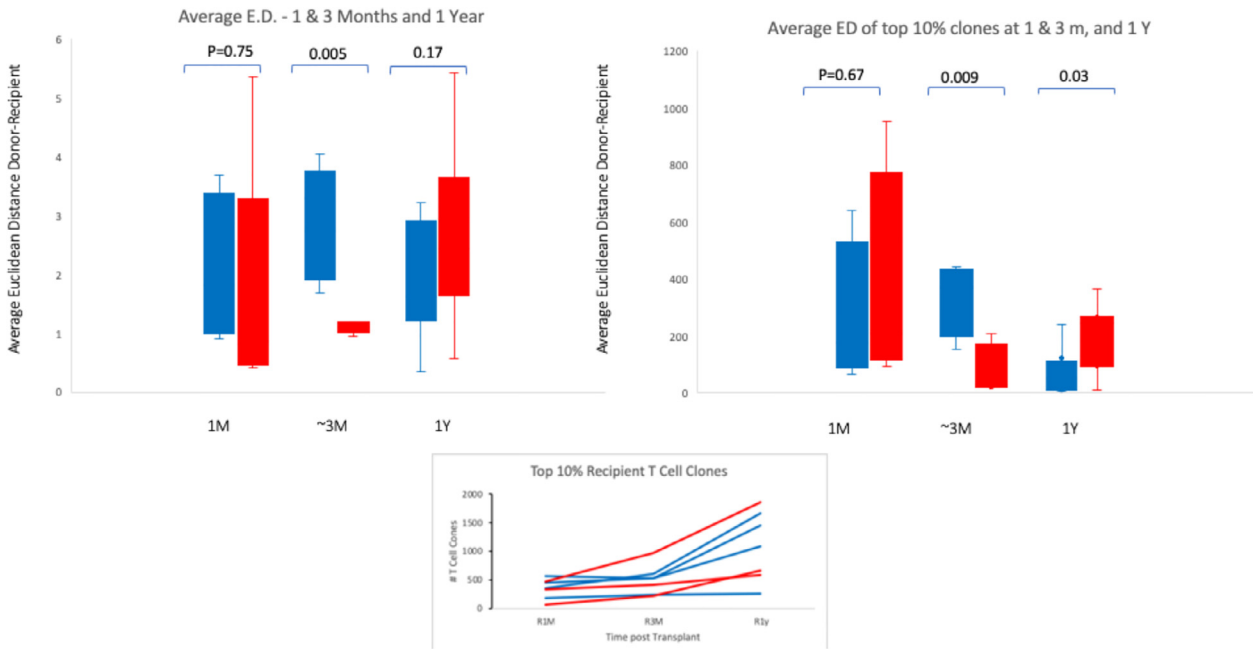


Figure 7. ED between donor and recipients at different times post-BMT. (A) The entire repertoire ED between donor and recipients in patients with no GVHD (blue box/whisker; n = 4 at 1 and 3 months, n = 8 at 1 year) versus those with GVHD (red; n = 4 at 1 and 3 months, 7 at 1 year); (B) ED between donor and recipient T cell repertoire for the 10% most frequent recipient T cell clones (no GVHD/GVHD, n = 4/4 at 1 month, 3/5 at 3 months, and 8/7 at 1 year). (C) Change in number of T cells constituting the top 10% of the recipient vector at each time point (n = 4 with no GVHD; n = 3 with GVHD).

features of self-similarity, periodicity, and translational symmetry.

Despite the T cell depletion observed in the recipient following post-transplantation cyclophosphamide administration compared with the donor, there is broad-based but proportional representation of V and J segments from the entire TRB locus. This organization spans the T cell subsets and shows growing complexity over time after BMT, as T cell repertoire recovers. Regardless of whether patients developed aGVHD, the basis of repertoire organization remained unchanged; however, within this broad clonal distribution, the hierarchy of T cell clones was changed as low-frequency donor clones proliferated after BMT. This presumably occurs on encountering new antigens in the recipient.

In this admittedly small cohort of patients, recipients who develop GVHD demonstrate a significant departure from the donor T cell clonal hierarchy among the dominant T cell clones. Conversely, the T cell repertoire assumes a more donor-like state in those who do not develop GVHD. Thus, GVHD appears to be associated with a polyclonal expansion of low-frequency potentially alloreactive donor T cells. In short, a dynamic and continuously evolving T cell repertoire is observed after BMT, which makes clonal tracking and monitoring of clonal frequency changes in the post-transplantation setting important to better understand the cellular basis of alloreactivity.

The TCRs consist of α and β subunits, each of which includes V and J, as well as D, segments. Their recombination, along with the addition of nontemplated nucleotides, confers clonal diversity on the T cell repertoire, enabling it to respond to changes in the antigenic landscape. This is a dynamic process, with T cell clonal hierarchy changing in response to the different antigens encountered. After BMT, a completely new library of recipient minor histocompatibility and microbial antigens is encountered by donor T cells, which is likely to change the T cell clonal hierarchy very significantly from the steady state present at the time of stem cell donation [20]. This hypothesis is supported by the finding of a change in the clonal hierarchy reported here, in that T cell clones dominant in the donor declined after transplantation and others present in low numbers proliferated in the recipient.

Also noteworthy is the significant dynamic change seen over time in the clonal makeup of the recipient when the T cell repertoire was compared over time after BMT. This hypothesis of alteration of overall clonal hierarchy is supported by the observation of aggregate T cell clonal frequency at 1-year post-BMT changes, such that the highest-ranking T cell clones circulating in a recipient who develops GVHD diverge significantly from the donor, with the opposite situation prevailing in those who do not develop GVHD. This skewing at the top of the T cell repertoire may be driven in part by the alloreactive antigens present in the recipients who develop GVHD.

The self-similar and symmetric qualities of the T cell repertoire imply that T cell responses occur with mathematical precision, and if the T cell antigen responses are subject to mathematical rules, then logically the process of TCR generation itself may be so founded. Processes governed by mathematical rules will yield quantifiable results; in other words, the T cell repertoire organization cannot be random. TCR sequencing allows evaluation of the T cell repertoire by quantifying the contribution of each gene segment to the receptor makeup of the T cell clones. When viewed from the standpoint of gene segment usage in the T cell population, fractal organization with self-similarity has been demonstrated previously [22,28]. This means that when the T cell clonal frequencies are measured using either J segment usage or VJ segment usage and beyond, similar, but not identical, clonal distribution is observed at each organizational level [22].

This fractal organization also has been observed at the level of the TCR locus [9]. In support of this concept, it also has become recognized that T cell repertoires are remarkably well conserved among individuals [30–32]. A property of self-similar systems is that they demonstrate symmetry. To identify the

symmetry at work in T cell repertoire generation, the process of VDJ recombination may be viewed with reference to the positions of the V and J segments on the TCR locus. This spatial symmetry is referred to as translational symmetry (Figure 1), which for the VDJ recombination process implies that each J segment will recombine with a particular V segment in proportion to its position on the locus and vice versa. This is evident in the radar plots depicted in this article, which plot the clonal frequencies of various recombinations and show concentric rings of V segment-defined clonal frequency when specific J segment containing clones are plotted. Note that these plots are on a log scale, owing to the exponential growth of the T cell clones in response to their cognate antigens. The presence of translational symmetry of VDJ recombination in the normal and post-transplantation T cell repertoires, with diminished complexity of the repertoires, supports the application of quantitative reasoning to more accurately measure broad T cell responses following various methods of immune suppression after BMT.

Although translational symmetry of TRB VDJ recombination (and logically TRA recombination) is relatively well preserved, the repertoire complexity is not. This is likely a consequence of the kinetics of adult lymphoid repopulation after immunoablative therapy in a post-thymic individual but, more significantly, may be an effect of the immunosuppressive therapy following BMT, in this instance post-transplantation cyclophosphamide given for GVHD prophylaxis. This effect was manifested in the loss of magnitude in both the symmetry radar plots and the TCR recombination potential calculations. This complexity was restored over time after transplantation, likely due to both individual clonal growth and increasing clonal diversity. The inference of increasing clonal diversity is supported by the observation that translational symmetry of VDJ recombination is maintained through this process. It is noteworthy that this set of measurements is not dramatically impacted by either GVHD or CMV serostatus; however, it should be noted that although these analyses account for the scaling bias introduced by unique clonal growth at the time of measurement and make it possible to compare measurements across several time points, the isolated clonal T cell expansion occurring in response to unique antigens may be obscured.

The consistent TCR usage observed within these studies among BMT donors and recipients regardless of post-transplantation time points to an ingrained order in which VDJ recombination occurs. One potential confounder of the foregoing results noted in the literature is that these may be the byproduct of PCR amplification bias. However, through the use of primer-optimized multiplex PCR, the sequencing platform used to generate these results can minimize bias through computational corrections [25]. Moreover, the observations reported here are very similar to those found in a previous model of J segment rearrangement biases that used a mathematical model to faithfully predict J segment usage in a murine model [23]. In that report, the authors used transcript data from antigen-experienced human T cells, as well as $CD4^+$ and $CD8^+$ cells, in an attempt to determine the relative frequencies of J segment usage in humans. The HTS results presented here are similar to those frequencies, indicating that gene segment usage both at the TRB V and J loci follows an inherent pattern of recombination and is not just an artifact of methodology. Moreover, this is appreciated even early on after BMT and evolves in this manner throughout the post-transplantation period in all donors and recipients studied. The driving force behind this is likely the organization of the DNA itself [9,26,33,34], whereby its physical structure and fractal arrangement may increase the probability of certain VDJ rearrangements occurring and thus contributing more substantially to immune reconstitution post-transplantation. It is conceivable then that this observation extends beyond TCR recombination and is important to consider in a variety of biological processes. It also explains why the overarching architecture seen within the total CD3 population also persists at the $CD4^+$ and $CD8^+$ subset levels.

The methodology presented herein for analyzing the TCR repertoire provides a novel means of comprehensively analyzing the entire repertoire and the dynamics of its evolution over time while accounting for its inherent complexity. aGVHD has been associated with changes in the TCR repertoire when analyzed via HTS [7,13,14,16]. In this cohort of HLA- matched related donor transplant recipients whose sole GVHD prophylaxis was post-transplantation high-dose cyclophosphamide, relative preservation of the T cell repertoire architecture in patients with GVHD is somewhat counterintuitive when a comparison is made with classical measures, such as clonality and repertoire diversity [6,7,14,16]. These widely used measures of clonal diversity usually summarize information by averaging out the changes in the T cell repertoire, and as such provide a low-resolution assessment of the problem at hand. On the other hand, clonality assays focus on high-frequency clones, which in a Power Law-distributed population would also give a biased view. The analysis presented here is an attempt to bridge the gap between these 2 broad methodologies. The preservation of symmetry observed here implies that pathogenic T cell clones do not have a preference for one TCR VDJ configuration over another.

This plurality of potentially pathogenic T cell clones mirrors the very large array of potentially alloreactive peptide antigens bound to HLA molecules in donor-recipient pairs [29,35]. Furthermore, the range of T cell clonal frequencies observed here is similar to the variability in HLA-binding affinities of the alloreactive peptides in these pairs. This also suggests that although T cell growth post-transplantation may follow mathematical precision, the likelihood of alloreactivity is still a stochastic function of alloreactive T cell clones present in the infused allograft and the antigens presented. All these observations support the notion that GVHD may be a polyclonal T cell response to antigenic disparity between donor and recipient, and not necessarily an oligoclonal process with a limited number of dominant T cell clones. This implies that even a “weak” antigen may trigger a pathological graft-versus-host response in a particular tissue. This tissue injury response may then be amplified by other T cell clones that target different alloreactive antigens presented after the initial insult.

Is it possible that these quantitative relationships are influenced by the GVHD prophylaxis regimen used? Because all the patients in this study were treated with post-transplantation cyclophosphamide only, there is a possibility that the lack of a significant difference in the T cell repertoire between patients with aGVHD and those without aGVHD could be a consequence of the lack of ongoing immunosuppression. Furthermore, the blood samples were drawn at fixed time points after transplantation, as opposed to at the onset of GVHD, which might have limited the ability to discern the true magnitude of clonal difference induced by alloreactivity. The relatively small cohort of patients and absence of data on TCR-α-defined clonality may further diminish the ability to identify strong clinical associations. Despite these limitations, there was consistent evidence of the lower-ranked donor clones becoming dominant in the recipient to a greater extent in patients who developed aGVHD, which was particularly evident at 1 year

post- transplantation. These analyses do demonstrate a difference in T cell clonal hierarchy in patients with aGVHD. The findings reported here are also limited to the T cells present in circulation, and it should be noted that these might not accurately reflect the clonal distribution in the lymph nodes and target tissues. Therefore, correlative analysis with tissue samples [16,36] in the case of GVHD, considered along with GVHD predictive models using exome sequencing data to elucidate the antigenic background in the context of specific HLAs, may help develop an understanding of GVHD pathophysiology at a cellular level and better predict those patients at risk for GVHD [20,29,35,37].

The conclusions drawn from the study presented here need to be confirmed in a larger cohort of patients in which the samples are drawn at the time that GVHD is observed clinically and viral reactivation and microbiome makeup are accounted for. Furthermore, it is also necessary to study these patterns in patients who receive alternative immunosuppressive regimens rather than post-transplantation cyclophosphamide, to look for any possible differences in magnitude and organization between patients undergoing different GVHD prophylaxis regimens. These limitations of the present study aside, it is important to recognize that this method for TCR repertoire analysis based on the organization of the TCR locus yields an intuitive and mechanistically interpretable framework of repertoire generation, rather than analyzing it as a random collection of T cell clones. Furthermore, this method helps unveil specific quantitative principles governing immune responses such as alloreactivity, based on the TCR recombination process.

Clinical utility aside, this analytic methodology describes several tangible physical qualities of the T cell repertoire when viewed from the TRB VJ recombination perspective, including fractal organization and self-similarity with respect to TRB gene segment representation in the repertoire, translational symmetry of the VDJ recombination process, and another mathematically definable quality, the periodicity of cell clonal frequency across the TRB locus [9,27]. This periodicity occurs at a scale vastly greater than the size of histone subunits composing the TRB locus. It is noteworthy that the resulting clonal frequencies are scaled logarithmically, as is the TRB locus. The nucleotides composing the TRB locus, and by extension DNA molecules in general, are usually represented as numbers on a real number line (positive or negative numbers); however, considering the fractal, periodic organization of the TCR locus [9] and the translational symmetry observed in the recombination process, an inescapable conclusion is that these nucleotide positions on DNA molecules may be modeled using the complex number system. Complex numbers have a real component as well as an imaginary component, taking the form $x + a\sqrt{-1}$, and are traditionally depicted as coordinates in a complex number plane in which the real number component spans the x-axis and the imaginary number component spans the y-axis. Complex numbers are linked to real numbers by the famous Euler identity, $e^{i\pi} + 1 = 0$, where $i = \sqrt{-1}$, and are widely used to represent periodic phenomenon [38]. Considering DNA molecules as oscillating structures makes it possible to apply wave mechanical phenomena, such as interference of different waves, to understand how the TCR recombination process may generate a repertoire thus ordered.

In conclusion, the findings reported here demonstrate that T cell repertoire is not a random collection of stochastically generated receptor-bearing T cells. Instead, these may be described using mathematically definable rules that depend in part on TCR gene segment localization on the TRB locus. Therefore, it is also unlikely that the immune responses following BMT are random events; rather, the evolution of post-transplantation T cell repertoire is a process akin to a dynamical system susceptible to computation. These analytic methods broadly applied in a larger cohort of patients may help develop a better understanding of alloreactivity following BMT.

ACKNOWLEDGMENTS

Financial disclosure: A.A.T. was supported by research funding from National Cancer Center Support Grant P30-CA016059.

Conflict of interest statement: There are no conflicts of interest to report.

Authorship statement: J.M., M.H., and A.A.T. developed the study concept; L.L. and C.K. designed the clinical trial; J.M., M.H., M.F., H.A., A.T., and A.A. collected data; D.C., A.T., and W.H. performed the sequencing; J.M., M.H., M.F., H.A., A.T., A.A., and A.A.T. performed analyses; and J.M., M.H., M.F., H.A., A.T., A.A., D.C., A.T., W.H., L.L., C.K., A.K., J.R., and A.A.T. wrote the manuscript.

SUPPLEMENTARY DATA

Supplementary data related to this article can be found online at doi:10.1016/j.bbmt.2019.01.021.

APPENDIX 1: DEFINITIONS

Self-similarity: The property that an object or quantity is similar, not necessarily identical, at different scales of magnification.

Power law: The rule where a change in one quantity (s) produces a proportional change in the other quantity (N) in accordance to a power (d) of the first variable:

$$N = s^d.$$

These quantities plotted on a log-log scale will yield a straight line, and the slope of the line is the Hausdorf or fractal dimension,

$$d = \frac{\ln N}{\ln s}.$$

Fractal: An object that demonstrates self-similarity. In geometry, Koch snowflake or Mandelbrot set; in nature, coastlines, trees, or vascular branching.

Symmetry: The property that an object remains unchanged under a transformation, such as rotation or reflection.

Translational symmetry: A symmetry in which a quantity or an object will remain unchanged when shifted or moved a certain distance, such as the repeating pattern in a beehive or on a wallpaper.

Function: A quantity that varies in correspondence with another changing quantity.

Periodic function: A function that repeats its values at regular intervals, such as trigonometric functions.

Euclidean distance: The Euclidean distance between 2 points in either a plane or in n-dimensional space measures the length of a segment connecting the 2 points in the plane or the sum of the lengths of the line segments connecting the n points. It is given by the Pythagoras theorem.

APPENDIX 2

To define self-similarity or fractal dimension with respect to TCR beta V segment- and J segment-defined clones, a previous report had examined the relationship between the number of TCR gene segments involved in determining the sequence

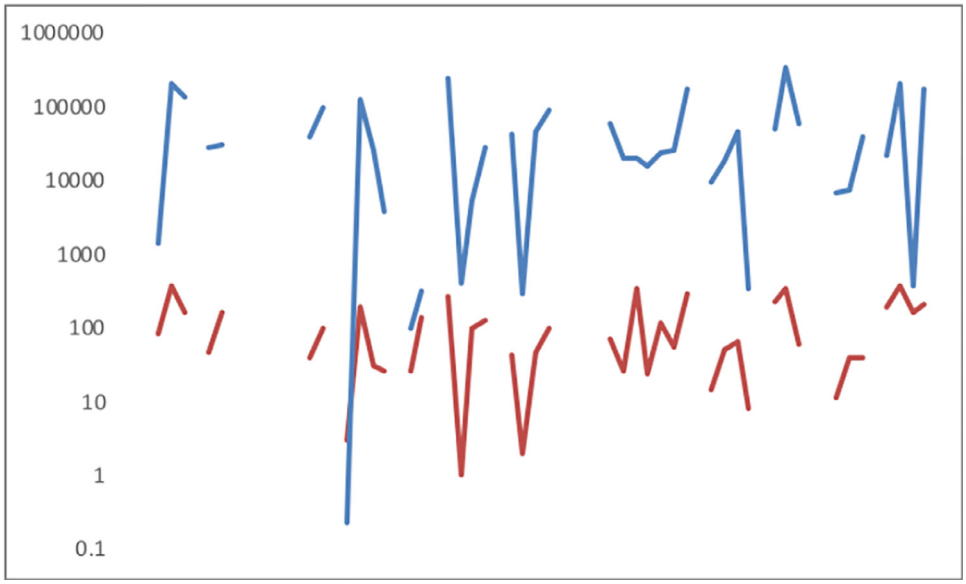


Figure. TRBJ01-01 clonal frequency in combination with each of the TRB V segments measured from genomic DNA (red) and compared with normalized values to approximate cDNA measurements (blue). The *x*-axis depicts sequential TRB V segments as arranged across the locus.

identity and the frequency of individual clones [22]. Specifically, the relative proportion between the clonal populations composed of either specific J segments or VJ segments was examined. To do this, the following relationship was used for J segments:

$$SSJ_{ij} = \underset{k}{mean}\left(\log\left(\sum_{l=1}^{47} X_{ijkl}\right) \times \log(2)\right).$$

For the VJ-containing clones,

$$SSV_{ij} = \underset{k}{mean}\left(\underset{l}{mean}\left(\log(X_{ijkl}) \times \log(3)\right)\right).$$

Here X_{ijkl} represent the sequence count for patient i at time j, for the kth J segment and lth V segment. The current sequencing data used genomic DNA input, compared with the cDNA used previously. Therefore, the sequence reads and reported clonal frequencies were substantially different in the 2 datasets. To normalize this difference in sequence counts between the present study and the previous report, a correction factor was derived as follows. Because coding segments account for only 1% of genomic DNA, this resulted in an initial normalization factor of 100. The current sequencing results were also generated from a PBMC sample as opposed to a pure CD3 population. Because CD3 cells represent roughly 50% of the PBMC pool, this results in an additional normalization factor of 2. Finally, the sequencing depth in the present study was survey level, compared with deep sequencing in the previous report. This resulted in a roughly 5-fold difference in terms of sequencing reads resulting in an additional normalization factor of 5. Because the intralocus distance between segments is important for determining the frequency of recombination events, the 1,000-fold normalization factor was multiplied by $\cos^2\theta$ for each VJ recombination to determine the final correction factor:

Normalization factor for each TRB gene segment

$$= 100 \times 2 \times 5 \times \cos^2\theta V \text{ or } J.$$

Clonal frequency was multiplied by the normalization factor for each data point and average taken as above. The derived self-similarity score for the data presented here (Table 1) approximated the previously reported scores of 1.65 for TRB J segment-containing clones and 1.52 for the VJ-defined clones. The normalized values followed the actual measured values closely as the values were plotted across the locus (Figure). These findings support the overall findings reported herein, reinforcing the hypothesis that T cell repertoire organization has a mathematically determined foundation.

APPENDIX 3: CALCULATING THE ED

The overall goal is to investigate whether there is a relationship between the changes in the clonal frequency of the T cells in the complete T cell repertoire of a donor-recipient pair at 3 different time points to the clinical outcomes of such recipients. To accomplish this, matrices are used to describe the various T cell repertoires studied.

A Euclidean vector is a geometric object with magnitude and direction. We used n x 1 Euclidean column vectors, where n denotes the number of different T cell clones, to represent all the T cell clones present in a donor or recipient. Each component of the column vector describes a specific T cell clone, and the corresponding value describes the frequency of that clone (number of reads in the sample). In this instance, each T cell repertoire was considered as a multicomponent vector evolving in an abstract immune space, as described previously by Abdul Razzaq at el [20] and Koparde et al [29].

To determine how a given T cell repertoire and its associated clones in a donor change in frequency in a recipient posttransplantation at different time points (1 month, 3 months, and 1 year), the distance between the vectors that represent a donor versus the vectors that express a recipient at a specific time point was measured. We did so, by calculating the ED between such vectors. ED is calculated by measuring the change in the dimensions of the vector components (T cell clones in this instance). The vector components are given by

the Pythagoras theorem,

$$c = \sqrt{a^2 + b^2},$$

where c is the vector magnitude and a and b are its components. ED will be calculated as the vector evolves from c to c' and is given by

$$c - c' = \sqrt{(a - a')^2 + (b - b')^2}.$$

The ED for a multicomponent T cell vector will be given by

$$EDT_n = \sqrt{\sum_{1}^{n} (T_i - T_i')^2},$$

Here EDT_n is the ED between the two T cell repertoire, composed of all n T cell clones together, where the ith T cell clone T_i changes (transforms) to T_i' over time. EDT_n may be divided by the n to yield an average EDT_n.

This analysis was performed to determine whether there was a relationship between the change in the T cell repertoire vector magnitude as measured by the ED over time in donor-recipient pairs and the clinical outcome in individual patients.

To do so, the vectors were built for each donor-recipient pair. The T cell repertoire data were imported to a Excel spreadsheet (Microsoft, Redmond, WA) , and unique rearrangements (describing the various T cell clones) were isolated, recording the number of reads (frequency) of T cell clones for the donor and the recipient at 1 month, 3 months, and 1 year post-transplantation. This yielded 4 different vectors of different sizes corresponding to different donor and recipient time points.

To determine the ED between the different multicomponent vectors for specific time points (ie, the change in T cell repertoire between those points), the vectors corresponding to those time points were merged. For the ED calculations for the donor and recipient T cell vectors, this was done by comparing the donor vector with a recipient vector at a specific time point (*Rx*, where *x* = 1 month, 3 months, or 1 year) and removing the shared clones from the recipient. The recipient column without the shared clones was denoted as *Rx, Ry*, or *Rz*. Next, because this new column vector had T cell clones present in the recipient at a measurable frequency but were not detected in the donor sample (likely representing low frequency), a value of 0 was ascribed to these, and they were included along with the donor vector, to make a combined donor vector denoted as *D1, D2*, or *D3*. The same operation was performed for the recipient vectors. This allowed the construction of vectors of similar size, as in the following example:

D0 = Donor

R1M = Recipient 1 month post-transplantation

R3M = Recipient 3 months post-transplantation

R1Y = Recipient 1 year post-transplantation

Rx = R1M \ D0 (Recipient vector − shared clones)

D1 = D0 + Rx = Combined donor with nonshared sequences from recipient 1 month post-transplantation

Full R1M = R1M + D0\R1M

Ry = R3M\D0 (recipient vector − shared clones)

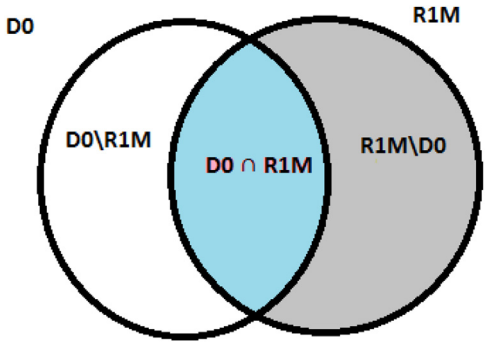
D2 = D0 + Ry = Combined donor with nonshared sequences from recipient 3 months post-transplantation

Full R3M = R3M + D0\R3M

Rz = R1Y\D0 (recipient vector − shared clones)

D3 = D0 + Rz = Combined donor with nonshared sequences from recipient 1 year post-transplantation

Full R1Y = R1Y+ D0\R1Y.



Excel was used to build the vectors. After creating the initial donor and recipient vectors, the TCR rearrangement columns were compared. The conditional formatting function in Excel was applied to highlight the shared clones between the donor and the recipient at specific time points. The filter function was used to filter out shared clones in the recipient. This left the clones that were not shared between 2 repertoires. These clones were given a value of 0 and included in the donor vector to build a combined donor vector (eg, D1) including all the T cell clones for the donor and the recipient time points being studied. The same process was followed to build the combined donor-recipient or recipient-recipient vectors at *x* months post-transplantation (eg, full R1m). To compare the ED between donors and recipients in the 10% highest-ranking T cell clones, the recipient T cell repertoire at specific times was arranged in a descending order of the number of reads, and the top 10% of T cell clones were selected. A comparator donor vector was derived by taking the shared T cell clones from this 10% recipient vector, and the clones not present in the original donor vector were added to this "10% donor vector" with 0 reads ascribed to these clones. This allowed estimation of the ED of the top-ranking recipient clones.

The combined donor and recipient vectors were sorted alphabetically, and the ED between the 2 vectors was calculated using a built-in function in Excel [=SQRT(SUMXMY2)]. Because both the combined donor and recipient vector are composed of the same clones, arranging them alphabetically aligns identical clones in the same order, which allowed for calculation of the ED.

APPENDIX 4: EFFECT OF CMV SEROSTATUS ON MEASURES OF T CELL CLONALITY

Previous studies have noted that CMV serostatus and reactivation are associated with an earlier recovery in lymphocyte counts, with a shift toward an increasingly clonal and less diverse repertoire [8,39]. In examining the organization of the repertoire in BMT recipients with either positive or negative CMV serostatus, J and V segment recombination potentials

showed similar results (Supplementary Figure 4C and D). There was also no clear difference in the complexity or translational symmetry of the repertoire in those patients who experienced CMV reactivation (Supplementary Figure 6A and B).

When CMV serostatus was evaluated as to its effect on the developing T cell repertoire, a greater overlap was observed between donor and recipient repertoires in recipients who had a positive CMV serostatus using this methodology (Supplementary Figure 7A, B, and E). This was accompanied by increased clonality in the repertoire over time, likely reflective of T cell clonal amplification by CMV antigens (Supplementary Figure 7C and D) as described previously [8]. Furthermore, like aGVHD, CMV reactivation does not distort the underlying organizational principles observed when comparing V or J segment recombination potentials (Supplementary Figure 8A and B). This indicates that the mechanism by which the TCR repertoire pool emerges is preserved regardless of disease state.

REFERENCES

1. Kanakry CG, Fuchs EJ, Luznik L. Modern approaches to HLA-haploidentical blood or marrow transplantation. *Nat Rev Clin Oncol*. 2016;13:10–24.
2. Mohty B, Mohty M. Long-term complications and side effects after allogeneic hematopoietic stem cell transplantation: an update. *Blood Cancer J*. 2011;1:e16.
3. Ogonek J, Kralj Juric M, Ghimire S, et al. Immune reconstitution after allogeneic hematopoietic stem cell transplantation. *Front Immunol*. 2016;7:507.
4. Robins HS, Campregher PV, Srivastava SK, et al. Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. *Blood*. 2009;114:4099–4107.
5. Robins HS, Srivastava SK, Campregher PV, et al. Overlap and effective size of the human CD8$^+$ T cell receptor repertoire. *Sci Transl Med*. 2010;2:47ra64.
6. van Heijst JW, Ceberio I, Lipuma LB, et al. Quantitative assessment of T cell repertoire recovery after hematopoietic stem cell transplantation. *Nat Med*. 2013;19:372–377.
7. Yew PY, Alachkar H, Yamaguchi R, et al. Quantitative characterization of T-cell repertoire in allogeneic hematopoietic stem cell transplant recipients. *Bone Marrow Transplant*. 2015;50:1227–1234.
8. Kanakry CG, Coffey DG, Towlerton AM, et al. Origin and evolution of the T cell repertoire after posttransplantation cyclophosphamide. *JCI Insight*. 2016;1:e86252.
9. Toor AA, Toor AA, Rahmani M, Manjili MH. On the organization of human T-cell receptor loci: log-periodic distribution of T-cell receptor gene segments. *J R Soc Interface*. 2016;13 20150911.
10. Mahe E, Pugh T, Kamel-Reid S. T cell clonality assessment: past, present and future. *J Clin Pathol*.. 2018;71:195–200.
11. DeWolf S, Sykes M. Alloimmune T cells in transplantation. *J Clin Invest*.. 2017;127:2473–2481.
12. Berrie JL, Kmieciak M, Sabo RT, et al. Distinct oligoclonal T cells are associated with graft versus host disease after stem-cell transplantation. *Transplantation*. 2012;93:949–957.
13. Alachkar H, Nakamura Y. Deep-sequencing of the T-cell receptor repertoire in patients with haplo-cord and matched-donor transplants. *Chimerism*. 2015;6:47–49.
14. Liu C, He M, Rooney B, Kepler TB, Chao NJ. Longitudinal analysis of T-cell receptor variable beta chain repertoire in patients with acute graft-versus-host disease after allogeneic stem cell transplantation. *Biol Blood Marrow Transplant*. 2006;12:335–345.
15. Koyama D, Murata M, Hanajiri R, et al. Quantitative assessment of TCR repertoire and allo-reactivity of human acute GVHD tissue-infiltrating T cells. *Blood*. 2017;130(Suppl 1):1982.
16. Meyer EH, Hsu AR, Liliental J, et al. A distinct evolution of the T-cell repertoire categorizes treatment refractory gastrointestinal acute graft-versus-host disease. *Blood*. 2013;121:4955–4962.
17. Kubo K, Yamanaka K, Kiyoi H, et al. Different T-cell receptor repertoires between lesions and peripheral blood in acute graft-versus-host disease after allogeneic bone marrow transplantation. *Blood*. 1996;87:3019–3026.
18. Suessmuth Y, Mukherjee R, Watkins B, et al. CMV reactivation drives post-transplant T-cell reconstitution and results in defects in the underlying TCRß repertoire. *Blood*. 2015;125:3835–3850.
19. Toor AA, Sabo RT, Roberts CH, et al. Dynamical system modeling of immune reconstitution after allogeneic stem cell transplantation identifies patients at risk for adverse outcomes. *Biol Blood Marrow Transplant*. 2015;21:1237–1245.
20. Abdul Razzaq B, Scalora A, Koparde VN, et al. Dynamical system modeling to simulate donor T cell response to whole exome sequencing-derived recipient peptides demonstrates different alloreactivity potential in HLA-matched and -mismatched donor-recipient pairs. *Biol Blood Marrow Transplant*. 2016;22:850–861.
21. Toor AA, Kobulnicky JD, Salman S, et al. Stem cell transplantation as a dynamical system: are clinical outcomes deterministic? *Front Immunol*. 2014;5:613.
22. Meier J, Roberts C, Avent K, et al. Fractal organization of the human T cell repertoire in health and after stem cell transplantation. *Biol Blood Marrow Transplant*. 2013;19:366–377.
23. Ndifon W, Gal H, Shifrut E, et al. Chromatin conformation governs T-cell receptor jbeta gene segment usage. *Proc Natl Acad Sci U S A*. 2012;109:15865–15870.
24. Kanakry CG, O'Donnell PV, Furlong T, et al. Multi-institutional study of post-transplantation cyclophosphamide as single-agent graft-versus-host disease prophylaxis after allogeneic bone marrow transplantation using myeloablative busulfan and fludarabine conditioning. *J Clin Oncol*. 2014;32:3497–3505.
25. Carlson CS, Emerson RO, Sherwood AM, et al. Using synthetic templates to design an unbiased multiplex PCR assay. *Nat Commun*. 2013;4:2680.
26. Annunziato AT. DNA packaging: nucleosomes and chromatin. *Nat Ed*. 2008;1:26.
27. Ritter J, Seitz V, Balzer H, et al. Donor CD4 T cell diversity determines virus reactivation in patients after HLA-matched allogeneic stem cell transplantation. *Am J Transplant*. 2015;15:2170–2179.
28. Naumov YN, Naumova EN, Hogan KT, Selin LK, Gorski J. A fractal clonotype distribution in the CD8$^+$ memory T cell repertoire could optimize potential for immune responses. *J Immunol*. 2003;170:3994–4001.
29. Koparde V, Abdul Razzaq B, Suntum T, et al. Dynamical system modeling to simulate donor T cell response to whole exome sequencing-derived recipient peptides: understanding randomness in alloreactivity incidence following stem cell transplantation. *PLoS One*. 2017;12 e0187771.
30. Miyama T, Kawase T, Kitaura K, et al. Highly functional T-cell receptor repertoires are abundant in stem memory T cells and highly shared among individuals. *Sci Rep*. 2017;7. 3663.
31. Hou X, Lu C, Chen S, et al. High throughput sequencing of T cell antigen receptors reveals a conserved TCR repertoire. *Medicine (Baltimore)*. 2016;95:e2839.
32. Britanova OV, Shugay M, Merzlyak EM, et al. Dynamics of individual T cell repertoires: from cord blood to centenarians. *J Immunol*. 2016;196:5005–5013.
33. Almassalha LM, Tiwari A, Ruhoff PT, et al. The global relationship between chromatin physical topology, fractal structure, and gene expression. *Sci Rep*. 2017;7:41061.
34. Metze K. Fractal dimension of chromatin: potential molecular diagnostic applications for cancer prognosis. *Expert Rev Mol Diagn*. 2013;13:719–735.
35. Jameson-Lee M, Koparde V, Griffith P, et al. In silico derivation of HLA-specific alloreactivity potential from whole exome sequencing of stem-cell transplant donors and recipients: understanding the quantitative immunobiology of allogeneic transplantation. *Front Immunol*. 2014;5:529.
36. Beck RC, Wlodarski M, Gondek L, et al. Efficient identification of T-cell clones associated with graft-versus-host disease in target tissue allows for subsequent detection in peripheral blood. *Br J Haematol*. 2005;129:411–419.
37. Hall CE, Koparde VN, Jameson-Lee M, et al. Sequence homology between HLA-bound cytomegalovirus and human peptides: a potential trigger for alloreactivity. *PLoS One*. 2017;12 e0178763.
38. Stipp D. *A most elegant equation: Euler's formula and the beauty of mathematics* 2017.
39. Suessmuth Y, Mukherjee R, Watkins B, et al. CMV reactivation drives post-transplant T-cell reconstitution and results in defects in the underlying TCRß repertoire. *Blood*. 2015;125:3835–3850.