

**TECHNICAL BRIEF**

# TermineR: Extracting information on endogenous proteolytic processing from shotgun proteomics data

Miguel Cosenza-Contreras<sup>1,2</sup>  | Adrianna Seredynska<sup>1,2</sup> | Daniel Vogele<sup>1,2,3</sup> |  
 Niko Pinter<sup>2</sup> | Eva Brombacher<sup>1,4,5,6</sup> | Ruth Fiestas Cueto<sup>1,2</sup> | Thien-Ly Julia Dinh<sup>1,2</sup> |  
 Patrick Bernhard<sup>2</sup> | Manuel Rogg<sup>2</sup> | Junwei Liu<sup>7</sup> | Patrick Willems<sup>8,9,10,11</sup> |  
 Simon Stael<sup>8,9,12</sup> | Pitter F. Huesgen<sup>1,4</sup>  | E. Wolfgang Kuehn<sup>7</sup> | Clemens Kreutz<sup>4,5</sup> |  
 Christoph Schell<sup>2,14</sup> | Oliver Schilling<sup>2,13,14</sup> 

<sup>1</sup>Faculty of Biology, University of Freiburg, Freiburg, Germany

<sup>2</sup>Faculty of Medicine, Institute for Surgical Pathology Medical Center—University of Freiburg, Freiburg, Germany

<sup>3</sup>ProtPath Research Training Group, University of Freiburg, Freiburg, Germany

<sup>4</sup>Centre for Integrative Biological Signaling Studies (CIBSS), Freiburg, Germany

<sup>5</sup>Faculty of Medicine and Medical Center, Institute of Medical Biometry and Statistics, University of Freiburg, Freiburg, Germany

<sup>6</sup>Spemann Graduate School of Biology and Medicine (SGBM), University of Freiburg, Freiburg, Germany

<sup>7</sup>Department of Medicine IV, Faculty of Medicine, Medical Center—University of Freiburg, Freiburg, Germany

<sup>8</sup>Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium

<sup>9</sup>VIB-UGent Center for Plant Systems Biology, Ghent, Belgium

<sup>10</sup>Department of Biomolecular Medicine, Ghent University, Ghent, Belgium

<sup>11</sup>Center for Medical Biotechnology, VIB, Ghent, Belgium

<sup>12</sup>Department of Molecular Sciences, Uppsala BioCenter, Swedish University of Agricultural Sciences and Linnean Center for Plant Biology, Uppsala, Sweden

<sup>13</sup>German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ), Heidelberg, Germany

<sup>14</sup>Freiburg Institute for Advanced Studies (FRIAS), University of Freiburg, Freiburg, Germany

## Correspondence

Oliver Schilling, Breisacher Str. 115A, 79106  
 Freiburg, Germany.  
 Email:  
[oliver.schilling@mol-med.uni-freiburg.de](mailto:oliver.schilling@mol-med.uni-freiburg.de)

## Funding information

Deutsche Forschungsgemeinschaft; DFG, projects, Grant/Award Numbers: 446058856, 466359513, 444936968, 405351425, 431336276, 431984000; SFB 1453 “NephGen”, Grant/Award Number: 441891347; SFB 1479 “OncoEscape”, Grant/Award Number: 423813989; GRK 2606 “ProtPath”, Grant/Award Number:

## Abstract

State-of-the-art mass spectrometers combined with modern bioinformatics algorithms for peptide-to-spectrum matching (PSM) with robust statistical scoring allow for more variable features (i.e., post-translational modifications) being reliably identified from (tandem-) mass spectrometry data, often without the need for biochemical enrichment. Semi-specific proteome searches, that enforce a theoretical enzymatic digestion to solely the N- or C-terminal end, allow to identify of native protein termini or those arising from endogenous proteolytic activity (also referred to as “neo-N-termini” analysis or “N-terminomics”). Nevertheless, deriving biological meaning from these search outputs can be challenging in terms of data mining and analysis. Thus,

**Abbreviations:** EBI, European Bioinformatics Institute; EtO, Hethanol; FFPE, Formalin-fixed paraffin embedded; HCD, high collision dissociation; KO, knock-out; ORF, Open Reading Frame; PKD, polycystic kidney disease; PSM, peptide-to-spectrum match; TCEP, tris(2-carboxyethyl)phosphine; TMT, tandem mass tag; WT, wild-type.

Miguel Cosenza-Contreras and Adrianna Seredynska contributed equally to this study

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2024 The Author(s). *PROTEOMICS* published by Wiley-VCH GmbH.

322977937; GRK 2344 "MeInBio"; ERA PerMed Program, Grant/Award Numbers: 01KU1916, 01KU1915A; German Consortium for Translational Cancer Research (project Impro-Rec); MatrixCode Research Group; investBW program; ERA TransCan program, Grant/Award Number: 01KT2201; BMBF KMUI program, Grant/Award Number: 13GW0603E; BMBF Cluster4Future program (nanodiag

we introduce *TermineR*, a data analysis approach for the (1) annotation of peptides according to their enzymatic cleavage specificity and known protein processing features, (2) differential abundance and enrichment analysis of N-terminal sequence patterns, and (3) visualization of neo-N-termini location. We illustrate the use of *TermineR* by applying it to tandem mass tag (TMT)-based proteomics data of a mouse model of polycystic kidney disease, and assess the semi-specific searches for biological interpretation of cleavage events and the variable contribution of proteolytic products to general protein abundance. The *TermineR* approach and example data are available as an R package at <https://github.com/MiguelCos/TermineR>.

#### KEYWORDS

data processing, polycystic kidney disease, proteolysis, terminomics

## 1 | INTRODUCTION

Proteolysis is an irreversible post-translational protein modification yielding truncated, stable products that result in either the mature form of a protein (i.e., zymogen activation or removal of signal peptide) or previously undescribed (non-canonical) cleavage with often altered functionality and representing novel N- or C-termini (Figure 1A) [1]. Proteolytic disturbance has been associated with diverse diseases such as heart failure [2], cancer [3], kidney disease [4], and others. Insights into the proteolytic activity in the context of various cell states or clinical conditions to infer regulatory targets for drugs and disease prognosis can be gained by large-scale probing of protein termini, termed "terminomics" analyses [5], conducted using liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). LC-MS/MS relies on the experimental proteolysis of proteins (i.e., trypsin digestion) to produce peptides with known specificities (i.e., tryptic peptides) that are then ionized and measured in the mass spectrometer to produce sequence-specific spectra [6]. Typically, the algorithm for spectra processing and peptide identification would be constrained in the search for peptides with certain enzymatic specificity ("specific search"), which purposely limits the search space to improve computation times and identification sensitivity, while overlooking non-specific peptides. In this context, proteolytic products (i.e., peptides arising from endogenous proteolysis) comprise a small fraction of a digested sample, and their identification would be ignored in shotgun experiments.

For this reason, a variety of experimental methods have been developed for the selective enrichment of protein N-termini or C-termini [5, 7, 8]; and efforts have been made towards standardized processing of N-terminomics search results from the widely used MaxQuant software [9]. These methods, although sensitive, are usually laborious, intrinsically include sample loss, and the inability to simultaneously explore proteolytic products in the wider proteomic context of the same sample and measurement.

On the other hand, semi-specific peptide searches can probe from peptides with a non-specific N- or C-terminal cleavage end (i.e., semi-tryptic) from shotgun LC-MS/MS experiments. This allows for the

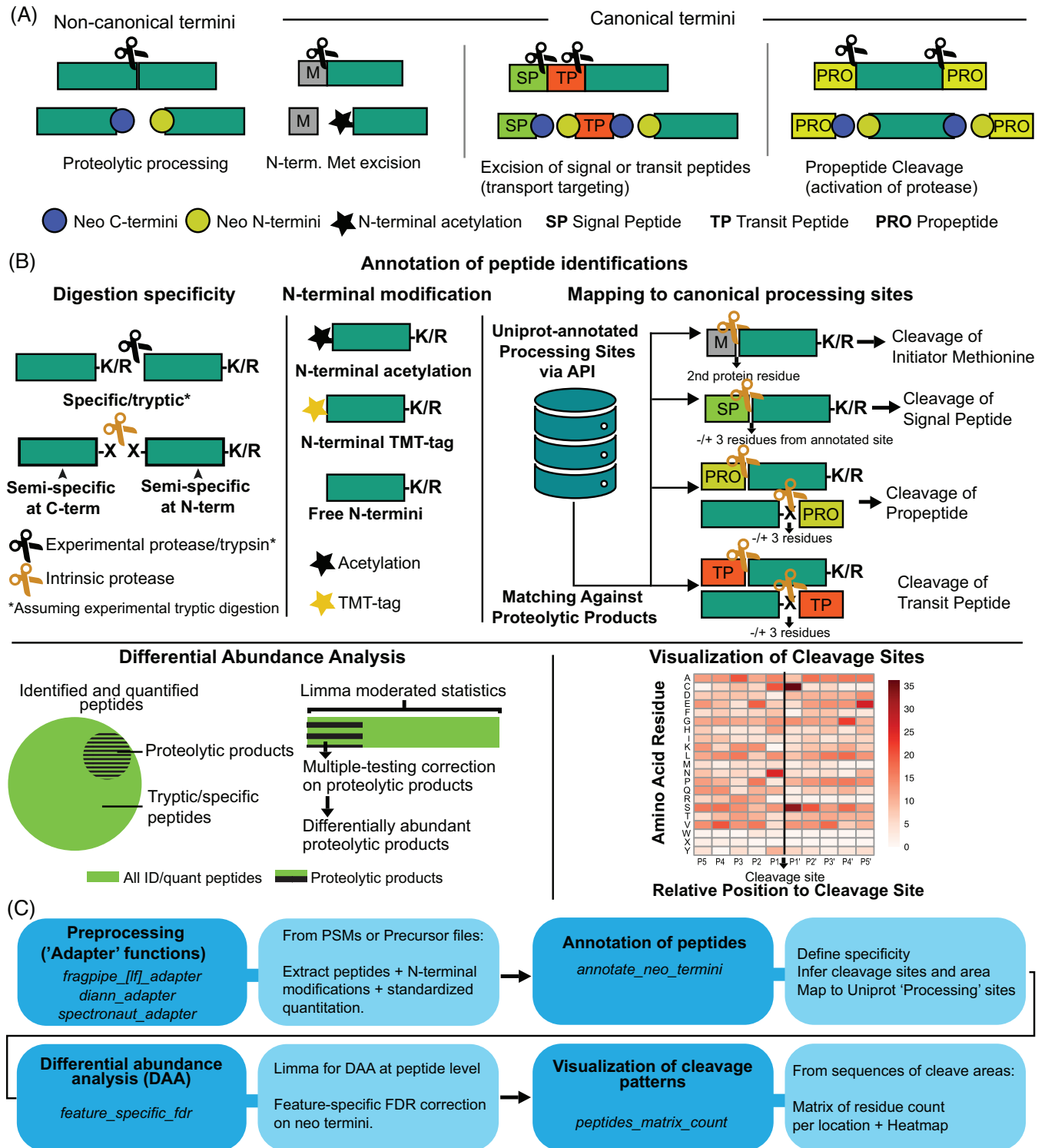
bioinformatic identification of peptides associated with truncated proteins after endogenous processing. This supports the observation of previously undescribed biological processes [10], without the requirement of biochemical enrichment of terminal peptides [11]. The use of modern search engines allows for fast peptide-to-spectrum matching (PSM) [12, 13] which, in combination with algorithms for probabilistic scoring of PSMs for false discovery rate (FDR) control [14, 15], enables the reliable identification of both semi-specific and fully specific peptides despite the increased search space. FragPipe offers an alternative for peptide/protein identification by integrating the MSFragger search engine ([12, 16], p. 202) with deep learning-assisted probabilistic scoring of PSMs [15, 17]. Nevertheless, due to the complexity of the data, there is a lack of standardized methods for the extraction of interpretable and biologically meaningful information from semi-specific searches related to both proteolytic processing and proteomic fingerprint.

Here we present *TermineR*, a reproducible data analysis framework for the annotation and quantitative evaluation of proteolytic products identified originating from semi-specific searches of shotgun proteomics and N-terminomics experimental setups. We wrapped the *TermineR* main data processing functionalities into an R package, freely accessible via GitHub, featuring functions for data preparation, annotation, statistical analysis, visualization of cleavage motifs, and annotation of proteolytic products based on publicly available UniProt processing annotation. We showcase the use of this method for the extraction of proteolytic processing information using a mouse model of polycystic kidney disease (PKD).

## 2 | MATERIALS AND METHODS

### 2.1 | Sample materials

Fresh frozen kidney tissue samples from 6 Pkd1<sup>fl/fl</sup>;Ksp-CrePax8rtTA;TetOCre mice (KO) [18] presenting enlarged cystic kidneys and as well from 5 wild-type mice (WT) were analyzed. Animal experiments were approved by the local animal ethics committee



**FIGURE 1** General outline of the *TermineR* approach for processing, annotation, statistical analysis, visualization, and biological interpretation of semi-specific search results. (A) Differences between canonical and non-canonical terminal peptides/proteolytic products. (B) Descriptive summary of *TermineR* approach. It (1) annotates the peptides based on specificity, label N-terminal modifications, (2) annotate the proteolytic products based on Uniprot processing sites, (3) perform differential abundance statistics on proteolytic products, and (4) visualize cleavage patterns. (C) Summary of *TermineR* data analysis workflow.

(Project Nr G-19/29). HeLa standards were obtained commercially as digested peptides (Thermo Scientific).

## 2.2 | Evaluation of the effect of protein extraction methods on the amount of semi-tryptic peptides

The effect of three tissue homogenization methods (beat beating [Precellys], sonicator [UIP400MTP, Hielscher], and bioruptor [Diagenode]) on spontaneous cleavage was evaluated by performing a pilot extraction on three fresh frozen tissue samples from each condition. The Precellys was set to 4000 rpm with two programs: 1 cycle, 30 s shaking and 5 s break, and 3 cycles, 30 s shaking and 10 s break. The Bioruptor is used in 10 cycles with 40/20 s on/off. The 96 well-plate sonicator (Hielscher) was set with 100 Wh, amplitude 80% with 10/10 s on/off. Protein extracts were further reduced, alkylated, and digested with trypsin (see Protein Extraction and Protein-Level Labeling section). Digested peptides were loaded into Evtotips (Evosep), injected using a 30SPD method (Evosep) into a Tims-TOF flex mass spectrometer, and measured in DIA mode (see LC-MS/MS section below) (Figure S1).

## 2.3 | Protein extraction and protein-level labeling

Tissues samples from mouse kidneys were homogenized using beat beating (Precellys) in 2% SDS in (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)) buffer. Reductive alkylation was performed using 5 mM tris(2-carboxyethyl) phosphine (TCEP) and 20 mM chloroacetamide followed by 30 min incubation at room temperature in dark. Subsequently, protein-level labeling of native and neo-protein-N-termini and lysine sidechains was conducted using TMT16plex (Thermo Scientific), by adding the isobaric tags to the respective protein extracts at a 1:8 (w/w) protein to TMT ratio and incubating for 3 h and 500 rpm at room temperature. A second incubation step was performed overnight at 37°C with agitation at 500 rpm and the labeling reaction was stopped by incubation at 80°C for 1 h and further addition of hydroxylamine at 2% with incubation at 37°C for 30 min. Labeling efficiency was evaluated by calculating the proportion of free C-terminal lysines (K) among all lysines in identified peptidofoms, after variable modification search of TMT at K (see Spectral Data Processing section) and evaluating the distribution of positional abundance of residues before cleavage (by trypsin or others), depending on the N-terminal modification (Figure S2).

## 2.4 | Protein digestion and high pH reversed-phase chromatography

Protein samples were reconstituted in 1.2% phosphoric acid and processed using S-Trap columns (Protifi) for digestion and clean-up following the manufacturer's instructions. In order to increase proteome coverage, sample fractionation was carried out on an Agilent 1100 HPLC system fitted with a XBRIDGE Peptide BEH C18, Agilent column (3.5  $\mu\text{m}$ , 130  $\text{\AA}$ , 1  $\times$  150 mm), operating at 42  $\mu\text{L}/\text{min}$ . The elution solvent system consisted of buffer A (10 mM ammonium

formate) and buffer B (10 mM ammonium formate in 70% acetonitrile); both solvents were adjusted to pH 10 with ammonium hydroxide solution. 160  $\mu\text{g}$  of lyophilized peptide mixtures were reconstituted in 200  $\mu\text{L}$  in buffer A and injected in the column at 30  $\mu\text{L}/\text{min}$  for 4 min. Prior to sample fractionation, the column was equilibrated for 60 min at a constant gradient of 20% buffer B. Subsequently, the fractionation gradient started as follows: 20% to 60% buffer B for 60 min, after which the collection of the fractions was stopped. Finally, the gradient was then increased to 100% buffer B within 2 min and held at this level for 1 min before being ramped down to 1%. A total of 48 fractions were collected and pooled into 22 samples following a concatenation strategy [19]. Fractions were dried by vacuum centrifugation and stored at  $-80^\circ\text{C}$  until tandem mass spectrometry measurement.

## 2.5 | Liquid chromatography—tandem mass spectrometry (LC-MS/MS)

Peptide fractions and commercial HeLa digests (Thermo scientific) were solubilized in 0.1% trifluoroacetic acid (TFA) 1  $\mu\text{g}$  of each peptide as together with 200 fmol of iRT peptides were injected into an EASY-nLC 1200 UHPLC system equipped with  $\mu\text{PAC}$  C18 Trapping column (5  $\mu\text{m}$  pillar diameter, 2.5  $\mu\text{m}$  inter-pillar distance, 18  $\mu\text{m}$  pillar length, 10 mm bed length, 100 to 200  $\text{\AA}$  pore sizes) and an  $\mu\text{PAC}$  C18 nano-LC analytical column (5  $\mu\text{m}$  pillar diameter, 2.5  $\mu\text{m}$  inter-pillar distance, 18  $\mu\text{m}$  pillar length, 50 cm bed length, 100–200  $\text{\AA}$  pore sizes). A multistep gradient of 6% to 55% buffer B (0.1% v/v formic acid, 80% v/v acetonitrile) in buffer A (0.1% v/v formic acid) was used for separation at 500 nL/min flow rate, followed by washing (100% buffer B) and reconditioning of the column to 6% B. The chromatography system was coupled to a Q-Exactive Plus mass spectrometer via a Nanospray Flex Ion source. Mass spectrometry (MS) data was obtained as previously described [20]. For pilot experiments on the effect of protein extraction, 300 ng of dried peptides, along with 200 fmol internal retention time standards (iRTs), were loaded into Evtotips (EV2001, Evosep) following the manufacturer's instructions. The Evosep One chromatography system was operated using a predefined 30SPD Method, utilizing a performance column (EV1137, 15 cm  $\times$  150  $\mu\text{m}$ ) packed with ReproSil-Pur C18 1.5  $\mu\text{m}$  beads. Mobile phases A and B consisted of 0.1 vol% formic acid in water and 0.1 vol% formic acid in acetonitrile (ACN), respectively. The liquid chromatography system was coupled online to a timsTOF Flex mass spectrometer (Bruker) using a CaptiveSpray nano-electrospray ion source. Measurements were conducted employing Data-Independent Acquisition (DIA)-PASEF [21].

## 2.6 | Spectral data processing

Peptide identification was conducted by the MSFragger search engine [12] based on the EBI mouse canonical proteome (release 2021\_04) using Arg-C protease specificity with only 1 enzymatic termini to

generate in-silico digested peptides (semi-specific search), noting that TMT labeling at the protein level would render the lysine unavailable for cleavage by trypsin. Precursor candidates were selected with a mass tolerance of  $-20/20$  ppm and a fragment mass tolerance of 20 ppm. Peptide N-terminal acetylation and peptide N-terminal TMT labeling were set as variable modifications. TMT labeling at K and carbamidomethylation of C were set as fixed modifications. For labeling efficiency testing, a search with TMT labeling at K as variable was performed. MSBooster [17] was used for deep-learning based predictions of retention time and spectra. Predicted features were used by Percolator [15] for post-processing scoring and FDR control. Results were summarized and filtered by Philosopher [22]. Spectral data obtained from HeLa samples were searched using FragPipe as described above, with the following differences: Peptides were identified against the EBI human canonical proteome (release 2021\_04), including reshuffled sequences. Reshuffled sequences were generated using DBToolkit [23], by generating randomized versions of all human sequences, and appending them to the canonical fasta. Trypsin/P protease specificity was used with only 1 enzymatic terminus. Only peptide N-terminal acetylation was allowed as variable modification. MaxQuant (v 2.4.0.0) search was performed with a precursor mass tolerance of  $-20/20$  ppm and fragment mass tolerance of 20 ppm. Trypsin/P was set as protease specificity with Semi-specific mode. N-terminal acetyl was set as variable modification. In all cases, the FDR cut-off for peptide identification was set to 1%.

### 3 | RESULTS

#### 3.1 | FragPipe excels at uncovering semi-tryptic peptides in shotgun data indicative of endogenous protein cleavage

As semi-specific peptide-to-spectrum (PSM) searches are time-consuming and of higher complexity than fully-specific searches, we initially tested the performance of two widely used proteomics search algorithms, FragPipe [12, 22] and MaxQuant [24], for uncovering protein cleavage products in shotgun proteomics data. An initial analysis of a single tryptic proteome profile of a human cell line (HeLa, Thermo Fisher) allows to evaluate the performance of both approaches for semi-specific searches. FragPipe outperforms MaxQuant in identifying peptides, proteins, semi-specific peptides, and N-terminal modifications, while only needing 10% of the execution time. FragPipe detects 34,848 PSMs (+63%), 29,166 peptides (+60%), and 4421 proteins (+31%), compared to MaxQuant (Figure S3A, Table S1). Notably, FragPipe excels in identifying semi-specific peptides, detecting 2645, whereas MaxQuant identifies 1598. Moreover, FragPipe demonstrates superior performance in the identification of N-terminal modifications, with 1685 acetylated N-termini detected, while MaxQuant identifies 369 (Figure S3B). We used an entrapment strategy consisting of a semi-specific search based on a sequence fasta file appended with absent *Escherichia coli* and shuffled sequences, to assess the potential of false identifications. Less than 0.1% of identified peptides mapped

to unexpected sequences (FragPipe detects 6 false hits vs. MaxQuant with 29). None of the unexpected identifications were semi-specific peptides (Figure S3C). This illustrates the speed and sensitivity of MSFragger for peptide-spectrum-matching together with Percolator and MSBooster for probabilistic rescoring. Taken together, this underscores the preferred usage of FragPipe for data (re-)analysis and the exploration of proteolytic events in shotgun proteomics data.

#### 3.2 | Description of terminomics analysis framework

In brief, we describe *TermineR* as a data analysis approach to extract information related to proteolytic processing (Figure 1B) from shotgun spectral data arising from typical sample preparation without prior biochemical enrichment, based on semi-specific searches. After the annotation of peptide identifications based on their specificity and biochemical modifications, we integrate sample annotations for differential abundance analyses. We match the identified proteolytic products to annotated processing information from UniProt and generate visualizations to assist in the biological interpretation of the results.

We organized *TermineR* into an R package, containing functions to process the semi-specific search results. Initially, PSM or precursor level search results are preprocessed by “adapter” functions, that summarize information on peptide sequence, N-terminal modification (currently TMT, acetylation, and dimethylation), and their respective standardized quantitative information per sample. Currently available adapters can process results from FragPipe, Spectronaut, and DIANN. Then we annotate the identified peptides based on their sequence specificity, infer potential cleavage sites and areas for semi-specific peptides and map them to processing sites annotated in UniProt, with an error range of  $\pm 4$  residues (i.e., Processing of Signal Peptide, Removal of Initiator Methionine, Propeptide, and intact Open Reading Frame (ORF). Intact ORF is defined as the N-terminal boundary of the protein sequence in the fasta file). The *annotate\_neo\_termini* function is used for this purpose, which requires the user to supply (1) the data frame generated from the adapter functions, (2) the location of the fasta file of identified protein sequences, and (3) the definition of the experimental cleavage specificity. The user can define the expected residues (i.e., for trypsin K or R) and the sense of the specificity (C for C-terminal specificity and N for N-terminal specificity; that is, trypsin has K and R as expected residues and sense “C”; TrypN has K and R as expected residues and sense “N”) (Figure 1C).

Subsequently, we perform differential abundance analysis per each peptide [25] and use the peptide annotation to apply multiple-testing correction specifically on proteolytic products, based on independent hypothesis weighting of the peptides defined as proteolytic products [26]. Finally, we can visualize the differentially abundant proteolytic products using a heatmap of the residues relative to the position of cleavage sites, using the sequences of the cleavage areas obtained from the *annotate\_neo\_termini* function (Figure 1C).

This workflow is aimed at providing researchers with a framework for extracting information regarding the subfraction of proteolytic products from any shotgun proteomics dataset. As such, the location of truncated labeled N-termini allows us to study patterns of intrinsic proteolytic activity and the location of the N-terminal acetylation can offer clues about shifted or non-canonical translation initiation sites. The use of this framework also allows for the identification of C-terminally truncated peptides, which are usually missed by traditional terminomics approaches [27]. However, in the present showcase study, we focus on N-terminal processing.

### 3.3 | Experimental design for terminomic analysis of murine polycystic kidney disease (PKD) model

We established an experimental setup to showcase the capabilities of our data analysis approach for extracting information related to proteolytic processing starting from a shotgun proteomics sample preparation without biochemical enrichment for N-terminal peptides. A cohort of fresh frozen tissue samples derived from a mouse model of PKD comprised of 6 wild type mice (WT/Healthy) and 5 *Pkd1<sup>fl/fl</sup>;Pax8rtTA;TetOCre* (KO) mice that developed enlarged cystic kidneys. Protein extraction was performed under denaturing conditions and proteins were labeled using TMT16plex isobaric tags followed by tryptic digestion, peptide clean-up and high pH reverse phase fractionation. Peptide fractions were measured using liquid chromatography followed by mass spectrometry in a Q-Exactive plus. Executing the TMT labeling before tryptic digestion was aimed to specifically tag native and neo-N-termini and differentiate them from the tryptically-generated N-termini. To determine labeling efficiency, we performed an additional search step outside of the *TerminerR* approach: N-terminal and lysine TMT-labeling was set to variable and a semi-specific tryptic search with up to 2 missed cleavages was executed. In total, 24,597 peptidofoms were identified, 16,540 of these carrying at least a lysine (K) in their sequence. Trypsin cleaves after free (unlabeled) lysine residues, hence we used the proportion of peptide C-terminal free lysines among all lysines identified, as an estimate of labeling efficiency. From a total of 19,098 lysines in these peptidofoms, 2425 were free C-terminal ones, indicating a labeling efficiency of at least 87% and a digestion efficiency showing 1.4% missed cleavages (Figure S2A). Moreover, the positional clustering highlights that most TMT-labeled N-termini are found within the first quarter of the annotated protein sequence, and most acetylated N-termini are associated to the start of the intact ORF (Figure S2B). These tests showed that this dataset was suitable for a showcase application of our *TerminerR* approach. Furthermore, the comparative quantitation of the TMT reporter ions allowed for the differential abundance analysis of proteolytic products. Although the *TerminerR* approach can be applied on any kind of shotgun proteomics dataset, the use of this setup allowed us to evaluate and showcase the capabilities of semi-specific searches with FragPipe for the study of proteolytic processing (Figure 2A).

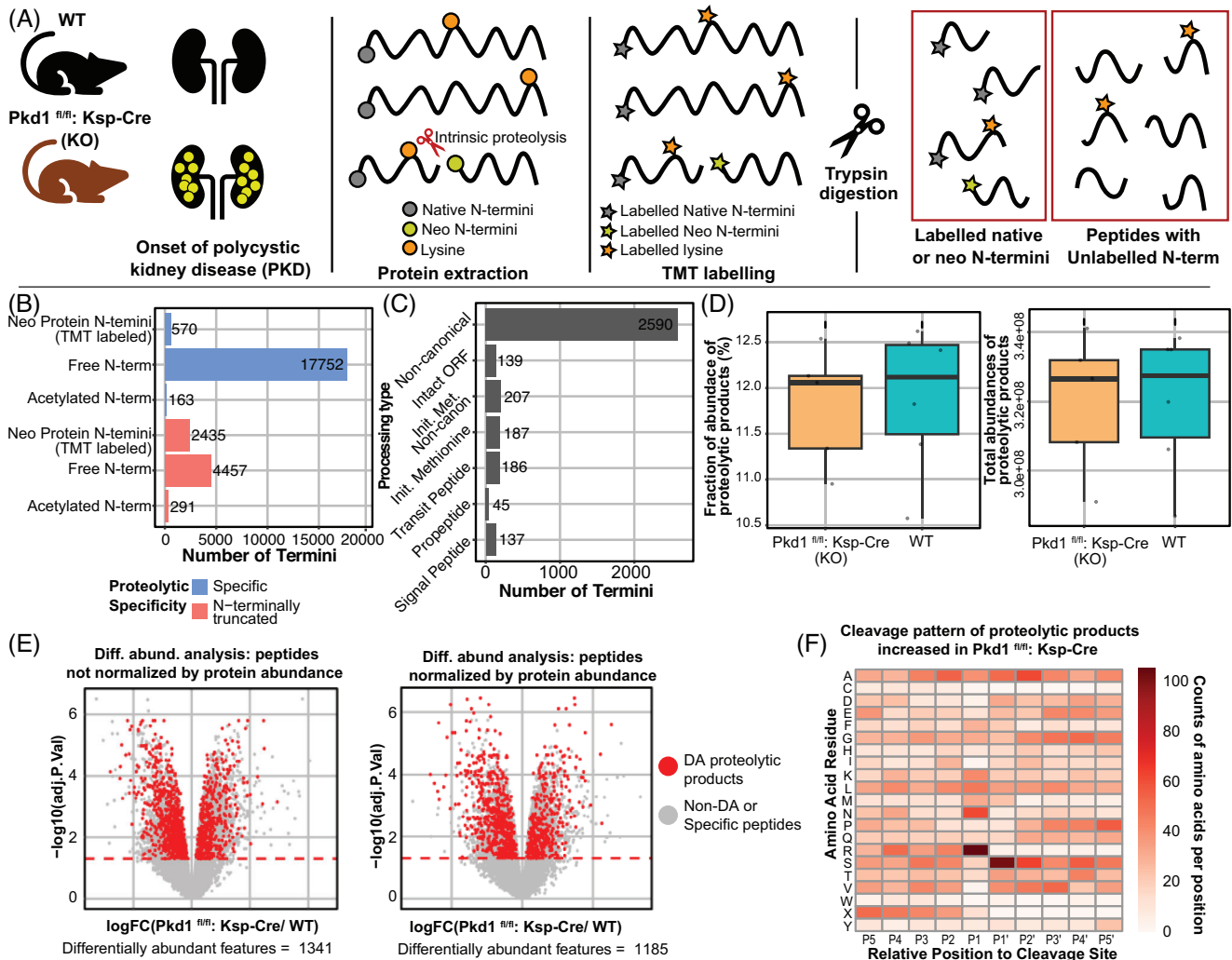
### 3.4 | *TerminerR* annotation and differential statistics of proteolytic products in murine polycystic kidney disease (PKD)

The peptide-centric analysis allowed us to identify 25,668 peptides and 5379 proteins in total (Figure 2B). Further data processing allowed us to classify 3005 peptides as semi-specific and/or bearing an N-terminal TMT arising from protein-level TMT labeling. An additional 454 acetylated N-termini were found, yielding an N-terminal coverage in excess of 3400 N-termini. This number is in the range of a recently published N-terminomic study based on biochemical enrichment [28]. Next, the annotation of proteolytic products mapped them to UniProt processing annotation. The cleavage sites of proteolytic products are mapped against their processing annotations, with aims to characterize potential canonical processing in the dataset. Most identified proteolytic products are not matched against any UniProt processing site (non-canonical, 2590 peptides, 76% of total proteolytic products), followed by clipping of initiator methionine with 394 (11%), and transit peptide, signal peptides and propeptides accounting for 368 peptides (10%).

### 3.5 | Differential abundance statistics

The subsequent quantitative analyses show a differential behavior in terms of proteolytic processing between the *Pkd1<sup>fl/fl</sup>;Ksp-Cre* and WT mice, albeit without a globally different abundance of proteolytic products between conditions (Figure 2D). We then executed *limma*-moderated *t*-tests on the abundance matrix of all quantified peptides while focusing the multiple-testing *p*-value correction on the subset of peptides labeled as proteolytic products. Based on our whole-proteome analysis, it is possible to evaluate the differential abundance of proteolytic products with and without normalization by protein abundance. This allows to differentiate between features whose differential abundance is associated with protein abundance from those proteolytic products with increased abundance due to increased proteolysis (Figure 2E). We then focused our quantitative analyses on proteolytic products normalized by protein abundance. Overall, we observed 2740 neo-N-termini consistently quantified and protein-normalized in all samples from which we identify 1185 differentially abundant proteolytic events after protein-abundance normalization, considered as differential proteolysis (Table S2), and 1341 without protein-normalized peptide abundances (Table S3). Among those truncated peptides representing differential proteolysis, 524 were found upregulated in *Pkd1<sup>fl/fl</sup>;Ksp-Cre* and 661 downregulated (Figure 2E).

For differential statistics of these products, we apply multiple-testing correction on those features considered as proteolytic products. This “feature-specific” multiple-testing correction approach is based on independent hypothesis testing [26], to decrease the penalization for multiple testing at the peptide level, for increased sensitivity. We consider that the nature of the peptide (specific, semi-specific, N-terminal acetylation, etc.) is independent of the moderated *t*-test



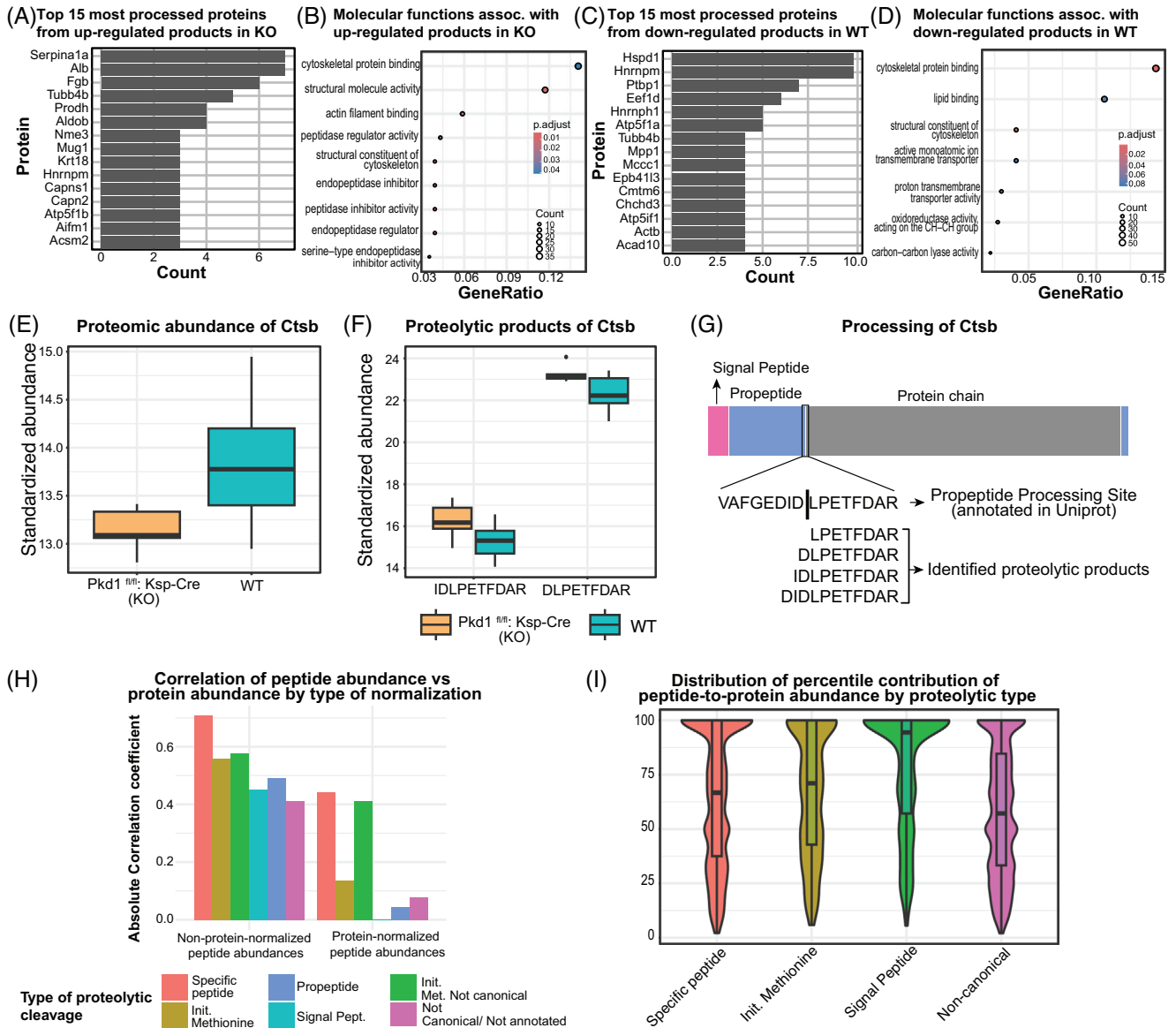
**FIGURE 2** The application of the *TermineR* approach identifies patterns of differential proteolysis in a model of polycystic kidney disease. (A) Description of example model for polycystic kidney disease (PKD) and TMT labeling at the protein level. (B) Identified peptides in PKD model, differentiating them by specificity and N-terminal modification. TMT-labeled N-termini are considered canonical or Neo-N-termini. (C) Uniprot-based annotation of proteolytic products. (D) Boxplots showing (1) the percentage of proteolytic products from total abundance (sum of reporter ion abundances of proteolytic products/total reporter ion abundances of all peptides) and (2) the total summed abundance of reporter ion intensities for semi-specific peptides, between conditions. (E) Differential abundance of proteolytic products between Pkd1<sup>fl/fl</sup>:Ksp-Cre and WT mouse kidneys; comparing types of peptide abundance normalization. Differentially abundant (DA) peptides are defined as those with adjusted *p*-values < 0.05 after *limma* statistics. (F) Heatmap showing proteolytic patterns of upregulated proteolytic products. We see a high number of increased cleavages with Ser and Cys at P1'.

applied to evaluate differential abundance and therefore it is feasible to apply multiple testing correction specifically on interesting features.

Of note, the quantitative accuracy of TMT reporter ions for differential abundance analysis was confirmed in our system with the use of a titration dataset of spiked-in *E. coli* protein extracts in known proportion onto a HEK background proteome, showing consistent differential abundances regardless of peptide specificity (Figure S4A,B, Table S4) [29]. We evaluated the impact of purity correction on the reporter ion abundances, resulting in an almost negligible effect (Figure S4C); likely because most peptides feature isobaric purity values bigger than 0.9 (Figure S4D).

### 3.6 | Visualization of cleavage patterns from differentially abundant proteolytic products

Differentially abundant proteolytic products can be explored for cleavage patterns by reconstructing its representative cleavage area. We use a heatmap to represent the count of amino acids in particular relative positions to the cleavage area for all differentially abundant features, with either increased or decreased abundance against the compared baseline (WT in this example) (Figure 2F). We can use this visualization to look for proteolytic patterns of known proteases.



**FIGURE 3** Analysis of proteolytic products in shotgun proteomics data offers novel biological insights. (A) Top 15 proteins by the number of associated up-regulated proteolytic products in KO mice (“highly processed” proteins). (B) Molecular Functions associated with highly processed proteins in KO mice. (C) Top 15 proteins by the number of associated up-regulated proteolytic products in WT mice. (D) Molecular Functions associated with highly processed proteins in WT mice. (E) Standardized abundances of *Ctsb* protein between experimental conditions. (F) Standardized abundances of proteolytic products/N-terminally TMT-tagged peptides of *Ctsb* between experimental conditions. (G) Diagram of the annotated propeptide processing sites of *Ctsb*, together with identified proteolytic products. (H) Bar plot showing Pearson’s correlation values between peptide and protein abundance between different processing types and abundance normalization. (I) Percentile contribution to protein abundance by identified, quantified, and annotated peptides by their associated proteolytic condition.

### 3.7 | Biological interpretation

Further analysis allows us to pinpoint motifs associated with increased proteolytic processing and/or degradation. In the context of PDK, the top five proteins associated with up-regulated proteolytic products are Serpina1a, Albumin, Fibrinogen (Fbg), Tubulin (Tubb4b), and Prodh (Figure 3A). A good proportion of up-regulated products can be related to serum proteins such as Albumin and Fibrinogen, and over-representation analyses relate increased cleavages to cytoskeletal protein binding and peptidase activity (Figure 3B) (Figure 3A; Table

S5). Down-regulated proteolytic products were associated mainly with Hspd1, Hnnp1, Ptbp1, Atp5f1a, and Hnnp1, (Figure 3C) showing a defined metabolic fingerprint (Figure 3D). Notably, the differential proteomic analysis shows an important downregulation of mitochondrial proteins in *Pkd1<sup>fl/fl</sup>:Ksp-Cre* kidneys (Table S5), a feature that has been described in PKD [30].

Previous studies have shown that loss Pkd1 can impair lysosomal activity potentially due to the disruption in the processing of lysosomal proteases, including Cathepsin B (*Ctsb*) [31]. In line with this, the proteomics analysis show a significantly decreased abundance of



Ctsb in Pkd1<sup>fl/fl</sup>:Ksp-Cre (Figure 3E, Table S5), with a tendency to increased abundance of proteolytic products from an annotated region of propeptide processing (Figure 3F). We identified four proteolytic products of Ctsb, showcasing a “ragging effect” (consecutive cleavage of individual residues in a directional manner) around the area of propeptide activation (Figure 3G), an observation that has been previously reported in an independent study [32]. This is in line with the postulated disruption of the proteolytic processing of Ctsb in PDK, and its effects on impaired lysosomal activity.

### 3.8 | Assessing the association between proteolytic products and protein abundance: The potential to evaluate cleavage stoichiometry

One of the advantages of performing terminomics analysis starting from a classical shotgun proteomics dataset, is the possibility to associate the abundance of proteolytic products with the abundance of their associated proteins (Table S5). This allows us to directly evaluate if the differential quantitation of a certain proteolytic product is due to differential processing or differential protein expression, and therefore act as a proxy for stoichiometric evaluation of substrate processing. In our present approach, we compare the correlation between peptide and protein intensities, with or without correction by protein abundance. It is noticeable that normalization by protein abundance promotes an exchange in the behavior of the correlation coefficients, which particularly affects non-canonical cleavage products (Figure 3H).

It is then possible to pinpoint proteolytic products whose quantitation proportionally differs from the abundance of the proteins they are associated with, and whose differential abundance between conditions can be related to differential proteolysis. Visualizing the normalized fold changes of differentially abundant proteolytic products against the fold changes of their associated proteins, help us understand the relationship between differential proteolysis and protein abundance for any given substrate. To illustrate this, we explored the proteolytic processing of cadherin-16 (Cdh16), Beta-2 microglobulin (B2m), and Fibronectin (Fn1) (Figure S5C,D). We observed that all these proteins showed differential proteolytic processing; with at least a fraction of their proteolytic products showing a differential quantitative behavior compared to protein abundances.

Finally, to evaluate the stoichiometric relationship between proteolytic products and protein, we calculated the percentile contribution of each unique peptide to their associated protein abundance. We aimed to see differences in median percentile contributions to protein abundance between different types of peptide processing. Of note: most specific peptides would show a diverse contribution to general protein abundance. This is expected if we assume that tryptic peptides should, in general, contribute similarly to protein abundance when accounting for sequence bias and protein coverage. This behavior is mostly mimicked by peptides risen from canonical processing. On the other hand, non-canonical proteolytic products tend to show a smaller contribution to their protein abundances, suggesting that these would tend to be in sub-stoichiometric abundance levels (Figure 3I).

## 4 | DISCUSSION

The current framework presents itself as a data analysis approach for tackling the challenge of extracting meaningful information on proteolytic processing from large-scale explorative proteomics data. Our bioinformatic approach adds to the portfolio of N-terminomic techniques that employ biochemical enrichment of protein N-termini. Our N-terminomic coverage based on TMT N-terminal-labeling is slightly reduced but still in the range of biochemical N-terminomic enrichment [9, 33–35]. We showcase our workflow using a TMT-labeling approach at the protein level, to offer a supporting layer of evidence for the presence of protein N-termini, in particular cleavage events. For this non-enriched experimental setup, other N-terminal labeling methods such as dimethylation (e.g.,  $\Delta = 28.031/36.075$  Da for the prototypical light/heavy setup) can be considered. Our adapter function for processing of corresponding FragPipe output is suited to extract this modification information provided the user includes the appropriate variable modifications in the search. Yet, the usage of TMT-like multiplexing or label-free quantitation is presently outpacing the usage of light/heavy 2-plex experiments (i.e., SILAC or dimethyl). For this reason—and to reduce complexity—it has remained beyond the scope of the present tool to fully integrate pre-configured workflows of this kind. Yet, label-free (DDA or DIA) proteomic experiments can integrate semi-specific searches and use our framework for the annotation of cleavage events without the prior requirement of N-terminal labeling.

Currently established methods for the large-scale probing of proteolytic activity (yielding neo-N- or C-termini) rely on the biochemical enrichment of N- or C-terminal peptides before their analysis via LC-MS/MS, for the study of differential proteolysis between biological conditions. These methods need specialized sample processing that is not widely standardized in most proteomics labs. These can also lead to sample loss and make it difficult to contextualize the differential abundance of proteolytic products with general protein abundance. We acknowledge that experimental approaches for the enrichment of proteolytic products offer deeper insights of cleavage events and increased identifications of native ORF N-terminal peptides (those mapping to the N-terminal boundary of the fasta protein sequence). In comparison with an experiment such as the one showcased in this study (non-enriched N-terminal protein-labeling), proteolytic products and ORF N-terminal peptides appear in comparably low abundance in relation to all tryptic peptides in the sample produced by experimental digestion. Nevertheless, the *TermineR* data analysis approach can be used to annotate and perform differential abundance analysis in both dataset types, coming from N-terminal enrichment or not.

Noticing the great sensitivity of currently available mass spectrometers and bioinformatics tools for peptide identification, we propose this data analysis approach as an alternative for the explorative analysis of proteolysis from classic shotgun proteomics experiments, that allows to probe both N- and C-terminal cleavage events within the wider context of protein abundance, in the same sample. We consider this approach to be particularly useful (but not limited to) for data reanalysis. Starting from spectral data acquired from any classic

shotgun experiment, a semi-specific search can identify peptides that arise from non-experimental cleavage events, and researchers can apply our framework for the annotation of these events as potential proteolytic products and perform differential abundance analyses. The potential of this application has been showcased in the analysis of differential proteolysis from shotgun proteomics data on recurrent glioblastoma multiforme [36].

## 5 | CONCLUSIONS

Shotgun proteomics LC-MS/MS data is a rich source for probing N- and C-terminal modifications, such as proteolytic processing. Here we present a robust data analysis approach wrapped as an R package, for the processing of semi-specific search results, enabling data preparation, annotation and dedicated statistical testing based on feature-specific multiple testing correction. Although described and initially developed to be applied to search results from the FragPipe bioinformatics suit, our approach for peptide annotation, differential abundance analysis and visualization can be easily adapted to process data from other data types and search engines (i.e., DIA-NN and Spectronaut), including label-free data. The R package with example usability code, data and documentation can be accessed in its latest version via GitHub.

## ACKNOWLEDGMENTS

O.S. acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, projects 446058856, 466359513, 444936968, 405351425, 431336276, 431984000 (SFB 1453 “NephGen”), 441891347 (SFB 1479 “OncoEscape”), 322977937 (GRK 2344 “MeInBio”), the ERA PerMed Program (BMBF, 01KU1916, 01KU1915A), the German Consortium for Translational Cancer Research (project Impro-Rec), the MatrixCode Research Group, FRIAS, Freiburg, the investBW program BW1\_1198/03, the ERA TransCan program (project 01KT2201, “PREDICO”), the BMBF KMUi program (project 13GW0603E, project ESTHER), and the BMBF Cluster4Future program (nanodiag). O.S. and P.F.H. acknowledge funding by DFG project 423813989 (GRK 2606 “ProtPath”). M.C.C. is a member of the GRK2344 (MeInBio), funded by the German Research Foundation—322977937-GRK234.

Open access funding enabled and organized by Projekt DEAL.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

## DATA AVAILABILITY STATEMENT

Spectral data from PKD model and titration data and intermediary parameters and search results can be accessed via Massive under the identifier MSV000094661. The *TermineR* pipeline can be accessed via Github as an R package with example data and documentation for its implementation (<https://github.com/MiguelCos/TerminER>), included necessary code to reproduce the analysis showcased here. We encourage users to use our issue tracker on Github (<https://github.com/MiguelCos/TerminER/issues>) when facing issues or for requests for specific features.

## ORCID

Miguel Cosenza-Contreras  <https://orcid.org/0000-0002-9727-3229>

Pitter F. Huesgen  <https://orcid.org/0000-0002-0335-2242>

Oliver Schilling  <https://orcid.org/0000-0001-7678-7653>

## REFERENCES

1. Klein, T., Eckhard, U., Dufour, A., Solis, N., & Overall, C. M. (2018). Proteolytic cleavage—mechanisms, function, and “omic” approaches for a near-ubiquitous posttranslational modification. *Chemical Reviews*, 118(3), 1137–1168. <https://doi.org/10.1021/acs.chemrev.7b00120>
2. Shahinian, J. H., Rog-Zielinska, E. A., Schlimpert, M., Mayer, B., Tholen, S., Kammerer, B., Biniossek, M. L., Beyersdorf, F., Schilling, O., & Siepe, M. (2020). Impact of left ventricular assist device therapy on the cardiac proteome and metabolome composition in ischemic cardiomyopathy. *Artificial Organs*, 44(3), 257–267. <https://doi.org/10.1111/aor.13566>
3. Eatemadi, A., Aiyelabegan, H. T., Negahdari, B., Mazlomi, M. A., Daraee, H., Daraee, N., Eatemadi, R., & Sadroddiny, E. (2017). Role of protease and protease inhibitors in cancer pathogenesis and treatment. *Biomedicine and Pharmacotherapy*, 86, 221–231. <https://doi.org/10.1016/j.biopha.2016.12.021>
4. Demir, F., Troldborg, A., Thiel, S., Lassé, M., Huesgen, P. F., Tomas, N. M., Wiech, T., & Rinschen, M. M. (2021). Proteolysis and inflammation of the kidney glomerulus. *Cell and Tissue Research*, 385(2), 489–500. <https://doi.org/10.1007/s00441-021-03433-8>
5. Lai, Z. W., Petrer, A., & Schilling, O. (2015). Protein amino-terminal modifications and proteomic approaches for N-terminal profiling. *Current Opinion in Chemical Biology*, 24, 71–79. <https://doi.org/10.1016/j.cbpa.2014.10.026>
6. Aebersold, R., & Mann, M. (2016). Mass-spectrometric exploration of proteome structure and function. *Nature*, 537(7620), 347–355. <https://doi.org/10.1038/nature19949>
7. Mintoo, M., Chakravarty, A., & Tilwala, R. (2021). N-terminomics strategies for protease substrates profiling. *Molecules (Basel, Switzerland)*, 26(15), 4699. <https://doi.org/10.3390/molecules26154699>
8. Wang, L., Main, K., Wang, H., Julien, O., & Dufour, A. (2021). Biochemical tools for tracking proteolysis. *Journal of Proteome Research*, 20(12), 5264–5279. <https://doi.org/10.1021/acs.jproteome.1c00289>
9. Demir, F., Kizhakkedathu, J. N., Rinschen, M. M., & Huesgen, P. F. (2021). MANTI: automated annotation of protein N-termini for rapid interpretation of N-terminome data sets. *Analytical Chemistry*, 93(13), 5596–5605. <https://doi.org/10.1021/acs.analchem.1c00310>
10. Fahrner, M., Kook, L., Fröhlich, K., Biniossek, M. L., & Schilling, O. (2021). A systematic evaluation of semispecific peptide search parameter enables identification of previously undescribed n-terminal peptides and conserved proteolytic processing in cancer cell lines. *Proteomes*, 9(2), 1–13. <https://doi.org/10.3390/proteomes9020026>
11. Bell, P. A., & Overall, C. M. (2023). No substrate left behind—mining of shotgun proteomics datasets rescues evidence of proteolysis by SARS-CoV-2 3CLpro main protease. *International Journal of Molecular Sciences*, 24(10), 8723. <https://doi.org/10.3390/ijms24108723>
12. Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D., & Nesvizhskii, A. I. (2017). MSFragger: Ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nature Methods*, 14(5), 513–520. <https://doi.org/10.1038/nmeth.4256>
13. Solntsev, S. K., Shortreed, M. R., Frey, B. L., & Smith, L. M. (2018). Enhanced Global Post-translational Modification Discovery with MetaMorpheus. *Journal of Proteome Research*, 17(5), 1844–1851. <https://doi.org/10.1021/acs.jproteome.7b00873>

14. Ma, K., Vitek, O., & Nesvizhskii, A. I. (2012). A statistical model-building perspective to identification of MS/MS spectra with PeptideProphet. *BMC Bioinformatics [Electronic Resource]*, 13(16):S1, 1–17. <https://doi.org/10.1186/1471-2105-13-S16-S1>
15. The, M., MacCoss, M. J., Noble, W. S., & Käll, L. (2016). Fast and accurate protein false discovery rates on large-scale proteomics data sets with percolator 3.0. *Journal of the American Society for Mass Spectrometry*, 27(11), 1719–1727. <https://doi.org/10.1007/s13361-016-1460-7>
16. Yu, S.-H., Kiriakidou, P., & Cox, J. (2020). Isobaric matching between runs and novel PSM-level normalization in MaxQuant strongly improve reporter ion-based quantification. *Journal of Proteome Research*, 19(10), 3954–3954. <https://doi.org/10.1021/acs.jproteome.0c00209>
17. Yang, K. L., Yu, F., Teo, G. C., Li, K., Demichev, V., Ralsner, M., & Nesvizhskii, A. I. (2023). MSBooster: Improving peptide identification rates using deep learning-based features. *Nature Communications*, 14(1), 4539. <https://doi.org/10.1038/s41467-023-40129-9>
18. Viau, A., Bienaimé, F., Lukas, K., Todkar, A. P., Knoll, M., Yakulov, T. A., Hofherr, A., Kretz, O., Helmstädter, M., Reichardt, W., Braeg, S., Aschman, T., Merkle, A., Pfeifer, D., Dumit, V. I., Gubler, M., Nitschke, R., Huber, T. B., Terzi, F., & Kuehn, E. W. (2018). Cilia-localized LKB 1 regulates chemokine signaling, macrophage recruitment, and tissue homeostasis in the kidney. *The EMBO Journal*, 37(15), e98615. <https://doi.org/10.15252/embj.201798615>
19. Yang, F., Shen, Y., Camp, D. G., & Smith, R. D. (2012). High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. *Expert Review of Proteomics*, 9(2), 129–134. <https://doi.org/10.1586/epr.12.15>
20. Baumert, H. M., Metzger, E., Fahrner, M., George, J., Thomas, R. K., Schilling, O., & Schüle, R. (2020). Depletion of histone methyltransferase KMT9 inhibits lung cancer cell proliferation by inducing non-apoptotic cell death. *Cancer Cell International*, 20(1), 1–13. <https://doi.org/10.1186/s12935-020-1141-2>
21. Meier, F., Brunner, A.-D., Frank, M., Ha, A., Bludau, I., Voytik, E., Kaspar-Schoenefeld, S., Lubeck, M., Raether, O., Bache, N., Aebersold, R., Collins, B. C., Röst, H. L., & Mann, M. (2020). diaPASEF: Parallel accumulation–serial fragmentation combined with data-independent acquisition. *Nature Methods*, 17(12), 1229–1236. <https://doi.org/10.1038/s41592-020-00998-0>
22. da Veiga Leprevost, F., Haynes, S. E., Avtonomov, D. M., Chang, H.-Y., Shanmugam, A. K., Mellacheruvu, D., Kong, A. T., & Nesvizhskii, A. I. (2020). Philosopher: A versatile toolkit for shotgun proteomics data analysis. *Nature Methods*, 17(9), 869–870. <https://doi.org/10.1038/s41592-020-0912-y>
23. Martens, L., Vandekerckhove, J., & Gevaert, K. (2005). DBToolKit: Processing protein databases for peptide-centric proteomics. *Bioinformatics (Oxford, England)*, 21(17), 3584–3585. <https://doi.org/10.1093/bioinformatics/bti588>
24. Tyanova, S., Temu, T., & Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature Protocols*, 11(12), 2301–2319. <https://doi.org/10.1038/nprot.2016.136>
25. Smyth Walter, G. K., & Hall, E. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *In Statistical Applications in Genetics and Molecular Biology*, 3(1), Article 3. <https://doi.org/10.2202/1544-6115.1027>
26. Bourgon, R., Gentleman, R., & Huber, W. (2010). Independent filtering increases detection power for high-throughput experiments. *Proceedings of the National Academy of Sciences of the United States of America*, 107(21), 9546–9551. <https://doi.org/10.1073/pnas.0914005107>
27. Schilling, O., Barré, O., Huesgen, P. F., & Overall, C. M. (2010). Proteome-wide analysis of protein carboxy termini: C terminomics. *Nature Methods*, 7(7), 508–511. Article 7. <https://doi.org/10.1038/nmeth.1467>
28. Hanna, R., Rozenberg, A., Saied, L., Ben-Yosef, D., Lavy, T., & Kleifeld, O. (2023). In-depth characterization of apoptosis N-terminome reveals a link between caspase-3 cleavage and posttranslational N-terminal acetylation. *Molecular & Cellular Proteomics*, 22(7), 100584. <https://doi.org/10.1016/j.mcpro.2023.100584>
29. Bernhard, P., Feilen, T., Rogg, M., Fröhlich, K., Cosenza-Contreras, M., Hause, F., Schell, C., & Schilling, O. (2022). Proteome alterations during clonal isolation of established human pancreatic cancer cell lines. *Cellular and Molecular Life Sciences*, 79(11), 561. <https://doi.org/10.1007/s00018-022-04584-9>
30. Padovano, V., Podrini, C., Boletta, A., & Caplan, M. J. (2018). Metabolism and mitochondria in polycystic kidney disease research and therapy. *Nature Reviews Nephrology*, 14(11), 678–687. <https://doi.org/10.1038/s41581-018-0051-1>
31. Peintner, L., Venkatraman, A., Waeldin, A., Hofherr, A., Busch, T., Voronov, A., Viau, A., Kuehn, E. W., Köttgen, M., & Borner, C. (2021). Loss of PKD1/polycystin-1 impairs lysosomal activity in a CAPN (calpain)-dependent manner. *Autophagy*, 17(9), 2384–2400. <https://doi.org/10.1080/15548627.2020.1826716>
32. Tholen, S., Biniossek, M. L., Geßler, A.-L., Müller, S., Weißer, J., Kizhakkedathu, J. N., Reinheckel, T., & Schilling, O. (2011). Contribution of cathepsin L to secretome composition and cleavage pattern of mouse embryonic fibroblasts. *Biological Chemistry*, 392(11), 961–971. <https://doi.org/10.1515/BC.2011.162>
33. Chang, C.-H., Chang, H.-Y., Rappsilber, J., & Ishihama, Y. (2021). Isolation of acetylated and unmodified protein N-terminal peptides by strong cation exchange chromatographic separation of TrypN-digested peptides. *Molecular & Cellular Proteomics*, 20, 100003. <https://doi.org/10.1074/mcp.TIR120.002148>
34. Tsumagari, K., Chang, C.-H., & Ishihama, Y. (2021). Exploring the landscape of ectodomain shedding by quantitative protein terminomics. *Science*, 24(4), 102259. <https://doi.org/10.1016/j.isci.2021.102259>
35. Weng, S. S. H., Demir, F., Ergin, E. K., Dirnberger, S., Uzozie, A., Tuscher, D., Nierves, L., Tsui, J., Huesgen, P. F., & Lange, P. F. (2019). Sensitive determination of proteolytic proteoforms in limited microscale proteome samples. *Molecular & Cellular Proteomics*, 18(11), 2335–2347. <https://doi.org/10.1074/mcp.TIR119.001560>
36. Cosenza-Contreras, M., Schäfer, A., Sing, J., Cook, L., Stillger, M. N., Chen, C.-Y., Villacorta Hidalgo, J., Pinter, N., Meyer, L., Werner, T., Bug, D., Haberl, Z., Kübeck, O., Zhao, K., Stei, S., Gafencu, A. V., Ionita, R., Brehar, F. M., Ferrer-Lozano, J., & Schilling, O. (2023). Proteometabolomics of initial and recurrent glioblastoma highlights an increased immune cell signature with altered lipid metabolism. *Neuro-Oncology*, 26(3), 488–502. <https://doi.org/10.1093/neuonc/noad208>

## SUPPORTING INFORMATION

Additional supporting information may be found online <https://doi.org/10.1002/pmic.202300491> in the Supporting Information section at the end of the article.

**How to cite this article:** Cosenza-Contreras, M., Seredynska, A., Vogele, D., Pinter, N., Brombacher, E., Cueto, R. F., Dinh, T.-L. J., Bernhard, P., Rogg, M., Liu, J., Willems, P., Stael, S., Huesgen, P., Kuehn, E. W., Kreutz, C., Schell, C., & Schilling, O. (2024). TermineR: extracting information on endogenous proteolytic processing from shotgun proteomics data. *Proteomics*, 24, e2300491. <https://doi.org/10.1002/pmic.202300491>