

# Origin and evolution of the T cell repertoire after posttransplantation cyclophosphamide

Christopher G. Kanakry,<sup>1</sup> David G. Coffey,<sup>2,3</sup> Andrea M.H. Towlerton,<sup>2</sup> Ante Vulic,<sup>1</sup> Barry E. Storer,<sup>2</sup> Jeffrey Chou,<sup>2,3</sup> Cecilia C.S. Yeung,<sup>2,4</sup> Christopher D. Gocke,<sup>1</sup> Harlan S. Robins,<sup>5</sup> Paul V. O'Donnell,<sup>2,3</sup> Leo Luznik,<sup>1</sup> and Edus H. Warren<sup>2,3</sup>

<sup>1</sup>Sidney Kimmel Comprehensive Cancer Center, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. <sup>2</sup>Clinical Research Division, Fred Hutchinson Cancer Research Center (FHCRC), Seattle, Washington, USA.

<sup>3</sup>Department of Medicine and <sup>4</sup>Department of Pathology, University of Washington, Seattle, Washington, USA.

<sup>5</sup>Public Health Sciences Division, FHCRC, Seattle, Washington, USA.

Posttransplantation cyclophosphamide (PTCy) effectively prevents graft-versus-host disease (GVHD), but its immunologic impact is poorly understood. We assessed lymphocyte reconstitution via flow cytometry ( $n = 74$ ) and antigen receptor sequencing ( $n = 35$ ) in recipients of myeloablative, HLA-matched allogeneic BM transplantation using PTCy. Recovering T cells were primarily phenotypically effector memory with lower T cell receptor  $\beta$  (*TRB*) repertoire diversity than input donor repertoires. Recovering B cells were predominantly naive with immunoglobulin heavy chain locus (*IGH*) repertoire diversity similar to donors. Numerical T cell reconstitution and *TRB* diversity were strongly associated with recipient cytomegalovirus seropositivity. Global similarity between input donor and recipient posttransplant repertoires was uniformly low at 1–2 months after transplant but increased over the balance of the first posttransplant year. Blood *TRB* repertoires at  $\geq 3$  months after transplant were often dominated by clones present in the donor blood/marrow memory CD8<sup>+</sup> compartment. Limited overlap was observed between the *TRB* repertoires of T cells infiltrating the skin or gastrointestinal tract versus the blood. Although public *TRB* sequences associated with herpesvirus- or alloantigen-specific CD8<sup>+</sup> T cells were detected in some patients, posttransplant *TRB* and *IGH* repertoires were unique to each individual. These data define the immune dynamics occurring after PTCy and establish a benchmark against which immune recovery after other transplantation approaches can be compared.

## Introduction

Posttransplantation cyclophosphamide (PTCy), when given on posttransplant days +3 and +4, is an effective prophylaxis for both severe acute and chronic graft-versus-host disease (GVHD) after allogeneic blood or marrow transplantation (alloBMT) using a variety of transplantation platforms (1–8). Furthermore, PTCy has been associated with favorable immune reconstitution marked by low incidence of invasive viral infections (2–4, 6, 8) or Epstein-Barr virus–related (EBV-related) posttransplantation lymphoproliferative disease (9, 10). Recent studies have shown that immune reconstitution in the first 1–2 months after PTCy is characterized by persistence of activated Tregs and phenotypically stem cell–like memory T cells that survive PTCy (11–14). However, the immune dynamics occurring after alloBMT using PTCy remain poorly understood.

Recently, next-generation sequencing of the T cell receptor  $\beta$  (*TRB*) chain locus (15, 16) has been applied to tracking T cell reconstitution after alloBMT (13, 14, 17–20). *TRB* repertoires were found to be more diverse after double umbilical cord blood transplantation (UCBT) than after T cell depletion (TCD) or calcineurin inhibitor–based (CNI-based) GVHD prophylaxis for alloBMT (18). Surprisingly, the *TRB* repertoires within both the CD4<sup>+</sup> and CD8<sup>+</sup> T cell compartments were more diverse in patients with acute GVHD (aGVHD) or patients who had received prior systemic steroids (18). Regardless of the donor type, the repertoires of CD4<sup>+</sup> T cells were much more diverse than the repertoires of CD8<sup>+</sup> T cells (18), similar

**Authorship note:** C.G. Kanakry, D.G. Coffey, A.M.H. Towlerton, and A. Vulic, as well as L. Luznik and E.H. Warren, contributed equally to this work.

**Conflict of interest:** H.S. Robins has consultancy, equity ownership, patents, and royalties with Adaptive Biotechnologies Inc., which performed much of the antigen receptor sequencing described in this work.

**Submitted:** December 28, 2015

**Accepted:** March 18, 2016

**Published:** April 21, 2016

**Reference information:**

*JCI Insight.* 2016;1(5):e86252.

doi:10.1172/jci.insight.86252.

**Table 1. Multivariable analyses of lymphocyte subset counts at 2–3 months after transplant**

Total Numbers	Recipient CMV Seropositivity			Donor CMV Seropositivity			CMV Peripheral Blood Reactivation			Acute GVHD		
	Effect Size	95% CI	<i>P</i>	Effect Size	95% CI	<i>P</i>	Effect Size	95% CI	<i>P</i>	Effect Size	95% CI	<i>P</i>
Lymphocytes	1.52	0.92–2.52	0.101	1.03	0.66–1.6	0.905	1.33	0.76–2.36	0.314	<b>0.48</b>	<b>0.3–0.77</b>	<b>0.003</b>
CD3 <sup>+</sup> cells	<b>2.39</b>	<b>1.16–4.90</b>	<b>0.019</b>	1.24	0.66–2.34	0.495	1.49	0.66–3.36	0.33	<b>0.38</b>	<b>0.19–0.74</b>	<b>0.006</b>
CD3 <sup>+</sup> CD4 <sup>+</sup> cells	<b>2.2</b>	<b>1.12–4.3</b>	<b>0.022</b>	1.09	0.6–1.96	0.777	1.18	0.55–2.52	0.661	<b>0.47</b>	<b>0.25–0.87</b>	<b>0.017</b>
Naive	1.01	0.37–2.76	0.976	<b>0.47</b>	<b>0.22–0.99</b>	<b>0.047</b>	0.72	0.24–2.2	0.557	1.09	0.48–2.45	0.836
CM	1.82	0.63–5.23	0.26	0.64	0.29–1.41	0.262	0.47	0.14–1.53	0.202	0.75	0.32–1.76	0.498
EM	<b>3.71</b>	<b>1.34–10.28</b>	<b>0.013</b>	1.16	0.54–2.49	0.693	0.82	0.26–2.55	0.721	0.53	0.23–1.22	0.130
TEMRA	1.99	0.75–5.28	0.16	0.69	0.33–1.44	0.316	0.85	0.29–2.52	0.762	0.79	0.36–1.74	0.544
CD3 <sup>+</sup> CD8 <sup>+</sup> cells	<b>2.65</b>	<b>1.11–6.33</b>	<b>0.029</b>	1.35	0.64–2.91	0.431	1.81	0.68–4.83	0.232	<b>0.33</b>	<b>0.15–0.74</b>	<b>0.008</b>
Naive	1.89	0.53–6.76	0.317	0.75	0.29–1.95	0.546	0.42	0.1–1.73	0.22	0.97	0.35–2.73	0.955
CM	4.0	0.88–18.16	0.071	1.02	0.33–3.17	0.965	0.56	0.1–3.01	0.488	0.52	0.15–1.78	0.291
EM	<b>5.22</b>	<b>1.37–19.86</b>	<b>0.017</b>	1.83	0.67–4.97	0.228	1.39	0.31–6.17	0.657	<b>0.28</b>	<b>0.09–0.82</b>	<b>0.021</b>
TEMRA	2.8	0.89–8.85	0.077	1.22	0.52–2.88	0.645	0.86	0.24–3.11	0.818	<b>0.36</b>	<b>0.14–0.91</b>	<b>0.032</b>
CD3 <sup>+</sup> CD56 <sup>+</sup> NK cells	1.08	0.59–1.97	0.803	0.67	0.41–1.09	0.104	1.02	0.53–1.95	0.962	0.95	0.57–1.59	0.84
CD19 <sup>+</sup> B cells	4.26	0.72–25.1	0.106	1.09	0.26–4.52	0.906	0.37	0.05–2.84	0.329	0.44	0.09–2.1	0.295
Naive mature	5.07	0.36–71.15	0.22	0.77	0.11–5.52	0.788	0.32	0.02–6.01	0.433	0.47	0.06–3.99	0.479
Transitional	16.33	0.72–370.5	0.078	0.95	0.09–9.78	0.964	0.44	0.01–14.27	0.635	<b>0.07</b>	<b>0.01–0.83</b>	<b>0.036</b>
Memory	0.63	0.08–5.14	0.659	1.41	0.29–6.76	0.659	1.24	0.12–12.87	0.852	3.35	0.61–18.36	0.157

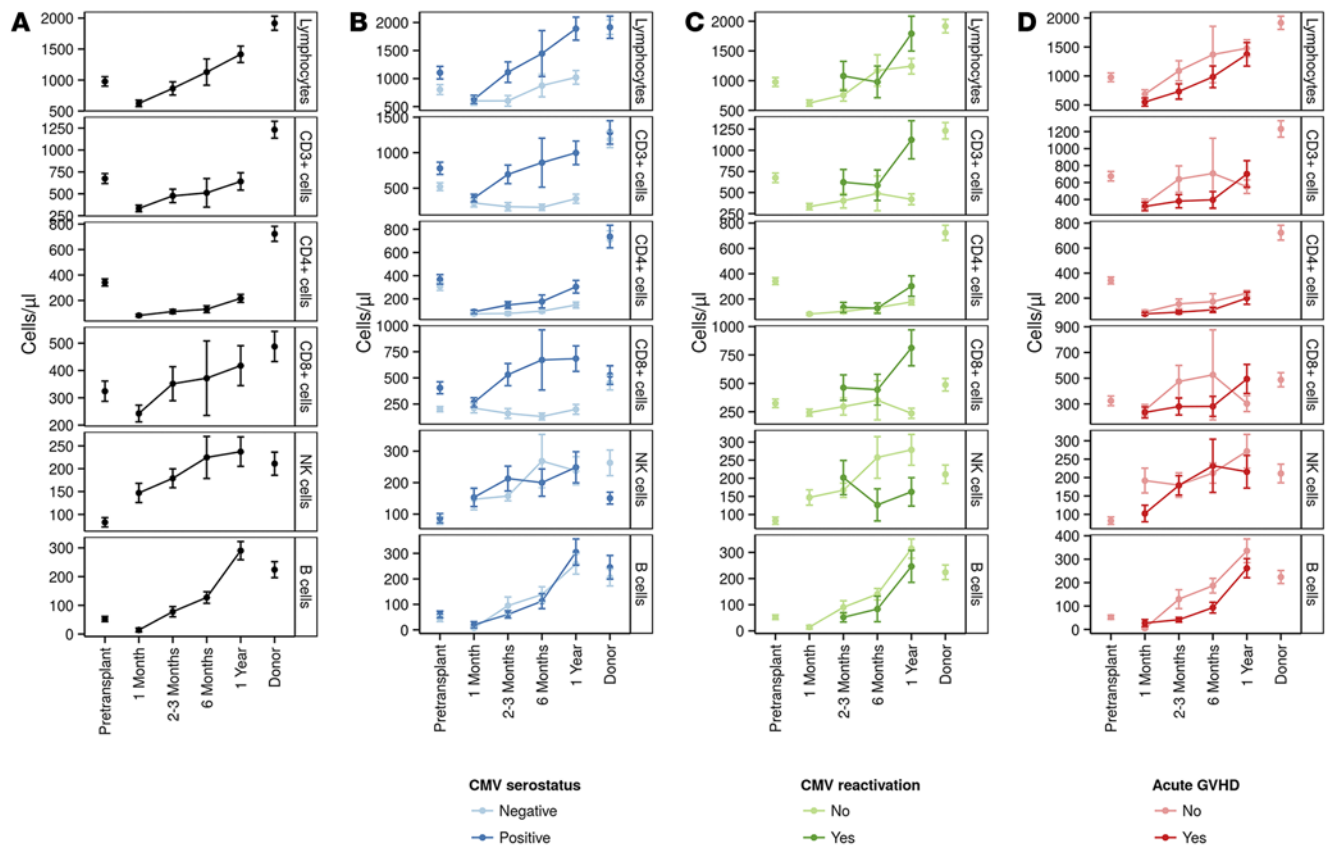
Bold numbers indicate statistically significant effect sizes with values of *P* < 0.05.

to differences in CD4<sup>+</sup> and CD8<sup>+</sup> T cell repertoire diversity observed after autologous transplantation (21). Greater diversity in CD4<sup>+</sup> T cells than CD8<sup>+</sup> T cells also has been observed in donors for alloBMT (22). Another study monitored *TRB* repertoires in patients with gastrointestinal (GI) aGVHD and found that the *TRB* repertoires in GI tissues were distinct among individuals, that these *TRB* repertoires were more similar across different GI sites within patients with steroid-refractory aGVHD, and that higher levels of prominent GI tissue-infiltrating clones were seen in the blood of steroid-refractory cases (19). Lastly, *TRB* sequencing recently has been applied to alloBMT employing PTCy in several patients where it was used to show the persistence both of naive T cells and antigen-specific memory T cells surviving PTCy (13, 14). These naive-derived T cells manifested a stem cell-like memory phenotype and seemed to preferentially reconstitute the circulating repertoire in the first month after transplant (13, 14).

We performed flow cytometric characterization of lymphocyte reconstitution in 74 patients treated on a multiinstitutional study of PTCy as single-agent GVHD prophylaxis after myeloablative, HLA-matched related or HLA-matched unrelated alloBMT (8). Using peripheral blood (PB), BM, and/or GVHD target organ biopsy samples from 35 of these patients and 16 of their related donors, we employed antigen receptor sequencing to assess the diversity of the *TRB* and immunoglobulin heavy chain locus (*IGH*) repertoires and track T cell clones in both blood and GVHD target tissues for up to 5 years after transplant.

## Results

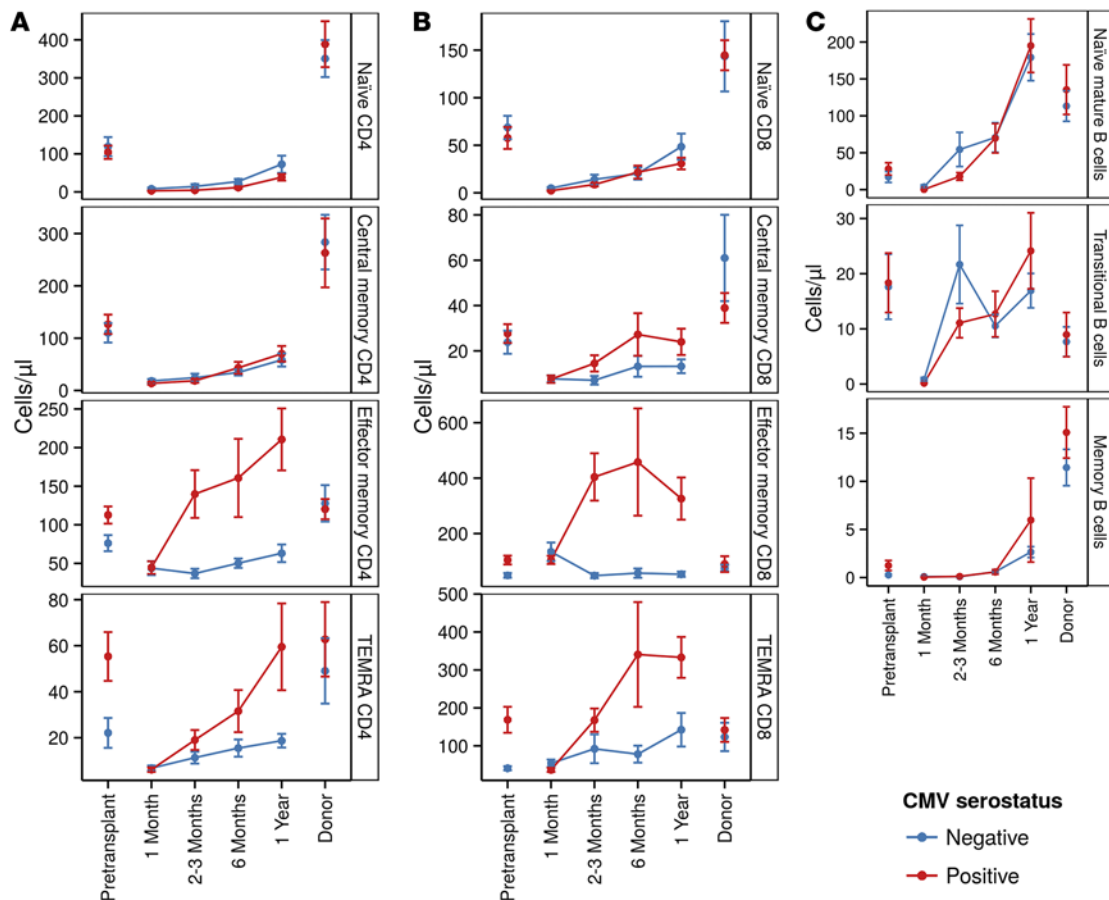
*After PTCy, the phenotype of recovering T cells is predominantly effector memory, while B cells are predominantly phenotypically naive.* Absolute lymphocyte counts (ALCs) increased progressively after transplant and recovered to the clinical laboratory-defined normal range (1,100–4,800 cells/μl) by 6–12 months after transplant (Figure 1A). However, CD3<sup>+</sup> T cell recovery lagged behind total lymphocyte recovery, largely as a result of slow CD3<sup>+</sup>CD4<sup>+</sup> T cell recovery (Figure 1A). Numerical reconstitution of CD3<sup>+</sup>CD8<sup>+</sup> T cells and CD3<sup>+</sup>CD56<sup>+</sup> NK cells was rapid (Figure 1A). CD19<sup>+</sup> B cells returned to median donor levels by 12 months after transplant (Figure 1A). Reconstitution of both CD4<sup>+</sup> and CD8<sup>+</sup> T cell subsets was driven largely by rapid recovery of phenotypically effector memory (EM; CD45RA<sup>+</sup>CCR7<sup>-</sup>) and terminally differentiated CD45RA<sup>+</sup> EM (TEMRA; CD45RA<sup>+</sup>CCR7<sup>-</sup>) cells (Figure 2, A and B). Conversely, naive T cells and recent thymic emigrants (RTEs), defined by CD31 expression on CD4<sup>+</sup>CD45RA<sup>+</sup> cells, were slow to recover (Figure 2, A and B, and Supplemental Figure 1; supplemental material available online with this article; doi:10.1172/jci.insight.86252DS1). The low levels of RTEs suggested that the vast majority of T cell reconstitution in



**Figure 1. The association of clinical factors with lymphocyte recovery after myeloablative, HLA-matched alloBMT using PTCy as single-agent GVHD prophylaxis.** Numerical reconstitution of lymphocytes, CD3<sup>+</sup> T cells, CD3<sup>+</sup>CD4<sup>+</sup> T cells, CD3<sup>+</sup>CD8<sup>+</sup> T cells, CD3<sup>+</sup>CD56<sup>+</sup> NK cells, and CD19<sup>+</sup> B cells is shown for (A) all patients and donors and stratified by (B) recipient cytomegalovirus (CMV) serostatus, (C) the detection of peripheral blood CMV reactivation after transplant occurring by the specified time point, or (D) the diagnosis of acute graft-versus-host disease (GVHD) by the specified time point. Error bars represent the standard error. Donor data in B are stratified by donor CMV serostatus. Donor and pretransplant recipient data are shown in D for comparison purposes only.

the first posttransplant year was attributable to proliferation of cells present in the donor graft that survived PTCy, rather than from T cells newly generated and selected in the thymus of these adult patients after myeloablative alloBMT. By contrast, recovering B cells were predominantly naive in phenotype throughout the first posttransplant year (Figure 2C).

*Numerical T cell reconstitution is associated with history of aGVHD and recipient CMV seropositivity.* Occurrence of GVHD (and its accompanying immunosuppression) and CMV are both known to influence immune recovery after allogeneic transplantation with CNI-based GVHD prophylaxis (23, 24); therefore, we explored whether immune reconstitution after PTCy was associated with the occurrence of aGVHD, recipient CMV serostatus, donor CMV serostatus, and/or the occurrence of detectable posttransplant CMV reactivation (Figures 1 and 2, and Tables 1 and 2). In multivariate analysis, recipient CMV seropositivity maintained a consistent positive association with both CD4<sup>+</sup> and CD8<sup>+</sup> T cell numerical reconstitution, while occurrence of aGVHD was associated with lower T cell numbers (Figure 1, B and D, and Tables 1 and 2). Recipients without detectable CMV reactivation had more rapid CD4<sup>+</sup> and CD8<sup>+</sup> T cell reconstitution if they were CMV seropositive before transplant than if they were CMV seronegative (Supplemental Figure 2), consistent with the hypothesis that CMV exposure (irrespective of detectable posttransplant CMV reactivation) was driving this expansion (Figure 1, B and C, and Figure 2, A and B). Even so, at 1 year — but not 2–3 months — after transplant, detectable CMV reactivation also was associated in multivariate analysis with higher numerical CD8<sup>+</sup> T cell reconstitution (Tables 1 and 2). The higher absolute T cell numbers in CMV-seropositive recipients was driven by rapid CD4<sup>+</sup> and CD8<sup>+</sup> EM recovery (Figure 1B; Figure 2, A and B; and Tables 1 and 2). Donor CMV serostatus did not seem to influence T cell reconstitution, except for lower naive CD4<sup>+</sup> T cell recovery at both 2–3 months and 1 year after transplant for



**Figure 2. Lymphocyte subset numerical reconstitution is significantly associated with recipient CMV seropositivity.** (A and B) Numerical reconstitution of naive (CCR7<sup>+</sup>CD45RA<sup>+</sup>), central memory (CCR7<sup>+</sup>CD45RA<sup>-</sup>), effector memory (CCR7<sup>-</sup>CD45RA<sup>+</sup>), and terminally differentiated CD45RA<sup>+</sup> effector memory (TEMRA, CCR7<sup>-</sup>CD45RA<sup>+</sup>) subsets of (A) CD3<sup>+</sup>CD4<sup>+</sup> and (B) CD3<sup>+</sup>CD8<sup>+</sup> T cell subsets is shown stratified by recipient pretransplant cytomegalovirus (CMV) serostatus. (C) Numerical reconstitution of CD19<sup>+</sup> B cell subsets (naive mature, CD27<sup>+</sup>CD24<sup>int</sup>CD38<sup>int</sup>; transitional, CD27<sup>-</sup>CD24<sup>hi</sup>CD38<sup>hi</sup>; and memory, CD27<sup>+</sup>IgD<sup>-</sup>CD24<sup>+</sup>CD38<sup>-/int</sup>) is shown stratified by recipient CMV serostatus. Donor data in all parts are stratified by donor CMV serostatus.

CMV-seropositive donors (Tables 1 and 2). History of aGVHD also was associated with lower naive CD4<sup>+</sup> and CD8<sup>+</sup> T cell recovery at 1 year after transplant (Table 2). There were no statistically significant associations between any of these 4 variables and NK cell or B cell recovery at 2–3 months or 1 year after transplant, except for lower transitional B cell numbers in recipients with a history of aGVHD (Tables 1 and 2).

*CMV has a more profound impact on the posttransplant T cell repertoire than does GVHD.* To complement the assessment of numerical and phenotypic T- and B-lymphocyte reconstitution by flow cytometry, we performed high-throughput sequencing of T cell and B cell receptor genes in both unfractionated PB mononuclear cells (PBMCs) and flow cytometrically sorted lymphocyte subsets in our cohort. T cell receptor  $\alpha$  chain locus (*TRA*) and *TRB* sequencing in a subset of 9 samples comprising both unfractionated and sorted PBMCs revealed that *TRA* and *TRB* repertoire clonality in individual samples was tightly correlated ( $r = 0.97$ ,  $P = 0.000015$ ) (Supplemental Figure 3); all subsequent T cell receptor (TCR) analyses consequently focused on assessment of *TRB* sequences. *TRB* sequencing was performed on PB specimens of all 34 patients for whom 1 year posttransplant specimens were available, as well as all 15 of their related BM donors for whom PB specimens obtained on the day of allograft donation were available (Table 3 and Supplemental Tables 1 and 2). Two distinct measures, clonality and Gini coefficient, were used to assess the degree of clonal skewing in the *TRB* repertoires (Figure 3, A–C, and Supplemental Figure 4). Clonality is a metric derived from the Shannon entropy of the frequency distribution of *TRB* sequences normalized by the  $\log_2$ (number of unique *TRB* sequences); it captures the clonal composition of a T cell population and has an adjustment for sequencing depth, thus making it less sensitive to this critical variable (25). The Gini coefficient is a graphically defined metric of repertoire diversity in the sequence frequency distribu-

**Table 2. Multivariable analyses of lymphocyte subset counts at 1 year after transplant**

Total Numbers	Recipient CMV Seropositivity			Donor CMV Seropositivity			CMV Peripheral Blood Reactivation			Acute GVHD		
	Effect Size	95% CI	P	Effect Size	95% CI	P	Effect Size	95% CI	P	Effect Size	95% CI	P
Lymphocytes	<b>1.95</b>	<b>1.25–3.07</b>	<b>0.005</b>	0.92	0.64–1.32	0.625	1.21	0.71–2.07	0.467	0.68	0.45–1.02	0.062
CD3 <sup>+</sup> cells	<b>2.25</b>	<b>1.06–4.81</b>	<b>0.037</b>	1.07	0.58–1.97	0.834	2.37	0.96–5.83	0.059	<b>0.5</b>	<b>0.25–1.0</b>	<b>0.049</b>
CD3 <sup>+</sup> CD4 <sup>+</sup> cells	<b>2.21</b>	<b>1.09–4.46</b>	<b>0.028</b>	0.75	0.43–1.33	0.313	1.59	0.69–3.65	0.268	<b>0.4</b>	<b>0.21–0.74</b>	<b>0.006</b>
Naive	1.08	0.45–2.6	0.851	<b>0.36</b>	<b>0.17–0.73</b>	<b>0.007</b>	1.13	0.4–3.14	0.813	<b>0.21</b>	<b>0.09–0.48</b>	<b>0.001</b>
CM	2.11	0.68–6.56	0.189	0.66	0.26–1.68	0.366	0.87	0.23–3.3	0.834	<b>0.31</b>	<b>0.1–0.9</b>	<b>0.033</b>
EM	<b>3.01</b>	<b>1.01–8.99</b>	<b>0.048</b>	1.06	0.43–2.6	0.902	1.79	0.5–6.43	0.359	0.42	0.15–1.2	0.103
TEMRA	2.08	0.78–5.51	0.135	0.56	0.25–1.26	0.153	1.64	0.52–5.12	0.383	0.7	0.27–1.76	0.427
CD3 <sup>+</sup> CD8 <sup>+</sup> cells	2.36	0.99–5.65	0.053	1.39	0.69–2.82	0.344	<b>2.82</b>	<b>1.0–7.94</b>	<b>0.049</b>	0.61	0.28–1.34	0.207
Naive	1.04	0.41–2.61	0.931	0.49	0.23–1.05	0.064	0.99	0.34–2.92	0.991	<b>0.35</b>	<b>0.15–0.84</b>	<b>0.02</b>
CM	2.44	0.57–10.4	0.217	1.04	0.31–3.44	0.946	0.84	0.15–4.59	0.836	0.49	0.12–1.95	0.297
EM	3.85	0.97–15.25	0.055	2.14	0.69–6.67	0.178	2.44	0.49–12.23	0.264	0.53	0.14–1.97	0.329
TEMRA	1.66	0.57–4.86	0.337	1.18	0.49–2.87	0.698	2.78	0.79–9.75	0.106	0.92	0.33–2.54	0.862
CD3 <sup>+</sup> CD56 <sup>+</sup> NK cells	1.46	0.66–3.24	0.335	0.57	0.29–1.09	0.086	0.47	0.19–1.2	0.11	0.84	0.39–1.78	0.63
CD19 <sup>+</sup> B cells	1.65	0.7–3.89	0.24	0.84	0.41–1.7	0.612	0.66	0.24–1.81	0.409	0.72	0.32–1.63	0.418
Naive Mature	1.75	0.61–5.01	0.286	1.06	0.44–2.54	0.896	0.52	0.15–1.79	0.286	0.6	0.22–1.67	0.313
Transitional	1.74	0.51–5.88	0.357	0.61	0.22–1.69	0.33	1.11	0.27–4.6	0.882	0.33	0.1–1.09	0.067
Memory	0.68	0.2–2.33	0.521	1.17	0.42–3.26	0.761	1.34	0.32–5.67	0.68	0.42	0.13–1.41	0.153

Bold numbers indicate statistically significant effect sizes with values of  $P < 0.05$ .

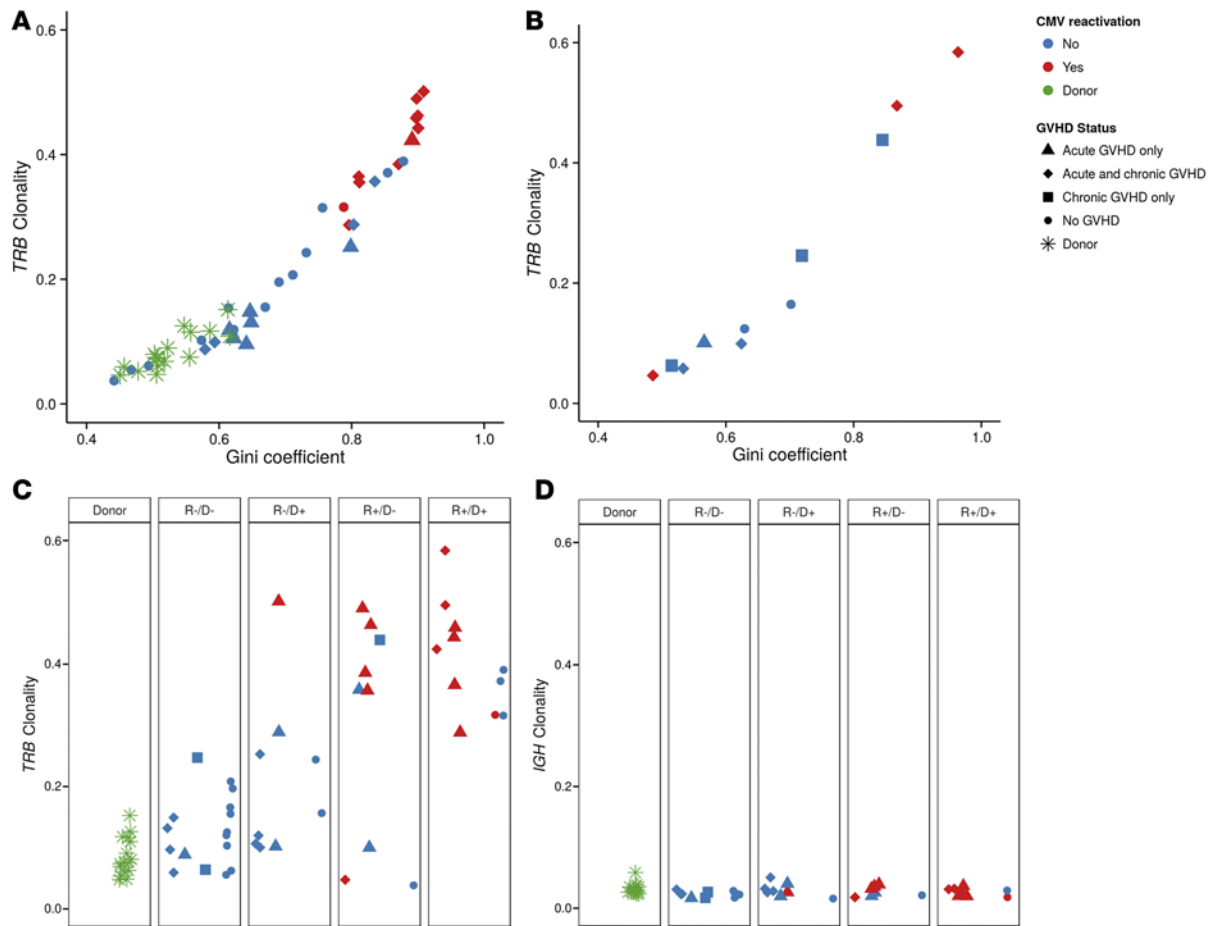
tion (Supplemental Figure 4 and ref. 26). These 2 measures of repertoire diversity were strongly correlated (Figure 3, A and B). Given this concordance, clonality was used as a measure of TCR repertoire diversity for all subsequent analyses.

*TRB* repertoires of PTCy-treated patients at 1 year after transplant were more clonal than those of donors (Figure 3, A and C). However, the *TRB* repertoire diversity in PTCy-treated patients was similar to that seen in a control cohort of 11 patients who underwent myeloablative, HLA-matched allogeneic BM transplantation using methotrexate and a CNI for GVHD prophylaxis (Figure 3B and Supplemental Tables 1 and 3). In all 45 patients (including both PTCy- and CNI-treated patients), irrespective of GVHD prophylaxis or the development of acute and/or chronic GVHD, recipient CMV seropositivity and detectable PB CMV reactivation were both highly associated with *TRB* clonality (Figure 3C). Indeed, the *TRB* repertoire diversity of recipients who were CMV seronegative before transplant approached levels seen in donors (Figure 3C). By contrast, *TRB* clonality had no relationship with the age of the recipient or donor ( $r = 0.15$ ,  $P = 0.24$ ) (Supplemental Figure 5).

Published data have demonstrated that CD4<sup>+</sup> T cell repertoires are more diverse than CD8<sup>+</sup> T cell repertoires in recipients of either allogeneic or autologous BMT (18, 21) and in healthy individuals (27). We performed *TRB* sequencing on flow cytometrically sorted CD4<sup>+</sup> and CD8<sup>+</sup> T cells from 13 recipient posttransplant PB samples and likewise found that CD4<sup>+</sup> *TRB* repertoires were consistently more diverse than CD8<sup>+</sup> *TRB* repertoires (Supplemental Figures 6 and 7). The difference in *TRB* diversity between CD4<sup>+</sup> and CD8<sup>+</sup> T cell subsets was durable and observed in PB samples obtained from 3 patients at 2.6–3.8 years after transplant (Supplemental Figure 6). CD4<sup>+</sup> and, to a greater extent, CD8<sup>+</sup> memory subsets sorted from 8 recipient or donor samples were more clonal than were CD4<sup>+</sup> and CD8<sup>+</sup> naive subsets, which were quite diverse (Supplemental Figure 7).

*Posttransplant B cell repertoires are highly diverse.* To characterize the B cell antigen receptor repertoire after PTCy, we sequenced the repertoire of *IGH* rearrangements in PB samples from 26 PTCy-treated patients at 1 year after transplant, PB samples from 14 BM donors for these patients, and PB samples from 8 CNI-treated control patients. Consistent with the flow cytometric results showing that most B cells were phenotypically naive, *IGH* sequence repertoires were exceedingly diverse in all groups; there were no discernible differences in *IGH* repertoire clonality between donor and recipient specimens or between patients depending on their GVHD or CMV status (Figure 3D).

*Posttransplant T cell repertoires are dynamic and unique to each individual.* Global similarity between antigen



**Figure 3. *TRB* and *IGH* sequence diversity in patients at 1 year after transplantation.** Unsorted peripheral blood mononuclear cells were assessed by survey-level sequencing for *TRB* and/or *IGH*. **(A and B)** There is a high correlation between 2 measurements of *TRB* diversity, clonality and Gini coefficient, for **(A)** patients treated with PTCy as single-agent GVHD prophylaxis ( $n = 34$ ) and their donors ( $n = 15$ ) (Pearson's correlation,  $r = 0.967$ ), as well as for **(B)** patients treated with methotrexate and a calcineurin inhibitor (CNI) for GVHD prophylaxis ( $n = 11$ ) (Pearson's correlation,  $r = 0.973$ ). **(C)** *TRB* clonality for all 45 patients (both PTCy-treated and CNI-treated) is more strongly associated with recipient CMV seropositivity and posttransplant CMV peripheral blood reactivation than with history of acute or chronic graft-versus-host disease (GVHD). **(D)** *IGH* clonality in PTCy-treated recipients ( $n = 26$ ) and CNI-treated recipients ( $n = 8$ ) is similar to that of donors ( $n = 14$ ), irrespective of CMV seropositivity, CMV reactivation, or GVHD history. R±/D± indicates recipient (R) and donor (D) CMV seropositivity or seronegativity.

receptor repertoires was assessed by calculating the Bhattacharyya coefficient, a metric that sums over all shared sequences the geometric means of their frequencies in the 2 repertoires (19, 28). This measurement was directly compared with the similarity score, which is a function of the number of *TRB* sequences found in both repertoires and the total number of *TRB* sequences found in the union of the 2 repertoires (25), and the 2 were found to be highly correlated ( $r = 0.97$ ,  $P < 2.2 \times 10^{-16}$ ) (Supplemental Figure 8). Either measurement can be used to assess the degree of overlap between any 2 *TRB* repertoires, such as the repertoires present in 2 different patients, the repertoires of donors and their marrow graft recipients, or the repertoires existing in a given patient at different times. Given prior use of the Bhattacharyya coefficient in other alloBMT studies (19), we used that parameter primarily for these analyses. Bhattacharyya coefficients range from 0, which indicates no shared sequences, to 1, which indicates complete identity. In practice, the calculated Bhattacharyya coefficient between independent samples of a single PB T cell repertoire is far from unity, due primarily to the fact that current antigen receptor sequencing platforms can sample, at best, a small fraction of the receptor diversity found in a typical blood repertoire (29). Indeed, in this study, we observed a Bhattacharyya coefficient of 0.514 between 2 independent samples from the same PBMC specimen (Supplemental Figure 9). Using the Bhattacharyya coefficient, the calculated *TRB* repertoire similarity between any 2 of the 16 donors or between any 2 of the 34 patients for whom 1-year posttransplant specimens were available was extremely low (Figure 4A), demonstrating that the *TRB* repertoires in any

**Table 3. Characteristics of PTCy-treated patients from whom PBMC samples were obtained at 1 year after transplant and assessed via antigen receptor sequencing**

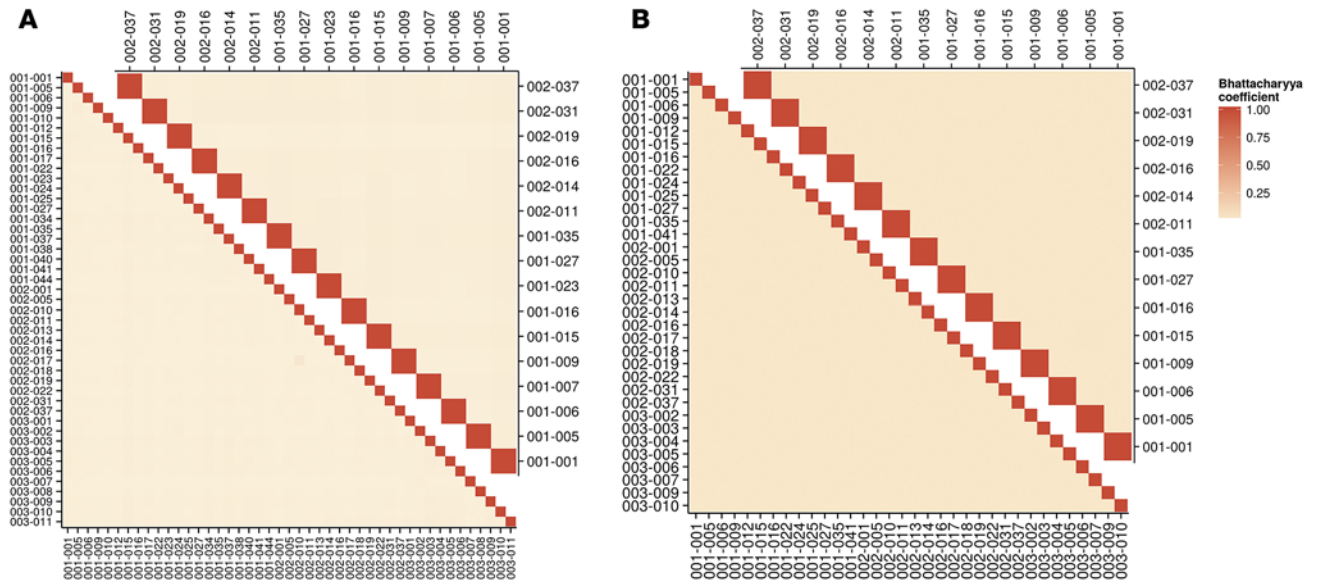
	No GVHD ( <i>n</i> = 14)	aGVHD only ( <i>n</i> = 13)	aGVHD and cGVHD ( <i>n</i> = 7)	All Patients ( <i>n</i> = 34)
Age, median (range)	45 (22–60)	57 (29–63)	45 (25–63)	49 (22–63)
Sex, female, <i>n</i> (%)	11 (79%)	8 (62%)	2 (29%)	21 (62%)
Disease, <i>n</i> (%)				
AML	8 (57%)	12 (92%)	3 (43%)	23 (68%)
MDS	0 (0%)	1 (8%)	1 (14%)	2 (6%)
ALL	6 (43%)	0 (0%)	3 (43%)	9 (26%)
Donor relatedness, <i>n</i> (%)				
Related	10 (71%)	6 (46%)	2 (29%)	18 (53%)
Unrelated	4 (29%)	7 (54%)	5 (71%)	16 (47%)
Donor age, median (range)	47 (17–58)	46 (22–67)	46 (31–61)	46 (17–67)
Donor sex, female, <i>n</i> (%)	9 (64%)	6 (46%)	4 (57%)	19 (56%)
Female-into-male allografting, <i>n</i> (%)	3 (21%)	2 (15%)	3 (43%)	8 (24%)
CMV serologies, <i>n</i> (%)				
D <sup>-</sup> R <sup>-</sup>	7 (50%)	1 (8%)	3 (43%)	11 (32%)
D <sup>+</sup> R <sup>-</sup>	2 (14%)	2 (15%)	3 (43%)	7 (21%)
D <sup>+</sup> R <sup>+</sup>	4 (29%)	4 (31%)	1 (14%)	9 (26%)
D <sup>-</sup> R <sup>+</sup>	1 (7%)	6 (46%)	0 (0%)	7 (21%)
CMV Reactivation, <i>n</i> (%)	1 (7%)	9 (69%)	1 (14%)	11 (32%)
Maximal aGVHD grade (%)				
0	14 (100%)	0 (0%)	0 (0%)	14 (41%)
I	0 (0%)	1 (8%)	0 (0%)	1 (3%)
II	0 (0%)	10 (77%)	4 (57%)	14 (41%)
III–IV	0 (0%)	2 (15%)	3 (43%)	5 (15%)
Steroids used for GVHD treatment, <i>n</i> (%)	N/A	12 (92%)	7 (100%)	19 (95%)
Nonsteroidal(s) used for GVHD treatment, <i>n</i> (%)	N/A	11 (85%)	7 (100%)	18 (90%)

Recipient 001-007 was excluded from this table because that individual suffered nonrelapse mortality on posttransplant day +64; therefore, no 1-year posttransplant PBMC sample was available. PTCy, posttransplantation cyclophosphamide; PBMC, peripheral blood mononuclear cell; GVHD, graft-versus-host disease; aGVHD, acute GVHD; cGVHD, chronic GVHD; AML, acute myeloid leukemia; MDS, myelodysplastic syndrome; ALL, acute lymphoblastic leukemia; CMV, cytomegalovirus; D, donor; R, recipient.

individual are largely unique to each individual recipient. There was likewise negligible similarity between the *IGH* repertoires of any 2 donors or any 2 recipients at 1 year after transplant (Figure 4B), demonstrating that B cell repertoires also are largely unique to each individual. Not surprisingly, there was little similarity between the CD4<sup>+</sup> and CD8<sup>+</sup> T cell repertoires observed in a given individual (not shown).

To assess how *TRB* repertoires present in individual patients evolved over time, sequencing was performed on serially acquired posttransplant PB samples from 12 patients. Nine of these 12 patients had related donors on whom PB *TRB* sequencing had also been performed, making it possible to evaluate how the recipient repertoire changed over time vis-à-vis the input donor repertoire. In all cases, the *TRB* repertoires observed at 1 month after transplant in the recipients' PB bore relatively low similarity to the repertoires observed in their own BM donors (Figure 5A). In fact, the mean Bhattacharyya coefficient between the donor and corresponding recipient 1-month posttransplant repertoires (*n* = 9) was only marginally higher than the mean Bhattacharyya coefficient observed between the donor and recipient pretransplant repertoires (*n* = 4) (Figure 5A). The early posttransplant repertoires were also often quite diverse. Serial profiling of the posttransplant recipient repertoire over the balance of the first posttransplant year revealed that, in most patients, the *TRB* repertoires seen at later time points demonstrated less diversity and increasing similarity with the corresponding donor repertoires, particularly in recipients with CMV-seropositive donors (Figure 5, A–F).

In one case (001-006), the rapid turnover and dynamic evolution of the blood *TRB* repertoire during the first posttransplant year were particularly dramatic (Figure 6). Absolute lymphocyte recovery was brisk in this recipient (Figure 6A). Grade I skin aGVHD developed on day +21, progressing to grade II skin aGVHD on day +48, for which the patient finally required corticosteroid treatment starting on day +62. The blood *TRB* repertoire present on day +62 was quite diverse (clonality 0.05) but dissimilar to that of the



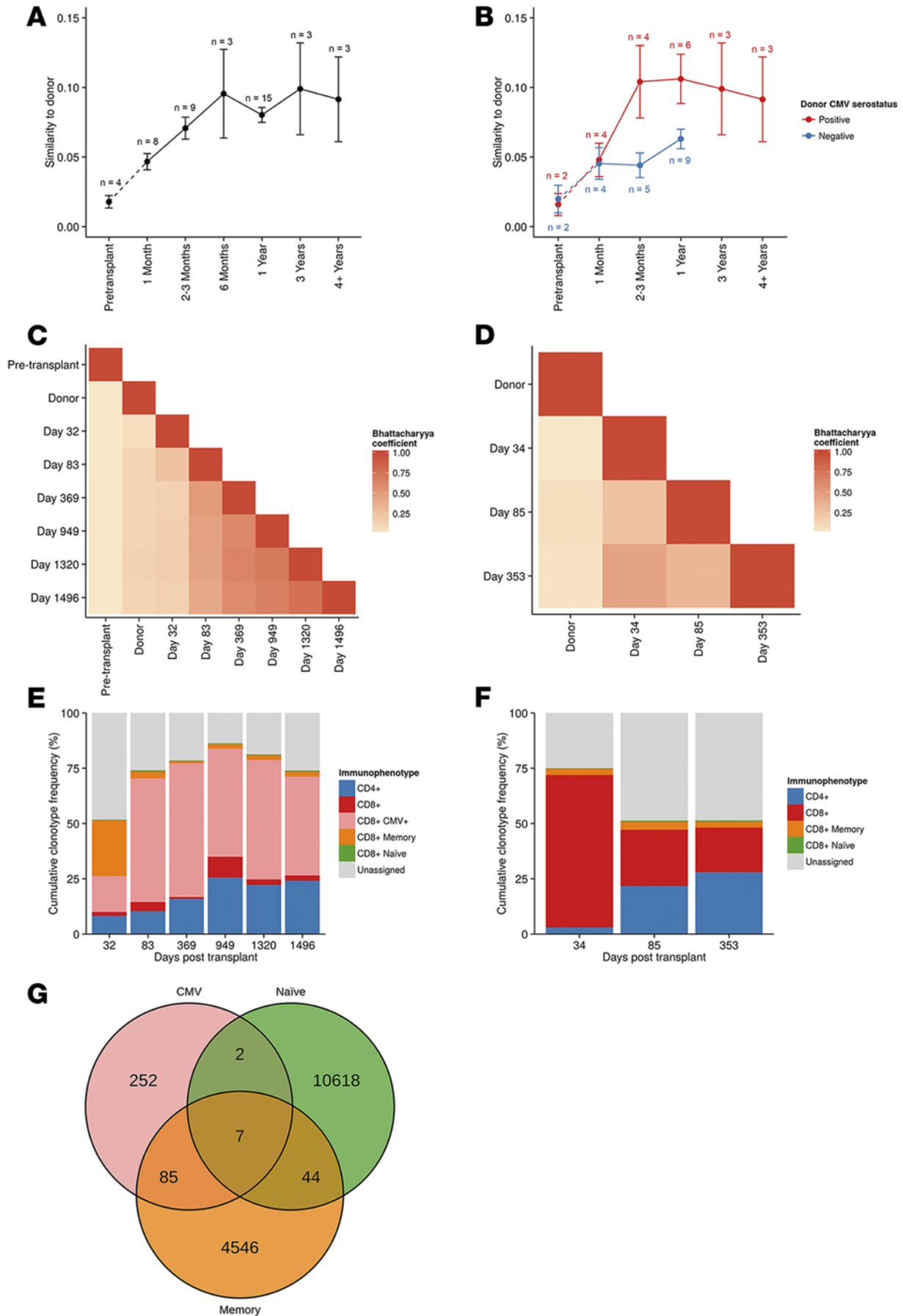
**Figure 4. Posttransplant *TRB* repertoires are unique to unrelated recipients.** Shown are pairwise similarity, calculated using the Bhattacharyya coefficient, between (A) *TRB* and (B) *IGH* repertoires of donors on the day of allograft donation (upper right in A and B) or of recipients at 1 year after transplant (lower left in A and B).

donor (Bhattacharyya coefficient 0.019). Around day +80, the ALC began to decline and reached a nadir of 101 cells/ $\mu$ l on day +100. Thereafter, the ALC began a slow recovery that continued for the balance of the first posttransplant year. *TRB* sequencing of the PB on days +188 and +363 revealed that the repertoire established during this period of recovery was less diverse (clonality 0.14 and 0.15, respectively) but also demonstrated far more similarity with the input donor repertoire (Bhattacharyya coefficients 0.111 and 0.117, respectively) (Figure 6A).

The evolution of the posttransplant *TRB* repertoire toward a more donor-like repertoire over the first posttransplant year was reflected in the expansion of *TRB* sequences present in donor PB that were observed at high frequency in the recipient at later posttransplant time points (representative data in Figure 6, C and D). Most such sequences were not observed in the initial recipient posttransplant blood samples obtained at 1–2 months and were first detected at  $\geq 3$  months after transplant. Although all patients received BM grafts, donor PB and BM displayed a high degree of similarity, with the same dominant clones generally present in both compartments (Figure 6B and Supplemental Figure 10); this consistency between donor PB and BM supports a true dynamic change in the *TRB* repertoire rather than an artifact of clone compartmentalization between PB and BM. Dominant clones in the PB of the recipient at  $\geq 3$  months after transplant frequently shared identity with dominant clones found in the CD8<sup>+</sup> memory compartment of the donor (Figure 5, E and G, and Figure 6, B–D). In several cases, the trend toward increasing similarity between input donor and posttransplant recipient repertoires over time was primarily due to expansion of donor CD8<sup>+</sup> memory-derived CMV-specific T cell clones (Figure 5, E and G). However, a similar pattern of repertoire evolution, with a transient initial repertoire bearing little similarity to that of the donor, followed by steadily increasing similarity to the input donor repertoire over the balance of the first posttransplant year, was observed in patients who were recipient and/or donor CMV seronegative (Figure 6) or who did not experience GVHD.

Overall similarity between the donor repertoire and the posttransplant recipient repertoire appeared to peak at 6–12 months after transplant. In samples obtained between 2.5–5.2 years after transplant from 3 recipients of related donor grafts, the similarity between donor and recipient appeared to stabilize or perhaps even slowly decline after 1 year; however, dominant, primarily CD8<sup>+</sup>, *TRB* sequences observed in the donor were still commonly detected, and the *TRB* repertoires remained significantly more clonal than the input donor repertoires (Figure 5, A–C, E, and G).

*Shared TRB sequences can serve as markers of antigen-specific T cell responses.* Although similarity analysis of donor and recipient *TRB* repertoires conclusively established the overwhelming unique nature of each

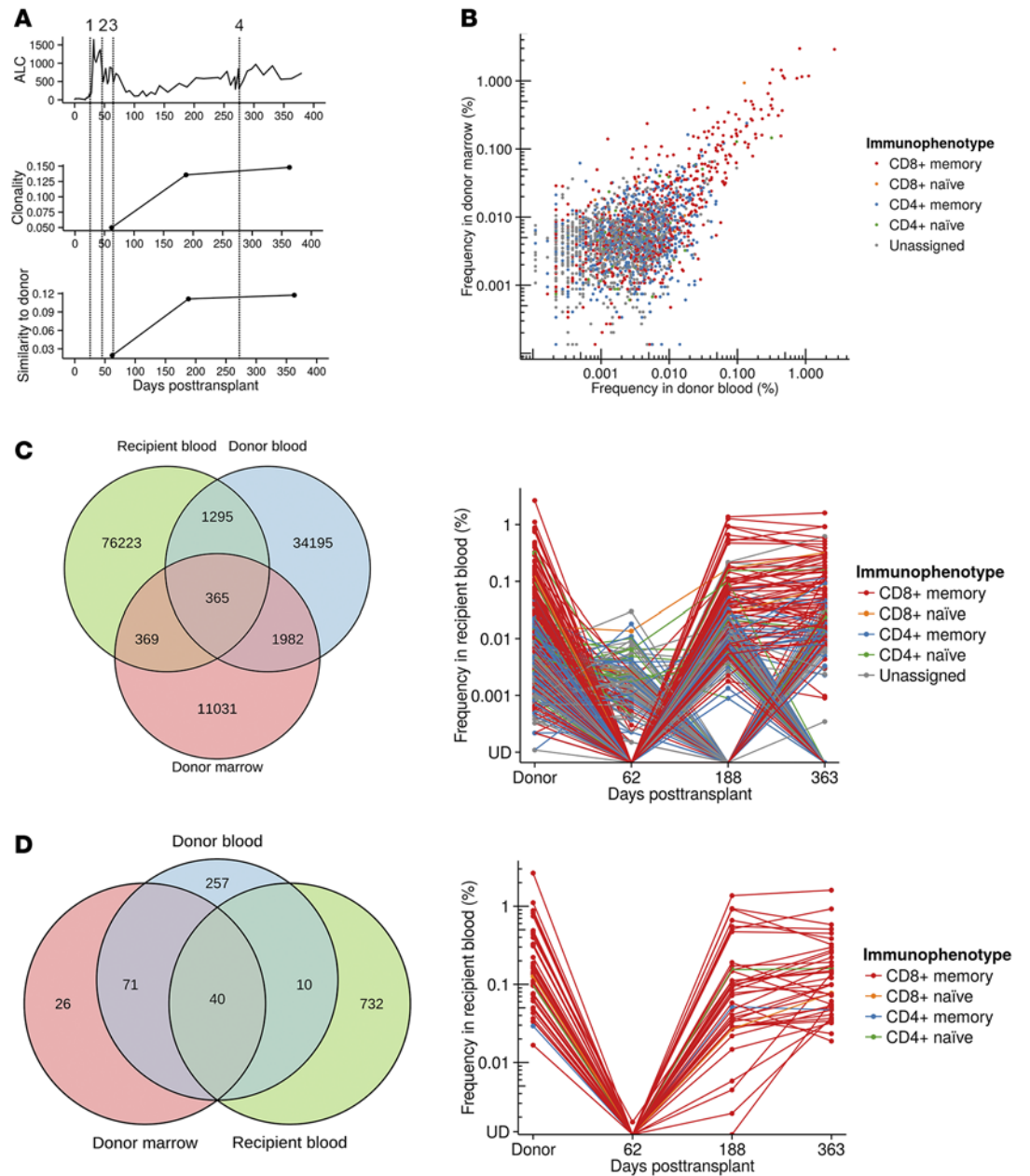


**Figure 5. Posttransplant *TRB* repertoires become more similar to input donor repertoires over time during the first posttransplant year. (A)** Mean Bhattacharyya coefficients between the *TRB* repertoires of recipients at the indicated time points before or after transplant and the *TRB* repertoires of their related donors on the day of marrow donation. The number of donor/recipient pairs for whom a Bhattacharyya coefficient at each time point could be calculated is indicated. **(B)** Mean Bhattacharyya coefficients from **A** separated into 2 groups by donor CMV serostatus. The evolution toward a more “donor-like repertoire” over the first posttransplant year was more prominent in recipients of allografts from CMV-seropositive donors, in part due to the high frequency of CMV-specific memory T cells in both the donor and posttransplant recipient. **(C)** Pairwise *TRB* repertoire similarity for donor and recipient 002-037. Both were CMV seropositive, and detectable CMV reactivation occurred in the recipient on day +41 after transplant. The donor repertoire was assessed on the day of allograft donation, and the recipient repertoire was assessed at the indicated time points before and after transplant. **(D)** Pairwise *TRB* repertoire similarity matrix for donor and recipient 002-031, who were both CMV seronegative. **(E)** Cumulative frequency of *TRB* sequences in the blood of recipient 002-037 on the indicated posttransplant days that could be presumptively mapped back to the donor’s peripheral blood (PB) repertoire. CD4<sup>+</sup>, CD8<sup>+</sup>, CD8<sup>+</sup> naive, and CD8<sup>+</sup> memory compartments were defined based on sorting of donor PB. The CD8<sup>+</sup> CMV pp65-specific tetramer<sup>+</sup> (HLA-A2- and HLA-B7-restricted) compartment was defined based on tetramer sorting of recipient samples at posttransplant days +83, +369, and +1,320. It was not possible to sort CMV pp65-specific tetramer<sup>+</sup> cells from donor PB due to insufficient sample availability. **(F)** Cumulative frequency of *TRB* sequences in the blood of recipient 002-031 on the indicated posttransplant days that could be presumptively mapped back to the CD4<sup>+</sup>, CD8<sup>+</sup>, CD8<sup>+</sup> naive, or CD8<sup>+</sup> memory compartments of donor 002-031’s PB repertoire. **(G)** Venn diagram illustrating the overlap between the recipient CMV pp65 tetramer<sup>+</sup> compartment and the *TRB* sequences that could be mapped back to the naive or memory compartment of donor 002-037, showing that nearly all CMV pp65 tetramer<sup>+</sup> T cells that could be tracked were able to be traced to the memory compartment.

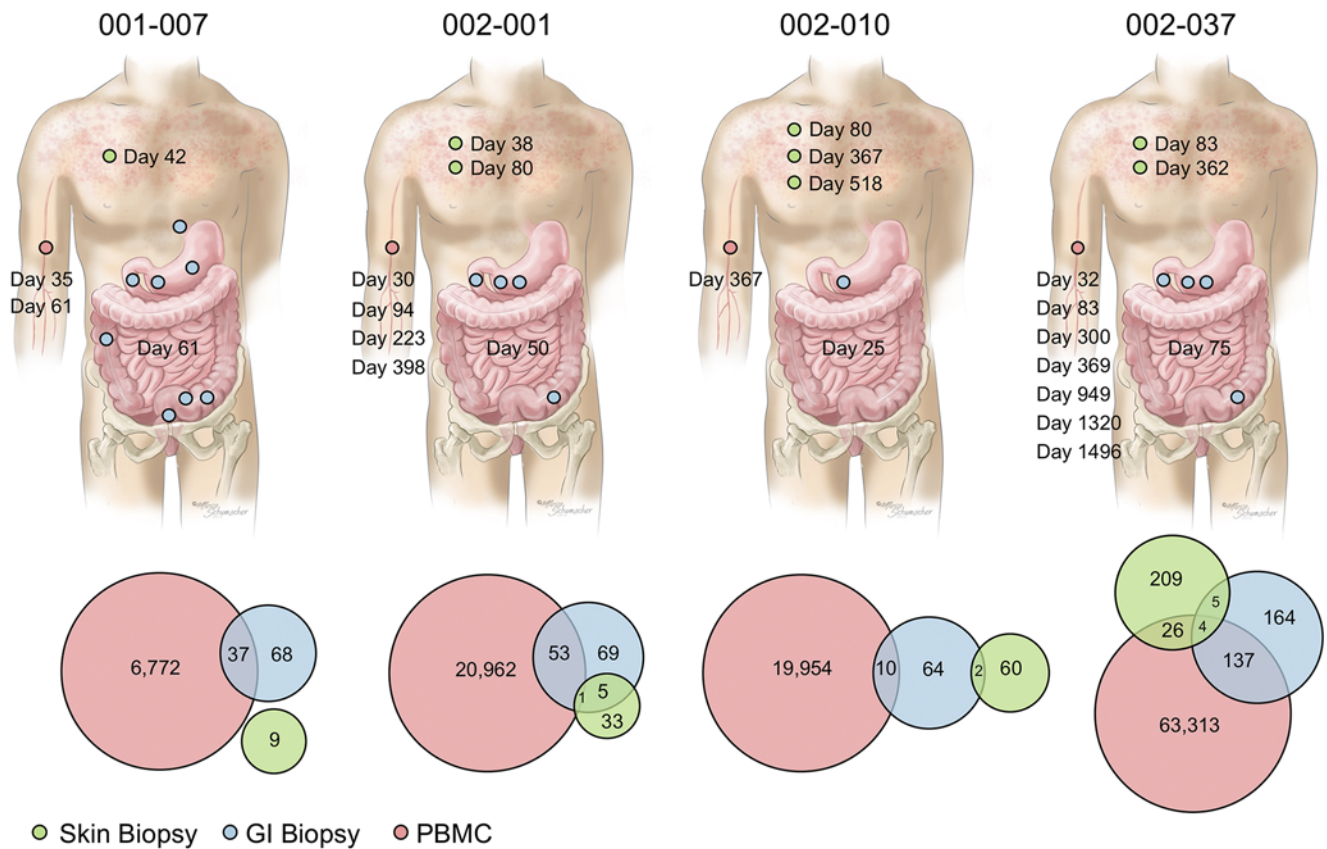
individual’s repertoire (Figure 4A), *TRB* sequences utilizing identical V, D, and J gene segments and encoding identical CDR3 amino acid sequences were often identified in 2 or more unrelated donors and/or recipients. The observation of many such “shared” or “public” *TRB* sequences is likely attributable, at least in part, to the high probability of their generation (30). However, *TRB* and *TRA* sequences associated with T cell responses to specific antigens, often CD8<sup>+</sup> responses to herpesvirus (e.g., CMV and EBV antigens), and observed in many individuals have frequently been described. Having assembled a database of more than 3,600 *TRB* sequences known to be associated with CD8<sup>+</sup> or CD4<sup>+</sup> responses to viral and other pathogens, tumor antigens, self-antigens, and alloantigens, we systematically searched for these public sequences in the ensemble of sequence datasets generated in this study, which collectively comprised more than 200 million sequence reads and more than 2.8 million unique productive *TRB* sequences.

*TRB* sequences associated with CD8<sup>+</sup> T cell responses to prototypic CMV epitopes, such as pp65<sub>265–275</sub>/HLA-B7 and pp65<sub>495–503</sub>/HLA-A2, were detected in unsorted PBMCs from several HLA-B7<sup>+</sup> or HLA-A2<sup>+</sup> donors and recipients, respectively. *TRB* sequencing of CD8<sup>+</sup> T cells sorted from several of these individuals using pp65<sub>265–275</sub>/HLA-B7 and pp65<sub>495–503</sub>/HLA-A2 tetramers confirmed that these CMV-specific public *TRB* sequences were readily identified in the appropriate tetramer<sup>+</sup> subsets (Supplemental Table 4). Public *TRB* sequences associated with responses to these 2 CMV antigens also were found in GI biopsy specimens from an HLA-A2<sup>+</sup>, -B7<sup>+</sup> recipient (002-037) who had developed CMV reactivation 1 month earlier (Supplemental Figure 11). In addition to CMV-associated *TRB* sequences, public *TRB* sequences associated with CD8<sup>+</sup> T cell responses to EBV were identified in several donors and recipients. Public *TRB* sequences associated with HLA-B8- or -B35-restricted CD8<sup>+</sup> T cell responses to EBV were consistently detected in the blood of an HLA-B8<sup>+</sup>, -B35<sup>+</sup> recipient (002-014) at 5 separate assessments over a period of 5 years after transplant, spanning days +32 and +1,755 (Supplemental Figure 12 and Supplemental Table 5). A transient increase in the frequency of these EBV-specific public *TRB* sequences in the blood on day +367 coincided with detection of low-level EBV plasma viremia (Supplemental Figure 12).

*TRB* sequences associated with CD8<sup>+</sup> T cell responses to autosomal- and Y chromosome-encoded (H-Y-encoded) minor histocompatibility antigens and to known tumor antigens were also observed in our cohort. *TRB* sequences carried in CD8<sup>+</sup> T cells recognizing the minor antigens HA-1 (31–34), HA-3 (35), and HA-8 (36); the prototypic H-Y antigen encoded by *KDM5D* (37); and the tumor antigen NY-ESO-1<sub>157–165</sub> were observed in blood or tissues of recipients carrying the corresponding MHC class I restricting allele (HLA-A\*02:01 or \*01:01). With few exceptions, however, these sequences were present at low frequency (<1% of unsorted T cells and generally <0.01%). They were also seen at low frequency (<0.01%) in some individuals who did not express the corresponding class I MHC restricting allele. In a small number of recipients, sequences associated with CD8<sup>+</sup> T cells recognizing defined minor histocompatibility or tumor antigens were observed at high frequency in either blood or tissue. For example, 3 different *TRB* sequences encoding the TCR β chain carried in a CD8<sup>+</sup> T cell clone specific for the prototypic H-Y antigen *KDM5D*<sub>311–319</sub>/HLA-A2 previously established in our lab accounted for 36% of sequences in a skin biopsy from the HLA-A2<sup>+</sup> male recipient (002-037) who had a female donor and experienced skin and GI aGVHD. This sequence was not found at frequency >0.01% in any other individual in the cohort. In a different HLA-A2<sup>+</sup> patient with AML (001-006), 11 distinct *TRB* nucleotide sequences encoding the CDR3



**Figure 6. Dynamic repertoire changes occur after transplant.** Data from donor and recipient 001-006, both of whom were CMV seronegative, are shown. T cells were sorted from donor peripheral blood (PB) and BM samples by FACS and assessed by *TRB* sequencing. Unfractionated peripheral blood mononuclear cells collected from the recipient on posttransplant days +62, +188, and +363 were also assessed by *TRB* sequencing. **(A)** Absolute lymphocyte count (ALC) (top), clonality of the recipient *TRB* repertoire (middle), and similarity (Bhattacharyya coefficient) of the 001-006 recipient repertoire to the input donor repertoire (bottom) during the first posttransplant year. Posttransplant days of onset of grade I (1) and grade II (2) skin acute GVHD, start of immunosuppression with corticosteroids (3), and chronic GVHD (4) are indicated by the dotted vertical lines. **(B)** Relative frequency in donor 001-006's PB and BM of *TRB* sequences that were observed in both physical compartments. The color of the point representing each sequence indicates the donor T cell phenotypic compartment in which that sequence was observed and from which it was sorted. **(C and D)** Left panels: Venn diagrams showing the overlap in unique *TRB* sequences between recipient 001-006 posttransplant PB, donor 001-006 PB, and donor 001-006 BM for **(C)** all *TRB* sequences or **(D)** the top first percentile of *TRB* sequences. The sequences observed in the 3 recipient posttransplant samples (days +62, +188, and +363) were grouped together for this analysis. Right panels: Relative frequency in the donor or in the recipient at the indicated time points of the *TRB* sequences that were observed in all 3 compartments (donor PB, donor BM, recipient PB). The color of the line representing each sequence indicates the donor T cell compartment (CD4<sup>+</sup> naïve, CD8<sup>+</sup> naïve, CD4<sup>+</sup> memory, and CD8<sup>+</sup> memory) in which that sequence was observed. *TRB* sequences that could not be unambiguously assigned to a specific donor T cell compartment are coded as "unassigned." UD, undetected.

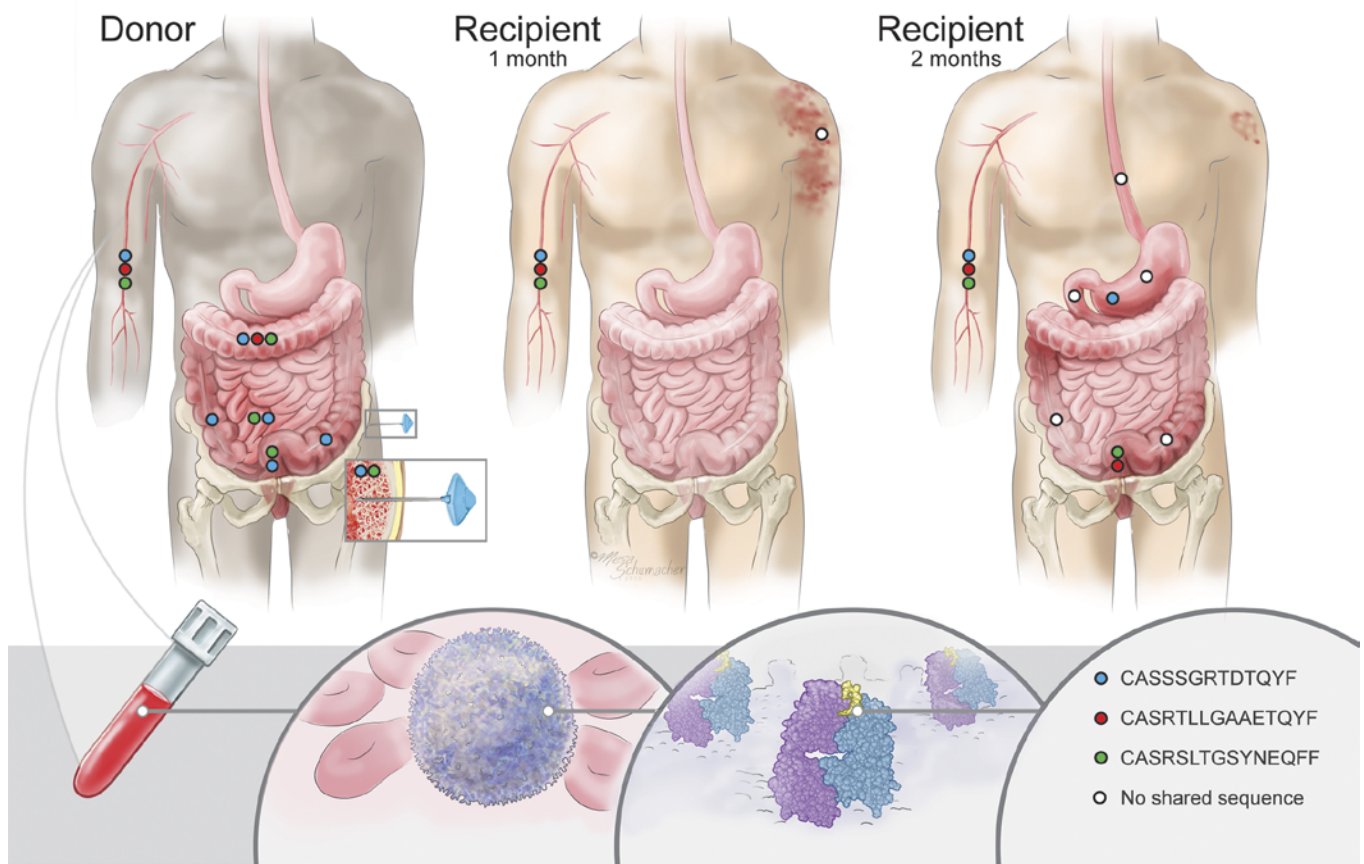


**Figure 7. Overlap between TRB repertoires observed in blood or during evaluation for GVHD in the gastrointestinal tract and skin.** TRB sequencing was retrospectively performed on biopsies obtained from 4 recipients on this study, identified by the codes along the top of the figure, who underwent evaluation of both GI tract and skin for suspected GVHD. The posttransplant days on which GI and skin biopsies were obtained from each recipient are superimposed on the abdomen and chest, respectively, of the cutaway torso representing each patient. Recipients 001-007, 002-001, and 002-010 were female, but a male torso is displayed for all recipients for the sake of simplicity. The approximate location from which each GI tract biopsy was obtained is indicated by the blue dots; the location from which the skin biopsies (green dots) were obtained is not indicated. The days below the red dot superimposed on the vein in the right arm of each torso indicate the posttransplant dates of collection of blood used for TRB sequencing of unfractionated and/or sorted cells. The Venn diagram below each torso indicates the total number of unique TRB sequences collectively found in the blood samples (PBMC, red), skin biopsies (green), or GI biopsies (blue) obtained from each recipient, as well as the number of TRB sequences that were shared between the 3 compartments.

region of a previously described NY-ESO-1<sub>157-165</sub>/HLA-A2-specific TCR (38) collectively comprised 4.4% of the TRB sequences in the blood at 1 year after transplant.

*T cell repertoires in GVHD target organs appear distinct from repertoires in the blood.* Although prospective analysis of the T cell repertoires in GVHD target organs was not performed in this study, TRB sequencing was performed retrospectively on DNA extracted from skin and/or GI tract biopsies obtained from a subset of 8 patients who were being evaluated for GVHD. In general, the mass of tissue from which DNA was extracted was extremely low; the median estimated number of T cell genomes present in the biopsy specimens was 98 for the skin biopsies ( $n = 16$ ; range = 12–508) and 117 for the GI tract biopsies ( $n = 22$ ; range = 3–1799). In contrast, the estimated number of T cell genomes sampled in unsorted blood and marrow samples ranged from approximately 5,000–100,000. Thus, the depth of sampling of the TRB repertoires of T cells infiltrating skin and GI tract in these 8 recipients was, on average, at least 50-fold lower than for blood or BM. Furthermore, the tissue biopsies were generally not acquired on the same day as blood samples, as the former were acquired for clinical care and the latter were acquired for research.

Nonetheless, analysis of the TRB sequences observed in the tissue biopsy specimens proved quite informative. Given that tissues involved with GVHD are inflamed and often hyperemic, one might expect that TRB sequences observed at high frequency in the blood of patients with GVHD might also be observed in skin and GI tract biopsies obtained at similar times. However, for a given individual, the global TRB repertoires present in blood had surprisingly limited overlap with the repertoires found in the GI tract and little apparent overlap with the repertoires found in the skin (Figure 7). Moreover, with few exceptions,



**Figure 8. Tracking T cell clones from donor blood, BM, and tissue to recipient blood and tissue.** A CMV-seropositive patient (001-007) with AML received an allograft from her CMV-seronegative sister who had a history of inflammatory bowel disease. *TRB* sequencing was performed on donor blood and BM collected on the day of donation, and on tissue from gastrointestinal biopsies performed 8 months prior to donation that showed lymphocytic colitis. Three *TRB* sequences, denoted by the amino acid sequences encoded in their CDR3 regions, were shared between the BM, blood, and gastrointestinal biopsies from the donor and between the blood and gastrointestinal biopsies from the recipient. Although all 3 sequences were observed in the posttransplant blood and gastrointestinal tract biopsies of the recipient, who died of complications of gastrointestinal grade III acute GVHD, none were observed in a skin biopsy from the recipient, who also showed acute skin GVHD. Circles indicate the site of tissue sampling, and the color of the circle indicates the *TRB* sequence detected at that site.

dominant clones in blood were infrequently observed in GI tract biopsies and almost never observed in skin biopsies. For example, the most frequent *TRB* sequence observed in the blood of subject 002-037 on day +83, encoding the CDR3 amino acid sequence CASSLQGREKLF, comprised 10.1% of the productive *TRB* sequences in the blood on day +83 but was not observed in any of the 4 GI tract biopsies obtained on day +75 from the gastric antrum (2 biopsy samples), duodenum, and sigmoid colon, nor was it observed in the skin on day +83. A *TRB* sequence encoding the CDR3 amino acid sequence CASSPSRNTEAFF was the most notable exception to this pattern. This sequence comprised 0.37% of the *TRB* reads in the CD8<sup>+</sup> memory compartment in donor PBMC. It comprised 1.73% of the sequences in the blood of subject 002-037 on day +83 and was found in 3 of the 4 day +75 GI biopsies at frequencies of 1.78% (antrum), 0.69% (duodenum), and 0.99% (sigmoid); it was not observed in the skin biopsies obtained on days +83 or +369. This CASSPSRNTEAFF-encoding *TRB* sequence is a “public” sequence associated with CD8<sup>+</sup> T cells specific for CMV pp65<sub>265-275</sub>/HLA-B7 (39); it was also observed in CD8<sup>+</sup>, CMV pp65<sub>265-275</sub>/HLA-B7-tetramer<sup>+</sup> cells sorted from the blood of this subject (who was HLA-B7<sup>+</sup> and experienced CMV reactivation on day +42) on days +83 and +369. Of note, none of these 4 GI biopsies were formally interpreted as exhibiting histopathologic evidence of CMV infection.

The failure to detect *TRB* sequences that were common in the blood in the *TRB* repertoires of skin biopsy specimens was complemented by the failure to detect in blood most of the sequences observed in skin. None of the 3 putative H-Y-specific *TRB* sequences that collectively comprised 36% of the sequences in the 002-037 day +369 skin biopsy were observed in the blood of this subject on days +32, +83, +300,

+369, +949, +1,320, or +1,496. Sequences encoding a TCR  $\beta$  chain identical to that carried in a CD8<sup>+</sup> T cell clone previously isolated in our lab and specific for the HLA-A1–restricted HA-3 minor histocompatibility antigen were observed in skin biopsies from 3 different HLA-A1<sup>+</sup> recipients — 002-014, 002-031, and 002-037 — but were not observed in the blood of these subjects at any time point. The HA-3 minor antigen is expressed in a wide range of tissues, including the skin (40). The highest frequencies of the HA-3–associated sequences were observed in the skin biopsies obtained on days +80 and +367 from subject 002-014, who was mismatched (donor negative and recipient positive) for the SNP (rs2061821) that creates the HA-3 minor antigen. These observations suggest a high degree of compartmentalization of the skin-resident T cell repertoires in recipients of HLA-matched alloBMT using PTCy, although limited sequencing depth possible in these skin samples precludes definitive conclusions.

*Potentially pathogenic T cell clones can be traced from donor into recipient.* Although our sequencing data allowed us to follow specific T cell clones from donors into blood or tissues of recipients, the presence of specific *TRB* sequences in GVHD target organs does not establish that the clones carrying those sequences were in fact pathogenic for GVHD. One particular case, however, provided suggestive evidence for an association of specific *TRB* sequences and GI pathology. In this case (001-007), a 60-year-old CMV-seropositive female with relapsed acute myeloid leukemia was transplanted with BM harvested from her 59-year-old HLA-matched CMV-seronegative sister. The sister had a history of inflammatory bowel disease (IBD) and, 8 months prior to marrow donation, had an IBD flare during which lower GI tract biopsies were obtained, confirming lymphocytic colitis involving the terminal ileum, colon, and rectum. The recipient did not experience CMV reactivation after transplant but did develop aGVHD that ultimately proved fatal. GVHD initially presented as grade II skin aGVHD on day +41 and progressed to grade III GI aGVHD on day +52, despite immunosuppressive therapy. On day +61, GI biopsies of the upper and lower GI tract were performed, confirming histopathologic evidence of moderate to severe GVHD. The patient expired on day +64 secondary to infection in the setting of severe GI aGVHD. *TRB* sequencing of blood, marrow, and GI biopsies from the donor — as well as blood, skin, and GI biopsies from the recipient — identified 3 *TRB* sequences that were shared between donor and recipient blood, marrow, and GI tissues. All 3 sequences were frequently detected in the PB and GI tracts of both donor and recipient (2%–4% of GI biopsy *TRB* sequences in the recipient) and were identical to sequences observed in sorted CD8<sup>+</sup> memory T cells in the donor's PB and BM (Figure 8), suggesting that they might be carried in pathogenic T cell clones targeting GI tissue that were passed from donor to recipient. None of these 3 clones was found in the skin biopsy obtained on day +42. Two of the 3 *TRB* sequences were not observed in any other donor or recipient in this study; the other (encoding the CDR3 amino acid sequence CASSSGRTDTQYF) was observed at low level (<0.007%) in PBMCs from 7 other individuals (3 donors, 4 recipients).

## Discussion

In this study, we comprehensively explored lymphocyte reconstitution in patients treated with myeloablative, HLA-matched alloBMT using PTCy as single-agent GVHD prophylaxis. The assessment of lymphocyte reconstitution with concurrent flow cytometry and antigen receptor sequencing yielded valuable complementary information that would not be as readily apparent if either modality had been used in isolation. Flow cytometric analysis revealed rapid numeric recovery of particular lymphocyte subsets and highlighted the dominant roles played by recipient CMV seropositivity and history of aGVHD/immunosuppressive treatment in shaping immune reconstitution. Our sequencing analysis showed that, despite rapid numerical lymphocyte reconstitution in our patients, the PB T cell repertoire is quite clonal during the first posttransplant year and remains so for multiple years after transplant. Furthermore, there is rapid turnover in the PB *TRB* repertoires between those seen within the first few posttransplant months and those from later posttransplant time points. During the first posttransplant year, there is progressive reversion to a more donor-like repertoire apparently deriving from the donor memory compartment. In the small subset of recipients whose *TRB* repertoires were evaluated at 2 or more years after transplant, the similarity between donor and recipient repertoire observed at 1 year after transplant appeared to stabilize or even slowly decline thereafter. Although the depth to which we could sample the *TRB* repertoires of GVHD target organs such as skin and GI tract was limited, our sequencing analysis identified a significant degree of compartmentalization of the repertoires present in blood and tissues (41).

Using primarily flow cytometric assessment, 2 groups recently reported that the vast majority of T cells recovering within the first 1–3 months after PTCy-based GVHD prophylaxis appear to derive from

the donor naive compartment (13, 14). These studies used PB stem cell allografts in which the number of donor T cells available for sequence analysis was much greater than available in our studies, which used marrow grafts. Given that naive T cell clones would be expected to be present in the donor at low frequency, our inability to demonstrate expansion of naive-derived clones in our patients is not surprising. Even so, we demonstrated that the naive-derived repertoire established in the first posttransplant month does not persist and that memory-derived clones regain dominance over longer followup, suggesting that some memory T cells also survive PTCy. These memory T cells apparently do not expand as robustly in the initial lymphopenic engraftment period but expand later after transplant, possibly as a result of pathogen antigen exposure.

Our flow cytometric data would suggest that CMV may be a particularly potent stimulator of this late expansion. Although dominant public CMV- and EBV-specific clones have been described (42, 43), the public component of T cell responses to herpesvirus antigens may be a small component of the total response (39). We were able to identify public *TRB* sequences associated with CMV- and EBV-specific TCRs in a subset of patients, but we did not find them in all recipients who were at risk for CMV reactivation by virtue of donor or recipient seropositivity. Given that no clinically significant CMV or EBV disease occurred in our cohort (8, 9), we infer that the CMV- and EBV-specific repertoires established in our patients were primarily unique to each individual.

The profound influence of CMV on the immune system in clinically healthy individuals is well documented (44–47). More recently, the substantial impact that CMV has on shaping lymphocyte reconstitution in patients undergoing alloBMT has been appreciated (23, 24, 48), although there has been no consensus achieved over whether CMV seropositivity or detectable CMV reactivation plays the dominant role. Consistent with a previous report (23), we found that recipient CMV seropositivity appears to have a more significant impact on CD8<sup>+</sup> and CD4<sup>+</sup> EM and TEMRA numerical recovery after transplant than does CMV reactivation, particularly in the early posttransplant period. This may reflect adequate maintenance of CMV-specific immunity after transplant such that low-level CMV reactivation in the blood and/or tissues is immediately controlled before it can become detectable by standard laboratory testing, while still being capable of providing sufficient antigen stimulation for CMV-specific CD8<sup>+</sup> T cell expansion. Although CMV may play a dominant role, posttransplant numerical reconstitution of T cells is also likely driven by other pathogens, including EBV (49, 50). Whether latent viral infections and the associated memory inflation that consequently occurs within the EM and TEMRA T cell subsets contribute to functional impairment in the T cell repertoires in our patients requires further evaluation (48, 51).

Pathogen-specific clones are not the only memory clones that can survive PTCy. In subject 001-007, CD8<sup>+</sup> memory clones detectable in the blood, marrow, and GI tract of the donor, who suffered from IBD, were detectable in both the blood and GI tract of the recipient who died of GI GVHD, suggesting that memory T cells reactive with autoantigens or perhaps microbial antigens can be transferred from donor to recipient and survive PTCy. The frequency of such events is unknown, but this observation supports murine data challenging the concept that only naive T cells are capable of causing GVHD (52).

In stark contrast with the dynamics of T cell reconstitution after PTCy, B cell recovery in our patients appears to have occurred nearly entirely by the de novo generation of diverse, naive B cells from the infused BM. Furthermore, there was no discernable impact of CMV on B cell recovery. Although the slightly delayed B cell recovery seen in patients with aGVHD was not statistically significant on multivariate analysis, any delay in B cell reconstitution that these patients may have experienced could be in large part attributable to the immunosuppression, particularly corticosteroids, which these patients received for GVHD treatment. Even so, the impact of B cells on chronic GVHD has recently been recognized (53), and it is therefore possible that any differences in B cell numbers or phenotype between patients with or without aGVHD could contribute to or be a consequence of GVHD itself. Although elevated levels of transitional B cells in patients without aGVHD most likely reflects de novo B cell output from the BM (53), regulatory B cells are enriched in this subset and may play a role in chronic GVHD prevention (54). Further studies would be necessary to assess whether the transitional B cells in our patients have regulatory capacity and whether regulatory B cell numeric or functional deficiencies could explain the occurrence of acute and/or chronic GVHD in patients with robust Treg reconstitution after PTCy (11).

The absence of additional immunosuppression beyond PTCy in our patients not experiencing GVHD did allow us to explore posttransplant lymphocyte reconstitution independent of the effects of CNIs. However, it is unclear whether these findings are specific for patients treated with PTCy or are more general-

izable to all patients undergoing alloBMT. In a study of T cell repertoire recovery after UCBT, patients with aGVHD had more diverse repertoires than patients without aGVHD (18), which is not consistent with our data. However, that study also demonstrated that CMV reactivation had a significant negative impact on T cell repertoire diversity (18). Given our finding that CMV appears to play a more prominent role than aGVHD in T cell repertoire skewing in both PTCy- and CNI-treated patients, CMV also may be a confounder for aGVHD in that umbilical cord blood study. As the dominant effect of CMV in shaping immune reconstitution has now been reported by several groups (23, 24, 48), it is likely that at least some of our findings pertain more broadly to other transplantation platforms.

Analysis of global similarity between the ensembles of *TRB* sequences observed in different recipients, in the same recipient at different times, and between recipients and their related donors highlights both the future promise and current challenges of using antigen receptor sequencing to study the dynamics of complex posttransplant phenomena such as infections and GVHD. Public *TRB* sequences associated with stereotypic responses to herpesvirus antigens were observed in many donors and recipients in this study, and the antigenic specificity associated with several such sequences was directly confirmed by *TRB* sequencing of CD8<sup>+</sup> T cells sorted with CMV pp65-specific tetramers. The observation that one of these CMV pp65-associated public TCRs was present in multiple GI biopsies of subject 002-037, who had previously developed CMV reactivation, demonstrates the potential of antigen receptor sequencing to enable in silico dissection of in vivo immune responses. A distinct *TRB* sequence that our previous studies have associated with a CD8<sup>+</sup> T cell response to the prototypic H-Y antigen KDM5D<sub>311-319</sub>/HLA-A2 comprised 36% of the sequences found in a skin biopsy from this same HLA-A2<sup>+</sup> subject at 1 year after transplant; the recipient was male and received a graft from his HLA-identical sister. Although we could not independently confirm that CD8<sup>+</sup> KDM5D<sub>311-319</sub>/HLA-A2-specific T cells were present in the skin biopsy due to inadequate tissue available for study, other investigators have demonstrated that T cells with this antigenic specificity are commonly present in the skin of HLA-A2<sup>+</sup> male recipients of female grafts (55). This suggests that antigen receptor sequencing may also have utility for dissecting GVHD, similar to its potential in assessing antiviral responses. However, although we observed other *TRB* sequences previously linked to alloantigen-specific or tumor antigen-specific T cell responses, the total number of such sequences was low. It is likely that the TCR repertoire of T cells reactive with a given alloantigen, both within an individual and between different individuals, will be quite diverse and that the subset of *TRB* sequences associated with a specific alloantigen response and shared by 2 or more individuals will be small. Nonetheless, as utilization of high-throughput antigen receptor sequencing to study both the diversity of TCRs reactive with specific antigens and lymphocyte dynamics and trafficking in BMT recipients grows, the number of public TCRs associated with responses to alloantigens, tumor antigens, and microbial antigens will steadily increase. As it does, antigen receptor sequencing could potentially become a powerful adjunct technique for in silico dissection of posttransplant lymphocyte dynamics.

## Methods

**Study design.** This study was designed to assess lymphocyte reconstitution by flow cytometry and high-throughput antigen receptor sequencing among patients treated on a multiinstitutional clinical study (8). The sample size was determined by the number of available specimens collected from patients treated at 2 of the 3 participating institutions (Johns Hopkins Hospital [JHH; Baltimore, Maryland, USA] and FHCRC). As part of the clinical study, PB samples were collected at predefined time points. Since samples were not always drawn on the prescribed day, samples collected at times more or less than 30% beyond the intended draw date were excluded. All patient specimens collected within the defined temporal parameters were included, except for posttransplant specimens in those experiencing graft failure ( $n = 7$ ) or maximal grade I aGVHD not requiring treatment ( $n = 4$ ). The latter group was excluded to provide optimal separation of patients with or without clinically significant aGVHD. Samples were deidentified and analyzed in a blinded fashion.

**Patients and samples.** Eighty patients were treated on a prospective clinical study (NCT00809276) at JHH ( $n = 42$ ) or FHCRC ( $n = 38$ ) with myeloablative once-daily i.v. busulfan (Otsuka Pharmaceutical) and fludarabine conditioning, HLA-matched related or unrelated BM alloBMT, and high-dose PTCy on days +3 and +4 as single-agent GVHD prophylaxis (8). Related donors of BM for these patients also provided PB samples on the day of BM donation. BM from 3 donors had been collected separately on a JHH cell banking protocol. PBMCs and BM mononuclear cells (BMMCs) were isolated using Ficoll-Hypaque density centrifugation and then cryopreserved.

Overall, of the 80 patients treated at these 2 institutions, available PB specimens from 74 patient/donor pairs were included from the following time points: donor ( $n = 26$ ), patient pretransplant ( $n = 70$ ), and patient posttransplant at month 1 ( $n = 59$ ; drawn on median day 31 [range 23–38]), month 2–3 ( $n = 57$ ; drawn on median day 72 [range 55–99] as intended to be day 60 at JHH and day 80 at FHCRC), month 6 ( $n = 27$ ; drawn on median day 187 [range 169–223]), and month 12 ( $n = 34$ ; drawn on median day 367.5 [range 343–450]). All available specimens were assessed by flow cytometry. All 34 patients with available 1-year posttransplant PB specimens had samples assessed by high-throughput antigen receptor sequencing (Table 3 and Supplemental Tables 1 and 2). Eighteen of these 34 patients had received an alloBMT from a related donor, and 15 of the 18 patients had donor PB samples available for sequencing. One additional patient who did not have a 1-year sample (due to nonrelapse mortality at day +64) was included in the sequencing studies due to the availability of donor PB, BM, and GI biopsy specimens. Twenty-six 1-year patient samples and 14 donor samples were assessed by *IGH* sequencing. *TRB* sequencing also was performed on up to 3 additional blood samples acquired between 2 and 5.2 years after transplant from each of 3 patients and on pretransplant blood samples acquired from 4 patients.

In order to provide control samples of patients treated with standard GVHD prophylaxis using methotrexate and a CNI, the FHCRC clinical transplant database was queried for patients who met the following criteria: (i) first alloBMT for a hematologic malignancy between 2001–2013; (ii) transplantation platform of myeloablative conditioning, HLA-matched related or unrelated BM allograft, and GVHD prophylaxis comprising methotrexate and a CNI; and (iii) survival until at least 1 year after transplantation. Among the 152 individuals who met these eligibility criteria, cryopreserved 1-year posttransplant samples previously obtained on a cell banking protocol were available for 11 patients (Supplemental Tables 1 and 3).

*Flow cytometry.* Flow cytometry was performed centrally on all samples using an LSR II (BD Biosciences). The monoclonal antibodies used are included in the Supplemental Methods. Dead cells were excluded using a viability dye (LIVE/DEAD Fixable Aqua Dead Cell Stain; Invitrogen). Gating was performed by an investigator blinded to the patients' clinical outcomes/parameters and independently verified by a second investigator who was blinded to the identity of each sample. For sequencing studies that involved sorted populations, FACS was performed on a FACSAria II (BD Biosciences). Among viable CD3<sup>+</sup>CD4<sup>+</sup> or CD3<sup>+</sup>CD8<sup>+</sup> cells, naive was defined as CCR7<sup>+</sup>CD45RA<sup>+</sup>CD45RO<sup>-</sup>CD95<sup>-</sup>CD27<sup>+</sup>CD28<sup>+</sup>, while memory was defined as CCR7<sup>+</sup>CD45RA<sup>-</sup>CD45RO<sup>+</sup> (central memory, CM), CCR7<sup>-</sup>CD45RA<sup>-</sup>CD45RO<sup>+</sup> (EM), or CCR7<sup>-</sup>CD45RA<sup>+</sup>CD45RO<sup>+</sup> (TEMRA). CMV pp65-specific CD8<sup>+</sup> T cells were sorted from PBMCs in a subset of patients using APC-conjugated NLVPMVATV/HLA-A\*02:01 or PE-conjugated RIPHERNGFTVL/HLA-B\*07:02 tetramers (FHCRC Immune Monitoring Lab).

*Antigen receptor sequencing and sequence analysis.* Sequencing of *TRA*, *TRB*, and *IGH* loci in blood, BM, and biopsies of skin and GI tract was performed exclusively on genomic DNA either at Adaptive Biotechnologies using the immunoSEQ assay at survey level resolution or at the FHCRC using the hsTCRB kit (Adaptive Biotechnologies) (15). Genomic DNA was isolated as described in the Supplemental Methods.

Initial analysis of *TRA*, *TRB*, and *IGH* sequence reads (Supplemental Tables 6–10) was performed at Adaptive Biotechnologies using the immunoSEQ suite of software tools (Adaptive Biotechnologies). Repertoire analyses and longitudinal tracking of sequence frequency over time were performed using a package of software tools (LymphoSeq; available for download from <http://www.bioconductor.org>) created by D.G. Coffey and implemented in the R statistical computing environment. The Shannon-Weaver entropy (56) (defined as:  $Entropy [bits] = -\sum_i p[x_i][\log_2 p(x_i)]$ ) where the sum is over all unique productive sequences, and  $p(x_i)$  is the frequency of sequence  $i$ ), Gini coefficient (26, 57), and clonality (1 – normalized entropy) (25) were calculated for each *TRA*, *TRB*, or *IGH* dataset. Originally developed to represent income distribution, the Lorenz curve also may be used to graphically represent *TRB* clonality, wherein the  $x$  axis represents the cumulative percentage of unique *TRB* sequences and the  $y$  axis represents the cumulative percentage of *TRB* reads (Supplemental Figure 4). A line passing through the origin with a slope of 1 reflects equal frequencies of all clones. The Gini coefficient is the ratio of the area between the line of equality and the observed Lorenz curve over the total area under the line of equality (Supplemental Figure 4). The higher the coefficient is, the more unequal the distribution (i.e., the more clonal the T cell population). The range of both the Gini coefficient and the clonality metric is 0–1, with 0 indicating a sequence distribution with uniform frequency and 1 indicating a monoclonal population. The amount of overlap or Bhattacharyya coefficient (28) between 2 *TRB* repertoires was defined as the sum over all shared sequences of the geometric means of their frequencies in the 2 repertoires (Bhattacharyya coefficient  $[A,B] = \sum_i \sqrt{p(x_i)q(x_i)}$ ), where  $p(x_i)$  = frequency of

sequence  $x_i$  in set A and  $q(x_i)$  = frequency of sequence  $x_i$  in set B. The range of the Bhattacharyya coefficient is also 0–1, with 0 indicating there are no sequences in common between the 2 samples and 1 indicating complete identity between the 2 samples. The raw sequence data have been made publicly available for download from Adaptive Biotechnologies (<http://adaptivebiotech.com/pub/Kanakry-2016-JCIInsight>).

*Assignment of TRB sequences to specific T cell subsets.* Since flow cytometric sorting of T cell subsets could not be consistently performed on all patient samples due to low lymphocyte counts, we inferred the immunophenotype of T cells in unsorted samples by mapping the TRB sequences from flow cytometrically sorted samples (e.g., CD8<sup>+</sup> memory) from the recipient or the recipient's donor onto unsorted samples collected from the same recipient. For example, to track the change in frequency of CD8<sup>+</sup> CMV pp65 tetramer-positive T cells in recipient 002-037, we mapped the TRB sequences of CD8<sup>+</sup> CMV pp65 tetramer-positive T cells sorted from the recipient's blood at one posttransplant time point onto the TRB repertoires of the patient's PBMC samples at other time points. To resolve cases where the same TRB sequence was observed in 2 or more flow cytometrically defined T cell immunophenotypic subsets, such as CD8<sup>+</sup> memory and naive, we assigned the T cell phenotype in the following order: CD8<sup>+</sup>CMV<sup>+</sup>, CD4<sup>+</sup> naive, CD8<sup>+</sup> naive, CD4<sup>+</sup> memory, CD8<sup>+</sup> memory, CD4<sup>+</sup>, CD8<sup>+</sup>.

*Statistics.* Univariate and multivariate linear regression was used to assess the association of clinical factors (recipient CMV serostatus, donor CMV serostatus, detectable PB CMV reactivation, and occurrence of aGVHD) with the flow cytometrically determined lymphocyte subsets. Cell concentrations (cells/ $\mu$ l) were calculated by multiplying the proportion of gated subsets by the ALCs determined by the clinical laboratories on the day of specimen collection. Total cell concentrations were log-transformed prior to analysis; cell concentrations of 0.0 were assigned a value of 0.001. Mean clonality was compared using 2-sample *t* tests. Two-sided *P* values < 0.05 were considered statistically significant, and no adjustment was made for multiple comparisons. Pearson's correlation coefficient was calculated to measure the degree of linear dependence between clonal frequencies of 2 samples. All statistical analyses and data visualization were performed in R version 3.2.2.

*Study approval.* The clinical study and cell banking protocols were reviewed and approved by the Johns Hopkins Medicine IRB (Baltimore, Maryland, USA) and/or the FHCRC/University of Washington Cancer Consortium IRB (Seattle, Washington, USA). All patients and healthy donors provided written informed consent prior to study enrollment and sample collection. GVHD target organ biopsies had been obtained for clinical diagnosis of GVHD at JHH or FHCRC and were used for these studies after IRB approval.

## Author contributions

CGK, LL, PVO, and EHW contributed to the study design. CGK, AMHT, and AV acquired data. CCSY and CDG obtained pathologic specimens for sequencing studies. CGK, DGC, AMHT, AV, JC, LL, and EHW analyzed data. CGK, DGC, AMHT, AV, JC, HSR, LL, and EHW interpreted data. DGC and BES performed statistical analyses. CGK, DGC, AMHT, and EHW designed and created the tables and figures. CGK, DGC, LL, and EHW wrote the manuscript, and all authors edited the manuscript.

## Acknowledgments

This work was supported by awards from the NIH (R01 HL 110907 [to L. Luznik and E.H. Warren] and T32 HL007093 [J. Abkowitz]), the Cancer Therapeutics Endowment at the FHCRC, and a grant from Otsuka Pharmaceutical (to L. Luznik). We thank all patients and donors who participated in this clinical study and contributed PB, BM, or tissue specimens that were analyzed in this work. We thank Mesa Schumacher for assistance with preparation of Figures 7 and 8 and the University of Washington Molecular Virology Laboratory for assessment of EBV viremia in recipient 002-014.

Address correspondence to: Edus H. Warren, 1100 Fairview Ave N, D4-100, P.O. Box 19024, Seattle, Washington 98109-1024, USA. Phone: 206.667.6441; E-mail: ehwarren@u.washington.edu. Or to: Leo Luznik, 1650 Orleans Street, Baltimore, Maryland 21287, USA. Phone: 410.502.7732; E-mail: luznik@jhmi.edu.

Christopher G. Kanakry's present address is: Experimental Transplantation and Immunology Branch, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, Maryland, USA.

1. Kanakry CG, Fuchs EJ, Luznik L. Modern approaches to HLA-haploidentical blood or marrow transplantation. *Nat Rev Clin Oncol*. 2016;13(1):10–24.
2. Luznik L, et al. HLA-haploidentical bone marrow transplantation for hematologic malignancies using nonmyeloablative conditioning and high-dose, posttransplantation cyclophosphamide. *Biol Blood Marrow Transplant*. 2008;14(6):641–650.
3. Luznik L, et al. High-dose cyclophosphamide as single-agent, short-course prophylaxis of graft-versus-host disease. *Blood*. 2010;115(16):3224–3230.
4. Bolanos-Meade J, et al. HLA-haploidentical bone marrow transplantation with posttransplant cyclophosphamide expands the donor pool for patients with sickle cell disease. *Blood*. 2012;120(22):4285–4291.
5. Bashey A, et al. T-cell-replete HLA-haploidentical hematopoietic transplantation for hematologic malignancies using post-transplantation cyclophosphamide results in outcomes equivalent to those of contemporaneous HLA-matched related and unrelated donor transplantation. *J Clin Oncol*. 2013;31(10):1310–1316.
6. Raiola AM, et al. Unmanipulated haploidentical bone marrow transplantation and posttransplantation cyclophosphamide for hematologic malignancies after myeloablative conditioning. *Biol Blood Marrow Transplant*. 2013;19(1):117–122.
7. Raj K, et al. Peripheral blood hematopoietic stem cells for transplantation of hematological diseases from related, haploidentical donors after reduced-intensity conditioning. *Biol Blood Marrow Transplant*. 2014;20(6):890–895.
8. Kanakry CG, 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(31):3497–3505.
9. Kanakry JA, et al. Absence of post-transplantation lymphoproliferative disorder after allogeneic blood or marrow transplantation using post-transplantation cyclophosphamide as graft-versus-host disease prophylaxis. *Biol Blood Marrow Transplant*. 2013;19(10):1514–1517.
10. Tischer J, et al. Virus infection in HLA-haploidentical hematopoietic stem cell transplantation: incidence in the context of immune recovery in two different transplantation settings. *Ann Hematol*. 2015;94(10):1677–1688.
11. Kanakry CG, et al. Aldehyde dehydrogenase expression drives human regulatory T cell resistance to posttransplantation cyclophosphamide. *Sci Transl Med*. 2013;5(211):211ra157.
12. Cieri N, et al. Post-transplantation cyclophosphamide sirolimus after haploidentical hematopoietic stem cell transplantation using a treosulfan-based myeloablative conditioning peripheral blood stem cells. *Biol Blood Marrow Transplant*. 2015;21(8):1506–1514.
13. Cieri N, et al. Generation of human memory stem T cells upon haploidentical T-replete hematopoietic stem cell transplantation. *Blood*. 2015;125(18):2865–2874.
14. Roberto A, et al. Role of naive-derived T memory stem cells in T cell reconstitution following allogeneic transplantation. *Blood*. 2015;125(18):2855–2864.
15. Robins HS, et al. Comprehensive assessment of T-cell receptor  $\beta$ -chain diversity in  $\alpha\beta$  T cells. *Blood*. 2009;114(19):4099–4107.
16. Robins HS, et al. Overlap and effective size of the human CD8<sup>+</sup> T cell receptor repertoire. *Sci Transl Med*. 2010;2(47):47ra64.
17. Warren EH, Matsen FA 4th, Chou J. High-throughput sequencing of B- and T-lymphocyte antigen receptors in hematology. *Blood*. 2013;122(1):19–22.
18. van Heijst JW, et al. Quantitative assessment of T cell repertoire recovery after hematopoietic stem cell transplantation. *Nat Med*. 2013;19(3):372–377.
19. Meyer EH, et al. A distinct evolution of the T-cell repertoire categorizes treatment refractory gastrointestinal acute graft-versus-host disease. *Blood*. 2013;121(24):4955–4962.
20. Kanda J, et al. Immune recovery in adult patients after myeloablative dual umbilical cord blood, matched sibling, and matched unrelated donor hematopoietic cell transplantation. *Biol Blood Marrow Transplant*. 2012;18(11):1664–1676.
21. Muraro PA, et al. T cell repertoire following autologous stem cell transplantation for multiple sclerosis. *J Clin Invest*. 2014;124(3):1168–1172.
22. Ritter J, et al. Donor CD4 T cell diversity determines virus reactivation in patients after HLA-matched allogeneic stem cell transplantation. *Am J Transplant*. 2015;15(8):2170–2179.
23. Itzykson R, Robin M, Moins-Teisserenc H, et al. Cytomegalovirus shapes long-term immune reconstitution after allogeneic stem cell transplantation. *Haematologica*. 2015;100(1):114–123.
24. Lugthart G, et al. Early cytomegalovirus reactivation leaves a specific and dynamic imprint on the reconstituting T cell compartment long-term after hematopoietic stem cell transplantation. *Biol Blood Marrow Transplant*. 2014;20(5):655–661.
25. Sherwood AM, et al. Tumor-infiltrating lymphocytes in colorectal tumors display a diversity of T cell receptor sequences that differ from the T cells in adjacent mucosal tissue. *Cancer Immunol Immunother*. 2013;62(9):1453–1461.
26. Gastwirth JI. Estimation of Lorenz-Curve and Gini-Index. *Rev Econ Stat*. 1972;54(3):306–316.
27. Qi Q, et al. Diversity and clonal selection in the human T-cell repertoire. *Proc Natl Acad Sci U S A*. 2014;111(36):13139–13144.
28. Bhattacharyya A. On a measure of divergence between two statistical populations defined by their probability distributions. *Bull Calcutta Mathematical Soc*. 1943;35:99–109.
29. Robins H, et al. Ultra-sensitive detection of rare T cell clones. *J Immunol Methods*. 2012;375(1–2):14–19.
30. Murugan A, et al. Statistical inference of the generation probability of T-cell receptors from sequence repertoires. *Proc Natl Acad Sci U S A*. 2012;109(40):16161–16166.
31. den Haan JM, et al. The minor histocompatibility antigen HA-1: a diallelic gene with a single amino acid polymorphism. *Science*. 1998;279(5353):1054–1057.
32. Goulmy E, Pool J, van den Elsen PJ. Interindividual conservation of T-cell receptor beta chain variable regions by minor histocompatibility antigen-specific HLA-A\*0201-restricted cytotoxic T-cell clones. *Blood*. 1995;85(9):2478–2481.
33. Tykodi SS, et al. Allogeneic hematopoietic cell transplantation for metastatic renal cell carcinoma after nonmyeloablative conditioning: toxicity, clinical response, and immunological response to minor histocompatibility antigens. *Clin Cancer Res*. 2004;10(23):7799–7811.
34. Dierselhuis MP, et al. Functionally different HA-1 specific T cells use the same TRBV7-9: a snake in the grass for T cell receptor

- transfer studies. *Immunother Insights*. 2012;4:1–7.
35. Spierings E, et al. The minor histocompatibility antigen HA-3 arises from differential proteasome-mediated cleavage of the lymphoid blast crisis (Lbc) oncoprotein. *Blood*. 2003;102(2):621–629.
  36. Brickner AG, et al. The immunogenicity of a new human minor histocompatibility antigen results from differential antigen processing. *J Exp Med*. 2001;193(2):195–206.
  37. Meadows L, et al. The HLA-A\*0201-restricted H-Y antigen contains a posttranslationally modified cysteine that significantly affects T cell recognition. *Immunity*. 1997;6(3):273–281.
  38. Zhao Y, et al. High-affinity TCRs generated by phage display provide CD4<sup>+</sup> T cells with the ability to recognize and kill tumor cell lines. *J Immunol*. 2007;179(9):5845–5854.
  39. Weekes MP, et al. The memory cytotoxic T-lymphocyte (CTL) response to human cytomegalovirus infection contains individual peptide-specific CTL clones that have undergone extensive expansion in vivo. *J Virol*. 1999;73(3):2099–2108.
  40. de Bueger M, et al. Tissue distribution of human minor histocompatibility antigens. Ubiquitous versus restricted tissue distribution indicates heterogeneity among human cytotoxic T lymphocyte-defined non-MHC antigens. *J Immunol*. 1992;149(5):1788–1794.
  41. Kubo K, 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(7):3019–3026.
  42. Argat VP, et al. Dominant selection of an invariant T cell antigen receptor in response to persistent infection by Epstein-Barr virus. *J Exp Med*. 1994;180(6):2335–2340.
  43. Trautmann L, et al. Selection of T cell clones expressing high-affinity public TCRs within Human cytomegalovirus-specific CD8 T cell responses. *J Immunol*. 2005;175(9):6123–6132.
  44. Chidrawar S, et al. Cytomegalovirus-seropositivity has a profound influence on the magnitude of major lymphoid subsets within healthy individuals. *Clin Exp Immunol*. 2009;155(3):423–432.
  45. Pourghesari B, et al. The cytomegalovirus-specific CD4<sup>+</sup> T-cell response expands with age and markedly alters the CD4<sup>+</sup> T-cell repertoire. *J Virol*. 2007;81(14):7759–7765.
  46. Khan N, et al. Cytomegalovirus seropositivity drives the CD8 T cell repertoire toward greater clonality in healthy elderly individuals. *J Immunol*. 2002;169(4):1984–1992.
  47. Brodin P, et al. Variation in the human immune system is largely driven by non-heritable influences. *Cell*. 2015;160(1–2):37–47.
  48. Suessmuth Y, et al. CMV reactivation drives post-transplant T cell reconstitution results in defects in the underlying TCR $\beta$  repertoire. *Blood*. 2015.
  49. Silins SL, et al. A functional link for major TCR expansions in healthy adults caused by persistent Epstein-Barr virus infection. *J Clin Invest*. 1998;102(8):1551–1558.
  50. Price DA, et al. Avidity for antigen shapes clonal dominance in CD8<sup>+</sup> T cell populations specific for persistent DNA viruses. *J Exp Med*. 2005;202(10):1349–1361.
  51. Nikolich-Zugich J. Ageing and life-long maintenance of T-cell subsets in the face of latent persistent infections. *Nat Rev Immunol*. 2008;8(7):512–522.
  52. Zheng H, Matte-Martone C, Jain D, et al. Central memory CD8<sup>+</sup> T cells induce graft-versus-host disease and mediate graft-versus-leukemia. *J Immunol*. 2009;182(10):5938–5948.
  53. Sarantopoulos S, Ritz J. Aberrant B cell homeostasis in chronic GVHD. *Blood*. 2015;125(11):1703–1707.
  54. Khoder A, et al. Regulatory B cells are enriched within the IgM memory and transitional subsets in healthy donors but are deficient in chronic GVHD. *Blood*. 2014;124(13):2034–2045.
  55. Kim YH, et al. In situ detection of HY-specific T cells in acute graft-versus-host disease-affected male skin after sex-mismatched stem cell transplantation. *Biol Blood Marrow Transplant*. 2012;18(3):381–387.
  56. Ekström M. Quantifying spatial patterns of landscapes. *Ambio*. 2003;32(8):573–576.
  57. Gini C. Sulla misura della concentrazione e della variabilità dei caratteri. *Atti R Ist Veneto Sci Lett Arti*. 1914;73:1203–1248.