USING NEXT GENERATION SEQUENCING OF FLOW CYTOMETRY CD MARKERS AND MACHINE LEARNING AS A REPLACEMENT TO FLOW CYTOMETRY ANALYSIS FOR THE DIAGNOSIS OF HEMATOLOGIC NEOPLASMS

Maher Albitar[1], Hong Zhang[1], Andrew Ip[2], Wanlong Ma[1], Lori Leslie[2], Tatyana Feldman[2], Jeffrey Estella[1], 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

Flow cytometry performs multi-parameter analysis of cells and analyzes surface and intracellular markers for accurate phenotypic characterization of a cell population. Flow cytometry is used extensively in the diagnosis and classification of various hematologic neoplasms. However, analysis of the generated data is time consuming and remains subjective requiring special skill and experience. Furthermore, some diagnostic classes, such as myeloproliferative neoplasms (MPN) and myelodysplastic syndrome (MDS), are difficult to diagnose using flow cytometry. The RNA levels of the CD markers used in flow cytometry can be reliably quantified using next generation sequencing (NGS). However, when all cells are jointly sequenced, studying subpopulation of cells is lost, which hinders accurate diagnosis. Machine learning algorithms are capable of multimarker normalizing and compensate for the loss of subclonal analysis..

AIM

To explore and validate the potential of using RNA expression profiling with machine learning in providing information similar to these provided by flow cytometry. We used the expression of 30 CD markers as determined by RNA sequencing of bone marrow or peripheral blood samples in random forest platform and compared findings with actual flow cytometry based diagnostic findings.

METHOD

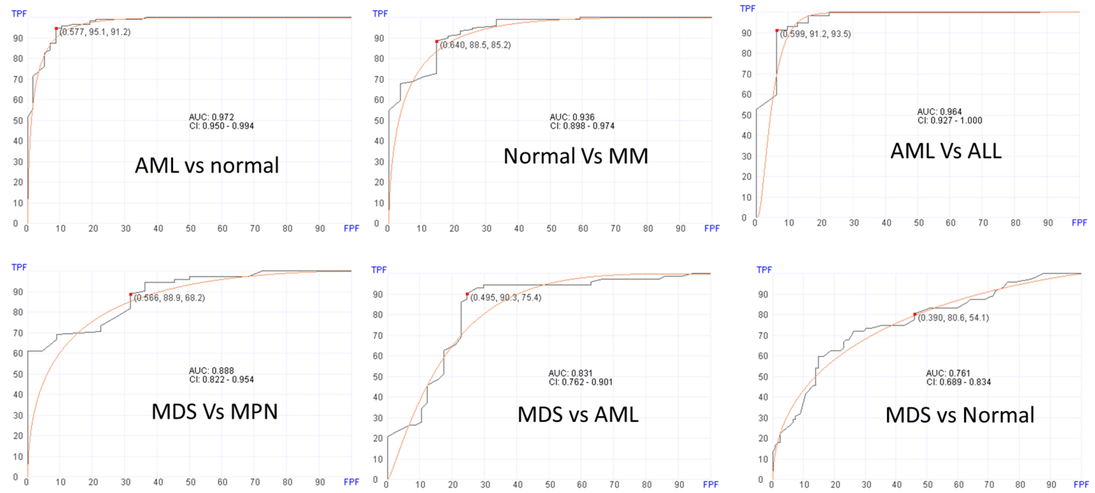
RNA was extracted from fresh bone marrow and peripheral blood samples from 172 acute myeloid leukemia (AML), 369 normal control, 68 MPN, 218 MDS, 93 acute lymphoblastic leukemia (ALL), 74 chronic lymphocytic leukemia (CLL), 38 mantle cell lymphoma, and 83 multiple myeloma. The samples were consecutive and collected without selection. 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). Machine learning algorithm (Random forest) is used for classifying diseases. Two thirds of samples were used for training the random forest algorithm and one third was used for testing.

RESULTS

Example of a diagnostic samples with various types of hematologic neoplasms showing RNA expression levels that are diagnostic

	Normal	B-ALL	T-cell	DLBCL	T-ALL	Mantle CD5-neg	CLL	AML-Mixed lineage
ANPEP (CD13)	1871	598	58	141	63	51	207	557
BCL2	227	1265	121	296	958	1557	6786	1064
CCND1	3	5	19	70	33	2391	12	14
CD14	715	40	129	384	410	167	11	1603
CD19	1	435	29	1004	42	1022	935	136
CD1A	1	0	1	15	1	46	0	5
CD2	30	98	1045	277	353	87	31	118
CD200	2	142	15	60	458	37	499	218
CD22	8	180	29	1149	92	2236	635	43
CD33	403	39	18	132	65	65	51	852
CD34	9	736	25	58	2159	28	4	1199
CD38	204	109	40	109	749	82	61	91
CD3D	107	400	1154	159	1456	59	50	263
CD3E	105	194	637	104	268	50	65	165
CD3G	33	83	358	70	736	14	14	71
CD4	236	37	89	323	184	43	27	237
CD5	39	184	1103	61	145	70	197	44
CD7	50	135	12	81	2232	42	29	92
CD79A	13	958	172	3132	299	4762	2704	161
CD79B	71	797	191	2238	221	955	268	370
CD8A	11	127	13	96	21	37	15	25
CD8B	16	80	8	84	9	86	12	30
CRLF2	5	24	1	18	23	12	3	10
DNTT(tdt)	1	4665	1	16	16	49	10	402
FCER2(CD23)	4	45	22	39	6	677	2255	108
FCGR1A(CD64)	202	28	14	28	6	49	6	418
FCGR3A(CD16)	672	118	14	30	10	33	23	500
KIT (CD117)	12	2	3	31	9	24	4	458
MKI67	1568	530	358	1270	1496	136	87	73
MME(CD10)	181	1834	12	302	16	4	49	43
MPO	8678	608	1	37	3	51	872	4613
MS4A1(CD20)	1	176	89	1620	178	4451	691	52
MYC	253	202	124	497	373	1406	78	1909
NCAM1(CD56)	14	6	2	17	1	20	2	18
SDC1(CD138)	5	16	33	52	24	57	1	1
TNFRSF17(BCMA)	2	3	25	72	28	122	21	1
TNFRSF8(CD30)	2	2	102	32	9	23	1	19
ZAP70	64	493	461	122	106	114	429	180

In samples with partial involvement machine learning is needed when for precise and high sensitivity prediction of diagnosis. Using random forest two third were used for training and one third is used for testing.



Receiver operating characteristic (ROC) curve of the testing set for classifying MPN Vs MDS/AML. TPF, true positive fraction (sensitivity); FPF, false positive fraction (specificity).

CONCLUSIONS

1. In diagnostic samples, NGS quantification of RNA from 30 CD markers can reliably be used to evaluate hematologic neoplasms and be used as a replacement to flow cytometry a fashion similar to flow cytometry.
2. In cases with low level neoplastic cells, NGS data of RNA of 30 CD markers when combined with machine learning is adequate for reliable diagnosis of various types of hematologic neoplasms.
3. Using NGS data of RNA from 30 CD markers when combined with machine learning can provide differential diagnosis in cases flow cytometry typically not useful such as differentiating between MDS and normal or myeloproliferative neoplasms.
4. This technology can be automated and less susceptible to human errors and does not require high level of specialization.

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 INFO

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

Jennifer Varca
Vice President, Business Development
Genomic Testing Cooperative
Phone: 714-401-3069
jvarca@genomictestingcooperative.com