LYMPHOMA DIAGNOSIS AND CLASSIFICATION USING NEXT GENERATION SEQUENCING OF 30 CD MARKERS AND MACHINE LEARNING AS AN ALTERNATIVE TO IMMUNOHISTOCHEMISTRY

Maher Albitar[1], Hong Zhang[1], Andrew Ip[2], Wanlong Ma[1], James McCloskey[2], Kathrine Lender[2], Jeffrey Estella[1], Jamie Koprivnikar[2], Noa Biran[2], David Siegel[2], Ahmad Charifa[1], Arash Mohtashamian[1], Andrew Pecora[2], and Andre Goy[2]

[1]Genomic Testing Cooperative, LCA, Irvine, CA 92618, USA . [2]John Theurer Cancer Center at Hackensack University Medical Center, Hackensack, NJ, USA

INTRODUCTION

Lymphoma diagnosis and classification requires pathologist interpretation of morphology and large number of immunohistochemistry (IHC) stains of various CD markers. This process is subjective and requires a significant amount of tissue. In contrast RNA quantification of the same CD markers used in IHC using next generation sequencing (NGS) requires little tissue and less influenced by the antigen retrieval process used in IHC. However, IHC staining and microscopic examination allows evaluation of the expression in various subpopulations and makes diagnosis possible. In contrast when total RNA is evaluated by NGS distinguishing between subpopulation is lost. Machine learning algorithms are capable of multimarker normalization and compensating for the loss of subpopulation analysis.

AIM

We used NGS RNA data generated from FFPE samples along with machine learning algorithm to explore the potential of obtaining information that are similar to those obtained by IHC and microscopic examination. We quantified 30 CD markers typically used by routine IHC using NGS and examined the potential of this data in the diagnosis and classification of various types of lymphoma.

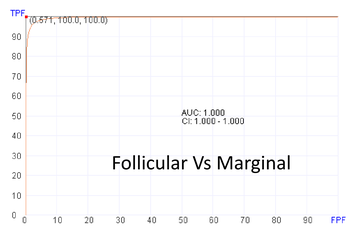
METHOD

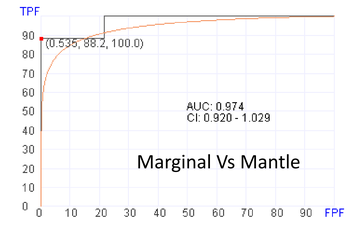
Formalin-fixed paraffin-embedded (FFPE) tissue from 130 diffuse large -cell lymphoma(DLBCL), 70 mantle cell lymphoma, 92 T-cell lymphoma, 48 follicular lymphoma, 36 Hodgkin lymphoma, and 52 marginal zone lymphoma were used for extracting mRNA. The studied samples were consecutive without selection and included mainly lymph node excisional biopsies or core biopsies. RNA sequencing was performed using a targeted hybrid capture panel that included CD1A, CD2, CD3D, CD3E, CD3G, CD4, CD5, CD7, CD8A, CD8B, CD10, CD14, CD19, CD20, CD22, CD33, CD34, CD38, CD40, CD44, CD47, CD68, CD70, CD74, CD79A, CD79B, CD81, CD138, CD200, CD274 genes. Salmon v1.4.0 software is used for expression quantification (TPM). Random Forest machine learning algorithm is used for predicting diagnosis. Randomly selected two thirds of samples were used for training and one third was used for testing.

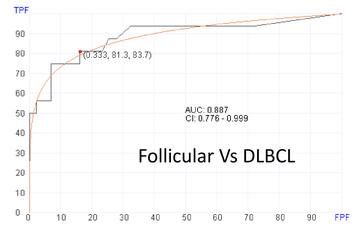
RESULTS

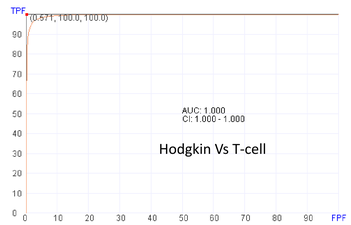
Example of a diagnostic FFPE samples showing RNA expression levels that are diagnostic

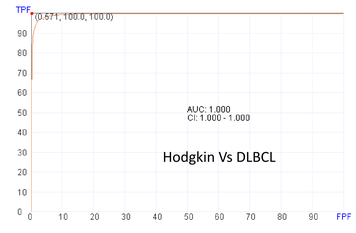
	Normal LN	DLBCL	Mantle	Marginal	Follicular	Double Hit	CLL	Plasmablastic	Hodgkin	T-Cell
CD57	63	0	5	6	15	25	1	2	0	6
BCL2	241	801	2812	929	1570	3803	1223	26	240	648
BCL6	464	363	101	503	834	2747	743	587	446	847
CCND1	370	469	8219	787	261	338	282	1282	279	1106
CD19	415	75	1000	1361	580	575	1387	18	172	118
CD2	2364	1138	113	969	593	1362	882	1311	1	1158
CD22	3371	1079	3289	5025	5245	4400	7868	95	766	865
CD274	210	182	62	101	33	158	64	287	319	247
CD3D	2137	263	436	637	361	851	765	302	0	3086
CD3E	1065	215	63	414	173	458	510	2222	1	669
CD3G	506	175	116	183	120	268	201	118	0	1183
CD4	669	867	405	833	250	749	428	636	1	790
CD5	469	584	421	257	82	180	1007	47	0	887
CD7	237	82	172	125	91	147	97	292	0	136
CD79A	1523	1391	5969	3626	2336	5000	7123	60	706	652
CD79B	621	643	1489	1523	1192	2493	2432	253	266	242
CD8A	244	166	41	73	69	261	48	219	180	68
CD8B	174	54	18	19	37	113	40	149	0	41
CD23	159	0	3	934	167	7	3205	3	0	87
CD25	444	43	80	527	85	74	72	445	460	4897
IRF4(MUM1)	302	302	495	852	265	100	518	1504	966	274
MKI67	871	762	1524	314	434	1557	64	1170	258	1062
MME(CD10)	62	108	3	33	236	258	3	45	0	47
CD20	4476	2486	6358	6304	4732	3357	4208	75	1	1882
MYC	306	558	780	272	166	1741	178	1338	351	683
CD56	97	31	9	58	37	6	4	528	9	309
CD138	66	15	94	422	262	5	1	7031	1	50
SOX11	10	0	291	0	0	0	0	0	0	1
BCMA	56	0	14	132	79	182	70	1331	175	13
CD30	31	11	1	9	8	4	6	6	0	35

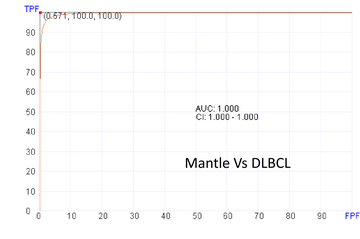












Using machine learning (random Forest) for precise and high sensitivity prediction of diagnosis. Two third were used for training and one third of cases were used for testing.
The study included 130 diffuse large B-cell lymphoma(DLBCL), 70 mantle cell lymphoma, 92 T-cell lymphoma, 48 follicular lymphoma, 36 Hodgkin lymphoma, and 52 marginal zone lymphoma.

CONCLUSIONS

- NGS quantification and quantification of RNA extracted from FFPE samples provides reliable data for evaluating the expression of various CD markers typicall used in IHC studies.
- RNA data from 30 CD markers when combined with machine learning are adequate for reliable classification of various types of lymphoma.
- The use quantitative RNA along with machine learning may resolve the diagnostic difficulties frequently pathologists have to deal with especially when tissue sample is scant (needle aspiration) or crushed.
- This technology can be automated and less susceptible to human errors and does not require high level of specialization
- This technology can be automated and less susceptible to human errors. RNA quantification using NGS has the potential to replace the need for IHC reducing cost and preserving precious tissue samples.

REFERENCES

1. Zhang H, Qureshi MA, Wahid M, Charifa A, Ehsan A, Ip A, De Dios I, Ma W, Sharma I, McCloskey J, Donato M, Siegel D, Gutierrez M, Pecora A, Goy A, Albitar M. Differential Diagnosis of Hematologic and Solid Tumors Using Targeted Transcriptome and Artificial Intelligence. Am J Pathol. 2023 Jan;193(1):51-59. doi: 10.1016/j.ajpath.2022.09.006. Epub 2022 Oct 13. PMID: 36243045.

2. Xu-Monette ZY, Zhang H, Zhu F, Tzankov A, Bhagat G, Visco C, Dybkaer K, Chiu A, Tam W, Zu Y, Hsi ED, You H, Huh J, Ponzoni M, Ferreri AJM, Møller MB, Parsons BM, van Krieken JH, Piris MA, Winter JN, Hagemeister FB, Shahbaba B, De Dios I, Zhang H, Li Y, Xu B, Albitar M, Young KH. A refined cell-of-origin classifier with targeted NGS and artificial intelligence shows robust predictive value in DLBCL. Blood Adv. 2020 Jul 28;4(14):3391-3404. doi: 10.1182/bloodadvances.2020001949. PMID: 32722783; PMCID: PMC7391158.

CONTACT INFORMATION

Maher Albitar, MD
Genomic Testing Cooperative, LCA.
Phone: 657-202-5950/FAX: 949-301-9719
malbitar@genomictestingcooperative.com