Table of Contents

Supplemental Table 1: Participating Centers

	Training (n=341)	Validation (n=374)
Bambino Gesu Children's Hospital (Rome, Italy)	0	18
Children's Hospital Los Angeles	0	7
City of Hope Comprehensive Cancer Center (Duarte, CA)	0	11
Columbia University Medical Center (New York, NY)	0	4
Emory University (Atlanta, GA)	4	18
Hospital for Sick Children (Toronto, ON)	0	1
Icahn School of Medicine at Mount Sinai (New York, NY)	17	34
King Chulalongkorn Memorial Hospital (Bangkok, Thailand)	7	2
Mayo Clinic (Rochester, MN)	21	23
Massachusetts General Hospital (Boston, MA)	0	36
Ohio State University	21	57
University Hospital Carl Gustav Carus Dresden	9	1
University of Erlangen	0	31
University of Hamburg	41	38
University of Michigan	168	4
University of Pennsylvania	11	20
University of Regensburg	32	36
University of Wurzburg	10	11
Vanderbilt University (Nashville, TN)	0	22

Supplemental Table 2: Patient Characteristics

Characteristic	Training (n = 341) 2004-2015	Validation (n = 374) 2015-2017	P-value
Median Age - yr (range)	53 (0-75)	53 (0-74)	0.65
Age (years) – n (%)			0.002
<18	36 (11%)	56 (15%)	
18-60 yrs	229 (67%)	203 (54%)	
>60	76 (22%)	115 (31%)	
Indication for HCT - n (%)			0.001
Acute leukemia	184 (54%)	196 (52%)	
MDS/MPS	69 (20%)	109 (29%)	
Lymphoma	44 (13%)	31 (8%)	
Non-malignant	11 (3%)	21 (6%)	
Other malignant	33 (10%)	17 (5%)	
Conditioning Regimen Intensity - n (%)			0.006
Myeloablative	236 (69%)	221 (59%)	
Reduced Intensity/non-myeloablative	105 (31%)	153 (41%)	
Donor type - n (%)			
Related	96 (28%)	94 (25%)	<0.001
Unrelated	237 (70%)	243 (65%)	
Haploidentical	8 (2%)	37 (10%)	
HLA-match - n (%)			<0.001
Matched	250 (73%)	267 (71%)	
Mismatched	83 (24%)	70 (19%)	
Haploidentical	8 (2%)	37 (10%)	
GVHD serotherapy - n (%)			<0.001
ATG	100 (29%)	159 (42%)	
No ATG	241 (71%)	215 (58%)	
GVHD prophylaxis - n (%)			<0.001
CNI/MTX ± other	199 (58%)	204 (54%)	
CNI/MMF ± other	121 (36%)	86 (23%)	
Tac + Sirolimus	1 (<1%)	10 (3%)	
Cyclophosphamide based	17 (5%)	56 (15%)	
T cell depletion	1 (<1%)	14 (4%)	
Other	2 (1%)	4 (1%)	
Stem cell source - n (%)			0.22
Marrow	48 (14%)	67 (18%)	
Peripheral blood	262 (77%)	282 (75%)	
Cord	31 (9%)	25 (7%)	
Diagnosis GVHD: median day (range)	26 (7-273)	28 (5-196)	<0.001
Late-onset GVHD, n (%)	9 (2.6%)	35 (9.3%)	<0.001

Diagnosis GVHD Grade - n (%)				0.98
	Grade I*	146 (43%)	159 (42%)	
	Grade II	134 (39%)	147 (39%)	
	Grade III	51 (15%)	55 (15%)	
	Grade IV	10 (3%)	13 (4%)	
Minnesota Risk at Dx - n (%)				0.682
	Standard	288 (84%)	321 (86%)	
	High	53 (16%)	53 (14)	
Target Organ Involvement at Dx - n (%)				0.34
	LGI+/-other	109 (32%)	109 (29%)	
	Liver only	2 (<1%)	4 (1%)	
	Skin only	191 (56%)	199 (53%)	
	UGI only	23 (7%)	36 (10%)	
	Other	16 (5%)	26 (7%)	
Maximum GVHD Grade III-IV, n (%)		129 (38%)	116 (31%)	0.07
Systemic treatment- n (%)				0.28
	Yes	302 (89%)	321 (88%)**	
	No	39 (11%)	53 (14%)	
12-month NRM (%)		24%	19%	0.08

ATG, anti-thymocyte globulin; CNI, calcineurin inhibitor; Dx, diagnosis; LGI, lower GI; MDS, myelodysplastic syndrome; MMF, mycophenolate mofetil; MPS, myeloproliferative syndrome; MTX, methotrexate; Tac, tacrolimus; UGI, upper GI.

*Includes 1 patient with biopsy-proven liver GVHD and total bilirubin <2 mg/dl.

** Includes 2 patients treated with sirolimus monotherapy

Supplemental Table 3: Additional Post-Transplant Cyclophosphamide Prophylaxis Patient Characteristics (validation cohort)

Characteristic	(n =77) 2020-2022
Median Age - yr (range)	59 (6-74)
Age (years) – n (%)	
<18	4 (5%)
18-60 yrs	38 (49%)
>60	35 (46%)
Indication for HCT - n (%)	
Acute leukemia	35 (46%)
MDS/MPS	21 (27%)
Lymphoma	7 (9%)
Non-malignant	6 (8%)
Other malignant	8 (10%)
Conditioning Regimen Intensity - n (%)	
Myeloablative	49 (64%)
Reduced Intensity/non-myeloablative	28 (36%)
Donor type - n (%)	
Related	46 (60%)
Unrelated	31 (40%)
Haploidentical	
HLA-match - n (%)	
Matched	17 (22%)
Mismatched	15 (20%)
Haploidentical	45 (58%)
GVHD serotherapy - n (%)	
ATG	8 (10%)
No ATG	69 (90%)
GVHD prophylaxis - n (%)	
Cyclophosphamide based	77 (100%)
Stem cell source - n (%)	
Marrow	17 (22%)
Peripheral blood	60 (78%)
Diagnosis GVHD: median day (range)	37 (5-105)
Late-onset GVHD, n (%)	1 (1%)
Diagnosis GVHD Grade - n (%)	
Grade I	30 (39%)
Grade II	32 (41%)
Grade III	12 (16%)
Grade IV	3 (4%)

Minnesota Risk at Dx - n (%)		
	Standard	65 (84%)
	High	12 (16%)
Target Organ Involvement at Dx - n (%)		
	LGI+/-other	29 (38%)
	Liver only	2 (3%)
	Skin only	43 (56%)
	UGI only	2 (3%)
	Other	1 (1%)
Maximum GVHD Grade III-IV, n (%)		24 (31%)
Systemic treatment- n (%)		
	Yes	77 (100%)
12-month NRM (%)		18%

Supplemental Table 4: Biomarker Algorithms

	Algorithm		P value*
ST2	$\log[-\log(1-\hat{p})] = -7.836 + 1.4127 \log_{10} ST2$	ST2	<0.001
REG3α	$\log[-\log(1-\hat{p})] = -3.0284 + 0.9749 \log_{10} REG3\alpha$	REG3α	<0.001
AREG	$\log[-\log(1-\hat{p})] = -2.7919 + 1.1743 \log_{10} AREG$	AREG	<0.001
ST2+AREG	$\log[-\log(1-\hat{p})] = -5.7536 + 0.7001 \log_{10} ST2 + 0.9543 \log_{10} AREG$	ST2 AREG	0.021 <0.001
ST2+REG3α	$\log[-\log(1-\hat{p})] = -7.3272 + 1.0205 \log_{10} ST2 + 0.7312 \log_{10} REG3\alpha$	ST2 REG3α	<0.001 <0.001
REG3α+AREG	$\log[-\log(1-\hat{p})] = -3.5277 + 0.5731 \log_{10} REG3\alpha + 0.947 \log_{10} AREG$	REG3α AREG	0.002 <0.001
ST2+ REG3α+ AREG	$\log[-\log(1-\hat{p})] = -5.6647 + 0.53\log_{10} ST2 + 0.4957 \log_{10} REG3\alpha + 0.8075 \log_{10} AREG$	ST2 REG3α AREG	0.088 0.009 <0.001

*** p-values represent the statistical significance for each biomarker in the model in the training cohort**

Supplemental Table 5: Comparison of AUCs between algorithms
Key algorithms are **bolded** (validation cohort, n=374)

Reference AUC	Comparator AUC	Adjusted P value
AREG (0.707)	**ST2+REG3α** (0.757)	0.23
	ST2+REG3α+AREG (0.752)	0.15
	ST2 (0.710)	0.97
	REG3α (0.711)	0.97
	ST2+AREG (0.734)	0.23
	REG3α+AREG (0.736)	0.23
ST2+REG3α (0.757)	**ST2+REG3α+AREG** (0.752)	0.94
	ST2 (0.710)	0.23
	REG3α (0.711)	0.23
	ST2+AREG (0.734)	0.50
	REG3α+AREG (0.736)	0.50
ST2+REG3α+AREG (0.752)	ST2 (0.710)	0.32
	REG3α (0.711)	0.23
	ST2+AREG (0.734)	0.32
	REG3α+AREG (0.736)	0.23
ST2 (0.710)	REG3α (0.711)	0.97
	ST2+AREG (0.734)	0.50
	REG3α+AREG (0.736)	0.64
REG3α (0.711)	ST2+AREG (0.734)	0.64
	REG3α+AREG (0.736)	0.50
ST2+AREG (0.734)	REG3α+AREG (0.736)	0.97

The false discovery rate (FDR) approach was used for multiplicity adjustment

Supplemental Table 6: Comparison of AUCs between algorithms, systemically treated subset of validation cohort

Key algorithms are **bolded (n=321)**

Reference AUC	Comparator AUC	Adjusted P value
AREG (0.693)	**ST2+REG3α** (0.739)	0.39
	ST2+REG3α+AREG (0.735)	0.39
	ST2 (0.694)	0.99
	REG3α (0.698)	0.96
	ST2+AREG (0.718)	0.39
	REG3α+AREG (0.721)	0.39
ST2+REG3α (0.739)	**ST2+REG3α+AREG** (0.735)	0.96
	ST2 (0.694)	0.39
	REG3α (0.698)	0.39
	ST2+AREG (0.718)	0.59
	REG3α+AREG (0.721)	0.65
ST2+REG3α+AREG (0.735)	ST2 (0.694)	0.39
	REG3α (0.698)	0.39
	ST2+AREG (0.718)	0.39
	REG3α+AREG (0.721)	0.39
ST2 (0.694)	REG3α (0.698)	0.96
	ST2+AREG (0.718)	0.59
	REG3α+AREG (0.721)	0.66
REG3α (0.698)	ST2+AREG (0.718)	0.76
	REG3α+AREG (0.721)	0.59
ST2+AREG (0.718)	REG3α+AREG (0.721)	0.96

The false discovery rate (FDR) approach was used for multiplicity adjustment

Supplemental Table 7: Performance characteristics of all algorithms stratified for risk by the concordance probability threshold

		Threshold	% High Risk	Sensitivity	Specificity	PPV	NPV	Balanced accuracy	Correctly classified
Validation cohort (n=374)	**AREG**	0.231	45%	0.69	0.61	0.29	0.89	0.65	62%
	ST2+REG3α	0.247	30%	0.63	0.77	0.39	0.90	0.70	75%
	ST2+REG3α+AREG	0.204	48%	0.77	0.59	0.30	0.92	0.68	60%
	ST2	0.225	30%	0.59	0.77	0.37	0.89	0.68	73%
	REG3α	0.223	52%	0.77	0.54	0.28	0.91	0.66	58%
	ST2+AREG	0.191	53%	0.76	0.52	0.27	0.90	0.64	56%
	REG3α+AREG	0.235	46%	0.74	0.61	0.30	0.91	0.68	63%
		Threshold	**% High Risk**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Balanced accuracy**	**Correctly classified**
Systemically treated subset (n=321)	**AREG**	0.231	47%	0.68	0.59	0.31	0.87	0.63	61%
	ST2+REG3α	0.247	34%	0.62	0.74	0.39	0.88	0.68	71%
	ST2+REG3α+AREG	0.204	52%	0.77	0.55	0.32	0.90	0.66	60%
	ST2	0.225	33%	0.59	0.74	0.39	0.87	0.67	71%
	REG3α	0.223	55%	0.77	0.50	0.30	0.89	0.64	56%
	ST2+AREG	0.191	56%	0.75	0.49	0.29	0.88	0.62	55%
	REG3α+AREG	0.235	49%	0.74	0.58	0.32	0.89	0.66	61%

Supplemental Table 8: Cumulative incidence of NRM and relapse and overall survival using concordance probability threshold

		CI 6 Month NRM	P-value*	CI 12 Month NRM	P-value*	CI 12 Month Relapse	P-value*	12 Month OS	P-value**
Validation cohort (n=374)	**AREG**	6% vs 24%	<0.001	11% vs 29%	<0.001	16% vs 12%	0.334	81% vs 64%	<0.001
	ST2+REG3α	5% vs 35%	<0.001	10% vs 39%	<0.001	15% vs 11%	0.338	82% vs 52%	<0.001
	ST2+REG3α+AREG	4% vs 25%	<0.001	8% vs 30%	<0.001	14% vs 13%	0.287	84% vs 61%	<0.001
	ST2	7% vs 32%	<0.001	11% vs 37%	<0.001	14% vs 14%	0.424	81% vs 54%	<0.001
	REG3α	9% vs 23%	<0.001	9% vs 28%	<0.001	17% vs 11%	0.274	83% vs 65%	<0.001
	ST2+AREG	5% vs 22%	<0.001	10% vs 27%	<0.001	15% vs 13%	0.191	83% vs 65%	<0.001
	REG3α+AREG	5% vs 25%	<0.001	9% vs 30%	<0.001	15% vs 13%	0.292	83% vs 62%	<0.001
		CI 6 Month NRM	**P-value***	**CI 12 Month NRM**	**P-value***	**CI 12 Month Relapse**	**P-value***	**12 Month OS**	**P-value****
Systemically treated subset (n=321)	**AREG**	8% vs 27%	<0.001	13% vs 31%	<0.001	15% vs 13%	0.721	78% vs 61%	<0.001
	ST2+REG3α	6% vs 37%	<0.001	12% vs 39%	<0.001	16% vs 11%	0.284	80% vs 51%	<0.001
	ST2+REG3α+AREG	5% vs 27%	<0.001	10% vs 32%	<0.001	14% vs 14%	0.874	81% vs 60%	<0.001
	ST2	8% vs 33%	<0.001	13% vs 39%	<0.001	14% vs 14%	0.933	79% vs 52%	<0.001
	REG3α	6% vs 25%	<0.001	11% vs 30%	<0.001	17% vs 12%	0.187	80% vs 62%	<0.001
	ST2+AREG	6% vs 24%	<0.001	12% vs 29%	<0.001	14% vs 14%	0.798	81% vs 62%	<0.001
	REG3α+AREG	6% vs 27%	<0.001	11% vs 33%	<0.001	15% vs 13%	0.710	80% vs 60%	<0.001

*Gray's test for comparing cumulative incidence function (CIF)

** Log-rank test

Supplemental Table 9: Day 28 overall response rate using concordance probability threshold, systemically treated subset of validation cohort
(n=321)

	Day 28 Overall Response Rate (%) LR vs HR	**Difference Between Groups**	**P-value**
AREG	76% vs 61%	15%	0.009
ST2+REG3α	78% vs 52%	26%	<0.001
ST2+REG3α+AREG	78% vs 61%	17%	0.001
ST2	75% vs 56%	19%	<0.001
REG3α	73% vs 66%	7%	0.202
ST2+AREG	76% vs 63%	13%	0.017
REG3α+AREG	76% vs 62%	14%	0.014

Supplemental Table 10: Performance characteristics using threshold corresponding to 80% specificity

		Threshold	% High Risk	Sensitivity	Specificity	PPV	NPV	Balanced Accuracy	Correctly classified
Validation cohort (n=374)	**AREG**	0.262	37%	0.66	0.70	0.34	0.90	0.68	69%
	ST2+REG3α	0.305	18%	0.49	0.88	0.49	0.88	0.69	81%
	ST2+REG3α+AREG	0.283	30%	0.60	0.76	0.37	0.89	0.68	74%
	ST2	0.333	14%	0.39	0.92	0.52	0.87	0.66	82%
	REG3α	0.295	32%	0.54	0.73	0.32	0.87	0.64	69%
	ST2+AREG	0.272	29%	0.54	0.77	0.36	0.88	0.66	73%
	REG3α+AREG	0.276	36%	0.63	0.71	0.33	0.89	0.67	69%
		Threshold	% High Risk	Sensitivity	Specificity	PPV	NPV	Balanced Accuracy	Correctly classified
Systemically treated subset (n=321)	**AREG**	0.262	39%	0.65	0.68	0.36	0.88	0.67	67%
	ST2+REG3α	0.305	21%	0.49	0.87	0.51	0.86	0.68	79%
	ST2+REG3α+AREG	0.283	34%	0.59	0.73	0.38	0.87	0.66	70%
	ST2	0.333	16%	0.39	0.90	0.53	0.84	0.65	79%
	REG3α	0.295	33%	0.54	0.72	0.35	0.85	0.63	68%
	ST2+AREG	0.272	31%	0.54	0.75	0.37	0.86	0.64	70%
	REG3α+AREG	0.276	38%	0.62	0.68	0.35	0.87	0.65	67%

Supplemental Table 11: Cumulative incidence of NRM and relapse and overall survival using threshold corresponding to 80% specificity

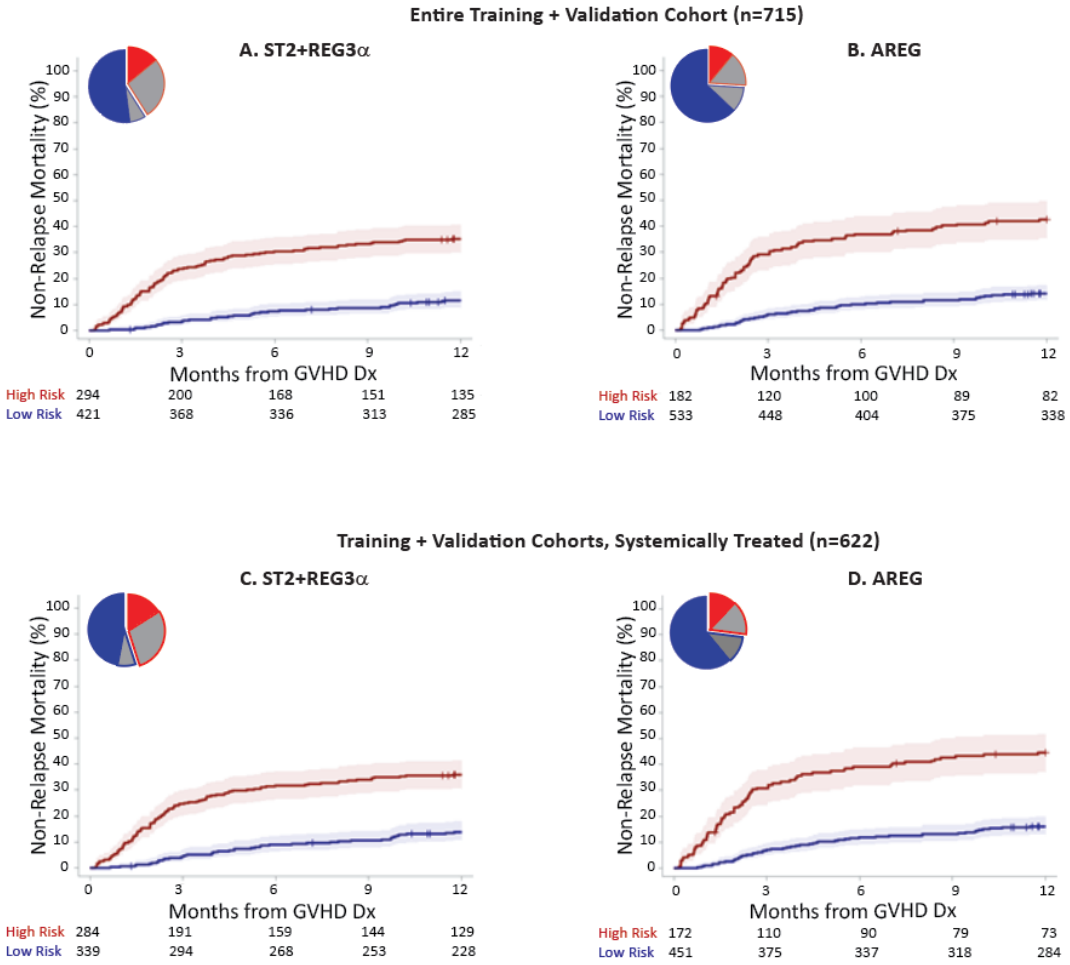
		CI 12 month NRM	P-value	CI 12 Month Relapse	P-value	12 Month OS	P-value
Validation cohort (n=374)	**AREG**	10% vs 34%	<0.001	15% vs 12%	0.555	81% vs 59%	<0.001
	ST2+REG3α	12% vs 49%	<0.001	15% vs 9%	0.183	80% vs 42%	<0.001
	ST2+REG3α+AREG	11% vs 37%	<0.001	15% vs 11%	0.389	81% vs 55%	<0.001
	ST2	13% vs 52%	<0.001	15% vs 8%	0.180	79% vs 40%	<0.001
	REG3α	13% vs 32%	<0.001	15% vs 12%	0.376	80% vs 60%	<0.001
	REG3α+AREG	11% vs 33%	<0.001	15% vs 13%	0.836	82% vs 58%	<0.001
	ST2+AREG	12% vs 36%	<0.001	14% vs 13%	0.676	81% vs 55%	<0.001
		CI 12 month NRM	**P-value**	**CI 12 Month Relapse**	**P-value**	**12 Month OS**	**P-value**
Systemically treated subset (n=321)	**AREG**	12% vs 36%	<0.001	14% vs 14%	0.836	79% vs 56%	<0.001
	ST2+REG3α	14% vs 51%	<0.001	15% vs 9%	0.192	78% vs 40%	<0.001
	ST2+REG3α+AREG	13% vs 38%	<0.001	15% vs 11%	0.132	78% vs 54%	<0.001
	ST2	16% vs 53%	<0.001	15% vs 12%	0.493	76% vs 40%	<0.001
	REG3α	15% vs 35%	<0.001	15% vs 12%	0.471	77% vs 57%	<0.001
	REG3α+AREG	15% vs 37%	<0.001	14% vs 14%	0.961	78% vs 53%	<0.001
	ST2+AREG	13% vs 35%	<0.001	14% vs 14%	0.947	79% vs 55%	<0.001

Supplemental Table 12: Cumulative incidence of NRM by Minnesota Risk, systemically treated subset of validation cohort
(n=321)

	NRM	P value
Minnesota Standard Risk	18%	-
ST2+REG3α	12% vs 35%	<0.001
AREG	13% vs 23%	0.022
Minnesota High Risk	42%	-
ST2+REG3α	26% vs 48%	0.116
AREG	8% vs 51%	0.011

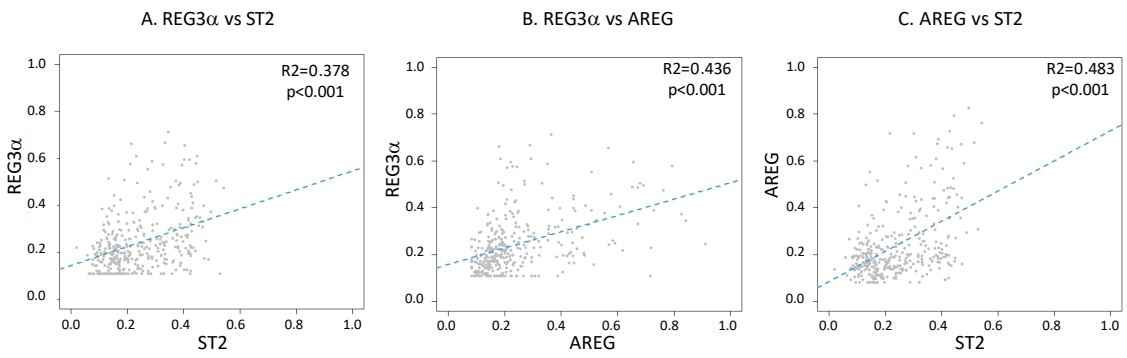
Supplemental Table 13: Area under the curve for algorithms by subsets
Validation cohort patients

		AUC for ST2+REG3α	AUC for AREG	p-value
Validation cohort (n=374)	Minnesota High Risk (n=53)	0.689	0.780	0.162
	Minnesota Standard Risk (n=321)	0.720	0.636	0.035
	LGI Involvement (n=109)	0.790	0.773	0.678
	Skin Only Involvement (n=199)	0.629	0.563	0.312
	Post-Transplant Cyclophosphamide Prophylaxis (n=133)	0.717	0.595	0.114
	Systemically Treated (n=321)	0.739	0.693	0.144
		AUC for ST2+REG3α	**AUC for AREG**	**p-value**
Systemically treated subset (n=321)	Minnesota High Risk (n=52)	0.679	0.780	0.126
	Minnesota Standard Risk (n=269)	0.700	0.617	0.052
	LGI involvement (n=104)	0.777	0.768	0.827
	Skin Only Involvement(n=158)	0.610	0.555	0.427
	Post-Transplant Cyclophosphamide Prophylaxis (n=125)	0.711	0.592	0.129



Supplemental Figure 1: Cumulative incidence of 12-month NRM using risk groups identified by published algorithms and accompanying thresholds.

Pie charts show the proportion of patients classified as high risk (HR, red border) and low risk (LR, blue border). The proportion correctly classified as HR or LR are shaded red or blue, respectively. The proportion incorrectly classified are shaded in gray. The cumulative incidence curves show 12-month NRM by risk group. **A and B**: Combined training + validation cohorts [n=715] **A.** ST2+REG3 by Ann Arbor 1 (LR) vs 2/3 (HR). 12-month NRM 12% vs. 35% (p<0.001). **B.** AREG <33 pg/ml (LR) vs ≥33 pg/ml (HR): 12-month NRM 14% vs. 43% (p<0.001). **C and D.** Systemically treated subset [n=622] **C.** ST2+REG3 by Ann Arbor 1 (LR) vs 2/3 (HR). 12-month NRM 14% vs. 36% (p<0.001). **D.** AREG <33 pg/ml (LR) vs ≥33 pg/ml (HR): 12-month NRM 16% vs. 45% (p<0.001).



Supplemental Figure 2: Correlation of predicted probabilities of AREG, REG3α, and ST2 algorithms in the training cohort.

A. REG3α and ST2. R^2 0.378, p<0.001 **B.** REG3α and AREG. R^2 0.436, p<0.001. **C.** AREG and ST2. R^2 0.489, p<0.001.