

# Prospective external validation of biomarkers to predict acute graft-versus-host disease severity

Marie Robin,<sup>1</sup> Raphael Porcher,<sup>2,3</sup> David Michonneau,<sup>1,4</sup> Laetitia Taurines,<sup>1</sup> Flore Sicre de Fontbrune,<sup>1</sup> Aliénor Xhaard,<sup>1</sup> Bastien Oriano,<sup>5</sup> Aurélien Sutra Del Galy,<sup>1</sup> Régis Peffault de Latour,<sup>1</sup> Gérard Socié,<sup>1,4,\*</sup> and Marie-Hélène Schlageter<sup>6,\*</sup>

<sup>1</sup>Service d'Hématologie-Greffe, Hôpital Saint-Louis, APHP, Université de Paris-Cité, Paris, France; <sup>2</sup>CRESS, INSERM, INRAE, Paris, France; and <sup>3</sup>Centre d'Epidémiologie Clinique, Hôpital Hôtel-Dieu, APHP, <sup>4</sup>INSERM U976, <sup>5</sup>Dermatologie, Hôpital Saint-Louis, APHP, and <sup>6</sup>Biologie Cellulaire, Hôpital Saint-Louis, U1131 INSERM, IRS, Université de Paris-Cité, Paris, France

## Key Points

- Biomarker panels at GVHD onset, independently from clinical parameters, were associated with survival and nonrelapse mortality.
- Using 3 different biostatistical tools, biomarkers only slightly improved prediction over clinical parameters.

Acute graft-versus-host disease (GVHD) is still the major contributor to comorbidities and mortality after allogeneic hematopoietic stem cell transplantation. The use of plasmatic biomarkers to predict early outcomes has been advocated in the past decade. The purpose of this prospective noninterventional study was to test the ability of panels including 7 biomarkers (Elafin, HGF, IL2RA, IL8, REG3, ST2, and TNFRI), to predict day 28 (D28) complete response to steroid, D180 overall survival, and D180 nonrelapse mortality (NRM). Using previous algorithms developed by the Ann Arbor/MAGIC consortium, 204 patients with acute GVHD were prospectively included and biomarkers were measured at GVHD onset for all of them. Initial GVHD grade and bilirubin level were significantly associated with all those outcomes. After adjustment on clinical variables, biomarkers were associated with survival and NRM. In addition to clinical variables, biomarkers slightly improved the prediction of overall survival and NRM (concordance and net reclassification indexes). The potential benefit of adding biomarkers panel to clinical parameters was also investigated by decision curve analyses. The benefit of adding biomarkers to clinical parameters was however marginal for the D28 nonresponse and mortality endpoints.

## Introduction

Hematopoietic stem cell transplantation (HSCT) remains the only curative treatment for many hematological diseases. HSCT's efficacy relies on its immune effect against the hematological malignancy referred as the graft-versus-leukemia effect. Unfortunately, this graft-versus-leukemia is also closely related to acute graft-versus-host disease (GVHD), which induces tissue damage, immune defect, and eventually mortality. Acute GVHD occurs in approximately 50% of patients. Age, type of donor, sex, and HLA matching between donor and recipient, type of regimen, and GVHD prophylaxis influence the risk of acute GVHD. Clinical and histological grades of acute GVHD are predictive for outcome.<sup>1-5</sup> Acute GVHD first-line therapy usually consists of high-dose corticosteroids. Unfortunately, up to 50% of patients will not achieve a sustainable complete remission (CR) with steroids and will be at higher risk of mortality. The main risk factor of GVHD-related mortality is the nonresponsiveness to steroids,<sup>6</sup> but response is obviously unknown at GVHD onset. Finding biomarkers able to better predict acute GVHD severity and response to corticosteroids would thus be highly useful to personalize patient management.

Submitted 4 March 2022; accepted 26 May 2022; prepublished online on *Blood Advances* First Edition 6 June 2022; final version published online 17 August 2022. DOI 10.1182/bloodadvances.2022007477.

\*G.S. and M.-H.S. are joint last authors.

Contact the corresponding author for data sharing: marie.robin@aphp.fr.

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

© 2022 by The American Society of Hematology. Licensed under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0), permitting only noncommercial, nonderivative use with attribution. All other rights reserved.

Candidate biomarkers associated with GVHD severity at disease onset have been previously reported (reviewed in Chen and Zeiser<sup>7</sup> and Adom et al<sup>8</sup>). Most of these biomarkers are cytokines or their soluble receptors, chemokines, growth factors, angiogenic factors, or molecules involved in inflammation. Some biomarkers are specifically associated with a GVHD target organ (skin, liver, or gut). Biomarker performances have been evaluated either separately, as composite panels, or as algorithms. Our group previously reported that fecal calprotectin was significantly associated with steroid-refractory gut GVHD and mortality<sup>9</sup> with a C-index discrimination of 0.86 for a clinico-biologic score that included calprotectin,  $\alpha$ -1 antitrypsin, and GVHD grade.

However, American investigators at Ann Arbor have sequentially reported for more than a decade plasma biomarkers related to skin GVHD (Elafin or trappin<sup>10</sup>), gastrointestinal GVHD (regenerating islet-derived 3  $\alpha$  [REG3 $\alpha$ ]<sup>11,12</sup>), liver GVHD (hepatocyte growth factor [HGF]<sup>12</sup>), or different combinations of biomarkers.<sup>13-21</sup> Specifically, these authors reported that 7 molecules could be potent GVHD biomarkers: Elafin, HGF, interleukin-2 receptor  $\alpha$  (IL-2RA), IL-8 (CXCL8), REG3 $\alpha$ , tumor necrosis factor  $\alpha$  receptor-1 (TNFR1), and suppression of tumorigenicity 2 (ST2, the IL33 receptor). These 7 markers have been extensively studied in panels of 2, 3, 4, or 6 biomarkers or alone for ST2.<sup>14-19</sup> Some of these panels have been internally or externally validated, usually using the same laboratory for biomarkers assays. External validations have been performed on different cohorts, but biomarkers assays have always been performed by authors who described seminal discoveries. Thus, our study is an external validation by independent investigators. Furthermore, it is worth noting that the time of measurement of these biomarkers varied significantly between these studies. Indeed, in what authors refers as panel 2 that includes ST2 and REG3 $\alpha$ , which has been elegantly validated in a large international consortium, sampling was done 7 days after GVHD treatment initiation,<sup>14</sup> whereas this same panel was also studied on samples taken 7 days after HSCT in another study.<sup>15</sup> Panels including the 3, 4, and 6 biomarkers and ST2 have been tested on samples at GVHD diagnosis. All these panels are reported to predict GVHD prognosis and mortality. Finally, most recently, a randomized multicenter trial even used Ann Arbor biomarker status panel 2 to risk-stratify patients.<sup>22</sup>

We conducted a prospective single-center study to assess the prognostic value of plasmatic biomarkers using a single-center French population and local assays to analyze biomarkers. The 7 biomarkers Elafin, HGF, IL-2RA, IL-8, REG3, ST2, and TNFR1 were measured at onset of GVHD, before corticosteroid initiation. The risks of noncomplete response to steroids, nonrelapse mortality (NRM), and mortality were tested for ST2 alone and for 4 previously described panels, and the additive value of biomarkers to clinical variables was tested.

## Methods

### Population and definitions

In this study, consecutive adult patients who agreed for the prospective protocol “plasma-inca” and signed the informed consent form were included between 2013 and 2016. The protocol was registered at [www.clinicaltrials.com](http://www.clinicaltrials.com) with the number NCT02254798 and was approved by ethics committee Ile de France IV (institutional review board 00003835). The protocol was supported by 2 national grants

**Table 1. Cytokines and panels tested**

Name	Cytokines	Formula	Cutoff
ST2 <sup>19</sup>	ST2	–	Q3*
Panel-2 <sup>15</sup>	ST2, REG3 $\alpha$	$\log[-\log(1 - p)] = -11.263 + 1.844 \times \log_{10}(\text{ST2}) + 0.577 \times \log_{10}(\text{REG3}\alpha)$	$P = .20$
Panel-3 <sup>16</sup>	ST2, REG3 $\alpha$ , TNFR1	$\log[-\log(1 - p)] = -9.169 + 0.598 \times \log_2(\text{TNFR1}) - 0.028 \times \log_2(\text{REG3}\alpha) + 0.189 \times \log_2(\text{ST2})$	Q3*†
Panel-4 <sup>18</sup>	IL-2R $\alpha$ , TNFR1, HGF, IL-8	$-4656 + 3.43 \times \text{IL-2R}\alpha + 0.48 \times \text{TNFR1} + 0.11 \times \text{HGF} - 0.45 \times \text{IL-8}$	Q3*‡
Panel-6 <sup>17</sup>	Elafin, IL-2R $\alpha$ , TNFR1, HGF, IL-8, REG3 $\alpha$	$0.31 \times \log(\text{elafin}) + 0.13 \times \log(\text{IL-2R}\alpha) + 0.80 \times \log(\text{TNFR1}) + 0.20 \times \log(\text{IL-8}) + 0.02 \times \log(\text{HGF}) + 0.19 \times \log(\text{REG3}\alpha)$	12.8

\*Third quartile.

†In the original publication, 2 cutoffs at  $P = .4$  and  $P = .66$  were used, but in the current data, values of  $P$  were  $> .66$ .

‡In the original publication, the panel score was 1000 times the linear combination presented, then back-transformed on a probability scale through reverse logit, but in the current data the range of values obtained makes it impossible for most observations.

related to the “Agence de la Biomédecine” (AOR 2015) and the “Programme de Recherche Translationnelle en Cancérologie” (PRTC2013). All patients had weekly plasma collection until week 7 and, at acute GVHD onset, if any, and clinical symptoms of GVHD were prospectively collected. The current study focuses on the 204/304 patients who developed acute GVHD requiring systemic high-dose steroid ( $\geq 1$  mg/kg). Patients with acute GVHD were a median of age 45 years and were mostly transplanted for acute leukemia, myelodysplastic syndrome, or myeloproliferative neoplasm ( $n = 146$ , 71%). Most patients (64%) presented with an intermediate Disease Risk Index scores.<sup>23</sup> Sixty patients (29%) were transplanted from an HLA-matched sibling and 136 patients (67%) from an unrelated donor, including 43 HLA-mismatched donors. Peripheral blood stem cells were the sources of stem cells in 162 patients (79%). GVHD prophylaxis was either cyclosporine and methotrexate (57%) or cyclosporine and mycophenolate mofetil (43%). Supplemental Table 1 shows details on patient, disease, and transplant characteristics. Acute GVHD was classified according to the modified Glucksberg criteria.<sup>24</sup> CR to first-line therapy (steroids) was assessed at day 28 after treatment. The study received institutional review board approval from the Comité d'éthique Paris IV and was conducted according to the Declaration of Helsinki.

### Cytokines assays

Blood samples were collected into ethylene-diamine tetra-acetic acid tubes. Plasma was separated from blood cells within 2 hours after blood collection by 2 successive centrifugations and were immediately frozen at  $-80^{\circ}\text{C}$ . Biomarkers Elafin, HGF, IL-2RA, IL-8, ST2, and TNFR1 were measured simultaneously in a multiplex magnetic Luminex Assay (reference LXSAM) from R&D Systems, whereas REG3 was measured by an enzyme-linked immunosorbent assay from MBL (reference 5323). Each sample was tested in duplicate for each cytokine. Luminex assays were read on a BioPlex MAGPIX Multiplex Reader from BioRad and enzyme-linked immunosorbent assays on a Biotek microplate reader. All assays were

**Table 2. Outcome of patients with acute GVHD according to clinical or biological markers**

Variable	N	Day 28 nonresponse		Day 180 survival		Day 180 nonrelapse mortality	
		No.	Estimate (95% CI)	No.*	Estimate (95% CI)	No.	Estimate (95% CI)
All patients	204	66	32% (26-39)	43	79% (74-85)	32	16% (11-21)
<b>Initial liver GVHD</b>							
No	171	41	24% (18-31)	22	87% (82-92)	13	8% (4-12)
Yes	33	25	76% (58-89)	21	36% (23-57)	19	58% (39-73)
<b>Acute GVHD initial grade</b>							
1-2	143	34	24% (17-32)	19	87% (81-92)	11	8% (4-13)
3	61	32	52% (39-65)	24	61% (50-74)	21	34% (23-46)
<b>Age</b>							
<50 y	115	29	25% (18-34)	18	87% (81-92)	14	12% (7-19)
≥50 y	89	37	42% (31-53)	25	61% (50-74)	18	20% (13-29)
<b>Fever</b>							
Yes	91	38	42% (32-53)	24	74% (65-83)	11	10% (5-16)
No	113	28	25% (17-34)	19	83% (77-90)	21	23% (15-32)
<b>No. of clinical variables†</b>							
0	49	8	16% (7-30)	3	94% (87-100)	1	2% (0-25)
1	83	22	27% (17-37)	15	82% (74-91)	10	12% (6-20)
2	58	23	40% (27-53)	17	71% (60-83)	13	22% (13-34)
3	14	13	93% (66->99)	8	43% (23-78)	8	57% (27-79)
<b>ST2</b>							
High	51	25	49% (35-63)	22	57% (45-72)	19	37% (24-50)
Low	153	41	27% (20-35)	21	86% (81-92)	13	8% (5-14)
<b>Panel 2</b>							
High	73	30	41% (30-53)	28	62% (51-74)	22	30% (20-41)
Low	131	36	27% (20-36)	15	89% (83-94)	10	8% (4-13)
<b>Panel 3</b>							
High	51	26	51% (37-65)	22	57% (45-72)	20	39% (26-52)
Low	153	40	26% (19-34)	11	86% (81-92)	12	8% (4-13)
<b>Panel 4</b>							
High	51	28	55% (40-69)	22	57% (45-72)	19	37% (24-50)
Low	153	38	25% (18-32)	21	86% (81-92)	13	8% (5-14)
<b>Panel 6</b>							
High	85	38	45% (34-56)	30	65% (55-76)	26	31% (21-41)
Low	119	28	24% (16-32)	13	89% (84-95)	6	5% (2-10)

\*Number of deaths by day 180.

†Clinical variables considered were initial liver GVHD, age 50 years or more, and initial GVHD grade 3.

performed according to the manufacturers' instructions. Cytokines and panels tested are reported in Table 1.

## Statistical analysis

Data are described as count and percent for categorical variables and median and interquartile range (IQR) for quantitative variables. ST2 and panels were classified as "high" or "low" (Table 1). For panels 2 and 6, the observed values were identical to the original publications and we used the published cutoff. For ST2, panel 3 and panel 4, the distributions of values were different from the original publications; the third quartile was thus chosen as the cutoff discriminating

patients with low or high panels (supplemental Figure 1). Because dichotomizing continuous variables is known to result in a loss of information, we added an analysis in which biomarker panels were analyzed on their original linear predictor scale without dichotomization.

Nonresponse at day 28 (D28) after steroids initiation (ie, not meeting CR criteria at that time) was analyzed as a binary variable and adjusted (multivariable) analyses were performed using logistic regression. This is similar to analyzing CR but allows the presentation of association measures similarly to the other outcomes and an odds ratio >1, indicating a factor associated with unfavorable outcome. Overall survival up to D180 was analyzed

**Table 3. Multivariable models for clinical risk factors**

Variable	Day 28 nonresponse		Day 180 survival		Day 180 nonrelapse mortality	
	OR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
<b>Initial liver GVHD</b>						
No	1		1		1	
Yes	7.11 (2.86-17.7)	<.001	5.18 (2.67-10.1)	<.001	7.07 (3.25-15.4)	<.001
<b>Acute GVHD initial grade</b>						
1-2	1		1		1	
3	2.52 (1.24-5.14)	.011	2.22 (1.14-4.32)	.019	2.74 (1.23-6.12)	.014
<b>Age</b>						
<50 y	1		1		1	
≥50 y	2.13 (1.10-4.15)	.026	1.66 (0.89-3.09)	.11	1.46 (0.71-2.99)	.30
<b>Donor</b>						
Matched sibling	–		1		1	
Others	–		2.00 (0.83-4.82)	.12	1.99 (0.68-5.81)	.21

CB, cord blood; BM, bone marrow; OR, odds ratio; PB, peripheral blood.

using Kaplan-Meier estimator and Cox models. The cumulative incidence of NRM up to D180 was estimated and the association of variables analyzed using cause-specific Cox models, with relapse as a competing risk. Because the aim was to evaluate the clinical utility of biomarkers panels for patients with acute GVHD and to avoid overfitting, a set of clinical predictors was predetermined (initial liver GVHD, acute GVHD grade 3, and age older than 50 years), and we checked that they were associated with nonresponse at D28 in a multivariable logistic model. Adding  $2 \times 2$  interactions between those 3 variables showed that the effects of age and the 2 other variables were not additive. Based on the observed regression coefficients, we simplified the clinical model and considered as high-risk patients with initial liver GVHD, or those aged older than 50 years, or with initial grade 3 GVHD. Because the only data-driven part of this score building was the grouping of combination of variables but not the coefficients or variable selection based on statistical testing, we considered that this would entail minimal optimism or risk of overfitting. The ability of clinical variables and biomarkers to predict outcomes was evaluated by the concordance (C) index. The C-index quantifies how well a marker or a model discriminates between patients with the event of interest and the others. For binary outcomes, it is equal to the area under the receiver operating characteristic curve, and for survival data, it can be viewed as its extension to the censored data setting. It varies from 0.5 to 1, where 0.5 indicates no better discrimination than would be obtained at random, and 1 perfect discrimination. For NRM, the competing risks version of the C-index was used.<sup>25</sup>

To quantify the additional value of biomarker panels compared with the clinical model, we computed the difference in C-index, as well as the category-less (or continuous) net reclassification index (NRIc).<sup>26</sup> In both cases, 95% confidence intervals (CI) were computed by bootstrapping, as recommended.<sup>27</sup> A lower confidence bound  $>0$  would confirm that a biomarker panel adds additional value for predicting the outcome, compared with a simplified clinical model. Further, we also used decision curves analysis (DCA) to assess the incremental value of using the biomarker panels for medical decision making.<sup>28,29</sup> Briefly, decision curves are a graphical

summary of the net benefit of using a risk prediction model for decision, weighing the expected benefit of a possible intervention or treatment for patients correctly identified as high risk with the costs or harms (in a general sense) associated with a false-positive decision. The relative weights of benefits and costs are reflected by the decision threshold on the probability scale.

Analyses were performed using the R software version 4.0.5 (The R Foundation for Statistical Computing, Vienna, Austria). Packages used included pec to compute C-indexes for censored and competing risks data, nrIc for NRIc and censored data, and rmda for decision curve analyses.

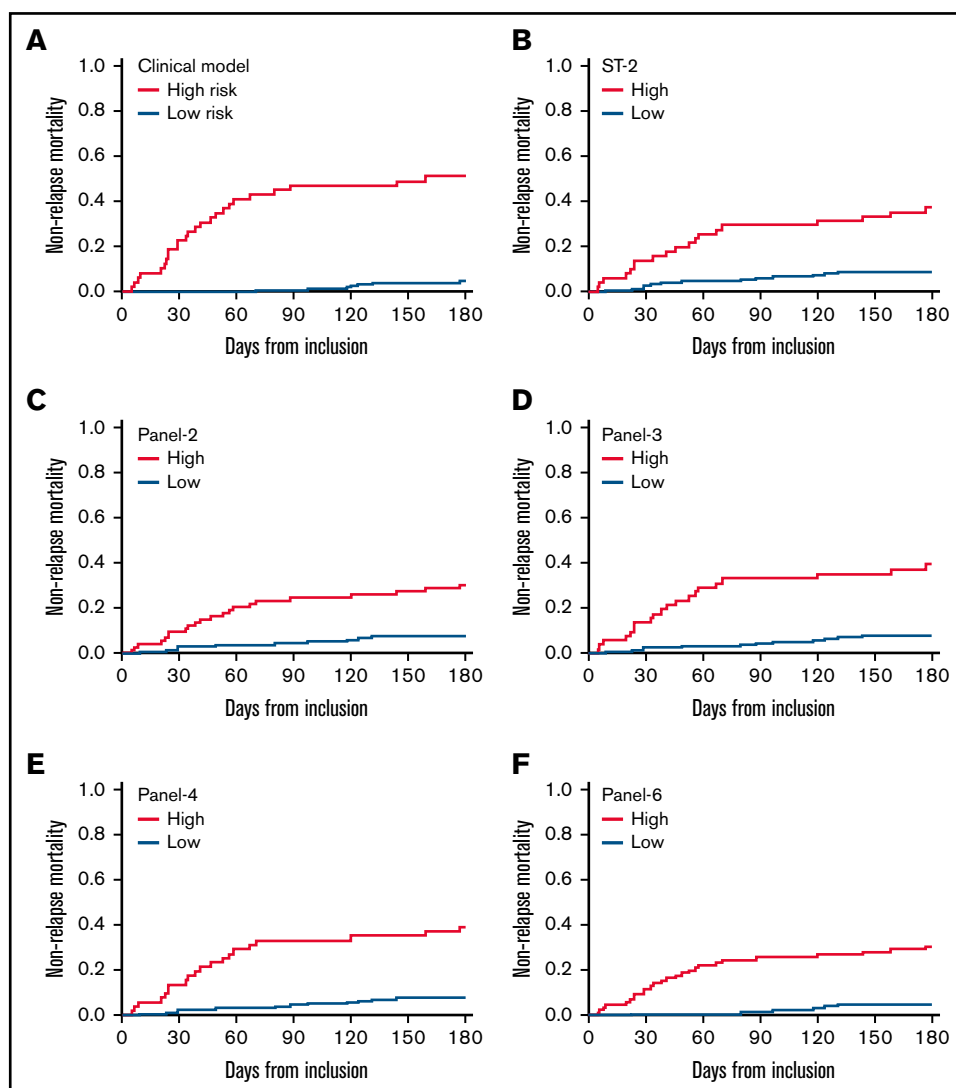
## Results

### Acute GVHD

Acute GVHD occurred in median 24 days (IQR: 15-36) after transplantation. At steroid initiation, patients had grade 1, 2, and 3 in 27 (13%), 116 (57%), and 61 (30%) cases, respectively. Skin, gut, and liver were involved in 169 (83%), 156 (76%), and 31 (15%) patients. Maximal grade 1, 2, 3, and 4 was observed in 19 (9%), 62 (30%), 95 (47%), and 28 (14%) patients, respectively (supplemental Methods; supplemental Table 2). GVHD diagnosis was associated with concomitant fever in 91 (46%) patients: 48 in grade 1-2 and 43 in grade 3. Median albuminemia level was 35 g/L (IQR: 32-39) and median protidemia was 59 g/L (IQR: 55-64) at time of GVHD diagnosis. Gut biopsies were performed in 83 patients with gut involvement, confirming specific signs of GVHD in 76 patients (92%).

### Outcome

**Association of clinical variables on outcome.** Table 2 shows the incidences of D28 nonresponse, D180 overall survival (OS), and D180 NRM: 32% (95% CI, 26-39), 79% (95% CI, 74-85), and 16% (95% CI, 11-21), respectively. Causes of nonrelapse death were attributed to GVHD in the majority of patients (supplemental Table 3). Univariate analysis shows that all 3 outcomes (nonresponse, OS, and NRM) were better in patients with initial grade



**Figure 1. Cumulative incidence of nonrelapse mortality according to clinical factors and cytokines panels.** In the clinical model, high-risk patients are those with initial liver GVHD or those aged 50 years or more with initial GVHD grade 3.

1-2 GVHD or without initial liver involvement (supplemental Table 4). Fever increased the risks of D28 nonresponse and D180 NRM. Age  $\geq 50$  years increased the risk of D28 nonresponse and decreased D180 OS. A transplantation from an unrelated donor increased the risk of D180 NRM. Multivariable analyses (MVA) showed that an initial grade 3 GVHD, age  $\geq 50$  years, and initial liver involvement increased the risk of D28 nonresponse (Table 3). MVA also showed that initial grade 3 and initial liver involvement were both significantly associated with OS and NRM (Table 3). Cumulative incidences of complete response and NRM, and of OS are given for the clinical score (considering significant clinical variables of the MVA) in Table 2, Figure 1A, and supplemental Tables 2A and 3A. The proportions of patients not in CR at D28 were 93%, 40%, 27%, and 16% with 3, 2, 1, or 0 clinical risk factors, respectively. OS was 43% with 3 risk factors vs 94% and 82% when none or 1 risk was present, respectively. NRM was 57% with the 3 clinical risk factors vs 2% or 12%, when none or 1 risk factor was present, respectively.

**Association of plasmatic biomarkers on outcome.** As explained in the statistics section, 25% of patients were classified as high risk according to ST2, panel 3, and panel 4, whereas 36% and 42% were at high risk according to panel 2 and panel 6.

Probabilities of D28 nonresponse, D180 survival, and D180 NRM according to biomarkers are given in Table 2 and corresponding survival curves are shown in Figure 1B-F and supplemental Figures 1 and 2. The univariate analysis showed that all biomarkers were significantly associated with response, OS, and NRM (Table 4). An MVA adjusted on the 3 clinical variables (age, initial GVHD grade, initial liver involvement) is shown in Table 4. None of the panels (or ST2) maintained their association with D28 response. In contrast, biomarkers maintained their significant association with D180 OS and NRM. For survival, the hazard ratio (HR) was the highest with panel 2 (HR: 2.65), followed by panel 6 (HR: 2.30), panel 4 (HR: 2.16), ST2 (HR: 2.12), and panel 3 (1.92). For NRM, the HR



**Table 4. Biomarkers as risk factors for outcomes**

Outcome/cytokine panel	Univariate analysis		Adjusted analysis*	
	OR or HR (95% CI)	P	OR or HR (95% CI)	P
<b>Day 28 nonresponse</b>				
High ST2	2.63 (1.36-5.06)	.004	1.37 (0.63-2.98)	.43
High panel 2	1.84 (1.01-3.37)	.048	1.07 (0.52-2.18)	.86
High panel 3	2.94 (1.52-5.67)	.001	1.28 (0.57-2.86)	.55
High panel 4	3.68 (1.90-7.14)	<.001	2.03 (0.94-4.35)	.070
High panel 6	2.63 (1.44-4.80)	.002	1.39 (0.69-2.81)	.36
<b>Day 180 survival</b>				
High ST2	3.89 (2.14-7.08)	<.001	2.12 (1.13-3.99)	.019
High panel 2	4.00 (2.14-7.50)	<.001	2.65 (1.39-5.06)	.003
High panel 3	4.07 (2.23-7.40)	<.001	1.92 (1.00-3.68)	.050
High panel 4	4.02 (2.21-7.31)	<.001	2.16 (1.14-4.07)	.017
High panel 6	3.92 (2.04-7.52)	<.001	2.30 (1.16-4.55)	.017
<b>Day 180 nonrelapse mortality</b>				
High ST2	5.16 (2.54-10.4)	<.001	2.43 (1.17-5.09)	.018
High panel 2	4.50 (2.13-9.51)	<.001	2.68 (1.25-5.77)	.011
High panel 3	6.07 (2.96-12.4)	<.001	2.45 (1.14-5.28)	.022
High panel 4	5.25 (2.59-10.6)	<.001	2.53 (1.21-5.28)	.014
High panel 6	6.92 (2.85-16.8)	<.001	3.61 (1.44-9.03)	.006

\*Adjusted on a simple clinical model where high-risk patients are those with initial liver GVHD or those aged 50 years or more with initial GVHD grade 3.

was highest with panel 6 (HR: 3.61), followed by panel 2 (HR: 2.68), panel 4 (2.53), panel 3 (HR: 2.45), and ST2 (HR: 2.43).

**Additional value of biomarkers over clinical factors.** We also analyzed the discriminative value of the biomarker panels and of clinical variables by measuring the C-index (Table 5). Biomarkers as well as clinical factors were both able to predict outcome. The difference between C-index ( $\Delta$  C-index) was used to test the added value of biomarkers to clinical parameters. As expected from the MVA, biomarkers did not improve the discrimination for D28 nonresponse (Table 5). However, all biomarkers significantly improved discrimination for OS and NRM over clinical model alone ( $\delta$  C-index > 0). This was confirmed by the NRIc showing positive values when adding biomarkers to clinical factors for D180 OS and NRM. NRIc was higher for NRM than OS, reflecting a better prediction of NRM than of OS. The next step was to determine if the additional value of biomarkers could help clinical management using DCA. These analyses rely on the hypothesis that predicting an outcome (nonresponse for instance), could help medical decision like adding new therapy to corticosteroids in acute GVHD. DCA, strategies based on clinical models (default strategy) or based on clinical models, and biomarkers were compared. Figure 2 shows DCA curves with clinical model and clinical model with ST2, panel 3, panel 4, and panel 6. Using DCA, biomarkers added only modest net benefit except for panel 6.

## Discussion

In the past decade, many studies have highlighted the association of plasmatic biomarkers with prognosis after HSCT. In our study,

we focused on patients who developed acute GVHD to determine if biomarkers at GVHD onset could better predict treatment response and NRM, as previously reported.<sup>10,14-21,30-36</sup> We conducted a single-center prospective study with the aim of externally validating the potent clinical usefulness of these biomarkers at GVHD onset. In our hands, biomarkers are predictive of D180 NRM and OS, thus confirming previous results,<sup>15-17,19</sup> in an independent center and using slightly different measurement methods of biomarkers.

Different classifications have been designed to predict GVHD severity but few have been tested at disease onset, as is the case in our study. To avoid overfitting, we had targeted a predefined limited set of clinical predictors to be adjusted for. Herein, grade at onset of GVHD and initial liver involvement were highly predictive for D28 response and NRM. Indeed, patients with initial grade 3 had a high probability of nonresponse on D28 (52%), whereas patients with grade 1/2 had a low probability of D180-NRM (8%). Furthermore, as we and others previously reported, an initial liver involvement had even a stronger impact on outcome with a D28 nonresponse reaching 76%.<sup>2,37</sup> We also looked at other clinical variables for comparison with previous studies. In our clinical model, albuminemia was not predictive for D28 response or D180 NRM. Age remained an important risk factor for response and survival. Fever, which may reflect cytokine release, was associated with response and NRM, but its impact was no longer significant in the adjusted model when liver involvement was included.

Our study confirms that biomarkers measured at GVHD onset predict response and NRM. We subsequently tested if they could

**Table 5. Discrimination of models and improvement in discrimination for outcomes**

Model	Parameter	Day 28 nonresponse	Day 180 survival	Day 180 nonrelapse mortality
Clinical model*	C-index	0.71 (0.65-0.78)	0.75 (0.68-0.81)	0.81 (0.73-0.86)
ST2	C-index	0.60 (0.53-0.67)	0.65 (0.58-0.72)	0.69 (0.60-0.77)
Panel 2	C-index	0.57 (0.51-0.64)	0.67 (0.60-0.74)	0.68 (0.59-0.75)
Panel 3	C-index	0.61 (0.54-0.68)	0.66 (0.58-0.73)	0.71 (0.62-0.78)
Panel 4	C-index	0.63 (0.56-0.70)	0.66 (0.58-0.73)	0.69 (0.61-0.78)
Panel 6	C-index	0.62 (0.56-0.69)	0.67 (0.59-0.73)	0.72 (0.64-0.78)
Clinical + ST2	C-index	0.71 (0.63-0.78)	0.78 (0.70-0.84)	0.84 (0.76-0.90)
	Δ C-index	−0.002 (−0.02 to 0.02)	0.03 (0.004-0.08)	0.04 (0.01-0.09)
	NRIc	0.38 (−0.35 to 0.67)	0.66 (0.30-0.99)	0.82 (0.42-1.16)
Clinical + panel 2	C-index	0.70 (0.58-0.76)	0.79 (0.70-0.85)	0.84 (0.73-0.89)
	Δ C-index	−0.02 (−0.04 to −0.006)	0.04 (0.006-0.09)	0.03 (−0.002 to 0.07)
	NRIc	0.29 (−0.29 to 0.61)	0.74 (0.42-1.05)	0.78 (0.39-1.09)
Clinical + panel 3	C-index	0.70 (0.61-0.77)	0.77 (0.69-0.84)	0.85 (0.75-0.90)
	Δ C-index	−0.01 (−0.03 to 0.01)	0.03 (<0.001-0.07)	0.04 (0.01-0.09)
	NRIc	0.43 (−0.44 to 0.70)	0.66 (−0.36 to 1.00)	0.89 (0.46-1.20)
Clinical + panel 4	C-index	0.73 (0.66-0.80)	0.78 (0.70-0.84)	0.85 (0.76-0.90)
	Δ C-index	0.02 (−0.01 to 0.05)	0.03 (<0.001-0.08)	0.04 (0.01-0.09)
	NRIc	0.52 (−0.29 to 0.77)	0.66 (−0.26 to 0.98)	0.82 (0.43-1.16)
Clinical + panel 6	C-index	0.72 (0.63-0.78)	0.77 (0.68-0.84)	0.84 (0.73-0.89)
	Δ C-index	0.002 (−0.03 to 0.03)	0.02 (−0.02 to 0.06)	0.03 (−0.001 to 0.08)
	NRIc	0.47 (−0.44 to 0.77)	0.71 (0.37-1.02)	0.94 (0.55-1.21)

Values are estimates with 95% confidence intervals. Δ C-index: difference in C-index compared with the clinical model; NRIc: category-less net reclassification index, using clinical model as comparator. Confidence intervals for Δ C-index and NRIc were obtained with 1000 bootstrap replicates. Confidence intervals that contain 0 indicate no evidence of improvement compared with the clinical model.

\*Clinical model where high-risk patients are those with initial liver GVHD or those aged 50 years or more with initial GVHD grade 3.

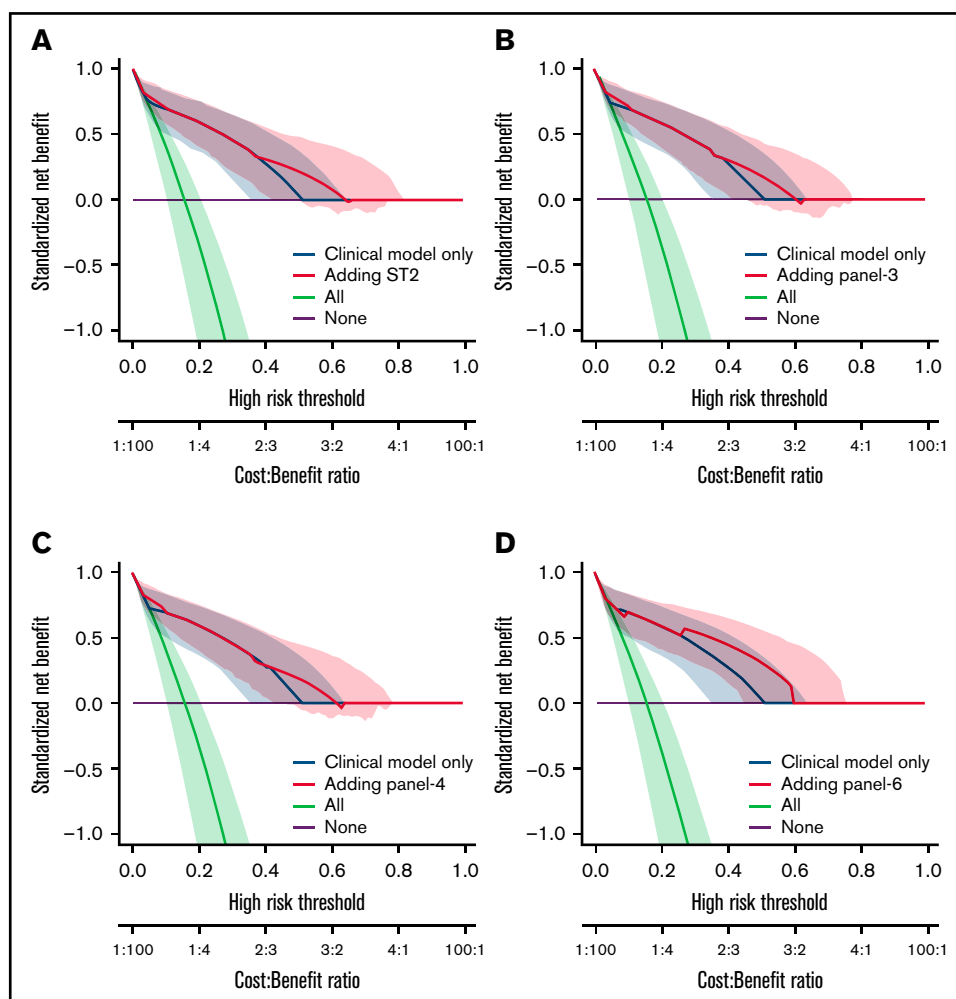
provide an additional value to the clinical predictors (grade, initial liver involvement, and age). First, to predict response, biomarkers did not add predictive value to clinical parameters in MVA in an adjusted Cox regression model. When comparing the discrimination models (C-index), we also confirmed that biomarkers did not add value to clinical model alone. This sharply contrasts with the results of the original panel 6 article, where none of the clinical parameters predicted response to treatment.<sup>17</sup> In contrast, NRM and mortality predictions were both improved by the addition of biomarkers to clinical parameters in the MVA models.

We next used the C-index to quantify how well a marker or a model discriminates between patients. Both clinical model and ST2/panels had C-indexes in the 0.6. to 0.7 range. As described in Table 5, Δ C-indexes and NRIc showed that none of the panels adds significant additional value to clinical model to predict response.

Conversely, there was statistical evidence of slight (though not impressive) improvements over clinical parameters alone to predict mortality. Results were also very similar when analyzing the biomarker panels as continuous variables, so that the limited improvement in prediction should not be explained by a loss of information due to dichotomizing the panels as “high” and “low.” Of note, in the prospective BMT-CTN randomized trial 1501, clinical low risk according to the Minnesota algorithm is highly correlated with low risk according to the Ann Arbor biomarker panel 2. Table 1 shows formula and cutoff for panel 2 used in

Pidala et al's study and in ours. In Pidala et al's study, only 5 of the 127 low-risk patients identified according to clinical criteria had higher risk according to the Ann Arbor biomarker panel.<sup>22</sup> Finally, we asked whether biomarkers would help patients' management using decision curve analyses. Decision curve analyses using panel 6 showed improved benefit ratio in predicting mortality. Overall, this study confirms that biomarkers maintained their significant association with D180 OS and NRM. However, in contrast with Ann Arbor results, we found that some clinical parameters had a strong influence on early endpoints. “In another recent study, it has been reported that biomarkers' additional values as compared with clinical features are only modest; in this study samples and clinical data were taken 14 days after steroid initiation which differed from our (day 0).”<sup>32</sup> This thus raises the points of their utility in clinical practice (from a medico-economic point of view), although their utility in randomized clinical trial to analyze results or risk-stratify patients surely warrants ongoing efforts.

Our study has some limits related to the number of patients that did not enable analyses in clinical subgroups where biomarkers may have a different impact. Also, we did not aim at building full risk prediction models using the biomarker panels and did not assess the calibration of models with those panels. We focused our study at disease onset, assuming that post-GVHD samples may be less predictive than clinical markers for steroid-refractory



**Figure 2. Decision curve analysis.** The decision curves display the net benefit of using a risk prediction model for decision, weighing the expected benefit of a possible intervention or treatment for patients correctly identified as high risk with the costs or harms associated with a false-positive decision. The relative weights of benefits and costs (cost:benefit ratio) are also reflected by the decision threshold on the probability scale. The dashed blue lines correspond to using the clinical model only and are identical on the 4 panels. The decision curves represent the standardized net benefit of using a model (dashed blue or plain red lines) relative to a default strategy of changing management for no patient ("None," black line), or all patients ("All," gray line). The horizontal line ("None") represents the default strategy of changing management for no patients (net benefit is 0 at all thresholds because there is no change in patient management). The line for changing management for all patients (black line) (ie, considering all patients at "positive") starts at 1 at a threshold of 0 because false positives are given no weight relative to true positives. Consequently, falsely tagging an individual as positive has no negative consequence. When the risk threshold (corresponding to a given cost:benefit ratio) increases, false positives are given more weight comparatively to true positives, so that changing management for all patients leads to a decreasing net benefit. The dashed blue or plain red lines, corresponding to using models to predict risk and changing management for those predicted at higher risk, allows us to assess the evidence in favor of using the risk-prediction model over the "None" strategy. The thin lines present bootstrap 95% confidence intervals.

GVHD, in contrast with the MAGIC consortium.<sup>21</sup> The aim of prediction studies may also be to adjust therapy at an earlier time point to prevent steroid-refractory GVHD mortality. Indeed, as a proof of concept, a prospective study has recently proposed a preemptive therapy with  $\alpha$ -1 antitrypsin based on the results of a panel of 2 biomarkers measured at day 7 or 14 after HSCT.<sup>33</sup>  $\alpha$ -1 antitrypsin was given to 30 patients with "high-risk disease" panel 2 level, whereas 90 other high-risk patients were not treated. This study did not show any difference in the outcome of the 2 groups.

To conclude, biomarkers previously published by the Ann Arbor/MAGIC consortium appear as objective tools able to predict the outcomes. However, in our hands, their predictivity was relatively

low ( $C\text{-index} \leq 0.80$ ) adding modest improvements to clinical parameters, casting caution on their use in daily practice.

## Acknowledgments

Funding support for this article was provided by the Institut National du Cancer (InCA) and Organisation de la Direction Générale de l'Offre de Soins (DGOS) with the grant "TRANSLA13-122" and Agence de Biomédecine with the grant "Recherche Greffe 2015".

## Authorship

Contribution: M.R., M.-H.S., and G.S. designed the study; M.-H.S. and L.T. performed the biological analysis; R.P. and B.O. performed



the statistical analysis; M.R., G.S., D.M., F.S.d.F., R.P.d.L., A.X., and A.S.D.G. enrolled and provided patients; M.R. and L.T. reviewed clinical data; M.R., M.H.S., G.S., and R.P. wrote the paper; and all coauthors reviewed the paper and added substantial contributions.

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

**ORCID profiles:** M.R., 0000-0003-1388-9876; D.M., 0000-0003-4553-3065; F.S.d.F., 0000-0003-2000-1556; A.S.D.G., 0000-0002-4716-1472; M.H.S., 0000-0002-9507-0428.

**Correspondence:** Marie Robin, Service d'hématologie – Greffe, Hôpital Saint-Louis, APHP, 1 avenue Claude Vellefaux, 75010 Paris; e-mail: marie.robin@aphp.fr.

## References

1. Ertault-Daneshpouy M, Leboeuf C, Lemann M, et al. Pericapillary hemorrhage as criterion of severe human digestive graft-versus-host disease. *Blood*. 2004;103(12):4681-4684.
2. 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.
3. Xhaard A, Rocha V, Bueno B, et al. Steroid-refractory acute GVHD: lack of long-term improved survival using new generation anticytokine treatment. *Biol Blood Marrow Transplant*. 2012;18(3):406-413.
4. Castilla-Llorente C, Martin PJ, McDonald GB, et al. Prognostic factors and outcomes of severe gastrointestinal GVHD after allogeneic hematopoietic cell transplantation. *Bone Marrow Transplant*. 2014;49(7):966-971.
5. Leisenring WM, Martin PJ, Petersdorf EW, et al. An acute graft-versus-host disease activity index to predict survival after hematopoietic cell transplantation with myeloablative conditioning regimens. *Blood*. 2006;108(2):749-755.
6. Deeg HJ. How I treat refractory acute GVHD. *Blood*. 2007;109(10):4119-4126.
7. Chen S, Zeiser R. Novel biomarkers for outcome after allogeneic hematopoietic stem cell transplantation. *Front Immunol*. 2020;11:1854.
8. Adom D, Rowan C, Adeniyi T, Yang J, Paczesny S. Biomarkers for allogeneic HCT outcomes. *Front Immunol*. 2020;11:673.
9. Rodriguez-Otero P, Porcher R, Peffault de Latour R, et al. Fecal calprotectin and alpha-1 antitrypsin predict severity and response to corticosteroids in gastrointestinal graft-versus-host disease. *Blood*. 2012;119(24):5909-5917.
10. 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):13ra2.
11. Ferrara JLM, Harris AC, Greenson JK, et al. Regenerating islet-derived 3-alpha is a biomarker of gastrointestinal graft-versus-host disease. *Blood*. 2011;118(25):6702-6708.
12. Harris AC, Ferrara JLM, Braun TM, et al. Plasma biomarkers of lower gastrointestinal and liver acute GVHD. *Blood*. 2012;119(12):2960-2963.
13. Wang H, Clouthier SG, Galchev V, et al. Intact-protein-based high-resolution three-dimensional quantitative analysis system for proteome profiling of biological fluids. *Mol Cell Proteomics*. 2005;4(5):618-625.
14. Major-Monfried H, Renteria AS, Pawarode A, et al. MAGIC biomarkers predict long-term outcomes for steroid-resistant acute GVHD. *Blood*. 2018;131(25):2846-2855.
15. 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.
16. 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.
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(16):3854-3860.
18. Paczesny S, Krijanovski OI, Braun TM, et al. A biomarker panel for acute graft-versus-host disease. *Blood*. 2009;113(2):273-278.
19. 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.
20. Choi SW, Kitko CL, Braun T, et al. Change in plasma tumor necrosis factor receptor 1 levels in the first week after myeloablative allogeneic transplantation correlates with severity and incidence of GVHD and survival. *Blood*. 2008;112(4):1539-1542.
21. Srinagesh HK, Özbek U, Kapoor U, et al. The MAGIC algorithm probability is a validated response biomarker of treatment of acute graft-versus-host disease. *Blood Adv*. 2019;3(23):4034-4042.
22. Pidala J, Hamadani M, Dawson P, et al. Randomized multicenter trial of sirolimus vs prednisone as initial therapy for standard-risk acute GVHD: the BMT CTN 1501 trial. *Blood*. 2020;135(2):97-107.
23. 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.
24. Przepiorka D, Weisdorf D, Martin P, et al. 1994 Consensus Conference on Acute GVHD Grading. *Bone Marrow Transplant*. 1995;15(6):825-828.
25. Wolbers M, Blanche P, Koller MT, Witteman JCM, Gerds TA. Concordance for prognostic models with competing risks. *Biostatistics*. 2014;15(3):526-539.

26. Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med*. 2011;30(1):11-21.
27. Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net reclassification indices for evaluating risk prediction instruments: a critical review. *Epidemiology*. 2014;25(1):114-121.
28. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26(6):565-574.
29. Kerr KF, Brown MD, Zhu K, Janes H. Assessing the clinical impact of risk prediction models with decision curves: guidance for correct interpretation and appropriate use. *J Clin Oncol*. 2016;34(21):2534-2540.
30. De Bock M, Beguin Y, Leprince P, et al. Comprehensive plasma profiling for the characterization of graft-versus-host disease biomarkers. *Talanta*. 2014;125:265-275.
31. Budde H, Papert S, Maas J-H, et al. Prediction of graft-versus-host disease: a biomarker panel based on lymphocytes and cytokines. *Ann Hematol*. 2017;96(7):1127-1133.
32. 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.
33. Gergoudis SC, DeFilipp Z, Özbek U, et al. Biomarker-guided preemption of steroid-refractory graft-versus-host disease with  $\alpha$ -1-antitrypsin. *Blood Adv*. 2020;4(24):6098-6105.
34. Ahmed SS, Wang XN, Norden J, et al. Identification and validation of biomarkers associated with acute and chronic graft versus host disease [published correction appears in *Bone Marrow Transplant*. 2016;51(16):890]. *Bone Marrow Transplant*. 2015;50(12):1563-1571.
35. San Min S, Mehra V, Clay J, et al. Composite biomarker panel for prediction of severity and diagnosis of acute GVHD with T-cell-depleted allogeneic stem cell transplants - single centre pilot study. *J Clin Pathol*. 2017;70(10):886-890.
36. Abu Zaid M, Wu J, Wu C, et al. Plasma biomarkers of risk for death in a multicenter phase 3 trial with uniform transplant characteristics post-allogeneic HCT. *Blood*. 2017;129(2):162-170.
37. Robin M, Porcher R, de Castro R, et al. Initial liver involvement in acute GVHD is predictive for nonrelapse mortality. *Transplantation*. 2009;88(9):1131-1136.