blood advances

Organ-specific response after low-dose interleukin-2 therapy for steroid-refractory chronic graft-versus-host disease

Haesook T. Kim,[1] John Koreth,[2] Jennifer Whangbo,[2] Sarah Nikiforow,[2] Carol G. Reynolds,[2] Peter Stowe,[2] Vincent T. Ho,[2] Corey Cutler,[2] Joseph H. Antin,[2] Robert J. Soiffer,[2] and Jerome Ritz[2]

[1]Department of Data Science, Dana-Farber Cancer Institute, Harvard School of Public Health, Boston, MA; and [2]Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA

Key Points

- For 105 patients enrolled in 5 trials, the overall response rate after 8 or 12 weeks' LD IL-2 was 48.6% and 53.3% with continued therapy.

- Skin involvement was the most frequent (84%); the organ-specific response rate was highest in liver (66.7%) and lowest in lung (19.2%).

Despite new therapeutic options, treatment of steroid-refractory chronic graft-versus-host disease (SR-cGVHD) remains challenging as organ involvement and clinical manifestations are highly variable. In previous trials of low-dose interleukin-2 (LD IL-2), we established the safety and efficacy of LD IL-2 for the treatment of SR-cGVHD. In the present report, we combined five phase 1 or 2 clinical trials conducted at our center to investigate organ-specific response rate, coinvolvement of organs, predictors of organ-specific response, and its possible association with immune response. For the 105 adult patients included in this report, the overall response rate after 8 or 12 weeks of LD IL-2 was 48.6% and 53.3%, including late responses in patients who continued treatment for extended periods. Skin was the most frequent organ involved (84%). The organ-specific response rate was highest in liver (66.7%) followed by the gastrointestinal tract (62.5%), skin (36.4%), joint/muscle/fascia (34.2%), and lung (19.2%). In multivariable analysis, shorter time from diagnosis of cGVHD to IL-2 initiation, shorter time from transplant to IL-2 initiation, and fewer prior therapies were associated with overall response as well as skin response. For immunologic correlates, the ratio of regulatory T cells:conventional T cells (ie, CD4Treg:CD4Tcon) ratio at 1 week was significantly higher in patients with overall and skin response; skin response was significantly associated with lower number of total CD3 T cells, CD4Tcon cells, and CD8 T cells and a higher number of B cells. For lung responders, terminal effector memory cell counts were lower within all T-cell populations compared with nonresponders. Organ-specific mechanisms of injury should be investigated, and organ-specific targeted therapies need to be developed.

Introduction

Chronic graft-versus-host disease (cGVHD) is a major cause of late morbidity and impaired quality of life for many survivors after allogeneic hematopoietic cell transplantation (alloHCT). cGVHD has a wide spectrum of clinical presentations affecting multiple target tissues, prominently including the skin, mouth, eyes, gastrointestinal (GI) tract, liver, lung, joints, musculoskeletal, and fasciae. cGVHD invariably results from impaired immune tolerance to recipient tissues after alloHCT. The mechanisms leading to diverse clinical manifestations are complex, however, and involve multiple phases and multiple immune cell types, including coordinated immune responses between T and B cells.[1-7]

The full-text version of this article contains a data supplement.

Of multiple mechanisms involved in developing cGVHD, we previously described impaired reconstitution of regulatory T cells (CD4Treg) in patients with cGVHD and hypothesized that preferential augmentation of CD4Treg may enhance immune tolerance resulting in control of cGVHD.[7,8] Interleukin-2 (IL-2) is the primary homeostatic regulator of CD4Treg differentiation, proliferation, and survival, and we therefore conducted multiple phase 1 or 2 clinical trials evaluating the ability of low-dose (LD) IL-2 to enhance CD4Treg function in patients with steroid-refractory cGVHD (SR-cGVHD).[9-13] These clinical trials showed that daily subcutaneous IL-2 administered at a dose of 1×10^6 IU/m^2 per day is safe and well tolerated for prolonged periods. Importantly, this regimen consistently induced preferential Treg expansion for the duration of IL-2 therapy and produced an objective clinical response in 50% to 60% of adults with SR-cGVHD.[9-12] In these trials, clinical response was assessed based on the global cGVHD score, and the size of each study was limited and insufficient to evaluate organ-specific response and its association with immune reconstitution. In the current study, we combined all 5 clinical trials conducted at our institute to evaluate organ-specific response rates, coinvolvement of target organs, predictors of organ-specific response, and possible association with immune reconstitution. Because cGVHD is a complex disease with highly variable clinical manifestations, injury to specific tissues and organs may be mediated by distinct as well as common immunologic pathways, and therapeutic agents may selectively affect different sites of disease.

Methods

Clinical trials

Five single-center phase 1 or 2 clinical trials conducted at the Dana-Farber Cancer Institute from 2008 through 2017 were aggregated (#NCT00529035, #NCT01366092, #NCT01937468, #NCT02318082, and #NCT02340676). Each study investigated the safety and efficacy of LD IL-2 therapy for the treatment of SR-cGVHD. The initial treatment period was 12 weeks for #NCT01366092 and 8 weeks for the other 4 trials. #NCT01937468 and #NCT02340676 combined LD IL-2 with donor Treg infusion and extracorporeal photopheresis (ECP), respectively. Upon completion of the initial treatment period, patients with clinical benefit (complete response/partial response [PR], stable disease with minor response) could continue daily IL-2 indefinitely. Detailed information of these trials is presented in supplemental Table 1.

Clinical assessments

cGVHD assessments using either 2005 or 2014 (depending on each trial) National Institutes of Health consensus criteria[14-17] were undertaken at baseline, at completion of either 8 or 12 weeks of IL-2 therapy, every 12 to 16 weeks during extended therapy, and after discontinuation of IL-2. Response definitions are provided in the supplemental Materials.

Flow cytometric analysis of lymphocyte subsets

Phenotypic analyses of lymphocyte subsets were performed at baseline; 1, 2, 3, 4, 6, and 8 weeks after starting IL-2 therapy; and every 3 to 6 months while receiving extended-duration IL-2. Detailed methods are provided in the supplemental Materials.

Statistical analyses

Logistic and Cox regression analyses were performed to investigate clinical factors that were associated with response and overall survival (OS), respectively. Immunologic parameters were analyzed primarily descriptively and compared by using the Wilcoxon rank sum test for group comparison. Principal component analysis (PCA) was performed to reduce dimension for highly correlated immune parameters. More detailed information is provided in the supplemental Materials.

Results

Patient characteristics

A total of 123 adult patients with SR-cGVHD were enrolled on five IL-2 clinical trials from 2008 through 2017. Fourteen patients were not evaluable for response and thus excluded from the study. Four patients were enrolled on multiple trials, and only the first enrollment for these patients was included. Baseline clinical and transplant characteristics of the 105 patients included in this study are presented in Table 1. The median age at study enrollment was 54 years (range, 22-76 years), 61% were male, median time from cGVHD onset to study enrollment was 1.7 years (range, 0.1-11.8 years), median number of prior therapies was 3 (range, 1-9), and median number of sites involved was 4 (range, 1-7). Forty-seven (45%) patients had prior acute GVHD. The median cGVHD global score at baseline was 6 (range, 2-10); 8.6% had mild, 60% moderate, and 31.4% severe cGVHD.

Organ-specific response

Overall response assessed at the end of either 8 or 12 weeks of LD IL-2 therapy (EOT) was PR 48.6%, stable disease 43.8% (21.9% with minor response and 21.9% with mixed response), and progressive disease 7.6% (Figure 1). The median global score was 5 (range, 0-9) at EOT, and 64% of patients had at least 1 score improvement. Fifty-nine patients (56.2%) with clinical benefit elected to continue extended-duration IL-2 therapy, and the majority of these patients (51%) exhibited continued improvement at various sites with decreasing global score during extended therapy. For all patients who received extended therapy, the median global score was reduced to 4 (range, 0-8). One patient with PR achieved complete response, and 5 additional patients achieved PR, resulting in an overall response rate of 53.3%.

Overall, skin was the most frequently involved site (84%) followed by joint/muscle/fascia (JMF; 75%), eyes (64%), mouth (50%), lung (45%), liver (17%), GI tract (15%), and genital tract (5%). Of 79 patients with JMF involvement, 77 had coinvolvement with skin, and JMF was inversely correlated with liver involvement ($r = -0.38$) (Figure 1B). Five organ sites were evaluated for response: skin, JMF, lung, GI tract, and liver. Mouth and eyes were also assessed; however, because protocols permitted concurrent topical treatments, responses to these sites were not included. Organ-specific response rate was the highest in liver (66.7%) and lowest in lung (19.2%). Response rate was 36.4% for skin, 34.2% for JMF, and 62.5% for GI tract (Figure 1C). Of 77 patients who had both skin and JMF involvement, 34 (44%) responded in skin and/or JMF: 19 responded in both sites, 8 in skin only, and 7 in JMF only. Of 11 patients who had both skin and liver involvement, 4 responded at both sites, and

Table 1. Baseline characteristics

Characteristic	Value
Total	105 (100%)
Age at enrollment, y	54 (22,76)
Time from HSCT to study entry, y	2.7 (0.5, 12.1)
Time from cGVHD to study entry, y	1.7 (0.1, 11.8)
No. of prior treatments	3 (1, 9)
No. of sites involved	4 (1, 7)
Global score at enrollment	6 (2, 10)
Sex	
Female	41 (39%)
Male	64 (61%)
ECOG PS at study enrollment	
0	6 (5.7%)
1	69 (65.7%)
2	28 (26.7%)
UNK	2 (1.9%)
HLA typing (A,B,C, DRB1)	
Matched, related	32 (30.5%)
Matched, unrelated	63 (60%)
Mismatch, unrelated	10 (9.5%)
Progenitor cell source	
BM	5 (4.8%)
BM and PBSC	1 (1%)
PBSC	98 (93.3%)
UNK	1 (1%)
Primary disease	
ALL	8 (7.6%)
AML	31 (29.5%)
CLL/SLL/PLL	12 (11.4%)
CML	5 (4.8%)
Hodgkin disease	2 (1.9%)
MDS	20 (19%)
MPD	4 (3.8%)
Mixed MDS/MPD	1 (1%)
Multiple myeloma	2 (1.9%)
Non–Hodgkin lymphoma	18 (17.1%)
Other acute leukemia	1 (1%)
Other	1 (1%)
Conditioning intensity	
Myeloablative	55 (52.4%)
Non-myeloablative	49 (46.7%)
UNK	1 (1%)
Prior acute GVHD	
None	58 (55.2%)
I	15 (14.3%)
II	24 (22.9%)
III	8 (7.6%)

Table 1. (continued)

Characteristic	Value
cGVHD severity at study enrollment	
Mild	9 (8.6%)
Moderate	63 (60%)
Severe	33 (31.4%)

Data are presented as median (range) for continuous variables and frequency (%) for categorical variables.

AML, acute myeloid leukemia; BM, bone marrow; CLL/SLL/PLL, chronic lymphocytic leukemia/small lymphocytic lymphoma/prolymphocytic leukemia; ECOG PS, Eastern Cooperative Oncology Group performance status; HSCT, hematopoietic stem cell transplantation; MDS, myelodysplastic syndrome; MPD, myeloproliferative disorder; PBSC, peripheral blood stem cell; UNK, unknown.

4 responded in liver only. Overall, 1 patient responded at 4 sites, 5 at 3 sites, 22 at 2 sites, and 29 responded at 1 site. As expected, overall response encompasses almost all site-specific responses except a few mixed site responses (Figure 1D).

In multivariable logistic regression analysis, shorter time to initiation of therapy (odds ratio [OR], 4.12 for <2.5 vs ≥2.5 years from alloHCT [P = .0025]; OR, 4.6 for <1.6 vs ≥1.6 years from cGVHD onset [P = .0015]) and fewer prior therapies (OR, 5.29 for <4 vs ≥4; P = .002) were associated with overall response. Similarly, shorter time from alloHCT (OR, 3.8; P = .01) and from cGVHD (OR, 2.8; P = .044), and fewer number of prior therapies (OR, 5.06; P = .01), were also associated with skin response (Table 2). Other baseline factors, including cGVHD severity at enrollment, were not associated with overall or skin response. When responses to difficult-to-treat organs (ie, lung, liver, GI tract) were combined, number of prior therapies was associated borderline with response (OR, 3.45; P = .056).

cGVHD symptom scores

Four of the 5 trials collected patient-reported cGVHD symptom scores (ie, Lee symptom scale[18]); the first trial (#NCT00529035) did not have this information collected. Paired pretreatment to posttreatment analysis was performed for each of the 7 subscales of these cGVHD symptom scores. For the entire cohort, skin and muscle symptom scores showed a significant improvement (median score from 5 to 3 for skin [P = .001]; median score from 6 to 4.5 for muscle [P = .015]), and the sum of all scores decreased significantly from a median of 28.7 at baseline (interquartile range, 18-42) to 21 (interquartile range, 13-33; P = .0025) at EOT (supplemental Figure 1). The response rate on the Lee symptom scale (defined as a ≥7 point reduction from baseline in total symptom score on the scale, with higher scores indicating worse symptoms) was therefore 40%. Improvement was noted regardless of organ-specific or overall response status. In addition, 39% of patients experienced reduction in the steroid dose at EOT, and 79% and 86% of patients who were on the extended follow-up experienced steroid dose reduction at 6 months and 1 year, respectively. Summary of prednisone dose and percent change at each time point and frequencies of all concurrent therapies are presented in supplemental Table 2. In our initial studies, concurrent immune suppressive therapies were
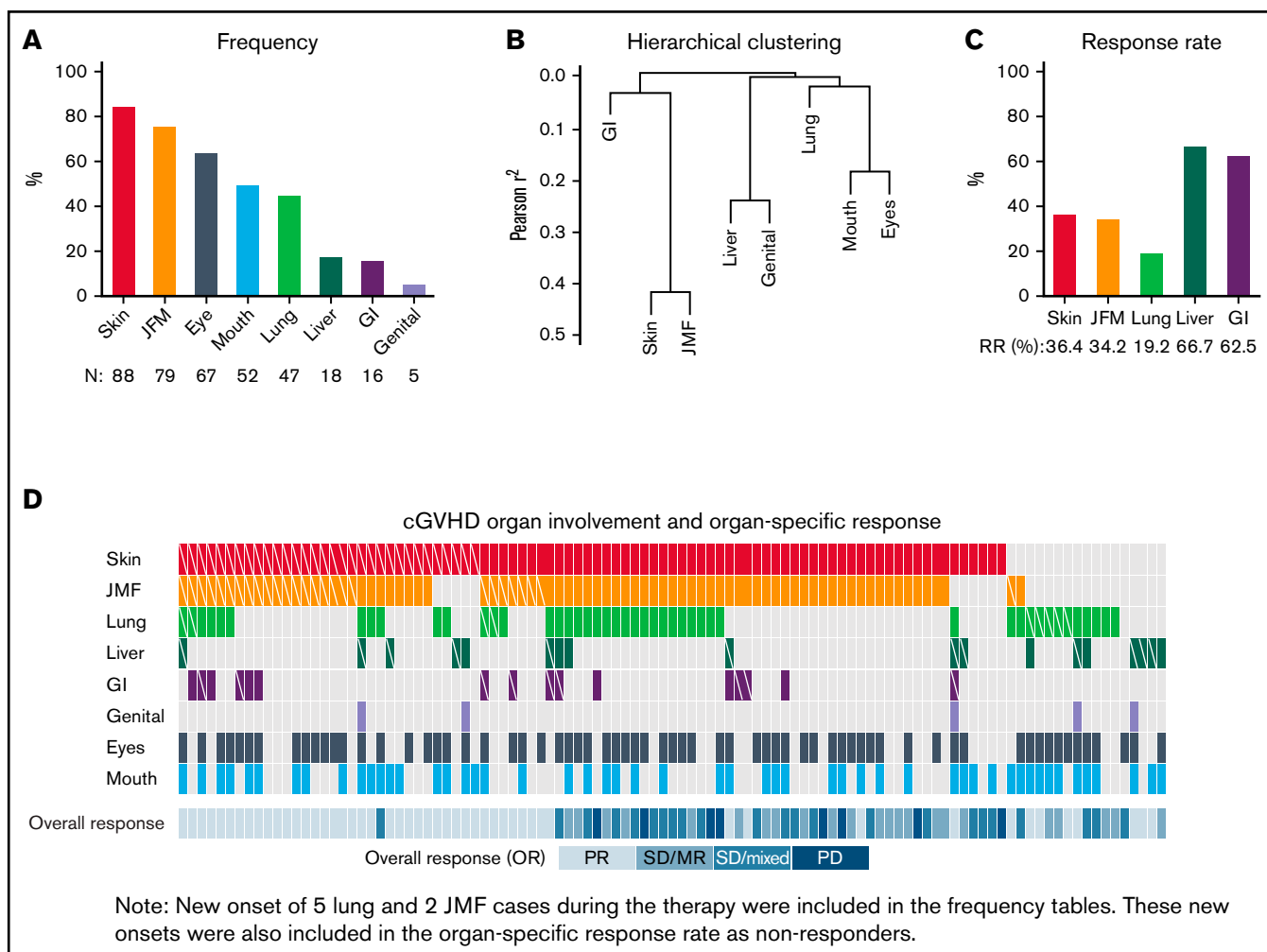
Figure 1. Organs involved in cGVHD. (A) Frequency of cGVHD organ involvement. (B) Hierarchical clustering of involved sites. (C) Organ-specific response rate. (D) Heatmap for organ involvement and organ-specific response for the entire cohort. Skin (red), JFM (orange), lung (green), liver (dark green), GI tract (purple), genital tract (light purple), eyes (dark gray), and mouth (blue) involvement. White diagonal line in each box indicates organ-specific responder. CR, complete response; Mixed, mixed response; MR, minor response; N, number of each site involved; PD, progressive disease; PR, partial response; RR, responder rate; SD, stable disease.

maintained at pretreatment levels when LD IL-2 was started to allow for accurate assessment of toxicities and efficacy associated with LD IL-2.

Survival outcomes

For the entire cohort, median follow-up among survivors was 69 months (range, 2-145 months) with 41 deaths; 19 (46%) died of GVHD, 5 of disease recurrence, and 5 of new malignancies (supplemental Table 3). One of these patients died of complications of disease recurrence and COVID-19. Five-year OS and progression-free survival (PFS) from study entry was 63% (95% confidence interval, 52-72) for both (supplemental Figure 2A). Five-year OS from stem cell infusion was 86% (95% confidence interval, 78-92). Neither specific organ involvement nor organ-specific response affected OS or PFS. Five-year OS for patients with lung/GI tract/liver involvement was 62% vs 64% (P = .95) for those without lung/GI tract/liver (supplemental Table 3; supplemental Figure 2B). Univariable and multivariable analyses were performed to identify risk factors for OS. Of all factors considered, only age was

significantly associated with OS (hazard ratio, 2.4 for ≥50 years vs <50 years; P = .028) (supplemental Figure 2C; supplemental Table 3).

Immunologic and organ-specific response

Daily LD IL-2 therapy led to a consistent increase in CD4Treg that peaked 2 to 4 weeks after starting treatment. CD4Treg counts subsequently declined but remained elevated above baseline for the entire duration of IL-2 therapy in both responders and nonresponders. Natural killer (NK) cell counts also increased during therapy and remained elevated after therapy (Figure 2A-B). There were no significant differences in absolute CD3 or CD8 T-cell counts between responders and nonresponders, but CD4 conventional T-cell (CD4Tcon) counts were significantly lower at 1 week and NK cell counts were significantly lower at 1 and 2 weeks in responders (Figure 2B; supplemental Figure 3). At 1 week, percent CD4Treg was higher (median, 19% vs 12%; P = .013) and %CD4Tcon was lower (median, 81% vs 88%; P = .012) in the response group. As reported previously,[9,10] the CD4Treg:CD4Tcon ratio was

Table 2. Multivariable logistic regression analysis for overall and organ-specific cGVHD response

| Response to | Contrast | | Multivariable logistic | |
		OR	95% CI	P
Overall response	Years from transplant to study	<2.5 vs ≥2.5	4.12	1.64-10.3	.0025
	No. of prior therapies	<4 vs ≥4	5.29	1.83-15.3	.002
	Years from cGVHD to study	<1.6 vs ≥1.6	4.60	1.79-11.8	.0015
Skin	Years from transplant to study	<2.5 vs ≥2.5	3.80	1.38-10.4	.0096
	No. of prior therapies	<4 vs ≥4	5.06	1.44-17.7	.011
	Years from cGVHD to study	<1.6 vs ≥1.6	2.8	1.03-7.6	.044
JMF	No. of prior therapies	<4 vs ≥4	2.63	0.86-8.02	.089
Lung/GI tract/liver	No. of prior therapies	<4 vs ≥4	3.45	0.97-12.3	.056

significantly higher in responders compared with nonresponders at 1 week (Figure 2C). Using previously defined CD4Treg:CD4Tcon threshold values, 40% of nonresponders and 70% of responders had a ratio >0.06 (median value) at baseline (P = .004), and 40.4% and 75% of nonresponders and responders, respectively, had a ratio ≥0.2 at 1 week (P = .0004) (Figure 2D); these findings indicate that the baseline and week 1 ratios are highly predictive of the overall response. Within CD4Treg, percent naive cells was significantly lower (P = .0004) and percent effector memory (EM) cells was higher in the response group at 1 week (median, 44% vs 28%; P = .06), and percent terminal effector memory (TEMRA) was significantly lower after week 8 through 1 year (Figure 2E). Also, percent EM CD8 T cells was significantly higher at various time points in the response group (supplemental Figure 3E).

Absolute immune cell counts according to the skin response are presented in Figure 3A and supplemental Figure 4. Absolute lymphocyte (supplemental Figure 4A), CD3$^+$, CD4Tcon, and CD8$^+$ T-cell counts were lower in skin responders up to 9 months after IL-2 therapy, whereas B-cell (CD19$^+$) counts were consistently higher, although the difference was not statistically significant (supplemental Figure 4B). Consistent with absolute cell counts, percent B cells was higher in the response group throughout 1-year follow-up, particularly 1 week after starting IL-2 therapy (median 12% vs 4%; P = .001) (supplemental Figure 4D). As in overall response, CD4Treg and NK cell expansion was similar in skin responders and nonresponders (Figure 3; supplemental Figure S4C). Major T-cell populations were highly correlated with total CD3 T cells throughout the study period but not with B cells (Figure 3B). We thus performed PCA, including CD3, CD4Treg, CD4Tcon, CD8, and CD19$^+$. The correlation among these major populations is also shown in component pattern analysis from the PCA. All T-cell populations were clustered at a high first principal component (PC1) score, whereas CD19$^+$ was located perpendicularly at a high second component (PC2) score. In the PCA, PC1 had high and even loadings of T-cell populations, whereas PC2 was almost entirely explained by CD19$^+$. As such, PC1 was lower (0.06, 0.04, 0.02 at week 0, 1, 2, and 4, respectively; P = .07) and PC2 was higher (P = .0002 at week 1) in skin responders compared with nonresponders after IL-2 therapy (Figure 3C), which is consistent with the absolute counts shown in Figure 3A.

As in the overall response, percent CD4Treg was higher (P = .03), percent CD4Tcon was lower (P = .03), and the CD4Treg:CD4Tcon ratio was higher at 1 week in skin responders (P = .02) (Figure 4A).

Although the CD4Treg:CD4Tcon ratio increased significantly from baseline in both groups, the increase from baseline in responders was significantly higher compared with nonresponders (P = .0095). Furthermore, a higher proportion of skin responders had CD4Treg:CD4Tcon ratio values above the threshold at baseline and at 1 week (40% vs 75% at baseline, P = .0046; 32.7% and 65.2% at week 1, P = .01, in nonresponders and responders, respectively). A similar observation was noted in the CD4Treg:CD8 ratio. Both skin responders and nonresponders had a significant increase in the CD4Treg:CD8 ratio at 1 week after IL-2 therapy (P < .001), but the increase in responders was significantly higher compared with that in nonresponders (P = .02). This observation is different from the overall response as CD8 cell counts are lower in skin responders than in nonresponders.

Patients with JMF largely overlapped with skin, and JMF responders showed similar results in absolute T and B cells and the CD4Treg:CD4Tcon ratio, which were similar in JMF responders and nonresponders. Still, a higher proportion of JMF responders had CD4Treg:CD4Tcon ratio values above the threshold at baseline and week 1 (44.7% vs 75% at baseline; P = .023; 32.6% vs 65% at week 1 [P = .028] in nonresponders and responders, respectively). In addition, the CD4Treg:CD8 ratio was significantly higher in JMF responders at weeks 1 and 2, and the CD4Treg:NK ratio was significantly higher at week 2 after starting LD IL-2 (Figure 4B).

Because there is no correlation between skin and lung involvement, and the response between these 2 sites is largely non-overlapping, we also examined immunologic response to LD IL-2 in patients with lung cGVHD. Overall, major T-cell populations showed a distinct pattern between lung responders and nonresponders. Absolute T-cell counts were lower in lung responders starting 4 weeks after LD IL2 although the difference was significant only at a few time points (supplemental Figure 5) owing to the limited number of patients with lung response. Absolute NK cell counts were lower in lung responders and significantly lower at week 8. There was no significant difference in CD4Treg:CD4Tcon ratio. Within CD4Treg, CD4Tcon, and CD8 T-cell populations, we also examined the fraction of cells that exhibited different levels of maturation: naive, central memory (CM), EM, and TEMRA. Consistent with the pattern seen in the major T-cell population, the TEMRA subset was significantly lower in lung responders compared with nonresponders starting week 4 in major T cell populations (Figure 5A). Furthermore, for lung responders, percent TEMRA CD4Treg cells was significantly lower at 8 weeks, 3 and 6 months, and 1 year after IL-2 therapy
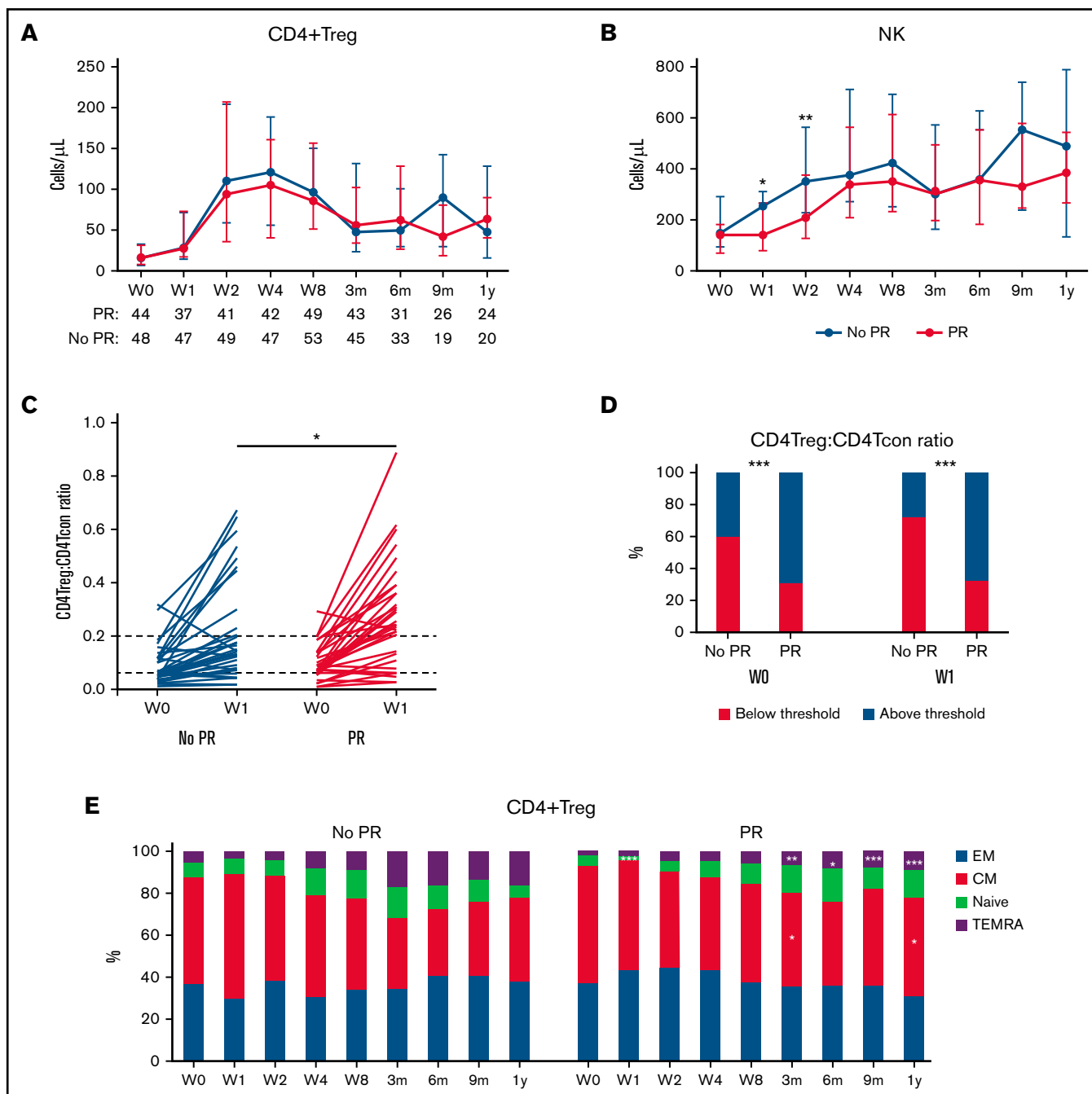
Figure 2. Immune reconstitution according to the overall response. (A) Median absolute CD4$^+$Treg cell counts with the interquartile range. (B) Median absolute NK cell counts with the interquartile range. PR (red) and no PR (blue). (C) CD4Treg:CD4Tcon ratio at week 0 (W0) and week 1 (W1). Each line indicates individual patient's data. Dotted lines indicate threshold values of the ratio at W0 and W1 (0.06 and 0.2, respectively). (D) Proportion of patients with CD4Treg:CD4Tcon ratio above and below the threshold at W0 and W1. (E) Percent CD4Treg subset populations. Median value of each subset. EM (blue), CM (red), naive (green), and TEMRA (purple). Number of subjects (N) at each time point is the same in every figure. *$P < .05$, **$P < .01$. ***$P < .005$. m, month; NR, nonresponder; R, responder.

(0.0056, 0.0094, and 0.03, respectively; $P = .01$). Percent CM CD4Treg cells (Figure 5B) was significantly higher at baseline, 4 and 8 weeks, 3 and 6 months, and 1 year after starting IL-2 therapy (0.035, 0.03, 0.004, 0.0047, and 0.014; $P = .02$) and percent naive CD4Treg was lower, particularly at 2 weeks after IL-2 therapy ($P = .04$). Within CD4Tcon, percent TEMRA CD4Tcon was significantly lower in lung responders at 6 and 9 months after LD IL-2

therapy ($P = .018$ and .039) and percent naive CD4Tcon was lower during therapy but was not statistically significant.

Discussion

Based on the critical role of CD4Treg in the maintenance of immune tolerance and the observation that patients with cGVHD had lower
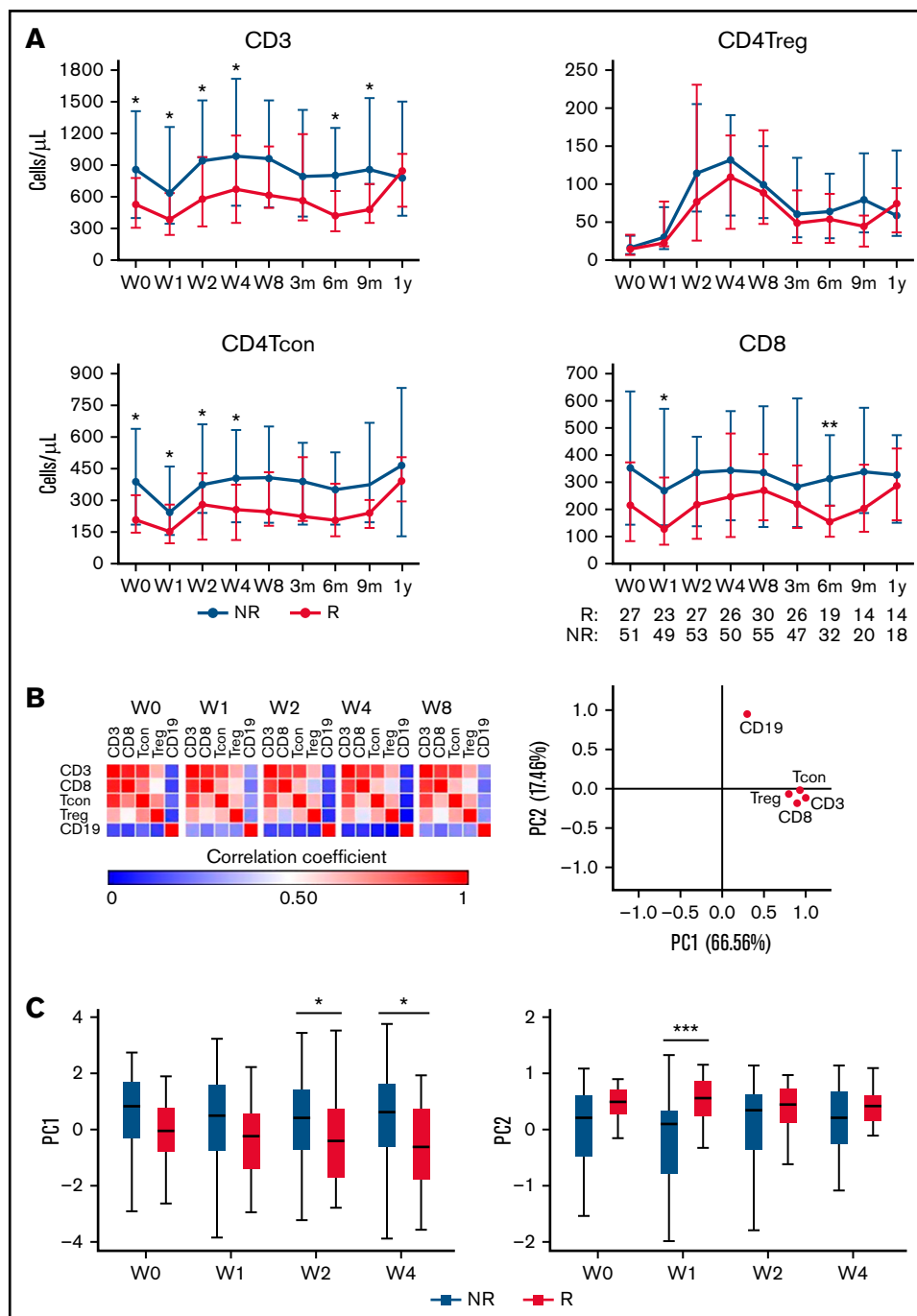
Figure 3. Immune markers associated with skin response. (A) Immunologic reconstitution according to skin response. Number of subjects (N) at each time point is the same in every figure. (B) Correlation between lymphocyte populations and component pattern analysis after LD IL-2 therapy. (C) PCA and clinical response. *$P < .05$, **$P < .01$, ***$P < .005$. m, month; NR, nonresponder; R, responder; W, week; y, year.

numbers of CD4Treg,[8] LD IL-2 was developed as a therapy for SR-cGVHD with the rationale that preferential in vivo expansion of CD4Treg and restoration of a favorable CD4Treg:CD4Tcon ratio would enhance immune tolerance and improve clinical manifestations of cGVHD. Indeed, in a series of 5 clinical trials, we showed safety and efficacy of daily LD IL-2 in the treatment of SR-cGVHD with a combined overall response rate of 53% in adults.[9-13] Limited experience in pediatric patients suggests that clinical responses

may be higher in this population.[11] The clinical response rate in adults with SR-cGVHD is comparable to that of other emerging agents. For example, response rates with ibrutinib, a Bruton's tyrosine kinase inhibitor, range from 67% to 69%,[19,20] 65% to 72% with belumosudil (a rho-associated coiled-coil–containing protein kinase-2 inhibitor[21,22]), and 49% to 77% with ruxolitinib (a JAK2 inhibitor).[23] Although similar, these response rates are not exactly comparable as inclusion criteria varied among these studies, and
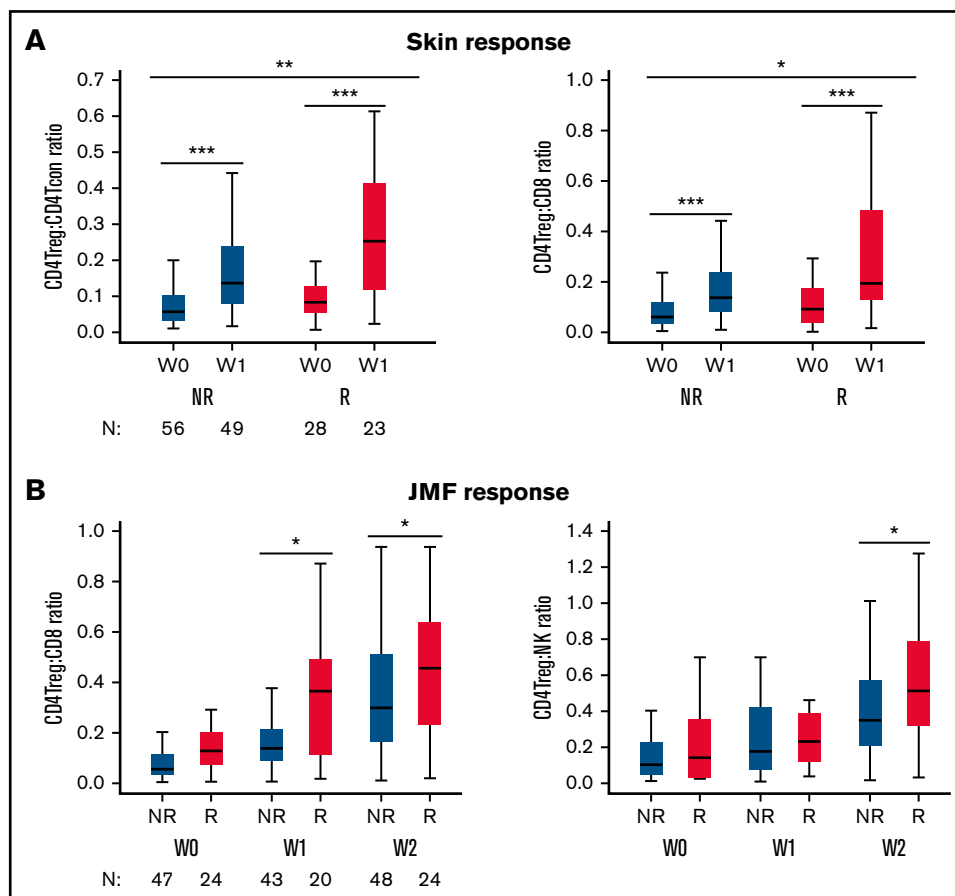
Figure 4. Regulatory cell/effector cell ratios and clinical response in patients with skin or JMF cGVHD. (A) CD4Treg:CD4Tcon ratio and CD4Treg:CD8 ratio according to skin response. (B) CD4Treg:CD4Tcon ratio and CD4Treg:NK ratio according to JMF response. Number of subjects (N) is the same in both panels in each row. *P < .05, **P < .01, ***P < .005. NR, nonresponder; R, responder; W, week.

best response rate during an unspecified observation period can be much higher than the overall response rate at a specific time point. Nevertheless, one common observation is that complete responses are relatively rare with all these agents. This suggests that targeting one pathway may not be sufficient to control SR-cGVHD, and development of combination therapies for targeting complementary pathways is needed. Although individual agents may be more effective in combination with steroids,[24] targeting different cGVHD pathways may also be important when these agents are tested as frontline therapy for newly diagnosed cGVHD. The development of specific organ-targeted therapies may also be beneficial, but this will depend on additional progress to define mechanisms responsible for organ-specific injury in cGVHD.

In the current study, we aggregated all five IL-2 clinical trials that we conducted at our center between 2008 and 2017 to investigate organ-specific response and its immunologic correlates. We found that initiating therapy early with fewer prior therapies is beneficial to overall response. Of sites investigated, skin is the most predominant site of disease, which is consistent with previous reports describing clinical manifestations of cGVHD.[23] As in overall response, shorter time to initiating therapy and fewer number of prior therapies were associated with response to skin cGVHD. For the entire cohort, survival outcome was excellent (2-year OS, 84% from study entry;

5-year OS and PFS, 62% from study entry; 5-year OS, 79% from cGVHD onset and 86% from stem cell infusion). Neither specific-organ involvement nor overall or organ-specific response was associated with OS. This result is consistent with the finding reported in the phase 3 ruxolitinib study.[23] In that study, OS values in the ruxolitinib and control arms are completely superimposable, although this result warrants a longer follow-up. This survival outcome can be compared with the 82% 2-year OS in a recently reported phase 2, rho-associated coiled-coil–containing protein kinase-2 inhibitor trial,[21] 71% 2-year OS in the follow-up study of ibrutinib,[20] 76% 2-year OS in a real-world experience of ibrutinib,[25] approximately 73% 2-year OS in the phase 3 ruxolitinib study,[23] and 77% and 83% in reports of real-world experience with ruxolitinib.[26,27]

Daily LD IL-2 led to rapid expansion of peripheral CD4Treg in all patients, with no significant changes in CD4Tcon or CD8 T cells. As reported previously,[9-13] a higher CD4Treg:CD4Tcon ratio at 1 week was associated with overall response. This result was due to both higher CD4Treg and lower CD4Tcon in patients with clinical responses at this time point. CD56[bright] NK cells also constitutively express a high-affinity IL-2 receptor, and LD IL-2 also leads to expansion of this NK cell subset in all patients.[28,29] Our analysis of this larger cohort found that nonresponders had higher levels of peripheral NK cells at 1 and 2 weeks after starting LD IL-2.
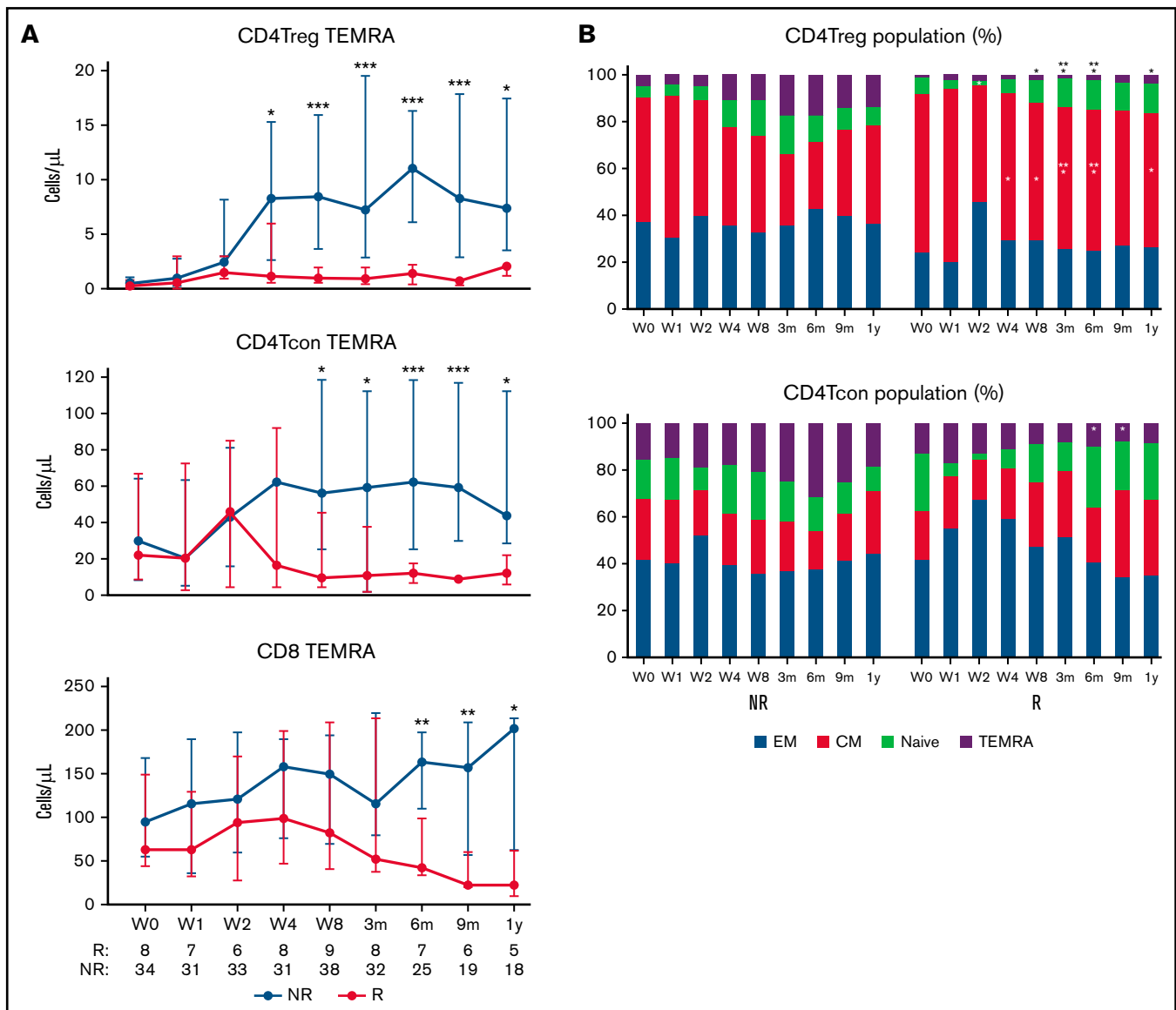
Figure 5. Immune correlates with lung cGVHD response. (A) Median absolute counts of CD4Treg TEMRA, CD4Tcon TEMRA, and CD8 TEMRA with the interquartile range. (B) Percent CD4Treg and %CD4Tcon populations according to lung response. Number of subjects (N) at each time point is the same in every figure. *$P < .05$, **$P < .01$, ***$P < .005$. NR, nonresponder; m, month; R, responder; W, week; y, year.

Although NK cells have not been shown to play an important role in the development of cGVHD, CD56^bright NK cells have predominately regulatory functions. By competing for exogenous IL-2, preferential expansion of this NK subset may impair the ability of CD4Treg to expand and promote immune tolerance.[30]

Consistent with the analysis of overall response, skin responders had higher CD4Treg:CD4Tcon and CD4Treg:CD8 ratios at 1 week after starting LD IL-2. A similar observation was noted for JMF responders, but the magnitude of the difference was less. Interestingly, skin responders had lower lymphocyte and CD3 T-cell counts before starting LD IL-2, and this difference persisted throughout the follow-up period. Levels of both CD4Tcon and CD8 T cells were lower at baseline in skin responders, but absolute levels of CD4Treg were similar at baseline, increased in all patients, and were not associated with clinical response. In contrast, B-cell counts were higher at baseline in responders, and this difference also persisted throughout the follow-up period. These observations emphasize the complexity of the immune responses that contribute to tissue damage in cGVHD and in this case highlight the potential role of autoreactive B cells that develop in the setting of prolonged B-cell lymphopenia.[5,31]

Because few patients presented with both skin and lung involvement, clinical responses in these 2 sites did not overlap. Although the smaller number of patients with lung involvement limited the statistical power of our analysis, we noted a distinct pattern of immune reconstitution in lung responders. In particular, TEMRA within CD4Treg, CD4Tcon, and CD8 T-cell populations was lower in the lung responders compared with nonresponders after 4 weeks of

the therapy. Also, percent CM CD4Treg was higher in responders and percent naive CD4Treg was lower at various time points. The impact of these differences on Treg function or distribution to sites of lung injury are not clear, and further studies will be needed to better understand the biologic implications of these phenotypic differences.

Although we analyzed aggregated data on 105 patients from 5 clinical trials, sample size still limited our ability to elucidate organ-specific responses and its impact on immunologic response. In our study, the predominant cGVHD site is skin followed by JMF. Because JMF involvement largely overlapped with skin, it was not possible to clearly distinguish organ-specific responses between these 2 sites. We were also unable to evaluate eyes and mouth because our studies permitted concurrent topical therapies. Lung responses showed unique features of immune reconstitution, but this finding requires validation in a larger study. The number of patients with GI tract, liver, and genital tract involvement was too small for any meaningful analysis. Lastly, the study was based on early-phase nonrandomized studies at a single center, and some of the studies included additional therapies such as ECP and Treg infusions. In both ECP lead-in and Treg infusion trials,[11,12] immune reconstitution was not affected by these additional therapies (supplementary materials). In patients who received ECP before IL-2, organ-specific responses were either sustained or 1-point improved after addition of LD IL-2. Larger studies focusing on tissue-specific responses that include multiple sites are needed to confirm the results of this analysis.

Although the biology of cGVHD is now better understood, and new agents are emerging, clinical responses remain suboptimal. Forty percent to 50% of patients do not respond to individual agents, and few patients with SR-cGVHD achieve complete responses to second-line therapy. Because the immunologic pathways targeted by new agents are very different, it may be possible to improve clinical responses by combining agents. For example, LD IL-2, which selectively targets CD4Treg, could be combined with ibrutinib, ruxolitinib, or belumosudil, which each specifically inhibits different signaling pathways in B cells or effector T cells. The observation that LD IL-2 is most effective when used earlier in the course of cGVHD and after fewer failed therapies suggests that Treg-based therapies may be more effective in suppressing the initial development of cGVHD. This is consistent with previous studies showing that delayed recovery of Treg 3 months after transplant is associated with subsequent development of cGVHD.[32] Thus, LD IL-2 therapy may be more effective when used for cGVHD prevention or in combination with corticosteroids as primary therapy for cGVHD before extensive fibrosis and tissue damage have developed. Developing rationale combinations that target different pathways and various sites of disease as well as treatment earlier in the development of cGVHD will hopefully lead to more effective and more complete responses and reduce reliance on long-term maintenance therapy with corticosteroids. Because LD IL-2 therapies are now also being evaluated in many autoimmune diseases, the results of these studies may be relevant to much larger populations of patients.

Authorship

Contribution: H.T.K. conceived and designed the study, performed statistical analysis, interpreted the data, and wrote the manuscript; J.R. conceived the study and wrote the manuscript; C.G.R. collected and assembled immunophenotypic data; P.S. compiled the outcome data; and all authors contributed to the manuscript review and approved the final version for submission.

Conflict-of-interest disclosure: H.T.K. has served as a consultant for Miltenyi, outside the submitted work. J.K. has served as a consultant for Amgen, Equillium, EMD Serono, Cugene, Cue BioPharma, GentiBio, and Moderna and advisory board member for Therakos/Mallinckrodt, Biolojic Design, and Gamida Cell; and received funding from BMS, Miltenyi, Clinigen, and Regeneron, outside the submitted work. C.C. has advisory/consultancy to Mallinckrodt, Deciphera, Jazz, Incyte, Sanofi, Bristol Myer Squibb, CTI Biopharma, Equillium, and Kadmon (pro bono). R.J.S. has served as a consultant for Gilead, Rheos Therapeutics, Jazz, Cugene, Mana Therapeutics, VOR, and Novartis; served on data safety monitoring committees for Juno; and served on board of directors for Kiadis and Be the Match/National Marrow Donor Program, outside the submitted work. J.R. receives research funding from Amgen, Equillium, Kite/Gilead and Novartis; and serves on Data Safety Monitoring Committees for AvroBio and Scientific Advisory Boards for Akron Biotech, Clade Therapeutics, Garuda Therapeutics, Immunitas Therapeutics, LifeVault Bio, Novartis, Rheos Medicines, Talaris Therapeutics, and TScan Therapeutics. The remaining authors declare no competing financial interests.

ORCID profiles: H.T.K., 0000-0003-4889-6849; J.W., 0000-0002-5759-2532; J.R., 0000-0001-5526-4669.

Correspondence: Haesook T. Kim, Department of Data Science, Dana-Farber Cancer Institute, 450 Brookline Ave, Boston, MA 02215; e-mail: htkimc@jimmy.harvard.edu.

References

1. Zeiser R, Blazar BR. Pathophysiology of chronic graft-versus-host disease and therapeutic targets. *N Engl J Med*. 2017;377(26):2565-2579.

2. Socié G, Kean LS, Zeiser R, Blazar BR. Insights from integrating clinical and preclinical studies advance understanding of graft-versus-host disease. *J Clin Invest*. 2021;131(12):e149296.

3. Cooke KR, Luznik L, Sarantopoulos S, et al. The biology of chronic graft-versus-host disease: a task force report from the National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host disease. *Biol Blood Marrow Transplant*. 2017;23(2):211-234.

4. Jaglowski SM, Blazar BR. How ibrutinib, a B-cell malignancy drug, became an FDA-approved second-line therapy for steroid-resistant chronic GVHD. *Blood Adv*. 2018;2(15):2012-2019.

5.   Zeiser R, Sarantopoulos S, Blazar BR. B-cell targeting in chronic graft-versus-host disease. *Blood.* 2018;131(13):1399-1405.

6.   Mahadeo KM, Masinsin B, Kapoor N et al. Immunologic resolution of human chronic graft-versus-host disease. *Biol Blood Marrow Transplant.* 2014;20(10):1508-1515.

7.   Matsuoka K, Kim HT, McDonough S, et al. Altered regulatory T cell homeostasis in patients with CD4[+] lymphopenia following allogeneic hematopoietic stem cell transplantation. *J Clin Invest.* 2010;120(5):1479-1493.

8.   Zorn E, Kim HT, Lee SJ, et al. Reduced frequency of FOXP3[+] CD4[+]CD25[+] regulatory T cells in patients with chronic graft-versus-host disease. *Blood.* 2005;106(8):2903-2911.

9.   Koreth J, Matsuoka K, Kim HT, et al. Interleukin-2 and regulatory T cells in graft-versus-host disease. *N Engl J Med.* 2011;365(22):2055-2066.

10. Koreth J, Kim HT, Jones KT, et al. Efficacy, durability, and response predictors of low-dose interleukin-2 therapy for chronic graft-versus-host disease. *Blood.* 2016;128(1):130-137.

11. Whangbo JS, Kim HT, Mirkovic N, et al. Dose-escalated interleukin-2 therapy for refractory chronic graft-versus-host disease in adults and children. *Blood Adv.* 2019;3(17):2550-2561.

12. Belizaire R, Kim HT, Poryanda SJ, et al. Efficacy and immunologic effects of extracorporeal photopheresis plus interleukin-2 in chronic graft-versus-host disease. *Blood Adv.* 2019;3(7):969-979.

13. Whangbo J, Nikiforow S, Kim HT, et al. A phase 1 study of donor regulatory T cell infusion plus low-dose interleukin-2 for steroid-refractory chronic graft-versus-host disease [published online ahead of print April 27, 2022]. *Blood Adv.* doi:10.1182/bloodadvances.2021006625.

14. Pavletic SZ, Martin P, Lee SJ, et al; Response Criteria Working Group. Measuring therapeutic response in chronic graft-versus-host disease: National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease: IV. Response Criteria Working Group report. *Biol Blood Marrow Transplant.* 2006;12(3):252-266.

15. Arai S, Jagasia M, Storer B, et al. Global and organ-specific chronic graft-versus-host disease severity according to the 2005 NIH Consensus Criteria. *Blood.* 2011;118(15):4242-4249.

16. Jagasia MH, Greinix HT, Arora M, et al. National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease: I. The 2014 Diagnosis and Staging Working Group report. *Biol Blood Marrow Transplant.* 2015;21(3):389-401.e1.

17. Lee SJ, Wolff D, Kitko C, et al. Measuring therapeutic response in chronic graft-versus-host disease. National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host disease: IV. The 2014 Response Criteria Working Group report. *Biol Blood Marrow Transplant.* 2015;21(6):984-999.

18. Lee S, Cook EF, Soiffer R, Antin JH. Development and validation of a scale to measure symptoms of chronic graft-versus-host disease. *Biol Blood Marrow Transplant.* 2002;8(8):444-452.

19. Miklos D, Cutler CS, Arora M, et al. Ibrutinib for chronic graft-versus-host disease after failure of prior therapy. *Blood.* 2017;130(21):2243-2250.

20. Waller EK, Miklos D, Cutler C, et al. Ibrutinib for chronic graft-versus-host disease after failure of prior therapy: 1-year update of a phase 1b/2 study. *Biol Blood Marrow Transplant.* 2019;25(10):2002-2007.

21. Jagasia M, Lazaryan A, Bachier CR, et al. ROCK2 inhibition with belumosudil (KD025) for the treatment of chronic graft-versus-host disease. *J Clin Oncol.* 2021;39(17):1888-1898.

22. Cutler C, Lee SJ, Arai S et al. Belumosudil for chronic graft-versus-host disease (cGVHD) after 2 or more prior lines of therapy: the ROCKstar study. *Blood.* 2021;138(22):2278-2289.

23. Zeiser R, Polverelli N, Ram R, et al; REACH3 Investigators. Ruxolitinib for glucocorticoid-refractory chronic graft-versus-host disease. *N Engl J Med.* 2021;385(3):228-238.

24. Pidala J, Onstad L, Martin PJ, et al. Initial therapy for chronic graft-versus-host disease: analysis of practice variation and failure-free survival. *Blood Adv.* 2021;5(22):4549-4559.

25. Chin KK, Kim HT, Inyang EA, et al. Ibrutinib in steroid-refractory chronic graft-versus-host disease, a single-center experience. *Transplant Cell Ther.* 2021;27(12):990.e1-990.e7.

26. Ferreira AM, Szor RS, Molla VC, et al. Long-term follow-up of ruxolitinib in the treatment of steroid-refractory chronic graft-versus-host disease. *Transplant Cell Ther.* 2021;27(9):777.e1-777.e6.

27. Redondo S, Esquirol A, Novelli S, et al. Efficacy and safety of ruxolitinib in steroid-refractory/dependent chronic graft-versus-host disease: real-world data and challenges. *Transplant Cell Ther.* 2022;28(1):43.e1-43.e5.

28. Caligiuri MA, Zmuidzinas A, Manley TJ, Levine H, Smith KA, Ritz J. Functional consequences of interleukin 2 receptor expression on resting human lymphocytes. Identification of a novel natural killer cell subset with high affinity receptors. *J Exp Med.* 1990;171(5):1509-1526.

29. Hirakawa M, Matos TR, Liu H, et al. Low-dose IL-2 selectively activates subsets of CD4[+] Tregs and NK cells. *JCI Insight.* 2016;1(18):e89278.

30. Fumagalli V, Venzin V, Di Lucia P, et al. Group 1 ILCs regulate T cell-mediated liver immunopathology by controlling local IL-2 availability. *Sci Immunol.* 2022;7(68):eabi6112.

31. Sarantopoulos S, Stevenson KE, Kim HT, et al. Altered B-cell homeostasis and excess BAFF in human chronic graft-versus-host disease. *Blood.* 2009;113(16):3865-3874.

32. Alho AC, Kim HT, Chammas MJ, et al. Unbalanced recovery of regulatory and effector T cells after allogeneic stem cell transplantation contributes to chronic GVHD. *Blood.* 2016;127(5):646-657.