



OPEN

iTRAQ proteomics of sentinel lymph nodes for identification of extracellular matrix proteins to flag metastasis in early breast cancer

Sheetal Pathania¹, Mohd Imran Khan¹, Sabyasachi Bandyopadhyay², Suneet Shekhar Singh², Komal Rani¹, Tanvi Ramesh Parashar¹, Jnaneshwari Jayaram³, Piyush Ranjan Mishra³, Anurag Srivastava³, Sandeep Mathur⁴, Smriti Hari⁵, Perumal Vanamail⁶ & Gururao Hariprasad¹✉

Patients with early breast cancer are affected by metastasis to axillary lymph nodes. Metastasis to these nodes is crucial for staging and quality of surgery. Sentinel Lymph Node Biopsy that is currently used to assess lymph node metastasis is not effective. This necessitates identification of biomarkers that can flag metastasis. Early stage breast cancer patients were recruited. Surgical resection of breast was followed by identification of sentinel lymph nodes. Fresh frozen section biopsy was used to assign metastatic and non-metastatic sentinel lymph nodes. Discovery phase included iTRAQ proteomics coupled with mass spectrometric analysis to identify differentially expressed proteins. Data is available via ProteomeXchange with identifier PXD027668. Validation was done by bioinformatic analysis and ELISA. There were 2398 unique protein groups and 109 differentially expressed proteins comparing metastatic and non-metastatic lymph nodes. Forty nine proteins were up-regulated, and sixty proteins that were down regulated in metastatic group. Bioinformatic analysis showed ECM-receptor interaction pathways to be implicated in lymph node metastasis. ELISA confirmed up-regulation of ECM proteins in metastatic lymph nodes. ECM proteins have requisite parameters to be developed as a diagnostic tool to assess status of sentinel lymph nodes to guide surgical intervention in early breast cancer.

Breast cancer is a common malignant disease among women population worldwide and is characterized by the highest cancer incidence, high recurrence rate, morbidity, mortality and poor prognosis¹. Globally breast cancer accounted for 11.7% of the estimated 19.3 million new cancer cases in 2020 in both the sexes and about 24.5% of all cancer cases among women². The International agency for cancer research has reported that breast cancer ranks number one cancer worldwide with an incidence of 22.5 lakh and a mortality rate of nearly seven lakh³. The latest cancer registry of Indian Council of Medical Research has reported that breast cancer ranks number one cancer in India with an incidence rate of 25.8 per lakh, and a mortality rate of 12.7 per lakh⁴. The breast cancer incidence cases are expected to increase by more than 46% by 2040⁵. As per the cancer death cause analysis, it is not the primary breast tumor but metastasis to distant organs that is responsible for the death of over 90% of breast cancer patients⁶. Breast cancer accounts for 11.6% of distant malignancies in both men and women globally⁷.

¹Department of Biophysics, All India Institute of Medical Sciences, Ansari nagar, New Delhi 110029, India. ²Proteomics Lab, Centralized Core Research Facility, All India Institute of Medical Sciences, New Delhi 110029, India. ³Department of Surgery, All India Institute of Medical Sciences, Ansari nagar, New Delhi 110029, India. ⁴Department of Pathology, All India Institute of Medical Sciences, Ansari nagar, New Delhi 110029, India. ⁵Department of Radiology, All India Institute of Medical Sciences, Ansari nagar, New Delhi 110029, India. ⁶Department of Obstetrics and Gynaecology, All India Institute of Medical Sciences, Ansari nagar, New Delhi 110029, India. ✉email: g.hariprasad@rediffmail.com

The TNM (tumor, node, metastasis) classification is a globally recognised standard to assign the stages for breast cancer. Stage 0 the non-invasive; stage I, IIA, IIB, IIIA describes early breast cancer which is confined to the breast with or without axillary lymph node involvement; stage IIIB, IIIC describes locally advanced and, stage IV describes breast cancer that has spread to other parts of the body like liver, lungs, brain and bones termed as metastatic breast cancer^{8,9}.

Axillary lymph nodes represent the main stay of lymphatic drainage from the breast, and the first lymph node among them to be affected is called the sentinel lymph node¹⁰. Sentinel lymph node biopsy (SLNB) is the standard diagnostic procedure for patients with early-stage breast cancer to assess the status of lymph node¹¹. Identification of this node is made by injecting radioactive tracer into the breast in the vicinity of the tumour, followed by its excision, and a biopsy that reflects the histological characteristics and status of nodes in the axilla^{12,13}. Surgeons carry out axillary lymph node dissection (ALND) in breast cancer for cases where SLNB is positive and abandon this intervention if it is negative^{14,15}. Identification and assessment of sentinel lymph node metastasis dictates the right surgical approach adopted and the prognosis in patients with early breast cancer¹⁶.

Recently, potential biomarkers have been identified to understand lymph node metastasis in early breast cancer^{17–20}. Unfortunately, each of these have poor sensitivity and specificity, and has therefore not been translated into a diagnostic tool as desired by the surgeon. A number of imaging techniques too have been used to pick up axillary lymph node metastasis, but these techniques are neither efficient, nor applicable during the surgical procedures^{21–26}.

Our team has been identifying potential protein biomarkers in various clinical phenotypes in the last few years^{27,28}. In the recent past, we have carried out a gel based proteomics experiment to identify biomarkers that can flag lymph node metastasis in early breast cancer²⁹. In this study, we seek to carry out an isobaric label based proteomic experiment to delineate protein signatures that can accurately reflect the metastatic state of the sentinel lymph nodes in early breast cancer.

Methodology

Ethics and patient recruitment. This study was conducted after approval was obtained from the Institute Ethics Committee at All India Institute of Medical Sciences New Delhi, India (Ref. No. IECPG- 27/23.01.2019, RT-03/28.02.2019). The procedures were followed as per the ethical standards formulated in the Helinski declaration. Early breast Cancer patients were screened and admitted at the Department of Surgery. Detailed information on the study was explained to the recruited patients and informed consent was obtained from them. Bedside examination included clinical history, symptoms, signs and general examination. Patients suspected of having early breast cancer were subjected to mammography imaging. As explained later in this section, sentinel lymph node biopsy tissues from these patients were sent to Department of Pathology for histopathology, which was the gold standard for assigning the clinical phenotypes of sentinel lymph node metastasis (SLNM+) and sentinel lymph node without metastasis (SLNM-). For the discovery phase of the proteomic experiment, we took 5 patients with SLNB+, 5 patients with SLNB- and, two benign breast tumor tissues as cancer controls. For the validation phase of experiment, we took 13 patients each with SLNM+ and SLNM-.

Patient inclusion and exclusion criteria. Staging of the breast cancer were determined as per to the American Joint Committee on Cancer (AJCC) cancer staging criteria. Inclusion and exclusion criteria were used just as in our previous study²⁹. Women with early invasive intra-ductal breast cancer as per the WHO classification of the tumor, and who had not undergone any therapeutic intervention were recruited into the study. Patients with advanced breast cancer, and patients with early breast cancer who had either received chemotherapy or radiotherapy were excluded from the study.

Identification and excision of sentinel lymph node. Technetium tagged sulphur colloid was injected intra-dermally into the lower inner quadrant of the affected breast, two hours prior to the surgery²⁹. In operation theatre, 1 ml of 1% fluorescent methylene blue dye diluted in 4 ml saline was injected at multiple sites intra-dermally around areolar region and in sub-areolar region. After five minutes of gentle massage, an incision was made on axillary skin crease at the site of maximum radioactivity. By using blunt and sharp dissection, methylene fluorescent lymphatics was identified using a blue light lamp, and blue lymphatics was identified by direct visualization³⁰. Lymph node having highest count was considered as sentinel node and was excised.

Sentinel lymph node tissue sample collection. After the excision of sentinel lymph node, adherent fat tissues were neatly removed, and blood stains were washed thoroughly with 1X PBS (pH 7.4)²⁹. The nodes were longitudinally sectioned to obtain 5 mm thick slices. One set of alternate slices was sent to the Department of Pathology for histopathological assessment. The other set of slices were taken to proteomics facility and stored at – 80 °C for iTRAQ based proteomic experiments.

Histopathology. Histopathological procedures that were followed were those standardized in our previous study²⁹. Lymph node slices were fixed into formalin fixed paraffin embedded blocks. These were further sectioned into 4 µm poly-L-lysine-coated slides. These paraffin sections were deparaffinised with three subsequent washes in xylene and then rehydrated by washing them stepwise in 100% ethanol, 90% ethanol, 70% ethanol and distilled water. The sections were stained with hematoxylin and washed in running water for 5 min. The slides were then stained in eosin solution for two minutes and then rinsed with 95% ethanol. The slides were then subjected to 100% ethanol for two minutes, twice. After final exposure to xylene, a drop of Distyrene Plasticizer Xylene (DPX) was used to mount the tissue on each slide and covered with a glass cover-slip. The slides

were examined under the microscope and sentinel lymph node tissue samples were annotated as either Sentinel Lymph Node Metastatic (SLNM+), or Sentinel Lymph Node Non-Metastatic (SLNM-).

Mammography. All recruited patients in the study underwent full-field digital mammography in cranio-caudal projection and medio-lateral oblique projection. The effective dose of a four view mammogram ranged from 4 to 6 mega gray. The evaluation of mammogram was done according to the Breast Imaging Reporting and Data System classification (BIRADS).

Sample phenotyping and protein isolation. Phenotyping and protein isolation was done using protocols standardized in our previous study²⁹. The sentinel lymph node tissue sections which were stored at -80°C were annotated as either SLNM+ or SLNM- based on histopathology finding. The tissue samples were minced and the proteins were solubilized in 120 μl of lysis buffer that contained 8 M urea, 2 M thiourea and 4% 3-[(3-Cholamidopropyl)dimethylammonio]1-propanesulfonate (CHAPS). The tissue was homogenized by sonication at an interval of 3 s and vortexed for 2 min. The samples were then centrifuged at 15,000 rpm for 20 min at 4°C , debris was discarded and the supernatant was transferred onto a fresh eppendorf tube. Protein extracted with lysis solution was buffer exchanged with 100 mM Triethylammonium bicarbonate (TEAB) using a 3 kDa cut off membrane filters to bring down the concentration of urea well below 0.1 M. Protein amount was quantified using the Bradford assay using 1 $\mu\text{g}/\mu\text{l}$ of Bovine serum albumin as a standard.

Isobaric tags for relative and absolute quantitation (iTRAQ) labelling. Five sets of SLNM+ tissue samples, five sets of SLNM- tissue samples and two cancer control benign breast tumor tissues were taken for iTRAQ experiment. The design of iTRAQ experiment is illustrated in Fig. 1. Each experiment composed of at least one SLNM+, one SLNM- and, one of either of these two phenotypes or a cancer control benign breast tumor tissue. The equimolar culmination of the three phenotypes was made into an internal standard for the sake of normalization for each of the four experiments²⁷. 80 μg of protein from each phenotype sample was reduced with 25 mM DTT for 30 min at 60°C and alkylated with 55 mM iodoacetamide for 20 min at room temperature. Each of these proteins samples were digested for 16 h with trypsin in 1:10 ratio at 37°C . Digested peptides were then labelled with iTRAQ 4-plex reagents, 114 (sentinel lymph node metastasis), 115 (sentinel lymph node without metastasis), 116 (one of either of these two phenotypes or cancer control benign breast tumor tissues), and 117 (internal standard) following the protocol provided by manufacturer (AB Sciex, Foster city, CA, USA). In brief, all vials of iTRAQ labelling tags were reconstituted in 70 μl of absolute ethanol. This was then added to each sample and incubated for 2 h at room temperature, and the reaction was quenched using 50 μl mili-Q water. iTRAQ labelled samples in each experiment were then pooled into a single vial and dried using speedvac. These samples were reconstituted in 8 mM ammonium formate buffer (pH: 3) and were fractionated by cation exchange using isotope coded affinity tag cartridge. Peptides were then eluted with 500 μl of gradient elution with 5 mM to 500 mM concentrations range of ammonium formate (pH: 3) to obtain a total of eleven fractions from each experiment. These 44 fractions from the four experiments were vacuum dried and taken for mass spectrometry analysis.

Mass spectrometry data acquisition. In-house protocols were used for Mass spectrometry data acquisition²⁷. Peptides from the 44 fractions were desalted and concentrated using reversed phase ZipTip, and reconstituted in 0.1% formic acid. The peptide fractions were loaded onto analytical column (Acclaim PepMap RSLC C18, 2 μm , 100 \AA , 50 μm x 15 cm; Thermo Scientific, Rockford, USA) associated with trap column (Acclaim PepMap C18, 3 μm , 100 \AA , 75 μm x 2 cm; Thermo Scientific, Rockford, USA). The peptide separation were performed using EASY-nLC 1200 which was coupled with Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Scientific, Rockford, USA) for Mass spectrometry analysis. The peptide fractions were premixed in loading buffer (Mobile phase A: 100% water and 0.1% formic acid) and 1 μg were loaded on a trap column with a flow rate of 300 nl/min. The retained peptides were washed iso-cractically by loading buffer for 45 min to remove excess salt. The peptides were then resolved on an analytical column with a multi step linear gradient of loading buffer and elution buffer (mobile phase B: 80% acetonitrile and 0.1% formic acid) at a flow rate of 250 nl/min. The gradient elution were initiated using 5% elution buffer and were held for 1 min, with linear increase rate of 10% for 10 min, 35% for 70 min and 50% for 80 min. The gradient elutions were held at 80% mobile phase B for 8 min before being re-equilibrated to 5% mobile phase B for 18 min. The mass spectrometer was operated in data dependent acquisition (DDA) mode. The full MS spectra was acquired in positive ion mode in m/z ratio of 350–2000 Da, with a 100 milli second MS accumulation time, whereas the MS/MS product ion scans were performed in the mass range of 100–2000 Da with a 7 milli second accumulation time in Orbitrap mass analyzer. The mass spectrometric setting included Ion spray voltage floating of 1900 V. For 3 s former target was excluded and 25 ions were monitored per MS cycle. DDA advanced ‘rolling collision energy’ were applied for subsequent MS/MS scans with normalized High energy Collision induced Dissociation (HCD) collision energy set to 35%.

Data analysis and reporter ion quantification. Raw files from the Orbitrap Fusion Mass Spectrometer were processed using Proteome Discoverer (version 2.4.1.15) analysis software³¹. Both MS and MS/MS spectra were searched using Sequest HT algorithm against a combined UniProt Human proteome database appended to a list of common contaminants provided by Thermo Scientific. Sequest HT parameters were specified as trypsin enzyme, two missed cleavages allowed, minimum peptide length of 6, precursor mass tolerance of 10 ppm and a fragment mass tolerance of 0.05 Daltons. The static modification was set to carbamidomethylation (+ 57.021 Da) of cysteine. The dynamic modifications on peptide terminus were set to methionine oxidation (+ 15.995 Da) and iTRAQ 4-plex (+ 144.102 Da) modification, on N-terminus and Lysine (K) residues. Since iTRAQ modifica-

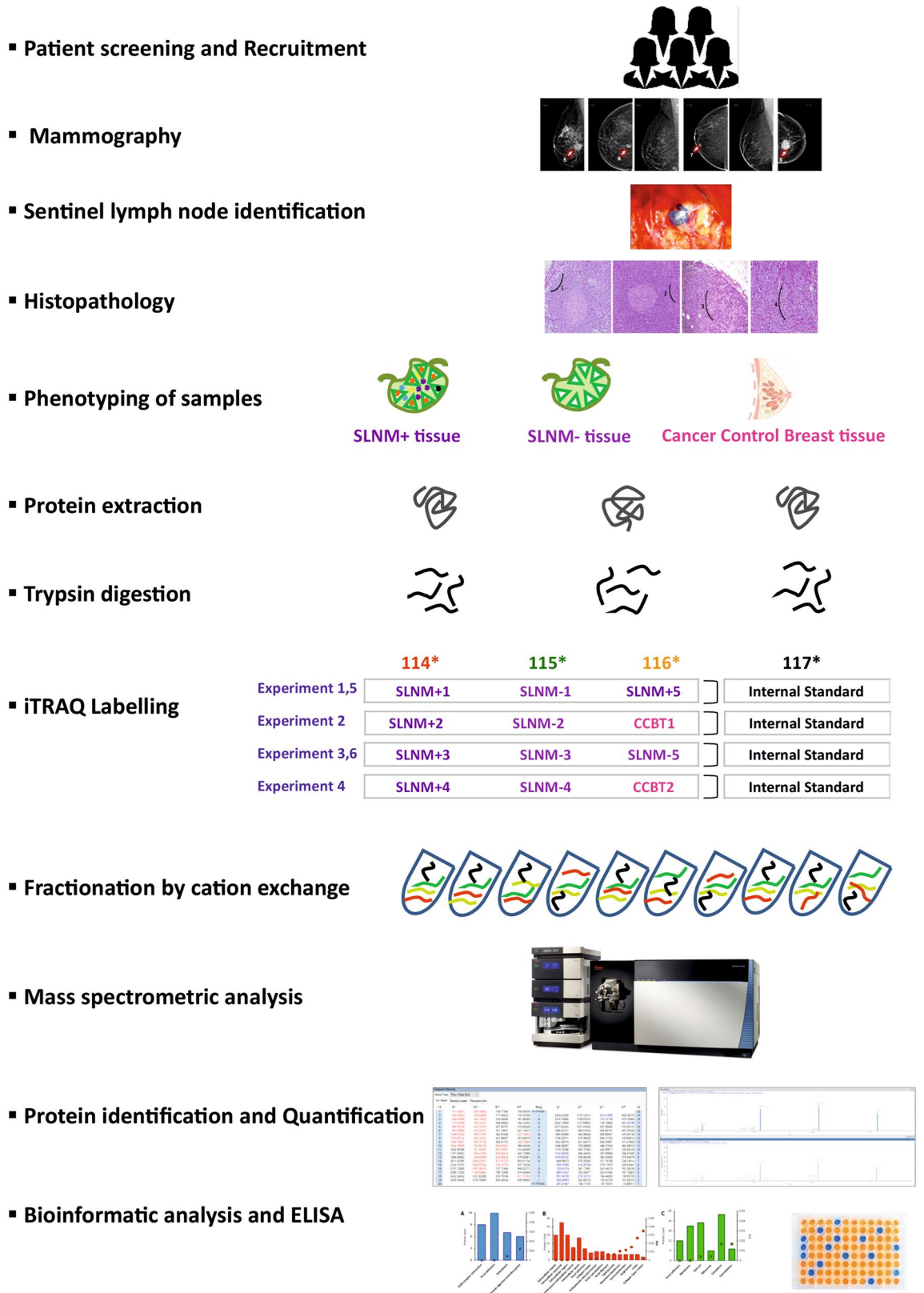


Figure 1. Flowchart depicting the methodology used in the study.

tion was used as dynamic modification, the unmodified or unlabelled peptides and associated proteins were removed from the analysis. Also, dynamic modification was assigned for acetylation (+42.011 Da) of protein's

N-terminus. Peptide spectral match (PSM) error rates were determined using the target-decoy strategy coupled to Percolator PSM validation node to trigger the positive and false matches. In Percolator node, the false discovery rate (FDR) was calculated based on the q -values of Decoy database search. Data were filtered at the peptide spectral match-level using a strict FDR cut off of 0.01 and relaxed FDR cut off of 0.05 as determined by Percolator. Contaminant and decoy proteins were removed from all data sets prior to downstream analysis. Reporter ion values were calculated on “Reporter Ions Quantifier” node using FTMS mass analyzer setting and HCD activation process. Reporter ions were quantified from MS/MS scans using an integration tolerance of 20 ppm with the most confident centroid setting. The following settings were used to obtain the quantification results: the protein ratio type was the ‘weighted’ geometric mean, normalization with summed intensities and outlier removal was ‘automatic’. The peptide threshold was ‘at least homology’ where peptide score does not exceed absolute threshold but is an outlier from the quasi-normal distribution of random scores. Minimum of two unique peptides were required to be the top ranking matches. In Consensus workflow in “Reporter Ion Quantifier” node, the following settings were applied to increase the quantification accuracy of the analyzed proteins: (i) only unique peptides were used for protein quantification, (ii) precursor co-isolation threshold was considered 25%, (iii) average reporter signal to noise ratio threshold was considered 10 and, (iv) peptide normalization was done with respect to the total peptide amount. At the level of protein analysis, further normalization was done where the protein abundance of individual sample was scaled by the abundance of the internal standard which was labelled with channel 117 of iTRAQ 4-plex reagent. On obtaining the results, multiple filter criteria were applied and only those proteins were considered for differential expression analysis which had: (i) FDR confidence threshold as medium during identification by Sequest-HT, (ii) presence of at least two unique peptides and, (iii) peak found at found in sample.

Expression fold change ratio of ≥ 1.5 and ≤ 0.66 were considered as up-regulated and down-regulated proteins. Proteins with fold change ratio between 1.5 and 0.66 were considered as house-keeping proteins. A multiple students t -test was applied to the whole set of differentially expressed proteins and a volcano plot with a p -value < 0.05 was generated to graphically represent the up-regulated, down-regulated and house-keeping proteins. Those proteins that had a consistent expression pattern in at least four of the six experiments were considered to be potential biomarker candidates to differentiate SLNM+ and SLNM-. The relative ratios of protein abundances of only the up-regulated proteins was compared between metastatic group and cancer control breast tissues was used to estimate the possible tissue source.

KEGG pathway and Gene Ontology analysis. Differentially expressed genes were imported on DAVID³² (Database for Annotation, Visualization and Integrated Discovery, version 6.8, <https://david.ncicfcrf.gov/tools.jsp>) functional annotation tool and Functional Enrichment analysis was done. Homo Sapiens was used as background species and the enrichment analysis was run for Cellular Component in Gene Ontology (GO_CC) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathways. Only those results that had FDR adjusted p -values ≤ 0.05 were considered.

ELISA. Twelve differentially expressed proteins: α -crystallin B chain, monoamine oxidase, caveolin-1, collagen α -1, desmin, fibrillin-1, long-chain-fatty-acid-coA ligase 1, laminin subunit α -4, heterogeneous nuclear ribonucleoprotein D, non-histone chromosomal protein, cathelicidin antimicrobial peptide, rho GDP-dissociation inhibitor 2 were chosen for validation phase of the experiment. The protein concentrations of these proteins were quantified in validation set of 13 SLNM+ and 12SLNM- patients using ELISA according to the manufacturer’s instructions (Bioassay Technology Laboratory, Shanghai, China). All determinations were performed in duplicates according to the manufacturer’s recommendations. Differences between SLNM+ and SLNM- groups of patients were calculated using independent Student t -test; values of $p < 0.05$ were considered significant.

Statistical analysis. Normalization of the proteins was done using MetaboAnalyst (version 5.0) software (<https://www.metaboanalyst.ca/MetaboAnalyst/ModuleView.xhtml>)³³ using sum of protein abundances. Data transformation using generalized logirithm and has been scaled using Pareto scaling option. Data analysis for ELISA was carried out using STATA version 16.0 version. Protein concentrations that could be used as cut-off to distinguish between metastatic state and non-metastatic state, were estimated based on Receiver Operating Characteristics (ROC) analysis of the ELISA data. The non-parameteric ROC analysis was carried out using DeLong method. Area under the curve (AUC) was obtained with 95% confidence limits. Optimum cut-off value was obtained at which Yuden Index (sensitivity + specificity-1) was maximum. Percentage correct classification and likelihood ratio values were computed. Statistical significance level $P < 0.05$ was adopted to test the significance of the AUC.

Ethics approval and consent to participate. This study was conducted after approval was obtained from the Institute Ethics Committee at All India Institute of Medical Sciences New Delhi, India (Ref. No. IECPG- 27/23.01.2019, RT-03/28.02.2019).

Result

Clinical profile. Based on the inclusion and exclusion criteria a total of twenty eight patients with early breast cancer were recruited to this study. From among these, thirteen patients were those who had sentinel lymph node metastasis; thirteen patients were those who did not have sentinel lymph node metastasis, and two patients who were diagnosed with fibroadenoma were recruited to procure breast tissue that would serve as cancer controls. From among these, five with SLNM+, five with SLNM-, and the two cancer control benign breast tumor were chosen for the discovery phase of proteomic experiments by iTRAQ. The clinical profile of

S. No	Patient id	Age (yrs)	Mammography	Size(cm)	ER/PR/Her2	Histopathology	Phenotype
1	KD42	42	Early breast cancer	2.5 × 2 × 1.8	+, +, -	Complete effacement of architecture	SLNM +
2	UG51	51	Early breast cancer	5.2 × 1.8 × 1.4	+, -, +	Infiltration by a metastatic cancer cells	SLNM +
3	PK47	47	Early breast cancer	3.5 × 2 × 1	+, -, -	Tumor cell nests in the sub-capsular area	SLNM +
4	KM50	50	Early breast cancer	2.5 × 1.5 × 1.5	+, +, -	Tumor cell nest replacing the normal lymph node parenchyma	SLNM +
5	SP52	52	Early breast cancer	3 × 2.5 × 2	+, +, -	Tumor cell invading into stroma	SLNM +
6	PD54	54	Early breast cancer	3.5 × 3.1 × 2.5	-, -, -	Reactive lymphadenopathy without any evidence of tumor cells	SLNM -
7	MK54	54	Early breast cancer	1.0 × 0.8 × 0.5	+, -, -	Reactive lymphadenopathy without any evidence of tumor cells	SLNM -
8	LW52	52	Early breast cancer	3.3 × 2.5 × 1.5	+, -, +	Reactive lymphadenopathy without any evidence of tumor cells	SLNM -
9	SM59	59	Early breast cancer	3 × 3 × 2.5	-, -, -	Reactive lymphadenopathy without any evidence of tumor cells	SLNM -
10	CD	47	Early breast cancer	1.5 × 1.5 × 5	+, -, -	Reactive lymphadenopathy without any evidence of tumor cells	SLNM -
11	SC36	36	Well circumscribed mass	4.3 × 3 × 2.5	-, -, -	Terminal duct lobular units with bilayered epithelium and a slant interveniumstroma	Benign breast tumor
12	SS59	59	Fat containing mass	3 × 3 × 2.5	+, +, -	Terminal duct lobular units with bilayered epithelium and a slant interveniumstroma	Benign breast tumor

Table 1. Clinical profile of patients recruited in the discovery phase of the iTRAQ experiment.

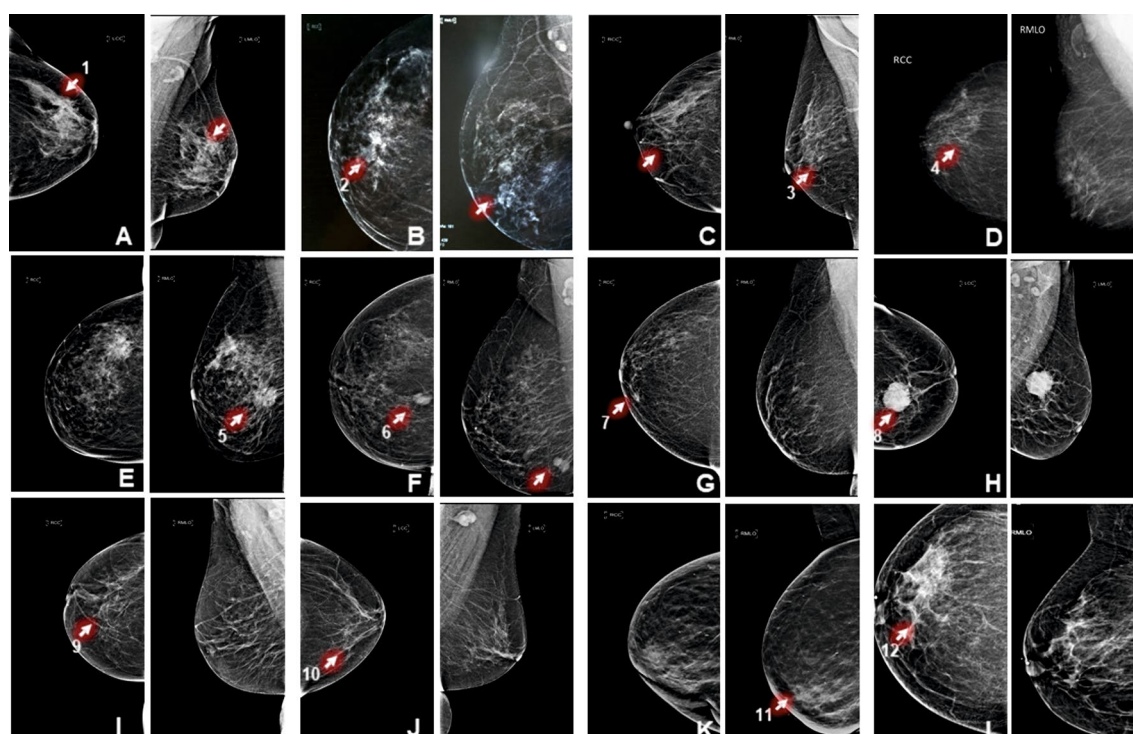


Figure 2. Medio-lateral oblique and cranio-caudal mammography images of patients recruited in the discovery phase of the study. (A)–(J) shows features of early invasive breast cancer and K–L shows features of benign breast tumor. (1) focal architectural disruption; (2) speculated mass; (3) micro-calcification; (4) micro-calcification along with subtle focal asymmetry; (5) multiple speculated masses; (6) macro-lobulated mass; (7) subtle area of focal architectural distortion; (8) circumscribed mass with focal speculation; (9) architectural distortion; (10) unilateral axillary lymphadenopathy; (11) well circumscribed masses; (12) fat containing masses.

these twelve patients is provided in Table 1. All patients with early breast cancer were confirmed by mammography and sentinel lymph node tissue samples annotation either with sentinel lymph node metastasis or without metastasis were done using histopathology evaluation. The mammography images of patients who were considered for the discovery phase of the study are shown in Fig. 2. The assessment of mammogram was performed according to scores of BIRADS (Breast Imaging Reporting and Data System) classification. The hematoxylin

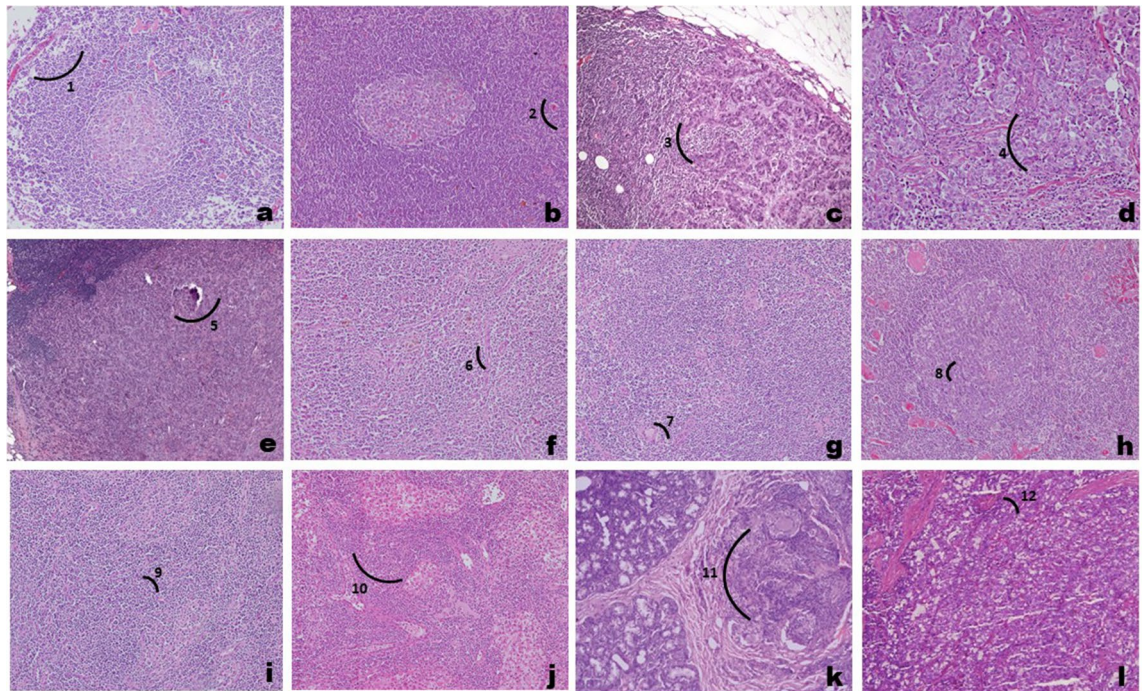


Figure 3. Photomicrographs showing sentinel lymph node with metastasis (a–e), sentinel lymph node without metastasis (f–j), and benign breast tumor tissue (k,l) in hematoxylin and eosin stained sections under light microscopy. (1) complete effacement of architecture by a metastatic cancer cells; (2) infiltration by a metastatic cancer cells; (3) presence of tumor cell nests in the sub-capsular area with a part of the lymph node seen at one edge; (4) presence of large area of tumor cell nest replacing the normal lymph node parenchyma; (5) complete effacement of architecture by metastatic cancer cells. (6–10) shows prominent reactive centre without any evidence of tumor cells in the lymph node parenchyma and depicts paracortical expansion with vascular proliferation indicative of reactive lymphadenopathy. (11–12) depicts benign breast lesion consisting of terminal duct lobular units with bi-layered epithelium and a slant inter venium stroma.

and eosin stained slides for the five metastatic, five non-metastatic sentinel lymph nodes and two cancer control benign breast tumor tissues are shown in Fig. 3.

iTRAQ labelling. In our study, proteins isolated from metastatic and non-metastatic sentinel lymph node tissues, and cancer control benign breast tumor tissues were labelled with isobaric tags 114, 115 and 116. Equimolar protein concentrations of proteins from each representative phenotype in the experiment were labelled with tag 117 for the sake of normalization. Out of 276,572 Peptide Spectrum Matches (PSMs), 269,279 (97.4%) are iTRAQ modified and out of a total of 42,105 peptides, 40,074 peptides (95.2%) were found to be iTRAQ labelled reflective of the labelling efficiency.

Protein expression. From the 44 fractionated iTRAQ four plex labelled samples, 6335 unique protein groups were identified and after applying stringent filter criteria 2398 proteins were taken up for further analysis. Distribution of these proteins is shown by a normal curve indicative of the quality of the normalization (Supplementary Fig. 1). This process has adjusted for differences among different samples, data transformation and scaling to make individual protein expressions comparable across metastatic and non-metastatic lymph node groups. A pair-wise multiple Student's *t*-test that incorporates p-value was used to arrive at 109 differentially expressed proteins between SLNM+ and SLNM- of which 49 proteins are up-regulated 60 proteins are down-regulated as shown in volcano plot (Fig. 4). The relative abundance ratios of only the up-regulated proteins were compared between metastatic group and cancer control breast tissues. Proteins such as desmin, fibrillin 1, tau-tubulin kinase, transgelin, calponin 1 and myosin 11 that had a ratio of more than one are ones that are native to the lymph node, and proteins such as heat shock protein 6, α -crystalline B, amine oxidase 3, caveolin 1, collagen α 1, fibrinogen gamma chain, GAPDH, long chain fatty acid Co-A ligase, laminin subunit α 4, membrane primary oxidase, microfibril-associated glycoprotein 4, perilipin 1 and redox regulatory protein that had a ratio of less than one are derived from breast tissue. Based on the functional annotation and their relevance in this study few of these proteins are discussed in Tables 2 and 3.

Validation by bioinformatic analysis and ELISA. KEGG pathway analysis highlights the role of the differentially expressed proteins in various pathways (Fig. 5a). The most interesting feature is the ECM-receptor interaction that implicates seven proteins that include various isoforms of collagen- α . Apart from ECM-receptor interaction pathways, focal adhesion and PI3K-Akt signalling feature prominently in the pathway analysis. Gene

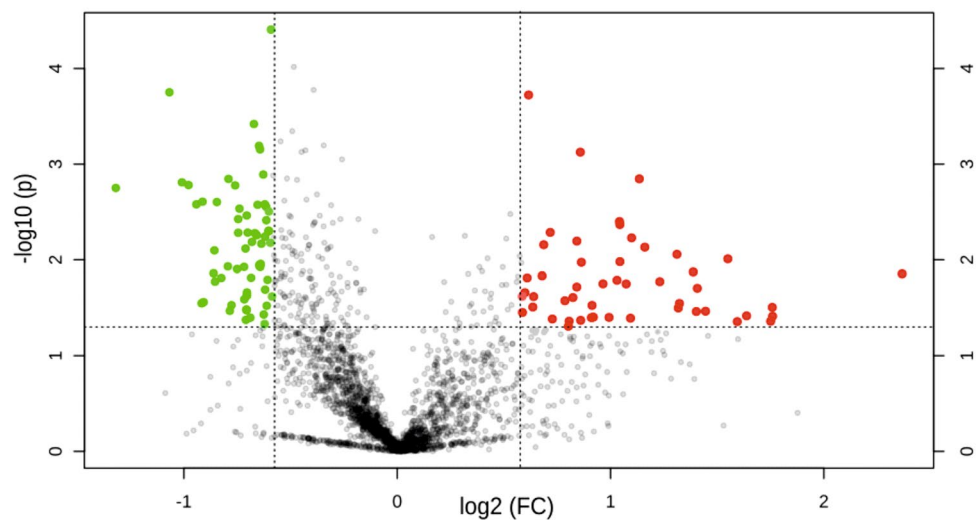


Figure 4. Volcano plot of proteins identified in iTRAQ experiment. Red dots represent up-regulated proteins with >1.5 fold change, green dots represent down-regulated proteins with <0.66 fold change. House-keeping proteins with expression fold between 1.5 and 0.66 is shown as grey dots. FC indicates Fold Change.

Ontology for Cellular Component (GO_CC) enrichment that was carried for the 49 upregulated proteins shows that majority of the proteins belong to extra cellular component (Fig. 5b). The next component with highest number of proteins was focal adhesion component. Focal adhesions are large macromolecular assemblies through which mechanical force and regulatory signals are transmitted between the extracellular matrix and an interacting cell. Implication of extracellular proteins is therefore quite evident in breast cancer metastasis. Gene Ontology for Cellular Component (GO_CC) enrichment that was carried for the 61 down-regulated proteins shows majority of the proteins being confined to cytoplasmic, cytosol and membrane component (Fig. 5c). Based on the results of bioinformatic studies, ELISA was performed on upregulated extracellular proteins in SLNM+. Caveolin 1, Desmin, Microfibrillar associated glycoprotein, Collagen $\alpha 4$ and Fibrillin 1 were confirmed to be elevated in SLNM+ as compared to SLNM- (Fig. 6) (Supplementary Table 1). These proteins have a minimum of two fold higher expression in SLNM+ as compared to SLNM-.

Discussion

The predominant age group of early breast cancer patient recruited in our study was 36–59 years. Most of the patients presented with lump, nipple discharge, nipple retraction, pain in one part of the breast. Progesterone, estrogen and HER-2 status did not have any correlation with the lymph node status of early breast cancer. Patients clinically diagnosed as early breast cancer were subjected for mammography screening to confirm the type of breast cancer. Twenty five patients were diagnosed with early breast cancer on the basis of their characteristic features like abnormal masses and collection of calcification. As normal control tissues could not be procured due to ethical concerns, Two fibroadenoma tissues that were closest representations of ‘cancer control’ breast tissues were used to know the source of the protein. Methylene blue was used to map the sentinel lymph node in axilla region of breast, followed by its excision. Histopathological analysis was done to annotate the tissue phenotype either SLNM+ or SLNM- or benign tumor.

Proteins from sentinel lymph node tissues were isolated, quantified, digested with trypsin and labelled with different isobaric tags combined into one sample mixture for identification and quantification by LC-MS/MS analysis. Isobaric tags for relative and absolute quantitation (iTRAQ) technology relies on the quantitation of low molecular mass reporter ion groups released from isobaric tags that are covalently bind to primary amines of tryptic peptides that need to be quantified via amine labelling. The final experiment so designed was to enable: (i) six phenotypic protein profile comparisons, (ii) intra-experimental normalization and (iii) understand the cellular source of the protein. While the Up-regulated proteins in SLNM+, are related to tumorigenesis, cell proliferation, motility, cell survival, progression and anti-apoptosis, the up-regulated proteins in SLNM- are involved in cell motility suppression and influencing decreased cell growth. Expression of five proteins caveolin-1, collagen $\alpha 1$, desmin, fibrillin-1, and microfibrillar associated glycoprotein 4 were validated and were found to be consistent with the discovery phase results. The functions of these proteins in the context of understanding sentinel lymph node metastasis are: (a) Caveolin-1 (Cav-1), a 22 kDa small oligomeric scaffolding protein encoded by CAV1 gene is a major structural protein of membranes called caveolae and plays very crucial role in many cellular processes, including endocytosis, receptor internalization, ECM organization, lipid transport, signal transduction^{86,87}; (b) Collagen $\alpha 1$ (COL1A1), a 138 kDa protein encoded by the COL1A1 gene is a most abundant protein of extracellular matrix forms a characteristic triple helix structure of three polypeptide chains, and contributes to the integrity, elasticity and strength of body's connective tissues, entrapment, local storage and delivery of growth factors and cytokines and therefore plays an important role during organ development, wound healing and tissue repair^{88,89}; (c) Desmin (DES), a 53 kDa protein encoded by DES gene is a muscle-specific protein and

S.No	Protein	Accession no.	Avg. ratio of SLNM +/SLNM-	Functions	Relevance in our study	References
1	Alpha-crystallin B chain	E9PS12	1.8	Inhibits the autoproteolytic cleavage of caspase-3, interacts with the pro-apoptotic proteins and prevents translocation to mitochondria	Promotes anti-apoptosis	34,35
				Inhibits P53 dependent apoptosis by inhibiting Ras activation, blocks UVA cell apoptosis and binds to X-linked inhibitor of apoptosis to inhibit caspase	Promotes anti-apoptosis	36
				Associated with transcriptional activation, regulates the Vascular Endothelial Growth Factor (VEGF) and confers VEGF resistance; Induces EGF and anchorage-independent growth through activation of the MEK/ERK pathway	Promotes tumorigenesis	37
2	Amine oxidase 3	K7EQZ5	3.4	Regulates myofibroblastic phenotype in a cancer-associated fibroblast	Promotes metastasis	38,39
3	Calponin 1	K7ENC5	3.0	Induces cytoskeleton changes that leads to EM	Promotes metastasis and invasion of tumor cell	40–44
4	Caveolin 1	C9JKI3	2.2	Regulates intrinsic, extrinsic apoptotic pathway proteins and downstream proteins; Cav-1 induced autophagy alteration influenced apoptosis	Promotes anti-apoptosis	45,46
				Influences the levels of EMT (Epithelial to mesenchymal transition) associated markers, transcription factors, MMP and RhoGTPases; Participates in Cav1 dependent trafficking of integrins	Promotes invasion and migration	47–49
				Endows anoikis resistance by inactivating caspase-8 and regulates the levels of metastasis-associated proteins like MT4-MMP, MMP9 and TLR4	Promotes metastasis	50,51
5	Collagen alpha-1	P02461	5.6	Induces mRNA level of MMP9 and regulates the target gene of Wnt/ β catenin pathways	Promotes invasion	52,53
6	Desmin	P17661	2.7	Activates Tumor stromal myofibroblasts and transforming growth factor (TGF) β that switches non-invasive to invasive cancer	Promotes angiogenesis	54
7	Fibrillin 1	P35555	3.8	Associates with TGF β signalling which induces EMT and in turn regulates the expression of E-cadherin, beta-catenin and MMPs	Promotes growth, proliferation and invasion	55–57
8	Glyceraldehyde-3-phosphate dehydrogenase	P21695	3.1	Interacts with Sp1 and enhances the expression of SNAIL, a transcriptional inducer of EMT	Promotes invasion and metastasis	58
9	Heat shock protein gene 6	K7EP04	2.2	Chaperone focal adhesion kinase, integrin linked kinase and the receptor tyrosine kinases ErbB2 and MET	Promotes metastasis	59,60
				Induces epithelial-mesenchymal transition, in which cells switch from a compact shape to a spindle shape	Increases cell motility	59,61
10	Laminin subunit alpha 4	Q16363	2.1	Involved in mitogen-activated protein kinases (MAPK) cascades and dual-specificity phosphatases in laminin signaling in human melanoma cells	Promotes tumor progression, tumor invasion, metastasis	62,63
11	Membrane primary amine oxidase	Q16853	3.3	MAOA stabilizes HIF1 α , activates the VEGF-A/NRP1 system, and induces expression of TWIST1, an EMT master transcription factor that participates in EMT promotion	Promotes migration, proliferation and invasion	64
12	Microfibril-associated glycoprotein 4	K7ES70	2.6	Inhibits Notch1 signaling and interacts with Jagged1	Promotes angiogenesis	65
				Interacts with α v β 3 integrin receptor	Promotes tumor cell motility and survival	66

Continued

S.No	Protein	Accession no.	Avg. ratio of SLNM +/SLNM-	Functions	Relevance in our study	References
13	Myosin-11	P35749	3.0	Link to actin filaments	Promotes breast cancer motility	67
				Response to impaired p53, cell adhesion inhibition, protrusion formation	Breast tumor progression	68
14	Perilipin-1	O60240	3.0	Regulates lipid network of fatty acid synthase (FASN) and sterol regulatory element-binding protein (SREBP) family	Promotes angiogenesis and metastasis in breast cancer	69
15	Redox-regulatory protein FAM213A	Q9BRX8	1.8	miR-211 and FAM213A represses TCF12(transcription factor 12)	Promotes oncogenesis	70
16	Transgelin	K7ERK4	3	Associates with TGFβ/Smad-3	Promotes invasion	71-73
				Regulates the production of MMP2	Promotes metastasis	72,73

Table 2. Proteins up-regulated in tissue of sentinel lymph node with metastasis as compared to sentinel lymph node without metastasis.

S. No	Proteins identified	Accession ID	Avg Ratio SLNM +/SLNM-	Functions	Relevance in the study	References
1	Coagulation factor XIII A chain	H0Y796	-1.5	Coagulation factor Xa inhibits cancer cell migration in a PAR-1 dependent manner	Inhibition of cancer cell migration	74
				Causes activation of LIMK1/2 which inactivatescofilin and leads to inhibition of actinide polymerization	Suppresses cancer motility	74
2	Heterogeneous nuclear ribonucleoprotein D	D6RBQ9	-4.0	Inhibits STAT3 and MMP2 via WNT/ TCF4 signalling pathways	Inhibits breast metastasis	75-77
				Increases the expression of E-cadherin and down-regulates Twist1 and Snail signalling pathways	Inhibits breast cancer progression	78
3	Non-histone chromosomal protein HMG-14	A6NEL0	-2.0	Transcriptional repression	Acts as tumor-suppressor gene	79
4	Histone H1.3	P16402	-2.1	Inhibits <i>H19</i> expression (non coding oncogene) by preferential occupancy at the ICR (imprinting control region) of H19 and regulating DNA methylation at this region	Decreases the growth rate, colony formation in cancer cell and suppression of epithelial carcinogenesis	80
5	Cathelicidin antimicrobial peptide	P49913	-4.7	Inhibit tumor-associated fibroblasts (TAFs) through suppression of epithelial-mesenchymal transition (EMT) and disruption of the cytoskeleton	Suppresses cancer cell proliferation	81
6	High mobility group nucleosome-binding domain-containing protein4	O00479	-2.5	HMG along with (miRNA) down-regulates SOX4 expression	Inhibits metastasis	82
7	Rho GDP-dissociation inhibitor 2	F5H6Q0	-1.6	GDI2 inhibits GTPase function by binding to Rac 1	Inhibition of metastasis and cancer growth	83
8	Thymosin beta-4	P62328	-1.7	Down-regulates STAT3-MMP2 signaling	Suppresses migration and invasion	84,85

Table 3. Proteins down-regulated in tissue of sentinel lymph node metastasis as compared to sentinel lymph node without metastasis.

a key subunit of the intermediate filament in cardiac, skeletal and smooth muscles, and plays an essential role in maintaining extracellular matrix interactions, cytoarchitecture, structural integrity and function of muscles by forming three dimensional scaffold across sarcomeres of smooth muscles^{90,91}; (d) Fibrillin-1 (FBN1), a 312 kDa protein encoded by FBN1 gene, is a large cysteine rich glycoprotein produced by fibroblasts, and is the principal structural component of extracellular matrix forming microfibrils in the connective-tissue, and Interacting with other components of the extracellular matrix (ECM), this ubiquitous glycoprotein exert pivotal roles in tissue development, homeostasis and repair. In addition to mechanical support, FBN networks also exhibit regulatory activities on growth factor signalling, ECM formation, cell behaviour and the immune response^{92,93}; and (e) Microfibrillar-associated glycoprotein 4 (MFAP4), a 36 kDa protein encoded by MFAP4 gene is an extracellular matrix protein that plays major role in elastin fiber formation and is associated with ECM remodeling processes during vascular injury, and interacts with other ECM proteins such as FBN1 that provides cell adhesion, intercellular interactions and the assembly and/or maintenance of elastic fibres^{94,95}.

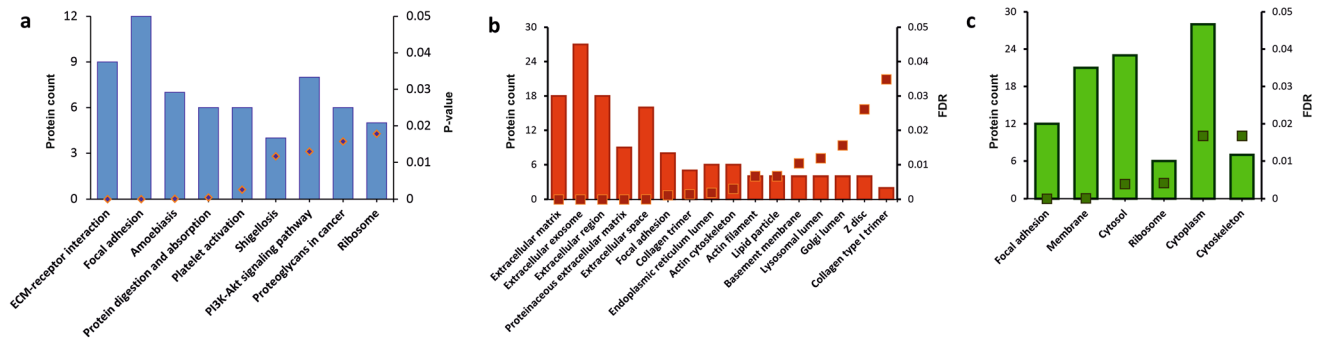


Figure 5. KEGG pathways analysis and Gene Ontology of differentially expressed proteins. (a) KEGG pathway analysis showing the various possible pathways regulated by the differentially expressed proteins; (b) Gene Ontology for cellular components showing distribution of 49 up-regulated proteins across various cellular organelles; (c) Gene Ontology for cellular components showing distribution of 60 down-regulated proteins across various cellular organelles. Square boxes designate the FDR corrected p-values.

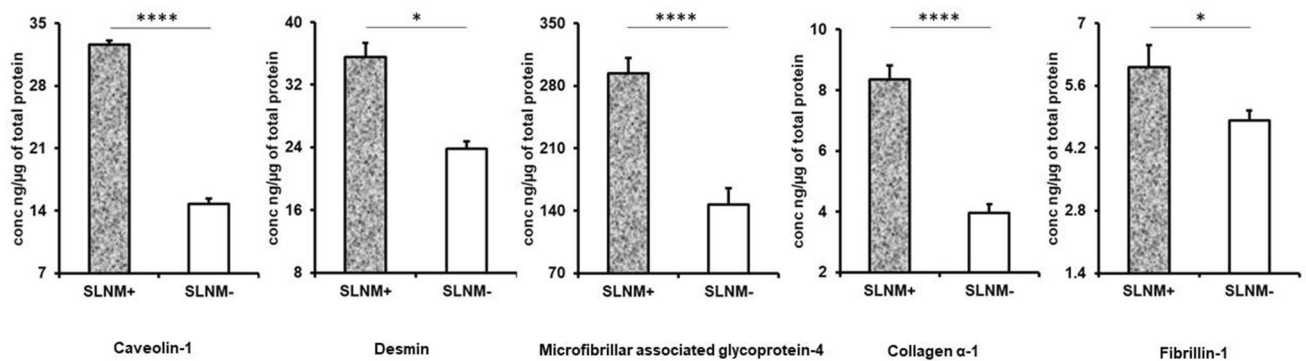


Figure 6. ELISA. Bar diagrams showing the ratio of protein expression of (a) caveolin-1, (b) desmin, (c) microfibrillar associated glycoprotein 4 (d) collagen α -1, and (e) fibrillin-1 in metastatic patients (black) and non-metastatic patients (grey). *indicates $P < 0.05$; ****indicates $P < 0.0001$.

Label based differential proteomic experiments, Pathway analysis, Gene Ontology studies, and ELISA experiments clearly establish the role of extracellular matrix proteins in sentinel lymph node metastasis. Extracellular matrices (ECMs) are highly specialized and dynamic three-dimensional (3D) scaffolds into which cells reside in tissues and its principal components are collagens, glycoproteins, and proteoglycans⁹⁶. Upon physiological and pathological triggers, ECM-degrading enzymes, matrikines, are released to remodel the ECM, to re-establish an appropriate functional meshwork and maintain cellular homeostasis^{97,98}. But in metastasis, ECM remodeling is hijacked and there is perturbation and degradation in ECM architecture by matrix metalloproteinases^{99,100}. Due to ECM degradation there is loss of endothelial integrity allowing cancer cells to escape from primary tumor to other tissues including lymph nodes¹⁰¹. During this process of metastasis, cancer cells undergo epithelial-to-mesenchymal transition (EMT), which can be induced by increased deposition of ECM proteins¹⁰². This action alters the phenotypic properties of cells and affects their propensity to escape primary tumor and cause metastasis¹⁰³. In addition, an increased regulation of ECM proteins through recruitment and activation of cancer-associated fibroblasts (CAF) results in activation of biophysical and biochemical oncogenic signalling pathways¹⁰⁴. The oncogenic signalling pathways of the identified extra cellular matrix proteins are (a) caveolin-1: PI3K/AKT and Ras/Raf signaling through the ERK; (b) Desmin: PI3K/AKT through caspase; (c) microfibril associated glycoprotein 4: ERK/MMP signalling through FAK and c-Jun; (d) collagen α -1: FAK signalling through PI3K/AKT and MAPK/ERK, and (e) fibrillin 1: SMAD2/3/4 and MEK pathway through ERK; which induces cell proliferation, survival, motility, angiogenesis, hypoxia, cancer stem cell activity, epithelial to mesenchymal transition and eventually lymph node metastasis¹⁰⁵⁻¹⁰⁹. Caveolin-1, Desmin and Collagen α -1 which mediate their functions through PI3K/Akt signalling which is represented along side ECM-receptor interaction in the pathway analysis. The overview of this detailed analysis is pictorially represented in Fig. 7. Therefore, it is this interplay between the up-regulated extra cellular matrix proteins, active growth factors of cancer cells, fibroblasts and signalling pathways, which together promote lymph node metastasis.

Caveolin-1, Desmin, microfibril associated glycoprotein 4, fibrillin 1 and collagen-1 have been identified as potential biomarkers that can discriminate metastatic from non-metastatic sentinel lymph nodes in early breast cancer. To understand their ability to differentiate the two clinical phenotypes AUCs were plotted based on the ELISA concentrations (Supplementary Fig. 2). The areas were estimated to be 0.81 and 1.0 for these five proteins. Diagnostic parameters of sensitivity, specificity, positive predictive value and Negative predictive values are over 80% for each making them fairly accurate as a translational tool (Table 4). Caveolin-1 and Desmin with cut-off

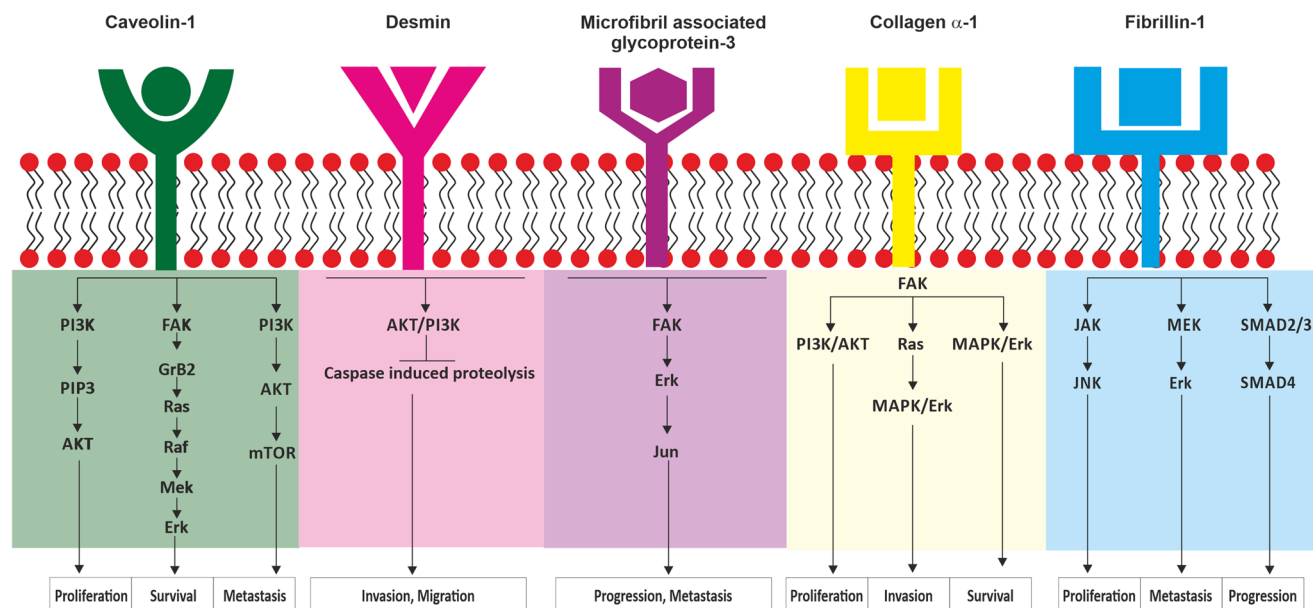


Figure 7. Diagrammatic representation showing oncogenic signalling pathways in breast cancer. Abbreviations: PI3K, Phosphatidylinositol 3-kinase; PIP3, Phosphatidylinositol (3,4,5)-trisphosphate; AKT, Protein kinase B; FAK, Focal adhesion kinase; GrB2, Growth factor receptor-bound protein 2; Ras, Rat sarcoma; Raf, Rapidly accelerated fibrosarcoma; Mek, Mitogen-activated protein kinase kinase; Erk, Extracellular signal-regulated kinase; mTOR, mechanistic target of rapamycin; PDK2, Pyruvate dehydrogenase kinase 2; Bcl2, B-cell lymphoma 2; NFkB, Nuclear factor kappa-light-chain-enhancer of activated B cells; JAK, Janus kinase; JNK, Jun N-terminal Kinase; SMAD, Small mothers against decapentaplegic.

Protein	Cut off values (ng/ μ g of total protein)	Statistical Parameters							
		AUC (95% CI: LL- UL)	Sensitivity (%)	Specificity (%)	LR+	LR-	PPV (%)	NPV (%)	Correctly classified (%)
Collagen alpha 1	≥ 5.98	0.98 (0.86–1.00)	92.31	91.67	11.07	0.08	92.30	91.66	92.00
Fibrillin-1	≥ 5.17	0.81 (0.59–0.93)	84.62	83.33	5.07	0.18	84.61	81.81	84.00
Microfibril associated glycoprotein 4	≥ 254.3	0.96 (0.79–0.99)	84.62	91.67	10.15	0.16	91.66	84.61	88.00
Caveolin-1	≥ 17.40	1.0 (0.86–1.00)	100	100	NA	NA	100	100	100
Desmin	≥ 28.50	1.0 (0.8–1.00)	100	100	NA	NA	100	100	100

Table 4. Diagnostic parameters of identified proteins to differentiate SLNM + from SLNM- in early breast cancer.

values of 17.4 and 28.5 ng/ μ g of tissue protein respectively, in particular are promising candidates with 100% values for all the diagnostic parameters. Since both the sensitivity and specificity measures are independent of prevalence rate, the practical utility of these two markers need to be validated among the population with varying prevalence rates.

Conclusion

iTRAQ based proteomic experiment is an ideal platform for comparative protein profiling in identification of potential biomarker candidates for sentinel lymph node metastasis in early breast cancer. Extra cellular matrix proteins caveolin-1, collagen α -1, desmin, fibrillin-1, and microfibrillar associated glycoprotein 4, have been identified as potential biomarkers that can differentiate the two metastatic states of sentinel lymph nodes. Each of these by themselves or as a collective panel offers translational scope for the design of 'on-table' diagnostics to flag sentinel lymph node metastasis in early breast cancer.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange¹¹⁰ Consortium via the PRIDE partner repository with the dataset identifier **PXD027668**.

Received: 17 May 2021; Accepted: 3 May 2022

Published online: 22 May 2022

References

1. Wu, H. *et al.* Isobaric tags for relative and absolute quantitation in proteomic analysis of potential biomarkers in invasive cancer, ductal carcinoma in situ, and mammary fibroadenoma. *Front. Oncol.* **10**, 574552 (2020).
2. IARC. New Global Cancer Data: GLOBOCAN 2020. *Int. Agency Res. Cancer* **6** (2020).
3. WHO. Globocan 2020. *International Agency for research* vol. 419 3–4 [https://ascopost.com/news/december-2020/globocan-2020-database-provides-latest-global-data-on-cancer-burden-cancer-deaths/#:~:text=Female breast cancer has now, with 685%2C000 deaths in 2020. \(2020\).](https://ascopost.com/news/december-2020/globocan-2020-database-provides-latest-global-data-on-cancer-burden-cancer-deaths/#:~:text=Female breast cancer has now, with 685%2C000 deaths in 2020. (2020).)
4. Sofi, N., Jain, M., Kapil, U. & Yadav, C. Epidemiological characteristics of breast cancer patients attending a tertiary health-care institute in the National Capital Territory of India. *J. Cancer Res. Ther.* **15**, 1087–1091 (2019).
5. Heer, E. *et al.* Global burden and trends in premenopausal and postmenopausal breast cancer: A population-based study. *Lancet Glob. Heal.* **8**, e1027–e1037 (2020).
6. Schwartz, R. S. & Erban, J. K. Timing of metastasis in breast cancer. *N. Engl. J. Med.* **376**, 2486–2488 (2017).
7. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **68**, 394–424 (2018).
8. Weledji, E. P. & Tambe, J. Breast cancer detection and screening. *Med. Clin. Rev.* <https://doi.org/10.21767/2471-299X.1000071> (2018).
9. Qiu, Y. *et al.* A multiple breast cancer stem cell model to predict recurrence of T1–3, N0 breast cancer. *BMC Cancer* **19**, 729 (2019).
10. Brenot-Rossi, I. *et al.* Nonvisualization of axillary sentinel node during lymphoscintigraphy: Is there a pathologic significance in breast cancer? *J. Nucl. Med.* **44**, 1232–1237 (2003).
11. Gradishar, W. J. *et al.* Clinical practice guidelines in oncology. *JNCCN J. Natl. Compr. Cancer Netw.* **16**, 310–320 (2018).
12. Kelley, M. C., Hansen, N. & McMasters, K. M. Lymphatic mapping and sentinel lymphadenectomy for breast cancer. *Am. J. Surg.* **188**, 49–61 (2004).
13. Krag, D. N., Weaver, D. L., Alex, J. C. & Fairbank, J. T. Surgical resection and radiolocalization of the sentinel lymph node in breast cancer using a gamma probe. *Surg. Oncol.* **2**, 335–340 (1993).
14. Noguchi, M. *et al.* The role of axillary lymph node dissection in breast cancer management. *Breast Cancer* **4**, 143–153 (1997).
15. Erb, K. M. & Julian, T. B. Completion of axillary dissection for a positive sentinel node: Necessary or not? *Curr. Oncol. Rep.* **11**, 15–20 (2009).
16. Borgstein, P. J. *et al.* Sentinel lymph node biopsy in breast cancer: Guidelines and pitfalls of lymphoscintigraphy and gamma probe detection. *J. Am. Coll. Surg.* **186**, 275–283 (1998).
17. van de Vijver, M. J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009 (2002).
18. Lorusso, G. & Rügge, C. New insights into the mechanisms of organ-specific breast cancer metastasis. *Semin. Cancer Biol.* **22**, 226–233 (2012).
19. Fry, S. A., Sinclair, J., Timms, J. F., Leatham, A. J. & Dwek, M. V. A targeted glycoproteomic approach identifies cadherin-5 as a novel biomarker of metastatic breast cancer. *Cancer Lett.* **328**, 335–344 (2013).
20. Zeng, L. *et al.* Identification of nucleobindin-2 as a potential biomarker for breast cancer metastasis using iTRAQ-based quantitative proteomic analysis. *J. Cancer.* **8**, 3062–3069 (2017).
21. Mokhtar, M. *et al.* Triple assessment of sentinel lymph node metastasis in early breast cancer using preoperative CTLG, intraoperative fluorescence navigation and OSNA. *Breast Cancer* **23**, 202–210 (2016).
22. Zhao, Q. L. *et al.* Elastasonography and two-dimensional ultrasonography in diagnosis of axillary lymph node metastasis in breast cancer. *Clin. Radiol.* **73**, 312–318 (2018).
23. Solon, J. G., Power, C., Al-Azawi, D., Duke, D. & Hill, A. D. K. Ultrasound-guided core biopsy: An effective method of detecting axillary nodal metastases. *J. Am. Coll. Surg.* **214**, 12–17 (2012).
24. Nandu, V. V. & Chaudhari, M. S. Efficacy of sentinel lymph node biopsy in detecting axillary metastasis in breast cancer using methylene blue. *Indian J. Surg. Oncol.* **8**, 109–112 (2017).
25. Harada, T., Tanigawa, N., Matsuki, M., Nohara, T. & Narabayashi, I. Evaluation of lymph node metastases of breast cancer using ultrasmall superparamagnetic iron oxide-enhanced magnetic resonance imaging. *Eur. J. Radiol.* **63**, 401–407 (2007).
26. Roumen, R. M. H., Valkenburg, J. G. M. & Geuskens, L. M. Lymphoscintigraphy and feasibility of sentinel node biopsy in 83 patients with primary breast cancer. *Eur. J. Surg. Oncol.* **23**, 495–502 (1997).
27. Gupta, A. K. *et al.* Cerebrospinal fluid proteomics for identification of α 2-macroglobulin as a potential biomarker to monitor pharmacological therapeutic efficacy in dopamine dictated disease states of Parkinson's disease and schizophrenia. *Neuropsychiatr. Dis. Treat.* **15**, 2853–2867 (2019).
28. Pokhriyal, R., Hariprasad, R., Kumar, L. & Hariprasad, G. Chemotherapy resistance in advanced ovarian cancer patients. *Biomark. Cancer.* **11**, 1179299X1986081 (2019).
29. Pathania, S. *et al.* Proteomics of sentinel lymph nodes in early breast cancer for identification of thymidylate synthase as a potential biomarker to flag metastasis: A preliminary study. *Cancer Manag. Res.* **12**, 4841–4854 (2020).
30. Kataria, K., Srivastava, A. & Qaiser, D. What Is a false negative sentinel node biopsy: Definition, reasons and ways to minimize it? *Indian J. Surg.* **78**, 396–401 (2016).
31. Orsburn, B. C. Proteome discoverer—a community enhanced data processing suite for protein informatics. *Proteomes* **9**, 1–13 (2021).
32. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
33. Pang, Z. *et al.* MetaboAnalyst 5.0: Narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* **49**, W388–W396 (2021).
34. Kamradt, M. C., Chen, F. & Cryns, V. L. The small heat shock protein α B-crystallin negatively regulates cytochrome c- and caspase-8-dependent activation of caspase-3 by inhibiting its autoproteolytic maturation. *J. Biol. Chem.* **276**, 16059–16063 (2001).
35. Mao, Y. W., Liu, J. P., Xiang, H. & Li, D. W. C. Human α A- and α B-crystallins bind to Bax and Bcl-Xs to sequester their translocation during staurosporine-induced apoptosis. *Cell Death Differ.* **11**, 512–526 (2004).
36. Liu, S. *et al.* As a novel p53 direct target, bidirectional gene HspB2/ α B-crystallin regulates the ROS level and Warburg effect. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1839**, 592–603 (2014).
37. Shi, C. *et al.* Alpha B-crystallin correlates with poor survival in colorectal cancer. *Int. J. Clin. Exp. Pathol.* **7**, 6056–6063 (2014).
38. Mellone, M. *et al.* Induction of fibroblast senescence generates a non-fibroblastic myofibroblast phenotype that differentially impacts on cancer prognosis. *Aging (Albany NY)* **9**, 114–132 (2017).
39. Hanley, C. J. *et al.* A subset of myofibroblastic cancer-associated fibroblasts regulate collagen fiber elongation, which is prognostic in multiple cancers. *Oncotarget* **7**, 6159–6174 (2016).
40. Li, Y. *et al.* miR-181a-5p inhibits cancer cell migration and angiogenesis via downregulation of matrix metalloproteinase-14. *Cancer Res.* **75**, 2674–2685 (2015).
41. Gonzalez-Avila, G. *et al.* Matrix metalloproteinases participation in the metastatic process and their diagnostic and therapeutic applications in cancer. *Crit. Rev. Oncol. Hematol.* **137**, 57–83 (2019).

42. Nieto, M. A., Huang, R. Y. J., Jackson, R. A. A. & Thiery, J. P. P. Emt: 2016. *Cell* **166**, 21–45 (2016).
43. Ju, Q. *et al.* Identification of a miRNA-mRNA network associated with lymph node metastasis in colorectal cancer. *Oncol. Lett.* **18**, 1179–1188 (2019).
44. Pike, L. J. Growth factor receptors, lipid rafts and caveolae: An evolving story. *Biochim. Biophys. Acta – Mol. Cell Res.* **1746**, 260–273 (2005).
45. Badana, A. K. *et al.* Lipid rafts disruption induces apoptosis by attenuating expression of LRP6 and survivin in triple negative breast cancer. *Biomed. Pharmacother.* **97**, 359–368 (2018).
46. Wang, R. *et al.* Caveolin-1 functions as a key regulator of 17 β -estradiol-mediated autophagy and apoptosis in BT474 breast cancer cells. *Int. J. Mol. Med.* **34**, 822–827 (2014).
47. Bailey, K. M. & Liu, J. Caveolin-1 up-regulation during epithelial to mesenchymal transition is mediated by focal adhesion kinase. *J. Biol. Chem.* **283**, 13714–13724 (2008).
48. Gai, X., Lu, Z., Tu, K., Liang, Z. & Zheng, X. Caveolin-1 is up-regulated by GLI1 and contributes to GLI1-driven EMT in hepatocellular carcinoma. *PLoS One*. **9**, e84551 (2014).
49. Joglekar, M., Elbazanti, W. O., Weitzman, M. D., Lehman, H. L. & van Golen, K. L. Erratum to: Caveolin-1 mediates inflammatory breast cancer cell invasion via the Akt1 pathway and RhoC GTPase: RhoC and caveolin -1 in inflammatory breast cancer. *J. Cell. Biochem.* **118**, 1273 (2017).
50. Li, S. *et al.* Shear stress promotes anoikis resistance of cancer cells via caveolin-1-dependent extrinsic and intrinsic apoptotic pathways. *J. Cell. Physiol.* **234**, 3730–3743 (2019).
51. Wang, K., Zhu, X., Chen, Y., Yin, Y. & Ma, T. Tubeimoside V sensitizes human triple negative breast cancer MDA-MB-231 cells to anoikis via regulating caveolin-1-related signaling pathways. *Arch. Biochem. Biophys.* **646**, 10–15 (2018).
52. Kim, S. H. *et al.* Role of secreted type I collagen derived from stromal cells in two breast cancer cell lines. *Oncol. Lett.* **8**, 507–512 (2014).
53. Liu, J. *et al.* Collagen 1A1 (COL1A1) promotes metastasis of breast cancer and is a potential therapeutic target. *Discov. Med.* **25**, 211–223 (2018).
54. De Wever, O. & Mareel, M. Role of tissue stroma in cancer cell invasion. *Journal of Pathology.* **200**, 429–447 (2003).
55. Zilberberg, L. *et al.* Specificity of latent TGF- β binding protein (LTBP) incorporation into matrix: Role of fibrillins and fibronectin. *J. Cell. Physiol.* **227**, 3828–3836 (2012).
56. Fuxe, J., Vincent, T. & De Herrerros, A. G. Transcriptional crosstalk between TGF β and stem cell pathways in tumor cell invasion: Role of EMT promoting Smad complexes. *Cell Cycle* **9**, 2363–2374 (2010).
57. Chaudhury, A. *et al.* TGF-B-mediated phosphorylation of hnRNP E1 induces EMT via transcript-selective translational induction of Dab2 and ILEI. *Nat. Cell Biol.* **12**, 286–293 (2010).
58. Liu, K. *et al.* Glyceraldehyde-3-phosphate dehydrogenase promotes cancer growth and metastasis through upregulation of SNAIL expression. *Int. J. Oncol.* **50**, 252–262 (2017).
59. Calderwood, S. K. & Gong, J. Heat shock proteins promote cancer: It's a protection racket. *Trends Biochem. Sci.* **41**, 311–323 (2016).
60. Tsutsumi, S., Beebe, K. & Neckers, L. Impact of heat-shock protein 90 on cancer metastasis. *Future Oncol.* **5**, 679–688 (2009).
61. Shiota, M. *et al.* Hsp27 regulates epithelial mesenchymal transition, metastasis, and circulating tumor cells in prostate cancer. *Cancer Res.* **73**, 3109–3119 (2013).
62. Kim, W. H., Lee, B. L., Kim, D. K. & Kleinman, H. K. Laminin-1-adherent cancer cells show increased proliferation and decreased apoptosis in vivo. *Anticancer Res.* **19**, 3067–3071 (1999).
63. Givant-Horwitz, V., Davidson, B. & Reich, R. Laminin-induced signaling in tumor cells. *Cancer Lett.* **223**, 1–10 (2005).
64. Wu, J. B. *et al.* Monoamine oxidase A mediates prostate tumorigenesis and cancer metastasis. *J. Clin. Invest.* **124**, 2891–2908 (2014).
65. Albig, A. R., Becenti, D. J., Roy, T. G. & Schiemann, W. P. Microfibril-associate glycoprotein-2 (MAGP-2) promotes angiogenic cell sprouting by blocking notch signaling in endothelial cells. *Microvasc. Res.* **76**, 7–14 (2008).
66. Yang, J. *et al.* Integrated analysis of microfibillar-associated proteins reveals MFAP4 as a novel biomarker in human cancers. *Epigenomics* **11**, 5–21 (2019).
67. Huang, Y., Arora, P., McCulloch, C. A. & Vogel, W. F. The collagen receptor DDR1 regulates cell spreading and motility by associating with myosin IIA. *J. Cell Sci.* **122**, 1637–1646 (2009).
68. Arjonen, A. *et al.* Mutant p53-associated myosin-X upregulation promotes breast cancer invasion and metastasis. *J. Clin. Invest.* **124**, 1069–1082 (2014).
69. Tirinato, L. *et al.* An overview of lipid droplets in cancer and cancer stem cells. *Stem Cells Int.* **2017**, 1–17 (2017).
70. Chen, Y. F. *et al.* MicroRNA-211 enhances the oncogenicity of carcinogen-induced oral carcinoma by repressing TCF12 and increasing antioxidant activity. *Cancer Res.* **76**, 4872–4886 (2016).
71. Yu, H. *et al.* Transgelin is a direct target of TGF- β /Smad3-dependent epithelial cell migration in lung fibrosis. *FASEB J.* **22**, 1778–1789 (2008).
72. Lin, Y. *et al.* Association of the actin-binding protein transgelin with lymph node metastasis in human colorectal cancer. *Neoplasia* **11**, 864–873 (2009).
73. Wu, X., Dong, L., Zhang, R., Ying, K. & Shen, H. Transgelin overexpression in lung adenocarcinoma is associated with tumor progression. *Int. J. Mol. Med.* **34**, 585–591 (2014).
74. Borensztajn, K., Peppelenbosch, M. P. & Spek, C. A. Coagulation Factor Xa inhibits cancer cell migration via LIMK1-mediated cofilin inactivation. *Thromb. Res.* **125**, e323–e328 (2010).
75. Zhou, Z. J. *et al.* Hnrnpab induces epithelial-mesenchymal transition and promotes metastasis of hepatocellular carcinoma by transcriptionally activating snail. *Cancer Res.* **74**, 2750–2762 (2014).
76. Stockley, J. *et al.* The RNA-binding protein hnRNP2 regulates β -catenin protein expression and is overexpressed in prostate cancer. *RNA Biol.* **11**, 755–765 (2014).
77. Deng, J. *et al.* Effects of hnRNP A2/B1 knockdown on inhibition of glioblastoma cell invasion, growth and survival. *Mol. Neurobiol.* **53**, 1132–1144 (2016).
78. Tauler, J., Zudaire, E., Liu, H., Shih, J. & Mulshine, J. L. hnRNP A2/B1 modulates epithelial-mesenchymal transition in lung cancer cell lines. *Cancer Res.* **70**, 7137–7147 (2010).
79. Bianchi, M. E. & Agresti, A. HMG proteins: Dynamic players in gene regulation and differentiation. *Curr. Opin. Genet. Dev.* **15**, 496–506 (2005).
80. Medrzycki, M. *et al.* Histone H1.3 suppresses H19 noncoding RNA expression and cell growth of ovarian cancer cells. *Cancer Res.* **74**, 6463–6473 (2014).
81. Wang, J. *et al.* Cathelicidin suppresses colon cancer metastasis via a P2RX7-dependent mechanism. *Mol. Ther. - Oncolytics.* **12**, 195–203 (2019).
82. Kang, M. *et al.* miR-129-2 suppresses proliferation and migration of esophageal carcinoma cells through downregulation of SOX4 expression. *Int. J. Mol. Med.* **32**, 51–58 (2013).
83. Dransart, E., Olofsson, B. & Cherfils, J. RhoGDIs revisited: Novel roles in Rho regulation. *Traffic* **6**, 957–966 (2005).
84. Hooper, J. A. *et al.* Purification and properties of Bovine Thymosin. *Ann. N. Y. Acad. Sci.* **249**, 125–144 (1975).

85. Fan, Y.-z, Chang, H., Ye, Y., Liu, J. & Wang, R. Thymosin $\alpha 1$ suppresses proliferation and induces apoptosis in human leukemia cell lines. *Peptides* **27**(9), 2165–2173. <https://doi.org/10.1016/j.peptides.2006.03.012> (2006).
86. Williams, T. M. & Lisanti, M. P. The caveolin proteins. *Genome Biol.* **5**, 214 (2004).
87. Xu, Y. *et al.* Endocytosis and membrane receptor internalization: Implication of F-BAR protein Carom. *Front. Biosci. - Landmark.* **22**, 1439–1457 (2017).
88. Gajko-Galicka, A. Mutations in type I collagen genes resulting in osteogenesis imperfecta in humans. *Acta Biochim. Pol.* **49**, 433–441 (2002).
89. Gelse, K., Pöschl, E. & Aigner, T. Collagens - Structure, function, and biosynthesis. *Adv. Drug Deliv. Rev.* **55**, 1531–1546 (2003).
90. Paulin, D. & Li, Z. Desmin: A major intermediate filament protein essential for the structural integrity and function of muscle. *Exp. Cell Res.* **301**, 1–7 (2004).
91. Shah, S. B. *et al.* Structural and functional roles of desmin in mouse skeletal muscle during passive deformation. *Biophys. J.* **86**, 2993–3008 (2004).
92. Sakai, L. Y., Keene, D. R., Renard, M. & De Backer, J. FBN1: The disease-causing gene for Marfan syndrome and other genetic disorders. *Gene* **592**, 279–291 (2016).
93. Schrenk, S., Cenzi, C., Bertalot, T., Conconi, M. T. & Di Liddo, R. Structural and functional failure of fibrillin-1 in human diseases (Review). *Int. J. Mol. Med.* **41**, 1213–1223 (2018).
94. Pilecki, B. *et al.* Characterization of microfibrillar-associated protein 4 (MFAP4) as a tropoelastin- and fibrillin-binding protein involved in elastic fiber formation. *J. Biol. Chem.* **291**, 1103–1114 (2016).
95. KIELTY, C. M. Fell-Muir lecture: Fibrillin microfibrils: structural tensometers of elastic tissues?. *Int. J. Exp. Pathol.* **98**, 172–190 (2017).
96. Morgado, F. N., da Silva, A. V. A. & Porrozzzi, R. Infectious diseases and the lymphoid extracellular matrix remodeling: A focus on conduit system. *Cells* **9**, 1–13 (2020).
97. Fares, J., Fares, M. Y., Khachfe, H. H., Salhab, H. A. & Fares, Y. Molecular principles of metastasis: A hallmark of cancer revisited. *Signal Transduct. Target. Ther.* <https://doi.org/10.1038/s41392-020-0134-x> (2020).
98. Wiig, H., Keskin, D. & Kalluri, R. Interaction between the extracellular matrix and lymphatics: Consequences for lymphangiogenesis and lymphatic function. *Matrix Biol.* **29**, 645–656 (2010).
99. Kaushik, S., Pickup, M. W. & Weaver, V. M. From transformation to metastasis: Deconstructing the extracellular matrix in breast cancer. *Cancer Metastasis Rev.* **35**, 655–667 (2016).
100. Winkler, J., Abisoye-Ogunniyan, A., Metcalf, K. J. & Werb, Z. Concepts of extracellular matrix remodelling in tumour progression and metastasis. *Nat. Commun.* **11**, 1–17 (2020).
101. Ecker, B. L. *et al.* Age-related changes in HAPLN1 increase lymphatic permeability and affect routes of melanoma metastasis. *Cancer Discov.* **9**, 82–95 (2019).
102. Paszek, M. J. *et al.* Tensional homeostasis and the malignant phenotype. *Cancer Cell* **8**, 241–254 (2005).
103. Frantz, C., Stewart, K. M. & Weaver, V. M. The extracellular matrix at a glance. *J. Cell Sci.* **123**, 4195–4200 (2010).
104. Erdogan, B. & Webb, D. J. Cancer-associated fibroblasts modulate growth factor signaling and extracellular matrix remodeling to regulate tumor metastasis. *Biochem. Soc. Trans.* **45**, 229–236 (2017).
105. Kortum, R. L. *et al.* Caveolin-1 is required for kinase suppressor of Ras 1 (KSR1)-mediated extracellular signal-regulated kinase 1/2 activation, H-Ras V12-induced senescence, and transformation. *Mol. Cell. Biol.* **34**, 3461–3472 (2014).
106. Zhang, H. *et al.* LAMB3 mediates apoptotic, proliferative, invasive, and metastatic behaviors in pancreatic cancer by regulating the PI3K/Akt signaling pathway. *Cell Death Dis.* <https://doi.org/10.1038/s41419-019-1320-z> (2019).
107. Xia, P., Liu, Y. & Cheng, Z. Signaling pathways in cardiac myocyte apoptosis. *Biomed. Res. Int.* **2016**, 1–22 (2016).
108. Mecham, R. P. & Gibson, M. A. The microfibril-associated glycoproteins (MAGPs) and the microfibrillar niche. *Matrix Biol.* **47**, 13–33 (2015).
109. Xiong, H. *et al.* Inhibition of JAK1, 2/STAT3 signaling induces apoptosis, cell cycle arrest, and reduces tumor cell invasion in colorectal cancer cells. *Neoplasia* **10**, 287–297 (2008).
110. Deutsch, E. W. *et al.* The ProteomeXchange consortium in 2020: Enabling 'big data' approaches in proteomics. *Nucleic Acids Res.* **48**, D1145–D1152 (2020).

Acknowledgements

Proteomics lab at Central Core Research Facility, AIIMS for the Mass spectrometric analysis is acknowledged.

Author contributions

S.P., M.I.K., K.R., S.B., S.S.S., T.R.P., G.H. performed the experiments; J.J., P.R.M. and A.S. performed the surgeries and procured the samples; S.M. carried out the histopathology analysis of the samples; S.H. carried out mammography on the recruited patients; P.V. Performed statistical analysis; A.S. and G.H. conceived the idea and designed the work flow of the experiment; S.B. and G.H. have done the proteomic analysis; S.P. and G.H. drafted the manuscript; All authors read and approved the final manuscript.

Funding

AIIMS (A:590).

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/s41598-022-12352-9>.

Correspondence and requests for materials should be addressed to G.H.

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) 2022