Journal of Proteomics 142 (2016) 114–121

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

# An integrated cell line-based discovery strategy identified follistatin and kallikrein 6 as serum biomarker candidates of breast carcinoma

Alain Mangé [a,b,c,d], Lampros Dimitrakopoulos [e], Antoninus Soosaipillai [e], Peter Coopman [a,b,c,d], Eleftherios P. Diamandis [e], Jérôme Solassol [a,b,c,d,*]

[a] IRCM, Institut de Recherche en Cancérologie de Montpellier, Montpellier, F-34298, France
[b] INSERM, U1194, Montpellier, F-34298, France
[c] Université de Montpellier, Montpellier, F-34090, France
[d] Institut régional du Cancer de Montpellier, Montpellier, F-34298, France
[e] Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada

ARTICLE INFO

*Article history:*
Received 11 January 2016
Received in revised form 7 April 2016
Accepted 27 April 2016
Available online 7 May 2016

*Keywords:*
Breast cancer
Biomarkers
Proteomics

ABSTRACT

Secreted proteins constitute a relevant source of putative cancer biomarkers. Here, we compared the secretome of a series of four genetically-related breast cancer cell lines as a model of aggressiveness using quantitative mass spectrometry. 537 proteins (59.5% of the total identified proteins) predicted to be released or shed from cells were identified. Using a scoring system based on i) iTRAQ value, ii) breast cancer tissue mRNA expression levels, and iii) immunohistochemical staining (public database), a short list of 10 candidate proteins was selected. Using specific ELISA assays, the expression level of the top five proteins was measured in a verification set of 56 patients. The four significantly differentially expressed proteins were then validated in a second independent set of 353 patients. Finally, follistatin (FST) and kallikrein 6 (KLK6) in serum were significantly higher ($p$-value < 0.0001) in invasive breast cancer patients compared with non-cancerous controls. Using specific cut-off values, FST distinguished breast cancer samples from healthy controls with a sensitivity of 65% and an accuracy of 68%, whereas KLK6 achieved a sensitivity of 55% and an accuracy of 61%. Therefore, we concluded that FST and KLK6 may have significance in breast cancer detection.

*Biological significance:* Discovery of new serum biomarkers that exhibit increased sensitivity and specificity compared to current biomarkers appears to be an essential field of research in cancer. Most biological markers show insufficient diagnostic sensitivity for early breast cancer detection and, for the majority of them, their concentrations are elevated only in metastatic forms of the disease. It is therefore essential to identify clinically reliable biomarkers and develop effective approaches for cancer diagnosis. One promising approach in this field is the study of secreted proteins through proteomic analysis of in vitro progression breast cancer models. Here we have shown that FST and KLK6 are elevated in breast cancer patient serum compared to healthy controls. We expect that our discovery strategy will help to identify cancer-specific and body-fluid-accessible biomarkers.



## 1. Introduction

Early detection of breast cancer, so as to diagnose and treat cancer in its state prior to clinical symptoms and/or metastasis, may greatly impact the treatment and prognosis of patients with this common, but deadly, malignancy. Clinical breast exam and breast self-exam did not show a clear benefit to increase early cancer detection. Diagnostic mammography can often help find breast cancer at an early stage. However, it can also miss some cancers. False-negative mammograms can be attributed to technical or interpretive reasons including the absence of calcifications and, more importantly, high breast density [1,2]. In addition, sometimes more tests are needed to find out if something found on a mammogram is or is not cancer. Therefore, there is an urgent need to develop complementary approaches to improve the sensitivity and specificity of early breast cancer detection.

As whole blood is considered to provide a dynamic representation of an individual's physiological and pathological status, human serum/plasma represents the most extensively studied biological matrix in the quest for cancer biomarkers [3]. Known serum-based tumor markers, such as CA15.3 or BR27.29, cannot be used for breast cancer detection. These markers, having insufficient predictive value as an early detection blood diagnostic assay, are only recommended for monitoring treatment response and disease recurrence of patients with metastatic disease [4]. Therefore, the search for specific disease-

*Abbreviations:* FDR, false discovery rate; GO, gene ontology; MMTS, methyl methanethiosulfonate; SCX, strong cation exchange; TCEP, tris-(2-carboxyethyl) phosphine.

* Corresponding author at: Department of Biopathology, CHU Arnaud de Villeneuve, 191 Avenue du Doyen Giraud, 34295 Montpellier Cedex 5, France.
*E-mail address:* j-solassol@chu-montpellier.fr (J. Solassol).

http://dx.doi.org/10.1016/j.jprot.2016.04.050
1874-3919/

associated biomarker signatures is of particular interest since they could be applied in a standard clinical setting. Biomarker discovery for this disease is still very much in its discovery phase.

Multiple approaches have been developed that hold promise for the identification of serum biomarkers. Among them, quantitative proteomics yields information that specifically recognizes the differences between samples. Numerous studies have already shown that this methodology can be used to uncover proteomic expression patterns linked with cancer, and some expression patterns have shown high promise to discover new biomarkers of early-stage cancers [5]. The challenges of blood proteomics, stemming from the complexity of the fluid, have led researchers to seek alternate sources for the discovery of circulating cancer biomarkers [6]. In tumor progression, the "secretome" proteins released or shed by cells, tissues, or organisms through various pathways, act as mediators of cancer cell-host communication in the cancer microenvironment [7]. These proteins may be detected by analyzing the conditioned media of cultured cell lines derived from specific cancer types. Interestingly, these proteins are often present at relative high concentration in proximity to tumor site but are also more likely to end up in body fluids such as serum/plasma in sufficient concentration to be detectable. However, search for cancer biomarkers using cell lines established from different individuals is complicated by differences among the cancer cell donors, their origin, the passage number and culture conditions, creating huge variations that may be unrelated to normal and malignant behaviors. A comparison of malignant and non-malignant cell line variants from the same lineage will avoid this problem.

Here, we report a progression-related differential secretome analysis between MCF10 genetically-related breast cancer cell lines [8–11] using 2D nanoLC-MS/MS and isobaric tags for relative and absolute quantitation (iTRAQ) labeling technology. We identified a total of 903 proteins. Among them, 109 were found to be present at significantly elevated levels in breast cancer cell lines compared to normal or premalignmant cell lines. The differential expression of selected proteins was further validated by specific ELISA assays in a large independent cohort of invasive breast cancer patients. We demonstrated that the serum levels of two proteins (follistatin and kallikrein 6) were significantly higher in breast cancer patients versus healthy controls.

## 2. Methods

### 2.1. Patient selection

Sera from female breast cancer patients were prospectively collected between 2005 and 2009 at the "Institut du Cancer de Montpellier", France, after obtaining written informed consent (Inserm RBM03–63 cohort). The study protocol was approved by the institutional review board. The healthy controls were collected at in the same institute during the same period, and were sex- and age-matched. All patients and healthy controls are Caucasian. The verification population set consisted of 56 subjects: 28 women whith histopathologic diagnosis of breast cancer, and 28 healthy controls with negative mammograms, negative physical breast exams for at least 4 years and no history of prior malignancy, immunodeficiency, autoimmune disorder, hepatitis or HIV infection. The validation population set included 353 patients: 241 patients with a histopathologic diagnosis of breast cancer, and 112 healthy controls exempt of any breast, infectious and/or inflammatory disease. Sera from patients with breast cancer were collected at the time of cancer diagnosis and just prior surgery. All samples were collected, processed, and stored in a similar fashion. Blood samples were centrifuged at 1250 g for 5 min, and sera were then stored at −80 °C. Detailed clinical and pathological information of breast cancer patients are listed in Table 1.

**Table 1**
Clinicopathologic characteristic of patients with invasive breast cancer.

| Characteristics | Verification Set | Validation set |
|---|---|---|
| | *N* = 28 (%) | *N* = 241 (%) |
| Age (years) median, [min–max] | 68 [38–86] | 72, [37–97] |
| Histotype | | |
| Ductal | 28 (100) | 222 (92.1) |
| Lobular | – | 19 (7.9) |
| Tumor size | | |
| T1 | 15 (53.6) | 115 (47.7) |
| T2 | 10 (35.7) | 112 (46.5) |
| T3 | 1 (3.6) | 4 (1.7) |
| T4 | 2 (7.1) | 10 (4.1) |
| Histological grade | | |
| I | 4 (14.3) | 39 (16.2) |
| II | 18 (64.3) | 126 (52.3) |
| III | 5 (17.8) | 72 (29.9) |
| Missing | 1 (3.6) | 4 (1.7) |
| Lymph node status | | |
| Negative | 14 (50.0) | 121 (50.2) |
| Positive | 14 (50.0) | 120 (49.8) |
| Estrogen receptor | | |
| Negative | 1 (3.6) | 46 (19.1) |
| Positive | 27 (96.4) | 195 (80.9) |
| Progesterone receptor | | |
| Negative | 6 (21.4) | 60 (24.9) |
| Positive | 19 (67.9) | 181 (75.1) |
| Missing | 3 (10.7) | 0 |
| Her-2 overexpression | | |
| Negative | 19 (67.9) | 138 (57.3) |
| Positive | 2 (7.1) | 25 (10.4) |
| Missing | 7 (25.0) | 78 (32.4) |
| Classification | | |
| Luminal A | 19 (67.8) | 127 (52.7) |
| Luminal B | 1 (3.6) | 14 (5.8) |
| Enriched Her-2 | 1 (3.6) | 11 (4.6) |
| Others | – | 11 (4.6) |
| Missing | 7 (25) | 78 (32.3) |

### 2.2. Cell cultures and conditioned medium

The human MCF10A (non-tumorigenic), MCF10·NeoT (premalignant; tumorigenic), MCF10.DCIS (tumorigenic and locally invasive) and MCF10·CA1d (tumorigenic and metastatic) breast cancer cell lines were purchased from Asterand (MCF10.DCIS), the Barbaba Ann Karmanos Cancer Institute (MCF10·NeoT and MCF10·CA1d) and ATCC (MCF10A). Cells were maintained in DMEM/F12 (1:1) supplemented with 5% horse serum, 10 μg/mL insulin, 25 ng/mL epidermal growth factor, 0.5 μg/mL hydrocortisone, and 100 ng/mL cholera toxin. Culture media were supplemented with penicillin/streptomycin (100 U/mL). Cells were cultured at 37 °C in 5% $CO_2$. DMEM/F12, horse serum and penicillin/streptomycin were obtained from Gibco and insulin, EGF, hydrocortisone and cholera toxin from Sigma-Aldrich. Cells were grown to 60% confluence in 100 mm culture dishes and were rinsed three times for 15 min with serum free medium. The cells were then incubated in the serum free medium at 37 °C for 18 h. The cells were 95–100% viable after the serum free growth, as determined by trypan blue exclusion counting (Supplemental Fig. 1A). The conditioned medium was centrifuged for 10 min at 800 g to remove suspended cells. Samples were then concentrated using Amicon Ultra Centrifugal Filter Unit with a 10 kDa cut-off (Millipore). The protein concentration was measured in triplicate using the Micro BCA Kit (Pierce). Equal loading of proteins onto protein gels showed the same pattern of bands for all the samples (Supplemental Fig.1B).

### 2.3. iTRAQ reagent labeling and mass spectrometry analysis

The experimental design used for this study is illustrated in Fig. 1. iTRAQ labeling and mass spectrometry analysis were performed as previously described [12]. Briefly, 80 μg protein of each sample were digested using trypsin before iTRAQ labeling. Labeled peptides were separated on an IPG Drystrip 24 cm, pH 3–10 using the Agilent 3100 OFFGEL Fractionator (Agilent), and all fractions were analyzed by nanoLC-MS/MS using a MALDI TOF/TOF 4800 mass spectrometer

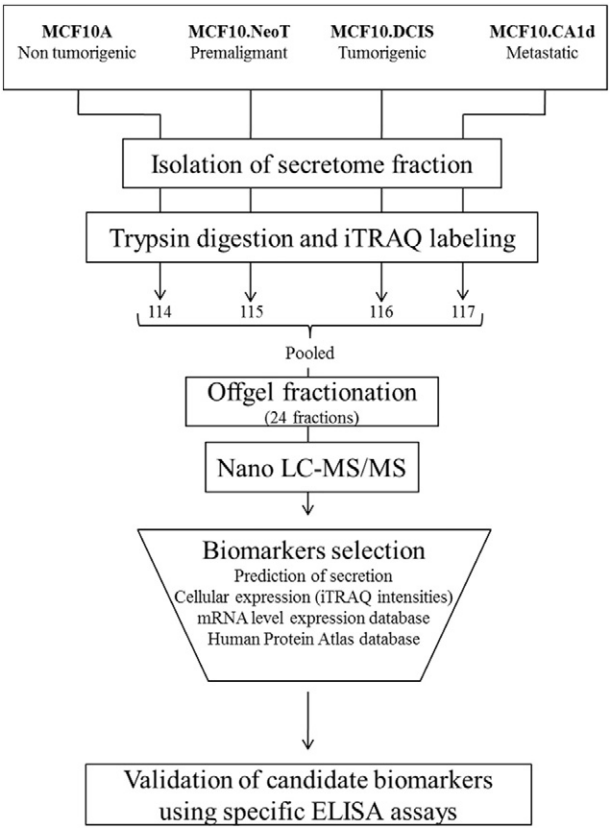


**Fig. 1.** Flow-chart illustrating the main steps of the study.

(ABSciex). Protein identification and quantification were performed with the ProteinPilot Software 2.0.1 (ABSciex). The search parameters for tryptic cleavage and accuracy are built-in functions of the software. The user-defined search parameters were as follows: (1) sample type, iTRAQ 4-plex (Peptide Labeled); (2) cysteine alkylation, MMTS; (3) digestion, trypsin; (4) instrument, 4800; (5) special factors, none; (6) species, *Homo sapiens*; (7) ID focus, biological modification; (8) database, UniProtKB/Swiss-Prot (release version 2014_08, http://www.uniprot.org); and (9) search effort, Thorough ID. A local false discovery rate (FDR) was estimated using the Proteomics System Performance Evaluation Pipeline (PSPEP) tool. Proteins with, at least, two peptides with a high confidence score (>99%) and a low FDR (estimated local FDR of 5%) were considered positively identified. The relative quantification was based on the ratio of the areas under the reporter peaks of MCF10·NeoT, MCF10.DCIS and MCF10·CA1d cell lines comparatively to the normal cell MCF10A. iTRAQ labeling followed by nanoLC-MS/MS analysis was repeated in duplicate to reduce the effect of experimental variation.

### 2.4. Selection of biomarker candidates

The following criteria were required to select a protein for further analysis: two or more unique peptides with a high confidence score (>99%), a *p*-value for protein quantification assigned by the ProteinPilot software <0.05, and greater than a 1.3-fold difference relative to the control sample (MCF10A). For this discovery phase, the *p*-value was not corrected with more stringent statistical analysis, such as adjustment for multiple testing, and the risk for false discovery association still exists at this step. We used SignalP 3.0 to analyze the identified proteins that are predicted to be secreted through the classical secretory pathway based on the presence of a signal peptide cleavage site in amino acid sequences [13]. This method predict the presence and the location of cleavage sites and signal peptide using a combination of artificial neural networks and hidden Markov model algorithms. Proteins potentially secreted through the non-classical secretory pathway were identified using SecretomeP 1.0 algorithm [14]. From these selections, we checked if the identified secreted proteins were found in plasma or serum in the Human Plasma Proteome Project database and Pride archive [15,16]. Biomarker candidates were then prioritized using scoring a system based on: (1) mRNA expression from a breast cancer tissue transcriptome dataset (GEO accession GSE29044) using the Gene Expression Omnibus (GEO) database [17]; (2) immunohistochemical (IHC) staining in breast cancer tissues obtained from Human Protein Atlas [18]; and (3) iTRAQ values from ProteinPilot analysis. A total of 7 points could be given. A two-fold change in mRNA expression was given one point. For protein expression in breast tissues, we used an intensity score based on a combination of IHC staining (negative, weak, moderate, and strong), quantity of cells stained (%) and numbers of samples analyzed. Depending on the protein expression in the malignant tissues compared to normal tissues, proteins were scored with 0 (low intensity score), 1 (medium score) or 2 (high score) points. Next, proteins with an iTRAQ ratio for MCF10.DCIS and MCF10·CA1d in the top 25% and in the top 50% were scored with 2 points and 1 point, respectively. Finally, selected proteins with a score >3 were searched using Pubmed to exclude proteins that had been extensively studied as breast cancer serum biomarkers.

### 2.5. Immunoassay and biochemical analyses

The human Leukemia inhibitory factor, LIF (RayBiotech, Inc.), Follistatin (FST) and Insulin-Like Growth Factor Binding Protein 2, IGFBP2 (R&D systems), and Tissue plasminogen activator, tPA (Abcam) concentrations were measured using a colorimetric sandwich enzyme linked immunosorbent assay (ELISA) as recommended by the manufacturers. Quantification of Kallikrein 6, KLK6 was performed using KLK6 in-house ELISA method [19]. The within-laboratory imprecision (CV) of the KLK6 assay is 5% at the range of 0.05–5 ng/mL.

### 2.6. Statistical analysis and gene ontology analysis

Statistical significance was determined using GraphPad Prism software version 5.03. Differences between conditions were analyzed using two-tailed Student's *t*-test or Mann-Whitney *U* test. A *p*-value of <0.05 was chosen for statistical significance. Results are expressed as medians and ranges. Individual performances of biomarkers were based on the receiver operating characteristic curves (ROC) and the area under the ROC curve (AUC). Cut-off values, sensitivities and accuracies for Follistatin and KLK6 were based on the ROC curve that discriminates invasive breast cancer patients from healthy controls, and were calculated with a predetermined specificity threshold of 80%. The linear combination (virtual marker) of markers was determined using a logistic regression model. The corresponding AUC, accuracy, sensitivity, and specificity of this combination were estimated. Gene ontology (GO) classification, including GO cellular process, GO cellular function and GO cellular component categories, was determined using the Generic GO Term Mapper (http://go.princeton.edu/cgi-bin/GOTermMapper).

## 3. Results

### 3.1. Identification and expression level of proteins released by MCF10 cells

The overall protein expression profile of secretome in MCF10 genetically-related models of breast cancer cell lines was compared using iTRAQ labelings and 2D nanoLC-MS/MS (Fig. 1). A total of 903 high confidence nonredundant proteins were identified using stringent criteria, including two or more peptides with a > 99% confidence score and 5% local FDR (see Supplemental Table S1). iTRAQ quantitation was performed using ProteinPilot software. The differentially expressed proteins in MCF10·NeoT, MCF10.DCIS and MCF10·CA1d cell lines

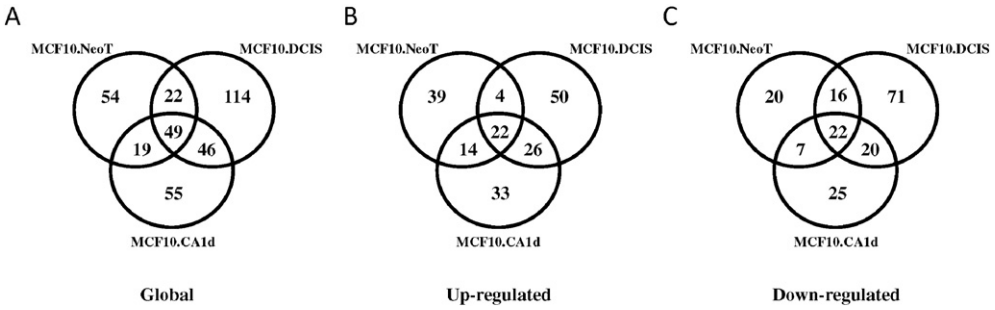


**Fig. 2.** Overlap of the differentially expressed proteins in conditioned medium of the three tumorigenic cell lines compared to the non tumorigenic MCF10A cell line. The overlap of global (A), up-regulated (B) and down-regulated (C) differentially expressed proteins is shown.

were determined using fold change (ratio) comparatively to the normal cell MCF10A. A *p*-value was assigned by the ProteinPilot software to assess the statistical significance of these variations. Only proteins with a differential expression of at least 1.3-fold relative to the control samples, and a $p$-value $< 0.05$ were considered in the filtering process. As a result, 359 proteins were found differentially expressed in MCF10·NeoT, MCF10.DCIS and MCF10·CA1d cell lines comparatively to MCF10A cell line (Fig. 2 and Supplemental Table S2). Among these 359 proteins, 178 were consistently up-regulated, 172 were consistently down-regulated, and 9 were differently regulated depending on the cell lines, as compared with MCF10A (Fig. 2 and Supplemental Table S2).

### 3.2. GO classification by cellular localization and biological function

The cellular localization, the biological processes and the molecular functions of all of the 903 proteins identified were classified using the GO classification system. Overall, the identified proteins were 75% cytoplasmic proteins, 74% extracellular region proteins, 31% plasma membrane proteins, and 47% nuclear proteins. Note that some proteins were classified in more than one compartment and that 7 proteins were unannotated (Fig. 3). As expected, the majority of identified proteins (79%) were annotated as extracellular with different origins (secretory origin or membrane shedding), and only a small proportion was annotated as intracellular (21%). The proteins were then functionally categorized on the basis of GO annotation terms, and linked to at least one molecular function and biological process categories. Functionally, the top three molecular functions were protein binding (80%), catalytic activity (38%) and metal ion binding (24%). Proteins were involved in various biological processes, including metabolic process (78%), response to stimulus (60%), cell differentiation (28%), regulation of cell death (20%), cell proliferation (14%) and cell communication (13%).

### 3.3. Biomarker selection

To narrow down the list of putative biomarkers, we first selected candidate biomarkers that are predicted to be secreted through the classical secretory pathway (endoplasmic reticulum/ Golgi dependent pathway based on the presence of a signal peptide) or non-classical secretory pathway using SignalP and SecretomeP algorithms, respectively. Among the 903 proteins identified, 537 proteins (59.5%) were predicted to be secreted, of which 293 in the classical secretory pathway (SignalP probability ≥0.50) and 244 released via the non-classical secretory pathway (Secretome score ≥0.50). From these 537 proteins, we then selected proteins that were significantly up-regulated in secretome of both malignant cell lines (MCF10.DCIS) compared to the secretome of normal and premalignant cell lines (MCF10A and MCF10·NeoT). As a result, 24 proteins correspond to these criteria. Of these 24 proteins, 23 were found in blood plasma or serum in the Human Plasma Proteome Project Database or Pride archive (Table 2). A scoring system based on mRNA expression levels and immunohistochemical staining in breast cancer tissues obtained from public databases, as well as the iTRAQ value from ProteinPilot analysis, were used to further select biomarkers for clinical validation. A breast cancer tissue transcriptome dataset (GSE29044) obtained from 73 cancer and 36 normal tissues were retrieved from the GEO database, and analyzed to rank our 23 biomarker candidates (Table 2). A two-fold change in mRNA expression was given one point. For protein expression in breast tissues, Human Protein Atlas database was used to rank biomarker candidates based

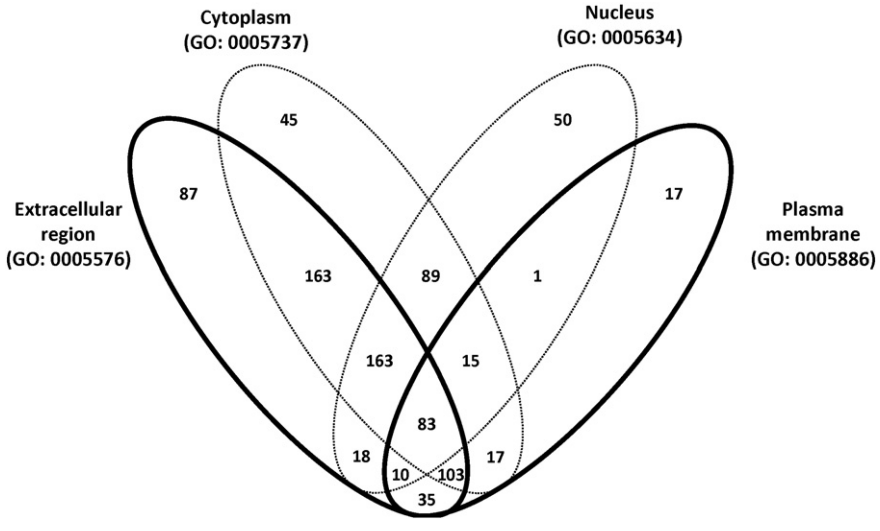


**Fig. 3.** Gene Ontology (GO) classification of the 903 identified proteins according to the cellular localization.

**Table 2**
List of 23 candidate biomarkers selected using a scoring system[a].

| Accession no | Gene | Proteins | SignalP | SecretomeP | Serum[b] | iTRAQ ratio compared to MCF10A | | | | mRNA expression | | Protein Atlas | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | MCF10.DCIS | | MCF10·CA1d | | FC[c] | Score | Score | |
| | | | | | | FC[c] | Score | FC[c] | Score | | | | |
| P07339 | CATD | Cathepsin D | Y | | Y | 1.63 | 0 | 2.75 | 2 | 3.01 | 1 | 2 | 5 |
| P00750 | TPA | Tissue-type plasminogen activator | Y | | Y | 3.11 | 2 | 1.45 | 0 | 0.74 | 0 | 2 | 4 |
| P15018 | LIF | Leukemia inhibitory factor | Y | | Y | 4.70 | 2 | 1.59 | 0 | 0.81 | 0 | 2 | 4 |
| P18065 | IFBP2 | Insulin-like growth factor-binding protein 2 | Y | | Y | 2.50 | 1 | 2.72 | 2 | 1.19 | 0 | 1 | 4 |
| Q92876 | KLK6 | Kallikrein-6 | Y | | Y | 1.85 | 2 | 2.42 | 2 | 1.84 | 0 | 0 | 4 |
| P19883 | FST | Follistatin | Y | | Y | 2.03 | 1 | 2.24 | 1 | 0.58 | 0 | 2 | 4 |
| O60911 | CATL2 | Cathepsin L2 | Y | | Y | 2.58 | 1 | 2.66 | 2 | 1.25 | 0 | 0 | 3 |
| P20061 | TCO1 | Transcobalamin-1 | Y | | Y | 3.83 | 2 | 2.37 | 1 | 0.45 | 0 | 0 | 3 |
| P28300 | LYOX | Protein-lysine 6-oxidase | Y | | Y | 1.96 | 0 | 2.57 | 2 | – | 0 | 1 | 3 |
| Q14574 | DSC3 | Desmocollin-3 | Y | | Y | 1.82 | 0 | 1.84 | 1 | 0.36 | 1 | 1 | 3 |
| P01011 | AACT | Alpha-1-antichymotrypsin | Y | | Y | 4.38 | 2 | 1.69 | 0 | 1.42 | 0 | 0 | 2 |
| P09486 | SPRC | SPARC | Y | | Y | 2.85 | 1 | 2.03 | 1 | 1.59 | 0 | 0 | 2 |
| P09668 | CATH | Cathepsin H | Y | | Y | 2.39 | 0 | 1.65 | 0 | 1.09 | 0 | 2 | 2 |
| P48745 | NOV | Protein NOV homolog | Y | | Y | 1.80 | 0 | 2.55 | 2 | – | 0 | 0 | 2 |
| Q02487 | DSC2 | Desmocollin-2 | Y | | Y | 1.77 | 0 | 2.39 | 1 | 0.58 | 0 | 1 | 2 |
| Q12907 | LMAN2 | Vesicular integral-membrane protein VIP36 | Y | | Y | 1.32 | 0 | 1.30 | 0 | 0.99 | 0 | 2 | 2 |
| Q92626 | PXDN | Peroxidasin homolog | Y | | Y | 2.55 | 1 | 1.51 | 0 | 2.81 | 1 | 0 | 2 |
| P31431 | SDC4 | Syndecan-4 | Y | | Y | 2.65 | 1 | 1.76 | 0 | 0.96 | 0 | 0 | 1 |
| Q9Y2A9 | B3GN3 | UDP-GlcNAc:betaGal beta-1.3-*N*-acetylglucosaminyltransferase 3 | N | Y | Y | 2.01 | 0 | 1.94 | 1 | – | 0 | 0 | 1 |
| P06865 | HEXB | Beta-hexosaminidase subunit alpha | Y | | Y | 1.41 | 0 | 1.40 | 0 | 1.24 | 0 | 1 | 1 |
| P07686 | HEXA | Beta-hexosaminidase subunit beta | N | Y | Y | 1.49 | 0 | 1.59 | 0 | 1.17 | 0 | 0 | 0 |
| O95274 | LYPD3 | Ly6/PLAUR domain-containing protein 3 | Y | | Y | 1.84 | 0 | 1.69 | 0 | 1.66 | 0 | 0 | 0 |
| P15514 | AREG | Amphiregulin | Y | | Y | 1.76 | 0 | 1.39 | 0 | 1.00 | 0 | 0 | 0 |

[a] Described in text.
[b] Found in plasma or serum in the Human Proteome Project database or Pride archive.
[c] FC, fold-change.

on their differential immunohistochemical expression in breast cancer tissues compared to normal tissues. Eleven proteins showed over-expression in malignant tissues that were scored with one or two points depending on the IHC staining (Table 2). Finally, biomarker candidates were scored based on their iTRAQ ratio between MCF10.DCIS or MCF10·CA1d, and MCF10A (proteins in the top 25% were given 2 points, whereas those in the top 50% were given 1 point). Applying these criteria led to the selection of 10 promising biomarker candidates with a score ≥ 3 points (with a maximum of 7 points). We excluded the first protein (Cathepsin D) that had been extensively studied as a breast cancer biomarker, and the next top five proteins were verified using ELISA assays.

### 3.4. Detection of secretome proteins in human sera

Using specific ELISA assays, we quantified LIF, FST, IGFBP2, KLK6 and tPA in 28 healthy control serum samples and 28 invasive breast cancer serum samples. For tPA, no significant difference ($p$-value > 0.05) was observed between healthy control serum samples and invasive cancer serum samples (Fig. 4 and Table 3). Interestingly, the concentration

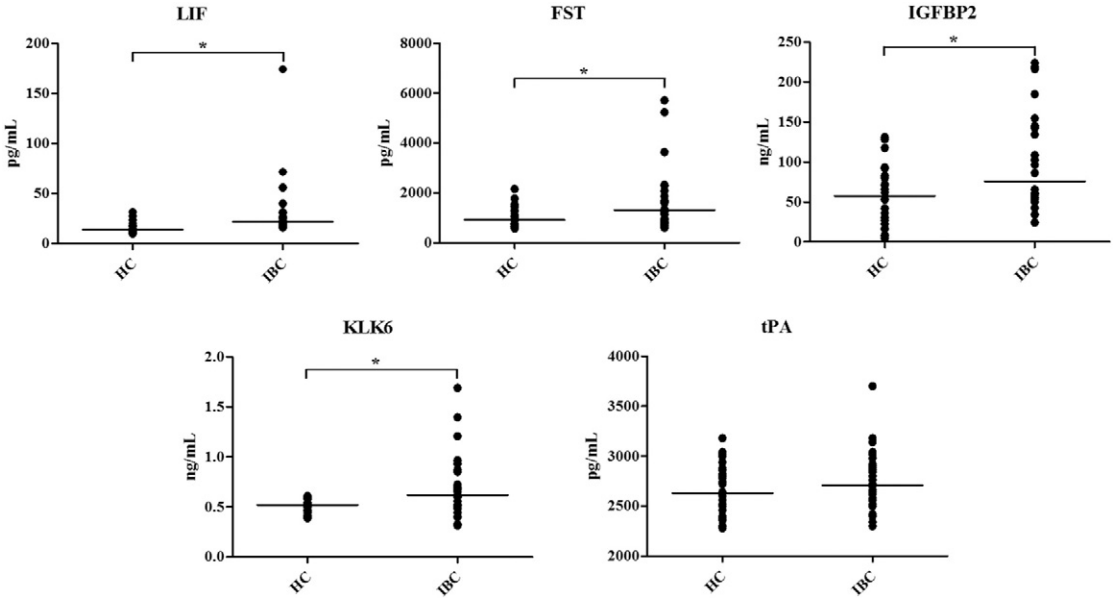


**Fig. 4.** Expression analysis of LIF, FST, IGFBP2, KLK6 and tPA in healthy control and invasive breast cancer samples. The concentration of the proteins in the verification population set of 28 healthy controls and 28 invasive breast cancer samples was determined using ELISA assay. The median values are indicated. IBC, invasive breast cancer; HC, healthy controls. *$p < 0.05$.

medians of LIF, FST, IGFBP2 and KLK6 were significantly higher ($p$-value < 0.0001, 0.01, 0.05, 0.01, respectively) in invasive breast cancer serum samples than in healthy control serum samples with 1.5, 1.4, 1.3, 1.2 fold-changes respectively, and their AUC ranged from 0.70 to 0.89 (Fig. 4 and Table 3).

To evaluate the potential of LIF, FST, IGFBP2 and KLK6 as blood-base biomarkers, proteins were measured in an independent set of 241 invasive breast cancer serum samples (120 node-negative breast cancers and 121 node-positive breast cancers) and 112 healthy controls (Fig. 5). FST and KLK6 serum levels were significantly higher ($p$-value < 0.0001) in invasive breast cancer patients (median [range]; 1231 pg/mL [302–5025] and 0.55 ng/mL [0.30–1.61], respectively) than those of healthy controls (915 pg/mL [296–1226] and 0.48 ng/mL [0.32–0.59], respectively). However, for LIF and IGFBP2, no significant differences were observed between breast cancer patients (28.1 pg/mL [3.0–77.5] and 58.5 [5.2–121.5], respectively) and healthy controls (31.0 pg/mL [10.4–1085.7] and 58.9 [26.6–267.9], respectively). To further investigate the diagnostic significance of FST and KLK6 in serum samples, we constructed ROC curves (Fig. 6) and determined an AUC of 0.77 [0.71–0.80] and 0.71 [0.64–0.76] for FST and KLK6, respectively, when we compared cancer samples and healthy controls. For a predetermined specificity of 80%, cut-off values of 1072 pg/mL and 0.535 ng/mL were defined for FST and KLK6, respectively. Using these cut-off values, FST distinguished breast cancer samples from healthy controls with a sensitivity of 65% and an accuracy of 68%, whereas KLK6 achieved a sensitivity of 55% and an accuracy of 61%. Interestingly, the combination of FST and KLK6 was able to discriminate breast cancer samples from healthy controls with the highest AUC (0.82 [0.76–0.86]; Fig. 6), yielding a sensitivity of 68% with a fixed specificity of 80%, and an accuracy of 71%. We next evaluated the correlation of both proteins with classic biological, histologic, and clinical variables in cancer group. We did not find a correlation with any of the variables, including lymph node criteria.

## 4. Discussion

Diagnostic mammography has been a mainstay to help diagnose suspicious lumps that have been found by a woman or her healthcare provider. However, many breast cancer cases especially in younger women having dense breast tissues, remain difficult to detect [20,21]. MRI has been suggested as an adjunct test to mammography mainly in high risk populations, however it is not an option for claustrophobic patients, patients with pacemakers or other metallic implants [20]. In addition, MRI exhibits a lower sensitivity in the detection of very small cancers when compared with mammography [22]. As a result, there has been much interest in development and validation of serum-based biomarkers. Biomarkers assisting in breast cancer detection and management have been proposed in the literature and can be mainly categorized as diagnostic, prognostic, predictive of response to treatment and biomarkers useful for monitoring for disease recurrence [23]. Proteomic biomarkers of breast carcinoma ideally should be able to be measured with a simple, cost-effective and minimally invasive assay. Thus, proteins “leaking” from the primary cancerous tissue to blood circulation would serve as excellent biomarkers for the early detection of breast cancer [24–26]. A typical preclinical study for the development of a biomarker set consists of three phases [27]. Upon hypothesis generation, a discovery phase in a small number of samples is undertaken, followed by a verification phase that usually measures prioritized biomarkers from the discovery set in pairs of case and control samples. Finally, a validation phase in a larger independent cohort of samples identifies promising biomarkers that can move on to clinical evaluation. Mass spectrometry is widely used for the discovery phase, whereas ELISA's are used for the validation phase and the subsequent clinical implementation of the developed tests [28,29]. However, like in all biomarker development pipelines, attention should be paid to use the appropriate cohorts of samples that serve the specific purpose that is pursued [27]. As an optimization guideline to reliably identify useful biomarkers, discovery research phase as well as validation phase might be more fruitful with use of specimen sets that adhere to PRoBE design principles [30].

Here, we report an integrated cell line-based discovery approach for the identification of protein biomarkers and the subsequent validation of two of them for the early detection of breast carcinoma that could be used in complement to mammography when unconclusive. A breast cancer progression model was used for the discovery phase. Specifically, a non tumorigenic and 3 tumorigenic (one premalignant, one locally invasive and one metastatic) cell lines were cultured and their secretome was analyzed using iTRAQ relative quantification. 359 proteins with at least 1.3-fold differential expression between the 3 tumorigenic cell lines and the non tumorigenic MCF10A cell line were identified as candidate breast cancer biomarkers. Further filtering for secreted proteins and prioritization based on gene expression data and immunohistochemical staining from breast cancer tissues combined with iTRAQ results provided a short list of 5 proteins. We then assessed their expression level in a verification cohort of 56 samples. We confirmed the significantly higher concentrations of KLK6, FST, LIF and IGFBP2 in the breast cancer group compared to the healthy controls, whereas tPA expression showed no significant difference between both groups. When an independent validation cohort of 241 invasive breast cancer serum samples and 112 healthy control samples was used, only KLK6 and FST protein expressions were found to be significantly higher in the breast cancer group, compared to healthy controls. ROC curves for those two proteins had AUC values 0.71 and 0.77 respectively. Sensitivities were 55% and 65% at a fixed specificity of 80%. Finally, the combination of FST and KLK6 yielded a sensitivity of 68%.

KLK6 is a 26.8 kDa serine protease with trypsin-like activity involved in a large variety of physiological and pathophysiological processes [31]. KLK6 mRNA was initially identified to be upregulated in primary mammary carcinoma lines when compared to normal tissues, but not in metastatic breast cancer cell lines [32]. The transcriptional regulation of KLK6 in breast cancer is not uniform in different tumors or cell lines, and has been shown to be driven by complex epigenetic events [33]. KLK6 has been implicated to play a role in the degradation of the extracellular matrix and thus facilitate tumor invasion and metastasis [31,34], but KLK6 has also a cancer-inhibiting activity in MDA-MB-231 cells [33,35]. However, this last effect was not observed in an subset of overexpressing KLK6 breast cancer tumors [35]. Interestingly, the stimulation of MCF7 breast cancer cell line by des[Arg$^9$]bradykinin causes release of KLK6 into the incubation medium and increases cell invasion into matrigel [36]. The presence of KLK6 in the conditioned medium has also been reported in other breast cancer cell lines such as MDA-MB-468 and BT474 [37]. Follistatin (FST) is a 38 kDa protein involved in the regulation of some members of the TGF-β superfamily such as activin and bone morphogenetic protein [38]. Activin plays a significant role in the malignant progression and has been found to have both oncogenic and tumor suppressor roles in cancer [39]. Specifically, activin exhibits an antiproliferative effect in early stages of breast cancer [40,41], and in prostate cancer [42,43], while activin is associated with increased proliferation and invasion in head and neck cancer [44]. FST has a high affinity for activin molecules, having thus the capacity to

**Table 3**
LIF, FST, IGFBP2, KLK6 and tPA concentration and AUC values assessed in the 56 healthy control and invasive breast cancer serum samples.

| | Median [range] | | AUC [95% CI] |
|---|---|---|---|
| | Healthy controls ($n = 28$) | Invasive breast cancer ($n = 28$) | |
| LIF | 14.8 pg/mL [9.6–31.6] | 21.9 pg/mL [15.6–174.1] | 0.89 [0.79–0.98] |
| FST | 926 pg/mL [576–2159] | 1319 pg/mL [604–5716] | 0.71 [0.57–0.85] |
| IGFBP2 | 57.5 ng/mL [4.6–131.2] | 75.9 ng/mL [24.2–224] | 0.70 [0.55–0.80] |
| KLK6 | 0.5 ng/mL [0.4–0.6] | 0.6 ng/mL [0.3–1.7] | 0.71 [0.56–0.86] |
| tPA | 2630 pg/mL [2280–3180] | 2710 pg/mL [2300–3700] | 0.57 [0.43–0.72] |

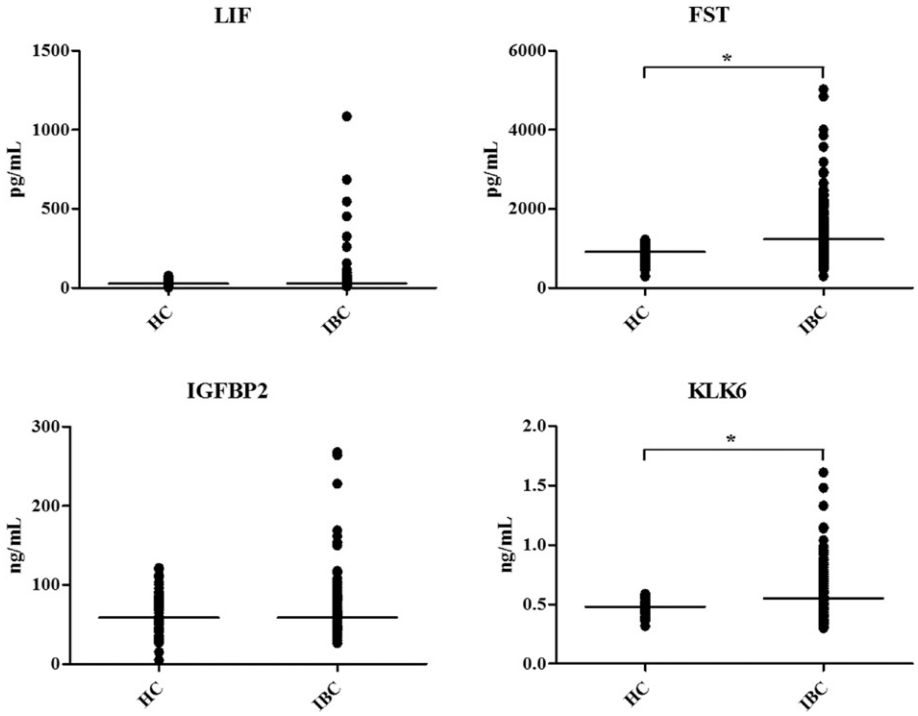


**Fig. 5.** Expression analysis of LIF, FST, IGFBP2 and KLK6 in healthy control and invasive breast cancer samples. The concentration of the proteins in the validation population set of 112 healthy controls and 241 invasive breast cancer samples was determined using ELISA assay. The median values are indicated. IBC, invasive breast cancer; HC, healthy controls. $^{*}p < 0.05$.

form inactive complexes and abrogate activin signaling pathway [41]. Some evidence supports that FST plays a role in the malignant progression of several cancers, including breast cancer [40], and may be also involved in the tumor angiogenic response, as well as in the tumor cell dissemination [38]. Taken together, our results show increased KLK6 and FST protein expression in both the conditioned medium of locally invasive and metastatic breast cancer cell lines, when compared with the non-tumorigenic cell lines, supporting that KLK6 and FST are pro-oncogenic in this breast cancer model and may favor proliferation, migration and invasion of cancer cells. This hypothesis is also supported by clinical observations, which show that KLK6 and FST are significantly increased in the sera of breast cancer patients compared to the healthy control group of patients.

To determine the usefulness of promising diagnostic biomarkers, candidates should be tested and validated by several independent studies using testing and training sets of both adequate size and patient populations [45]. These populations should ideally include sera from patients with benign breast lesions, early and very early stage breast cancers, as well as patients with high risk of breast cancer, different molecular types of breast cancer, and other control groups such as non-breast malignancies. Interestingly, the National Cancer Institute through the Early Detection Research Network develops suitable serum collection for both discovery and subsequent validation phases of early biomarker detection study. Such development of standardized resources should be largely favorite in France. As our population study do not satisfy all the criteria described in the ProBE design for clinical evaluation of biomarkers [30], we cannot exclude the possibility of confounding by unmeasured variables. However, we minimized risk of confounding variables by matching groups with age, race, site, and the timing of blood draw with respect to diagnosis, surgery, and treatment. Given the heterogeneity of results obtained with adjunct to mammography radiographic diagnostic tools and the complete absence of available diagnostic biomarkers in this early phase of the disease, we believe that our results are encouraging for the future development of blood assays for early detection of breast cancer. Based on these findings, we propose that KLK6 and FST could be considered as relevant breast cancer biomarkers that could be tested in future systematic and multi-institutional trials to investigate their clinical utility.

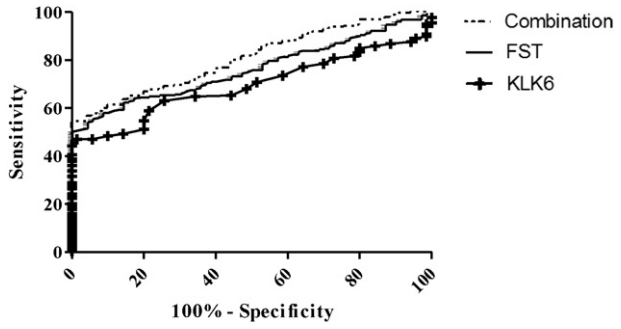


**Fig. 6.** Receiver operating characteristic curve analysis of FST, KLK6 and combination of both markers to discriminate invasive breast cancer patients ($n = 241$) from healthy controls ($n = 112$).

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2016.04.050.

## Conflict of interest

Authors declared no conflict of interest.

## Acknowledgements

This work was supported in part by grants from the Fondation Simone & Gustave Prévot.

## References

[1] J.G. Elmore, K. Armstrong, C.D. Lehman, S.W. Fletcher, Screening for breast cancer, J. Am. Med. Assoc. 293 (2005) 1245–1256.
[2] B.A. Virnig, T.M. Tuttle, T. Shamliyan, R.L. Kane, Ductal carcinoma in situ of the breast: a systematic review of incidence, treatment, and outcomes, J. Natl. Cancer Inst. 102 (2010) 170–178.
[3] S.M. Hanash, C.S. Baik, O. Kallioniemi, Emerging molecular biomarkers–blood-based strategies to detect and monitor cancer, Nat. Rev. Clin. Oncol. 8 (2011) 142–150.

[4] L. Harris, H. Fritsche, R. Mennel, L. Norton, P. Ravdin, S. Taube, et al., American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer, J. Clin. Oncol. 25 (2007) 5287–5312.
[5] E.P. Diamandis, Towards identification of true cancer biomarkers, BMC Med. 12 (2014) 156.
[6] A. Taguchi, S.M. Hanash, Unleashing the power of proteomics to develop blood-based cancer markers, Clin. Chem. 59 (2013) 119–126.
[7] S.T. Chen, T.L. Pan, H.F. Juan, T.Y. Chen, Y.S. Lin, C.M. Huang, Breast tumor microenvironment: proteomics highlights the treatments targeting secretome, J. Proteome Res. 7 (2008) 1379–1387.
[8] F. Basolo, J. Elliott, L. Tait, X.Q. Chen, T. Maloney, I.H. Russo, et al., Transformation of human breast epithelial cells by *c*-Ha-ras oncogene, Mol. Carcinog. 4 (1991) 25–35.
[9] F.R. Miller, S.J. Santner, L. Tait, P.J. Dawson, MCF10DCIS.com xenograft model of human comedo ductal carcinoma in situ, J. Natl. Cancer Inst. 92 (2000) 1185–1186.
[10] S.J. Santner, P.J. Dawson, L. Tait, H.D. Soule, J. Eliason, A.N. Mohamed, et al., Malignant MCF10CA1 cell lines derived from premalignant human breast epithelial MCF10AT cells, Breast Cancer Res. Treat. 65 (2001) 101–110.
[11] H.D. Soule, T.M. Maloney, S.R. Wolman, W.D. Peterson Jr., R. Brenz, C.M. McGrath, et al., Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10, Cancer Res. 50 (1990) 6075–6086.
[12] A. Mange, A. Goux, S. Badiou, L. Patrier, B. Canaud, T. Maudelonde, et al., HDL proteome in hemodialysis patients: a quantitative nanoflow liquid chromatography-tandem mass spectrometry approach, PLoS One 7 (2012) e34107.
[13] J.D. Bendtsen, H. Nielsen, G. von Heijne, S. Brunak, Improved prediction of signal peptides: SignalP 3.0, J. Mol. Biol. 340 (2004) 783–795.
[14] J.D. Bendtsen, L.J. Jensen, N. Blom, G. Von Heijne, S. Brunak, Feature-based prediction of non-classical and leaderless protein secretion, Protein Eng. Des. Sel. 17 (2004) 349–356.
[15] V. Nanjappa, J.K. Thomas, A. Marimuthu, B. Muthusamy, A. Radhakrishnan, R. Sharma, et al., Plasma proteome database as a resource for proteomics research: 2014 update, Nucleic Acids Res. 42 (2014) D959–D965.
[16] J.A. Vizcaino, R.G. Cote, A. Csordas, J.A. Dianes, A. Fabregat, J.M. Foster, et al., The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013, Nucleic Acids Res. 41 (2013) D1063–D1069.
[17] R. Edgar, M. Domrachev, A.E. Lash, Gene expression omnibus: NCBI gene expression and hybridization array data repository, Nucleic Acids Res. 30 (2002) 207–210.
[18] M. Uhlen, E. Bjorling, C. Agaton, C.A. Szigyarto, B. Amini, E. Andersen, et al., A human protein atlas for normal and cancer tissues based on antibody proteomics, Mol. Cell. Proteomics 4 (2005) 1920–1932.
[19] K. Oikonomopoulou, K.K. Hansen, A. Baruch, M.D. Hollenberg, E.P. Diamandis, Immunofluorometric activity-based probe analysis of active KLK6 in biological fluids, Biol. Chem. 389 (2008) 747–756.
[20] D. Euhus, P.A. Di Carlo, N.F. Khouri, Breast cancer screening, Surg. Clin. N. Am. 95 (2015) 991–1011.
[21] H. Joensuu, R. Asola, K. Holli, E. Kumpulainen, V. Nikkanen, L.M. Parvinen, Delayed diagnosis and large size of breast cancer after a false negative mammogram, Eur. J. Cancer 30A (1994) 1299–1302.
[22] A. Shimauchi, S.A. Jansen, H. Abe, N. Jaskowiak, R.A. Schmidt, G.M. Newstead, Breast cancers not detected at MRI: review of false-negative lesions, AJR Am. J. Roentgenol. 194 (2010) 1674–1679.
[23] J.A. Ludwig, J.N. Weinstein, Biomarkers in cancer staging, prognosis and treatment selection, Nat. Rev. Cancer 5 (2005) 845–856.
[24] L. Chung, R.C. Baxter, Breast cancer biomarkers: proteomic discovery and translation to clinically relevant assays, Expert Rev. Proteomics 9 (2012) 599–614.
[25] P. Gromov, J.M. Moreira, I. Gromova, Proteomic analysis of tissue samples in translational breast cancer research, Expert Rev. Proteomics 11 (2014) 285–302.
[26] J. Lacombe, A. Mange, J. Solassol, Use of autoantibodies to detect the onset of breast cancer, J. Immunol. Res. 2014 (2014) 574981.
[27] S. Surinova, R. Schiess, R. Huttenhain, F. Cerciello, B. Wollscheid, R. Aebersold, On the development of plasma protein biomarkers, J. Proteome Res. 10 (2011) 5–16.
[28] A.P. Drabovich, E. Martinez-Morillo, E.P. Diamandis, Toward an integrated pipeline for protein biomarker development, Biochim. Biophys. Acta 2015 (1854) 677–686.
[29] M.P. Pavlou, E.P. Diamandis, I.M. Blasutig, The long journey of cancer biomarkers from the bench to the clinic, Clin. Chem. 59 (2013) 147–157.
[30] M.S. Pepe, Z. Feng, H. Janes, P.M. Bossuyt, J.D. Potter, Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design, J. Natl. Cancer Inst. 100 (2008) 1432–1438.
[31] J. Bayani, E.P. Diamandis, The physiology and pathobiology of human kallikrein-related peptidase 6 (KLK6), Clin. Chem. Lab. Med. 50 (2012) 211–233.
[32] A. Anisowicz, G. Sotiropoulou, G. Stenman, S.C. Mok, R. Sager, A novel protease homolog differentially expressed in breast and ovarian cancer, Mol. Med. 2 (1996) 624–636.
[33] G. Pampalakis, G. Sotiropoulou, Multiple mechanisms underlie the aberrant expression of the human kallikrein 6 gene in breast cancer, Biol. Chem. 387 (2006) 773–782.
[34] M.C. Ghosh, L. Grass, A. Soosaipillai, G. Sotiropoulou, E.P. Diamandis, Human kallikrein 6 degrades extracellular matrix proteins and may enhance the metastatic potential of tumour cells, Tumour Biol. 25 (2004) 193–199.
[35] G. Pampalakis, E. Prosnikli, T. Agalioti, A. Vlahou, V. Zoumpourlis, G. Sotiropoulou, A tumor-protective role for human kallikrein-related peptidase 6 in breast cancer mediated by inhibition of epithelial-to-mesenchymal transition, Cancer Res. 69 (2009) 3779–3787.
[36] P. Ehrenfeld, L. Manso, M.F. Pavicic, C.E. Matus, C. Borquez, A. Lizama, et al., Bioregulation of kallikrein-related peptidases 6, 10 and 11 by the kinin B(1) receptor in breast cancer cells, Anticancer Res. 34 (2014) 6925–6938.
[37] V. Kulasingam, E.P. Diamandis, Proteomics analysis of conditioned media from three breast cancer cell lines: a mine for biomarkers and therapeutic targets, Mol. Cell. Proteomics 6 (2007) 1997–2011.
[38] M.V. Sepporta, F.M. Tumminello, C. Flandina, M. Crescimanno, M. Giammanco, M. La Guardia, et al., Follistatin as potential therapeutic target in prostate cancer, Target. Oncol. 8 (2013) 215–223.
[39] H.A. Loomans, C.D. Andl, Intertwining of activin a and TGFbeta signaling: dual roles in cancer progression and cancer cell invasion, Cancer 7 (2014) 70–91.
[40] E. Bloise, H.L. Couto, L. Massai, P. Ciarmela, M. Mencarelli, L.E. Borges, et al., Differential expression of follistatin and FLRG in human breast proliferative disorders, BMC Cancer 9 (2009) 320.
[41] C.A. Harrison, P.C. Gray, W.W. Vale, D.M. Robertson, Antagonists of activin signaling: mechanisms and potential biological applications, Trends Endocrinol. Metab. 16 (2005) 73–78.
[42] E. Gold, G. Risbridger, Activins and activin antagonists in the prostate and prostate cancer, Mol. Cell. Endocrinol. 359 (2012) 107–112.
[43] G.P. Risbridger, S.L. Mellor, S.J. McPherson, J.F. Schmitt, The contribution of inhibins and activins to malignant prostate disease, Mol. Cell. Endocrinol. 180 (2001) 149–153.
[44] K.P. Chang, H.K. Kao, Y. Liang, M.H. Cheng, Y.L. Chang, S.C. Liu, et al., Overexpression of activin a in oral squamous cell carcinoma: association with poor prognosis and tumor progression, Ann. Surg. Oncol. 17 (2010) 1945–1956.
[45] M.S. Pepe, C.I. Li, Z. Feng, Improving the quality of biomarker discovery research: the right samples and enough of them, Cancer Epidemiol. Biomark. Prev. 24 (2015) 944–950.