A Diagnostic Classifier for Pediatric Chronic Graft-Versus-Host Disease: Results of the ABLE / PBMTC 1202 Study

SUPPLEMENTARY DATA FILES

Geoffrey D.E. Cuvelier, MD; Bernard Ng, PhD; Sayeh Abdossamadi, PhD; Eneida R. Nemecek, MD, MS; Alexis Melton, MD; Carrie L. Kitko, MD; Victor A. Lewis, MD; Tal Schechter, MD; David A. Jacobsohn, MD; Andrew C. Harris, MD; Michael A. Pulsipher, MD; Henrique Bittencourt, MD, PhD; Sung Won Choi, MD; Emi H. Caywood, MD; Kimberly A. Kasow, DO; Monica Bhatia, MD; Benjamin R. Oshrine, MD; Sonali Chaudhury, MD; Donald Coulter, MD; Joseph H. Chewning, MD; Michael Joyce, MD, PhD; Süreyya Savaşan, MD; Anna B. Pawlowska, MD; Gail C. Megason, MD; David Mitchell, MD; Alexandra C. Cheerva, MD; Anita Lawitschka, MD; Elena Ostroumov, PhD; and Kirk R. Schultz, MD.

Supplementary Table 1. Baseline characteristics. ALL, Acute Lymphoblastic Leukemia; AML, Acute Myelogenous Leukemia; HLA, Human Leukocyte Antigen; MDS, Myelodysplastic Syndrome

Characteristic	**Chronic GVHD Group**	**Control Group**
	(N=44)	**(N=190)**
***Age at HCT**, years (median)*	12.0	8.9
(range)	(2.0 – 18)	(0.2 – 18.1)
Sex		
Male	28 (64%)	101 (53%)
Female	16 (36%)	89 (47%)
Indication for Transplant		
Malignant (n=161, 69%)	N=35 (80%)	N=126 (66%)
ALL	14	56
MDS/AML	14	57
Other	7	13
Non-Malignant (n=73, 31%)	N=9 (20%)	N=64 (34%)
Inherited Marrow Failure	2	8
Primary Immune Deficiency	1	15
Severe Aplastic Anemia	2	13
Sickle Cell Anemia	3	16
Thalassemia	0	5
Metabolic	1	7
Graft Source		
Bone Marrow	22 (50%)	126 (66%)
Peripheral Blood Stem Cells	16 (36%)	32 (17%)
Umbilical Cord	6 (14%)	32 (17%)
Type of Donor / HLA Match		
Matched Sibling	8 (18%)	66 (35%)
Matched-Related Non-Sibling	0 (0%)	6 (3%)
Matched Unrelated	19 (43%)	80 (42%)
Mismatched Unrelated	16 (36%)	34 (18%)
Haploidentical Family Member	1 (2%)	4 (2%)
Conditioning Regimen		
Myeloablative	39 (89%)	164 (86%)
Reduced Intensity	5 (11%)	26 (14%)

	Chronic GVHD Group	Control Group
Serotherapy		
Anti-Thymocyte Globulin	19 (43%)	83 (44%)
Alemtuzumab	2 (5%)	24 (12%)
None	23 (52%)	83 (44%)
GVHD in First Year After Transplant		
None (no acute or chronic)	N/A	87 (46%)
Acute GVHD Only Before Day 100	N/A	45 (24%)
Late Acute GVHD After Day 100	N/A	58 (30%)
Maximal Grade Acute GVHD (n=45)		
Grade 1	N/A	19 (42%)
Grade 2	N/A	17 (38%)
Grade 3	N/A	7 (16%)
Grade 4	N/A	2 (4%)
Type of Late Acute GVHD (n=58)		
De Novo *(no previous acute)*	N/A	12 (21%)
Recurrent *(previous acute that resolved then returned as late acute GVHD)*	N/A	18 (31%)
Progressive *(acute GVHD that progressed after day 100 into late acute GVHD)*	N/A	28 (48%)
Type of Chronic GVHD (N=44)		
De Novo *(no previous history of any GVHD before onset of chronic GVHD)*	7 (16%)	N/A
Quiescent *(history of previous acute GVHD that resolved and later developed chronic GVHD)*	19 (43%)	N/A
Progressive *(history of acute or late acute GVHD that developed into chronic GVHD, including all cases of overlap syndrome)*	18 (41%)	N/A
Maximal NIH Severity of Chronic GVHD in First Year Post Transplant (n=44)		
Mild	10 (23%)	N/A
Moderate	16 (36%)	N/A
Severe	18 (41%)	N/A

I. Flow Cytometry / Phenotyping Methods and Cellular Populations

Blood was drawn from the study participant and shipped overnight to arrive in the Schultz laboratory the following day. Phenotyping was usually performed on the day of sample arrival in the Schultz laboratory, and always within 5-days. Six flow cytometry panels were designed to evaluate for various subpopulations of T, T_{REGS}, B, NK cells, and myeloid cells. All antibodies, corresponding conjugated dyes, and vendors are provided here:

Supplementary Table 2. Flow cytometry antibodies used per panel

Panel 1 (B-Cell)

	Fluorophore	Antibody	Catalog Number	Manufacturer
1	Pacific Blue	Anti-human CD19	302232	BioLegend
2	FITC	Anti-human CD21	354910	BioLegend
3	PE	Anti-human CD38	303506	BioLegend
4	APC	Anti-human CD10	312210	BioLegend
5	Brilliant Violet 785	Anti-human CD27	302832	BioLegend
6	PerCP	Anti-human CD5	300618	BioLegend
7	Brilliant Violet 510	Anti-human IgD	348220	BioLegend

Panel 2 (NK Cell)

	Fluorophore	Antibody	Catalog Number	Manufacturer
1	PerCP	Anti-human CD3	300428	BioLegend
2	Alexa Fluor 488	Anti-human CD56	304611	BioLegend
3	Brilliant Violet 421	Anti-human CD335	331914	BioLegend
4	Alexa Fluor 647	Anti-human CD69	310918	BioLegend
5	PE	Anti-human CD337	325208	BioLegend
6	APC Cy7	Anti-human CXCR3	353722	BioLegend

Panel 3 (T cytotoxic)

	Fluorophore	Antibody	Catalog Number	Manufacturer
1	PE	Anti-human CD56	304606	BioLegend
2	APC Cy7	Anti-human CD3	300426	BioLegend
3	Pacific Blue	Anti-human CD8	301033	BioLegend
4	Alexa Fluor 488	Anti-human Perforin	308108	BioLegend
5	Alexa Fluor 647	Anti-human Granzyme B	515406	BioLegend

Panel 4 (TCM/TEM)

	Fluorophore	Antibody	Catalog Number	Manufacturer
1	APC Cy7	Anti-human CD3	300426	BioLegend
2	Alexa Fluor 488	Anti-human CD4	344618	BioLegend
3	Pacific Blue	Anti-human CD45RA	304123	BioLegend
4	PE	Anti-human CCR7	353204	BioLegend
5	Brilliant Violet 510	Anti-human CD8	301048	BioLegend

6	Brilliant Violet 785	Anti-human CD27	302832	BioLegend
7	Alexa Fluor 647	Anti-human PD1	329910	BioLegend

Panel 5 (Treg)

	Fluorophore	Antibody	Catalog Number	Manufacturer
1	APC Cy7	Anti-human CD3	300426	BioLegend
2	Brilliant Violet 785	Anti-human CD4	317442	BioLegend
3	FITC	Anti-human CD45RA	304106	BioLegend
4	PE	Anti-human CD25	356104	BioLegend
5	PerCP Cy5.5	Anti-human CD127	351322	BioLegend
6	Alexa Fluor 647	Anti-human PD1	329910	BioLegend
7	Brilliant Violet 421	Anti-human CD31	303124	BioLegend

Panel 6 (Leukocytes)

	Fluorophore	Antibody	Catalog Number	Manufacturer
1	Alexa Fluor 647	Anti-human CD45	304018	BioLegend
2	PerCP	Anti-human CD3	300428	BioLegend
3	Brilliant Violet 785	Anti-human CD19	302240	BioLegend
4	Alexa Fluor 488	Anti-human CD56	304611	BioLegend
5	PE	Anti-human CD66b	305106	BioLegend
6	APC-eFluor 780	Anti-human CD14	47-0149-42	Invitrogen
7	Brilliant Violet 421	Anti-human CD13	301716	BioLegend

Samples were stained in the dark for 12 minutes at room temperature followed by treatment with fix/red blood cell lyze solution (eBiosceinces, Thermo Fisher Scientific, Waltham, MA). For intracellular staining, cells were made permeable using BD Perm II solution (BD Biosciences Mississauga, ON, Canada). Flow cytometry data were acquired using BD LSR Fortessa X-20 Special Order four channel flow cytometer (BD Biosciences, San Jose, CA). A minimum of 300,000 events were acquired for all panels. Instrument settings was also standardized using SPHERO Rainbow Calibration particles 6 peaks (Sphereotech, Lake Forest, IL) to adjust laser power drifts over time. Flow cytometry analysis files were analyzed using Kaluza software v2 (Beckman Coulter, Mississauga, ON, Canada). Flow cytometry accuracy and reproducibility were ensured by the approaches described in detail in the supplement of Schultz et al. Blood 2020: 135(15): 1287-1298 (reference 17 in manuscript) with minimal batch variability.

Cellular subpopulations were defined as follows (adapted from Schultz et al. Blood 2020: 135(15): 1287-1298):

Supplementary Table 3: Immunophenotype and Cell Type Subpopulation

Naive T helper cell populations	
Naïve Th cells	$CD31^{-}CD45RA^{+}CD4^{+}$ T cells
RTE Naïve Th cells	$CD31^{+}CD45RA^{+}CD4^{+}$ T cells
$PD1^{+}$ Naïve Th cells	$PD1^{+}$ $CD45RA^{+}$ $CD4^{+}$ T cells
$PD1^{-}$ Naïve Th cells	$PD1^{-}$ $CD45RA^{+}$ $CD4^{+}$ T cells
$CCR7^{+}$ Naïve Th cells	$CCR7^{+/-}$ $CD45RA^{+}$ $CD4^{+}$ T cells
$CCR7^{-}$ Naïve Th cells	$CCR7^{-}$ CD45RA+ $CD4^{+}$ T cells

$CD27^{+}$ Naïve Th cells	$CD27^{+}CD45RA^{+}CD4^{+}$ T cells
$CD27^{-}$ Naïve Th cells	$CD27^{-}CD45RA^{+}CD4^{+}$ T cells
Follicular Th cells	$PD1^{++}$ $CD45RA^{+}$ $CD4^{+}$ T cells
Memory T helper cell populations	
$PD1^{+/-}$ memory Th cells	$PD1^{+/-}$ $CD45RA^{-}$ $CD4^{+}$ T cells
$CCR7^{+}$ Memory Th cells	$CCR7^{+}CD45RA^{-}$ $CD4^{+}$ T cells
$CCR7^{-}$ Memory Th cells	$CCR7^{-}CD45RA^{-}$ $CD4^{+}$ T cells
$CD27^{+/-}$ Memory Th cells	$CD27^{+/-}$ $CD45RA^{-}CD4^{+}$ T cells
$CD31^{-}$ Memory Th cells	$CD31^{-}$CD45RA-$CD4^{+}$ T cells
RTE Memory Th cells	$CD31^{+}CD45RA^{-}CD4^{+}$ T cells
Naïve Tc cell populations	
Naïve Tc cells	$CD31^{-}CD45RA^{+}CD8^{+}$ T cells
RTE Naïve Tc cells	$CD31^{+}CD45RA^{+}CD8^{+}$ T cells
$PD1^{+}$ Naïve Tc cells	$PD1^{+}$ $CD45RA^{+}$ $CD8^{+}$ T cells
$PD1^{-}$ Naïve Tc cells	$PD1^{-}$ $CD45RA^{+}$ $CD8^{+}$ T cells
$CCR7^{+}$ Naïve Tc cells	$CCR7^{+}$ $CD45RA^{+}$ $CD8^{+}$ T cells
$CCR7^{-}$ Naïve Tc cells	$CCR7^{-}$ $CD45RA^{+}$ $CD8^{+}$ T cells
$CD27^{+}$ Naïve Tc cells	$CD27^{+}CD45RA^{+}CD8^{+}$ T cells
$CD27^{-}$ Naïve Tc cells	$CD27^{-}CD45RA^{+}CD8^{+}$ T cells
Memory Tc cell populations	
$PD1^{+/-}$ memory Tc cells	$PD1^{+/-}$ $CD45RA^{-}$ $CD8^{+}$
$CCR7^{+/-}$ Memory Tc cells	$CCR7^{+/-}CD45RA^{-}$ $CD8^{+}$ T cells
Cytolytic Tc cells	$Perforin^{+}$ Granzyme B^{+} $CD8^{+}$
$CD27^{+/-}$ Tc cells	$CD27^{+/-}$ $CD45RA^{-}CD4^{+}$ T cells
Treg cells	
All Memory T_{reg} populations	$CD45RA^{-}$ Treg cells $CD31^{-}CD45RA^{-}$ T_{reg} cells
RTE Memory T_{reg} populations	$CD31^{+}CD45RA^{-}$ T_{reg} cells
$PD1^{-}$Memory Treg	$PD1^{-}CD45RA^{-}$ Treg cells
$PD1^{+}$Memory Treg	$PD1^{+}CD45RA^{+}$ Treg cells
$PD1^{+}$ Naïve Treg cells	$PD1^{+}CD45RA^{+}$ T_{reg} cells
$PD1^{-}$ Naïve Treg	$PD1^{-}CD45RA^{+}$ T_{reg} cells
Naïve T_{reg} populations	$CD31^{-}CD45RA^{+}$ T_{reg} cells
RTE Naïve T_{reg} populations	$CD31^{+}CD45RA^{+}$ T_{reg} cells
B cell populations	
T1 –Transitional consistent with Breg cells	$CD10^{high}CD38^{high}CD19^{+}$ B cells
CD21 ^{low} B cells	$CD21^{low}CD19^{+}$ B cells
T2 transitional	$CD38^{int}CD10^{int}$ $CD19^{+}$B cells
T3 transitional	$CD38^{dim}CD10^{low}$ $CD19^{+}$ B cells
Mature Naïve B cells	$IgD^{+}CD27^{-}CD19^{+}$ B cells
Unswitched memory/ Marginal-zone like	IgD^{+} $CD27^{+}CD19^{+}$ B cells
Classic Switched memory	$IgD^{-}CD27^{+}CD19^{+}$ B cells
Late Memory B cell	$IgD^{-}CD27^{-}CD19^{+}$ B cells
Plasma cells	$CD38^{high}CD10^{-}CD19^{+}$ B cells
Regulatory NK cells (noncytolytic)	
NK_{reg} cells	$CD56^{high}Perforin^{low}$ NK cells $CD56^{high}$ $CD335^{high}$ NK cells $CD56^{high}$ Granzyme B^{low} NK cells
$CD56^{bright}$ cytolytic NK cells	$CD56^{high}$ $CD335^{high}$ NK cells $CD56^{high}$ $^{-}Perforin^{high}$ NK cells

Activated CD56^{bright} NK cells	CD56^{high} Granzyme B^{high} NK cells CD56^{high} CD69^{+} NK cells
Classic NK cells *Activated classic NK cells*	CD56^{low} CD335^{low} NK cells CD56^{low} Perforin^{high} NK cells CD56^{low} Granzyme B^{high} NK cells CD56^{low} CD69^{+} NK cells
NKT cells	CD56^{+} CD3^{+}
Myeloid Population	
Monocytes *Monocytes/Neutrophils*	CD14^{+}CD45^{+} CD66b^{+}CD45^{+}

II. Development of Machine Learning-Based Classifier for Chronic GVHD Diagnosis

In addition to analyzing each marker in a univariate manner, we developed a machine learning-based classifier that combines multiple cellular and plasma markers along with clinical factors for diagnosing whether a patient has cGVHD. The approach is summarized in Figure 4 of the manuscript. We first randomly selected 10 cGVHD samples and 10 control samples as the test set for classifier evaluation. If a control sample was from one of the selected cGVHD subjects (i.e., a measurement made prior to cGvHD onset) or another sample had already been drawn from the same control subject, we randomly drew another control sample, since in real clinical settings, we would immediately diagnose a subject when marker measurements become available. All remaining samples, except those acquired at later time points from subjects in the test set, were used for classifier training. To account for outlier marker values, we performed winsorization by first estimating the median and median absolute deviation (MAD) of each marker based on the training samples. We then clipped all marker values at 3 standard deviations away from the training median with standard deviation estimated as $1.483 \cdot \text{MAD}$.[1] To deal with missing marker values in the training set, we applied k-nearest neighbors (k=15) to impute the

missing values. To reduce the number of markers, we performed feature selection using a bootstrapping approach. Specifically, we extracted 1,000 bootstrap samples from the training set. For each bootstrap sample, we applied a Student's t-test to compare the marker values of cGVHD samples against controls for each marker. Markers with $p<0.05$ for >99% of the bootstrap samples (i.e., selection frequency >0.99) were selected for classifier training. We trained a support vector machine (SVM) classifier with the selected markers and all clinical factors used in the regression analysis. SVM finds an optimal weighting of the selected markers and clinical factors that best separates cGVHD samples from control samples. To account for class imbalance (i.e., the training set had many more control samples than cGVHD samples), we set the penalty weight for misclassifying cGvHD (controls) to the number of training samples over the number of cGvHD (control) samples. Setting a higher penalty for misclassifying cGVHD reduces the bias towards classifying samples as controls. To deal with missing marker values in the test set, we applied k-nearest neighbors (k=15) using only values of the selected markers from training samples to impute the missing values in each test sample. Imputing each test sample separately without using other test samples better emulates real clinical settings. To evaluate the classifier, we applied it to the test set and computed its positive predictive value (PPV), negative predictive value (NPV), and area under the receiver operator characteristic curve (AUC). PPV is the proportion of test samples classified as cGvHD that are truly cGvHD, and NPV is the proportion of test samples classified as controls that are truly controls. The described procedures were repeated 1,000 times with different random train-test sample splits to assess variability in classification performance arising from sample variability. The average training AUC over the 1,000 random train-test sample splits was 0.95 ($\pm$ 0.01). We note that the selected markers would

normally vary across random train-test sample splits due to sample variability. However, since we used a bootstrapping procedure with an extremely high selection frequency threshold of 99%, the selected markers highly overlapped across the random train-test sample splits. To report one representative set of markers for future validation, two common strategies are often taken in practice. One strategy is to report markers that are selected in majority (e.g. 99%) of the random train-test sample splits. The other strategy is to apply the marker selection procedure to all samples. Since we used bootstrapping with a selection frequency threshold of 99% as our marker selection procedure, the selected markers were identical with both strategies. The reported selection frequencies correspond to applying our marker selection procedure to all samples.

Reference:

1. I. Gijbels, M. Hubert: Robust and Nonparametric Statistical Methods, in Brown SD, Tauler R, and Walczak B (eds): Comprehensive Chemometrics. Chemical and Biochemical Data Analysis. Elseiver, 2009, volume 1, pp 189-211.

Supplementary Table 4. ***Mixed effect modeling of*** **cellular and plasma chronic GVHD diagnostic biomarkers (present at the onset of chronic GVHD) compared to control individuals without chronic GVHD.** Chronic GVHD patients included all chronic GVHD severities (mild, moderate, severe) according to the NIH consensus criteria. Biomarkers highlighted in blue also met the same criteria at all three time points in the fixed effect model. Biomarkers highlighted in yellow met the same criteria in one or two time points (but not all three time points) in the fixed effect model . AUC, Area Under Curve; NK_{REG}, Regulatory NK cell; RTE, Recent Thymic Emigrant ($CD4^{+}CD45RA^{+}CD31^{+}$); Th, Helper T cell ($CD4^{+}$); T_{REG}, Regulatory T cell ($CD4^{+}CD127^{Low}CD25^{+}$).

Cellular Biomarkers					
Diagnostic Biomarker (As a % of Parent Cell Population)	**Relationship cGvHD to Controls**	**Mean Absolute Values (All cGvHD Cases Compared to Controls)**	**Effect Ratio**	**P-value**	**AUC**
Natural Killer Cells					
CD56+ NK cells (% of lymphocytes)	Decreased	9.7% vs 13.6%	0.71	0.0001	0.71
Regulatory Natural Killer Cells (NK_{REG})					
$CD56^{bright}$ $Perforin^{Negative}$ (%CD56) (NK_{REG}, Non-Cytolytic)	Decreased	13.6% vs 25.7%	0.53	2.5×10^{-4}	0.68
Naïve Helper T Cells (Naïve Th)					
CD4+CD45RA+ (%CD3)	Decreased	6.4% vs 13.9%	0.46	3.5×10^{-6}	0.75
CD4+CD45RA+ CCR7+ (%CD4)	Decreased	9.6% vs 25.9%	0.37	9.3×10^{-8}	0.78
CD4+CD45RA+PD1- (%CD4)	Decreased	13.1% vs 25.7%	0.51	4.9×10^{-6}	0.73
CD4+CD45RA+CD27+ (%CD4)	Decreased	11.7% vs 26.6%	0.44	6.6×10^{-8}	0.77
CD4+CD45RA+CD31+ (%CD4) (Recent Thymic Emigrants)	Decreased	8.5% vs 23.7%	0.36	1.1×10^{-6}	0.76

Naïve Regulatory T Cells (T_{REG})					
CD45RA+PD1- T_{REG} (%T_{REG})	Decreased	10.2% vs 23.5%	0.43	3.5×10^{-6}	0.73
CD45RA+CD31+ T_{REG} (%T_{REG}) (Recent Thymic Emigrant, Naïve T_{REG})	Decreased	5.3% vs 15.5%	0.34	1.7×10^{-5}	0.72
Memory Helper T Cells					
CD4+ CD45RA-CCR7- (%CD4) (Effector Memory Th)	Increased	45.4% vs 32.3%	1.40	5.4×10^{-5}	0.68
Plasma Biomarkers					
Diagnostic Biomarker	**Relationship cGvHD to Controls**	**Mean Absolute Values (All cGvHD Cases Compared to Controls)**	**Effect Ratio**	**P-value**	**AUC**
CXCL9 (pg/mL)	Increased	421 vs 162 pg/mL	2.59	1×10^{-16}	0.77
CXCL10 (pg/mL)	Increased	609 vs 250 pg/mL	2.44	1×10^{-16}	0.71
CXCL11 (pg/mL)	Increased	2132 vs 921 pg/mL	2.31	1.5×10^{-9}	0.70
ICAM-1 (ng/mL)	Increased	510 vs 364 ng/mL	1.4	1×10^{-7}	0.71
ST2 (pg/mL)	Increased	59,644 vs 23,669 pg/mL	2.52	1×10^{-11}	0.73
sCD13 (Aminopeptidase N) Enzyme Activity mU/mL	Increased	1.02 vs 0.68 mU/mL	1.49	5.6×10^{-7}	0.71

Supplementary Table 5. ***Mixed effect modeling of*** **cellular and plasma chronic GVHD diagnostic biomarkers (present at the onset of chronic GVHD) in moderate-severe chronic NIH consensus criteria GVHD (n=34) compared to control individuals without chronic GVHD.** Patients with mild chronic GVHD according to the NIH consensus criteria were removed from analysis. AUC, Area Under Curve; NK_{REG}, Regulatory NK cell; RTE, Recent Thymic Emigrants ($CD4^+CD45RA^+CD31^+$); Th, Helper T cell ($CD4^+$); T_{REG}, Regulatory T cell ($CD4^+CD127^{Low}CD25^+$).

Cellular Biomarkers					
Diagnostic Biomarker (As a % of Parent Cell Population)	**Relationship cGvHD to Controls**	**Mean Absolute Values (All cGvHD Cases compared to controls)**	**Effect Ratio**	**P-value**	**AUC**
Regulatory Natural Killer Cells (NK_{REG})					
$CD56^{Bright}Perforin^{NEGATIVE}$ (%CD56) (NK_{REG} Non-Cytolytic)	Decreased	12% vs 25.8%	0.47	1.2×10^{-5}	0.73
$CD56^{Bright}$Granzyme $B^{NEGATIVE}$ (%CD56) (NK_{REG} Non-Cytolytic)	Decreased	6.4% vs 16.9%	0.38	0.0003	0.69
Naïve Helper T Cells (Naïve Th)					
CD4+CD45RA+ (%CD3)	Decreased	6.3% vs 13.9%	0.45	7.4×10^{-5}	0.74
CD4+CD45RA+CCR7+ (%CD4)	Decreased	9.9% vs 25.9%	0.38	6.3×10^{-7}	0.78
CD4+CD45RA+PD1- (%CD4)	Decreased	12.9% vs 25.7%	0.50	1.7×10^{-5}	0.74
CD4+CD45RA+CD27+ (%CD4)	Decreased	12.2% vs 26.7%	0.46	5.2×10^{-7}	0.77
CD4+CD45RA+CD31+ (%CD4) (Recent Thymic Emigrants)	Decreased	8.9% vs 23.7%	0.37	1.2×10^{-5}	0.75
Naïve Regulatory T Cells (T_{REG})					
CD45RA+PD1- T_{REG} (%T_{REG})	Decreased	9.6% vs 23.5%	0.41	8.1×10^{-6}	0.74
CD45RA+CD31+ T_{REG} (%T_{REG}) (Recent Thymic Emigrant, Naïve T_{REG})	Decreased	5.4% vs 15.6%	0.35	0.0001	0.72
Memory Helper T Cells					
CD4+CD45RA-CCR7- (%CD4) (Effector Memory Th)	Increased	43.8% vs 32.3%	1.36	0.0001	0.69
Plasma Biomarkers					
Diagnostic Biomarker	**Relationship cGvHD to Controls**	**Mean Absolute Values (All cGvHD Cases compared to controls)**	**Effect Ratio**	**P-value**	**AUC**
CXCL9 (pg/mL)	Increased	448 vs 163	2.74	1×10^{-16}	0.78

CXCL10 (pg/mL)	Increased	652 vs 250 pg/mL	2.6	1×10^{-16}	0.72
CXCL11 (pg/mL)	Increased	2235 vs 927 pg/mL	2.4	2.4×10^{-9}	0.72
ICAM-1 (ng/mL)	Increased	511 vs 364 ng/mL	1.4	1.1×10^{-7}	0.73
ST2 (pg/mL)	Increased	63,984 vs 23,376 pg/mL	2.73	8×10^{-11}	0.74
sCD13 (Aminopeptidase N) Enzyme Activity mU/mL	Increased	1.04 vs 0.68 mU/mL	1.53	5.3×10^{-6}	0.72

Supplementary Table 6. ***Fixed effect modeling of*** **cellular and plasma chronic GVHD diagnostic biomarkers (present at the onset of chronic GVHD) compared to all control individuals without chronic GVHD.** Chronic GVHD patients included all chronic GVHD severities (mild, moderate, severe), according to the NIH consensus criteria. Markers highlighted in blue met these three criteria at all three time points of chronic GVHD onset (early-, mid-, and late-onset). Markers highlighted in yellow met the three criteria at one or two (but not all three) time points for chronic GVHD onset. AUC, Area Under Curve; NK_{REG}, Regulatory NK cell; RTE, Recent Thymic Emigrants ($CD4^{+}CD45RA^{+}CD31^{+}$); Th, Helper T cell ($CD4^{+}$); T_{REG}, Regulatory T cell ($CD4^{+}CD127^{Low}CD25^{+}$).

Diagnostic Marker	Early Onset cGvHD (<4 months) (n=11)			Mid Onset cGvHD (4-8 months) (n=24)			Late Onset cGvHD (≥8 months) (n=9)		
	Relationship of cGvHD to Control	Mean Absolute Values (Effect Ratio)	p-value (AUC)	Relationship of cGvHD to Control	Mean Absolute Values (Effect Ratio)	p-value (AUC)	Relationship of cGvHD to Control	Mean Absolute Values (Effect Ratio)	p-value (AUC)
Cellular Biomarkers									
Natural Killer Cells									
CD56+ NK cells (% of lymphocytes)	Decreased	9.9% vs 20.0% (0.50)	0.003 (0.81)	Decreased	9.9% vs 13.4% (0.74)	0.02 (0.65)	Not Significant	9.4% vs 7.7% (1.23)	0.99 (0.54)
Regulatory Natural Killer Cells (NK_{REG})									
$CD56^{Bright}$ $Perforin^{Negative}$ (%CD56) (NK_{REG} Non-Cytolytic)	Decreased	20% vs 32.2% (0.62)	0.03 (0.69)	Decreased	14.9% vs 27.8% (0.54)	0.0008 (0.70)	Decreased	5.9% vs 17.1% (0.34)	0.004 (0.80)
$CD56^{Bright}$ $Granzyme\ B^{Negative}$ (%CD56) (NK_{REG} Non-Cytolytic)	Not Significant	11.1% vs 17.7% (0.63)	0.16 (0.61)	Decreased	8.8% vs 19% (0.46)	0.005 (0.68)	Decreased	2.6% vs 14% (0.19)	0.008 (0.86)
$CD56^{Bright}$ $CD335^{High}$ (%CD56) (NK_{REG} Non-Cytolytic)	Not Significant	24% vs 33.7% (0.71)	0.07 (0.66)	Decreased	18.7% vs 28.9% (0.65)	0.008 (0.65)	Decreased	6.8% vs 17.3% (0.40)	0.02 (0.73)
Naïve Helper T Cells (Naïve Th)									
CD4+CD45RA+ (%CD3)	Decreased	3.8% vs 8.8% (0.43)	0.008 (0.74)	Decreased	5.8% vs 10.2% (0.57)	0.04 (0.62)	Decreased	9.7% vs 22.7% (0.43)	0.05 (0.70)
CD4+CD45RA+CCR7+ (%CD4)	Decreased	5.6% vs 13.6% (0.41)	0.005 (0.73)	Decreased	11.8% vs 21.6% (0.54)	0.03 (0.62)	Decreased	11.4% vs 42.4% (0.27)	0.001 (0.84)

CD4+CD45RA+PD1- (%CD4)	Decreased	8.4% vs 14.0% (0.60)	0.04 (0.70)	Decreased	12.9% vs 21.4% (0.60)	0.04 (0.61)	Decreased	17.8% vs 41.7% (0.43)	0.01 (0.77)
CD4+CD45RA+CD27+ (%CD4)	Decreased	6.5% vs 14.3% (0.45)	0.007 (0.72)	Decreased	12.3% vs 22.4% (0.55)	0.02 (0.64)	Decreased	16.3% vs 43.1% (0.38)	0.006 (0.77)
CD4+CD45RA+CD31+ (%CD4) (Recent Thymic Emigrants)	Not Significant	6.0% vs 11.1% (0.54)	0.06 (0.68)	Decreased	9.2% vs 20.7% (0.44)	0.009 (0.66)	Decreased	10.4% vs 39.2% (0.27)	0.007 (0.80)
Naïve Regulatory T Cells (T_{REG})									
CD45RA+PD1- T_{REG} (%T_{REG})	Decreased	9.0% vs 17.9% (0.50)	0.009 (0.77)	Decreased	11.5% vs 19.2% (0.60)	0.01 (0.65)	Decreased	10% vs 33.5% (0.30)	0.02 (0.78)
CD45RA+CD31+ T_{REG} (%T_{REG}) (RTE Naïve T_{REG})	Decreased	3.8% vs 9.3% (0.41)	0.02 (0.72)	Decreased	6.3% vs 12.6% (0.50)	0.02 (0.63)	Decreased	5.9% vs 24.8% (0.24)	0.03 (0.77)
Memory Helper T Cells									
CD4+CD45RA-CCR7- (%CD4) (Effector Memory Th)	Not Significant	51.0% vs 42.7% (1.19)	0.10 (0.64)	Increased	44.3% vs 34.2% (1.3)	0.03 (0.65)	Increased	40.7% vs 20% (2.04)	0.0009 (0.83)
Plasma Biomarkers									
CXCL9 (pg/mL)	Increased	332 vs 128 pg/mL (2.6)	1×10^{-5} (0.73)	Increased	414 vs 197 pg/mL (2.1)	0.003 (0.68)	Increased	516 vs 164 pg/mL (3.15)	3.6×10^{-8} (0.89)
CXCL10 (pg/mL)	Increased	416 vs 206 pg/mL (2.02)	0.0002 (0.68)	Increased	693 vs 261 pg/mL (2.66)	3×10^{-7} (0.70)	Increased	719 vs 282 pg/mL (2.54)	0.01 (0.62)
CXCL11 (pg/mL)	Increased	1312 vs 768 pg/mL (1.71)	0.02 (0.60)	Increased	2428 vs 870 pg/mL (2.79)	8.3×10^{-8} (0.77)	Increased	2656 vs 1135 pg/mL (2.34)	0.02 (0.74)
ICAM-1 (ng/mL)	Increased	435 vs 335 ng/mL (1.3)	0.03 (0.71)	Not Significant	469 vs 390 ng/mL (1.20)	0.01 (0.65)	Increased	626 vs 369 ng/mL (1.7)	0.00003 (0.77)
ST2 (pg/mL)	Not Significant	34,187 vs 28,594 pg/mL (1.2)	0.57 (0.57)	Increased	53,642 vs 22,922 pg/mL (2.34)	3.6×10^{-9} (0.72)	Increased	91,102 vs 19,512 (4.67)	1.4×10^{-11} (0.75)
Soluble CD13 (Aminopeptidase N) (Enzyme Activity, mU/mL)	Increased	1.02 vs 0.52 mU/mL (1.94)	1.1×10^{-6} (0.89)	Not Significant	0.84 vs 0.80 mU/mL (1.05)	0.58 (0.51)	Increased	1.19 vs 0.72 mU/mL (1.64)	9×10^{-6} (0.79)

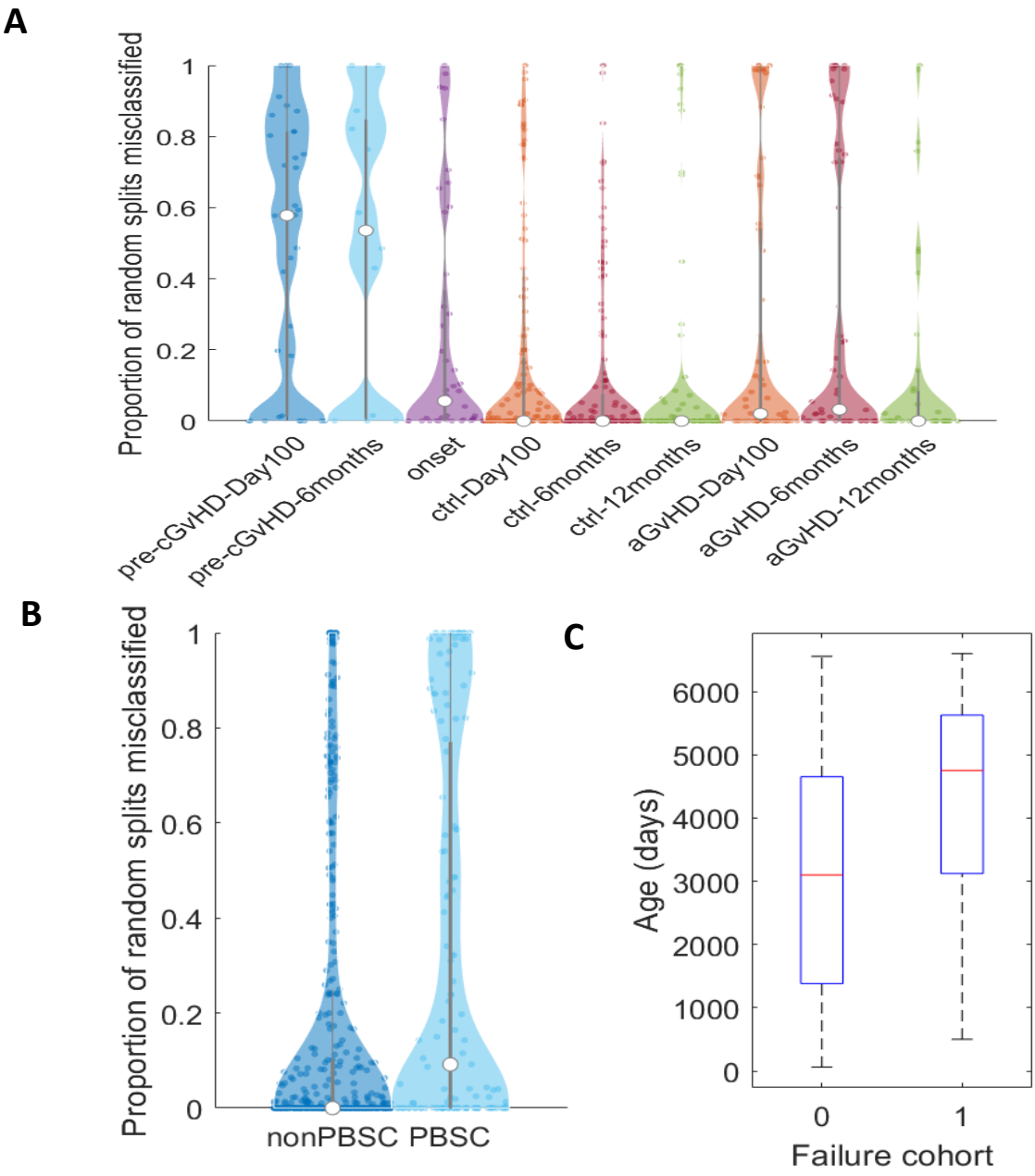
Supplementary Table 7. Classifier weights of various cellular, plasma, and clinical factors used in the diagnostic classifier.

Variables	Classifier Weights
$CD56^{Bright}$ $Perforin^{Negative}$	-0.2652
$CD56^{Bright}$ Granzyme $B^{Negative}$	-0.1645
CD4+ CD45RA+ CCR7+	0.1240
CD4+ CD45RA+ PD1-	0.1572
CD4+ CD45RA+ CD27+	0.0467
CD8+ CD45RA+ CCR7+	-0.1375
CD45RA+ PD1– T_{REG}	-1.1309
CD45RA+ CD31+ T_{REG}	-0.6028
CD4+ CD45RA+ CD31+ (RTE)	-0.0276
ICAM-1	0.2784
CXCL10 (IP10)	0.7568
TIM-3	-0.4277
ST2	0.5265
CXCl11	0.1064
CXCL9	0.3146
Malignant	-0.8049
PBSC	0.4668
Bone Marrow	-0.1481
Sibling	0.3665
Unrelated	0.7533
HLA Antigen Match	-0.1560
M:M (donor:recipient)	0.4341
M:F (donor:recipient)	0.3455
F:M (donor:recipient)	0.7885
ABO match	0.4053
Myeloablative	0.7243
Serotherapy	-0.2899
Age	0.6145
Total Body Irradiation	-0.2387
Days Post-BMT	-0.4817
Classifier offset	-1.7154

Supplementary Table 8. Exploratory *Mixed effect modeling of* cellular and plasma chronic GVHD diagnostic biomarkers (present at the onset of chronic GVHD) by type of chronic GVHD compared to all control individuals without chronic GVHD. All three criteria to be a biologically relevant potential diagnostic biomarker had to be met, including: (1) an effect ratio ≥ 1.3 or ≤0.75, meaning the mean percentage in the chronic GVHD group had to either 30% greater or 30% lower (1.0 / 1.3 = ≤0.75) compared to the mean value of the control group; (2) the area under the curve (AUC) on the receiver operator curve had to be ≥0.60; and (3) $p < 0.05$. None of the markers met the p-value criteria for Bonferroni correction. AUC, Area Under Curve; NK_{REG}, Regulatory NK cell; RTE, Recent Thymic Emigrants ($CD4^+CD45RA^+CD31^+$); Th, Helper T cell ($CD4^+$); T_{REG}, Regulatory T cell ($CD4^+CD127^{Low}CD25^+$).

Diagnostic Biomarker	Pulmonary Phenotype (n=12)		De Novo cGVHD (n=7)		Progressive cGVHD (n=18)	
	Relationship cGVHD to Controls	Mean Values (Effect Ratio) P-Value (AUC)	Relationship cGVHD to Controls	Mean Values (Effect Ratio) P-Value (AUC)	Relationship cGVHD to Controls	Mean Values (Effect Ratio) P-Value (AUC)
Cellular						
CD19 B cells (% of total lymphocytes)	Not Significant	18.2% vs 25% (ER: 0.73) p=0.99 (AUC: 0.51)	Not Significant	28.7% vs 22.0% (ER: 1.3) p=0.19 (AUC: 0.62)	Decreased	14.6% vs 29.2% (ER: 0.50) p=0.01 (AUC: 0.74)
$CD19+CD38^{Low}CD10-$ (% of CD19 B cells) (Transitional T3 B Cell)	Not Significant	36.7% vs 42.8% (ER: 0.86) p=0.82 (AUC: 0.51)	Not Significant	46.6% vs 40% (ER: 1.16) p=0.31 (AUC: 0.67)	Decreased	29.4% vs 49.6% (ER: 0.59) p=0.005 (AUC: 0.78)
CD19+IgD+CD27- (% of CD19 B cells) (Mature naïve B cells)	Not Significant	72.7% vs 79.8% (ER: 0.91) p=0.57 (AUC: 0.55)	Not Significant	79.9% vs 77.4% (ER: 1.03) p=0.86 (AUC: 0.50)	Decreased	61.8% vs 89.3% (ER: 0.69) p=0.004 (AUC: 0.74)
NK T Cells (% of total lymphocytes)	Decreased	1.2% vs 1.9% (ER: 0.65) p=0.04 (AUC: 0.73)	Not Significant	1.2% vs 1.8% (ER: 0.65) p=0.94 (AUC: 0.54)	Not Significant	1.4% vs 1.9% (ER: 0.76) p=0.19 (AUC: 0.64)
$CD56^{Dim}CD69+$ (% of NK cells) (Activated $CD56^{Dim}$ Cytolytic NK cells)	Not Significant	29.1% vs 20% (ER: 1.44) p=0.07 (AUC: 0.66)	Not Significant	21.4% vs 23% (ER: 0.94) p=0.53 (AUC: 0.56)	Increased	29.5% vs 18% (ER: 1.64) p=0.01 (AUC: 0.73)
$CD56^{Bright}CD69^+$ (% of NK cells) (Activated Cytolytic $CD56^{bright}$ NK cells)	Decreased	1.3% vs 2.2% (ER: 0.60) p=0.01 (AUC: 0.77)	Decreased	0.9% vs 2.2% (ER: 0.39) p=0.007 (AUC: 0.86)	Not Significant	2.6% vs 1.5% (ER: 1.7) p=0.15 (AUC: 0.64)
$CD56^{Bright}Perforin^{High}$ (% of NK cells)	Not Significant	1.5% vs 1.6%	Decreased	0.8% vs 1.8% (ER: 0.43)	Not Significant	1.9% vs 1.4% (ER: 1.4)

(CD56^{Bright} Cytolytic NK cells)		(ER: 0.91) p=0.53 (AUC: 0.61)		p=0.03 (AUC: 0.79)		p=0.17 (AUC: 0.64)
CD56^{Bright}Granzyme B^{High} (% of NK cells) (CD56^{Bright} Cytolytic NK cells)	Not Significant	8.7% vs 8.1% (ER: 1.08) p=0.42 (AUC: 0.62)	Decreased	5.9% vs 8.7% (ER: 0.68) p=0.03 (AUC: 0.79)	Not Significant	9.5% vs 7.3% (ER: 1.3) p=0.68 (AUC: 0.53)
CD4+CD45RA+CCR7+ (% of CD4 T cells) (Naïve helper T cells)	Not Significant	5% vs 12.1% (ER: 0.42) p=0.34 (AUC: 0.62)	Not Significant	11.5% vs 9.8% (ER: 1.17) p=0.53 (AUC: 0.61)	Decreased	5.5% vs 13.4% (ER: 0.41) p=0.02 (AUC: 0.72)
CD4+CD45RA-PD1+ (% of CD4 T cells) (PD1+ memory helper T cells)	Not Significant	51% vs 34% (ER: 1.50) p=0.13 (AUC: 0.65)	Increased	50.5% vs 36% (ER: 1.40) p=0.01 (AUC: 0.79)	Not Significant	37.1% vs 39.4% (ER: 0.94) p=0.30 (AUC: 0.61)
Plasma						
ICAM-1 (ng/mL)	Increased	1039 vs 532 (ER: 1.95) p=0.005 (AUC: 0.76)	Not Significant	579 vs 682 (ER: 0.85) p=0.75 (AUC: 0.51)	Not Significant	775 vs 583 (ER: 1.33) p=0.054 (AUC: 0.66)



Supplementary Figure 1. Failure cohort not captured by the classifier. For different random train-test sample splits, different samples would be misclassified. We divided the samples based on the proportion of random splits for which they were misclassified, with 0.8 as the threshold to define the failure cohort. (A) The proportion of random splits misclassified for each sample is represented by a dot within the violin plots. The white circle is the median. A key observation (as would be expected) is that samples from cGvHD subjects prior to their onset of cGVHD (samples drawn at day 100 +/-14 days and 6-months +/- 1 month) were more often misclassified. (B) Peripheral blood stem cell grafts (PBSC) compared to non-PBSC grafts have a higher proportion of random splits misclassified. (C) The failure cohort (denoted as "1") is on average older.

.