Estimation of ALU repetitive element in plasma as a cost-effective liquid biopsy tool for disease prognosis in breast cancer

Madhumathy G Nair[1*#], Rakesh S Ramesh[2*], Chandrakala M[1], Apoorva D[1], Snijesh VP[1], Vishakha Ramamurthy[1], Vidya M Somashekaraiah[1], Anupama CE[1], Kiruthiga Raghunathan[3], Anuradha Panigrahi[3], Manjula Das[3], Sujan K Dhar[3], Jyothi S Prabhu[1#]

[1]Division of Molecular Medicine, St. John's Research Institute, St. John's Medical College, Bangalore, Karnataka, India

[2]Department of Surgical Oncology, St. John's Medical College and Hospital, Bangalore, Karnataka, India

[3]Molecular Immunology Program, MSMF, Narayana Health City, Bangalore 560099, India.

* These authors contributed equally to this work

Corresponding authors- Email id: madhumathy@sjri.res.in, jyothi@sjri.res.in

Keywords- Breast Cancer**,** Liquid biopsy, ctDNA, ALU 247, Prognosis, Disease progression

Abstract

Background: Liquid biopsy is widely recognised as an efficient diagnostic method in oncology for disease detection and monitoring. Though examination of circulating tumor cell (CTC) is mostly implemented for assessment of genomic aberrations, need of complex methodologies for their detection has impeded its acceptance in low resource settings. We evaluated the cell free DNA (cfDNA) as a liquid biopsy tool and investigated its utility in breast cancer patients.

Methods: Total cell free DNA was extracted from plasma of breast cancer patients (n=167) with median follow up of more than 5 years, at various stages of the disease. Quantitative PCR was performed to estimate the copy numbers of two fractions of ALU repetitive elements (ALU 115 and ALU 247) and DNA Integrity (DI) was calculated as the ratio of ALU 247/115. Mutations in TP53 and PIK3CA in the cfDNA was estimated by next-gen sequencing (NGS) in a subset of samples. Association of the levels of both the ALU fragments with various clinico-pathological factors and disease-free survival at various stages were examined. Nomogram models were constructed with clinical variables and ALU 247 levels to predict disease-free survival and best performing model was evaluated by decision curve analysis.

 doi:10.20944/preprints202212.0400.v1

Results: DI and ALU 247 levels were significantly lower ($p<0.0001$) in the post-operative plasma when compared to their pre-surgery levels. DI and ALU 247 were found to be significantly higher in patients with metastasis ($p<0.05$). Patients with higher levels of ALU 247 in their post-operative plasma had significantly poor disease-free survival ($p=0.005$). Higher level of ALU 247 in circulation also correlated with low tumor-infiltrating lymphocytes (TIL) within their primary tumors in the ER-negative breast cancer subtype ($p=0.01$). Cox proportional hazard analysis confirmed ALU 247 as an independent variable of disease-free survival both in univariate and multivariate analysis [HR 1.3 (95% CI 1.047 to 1.613, p= 0.017)]. Nomogram model showed addition of ALU 247 with other variables significantly improved (C-index-0.823) the predictive ability of the model.

Conclusion: Our results confirm the utility of cfDNA as an evolving liquid biopsy tool for molecular analysis. Evaluation of larger fragment of cfDNA estimated through ALU 247 can provide vital information concurrent with the pathological process of disease evolution in breast cancer and warrants expansion to other cancer types.

Introduction:

Evolution of targeted therapies and advanced imaging modalities have enhanced the survival rate of breast cancer patients, however 20-30% of early breast cancer patients die of metastatic disease. There is a requisite for identifying specific biomarkers and/or methods through which breast cancer disease progression could be serially monitored and the associated prognosis could be predicted early on. In the recent years, liquid biopsy has emerged as a cost-effective and minimally invasive technique and is getting identified as a competent method for diagnosis and screening of biomarkers in breast cancer. Liquid biopsy or fluid biopsy for cancer diagnosis, is a method of analysing the blood samples for the presence of circulating tumour cells (CTCs), circulating cell free DNA, RNA (ccfDNA, ccfRNA), circulating tumour DNA (ctDNA), exosomes etc. for monitoring cancer progression. The techniques are minimally invasive and can be utilized for longitudinal serial monitoring of patients. The source of ctDNA in a cancer patient can be from CTCs, primary tumor cells or metastatic cells lodged at distant metastatic sites in the body and can be used for detecting genomic patterns and alterations that evolve during treatment. Serum markers CEA or CA15-3 are still in use to follow-up disease but has a low specificity.

The major source of ccfDNA in healthy individuals comprises of short fragments of DNA which are <200 bp in length, are most often associated with histone proteins and hence deemed to be released by the process of apoptosis. In cancer patients however, necrotic events and endonuclease activities engage in fragmenting the chromatin into longer, nucleosomal units of 180 bp up to 1 kb in length and may be considered as representative of the ctDNA. This leads to increasing levels of ccfDNA with elevated levels of longer fragments in serum or plasma of cancer patients than in healthy controls and has been reported as an effective blood-based biomarker for various types of cancer (1-3).

ctDNA could be discerned from ccfDNA, qualitatively, by evaluating the tumour-specific mutations, methylations and gene amplifications of these fragments (4). However, only a

 doi:10.20944/preprints202212.0400.v1

limited proportion of patients carry such known mutations, for instance, only 30-40% of breast cancer patients carry PI3KCA mutations (4). Other known approaches also focus on quantitating the total ccfDNA levels using target genes such as GAPDH, β-globin with higher levels observed in malignant disease (4). However, these results may be influenced by other co-existing conditions like infection or inflammation which may have a likely effect on the ccfDNA levels. Another well-established methodology is a quantitative approach to assess the copy number of tumour DNA associated gene elements. In this regard, LINE elements, a group of self-replicating, retrotransposons, are the most used targets accounting for its high copy number, occupying nearly 15% of the total genome. Among them, ALU sequences ALU 115 and ALU 247 copy numbers are associated with the levels of ccfDNA and ctDNA respectively (5). The advent of real-time quantitative PCR (q-PCR) has enabled a quick and cost-effective method of detecting the target biomarker ALU elements in ccfDNA from the serum or plasma of cancer patients. The ratio between the long and short ccfDNA fragments (DI) have been used as a potential diagnostic marker in multiple cancer types including breast cancer (4, 6-8). In this study, we have used liquid biopsy to evaluate prognosis and monitor disease progression in breast cancer patients by assessing the small and large ALU repetitive element levels. Levels of ALU 247 and the DI has been used as a measure to gauge ctDNA levels which was then correlated with various clinico-pathological features in an attempt to identify its utility to predict disease prognosis.

Methods:

Patient recruitment:

Blood samples (n=254) were accessed from 167 breast cancer patients at various stages during disease diagnosis and treatment at St. John's Medical College and Hospital, Bengaluru. 111 samples were collected pre-operatively (treatment naïve) immediately after diagnosis. 110 were post-operative, 20 were metastatic and 13 were collected from patients who were under remission at median follow up of 47 months. For 88 patients, we had access to paired pre and post-operative samples. 66 patients with pre-operative samples and 61 patients with post-operative samples had long-term follow up with median of 72 months. Blood samples collected from metastatic patients either presented with metastasis or were at median follow up of 32 months after their primary diagnosis. A written informed consent was obtained from all participants and the Institutional Ethics Committee approved the study (IERB No: 167/2019, 62/2008).

Plasma preparation and DNA extraction:

5 ml of venous blood samples were collected from patients into EDTA tubes, centrifuged for 10 minutes at 2500 rpm. The plasma was carefully separated and stored in -80° C until further use. All samples were processed within 2 hours of collection. DNA extraction was carried out using QIAamp DNA Blood Mini Kit (QIAGEN, Germany) as per manufacturer's protocol. The DNA was subjected to Qubit-based quantitation and stored in -20° C.

Quantitative PCR of ALU repeats:

Quantitative PCR for ALU repeats were performed as described previously in Umetani et al., and Iqbal et al., (9, 10). Briefly primer sequences were designed to amplify ALU sequences of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0400.v1

ALU 115- a 115 bp DNA amplicon and ALU 247- a 247 bp DNA amplicon. The primer sequences are; ALU 115 Forward primer- 5'-CCTGAGGTCAGGAGTTCGAG-3'; Reverse primer- 5'-CCCGAGTAGCTGGGATTACA-3'; ALU 247 Forward primer- 5'-GTGGCTCACGCCTGTAATC-3'; Reverse primer- 5'-CAGGCTGGAGTGCAGTGG-3'. The ALU 115 amplifies both short and long fragments of the DNA and hence represents the total amount of circulating free DNA (cfDNA) in the plasma that is released from apoptotic cells while ALU 247 amplifies the long fragments of the DNA hence represents the DNA fragments released from cells because of necrosis (ctDNA). DNA levels were measured by q-PCR in duplicates with 100pg template per reaction, using SYBR Green master mix on the LightCycler 480 II (Roche Diagnostics) with a 3 pM concentration of ALU 115 and ALU 247 primers making the reaction volume 10 µL. The DNA copy number corresponding to amplified ALU 115 and ALU 247 were evaluated by normalisation with known concentrations of genomic DNA by deriving a standard curve, and the DI index was calculated as a ratio of ALU 247 to ALU 115 copy numbers.

Next-Gen Sequencing- primer synthesis, library preparation and analysis:

The mutations identified in DNA samples from breast cancer patients were verified through database analysis to detect concurrent mutations common in breast cancer patients. Through cBioPortal, the TCGA data for breast cancer patients of 818 patients were checked for mutations using TP53, and PIK3CA as the query genes. cfDNA samples were sequenced by amplicon sequencing for specific variants of *TP53* and *PIK3CA* using primers as listed in Supplementary Table 1. Primers were designed based on the guideline provided in 16S metagenomic sequencing library preparation (15044223 B) manual. Amplification of the target regions in the samples was carried out with JumpStart kit (P2893-100RXN) by multiplexing all primers. The library was prepared by ligating unique barcodes from IDT for Illumina DNA/RNA UD Indexes (20026121). The pooled libraries were sequenced on NovaSeq 6000 (Illumina) platform using 2 X 150 bp chemistry. Reads were quality filtered using Trimmomatic Bolger, Lohse (11) and aligned to the GRCh38 reference using Bowtie2 (12) From the alignment map variants were called using multiple pipelines including freebayes (13) and bcftools (14). Called variants were analysed and annotated using wANNOVAR (15) and Ensembl Variant Effect Predictor (16) web-based tools.

Table 1: Association between ALU 115, ALU 247 and DNA integrity ratio with various clinico-pathological features of primary tumours from n=132 breast cancer patients

Sl No.	Parameter (n)	Category	Rel. frequency /category (%)	qPCR ALU115 Mean	*p* value (Mann-Whitney)	qPCR ALU247 Mean	*p* value (Mann-Whitney)	qPCR DNA Integrity ratio Mean	*p* value (Mann-Whitney)
1	Age (129)	<50	29	0.227	0.837	0.611	0.360	1.485	0.464
		>50	71	0.233		0.364		0.966	
2	Age at Menarche (118)	Early	14	0.197	0.259	0.402	0.453	1.029	0.462
		Late	86	0.252		0.490		1.175	

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0400.v1

3	Menopausal status (129)	Pre Post	26 74	0.225 0.233	0.912	0.554 0.398	0.520	1.242 1.071	0.864
4	Breast side (129)	Left Right	55 43	0.227 0.234	0.846	0.488 0.350	0.679	1.304 0.861	0.436
5	Diabetic (109)	Yes No	33 67	0.240 0.253	0.585	0.295 0.597	0.087	0.733 1.477	0.053
7	Hypertension (114)	Yes No	48 52	0.250 0.241	0.986	0.433 0.518	0.313	1.097 1.376	0.372
8	Tumour grade (128)	1 2 3	19 55 21	0.228 0.244 0.195	1 v/s 2- 0.641 1 v/s 3- 0.788 2 v/s 3- 0.353	0.155 0.561 0.300	1 v/s 2- 0.704 1 v/s 3- 0.952 2 v/s 3- 0.623	0.464 1.396 0.873	1 v/s 2- 0.494 1 v/s 3- 0.886 2 v/s 3- 0.577
9	Tumour size (127)	<3cm >3cm	40 60	0.185 0.250	**0.048**	0.437 0.414	0.409	1.094 1.102	0.639
10	Lymph node status (125)	Positive Negative	42 58	0.249 0.221	0.407	0.394 0.471	0.749	0.964 1.202	0.348
11	Lympho-vasular invasion (108)	Present Absent	40 60	0.244 0.230	0.597	0.376 0.489	0.844	0.953 1.273	0.526
12	Tumour stage (117)	0 1 2 3	3 15 50 32	0.170 0.161 0.238 0.291	1 v/s 0- 0.953 2 v/s 0- 0.344 3 v/s 0- 0.166 2 v/s 1- 0.121 3 v/s 1- **0.025** 3 v/s 2- 0.209	0.048 0.435 0.515 0.433	1 v/s 0- 0.953 2 v/s 0- 0.210 3 v/s 0- 0.439 2 v/s 1- 0.141 3 v/s 1- 0.328 3 v/s 2- 0.828	0.287 1.126 1.340 1.038	1 v/s 0- 0.961 2 v/s 0- 0.513 3 v/s 0- 0.903 2 v/s 1- 0.304 3 v/s 1- 0.644 3 v/s 2- 0.387
13	Ki-67 index (91)	<15 >15	40 60	0.279 0.188	**0.044**	0.531 0.408	0.218	1.290 1.153	0.890
14	Immune infiltrate (120)	Mild High	43 43	0.252 0.218	0.309	0.572 0.374	0.263	1.456 0.980	0.098

Statistical analysis:

Descriptive statistics was used for all clinical variables. Differences in estimated levels of DNA fragments and DNA integrity (DI) between the blood samples collected during different stages of disease as mentioned above (in the section-patient recruitment) were assessed using Mann-Whitney U test or two-tailed student's t-test. Kaplan-Meier analysis was used to examine the estimated differences in disease-free survival. Disease-free survival was calculated as the time from the date of first diagnosis to the time when either a local recurrence or a distance metastasis occurred. Nelson-Aalen analysis, a non-parametric estimator of the cumulative hazard function was also performed to analyse the survival function. Patients without an event or had succumbed to non-breast cancer-related causes were right censored. Log-rank test (Mantel-Cox) was used to compare the survival between groups. Both univariate and multivariate cox-proportional hazard analysis were done to validate the prognostic importance of ALU 247 and DI in comparison to other clinico-pathological characteristics. A nomogram was constructed using R software (version 4.1.3) with package "rms" (version 6.3.0). The performance of the nomogram was assessed by the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0400.v1

Harrell's concordance index (C-index). The C-index has a range from 0.5 to 1.0, with 0.5 indicating random chance and 1.0 considered perfect discrimination. It was used to assess the accuracy and identification abilities of the predictive factors. To estimate the clinical utility of the nomogram, decision curve analysis using r packge "dcurves" (version 0.3.0) was performed by calculating the net benefits for a range of threshold probabilities. Predicting variables were evaluated for their association with the occurrence of an event (metastasis or death due to metastasis) using multivariate logistic regression models and LASSO regression models. For all tests, a p value of <0.05 was considered to be statistically significant.

Results:

Patient characteristics

A total of 254 plasma samples isolated from 167 patients belonging to various stages of disease were subjected to cfDNA analysis by q-PCR. 44% of the samples were treatment naïve collected before surgery (pre-operative), 43% were collected within 48 hours after surgery (post-operative). 8% were collected from patients with metastasis at presentation. Samples were also collected from patients under remission at varying time points. Samples collected anytime from 24 months to 56 months were grouped as short-term remission (2%) and from 56 months to 13 years as long-term remission (3%) (Supplementary Figure 1). The mean age of all the patients were 56 ± 13 years. The mean tumor size in treatment naïve patients undergoing surgery was 3.5 ± 1.8 cm. The Hormone Receptor (HR+) and triple negative (TNBC) subtype accounts for 55% and 26% of these tumors.

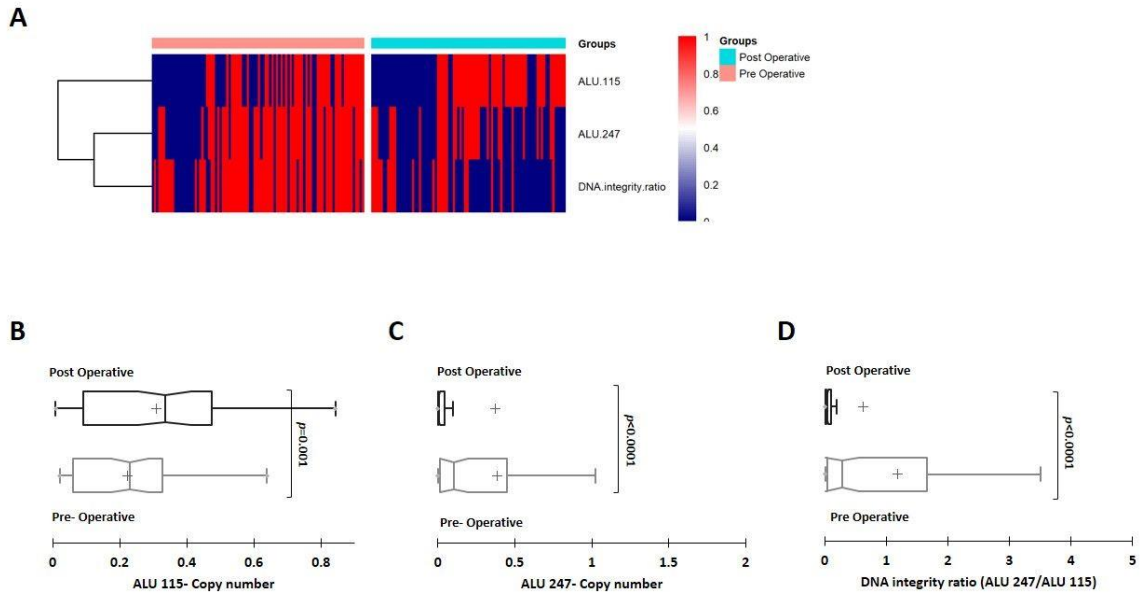

Figure 1.A: Heat map depicting the levels of ALU 115, ALU 247 and the DI index among post-operative and pre-operative samples. B, C, D: Univariate analysis depicting levels of ALU 115, ALU 247 and DI ratio, respectively, among matched/paired post-operative and pre-operative samples (n=88 each).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0400.v1

Evaluation of ctDNA & DI index; ALU 247 levels decrease post-surgical intervention and is higher in metastatic patients

After computing the copy numbers of ALU 115, ALU 247 and the DI index, we first compared them between all pre-operative and post-operative samples. ALU 115 levels (representing the total cfDNA) were observed to be higher in the post-operative group, but this difference was not statistically significant (p=0.1). However, the ALU 247 (representing the ctDNA) and the DI index was significantly lower in the post-operative groups (p<0.0001) (Figure 1A). ALU 115, ALU 247 and the DI index was also compared between matched/paired pre and post-operative samples (n=88). We observed similar trends with ALU 115 being higher in the post-operative group (p<0.05) in these matched samples. As observed earlier, ALU 247 and the DI index was significantly lower in the post-operative samples (p<0.0001) (Figure 1B, C, D).

The levels of ALU 115, ALU 247 and the DI index was then compared between the samples collected during the various stages of disease in different breast cancer patients. ALU 115 was found to be significantly lower in patients under remission (combined short-term and long-term) when compared to pre-operative (p=0.002) and post-operative (p=0.003) groups. This was also lower in the metastatic group when compared to the post-operative samples (p=0.01), and although not statistically significant, lower when compared to the pre-operative samples (p=0.08). ALU 247 and the DI index was, however, significantly higher in the metastatic group when compared to the post-operative (p=0.06 and p=0.01 respectively) (Figure 2A, B, C) group.

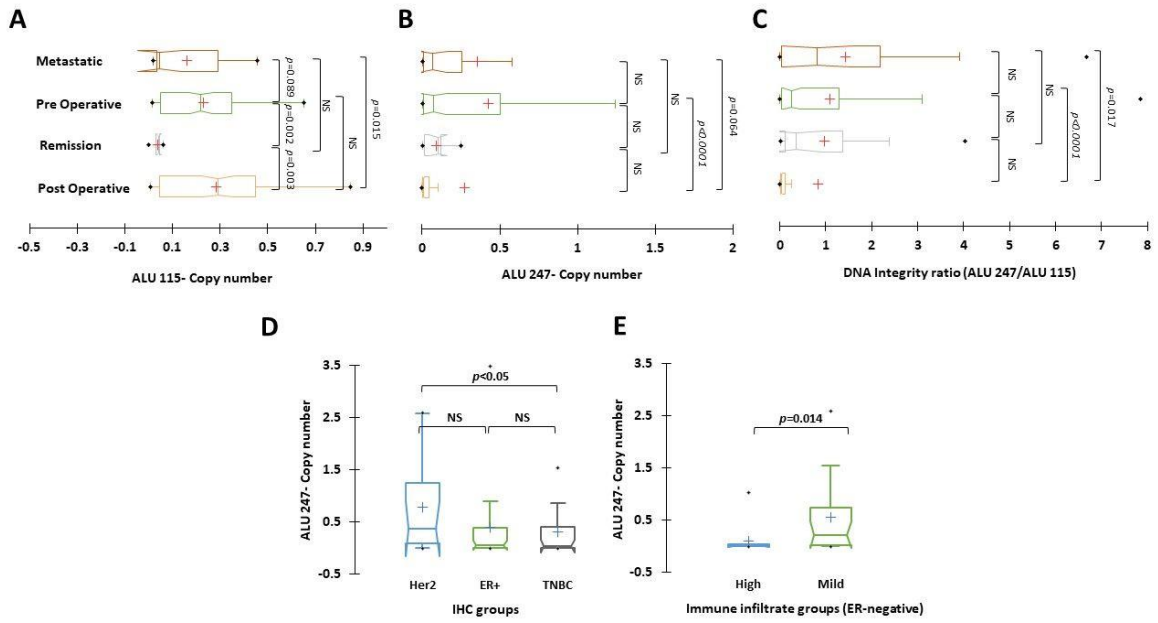

Figure 2. A, B, C: Univariate analysis depicting levels of ALU 115, ALU 247 and DI ratio, respectively, among metastatic (n=20), pre-operative (n=111), remission (n=12) and post-operative samples (n=108).

 doi:10.20944/preprints202212.0400.v1

D: Comparative analysis of ALU 247 levels among IHC groups Her2 (n=16), ER+ (n=57) and TNBC (n=30) and E: between dense/moderate (n=16) and mild (n=23) immune infiltrate groups of the ER-negative subtype.

Implication of baseline ALU 115, ALU 247 and DI levels on prognosis and association with tumor clinico-pathological characteristics

To verify the prognostic importance of the DI index, we performed Kaplan-Meier survival analysis on a subset of samples where follow-up information was available. We first examined the pre-operatively collected samples to evaluate the prognostic implication of the ALU markers and DI index in them. To maximize the specificity, we chose the cut-off at third quartile (0.492 for ALU 115, 0.042 for ALU 247 & 0.08 for DI index) and divided the samples into high and low groups based on the cut-off value. In the 50 samples where the analysis was performed, we did not observe any statistically significant difference in survival based on stratification by ALU 115, ALU 247 or DI ($p>0.05$). Further to examine association of ALU levels and DNA index with tumor clinico-pathological characteristics, we analysed the known clinical features associated with pre-operatively collected tumors (treatment naïve) and performed a correlative analysis. The pre-operative samples were stratified based on IHC classification into Hormone Receptor Positive (HR+); n=57, HER2+; n=16, and the Triple Negative Subtype (TNBC); n=30. DI index and ALU 247 were significantly higher in the HER2 subtype when compared to the TNBC ($p<0.05$) and relatively higher than the HR+ subtypes although not statistically significant (Figure 2D, Supplementary Figure 2). There was no significant difference in the distribution of ALU 115 levels across the various subtypes. ALU 115 levels however increased with higher tumor stage and size, but lower levels were associated with a higher Ki67 index ($p<0.05$) (Table 1). We did not observe any significant associations with other clinico-pathological features like tumor size, grade, stage or lymph node status. (Table 1 and Supplementary Table 2). Although, the DI index was found to be higher in tumors with a higher grade and stage, the observation was not statistically significant ($p>0.05$). The samples were then stratified based on ER-positivity and the same analysis was performed between ER-negative and ER-positive tumors. We did not observe any associations with any of the clinico-pathological features in the ER-positive subtype. However, we observed that the DI index and ALU 247 levels were positively associated with a mild immune infiltrate in the ER-negative subtype (n=39). The tumors were separated into groups (Dense/Moderate & Mild) based on the immune filtrate as recorded in the pathological reports. Tumors with mild immune infiltrate were associated with a significantly higher DI index ($p=0.01$) and higher amount of circulating tumor DNA as represented by ALU 247 ($p=0.01$) (Figure 2E, Supplementary Figure 3).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0400.v1

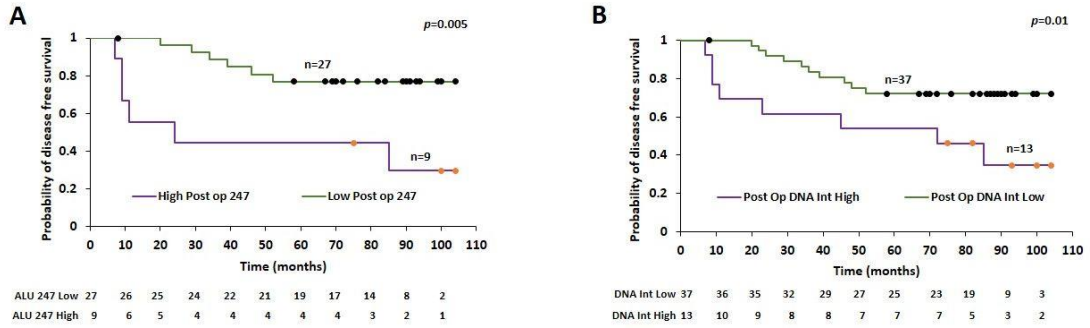

Figure 3. A: Disease-free survival between ALU 247 low (n=27) and ALU 247 high (n=9) samples and B: between DI low (n=37) and DI high (n=13) samples.

We then evaluated the prognostic implication of ALU 115, ALU 247 or DI on post-operatively collected samples by Kaplan-Meier analysis. The third quartile cut-off was again used for all the analysis (0.498 for ALU 115, 0.03 for ALU 247 & 0.07 for DI index) and samples divided into high and low groups based on the cut-off. ALU 115 did not demonstrate any statistically significant separation based on survival (p=0.1). However, stratification of the samples by both ALU 247 levels (log rank p=0.005) and the DI index (log rank p=0.01) produced significant separation of the groups based on disease-free survival. The disease-free survival at 6-year mean follow-up dropped from 84 months in ALU 247 low samples to 44 months in ALU 247 high samples and from 82 months in DI low samples to 58 months in DI high tumors (Figure 3A & B).

Table 2. Univariate and Multivariate Cox-proportional hazard analysis done to validate the prognostic importance of ALU 247 in comparison to other clinico-pathological characteristics.

	All; n = 36			
	Univariate		**Multivariate**	
	HR (95% CI)	p-value	HR (95% CI)	p-value
Age	1.006 (0.965-1.05)	0.76	1.013 (0.972-1.056)	0.537
T-size	1.416 (0.911-2.203)	0.12	1.604 (0.990-2.601)	**0.05**
Lymph Node status				
Negative	Reference			
Positive	9.441 (2.026-44.003)	**0.004**	8.073 (1.611-40.449)	**0.01**
Grade				
I and II	Reference			
III	0.808(0.175-3.735)	0.78	1.347(0.248-7.300)	0.73
Post-Operative ALU 247	1.302 (1.074 – 1.578)	**0.007**	1.3 (1.047-1.613)	**0.017**

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0400.v1

Detection of mutations from ctDNA by NGS

To confirm if the DNA fragments evaluated by ALU 247 were derived from tumor DNA, we examined them for commonly observed mutations in breast cancer using NGS approach. An NGS based short panel analysis was done on 19 random samples with high ALU 247 levels as described in materials and methods to detect concurrent mutations commonly observed in *TP53* and *PIK3CA* among breast cancer patients. PIK3CA variant [179234250-251 (TG>CT)] was detected in 17/19 samples and TP53 variant [7673432-434 (TGG>GGC)] was detected in 4/19 blood samples tested. This result confirmed that DNA identified by higher ALU 247 levels are likely to represent the breast tumor DNA in circulation.

Utility of estimating post-surgical ALU 247 levels for predicting prognosis

Encouraged by the observation that ALU 247 levels and DI estimated post-operatively, identified patients with poor prognosis based on disease-free survival analysis we wanted to further probe the clinical utility of these for predicting disease progression. The prognostic value of ALU 247 and DI was further verified using univariate cox-proportional hazard analysis, (Table 2). Risk associated with high ALU 247, and DI levels were estimated with other clinical variables including age, tumor size, grade and lymph node status. ALU 247 levels and lymph node status emerged as highly significant independent variables after univariate analysis, while, DI, age, tumor size and grade were of no significance – ALU 247: HR 1.302 (95% CI 1.074 – 1.578, p=0.007); Lymph node status: HR 9.441 (95% CI 2.026 to 44.003, p=0.004) DI: HR 0.856 (95% CI 0.659 – 1.110, p=0.24). Since only ALU 247 and not DI emerged as a significant variable, further analysis was performed with ALU 247. Nelson-Aalen analysis, a non-parametric estimator of the cumulative hazard function also showed significant stratification of the ALU 247 high and low groups based on disease-free survival (log-rank; p=0.005 and Wilcoxon; p=0.004). In multivariate analyses, with the other prognostic markers, ALU 247 levels measured post surgically evolved as an independent prognostic factor [ALU 247: HR 1.3 (95% CI 1.047 to 1.613, p=0.017)] (Table 2).

A Lasso regression model was then used to assess weight of the ALU 247 towards determining the occurrence of an event (metastasis or death due to disease) along with other clinical parameters of age, tumor size, grade and lymph node status. The model predicted ALU 247 levels to be the most influential factor with the highest variable importance score and co-efficient value of 1.789 (Supplementary Figure 4) along with other clinical parameters. The Lasso regression model by including ALU 247 gave an improved R^2 of 0.36 and a lower Root Mean Square Error (RMSE) of 0.38 when compared to the model with all other conventional parameters of age, tumor size, grade and lymph node status excluding ALU 247 (R^2 of 0.30 and a RMSE of 0.399). Cross-validation confirmed that the model where ALU 247 is combined with conventional parameters, produced a lower lambda value of 0.07 (when compared to 0.087 for the conventional model) and hence emerges as the more robust performing model. Further a logistic regression model and analysis performed by including ALU 247 with other parameters improves the sensitivity of the assay by 8% and renders a higher AUC of 0.875 (Figure 4A).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0400.v1

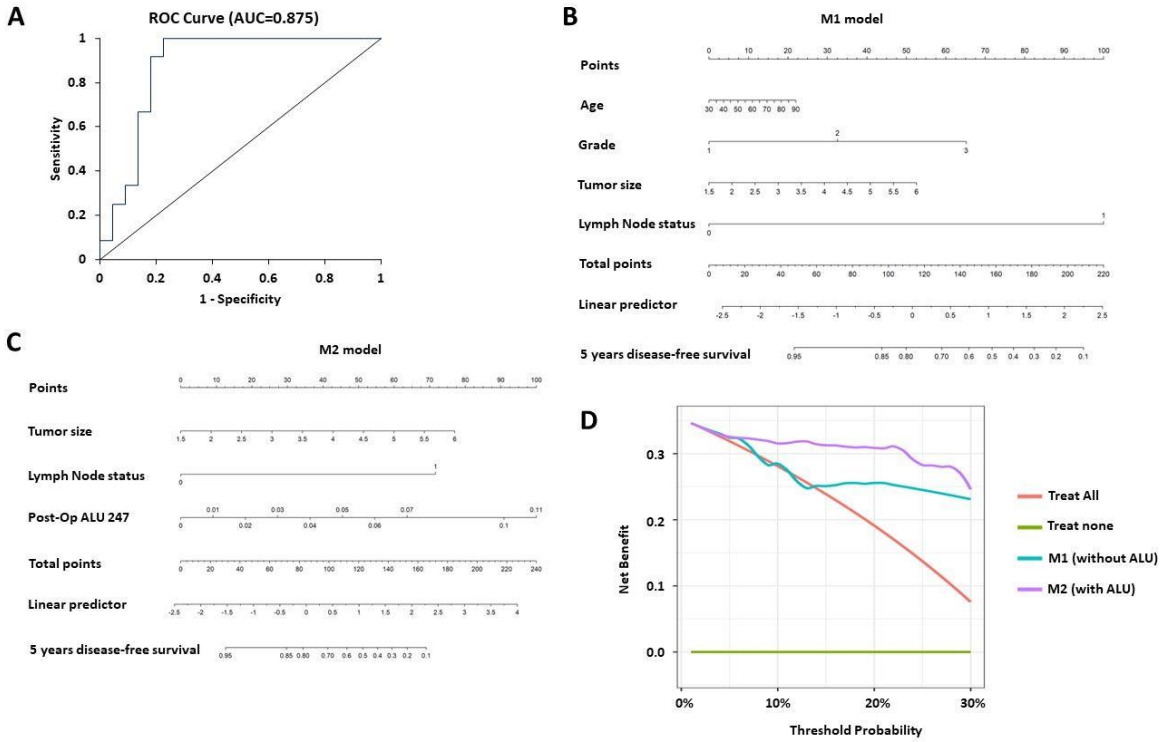

Figure 4. A: ROC curve for ALU 247 with other parameters with AUC analysis derived by logistic regression. B: Nomogram analysis with M1 model depicting disease-free survival based on conventional parameters. C: Nomogram analysis with M2 model depicting disease-free survival based on conventional parameters along with post-operative ALU 247 levels. D: Decision curve analysis derived for the nomogram M1 and M2 depicting net benefit of the model across different threshold probabilities.

Nomogram and decision curve analysis for prediction of prognosis

Next, we constructed various nomograms using conventional parameters of age, tumor size, lymph node status and grade. Different models were constructed, using clinical parameters alone (M1) and using 2 clinical parameters along with post-operative ALU 247 that exhibited significant differences in the multivariate analysis: tumor size and lymph node status (M2). Each factor was ascribed a weighted point, and the total points indicated the risk of disease-free survival. The variables used as factors and final nomogram model are presented in Figure 4B & C. To evaluate the predictive accuracy of the nomogram prediction system, the C-index was calculated and validated between the 2 models. For the nomogram model built with the conventional parameters (M1), the C-index was 0.772. For the model M2 after inclusion of post-operative ALU 247, C-index improved to 0.823 which indicates a better ability to assign patients to an accurate risk stratification and a higher C-index of more than 0.8 suggesting a stronger model. A decision curve analysis was then derived to assess net benefit of the model across different threshold probabilities. The decision curve analysis for the 2 models is presented in Figure 4D. The analysis predicted that the model integrating ALU 247 along with other clinical parameters was of value at a threshold of 25% at 5 years with a net benefit of 0.039. These results are suggestive that ALU 247 levels measured post-operatively can be

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0400.v1

used as a tool to predict prognosis in breast cancer. It adds value to already employed clinical parameters that are routinely used to assess risk of progressive disease after surgical intervention.

Discussion

Biopsy is still deemed the right tool for confirmation of cancer diagnosis but has apparent deficiencies that include inability to capture the dynamic changes incurred by anti-cancer therapies and other genomic insults that occur during the course of treatment. In addition, serial biopsies may not always be feasible in the clinical setting. The use of circulating biomarkers from the peripheral blood as a liquid biopsy tool has been subjected to exploration over the past several years due to the apparent advantage of being minimally invasive. Analysis of circulating nucleic acids and CTCs provide a diversity of data including genomic changes, treatment response, and prognosis which can not only be used for detection but also to monitor the disease progression. The ratio between the long and short cfDNA fragments (DI) has been shown to be a potential diagnostic marker in multiple cancer types as described below. DI was found to be significantly increased in patients with colorectal cancer (17, 18) and found to be correlated with poor prognosis in these patients (19). In non-small cell lung cancer higher levels of ALU 247 and 115 were associated with higher stages and metastatic disease (20). High levels of ctDNA are associated with a more aggressive, potentially resistant disease and have been detected both in early and later stages in breast cancer (21-23). We have estimated the levels of ALU 115, ALU 247 and the DI index in breast cancer patients at various stages of disease and observed a significant reduction in ALU 247 values (representing the ctDNA) and the DI index in the post-operative period. This result could be in resonance and reflect effective surgical removal of the primary tumor. A previous study by Iqbal et al., has reported similar results where they observed a significant decline in DI after surgery while this change was not observed for ALU 247 (24). Elhelaly et al., and Hassan et al., have also recently reported similar findings by assessing cell free DNA quantity post-surgery, and there was a drop in DNA concentrations after elimination of macroscopic tumor burden (25, 26).

We also observed a significant reduction in the total cell free DNA levels identified by ALU 115 and a marked increase in the levels of ALU 247 and the DI index in the metastatic patients when compared to post-operative samples. The increase in the baseline DI in relapsed patients than in patients who were free of disease was also reported by Iqbal et al., (24). In another study, using LINE-1 and ALU markers, baseline cfDNA and DI was shown to be independent prognostic markers in metastatic breast cancer patients (27). Ability of DI alone to detect metastasis has been however contradictory as another report observed that evaluating levels of both plasma ALU 247 and ALU 115 were useful in identifying patients who developed metastasis in breast cancer, but cfDNA integrity failed to do this (28). Changes in sample processing, experimental procedures and methodologies involved in DNA extraction could have contributed to these variations. To overcome this, we have employed the ALU primer pairs that have been widely used by other groups and standard methodologies to confirm (18, 28-32) the authenticity of the results we have obtained.

 doi:10.20944/preprints202212.0400.v1

Further evaluation of the prognostic implication of the baseline (pre-operative levels) ALU and the DI indices did not yield statistically significant difference in survival in our study, as opposed to earlier reports in a Finnish cohort where DI was measured using a fluorometer and another meta-analysis of 8 different studies which performed ctDNA gene variation detection (22, 33). In these studies levels of pre-operative ctDNA levels were significantly associated with shorter disease-free survival. Further correlation of ALU levels and DI with tumor clinico-pathological characteristics, showed that DI was significantly higher in the HER2 subtype. This observation is in accordance with Hussein et al., (28) where they observed a marked increase of cfDNA integrity in breast cancer patients with HER2 amplification which may be attributed to increased proliferation associated with the HER2 subtype.

We did not observe any associations with other clinico-pathological features like tumor size, grade, stage, or lymph node status in accordance with Hussein et al., however, it is noteworthy that some studies have reported a positive correlation of DI index with TNM staging (23), to size of invasive cancers and the presence of lymphovascular invasion or lymph node status (34). Interestingly, very few studies have recorded the correlation of immunotherapy resistance and ctDNA levels in breast cancer. Prescence of high microsatellite instability detected in ctDNA is reported to be associated with a good response to immunotherapy and assessment of dynamic levels of ctDNA during pembrolizumab treatment revealed that a rise of ctDNA levels during treatment correlated with a progressive disease and poor survival (35, 36). We report for the first time a positive correlation between a milder immune infiltrate and ALU 247 & DI levels in the ER-negative tumors. A milder immune infiltrate is associated with a poorer prognosis in the ER-negative subtype of breast cancer (37) and higher levels of ctDNA as measured by our assay may aid in stratifying tumors based on immune infiltrate. This observation needs additional validation on a larger set of tumors and if proven useful, may profoundly benefit in identifying tumors that may qualify for immunotherapy especially in the metastatic setting.

Although our results support the prognostic potential of the DI index, with multivariate survival analysis DI index failed to emerge as an independent prognostic factor. The DI index is computed as a relative proportion and is contingent on ALU 115 levels. The ALU 115 represents cfDNA, and the levels of the same may vary depending on multiple physiological factors including inflammation and other co-morbidities. This could partially explain why DI index failed to emerge as an independent prognostic factor. However, ALU 247 emerged as an independent prognostic factor with Cox regression survival analysis as larger fragments are likely to be derived from tumor DNA, further confirmed by mutations observed by sequencing. This prompted us to explore the possibility of inclusion of ALU 247 along with other conventional parameters to assess disease progression in breast cancer. Inclusion of ALU 247 levels in the lasso and multivariate logistic regression model improved the predictive performance of the model. Cross-validation confirmed that a model where ALU 247 was combined with conventional parameters provided a more robust performance and a lesser mean squared error. Nomogram based decision curve analysis suggests that the combined use of established clinico-pathological features and liquid biopsy could help to discriminate breast cancer patients prone for poor disease-free survival during the immediate post-operative period.

 doi:10.20944/preprints202212.0400.v1

This study has several limitations including, a retrospective study design and limited number of samples used for survival and model construction. Availability of the long term follow up information in only a small number of patients was a disadvantage. Nucleic acid extraction from blood samples poses significant analytical and experimental challenges and despite having accessed stored blood samples, we have followed strict quality control measures and protocols to eliminate such hitches. To prevent genomic DNA contamination, we have employed plasma as the source of ctDNA which is in line with other studies performed (38). Though our results indicate that the levels of ALU 247 are likely to be derived from tumor DNA, this validation was done for limited number of mutations in a small subset of samples. This needs further validation with gold standard methods such as analysis of a larger panel of frequently mutated loci of cancer associated genes through NGS.

Conclusion: Our observations support the utility of evaluating the large circulating DNA fragments in the plasma of breast cancer patients for prognosis and monitoring disease progression. Together with conventional clinico-pathological features, this non-invasive and easily reproducible liquid biopsy assay can be used to identify patients prone for poor prognosis. Techniques employed in our study are rapid, low-cost, and accessible diagnostics and can be deployed as supplementary measures to the consensus-based gold standard sequencing approaches employed in liquid biopsy.

Reference

1. Lone SN, Nisar S, Masoodi T, Singh M, Rizwan A, Hashem S, et al. Liquid biopsy: a step closer to transform diagnosis, prognosis and future of cancer treatments. Molecular Cancer. 2022;21(1):79.
2. Perakis S, Speicher MR. Emerging concepts in liquid biopsies. BMC Medicine. 2017;15(1):75.
3. Crowley E, Di Nicolantonio F, Loupakis F, Bardelli A. Liquid biopsy: monitoring cancer-genetics in the blood. Nature Reviews Clinical Oncology. 2013;10(8):472-84.
4. Sobhani N, Generali D, Zanconati F, Bortul M, Scaggiante B. Cell-free DNA integrity for the monitoring of breast cancer: Future perspectives? World J Clin Oncol. 2018;9(2):26-32.
5. Buono G, Gerratana L, Bulfoni M, Provinciali N, Basile D, Giuliano M, et al. Circulating tumor DNA analysis in breast cancer: Is it ready for prime-time? Cancer Treat Rev. 2019;73:73-83.
6. Shaban SA, Al-Rahim AM, Suleiman AA. ALU repeat as potential molecular marker in the detection and prognosis of different cancer types: A systematic review. Mol Clin Oncol. 2022;16(4):86.
7. Khani M, Hosseini J, Mirfakhraie R, Habibi M, Azargashb E, Pouresmaeili F. The value of the plasma circulating cell-free DNA concentration and integrity index as a clinical tool for prostate cancer diagnosis: a prospective case-control cohort study in an Iranian population. Cancer Manag Res. 2019;11:4549-56.
8. Kumari S, Husain N, Agarwal A, Neyaz A, Gupta S, Chaturvedi A, et al. Diagnostic Value of Circulating Free DNA Integrity and Global Methylation Status in Gall Bladder Carcinoma. Pathology & Oncology Research. 2019;25(3):925-36.
9. Iqbal S, Vishnubhatla S, Raina V, Sharma S, Gogia A, Deo SS, et al. Circulating cell-free DNA and its integrity as a prognostic marker for breast cancer. Springerplus. 2015;4:265.

10. Umetani N, Giuliano AE, Hiramatsu SH, Amersi F, Nakagawa T, Martino S, et al. Prediction of breast tumor progression by integrity of free circulating DNA in serum. J Clin Oncol. 2006;24(26):4270-6.
11. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20.
12. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Methods. 2012;9(4):357-9.
13. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:12073907. 2012.
14. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2).
15. Chang X, Wang K. wANNOVAR: annotating genetic variants for personal genomes via the web. J Med Genet. 2012;49(7):433-6.
16. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biology. 2016;17(1):122.
17. Leszinski G, Lehner J, Gezer U, Holdenrieder S. Increased DNA integrity in colorectal cancer. In Vivo. 2014;28(3):299-303.
18. El-Gayar D, El-Abd N, Hassan N, Ali R. Increased Free Circulating DNA Integrity Index as a Serum Biomarker in Patients with Colorectal Carcinoma. Asian Pac J Cancer Prev. 2016;17(3):939-44.
19. Zhu F, Ma J, Ru D, Wu N, Zhang Y, Li H, et al. Plasma DNA Integrity as a Prognostic Biomarker for Colorectal Cancer Chemotherapy. Journal of Oncology. 2021;2021:5569783.
20. Soliman SE, Alhanafy AM, Habib MSE, Hagag M, Ibrahem RAL. Serum circulating cell free DNA as potential diagnostic and prognostic biomarker in non small cell lung cancer. Biochem Biophys Rep. 2018;15:45-51.
21. Madhavan D, Wallwiener M, Bents K, Zucknick M, Nees J, Schott S, et al. Plasma DNA integrity as a biomarker for primary and metastatic breast cancer and potential marker for early diagnosis. Breast Cancer Research and Treatment. 2014;146(1):163-74.
22. Cullinane C, Fleming C, O'Leary DP, Hassan F, Kelly L, O'Sullivan MJ, et al. Association of Circulating Tumor DNA With Disease-Free Survival in Breast Cancer: A Systematic Review and Meta-analysis. JAMA Network Open. 2020;3(11):e2026921-e.
23. Kamel AM, Teama S, Fawzy A, El Deftar M. Plasma DNA integrity index as a potential molecular diagnostic marker for breast cancer. Tumor Biology. 2016;37(6):7565-72.
24. Iqbal S, Vishnubhatla S, Raina V, Sharma S, Gogia A, Deo SSV, et al. Circulating cell-free DNA and its integrity as a prognostic marker for breast cancer. SpringerPlus. 2015;4(1):265.
25. Elhelaly R, Effat N, Hegazy MAE, Abdelwahab K, Hamdy O, Abo Hashem EM, et al. Circulating Cell Free DNA and DNA Integrity Index as Discriminating Tools between Breast Cancer and Benign Breast Disease. Asian Pac J Cancer Prev. 2022;23(2):545-52.
26. Hassan F, Wang JH, Cullinane C, Ita M, Corrigan M, O'Leary DP, et al. Assessment of cell-free DNA (cfDNA) concentrations in the perioperative period can predict risk of recurrence in patients with non-metastatic breast cancer. Surg Oncol. 2022;42:101753.
27. Cheng J, Holland-Letz T, Wallwiener M, Surowy H, Cuk K, Schott S, et al. Circulating free DNA integrity and concentration as independent prognostic markers in metastatic breast cancer. Breast Cancer Research and Treatment. 2018;169(1):69-82.
28. Hussein NA, Mohamed SN, Ahmed MA. Plasma ALU-247, ALU-115, and cfDNA Integrity as Diagnostic and Prognostic Biomarkers for Breast Cancer. Applied Biochemistry and Biotechnology. 2019;187(3):1028-45.
29. Arko-Boham B, Aryee NA, Blay RM, Owusu EDA, Tagoe EA, Doris Shackie E-S, et al. Circulating cell-free DNA integrity as a diagnostic and prognostic marker for breast and prostate cancers. Cancer Genetics. 2019;235-236:65-71.

 doi:10.20944/preprints202212.0400.v1

30. Fawzy A, Sweify KM, El-Fayoumy HM, Nofal N. Quantitative analysis of plasma cell-free DNA and its DNA integrity in patients with metastatic prostate cancer using ALU sequence. J Egypt Natl Canc Inst. 2016;28(4):235-42.
31. Vizza E, Corrado G, De Angeli M, Carosi M, Mancini E, Baiocco E, et al. Serum DNA integrity index as a potential molecular biomarker in endometrial cancer. J Exp Clin Cancer Res. 2018;37(1):16.
32. Kumari S, Husain N, Agarwal A, Neyaz A, Gupta S, Chaturvedi A, et al. Diagnostic Value of Circulating Free DNA Integrity and Global Methylation Status in Gall Bladder Carcinoma. Pathol Oncol Res. 2019;25(3):925-36.
33. Lamminaho M, Kujala J, Peltonen H, Tengström M, Kosma V-M, Mannermaa A. High Cell-Free DNA Integrity Is Associated with Poor Breast Cancer Survival. Cancers. 2021;13(18):4679.
34. Umetani N, Giuliano AE, Hiramatsu SH, Amersi F, Nakagawa T, Martino S, et al. Prediction of Breast Tumor Progression by Integrity of Free Circulating DNA in Serum. Journal of Clinical Oncology. 2006;24(26):4270-6.
35. Bratman SV, Yang SYC, Iafolla MAJ, Liu Z, Hansen AR, Bedard PL, et al. Personalized circulating tumor DNA analysis as a predictive biomarker in solid tumor patients treated with pembrolizumab. Nat Cancer. 2020;1(9):873-81.
36. Weiss GJ, Beck J, Braun DP, Bornemann-Kolatzki K, Barilla H, Cubello R, et al. Tumor Cell-Free DNA Copy Number Instability Predicts Therapeutic Response to Immunotherapy. Clin Cancer Res. 2017;23(17):5074-81.
37. Ali HR, Chlon L, Pharoah PD, Markowetz F, Caldas C. Patterns of Immune Infiltration in Breast Cancer and Their Clinical Implications: A Gene-Expression-Based Retrospective Study. PLoS Med. 2016;13(12):e1002194.
38. Yu D, Tong Y, Guo X, Feng L, Jiang Z, Ying S, et al. Diagnostic Value of Concentration of Circulating Cell-Free DNA in Breast Cancer: A Meta-Analysis. Frontiers in Oncology. 2019;9.