MASS SPECTROMETRY-BASED PROTEOME PROFILE MAY BE USEFUL TO DIFFERENTIATE ADENOID CYSTIC CARCINOMA FROM POLYMORPHOUS ADENOCARCINOMA OF SALIVARY GLANDS

Felipe Paiva Fonseca[1,2], Carolina Carneiro Soares Macedo[1], Sara Ferreira dos Santos Costa[2], Adriana Franco Paes Leme[3], Romênia Ramos Rodrigues[3], Hélder Antônio Rebelo Pontes[4], Albina Altemani[5], Willie F. P. van Heerden[6], Manoela Domingues Martins[1,7], Oslei Paes de Almeida[1], Alan Roger Santos-Silva[1], Márcio Ajudarte Lopes[1] and Pablo Agustin Vargas[1,6]

1. Department of Oral Diagnosis, Oral Pathology Division, Piracicaba Dental School, University of Campinas (UNICAMP), Piracicaba, Brazil.
2. Department of Surgery and Pathology, School of Dentistry, Universidade Federal de Minas Gerais (Belo Horizonte, Brazil).
3. Laboratório Nacional de Biociências, LNBio (Campinas, Brazil).
4. Service of Oral Pathology, João de Barros Barreto University Hospital, Federal University of Pará (Belém, Brazil).
5. Department of Pathology, School of Medical Sciences, University of Campinas (UNICAMP) (Campinas, Brazil).
6. Department of Pathology, School of Dentistry, University of Pretoria (Pretoria, South Africa).
7. Department of Pathology, School of Dentistry, Federal University of Rio Grande do Sul (Porto Alegre, Brazil).

Corresponding author

Prof. Dr. Pablo Agustin Vargas – DDS, PhD, FRCPath

Department of Oral Diagnosis, Piracicaba Dental School, University of Campinas

Address: Av. Limeira, 901 Postal Code: 13414-903. Piracicaba - São Paulo - Brazil.

Tel.: +55 19 21065319; E-mail: pavargas@fop.unicamp.br

Declarations of interest: none.

Funding: This study was funded by São Paulo State Research Foundation (FAPESP) multi-user Project FAPESP 2009/54067-3 and a post-doctoral scholarship Project FAPESP 2015/16056-0.

Word count for the abstract: 200. **Manuscript word count:** 3,350. **Number of references:** 41. **Figures:** 4. **Tables:** 2. **Supplementary elements:** 7.

The current study was presented as a poster during the 44th Brazilian Meeting of Oral Pathology and Medicine in Rio de Janeiro.

Abstract

Objective: The aim of this study was to determine the proteome of adenoid cystic carcinoma (AdCC) and polymorphous adenocarcinoma (PAc) and to identify a protein signature useful to distinguish both neoplasms.

Study design: Ten cases of AdCC and 10 cases of PAc were microdissected for enrichment of neoplastic tissue. The samples were submitted to liquid chromatography-tandem mass spectrometry (LC-MS/MS) and the proteomics data were analyzed using the MaxQuant software. LC-MS/MS spectra were searched against the Human UniProt database, whereas statistical analyses were performed with Perseus software. Bioinformatic analyses were performed using discovery-based proteomic data obtained from both tumors.

Results: LC-MS/MS analysis identified 1,957 proteins. The tumors shared 1,590 proteins, whereas 261 were exclusively identified in AdCC and 106 in PAc. Clustering analysis of the statistically significant proteins clearly separated AdCC from PAc. Proteins expression higher than 10 times in one group than in the other led to a signature of 16 proteins, 6 upregulated in AdCC and 10 in PAc. A new clustering analysis showed a reverse regulation and also differentiated both tumors.

Conclusion: Global proteomics may be useful to discriminate these two malignant salivary neoplasms that frequently show clinical and microscopic overlaps, but additional validation studies are still necessary to determine the diagnostic potential of the protein signature obtained.

Keywords: Proteomics; mass spectrometry; salivary gland tumor; adenoid cystic carcinoma; polymorphous adenocarcinoma.

Introduction

Salivary gland tumors (SGTs) constitute a heterogeneous group of neoplasms with complex clinical behaviors and microscopic presentations, and account for approximately 5% of all neoplasms diagnosed in the head and neck region.[1] Adenoid cystic carcinoma (AdCC) is one of the most common salivary gland malignancies, predominantly affecting the parotid gland and the minor salivary glands.[2] It is considered a high-grade neoplasm, given its significant association with late recurrences and distant metastases, leading to an unsatisfactory long-term survival.[2, 3] Recently, Persson et al.[4] demonstrated that chromosomal translocation t(6;9)(q22-23;p23-24) resulting in a fusion of genes encoding the transcription factors *MYB* [v-myb myeloblastosis viral oncogene homolog (avian)] and *NFIB* (nuclear factor IB) would represent a specific molecular event in the context of SGTs, being associated with AdCC development.

Although t(6;9)(q22-23;p23-24) may not represent a prognostic determinant for AdCC[5], the presence of this translocation contributes to distinguish AdCC from its main differential diagnoses, which frequently represents a difficult approach given the well known clinical and microscopic overlaps of salivary gland neoplasms. However, a high number of AdCC cases remains free of t(6;9)(q22-23;p23-24) and the search for molecular markers that could better aid in its diagnosis is warranted[6].

Polymorphous adenocarcinoma (PAc) used to be considered part of the spectrum of AdCC until 1983 when it was first recognized as an independent entity, representing however, one of the main mimickers of AdCC thereafter[7]. PAc is virtually restricted to the intra-oral minor salivary glands, especially from the palate, where it usually presents as a slow-growing, non-aggressive tumor[8]. Molecular alterations have also been

described for PAc including mutation in the PRKD-1 gene[9]. However, its frequency, specificity and clinical importance must be addressed.

The search for new molecular markers that could aid in the differential diagnosis of these salivary neoplasms is still necessary, especially when only small incisional biopsies are available, which is a common scenario in histopathology practice. Recently, different groups have used proteomics to discover new prognostic and diagnostic markers for several human neoplastic and non-neoplastic diseases[10,11]. Mass spectrometry-based proteomics is an approach that provides large-scale data on the protein profile of tumor cells, tissues, and body fluids, it has already been employed to detect the proteome of melanomas, head and neck, pancreatic, lung, breast, colonic and other human cancers, assisting in the discovery of unknown biomarkers to improve their early detection and to better understand their pathogenesis[11-16]. Moreover, research groups have also used proteomics to obtain protein signatures or isolated proteins with potential to distinguish different neoplasms[14,17-19]. However, the proteome of SGT is largely unknown and the general protein expression of AdCC and PAc remains to be determined.

Therefore, to investigate the hypothesis that AdCC and PAc could be differentiated by their proteome profile, the objective of the present study is to identify a protein signature that could be useful to distinguish both neoplasms.

Material and methods

Sample collection

This study was approved by the ethics committee of the Universidade Federal de Minas Gerais (process number CAAE 65798717.3.0000.5149). We used a convenience sample consisting of formalin-fixed, paraffin-embedded (FFPE) tissues of 10 cases

diagnosed as AdCC and 10 cases diagnosed as PAc retrieved from the pathology files of the Piracicaba Dental School – University of Campinas (Piracicaba-Brazil) and the João de Barros Barreto University Hospital – Federal University of Pará (Belém-Brazil). Two authors jointly confirmed the original diagnoses using new hematoxylin and eosin (H&E)-stained slides following the current World Health Organization guidelines[20] for classification of salivary gland tumors. Clinical data regarding sex, age, tumor location and symptomatology were retrieved from the patient's medical charts.

Immunohistochemistry

Immunohistochemical reactions against low-molecular weight cytokeratin (CK7), high-molecular weight cytokeratin (CK14), S100, p63 and α-smooth muscle actin (α-SMA), which allowed the identification of luminal and myoepithelial components were used to better characterize the cases used. Briefly, the reactions were done in 3µm sections of the original FFPE tissues that were de-waxed with xylene and then hydrated in a descending ethanol series. The antigen retrieval was done and the endogenous peroxidase activity was blocked using 10% hydrogen peroxide in five baths, each of 5 minutes. After washing in PBS buffer (pH 7.4), slides were incubated overnight with primary antibodies anti-human CK7 (Leica biosystems – Wetzlar/Germany; clone OV-TL 12/30; dilution 1:200), anti-human CK14 (Leica biosystems – Wetzlar/Germany; clone LL002; dilution 1:200), anti-human-S100 (Dako – Glostrup/Denmark; polyclonal; dilution 1:100), anti-human p63 (Abcam – California/USA; clone 4A4; dilution 1:100) and anti-human α-SMA (Dako – Glostrup/Denmark; clone 1A4; dilution 1:200). All slides were subsequently exposed to avidin-biotin complex and horseradish peroxidase reagents (LSAB Kit - DakoCytomation, Denamark) and diaminobenzidin tetrahydrochloride (DAB, Sigma,

St. Louis, MO, USA), and subsequently counterstained with Carazzi hematoxylin. Sections of lung carcinoma, normal skin, normal nerve, fibrous hyperplasia with overlying surface epithelium and endometrium were used as positive controls for each antibody, respectively. The negative control was obtained by omitting the primary specific antibody. Results were evaluated by one oral pathologist.

Sample preparation and laser microdissection (LMD)

A new histological slide was prepared for each of the 20 SGT investigated. Paraffin blocks were cut into microtome with a thickness of 5 μm and another slide with 10 μm of thickness. Histological section of 5 μm was stained with H&E in order to guide LMD. The other section (10 μm) was prepared using specific membrane slides for LMD (PEN Arcturus® Membrane, Life Technologies), stained with hematoxylin, submitted to dehydration in ethanol 90% and 100% and stored in containers before drying LMD.

Samples were processed using Leica Laser Microdissection Systems. Only neoplastic cells were microdissected and collected. An average of 6.000.000 μm^2 of each tissue was obtained.

All samples were collected in 600 μL microtubes and stored at -80°C. LMD was standardized for FFPE tissues with a slice depth of 10 μm. The adjustable parameters in LMD Laser Microdissection Leica software are considered optimum for these samples and the area (μm^2) cut for each patient was recorded.

Protein extraction and trypsin digestion

Protein extraction and digestion followed the protocol developed previously[21-24]. Samples were treated with 8M urea, followed by protein reduction with dithiothreitol (5 mM for 25 minutes at 56°C) and alkylation with iodoacetamide (14 mM for 30 minutes

at room temperature in the dark). For protein digestion, urea was diluted to a final concentration of 1.6 M and 1 mM of calcium chloride was added to samples for trypsin digestion during 16 h at 37°C (2 µg of trypsin). The reaction was quenched with 0.4% formic acid and peptides were desalted by stage tips C18[25].

Liquid chromatography – tandem mass spectrometry (LC-MS/MS) analysis

An aliquot containing 4.5 µl of peptide mixture was analyzed on a LTQ Orbitrap Velos (Thermo Fisher Scientific) mass spectrometer coupled to nanoflow liquid chromatography on EASY-nLC system (Proxeon Biosystems) through a Proxeon nanoelectrospray ion source. Peptides in 0.1% formic acid were separated by a 2-90% acetonitrile gradient in a PicoFrit analytical column (20 cm x ID75, 5 µm particle size, New Objective), with a flow rate of 300 nl/min over 212 minutes as previously described[12].

Proteomics Data analysis

The raw files were processed using the MaxQuant v1.3.0.3 software[24] and MS/MS spectra were searched against the Human UniProt database (release May, 2017, 92,646 sequences, 36,874,315 residues) using the Andromeda search engine[26]. A tolerance of 6 ppm was considered for precursor ions and 0.5 Da for fragment ions, Oxidation of methionine and protein N-terminal acetylation was set as variable modifications and carbamidomethylation of cysteine as fixed modification and with 2 missed trypsin cleavage. Label-free quantification (LFQ) was used for protein quantification, with a 2 min window for matching between runs and minimal ratio count set as 1. The FDR for protein and peptide was set to 1%. Statistical analysis was performed with Perseus v1.2.7.4 software[24]. Protein dataset were processed excluding

reverse sequences and 'only identified by site' entries, and then the data were transformed by log2. In addition, it was considered in the final analysis a minimum of three valid values in at least one group and statistical significance was assessed by Student's *t*-test to indicate the differentially expressed proteins ($p<0.05$).

Additionally, the files containing the identified proteins and their LFQ intensity values were used for clustering and heat map generation. Heat maps, hierarchical clustering, and principal component analysis (PCA) were constructed in the web-based chemometrics platform MetaboAnalyst 3.0 using the Pearson distance measure.

Bioinformatics analysis

Proteins with differential expression between the AdCC and PAc groups were submitted to an enrichment analysis in order to gain biological information from this list of the identified proteins. Uniprot IDs of differential proteins were submitted to the Integrated Interactome System (IIS) platform[27] to perform the enrichment analysis for the GO (Gene Ontology)[28] database. Only significantly biological processes (p-value < 0.05) were considered in the results. Protein-protein association networks were constructed in the STRING database[29].

Results

Clinicopathological data from patients

The clinicopathological features of the 20 cases investigated in this study are detailed in **Table 1.** Briefly, females predominated in both groups (7:3 in the AdCC group and 6:4 in the PAc group), with a mean age of 48.2 year-old and 54.4 years-old in the AdCC and PAc groups, respectively. The palate was the most affected site in both tumors and a painful symptomatology was present in 5 patients affected by AdCC and

in 3 affected by PAc **(Figure 1)**. Microscopically, AdCC cases demonstrated a predominance of the cribriform architectural pattern characterized by pseudo-cystic spaces filled with homogeneous hyaline substance and surrounded by neoplastic cells exhibiting scarce cytoplasm and hypercromatic nuclei. Bilayered ductal structures were also present. PAc was predominantly composed by the lobular architectural pattern with the so-called targetoid growth pattern present in three of them. Single line arrangement of neoplastic cells was also found. Tumor cells were round, with pale stained nuclei **(Figure 2)**. Immunohistochemistry was done to better characterize the cellular components of both tumors. AdCC exhibited strong positivity for CK7 (in the luminal cells of the ductal structures) and for CK14 (myoepithelial cells), as well as a strong positivity for p63 and α-SMA in the myoepithelial cells. S100 was positive only in myoepithelial cells. PAc was strongly and diffusely positive for CK7 and S100, with focal positivity to p63 and CK14, and negativity for α-SMA **(Figure 3)**.

Table 1. Clinical data of the 20 cases included in the study.

No.	Diagnosis	Sex	Age	Site	Symptomatology
1	AdCC	M	46	Minor gland - Palate	Painless
2	AdCC	F	41	Minor gland - lip	Painful
3	AdCC	F	58	Minor gland – Floor of the mouth	Painless
4	AdCC	F	69	Minor gland - Palate	Painless
5	AdCC	M	34	Minor gland – buccal mucosa	Painful
6	AdCC	F	76	Parotid gland	Painful
7	AdCC	M	52	Minor gland - Palate	Painful
8	AdCC	F	39	Minor gland - Palate	Painful
9	AdCC	F	37	Submandibular gland	Painless
10	AdCC	F	30	Minor gland - Palate	Painless
11	PAc	M	70	Minor gland - palate	Painless
12	PAc	F	51	Minor gland - palate	Painful
13	PAc	F	48	Minor gland - Palate	Painless
14	PAc	F	48	Minor gland - Palate	Painful
15	PAc	M	80	Minor gland - Upper lip	Painless
16	PAc	F	64	Minor gland - Buccal mucosa	Painless
17	PAc	F	40	Minor gland - Palate	Painless
18	PAc	M	45	Minor gland - Palate	Painless
19	PAc	F	60	Minor gland - palate	Painful
20	PAc	M	38	Minor gland – upper lip	Painless

AdCC.: Adenoid cystic carcinoma. PAc.: Polymorphous adenocarcinoma. M: Male. F: Female.

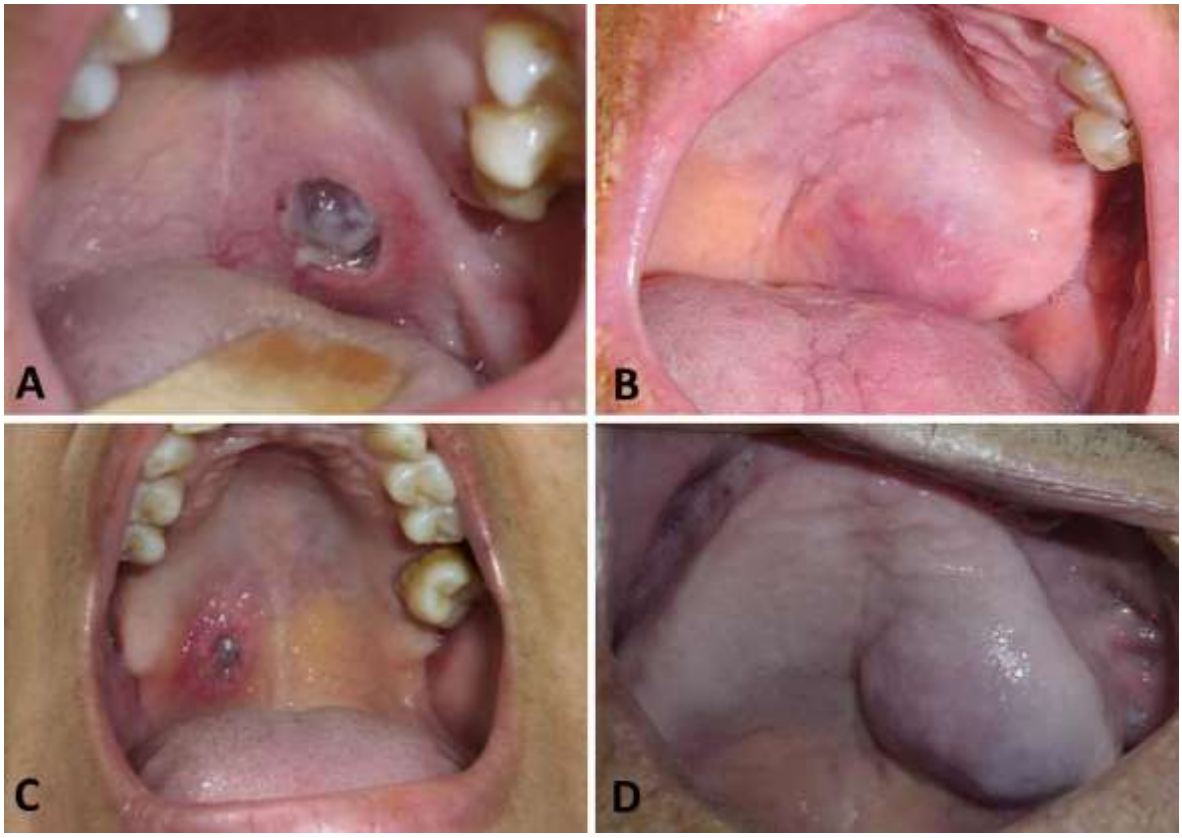

Figure 1. Representative images of AdCC and PAc demonstrating the important overlap observed in the clinical presentation of these malignant neoplasm. **A)** AdCC may clinically present as a painful, ulcerative lesion in the palate. **B)** However, it is more commonly diagnosed as an asymptomatic, slow-growing swelling affecting the minor glands of the palate. **C)** PAc may also show an ulcer or **D)** present as a slow-growing painless tumor in the oral cavity.

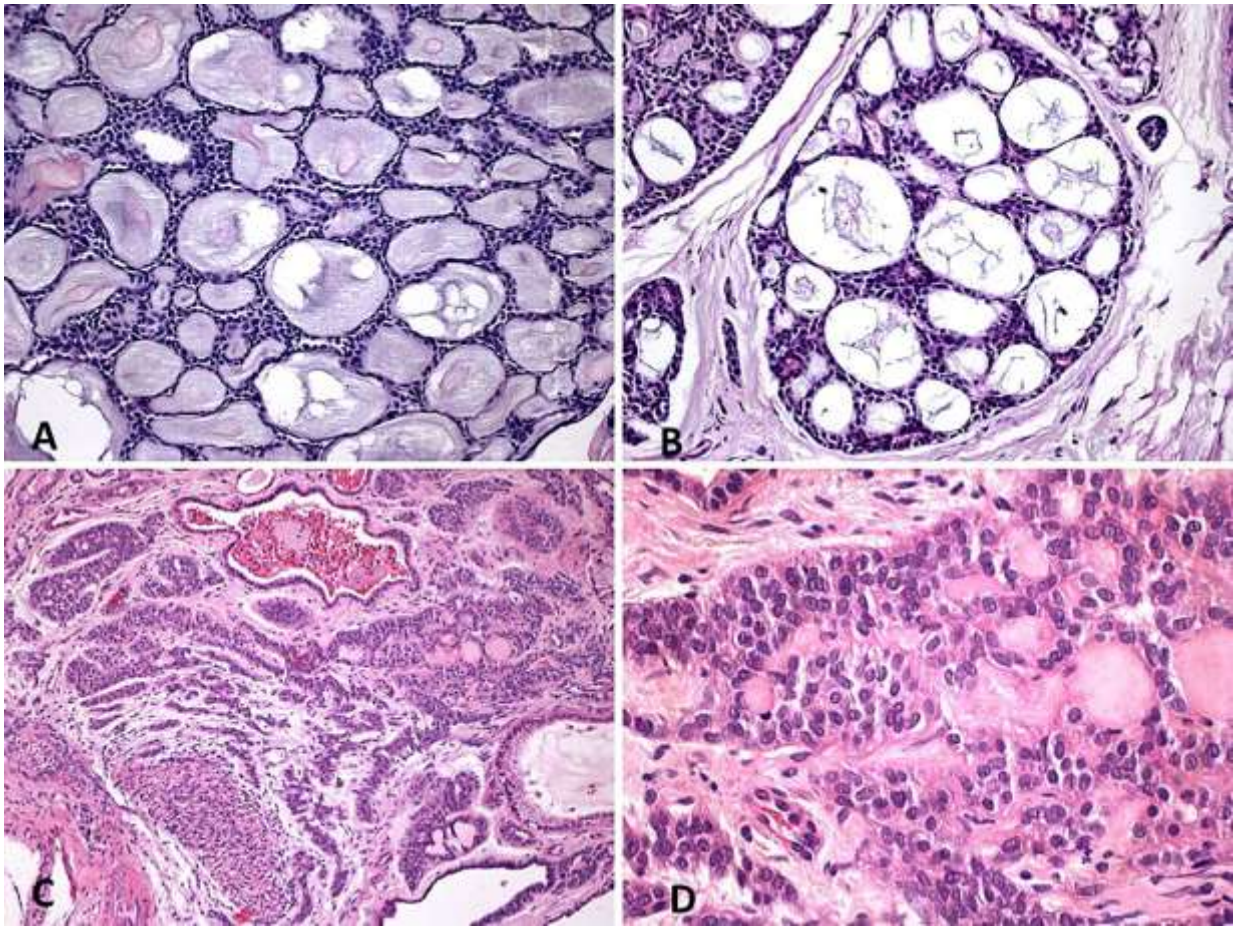

Figure 2. Microscopic aspect of the AdCC and PAc cases used in this study. **A)** Characteristic cribriform growth pattern of AdCC (H&E; 100X). **B)** In the cribriform structures of AdCC it was possible to observe more eosinophilic luminal epithelial cells forming small ductal structures (H&E; 100X). **C)** PAc case containing neoplastic cells that exhibited foci of neural involvement (H&E; 100X). **D)** Under higher-magnification it is possible to observe the pale staining and the blend appearance of PAc cells (H&E; 200X).

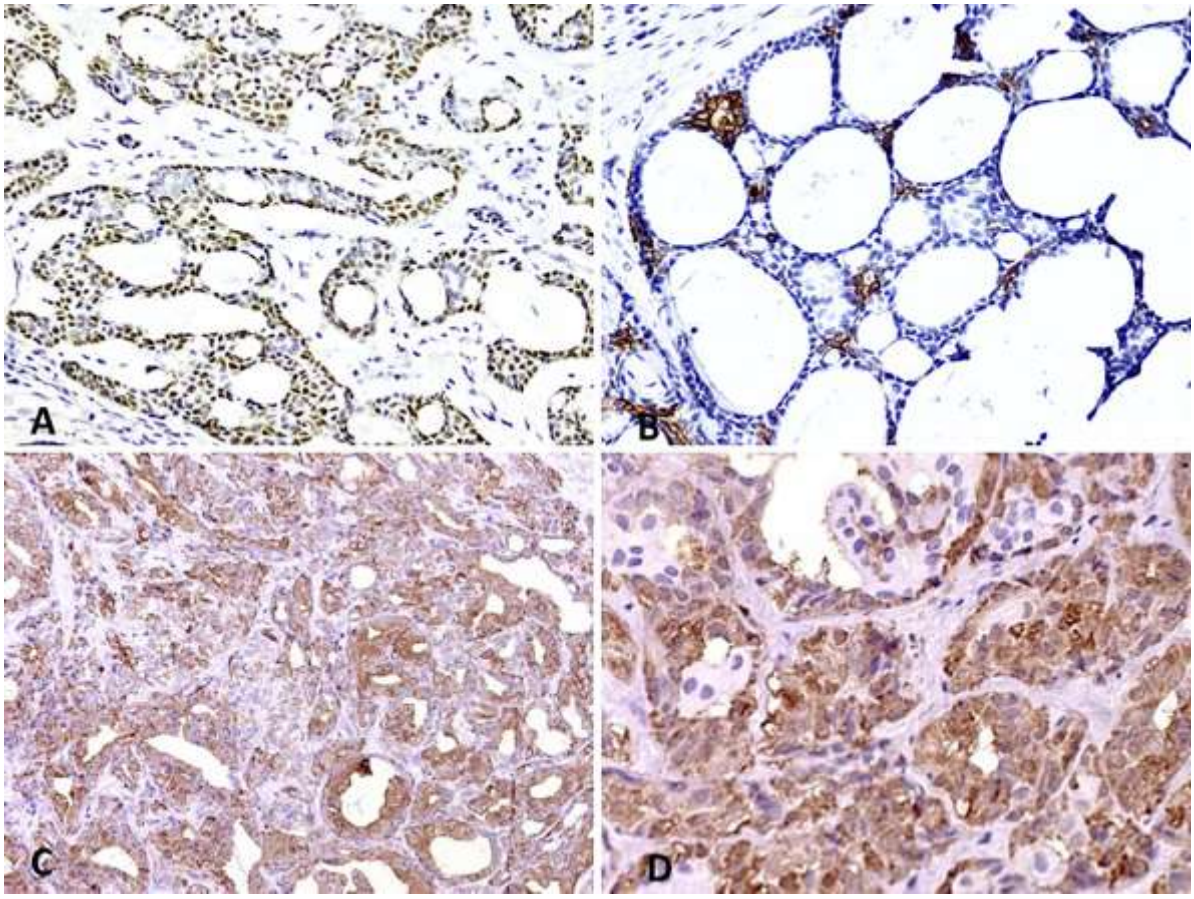

Figure 3. Immunohistochemical reactions used to better characterize the cellular components of AdCC and PAc cases. A**)** p63 strong nuclear positivity in the myoepithelial cells of AdCC (DAB; 200X). B) CK7 stained the luminal cells of small ductal structures observed in AdCC cases (DAB; 200X). C) In PAc CK7 revealed a more diffuse positive staining (DAB; 200X). D) Similarly, PAc cases also demonstrated a strong and diffuse positivity for S100 (DAB; 200X).

Association of laser microdissection (LMD) and LC-MS/MS is able to identify the proteome of AdCC and PAc

LMD approach was able to individualize the tumor regions from AdCC and PAc, thus separating the neoplastic cells from adjacent normal tissue.

LC-MS/MS analysis identified a total of 1,957 proteins **(Supplementary Tables 1 - 3)**. In Venn diagram, it is possible to verify the number of shared proteins between AdCC and PAc, as well as the proteins found exclusively in each tumor **(Figure 4a)**. A total of 1,590 proteins were shared between the tumors studied, with 261 proteins

identified only in the AdCC cases **(Supplementary Table 4)** and 106 proteins only in PAc **(Supplementary Table 5)**.

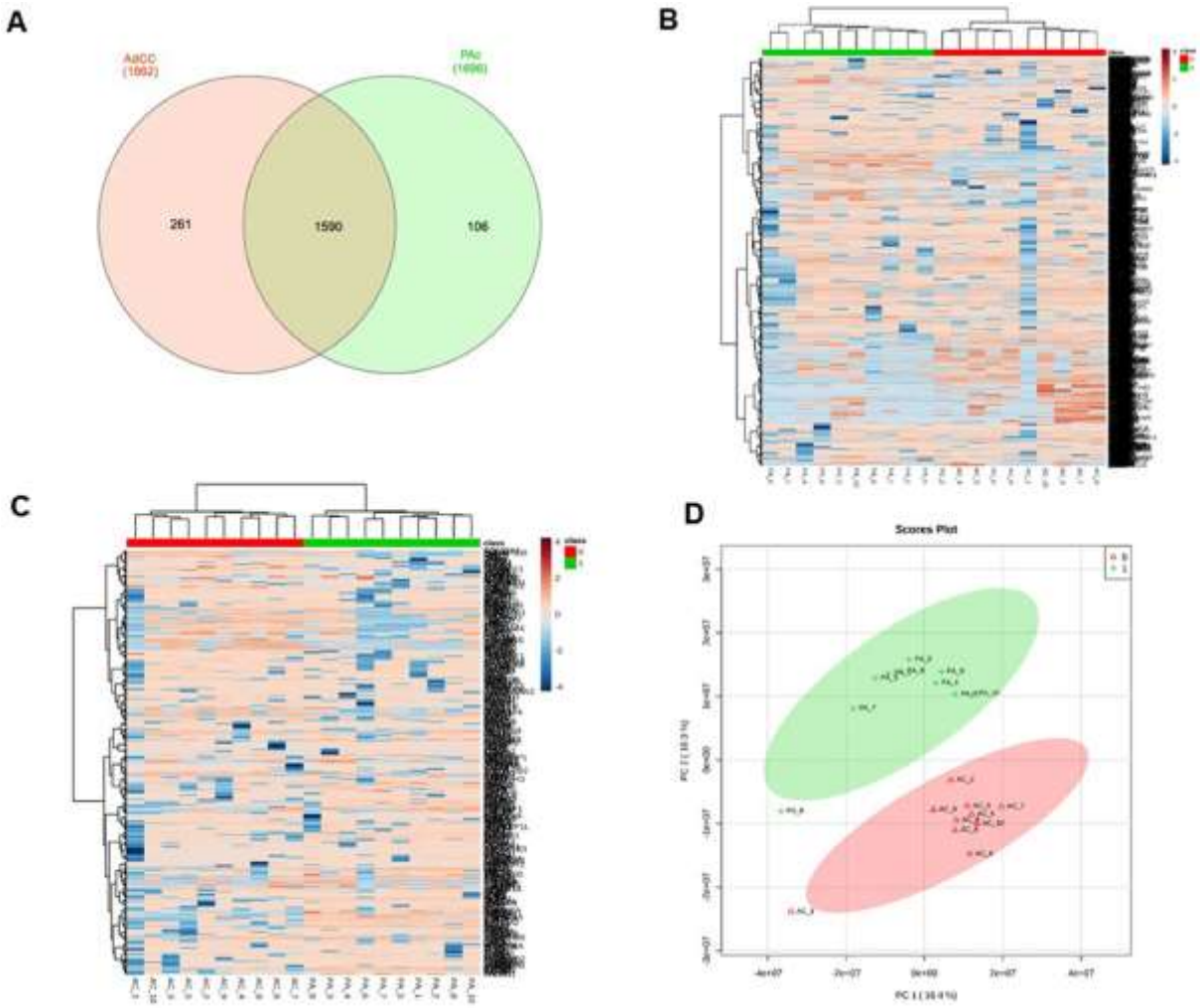

Figure 4. Quantitative proteome analysis differentiates adenoid cystic carcinoma from polymorphous adenocarcinoma. **A)** Venn diagram of common and exclusive proteins identified for adenoid cystic carcinoma and polymorphous adenocarcinoma. **B)** Analysis of clustering of the 1957 proteins identified in both tumors. The values for each protein (lines) and for each microdissected sample (columns) are colored based on the abundance of protein, in which high (red) and low (blue) values (using the Z-score LFQ values) are indicated in the scale color bar at the bottom. The hierarchical grouping was performed in the MetaboAnalyst 3.0 software using Pearson's correlation for data. **C)** Heat map of the differentially expressed proteins between adenoid cystic carcinoma and polymorphous adenocarcinoma. Proteins identified in tumors were hierarchically clustered by using the Z-score LFQ values with MetaboAnalyst 3.0 software. This cluster analysis revealed significant differences between adenoid cystic carcinoma and polymorphous adenocarcinoma. **D)** Score-plot of the PC analysis performed also in Metaboanalyst 3.0 showed the variance in the data among the two components that were derived from this analysis, with clear separation. PC 1 present 16.4% of the variance, PC 2 present 10.3% of the variance.

LC-MS/MS analysis reveals differentially abundant proteins between AdCC and PAc

LC-MS/MS analysis revealed differentially expressed proteins between AdCC and PAc. Student's t-test analyses performed with the Perseus software identified differentially expressed proteins ($p \leq 0.05$) between these two malignant SGT.

Of the 1,957 total proteins identified, 394 proteins showed a statistically significant difference in their expression between AdCC and PAc **(Supplementary Table 6)**. Clustering analyses first using all proteins identified and then using statistically significant proteins showed a clear separation of the tumors, with distinct protein expression patterns **(Figure 4b and 4c)**. Moreover, PC analysis was performed to transform and cluster AdCC and PAc global datasets in unsupervised fashion. PC 1 and PC 2 presented 16.4% and 10.3% of variance, respectively, demonstrating a clear data segregation. Therefore, PC analysis confirmed the distinct proteomes exhibited by AdCC and PAc (**Figure 4d**).

Proteins with statistically significant difference in expression were then correlated with biological processes in an enriched analysis. Thus, 243 proteins showed statistically significant association with 80 biological processes ($p \leq 0.05$) (**Supplementary Table 7**). The top 10 enriched biological processes were associated with proteins that plays a role in tumor progression such as small molecule metabolic process, extracellular matrix organization, extracellular matrix disassembly, mRNA splicing, via spliceosome, RNA splicing, osteoblast differentiation, immune response, cell adhesion and gene expression (**Figure 5**).

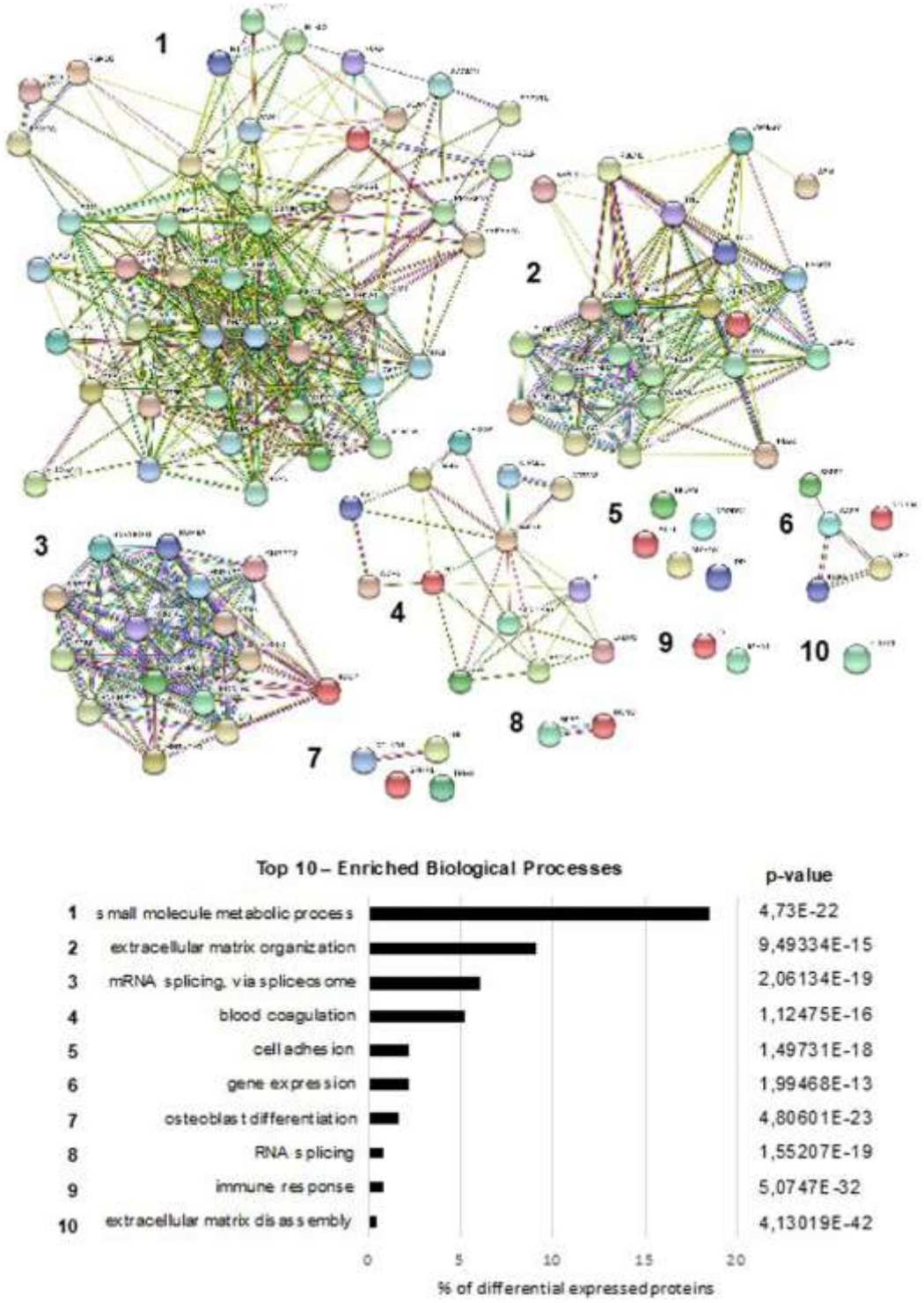

Figure 5. Top 10 enriched biological processes. The enrichment analysis for the GO (Gene Ontology) database identified 80 significant biological processes in adenoid cystic carcinoma and polymorphous adenocarcinoma, involving 243 differentially expressed proteins (Supplementary table 7). A schematic representation of the top 10 enriched biological processes shared by both tumors is showed, with protein-interaction networks constructed in the STRING database.

Table 2. Upregulated and downregulated proteins identified in AdCC and PAc of salivary glands (minimun fold change of 10).

Protein ID	UniProt ID	Protein name	Top enriched BP	Ratio in AcCC*	Ratio in PAc**	Expression	*p*-values
COL7A1	Q02388	Collagen alpha-1(VII) chain	Extracellular matrix organization	143,9470886	0,006946997	Upregulated in AdCC	0,00046386
HAPLN1	P10915	Hyaluronan and proteoglycan link protein 1	Extracellular matrix organization	56,40366049	0,017729346	Upregulated in AdCC	2,3543E-06
ACTA2	P62736	Actin, aortic smooth muscle	Muscle contraction	24,66728711	0,040539521	Upregulated in AdCC	0,02137686
TUBA1A	Q71U36	Tubulin alpha-1A chain	Mitotic cell cycle	15,22867117	0,065665611	Upregulated in AdCC	0,04970298
FBLN1	P23142	Fibulin-1	Extracellular matrix organization	15,17372016	0,065903417	Upregulated in AdCC	0,00180606
PKP1	Q13835	Plakophilin-1	Cell adhesion	12,59913473	0,07937053	Upregulated in AdCC	0,04949547
NCAN	O14594	Neurocan core protein	Small molecule metabolic process	0,010056395	99,43921476	Upregulated in PAc	0,00268849
ALDH1A1	P00352	Retinal dehydrogenase 1	Small molecule metabolic process	0,016975038	58,9100287	Upregulated in PAc	1,7472E-07
S100A1	Q5T7Y6	Protein S100-A1	Regulation of heart contraction	0,029050413	34,42291807	Upregulated in PAc	0,04659331
SLC12A2	P55011	Solute carrier family 12 member 2	Transmembrane transport	0,031889967	31,35782518	Upregulated in PAc	8,4007E-07
PRKAR1A	P10644	cAMP-dependent protein kinase type I-alpha regulatory subunit	Small molecule metabolic process	0,057774688	17,30861806	Upregulated in PAc	2,9863E-07
ATP1B1	P05026	Sodium/potassium-transporting ATPase subunit beta-1	Blood coagulation	0,062138003	16,09321107	Upregulated in PAc	0,01256684
TNC	P24821	Tenascin	Extracellular matrix organization	0,063413384	15,76954162	Upregulated in PAc	4,0042E-07
COPS7A	Q9UBW8	COP9 signalosome complex subunit 7a	Cullin deneddylation	0,069808625	14,32487744	Upregulated in PAc	6,2297E-05
SYNM	O15061	Synemin	Intermediate filament cytoskeleton organization	0,072650241	13,76457924	Upregulated in PAc	9,4936E-07
NDRG1	Q92597	NDRG1	Cell death	0,086140994	11,60887469	Upregulated in PAc	4,3605E-06

BP: Biological processes. AdCC: Adenoid cystic carcinoma. Pac: Polymorphous adenocarcinoma. Values represent the fold of expression in adenoid cystic carcinoma in relation to polymorphous adenocarcinoma and in polymorphous adenocarcinoma in relation to adenoid cystic carcinoma. *Upregulated in adenoid cystic carcinoma and downregulated in polymorphous adenocarcinoma. **Upregulated in polymorphous adenocarcinoma and downregulated in adenoid cystic carcinoma.

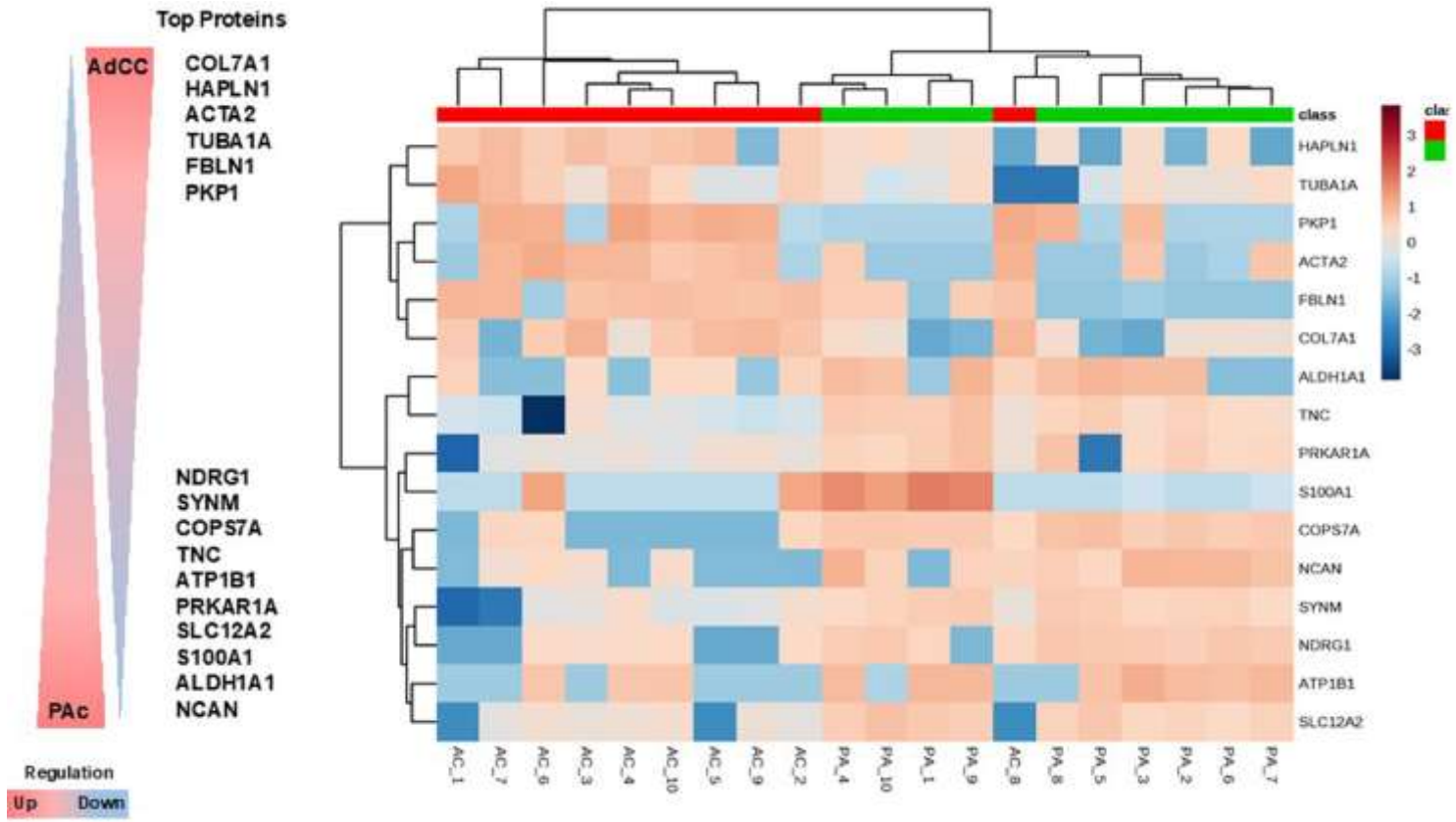

Figure 6. Reverse regulation of the signature proteins differentially expressed between adenoid cystic carcinoma (AdCC) and polymorphous adenocarcinoma (PAc). Heat map of the 16 top differentially expressed proteins between AdCC and PAc. Signature proteins identified in tumors were hierarchically clustered by using LFQ values with MetaboAnalyst 3.0 software. This cluster analysis revealed a reverse regulation of the proteins, likewise represented in the schematic figure: while COL7A1, HAPLN1, ACTA2, TUBA1A, FBLN1 e PKP1 were upregulated in AdCC, they were also downregulated in PAc. In the same way, while NCAN, ALDH1A1, S100A1, SLC12A2, PRKAR1A, ATP1B1, TNC, COPS7A, SYNM and NDRG1 were upregulated in PAc, they were also downregulated in AdCC.

Proteome annotation indicated signature proteins and distinct protein expression to differentiate AdCC from PAc

We combined the data from Student's t-test and enriched biological process with the expression of proteins in each tumor. Based on previous publications, it was considered a fold-change higher than 10 times in one group than in the other as a cut-off value in order to decrease the number of candidate proteins and to create a protein signature with a higher potential of diagnostic determination[30-32]. Through this combination of data, we arrived at a signature of 16 proteins, 6 upregulated in AdCC and 10 in PAc **(Table 2)**. To confirm the significant difference in the expression of these 16 proteins between both tumors, a new clustering analysis was done, showing separation of the groups and a trend towards a reverse regulation, for instance, COL7A1 was upregulated in AdCC and downregulated in PAc. This cluster analysis with the 16 signature proteins revealed clear differences between AdCC and PAc **(Figure 6)**.

Discussion

Diagnosis of SGT frequently represents a challenge even for experienced diagnosticians given their highly heterogeneous microscopic aspects and similar clinical manifestations. Moreover, the lack of reliable diagnostic markers also contributes to make the recognition of these tumors more difficult. Therefore, the use of strategies that would lead to the discovery of new molecules that could contribute to improve the diagnosis of these neoplasms is mandatory. In this study, we performed a comparative proteome analysis to examine the global differences in the protein expression patterns of AdCC and PAc, and our results demonstrated that mass spectrometry approach was useful to obtain a proteomic signature that significantly differentiated both tumors.

AdCC and PAc are two frequent malignant SGT that usually demonstrate clinical and microscopic overlaps[7]. These neoplasms usually manifest as slow growing swellings that more commonly affect parotid glands and the minor salivary glands of the oral cavity, respectively[8,33]. Pain may be reported, whereas ulcers are only occasionally observed. Microscopically, in both tumors neural invasion and the cribriform growth pattern are common findings[33,34]. Although some histologic features and the use of some non-specific immunohistochemical markers may contribute to differentiate these malignancies, this may be exceedingly difficult in some cases, especially in those with limited tissue availability, which represents an important issue given the worse prognosis carried by AdCC, frequently associated with late recurrences and distant metastases[8,33,35].

Although recent studies have demonstrated that these tumors are characterized by the presence of different and specific genetic events like the translocation involving the genes Myb and NFIB on chromosomes 6 and 9 in AdCC and mutations in the PRKD1 gene in PAc, these molecular imbalances are not always present and the access to laboratorial facilities necessary to identify these mutations for routine microscopic diagnosis is still limited[6]. The use of proteomics to find new molecular markers not only with diagnostic, but also with prognostic potential, as well as those involved with the pathogenesis of different cancerous and non-cancerous diseases that could later be identified by more accessible laboratorial techniques like immunohistochemistry, has been described by different groups[36,37].

Using mass spectrometry, Flores et al[38] identified 388 proteins exclusively expressed in oral squamous cell carcinoma (OSCC) when compared to normal oral mucosa, subsequently demonstrating that one of these proteins, EEF1D, would play an important role in the pathogenesis of this aggressive cancer. Similarly, Kawahara et al[10]

used targeted proteomics to measure in saliva a panel of biomarker candidates for increased risk of OSCC and observed that CFB, C3, C4B, SERPINA1 and LRG1 were associated with a higher risk of developing this malignancy. Johnston et al[37] using normal B-lymphocytes, recently showed the ability of proteomics to identify previously unknown potential molecular targets for chronic lymphocytic leukemia, as well as a protein expression signature for this neoplasm. These results are in accordance to Giusti et al[17] that using fine-needle aspiration fluids, found 17 differentially expressed proteins between two variants of papillary thyroid cancers.

In the context of SGT, Donadio et al[18] investigated the proteomic profile of pleomorphic adenoma and Warthin tumor. The authors demonstrated a total of 26 differentially expressed proteins, further validating 9 of them as potentially diagnostic markers. To date, this was the only study investigating the proteomic profile of SGT available in literature. There seems to be no previous attempts to determine the proteome of AdCC and PAc of salivary glands, and our results corroborate the findings of Donadio et al[18] demonstrating that it is possible to obtain a global protein profile able to differentiate both tumors and to obtain a protein signature of 16 molecules with a significantly different expression pattern between AdCC and PAc, indicating that the use of this combination of differentially expressed proteins could be an useful diagnostic auxiliary for those cases where routine histology is not sufficient.

Furthermore, we observed that this protein signature of 16 biomarkers demonstrated a trend towards a reverse regulation between both tumors. For instance, the collagen type VII α 1 chain (COL7A1) protein exhibited an expression 143 times higher in AdCC than in PAc. COL7A1 is a basement membrane protein responsible to form anchoring fibrils that may contribute to the organization and adhesion of the epithelial basement membrane, interacting with extracellular matrix proteins, and was

associated with esophageal squamous cell carcinoma development. In a recent study, Tao et al[39] investigated the single fused RNA sequence of COL7A1 and urocortin 2 (UCN2) in laryngeal cancer and found that this chimera COL7A1-UCN2 affected cancer stem cell transition, promoted epithelial-mesenchymal transition, and resulted in poorer prognosis.

On the other hand, NCAN (neurocan core protein) protein had an expression 99 times higher in PAc than in AdCC. NCAN is a chondroitin sulfate proteoglycan that modulates neuronal adhesion and has recently been associated with neuroblastoma progress. Su et al[40] studied the role of NCAN in neuroblastomas, observing that this protein was highly upregulated in malignant cells, promoting malignant phenotype aquisition and cell division. To date, no study investigated COL71A and NCAN proteins expression in AdCC and PAc.

Also important to observe that S100A1 protein was shown to be upregulated in PAc cases if compared to AdCC, which would be in accordance to the observed in immunohistochemical analyses and already reported in literature[41], since S100 positivity in AdCC is restricted to its myoepithelial component and diffusely present in PAc tumor cells. On the other hand, our proteomics results refer more specifically to the S100A1 subtype, whose distribution in AdCC and PAc remains to be investigated.

Although laser microdissection does not avoid that some occasional stromal tissue was included in the samples submitted to spectrometry analysis, we attempted in this study to enrich as much as possible our samples for neoplastic cells only; therefore, we assume that any diagnostic potential of the 16 biomarkers present in the protein signature obtained may rely in their expression in the parenchymal component of the neoplasms, not in their stromal compartment, which deserves to be evaluated in future investigations. In addition, an important component of a high throughput study like

proteomics analysis is the follow-up validations, also because we used a small convenience sample of 10 cases of each tumor to determine their proteomic profile; therefore, we understand that the diagnostic potential of the proteins exclusively expressed by each tumor and those differentially expressed in the neoplasms described in our study must be validated in a larger independent sample of AdCC and PAc cases using other laboratorial methodologies like immunohistochemistry and western blot, but we also believe that these preliminary results presented here may be useful for other groups to conduct their own analyses on the expression of different proteins in these two neoplasms.

In conclusion, our study demonstrates that global proteomics may be useful to discriminate two malignant SGT that frequently show very similar clinical and microscopic features, although additional validation studies are still necessary to confirm and to determine the diagnostic potential of the protein signature obtained here.

Acknowledgments

The authors acknowledge that Dr. Felipe Paiva Fonseca received a post-doctoral schorlarship from the São Paulo State Research Foundation (FAPESP Process# 2015/16056-0). The authors also acknowledge the mass spectrometry group at Brazilian Biosciences National Laboratory (LNBio), CNPEM, Campinas, Brazil for their support with the use of NanoLC PROXEON EASY-nLC II coupled with LTQ Orbitrap Velos ETD mass spectrometer, Thermo Scientific (Financial Support: FAPESP multi-user Project 2009/54067-3).

References

1 Fonseca FP, Carvalho MV, de Almeida OP, Range ALCA, Takizawa MC, Bueno AG, et al. Clinicopathologic analysis of 493 cases of salivary gland tumors in a Southern Brazilian population. Oral Surg Oral Med Oral Pathol Oral Radiol 2012; 114: 230-239. doi: 10.1016/j.oooo.2012.04.008.

2 Coca-Pelaz A, Rodrigo JP, Bradley PJ, Vander Poorten V, Triantafyllou A, Hunt JL, et al. Adenoid cystic carcinoma of the head and neck - An update. Oral Oncol 2015; 51: 652-661. doi: 10.1016/j.oraloncology.2015.04.005.

3 Agarwal JP, Jain S, Gupta T, Tiwari M, Laskar SG, Dinshaw KA, et al. Intraoral adenoid cystic carcinoma: prognostic factors and outcome. Oral Oncol 2008; 44: 986-993. doi: 10.1016/j.oraloncology.2008.01.004.

4 Persson M, Andrén Y, Mark J, Horlings HM, Persson F, Stenman G. Recurrent fusion of MYB and NFIB transcription factor genes in carcinomas of the breast and head and neck. Proc Natl Acad Sci U S A 2009; 106: 18740-18744. doi: 10.1073/pnas.0909114106.

5 Almeida-Pinto YD, Costa SFS, Andrade BAB, Altemani A, Vargas PA, Abreu LG, et al. t(6;9)(MYB-NFIB) in head and neck adenoid cystic carcinoma: a systematic review with meta-analysis. Oral Dis. 2019. doi: 10.1111/odi.12984.

6 Fonseca FP, Sena Filho M, Altemani A, Speight PM, Vargas PA. Molecular signature of salivary gland tumors: potential use as diagnostic and prognostic marker. J Oral Pathol Med 2016; 45: 101-110. doi: 10.1111/jop.12329.

7 Fonseca FP, Brierley D, Wright JM, Santos-Silva AR, Almeida OP, Rocha AC, et al. Polymorphous low-grade adenocarcinoma of the upper lip: 11 cases of an uncommon diagnosis. Oral Surg Oral Med Oral Pathol Oral Radiol 2015; 119: 566-571. doi: 10.1016/j.oooo.2015.01.001.

8 Elhakim MT, Breinholt H, Godballe C, Andersen LJ, Primdahl H, Kristensen CA, et al. Polymorphous low-grade adenocarcinoma: A Danish national study. Oral Oncol 2016; 55: 6-10. doi: 10.1016/j.oraloncology.2016.02.005.

9 Weinreb I, Piscuoglio S, Martelotto LG, Waggott D, Ng CK, Perez-Ordonez B, et al. Hotspot activating PRKD1 somatic mutations in polymorphous low-grade adenocarcinomas of the salivary glands. Nat Genet 2014; 46: 1166-1169. doi: 10.1038/ng.3096.

10 Kawahara R, Bollinger JG, Rivera C, Ribeiro AC, Brandão TB, Paes Leme AF, et al. A targeted proteomic strategy for the measurement of oral cancer candidate biomarkers in human saliva. Proteomics 2016; 16: 159-173. doi: 10.1002/pmic.201500224.

11 Mueller C, Haymond A, Davis JB, Williams A, Espina V. Protein biomarkers for subtyping breast cancer and implications for future research. Expert Rev Proteomics 2018; 15: 131-152. doi: 10.1080/14789450.2018.1421071.

12 Kawahara R, Meirelles GV, Heberle H, Domingues RR, Granato DC, Yokoo S, et al. Integrative analysis to select cancer candidate biomarkers to targeted validation. Oncotarget 2015; 6: 43635-43652. doi: 10.18632/oncotarget.6018.

13 Rodríguez-Cerdeira C, Molares-Vila A, Carnero-Gregorio M, Corbalan-Rivas A. Recent advances in melanoma research via "omics" platforms. J Proteomics 2018; 188: 152-166. doi: 10.1016/j.jprot.2017.11.005.

14 Shoshan-Barmatz V, Bishitz Y, Paul A, Krelin Y, Nakdimon I, Peled N, et al. A molecular signature of lung cancer: potential biomarkers for adenocarcinoma and squamous cell carcinoma. Oncotarget 2017; 8: 105492-105509. doi: 10.18632/oncotarget.22298.

15 Park J, Lee E, Park KJ, Park HD, Kim JW, Woo HI, et al. Large-scale clinical validation of biomarkers for pancreatic cancer using a mass spectrometry-based proteomics approach. Oncotarget 2017; 8: 42761-42771. doi: 10.18632/oncotarget.17463.

16 Hu HF, Xu WW, Wang Y, Zheng CC, Zhang WX, Li B, et al. Comparative proteomics analysis identifies Cdc42-Cdc42BPA signaling as prognostic biomarker and therapeutic target for colon cancer invasion. J Proteome Res 2018; 17: 265-275. doi: 10.1021/acs.jproteome.7b00550.

17 Giusti L, Iacconi P, Ciregia F, Giannaccini G, Donatini GL, Basolo F, et al. Fine-needle aspiration of thyroid nodules: proteomic analysis to identify cancer biomarkers. J Proteome Res 2008; 7: 4079-4088. doi: 10.1021/pr8000404.

18 Donadio E, Giusti L, Seccia V, Ciregia F, da Valle Y, Dallan I, et al. New insight into benign tumours of major salivary glands by proteomic approach 2013; 8: e71874. doi: 10.1371/journal.pone.0071874.

19 Ciregia F, Giusti L, Molinaro A, Niccolai F, Mazzoni MR, Rago T, et al. Proteomic analysis of fine-needle aspiration in differential diagnosis of thyroid nodules. Transl Res 2016; 176: 81-94. doi: 10.1016/j.trsl.2016.04.004.

20 Bell D, Bullerdiek J, Gnepp DR, et al. Salivary gland tumours. In: El-Naggar AK, Chan JKC, Grandis JR, Takata T, Slootweg PJ, editors. World Health Organization classification of head and neck tumours. Lyon: IARC; 2017. p. 33.

21 Ostasiewicz P, Zielinska DF, Mann M, Wisniewski JR. Proteome, phosphoproteome, and n-glycoproteome are quantitatively preserved in formalin-fixed paraffin-embedded tissue and analyzable by high-resolution mass spectrometry. J Proteome Res 2010; 9: 3688-3700. doi: 10.1021/pr100234w.

22 Villén J, Gygi SP. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. Nat Protoc 2008; 3: 1630-1638. doi: 10.1038/nprot.2008.150.

23 Carnielli CM, Macedo CCS, De Rossi T, Granato DC, Rivera C, Domingues RR, et al. Combining discovery and targeted proteomics reveals a prognostic signature in oral cancer. Nat Commun; 2018: 3598. doi: 10.1038/s41467-018-05696-2.

24 Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 2008; 26: 1367-1372. doi: 10.1038/nbt.1511.

25 Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc 2007; 2: 1896-1906. doi: 10.1038/nprot.2007.261.

26 Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res 2011; 10: 1794-1805. doi: 10.1021/pr101065j.

27 Carazzolle MF, Carvalho LM, Slepicka HH, Vidal RO, Peeira GA, Kobarg J, et al. IIS-Integrated Interactome System: a web-based platform for the annotation, analysis and visualization of protein–metabolite–gene–drug interactions by integrating a variety of data sources and tools. PLoS One 2014; 9: e100385. doi: 10.1371/journal.pone.0100385.

28 Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Chery JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000; 25: 25-29. doi: 10.1038/75556

29 Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res 2017; 45: D362-D368. doi: 10.1093/nar/gkw937.

30 Molloy MP, Donohoe S, Brzezinski EE, Kilby GW, Stevenson TI, Baker JD, et al. Large-scale evaluation of quantitative reproducibility and proteome coverage using acid cleavable isotope coded affinity tag mass spectrometry for proteomic profiling. Proteomics 2005; 5: 1204-1208. doi: 10.1002/pmic.200400994

31 Gygi SP, Rist B, Griffin TJ, Eng J, Aebersold R. Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags. J Proteome Res 2002; 1: 47-54. doi: 10.1021/pr015509n.

32 Griffin TJ, Lock CM, Li XJ, Patel A, Chervetsova I, Lee H, et al. Abundance ratio-dependent proteomic analysis by mass spectrometry. Anal Chem 2003; 75: 867-874. doi: 10.1021/ac026127j.

33 Zhang CY, Xia RH, Han J, Wang BS, Tian WD, Zhong LP, et al. Adenoid cystic carcinoma of the head and neck: Clinicopathologic analysis of 218 cases in a Chinese population. Oral Surg Oral Med Oral Pathol Oral Radiol 2013; 115: 368-375. doi: 10.1016/j.oooo.2012.11.018.

34 Ju J, Li Y, Chai J, Ma C, Ni Q, Shen Z, et al. The role of perineural invasion on head and neck adenoid cystic carcinoma prognosis: a systematic review and meta-analysis. Oral Surg Oral Med Oral Pathol Oral Radiol 2016; 122: 691-701. doi: 10.1016/j.oooo.2016.08.008.

35 van Weert S, van der Waal I, Witte BI, Leemans CR, Bloemena E. Histopathological grading of adenoid cystic carcinoma of the head and neck:

analysis of currently used grading systems and proposal for a simplified grading scheme. Oral Oncol 2015; 51: 71-76. doi: 10.1016/j.oraloncology.2014.10.007.

36 Li JQ, Xu BJ, Shakhtour B, Deane N, Merchant N, Heslin MJ, et al. Variability of in situ proteomic profiling and implications for study design in colorectal tumors. Int J Oncol 2007; 31: 103-111. doi: 10.3892/ijo.31.1.103.

37 Johnston HE, Carter MJ, Larrayoz M, Clarke J, Garbis SD, Oscier D, et al. Proteomics profiling of CLL versus healthy B-cells identifies putative therapeutic targets and a subtype-independent signature of spliceosome dysregulation. Mol Cell Proteomics 2018; 17: 776-791. doi: 10.1074/mcp.RA117.000539.

38 Flores IL, Kawahara R, Miguel MCC, Granato DC, Domingues RR, Macedo CC, et al. EEF1D modulates proliferation and epithelial–mesenchymal transition in oral squamous cell carcinoma. Clin Sci 2016; 130: 785-799. doi: 10.1042/CS20150646.

39 Tao Y, Gross N, Fan X, Yang J, Teng M, Li X, et al. Identification of novel enriched recurrent chimeric COL7A1-UCN2 in human laryngeal cancer samples using deep sequencing. BMC Cancer 2018; 18: 248. doi: 10.1186/s12885-018-4161-8.

40 Su Z, Kishida S, Tsubota S, Sakamoto K, Cao D, Kiyonari S, et al. Neurocan, an extracellular chondroitin sulfate proteoglycan, stimulates neuroblastoma cells to promote malignant phenotypes. Oncotarget 2017; 8: 106296-106310. doi: 10.18632/oncotarget.22435.

41 Rooper LM. Challenges in minor salivary gland biopsies: a practical approach to problematic histologic patterns. Head Neck Pathol 2019. doi: 10.1007/s12105-019-01010-8