

Acute graft-versus-host disease biomarkers measured during therapy can predict treatment outcomes: a Blood and Marrow Transplant Clinical Trials Network study

John E. Levine,¹ Brent R. Logan,² Juan Wu,³ Amin M. Alousi,⁴ Javier Bolaños-Meade,⁵ James L. M. Ferrara,¹ Vincent T. Ho,⁶ Daniel J. Weisdorf,⁷ and Sophie Paczesny¹

¹University of Michigan Blood and Marrow Transplant Program, Ann Arbor, MI; ²Medical College of Wisconsin Division of Biostatistics, Milwaukee, MN; ³The EMMES Corporation, Rockville, MD; ⁴The University of Texas MD Anderson Cancer Center, Department of Stem Cell Transplantation, Houston, TX; ⁵Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins Hospital, Baltimore, MD; ⁶Dana-Farber/Harvard Cancer Institute, Boston, MA; and ⁷University of Minnesota Medical Center Blood and Marrow Transplant Program, Minneapolis, MN

Acute graft-versus-host disease (GVHD) is the primary limitation of allogeneic hematopoietic cell transplantation, and once it develops, there are no reliable diagnostic tests to predict treatment outcomes. We hypothesized that 6 previously validated diagnostic biomarkers of GVHD (IL-2 receptor- α ; tumor necrosis factor receptor-1; hepatocyte growth factor; IL-8; elafin, a skin-specific marker; and regenerating islet-derived 3- α , a gastrointestinal tract-specific marker) could

discriminate between therapy responsive and nonresponsive patients and predict survival in patients receiving GVHD therapy. We measured GVHD biomarker concentrations from samples prospectively obtained at the initiation of treatment, day 14, and day 28, on a multicenter, randomized, 4-arm phase 2 clinical trial for newly diagnosed acute GVHD. We found that at each of 3 time points, GVHD onset, 2 weeks into treatment, and 4 weeks into treatment, a 6-protein bio-

marker panel predicted for the important clinical outcomes of day 28 posttherapy nonresponse and mortality at day 180 from onset. GVHD biomarker panels can be used for early identification of patients at high or low risk for treatment nonresponsiveness or death, and they may provide opportunities for early intervention and improved survival after hematopoietic cell transplantation. The study was registered in clinicaltrials.gov as NCT00224874. (*Blood*. 2012;119(16):3854-3860)

Introduction

Acute graft-versus-host disease (GVHD) is the primary limitation of allogeneic hematopoietic cell transplantation.¹⁻³ Resistance to GVHD therapy is associated with high transplant-related mortality and low overall survival.^{4,5} We and others have reported that clinical findings at diagnosis of GVHD, together with early assessment of treatment response, have considerable prognostic value for important outcomes such as GVHD severity 1 month after diagnosis and, most importantly, survival.⁶⁻⁸ In these studies, partial or complete treatment response on day 28 was strongly predictive of nonrelapse mortality and overall survival. However, even with this new knowledge, individualized GVHD treatment, based on an accurate understanding of likely long-term outcomes for a specific patient, remains an unmet goal.

Recently, investigators have applied proteomics discovery and validation strategies to identify biomarkers with relevance for GVHD. Among the first of these studies was a report of a panel of 4 informative biomarkers (tumor necrosis factor receptor-1 [TNFR1], IL-2 receptor- α [IL-2R α], IL-8, and hepatocyte growth factor [HGF]) with diagnostic and prognostic value for acute GVHD.⁹ Subsequently, GVHD target organ-specific biomarkers have been identified and validated by the same group. The GVHD of the skin-specific marker elafin demonstrated significant diagnostic and prognostic ability.¹⁰ Likewise, the gastrointestinal GVHD-specific biomarker regenerating islet-derived 3- α (reg3 α) has been reported to possess prognostic significance.¹¹ However, blood tests

are still not able to predict a patient's response to treatment, and these 6 biomarkers have not been validated in a blinded manner using samples from different institutions.

We therefore hypothesized that these 6 GVHD biomarkers, measured on samples obtained during treatment, would help discriminate between therapy-responsive and nonresponsive patients and predict survival. We used samples prospectively obtained on a multicenter, randomized, 4-arm phase 2 clinical trial for newly diagnosed acute GVHD (BMT CTN 0302). In brief, patients with newly diagnosed acute GVHD were randomized to receive methylprednisolone (2 mg/kg/day) plus 1 of 4 second agents (etanercept, mycophenolate mofetil, denileukin diftitox, or pentostatin), with the exception that patients who had received mycophenolate mofetil for GVHD prophylaxis within 7 days were randomized to one of the other arms. The results of this study have been reported previously.¹² We measured GVHD biomarker concentrations from serum samples obtained at treatment initiation (day 0), day 14, and day 28 of treatment. We then used logistic regression to create biomarker panels for each time point that assigned a relative weight to each individual biomarker such that the combined panel defined a threshold that best categorized patients on the basis of likelihood to experience the key clinical outcomes of day 28 treatment response or survival. We found that biomarker concentrations obtained during treatment were predictive of likelihood of both treatment response and survival either as a stand-alone test or in combination with clinical information.

Submitted January 9, 2012; accepted February 25, 2012. Prepublished online as *Blood* First Edition paper, March 1, 2012; DOI 10.1182/blood-2012-01-403063.

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.

The online version of the article contains a data supplement.

© 2012 by The American Society of Hematology

Table 1. Patient characteristics

Characteristic	Treatment arm				Total (N = 112)
	Etanercept (N = 29)	MMF (N = 29)	Denileukin diftitox (N = 26)	Pentostatin (N = 28)	
Female sex	10 (34)	11 (38)	9 (35)	12 (43)	42 (38)
Median age, y (range)	52 (20-70)	45 (13-63)	49 (11-65)	55 (35-68)	51 (11-70)
Disease					
AML/MDS	16 (55)	13 (45)	12 (46)	12 (43)	53 (47)
ALL	3 (10)	5 (17)	2 (7)	5 (18)	15 (13)
CML	0 (0)	2 (7)	1 (4)	0 (0)	3 (3)
Lymphoma	3 (10)	4 (14)	3 (11)	3 (11)	13 (12)
Other	7 (24)	5 (17)	8 (31)	8 (29)	28 (25)
Conditioning regimen					
Myeloablative	18 (62)	24 (83)	13 (50)	14 (50)	69 (62)
Nonmyeloablative	11 (38)	5 (17)	13 (50)	14 (50)	43 (38)
Donor type					
Related	13 (45)	16 (55)	9 (35)	12 (43)	50 (45)
Unrelated	16 (55)	13 (45)	17 (65)	16 (57)	62 (55)
Stem cell source					
Bone marrow	4 (14)	11 (38)	11 (42)	7 (25)	33 (30)
Peripheral blood	21 (72)	17 (59)	7 (27)	18 (64)	63 (56)
Umbilical cord blood	4 (14)	1 (3)	8 (31)	3 (11)	16 (14)
MMF prophylaxis					
Yes	9 (31)	0 (0)	12 (46)	7 (25)	28 (25)
No	20 (69)	29 (100)	14 (54)	21 (75)	84 (75)
Median days from transplantation to randomization (range)	30 (13-147)	29 (17-87)	35 (14-80)	36 (12-97)	32 (12-147)
Enrollment acute GVHD					
Grade I/II	24 (83)	20 (69)	18 (69)	21 (75)	83 (74)
Grade III/IV	5 (17)	9 (31)	8 (31)	7 (25)	29 (26)
Organ involvement at randomization					
Skin	25 (86)	23 (79)	20 (77)	24 (86)	92 (82)
GI tract (lower)	7 (24)	7 (24)	7 (27)	10 (36)	31 (28)
GI tract (upper)	4 (14)	9 (31)	8 (31)	8 (29)	29 (26)
Liver	4 (14)	6 (21)	3 (11)	3 (11)	16 (14)
Response to treatment at day 28					
CR/PR	17 (59)	22 (76)	22 (85)	20 (71)	81 (72)
NR	12 (41)	7 (24)	4 (15)	8 (29)	31 (28)
Survival status at day 180					
Alive	20 (69)	21 (72)	19 (73)	17 (61)	77 (69)
Dead	9 (31)	8 (28)	7 (27)	11 (39)	35 (31)

All values are presented as N (%).

AML indicates acute myeloid leukemia; MDS, myelodysplastic syndrome; ALL, acute lymphocytic leukemia; CML, chronic myelogenous leukemia; MMF, mycophenolate mofetil; GI, gastrointestinal; CR, complete response; PR, partial response; and NR, nonresponse.

Methods

Patients and samples

Peripheral blood samples were collected from patients who provided written informed consent in accordance with the Declaration of Helsinki at the time of study entry (day 0) and on days 14 and 28 of therapy. The study was approved by the institutional review boards of all participating clinical centers. We processed 7-mL peripheral blood samples at the clinical centers on the day of collection. The tubes were centrifuged at 1000-1300g for 10 minutes, and then serum was removed, distributed into cryovials, and frozen at a minimum of -20°C . The frozen serum samples were then batch-shipped quarterly to the National Heart, Lung, and Blood Institute Sample Repository and stored at -80°C for protocol-defined ancillary studies. At least 1 serum sample was available for analysis from 112 patients of 180 that participated in the clinical trial. Nine patients did not provide day 14 samples, and 8 patients did not provide day 28 samples. All available samples were included in the analysis. Clinical responses to treatment were the same as described previously and were scored by a study committee blinded to the treatment assignment.^{6,12} Enzyme-linked immunosorbent assays (ELISAs) were performed as reported previously.⁹⁻¹¹ Similarly, the ELISAs were performed in a blinded manner,

with the investigators blinded to all clinical information including outcomes.

Statistical methods

Logarithmic transformed biomarker values were used in all analyses because of skewness of the raw values. Spearman rank correlations were used to examine the association between the different biomarkers. Biomarker values at day 0, 14, and 28 were compared in a univariate manner between responders and nonresponders at day 28 or between patients alive or not at day 180 using side-by-side boxplots. Separate biomarker panels were constructed using biomarkers measured at day 0 and day 14 to predict nonresponse to therapy by day 28, and using biomarkers measured at day 0 and day 28 to predict mortality by 180 days. These panels were done by fitting a logistic regression model for each outcome using all the continuous biomarkers measured at the time of interest, and by constructing the composite biomarker panel as the linear combination of the logistic regression coefficients and the corresponding biomarker values. The intercepts for these models also were determined. Relative effects of each individual biomarker were summarized using odds ratios obtained from univariate logistic regression models. Because of the different scales of each continuous biomarker, the odds ratios were rescaled to correspond to a similar change in each biomarker, namely, a change from the median biomarker level to the third quartile. To further aid clinical interpretability,

an optimal threshold defining high versus low values for each panel was constructed based on maximizing the likelihood function in the logistic regression model. Univariate panel results were summarized using this high versus low definition with simple frequencies for response status and Kaplan-Meier curves for overall survival. The panels were evaluated in multivariate logistic regression models by adjusting for clinical variables previously identified as important predictors of day 28 nonresponse or day 180 mortality in the larger clinical study from which these samples were obtained.¹²

Results

Patient characteristics

The characteristics of the 112 patients (Table 1) who provided samples for this study did not significantly differ from the 180 patients who participated in the parent study with respect to any of the baseline clinical characteristics or clinical study outcomes. Samples were available from 112 patients at study entry (day 0) and 104 patients at day 14 and day 28.

Biomarker concentrations at GVHD study entry (day 0) independently predicted day 28 response status and day 180 survival status

Biomarker concentrations at day 0 did not significantly correlate with any of the patient characteristics listed in Table 1. We also found no significant differences between individual biomarker concentrations and the treatment arm assigned to the patients. However, we did find that the concentrations of the biomarkers correlated with each other; that is, high concentrations of one biomarker were frequently associated with high concentrations of other biomarkers. Most prominent of these were concentrations of TNFR1, which strongly correlated with concentrations of each of the other biomarkers ($P < .001$ for each biomarker) except for HGF (supplemental Table 1, available on the *Blood* Web site; see the Supplemental Materials link at the top of the online article). By contrast, concentrations of HGF only correlated with concentrations of IL-8 ($P = .006$; supplemental Table 1).

We first tested whether the concentrations of each of the biomarkers at study entry correlated with nonresponse status on day 28 of treatment (supplemental Figure 1). Mean concentrations of TNFR1 were significantly higher at study entry in patients who did not respond to treatment on day 28, but otherwise there were no significant differences between mean biomarker concentrations at study entry when categorized by day 28 response to treatment. Next, we used logistic regression to determine the odds ratio for an increase in the concentration of each biomarker individually for nonresponse at day 28. To facilitate comparisons, as mentioned in the methods, the odds ratios were calibrated across the various biomarkers so that each odds ratio corresponds to a similar change in each biomarker, which is a change from the median to the third quartile. We found that only higher concentrations of TNFR1 predicted for an increased likelihood of nonresponse at day 28 in univariate analysis (odds ratio, 1.51, $P = .028$; supplemental Table 2).

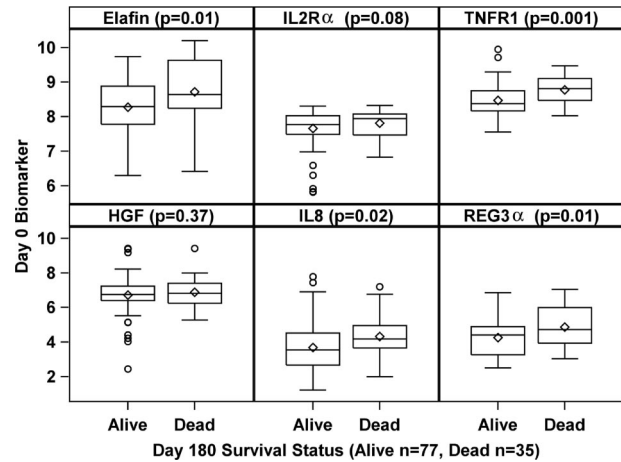


Figure 1. Biomarker concentrations at study entry (day 0) according to day 180 survival status. The boxplots show the range, median, and mean (diamond) graphed on a log-scale for each of the 6 biomarkers measured on day 0 according to the day 180 survival status (alive or dead). P values are shown for the comparison of median concentrations for each biomarker.

We similarly analyzed the day 0 concentrations of the 6 biomarkers using day 180 mortality as the end point. In contrast to the treatment response outcome where only TNFR1 concentrations were significantly different at day 0, the mean concentrations of multiple biomarkers (TNFR1, IL-8, elafin, and reg3 α) were significantly higher at study entry in patients who did not survive to day 180 from start of treatment compared with patients who were alive at day 180 (Figure 1). We used logistic regression again to determine the odds ratio for an increase in the concentration of each biomarker for mortality at day 180. Patients were significantly more likely not to survive to day 180 if they had high concentrations of TNFR1, IL-8, elafin, and reg3 α at study entry (supplemental Table 3).

Given that we are primarily interested in post-GVHD treatment survival, we then used logistic regression to create a formula that uses all 6 biomarker concentrations measured on day 0 to predict for mortality at day 180 (Table 2). In this formula, a relative weight is calculated for each biomarker. Thus, the day 0 biomarker panel is the sum of the value of the individual weighted biomarkers measured at study entry. Notably, although TNFR1 was the most important contributor to the day 0 biomarker panel, the combined panel outperformed the individual biomarkers. To aid in the clinical interpretation of the day 0 biomarker panel, we used the formula shown in Table 2 to identify a threshold that optimized the discrimination between mortality at day 180. Using the optimal threshold, 80 patients had a low day 0 panel and 32 had a high day 0 panel. As shown in Figure 2, the 32 patients with a high day 0 panel experienced worse outcomes (day 180 mortality, 56% vs 21%; $P < .0001$) corresponding to an odds ratio for death of 4.76. We had previously shown that GVHD grade at onset (III/IV vs I/II), stem cell source (peripheral blood, BM, or cord), and donor type (unrelated vs related) are predictors for day 180 mortality.¹² We therefore used multivariate analysis to test whether the day 0 biomarker panel possessed independent prognostic value for day 180 mortality after taking these clinical characteristics into account. As shown in Table

Table 2. Logistic regression formulas for biomarker panels and optimal thresholds

Biomarker panel	Formula	Optimal threshold
Day 0	$0.31 \cdot \ln(\text{elafin}) + 0.13 \cdot \ln(\text{IL-2R}\alpha) + 0.80 \cdot \ln(\text{TNFR1}) + 0.20 \cdot \ln(\text{IL-8}) + 0.02 \cdot \ln(\text{HGF}) + 0.19 \cdot \ln(\text{reg3}\alpha)$	12.8
Day 14	$0.98 \cdot \ln(\text{elafin}) + 0.19 \cdot \ln(\text{IL-2R}\alpha) - 0.86 \cdot \ln(\text{TNFR1}) + 0.26 \cdot \ln(\text{IL-8}) + 0.74 \cdot \ln(\text{HGF}) - 0.10 \cdot \ln(\text{reg3}\alpha)$	8
Day 28	$-0.17 \cdot \ln(\text{elafin}) + 0.23 \cdot \ln(\text{IL-2R}\alpha) + 2.27 \cdot \ln(\text{TNFR1}) + 0.19 \cdot \ln(\text{IL-8}) + 0.33 \cdot \ln(\text{HGF}) + 0.14 \cdot \ln(\text{reg3}\alpha)$	23.5

Table 3. Multivariate analysis of predictors for day 180 mortality at day 0

Variable	Clinical characteristics alone		Clinical characteristics plus biomarker panel	
	Odds ratio	P	Odds ratio	P
Onset GVHD grade (III/IV vs I/II)	1.11	.828	1.10	.849
Donor (unrelated vs related)	2.16	.098	2.05	.142
Stem cell source (cord vs peripheral blood)	0.57	.386	0.65	.533
Stem cell source (bone marrow vs peripheral blood)	0.57	.256	0.63	.380
Day 0 biomarker panel (high vs low)			4.61	< .001

3, the day 0 biomarker panel independently predicted mortality by day 180 (odds ratio, 4.61; $P < .001$). The corresponding receiver operating characteristic curve is shown in supplemental Figure 2.

Next, we evaluated how well the threshold value that optimized the discrimination for day 180 mortality predicted for day 28 nonresponse. As before, the 32 patients with a high day 0 panel experienced higher rate of nonresponse by day 28 (44% vs 21%; $P = .016$), corresponding to an odds ratio for treatment nonresponse of 2.88. We used multivariate analysis to adjust for GVHD grade at onset, the only clinical factor previously shown to predict day 28 nonresponse.¹² As shown in Table 4, the day 0 biomarker panel independently predicted nonresponse at day 28 (odds ratio, 2.98; $P = .017$). Finally, we repeated the statistical model with treatment arm assignment as a variable. There were no significant differences in the results when treatment assignment was included in the 2 regression models used in this study.

Biomarker concentrations at day 14 of treatment independently predict day 28 response status

We also were interested in whether biomarker concentrations measured during treatment provided prognostic information beyond what could be determined based on the patient’s clinical characteristics and observed clinical response. First, in this subset of patients from the parent study that provided samples for biomarker analysis, we confirmed, as reported previously,⁶ that response status on day 14 of treatment (nonresponse vs partial or complete response) strongly predicted nonresponse status on day 28 (odds ratio, 21.79; $P < .001$), whereas onset GVHD grade (III/IV vs I/II) did not (Table 5). We then analyzed each individual day 14 biomarker for day 28 nonresponse status. The mean concentrations of 3 biomarkers (elafin, HGF, and IL-8) were significantly different on day 14 based on treatment nonresponse at day 28 (supplemental Figure 3). As before, we used logistic regression to determine the odds ratio for an increase in the concentration of each individual biomarker for day 28 treatment nonresponse. Patients with higher concentrations of elafin, IL-8, and reg3α on day 14 were significantly more likely to demonstrate treatment failure at day 28 (supplemental Table 4). As before, logistic regression was used to identify the optimal threshold for a day 14 biomarker panel that gave maximal discrimination for treatment responsiveness at day 28 (Table 2). The 32 patients with a high day 14 biomarker panel were significantly more likely to demonstrate treatment nonresponse at day 28 than the 71 patients with a low day 14 biomarker panel (56% vs 17%; $P < .001$),

corresponding to an odds ratio of 6.32. Concentrations of elafin, a skin GVHD-specific biomarker, on day 14 were the most important contributor to the performance of the panel. This is perhaps not surprising, given that at study entry, skin GVHD was present in 82% of the patients.

We next used multivariate analysis to test whether the day 14 biomarker panel possessed independent prognostic value for day 28 treatment nonresponse after taking clinical characteristics into account. As shown in Table 5, both day 14 nonresponse status and the biomarker panel independently predicted day 28 nonresponse status. The corresponding receiver operating characteristic curve is shown in supplemental Figure 4. We again confirmed that inclusion of treatment assignment as a variable in logistic regression model did not significantly alter the results.

Biomarker concentrations at day 28 of treatment independently predict day 180 survival status

We also tested whether biomarker concentrations measured after 28 days of GVHD treatment provided prognostic survival information beyond what could be determined by the patient’s clinical characteristics. First, as in the previous analyses, we analyzed each individual day 28 biomarker for day 180 mortality. The mean concentrations of 5 biomarkers (TNFR1, IL-8, IL-2Rα, elafin, and reg3α) were significantly different on day 28 based on survival status at day 180 (supplemental Figure 5). Using the same logistic regression approach, we determined the odds ratio for an increase in the concentration of each individual biomarker for day 180 mortality and also created a day 28 biomarker panel that used the concentrations of each of the individual biomarkers. As shown in supplemental Table 5, patients with high concentrations of multiple individual biomarkers were significantly more likely to have died by day 180. The optimal threshold value for a high (N = 41) versus low (N = 63) day 28 biomarker panel was then determined. As shown in Figure 3, using this threshold resulted in an almost 40% difference in survival at day 180 (49% vs 87%; $P < .0001$), corresponding to an odds ratio of 7.22.

Next, we quantified the additional prognostic value the day 28 biomarker panel provided beyond what could be determined using the clinical characteristics alone. To do this, we first created a logistic regression model that included the previously reported clinical characteristics predictive of survival at day 180. These included GVHD grade at study entry, donor type, stem cell source,

Table 4. Multivariate analysis of predictors for day 28 nonresponse at day 0

Variable	Clinical characteristics alone		Clinical characteristics plus biomarker panel	
	Odds ratio	P	Odds ratio	P
Onset GVHD grade (III/IV vs I/II)	1.92	.155	2.03	.136
Day 0 biomarker panel (high vs low)			2.98	.017

Table 5. Multivariate analysis of predictors for day 28 response status at day 14

Variable	Clinical characteristics alone		Clinical characteristics plus biomarker panel	
	Odds ratio	P	Odds ratio	P
Onset GVHD grade (III/IV vs I/II)	1.17	.80	1.06	.931
Day 14 response status (NR vs CR/PR)	21.79	< .001	25.34	< .001
Day 14 biomarker panel (high vs low)			7.67	.002

and day 28 response status. As shown in Table 6, in this multivariate model, only day 28 response status predicted day 180 mortality (odds ratio, 7.22; $P < .001$). We then found that including the day 28 biomarker panel in this multivariate analysis provided an additional independent predictor of day 180 mortality (odds ratio, 7.43; $P < .001$). The corresponding receiver operating characteristic curve is shown in supplemental Figure 6. We again confirmed that inclusion of treatment assignment as a variable in logistic regression model did not significantly alter the results.

Discussion

Since the publication of the first GVHD biomarker panel,⁹ the potential utility of biomarkers to guide clinical therapy for patients with GVHD has been the source of considerable interest.¹³ In this report, we have demonstrated that biomarker panels measured during the course of GVHD treatment provide valuable prognostic information, independent and additive to the observable clinical status of the patient at the time of measurement. At each of 3 time points, start of GVHD treatment, 2 weeks into treatment, and 4 weeks into treatment, a 6-protein composite biomarker panel predicted for the highly important clinical outcomes of treatment response and mortality. Perhaps most importantly, after weighting each biomarker according to the logistic regression model we developed, a high GVHD biomarker panel measured at GVHD onset predicted for both death by day 180 and treatment failure at day 28. We have shown previously that although treatment failure by day 28 is predictive of mortality, it does not entirely account for survival outcomes at day 180.⁶ Thus, it seems that the day 0 concentra-

tions of several GVHD biomarkers contribute to predicting causes of death other than GVHD treatment failure. Because cause of death was not collected as part of this clinical trial, we can only speculate on the identity of these other causes, which might be late GVHD flares, or possibly other complications related to GVHD treatment. Further prospective clinical trials with the collection of additional samples and clinical data are needed to more fully explain this finding.

We recognize that there are some limitations to this analysis. First, the majority of patients had skin GVHD as the indication for study entry, and this may account for the relatively high contribution of the skin GVHD biomarker elafin to the performance of the day 14 biomarker panel. It is likely that a biomarker panel generated from patients with predominantly gastrointestinal involvement would have different characteristics. Thus, we included all 6 biomarkers when constructing the panels, recognizing that that larger prospective validation trials will be needed to ultimately determine which combination of biomarkers provides the greatest clinical utility. Second, patients who participated in the parent clinical trial may have other differences compared with the general acute GVHD population. For example, patients with active infections, such as CMV, or less easily discernible GVHD symptoms may not have enrolled on the clinical trial, providing additional justification for the need to conduct larger prospective validation trials before implementing biomarker assays into clinical management. Third, because of the very nature of the clinical trial that provided the samples for the analyses, all of the patients were treated with both steroids and 1 of 4 other drugs. We are not able to account for the potential differences that might be observed if only steroids had been administered as the treatment for newly diagnosed acute GVHD. However, it is reassuring that the inclusion of

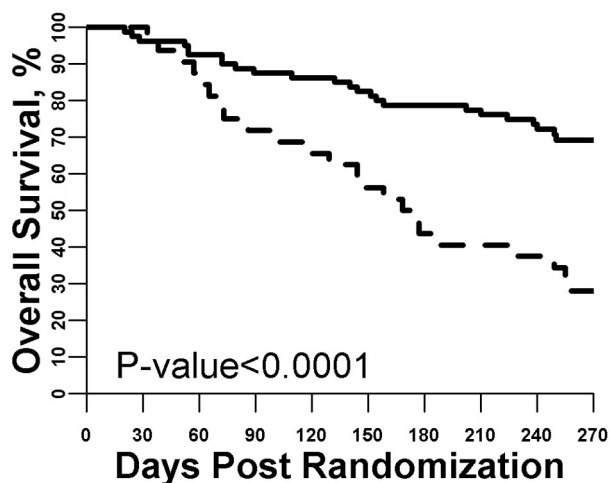


Figure 2. Survival from study entry stratified by day 0 biomarker panel. The optimized threshold for the biomarker panel measured at study entry was used to define a high versus low panel result. Patients with a high panel (dashed line, $N = 32$) were much more likely to die by day 180 compared with patients with a low panel (solid line, $N = 80$; $P < .001$).

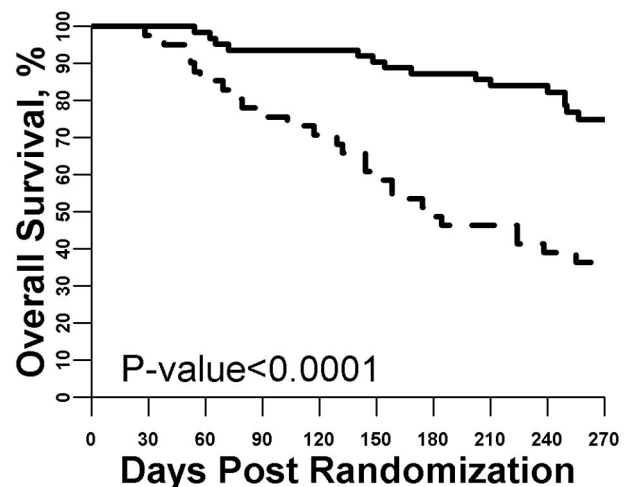


Figure 3. Survival from study entry stratified by day 28 biomarker panel. The optimized threshold for the biomarker panel measured at day 28 of GVHD treatment was used to define a high versus low panel result. Patients with a high panel (dashed line, $N = 41$) were much more likely to die by day 180 compared with patients with a low panel (solid line, $N = 63$; $P < .001$).

Table 6. Multivariate analysis of predictors for day 180 mortality at day 28

Variable	Clinical characteristics alone		Clinical characteristics plus biomarker panel	
	Odds ratio	P	Odds ratio	P
Onset GVHD grade (III/IV vs I/II)	1.03	.96	1.01	.98
Donor (unrelated vs related)	2.02	.18	2.09	.20
Stem cell source (cord vs peripheral blood)	0.66	.57	0.57	.48
Stem cell source (bone marrow vs peripheral blood)	0.80	.68	1.07	.92
Day 28 response status (NR vs CR/PR)	3.65	.01	3.58	.02
Day 28 biomarker panel (high vs low)			7.43	< .001

NR indicates nonresponse; CR, complete response; and PR, partial response.

treatment assignment in the statistical modeling did not significantly alter the results. Fourth, there might be other clinical factors, such as rapidity of treatment response, that were not captured as part of this study but that might be helpful in predicting treatment outcomes and survival. Finally, although we did not find any evidence that the patients who provided samples were significantly different from the overall patient population, the failure to obtain samples from all the patients who participated in the clinical trial led to a relatively small number of patients available for analysis. Nonetheless, the magnitude of the odds ratios and highly statistically significant results strongly suggest that measuring biomarker concentrations during treatment will prove clinically useful in other GVHD patient populations.

The validity of the results is strengthened by several factors. First, the investigators were blinded to the patient outcomes. Second, the samples were prospectively obtained from patients treated at multiple centers. Although the biomarker concentrations were measured long after treatment on the clinical trial was completed, the technology used, ELISA assays, is widely available and there is no inherent barrier to real-time measurement of the individual proteins that were included in the biomarker panels. Thus, incorporation of ELISA-based GVHD biomarker panels into clinical care should be highly feasible.

This study represents the first demonstration of how GVHD biomarker panels may ultimately be incorporated into clinical care. We envision that measuring GVHD biomarker concentrations will be useful to the clinical management of patients undergoing treatment similar to tumor markers that are well established for the management of certain cancers, such as blood measurements of α -fetoprotein, human chorionic gonadotropin, and lactate dehydrogenase for staging, prognosis determination, and therapy monitoring of testicular cancer,¹⁴ or serum measurements of CA 15-3, CA 27.29, or carcinoembryonic antigen to guide treatment for breast cancer.¹⁵ Cancer biomarkers supplement clinical measures of tumor activity at diagnosis, or residual tumor activity during treatment, whereas GVHD biomarkers seem to supplement clinical measures of disease activity such as degree of skin rash or volume of diarrhea. Therefore, one possible application of a GVHD biomarker panel is to use a threshold to categorize patients as having a high or low biomarker panel to risk stratify for GVHD outcomes, a strategy that may prove useful for the development of treatment algorithms at the onset of GVHD or during treatment. Along these same lines, it may ultimately be possible to use biomarker panel results to predict the likelihood of an outcome for an individual patient, the goal of personalized medicine. To illustrate this point, we selected a patient with a high day 14 biomarker panel result and applied the formula shown in Table 2 to determine the probability of nonresponse on day 28. For this specific patient, who was in response on day 14, we predicted a 73% likelihood of nonresponse on

day 28, despite their favorable day 14 clinical response. In actual fact, the patient was no longer in response on day 28. Information of this sort may be useful in guiding clinical decisions. For example, in cases similar to the specific patient just presented, the biomarker panel result could highlight that residual GVHD is greater than suspected from the clinical symptoms, and therefore, a slow steroid taper may be indicated. The inclusion of both clinical and laboratory values in future GVHD treatment algorithms may therefore help determine the optimal therapy for individual patients.

Acknowledgments

The authors thank the participating centers and coinvestigators for the BMT CTN 0302 study: MD Anderson Cancer Center (Amin M. Alousi), Johns Hopkins University (Javier Bolaños-Meade and Georgia Vogelsang), University of Minnesota (Daniel Weisdorf), Dana-Farber Brigham Women's Partners (Vincent Ho, Robert Soiffer, and Joseph Antin), University of Michigan (John Levine and James L. Ferrara), Oregon Health Sciences University (Eneida Nemecek), University of Florida College of Medicine (John Wingard), University of Pennsylvania Hospital Center (Steven Goldstein and Edward Stadtmauer), University of Nebraska Medical Center (Marcel Devetten), Washington University of St. Louis (John DiPersio and Peter Westervelt), Stanford Hospital and Clinics (Laura Johnston), University of California at San Diego (Edward Ball), Duke University Medical Center (Nelson Chao and Joanne Kurtzberg), University Hospitals of Cleveland (Case Western Reserve University; Hillard Lazarus), Memorial Sloan-Kettering Cancer Center (Nancy Kernan and Miguel-Angel Perales), Texas Transplant Institute (Carlos Bachier, Michael Grimley, and Paul Shaughnessy), City of Hope National Medical Center (Pablo Parker), Fred Hutchinson Cancer Research Center (Richard Nash), and Hackensack University Medical Center (Joel Brochstein).

This work was supported in part by the Blood and Marrow Transplant Clinical Trials grant U01HL069294 from the National Heart, Lung, and Blood Institute and the National Cancer Institute and by the National Institute of Allergy and Infectious Diseases and National Institutes of Health National Marrow Donor Program subcontract 065528.

Authorship

Contribution: J.E.L. and S.P. designed and performed research, analyzed and interpreted data, and wrote the paper; D.J.W. and J.L.M.F. analyzed and interpreted data and wrote the paper; B.R.L. designed the research, performed statistical analysis, analyzed and

interpreted the data, and wrote the paper; J.W. performed statistical analysis, analyzed and interpreted the data, and wrote the paper; and A.M.A., J.B.-M., and V.T.H. wrote the paper.

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

Correspondence: John E. Levine or Sophie Paczesny, Blood and Marrow Transplant Program, University of Michigan Comprehensive Cancer Center, Rm 6410, 1500 E Medical Center Dr, Ann Arbor, MI 48109-5942; e-mail: jelevine@med.umich.edu or sophiep@med.umich.edu.

References

- Cutler C, Antin JH. Manifestation and treatment of acute graft-versus-host-disease. In: Appelbaum F, Forman SJ, Negrin RS, Blume KG, eds. *Thomas' Hematopoietic Cell Transplantation*. Oxford: Blackwell Publishing Ltd; 2009:1287-1303.
- Ferrara JL, Levine JE, Reddy P, Holler E. Graft-versus-host disease. *Lancet*. 2009;373(9674):1550-1561.
- Welniak LA, Blazar BR, Murphy WJ. Immunobiology of allogeneic hematopoietic stem cell transplantation. *Annu Rev Immunol*. 2007;25:139-170.
- 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.
- Deeg HJ. How I treat refractory acute GVHD. *Blood*. 2007;109(10):4119-4126.
- Levine JE, Logan B, Wu J, et al. Graft-versus-host disease treatment: predictors of survival. *Biol Blood Marrow Transplant*. 2010;16(12):1693-1699.
- MacMillan ML, DeFor TE, Weisdorf DJ. The best endpoint for acute GVHD treatment trials. *Blood*. 2010;115(26):5412-5417.
- Saliba RM, Couriel DR, Giralt S, et al. Prognostic value of response after upfront therapy for acute GVHD. *Bone Marrow Transplant*. 2012;47(1):125-131.
- Paczesny S, Krijanovski OI, Braun TM, et al. A biomarker panel for acute graft-versus-host disease. *Blood*. 2009;113(2):273-278.
- Paczesny S, Braun TM, Levine JE, et al. Elafin is a biomarker of graft-versus-host disease of the skin. *Sci Transl Med*. 2010;2(13):13ra12.
- 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):6702-8.
- Alousi AM, Weisdorf DJ, Logan BR, et al. Etanercept, mycophenolate, denileukin or pentostatin plus corticosteroids for acute graft vs. host disease: a randomized phase II trial from the BMT CTN. *Blood*. 2009;114(3):511-517.
- Socié G. Graft-versus-host disease: proteomics comes of age. *Blood*. 2009;113(2):271-272.
- Sturgeon CM, Duffy MJ, Stenman UH, et al. National Academy of Clinical Biochemistry laboratory medicine practice guidelines for use of tumor markers in testicular, prostate, colorectal, breast, and ovarian cancers. *Clin Chem*. 2008;54(12):e11-e79.
- Harris L, Fritsche H, Mennel R, et al. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol*. 2007;25(33):5287-5312.