TO THE EDITOR:

Validation of genetic associations with acute GVHD and nonrelapse mortality in DISCOVeRY-BMT

Hancong Tang,1 Theresa Hahn,2 Ezgi Karaesmen,1 Abbas A. Rizvi,1 Junke Wang,1 Sophie Paczesny,3 Tao Wang,4,5 Leah Preus,1 Qianqian Zhu,6 Yiwen Wang,1 Christopher A. Haiman,7 Daniel Stram,7 Lorell Pooler,7 Xin Sheng,7 David Van Den Berg,7 Guy Brock,8 Amy Webb,6 Marcelo C. Pasquini,4 Philip L. McCarthy,2 Stephen R. Spellman,9 and Lara E. Sucheston-Campbell1,10

1Pharmaceutics and Pharmaceutical Chemistry, College of Pharmacy, The Ohio State University, Columbus, OH; 2Department of Medicine, Roswell Park Comprehensive Cancer Center, Buffalo, NY; 3Department of Pediatrics, Indiana University School of Medicine, Indianapolis, IN; 4Center for International Blood and Marrow Transplant Research, Department of Medicine, and 5Division of Biostatistics, Institute for Health and Society, Medical College of Wisconsin, Milwaukee, WI; 6Department of Biostatistics, Roswell Park Comprehensive Cancer Center, Buffalo, NY; 7Department of Preventive Medicine, University of Southern California, Los Angeles, CA; 8Department of Biomedical Informatics, The Ohio State University, Columbus, OH; 9Center for International Blood and Marrow Transplant Research, National Marrow Donor Program, Minneapolis, MN; and 10College of Veterinary Medicine, The Ohio State University, Columbus, OH

In their recent report in Blood Advances, Martínez-Laperche et al1 proposed a predictive model for the risk of acute and chronic graft-versus-host disease (GVHD) based on the selection of clinical variables and 25 single-nucleotide polymorphisms (SNPs), spanning different cytokines previously found relevant in the biology of GVHD. The study included 509 patients and their sibling donors from the Spanish Group for Hematopoietic Stem Cell Transplantation (GETH). We attempted to validate their SNP associations with acute GVHD (aGVHD) and nonrelapse mortality (NRM) in DISCOVeRY-BMT, a 2-cohort study of almost 3000 8/8 HLA-matched unrelated donor-recipient paired samples from individuals of European American ancestry.2-5

The analyses of the GETH data were performed using univariate logistic and least absolute shrinkage and selection operator (LASSO) models.6 Univariate logistic regression models for grade 2 to 4 aGVHD, grade 3 to 4 aGVHD, and NRM were constructed via regression of clinical and genetic variables. LASSO models were constructed for all outcomes using genetic and clinical variables from a training set of 85% of patients, with the remaining 15% comprising the test set. Risk scores were calculated using the LASSO outcomes, and recipients were stratified into high- and low-risk groups based on the proportion of events in the total data. To validate the clinical and genetic associations seen with grade 2 to 4 aGVHD, grade 3 to 4 aGVHD, and NRM, we constructed univariate clinical logistic regression models identical to those of Martínez-Laperche et al1 and LASSO models for grade 2 to 4 aGVHD, grade 3 to 4 aGVHD, and NRM. Twenty-four of the 25 SNPs tested were available for univariate and LASSO analyses in DISCOVeRY-BMT; rs9267487 was used as a surrogate for rs3619024 in TNF, because the SNPs are in linkage disequilibrium at r² = 0.82.7

Clinical characteristics of both GETH and DISCOVeRY-BMT data are listed in Table 1. In GETH data, univariate clinical models showed total-body irradiation and female donor/male recipient increased the odds of aGVHD and NRM, respectively (P < .05). In DISCOVeRY-BMT, total-body irradiation increased the odds of grade 2 to 4 aGVHD (odds ratio [OR], 1.3; 95% confidence interval [CI], 1.12-1.51; P = 5.9 × 10⁻⁴) and grade 3 to 4 aGVHD (OR, 1.22; 95% CI, 1.01-1.47; P = .04), whereas female donor/ male recipient was not associated with NRM at P < .05.

We selected 6 SNPs at P < .05, with 95% CIs that did not include 1, and tested them for association with aGVHD or NRM in DISCOVeRY-BMT. We assumed the same transmission models (dominant, additive, recessive, or codominant) for donors and recipients presented in the report by Martínez-Laperche et al1 and adjusted for significant clinical variables but not multiple comparisons. Donor GG/AG vs AA at rs3819024 was associated with NRM in DISCOVeRY-BMT cohort 1 at P < .05 (OR, 0.79; 95% CI, 0.64-0.99; P = .04) but not cohort 2 (OR, 0.89; 95% CI, 0.61-1.29; P = .53), although the effect direction was also risk reducing, as in the GETH data (Figure 1). The dominant models for rs16944, rs1143627, and rs2275913 trended toward a risk reduction in both cohorts in DISCOVeRY-BMT, with an OR of <1, whereas these genetic models showed an OR of >1 in the GETH cohort (Figure 1).

We assessed effect sizes detectable in DISCOVeRY-BMT and GETH for the 25 SNPs tested ranging in MAF from 0.4% to 45%. After correcting for testing in 22 (of 25) independent SNPs...
(α = 0.05/22 = 0.0023), DISCOVeRY-BMT was powered (β = 0.80) to detect hazard ratios from 4.0 to 1.2 (NRM and aGVHD 3-4) and 2.4 to 1.1 (aGVHD 2-4) for MAFs from 0.4% to 45%, respectively.3 We were therefore powered to validate the effect sizes detected in the Martínez-Laperche et al1 report.3 In contrast, the GETH cohort was powered (β = 0.80) to detect medium effect sizes (OR, 1.6) for a MAF of 45% in grade 2 to 4 aGVHD; however, for variants at the bottom end of the MAF range, the detectable OR was ~40, with further reduced power for NRM and grade 3 to 4 aGVHD outcomes. Compounding this sample size challenge was the fact that GETH SNP data were analyzed for all outcomes using logistic regression. For aGVHD, these events occur early on, and this model may be reasonable. However, failure to analyze overall survival or NRM with survival models can result in information loss, and the use of survival models is preferable.5,8-11

The well-described LASSO and prediction modeling performed in the GETH cohorts provides us the opportunity to consider important issues in association testing and prediction modeling of transplantation outcomes using genetic variables.6,12 Because of problems with interpretations of the proposed LASSO models in the report, we did not attempt validation of the exact model in DISCOVeRY-BMT. Specifically, the same SNPs were included under multiple transmission assumptions for all LASSO models. For example, in the LASSO model for grade 3 to 4 aGVHD, 7 of 11 SNPs were included in 2 different transmission models, and 1 SNP was included in...
Figure 1. Validation in DISCOVeRY-BMT of significant SNP associations in Martınez-Laperche et al report. This figure shows the ORs, 95% CIs, and P values as reported by Martınez-Laperche et al (literature) and validation of these associations in DISCOVeRY-BMT cohort 1 and cohort 2. Martınez-Laperche et al reported 6 SNPs with P < .05 and 95% CIs that did not include OR = 1, which are shown along the y-axis. The results are grouped by those shown in the Martınez-Laperche et al report (literature) and DISCOVeRY-BMT (cohort 1 and cohort 2), as indicated in the gray boxes displayed along the top of the figure. Each circle, square, and triangle represents an OR from regression analysis, with the shapes corresponding to grade 2 to 4 aGVHD, grade 3 to 4 aGVHD, or NRM, respectively. The colors represent P values of the analyses: .05 > P ≥ .01 (blue) and P > .05 (red). rs2275913 and rs3819024 are donor variants; the remaining 4 SNPs are recipient associations. *SNPs with a 5% and **SNPs with a 10% difference in minor allele frequency (MAF) between DISCOVeRY-BMT and GETH, respectively; all other MAFs were comparable between the groups. For example, (recipient) SNP rs4711998 was associated with increased risk of grade 2 to 4 aGVHD in GETH (blue and to the right of OR = 1) but resides on OR = 1 and P > .05 in DISCOVeRY BMT.

To consider the 25 SNPs in aggregate, we constructed 3 separate LASSO models for grade 2 to 4 aGVHD, grade 3 to 4 aGVHD, and NRM under an additive genetic model. All 25 recipient and donor SNPs for cohorts 1 and 2 were possible variables (this approach most closely represents Table 4 in the Martınez-Laperche et al report); each cohort was divided into a training (85%) and test (15%) set. To select the variables with the smallest prediction errors, we built logistic LASSO regression using fivefold cross validation to find the best penalty parameter λ, and we repeated this cross validation 50 times to give a robust estimate of λ by taking the median. The final LASSO model selected the SNPs with the best λ. LASSO models for 2 outcomes selected 1 SNP; however, the coefficients were almost 0 (≤ 5 × 10⁻¹⁰). Thus, for unrelated donor-recipient pairs, we concluded the 25 SNPs were not associated with either aGVHD or NRM and therefore did not pursue additional predictions. In addition, the predictive models specified in the report cannot be generalized to other transplantation patient cohorts or applied in a prospective setting, because high- and low-risk cut points were determined by the proportion of events in the GETH cohort. The successful stratification of risks for aGVHD and NRM outcomes in 1 cohort may not necessarily be carried over to another cohort when prediction is needed.

Our inability to validate the univariate associations or find SNPs predictive of either aGVHD or NRM may have been driven by the differences in transplantation type (related sibling vs unrelated donor), distribution of disease, genomic ancestry, and/or event rates between DISCOVeRY-BMT and GETH. It is important to consider that these single SNPs were initially identified as important in small expression studies or were selected because they reside in gene promoter regions. However, recent large-scale functional studies of SNPs in thousands of samples can now be leveraged. For example, rs3819024, although selected as an IL17A donor variant, is correlated with the expression of PAQR8, not IL17A, and only in whole blood, although tested in >70 tissues. Therefore, collectively, these variants may not be informative for the genes of interest. Irrespective of this, we must consider that the DISCOVeRY-BMT results show these SNP findings are not generalizable to other transplantation populations, and to start building successful prediction models of transplantation outcomes, we need larger homogeneous studies across multiple patient populations.

Acknowledgments: This work was supported by grants from the National Heart Lung and Blood Institute (NHLBI) (1R01HL102278)
and National Cancer Institute (NCI) (1R03CA188733), National Institutes of Health (NIH) (L.E.S.-C. and T.H.). E.K. is supported by the Pelotonia Foundation Graduate Student Fellowship. S.P. is supported by the NCI (R01CA168814), National Institute of Child Health and Human Development (R01HD074587 and U54HD090215), and Leukemia & Lymphoma Society (1293-15). The Center for International Blood and Marrow Transplant Research is supported by public health service grant/cooperative agreement 5U24-CAN076518 from the NCI, NHLBI, and National Institute of Allergy and Infectious Diseases; grant/cooperative agreement SU10HL069294 from the NHLBI and NCI; contract HHSN250201200016C with the Health Resources and Services Administration; grants N00014-15-1-0848 and N00014-16-1-2020 from the Office of Naval Research; grants from Alexion, Amgen Inc. (corporate member), Astellas Pharma US, AstraZeneca, Be the Match Foundation, Bluebird Bio Inc. (corporate member), Bristol-Myers Squibb Oncology (corporate member), Celgene Corporation (corporate member), Cellular Dynamics International Inc., Chimerix Inc. (corporate member), Fred Hutchinson Cancer Research Center, Gamida Cell Ltd., Genentech Inc., Genzyme Corporation, Gilead Sciences Inc. (corporate member), Jeff Gordon Children’s Foundation, Leukemia & Lymphoma Society, Medac GmbH, MedImmune, Medical College of Wisconsin, Merck & Co Inc. (corporate member), Mesoblast, MesoScale Diagnostics Inc., Miltenyi Biotec Inc. (corporate member), National Marrow Donor Program, Neovii Biotech NA Inc., Novartis Pharmaceuticals Corporation, Onyx Pharmaceuticals, Optum Health Solutions Inc., Otsuka America Pharmaceutical Inc., Otsuka Pharmaceutical Co Ltd.–Japan, Patient-Centered Outcomes Research Institute, Perkin Elmer Inc., Pfizer Inc., Sanofi US (corporate member), Seattle Genetics (corporate member), Spectrum Pharmaceuticals Inc. (corporate member), St. Baldrick’s Foundation, Sunesis Pharmaceuticals Inc. (corporate member), Swedish Orphan Biovitrum Inc., Takeda Oncology, Telomere Diagnostics Inc., University of Minnesota, and Wellpoint Inc. (corporate member); and an anonymous donation to the Medical College of Wisconsin.

Opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect those of the Pelotonia Fellowship Program or The Ohio State University. The views expressed in this article do not reflect the official policy or position of the National Institutes of Health, the Department of the Navy, the Department of Defense, the Health Resources and Services Administration, or any other agency of the US government.

**Conflict-of-interest disclosure:** S.P. has a patent on "Methods of detection of graft-versus-host disease" (US 20130115523A1, WO2013066369A3) licensed to Viracor-IBT Laboratories. The remaining authors declare no competing financial interests.

**ORCID profiles:** E.K., 0000-0002-0881-7424; S.P., 0000-0001-5571-2775; Q.Z., 0000-0003-2516-6675; L. Pooler, 0000-0001-5388-6640; X.S., 0000-0002-6844-2740; M.C.P., 0000-0003-1579-2293; P.L.M., 0000-0002-9577-3879.

**Correspondence:** Lara E. Sucheston-Campbell, The Ohio State University, 496 W 12th Ave, 604 Riffe Building, Columbus, OH 43210; e-mail: sucheston-campbell.1@osu.edu.

**References**


DOI 10.1182/bloodadvances.2019000052

© 2019 by The American Society of Hematology