



The 65th ASH Annual Meeting Abstracts

ORAL ABSTRACTS

722.ALLOGENEIC TRANSPLANTATION: ACUTE AND CHRONIC GVHD, IMMUNE RECONSTITUTION

Machine Learning Validates Risk Biomarkers of Chronic Graft-Versus-Host Disease in 936 Patients from BMT CTN 0201 & 1202 CohortsMichael J Martens, PhD^{1,2}, Jianqun Kou, MS¹, Brent R. Logan, PhD^{3,2}, Sophie Paczesny⁴¹ CIBMTR® (Center for International Blood and Marrow Transplant Research), Department of Medicine, Medical College of Wisconsin, Milwaukee, WI² Division of Biostatistics, Institute for Health and Equity, Medical College of Wisconsin, Milwaukee, WI³ CIBMTR® (Center for International Blood and Marrow Transplant Research), Medical College of Wisconsin, Milwaukee⁴ Medical University of South Carolina, Charleston, SC**Introduction**

Our previous work identified susceptibility/risk biomarkers for developing chronic GVHD (cGVHD) in individuals without clinically apparent disease (Logan et al. 2023, JCI), which is a major contributor to morbidity and mortality following allogeneic hematopoietic cell transplant (HCT). In this study, we used machine learning (ML) algorithms to build biomarker-based prediction models for development of cGVHD and non-relapse mortality (NRM).

Methods

Data were obtained on 9 pre-transplant factors and 7 plasma proteins at Day 90 post HCT in 936 HCT recipients from the BMT CTN 0201 and 1202 studies. The study population was split into separate training (80%) and validation (20%) datasets. Associations of each protein marker with the hazards of cGVHD and NRM were evaluated using Cox proportional hazards (PH) models. To evaluate the predictive ability of the proteins, ML models were constructed from them and the time varying Area Under the ROC curve (AUC_t) was estimated at Days 180, 270, 360, and 540 post HCT. We considered ML approaches within the PH model framework, including Boosting (XGBoost), Group SCAD, and Adaptive Group Lasso, as well as methods avoiding the PH assumption, including Random Survival Forests and Bayesian Additive Regression Trees (BART). DeepSurv and DeepHit models were also used in the first application of deep learning to GVHD biomarker risk evaluation to our knowledge. A p-value cutoff of 0.05 determined statistical significance.

Results

Of the 7 proteins tested, 5 (CXCL9, CXCL10, MMP3, DKK3, and CD163) were associated with cGVHD risk, while 4 (MMP3, DKK3, ST2, CD163) were correlated with NRM ($p < 0.05$ for all). Moreover, effects of some markers varied depending on the graft type, with higher MMP3 found to have a larger, deleterious effect on cGVHD risk in patients with bone marrow (BM) compared to peripheral blood (PB) graft ($p = 0.03$). CD163 had a similar interaction with graft type such that its effects on cGVHD and NRM risk are significant with BM graft but not with PB graft ($p < 0.05$ for both).

The abilities of the ML methods to risk stratify cGVHD and NRM were assessed and contrasted to a Group SCAD model having clinical factors only. Because graft type was found to impact both the risks of cGVHD and NRM as well as the biomarkers' effects, the ML models considered markers, graft type, and their interactions. In modeling cGVHD risk, the Group SCAD model with only clinical factors selected 4 variables: age, graft type, donor-recipient sex-mismatch, and GVHD prophylaxis; while Adaptive Group Lasso and Group SCAD with biomarkers selected 5 proteins: CXCL9, CXCL10, MMP3, ST2, and CD163. In the validation dataset, all ML methods with biomarkers provided similar or higher AUC_t for cGVHD after Day 270 compared to clinical factors only, suggesting that plasma proteins measured as early as Day 90 can inform underlying cGVHD biology that has not manifested clinically (**Fig**). BART and Boosting yielded the highest AUC_t and were better than the model with only clinical factors, attaining AUC_t > 0.60 from Day 270-540. To assess the variables' importance in ML, we examined BART's posterior probabilities of selection, which describe how frequently each variable is used in the BART tree ensemble. For cGVHD, variables with selection probabilities > 0.10 were graft type, CXCL9, and MMP3. Notably, the protein markers that were significant in separate models also had the largest selection probabilities of all markers. The AUC_t for NRM was stable across time points in the 0.65-0.72 range and superior to clinical factors after Day 180 (**Fig**). For NRM, variables with selection probabilities > 0.10 were graft type, ST2, MMP3, CD163, and DKK3.

Conclusions

ML methods using non-invasive plasma proteins can successfully identify and validate risk biomarkers of cGVHD. ML algorithms using objective and early measurements of soluble markers including CXCL9, CXCL10, MMP3, DKK3, and ST2 perform better in predicting cGVHD than ML with known clinical factors only. ML with MMP3, DKK3, ST2, and CD163 improved NRM prediction. Deep learning models did not outperform more classical ML methods, possibly due to a limited sample size that could not reveal intricate relationships between proteins and cGVHD/NRM risk. Several proteins represent potential therapeutic targets. These data support future research to further validate these biomarkers and to develop ML algorithms to identify patients at risk for developing cGVHD and NRM.

Disclosures Logan: Enlivex: Consultancy. **Paczesny:** Indiana University Research and Technology Corporation: Patents & Royalties: Biomarkers and assays to detect chronic graft versus host disease (U.S. Patent 10,571,478 B2).

Figure. Time-varying Area Under the Curve (AUC_t) for Machine Learning Models

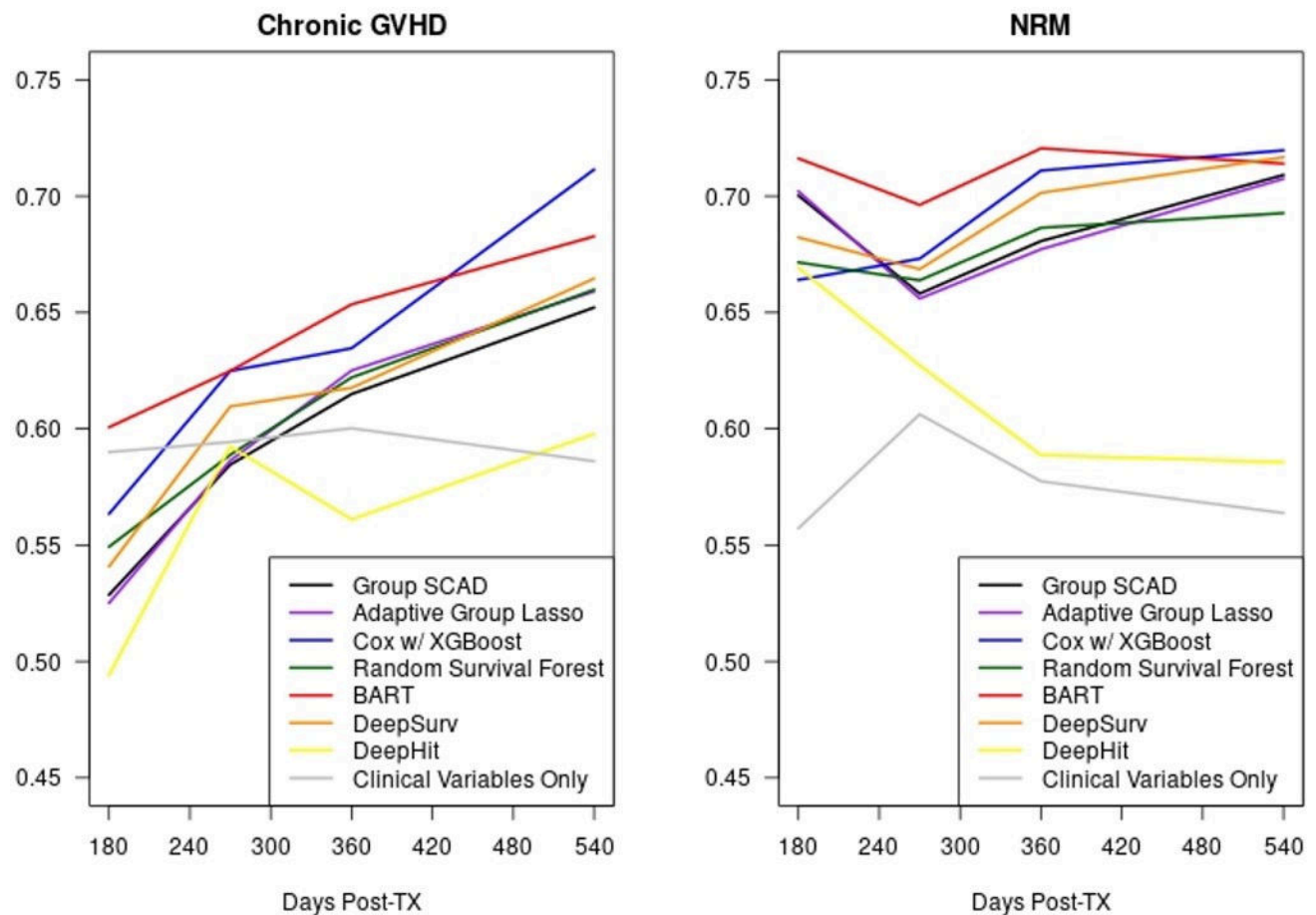


Figure 1

<https://doi.org/10.1182/blood-2023-182438>

Downloaded from http://ashpublications.net/blood/article-pdf/142/Supplement_1/479/2189805/blood-5083-main.pdf by guest on 17 May 2024