DEVELOPING ARTIFICIAL INTELLIGENCE-BASED TRANSCRIPTOMIC SIGNATURE FOR THE DIAGNOSIS OF DARK ZON LYMPHOMA IN PATIENTS WITHOUT MYC GENE REARRANGEMENT

Andrew IP[1], Maher Albitar[2], Hong Zhang[3], Sally Agersborg[2], Adam Albitar[2], Ahmed Charifa[2], Andrew Pecora[1], David Siegel[1], Lori Leslie[1], Andre Goy[1], Tatyana Feldman[1]

[1]John Theurer Cancer Center, Hackensack, NJ, United States, [2]Genomic Testing Cooperative, CA, United State, [3]Georgia Southern University, Statesboro, GA, United States.

INTRODUCTION

Double hit diffuse large B-cell lymphoma (DLBCL) and Burkitt lymphoma (BL) are currently defined as Dark zone lymphoma (DZL) because the tumor cells are believed to originate from the densely packed B-cells in germinal centers. These cases are associated with poor outcome and require more aggressive treatment protocols. It has been established that DZ lymphoma have specific expression profile that distinguish them from other types of DLBCL.

AIM

We explored the potential of using transcriptomic data with artificial intelligence (AI) to first establish an AI defined transcriptomic signature. Then we explored if this specific signature can be found in garden variety DLBCL that show no evidence MYC rearrangement. The presence of such signature in Any DLBCL should be the basis for classification of DZL and this should dectate specific therapeutic approach.

METHOD

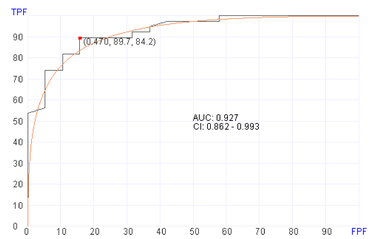
RNA was extracted from the lymph node samples of 363 cases with high-grade lymphoma. This included 40 cases double-hit lymphoma, 19 Burkitt lymphoma and 304 cases with DLBCLn. The RNA was sequenced by next generation sequencing (NGS) using a targeted RNA panel of 1600 genes. Hybrid capture sequencing library preparation was used and RNA was quantified using transcript per million (TPM). An independent set of the double-hit, Burkitt (total 59) and 117 DLBCLn is used to establish the DZ signature and a set of 187 DLBCLn was used for testing. Bayesian statistics were used to rank the genes that distinguish between DZL and DLBCLn, then eXtreme Gradient Boosting (XGBoost) was used to establish the DZL signature. Two thirds of the first set of cases were used for training and one third was used for testing the model. A score for the combination of relevant genes with a cut-off point was established that distinguish DZL from DLBCLn. The same Bayesian/XGBoost algorithm was used to test the rest of the DLBCLn cases and to stratify as DZ signature positive vs negative.

RESULTS

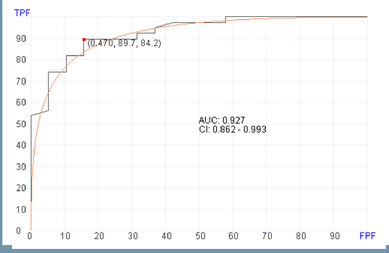
Development of DZL Signature:

Using 6 gene in XGBoost Model: MYC, SMARCA4, PATZ1, MEN1, TCF3, MSI2

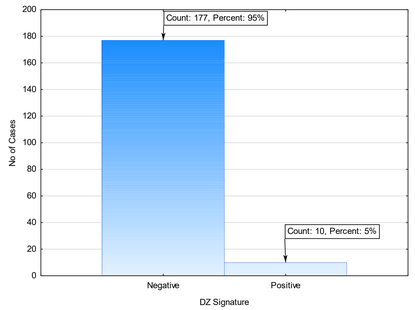
Testing set



Using 50 genes

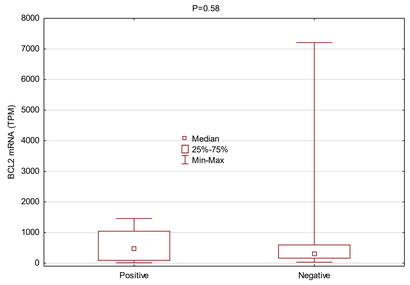


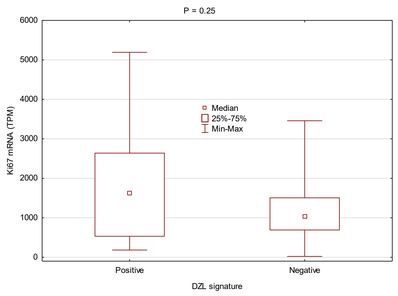
Using 50 gene with the AI model, we tested 187 without MYC rearrangement (DLBCLn) for the presence or absence of such signature. 10% were positive for the signature.



- MYC as the core oncogenic driver.
- MSI2: regulate translation and enhance MYC translation
- TCF3 : BHLH transcription factor enhances B-cell and stem-cell differentiation
- SMARCA4: chromatin accessibility enabling MYC program
- PATZ1: Transcription regulator enhances stem-cell differentiation
- MEN1 Tumor-suppressor interact with KMT2A for regulating HOX_MEIS signaling

No significant Difference in BCL2 or Ki67 between DZL-Like cases and DZL-negative cases by signature.





Top 30 genes showing significant difference between DZL signature-positive and DZL-signature negative.

Gene	DZL-positive (Median)	DZL-negative (Median)	LogFDR
BACH2	5.4311	3.7591	-Infinity
MYC	7.4032	6.0483	-Infinity
SMARCA4	6.6654	5.8382	-12.7785
MSI2	6.2137	5.3961	-12.0795
TCF3	6.2102	5.4161	-11.4775
TERT	3.9369	2.5159	-11.2733
CPSF6	6.8199	6.2642	-10.1754
RPS21	9.5457	8.85	-10.1647
MFNG	5.5533	4.8084	-8.89809
SRSF2	6.6024	6.2055	-8.65517
RBM15	6.4792	5.9494	-8.26347
SSBP2	4.5936	3.484	-8.21019
PATZ1	5.286	4.76	-7.90916
CD24	6.1582	4.9384	-7.80987
STAT3	5.1676	5.9661	-7.72984
CCNB1IP1	5.2853	4.7319	-7.72875
H1-4(HIST1H1E)	9.1831	8.6127	-7.72328
BCR	5.256	4.2703	-7.65465
MBTD1	4.6729	4.1504	-7.55094
BCL7A	4.5674	3.6658	-7.49587
CD38	6.0648	5.2892	-7.40251
EXOSC6	4.1225	3.4991	-7.36271
BCL2A1	5.148	6.6693	-7.29286
ACVR1	2.4289	4.2259	-7.24526
MAP2K7	4.2062	3.861	-7.20679
SUGP2	6.6376	6.1665	-7.17605
WDR90	4.4302	3.8685	-7.14626
USP7	6.3547	5.9106	-7.1077
RPL22	7.4487	6.9128	-7.06441
SRSF3	6.7016	6.392	-6.89333

CONCLUSIONS

- DZL can be defined by a specific transcriptomic signature irrespective of the presence or absence of MYC gene rearrangement
- 5% of DLBCL (14% DZL) are misdiagnosed and not called DZL if diagnosis is strictly based on FISH classification.
- Using RNA expression of 50 genes in AI model provides a reliable approach to identify these DLBCL patients
- While 6 genes or 50 genes are adequate to reliably distinguish DZL from other types of DLBCL, more than 150 gene show significant (LogFDR <-2) difference between DZL and the rest of DLBCL
- BCL2 and Ki67 are not significantly different between DZL and the rest of DLBCL.
- Most of the DZL cases showed GCB cell of origin
- Clinical trials using escalated therapy for patients with DZL signature are needed to confirm the value of using transcriptomic signature for the classification of DLBCL.

CONTACT INFORMATION

Maher Albitar, MD

Genomic Testing Cooperative, LCA.

25371 Commercentre Dr., Lake Forest, CA 92630

Phone: 657-202-5950/FAX: 949-301-9719

Mobile: 949.275.7564

malbitar@genomictestingcooperative.com