ORIGINAL ARTICLE

ST2 as a Marker for Risk of Therapy-Resistant Graft-versus-Host Disease and Death

Mark T. Vander Lugt, M.D., Thomas M. Braun, Ph.D., Samir Hanash, M.D., Ph.D., Jerome Ritz, M.D., Vincent T. Ho, M.D., Joseph H. Antin, M.D., Qing Zhang, M.S., Chee-Hong Wong, M.S., Hong Wang, M.S., Alice Chin, M.S., Aurélie Gomez, M.D., Andrew C. Harris, M.D., John E. Levine, M.D., Sung W. Choi, M.D., Daniel Couriel, M.D., Pavan Reddy, M.D., James L.M. Ferrara, M.D., D.Sc., and Sophie Paczesny, M.D., Ph.D.

ABSTRACT

BACKGROUND

No plasma biomarkers are associated with the response of acute graft-versus-host disease (GVHD) to therapy after allogeneic hematopoietic stem-cell transplantation.

METHODS

We compared 12 biomarkers in plasma obtained a median of 16 days after therapy initiation from 10 patients with a complete response by day 28 after therapy initiation and in plasma obtained from 10 patients with progressive GVHD during therapy. The lead biomarker, suppression of tumorigenicity 2 (ST2), was measured at the beginning of treatment for GVHD in plasma from 381 patients and during the first month after transplantation in three independent sets totaling 673 patients to determine the association of this biomarker with treatment-resistant GVHD and 6-month mortality after treatment or transplantation.

RESULTS

Of the 12 markers, ST2 had the most significant association with resistance to GVHD therapy and subsequent death without relapse. As compared with patients with low ST2 values at therapy initiation, patients with high ST2 values were 2.3 times as likely to have treatment-resistant GVHD (95% confidence interval [CI], 1.5 to 3.6) and 3.7 times as likely to die within 6 months after therapy (95% CI, 2.3 to 5.9). Patients with low ST2 values had lower mortality without relapse than patients with high ST2 values, regardless of the GVHD grade (11% vs. 31% among patients with grade I or II GVHD and 14% vs. 67% among patients with grade III or IV GVHD, $P<0.001$ for both comparisons). Plasma ST2 values at day 14 after transplantation were associated with 6-month mortality without relapse, regardless of the intensity of the conditioning regimen.

CONCLUSIONS

ST2 levels measured at the initiation of therapy for GVHD and during the first month after transplantation improved risk stratification for treatment-resistant GVHD and death without relapse after transplantation. (Funded by the National Institutes of Health.)

From the Departments of Pediatrics (M.T.V.L., A.G., A.C.H., J.E.L., S.W.C., J.L.M.F., S.P.), Biostatistics (T.M.B.), and Internal Medicine (D.C., P.R., J.L.M.F.), University of Michigan, Ann Arbor; the Molecular Diagnostics Program, Fred Hutchinson Cancer Research Center, Seattle (S.H., Q.Z., C.-H.W., H.W., A.C., S.P.); and the Dana–Farber Cancer Institute, Boston (J.R., V.T.H., J.H.A.). Address reprint requests to Dr. Paczesny at the Department of Pediatrics, Bone Marrow and Stem Cell Transplantation Program, Indiana University Melvin and Bren Simon Cancer Center and Wells Center for Pediatric Research, 1044 W. Walnut St., Rm. 425, Indianapolis, IN 46202, or at sophpacz@iu.edu.

N Engl J Med 2013;369:529-39.
DOI: 10.1056/NEJMoa1213299

Although mortality related to graft-versus-host disease (GVHD) after allogeneic hematopoietic stem-cell transplantation has been reduced,[1,2] acute GVHD remains a major complication of allogeneic transplantation, occurring in approximately half the transplant recipients.[3,4] High-dose systemic glucocorticoids remain the first-line therapy for GVHD,[5-9] although just half of patients have complete resolution of GVHD by day 28 after therapy initiation.[6] Patients who do not have a response to GVHD therapy are at high risk for death without relapse of the primary disease for which the transplantation was performed within 6 months after therapy initiation.[10-13] We previously reported that a model with six biomarkers diagnostic of GVHD (interleukin-2 receptor α, tumor necrosis factor receptor 1, hepatocyte growth factor, and interleukin-8 for systemic GVHD[14]; elafin for skin GVHD[15]; and regenerating islet-derived 3α [REG3α] for gastrointestinal GVHD[16]) was associated with response to GVHD therapy by day 28 and mortality at 6 months after therapy initiation.[17] However, none of these studies were designed to identify biomarkers of glucocorticoid resistance.

In the current study, we used a plasma proteomic approach to compare patients who had a response with patients who did not and identified suppression of tumorigenicity 2 (ST2) as having the strongest association with risk, of the 12 markers of nonresponse to GVHD therapy and subsequent death without relapse that were assessed. We then evaluated ST2 levels at therapy initiation and at day 14 after transplantation as predictors of therapy-resistant GVHD and death.

Methods

Study Design and Oversight

The study was approved by the institutional review boards at the University of Michigan and the Dana–Farber Cancer Institute. No commercial sponsor was involved in the study. Written informed consent was obtained from all patients or their legal guardians before hematopoietic stem-cell transplantation and sample collection. Only those persons listed as authors contributed to the writing of the manuscript. All authors vouch for the completeness and accuracy of the data and the fidelity of the study to the protocol.

Sample Collection and Preparation

Plasma samples were collected prospectively between 2000 and 2010 from patients who underwent allogeneic hematopoietic stem-cell transplantation (hereafter referred to as transplantation). Samples were collected weekly throughout the first month after transplantation, monthly to day 100 after transplantation, at the time of key clinical events (including the development of GVHD symptoms), and in the 48-hour window before initiation of glucocorticoid therapy. GVHD was graded according to modified Glucksberg criteria.[18] Details are summarized in the Supplementary Appendix, available with the full text of this article at NEJM.org.

Discovery Set

To see the greatest difference in levels of protein expression, we compared plasma obtained a median of 16 days after therapy initiation from 10 patients in whom grade II to IV GVHD developed and who had a complete response by day 28 after therapy initiation with plasma obtained from 10 patients in whom grade II or III GVHD developed and who had progressive GVHD by day 28. Characteristics of these patients are shown in Table S1 in the Supplementary Appendix.

Response-to-Treatment Set

Biomarkers, including the lead candidate ST2, were measured in plasma from 381 patients at the beginning of treatment for GVHD to determine the association of the biomarkers with treatment-resistant GVHD and 6-month mortality after treatment. The characteristics of 381 patients in whom GVHD developed and who received high-dose systemic glucocorticoids for treatment of grade I to IV GVHD before day 180 after transplantation are shown in Table S2 in the Supplementary Appendix. Patients were divided into two groups on the basis of response status at day 28, which is a commonly accepted surrogate for 6-month survival.[10-13] Patients were considered to have had a response if they had improvement in at least one organ without progression in any other organs and if additional therapy was not required. Patients were considered not to have had a response if they had stable or progressive GVHD or if the subsequent addition of secondary therapy was required.

EARLY-STRATIFICATION SET

We measured ST2 concentrations during the first month after transplantation in plasma from three independent sets totaling 673 patients to determine the association of ST2 and 6-month mortality after transplantation. We first measured ST2 concentrations in a pilot group of 296 patients with samples available at days 0, 14, and 21 after transplantation and in whom GVHD did not develop before day 35 (Table S3 in the Supplementary Appendix). The greatest difference between those who were alive and those who had died at 6 months after transplantation was seen in samples obtained on day 14 after transplantation. We then measured ST2 concentrations at day 14 after transplantation in 302 separate patients: 94 with no GVHD and 208 with grade I to IV acute GVHD between day 18 and day 100. We also measured ST2 concentrations at day 14 after transplantation in an independent set of 75 patients who underwent unrelated-donor transplantation involving total-body irradiation at the Dana–Farber Cancer Institute. Characteristics of these patients are shown in Tables S4 and S5 in the Supplementary Appendix.

PROTEOMIC ANALYSIS AND ENZYME-LINKED IMMUNOSORBENT ASSAY

Methods used for the intact-protein analysis system (IPAS) have been reported previously[15,19,20] and are summarized in the Supplementary Appendix. The procedures used for enzyme-linked immunosorbent assay (ELISA) and the assay parameters are described in the Supplementary Appendix, including Table S6.

STATISTICAL ANALYSIS

Between-group differences were assessed with the use of either a chi-square test or a Wilcoxon rank-sum test. Log-transformed biomarker values were used in all analyses owing to skewed raw values. Pearson's product-moment correlation was used to estimate the pairwise risk association of the biomarkers. Logistic regression, Cox regression, and competing-risks regression[21] were used to evaluate the association of the six-biomarker panel and ST2 alone with response status at day 28, overall survival, mortality without relapse, and mortality with relapse. Competing models were compared with the use of three metrics: hazard ratios and their 95% confidence intervals; the area under the curve (AUC), the area under the receiver-operating-characteristic (ROC) curve, or the C statistic, a scaled version of the Wilcoxon rank-sum statistic; and a continuous form of the net reclassification index as described by Pencina et al.[22] A P value of less than 0.10 was considered to indicate statistical significance for the proteomic analysis, and a P value of less than 0.05 was considered to indicate statistical significance for all other analyses.

RESULTS

CANDIDATE BIOMARKERS OF NONRESPONSE TO GVHD THERAPY

Using IPAS technology, we compared pooled plasma obtained a median of 16 days after therapy initiation from 10 patients who had a complete response to GVHD therapy and 10 with progressive GVHD. Of 571 proteins identified and quantified, 197 were increased by a factor of at least 1.5 in patients with progressive GVHD, and 12 had commercially available antibodies that were suitable for use in an ELISA (Table S7 in the Supplementary Appendix). We chose for further evaluation biomarkers that were significantly elevated in the individual plasma samples from patients with progressive GVHD as compared with those with a complete response ($P<0.10$) and that had an AUC of the ROC that was significant ($P<0.10$) for the comparison of patients who had a complete response with patients who had progressive disease. Six biomarkers met both these criteria: ST2, macrophage inhibitory factor, interleukin-1 receptor 2, lipocalin 2, lymphatic-vessel endothelial hyaluronan receptor 1, and REG3α (Table S8 and Fig. S1 in the Supplementary Appendix). Each of these six biomarkers had an AUC of at least 0.85 by ROC analysis.

DEVELOPMENT OF A BIOMARKER PANEL

We measured the six best biomarkers from the discovery set, listed above, in 381 patients in whom GVHD developed and who received treatment; the characteristics of these patients are summarized in Table S2 in the Supplementary Appendix. This group included all patients with samples available at the onset of GVHD who were treated with systemic glucocorticoids. We divided patients into two subgroups according to response status at day 28, as detailed above. Grade

III or IV GVHD, gastrointestinal GVHD, and treatment with glucocorticoids alone were overrepresented in patients who did not have a response. The median day of GVHD onset and the median day of sample acquisition were earlier in patients who did not have a response. There was also a trend toward overrepresentation of HLA-mismatched donors and full-intensity (myeloablative) conditioning among patients who did not have a response.

Biomarkers do not improve the prediction of outcomes unless they are not correlated with each other or with clinical predictors.[14,17] Concentrations of the six lead biomarkers correlated with each other weakly ($r<0.43$) (Fig. S2 and Table S9 in the Supplementary Appendix). We determined the AUC for each marker to predict response to therapy by day 28, an established surrogate for 6-month mortality without relapse after treatment,[10-13] and the actual outcome of interest, 6-month mortality without relapse (Table S10 in the Supplementary Appendix). We used these six biomarkers in a panel for further analyses and compared the panel with ST2 alone.

BIOMARKER VALUES AT THERAPY INITIATION AND RISK OF NONRESPONSE

We first performed univariate analyses to predict response to treatment by day 28 with the full panel of biomarkers, ST2 alone, and the following demographic and clinical characteristics: age, disease status, conditioning intensity, donor source, HLA match, GVHD grade at therapy initiation, and initial therapy for GVHD. Conditioning intensity, HLA match, GVHD grade at therapy initiation, and initial therapy for GVHD were associated with nonresponse by day 28 and were included in the model. We then divided patients into high-risk and low-risk groups on the basis of the value of the full biomarker panel or ST2 alone. Owing to the poor prognosis for the patients who did not have a response, we defined a high biomarker value as a biomarker concentration that was at least 50% higher than the median value in patients who had a response and a high panel value as high concentrations of at least three of the six biomarkers measured. In a multivariate analysis, either a high panel value or a high ST2 value alone increased the likelihood of nonresponse. As compared with patients with low ST2 values at therapy initiation, patients with high ST2 values were 2.3 times as likely to have treatment-resistant GVHD (95% confidence interval, 1.5 to 3.6) (Table S11A in the Supplementary Appendix). An analysis of the stage of involvement in the target organ, which is also associated with response to therapy,[23,24] yielded similar results (Table S11B in the Supplementary Appendix).

We then compared ROC curves for the significant demographic and clinical characteristics alone with those of the clinical characteristics combined with either the full panel or ST2 alone to predict nonresponse by day 28 and found no significant increase in the AUC (0.64 vs. 0.68) (Table S11A in the Supplementary Appendix). We calculated the continuous-form net reclassification index[22] to evaluate the benefit of the full panel or ST2 alone as compared with the model that included demographic and clinical characteristics only. The net reclassification index for the biomarker panel is the sum of the proportion of patients without a response for whom the predicted probability increased with the addition of the panel and the proportion of patients with a response for whom the predicted probability decreased with the addition of the panel. The same approach was used to calculate the net reclassification index for ST2 alone. The addition of either the panel or ST2 alone correctly reclassified 44% and 45% of patients, respectively ($P<0.001$ for both comparisons) (Table S11C in the Supplementary Appendix). All net-reclassification-index values reported here were calculated with the use of this algorithm.

We have previously reported that the biomarker REG3α is associated with response to therapy in patients with lower gastrointestinal GVHD.[16] We therefore compared ST2 with REG3α in this subgroup of patients. Inclusion of both ST2 and REG3α in the model improved the accuracy of the risk stratification as compared with either biomarker alone. The net reclassification index was 71% as compared with the clinical characteristics alone (Table S12 in the Supplementary Appendix).

BIOMARKER VALUES, 6-MONTH MORTALITY WITHOUT RELAPSE, AND SURVIVAL

Nonresponse to therapy by day 28 is an established surrogate for 6-month mortality without relapse; however, a surrogate end point does not always indicate all the effects of the biomarkers

on the clinical end point.[25,26] Therefore, we analyzed the association of the clinical characteristics and biomarkers with 6-month post-therapy mortality without relapse. In univariate analyses, age, HLA match, GVHD grade at therapy initiation, and initial therapy for GVHD were associated with 6-month post-therapy mortality without relapse. After adjustment for the significant demographic and clinical characteristics, patients with a high panel or ST2 value were significantly more likely to die than patients with lower values. For patients with a high panel value and those with a high ST2 value, 6-month mortality without relapse was 44% and 42%, respectively, whereas for patients with a low panel value and those with a low ST2 value, the rates were 17% and 12%, respectively (hazard ratio, 2.7 and 3.7, respectively) (Table 1). The addition of either the panel value or the ST2 value alone did not significantly increase the AUCs; however, the addition of either the panel value or the ST2 value alone significantly improved the net reclassification index (62% and 76%, respectively; P<0.001 for both comparisons) (Table 2). ST2 was superior to the six-biomarker panel in predicting death without relapse, and we used ST2 alone for all further analyses.

RISK STRATIFICATION FOR DEATH WITHOUT RELAPSE ACCORDING TO ST2 VALUE AND GVHD GRADE

We stratified patients in the response-to-treatment set according to risk of death without relapse at initiation of therapy, using the strongest clinical predictor, GVHD grade, and the strongest biomarker, ST2. This categorization divided patients into four groups: low ST2 concentration and grade I or II GVHD (group 1), low ST2 concentration and grade III or IV GVHD (group 2), high ST2 concentration and grade I or II GVHD (group 3), and high ST2 concentration and grade III or IV GVHD (group 4). Patients in group 1 and group 2 had low mortality without relapse (11% and 14%, respectively). Patients in group 3, who would be expected to do well solely on the basis

Table 1. Effect of Potential Risk Factors, Including Suppression of Tumorigenicity 2 (ST2) Concentration at the Initiation of Therapy, on the Risk of Death without Relapse of the Primary Disease for Which the Transplantation Was Performed, at 6 Months after Therapy.*

Variable	Univariate Analysis		Multivariate Analysis with Biomarker Panel Value†		Multivariate Analysis with ST2 Value Alone‡	
	Odds Ratio (95% CI)	P Value	Odds Ratio (95% CI)	P Value	Odds Ratio (95% CI)	P Value
Age: ≥55 yr vs. <55 yr	1.8 (1.2–2.6)	0.004	2.1 (1.4–3.1)	<0.001	2.0 (1.4–3.0)	<0.001
HLA match: mismatched vs. matched	1.5 (1.0–2.2)	0.049	1.4 (1.0–2.1)	0.09	1.3 (0.9–1.9)	0.24
GVHD grade: III or IV vs. I or II	3.0 (2.1–4.5)	<0.001	2.8 (1.9–4.2)	<0.001	2.6 (1.7–3.8)	<0.001
Initial therapy for GVHD: glucocorticoids only vs. glucocorticoids and another agent	0.6 (0.4–0.8)	0.003	0.6 (0.4–0.9)	0.01	0.7 (0.5–1.1)	0.10
Panel value: high vs. low§	3.1 (2.1–4.6)	<0.001	2.7 (1.9–4.0)	<0.001	NA	NA
ST2 value alone: high vs. low	4.4 (2.8–7.1)	<0.001	NA	NA	3.7 (2.3–5.9)	<0.001

* NA denotes not applicable.

† The six-biomarker panel consisted of ST2, macrophage inhibitory factor, interleukin-1 receptor 2, lipocalin 2, lymphatic-vessel endothelial hyaluronan receptor 1, and regenerating islet-derived 3α (REG3α). The effect of the panel value was calculated with the use of regression models adjusted for age, HLA match, GVHD grade, and initial therapy for GVHD. The area under the curve (AUC) was 0.75 in the multivariate model including demographic and clinical characteristics and the panel value, as compared with 0.71 in the model including demographic and clinical characteristics alone (P=0.31); the net reclassification index was 62% (P<0.001).

‡ The effect of the ST2 value alone was calculated with the use of regression models adjusted for age, HLA match, GVHD grade, and initial therapy for GVHD. The AUC was 0.77 in the multivariate model including demographic and clinical characteristics and the ST2 value, as compared with 0.71 in the model including demographic and clinical characteristics alone (P=0.13); the net reclassification index was 76% (P<0.001).

§ A high panel value was defined as high concentrations of at least three of the six biomarkers measured. A high concentration was defined as one that was at least 50% higher than the median value in patients who had a response. High concentrations were as follows: ST2, 740 pg per milliliter or more; interleukin-1 receptor 2, 12,000 pg per milliliter or more; macrophage inhibitory factor, 760 pg per milliliter or more; lipocalin 2, 125 ng per milliliter or more; lymphatic-vessel endothelial hyaluronan receptor 1, 650 ng per milliliter or more; and REG3α, 120 ng per milliliter or more. Of 142 patients with a high panel value, 122 had a high ST2 value, 82 had a high value of interleukin-1 receptor 2, 97 had a high value of macrophage inhibitory factor, 77 had a high value of lipocalin 2, 71 had a high value of lymphatic-vessel endothelial hyaluronan receptor 1, and 101 had a high REG3α value.

Table 2. Reclassification of the Risk of Death without Relapse at 6 Months after Therapy in the Response-to-Treatment Set after the Addition of the Panel Value or ST2 Value to Demographic and Clinical Characteristics.

Variable	Patients Who Did Not Die within 6 Months after Therapy	Patients Who Died within 6 Months after Therapy
Panel value added		
Expected (no.)	278	103
Reclassified to lower risk (no.)	198	41
Reclassified to higher risk (no.)	80	62
Net reclassification index (%)*	42	20
ST2 value alone added		
Expected (no.)	278	103
Reclassified to lower risk (no.)	166	22
Reclassified to higher risk (no.)	112	81
Net reclassification index (%)†	19	57

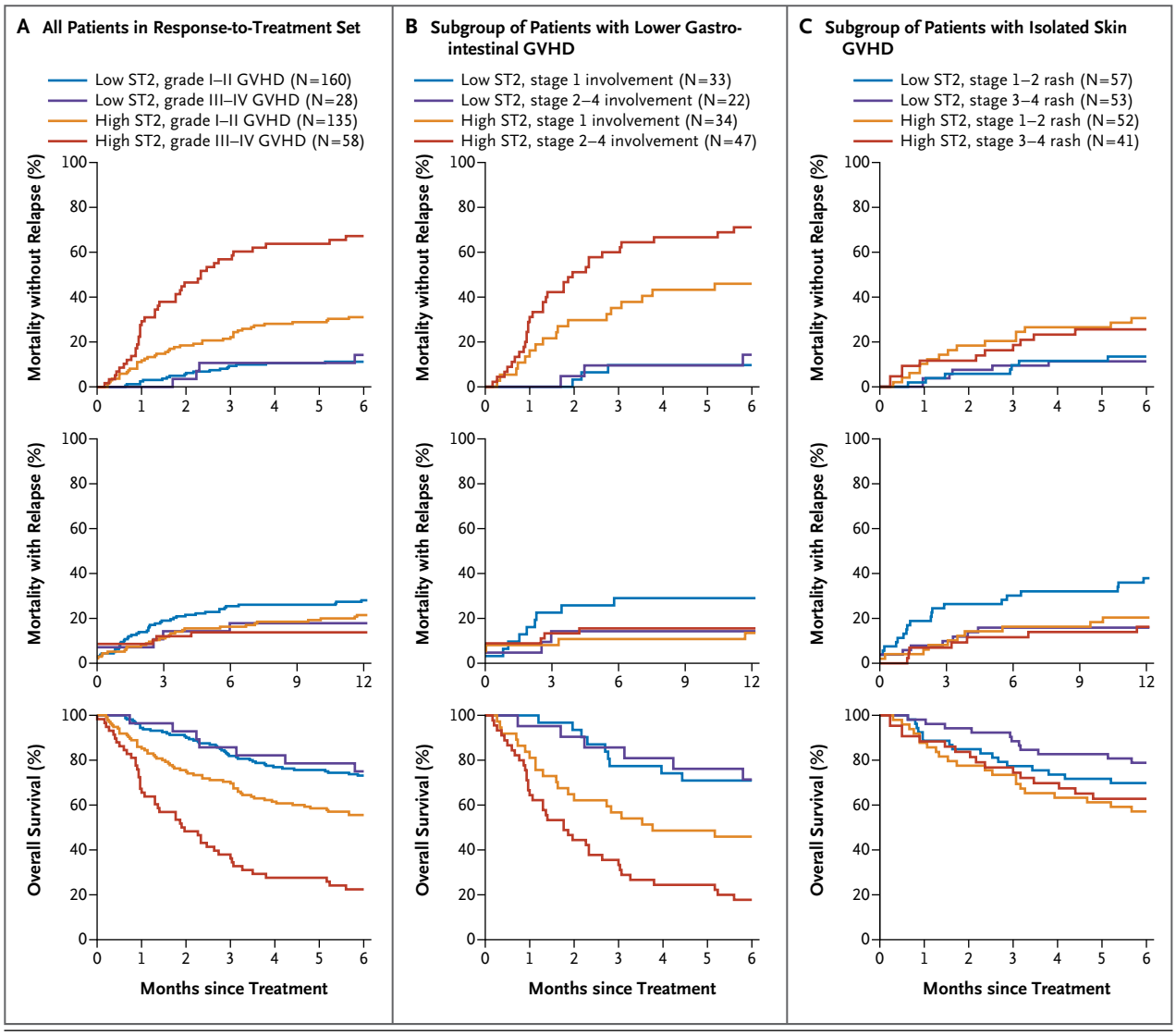
* The net reclassification index was 42% (118 of 278) among patients who did not die within 6 months after therapy, 20% (21 of 103) among patients who died within 6 months after therapy, and 62% overall.

† The net reclassification index was 19% (54 of 278) among patients who did not die within 6 months after therapy, 57% (59 of 103) among patients who died within 6 months after therapy, and 76% overall.

of GVHD grade, had significantly higher mortality without relapse (31%) than patients in group 1 (P<0.001). Patients in group 4 had the highest mortality without relapse, at 67% (Fig. 1A, and Table S13A in the Supplementary Appendix). Thus, ST2 concentration significantly stratified the risk of death without relapse in each clinical group. The survival difference of 53 percentage points between the two groups of patients with a high grade of GVHD (groups 2 and 4, P<0.001) is particularly striking, showing that the biomarker value was more important than the clinical grade in group 2. Mortality with relapse did not differ significantly among the four risk groups (Fig. 1A, and Table S13B in the Supplementary Appendix). As a result, overall survival was significantly lower in groups 3 and 4 than in groups 1 and 2 (Fig. 1A, and Table S13C in the Supplementary Appendix). The causes of death are shown in Table S14 in the Supplementary Appendix. The findings were nearly identical for patients with lower gastrointestinal GVHD, the target organ most responsible for death without relapse (Fig. 1B; and Tables S13D, S13E, and S13F in the Supplementary Appendix). The ST2 concentration was not associated with mortality without re-

Figure 1 (facing page). Outcomes According to Plasma Concentration of Suppression of Tumorigenicity 2 (ST2) and Grade of GVHD at Therapy Initiation.

The 381 patients in the response-to-treatment set (Panel A) were divided into four groups on the basis of plasma ST2 concentration (with low defined as <740 pg per milliliter and high as ≥740 pg per milliliter) and GVHD grade at therapy initiation. The 6-month mortality without relapse after therapy was as follows: patients with low ST2 and grade I or II GVHD, 11% (95% confidence interval [CI], 7 to 17); patients with low ST2 and grade III or IV GVHD, 14% (95% CI, 6 to 34); patients with high ST2 and grade I or II GVHD, 31% (95% CI, 24 to 40); and patients with high ST2 and grade III or IV GVHD, 67% (95% CI, 55 to 79). The 1-year mortality with relapse was 28% (95% CI, 22 to 36), 18% (95% CI, 8 to 38), 21% (95% CI, 15 to 29), and 14% (95% CI, 7 to 26) in the respective groups. Overall survival at 6 months after therapy initiation was 73% (95% CI, 67 to 80), 75% (95% CI, 61 to 93), 56% (95% CI, 48 to 65), and 22% (95% CI, 14 to 36) in the respective groups. Among the 136 patients with lower gastrointestinal GVHD (Panel B), the 6-month mortality without relapse was as follows: patients with low ST2 and stage 1 gastrointestinal involvement, 10% (95% CI, 5 to 18); patients with low ST2 and stage 2 to 4 gastrointestinal involvement, 14% (95% CI, 7 to 27); patients with high ST2 and stage 1 gastrointestinal involvement, 46% (95% CI, 42 to 50); and patients with high ST2 and stage 2 to 4 gastrointestinal involvement, 71% (95% CI, 68 to 74). The 1-year mortality with relapse was 29% (95% CI, 24 to 35), 14% (95% CI, 7 to 27), 11% (95% CI, 7 to 17), and 16% (95% CI, 12 to 20) in the respective groups. Overall survival at 6 months after therapy initiation was 71% (95% CI, 57 to 89), 71% (95% CI, 54 to 94), 46% (95% CI, 32 to 65), and 18% (95% CI, 9 to 33) in the respective groups. Among the 203 patients with isolated skin GVHD (Panel C), the 6-month mortality without relapse was as follows: patients with low ST2 and stage 1 or 2 rash, 17% (95% CI, 7 to 27); patients with low ST2 and stage 3 or 4 rash, 15% (95% CI, 7 to 37); patients with high ST2 and stage 1 or 2 rash, 27% (95% CI, 6 to 48); and patients with high ST2 and stage 3 or 4 rash, 25% (95% CI, 9 to 41). The 1-year mortality with relapse was 26% (95% CI, 24 to 29), 14% (95% CI, 11 to 18), 19% (95% CI, 15 to 24), and 14% (95% CI, 9 to 20) in the respective groups. Overall survival at 6 months after therapy initiation was 68% (95% CI, 57 to 80), 78% (95% CI, 68 to 89), 57% (95% CI, 43 to 75), and 61% (95% CI, 47 to 79) in the respective groups. All outcomes were determined with the use of the Kaplan–Meier procedure. Mortality without relapse on the ordinate refers to death in the absence of relapse of the primary disease for which the transplantation was performed. Detailed hazard ratios for these outcomes are shown in Tables S13A through S13I in the Supplementary Appendix. GVHD was graded according to modified Glucksberg criteria[18] (summarized in the Supplementary Appendix).



lapse or overall survival among patients with skin GVHD (Fig. 1C; and Tables S13G, S13H, and S13I in the Supplementary Appendix), findings that are consistent with the fact that the skin is the GVHD target organ that is most responsive to therapy.

ST2 AT DAY 14 AFTER TRANSPLANTATION AND MORTALITY IN THREE INDEPENDENT SETS

We next evaluated the association of ST2 concentrations during the first month after transplantation and mortality at 6 months after transplantation. We measured ST2 concentrations on days 0, 14, and 21 after transplantation in a University of Michigan pilot set of 296 patients in whom GVHD did not develop before day 35 (Table S3 in the Supplementary Appendix). ST2 levels on days 0, 14, and 21 after transplantation were two to four times as high in patients receiving myeloablative preparative therapy as in patients receiving reduced-intensity conditioning, so a separate analysis of each conditioning intensity was necessary. Samples obtained on day 14 after transplantation were most closely correlated with 6-month mortality in both conditioning groups (Fig. 2A and 2B).

To further develop the early-stratification model, we analyzed data from another set of 302 patients at the University of Michigan, who did not have GVHD before day 18 and who had

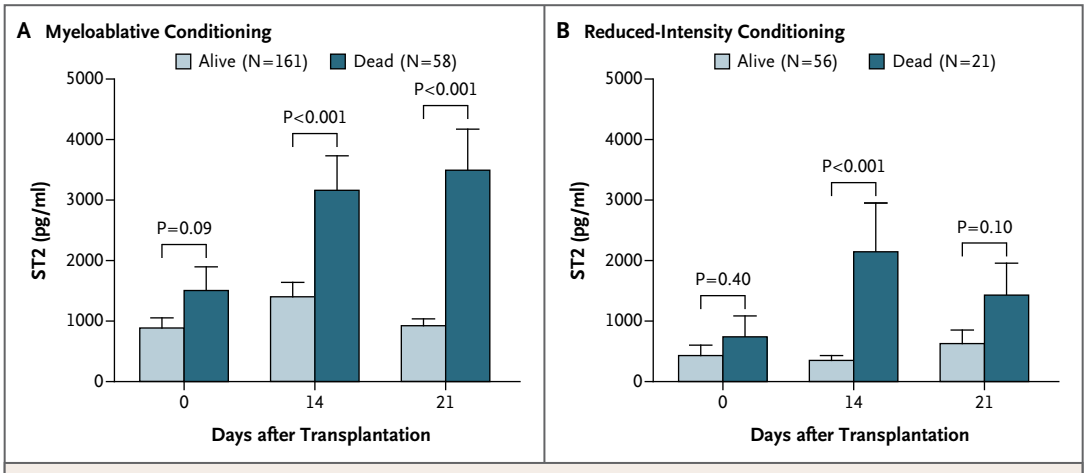


Figure 2. Plasma ST2 Concentrations within 21 Days after Transplantation According to Conditioning Intensity and Survival.

Panel A shows the mean (±SE) plasma concentration of ST2 in patients in the pilot group receiving myeloablative conditioning who were alive at 180 days after transplantation and those who were dead at that time. Values at day 0 were as follows: alive at 180 days, 880±170 pg per milliliter; dead at 180 days, 1500±390 pg per milliliter (P=0.09). Values at day 14 were 1400±240 pg per milliliter and 3200±570 pg per milliliter, respectively (P<0.001), and values at day 21 were 920±115 pg per milliliter and 3500±680 pg per milliliter, respectively (P<0.001). Panel B shows the mean (±SE) plasma concentration of ST2 in patients in the pilot group receiving reduced-intensity conditioning who were alive at 180 days after transplantation and those who were dead at that time. Values at day 0 were as follows: alive at 180 days, 430±170 pg per milliliter; dead at 180 days, 740±350 pg per milliliter (P=0.40). Values at day 14 were 350±80 pg per milliliter and 2150±800 pg per milliliter, respectively (P<0.001), and values at day 21 were 630±225 pg per milliliter and 1430±525 pg per milliliter, respectively (P=0.10).

samples available on day 14 (Table S4 in the Supplementary Appendix). We also separately analyzed data from an independent external validation set of 75 patients receiving myeloablative conditioning with total-body irradiation at the Dana–Farber Cancer Institute (Table S5 in the Supplementary Appendix). University of Michigan patients in whom GVHD developed were older and more likely to have received HLA-mismatched or unrelated-donor transplants than patients without GVHD. Patients receiving methotrexate as GVHD prophylaxis were overrepresented in the no-GVHD group. Dana–Farber patients who received sirolimus as GVHD prophylaxis were overrepresented in the no-GVHD group.

Because ST2 concentrations varied across levels of conditioning intensity, we used three models to stratify risk, with the following cutoff points: the median ST2 concentration in University of Michigan patients receiving myeloablative conditioning, the median concentration in those receiving reduced-intensity conditioning, and the median concentration in Dana–Farber patients receiving myeloablative conditioning with total-body irradiation. In a multivariate analysis adjusted for age, disease status, donor source, and HLA match, a high ST2 concentration on day 14 was significantly associated with an increased risk of death without relapse within 6 months after transplantation for all conditioning regimens. The addition of ST2 to the clinical risk factors with the use of the net reclassification index significantly reclassified 78% of patients receiving chemotherapy-based myeloablative conditioning (P<0.001) and 58% of patients receiving myeloablative conditioning with total-body irradiation (P=0.02); it also reclassified 71% of patients receiving reduced-intensity conditioning (P<0.001) (Table 3, and Tables S15A and S15B in the Supplementary Appendix). High ST2 values were not associated with an increased risk of death with relapse within 1 year after transplantation; because high ST2 values were associated with an increased risk of death without relapse, overall survival at 1 year was decreased (Tables S16A through S16D in the Supplementary Appendix). The causes of death are shown in Table S17 in the Supplementary Appendix. Sensitivities and specificities for several cutoff points of ST2

Table 3. Effect of Potential Risk Factors, Including ST2 Concentrations at Day 14 after Transplantation, on the Risk of Death without Relapse at 6 Months after Transplantation.*

Variable	Chemotherapy-Based Myeloablative Conditioning (N = 195)				Chemotherapy-Based Reduced-Intensity Conditioning (N = 107)				Myeloablative Conditioning with Total-Body Irradiation (N = 75)			
	Univariate Analysis		Multivariate Analysis†‡		Univariate Analysis		Multivariate Analysis†§		Univariate Analysis		Multivariate Analysis¶	
	hazard ratio (95% CI)	P value	hazard ratio (95% CI)	P value	hazard ratio (95% CI)	P value	hazard ratio (95% CI)	P value	hazard ratio (95% CI)	P value	hazard ratio (95% CI)	P value
Age: ≥55 yr vs. <55 yr	2.6 (1.3–5.1)	0.008	2.4 (1.1–5.1)	0.02	1.8 (0.5–6.8)	0.36	1.3 (0.3–5.1)	0.72	0.7 (0.2–2.3)	0.53	0.7 (0.2–2.1)	0.49
Disease status: high risk vs. low or intermediate risk	1.2 (0.6–2.2)	0.67	1.1 (0.6–2.1)	0.82	4.0 (1.3–12.8)	0.02	3.5 (1.0–12.4)	0.05				
Donor: unrelated vs. related	1.6 (0.8–3.2)	0.15	1.3 (0.6–2.7)	0.49	1.3 (0.4–3.8)	0.67	0.7 (0.2–2.2)	0.56				
HLA match: mismatched vs. matched	1.9 (1.0–3.6)	0.06	1.6 (0.8–3.2)	0.22	1.0 (0.3–3.7)	0.95	0.7 (0.2–3.1)	0.68	2.0 (0.8–5.1)	0.15	1.9 (0.7–4.9)	0.21
ST2 value: high vs. low‖	5.1 (2.3–11.5)	<0.001	4.4 (1.9–10.4)	0.001	4.4 (1.2–15.5)	0.02	3.9 (1.1–14.1)	0.04	2.8 (1.2- 6.7)	0.02	2.6 (1.1–6.5)	0.04

* Included are data from 302 patients who received chemotherapy-based myeloablative conditioning or reduced-intensity conditioning at the University of Michigan and an independent cohort of 75 patients who received myeloablative conditioning with total-body irradiation at the Dana–Farber Cancer Institute.

† The effect of the ST2 value was calculated with the use of regression models adjusted for age, disease status, donor source, and HLA match.

‡ The AUC was 0.76 in the multivariate model including demographic and clinical characteristics and the ST2 value, as compared with 0.66 in the model including demographic and clinical characteristics alone (P=0.18); the net reclassification index was 78% (P<0.001).

§ The AUC was 0.74 in the multivariate model including demographic and clinical characteristics and the ST2 value, as compared with 0.62 in the model including demographic and clinical characteristics alone (P=0.14); the net reclassification index was 71% (P<0.001).

¶ The effect of the ST2 value was calculated with the use of regression models adjusted for age and HLA match. The AUC was 0.68 in the multivariate model including demographic and clinical characteristics and the ST2 value, as compared with 0.58 in the model including demographic and clinical characteristics alone (P=0.32); the net reclassification index was 58% (P=0.02).

‖ A high ST2 value was defined as more than 1100 pg per milliliter for patients receiving chemotherapy-based myeloablative conditioning, more than 300 pg per milliliter for patients receiving chemotherapy-based reduced-intensity conditioning, and more than 1660 pg per milliliter for patients receiving myeloablative conditioning with total-body irradiation.

are shown in Tables S18A and S18B in the Supplementary Appendix.

DISCUSSION

ST2 was the best single biomarker of nonresponse to GVHD therapy and subsequent death without relapse, and it significantly improved risk stratification according to clinical grade and target-organ stage, which have been shown to be important predictors of risk, particularly in the subgroup of patients with lower gastrointestinal GVHD.[23,24] When ST2 was examined with REG3α in this subgroup, they had a strong additive effect, as compared with the clinical characteristics alone (net reclassification index, 71%) (Table S12 in the Supplementary Appendix).

ST2 is a recently discovered member of the interleukin-1 receptor family,[27] whose only known ligand is interleukin-33.[28] Plasma interleukin-33 concentrations were similar in patients who had a response and those who did not (data not shown). There are two ST2 isoforms: a membrane-bound form expressed on hematopoietic cells, particularly type 2 helper T (Th2) cells,[29-31] which plays a role in Th2-mediated diseases such as asthma,[32-34] and a soluble form, secreted by endothelial cells, epithelial cells, and fibroblasts in response to inflammatory stimuli.[35,36] Soluble ST2 acts as a decoy receptor for interleukin-33[32,33,37] and drives Th2 cells toward a type 1 helper T (Th1)–cell phenotype, which may be important in the pathophysiology of GVHD.

There are several limitations of this study. Although the concentrations of ST2 at the initiation of GVHD therapy were similar across conditioning intensities, soluble ST2 concentrations were two to four times as high in patients receiving myeloablative conditioning as in patients receiving reduced-intensity conditioning, necessitating separate thresholds for each conditioning intensity. These differences may be due to the degree of injury to endothelial, epithelial, and mesenchymal tissues that release ST2, with the greatest amount of injury occurring after myeloablative total-body irradiation. Although the samples were collected prospectively, retrospectively defined data sets were used to develop prediction models; thus, the discriminatory ability of ST2 might be overestimated. The metrics we chose to compare models are only a subset of metrics that are used to evaluate new biomarkers. For example, we did not examine model calibration, which is usually assessed with a Hosmer–Lemeshow test that requires a categorization of fitted probabilities (i.e., deciles). We used a continuous net reclassification index rather than the noncontinuous (traditional) net reclassification index, which requires risk categories, since the risk categories in our setting are arbitrary and not well defined.

Early identification of patients who will not have a response to GVHD therapy is important because these patients are more likely to die from GVHD than are those who have a response.[10-13] GVHD responded poorly to treatment in patients with high ST2 concentrations at the initiation of therapy, and these patients were at increased risk for death, independent of the GVHD grade. The improved risk stratification of patients with GVHD with the use of ST2 may permit early evaluation of additional therapies, before the development of resistant disease. When measured as early as day 14 after transplantation, ST2 concentration was a better predictor of the risk of death than were the other known risk factors, including the age of the recipient, conditioning intensity, donor source, and HLA match.[38] The ability to identify high-risk patients with the use of ST2 soon after transplantation, before the development of GVHD, may permit more stringent monitoring and preemptive interventions. We believe that our results may affect the assessment of GVHD risk before the development of GVHD and at the onset of the clinical signs of GVHD; however, a generalizable definition of high risk has yet to be developed and will require larger studies.

Supported by grants (RC1HL101102, R01CA168814, P01-CA039542, and T32-HL007622) from the National Institutes of Health. The linear trap quadrupole–Fourier-transform mass spectrometer used in this work was funded in part by the M.J. Murdock Charitable Trust.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank the patients and their families; the staff at the University of Michigan Bone Marrow Transplantation Biobank (funded by the Doris Duke Charitable Foundation) and the Pasquarello Tissue Bank at the Dana–Farber Cancer Institute; the clinicians at the University of Michigan Blood and Marrow Transplantation (BMT) Division (Shin Mineishi, John Magenau, Attaphol Pawarode, Gregory Yanik, Carrie Kitko, Jim Connelly, Craig Byersdorfer, Ed Peres, and Steven Goldstein); the University of Michigan clinical-research support staff (Joel Whitfield, Dawn Jones, and Connie Varner); the University of Michigan BMT Division data managers (Pamela Jones, Rachel Young, and Charlotte Zinkus); and the members of Dr. Hanash's laboratory (Vitor Faca and Sharon Pitteri).

REFERENCES

1. Gooley TA, Chien JW, Pergam SA, et al. Reduced mortality after allogeneic hematopoietic-cell transplantation. N Engl J Med 2010;363:2091-101.
2. Jenq RR, van den Brink MR. Allogeneic haematopoietic stem cell transplantation: individualized stem cell and immune therapy of cancer. Nat Rev Cancer 2010;10:213-21.
3. Ferrara JL, Levine JE, Reddy P, Holler E. Graft-versus-host disease. Lancet 2009; 373:1550-61.
4. Weisdorf D, Zhang MJ, Arora M, Horowitz MM, Rizzo JD, Eapen M. Graft-versus-host disease induced graft-versus-leukemia effect: greater impact on relapse and disease-free survival after reduced intensity conditioning. Biol Blood Marrow Transplant 2012;18:1727-33.
5. Cragg L, Blazar BR, Defor T, et al. A randomized trial comparing prednisone with antithymocyte globulin/prednisone as an initial systemic therapy for moderately severe acute graft-versus-host disease. Biol Blood Marrow Transplant 2000; 6:441-7.
6. 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;8:387-94.
7. Bacigalupo A. Management of acute graft-versus-host disease. Br J Haematol 2007;137:87-98.
8. 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;18:1150-63.
9. 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;18:406-13.
10. Levine JE, Logan B, Wu J, et al. Graft-versus-host disease treatment: predictors of survival. Biol Blood Marrow Transplant 2010;16:1693-9.
11. MacMillan ML, DeFor TE, Weisdorf DJ. The best endpoint for acute GVHD treatment trials. Blood 2010;115:5412-7.
12. Westin JR, Saliba RM, De Lima M, et al. Steroid-refractory acute GVHD: predictors and outcomes. Adv Hematol 2011; 2011:601953.
13. Saliba RM, Couriel DR, Giralt S, et al. Prognostic value of response after upfront therapy for acute GVHD. Bone Marrow Transplant 2012;47:125-31.
14. Paczesny S, Krijanovski OI, Braun TM, et al. A biomarker panel for acute graft-versus-host disease. Blood 2009;113: 273-8.
15. Paczesny S, Braun TM, Levine JE, et al. Elafin is a biomarker of graft-versus-host disease of the skin. Sci Transl Med 2010;2:13ra2.
16. Ferrara JL, Harris AC, Greenson JK, et al. Regenerating islet-derived 3-alpha is a biomarker of gastrointestinal graft-versus-host disease. Blood 2011;118:6702-8.
17. 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;119:3854-60.
18. Przepiorka D, Weisdorf D, Martin P, et al. 1994 Consensus Conference on Acute GVHD Grading. Bone Marrow Transplant 1995;15:825-8.
19. Faca V, Coram M, Phanstiel D, et al. Quantitative analysis of acrylamide labeled serum proteins by LC-MS/MS. J Proteome Res 2006;5:2009-18.
20. Faca V, Pitteri SJ, Newcomb L, et al. Contribution of protein fractionation to depth of analysis of the serum and plasma proteomes. J Proteome Res 2007;6: 3558-65.
21. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc 1999;94: 496-509.
22. Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med 2011;30:11-21.
23. Robin M, Porcher R, de Castro R, et al. Initial liver involvement in acute GVHD is predictive for nonrelapse mortality. Transplantation 2009;88:1131-6.
24. MacMillan ML, DeFor TE, Weisdorf DJ. What predicts high risk acute graft-versus-host disease (GVHD) at onset? Identification of those at highest risk by a novel acute GVHD risk score. Br J Haematol 2012;157:732-41.
25. Paczesny S. Discovery and validation of graft-versus-host disease biomarkers. Blood 2013;121:585-94.
26. De Gruttola VG, Clax P, DeMets DL, et al. Considerations in the evaluation of surrogate endpoints in clinical trials: summary of a National Institutes of Health workshop. Control Clin Trials 2001;22:485-502.
27. Dinarello CA. Interleukin-1 in the pathogenesis and treatment of inflammatory diseases. Blood 2011;117:3720-32.
28. Schmitz J, Owyang A, Oldham E, et al. IL-33, an interleukin-1-like cytokine that signals via the IL-1 receptor-related protein ST2 and induces T helper type 2-associated cytokines. Immunity 2005;23:479-90.
29. Yanagisawa K, Naito Y, Kuroiwa K, et al. The expression of ST2 gene in helper T cells and the binding of ST2 protein to myeloma-derived RPMI8226 cells. J Biochem 1997;121:95-103.
30. Löhning M, Stroehmann A, Coyle AJ, et al. T1/ST2 is preferentially expressed on murine Th2 cells, independent of interleukin 4, interleukin 5, and interleukin 10, and important for Th2 effector function. Proc Natl Acad Sci U S A 1998;95: 6930-5.
31. Xu D, Chan WL, Leung BP, et al. Selective expression of a stable cell surface molecule on type 2 but not type 1 helper T cells. J Exp Med 1998;187:787-94.
32. Oshikawa K, Yanagisawa K, Tominaga S, Sugiyama Y. Expression and function of the ST2 gene in a murine model of allergic airway inflammation. Clin Exp Allergy 2002;32:1520-6.
33. Hayakawa H, Hayakawa M, Kume A, Tominaga S. Soluble ST2 blocks interleukin-33 signaling in allergic airway inflammation. J Biol Chem 2007;282:26369-80.
34. Kurowska-Stolarska M, Kewin P, Murphy G, et al. IL-33 induces antigen-specific IL-5+ T cells and promotes allergic-induced airway inflammation independent of IL-4. J Immunol 2008;181:4780-90. [Erratum, J Immunol 2008;181:8170.]
35. Rössler U, Thomassen E, Höltner L, Baier S, Danescu J, Werenskiold AK. Secreted and membrane-bound isoforms of T1, an orphan receptor related to IL-1-binding proteins, are differently expressed in vivo. Dev Biol 1995;168:86-97.
36. Kumar S, Tzimas MN, Griswold DE, Young PR. Expression of ST2, an interleukin-1 receptor homologue, is induced by proinflammatory stimuli. Biochem Biophys Res Commun 1997;235:474-8.
37. Sanada S, Hakuno D, Higgins LJ, Schreiter ER, McKenzie AN, Lee RT. IL-33 and ST2 comprise a critical biomechanically induced and cardioprotective signaling system. J Clin Invest 2007;117:1538-49.
38. Jagasia M, Arora M, Flowers ME, et al. Risk factors for acute GVHD and survival after hematopoietic cell transplantation. Blood 2012;119:296-307.