

Clinical verification of plasma messenger RNA as novel noninvasive biomarker identified through bioinformatics analysis for lung cancer

Dan Zhou^{1,2}, Weiwei Tang¹, Xinli Liu¹, Han-Xiang An¹ and Yun Zhang^{1,2}

¹Department of Medical Oncology, The First Affiliated Hospital of Xiamen University, Xiamen, Fujian, China

²Xiamen Institute of Rare Earth Materials, Chinese Academy of Sciences and Key Laboratory of Design and Assembly of Functional Nanostructures, Fujian Provincial Key Laboratory of Nanomaterials, Fujian Institute of Research on the Structure of Matter, Chinese Academy of Sciences, Xiamen, Fujian, China

Correspondence to: Yun Zhang, **email:** zhangy@fjirsm.ac.cn
Han-Xiang An, **email:** anhanxiang@xmu.edu.cn

Keywords: lung cancer, biomarker, cell-free RNA, HJURP, ADAMTS8

Received: November 30, 2016

Accepted: March 08, 2017

Published: March 30, 2017

Copyright: Zhou et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Lung cancer (LC) remains associated with significant mortality worldwide. The lack of reliable noninvasive biomarkers and targeted therapies contributes to poor survival rate. Herein, we initially took advantage of the public microarray data from Oncomine database to filter messenger RNAs (mRNAs) as potential biomarkers. Subsequently, clinical validation was applied to identify candidate noninvasive biomarkers in plasma from patients with LC. Through comprehensive analysis of transcriptional expression profiles across 12 studies, top 6 over- and underexpressed mRNAs were generated. Then, a pair of matched plasma samples from LC patient and normal control was detected by RT-PCR, and three genes with positive bands were selected for further validation. Finally, qPCR was conducted to further assess values of the three identified genes. We displayed with high confidence that two cell-free mRNAs (HJURP and ADAMTS8) were expressed at significantly levels compared to normal controls. Receiver-operating characteristic (ROC) curves on the diagnostic efficacy of plasma HJURP and ADAMTS8 mRNAs in LC diagnosis showed that the area under the ROC (AUC) was 0.6960 and 0.6877; sensitivity was 66.0% and 83.7%; specificity was 78.6% and 71.4%, respectively. Combined ROC analyses using these two biomarkers revealed an elevated AUC of 0.75. Furthermore, the higher HJURP level could be associated with early-stage LC while lower ADAMTS8 level could be correlated with non-small cell lung cancer. Collectively, circulating HJURP and ADAMTS8 mRNAs are promising noninvasive biomarkers for LC diagnosis. Our integrative strategy provides new insights into novel noninvasive biomarker identification for other types of cancer.

INTRODUCTION

Lung cancer (LC) is currently ranked as the most common cancer and the leading cause of cancer-related mortality in males worldwide with roughly 1.8 million new cases diagnosed in 2012 (13% of all cancers) [1]. It is well established that LC is a clinically and pathologically heterogeneous disease and has been categorized into two main histological types non-small cell lung

cancer (NSCLC) including squamous cell carcinoma, adenocarcinoma as well as large cell carcinoma and small cell lung cancer based on the origin of epithelial-cell precursors [2]. Risk factors including exposure to environmental and occupational carcinogens, have been associated with an increased incidence of LC [3]. The overall survival rate of LC patients is remarkably improved if diagnosis is confirmed at early stages, but limited access to effective screening with X-ray or computed tomography

remains a problem [4]. Accordingly, identification of novel biomarkers for better diagnosis and for gaining insights into the molecular makeup of LC would be beneficial.

There is a growing interest in cell-free nucleic acids isolated from body fluids as diagnostic indicators for cancer pathology [5]. Circulating cell-free messenger RNA (mRNA) containing tumor-associated genetic alterations were first demonstrated in the 1990s in the plasma of patients with nasopharyngeal carcinoma [6]. A few years later, the investigation was expanded to patients with different type of cancers including breast cancer, colorectal cancer and lung cancer [7–9]. Collectively, these findings provide opportunities for cell-free mRNA to be served as appealing non-invasive biomarker candidates in various cancers including lung cancer. However, until now there have been only limited studies on the identification of cell-free mRNA level and the correlation between the presence of cell-free mRNA and the clinicopathological characteristics of LC patients.

One potential approach to obtain useful candidate biomarkers is the large body of publicly available microarray data [10]. The origin of circulating nucleic acids has been shown to be released from apoptotic and necrotic cancer cells as well as tissues [11]. Thus, we can collect a series of potential biomarkers on the basis of publicly tissue microarray data. In this study, we first took advantage of the large set of existing microarray to analyzed mRNA

expression levels in LC tissues from Oncomine database via a cancer vs. normal analysis yielding a list of candidate mRNAs differentially expressed in LC. This was followed by a clinical validation study with multiple patient plasma samples that ultimately led to confirmation of several cell-free mRNAs as novel noninvasive biomarkers.

RESULTS

Identification of candidate mRNAs from Oncomine database

Using the Oncomine database we compared the mRNA expression level in lung cancer vs. normal according to a flow chart in Figure 1. In total, 892 lung cancer samples including 714 adenocarcinoma samples, 114 squamous cell lung carcinoma samples, 30 large cell lung carcinoma samples, 22 lung carcinoid tumor samples and 12 small cell lung carcinoma samples, along with 285 normal samples from 12 pioneering works on lung cancer were analyzed [12–23]. As shown in Figure 2, the Oncomine data yielded the following top ten overexpressed mRNAs: Top2A, GPT2, GINS2, HJURP, TK1, CDCA5, ZWINT, NUSAP1, LOC100289612 and CCNB1. Similarly, another set of the top ten underexpressed mRNAs resulted from the Oncomine analyses: AGER, FHL1, CLDN18, ADAMTS8, ADH1B, GPIHBP1, CLEC3B, TCF21, CLIC5 and SHROOM4.

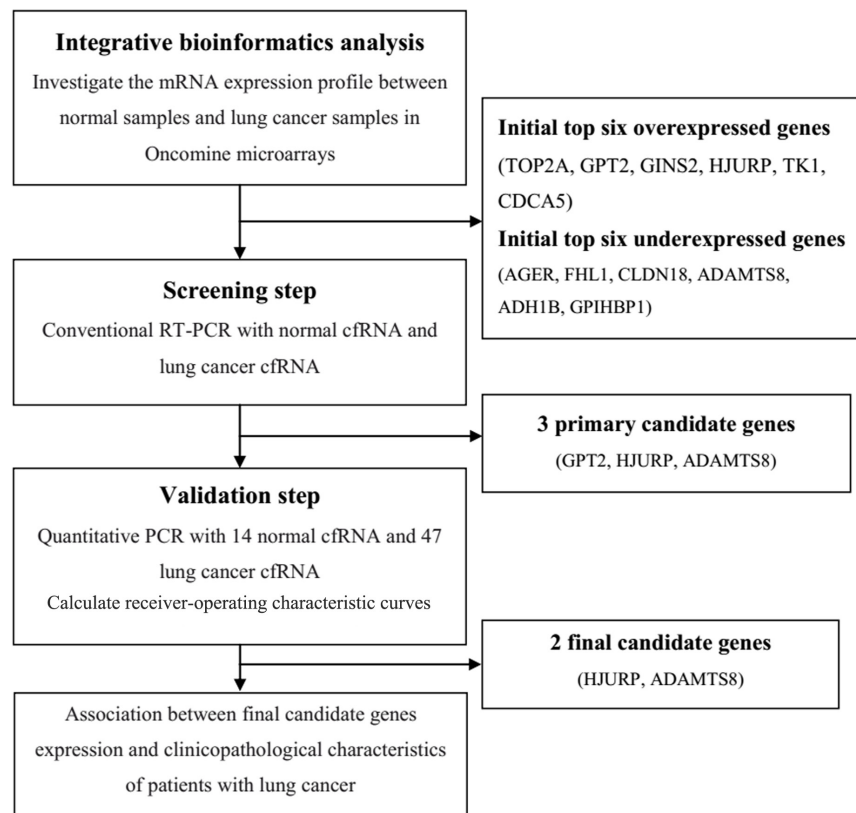


Figure 1: Schematic diagram of the overall procedure.

In general there is concordance between the tumor-associated genetic alterations in plasma and those changes in the primary tumor. However, exceptions to this have also yielded. This implies that experimental validation of candidate biomarkers in cell-free mRNA (cfRNA) derived from analysis of the tissue microarray data using well-defined plasma samples is an essential step for reliably indentifying noninvasive biomarkers.

Experimental validation of the candidate noninvasive biomarkers

As a result of the limited amount of the cfRNA, we selected 12 genes including top six overexpressed and underexpressed mRNAs for experimental validation. The validation procedures contained two procedures. Initially, conventional RT-PCR analysis was carried out for the 12

candidate mRNAs using a pair of matched plasma samples from LC patient and normal control. Three mRNAs were selected according to the existence of RT-PCR bands. Then, qPCR experiments were conducted for the three genes using cfRNA samples from 47 lung cancer patients and 14 healthy subjects.

Conventional RT-PCR results

We employed conventional RT-PCR to examine 12 mRNAs expression including top six overexpressed (TOP2A, GPT2, GINS2, HJURP, TK1, and CDCA5) and underexpressed (AGER, FHL1, CLDN18, ADAMTS8, ADH1B, and GPIHBP1) mRNAs from the Oncomine analyses. As shown in Figure 3, 9 genes (TOP2A, GINS2, TK1, CDCA5, AGER, FHL1 CLDN18, ADH1B, and GPIHBP1) were excluded from further validation efforts

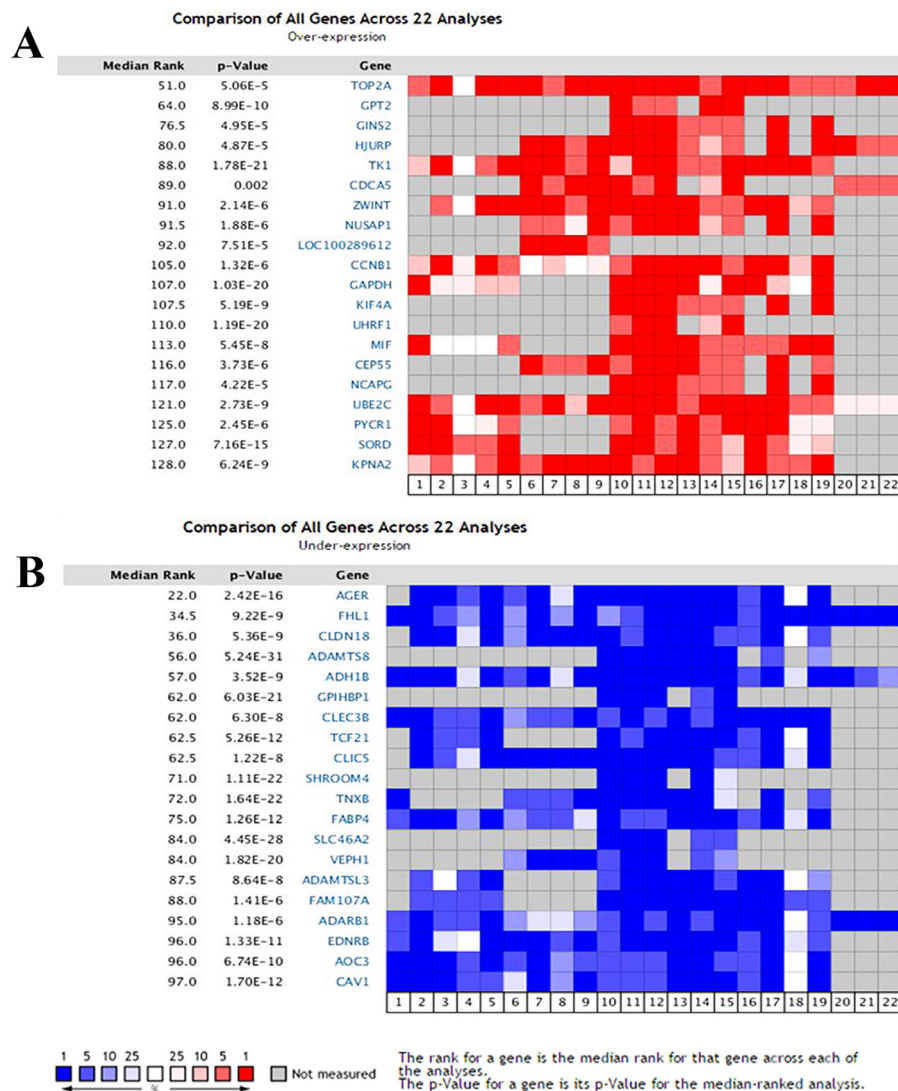


Figure 2: Transcriptional heat map of the top 20 over- and underexpressed genes in lung cancer samples compared with the normal samples through Oncomine analysis. The level plots depict the frequencies (%) of (A) over- and (B) underexpressed candidate genes in 22 analyses from 12 included studies. Red cells represent overexpression. Blue cells represent underexpression. Gray cells represent not measured.

due to the absence of the RT-PCR bands compared with positive control. In contrast, weakly positive bands of GPT2, HJURP and ADAMTS8 were observed in cfRNA samples from the healthy subject and matched LC patient in accordance with positive control. Taken together, these results indicated that GPT2, HJURP and ADAMTS8 might be closely associated with LC.

QPCR results

To further validate transcriptional profiles of GPT2, HJURP and ADAMTS8 in LC, qPCR was conducted

on cfRNA samples from 47 LC patients and 14 healthy subjects.

The expression of HJURP and GPT2 mRNA was detected in all LC plasma samples, whereas the expression of ADAMTS8 mRNA could only be detected in 43/47 LC samples. As shown in Figure 4A, 4B and 4C, the change in the expression of HJURP ($p=0.0275$) and ADAMTS8 ($p=0.0370$) but not GPT2 ($p=0.3959$) was significantly different between the LC cfRNA samples and healthy controls. In addition, the diagnostic efficacy of plasma HJURP and ADAMTS8 mRNAs in LC diagnosis was performed using receiver-operating characteristic (ROC)

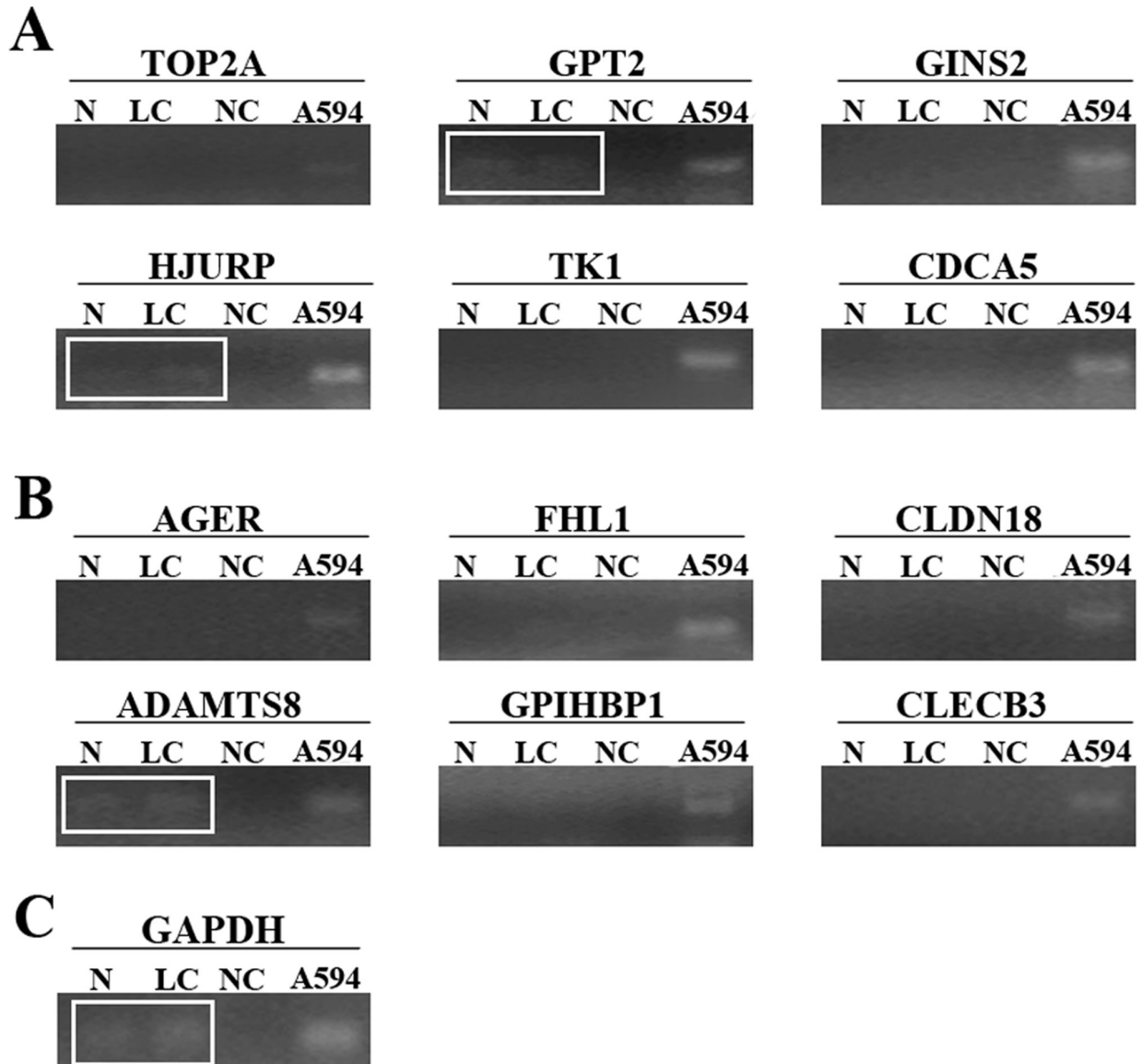


Figure 3: The amplified fragments of RT-PCR for top six overexpressed (A) including TOP2A, GPT2, GINS2, HJURP, TK1, and CDCA5 and underexpressed (B) including AGER, FHL1, CLDN18, ADAMTS8, ADH1B, and GPIHBP1 mRNAs. Weakly positive bands of GPT2, HJURP and ADAMTS8 were identified in cfRNA samples from lung cancer patient and healthy subject after separation in an agarose gel respectively. (C) GAPDH was amplified as the control PCR product. N represents cfRNA from healthy subject. LC represents cfRNA from lung cancer patient. NC represents no template control. A594 represent total RNA from A594 cells as positive control.

curves. At the optimal cut-off (<0.7104), HJURP had a 66.0% sensitivity and a 78.6% specificity in separating LC patients from healthy subjects with an AUC of 0.6960 (Figure 4D). Similarly, as shown in Figure 4E, a cut-off (<0.9482) for ADAMTS8 could be determined with an 83.7% sensitivity and a 71.4% specificity for separating LC patients from healthy subjects (AUC=0.6877). Additionally, the result of using a logistic regression model to combine the data from the two markers showed that the AUC for combined ROC was significantly improved (AUC=0.75) compared with the AUC for HJURP or ADAMTS8 alone. The results suggested that the abnormality of plasma HJURP and ADAMTS8 levels could be useful molecular markers for LC diagnosis. Also, ROC curve analysis further indicated that there was no significant difference of GPT2 expression between LC patients and healthy controls (AUC=0.5760).

Clinicopathologic characteristics of candidate biomarkers

The clinicopathologic characteristics of circulating HJURP and ADAMTS8 were shown in Table 1. Interestingly, the lower expression of HJURP was significantly associated with stage IV LC patients

($p=0.0422$). However, no significant difference was found between HJURP expression and patient sex, age, tumor location, or histology types ($p>0.05$). Specifically, ADAMTS8 expression tended to be remarkably lower in non-small cell lung cancer group than small cell lung cancer group. Similarly, there was no correlation between ADAMTS8 expression and other characteristics such as age, sex, tumor stage, or tumor location ($p>0.05$).

Transcriptional expression level of HJURP and ADAMTS8 mRNA regarding various tumor entities and coexpression genes

To further explore the potential roles of circulating HJURP and ADAMTS8 mRNAs in different tumor entities, expression level of HJURP and ADAMTS8 in 20 common cancers was compared with normal samples using Oncomine database. As shown in Figure 5A, expression of HJURP was inconsistent, being overexpressed in 14 of 20 cancer types, especially colorectal cancer, breast cancer, sarcoma as well as lung cancer and underexpressed in 4 types. However, our qPCR assay showed that HJURP expression level was significantly underexpressed in cfRNA of LC patients compared with that of healthy controls. The inconsistency of HJURP expression trend

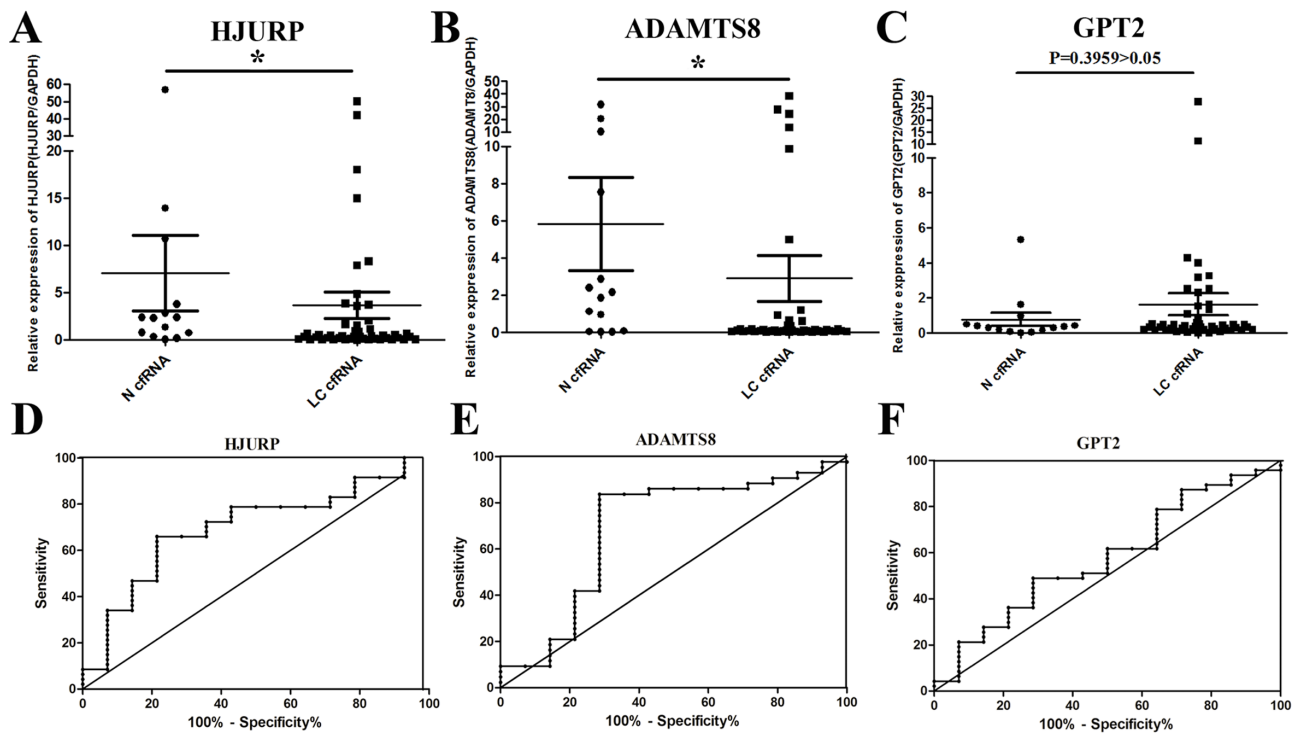


Figure 4: Clinical validation of candidate noninvasive biomarkers. The change in circulating mRNA levels of HJURP (A), ADAMTS8 (B) and GPT2 (C) mRNAs between lung cancer patients and normal subjects detected by qPCR. Data was shown as scatter plots and the intersecting line represents the median value with the interquartile range. Receiver-operating characteristic curves (ROC) showed the performances of fold-change in HJURP (D), ADAMTS8 (E) and GPT2 (F) mRNA expression in predicting the lung cancer. Results were shown with means \pm SEM. * $p<0.05$.

Table 1: Correlations between HJURP and ADAMTS8 expressions in lung cancer cfRNA and clinicopathologic characteristics

| Variables | HJURP | | | ADAMTS8 | | |
|----------------------------|-------|----------------------|---------------------|---------|-----------------------|---------------------|
| | n | Expression status | P | n | Expression status | P |
| Age | | | | | | |
| >60 | 25 | 5.54(0.0117-50.00) | 0.774 | 23 | 5.027(0.0212-38.28) | 0.970 |
| ≤60 | 22 | 0.9423(0.0309-4.828) | | 20 | 0.4856(0.0246-5.003) | |
| Sex | | | | | | |
| Male | 35 | 4.603(0.0117-50) | 0.347 | 32 | 3.857(0.0212-38.28) | 0.309 |
| Female | 12 | 1.014(0.0514-4.828) | | 11 | 0.1725(0.0246-0.6549) | |
| Stage | | | | | | |
| I+II+III | 11 | 4.632(0.1211-17.99) | 0.042 ^{*a} | 9 | 0.2079(0.0212-0.934) | 1 |
| IV | 31 | 3.524(0.0117-50) | | 29 | 2.904(0.0246-38.28) | |
| Tumor location | | | | | | |
| Left lung | 18 | 1.241(0.0117-14.98) | 0.185 | 16 | 0.4144(0.0212-5.003) | 0.286 |
| Right lung | 29 | 5.205(0.0309-50) | | 27 | 4.397(0.0349-38.28) | |
| Histology types | | | | | | |
| Non-small cell lung cancer | 40 | 2.481(0.0117-42.11) | 0.226 | 36 | 1.320(0.0212-27.76) | 0.011 ^{*a} |
| Small cell lung cancer | 5 | 4.058(0.0309-8.319) | | 9 | 7.84(0.1338-24.12) | |

*a: p value<0.05.

between tissue RNA and plasma circulating mRNA may be due to several reasons, which would be discussed next. Meanwhile, we found that ADAMTS8 was downregulated in a variety of cancers including bladder cancer, brain cancer and lung cancer. In addition, the protein-protein interaction networks for HJURP and ADAMTS8 in the most significant modules among different species were shown in Figure 5B and 5C.

DISCUSSION

Lung cancer remains associated with significant mortality worldwide. The lack of specific and sensitive biomarkers for early diagnosis as well as targeted therapies contributes to poor overall survival rate. A blood-based biomarker is a cost- and effective tool because blood is easily accessible in a noninvasive manner and measurements could be repeated over time [24]. Recently, numerous research groups have investigated the potential use of circulating mRNA as biomarker for cancer diagnosis [25, 26]. Identification of biomarkers for multiple cancer types can be carried out in various manners. The general experimental strategy for potential circulating mRNA biomarkers discovery is to employ microarray

technology, which is then followed by validation using RT-PCR. However, a systematic study deciphering the transcriptional expression of the candidate circulating mRNA biomarkers by integrating diverse microarray data is still lacking. Herein, we describe a novel approach to biomarker discovery that take advantage of underused public data. Meanwhile, we believe through this study that the strategy initially developed for lung cancer can be extended to other kinds of cancer. To our knowledge, this is the first systematic and comprehensive study to identify potential circulating mRNA biomarkers.

RNA was commonly thought to be easily degradable, highly labile and rarely abundant. However, numerous recent studies have demonstrated that RNA released into the circulation is surprisingly stable in spite of the point that increased amounts of RNases circulate in the blood of cancer patients and various kinds of RNA, including mRNA and non-coding RNA can be extracted and detected in the circulating plasma, serum and other bodily fluids [27–30]. RNA protection in the blood may be provided by inclusion into lipoproteic complexes or phospholipids [30]. In addition, the inability to obtain primary tumor tissues, particularly through repeat biopsy, from patients with advanced-stage lung cancer makes

the use of plasma as a surrogate fashion for genetic analysis clinically effective. In the present study, we found that the levels of plasmatic HJURP and ADAMTS8 were significantly downregulated in LC patients than in healthy controls, with remarkable diagnostic value for LC. Specifically, HJURP and ADAMTS8 showed statistically significant association with certain clinical aspects of LC. The mechanism accounting for the release of plasma mRNA is not well understood. The most common hypothesis is that they come from apoptosis or necrosis of tumor cells at the tumor site, lysis or apoptosis of circulating tumor cells, or an active release of nucleic acids from the tumor, or even from degenerative normal cells adjacent to tumor cells [31, 32].

HJURP (Holiday Junction Recognizing Protein) located at chromosome 2q37 is a novel tumor suppressor shown to be an independent prognostic of death risk for cancer patients [33]. Recent studies have reported that HJURP is responsible for centromere protein A (CENPA) locationization for loading of new CENPA nucleosomes during the cell cycle and involved in repairing double-strand DNA breaks [34]. HJURP downregulation resulted in a dramatic loss of CENPA from centromeres and induced genomic instability involving chromosome segregation defect during mitosis and then promote tumorigenesis [35]. Several studies have reported that HJURP overexpression was observed in a majority of lung cancer and breast cancer tissue samples [36]. However,

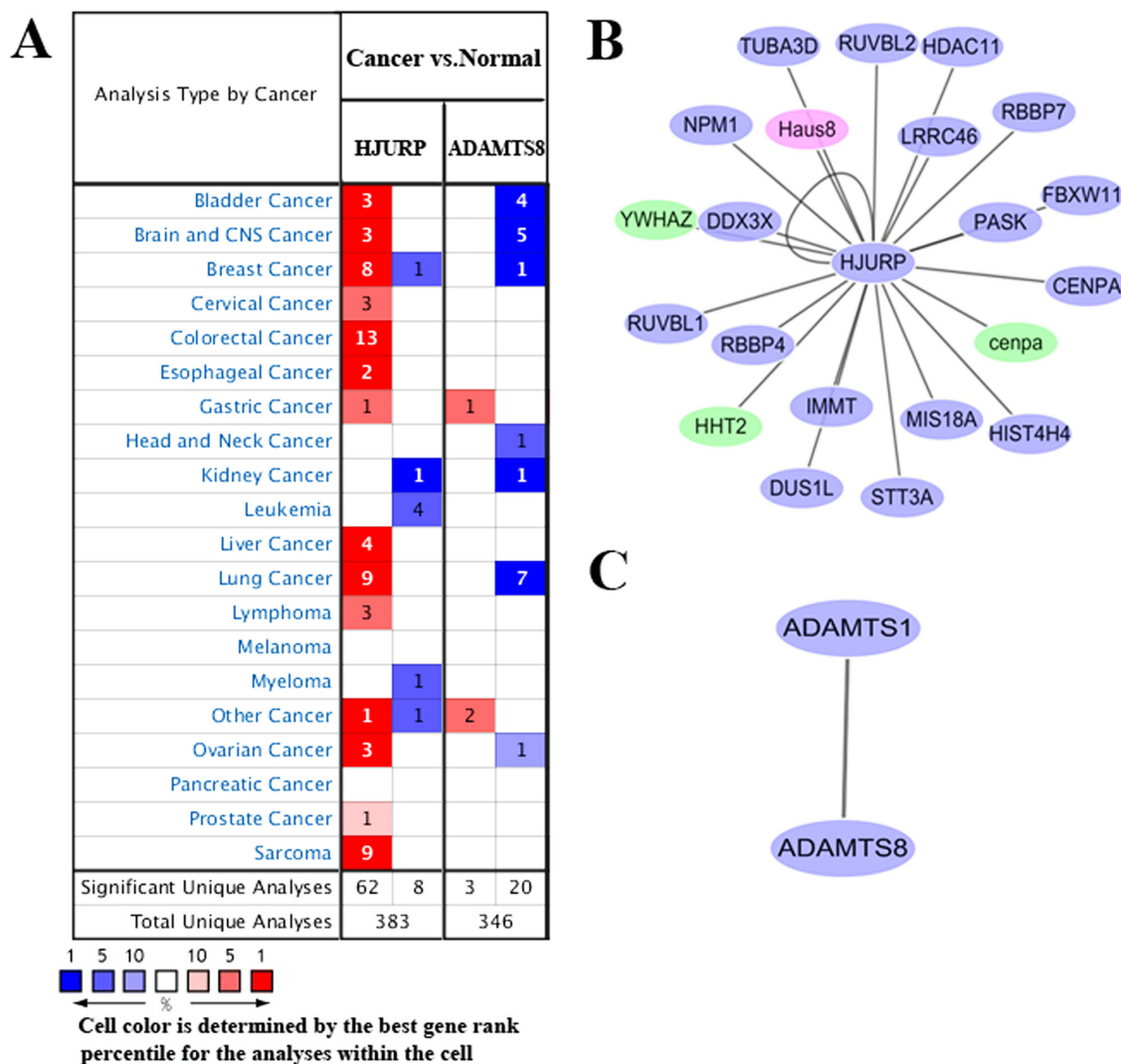


Figure 5: (A) OncoPrint analysis of HJURP and ADAMTS8 gene expression in normal and cancer tissues of different types. Gene summary views for HJURP and ADAMTS8 genes were shown, which indicated the number of significant unique microarray results with changes >2.0 fold ($P < 1E-4$ by two tailed t-test). The bottom rows in each case indicate the total number of unique analyses and the number of unique analyses that show significant overexpression (red) or underexpression (blue) of the target gene in the cancer samples relative to the normal tissue samples. Genes coexpressed with HJURP (B) and ADAMTS8 (C) were analyzed using Cytoscape 3.3.0 from different species. Blue node represents gene from Homo species. Red node represents gene from Mus species. Green node represents gene from Bos taurus species.

Table 2: Clinical characteristics of patients with Lung cancer and healthy controls

| Clinicopathologic factors | Serum study | |
|-----------------------------------|-------------|------------------|
| | LC Cases | Healthy controls |
| Total | 47 | 14 |
| Mean Age±SD | 59.23±9.26 | 43.64±14.91 |
| Sex | | |
| Male | 35 | 9 |
| Female | 12 | 5 |
| Stage | | |
| I | 1 | |
| II | 2 | |
| III | 8 | |
| IV | 31 | |
| NA | 5 | |
| Histology types | | |
| Non-small cell lung cancer | | |
| Adenocarcinoma | 26 | |
| Squamous cell carcinoma | 13 | |
| Large cell carcinoma | 1 | |
| Small cell lung cancer | 5 | |
| NA | 2 | |
| Tumor location | | |
| Left lung | 18 | |
| Right lung | 29 | |

in our study the expression of HJURP was significantly downregulated in cfRNA samples from patients with LC, compared with normal controls. The reasons accounting for the inconsistency could be explained by the differences in the number of patients studied, age, stage (early stage vs. metastatic stage), performance status, molecular characteristics, or other technical variables, including the amount of blood collected, storage, and RT-PCR method (primers, reagents and PCR conditions) [37].

ADAMTS8 located at chromosome 11q24 is a member of the ADAMTS (a disintegrin and metalloprotease with thrombospondin-1-like motifs) family with antiangiogenic property [38]. Unlike conventional ADAM family members, the ADAMTS family members do not have transmembrane regions and are secreted from cells in a catalytically active form. ADAMTS8 is also known to be highly expressed in both adult and fetal lung, and is found at lower levels in the brain, heart, stomach and kidney. Downregulation of ADAMTS8 through epigenetic silencing has been detected

in multiple tumors including brain cancer, breast cancer and non-small cell lung cancer [39].

In conclusion, we identified candidate noninvasive biomarkers for lung cancer through bioinformatic analysis of the public Oncomine database and validated their potential using clinical specimens. In addition, our paper is the first study on plasma levels of HJURP and ADAMTS8 mRNA in LC patients. Nevertheless, sample size is a crucial factor to gain statistically significant results. The clinical samples employed in this study were relatively limited. Hence, large-scale studies should be performed to investigate the potential role of HJURP and ADAMTS8 in LC in the future.

MATERIALS AND METHODS

Ethics statement

All experimental protocols were approved by the Clinical Research Ethics Committee of the First Affiliated

Hospital of Xiamen University. All methods were performed in accordance with the Declaration of Helsinki. Written informed consent was obtained from all human participants after complete description of the study.

Genome-wide expression analyses by oncomine

A bioinformatics flow was established to analyze lung cancer microarrays in the Oncomine database (Figure 1). Oncomine platform is an industry-standard tool aimed at computing gene expression signatures and extracting biological insights from the database [40]. Differential genome-wide expression analysis on the basis of microarray data comparing the most common types of cancer with their respective normal controls as well as various cancer subtypes are ready to be explored [41]. A simple normalization strategy including a log₂ transformation and median centering is employed to all microarray datasets regardless of the platforms, the pre-treatment methods and other variables [42]. For our analysis of cancer vs. normal tissue mRNA, we focused on primary tumors and the following cut-offs were employed p-value ≤ 10⁻⁴ and fold change ≥ 2. Lists of overexpressed and underexpressed mRNAs in lung cancer were available from each study.

Clinical specimens

Blood specimens from 47 primary LC patients and 14 healthy subjects were drawn before therapeutic intervention by venapuncture and processed within 2hr. Plasma was collected from the 8 ml blood specimens after centrifugation at 1,600×g for 10 min and 10,000×g for 10 min both at 4°C and stored at -80°C until processing for RNA extraction. Demographic, clinical and histopathological parameters of all these cases were shown in Table 2.

RNA extraction

RNA was isolated from 2ml plasma using TRIzol LS reagent (cat#10296018, Thermo Fisher Scientific Inc.) according to manufacturer's instructions as previously described [43, 44]. In brief, 2 ml plasma was separated into eight aliquots (250µl), and each aliquot was mixed with 750µl of TRIzol. After 5min incubation at 4°C, 400µl chloroform was added, followed by 30 sec of violent shaking. The mixture was immediately centrifuged at 12,000g for 5min at 4°C. The above aqueous layer was transferred in to a fresh tube containing 800µl isopropyl alcohol. The RNA in aqueous layer was precipitate for 12-16hr at -20°C and washed with 1 ml 75% ethanol. Lastly, the RNA pellet was dried for 3-5min at room temperature, dissolved in 10µl Rnase-free water and then the all aliquots were pooled together. The RNA concentration was quantitated using Qubit RNA HS Assay Kit and Qubit 3.0 fluorometer (ThermoFisher Scientific Inc. cat#Q32851).

The concentration of RNA isolated from plasma ranged from 2.12 to 12.26ng/µl.

Conventional RT-PCR analysis

The extracted total RNA was reverse-transcribed into cDNA using PrimeScript RT reagent Kit (TAKARA cat#RR047A) according to the manufacturer's protocol in triplicates. The resulting cDNA was pooled for next PCR amplification. Primer sequences are shown in Supplementary Table 1. The target sequences were amplified by PCR in 50µl of 1×Taq buffer containing 0.3µM of each primer, 1.5 mM magnesium chloride, 200µM dNTP mixture, 2.5 units of Taq polymerase and 1µl (10µg) of each cDNA by Ex Taq PCR kit (TAKARA cat#RR001A). The reaction was started after 5min denaturation of cDNA at 94°C (hot start). DNA amplification in a T100 cycler (Bio-Rad Laboratories, Inc) was followed by a final extension for 8min at 72°C for HJURP, ADAMTS8 and GPT2 (94°C 30s, 58.5°C 1min; 40 cycles). Although a range of annealing temperatures (from 5°C below the T_m to 5°C above the T_m) had been tried, the PCR products of other 9 mRNAs (TOP2A, GINS2, TK1, CDCA5, AGER, FHL1 CLDN18, ADH1B, and GPIHBP1) were rare compared with positive control. GAPDH expression was used as an internal control. All PCR products were visualized by 2% agarose gels electrophoresis. Water negative controls contained all components for the RT-PCR reaction without target RNA. Positive controls of RNA were extracted from A594 cells obtained from the Cancer Center of Xiamen University (Xiamen, China).

Quantitative PCR (qPCR)

For further validation, qPCR was carried out in duplicate at 50°C for 2min, denaturing at 95°C for 5min, followed by 40 cycles of 95°C for 30s, 58.5°C for 1min by using ABI ViiA 7 Real-Time PCR System (Applied Biosystems) with melting curve analysis. GAPDH was employed as a reference gene applying the 2^{-ΔCT} algorithm (ΔCT=Ct. target-Ct. reference) [45].

Statistical analysis

Kolmogorov–Smirnov test was applied to determine the distribution of the samples of each group. Data were showed as median and range. Expression levels of plasma cell-free mRNAs between two groups were compared using the Mann–Whitney U test. The ratios of qPCR expression values of normal and cancer samples were used after normalization by the expression value of internal control GAPDH gene. The association between relevant gene expression levels and clinical parameters was addressed by Mann–Whitney U test. The diagnostic performance of HJURP and ADAMTS8 was identified using receiver operating characteristic (ROC) curves

and the area under the curve (AUC). Cut-offs for the two biomarkers were estimated at various sensitivities and specificities and at the maximum Youden's index (sensitivity+specificity-1) [46]. Moreover, the two markers were multiplied by their logistic regression coefficient and added to give a combined AUC value. Statistical analyses were carried out using GraphPad Prism 7.0 (GraphPad Soft-ware Inc., La Jolla, CA, USA) and MedScale software. Summary data were reported as mean±SEM. A p value of less than 0.05 was considered statistically significant.

Author contributions

Z.D. designed the experimental procedure, finished the clinical validation and wrote the initial manuscript. T.W.W. carried out the statistical analysis and revised the figures. L.X.L. helped to carry out statistical analysis. A.H.X helped to prepare the figures and revised the manuscript. Z.Y. participated in the design of this study.

CONFLICTS OF INTEREST

The authors declare there is no potential conflicts of interest.

FUNDING

This study was supported by the National Natural Science Foundation of China (No. 31400699), the Natural Science Foundation of Fujian Province of China under grant (No. 2014J01142 and NO.2017Y0084), the Fujian Provincial Health and Family Planning Commission Foundation of Youth scientific research project (No. 2015-2-42) and Xiamen Science and Technology Bureau Foundation of science and technology project for the benefit of the people (Grant No. 3502Z20164010).

REFERENCES

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015; 136: E359-86. doi: 10.1002/ijc.29210.
2. Devesa SS, Bray F, Vizcaino AP, Parkin DM. International lung cancer trends by histologic type: male:female differences diminishing and adenocarcinoma rates rising. *Int J Cancer*. 2005; 117: 294-9. doi: 10.1002/ijc.21183.
3. Swiatkowska B, Szubert Z, Sobala W, Szeszenia-Dabrowska N. Predictors of lung cancer among former asbestos-exposed workers. *Lung Cancer*. 2015; 89: 243-8. doi: 10.1016/j.lungcan.2015.06.013.
4. Infante M, Lutman FR, Cavuto S, Brambilla G, Chiesa G, Passera E, Angeli E, Chiarenza M, Aranzulla G, Cariboni

- U, Alloisio M, Incarbone M, Testori A, et al. Lung cancer screening with spiral CT: baseline results of the randomized DANTE trial. *Lung Cancer*. 2008; 59: 355-63. doi: 10.1016/j.lungcan.2007.08.040.
5. Pinzani P, Salvianti F, Pazzagli M, Orlando C. Circulating nucleic acids in cancer and pregnancy. *Methods*. 2010; 50: 302-7. doi: 10.1016/j.ymeth.2010.02.004.
6. Lo KW, Lo YM, Leung SF, Tsang YS, Chan LY, Johnson PJ, Hjelm NM, Lee JC, Huang DP. Analysis of cell-free Epstein-Barr virus associated RNA in the plasma of patients with nasopharyngeal carcinoma. *Clin Chem*. 1999; 45: 1292-4.
7. Wong SC, Lo SF, Cheung MT, Ng KO, Tse CW, Lai BS, Lee KC, Lo YM. Quantification of plasma beta-catenin mRNA in colorectal cancer and adenoma patients. *Clin Cancer Res*. 2004; 10: 1613-7.
8. Chen XQ, Bonnefoi H, Pelte MF, Lyautey J, Lederrey C, Movarekhi S, Schaeffer P, Mulcahy HE, Meyer P, Stroun M, Anker P. Telomerase RNA as a detection marker in the serum of breast cancer patients. *Clin Cancer Res*. 2000; 6: 3823-6.
9. Kopreski MS, Benko FA, Gocke CD. Circulating RNA as a tumor marker: detection of 5T4 mRNA in breast and lung cancer patient serum. *Ann N Y Acad Sci*. 2001; 945: 172-8.
10. Petty RD, Nicolson MC, Kerr KM, Collie-Duguid E, Murray GI. Gene expression profiling in non-small cell lung cancer: from molecular mechanisms to clinical application. *Clin Cancer Res*. 2004; 10: 3237-48. doi: 10.1158/1078-0432.CCR-03-0503.
11. Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, Hesch RD, Knippers R. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res*. 2001; 61: 1659-65.
12. Beer DG, Kardia SL, Huang CC, Giordano TJ, Levin AM, Misek DE, Lin L, Chen G, Gharib TG, Thomas DG, Lizyness ML, Kuick R, Hayasaka S, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med*. 2002; 8: 816-24. doi: 10.1038/nm733.
13. Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M, Loda M, Weber G, Mark EJ, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U S A*. 2001; 98: 13790-5. doi: 10.1073/pnas.191502998.
14. Garber ME, Troyanskaya OG, Schluens K, Petersen S, Thaesler Z, Pacyna-Gengelbach M, van de Rijn M, Rosen GD, Perou CM, Whyte RI, Altman RB, Brown PO, Botstein D, et al. Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci U S A*. 2001; 98: 13784-9. doi: 10.1073/pnas.241500798.
15. Hou J, Aerts J, den Hamer B, van Ijcken W, den Bakker M, Riegman P, van der Leest C, van der Spek P, Foekens JA,

- Hoogsteden HC, Grosveld F, Philipsen S. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. *PLoS One*. 2010; 5: e10312. doi: 10.1371/journal.pone.0010312.
16. Landi MT, Dracheva T, Rotunno M, Figueroa JD, Liu H, Dasgupta A, Mann FE, Fukuoka J, Hames M, Bergen AW, Murphy SE, Yang P, Pesatori AC, et al. Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLoS One*. 2008; 3: e1651. doi: 10.1371/journal.pone.0001651.
 17. Okayama H, Kohno T, Ishii Y, Shimada Y, Shiraiishi K, Iwakawa R, Furuta K, Tsuta K, Shibata T, Yamamoto S, Watanabe S, Sakamoto H, Kumamoto K, et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res*. 2012; 72: 100-11. doi: 10.1158/0008-5472.CAN-11-1403.
 18. Selamat SA, Chung BS, Girard L, Zhang W, Zhang Y, Campan M, Siegmund KD, Koss MN, Hagen JA, Lam WL, Lam S, Gazdar AF, Laird-Offringa IA. Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression. *Genome Res*. 2012; 22: 1197-211. doi: 10.1101/gr.132662.111.
 19. Stearman RS, Dwyer-Nield L, Zerbe L, Blaine SA, Chan Z, Bunn PA Jr, Johnson GL, Hirsch FR, Merrick DT, Franklin WA, Baron AE, Keith RL, Nemenoff RA, et al. Analysis of orthologous gene expression between human pulmonary adenocarcinoma and a carcinogen-induced murine model. *Am J Pathol*. 2005; 167: 1763-75. doi: 10.1016/S0002-9440(10)61257-6.
 20. Su LJ, Chang CW, Wu YC, Chen KC, Lin CJ, Liang SC, Lin CH, Whang-Peng J, Hsu SL, Chen CH, Huang CY. Selection of DDX5 as a novel internal control for Q-RT-PCR from microarray data using a block bootstrap re-sampling scheme. *BMC Genomics*. 2007; 8: 140. doi: 10.1186/1471-2164-8-140.
 21. Talbot SG, Estilo C, Maghami E, Sarkaria IS, Pham DK, P Oc, Socci ND, Ngai I, Carlson D, Ghossein R, Viale A, Park BJ, Rusch VW, et al. Gene expression profiling allows distinction between primary and metastatic squamous cell carcinomas in the lung. *Cancer Res*. 2005; 65: 3063-71. doi: 10.1158/0008-5472.CAN-04-1985.
 22. Wachi S, Yoneda K, Wu R. Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. *Bioinformatics*. 2005; 21: 4205-8. doi: 10.1093/bioinformatics/bti688.
 23. Yamagata N, Shyr Y, Yanagisawa K, Edgerton M, Dang TP, Gonzalez A, Nadaf S, Larsen P, Roberts JR, Nesbitt JC, Jensen R, Levy S, Moore JH, et al. A training-testing approach to the molecular classification of resected non-small cell lung cancer. *Clin Cancer Res*. 2003; 9: 4695-704.
 24. Patz EF Jr, Campa MJ, Gottlin EB, Kusmartseva I, Guan XR, Herndon JE 2nd. Panel of serum biomarkers for the diagnosis of lung cancer. *J Clin Oncol*. 2007; 25: 5578-83. doi: 10.1200/JCO.2007.13.5392.
 25. Wang H, Zhang X, Wang L, Zheng G, Du L, Yang Y, Dong Z, Liu Y, Qu A, Wang C. Investigation of cell free BIRC5 mRNA as a serum diagnostic and prognostic biomarker for colorectal cancer. *J Surg Oncol*. 2014; 109: 574-9. doi: 10.1002/jso.23526.
 26. March-Villalba JA, Martinez-Jabaloyas JM, Herrero MJ, Santamaria J, Alino SF, Dasi F. Cell-free circulating plasma hTERT mRNA is a useful marker for prostate cancer diagnosis and is associated with poor prognosis tumor characteristics. *PLoS One*. 2012; 7: e43470. doi: 10.1371/journal.pone.0043470.
 27. Schwarzenbach H, Hoon DS, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. *Nat Rev Cancer*. 2011; 11: 426-37. doi: 10.1038/nrc3066.
 28. Tsui NB, Ng EK, Lo YM. Stability of endogenous and added RNA in blood specimens, serum, and plasma. *Clin Chem*. 2002; 48: 1647-53.
 29. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, Peterson A, Noteboom J, O'Briant KC, Allen A, Lin DW, Urban N, Drescher CW, et al. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A*. 2008; 105: 10513-8. doi: 10.1073/pnas.0804549105.
 30. El-Hefnawy T, Raja S, Kelly L, Bigbee WL, Kirkwood JM, Luketich JD, Godfrey TE. Characterization of amplifiable, circulating RNA in plasma and its potential as a tool for cancer diagnostics. *Clin Chem*. 2004; 50: 564-73. doi: 10.1373/clinchem.2003.028506.
 31. Garcia JM, Garcia V, Pena C, Dominguez G, Silva J, Diaz R, Espinosa P, Citores MJ, Collado M, Bonilla F. Extracellular plasma RNA from colon cancer patients is confined in a vesicle-like structure and is mRNA-enriched. *RNA*. 2008; 14: 1424-32. doi: 10.1261/rna.755908.
 32. Valadi H, Ekstrom K, Bossios A, Sjostrand M, Lee JJ, Lotvall JO. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat Cell Biol*. 2007; 9: 654-9. doi: 10.1038/ncb1596.
 33. Huang W, Zhang H, Hao Y, Xu X, Zhai Y, Wang S, Li Y, Ma F, Li Y, Wang Z, Zhang Y, Zhang X, Liang R, et al. A Non-Synonymous Single Nucleotide Polymorphism in the HJURP Gene Associated with Susceptibility to Hepatocellular Carcinoma among Chinese. *PLoS One*. 2016; 11: e0148618. doi: 10.1371/journal.pone.0148618.
 34. Sanchez-Pulido L, Pidoux AL, Ponting CP, Allshire RC. Common ancestry of the CENP-A chaperones Scm3 and HJURP. *Cell*. 2009; 137: 1173-4. doi: 10.1016/j.cell.2009.06.010.
 35. Dunleavy EM, Roche D, Tagami H, Lacoste N, Ray-Gallet D, Nakamura Y, Daigo Y, Nakatani Y, Almouzni-Pettinotti G. HJURP is a cell-cycle-dependent maintenance and

- deposition factor of CENP-A at centromeres. *Cell*. 2009; 137: 485-97. doi: 10.1016/j.cell.2009.02.040.
36. Kato T, Sato N, Hayama S, Yamabuki T, Ito T, Miyamoto M, Kondo S, Nakamura Y, Daigo Y. Activation of Holliday junction recognizing protein involved in the chromosomal stability and immortality of cancer cells. *Cancer Res*. 2007; 67: 8544-53. doi: 10.1158/0008-5472.CAN-07-1307.
 37. Klein CA, Blankenstein TJ, Schmidt-Kittler O, Petronio M, Polzer B, Stoecklein NH, Riethmuller G. Genetic heterogeneity of single disseminated tumour cells in minimal residual cancer. *Lancet*. 2002; 360: 683-9. doi: 10.1016/S0140-6736(02)09838-0.
 38. Dunn JR, Panutsopoulos D, Shaw MW, Heighway J, Dormer R, Salmo EN, Watson SG, Field JK, Liloglou T. METH-2 silencing and promoter hypermethylation in NSCLC. *Br J Cancer*. 2004; 91: 1149-54. doi: 10.1038/sj.bjc.6602107.
 39. Porter S, Scott SD, Sassoon EM, Williams MR, Jones JL, Girling AC, Ball RY, Edwards DR. Dysregulated expression of adamalysin-thrombospondin genes in human breast carcinoma. *Clin Cancer Res*. 2004; 10: 2429-40.
 40. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, Barrette TR, Anstet MJ, Kincead-Beal C, Kulkarni P, Varambally S, Ghosh D, Chinnaiyan AM. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia*. 2007; 9: 166-80.
 41. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia*. 2004; 6: 1-6.
 42. Syring I, Klumper N, Offermann A, Braun M, Deng M, Boehm D, Queisser A, von Massenhausen A, Bragelmann J, Vogel W, Schmidt D, Majores M, Schindler A, et al. Comprehensive analysis of the transcriptional profile of the Mediator complex across human cancer types. *Oncotarget*. 2016; 7: 23043-55. doi: 10.18632/oncotarget.8469.
 43. Zhang X, Wang C, Wang L, Du L, Wang S, Zheng G, Li W, Zhuang X, Zhang X, Dong Z. Detection of circulating Bmi-1 mRNA in plasma and its potential diagnostic and prognostic value for uterine cervical cancer. *Int J Cancer*. 2012; 131: 165-72. doi: 10.1002/ijc.26360.
 44. Pucciarelli S, Rampazzo E, Briarava M, Maretto I, Agostini M, Digito M, Keppel S, Friso ML, Lonardi S, De Paoli A, Mescoli C, Nitti D, De Rossi A. Telomere-specific reverse transcriptase (hTERT) and cell-free RNA in plasma as predictors of pathologic tumor response in rectal cancer patients receiving neoadjuvant chemoradiotherapy. *Ann Surg Oncol*. 2012; 19: 3089-96. doi: 10.1245/s10434-012-2272-z.
 45. Maru DM, Singh RR, Hannah C, Albarracin CT, Li YX, Abraham R, Romans AM, Yao H, Luthra MG, Anandasabapathy S, Swisher SG, Hofstetter WL, Rashid A, et al. MicroRNA-196a is a potential marker of progression during Barrett's metaplasia-dysplasia-invasive adenocarcinoma sequence in esophagus. *Am J Pathol*. 2009; 174: 1940-8. doi: 10.2353/ajpath.2009.080718.
 46. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950; 3: 32-5.