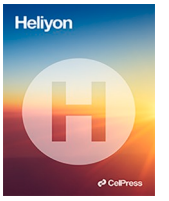
Heliyon 9 (2023) e16261

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Combining cell-free RNA with cell-free DNA in liquid biopsy for hematologic and solid tumors

Maher Albitar [a,*], Hong Zhang [a], Ahmad Charifa [a], Andrew Ip [b], Wanlong Ma [a], James McCloskey [b], Michele Donato [b], David Siegel [b], Stanley Waintraub [b], Martin Gutierrez [b], Andrew Pecora [b], Andre Goy [b]

[a] *Genomic Testing Cooperative, LCA, Irvine, CA, 92618, USA*
[b] *John Theurer Cancer Center at Hackensack University Medical Center, Hackensack, NJ, 07601, USA*

ARTICLE INFO

Keywords:
Liquid biopsy
Cell-free DNA
Cell-free RNA
Biomarkers
Mutations
Machine learning
Next generation sequencing
Expression profiling
Cancer diagnosis
Minimal residual disease

ABSTRACT

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. We developed an approach to isolate cell-free total nucleic acid (cfDNA) and used targeted next generation sequencing to sequence cell-free RNA (cfRNA) and cfDNA as new approach in liquid biopsy. We demonstrate that cfRNA is overall more sensitive than cfDNA in detecting mutations. We show that cfRNA is reliable in detecting fusion genes and cfDNA is reliable in detecting chromosomal gains and losses. cfRNA levels of various solid tumor biomarkers were significantly higher ($P < 0.0001$) in samples from solid tumors as compared with normal control. Similarly, cfRNA lymphoid markers and cfRNA myeloid markers were all higher in lymphoid and myeloid neoplasms, respectively as compared with control ($P < 0.0001$). Using machine learning we demonstrate cfRNA was highly predictive of diagnosis (AUC >0.98) of solid tumors, B-cell lymphoid neoplasms, T-cell lymphoid neoplasms, and myeloid neoplasms. In evaluating the host immune system, cfRNA CD4:CD8B and CD3D:CD19 ratios in normal controls were as expected (median: 5.92 and 6.87, respectively) and were significantly lower in solid tumors ($P < 0.0002$). This data suggests that liquid biopsy combining analysis of cfRNA with cfDNA is practical and may provide helpful information in predicting genomic abnormalities, diagnosis of neoplasms and evaluating both the tumor biology and the host response.

1. Introduction

Numerous studies have shown that peripheral blood contains tumor-specific cell-free DNA (cfDNA) that can be used as an alternative to tissue-based DNA testing. The concept of liquid biopsy was evolved around analyzing cfDNA for mutations. Liquid biopsy currently is used for cancer diagnosis, therapy recommendation and monitoring of therapy and relapse [1–4]. Detection of mutations and methylation patterns using next generation sequencing (NGS) are currently the main approaches used in liquid biopsy [1–4]. Blood is accessible and frequently used when tissue is not easily accessible or limited in quantity. The circulating DNA captures heterogeneity in the tumor, and it represents the entire tumor in the body minimizing sampling errors. However, because liquid biopsy

* Corresponding author. Genomic Testing Cooperative, LCA. 175 Technology Dr. #100, Irvine, CA, 92618, USA.
E-mail address: malbitar@genomictestingcooperative.com (M. Albitar).

https://doi.org/10.1016/j.heliyon.2023.e16261
Received 16 March 2023; Received in revised form 7 May 2023; Accepted 11 May 2023
Available online 16 May 2023

in its current format is depending on analyzing cfDNA it has multiple challenges. This includes variation between tumors in shedding their DNA, low sensitivity especially in early stage of cancer, difficulty in detecting fusion genes and inability to reflect the numerous biological process that modify RNA expression levels such as alternative splicing, stability, and allele-specific methylation. Recent studies showed that RNA testing provides another level of biological information related to the tumor and its microenvironment [5–8]. RNA expression profiling in tissue is helpful in diagnosis, selecting therapy, determining immunophenotype and microenvironment changes, fusion genes, alternative splicing and enhancer hijacking [4–10]. This deficiency in relying on cfDNA is currently significant and hampering progress in liquid biopsy testing.

RNA is typically considered less stable than DNA and more amenable to degradation especially in peripheral blood circulation. Therefore, the current concept assume it would be difficult to analyze cfRNA. However, the principle of NGS technology is using fragmented DNA or RNA in sequencing then align these fragments against the reference genome. RNA analysis using NGS from FFPE (formalin-fixed paraffin-embedded) tissue is routinely now used in clinical testing and had shown to be highly informative and reliable. NGS technology by its principle relies on fragmenting DNA or RNA before sequencing. Aligning the sequence of the fragmented DNA or RNA with proper bioinformatic system allows analysis and quantification of the sequenced nucleic acid. Therefore, sequencing fragmented cell-free RNA should be possible in a fashion similar to sequencing RNA in FFPE tissue [11–13].

We explored using NGS to analyze cfRNA for evaluating mutations, expression levels, detecting fusion genes and compared with cfDNA. However, since typically few highly expressed genes might overshadow the sequencing system, instead we used targeted RNA sequencing using a panel of 1459 genes selected specifically to cover clinically relevant for diagnosis and oncogenesis. Developing methods for analyzing cell-free RNA (cfRNA) is unmet need that may change how liquid biopsy is currently performed and used in clinical setting. We demonstrate that cfRNA when combined with cfDNA analysis enhance significantly the sensitivity of liquid biopsy and can provide information that highly useful clinically covering mutations, tumor-specific biomarkers, immunoprofiling, chromosomal structural abnormalities or copy number variation (CNV) and fusion data.

2. Methods

2.1. Patients' samples

Peripheral blood samples were collected randomly from patients with confirmed history of B-cell lymphoid neoplasms (#81), T-cell neoplasms (#14), myeloid neoplasms (#71), solid tumors (#45), as well as from normal individuals (#51) and patients post-transplant as reactive samples (#137). The tested samples from patients with hematologic and solid tumors included various diseases and at different stage. These patients were chosen randomly without any selection. Some of these patients had active disease and some without clinically active disease. The reactive samples were collected from patients after hematopoietic stem cell transplant with no clinical evidence of residual neoplasm. These post-transplant samples were used to explore the significance of the immune reactive processes in models for detecting diseases using cfRNA. Total nucleic acid was extracted from 1 ml plasma. Blood samples were collected in EDTA (Ethylenediaminetetraacetic acid). Total nucleic acid was extracted within 72 h of collection. Samples that their RNA did not meet quality control criteria (see below) were excluded. The failure of meeting the quality RNA criteria was 9%. Study samples and data was collected under approved IRB (WCG IRB # 1-1476184-1).

2.2. Nucleic acid extraction

To capture small, fragmented RNA and DNA, we used an extraction kit optimized for capturing microRNA in circulation. We used Apostle MiniMax High Efficiency cfRNA/cfDNA isolation kit and followed the recommended protocol. After extraction, half of the cfDNA was treated with DNase to obtain cfRNA and the other half was used for DNA studies.

2.3. Targeted cfRNA and cfDNA next generation sequencing

Each patient's cfDNA and cfRNA were processed in parallel using KAPA RNA HyperPrep kit. The cfRNA steps include: 1. First strand synthesis; 2. second strand synthesis and A-tailing; 3. ligating adapters; 4. library amplification; 5. cleaning up; 6. quantification; 7. preparing each 8-plex DNA sample library pool; 8. Adding designed KAPA Target Enrichment Probes (1459 genes, Table S1, supplement) for hybridization overnight; 9. Using KAPA bead to capture multiplexed DNA library; 10. Amplifying enriched multiplex DNA library; 11. Cleaning up amplified enriched multiplexed DNA library; 12. Checking multiplexed library by running TapeStation; 13. Normalizing and Pooling; 14. Denaturing and Load Libraries on NovaSeq 6000 sequencer for Sequencing using pair-end 100X2 cycles. For cfDNA, similar steps were followed except for the initial reverse transcription step.

After the sequence bcl2fastq2 Software v.2.20.0 was used for de-multiplexing and conversion to FASTQ. Dragen 3.8 RNA seq pipeline for fusion calls, Salmon v1.4.0 for expression (TPM), CNVkit for CNV calls and RNA-Seq Alignment v.2.0.2 – BaseSpace Sequence Hub App for VCf for mutation calls. For RNA sequencing, thirty million reads per sample were required. As a quality control, cases with DNA contamination in the RNA sequencing were excluded. Minimal of 30 million reads were required for RNA sequencing. The RNA sequencing considered contaminated by DNA and the cases were excluded from the study if the percent of splice junction reads was $<20\%$.

2.4. Using machine learning algorithm for analyzing RNA data

The RNA expression data is used in machine learning algorithm to distinguish between various diagnostic groups. We first selected genes that distinguish between two classes using standard Naïve Bayesian classifier on each gene with k-fold cross validation.

$$a = \frac{1}{m}\sum_{i=1}^{m}\frac{t_i}{n_i}$$

where m is the number of classes, n_i is the number of cases in the class, and t_i the number of correctly classified cases in class i estimated with a k-fold cross validation.

When the number of classes $m = 2$, the measure is the average of sensitivity and specificity. In general, it is the average of the accuracies on individual classes. The overall accuracy is not appropriate for gene selection because it can be misleading in the datasets with unbalanced classes. For example, in a dataset with 80 negative and 20 positive cases, a trivial classification of all negative would yield an 80% accuracy.

The coefficient $\frac{1}{m}$ is usually ignored in our implementation since m is a constant and will not affect the ranking. The k-fold cross validation is usually implemented with $k = n$ (i.e. leave-one-out). Even though the leave-one-out method is more computationally intensive, the efficiency of the Naïve Bayes algorithm would still make the selection process reasonably fast.

This accuracy value provides a direct measure of the gene in classifying the groups. However, it does not give confidence information. For confidence measure for gene selection we relied the p-value of a gene in differentiating the classes. Analysis of Variance (ANOVA) is applied to compute the p-value for a gene to discriminate the groups.

$$F = \frac{MSB}{MSW}$$

where *MSB* is the mean sum of squares between the groups, *MSW* the mean sum of squares within the groups, and F is the ANOVA coefficient following the F-distribution. The p-value can be obtained from the F value. This confidence value provides a measure of stability and robustness of the gene in classifying the groups. It does not give the concrete classification accuracy, but it provides an overall confidence in the differences of the class means. Both criteria provide quantitative measures on the relevancy of the gene for classification. However, the two relevance measures do not always provide the same rankings. Applying both measures would produce effective and stable gene selection methods for machine learning based classification systems.

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. In particular,

$$P(x_1, x_2, \ldots, x_d | C_j) = \sqrt[d]{[P(x_1|C_j)P(x_2|C_j)\ldots P(x_d|C_j)]}$$

This formula represents the geometric mean of the conditional probabilities. We proved that the GMNB method resolves the underflow problem for high dimensional data, by showing the expected value of such a likelihood will approach $1/e$ when the dimension $d \to \infty$. We also proved that such a function is essentially unique, up to a constant multiple of the exponent [14].

To reduce the effects of noises and avoid overfitting in selecting these genes, we employed a leave-one-out cross validation to obtain a robust performance measure. For an individual gene, a GMNB is constructed on the training subset and tested on the testing subset. The complement of the cross-validation error rate is used as the discriminant measure for the bins.

$$d = \sum_{c=1}^{k} 1 - \frac{error_c}{n_c}$$

Instead of the overall error rate, the value d takes a sum of the error rates of the individual classes. This definition would avoid the bias when the sample sizes are not balanced for different classes. The genes were ranked by d with higher values corresponding to better performing genes for classifying the two classes. To address stability issues, we used the t-test to measure the significance of a bin in separating the 2 classes. By setting a p-value threshold, insignificant bins can be filtered out. The selected genes were used to distinguishing between two classes with k-fold cross-validation procedure (with $k = 12$). A naïve Bayesian classifier was constructed on the training of k-1 subsets and tested on the other testing subset. The training and testing subsets then rotated, and the average of the classification errors was used to measure the relevancy of the gene. The classification system was trained with the selected subset of most relevant genes. The processes of Gene selection and class selection status were applied iteratively to obtain an optimal classification system and a subset of genes relevant to distinguishing between the two classes were defined and isolated.

3. Results

3.1. Higher number of mutations detected in cfRNA

Number of mutations detected in solid tumors and hematologic neoplasms were significantly ($P > 0.0001$) higher in cfRNA (No. =

1229) than in cfDNA (No. = 1004) (P < 0.0001) (Fig. 1A). However, numerous mutations detected by RNA were not detected by cfDNA and vice versa. In general nonsense mutations were more likely to be detected by cfDNA than by cfRNA and at higher VAF (Fig. S1). Almost all missense mutations were detected in both cfDNA and cfRNA. However, low level mutations (VAF< 10%) were more likely to be detected by cfRNA than by cfDNA. We detected 136 mutations in TP53 gene using cfRNA and only 70 mutations in cfDNA (Fig. 2A) (P < 0.0001). KRAS mutations were also higher in cfRNA (#33) as compared with cfDNA (#21) (P = 0.08) (Fig. 2B). In contrast, when most of the mutations were nonsense, as in ASXL1 gene, slightly more mutations were detected by cfDNA (#24) as compared with cfRNA (#23) (Fig. 2 C). When we evaluated mutation in normal individuals we detected more CHIP (clonal hematopoiesis of indeterminate potential) mutations in cfDNA as compared with cfRNA (P < 0.0001) (Fig. 1B). Most of the mutations in CHIP are nonsense in epigenetic genes (ASXL1, TET2, and DNMT3A). When mutations were detected in both cfRNA and cfDNA, mutation load (level of mutant copies) was overall slightly higher in cfDNA (P = 0.06), likely due to higher degradation of RNA, but varied significantly dependent on the type of mutated gene and type of mutation. Overall variant allele frequency (VAF) was significantly higher in cfRNA than in cfDNA (P < 0.0001) due to the higher number of mutations detected in cfRNA as compared to cfDNA. This was true in general, when we evaluated myeloid, lymphoid and solid tumors are separately (Fig. S1). To eliminate cases with DNA contamination.

A. Bar graphs showing significantly more mutations detected cfRNA than in cfDNA. B. Bar graphs showing significantly more CHIP (clonal hematopoiesis of indeterminate potential) detected in cfDNA than in cfRNA.

A. Significantly higher number of mutations in TP53 are detected in cfRNA than in cfDNA. B. Only 21 mutations in KRAS gene are detected in cfDNA, while 33 mutations are detected in the cfRNA. C. Slightly more mutations in ASXL1 gene were detected in cfDNA as compared with cfRNA (24 vs 23). The levels of these mutations are shown on the Y-axis.

3.2. Detection of fusions using cfRNA

RNA enabled us to detect fusion genes using MANTA and DRAGEN (Illumina). Fusions involving ROS1 in lung cancer, BCL6 in lymphoma, ETV6 (Fig. 3A), RUNX1 (Fig. 3B), RARA, FPM2, DEK, and EP300 were detected (Table S2). All these fusions were detected in tissue/cell samples. The sensitivity of detecting fusions is difficult to evaluate due to variation in the levels of the fused genes. In general, most of the detected fusions were detected in samples with relatively high tumor load in circulation (Table S2). In a lung sample, we detected SLC34A2-ROS1 fusion in despite no somatic cancer-specific mutations were detected which suggests high sensitivity in detecting fusions (Table S2). All these fusions were also detected in tissue testing by cytogenetic studies or by FISH (fluorescent in-situ hybridization) testing.

A. RUNX1-ETV6 fusion transcript is detected in the cfRNA from a patient with acute lymphoblastic leukemia. B. RUNX1-RUNX1T1 Fusion transcript detected in the cfRNA from a patient with acute myeloid leukemia. The fusion transcripts are visualized using FusionInspector.

3.3. Detection of chromosomal structural gains and losses

The cfDNA testing allowed us to detect chromosomal structural abnormalities in hematologic and solid tumors (Fig. 4A and B) that were also detected in corresponding tissue samples (Fig. 4C and D). The clinical utility of detecting chromosomal structural abnormalities using liquid biopsy and targeted gene DNA panel has been established [15]. Using large number of genes makes this detection more reliable.

A. Diagram resulting from CNVkit analysis cfDNA from a patient with colorectal cancer showing multiple chromosomal abnormalities. Gain is illustrated in red, and loss is illustrated in blue. B. Scatter plot of the same data showing the same abnormalities. C and D showing the same findings as detected in the corresponding tissue sample. The Y-axis shows the log2 ratio of the patient's sample to normal control. The X-axis shows the various chromosomes.

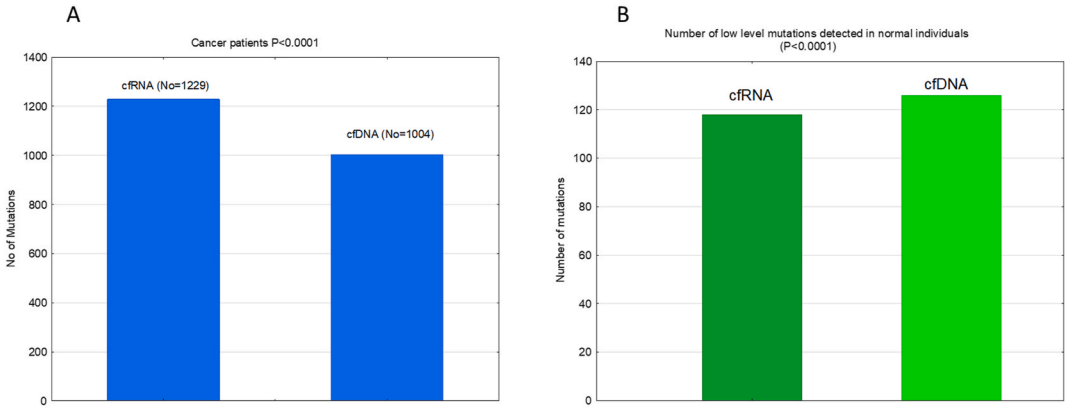


Fig. 1. Higher number of mutations detected in cfRNA as compared with cfDNA in patients with cancer.

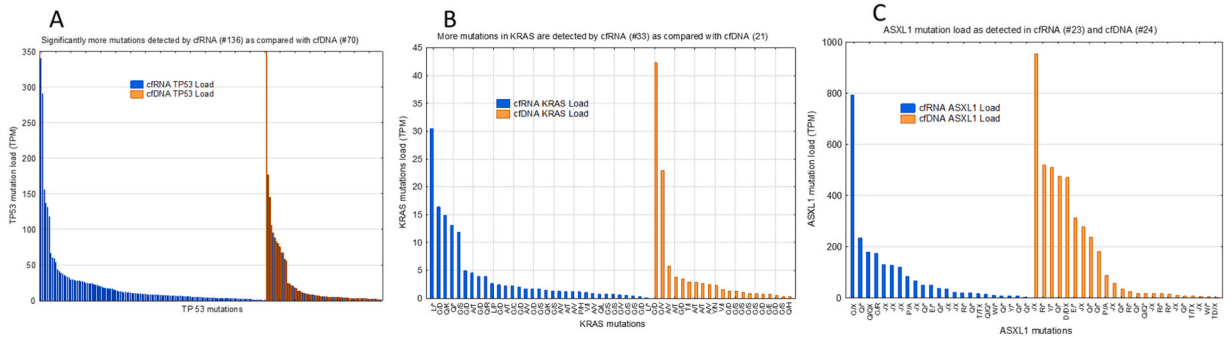


Fig. 2. Levels and numbers of detected mutations vary between cfRNA and cfDNA dependent on type of mutated genes and type of mutations.

3.4. cfRNA contains numerous biomarkers reflecting tumor characteristics

Expression levels of various genes as measured in cfRNA showed significant difference from normal sample. Expression of known solid tumors biomarkers (CA-125, CA-15-3, CEA 8, Keratin19, Keratin6A …) [16,17] were significantly higher ($P < 0.0001$) in samples from solid tumors as compared with normal control (Fig. S2). Evaluating GATA3, CA15-3, CA125 and ERBB2 (HER2) in samples with breast cancer (#8) shows significantly higher levels of these biomarkers as compared with normal control (Fig. S3). In lymphoid neoplasms, lymphoid markers (CD19, CD22, CD79A, and CD79B) [18] were also significantly higher ($P < 0.0001$) than in normal control sample (Fig. S4). Myeloid markers (CD33, CD14, CD117, CD56) [19] were also higher ($P < 0.0001$) than in normal control (Fig. S5). The CD34 mRNA was not different in myeloid neoplasms as compared with normal control ($P = 0.16$). This is most likely due to the relatively overall low level blast cells in the myeloid group, which included myelodysplastic and myeloproliferative cases. Also, possibly that due to CD34 is expressed in endothelial cells the level of CD34 is relatively high in normal samples (Fig. S5).

3.5. Utilizing cfRNA in the differential diagnosis of neoplasms

To explore the value of the cfRNA expression in determining the differential diagnosis of specific neoplasm, we used an algorithm in which we first select the specific genes that their expression can distinguish between two diagnostic classes then we used machine learning in combination of genes in distinguishing between various diagnostic classes. As shown in Table 1 and Fig. 5, distinguishing between normal group and various diagnostic classes (B-lymphoid, myeloid, solid tumors, and T-lymphoid) can be achieved with high sensitivity and specificity. Distinguishing between normal control and individuals with reactive process can also be achieved with high accuracy (AUC = 0.998, 95% CI: 0.988–1.0). As shown in Supplement Table S3, significant difference in the levels of various genes is easily demonstratable between various diagnostic groups (Table S3). Despite that large number of genes show significant difference in levels between two diagnostic classes (Table S3), the number of genes needed for good separation is small varied between 10 and 100 (Table 1). As expected, most the relative genes that distinguish each diagnostic class from normal specific for that diagnosis. However, when each of the neoplastic diagnostic classes is compared to normal and reactive groups combined, the prediction became significantly less accurate, but remained very useful (Table 1 and Fig. 5). The genes that distinguish between these diagnostic groups are listed in Supplement Table S3. The relative significant decrease in the accuracy of predicting the diagnostic class upon adding reactive groups to the normal individuals reflects the importance of the host reactive response in predicting diagnosis.

3.6. cfRNA immune biomarkers reflect host response to the neoplastic process

To explore the potential of cfRNA in evaluating the host immune system, we evaluated cfRNA levels of known markers of lymphoid cells such as CD4:CD8 and CD3:CD19 rations. As shown in Table 2, normal control shows ratio of CD4 to CD8 and CD3 to CD19 cfRNA comparable to those expected ratios from cell in peripheral blood. However, significant variation is seen when various neoplasms and in reactive processes are considered (Table 2). There was significant difference in CTLA4:CD8B cfRNA ratios in various diagnostic groups as compared with control. In contrast, there is no difference in CD274(PD-L1):CD8B cfRNA ratios between control and other diagnostic groups. Similar distinct patterns are noted for various cytokine and chemokines (data not shown).

4. Discussion

The current concept of liquid biopsy is centered around isolating cfDNA from peripheral blood and analyzing DNA for mutations, structural abnormalities, and methylation. Little work has been reported on using cfRNA. RNA is, in general, more labile, therefore, few attempts have been made to use RNA in liquid biopsy. In a previous publication, we have demonstrated that patients with chronic myeloid leukemia can be monitored using reverse transcription/polymerase chain reaction (RT/PCR) of cfRNA [20]. cfRNA evaluation using RT/PCR is also reported in multiple studies as a source for providing biomarkers for evaluating specific cancers [21]. cfRNA is also reported in evaluating pregnancy and complication of pregnancy [22]. Recent work by Moufarrej et al. showed the ability of

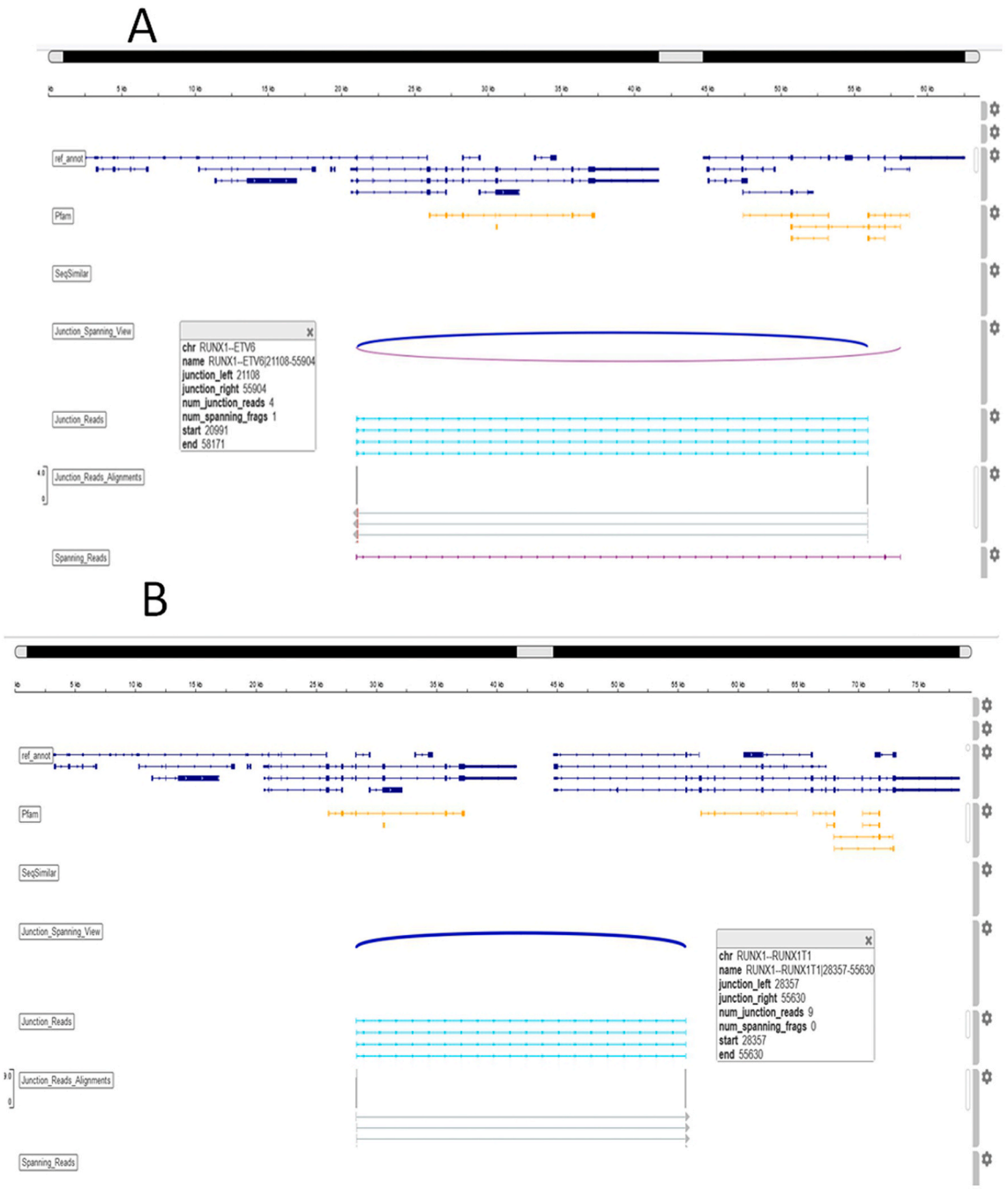


Fig. 3. Examples of detected fusion transcripts using cfRNA.

cfRNA expression pattern as determined by NGS to predict preeclampsioa [23].

Most of the attempts to utilize cfRNA in NGS have been based on whole transcriptome. As expected cfRNA is highly degraded, which compromises capturing intact mRNA for whole transcriptome sequencing. Using targeted sequencing and hybrid capturing of the mRNA of genes of interest does not only overcome the problem of degradation, but also allows capturing of low expressed genes increasing the dynamic range of detection and quantification and improves quantification significantly.

In this study, we extracted cfDNA from peripheral blood plasma of samples. A portion of the cfDNA was treated with DNAse to isolate RNA and the RNA was sequenced using a targeted RNA panel of 1459 genes. The remaining cfDNA was also sequenced using the same hybrid capture panel of 1459 genes. Upon comparing the number of mutations detected in the cfRNA with those detected by cfDNA, significantly more mutations were detected using cfRNA than cfDNA (Fig. 1). The higher number of mutations in the cfRNA

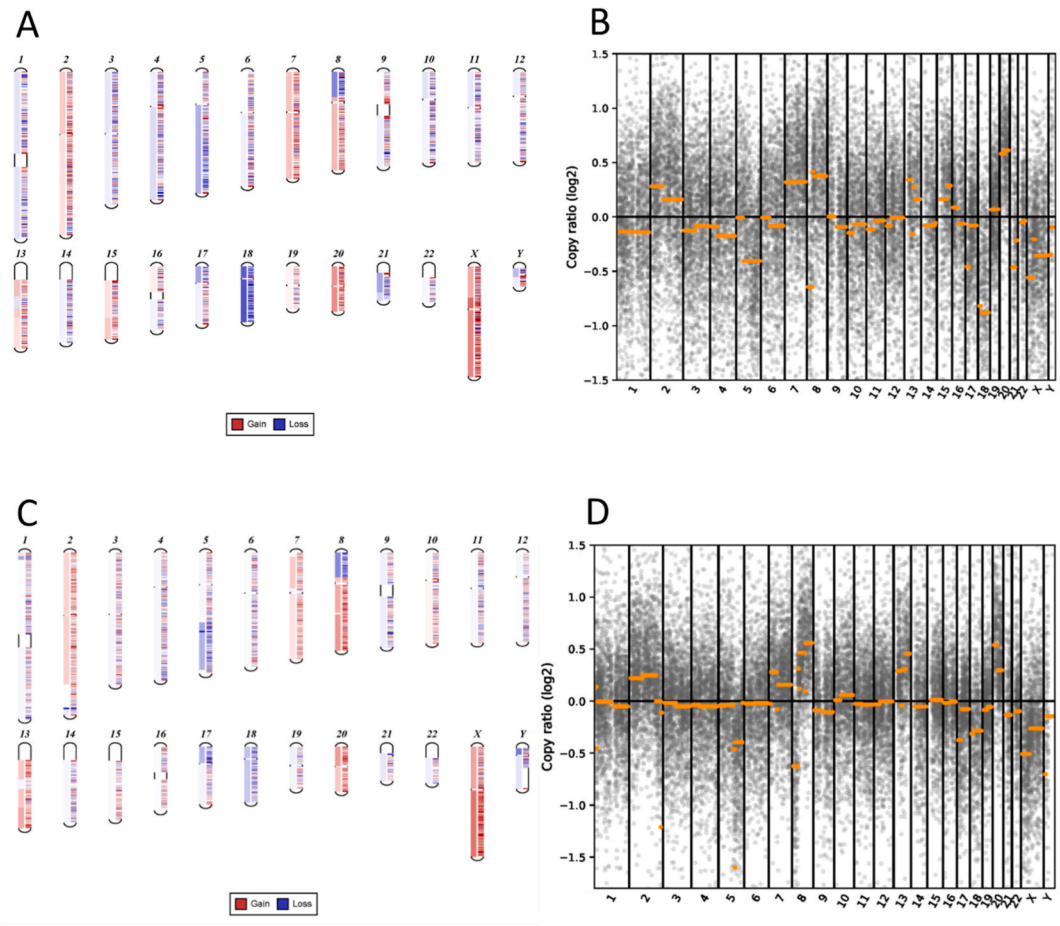


Fig. 4. Example of chromosomal structural loss and gain using cfDNA as compared with tissue-based detection.

Tables 1
Prediction of diagnosis using cfRNA in 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

most likely due to the well-documented allele-specific methylation and lower expression of the wild-type RNA enriching the relative level of mutant RNA as compared with mutant DNA [24–28]. However, this phenomenon is not necessarily true for all types of mutations. RNA stability has been reported to be significantly less when mutations lead to early termination of protein synthesis [29, 30]. This is particularly relevant for loss of function mutations that are common in tumor suppressor genes and epigenetic genes. When we compared number of mutations detected in ASXL1 gene, most of which are nonsense mutations, slightly more mutations were detected in cfDNA than in cfRNA (24 vs 23). In contrast more mutations in KRAS gene were detected in cfRNA as compared with cfDNA (33 vs 21, $P = 0.08$). TP53 mutations were significantly more (#136) in cfRNA than in cfDNA (#70) ($P < 0.0001$). This difference was

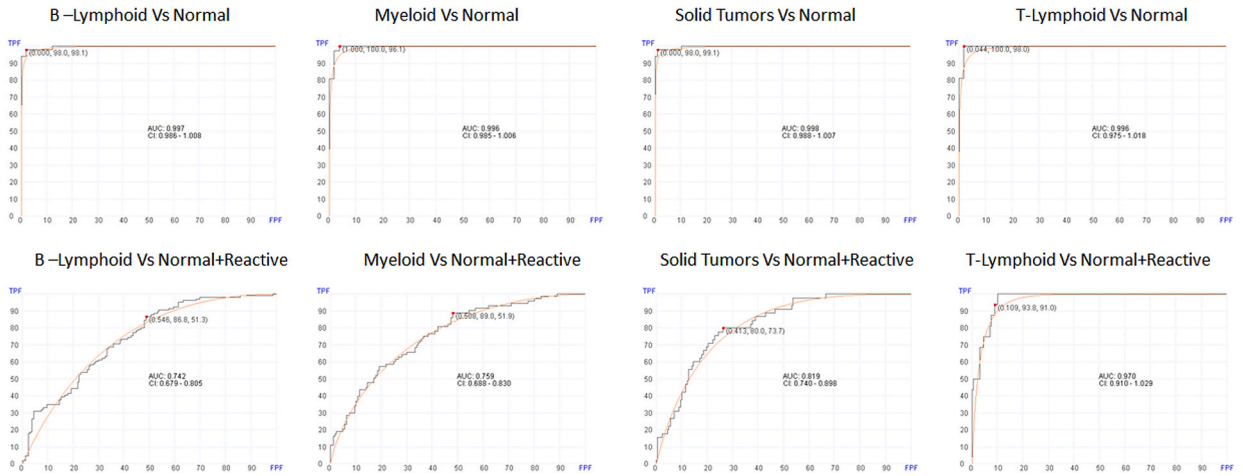


Fig. 5. Receiver operating characteristic (ROC) curve for prediction of diagnosis using cfRNA in artificial intelligence (AI)-based models. The area under the curve (*AUC*) and the 95% confidence interval are shown for various diagnostic classes. Only cfRNA expression levels are used in this modeling. TPF, true positive fraction (sensitivity); FPF, false positive fraction (specificity).

Table 2
Relative cfRNA expression levels of various immune cell markers and comparison of various neoplastic processes with normal control.

	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

more conspicuous in normal individuals with CHIP mutations. CHIP mutations were detected at higher rate in cfDNA as compared with cfDNA ($P < 0.0001$).

Using RNA sequencing has been established to be more sensitive in detected fusions [31–34]. Therefore, using cfRNA for the purpose of detecting fusions is more sensitive than DNA. cfRNA allowed us to detect fusions in samples from hematologic and solid tumors with relatively high sensitivity as some samples showed fusion with significantly low level mutations or no mutations.

The large panel of 1459 genes scattered throughout all chromosomes allows us to detect chromosomal structural abnormalities reliably in cfDNA. Th1459 gene panel covers significantly more genomic areas than our previous panel of 275 genes, which showed reliability in detecting chromosomal losses and gains [15].

We also demonstrate that the cfRNA is highly valuable in evaluating expression levels of important genes that can reflect the biology of the cancer and the host response when patients with various types of cancers are tested. We demonstrate that patients with

B-cell lymphoid neoplasms show significantly higher levels of biomarkers associated with B-cell neoplasms such as CD19, CD22, CD79B, CD79A and others. Similarly, cfRNA in patients with myeloid neoplasms show the expression of biomarkers specific for myeloid neoplasms. This data suggests that cfRNA can provide information comparable to those typically obtained from flow cytometry analysis of hematologic neoplasms. In solid tumors, cfRNA can reflect the typical biomarkers detected in solid tumors (CA-125, CA-15-3, CEA and others). In fact evaluating expression of biomarkers typically evaluated by immunohistochemistry (IHC), such as various keratin types, can also be performed using cfRNA and specific biomarkers evaluated by chemistry tests (CEA, PSA, CA-125 …) can also be performed by cfRNA. Using profiling with machine learning algorithms enhances the ability of the cfRNA profiles in detecting specific diseases or processes in patients. Using the expression profiles of 30–60 genes in machine learning system can distinguish patients with various neoplasms from normal control with high accuracy. The number of cases is small in this study but distinguishing between normal and a specific disease seems to be reliable. Clearly it would be more reliable if DNA and RNA mutation profile is integrated into the system as well. Distinguishing between reactive process and normal control was also reliable using only 10 biomarkers. However, distinguishing neoplastic diseases from combined reactive and normal cases was less reliable. This suggests that the immune system markers play a major rule in distinguishing between normal and diseased status.

Immune system genes are between the list of genes (Table S3) used by the machine learning algorithms to distinguish specific neoplasms from normal control. The cfRNA in normal individuals shows CD4:CD8 and T-cell:B-cell cfRNA markers ratios comparable to those expected when actual cells are evaluated in peripheral blood by flow cytometry. This suggests that models can be developed to establish the norm for cfRNA immune markers so abnormal relative immune biomarkers can be defined and used to evaluate the immune system and its relevance and response to diseases and immune therapy. This data also suggests that there is a possibility that the cfRNA can be used in reactive processes, autoimmune diseases and wellness programs.

The data presented in this paper is proof of principle and indicates that cell-free RNA can be reliably evaluated in health and disease. cfRNA provides valuable information for predicting mutations that sometimes missed by cfDNA testing. cfRNA is reliable in detecting fusion mRNA. This data suggests that for improving the capabilities of liquid biopsy, cfRNA should be evaluated along with cfDNA or cfDNA to obtain comprehensive data on mutation profiling, chromosomal structural abnormalities and fusion genes. This approach has a potential to increase sensitivity and specificity significantly for detecting minimal residual disease (MRD). Liquid biopsy evaluating cfRNA and cfDNA also provide highly informative data that can be used for establishing specific diagnosis and evaluating the immune response of the host as well as predicting discovering biomarkers for predicting response to various therapies including immunotherapy. Furthermore, since the first indication of a neoplastic process is immune response [35,36] to the neoplastic process cfRNA + cfDNA liquid biopsy evaluating host response and cancer biomarkers as well as mutations profile has the potential of being used for early diagnosis and for screening. However, larger study is needed for confirming these findings and for developing proper models for clinical applications of such technology and approach.

Informed consent

Study samples and data were collected under approved IRB (WCG IRB # 1-1476184-1).

Author contribution statement

Maher Albitar: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper. Hong Zhang; Andrew Ip: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper. Ahmad Charifa: Analyzed and interpreted the data; Wrote the paper. Wanlong Ma: Performed the experiments; Wrote the paper. James McCloskey; Michele Donato; Andrew Pecora; David Siegel: Contributed reagents, materials, analysis tools or data; Wrote the paper. Martin Gutierrez; Stanley Waintraub: Contributed reagents, materials, analysis tools or data. Andre Goy: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Data availability statement

Data will be made available on request.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

MA, AC, and WM work for a diagnostic company offering liquid biopsy testing; MA, HZ, WM, AP, and AG own stocks in a company offering liquid biopsy testing. JM Served on speakers bureau for Amgen, Bristol Myers Squibb, Incyte, Jazz Pharmaceuticals, Stemline, and Takeda, and has served as a consultant for AbbVie, CTI BioPharma, and Novartis. The rest of authors have no relevant competing interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e16261.

References

[1] R.B. Corcoran, B.A. Chabner, Application of cell-free DNA analysis to cancer treatment, N. Engl. J. Med. 379 (18) (2018) 1754–1765, https://doi.org/10.1056/NEJMra1706174.

[2] M. Ignatiadis, G.W. Sledge, S.S. Jeffrey, Liquid biopsy enters the clinic - implementation issues and future challenges, Nat. Rev. Clin. Oncol. 18 (5) (2021) 297–312, https://doi.org/10.1038/s41571-020-00457-x.

[3] C. Alix-Panabières, K. Pantel, Liquid biopsy: from discovery to clinical application, Cancer Discov. 11 (4) (2021) 858–873, https://doi.org/10.1158/2159-8290.CD-20-1311. PMID: 33811121.

[4] C. Rolfo, P. Mack, G.V. Scagliotti, C. Aggarwal, M.E. Arcila, F. Barlesi, et al., Liquid biopsy for advanced NSCLC: a consensus statement from the international association for the study of lung cancer, J. Thorac. Oncol. 16 (10) (2021) 1647–1662, https://doi.org/10.1016/j.jtho.2021.06.017.

[5] R. Kamps, R.D. Brandão, B.J. Bosch, A.D. Paulussen, S. Xanthoulea, M.J. Blok, et al., Next-generation sequencing in oncology: genetic diagnosis, risk prediction and cancer classification, Int. J. Mol. Sci. 18 (2) (2017) 308, https://doi.org/10.3390/ijms18020308. PMID: 28146134.

[6] A.H. Saeidian, L. Youssefian, H. Vahidnezhad, J. Uitto, Research techniques made simple: whole-transcriptome sequencing by RNA-seq for diagnosis of monogenic disorders, J. Invest. Dermatol. 140 (6) (2020) 1117–1126.e1, https://doi.org/10.1016/j.jid.2020.02.032. PMID: 32446329.

[7] PCAWG Transcriptome Core Group, C. Calabrese, N.R. Davidson, D. Demircioğlu, N.A. Fonseca, Y. He, et al., Genomic basis for RNA alterations in cancer, Nature 578 (7793) (2020) 129–136, https://doi.org/10.1038/s41586-020-1970-0.

[8] A.M. Tsimberidou, E. Fountzilas, M. Nikanjam, R. Kurzrock, Review of precision cancer medicine: evolution of the treatment paradigm, Cancer Treat Rev. 86 (2020), 102019, https://doi.org/10.1016/j.ctrv.2020.102019.

[9] F. Pastor, P. Berraondo, I. Etxeberria, J. Frederick, U. Sahin, E. Gilboa, I. Melero, An RNA toolbox for cancer immunotherapy, Nat. Rev. Drug Discov. 17 (10) (2018) 751–767, https://doi.org/10.1038/nrd.2018.132.

[10] F.F. Hu, C.J. Liu, L.L. Liu, Q. Zhang, A.Y. Guo, Expression profile of immune checkpoint genes and their roles in predicting immunotherapy response, Briefings Bioinf. 22 (3) (2021) bbaa176, https://doi.org/10.1093/bib/bbaa176. PMID: 32814346.

[11] M. Sorokin, A. Gorelyshev, V. Efimov, E. Zotova, M. Zolotovskaia, E. Rabushko, D. Kuzmin, et al., RNA sequencing data for FFPE tumor blocks can Be used for robust estimation of tumor mutation burden in individual biosamples, Front. Oncol. 11 (2021), 732644, https://doi.org/10.3389/fonc.2021.732644. PMID: 34650919.

[12] A. Talebi, J.P. Thiery, M.A. Kerachian, Fusion transcript discovery using RNA sequencing in formalin-fixed paraffin-embedded specimen, Crit. Rev. Oncol. Hematol. 160 (2021), 103303, https://doi.org/10.1016/j.critrevonc.2021.103303.

[13] Y. Zhao, M. Mehta, A. Walton, K. Talsania, Y. Levin, J. Shetty, E.M. Gillanders, B. Tran, et al., Robustness of RNA sequencing on older formalin-fixed paraffin-embedded tissue from high-grade ovarian serous adenocarcinomas, PLoS One 14 (5) (2019), e0216050, https://doi.org/10.1371/journal.pone.0216050. PMID: 31059554.

[14] M. Albitar, H. Zhang, A. Goy, et al., Determining clinical course of diffuse large B-cell lymphoma using targeted transcriptome and machine learning algorithms, Blood Cancer J. 12 (2022) 25, https://doi.org/10.1038/s41408-022-00617-5.

[15] A. Ip, M. Albitar, J. Lofters, J. Behrmann, D. Patel, A.H. Goy, J. Estella, I. De Dios, W. Ma, et al., Reliability of cell-free DNA (cfDNA) and targeted next generation sequencing in predicting chromosomal structural abnormalities, Front. Oncol. 12 (2022), 923809, https://doi.org/10.3389/fonc.2022.923809. PMID: 35774119; PMCID: PMC9238409.

[16] M.K. Accordino, J.D. Wright, S. Vasan, A.I. Neugut, A. Tergas, J.C. Hu, D.L. Hershman, Serum tumor marker use in patients with advanced solid tumors, J Oncol Pract 12 (1) (2016) 65, https://doi.org/10.1200/JOP.2015.005660, 6, e36-43.

[17] J.A. Sandoval, L.H. Malkas, R.J. Hickey, Clinical significance of serum biomarkers in pediatric solid mediastinal and abdominal tumors, Int. J. Mol. Sci. 13 (1) (2012) 1126–1153, https://doi.org/10.3390/ijms13011126.

[18] C. Debord, S. Wuillème, M. Eveillard, O. Theisen, C. Godon, Y. Le Bris, M.C. Béné, Flow cytometry in the diagnosis of mature B-cell lymphoproliferative disorders, Int J Lab Hematol 42 (Suppl 1) (2020) 113–120, https://doi.org/10.1111/ijlh.13170. PMID: 32543070.

[19] G.J. Ossenkoppele, A.A. van de Loosdrecht, G.J. Schuurhuis, Review of the relevance of aberrant antigen expression by flow cytometry in myeloid neoplasms, Br. J. Haematol. 153 (4) (2011) 421–436, https://doi.org/10.1111/j.1365-2141.2011.08595.x.

[20] W. Ma, R. Tseng, M. Gorre, I. Jilani, M. Keating, H. Kantarjian, J. Cortes, et al., Plasma RNA as an alternative to cells for monitoring molecular response in patients with chronic myeloid leukemia, Haematologica 92 (2) (2007) 170–175, https://doi.org/10.3324/haematol.10360.

[21] H. Markus, Drag and Tuomas O. Kilpeläinen. Cell-free DNA and RNA—measurement and applications in clinical diagnostics with focus on metabolic disorders, Physiol. Genom. 53 (1) (2021) 33–46, https://doi.org/10.1152/physiolgenomics.00086.2020.

[22] M.N. Moufarrej, R.J. Wong, G.M. Shaw, D.K. Stevenson, S.R. Quake, Investigating pregnancy and its complications using circulating cell-free RNA in women's blood during gestation, Front. Pediatr. 8 (2020), 605219, https://doi.org/10.3389/fped.2020.605219.

[23] M.N. Moufarrej, S.K. Vorperian, R.J. Wong, A.A. Campos, C.C. Quaintance, R.V. Sit, M. Tan, et al., Early prediction of preeclampsia in pregnancy with cell-free RNA, Nature 602 (7898) (2022) 689–694, https://doi.org/10.1038/s41586-022-04410-z.

[24] M. Albitar, Z. Yidan Xu-Monette, M. Ma, Y. Wang, D. Manman, A. Tzankov, C. Visco, et al., Higher stability of mutant mRNA as compared to wild-type mRNA in diffuse large B-cell lymphoma, Blood 134 (Supplement_1) (2019) 1499, https://doi.org/10.1182/blood-2019-128516.

[25] A. Kazanets, T. Shorstova, K. Hilmi, M. Marques, M. Witcher, Epigenetic silencing of tumor suppressor genes: paradigms, puzzles, and potential, Biochim. Biophys. Acta 1865 (2016) 275–288, https://doi.org/10.1016/j.bbcan.2016.04.001.

[26] E.A. Clayton, S. Khalid, D. Ban, L. Wang, K. Jordan, J.E. McDonald, Tumor suppressor genes and allele-specific expression: mechanisms and significance, Oncotarget 11 (4) (2020) 462–479, https://doi.org/10.18632/oncotarget.27468. PMCID: PMC6996918. PMID: 32064050.

[27] J.L. Lewis Stern, R.D. Paucek, F.W. Huang, M. Ghandi, R. Nwumeh, J.C. Costello, T.R. Cech, et al., Allele-specific DNA methylation and its interplay with repressive histone marks at promoter-mutant, TERT Genes 21 (13) (2017) 3700–3707.

[28] S.K. Myöhänen, S.B. Baylin, J.G. Herman, Hypermethylation can selectively silence individual p16ink4A alleles in neoplasia, Cancer Res. 58 (4) (1998) 591–593, https://doi.org/10.1016/j.celrep.2017.12.001. PMID: 9485004.

[29] S. Das, D. Sarkar, B. Das, The interplay between transcription and mRNA degradation in Saccharomyces cerevisiae, Microbial Cell 4 (7) (2017) 212–228, https://doi.org/10.15698/mic2017.07.580.

[30] Toshimichi Yamada, Masami Nagahama, Nobuyoshi Akimitsu, Interplay between Transcription and RNA Degradation, InTech, 2018, https://doi.org/10.5772/intechopen.71862.

[31] R. Benayed, M. Offin, K. Mullaney, P. Sukhadia, K. Rios, P. Desmeules, et al., High yield of RNA sequencing for targetable kinase fusions in lung adenocarcinomas with no mitogenic driver alteration detected by DNA sequencing and low tumor mutation burden, Clin. Cancer Res. 25 (15) (2019) 4712–4722.

[32] B.C. Haynes, R.A. Blidner, R.D. Cardwell, R. Zeigler, S. Gokul, J.R. Thibert, et al., An integrated next-generation sequencing system for analyzing DNA mutations, gene fusions, and RNA expression in lung cancer, Transl Oncol 12 (6) (2019) 836–845, https://doi.org/10.1016/j.tranon.2019.02.012.

[33] E.E. Heyer, I.W. Deveson, D. Wooi, C.I. Selinger, R.J. Lyons, V.M. Hayes, et al., Diagnosis of fusion genes using targeted RNA sequencing, Nat. Commun. 10 (1) (2019) 1388, https://doi.org/10.1038/s41467-019-09374-9.

[34] C. Heydt, C.B. Wölwer, O.V. Camacho, S. Wagener-Ryczek, R. Pappesch, Detection of gene fusions using targeted next-generation sequencing: a comparative evaluation, BMC Med. Genom. 14 (1) (2021) 62, https://doi.org/10.1186/s12920-021-00909-y. PMID: 33639937; PMCID: PMC7912891.

[35] Hugo Gonzalez Catharina Hagerling1, Zena Werb, Roles of the immune system in cancer: from tumor initiation to metastatic progression, Genes Dev. 32 (19–20) (2018) 1267–1284, https://doi.org/10.1101/gad.314617.118.

[36] D. Hanahan, R.A. Weinberg, Hallmarks of cancer: the next generation, Cell 144 (5) (2011) 646–674, https://doi.org/10.1016/j.cell.2011.02.013, 2011.