

TRANSPLANTATION

CME Article

MAGIC biomarkers predict long-term outcomes for steroid-resistant acute GVHD

Hannah Major-Monfried,^{1,*} Anne S. Renteria,^{1,*} Attaphol Pawarode,² Pavan Reddy,² Francis Ayuk,³ Ernst Holler,⁴ Yvonne A. Efebera,⁵ William J. Hogan,⁶ Matthias Wöflfl,⁷ Muna Qayed,⁸ Elizabeth O. Hexner,⁹ Kitsada Wudhikarn,¹⁰ Rainer Ordemann,¹¹ Rachel Young,¹ Jay Shah,¹ Matthew J. Hartwell,¹ Mohammed S. Chaudhry,¹ Mina Aziz,¹ Aaron Etra,¹ Gregory A. Yanik,² Nicolaus Kröger,³ Daniela Weber,⁴ Yi-Bin Chen,¹² Ryotaro Nakamura,¹³ Wolf Rösler,¹⁴ Carrie L. Kitko,¹⁵ Andrew C. Harris,¹⁶ Michael Pulsipher,¹⁷ Ran Reshef,¹⁸ Steven Kowalyk,¹ George Morales,¹ Ivan Torres,¹ Umut Özbek,^{19,†} James L. M. Ferrara,^{1,†} and John E. Levine^{1,†}

¹Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY; ²Blood and Marrow Transplantation Program, University of Michigan, Ann Arbor, MI; ³Department of Stem Cell Transplantation, University Medical Center, Hamburg-Eppendorf, Germany; ⁴Blood and Marrow Transplantation Program, University of Regensburg, Regensburg, Germany; ⁵Blood and Marrow Transplantation Program, Ohio State University, Columbus, OH; ⁶Blood and Marrow Transplantation Program, Mayo Clinic, Rochester, MN; ⁷Pediatric Blood and Marrow Transplantation Program, Children's Hospital, University of Würzburg, Würzburg, Germany; ⁸Pediatric Blood and Marrow Transplantation Program, Aflac Cancer and Blood Disorders Center, Emory University and Children's Healthcare of Atlanta, Atlanta, GA; ⁹Blood and Marrow Transplantation Program, University of Pennsylvania, Philadelphia, PA; ¹⁰Blood and Marrow Transplantation Program, Chulalongkorn University, Bangkok, Thailand; ¹¹Medical Department 1, University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany; ¹²Bone Marrow Transplantation Program, Massachusetts General Hospital, Boston, MA; ¹³Hematology and Hematopoietic Cell Transplantation, City of Hope Medical Center, Duarte, CA; ¹⁴Department of Internal Medicine 5, Hematology/Oncology, University Hospital Erlangen-Nuremberg, Erlangen, Germany; ¹⁵Pediatric Blood and Marrow Transplantation Program, Vanderbilt University Medical Center, Nashville, TN; ¹⁶Blood and Marrow Transplantation Program, University of Utah, Salt Lake City, UT; ¹⁷Blood and Marrow Transplantation Program, Children's Hospital of Los Angeles, Los Angeles, CA; ¹⁸Blood and Marrow Transplantation Program, Columbia University, New York, NY; and ¹⁹Biostatistics Shared Resource Facility, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY

KEY POINTS

- Biomarker scores generated after 1 week of steroid treatment of GVHD are prognostic.
- Biomarkers reflect prognosis better than early clinical response to GVHD treatment.

Acute graft-versus-host disease (GVHD) is treated with systemic corticosteroid immunosuppression. Clinical response after 1 week of therapy often guides further treatment decisions, but long-term outcomes vary widely among centers, and more accurate predictive tests are urgently needed. We analyzed clinical data and blood samples taken 1 week after systemic treatment of GVHD from 507 patients from 17 centers of the Mount Sinai Acute GVHD International Consortium (MAGIC), dividing them into a test cohort (n = 236) and 2 validation cohorts separated in time (n = 142 and n = 129). Initial response to systemic steroids correlated with response at 4 weeks, 1-year nonrelapse mortality (NRM), and overall survival (OS). A previously validated algorithm of 2 MAGIC biomarkers (ST2 and REG3α) consistently separated steroid-resistant patients into 2 groups with dramatically different NRM and OS (P < .001 for all 3 cohorts). High biomarker probability, resistance to steroids, and GVHD severity (Minnesota risk) were all significant predictors of NRM in multivariate analysis. A direct comparison of receiver operating characteristic curves showed that the area under the curve for biomarker probability (0.82) was significantly greater than that for steroid response (0.68, P = .004) and for Minnesota risk (0.72, P = .005). In conclusion, MAGIC biomarker probabilities generated after 1 week of systemic treatment of GVHD predict long-term outcomes in steroid-resistant GVHD better than clinical criteria and should prove useful in developing better treatment strategies. (*Blood*. 2018;131(25):2846-2855)

resistance to steroids, and GVHD severity (Minnesota risk) were all significant predictors of NRM in multivariate analysis. A direct comparison of receiver operating characteristic curves showed that the area under the curve for biomarker probability (0.82) was significantly greater than that for steroid response (0.68, P = .004) and for Minnesota risk (0.72, P = .005). In conclusion, MAGIC biomarker probabilities generated after 1 week of systemic treatment of GVHD predict long-term outcomes in steroid-resistant GVHD better than clinical criteria and should prove useful in developing better treatment strategies. (*Blood*. 2018;131(25):2846-2855)



JOINTLY ACCREDITED PROVIDER™
INTERPROFESSIONAL CONTINUING EDUCATION

Medscape Continuing Medical Education online

In support of improving patient care, this activity has been planned and implemented by Medscape, LLC and the American Society of Hematology. Medscape, LLC is jointly accredited by the Accreditation Council for Continuing Medical Education (ACCME), the Accreditation Council for Pharmacy Education (ACPE), and the American Nurses Credentialing Center (ANCC), to provide continuing education for the healthcare team.

Medscape, LLC designates this Journal-based CME activity for a maximum of 1.00 AMA PRA Category 1 Credit(s)[™]. Physicians should claim only the credit commensurate with the extent of their participation in the activity.

All other clinicians completing this activity will be issued a certificate of participation. To participate in this journal CME activity: (1) review the learning objectives and author disclosures; (2) study the education content; (3) take the post-test with a 75% minimum passing score and complete the evaluation at <http://www.medscape.org/journal/blood>; and (4) view/print certificate. For CME questions, see page 2870.

Disclosures

CME questions author Laurie Barclay, freelance writer and reviewer, Medscape, LLC, owns stock, stock options, or bonds from Pfizer. Authors Umut Özbek, James L. M. Ferrara, and John E. Levine are joint inventors on a graft-versus-host disease biomarker patent. Associate Editor Robert Zeiser and the remaining authors declare no competing financial interests.

Learning objectives

Upon completion of this activity, participants will be able to:

1. Describe the prognostic value of biomarker scores and clinical features after 1 week of steroid treatment for graft-versus-host disease (GVHD), based on a clinical study
2. Compare the prognostic value of these biomarker scores with that of early clinical response to GVHD treatment
3. Determine the clinical implications of the prognostic value of biomarker scores generated after 1 week of steroid treatment for GVHD

Release date: June 21, 2018; Expiration date: June 21, 2019

Introduction

Improvements in survival following allogeneic hematopoietic cell transplantation (HCT) have led to its increasing use to cure hematologic malignancies and other disorders,¹ but the leading cause of nonrelapse mortality (NRM) after HCT remains graft-versus-host disease (GVHD).^{2,3} Many patients do not respond to primary therapy, high-dose systemic corticosteroids; therefore, survival for patients with treatment-resistant GVHD remains particularly poor.⁴ The clinical response after 4 weeks of treatment is a validated surrogate for long-term survival,⁵ but physicians usually cannot wait 1 month before deciding whether to modify therapy, particularly for patients who do not achieve or maintain a convincing clinical response. In clinical practice, 1 week of treatment is commonly used to determine escalation or de-escalation of immunosuppressive therapy, but this early response correlates poorly with long-term outcomes.⁶ Patient who do not respond early to systemic steroids have a generally poor prognosis, but results are inconsistent among transplant centers, and biomarkers that accurately predict long-term outcomes in this highly immunosuppressed population are urgently needed.

The Mount Sinai Acute GVHD International Consortium (MAGIC) comprises 17 HCT centers and was established to provide consistent multicenter monitoring of acute GVHD severity during treatment, as well as to obtain samples that could be interrogated for potential predictive biomarkers. We have previously validated an algorithm that uses the serum concentrations of the GVHD biomarkers suppressor of tumorigenicity-2 (ST2) and regenerating islet-derived protein 3- α (REG3 α) to generate a probability for NRM and predict resistance to treatment.^{7,8} In this study, we first determined the extent to which early clinical responses to steroid treatment could predict long-term outcome of patients with acute GVHD. We then evaluated biomarkers obtained at the time of the clinical evaluation. We hypothesized that the probabilities would predict long-term outcomes, even if the biomarkers are measured at a time when the initial response to treatment was already known.

Methods

Study design

Patients from the 17 centers in MAGIC underwent their first allogeneic HCT between May 2001 and December 2016 and

provided blood samples for a biorepository 7 days after initiation of corticosteroid treatment for newly diagnosed acute GVHD. All patients consented to participation in an Institutional Review Board–approved protocol at each MAGIC center (supplemental Table 1, available on the *Blood* Web site). Patients transplanted before 2016 whose data and GVHD-onset samples had contributed to the development of the initial algorithm formed the test cohort (n = 236),^{7,8} whereas patients not previously analyzed formed the first validation cohort (n = 142). Patients transplanted in 2016 (n = 129) formed the second validation cohort.

GVHD was staged using published guidelines.⁹ MAGIC centers were trained via webinar in the use of these guidelines, and data entry personnel needed to pass a GVHD staging test prior to enrolling patients in the MAGIC protocol. GVHD staging guidance was reinforced in monthly webinars.

Nonrelapse deaths were considered related to GVHD only if the patient died of GVHD itself or from an infection that developed while receiving systemic steroids (≥ 10 mg of prednisone daily or equivalent) for the treatment of GVHD. Noninfectious contributing causes of death from GVHD included cardiac and pulmonary events, multiorgan failure, and hemorrhage. Clinical response to treatment was determined at 1 and 4 weeks after the start of treatment, according to published criteria.¹⁰ Complete response (CR) was defined as the complete resolution of acute GVHD manifestations in all organs. Partial response (PR) was defined as improvement, but not complete resolution, of GVHD in all initially affected organs without new target organ involvement. Nonresponse was defined as all other responses or death before response assessment. Relapse risk was assessed according to published criteria.¹¹

Biomarker determination and statistical analyses

Samples were shipped to a central laboratory where they were analyzed in batches for ST2 and REG3 α by enzyme-linked immunosorbent assay, as previously described.^{12,13} We then created a competing risks model, with relapse as the competing risk, that predicted 6-month NRM after 1 week of systemic GVHD treatment using the concentrations of ST2

Table 1. Patient characteristics

Characteristic	Test cohort (n = 236)	Validation cohort 1 (n = 142)	Validation cohort 2 (n = 129)	P
Median age, y (range)	51 (1-73)	49.5 (1-74)	48 (1-74)	.840
Indication for HCT, n (%)				.004
Acute leukemia	121 (51.3)	73 (51.4)	65 (50.4)	
MDS/MPN	48 (20.3)	24 (16.9)	43 (33.2)	
Lymphoma	38 (16.1)	19 (13.4)	6 (4.7)	
Other malignant	22 (9.3)	20 (14.1)	9 (7.0)	
Nonmalignant	7 (3.0)	6 (4.2)	6 (4.7)	
Disease risk index at HCT, n (%)				.326
Low	16 (6.8)	7 (4.9)	3 (2.3)	
Intermediate	120 (50.8)	78 (54.9)	81 (62.8)	
High	61 (25.8)	38 (26.8)	31 (24.0)	
Very high	20 (8.5)	7 (4.9)	8 (6.2)	
Unknown	19 (8.1)	12 (8.5)	6 (4.7)	
Donor type, n (%)				.529
Related	67 (28.4)	46 (32.4)	34 (26.4)	
Unrelated	169 (71.6)	96 (67.6)	95 (73.6)	
HLA match, n (%)				.485
Matched	159 (67.4)	95 (66.9)	94 (72.9)	
Mismatched	77 (32.6)	47 (33.1)	35 (27.1)	
Stem cell source, n (%)				.433
Marrow	36 (15.3)	30 (21.1)	19 (14.7)	
Peripheral blood	179 (75.8)	104 (73.2)	101 (78.3)	
Cord blood	21 (8.9)	8 (5.7)	9 (7.0)	
Conditioning regimen intensity, n (%)				.007
Full	194 (82.2)	104 (73.2)	88 (68.2)	
Reduced	42 (17.8)	38 (26.8)	41 (31.8)	
Antithymocyte globulin in conditioning, n (%)				.0001
Yes	58 (24.6)	30 (21.1)	55 (42.6)	
No	178 (75.4)	112 (78.9)	74 (57.4)	
GVHD prophylaxis, n (%)				.024
CNI/MTX ± other	144 (61.0)	84 (59.2)	61 (47.3)	
CNI/MMF ± other	77 (32.6)	47 (33.1)	47 (36.4)	
CNI/sirolimus	1 (0.4)	0 (0)	2 (1.6)	
Posttransplant cyclophosphamide ± other	8 (3.4)	9 (6.3)	13 (10)	
T-cell depleted	2 (0.9)	0 (0)	2 (1.6)	
Other	4 (1.7)	2 (1.4)	4 (3.1)	
Onset GVHD: median day (range)	26 (9-275)	31 (7-204)	26 (8-180)	.0005
Onset GVHD: organ distribution, n (%)				.891
Isolated skin	116 (49.2)	67 (47.2)	68 (52.7)	
Isolated GI (upper and/or lower)	63 (26.7)	39 (27.5)	34 (26.3)	
Isolated liver	2 (0.8)	3 (2.1)	2 (1.6)	
≥2 organs involved	55 (23.3)	33 (23.2)	25 (19.4)	
Onset GVHD grade, n (%)				.112
1	73 (30.9)	39 (27.5)	51 (39.6)	
2	105 (44.5)	64 (45.1)	57 (44.2)	
3	50 (21.2)	31 (21.8)	14 (10.8)	
4	8 (3.4)	8 (5.6)	7 (5.4)	

CNI, calcineurin inhibitor; GI, gastrointestinal; MMF, mycophenolic acid; MTX, methotrexate.

Table 1. (continued)

Characteristic	Test cohort (n = 236)	Validation cohort 1 (n = 142)	Validation cohort 2 (n = 129)	P
Minnesota risk score (onset), n (%)				.016
Standard	183 (77.5)	107 (75.4)	114 (88.4)	
High	53 (22.5)	35 (24.6)	15 (11.6)	
Week 1 response, n (%)				.674
CR or PR	114 (48.3)	62 (43.7)	61 (47.3)	
Nonresponse	122 (51.7)	80 (56.3)	68 (52.7)	
Minnesota risk score (week 1), n (%)				.045
Standard	187 (79.2)	110 (77.5)	114 (88.4)	
High	49 (20.8)	32 (22.5)	15 (11.6)	
Long-term outcomes by cohort, %				
1-year NRM	31.3	28.9	20.5	.084
1-year relapse rate	20.3	16.3	8.9	.030
1-year OS	56.9	59.8	72.2	.019

CNI, calcineurin inhibitor; GI, gastrointestinal; MMF, mycophenolic acid; MTX, methotrexate.

and REG3 α . We compared the performance of this model with the previously published MAGIC prediction model by calculating the area under the curve (AUC) of the receiver operating characteristic curve of each model in the validation cohort. The AUC was the same for both models (0.82, $P = .977$); therefore, we used the previously published MAGIC prediction model $\log[-\log(1-\hat{p})] = -11.263 + 1.844(\log_{10}ST2) + 0.577(\log_{10}REG3\alpha)$ to calculate the probability value (\hat{p}) for each patient.⁸ An unsupervised learning algorithm, K-medoids clustering, which maximized the differences between groups while minimizing the differences in probabilities within each group, was used to identify the threshold that best separated test cohort patients into 2 groups.¹⁴

Clinical characteristics of patients between cohorts were compared using the χ^2 or Wilcoxon rank-sum test, as appropriate. Competing risks regression, with relapse as the competing risk, was used to model 1-year NRM, with early treatment response, clinical severity (Minnesota risk staging),¹⁰ and biomarkers as predictors. Patients with complete resolution of GVHD symptoms after 1 week of treatment were assigned to Minnesota standard risk. Logistic regression was used to model week 4 resistance to treatment. Differences in the cumulative incidence of NRM and relapse between groups were calculated by the Gray test. OS was estimated by the Kaplan-Meier method, and differences between groups were calculated using the log-rank test. Areas under the receiver operating characteristic curves were compared using the DeLong method.¹⁵ Univariate analyses for NRM were performed on the combined validation cohort patients and included pretransplant characteristics that are important risk factors for GVHD (Table 1), clinical severity after 1 week of treatment, initial response to treatment, and either biomarker concentrations or the categorical variable of the probability group. Multivariate analyses included all variables that are risk factors for GVHD and that were statistically significant on univariate analysis and were performed on the combined validation cohort patients. All analyses were performed using R statistical package version 3.4.0 (R Development Core Team 2017). Error bars represent the standard error of proportion in all figures in which error bars are shown.

Results

Patient characteristics

Clinical data and samples were available from 507 patients with acute GVHD who were treated with systemic corticosteroids. Patients were divided into a test cohort ($n = 236$) and 2 validation cohorts ($n = 142$ and $n = 129$). The median starting dose of steroids was 2.0 mg/kg per day for patients with grade 2-4 disease and 1.0 mg/kg per day for grade 1 disease. All GVHD treatments are listed in supplemental Table 2.

Clinical response after 1 week of treatment predicts outcomes

Because the clinical response after 1 week of systemic steroid treatment often guides further treatment,^{16,17} we first determined whether clinical response alone could predict NRM at 1 year. Patients with CRs and PRs in the test cohort had similar NRM (supplemental Figure 1), and we categorized these patients as early treatment sensitive, whereas all other patients were categorized as early treatment resistant. NRM at 1 year was significantly higher in patients with early treatment resistance (Figure 1A). Relapse did not consistently correlate with response (supplemental Table 3), with the result that the early treatment resistance group experienced significantly worse OS (Figure 1A). Some early treatment-resistant patients eventually showed responses to treatment by 4 weeks, an important surrogate end point for long-term outcomes,^{5,6,18} but most early treatment-resistant patients remained resistant at 4 weeks. Similarly, a significant minority of early treatment-sensitive patients became resistant to treatment at 4 weeks. Results were the same in both validation cohorts, demonstrating that the response to systemic steroids at 1 week can reliably predict long-term outcomes (Figure 1B-C).

Biomarker stratification

We have recently shown that serum biomarker concentrations can be used to predict long-term outcomes of patients at the onset of GVHD.^{7,8} We hypothesized that an algorithm of the same 2 biomarkers, ST2 and REG3 α , would predict long-term outcomes

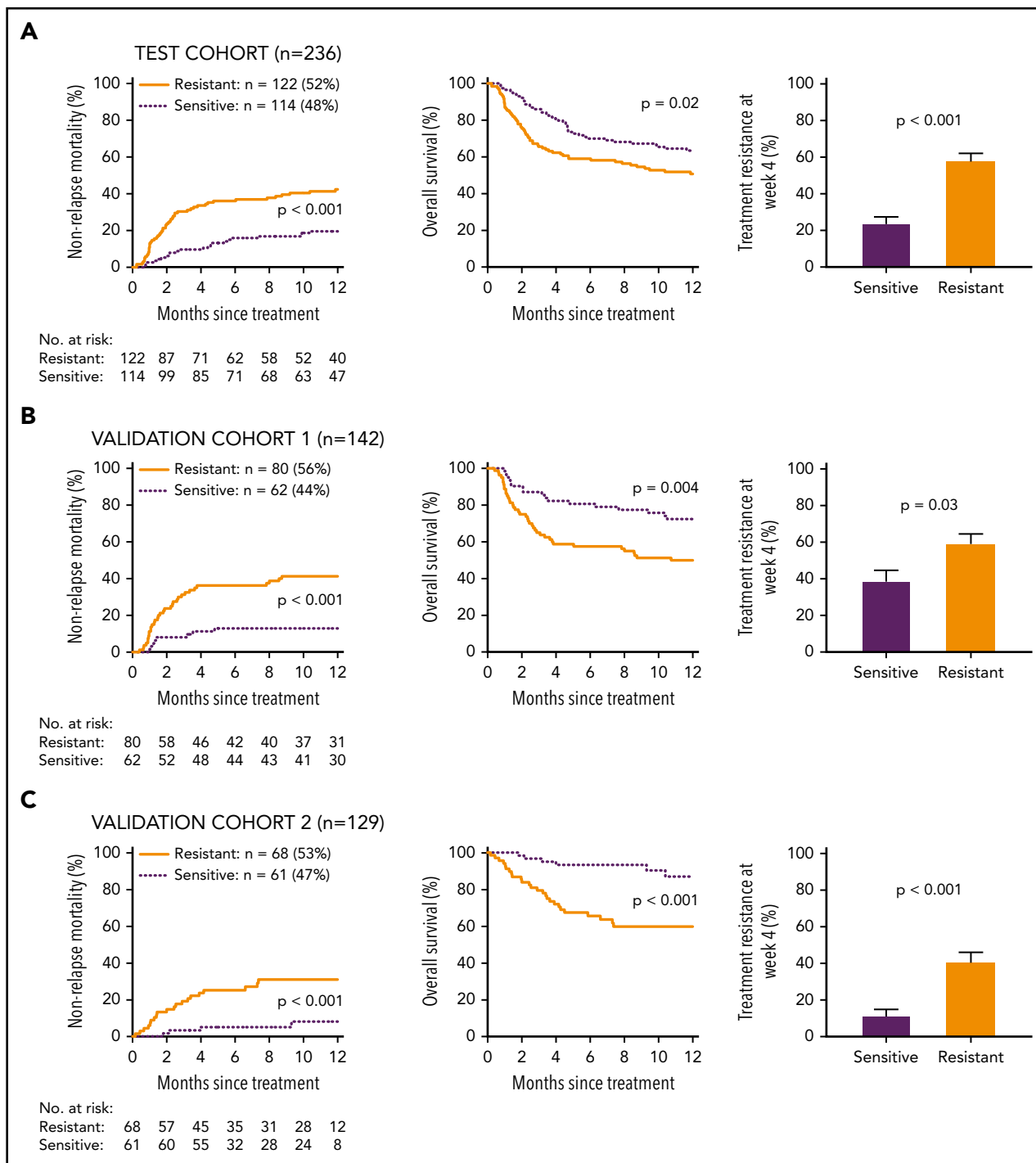


Figure 1. Long-term outcomes by clinical response to 1 week of treatment in all patients. Patients were divided into 2 groups based on response to treatment: early treatment sensitive (ETS; dotted line) and early treatment resistant (ETR; solid line). (A) Test cohort (n = 236). Twelve-month cumulative incidence of NRM (ETS 20% vs ETR 42%, $P < .001$) and OS (ETS 63% vs ETR 51%, $P = .02$) and proportion of patients resistant to treatment at week 4 (ETS 24% vs ETR 57%, $P < .001$). (B) Validation cohort 1 (n = 142). Twelve-month cumulative incidence of NRM (ETS 13% vs ETR 41%, $P < .001$) and OS (ETS 72% vs ETR 50%, $P = .004$) and proportion of patients resistant to treatment at week 4 (ETS 39% vs ETR 59%, $P = .03$). (C) Validation cohort 2 (n = 129). Twelve-month cumulative incidence of NRM (ETS 8% vs ETR 31%, $P = .001$) and OS (ETS 87% vs ETR 60%, $P < .001$) and proportion of patients resistant to treatment at week 4 (ETS 11% vs ETR 41%, $P < .001$).

when measured in patients after 1 week of treatment, when the early clinical response was already known. Because the accuracy of the newly derived algorithm was the same as the previously validated algorithm (see “Methods”), we used the original algorithm in all of our analyses. We first determined whether patients could be segregated into 2 groups (low and high

probability) based only on the predicted probabilities of NRM generated by the biomarker algorithm and without reference to any known clinical characteristics or outcomes (see “Methods”). A threshold of $\hat{p} \leq 0.291$ best separated patients into groups with low probability (n = 143, 61%) and high probability (n = 93, 39%) (supplemental Figure 2).

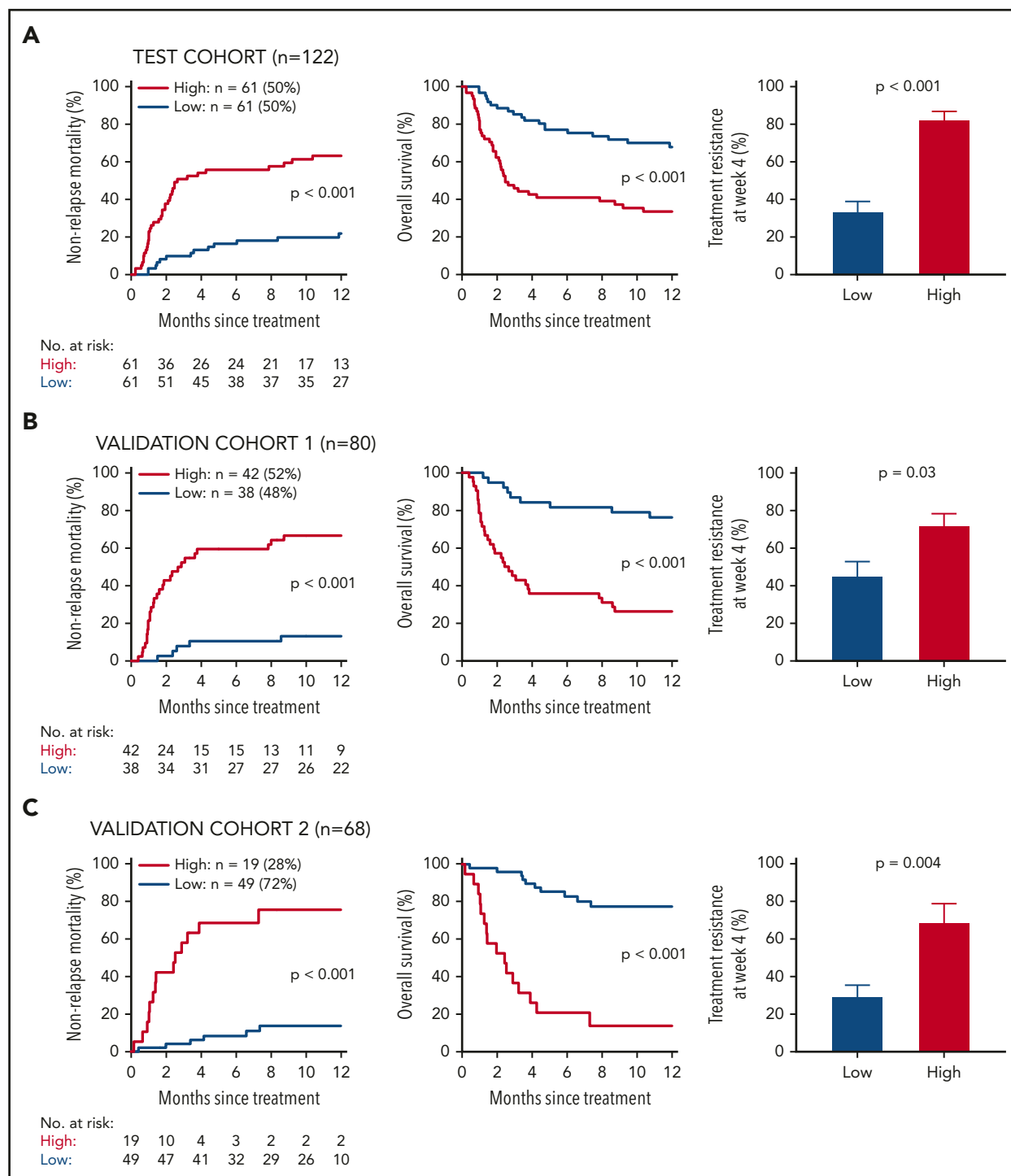


Figure 2. Long-term outcomes by biomarker probabilities in early treatment-resistant patients. Early treatment-resistant patients were subdivided based on biomarker probabilities into low and high groups. (A) Test cohort of patients (n = 122). Twelve-month cumulative incidence of NRM (low 22% vs high 63%, $P < .001$) and OS (low 68% vs high 34%, $P < .001$) and proportion of patients resistant to treatment at week 4 (low 33% vs high 82%, $P < .001$). (B) Validation cohort 1 (n = 80). Twelve-month cumulative incidence of NRM (low 13% vs high 67%, $P < .001$) and OS (low 76% vs high 26%, $P < .001$) and proportion of patients resistant to treatment at week 4 (low 45% vs high 71%, $P = .03$). (C) Validation cohort 2 (n = 68). Twelve-month cumulative incidence of NRM (low 14% vs high 75%, $P < .001$) and OS (low 78% vs high 14%, $P < .001$) and proportion of patients resistant to treatment at week 4 (low 29% vs high 68%, $P = .004$).

Biomarkers predict NRM within response groups

We evaluated the long-term outcomes of early treatment-resistant patients by their probability status. The algorithm analysis identified an unexpectedly large proportion (48%-72%) of the early treatment-resistant patients as low probability

and who experienced strikingly less NRM than the high-probability group in all 3 cohorts (Figure 2). Relapse was not consistently different between probability groups (supplemental Table 3), resulting in dramatically better OS in the low-probability group that was similar to that of steroid-sensitive patients.

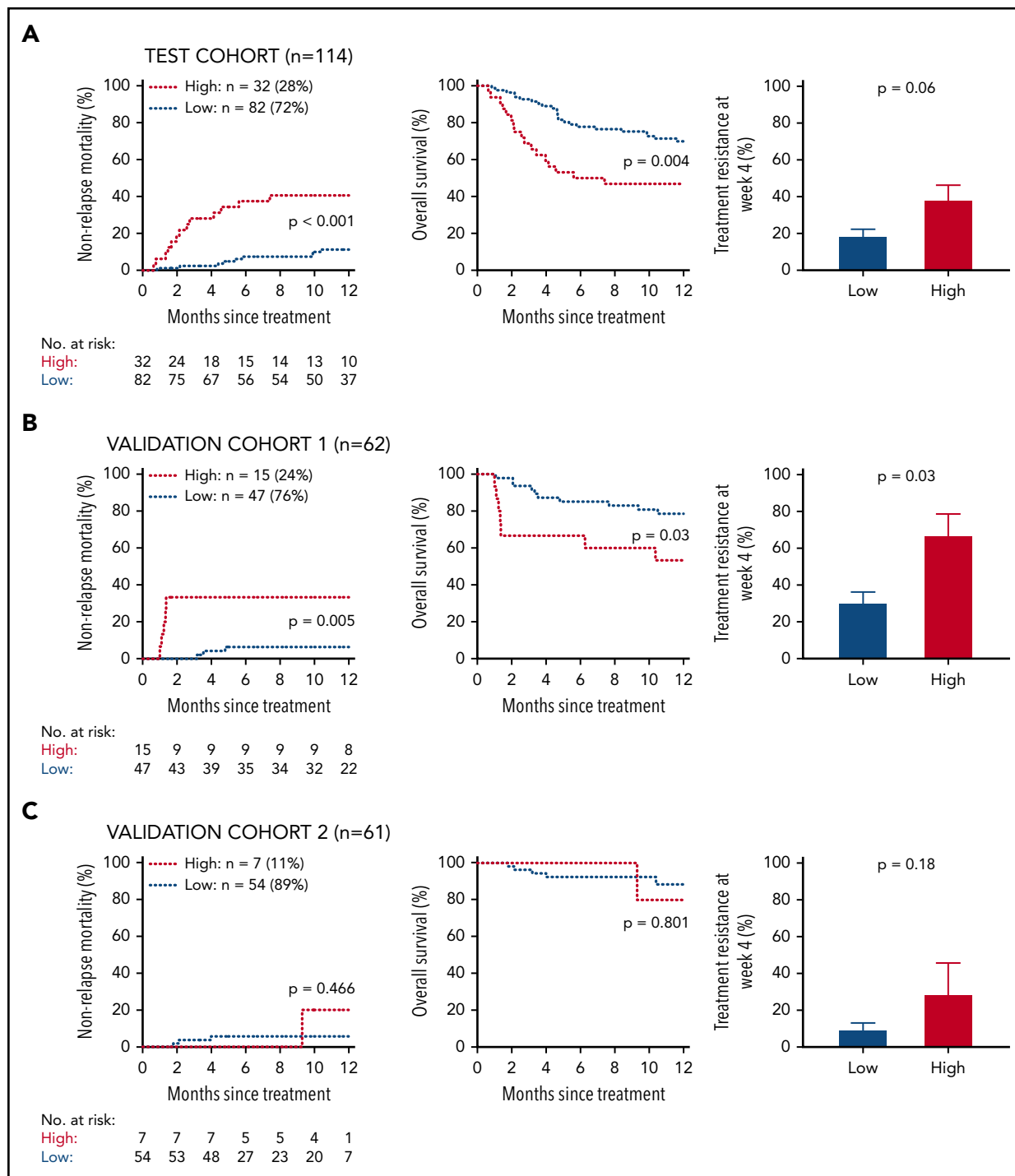


Figure 3. Long-term outcomes by biomarker probabilities in early treatment sensitive patients. Early treatment-sensitive patients were subdivided based on biomarker probabilities into low and high groups. (A) Test cohort of patients (n = 114). Twelve-month cumulative incidence of NRM (low 11% vs high 41%, $P < .001$) and OS (low 70% vs high 47%, $P = .004$) and proportion of patients resistant to treatment at week 4 (low 18% vs high 38%, $P = .06$). (B) Validation cohort 1 (n = 62). Twelve-month cumulative incidence of NRM (low 6% vs high 33%, $P = .005$) and OS (low 79% vs high 53%, $P = .03$) and proportion of patients resistant to treatment at week 4 (low 30% vs high 67%, $P = .03$). (C) Validation cohort 2 (n = 61). Twelve-month cumulative incidence of NRM (low 6% vs high 20%, $P = .46$) and OS (low 88% vs high 80%, $P = .80$) and proportion of patients resistant to treatment at week 4 (low 8% vs high 28%, $P = .18$).

Low-probability patients were also significantly less likely to remain resistant to treatment at week 4 compared with high-probability patients. As expected, GVHD was the leading cause of death in early treatment-resistant patients (supplemental Table 4).

We then evaluated biomarker stratification for patients whose GVHD was sensitive to the first week of systemic treatment. The biomarker algorithm again separated patients into 2 groups, with different outcomes in the test and first validation cohorts, but not in the second validation cohort, where the NRM of the

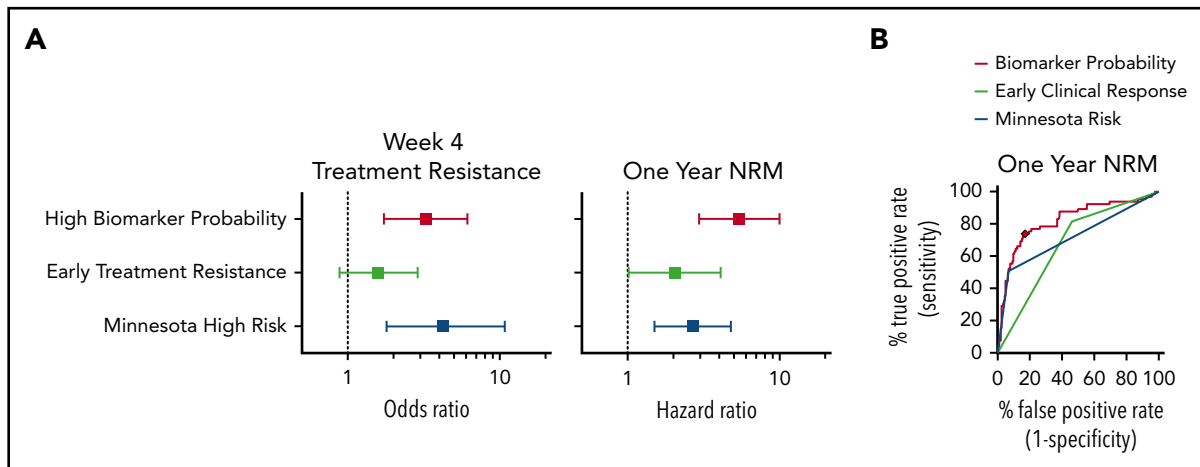


Figure 4. Prediction of long-term outcomes by early clinical response and biomarker probability status. (A) Forest plots. Left panel: Effect of early treatment resistance, Minnesota high-risk and high biomarker probability status on odds of resistance to treatment at week 4. Right panel: Effect of early treatment resistance, Minnesota high risk and high biomarker probability status on hazard of NRM at 1 year. Data are ratios and 95% confidence intervals. (B) Receiver operating characteristic curves to predict NRM. Curves are shown for early treatment response, biomarker probabilities, and Minnesota risk. The diamond (◆) indicates the threshold that defines low- versus high-risk groups. AUC for early treatment response = 0.68 ($P = .004$ compared with biomarker probability), for Minnesota risk = 0.72 ($P = .005$ compared with biomarker probability), and for biomarker probability = 0.82.

early-response group was only 8% (Figure 3). Again, relapse rates were not different between groups (supplemental Table 3). A similar pattern was seen for prediction of resistance to treatment at week 4. Thus, the biomarker algorithm did not reliably segregate patients into distinct risk groups when the NRM of the group was very low.

Biomarker probability scores predict NRM better than initial response to treatment or Minnesota clinical severity

The Minnesota clinical staging system has been shown to predict patient outcomes at the onset of GVHD symptoms, but its validity at later time points during treatment is unknown.¹⁰ We found that Minnesota risk after 1 week of treatment was highly significant in predicting resistance to treatment at week 4 and NRM at 1 year (supplemental Table 5). Univariate analyses of important pretransplant and GVHD clinical variables showed that only biomarker probability, early response to treatment, and Minnesota risk consistently significantly predicted response to treatment at 4 weeks and 1-year NRM. When these 3 variables were included in multivariate analysis of long-term outcomes in the combined validation cohorts, the biomarker and Minnesota risk ratios remained highly significant (Figure 4A). We then directly compared the ability of each variable to predict NRM by creating receiver operating characteristic curves (Figure 4B). The AUC for the biomarker probabilities was 0.82, significantly higher than 0.68 for early clinical response ($P = .004$) and 0.72 for Minnesota clinical risk ($P = .005$). The sensitivity, specificity, positive predictive value, and negative predictive values for 1-year NRM at the threshold used were 74%, 83%, 58%, and 91%, respectively, for the combined validation cohorts. These values are shown in Table 2 for each cohort and for the combined validation cohorts stratified by GVHD severity and initial treatment response. A separate algorithm that included the concentrations of ST2 and REG3 α , early treatment response, and Minnesota risk to predict 1-year NRM produced an AUC of 0.84, which was slightly better than the 0.82 for the biomarkers-only algorithm ($P = .024$); however, the individual clinical characteristics were not statistically significant elements of the algorithm (supplemental Table 6).

ST2 and REG3 α reflect damage to lower gastrointestinal (GI) mucosa, particularly in the crypts,^{12,19} and patients with high probabilities eventually experience sixfold more GI GVHD (supplemental Table 7). It should be noted that approximately one third of patients in the study were treated with systemic steroids for grade 1 GVHD (limited skin disease). If one excludes these patients from the analysis (some centers do not routinely use systemic treatment for isolated skin GVHD), the biomarker probabilities continue to divide patients into 2 groups, with significantly different outcomes based on whether patients were resistant to early systemic treatment (supplemental Figure 3). Patients with lower-GI symptoms (≥ 500 mL of diarrhea) demonstrated a significantly higher 1-year NRM than those with no diarrhea (55% vs 17%, $P < .001$); however, biomarker probabilities continue to separate both of these populations into 2 groups, with highly significant differences in NRM (supplemental Figure 4).

Discussion

Little progress has been made during the last several decades in validating new treatments for GVHD, for several reasons. The immune systems of all bone marrow transplantation patients have been largely eradicated by the pretransplant conditioning regimens to prevent rejection of the stem cell graft, and half of GVHD develops within the first month of transplant when immunologic reconstitution from the new graft is in its earliest phases.^{20,21} All strategies to treat acute GVHD suppress multiple elements of the immune system, further diminishing the immunologic competence of the patient. Even successful treatment requires at least a month of additional immunosuppression, increasing vulnerability to potentially fatal opportunistic infections. Thus, treatment of GVHD leads to severe infections that are directly related to the cumulative steroid dose.^{22,23} Additionally, acute GVHD symptom severity can fluctuate widely on a day-to-day basis, introducing significant uncertainty into assessments of response to treatment.⁹ GVHD can progress rapidly if not adequately treated, and clinicians often react quickly to worsening symptom severity, even though these might resolve without further intervention. Thus, accurate prediction of durable responses and long-term outcomes is key to escalation and de-escalation of immunosuppressive therapy.

Table 2. Sensitivity, specificity, positive-predictive value, and negative-predictive value analyses for 1-year NRM

		Sensitivity	Specificity	PPV	NPV
Test cohort		0.70	0.74	0.55	0.85
Validation cohort 1		0.80	0.76	0.58	0.91
Validation cohort 2		0.63	0.93	0.58	0.91
Minnesota risk	Early treatment response				
Standard	Response	0.50	0.86	0.29	0.94
Standard	Nonresponse	0.55	0.84	0.46	0.88
High*	Response	NA	NA	NA	NA
High	Nonresponse	0.94	0.54	0.84	0.78

NPV, negative-predictive value; PPV, positive-predictive value.

*Too few patients with Minnesota high risk and early treatment response to calculate sensitivity, specificity, positive-predictive value, and negative-predictive value

Previous studies have found that early clinical response to GVHD treatment correlates poorly with later clinical response and long-term survival.^{6,24} In this study, we found that, although early clinical responses continued to evolve, overall response at 1 week continued to have a significant predictive value for long-term outcomes in multivariate analysis (Figure 4). Measurement of serum concentrations of 2 MAGIC biomarkers (ST2 and REG3 α) at the same time as the clinical response segregated patients with steroid-resistant GVHD into 2 groups with strikingly different outcomes in all 3 patient cohorts. Patients with a low probability score who have not yet responded to treatment may be slow responders and may not require escalation of immunosuppression, despite the appearance of steroid resistance. Given the serious infectious risks of further immunosuppression, clinicians might adopt a stance of watchful waiting for such patients, an approach that could be addressed in a carefully controlled clinical trial.

We hypothesize that the accuracy of biomarker probabilities reflect the ability of serum concentrations of ST2 and REG3 α to measure immunologically mediated changes in tissue more accurately than the severity of the clinical disease or resistance to systemic steroids after 1 week of treatment. The superior predictive accuracy of biomarkers, when added to the presence of clinical symptoms, may be attributed to the number of processes that can simultaneously contribute to clinical GVHD symptoms. For example, diarrhea worsened by a concomitant viral gastroenteritis may be treated as GVHD with systemic steroids that, in fact, may prolong or intensify the viral disease; however, infection does not increase the serum concentration of REG3 α and, thus, would not raise the biomarker probability.¹² In pediatric patients, increased levels of ST2 correlate with transplant-associated thrombotic microangiopathy, as well as 6-month NRM.²⁵ Transplant-associated thrombotic microangiopathy and damage to the endothelium have been associated with acute GVHD,^{26,27} but whether such an association also exists in this study cannot be determined due to lack of relevant data.

Previous studies have found prognostic value for combinations of different biomarkers after 2 weeks of systemic steroid treatment, but none have determined the prognostic utility of biomarkers obtained earlier than 14 days into treatment.^{28,29} In 1 single-center study of 165 patients, the clinical status after 2 weeks of systemic steroid treatment of GVHD was a slightly worse predictor for 1-year NRM than 2 biomarkers (TIM3 and TNFR1)

measured at the same time (AUC of 0.81 and 0.85, respectively). It is not surprising that the clinical status after 2 weeks of therapy better reflects long-term outcomes than after 1 week, but some physicians might feel uncomfortable with such a delay, particularly if the patient has not responded to therapy. Thus, an important strength of the current study is the time of analysis at 1 week after therapy, when information is likely to be actionable. A second strength is the fact that the patients contributing data and samples from multiple centers were treated without prescriptive directives, including the second multicenter validation cohort that represents current practice and, therefore, reflects the heterogeneity of “real-life” GVHD treatment practices. A third strength of this study is that, although biomarkers were measured only at a single time point, the same algorithm that generated the probability score following treatment can predict outcomes when used prior to (and at) the onset of GVHD symptoms.⁸ The use of the same algorithm should enable the comparisons of probabilities generated serially at multiple time points following HCT, facilitating the incorporation of biomarker measurements into clinical practice. One limitation of this study is that the GVHD prophylaxis regimens of posttransplant cyclophosphamide, T-cell depletion, and tacrolimus/sirolimus were used in very few of the patients (Table 1); thus, these results should be applied with caution in these patients populations. It is also important to note that this study has not demonstrated that therapeutic decisions based on biomarker probabilities can change the outcome for patients with GVHD. But such probabilities should prove to be valuable clinical research tools because even relatively large clinical trials for acute GVHD generally do not enroll more than a few hundred patients.^{24,30} As a result, experimental therapies need to demonstrate a large benefit to prove beneficial, an outcome that has proved elusive over the past 40 years. The inclusion of patients who are likely to respond to standard therapy in a placebo-controlled trial reduces the likelihood of detecting a difference between arms; stratification of treatments by biomarker probability will avoid such a pitfall. Thus, the ability of the biomarker probability to risk stratify for patients who are not responding to systemic therapy should prove useful in clinical trial design and, ultimately, may help to tailor GVHD treatment to the risks and benefits for individual patients.

Acknowledgments

The authors thank the patients, their families, and the research staff for their participation.

This work was supported by grants from the National Institutes of Health, National Cancer Institute (R21CA173459, P01CA03942, and P30CA196521), an American Cancer Society Clinical Research Professorship (J.L.M.F.), and a Doris Duke Charitable Foundation Clinical Research Mentorship (M.J.H.).

Authorship

Contribution: H.M.-M., A.S.R., U.Ö., J.L.M.F., and J.E.L. designed the study, analyzed data, and wrote the paper; H.M.-M., M.J.H., M.A., S.K., G.M., and I.T. performed laboratory analyses; U.Ö., M.S.C., J.L.M.F., and J.E.L. created the figures; H.M.-M., A.S.R., A.P., F.A., E.H., Y.A.E., W.J.H., M.W., M.Q., E.O.H., K.W., R.O., R.Y., J.S., A.E., G.A.Y., D.W., A.C.H., M.P., and R.R. collected data; and all authors interpreted the data and contributed to the writing of the paper.

Conflict-of-interest disclosure: J.E.L., J.L.M.F., and U.Ö. are joint inventors on a GVHD biomarker patent. The remaining authors declare no competing financial interests.

Correspondence: John E. Levine, The Tisch Cancer Institute, Mount Sinai School of Medicine, 1 Gustave Levy Place, Box 1410, New York, NY 10029; e-mail: john.levine@mssm.edu.

Footnotes

Submitted 2 January 2018; accepted 14 March 2018. Prepublished online as *Blood* First Edition paper, 15 March 2018; DOI 10.1182/blood-2018-01-822957.

*H.M.-M. and A.S.R. contributed equally to this work.

†U.Ö., J.L.M.F., and J.E.L. contributed equally to this work.

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.

REFERENCES

- Majhail NS, Chitphakdithai P, Logan B, et al. Significant improvement in survival after unrelated donor hematopoietic cell transplantation in the recent era. *Biol Blood Marrow Transplant*. 2015;21(1):142-150.
- Hahn T, Sucheston-Campbell LE, Preus L, et al. Establishment of definitions and review process for consistent adjudication of cause-specific mortality after allogeneic unrelated-donor hematopoietic cell transplantation. *Biol Blood Marrow Transplant*. 2015;21(9):1679-1686.
- Anasetti C, Logan BR, Lee SJ, et al; Blood and Marrow Transplant Clinical Trials Network. Peripheral-blood stem cells versus bone marrow from unrelated donors. *N Engl J Med*. 2012;367(16):1487-1496.
- Deeg HJ. How I treat refractory acute GVHD. *Blood*. 2007;109(10):4119-4126.
- MacMillan ML, DeFor TE, Weisdorf DJ. The best endpoint for acute GVHD treatment trials. *Blood*. 2010;115(26):5412-5417.
- Saliba RM, Couriel DR, Giral S, et al. Prognostic value of response after upfront therapy for acute GVHD. *Bone Marrow Transplant*. 2012;47(1):125-131.
- Levine JE, Braun TM, Harris AC, et al; Blood and Marrow Transplant Clinical Trials Network. A prognostic score for acute graft-versus-host disease based on biomarkers: a multicentre study. *Lancet Haematol*. 2015;2(1):e21-e29.
- Hartwell MJ, Özbek U, Holler E, et al. An early-biomarker algorithm predicts lethal graft-versus-host disease and survival. *JCI Insight*. 2017;2(3):e89798.
- Harris AC, Young R, Devine S, et al. International, multicenter standardization of acute graft-versus-host disease clinical data collection: A report from the Mount Sinai Acute GVHD International Consortium. *Biol Blood Marrow Transplant*. 2016;22(1):4-10.
- 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.
- Armand P, Kim HT, Logan BR, et al. Validation and refinement of the Disease Risk Index for allogeneic stem cell transplantation. *Blood*. 2014;123(23):3664-3671.
- 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.
- 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.
- Reynolds AP, Richards G, de la Iglesia B, Rayward-Smith VJ. Clustering rules: a comparison of partitioning and hierarchical clustering algorithms. *J Math Model Algorithms*. 2006;5(4):475-504.
- Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12(1):77.
- MacMillan ML, Weisdorf DJ, Wagner JE, et al. Response of 443 patients to steroids as primary therapy for acute graft-versus-host disease: comparison of grading systems. *Biol Blood Marrow Transplant*. 2002;8(7):387-394.
- Martin PJ, Rizzo JD, Wingard JR, et al. First- and second-line systemic treatment of acute graft-versus-host disease: recommendations of the American Society of Blood and Marrow Transplantation. *Biol Blood Marrow Transplant*. 2012;18(8):1150-1163.
- Levine JE, Logan B, Wu J, et al; Blood and Marrow Transplant Clinical Trials Network. Graft-versus-host disease treatment: predictors of survival. *Biol Blood Marrow Transplant*. 2010;16(12):1693-1699.
- Zhang J, Ramadan AM, Griesenauer B, et al. ST2 blockade reduces sST2-producing T cells while maintaining protective mST2-expressing T cells during graft-versus-host disease. *Sci Transl Med*. 2015;7(308):308ra160.
- Jagasia M, Arora M, Flowers ME, et al. Risk factors for acute GVHD and survival after hematopoietic cell transplantation. *Blood*. 2012;119(1):296-307.
- Mehta RS, Rezvani K. Immune reconstitution post allogeneic transplant and the impact of immune recovery on the risk of infection. *Virulence*. 2016;7(8):901-916.
- Matsumura-Kimoto Y, Inamoto Y, Tajima K, et al. Association of cumulative steroid dose with risk of infection after treatment for severe acute graft-versus-host disease. *Biol Blood Marrow Transplant*. 2016;22(6):1102-1107.
- Miller HK, Braun TM, Stillwell T, et al. Infectious risk after allogeneic hematopoietic cell transplantation complicated by acute graft-versus-host disease. *Biol Blood Marrow Transplant*. 2017;23(3):522-528.
- Bolaños-Meade J, Logan BR, Alousi AM, et al. Phase 3 clinical trial of steroids/mycophenolate mofetil vs steroids/placebo as therapy for acute GVHD: BMT CTN 0802. *Blood*. 2014;124(22):3221-3227, quiz 3335.
- Rotz SJ, Dandoy CE, Davies SM. ST2 and endothelial injury as a link between GVHD and microangiopathy. *N Engl J Med*. 2017;376(12):1189-1190.
- Dietrich S, Falk CS, Benner A, et al. Endothelial vulnerability and endothelial damage are associated with risk of graft-versus-host disease and response to steroid treatment. *Biol Blood Marrow Transplant*. 2013;19(1):22-27.
- Penack O, Socié G, van den Brink MR. The importance of neovascularization and its inhibition for allogeneic hematopoietic stem cell transplantation. *Blood*. 2011;117(16):4181-4189.
- McDonald GB, Tabellini L, Storer BE, et al. Predictive value of clinical findings and plasma biomarkers after fourteen days of prednisone treatment for acute graft-versus-host disease. *Biol Blood Marrow Transplant*. 2017;23(8):1257-1263.
- Levine JE, Logan BR, Wu J, et al. Acute graft-versus-host disease biomarkers measured during therapy can predict treatment outcomes: a Blood and Marrow Transplant Clinical Trials Network study. *Blood*. 2012;119(16):3854-3860.
- Hockenbery DM, Cruickshank S, Rodell TC, et al. A randomized, placebo-controlled trial of oral beclomethasone dipropionate as a prednisone-sparing therapy for gastrointestinal graft-versus-host disease. *Blood*. 2007;109(10):4557-4563.