

## TRANSPLANTATION

# Plasma biomarkers of acute GVHD and nonrelapse mortality: predictive value of measurements before GVHD onset and treatment

George B. McDonald,<sup>1,2</sup> Laura Tabellini,<sup>1</sup> Barry E. Storer,<sup>1,3</sup> Richard L. Lawler,<sup>1</sup> Paul J. Martin,<sup>1,2</sup> and John A. Hansen<sup>1,2</sup>

<sup>1</sup>Division of Clinical Research, Fred Hutchinson Cancer Research Center, Seattle, WA; and <sup>2</sup>Department of Medicine, and <sup>3</sup>Department of Biostatistics, University of Washington School of Medicine, Seattle, WA

## Key Points

- Before GVHD treatment, informative plasma biomarkers included TIM3, IL6, sTNFR1 (for grade 3-4 GVHD), and ST2 and sTNFR1 (for NRM at 1 year).
- In a day 14 landmark analysis, plasma TIM3 was predictive of grade 3-4 GVHD.

We identified plasma biomarkers that presaged outcomes in patients with gastrointestinal graft-versus-host disease (GVHD) by measuring 23 biomarkers in samples collected before initiation of treatment. Six analytes with the greatest accuracy in predicting grade 3-4 GVHD in the first cohort (74 patients) were then tested in a second cohort (76 patients). The same 6 analytes were also tested in samples collected at day 14 ± 3 from 167 patients free of GVHD at the time. Logistic regression and calculation of an area under a receiver-operating characteristic (ROC) curve for each analyte were used to determine associations with outcome. Best models in the GVHD onset and landmark analyses were determined by forward selection. In samples from the second cohort, collected a median of 4 days before start of treatment, levels of TIM3, IL6, and sTNFR1 had utility in predicting development of peak grade 3-4 GVHD (area under ROC curve, 0.88). Plasma ST2 and sTNFR1 predicted nonrelapse mortality within 1 year after transplantation (area under ROC curve, 0.90). In the landmark analysis, plasma TIM3 predicted subsequent grade 3-4

GVHD (area under ROC curve, 0.76). We conclude that plasma levels of TIM3, sTNFR1, ST2, and IL6 are informative in predicting more severe GVHD and nonrelapse mortality. (*Blood*. 2015;126(1):113-120)

## Introduction

The frequency of acute graft-versus-host disease (GVHD) after allogeneic hematopoietic cell transplantation (HCT) is in the 50% to 70% range, depending on the conditioning regimen, donor characteristics, and prophylaxis strategies.<sup>1</sup> Although the overall frequency of GVHD has remained stable during the past decade, its presentation has shifted toward gastrointestinal involvement as the major cause of morbidity and away from severe damage to the skin and liver.<sup>1,2</sup> The result of these clinical trends has been a reduction in the frequency of grade 3-4 GVHD to <10% in most centers, along with a 50% reduction in nonrelapse mortality (NRM).<sup>1</sup>

Retrospective analyses demonstrate that patients with more severe peak symptoms and especially more prolonged acute GVHD have substantially higher mortality rates than those with less severe and shorter-duration GVHD.<sup>3</sup> Recognition of the ultimate severity of GVHD often becomes apparent within the first 2 weeks of the onset of signs and symptoms, marked by the absence of improvement during initial prednisone therapy and the development of gastrointestinal mucosal necrosis and jaundice.<sup>4,5</sup> In patients with these adverse prognostic signs, secondary immune suppressive therapy provides suboptimal benefit, and mortality rates are high.<sup>5,6</sup>

If it were possible to predict the ultimate severity of GVHD before or at the onset of symptoms, preemptive immune suppressive therapy could be administered in an effort to blunt the intensity of tissue damage, especially in the gastrointestinal tract.<sup>2,7</sup> Research on the predictive value of plasma biomarkers has yielded several candidate

analytes that have been measured at higher levels in patients with GVHD than in allografted controls with no GVHD or less severe GVHD.<sup>2,7-13</sup> In the study reported here, 2 cohorts of patients provided frequent blood samples after allogeneic transplantation, and we measured plasma levels of 23 analytes previously reported to be elevated in patients with GVHD. In plasma samples from patients in the first cohort, we identified 6 analytes with the greatest accuracy in predicting more severe GVHD. We then measured the levels of these 6 analytes in a second cohort of patients. Data were analyzed in 2 ways. The first analysis examined the predictive value of biomarkers in plasma samples from the onset period, before initiation of treatment of GVHD, and the second was a landmark analysis based on samples collected 11 to 17 days after HCT (day 14 ± 3 days). The purpose of this work was to identify biomarkers during the onset phase of GVHD whose sensitivity and specificity could be translated into clinical utility in predicting more severe GVHD and a higher risk of NRM.

## Methods

### Allogeneic hematopoietic cell transplantation

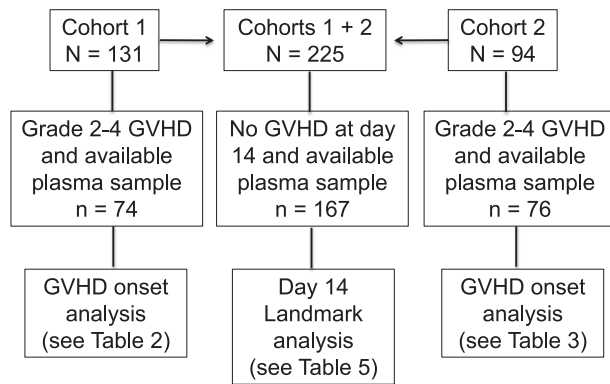
All patients except one received a myeloablative conditioning regimen followed by infusion of donor cells. Myeloablative conditioning regimens generally contained high-dose cyclophosphamide with busulfan or 12 to 13.2 Gy total

Submitted March 25, 2015; accepted May 6, 2015. Prepublished online as *Blood* First Edition paper, May 18, 2015; DOI 10.1182/blood-2015-03-636753.

The online version of this article contains a data supplement.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

© 2015 by The American Society of Hematology



**Figure 1.** Flowchart illustrating the number of patients in each analysis cohort (cohorts 1, 2, and the combined cohort landmark analysis at day 14 ± 3 days).

body irradiation. The day of donor cell infusion was day 0. Recipients were given immunosuppressive drugs, usually a calcineurin inhibitor plus methotrexate to prevent GVHD. Prophylaxis for infections included low-dose acyclovir, trimethoprim/sulfamethoxazole or dapsone, an antifungal agent, pre-emptive therapy with ganciclovir for patients with cytomegalovirus antigenemia or DNAemia, and antibiotics for patients with neutropenia. Ursodiol was given as a prophylaxis against cholestasis.

### Acute gastrointestinal GVHD

The peak stage of gastrointestinal GVHD, the peak grade of GVHD, and the date of onset of clinical signs and symptoms were independently scored, according to the extent of rash, the total serum bilirubin, the presence of upper gastrointestinal symptoms, and the daily stool volume.<sup>3,14</sup> Patients were classified according to peak GVHD grade (2-4, 2b-4, or 3-4).<sup>15</sup> NRM at 1 year after HCT was also noted.

### Study design

Patients gave informed consent for blood collection and for analysis of clinical data under protocols approved by the Fred Hutchinson Cancer Research Center Institutional Review Board. Blood samples were collected weekly from day 7 through day 70, from patients in 2 cohorts. Cohort 1 consisted of 131 consenting patients who received allografts from April 2003 to December 2008; 74 of 131 patients had gastrointestinal GVHD grade 2-4 and a plasma sample collected in the defined time window, and 57 of 131 had no gastrointestinal GVHD or no plasma sample collected in the defined window (Figure 1). Cohort 2 was composed of 94 patients who received allografts from August 2007 to April 2010<sup>16</sup>; 76 of 94 patients had gastrointestinal GVHD grade 2-4 and a plasma sample collected in the defined time window, and 18 of 94 had no gastrointestinal GVHD or no plasma sample collected in the defined window (Figure 1). Plasma from samples collected before or at GVHD onset and before the initiation of systemic treatment of GVHD from cohort 1 patients was analyzed for 23 candidate biomarkers (median, 2 days before start of treatment; range, -15 to +1). After calculation of the area under receiver-operating characteristic curves (ROC), 6 analytes from cohort 1 patients that were correlated with grade 3-4 GVHD and NRM at 1 year were identified. These 6 analytes were then measured in plasma samples collected before or at onset and before initiation of treatment of GVHD from cohort 2 patients (median, 4 days before start of treatment; range -7 to -1). In addition, a landmark analysis was undertaken in which the same 6 analytes were measured in 167 plasma samples that had been collected at day 14 ± 3 days (day 11-17) after transplantation in 167 patients from cohorts 1 and 2 who had no clinical evidence of acute GVHD when blood samples were drawn (Figure 1).

### Collection and processing of blood samples

Blood samples were collected from central venous access catheters and centrifuged, and plasma was aliquoted in 0.5-mL tubes for cryopreservation at -80°C. Plasma was collected from 48 healthy, consenting, normal adults who were free of fever or any respiratory or influenzalike symptoms for at least 7 days

before phlebotomy, and who had received no nonsteroidal antiinflammatory drugs, glucocorticoid medication, or antibiotics for at least 48 hours. This normal control cohort included 29 women and 19 men, with a median age of 37 years (range, 23-67).

**Assay methods for plasma proteins: Luminex microbead assay and enzyme-linked immunosorbent assay.** The Luminex microbead method (Luminex, Austin, TX) was used for measurement of 21 analytes, as previously described.<sup>12</sup> A sandwich enzyme-linked immunosorbent assay method was used for transforming growth factor-β1 and REG3α, as previously described.<sup>17</sup> Supplemental Table 1 lists the reagents used in these assays along with their respective lower levels of detection.

**Statistical methods.** All biomarker values were log<sub>10</sub>-transformed before analysis. Values at the lower limit of detection were assigned that value. Univariate logistic regression was used to evaluate the association of each biomarker with the outcomes of interest, including the calculation of area under the ROC curve. Odds ratios from these models refer to the increase in odds of the outcome for a tenfold increase in the analyte. Best models in the cohort 2 onset analysis and the combined cohort landmark analysis were determined by forward selection at the .05 level of significance. Backward selection yielded the same model in all cases. All *P* values are derived from likelihood ratio statistics and are 2-sided. All statistical analysis was carried out using SAS version 8 (SAS Systems, Cary, NC).

**Table 1. Characteristics of patients in the GVHD onset analyses (cohorts 1 and 2) and the landmark analysis at 11 to 17 days (day 14 ± 3) after transplantation**

	Cohort 1 onset analysis (n = 74)	Cohort 2 onset analysis (n = 76)	Landmark analysis (n = 167)
Median age, y (range)	42 (18-66)	45 (13-65)	45 (10-65)
Median day of GHVD onset (range)	20 (7-64)	24 (11-82)	26 (15-99)*
<b>Diagnosis, n (%)</b>			
Acute leukemia	38 (51)	54 (58)	93 (56)
Myelodysplastic syndrome	16 (22)	20 (26)	40 (24)
Lymphoma	7 (9)	—	6 (4)
Chronic myeloid leukemia	12 (16)	11 (14)	25 (15)
Other diagnosis	1 (1)	1 (1)	3 (2)
<b>Donor type, n (%)</b>			
Related	33 (45)	25 (33)	71 (43)
Unrelated	41 (55)	51 (67)	96 (57)
<b>Graft type, n (%)</b>			
Marrow	6 (8)	16 (21)	28 (17)
Peripheral blood	68 (92)	60 (79)	139 (83)
<b>Sex match, n (%)</b>			
Female to male	17 (23)	22 (29)	59 (35)
Other	57 (77)	54 (71)	108 (65)
<b>Conditioning regimens, n (%)</b>			
Myeloablative	73 (99)	76 (100)	167 (100)
Reduced intensity	1 (1)	—	—
<b>CMV serostatus before transplant (donor/recipient), n (%)</b>			
+/+	21 (28)	17 (22)	38 (23)
-/+	18 (24)	18 (24)	38 (23)
+/-	8 (11)	6 (8)	24 (14)
-/-	27 (36)	35 (46)	67 (40)
<b>GVHD prophylaxis, n (%)</b>			
Calcineurin inhibitor + methotrexate	65 (88)	74 (97)	156 (93)
Calcineurin inhibitor + methotrexate + mycophenolate mofetil	4 (5)	2 (3)	7 (4)
Calcineurin inhibitor + mycophenolate mofetil	3 (4)	—	2 (1)
Other regimens	2 (3)	—	2 (1)

\*For the 100 patients who subsequently developed GVHD.

**Table 2. Values for 23 analytes in samples drawn before the initiation of treatment of GVHD in patients from cohort 1**

	Normal median values (Q1-Q3) (N = 48*)	Values in patients with acute GVHD (Q1-Q3) (N = 74†)	Area under ROC curve‡ (P)		NRM at 1 y (n = 12) vs not NRM
			Peak grade 2b-4 (n = 25) vs grade 2a (n = 49)	Peak grade 3-4 (n = 12) vs grade 2 (n = 62)	
HGF	115 (90-129)	182 (135-341)	0.63 (.06)	0.72 (.01)§	0.81 (.001)§
IFN $\gamma$	0.3 (0.3-0.3)	2.93 (1.85-4.06)	0.61 (.05)	0.51 (.53)	0.61 (.13)
IL1 $\alpha$	2.2 (1.0-4.4)	1.00 (1.00-1.00)	0.52 (.30)	0.51 (.53)	0.57 (.10)
IL2	35.7 (16.0-62.8)	22.8 (3.6-64.9)	0.71 (.001)§	0.61 (.23)	0.60 (.19)
IL6	19.1 (8.7-36.6)	27.2 (17.0-47.5)	0.69 (.005)	0.75 (.02)§	0.73 (.03)§
IL8	3.8 (1.7-8.9)	34.4 (22.9-58.9)	0.69 (.001)	0.67 (.02)	0.67 (.03)
IL10	90.9 (48.5-164.5)	11.9 (5.7-22.3)	0.45 (.81)	0.69 (.47)	0.60 (.69)
IL12p70	10.2 (6.8-17.7)	39.2 (16.8-74.6)	0.55 (.64)	0.50 (.96)	0.55 (.52)
IL17	9.1 (2.7-13.6)	9.42 (1.00-29.67)	0.63 (.11)	0.51 (.58)	0.48 (.87)
IL18	18.1 (11.7-21.0)	64.8 (32.4-124.0)	0.64 (.07)	0.67 (.06)	0.63 (.10)
IL21	115 (8-177)	10.6 (8.0-75.5)	0.62 (.09)	0.51 (.92)	0.61 (.17)
IL22	240 (240-240)	240 (240-240)	0.52 (.20)	0.52 (.40)	0.52 (.40)
IL32	12.3 (8.0-31.0)	8.00 (8.00-8.00)	0.52 (.60)	0.48 (.98)	0.49 (.94)
IL33	496 (280-840)	499 (240-1107)	0.64 (.04)	0.48 (.84)	0.56 (.34)
IL2Ra	120 (120-120)	195 (120-449)	0.58 (.39)	0.49 (.91)	0.57 (.68)
sTNFR1	188 (120-448)	1720 (1080-2900)	0.61 (.17)	0.73 (.007)§	0.73 (.005)§
TNF $\alpha$	20.7 (4.8-37.4)	1.52 (1.00-5.14)	0.48 (.97)	0.71 (.03)§	0.68 (.10)
TSLP	2.0 (2.0-17.8)	2.00 (2.00-2.00)	0.55 (.67)	0.54 (.94)	0.54 (.79)
TGF $\beta$ 1	19.0 (11.2-33.1)	39.4 (16.4-72.9)	0.67 (.03)	0.54 (.46)	0.52 (.71)
CRP	329 (168-776)	8880 (1860-20680)	0.71 (.0009)§	0.68 (.03)	0.68 (.03)
TIM3	2290 (1860-2670)	5570 (2960-10150)	0.66 (.01)	0.75 (.002)§	0.66 (.03)
REG3a	21.5 (15.2-28.8)	36.7 (20.5-83.7)	0.62 (.22)	0.56 (.52)	0.71 (.009)§
ST2	15.0 (15.0-21.5)	21.1 (15.0-58.4)	0.77 (.0002)§	0.81 (.002)§	0.77 (.002)§

The operating characteristics for predicting more severe outcomes in patients with acute GVHD, and the relevant P values, are provided for each analyte.

\*TIM3 concentrations were measured in samples from 48 healthy controls, and REG3a and ST2 concentrations were measured in 47 healthy controls.

†REG3a and ST2 concentrations were measured in samples from 62 patients with acute GVHD.

‡ROC curves were derived for less vs more severe GVHD and for deaths before relapse vs survival among patients with GVHD.

§With area under an ROC curve  $\geq$ 0.70.

## Results

### Demographics of patients analyzed from cohorts 1 and 2 and in the landmark analysis

Table 1 summarizes characteristics of patients in this study, by analysis groups. Patients selected from cohorts 1 and 2 are similar except for a greater proportion of unrelated donors and marrow as a source of donor cells in patients from cohort 2.

### Biomarker concentrations before the initiation of treatment of acute GVHD

Values for analytes in plasma near to the time of symptomatic onset and before treatment of GVHD in patients from cohorts 1 and 2 are displayed in Tables 2 and 3, respectively. As judged by the area under an ROC curve of 0.70 or greater, the most informative analytes for predicting grade 3-4 vs grade 2 GVHD in cohort 1 patients were HGF, IL6, ST2, TIM3, tumor necrosis factor-alpha (TNF $\alpha$ ), and sTNFR1 (Table 2). The most informative analytes for predicting NRM at 1 year were HGF, IL6, ST2, sTNFR1, and Reg3a (Table 2).

When the 6 most informative analytes derived from cohort 1 were applied to the analysis of plasma samples drawn before initiation of treatment of GVHD in patients from cohort 2, IL6 was the most informative analyte(s) for predicting grade 2b-4 GVHD vs grade 2a GVHD. IL6, TIM3, and sTNFR1 were the most informative for predicting grade 3-4 from grade 2 GVHD. ST2 and sTNFR1 were most informative for predicting NRM vs survival at 1 year (Table 3 and Figure 2).

Table 4 displays the sensitivity and specificity of the best-model analytes in cohort 2, derived from the data in Table 3 and Figure 2, with

regard to the prediction of more severe GVHD (Table 4A) and NRM (Table 4B).

### Analytes in plasma at transplant day 11 to 17 (day 14 $\pm$ 3): landmark analysis

The 6 informative analytes derived from cohort 1 were measured in plasma samples collected at 11 to 17 days after transplantation from patients in both cohorts who were without symptoms of GVHD when the blood was drawn. Note that these analyses included patients who subsequently developed GVHD after transplantation and those who did not. Of the 167 patients without GVHD at the time of the landmark sample, 100 developed grade 2-4 acute GVHD. The onset of GVHD symptoms occurred at a median 26 days (range, 15-99) after HCT. The most informative analyte for predicting more severe from less severe GVHD was TIM3 (Table 5, Figure 3, and Table 4C).

### Pro- and antiinflammatory cytokines and other analytes across time, by presence or absence of gastrointestinal GVHD

Figure 4 displays plots of values for 23 analytes in blood drawn at weekly intervals after transplant from patients in cohort 1. Patients who developed GVHD with gastrointestinal involvement are represented in red, and patients who never developed GVHD are represented in black.

## Discussion

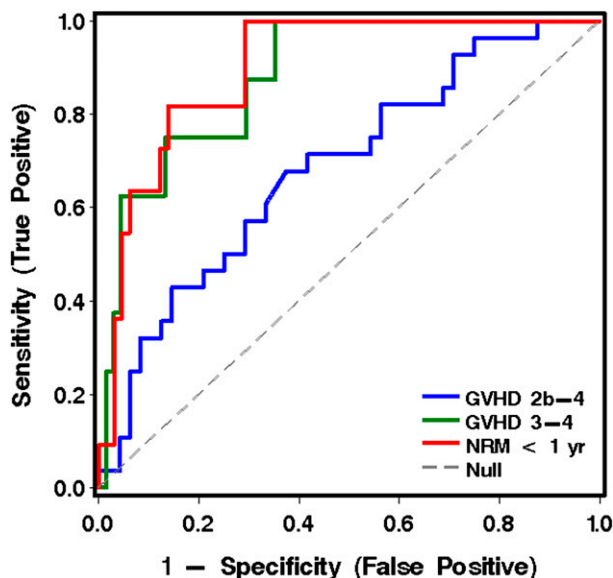
The purpose of this study was to identify biomarkers before the development of symptoms that would predict grade 3-4 GVHD and NRM

**Table 3. Analysis of the predictive value of 6 analytes in 76 samples drawn before the initiation of treatment of GVHD in patients in cohort 2**

	Area under ROC curve, P value, OR (95% CI), per log increase in analyte		
	Peak grade 2b-4 (n = 28) vs grade 2a (n = 48)	Peak grade 3-4 (n = 8) vs grade 2 (n = 68)	NRM at 1 y (n = 11) vs not NRM
<b>HGF</b>			
AUC	0.58	0.61	0.73
P	.23	.40	.12
OR (95% CI)	1.9 (0.6-5.8)	2.0 (0.4-9.0)	3.0 (0.8-11)
<b>IL6</b>			
AUC	0.69	0.83	0.80
P	.007	.009	.003
OR (95% CI)	3.2 (1.3-7.8)	4.7 (1.4-16)	4.9 (1.6-15)
<b>ST2</b>			
AUC	0.64	0.78	0.73
P	.04	.01	.003
OR (95% CI)	5.5 (1.0-31)	26.3 (1.9-367)	29.2 (2.6-335)
<b>TIM3</b>			
AUC	0.64	0.81	0.82
P	.04	.004	.0007
OR (95% CI)	5.0 (1.0-24)	46.1 (2.3-926)	59.8 (3.8-946)
<b>TNF<math>\alpha</math></b>			
AUC	0.52	0.61	0.69
P	.85	.38	.07
OR (95% CI)	1.2 (0.2-9.3)	4.1 (0.2-96)	13.7 (0.7-260)
<b>sTNFR1</b>			
AUC	0.60	0.67	0.88
P	.26	.21	$\leq .0001$
OR (95% CI)	3.4 (0.4-30)	8.3 (0.3-235)	999+ (19-999+)
<b>Best model by forward selection</b>			
AUC	.69	.88	.90
P	.007	.001	<.0001
OR (95% CI)	3.2 (1.3-7.8)	6.8 (1.3-35), 519 (3.6-999+), 0.0 (0.0-1.5), respectively	17.2 (1.3-232), 999+ (13.3-1000+), respectively

AUC, area under a ROC curve.

with sufficient accuracy that would allow preemptive therapy. For the time period before or at the onset of symptoms of GVHD with gastrointestinal involvement (median of 4 days before start of



**Figure 2. ROC curves for analytes in plasma for the prediction of more severe acute GVHD and NRM for patients in cohort 2, using the best model analytes given in Table 3.**

**Table 4. Sensitivity and specificity of measurement of analytes before GVHD treatment**

A. Derived from Figure 2, for prediction of grade 3-4 GVHD			
	Yes GVHD 3-4	No GVHD 3-4	
Predicted GVHD 3-4	6	9	PPV 40%
Predicted GVHD 2	2	59	
Sensitivity 75%		Specificity 87%	
B. Derived from Figure 2, for prediction of NRM			
	Yes NRM	No NRM	
Predicted NRM	9	9	PPV 50%
Predicted Negative	2	56	
Sensitivity 82%		Specificity 86%	
C. Derived from Figure 3, landmark analysis for prediction of grade 3-4 GVHD			
	Yes GVHD 3-4	No GVHD 3-4	
Predicted GVHD 3-4	8	42	PPV 16%
Predicted GVHD 0-2	3	114	
Sensitivity 73%		Specificity 73%	

PPV, positive predictive value.

treatment), the principal findings are that (1) measurement of plasma IL6, TIM3, and sTNFR1 has utility in predicting development of grade 3-4 vs grade 2 GVHD (area under a ROC curve, 0.88); and (2) measurement of ST2 and sTNFR1 has predictive value with regard to NRM (area under a ROC curve, 0.90). In our day, 14 landmark analyses of samples collected between day 11 and 17, measurement of TIM3 was the most useful analyte for predicting grade 3-4 vs grade 0-2

**Table 5. Analysis of 6 analytes in 167 samples collected at 11 to 17 d (day 14  $\pm$  3) after transplantation**

	Area under ROC curve, P value OR (95% CI), per log increase in analyte		
	Peak grade 2-4 (n = 100) vs grade 0-1	Peak grade 2b-4 (n = 37) vs grade 0-2a	Peak grade 3-4 (n = 11) vs grade 0-2
<b>HGF</b>			
AUC	0.56	0.58	0.63
P	.31	.29	.17
OR (95% CI)	1.6 (0.6-4.0)	1.7 (0.6-4.7)	2.9 (0.7-13)
<b>IL6</b>			
AUC	0.57	0.62	0.68
P	.10	.008	.05
OR (95% CI)	1.5 (0.9-2.7)	2.2 (1.2-4.0)	2.4 (1.0-5.8)
<b>ST2</b>			
AUC	0.52	0.52	0.56
P	.66	.73	.32
OR (95% CI)	1.2 (0.5-2.8)	1.2 (0.5-3.1)	2.1 (0.5-8.3)
<b>TIM3</b>			
AUC	0.64	0.70	0.71
P	.002	.0004	.01
OR (95% CI)	6.0 (1.9-19)	9.9 (2.7-36)	12.1 (1.6-92)
<b>TNF<math>\alpha</math></b>			
AUC	0.53	0.59	0.61
P	.64	.16	.20
OR (95% CI)	1.4 (0.4-5.0)	2.8 (0.7-12.2)	4.2 (0.5-36)
<b>sTNFR1</b>			
AUC	0.61	0.63	0.70
P	.02	.01	.03
OR (95% CI)	5.4 (1.3-23)	7.0 (1.5-33)	12.9 (1.3-130)
<b>Best model by forward selection</b>			
AUC	0.64	0.70	0.71
P	.002	.0004	.01
OR (95% CI)	6.0 (1.9-19)	9.9 (2.7-34)	12.1 (1.6-92)

Patients did not have GVHD when the blood sample was drawn. End points are defined by onset of GVHD with gut involvement requiring systemic therapy.

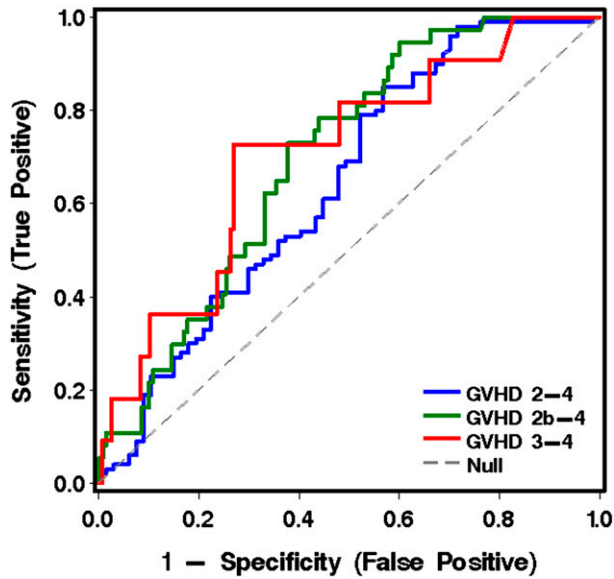


Figure 3. ROC curves for analytes in plasma at transplant day 14 (landmark analysis, cohorts 1 and 2 combined), using the best model analytes given in Table 5 for predicting more severe from less severe GVHD to day 100.

GVHD to day 100 post-transplant (area under an ROC curve, 0.76). Analysis of inflammatory cytokines and markers of inflammation across time (Figure 4) suggests that most allograft recipients experience

significant systemic inflammation to day 100. Our findings are consistent with recent studies that have identified TNFR1 and ST2 as useful biomarkers for more severe GVHD<sup>2,7-9</sup> but that are inconsistent with regard to REG3 $\alpha$ 11 (which was less predictive than other analytes) and TIM3<sup>12</sup> (which was more predictive).

Despite the statistically significant operating characteristics of the analytes that we have identified in predicting more severe GVHD and NRM, the use of these data in clinical trials designed to improve outcomes is likely to encounter difficulties. The day of onset of GVHD symptoms is unknown a priori, and thus blood samples would have to be drawn at frequent intervals and analytes measured in real time to provide data that would inform treatment decisions. Landmark data would be easier to generate in clinical practice but may not be as accurate. Most centers currently report a low incidence of more severe gastrointestinal GVHD. As can be seen in the 2  $\times$  2 tables (Table 4), the sensitivity, specificity, and positive predictive value of our “best” analytes and combination of analytes show that the number of false positives would still outnumber the true positives, and thus some patients who were not destined to develop more severe GVHD would receive inappropriate preemptive therapy. However, the outcome in patients with more severe, protracted gastrointestinal GVHD is dismal. Therefore, multicenter clinical trials are warranted to examine the hypothesis that preemptive treatment of appropriately selected patients could alter the course of the disease and decrease the risk of mortality. The treatment that should be tested in clinical trials of preemptive therapy is an open question. More aggressive preemptive

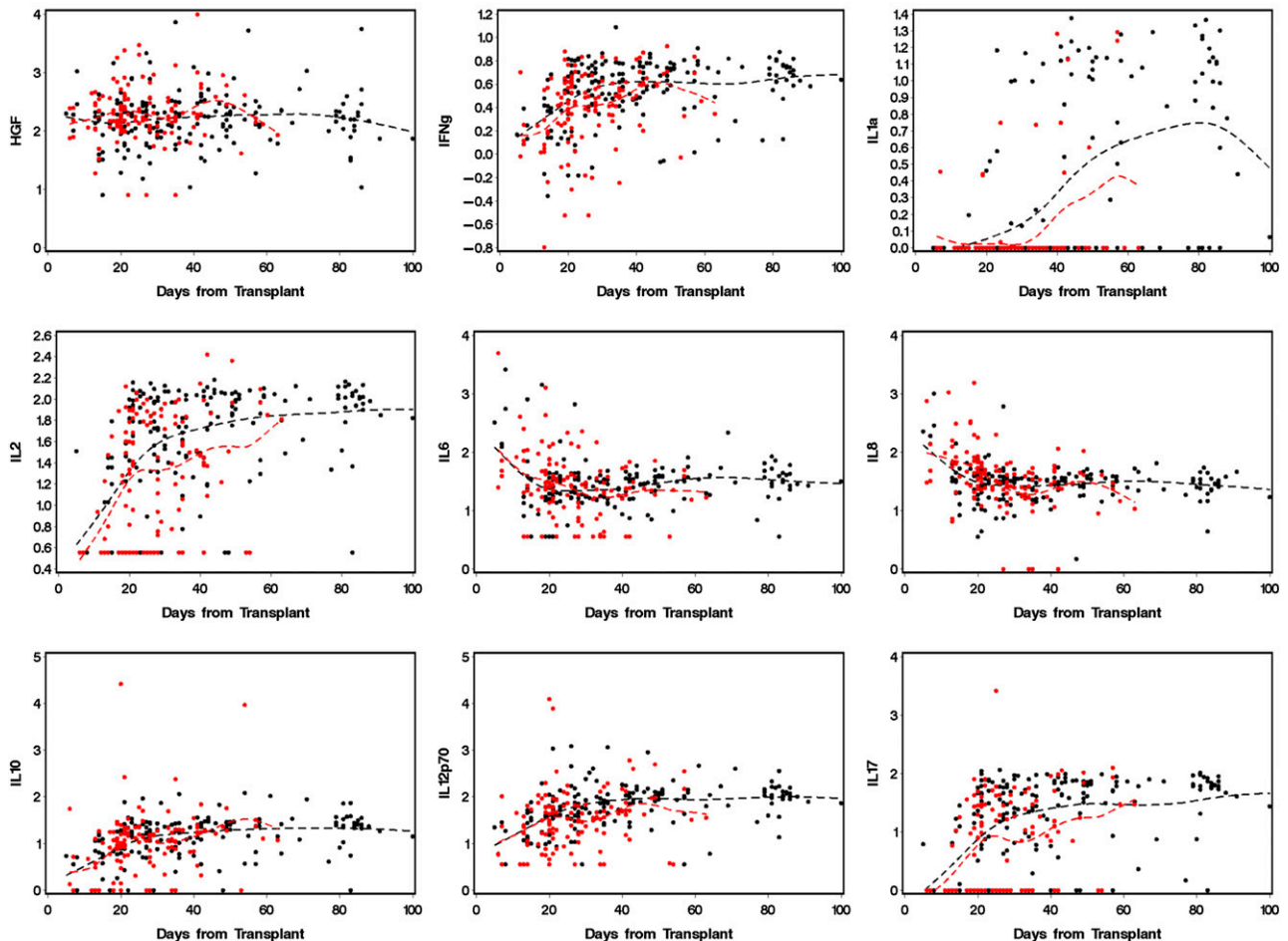


Figure 4. Values for analytes found in blood samples from patients in cohort 1, drawn at weekly intervals. The dotted lines are cubic spline curves, for GVHD cases in red and non-GVHD cases in black.

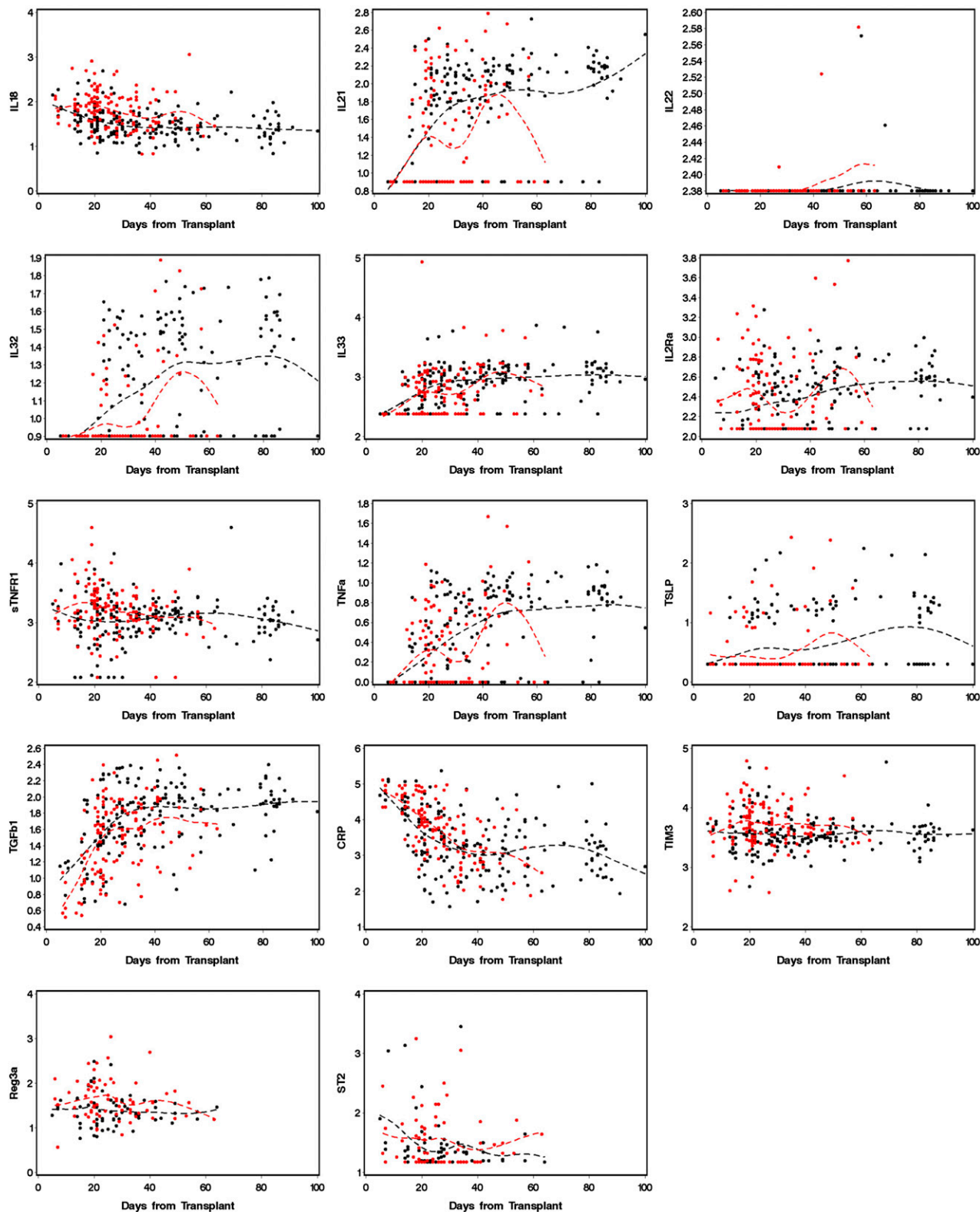


Figure 4. (Continued).

approaches—for example, using anti-T cell therapy—might be effective in improving outcomes for patients truly destined to develop more severe GVHD, but might pose the unnecessary risk of greater immune suppression among patients who were never destined to develop more severe GVHD.

The analyte TIM3 provided the most predictive utility in our landmark analysis of samples drawn at approximately day 14 after transplantation (Table 5 and Figure 3). However, the grade 3-4 GVHD–positive predictive value of TIM3 measurement at day 14 was only 16%. Our choice of day 14 for the landmark analysis was arbitrary,

designed to capture the largest number of patients before clinical signs of GVHD developed. One might choose a later day if patients had received reduced-intensity conditioning regimens, because their onset of GVHD tends to be later than GVHD after myeloablative conditioning regimens.<sup>18</sup> A recent study of cord blood recipients with a day 28 landmark showed that the analyte ST2 was predictive of transplant-related mortality and grade 3-4 GVHD.<sup>9</sup> ST2 has been associated with more severe GVHD in another study,<sup>8</sup> but at day 14, our data show TIM3, and not ST2, as the most useful analyte for grade 3-4 GVHD to day 100. Previous studies have not included TIM3 in the panel of analytes. A well-powered study has examined the utility of measuring 3 biomarkers (TNFR1, Reg3 $\alpha$ , and ST2) at the time of GVHD diagnosis to create an algorithm that predicts NRM 6 months later.<sup>2</sup> The resulting evidence-based Ann Arbor GVHD grading system addresses flaws in grading the severity of GVHD by its peak signs and symptoms. Alternative GVHD risk scoring systems for predicting outcomes, including NRM, rely solely on clinical parameters.<sup>3,19</sup> Other potential prognostic markers of gastrointestinal GVHD at the time of initial diagnosis include fecal samples (for calprotectin and  $\alpha$ -1-antitrypsin), falling serum albumin as a reflection of gut protein loss, endoscopic appearance, specific histologic changes, and circulating angiogenic factors.<sup>10,13,20-23</sup> Reconciling results of studies predicting the outcome of GVHD based on clinical parameters, plasma and fecal samples, and mucosal histology—done at different centers—is difficult, because there are differences in timing, handling of specimens, storage time, analytical techniques, choice of biomarkers, and statistical methods. It is also not clear whether prognostic markers for gastrointestinal GVHD will vary with the intensity of conditioning therapy, choice of GVHD prophylaxis, and source of hematopoietic cells.

The display of longitudinal data of analytes across time (Figure 4) suggests that most allograft recipients have biochemical evidence of systemic inflammation long after the clinical effects of conditioning therapy have resolved. A prospective study of colon biopsies demonstrated restoration of mucosal epithelium at around day 16 after myeloablative conditioning therapy.<sup>24</sup> Oral mucositis resulting from conditioning therapy also resolves in this time frame. A clinical study of eating behavior after myeloablative therapy also showed improvements in oral caloric intake at around day 20.<sup>17</sup> The graphs in Figure 4 suggest that a systemic inflammatory milieu persists after day 20, even among patients who do not have clinical signs or symptoms of gastrointestinal GVHD. We speculate that virtually all allograft recipients experience graft-vs-host reactions as the cause of systemic inflammation, but not all will develop clinically significant disease.<sup>25</sup>

In summary, we have identified plasma levels of TIM3, IL6, sTNFR1, and ST2 as informative analytes in predicting the develop-

ment of more severe GVHD and NRM among patients with gastrointestinal GVHD. Our study has several limitations. The numbers of patients in the outcome categories grade 3-4 GVHD and NRM are relatively small, reflecting current incidence rates. The study design, using consecutive cohorts of patients rather than randomizing to development and validation sets, was dictated by chronology—that is, analysis of plasma from cohort 1 patients was underway before samples from cohort 2 patients became available. Implementing a protocol that would lead to preemptive or initial therapy in those patients predicted to develop more severe GVHD will require a multicenter approach, uniform specimen handling, a central analytical laboratory, integration of clinical and biomarker parameters, and a primary end point of GVHD-related mortality, or its converse, GVHD-free survival.<sup>26</sup>

## Acknowledgments

This study was supported by grants from the National Institutes of Health, National Cancer Institute (CA18029, CA15704), National Institute of Allergy and Infectious Diseases (AI33484), and National Heart, Lung, and Blood Institute (HL094260).

## Authorship

Contribution: G.B.M. designed the research, supervised the Cytokine Shared Resource Laboratory, analyzed data, and wrote the paper; L.T. supervised the collection and archiving of plasma samples, recorded clinical data, and assisted in the preparation of the manuscript; B.E.S. assisted in study design and served as the statistician responsible for analysis of data; R.L.L. was the lead technician in the Cytokine Shared Resources Laboratory who developed and executed Luminex and ELISA assays for this research; P.J.M. designed the study, provided plasma specimens for Cohort 2, and assisted in the analysis of data and writing of the manuscript; and J.A.H. designed the study and assisted in the analysis of data and in the writing of the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: George B. McDonald, Clinical Research Division (D5-114), Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109-1024.

## References

- Gooley TA, Chien JW, Pergam SA, et al. Reduced mortality after allogeneic hematopoietic-cell transplantation. *N Engl J Med*. 2010;363(22):2091-2101.
- Levine JE, Braun TM, Harris AC, et al; for the Blood and Marrow Transplant Clinical Trials Network. A prognostic score for acute graft-versus-host disease based on biomarkers: A multicenter study. *Lancet Haematol*. 2015;2(1):e21-e29.
- Leisenring WM, Martin PJ, Petersdorf EW, et al. An acute graft-versus-host disease activity index to predict survival after hematopoietic cell transplantation with myeloablative conditioning regimens. *Blood*. 2006;108(2):749-755.
- Van Lint MT, Milone G, Leotta S, et al. Treatment of acute graft-versus-host disease with prednisolone: significant survival advantage for day +5 responders and no advantage for nonresponders receiving anti-thymocyte globulin. *Blood*. 2006;107(10):4177-4181.
- Castilla-Llorente C, Martin PJ, McDonald GB, et al. Prognostic factors and outcomes of severe gastrointestinal GVHD after allogeneic hematopoietic cell transplantation. *Bone Marrow Transplant*. 2014;49(7):966-971.
- Xhaard A, Rocha V, Bueno B, et al. Steroid-refractory acute GVHD: lack of long-term improved survival using new generation anticytokine treatment. *Biol Blood Marrow Transplant*. 2012;18(3):406-413.
- Paczesny S. Discovery and validation of graft-versus-host disease biomarkers. *Blood*. 2013;121(4):585-594.
- Vander Lugt MT, Braun TM, Hanash S, et al. ST2 as a marker for risk of therapy-resistant graft-versus-host disease and death. *N Engl J Med*. 2013;369(6):529-539.
- Ponce DM, Hilden P, Mumaw C, et al. High day 28 ST2 levels predict for acute graft-versus-host disease and transplant-related mortality after cord blood transplantation. *Blood*. 2015;125(1):199-205.
- Rezvani AR, Storer BE, Storb RF, et al. Decreased serum albumin as a biomarker for severe acute graft-versus-host disease after reduced-intensity allogeneic hematopoietic cell transplantation. *Biol Blood Marrow Transplant*. 2011;17(11):1594-1601.
- Ferrara JL, Harris AC, Greenson JK, et al. Regenerating islet-derived 3-alpha is a biomarker of gastrointestinal graft-versus-host disease. *Blood*. 2011;118(25):6702-6708.

12. Hansen JA, Hanash SM, Tabellini L, et al. A novel soluble form of Tim-3 associated with severe graft-versus-host disease. *Biol Blood Marrow Transplant*. 2013;19(9):1323-1330.
13. Holtan SG, Verneris MR, Schultz KR, et al. Circulating angiogenic factors associated with response and survival in patients with acute Graft-versus-Host Disease: Results from Blood and Marrow Transplant Clinical Trials Network 0302 and 0802. *Biol Blood Marrow Transplant*. 2015; 21(6):1029-1036.
14. Przepiorka D, Weisdorf D, Martin P, et al. 1994 Consensus Conference on Acute GVHD Grading. *Bone Marrow Transplant*. 1995;15(6):825-828.
15. Mielcarek M, Storer BE, Boeckh M, et al. Initial therapy of acute graft-versus-host disease with low-dose prednisone does not compromise patient outcomes. *Blood*. 2009;113(13): 2888-2894.
16. Martin PJ, Furlong T, Rowley SD, et al. Evaluation of oral beclomethasone dipropionate for prevention of acute graft-versus-host disease. *Biol Blood Marrow Transplant*. 2012;18(6): 922-929.
17. Malone FR, Leisenring WM, Storer BE, et al. Prolonged anorexia and elevated plasma cytokine levels following myeloablative allogeneic hematopoietic cell transplant. *Bone Marrow Transplant*. 2007;40(8):765-772.
18. Mielcarek M, Martin PJ, Leisenring W, et al. Graft-versus-host disease after nonmyeloablative versus conventional hematopoietic stem cell transplantation. *Blood*. 2003;102(2):756-762.
19. MacMillan ML, Robin M, Harris AC, et al. A refined risk score for acute graft-versus-host disease that predicts response to initial therapy, survival, and transplant-related mortality. *Biol Blood Marrow Transplant*. 2015;21(4):761-767.
20. Rodriguez-Otero P, Porcher R, Peffault de Latour R, et al. Fecal calprotectin and alpha-1 antitrypsin predict severity and response to corticosteroids in gastrointestinal graft-versus-host disease. *Blood*. 2012;119(24):5909-5917.
21. Abraham J, Janin A, Gornet JM, et al. Clinical severity scores in gastrointestinal graft-versus-host disease. *Transplantation*. 2014;97(9): 965-971.
22. Levine JE, Huber E, Hammer ST, et al. Low Paneth cell numbers at onset of gastrointestinal graft-versus-host disease identify patients at high risk for nonrelapse mortality. *Blood*. 2013;122(8): 1505-1509.
23. Socié G, Mary J-Y, Lemann M, et al. Prognostic value of apoptotic cells and infiltrating neutrophils in graft-versus-host disease of the gastrointestinal tract in humans: TNF and Fas expression. *Blood*. 2004;103(1):50-57.
24. Epstein RJ, McDonald GB, Sale GE, Shulman HM, Thomas ED. The diagnostic accuracy of the rectal biopsy in acute graft-versus-host disease: a prospective study of thirteen patients. *Gastroenterology*. 1980;78(4):764-771.
25. DiCarlo J, Agarwal-Hashmi R, Shah A, et al. Cytokine and chemokine patterns across 100 days after hematopoietic stem cell transplantation in children. *Biol Blood Marrow Transplant*. 2014; 20(3):361-369.
26. Holtan SG, DeFor TE, Lazaryan A, et al. Composite end point of graft-versus-host disease-free, relapse-free survival after allogeneic hematopoietic cell transplantation. *Blood*. 2015; 125(8):1333-1338.