Bone Marrow-Based Biomarkers for Predicting aGVHD Using Targeted RNA Next Generation Sequencing and Machine Learning

Maher Albitar, Hong Zhang, Andrew Pecora, Andrew IP, Andre Goy, Spiraggelos Antzoulatos, Ivan De Dios, Wanlong Ma, Sukhdeep Kaur, and Hyoung Suh, Michele Donato, Scott D. Rowley

Genomic Testing Cooperative, Irvine, CA and John Theurer Cancer Center, Hackensack Meridian Health

INTRODUCTION

Acute graft-vs.-host disease (aGVHD) remains a major diagnostic and clinical problem in patients after allogenic hematopoietic stem cell transplant (HSCT). Currently there are no reliable biomarkers for prediction of aGVHD. HLA disparities between donor and recipient, age of patient, conditioning intensity, and the type of disease being treated are some of the factors believed to be relevant for the development of aGVHD.

Other complexity of aGVHD is the difficulty in the diagnosis of this disease, especially when it is in early stage. Typically this disease target skin, liver, and gastrointestinal tract. Therefore, biopsies of these organs are currently the major diagnostic tools.

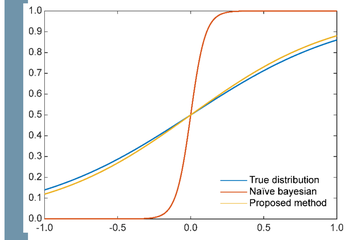
Finding biomarkers that play a role in aGVHD not only helps in predicting and diagnosing aGVHD, but might help in developing prophylaxis and therapeutic approaches. Using Next Generation Sequencing (NGS) and targeted RNA sequencing along with a machine learning approach to predict, we investigated the potential of discovering new biomarkers that can predict aGVHD.

AIM

Discovering new biomarkers and developing an assay to predict aGVHD using Next Generation Sequencing (NGS) and targeted RNA sequencing along with a machine learning algorithms.

RESULTS

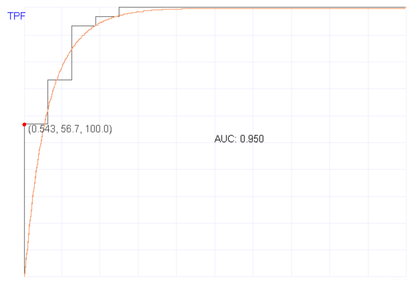
Number of patients	46
Age at Transplant	Median: 61, range: 26-76
Donor's age	Median: 31, range: 9-63
DX	
AML	16
MPAL	1
CMML	5
MDS	10
MPN	4
AA	1
ALL	9
Sex	
Female	24 (52%)
Male	22 (48%)
HLA	
Match	32 (69.5%)
Haplo	14 (30.5%)



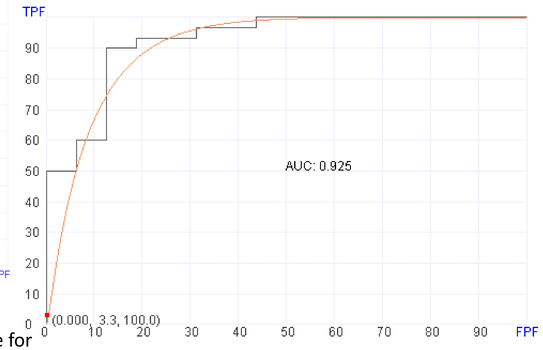
Using Geometric Mean Naïve Bayesian (GMNB) to eliminate the underflow problem commonly associated with the standard Naïve Bayesian classifiers for ranking the relevant biomarkers.

Using k-fold cross validation procedure (with k=10) for an individual gene the Naïve Bayesian classifier was constructed on the training of k-1 subsets and tested on the other testing subset for vinal model generating ROC curve

		95% Confidence Interval	
		Lower Limit	Upper Limit
Sensitivity	93%	76%	99%
Specificity	88%	60%	98%
PPV	93%	76%	99%
NPV	88%	60%	98%



Receiver operating characteristic (ROC) curve for prediction of aGVHD. The area under the curve (*AUC*) is 0.950



Receiver operating characteristic (ROC) curve for validating prediction of aGVHD using "leave one out, LOO). The area under the curve (*AUC*) is 0.925

Top 7 Biomarkers selected by the Bayesian model for predicting aGVHD

CIITA	Class II Major Histocompatibility Complex Transactivator
LMO7	Encodes a protein containing a calponin homology (CH) domain, a PDZ domain, and a LIM domain, and may be involved in protein-protein interactions.
IKZF3	Three members of this protein family (Ikaros, Aiolos and Helios) are hematopoietic-specific transcription factors involved in the regulation of lymphocyte development. This gene product is a transcription factor that is important in the regulation of B lymphocyte proliferation and differentiation.
CD19	Encodes a member of the immunoglobulin gene superfamily. Expression of this cell surface protein is restricted to B cell lymphocytes.
TCL1A	Expressed in CD4-/CD8- cells, but not in cells at later stages of differentiation. TCL1 functions as a coactivator of the cell survival kinase AKT
ERCC3	Encodes an ATP-dependent DNA helicase that functions in nucleotide excision repair. The encoded protein is a subunit of basal transcription factor 2 (TFIIH) and, therefore, also functions in class II transcription. Mutations in this gene are associated with Xeroderma pigmentosum B,
CD22	Mediates B-cell B-cell interactions. May be involved in the localization of B-cells in lymphoid tissues. Binds sialylated glycoproteins; one of which is CD45.

METHOD

RNA extracted from bone marrow aspiration samples collected around day 90 post HSCT from 46 patients were sequenced using 1408 targeted genes. cDNA was first generated, then adapters were ligated. 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 an Illumina NextSeq 550 platform. Ten million reads per sample in a single run were required. Read length was 2 × 150 bp. Expression profile was generated using Cufflinks. A machine learning system is developed to predict the GvHD cases and to discover the relevant genes. A subset of genes relevant to GvHD is automatically selected for the classification system, based on a k-fold cross validation procedure (with k=10). For an individual gene, a Naïve Bayesian classifier was constructed on the training of k-1 subsets and tested on the other testing subset. To eliminate the underflow problem commonly associated with the standard Naïve Bayesian classifiers, we applied Geometric Mean Naïve Bayesian (GMNB) as the classifier to predict GvHD. The processes of gene selection and GVHD classification are applied iteratively to obtain an optimal classification system and a subset of genes relevant to GvHD.

CONCLUSIONS

1. Targeted RNA profiling of bone marrow cells using NGS provides valuable information for predicting aGVHD
2. RNA Profiling along with machine learning algorithm shows that aGVHD can be predicted with high sensitivity and specificity.
3. RNA expression levels of 7 genes (CIITA, CD19, CD22, TCL1A, IKZF3, LMO7, and ERCC3) involved in MHC, B- and T-cell proliferation and interaction are adequate for the prediction of aGVHD.
4. Further testing of large number of cases is needed for validation of this approach

Just highlight this text and replace with your own text.

REFERENCES

1. Zeiser R, Blazar BR. Acute graft-*versus*-host disease - biologic process, prevention, and therapy. N Engl J Med 2017; 377: 2167–2179.
2. MacMillan ML, Robin M, Harris AC, et al. A refined risk score for acute graft-*versus*-host disease that predicts response to initial therapy, survival, and transplant-related mortality. Biol Blood Marrow Transplant 2015; 21: 761–767.
Srinagesh HK, Özbek U, Kapoor U, Ayuk F, Aziz M, Ben-David K, Choe HK, DeFilipp Z, Etra A, Grupp SA, Hartwell MJ, Hexner EO, Hogan WJ, Karol AB, Kasikis S, Kitko CL, Kowalyk S, Lin JY, Major-Monfried H, Mielke S, Merli P, Morales G, Ordemann R, Pulsipher MA, Qayed M, Reddy P, Reshef R, Rösler W, Sandhu KS, Schechter T, Shah J, Sigel K, Weber D, Wölfl M, Wudhikarn K, Young R, Levine JE, Ferrara JLM. The MAGIC algorithm probability is a validated response biomarker of treatment of acute graft-versus-host disease. Blood Adv. 2019 Dec 10;3(23):4034-4042. doi: 10.1182/bloodadvances.2019000791. PMID: 31816061; PMCID: PMC6963240.
3. Kohler S, Hendrickson MR, Chao NJ, et al. Value of skin biopsies in assessing prognosis and progression of acute graft-*versus*-host disease. Am J Surg Pathol 1997; 21: 988–996
4. Kanis JA, Johansson H, Oden A, et al. A meta-analysis of prior corticosteroid use and fracture risk. J Bone Miner Res 2004; 19: 893–899.
5. van Staa TP, Leufkens HG, Cooper C, et al. The epidemiology of corticosteroid-induced osteoporosis: a meta-analysis. Osteoporos Int 2002; 13: 777–787.
6. Weinstein RS. Glucocorticoid-induced osteonecrosis. Endocrine 2012; 41: 183–190.
7. Socié G, Vigouroux S, Yakoub-Agha I, et al. A phase 3 randomized trial comparing inolimomab vs usual care in steroid-resistant acute GVHD. Blood 2017; 129: 643–649.
8. MacMillan ML, DeFor TE, Weisdorf DJ. The best endpoint for acute GVHD treatment trials. Blood 2010; 115: 5412–5417.
9. Martin PJ, Bachier CR, Klingemann H-G, et al. Endpoints for Clinical trials testing treatment of acute graft-*versus*-host disease: a consensus document. Biol Blood Marrow Transplant 2009; 15: 777.

CONTACT INFORMATION

Maher Albitar, MD
Genomic Testing Cooperative, LCA.
175 Technology Dr. #100, Irvine, CA 92618 USA
Phone: 657-202-5950/FAX: 949-301-9719
Mobile: 949.275.7564
malbitar@genomictestingcooperative.com

Jennifer Varca
Genomic Testing Cooperative
175 Technology Drive, Suite 100
Irvine, Ca 92618
Phone: 714-401-3069
jvarca@genomictestingcooperative.com