Supplemental Appendix:

Assessment of Systemic and Gastrointestinal Tissue Damage Biomarkers for GVHD Risk Stratification

<u>**Supplemental Tables:**</u>

<u>**Supplemental Figures:**</u>

Table S1: Patient distribution by center

	Training (n=352)	Validation (n=378)
Bambino Gesu Children's Hospital	0	18
Children's Hospital of Los Angeles	0	7
City of Hope Comprehensive Cancer Center	0	11
Columbia University Medical Center	0	5
Emory University	4	18
Hospital for Sick Children	0	1
Icahn School of Medicine at Mount Sinai	17	34
King Chulalongkorn Memorial Hospital	7	2
Mayo Clinic	21	23
Massachusetts General Hospital	0	36
Ohio State University	22	57
University Hospital Carl Gustav Carus Dresden	9	1
University of Erlangen	0	31
University of Hamburg	41	39
University of Michigan	178	4
University of Pennsylvania	11	22
University of Regensburg	32	36
University of Würzburg	10	11
Vanderbilt University	0	22
Total	**352**	**378**

Table S2: AICs for each algorithm in the training cohort.

1 Biomarker	AIC	2 Biomarker	AIC	3 Biomarker	AIC	4 Biomarker	AIC
ST2	807.7	+ REG3	792.2	+ TNFR1	**792.5**	+ TIM3	**793.0**
				+ TIM3	**793.9**		
		+ TNFR1	803.3	+ TIM3	**803.5**		
		+ TIM3	**809.7**				
REG3	799.8	+ TNFR1	795.8	+ TIM3	**796.8**		
		+ TIM3	**801.8**				
TNFR1	809.7	+ TIM3	810.3				
TIM3	829.3						

The first column contains algorithms of single biomarkers. Additional biomarkers are added sequentially in each cell to the right.

Table S3: TNFR1+TIM3 algorithm and threshold performance.

Algorithm*: $\log[-\log(1-\hat{p})] = -9.266 + 2.470 \log_{10} \text{TNFR1} - 0.375 \log_{10} \text{TIM3}$

Threshold	Sensitivity	Specificity	Product
0.19	0.65	0.60	0.390
0.2	0.63	0.65	0.410
0.21	**0.60**	**0.70**	**0.420**
0.22	0.56	0.72	0.403
0.23	0.53	0.75	0.398

Table S4: TNFR1+ST2 algorithm and threshold performance.

Algorithm*: $\log[-\log(1-\hat{p})] = -11.287 + 1.062 \log_{10} \text{ST2} + 1.329 \log_{10} \text{TNFR1}$

Threshold	Sensitivity	Specificity	Product
0.2	0.74	0.64	0.474
0.21	0.72	0.66	0.475
0.22	**0.68**	**0.70**	**0.476**
0.23	0.65	0.72	0.468
0.24	0.63	0.75	0.473

Table S5: TNFR1+REG3α algorithm and threshold performance.

Algorithm*: $\log[-\log(1-\hat{p})] = -7.543 + 1.218 \log_{10} \text{TNFR1} + 0.884 \log_{10} \text{REG3}\alpha$

Threshold	Sensitivity	Specificity	Product
0.14	0.88	0.45	0.396
0.15	0.82	0.52	0.426
0.16	**0.81**	**0.57**	**0.462**
0.17	0.74	0.60	0.444
0.18	0.67	0.64	0.429

Table S6: ST2+REG3α algorithm and threshold performance.

Algorithm*: $\log[-\log(1-\hat{p})] = -7.823 + 1.027 \log_{10} \text{ST2} + 0.875 \log_{10} \text{REG3}\alpha$

Threshold	Sensitivity	Specificity	Product
0.19	0.68	0.66	0.449
0.2	0.68	0.68	0.462
0.21	**0.65**	**0.72**	**0.468**
0.22	0.63	0.72	0.454
0.23	0.61	0.74	0.451

* $\hat{p}$ = estimated predicted probability of six month NRM

Table S7: AUC of the ROC curves for individual biomarker algorithms in the validation cohort.

Biomarker	AUC
TNFR1	0.58
TIM3	0.56
IL6	0.55
ST2	0.73
REG3α	0.75

Table S8: Performance of two biomarker algorithms using the thresholds that provide sensitivity closest to 0.8

Algorithm	Threshold	Sensitivity	Specificity	PPV	NPV
TNFR1+TIM3	0.13	**0.82**	0.24	0.15	0.89
TNFR1+ST2	0.12	**0.82**	0.34	0.17	0.92
TNFR1+REG3α	0.17	**0.80**	0.53	0.22	0.94
ST2+REG3α	0.15	**0.84**	0.59	0.26	0.96

Table S9: Performance of two biomarker algorithms using the thresholds that provide specificity closest to 0.8

Algorithm	Threshold	Sensitivity	Specificity	PPV	NPV
TNFR1+TIM3	0.26	0.33	**0.81**	0.23	0.88
TNFR1+ST2	0.22	0.55	**0.80**	0.31	0.91
TNFR1+REG3α	0.28	0.47	**0.80**	0.29	0.9
ST2+REG3α	0.24	0.67	**0.82**	0.39	0.94

Table S10: ST2+REG3α algorithm performance characteristics in patients presenting with LGI ± other or skin only GVHD

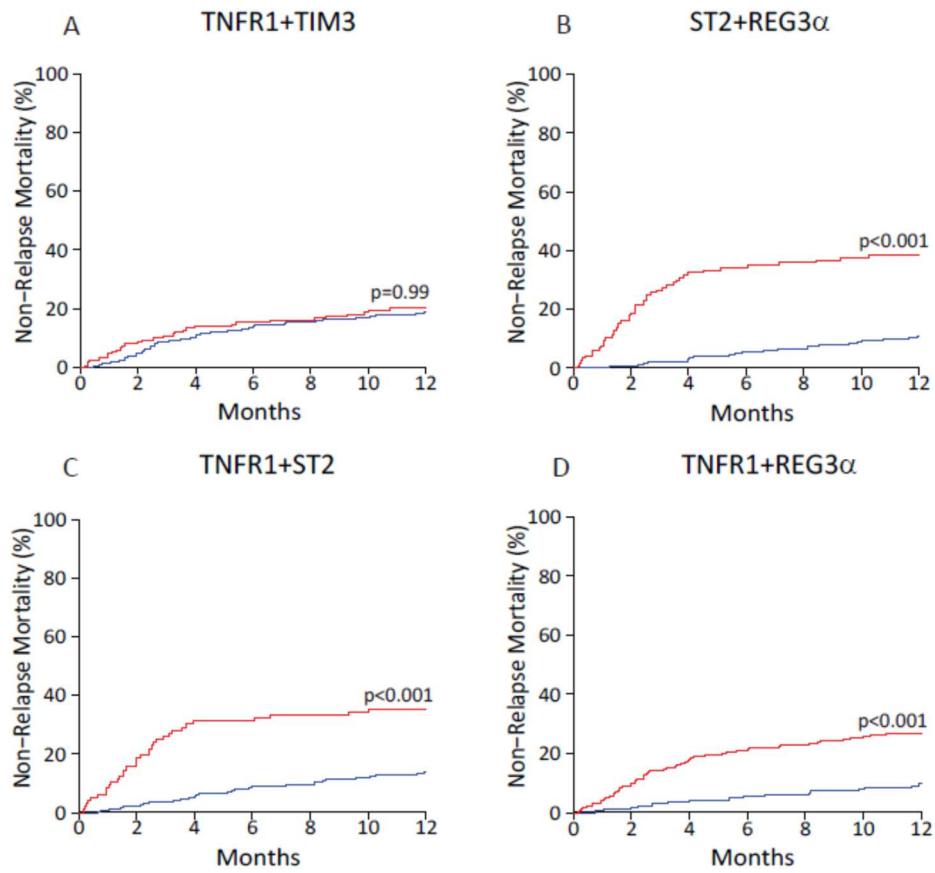|	Sensitivity	Specificity	PPV	NPV
LGI ± other	0.93	0.61	0.48	0.96
Skin only	0.47	0.81	0.17	0.95

Figure S1: 12 month NRM for risk groups defined by two biomarker algorithms. The cumulative incidence of 12 month NRM is shown for high (red) and low (blue) risk patients (A) TNFR1+TIM3 (20% vs 19%, p=0.99); (B) ST2+REG3α (38% vs 11%, p<0.001); (C) TNFR1+ST2 (35% vs 14%, P<0.001); and (D) TNFR1+REG3α (27% vs 10%, p<0.001).
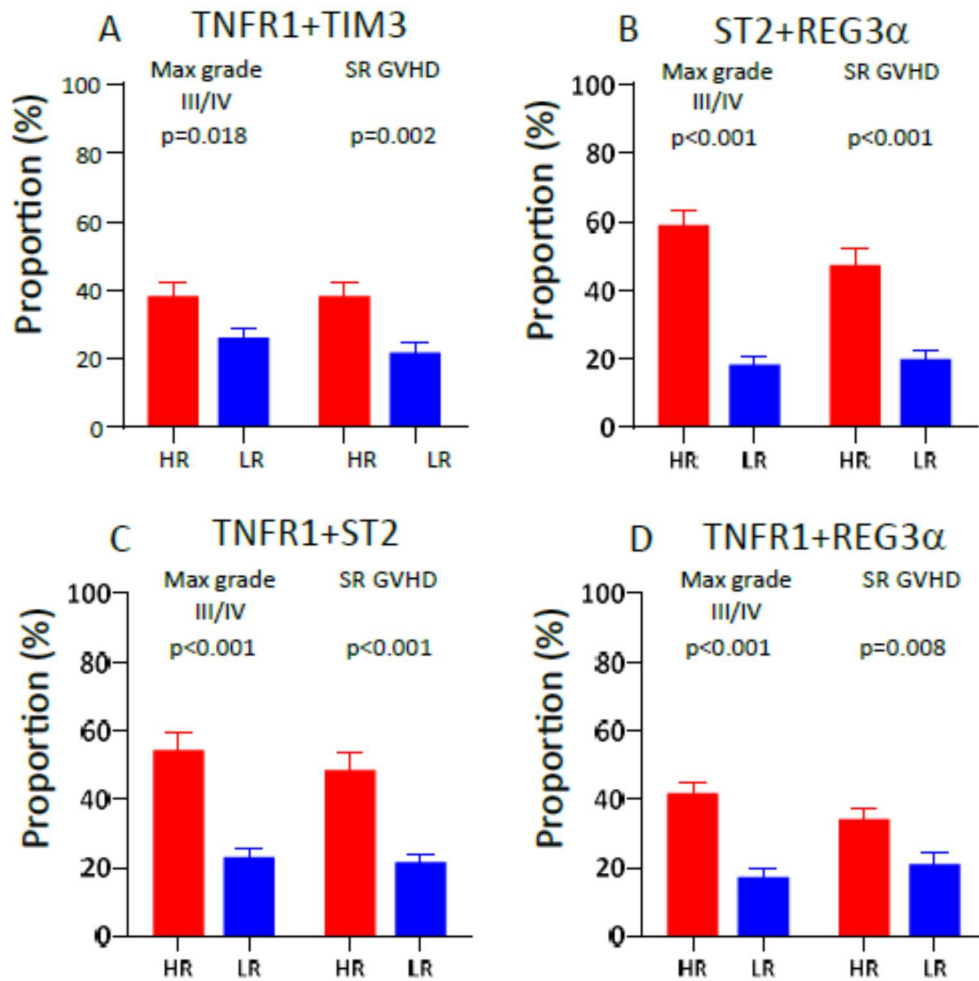
Figure S2: GVHD outcomes in risk groups defined by two biomarker algorithms. The proportion of patients with maximum grade III/IV GVHD for high risk patients (red) compared to low risk patients (blue) for each algorithm: (A) TNFR1+TIM3 (38% vs 26%, p=0.018), (B) ST2+REG3α (59% vs 18%, p<0.001), (C) TNFR1+ST2 (54% vs 23%. P <0.001), and (D) TNFR1+REG3α (42% vs 17%, p <0.001). The proportion of patients who developed steroid refractory GVHD for high risk patients compared to low risk patients for each algorithm was: (A) TNFR1+TIM3 (38% vs 22%, p=0.002), (B) ST2+REG3α (47% vs 20%, p<0.001), (C) TNFR1+ST2 (48% vs 21%. p <0.001), and (D) TNFR1+REG3α (34% vs 21%, p=0.008).
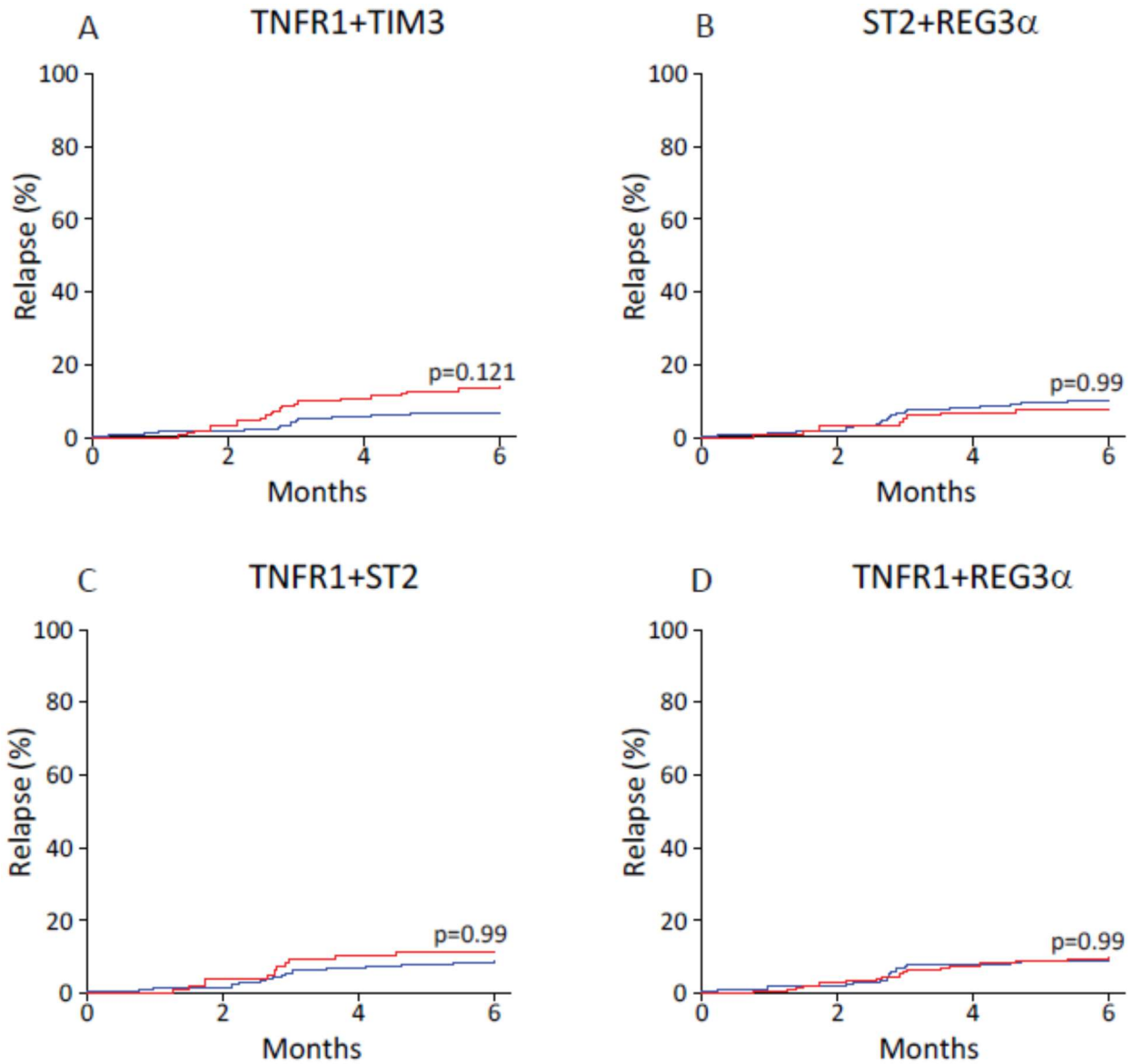
Figure S3: Cumulative incidence of relapse in risk groups defined by two biomarker algorithms. The cumulative incidence of relapse is shown for high (red) and low (blue) risk patients: (A) TNFR1+TIM3 (14% vs 7%, p=0.121); (B) ST2+REG3α (8% vs 10%, p=0.99); (C) TNFR1+ST2 (11% vs 9%. P=0.99); and (D) TNFR1+REG3α (10% vs 9%, p=0.99).
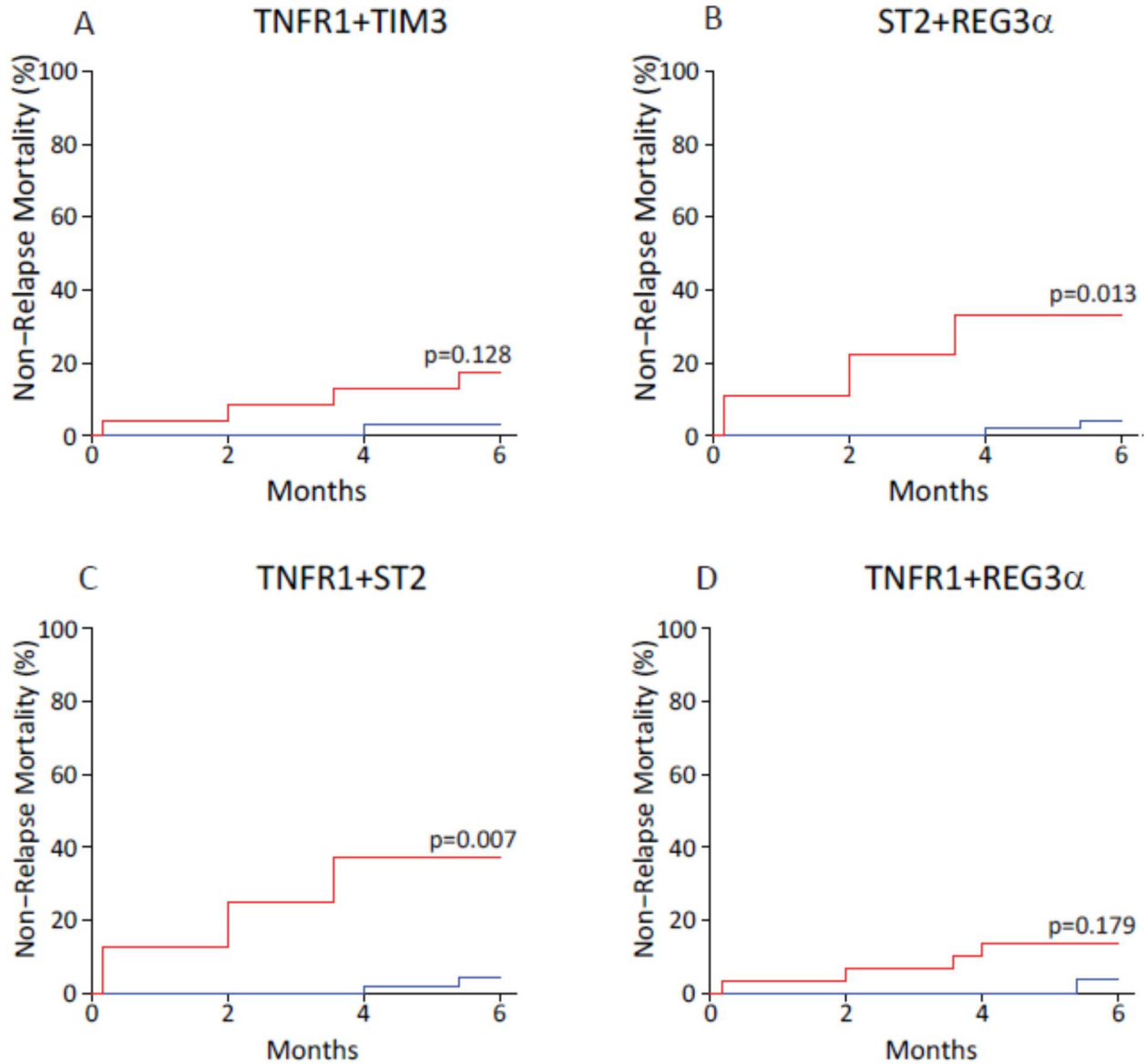
Figure S4: 6 month NRM for risk groups defined by two biomarker algorithms in patients who received post-transplant cyclophosphamide as GVHD prophylaxis. The cumulative incidence of 6 month NRM is shown for high (red) and low (blue) risk patients (A) TNFR1+TIM3 (17% vs 3%, p=0.128); (B) ST2+REG3α (33% vs 4%, p=0.013); (C) TNFR1+ST2 (38% vs 4%. P=0.007); and (D) TNFR1+REG3α (14% vs 4%, p=0.179)