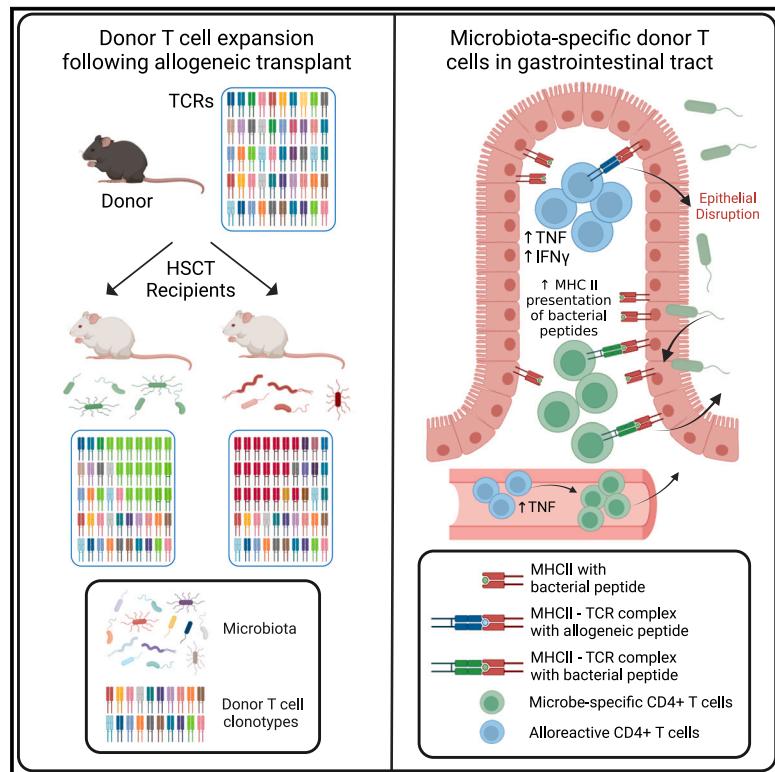


# Immunity

## Microbiota dictate T cell clonal selection to augment graft-versus-host disease after stem cell transplantation

### Graphical abstract



### Authors

Albert C. Yeh, Motoko Koyama,  
Olivia G. Waltner, ..., Charles O. Elson,  
Philip H. Bradley, Geoffrey R. Hill

### Correspondence

ayeh@fredhutch.org (A.C.Y.),  
grhill@fredhutch.org (G.R.H.)

### In brief

Allogeneic T cell expansion is the primary determinant of graft-versus-host disease (GVHD), and current dogma dictates that this is driven by histocompatibility antigen disparities between donor and recipient. Yeh et al. found that a subset of donor CD4<sup>+</sup> T cell clonal expansion was microbiota dependent and that microbiota-specific T cells augmented GVHD lethality independent of donor-recipient genetics.

### Highlights

- The recipient microbiota dictates clonal selection of donor T cells post HSCT
- Microbiota-specific donor T cells augment GVHD in the presence of cognate antigen
- Allogeneic T cells promote activation and TNF response in microbiota-specific T cells
- Allogeneic T cells facilitate microbial antigen presentation by recipient GI epithelium

Article

# Microbiota dictate T cell clonal selection to augment graft-versus-host disease after stem cell transplantation

Albert C. Yeh,<sup>1,2,\*</sup> Motoko Koyama,<sup>1</sup> Olivia G. Waltner,<sup>1</sup> Simone A. Minnie,<sup>1</sup> Julie R. Boiko,<sup>1</sup> Tamer B. Shabaneh,<sup>1</sup> Shuichiro Takahashi,<sup>1</sup> Ping Zhang,<sup>1</sup> Kathleen S. Ensley,<sup>1</sup> Christine R. Schmidt,<sup>1</sup> Samuel R.W. Legg,<sup>1</sup> Tomoko Sekiguchi,<sup>1</sup> Ethan Nelson,<sup>1</sup> Shruti S. Bhise,<sup>1</sup> Andrew R. Stevens,<sup>1</sup> Tracy Goodpaster,<sup>3</sup> Saranya Chakka,<sup>1</sup> Scott N. Furlan,<sup>1</sup> Kate A. Markey,<sup>1,2</sup> Marie E. Bleakley,<sup>1,4</sup> Charles O. Elson,<sup>5</sup> Philip H. Bradley,<sup>6,7</sup> and Geoffrey R. Hill<sup>1,2,8,\*</sup>

<sup>1</sup>Translational Science and Therapeutics Division, Fred Hutchinson Cancer Center, Seattle, WA, USA

<sup>2</sup>Division of Medical Oncology, Department of Medicine, University of Washington, Seattle, WA, USA

<sup>3</sup>Experimental Histopathology Core, Fred Hutchinson Cancer Center, Seattle, WA, USA

<sup>4</sup>Division of Hematology, Oncology, and Bone Marrow Transplantation, Department of Pediatrics, University of Washington, Seattle, WA, USA

<sup>5</sup>Department of Medicine, The University of Alabama at Birmingham, Birmingham, AL, USA

<sup>6</sup>Basic Sciences Division, Fred Hutchinson Cancer Center, Seattle, WA, USA

<sup>7</sup>Department of Biochemistry and Institute for Protein Design, University of Washington, Seattle, WA, USA

<sup>8</sup>Lead contact

\*Correspondence: [a.yeh@fredhutch.org](mailto:a.yeh@fredhutch.org) (A.C.Y.), [g.hill@fredhutch.org](mailto:g.hill@fredhutch.org) (G.R.H.)

<https://doi.org/10.1016/j.jimmuni.2024.05.018>

## SUMMARY

Allogeneic T cell expansion is the primary determinant of graft-versus-host disease (GVHD), and current dogma dictates that this is driven by histocompatibility antigen disparities between donor and recipient. This paradigm represents a closed genetic system within which donor T cells interact with peptide-major histocompatibility complexes (MHCs), though clonal interrogation remains challenging due to the sparseness of the T cell repertoire. We developed a Bayesian model using donor and recipient T cell receptor (TCR) frequencies in murine stem cell transplant systems to define limited common expansion of T cell clones across genetically identical donor-recipient pairs. A subset of donor CD4<sup>+</sup> T cell clonotypes differentially expanded in identical recipients and were microbiota dependent. Microbiota-specific T cells augmented GVHD lethality and could target microbial antigens presented by gastrointestinal epithelium during an alloreactive response. The microbiota serves as a source of cognate antigens that contribute to clonotypic T cell expansion and the induction of GVHD independent of donor-recipient genetics.

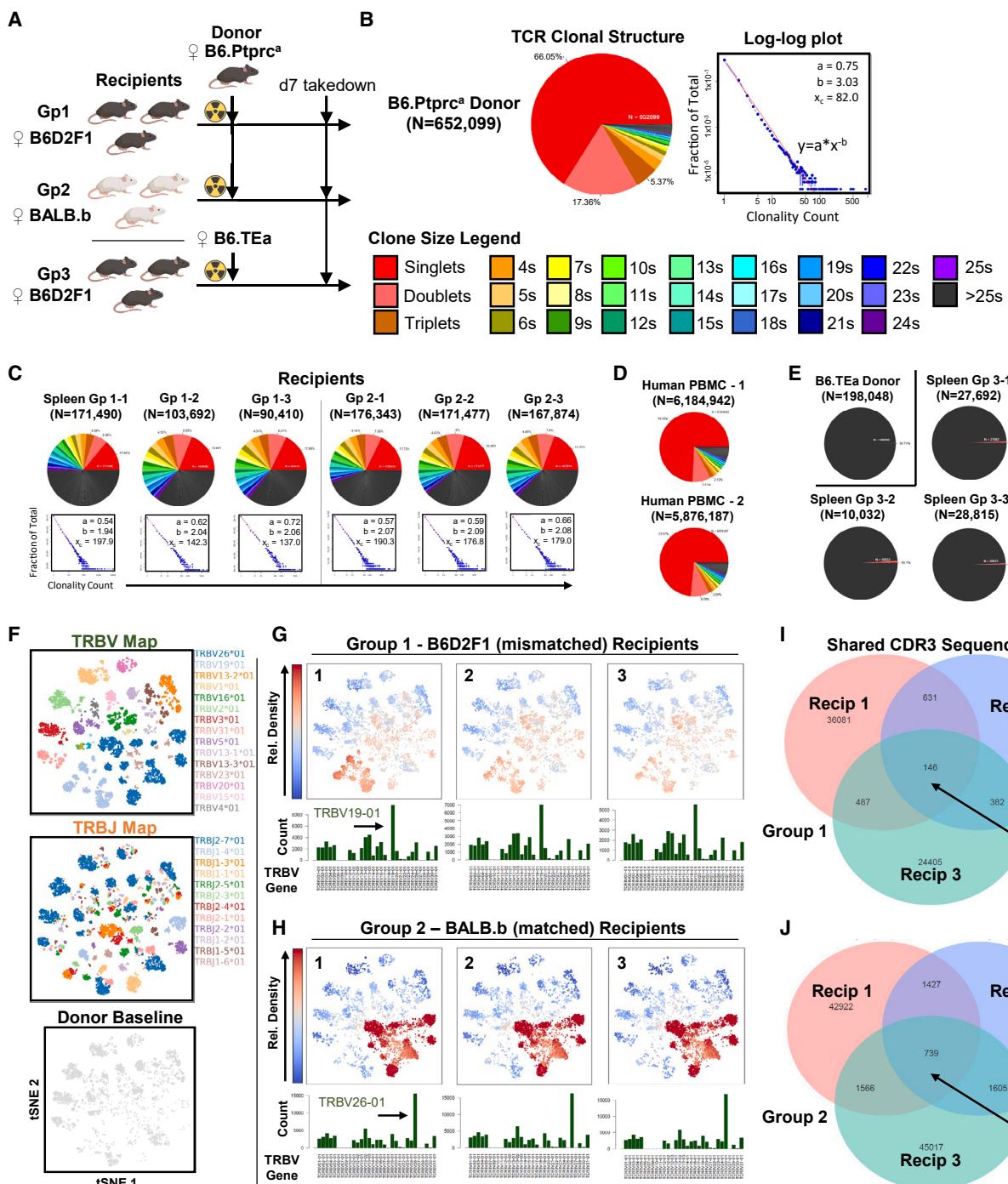
## INTRODUCTION

Graft-versus-host disease (GVHD) remains a major barrier to the successful application of hematopoietic stem cell transplant (HSCT) for a wide range of benign conditions and hematologic malignancies. The immunologic basis of acute GVHD fundamentally involves alloreactive donor T cells that recognize foreign major histocompatibility complex (MHC)-peptide structures derived from both major and minor antigen mismatches with the host.<sup>1,2</sup> Within this genetic paradigm, the relationship between the donor and recipient represents a fixed system that dictates the potential alloreactivity of a given T cell receptor (TCR), raising the question of whether there are predictable aspects of T cell reconstitution associated with GVHD at a clonal level. Exploring the assumptions of this genetic paradigm requires overcoming several challenges, including the vast combinatorial diversity of  $\alpha\beta$  TCRs,<sup>3</sup> the high prevalence of low-copy-number TCRs<sup>4</sup> that appear to characterize alloreactive clones,<sup>5</sup> and sampling constraints—all of which render tracking and

comparing TCR expansion between the donor and host difficult.<sup>6</sup>

Application of high-throughput next-generation TCR amplicon sequencing has enabled insight into the complexity of the reconstituting TCR repertoire in the post-transplant setting.<sup>5–12</sup> In particular, studies evaluating the reproducibility of TCR clonotype responses following stem cell transplant within genetically defined donor-recipient systems in both human<sup>5,11</sup> and murine models<sup>6</sup> suggest highly individualized repertoires for any given donor-recipient pair with minimal overlap across transplant pairs, giving the appearance of a stochastic process characterizing alloreactive T cell expansion.<sup>6</sup> This finding is not surprising given the diversity and sparseness that characterize any given donor TCR repertoire.

We used established murine transplant systems and modeled the probable expansion rates of all identifiable donor TCRs in each system by applying a Bayesian approach incorporating both donor and recipient T cell clonotype frequencies. This strategy enabled the identification of “selective expanders,” which



**Figure 1. A CD4<sup>+</sup>-dependent GVHD system predictably models T cell clonal structure across matched and mismatched transplants**

(A) B6.Ptprc<sup>a</sup> (CD45.1<sup>+</sup>, H-2<sup>b</sup>, I-A<sup>b</sup>, I-E<sup>Nul</sup>) mice were used as donor grafts for three B6D2F1 (F1) (group 1, CD45.2<sup>+</sup>, H-2<sup>b/d</sup>, I-A<sup>b/d</sup>, I-E<sup>d</sup>) and three BALB.b (group 2, H-2<sup>b</sup>, I-A<sup>b</sup>, I-E<sup>Nul</sup>) littermates. As control, B6.TEa (CD45.1<sup>+</sup>, Rag1<sup>-/-</sup>, H-2<sup>b</sup>, I-A<sup>b</sup>, I-E<sup>Nul</sup>) mice were used as donor grafts for three F1 recipients.  $2 \times 10^6$ ,  $5 \times 10^6$ , and  $2 \times 10^6$  CD4<sup>+</sup> magnetic-activated cell sorting (MACS)-sorted T cells were injected into each mouse in groups 1, 2, and 3, respectively, following TBI conditioning at day minus 1. Recipient spleens from each group were collected at day 7 post-transplant for TCR<sup>β</sup> amplicon sequencing.

(B) TCR<sup>β</sup> clonal structure derived from sequencing the same donor pool used for groups 1 and 2 with 652,099 donor templates modeled by an exponential decay function between the TCR clone size (x axis) and clone frequency (y axis). These log-log plots exhibit a slope representing the transformed rate of decay.

(C) TCR clonal structure of sequenced donor CD4<sup>+</sup> T cells from both F1 and BALB.b recipients. Recipients showed a slower rate of decay, suggesting a more highly expanded repertoire compared with the donor pool. Data from 1 experiment.

(D) Clonal structure of TCR<sup>β</sup> seq results from two healthy human peripheral blood samples. Publicly available human TCR<sup>β</sup> datasets were used (see [method details](#)). Approximately 74% of TCRs from each patient were singlets.

(legend continued on next page)

represented T cell clones that differentially expand in genetically identical recipients. We demonstrated that alterations in microbiota composition contributed to differential expansion of T cells following HSCT and observed the selective expansion of donor T cells specific to microbial antigen proportionate to the burden of cognate antigen in the recipient. Furthermore, microbiota-specific T cells were capable of exacerbating GVHD and compromising transplant outcomes. Their pathogenicity was enhanced by the presence of an alloreactive T cell response through tumor necrosis factor (TNF) signaling and the augmentation of antigen presenting capacity by recipient gut epithelium. These results refine our current understanding of T cell expansion following stem cell transplant and suggest that, for a given transplant system, the potential expansion of donor T cells is not solely dictated by major and minor histocompatibility mismatch but also by the recipient microbiota, which effectively act as a source of minor antigen.

## RESULTS

### A CD4<sup>+</sup>-dependent GVHD system predictably models T cell clonal structure across matched and mismatched transplants

As proof of concept, we utilized a transplant platform in which GVHD is dependent on CD4<sup>+</sup> donor T cells because of the dominant CD4<sup>+</sup> response in driving GVHD pathology in MHC disparate systems.<sup>13</sup> We purified CD4<sup>+</sup> splenic T cells from B6.Ptprc<sup>a</sup> (CD45.1<sup>+</sup>) donors as a graft for three B6D2F1 (F1) recipients (CD45.2<sup>+</sup>) and three BALB.b (CD45.2<sup>+</sup>) mice, representing mismatched and matched transplant systems, respectively. As control, we utilized a B6.TEa (Rag1<sup>-/-</sup>, CD45.1<sup>+</sup>) to F1 transplant system where the donor T cell pool consists of a single clone that recognizes the E<sub>x</sub> peptide bound to I-A<sup>b</sup>.<sup>14</sup> Recipient spleens from each group were collected at day 7 post-transplant (Figure 1A), when T cell expansion is approximately at the peak of the initial growth phase,<sup>15</sup> and donor CD4<sup>+</sup> T cells (CD45.1<sup>+</sup>CD45.2<sup>neg</sup>) were sorted for TCR sequencing.

Consistent with prior reports, the clonal structure of both donor and recipient T cell pools followed an exponential decay<sup>17</sup> (Figures 1B and 1C). Notably, out of the 652,099 donor TCRs sequenced, approximately 66% were “singlets” (Figure 1B). The frequency of singlets was similar to that seen in TCR sequences derived from healthy human samples (Figure 1D). In contrast to the donor, both F1 (group 1) and BALB.b (group 2) recipients exhibited clonal structures with a slower rate of clone size decay (Figure 1C), with less than 20% singlets. Sort purity prior to sequencing was over 99.7% for the donor pool and 99.1% for the recipient pools (Figure 1E). We thus confirm in our transplant system that donor TCR repertoire behavior follows a predictable expansion structure despite the sparseness of individual clonotypes.

### MHC matching predictably influences TRBV and TRBJ gene usage post-transplant, despite minimal overlap of the CDR3 region

We next assessed the similarity of alloexpanded T cells among both MHC matched and mismatched transplants. Overall T cell receptor beta variable (TRBV) and T cell receptor beta joining (TRBJ) gene usage was similar among replicates of each recipient group, consistent with prior observations that MHC genotype can predictably skew TCR VDJ gene usage<sup>18</sup> (Figures 1F–1H). Compared with the V- and J-gene usage in the donor pool (Figure 1F), F1 (mismatched) recipients had a higher usage of TCRBV19-01, while BALB.b (matched) recipients with TCRBV26-01 (Figures 1G and 1H). However, CDR3 nucleotide sequence overlap was relatively small, with only 0.16% of unique TCRs shared among all three F1 mice and <2% shared among any given pair (Figure 1I). A small overlapping fraction was seen among BALB.b recipients as well, with only 0.52% unique TCRs shared among the triplicates and <3% shared among any given pair (Figure 1J).

Among TCR clonotypes that were shared in all three recipient triplicates in a group, TCR frequencies varied and were occasionally heavily skewed toward one or more recipients, as illustrated by examining the 146 unique, overlapping TCRs from group 1 (Figure 2A). Of particular interest, we found that a small fraction of clones have widely disparate frequencies among identical recipients. In one instance, a single TCR clonotype was found in 9,739 copies in a group 1 recipient, 129 copies in a second recipient, and 0 copies in the third recipient (Figure 2B). Such behavior could be seen in both the mismatched and matched transplant systems (Figure 2C). However, given the known sparseness of the majority of donor clonotypes (Figure 1B), the TCR count discrepancy seen across identical recipients could be a reflection of donor partitioning stochasticity as opposed to differential expansion within each recipient clone (Figure 2D). We next sought to better understand the nature of these differentially expanded clonotypes by developing a probabilistic model to account for partitioning stochasticity that can heavily influence the distribution of low-copy-number TCRs.

### Probabilistic simulation of post-transplant T cell expansion identifies selective expanders among genetically identical recipients

We set out to test the hypothesis that, if each genetically identical mouse receives a statistically equivalent quantity of a particular T cell clonotype, then clonal expansion should be similar among each recipient. We constructed a series of “twin-transplant” systems, the basic analytic unit consisting of one donor pool and two co-housed recipient littermates of the same strain. Instead of applying traditional beta-diversity metrics to assess TCR clonotype similarity between recipient pairs, we utilized a Bayesian approach that incorporated input clone frequencies data from the donor pool to infer the probability of clonotypes experimentally

(E) TCR sequencing of B6.TEa donor mice and F1 recipients. Recovery of rearranged TCR was over 99.7% in the donor and 99.1% in all recipients.

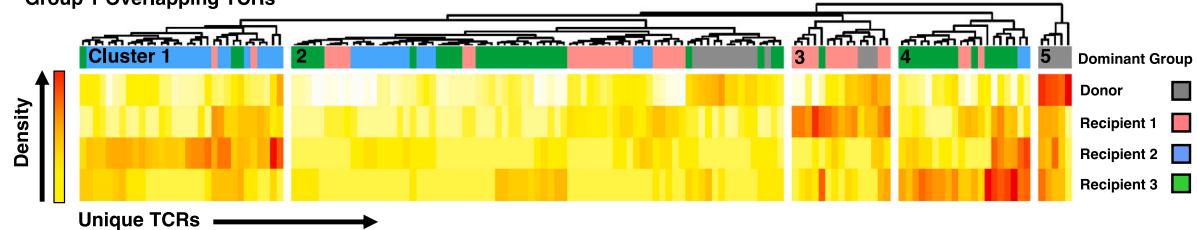
(F) Overview of TRBV and TRBJ gene usage for all samples and tSNE plot of B6.Ptprc<sup>a</sup> donor sample using TCRdist.<sup>16</sup>

(G) TRBV gene usage for all B6D2F1 recipients shows skewing toward TRBV19-01. Color map is based on relative density to donor pool.

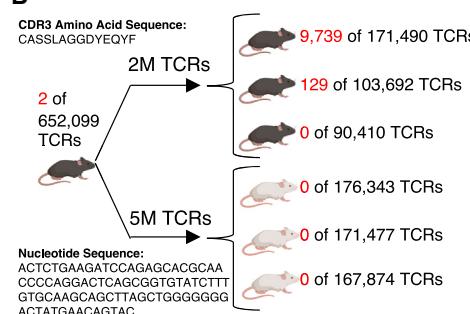
(H) TRBV gene usage for all BALB.b recipients shows skewing toward TRBV26-01.

(I and J) Overlap of unique CDR3 sequences among group 1 (F1) and group 2 (BALB.b) recipients show that 0.16% and 0.52% of TCR clonotypes are present in all recipients within each group, respectively.

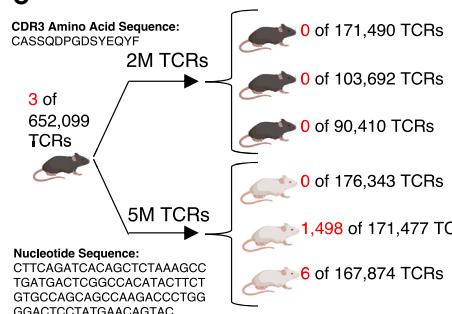
**A Group 1 Overlapping TCRs**



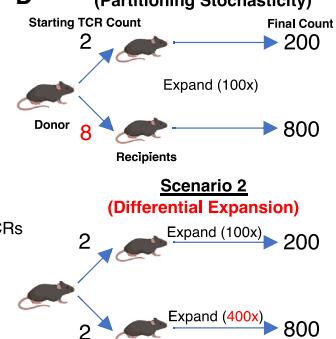
**B**



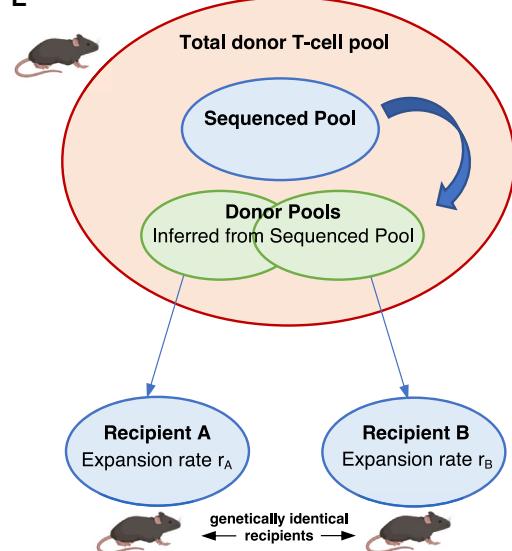
**C**



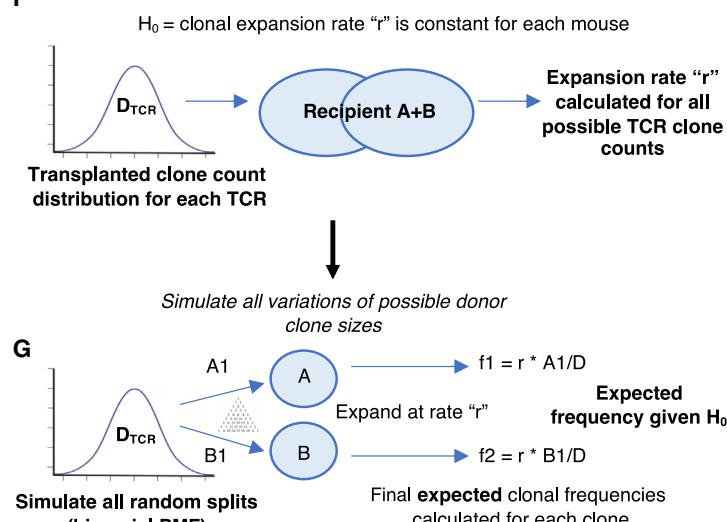
**D** Scenario 1 (Partitioning Stochasticity)



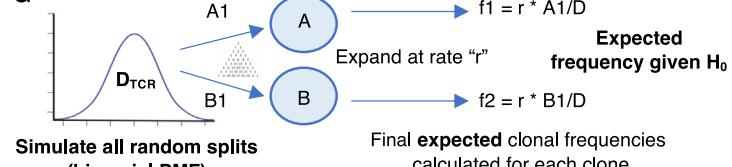
**E**



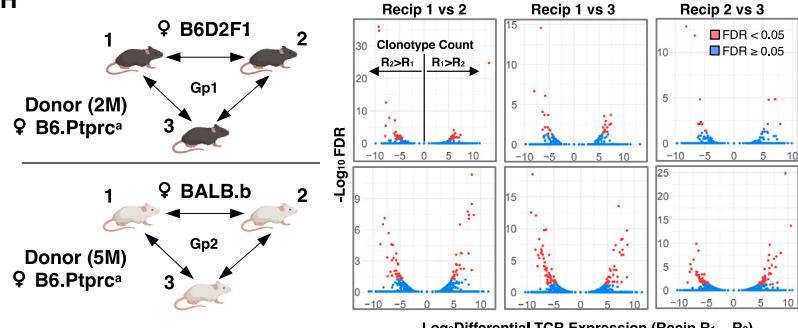
**F**



**G**



**H**



**I**

AA Sequence	Donor	R1	R2	P-value	FDR
CASGDTOQAPLF	0	1	607	2.08E-40	8.99E-37
CASSDGQAGASGYTF	0	1	581	8.42E-39	1.82E-35
CASSLAGGDYEQYF	2	9739	129	1.04E-28	1.50E-25
CASTVLGRHEQYF	0	1	216	2.17E-16	1.34E-13
CASSRDWQIEQYF	0	1	136	1.51E-11	1.31E-08
CASSQDWGQEYOF	26	0	61	1.05E-10	7.54E-08
CTCSGDTNERLFF	4	1	226	6.13E-09	3.78E-06
CASSLDQGNTLYF	34	66	0	1.45E-07	7.80E-05
CASSLEGNTVEFF	16	3	58	5.98E-07	2.87E-04
CASSLGQGANTGQLYF	0	3	71	1.90E-06	7.78E-04
CASSGDQDWGQEYOF	12	83	0	1.98E-06	7.78E-04
CASSDRNTEVFF	12	76	0	2.35E-06	8.46E-04
CASSLDSSYEQYF	11	171	0	5.98E-06	1.99E-03
CASSLDYEQYF	0	1	43	1.01E-05	3.11E-03
CASSLDYEQYF	12	56	0	1.12E-05	3.24E-03
CASSPONGSGNTLYF	0	4	62	1.40E-05	3.79E-03
CASSQQGQANTGQLYF	0	67	1	1.62E-05	3.94E-03
CASSQDWGQEYOF	15	0	30	1.64E-05	3.94E-03
CASGDAGEYQF	10	0	241	2.37E-05	5.39E-03
CASGDWGGAEYOF	0	72	2	3.30E-05	7.13E-03

(legend on next page)

transferred to each recipient mouse (Figure 2E). We then calculated the rate of expansion for each clonotype and modeled the distribution of the donor T cells by simulating all possible partitioning sizes for each clonotype (Figures 2F and 2G), which enabled identification of differentially expanded clonotypes between recipient pairs (see Figures S1A–S1E and *method details*). As a control, we performed syngeneic transplants using 3 female B6 mice and 1 male B6 mouse as recipients, with the addition of a fixed quantity of spiked-in donor CD4<sup>+</sup> T cells targeting the H-Y male antigen (Marilyn T cells) to demonstrate differential expansion of the Marilyn clonotype only between male and female recipient pairs (Figures S1F–S1H).

We next applied our model to identify the most differentially expanded TCRs within all twin-transplant pairs for both matched and mismatched systems from Figure 1A (Figure 2H). We defined selective expanders, which represent differentially expressed TCRs among genetically identical recipients, as TCRs that met a false discovery rate (FDR) cutoff of  $q < 0.05$  (Figure 2I; Table S1). In both matched and mismatched transplant systems, we saw a small fraction of selective expanders comprising ~2%–6% of the top 100 most expanded clonotypes for each twin transplant (Table S1). To demonstrate that selective expanders were unlikely to arise simply by random sampling variation, we simulated twin-transplant systems, imputing donor TCR clonality structures established from experimental observations (Figures S1I–S1L).

### Ablation of the microbiota inhibits the generation of selective expanders

The presence of selective expanders (Figure 3A) among the identical recipients suggested that MHC-independent factors influenced clonal expansion. It is well known that conditioning regimen intensity as well as the regulation of gastrointestinal (GI) microbiota can influence the presence and severity of GVHD via alterations in systemic and local cytokine composition.<sup>19,20</sup> We transplanted 12 female F1 mice with B6.Ptprc<sup>a</sup> CD4<sup>+</sup> T cells using four separate conditions: 900 versus 1,300 cGy total body irradiation (TBI) conditioning and peri-transplant

versus no antibiotic treatment (Figure 3B). We injected a fixed quantity of transgenic CD90.1<sup>+</sup> Marilyn CD4<sup>+</sup> T cells, which are not alloantigen-specific in this all-female system, into each recipient the day prior to organ harvest (Figure S2A). Donor T cells isolated from the spleen of each recipient were collected for TCR sequencing (Figures S2B–S2F). We also harvested small intestine lamina propria (SILP) T cells, though total productive donor TCR templates were lower in the 900 cGy group (range  $n = 102$ –4,611) compared with the 1,300 cGy group (range  $n = 119,738$ –183,470), limiting our analyses of SILP TCRs to the 1,300 cGy group (Figures S2G–S2J).

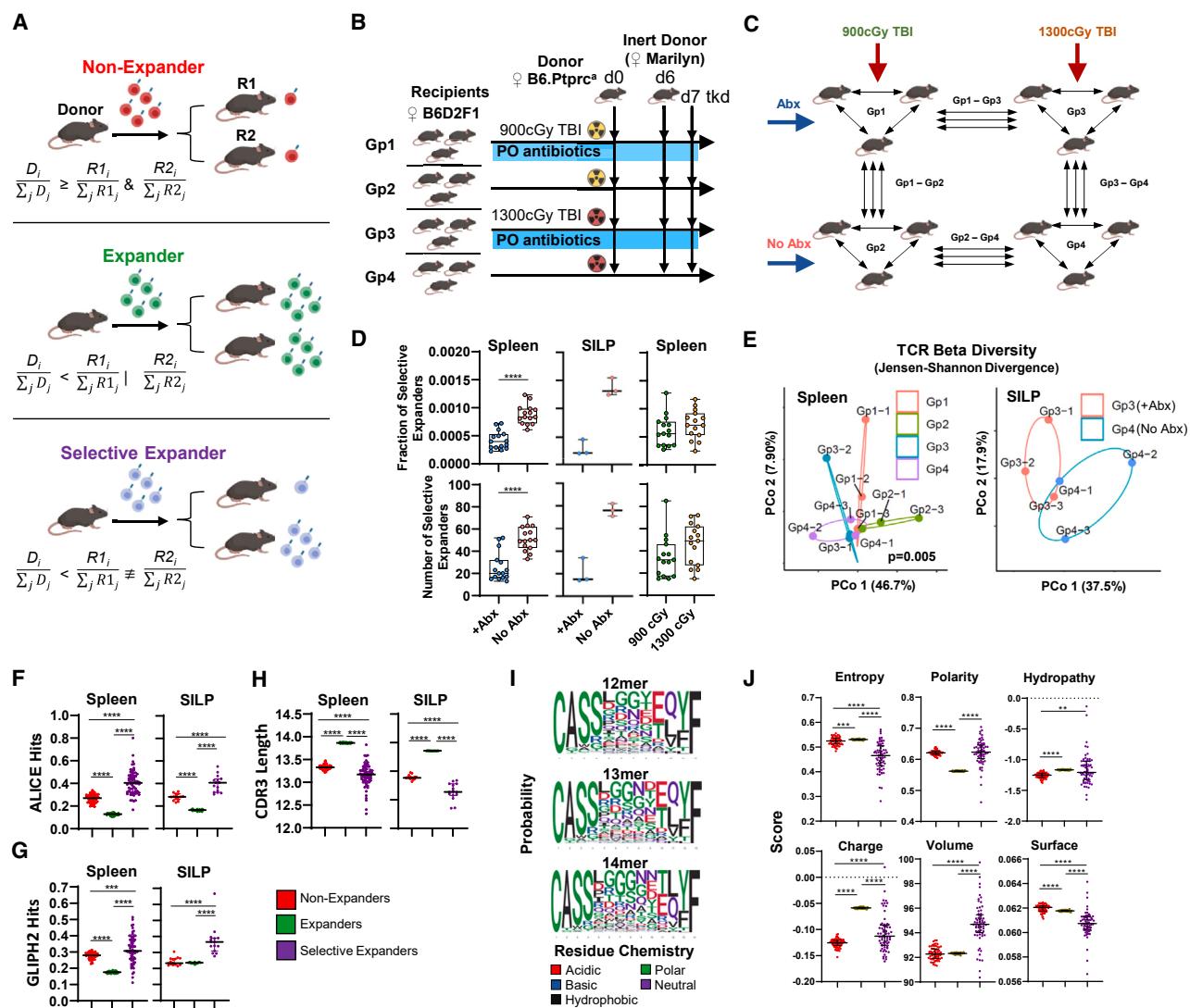
We examined the fraction of selective expanders identified between all recipient pairs with respect to antibiotic versus no antibiotic treatment as well as 900 versus 1,300 cGy conditioning (Figure 3C). Treatment with antibiotics diminished the fraction of selective expanders in both the spleen ( $p < 0.0001$ ) and SILP (Figures 3D and S3A–S3D; Table S2), though augmenting TBI conditioning intensity from 900 to 1,300 cGy, which increases systemic inflammation and lipopolysaccharide exposure,<sup>19</sup> did not result in an increase in selective expanders (Figures 3D, S3E, and S3F; Table S2). A repeat experiment focused on peri-transplant antibiotic treatment at 1,300 cGy (Figures S3G–S3I) confirmed a reduction of selective expanders (Figures S3J–S3L; Table S2). These results held after accounting for background sequencing noise (Figures S4A and S4B) and remained consistent when recipient TCR pool sizes were down-sampled to that of the lowest individual recipient count to account for variations in sampling size, an adjustment applied to subsequent analyses (Figures S4C–S4E). Of note, beta-diversity metrics, which does not consider donor sequence information, suggested different splenic TCR repertoires across groups<sup>21</sup> (Figure 3E).

Selective expanders were also identified in syngeneic transplant systems. When we treated B6 recipients with peri-transplant antibiotics (Figures S5A–S5C), selective expanders were found at a higher frequency in the non-antibiotic group, as the antibiotic treatment significantly reduced the number of recoverable donor TCRs in the recipient (median count 19,501 versus 99,431,  $p < 0.01$ ; Figures S5D–S5G; Table S2). Taken together,

**Figure 2. A probabilistic model of post-transplant TCR expansion using a twin-transplant system identifies differentially expanded clonotypes across genetically identical donor-recipient pairs**

- (A) TCR clonotypes represented in the donor pool and all three F1 recipients from Figure 1A showed clusters of TCRs with enriched densities in recipient subsets (gray, donor; red, recipient 1; blue, recipient 2; green, recipient 3). Hierarchical clustering of normalized cell counts showed TCRs enriched in each subgroup (donor: cluster 5; recipient 1: cluster 3; recipient 2: cluster 1; recipient 3: cluster 4).
- (B) Sample TCR clonotype present in F1 recipients but not in BALB.b recipients. Within the F1 recipients this clonotype comprised 5,679, 124, and 0 templates per 100,000 TCRs in respective recipients.
- (C) Sample TCR clonotype present in BALB.b recipients but not in F1 recipients. Within the BALB.b recipients, this clonotype comprised 0, 874, and 3.6 templates per 100,000 TCRs in respective recipients.
- (D) Schematic illustrating the effects of partitioning stochasticity (scenario 1) from donor T cell variation versus differential expansion (scenario 2) driving disparate TCR frequencies seen between identical recipients.
- (E) Schematic of a twin-transplant system with a single donor pool and two genetically identical co-housed recipients. Count size probabilities for all donated TCR clonotypes are inferred by deep sequencing of the donor T cell pool.
- (F) Expansion rate calculated for each TCR clonotype in the system, assuming equal expansion (null hypothesis) between the two recipient mice.
- (G) The probability distribution of all final expected frequencies found in each recipient is calculated for each TCR clonotype after simulation of possible donor count splits.
- (H) Analytic schema for all twin-transplant systems in F1 and BALB.b recipients. All combinations of recipients in each group were analyzed for differentially expanded TCR clonotypes. Clonotypes with an FDR < 0.05 are denoted in red (y axis). Differences in TCR expression between each recipient pair are shown (x axis). Data from 1 experiment.
- (I)  $p$  values and sequenced TCR frequencies are shown for the top 20 differentially expressed TCRs (selective expanders) between a recipient pair (group 1-1 and group 1-2) as referenced from (H). Clonotype counts denoted for donor, recipient 1 (R1), and recipient 2 (R2) (Table S1 for all pairs).

See also Figure S1.



**Figure 3. Antibiotic treatment depletes selectively expanded donor T cells**

(A) Classification of donor TCR clonotype behavior for a twin-transplant system. Non-expanders (both recipient TCR fractions  $\leq$  donor fraction), expander (either recipient TCR fraction  $>$  donor fraction), and selective expanders (differential expansion of donor TCRs between recipient pair).

(B) B6.Ptprc<sup>a</sup> mice were used as donor grafts for 12 female F1 recipients co-housed into 4 groups. Two groups were pre-treated with oral antibiotics from day minus 14 to day 7 and two groups were treated with low-dose (900 cGy TBI on day minus 1) instead of high-dose (1,300 cGy TBI) conditioning. At day 0, each recipient received  $2 \times 10^6$  CD4<sup>+</sup> donor T cells (CD45.1<sup>+</sup>) and  $5 \times 10^6$  bone marrow cells (B6, CD45.2<sup>+</sup>). At day 6, all recipients received an injection of  $5 \times 10^5$  non-alloreactive Marilyn CD4<sup>+</sup> T cells. Spleen and SILP were harvested at day 7 for donor T cell sequencing.

(C) Analysis schema representing individual twin-transplant systems (black arrows) for possible recipient pairs.

(D) Both the number and fraction of selective expanders (out of unique recipient TCR clonotypes) identified among twin-transplant systems showed that mice treated with peri-transplant antibiotics harbor significantly fewer selective expanders. Each point represents a result from a twin-transplant system. This effect was seen in both the spleen and SILP compartments. Conditioning intensity differences did not alter the fraction of selective expanders across twin-transplant systems. Data from 1 experiment.

(E) Beta-diversity metrics were calculated using the Jensen-Shannon divergence<sup>21</sup> between TCR repertoire pairs (spleen n = 57,983 random TCRs per sample, PERMANOVA<sup>22</sup> p = 0.005; SILP n = 119,738, p = 0.1) and illustrated via principal coordinate analysis (PCoA).

(F) Identification of CDR3 clonotypes using ALICE showed that selective expanders are significantly enriched in TCRs that are more likely to respond to specific immune stimuli compared with non-expanders and expanders in both splenic and SILP compartments.

(G) Clustering of TCRs that appear in specificity groups using GLIPH2 showed a similar enrichment in selective expanders compared with non-expanders and expanders.

(H) Selective expanders have a significantly shorter CDR3 length compared with non-expanders and expanders.

(I) 12–14 amino acid length CDR3 regions represented for all TCRs sequenced across the 12 F1 recipient spleen samples.

(J) The first and last 4 amino acid residues were truncated to calculate the average entropy, polarity, hydrophobicity, charge, volume, and surface propensity for the remaining amino acid residues. Statistical analyses by Mann-Whitney U test (D), Wilcoxon matched-pairs signed-rank test (F–H, J). \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001, \*\*\*\*p < 0.0001.

See also Figures S2–S4.

these data indicate that reduction in microbial load results in a more homogeneous and compact post-transplant TCR repertoire with fewer selective expanders in both the gastrointestinal (GI) and systemic compartments.

### Sequence analysis predicts the generation of selective expanders in response to antigenic stimuli

The lack of change in selective expander frequencies seen with the increase in total body irradiation raised the hypothesis that the difference detected when sterilizing the microbiome may not depend solely on systemic inflammation but also on differences in exposure to microbial antigenic stimuli.<sup>23</sup> To test the hypothesis that selective expanders are generated via an antigenic response, we utilized two TCR clustering algorithms developed for this purpose, antigen-specific lymphocyte identification by clustering of expanded sequences (ALICE),<sup>24</sup> and grouping of lymphocyte interactions by paratope hotspots (GLIPH2).<sup>25</sup> We compared non-expanders, expanders, and selective expanders (Figure 3A) within all recipient pairs in our F1 system using spleen ( $n = 66$ ) or lamina propria ( $n = 15$ ) outputs with respect to the fraction of ALICE “hits,” with a higher fraction suggesting a larger proportion of antigen-specific T cells. Selective expanders in both spleen and GI compartments had significantly more ALICE hits compared with both non-expanders ( $p < 0.0001$ ) and expanders ( $p < 0.0001$ ) (Figure 3F; Data S1). We also compared the same populations with respect to fraction of TCRs that appear in specificity groups identified using GLIPH2 and demonstrated a similar enrichment in selective expanders (Figure 3G; Data S2).

As shorter TCR beta-chains are enriched during antigen-driven selection,<sup>26</sup> we analyzed CDR3 amino acid length and found that selective expanders have significantly shorter CDR3 regions compared with other TCR types in both splenic and SILP compartments ( $p < 0.0001$ ) (Figure 3H). We next asked whether selective expanders harbor CDR3 regions with distinct biochemical properties<sup>27–29</sup> (Table S3). We calculated the mean entropy, polarity, hydropathy, charge, volume, and surface propensity of all recipient CDR3 sequences (Figures 3I and 3J) and demonstrated that selective expanders exhibited CDR3 regions characterized by lower conformational entropy, larger volume, and lower surface propensity compared with expanders and non-expanders ( $p < 0.0001$ ). These data suggest that selective expanders represent a subpopulation of TCRs likely enriched in antigen specificity.

### Diversification of the gut microbiota increases the fraction of selective expanders

If microbiota-derived antigens can generate selective expanders, we surmised that diversification of the microbiome would increase their frequencies. We examined this hypothesis by co-housing F1 mice from Jackson Laboratories for 2 weeks. We then divided these F1 mice into two groups—the first group remained co-housed with each other for 4 weeks, while the second group was co-housed with an equal number of B6 mice derived from Taconic, which harbor a different microbiota compared with the Jackson mice.<sup>30</sup> After this 6-week period, all Jackson F1 recipients received B6.Ptprc<sup>a</sup> CD4<sup>+</sup> T cells from the same donor pool (Figure 4A). Donor T cells isolated from the initial donor pool as well as from the spleen and SILP of each recipient at day 7 were sequenced (Figures S6A–S6C).

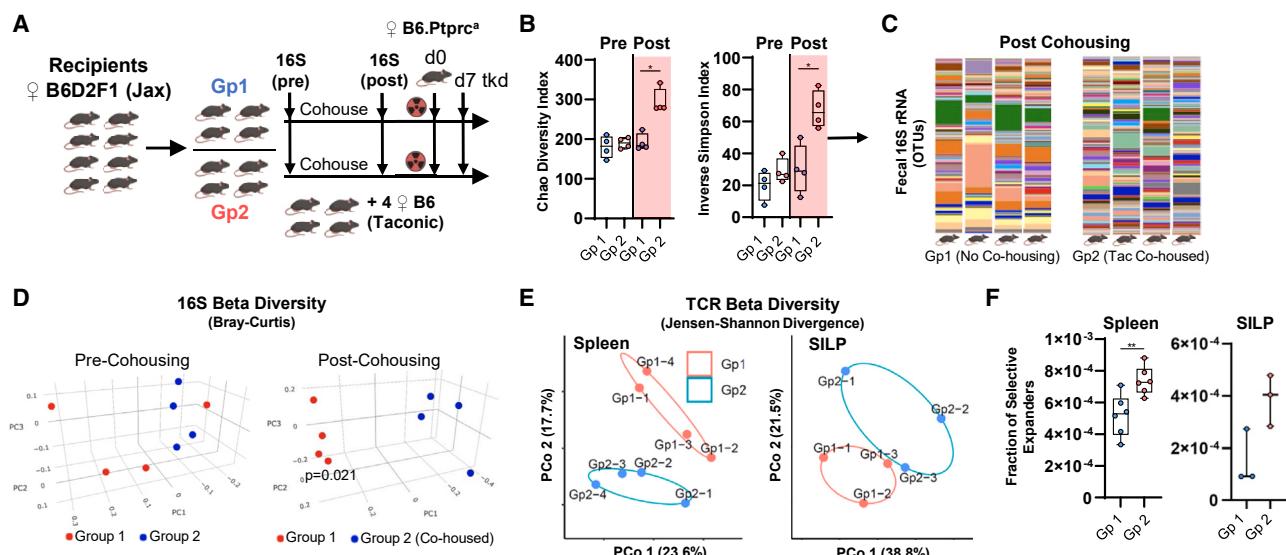
Analysis of fecal material before and after co-housing showed higher alpha diversity indices in the Jackson F1 mice co-housed with Taconic mice compared with the non-co-housed Jackson mice, indicating an increased number of distinct bacterial species present in the co-housed group (Figures 4B and 4C; Table S2). Fecal 16S beta-diversity indices showed a significant difference among co-housed recipient mice that was not detected pre-co-housing (Figure 4D). Although TCR repertoire beta diversity suggested repertoire similarity among co-housed and non-co-housed groups (Figure 4E), probabilistic modeling identified an increased fraction of selective expanders in co-housed recipients, which have a more diversified microbiome compared with non-co-housed recipients (Figures 4F and S6D–S6G; Table S4).

### Microbiota-specific donor T cells selectively expand in the presence of cognate antigen and augment GVHD lethality

To confirm whether differences in microbiome-derived antigenic stimuli trigger selectively expanded T cell clonotypes, we next introduced transgenic donor CD4<sup>+</sup> T cells (segmented filamentous bacteria [SFB] Tg), specific for SFB in the context of I-A<sup>b</sup>, into mice with differing burdens of cognate antigen.<sup>31</sup> SFB is well described to be present in larger quantities in the Taconic versus Jackson B6 murine GI microbiota.<sup>30</sup> At baseline, CD4<sup>+</sup> T cells in the SILP of Taconic B6 and B6D2F1 mice have a higher fraction of SFB-tetramer-positive cells compared with Jackson B6 and F1 SILP (Figure 5A). Either  $1 \times 10^4$  or  $1 \times 10^6$  donor CD4<sup>+</sup> SFB Tg T cells (CD45.2<sup>+</sup>) were subsequently transplanted into Jackson and Taconic B6 (CD45.1<sup>+</sup>) or F1 mice along with polyclonal donor CD4<sup>+</sup> T cells (CD45.1<sup>+</sup>) (Figures 5B and 5C). Donor SFB T cells were significantly expanded in the SILP of Taconic versus Jackson mice in both B6 and F1 recipients by day 8, as indicated by Vβ14<sup>+</sup>SFB<sup>+</sup> cell subsets, at both  $1 \times 10^4$  and  $1 \times 10^6$  donor SFB T cell titrations (Figures 5D and 5E).

To test whether donor anti-microbial T cells have an impact on transplant outcomes, we utilized a sublethal GVHD system with male B6 mice as recipients that develop GVHD when transplanted with Marilyn (Allo) Tg T cells (Figure 5F). Taconic B6 recipients developed lethal GVHD when both donor Allo T cells and SFB T cells were introduced, but not when either Tg T cell was introduced alone (Figures 5G and 5H,  $p < 0.0001$ ). In contrast, Jackson B6 recipients receiving both Allo and SFB T cells did not develop lethal GVHD (Figures 5G and 5H,  $p < 0.0001$ ), confirming that donor SFB T cells lethally augmented GVHD in Taconic (SFB-bearing) mice but not in Jackson (SFB-absent) mice.

We introduced a second transgenic donor CD4<sup>+</sup> T cell (CBir1 Tg) specific for commensal bacterial flagellin<sup>32–34</sup> in the context of I-A<sup>b</sup> into Taconic B6 recipients using the same sublethal GVHD system (Figures 5F and 5I). Recipient mice developed lethal GVHD when CBir1 T cells were introduced along with Marilyn T cells but not when either Tg T cell was introduced alone (Figures 5J and 5K,  $p < 0.0001$ ). To test whether lethal augmentation of the Allo Tg graft was secondary to nonspecific activation of additional T cells, we introduced a similar quantity of B6 polyclonal CD4<sup>+</sup> (non-alloreactive) T cells instead of CBir1 T cells, along with Allo T cells. Lethal augmentation of the alloreactive graft occurred only with the addition of CBir1 T cells and not with polyclonal T cells (Figures 5J and 5K). These data



**Figure 4. Microbiota diversification increases selectively expanded donor T cells**

(A) Eight F1 Jackson-derived recipients were co-housed for 2 weeks prior to separation into two groups (group 1, additional 4 weeks co-housing among each other; group 2, additional 4 weeks co-housing with 4 B6 Taconic-derived mice).  $5 \times 10^6$  bone marrow cells derived from B6 donor mice along with  $2 \times 10^6$  B6.Ptprc<sup>a</sup> CD4<sup>+</sup> splenic T cells were transferred on day 0, with recipient 1,300 cGy total body irradiation (TBI) conditioning on day minus 1. Donor T cells were sorted from the spleen on day 7. Fecal material was taken from each recipient before and after co-housing for 16S rRNA sequencing.

(B) Alpha diversity indices of fecal material showed that co-housed Jackson mice (group 2) exhibited a higher level of diversity following co-housing compared with non-co-housed Jackson mice (group 1).

(C) Sequencing results for groups 1 and 2 post co-housing (operational taxonomic units [OTUs] denoted by color).

(D) 16S beta diversity among pre- and post-co-housing time points between groups 1 and 2 measured via Bray-Curtis dissimilarity (pre-co-housing PERMANOVA  $p = 0.09$ ; post co-housing  $p = 0.021$ ), illustrated using PCoA.

(E) TCR beta-diversity metrics in the post-co-housing time point between groups 1 and 2 were calculated via Jensen-Shannon divergence between repertoire pairs (spleen  $n = 75,000$  random TCRs per sample, PERMANOVA  $p = 0.13$ ; SILP  $n = 10,799$ ,  $p = 0.3$ ).

(F) Compared with non-co-housed Jackson mice, co-housed mice receiving donor grafts contained a higher proportion of selective expanders in the spleen and SILP. Data from 1 experiment. Statistical analyses by Mann-Whitney test (B, F). \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ .

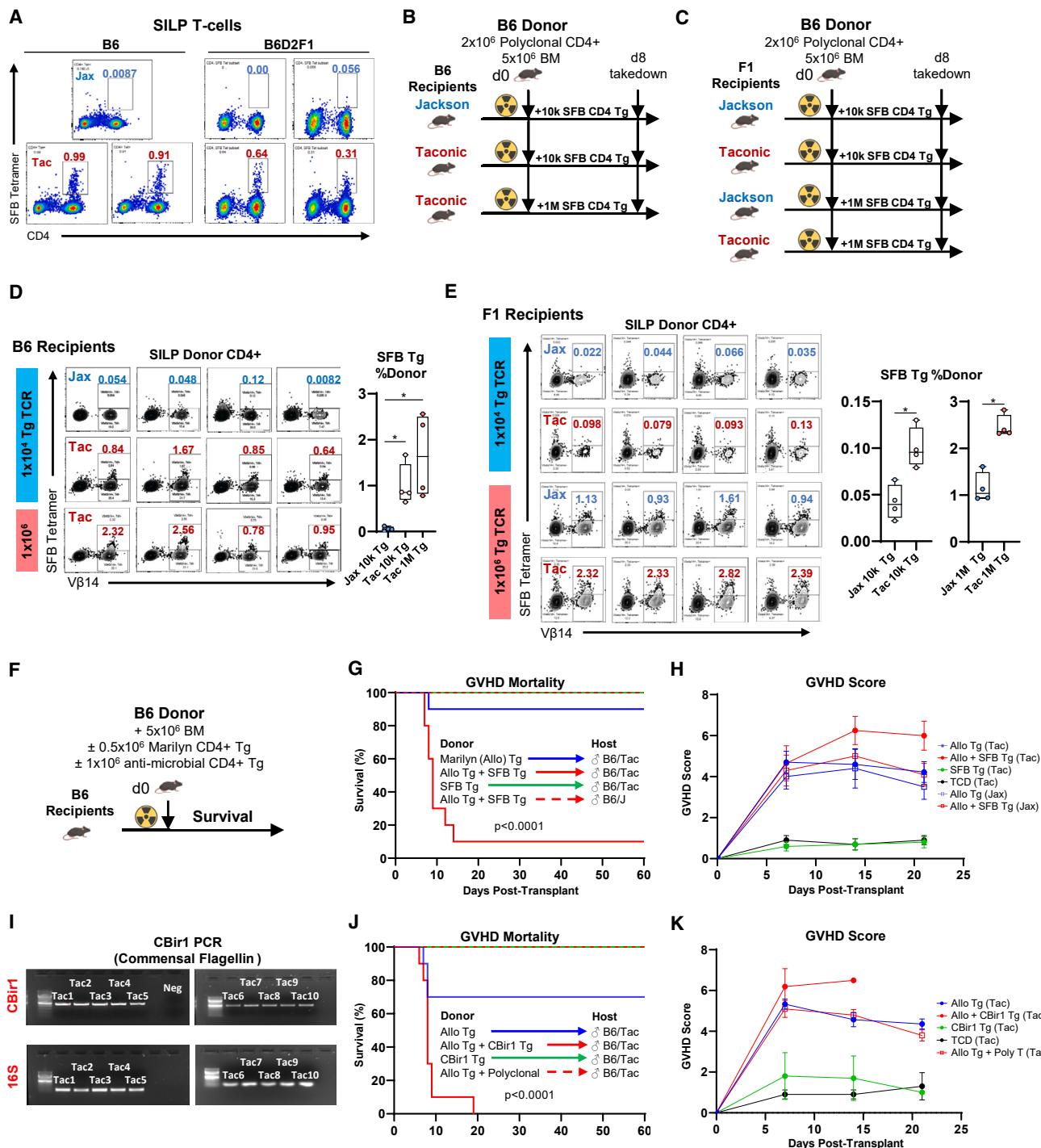
See also Figure S6.

suggest that microbiota-specific donor T cells play a consequential role in the evolution of GVHD and demonstrate preferential expansion in the context of cognate antigen.

#### Alloreactivity promotes early activation phenotype and TNF response in microbiota-specific donor T cells

Although we identified selective expanders in both syngeneic and allogeneic transplant systems (Figures 3D and S5E), the absence of GVHD lethality generally observed in syngeneic systems and the lack of GVHD severity induced by microbiota-specific T cells alone led us to surmise that the function of microbiota-specific T cells may be potentiated by a concurrent alloreactive T cell response. We performed single-cell RNA sequencing, comparing alloreactive Marilyn (Allo) T cells ( $CD90.1^+CD45.2^+$ ) and microbiota-specific CBir1 T cells ( $CD90.1^{neg}CD45.2^+$ ) at day 6 post-transplant when these cells were introduced in isolation or concurrently as part of the donor graft (Figure 6A; Data S3). Although Allo T cells were found in similar uniform manifold approximation and projection (UMAP) space when given in isolation or together with CBir1 T cells (Figure 6B) and did not differ in composition (group 1 versus 2, Figure 5C), the phenotype of CBir1 T cells was significantly altered by the presence of concurrent Allo T cells (group 3 versus 2, Figure 6C). In particular, there was a notable shift in naive CBir1

T cells (cluster 1, green,  $Sell^{high}CD44^{low}CD69^{low}$ ) when transplanted without Allo T cells to a predominantly early-activated phenotype when Allo T cells were concurrently transplanted (cluster 0, dark green,  $Sell^{int}CD44^{low}CD69^{high}$ ), characterized by elevated expression of Fos and JunB and downregulation of Ccr7 and Tcf7 (Figure 6D). Gene set enrichment analysis (GSEA) using the MSigDB Hallmark gene set showed that activated CBir1 T cells (cluster 0) exhibited an increased TNF response signature and decreased oxidative phosphorylation compared with naive CBir1 T cells (cluster 1) (Figure 6E), suggesting that during an allogeneic response, CBir1 T cells are primed by TNF, a known major mediator of GI GVHD.<sup>35</sup> Supporting this observation, serum TNF concentrations were higher in mice receiving CBir1 with Allo T cells compared with mice that received CBir1 T cells alone or in conjunction with syngeneic T cells (Figures 6F and 6G). Signaling through the TNF pathway has been previously shown to inhibit oxidative phosphorylation,<sup>36</sup> activate the AP-1 transcriptional complex,<sup>37</sup> and lower the TCR signaling threshold.<sup>38</sup> Upon stimulation, CBir1 T cells augmented expression of interferon gamma (IFN $\gamma$ ) in the presence of Allo T cells (Figure 6H), whereas there was no difference seen in Allo T cells in the presence or absence of CBir1 T cells. Together, these data demonstrate that an alloreactive T cell response drives concurrent activation and TNF response in



**Figure 5. Microbiota-targeting donor T cells selectively expand in recipients with increased antigenic burden and augment GVHD**

(A) Taconic B6 and F1 mice harbored increased abundance of SFB-specific CD4<sup>+</sup> T cells (DVQFSGAVPNKTD I-A<sup>b</sup> tetramer<sup>+</sup>) compared with Jackson B6 and F1 SILP at baseline (Jax  $n = 3$ , Tac  $n = 4$ ).

(B and C)  $1 \times 10^4$  or  $1 \times 10^6$  donor B6.SFB Tg T cells were transplanted into Jackson and Taconic B6 (1,100 cGy day minus 1 conditioning) or F1 (1,300 cGy) mice along with  $2 \times 10^6$  polyclonal donor CD4<sup>+</sup> T cells.

(D and E) Donor SFB T cells were found to be significantly more expanded in the SILP of Taconic versus Jackson mice in both B6 and F1 recipients at day 8 at  $1 \times 10^4$  donor titrations (B6 and F1, Tac  $n = 4$ , Jax  $n = 4$ ) and at  $1 \times 10^6$  donor titrations (F1, Tac  $n = 4$ , Jax  $n = 4$ ). Data from 1 experiment.

(F)  $0.5 \times 10^6$  CD4<sup>+</sup> donor Marilyn (Allo) T cells and/or  $1 \times 10^6$  donor MACS-sorted CD4<sup>+</sup> microbiota-specific (CBir1 or SFB) T cells were transplanted into male B6 recipients. All recipients were conditioned with 1,100 cGy on day minus 1 and received  $5 \times 10^6$  bone marrow cells on day 0.

(legend continued on next page)

microbiota-specific CBir1 T cells absent during a syngeneic setting.

### Alloreactive T cell responses promote microbial antigen presentation by gut epithelium

To further examine how an alloreactive response can drive tissue-specific changes in the behavior of microbiota-specific donor T cells, we examined the frequency of CBir1 T cells in the spleen and mesenteric lymph node (mLN) after transplant. We found that in the presence of Allo T cells, donor CBir1 T cells were found at similar frequencies in the spleen but were more abundant in the mLN (Figure 6I). Similar to the cytokine profile seen in splenic CBir1 T cells (Figure 6H), CBir1 T cells in the mLN augmented IFN $\gamma$  expression upon stimulation in the presence of Allo T cells (Figure 6J). We also examined infiltration of donor CBir1 T cells in recipient tissue and found that a higher fraction of donor T cells were CBir1-specific in the ileum and colon compared with the skin and liver (Figure 6K), demonstrating relative expansion in compartments containing increased access to cognate microbial antigen.

Given the proximity of the GI microbiota to the epithelium, we hypothesized that alterations to the GI epithelium induced by an alloreactive response may result in increased ability for microbiota-specific T cells to recognize cognate bacterial antigen. To examine this, we expressed the SFB TCR in an NFAT–GFP+ murine hybridoma line that reports TCR signaling<sup>39</sup> (Figures 7A and 7B). We co-cultured intestinal epithelial cells (IECs) harvested at day 6 from B6 recipient mice, with or without an alloreactive donor graft, and demonstrated that, in the setting of an alloreactive response, IECs upregulated MHC class II and induced hybridoma GFP expression in the presence of bacterial peptide (Figures 7C–7E). Furthermore, we demonstrated using Villin-CreERT2 I-A<sup>b</sup> flox recipients that ablation of MHC class II on the gut epithelial surface abrogated hybridoma GFP expression (Figures 7F–7H), confirming that activation of SFB-specific TCR by IECs depended on epithelial class II expression.

To directly examine the ability for microbiota-specific T cells to localize to the GI epithelium, we next performed RNA *in situ* hybridization of the CBir1 CDR3 alpha region (V4J33 probe) in the terminal ileum of mice receiving CBir1 T cells alone or in concert with alloreactive Marilyn T cells. V4J33 probe in pre-transplant CD4 $^{+}$  splenocytes distinguished CBir1 from wild-type B6 T cells (Figure 7I). In contrast to the sparse T cell infiltration observed at day 6 around terminal ileum crypts of mice receiving only microbiota-specific T cells, there was increased CBir1 T cell infiltration when alloreactive T cells were concurrently introduced (Figures 7J and 7K).

As antigen presenting cell (APC) activation plays a central role in GVHD pathogenesis by activating alloreactive T cells, we also examined whether introduction of microbiota-specific T cells augmented APC activation as an indirect mechanism of injury by increasing the alloreactive response. We analyzed surface expression of MHC class II and CD80 on both recipient and donor hematopoietic APCs in the recipient spleen, mLN, ileum, and colon as well as on recipient non-hematopoietic APCs (ileum and colon IECs) and demonstrated that the addition of CBir1 T cells to an allogeneic graft did not increase MHC class II or CD80 expression in recipient or donor cell subsets (Figures S7A–S7C). We also performed secondary transfer of CBir1 T cells or polyclonal syngeneic T cells along with alloreactive T cells to recipient mice on day 5, chosen as a time point following activation of alloreactive T cells.<sup>13</sup> We observed that CBir1 T cells are still able to preferentially augment transplant lethality (Figures S7D and S7E), thus favoring a direct rather than indirect mechanism of injury.

These data, taken together, demonstrate that clonotypic T cell expansion following allogeneic transplantation is not only driven by donor-recipient genetics, as reflected by major and minor histocompatibility matching, but also by differences in microbiota composition that can serve as a source of antigenic stimulation, providing a feedforward mechanism to propagate GVHD. Furthermore, in the context of an alloreactive T cell response that damages gut integrity, microbial antigens may be presented by epithelia, allowing the gut to be concurrently targeted by microbiota-specific T cells whose function is additionally enhanced by the alloreactive response.

### DISCUSSION

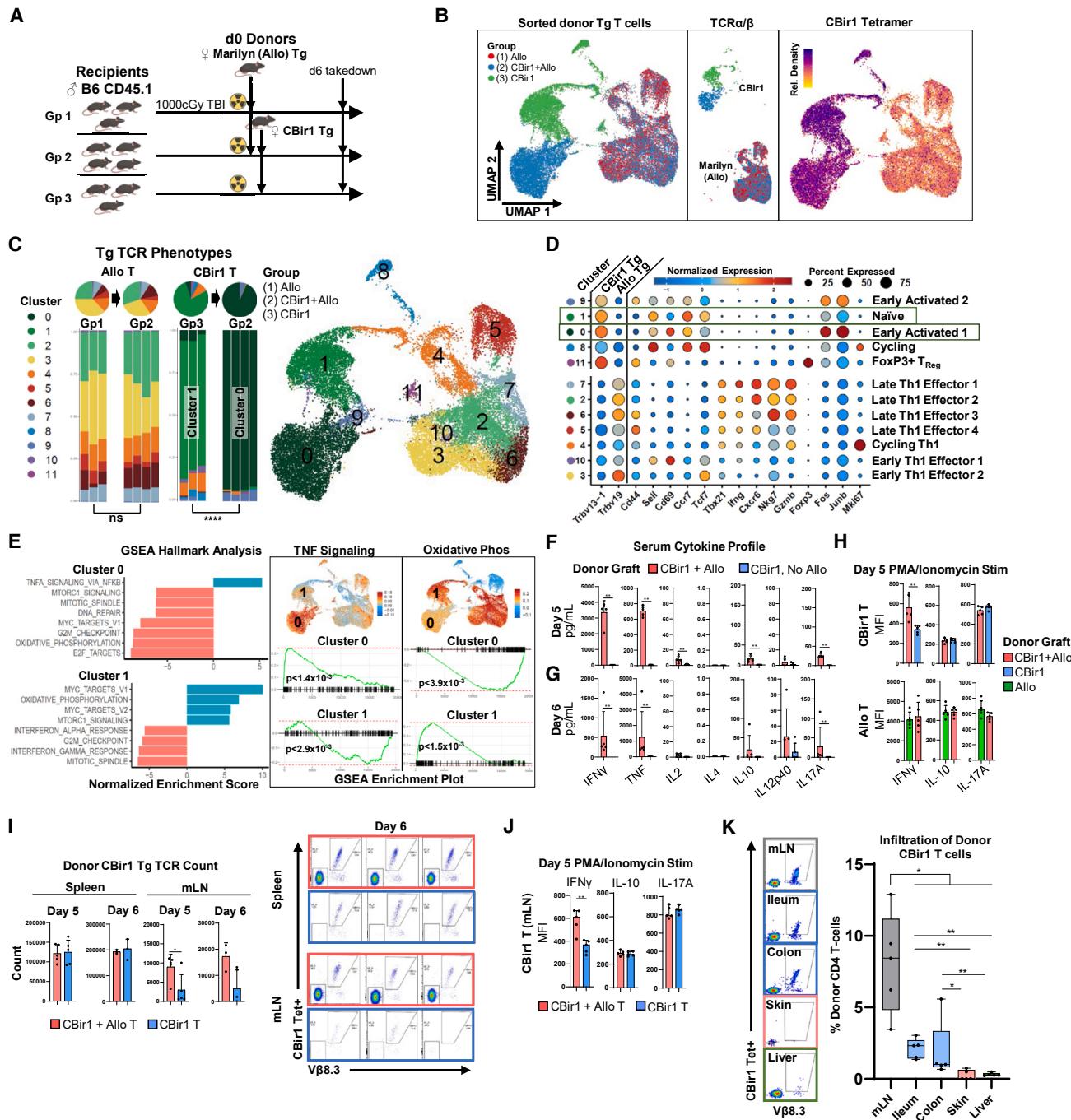
In this study, we examined a basic underlying assumption governing clonotypic T cell expansion in the context of allogeneic stem cell transplant—that major and minor antigen mismatch between donor and recipient dictate the potential expansion of donor T cells. We applied a computational approach utilizing Bayesian inference of donor TCR counts in well-established murine transplant systems to show that a subset of CD4 $^{+}$  T cells differentially expand, even among genetically identical recipients, in a highly controlled environment and that this phenomenon is in part driven by antigenic stimuli from the microbiota.

Although pre-clinical and clinical studies have established a connection between the microbiota and acute GVHD,<sup>20,40–44</sup> the cause-effect relationship between changes in the gut microbiota and GVHD remains unclear. Efforts at mitigating GVHD through targeted manipulation of the microbiome have produced inconsistent and mixed results, and further mechanistic insight is needed.<sup>45,46</sup> Current models aimed at explaining how the gut

(G and H)  $0.5 \times 10^6$  CD4 $^{+}$  Marilyn (Allo) T cells (blue, solid,  $n = 10$ ),  $1 \times 10^6$  SFB T cells (green, solid,  $n = 5$ ), or both Allo and SFB T cells (red, solid,  $n = 10$ ) were transplanted into Taconic B6 male recipients or Jackson B6 male recipients (red, dashed,  $n = 5$ ). (G) GVHD survival. There was a decrease in overall survival in the Taconic recipients receiving with both Allo and SFB T cells compared with other subgroups. Data from 2 replicate experiments. (H) GVHD scores including T cell depleted (TCD,  $n = 5$ ), control (black,  $n = 5$ ), and Jackson B6 receiving only Allo T cells (blue, empty,  $n = 5$ ).

(I) PCR results for B6 Taconic recipients using CBir1 (F,R) and 16S (F,R) control primers.

(J and K)  $0.5 \times 10^6$  Allo T cells (blue, solid,  $n = 10$ ) alone, with  $1 \times 10^6$  CBir1 T cells (red, solid,  $n = 10$ ), with  $1 \times 10^6$  CD4 $^{+}$  polyclonal B6 T cells (red, dashed,  $n = 5$ ), or  $1 \times 10^6$  CBir1 T cells alone (green, solid,  $n = 5$ ) were transplanted into Taconic male B6 recipients. All recipients received  $5 \times 10^6$  B6 bone marrow cells. (J) GVHD survival. There was a decrease in overall survival in the Taconic male recipients treated with both Allo and CBir1 T cells compared with Allo T cells alone, Allo T with polyclonal T cells or CBir1 T cells alone. Data from 2 replicate experiments. (K) GVHD scores including TCD control (black, solid,  $n = 5$ ). Statistical analyses by Mann-Whitney test (D, E), log-rank test (G, J). \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ .



**Figure 6. Alloreactivity promotes early activation phenotype and TNF response in microbiota-specific donor T cells that preferentially expand in the GI compartment**

(A) Male B6.Ptprc<sup>a</sup> (CD45.1, Taconic) were divided into 3 groups representing exposure to different donor grafts: group 1,  $n = 3$ :  $0.5 \times 10^6$  CD4<sup>+</sup> Marilyn (Allo) Tg (CD45.2<sup>+</sup>CD90.1<sup>+</sup>); group 2,  $n = 4$ :  $0.5 \times 10^6$  Allo T with  $1 \times 10^6$  CD4<sup>+</sup> CBir1 T (CD45.2<sup>+</sup>CD90.2<sup>+</sup>); group 3,  $n = 3$ :  $1 \times 10^6$  CD4<sup>+</sup> CBir1 T. All recipients received  $5 \times 10^6$  B6 (CD45.1<sup>+</sup>) bone marrow on day 0 and conditioning with 1,000 cGy on day minus 1. Hashed and sorted splenocytes were harvested at day 6 to isolate Allo T (CD45.2<sup>+</sup>CD45.1<sup>neg</sup>CD90.1<sup>+</sup>) and CBir1 T (CD45.2<sup>+</sup>CD45.1<sup>neg</sup>CD90.1<sup>neg</sup>Vβ14<sup>+</sup>) T cells for single-cell RNA sequencing.

(B) Dimensional reduction using UMAP to visualize gene expression profiling for all groups, including subplots of CBir1 and Allo T cells expressing the corresponding TCRα/β chain and CBir1 tetramer binding.

(C) Analysis of single-cell clusters showed a significant shift of donor CBir1 T cell phenotype when Allo T cells are concurrently introduced (CBir1 T group 2 versus 3 for cluster 1,  $p < 1 \times 10^{-5}$ ; CBir1 T group 2 versus 3 for cluster 0,  $p < 1 \times 10^{-5}$ ). Data from 1 experiment.

(D) Percentage and intensity of gene expression shown by cluster. Cluster 0 (dark green, early activated T cells) and cluster 1 (green, naive T cells) outlined.

(legend continued on next page)

microbiota may influence the development of GVHD have focused on induction of inflammation and other microbiota-dependent effects that may non-specifically shape the immune and metabolic environment.<sup>2,43,47</sup> A major unexplored area in this field relates to how specific T cell clonotypes are shaped directly by the microbiome and potentially contribute to GVHD pathology.

The GI lamina propria harbors an abundant and diverse population of CD4<sup>+</sup> T cells with different functionality and antigen specificities,<sup>48</sup> and reactivity to intestinal bacteria is a normal property of the human CD4<sup>+</sup> T cell repertoire.<sup>49</sup> Evidence that peptide sequences from the gut microbiota acting as an antigenic source for CD4<sup>+</sup> T cells have been demonstrated for several specific bacterial species,<sup>31,50–55</sup> and cognate recognition of microbiota-derived antigens can drive a range of CD4<sup>+</sup> T cell phenotypes with both regulatory and inflammatory functions. In the context of GVHD, CD4<sup>+</sup> T cells play a critical role in the induction of alloreactivity via effector molecules such as TNF, IFN $\gamma$ , and interleukin-17 (IL-17) during a Th1/Th17-mediated response<sup>56</sup> or by providing help to a cytolytic CD8<sup>+</sup> T cell response.

As proof of concept that microbiota-specific T cells play a relevant role in transplantation outcomes, we used CD4<sup>+</sup> T cell clonotypes directed against previously characterized bacterial antigen to demonstrate selective expansion associated with the burden of cognate antigen and an ability to augment GVHD in the context of an alloreactive response. Our data suggest that the interplay between microbiota-responsive and alloreactive CD4<sup>+</sup> T cells in the immune pathology following allogeneic transplantation deserves further exploration. Although we do not examine the CD8<sup>+</sup> T cell response in this setting, our focus on the CD4<sup>+</sup> T cell response builds on recent reports highlighting the importance of MHC class II-dependent mechanisms. We surmise that two potential CD4<sup>+</sup> T cell-driven pathways may explain this link. First, donor CD4<sup>+</sup> T cells recognizing cognate microbial antigen presented by host MHC class II can preferentially differentiate into effector cells during an alloreactive response, augmenting the existing inflammatory cycle. Second, when gut barrier function is compromised by GVHD induced by alloreactive T cells, anti-microbial CD4<sup>+</sup> T cells may also mediate direct tissue injury by direct recognition of microbial antigen presented on MHC class II by the intestinal epithelium. Ongoing inflammation leads to upregulation of MHC class II by intestinal epithelium, as demonstrated here and previously reported by our group,<sup>57</sup> which is juxtaposed to the gut microbiome when barrier function is compromised by

GVHD, thus serving as a context by which these CD4<sup>+</sup> T cells can target epithelial cells presenting cognate microbial antigen. In this model, microbiota-specific CD4<sup>+</sup> T cells targeting MHC-class-II-dependent antigen expressed on local tissue represent an important feedforward mediator of gut injury and lethal GVHD. These mechanisms deserve further downstream characterization and are supported by recent observations suggesting that donor T cells in GVHD tissue are locally maintained and can diverge over time following transplant.<sup>58</sup>

As humans are estimated to co-exist with over 1,000 bacterial species in the gut,<sup>59</sup> our understanding of the allogeneic response in transplantation should not be confined to donor and recipient genetics but also include potential cognate interactions with the GI microbiota. Our data provide a potentially underappreciated mechanism for the well-established observation between dysbiosis and transplant outcomes that centers on the control of clonotypic T cell expansion.

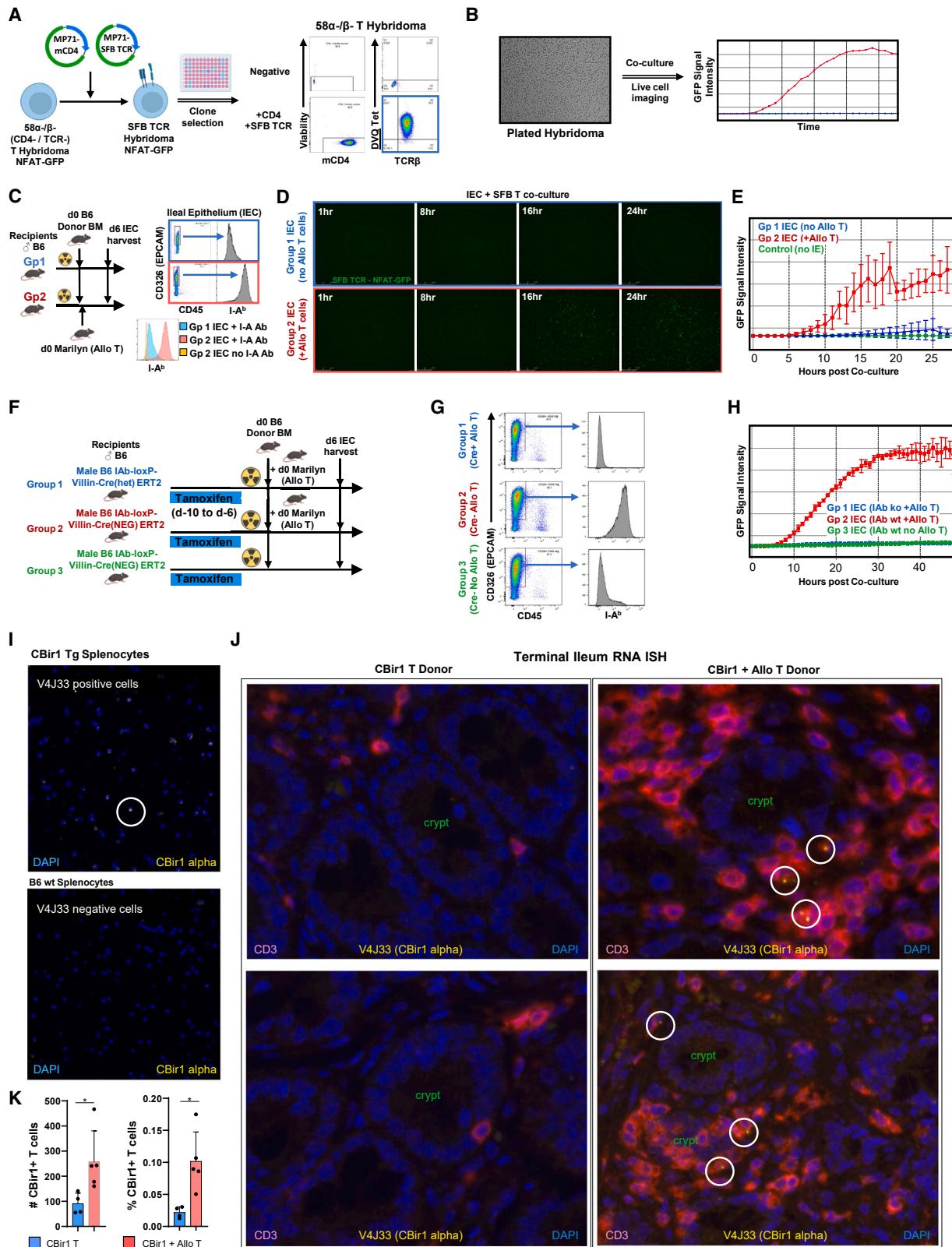
### Limitations of the study

A limitation of our modeling approach is the inability to characterize TCRs that occur very few times in the donor pool, as these T cells are likely to be introduced into only one recipient by chance, rendering comparison of clonal expansion between recipients impossible. Given the sparseness of the majority of donor TCR clonotypes, we are therefore likely underestimating the true fraction of selective expanders. As we employed TCR $\beta$  bulk sequencing due to the depth required for modeling, we were unable to reconstruct selective expanders for antigen screening and instead utilized two microbiota-targeted T cell clonotypes that have been well-characterized. In the normal setting we would expect a more diverse pool of potential microbiota-specific clonotypes present at lower frequencies per clone than that used in our model systems. Further identification of T cell clonotypes directed at both commensal and pathogenic bacteria, and their propensity for various differentiation pathways, would enable more sophisticated modeling that may provide insight into differential responses with respect to pathogenic versus protective immunity.

### STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- (E) GSEA Hallmark analysis including cluster 0 and 1 showed an increase in TNF response genes and a decrease in oxidative phosphorylation in cluster 0. Pathways with FDR < 0.05 and |NES| (normalized enrichment score) > 5 are shown.
- (F) Serum cytokine analysis at day 5 for male B6 mice receiving CBir1 + Allo T cells (red) or CBir1 T cells (blue). Data from 1 experiment ( $n = 5$  each group).
- (G) Cytokine analysis at day 6 for male B6 mice receiving CBir1 + Allo T cells (red), and female B6 mice receiving CBir1 + Marilyn (syngeneic) T cells (blue). Data from 1 experiment ( $n = 5$  each group).
- (H) Intracellular cytokine analysis was performed on CBir1 T cells (CD45.2<sup>+</sup>CD45.1<sup>neg</sup>CD90.1<sup>neg</sup>CD4<sup>+</sup>CD8<sup>neg</sup>V $\beta$ 14<sup>+</sup>CBir1 Tet<sup>+</sup>) and Allo T cells (CD45.2<sup>+</sup>CD45.1<sup>neg</sup>CD90.1<sup>neg</sup>CD8<sup>neg</sup>V $\beta$ 6<sup>+</sup>) isolated from recipient spleen 5 days post-transplant. Cells were stimulated for 4 h with phorbol myristate acetate (PMA)/ionomycin. CBir1 T cells expressed higher levels of IFN $\gamma$  in the presence versus absence of Allo T cells. Data from 1 experiment ( $n = 5$  each group).
- (I)  $1 \times 10^6$  CBir1 T cells with  $0.5 \times 10^6$  Marilyn (Allo) T cells (red) or without Allo T cells (blue) were transplanted into Taconic B6.Ptprc<sup>a</sup> male recipients. Recipients conditioned with 1,100 cGy on day minus 1 and received  $5 \times 10^6$  B6.Ptprc<sup>a</sup> bone marrow cells on day 0. Quantification of donor CBir1 T cells in recipient spleen and mesenteric lymph node (mLN) at days 5 and 6 post-transplant are shown. Data from 2 experiments (day 5,  $n = 5$  each group; day 6,  $n = 3$  each group).
- (J) Intracellular cytokine analysis was performed on CBir1 and Allo T cells isolated from recipient mLN on day 5. Cells were stimulated for 4 h with PMA/ionomycin prior to analysis. CBir1 T cells expressed higher levels of IFN $\gamma$  in the presence versus absence of Allo T cells. Data from 1 experiment ( $n = 5$  each group).
- (K) 5 recipient male B6.Ptprc<sup>a</sup> mice received  $5 \times 10^6$  B6 bone marrow,  $0.3 \times 10^6$  Allo T cells, and  $5 \times 10^6$  CBir1 CD4<sup>+</sup> T cells. Quantification of donor CBir1 T cells in recipient mLN, ileum, colon, skin, and liver at day 9 is shown as fraction of all donor T cells detected. Data from 1 experiment ( $n = 5$  each group). Statistical analyses by Mann-Whitney test (F, H–K) (mean  $\pm$  SD). \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ .



(legend on next page)

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Mice
- METHOD DETAILS
  - Stem cell transplantation
  - Probabilistic simulation of post-transplant TCR expansion
  - Antibodies and flow cytometry
  - MHC Class II tetramer analysis
  - Serum cytokine analysis
  - Cell isolation from spleen and small intestine
  - Gut decontamination
  - TCR amplicon sequencing
  - Publicly available TCR amplicon sequencing dataset
  - PCR and 16S rRNA sequencing
  - Single cell RNA sequencing
  - Generation of TCR hybridomas
  - Live cell imaging
  - RNA in-situ hybridization (ISH) imaging
- QUANTIFICATION AND STATISTICAL ANALYSIS

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.jimmuni.2024.05.018>.

## ACKNOWLEDGMENTS

We acknowledge the Comparative Medicine animal facility team, Experimental Histopathology (Shared Resource of the FHCC/UW Cancer Consortium, grant P30 CA015704), and the Scientific Computing Infrastructure at FHCC (Office of Research Infrastructure Programs, grant no. S10OD028685) for support and Effie Petersdorf and Stephanie Lee for manuscript comments. Funding: NIH K08 HL167161 (A.C.Y.), NIH K12 CA076930 (A.C.Y.), ASTCT New Investigator Award (A.C.Y.), and NIH R01 HL148164 (G.R.H.).

## AUTHOR CONTRIBUTIONS

Conceptualization, A.C.Y., P.H.B., and G.R.H.; methodology, A.C.Y., P.H.B., G.R.H., and C.O.E.; validation, A.C.Y., P.H.B., and G.R.H.; formal analysis, A.C.Y. and P.H.B.; investigation, A.C.Y., M.K., O.G.W., S.A.M., J.R.B., T.B.S., S.T., P.Z., K.S.E., C.R.S., S.R.W.L., T.S., E.N., A.R.S., S.S.B., T.B.S., and S.C.; project administration, A.C.Y., K.S.E., C.R.S., and G.R.H.; funding acquisition, A.C.Y. and G.R.H.; supervision, P.H.B. and G.R.H.; writing—original draft, A.C.Y., P.H.B., and G.R.H.; writing—review and editing, A.C.Y., S.N.F., K.A.M., M.E.B., P.H.B., and G.R.H.

## DECLARATION OF INTERESTS

G.R.H. has consulted for Generon Corporation, Napajen Pharma, iTeos Therapeutics, and Neoleukin Therapeutics and has received unrelated research funding from Compass Therapeutics, Syndax Pharmaceuticals, Applied Molecular Transport, Serplus Technology, Heat Biologics, Laevoroc Oncology, iTeos Therapeutics, Genentech, and CSL Behring.

Received: June 26, 2023

Revised: February 9, 2024

Accepted: May 20, 2024

Published: June 13, 2024

## REFERENCES

1. Goulimy, E., Schipper, R., Pool, J., Blokland, E., Falkenburg, J.H., Vossen, J., Gratwohl, A., Vogelsang, G.B., van Houwelingen, H.C., and van Rood, J.J. (1996). Mismatches of minor histocompatibility antigens between HLA-identical donors and recipients and the development of graft-versus-host disease after bone marrow transplantation. *N. Engl. J. Med.* 334, 281–285. <https://doi.org/10.1056/NEJM199602013340501>.
2. Zeiser, R., and Blazar, B.R. (2017). Acute graft-versus-host disease - biological process, prevention, and therapy. *N. Engl. J. Med.* 377, 2167–2179. <https://doi.org/10.1056/NEJMra1609337>.
3. Davis, M.M., and Boyd, S.D. (2019). Recent progress in the analysis of  $\alpha\beta$ T cell and B cell receptor repertoires. *Curr. Opin. Immunol.* 59, 109–114. <https://doi.org/10.1016/j.coi.2019.05.012>.
4. Meier, J., Roberts, C., Avent, K., Hazlett, A., Berrie, J., Payne, K., Hamm, D., Desmarais, C., Sanders, C., Hogan, K.T., et al. (2013). Fractal

## Figure 7. Presentation of bacterial peptide by ileal epithelium is augmented by the alloreactive response and depends on MHC class II expression

(A) SFB TCR reporter hybridoma was generated from transducing 58 $\alpha$ -/β-NFAT-GFP murine T hybridoma cell line with MP71 vector containing 7B8 SFB TCR  $\alpha$ /β and mCD4 followed by clonal selection (blue square).

(B) SFB reporter hybridomas were plated on a 96-well plate at a density of  $5 \times 10^4$  cells per well and were imaged in co-culture for GFP expression over time.

(C-E) Two groups of recipient male B6 mice received  $5 \times 10^6$  B6 bone marrow with  $0.5 \times 10^6$  Allo T cells (group 2, red,  $n = 4$ ) or without Allo T cells (group 1, blue,  $n = 3$ ) on day 0 with 1,000 cGy conditioning on day minus 1. The intestinal epithelial cells (IECs) of the terminal ileum were harvested at day 6 and sorted (CD326 $^+$  CD45 $^{neg}$ ) for downstream co-culture. (C) I-A $b$  expression was higher in the epithelium of recipient mice in the presence of an alloreactive response. Data from 2 replicate experiments. (D) SFB reporter hybridomas were plated at  $5 \times 10^4$  cells per well with 1  $\mu$ g/mL of DVQ SFB peptide, followed by addition of  $5 \times 10^5$  epithelial cells (group 1, group 2,  $n = 4$ ) or no additional cells (control,  $n = 2$ ). Reporter hybridomas were imaged for GFP expression. Representative images taken from the GFP channel at 1-, 8-, 16-, and 24-h post co-culture. (E) Hybridoma GFP expression analyzed over time showed increased signal when co-cultured with IECs derived from Allo T recipients. Data from 2 replicate experiments.

(F-H) (F) I-A $b$ -loxP × Villin-Cre ERT2 recipient mice treated with 1 mg tamoxifen intraperitoneally (i.p.) daily from day minus 10 to day minus 6 prior to transplant. Group 1 (blue,  $n = 3$ ) was Villin-Cre $^+$  and received  $0.5 \times 10^6$  donor Marilyn (Allo) T cells. Group 2 (red,  $n = 3$ ) was Villin-Cre $^{neg}$  and received  $0.5 \times 10^6$  Allo T cells. Group 3 (green,  $n = 2$ ) was Villin-Cre $^+$  without Allo T cells. All recipients were conditioned with 1,000 cGy on day minus 1. The epithelial cell layer of the ileum was harvested at day 6 and sorted for downstream co-culture. (G) I-A $b$  expression on sorted epithelial cells from the terminal ileum (CD45 $^{neg}$ CD326 $^+$ ) confirmed class II knockout in group 1 (Villin-Cre $^+$ ) versus group 2 (Villin-Cre $^{neg}$ ), both of which received Allo T cells. Group 3 (Villin-Cre $^{neg}$ ) did not receive Allo T cells and had low-intermediate I-A $b$  expression. (H) Hybridoma GFP expression showed increased signal when co-cultured with DVQ peptide and IECs derived from Allo T Cre negative but not Cre $^+$  recipients. Data from 1 experiment.

(I) V4J33 RNA *in situ* hybridization (ISH) probe used to visualize CBir1 Tg donor T cells versus B6 wild-type donor T cells.

(J) 40 $\times$  cross-sectional representative imaging of terminal ileum harvested at day 6 from male B6 mice receiving CBir1 T cells alone or CBir1 with Allo T cells. Visualization of DAPI (blue), CD3 (red), and V4J33 RNA ISH probe (yellow).

(K) The number of CBir1+ T cells as detected by co-localization of V4J33 probe and CD3 stain and fraction of total nucleated cells are shown for recipients of CBir1 T cells (blue,  $n = 4$ ) and both CBir1 and Allo T cells (red,  $n = 5$ ). Data from 1 experiment. Statistical analyses by Mann-Whitney test (K) (mean  $\pm$  SD). \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ .

- organization of the human T cell repertoire in health and after stem cell transplantation. *Biol. Blood Marrow Transplant.* 19, 366–377. <https://doi.org/10.1016/j.bbmt.2012.12.004>.
5. DeWolf, S., Grinshpun, B., Savage, T., Lau, S.P., Obradovic, A., Shonts, B., Yang, S., Morris, H., Zuber, J., Winchester, R., et al. (2018). Quantifying size and diversity of the human T cell alloresponse. *JCI Insight* 3, e121256. <https://doi.org/10.1172/jci.insight.121256>.
  6. Zheng, P., Tamaresis, J., Thangavelu, G., Xu, L., You, X., Blazar, B.R., Negrin, R.S., Zehnder, J.L., Iliopoulos, B.P., and Meyer, E.H. (2020). Recipient-specific T-cell repertoire reconstitution in the gut following murine hematopoietic cell transplant. *Blood Adv.* 4, 4232–4243. <https://doi.org/10.1182/bloodadvances.2019000977>.
  7. Meyer, E.H., Hsu, A.R., Liliental, J., Löhr, A., Florek, M., Zehnder, J.L., Strober, S., Lavori, P., Miklos, D.B., Johnson, D.S., and Negrin, R.S. (2013). A distinct evolution of the T-cell repertoire categorizes treatment refractory gastrointestinal acute graft-versus-host disease. *Blood* 121, 4955–4962. <https://doi.org/10.1182/blood-2013-03-489757>.
  8. Kanakry, C.G., Coffey, D.G., Towleron, A.M.H., Vulic, A., Storer, B.E., Chou, J., Yeung, C.C.S., Gocke, C.D., Robins, H.S., O'Donnell, P.V., et al. (2016). Origin and evolution of the T cell repertoire after posttransplantation cyclophosphamide. *JCI Insight* 1, e86252. <https://doi.org/10.1172/jci.insight.86252>.
  9. Koyama, D., Murata, M., Hanajiri, R., Akashi, T., Okuno, S., Kamoshita, S., Julamane, J., Takagi, E., Miyao, K., Sakemura, R., et al. (2019). Quantitative assessment of T cell clonotypes in human acute graft-versus-host disease tissues. *Biol. Blood Marrow Transplant.* 25, 417–423. <https://doi.org/10.1016/j.bbmt.2018.10.012>.
  10. Meier, J.A., Haque, M., Fawaz, M., Abdeen, H., Coffey, D., Towleron, A., Abdeen, A., Toor, A., Warren, E., Reed, J., et al. (2019). T cell repertoire evolution after allogeneic bone marrow transplantation: an organizational perspective. *Biol. Blood Marrow Transplant.* 25, 868–882. <https://doi.org/10.1016/j.bbmt.2019.01.021>.
  11. Shah, O., Tamaresis, J.S., Kenyon, L.J., Xu, L., Zheng, P., Gupta, P., Rangarajan, K., Lee, S., Spellman, S., Nikiforow, S., et al. (2020). Analysis of the whole CDR3 T cell receptor repertoire after hematopoietic stem cell transplantation in 2 clinical cohorts. *Biol. Blood Marrow Transplant.* 26, 1050–1070. <https://doi.org/10.1016/j.bbmt.2020.01.020>.
  12. Pagliuca, S., Gurnari, C., Hong, S., Zhao, R., Kongkiatkamon, S., Terkawi, L., Zawit, M., Guan, Y., Awada, H., Kishtagari, A., et al. (2021). Clinical and basic implications of dynamic T cell receptor clonotyping in hematopoietic cell transplantation. *JCI Insight* 6, 149080. <https://doi.org/10.1172/jci.insight.149080>.
  13. Koyama, M., Kuns, R.D., Olver, S.D., Raffelt, N.C., Wilson, Y.A., Don, A.L.J., Lineburg, K.E., Cheong, M., Robb, R.J., Markey, K.A., et al. (2011). Recipient nonhematopoietic antigen-presenting cells are sufficient to induce lethal acute graft-versus-host disease. *Nat. Med.* 18, 135–142. <https://doi.org/10.1038/nm.2597>.
  14. Ochando, J.C., Homma, C., Yang, Y., Hidalgo, A., Garin, A., Tacke, F., Angeli, V., Li, Y., Boros, P., Ding, Y., et al. (2006). Alloantigen-presenting plasmacytoid dendritic cells mediate tolerance to vascularized grafts. *Nat. Immunol.* 7, 652–662. <https://doi.org/10.1038/ni1333>.
  15. Beilhack, A., Schulz, S., Baker, J., Beilhack, G.F., Wieland, C.B., Herman, E.I., Baker, E.M., Cao, Y.-A., Contag, C.H., and Negrin, R.S. (2005). In vivo analyses of early events in acute graft-versus-host disease reveal sequential infiltration of T-cell subsets. *Blood* 106, 1113–1122. <https://doi.org/10.1182/blood-2005-02-0509>.
  16. Dash, P., Fiore-Gartland, A.J., Hertz, T., Wang, G.C., Sharma, S., Souquette, A., Crawford, J.C., Clemens, E.B., Nguyen, T.H.O., Kedzierska, K., et al. (2017). Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature* 547, 89–93. <https://doi.org/10.1038/nature22383>.
  17. Naumov, Y.N., Naumova, E.N., Hogan, K.T., Selin, L.K., and Gorski, J. (2003). A fractal clonotype distribution in the CD8+ memory T cell repertoire could optimize potential for immune responses. *J. Immunol.* 170, 3994–4001. <https://doi.org/10.4049/jimmunol.170.8.3994>.
  18. Sharon, E., Sibener, L.V., Battle, A., Fraser, H.B., Garcia, K.C., and Pritchard, J.K. (2016). Genetic variation in MHC proteins is associated with T cell receptor expression biases. *Nat. Genet.* 48, 995–1002. <https://doi.org/10.1038/ng.3625>.
  19. Hill, G.R., Crawford, J.M., Cooke, K.R., Brinson, Y.S., Pan, L., and Ferrara, J.L. (1997). Total body irradiation and acute graft-versus-host disease: the role of gastrointestinal damage and inflammatory cytokines. *Blood* 90, 3204–3213. <https://doi.org/10.1182/blood.V90.8.3204>.
  20. Jenq, R.R., Ubeda, C., Taur, Y., Menezes, C.C., Khanin, R., Dudakov, J.A., Liu, C., West, M.L., Singer, N.V., Equinda, M.J., et al. (2012). Regulation of intestinal inflammation by microbiota following allogeneic bone marrow transplantation. *J. Exp. Med.* 209, 903–911. <https://doi.org/10.1084/jem.20112408>.
  21. Yokota, R., Kaminaga, Y., and Kobayashi, T.J. (2017). Quantification of inter-sample differences in T-cell receptor repertoires using sequence-based information. *Front. Immunol.* 8, 1500. <https://doi.org/10.3389/fimmu.2017.01500>.
  22. Oksanen, J., Simpson, G., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, R., Solymos, P., Stevens, M., Szoebs, E., et al. (2022). Vegan: Community Ecology Package. <http://cran.r-project.org/package=vegan>.
  23. Sorini, C., Cardoso, R.F., Gagliani, N., and Villablanca, E.J. (2018). Commensal bacteria-specific CD4+ T cell responses in health and disease. *Front. Immunol.* 9, 2667. <https://doi.org/10.3389/fimmu.2018.02667>.
  24. Pogorelyy, M.V., Minervina, A.A., Shugay, M., Chudakov, D.M., Lebedev, Y.B., Mora, T., and Walczak, A.M. (2019). Detecting T cell receptors involved in immune responses from single repertoire snapshots. *PLoS Biol.* 17, e3000314. <https://doi.org/10.1371/journal.pbio.3000314>.
  25. Huang, H., Wang, C., Rubelt, F., Scriba, T.J., and Davis, M.M. (2020). Analyzing the *Mycobacterium tuberculosis* immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. *Nat. Biotechnol.* 38, 1194–1202. <https://doi.org/10.1038/s41587-020-0505-4>.
  26. Hou, X., Zeng, P., Zhang, X., Chen, J., Liang, Y., Yang, J., Yang, Y., Liu, X., and Diao, H. (2019). Shorter TCR β-chains are highly enriched during thymic selection and antigen-driven selection. *Front. Immunol.* 10, 299. <https://doi.org/10.3389/fimmu.2019.00299>.
  27. Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hippis, K.W., et al. (2001). Intrinsically disordered protein. *J. Mol. Graph. Model.* 19, 26–59. [https://doi.org/10.1016/s1093-3263\(00\)00138-8](https://doi.org/10.1016/s1093-3263(00)00138-8).
  28. Martin, J., and Lavery, R. (2012). Arbitrary protein-protein docking targets biologically relevant interfaces. *BMC Biophys.* 5, 7. <https://doi.org/10.1186/2046-1682-5-7>.
  29. Shugay, M., Bagaev, D.V., Turchaninova, M.A., Bolotin, D.A., Britanova, O.V., Putintseva, E.V., Pogorelyy, M.V., Nazarov, V.I., Zvyagin, I.V., Kirgizova, V.I., et al. (2015). VDJtools: unifying post-analysis of T cell receptor repertoires. *PLoS Comp. Biol.* 11, e1004503. <https://doi.org/10.1371/journal.pcbi.1004503>.
  30. Ivanov, I.I., Atarashi, K., Manel, N., Brodie, E.L., Shima, T., Karaoz, U., Wei, D., Goldfarb, K.C., Santee, C.A., Lynch, S.V., et al. (2009). Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell* 139, 485–498. <https://doi.org/10.1016/j.cell.2009.09.033>.
  31. Yang, Y., Torchinsky, M.B., Gobert, M., Xiong, H., Xu, M., Linehan, J.L., Alonso, F., Ng, C., Chen, A., Lin, X., et al. (2014). Focused specificity of intestinal TH17 cells towards commensal bacterial antigens. *Nature* 510, 152–156. <https://doi.org/10.1038/nature13279>.
  32. Lodes, M.J., Cong, Y., Elson, C.O., Mohamath, R., Landers, C.J., Targan, S.R., Fort, M., and Hershberg, R.M. (2004). Bacterial flagellin is a dominant antigen in Crohn disease. *J. Clin. Invest.* 113, 1296–1306. <https://doi.org/10.1172/JCI20295>.
  33. Cong, Y., Feng, T., Fujihashi, K., Schoeb, T.R., and Elson, C.O. (2009). A dominant, coordinated T regulatory cell-IgA response to the intestinal microbiota. *Proc. Natl. Acad. Sci. USA* 106, 19256–19261. <https://doi.org/10.1073/pnas.0812681106>.

34. Chiaranunt, P., Tometich, J.T., Ji, J., and Hand, T.W. (2018). T cell proliferation and colitis are initiated by defined intestinal microbes. *J. Immunol.* 201, 243–250. <https://doi.org/10.4049/jimmunol.1800236>.
35. Choi, S.W., and Reddy, P. (2014). Current and emerging strategies for the prevention of graft-versus-host disease. *Nat. Rev. Clin. Oncol.* 11, 536–547. <https://doi.org/10.1038/nrclinonc.2014.102>.
36. Samavati, L., Lee, I., Mathes, I., Lottspeich, F., and Hüttemann, M. (2008). Tumor necrosis factor alpha inhibits oxidative phosphorylation through tyrosine phosphorylation at subunit I of cytochrome c oxidase. *J. Biol. Chem.* 283, 21134–21144. <https://doi.org/10.1074/jbc.M801954200>.
37. Kyriakis, J.M. (1999). Activation of the AP-1 transcription factor by inflammatory cytokines of the TNF family. *Gene Expr.* 7, 217–231.
38. Kim, E.Y., and Teh, H.S. (2001). TNF type 2 receptor (p75) lowers the threshold of T cell activation. *J. Immunol.* 167, 6812–6820. <https://doi.org/10.4049/jimmunol.167.12.6812>.
39. Ise, W., Kohyama, M., Nutsch, K.M., Lee, H.M., Suri, A., Unanue, E.R., Murphy, T.L., and Murphy, K.M. (2010). CTLA-4 suppresses the pathogenicity of self antigen-specific T cells by cell-intrinsic and cell-extrinsic mechanisms. *Nat. Immunol.* 11, 129–135. <https://doi.org/10.1038/ni.1835>.
40. Guthery, S.L., Heubi, J.E., and Filipovich, A. (2004). Enteral metronidazole for the prevention of graft versus host disease in pediatric marrow transplant recipients: results of a pilot study. *Bone Marrow Transplant.* 33, 1235–1239. <https://doi.org/10.1038/sj.bmt.1704474>.
41. Holler, E., Butzhammer, P., Schmid, K., Hundsrucker, C., Koestler, J., Peter, K., Zhu, W., Sporrer, D., Hehlgans, T., Kreutz, M., et al. (2014). Metagenomic analysis of the stool microbiome in patients receiving allogeneic stem cell transplantation: loss of diversity is associated with use of systemic antibiotics and more pronounced in gastrointestinal graft-versus-host disease. *Biol. Blood Marrow Transplant.* 20, 640–645. <https://doi.org/10.1016/j.bbmt.2014.01.030>.
42. Jenq, R.R., Taur, Y., Devlin, S.M., Ponce, D.M., Goldberg, J.D., Ahr, K.F., Littmann, E.R., Ling, L., Gobourne, A.C., Miller, L.C., et al. (2015). Intestinal Blautia is associated with reduced death from graft-versus-host disease. *Biol. Blood Marrow Transplant.* 21, 1373–1383. <https://doi.org/10.1016/j.bbmt.2015.04.016>.
43. Mathewson, N.D., Jenq, R., Mathew, A.V., Koenigsknecht, M., Hanash, A., Toubei, T., Oravecz-Wilson, K., Wu, S.-R., Sun, Y., Rossi, C., et al. (2016). Gut microbiome-derived metabolites modulate intestinal epithelial cell damage and mitigate graft-versus-host disease. *Nat. Immunol.* 17, 505–513. <https://doi.org/10.1038/ni.3400>.
44. Staffas, A., Burgos da Silva, M., and van den Brink, M.R.M. (2017). The intestinal microbiota in allogeneic hematopoietic cell transplant and graft-versus-host disease. *Blood* 129, 927–933. <https://doi.org/10.1182/blood-2016-09-691394>.
45. Fredricks, D.N. (2019). The gut microbiota and graft-versus-host disease. *J. Clin. Invest.* 129, 1808–1817. <https://doi.org/10.1172/JCI125797>.
46. Van Lier, Y.F., Van den Brink, M.R.M., Hazenberg, M.D., and Markey, K.A. (2021). The post-hematopoietic cell transplantation microbiome: relationships with transplant outcome and potential therapeutic targets. *Haematologica* 106, 2042–2053. <https://doi.org/10.3324/haematol.2020.270835>.
47. Stein-Thoeringer, C.K., Nichols, K.B., Lazrak, A., Docampo, M.D., Slingerland, A.E., Slingerland, J.B., Clurman, A.G., Armijo, G., Gomes, A.L.C., Shono, Y., et al. (2019). Lactose drives Enterococcus expansion to promote graft-versus-host disease. *Science* 366, 1143–1149. <https://doi.org/10.1126/science.aax3760>.
48. Kiner, E., Willie, E., Vijaykumar, B., Chowdhary, K., Schmutz, H., Chandler, J., Schnell, A., Thakore, P.I., LeGros, G., Mostafavi, S., et al. (2021). Gut CD4+ T cell phenotypes are a continuum molded by microbes, not by TH archetypes. *Nat. Immunol.* 22, 216–228. <https://doi.org/10.1038/s41590-020-00836-7>.
49. Hegazy, A.N., West, N.R., Stubbington, M.J.T., Wendt, E., Suijker, K.I.M., Datsi, A., This, S., Danne, C., Campion, S., Duncan, S.H., et al. (2017). Circulating and Tissue-Resident CD4+ T Cells With Reactivity to Intestinal Microbiota Are Abundant in Healthy Individuals and Function Is Altered During Inflammation. *Gastroenterology* 153, 1320–1337.e1316. <https://doi.org/10.1053/j.gastro.2017.07.047>.
50. Lathrop, S.K., Bloom, S.M., Rao, S.M., Nutsch, K., Lio, C.-W., Santacruz, N., Peterson, D.A., Stappenbeck, T.S., and Hsieh, C.-S. (2011). Peripheral education of the immune system by colonic commensal microbiota. *Nature* 478, 250–254. <https://doi.org/10.1038/nature10434>.
51. Xu, M., Pokrovskii, M., Ding, Y., Yi, R., Au, C., Harrison, O.J., Galan, C., Belkaid, Y., Bonneau, R., and Littman, D.R. (2018). c-MAF-dependent regulatory T cells mediate immunological tolerance to a gut pathobiont. *Nature* 554, 373–377. <https://doi.org/10.1038/nature25500>.
52. Ansaldi, E., Slayden, L.C., Ching, K.L., Koch, M.A., Wolf, N.K., Plichta, D.R., Brown, E.M., Graham, D.B., Xavier, R.J., Moon, J.J., and Barton, G.M. (2019). Akkermansia muciniphila induces intestinal adaptive immune responses during homeostasis. *Science* 364, 1179–1184. <https://doi.org/10.1126/science.aaw7479>.
53. Węgorzewska, M.M., Glowacki, R.W.P., Hsieh, S.A., Donermeyer, D.L., Hickey, C.A., Horvath, S.C., Martens, E.C., Stappenbeck, T.S., and Allen, P.M. (2019). Diet modulates colonic T cell responses by regulating the expression of a *Bacteroides thetaiotaomicron* antigen. *Sci. Immunol.* 4, eaau9079. <https://doi.org/10.1126/sciimmunol.aau9079>.
54. Pedersen, T.K., Brown, E.M., Plichta, D.R., Johansen, J., Twardus, S.W., Delorey, T.M., Lau, H., Vlamakis, H., Moon, J.J., Xavier, R.J., and Graham, D.B. (2022). The CD4(+) T cell response to a commensal-derived epitope transitions from a tolerant to an inflammatory state in Crohn's disease. *Immunity* 55, 1909–1923.e6. <https://doi.org/10.1016/j.jimmuni.2022.08.016>.
55. Nagashima, K., Zhao, A., Atabakhsh, K., Bae, M., Blum, J.E., Weakley, A., Jain, S., Meng, X., Cheng, A.G., Wang, M., et al. (2023). Mapping the T cell repertoire to a complex gut bacterial community. *Nature* 621, 162–170. <https://doi.org/10.1038/s41586-023-06431-8>.
56. Coghill, J.M., Sarantopoulos, S., Moran, T.P., Murphy, W.J., Blazar, B.R., and Serody, J.S. (2011). Effector CD4+ T cells, the cytokines they generate, and GVHD: something old and something new. *Blood* 117, 3268–3276. <https://doi.org/10.1182/blood-2010-12-290403>.
57. Koyama, M., Mukhopadhyay, P., Schuster, I.S., Henden, A.S., Hülsdünker, J., Varelias, A., Vetzou, M., Kuns, R.D., Robb, R.J., Zhang, P., et al. (2019). MHC Class II antigen presentation by the intestinal epithelium initiates graft-versus-host disease and is influenced by the microbiota. *Immunity* 51, 885–898.e7. <https://doi.org/10.1016/j.jimmuni.2019.08.011>.
58. Sacirbegovic, F., Günther, M., Greco, A., Zhao, D., Wang, X., Zhou, M., Rosenberger, S., Oberbarnscheidt, M.H., Held, W., McNiff, J., et al. (2023). Graft-versus-host disease is locally maintained in target tissues by resident progenitor-like T cells. *Immunity* 56, 369–385.e6. <https://doi.org/10.1016/j.jimmuni.2023.01.003>.
59. Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65. <https://doi.org/10.1038/nature08821>.
60. Denman, S.E., and McSweeney, C.S. (2006). Development of a real-time PCR assay for monitoring anaerobic fungal and cellulolytic bacterial populations within the rumen. *FEMS Microbiol. Ecol.* 58, 572–582. <https://doi.org/10.1111/j.1574-6941.2006.00190.x>.
61. Morita, S., Kojima, T., and Kitamura, T. (2000). Plat-E: an efficient and stable system for transient packaging of retroviruses. *Gene Ther.* 7, 1063–1066. <https://doi.org/10.1038/sj.gt.3301206>.
62. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell* 184, 3573–3587.e29. <https://doi.org/10.1016/j.cell.2021.04.048>.
63. Cooke, K.R., Kobzik, L., Martin, T.R., Brewer, J., Delmonte, J., Crawford, J.M., and Ferrara, J.L. (1996). An experimental model of idiopathic

- pneumonia syndrome after bone marrow transplantation: I. The roles of minor H antigens and endotoxin. *Blood* 88, 3230–3239. <https://doi.org/10.1182/blood.V88.8.3230.bloodjournal8883230>.
64. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
65. Thorne, T. (2018). Approximate inference of gene regulatory network models from RNA-Seq time series data. *BMC Bioinformatics* 19, 127. <https://doi.org/10.1186/s12859-018-2125-2>.
66. Hülsdünker, J., Ottmüller, K.J., Neeff, H.P., Koyama, M., Gao, Z., Thomas, O.S., Follo, M., Al-Ahmad, A., Prinz, G., Duquesne, S., et al. (2018). Neutrophils provide cellular communication between ileum and mesenteric lymph nodes at graft-versus-host disease onset. *Blood* 131, 1858–1869. <https://doi.org/10.1182/blood-2017-10-812891>.
67. Barennes, P., Quiniou, V., Shugay, M., Egorov, E.S., Davydov, A.N., Chudakov, D.M., Uddin, I., Ismail, M., Oakes, T., Chain, B., et al. (2021). Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases. *Nat. Biotechnol.* 39, 236–245. <https://doi.org/10.1038/s41587-020-0656-3>.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
PE/Cy7 anti-mouse CD3ε (Clone 145-2C11)	BioLegend	Cat #100319; RRID:AB_312684
BUV737 anti-mouse CD3ε (Clone 145-2C11)	BD Biosciences	Cat #612803; RRID:AB_2870130
Pac Blue anti-mouse CD4 (Clone RM4-5)	BioLegend	Cat #100534; RRID:AB_493375
BUV496 anti-mouse CD4 (Clone GK1.5)	BD Biosciences	Cat #612952; RRID:AB_2813886
APC/Cy7 anti-mouse CD8a (Clone 53-6.7)	BioLegend	Cat #100713; RRID:AB_312752
BUV805 anti-mouse CD8a (Clone 53-6.7)	BD Biosciences	Cat #612898; RRID:AB_2870186
PE/Cy5 anti-mouse CD19 (Clone 6D5)	BioLegend	Cat #115510; RRID:AB_313645
BV510 anti-mouse NK1.1 (Clone PK136)	BD Biosciences	Cat #563096; RRID:AB_2738002
BUV563 anti-mouse CD49b (Clone HMa2)	BD Biosciences	Cat #741280; RRID:AB_2870819
AF700 anti-mouse Ly6C (Clone HK1.4)	BioLegend	Cat #128024; RRID:AB_10640119
APC/Cy7 anti-mouse Ly6G (Clone 1A8)	BioLegend	Cat #127624; RRID:AB_10640819
PerCP/Cy5.5 anti-mouse CD11b (Clone M1/70)	BioLegend	Cat #101228; RRID:AB_893233
BV605 anti-mouse CD11c (Clone N418)	BioLegend	Cat #117334; RRID:AB_11204262
BUV737 anti-mouse SiglecH (Clone 440c)	BD Biosciences	Cat #748293; RRID:AB_2872719
PE/Cy7 anti-mouse CD64 (Clone X54-5/7.1)	BioLegend	Cat #139314; RRID:AB_2563903
PE anti-mouse CD80 (Clone 16-10A1)	BioLegend	Cat #104708; RRID:AB_313128
PE/Cy anti-mouse CD45.1 (Clone A20)	BioLegend	Cat #110730; RRID:AB_1134168
BUV395 anti-mouse CD45.1 (Clone A20)	BD Biosciences	Cat #565212; RRID:AB_2722493
FITC anti-mouse CD45.2 (Clone A20)	BD Biosciences	Cat #561874; RRID:AB_10894189
BV510 anti-mouse CD45.2 (Clone 104)	BioLegend	Cat #109837; RRID:AB_2561393
PE anti-mouse CD90.1 (Clone HIS51)	Thermo Fisher Scientific	Cat #12-0900-81; RRID:AB_465773
APC/Fire750 anti-mouse H-2Db (Clone KH95)	BioLegend	Cat #111519; RRID:AB_2728132
AF647 anti-mouse H-2Dd (Clone 34-2-12)	BioLegend	Cat #110612; RRID:AB_492913
Pac Blue anti-mouse I-A/I-E (Clone M5/114.15.2)	BioLegend	Cat #107619; RRID:AB_493528
AF647 anti-mouse CD326/EPCAM (Clone G8.8)	BioLegend	Cat #118212; RRID:AB_1134101
BV421 anti-mouse IFNγ (Clone XMG1.2)	BioLegend	Cat #505829; RRID:AB_10897937
PE/Dazzle594 anti-mouse IL10 (Clone JESS-16E3)	BioLegend	Cat #505033; RRID:AB_2566328
BV605 anti-mouse IL-17A (Clone TC11-18H10.1)	BioLegend	Cat #506927; RRID:AB_11126144
APC/Cy7 anti-mouse Vα2 (Clone B20.1)	BioLegend	Cat #127818; RRID:AB_10682897
APC anti-mouse Vβ6 (Clone RR4-7)	BioLegend	Cat #140005; RRID:AB_2564056
PE anti-mouse Vβ8.3 (Clone 1B3.3)	BioLegend	Cat #156303; RRID:AB_2800699
FITC anti-mouse Vβ14 (Clone 14-2)	BD Biosciences	Cat #553258; RRID:AB_394738
TotalSeqC anti-mouse Hashtag 1 (C0301)	BioLegend	Cat #155861; RRID:AB_2800693
TotalSeqC anti-mouse Hashtag 2 (C0302)	BioLegend	Cat #155863; RRID:AB_2800694
TotalSeqC anti-mouse Hashtag 3 (C0303)	BioLegend	Cat #155865; RRID:AB_2800695
TotalSeqC anti-mouse Hashtag 4 (C0304)	BioLegend	Cat #155867; RRID:AB_2800696
TotalSeqC anti-mouse Hashtag 5 (C0305)	BioLegend	Cat #155869; RRID:AB_2800697
TotalSeqC anti-mouse Hashtag 6 (C0306)	BioLegend	Cat #155871; RRID:AB_2819910
TotalSeqC anti-mouse Hashtag 7 (C0307)	BioLegend	Cat #155873; RRID:AB_2819911
TotalSeqC anti-mouse Hashtag 8 (C0308)	BioLegend	Cat #155875; RRID:AB_2819912
TotalSeqC anti-mouse Hashtag 9 (C0309)	BioLegend	Cat #155877; RRID:AB_2819913
TotalSeqC anti-mouse Hashtag 10 (C03010)	BioLegend	Cat #155879; RRID:AB_2819914
TotalSeqC anti-APC (C0987)	BioLegend	Cat #408007; RRID:AB_2820077
Rabbit monoclonal Anti-CD3ε	Abcam	Cat #ab215212; RRID: AB_3101752

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
7-Aminoactinomycin D (7AAD)	Sigma Aldrich	Cat #A9400-5MG
LIVE/DEAD Fixable Aqua Dead Cell Stain Kit	Thermo Fisher Scientific	Cat #L34957
Phorbol 12-myristate 13-acetate (PMA)	Sigma Aldrich	Cat #P1585-1MG
Ionomycin	Sigma Aldrich	Cat #I0634-1MG
Brefeldin A Solution (1000X)	BioLegend	Cat #420601
Tamoxifen	MP Biomedicals	CAS #10540-29-1
RectroNectin	Takara Bio	Cat #T100A
DNA Primer: CBir1 (F, R)	Integrated DNA Technologies; Dr. C. Elson, UAB	N/A
DNA Primer: 16S RNA (F, R)	Integrated DNA Technologies; Denman et al. <sup>60</sup>	N/A
DNA Primer: TS41/SS90 (F)	Integrated DNA Technologies	N/A
DNA Primer: TS42/SS91 (R)	Integrated DNA Technologies	N/A
BA-Mm-Tra-V4J33-CDR3-C1 (RNA Probe)	ACD	1210058-C1
BA-Mm-Ppb1-1zz (RNA probe)	ACD	Cat #712358
BA-DapB-zz (RNA probe)	ACD	Cat #701028
DVQ (SFB) peptide – I-A <sup>b</sup> tetramer	National Institutes of Health; Yang et al. <sup>31</sup>	N/A
YSN (CBir1) peptide – I-A <sup>b</sup> tetramer	National Institutes of Health; Chiaranunt et al. <sup>34</sup>	N/A
DVQ peptide	Genscript; Yang et al. <sup>31</sup>	N/A
MP71 (plasmid)	Dr. S. Riddell, Fred Hutchinson Cancer Center, Seattle, USA	N/A
pBluescript mCD4 (plasmid)	Addgene	Cat #14613
<b>Critical Commercial Assays</b>		
eBiosciences Foxp3 / Transcription Factor Staining Buffer Set	eBioscience	Cat #00-5523-00
Cytometric Bead Array; IFN $\gamma$ , TNF, IL-2, IL-4, IL-10, IL12p40, IL-17A	BD Biosciences	Cat #560485
Lamina Propria Dissociation Kit, mouse	MACS Miltenyi Biotec	Cat #130-097-410
QIAamp DNA Mini Kit	QIAGEN	Cat #51304
QIAamp Fast DNA Stool Mini Kit	QIAGEN	Cat #51604
RNeasy Mini Kit	QIAGEN	Cat #74106
SMARTer Mouse TCR a/b Profiling Kit	Takara Bio	Cat #634402
Chromium Next GEM Single Cell 5'Kit (v2)	10x Genomics	Cat #1000263
Chromium Single Cell Mouse TCR Amplification Kit (v2)	10x Genomics	Cat #1000254
Chromium Next GEM Chip K Single Cell Kit	10x Genomics	Cat #1000287
Nextseq 2000 P3 Reagents (100 cycles)	Illumina	Cat #20040559
Xfect Transfection Reagent	Takara Bio	Cat #631318
Basescope LS Reagent Kit	ACD	Cat #323600
Bond Polymer Refine Red Detection	Leica Biosystems	Cat #DS9390
Opal Anti-RB HRP Kit	Akoya Biosciences	Cat #ARR1001KT
Opal 690 Reagent Pack	Akoya Biosciences	Cat #FP1497001KT
1x Plus Automation Amplification Diluent	Akoya Biosciences	Cat #FP1609
ProLong Gold Antifade Mountant	Invitrogen	Cat #P36930
<b>Deposited Data</b>		
RNAseq data	GEO	GEO: GSE267178; Mendeley Data: <a href="https://www.doi.org/10.17632/jvcnngktk6.1">https://www.doi.org/10.17632/jvcnngktk6.1</a>

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Sequencing analysis / code	Github	Github: <a href="https://github.com/acyeh-lab/2023">github.com/acyeh-lab/2023</a> ; Zenodo: <a href="https://doi.org/10.5281/zenodo.11081846">https://doi.org/10.5281/zenodo.11081846</a>
ALICE output	Zenodo	Zenodo: <a href="https://doi.org/10.5281/zendodo.7402922">https://doi.org/10.5281/zendodo.7402922</a>
GLIPH2 output	Zenodo	Zenodo: <a href="https://doi.org/10.5281/zendodo.7402790">https://doi.org/10.5281/zendodo.7402790</a>
<b>Experimental Models: Cell Lines</b>		
NFAT-GFP 58 $\alpha$ - $\beta$ - murine T hybridoma	Dr. K. Murphy, Washington University, St. Louis, USA; Ise et al. <sup>39</sup>	N/A
PlatE packaging cells	Dr. S. Riddell, Fred Hutchinson Cancer Center, Seattle, USA; Morita et al. <sup>61</sup>	N/A
<b>Experimental Models: Organism/Strain</b>		
Mouse: B6.WT: C57BL/6J	The Jackson Laboratory	Strain #000664
Mouse: B6.WT: C57BL/6Tac	Taconic Biosciences	B6-M, B6-F
Mouse: B6.Ptprc <sup>a</sup>	Fred Hutchinson Cancer Center	N/A
Mouse: B6D2F1: B6D2F1/J	The Jackson Laboratory	Strain #100006
Mouse: BALB.b	Fred Hutchinson Cancer Center	N/A
Mouse: B6.Marilyn (Rag2 <sup>-/-</sup> )	Dr. P. Matzinger, NIH, Bethesda, MD, USA	N/A
Mouse: B6.TEa (Rag1 <sup>-/-</sup> )	Dr. J. Bromberg, Mt. Sinai School of Medicine, NY, USA; Ochando et al. <sup>14</sup>	N/A
Mouse: B6.MHCII <sup>-/-</sup>	Fred Hutchinson Cancer Center	N/A
Mouse. Villin Cre-ER <sup>T2</sup> I-A <sup>b-f/f</sup>	QIMR Berghofer	N/A
Mouse: B6.CBir1	Dr. C. Elson, University of Alabama, Birmingham, AL; Cong et al. <sup>33</sup>	N/A
Mouse: B6.SFB (7B8): C57BL/6-Tg(Tcrα,Tcrβ) 2Lii/J	The Jackson Laboratory	Strain #027230
<b>Software and Algorithms</b>		
BD FACSDiva software version 8	BD Biosciences	<a href="https://www.bd biosciences.com/en-us/products/software/instrument-software/bd-facsdiva-software">https://www.bd biosciences.com/en-us/products/software/instrument-software/bd-facsdiva-software</a>
FlowJo v10	Tree Star	<a href="https://www.flowjo.com/">https://www.flowjo.com/</a>
R statistical software (v3.5.3)	The R Project for Statistical Programming	<a href="https://www.r-project.org/">https://www.r-project.org/</a>
Seurat (v4.3.0)	<a href="https://github.com/satijalab/seurat/">https://github.com/satijalab/seurat/</a> ; Hao et al. <sup>62</sup>	N/A
GSEA Analysis	<a href="https://github.com/ctlab/fgsea">https://github.com/ctlab/fgsea</a>	<a href="https://bioconductor.org/packages/release/bioc/html/fgsea.html">https://bioconductor.org/packages/release/bioc/html/fgsea.html</a>
RECOLD	<a href="https://github.com/Q-bio-at-IIS/RECOLD/blob/master/codes;">https://github.com/Q-bio-at-IIS/RECOLD/blob/master/codes/</a> ; Yokota et al. <sup>21</sup>	N/A
VEGAN	<a href="https://github.com/vegandevels/vegan">https://github.com/vegandevels/vegan</a> ; Oksanen et al. <sup>22</sup>	N/A
ALICE	<a href="https://github.com/pogorely/ALICE">https://github.com/pogorely/ALICE</a> ; Pogorely et al. <sup>24</sup>	N/A
GLIPH (v2)	<a href="http://50.255.35.37:8080">http://50.255.35.37:8080</a> ; Huang et al. <sup>25</sup>	N/A
GraphPad Prism (v9.4.0)	GraphPad Software	<a href="https://www.graphpad.com/scientific-software/prism/">https://www.graphpad.com/scientific-software/prism/</a>
HALO image analysis software (v3.5), FISH-IF module (v2.2.5)	Indica Labs	<a href="https://indicalab.com/halo/">https://indicalab.com/halo/</a>

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Other		
BD FACSymphony A3	BD Biosciences	<a href="https://www.bd-biosciences.com/en-us/products/instruments/flow-cytometers/research-cell-analyzers/bd-facsymphony-a3">https://www.bd-biosciences.com/en-us/products/instruments/flow-cytometers/research-cell-analyzers/bd-facsymphony-a3</a>
BD FACSymphony S6	BD Biosciences	<a href="https://www.bd-biosciences.com/en-us/products/instruments/flow-cytometers/research-cell-sorters/bd-facsymphony-s6">https://www.bd-biosciences.com/en-us/products/instruments/flow-cytometers/research-cell-sorters/bd-facsymphony-s6</a>
Sony MA900	Sony	<a href="https://www.sonybiotechnology.com/us/instruments/ma900-multi-application-cell-sorter/">https://www.sonybiotechnology.com/us/instruments/ma900-multi-application-cell-sorter/</a>
xCELLigence Real-Time Cell Analyzer	Agilent	<a href="https://www.agilent.com/en/product/cell-analysis/real-time-cell-analysis/rtdca-analyzers/xcelligence-rtca-esight-imaging-impedance-741228">https://www.agilent.com/en/product/cell-analysis/real-time-cell-analysis/rtdca-analyzers/xcelligence-rtca-esight-imaging-impedance-741228</a>
Leica Bond Rx Autostainer	Leica	<a href="https://www.leicabiosystems.com/ihc-ISH/ihc-ISH-instruments/bond-rx">https://www.leicabiosystems.com/ihc-ISH/ihc-ISH-instruments/bond-rx</a>
Illumina Nextseq 2000	Illumina	<a href="https://www.illumina.com/systems/sequencing-platforms/nextseq-1000-2000/specifications.html">https://www.illumina.com/systems/sequencing-platforms/nextseq-1000-2000/specifications.html</a>
Illumina MiSeq v3	Illumina	<a href="https://www.illumina.com/systems/sequencing-platforms/miseq/specifications.html">https://www.illumina.com/systems/sequencing-platforms/miseq/specifications.html</a>
Mouse TCRβ assay (v3) (service)	Adaptive Biotechnologies	<a href="https://adaptivebiotech.com/adaptive-immunosequencing/">https://adaptivebiotech.com/adaptive-immunosequencing/</a>
16S Targeted Sequencing (service)	Zymo Research	<a href="https://www.zymoresearch.com/pages/16s-its-amplicon-sequencing">https://www.zymoresearch.com/pages/16s-its-amplicon-sequencing</a>
TCR 7B8 Cassette Synthesis (service)	Twist Biosciences	<a href="https://www.twistbioscience.com/products/genes">https://www.twistbioscience.com/products/genes</a>

**RESOURCE AVAILABILITY**

**Lead contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Geoffrey R. Hill ([grhill@fredhutch.org](mailto:grhill@fredhutch.org)).

**Materials availability**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact.

**Data and code availability**

- RNA sequencing data can be accessed via GEO: GSE267178 and Mendeley Data: <https://www.doi.org/10.17632/jvcnngktk6.1>. Sample data sets and output files can be found in associated Zenodo repositories referred to in Supplemental Materials and is publicly available as of the date of publication.
- Source code can be found at “[github.com/acyeh-lab/2023/tree/main/tcr-simulation](https://github.com/acyeh-lab/2023/tree/main/tcr-simulation)” and “[github.com/acyeh-lab/2023/tree/main/scseq](https://github.com/acyeh-lab/2023/tree/main/scseq)”. Please also refer to Zenodo DOI: “<https://doi.org/10.5281/zenodo.11081846>”.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

**EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**

**Mice**

All mice were housed in the animal facility at the Fred Hutchinson Cancer Center. C57BL/6J (B6.WT, CD45.2, H-2<sup>b</sup>, I-A<sup>b</sup>, I-E<sup>null</sup>), C57BL/6J (B6.Ptprc<sup>a</sup>, CD45.1), B6D2F1 (CD45.2, H-2<sup>b/d</sup>, I-A<sup>b/d</sup>, I-E<sup>d</sup>), and BALB.b (CD45.2, H-2<sup>b</sup>) mice were either bred inhouse or purchased from Taconic Biosciences or Jackson Laboratory. Transgenic and knockout mice on a B6 background originated

as follows: TEa Tg (CD45.1, Rag1<sup>-m</sup>, H-2<sup>b</sup>, I-A<sup>b</sup>, I-E<sup>Null</sup>), Marilyn Tg (CD45.2, CD90.1, Rag2<sup>-</sup>, H-2<sup>b</sup>, I-A<sup>b</sup>, I-E<sup>Null</sup>; Dr. P Matzinger, NIH, Bethesda MD, USA), 7B8 SFB Tg (CD45.2; Jackson Laboratory ID# 027230), CBir1 Tg<sup>33</sup> (CD45.2; Dr. C. Elson, University of Alabama, Birmingham, AL). SPF Mice were kept in sterilized microisolator cages and received acidified autoclaved water (pH 2.5) after transplantation. All transplant recipients were between 8-12 weeks of age, and recipient age were matched between experimental groups. Experiments were performed using littermates where possible.

## METHOD DETAILS

### Stem cell transplantation

Detailed individual experimental schematics are described in the text and supplement. Delivery of total body irradiation was split between two fractions and performed using a Cesium-137 irradiator at 84cGy/min on day -1. Donor T cells were obtained by pooling splenocytes from up to six donor mice in any given experiment. CD4<sup>+</sup> T cells were purified using magnetic-activated cell sorting (MACS, Miltenyi Biotec). Between 0.5 to 5x10<sup>6</sup> donor CD4<sup>+</sup> T cells were given intravenously through the tail vein to each recipient depending on the system. Animal procedures were undertaken using protocols approved by the institutional (Fred Hutchinson Cancer Center) animal ethics committee. In female transplant systems receiving CD90.1<sup>+</sup> transgenic Marilyn CD4<sup>+</sup> T cells, between 0.2 to 0.5x10<sup>6</sup> Tg TCRs were introduced intravenously through the tail vein 1 day before planned sacrifice. The severity of systemic GVHD was assessed by scoring as previously described (maximum index = 10).<sup>63</sup> For survival experiments, transplanted mice were monitored daily and those with GVHD clinical scores  $\geq 6$  were sacrificed and the date of death registered as the next day, in accordance with institutional guidelines. Experimental schematics illustrated using BioRender ([biorender.com](http://biorender.com)).

### Probabilistic simulation of post-transplant TCR expansion

#### Definitions

We define a “twin transplant” system consisting of a single donor pool of TCRs distributed to two co-housed, genetically identical murine recipients. In each such system, bulk TCR $\beta$  amplicon sequencing is performed on a subset of donor pool TCRs, with sequenced pool size  $T_0$  and individual sequenced clonotypic frequencies  $C_0$ , where  $T_0 = \sum_1^i C_0$ , where  $i = \text{index of each unique clonotype in } C_0$ . Donor TCRs harvested at pre-defined timepoints from each recipient are denoted to have sequenced pool sizes of  $T_1$  and  $T_2$  (recipient 1 and 2, respectively), as well as clonotypic frequencies  $C_1$ , and  $C_2$ , where  $T_1 = \sum_1^j C_1$ , and  $T_2 = \sum_1^k C_2$ . For convenience of notation in all subsequent steps, we will assume  $C_0$ ,  $C_1$ , and  $C_2$  to represent the respective sums of all unique clonotypes (1,..., i) in each group. We further denote  $D_1$  and  $D_2$  as counts representing all TCRs donated to recipients 1 and 2, respectively (Figure S1A).

#### TCR Selection

In each system, let  $n = \text{index of all unique clonotypes}$  identified across  $C_0$ ,  $C_1$ , and  $C_2$ . TCR clonotypes are selected for further analysis if their sequenced frequency exceeds a pre-specified combined count threshold between the two recipients ( $C_1 + C_2$ )  $\geq 10$ , below which the probability of finding meaningful clone size differences was highly unlikely. This was determined *a priori* to analysis.

#### Generating probability distributions of individual donor clonotype frequencies

We define  $M = \text{merged donor pool}$  consisting of all TCRs in  $C_0$ ,  $D_1$ , and  $D_2$  ( $M = C_0 + D_1 + D_2$ ), and  $N = \text{individual clonotype frequencies}$  in the merged pool  $M$  (Figure S1B). In order to estimate the probability distribution of a range of frequencies for each individual clonotype,  $P(N|C_0)$ , we apply Bayes’ theorem:

$$P(N|C_0) = \frac{P(C_0|N)P(N)}{P(C_0)} \quad (\text{Equation 1})$$

$P(N)$  is the prior expectation of any given TCR clone size seen in a naïve (non-transplant) donor pool and is estimated from the exponential decay model as shown in Figure 1B for each transplant.  $P(C_0|N)$  represents the probability of sampling  $C_0$  counts for a particular TCR clonotype given  $N$  and is generated using the hypergeometric distribution probability mass function (PMF) to simulate sub-setting TCR pools without replacement:

$$P(C_0|N) = \frac{\binom{N}{C_0} \binom{M - N}{T_0 - C_0}}{\binom{M}{T_0}} \quad (\text{Equation 2})$$

Of note, the likelihood of  $N$  counts occurring decreases exponentially and is thus capped for computational efficiency when calculating  $P(C_0|N)$ , with an upper-bound  $L$  set at approximately 10 times the largest clone size identified in  $C_0$ . Finally,  $P(C_0)$  is calculated over all values of  $N$  considered,  $P(C_0) = \sum_{N=1}^L P(C_0|N) * P(N)$ , and represents a normalizing constant for each  $C_0$ .

### Defining null hypothesis for TCR expansion rates

For each given TCR clonotype in our twin transplant system, we define  $r$  being its overall expansion rate, with  $f_{\text{start}}$  and  $f_{\text{final}}$  being the fraction of TCR counts represented by that clonotype derived from sequencing the donor pool and both recipient pools, respectively (Figure S1C). Thus:

$$f_{\text{start}} = \frac{N - C_0}{D1 + D2}, f_{\text{final}} = \frac{C_1 + C_2}{T_1 + T_2}, \text{ and } r = f_{\text{final}} / f_{\text{start}} \quad (\text{Equation 3})$$

where  $\mathbf{C}$ ,  $\mathbf{D}$ ,  $\mathbf{T}$  are defined in (section A), and  $\mathbf{N}$  defined in Equation 1

We next define  $r1$  and  $r2$  as being the clonotypic expansion rate in each of recipients 1 and 2. The null hypothesis is subsequently defined as:

$$H_0 : r = r1 = r2 \text{ for all TCR clonotypes} \quad (\text{Equation 4})$$

This represents the default assumption that each clonotype expands at a similar rate within each transplant recipient.

### Modeling expected clonotype frequencies post-transplant

For each donor TCR clonotype found in the merged donor pool  $\mathbf{M}$ , we have derived a probability distribution  $P(N|C_0)$  of clone size frequencies as defined in (3) above. We next model final expected donor clonotype frequencies  $f1$  and  $f2$  found in recipients 1 and 2, respectively, while accounting for all possible ways each clonotype can be distributed between the two recipients (Figure S1D). To calculate this, we first assign  $\mathbf{A1}$  and  $\mathbf{A2}$  to represent all possible combinations of count partitions for each donor clonotype to recipient 1 and recipient 2, respectively, where the starting clone size is represented by clone frequency  $\mathbf{N}$  in the merged pool minus sequenced counts from the starting donor pool:

$$\mathbf{A1} + \mathbf{A2} = \mathbf{N} - \mathbf{C}_0 \text{ and} \quad (\text{Equation 5})$$

$$P(A1|N) = \binom{N - C_0}{A1} \left(\frac{1}{2}\right)^{A1} \left(\frac{1}{2}\right)^{N - C_0 - A1}$$

(binomial distribution PMF with probability of “success” set at 0.5)

Using the expansion rate  $r$  as defined by the null hypothesis (eq 4), expected donor clonotype frequencies are then defined as:

$$f1 = r \left( \frac{A1}{D1} \right) \text{ and } f2 = r \left( \frac{A2}{D2} \right) \quad (\text{Equation 6})$$

where  $\mathbf{D1}$  and  $\mathbf{D2}$  are defined in (section A)

### Calculating probability of observed clone size counts

For each TCR clonotype in our twin transplant system and for each possible split, we observe final clone size frequencies of  $C_1/T_1$  and  $C_2/T_2$  in recipients 1 and 2 as defined in (1). We arrange recipient 1 and 2 such that the clone size frequency in recipient 1 is less than that seen in recipient 2 (i.e.  $C_1/T_1 \leq C_2/T_2$ ). We next iterate over all possible A1 and A2 split as modeled in (5) and define  $P(C_1) \leq C_1$  observed representing the probability of seeing at most  $C_1$  copies of that TCR clonotype given the expected frequency  $f1$ . In choosing an appropriate probability density function to model our p value, we selected the negative binomial distribution over the poisson distribution as the former allows adjustment of model variance to account for overdispersion (e.g. noise in read counts) whereas the latter assumes variance equal to the mean,<sup>64,65</sup> which does not fit our empiric sequencing data showing an estimated model variance of approximately  $2^*\mu$ . Thus,

$$P(C_1|A1, N) = \binom{k+n-1}{n-1} (1-p)^k p^n \quad (\text{Equation 7})$$

where  $k = C_1$ ,  $n = \mu * p / (1-p)$ ,  $p = \mu / \sigma^2$ ,  $\mu = T_1 * f1$  as defined in (5), and  $\sigma^2 = 2^*\mu$

To obtain a p value,  $p'$ , that represents the likelihood of the observed clone count distribution between the two recipients, we multiple the probability of each starting clones size  $P(N)$  by the probability of each binomial split  $P(A1, A2)$ , and the probability  $P(C_1)$ :

$$p' = \sum_{N=1}^L \left( P(N) * \sum_{1}^N P(A1|N) P(C_1|A1, N) \right) \quad (\text{Equation 8})$$

where  $\mathbf{N}$  represents all modeled clone sizes for each clonotype as defined in (eq 1),  $P(A1, A2)$  is defined in (eq 5), and  $P(C_1)$  is defined in Equation 7

As we have arbitrarily selected recipient 1 as having the lower frequency TCR count (6), we multiply  $p'$  by 2 to obtain a 2-sided p value. Thus, we obtain final value of  $p = 2*p'$  for each TCR clonotype (Figure S1E). Model source code can be found at “<https://github.com/acyeh-lab/2023/tree/main/tcr-simulation>” along with accompanying data set in the Zenodo Repository “<https://doi.org/10.5281/zenodo.7406937>”.

### Antibodies and flow cytometry

The following antibodies were from BioLegend or BD Biosciences: V $\alpha$ 2 (B20.1), V $\beta$ 6 (RR4-7), V $\beta$ 8.3 (1B3.3), V $\beta$ 14 (14-2), CD3 (145-2C11), CD4 (RM4-5 or GK1.5), CD8 (53-6.7), CD45.1 (A20), CD45.2 (104), CD90.1 (HIS51), H-2Db (KH95), H-2Dd (34-2-12), I-A/I-E (M5/114.15.2), CD326/EPCAM (G8.8), IFN $\gamma$  (XMG1.2), IL10 (JESS-16E3), IL17A (TC11-18H10.1), CD80 (16-10A1), CD19 (6D5), NK1.1 (PK136), CD49b (HMa2), Ly6C (HK1.4), Ly6G (1A8), CD11b (M1/70), CD11c (N418), SiglecH (440c), CD64/Fc $\gamma$ RI (X54-5/7.1). 7AAD (Sigma) was added before cell acquisition. Cytokine staining was performed following 4-hour stimulation with phorbol myristate acetate (PMA, Sigma-Aldrich) 50ng/mL, ionomycin (Sigma-Aldrich) 1ug/mL, and Brefeldin A (Biolegend). Samples were acquired with BD FACSsymphony A3 and analysis was performed with FlowJo v10 (BD) software. All flow sorting experiments were performed on the Sony MA900 Multi-Application Cell Sorter or BD FACSsymphony S6. Sorted cells were collected in 5mL polypropylene round-bottom tubes that contained 1mL of RPMI with 10% FCS. Donor T cells were distinguished from recipient T cells by gating on CD45.1/CD45.2 and/or H-2Db/H-2Dd. CD3 $^+$ CD4 $^+$ CD8 $^{neg}$  donor T cells were collected for sequencing.

### MHC Class II tetramer analysis

I-A(b) tetramers were custom made by the National Institutes of Health (NIH) tetramer core facility. These include I-A(b) bound to candidatus arthromitus (SFB) DVQFSGAVPNKTD – PE (DVQ, SFBNYU\_003340 568-880)<sup>31</sup> and I-A(b) bound to bacterial flagellin to YSNANILSA – APC (YSN, CBir1 flagellin 464-472).<sup>34</sup> Control tetramer I-A(b) with peptide PVSKMRMATPLLMQA (CLIP peptide 87-101) was also provided by the NIH. Tetramer staining was performed using 1:500 dilution.

### Serum cytokine analysis

Serum cytokine analysis was performed using BD Biosciences Cytometric Bead Array kits and included interleukin-1 $\beta$  (IL-1 $\beta$ ), (IL-2), IL-4, IL-10, IL-12p40, IL-17A, interferon gamma (IFN $\gamma$ ) and tumor necrosis factor (TNF).

### Cell isolation from spleen and small intestine

Recipient spleen or small intestine samples were collected at designated timepoints. We sectioned the small intestine into thirds and collected the distal third (ileum) for downstream analysis. Splenocytes were obtained following 70 $\mu$ m filtration of mechanically ground spleen and treatment with Gey's solution for RBC lysis. Longitudinally sectioned pieces of the small intestine were processed using a gentleMACS Dissociator and mouse lamina propria dissociation kit (Miltenyi Biotec), according to the manufacturer's protocol. Dithiothreitol was excluded from the entire process. Lamina propria cells were isolated from the distal third of the small intestine (ileum SILP).

### Gut decontamination

For eradication of commensal microbiota, mice were allowed to drink *ad libitum* from a cocktail containing quadruple antibiotics, which consists of 1mg/mL of PO vancomycin, gentamycin, cefoxitin, and metronidazole,<sup>57,66</sup> starting 14 days before TBI. Antibiotics were changed twice a week, alternating between vancomycin/gentamycin and metronidazole/cefoxitin. Vancomycin/gentamycin was dissolved in sterile autoclaved water, and metronidazole/cefoxitin was dissolved in sterile 30% sucralose solution. Control mice received autoclaved water and 30% sucralose solution without antibiotics, replaced at the same intervals as the antibiotic group.

### TCR amplicon sequencing

Genomic DNA extraction from flow-sorted cells was performed using the QIAamp DNA Mini Kit (QIAGEN, Cat#51304) as described in the product protocol, and TCR sequencing was performed using the mouse TCRB assay (v3) by Adaptive Biotechnologies. We choose to use the Adaptive platform as it is the most reliable method for DNA-based TCR sequencing in terms of sensitivity and reproducibility.<sup>67</sup>

### Publicly available TCR amplicon sequencing dataset

Publicly available dataset used for Figure 1D can be found on Adaptive Biotechnologies immuneACCESS platform (J. Gil, TCRB technical replicates of PBMC from four donors, <https://doi.org/10.21417/ADPT2017TR>, last accessed June 23, 2023). Sequencing results were accessed for donor 1 and donor 2 by combining 42 deep sequencing data sets used for each donor.

### PCR and 16S rRNA sequencing

Fecal DNA extraction was performed using QIAamp Fast DNA Stool Mini Kit (Qiagen, #51604). Purified DNA was sent to Zymo Research for 16S rRNA sequencing. PCR primers were purchase from IDT (Integrated DNA Technologies) and were used to detect CBir1 flagella: (F) 5'-GCT GAC ACA GGA AAT CGA TCG T-3' and (R) 5'GAG AGT ATA CAT CAC CCG TCG CAT-3', Tm 65°C (Provided by Dr. Charles Elson, UAB); and 16S RNA<sup>60</sup>: BAC 1114 (F) 5'-CGG CAA CGA GCG CAA CCC -3' and BAC 1275 (R) 5'-CCA TTG TAG CAC GTG TGT AGC C-3', Tm 58°C, amplicon size 130bp.

### Single cell RNA sequencing

Naïve CD45.1 $^+$  C57BL/6 mice from Taconic Biosciences were transplanted with 5x10 $^6$  bone marrow cells derived from CD45.1 $^+$  C57BL/6 mice and with either 5x10 $^5$  CD45.2 $^+$ CD90.1 $^+$  Marilyn Tg CD4 $^+$  T cells (group 1), 1x10 $^6$  CD45.2 $^+$ CD90.2 $^+$  CBir1 Tg

CD4<sup>+</sup> T cells (group 3), or both Tg T cells (group 2). Spleens were harvested at D+6 post-transplant (three mice pooled for group 1, four for group 2, and three for group 3). T cells from each mouse were incubated with APC-labeled CBir1 tetramer (NIH, above), washed, and labeled with TotalSeq C hashtag antibodies (C0301-C0310) along with TotalSeq C0987 (anti-AP). Following staining, T cells were sort-purified (Marilyn: CD45.1<sup>neg</sup>CD45.2<sup>+</sup>CD90.1<sup>+</sup>CD90.2<sup>neg</sup>CD4<sup>+</sup>CD8<sup>neg</sup>; CBir1: CD45.1<sup>neg</sup>CD45.2<sup>+</sup>CD90.1<sup>neg</sup>CD90.2<sup>+</sup>CD4<sup>+</sup>CD8<sup>neg</sup>) at a 4:1 ratio for cell capture and library construction according to the 10x Genomics Chromium User Guide CG000424 (CG000424\_ChromiumNextGEMSingleCell5\_HTv2\_CellSurfaceProtein\_UserGuide\_RevD.pdf). Libraries were sequenced using Illumina Nextseq 2000 targeting a depth of 25,000 reads per cell for the 5' gene expression library, 5,000 reads/cell for the V(D)J library, and 5,000 reads/cell for the cell surface protein library.

Reads were demultiplexed and processed using cellranger-arc v1.0.1 aligning reads to GENCODE vM23/Ensembl98. Cells meeting the following criteria were retained for downstream analysis:  $3.25 < \log_{10}(n\text{Count\_RNA}) < 4.25$  & percent.mt < 5%. Dimensionality reduction using principal components analysis (PCA) was performed and clusters were identified using the standard Seurat workflow.<sup>62</sup> We used the CLR method to normalize HTO data and used HTODemux to ID experimental groups. Two clusters enriched in hemoglobin and myeloid genes were removed, as well as one cluster expressing Trbv29 (non-Marilyn, non-CBir1 TCR), prior to generating the final embedding for analysis. Gene set enrichment analysis (GSEA) using the Hallmark gene set form MSigDB (<https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp>) was performed using the R package fgsea (<https://github.com/ctlab/fgsea>). Source code can be found at <https://github.com/acyeh-lab/2023/tree/main/scseq>.

### Generation of TCR hybridomas

The NFAT-GFP 58 $\alpha$ - $\beta$ <sup>-</sup> hybridoma cell line was provided by K. Murphy.<sup>39</sup> To reconstruct TCRs, we used a self-cleavage sequence of 2A to link cDNAs of TCR $\alpha$  and TCR $\beta$  generated from annealing oligonucleotides (TCR $\alpha$ -p2A-TCR $\beta$ ). This cassette was transferred into a MP71 retroviral vector. Sequence confirmation was performed using forward primer pF TS41 (TCTTGCTGCTGCAGCATCGT) and reverse primer pR TS42 (TCCCATAAAATGAAATGATTGCCACC). The retroviral vectors were transfected into packaging PlatE cells<sup>61</sup> using Xfect Transfection Reagent (Takara Biosciences). The full sequence for 7B8 SFB Tg TCR was generated by RNA sequencing of the TCR $\alpha$ / $\beta$  V(D)J region (SMARTer Mouse TCR a/b Profiling Kit, Takara Biosciences) from the corresponding 7B8 SFB Tg mouse. Libraries constructs were sequenced on an Illumina MiSeq v3 with 300bp paired-end reads. The final codon-optimized cassette was synthesized via Twist Biosciences (pTwist Amp High Copy) with additional 5'-NotI and 3'-EcoRI sites. A similar strategy was used to transfer an mCD4 cassette derived from pBluescript (Addgene #14613) into MP71 retroviral vector.

### Live cell imaging

All live imaging was performed using the xCELLigence Real-Time Cell Analyzer (Agilent) at 37°C and 5% CO<sub>2</sub>. 96-well plates were coated with 12.5 $\mu$ g/mL retronec tin, incubated at 4°C overnight, and aspirated prior to cell plating. Hybridomas were plated at a density of 50k/well and centrifuged at 800g for 1 minute to facilitate adherence. Images were taken at 60-minute intervals in both bright-field and GFP channel (500ms exposure). Analysis was performed using the xCELLigenace RTCA software package.

### RNA in-situ hybridization (ISH) imaging

Formalin-fixed paraffin-embedded tissues were sectioned at 4 microns onto positive-charged slides and baked for 1 hour at 60°C. The slides were loaded to the Leica Bond Rx Autostainer platform (Leica, Buffalo Grove, IL) to start the run. The slides were baked and dewaxed using Leica Bond reagents for dewaxing (Dewax Solution). Antigen retrieval was performed at 95°C for 15 minutes using Leica Epitope Retrieval Solution 2 followed with 0.5% Triton-X/PBS solution pretreatment at 40°C for 60 minutes. After the pretreatment steps, the slides were exposed to hydrogen peroxide for 10 minutes and then incubated with ACD Target probe, BA-Mm-Tra-V4J33-CDR3-C1 (ACD #1210058-C1, sequence GCTGTGTACTACTGTGCTCACAGGAAACTACAAATACGTC TTTGGAG, CBir1 Tg CDR3 region), Positive control probe BA-Mm-Ppib-1zz (ACD #712358) or Negative control probe BA-DapB-zz (ACD #701028) at 42°C for 120 minutes. After the probe incubation, the staining continues with the Basescope LS Reagent Kit (ACD #323600) for the amplification and detection steps, followed by use of the Bond Polymer Refine Red detection kit (Leica #DS9390). To perform the CD3 immunofluorescence staining, protein blocking with TCT buffer (0.05M Tris, 0.15M NaCl, 0.25% Casein, 0.1% Tween 20, pH 7.6 +/- 0.1) for 10 minutes was used. The CD3 (Abcam #ab215212) primary antibody at 1:200 was applied for 60 minutes followed by the OPAL Polymer Anti-Rabbit HRP (Akoya Biosciences #ARR1001KT) for 20 minutes. The tertiary TSA-amplification reagent Opal fluor 690 (Akoya Biosciences #OP-001006) was added at 1:100 in Akoya 1x Plus Automation Amplification Diluent (Akoya #FP1609) for 20 minutes. Slides were removed from the Bond Autostainer and stained with DAPI (10 $\mu$ g/ml distilled water) for 5 minutes, rinsed in water, and coverslipped using Prolong Gold Antifade reagent (Invitrogen #P36930). The slides were cured at room temperature and the images of the slides were acquired on the Vectra Polaris Imaging System.

Images were analyzed with HALO image analysis software v3.5 (Indica Labs, Cooles, NM) using the FISH-IF module, v2.2.5. Cellular analysis of the images was performed by first identifying cells based on nuclear recognition (DAPI stain), then measuring fluorescence intensity of the estimated cytoplasmic areas of each cell. A mean intensity threshold above background was used to determine positivity for each fluorochrome within a cytoplasm radius, thereby defining cells as either positive or negative for each marker. The positive cell data was then used to define colocalized populations. A random forest classifier was incorporated to define areas of interest and exclude the intestinal lumen and mesentery.

## QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis was performed using R statistical software (v3.5.3) or GraphPad Prism (v 9.4.0). To account for the TCR repertoire sparseness, TCR beta diversity metrics were calculated using the Jensen-Shannon divergence between repertoire pairs as applied via Repertoire Comparison in Low Dimensions (RECOLD).<sup>21</sup> Analysis was run via default parameters (manifold learning using tSNE). Results were illustrated via principal coordinate using the “stats” R package. Fecal 16S beta diversity metrics were calculated using the Bray-Curtis index and provided by Zymo Research. *P*-values for beta diversity metrics were calculated via PERMANOVA using the “vegan” R package.<sup>22</sup> Antigen-specific Lymphocyte Identification by Clustering of Expanded Sequences (ALICE) was performed as described in Pogorelyy et al.,<sup>24</sup> using publicly available code (<https://github.com/pogorely/ALICE>, last accessed 1/16/2024). Each analysis was run over 100 iterations consisting of  $1 \times 10^6$  sequences/iteration for generative probability estimation simulation for each VJ class, as recommended in the original manuscript. Grouping of Lymphocyte Interactions by Paratope Hotspots (GLIPH2) was performed as described in Huang et al.<sup>25</sup> using the publicly available web portal (<http://50.255.35.37:8080/>, last accessed 1/16/2024, reference file mouse v1.0). We focused on specificity groups that contained at least 3 unique TCRs and exhibited significant V-gene bias ( $p < 0.05$ ), cutoffs suggested in the original manuscript. CDR3 amino acid logo motifs were generated using the “ggseqlogo” R package. Calculation of selective expanders is described in detail above and also in the main text. Comparison of selective expander frequencies across groups were performed using the Mann-Whitney U test. Survival analysis was performed using log-rank test.