TECHNICAL REPORTS

nature medicine

Quantitative assessment of T cell repertoire recovery after hematopoietic stem cell transplantation

Jeroen W J van Heijst[1], Izaskun Ceberio[2,3], Lauren B Lipuma[4], Dane W Samilo[1], Gloria D Wasilewski[2], Anne Marie R Gonzales[2], Jimmy L Nieves[2], Marcel R M van den Brink[1,2,5], Miguel A Perales[2,5] & Eric G Pamer[1,4–6]

Delayed T cell recovery and restricted T cell receptor (TCR) diversity after allogeneic hematopoietic stem cell transplantation (allo-HSCT) are associated with increased risks of infection and cancer relapse. Technical challenges have limited faithful measurement of TCR diversity after allo-HSCT. Here we combined 5′ rapid amplification of complementary DNA ends PCR with deep sequencing to quantify TCR diversity in 28 recipients of allo-HSCT using a single oligonucleotide pair. Analysis of duplicate blood samples confirmed that we accurately determined the frequency of individual TCRs. After 6 months, cord blood–graft recipients approximated the TCR diversity of healthy individuals, whereas recipients of T cell–depleted peripheral-blood stem cell grafts had 28-fold and 14-fold lower CD4$^+$ and CD8$^+$ T cell diversities, respectively. After 12 months, these deficiencies had improved for the CD4$^+$ but not the CD8$^+$ T cell compartment. Overall, this method provides unprecedented views of T cell repertoire recovery after allo-HSCT and may identify patients at high risk of infection or relapse.

Allo-HSCT is a potentially curative treatment for a variety of hematologic diseases, including lymphoid and myeloid malignancies. Before transplantation, patients undergo conditioning with chemotherapy with or without irradiation, which results in severe immunodeficiency that can take months or years to restore, particularly in the T cell compartment[1,2]. This prolonged T cell deficiency predisposes patients to infection and cancer relapse[3–6]. Strategies that improve T cell reconstitution and recovery of high TCR diversity could therefore greatly reduce transplant-associated morbidity and mortality[7].

Restoration of TCR diversity after allo-HSCT depends heavily on the thymic generation of new naive T cells[8–10]. Thymic function, however, diminishes markedly after the onset of puberty and, in the setting of allo-HSCT, is further impaired by conditioning-associated damage and graft-versus-host disease (GVHD)[11,12]. Thus, it is unclear how well TCR diversity can be restored, particularly in older patients.

Over the past two decades, several strategies have been developed to probe human TCR diversity. One strategy aims to identify the presence of different TCR families by using flow cytometry or PCR to determine the usage of different TCR variable (V) genes[13,14]. A second strategy, called CDR3 size spectratyping, aims to determine polyclonality of the repertoire by using fluorescent primers to measure length variation of the CDR3 region within each TCR V family[15,16]. Spectratyping has been useful in documenting substantial abnormalities in T cell repertoire composition after allo-HSCT[17–19]. However, as neither of these strategies is able to measure the frequency of individual TCRs, they can only provide an estimate of repertoire complexity. With the advent of deep sequencing technology, it has now become possible to directly measure TCR diversity with high resolution[20–26]. Here we built on this approach to address two fundamental questions related to T cell reconstitution after allo-HSCT: how TCR diversity recovers over time and as a function of different sources of stem cells[27,28].

RESULTS

Strategy and reproducibility of T cell repertoire analysis

T cells typically express only one TCR-β chain, making sequence analysis of TCR-β complementary DNA (cDNA) a useful measure of TCR diversity. To faithfully evaluate the TCR-β repertoire, we made use of 5′ rapid amplification of cDNA ends (RACE) PCR, which allows amplification of all 48 Vβ genes using a single oligonucleotide pair. To test the reproducibility of this approach[29], we separately amplified two ~8-ml blood samples from a single patient 138 d after T cell–depleted (TCD) peripheral-blood stem cell transplantation (TCD patient 1; **Supplementary Table 1**). For comparison, we also amplified two blood samples from each of four healthy donors (healthy donors 1–4). We analyzed all samples by Roche/454 sequencing, using a Phred quality score average of 30 to minimize sequence errors[30]. We found all 48 Vβ genes among the total TCR sequences (**Supplementary Fig. 1**), confirming that our approach covers the entire TCR-β repertoire. Comparison of the two blood samples from TCD patient 1 showed a highly reproducible pattern of Vβ usage, which differed markedly from the Vβ usage in the healthy donors (**Fig. 1a**). This suggested substantial clonal expansions in the repertoire of the patient, which we confirmed by digital CDR3 size spectratype profiles (**Fig. 1b**).

[1]Immunology Program, Sloan-Kettering Institute, Memorial Sloan-Kettering Cancer Center, New York, New York, USA. [2]Adult Bone Marrow Transplantation Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York, USA. [3]Complejo Hospitalario de Navarra, Pamplona, Spain. [4]Lucille Castori Center, Molecular Microbiology Core Facility, Memorial Sloan-Kettering Cancer Center, New York, New York, USA. [5]Weill Cornell Medical College of Cornell University, New York, New York, USA. [6]Infectious Diseases Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York, USA. Correspondence should be addressed to J.W.J.v.H. (vanheijj@mskcc.org) or E.G.P. (pamere@mskcc.org).

Received 23 November 2011; accepted 16 July 2012; published online 24 February 2013; doi:10.1038/nm.3100

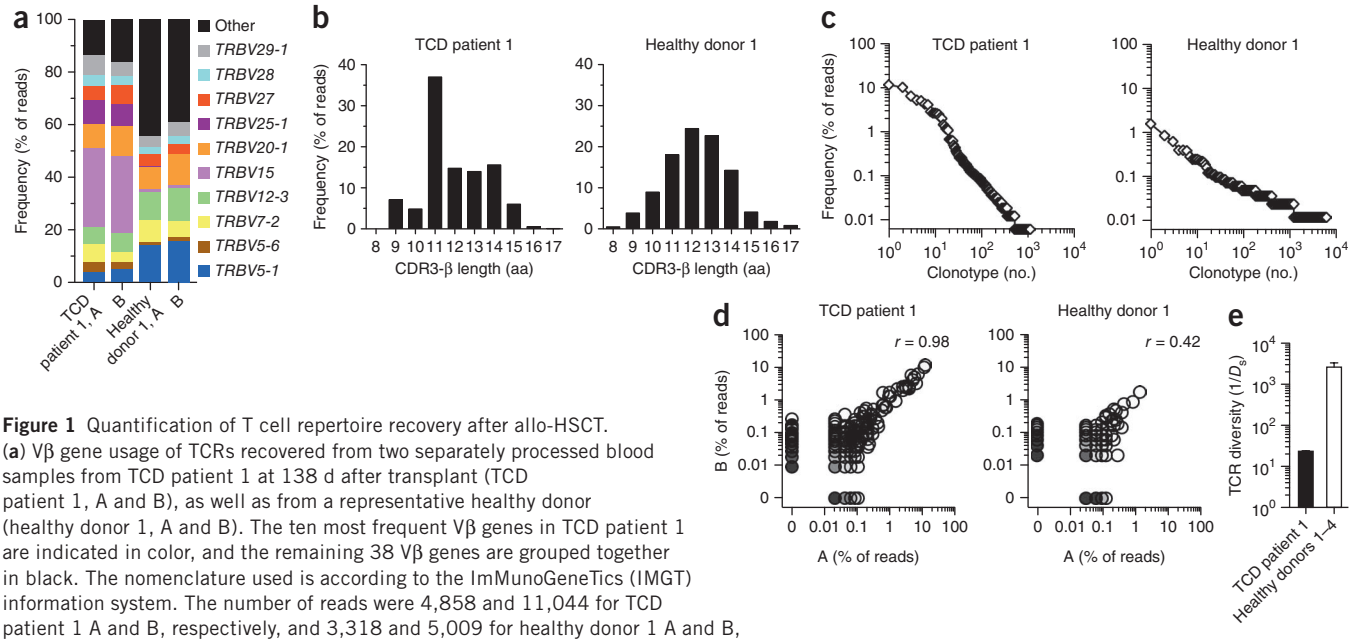

Figure 1 Quantification of T cell repertoire recovery after allo-HSCT. (**a**) Vβ gene usage of TCRs recovered from two separately processed blood samples from TCD patient 1 at 138 d after transplant (TCD patient 1, A and B), as well as from a representative healthy donor (healthy donor 1, A and B). The ten most frequent Vβ genes in TCD patient 1 are indicated in color, and the remaining 38 Vβ genes are grouped together in black. The nomenclature used is according to the ImMunoGeneTics (IMGT) information system. The number of reads were 4,858 and 11,044 for TCD patient 1 A and B, respectively, and 3,318 and 5,009 for healthy donor 1 A and B, respectively. (**b**) Digital CDR3 size spectratype plots of total TCR-β sequences from TCD patient 1 (15,902 reads) and healthy donor 1 (8,327 reads). CDR3-β length is defined as all amino acids (aa) between the conserved 5′ cysteine and the 3′ phenylalanine of the CDR3-β region. (**c**) Clonotype distribution plots of total TCR-β sequences from TCD patient 1 and healthy donor 1. Each diamond represents a distinct CDR3-β amino acid sequence. (**d**) Dot plots comparing the clonotype distribution of two blood samples (A and B) from either TCD patient 1 or healthy donor 1. Each dot represents a distinct TCR-β clonotype. Dot opacity reflects multiple clonotypes of the same frequency. Values in the upper right corners are the Pearson correlations. (**e**) TCR diversity of TCD patient 1, as well as the average TCR diversity of four individually measured healthy donors (healthy donors 1–4). Error bars, 95% confidence intervals (CI).

We then determined the frequency at which each distinct TCR-β chain or TCR-β clonotype was present. In 15,902 reads obtained from both blood samples from TCD patient 1, we detected 1,097 clonotypes, with the most frequent clonotype comprising 11.7% of the reads (**Fig. 1c**). In fact, 19 clonotypes were present at frequencies above 1% and together comprised 70.8% of the reads. In contrast, the most abundant clonotype in healthy donors comprised just 2.8% of the reads, and we found on average only three clonotypes at frequencies above 1% (**Supplementary Fig. 1**). To establish the accuracy of the clonotype frequencies, we compared the clonotype distribution of both blood samples from TCD patient 1. Notably, we found that abundant clonotypes in one blood sample had an almost identical frequency in the other blood sample, resulting in a near-perfect correlation of both clonotype distributions ($r = 0.98$; **Fig. 1d**). In healthy donors, we also detected expanded clonotypes with high reproducibility; however, the average correlation between two blood samples from the same individual was lower ($r = 0.44$; **Supplementary Fig. 1**), primarily because fewer clonotypes passed the threshold for physical presence in a second blood sample (8 ml of ~5 l total, or ~0.16%).

To quantify TCR diversity, we used the inverse Simpson's diversity index ($1/D_s$), which sums the frequency of each clonotype[31]. This index ranges from 1 to ∞ (that is, from no diversity to infinite diversity) and is highest when all clonotypes are equally distributed. To test the usefulness of this index, we sorted naive and memory CD8$^+$ T cells from healthy donor 1 and found a 20-fold higher TCR diversity in the naive T cell compartment (**Supplementary Fig. 2**). Analysis of the repertoire of TCD patient 1 revealed a very low TCR diversity ($1/D_s = 23$), which was more than 100-fold lower than the average diversity of the four healthy donors ($1/D_s = 2{,}525$; **Fig. 1e**). Therefore, at 138 d after transplant, the T cell repertoire of this patient was highly restricted.

To monitor repertoire recovery over time, we measured TCD patient 1 at three additional time points: days 147, 194 and 377 after transplant. To our surprise, TCR Vβ usage was very different at each time point examined (**Fig. 2a**), indicating high variability of the T cell repertoire. To evaluate individual clonotypes, we first determined whether we had reliably measured the clonotype frequencies. On day 147, we again found a near-perfect correlation between the two blood samples from TCD patient 1 ($r = 0.99$), and the same held true for days 194 and 377 ($r = 1$ and $r = 0.98$, respectively; **Fig. 2b**). Comparison of the repertoires on days 138 and 147, however, revealed remarkable shifts in clonotype frequencies (**Fig. 2c**). Although some clonotypes were present at roughly similar frequencies on both of these days, several others differed by more than 100-fold, resulting in a low similarity between the T cell repertoires measured just 9 d apart ($r = 0.24$). Notably, these repertoire shifts coincided with Epstein-Barr virus (EBV) reactivation in the patient, which we first detected on day 145 and which peaked on day 147 (**Supplementary Fig. 3**). We identified the ninth most abundant clonotype on day 147 (*TRBV29-1* and CDR3-β CSVGTGGTNEKLFF) as being specific for the HLA-A2–restricted $BMLF1_{280}$ epitope from EBV[32]. Although undetectable on days 138 and 194, this clonotype comprised 2.9% of the reads on day 147 and 0.1% of the reads on day 377, highlighting the potential of our method to track antigen-specific clonotypes. Despite the apparent resolution of EBV reactivation by day 158, subsequent blood samples continued to reveal clonotype frequencies fluctuating over orders of magnitude. Thus, the repertoires of two samples on the same day were highly similar ($r = 0.99$), whereas on different days, they were markedly distinct ($r = 0.25$; **Fig. 2d**). A notable exception was the repertoires on days 138 and 377, which revealed a surprisingly large degree of similarity ($r = 0.87$; **Supplementary Fig. 3**).

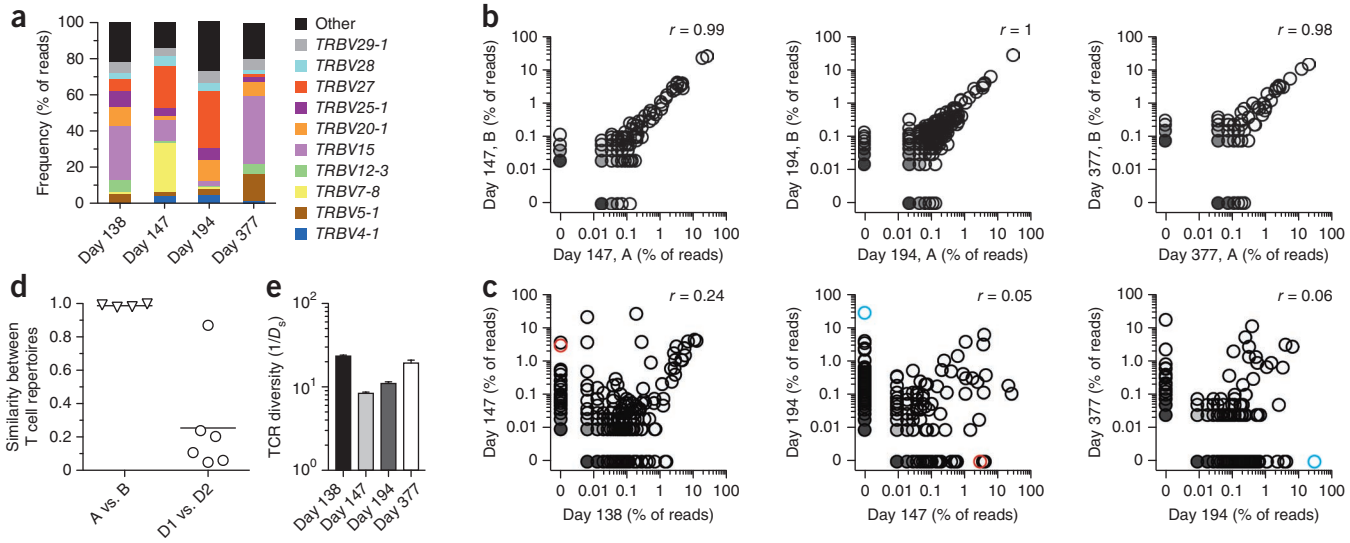

Figure 2 T cell repertoire dynamics during the first year after allo-HSCT. (**a**) Vβ gene usage of TCRs recovered from TCD patient 1 at the indicated time points after transplant. The numbers of reads were 15,902 on day 138, 10,732 on day 147, 11,220 on day 194 and 3,980 on day 377. (**b**) Dot plots comparing the clonotype distribution of two blood samples (A and B) obtained on the same day from TCD patient 1 at the indicated time points. The numbers of reads were 5,644 for A and 5,088 for B on day 147, 4,445 for A and 6,775 for B on day 194 and 2,607 for A and 1,373 for B on day 377. (**c**) Dot plots comparing the clonotype distributions of blood samples obtained on different days from TCD patient 1 at the indicated time points. The red clonotype (*TRBV29-1* and CDR3-β CSVGTGGTNEKLFF) is specific for the HLA-A2–restricted $BMLF1_{280}$ epitope from EBV. The cyan clonotype was below the limit of detection on days 138 and 147, comprised 28% of the T cell repertoire on day 194 and was again below the limit of detection on day 377. The numbers in the upper right corners are the Pearson correlations. (**d**) Similarity of T cell repertoires recovered from blood samples of TCD patient 1 obtained on either the same day (A vs. B) or different days (D1 vs. D2). Values are the Pearson correlations. Horizontal bars depict the group means. (**e**) Diversity of T cell repertoires recovered from TCD patient 1 at the indicated time points. Error bars, 95% CI.

Despite substantial changes in repertoire composition, TCR diversity did not increase over time ($1/D_s = 23$ and $1/D_s = 19$ for days 138 and 377, respectively; **Fig. 2e**). Therefore, between 4.5 and 12.5 months after transplant, the complexity of this patient's T cell repertoire did not improve.

T cell repertoire recovery by different stem cell sources

To evaluate whether our method might enable the stratification of allo-HSCT recipients according to their T cell repertoire, we measured TCR diversity in 27 patients at either 6 or 12 months after conventional or TCD peripheral-blood stem cell transplantation or double-unit

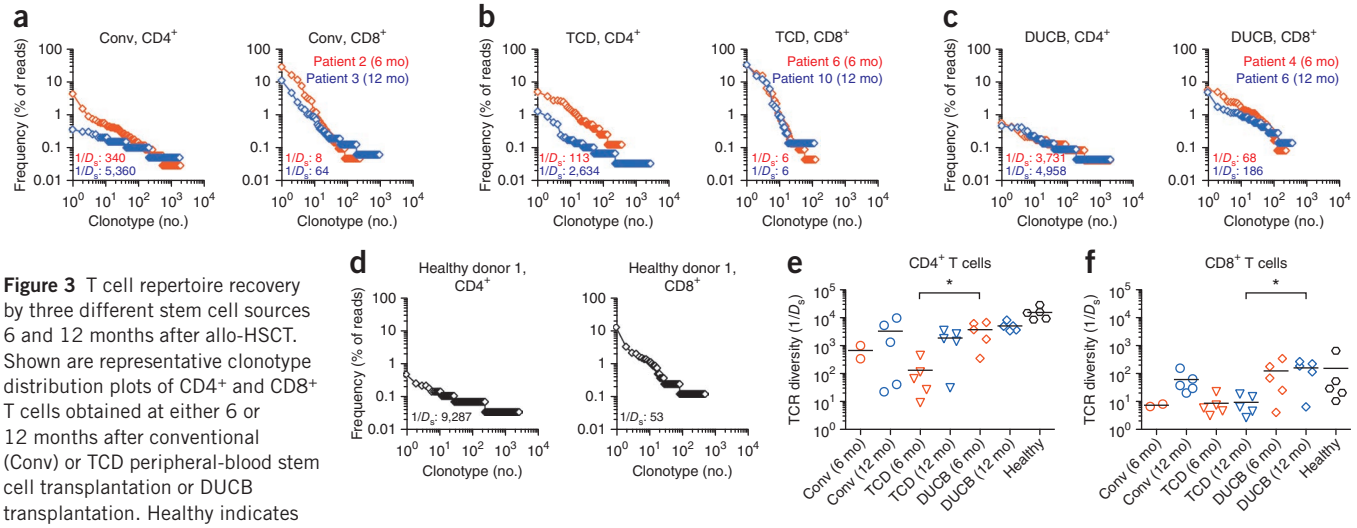

Figure 3 T cell repertoire recovery by three different stem cell sources 6 and 12 months after allo-HSCT. Shown are representative clonotype distribution plots of CD4+ and CD8+ T cells obtained at either 6 or 12 months after conventional (Conv) or TCD peripheral-blood stem cell transplantation or DUCB transplantation. Healthy indicates age-matched healthy donors. (**a**) Clonotype distribution plots of conventional transplant patients 2 (6 months; red) and 3 (12 months; blue). Values in the lower left corners are the TCR diversities. The numbers of reads were 3,379 for CD4+ and 1,985 for CD8+ for conventional transplant patient 2 and 1,954 for CD4+ and 1,515 for CD8+ for conventional transplant patient 3. (**b**) Clonotype distribution plots of TCD patients 6 (6 months; red) and 10 (12 months; blue). The numbers of reads were 793 for CD4+ and 2,141 for CD8+ for TCD patient 6 and 2,889 for CD4+ and 694 for CD8+ for TCD patient 10. (**c**) Clonotype distribution plots of DUCB patients 4 (6 months; red) and 6 (12 months; blue). The numbers of reads were 2,312 for CD4+ and 1,138 for CD8+ for DUCB patient 4 and 2,173 for CD4+ and 680 for CD8+ for DUCB patient 6. (**d**) Clonotype distribution plots of healthy donor 1. The numbers of reads were 2,856 for CD4+ and 800 for CD8+. (**e**) CD4+ T cell diversity in the indicated groups. Symbols represent individual subjects, and bars depict the group means. *$P = 0.033$ (one-way analysis of variance (ANOVA)). (**f**) CD8+ T cell diversity in the indicated groups. *$P = 0.012$ (one-way ANOVA).

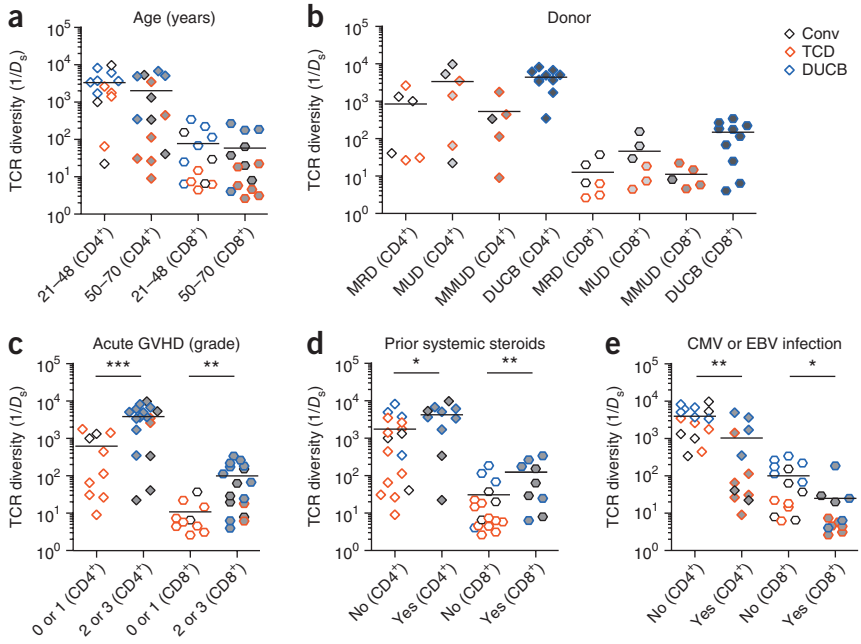

Figure 4 T cell repertoire recovery after allo-HSCT as a function of clinical variables. CD4+ and CD8+ T cell diversity of 27 allo-HSCT recipients was divided according to clinical parameters that could influence T cell repertoire recovery. (**a**) TCR diversity in patients either 21–48 years old ($n = 13$) or 50–70 years old ($n = 14$). (**b**) TCR diversity in patients that received transplantation from a matched related donor (MRD; $n = 6$), matched unrelated donor (MUD; $n = 6$), mismatched unrelated donor (MMUD; $n = 5$) or DUCB ($n = 10$). (**c**) TCR diversity in patients that had no (grade 0) GVHD or grade 1 acute GVHD ($n = 12$) or had grade 2 or 3 acute GVHD ($n = 15$). $^{**}P = 0.003$, $^{***}P < 0.001$ (Student's t test). (**d**) TCR diversity in patients that either received ($n = 10$) or did not receive ($n = 17$) prior systemic steroid treatment. $^{*}P = 0.023$, $^{**}P = 0.006$ (Student's t test). (**e**) TCR diversity in patients either infected ($n = 12$) or not infected ($n = 15$) with cytomegalovirus (CMV) or EBV. $^{*}P = 0.033$, $^{**}P = 0.004$ (Student's t test). Throughout the figure, each symbol represents an individual patient, and the bars depict the group means.

umbilical cord blood (DUCB) transplantation (**Fig. 3a–c**). As analysis of TCD patient 1 suggested substantially greater TCR diversity in CD4+ compared to CD8+ T cells (**Supplementary Fig. 4**), we separately analyzed both T cell compartments in all additional patients, as well as in five healthy donors (**Fig. 3d**). After determining the TCR diversity of each individual, we established the following: first, for all stem cell sources as well as in healthy donors, CD4+ T cell diversity was ~50 times higher than CD8+ T cell diversity ($1/D_s$ = 4,665 and

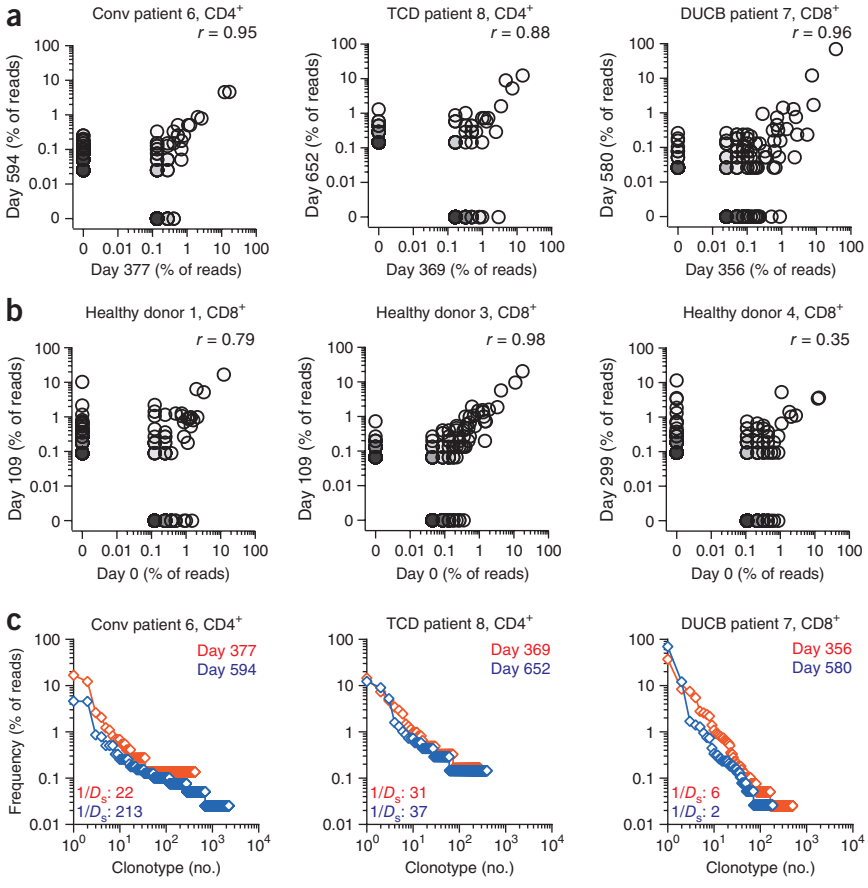

Figure 5 Monitoring patients with poor T cell repertoire recovery after allo-HSCT. Shown are the results from three patients with low TCR diversity after 12 months that were reanalyzed after 19–21 months. (**a**) Dot plots comparing the clonotype distribution of T cells isolated on different days from conventional (Conv) transplant patient 6 (CD4+ T cells, 218 d apart), TCD patient 8 (CD4+ T cells, 284 d apart) and DUCB patient 7 (CD8+ T cells, 225 d apart). The numbers of reads were 731 on day 377 and 3,940 on day 594 for conventional transplant patient 6,610 on day 369 and 688 on day 652 for TCD patient 8 and 3,932 on day 356 and 3,825 on day 580 for DUCB patient 7. (**b**) Dot plots comparing the clonotype distribution of CD8+ T cells isolated on different days from healthy donors 1 and 3 (both 109 d apart) and healthy donor 4 (299 d apart). The numbers of reads were 800 on day 0 and 1,120 on day 109 for healthy donor 1, 2,267 on day 0 and 1,508 on day 109 for healthy donor 3 and 917 on day 0 and 1,068 on day 299 for healthy donor 4. (**c**) Clonotype distribution plots of CD4+ T cells isolated from conventional transplant patient 6 and TCD patient 8, as well as CD8+ T cells isolated from DUCB patient 7. The values in the lower left corners are the TCR diversities. Over 218 d, conventional transplant patient 6 had a 1.6-fold increase in CD4+ T cell count (from 452 to 722 cells μl^{-1}) and a 9.7-fold increase in CD4+ T cell diversity. In contrast, over 284 d, TCD patient 8 had no increase in CD4+ T cell count (from 212 to 207 cells μl^{-1}) and no increase in CD4+ T cell diversity.

$1/D_s$ = 81, respectively; **Fig. 3e**,**f**). Second, healthy donors had the highest TCR diversity of CD4$^+$ T cells ($1/D_s$ = 15,470), followed by DUCB recipients after 12 months ($1/D_s$ = 5,069), DUCB recipients after 6 months ($1/D_s$ = 3,745), conventional transplant recipients after 12 months ($1/D_s$ = 3,298), TCD recipients after 12 months ($1/D_s$ = 1,871), conventional transplant recipients after 6 months ($1/D_s$ = 674) and TCD recipients after 6 months ($1/D_s$ = 132). Therefore, DUCB recipients had the highest TCR diversity of all patients and had a significantly (28-fold; $P = 0.033$) more diverse CD4$^+$ T cell repertoire compared to TCD recipients after 6 months. Notably, this increased TCR diversity also correlated with a substantially greater fraction of naive CD4$^+$ T cells in DUCB recipients compared to TCD recipients (**Supplementary Fig. 5**). Although TCD recipients had limited CD4$^+$ T cell diversity after 6 months, this diversity was 14-fold higher after 12 months, reducing the difference compared to DUCB recipients to threefold. Third, DUCB recipients also had the highest TCR diversity in CD8$^+$ T cells of all patients, which was 14-fold higher than TCD recipients after 6 months and 17-fold higher after 12 months, thereby reaching statistical significance ($P = 0.012$).

Using the above data, we investigated several clinical parameters that could influence T cell repertoire recovery. We found no significant impact of age or donor on TCR diversity (**Fig. 4a**,**b**). Notably, acute GVHD (grade 2 or 3) and prior systemic steroid treatment were associated with higher TCR diversity, suggesting that these variables do not restrict repertoire recovery (**Fig. 4c**,**d**). In contrast, infection with cytomegalovirus or EBV was associated with lower TCR diversity (**Fig. 4e**).

Within each stem cell group, we identified patients who had normal T cell counts but very low TCR diversity at 1 year after transplant (**Supplementary Fig. 6**). To investigate repertoire recovery during the second year after transplant, we reanalyzed TCR diversity in three of these patients at 19–21 months. For comparison, we also reanalyzed three healthy donors. We found stability in the CD4$^+$ T cell repertoires of conventional transplant patient 6 and TCD patient 8 and in the CD8$^+$ T cell repertoire of DUCB patient 7 (**Fig. 5a**). We found similar stability in the CD8$^+$ T cell repertoires of healthy donors when measured over 109 d, whereas over 299 d, there was somewhat greater divergence (**Fig. 5b**). Despite occasional changes in clonotype frequencies, TCR diversity in the healthy donors was stable (**Supplementary Fig. 7**). In conventional transplant patient 6 the frequency of abundant clonotypes substantially decreased over time, resulting in a tenfold higher CD4$^+$ T cell diversity (**Fig. 5c**). In contrast, there was no diversification in the other two patients. Together, these data illustrate the potential of our method to identify patients and transplant protocols that are associated with either more or less T cell repertoire recovery.

DISCUSSION

In this study we established a method to reproducibly and accurately measure human TCR diversity. By combining 5′ RACE PCR with deep sequencing, this method assesses the entire TCR-β repertoire using a single oligonucleotide pair, thereby eliminating amplification bias. This contrasts with TCR sequencing methods based on genomic DNA[20–23], which use many different oligonucleotides for amplification, making some degree of bias unavoidable. Although 5′ RACE PCR provides a clear advantage, a limitation is that it requires RNA, and thus changes in TCR transcription could skew the frequency of particular clonotypes. Although we used Roche/454 sequencing here, our approach should be readily adaptable to other platforms with greater sequence coverage, which could help identify infrequent TCRs.

We validated our method by measuring T cell repertoire recovery in allo-HSCT recipients, in whom limited TCR diversity is linked to susceptibility to infection and cancer relapse. Although we found significant improvement in TCR diversity over time, there was substantial variability in the rate of recovery between different stem cell sources. Most notably, DUCB recipients had a 28-fold higher CD4$^+$ T cell diversity compared to TCD recipients after 6 months. This is consistent with clinical findings, which have shown that after 6 months, DUCB recipients have a low incidence of infection[33] and have higher CD4$^+$ T cell numbers and a lower rate of leukemia relapse compared to TCD recipients[27,33,34]. Although many variables could contribute to this differential repertoire recovery, it can be explained at least partially by the fact that DUCB recipients receive ~7,000-fold more T cells in their graft, and DUCB transplantation is performed without T cell–depleting regimens[33].

Besides allo-HSCT, this method should be useful in characterizing T cell immunity in other clinical settings of immune deficiency, autoimmunity and tumor immunity. Ultimately, the ability to measure T cell repertoire complexity with great precision should guide the way for new therapeutic approaches aimed at immune regeneration.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank all patients and healthy donors who volunteered to participate in this study. We thank G. Redelman-Sidi for collection of blood samples, I. Leiner for processing of blood samples, C. Ubeda-Morant for assistance with data analysis, J. Li and A. Viale of the Memorial Sloan-Kettering Cancer Center Genomics Core Laboratory for performing Roche/454 sequencing and R. Jenq for critical reading of the manuscript. This work was supported by the Netherlands Organisation for Scientific Research (NWO) (Rubicon fellowship to J.W.J.v.H.), the Asociación Española Contra el Cáncer (AECC) (PAO grant to I.C.), the US National Institutes of Health (NIH) (grants HL069929, CA023766, AI080455, AI100288 and AI101406 to M.R.M.v.d.B. and AI042135, AI039031 and CA023766 to E.G.P.) and The Tow Foundation (to E.G.P.). Additional support was received from the US Department of Defense (USAMRAA Award W81XWH-09-1-0294), the Radiation Effects Research Foundation (RERF-NIAID), The Lymphoma Foundation, Alex's Lemonade Stand, The Geoffrey Beene Cancer Research Center at Memorial Sloan-Kettering Cancer Center and The Susan and Peter Solomon Divisional Genomics Program (to M.R.M.v.d.B.), The Translational and Integrative Medicine Research Fund of Memorial Sloan-Kettering Cancer Center, Cycle for Survival, The New York Community Trust and When Everyone Survives (to M.A.P.) and The Experimental Therapeutics Center of Memorial Sloan-Kettering Cancer Center funded by W.H. Goodwin and A. Goodwin (to M.R.M.v.d.B. and M.A.P.).

AUTHOR CONTRIBUTIONS

J.W.J.v.H. designed and performed the study, analyzed data and wrote the manuscript. I.C. collected clinical data. L.B.L. performed Roche/454 sequencing. D.W.S. compared naive and memory CD8$^+$ T cells. G.D.W., A.M.R.G. and J.L.N. provided patient samples. M.R.M.v.d.B. designed and supervised the study. M.A.P. designed and supervised the study and provided patient samples. E.G.P. designed and supervised the study, analyzed data and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/nm.3100.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Storek, J. *et al.* Reconstitution of the immune system after hematopoietic stem cell transplantation in humans. *Semin. Immunopathol.* **30**, 425–437 (2008).
2. Seggewiss, R. & Einsele, H. Immune reconstitution after allogeneic transplantation and expanding options for immunomodulation: an update. *Blood* **115**, 3861–3868 (2010).

3. Deeg, H.J. & Socie, G. Malignancies after hematopoietic stem cell transplantation: many questions, some answers. *Blood* **91**, 1833–1844 (1998).
4. Small, T.N. *et al.* Comparison of immune reconstitution after unrelated and related T-cell–depleted bone marrow transplantation: effect of patient age and donor leukocyte infusions. *Blood* **93**, 467–480 (1999).
5. Maury, S. *et al.* Prolonged immune deficiency following allogeneic stem cell transplantation: risk factors and complications in adult patients. *Br. J. Haematol.* **115**, 630–641 (2001).
6. Nikolich-Zugich, J., Slifka, M.K. & Messaoudi, I. The many important facets of T-cell repertoire diversity. *Nat. Rev. Immunol.* **4**, 123–132 (2004).
7. Goldberg, G.L., Zakrzewski, J.L., Perales, M.A. & van den Brink, M.R. Clinical strategies to enhance T cell reconstitution. *Semin. Immunol.* **19**, 289–296 (2007).
8. Dumont-Girard, F. *et al.* Reconstitution of the T-cell compartment after bone marrow transplantation: restoration of the repertoire by thymic emigrants. *Blood* **92**, 4464–4471 (1998).
9. Douek, D.C. *et al.* Assessment of thymic output in adults after haematopoietic stem-cell transplantation and prediction of T-cell reconstitution. *Lancet* **355**, 1875–1881 (2000).
10. Roux, E. *et al.* Recovery of immune reactivity after T-cell–depleted bone marrow transplantation depends on thymic activity. *Blood* **96**, 2299–2303 (2000).
11. Weinberg, K. *et al.* Factors affecting thymic function after allogeneic hematopoietic stem cell transplantation. *Blood* **97**, 1458–1466 (2001).
12. Lynch, H.E. *et al.* Thymic involution and immune reconstitution. *Trends Immunol.* **30**, 366–373 (2009).
13. Langerak, A.W. *et al.* Molecular and flow cytometric analysis of the Vβ repertoire for clonality assessment in mature TCRαβ T-cell proliferations. *Blood* **98**, 165–173 (2001).
14. Gaspar, H.B. *et al.* Long-term persistence of a polyclonal T cell repertoire after gene therapy for X-linked severe combined immunodeficiency. *Sci. Transl. Med.* **3**, 97ra79 (2011).
15. Gorski, J. *et al.* Circulating T cell repertoire complexity in normal individuals and bone marrow recipients analyzed by CDR3 size spectratyping. Correlation with immune status. *J. Immunol.* **152**, 5109–5119 (1994).
16. Memon, S.A., Sportes, C., Flomerfelt, F.A., Gress, R.E. & Hakim, F.T. Quantitative analysis of T cell receptor diversity in clinical samples of human peripheral blood. *J. Immunol. Methods* **375**, 84–92 (2012).
17. Verfuerth, S. *et al.* Longitudinal monitoring of immune reconstitution by CDR3 size spectratyping after T-cell–depleted allogeneic bone marrow transplant and the effect of donor lymphocyte infusions on T-cell repertoire. *Blood* **95**, 3990–3995 (2000).
18. Wu, C.J. *et al.* Reconstitution of T-cell receptor repertoire diversity following T-cell depleted allogeneic bone marrow transplantation is related to hematopoietic chimerism. *Blood* **95**, 352–359 (2000).
19. Talvensaari, K. *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* **99**, 1458–1464 (2002).
20. Wang, C. *et al.* High throughput sequencing reveals a complex pattern of dynamic interrelationships among human T cell subsets. *Proc. Natl. Acad. Sci. USA* **107**, 1518–1523 (2010).
21. Klarenbeek, P.L. *et al.* Human T-cell memory consists mainly of unexpanded clones. *Immunol. Lett.* **133**, 42–48 (2010).
22. Robins, H.S. *et al.* Overlap and effective size of the human CD8+ T cell receptor repertoire. *Sci. Transl. Med.* **2**, 47ra64 (2010).
23. Sherwood, A.M. *et al.* Deep sequencing of the human TCRγ and TCRβ repertoires suggests that TCRβ rearranges after αβ and γδ T cell commitment. *Sci. Transl. Med.* **3**, 90ra61 (2011).
24. Venturi, V. *et al.* A mechanism for TCR sharing between T cell subsets and individuals revealed by pyrosequencing. *J. Immunol.* **186**, 4285–4294 (2011).
25. Warren, R.L. *et al.* Exhaustive T-cell repertoire sequencing of human peripheral blood samples reveals signatures of antigen selection and a directly measured repertoire size of at least 1 million clonotypes. *Genome Res.* **21**, 790–797 (2011).
26. Benichou, J., Ben-Hamo, R., Louzoun, Y. & Efroni, S. Rep-Seq: uncovering the immunological repertoire through next-generation sequencing. *Immunology* **135**, 183–191 (2012).
27. Ponce, D.M. *et al.* Reduced late mortality risk contributes to similar survival after double-unit cord blood transplantation compared with related and unrelated donor hematopoietic stem cell transplantation. *Biol. Blood Marrow Transplant.* **17**, 1316–1326 (2011).
28. Jacobson, C.A. *et al.* Immune reconstitution after double umbilical cord blood stem cell transplantation: comparison with unrelated peripheral blood stem cell transplantation. *Biol. Blood Marrow Transplant.* **18**, 565–574 (2012).
29. Schumacher, T.N., Gerlach, C. & van Heijst, J.W. Mapping the life histories of T cells. *Nat. Rev. Immunol.* **10**, 621–631 (2010).
30. Nguyen, P. *et al.* Identification of errors introduced during high throughput sequencing of the T cell receptor repertoire. *BMC Genomics* **12**, 106 (2011).
31. Venturi, V., Kedzierska, K., Turner, S.J., Doherty, P.C. & Davenport, M.P. Methods for comparing the diversity of samples of the T cell receptor repertoire. *J. Immunol. Methods* **321**, 182–195 (2007).
32. Lim, A. *et al.* Frequent contribution of T cell clonotypes with public TCR features to the chronic response against a dominant EBV-derived epitope: application to direct detection of their molecular imprint on the human peripheral T cell repertoire. *J. Immunol.* **165**, 2001–2011 (2000).
33. Sauter, C. *et al.* Serious infection risk and immune recovery after double-unit cord blood transplantation without antithymocyte globulin. *Biol. Blood Marrow Transplant.* **17**, 1460–1471 (2011).
34. Jakubowski, A.A. *et al.* T cell depleted stem-cell transplantation for adults with hematologic malignancies: sustained engraftment of HLA-matched related donor grafts without the use of antithymocyte globulin. *Blood* **110**, 4552–4559 (2007).

ONLINE METHODS

Patients. Twenty-eight patients underwent allo-HSCT at Memorial Sloan-Kettering Cancer Center between April 2010 and September 2011. Patient and treatment characteristics are summarized in **Supplementary Table 1**. Pre-transplant conditioning varied according to patient age, diagnosis, remission status, extent of previous therapies and comorbidities and consisted of high-dose, reduced-intensity myeloablative and nonmyeloablative regimens[35]. GVHD prophylaxis for peripheral-blood stem cell transplantation was either with T cell depletion[34] or was calcineurin-inhibitor based, and antithymocyte globulin was used according to protocol or physician preference. Cord-blood recipients received mycophenolate mofetil and calcineurin inhibitors[36]; however, none of these patients received antithymocyte globulin[33]. Granulocyte colony-stimulating factor was used in all patients after transplant. Acute and late-acute or chronic GVHD were diagnosed clinically with histological confirmation when possible. Staging of GVHD was graded according to standard criteria[37,38]. All subjects provided informed consent for the collection of blood samples, which was approved by the Institutional Review Board of Memorial Sloan-Kettering Cancer Center. Graft samples were not available for analysis.

T cell isolation and flow cytometry. From each ~8-ml heparinized blood sample, mononuclear cells were isolated by density centrifugation (Lymphocyte Separation Medium, MP Biomedicals). Recovered cells were lysed in RLT buffer (QIAGEN), homogenized using QIAshredder columns (QIAGEN) and stored at −80 °C. For $CD4^+$ and $CD8^+$ T cell separation, two blood samples were pooled, and then the mononuclear cell fraction was isolated. Recovered cells were split into two fractions and incubated with either CD4 or CD8 MicroBeads (Miltenyi Biotec). $CD4^+$ and $CD8^+$ T cells were separated using MS columns (Miltenyi Biotec). Eluted cells were lysed, homogenized and stored as described above. To determine the efficiency of T cell separation, eluted cells were stained with antibody to CD14 (clone M5E2, 1:5), antibody to CD4 (clone SK3, 1:20) and antibody to CD8 (clone RPA-T8, 1:5) (all BD Pharmingen) and measured on an LSRII flow cytometer (BD Biosciences). Data were analyzed using FlowJo software (TreeStar). For separation of naive and memory $CD8^+$ T cells, isolated mononuclear cells were stained with antibody to CD45RA (clone HI100, 1:5), antibody to CD45RO (clone UCHL1, 1:5) (both BD Pharmingen) and antibody to CD8. Cells were sorted using a FACSAria cell sorter (BD Biosciences) into $CD8^+CD45RA^+CD45RO^-$ (naive) and $CD8^+CD45RA^-CD45RO^+$ (memory) fractions.

5′ RACE PCR and Roche/454 sequencing. Total RNA from frozen homogenates was extracted using an RNeasy mini kit (QIAGEN). RACE-Ready cDNA was generated using a SMARTer RACE cDNA Amplification kit (Clontech) and oligo(dT) or random (N-15) primers. 5′ RACE PCR was performed using Advantage 2 Polymerase (Clontech) with Clontech's universal forward primer and a self-designed universal TCR-β–constant reverse primer compatible with both *TRBC* genes (5′-GCACACCAGTGTGGCCTTTTGGG-3′). Amplification was performed on a Mastercycler pro (Eppendorf) and was 1 min at 95 °C; 5 cycles of 20 s at 95 °C and 30 s at 72 °C; 5 cycles of 20 s at 95 °C, 30 s at 70 °C and 30 s at 72 °C; 25 cycles of 20 s at 95 °C, 30 s at 60 °C and 30 s at 72 °C; and 7 min at 72 °C. PCR products were loaded on 1.2% agarose gels (Bio-Rad), and bands centered at ~600 bp were excised and purified using a MinElute Gel Extraction kit (QIAGEN). Purified products were subjected to a second round of amplification to introduce adaptor sequences compatible with unidirectional Roche/454 sequencing. One-fiftieth of the first-round PCR product was amplified using Advantage 2 Polymerase with a hybrid forward primer consisting of Roche's Lib-L primer B and Clontech's nested universal primer (5′-CCTATCCCCTGTGTGCCTTGGCAGTCTC AGAAGCAGTGGTATCAACGCAGAGT-3′) and a hybrid reverse primer consisting of Roche's Lib-L primer A and a self-designed nested universal TCR-β–constant primer (5′-CCATCTCATCCCTGCGTGTCTCCGACTCA G-MID-AACACAGCGACCTCGGGTGGGAA-3′, where MID represents the multiplex identifier used to separate pooled samples during sequence analysis). Multiplex identifiers were 6–7 bp long. Amplification was 1 min at 95 °C; 25 cycles of 20 s at 95 °C, 30 s at 68 °C and 30 s at 72 °C; and 7 min at 72 °C. PCR products were purified from agarose gels as described above. Purified products were sequenced using the GS Junior 454 platform (Roche).

Sequence analysis. Raw sequence data were converted to FASTA format using MOTHUR software[39]. Sequences shorter than 125 bp, those with uncalled bases, those with a Phred quality score average <30 (base call accuracy <99.9%)[30] or those with no exact match to the TCR-β–constant primer or a multiplex identifier were discarded. The resulting FASTA files were uploaded to the IMGT/HighV-QUEST database (http://www.imgt.org/HighV-QUEST/index.action)[40]. Using IMGT summary files, sequences with out-of-frame rearrangements, a V- and J-region identity <80%, V-region pseudogenes or a CDR3-β amino acid junction lacking a 5′ cysteine and 3′ phenylalanine were discarded. Resulting sequences were sorted using Excel (Microsoft) and graphed using Prism 5 software (GraphPad). The inverse Simpson's diversity index ($1/D_s$) was calculated using MOTHUR.

Statistical analyses. TCR diversity was compared using an unpaired Student's *t* test (two groups) or one-way ANOVA with Bonferroni's multiple comparison test (three or more groups). $P < 0.05$ was considered statistically significant.

35. Bacigalupo, A. *et al.* Defining the intensity of conditioning regimens: working definitions. *Biol. Blood Marrow Transplant.* **15**, 1628–1633 (2009).
36. Barker, J.N. *et al.* A "no-wash" albumin-dextran dilution strategy for cord blood unit thaw: high rate of engraftment and a low incidence of serious infusion reactions. *Biol. Blood Marrow Transplant.* **15**, 1596–1602 (2009).
37. Rowlings, P.A. *et al.* IBMTR Severity Index for grading acute graft-versus-host disease: retrospective comparison with Glucksberg grade. *Br. J. Haematol.* **97**, 855–864 (1997).
38. Filipovich, A.H. *et al.* National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host disease: I. Diagnosis and staging working group report. *Biol. Blood Marrow Transplant.* **11**, 945–956 (2005).
39. Schloss, P.D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
40. Alamyar, E., Giudicelli, V., Li, S., Duroux, P. & Lefranc, M.P. IMGT/HighV-QUEST: the IMGT(R) web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. *Immunome Res.* **8**, 26 (2012).

doi:10.1038/nm.3100