Cell-Free RNA in Liquid Biopsy and Biomarkers Profiling of Hematologic and Solid Tumors

Maher Albitar1, Hong Zhang1, Ahmad Charifa1, Andrew IP2, Ivan De Dios1, Wanlong Ma1, James McCloskey2, Michele Donato2, David Siegel2, Stanley Waintraub2, Martin Gutierrez2, Andrew Pecora2, Andre Goy2; 1 Genomic Testing Cooperative, LCA, Irvine, CA, 2 John Theurer Cancer Center at Hackensack University Medical Center, Hackensack, NJ, USA

Background:

Current use of liquid biopsy is based on cell-free DNA (cfDNA) and the evaluation of mutations or methylation pattern. However, expressed RNA can capture mutations, changes in expression levels due to methylation, and provide information on cell of origin, growth, and proliferation status. RNA is typically considered less stable than DNA and is more prone to degradation, especially in the peripheral blood circulation. However, the principle underlying NGS technology is to use fragmented DNA or RNA for sequencing in order to align these fragments against the reference genome. Using targeted transcriptome and analyzing cell-free RNA (cfRNA) by NGS is a practical approach that may change how liquid biopsies are currently performed and used in clinical settings. RNA expression profiling (transcriptome) provides important dynamic information on the tested tumor and the host response to the tumor.

Methods:

1 ml of plasma collected in EDTA was used for extracting RNA. The steps for processing cfRNA were as follows: 1) first-strand synthesis, 2) second-strand synthesis and A-tailing, 3) ligating adapters, 4) library amplification, 5) cleaning, 6) quantification, 7) preparing the 8-plex DNA sample library pool, 8) adding the designed KAPA target enrichment probes (1,459 genes, Supplementary Table S1) for overnight hybridization, 9) implementing KAPA beads in order to capture a multiplexed DNA library, 10) amplifying the enriched multiplex DNA library, 11) cleaning the amplified enriched multiplexed DNA library, 12) checking the multiplexed library by running TapeStation, 13) normalizing and pooling, and 14) denaturing and loading libraries on a NovaSeq 6000 sequencer.

For machine learning, we first selected genes that distinguish between two classes using standard Naïve Bayesian classifier on each gene with k-fold cross validation. After selecting individual genes, we used Naïve Bayesian classifier to distinguish between diagnostic classes using multiple selected genes using both confidence and P values. However, since Naïve Bayesian classifier suffers from severe numerical underflow problem when the dimension of data is high, we developed the Geometric Mean Naïve Bayesian (GMNB) classifier that eliminates the underflow problem by applying a multiplicative positive increasing function to the likelihood. The Geometric Mean Naïve Bayesian (GMNB) classifier was also used in classifying each sample against multiple diagnostic classes.

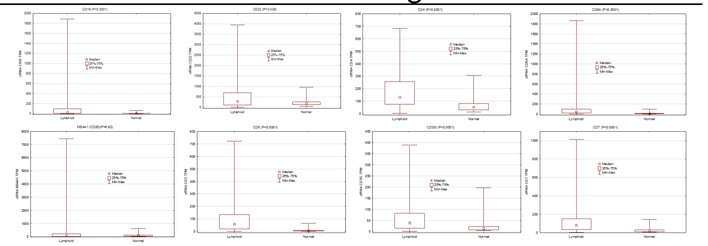
Conclusions:

- Targeted transcriptome using cfRNA is a reliable source of biomarkers to evaluate the underlying cancer and the host response
- Liquid Transcriptome can be used for diagnosis and staging, selecting therapy, predicting prognosis and monitoring disease.

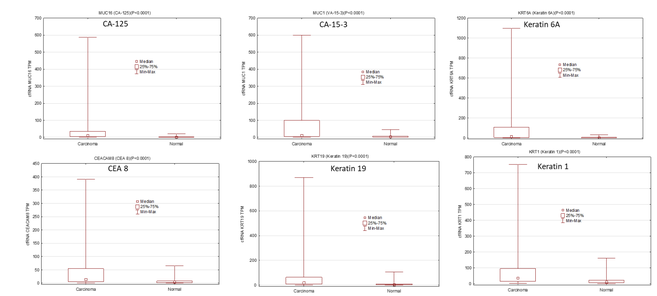
Results:

cfRNA contains numerous biomarkers reflecting tumor characteristics

Box plots showing various surface cell markers typically detected by flow cytometry (compared between B-cell neoplasms and normal control).

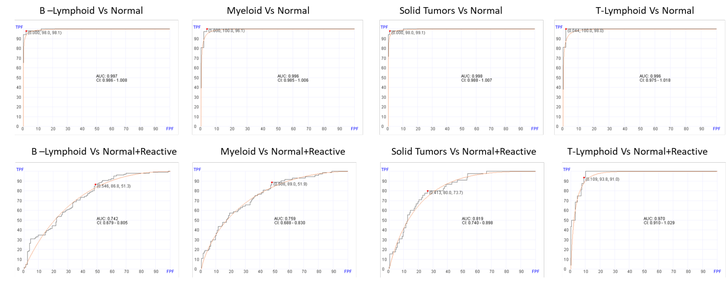


High levels of solid tumor biomarkers (CA-125, CA-15-3, Keratin 6A-3, CEA 8, Keratin 19, and Keratin 1) in solid tumors (compared with normal control)



Utilizing cfRNA in the differential diagnosis of neoplasms

Receiver operating characteristic (ROC) curve for the prediction of diagnoses using cfRNA according to AI-based models. The AUC and 95% CI are shown for various diagnostic classes. Only cfRNA expression levels were used in this modeling.
TPF, true positive fraction (sensitivity);
FPF, false positive fraction (specificity)



Predictions of diagnoses using cfRNA in an artificial intelligence (AI)-based model.

Diagnostic classes	Training					Leave one out			
	AUC	95% Confidence Interval	Sensitivity (%)	Specificity (%)	No of genes	AUC	95% Confidence Interval	Sensitivity (%)	Specificity (%)
Normal control Vs									
B-Lymphoid	0.997	0.986-1.0	98	98	60	0.988	0.966-1.0	98	96.2
Myeloid	0.996	0.985-1.0	100	96.1	30	0.994	0.981-1.0	100	96.1
Solid tumors	0.998	0.988-1.0	98	99.1	30	0.989	0.968-1.0	98	991
T-lymphoid	0.996	0.975-1.0	100	98	30	0.987	0.963-1.0	98	100
Reactive	0.998	0.988-1.0	98	100	10	0.995	0.982-1.0	98	99
Reactive+Normal Vs									
B-Lymphoid	0.742	0.679-0.805	86.8	51.3	100	0.693	0.627-0.760	83	49.4
Myeloid	0.759	0.688-0830	89	51	70	0.759	0.688-0.830	89	51.9
Solid tumors	0.819	0.740-0.898	80	73.7	90	0.781	0.696-0.866	80	67.3
T-lymphoid	0.97	0.910-1.0	93.8	91	100	0.877	0.765-0.989	93.8	68.6

	Median	Minimum	Maximum	Std.Dev.	Difference from Normal P-value
CD4:CD8B ratio					
Normal	5.93	1.62	63.77	10.56	
B-Lymphoid	3.53	0.44	27.75	4.71	<0.0001
Myeloid	4.65	0.64	179.77	28.07	0.038
Solid tumors	3.41	0.80	19.84	4.62	0.0002
T-lymphoid	3.18	1.43	9.95	2.84	0.003
Reactive	3.32	0.29	34.86	6.29	< 0.0001
CD3D:CD19					
Normal	6.88	1.45	20.13	4.91	
B-Lymphoid	3.18	0.01	390.76	42.99	0.0001
Myeloid	1.85	0.20	177.15	24.55	< 0.0001
Solid tumors	2.23	0.16	17.63	4.25	< 0.0001
T-lymphoid	6.28	0.34	16.63	4.38	0.36
Reactive	4.15	0.15	2529.89	316.25	0.01
CD274:CD8B					
Normal	1.77	0.19	30.69	4.92	
B-Lymphoid	1.32	0.14	57.77	7.01	0.41
Myeloid	1.86	0.10	33.28	5.01	0.51
Solid tumors	1.45	0.09	11.44	2.16	0.96
T-lymphoid	1.80	0.31	5.89	1.56	0.92
Reactive	1.37	0.00	23.40	3.04	0.19
CTLA4:CD8B					
Normal	0.19	0.00	8.14	1.27	
B-Lymphoid	0.54	0.01	937.78	96.06	< 0.0001
Myeloid	0.59	0.00	4.91	0.87	< 0.0001
Solid tumors	0.74	0.00	3.38	0.72	0.0001
T-lymphoid	0.58	0.01	2.66	0.66	0.024
Reactive	0.41	0.00	9.74	1.21	0.004

cfRNA immune biomarkers reflect the host response to the neoplastic process

Relative cfRNA expression levels of various immune cell markers and comparisons of various neoplastic processes with normal control.