

Machine learning–based scoring models to predict hematopoietic stem cell mobilization in allogeneic donors

Jingyu Xiang,¹ Min Shi,^{2,3} Mark A. Fiala,¹ Feng Gao,⁴ Michael P. Rettig,¹ Geoffrey L. Uy,¹ Mark A. Schroeder,¹ Katherine N. Weilbaecher,¹ Keith E. Stockerl-Goldstein,¹ Shamim Mollah,^{2,3} and John F. DiPersio¹

¹Division of Oncology, Department of Medicine, ²Department of Genetics, ³Institute for Informatics, and ⁴Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine, St. Louis, MO

Key Points

- A single-institution experience with G-CSF and five alternative PB mobilization strategies in related allogeneic donors.
- Machine learning models were established to predict G-CSF–induced PB mobilization in allogeneic donors with an accuracy of up to 83%.

Mobilized peripheral blood has become the primary source of hematopoietic stem cells for both autologous and allogeneic stem cell transplantation. Granulocyte colony-stimulating factor (G-CSF) is currently the standard agent used in the allogeneic setting. Despite the high mobilization efficacy in most donors, G-CSF requires 4–5 days of daily administration, and a small percentage of the donors fail to mobilize an optimal number of stem cells necessary for a safe allogeneic stem cell transplant. In this study, we retrospectively reviewed 1361 related allogeneic donors who underwent stem cell mobilization at Washington University. We compared the standard mobilization agent G-CSF with five alternative mobilization regimens, including GM-CSF, G-CSF+GM-CSF, GM-CSF + Plerixafor, Plerixafor and BL-8040. Cytokine-based mobilization strategies (G-CSF or in combination with GM-CSF) induce higher CD34 cell yield after 4–5 consecutive days of treatment, while CXCR4 antagonists (plerixafor and BL-8040) induce significantly less but rapid mobilization on the same day. Next, using a large dataset containing the demographic and baseline laboratory data from G-CSF–mobilized donors, we established machine learning (ML)–based scoring models that can be used to predict patients who may have less than optimal stem cell yields after a single leukapheresis session. To our knowledge, this is the first prediction model at the early donor screening stage, which may help identify allogeneic stem cell donors who may benefit from alternative approaches to enhance stem cell yields, thus ensuring safe and effective stem cell transplantation.

Introduction

Peripheral blood (PB) has largely replaced bone marrow as the predominant source of hematopoietic stem cells (HSCs) for hematopoietic stem cell transplantation (HSCT).¹ In healthy donors, HSC mobilization is generally induced by 4–5 days consecutive daily doses of 10 µg/kg subcutaneously (SC) of granulocyte colony-stimulating factor (G-CSF).² While G-CSF is effective and is the standard of care for stem cell mobilization in the allogeneic setting, other mobilization strategies such as granulocyte-macrophage colony-stimulating factor (GM-CSF), CXCR4 antagonists or the combination of both cytokine and CXCR4 antagonists have been evaluated in clinical trials to assess their efficacy in stem cell mobilization.³ Plerixafor, a selective CXCR4 antagonist that reversibly competes and binds to CXCR4, is one of the first agents approved for stem cell mobilization in the autologous setting.^{4,5} Other peptide CXCR4 antagonists, such as BL-8040 (Motixafortide), also induce rapid stem cell mobilization in healthy

Submitted 3 May 2021; accepted 16 July 2021; prepublished online on *Blood Advances* First Edition 23 September 2021; final version published online 25 March 2022. DOI 10.1182/bloodadvances.2021005149.

For data sharing, contact the corresponding author: jdpersi@wustl.edu.

The full-text version of this article contains a data supplement.

© 2022 by The American Society of Hematology. Licensed under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0), permitting only noncommercial, nonderivative use with attribution. All other rights reserved.

volunteers and allogeneic stem cell donors.⁶ This report summarizes one of the largest single-institution experiences with stem cell mobilization of related allogeneic donors with multiple cytokines, chemokine antagonists and cytokine/chemokine antagonist combinations.

It is widely accepted that a minimum of 2×10^6 cells/kg of recipient body weight is needed to ensure multilineage engraftment in recipients.⁷ The optimal number of stem cells is 5×10^6 cells/kg, which leads to earlier and more consistent neutrophil and platelet engraftment and reduced resource utilization.⁸ Remobilization is costly, and stem cell yield is usually less than the initial mobilization.⁹ Several algorithms have shown success in predicting and preventing mobilization failure through the preemptive approach based on well-established post-mobilization PB CD34 count and stem cell yield. Most of these focus on the pre-apheresis stage after the mobilization regimen has been completed but before apheresis.¹⁰⁻¹²

In recent years, with the increasing amount and complexity of patient data, machine learning (ML) has become popular due to its ability to quickly analyze large complex datasets and recognize the intrinsic pattern within the input data.¹³ ML has also been used in the context of HSCT, for example, to predict acute graft-versus-host disease following allogeneic transplantation.¹⁴ Nevertheless, ML-based prediction of stem cell mobilization has not been explored yet.

In this study, we retrospectively reviewed 1361 related allogeneic healthy donors from Washington University (WU) Blood and Marrow Transplant (BMT) program from 1995 to 2018. We compared the impact of 6 mobilization strategies on their efficacy to mobilize stem cells in allogeneic donors. Using demographic data and baseline screening laboratory results from the G-CSF–mobilized normal donors, we established ML-based scoring models to predict donors who may have poor/suboptimal stem cell yield (CD34 cell count $<40/\mu\text{L}$) following standard G-CSF. We expected that this machine learning approach would enable us to detect donors with suboptimal mobilization response at the early donor screening stage, at a time when potentially alternative approaches could be pursued to ensure optimal stem cell yields for HSCT. These alternative approaches may include selecting alternative related donors if available, increasing dose or frequency of G-CSF dosing, the addition of chemokine antagonists or GM-CSF to G-CSF.

Methods

Patient population

The Washington University School of Medicine institutional review board approved this retrospective study. The study was conducted in accordance with the Declaration of Helsinki. We retrospectively reviewed 1361 related allogeneic adult (≥ 18 years old) stem cell donors who underwent stem cell mobilization at WU Siteman Cancer Center between 1995 and 2018. Donors were mobilized with the standard of care G-CSF (G, N = 1025) or various other mobilization strategies evaluated through phase I/II clinical trials at Washington University, including: GM-CSF (GM, N = 40), G-CSF + GM-CSF (G+GM, N = 167), GM-CSF + Plerixafor (GM+P, N = 23), Plerixafor (N = 88) and BL-8040 (N = 18).

Mobilization regimens

Cytokine-based mobilization agents including G-CSF or GM-CSF involve 4–5 days of consecutive treatment before apheresis and stem cell collection, whereas CXCR4 antagonists (Plerixafor and

BL-8040) induce a rapid mobilization of stem cells, and the apheresis is completed in the same day. The standard mobilization regimen G-CSF was administered at $10 \mu\text{g}/\text{kg}$ SC once daily with leukapheresis (LP) commencing on day 5. Administration of GM and G+GM were described previously¹⁵: GM alone was administered at 10 or $15 \mu\text{g}/\text{kg}$ SC once daily with LP commencing on day 6; in the G+GM cohort, G ($10 \mu\text{g}/\text{kg}$) and GM ($5 \mu\text{g}/\text{kg}$) were administered SC once daily with LP commencing on day 5. In the GM+P cohort (ClinicalTrials.gov Identifier: NCT01158118), GM was administered at $5 \mu\text{g}/\text{kg}$ SC once daily with one dose of Plerixafor at $320 \mu\text{g}/\text{kg}$ IV before the LP on day 5. In the Plerixafor alone cohort, Plerixafor was administered at $320 \mu\text{g}/\text{kg}$ IV (ClinicalTrials.gov Identifier: NCT00914849) or $240 \mu\text{g}/\text{kg}$ by one SC dosing (ClinicalTrials.gov Identifier: NCT01696461), LP was commenced 4 hours post mobilization. In our analysis, donors receiving either SC or IV of Plerixafor were combined into one cohort as previous studies from our group showed that SC or IV administration of Plerixafor had similar mobilization kinetics and efficacy.¹⁶ In the BL-8040 cohort (ClinicalTrials.gov Identifier: NCT02639559), a single dose of BL-8040 was administered at $1.25 \text{ mg}/\text{kg}$ by SC, and LP was commenced 3 hours (up to 270 minutes) post mobilization.

Data collection

Donor demographics, including age at donation, gender, BMI at baseline, and donor baseline laboratory results (supplemental Table 1), including CBC with differential, blood chemistry was collected by retrospective chart review in electronic health record (EHR). Day 1 (of LP) CD34 yield per kg of the recipient body weight and PB CD34 count/ μL were collected from EHR as the primary outcomes of the study. Protected health information (PHI) was removed before proceeding to the subsequent modeling analysis.

Machine learning modeling

The modeling process was implemented in Scikit-learn.¹⁷ Data from 799 G-mobilized donors (Figure 2) were used to establish the predictive models for stem cell mobilization. The binary outcome of the prediction model was defined as on day 1 of LP: (1) class 0, donors who failed to reach the goal of optimal stem cell mobilization of $40 \text{ cells}/\mu\text{L}$ (poor/less-than-optimal mobilizer); (2) class 1, donors who reached the optimal stem cell mobilization of $40 \text{ cells}/\mu\text{L}$ (good mobilizer). One donor with missing CD34 count was manually assigned to the good mobilizer category based on the corresponding day 1 CD34 collection yield. Raw data underwent several pre-processing steps, including data normalization, mobilization class mapping and class distribution rebalancing with the synthetic minority oversampling technique (SMOTE) algorithm (described separately below). To obtain stable prediction models, different demographic features are linearly normalized by: $z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$, where x_i represents the feature value of the i^{th} donor, $\min(x)$ and $\max(x)$ are the minimum and maximum across all feature values under feature x .

Subsequently, different prediction methods were established, including a feature selection process to achieve improved results. In addition, a grid searching process is used to automatically select appropriate hyper-parameters for each prediction model. Seven machine learning algorithms, including Decision Tree, Linear Regression, Random Forest, Support Vector Machine, Feedforward Neural Networks, AdaBoost and Gradient boosting, were adopted to build the prediction model. Each model was trained and validated using

Table 1. Patient characteristics of 1361 related allogeneic donors

Regimens		All patients	G-CSF	G-CSF + GM-CSF	GM-CSF	GM-CSF + Plerixafor	Plerixafor	BL-8040
Total, N		1361	1025	167	40	23	88	18
Age	Median (range)	49 (18-79)	50 (18-78)	44 (18-77)	49 (29-79)	52 (21-65)	51 (21-67)	55 (20-69)
Sex, n (%)	Female	611 (44.9)	462 (45.1)	74 (44.3)	19 (47.5)	13 (56.5)	36 (40.9)	7 (38.9)
	Male	750 (55.1)	563 (54.9)	93 (55.7)	21 (52.5)	10 (43.5)	52 (59.1)	11 (61.1)
Race, n (%)	White	1209 (88.8)	904 (88.2)	148 (88.6)	38 (95)	21 (91.3)	82 (93.2)	16 (88.9)
	African American	96 (7.1)	78 (7.6)	10 (6)	1 (2.5)	2 (8.7)	5 (5.7)	–
	Others	41 (3)	31 (3)	8 (4.8)	–	–	1 (1.1)	1 (5.6)
	Unknown	15 (1.1)	12 (1.2)	1 (0.6)	1 (2.5)	–	–	1 (5.6)
CD34 yield	Day 1 CD34 per kg (10^6)	6.97 (0.1-62.65)	7.57 (0.3-62.65)	7.6 (0.3-57.4)	2.9 (0.1-20.6)	2.82 (0.79-15.41)	2.27 (0.2-19.73)	2.25 (0.4-9.42)
	Day 1 CD34 per μ L	64 (0.5-528)	70 (4.6-503)	79.7 (1.5-528)	26 (0.5-150)	23 (8-144)	16 (2-107)	27 (6-50)

10-fold cross-validation. Finally, the prediction performance of different methods was compared using accuracy, F1 score, and area under the ROC curve (AUC) metrics; and the prediction scores of each donor belonging to the 2 classes of “less-than-optimal mobilizer” and “good mobilizer” were generated as the output of the prediction model. A threshold of 0.5 was adopted to classify the donor as good mobilizer (mobilization score > 0.5) or poor/less-than-optimal mobilizer (mobilization score < 0.5).¹⁸ Detailed description for each individual algorithm is included in the supplemental Methods.

Oversampling with SMOTE

Since the donor distribution in class 0 and class 1 are highly skewed (1:3.7, class 0 is a minority while class 1 is a majority), this constitutes a class imbalance problem.¹⁹ To mitigate this problem, we performed a class distribution rebalancing process using SMOTE.²⁰ SMOTE is an over-sampling method in which the minority class is over-sampled by creating “synthetic” samples. Using SMOTE, 458 minority samples were synthesized and added to class 0.

Statistical analysis

Descriptive statistics were used to examine the patient's basic characteristics and proportion of different mobilization efficiency. Inferential statistics between different mobilization strategies were performed using nonparametric Kruskal-Wallis tests. The Pearson correlation coefficient test was adopted to determine statistically significant correlations between the 2 groups. As a comparison, a conventional multivariate logistics regression (good mobilization or not) was also constructed using the least absolute shrinkage and selection operator (LASSO). The linearity of each predictor was assessed by generalized additive models using nonparametric smoothing splines. The model performance was validated using 10-fold cross-validation. Odds ratios (OR) and their 95% confidence intervals were used to summarize the strength of association between each predictor and outcome. The predictive ability of the model was

described by AUC. Data analyses were performed using SAS 9.4 (SAS Institutes, Cary, NC) for logistic regression and using Prism 8 (GraphPad Software, Inc.) and Statistics (IBM) for all others. Results were considered to reach significance at $P \leq .05$ and are indicated with asterisks ($*P \leq .05$; $**P \leq .01$; $***P \leq .001$; $****P \leq .0001$). The source code used in our ML modeling is deposited and maintained on GitHub website (<https://github.com/MollahLab/PBHSCM>).

Results

Allogeneic donor characteristics

We identified 1361 related healthy adult (≥ 18 years old) stem cell donors who underwent their first stem cell mobilization at the BMT program of Washington University from 1995 to 2018. The baseline donor characteristics are shown in Table 1. The median age of all donors at the time of mobilization was 49 (range: 18-79). Age was comparable between different mobilization regimens, except that the donors in the G+GM cohort were significantly younger than the G-alone group ($P < .0001$, Kruskal-Wallis test) and Plerixafor alone group ($P = .0075$, Kruskal-Wallis test). The majority of the donors ($N = 1025$, 75.3%) were mobilized with the standard mobilization agent G-CSF. 336 healthy adult donors participated in the clinical trials (detailed in the methods) of alternative allogeneic mobilization strategies including: GM-CSF ($N = 40$, 2.9%), Plerixafor ($N = 88$, 6.5%), BL-8040 ($N = 18$, 1.3%), G-CSF + GM-CSF ($N = 167$, 12.3%) and GM-CSF + Plerixafor ($N = 23$, 1.7%). 611 (44.9%) of the donors were female, and 750 (55.1%) of the donors were male. The percentage of female donors was highest in the GM + Plerixafor (56.5%) cohort and lowest in the BL-8040 cohort (38.9%). The race of the majority of the donors was White, comprising 88.8% of all donors and ranging from 88.2% to 95% in each treatment group. The race of the rest of the donors was African American ($N = 96$, 7.1%), others ($N = 41$, 3%), and unknown ($N = 15$,

Table 2. Definition of “poor,” “less-than-optimal,” and “good” mobilizer

Day 1 CD34 yield/kg	Day 1 CD34/ μ L	Day 1 CD34/ μ L	Day 1 CD34 yield/kg (10^6)
Good, $\geq 5 \times 10^6$ /kg	87 (29-503)	Good (≥ 40)	9.03 (0.77-62.65)
Less-than-optimal, $2-5 \times 10^6$ /kg	36 (7-144)	Less-than-optimal (20-40)	3.48 (0.7-8.9)
Poor, $\leq 2 \times 10^6$ /kg	18 (4.6-174)	Poor (≤ 20)	1.7 (0.3-3.65)

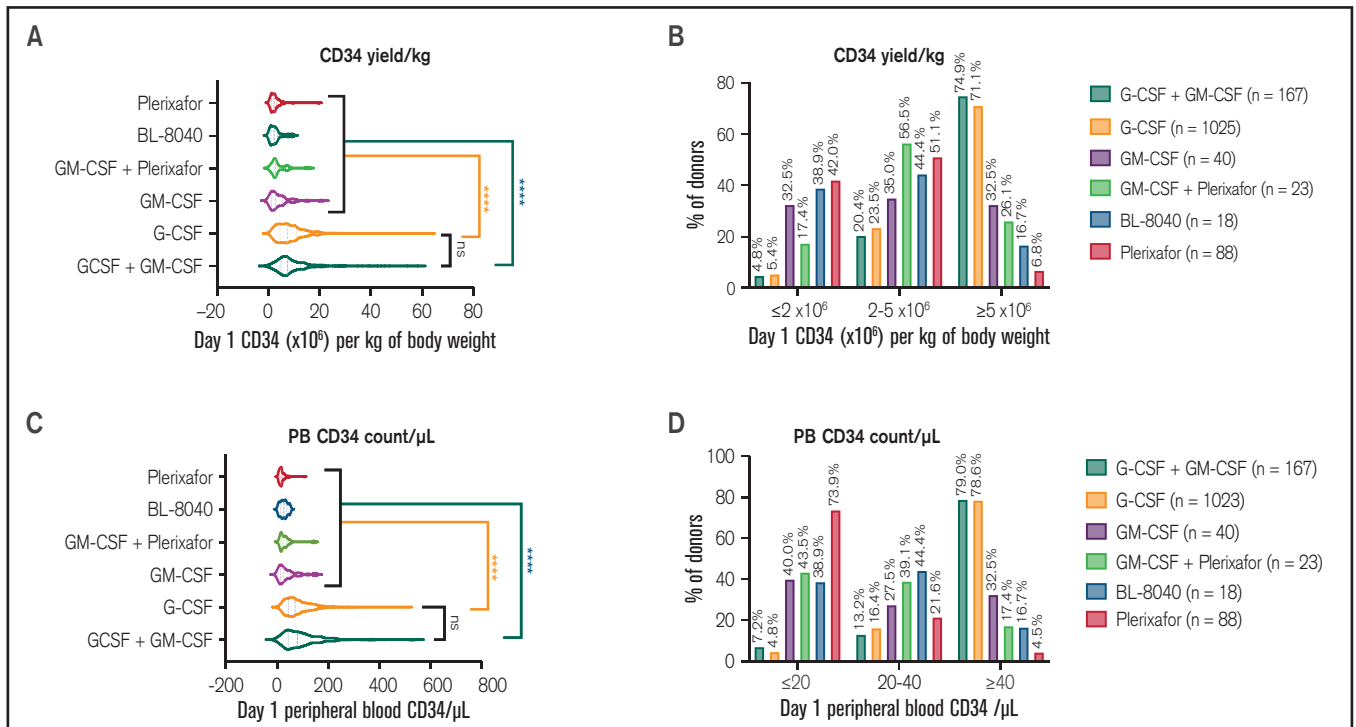


Figure 1. Impact of mobilization strategies on normal donor stem cell yield. Allogeneic donors were mobilized with cytokine (G-CSF or GM-CSF or G-CSF+GM-CSF), CXCR4 antagonists (Plerixafor or BL-8040) or combination of cytokine and CXCR4 antagonist (GM-CSF+Plerixafor). (A) Day 1 (of apheresis) CD34 stem cell yield per kg of recipient body weight in donors mobilized with different mobilization regimens. (B) Percentage of “poor mobilizer” ($\leq 2 \times 10^6$ cells/kg), “less-than-optimal mobilizer” (2.5×10^6 cells/kg), and “good mobilizer” ($\geq 5 \times 10^6$ cells/kg) in different mobilization regimens. (C) Day 1 peripheral blood (PB) CD34 cell count/ μL in donors mobilized with different mobilization regimens. (D) Percentage of “poor mobilizer” (PB CD34 count $\leq 20/\mu\text{L}$), “less-than-optimal” (PB CD34 count 20-40/ μL), and “good mobilizer” (PB CD34 count $\geq 40/\mu\text{L}$) in different mobilization regimens.

1.1%). The median CD34 yield/kg was 6.97×10^6 , and the median PB CD34 count was 64 cells/ μL on day 1 of LP.

The minimum threshold of CD34 stem cells for transplantation is $2 \times 10^6/\text{kg}$, below which is defined as a “poor mobilizer.” The optimal dose is $5 \times 10^6/\text{kg}$, and donors who failed to reach this target were defined as “less-than-optimal mobilizer” ($2.5 \times 10^6/\text{kg}$). However, the donor: recipient weight difference can vary greatly and confound the analysis. Thus, we examined the equivalent PB CD34 count of “poor mobilizers,” “less-than-optimal mobilizers,” and “good mobilizers” in our G-CSF cohort ($N = 1025$). In G-CSF mobilized donors ($N = 1023$), PB CD34 count was strongly correlated with the CD34 yield ($r = 0.81$, $P < .0001$) (supplemental Figure 1). In Table 2, we found that the median PB CD34 count is 18 cells/ μL in donors with less than $2 \times 10^6/\text{kg}$, which is in line with the 20 cells/ μL that others have reported,⁷ and the median PB CD34 count for “less-than-optimal mobilizers” was 36 cells/ μL . Using this information, we defined donors with day 1 PB CD34 count ≤ 20 cells/ μL as “poor mobilizers,” donors with day 1 CD34 count between 20-40 as “less-than-optimal mobilizers” and donors with day 1 CD34 count ≥ 40 as “good mobilizers.”

Impact of mobilization regimen on stem cell yield

We first compared the day 1 CD34 yield of a single 20 L (median collection volume) apheresis from the six different mobilization strategies. All of these mobilization agents were well-tolerated in patients. Among six different mobilization approaches (Figure 1A,C), G-CSF

(both G-CSF alone or G + GM) induced significantly higher CD34 cells/kg ($P < .0001$) than any of the other groups. In line with this finding, the majority of the donors mobilized with G-CSF alone (71.1%) and G + GM (74.9%) had good mobilization ($> 5 \times 10^6/\text{kg}$), with failure rates only 5.4% and 4.8% in the G-CSF or GM mobilized donors, respectively (Figure 1B). In contrast, single-day mobilization with CXCR4 antagonist alone (Plerixafor or BL-8040) was associated with a high poor mobilizer rate ranging from 38.9% to 42%. This was also consistent when the mobilizer status was defined by PB CD34 count per μL (Figure 1D). Interestingly, although there was no statistical difference in CD34 yields (Figure 1A), there was a higher percentage of good mobilizers, and a lower percentage of poor mobilizers from BL-8040 mobilized donors, compared with Plerixafor. GM alone or in combination with Plerixafor showed comparable efficacy as the CXCR4 antagonists alone. Together, these data showed that mobilization with G-CSF is effective, other mobilization strategies such as CXCR4 antagonists can induce a rapid mobilization in one day, but they fail to induce minimum stem cell yields in 30% to 50% of the donors after a single apheresis.

Patient selection for ML modeling

After removing donors with missing demographic or laboratory data, 799 G-CSF mobilized donors (from 2003 to 2018) were identified for the predictive analysis (Figure 2), among which 171 donors had CD34 count $< 40/\mu\text{L}$ (poor/less-than-optimal mobilizer) and 628

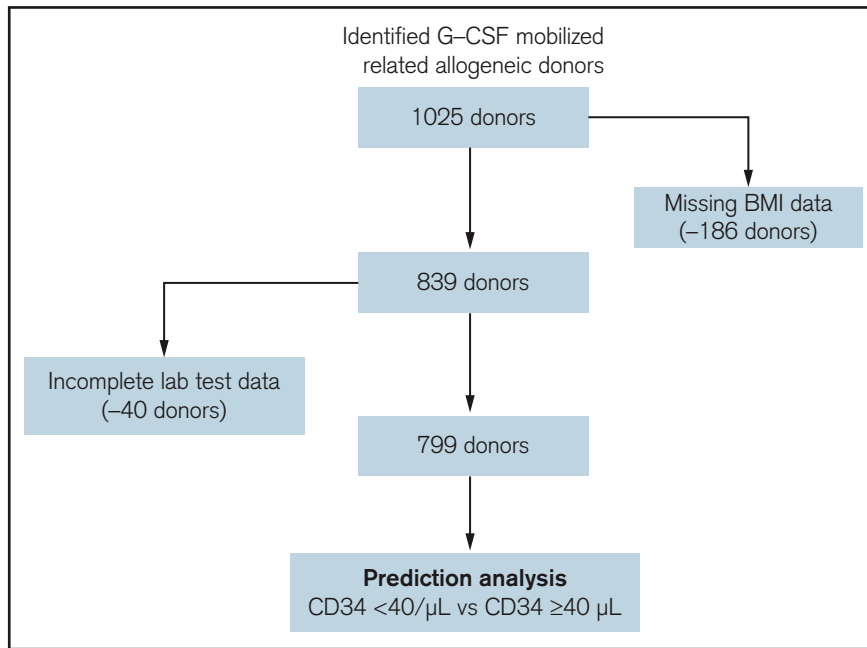


Figure 2. Flow-chart of patient selection for the prediction analysis. One thousand twenty-five G-CSF mobilized allogeneic donors were identified in the EHR database. Donors with key demographical data (N = 186) missing or incomplete donor screening laboratory test data (N = 40) were removed. The remaining 799 donors were selected for the subsequent prediction analysis to identify donors with “poor/less-than-optimal” mobilization (Day 1 PB CD34 cell count <40/μL) or “good” mobilization (Day 1 PB CD34 cell count ≥40/μL).

donors had CD34 count ≥ 40/μL (good mobilizer). Patient characteristics used in the predictive modeling were summarized in supplemental Table 1. 70.6% of the donors (N = 564) were matched siblings, and 29% were haplo-identical donors (N = 232) (supplemental Table 2). From an initial assessment of the baseline laboratory and demographical data using Pearson correlation test from these donors, we found that age and BMI, as well as basal laboratory results including WBC, platelet counts, lymphocyte count, monocyte counts, uric acid and a few others, were significantly correlated with the day 1 PB CD34 count (Table 3). This prompted us to explore if we can use these data to predict donors who may have poor or less-than-optimal mobilization (CD34 count < 40/μL) in response to G-CSF at the early donor screening stage. We took advantage of the large dataset of G-CSF-mobilized donors and used a novel unbiased machine learning approach.

ML model predicts response to G-CSF-mediated stem cell mobilization

We adopted a well-established machine learning pipeline (Figure 3A) to predict whether the donors may have poor or less-than-optimal mobilization (CD34 count < 40/μL), based on demographics and baseline laboratory data. A tree-based feature selection revealed 18 highly discriminant features for the linear regression and decision tree models (Figure 3B). The top 3 variables also showed significant correlation with the day 1 CD34 count. There was a significant negative correlation ($r = -0.25$) between age and day 1 CD34 count (Figure 3C) and a positive correlation between platelet count ($r = 0.21$) and BMI ($r = 0.27$) and day 1 CD34 count (Figure 3D-E). When stratified by BMI, overweight and obese donors had significantly higher CD34 yield (Figure 3F). These key clinical features, as well as the rest of the features, were adopted and the

prediction models were developed using seven machine learning algorithms. We then evaluated different models using a 10-fold cross-validation technique: dataset containing all 799 donors were first partitioned into 10 equal-size subsets, then in each fold, a different subset was iteratively retained as the validation set while the remaining 9 subsets were combined as the training set. The

Table 3. Top variables correlated with Day 1 PB CD34 count/μL

Pearson correlation	<i>r</i>	<i>P</i> , 2-tailed, <i>P</i> summary
PB CD34 count vs BMI	0.2744	<.0001****
PB CD34 count vs WBC	0.2303	<.0001****
PB CD34 count vs Platelets	0.2094	<.0001****
PB CD34 count vs Lymphs, absolute	0.1891	<.0001****
PB CD34 count vs Mono, absolute	0.1883	<.0001****
PB CD34 count vs Uric acid	0.1834	<.0001****
PB CD34 count vs Neutrophil, absolute	0.178	<.0001****
PB CD34 count vs RBC	0.1759	<.0001****
PB CD34 count vs ALT	0.1378	<.0001****
PB CD34 count vs BASO, absolute	0.1196	.0007***
PB CD34 count vs EOS, absolute	0.1139	.0013**
PB CD34 count vs MPV	0.09903	.0051**
PB CD34 count vs PTT	0.08833	.0126*
PB CD34 count vs Hematocrit	0.07807	.0274*
PB CD34 count vs BUN	-0.1216	.0006***
PB CD34 count vs Age	-0.2475	<.0001****
PB CD34 count vs MCV	-0.1708	<.0001****
PB CD34 count vs MCH	-0.1613	<.0001****

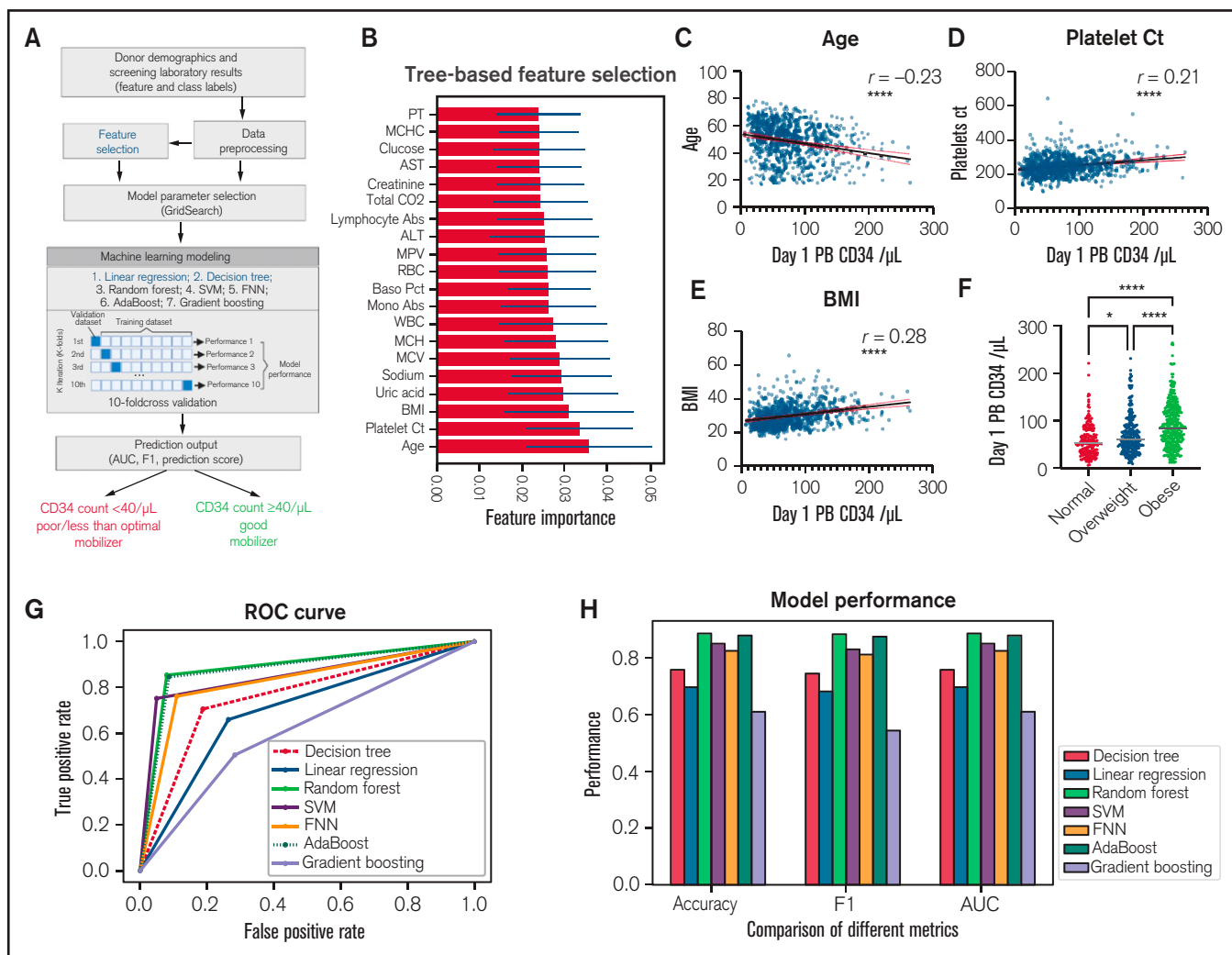


Figure 3. Machine learning-based modeling to predict “poor/less-than-optimal” vs “good” mobilizer. (A) Schema of machine learning modeling pipeline to predict “poor/less-than-optimal” vs “good” mobilizer. (B) Tree-based feature selection for linear regression and decision tree modeling, ranked by feature importance. Correlation of age (C), platelet count (D), and BMI (E) with Day 1 PB CD34 count/ μL ; Day 1 PB CD34 count/ μL in normal (BMI: 18.5-25), overweight (BMI: 25-30) and obese (BMI: >30) donors (F); ROC curve of seven machine learning prediction algorithms, including: Decision tree, Linear Regression, Random Forest, Support Vector Machine, Feedforward Neural Networks, AdaBoost and Gradient boosting (G). (H) Model performance evaluated by accuracy, F1 score, and AUC.

prediction abilities of seven produced prediction models were compared by the receiver operating characteristic (ROC) curves (Figure 3G). AUC is the gold standard to evaluate the predictive performance of the prediction models. We also used accuracy and F1 score to compare the average prediction results quantitatively over the 10-fold runs (Figure 3H). According to all 3 metrics, Random Forest and AdaBoost achieved the best performance, with AUCs of 0.89 and 0.88, respectively (Table 4).

Prediction scores for each donor in the validation cohort were calculated from the Random Forest or AdaBoost algorithms. The prediction score that predicts the good mobilizers was called the “mobilization score,” which represented the likelihood of having CD34 count ≥ 40 cells/ μL . The higher the mobilization score, the higher the likelihood the CD34 counts would be higher than 40 cells/ μL , as shown in Figure 4A,D. In addition, we examined the distribution of the mobilization scores for all 799 donors. Both

Random Forest and AdaBoost based showed good discrimination of good mobilizers vs less-than-optimal mobilizers (Figure 4B,E). Finally, the performance of the two models with all 799 donors was constructed in confusion matrices (Figure 4C,F). In the Random

Table 4. Table of AUCs from seven ML prediction models.

Machine learning algorithms	AUC
Decision tree	0.76
Linear regression	0.7
Random forest	0.89
Support vector machine	0.85
Feedforward neural network	0.83
Adaboost	0.88
Gradient boosting	0.61

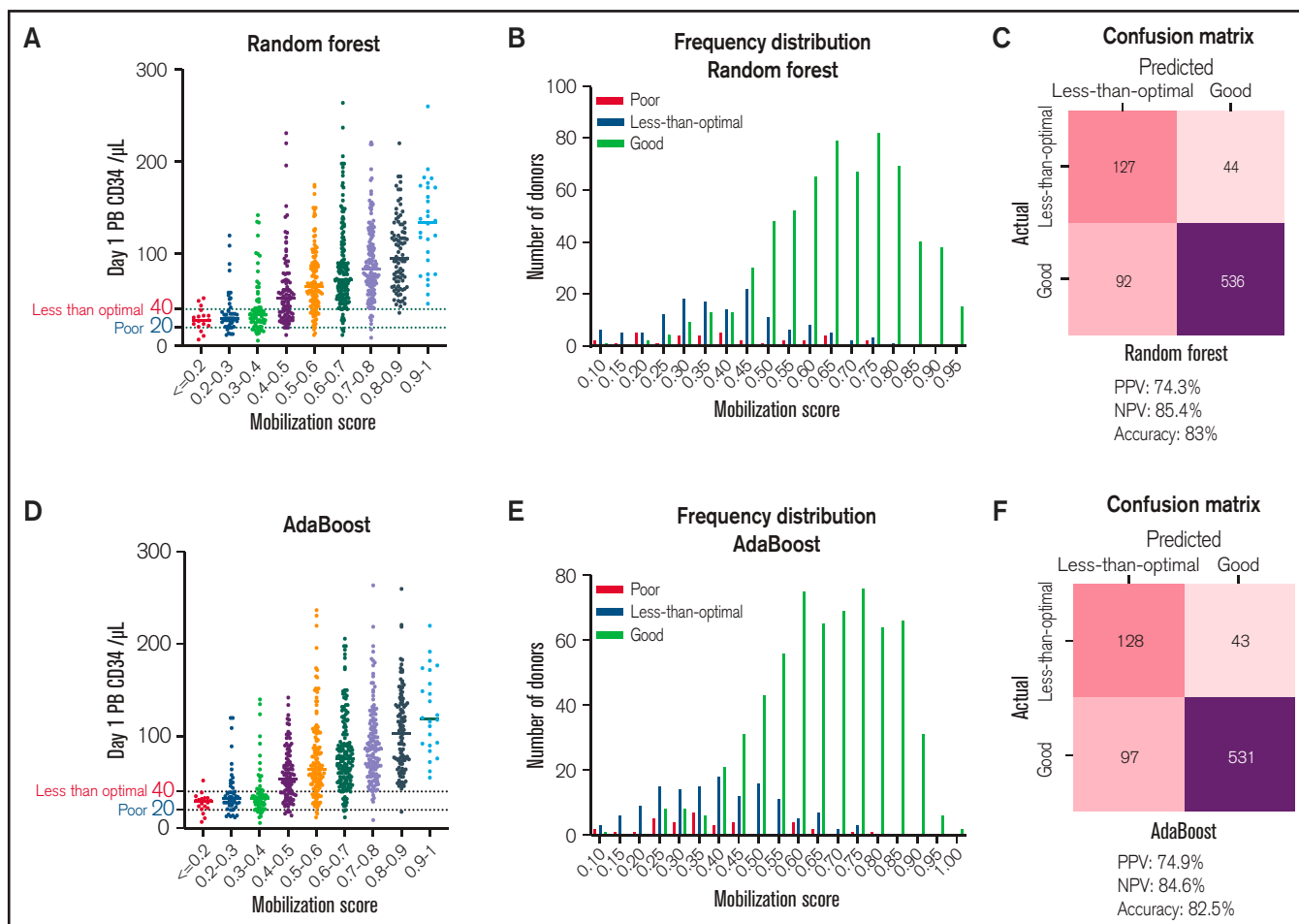


Figure 4. ML-based mobilization score predicts response to G-CSF-mediated stem cell mobilization from pre-mobilization data. (A) Mobilization score derived from the Random Forest algorithm-based ML model vs Day 1 PB CD34 count in G-CSF mobilized donors (N = 798). (B) Distribution of “poor mobilizer” (Day 1 PB CD34 cells <40/ μ L), “less-than-optimal mobilizer” (Day 1 PB CD34 cells 20-40/ μ L) and “good-mobilizer” (Day 1 PB CD34 cells >40/ μ L) based on their mobilization score calculated by Random Forest based ML model (N = 798). (C) Confusion matrix, PPV, NPV and accuracy of the prediction in all patients (N = 799) from the Random Forest-based model. One hundred twenty-seven out of 171 “poor/less-than-optimal mobilizers” and 536 out of 628 “good mobilizers” were correctly predicted. (D) Mobilization score derived from the AdaBoost algorithm vs Day 1 PB CD34 count in G-CSF mobilized donors (N = 798). (E) Distribution of “poor mobilizer” (Day 1 PB CD34 count <40/ μ L), “less-than-optimal mobilizer” (Day 1 PB CD34 count 20-40/ μ L) and “good mobilizer” (Day 1 PB CD34 count >40/ μ L) based on their mobilization score calculated by AdaBoost based ML model. (F) Confusion matrix, PPV, NPV and accuracy of the prediction in all patients (N = 799) from the AdaBoost-based model. 128 out of 171 “poor/less-than-optimal mobilizers” and 531 out of 628 “good mobilizers” were correctly predicted.

Forest-based machine learning model, 127 out of 171 “poor/less-than-optimal mobilizers” and 536 out of 628 “good mobilizers” were correctly predicted (positive prediction value [PPV]: 74.3%, negative prediction value [NPV]: 85.4%, accuracy: 83%). The AdaBoost-based machine learning model showed similar prediction performance, as 128 out of 171 “poor/less-than-optimal mobilizers” and 531 out of 628 “good mobilizers” were correctly predicted (PPV: 74.9%, NPV: 84.6%, accuracy: 82.5%). Together, these machine learning-based models demonstrated strong predictive power for predicting donor’s response to G-CSF-induced stem cell mobilization.

Discussion

G-CSF is currently the standard of care for mobilization and is well-tolerated in patients. However, it requires 4 to 5 days of daily

treatment and is associated with side effects such as skeletal pain, fever, fatigue, nausea, splenic enlargement and rarely, splenic rupture.² In an effort to find more rapid, low-risk mobilization regimens for donors, the BMT program at Washington University has initiated a number of phase I/II trials to evaluate novel mobilization strategies. In this retrospective review, we did a side-by-side comparison of the mobilization efficacy of G-CSF with three alternative mobilization strategies, including (1) cytokine-based regimens: GM and GM+G; (2) CXCR4 antagonist-based regimens: Plerixafor and BL-8040; (3) combination of GM+P. Based on the CD34 stem cell yield, mobilization strategies involving G-CSF were the most effective mobilization agent, with over 70% of donors achieving the $>5 \times 10^6$ /kg goal after the first LP session. The addition of GM-CSF resulted in a slight increase of CD34 count/ μ L, but this was not statistically significant (Figure 1C), despite the significantly younger age in this cohort (Table 1). GM alone has been shown to be a relatively weak

mobilizer as a single agent and had to be adjusted to a higher dose of 15 $\mu\text{g}/\text{kg}$ to ensure the minimum $2 \times 10^6/\text{kg}$ goal was met.¹⁵ However, the rate of grade II-IV GvHD in recipients was only 13% from the cells mobilized from GM donors, compared with 49% and 69% from the G-CSF alone and G+GM cohorts, respectively.¹⁵ In contrast to the 4–5 days of cytokine (G-CSF or G+GM) mediated mobilization, CXCR4 antagonists directly disrupts the interaction of CXCR4-CXCL12 axis and induces a rapid mobilization of stem and progenitor cells into peripheral blood within hours. However, as single agents, CXCR4 antagonists mobilize significantly fewer CD34 cells/ μL than the standard G-CSF approach (Figure 1A,C). Interestingly, BL-8040 mobilizes slightly more CD34 cells compared with Plerixafor (Figure 1C). This may be explained by prior studies demonstrating that BL-8040 binds to CXCR4 with a higher affinity and longer CXCR4 occupancy (lower off-rate) compared with Plerixafor.²¹

A recent study using the National Marrow Donor Program (NMDP) dataset found that higher BMI predicts a better stem cell mobilization response to G-CSF. The maximum effective G-CSF dose was achieved with obese donors, above which higher doses of G-CSF resulted in no obvious increase in stem cell yield. In contrast, patients with normal or overweight BMI may benefit from increasing doses of G-CSF to enhance stem cell yields.²² However, these increasing doses of G-CSF may also be associated with increased side effects.² In our dataset, we have also observed a significantly higher CD34 cell yield in obese donors (Figure 3F). In our G-CSF data subset (supplemental Table 1), about half of the donors who had a CD34 count over 40 cells/ μL were obese, compared with 27% in donors who had less than 40 cells/ μL . The combinational use of BMI and mobilization score therefore would enable us to narrow down the subset of nonobese, low mobilization score donors who may benefit from increased G-CSF dosing.

We have previously demonstrated that novel combination therapy such as G-CSF +Plerixafor significantly improved the efficacy of stem cell mobilization and yield in cancer patients, which led to the FDA approval to use this regimen in patients with lymphoma and multiple myeloma.^{4,5} However, G-CSF remains the only FDA-approved method for mobilization of allogeneic donors. Donors who do not mobilize adequate stem cell yield on day 1 of LP may undergo additional apheresis on day 2. In our dataset, the CD34 yield on day 2 is statistically lower than the day 1 yield ($P < .0001$, Wilcoxon Matched-Pairs Signed Ranks Test, supplemental Figure 2). Regardless, there is also the high cost of the Plerixafor treatment and predictive algorithms to determine the small subset of donors who may benefit from the new combination therapy have not been developed yet. Taking the cost of remobilization into account, plerixafor is better and cheaper if used “on-demand” than within a subsequent remobilization in multiple myeloma and lymphoma patients for autologous stem cell transplantation.²³ In some centers, the “preemptive” strategy based on the PB CD34 count showed some success to salvage the initial mobilization with G-CSF with the addition of Plerixafor.²⁴ Therefore, we hope that the use of our mobilization score could further improve this clinical need.

The present study demonstrates the use of machine learning as a prediction tool to assist clinical decision making in the context of PB mobilization for HSCT. To our knowledge, this is the first ML-based model for stem cell mobilization prediction and one of the largest single-center datasets for allogeneic stem mobilization. Notably, the

data used in the prediction modeling were usually acquired 1–2 weeks prior to the mobilization, therefore allowing for early detection of donors who may have poor or less than optimal stem cell yields. Compared with the traditional statistical analysis, which is model-driven, our ML approach shows higher predictability power. In our parallel analysis of the same dataset using a multivariate logistic regression model (supplemental Figure 3, supplemental Table 3), an AUC of 0.74 was achieved, compared with the 0.89 in the ML approach. However, since the ML model is created from the data by the algorithms, it is often presented as a “black box” that is less interpretable as to what the model has learned during the process and how much each feature contributes to the final prediction.²⁵ In further studies, machine learning algorithms such as Naïve Bayes (NB) and alternative decision tree (ADTree) that provide interpretable model structure will also be explored.

One limitation of the study is that this is a single-center study and spans over two decades. While the single-center dataset showed benefit for our comparison of different mobilization regimens by eliminating inter-center variabilities, for our prediction model, additional datasets from other centers, the NDMP or CIBMTR datasets may further improve the accuracy and generalizability of our models. In spite of the significant technological advances in apheresis during the study time, there were minimal changes in the proportion of good versus poor mobilizers after the 1995–2000 period, as determined by both the CD34 yield and PB CD34 count (supplemental Figure 4A–B). Additional datasets from the above sources in recent years may further strengthen our study. Another limitation of the study is that while our prediction model can discriminate most of the donors from less or over 40 cells/ μL CD34 cells, it cannot separate the poor mobilizer (<20) vs suboptimal mobilizer (between 20 and 40) due to the small percentage of poor mobilizers in our dataset. Additional datasets and clinical features will be included in future studies. More importantly, a similar approach will also be extended to autologous cancer patients, who have a much higher failure rate (up to 20%) during the stem cell mobilization.²⁶

In conclusion, in this study we evaluated the impact of six strategies for stem cell mobilization. G-CSF-based mobilization strategies were the most efficient resulting in the highest CD34 cells/ μL after 4 to 5 days of G-CSF and in the first apheresis collection, while CXCR4 antagonists as a mobilizing strategy results in significantly less but more rapid mobilization in the same day. Using donor demographic data and baseline laboratory results from the G-CSF mobilized normal donors, we developed ML-based mobilization scoring models that enable us to predict allogeneic donor response to G-CSF for stem cell mobilization at the early donor screening stage. The current study provides proof of concept that the ML-based approach can be used to predict G-CSF-induced stem cell mobilization in normal allogeneic donors and would serve as a base for the future validation using other datasets, alternative mobilizing agents (CXCR4 antagonists) or for similar analyses in patients undergoing autologous stem cell transplantation.

Acknowledgments

The authors thank David Warren, Joel Eissenberg, Brian Gage, Philip Payne, Randi Foraker, and Aixia Guo for their advice and support.

This work was supported by National Cancer Institutes (NCI), National Institutes of Health: R35 CA210084 NCI Outstanding Investigator Award (principal investigator [PI]: DiPersio). 5R50CA211466-05, bone marrow (BM) Niche Disruption and Immunotherapy in Hematological Malignancies (PI: Rettig).

Authorship

Contribution: J.X. and J.F.D. conceived and designed the study; J.X., M.S., and S.M. developed the methodology; J.X. and M.A.F. acquired the data; J.X., M.S., F.G., S.M., and J.F.D. analyzed and

interpreted data; and all authors wrote, reviewed, and/or revised the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

ORCID profiles: M.S., 0000-0003-4017-5377; G.L.U., 0000-0002-7809-0996; M.A.S., 0000-0003-4924-8990; K.N.W., 0000-0002-7137-4742; S.M., 0000-0001-9178-8339.

Correspondence: John F. DiPersio, Washington University School of Medicine, Campus Box 8007, 660 South Euclid Ave, St. Louis, MO 63110. e-mail: jdpersi@wustl.edu.

References

1. Anasetti C, Logan BR, Lee SJ, et al; Blood and Marrow Transplant Clinical Trials Network. Peripheral-blood stem cells versus bone marrow from unrelated donors. *N Engl J Med*. 2012;367(16):1487-1496.
2. Hölig K. G-CSF in healthy allogeneic stem cell donors. *Transfus Med Hemother*. 2013;40(4):225-235.
3. Rettig MP, Ansstas G, DiPersio JF. Mobilization of hematopoietic stem and progenitor cells using inhibitors of CXCR4 and VLA-4. *Leukemia*. 2011;26(1):34-53.
4. DiPersio JF, Micallef IN, Stiff PJ, et al; 3101 Investigators. Phase III prospective randomized double-blind placebo-controlled trial of plerixafor plus granulocyte colony-stimulating factor compared with placebo plus granulocyte colony-stimulating factor for autologous stem-cell mobilization and transplantation for patients with non-Hodgkin's lymphoma. *J Clin Oncol*. 2009;27(28):4767-4773.
5. DiPersio JF, Stadtmauer EA, Nademanee A, et al; 3102 Investigators. Plerixafor and G-CSF versus placebo and G-CSF to mobilize hematopoietic stem cells for autologous stem cell transplantation in patients with multiple myeloma. *Blood*. 2009;113(23):5720-5726.
6. Abraham M, Pereg Y, Bulvik B, et al. Single dose of the CXCR4 antagonist BL-8040 induces rapid mobilization for the collection of human CD34⁺ cells in healthy volunteers. *Clin Cancer Res*. 2017;23(22):6790-6801.
7. To LB, Levesque JP, Herbert KE. How I treat patients who mobilize hematopoietic stem cells poorly. *Blood*. 2011;118(17):4530-4540.
8. Bensinger W, Appelbaum F, Rowley S, et al. Factors that influence collection and engraftment of autologous peripheral-blood stem cells. *J Clin Oncol*. 1995;13(10):2547-2555.
9. Fiala MA, Park S, Slade M, DiPersio JF, Stockerl-Goldstein KE. Remobilization of hematopoietic stem cells in healthy donors for allogeneic transplantation. *Transfusion*. 2016;56(9):2331-2335.
10. Duong HK, Bolwell BJ, Rybicki L, et al. Predicting hematopoietic stem cell mobilization failure in patients with multiple myeloma: a simple method using day 1 CD34⁺ cell yield. *J Clin Apher*. 2011;26(3):111-115.
11. Costa LJ, Miller AN, Alexander ET, et al. Growth factor and patient-adapted use of plerixafor is superior to CY and growth factor for autologous hematopoietic stem cells mobilization. *Bone Marrow Transplant*. 2010;46(4):523-528.
12. Costa LJ, Alexander ET, Hogan KR, Schaub C, Fouts TV, Stuart RK. Development and validation of a decision-making algorithm to guide the use of plerixafor for autologous hematopoietic stem cell mobilization. *Bone Marrow Transplant*. 2010;46(1):64-69.
13. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44-56.
14. Arai Y, Kondo T, Fuse K, et al. Using a machine learning algorithm to predict acute graft-versus-host disease following allogeneic transplantation. *Blood Adv*. 2019;3(22):3626-3634.
15. Devine SM, Brown RA, Mathews V, et al. Reduced risk of acute GVHD following mobilization of HLA-identical sibling donors with GM-CSF alone. *Bone Marrow Transplant*. 2005;36(6):531-538.
16. Schroeder MA, Rettig MP, Lopez S, et al. Mobilization of allogeneic peripheral blood stem cell donors with intravenous plerixafor mobilizes a unique graft. *Blood*. 2017;129(19):2680-2692.
17. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825-2830.
18. Freeman EA, Moisen GG. A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecol Modell*. 2008;217(1-2):48-58.
19. Daskalaki S, Kopanas I, Avouris N. Evaluation of classifiers for an uneven class distribution problem. *Appl Artif Intell*. 2006;20(5):381-417.
20. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. *J Artif Intell Res*. 2002;16:321-357.
21. Peled A, Abraham M, Avivi I, et al. The high-affinity CXCR4 antagonist BKT140 is safe and induces a robust mobilization of human CD34⁺ cells in patients with multiple myeloma. *Clin Cancer Res*. 2013;20(2):469-479.
22. Farhadfar N, Hsu JW, Logan BR, et al. Weighty choices: selecting optimal G-CSF doses for stem cell mobilization to optimize yield. *Blood Adv*. 2020;4(4):706-716.

23. Tichopád A, Vítová V, Kořístek Z, Lysák D. Cost-effectiveness of hematopoietic stem cell mobilization strategies including plerixafor in multiple myeloma and lymphoma patients. *J Clin Apher.* 2013;28(6):395-403.
24. Horwitz ME, Chute JP, Gasparetto C, et al. Preemptive dosing of plerixafor given to poor stem cell mobilizers on day 5 of G-CSF administration. *Bone Marrow Transplant.* 2011;47(8):1051-1055.
25. Castelveccchi D. Can we open the black box of AI? *Nature.* 2016;538(7623):20-23.
26. Giralt S, Costa L, Schriber J, et al. Optimizing autologous stem cell mobilization strategies to improve patient outcomes: consensus guidelines and recommendations. *Biol Blood Marrow Transplant.* 2014;20(3):295-308.