



Article

Prediction of Oral Cancer Biomarkers by Salivary Proteomics Data

Veronica Remori ¹, Manuel Airoidi ², Tiziana Alberio ¹, Mauro Fasano ¹ and Lorenzo Azzi ^{2,*}

¹ Department of Science and High Technology, University of Insubria, 22100 Como, Italy; vremori@uninsubria.it (V.R.); tiziana.alberio@uninsubria.it (T.A.); mauro.fasano@uninsubria.it (M.F.)

² Department of Medicine and Technological Innovation, University of Insubria, 21100 Varese, Italy; manuel.airoidi@uninsubria.it

* Correspondence: l.azzi@uninsubria.it

Abstract: Oral cancer, representing 2–4% of all cancer cases, predominantly consists of Oral Squamous Cell Carcinoma (OSCC), which makes up 90% of oral malignancies. Early detection of OSCC is crucial, and identifying specific proteins in saliva as biomarkers could greatly improve early diagnosis. Here, we proposed a strategy to pinpoint candidate biomarkers. Starting from a list of salivary proteins detected in 10 OSCC patients and 20 healthy controls, we combined a univariate approach and a multivariate approach to select candidates. To reduce the number of proteins selected, a Protein–Protein Interaction network was built to consider only connected proteins. Then, an over-representation analysis (ORA) determined the enriched pathways. The network from 172 differentially abundant proteins highlighted 50 physically connected proteins, selecting relevant candidates for targeted experimental validations. Notably, proteins like Heat shock 70 kDa protein 1A/1B, Pyruvate kinase PKM, and Phosphoglycerate kinase 1 were suggested to be differentially regulated in OSCC patients, with implications for oral carcinogenesis and tumor growth. Additionally, the ORA revealed enrichment in immune system, complement, and coagulation pathways, all known to play roles in tumorigenesis and cancer progression. The employed method has successfully identified potential biomarkers for early diagnosis of OSCC using an accessible body fluid.

Keywords: OSCC; proteins; salivary biomarkers; differential abundance analysis; PPI network; ORA



Citation: Remori, V.; Airoidi, M.; Alberio, T.; Fasano, M.; Azzi, L. Prediction of Oral Cancer Biomarkers by Salivary Proteomics Data. *Int. J. Mol. Sci.* **2024**, *25*, 11120. <https://doi.org/10.3390/ijms252011120>

Academic Editor: Shun-Fa Yang

Received: 12 September 2024

Revised: 12 October 2024

Accepted: 14 October 2024

Published: 16 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Oral Squamous Cell Carcinoma (OSCC) accounts for approximately 90% of oral malignancies and is the most representative tumor in the group of Head and Neck Squamous Cell Carcinoma (HNSCC) [1]. With 377,713 new cases recorded in 2020, OSCC represents the 16th most common malignancy worldwide [2]. Risk factors commonly associated with the onset of the disease are tobacco, alcohol, and betel quid consumption, but 10–15% of patients have not been exposed to traditional risks [3]. In addition, there is an alarming increase in the incidence of OSCC in young adults, and this trend is independent of Human Papilloma Virus (HPV) infection, a feature that differentiates this tumor from oropharyngeal cancer [4]. Unfortunately, the most recent advances in the field of cancer therapy, like immunotherapeutic approaches, have not changed the clinical scenario of the disease, with the 5-year survival rate steadily assessed around 50% over the last few decades [5]. Therefore, the only effective strategy to reduce the morbidity and mortality associated with this tumor is early detection, which relies on clinical examination and histopathological assessment of suspected lesions. Within this frame, the detection of surrogate biomarkers in biological fluids associated with the early stages of the disease could represent an innovative auxiliary approach that supports the clinician in adopting an effective prevention strategy [6]. In the case of oral cancer, saliva could represent the biological fluid that is best indicated for the detection of biomarkers, since tumors of the oral cavity are constantly in contact with the secreted saliva for a prolonged period [7].

The interest in salivary proteins has significantly increased in the last five years, leading to the creation of several databases containing data retrieved from untargeted and targeted quantitative proteomics approaches, such as shotgun proteomics and multiple reaction monitoring, respectively. Among them, the Human Salivary Proteome Wiki (HSP Wiki) [8] stores custom-curated experimental data of the saliva proteome. It catalogues all data according to several pieces of information, such as sample ID, sample title, tissue type, institution, disease, and protein count. The HSP Wiki also offers analytic tools to perform differential expression analysis comparing protein levels in selected datasets. The analysis of differential protein abundance is crucial in finding biomarkers that are significantly different between two groups of subjects, namely healthy subjects and disease carriers. The classical technique identifies which proteins are up-regulated and down-regulated through the comparison between the abundance of each protein in two groups [9]. Even if this technique has already been employed in several proteomics works [10,11], the univariate comparison of the protein abundance lacks information on the inter-relationship among proteins, which is a fundamental factor when considering multifactorial diseases, such as OSCC. Thus, it is essential to consider new methods, including the use of multivariate analysis techniques, in conjunction with Network Medicine approaches.

Indeed, in recent years, Network Medicine has been increasingly exploited to obtain global information on the proteins involved in a disease and to identify protein targets for therapy [12]. In this context, Protein–Protein Interaction (PPI) networks are useful to consider the physical interactions (edges) between proteins (nodes) on a genome-wide scale. PPIs are biologically meaningful since many functionally similar proteins form complexes and often carry out their function through interactions with other proteins [13].

PPI networks may be further analyzed in terms of pathways to show how protein clusters are functionally linked to specific biological processes [14]. Pathway analysis is employed to identify activated metabolic or signaling mechanisms from omics data, considering the actual functional factors involved in regulatory pathways [15], and it is usually performed by using an over-representation analysis (ORA). An ORA is a statistical method used to highlight relevant pathways and provide functional explanations for changes in protein abundance revealed by differential quantitative proteomics [16]. ORA determines whether genes from pre-defined datasets are present more than it would be expected in a subset of data. It assumes that relevant pathways can be detected if the number of differentially expressed genes exceeds the number of genes randomly expected within a given pathway or biological process. The probability value for the null hypothesis is typically computed by the Fisher exact test with Benjamini–Hochberg multiple-test correction [17].

Here, using available curated proteomic data, we integrated univariate and multivariate analyses in order to select potential salivary biomarkers for OSCC.

2. Results

A dataset comprising 2815 proteins detected in the saliva of 10 OSCC patients and 20 healthy controls was downloaded from the HSP Wiki (Table S1, Figure S1).

As a first approach, differential abundance analysis was performed, resulting in the identification of 150 differentially abundant proteins with a \log_2 fold change higher than 1.3 or lower than -1.3 , representing 30% variation (75 up-regulated, 75 down-regulated) and with an adjusted p -value lower than 0.05. This subset of proteins represented around 5.3% of the input list (Figure 1, Table S2).

However, the classic univariate approach identifies all those proteins whose level is significantly different between OSCC and healthy controls. Several groups of proteins may change in a correlated way, thus reflecting the presence of protein communities characterized by inter-relationships. To consider such proteins' inter-relationships, we applied a multivariate approach, based on the Sparse Partial Least Squares Discriminant Analysis (sPLS-DA). After Principal Component Analysis (Figure S2) and the evaluation of the best number of components (Figure S3), the sPLS-DA was performed considering

only two components (Figure 2a). The sPLS-DA coupled with recursive feature elimination (RFE) was iteratively applied until 141 proteins were left (5% of the input list). At the end of the process, the first component explained 45% of the variance between the two groups (Figure 2b).

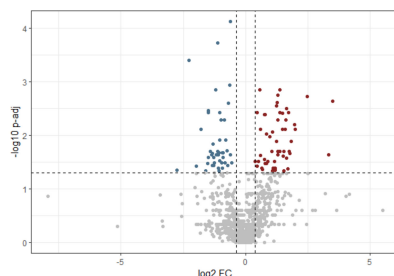


Figure 1. Volcano plot of differential abundance analysis results obtained by using DE analysis tool from HSP Wiki. Red = up-regulated proteins, blue = down-regulated proteins, gray = non-significant differentially abundant proteins (\log_2 (FC) higher than 1.3 or lower than -1.3 , representing 30% variation; $p\text{-adj} > 0.05$).

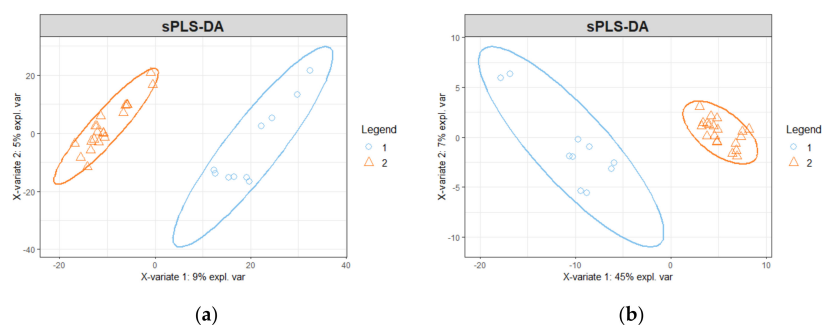


Figure 2. sPLS-DA analysis. Group: 1 = OSCC patients; 2 = healthy controls. (a) sPLS-DA performed on initial matrix. X axis = component 1 (it explains 9% of variance). Y axis = component 2 (it explains 5% of variance). (b) sPLS-DA performed on final matrix after RFE. X axis = component 1 (it explains 45% of variance). Y axis = component 2 (it explains 7% of variance).

Among the proteoforms selected by the univariate and multivariate approaches, 119 were in common (Figure 3), including 38 isoforms. The combination of the two lists comprised 172 differentially abundant proteoforms. Out of them, 54 were isoforms. Thus, the 118 unique IDs (not considering isoforms) were used as input to query the IMEx database to generate a PPI network. After removing the first interactors, self-loops, and duplicated edges, the resulting PPI network was composed of 48 connected proteins (30 up-regulated, 18 down-regulated) (Figures 4 and 5). These proteins are listed in Table 1. Thus, the PPI network shows proteins directly interacting together among the biomarkers list and were those selected for the further analysis.

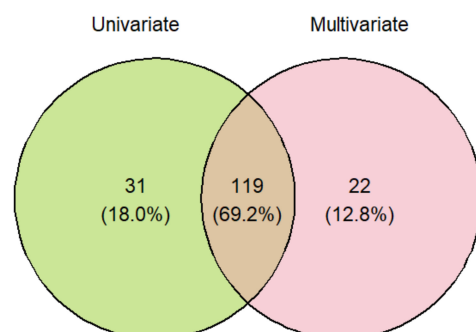


Figure 3. A Venn diagram of the proteoforms selected in the two approaches.

Table 1. The list of connected proteins annotated with the log₂ FC value and the FDR of the univariate analysis ^a.

Uniprot Accession	Protein Name	log ₂ FC	FDR
P08107	Heat shock 70 kDa protein 1A/1B	−0.62	0.000075
P14618	Pyruvate kinase PKM	−1.12	0.000187
P00558	Phosphoglycerate kinase 1	−2.28	0.000396
P0C0L5	Complement C4-B	2.48	0.001869
P07237	Protein disulfide-isomerase	−0.68	0.002520
P02671	Fibrinogen alpha chain	1.73	0.003736
P02748	Complement component C9	1.39	0.003736
P05156	Complement factor I	0.47	0.003736
P04406	Glyceraldehyde-3-phosphate dehydrogenase	−1.50	0.003736
P02675	Fibrinogen beta chain	1.48	0.003894
P02751	Fibronectin	1.63	0.005073
P01031	Complement C5	0.59	0.007604
P02679	Fibrinogen gamma chain	1.60	0.007648
P05155	Plasma protease C1 inhibitor	1.49	0.007648
P01024	Complement C3	1.08	0.008591
P08603	Complement factor H	0.97	0.010488
P10599	Thioredoxin	−1.02	0.012139
P05164	Myeloperoxidase	1.23	0.012872
P34932	Heat shock 70 kDa protein 4	−0.73	0.019327
P02768	Serum albumin	1.05	0.019750
P62805	Histone H4	1.38	0.020198
P18669	Phosphoglycerate mutase 1	−1.07	0.021136
P04075	Fructose-bisphosphate aldolase A	−0.89	0.021136
P02652	Apolipoprotein A-II	3.34	0.023220
P60709	Actin, cytoplasmic 1	−0.60	0.023220
Q96HE7	ERO1-like protein alpha	−0.90	0.025640
P06733	Alpha-enolase	−1.04	0.029925
P12956	X-ray repair cross-complementing protein 6	0.40	0.030974
P11142	Heat shock cognate 71 kDa protein	−0.56	0.032450
P00338	L-lactate dehydrogenase A chain	−0.83	0.032450
Q99829	Copine-1	0.85	0.032993
P17931	Galectin-3	−1.32	0.036387
P61158	Actin-related protein 3	−1.28	0.036525
Q99623	Prohibitin-2	0.66	0.040969
P99999	Cytochrome c	1.19	0.040969
O75083	WD repeat-containing protein 1	−1.08	0.040969
P12724	Eosinophil cationic protein	1.69	0.042247
P19827	Inter-alpha-trypsin inhibitor heavy chain H1	0.69	0.043209
P00738	Haptoglobin	1.18	0.044182
P62318	Small nuclear ribonucleoprotein Sm D3	1.09	0.045532
P61457	Pterin-4-alpha-carbinolamine dehydratase	0.87	0.050922
P02654	Apolipoprotein C-I	2.52	0.051262
P00736	Complement C1r subcomponent	0.33	0.051262
P08311	Cathepsin G	1.38	0.053989
O14732	Inositol monophosphatase 2	−0.43	0.101858
P0C0L4	Complement C4-A	1.60	0.115126
P36222	Chitinase-3-like protein 1	1.09	0.125089
P27105	Erythrocyte band 7 integral membrane protein	1.50	0.150589

^a Proteins with an FDR > 0.05 are those added by the multivariate approach.

To investigate the functional roles of the mapped proteins within known pathways, we performed an over-representation analysis interrogating KEGG and Reactome as pathway databases, using the 48 proteins listed in Table 1 as the input list. KEGG analysis revealed 12 enriched categories (three after redundancy reduction), while Reactome revealed 156 enriched categories (16 after redundancy reduction). Both analyses showed enrichment in pathways related to immune processes (innate immunity, neutrophil degranulation, complement, and coagulation) among those with a lower FDR (Figure 6). GO cellular component analysis revealed eight enriched categories (three after redundancy reduction), with significant enrichment of proteins localized in the extracellular region (Figure 7). These results showed additional significantly enriched pathways with respect to the ORA performed on the 119 proteoforms resulting from the intersection of the univariate and multivariate analyses (Figure S4).

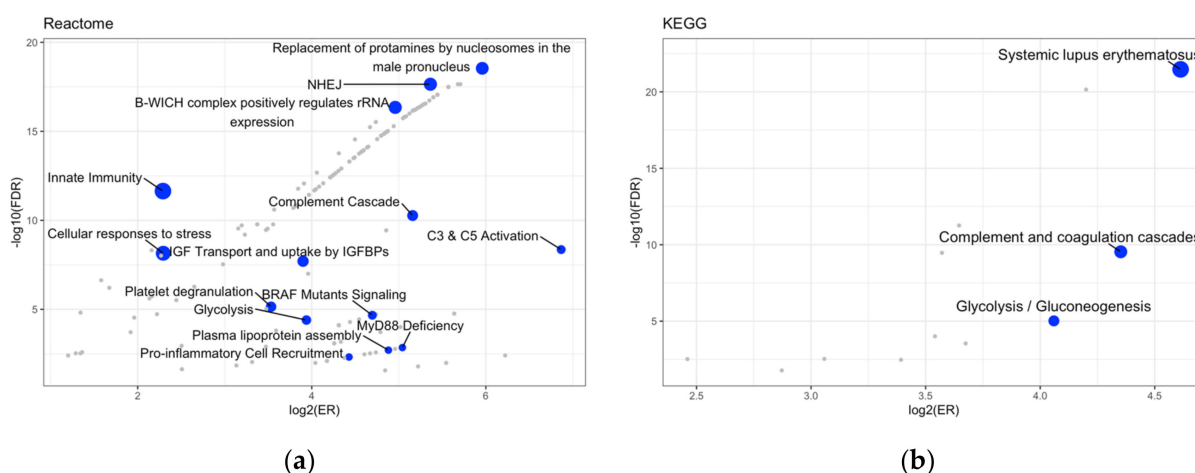


Figure 6. Volcano plots of the ORA results. The size of each dot is related to the number of proteins present in the overlap between the input data and the signature. Blue = pathways selected through affinity propagation. Gray = other enriched pathways. (a) Enriched pathways in the Reactome database. (b) Enriched pathways in the KEGG database.

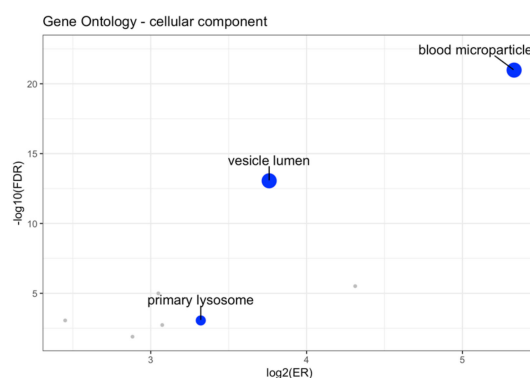


Figure 7. A volcano plot of the ORA results in the Gene Ontology cellular component database. The size of each dot is related to the number of proteins present in the overlap between the input data and the signature. Blue = the GO terms selected through affinity propagation. Gray = other enriched GO terms.

3. Discussion

Oral Squamous Cell Carcinoma represents the most common malignant tumor of the oral cavity, accounting for 90% of oral malignancies. Despite traditional risk factors like tobacco and alcohol consumption, a significant number of patients remain unexposed to these factors. The increasing incidence of OSCC among young adults, regardless of HPV infection, presents a challenge to current therapeutic approaches, with a stagnant 5-year survival rate of approximately 50%. Early detection through biomarker identification in saliva offers a promising avenue for effective prevention strategies against this disease.

In recent years, there has been a surge of interest in salivary proteins, leading to the development of multiple databases. The examination of differential protein abundance plays a crucial role in identifying biomarkers that show significant differences between healthy individuals and those with the disease. By comparing protein abundance levels between these two groups, traditional techniques can identify up-regulated or down-regulated proteins, aiding in the search for effective biomarkers or evidencing proteins involved in the pathogenesis.

However, when studying OSCC, it is imperative to move beyond traditional univariate analysis, which solely compares individual protein abundance levels between OSCC patients and healthy controls. While this approach provides valuable insights into specific

proteins, it overlooks the interconnectedness and collective impact of these proteins within the biological system. To gain a more comprehensive understanding of the molecular landscape and how these proteins collectively contribute to OSCC pathogenesis, a multivariate approach is essential.

Thus, the integration of univariate and multivariate approaches is valuable for identifying proteins that not only show significant differences in abundance between the two groups, but also play a crucial role in discriminating between them and evidencing disease molecular cues. The PPI network obtained from the list of 172 differentially abundant proteoforms highlighted 48 proteins as physically connected. This workflow allowed us to select features probably more relevant to the disease and to obtain a list of candidate markers for experimental validation. Indeed, retaining a smaller protein subset could be useful for performing targeted analyses.

Furthermore, the list contained proteins that were already suggested to be differentially regulated in OSCC patients. For instance, starting from the top of the list, the protein with the lowest FDR value was the Heat Shock 70 kDa protein 1A/1B (P08107). Several studies demonstrated that it provides a crucial contribution to oral carcinogenesis, and its expression could potentially be used as predictor of progression and recurrence of OSCC [18]. Also, its down-regulation in cancer cell leads to phenotypic alterations [19]. Additionally, the level of two key glycolytic enzymes, Pyruvate kinase PKM (P14618) and Phosphoglycerate kinase 1 (P00558), was also found to be altered in saliva. These proteins were related to carcinogenesis and OSCC beyond their role in glycolysis [20,21]. Moreover, they contribute to cancer progression and angiogenesis. In addition, the fragments containing Complement C4-B (P0C0L5) can be detected in oral primary tumors and are increased in saliva from patients with OSCC [22]. This protein is one of the chemical barrier proteins suggested to show altered expression between patients with OSCC and controls, along with Fibrinogen alpha chain (P02671), Complement component C9 (P02748), Complement factor I (P05156), Fibrinogen beta chain (P02675), Complement C5 (P01031), Fibrinogen gamma chain (P02679), Plasma protease C1 inhibitor (P05155), Complement C3 (P01024), Complement factor H (P08603), Apolipoprotein A-II (P02652), Eosinophil cationic protein (P12724), Inter-alpha-trypsin inhibitor heavy chain H1 (P19827), Haptoglobin (P00738), and Complement C1r subcomponent (P00736). All of these proteins were more present in the saliva of patients with OSCC compared with controls [23].

To the best of our knowledge, the other proteins listed in Table 1 have not yet been associated with OSCC and may be considered new potential biomarkers that require further validation.

Finally, to understand the biological significance of the identified proteins, an ORA was performed. The 48 connected and differentially expressed proteins were enriched in pathways related to immune system, complement, and coagulation processes. Alterations in immune system processes are known to enhance tumorigenesis and cancer progression [24]. Furthermore, complement processes (as part of immune processes) have a role in the initiation of OSCC and other tumor types [25,26]. Coagulation is the process by which a blood clot is generated. OSCC and other cancer types are frequently associated with hemostatic derangement [27,28]. However, given that these pathways are very generic, the specificity of our signature should be verified. For instance, it would be crucial to distinguish between minor oral lesions and actual tumor lesions. To this purpose, it would be useful to collect more data by expanding the sample size, especially of the control group, in order to enhance the specificity and the diagnostic utility of this signature.

As expected, since our data are related to saliva, Gene Ontology analysis demonstrated an enrichment in proteins localized to the extracellular region. Worthy of note is that the most significant term is “blood microparticles”. Microparticles are phospholipid microvesicles derived from different cell types, e.g., blood cells and platelets, recently related to coagulation and angiogenesis in OSCC patients [29].

In conclusion, the combined univariate and multivariate approach, along with the construction of a Protein–Protein Interaction network, has proven to be a robust and

effective method for identifying potential biomarkers in OSCC. By narrowing down the list of proteins to those physically connected in the network, the analysis has provided a focused set of candidates for further experimental validation. To this end, in the near future, we plan to analyze a homogeneous cohort of independently enrolled patients to assess the sensitivity and specificity of our proposed signature. This will be achieved using machine learning approaches, such as the Sparse Partial Least Squares Discriminant Analysis with a training set and a test set for cross validation [30,31].

Moreover, the over-representation analysis further strengthened the validity of the findings by revealing enrichment in key pathways related to tumorigenesis and cancer progression. This method not only offers insights into the molecular mechanisms underlying oral cancer but also holds promise for improving early diagnosis and personalized treatment strategies in the future. Indeed, new signatures specific for treatment response could be obtained by enrolling new subjects stratified in terms of treatment response.

4. Materials and Methods

4.1. Differential Abundance Analysis

Lists of proteins detected in saliva in 10 OSCC patients and 20 healthy subjects were downloaded from the HSP Wiki, selecting samples provided by the J. Craig Venter Institute. The inclusion criteria for their study required that patients had undergone surgery as their primary treatment, while those who had received prior chemotherapy, radiation, or other cancer medications were excluded [32].

To retrieve more information, two different approaches were adopted to analyze the differential protein abundance: a univariate approach and a multivariate approach. Using the DE analysis tool from the HSP Wiki, the expression in the two groups (OSCC and healthy) was compared for each protein by computing the \log_2 mean fold changes and employing the Wilcoxon rank-sum test. The fold change value was computed as the ratio between the average abundance of the two groups (OSCC/healthy). The resulting p -values were corrected to the false discovery rate (FDR) according to the Benjamini–Hochberg procedure. Only proteins with an expression difference higher than $\pm 30\%$ (up-regulated or down-regulated) and with an adjusted p -value lower than 0.05 were selected.

Then, we developed a novel method using a script in R (version 4.2.0.) to identify proteins that significantly influenced the distinction between the two groups (OSCC and healthy) using the Sparse Partial Least Squares Discriminant Analysis (sPLS-DA) [31]. This classifier is a variation of the sPLS regressor [33]. In essence, the algorithm maximized the distance between scores assigned to individuals grouped by a specific factor. The scores (S_j) were derived through a unitary transformation of the covariance matrix, ensuring that the elements were not part of a single eigenstructure. The diagonalization of such matrix led to the transformation matrix U and the diagonal matrix D so that $U \times Z = D \times Z$, where Z is the penalized covariance matrix. Therefore, U contained the eigenvectors \rightarrow_{u_i} and D the eigenvalues d_{ii} . Each sample j could be represented in a low-dimensionality space with one or two scores that were computed as follows:

$$S_j = \sum_i u_{i1} r_{ij}$$

where r_{ij} indicates the abundance of protein i in sample j . The u_{i1} values are the loadings on the first component, i.e., the vector with the weights of the contribution of each original variable to the corresponding latent component. A recursive feature elimination (RFE) technique was incorporated into the sPLS-DA to eliminate proteins with less impact on the discrimination between the two groups, selecting the protein with the lower loading u_{i1} . The sPLS-DA coupled with RFE was iteratively applied until 141 proteins were left (5% of the input list).

Finally, the two lists of proteins obtained from the two approaches were merged to consider proteins that are significantly differentially abundant and that also contribute the most to the difference between the two groups.

4.2. PPI Network Generation Workflow

Cytoscape 3.9.1 was used to generate a PPI network starting from the selected proteins to investigate their physical interactions [34]. The public database of the International Molecular Exchange Consortium (IMEx) was queried through Cytoscape using the PSIC-QUIC standard (the Proteomics Standard Initiative Common QUery InterfaCe) created by the Human Proteome Organization Proteomics Standards Initiative (HUPO-PSI) in 2008 [35]. First, the list of proteins was used as an input list to generate a PPI network encompassing all of their first interactors. The network was filtered for interactor type and for taxonomy ID 9606 (*Homo sapiens*) to remove homology inferences. Then, the first interactors were removed to keep only interactions between the input proteins. Also, duplicated edges and self-loops were removed. Only connected proteins were considered under the assumption that interacting proteins may participate in the same biochemical process at a certain time.

4.3. Over-Representation Analysis

An over-representation analysis was performed utilizing Webgestalt, a web application designed for enrichment analysis [36]. Webgestalt enables users to extract biological information from a list of genes (or proteins) derived from omics studies. The ORA was performed using the list of proteins retrieved from our analysis as input, setting *Homo sapiens* as the organism of interest. As functional databases, Kyoto Encyclopedia of Genes and Genomes (KEGG) [37] and Reactome [38] were selected as pathway databases. Furthermore, the Gene Ontology cellular component database [39] was incorporated. As a genomic landscape for statistical testing, “genome protein coding” was selected. Enriched categories with a Benjamini–Hochberg false discovery rate (FDR) ≤ 0.05 were considered statistically significant. To remove redundancy in the output of the ORA, the affinity propagation algorithm was employed, and the remaining pathways were selected.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms252011120/s1>.

Author Contributions: Conceptualization, V.R. and M.F.; methodology, V.R.; formal analysis, V.R. and M.A.; investigation, V.R.; data curation, V.R. and M.A.; writing—original draft preparation, V.R. and M.A.; writing—review and editing, M.F., T.A. and L.A.; visualization, V.R.; supervision, M.F., T.A. and L.A.; project administration, L.A.; funding acquisition, L.A. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by intramural funding to LA, University of Insubria, Varese. In addition, it was supported by a liberal donation provided by “Romeo De Molli” Trust, Castronno (VA), Italy.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study were derived from the following resources available in the public domain: <https://www.salivaryproteome.org/salivary-protein> (accessed on 30 July 2024).

Acknowledgments: We are grateful to the “Romeo De Molli” Trust for the invaluable support provided for our study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Chamoli, A.; Gosavi, A.S.; Shirwadkar, U.P.; Wangdale, K.V.; Behera, S.K.; Kurrey, N.K.; Kalia, K.; Mandoli, A. Overview of Oral Cavity Squamous Cell Carcinoma: Risk Factors, Mechanisms, and Diagnostics. *Oral Oncol.* **2021**, *121*, 105451. [[CrossRef](#)] [[PubMed](#)]
2. International Agency for Research on Cancer (IARC). Global Cancer Observatory. Available online: <https://gco.iarc.fr/> (accessed on 10 August 2024).

3. Wyss, A.; Hashibe, M.; Chuang, S.-C.; Lee, Y.-C.A.; Zhang, Z.-F.; Yu, G.-P.; Winn, D.M.; Wei, Q.; Talamini, R.; Szeszenia-Dabrowska, N.; et al. Cigarette, Cigar, and Pipe Smoking and the Risk of Head and Neck Cancers: Pooled Analysis in the International Head and Neck Cancer Epidemiology Consortium. *Am. J. Epidemiol.* **2013**, *178*, 679–690. [[CrossRef](#)]
4. Tran, Q.; Maddineni, S.; Arnaud, E.H.; Divi, V.; Megwalu, U.C.; Topf, M.C.; Sunwoo, J.B. Oral Cavity Cancer in Young, Non-Smoking, and Non-Drinking Patients: A Contemporary Review. *Crit. Rev. Oncol./Hematol.* **2023**, *190*, 104112. [[CrossRef](#)] [[PubMed](#)]
5. Nindra, U.; Hurwitz, J.; Forstner, D.; Chin, V.; Gallagher, R.; Liu, J. A Systematic Review of Neoadjuvant and Definitive Immunotherapy in Locally Advanced Head and Neck Squamous Cell Carcinoma. *Cancer Med.* **2023**, *12*, 11234–11247. [[CrossRef](#)]
6. Gissi, D.B.; Rossi, R.; Lenzi, J.; Tarsitano, A.; Gabusi, A.; Balbi, T.; Montebugnoli, L.; Marchetti, C.; Foschini, M.P.; Morandi, L. Thirteen-Genes DNA Methylation Analysis of Oral Brushing Samples: A Potential Surveillance Tool for Periodic Monitoring of Treated Patients with Oral Cancer. *Head Neck* **2024**, *46*, 728–739. [[CrossRef](#)]
7. Raslan, S.; Smith, D.H.; Reis, I.M.; Peifer, S.J.; Forman, G.; Ezeh, U.C.; Joshi, P.; Koester, M.; Buitron, I.; Al-Awady, A.; et al. Soluble CD44 in Oral Rinses for the Early Detection of Cancer: A Prospective Cohort Study in High-Risk Individuals. *BMC Oral Health* **2024**, *24*, 820. [[CrossRef](#)]
8. Lau, W.W.; Hardt, M.; Zhang, Y.H.; Freire, M.; Ruhl, S. The Human Salivary Proteome Wiki: A Community-Driven Research Platform. *J. Dent. Res.* **2021**, *100*, 1510–1519. [[CrossRef](#)]
9. Grishin, N.V. Fold Change in Evolution of Protein Structures. *J. Struct. Biol.* **2001**, *134*, 167–185. [[CrossRef](#)]
10. Riccardi, G.; Bellizzi, M.G.; Fatuzzo, I.; Zoccali, F.; Cavalcanti, L.; Greco, A.; de Vincentiis, M.; Ralli, M.; Fiore, M.; Petrella, C.; et al. Salivary Biomarkers in Oral Squamous Cell Carcinoma: A Proteomic Overview. *Proteomes* **2022**, *10*, 37. [[CrossRef](#)]
11. Chu, H.-W.; Chang, K.-P.; Hsu, C.-W.; Chang, I.Y.-F.; Liu, H.-P.; Chen, Y.-T.; Wu, C.-C. Identification of Salivary Biomarkers for Oral Cancer Detection with Untargeted and Targeted Quantitative Proteomics Approaches*[S]. *Mol. Cell. Proteom.* **2019**, *18*, 1796–1806. [[CrossRef](#)]
12. Barabási, A.-L.; Gulbahce, N.; Loscalzo, J. Network Medicine: A Network-Based Approach to Human Disease. *Nat. Rev. Genet.* **2011**, *12*, 56–68. [[CrossRef](#)] [[PubMed](#)]
13. Wang, S.; Wu, R.; Lu, J.; Jiang, Y.; Huang, T.; Cai, Y.-D. Protein-Protein Interaction Networks as Miners of Biological Discovery. *Proteomics* **2022**, *22*, e2100190. [[CrossRef](#)] [[PubMed](#)]
14. Monti, C.; Zilocchi, M.; Colognat, I.; Alberio, T. Proteomics Turns Functional. *J. Proteom.* **2019**, *198*, 36–44. [[CrossRef](#)]
15. Curtis, R.K.; Oresic, M.; Vidal-Puig, A. Pathways to the Analysis of Microarray Data. *Trends Biotechnol.* **2005**, *23*, 429–435. [[CrossRef](#)]
16. Khatri, P.; Sirota, M.; Butte, A.J. Ten Years of Pathway Analysis: Current Approaches and Outstanding Challenges. *PLoS Comput. Biol.* **2012**, *8*, e1002375. [[CrossRef](#)]
17. Yeganeh, P.N.; Mostafavi, M.T. Causal Disturbance Analysis: A Novel Graph Centrality Based Method for Pathway Enrichment Analysis. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2020**, *17*, 1613–1624. [[CrossRef](#)]
18. Elhendawy, H.A. Clinical Implications of Heat Shock Protein 70 in Oral Carcinogenesis and Prediction of Progression and Recurrence in Oral Squamous Cell Carcinoma Patients: A Retrospective Clinicopathological Study. *Eur. J. Med. Res.* **2023**, *28*, 464. [[CrossRef](#)]
19. Kasioumi, P.; Vrazeli, P.; Vezyraki, P.; Zerikiotis, S.; Katsouras, C.; Damalas, A.; Angelidis, C. Hsp70 (HSP70A1A) Downregulation Enhances the Metastatic Ability of Cancer Cells. *Int. J. Oncol.* **2019**, *54*, 821–832. [[CrossRef](#)]
20. Kurihara-Shimomura, M.; Sasahira, T.; Nakashima, C.; Kuniyasu, H.; Shimomura, H.; Kirita, T. The Multifarious Functions of Pyruvate Kinase M2 in Oral Cancer Cells. *Int. J. Mol. Sci.* **2018**, *19*, 2907. [[CrossRef](#)]
21. Chen, Y.; Cen, L.; Guo, R.; Huang, S.; Chen, D. Roles and Mechanisms of Phosphoglycerate Kinase 1 in Cancer. *Bull. Du Cancer* **2022**, *109*, 1298–1307. [[CrossRef](#)]
22. Ajona, D.; Pajares, M.; Chiara, M.; Rodrigo, J.; Jantus-Lewintre, E.; Camps, C.; Suarez, C.; Bagán, J.; Montuenga, L.; Pio, R. Complement Activation Product C4d in Oral and Oropharyngeal Squamous Cell Carcinoma. *Oral Dis.* **2015**, *21*, 899–904. [[CrossRef](#)] [[PubMed](#)]
23. Kalló, G.; Bertalan, P.M.; Márton, I.; Kiss, C.; Csósz, É. Salivary Chemical Barrier Proteins in Oral Squamous Cell Carcinoma—Alterations in the Defense Mechanism of the Oral Cavity. *Int. J. Mol. Sci.* **2023**, *24*, 13657. [[CrossRef](#)] [[PubMed](#)]
24. Caponio, V.C.A.; Zhurakivska, K.; Muzio, L.L.; Troiano, G.; Cirillo, N. The Immune Cells in the Development of Oral Squamous Cell Carcinoma. *Cancers* **2023**, *15*, 3779. [[CrossRef](#)] [[PubMed](#)]
25. Afshar-Kharghan, V. The Role of the Complement System in Cancer. *J. Clin. Investig.* **2017**, *127*, 780–789. [[CrossRef](#)] [[PubMed](#)]
26. Gallenkamp, J.; Spanier, G.; Wörle, E.; Englbrecht, M.; Kirschfink, M.; Greslechner, R.; Braun, R.; Schäfer, N.; Bauer, R.J.; Pauly, D. A Novel Multiplex Detection Array Revealed Systemic Complement Activation in Oral Squamous Cell Carcinoma. *Oncotarget* **2017**, *9*, 3001–3013. [[CrossRef](#)]
27. Falanga, A.; Marchetti, M.; Vignoli, A. Coagulation and Cancer: Biological and Clinical Aspects. *J. Thromb. Haemost.* **2013**, *11*, 223–233. [[CrossRef](#)]
28. Racine, F.; Soudet, S.; Sevestre, M.-A.; Galmiche, A.; Saidak, Z. The Coagulum of Oral Squamous Cell Carcinoma: Examining the Role and Regulation of Coagulation in Oral Cancers Using a Systems Approach. *Curr. Opin. Otolaryngol. Head Neck Surg.* **2023**, *31*, 73. [[CrossRef](#)]

29. Ren, J.G.; Zhang, W.; Liu, B.; Man, Q.W.; Xiong, X.P.; Li, C.; Zhu, J.Y.; Wang, W.M.; Jia, J.; Sun, Z.J.; et al. Clinical Significance and Roles in Angiogenesis of Circulating Microparticles in Oral Cancer. *J. Dent. Res.* **2016**, *95*, 860–867. [[CrossRef](#)]
30. Ogata, S.; Masuda, T.; Ito, S.; Ohtsuki, S. Targeted Proteomics for Cancer Biomarker Verification and Validation. *Cancer Biomark.* **2022**, *33*, 427–436. [[CrossRef](#)]
31. Lualdi, M.; Fasano, M. Statistical Analysis of Proteomics Data: A Review on Feature Selection. *J. Proteom.* **2019**, *198*, 18–26. [[CrossRef](#)]
32. Lin, Y.-H.; Eiguez, R.V.; Torralba, M.G.; Singh, H.; Golusinski, W.; Masternak, M.; Nelson, K.E.; Freire, M.; Yu, Y. Self-Assembled STrap for Global Proteomics and Salivary Biomarker Discovery. *J. Proteome Res.* **2019**, *18*, 1907–1915. [[CrossRef](#)] [[PubMed](#)]
33. Wold, S.; Ruhe, A.; Wold, H.; Dunn, W.J., III. The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM J. Sci. Stat. Comput.* **1984**, *5*, 735–743. [[CrossRef](#)]
34. Su, G.; Morris, J.H.; Demchak, B.; Bader, G.D. Biological Network Exploration with Cytoscape 3. *Curr. Protoc. Bioinform.* **2014**, *47*, 8–13. [[CrossRef](#)] [[PubMed](#)]
35. Aranda, B.; Blankenburg, H.; Kerrien, S.; Brinkman, F.S.L.; Ceol, A.; Chautard, E.; Dana, J.M.; De Las Rivas, J.; Dumousseau, M.; Galeota, E.; et al. PSICQUIC and PSIScore: Accessing and Scoring Molecular Interactions. *Nat. Methods* **2011**, *8*, 528–529. [[CrossRef](#)]
36. Elizarraras, J.M.; Liao, Y.; Shi, Z.; Zhu, Q.; Pico, A.R.; Zhang, B. WebGestalt 2024: Faster Gene Set Analysis and New Support for Metabolomics and Multi-Omics. *Nucleic Acids Res.* **2024**, *52*, W415–W421. [[CrossRef](#)]
37. Kanehisa, M.; Furumichi, M.; Sato, Y.; Kawashima, M.; Ishiguro-Watanabe, M. KEGG for Taxonomy-Based Analysis of Pathways and Genomes. *Nucleic Acids Res.* **2022**, *51*, D587–D592. [[CrossRef](#)]
38. Milacic, M.; Beavers, D.; Conley, P.; Gong, C.; Gillespie, M.; Griss, J.; Haw, R.; Jassal, B.; Matthews, L.; May, B.; et al. The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Res.* **2024**, *52*, D672–D678. [[CrossRef](#)]
39. The Gene Ontology Consortium; Aleksander, S.A.; Balhoff, J.; Carbon, S.; Cherry, J.M.; Drabkin, H.J.; Ebert, D.; Feuermann, M.; Gaudet, P.; Harris, N.L.; et al. The Gene Ontology Knowledgebase in 2023. *Genetics* **2023**, *224*, iyad031. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.