SUPPLEMENTAL MATERIAL APPENDIX:

Legend:

Receiver operating characteristic curve for the day +7 MAGIC algorithm for prediction of 6-month non-relapse mortality. The diamond indicates the threshold that defines low versus high risk groups. PPV = positive predictive value, NPV = negative predictive value

FIGURE S2: RELAPSE AND OVERALL SURVIVAL BY DONOR TYPE
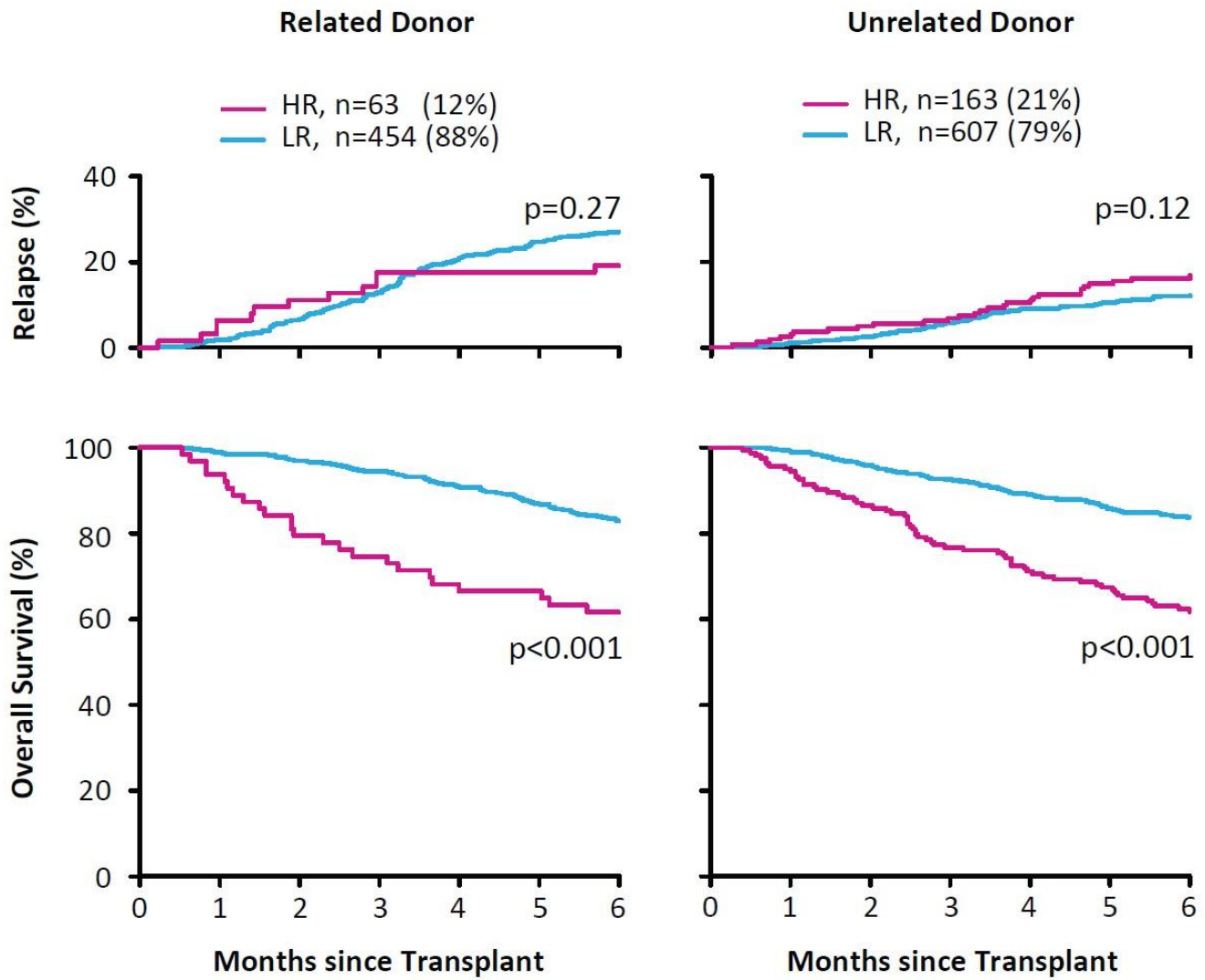
Legend: 6-month relapse rates for all patients (n=1287) were as follows: Related donor: High risk (HR) 19%; Low risk (LR) 27%; Unrelated donor: HR 17%; LR 13%. Gray's test was used for statistical comparisons between groups. 6-month overall survival rates were as follows: Related donor: HR 62%; LR 83%; Unrelated donor: HR 62%; LR 84%. Log-rank test was used for statistical comparisons between groups.

FIGURE S3: RELAPSE AND OVERALL SURVIVAL BY HLA HISTOCOMPATIBILITY
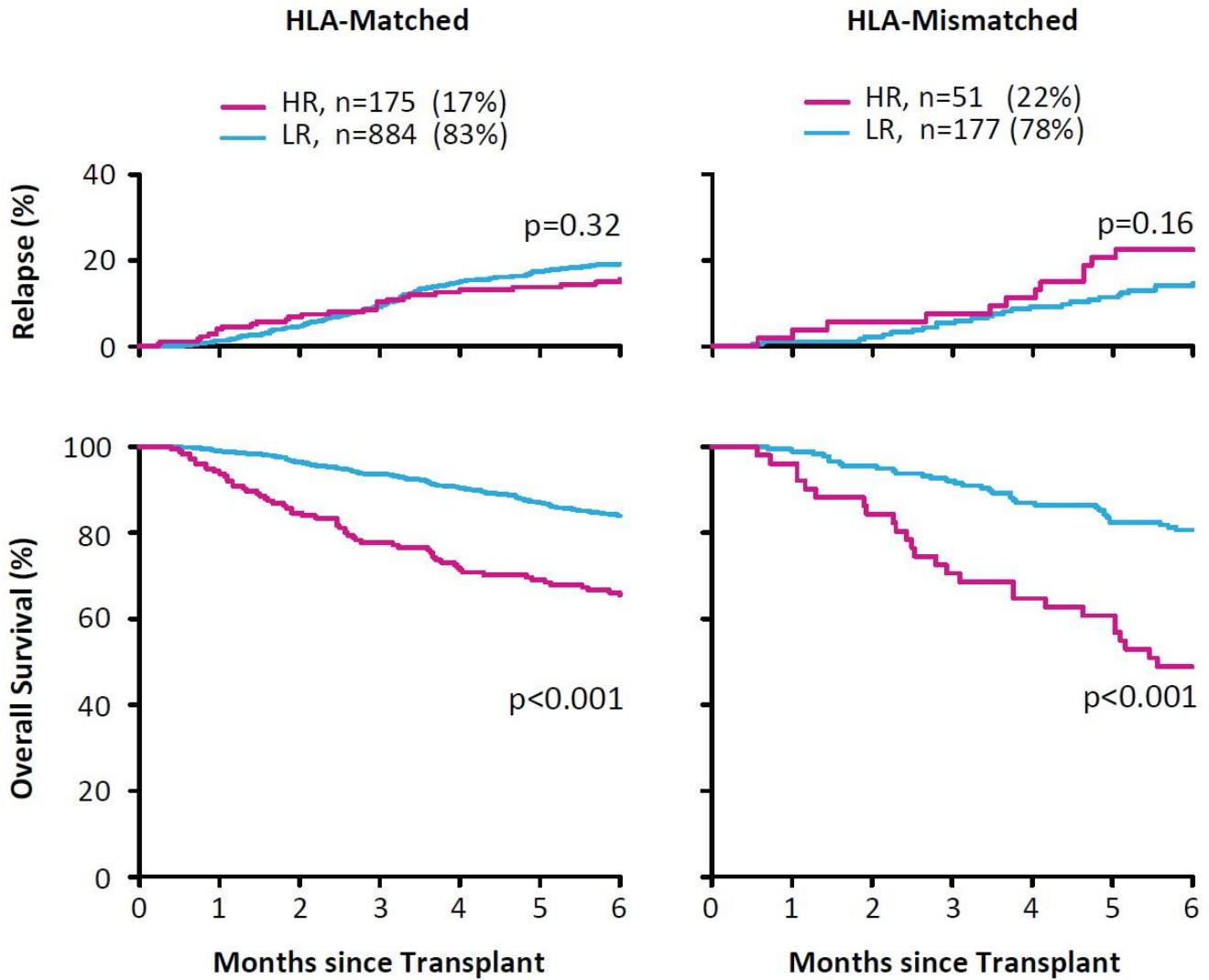
Legend: 6-month relapse rates for all patients (n=1287) were as follows: HLA-matched: High risk (HR) 16%; Low risk (LR) 19%; HLA- mismatched: HR 24%; LR 15%. Gray's test was used for statistical comparisons between groups. 6-month overall survival rates were as follows: HLA-matched: HR 65%; LR 84%; HLA-mismatched: HR 49%; LR 81%. Log-rank test was used for statistical comparisons between groups.
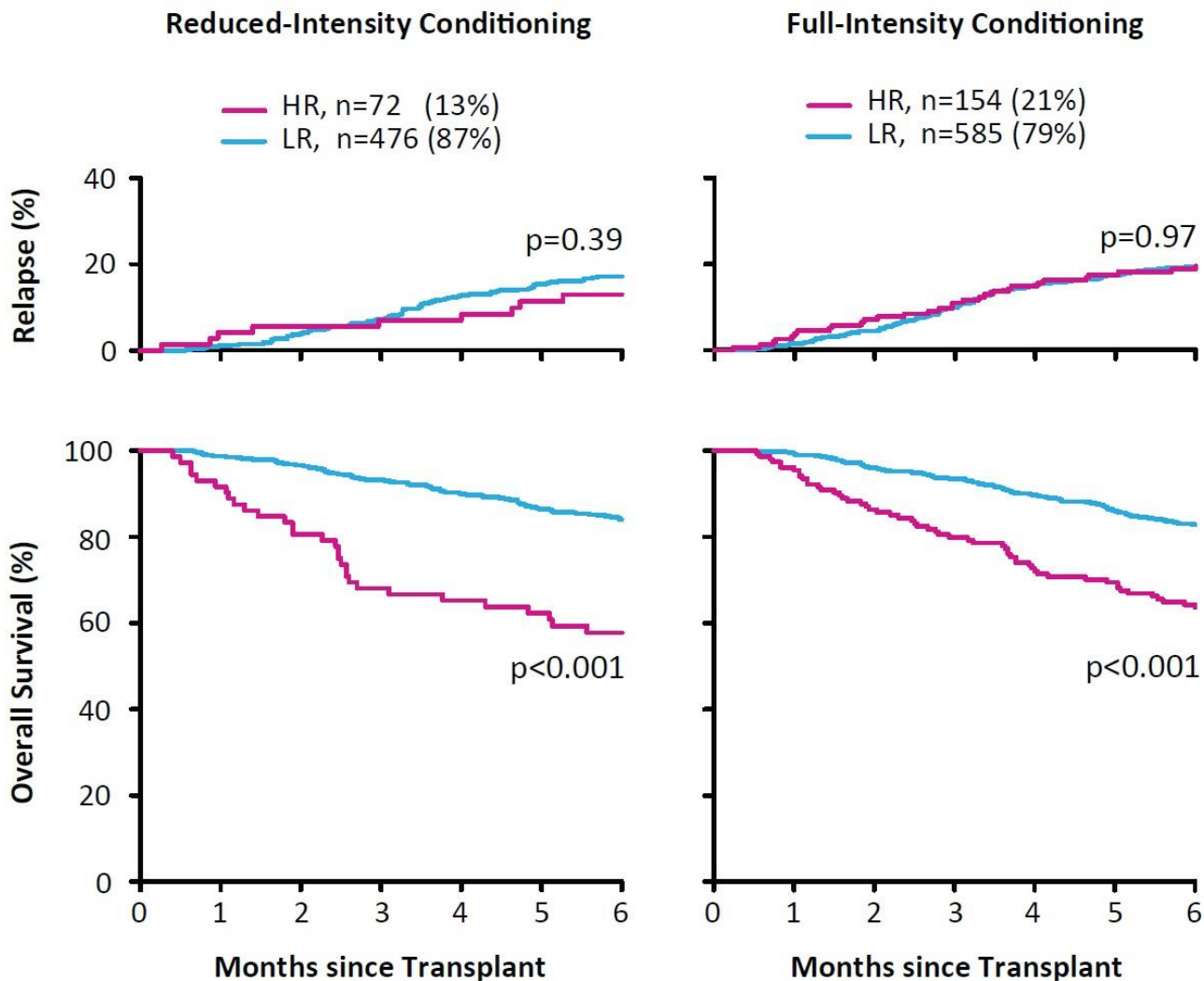
Legend: 6-month relapse rates for all patients (n=1287) were as follows: Reduced-intensity conditioning: High risk (HR) 13%; Low risk (LR) 17%; Full-intensity conditioning: HR 20%; LR 20%. Gray's test was used for statistical comparisons between groups. 6-month overall survival rates were as follows: Reduced-intensity conditioning: HR 58%; LR 84%; Full-intensity conditioning: HR 64%; LR 83%. Log-rank test was used for statistical comparisons between groups.

FIGURE S5: GVHD-RELATED MORTALITY BY MAGIC RISK STRATIFICATION

A   Training Set (N=620)

B   Test Set (N=309)

C   Validation Set (N=358)
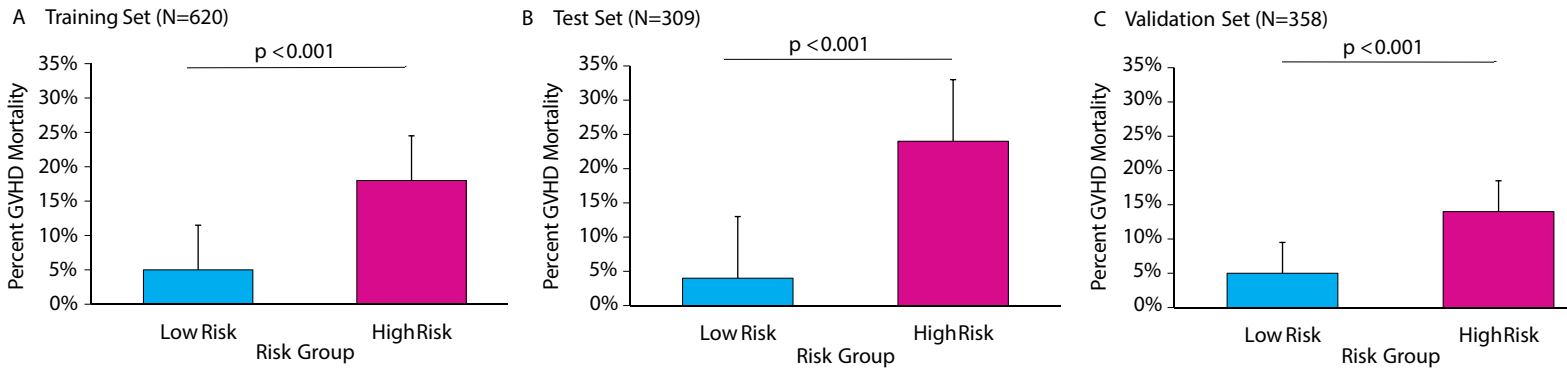
Legend:

Graft-vs-host disease (GVHD)-related mortality was assessed for High risk (HR) and low risk (LR) patients in each patient set. (A) Training set: HR 18% vs LR 5%, p<0.001. (B) Test set: HR 24% vs LR 4%, p<0.001. (C) Validation set: HR 14% vs LR 5%, p<0.001. The chi-squared test was used to compare proportions between groups.

FIGURE S6: ORGAN SPECIFIC GVHD OUTCOMES BY DAY 7 RISK STRATIFICATION
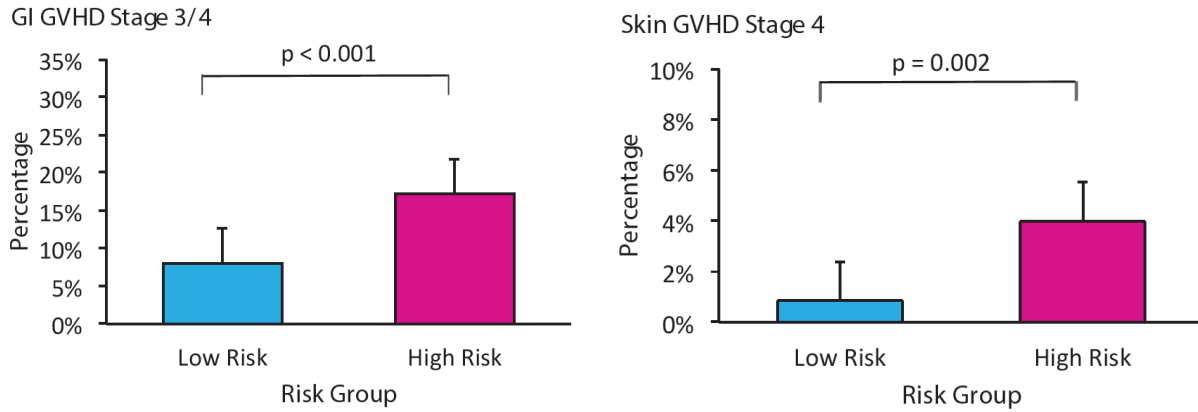
Legend:

Graft-vs-host disease (GVHD) outcomes were assessed for all 1287 patients according to risk group, HR (n=226) and LR (n=1061). (A) Severe gastrointestinal (GI) GVHD (Stage 3-4): HR 17% vs LR 8%, p<0.001. (B) Severe skin GVHD (Stage 4): HR 4% vs LR 0.7%, p=0.002. Chi-squared test was used to compare the proportions between groups.

TABLE S1: PATIENT OUTCOMES BY SET

Characteristic	Training set (N=620)	Test set (N=309)	Validation set (N=358)
Overall 6-month NRM - no. (%)	68 (11.0)	37 (12.0)	47 (13.1)
GVHD grade at maximum - no. (%)			
No GVHD	274 (44.2)	133 (43.0)	180 (50.3)
I	113 (18.2)	40 (12.9)	55 (15.4)
II	138 (22.3)	81 (26.2)	70 (19.6)
III	50 (8.1)	32 (10.4)	35 (9.8)
IV	45 (7.3)	23 (7.4)	18 (5.0)
GVHD onset - median day (range)	28 (7-175)	29 (8-168)	29 (10-177)
GVHD onset ≥ 14 days post HCT - no. (%)	311 (90)	158 (90)	166 (93)

TABLE S2: BIOMARKER MEDIAN AND RANGE BY SET

	Training				Test				Validation			
	REG3α	IL2Rα	TNFR1	ST2	REG3α	IL2Rα	TNFR1	ST2	REG3α	IL2Rα	TNFR1	ST2
Low Risk	27 (8-889)	971 (31-10928)	2594 (362-11602)	15266 (1495-72815)	24 (8-658)	938 (31-6092)	2599 (222-8716)	15979 (1771-73669)	24 (8-458)	957 (31-8303)	2175 (679-12226)	16972 (2977-74233)
High Risk	85 (8-2111)	2418 (31-12745)	5011 (1078-19708)	81684 (22246-177566)	96 (8-5072)	2279 (31-9486)	4742 (2152-25691)	76792 (21062-172851)	98 (8-1141)	2315 (571-9155)	4615 (2234-16886)	64709 (24069-199143)

TABLE S3: NON-RELAPSE MORTALITY AT 9 AND 12 MONTHS BY SET

	Training	Test	Validation
9-month			
HR	0.34	0.35	0.31
LR	0.10	0.09	0.15
12-month			
HR	0.39	0.39	0.31
LR	0.11	0.12	0.19

TABLE S4: UNIVARIATE ANALYSIS OF CLINICAL RISK FACTORS FOR NRM

Variable	Training Set (n=620)	
	Hazard Ratio (95% CI)	p-value
Donor: unrelated vs. related	3.07 (1.65-5.73)	<0.001
HLA match: matched vs. mismatched	1.89 (1.10-3.25)	0.02
Conditioning: reduced vs. full	0.78 (0·47-1.29)	0.34
Age†	1.12 (0.93-1.35)	0.24

† Reflects 10-year increases in age

TABLE S5: NON-RELAPSE MORTALITY FOR RISK GROUPS STRATIFIED BY INDIVIDUAL SETS

	MAGIC Risk Group	Training Set (N=620)				Test Set (N=309)				Validation Set (N=358)			
		Sample Size	6-Month NRM (%)	95% CI (%)	p-value	Sample Size	6-Month NRM (%)	95% CI (%)	p-value	Sample Size	6-Month NRM (%)	95% CI (%)	p-value
Donor Type													
Related	Low	217	3	1-6	p=0.001	112	4	1-8	p<0.001	125	10	5-16	p=0.013
	High	29	18	6-34		17	35	14-58		17	29	10-52	
Unrelated	Low	303	10	7-14	p<0.001	143	11	6-16	p=0.001	161	11	6-16	p=0.006
	High	71	33	22-44		37	32	18-48		55	26	15-38	
HLA													
Matched	Low	431	6	4-9	p<0.001	219	6	4-10	p<0.001	234	9	6-14	p=0.004
	High	82	26	17-36		37	30	16-45		56	23	13-35	
Mismatched	Low	89	12	7-20	p=0.005	36	14	5-27	p=0.036	52	13	6-24	p=0.026
	High	18	39	16-61		17	41	18-64		16	38	15-61	
Conditioning Intensity													
Reduced	Low	234	6	3-10	p<0.001	120	11	6-17	p=0.001	122	11	6-17	p=0.001
	High	30	35	18-53		16	44	19-67		26	35	17-53	
Full	Low	286	8	5-12	p<0.001	135	4	2-9	p<0.001	164	10	6-15	p=0.028
	High	70	26	16-36		38	29	15-44		46	22	11-35	
Age													
≤ 21	Low	62	5	1-12	p<0.001	23	9	1-25	p=0.470	33	6	1-18	N/A
	High	7	57	13-86		12	17	2-43		3	N/A	N/A	
> 21	Low	458	7	5-10	p<0.001	232	7	4-11	p<0.001	253	11	7-15	p<0.001
	High	93	26	18-36		42	38	23-53		69	28	18-38	

TABLE S6: CAUSES OF NON-RELAPSE MORTALITY BY RISK GROUP (N=1287)

Cause – no. (%)	Low Risk (N = 1061)	High Risk (N = 226)
Acute GVHD	47 (4.43)	41 (18.14)
Chronic GVHD	3 (0.28)	2 (0.88)
Infection unrelated to GVHD	15 (1.41)	7 (3.10)
Graft failure	2 (0.19)	2 (0.88)
Pulmonary events*	10 (0.94)	3 (1.33)
Cardiac events	2 (0.19)	2 (0.88)
Neurologic events	1 (0.09)	3 (1.33)
New malignancy**	3 (0.28)	0
VOD/SOS	0	2 (0.88)
Other	0	2 (0.88)
Total	83 (7.82)	64 (28.32)

Includes diffuse alveolar hemorrhage, idiopathic pneumonia syndrome, and respiratory failure not otherwise specified

*** Includes 2 post-transplant lymphoproliferative disorder, 1 metastatic gastric adenocarcinoma*

TABLE S7: THREE BIOMARKER ALGORITHM AT ONSET OF GVHD

Ann Arbor Score	Three biomarker Algorithm (Original)		Three biomarker Algorithm (Re-derived)	
	N (%) (Total N=328)	6-month NRM	N (%) (Total N=589)	6-month NRM
AA1	131 (40%)	9%	85 (14%)	7%
AA2	119 (36%)	28%	388 (66%)	18%
AA3	78 (24%)	47%	116 (20%)	45%

The ELISA test for ST2 (R&D Systems, Minneapolis, MN) used to derive the original three biomarker algorithm was discontinued and was no longer available for use in this study. We used the new ELISA test (R&D Systems, Minneapolis, MN) to measure all ST2 concentrations in this study. We re-derived the original algorithm to predict 6-month non-relapse mortality (NRM) using samples from 589 patients taken at the onset of GVHD symptoms. We employed an identical approach, fitting a competing risks regression model that used log-transformed concentrations of all three biomarkers [ST2, Reg3α and TNFR1] and which produced the following algorithm:

$$\log[-\log(1-p)] = -10.925 + 0.642 \, (\log_{10}TNFR1) + 0.845 \, (\log_{10}Reg3\alpha) + 1.182 \, (\log_{10}ST2)$$

Using this algorithm we computed a predicted probability, *p*, of non-relapse mortality within 6 months for each patient. We then rank ordered *p* and identified two thresholds (0.08 and 0.32) so that the 6-month NRM of the low risk (Ann Arbor 1) and high risk (Ann Arbor 3) groups were similar to the 6-month NRM of the original Ann Arbor groups, but without consideration of the proportion of patient in each group, since that was not a consideration in generating the original algorithm. This table shows the NRM and proportion of patients for each Ann Arbor (AA) score as originally published and the re-derived three biomarker algorithm.

TABLE S8: COMPARISON OF ALGORITHMS WITH AND WITHOUT CLINICAL RISK FACTORS*

	MAGIC Risk Group	A. Training set (N=620)				B. Test set (N=309)				C. Validation set (N=358)			
		Proportion	6-month NRM	Δ NRM	p-value	Proportion	6-month NRM	Δ NRM	p-value	Proportion	6-month NRM	Δ NRM	p-value
Biomarkers without clinical risk factors	High	16%	28%			17%	33%			20%	26%		
	Low	84%	7%	21%	<0.001	83%	7%	26%	<0.001	80%	10%	16%	<0.001
Biomarkers with clinical risk factors	High	19%	30%			16%	26%			24%	24%		
	Low	81%	6%	24%	<0.001	82%	7%	21%	<0.001	76%	10%	14%	<0.001

*Clinical risk factors included based on univariate analyses (Table S3) were donor type and HLA-match

TABLE S9: MAGIC CENTERS REPRESENTED IN THE MULTICENTER VALIDATION SET

SET	CENTER	LOCATION	SAMPLES
Training (N=620)	University of Michigan	Ann Arbor, Michigan	428
	University of Regensburg	Regensburg, Germany	192
Test (N=309)	University of Michigan	Ann Arbor, Michigan	214
	University of Regensburg	Regensburg, Germany	95
Validation (N=358)	Mayo Clinic	Rochester, Minnesota	100
	University of Hamburg	Hamburg, Germany	90
	Ohio State University	Columbus, Ohio	47
	University of Pennsylvania	Philadelphia, Pennsylvania	29
	King Chulalongkorn Memorial Hospital - Bangkok	Bangkok, Thailand	27
	Emory University	Atlanta, Georgia	21
	Icahn School of Medicine at Mount Sinai	New York, New York	18
	University of Würzburg	Würzburg, Germany	14
	University Hospital Carl Gustav Carus Dresden	Dresden, Germany	12

TABLE S10: REPRESENTATIVE THRESHOLDS THAT DEFINE RISK GROUPS

Thresholds	MAGIC Risk Group	A. Training set (N=620)				B. Test set (N=309)				C. Validation Set (N=358)			
		Proportion	6-month NRM	Δ NRM	p-value	Proportion	6-month NRM	Δ NRM	p-value	Proportion	6-month NRM	Δ NRM	p-value
0.14	High	21%	24%	17%	<0.001	22%	27%	19%	<0.001	25%	24%	14%	<0.001
	Low	79%	7%			78%	8%			75%	10%		
0.15	High	19%	26%	19%	<0.001	19%	30%	22%	<0.001	24%	26%	16%	<0.001
	Low	81%	7%			81%	8%			76%	10%		
0.16	High	16%	28%	21%	<0.001	17%	33%	26%	<0.001	20%	26%	16%	<0.001
	Low	84%	7%			83%	7%			80%	10%		
0.17	High	15%	29%	22%	<0.001	15%	33%	25%	<0.001	19%	25%	14%	<0.001
	Low	85%	7%			85%	8%			81%	11%		
0.18	High	14%	28%	20%	<0.001	14%	33%	24%	<0.001	16%	27%	16%	<0.001
	Low	86%	8%			86%	9%			84%	11%		