Journal Homepage: -www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI: 10.21474/IJAR01/20771
DOI URL: http://dx.doi.org/10.21474/IJAR01/20771

RESEARCH ARTICLE

PROTEOMIC PROFILES AND CLINICOPATHOLOGICAL FACTORS OF OROPHARYNGEAL CANCER IN SENEGAL.

Camara Moussa[1], Cissokho Boubacar[2], Ndiaye Ndeye Aissatou[3], Guillou Clément[4], Toure Silly[5], Ndiaye Mouhamadou Makhtar[5], Sembene M backé[2], Cosette Pascal[6] and Fall Malick[1]

1. Laboratory of General Parasitology, Department of Animal Biology, Faculty of Sciences and Technology, UniversitéCheikh Anta Diop, Dakar-Sénégal.
2. Genomics Laboratory, Department of Animal Biology, Faculty of Science and Technology, Université Cheikh Anta Diop, Dakar-Senegal.
3. Animal Physiology laboratory, Department of Animal Biology, Faculty of Science and Technology, Université Cheikh Anta Diop, Dakar-Sénégal.
4. INSERM US 51, CNRS UAR 2026, HeRacLeS PISSARO, UniversitédeRouen, France.
5. Department of Stomatology and Maxillofacial Surgery, Aristide Le Dantec Hospital, Dakar-Sénégal.
6. INSA Rouen, CNRS, PBS UMR 6270, Université de Rouen, France.

Manuscript Info

Manuscript History
Received: 17 Febuary 2025
Final Accepted: 21 March 2025
Published: April 2025

Key words:-
Proteomics; clinicopathological; Biomarkers; Oral cancers; Senegal

Abstract

Background: The early diagnosis of cancers around the world represents a major challenge. Among the most common cancers, those affecting theupper aerodigestive tract (UADT) are on the rise in Senegal with a very high mortality rate. To decipher the mechanisms associated with protein profile and clinicopathological parameters, we performed a correlation study of tumor samples collectedfrom Senegalese patients.

Methods: We referred to quantitative proteomics to detect protein profiles.And for explore the relationships between proteins and clinicopathological parameters, correlation analysis was performed. A correlation matrix was generated using Pearson's correlation coefficient.

Results: The correlative analyses allowed to identify for the gender parameter, 44 proteins statistically deregulated between men and women. Considering patient age, 23 proteins were significantly deregulated between patients under and over 50 years of age. Analysis of patients' tobacco consumption showed that 83 proteins were statistically deregulated between smokers and non-smokers. Regarding alcohol consumption, we found 28 proteins that were significantly deregulated between alcohol consumers and abstainers. Finally, quantitative analysis of tumor location showed that 76 proteins were statistically deregulated between cheek samples and all other locations

Conclusion: This study highlighted numerous proteins with high prognostic value that were significantly overexpressed according to clinicopathological parameters. Some of these proteins have already been reported in other cancers; however, to the first time (to our knowledge) we have associated them with oropharyngeal cancer.

Corresponding Author:-Camara Moussa
Address:-Laboratory of General Parasitology, Department of Animal Biology, Faculty of Sciences and Technology, UniversitéCheikh Anta Diop, Dakar-Sénégal.

Introduction:-

Cancer, one of the leading causes of death worldwide, is a major public health problem according to the World Health Organization. In 2022, an estimated 20 million cases of cancer and 9.7 million deaths were recorded(1). Among the most common oropharyngeal cancers, also known as cancers of the ENT (ear, nose and throat) sphere, are malignant tumors found in the mucous membranes of the upper aerodigestive tract (paranasal sinuses, pharynx, larynx, oral cavity, nostrils)(2–4). Etiological factors for oral cavity cancers include excessive tobacco and alcohol consumption, infection with the human papillomavirus (HPV) and other factors (5,6).

The early diagnosis of cancer remains a major challenge in patient management. The clinicopathological characteristics of oropharyngeal cancer, such as tumor location, age, sex and expression of biomarkers specific to cancer biology, are key factors in prognosis and therapeutic choice. Therefore, identification of these parameters is essential for the diagnosis of cancer(7).

The proteomic profile is defined as information on all the proteins produced by the organism at a given time, and can be used to detect and diagnose a disease and to assess the organism's response to treatment (8). Mass spectrometry (MS)-based proteomics is a powerful technology for detecting protein abundance levels in patients; therefore it is ideal tool for biomarker discovery (9). Thus, our initial mass spectrometry study of oropharyngeal tumors identified potential biomarkers. We analyzed the proteins expressed in the tumor and healthy tissues in the same patient. We found that 15 proteins couldbe considered potential biomarkers, and functional analysis suggested altered biological pathways and revealed the relevance of our protein profiling approach. However, this was done with the following proteins of interest: MMP9, GILT, BAG6, PPIF, CALD1, FXR1, SF3B2, ISG15, DDX46, SRSF1, MIME, PGS2, FRELP, RS9, and TARSH. These findings were not correlated with the clinicopathological characteristics of the patients. This correlation could inform us and lead us to anticipate therapy as a tool for early diagnosis of oropharyngeal cancers.

The aim of this study is to analyze and evaluate the association of the protein profile of cancerous tissues and their correlations with patients' clinical and pathological parameters to better understand the protein signatures that can be used to diagnose oropharyngeal cancers.

Materials and Methods

In this study, the database acquired during the initial differential proteomics analysis was used. The complete methodology used to obtain this database is summarized below.

Study Setting And Population

The study was conducted from December 2019 to February 2025 through collaboration between the Stomatology and Maxillofacial Surgery Department of Aristide Le Dantec Hospital in Dakar for patient recruitment, the General Parasitology Laboratory of the Animal Biology Department of the Faculty of Science and Technology of Cheikh Anta Diop University in Dakar for curation and statistical analyses, and the PISSARO Proteomic Platform of the University of Normandy for differential quantitative proteome analysis.The research protocol was approved by the Ethics Committee of Cheikh Anta Diop University under reference 0320/2019/CER/UCAD.

Patients referred to the Stomatology and Maxillofacial Surgery Department of Aristide Le Dantec Hospital for confirmation of the diagnosis of oral cancer were recruited after providing their consent. For each of the 20 patients, a biopsy representing the cancerous tissue was taken, as well as a small portion of healthy tissue for control purposes. Samples were immediately stored at +4°C before being frozen at -80°C within 24 h. Each patient completed a questionnaire providing information on age, sex, alcoholism, smoking, and tumor location.

Protein Extraction And Concentration

Protein extraction was performed as previously described by Chalfouh et *al.*(10)with slight modifications. Seven hundred and three hundred microliters (depending on tissue weight) of solubilization buffer (7M urea, 2M thiourea 2M, 20mM DTT, 2% CHAPS, 0.5%C7BzO, antiprotease,Sigma-Aldrich;Tributylphosphin 5µL, FlukaChemika, Zurich, Switzerland) was added to each sample inLysingMatrixD (Bio-Rad, Hercules, CA). The samples were homogenized in aFastPrep® device (grinding time=40s; speed=6m/s; centrifugation = 10000 g/7min). The supernatants were transferred to a clean 1.5ml microcentrifuge tube and stored at -20°C.

To estimate the protein concentration of each extract, Bradford method (Protein Assay Dye Reagent Concentrate) was used according to the manufacturer's instructions (Bio-Rad). The linear range of the BSA standard was 0.1-1 mg/mL. All samples were diluted 1:1, 1:5, and 1:15 and solubilization buffer was used as the negative control. The absorbance at 595nm was measured using a Perkin Elmer Wallac 1420 Victor. To validate the protein assays,

12.5 % sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed. Finally, silver nitrate staining was performed to determinethe protein profiles.

Enzymatic Digestion Of Protein Extracts

For each extract, 20 µg of protein was loaded onto a 7 % polyacrylamide gel (acrylamide/bis-acrylamide 30 % [29:1], Sigma-Aldrich) followed by migration (90 min at 10-20 mA/gel). After Coomassie Blue staining, the protein bands were excised and washed with H2O MilliQ before undergoing a reduction/alkylation process. After washing with acetonitrile, the gel bands were subjected to protein digestion with 1 µg of trypsin (Promega, Madison, WI, USA) solubilized in carbonate buffer (50mM ammonium carbonate, pH 8). After overnight incubation at 37°C, several peptide extractions performed using acetonitrile. Peptide fractions were solubilized in 20 µL H2O/formic acid [100:0.1]) and a peptide assay (Thermo Fisher) was performed to estimate peptide concentration according to the manufacturer's instructions.

Nano LC-MS/MS Analysis.

All samples were analyzed on a Q-exactive Pus OrbitrapTM mass spectrometer (Thermo Scientific) equipped with an Easy nanoLC II system (Thermo Scientific). The peptides were injected into an enrichment column (C18 Pepmap100, Thermo Scientific). The separation was performed using an analytical column needle (NTCC-360/100-5-153, Nikyo-Technos). The flow rate was 0.3 µL/min and the mobile phase was H2O/FA (100/0.1) [buffer A] and ACN/H2O/FA (80/20/0.1) [buffer B]. The elution gradient lasted 120 min: 0-106 min, 2-55 % B; 106-110 min, 55-100% B; 110-125 min, 100% B. The mass spectrometer operated in positive mode with HCD fragmentation. The capillary voltage was set at 1.9 kV, capillary temperature was 275°C, and m/z detection range was 400-1800. The resolutionswere 70000 and 17500 for MS1 and MS2 respectively. The ten most intense peptide ions were selected, and fragmentation occurred at a normalized collision energy of 28. Dynamic exclusion of the fragmented precursor ions was applied for 20 s.

Quantification.

For protein quantification, a label-free experiment was carried out as previously described by Escorcia et *al.*(11),with slight modifications. Briefly, after MS analysis, the raw data were imported into the Progenesis LC-MS software (Nonlinear Dynamics, Progenesis QI for Proteomics, Newcastle, UK). After alignment and normalization, statistical analysis were performed usingone-way ANOVA calculations.MS/MS spectra of selected peptides were exported for peptide identification with Mascot (Matrix Science, version 2.6.0). The results of the Mascot search were imported into Progenesis. For each growth condition, the total cumulative protein abundance was calculated by summing the peptide abundances. Proteins identified as having fewer than two peptides were discarded.

Statistical Analysis

All statistical analyses were performed using Excel software (version 2013) and R software (version 4.4.3). A one-way analysis of variance (ANOVA) was used for proteomic analysis. Libraries used with R software included *dplyr* for data cleaning, *corrplot* for the visualization of correlation matrices, and *Factoextra* and *Factominer* for principal component analysis.

Preparing and Cleaning Data

Onlyproteins with at least two peptides were used in the database obtained. Excel software was used to analyze and calculate the p-value, Fold Change (FC), and the group where the protein was most abundant for each clinicopathological feature of the study population. This made it possible to visualize theproteins that were statistically deregulated for each feature. Thus, for a more in-depth analysis, proteins with a p-value < 0.05 were retained for the characteristic under study. Finally, proteins with FC > 2 were targeted as significantly overexpressed for the parameter under consideration(12).

Correlation Analysis

To explore the relationships between proteins and clinicopathological parameters, correlation analysis was performed. The correlation matrix was generated using Pearson's correlation coefficient to measure the intensity and direction of the linear associations between different quantitative variables. The correlation matrix was calculated using the cor () function in R software, and results tabulated with coefficients rounded to two decimal places for clarity. For pairs of variables with significant correlation coefficients (p< 0.05), special attention was paid to better understand the interactions between these factors.

Principal component analysis (PCA)

Principal Component Analysis (Pca) Was Used To Explore The Data Structure And Reduce Protein Dimensionality. This Technique Was Used To Identify The Main Components Explaining The Variance In The Data, Particularly The Complex Relationships Among The Epidemiological Characteristics Of The Patients. When Data Are Represented In Different Planes, Proximities Between Variables Are Interpreted In Terms Of Correlation, Whereas Proximities Between Individuals Are Interpreted In Terms Of The Similarity Of The Observed Values.

Results.

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

For the sex parameter, we found 44 proteins that were statistically deregulated between men and women. Among these proteins, 31 were overexpressed in men, and only 13 were overexpressed in women.

Considering patient age, 23 proteins were significantly deregulated between patients under and over 50 years of age. Among these proteins, 18 were overexpressed in patients aged < 50 years and 5 in those aged > 50 years.

Analysis of patients' tobacco consumption showed that 83 proteins were statistically deregulated between smokers and non-smokers, and over 79 % of these proteins (n=66) were overexpressed in smokers, compared to only 17 in non-smokers.

Regarding alcohol consumption, we found 28 proteins statistically deregulated between alcohol consumers and abstainers, with over 85 % of these proteins (n=24) overexpressed in alcohol drinkers. Finally, quantitative analysis of tumor location showed that 76 proteins were significantly deregulated between cheek samples and all other locations. Of these proteins, 58 were overexpressed in the cheek, compared to 18 that were overexpressed in other tumor sites.

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

For all five patient parameters considered in this study, 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 (Figure 1). This enabled us to retain the first two dimensions (Dim 1 and Dim 2) for the principal component analyses (PCA), as they explain approximately 50 % of the information for almost all parameters. The first factorial plane of PCAs contains the maximum amount of information and provides the best quality of representation to explain the protein profile of patients as a function of clinicopathological factors.

Study Of Protein Profile As A Function Of Patient Gender

By selecting proteins statistically deregulated between males and females with a fold change greater than 2, we obtained 15 significantly overexpressed proteins, including 11 proteins overexpressed in males and four in females (Table 2).

PCA, contribution of variables according to patient gender

The first two dimensions of PCA explained 53.9 % of the total variance in proteins significantly overexpressed in the patient sex parameter (Figure 2A). Dim 1 explained 37.4 % of all these proteins. With regard to male sex, we noted a strong contribution and positive correlation between proteins such as HV373 (p=1.53E-02; FC=5.07), ATPB, AL3A2 and EHD1 on the one hand, and between proteins NUCB2 (p=7.85E-03; FC=2.34), FLNA and UBP14 on the other (Figure 2B). ARPC3, had little contribution.

Similarly, for females we noted a strong contribution and positive correlation between CCAR2, TRAP1 and AL1A1 proteins.

Study Of Protein Profile As A Function Of Patient Age

Taking patient age into account, quantitative analysis revealed that 10 proteins could be considered significantly overexpressed, of which seven proteins were overexpressed in patients under 50 and three in older patients (Table 3).

PCA, contribution of variables according to patient age

The PCA correlation circle showing the projection of proteins onto the first two dimensions (Dim 1 and Dim 2) explained a good proportion (57.7 %) of the total variance (Figure 3A). On Dim 1, APEX1 and CYB5 proteins were highly correlated, while FBLN5 protein was strongly correlated with the ZG16B, TPTC and RALA proteins (Figure 3B). These proteins also strongly contributedto this dimension. In contrast, CRK protein was negatively correlated with FBLN5 and ZG16B proteins. We also noted a weak contribution from the HV309, CTNA1 and STAT3 proteins.

Study Of Protein Profile As A Function Of Patients' Smoking Status

Quantitative analysis of proteins as a function of patients' smoking and non-smoking status and considering only proteins with a fold change greater than 2, revealed that 42 proteins were significantly overexpressed (Table 4). Of these, 36 were overexpressed in smokers, with two high-scoring proteins, GALT6 (p=1.73E-04; FC=79.21) and ITAM (p=1.47E-03; FC=18.47). In addition, six proteins were found to be overexpressed in non-smokers.

PCA, contribution of variables according to patients' smoking habits

The PCA correlation circle, showing the projection of proteins onto the first two dimensions (Dim 1 and Dim 2), explained 63.9 % of the total variance (Figure 4). A significant number of proteins with a fold change greater than two, obtained after quantitative analysis as a function of smoking, saturated the correlation circle. Nevertheless, GALT6, NUCB2, RD23A, IGKC and TAGL proteinscontributedsignificantlyto dimension 1, while PYR1, KLK10, SYIC and TITIN proteinscontributed strongly to both dimensions. We also noted a positive correlation between the GALT6 protein and the IGJ, RCN3, TAGL, HEXB and NUCB2 proteins (Supplementary Table 1). Finally, there was a weak contribution from the EHD1, PCY1A and ATLA3 proteins.

Study Of Protein Profile As A Function Of Patients' Alcohol Consumption Status

About patients' alcoholic and non-alcoholic status, quantitative protein analysis, considering only proteins with a fold change greater than 2, revealed that 24 proteins could be considered significantly overexpressed (Table 5). Of these proteins, 22 were overexpressed in alcohol consumers, with three proteins scoring high ($p < 0.01$ and FC > 5): CO7 (p = 9.72E-04; FC = 27.89), MYH7 (p = 5.90E-03; FC = 5.36) and RS7 (p = 9.84E-03; FC = 5.36). Only two proteins were overexpressed in non-alcoholics.

PCA, contribution of variables according to patients' alcohol status

The PCA correlation circle showing the projection of proteins onto the first two dimensions (Dim 1 and Dim 2) explained a good proportion (67.1 %) of the total variance (Figure 5A). Thus, in Dim 1 (which alone explained 55.5 % of the total variance), the CO7, PLMN, C4BPA, HSPB6 and COCA1 proteins made a strong contribution. In dimension 2, the JKIP3, TPP1 and RS7 proteins made a strong contribution. A positive correlation was noted between the CO7, HSPB6, COCA1, SYNC, PLMN, FA83H, TNPO1 and PNPT1 proteins (Figure 5B). In contrast, GLU2B level was negatively correlated with JKIP3, TPP1, MYH7 and UBA1 proteins.

Study Of Protein Profile As A Function Of Tumor Location In The Patient

Taking tumor location alone into account, quantitative analysis of statistically deregulated proteins with a fold change greater than 2 revealed that 27 proteins were significantly overexpressed. Of these, 18 were overexpressed at the cheek site and 9 at all other sites (Table 6).

PCA, contribution of variables according to tumor location

The PCA correlation circle, showing the projection of proteins on the first two dimensions (Dim 1 and Dim 2), explained a good proportion (49.2 %) of the total variance (Figure 6). Thus, RL5, THIO, and H2AY were highly correlated, with a strong contribution to both dimensions. In addition, MYL6B, MYG, MYL1, and ACT proteins were also correlated, but with a strong contribution to dimension 1 only. In contrast, all proteins overexpressed at the other sites had a low contribution to both dimensions.

PCA contribution of individuals

Using the proteins identified with qualitative criteria, a quantitative analysis was carried out to highlight the proteins that were statistically deregulated between the two populations constituting each parameter. Thus, principal component analysis revealed good discrimination between the two populations at the level of all parameters(Figure 7).

Amino acid composition of statistically deregulated proteins

By studying the amino acid composition of the highest-scoring proteins for all parameters (best p-value and fold-change scores: HV373, NUCB2, APEX1, ZG16B, GALT6, ITAM, CO7, MYH7, RL5 and H2AY proteins), we found that four amino acids were present in the majority. These were leucine (9.65 %), glutamate (8.29 %), lysine (8.06 %), and glycine (7.07 %) (Figure 8).

Discussion.

To the best of our knowledge, this is the first study to combine a proteomic approach with differential quantitative analysis of clinicopathological features in Senegalese patients with oral cancer. The high mortality caused by these cancers calls for the development of effective biomarkers for treatment planning and prognostic prediction, to improve patient management and long-term survival. To better understand the relationship between the pathogenesis of oral cancers and the existence of potential biomarkers according to each underlying clinicopathological

parameter, stricter criteria ($p < 0.02$ and FC > 2) were used to statistically evaluate the deregulated proteins in this study (13).For each parameter chosen, proteins with a good signature (best *p*- and fold-change scores) will be discussed.

Thus, considering the patients' gender characteristics, 15 proteins were targeted, among which the immunoglobulin heavy variable 3-73 (HV373: $p = 1.53E-02$; FC = 5.07) showed the best signature. HV373 participates in antigen recognition under normal conditions (14). However, studies have shown that these immunoglobulins are produced and present in tumor tissues in cancers of epithelial origin, including oropharyngeal, breast, liver, colon, cervical, thyroid, gastric, kidney, and bladder cancers(15,16). Our results showed overexpression of the HV373 protein and are in line with those of other researchers, who have found that overexpression of this protein is an indicator of tumor progression (16). This protein is involved in the proliferation, migration, invasion, resistance to apoptosis, and metastasis of these different types of cancer(17).These roles of the HV373 protein make it a good biomarker for the diagnosis and prognosis of disease progression, but no significant positive association with patient gender was found(16,17).

We also found a good protein signature for nucleobindin-2 (NUCB2: $p = 7.85E-03$; FC = 2.34). This protein was first identified in 2006 by Ohi et *al*. and its function in carcinogenesis was studied years later(18,19). Nucleobindin-2 is a precursor of nesfatin-1. As NUCB2 and nesfatin-1 are colocalized in each tissue, their expression is often analyzed together under the name NUCB2, which plays an important role in cancer development. NUCB2 overexpression is associated with poor prognosis in various cancers, including breast, prostate and colon cancers(20–22).These results are in line with those of Ming et *al*. and Suziki et *al*. showing that NUCB2 expression is associated with tumor growth, recurrence after remission, and metastases (23,24).Numerous studies have shown that the NUCB2 protein can be used as a potential biomarker of metastasis in breast cancer and proliferation in liver cancer(21,25,26).

Correlative analysis of the proteins with the patients' age parameters enabled us to target 10 proteins that were significantly overexpressed in patients aged < 50 and > 50 years. Apurinic-apyrimidine endonuclease 1/redox factor-1 (APEX1/Ref-1) was among those with high scores ($p = 1.36E-02$; FC = 2.73). It is a multifunctional protein involved in cell growth, transcriptional regulation, response to stress, and genome stability, and its deregulation is associated with cancer through the inhibition of apoptosis(27,28). Numerous studies have identified overexpression of the APEX1 in various cancers and this protein has been recommended as a prognostic biomarker, particularly for oropharyngeal cancer (29,30). In 2010, Zhang et *al*. showed that the APEX1 protein was linked to age. In 2013 and 2014, Sibylle et *al*. and Park et *al*. showed that the APEX1 protein is not only a prognostic factor in cancer but is also linked to age, since its reduction correlates with aging. These studies corroborate our findings concerning the overexpression of APEX1 protein in patients aged < 50 years(31–33). In addition, overexpression of this protein has been identified as a factor promoting tumor progression, migration and apoptosis in ENT cancers, and is associated with metastases, thus representing a poor prognostic factor and a biomarker(34–36).

Zymogen granule 16B (ZG16B) also showeda good protein signature ($p = 6.62E-03$; FC = 2.44). ZG16B, also known as PAUF (Pancreatic adenocarcinoma upregulated factor), is a newly identified secretory protein in pancreatic cancer. It is also overexpressed in ENT cancers, where it is involved in cell migration, invasion, proliferation, and metastasis(37). In 2021, Lu et *al*. showed that the ZG16B protein is overexpressed in breast cancer and that it represents a biomarker, an early diagnosis factor and a prognostic factor with no significant relationship with the age of patients(38).Other studies have supported this logic and have shown the involvement of ZG16B protein in prostate cancer and ovarian cancer (39,40).

Quantitative analysis of proteins according to whether patients were smokers or non-smokers revealed that 42 proteins were significantly overexpressed, two of which attracted our attention because of their protein signature. These were N-acetylgalactosaminyltransferase 6 polypeptide (GALT6: $p = 1.73E-04$; FC = 79.21) and integrin alpha M (ITAM: $p = 1.47E-03$; FC = 18.47). The polypeptide N-acetylgalactosaminyltransferase, overexpressed in smokers, is widely expressed in many tissues and is well-studied in cancer (41).Various studies have shown that GALT6 protein has considerable potential as a biomarker in gastric, pancreatic, and breast cancers, and its expression was first described in 2006 by Berois et *al*. Overexpression of GALT6 in tissues is an indicator of prognosis(42–44).This was made possible by the identification of GALT6 protein as a potential therapeutic target cancer treatment(41,45).

The alpha M integrin protein (ITAM: $p = 1.47E-03$; FC = 18.47), CD11b, was also overexpressed in smokers. It plays an essential role in various cancers, particularly as a potential predictive biomarker for oral cancer (46,47).Furthermore, overexpression of the ITAM protein is correlated with poor prognosis in oral squamous cell carcinoma, promoting tumor growth and metastasis(47–49).

Regarding the alcoholic and non-alcoholic status of patients, quantitative analysis revealed that 24 proteins were significantly overexpressed. Two of these proteins, because of their high scores compared to the others, are

discussed in this section. Protein Complement Component C7 (CO7: p = 9.72E-04; FC = 27.89) was overexpressed in alcohol consumers. The complement system, which is a component of the immune response, is involved in cancer progression. The expression level of CO7 was inversely correlated with prostate and ovarian cancers, and its low expression level was a poor prognostic factor (50–53).This information provides the status of biomarkers and therapeutic targets. In contrast to our study, Wang et *al.* in 2022 found that the CO7 protein is downregulated in gastric cancer tissues irrespective of patients' alcohol consumption status (54). CO7 protein is overexpressed in liver tumor tissue(55), and promotes the development of breast cancer, giving it a reliable prognostic value and insensitivity to chemotherapy (56,57).

We also found a good signature for Myosin-7 protein (MYH7: p = 5.90E-03; FC = 5.36). This protein is involved in immune infiltration in ENT cancers. In line with our results, other researchers have recently found that overexpression of the MYH7 protein correlates with poor prognosis in ENT cancer and constitutes a specific biomarker for this type of cancer (58,59). In addition to cancer, the MYH7 protein, which is overexpressed in cardiac pathologies, is a potential biomarker, particularly in heart failure and myocarditis (60,61).

In this study, the correlation of the quantitative analysis of proteins according to tumor location enabled us to target 27 significantly overexpressed proteins. The 60S L5 ribosomal protein (RL5) showed a high score (p = 3.39E-03; FC = 2.15) and was overexpressed in cheek tumors. The uL18 protein of the large ribosomal subunitis involved in cancer pathology. RL5 protein is an activator of the tumor suppressor protein p53 and plays a specific role in regulating p53 by inducing apoptosis (62–64).RL5 also plays an essential role in normal cell proliferation (65,66).

In the same dynamic, the Core histone macro-H2A.1 protein (H2AY: p = 2.50E-03; FC = 2.09) was overexpressed in samples taken from the cheek. It is also known as H2AFY, which is a member of the histone H2A family and is abnormally expressed in various pathologies, including tumors(67,68).In line with our results, H2AY protein is overexpressed in liver cancer and this overexpression predicts poor diagnosis, giving it the status of a biomarker for the diagnosis and prognosis of liver cancer(68,69).Similarly, overexpression of H2AY protein is also a tool for predicting recurrence in lung cancer and is therefore a useful prognostic biomarker(70).

Amino acids, the raw materials for protein synthesis and products of protein metabolism, enter the body through diet or are synthesized endogenously. Of the 20 proteinogenic amino acids, four were predominant in the composition of the proteins discussed in this study. Leucine, known for its role in the production of cellular energy in the form of adenosine triphosphate ATP), is an essential amino acid (71). A recent study carried out in 2024 showed that leucine has two properties (antitumor and protumor)(72).Leucine has significant anti-cancer activity, leading to apoptosis and regulation of immune signaling pathways. On the other hand, leucine also demonstrates pro-oncogenic activity, promoting the proliferation of tumor cells in various cancers including liver cancer. Glutamate, a nonessential amino acid, is one of the main neurotransmitters in the mammalian central nervous system. There is considerable evidence that glutamate plays an important role in tumor development by acting as a growth factor and signal mediator in both neuronal and non-neuronal tumor tissues(73). Lysine is an essential amino acid that is also involved in the production of energy in the form of ATP and in the immune system. High concentrations of lysine have been shown to stimulate the growth of tumor cells (71,74). Glycine, a non-essential amino acid, plays a role in DNA synthesis. Glycine overexpression correlates with the proliferation rate of cancer cells(75).

Conclusion.

To our knowledge, this work is the first proteomic study to correlate clinicopathological parameters of Senegalese patients with oropharyngeal cancer with overexpressed proteins in cancerous tissues. This work was carried out with an interesting number of patients for comparative study. Our findings highlighted numerous proteins of high prognostic value associated with robust statistics and significantly overexpressed according to clinicopathological parameters. Some of these proteins have already been reported in other cancers, but for the first time (to our knowledge) we associated them with oropharyngeal cancer. The next step would be to evaluate the expression of these proteins in cancerous tissue by immunohistochemistry, and validate their potential as biomarker for oropharyngeal cancer.

Conflict of Interest.

No potential conflicts of interest were disclosed.

References.

1. OMS. Cancer : une charge toujours plus lourde dans le monde et des besoins en services croissants. 2022.
2. Mesia R, Iglesias L, Lambea J, Martínez-Trufero J, Soria A, Taberna M, et al. SEOM clinical guidelines for the treatment of head and neck cancer (2020). Clin Transl Oncol. mai 2021;23(5):913-21.
3. Fayette J. Brochure : les cancers des voies aérodigestives supérieures | Fondation ARC pour la recherche sur le cancer. https://www.fondation-arc.org/support-information/brochure-les-cancers-des-voies-aerodigestives-superieures. (accessed 2024-11-04).
4. Rivera C. Essentials of oral cancer. Oral cancer.Int J ClinExpPathol 2015;8(9):11884-11894.
5. Chaturvedi AK, Freedman ND, Abnet CC. The Evolving Epidemiology of Oral Cavity and Oropharyngeal Cancers. Cancer Research. 16 août 2022;82(16):2821-3.
6. Badoual C, Péré H, Roussel H, Si Mohamed A, Tartour É. Les cancers des voies aérodigestives supérieures associés aux papillomavirus. Med Sci (Paris). janv 2013;29(1):83-8.
7. Güneri P, Epstein JB. Late stage diagnosis of oral cancer: Components and possible solutions. Oral Oncology. déc 2014;50(12):1131-6.
8. Kwon YW, Jo HS, Bae S, Seo Y, Song P, Song M, et al. Application of Proteomics in Cancer: Recent Trends and Approaches for Biomarkers Discovery. Front Med. 22 sept 2021;8:747333.
9. Karayel O, Virreira Winter S, Padmanabhan S, Kuras YI, Vu DT, Tuncali I, et al. Proteome profiling of cerebrospinal fluid reveals biomarker candidates for Parkinson's disease. Cell Reports Medicine. juin 2022;3(6):100661.
10. Chalfouh C, Guillou C, Hardouin J, Delarue Q, Li X, Duclos C, et al. The Regenerative Effect of Trans-spinal Magnetic Stimulation After Spinal Cord Injury: Mechanisms and Pathways Underlying the Effect. Neurotherapeutics. oct 2020;17(4):2069-88.
11. Simancas Escorcia V, Guillou C, Abbad L, Derrien L, Rodrigues Rezende Costa C, Cannaya V, et al. Pathogenesis of Enamel-Renal Syndrome Associated Gingival Fibromatosis: A Proteomic Approach. Front Endocrinol. 29 oct 2021;12:752568.
12. Vo K, Sharma Y, Paul A, Mohamadi R, Mohamadi A, Fields PE, et al. Importance of Transcript Variants in Transcriptome Analyses. Cells. 8 sept 2024;13(17):1502.
13. Dalman MR, Deeter A, Nimishakavi G, Duan ZH. Fold change and p-value cutoffs significantly alter microarray interpretations. BMC Bioinformatics. déc 2012;13(S2):S11.
14. Lefranc MP. Immunoglobulin and T Cell Receptor Genes: IMGT® and the Birth and Rise of Immunoinformatics. Front Immunol. 2014;5*:22.*
15. Zhu X, Li C, Sun X, Mao Y, Li G, Liu X, et al. Immunoglobulin mRNA and Protein Expression in Human Oral Epithelial Tumor Cells. Applied Immunohistochemistry & Molecular Morphology. mai 2008;16(3):232-8.
16. Cui M, Huang J, Zhang S, Liu Q, Liao Q, Qiu X. Immunoglobulin Expression in Cancer Cells and Its Critical Roles in Tumorigenesis. Front Immunol. 24 mars 2021;12:613530.
17. Peng J, Wang HC, Liu Y, Jiang JH, Lv WQ, Yang Y, et al. Involvement of non-B cell-derived immunoglobulin G in the metastasis and prognosis of salivary adenoid cystic carcinoma. Oncology Letters. oct 2017;14(4):4491-8.
18. Oh-I S, Shimizu H, Satoh T, Okada S, Adachi S, Inoue K, et al. Identification of nesfatin-1 as a satiety molecule in the hypothalamus. Nature. oct 2006;443(7112):709-12.
19. Zhao J, Yun X, Ruan X, Chi J, Yu Y, Yigong L, et al. High expression of NUCB2 promotes papillary thyroid cancer cells proliferation and invasion. OTT. févr 2019;Volume 12:1309-18.
20. Kmiecik AM, Dzięgiel P, Podhorska-Okołów M. Nucleobindin-2/Nesfatin-1—A New Cancer Related Molecule? IJMS. 2 août 2021;22(15):8313.
21. Skorupska A, Lenda R, Ożyhar A, Bystranowska D. The Multifaceted Nature of Nucleobindin-2 in Carcinogenesis. IJMS. 26 mai 2021;22(11):5687.
22. Zhang H, Qi C, Li L, Luo F, Xu Y. Clinical significance of NUCB2 mRNA expression in prostate cancer. J Exp Clin Cancer Res. déc 2013;32(1):56.
23. Liu GM, Xu ZQ, Ma HS. Nesfatin-1/Nucleobindin-2 Is a Potent Prognostic Marker and Enhances Cell Proliferation, Migration, and Invasion in Bladder Cancer. Disease Markers. 19 sept 2018;2018:1-9.
24. Suzuki S, Takagi K, Miki Y, Onodera Y, Akahira J, Ebata A, et al. Nucleobindin 2 in human breast carcinoma as a potent prognostic factor. Cancer Science. janv 2012;103(1):136-43.
25. Zeng L, Zhong J, He G, Li F, Li J, Zhou W, et al. Identification of Nucleobindin-2 as a Potential Biomarker for Breast Cancer Metastasis Using iTRAQ-based Quantitative Proteomic Analysis. J Cancer. 2017;8(15):3062-9.
26. Wang Y, Sun B, Wei W, Han T, Ma J, Li X, et al. NUCB2 promotes hepatocellular carcinoma cell growth and metastasis by activating the E2F4/PTGR1 axis. Int J Biol Sci. 2024;20(12):4767-80.

27. Dyballa-Rukes N, Jakobs P, Eckers A, Ale-Agha N, Serbulea V, Aufenvenne K, et al. The Anti-Apoptotic Properties of APEX1 in the Endothelium Require the First 20 Amino Acids and Converge on Thioredoxin-1. Antioxidants & Redox Signaling. 20 avr 2017;26(12):616-29.
28. Oliveira TT, Fontes-Dantas FL, De Medeiros Oliveira RK, Pinheiro DML, Coutinho LG, Da Silva VL, et al. Chemical Inhibition of Apurinic-Apyrimidinic Endonuclease 1 Redox and DNA Repair Functions Affects the Inflammatory Response via Different but Overlapping Mechanisms. Front Cell Dev Biol. 20 sept 2021;9:731588.
29. Thakur S, Sarkar B, Cholia RP, Gautam N, Dhiman M, Mantha AK. APE1/Ref-1 as an emerging therapeutic target for various human diseases: phytochemical modulation of its functions. Exp Mol Med. 18 juill 2014;46(7):e106-e106.
30. Kumar S, Talluri S, Pal J, Yuan X, Lu R, Nanjappa P, et al. Role of apurinic/apyrimidinic nucleases in the regulation of homologous recombination in myeloma: mechanisms and translational significance. Blood Cancer Journal. 25 sept 2018;8(10):92.
31. Zhang Y, Zhang L, Zhang L, Bai J, Ge H, Liu P. Expression changes in DNA repair enzymes and mitochondrial DNA damage in aging rat lens.Molecular Vision 2010; 16:1754-1763.
32. Park JS, Kim HL, Kim YJ, Weon JI, Sung MK, Chung HW, et al. Human AP Endonuclease 1: A Potential Marker for the Prediction of Environmental Carcinogenesis Risk. Oxidative Medicine and Cellular Longevity. 2014;2014:1-15.
33. Madlener S, Ströbel T, Vose S, Saydam O, Price BD, Demple B, et al. Essential role for mammalian apurinic/apyrimidinic (AP) endonuclease Ape1/Ref-1 in telomere maintenance. Proc Natl Acad Sci USA. 29 oct 2013;110(44):17844-9.
34. Malfatti MC, Bellina A, Antoniali G, Tell G. Revisiting Two Decades of Research Focused on Targeting APE1 for Cancer Therapy: The Pros and Cons. Cells. 20 juill 2023;12(14):1895.
35. Hsia K, Liu C, Mar K, Lin L, Lin C, Cheng M, et al. Impact of apurinic/apyrimidinic endonuclease 1/redox factor-1 on treatment response and survival in oral squamous cell carcinoma. Head & Neck. avr 2016;38(4):550-9.
36. Wicker CA, Takiar V, Suganya R, Arnold SM, Brill YM, Chen L, et al. Evaluation of antioxidant network proteins as novel prognostic biomarkers for head and neck cancer patients. Oral Oncology. déc 2020;111:104949.
37. Sasahira T, Kurihara M, Nishiguchi Y, Nakashima C, Kirita T, Kuniyasu H. Pancreatic adenocarcinoma up-regulated factor has oncogenic functions in oral squamous cell carcinoma. Histopathology. mars 2017;70(4):539-48.
38. Lu H, Shi C, Liu X, Liang C, Yang C, Wan X, et al. Identification of ZG16B as a prognostic biomarker in breast cancer. Open Medicine. 25 nov 2020;16(1):1-13.
39. Zhang T, Wang Y, Dong Y, Liu L, Han Y, Wang H, et al. Identification of Novel Diagnostic Biomarkers in Prostate Adenocarcinoma Based on the Stromal-Immune Score and Analysis of the WGCNA and ceRNA Network. Wang F, éditeur. Disease Markers. 15 janv 2022;2022:1-10.
40. Choi CH, Kang TH, Song JS, Kim YS, Chung EJ, Ylaya K, et al. Elevated expression of pancreatic adenocarcinoma upregulated factor (PAUF) is associated with poor prognosis and chemoresistance in epithelial ovarian cancer. Sci Rep. 15 août 2018;8(1):12161.
41. Reis JSD, Santos MARDC, Da Costa KM, Freire-de-Lima CG, Morrot A, Previato JO, et al. Increased expression of the pathological O-glycosylated form of oncofetal fibronectin in the multidrug resistance phenotype of cancer cells. Matrix Biology. avr 2023;118:47-68.
42. Guo Y, Shi J, Zhang J, Li H, Liu B, Guo H. Polypeptide N-acetylgalactosaminyltransferase-6 expression in gastric cancer. OTT. juill 2017;Volume 10:3337-44.
43. Li Z, Yamada S, Inenaga S, Imamura T, Wu Y, Wang KY, et al. Polypeptide N-acetylgalactosaminyltransferase 6 expression in pancreatic cancer is an independent prognostic factor indicating better overall survival. Br J Cancer. juin 2011;104(12):1882-9.
44. Berois N, Mazal D, Ubillos L, Trajtenberg F, Nicolas A, Sastre-Garau X, et al. UDP- *N* -Acetyl- D -Galactosamine: Polypeptide *N* -Acetylgalactosaminyltransferase-6 as a New Immunohistochemical Breast Cancer Marker. J Histochem Cytochem. mars 2006;54(3):317-28.
45. Deng B, Tarhan YE, Ueda K, Ren L, Katagiri T, Park JH, et al. Critical Role of Estrogen Receptor Alpha O-Glycosylation by N-Acetylgalactosaminyltransferase 6 (GALNT6) in Its Nuclear Localization in Breast Cancer Cells. Neoplasia. oct 2018;20(10):1038-44.
46. Zhang Y ping, Wu J, Han Y fang, Shi Z rui, Wang L. Pathogenesis of cutaneous lupus erythema associated with and without systemic lupus erythema. Autoimmunity Reviews. juill 2017;16(7):735-42.

47. Jiang Y, Wang C, Wang Y, Zhang W, Liu L, Cheng J. Prognostic role of CD11b+ myeloid-derived suppressor cells in oral squamous cell carcinoma. Arch Med Sci. 25 mars 2021; (1): 171-179.
48. Lo Y, Lee AY, Liu Y, Ko H, Peng H, Lee H, et al. β-glucan therapy converts the inhibition of myeloid-derived suppressor cells in oral cancer patients. Oral Diseases. sept 2022;28(6):1484-95.
49. Sugiura K, Nakajima S, Kato I, Okubo-Sato M, Nakazawa Y, Mitsudo K, et al. Hypoxia and CD11b+ Cell Influx Are Strongly Associated With Lymph Node Metastasis of Oral Cancer. Anticancer Res. déc 2020;40(12):6845-52.
50. Zhou Z, Jia D, Kwon O, Li S, Sun H, Roudier MP, et al. Androgen-regulated stromal complement component 7 (C7) suppresses prostate cancer growth. Oncogene. 4 août 2023;42(32):2428-38.
51. Ying L, Zhang F, Pan X, Chen K, Zhang N, Jin J, et al. Complement component 7 (C7), a potential tumor suppressor, is correlated with tumor progression and prognosis. Oncotarget. 27 déc 2016;7(52):86536-46.
52. Loveridge CJ, Slater S, Campbell KJ, Nam NA, Knight J, Ahmad I, et al. BRF1 accelerates prostate tumourigenesis and perturbs immune infiltration. Oncogene. 20 févr 2020;39(8):1797-806.
53. Chen Z, Yan X, Du GW, Tuoheti K, Bai XJ, Wu HH, et al. Complement C7 (C7), a Potential Tumor Suppressor, Is an Immune-Related Prognostic Biomarker in Prostate Cancer (PC). Front Oncol. 25 août 2020;10:1532.
54. Wang S, Hu W, Xie Y, Wu H, Jia Z, Zhang Z, et al. Functional genetic variants in complement component 7 confer susceptibility to gastric cancer. PeerJ. 18 janv 2022;10:e12816.
55. Seol HS, Lee SE, Song JS, Rhee JK, Singh SR, Chang S, et al. Complement proteins C7 and CFH control the stemness of liver cancer cells via LSF-1. Cancer Letters. mars 2016;372(1):24-35.
56. Zabihi MR, Farhadi B, Akhoondian M. Complement protein expression changes in various conditions of breast cancer: in-silico analyses—experimental research. Annals of Medicine & Surgery. sept 2024;86(9):5152-61.
57. Zhang H, Zhao Y, Liu X, Fu L, Gu F, Ma Y. High Expression of Complement Component C7 Indicates Poor Prognosis of Breast Cancer and Is Insensitive to Taxane-Anthracycline Chemotherapy. Front Oncol. 24 sept 2021;11:724250.
58. Li C, Guan R, Li W, Wei D, Cao S, Chang F, et al. Analysis of myosin genes in HNSCC and identify MYL1 as a specific poor prognostic biomarker, promotes tumor metastasis and correlates with tumor immune infiltration in HNSCC. BMC Cancer. 7 sept 2023;23(1):840.
59. Ju G, Yao Z, Zhao Y, Zhao X, Liu F. Data mining on identifying diagnosis and prognosis biomarkers in head and neck squamous carcinoma. Sci Rep. 20 juin 2023;13(1):10020.
60. Chugh S, Ouzounian M, Lu Z, Mohamed S, Li W, Bousette N, et al. Pilot study identifying myosin heavy chain 7, desmin, insulin-like growth factor 7, and annexin A 2 as circulating biomarkers of human heart failure. Proteomics. août 2013;13(15):2324-34.
61. Rasheed S, Hashim R, Yan JS. Possible Biomarkers for the Early Detection of HIV-associated Heart Diseases: A Proteomics and Bioinformatics Prediction. Computational and Structural Biotechnology Journal. 2015;13:145-52.
62. Koçak E, Çelebier M, Haznedaroglu IC, Altınöz S. Analysis of the Antiproliferative Effect of Ankaferd Hemostat on Caco-2 Colon Cancer Cells via LC/MS Shotgun Proteomics Approach. BioMed Research International. 21 mai 2019;2019:1-11.
63. Zhang Y, Lu H. Signaling to p53: Ribosomal Proteins Find Their Way. Cancer Cell. nov 2009;16(5):369-77.
64. Fumagalli S, Ivanenkov VV, Teng T, Thomas G. Suprainduction of p53 by disruption of 40S and 60S ribosome biogenesis leads to the activation of a novel G2/M checkpoint. Genes Dev. 15 mai 2012;26(10):1028-40.
65. Teng T, Mercer CA, Hexley P, Thomas G, Fumagalli S. Loss of Tumor Suppressor RPL5/RPL11 Does Not Induce Cell Cycle Arrest but Impedes Proliferation Due to Reduced Ribosome Content and Translation Capacity. Molecular and Cellular Biology. 1 déc 2013;33(23):4660-71.
66. Donati G, Peddigari S, Mercer CA, Thomas G. 5S Ribosomal RNA Is an Essential Component of a Nascent Ribosomal Precursor Complex that Regulates the Hdm2-p53 Checkpoint. Cell Reports. juill 2013;4(1):87-98.
67. Corujo D, Buschbeck M. Post-Translational Modifications of H2A Histone Variants and Their Role in Cancer. Cancers. 27 févr 2018;10(3):59.
68. Ma X, Ding Y, Zeng L. The diagnostic and prognostic value of H2AFY in hepatocellular carcinoma. BMC Cancer. déc 2021;21(1):418.
69. Rappa F, Greco A, Podrini C, Cappello F, Foti M, Bourgoin L, et al. Immunopositivity for Histone MacroH2A1 Isoforms Marks Steatosis-Associated Hepatocellular Carcinoma. Folli F, éditeur. PLoS ONE. 23 janv 2013;8(1):e54458.
70. Sporn JC, Kustatscher G, Hothorn T, Collado M, Serrano M, Muley T, et al. Histone macroH2A isoforms predict the risk of lung cancer recurrence. Oncogene. 24 sept 2009;28(38):3423-8.

71. Elie F. protéines et acides aminés. 15 dec 2022;[Internet]. http://fred.elie.free.fr:1-18 ((accessed 2025-04-24).
72. Akbay B, Omarova Z, Trofimov A, Sailike B, Karapina O, Molnár F, et al. Double-Edge Effects of Leucine on Cancer Cells. Biomolecules. 4 nov 2024;14(11):1401.
73. Stepulak A, Rola R, Polberg K, Ikonomidou C. Glutamate and its receptors in cancer. J Neural Transm. août 2014;121(8):933-44.
74. Lukasheva EV, Babayeva G, Karshieva SSh, Zhdanov DD, Pokrovsky VS. L-Lysine α-Oxidase: Enzyme with Anticancer Properties. Pharmaceuticals. 22 oct 2021;14(11):1070.
75. Amelio I, Cutruzzolá F, Antonov A, Agostini M, Melino G. Serine and glycine metabolism in cancer. Trends in Biochemical Sciences. avr 2014;39(4):191-8.

Table 1: List of all proteins statistically deregulated for each clinicopathological parameter.

Parameters	All proteins	Proteins$p$<005	Total proteins$p$<005
Man	677	31	
Woman	540	13	44
Under 50 years	599	18	
Over 50 years	618	5	23
Smokers	592	66	
No-smokers	625	17	83
Alcohol	296	24	
No-Alcohol	921	4	28
Cheek	745	58	
Other location	472	18	76

Table 2: List of significantly overexpressed proteins by gender (p< 0.05; FC > 2).

Accession	Description	Anova (p)	Max fold change	Highestmean condition
AL3A2_HUMAN	Fattyaldehydedehydrogenase	2,51E-02	36,78	Man
HV373_HUMAN	Immunoglobulinheavy variable	1,53E-02	5,07	Man
TF65_HUMAN	Transcription factor p65	3,92E-02	4,52	Man
ATPB_HUMAN	ATP synthasesubunit beta	3,39E-02	2,59	Man
FLNA_HUMAN	Filamin-A	3,45E-02	2,40	Man
NUCB2_HUMAN	Nucleobindin-2	7,85E-03	2,34	Man
ARPC3_HUMAN	Actin-relatedprotein 2/3 complexsubunit 3	4,24E-02	2,18	Man
EHD1_HUMAN	EH domain-containingprotein 1	1,31E-02	2,15	Man
CTHR1_HUMAN	Collagen triple helix repeat-containing	4,81E-02	2,04	Man
UBP14_HUMAN	Ubiquitincarboxyl-terminal hydrolase 14	9,70E-03	2,03	Man
ANXA2_HUMAN	Annexin A2	3,32E-02	2,01	Man
TRAP1_HUMAN	Heatshockprotein 75 kDa	4,86E-02	3,13	Woman
IRGQ_HUMAN	Immunity-related GTPase family Q protein	2,78E-02	2,76	Woman
CCAR2_HUMAN	Cell cycle and apoptosis regulator protein 2	2,32E-02	2,69	Woman
AL1A1_HUMAN	Retinaldehydrogenase 1	2,11E-02	2,41	Woman

Table 3: List of proteins significantly overexpressed as a function of age ($p < 0.05$; FC > 2)

Accession	Description	Anova (p)	Max fold change	Highestmean condition
CYB5_HUMAN	Cytochrome b5	2,05E-02	7,96	Under 50 years
STAT3_HUMAN	Signal transducer and activator of transcription 3	3,01E-02	2,90	Under 50 years
APEX1_HUMAN	DNA-(apurinic or apyrimidinic site) lyase	1,36E-02	2,73	Under 50 years
ZG16B_HUMAN	Zymogen granule protein 16 homolog B	6,62E-03	2,44	Under 50 years
TCTP_HUMAN	Translationally-controlledtumorprotein	4,39E-02	2,29	Under 50 years
FBLN5_HUMAN	Fibulin-5	1,81E-03	2,11	Under 50 years
RALA_HUMAN	Ras-relatedproteinRal-A	3,83E-02	2,10	Under 50 years
HV309_HUMAN	Immunoglobulinheavy variable 3-9	3,36E-02	2,89	Over 50 years
CRK_HUMAN	Adapter moleculecrk	3,85E-02	2,68	Over 50 years
CTNA1_HUMAN	Catenin alpha-1	3,68E-02	2,10	Over 50 years

Table 4: Proteins significantly overexpressed as a function of smoking ($p< 0.05$; FC > 2).

Accession	Description	Anova (p)	Max fold change	Highestmean condition
GALT6_HUMAN	Polypeptide N-acetylgalactosaminyltransferase 6	1,73E-04	79,21	Tobacco
ITAM_HUMAN	Integrin alpha-M	1,47E-03	18,47	Tobacco
GSTA1_HUMAN	Glutathione S-transferase A1	4,13E-02	8,95	Tobacco
PYR1_HUMAN	CAD protein	4,55E-02	7,68	Tobacco
CTHR1_HUMAN	Collagen triple helix repeat-containing protein 1	1,33E-02	4,43	Tobacco
AL3A1_HUMAN	Aldehyde dehydrogenase, dimeric NADP-preferring	3,39E-02	4,08	Tobacco
KLK10_HUMAN	Kallikrein-10	3,73E-02	4,07	Tobacco
CEAM8_HUMAN	Carcinoembryonic antigen-related cell adhesion molecule 8	1,05E-02	3,85	Tobacco
IGJ_HUMAN	Immunoglobulin J chain	3,49E-03	3,79	Tobacco
SYIC_HUMAN	Isoleucine--tRNA ligase, cytoplasmic	1,49E-02	3,59	Tobacco
RCN3_HUMAN	Reticulocalbin-3	3,36E-02	3,56	Tobacco
EST1_HUMAN	Livercarboxylesterase 1	3,32E-02	3,55	Tobacco
IPO5_HUMAN	Importin-5	4,68E-02	3,45	Tobacco
TITIN_HUMAN	Titin	3,40E-02	2,96	Tobacco
ENDOU_HUMAN	Poly(U)-specificendoribonuclease	7,84E-04	2,73	Tobacco
TAGL_HUMAN	Transgelin	2,81E-02	2,68	Tobacco
HEXB_HUMAN	Beta-hexosaminidasesubunit beta	2,82E-02	2,66	Tobacco
SH3L1_HUMAN	SH3 domain-binding glutamic acid-rich-like protein	7,01E-03	2,58	Tobacco
CILP1_HUMAN	Cartilage intermediatelayerprotein 1	1,73E-02	2,57	Tobacco
CATC_HUMAN	Dipeptidyl peptidase 1	4,41E-03	2,48	Tobacco
PERM_HUMAN	Myeloperoxidase	3,61E-02	2,46	Tobacco
NUCB2_HUMAN	Nucleobindin-2	3,98E-03	2,42	Tobacco
RD23A_HUMAN	UV excision repairprotein RAD23 homolog A	3,13E-02	2,37	Tobacco
PIGR_HUMAN	Polymericimmunoglobulinreceptor	4,18E-02	2,31	Tobacco
SGTA_HUMAN	Small glutamine-rich tetratricopeptiderepeat-containing protein alpha	4,28E-04	2,22	Tobacco
TKFC_HUMAN	Triokinase/FMN cyclase	2,68E-02	2,21	Tobacco
KVD11_HUMAN	Immunoglobulin kappa variable 3D-11	5,76E-03	2,20	Tobacco
TRAP1_HUMAN	Heat shock protein 75 kDa, mitochondrial	2,88E-03	2,19	Tobacco
IGKC_HUMAN	Immunoglobulin kappa constant	4,47E-02	2,16	Tobacco
AT2A1_HUMAN	Sarcoplasmic/endoplasmic reticulum calcium ATPase 1	8,18E-03	2,16	Tobacco
MYOM2_HUMAN	Myomesin-2	3,34E-02	2,16	Tobacco
LV39_HUMAN	Immunoglobulin lambda variable 3-9	1,24E-02	2,15	Tobacco
CATZ_HUMAN	Cathepsin Z	4,36E-03	2,12	Tobacco
FRIH_HUMAN	Ferritinheavychain	3,03E-02	2,09	Tobacco
GILT_HUMAN	Gamma-interferon-inducible lysosomalthiolreductase	2,15E-02	2,07	Tobacco
SPTN1_HUMAN	Spectrin alpha chain, non-erythrocytic 1	2,93E-03	2,06	Tobacco
SPB6_HUMAN	Serpin B6	4,28E-02	3,85	No Tobacco
EHD1_HUMAN	EH domain-containingprotein 1	1,66E-02	2,53	No Tobacco
PCY1A_HUMAN	Choline-phosphate cytidylyltransferase A	3,17E-02	2,51	No Tobacco
ATLA3_HUMAN	Atlastin-3	1,15E-02	2,43	No Tobacco
ANXA8_HUMAN	Annexin A8	4,27E-02	2,10	No Tobacco
SUN2_HUMAN	SUN domain-containingprotein 2	4,02E-02	2,07	No Tobacco

Table 5: Proteins significantly overexpressed as a function of alcohol status ($p < 0.05$; FC > 2)

Accession	Description	Anova (p)	Max fold change	Highestmean condition
CO7_HUMAN	Complement component C7	9,72E-04	27,89	Alcohol
MYH7_HUMAN	Myosin-7	5,90E-03	5,36	Alcohol
RS7_HUMAN	40S ribosomal protein S7	9,84E-03	5,36	Alcohol
TNPO1_HUMAN	Transportin-1	5,84E-03	4,65	Alcohol
AL9A1_HUMAN	4-trimethylaminobutyraldehyde dehydrogenase	9,36E-03	4,12	Alcohol
HSPB6_HUMAN	Heatshockprotein beta-6	1,24E-02	3,92	Alcohol
PLMN_HUMAN	Plasminogen	5,69E-03	3,83	Alcohol
UBA1_HUMAN	Ubiquitin-like modifier-activating enzyme 1	1,65E-02	3,46	Alcohol
COCA1_HUMAN	Collagen alpha-1(XII) chain	1,11E-02	3,24	Alcohol
MFAP4_HUMAN	Microfibril-associatedglycoprotein 4	3,93E-02	2,96	Alcohol
C4BPA_HUMAN	C4b-binding protein alpha chain	2,89E-02	2,76	Alcohol
ILF3_HUMAN	Interleukinenhancer-binding factor 3	8,56E-03	2,75	Alcohol
ITIH4_HUMAN	Inter-alpha-trypsin inhibitor heavy chain H4	1,02E-02	2,71	Alcohol
FA83H_HUMAN	Protein FAM83H	1,21E-02	2,71	Alcohol
LAMA3_HUMAN	Lamininsubunit alpha-3	2,54E-02	2,53	Alcohol
JKIP3_HUMAN	Janus kinase and microtubule-interacting protein 3	4,60E-02	2,48	Alcohol
PNPT1_HUMAN	Polyribonucleotidenucleotidyltransferase 1	8,63E-03	2,41	Alcohol
SYNC_HUMAN	Asparagine--tRNA ligase, cytoplasmic	1,17E-02	2,39	Alcohol
OPA1_HUMAN	Dynamin-like 120 kDaprotein, mitochondrial	1,85E-03	2,25	Alcohol
TPP1_HUMAN	Tripeptidyl-peptidase 1	1,20E-02	2,12	Alcohol
SEPT2_HUMAN	Septin-2	2,44E-02	2,10	Alcohol
TTHY_HUMAN	Transthyretin	2,84E-02	2,02	Alcohol
DC1L2_HUMAN	Cytoplasmic dynein 1 light intermediate chain 2	2,58E-02	11,35	NoAlcohol
GLU2B_HUMAN	Glucosidase 2 subunit beta	4,03E-02	4,32	NoAlcohol

Table 6: List of proteins significantly overexpressed according to tumor location ($p < 0.05$; FC > 2)

Accession	Description	Anova (p)	Max fold change	Highestmean condition
MYG_HUMAN	Myoglobin	1,30E-02	3,56	Cheek
KCRS_HUMAN	Creatine kinase S-type, mitochondrial	1,36E-02	3,11	Cheek
MYL6B_HUMAN	Myosin light chain 6B	2,99E-02	3,10	Cheek
ACTS_HUMAN	Actin, alpha skeletal muscle	1,40E-02	2,86	Cheek
DQB1_HUMAN	HLA class II histocompatibility antigen, DQ beta 1 chain	4,26E-02	2,72	Cheek
MYL1_HUMAN	Myosin light chain 1/3, skeletal muscle isoform	3,04E-02	2,58	Cheek
PGAM2_HUMAN	Phosphoglyceratemutase 2	9,25E-03	2,37	Cheek
RL6_HUMAN	60S ribosomal protein L6	1,07E-02	2,36	Cheek
SPB9_HUMAN	Serpin B9	1,77E-02	2,32	Cheek
MZB1_HUMAN	Marginal zone B- and B1-cell-specific protein	2,36E-02	2,22	Cheek
RL5_HUMAN	60S ribosomal protein L5	3,39E-03	2,15	Cheek
THIO_HUMAN	Thioredoxin	1,45E-02	2,10	Cheek
URP2_HUMAN	Fermitinfamilyhomolog 3	2,47E-02	2,09	Cheek
H2AY_HUMAN	Core histone macro-H2A.1	2,50E-03	2,09	Cheek
CMA1_HUMAN	Chymase	3,23E-02	2,09	Cheek
RU17_HUMAN	U1 small nuclear ribonucleoprotein 70 kDa	4,69E-02	2,08	Cheek
MYOTI_HUMAN	Myotilin	2,17E-02	2,08	Cheek
RS17_HUMAN	40S ribosomal protein S17	3,04E-02	2,02	Cheek
MMP1_HUMAN	Interstitialcollagenase	3,40E-02	8,28	Other location
PTK7_HUMAN	Inactive tyrosine-protein kinase 7	2,27E-02	3,10	Other location
ACON_HUMAN	Aconitatehydratase, mitochondrial	3,49E-02	2,78	Other location
HDGF_HUMAN	Hepatoma-derivedgrowth factor	2,50E-02	2,66	Other location
RO60_HUMAN	60 kDa SS-A/Ro ribonucleoprotein	3,56E-03	2,29	Other location
NDRG1_HUMAN	Protein NDRG1	3,88E-02	2,24	Other location
LAD1_HUMAN	Ladinin-1	3,94E-02	2,19	Other location
IRGQ_HUMAN	Immunity-related GTPase family Q protein	2,82E-03	2,06	Other location
PEPL_HUMAN	Periplakin	4,53E-02	2,06	Other location

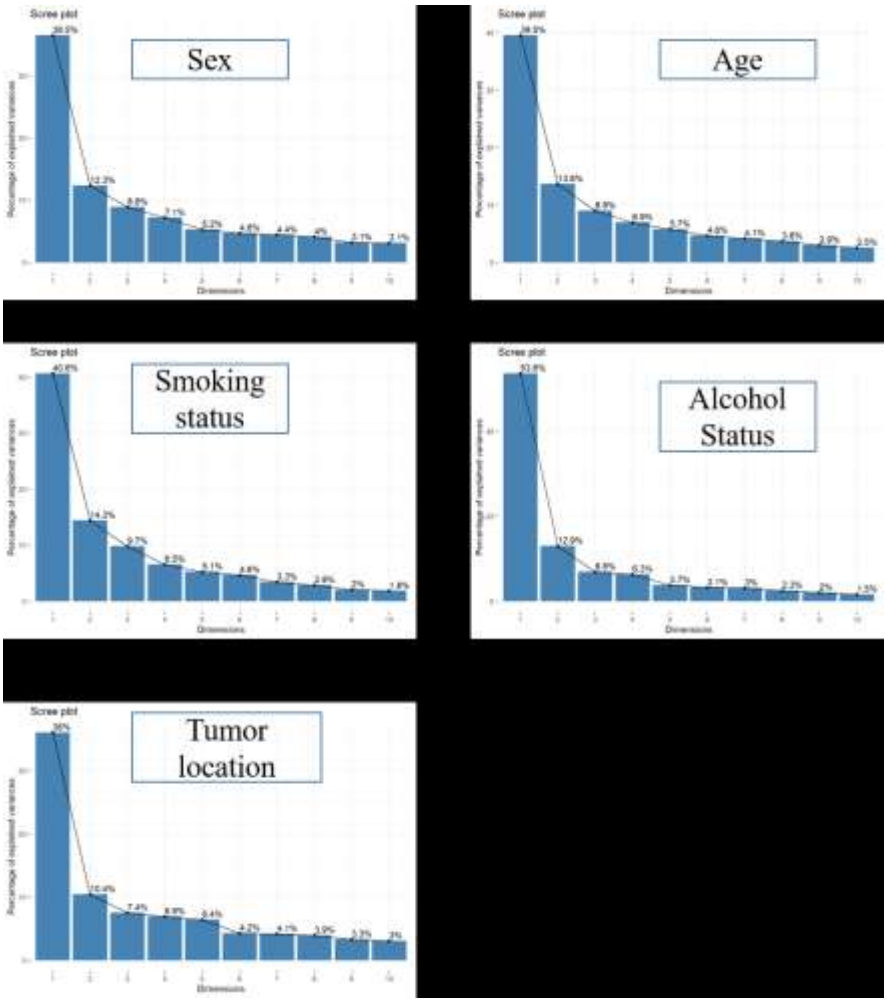

Figure 1: Screeplots of eigenvalues for all parameters. The eigenvalue score revealed that the drop-off (elbow) after the first two axes was followed by a steady decrease from the third axis onwards for all parameters.

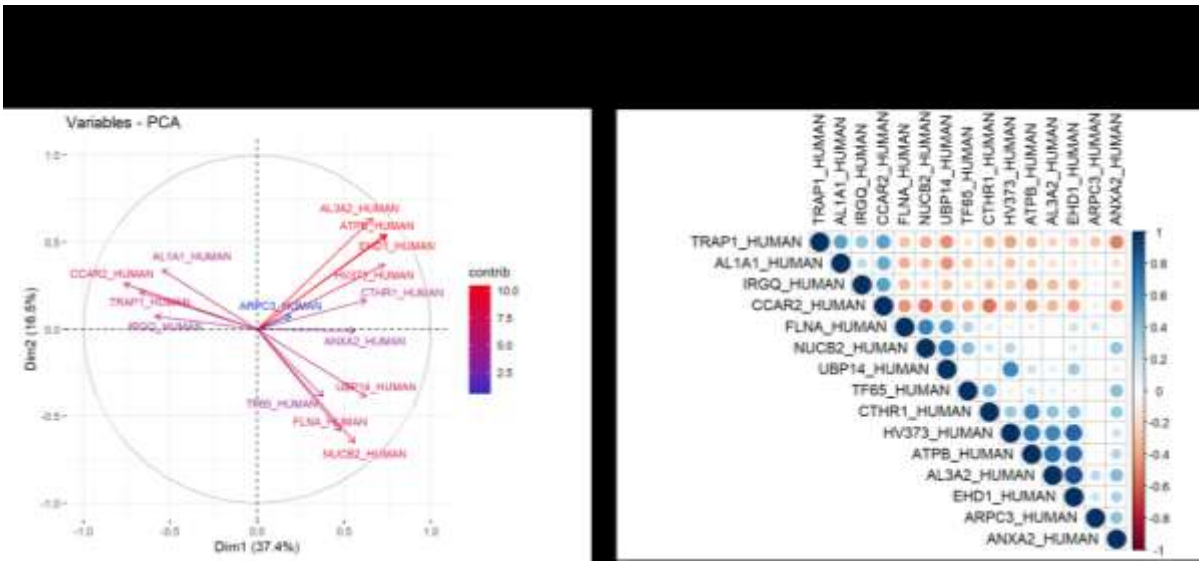

Figure 2: Correlation and contribution of proteins and patient gender to the construction of the two dimensions. Correlation circle of principal component analysis with high-contribution variables in red arrow and low-contribution variables in blue arrow. Proximities between variables are interpreted in terms of correlation (A). Correlation matrix with Positive correlations are shown in blue and negative correlations in red. Color intensity and circle size are proportional to correlation coefficients (B).

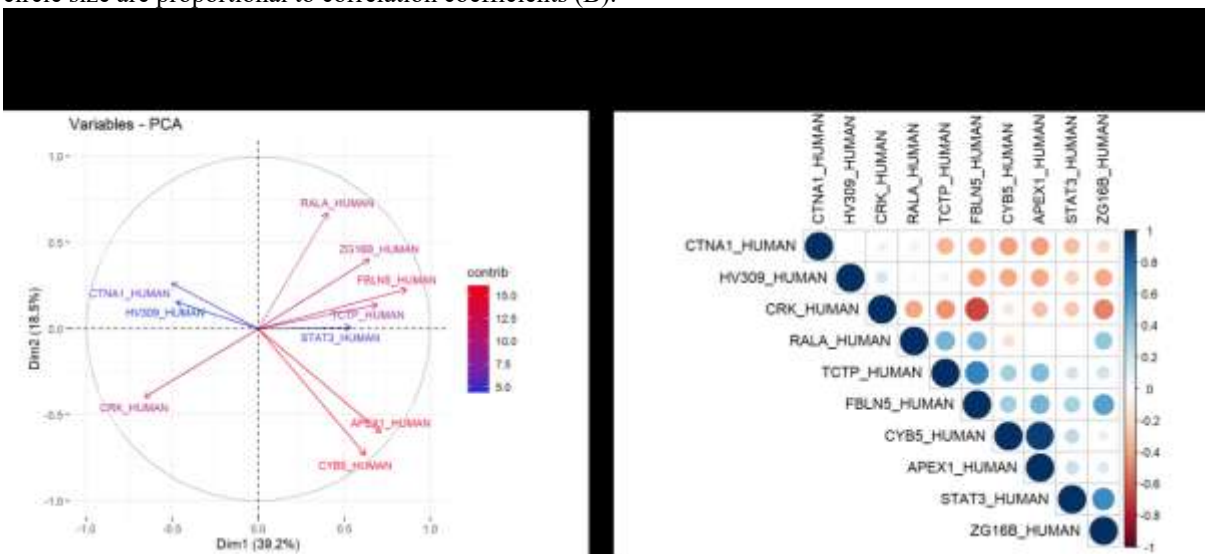

Figure 3: Correlation and contribution of proteins and patient age to the construction of the two dimensions. Correlation circle of principal component analysis (A) and Correlation matrix (B).

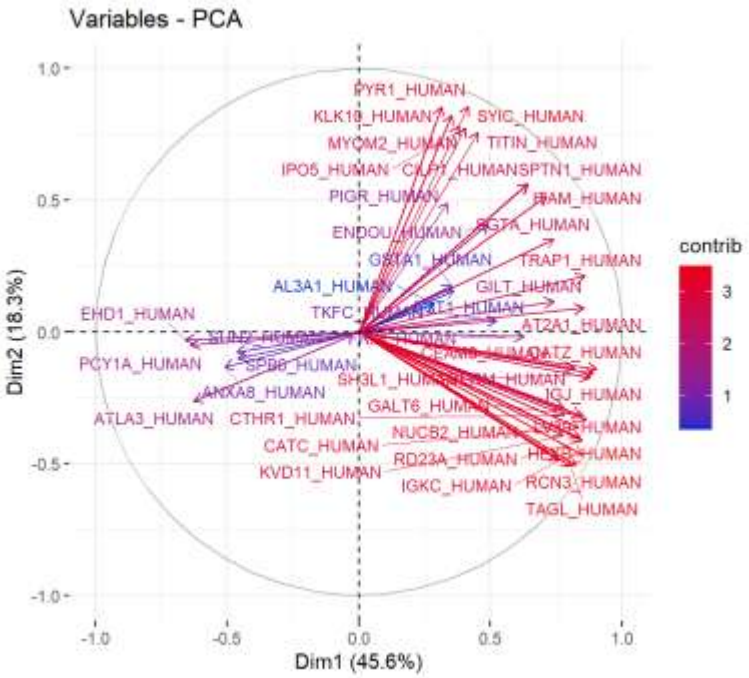

Figure 4: Correlation circle of principal component analysis of proteins and smoking status to the construction of the two dimensions.Correlation circle of principal component analysis with high-contribution variables in red arrow and low-contribution variables in blue arrow. Proximities between variables are interpreted in terms of correlation.

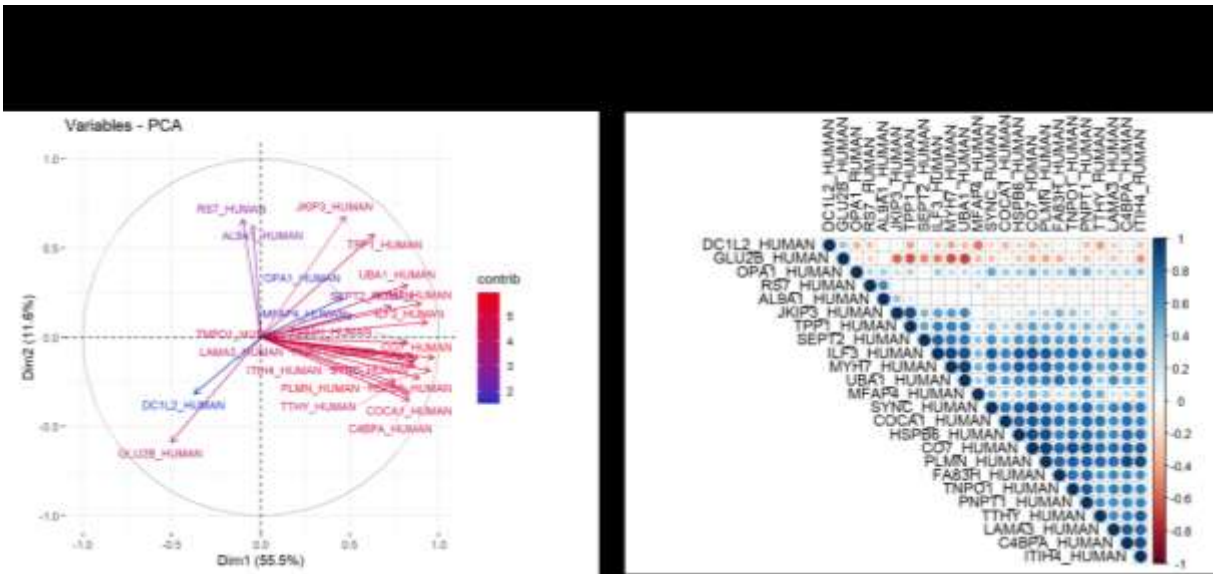

Figure 5: Correlation and contribution of proteins and patients' alcohol status to the construction of the two dimensions. Correlation circle of principal component analysis (A) and Correlation matrix (B).

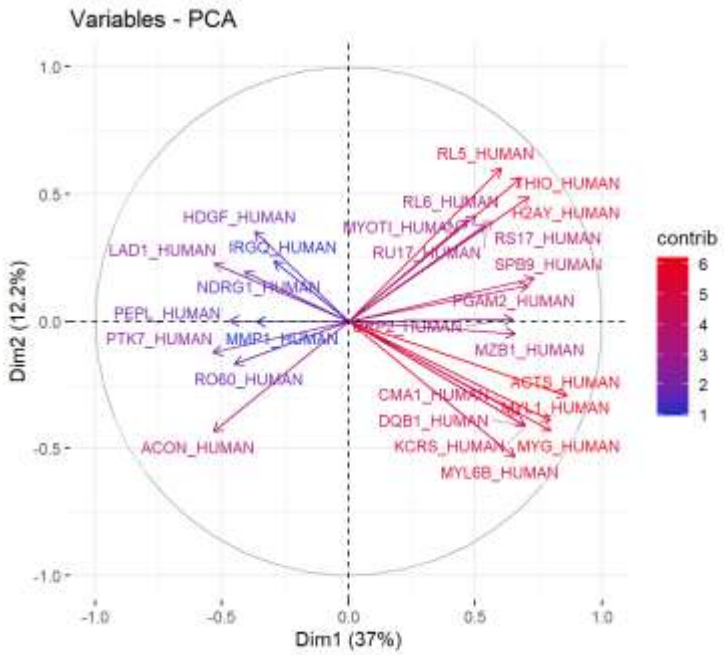

Figure 6: Correlation circle of principal component analysisof proteins and tumor location to the two-dimensional construct.

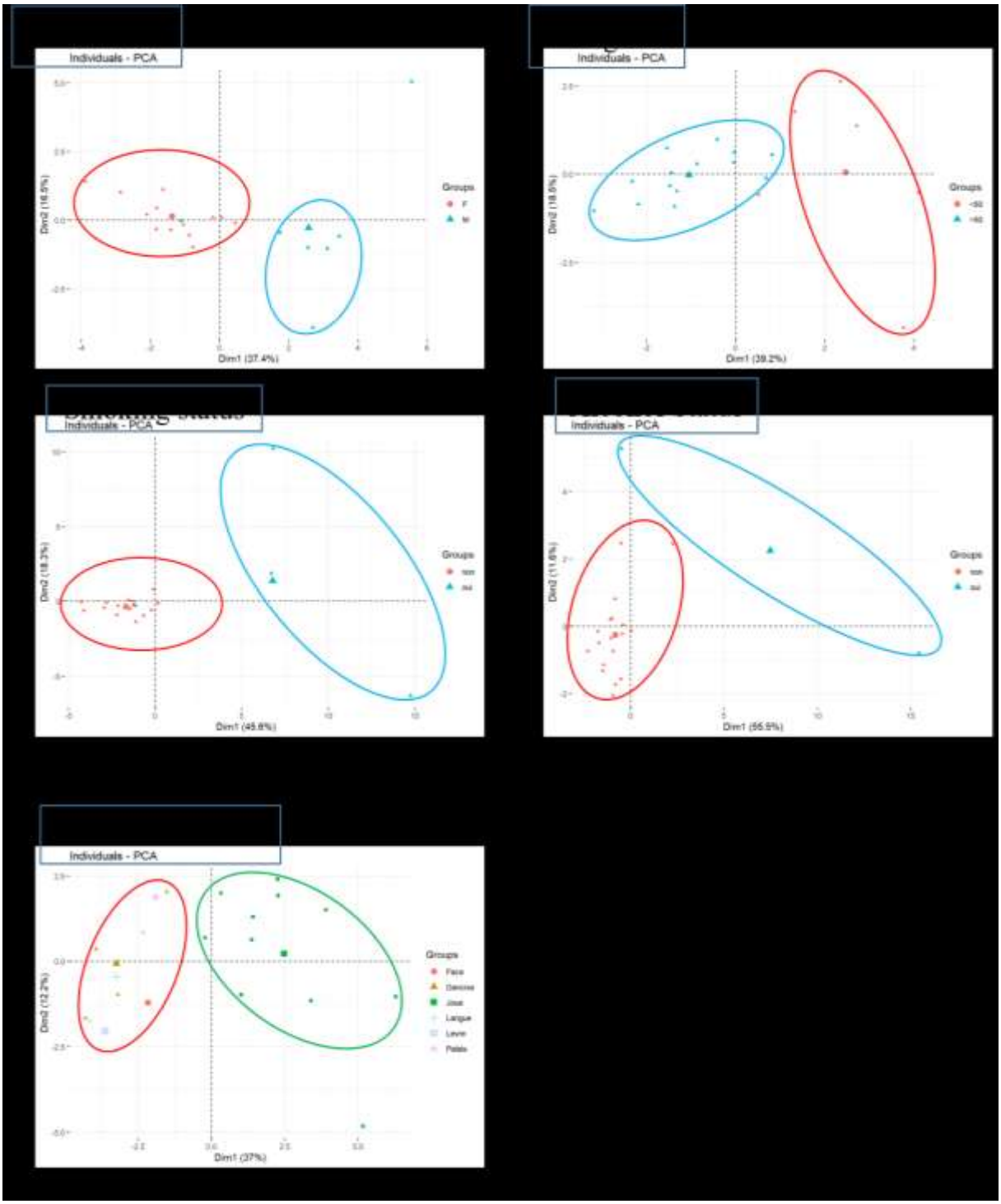

Figure 7: Graph of principal component analysis for individuals. Principal component analysis including all statistically significant proteins detected in the two populations making up each parameter (blue, red and green dots).Proximities between individuals are interpreted in terms of similarity of observed values.

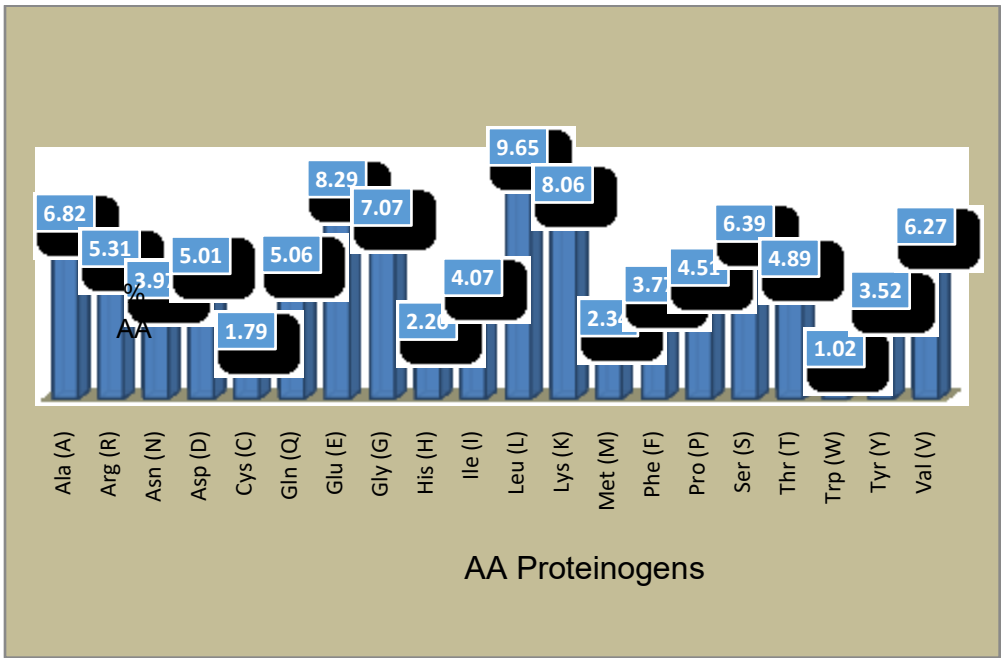

Figure 8: Percentage of amino acid composition of proteins with a good signature for clinicopathological parameters (https://www.uniprot.org/uniprotkb). Distribution of amino acid frequencies in protein composition. Leucine, Glutamate, Lysine and Glycine are the most represented.