Article

Peptidomics Unveils Distinct Acetylation Patterns of Histone and Annexin A1 in Differentiated Thyroid Cancer

Margarida Coelho [1,2,3,4], João Capela [5], Vera M. Mendes [1,2,3], João Pacheco [5], Margarida Sá Fernandes [5], Isabel Amendoeira [5,6,7], John G. Jones [1,2,3], Luís Raposo [8,9] and Bruno Manadas [1,2,3,*]

1   CNC—Center for Neurosciences and Cell Biology, University of Coimbra, 3004-504 Coimbra, Portugal; mmcoelho@cnc.uc.pt (M.C.)
2   CIBB—Centre for Innovative Biomedicine and Biotechnology, University of Coimbra, 3004-504 Coimbra, Portugal
3   III Institute for Interdisciplinary Research (IIIUC), University of Coimbra, 3030-789 Coimbra, Portugal
4   Department of Chemistry, Faculty of Sciences and Technology, University of Coimbra, 3004-535 Coimbra, Portugal
5   Centro Hospitalar Universitário São João, 4200-319 Porto, Portugal
6   I3S, Instituto de Investigação e Inovação em Saúde, 4200-135 Porto, Portugal
7   Ipatimup, Instituto de Patologia e Imunologia Molecular da Universidade do Porto, 4200-465 Porto, Portugal
8   Portuguese Society of Endocrinology, Diabetes and Metabolism, 1600-892 Lisbon, Portugal
9   EPIUnit, Institute of Public Health, University of Porto, 4050-600 Porto, Portugal
*   Correspondence: bmanadas@cnc.uc.pt

Abstract: Thyroid cancer is a common malignancy of the endocrine system. Nodules are routinely evaluated for malignancy risk by fine needle aspiration biopsy (FNAB), and in cases such as follicular lesions, differential diagnosis between benign and malignant nodules is highly uncertain. Therefore, the discovery of new biomarkers for this disease could be helpful in improving diagnostic accuracy. Thyroid nodule biopsies were subjected to a precipitation step with both the insoluble and supernatant fractions subjected to proteome and peptidome profiling. Proteomic analysis identified annexin A1 as a potential biomarker of thyroid cancer malignancy, with its levels increased in malignant samples. Also upregulated were the acetylated peptides of annexin A1, revealed by the peptidome analysis of the supernatant fraction. In addition, supernatant peptidomic analysis revealed a number of acetylated histone peptides that were significantly elevated in the malignant group, suggesting higher gene transcription activity in malignant tissue. Two of these peptides were found to be robust malignancy predictors, with an area under the receiver operating a characteristic curve (ROC AUC) above 0.95. Thus, this combination of proteomics and peptidomics analyses improved the detection of malignant lesions and also provided new evidence linking thyroid cancer development to heightened transcription activity. This study demonstrates the importance of peptidomic profiling in complementing traditional proteomics approaches.

Keywords: peptidomics; proteomics; tissue; LC-MS/MS; biomarker; thyroid cancer

1. Introduction

Thyroid nodules are extremely common, but the majority are benign, with a malignancy rate of only 1% [1]. Most thyroid malignancies are differentiated thyroid carcinomas [2] from follicular cells, either papillary (80–85% of cases) or follicular carcinomas (10–15% of cases). The standard diagnostic procedure is fine needle aspiration biopsy (FNAB) followed by a cytological analysis with diagnosis being made according to the categories of the Bethesda System for Reporting Thyroid Cytopathology. These six diagnostic categories recommend management strategies (e.g., molecular testing, repeat FNAB vs. surgery), and their implied risks of malignancy. The stated goals of this protocol are to inform conservative management of thyroid nodules unlikely to cause harm, but at the same time evaluate the probabilities of malignancy at each stage of analysis [3].

Considering that the initial FNAB cytological analysis is currently inefficient at resolving malignant from benign nodules—mainly in follicular patterned lesions—the development of complementary diagnostic methods that could more effectively discriminate between malignant and benign lesions would be useful.

Although thyroid cancer has been studied by proteomics [4,5], peptidomic approaches remain relatively unexplored, particularly in tissue [6]. Peptides can either originate from the breakdown of a protein or be intentionally expressed to play independent regulatory roles [7]. Peptidomics is the screening of such endogenous peptides, which among other things can potentially provide additional molecular insights into disease etiologies.

This work analysed thyroid nodules with both proteomics and peptidomics approaches using liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS). Alongside the conventional approach of identifying proteins whose levels were different between malignant and benign tissues, this approach also discovered naturally occurring peptides that, in addition to informing gene transcription processes in malignant cells, are also strong candidates for standalone biomarkers of malignancy.

2. Results

Thyroid samples were subjected to a protein precipitation procedure, resulting in a sediment composed mainly of proteins and a supernatant composed mainly of peptides and other small molecules (Figure 1a). A total of 2423 proteins were quantified in the proteomics screening, while 149 peptides belonging to 59 proteins were quantified from the supernatant and analysed in a peptidomic screening. The separation of malignant from benign samples observed in the PCAs of both proteomics and peptidomics screening reveals potential differences to be further explored in both datasets (Figure 1b,c). The top 20 statistically different proteins and peptides relative to the *p*-value are depicted in Table S2 and in a heatmap in Figure S1.

Of the 1532 proteins that were considered statistically different, annexin A1 was increased in the malignant group with a fold change of 5.4 and a *p*-value below 0.0001 and presented the highest area under the curve (AUC) in the ROC analysis with 0.992 (95% confidence interval (CI) 0.967–1), demonstrating the potential of this protein to distinguish benign from malignant thyroid tissues [8]. When searching if peptides of this protein were found in the supernatant fraction, three non-tryptic peptides were detected, all statistically different with similar trends as the protein found in the pellet (Figure 2). However, none of these peptides found in the supernatant under non-denaturing conditions was identified in the 14 peptides that were used to quantify this protein in the pellet, nor were they identified in any proteomics sample. Interestingly, these three peptides have a common amino acid sequence and an acetylation.
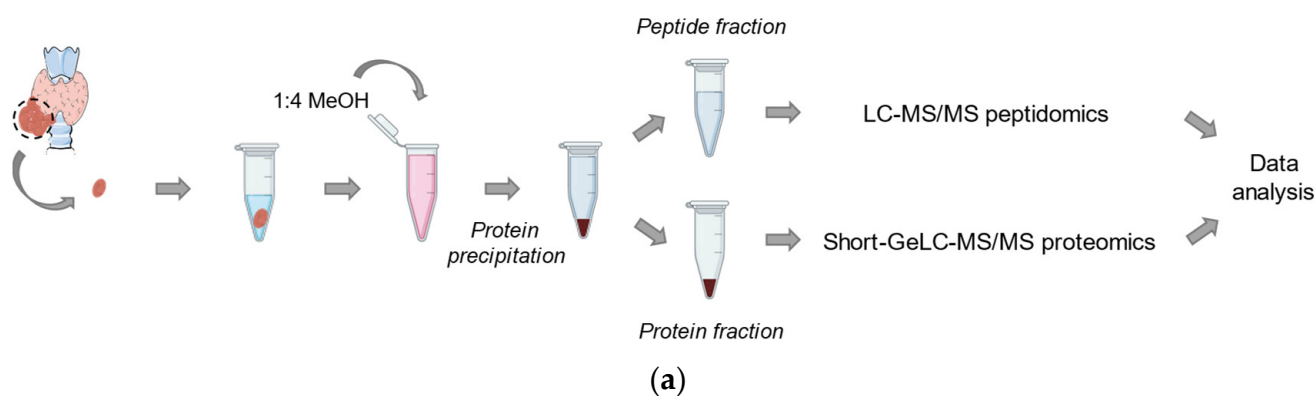

(a)

Figure 1. *Cont.*
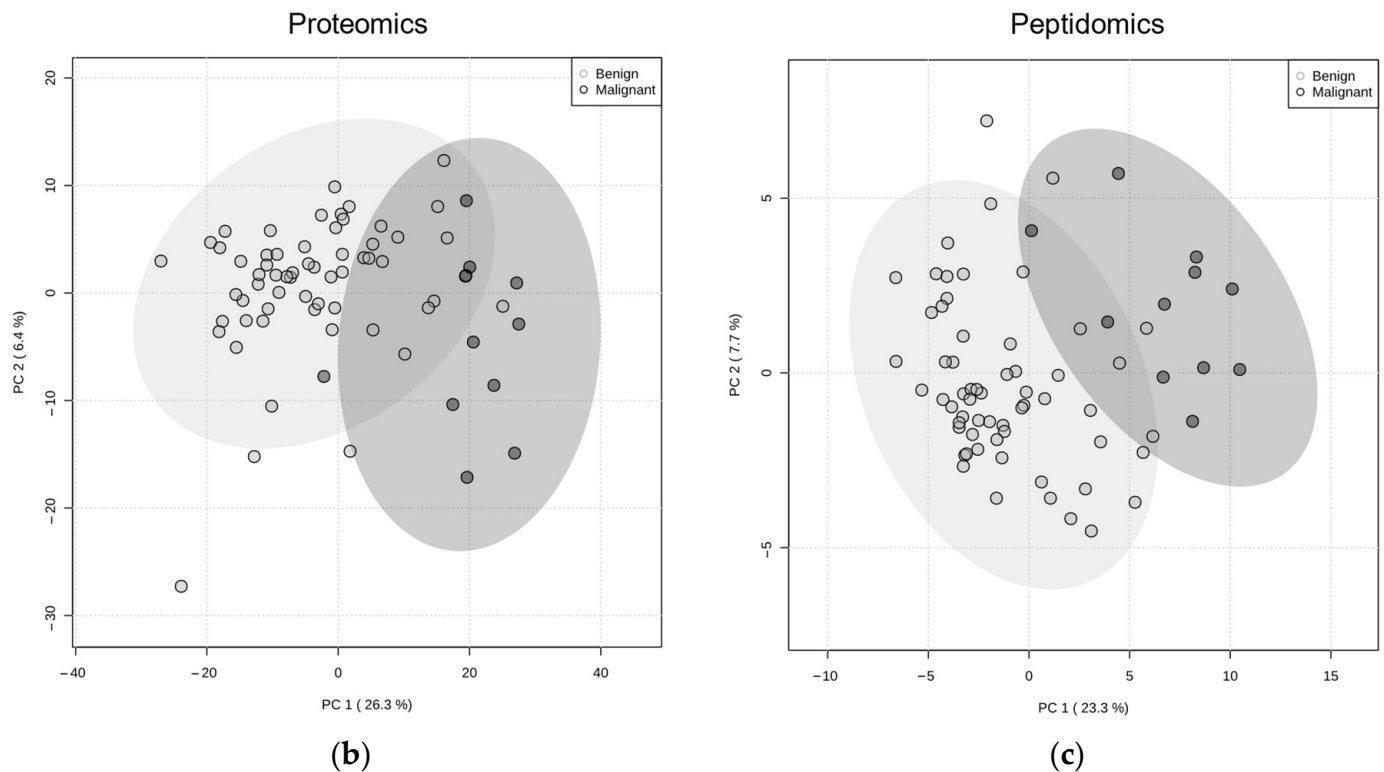
(b)

(c)

Figure 1. Workflow and principal component analysis (PCA) of benign and malignant samples in both the proteomics screening from the pellet fraction and the peptidomics screening of the supernatant: (**a**) schematic of sample preparation and analysis of the protein and peptide fractions; PCA scores plot computed using multivariate analysis of (**b**) proteomics data and (**c**) peptidomics data for benign (grey) and malignant (black) thyroid lesions.

In the peptidomics screening of the supernatant fraction, 75 peptides were considered statistically different. A total of 11 peptides from three histones appeared elevated in the malignant group compared to the benign (Figure 3). With the exception of one peptide, the other ten were acetylated. After individual ROC analysis, two peptides presented an AUC greater than 0.95: peptide [1Ac]-SETAPLAPTIPAPAEKTPVKK from histone H1.3 with an AUC of 0.967 (95% CI 0.901–1) and peptide SETAPAAPAAPAPAEKTPVK from histone H1.4 with an AUC of 0.956 (95% CI 0.889–1). Proteomics screenings depend highly on protein dynamic range. In our case, these histone peptides were not quantified in the proteomics analysis of the pellet fraction, although the histone H1.4 peptide has been identified in a few proteomics samples, emphasizing the importance of the peptidomics analysis of the supernatant.
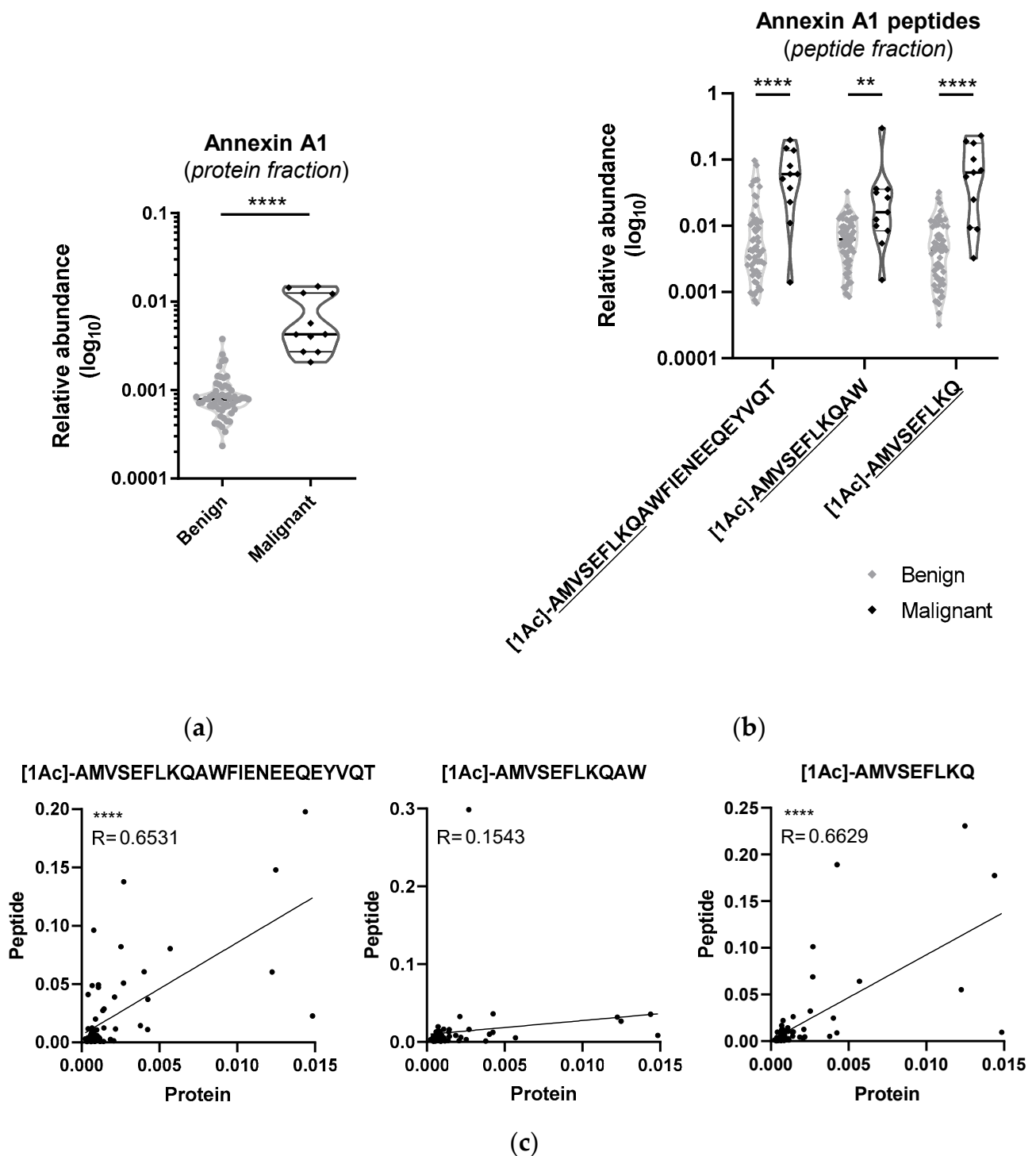
Figure 2. Annexin A1 peptides found in the supernatant were increased, similarly to the protein quantification obtained in the proteomics screening: violin plots with individual sample representation of (**a**) annexin A1 protein from the proteomics analysis and (**b**) peptides of annexin A1 from the peptidomics analysis of the peptide fraction of the same tissue samples. Peptide sequences are represented by amino acid one letter code and acetylation at the N-termini is represented by [1Ac]. Benign samples are represented in grey and malignant ones in black. ** $p \leq 0.01$, **** $p \leq 0.0001$ (Mann–Whitney test). (**c**) Correlations between annexin A1 protein from the proteomics analysis and each peptide from the peptidomics analysis. **** $p \leq 0.0001$ (Pearson correlation).
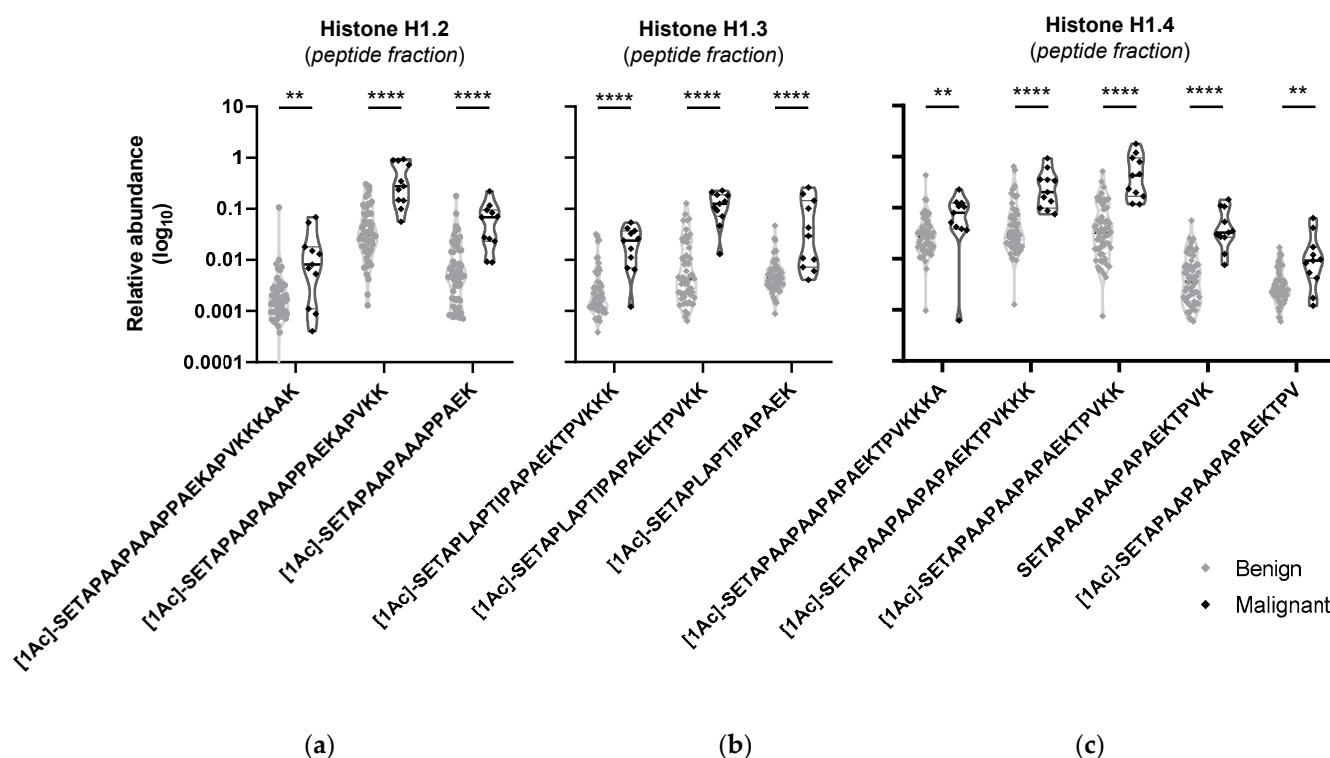
Figure 3. Peptides and acetylated peptides from histones were found increased in malignant samples: violin plots with individual sample representation of peptides of histones (**a**) H1.2, (**b**) H1.3 and (**c**) H1.4 from the peptidomics analysis of the peptide fraction of tissue samples of benign versus malignant groups. Peptide sequences are represented by amino acid one letter code and acetylation at the N-termini is represented by [1Ac]. Benign samples are represented in grey and malignant ones in black. ** $p \leq 0.01$, **** $p \leq 0.0001$ (Mann–Whitney test).

Both datasets were also compared to find which peptide sequences presented contradictory behaviours. Although Poly(rC)-binding protein 1 was not found statistically different in the proteomics analysis, the peptide with the sequence FAGIDSSSPEVKG of this protein was upregulated in the supernatant fraction (Figure 4b). In contrast, the peptide QQSHFAMMHGGTG<u>FAGIDSSSPEVK</u> of this protein, with part of the same amino acid sequence (underlined), was found downregulated in the protein fraction (Figure 4a). In addition, no correlation was found between this peptide and the protein (Figure 4c). Considering that this analysis is based on naturally occurring peptides, the cleavage of the peptide before the last phenylalanine could be performed by pepsin according to the Expasy Peptide Cutter tool and by cathepsin B according to TopFIND and MEROPS. The shift in regulation from one peptide to the other suggests that this enzyme could be more active in the malignant participants than in the benign.
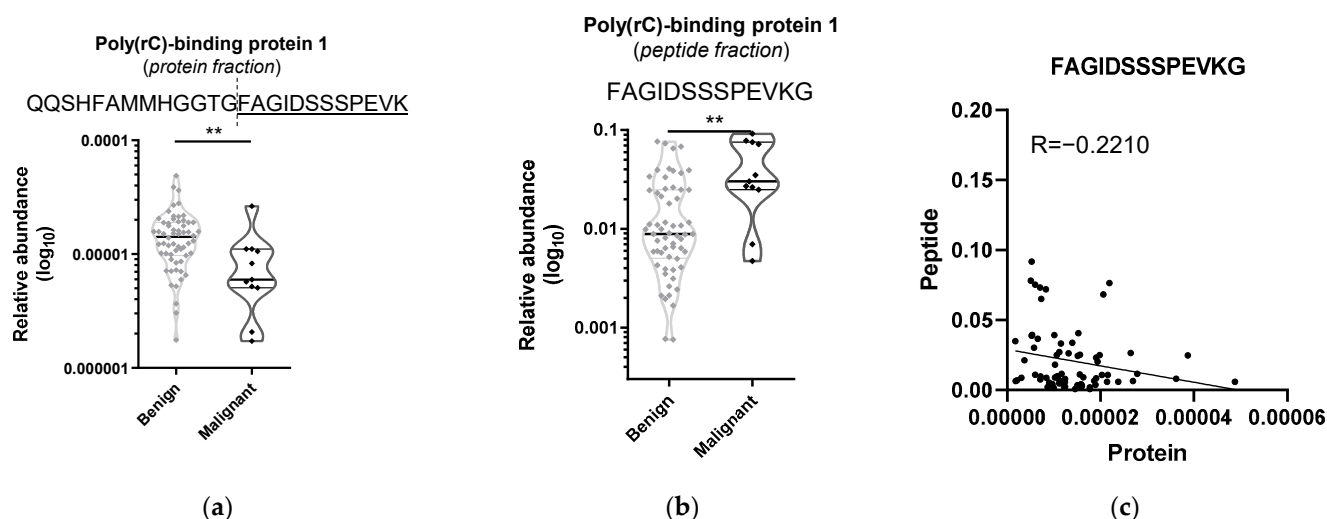
Figure 4. Differential abundance in the same amino acid sequence of Poly(rC)-binding protein 1 suggests differences in enzymatic activity: violin plots with individual sample representations of (**a**) Poly(rC)-binding protein 1 peptide from the proteomics analysis and (**b**) part of that same peptide from the peptidomics analysis of the peptide fraction of the same tissue samples, of benign versus malignant groups. Peptide sequences are represented by amino acid one letter code. Benign samples are represented in grey and malignant ones in black. The dashed vertical line in the peptide sequence represents a possible cleavage site. ** $p \leq 0.01$ (Mann–Whitney test). (**c**) Correlation between poly(rC)-binding protein 1 from the proteomics analysis and peptide FAGIDSSSPEVKG from the peptidomics analysis. Non-significant (Pearson correlation).

3. Discussion

Proteomics screenings have been previously applied in the study of thyroid cancer tissues [4,5,9,10], but to our knowledge there are no reports of peptidomics screening. Single peptides can be produced by the breakdown and turnover of regular proteins or by cleavage of a pro-hormone sequence generating a bioactive entity on its own [6]. Therefore, alterations to these peptides can also inform about disease processes such as thyroid cancer. In this work, proteomics and peptidomics approaches were applied to look for differences between benign and malignant thyroid nodules and explore the relationships between proteins and native peptides. Differences between the benign and malignant groups were found in both the proteome and the peptide screenings of these samples (Figure 1, Table S2, Figure S1).

A specific upregulation of annexin A1 in carcinomas of follicular cell origin has been reported previously [11–13] and suggests that this protein could be a diagnostic and prognostic biomarker for thyroid cancer. This calcium and phospholipid-binding protein participates in inflammatory processes, cell proliferation modulation, cell death regulation and tumour formation and development. The proteomics analysis in this work confirms an upregulation of annexin A1 in the malignancy group, but the peptidomics analysis of the same samples revealed corresponding alterations in stand-alone peptides that were not detected in the proteomics screening (Figure 2).

The peptides of three histones were found to be statistically different in malignancy (Figure 3). Moreover, most of these peptides were acetylated, and it is known that this post-translational modification has a role in regulating inflammation, one of the major hallmarks of cancer [14]. Post-translational modifications in histones regulate accessibility to DNA and consequently gene transcription. Acetyl groups neutralize the histones' positive charges, opening the chromatin structure and allowing transcription machinery to bind to DNA. In this case, a higher histone acetylation expression, such as in most histone peptides identified in this work, suggests higher gene transcription in malignant thyroid nodules. Moreover, this modification has been previously found in the histones

of cancer cells, including thyroid cancer [15,16]. Based on these results, it would also be interesting to assess the activity of histone acetyltransferases, as it has been implicated in thyroid cancer [17,18]. Furthermore, histone cleavage/regulation by enzymatic activity as a mechanism to increase gene transcription deserves further study.

The study of peptides can be informative based on their individual roles in disease, and analysis of their amino acid sequence can also inform about differential enzyme activity. Part of a peptide of poly(rC)-binding protein 1 detected in the pellet fraction was detected in the supernatant but with an opposite profile (Figure 4). Considering which enzyme(s) could have been responsible for this specific cut, pepsin is at first sight an unlikely candidate since it is expressed in the stomach. Nonetheless, considering that thyroid hormones can influence pepsin secretion, pepsin activity could be modified in thyroid cancer [19]. Pepsin has also been associated with laryngopharyngeal cancer [20,21], not only by damaging epithelial cells but also by disruption of signalling pathways [22]. Such mechanisms might also be active in thyroid malignancy. However, cathepsin B is the most likely candidate since this proteinase is frequently found to be increased in cancer and is thought to contribute to invasive and metastatic properties of cancer [23], and it has also been previously associated with thyroid cancer [24,25]. The results of this work suggest a higher enzymatic activity in malignant patients, which requires further validation. On the other hand, poly(rC)-binding protein 1 is an RNA-binding protein that can function as a tumour suppressor, having been found to be downregulated in many cancer types [26]. Therefore, increased degradation of this protein could dysregulate gene transcription to promote tumorigenesis, raising the question of whether inhibition of the protease that cleaves poly(rC)-binding protein can act as a tumour suppressor.

Peptides are more susceptible to degradation in biological matrices compared to proteins. In that sense, one limitation of this work is related to the difficulties in performing rapid freezing or heat inactivation to decrease enzyme activity. However, samples were all processed using the same protocol and kept at low temperatures until storage at $-80\ °C$, thereby limiting enzymatic activity. Also, protease inhibitors were used to prevent further protein degradation when first thawed.

4. Materials and Methods

4.1. Patients

This study was approved by the Ethics Committee of the Centro Hospitalar Universitário São João/Faculdade de Medicina da Universidade do Porto (approval ID 125/18). Informed consent was obtained from all participants. The cohort of thyroid tissue lesions consisted of 71 nodules from 43 patients. The benign group consisted of follicular adenoma and follicular nodular disease, while the malignant group was comprised of differentiated carcinomas of the follicular cells (papillary carcinoma and follicular carcinoma). In some cases, more than one nodule per individual was studied. Groups were gender- and age-matched (Table 1). Tissue samples were obtained during surgical resection and immediately stored at $-80\ °C$ until analysis. The final diagnosis was obtained after a postoperative histopathological examination of the same lesion (Table S1).

Table 1. Demographic summary of the sample cohort.

Group	Sex	Female (%)	Age (Years)	BMI (kg/m^2)	Free T4 (ng/dL)	TSH (um/L)
Benign	Female = 24 Male = 9	72.7	58.4 ± 2.3	28.7 ± 0.9 **	1.11 ± 0.07	0.83 ± 0.12
Malignant	Female = 8 Male = 2	80.0	53.2 ± 7.1	23.7 ± 1.0	1.08 ± 0.04	1.28 ± 0.29

Note: Data represents mean ± standard error of mean. ** $p \leq 0.01$ (Mann–Whitney test). Abbreviations: body mass index (BMI), thyroid-stimulating hormone (TSH) and Thyroxine (T4).

4.2. Peptidomics

Samples were initially analysed by non-destructive high-resolution magic angle spinning (HR-MAS) ^{1}H nuclear magnetic resonance (NMR) for metabolomics screening (not shown in this manuscript), in combination with proteomics. Nodules were recovered from the rotor and stored in a 100 μL 0.5 M triethylammonium bicarbonate (TEAB) solution with protease inhibitors (cOmplete™, ethylenediamine tetraacetic acid (EDTA)-free Protease Inhibitor Cocktail, Roche). Samples not analysed by HR-MAS NMR were also added to the same solution. Tissue samples were homogenised with the Dispersing-aggregates POLYTRON® PT1200 E with a 3 mm tip (Kinematica AG, Malters, Switzerland). The homogenised mixture was centrifuged at $5000 \times g$ for 5 min at 4 °C. The supernatant was harvested, with 5 μL being used for total protein content assessment with the Pierce™ 660 nm Protein Assay Reagent (ThermoFisher™, Waltham, MA, USA), according to the manufacturer's instructions. A volume corresponding to approximately 100 μg was harvested from each sample to continue sample processing. Moreover, pools of benign and malignant lesions were created using 5 μL from selected samples. To each individual and pooled sample, 2 μg of the recombinant protein green fluorescent protein and maltose-binding periplasmic protein (MBP-GFP) were added [27]. Protein precipitation was performed using 400 μL of cold methanol. Samples were incubated overnight at −80 °C and then centrifuged at $20{,}000 \times g$ for 20 min at 4 °C. The supernatant was harvested for peptidomics analysis, while the pellet was used in the proteomics analysis. After evaporation of the supernatant using the Concentrator Plus/Vacufuge® (Eppendorf, Hamburg, Germany), the supernatant fraction was stored at −80 °C until analysis, where it was resolubilized in 35 μL of 2% acetonitrile and 0.1% formic acid containing 1 μM of the standard sulfamethazine-^{2}H$_4$ (Toronto Research Chemicals, North York, ON, Canada) aided by sonication with a Vibra cell 75041 cup horn (Bioblock Scientific, Illkirch, France) at 20% amplitude every 1 s pulse for 2 min. After centrifugation at $14{,}100 \times g$ for 5 min at room temperature, samples were transferred to vials for LC-MS analysis. After individual sample resuspension, five additional pools were created from 5 μL of a selection of prepared samples for identification purposes. Samples were analysed on a NanoLC™ 425 System (Eksigent®, Framingham, MA, USA) coupled to a TripleTOF™ 6600 System (Sciex®, Framingham, MA, USA) using data-dependent acquisition (DDA) on pooled samples for peptide identification and data-independent acquisition (DIA/SWATH-MS) acquisition of each individual sample for peptide quantification. Detailed procedures of data acquisition are described in Appendix A (Materials and Methods A1). Peptide identification and library generation were performed with ProteinPilot™ 5.0 (Sciex®), while data processing for quantification was performed using the SWATH™ processing plug-in for PeakView™ 2.2 (ABSciex®). Detailed procedures of data processing are described in Appendix A (Materials and Methods A2). The peptidomics data have been deposited to the EMBL-EBI MetaboLights database [28] with the identifier MTBLS5206.

4.3. Proteomics

The pellet fraction, containing the proteins, was resuspended in 30 μL of 2× Laemmli Sample Buffer by sonication with a Vibra cell 75041 cup horn (Bioblock Scientific, Illkirch, France). Samples were incubated at 95 °C for 5 min in a Thermomixer comfort (Eppendorf, Hamburg, Germany) and 2 μL of 40% acrylamide (Bio-Rad Laboratories, Lda., Hercules, CA, USA) were added as an alkylating agent. Individual and pooled samples were loaded into an SDS-PAGE 4–20% Mini-PROTEAN® TGX™ precast gel and resolved at 110 V [29]. Gel staining was performed as previously described [30]. Each lane was divided into fractions, and each was divided into smaller pieces with the help of a scalpel and added to a 96 multi-well plate containing ddH$_2$O. After destaining the gel pieces with a 50 mM ammonium bicarbonate solution and 30% acetonitrile, in-gel digestion and peptide extraction were performed as previously described [30]. Peptides were evaporated in the Concentrator Plus/Vacufuge® (Eppendorf, Hamburg, Germany) and resolubilized in 30 μL of 2% acetonitrile and 0.1% formic acid aided by sonication with a Vibra cell 75,041 cup horn (Bioblock

Scientific, Illkirch, France) at 20% amplitude every 1 s pulse for 2 min. After centrifugation at $14,100 \times g$ for 5 min at room temperature, samples were transferred to vials for LC-MS analysis. Samples were analysed on a NanoLC™ 425 System (Eksigent®, Framingham, MA, USA) coupled to a TripleTOF™ 6600 System (Sciex®) using DDA for each fraction of the pooled samples for protein identification and DIA/SWATH-MS acquisition of each individual sample for protein quantification. Detailed procedures of data acquisition are described in Appendix A (Materials and Methods A3). Peptide identification and library generation were performed with ProteinPilot™ 5.0 (Sciex®), while data processing for quantification was performed using the SWATH™ processing plug-in for PeakView™ 2.2 (ABSciex®). Detailed procedures of data processing are described in Appendix A (Materials and Methods A4). The proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [31] partner repository with the dataset identifier PXD035583.

4.4. Data Analysis

Multivariate analysis was performed in Metaboanalyst 5.0 (https://www.metaboanalyst.ca) [32]. Log-transformation and Pareto scale were performed for principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA). All ellipses in the scores plots for both PCA and PLS models, were drawn at the 95% confidence level. Receiver operating characteristic (ROC) analysis was also performed on this platform, with the same normalization parameters. Multivariate ROC curves were based on a linear support vector machine (SVM) classification model and SVM built-in feature ranking method. Heatmaps with hierarchical clustering used a Euclidean distance measure and Ward clustering method.

Correlations and linear regressions were performed on GraphPad Prism 6.0 (GraphPad Software Inc., San Diego, CA, USA). Due to the small number of samples in each group and the lack of normal distribution of the populations, a Mann–Whitney test was applied to test differences between groups using IBM SPSS Statistics 23 (IBM®, Armonk, NY, USA).

Prediction of potential protease or chemical cleavage sites in a protein sequence was performed on Expasy Peptide Cutter web tool [33], TopFIND [34] and MEROPS [35].

5. Conclusions

This work demonstrates how different analytical approaches on the same samples can provide additional data, in this case uncovering small peptides that would otherwise have gone unnoticed with proteomics alone. In addition, the comparison of peptidomics and proteomics results from thyroid nodules revealed different results at the peptide level that can also be explored in the future. The peptidomics approach can also indicate potential enzymatic activity, which requires further validation. Regarding biomarker potential, the peptides themselves could be searched in the aspirates of FNAB and/or in blood samples as tissue or as peripheral biomarkers using targeted approaches.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms25010376/s1. References [3,36] are cited in the supplementary materials.

Author Contributions: Conceptualization, M.C., J.C., V.M.M., J.G.J., L.R. and B.M.; formal analysis, M.C., J.C., V.M.M., J.P., M.S.F. and I.A.; investigation, M.C., J.C., J.G.J., L.R. and B.M.; resources, J.G.J. and B.M.; writing—original draft preparation, M.C.; writing—review and editing, J.G.J. and B.M. All authors have read and agreed to the published version of the manuscript.

Appendix A

Materials and Methods A1. Peptidomics (data acquisition): Samples were analyzed on a NanoLC™ 425 System (Eksigent®, Framingham, MA, USA) coupled to a TripleTOF™ 6600 System (Sciex®, Framingham, MA, USA) using data-dependent acquisition (DDA) for pooled sample for peptide identification and data-independent acquisition (DIA/SWATH-MS) of each individual sample for peptide quantification. Mass spectrometer operated by Analyst® TF 1.8.1 software (Sciex®). Samples were loaded onto a YMC-Triart C18 Capillary Guard Column 1/32″ (12 nm, S-3μm, 5×0.5 mm) (YMC, Kyoto, Japan) at 5 μL/min of 5% of mobile phase B for 4 min and the separation of the compounds was carried out by micro-flow liquid chromatography using a YMC-Triart C18 Capillary Column (12 nm, S-3μm, 150×0.3 mm) (YMC, Kyoto, Japan) at 50 °C. The flow rate was set to 5 μL/min and mobile phases A and B were 0.1% formic acid in water and 0.1% formic acid in acetonitrile, respectively. The LC program was performed as follows: 5–50% of B (0–20 min), 50–95% of B (20–25 min), 95% of B (25–30 min), 95–5% of B (30–31 min), 5% of B (31–40 min). Samples were eluted into the mass spectrometer using an electrospray ionization source (DuoSpray™ Source, ABSciex®, Framingham, MA, USA) with a 25 μm internal diameter hybrid PEEKsil/stainless steel emitter (ABSciex®, Framingham, MA, USA). The ionization source was operated in the positive mode set to an ion spray voltage of 5500 V, 25 psi for nebulizer gas 1 (GS1), 10 psi for nebulizer gas 2 (GS2), 25 psi for the curtain gas (CUR), and source temperature (TEM) at 100 °C. For DDA experiments, the mass spectrometer was set to scanning full spectra (m/z 50–2250) for 250 ms, followed by up to 40 MS2 scans (m/z 10–1500) per cycle, to maintain a cycle time of 1.9 s. The accumulation time of each MS2 scan was adjusted per the precursor intensity (minimum of 40 ms for precursor above the intensity threshold of 2000). Candidate ions with counts above a minimum threshold of 50 counts per second were isolated for fragmentation and one MS2 spectrum was collected before adding those ions to the exclusion list for 10 s. The collision energy (CE) applied to each m/z was 40 eV CE and the collision energy spread (CES) was 30 eV. For DIA/SWATH-MS experiments, the mass spectrometer was operated in a looped product ion mode 4 and specifically tuned to a set of 73 overlapping windows, covering the precursor mass range of 50–1500 m/z, with the same chromatographic conditions used in the DDA run described above. A 250 ms survey scan (m/z 50–2500) was acquired at the beginning of each cycle and SWATH MS2 spectra were collected from 50–2500 m/z for 23 ms resulting in a cycle time of 2.0 s. The CE applied to each m/z window was 40 eV and the CES was 30 eV for each m/z window.

Materials and Methods A2. Peptidomics (data processing): A specific library of precursor masses and fragment ions was created by combining all files from the DDA experiments (the five fractions from the three pools) to be used for subsequent SWATH processing. Peptide identification and library generation were performed with ProteinPilot™ 5.0 software

(Sciex), using the following parameters: (i) search against a database composed of the Swiss-Prot's Homo sapiens proteome (downloaded in March 2021) and MBP-GFP (internal standard); (ii) focus on biological modifications; (iii) no cysteine alkylation or digestion. An independent False Discovery Rate (FDR) analysis, using the target-decoy approach provided by ProteinPilot™, was used to assess the quality of the identifications and confident identifications were considered when identified proteins reached a 5% local FDR [37,38]. Data processing for quantification was performed using the SWATH™ processing plug-in for PeakView™ 2.2 (ABSciex®). All identified peptides, with up to five fragments each, were chosen per protein, and quantitation was executed for all confidently identified peptides (FDR > 5%) from the ProteinPilot™ search. Peptides' confidence threshold was determined based on an FDR analysis using the target-decoy approach and those that met the 1% FDR threshold in at least 30% of one group of biological replicates were retained, and the peak areas of the target fragment ions of those peptides were extracted across the experiments using an extracted-ion chromatogram (XIC) window of 5 min with 100 ppm XIC width. Peptide's relative abundance was estimated by summing all the filtered transitions for a given peptide and normalized to the sum of all detected features (m/z and retention time information) by the MarkerView™ 1.3.1. software (ABSciex®) with a retention time tolerance of 0.6 min, 50 ppm mass tolerance, 100 intensity threshold, 10 scans of subtraction offset, 1.3 subtraction mult. factor, 100 noise threshold, 0.01 Da min. spectral peak width and 5 scans min. RT peak width.

Materials and Methods A3. Proteomics (data acquisition): Samples were analyzed on a NanoLC™ 425 System (Eksigent®, Framingham, MA, USA) coupled to a TripleTOF™ 6600 System (Sciex®, Framingham, MA, USA) using DDA for each fraction of the pooled samples for protein identification and DIA/SWATH-MS acquisition of each individual sample for protein quantification. The mass spectrometer was operated by Analyst® TF 1.8.1 software (Sciex®). Samples were loaded onto a YMC-Triart C18 Capillary Guard Column "1/32" (12 nm, S-3 μm, 5 × 0.5 mm) (YMC, Kyoto, Japan) at 5 μL/min of 5% of mobile phase B for 8 min and the peptides separation was carried out by micro-flow liquid chromatography using a YMC-Triart C18 Capillary Column (12 nm, S-3 μm, 150 × 0.3 mm) (YMC) at 50 °C. The flow rate was set to 5 μL/min and mobile phases A and B were 5% dimethyl sulfoxide (DMSO) plus 0.1% formic acid in water and 5% DMSO plus 0.1% formic acid in acetonitrile, respectively. The LC program was performed as follows: 5–30% of B (0–50 min), 30–98% of B (50–52 min), 98% of B (52–54 min), 98–95% of B (54–56 min), 95% of B (56–65 min). Peptides were eluted into the mass spectrometer using an electrospray ionization source (DuoSpray™ Source, ABSciex®, Framingham, MA, USA) with a 25 μm internal diameter hybrid PEEKsil/stainless steel emitter (ABSciex®). The ionization source was operated in the positive mode set to an ion spray voltage of 5500 V, 20 psi for nebulizer gas 1 (GS1), 15 psi for nebulizer gas 2 (GS2), 25 psi for the curtain gas (CUR), and source temperature (TEM) at 100 °C. For DDA experiments, the mass spectrometer was set to scanning full spectra (m/z 350–2250) for 250 ms, followed by up to 100 MS2 scans (m/z 100–1500) per cycle, to maintain a cycle time of 3.295 s. The accumulation time of each MS2 scan was adjusted per the precursor intensity (minimum of 25 ms for precursor above the intensity threshold of 2000). Candidate ions with a charge state between +1 and +5 and counts above a minimum threshold of 10 counts per second were isolated for fragmentation and one MS2 spectrum was collected before adding those ions to the exclusion list for 15 s. The rolling collision was used with a collision energy spread (CES) of 5. The mass spectrometer was operated in a looped product ion mode 4 with the same chromatographic conditions used in the DDA run described above for SWATH-MS experiments. A set of 160 windows of variable width (containing an m/z of 1 for the window overlap) was constructed, covering the precursor mass range of m/z 350–2250. A 50 ms survey scan (m/z 350–2250) was acquired at the beginning of each cycle for instrument calibration and SWATH MS2 spectra were collected from the precursors ranging from m/z 350 to 2250 for m/z 100–1500 for 20 ms resulting in a cycle time of 3.3 s The collision energy (CE) applied

to each m/z window was determined considering the appropriate CE for a +2 ion centered upon this window and the collision energy spread was also adapted to each m/z window.

Materials and Methods A4. Proteomics (data processing): A specific library of precursor masses and fragment ions was created by combining all files from the DDA experiments (the five fractions from the three pools) to be used for subsequent SWATH processing. Peptide identification and library generation were performed with ProteinPilot™ 5.0 software (Sciex), using the following parameters: (i) search against a database composed of the Swiss-Prot's Homo sapiens proteome (downloaded in March 2021) and MBP-GFP (internal standard); (ii) focus on biological modifications; (iii) acrylamide alkylated cysteines as fixed modification; (iv) trypsin as digestion enzyme. An independent False Discovery Rate (FDR) analysis, using the target-decoy approach provided by ProteinPilot™, was used to assess the quality of the identifications and confident identifications were considered when identified proteins reached a 5% local FDR [37,38]. Data processing for quantification was performed using the SWATH™ processing plug-in for PeakView™ 2.2 (ABSciex®). After retention time (RT) adjustment using the MBP-GFP peptides, up to 20 peptides, with up to five fragments each, were chosen per protein, and quantitation was executed for all confidently identified proteins (FDR > 5%) from the ProteinPilot™ search. Peptides' confidence threshold was determined based on an FDR analysis using the target-decoy approach and those that met the 1% FDR threshold in at least 30% of one group of biological replicates were retained. The peak areas of the target fragment ions of those peptides were extracted across the experiments using an extracted-ion chromatogram (XIC) window of 5 min with 100 ppm XIC width. The protein's relative abundance was estimated by summing all the filtered transitions from all the filtered peptides for a given protein and normalizing to the sum of all quantified proteins.

References

1. Grussendorf, M.; Ruschenburg, I.; Brabant, G. Malignancy rates in thyroid nodules: A long-term cohort study of 17,592 patients. *Eur. Thyroid. J.* **2022**, *11*, e220027. [CrossRef] [PubMed]
2. Nikiforov, Y.E.; Nikiforova, M.N. Molecular genetics and diagnosis of thyroid cancer. *Nat. Rev. Endocrinol.* **2011**, *7*, 569–580. [CrossRef] [PubMed]
3. Ali, S.Z.; Baloch, Z.W.; Cochand-Priollet, B.; Schmitt, F.C.; Vielh, P.; VanderLaan, P.A. The 2023 Bethesda System for Reporting Thyroid Cytopathology. *Thyroid* **2023**, *33*, 1039–1044. [CrossRef] [PubMed]
4. Damante, G.; Scaloni, A.; Tell, G. Thyroid tumors: Novel insights from proteomic studies. *Expert Rev. Proteom.* **2009**, *6*, 363–376. [CrossRef] [PubMed]
5. Pagni, F.; L'Imperio, V.; Bono, F.; Garancini, M.; Roversi, G.; De Sio, G.; Galli, M.; Smith, A.J.; Chinello, C.; Magni, F. Proteome analysis in thyroid pathology. *Expert Rev. Proteom.* **2015**, *12*, 375–390. [CrossRef]
6. Foreman, R.E.; George, A.L.; Reimann, F.; Gribble, F.M.; Kay, R.G. Peptidomics: A Review of Clinical Applications and Methodologies. *J. Proteome Res.* **2021**, *20*, 3782–3797. [CrossRef]
7. Schulz-Knappe, P.; Zucht, H.D.; Heine, G.; Jurgens, M.; Hess, R.; Schrader, M. Peptidomics: The comprehensive analysis of peptides in complex biological mixtures. *Comb. Chem. High. Throughput Screen.* **2001**, *4*, 207–217. [CrossRef]
8. Coelho, M.; Capela, J.; Anjo, S.I.; Pacheco, J.; Fernandes, M.S.; Amendoeira, I.; Jones, J.G.; Raposo, L.; Manadas, B. Proteomics Reveals mRNA Regulation and the Action of Annexins in Thyroid Cancer. *Int. J. Mol. Sci.* **2023**, *24*, 14542. [CrossRef]
9. Krause, K.; Jessnitzer, B.; Fuhrer, D. Proteomics in thyroid tumor research. *J. Clin. Endocrinol. Metab.* **2009**, *94*, 2717–2724. [CrossRef]
10. Ucal, Y.; Ozpinar, A. Proteomics in thyroid cancer and other thyroid-related diseases: A review of the literature. *Biochim. Biophys. Acta Proteins Proteom.* **2020**, *1868*, 140510. [CrossRef]
11. Liu, X.; Yang, M.; Guo, Y.; Lu, X. Annexin A10 is a novel prognostic biomarker of papillary thyroid cancer. *Ir. J. Med. Sci.* **2021**, *190*, 59–65. [CrossRef] [PubMed]
12. Petrella, A.; Festa, M.; Ercolino, S.F.; Zerilli, M.; Stassi, G.; Solito, E.; Parente, L. Annexin-1 downregulation in thyroid cancer correlates to the degree of tumor differentiation. *Cancer Biol. Ther.* **2006**, *5*, 643–647. [CrossRef] [PubMed]
13. Zhao, X.; Ma, W.; Li, X.; Li, H.; Li, J.; Li, H.; He, F. ANXA1 enhances tumor proliferation and migration by regulating epithelial-mesenchymal transition and IL-6/JAK2/STAT3 pathway in papillary thyroid carcinoma. *J. Cancer* **2021**, *12*, 1295–1306. [CrossRef] [PubMed]
14. Lin, Y.; Qiu, T.; Wei, G.; Que, Y.; Wang, W.; Kong, Y.; Xie, T.; Chen, X. Role of Histone Post-Translational Modifications in Inflammatory Diseases. *Front. Immunol.* **2022**, *13*, 852272. [CrossRef]
15. Puppin, C.; Passon, N.; Lavarone, E.; Di Loreto, C.; Frasca, F.; Vella, V.; Vigneri, R.; Damante, G. Levels of histone acetylation in thyroid tumors. *Biochem. Biophys. Res. Commun.* **2011**, *411*, 679–683. [CrossRef]

16. Zhang, L.; Xiong, D.; Liu, Q.; Luo, Y.; Tian, Y.; Xiao, X.; Sang, Y.; Liu, Y.; Hong, S.; Yu, S.; et al. Genome-Wide Histone H3K27 Acetylation Profiling Identified Genes Correlated With Prognosis in Papillary Thyroid Carcinoma. *Front. Cell Dev. Biol.* **2021**, *9*, 682561. [CrossRef]
17. West, A.C.; Johnstone, R.W. New and emerging HDAC inhibitors for cancer treatment. *J. Clin. Investig.* **2014**, *124*, 30–39. [CrossRef]
18. Zhang, Z.; Liu, D.; Murugan, A.K.; Liu, Z.; Xing, M. Histone deacetylation of NIS promoter underlies BRAF V600E-promoted NIS silencing in thyroid cancer. *Endocr. Relat. Cancer* **2014**, *21*, 161–173. [CrossRef]
19. Hays, M.T. Thyroid hormone and the gut. *Endocr. Res.* **1988**, *14*, 203–224. [CrossRef]
20. Niu, K.; Guo, C.; Teng, S.; Zhou, D.; Yu, S.; Yin, W.; Wang, P.; Zhu, W.; Duan, M. Pepsin promotes laryngopharyngeal neoplasia by modulating signaling pathways to induce cell proliferation. *PLoS ONE* **2020**, *15*, e0227408. [CrossRef]
21. Sereg-Bahar, M.; Jerin, A.; Hocevar-Boltezar, I. Higher levels of total pepsin and bile acids in the saliva as a possible risk factor for early laryngeal cancer. *Radiol. Oncol.* **2015**, *49*, 59–64. [CrossRef] [PubMed]
22. Samuels, T.L.; Zimmermann, M.T.; Zeighami, A.; Demos, W.; Southwood, J.E.; Blumin, J.H.; Bock, J.M.; Johnston, N. RNA Sequencing Reveals Cancer-Associated Changes in Laryngeal Cells Exposed to Non-Acid Pepsin. *Laryngoscope* **2021**, *131*, 121–129. [CrossRef] [PubMed]
23. Aggarwal, N.; Sloane, B.F. Cathepsin B: Multiple roles in cancer. *Proteom. Clin. Appl.* **2014**, *8*, 427–437. [CrossRef] [PubMed]
24. Kim, E.K.; Song, M.J.; Jang, H.H.; Chung, Y.S. Clinicopathologic Analysis of Cathepsin B as a Prognostic Marker of Thyroid Cancer. *Int. J. Mol. Sci.* **2020**, *21*, 9537. [CrossRef] [PubMed]
25. Kusunoki, T.; Nishida, S.; Nakano, T.; Funasaka, K.; Kimoto, S.; Murata, K.; Tomura, T. Study on cathepsin B activity in human thyroid tumors. *Auris Nasus Larynx* **1995**, *22*, 43–48. [CrossRef]
26. Zhang, X.; Di, C.; Chen, Y.; Wang, J.; Su, R.; Huang, G.; Xu, C.; Chen, X.; Long, F.; Yang, H.; et al. Multilevel regulation and molecular mechanism of poly (rC)-binding protein 1 in cancer. *FASEB J.* **2020**, *34*, 15647–15658. [CrossRef]
27. Anjo, S.I.; Simoes, I.; Castanheira, P.; Graos, M.; Manadas, B. Use of recombinant proteins as a simple and robust normalization method for untargeted proteomics screening: Exhaustive performance assessment. *Talanta* **2019**, *205*, 120163. [CrossRef]
28. Haug, K.; Cochrane, K.; Nainala, V.C.; Williams, M.; Chang, J.; Jayaseelan, K.V.; O'Donovan, C. MetaboLights: A resource evolving in response to the needs of its scientific community. *Nucleic Acids Res.* **2020**, *48*, D440–D444. [CrossRef]
29. Anjo, S.I.; Santa, C.; Manadas, B. Short GeLC-SWATH: A fast and reliable quantitative approach for proteomic screenings. *Proteomics* **2015**, *15*, 757–762. [CrossRef]
30. Anjo, S.I.; Santa, C.; Manadas, B. SWATH-MS as a tool for biomarker discovery: From basic research to clinical applications. *Proteomics* **2017**, *17*, 1600278. [CrossRef]
31. Perez-Riverol, Y.; Bai, J.; Bandla, C.; Garcia-Seisdedos, D.; Hewapathirana, S.; Kamatchinathan, S.; Kundu, D.J.; Prakash, A.; Frericks-Zipper, A.; Eisenacher, M.; et al. The PRIDE database resources in 2022: A hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **2022**, *50*, D543–D552. [CrossRef] [PubMed]
32. Chong, J.; Wishart, D.S.; Xia, J. Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Curr. Protoc. Bioinform.* **2019**, *68*, e86. [CrossRef] [PubMed]
33. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.e.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*; Walker, J.M., Ed.; Humana Press: Totowa, NJ, USA, 2005; pp. 571–607.
34. Fortelny, N.; Yang, S.; Pavlidis, P.; Lange, P.F.; Overall, C.M. Proteome TopFIND 3.0 with TopFINDer and PathFINDer: Database and analysis tools for the association of protein termini to pre- and post-translational events. *Nucleic Acids Res.* **2015**, *43*, D290–D297. [CrossRef] [PubMed]
35. Rawlings, N.D.; Barrett, A.J.; Thomas, P.D.; Huang, X.; Bateman, A.; Finn, R.D. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* **2018**, *46*, D624–D632. [CrossRef]
36. Lloyd, R.; Osamura, R.; Rosai, J. *WHO Classification of Tumours Editorial Board. Endocrine and Neuroendocrine Tumours*; IARC: Lyon, France, 2022.
37. Tang, W.H.; Shilov, I.V.; Seymour, S.L. Nonlinear fitting method for determining local false discovery rates from decoy database searches. *J. Proteome Res.* **2008**, *7*, 3661–3667. [CrossRef]
38. Sennels, L.; Bukowski-Wills, J.C.; Rappsilber, J. Improved results in proteomics by use of local and peptide-class specific false discovery rates. *BMC Bioinform.* **2009**, *10*, 179. [CrossRef]