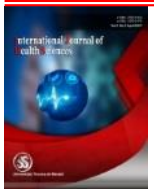


International Journal of Health Sciences

Available online at www.sciencescholar.us
Vol. 9 No. 3, December 2025, pages: 961-980
E-ISSN: 2550-696X
https://doi.org/10.53730/ijhs.v9n3.15835

Proteomic Mapping Associated with Clinical and Sociodemographic Parameters of Breast Cancer in Senegal

Malick FALL [a], Moussa CAMARA [b], Omar NIANG [c], Fatimata MBAYE [d], Ahmadou DEM [e], Sidy KA [f], Mamadou Aliou DIALLO [g], Julie HARDOUIN [h], Silly TOURE [i], Ndongo DIA [j], Mbacké SEMBENE [k], Pascal COSETTE [l], Emmanuel CORNILLOT [m]

Manuscript submitted: 27 September 2025, Manuscript revised: 18 November 2025, Accepted for publication: 08 December 2025

Corresponding Author [a]

Keywords

Biomarkers;
Breast cancers;
clinical
sociodemographic;
Proteomics;
Senegal;

Abstract

Breast cancer refers to a malignant tumour resulting from the uncontrolled proliferation of epithelial cells in the mammary gland. It is the leading cause of cancer in women. In Senegal, regional disparities remain marked by differences in access to screening, diagnosis, and treatment. Proteomics provides a direct reflection of the functional state of tissues and biological pathways and captures the functional effects of molecular alterations. In order to better understand the relationship between the pathogenesis of breast cancer and the existence of potential biomarkers based on each underlying clinical and sociodemographic parameter, this study performs correlation analyses. Proteins were extracted from healthy and cancerous tissues. The analytical workflow showed 30 proteins that were statistically deregulated between those under and over 50 years of age, 5 proteins between married and unmarried patients, 37 proteins between women with fewer than 7 children and those with more than 7 children, six proteins between the early stage and the locally advanced stage, and treatment response showed that 17 proteins were statistically deregulated. The results of this study have identified numerous proteins with high prognostic value associated with robust statistics and significantly overexpressed according to the parameters.

[a] Department of Animal Biology, Faculty of Science and Technology, Cheikh Anta Diop University, Dakar, Senegal
[b] Department of Animal Biology, Faculty of Science and Technology, Cheikh Anta Diop University, Dakar, Senegal
[c] Department of Animal Biology, Faculty of Science and Technology, Cheikh Anta Diop University, Dakar, Senegal
[d] Department of Animal Biology, Faculty of Science and Technology, Cheikh Anta Diop University, Dakar, Senegal
[e] Cancer Institute, Faculty of Medicine, Pharmacy and Stomatology, Cheikh Anta Diop University, Dakar, Senegal
[f] Cancer Institute, Faculty of Medicine, Pharmacy and Stomatology, Cheikh Anta Diop University, Dakar, Senegal
[g] Department of Animal Biology, Faculty of Science and Technology, Cheikh Anta Diop University, Dakar, Senegal
[h] UMR CNRS 6270, PISSARO Proteomic Platform (IRIB), University of Rouen, 76821 Mont-Saint-Aignan, France
[i] Department of Maxillofacial Surgery and Stomatology University Hospital Center Aristide le Dantec
[j] Institute Pasteur Dakar, Senegal, BP220
[k] Department of Animal Biology, Faculty of Science and Technology, Cheikh Anta Diop University, Dakar, Senegal
[l] UMR CNRS 6270, PISSARO Proteomic Platform (IRIB), University of Rouen, 76821 Mont-Saint-Aignan, France
[m] Institute of Computational Biology (IBC), Cancer Research Institute of Montpellier (IRCM - INSERM U1194), Montpellier Regional Cancer Institute (ICM) & University of Montpellier, France

Contents

1 Introduction

Breast cancer refers to a malignant tumour resulting from the uncontrolled proliferation of epithelial cells in the mammary gland. It is the leading cause of cancer in women (Smolarz *et al.,* 2022). According to the GLOBOCAN report, in 2022, this cancer surpassed lung cancer as the most frequently diagnosed cancer worldwide, with approximately 2.3 million new cases and an estimated 667,000 deaths (Sung *et al.,* 2021). In Africa, breast cancer is the most common cancer among women. An estimated 198,553 new cases and 91,252 deaths were reported on the continent in 2022, with a high mortality/incidence ratio reflecting late diagnosis and more limited access to specialized care. In Senegal, there were 11,841 new cases of cancer across all sites in 2022. Among women, breast cancer ranked second among all cancers, accounting for more than 24% of new cases (1,838) and 12% of deaths (976) (Ferlay *et al.,* 2021). Beyond this burden, regional disparities remain marked by differences in access to screening, diagnosis, and treatment (Castaldi *et al.,* 2022). These inequalities necessitate integrative approaches capable of linking tumour biology to clinical and social contexts. This reality poses the challenge of integrating these vast amounts of data to accurately predict complex pathophysiology and translate this complexity into clinically actionable insights (Heo *et al.,* 2021; Kidd *et al.,* 2015). These figures illustrate the need for practical molecular tools to improve prevention, early detection, and management. This burden is expected to increase further by 2040, with rapid growth anticipated in sub-Saharan Africa if screening and treatment capacities are not further expanded (Ngwa *et al.,* 2022; Hidig & Kitaghenda, 2025).

Proteomics studies proteins, their abundance, modifications, and interactions on a large scale. It provides a direct reflection of the functional state of tissues and biological pathways. Unlike genomics, which mainly provides information on potential, proteomics captures the actual activity of signalling networks and the functional effects of molecular alterations (Al-Amrani *et al.,* 2021). Modern platforms (LC-MS/MS, label-based or label-free quantification) make it possible to map thousands of proteins and analyse their variations according to clinical characteristics, paving the way for the identification of targets and disease signatures (Megger *et al.,* 2013).

In this context, protein biomarkers play a central role in precision medicine. They can be diagnostic (early detection), prognostic (risk stratification), and predictive (probability of response to treatment) (Bedore *et al.,* 2024). Recent studies show that clinical proteomics, including real-time proteomics, can refine therapeutic decisions, demonstrating its growing maturity in oncology (AlDoughaim *et al.,* 2024).

This study is part of that effort. We are analysing the correlations between tumour protein abundance and clinical and sociodemographic parameters (age, marital status, number of children, stage, response to treatment) to identify proteins associated with these dimensions. Our goal is to identify biomarkers for

stratification, response prediction, and, ultimately, improved care pathways in the Senegalese and African context.

2 Materials and Methods

This study involves performing correlation analyses based on data from a previous differential proteomics study. The complete methodology used to obtain this database is summarized below.

Study population and ethics
Healthy and cancerous tissue samples were collected at the Juliot Curie Institute, Aristide le Dantec Hospital, Dakar, Senegal. They were sent to the molecular biology laboratory at the Institute for Research and Development (IRD). Clinical and pathological data were obtained from patient records, and this study was approved by the ethics committee of Cheikh Anta Diop University. We confirm that written informed consent was obtained from the donor or their relatives for the use of this data.

Protein extraction
Proteins were extracted from healthy and cancerous tissues, then suspended in a lysis buffer containing 7 M urea, 4 M thiourea, 4% Chaps, 25 mM Tris, 200 mM DTT, and 500 µl protease inhibitors (Roche) at room temperature for 4 hours. The proteins in the supernatants were collected and quantified using the Bradford method.

Label Free
25 µg of protein from healthy and cancerous tissue were separated by 7% SDS-PAGE. Bands of interest were excised, the proteins were digested with trypsin (Promega), and the peptide fractions were recovered. Mass spectrometry analyses were performed using a linear ion trap/Orbitrap hybrid mass spectrometer (LTQ nanochromatography system (Easy-nLC II, Thermo Scientific).

Protein identification and peptide quantification
The raw data files were processed using Proteome Discoverer 1.3 software (Thermo Scientific), and a list of peaks was identified using the Mascot search engine (version 2.2, Matrix Science) against the Swiss Prot Homo sapiens database (V55.6; 390,696 sequences). Raw files (.raw) were processed using ProGenesis LC-MS software (Nonlinear Dynamics; V4.0.4441.29989) to generate a 2D peptide mass map showing retention time (RT) versus m/z.

Statistical analyses
In this study, analyses were performed using **R** software (R.4.4.2 2021.10.31) and **RStudio** (V.2024.09.1). The main packages used were *limma, dplyr, ggplot2*, and *openxlsx*, with reproducible scripts applied identically to each clinical variable. The significance threshold was $p<0.05$.

Data and organisation
The raw proteomic data were reorganized into two tables: an abundance matrix (rows = proteins, columns = samples) and a clinical table (1 row per sample) containing the variables of interest. Sample identifiers were harmonized between the two tables. For more in-depth analysis, proteins with a p-value <0.05 were selected for the characteristic studied. Finally, proteins with an FC >2 were targeted as significantly overexpressed for the parameter considered.

Definition of clinical variables
The variables were analysed separately: age (<50 vs. ≥50 years), number of children (<7 vs. ≥7), marital status (married vs. unmarried (single and widowed)), stage (early (I–II) vs. locally advanced (III)), and response to treatment (100% vs. <100%).

Differential analysis (limma)
For each variable, we fitted a linear model per protein using *limma* (R), coding the variable of interest as a factor and constructing the required contrast.
The moderated statistics from *limma* (moderated t-test, eBayes) provided for each protein logFC (log2 difference in means between groups in the direction of the contrast); p-value, and q-value (*Benjamini–Hochberg*, adj.P.Val column). The directional fold change was calculated as FC = 2^logFC. For each contrast, we reported a complete table (all proteins) and significant lists (q<0.05, logFC sign) as well as Top 10 (sorted by q then |logFC|), ready for insertion.

Correlations and principal component analysis (PCA)
For each parameter, we selected the most informative proteins (Top related to the parameter according to *limma*) and calculated protein-protein correlations (Pearson) from log2 values (visualized in upper triangular matrices). An exploratory principal component analysis (PCA) was performed on these same proteins (centered-reduced data), with visualization in a correlation circle (variable map).

3 Results and Discussions

3.1 Results

In the first part of this study, we filtered all proteins that were identified with at least two peptides. This highlighted statistically deregulated proteins (p-value < 0.05), considering each of the five parameters included in the study (**Table 1**).

For the age parameter, we found 30 proteins that were statistically deregulated between those under and over 50 years of age. Among these proteins, 22 were overexpressed in postmenopausal women, compared with only 8 in women under 50. Considering marital status, only 5 proteins were statistically deregulated between married and unmarried patients. Among these proteins, 3 were overexpressed in unmarried women and 2 in married women.

Analysis of the number of children per patient showed that 37 proteins were statistically deregulated between women with fewer than 7 children and those with more than 7 children. Thus, more than 89% of these proteins (n=33) were overexpressed in women with fewer than 7 children. As for the stage of cancer, we found six proteins that were statistically deregulated between the early stage and the locally advanced stage, with five proteins overexpressed in patients at the early stage.

Finally, the filters applied to the quantitative analysis of treatment response showed that 17 proteins were statistically deregulated, all of which were overexpressed in patients who achieved an optimal response to treatment.

Table 1
List of all statistically deregulated proteins for each of the clinical-pathological and sociodemographic parameters

Proteins	All proteins	Proteins *p<005*	Total proteins *p<005*
]0;50 years old]	225	8	
[50 years old;+…[	343	22	30
Unmarried	430	3	
Married	138	2	5
Fewer than 7 children	411	33	
More than 7 children	157	4	37
Early stage (I and II)	443	5	
Locally advanced stage (III)	125	1	6
Suboptimal response (<100%)	122	0	
Optimal response (100%)	446	17	17

For better visualization, we generated a heat map (**Figure 1**) representing the normalized expression levels of the most differentially expressed proteins. Each column corresponds to a patient, and each row to an identified protein. The color indicates the relative intensity of expression (red = overexpression, blue = under expression, yellow = intermediate expression). Hierarchical clustering, applied to proteins and patients, highlights distinct groupings according to protein profiles. The clinical annotations above the columns indicate treatment response (Yes/No), tumour stage (0–3), number of children (0–8), marital status (married, single, widowed), and age (40–80 years). This figure suggests the existence of specific protein signatures associated with therapeutic response and tumour stage, while sociodemographic variables appear to be less decisive.

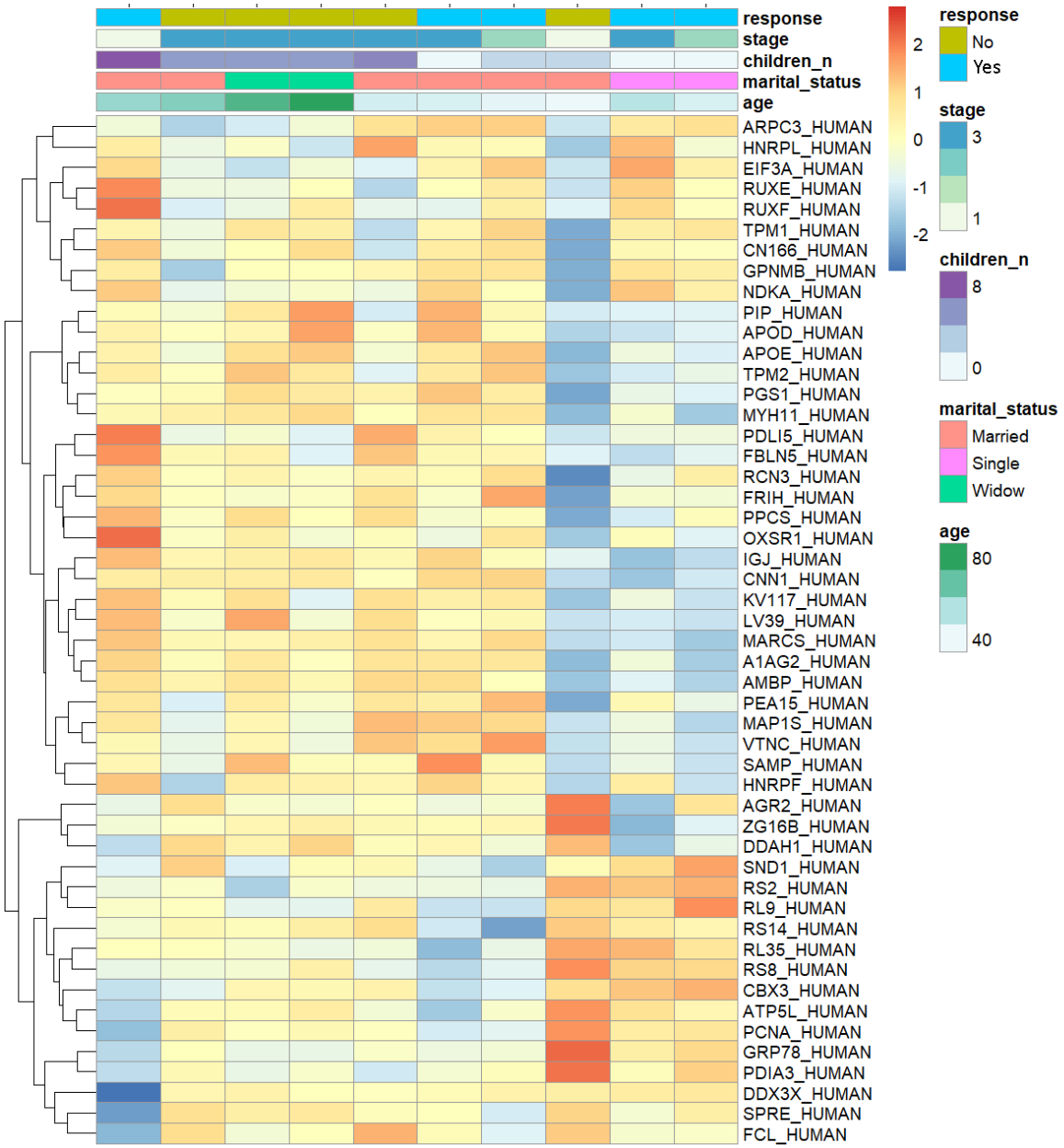


Figure 1. Heat map of the most discriminating proteins according to clinical and sociodemographic parameters. Heat map representing normalised expression levels (Z-score per protein) of the proteins showing the greatest differences between responders and non-responders. Each column corresponds to a

patient, and each row to an identified protein. The colour indicates the relative intensity of expression (red = overexpression, blue = under expression, yellow = intermediate expression). Hierarchical clustering applied to proteins highlights distinct groupings according to protein profiles. The clinical annotations above the columns indicate treatment response (Yes/No), tumour stage (0–3), number of children (0–8), marital status (married, single, widowed), and age (40–80 years).

Choice of the number of factorial axes for principal component analysis (PCA)

In this analysis, the scatter plots of the eigenvalues for all five patient parameters considered revealed that the break (Elbow) after the first two axes is followed by a steady decline from the third axis onwards (**Figure 2**). This allows the first two dimensions (Dim 1 and Dim 2) to be retained for principal component analysis (PCA) because they explain more than 50% of the information for almost all parameters. The first ACP factorial plan contains the maximum amount of information and provides the best quality representation for explaining the protein profile of patients based on clinical, pathological, and sociodemographic factors.

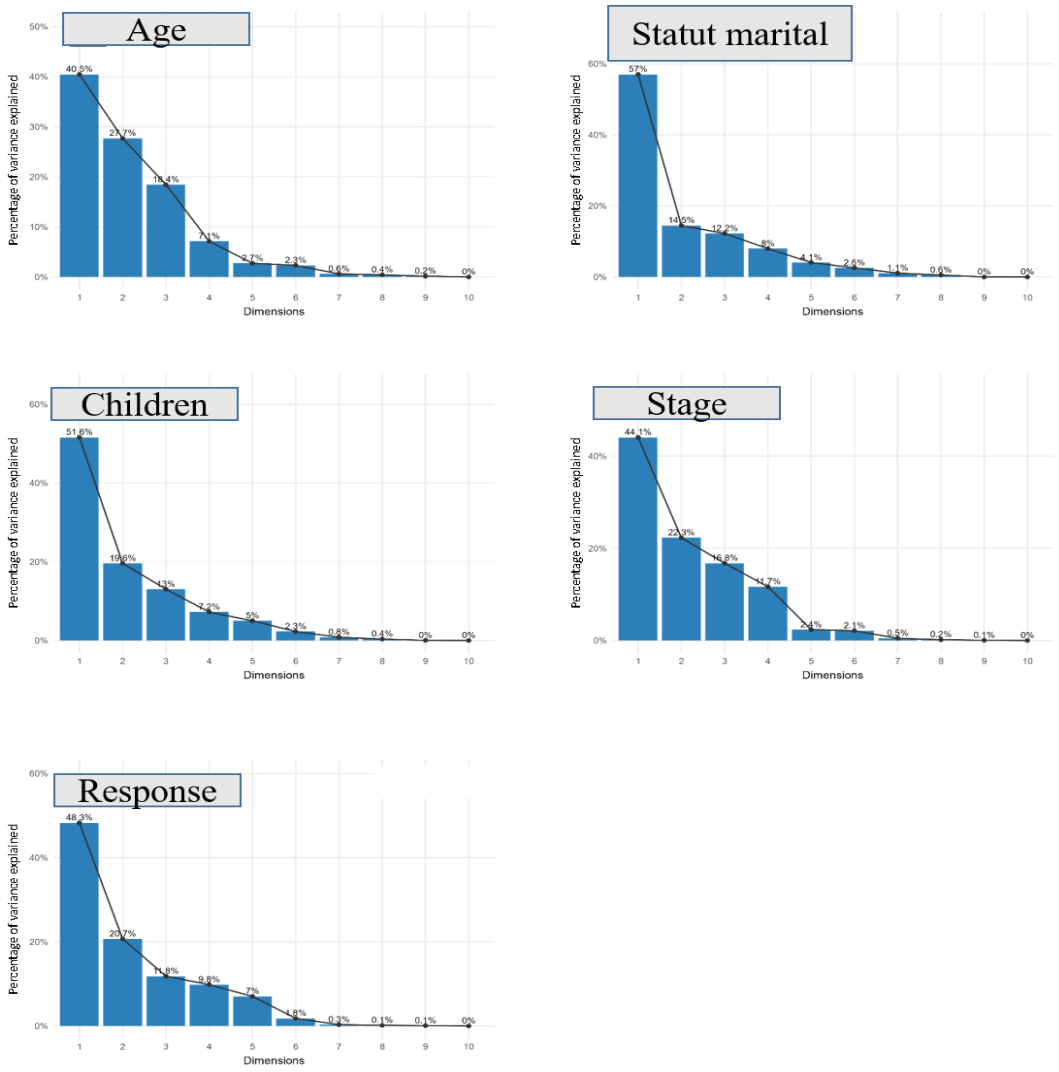


Figure 2. Scree plots of the eigenvalues of all clinical and sociodemographic parameters. The eigenvalue score revealed that the drop-off (elbow) after the first two axes was followed by a steady decrease from the third axis onwards for all parameters.

Study of protein profiles according to patient age

By selecting proteins that were statistically deregulated between patients under 50 and those over 50 and whose fold change was greater than 2, we found that all proteins (n=22) were significantly overexpressed in postmenopausal women (**Table 2**).

Table 2
List of proteins significantly overexpressed according to age ($p < 0.05$; FC>2)

Protein_id	*p* value	Fold_change	Highest mean condition
UBC9_HUMAN	9.E-04	866.52	≥50
PIP_HUMAN	3.E-03	894.20	≥50
ZO1_HUMAN	5.E-03	22.91	≥50
LYAG_HUMAN	6.E-03	671.93	≥50
NLTP_HUMAN	6.E-03	452.57	≥50
MATR3_HUMAN	6.E-03	89.55	≥50
ABHDA_HUMAN	7.E-03	702.97	≥50
ILF2_HUMAN	1.E-02	34.50	≥50
CHM4B_HUMAN	1.E-02	106.95	≥50
RABP2_HUMAN	1.E-02	33.56	≥50
RBBP7_HUMAN	2.E-02	234.56	≥50
LKHA4_HUMAN	2.E-02	2757.38	≥50
RAB10_HUMAN	2.E-02	217.47	≥50
HMGB2_HUMAN	2.E-02	113.48	≥50
SMD3_HUMAN	3.E-02	15.35	≥50
UBA1_HUMAN	3.E-02	11.55	≥50
IFM1_HUMAN	3.E-02	91.87	≥50
CBPB1_HUMAN	4.E-02	25.97	≥50
SMD2_HUMAN	4.E-02	7.29	≥50
AATM_HUMAN	4.E-02	15.10	≥50
DX39A_HUMAN	4.E-02	6.09	≥50
VDAC3_HUMAN	4.E-02	11.45	≥50

Correlations between age-related proteins

PCA on age-associated proteins shows that the first two axes explain 68.2% of the variance (Dim1 = 40.5%, Dim2 = 27.7%) (**Figure 3A**). Axis 1 (Dim1) captures a major age-related gradient. The proteins UBC9_HUMAN, CHM4B_HUMAN, and MATR3_HUMAN (long red arrows) load positively on this axis, indicating that they co-vary and contribute strongly to the age signal. In contrast, MAP1S_HUMAN and, to a lesser extent, CRP_HUMAN point in the opposite direction, suggesting a trend opposite to that of the previous group. On axis 2 (Dim2), RL10A_HUMAN, HNRPM_HUMAN, and SC22B_HUMAN (marked contributions) stand out, adding a complementary dimension to the age signal within the same set. Finally, the proteins GPNMB_HUMAN and CBPB1_HUMAN (short/blue vectors) are less well represented by these two axes and contribute less to the main structure.

The correlation matrix (Pearson) calculated for the proteins most associated with age shows a strongly correlated module observed in particular between the proteins MATR3_HUMAN, CHM4B_HUMAN, and UBC9_HUMAN (**Figure 3B**). We also note a strong correlation between RL10A_HUMAN and HNRPM_HUMAN. These proteins vary in concert in the samples (dark blue discs), suggesting a shared biological process related to aging.

Conversely, MAP1S_HUMAN shows marked negative correlations with this module (red discs), indicating an opposite trend: when the proteins in the module increase or decrease with age, MAP1S follows the opposite direction. CRP_HUMAN shows more modest negative correlations, indicating partial involvement in the age-related process.

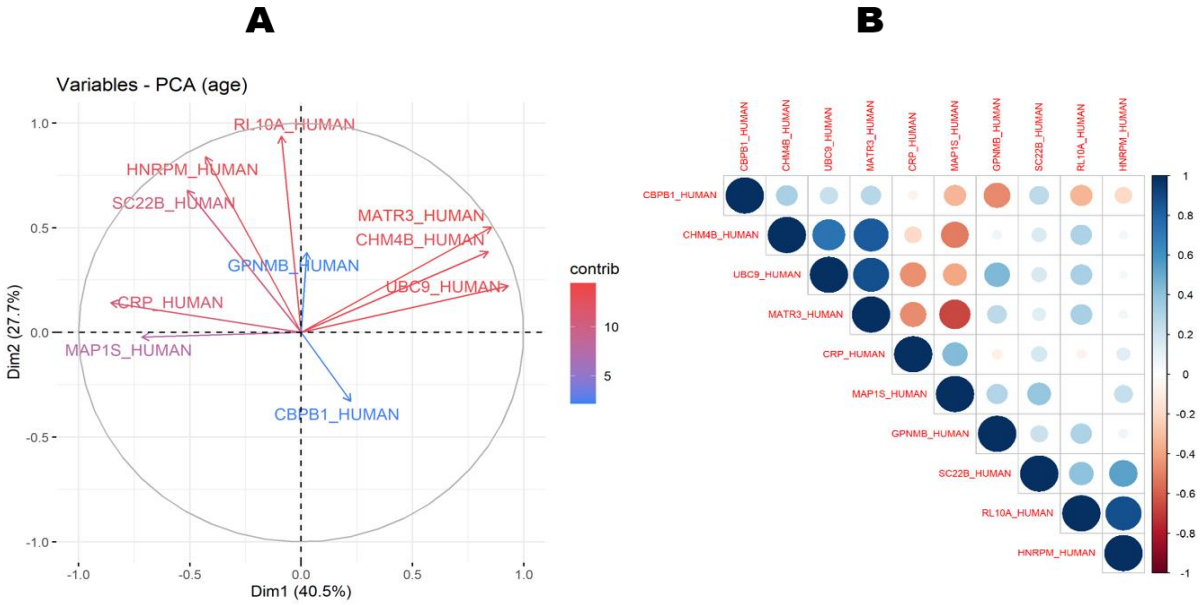


Figure 3. Correlation between proteins associated with patient age. Correlation circle of principal component analysis with high-contribution variables in red arrow and low-contribution variables in blue arrow. Proximities between variables are interpreted in terms of correlation (**A**). Correlation matrix with Positive correlations is shown in blue and negative correlations in orange. Color intensity and circle size are proportional to correlation coefficients (**B**).

Study of protein profiles according to patients' marital status

By selecting proteins that were statistically deregulated between married and unmarried patients and had a fold change greater than 2, we found that both target proteins were significantly overexpressed in married women (**Table 3**).

Table 3
List of proteins significantly overexpressed according to marital status ($p < 0.05$; FC>2)

Protein_id	p value	Fold_change	Highest mean condition
ALS_HUMAN	1.E-02	46.71	Married
SPTB1_HUMAN	4.E-02	6.26	Married

Correlations between proteins associated with marital status

PCA applied to proteins associated with marital status shows that the first two axes explain 71.5% of the variance (Dim1 = 57.0%, Dim2 = 14.5%) (**Figure 4A**). Axis 1 (Dim1) carries most of the signal and contrasts a ribosomal/nuclear group (RL35_HUMAN, RS23_HUMAN, RL9_HUMAN, RL10A_HUMAN, CBX3_HUMAN (long arrows to the right, strong contribution) to a secretory/adhesion-immune group (IGJ_HUMAN, SPTB1_HUMAN, BCAM_HUMAN) oriented to the left. This suggests two antagonistic signatures depending on marital status. Axis 2 (Dim2) mainly distinguishes ZG16B_HUMAN (high contribution, upward) and, to a lesser extent, PIP_HUMAN.

This trend is confirmed by the correlation matrix (Pearson, upper triangle) on the proteins most associated with marital status, revealing two antagonistic modules (**Figure 4B**): the proteins RS23_HUMAN, RL35_HUMAN, RL10A_HUMAN, RL9_HUMAN, and CBX3_HUMAN are strongly positively correlated (dark blue). And the proteins PIP_HUMAN, IGJ_HUMAN, ZG16B_HUMAN, SPTB1_HUMAN, and BCAM_HUMAN also show marked negative correlations between them.

These results suggest that the distinction between married and unmarried women is accompanied by a proteomic axis contrasting ribosomal/nuclear signatures with secretory/adhesion/immune signatures.

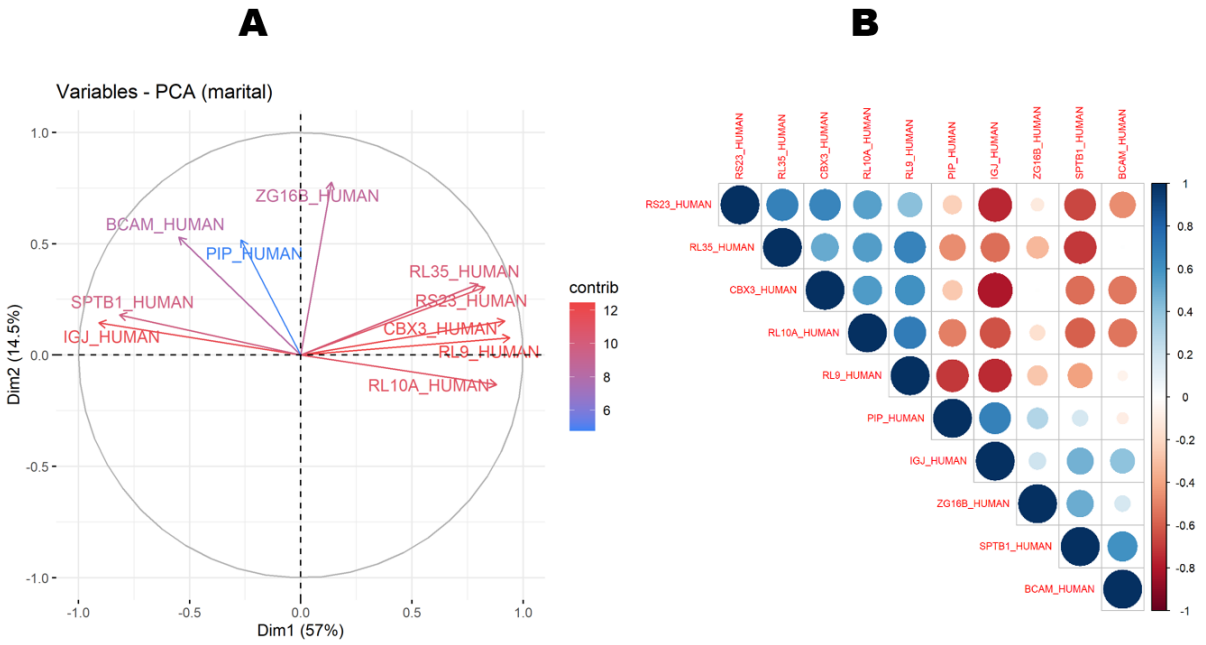


Figure 4. **Correlations between proteins associated with patients' marital status.** Correlation circle of principal component analysis **(A)** and Correlation matrix **(B).**

Study of protein profiles based on the number of children of patients

By selecting proteins that were statistically deregulated between patients with fewer than 7 children and those with more than 7 children and whose fold change was greater than 2, we also found that all 4 targeted proteins were significantly overexpressed in patients with more than 7 children (**Table 4**).

Table 4
List of proteins significantly overexpressed according to the number of children ($p < 0.05$; FC>2)

Protein_id	Fold_change	p value	Highest mean condition
FBLN5_HUMAN	25.42	1.E-02	Children ≥7
GSHB_HUMAN	11.46	4.E-02	Children ≥7
GLNA_HUMAN	9.55	2.E-02	Children ≥7
RAN_HUMAN	7.45	3.E-02	Children ≥7

Correlations between proteins associated with the number of children of patients

PCA on proteins associated with the number of children shows that the first two axes explain 71.2% of the variance (Dim1 = 51.6%, Dim2 = 19.6%) (**Figure 5A**). Axis 1 (Dim1) clearly separates two antagonistic sets: FBLN5_HUMAN, PDLI5_HUMAN, HNRPL_HUMAN, GPNMB_HUMAN (long arrows on the right) vs. UBC9_HUMAN, EWS_HUMAN, GANAB_HUMAN (arrows on the left). These groups vary in opposite directions within the gradient related to the number of children. Axis 2 (Dim2) mainly reinforces GPNMB_HUMAN and HNRPL_HUMAN (upward arrows), indicating a complementary dimension of variation beyond the main axis. We also found proteins that are poorly represented by these two axes: A1AG2_HUMAN and CRP_HUMAN. These proteins are close to the origin (short arrows), suggesting a weaker contribution to the main structure.

The matrix (Pearson, upper triangle) highlights two coherent subsets (**Figure 5B**): EWS_HUMAN, UBC9_HUMAN, and GANAB_HUMAN on the one hand, and FBLN5_HUMAN with CRP_HUMAN and

A1AG2_HUMAN to a lesser extent on the other, show strong positive correlations between them (dark blue discs), suggesting that they vary in concert according to the category studied (number of children).

Conversely, we note that the proteins CRP_HUMAN, FBLN5_HUMAN, A1AG2_HUMAN, and PPCS_HUMAN form a second group that is negatively correlated with each of the proteins EWS_HUMAN, UBC9_HUMAN, and GANAB_HUMAN. This indicates opposite behaviour within the gradient related to the number of children.

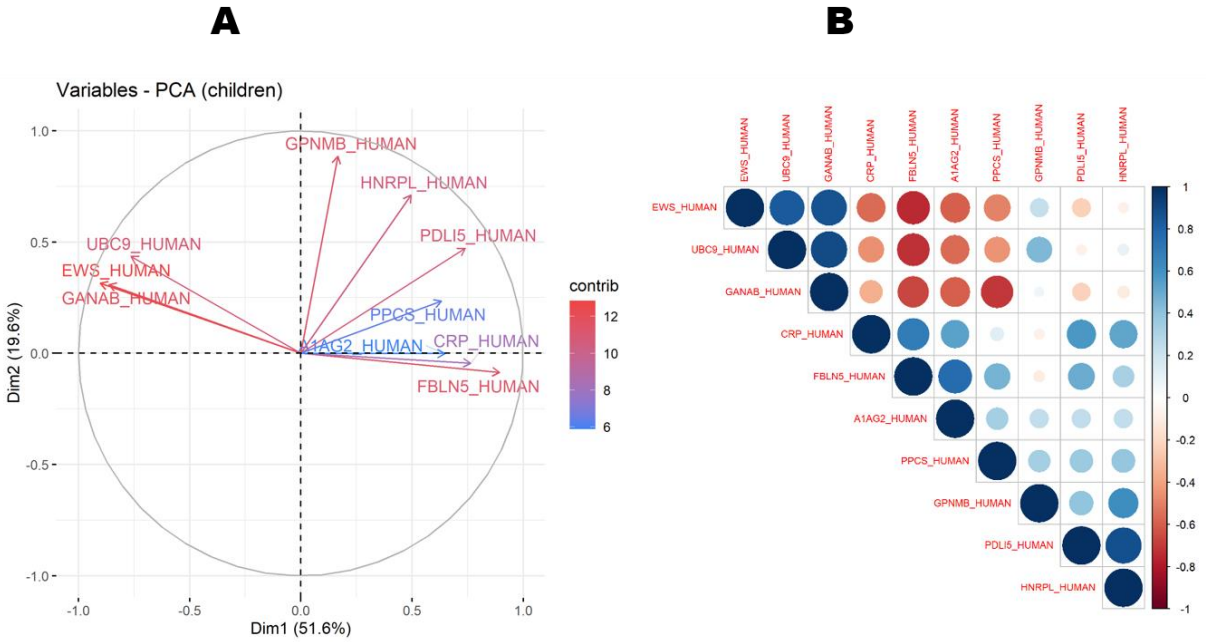


Figure 5. **Correlations between proteins associated with the number of children of patients.** Correlation circle of principal component analysis **(A)** and Correlation matrix **(B).**

Study of protein profiles according to cancer stage

By selecting proteins that were statistically deregulated between the different stages of cancer and whose fold change was greater than 2, we also found that only one protein was significantly overexpressed, and that was at the locally advanced stage (**Table 5**).

Table 5
List of proteins significantly overexpressed according to cancer stage ($p < 0.05$; FC>2)

Protein_id	p value	Fold_change	Highest mean condition
IC1_HUMAN	7.E-03	3.37	Locally advanced (III)

Correlations between proteins associated with cancer stage

PCA applied to stage-associated proteins explains 66.4% of the variance (Dim1= 44.1%, Dim2= 22.3%) (**Figure 6A**). Axis 1 (Dim1) clearly contrasts two signatures: on the right, CRP_HUMAN, PGS1_HUMAN, ANGT_HUMAN, A1AG2_HUMAN, and MAP1S_HUMAN (long arrows, strong contribution), and on the left, LYAG_HUMAN and NLTP_HUMAN. These groups vary inversely along the stage gradient. Axis 2 (Dim2) reinforces the opposition between CRP_HUMAN (downward-oriented) and LYAG_HUMAN / NLTP_HUMAN (upward-oriented), adding a complementary dimension beyond Dim1. The shorter vectors (CBPQ_HUMAN, GPNMB_HUMAN and HNRPL_HUMAN) are less well represented by these two axes.

The correlation matrix (Pearson, upper triangle) highlights two strongly correlated and antagonistic modules (**Figure 6B**): PGS1_HUMAN, ANGT_HUMAN, A1AG2_HUMAN, and MAP1S_HUMAN show high positive

correlations between them (dark blue discs), indicating robust co-variation within the cohort. Conversely, all these proteins are negatively correlated with the proteins LYAG_HUMAN and NLTP_HUMAN.

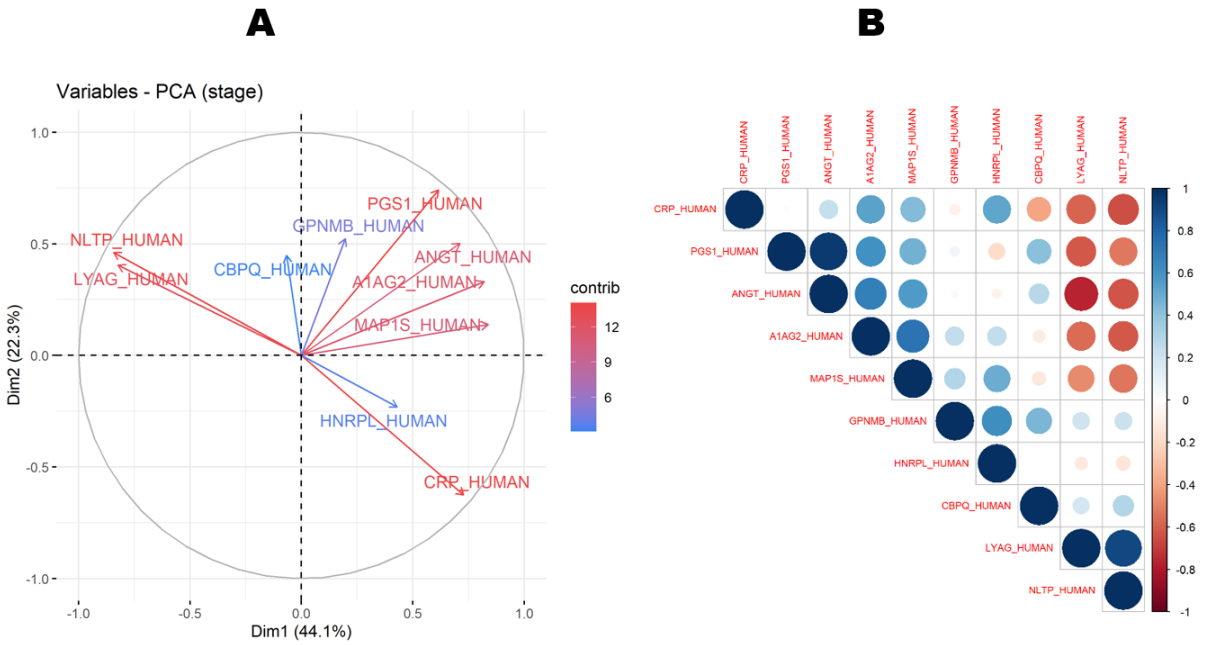


Figure 6. **Correlations between proteins associated with cancer stage.** Correlation circle of principal component analysis **(A)** and Correlation matrix **(B).**

Study of protein profiles based on response to treatment

By selecting proteins that were statistically deregulated between treatment responses and had a fold change greater than 2, we also found that all targeted proteins (n=17) were significantly overexpressed in patients with an optimal response to treatment (**Table 6**).

Table 6
List of proteins significantly overexpressed based on response to treatment ($p < 0.05$; FC>2)

Protein_id	p value	Fold_change	Highest mean condition
HTRA1_HUMAN	2.E-02	231.12	Response 100%
EIF3A_HUMAN	9.E-03	110.64	Response 100%
RUXF_HUMAN	2.E-02	102.14	Response 100%
GPNMB_HUMAN	4.E-04	62.25	Response 100%
RL10_HUMAN	3.E-02	38.88	Response 100%
ARPC3_HUMAN	2.E-02	10.78	Response 100%
COCA1_HUMAN	1.E-02	6.28	Response 100%
RUXE_HUMAN	1.E-02	6.03	Response 100%
TPM1_HUMAN	2.E-02	4.77	Response 100%
NDKA_HUMAN	3.E-03	4.42	Response 100%
RCN3_HUMAN	2.E-02	4.19	Response 100%
RHOA_HUMAN	4.E-02	3.81	Response 100%
TAGL2_HUMAN	3.E-02	3.58	Response 100%
PEA15_HUMAN	2.E-02	3.29	Response 100%
IQGA1_HUMAN	2.E-02	3.21	Response 100%
CN166_HUMAN	4.E-02	3.16	Response 100%
CNN3_HUMAN	4.E-02	2.76	Response 100%

Correlations between proteins associated with treatment response

PCA performed on response-associated proteins explains 69.0% of the variance (Dim1= 48.3%, Dim2=20.7%) (**Figure 7A**). Axis 1 (Dim1) contrasts two antagonistic signatures: on the right, APOE_HUMAN, PGS1_HUMAN and A1AG2_HUMAN, and on the left, SPRE_HUMAN, DDAH1_HUMAN. Axis 2 (Dim2) mainly highlights PIP_HUMAN (long arrow pointing upwards) and, in the opposite quadrant, SPRE/DDAH1. This suggests a complementary dimension of variation beyond the main opposition captured by Dim1. The proteins APOE_HUMAN, PGS1_HUMAN, PIP_HUMAN (coloured redder) make the strongest contributions to the axes; vectors close to the origin (PEA15/A1AG2/MAP1S) are less well represented on both dimensions.

The correlation matrix (Pearson, upper triangle) highlights two strongly correlated and opposing modules (**Figure 7B**). The first group, consisting of the proteins PIP_HUMAN, APOE_HUMAN, GPNMB_HUMAN, PEA15_HUMAN, A1AG2_HUMAN, and MAP1S_HUMAN, shows high positive correlations between them (dark blue discs), indicating robust co-variation related to the response. The second group, consisting of the proteins ZG16B_HUMAN, SPRE_HUMAN, and DDAH1_HUMAN, is strongly correlated with each other and negatively correlated with the first module (red discs between groups).

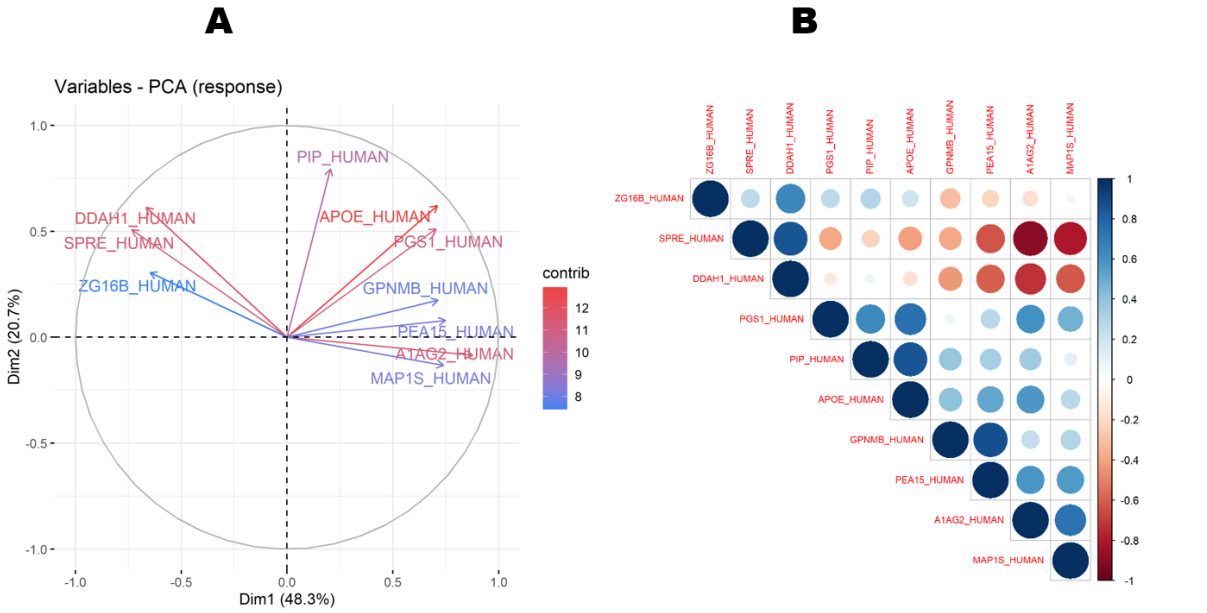


Figure 7. **Correlations between proteins associated with treatment response.** Correlation circle of principal component analysis **(A)** and Correlation matrix **(B).**

3.2 Discussions

This work is one of the few studies combining a differential proteomics approach with the clinical, pathological, and sociodemographic characteristics of breast cancer in Senegalese patients. The high mortality rate caused by these cancers requires the development of effective diagnostic and prognostic biomarkers to improve patient care and long-term survival. To this end, to better understand the relationship between the pathogenesis of breast cancer and the existence of potential biomarkers based on each underlying clinical and sociodemographic parameter, stricter criteria ($p < 0.02$ and $FC > 2$) were used to statistically evaluate the deregulated proteins in this study (Dalman *et al.,* 2012). For each parameter selected, proteins with a good signature (best p-values and fold change) will be discussed.

Thus, considering the age of the patients, 22 proteins were targeted, including ubiquitin-conjugating enzyme 9 (UBC9_HUMAN: p=9.04; FC=866.52) showed a better signature. This enzyme is involved in many cellular functions, and its dysregulation has been associated with various malignant tumours. In breast, bladder, and stomach cancers, UBC9 overexpression is closely linked to tumour cell proliferation and invasion (Huang *et*

al., 2020; Wang *et al.,* 2025). Our findings on the overexpression of the UBC9 protein in tumour tissue according to age are consistent with recent studies, which have also shown that overexpression of this protein, associated with a poor prognosis in breast, lung, and pancreatic cancer, is significantly correlated with patient age (Li *et al.,* 2025; Zhang *et al.,* 2025). All of the above-mentioned studies confirm that this protein is a good diagnostic and prognostic biomarker.

We also found a good protein signature for matrin 3 (MATR3_HUMAN: p=6.03; FC=89.55). This is an abundant nuclear protein with DNA and RNA binding domains, which in breast cancer could potentially be a prognostic biomarker for predicting better survival in patients. Indeed, the MATR3 protein inhibits tumourigenicity, induces cell death by apoptosis, and inhibits the migration and invasion of basal breast cancer cells (Yang *et al.,* 2020). However, other studies have shown that the MATR3 protein has a greater influence on the progression and metastasis of breast cancer and liver cancer. This shows a correlation between this protein and a poor prognosis (Luo *et al.,* 2023; Xiao *et al.,* 2024). The same is true for lung cancer, where the MATR3 protein is significantly overexpressed in cancerous tissue and associated with poorer overall patient survival. It is involved in carcinogenesis and tumour progression (Durślewicz *et al.,* 2022).

Correlative analysis of protein abundance based on the marital status of patients enabled us to identify two proteins that were significantly overexpressed in married women. The labile acid subunit of the insulin-like growth factor-binding protein complex (ALS_HUMAN: p=1.02; FC=46.71) scored highly. This protein is synthesised mainly by plasma cells in the liver and is capable of binding to insulin-like growth factors. It is therefore implicated in liver cancer (Minghui *et al.,* 2019). Various studies conducted on liver cancer corroborate our findings on ALS protein overexpression. In these studies, the ALS protein was identified as being involved in tumour prognosis and progression in liver tissue. It is also associated with a poor prognosis. This allows it to be classified as a potential biomarker and prognostic indicator for liver cancer (Chen *et al.,* 2021; Xu *et al.,* 2024).

In addition, the erythrocyte spectrin beta chain (SPTB1_HUMAN: p=4.02; FC=6.26) also had a good protein signature. This protein is important for determining cell shape and the functional arrangement of transmembrane proteins. Its overexpression is sometimes linked to tumour suppression, but in most studies it plays a pro-oncogenic role, particularly in colorectal, gastric, lung, breast, and skin cancers (Lin *et al.,* 2021; Sreeja *et al.,* 2020). Other studies on breast and liver cancer have corroborated our findings that overexpression of the SPTB1 protein is associated with a poor prognosis. This protein is involved in numerous signalling pathways linked to tumour development and metastasis (Huang *et al.,* 2020; Wu *et al.,* 2021).

By correlating protein abundance with the number of children the patients had, we identified four proteins that were significantly overexpressed in patients with more than seven children. Among these proteins was fibulin-5 (FBLN5_HUMAN: p=1.02; FC=25.42). This protein, also known as DANCE, EVEC, or UP50, is a secreted 66 kDa glycoprotein belonging to the fibulin family, whose members play an essential role in tissue growth and development. It thus contributes to creating a microenvironment conducive to tumour growth, but also stimulates various mechanisms that can hinder its progression, demonstrating its tumour-promoting and protective functions. However, studies have shown that the FBLN5 protein is involved in tumour proliferation in breast and cervical cancers (Mohamedi *et al.,* 2016, 2019). Researchers have recently highlighted the association between the FBLN5 protein and differentiation and prognosis in gastric and breast cancers. These authors have suggested the FBLN5 protein as a prognostic biomarker and therapeutic target because its overexpression is associated with low survival rates in patients (Choi *et al.,* 2025; Karanis *et al.,* 2019).

In this study, correlation analysis of protein abundance according to cancer stage identified a specific protein, namely plasma C1 protease inhibitor (IC1_HUMAN: p=7.03; FC=3.37). This protein, also known as Serpin G1, is a reference biomarker for diagnosis and prognosis in numerous studies on cancers, particularly prostate cancer and oropharyngeal cancer. This protein was significantly overexpressed in cancerous tissue and associated with a poor prognosis (Hsieh *et al.,* 2025; Shen *et al.,* 2025; Zebene *et al.,* 2024). Our findings on IC1 protein overexpression are corroborated by a study on lung cancer. In this study, overexpression of our protein of interest was significantly associated with metastasis (Yuan *et al.,* 2025).

Finally, this study aims to correlate protein abundance with response to treatment. The analyses identified 17 proteins with a good signature, including transmembrane glycoprotein NMB (GPNMB_HUMAN: p=4.04; FC=62.25). This is a type 1 transmembrane protein involved in the malignant progression of various cancers,

including melanoma, glioma, and breast cancer. Its high expression is considered an unfavourable prognostic factor. It can promote primary tumour growth and metastasis, and its expression is correlated with shorter recurrence-free survival in patients (Lazaratos *et al.,* 2022; Manevich *et al.,* 2022). Our findings on GPNMB protein overexpression in patients with optimal response to treatment confirm the work of Biondini et al. (2022), also focusing on breast cancer. These researchers have shown that GPNMB levels increase in response to standard and experimental treatments for several subtypes of breast cancer. Khan *et al.* (2021), also showed in a study on lung cancer that the IC1 protein was associated with treatment response in patients.

4 Conclusion

This work is one of the few proteomic studies correlating the clinical, pathological, and sociodemographic parameters of Senegalese breast cancer patients with the proteins overexpressed in cancerous tissue. By linking tumour biology to clinical and social determinants, our objective is to identify biomarkers for age, marital status, number of children, and stratification and prediction of response to treatment in a Senegalese and African context. The results of this study have identified numerous proteins with high prognostic value associated with robust statistics and significantly overexpressed according to the parameters. Some of these proteins have already been reported in many cancers, and the next step would be to evaluate the expression of these proteins in cancerous tissues using orthogonal ELISA and immunohistochemistry methods to validate their potential as breast cancer biomarkers.

Acknowledgments
We are grateful to two anonymous reviewers for their valuable comments on the earlier version of this paper.

References

Al-Amrani, S., Al-Jabri, Z., Al-Zaabi, A., Alshekaili, J., & Al-Khabori, M. (2021). Proteomics: Concepts and applications in human medicine. *World journal of biological chemistry*, *12*(5), 57-69.

AlDoughaim, M., AlSuhebany, N., AlZahrani, M., AlQahtani, T., AlGhamdi, S., Badreldin, H., & Al Alshaykh, H. (2024). Cancer biomarkers and precision oncology: A review of recent trends and innovations. *Clinical Medicine Insights: Oncology*, *18*, 11795549241298541.

Bedore, S., van der Eerden, J., Boghani, F., Patel, S. J., Yassin, S., Aguilar, K., & Lokeshwar, V. B. (2024). Protein-Based Predictive Biomarkers to Personalize Neoadjuvant Therapy for Bladder Cancer—A Systematic Review of the Current Status. *International journal of molecular sciences*, *25*(18), 9899.

Biondini, M., Kiepas, A., El-Houjeiri, L., Annis, M. G., Hsu, B. E., Fortier, A. M., ... & Siegel, P. M. (2022). HSP90 inhibitors induce GPNMB cell-surface expression by modulating lysosomal positioning and sensitize breast cancer cells to glembatumumab vedotin. *Oncogene*, *41*(12), 1701-1717.

Castaldi, M., Smiley, A., Kechejian, K., Butler, J., & Latifi, R. (2022). Disparate access to breast cancer screening and treatment. *BMC Women's Health*, *22*(1), 249.

Chen, W., Desert, R., Ge, X., Han, H., Song, Z., Das, S., ... & Nieto, N. (2021). The matrisome genes from hepatitis B–related hepatocellular carcinoma unveiled. *Hepatology Communications*, *5*(9), 1571-1585.

Choi, J., Kwak, Y., Park, M., Jo, J. Y., Kang, J. H., Myeong-Cherl, K., ... & Cho, S. J. (2025). Cancer-associated fibroblast-derived fibulin-5 promotes epithelial–mesenchymal transition in diffuse-type gastric cancer via cAMP response element-binding protein pathway, showing poor prognosis. *Experimental & Molecular Medicine*, 1-14.

Dalman, M. R., Deeter, A., Nimishakavi, G., & Duan, Z. H. (2012). Fold change and p-value cutoffs significantly alter microarray interpretations. *BMC bioinformatics*, *13*(Suppl 2), S11.

Durślewicz, J., Klimaszewska-Wiśniewska, A., Jóźwicki, J., Antosik, P., Kozerawski, K., Grzanka, D., & Braun, M. (2022). Prognostic significance of MATR3 in stage I and II non-small cell lung cancer patients. *Journal of Cancer Research and Clinical Oncology*, *148*(12), 3313-3322.

Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D. M., Piñeros, M., Znaor, A., & Bray, F. (2021). Cancer statistics for the year 2020: An overview. International journal of cancer, 149(4), 778-789.

Heo, Y. J., Hwa, C., Lee, G. H., Park, J. M., & An, J. Y. (2021). Integrative multi-omics approaches in cancer research: from biological networks to clinical subtypes. *Molecules and cells*, *44*(7), 433-443. https://doi.org/10.14348/molcells.2021.0042

Hsieh, C. C., Wu, Y. H., Chen, Y. L., Wang, C. I., Li, C. J., Liu, I. H., ... & Chen, C. Y. (2025). SERPING1 Reduces Cell Migration via ERK-MMP2-MMP-9 Cascade in Sorafenib-Resistant Hepatocellular Carcinoma. *Environmental toxicology*, *40*(2), 318-327.

Huang, J., Tang, Y., Zou, X., Lu, Y., She, S., Zhang, W., ... & Hu, H. (2020). Identification of the fatty acid synthase interaction network via iTRAQ-based proteomics indicates the potential molecular mechanisms of liver cancer metastasis. *Cancer Cell International*, *20*(1), 332.

Huang, X., Tao, Y., Gao, J., Zhou, X., Tang, S., Deng, C., ... & Li, T. (2020). UBC9 coordinates inflammation affecting development of bladder cancer. *Scientific Reports*, *10*(1), 20670.

Karanis, M., Koksal, H., Ates, E., & Dogru, O. (2019). Clinical importance of fibulin-5 immunohistochemical staining in breast lesions. *Polish Journal of Pathology*, *70*(4), 259-263.

Khan, S. A., Sun, Z., Dahlberg, S., Malhotra, J., Keresztes, R., Ikpeazu, C., ... & Pillai, R. (2021). Efficacy and safety of glembatumumab vedotin in patients with advanced or metastatic squamous cell carcinoma of the lung (PrECOG 0504). *JTO clinical and research reports*, *2*(5), 100166. https://doi.org/10.1016/j.jtocrr.2021.100166

Kidd, B. A., Readhead, B. P., Eden, C., Parekh, S., & Dudley, J. T. (2015). Integrative network modeling approaches to personalized cancer medicine. Personalized medicine, 12(3), 245-257.

Lazaratos, A. M., Annis, M. G., & Siegel, P. M. (2022). GPNMB: a potent inducer of immunosuppression in cancer. *Oncogene*, *41*(41), 4573-4590.

Li, F., Dai, Y., Tang, C., Peng, L., Huang, H., Chen, Y., ... & Lin, Y. (2025). Elevated UBC9 expression and its oncogenic role in colorectal cancer progression and chemoresistance. *Scientific Reports*, *15*(1), 9123.

Lin, L., Chen, S., Wang, H., Gao, B., Kallakury, B., Bhuvaneshwar, K., ... & He, A. R. (2021). SPTBN1 inhibits inflammatory responses and hepatocarcinogenesis via the stabilization of SOCS1 and downregulation of p65 in hepatocellular carcinoma. *Theranostics*, *11*(9), 4232.

Luo, D., Liang, Y., Wang, Y., Ye, F., Jin, Y., Li, Y., ... & Yang, Q. (2023). Long non-coding RNA MIDEAS-AS1 inhibits growth and metastasis of triple-negative breast cancer via transcriptionally activating NCALD. *Breast Cancer Research*, *25*(1), 109.

Manevich, L., Okita, Y., Okano, Y., Sugasawa, T., Kawanishi, K., Poullikkas, T., ... & Kato, M. (2022). Glycoprotein NMB promotes tumor formation and malignant progression of laryngeal squamous cell carcinoma. *Cancer Science*, *113*(9), 3244-3254.

Megger, D. A., Bracht, T., Meyer, H. E., & Sitek, B. (2013). Label-free quantification in clinical proteomics. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, *1834*(8), 1581-1590. https://doi.org/10.1016/j.bbapap.2013.04.001

Minghui, R., Malecela, M. N., Cooke, E., & Abela-Ridder, B. (2019). WHO's Snakebite Envenoming Strategy for prevention and control. *The Lancet Global Health*, *7*(7), e837-e838.

Mohamedi Munarriz, Y., Fontanil López, T., Cobo Díaz, T., Vega Álvarez, J. A., Cobo Plana, J. M., Pérez Basterrechea, M., ... & Cal Miguel, S. J. (2019). Antitumor potential of Fibulin-5 in breast cancer cells depends on its RGD cell adhesion motif. *Cellular physiology and biochemistry: international journal of experimental cellular physiology, biochemistry, and pharmacology*.

Mohamedi, Y., Fontanil, T., Solares, L., Garcia-Suarez, O., Garcia-Piqueras, J., Vega, J. A., ... & Obaya, A. J. (2016). Fibulin-5 downregulates Ki-67 and inhibits proliferation and invasion of breast cancer cells. *International journal of oncology*, *48*(4), 1447-1456.

Ngwa, W., Addai, B. W., Adewole, I., Ainsworth, V., Alaro, J., Alatise, O. I., ... & Kerr, D. (2022). Cancer in sub-Saharan Africa: a lancet oncology commission. *The Lancet Oncology*, *23*(6), e251-e312.

Hidig, S. M., & Kitaghenda, F. K. (2025). Next-generation Cancer Screening and Prevention: The Convergence of Liquid Biopsy Diagnostics. *Artificial Intelligence and Precision Medicine. Compr Cancer Detect Prev*, *1*(1), 001-005.

Shen, Y., Dong, X., Li, X., Shi, Z., Shao, T., Jiang, J., & Song, J. (2025). WNT inhibitor SP5-mediated SERPING1 suppresses lung adenocarcinoma progression via TSC2/mTOR pathway. *Cell Death & Disease*, *16*(1), 103.

Smolarz, B., Nowak, A. Z., & Romanowicz, H. (2022). Breast cancer—epidemiology, classification, pathogenesis and treatment (review of literature). *Cancers*, *14*(10), 2569.

Sreeja, J. S., John, R., Dharmapal, D., Nellikka, R. K., & Sengupta, S. (2020). A fresh look at the structure, regulation, and functions of fodrin. *Molecular and Cellular Biology*, *40*(17), e00133-20.

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, *71*(3), 209-249.

Wang, Q., Li, S., Xu, Y., Chen, Y., Xu, C., He, Q., ... & Lin, Y. (2025). UBC9 overexpression promotes proliferation and metastasis in gastric cancer via ATF2. *World Journal of Surgical Oncology*, *23*(1), 270.

Wu, H., Chen, S., Liu, C., Li, J., Wei, X., Jia, M., ... & Zhi, X. (2021). SPTBN1 inhibits growth and epithelial-mesenchymal transition in breast cancer by downregulating miR-21. *European journal of pharmacology*, *909*, 174401. https://doi.org/10.1016/j.ejphar.2021.174401

Xiao, Z., Chen, H., Xu, N., Chen, Y., Wang, S., & Xu, X. (2024). MATR3 promotes liver cancer progression by suppressing DHX58–mediated type I interferon response. *Cancer Letters*, *604*, 217231. https://doi.org/10.1016/j.canlet.2024.217231

Xu, L., Xiong, L., Chen, Y., Chen, J., Liu, X., Xu, Y., ... & Xu, X. (2024). IGFALS suppresses hepatocellular carcinoma progression by stabilizing PPAR-γ. *International Immunopharmacology*, *143*, 113414. https://doi.org/10.1016/j.intimp.2024.113414

Yang, J., Lee, S. J., Kwon, Y., Ma, L., & Kim, J. (2020). Tumor suppressive function of Matrin 3 in the basal-like breast cancer. *Biological research*, *53*.

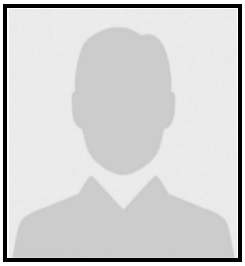
Yuan, Y., Jiang, H., Xue, R., Feng, X. J., Liu, B. F., Li, L., ... & Zhang, X. E. (2025). Identification of a Biomarker Panel in Extracellular Vesicles Derived From Non-Small Cell Lung Cancer (NSCLC) Through Proteomic Analysis and Machine Learning. *Journal of Extracellular Vesicles*, *14*(5), e70078.

Zebene, E. D., Lombardi, R., Pucci, B., Medhin, H. T., Seife, E., Di Gennaro, E., ... & Woldemichael, G. B. (2024). Proteomic Analysis of Biomarkers Predicting Treatment Response in Patients with Head and Neck Cancers. *International Journal of Molecular Sciences*, *25*(23), 12513.

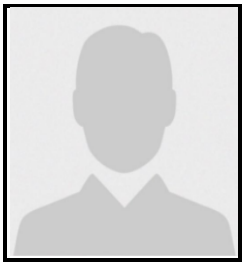
Zhang, H., Wu, J., Hu, H., Tang, H., Tan, K., Hu, M., & Zhu, G. (2025). UBC9: a novel therapeutic target in papillary thyroid carcinoma. *Journal of Endocrinological Investigation*, 1-19.

Biography of Authors

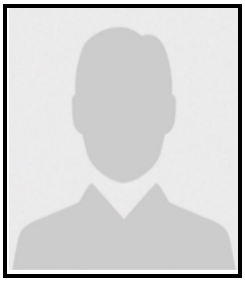
Malick FALL (Corresponding author)
Department of Animal Biology, Faculty of Science and Technology, Cheikh Anta Diop University, Dakar, Senegal
Email: malick.fall@ucad.edu.sn; malickfal@yahoo.fr



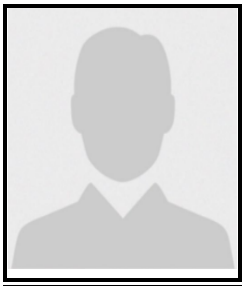
Moussa CAMARA
Department of Animal Biology, Faculty of Science and Technology, Cheikh Anta Diop University, Dakar, Senegal
Email: camara.msg@gmail.com



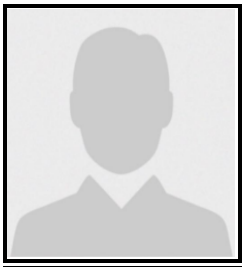
Omar NIANG
Department of Animal Biology, Faculty of Science and Technology, Cheikh Anta Diop University, Dakar, Senegal
Email: contact.niang@gmail.com



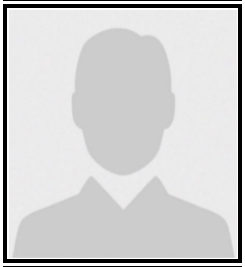
Fatimata MBAYE
Department of Animal Biology, Faculty of Science and Technology, Cheikh Anta Diop University, Dakar, Senegal
Email: fatimata2.mbaye@ucad.edu.sn



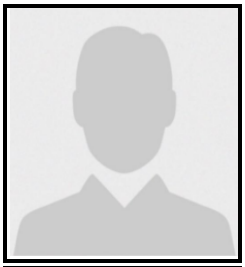
Ahmadou DEM
Cancer Institute, Faculty of Medicine, Pharmacy and Stomatology, Cheikh Anta Diop University, Dakar Senegal
Email: ahmadou1.dem@ucad.edu.sn



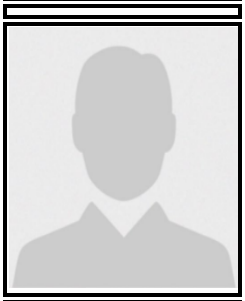
Sidy KA
Cancer Institute, Faculty of Medicine, Pharmacy and Stomatology, Cheikh Anta Diop University, Dakar Senegal
Email: sidy.ka@ucad.edu.sn



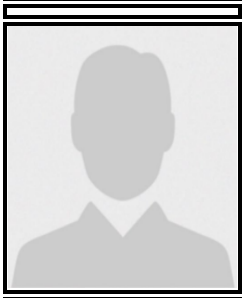
Mamadou Aliou DIALLO
Department of Animal Biology, Faculty of Science and Technology, Cheikh Anta Diop University, Dakar, Senegal
Email: hbadral@gmail.com



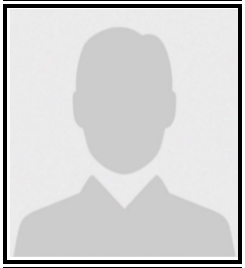
Julie HARDOUIN
UMR CNRS 6270, PISSARO Proteomic Platform (IRIB), University of Rouen, 76821 Mont-Saint-Aignan, France
Email: julie.hardouin@univ-rouen.fr



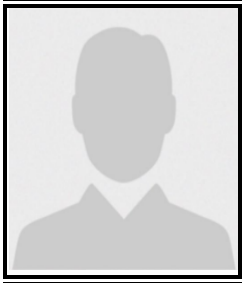
Silly TOURE
Department of Maxillofacial Surgery and Stomatology University Hospital Center Aristide le Dantec
Email: silly.toure@ucad.edu.sn



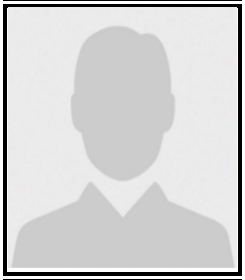
Ndongo DIA
Institute Pasteur Dakar, Senegal, BP220
Email: ndongo.dia@pasteur.sn



Mbacké SEMBENE
Department of Animal Biology, Faculty of Science and Technology, Cheikh Anta Diop University, Dakar, Senegal
Email: mbacke.sembene@ucad.edu.sn



Pascal COSETTE
UMR CNRS 6270, PISSARO Proteomic Platform (IRIB), University of Rouen, 76821 Mont-Saint-Aignan, France
Email: pascal.cosette@univ-rouen.fr



Emmanuel CORNILLOT
Institute of Computational Biology (IBC), Cancer Research Institute of Montpellier (IRCM - INSERM U1194), Montpellier Regional Cancer Institute (ICM) & University of Montpellier, France
Email: emmanuel.cornillot@umontpellier.fr