Determining Clinical Course of Diffuse Large B-cell lymphoma Using Targeted Transcriptome and Machine Learning Algorithms

Maher Albitar, Hong Zhang, Andre Goy, Zijun Y. Xu-Monette, Govind Bhagat, Carlo Visco, Alexandar Tzankov, Xiaosheng Fang, Feng Zhu, Karen Dybkaer, April Chiu, Wayne Tam, Youli Zu, Eric D. Hsi, Fredrick B. Hagemeister, Jooryung Huh, Maurilio Ponzoni, Andrés J.M. Ferreri, Michael B. Møller, Benjamin M. Parsons, J. Han van Krieken, Miguel A. Piris, Jane N. Winter, Yong Li, Bing Xu, and Ken H. Young

Genomic Testing Cooperative, LCA, Irvine, CA 92618, USA; John Theurer Cancer Center at Hackensack University Medical Center, Hackensack, New Jersey, NJ 07601, USA; Duke University Medical Center, Durham, NC 27710, USA; Columbia University Medical Center, New York, NY 10027, USA; University of Verona, Verona 37129, Italy; Institute of Pathology, University Hospital Basel 4054, Switzerland ; alborg University Hospital, Aalborg 5000–5270, Denmark; Mayo Clinic, Rochester, MN 55905, USA; Weill Medical College of Cornell University, New York, NY 10065, USA; The Methodist Hospital, Houston, TX 77030, USA; Wake Forest University Medical Center, NC 77055, USA; The University of Texas MD Anderson Cancer Center, Houston, TX 22030, USA; Asan Medical Center, Ulsan University College of Medicine, Seoul 05505, Korea; San Raffaele H. Scientific Institute, Milan 20132, Italy; Odense University Hospital, Odense 5000–5270, Denmark; Gundersen Lutheran Health System, La Crosse, WI 54601, USA; Radboud University Nijmegen Medical Centre, Nijmegen 6500, Netherlands; Hospital Universitario Marqués de Valdecilla, Santander 39008, Spain; Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA; Baylor College of Medicine, Houston, TX 77030, USA; The First Affiliated Hospital of Xiamen University, Xiamen, Fujian 361004, China
Duke Cancer Institute, Durham, NC 27710, USA

INTRODUCTION

Multiple studies have demonstrated that diffuse large B-cell lymphoma (DLBCL) can be divided into subgroups based on their biology. However, these biological subgroups overlap clinically. While R-CHOP (rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone) remains the standard of care for treating patients with DLBCL, predicting which patients will not benefit from such therapy is important so that alternative therapy or clinical trials can be considered. Most of the studies stratifying patients select biomarkers first, then explore how these biomarkers can stratify patients based on outcome. We explored the potential of using machine learning to first group patients with DLBCL based on survival, then isolating the biomarkers necessary for predicting these survival subgroups.

AIM

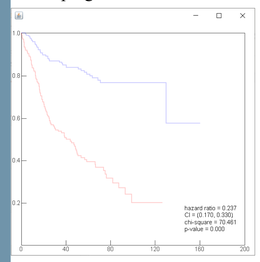
Using machine learning we sought to optimally stratify patients with DLBCL treated with R-CHOP based on their overall survival. Using another machine learning approach, we select RNA biomarkers generated from targeted NGS to predict these survival subgroups accurately.

METHOD

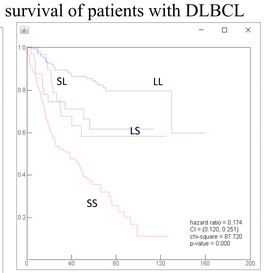
RNA was extracted from tissue paraffin blocks from 379 R-CHOP treated patients with *de novo* DLBCL, and from 247 patients with extranodal DLBCL. A targeted hybrid capture RNA panel of 1408 genes was used for next generation sequencing (NGS). Sequencing was performed using an Illumina NextSeq 550 System platform. Ten million reads per sample in a single run were required, and the read length was 2 × 150 bp. An expression profile was generated from the sequencing coverage profile of each individual sample using Cufflinks. A machine learning system was developed to classify patients into four groups based on their overall survival. This machine learning approach based on naïve Bayesian algorithm was also used to discover the relevant subset of genes with which to classify patients into each of the four survival groups. 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 the survival group for each patient.

RESULTS

Developing a naïve model for the survival of patients with DLBCL



Step 1

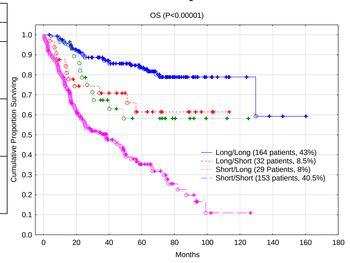


Step 2

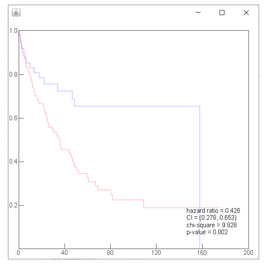
Patient Characterstics

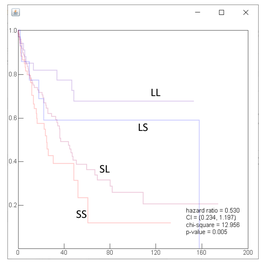
Characterstics	Number (%)
No. of patients	379
Age	
<60	164 (43%)
≥60	215 (57%)
Sex	
Male	210 (55%)
Female	169 (45%
Cell of origin	
GCB	198 (52%)
ABC	181 (48%)

Selecting 180 biomarkers for predicting survival groups using machine learning and validation of the prediction model

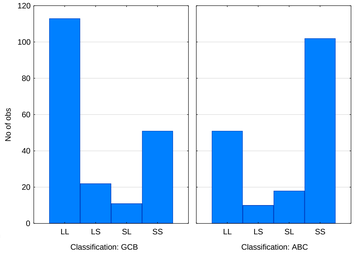


Validation of the approach using an independent group of patients

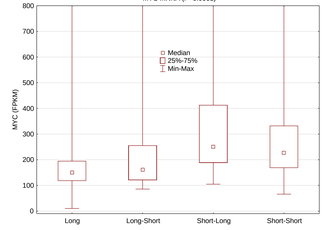




Correlation with cell of origin (COO) classification and other clinical prognostic markers



Levels of MYC mRNA in various survival groups



Multivariate survival analysis

N=379	Beta	Standard Error	Beta/Coefficient 95% lower	Beta/Coefficient 95% upper	p	Hazard ratio	Hazard ratio 95% lower	Hazard ratio 95% upper
Covariates: Survival groups, cell of origin, and IPI (>2)								
Survival classification	0.55	0.07	0.40	0.69	0.000000	1.73	1.49	2.00
GCB vs ABC	-0.03	0.18	-0.37	0.32	0.869145	0.97	0.69	1.37
IPI	0.88	0.17	0.54	1.21	0.000000	2.41	1.72	3.36
Covariates: Survival groups, IPI (>2), cell of origin, and TP53 mutation								
Survival classification	0.53	0.08	0.39	0.68	0.000000	1.71	1.47	1.98
IPI	0.84	0.17	0.50	1.18	0.000001	2.32	1.66	3.24
COO Classification	0.04	0.18	-0.31	0.40	0.816543	1.04	0.73	1.49
Mute.TP53	0.37	0.19	0.00	0.73	0.048156	1.44	1.00	2.08
Covariates: Survival groups, IPI (>2), cell of origin, Mutations in MYD88, CD79B, and TP53 mutation								
Survival classification	0.54	0.08	0.40	0.69	0.000000	1.72	1.49	2.00
IPI	0.87	0.17	0.53	1.21	0.000000	2.39	1.71	3.36
COO Classification	0.13	0.19	-0.24	0.50	0.491261	1.14	0.79	1.64
Mute.MYD88	-0.45	0.22	-0.88	-0.02	0.041843	0.64	0.41	0.98
Mute.CD79B	0.06	0.32	-0.55	0.68	0.841714	1.06	0.57	1.98
Mute. TP53	0.38	0.19	0.01	0.74	0.044687	1.46	1.01	2.10
Covariates: Survival groups, IPI (>2), cell of origin, TP53 mutation, and MYC expression (above upper 25 percentile)								
Survival classification	0.54	0.08	0.39	0.69	0.000000	1.71	1.47	1.99
IPI	0.84	0.17	0.51	1.18	0.000001	2.32	1.66	3.24
Classification	0.04	0.18	-0.31	0.40	0.816151	1.04	0.73	1.49
Mute.TP53	0.37	0.19	0.00	0.74	0.048720	1.45	1.00	2.11
MYC U25%	-0.03	0.18	-0.39	0.33	0.878706	0.97	0.68	1.39
Covariates: Survival groups, IPI (>2), cell of origin, TP53 mutation, and MYC expresion (continuous variable)								
Survival classification	0.55	0.08	0.41	0.70	0.000000	1.74	1.50	2.02
IPI	0.85	0.17	0.52	1.19	0.000001	2.35	1.68	3.28
Classification	0.03	0.18	-0.33	0.38	0.886603	1.03	0.72	1.46
Mute.TP53	0.41	0.19	0.04	0.78	0.028204	1.51	1.04	2.18
MYC	0.00	0.00	0.00	0.00	0.150307	1.00	1.00	1.00
Covariates: Survival groups, IPI (>2), cell of origin, TP53 mutation, and expression of MYC and IRF4 (continuous)								
Survival Classification	0.59	0.08	0.43	0.74	0.000000	1.80	1.54	2.09
IPI	0.85	0.17	0.51	1.18	0.000001	2.33	1.67	3.26
COO classification	0.21	0.21	-0.19	0.61	0.308746	1.23	0.82	1.84
Mute.TP53	0.43	0.19	0.06	0.80	0.022837	1.54	1.06	2.22
MYC mRNA	0.00	0.00	0.00	0.00	0.124518	1.00	1.00	1.00
IRF4 mRNA	0.00	0.00	0.00	0.00	0.066811	1.00	1.00	1.00

Cox proportional hazard regression multivariate model incorporating the survival classification with COO and the IPI (IPI ≤2 vs IPI >2), survival classification and IPI were the only independent predictors of survival. In this model, COO was no longer a predictor of survival. In a multivariate model incorporating age without IPI, age was significant independent predictor of survival (P = 0.01). Poor survival subgroup (SS) had significantly (P = 0.01) higher percentage of patients at age above 60. This raises the possibility that age and possible death from causes other than lymphoma--and not only biology-contribute to the poor survival in the SS subgroup.

CONCLUSIONS

1) Patients with DLBCL can be stratified to three groups with significantly different outcome upon treatment with R-CHOP
2) Expression profile of 180 genes using NGS is adequate for distinguishing between these three subgroups
3) Patients with intermediate outcome are heterogenous group and the use of different biomarkers is necessary for distinguishing these subgroups.
4) Only TP53 abnormalities and IPI remained independent predictor of survival in addition to the expression model.
5) Cell of origin, MYC and BCL2 abnormalities are weak predictor of survival and not independent of the expression model.

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

.