Supplemental Appendix

The MAGIC Algorithm Probability (MAP) is a response biomarker of treatment for acute graft-versus-host disease

Authors:
Hrishikesh K. Srinagesh, B.A., Umut Özbek, Ph.D., Urvi Kapoor, M.D., Francis Ayuk, M.D., Mina Aziz, M.D., Kaitlyn Ben-David, B.A., Hannah K. Choe, M.D., Zachariah DeFilipp, M.D., Aaron Etra, M.D., Stephan A. Grupp, M.D., Matthew J. Hartwell, M.D., Elizabeth O. Hexner, M.D., William J. Hogan, M.B., B.Ch, Alexander B. Karol, B.A., Stelios Kasikis, M.D., Carrie L. Kitko, M.D., Steven Kowalyk, B.S., Hannah Major-Monfried, M.D., Jung-Yi Lin, M.S., Stephan Mielke, M.D., Pietro Merli, M.D., George Morales, B.S., Rainer Ordemann, M.D., Michael A. Pulsipher, M.D., Muna Qayed, M.D., Pavan Reddy, M.D., Ran Reshef, M.D., Wolf Rösler, M.D., Karamjeet S. Sandhu, M.D., Tal Schechter, M.D., Jay Shah, M.B., B.S., Keith Sigel, M.D., Daniela Weber, M.D., Matthias Wölfl, M.D., Kitsada Wudhikarn, M.D., Rachel Young, B.A., John E. Levine*, M.D., James L.M. Ferrara*, M.D.

* JEL and JLMF contributed equally to this work

Supplemental Tables

Supplemental Figures

Table S1. Patient numbers from each center.

Participating MAGIC Centers	Training Cohort (n=248)	Validation Cohort (n=367)
Bambino Gesù Children's Hospital, Rome, Italy	-	7
Children's Hospital of Los Angeles, Los Angeles, CA	-	8
Children's Hospital of Philadelphia, Philadelphia, PA	-	1
City of Hope Comprehensive Cancer Center, Duarte, CA	-	15
Columbia University Medical Center, New York, NY	-	11
Emory University, Atlanta, GA	6	18
Erlangen University, Erlangen, Germany	-	42
Hospital for Sick Children, Toronto, Canada	-	3
Icahn School of Medicine at Mount Sinai Hospital, New York, NY	14	32
King Chulalongkorn Memorial Hospital, Bangkok, Thailand	5	2
Massachusetts General Hospital, Boston, MA	3	41
Mayo Clinic, Rochester, MN	6	15
Ohio State University, Columbus, OH	14	40
University Hospital Carl Gustav Carus, Dresden, Germany	3	1
University Medical Center Hamburg, Hamburg, Germany	37	52
University of Michigan, Ann Arbor, MI	127	3
University of Pennsylvania Health System, Philadelphia, PA	10	11
University of Regensburg, Regensburg, Germany	17	44
Würzburg University Medical Center, Würzburg, Germany	6	10
Vanderbilt University, Nashville, TN	-	11

Table S2. Systemic therapies for acute GVHD.

A. First-Line GVHD Therapy

Primary Therapy for GVHD	Training Cohort (n, %)	Validation Cohort (n,%)
Total	248 (100%)	367 (100%)
Steroids only	198 (80%)	344 (94%)
Additional First-Line Agent:		
Anti-TNF agent	18 (7%)	8 (2%)
ECP	11 (4%)	0 (0%)
MMF	15 (6%)	3 (1%)
Sirolimus	1 (1%)	3 (1%)
Ruxolitinib	0 (0%)	4 (1%)
Others[1]	5 (2%)	5 (1%)

B. Second-Line GVHD Therapy Administered in First Month of Treatment

Second Line Agent	Training Cohort (n=43)	Validation Cohort (n=47)
Alemtuzumab	1	0
Alpha-1 antitrypsin	1	1
ATG	3	4
ATG + etanercept	1	0
Basiliximab	2	0
ECP	10	8
Etanercept	4	4
Etanercept + ECP	2	1
Etanercept + ruxolitinib	0	1
Infliximab	4	1
Infliximab + ECP	1	0
MTX	0	1
MMF	4	5
MMF + Sirolimus + ECP	1	0
Ruxolitinib	5	14
Ruxolitinib + ECP	1	0
Tocilizumab	3	4
Vedolizumab	0	3

Abbreviations: TNF – Tumor Necrosis Factor, ECP – Extra Corporeal Photopheresis, MMF – Mycophenolate mofetil, MTX - methotrexate

1. Other agents included vorinostat, tocilizumab, anti-thymocyte globulin, rituximab, natalizumab, MMF vs placebo, itacitinib vs placebo

Table S3. Causes of death for all patients.

Cause of Death	Training Cohort (n=248)	Validation Cohort (n=367)
Total n (%)	89 (100%)	105 (100%)
Acute GVHD with and without infection	50 (56%)	61 (58%)
Chronic GVHD	4 (4%)	7 (7%)
Other:	8 (9%)	14 (13%)
Infection Unrelated to GVHD	*4 (4%)*	*8 (8%)*
Cardiac Event	*2 (2%)*	*1 (1%)*
Pulmonary Event	*1 (1%)*	*1 (1%)*
Rejection/Poor Graft Function	*1 (1%)*	*1 (1%)*
TMA	*0 (0%)*	*1 (1%)*
Unknown	*0 (0%)*	*2 (2%)*
Relapse	26 (29%)	23 (22%)

TMA: thrombotic microangiopathy. The percent of non-relapse deaths attributable to GVHD was 86% in the training cohort and 83% in the validation cohort.

Table S4. Univariable analysis to predict six month NRM in the training cohort (n=248).

	Hazard Ratio (95% CI)	P-value
Age (every 10 years)	**1.18** (1.00 – 1.38)	**0.049**
Cell source		
PBSC vs BM	1.86 (0.75 – 4.62)	0.36
Cord vs BM	2.61 (0.89 – 7.67)	0.16
Indication		
Lymphoma vs Acute leukemia	1.06 (0.42 – 2.66)	0.96
MDS/MPN vs Acute leukemia	1.41 (0.72 – 2.78)	0.96
Non-Malignant vs Acute leukemia	0.45 (0.06 – 3.60)	0.96
Other Malignant vs Acute leukemia	2.16 (1.06 – 4.40)	0.14
Initial dose of corticosteroid (per mg/kg methylprednisolone equivalent)	1.42 (0.90 – 2.23)	0.13
Minnesota risk at treatment initiation		
High vs. standard	**3.96** (2.31 – 6.77)	**<0.01**
Conditioning Regimen Intensity		
Full vs. reduced	0.58 (0.33 – 1.03)	0.06
Donor type		
Unrelated vs Related	1.10 (0.58 – 2.08)	0.78
HLA match		
Mismatched vs Matched	1.59 (0.93 – 2.73)	0.089
Prophylaxis		
CNI / MMF ± Other vs CNI / MTX ± others	1.38 (0.79 – 2.42)	0.52
Other vs CNI / MTX ± others	2.44 (0.97 – 6.10)	0.11

Abbreviations: PBSC – peripheral blood stem cells; BM: bone marrow; MDS: myelodysplastic syndromes; MPN: myeloproliferative neoplasms; CNI: calcineurin inhibitor; MMF: mycophenolate mofetil; MTX: methotrexate. P-values were adjusted for multiple comparisons using the Holm-Bonferroni method. Variables associated with significant p-values are shown in bold.

Table S5. Multivariable analyses to predict six month NRM (validation cohort, n = 367).
A. Week 4 MAP (> 0.290) and age at transplant

	Hazard Ratio (95% CI)	**P-value**
Age at transplant (10 year increments)	1.1 (0.91 – 1.2)	0.53
High vs. Low MAP	**9.7** (5.7 – 15.6)	**<0.01**

B. Week 4 MAP (>0.290) and Minnesota risk at treatment initiation

	Hazard Ratio (95% CI)	**P-value**
High vs standard Minnesota risk	**2.00** (1.1 – 3.6)	**0.02**
High vs low MAP	**8.45** (4.8 – 14.8)	**<0.01**

Table S6. Crude proportion of patients who experience NRM for a range of thresholds.

Week 4 MAP Threshold	High MAP (n)	Low MAP (n)	NRM High MAP	NRM Low MAP	P-value
0.270	80	287	50.0%	6.3%	<0.0001
0.275	77	290	51.9%	6.2%	<0.0001
0.280	74	293	52.7%	6.5%	<0.0001
0.285	72	295	51.4%	7.1%	<0.0001
0.290	71	296	50.7%	7.4%	<0.0001
0.295	70	297	51.4%	7.4%	<0.0001
0.300	68	299	51.5%	7.7%	<0.0001
0.305	66	301	53.0%	7.6%	<0.0001

Table S7. Sensitivity, specificity, positive predictive value, and negative predictive value of six month NRM for MAPs and clinical response after four weeks of treatment.

	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
MAP	**62.1** (48.4,74.5)	**88.7** (84.7,92.0)	**50.7** (38.6,62.8)	**92.6** (89.0,95.3)
Clinical Response	**60.3** (46.6,73.0)	**78.7** (73.7,83.1)	**34.7** (25.5,44.8)	**91.4** (87.4,94.5)

Table S8. Cumulative incidences of six month NRM and MAP by Ann Arbor Score.

Ann Arbor Score	Average MAP (range)	6 Month NRM (95% CI)
1	0.08 (0.017 – 0.138)	6% (3% – 10%)
2	0.21 (0.141 – 0.289)	20% (13% - 29%)
3	0.43 (0.293 – 0.762)	49% (34% - 62%)

Table S9. Algorithm combining biomarker concentrations and clinical response to four weeks of therapy.

$\log[-\log(1-\hat{p})] = -15.938 + 2.681(\log_{10}\text{ST2}) + 0.474(\log_{10}\text{REG3}\alpha) + 1.454(\text{Week 4 Response})$

Each variable was a significant predictor or NRM. The concentrations of ST2 are reported as pg/ml and of REG3α as ng/ml.

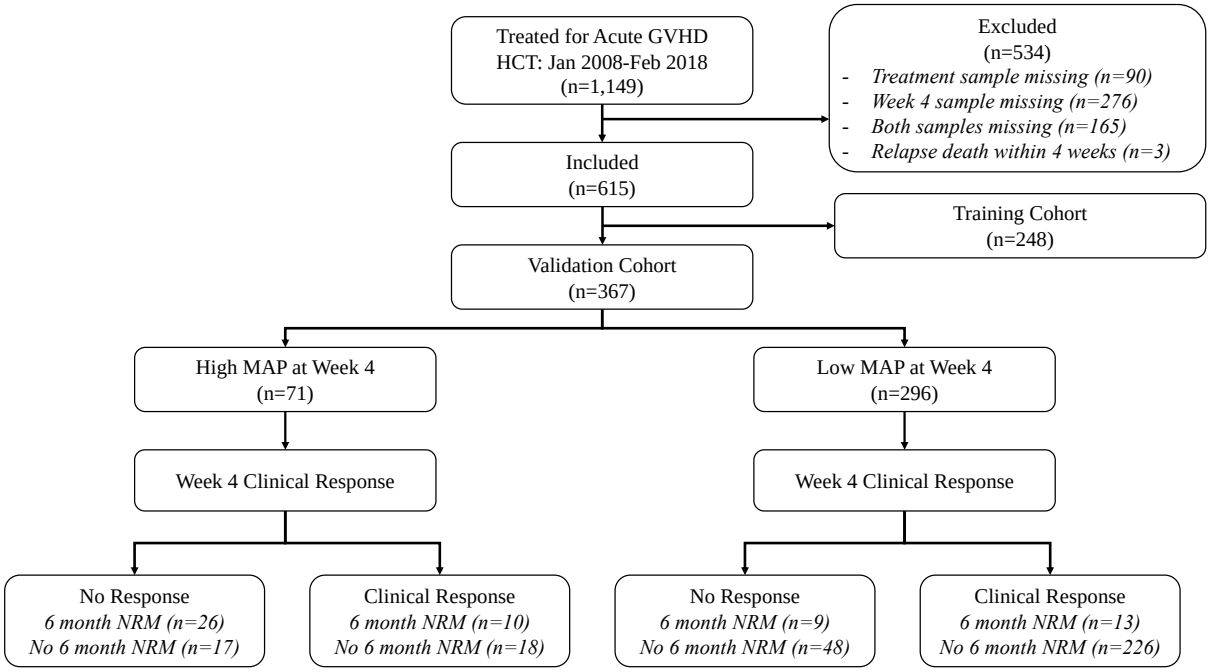

Figure S1. Standards for the Reporting of Diagnostic Accuracy (STARD) flow chart for validation of MAP. Patients contributing clinical data and samples to the MAGIC database and biorepository were enrolled at the time of HCT and monitored for six months for the development of acute GVHD. Patients who were treated systemically for acute GVHD were consecutively enrolled in this study. Patients were excluded if they were missing a treatment sample (n=90), missing a sample after four weeks of treatment despite surviving for four weeks (n=276), missing both samples (n=165), or relapsed and died within four weeks of treatment for GVHD (n=3). Patients who were missing a week four sample due to a non-relapse death were included. Key clinical characteristics of GVHD clinical severity and six month NRM are shown in Figure S2.

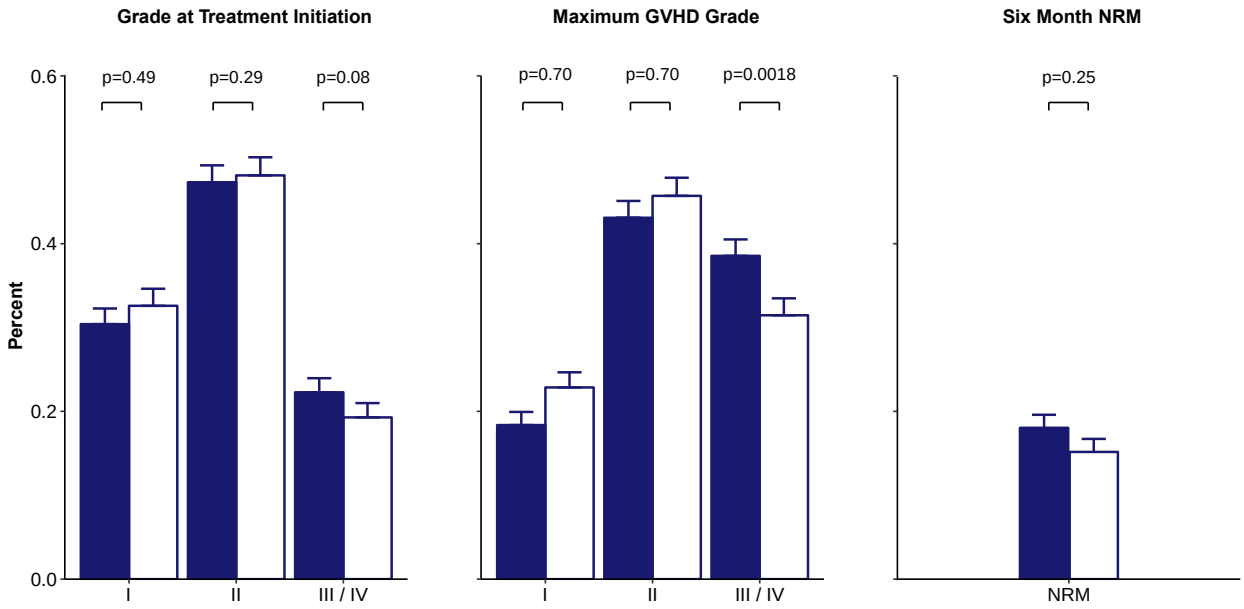

Figure S2. Key GVHD parameters in patients included and excluded from the study. Distributions of Glucksberg grade **(A)** at treatment initiation, **(B)** at maximum during the first month. **(C)** Crude proportion of six month NRM. Patients included in the study (■, n = 615) or excluded (☐, n = 534) because they were missing a sample at treatment initiation, after four weeks of treatment, at both timepoints, or if they relapsed and died within the first month of treatment for GVHD. Error bars represent one standard error of the proportion.

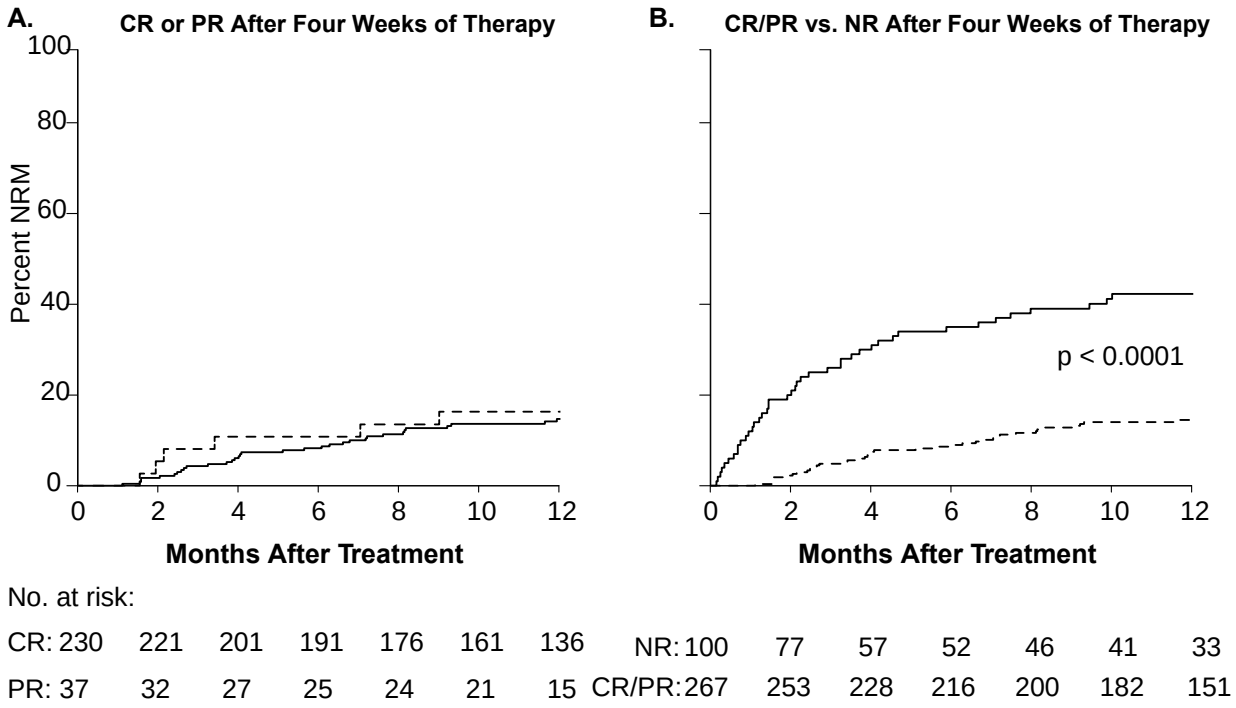

Figure S3. Cumulative incidence of one-year NRM in patients according to response to therapy. (A) 267 patients in the validation cohort experienced a clinical response to systemic therapy with either complete (—, n=230) or partial (- - -, n=37) resolution of GVHD symptoms. There was no difference in the cumulative incidence of NRM between the two groups (15% vs. 16%). **(B)** Cumulative incidence of NRM for patients in the validation cohort (n=367) with clinical responses (- - -, n=267) or with no response (—, n = 100).

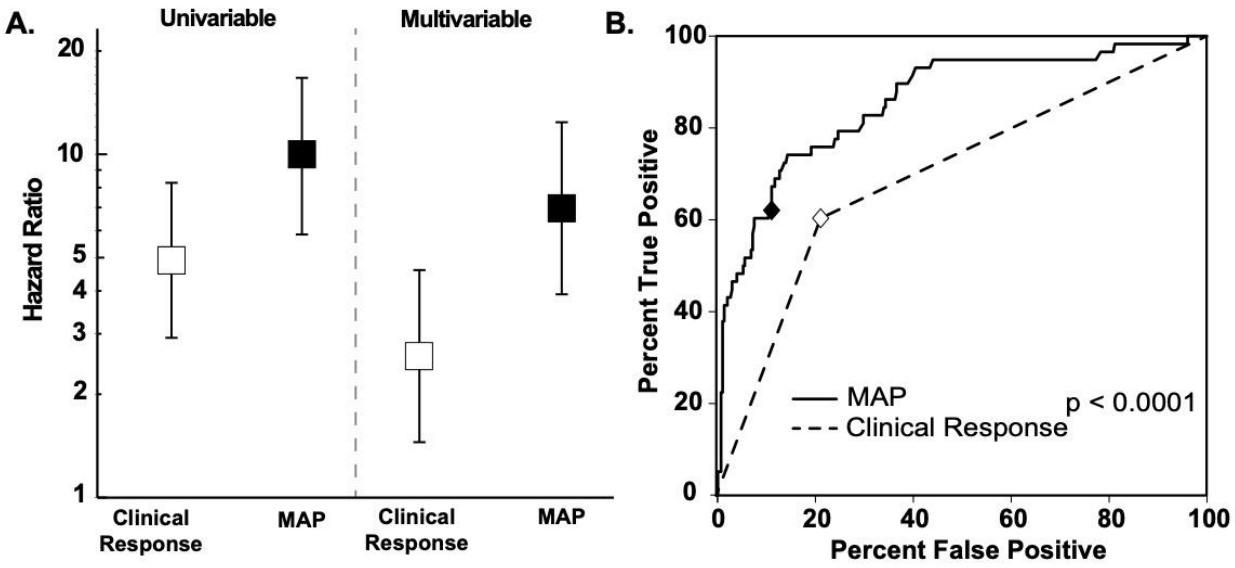

Figure S4. Clinical response and MAP prediction of six month NRM. (A) Hazard ratios (HR) of six month NRM with 95% confidence intervals were determined according to clinical responses (☐) and MAPs (■) in a univariable model (left) and multivariable model (right). **(B)** ROC curves of six month NRM for MAPs (AUC=0.86) and clinical responses (AUC=0.70) after four weeks of systemic therapy ($p<0.0001$). ♦ indicates the post-treatment threshold between high and low MAPs and ◇ indicates clinical response vs. no response.

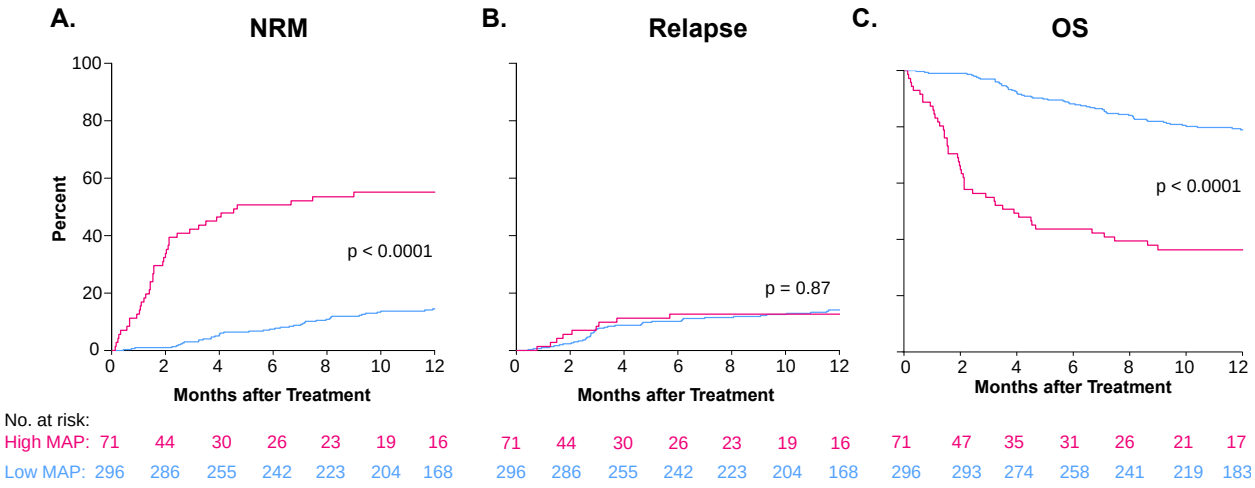

Figure S5. Incidence of NRM, relapse, and OS by MAP. Cumulative incidence of NRM **(A)**, relapse **(B)**, and overall survival (OS) **(C)** in patients of the validation cohort with high (—, n=71) or low (—, n=296) MAPs after four weeks of treatment.

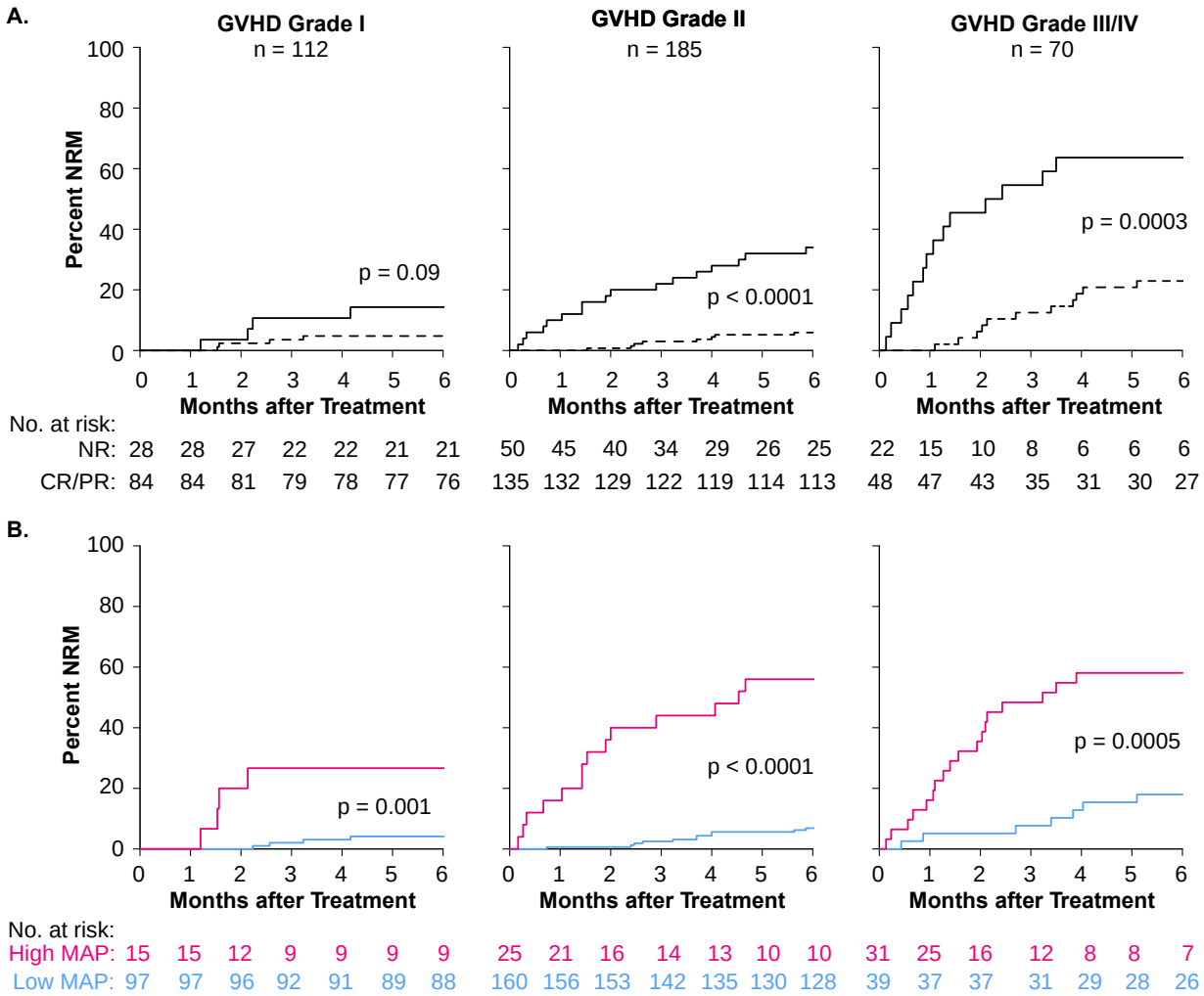

Figure S6. NRM according to Glucksberg grade at the initiation of treatment analyzed by clinical response and by MAP after four weeks of therapy. Cumulative incidences of NRM according to initial grade and analyzed by clinical response **(A)** or MAP **(B)** after four weeks of treatment. **(A)** Left: patients with Glucksberg GVHD grade I with a clinical response (- - -, n=84) or no response (—, n=28). Center: patients with Glucksberg GVHD grade II disease with a clinical response (- - -, n=135) or no response (—, n=50). Right: patients with Glucksberg GVHD grade III or IV disease with a clinical response (- - -, n=48) or no response (—, n=22). **(B)** Left: patients with Glucksberg GVHD grade I whose MAPs after four weeks of GVHD therapy were below (—, n=97) or above (—, n=15) the post-treatment threshold (0.290). Center: patients with Glucksberg GVHD grade II with MAPs below (—, n=160) or above (—, n=25) the threshold. Right: patients with Glucksberg GVHD grade III or IV disease with MAPs were below (—, n=39) or above (—, n=31) the threshold.

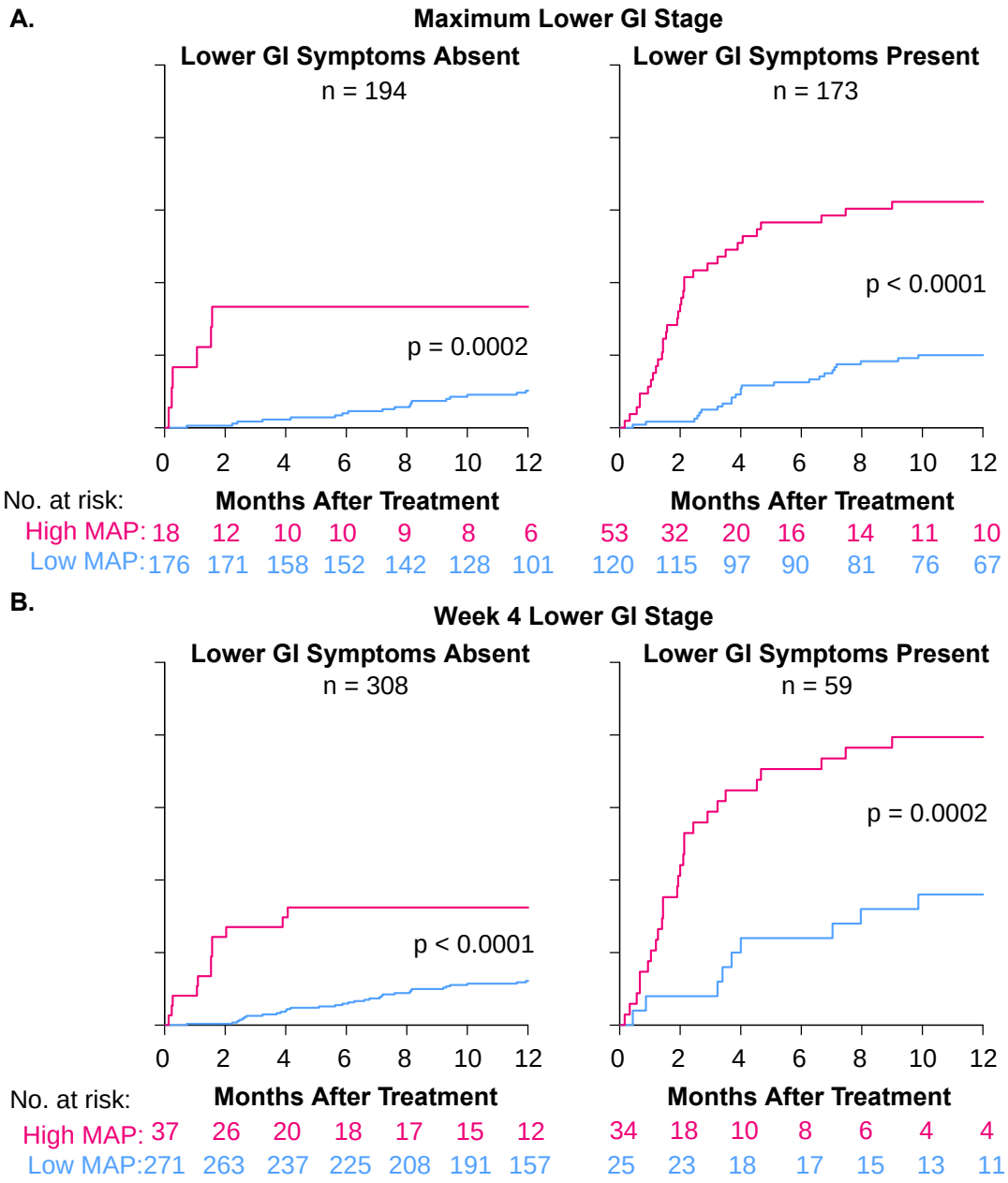

Figure S7. Non-relapse mortality in patients according to lower gastrointestinal symptoms during the first month of therapy analyzed by MAP. Patients of the validation cohort (n=367) were classified based on absence or presence of significant diarrhea (> 500 cc/day) during either any week during of therapy **(A)** or after four weeks of therapy **(B)**. Patients were analyzed according to high (—) and low (—) MAPs after four weeks of treatment. **(A)** Left: Patients with no lower GI symptoms and either high (—, n=18) or low MAPs (—, n=176). Right: Patients with lower GI symptoms during the first month of therapy and either high (—, n=53) or low MAPs (—, n=120). **(B)** Left: Patients without lower GI symptoms at four weeks after treatment and either high (—, n=37) or low MAPs (—, n=271). Right: Patients lower GI symptoms at four weeks after treatment and either high (—, n=34) or low MAPs (—, n=25).

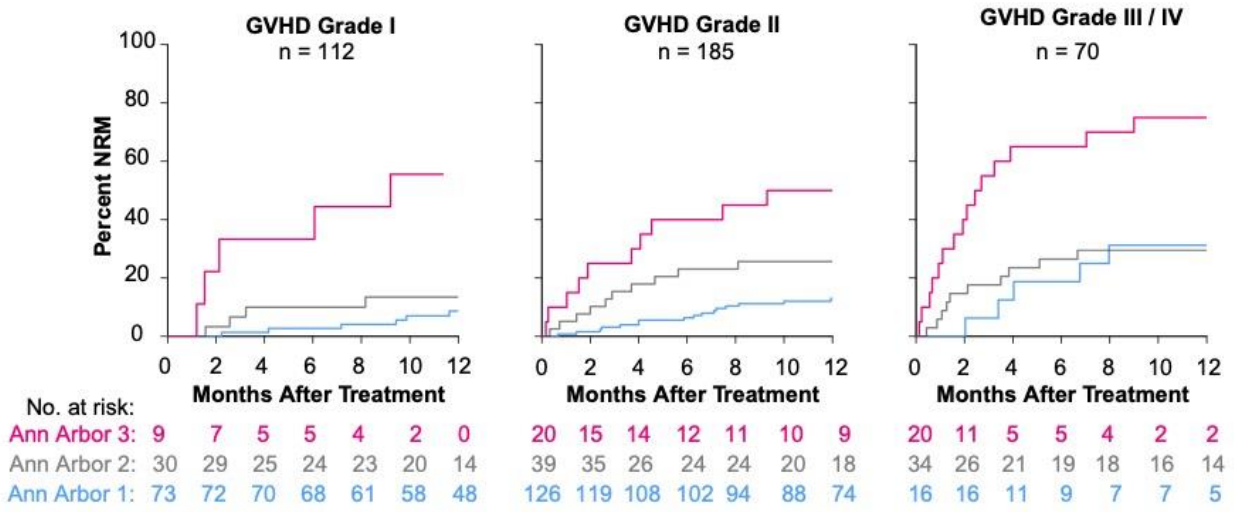

Figure S8. NRM analyzed by Ann Arbor score. Cumulative incidences of twelve month NRM according to Glucksberg grade at treatment initiation. Ann Arbor 1 (—, MAP<0.141) Ann Arbor 2 (—, 0.141 ≤ MAP ≤ 0.290) and Ann Arbor 3 (—, MAP > 0.290) are shown for Glucksberg grade I (left), grade II (center), and grade III / IV (right).

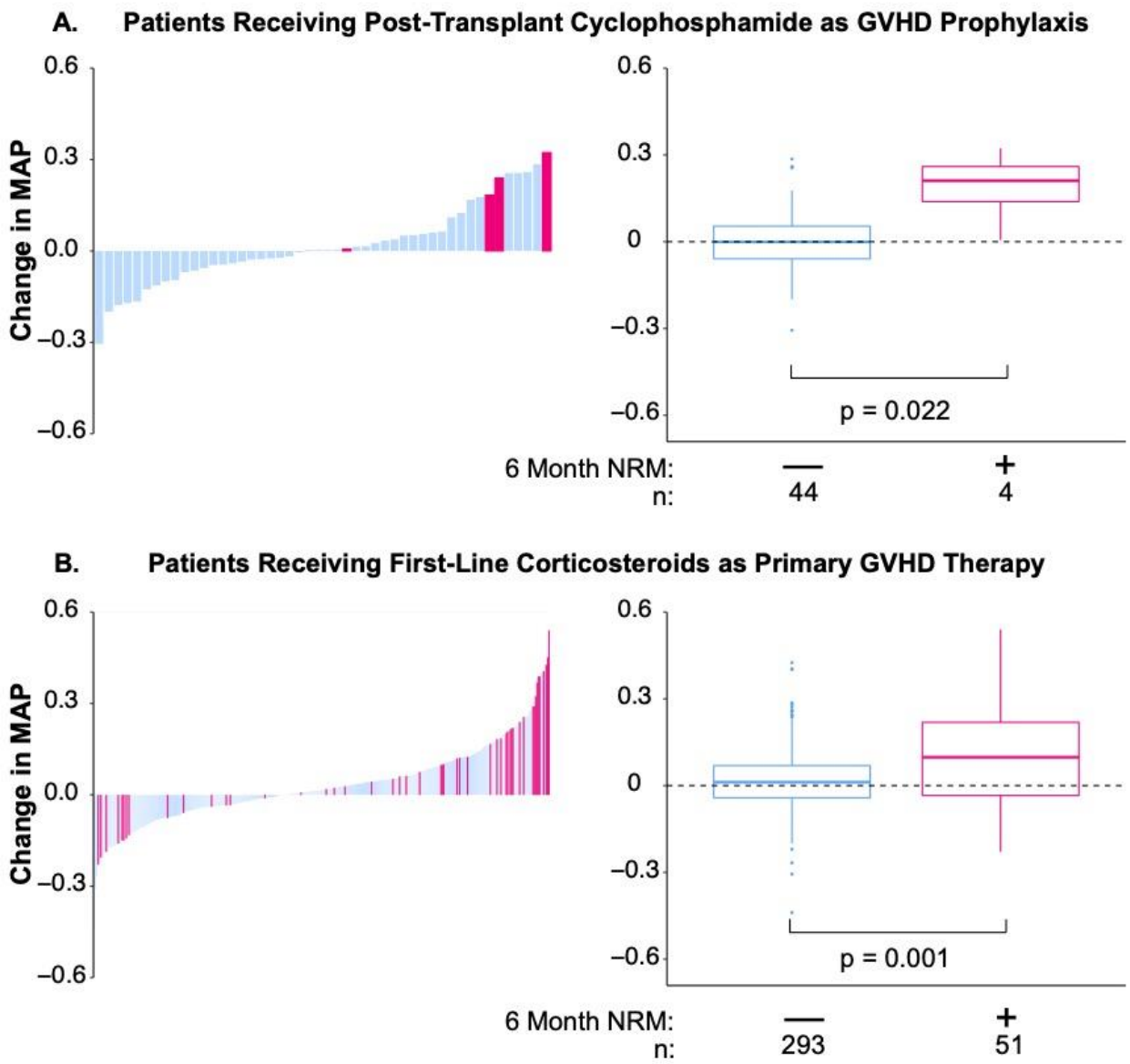

Figure S9. Subset analyses of changes in MAPs. (A) Changes in MAPs for patients receiving post-transplant cyclophosphamide as GVHD prophylaxis who either experienced six month NRM (—) or did not (—) in reverse waterfall plots (left) and box and whisker plots (right). **(B)** Changes in MAPs for patients whose first-line treatment for GVHD consisted of corticosteroids alone who either experienced six month NRM (—) or did not (—) in reverse waterfall plots (left) and box and whisker plots (right).

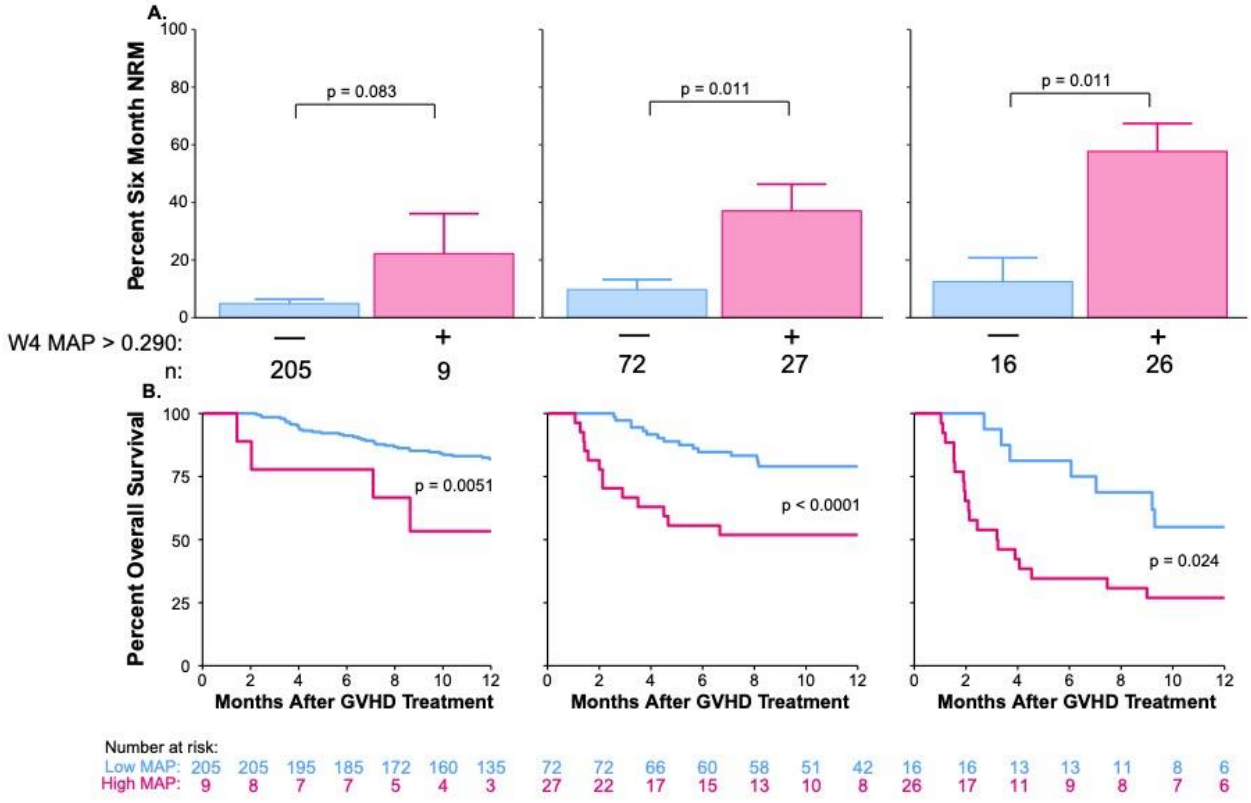

Figure S10. Landmark analysis of long-term mortality by MAP threshold (0.290) in patients surviving until day 28 post-treatment. (A) Crude proportions of six month non-relapse mortality according to Ann Arbor score for patients who survived four weeks after GVHD treatment initiation and whose MAPs after four weeks of treatment rose/remained above (—) or fell/remained below (—) the threshold of 0.290. Ann Arbor scores were determined as in Figure 3. **(B)** Kaplan-Meier estimates of overall survival of the same patients. Error bars represent one standard error of the proportion.

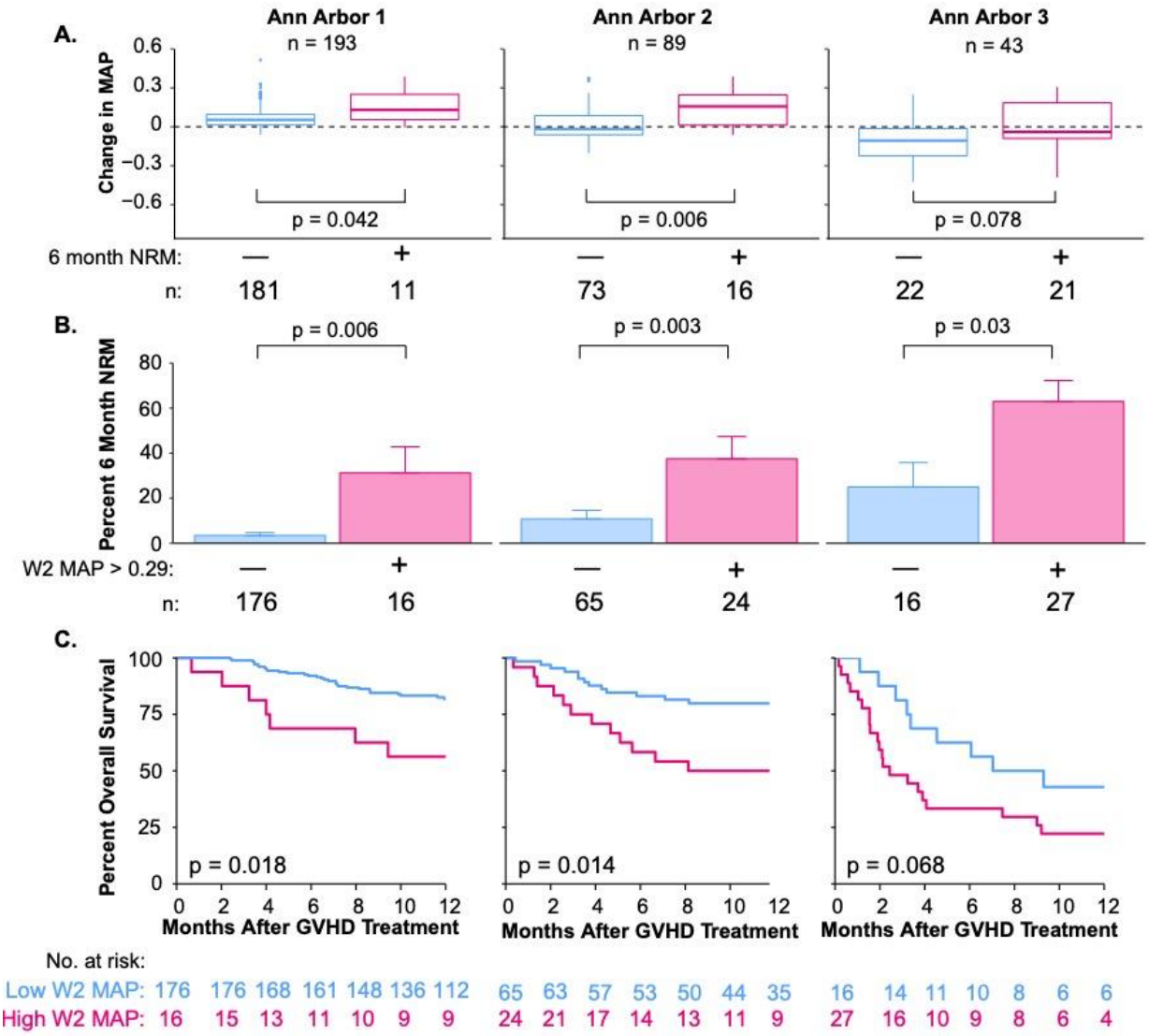

Figure S11. Association between long-term outcomes and changes in MAPs after two weeks. (A) Box and whisker plots of change in MAP in consecutive patients who provided samples before and after two weeks of treatment according to initial Ann Arbor score in patients with (—) and without (—) six month NRM. Left: Ann Arbor 1 patients (MAP < 0.141 at treatment initiation). Center: Ann Arbor 2 patients (0.141 ≤ MAP ≤ 0.290 at treatment initiation). Right: Ann Arbor 3 patients (MAP > 0.290 at treatment initiation). Patients with a biomarker evaluation at baseline only due to death prior to a second measurement (n=3) were excluded for the analysis. - - - indicates no change in MAP. **(B)** Crude proportion of six month NRM (+ standard error of proportion) for each clinical GVHD grade after four weeks of treatment according or low (■) or high (■) MAP. **(C)** Kaplan-Meier estimates of overall survival for patients whose MAPs after two weeks of treatment rose/remained above (—) or dropped/remained below (—) the threshold of 0.290.

Supplemental References

1. Hartwell MJ, Özbek U, Holler E, et al. An early-biomarker algorithm predicts lethal graft-versus-host disease and survival. JCI Insight 2018;2(3):e89798–e89798.