scientific reports

Check for updates

OPEN

A novel algorithm to differentiate between primary lung tumors and distant liver metastasis in lung cancers using an exosome based multi gene biomarker panel

Kanisha A. Shah[1] & Rakesh M. Rawal[2]✉

The lack of non-invasive methods for detection of early metastasis is a crucial reason for the poor prognosis of lung cancer (LC) liver metastasis (LM) patients. In this study, the goal was to identify circulating biomarkers based on a biomarker model for the early diagnosis and monitoring of patients with LCLM. An 8-gene panel identified in our previous study was validated in CTC, cfRNA and exosomes isolated from primary lung cancer with & without metastasis. Further multivariate analysis including PCA & ROC was performed to determine the sensitivity and specificity of the biomarker panel. Model validation cohort ($n = 79$) was used to verify the stability of the constructed predictive model. Further, clinic-pathological factors, survival analysis and immune infiltration correlations were also performed. In comparison to our previous tissue data, exosomes demonstrated a good discriminative value with an AUC of 0.7247, specificity (72.48%) and sensitivity (96.87%) for the 8-gene panel. Further individual gene patterns led us to a 5- gene panel that showed an AUC of 0.9488 ($p = < 0.001$) and 0.9924 ($p = < 0.001$) respectively for tissue and exosomes. Additionally, on validating the model in a larger cohort a risk score was obtained (RS > 0.2) for prediction of liver metastasis with an accuracy of 95%. Survival analysis and immune filtration markers suggested that four exosomal markers were independently associated with poor overall survival. We report a novel blood-based exosomal biomarker panel for early diagnosis, monitoring of therapeutic response, and prognostic evaluation of patients with LCLM.

Keywords Lung cancer, Liver metastasis, Exosomes, Predictive biomarkers, Liquid biopsy, Non-invasive, Disease progression, Risk score, Early diagnosis, Prognostic factor

Lung cancer is one of the most common tumors worldwide with high morbidity and mortality[1]. Non-small-cell lung carcinoma (NSCLC), is the most prevalent subtype and patients with advanced NSCLC die within 18 months of diagnosis, with an overall five-year survival rate of only 15%. Lung cancer metastasis is the primary cause of death of the patients with liver being the most common metastatic organ for NSCLC[2]. Since there is a long incubation period from primary tumor to diagnosis of liver metastasis, many patients develop advanced liver metastasis at diagnosis[3]. Moreover, these patients frequently develop metastasis to the other organs and lead to shorter survival and high mortality in patients with NSCLC liver metastasis[4,5].

Histopathological evaluation using tissue biopsy from the site of the tumor along with radiological investigations are currently available diagnostic methods for lung cancer liver metastasis which is an invasive procedure. Some of the major limitations of invasive tissue biopsies are inadequacy of tissue, difficulty in accessing deep sites, and potential risk for the patient[6]. These currently available multistep, extended and invasive procedures are not unsuitable as a screening tool. Therefore, it is crucial to identify non-invasive, sensitive and specific circulatory biomarkers that will be extremely useful for screening and early detection of the disease[7]. One of the promising approach to identify the potential biomarkers is to analyse the cancer-related biomolecules in bodily fluids. In

[1]Division of Biological and Life Science, Ahmedabad University, Ahmedabad, Gujarat, India. [2]Department of Life Science, School of Sciences, Gujarat University, Navrangpura, Ahmedabad, Gujarat 380009, India. ✉email: rakeshmrawal@gmail.com; rakeshrawal.gu@gmail.com; rakeshrawal@gujaratuniversity.ac.in

addition, the non-invasive procedure for collection of blood makes biomarkers ideal as a screening tool for liver metastasis in lung cancer patients.

To identify novel metastasis-associated genes, our previous study detected differentially expressed profile of a multi gene panel specific for liver metastasis in primary lung tumor. The gene expression data acquired by real time PCR was further subjected to multivariate analysis like PCA and LDA which ultimately led us to the generation of a specific model[8]. We found that out of the 5 models generated model 4 holds impressive potential for prediction of liver metastasis in primary lung cancer patients. Further, we also studied the individual sensitivity, specificity and cut-off values of each gene that were implicated in the model.

In this study, we validated 8 gene panel, in CTC, cfRNA and exosomes using real time PCR in primary lung cancer patients with & without metastasis and healthy controls, identified in our previous study in the tissue samples of the same cohort. ROC curve analysis was used to determine the sensitivity and specificity of the shortlisted gene panel and its correlation with the most accurate source of liquid biopsy component for early prediction of liver metastasis. Further DEP expression, clinic-pathological factors, PCA analysis and survival analysis correlations were performed. Using this model absolute quantification of candidate markers was done in exosomes from blood of primary lung cancer patients (n = 79) resulting in the identification of potential biomarker panel with high sensitivity and specificity that could predict liver metastasis in high risk cohort (Fig. 1).

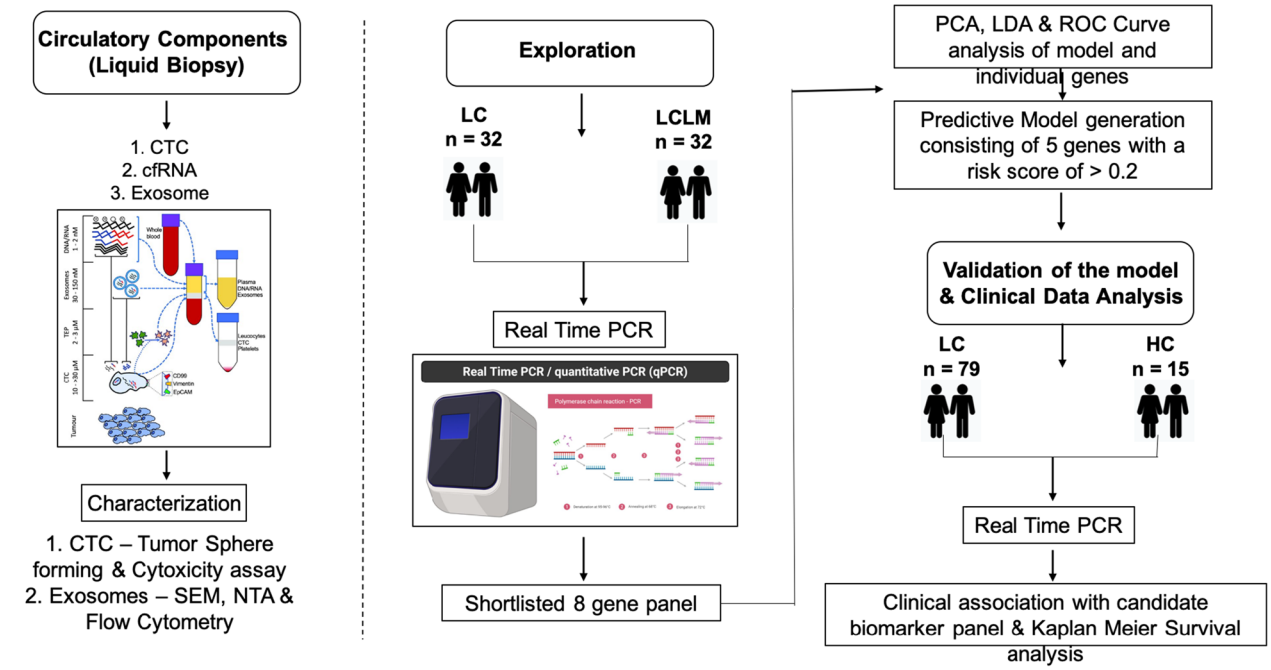
Materials and methods

Patient characteristics

A total of 156 specimens from patients that passed the inclusion/exclusion criteria for the test and validation set were recruited in the study from December 2016 to July 2017. Peripheral blood (9 ml) was collected from patients with primary NSCLC with liver metastasis (n = 32), primary NSCLC without liver metastasis (n = 30) and healthy individuals (n = 15) for the test cohort. For the validation cohort, peripheral blood (4 ml) was collected from primary NSCLC patients (n = 79). Patients with only liver metastasis from primary lung cancer were included in the metastatic cohort and for the primary cohort only those patients with NSCLC subtype were included. Patients with multiple metastasis, HIV/HBsAg positive other malignant tumors or incomplete clinical data were excluded from the study. All the patients recruited in the test and the validation cohort in the study provided written consents for their sample procurement and use of them for research purpose. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional ethics committee, and were approved and reviewed by The Gujarat University ethics committee (GU/IEC/07/2018)] and the The Gujarat Cancer Research Institute /Gujarat Cancer Society Internal Ethics Review board (EC-O-92-2014) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Culture of peripheral blood mononuclear cells for isolation of CTCs

Peripheral blood mononuclear cells (PBMNCs) were isolated by Ficoll-Histopaque density gradient centrifugation. The isolated PBMCs were cultured in RPMI cell culture media with 10% FBS (Fetal Bovine Serum) and 100 U/ml Penicillin—Streptomycin solution in low attachment conditions, at 5% CO2 incubator at 37 °C. After 24 h, the adherent cells were further expanded ex-vivo and were selected for subculture by trypsinization with 0.25%



CTC - Circulating Tumor Cells, cfRNA – Cell free RNA, SEM – Scanning Electron Microscopy, NTA – Nanoparticle Tracking Analysis, LC – Lung Cancer, LCLM – Lung Cancer Liver Metastasis, PCA – Principal Component analysis, LDA – Linear Discriminant analysis, ROC – Receiver Operating Curve, HC – Healthy Controls

Figure 1. Workflow incorporated for designing the entire study.

trypsin and 0.02% EDTA (Life Technologies Inc.). The culture medium was changed twice a week, and cellular homogeneity evaluated microscopically using a phase contrast microscope (Nikon Eclipse TS100) every 24–48 h for a period of 30 days. Adherent cells that appeared CTC-like were counted manually using a microscope to calculate the number of cells present in each cultured sample. The experiments were performed in triplicates.

Tumor sphere formation and cytotoxicity assay

Cells were plated to 6 well ultra-low attachment plates (Corning; New York, USA) at a density of 5×10^3 cells/well and further expanded with RPMI-1640 supplemented growth media. Wells containing spheres were counted manually under inverted phase contrast microscope and the percentage of cells exhibiting sphere-forming capacity was calculated by dividing the number of spheres by the number of cells seeded per well. For the cytotoxicity assay the isolated CTCs were seeded at a density of 1×10^4 cells/well in 96-well plate and exposed to Cisplatin and Carboplatin drugs for 24 h in increased concentrations ranging from 0.25 to 1.0 µg/ml. Further, 5 mg/ml MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide) (Hi-Media, India) and 100 µl of DMSO were added to each well for formazon crystal formation and the cells were incubated at 37 °C. Absorbance was taken at 590 nm with a reference filter of 620 nm using an ELISA reader (Multiskan Spectrum Microplate Reader, Thermo Scientific). All data were normalized against DMSO blank controls.

Flow cytometry

The purity of cultured CTC subpopulation was validated by gating the sorted cells with conjugated primary antibodies of CD44-FITC (Stem Cell Technologies), CD24-PE, CK-FITC and CD45-PE (BD Biosystems, USA) using flow cytometry analysis and for characterization of exosomes, exosome-coated beads (Stem Cell Technologies; USA), were stained with primary mouse monoclonal antibodies directed against CD63, CD81 and CD9. The percentage of each population was calculated by acquiring the sample in FACS Canto II instrument and result was analyzed using FACS Diva software (BD Biosystems, USA) (https://research.pasteur.fr/en/software/bd-facs-diva-software/).

Exosome isolation and characterization

Exosomes were isolated from 1 ml of serum using a Total Exosome Isolation Kit (Invitogen, Thermo Fisher Scientific Inc., USA), according to the manufacturers protocol. Exosomal pellets were split and resuspended (i) in 300 µl PBS for characterization, (ii) in 500 µl PBS for RNA isolation and real time PCR and the remaining exosomes were stored at – 80 °C. The purified exosomes were fixed with 2% (w/v) para-formaldehyde (PFA) in PBS (pH 7.4); dropped onto a carbon-coated grid and left to dry at room temperature. Further they were fixed with 1% (w/v) glutaraldehyde and stained with saturated aqueous uranyl oxalate. Exosomes were then re-suspended in a solution of 0.4% (w/v) uranyl acetate & 1.8% (w/v) methylcellulose, incubated on ice, excess liquid was drained off and the grid was air dried. Exosomes were viewed using a Field Emission Scanning Electron microscope (JSM-7600F; JEOL, Tokyo, Japan)[9]. Nanoparticle Tracking Analysis (NTA) was performed using a Nanosight LM10 and NTA 2.3 Software (NanoSight, Wiltshire, UK) (https://www.malvernpanalytical.com/en/support/product-support/software/nanosight-nta-software-update-v3-2). EV samples were resuspended in filtered PBS. Three 60-s videos were recorded of each sample with camera level and detection threshold set at 10. Temperature was monitored throughout the measurements.

Quantitative real time PCR (qRT-PCR)

Total RNA from CTCs was isolated using RNeasy kit (Qiagen 74106) following manufacturer's instructions whereas TRIzol reagent (Invitrogen, USA) was used for total RNA isolation from exosomes and cell free RNA from serum of patients and healthy individuals. 1 µg total RNA was reverse transcribed using the cDNA archive kit (Applied Biosystems—ABI, USA) following manufacturer's instructions and the resulting cDNA was used as a template for Real Time PCR analysis. qRT-PCR was performed using AriaMX (Agilent Technologies, USA) with a SYBR Green PCR Master Mix (Primer sequences are provided in Supp Table 1). The generation of a specific PCR product was tested using the ΔΔCT method with β-actin as the endogenous control and average C_T value was calculated for the quantification fold change analysis.

Principal component analysis

To obtain a reliable model for prediction of liver metastasis, data was analyzed statistically by applying Principal Component Analysis (PCA) method. Data dimensionality was reduced by using an orthogonal transformation to convert correlated variables into uncorrelated variables, which are termed principal components. PCA score was obtained using the following formulae:

$$\text{PCA score} = C1V1 + C2V2 + C3V3 + C4V4 + \cdots + CnVn$$

where C1, C2, C3 … Cn are coefficients of each of the variables and V1, V2, V3 … Vn are the values of original variables.

The accuracy, specificity and sensitivity of the PCA based model was securitized to check for the robustness of the model. For this, patient stratification process was simulated 100 times by resampling. The patients were divided into primary lung cancer set (30 patients, 48%) and liver metastatic test set (32 patients, 52%) groups. PCA was applied to the data sets to determine the coefficient of each variable, and a PCA score was generated for each patient. The data sets were then stratified into groups based on the mean value of the PCA scores. Statistical analyses and the PCA algorithm were performed using SPSS 20.00 software (https://www.ibm.com/support/

pages/downloading-ibm-spss-statistics-20) and R (https://cran.r-project.org/bin/macosx/) and a p-value of < 0.05 was considered statistically significant.

Linear discriminant analysis and receiver operating characteristics curve analysis

Linear Discriminant Analysis (LDA) was used to determine whether mRNA expression patterns could accurately discriminate liver metastasis in an independent data set. The accuracy of the predicted model was calculated using 1000 repetitions of a random partitioning process to regulate the number and proportion of false discoveries[10]. For diagnostic accuracy and discriminating metastatic tumor from primary tumor, a held-out test set from each database was utilized to evaluate the performance of each of the different classifiers. Receiver operator characteristics curves (ROC) were generated and AUCs of each classifier were calculated using MedCalc (Belgium, Europe) (https://www.medcalc.org). To understand the false positives and/or weaknesses of our classifiers, images frequently misclassified by the classifiers were also reviewed.

Connections between liver metastasis-associated genes, immune cell infiltration and immune checkpoints

TIMER (https://cistrome.shinyapps.io/timer/) is a computational tool that aids in comprehensively exploring the molecular characterization of tumor immune interactions across diverse tumors[11]. It calculates the abundances of six immune infiltrates (CD8 + T cells, CD4 + T cells, B cells, neutrophils, macrophages, and dendritic cells) based on RNA-Seq expression profiles data. We used TIMER to evaluate correlations between liver metastasis-associated gene expression and six immune cells and CTLA4, CD274, PDCD1, & PDCD1LG2 checkpoint genes.

Immunohistochemistry and scoring

All specimens were routinely fixed in 10% buffered neutral formalin and embedded in paraffin. Each Sect. (5.0 µm) was stained with the standard hematoxylin and eosin method for screening review by two experienced pathologists. Immunohistochemistry was performed using an EnVision two-step method. Primary antibodies included anti-napsin A, anti-KRT7, anti- CK20, anti-CDX2 and anti-TTF-1 antibody (all at a 1:200 dilution, mouse monoclonal antibody; Dako, Denmark). The paraffin was removed from the slides by xylene, and the tissue was rehydrated in various concentrations of ethanol. Hydrogen peroxide (3%) was added to eliminate endogenous peroxidase activity. The slides were then processed using steam-heat retrieval for 30 min. The antibody was incubated for 30 min at 37 °C. The above mentioned steps were performed automatically in a Dako automatic immunostainer. Negative controls for napsin A and TTF-1 were carried out by omitting the primary antibody. For napsin A & KRT7, only a granular cytoplasmic staining pattern was accepted as positive, while for TTF-1, only a nuclear staining pattern was considered to be positive. The stained slides were observed microscopically by two experienced pathologists. Results were scored in accordance with the standard criteria and expressed on a plus-minus scale based on the proportion of tumor cells stained: − , 0–10%; + + , 10–50% and + + + , 50–100%.

Statistical analysis

All experiments were reproducible and each set was repeated at least three times to check the reproducibility. All data were recorded as mean ± standard deviation (SD) unless stated otherwise. PCA and LDA were performed using SPSS 20.00 (https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-20) statistical software (Chicago, IL, USA). Kaplan Meier Survival curve analysis was performed using Graph Pad Prism 7.0 software (GraphPad Software Inc.) (https://www.graphpad.com/support/prism-7-updates/). Chi Square test and Mann Whitney U test were performed using the SPSS 29 software (IBM, USA) (https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-29) for clinical correlation with differential gene expression data. P-value < 0.05 was considered to be statistically significant.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional ethics committee, and were approved and reviewed by The Gujarat University ethics committee (GU/IEC/07/2018)] and the The Gujarat Cancer Research Institute /Gujarat Cancer Society Internal Ethics Review board (EC-O-92-2014) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Written Informed consent was obtained from all individual participants included in the study.

Results

Isolation and characterization of CTCs from PBMNCs of primary and metastatic tumors

PBMCs were isolated from peripheral blood of NSCLC patients with and without liver metastasis and cells were expanded ex-vivo after incubation for 24 h. The cells were cultured for a period of 15–20 days and circulating tumor stem like cells were observed as circular cells with spikes on their circumference. These cells were considered to be CTCs and their morphology and other characteristics were examined by phase contrast microscopy. Furthermore, we characterized these CTCs to check the purity by analyzing the presence of CSC and epithelial markers like CD44, CD24, CD45 and CK using flow cytometry. Based on the flow cytometry analysis, we identified three distinctive sub-populations: (i) CD44 + CD24-; (ii) CD44 + CK7 + ; (iii) CK7 + [Supp Fig. 1(A)] for primary lung cancer patients whereas the other hand, CTCs isolated from patients with lung cancer liver metastasis exhibited three distinctive subpopulations: (i) CD44 + CD24 + ; (ii) CD44 + CD24-; and (iii) CD24 + [Supp Fig. 1(B)]. However, absence of CD45 + cells in the sorted population confirmed the non-lymphocytic characteristic of these cancer stem like cell population.

Increased self-renewal potential and drug resistance properties of CTCs:

Intrinsic sphere forming ability of the isolated CTCs was assessed and they demonstrated enhanced colony forming ability with a significant increase in the number and size of spheres of metastatic patients [Supp Fig. 1(D)] as compared to CTCs isolated from primary tumors [Supp Fig. 1(C)]. Further, on assessing the proliferation rate, CTCs isolated from both the primary and metastatic patients demonstrated a colony forming efficiency of nearly 60–70% with the colonies exponentially increasing at regular intervals, suggestive of the fact that both the subpopulations possess the intrinsic self-renewal potential but the metastatic subpopulation had a 15% higher proliferation efficiency as compared to the other counterpart [Supp Fig. 1(E) & (F)].

Drug resistance is an intrinsic property of CSC population as these slow growing cells tend to often escape the conventional therapeutic regimen unlike the proliferating cells. Thus, we assessed the cytotoxic effect of carboplatin and cisplatin on CTCs isolated from primary lung cancer with and without metastasis. The CTCs from lung cancer and liver metastasis demonstrated 88.2% and 89.32% viability respectively as compared to merely 36.43% viability of MNC cells from healthy individuals even at higher concentrations of Cisplatin (100 µM; p-value < 0.001) Moreover, lung cancer derived CTCs showed viability of 85% and those of metastasis demonstrated a viability of 88% at higher concentrations of Carboplatin (100 µM; p-value < 0.001) as compared to the healthy MNC cells (viability 41.8%) [Supp Fig. 1(G) & (H), Supp Table 1 & 2]. Collectively, these results are suggestive of the fact that the CTCs were significantly chemo-resistant and showed dose dependent drug resistance to advanced chemotherapeutic drugs—Cisplatin and Carboplatin.

Characterization of the exosomes derived from primary lung cancer cells with and without liver metastasis

To investigate the role of exosomes in lung cancer liver metastasis, exosomes were isolated from the serum of lung patients with and without liver metastasis and their identity was confirmed by electron microscopy, flow cytometry and NTA. The vesicles isolated from the serum appeared as round-shaped 40–50 nm diameter vesicles under electron microscopy [Supp Fig. 2(A)]. We further confirmed whether these vesicles were exosomes by performing NTA analysis [Supp Fig. 2(B)]. Results of flow cytometry also showed CD63 and CD81to be present in exosomes derived from serum of lung cancer patients with and without liver metastasis which is are commonly used exosomal markers [Supp Fig. 2(C)].

Differential gene expression analysis in CTC, Exosomes and cfRNA: identifying a better marker for liquid biopsy

Quantitative gene expression patterns of CXCL12, CK7, CDH1, CTNNB1, HIF1A, MUC16, TGFBR2 and CD44v6 were analyzed from CTC, exosomes and cell free RNA of lung cancer patients with and without liver metastasis. CXCL12, CK7, CDH1, CTNNB1, TGFBR2 and CD44v6 were upregulated in CTC as well as cfRNA (Fig. 2A and C) whereas all these genes were downregulated in exosomes of primary lung cancer with and without metastasis (Fig. 2B). Interestingly, it was observed that HIF1A was downregulated in primary lung cancer but was upregulated in liver metastasis in CTCs and exosomes (Fig. 2A and B). Moreover, MUC16 was only expressed in CTCs derived from liver metastasis but were not expressed in CTCs of primary lung cancer (Fig. 2A). In exosomes, it was seen that MUC16 was upregulated in primary lung cancer whereas it was downregulated in liver metastasis (Fig. 2B). In cfRNA the expression of all the genes were observed to be upregulated in both the cohorts (Fig. 2C).

Generation of candidate biomarker panel

ROC analysis, was performed, using the values, to determine and confirm robustness of the models generated using different sample source (Exosomes, cfRNA and CTC) comprising of the genes (CD44v6, CXCL12, CTNNB1, CK7, HIF1A, TGFBR2, MUC16, CDH1) that were shortlisted for the tissue model (Fig. 2D)[8]. Area under the curve (AUC) specificity and sensitivity are mentioned in Supp Table 2. In order to determine if the multigene signature has any practical application, we performed ROC analysis for individual genes as seen in Supp Fig. 3 and the genes with maximum significant sensitivity, specificity & AUC in exosomes, cfRNA and CTC were further considered for the final biomarker model. From the ROC curve analysis of individual genes, we further speculated that a more precise model for prediction of lung liver metastasis could be formed with 5 genes (CXCL12, CK7, CD44v6, HIF1A, and TGFBR2) algorithm as described in Fig. 2E and Supp Table 3. Moreover, it was observed that exosomal model came out as the best model and was at par with the tissue model and hence can be used for predicting liver metastasis non-invasively with the highest AUC of 0.975 a Specificity and Sensitivity of 90 and 96.87% respectively showing good discriminative ability between primary and metastatic tumors. CTC and cfRNA did not show significant AUC, sensitivity and specificity as a diagnostic model and hence cannot be used for early prediction of liver metastasis in lung cancer patients.

PCA Analysis distinguishing between primary and metastatic carcinomas.

The liver is one of the most frequent sites of metastasis. Distinguishing between primary and metastatic carcinomas is important due to differences in their management. Therefore, genes from FNAC samples were profiled and added to the primary lung carcinoma dataset. The metastases originated from lung carcinomas. When the combined dataset was analyzed by PCA, primary and metastatic adenocarcinomas were clearly distinct in the tissue and the exosome cohorts (Fig. 3C and D). The small sample size for metastatic adenocarcinomas precluded us from performing class prediction analysis after training-to-test set allocation. Instead, we performed cross-validation analysis using all samples as a training set. The permutation p-values for the misclassification rate (at feature selection $p < 10^{-6}$) were CCP: 0.02, LDA: 0.01, 1-NN: < 0.01, 3-NN: 0.01, NC: 0.01 and SVM: < 0.01. In primary *vs.* metastatic carcinomas, *CXCL12, CD44v6, CK7, HIF1A and TGFBR2* were differentially expressed

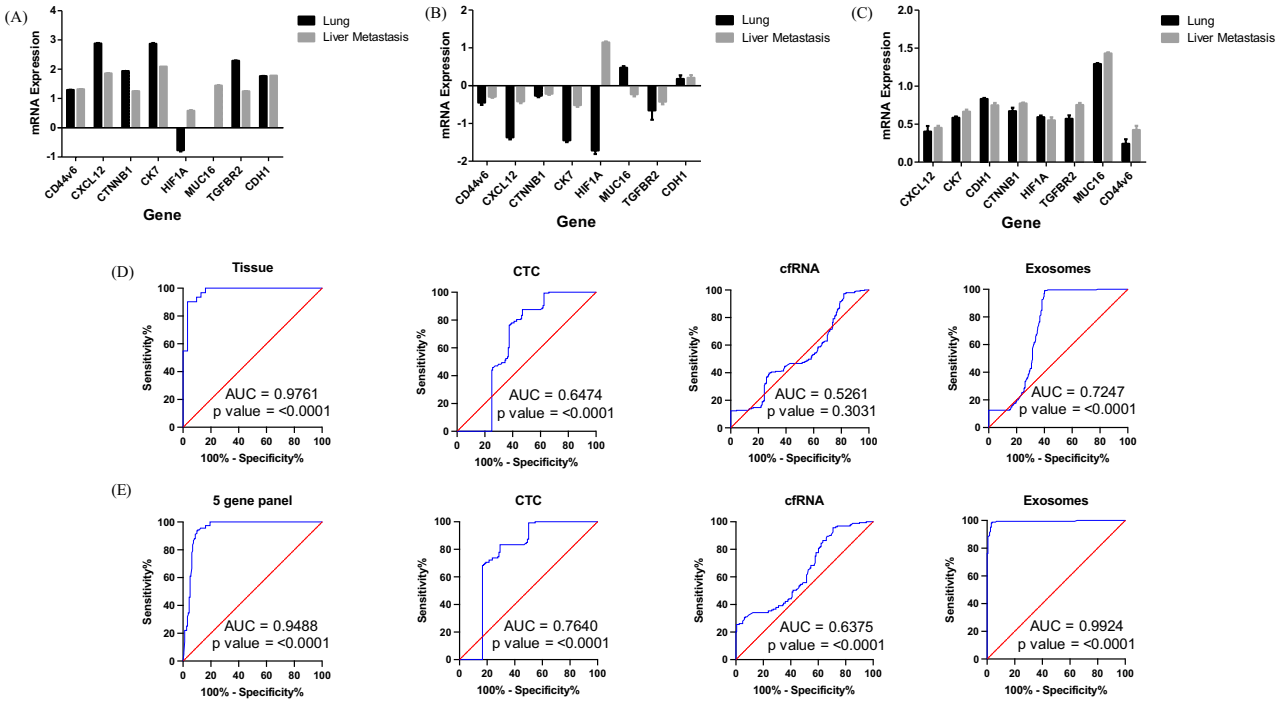


Figure 2. Gene expression analysis and performance analysis of the 8-gene & 5 – gene panel by quantitative RT-PCR in—(**A**) CTC, (**B**) Exosomes and (**C**) cfRNA derived from primary lung tumors and advanced stage lung cancer with liver metastasis im comparisison to healthy controls and (**D**) Performance of the 8-gene panel in predicting liver metastasis using multivariate receiver operating characteristics curve for CTC, cfRNA and exosomes in comparison to the matched tissue samples. (**E**) Performance of the 5-gene panel in predicting liver metastasis using multivariate receiver operating characteristics curve for CTC, cfRNA and exosomes in comparison to the matched tissue samples.

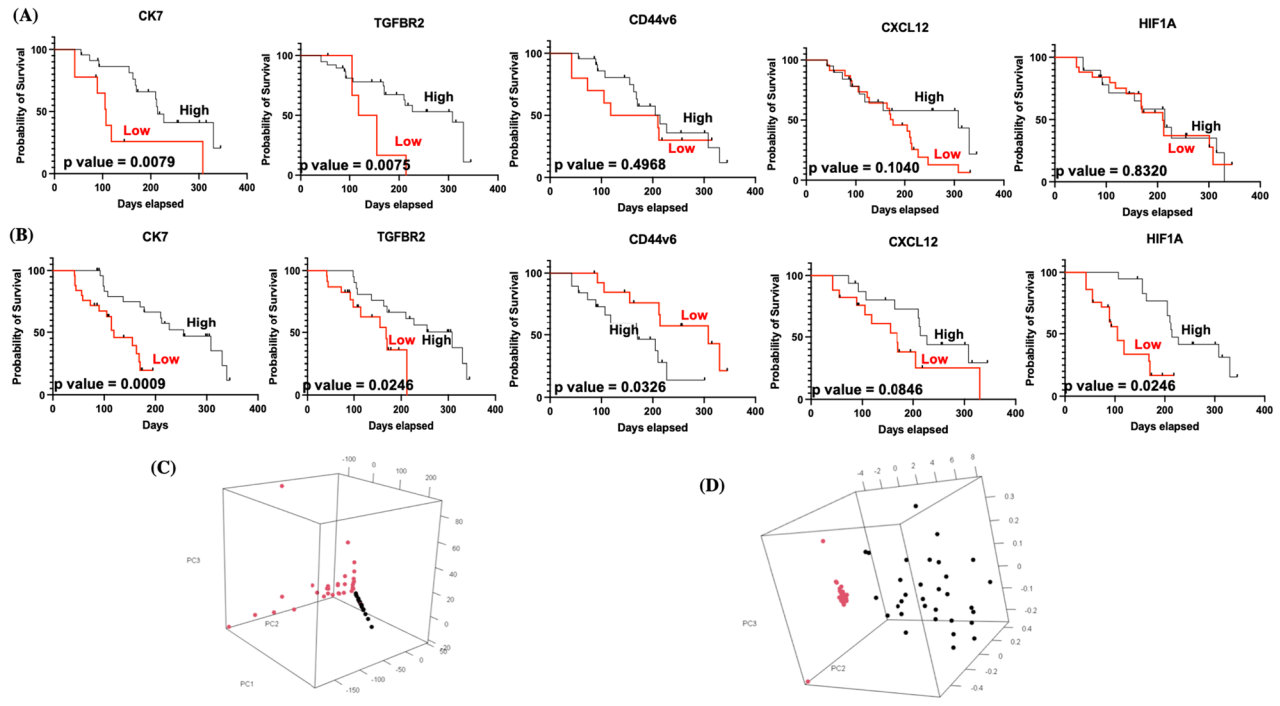


Figure 3. Kaplan Meier Survival Curves for (**A**) Tissue expression of CD44v6, CXCL12, CK7, HIF1A and TGFBR2 of lung cancer patients with liver metastasis, (**B**) exosome expression of CD44v6, CXCL12, CK7, HIF1A and TGFBR2 of lung cancer patients with liver metastasis. PCA analysis for differentiating between primary and the metastatic cohorts—3D score plot according to primary and metastatic sites in PCA for (**C**) tissue and (**D**) exosomes respectively, Red: Primary site expression Black: metastatic site expression.

at $p < 10^{-6}$; the level of expression of all the above-mentioned genes from exosomes of metastatic carcinoma were downregulated as compared to tissue samples where CD44v6 was upregulated and the other genes had a downregulated expression.

Risk Score equation for lung cancer liver metastasis

$$(0.713 * 0.47) + (0.570 * 0.340) + (0.632 * 0.346) + (0.774 * 10.305) + (0.588 * 0.468) = 8.689686$$

Risk Score equation for primary lung tumor

$$(-0.711 * 0.076) + (0.927 * 0.043) + (0.908 * 0.035) + (0.302 * 0.019) + (0.791 * 0.166) = 0.154649$$

When we further performed the ROC for the model we had got a cut off value of > 0.2 for predicting metastasis (mentioned in Supp Table 3). Our results comply with the associated criteria and thus from the above equation we can say that these markers together can be used for predicting liver metastasis in lung cancer patients. Moreover, from the above equation we observed that the metastasis patients have an average value of 8.689686 which is higher than the cut off value whereas for the primary patients the value obtained was 0.154649 which is lower than the cut off value for prediction. The combination variables had the best overall accuracy, resulting in an AUC of 0.80 (95% CI = 0.74 to 0.86). Next, ideal cutoff points were assessed as maximum sum of sensitivity and specificity. This revealed both a sensitivity and a specificity of 98.2% and 95.29% respectively and the AUC of 0.992 (95% CI = 0.968 to 0.996).

Gene expression and clinical/pathological correlation in metastatic tissue and exosome cohorts along with Immunohistochemistry marker expression

We further analysed the correlation between CXCL12, CD44v6, HIF1A, TGFBR2 and KRT7 expression in metastatic tumor tissues & exosomes with sex, age, habit, Tumor Differentiation grade, TNM stage, EGFR mutation status, AE1, Chromogranin, CDX2, CEA, CK20, TTF-1 & CK7. Results showed the correlation between CXCL12 and differentiation grade, CEA and TTF-1 protein expression, CD44v6 was associated with tumor stage, chromogranin expression whereas HIF1A and KRT7 were significantly associated with CEA protein expression in the metastatic tissues. On the other hand, in the exosome cohort it was observed that CD44v6 was associated with stage, Chromogranin, CK20 & TTF-1 protein expression, CXCL12 was significantly correlated to EGFR mutations, TGFBR2 was significantly associated with AE1 protein expression and HIF1A with CEA protein expression (p value < 0.05) (Tables 1 & 2). Further as seen in Fig. 4 the Immunohistochemistry results suggested that Napsin A and KRT7 showed Strong Positive staining in the tissue samples of the lung cancer liver metastatic patients as compared to TTF-1 and CDX2 that showed Medium positive expression. However, in the same samples CK20 showed no staining in the same set of samples.

Clinical validation, risk stratification and heatmap generation

The five-marker signature was tested in the exosomes isolated from the blood of patients in the validation set (n = 79). The score plot demonstrated that patients at high-risk had poorer outcomes and developed liver metastasis as compared to patients at low-risk. Figure shows the distribution of the Risk Score, and heatmap of the 5-gene signature in the validation set (Fig. 5A and B). It was observed that out of 79 patients 44 patients had a high risk for developing liver metastasis as compared to the other 35 patients. In the high-risk cohort, it was observed that out of 44 patients 39 developed liver metastases, 3 patients developed multiple metastasis (brain, bone & liver), whereas 2 patients did not develop any distant metastasis. In the other 35 patients 9 patients could not be evaluated at follow up as there was no data available whereas the other 26 did not develop any distant metastasis. The survival prognosis was significantly worse for the high-risk group as compared to the low risk cohort (Fig. 5C). These results demonstrate that the algorithm developed could accurately predict liver metastasis in 42 out of 44 patients (95% accuracy) with high-risk score based on the equation derived.

Clinical relevance and Kaplan–Meier survival analyses of the metastasis associated genes in tissue and exosome cohorts

To further demonstrate the clinical significance of the 5-gene signature in patients with lung cancer liver metastasis, the association between gene expressions and various clinico-pathological variables was investigated by real-time quantitative PCR in 32 lung cancer liver metastasis patients. We found that in the tissue cohort CXCL12 expression was strongly correlated with differentiation, CEA & TTF-1 protein expression, CD44v6 was associated with tumor stage and chromogranin protein expression and HIF1A & KRT7 showed positive correlation with CEA expression alone (Table 1). In the exosome cohort CXCL12 expression was strongly correlated with EGFR status, CD44v6 was associated with tumor stage, TTF-1 expression & chromogranin protein expression and HIF1A & TGFBR2 showed positive correlation with CEA expression & CK20 expression (Table 2). The Kaplan–Meier survival analyses of the metastasis associated genes in lung cancer patients with liver metastasis was evaluated. Among all the patients, low expression of KRT7 ($P = 0.0079$) & TGFBR2 ($P = 0.0075$) exhibited significantly worse Overall Survival (OS) compared to high expression in tissue cohort (Fig. 3A). Contradictorily it was observed that low expression of KRT7 ($P = 0.0009$), TGFBR2 ($P = 0.024$), & HIF1A ($P = 0.0246$) and high expression of CD44v6 ($P = 0.0326$) was associated with worse OS in the exosome cohort (Fig. 3B).

Sr. No		n values (%)	CXCL12 (Mean rank values)	*p*	TGFBR2	*p*	CD44v6	*p*	HIF1A	*p*	KRT7	*p*
1	Age	**33**		0.417		0.857		0.259		0.787		0.1074
	> = 59	**18**	15.72		16.69		15.25		16.56		14.5	
	< 59	**15**	18.53		17.37		19.1		17.53		20	
2	Habit			0.096		0.689		0.253		0.0818		0.624
	Smoking	**23**	18.73		17.44		18.3		15.04		16.43	
	Non-smoking	**10**	12.39		15.83		14		21.5		18.3	
3	Differentiation		**0.048**		0.685		0.671		0.896		0.659	
	Poorly diff	**12**	12.58		16.58		16		16.67		15.91	
	Diff unknown	**21**	19.52		17.24		17.57		17.19		17.55	
4	Stage			0.681		0.96		**0.008**		0.857		0.303
	III	**15**	17.88		17.14		21.8		16.58		14.81	
	IV	**18**	16.42		16.89		13		17.27		18.42	
5	EGFR Status		0.275		0.322		0.427		0.645		1	
	Wildtype	**12**	19.64		14.59		15.21		15.86		17.05	
	Mutation	**21**	15.68		18.2		18.02		17.57		16.98	
6	AE1			0.417		0.303		0.899		0.528		0.496
	Positive	**13**	15.27		19.19		17.31		18.35		18.46	
	Negative	**20**	18.12		15.58		16.8		16.12		16.05	
7	Chromogranin		0.818		0.896		**0.043**		0.802		0.603	
	Positive	**11**	16.56		18.28		13.45		17.47		17.94	
	Negative	**22**	17.41		17.76		20.77		16.56		16.12	
8	CEA			**0.031**		0.96		0.735		**0.0093**		**0.0423**
	Positive	**15**	21		16.87		17.63		21.83		12.64	
	Negative	**18**	13.67		17.11		16.47		12.97		19.5	
9	CK20			0.541		0.638		0.696		0.1556		0.483
	Positive	**6**	14.75		18.75		15.75		22.17		19.58	
	Negative	**27**	17.5		16.61		17.4		15.85		16.43	
10	CK7			0.173		1		0.643		0.757		0.152
	Positive	**28**	17.98		17.02		16.64		17.09		18.04	
	Negative	**5**	11.5		16.9		19		19.9		11.2	
11	CDX2			0.218		0.794		0.899		0.332		0.238
	Positive	**12**	14.62		16.38		17.31		19.21		19.67	
	Negative	**21**	19.02		17.36		16.8		15.74		15.48	
12	TTF-1			**0.033**		0.802		0.708		0.322		0.158
	Positive	**15**	20.97		17.5		16.3		15.13		14.37	
	Negative	**18**	13.69		16.58		17.58		18.56		19.19	

Table 1. Correlation between tissue gene panel expression levels and clinical characteristics. Significant values are in bold.

Association of marker panel genes to immune cell infiltration in LCLM

To ascertain if correlations existed between tumor infiltrating immune cells and metastasis-associated gene levels, we examined associations using TIMER. KRT7 expression was positively correlated with infiltrating CD8 + T cells ($P = 0.00923$), B cells ($P = 6.50e{-}10$), and myeloid dendritic cells ($P = 0.0048$). Similarly, HIF1A was also positively correlated with infiltrating B cell ($P = 0.03$), CD8 + T cells ($P = 8.14e{-}08$), macrophages ($P = 7.32e{-}07$), neutrophil ($P = 3.66e{-}18$), and myeloid dendritic cells ($P = 4.24e{-}09$). Moreover, CD44, CXCL12 and TGFBR2 were positively correlated with all tumor infiltrating immune cells (Fig. 6). As current immunotherapy strategies rely on immunological checkpoint inhibitors[12,13] we used TIMER to investigate co-expression relationship of all 5 genes with immune checkpoint-related genes. CD44, CXCL12 and HIF1A displayed strong co-expression relationships with CD274, CTLA4, PDCD1, and PDCD1LG2, whereas KRT7 had co-expression relationships with CTLA4, & PDCD1LG2 and TGFBR2 has co-expression with CTLA4, CD274 and PDCD1LG2 (Fig. 7).

Discussion

Non-small cell carcinoma (NSCLC) is one of the major causes of cancer death globally. Liver metastasis is one of the most critical prognostic factor for NSCLC[14]. Early detection of heterochronous lung cancer liver metastasis, may provide useful information for designing treatment strategy to improve patient survival. Most patients with early lung cancer have a 70% survival rate after appropriate surgical resections, yet many patients already develop distant metastasis, which is difficult to detected solely by imaging methods[15]. Moreover, patients with

Sr. No		n values (%)	CXCL12 (Mean rank values)	*p*	TGFBR2	*p*	CD44v6	*p*	HIF1A	*p*	KRT7	*p*
1	Age	**34**		0.794		0.417		0.851		0.849		0.7188
	> = 59	**18**	17.94		16.17		17.19		17.17		18.11	
	< 59	**16**	17		19		17.84		17.88		16.81	
2	Habit			0.234		0.865		0.828		0.378		0.603
	Smoking	**23**	18.35		17.72		17.24		18.57		18.13	
	Non-smoking	**11**	13.9		17.05		18.05		15.27		16.18	
3	Differentiation		0.535		0.183		0.944		0.617		0.872	
	Poorly diff	**13**	18.42		14.58		17.35		18.62		17.88	
	Diff unknown	**21**	16.19		19.31		17.6		16.81		17.26	
4	Stage			0.96		0.096		**0.03**		0.275		0.928
	III	**16**	17.13		13.5		19.88		14.69		16.79	
	IV	**18**	16.89		19.27		14.39		18.5		17.16	
5	EGFR Status		**0.043**		0.141		0.606		0.214		0.936	
	Wildtype	**12**	21.64		13.45		16.25		20		17.23	
	Mutation	**22**	14.68		18.77		18.18		15.5		16.89	
6	AE1			0.984		**0.043**		0.274		0.368		0.327
	Positive	**14**	17.08		12.92		19.75		18.92		14.92	
	Negative	**20**	16.95		19.65		15.93		15.75		18.35	
7	Chromogranin		0.624		0.447		**0.024**		0.322		0.368	
	Positive	**11**	16.12		18.34		20.59		15.25		15.41	
	Negative	**23**	17.82		15.7		14.02		18.65		18.5	
8	CEA			0.092		0.207		0.443		**0.0009**		0.303
	Positive	**16**	13.87		19.81		18.94		23.5		18.93	
	Negative	**18**	19.61		15.44		16.22		12.17		15.39	
9	CK20			0.378		**0.013**		0.413		**0.037**		0.748
	Positive	**6**	14.17		10.83		14.88		24.33		16.25	
	Negative	**28**	18.21		19.71		18.31		16.04		17.77	
10	CK7			0.352		0.118		0.644		0.126		0.596
	Positive	**28**	16.32		18.12		17.13		15.89		16.61	
	Negative	**6**	20.8		10.7		19.25		23.2		19.2	
11	CDX2			0.727		0.293		0.691		0.1443		0.373
	Positive	**12**	18.33		16.14		16.64		20.92		15.96	
	Negative	**22**	17.05		20		18.1		15.64		19.2	
12	TTF-1			0.0929		0.667		**0.015**		0.126		0.378
	Positive	**15**	13.87		18.37		14.16		20.47		15.77	
	Negative	**19**	20.61		16.82		20.58		15.16		18.87	

Table 2. Correlation between exosome gene panel expression levels and clinical characteristics. Significant values are in bold.

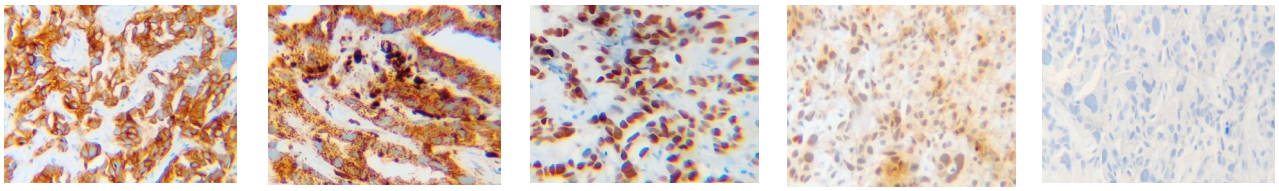


Figure 4. Immunohistochemistry staining of KRT7, Napsin A, TTF-1, CDX2 and CK20 in the lung cancer liver metastatic tissue samples.

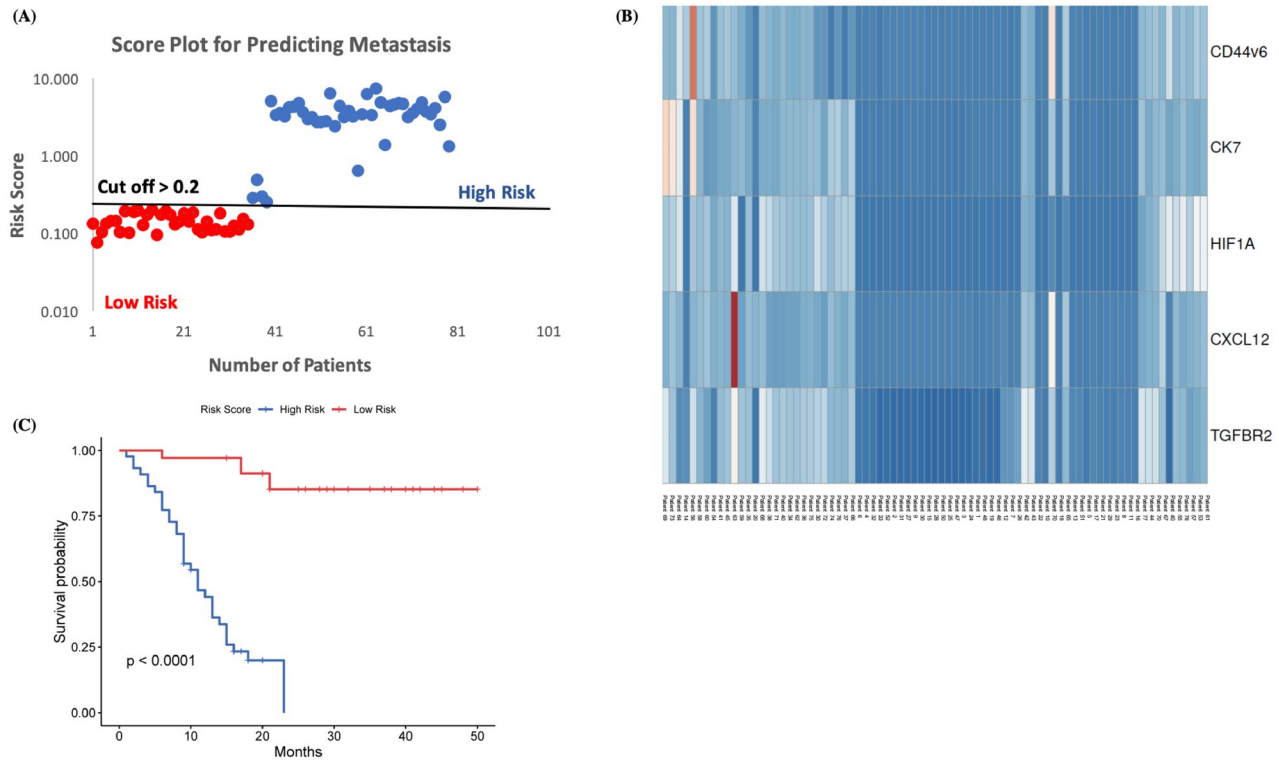


Figure 5. Score Plot and Heat Map generation for the validation study. Distribution of (**A**) risk score and (**B**) heat map of five genes expression (**C**) Kaplan Meier Survival curve for the high-risk group and low risk group.

poor tumor differentiation and higher TNM stage develop distant metastasis, and their average survival period is 4 months[16]. However, TNM staging is unable to facilitate accurate prognosis and hence it becomes difficult to determine treatment. It has been recommended that the traditional TNM staging should include number of involved metastatic sites, number of metastatic foci per involved site and the diameter of each metastatic focus. This will help in prediction of metastasis for more accurate tumor staging, which could improve treatment efficacy and increase survival rates[17].

Presently, serum biomarkers including sCEA, sCA125, sCA199 and sNSE have been widely used in the diagnosis and monitoring the progression of lung cancer liver metastasis[18]. However, the sensitivity of CEA and CA199 have always been lower compared to that of sCA125 and NSE. Unfortunately, the results of predictive ability of these markers or in combination with other indices are greatly discrepant[19,20]. Therefore, identification of non-invasive biomarkers for early diagnosis of NSCLC LM and dynamic monitoring of disease status is a key research imperative. There is a growing consensus that biomarker panels have higher specificity and sensitivity than single biomarkers and may be more effective in detecting cancer[20,21]. Herein, we sought to discover the unique patterns of circulating markers in NSCLC patients with LM, and to identify biomarkers with sufficient sensitivity and specificity for use as a supplement or substitute for invasive biopsy in clinical practice. In the present study, exosomal CXCL12, TFBR2, CD44v6, HIF1A and KRT7 were identified as novel diagnostic biomarkers for NSCLC LM. Moreover, the combination of this marker panel may be a sensitive biomarker for follow-up monitoring of therapeutic response, as well as a prognostic marker in patients with LM.

In this study, we used a previously established multi-gene panel for generating a model specific for prediction of liver metastasis. We generated the model using PCA and LDA and further calculated the accuracy of the model generated by ROC curve analysis that showed accuracy greater than 90%. Additionally, the Kaplan Meier survival data also showed a significance ($p < 0.0001$) amongst the gene expression patterns of the different genes and their association with overall survival (OS) in metastatic patients. Moreover, the developed equation with the co-efficient and the risk score aided in predicting metastasis in the primary lung cancer validation cohort with an accuracy of 92%. Furthermore, there was a significant association between the gene panel expression and the tumor infiltrating markers signifying the important role of these biomarkers in the development and diagnosis of liver metastasis. Thus, the five-gene panel (CXCL12, KRT7, CD44v6, TGFBR2 and HIF1A) can be a highly significant predictor of liver metastasis outcome independent of the standard prognostic criteria. Additionally, the ability of this panel to accurately predict recurrence in the liver independent of stage of the primary tumor is likely to be a useful enhancement to routine staging.

The biomarker panel identified in this study are not conventional tumor-derived cancer biomarkers but their combination generating a cut off score that aids in early detection of liver metastasis. This panel shows better diagnostic accuracy in comparison to the individual markers. Stromal cell-derived factor 1 (SDF-1) also known as CXCL12 is a crucial chemotactic factor that stimulates proliferation, adhesion, dissociation, migration, survival of tumor cells and the formation of tumor- associated vessels and invasion in a wide variety of tumor cells,

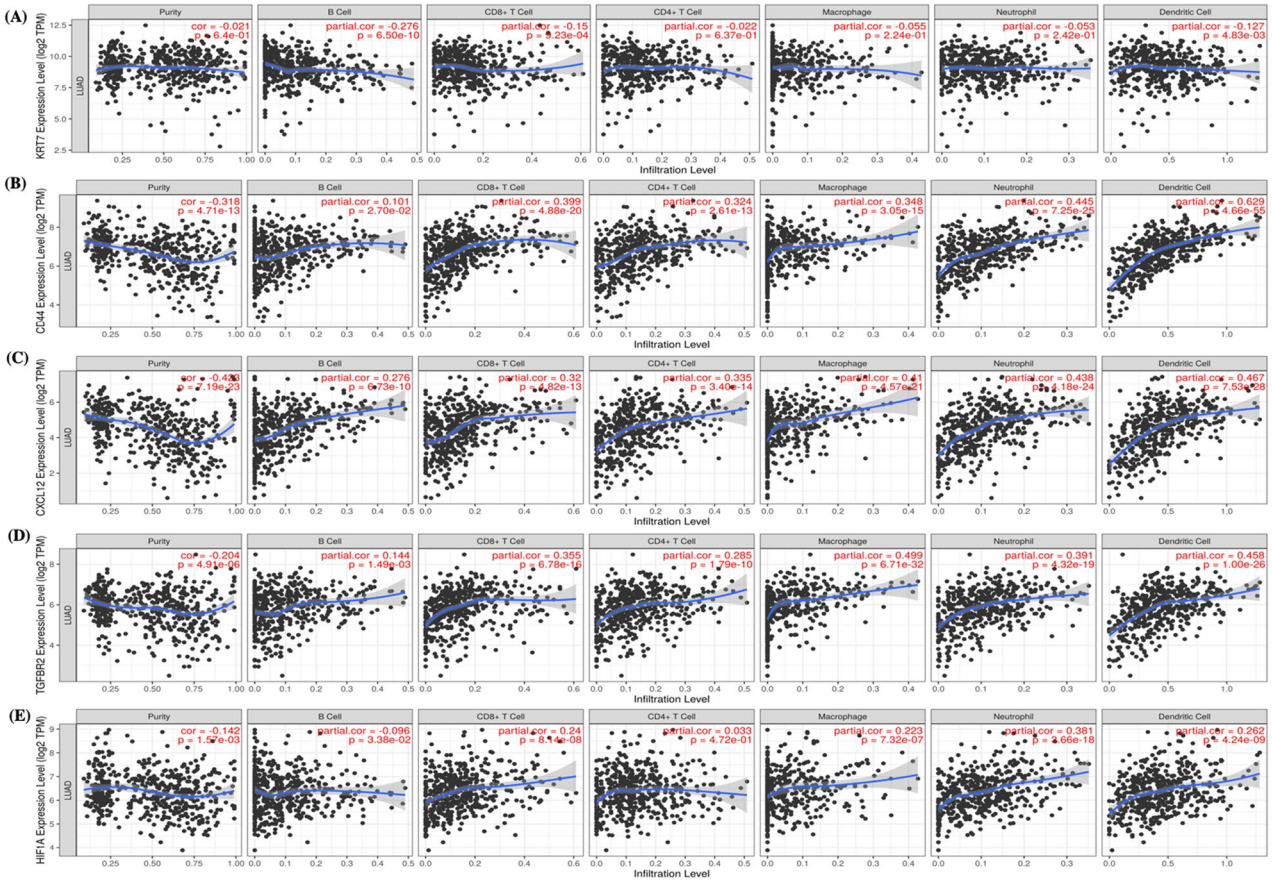


Figure 6. Analysis of immune cell infiltration. KRT7 (**A**), CD44 (**B**), CXCL12 (**C**), TGFBR2 (**D**) and HIF1A (**E**) expression correlations with tumor purity and infiltrating levels of B cells, CD8 + T cells, CD4 + T cells, macrophages, neutrophils, and dendritic cells in TIMER database. The scatter plots displayed the purity-adjusted spearman's rho value and statistical significance.

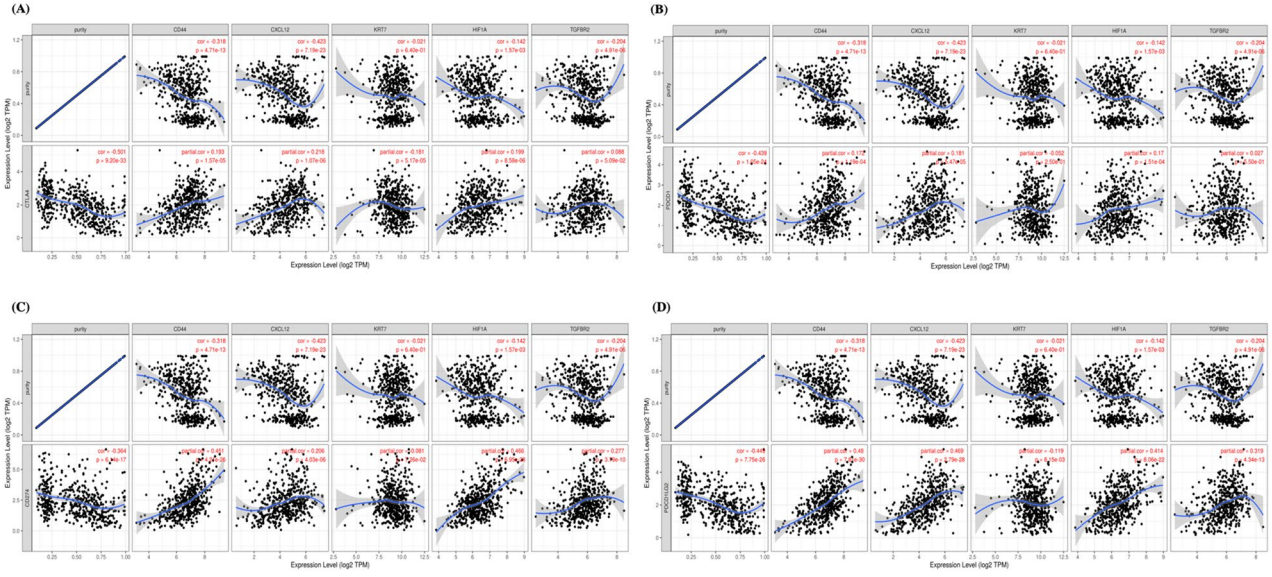


Figure 7. Association between KRT7, CD44, CXCL12, TGFBR2 and HIF1A expression level and immune checkpoints genes (**A**) CTLA4, (**B**) PDCD1, (**C**) CD274, and (**D**) PDCD1LG2) using TIMER ($n = 259$).

especially reported in liver metastasis[22]. Furthermore, the CXCL12/CXCR4 has been recognized as a prognostic marker in different cancers and preclinical models; signifying that metastasis is mediated by CXCR4 activation and migration of tumor cells towards CXCL12 expressing organs. Moreover, KRT7 is a membrane-cytoskeletal linker which contributes to regulation of cell adhesion and overexpression is associated with cellular transformation. KRT7 plays an important role in preserving tissue architecture, tumor progression, invasion, and metastasis. Cytoskeletal component levels, such as the degree of keratinization, have been reported to be valuable in identifying patients at risk for developing regional metastases[23]. CD44 transmembrane glycoproteins are cell adhesion molecules that have been associated with aggressiveness and metastasis. Through HIF-1/2α, hypoxia transcriptionally controls a host of factors that contribute to increased resistance to radiation and chemotherapy, and to the emergence of a more aggressive phenotype by regulating CD44v6 in metastasis[24–27]. The functional effects of the upregulation of CD44v6 and its variant isoforms under hypoxic conditions should be considered, since cell signaling events that promote anchorage-independent tumor cell growth, survival, migration and metastasis occur through the binding of hyaluronan with CD44[28]. Upregulation of both CD44 and hyaluronan under hypoxic conditions most likely amplify the signaling events and lead to tumor progression[29]. Hypoxia Inducing Factor 1α (HIF1A) plays an important role in the formation of liver metastasis. Further studies have reported that HIF1A overexpression enhances ZEB1 trans-activity by binding to its promoter leading to a loss in E-cadherin causing increased invasion, migration and tumor progression[30]. Lastly, TGFB signaling is also suggested to be involved in the malignant progression of tumors and can induce epithelial–mesenchymal transition and promote tumor cell invasion, and may also have angiogenic and immunosuppressive effects on the tumor microenvironment, all of which promote metastasis.8 It has been observed that reduced TGFBR2 expression significantly is correlated with the aggressive features in Hepatocellular carcinomas as well as other tumors. Morover, TGFBR2 has also been reported in liver metastasis arising from colorectal tumors[31]. Additionally, all these markers have an important role in liver metastasis and hence the generated model will assist in the early diagnosis of liver metastasis.

Although combination of exosomal serum levels of CD44v6, HIF1A, TGFBR2, KRT7 and CXCL12 seem to be promising biomarkers and may guide in the development of novel targeted drugs, some limitations of our study should be considered while interpreting the results. For example, there might be some inherent biases since clinical parameters are variable between institutions and individual clinicians. Therefore, a well-designed and large-scale multicenter follow-up cohort study is warranted to provide more robust evidence.

Conclusion

In summary, using a multiplexed, molecularly driven approach, we have identified a panel comprising CXCL12, KRT7, CD44v6, TGFBR2 and HIF1A that can predict recurrence in the liver independent of conventional prognostic criteria and identify patients with lung cancer who will develop liver metastasis despite undergoing definitive surgery and/or treatment. Increasing numbers of alterations in these genes predict poorer prognosis. As an outcome of this study, we report a sensitive biomarker panel which can be developed into a multi-parameter rapid testing kit to explore its potential in clinical settings. However, the study being a pilot study in nature, the outcome needs to be evaluated in a larger patient population. For future directions, cross-validation of this prediction model in terms of accuracy, precision, sensitivity, specificity and positive and negative predictive values using a separate large cohort of lung cancer liver metastasis cases, disease controls and healthy controls is needed so that the potential value of this prediction model as a biomarker panel in the clinical setting can be explored and a rapid detection kit can be developed for the model to facilitate population screening.

Data availability

All data generated or analyzed during this study are included in this published article and its supplementary information files. Further for the real time raw data the datasets generated during the current study will be available from the corresponding author on reasonable request.

Received: 7 February 2024; Accepted: 27 May 2024
Published online: 14 June 2024

References

1. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics. *Cancer J. Clin.* **71**(2021), 7–33 (2021).
2. Brodt, P. Role of the microenvironment in liver metastasis: From pre- to prometastatic niches. *Clin. Cancer Res.* **22**, 5971–5982 (2016).
3. Tamura, T. *et al.* Specific organ metastases and survival in metastatic non-small-cell lung cancer. *Mol. Clin. Oncol.* **3**, 217–221 (2015).
4. Yang, J. *et al.* The prognostic impact of lymph node metastasis in patients with non-small cell lung cancer and distant organ metastasis. *Clin. Exp. Metastasis* **36**, 457–466 (2019).
5. Jiang, C. *et al.* Extracellular vesicles promotes liver metastasis of lung cancer by ALAHM increasing hepatocellular secretion of HGF. *iScience* **25**(3), 103984. https://doi.org/10.1016/j.isci.2022.103984 (2022).
6. Shah, K., & Rakesh, R. A multi-gene predictor for detection and monitoring liver metastasis in lung cancer patients: An algorithmic, *in-vitro* and in-vivo approach [abstract]. In *Proceedings of the AACR Special Conference: Cancer Metastasis*; 2022 Nov 14–17; Portland, OR. Philadelphia (PA): AACR; Cancer Res 2022;83(2 Suppl_2):Abstract nr B031.
7. Jain, A. *et al.* Identification of potential salivary biomarker panels for oral squamous cell carcinoma. *Sci. Rep.* **11**, 3365. https://doi.org/10.1038/s41598-021-82635-0 (2021).
8. Shah, K., Patel, S., Mirza, S. & Rawal, R. M. A multi-gene expression profile panel for predicting liver metastasis: An algorithmic approach. *PLoS ONE* **13**(11), e0206400. https://doi.org/10.1371/journal.pone.0206400 (2018).
9. Wang, X. *et al.* Investigation of the roles of exosomes in colorectal cancer liver metastasis. *Oncol. Rep.* **33**(5), 2445–2453. https://doi.org/10.3892/or.2015.3843 (2015).

10. Roessler, S. *et al.* A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. *Cancer Res.* **70**(24), 10202–10212 (2010).
11. Luan, H. *et al.* The identification of liver metastasis- and prognosis-associated genes in pancreatic ductal adenocarcinoma. *BMC Cancer* **22**, 463. https://doi.org/10.1186/s12885-022-09577-2 (2022).
12. Tomita, Y. *et al.* Nivolumab versus everolimus in advanced renal cell carcinoma: Japanese subgroup 3-year follow-up analysis from the Phase III checkmate 025 study. *Jpn. J. Clin. Oncol.* **49**(6), 506–514 (2019).
13. Vaddepally, R. K., Kharel, P., Pandey, R., Garje, R. & Chandra, A. B. Review of indications of FDA-approved immune checkpoint inhibitors per NCCN guidelines with the level of evidence. *Cancers (Basel)* **12**(3), 738 (2020).
14. Yang, X. & Meng, G. Establishment of a non-small-cell lung cancer-liver metastasis patient-derived tumor xenograft model for the evaluation of patient-tailored chemotherapy. *Biosci. Rep.* **39**(7), 20182082. https://doi.org/10.1042/BSR20182082 (2019).
15. Hess, K. R. *et al.* Metastatic patterns in adenocarcinoma. *Cancer* **106**(7), 1624–1633. https://doi.org/10.1002/cncr.21778 (2006).
16. Wu, B. *et al.* Comparison of the survival time in the non-small cell lung cancer patients with different organ metastasis. *Zhongguo Fei Ai Za Zhi* **22**(2), 105–110. https://doi.org/10.3779/j.issn.1009-3419.2019.02.05 (2019).
17. Goldstraw, P. *et al.* The IASLC lung cancer staging project: Proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer. *J. Thoracic Oncol.* **11**(1), 39–51. https://doi.org/10.1016/j.jtho.2015.09.009 (2016).
18. Wang, C. F. *et al.* The combination of CA125 and NSE is useful for predicting liver metastasis of lung cancer. *Dis Mark.* **9**(2020), 8850873. https://doi.org/10.1155/2020/8850873 (2020).
19. Mor, G. *et al.* Serum protein markers for early detection of ovarian cancer. *Proc. Natl. Acad. Sci. USA* **102**, 7677–7682 (2005).
20. Wang, D. *et al.* Serum CCL20 combined with IL-17A as early diagnostic and prognostic biomarkers for human colorectal cancer. *J. Transl. Med.* **17**, 253 (2019).
21. Wei, S. *et al.* Cathepsin F and Fibulin-1 as novel diagnostic biomarkers for brain metastasis of non-small cell lung cancer. *Br. J. Cancer* **126**, 1795–1805. https://doi.org/10.1038/s41416-022-01744-3 (2022).
22. Xin, Q. *et al.* CXCR7/CXCL12 axis is involved in lymph node and liver metastasis of gastric carcinoma. *World J. Gastroenterol.* **23**(17), 3053–3065. https://doi.org/10.3748/wjg.v23.i17.3053] (2017).
23. Jiang, L. *et al.* Differential gene expression identifies KRT7 and MUC1 as potential metastasis-specific targets in sarcoma. *Cancer Manag. Res.* **9**(11), 8209–8218. https://doi.org/10.2147/CMAR.S218676 (2019).
24. Mook, O. R. *et al.* Visualization of early events in tumour formation of eGFP-transfected rat colon cancer cells in liver. *Hepatology* **38**, 295–304 (2003).
25. Dittmar, T., Heyder, C., Gloria-Maercker, E. & Hatzmann, W. Zänker KSAdhesion molecules and chemokines: The navigation system for circulating tumour (stem) cells to metastasize in an organ-specific manner. *Clin. Exp. Metastasis* **25**, 11–32 (2008).
26. Gassmann, P. *et al.* CXCR4 regulates the early extravasation of metastatic tumour cells in vivo. *Neoplasia* **11**, 651–661 (2009).
27. Qiu, H. *et al.* Arrest of B16 melanoma cells in the mouse pulmonary microcirculation induces endothelial nitric oxide synthase-dependent nitric oxide release that is cytotoxic to the tumour cells. *Am. J. Pathol.* **162**, 403–412 (2003).
28. Kirui, J. K. *et al.* Gbetagamma signaling promotes breast cancer cell migration and invasion. *J. Pharmacol. Exp. Ther.* **333**, 393–403 (2010).
29. Wendel, C. *et al.* CXCR4/CXCL12 participate in extravasation of metastasizing breast cancer cells within the liver in a rat model. *PLoS One* **7**(1), e30046. https://doi.org/10.1371/journal.pone.0030046 (2012).
30. Zhang, W. *et al.* HIF-1α promotes epithelial-mesenchymal transition and metastasis through direct regulation of ZEB1 in colorectal cancer. *PLoS One.* **10**(6), e0129603 (2015).
31. Mamiya, T. *et al.* Reduced transforming growth factor-beta receptor II expression in hepatocellular carcinoma correlates with intrahepatic metastasis. *Lab. Invest.* **90**(9), 1339–1345. https://doi.org/10.1038/labinvest.2010.105 (2010).

Acknowledgements

I would like to acknowledge DHR – ICMR for my current project as a Young Scientist (File No: File No.R.12014/52/2022-HR). All authors have read the journal's policy on disclosure of potential conflicts of interest and a statement that all authors have disclosed any financial or personal relationship with organizations that could potentially be perceived as influencing the described research.

Author contributions

K.S. collected the data. K.S. and R.M.R. designed the research study. K.S. and R.M.R. analysed the data and drafted the manuscript. R.M.R. proofread the final manuscript. All authors approved the final version of the manuscript.

Funding

This work was supported by the grant from Indian Council of Medical Research (File No: 5/3/8/51/ITR-F/2018-ITR).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-63252-z.

Correspondence and requests for materials should be addressed to R.M.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2024