Predicting PD-L1 Status in Solid Tumors Using Transcriptomic Data and Artificial Intelligence Algorithms

#4337

Ahmad Charifa, MD[1]; Alfonso Lam, PhD[1]; Hong Zhang, MD[1]; Andrew Ip, MD[2]; Andrew Pecora, MD[2] Stanley Waintraub, MD[2]; Deena Graham, MD[2]; Donna McNamara, MD[2]; Martin Gutierrez MD[2]; Andrew Jennis, MD[2]; Ipsa Sharma, Ms[1]; Jeffrey Estella, Ms[1]; Wanlong Ma, Ms[1]; Andre Goy, MD[2]; Maher Albitar, MD[1]
1 Genomic Testing Cooperative, Irvine, CA, 2 John Theurer Cancer Center at Hackensack University Medical Center, Hackensack, NJ

Introduction

PD-L1 immunohistochemistry (IHC) is routinely used to predict the clinical response to immune checkpoint inhibitors (ICIs); however, multiple assays and antibodies have been used. This study aimed to evaluate the potential of targeted transcriptome and artificial intelligence (AI) to determine PD-L1 RNA expression levels and predict the ICI response compared to traditional IHC.

Methods and Materials

RNA from 396 solid tumors samples was sequenced using next-generation sequencing (NGS) with a targeted 1408-gene panel. RNA expression and PD-L1 IHC were assessed across a broad range of PD-L1 expression levels. The geometric mean naïve Bayesian (GMNB) classifier was used to predict the PD-L1 status. PD-L1 RNA levels assessed by NGS demonstrated robust linearity across high and low expression ranges, and those assessed using NGS and IHC (Tumor cells (TC), The samples were selectively enriched for 1408 cancer-associated genes. cDNA was generated from the cleaved RNA fragments using random primers during first- and secondstrand synthesis. The sequencing adapters were ligated into the resulting double-stranded cDNA fragments. The coding regions of expressed genes were captured from this library using sequence-specific probes to create the final library. Sequencing was performed using an Illumina Novaseq (Illumina, San Diego, CA, USA). A minimum of ten million reads per sample was obtained in a single run, and the read length was 2 × 75 bp. An expression profile was generated from the sequencing coverage profile of each sample using the Cufflinks software. Expression levels were measured as fragments per kilobase of transcripts per million.

Pair of Variables	Number of cases (N)	R^2	p-value
CD19 & CD274	396	0.126502	0.000000
CD22 & CD274	396	-0.048253	0.000092
CD8A & CD274	396	0.242821	0.000000
Immune cells (IC) % & CD274	262	0.266951	0.000000
Tumor Cells (TC) % & CD274	313	0.568419	0.000000
Combined Positive Score (CPS) & CD274	346	0.605999	0.000000
CTLA4 & CD274	396	0.352428	0.000000
PDCD1 (PD-1) (CD279) & CD274	396	0.335572	0.000000
PDCD1LG2 (PD-L2) & CD274	396	0.872508	0.000000

Table 1. RNA sequencing provided in-depth information on the tumor microenvironment and immune response, including CD19, CD22, CD8A, CTLA4, and PD-L2 expression status.

Results

IHC test results	Variable	Cases (N)	Mean	Median	Range	Lower Quartile	Upper Quartile	10th Percentile	90th Percentile	Std. Dev.
TC<1%	CD274	223.00	4.49	2.97	0.00 - 25.99	1.79	5.73	1.07	9.16	4.32
TC>1%	CD274	90.00	14.87	10.11	0.62 - 77.53	4.60	18.62	2.95	32.35	15.75
TC<10%	CD274	267.00	5.17	3.36	0.00 - 72.72	1.94	6.43	1.14	10.71	6.14
TC>10%	CD274	46.00	20.81	14.03	2.90 - 77.53	8.67	26.32	4.21	51.02	17.33
IC<1%	CD274	133.00	3.95	2.66	0.00 - 25.99	1.71	4.03	0.99	7.95	4.56
IC>1%	CD274	129.00	9.38	5.43	0.29 - 77.53	3.21	10.31	1.74	19.05	12.13
IC<10%	CD274	226.00	5.69	3.03	0.00 - 72.72	1.92	6.09	1.15	12.25	8.24
IC>10%	CD274	36.00	12.50	8.52	0.29 - 77.53	4.72	13.80	4.18	31.83	13.95
TPS<1%	CD274	143.00	3.36	2.35	0.00 - 25.99	1.41	3.58	0.53	7.40	3.87
TPS>1%	CD274	207.00	10.19	5.63	0.29 - 133.81	3.03	11.65	1.74	22.25	14.74
TPS<10%	CD274	252.00	4.25	2.87	0.00 - 72.72	1.74	4.96	0.92	8.70	5.81
TPS>10%	CD274	98.00	15.50	9.67	0.29 - 133.81	4.91	18.74	2.90	31.87	18.57
TPS<30%	CD274	319.00	5.35	3.32	0.00 - 72.72	1.94	6.43	1.06	12.00	6.48
TPS>30%	CD274	31.00	28.50	18.62	2.90 - 133.81	11.47	35.94	6.67	52.69	27.31

Table 2. Correlation of the PD-L1 expression levels with the PD-L1 IHC results covering the high and low expression ranges (tumor cells (TC) <1%, >1%, <10%, and >10%; immune cells (IC) <1%, >1%, <10%, and >10%; and tumor proportion score (TPS) <1%, >1%, <10%, >10%, <30%, and >30%).

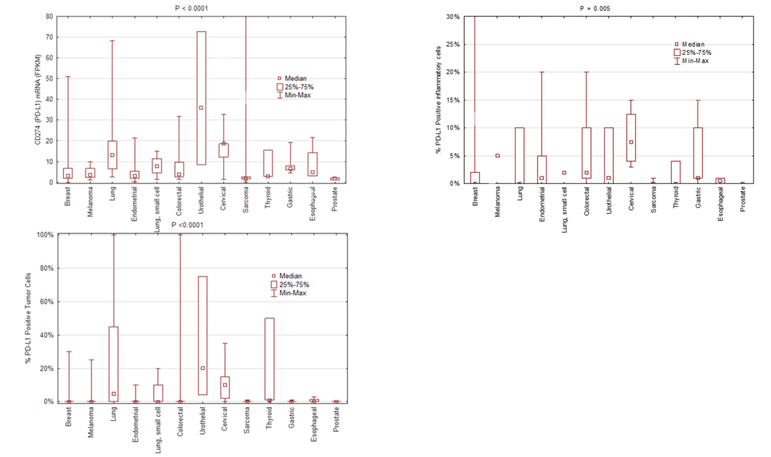

Figure1. Statistical tests demonstrated a significant correlation between PD-L1 RNA expression and IHC results across the various tumor types tested.

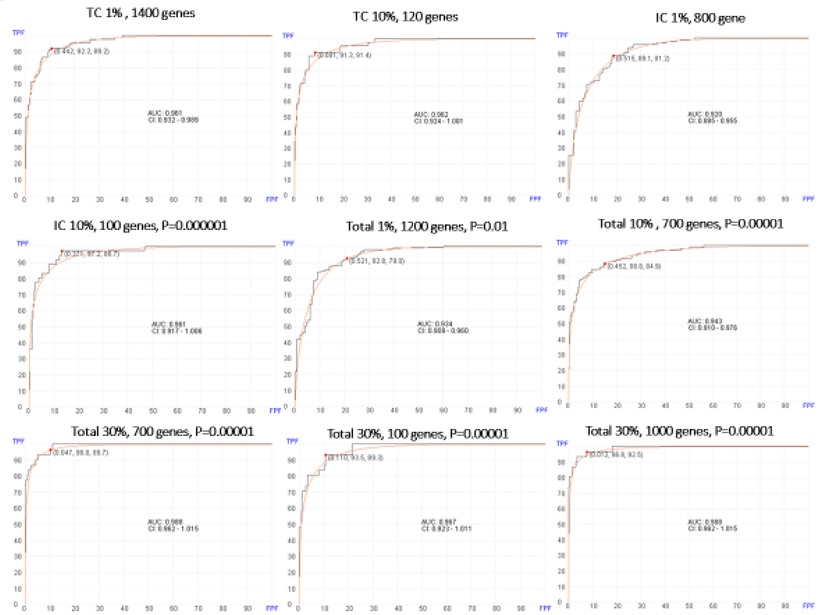

Figure 2. Machine learning showed high accuracy in predicting PD-L1 status, with the area under the curve varying between 0.988 and 0.920.

Conclusions

In summary, this study demonstrated the potential of combining AI with genomics in the routine practice of oncology and for predicting PD-L1 status. Therefore, clinical decisions can be made based on objective data. This approach was realized practically by developing a software that can be used to feed RNA data for automated quantification and prediction of biomarker status. Furthermore, this software and algorithms can be continuously trained by adding additional samples or new biomarkers. The limitation of this study was its retrospective nature. Therefore, PD-L1 status determined by RNA sequencing and machine learning algorithms needs to be further validated in future prospective clinical trials.