

*cancers*



Article

---

# Using Targeted Transcriptome and Machine Learning of Pre- and Post-Transplant Bone Marrow Samples to Predict Acute Graft-versus-Host Disease and Overall Survival after Allogeneic Stem Cell Transplantation

---







Scott D. Rowley, Thomas S. Gunning, Michael Pelliccia, Alexandra Della Pia, Albert Lee, James Behrmann, Ayrton Bangolo, Parul Jandir, Hong Zhang, Sukhdeep Kaur et al.



<https://doi.org/10.3390/cancers16071357>

## Article

# Using Targeted Transcriptome and Machine Learning of Pre- and Post-Transplant Bone Marrow Samples to Predict Acute Graft-versus-Host Disease and Overall Survival after Allogeneic Stem Cell Transplantation

Scott D. Rowley<sup>1,2,\*</sup>, Thomas S. Gunning<sup>3</sup>, Michael Pelliccia<sup>3</sup>, Alexandra Della Pia<sup>2</sup>, Albert Lee<sup>3</sup>, James Behrmann<sup>3</sup>, Ayrton Bangolo<sup>3</sup>, Parul Jandir<sup>3</sup>, Hong Zhang<sup>4</sup>, Sukhdeep Kaur<sup>2,3</sup>, Hyung C. Suh<sup>2,3</sup>, Michele Donato<sup>2,3</sup>, Maher Albitar<sup>4</sup> and Andrew Ip<sup>2,3</sup>

<sup>1</sup> Georgetown University School of Medicine, Washington, DC 20007, USA

<sup>2</sup> John Theurer Cancer Center, Hackensack, NJ 07601, USA; alexandra.dellapia@hmn.org (A.D.P.); sukhdeep.kaur@hmn.org (S.K.); hyung.suh@hmn.org (H.S.); michele.donato@hmn.org (M.D.)

<sup>3</sup> Hackensack Meridian School of Medicine, Nutley, NJ 07110, USA; thomas.gunning@hmn.org (T.S.G.); michael.pelliccia@hmn.org (M.P.); albert.lee@hmn.org (A.L.); james.behrmann@hmn.org (J.B.); ayrton.bangolo@hmn.org (A.B.); parul.jandir@hmn.org (P.J.)

<sup>4</sup> Genomic Testing Cooperative, Irvine, CA 92618, USA; hzhang@georgiasouthern.edu (H.Z.); malbitar@genomictestingcooperative.com (M.A.)

\* Correspondence: sdr62@georgetown.edu



**Citation:** Rowley, S.D.; Gunning, T.S.; Pelliccia, M.; Della Pia, A.; Lee, A.; Behrmann, J.; Bangolo, A.; Jandir, P.; Zhang, H.; Kaur, S.; et al. Using Targeted Transcriptome and Machine Learning of Pre- and Post-Transplant Bone Marrow Samples to Predict Acute Graft-versus-Host Disease and Overall Survival after Allogeneic Stem Cell Transplantation. *Cancers* **2024**, *16*, 1357. <https://doi.org/10.3390/cancers16071357>

Academic Editors: Francesco Bertolini and Miguel A. Sanz

Received: 21 February 2024

Revised: 27 March 2024

Accepted: 28 March 2024

Published: 29 March 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Simple Summary:** Acute graft-versus-host disease (aGvHD) remains a major cause of morbidity and mortality after allogeneic hematopoietic stem cell transplantation (HSCT), occurring to some degree in over 50% of patients and being a direct cause of death in about 20% of patients. This complication occurs even despite a better understanding of donor selection and GvHD prophylaxis regimens. aGvHD is a complex event in which multiple contributing factors are involved. We performed RNA transcriptome analysis of 1408 genes in bone marrow samples obtained before and after transplantation using machine learning to predict the risk of aGvHD and post-transplant survival for a cohort of patients undergoing HSCT. Differential gene expression identified several signaling pathways in the bone marrow microenvironment that may be major regulators of the complex biology of GvHD, and identified targets of intervention to ameliorate the risk of aGvHD and improve patient survival.

**Abstract:** Acute graft-versus-host disease (aGvHD) remains a major cause of morbidity and mortality after allogeneic hematopoietic stem cell transplantation (HSCT). We performed RNA analysis of 1408 candidate genes in bone marrow samples obtained from 167 patients undergoing HSCT. RNA expression data were used in a machine learning algorithm to predict the presence or absence of aGvHD using either random forest or extreme gradient boosting algorithms. Patients were randomly divided into training (2/3 of patients) and testing (1/3 of patients) sets. Using post-HSCT RNA data, the machine learning algorithm selected 92 genes for predicting aGvHD that appear to play a role in PI3/AKT, MAPK, and FOXO signaling, as well as microRNA. The algorithm selected 20 genes for predicting survival included genes involved in MAPK and chemokine signaling. Using pre-HSCT RNA data, the machine learning algorithm selected 400 genes and 700 genes predicting aGvHD and overall survival, but candidate signaling pathways could not be specified in this analysis. These data show that NGS analyses of RNA expression using machine learning algorithms may be useful biomarkers of aGvHD and overall survival for patients undergoing HSCT, allowing for the identification of major signaling pathways associated with HSCT outcomes and helping to dissect the complex steps involved in the development of aGvHD. The analysis of pre-HSCT bone marrow samples may lead to pre-HSCT interventions including choice of remission induction regimens and modifications in patient health before HSCT.

**Keywords:** allogeneic transplantation; acute GvHD; post-transplant survival; transcriptomics; machine learning; signaling pathways

---

## 1. Introduction

Acute graft-versus-host disease (aGvHD) is the leading cause of non-relapse mortality (NRM) in patients after allogeneic hematopoietic stem cell transplantation (allo-HSCT), with 6-month NRM occurring in up to 18% of patients with high-risk aGvHD [1,2]. For several decades, donor selection algorithms and prophylactic post-transplant chemotherapy regimens, with the later addition of calcineurin inhibitors, remained the standard approach to reduce the risk of aGvHD. Despite these routine but imprecise prevention strategies, nearly 40% to 60% of allo-HSCT recipients develop aGvHD, and even patients predicted to be at a low risk of developing severe aGvHD may still develop dire, treatment-refractory, and life-threatening disease [3,4].

The pathophysiology of aGvHD involves the activation of donor T-cells, macrophages, monocytes, and neutrophils, and the condition arises when these donated immune cells attack host tissues (e.g., the skin, gastrointestinal tract, and liver), resulting in potentially life-threatening complications [3,5]. These complications range from mild to severe, with the incidence, severity, and morbidity of aGvHD increasing with greater donor-to-host human leukocyte antigen (HLA) disparity. Historical methods for diagnosing aGvHD include symptom evaluation to assign a clinical grade and tissue biopsy to exclude other causes of pathology. However, these strategies are often difficult to quantify, invasive, and lack specificity [6]. The ability to predict the development of aGvHD may allow for a better selection of patients and donors, and the choice of conditioning and immunosuppression regimens. Previous studies in predicting aGvHD focused on the characteristics and effects of donor cells rather than the host microenvironment [7]. Several studies identified that both donor and host characteristics are associated with an increased risk of aGvHD and, more recently, an effort has been made to identify biomarkers associated with or predictive of the development of aGvHD before the onset of clinical manifestations [8–10]. For example, the Mount Sinai Acute GvHD International Consortium (MAGIC) algorithm probability (MAP) uses two biomarkers measured pre- and post-transplant to detect changes in the gastrointestinal tract as a response biomarker and correlated these findings with NRM [11]. Still, no method using one or a combination of biomarkers has yet been widely adopted for the prediction or characterization of aGvHD [7,12].

Advancements in machine learning are generating interest in developing an integrated, data-driven approach to predict and manage medical events [13,14]. The ability to measure gene expression in various cell populations using transcriptome analysis allows for the exploration of cell pathways involved in various biological events, which is critical as aGvHD is a complex event in which multiple contributing factors are involved. Machine learning employs data-driven statistical modeling approaches that can identify underlying patterns without predefined assumptions. Transcriptome analysis with machine learning has been used to identify gene expression profiles (GEP) including both upregulated and suppressed genes associated with an event. Transcriptome analysis has been used in the analysis of gastrointestinal aGvHD and cutaneous chronic GvHD (cGvHD), identifying potential pathways of T-cell activation that differ from those observed in other inflammatory disease of these organs [15,16].

Given the severity of aGvHD as one of the most fatal complications following allo-HSCT, there is a significant need for better and more objective means for the diagnosis and prediction of aGvHD. Using expression data collected by targeted RNA next-generation sequencing (NGS) of bone marrow (BM) samples obtained before and after transplantation, we explored the potential of RNA data using machine learning to predict aGvHD and overall survival (OS).

## 2. Materials and Methods

### 2.1. Patients

This is a retrospective, single-center study of patients undergoing allo-HSCT between 1 November 2019 and 31 August 2022, allowing for at least 6 months of follow-up for the determination of late-onset aGvHD and survival at the time of database closure (29 May 2023). Patients eligible for this study included all patients with either malignant or nonmalignant hematological disease who underwent allo-HSCT and had BM samples analyzed by RNA gene expression using next-generation sequencing (NGS). Institutional Review Board (IRB) approval for this study was obtained from WCG IRB (Pro2020-1406). The study was conducted under the International Conference on Harmonization Good Clinical Practice Guidelines and according to the Declaration of Helsinki. The requirement for patient informed consent (verbal or written) was waived by the IRB as this project represented a non-interventional study using routinely collected data for secondary research purposes.

Eligibility for transplantation, choice of conditioning and GvHD prophylaxis regimens, and donor and graft sources were not prescribed for enrollment into this study and were per the discretion of the physician caring for the patient. All patients met standard eligibility criteria for allo-HSCT at this center, including age below 80 years, Karnofsky performance status  $\geq 70\%$ , having a readily available HLA matched or mismatched related or unrelated donor (URD), and with appropriate pulmonary, cardiac, hepatic, and renal functions. Stem cell grafts were collected using standard techniques, and no processing of grafts (other than cryopreservation for donor management purposes and red cell depletion of ABO major incompatible bone marrow grafts) was performed. The collection goal for peripheral blood stem cell (PBSC) grafts was  $>4 \times 10^6$  and  $<10 \times 10^6$  CD34+ cells/kg. The target for bone marrow harvesting was  $>3 \times 10^8$  nucleated cells per kilogram, but cell quantities above or below this target were infused without adjustment in number. URD grafts were obtained through the National Marrow Donor Program or similar registries. Per institutional practices, male sex and younger age were given priority in donor selection. Day 0 was defined as the day the cell product infusion was completed.

Patients received pre-transplant conditioning using non-myeloablative (NMA), reduced-intensity (RIC), or myeloablative conditioning (MAC) regimens (Supplemental Table S1) [17]. In general, patients over 60 years of age were conditioned with RIC or NMA regimens. GvHD prophylaxis regimens were determined by the conditioning regimen and donor selected (Supplemental Table S2). Most patients undergoing transplantation using grafts from URD or HLA-matched related donors received a regimen of tacrolimus and methotrexate. A regimen of cyclosporine, sirolimus, and mycophenolate mofetil (MMF) was used for all patients receiving HSC from a non-haploidentical donor after a NMA regimen with single-fraction total body irradiation (TBI, Supplemental Table S1). All patients receiving HSC from related haploidentical donors received a standard regimen of cyclophosphamide, tacrolimus, and MMF (Supplemental Table S2). Two patients treated for aplastic anemia received a GvHD prophylaxis regimen of cyclosporine and methotrexate. Abatacept (Supplemental Table S2) could be added to the GvHD prophylaxis for recipients of haploidentical or unrelated donors. Rabbit anti-thymocyte globulin (ATG, Supplemental Table S2) was added to the regimens for recipients of HLA-matched siblings or unrelated donors receiving busulfan-based MAC regimens and all recipients of reduced-intensity busulfan.

Patients were hospitalized until neutrophil engraftment, control of any infectious complications, and resolution of severe regimen-related complications. Patients were then seen in post-transplantation follow-up at least weekly through day +100 after transplantation, bi-weekly through day +180, and then at least monthly through 12 months or until resolution of cGvHD. All patients received a standard antimicrobial prophylaxis regimen starting on initiation of the transplantation conditioning regimen and post-transplant cytokine support (Supplemental Methods S1). Patients were censored from follow-up for aGvHD on day of event for death, relapse, or onset of cGvHD, or infusion of donor lymphocytes (DLI).

Patients were censored from follow-up of survival on the day of event for death or relapse or last visit if lost to follow-up.

## 2.2. Diagnosis of aGvHD

Consensus criteria were used to diagnose, stage, and grade aGvHD at least weekly through day +100 after transplantation and then at least every other week through to day +180, with scoring maintained in the electronic medical record [18]. Treatment for aGvHD was not prescribed for this study and was determined by the medical team. Time to aGvHD is the day of onset of any grade; the maximal grade and stage of aGvHD were also recorded for this analysis. No attempt was made to analyze these RNA gene expression data stratified relative to the timing of aGvHD onset and the day of the post-transplant bone marrow sampling. For this analysis, the onset, presence, stage, and grade of aGvHD were adjudicated by a single reviewer (SDR) to confirm these details but using MAGIC criteria [19].

## 2.3. Sample Collection

BM samples for disease staging were usually collected per program standard practice within 28 days before initiation of transplant conditioning and at 84–100 days after allo-HSCT with samples outside these times also included.

Fresh BM samples were collected in EDTA tubes and transported at room temperature for initial processing. Pre-transplant was defined as BM samples taken from patients before the start of pre-transplant conditioning. Post-transplant was defined as BM samples taken from patients after allo-HSCT.

## 2.4. RNA Extraction and Next Generation Sequencing (NGS)

BM samples were processed within 72 h of collection. RNA was extracted using an automated Maxwell instrument (Promega, Madison, WI, USA). RNA was first converted to cDNA using random primers. RNA samples were selectively enriched for 1408 cancer-associated genes using the reagents provided in the Illumina<sup>®</sup> TruSight<sup>®</sup> RNA pan-cancer panel (Illumina, San Diego, CA, USA; Supplemental Table S3). Sequencing adapters were ligated into the resulting double-stranded cDNA fragments. The coding regions of the expressed genes were captured from this library using sequence-specific probes to create the final library. Sequencing was performed using the Illumina NovaSeq platform (Illumina, San Diego, CA, USA). Ten million reads per sample were performed in a single run, and the read length was  $2 \times 150$  bp. For any sequence to be accepted, it must show >20% reads with splice junction. Expression levels were measured as fragments per kilobase of transcripts per million.

## 2.5. Machine Learning Algorithm for Predicting aGvHD

The RNA expression data were used in the machine learning algorithm to predict the presence or absence of aGvHD using either random forest or extreme gradient boosting algorithms. Patients were randomly divided into training (2/3 of patients) and validation (1/3 of patients; Supplemental Table S4a) sets. Randomization was independently performed by computer for the pre- and post-transplant cohorts and, accordingly, different patients from each cohort could be randomly assigned to the training and analysis sets. The Entrez symbols of the genes selected by the algorithm were search for their pathway involvement using the KEGG database and software [20–22].

## 2.6. Statistical Analysis

Primary clinical endpoints are the development of grade I-IV aGvHD with censoring for relapse or death, and OS censored for relapse. Patients' characteristics were summarized but not compared between the pre- and post-transplant groups (or between training or analysis subsets). Continuous variables were summarized with median and range and categorical variables were summarized with counts and percentages.

We developed a machine learning algorithm that first selected the relative genes based on the performance of each gene with cross-validation and based on stability measures using statistical significance tests. The selected genes were then used to predict aGvHD or survival with k-fold cross validation procedures ( $k = 12$ ). A naïve Bayesian classifier was constructed on the training of  $k - 1$  subsets and tested on the other resting subset. We applied geometric mean naïve Bayesian (GMB) as the classifier for prediction. The details of this machine learning platform were previously reported [23,24].

### 3. Results

#### 3.1. Clinical Characteristics

Samples for analysis were available for 167 patients (Table 1), with pre-transplant samples available for 132 patients collected at a median of 29 days before transplantation (range, 14–170 days) and post-transplant samples available for 119 patients collected at a median of 84 days (range, 29–141 days). Patient and donor demographics, transplant diagnoses, conditioning and GvHD regimens, and use of ATG or abatacept are summarized (Table 1, Supplemental Table S4a–c). The median follow-up for the entire group at time of analysis was 344 days (range, 7–925 days). Thirty-six patients suffered disease relapse at a median of 87.5 days (range, 27–718 days) after transplantation. Fifty-nine patients expired of disease relapse or complications of treatment at a median of 165 days (range, 7–560 days).

**Table 1.** Patient characteristics.

Characteristics	All Patients N = 167 n (%)	Pre-Transplant Patients N = 132 n (%)	Post-Transplant Patients N = 119 n (%)
Recipient age, median years (range)	63.0 (20.8–79.0)	64.0 (20.8–79.0)	63.7 (20.8–79.0)
Recipient sex			
Male	89 (53)	71 (54)	62 (52)
Indication for allo-HSCT			
ALL	20 (12)	15 (11)	15 (13)
AML	57 (34)	40 (30)	47 (39)
CML	3 (2)	2 (2)	3 (3)
MDS	44 (26)	37 (28)	31 (26)
Myelofibrosis/CMML	32 (19)	27 (20)	19 (16)
SAA	4 (2)	4 (3)	1 (1)
NHL	7 (4)	7 (5)	3 (3)
Graft source			
Bone marrow	130 (78)	30 (23)	23 (19)
Peripheral blood	37 (22)	102 (77)	96 (81)
HLA compatibility			
Unrelated donor HLA match	84 (50)	64 (48)	65 (55)
Unrelated donor HLA mismatch	17 (10)	13 (10)	13 (11)
HLA matched related donor	14 (8)	11 (8)	9 (8)
Related donor, haploidentical	52 (31)	44 (33)	32 (27)
Donor age, median years (range)	28.3 (14.0–63.7)	28.1 (17.6–63.7)	28.0 (14.0–59.0)
Donor age < 35 years	120 (72)	98 (74)	89 (75)
Donor sex			
Male	109 (65)	82 (62)	77 (65)
Conditioning regimen			
Myeloablative	43 (26)	30 (23)	33 (28)
Non-myeloablative	47 (28)	41 (31)	31 (26)
Reduced intensity	77 (46)	61 (46)	55 (46)



Table 1. Cont.

Characteristics	All Patients N = 167 n (%)	Pre-Transplant Patients N = 132 n (%)	Post-Transplant Patients N = 119 n (%)
aGvHD prophylaxis regimen			
PtCy	73 (44)	62 (47)	49 (41)
TacMTX	75 (45)	55 (42)	58 (49)
RapaCspMMF	19 (11)	15 (11)	12 (10)
Addition of abatacept	21 (13)	18 (14)	16 (13)
Addition of anti-thymocyte globulin	38 (23)	26 (20)	31 (26)
Diagnosed with aGvHD			
Stage 1–4	109 (65)	87 (66)	80 (67)
Stage 3–4	7 (4)	6 (5)	1 (1)
Site of aGvHD			
Gastrointestinal	65 (39)	51 (71)	49 (40)
Lower	17 (10)	12 (30)	10 (8)
Upper	48 (29)	39 (41)	39 (32)
Liver	5 (3)	4 (3)	4 (3)
Skin	63 (38)	54 (41)	51 (43)

Allo—allogeneic; ALL—acute lymphoblastic leukemia; AML—acute myeloid leukemia; CML—chronic myeloid leukemia; CMML—chronic myelomonocytic leukemia; CSPMMF—cyclosporine and mycophenolate mofetil; MDS—myelodysplastic syndromes; NHL—non-Hodgkin’s lymphoma; PtCy—post-transplant cyclophosphamide; RapaCSPMMF—rapamycin, cyclosporine, and mycophenolate mofetil; RIC—reduced intensity conditioning; SAA—severe aplastic anemia; TacMTX—tacrolimus and methotrexate.

### 3.2. Prediction of aGvHD Using Post-Transplant Samples

Of the 119 patients with post-transplant BM samples available, 80 patients developed aGvHD of any stage (Table 1) at a median of 37.5 days. Using post-transplant BM RNA data, the machine learning algorithm selected 92 genes (Table 2) for predicting aGvHD in the training set with an AUC of 0.999 (95% confidence interval [CI], 0.992–1.007;  $p = 0.03$ ), 100% sensitivity, and 98.1% specificity (Figure 1a). In the validation set, the machine learning algorithm showed an AUC of 0.721 (95% CI, 0.542–0.900;  $p = 0.03$ ), 76.9% sensitivity, and 63% specificity using a cut-off score of 0.423 (Figure 1b). The genes selected by the algorithm for predicting aGvHD appear to play roles involved in PI3K/AKT signaling, MAPK signaling, FOXO signaling, and cancer-associated microRNA.

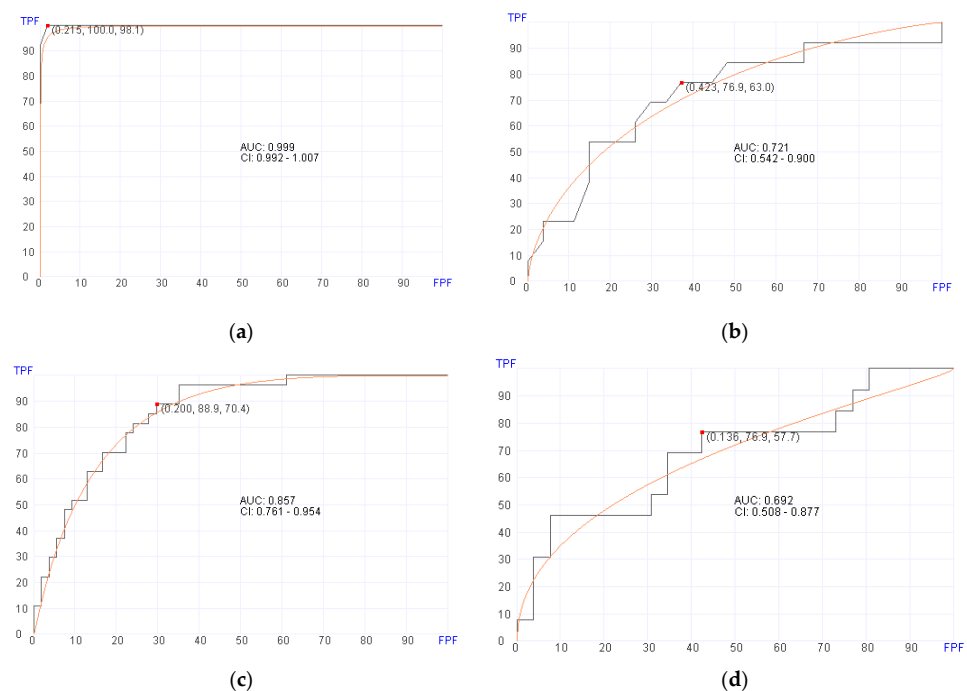
Table 2. 92 Genes Predicting aGvHD in Post-transplant Samples.

92 Genes Predicting GvHD			
1–23	24–46	47–69	70–92
DUSP2	CDKN1A	NEURL1	SUZ12
CD22	TFRC (CD71)	TNFRSF17 (BCMA)	TRIM33
FLNA	DLL3	BCL7A	CDK9
PAX8	SSBP2	YTHDF2	FLYWCH1
ARHGEF12	TRAF3	KIF5B	HIST1H2BC
AKAP9	PSIP1	IRS1	MAPK1
DLL4	43717SEPT9	DGKZ	RAC2
AIP	SPTBN1	CENPU	TCF7L2

Table 2. Cont.

92 Genes Predicting GvHD			
1–23	24–46	47–69	70–92
CDC14B	HIST1H2AC	STIL	USP42
FOXO3	TFDP1	XKR3	FGFR1OP
EGR4	TRAF5	CCT6B	MTCP1
MUTYH	BACH2	CD28	PTPRO
SS18L1	TNFRSF10D	OLIG1	SH3D19
PRKCG	SLC45A3	CCND2	CTDSP2
HOOK3	NACA	GID4	ID3
TCEA1	ASPH	STYK1	SMAP1
UBE2C	ZBTB16	ATF3	STL
FIGF	EPHA2	FGF9	TAL1
TOP1	APOD	ZNF703	DNMT3A
DTX1	KAT2B	AKAP12	IKBKE
TNF	ETV5	PTCRA	IKZF3
CCNE1	FGF13	SMAD6	AKT3
BAIAP2L1	FLT3LG	DNAJB1	HSPA4

Shown are the 92 genes identified in post-transplant marrow samples that associated with the development of aGvHD. Genes are listed in order of expression.



**Figure 1.** Receiver operator characteristics (ROC) curves for predicting aGvHD in pre- and post-transplant patient cohorts. (a) Post-transplant aGvHD prediction using 92 genes (AUC = 0.999,  $p = 0.03$ ) in the training set. (b) Post-transplant aGvHD prediction (AUC = 0.721,  $p = 0.03$ ) in the validation set. (c) Pre-transplant aGvHD prediction using 400 genes (AUC = 0.857,  $p = 0.02$ ) in the training set. (d) Pre-transplant aGvHD prediction using (AUC = 0.692,  $p = 0.02$ ) in the validation set.



### 3.3. Prediction of aGvHD Using Pre-Transplant Samples

Of the 132 patients with pre-transplant BM samples available, 87 patients developed aGvHD of any stage (Table 1). Using pre-transplant BM RNA data, the machine learning algorithm selected 400 genes (Supplemental Table S5) for predicting aGvHD in the training set with an AUC of 0.857 (95% CI, 0.761–0.954;  $p = 0.02$ ), 88.9% sensitivity, and 70.4% specificity (Figure 1c). In the validation set, the machine learning algorithm showed an AUC of 0.692 (95% CI, 0.508–0.877;  $p = 0.02$ ), 76.9% sensitivity, and 57.7% specificity using a cutoff of 0.136 (Figure 1d).

### 3.4. Prediction of Overall Survival (OS) Using Post-Transplant Samples

Of the 119 patients with post-transplant BM samples available, 87 were alive at a median follow up of 14 months at the time of data analysis. Using post-transplant BM RNA data, the machine learning algorithm selected 20 genes (Table 3) for predicting OS in the training set with an AUC of 0.918 (95% CI, 0.860–0.975;  $p = 0.02$ ), 86.8% sensitivity, and 89.5% specificity (Figure 2a). In the validation set, the machine learning algorithm showed an AUC of 0.716 (95% CI, 0.565–0.867;  $p = 0.02$ ), 73.1% sensitivity, and 66.7% specificity using a cut-off score at 0.676 (Figure 2b). The genes that are selected by the algorithm for predicting survival included genes involved in MAPK signaling pathway and chemokine signaling.

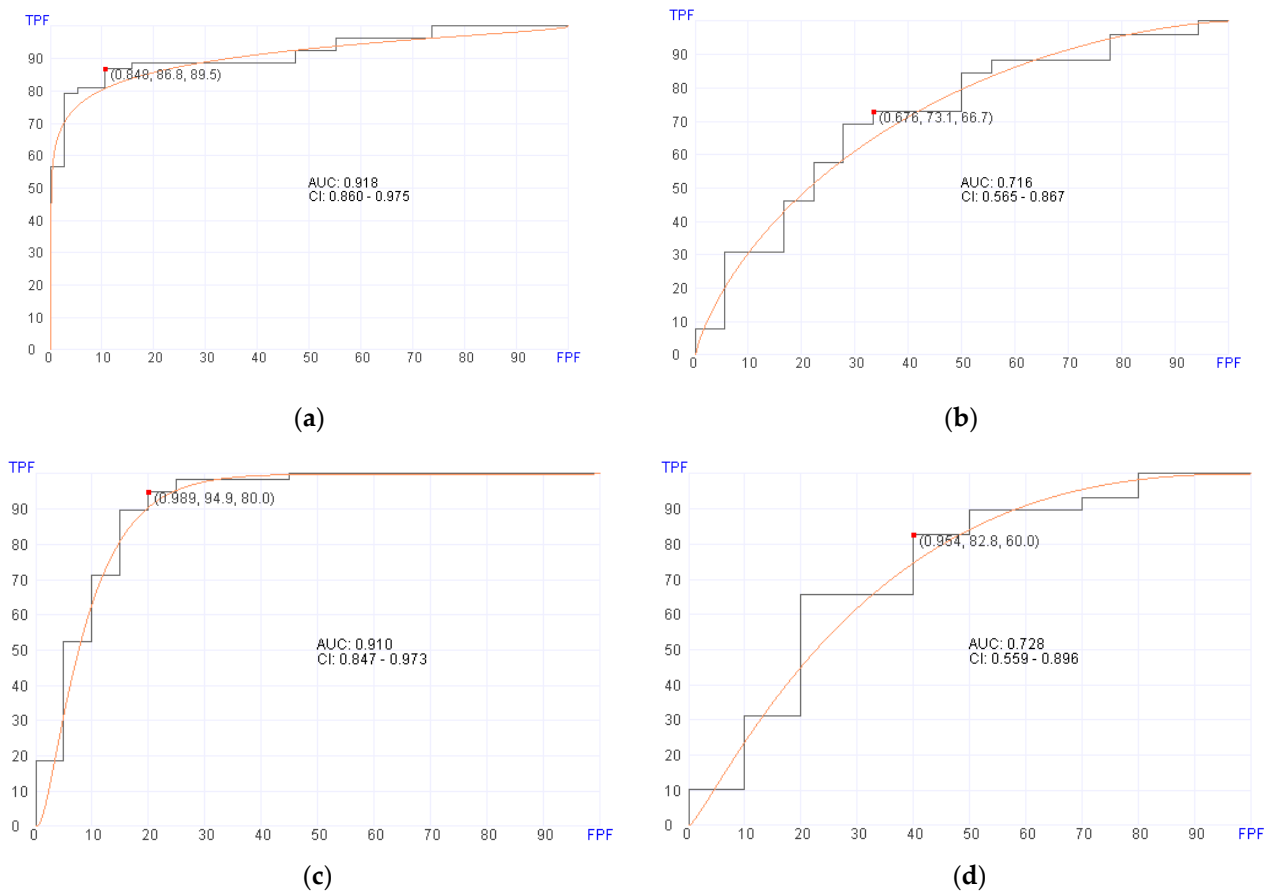
**Table 3.** 20 Genes Predicting Survival in Post-transplant Samples.

Genes Predicting Survival	
1–10	11–20
ATIC	TGFBI
PLAG1	BRSK1
CD36	KIT (CD117)
HSP90AB1	MSH6
DNMT1	HIST1H1D
WDR1	HEY1
CDC14A	FOXO1
MALT1	PRKCA
SP3	CCNB1IP1
MAP3K14	FANCC

Shown are the 20 genes identified in post-transplant marrow samples that associated with the overall survival. Genes are listed in order of expression.

### 3.5. Prediction of Overall Survival (OS) Using Pre-Transplant Samples

Using pre-transplant BM RNA data, the machine learning algorithm selected 700 genes (Supplemental Table S6) for predicting aGvHD in the training set with an AUC of 0.910 (95% CI, 0.847–0.973;  $p = 0.02$ ), 94.9% sensitivity, and 80% specificity (Figure 2c). In the validation set, the machine learning algorithm showed an AUC of 0.728 (95% CI, 0.559–0.896;  $p = 0.02$ ), 82.8% sensitivity, and 60% specificity using a cut-off score at 0.954 (Figure 2d).



**Figure 2.** Receiver operator characteristics (ROC) curves for predicting overall survival in pre- and post-transplant patient cohorts. (a) Post-transplant survival prediction using 20 genes (AUC = 0.910,  $p = 0.02$ ) in the training set. (b) Post-transplant survival prediction using 20 genes (AUC = 0.716,  $p = 0.02$ ) in the validation set. (c) Pre-transplant survival prediction using 700 genes (AUC = 0.910,  $p = 0.02$ ) in the training set. (d) Pre-transplant survival prediction using 700 genes (AUC = 0.728,  $p = 0.02$ ) in the validation set.

#### 4. Discussion

In this study, we analyzed gene expression profiles of BM samples that were obtained before and after transplantation for patients undergoing allo-HSCT. Patients were randomly assigned into training and testing cohorts independently for the pre- and post-HSCT samples. Using pre- and post-transplant BM samples, we developed machine learning algorithms that selected for a number of genes to predict aGvHD and OS.

The expressions of more than 1400 genes were used in this study to evaluate bone marrow environment in pre- and post-transplant samples. Using a machine learning algorithm is necessary to adjust for the multiple variables that may contribute to the prediction of the presence or absence of aGvHD and eliminate statistically insignificant markers. A Bayesian approach is specifically appropriate to use when the number of cases is limited. Furthermore, a Bayesian approach allows us to define the specific biomarkers that are relevant for the classification so the classification is less of a “blackbox” as compared with other classifiers such as random forest or extreme gradient booster. The expression of only 20 genes were adequate for the prediction of survival using post-transplant samples, reflecting significant changes in bone marrow that are detrimental to the survival of the patients. Similarly, for predicting aGvHD in post-transplant bone marrow samples, only 92 genes are needed. In contrast, for predicting aGvHD and survival in pre-transplant bone marrow samples, 400 genes and 700 genes are needed, respectively. This suggests that

pre-transplant, there are more bone marrow microenvironment factors that play a role in future development of aGvHD and subsequent survival.

The post-HSCT samples showed a high correlation with the presence of aGvHD. These samples were mostly obtained after the onset of aGvHD and, therefore, we cannot propose that this analysis serves as a biomarker predictive for the development of this complication of HSCT. Yet identification of the involved pathways may facilitate the development of GvHD prophylaxis regimens beyond the currently widely used calcineurin-based prophylaxis regimens that could be effective in suppressing the onset of aGvHD, facilitating the development of tolerance, and reducing the risk of off-target toxicities. This analysis may be equally valuable in dissecting the pathways involved with aGvHD, leading to more-specific non-steroid treatments including targeted treatments to manage steroid-resistant aGvHD (or cGvHD). Examples of such targeted therapies includes inhibition of the Janus kinase (JAK) [25–27], Bruton’s tyrosine kinase (BTK) [28–30], and Rho-associated coiled-coil-containing protein kinase 2 (ROCK2) [31] signaling pathways, each of which now have FDA-approved medications for management of steroid-refractory acute and/or chronic GvHD [32]. Our findings are in agreement with other reports such as the association of microRNA with acute and chronic GvHD [33].

The results of the pre-HSCT sample analysis also show that transcriptome analysis of the bone marrow microenvironment is predictive of aGvHD and OS, although we could not define specific signaling pathways. With further investigation, it may be possible to use transcriptome analysis to develop biomarker(s) predictive of the development of aGvHD and OS, allowing for modifications in the planned transplant treatment plan, and thereby improving transplant outcomes [4]. We hypothesize that transcriptome analysis early in the course of the disease could lead to modifications in patient care during initial remission induction and consolidation cycles before referral for transplantation. We further hypothesize that our findings may correspond to the immune microenvironment possibly being influenced by microbiome effects on transplant outcomes, for example, with the potential for prevention of aGvHD [34,35]. The pre-transplant BM specimens would be affected by therapies given in the control of disease in anticipation of allo-HSCT.

Our analysis of OS using post-transplant samples is complicated by the presence of aGvHD in a large proportion of the patient cohort, and we cannot ascertain, using this dataset, if the signaling pathways associated with OS are distinct from or overlap the signaling pathways associated with aGvHD. We also did not attempt to correlate our analysis with the onset or severity of cGvHD. That transcriptome analysis of the pre-transplant samples also predicts OS demands further, in-depth analysis of the patient populations being referred for transplantation, hopefully leading to improvements in the transplant process.

Numerous investigators are exploring the gene expression profile associated with the immunological GvHD and graft-versus-disease (GvL) events of HSCT. Such studies led to clinical studies of JAK, BTK, and ROCK2 inhibitors, resulting in FDA approval for these therapies [22–29]. Most of the studies, however, combine analysis of gene expression profile of certain lymphocyte populations to explore the biology of immunological reconstitution after HSCT, such as the study by McCurdy et al. of patients receiving PTCy for GvHD prophylaxis using machine learning and RNASeq analysis of blood lymphocyte subsets at day +28 after transplantation, which found 56 differentially expressed genes (DEGs) in regulatory T cells in patients who developed aGvHD [36]. Our analysis, especially of pre-transplant samples, may identify signaling pathways that will lead to more in-depth analysis of the immunological events occurring during HSCT.

The primary limitation of our study is the single-center, retrospective design and the relatively small study populations in both the training and testing cohorts. The primary advantage in this study is the large variation in patient and treatment characteristics, showing that this approach may be valid over a wide range of patients. The strong correlation across a number of patient variables including diagnosis, donor type, transplant conditioning regimen, and GvHD prophylaxis emphasizes the strength of our findings. Furthermore, our

analysis of over 1400 genes is not restricted to specific signaling pathways that have been the subject of study in the previously reported analyses of acute and chronic GvHD. While we do not propose that eligibility for transplant should be based at this time on testing bone marrow samples using the approach used in this study, the current study points out the importance of the pre-transplant bone marrow microenvironment in the potential of developing aGvHD and overall survival. Further studies are necessary and appropriate to explore the bone marrow microenvironment and to improve the management of aGvHD.

## 5. Conclusions

In conclusion, our study shows that targeted transcriptome analysis of pre- and post-transplant BM samples can predict aGvHD and OS with relatively high accuracy when a large number of genes are used. Although the accuracy of this prediction is higher when post-transplant transcriptomic data are used, the pre-transplant BM microenvironment is very important and relevant for the future development of aGvHD and for overall survival. This confirms that both the host BM microenvironment and the donor cells may play a significant role in the development of aGvHD and OS in patients undergoing allo-HSCT. Although preliminary, our study demonstrates expression data collected by targeted RNA NGS using machine learning can predict aGvHD and survival. Future studies are needed to validate our findings.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cancers16071357/s1>, Method S1: Post-transplant Supportive Care Policies and Procedures; Table S1: Description of Conditioning Regimens; Table S2: GvHD Prophylaxis Regimens; Table S3: List of Genes Included in the Analysis of Bone Marrow Samples; Table S4a: Patient and Donor Demographics; Table S4b: Transplant Source and Conditioning and GvHD Regimens; Table S4c: aGvHD and Survival; Table S5: 400 Genes Predicting aGvHD in Pre-transplant Samples; Table S6: 700 Genes Predicting Overall Survival in Pre-transplant Samples

**Author Contributions:** Conceptualization, S.D.R. and M.A.; methodology, M.A. and H.Z.; formal analysis, M.A. and H.Z.; investigation, S.D.R., M.P. and T.S.G.; resources, S.D.R., S.K., H.C.S. and M.D.; data curation, S.D.R., M.P. and T.S.G.; writing—original draft preparation, S.D.R., T.S.G., M.P., A.D.P., A.L., J.B., A.B. and P.J.; writing—review and editing, all authors.; visualization, S.D.R., M.P., T.S.G., M.A. and H.Z.; supervision, S.D.R. and A.I. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** This study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of Hackensack Meridian Health (protocol code Pro2020-1406).

**Informed Consent Statement:** Patient consent was waived as this project represented a non-interventional study utilizing routinely collected data for secondary research purposes.

**Data Availability Statement:** The complete datasets used and/or analyzed during this study are available from the corresponding author upon request. Requests can be made through the corresponding author or directly to representatives of Hackensack Meridian Health (Scott D. Rowley; Email: sdr62@georgetown.edu).

**Conflicts of Interest:** T.S.G., M.P., A.D.P., A.L., J.B., A.B., P.J., S.K. and M.D. declare no conflict of interest. M.A. and H.Z. work and own stocks in a diagnostic company that offers RNA sequencing using artificial intelligence. S.D.R.: Consultancy: ReAlta Life Sciences; Advisory Board: SIRPant Immunotherapeutics; Stock: COTA, Genomic Testing Cooperative. H.C.S.: Honoraria: Kite Pharma; Consultancy: Kite Pharma. A.I.: Consultancy, Honoraria, and/or Speaker Bureau: TG Therapeutics and AstraZeneca; Advisory Board: Secura Bio.

## References

1. Ramdial, J.L.; Mehta, R.S.; Saliba, R.M.; Alousi, A.M.; Bashir, Q.; Hosing, C.; Kebriaei, P.; Olson, A.L.; Oran, B.; Qazilbash, M.H.; et al. Acute graft-versus-host-disease is the foremost cause of late nonrelapse mortality. *Bone Marrow Transplant.* **2021**, *56*, 2005–2012. [[CrossRef](#)] [[PubMed](#)]
2. Hartwell, M.J.; Ozbek, U.; Holler, E.; Renteria, A.S.; Major-Monfried, H.; Reddy, P.; Aziz, M.; Hogan, W.J.; Ayuk, F.; Efebera, Y.A.; et al. An early-biomarker algorithm predicts lethal graft-versus-host disease and survival. *JCI Insight* **2017**, *2*, e89798, Erratum in *JCI Insight* **2018**, *3*, e124015. [[CrossRef](#)] [[PubMed](#)]
3. Choi, S.W.; Reddy, P. Current and emerging strategies for the prevention of graft-versus-host disease. *Nat. Rev. Clin. Oncol.* **2014**, *11*, 536–547. [[CrossRef](#)] [[PubMed](#)]
4. Rimando, J.; McCurdy, S.R.; Luznik, L. How I prevent GVHD in high-risk patients: Posttransplant cyclophosphamide and beyond. *Blood* **2023**, *141*, 49–59. [[CrossRef](#)] [[PubMed](#)]
5. Ghimire, S.; Weber, D.; Mavin, E.; Wang, X.N.; Dickinson, A.M.; Holler, E. Pathophysiology of GvHD and other HSCT-related major complications. *Front. Immunol.* **2017**, *8*, 79. [[CrossRef](#)] [[PubMed](#)]
6. Srinagesh, H.K.; Levine, J.E.; Ferrara, J.L.M. Biomarkers in acute graft-versus-host disease: New insights. *Ther. Adv. Hematol.* **2019**, *10*, 2040620719891358. [[CrossRef](#)] [[PubMed](#)]
7. Zhao, X.S.; Huang, X.J. Seeking biomarkers for acute graft-versus-host disease: Where we are and where we are heading? *Biomark. Res.* **2019**, *7*, 17. [[CrossRef](#)] [[PubMed](#)]
8. Harris, A.C.; Ferrara, J.L.; Levine, J.E. Advances in predicting acute GVHD. *Br. J. Haematol.* **2013**, *160*, 288–302. [[CrossRef](#)] [[PubMed](#)]
9. Jagasia, M.; Arora, M.; Flowers, M.E.; Chao, N.J.; McCarthy, P.L.; Cutler, C.S.; Urbano-Ispizua, A.; Pavletic, S.Z.; Haagenson, M.D.; Zhang, M.-J.; et al. Risk Factors for acute GVHD and survival after hematopoietic cell transplantation. *Blood* **2012**, *119*, 296–307. [[CrossRef](#)]
10. Paczesny, S. Discovery and validation of graft-versus-host disease biomarkers. *Blood* **2013**, *121*, 585–594. [[CrossRef](#)]
11. Srinagesh, H.K.; Ozbek, U.; Kapoor, U.; Ayuk, F.; Aziz, M.; Ben-David, K.; Choe, H.K.; DeFilipp, Z.; Etra, A.; Grupp, S.A.; et al. The MAGIC algorithm probability is a validated response biomarker of treatment of acute graft-versus-host disease. *Blood Adv.* **2019**, *3*, 4034–4042. [[CrossRef](#)] [[PubMed](#)]
12. Tang, S.; Chappell, G.T.; Mazzoli, A.; Tewari, M.; Choi, S.W.; Wiens, J. Predicting acute graft-versus-host disease using machine learning and longitudinal vital sign data from electronic health records. *JCO Clin. Cancer Inform.* **2020**, *4*, 128–135. [[CrossRef](#)]
13. Reel, P.S.; Reel, S.; Pearson, E.; Trucco, E.; Jefferson, E. Using machine learning approaches for multi-omics data analysis: A review. *Biotechnol. Adv.* **2021**, *49*, 107739. [[CrossRef](#)]
14. Sen, S.K.; Green, E.D.; Hutter, C.M.; Craven, M.; Ideker, T.; Di Francesco, V. Opportunities for basic, clinical, and bioethics research at the intersection of machine learning and genomics. *Cell Genom.* **2024**, *4*, 100466. [[CrossRef](#)] [[PubMed](#)]
15. Khandelwal, P.; Louder, D.T.; Bartlett, A.; Haberman, Y.; Jegga, A.G.; Ghandikota, S.; Koo, J.; Luebbering, N.; Leino, D.; Abdullah, S.; et al. Transcriptome analysis in acute gastrointestinal graft-versus-host disease reveals a unique signature in children and shared biology with pediatric inflammatory bowel disease. *Haematologica* **2023**, *108*, 1803–1816. [[CrossRef](#)] [[PubMed](#)]
16. Zouali, H.; Lemasson, J.; Calugareanu, A.; Battail, C.; Michonneau, D.; le Buanec, H.; Grolleau, C.; Cassius, C.; Robin, M.; Merandet, M.M.; et al. RNA sequencing of chronic GVHD skin lesions defines shared and unique inflammatory pathways characterizing lichen planus and morphea. *Blood Adv.* **2022**, *6*, 2805–2811. [[CrossRef](#)] [[PubMed](#)]
17. Bacigalupo, A.; Ballen, K.; Rizzo, D.; Giral, S.; Lazarus, H.; Ho, V.; Apperley, J.; Slavin, S.; Pasquini, M.; Sandmaier, B.M.; et al. Defining the intensity of conditioning regimens: Working definitions. *Biol. Blood Marrow Transplant.* **2009**, *15*, 1628–1633. [[CrossRef](#)] [[PubMed](#)]
18. Przepiorka, D.; Weisdorf, D.; Martin, P.; Klingemann, H.G.; Beatty, P.; Hoes, J.; Thomas, E.D. 1994 Consensus Conference on Acute GVHD Grading. *Bone Marrow Transplant.* **1995**, *15*, 825–828. [[PubMed](#)]
19. Harris, A.C.; Young, R.; Devine, S.; Hogan, W.J.; Ayuk, F.; Bunworasate, U.; Chanswangphuwana, C.; Efebera, Y.A.; Holler, E.; Litzow, M.; et al. International, multicenter standardization of acute graft-versus-host disease clinical data collection: A report from the Mount Sinai Acute GVHD International Consortium. *Biol. Blood Marrow Transplant.* **2016**, *22*, 4–10. [[CrossRef](#)]
20. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [[CrossRef](#)]
21. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **2019**, *28*, 1947–1951. [[CrossRef](#)] [[PubMed](#)]
22. Kanehisa, M.; Furumichi, M.; Sato, Y.; Kawashima, M.; Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **2023**, *51*, D587–D592. [[CrossRef](#)]
23. Zhang, H.; Qureshi, M.A.; Wahid, M.; Charifa, A.; Ehsan, A.; Ip, A.; De Dios, I.; Ma, W.; Sharma, I.; McCloskey, J.; et al. Differential diagnosis of hematologic and solid tumors using targeted transcriptome and artificial intelligence. *Am. J. Pathol.* **2023**, *193*, 51–59. [[CrossRef](#)] [[PubMed](#)]
24. Albitar, M.; Zhang, H.; Pecora, A.; Waintraub, S.; Graham, D.; Hellmann, M.; McNamara, D.; Charifa, A.; De Dios, I.; Ma, W.; et al. Homologous recombination abnormalities associated with *BRCA1/2* mutations as predicted by machine learning of targeted next-generation sequencing data. *Breast Cancer* **2023**, *17*, 11782234231198979. [[CrossRef](#)] [[PubMed](#)]



25. Jagasia, M.; Perales, M.A.; Schroeder, M.A.; Ali, H.; Shah, N.N.; Chen, Y.-B.; Fazal, S.; Dawkins, F.W.; Arbushites, M.C.; Tian, C.; et al. Ruxolitinib for the treatment of steroid-refractory acute GVHD (REACH1): A multicenter, open-label phase 2 trial. *Blood* **2020**, *135*, 1739–1749. [[CrossRef](#)] [[PubMed](#)]
26. Zeiser, R.; Polverelli, N.; Ram, R.; Hashmi, S.K.; Chakraverty, R.; Middeke, J.M.; Musso, M.; Giebel, S.; Uzay, A.; Langmuir, P.; et al. Ruxolitinib for glucocorticoid-refractory chronic graft-versus-host disease. *N. Engl. J. Med.* **2021**, *385*, 228–238. [[CrossRef](#)] [[PubMed](#)]
27. Choi, J.; Cooper, M.L.; Alahmari, B.; Ritchey, J.; Collins, L.; Holt, M.; DiPersio, J.F. Pharmacologic blockade of JAK1/JAK2 reduces GvHD and preserves the graft-versus-leukemia effect. *PLoS ONE* **2014**, *9*, e109799. [[CrossRef](#)]
28. Dubovsky, J.A.; Flynn, R.; Du, J.; Harrington, B.K.; Zhong, Y.; Kaffenberger, B.; Yang, C.; Towns, W.H.; Lehman, A.; Johnson, A.J.; et al. Ibrutinib treatment ameliorates murine chronic graft-versus-host disease. *J. Clin. Investig.* **2014**, *124*, 4867–4876. [[CrossRef](#)] [[PubMed](#)]
29. Schutt, S.D.; Fu, J.; Nguyen, H.; Bastian, D.; Heinrichs, J.; Wu, Y.; Liu, C.; McDonald, D.G.; Pidala, J.; Yu, X.-Z. Inhibition of BTK and ITK with ibrutinib is effective in the prevention of chronic graft-versus-host disease in mice. *PLoS ONE* **2015**, *10*, e0137641. [[CrossRef](#)]
30. Waller, E.K.; Miklos, D.; Cutler, C.; Arora, M.; Jagasia, M.H.; Pusic, I.; Flowers, M.E.; Logan, A.C.; Nakamura, R.; Chang, S.; et al. Ibrutinib for chronic graft-versus-host disease after failure of prior therapy: 1-year update of a phase 1b/2 study. *Biol. Blood Marrow Transplant.* **2019**, *25*, 2002–2007. [[CrossRef](#)]
31. Cutler, C.; Lee, S.J.; Arai, S.; Rotta, M.; Zoghi, B.; Lazaryan, A.; Ramakrishnan, A.; DeFilipp, Z.; Salhotra, A.; Chai-Ho, W.; et al. Belumosudil for chronic graft-versus-host disease after 2 or more prior lines of therapy: The ROCKstar Study. *Blood* **2021**, *138*, 2278–2289, Erratum in *Blood* **2022**, *139*, 1772. [[CrossRef](#)] [[PubMed](#)]
32. Martini, D.J.; Chen, Y.B.; DeFilipp, Z. Recent FDA approvals in the treatment of graft-versus-host disease. *Oncologist* **2022**, *27*, 685–693. [[CrossRef](#)] [[PubMed](#)]
33. Pitea, M.; Canale, F.A.; Porto, G.; Verduci, C.; Utano, G.; Policastro, G.; Alati, C.; Santoro, L.; Imbalzano, L.; Martino, M. The Role of MicroRNA in Graft-Versus-Host-Disease: A Review. *Genes* **2023**, *14*, 1796. [[CrossRef](#)] [[PubMed](#)]
34. Shono, Y.; van den Brink, M. Gut microbiota injury in allogeneic haematopoietic stem cell transplantation. *Nat. Rev. Cancer* **2018**, *18*, 283–295. [[CrossRef](#)] [[PubMed](#)]
35. Rashidi, A.; Ebadi, M.; Rehman, T.U.; Elhousseini, H.; Kazadi, D.; Halaweish, H.; Khan, M.H.; Hoeschen, A.; Cao, Q.; Luo, X.; et al. Potential of fecal microbiota transplantation to prevent acute GVHD: Analysis from a phase II trial. *Clin. Cancer Res.* **2023**, *29*, 4920–4929. [[CrossRef](#)]
36. McCurdy, S.R.; Radojicic, V.; Tsai, H.L.; Vulic, A.; Thompson, E.; Ivcevic, S.; Kanakry, C.G.; Powell, J.D.; Lohman, B.K.; Adom, D.; et al. Signatures of GVHD and relapse after posttransplant cyclophosphamide revealed by immune profiling and machine learning. *Blood* **2022**, *139*, 608–623. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.