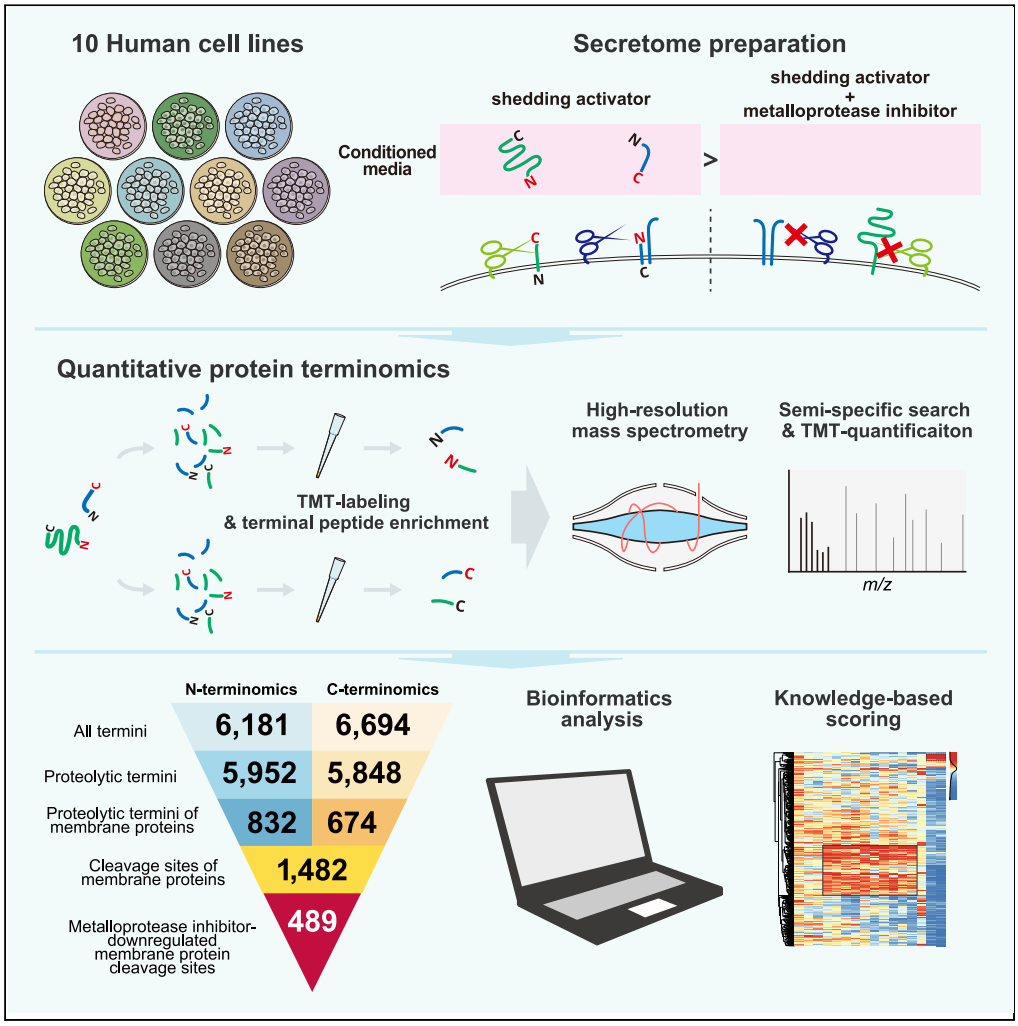
iScience

CellPress
OPEN ACCESS

Article

Exploring the landscape of ectodomain shedding by quantitative protein terminomics



Kazuya Tsumagari, Chih-Hsiang Chang, Yasushi Ishihama

yishiham@pharm.kyoto-u.ac.jp

HIGHLIGHTS

Secretomes across 10 human cell lines were investigated by protein terminomics

Cell-type-specific maps of shedding cleavage sites were generated

Most of the cleavage sites were identified in a cell-type-specific manner

Knowledge-based scoring enabled prediction of responsible sheddases

Tsumagari et al., iScience *24*, 102259
April 23, 2021 © 2021 The Author(s).
https://doi.org/10.1016/j.isci.2021.102259

iScience

CellPress
OPEN ACCESS

Article

Exploring the landscape of ectodomain shedding by quantitative protein terminomics

Kazuya Tsumagari,[1,2] Chih-Hsiang Chang,[1] and Yasushi Ishihama[1,3,4,*]

SUMMARY

Ectodomain shedding is a proteolytic process that regulates the levels and functions of membrane proteins. Dysregulated shedding is linked to severe diseases, including cancer and Alzheimer's disease. However, the exact cleavage sites of shedding substrates remain largely unknown. Here, we explore the landscape of ectodomain shedding by generating large-scale, cell-type-specific maps of shedding cleavage sites. By means of N- and C-terminal peptide enrichment and quantitative mass spectrometry, we quantified protein termini in the culture media of 10 human cell lines and identified 489 cleavage sites on 163 membrane proteins whose proteolytic terminal fragments are downregulated in the presence of a broad-spectrum metalloprotease inhibitor. A major fraction of the presented cleavage sites was identified in a cell-type-specific manner and mapped onto receptors, cell adhesion molecules, and protein kinases and phosphatases. We confidently identified 86 cleavage sites as metalloprotease substrates by means of knowledge-based scoring.

INTRODUCTION

Membrane proteins have critical physiological roles, and their abundance and functions are tightly controlled through multiple mechanisms, including ectodomain shedding (shedding), which is a form of limited proteolysis that liberates the extracellular domain of membrane proteins (Huovila et al., 2005; Lichtenthaler et al., 2018; Weber and Saftig, 2012). Shedding contributes to cellular interactions with the environment by releasing active cytokines, growth factors or other mediators from their membrane-bound precursors, or conversely, by reducing the levels of receptors and adhesion proteins at the cell surface (Huovila et al., 2005; Weber and Saftig, 2012). Shedding often triggers further proteolysis within the transmembrane domain – known as regulated intramembrane proteolysis (RIP) – thereby releasing the intracellular domain into the cytoplasmic region. These processes enable bidirectional signal transduction (Beard et al., 2019; Reiss and Saftig, 2009; Weber and Saftig, 2012).

A protease involved in shedding is specifically referred to as a sheddase. Distinct sheddases may cleave the same substrate at different sites, generating multiple proteoforms that may exhibit different biological functions and activities (Niedermaier and Huesgen, 2019). For instance, the difference of cleavage sites of amyloid precursor protein (APP) by a disintegrin and metalloprotease 10 (ADAM10) and by beta-site APP cleaving enzyme 1 (BACE1) is critical for the pathogenesis of Alzheimer's disease (Chow et al., 2010). Therefore, it is important to identify cleavage sites in order to achieve a precise understanding of the physiological roles of shedding.

Discovery and analysis of large numbers of proteoforms is the exclusive domain of liquid chromatography/tandem mass spectrometry (LC/MS/MS)-based proteomics (Aebersold et al., 2018; Olsen and Mann, 2013). For the identification of shedding substrates, it is preferable to investigate the proteins secreted into the cell culture media, rather than those remaining in the membrane, because the secretome is much less complex than the membrane proteome. Additionally, the fragments remaining in the membrane are often unstable due to further proteolytic events such as RIP and lysosomal degradation (Güner and Lichtenthaler, 2020; Merilahti and Elenius, 2019). By using established hydrazide chemistry, 18 metalloprotease substrates in the culture media were identified (Tsumagari et al., 2017); most of the membrane proteins are glycosylated (Apweiler et al., 1999). However, this approach did not lead to precise identification of the cleavage sites. In fact, knowledge of the cleavage sites of shedding substrates is limited, despite its importance. In standard shotgun proteomics, only peptides generated in accordance with the specificity of the employed

[1]Department of Molecular and Cellular BioAnalysis, Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto 606-8501, Japan

[2]Eisai-Keio Innovation Laboratory for Dementia, Center for Integrated Medical Research, Keio University School of Medicine, Tokyo 160-8582, Japan

[3]Department of Proteomics and Drug Discovery, Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto 606-8501, Japan

[4]Lead contact

*Correspondence: yishiham@pharm.kyoto-u.ac.jp

https://doi.org/10.1016/j.isci.2021.102259

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

digestive enzymes at both N-and C-termini are considered. On the other hand, it is essential for identifying shedding-originated peptide termini to employ semi-specific searches in which peptides may result from specific cleavage by the employed digestive enzymes at only one terminus, while the other terminus may be generated by a non-specific cleavage event, such as shedding (Niedermaier and Huesgen, 2019). Thus, terminal peptide enrichment from secreted proteins coupled with semi-specific search should be employed in order to efficiently identify shedding cleavage sites.

Several methods have been reported for protein terminomics. Prudova et al. employed a method for N-terminal peptide enrichment called TAILS (terminal amine isotopic labeling of substrates) (Kleifeld et al., 2010), and identified 201 and 19 cleavage products formed by recombinant MMP-2 and MMP-9, respectively (Prudova et al., 2010). An alternative strategy is to utilize an engineered enzyme to selectively label free protein N-termini with biotin-tags that enable positive selection of N-terminal peptides (Mahrus et al., 2008; Weeks and Wells, 2018). However, there are still obstacles to the large-scale detection of cleavage sites generated by endogenous sheddases in living cells. First, terminal peptide enrichment strategies generally consist of multiple steps including some chemical reactions, and therefore relatively large amounts of samples are needed. Second, the protein amounts in the culture media are much smaller than those obtained from cell lysates, and thus it is hard to prepare sufficiently large samples for terminal peptide enrichment. Finally, C-terminomics specifically remains a challenging task because the enrichment efficiency is greatly inferior to that of the N-terminal counterpart due to the difficulty in chemically modifying the carboxy group (Niedermaier and Huesgen, 2019).

Recently, our group developed a simple and rapid methodology for N-terminal peptide enrichment, in which N-terminal peptides of TrypN digests are isolated by strong cation exchange (SCX) chromatography at low pH (Chang et al., 2021). Earlier reports have demonstrated that C-terminal peptides generated by trypsin digestion are eluted first in SCX, together with acetylated N-terminal peptides (Alpert et al., 2010; Gauci et al., 2009; Helbig et al., 2010), and we discovered here this feature enables the identification and quantification of C-terminal peptides on a comparable scale to that of the N-terminal counterparts with the same amount of input (Figures S1A and S1B). Here, we applied these sensitive terminomics methodologies to conduct the first large-scale study of shedding cleavage sites targeted by endogenous metalloproteases, which have emerged as the major sheddase family. We quantitatively identify putative metalloprotease-regulated ectodomain shedding cleavage sites based on the results of broad-spectrum metalloprotease inhibitor treatment, and provide an overview of their positional and functional landscape.

RESULTS

Sensitive terminomics workflow enables efficient identification and reproducible quantification of N- and C-termini in the secretome

Protein shedding can be transiently upregulated by certain agonists, such as phorbol 12-myristate 13-acetate (PMA), activating mainly ADAM17 (Huovila et al., 2005). To achieve large-scale detection of membrane protein cleavage sites by members of the metalloprotease family, we employed a quantitative terminomics workflow consisting of shedding activation with PMA, broad-spectrum metalloprotease inhibitor BB-94 (batimastat) treatment, SCX-based terminal peptide enrichment (Alpert et al., 2010; Chang et al., 2021; Helbig et al., 2010), TMT labeling and nanoLC/MS/MS measurement (Figures S1A–S1D).

We investigated secreted protein fractions across a panel of ten human cell lines stimulated with PMA following BB-94 treatment (Figures 1A and 1B). Samples were prepared in triplicate for each condition. The digest of 10 μg protein per replicate was subjected to terminal peptide enrichment: thus, in total only 1.2 mg of protein was utilized in the whole of this study. Terminal peptide enrichment was performed by SCX-StageTip (Adachi et al., 2016; Rappsilber et al., 2007). Briefly, for N-terminal peptide enrichment, TrypN is first employed to generate protein N-terminal peptides without Lys or Arg. Then, the N-terminal peptides can be separated from the internal peptides by SCX due to their weaker retention (Figures S1A and S1B). Similarly, LysC and trypsin are used to generate C-terminal peptides without Lys or Arg for isolation of C-terminal peptides (Figures S1A and S1B). Among trypsin/LysC-digested peptides, C-terminal peptides and acetylated N-terminal peptides have similar charge distributions (Figure S1A), and therefore these terminal peptides are isolated together (Figure S1E). We eliminated acetylated N-terminal peptides generated by trypsin/LysC and TrypN from the identification list in this study, since peptides cleaved by sheddase should have unmodified termini. Note that N-terminal peptides were TMT-labeled after enrichment, since the TMT tag affects the enrichment efficiency (Figure 1C). Also note that acetylated N-terminal

CellPress
OPEN ACCESS

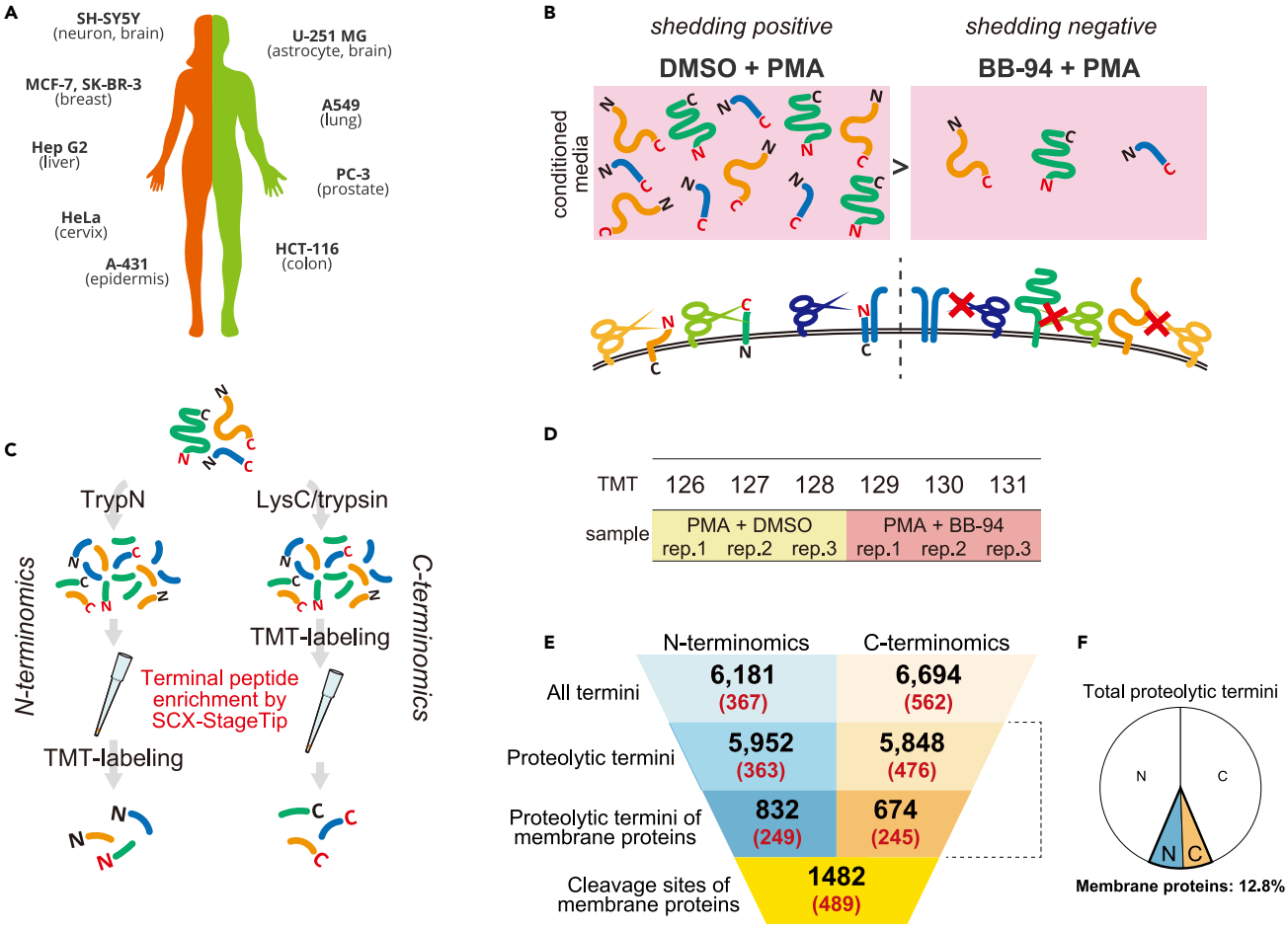


Figure 1. Experimental design and quantitative terminomics workflow for large-scale analysis of cleavage sites by metalloproteases

(A) List of investigated cultured cell lines (10 human cancer cell lines).

(B) Experimental design. Cells were treated with DMSO or BB-94 for 1 hr, followed by PMA treatment for 1 hr. Triplicate samples were prepared for each condition.

(C) Workflow of sample preparation. Proteins were digested with either TrypN or LysC and trypsin. Note that N-terminal peptides were TMT-labeled after terminal peptide enrichment, while C-terminal peptide enrichment was performed following TMT-labeling. See Figures S1C and S1D for details.

(D) Triplicate shedding-positive samples (PMA + DMSO) and shedding-negative samples (PMA + BB-94) were labeled as shown with TMT channels and form a 6-plexed TMT.

(E) Summary of the number of quantified termini and resulting cleavage sites. The parenthesized numbers show those significantly downregulated by BB-94. All termini: the total numbers of N- or C- termini (native protein termini and proteolytic termini). Proteolytic termini: the numbers of N- or C-termini that are presumably generated by endogenous proteases. Proteolytic termini on membrane proteins: the numbers of proteolytic N- or C-termini that are mapped on membrane proteins. Cleavage sites of membrane proteins: the total number of cleavage sites presented by proteolytic N- and C-termini on membrane proteins.

(F) Pie chart depicting the ratio of proteolytic termini mapped on membrane proteins in total proteolytic termini. The white areas show non-membrane proteins, and the areas highlighted with colors show membrane proteins. N, N-termini; C, C-termini.

peptides generated by TrypN cannot be quantified in our workflow because these peptides do not have TMT-reactive sites (Figure S1A). For C-terminomics, the flow-through and the 0.5% TFA-eluted fraction were separately collected and subjected to LC/MS/MS (Figure S1D). C-Terminal peptides were identified with the selectivity of 38% and 13% on average for the flow-through and the 0.5% TFA-eluted fractions, respectively (Figure S1E). Triplicates of controls and of BB-94-treated samples were multiplexed using 6-plexed TMT reagents (Figure 1D). We analyzed the TMT-labeled terminal peptides by high-resolution Orbitrap mass spectrometry considering a wide search space in semi-specific search mode (Niedermaier and Huesgen, 2019). Semi-specific search was performed using the Andromeda search engine on MaxQuant (Cox and Mann, 2008; Cox et al., 2011) against the SwissProt human protein database, including isoform sequences. TMT-reporter intensities were normalized by the trimmed mean of M values (TMM) method

(Robinson and Oshlack, 2010). We excluded peptides with missed cleavages for further analyses in order to simplify the relationship between the cleavage sites and the corresponding peptides.

Our strategy led to the quantification of 6,181 N-termini and 6,694 C-termini in total (Figure 1 E and Tables S1 and S2). In order to find peptides containing cleavage sites targeted by endogenous proteases, we extracted peptides with termini not cleaved by the spiked protease used for sample preparation, so-called "semi-specific peptides" (Figure S1F). Such terminal peptides would be derived from endogenous proteolysis events, such as shedding. In this study, we refer to such termini, presumably generated by endogenous proteolysis, as "proteolytic termini" and the terminal peptides presenting the proteolytic termini as "proteolytic peptides" (Figure S1F). Finally, our dataset included 5,952 proteolytic N-termini and 5,848 proteolytic C-termini (Figure 1E and Tables S1 and S2). Of identified proteolytic termini, 832 N- and 674 C-termini were derived from membrane proteins (14.0% and 11.5% of the respective proteolytic terminome), which were defined using the UniProtKB Keywords "transmembrane" or "GPI-anchor". Figure S1G shows the protein yields for each cell line to perform N, C-terminomics, and Figures S2 and S3 summarize the reproducibility in triplicate analyses. For individual conditions, there was excellent reproducibility in the peak intensity of identified peptides, with Pearson correlation coefficients in the ranges of $R > 0.93$ (0.98 on average) for N-terminomics and $R > 0.97$ for C-terminomics (0.99 on average). We created volcano plots with truncation at the false discovery rate of 0.05 and artificial within-groups variance (S_0) of 0.1 (default parameters) using Perseus (Tyanova et al., 2016) (Figures S4 and S5), which yielded 363 proteolytic N-termini and 476 proteolytic C-termini that were significantly downregulated upon BB-94 treatment (Figure 1E). Importantly, these included 249 proteolytic N-termini (68.6% of downregulated proteolytic N-termini) and 245 proteolytic C-termini (51.5% of downregulated proteolytic C-termini) mapped on membrane proteins, affording a total of 489 cleavage sites (Table S3). Note that five downregulated membrane protein cleavage sites were presented by both proteolytic N- and C-terminal peptides. As our results show (Figure 1F), proteolytic terminal peptides derived from membrane proteins account for a small fraction of the total proteolytic terminal peptides, which is one reason why they are so difficult to detect with MS. Our workflow enabled efficient identification and quantification of both proteolytic N- and C-termini of membrane proteins, highlighting a major strength of our methodology for achieving deep analysis of membrane protein cleavage sites with a limited amount of material.

BB-94 selectively downregulates proteolytic termini of membrane proteins in the secretome

The number of downregulated termini varied among investigated cell lines (Figure 2A). U-251 MG cells have the highest number of downregulated termini, while MCF-7 has the lowest number. First, we examined the subcellular distribution of downregulated proteins using Gene Ontology (GO) term enrichment analysis by DAVID (v6.8; https://david.ncifcrf.gov/) (Huang et al., 2009). We found strong enrichment for terms that evoke membrane proteins, such as plasma membrane and cell surface (Figure 2B). In addition, in most samples, except the N-terminome of SK-BR-3, proteolytic termini of membrane proteins were downregulated by BB-94, while the abundances of proteolytic termini of non-membrane proteins were not changed (Figures 2C and 2D). These results indicate that BB-94 broadly and selectively targets the cleavage of membrane proteins, which would be useful for studying metalloprotease-dependent shedding. In addition, the provided cleavage site information in living cells should facilitate an understanding of the mechanism of action of BB-94.

Topological analysis of downregulated membrane proteins and positional analysis of the cleavage sites

We found that some of the downregulated membrane protein cleavage sites are localized at or around signal peptide cleavage sites (Figure S6A). Note that the sequences around the signal peptide cleavage sites showed overrepresentation of AxA at P3-P1 (three to one amino acids upstream of the cleavage sites) (Figure S6B), which is the motif of canonical signal peptidases (Paetzel et al., 2002). Because it is not clear whether these cleavages occurred as metalloprotease-dependent shedding, signal peptidases or the combined events, we excluded 56 sites within ± 5 amino acids from the signal peptide cleavage sites for further analyses in this study. Topological analysis of the remaining 433 sites based on the UniProtKB (https://www.uniprot.org/) annotation revealed that single-pass type I membrane proteins, which have their N-termini in the extracellular region, are the majority (378 sites, 87.3%), followed by GPI-anchor proteins (21 sites, 4.8%), single-pass type II membrane proteins, which have the C-termini in the extracellular region (17 sites, 3.9%), and multi-pass membrane proteins, which span the membrane more than once (10 sites, 2.3%) (Figure 3A).

CellPress
OPEN ACCESS

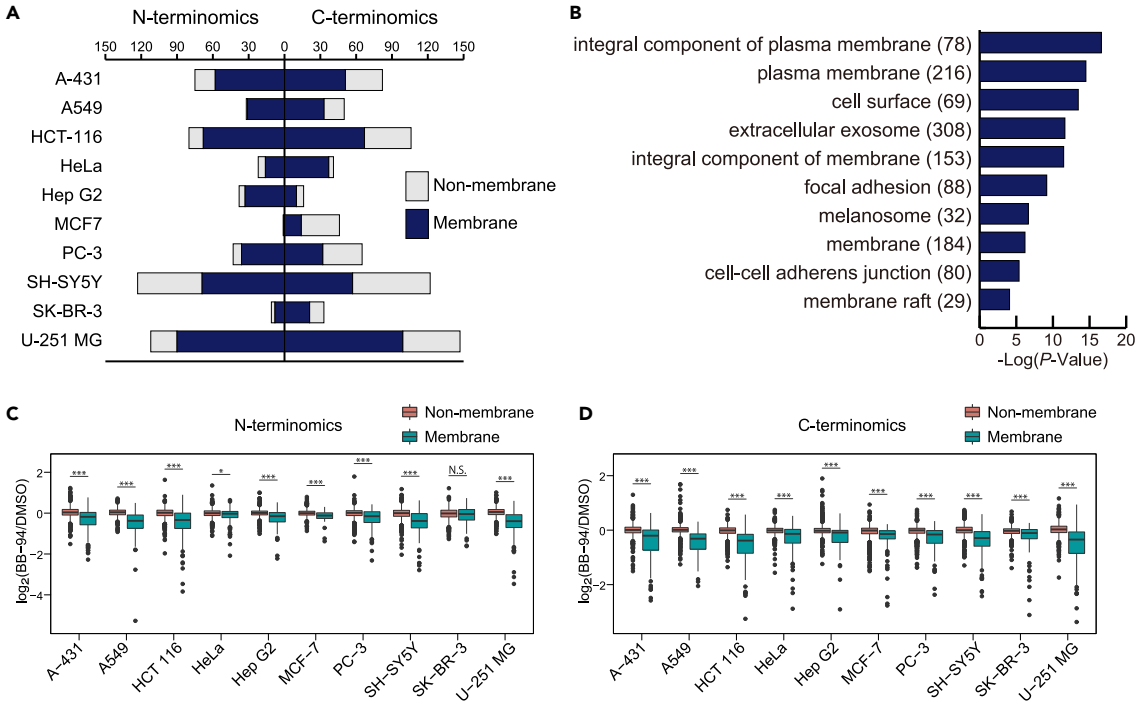


Figure 2. Overview of the BB-94-downregulated proteolytic terminome in the supernatant

(A) Distribution of the number of downregulated proteolytic termini mapped on membrane proteins and non-membrane proteins in the respective cell lines.

(B) Enrichment analysis on GO term cellular components performed by DAVID (v8.0). The top ten significant terms are shown, with Bonferroni-adjusted p values. The parenthesized numbers show the number of proteins.

(C and D) The $\log_2$-transformed ratios (BB-94/DMSO) of proteolytic termini are compared between membrane proteins and non-membrane proteins in the respective cell lines. The p values were calculated with the Wilcoxon rank-sum test and a Bonferroni adjustment was applied. ***, $p < 0.005$; **, $p < 0.001$; *, $p < 0.05$; N.S., not significant.

We examined the cleavage positions of single-pass type I and type II membrane proteins and GPI-anchor proteins. Of the 416 cleavage sites, 394 sites (94.7%) were distributed on the extracellular domain, confirming that BB-94 targets "ectodomain shedding" (Figure 3B). Notably, we found that metalloprotease-regulated proteolysis occurred in close proximity to the cell surface: 174 sites (44.2%) were within 200 amino acids from the transmembrane domain or GPI-anchor site, which evidently suggests that large portions of the extracellular domains are released by shedding. This trend is consistent with previous findings on individual shedding substrates (Hinkle et al., 2004; Shirakabe et al., 2011; Zheng et al., 2004), and thus validates our results. On the other hand, 15 BB-94-downregulated sites were mapped within the transmembrane domain (Figure 3B), including three APP cleavage sites considered to be cleaved by γ-secretase (corresponding to amyloid β 37, 38, and 40) (Takami et al., 2009) (Figure S6C), and this suggests that our dataset includes cleavage sites generated downstream of metalloprotease-dependent shedding (Brown et al., 2000; Edwards et al., 2008; Reiss and Saftig, 2009). Interestingly, these downregulated cleavage sites inside the transmembrane domain were likely to be located at the extracellular side (Figure 3C). γ-Secretase cleaves APP in a stepwise manner, in which APP is first cleaved at the membrane-cytoplasm boundary, followed by successive tri- or tetrapeptide trimming (Takami et al., 2009). Consequently, the resulting terminus may be positioned at the extracellular side. Notably, in addition to APP, we identified five downregulated proteins (ALCAM in A431 and HCT-116; GLG1 in HCT-116; PIGR in A431; SDC1 in A431, A-549, HCT-116, HeLa, PC-3, SH-SY5Y, SK-BR-3, and U-251 MG; SDC4 in HCT-116) with cleavage sites on the extracellular domain and inside the transmembrane domain in the same cell line (Table S3). These proteins could have roles in bidirectional signal transduction across the cell membrane (Beard et al., 2019; Reiss and Saftig, 2009; Weber and Saftig, 2012). It is noteworthy that we found only 7 sites (2 sites by N-terminomics and 5 sites by C-terminomics) were mapped on the cytoplasmic domain (Figure 3B), which might originate from dead cells.

We identified 10 BB-94-downregulated cleavage sites on 7 multi-pass membrane proteins (Table S3), and investigated their localization using the TMHMM Server (v2.0; http://www.cbs.dtu.dk/services/TMHMM-2.0/). We found 7 sites on 4 proteins (PLP2, GPR126, IGSF1, CELSR2) were mapped on their extracellular

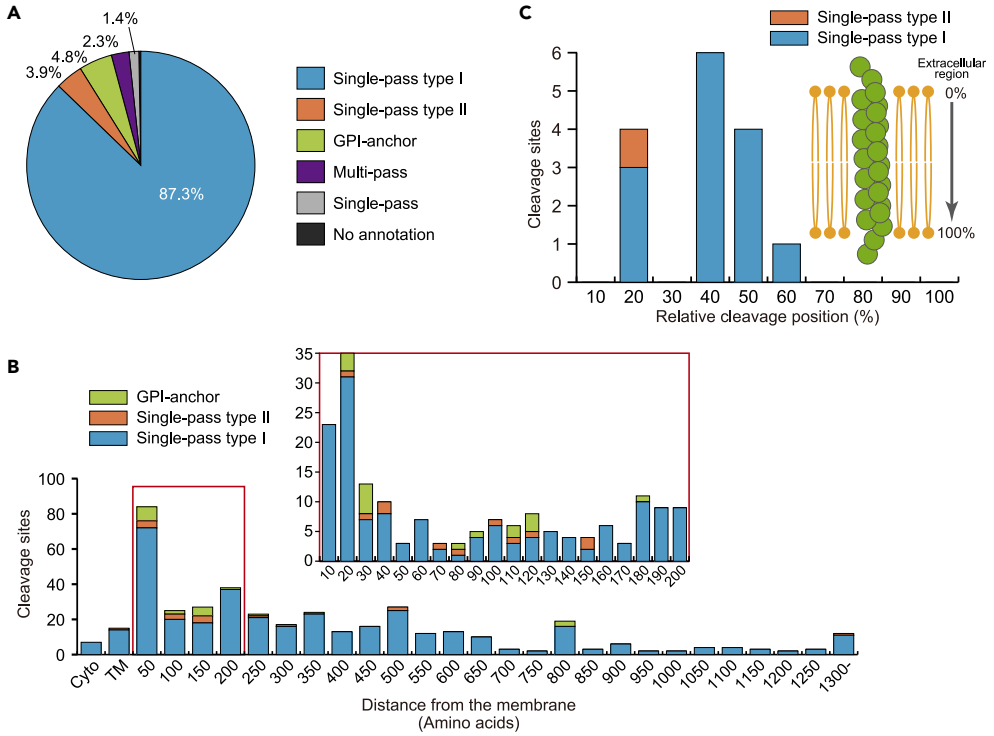


Figure 3. Topological analysis of downregulated membrane proteins and positional analysis of the cleavage sites

(A) Percentages of the various topologies of membrane proteins identified with downregulated cleavage sites. The categories of membrane proteins follow the notation in UniProt.

(B) The distribution of the distance (number of amino acids) from the transmembrane domain or GPI-anchor site to the cleavage site is depicted for single-pass type-I and type-II membrane proteins and GPI-anchored proteins. The section surrounded with a red box is shown in detail above. TM: cleavage sites within the transmembrane domain. Cyto: cleavage sites within the cytoplasmic domain.

(C) Relative cleavage positions of the cleavage sites within the transmembrane domain of single-pass type I and type II membrane proteins. Values represent the length (number of amino acids) from the extracellular region to the cleavage site, divided by the total length (number of amino acids) of the transmembrane domain.

domains (Table S3). Among them, 6 sites were localized at the extracellular region with only one transmembrane domain, as exemplified by CELSR2 F815↓L816 (Figure S6D). These single cleavages can lead to ectodomain shedding. On the other hand, the cleavage site of G-protein coupled receptor 126 (F840↓T841) is located on the extracellular region between two membrane-spanning domains (Figure S6E). This site was identified with C-terminal proteolytic peptides, indicating that this ectodomain was shed as a result of additional cleavage at the N-terminal side of F840. Of the cleavage sites on multi-pass membrane proteins, 3 sites were mapped on their cytoplasmic domains. As mentioned above, these sites might be artifacts. Overall, in our terminomics, most of the BB-94-downregulated sites on membrane proteins (416 sites (97.7%) out of 426 membrane protein cleavage sites) were mapped on their extracellular domains or transmembrane domains, supporting the reliability of our data set.

Functional characterization of ectodomain shedding

We functionally characterized the downregulated cleavage sites on the extracellular domain as metalloprotease-regulated shedding cleavage sites (394 cleavage sites on 119 proteins; Figure 3B). DAVID enrichment analysis (Huang et al., 2009) for GO term biological processes revealed significant enrichment for cell adhesion proteins (Figure 4), which is consistent with a previous study (Tsumagari et al., 2017). In addition, we found enrichment of proteins related to cell migration, including CD44, whose shedding is important for CD44-dependent cell migration (Nagano and Saya, 2004). Analysis of GO terms of molecular function showed that shedding targets central components of signal transduction, such as receptors and protein tyrosine kinases/phosphatases. Importantly, we observed that shedding liberates a large portion of the extracellular domain (Figure 3B), which is the main functional region of membrane proteins, on a

CellPress
OPEN ACCESS

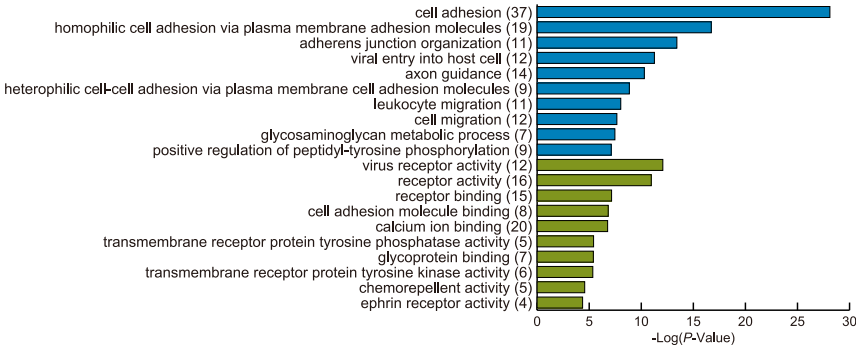


Figure 4. Functional analysis of metalloprotease-regulated shedding substrates

DAVID enrichment analysis of metalloprotease-regulated shedding substrates according to GO term biological processes (blue) and molecular functions (green). The top ten significant terms are shown, with Bonferroni-adjusted p values. The parenthesized numbers show the numbers of proteins.

proteomic scale. Thus, overall, our results indicate that shedding regulates fundamental biological events by modulating membrane protein functions.

Cell-type-specificity of shedding

Differently processed proteoforms can exhibit distinct physiological properties (Chow et al., 2010; Hu et al., 2006; La Marca et al., 2011; Willem et al., 2006). Whereas most BB-94-downregulated membrane proteins contain one cleavage site, interestingly no less than 65 cleavage sites were identified on neuronal cell adhesion molecule (NrCAM) (Figure 5A). This indicates that proteolysis can produce an impressive number of proteoforms and may critically contribute to the complexity of the human proteome. A large population of the identified shedding substrates was cell-type-specific rather than globally identified, at both the cleavage site and protein levels (Figures 5B and 5C), suggesting that shedding occurs depending on the individual functions of the cells. Sixty proteins were found in one cell line, while only one protein, 36-kDa vesicular integral-membrane protein VIP36 coded by the LMAN2 gene, was found commonly in all investigated cell lines. VIP36 is a validated shedding substrate (Shirakabe et al., 2011). Intriguingly, the VIP36 protein was cleaved at the same site (F298↓L299) across all analyzed cell lines. In the previous study, western blotting in combination with amino acid substitution at expected cleavage sites failed to identify the precise cleavage site of VIP36 (Shirakabe et al., 2011), but the site we identified here is positioned just within the expected region (Figure 5D), which again validates our result and underscores the strength of our strategy. VIP36 is a lectin domain-containing transmembrane protein that functions as a cargo receptor transporting glycoproteins. While shedding of the VIP36 protein plays a critical role in phagocytosis in Raw 264.7 cells (Shirakabe et al., 2011), the role of VIP36 shedding in other cell types is unknown.

We found 36 proteins were cleaved at different sites depending on the cell lines (Figure 5E). Of these, 5 proteins have proteolytic peptides upregulated upon BB-94 treatment, indicating these peptides might be generated by multiple proteolytic events (see later discussion for syndecan-1).

We identified cell-line-specific shedding substrate proteins in nine cell lines, except for MCF-7. Astrocytoma-derived U-251 MG cells yielded the highest number of substrates (Figure 5F), and 24 proteins were uniquely identified in U-251 MG cells. To investigate the functions of these U-251 MG cell-specific substrates, we investigated protein-protein interactions of these substrate proteins using the STRING database (v11.0; https://string-db.org/) (Szklarczyk et al., 2019), which resulted in significantly more interactions than would be expected for a random protein set of similar size with a p value of 6.13×10^{-11} (calculated by STRING) (Figure 5G). Interestingly, enrichment analysis revealed that these U-251 MG cell-specific substrates are related to typical functions of neurons, such as axon development and axon guidance, rather than those of astrocytes (Figure 5H). This may indicate that shedding contributes to cellular communication across different cell types in the central nervous system.

Evaluation of metalloprotease-regulated shedding cleavage sites using PWM scoring

As mentioned above, our results include cleavage sites downstream of the initial metalloprotease cleavages. Therefore, to identify the direct cleavage sites by metalloproteases with greater confidence, we

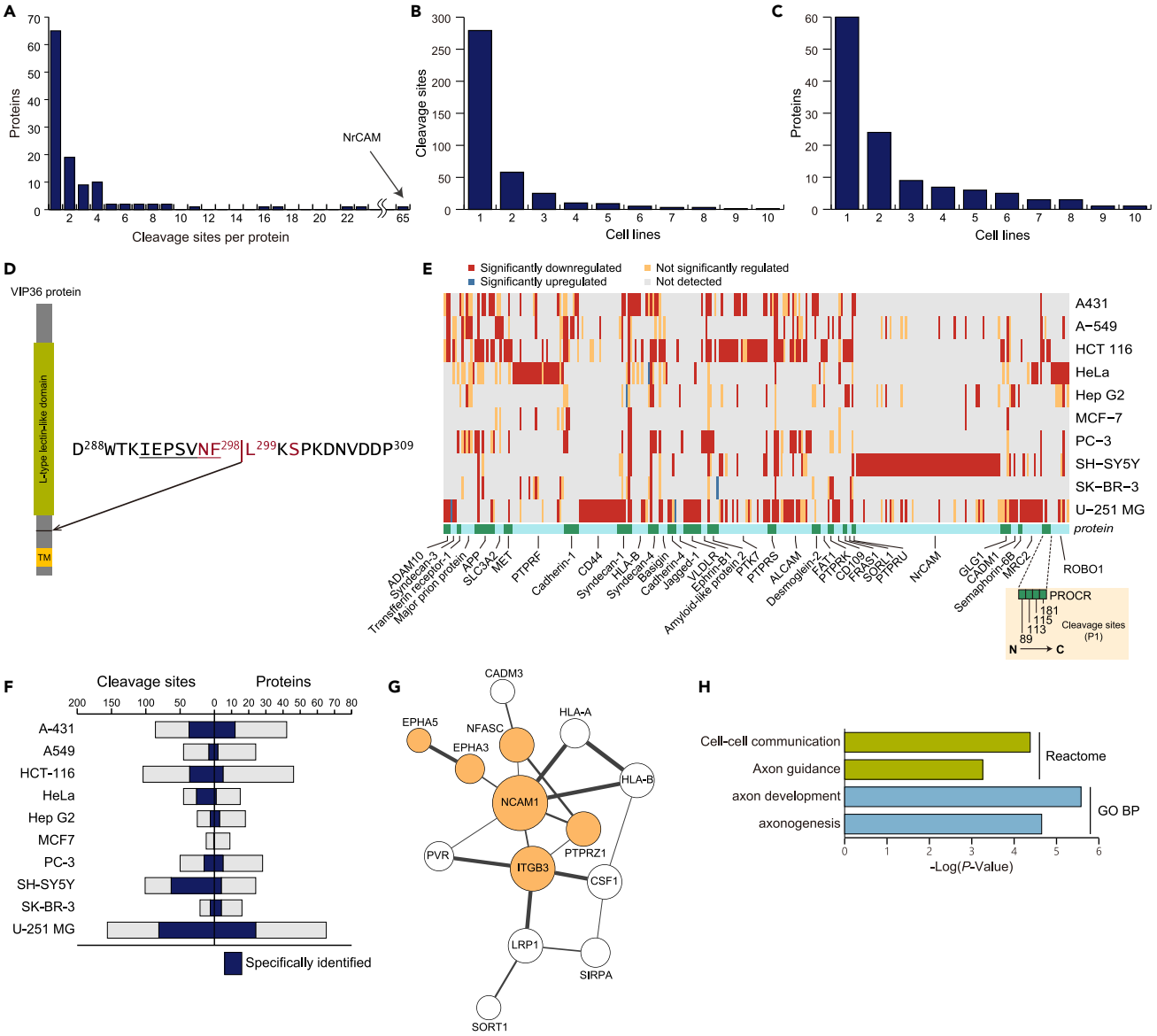


Figure 5. Cell-type-specificity of shedding

(A) The distribution of the number of cleavage sites per protein is depicted for single-pass type-I and type-II transmembrane proteins and GPI-anchored proteins.

(B and C) Histograms depicting the number of shedding substrate identified in 1–10 cell lines at the cleavage site level (B) and at the protein level (C).

(D) VIP36 cleavage site. The amino acids that were previously suggested to be essential for cleavage are highlighted in red (Shirakabe et al., 2011). The sequence of the peptide identified in this study is underlined. TM, transmembrane domain.

(E) Heatmap depicting membrane proteins shed by cleavage at differential sites in different cell lines (36 proteins). Each gray, yellow, red, or blue box designates a cleavage site. Cleavage sites are arranged in the order from N- to C-terminus from left to right in each protein, as the example of PROCR shows. At the bottom, to visualize the boundaries of proteins, green and light-blue boxes are alternately arranged in alphabetical order of UniProt accession from left to right.

(F) Distribution of the number of metalloprotease-regulated cleavage sites in the respective cell lines. The number of cleavage sites specifically identified in each cell type is highlighted in color.

(G and H) Protein interaction network of shedding substrate proteins specifically identified in U-251 MG cells (G). Proteins without any interactions are excluded. Significantly enriched reactome and GO terms of interest are shown with the Benjamini-adjusted p values (H), and the proteins annotated with these terms are highlighted in color (F). All analyses were performed by STRING (v11). The network was visualized using Cytoscape (v3.8.0). Node size reflects the number of connections (direct edges), and edge line width reflects the combined score calculated by STRING.

CellPress
OPEN ACCESS

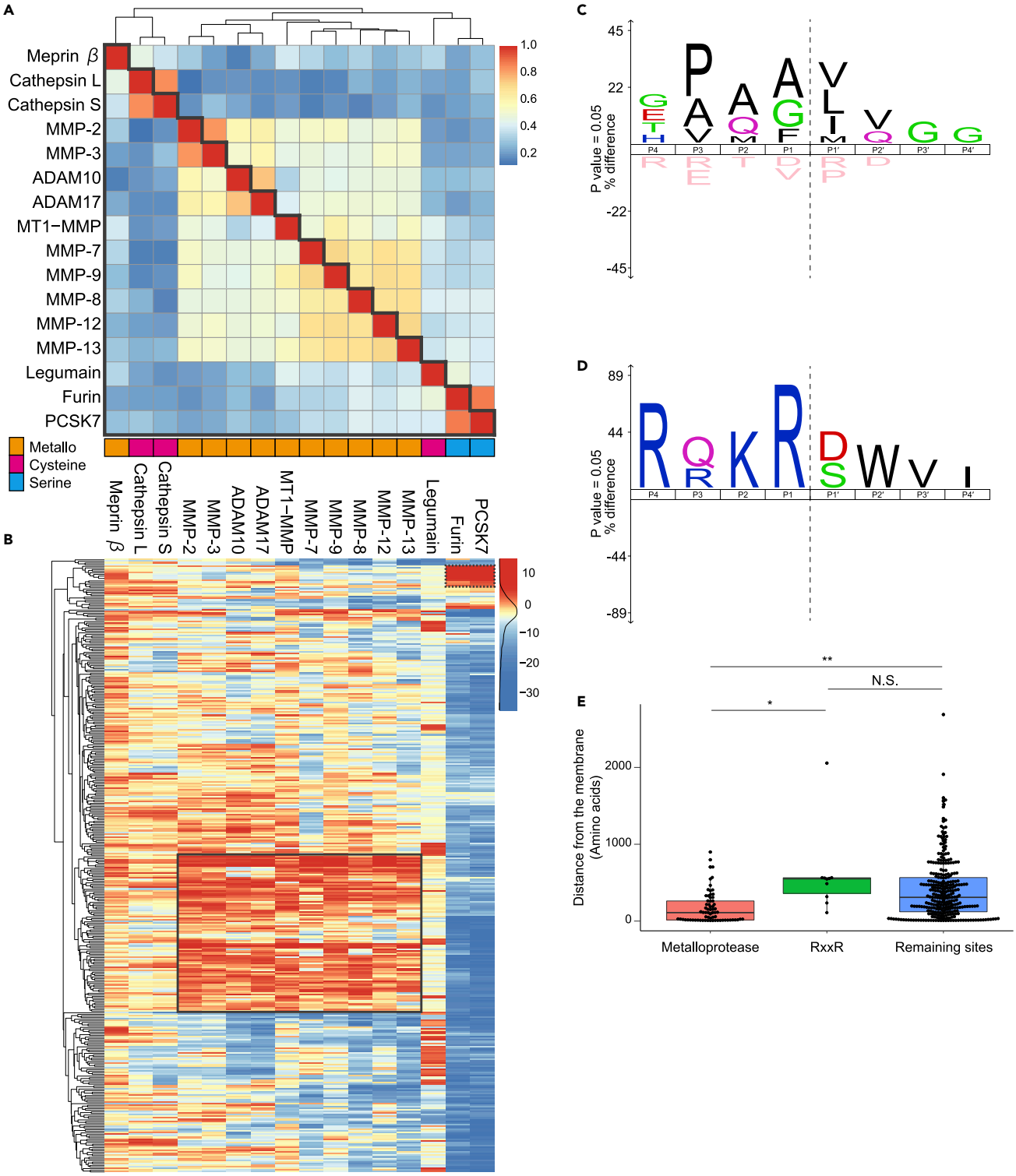


Figure 6. PWM scoring of downregulated cleavage sites

(A) Heatmap depicting the cosine similarity between PWMs for 16 selected sheddases shown in Figure S7. High similarity is highlighted in red, while low similarity is highlighted in blue. Sheddases are ordered by clustering based on the cosine distances (1 – cosine similarity).

(B) Hierarchical clustering analysis of PWM scores. High-scored sites are highlighted in red, while low-scored sites are highlighted in blue. A cluster of cleavage sites high-scored for metalloproteases (metalloprotease cluster) and a cluster for furin and PCSK7 (RxxR cluster) are enclosed with solid-line and dashed-line boxes, respectively.

Figure 6. *Continued*

(C and D) Sequence logos of cleavage sites in the metalloprotease cluster (C) and in the RxxR cluster (D) generated by iceLogo (https://iomics.ugent.be/icelogoserver/). Dashed line shows the cleavage site.

(E) The distances from the transmembrane domain or GPI-anchor site to the cleavage site are compared between the metalloprotease cluster, the RxxR motif cluster, and the remaining sites. The p values were calculated with the Wilcoxon rank-sum test and Bonferroni adjustment was applied. **, $p = 1.26 \times 10^{-7}$; *, $p = 8.82 \times 10^{-4}$; N.S., not significant.

evaluated the identified ectodomain cleavage sites of type-I membrane proteins and GPI-anchored proteins, whose N-terminal regions are located on the extracellular surface, using position weight matrix (PWM) scoring (Imamura et al., 2017) with the accumulated substrate information in the MEROPS protease database (v.12.1, https://www.ebi.ac.uk/merops/index.shtml) (Rawlings et al., 2018). Although ADAM17 and ADAM10 are major sheddases (Huovila et al., 2005), the numbers of their registered substrate cleavage sites are much fewer than those of other metalloproteases due to the relative absence of studies on their cleavage sites, despite their physiological importance. We concatenated *in vitro* substrate information to the MEROPS-registered substrates (Tucher et al., 2014), which yielded 225 and 381 cleavage site sequences for ADAM10 and ADAM17, respectively. In addition we further selected 14 sheddases (matrix metalloprotease (MMP) −2, −3, −7, −8, −9, −12, −13, MT1-MMP, legumain, meprin β, cathepsin S, cathepsin L, furin and proprotein convertase subtilisin/kexin 7 (PCSK7)) from various families, based on following criteria: described as canonical or part-time sheddases in a recent review (Lichtenthaler et al., 2018); having a number of registered substrates in MEROPS greater than one hundred.

We computed substrate sequence PWMs for these individual sheddases (Figures 6A and S7). Importantly, individual sheddases showed distinct substrate preference signatures. Overall, most metalloproteases, except meprin β, commonly exhibit a preference for hydrophobic residues such as Leu and Ile at the P1′ position (one amino acid downstream of the cleavage site), and for Pro and other hydrophobic residues at the P3 position (three amino acids upstream of the cleavage sites); these trends are consistent with previous reports (Eckhard et al., 2016; Tucher et al., 2014). ADAM10 characteristically exhibits a preference for bulky hydrophobic residues such as Tyr, Trp and Phe, which reflects their structural differences (Seegar et al., 2017). Furin and PCSK7 showed known RxxR or RxKR cleavage motifs at P4-P1 (Seidah et al., 2013). Legumain and meprin β showed preferences for Asp/Asn at P1 and acidic residues at P1′, respectively. Given that the PWMs are validated, they can provide a basis for scoring identified putative shedding sites of the respective sheddases.

We calculated PWM scores between more than 6000 relationships (16 sheddases x 378 cleavage sites) (Figure 6B, Table S4). Clustering analysis revealed a group of 86 cleavage sites high-scored for metalloproteases, including the α cleavage site of APP mediated by ADAM10 (Figures 6B and 6C; Table 1). This cluster also included the VIP36 cleavage site that was commonly identified in the analyzed cultured cells (Figure 5D). Notably, although these proteins were previously known to undergo shedding, in many cases, the exact cleavage sites that we identified here have not been reported. We also found another cluster, which involving the RxxR motif of furin and PCSK7 (Figures 6B and 6D). These serine proteases are considered as part-time sheddases, which mainly function as proprotein convertases, but can additionally participate in shedding (Lichtenthaler et al., 2018). The downregulation of these cleavages upon BB-94 treatment might be a secondary or later event following the inhibition of metalloproteases, rather than being due to direct inhibition by BB-94.

Comparison of the distances from the transmembrane domain to the cleavage site in the metalloprotease cluster, the RxxR cluster, and the remaining sites revealed that the cleavage sites in the metalloprotease cluster are located closer to the cell surface (Figure 6E). Furthermore, the cleavage sites close to the membrane (<200 amino acids in Figure 3B) are significantly enriched in metalloprotease cluster cleavage sites with the p value of 2.5×10^{-3} (Fisher's exact test). These results confirm that metalloproteases shed their substrates at positions close to the membrane.

Syndecan-1 with multiple cleavages

Of the 378 sites subjected to PWM scoring, 18 sites on 13 proteins were confirmed as proteolytic sites by the registered data in MEROPS (Table S3). For further validation, we focused on syndecan-1, and compared the quantitative terminomics results (BB-94/DMSO) with the expression profiles of the putative responsible sheddases in MEROPS with PWM scores. We obtained the mRNA expression profiles of these sheddases for nine of ten cell lines, except HCT-116, from The Human Protein Atlas (https://www.proteinatlas.org/).

CellPress
OPEN ACCESS

Table 1. List of cleavage sites grouped into the metalloprotease cluster

UniProt acc.[a]	Gene name[a]	Protein name[a]	Cleavage window[a]	P1[a]	ADAM10 PWM score[b]	ADAM17 PWM score[b]
P55283	CDH4	Cadherin-4	TIGA↓VAAA	724	5.10	8.23
Q9H3T3	SEMA6B	Semaphorin-6B	GPGR↓LTPA	55	3.64	2.35
P04156	PRNP	Major prion protein	GAAA↓AGAV	117	4.54	5.51
P04156	PRNP	Major prion protein	AAAG↓AVVG	119	3.74	4.51
P04156	PRNP	Major prion protein	AGAA↓AAGA	116	4.24	4.67
P04156	PRNP	Major prion protein	AAGA↓VVGG	120	6.91	9.05
P55291	CDH15	Cadherin-15	GAAA↓LLAG	594	8.07	6.95
Q96NY8	PVRL4	Nectin-4	DPQE↓DSGK	337	−6.80	−4.09
P18827	SDC1	Syndecan-1	GPKE↓GEAV	101	−6.35	−5.73
P98172	EFNB1	Ephrin-B1	GASG↓GSSG	224	−0.33	−2.67
P10586	PTPRF	Receptor-type tyrosine-protein phosphatase F	GPFQ↓EVDG	364	−4.41	−2.74
Q13308	PTK7	Inactive tyrosine-protein kinase 7	VPEE↓SEGP	689	−6.23	−1.29
Q13308	PTK7	Inactive tyrosine-protein kinase 7	HPAS↓EAEI	136	−6.89	−4.66
P30466	HLA-B	HLA class I histocompatibility antigen, B-18 alpha chain	ISVG↓YVDG	50	2.44	−0.77
P05067	APP	Amyloid beta A4 protein	HHQK↓LVFF	687	6.61	1.80
P22223	CDH3	Cadherin-3	EVQR↓LTVT	348	2.66	3.67
P0C7U0	ELFN1	Protein ELFN1	YAAE↓VVGP	258	2.81	7.02
P09603	CSF1	Macrophage colony stimulating factor 1	KAFL↓LVQD	87	5.31	4.01
P04156	PRNP	Major prion protein	HMAG↓AAAA	114	3.98	5.00
Q5ZPR3	CD276	CD276 antigen	VSLQ↓VAAP	137	3.27	5.51
Q8WVN6	SECTM1	Secreted and transmembrane protein 1	GAEP↓QSAP	138	1.12	−0.59
P05067	APP	Amyloid beta A4 protein	VLAN↓MISE	579	2.09	0.35
O60462	NRP2	Neuropilin-2	PISA↓FAGE	806	3.02	2.37
P29317	EPHA2	Ephrin type-A receptor 2	QVQA↓LTQE	509	3.96	3.43
Q13308	PTK7	Inactive tyrosine-protein kinase 7	TPAG↓SIEA	488	2.16	1.41
Q9BRK3	MXRA8	Matrix-remodeling-associated protein 8	AAAG↓SSVV	33	1.45	1.31
P98172	EFNB1	Ephrin-B1	ASGG↓SSGD	225	0.27	−0.01
O75056	SDC3	Syndecan-3	GSSA↓AQLP	371	0.53	0.62
Q14126	DSG2	Desmoglein-2	VLEG↓MVEE	279	−2.71	−3.98
Q7Z5N4	SDK1	Protein sidekick-1	AVSA↓QVEA	1998	1.79	−1.62
P27824	CANX	Calnexin	VVGQ↓MIEA	472	0.93	−0.12
P16422	EPCAM	Epithelial cell adhesion molecule	DVAY↓YFEK	214	0.56	−1.84
P09758	TACSTD2	Tumor-associated calcium signal transducer 2	DAAY↓YFER	224	−0.05	−1.47
P55287	CDH11	Cadherin-11	NAEA↓YILN	608	2.10	3.07
P18827	SDC1	Syndecan-1	EGEA↓VVLP	104	−0.61	3.95
Q92823	NRCAM	Neuronal cell adhesion molecule	FVPY↓LIKV	817	2.38	3.18
Q9Y6N7	ROBO1	Roundabout homolog 1	VIPF↓LVPG	841	1.12	0.81
Q8N126	CADM3	Cell adhesion molecule 3	PTAM↓IRPD	236	1.10	2.31
P30480	HLA-B; HLA-A	HLA class I histocompatibility antigen, B-7 alpha chain	TLQS↓MYGC	121	−1.28	−1.22
P22223	CDH3	Cadherin-3	EPVC↓VYTA	460	−1.00	2.00
P04626	ERBB2	Receptor tyrosine-protein kinase erbB-2	TACP↓YNYL	300	2.36	−1.11

(*Continued on next page*)

Table 1. *Continued*

UniProt acc.[a]	Gene name[a]	Protein name[a]	Cleavage window[a]	P1[a]	ADAM10 PWM score[b]	ADAM17 PWM score[b]
Q13740	ALCAM	CD166 antigen	EADE↓ISDE	513	−4.86	−1.57
Q9Y6N7	ROBO1	Roundabout homolog 1	PQPA↓IFWR	381	−2.50	−2.20
Q6YHK3	CD109	CD109 antigen	PSEA↓ISLS	1330	−1.14	1.18
Q08174	PCDH1	Protocadherin-1	NAEL↓VYSL	541	0.24	4.81
Q9Y624	F11R	Junctional adhesion molecule A	TSNA↓VRME	226	−1.04	0.15
P78504	JAG1	Protein jagged1	HPCY↓NSGT	792	1.70	−0.26
P10586	PTPRF	Receptor-type tyrosine-protein phosphatase F	QPNT↓EYSF	1071	−8.26	−4.91
P16070	CD44	CD44 antigen	DIYP↓SNPT	170	−3.85	−4.47
Q15262-2	PTPRK	Receptor-type tyrosine-protein phosphatase kappa	PDPA↓KQTD	745	−5.87	−4.31
Q92896	GLG1	Golgi apparatus protein 1	SDLA↓MQVM	1135	−3.57	−1.88
Q13308	PTK7	Inactive tyrosine-protein kinase 7	NSCN↓IKHT	671	−6.94	−1.51
P98155	VLDLR	Very low-density lipoprotein receptor	ICIN↓LKGG	412	−0.49	−2.39
Q92823	NRCAM	Neuronal cell adhesion molecule	LIIN↓IMSE	107	−2.44	−3.91
Q92823	NRCAM	Neuronal cell adhesion molecule	APQN↓LVLS	369	2.19	1.59
Q13332	PTPRS	Receptor-type tyrosine-protein phosphatase S	DPQP↓IVDG	1275	5.02	4.15
Q92823	NRCAM	Neuronal cell adhesion molecule	HPNG↓ILTE	980	0.61	0.51
P12830	CDH1	Cadherin-1	GAAG↓VCRK	693	−7.43	−0.27
P22223	CDH3	Cadherin-3	DPAG↓WLAM	484	−0.57	1.40
P18827	SDC1	Syndecan-1	ATGA↓SQGL	242	0.99	3.08
P30480	HLA-C	HLA class I histocompatibility antigen, Cw-7 alpha chain	DLRS↓WTAA	156	−3.91	1.61
Q14574	DSC3	Desmocollin-3	TPAA↓QYVR	475	2.20	0.76
Q14517	FAT1	Protocadherin Fat 1	PPFF↓FTIV	3477	2.92	2.73
P32926	DSG3	Desmoglein-3	TPMF↓LLSR	209	3.29	3.82
Q13740	ALCAM	CD166 antigen	NVSA↓ISIP	502	2.41	4.33
Q92896	GLG1	Golgi apparatus protein 1	LAMQ↓VMTS	1137	2.46	3.50
Q92823	NRCAM	Neuronal cell adhesion molecule	PAMA↓SRQV	1158	1.61	4.36
Q14210	LY6D	Lymphocyte antigen 6D	PSYT↓LQGQ	72	0.27	−0.23
P35613	BSG	Basigin	EPMG↓TANI	211	−0.01	2.09
P58658	EVA1C	Protein eva-1 homolog C	DPSG↓SKVL	294	−2.14	−0.98
P10586	PTPRF	Receptor-type tyrosine-protein phosphatase F	TPAQ↓QQEE	1254	−0.07	−4.15
Q12907	LMAN2	Vesicular integral-membrane protein VIP36	SVNF↓LKSP	298	3.04	3.67
P18827	SDC1	Syndecan-1	ASQG↓LLDR	245	0.55	−1.12
P35052	GPC1	Glypican-1	HPQL↓LLPD	182	3.64	0.73
Q14126	DSG2	Desmoglein-2	EIQF↓LISD	570	5.13	4.76
P15151	PVR	Poliovirus receptor	EVQK↓VQLT	153	1.30	3.26
Q14210	LY6D	Lymphocyte antigen 6D	LQGQ↓VSSG	76	0.94	3.05
P98172	EFNB1	Ephrin-B1	GPGA↓SGGS	222	1.06	2.28
P78504	JAG1	Protein jagged1	AVAE↓VRVQ	1054	−1.10	3.13
Q92823	NRCAM	Neuronal cell adhesion molecule	EVSG↓TQTT	702	−1.68	0.85

(*Continued on next page*)

CellPress
OPEN ACCESS

Table 1. *Continued*

UniProt acc.[a]	Gene name[a]	Protein name[a]	Cleavage window[a]	P1[a]	ADAM10 PWM score[b]	ADAM17 PWM score[b]
P05067	APP	Amyloid beta A4 protein	ELLP↓VNGE	608	1.35	3.06
Q92823	NRCAM	Neuronal cell adhesion molecule	TPEG↓VPSA	950	−0.12	3.10
P30508	HLA-C	HLA class I histocompatibility antigen, Cw-12 alpha chain	EPRF↓IAVG	46	0.07	4.02
P30480	HLA-C	HLA class I histocompatibility antigen, Cw-7 alpha chain	EPRF↓ISVG	46	−1.41	2.74
Q08174	PCDH1	Protocadherin-1	KKYF↓LQTT	457	−0.81	−2.64
P08581	MET	Hepatocyte growth factor receptor	GFMF↓LTDQ	228	6.58	1.36

The cleavage sites grouped into the metalloprotease cluster shown in Figure 6B are listed. The order from top to bottom is the same to that of the cluster.
[a]Accessions, gene names, protein names, cleavage window sequences and P1 positions (the first upstream amino acid of the cleavage site) follow those in the SwissProt database utilized for database search. Cleavage sites are indicated with "↓".
[b]PWM scores for ADAM10 and ADAM17 shown in Figure 6B.

We found three MEROPS-registered cleavage sites, R230↓N231, A242↓S243 and G245↓L246 (Manon-Jensen et al., 2013) of syndecan-1, in our dataset (Figures S8B and S8C). Notably, MT-MMP1 and MMP-7 gave high PWM scores (>2) to G245↓L246, and these sheddases have been reported as responsible sheddases for this site (Figure S8C). Cleavage at A242↓S243, possibly by ADAM17, MMP-9 and MMP-12 based on PWM scores>2, was observed across all analyzed cell lines, and was significantly downregulated in 9 cell lines including HCT-116, but not Hep G2 (Figure S8B). In addition to these downregulated cleavages, we found the cleavage site at R230↓N231 on syndecan-1, which has been reported to be mediated by plasmin (Manon-Jensen et al., 2013) was upregulated (Figures 5E and S8A). Interestingly, significant upregulation of the cleavage at R230↓N231 was observed only in Hep G2 cells (Figure S8A). Moreover, plasmin is expressed only in Hep G2 cells, indicating that the cleavage at R230↓N231 in Hep G2 cells would be mediated by plasmin. Furthermore, the proteolytic peptide having this upregulated site harbors two identified metalloprotease-dependent cleavage sites (Figure S8D). This proteolytic peptide (231-248) was predominantly secreted in Hep G2 cells in the presence of BB-94 (Figure S8E), indicating that the suppression of double cleaved peptides (231-242 or 231-245) by BB-94 results in the generation of the plasmin-cleaved peptide (231-248).

In vitro assay verifying the cleavage prediction by PWM scores

Finally, to demonstrate the accuracy of prediction based on our PWM scoring, we performed *in vitro* assay using peptide substrates and recombinant sheddases. We selected and tested 6 cleavage site-sheddase pairs with high PWM scores and prepared two synthetic peptides containing 3 cleavage sites for each (Figure 7A). As expected, all tested sites were successfully cleaved by the corresponding sheddases (Figures 7B–7H). Legumain has a clear D/N motif at the P1 position (Figure S7), and we also confirmed that legumain cleaved substrate peptide-2 at other D sites by quantifying peptides such as PGEDGTLIDPQPIVD, GRRKRSAAL, and GTLIDPQPIVD, in addition to the expected peptide (Figure 7F). Meprin β incubated with substrate peptide-1 generated abundant DDVSLMFKN peptide (Figure 7C) rather than DDVSLMFKNAPTSPDRVFNT, the expected product. To evaluate the cleavage selectivity, we computed PWM scores at every site of substrate peptide-1 for meprin β. We found the MFKN↓APTS sequence has a relatively higher score (Figure 7I), in accordance with the observation that the major cleavage product was DDVSLMFKN. Notably, we also detected the DVSLMFKN peptide corresponding to the second highest score (PTDD↓DVSL) (Figure 7I). Overall, these experimental results strongly validate our data set and support the accuracy of the scoring method.

DISCUSSION

Here, we have demonstrated that SCX-based N- and C-terminal peptide enrichment in combination with high-resolution mass spectrometry allows reproducible quantification of thousands of protein termini derived from membrane proteins in culture supernatant, with limited amounts of materials. One major advantage of our workflow is that only a very short conditioning time (1 hr) is needed for collecting the secreted proteins, which would reduce undesired alterations arising from elimination of serum (Eichelbaum

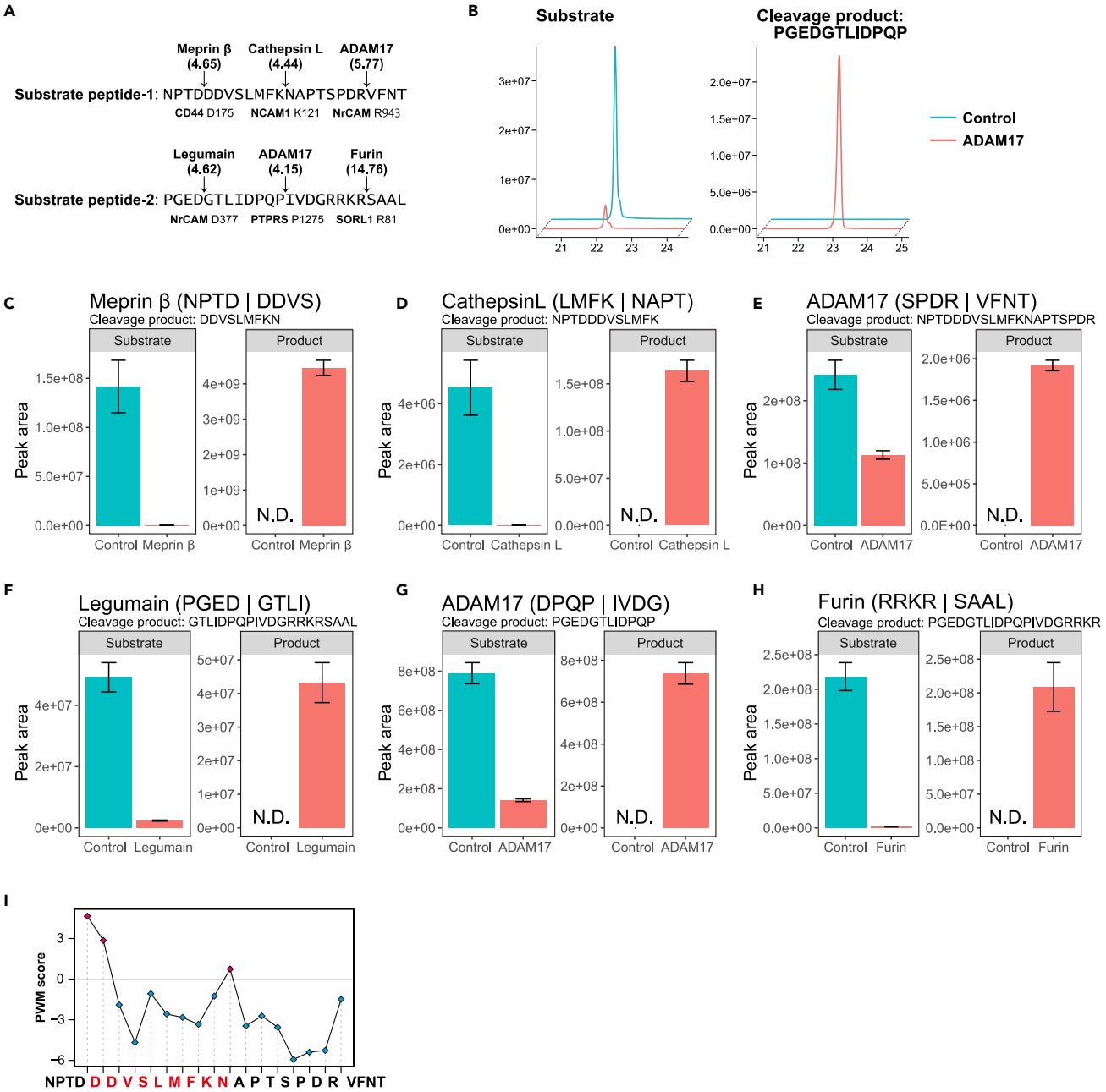


Figure 7. Validation of identified cleavage sites by *in vitro* assay

(A) Sequences of utilized synthetic substrate peptides. The substrates were designed as the flanking ±4 residues surrounding each cleavage site. Three substrate sequences were concatenated and synthesized as a 24-mer peptide. Arrows show the tested cleavage sites and sheddases. Parenthesized numbers show PWM scores. Under the sequences, protein names are shown with the P1 amino acid number, respectively.

(B) Detection of substrate and cleavage product by LC/MS is illustrated with their extracted ion chromatograms for ADAM17 and the DPQPIVDG sequence of PTPRS as an example.

(C–H) The peak areas of the substrate and the cleavage product peptides are calculated. The sequence of cleavage product utilized for quantification is shown above. The experiments were done in triplicate, and the data are presented as mean ± standard error of the mean (SEM). N.D., not detected.

(I) PWM scores for meprin β at respective positions of substrate peptide-1. Positive scores are highlighted in magenta and negative scores are highlighted in blue. The sequence of the cleavage product utilized for the quantification of cleavage by meprin β is highlighted in red on the x axis.

et al., 2012), whereas many previous shedding studies have employed longer incubation times (Johnson et al., 2017; Tsumagari et al., 2017). To avoid non-physiological alterations caused by the serum-free condition, smart technologies have been developed that enable shedding substrate identification in the presence of serum, such as hiSPECS (Tüshaus et al., 2020). Application of our sensitive terminal peptides enrichment workflow to the shedding substrate protein samples obtained by hiSPECS might make it possible to more reliably identify physiological cleavage sites. As our protocol of N, C-terminomics consists of simple steps, including isobaric tagging and tip-based SCX, it should prove to be a practically useful platform not only for identification of shedding substrates in the supernatant, but also for conventional N, C-terminomics of cells and tissues.

The results were validated by the observation of membrane-protein-selective downregulation by BB-94. Broadly speaking, shedding occurs in the vicinity of the cell surface, thus liberating a large portion of the extracellular domain. Also, shedding targets the main components of signal transduction including protein tyrosine kinases/phosphatases and receptors. We successfully identified the cleavage site of VIP36, which could not be fully clarified by amino acid substitution and western blotting studies (Shirakabe et al., 2011), and interestingly this was the only cleavage site commonly identified across the 10 investigated cell lines. The importance of investigating diverse cell types is highlighted by the substantial differences of shedding pattern across the analyzed cell lines. We focused on the function of U-251 MG cell-unique shedding and found that the released proteins are important for the environment, rather than the cells themselves. Overall, our results provide further support for the idea that shedding contributes to cellular communication. Many of the identified putative substrate proteins were previously reported to undergo shedding, and here we have unveiled the exact cleavage sites. We wish to emphasize that this is the first study in which native shedding is overviewed at the cleavage site level on a proteome-wide scale across multiple cell types.

Our data set contains cleavage sites generated downstream of metalloprotease-dependent shedding. In addition, all fragments may be further processed by a wide range of extracellular endo- and exopeptidases. We therefore evaluated identified metalloprotease-regulated cleavage sites by PWM scoring, which we previously employed to evaluate protein kinase-phosphorylation site relationships (Imamura et al., 2017). We successfully found cleavage sites high-scored for metalloproteases, which could represent the direct substrates of the metalloproteases. In addition, we found a cluster high-scored for furin and PCSK7, which have a clear cleavage motif, RxxR. Proteolytic processing by these proteases is mainly involved in the maturation of membrane proteins. These proteases might be activated as a result of drastic upregulation of membrane protein turnover by PMA-activated shedding. We observed distinct signatures of PWMs even for ADAMs and MMPs, which have extremely similar preferences. Importantly, we demonstrated that identified shedding site sequences with high PWM scores can be indeed cleaved, at least *in vitro*. In the future, accumulating knowledge of cleavage sites should enable even more powerful approaches, such as machine learning, for confident prediction of sheddase-cleavage site relationships.

Our data set provides a useful resource to support hypothesis-driven studies on shedding. It seems clear that a technological breakthrough in proteomic identification of shedding cleavage sites is needed, not only from a scientific viewpoint, but also for therapeutic purposes, given the involvement of shedding in various diseases, including cancer, cardiovascular disease, inflammation, and neurodegeneration (Edwards et al., 2008; Herrlich and Herrlich, 2017; Weber and Saftig, 2012). It has been suggested that cells shed different membrane proteins depending upon external stimulation (Dang et al., 2013; Huovila et al., 2005), i.e., cells selectively shed substrates in response to different stimuli. However, how shedding is made specific has long been enigmatic. Most shedding is mediated by ADAM10 and ADAM17, and these sheddases share many substrates (Gooz, 2010; Huovila et al., 2005). In addition, sheddases generally have broad amino acid preferences, and do not require a specific amino acid at any position surrounding the cleavage site. Consequently, it is extremely difficult to understand the basal mechanisms that determine shedding targets. Our strategy presented here enables systematic investigation of changes of shedding at the site level and should facilitate work to identify critical factors that regulate shedding by means of proteomics studies.

Limitations of the study

First, termini that are not accessible by TrypN or LysC/trypsin, e.g. due to too short or too long terminal peptides, would not be captured in our methodology. Second, we have not performed

biochemical validation of the specificity of spiked proteases utilized for sample preparation. Thus, there is a possibility that identified semi-specific peptides may have been generated by digestion at unpredicted sites during sample preparation; however, this seems unlikely, as previous studies have demonstrated extremely high specificity, particularly for trypsin (Olsen et al., 2004; Wilson et al., 2020). Third, we utilized a short pulse of PMA to stimulate shedding, which is an artificial stimulus, so that physiological cleavage sites of the identified proteins may be different from the identified sites. Forth, we only used *in vitro* peptide assay by MS to validate the sheddase-substrate relationship since we and others have experienced that traditional western blotting coupled with or without single amino acid substitution at expected cleavage sites of the substrates did not work due to the unexpected cleavage selectivity (Shirakabe et al., 2011; Tsumagari et al., 2017). Finally, as already discussed, serum-free conditions are required, which may lead to non-physiological changes, even though the duration is short in our method.

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Yasushi Ishihama (yishiham@pharm.kyoto-u.ac.jp).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The raw LC/MS/MS data generated during this study have been deposited with the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the jPOST partner repository (http://jpost.org) with the data set identifier PXD021378 (JPST000632).

METHODS

All methods can be found in the accompanying transparent methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.102259.

ACKNOWLEDGMENTS

We thank Eito Yamamoto (Kyoto University) for assisting in the development of the C-terminal peptide enrichment protocol and Atsuko Sehara-Fujisawa (Kyoto University) for providing SH-SY5Y cells. K.T. was supported by a fellowship for young scientists from the Japan Society for the Promotion of Science (JSPS). This work was supported by the JST Strategic Basic Research Program, CREST (grant No. 18070870), and by a JSPS Grant-in-Aid for Scientific Research (No. 17H05667).

AUTHOR CONTRIBUTIONS

Conceptualization, K.T. and Y.I.; methodology, K.T. and C-H.C.; formal analysis, K.T.; investigation, K.T.; resources, K.T.; writing – original draft, K.T.; writing – review & editing, Y.I.; visualization, K.T.; supervision, Y.I.; project administration, Y.I.; funding acquisition, K.T. and Y.I.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 15, 2020
Revised: January 18, 2021
Accepted: February 26, 2021
Published: April 23, 2021

CellPress
OPEN ACCESS

REFERENCES

Adachi, J., Hashiguchi, K., Nagano, M., Sato, M., Sato, A., Fukamizu, K., Ishihama, Y., and Tomonaga, T. (2016). Improved proteome and phosphoproteome analysis on a cation exchanger by a combined acid and salt gradient. Anal. Chem. *88*, 7899–7903.

Aebersold, R., Agar, J.N., Amster, I.J., Baker, M.S., Bertozzi, C.R., Boja, E.S., Costello, C.E., Cravatt, B.F., Fenselau, C., Garcia, B.A., et al. (2018). How many human proteoforms are there? Nat. Chem. Biol. *14*, 206–214.

Alpert, A.J., Petritis, K., Kangas, L., Smith, R.D., Mechtler, K., Mitulović, G., Mohammed, S., and Heck, A.J.R. (2010). Peptide orientation affects selectivity in ion-exchange chromatography. Anal. Chem. *82*, 5253–5259.

Apweiler, R., Hermjakob, H., and Sharon, N. (1999). On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochim. Biophys. Acta *1473*, 4–8.

Beard, H.A., Barniol-Xicota, M., Yang, J., and Verhelst, S.H.L. (2019). Discovery of cellular roles of intramembrane proteases. ACS Chem. Biol. *14*, 2372–2388.

Brown, M.S., Ye, J., Rawson, R.B., and Goldstein, J.L. (2000). Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans. Cell *100*, 391–398.

Chang, C.-H., Chang, H.-Y., Rappsilber, J., and Ishihama, Y. (2021). Isolation of acetylated and unmodified protein N-terminal peptides by strong cation exchange chromatographic separation of TrypN-digested peptides. Mol. Cell. Proteomics *20*, 100003.

Chow, V.W., Mattson, M.P., Wong, P.C., and Gleichmann, M. (2010). An overview of APP processing enzymes and products. Neuromolecular Med. *12*, 1–12.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. J. Proteome Res. *10*, 1794–1805.

Dang, M., Armbruster, N., Miller, M.A., Cermeno, E., Hartmann, M., Bell, G.W., Root, D.E., Lauffenburger, D.A., Lodish, H.F., and Herrlich, A. (2013). Regulated ADAM17-dependent EGF family ligand release by substrate-selecting signaling pathways. Proc. Natl. Acad. Sci. U S A *110*, 9776–9781.

Eckhard, U., Huesgen, P.F., Schilling, O., Bellac, C.L., Butler, G.S., Cox, J.H., Dufour, A., Goebeler, V., Kappelhoff, R., and Auf dem Keller, U. (2016). Active site specificity profiling of the matrix metalloproteinase family: proteomic identification of 4300 cleavage sites by nine MMPs explored with structural and synthetic peptide cleavage analyses. Matrix Biol. *49*, 37–60.

Edwards, D.R., Handsley, M.M., and Pennington, C.J. (2008). The ADAM metalloproteinases. Mol. Aspects Med. *29*, 258–289.

Eichelbaum, K., Winter, M., Diaz, M.B., Herzig, S., and Krijgsveld, J. (2012). Selective enrichment of newly synthesized proteins for quantitative secretome analysis. Nat. Biotechnol. *30*, 984–990.

Gauci, S., Helbig, A.O., Slijper, M., Krijgsveld, J., Heck, A.J.R., and Mohammed, S. (2009). Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach. Anal. Chem. *81*, 4493–4501.

Gooz, M. (2010). ADAM-17: the enzyme that does it all. Crit. Rev. Biochem. Mol. Biol. *45*, 146–169.

Güner, G., and Lichtenthaler, S.F. (2020). The substrate repertoire of γ-secretase/presenilin. Semin. Cell Dev. Biol. *105*, 27–42.

Helbig, A.O., Gauci, S., Raijmakers, R., Van Breukelen, B., Slijper, M., Mohammed, S., and Heck, A.J.R. (2010). Profiling of N-acetylated protein termini provides in-depth insights into the N-terminal nature of the proteome. Mol. Cell. Proteomics *9*, 928–939.

Herrlich, P., and Herrlich, A. (2017). ADAM metalloprotease-released cancer biomarkers. Trends Cancer *3*, 482–490.

Hinkle, C.L., Sunnarborg, S.W., Loiselle, D., Parker, C.E., Stevenson, M., Russell, W.E., and Lee, D.C. (2004). Selective roles for tumor necrosis factor α-converting enzyme/ADAM17 in the shedding of the epidermal growth factor receptor ligand family. The juxtamembrane stalk determines cleavage efficiency. J. Biol. Chem. *279*, 24179–24188.

Hu, X., Hicks, C.W., He, W., Wong, P., MacKlin, W.B., Trapp, B.D., and Yan, R. (2006). Bace1 modulates myelination in the central and peripheral nervous system. Nat. Neurosci. *9*, 1520–1525.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Huovila, A.-P.J., Turner, A.J., Pelto-Huikko, M., Kärkkäinen, I., and Ortiz, R.M. (2005). Shedding light on ADAM metalloproteinases. Trends Biochem. Sci. *30*, 413–422.

Imamura, H., Wagih, O., Niinae, T., Sugiyama, N., Beltrao, P., and Ishihama, Y. (2017). Identifications of putative PKA substrates with quantitative phosphoproteomics and primary-sequence-based scoring. J. Proteome Res. *16*, 1825–1830.

Johnson, N., Březinová, J., Stephens, E., Burbridge, E., Freeman, M., Adrain, C., and Strisovsky, K. (2017). Quantitative proteomics screen identifies a substrate repertoire of rhomboid protease RHBDL2 in human cells and implicates it in epithelial homeostasis. Sci. Rep. *7*, 7283.

Kleifeld, O., Doucet, A., Auf Dem Keller, U., Prudova, A., Schilling, O., Kainthan, R.K., Starr, A.E., Foster, L.J., Kizhakkedathu, J.N., and Overall, C.M. (2010). Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. Nat. Biotechnol. *28*, 281–288.

Lichtenthaler, S.F., Lemberg, M.K., and Fluhrer, R. (2018). Proteolytic ectodomain shedding of membrane proteins in mammals—hardware, concepts, and recent developments. EMBO J. *37*, e99456.

Mahrus, S., Trinidad, J.C., Barkan, D.T., Sali, A., Burlingame, A.L., and Wells, J.A. (2008). Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini. Cell *134*, 866–876.

Manon-Jensen, T., Multhaupt, H.A.B., and Couchman, J.R. (2013). Mapping of matrix metalloproteinase cleavage sites on syndecan-1 and syndecan-4 ectodomains. FEBS J. *280*, 2320–2331.

La Marca, R., Cerri, F., Horiuchi, K., Bachi, A., Feltri, M.L., Wrabetz, L., Blobel, C.P., Quattrini, A., Salzer, J.L., and Taveggia, C. (2011). TACE (ADAM17) inhibits Schwann cell myelination. Nat. Neurosci. *14*, 857–865.

Merilahti, J.A.M., and Elenius, K. (2019). Gamma-secretase-dependent signaling of receptor tyrosine kinases. Oncogene *38*, 151–163.

Nagano, O., and Saya, H. (2004). Mechanism and biological significance of CD44 cleavage. Cancer Sci. *95*, 930–935.

Niedermaier, S., and Huesgen, P.F. (2019). Positional proteomics for identification of secreted proteoforms released by site-specific processing of membrane proteins. Biochim. Biophys. Acta *1867*, 140138.

Olsen, J.V., and Mann, M. (2013). Status of large-scale analysis of posttranslational modifications by mass spectrometry. Mol. Cell. Proteomics *12*, 3444–3452.

Olsen, J.V., Ong, S.E., and Mann, M. (2004). Trypsin cleaves exclusively C-terminal to arginine and lysine residues. Mol. Cell. Proteomics *3*, 608–614.

Paetzel, M., Karla, A., Strynadka, N.C.J., and Dalbey, R.E. (2002). Signal peptidases. Chem. Rev. *102*, 4549–4580.

Prudova, A., Auf dem Keller, U., Butler, G.S., and Overall, C.M. (2010). Multiplex N-terminome analysis of MMP-2 and MMP-9 substrate degradomes by iTRAQ-TAILS quantitative proteomics. Mol. Cell. Proteomics *9*, 894–911.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat. Protoc. *2*, 1896–1906.

Rawlings, N.D., Barrett, A.J., Thomas, P.D., Huang, X., Bateman, A., and Finn, R.D. (2018). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Res. *46*, D624–D632.

Reiss, K., and Saftig, P. (2009). The "A Disintegrin and Metalloprotease" (ADAM) family of

sheddases: physiological and cellular functions. Semin. Cell Dev. Biol. *20*, 126–137.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. *11*, R25.

Seegar, T.C.M., Killingsworth, L.B., Saha, N., Meyer, P.A., Patra, D., Zimmerman, B., Janes, P.W., Rubinstein, E., Nikolov, D.B., Skiniotis, G., et al. (2017). Structural basis for regulated proteolysis by the α-secretase ADAM10. Cell *171*, 1638–1648.e7.

Seidah, N.G., Sadr, M.S., Chrétien, M., and Mbikay, M. (2013). The multifaceted proprotein convertases: their unique, redundant, complementary, and opposite functions. J. Biol. Chem. *288*, 21473–21481.

Shirakabe, K., Hattori, S., Seiki, M., Koyasu, S., and Okada, Y. (2011). VIP36 protein is a target of ectodomain shedding and regulates phagocytosis in macrophage Raw 264.7 cells. J. Biol. Chem. *286*, 43154–43163.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. *47*, D607–D613.

Takami, M., Nagashima, Y., Sano, Y., Ishihara, S., Morishima-Kawashima, M., Funamoto, S., and Ihara, Y. (2009). γ-Secretase: successive tripeptide and tetrapeptide release from the transmembrane domain of β-carboxyl terminal fragment. J. Neurosci. *29*, 13042–13052.

Tsumagari, K., Shirakabe, K., Ogura, M., Sato, F., Ishihama, Y., and Sehara-Fujisawa, A. (2017). Secretome analysis to elucidate metalloprotease-dependent ectodomain shedding of glycoproteins during neuronal differentiation. Genes Cells *22*, 237–244.

Tucher, J., Linke, D., Koudelka, T., Cassidy, L., Tredup, C., Wichert, R., Pietrzik, C., Becker-Pauly, C., and Tholey, A. (2014). LC-MS based cleavage site profiling of the proteases ADAM10 and ADAM17 using proteome-derived peptide libraries. J. Proteome Res. *13*, 2205–2214.

Tüshaus, J., Müller, S.A., Kataka, E.S., Zaucha, J., Sebastian Monasor, L., Su, M., Güner, G., Jocher, G., Tahirovic, S., and Frishman, D. (2020). An optimized quantitative proteomics method establishes the cell type-resolved mouse brain secretome. EMBO J. *39*, e105693.

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat. Methods *13*, 731–740.

Weber, S., and Saftig, P. (2012). Ectodomain shedding and ADAMs in development. Development *139*, 3693–3709.

Weeks, A.M., and Wells, J.A. (2018). Engineering peptide ligase specificity by proteomic identification of ligation sites. Nat. Chem. Biol. *14*, 50–57.

Willem, M., Garratt, A.N., Novak, B., Citron, M., Kaufmann, S., Rittger, A., DeStrooper, B., Saftig, P., Birchmeier, C., and Haass, C. (2006). Control of peripheral nerve myelination by the β-secretase BACE1. Science *314*, 664–666.

Wilson, J.P., Ipsaro, J.J., Del Giudice, S.N., Turna, N.S., Gauss, C.M., Dusenbury, K.H., Marquart, K., Rivera, K.D., and Pappin, D.J. (2020). Tryp-N: a thermostable protease for the production of N-terminal argininyl and lysinyl peptides. J. Proteome Res. *19*, 1459–1469.

Zheng, Y., Saftig, P., Hartmann, D., and Blobel, C. (2004). Evaluation of the contribution of different ADAMs to tumor necrosis factor α (TNFα) shedding and of the function of the TNFα ectodomain in ensuring selective stimulated shedding by the TNFα convertase (TACE/ADAM17). J. Biol. Chem. *279*, 42898–42906.

iScience, Volume *24*

Supplemental information

Exploring the landscape of ectodomain shedding by quantitative protein terminomics

Kazuya Tsumagari, Chih-Hsiang Chang, and Yasushi Ishihama

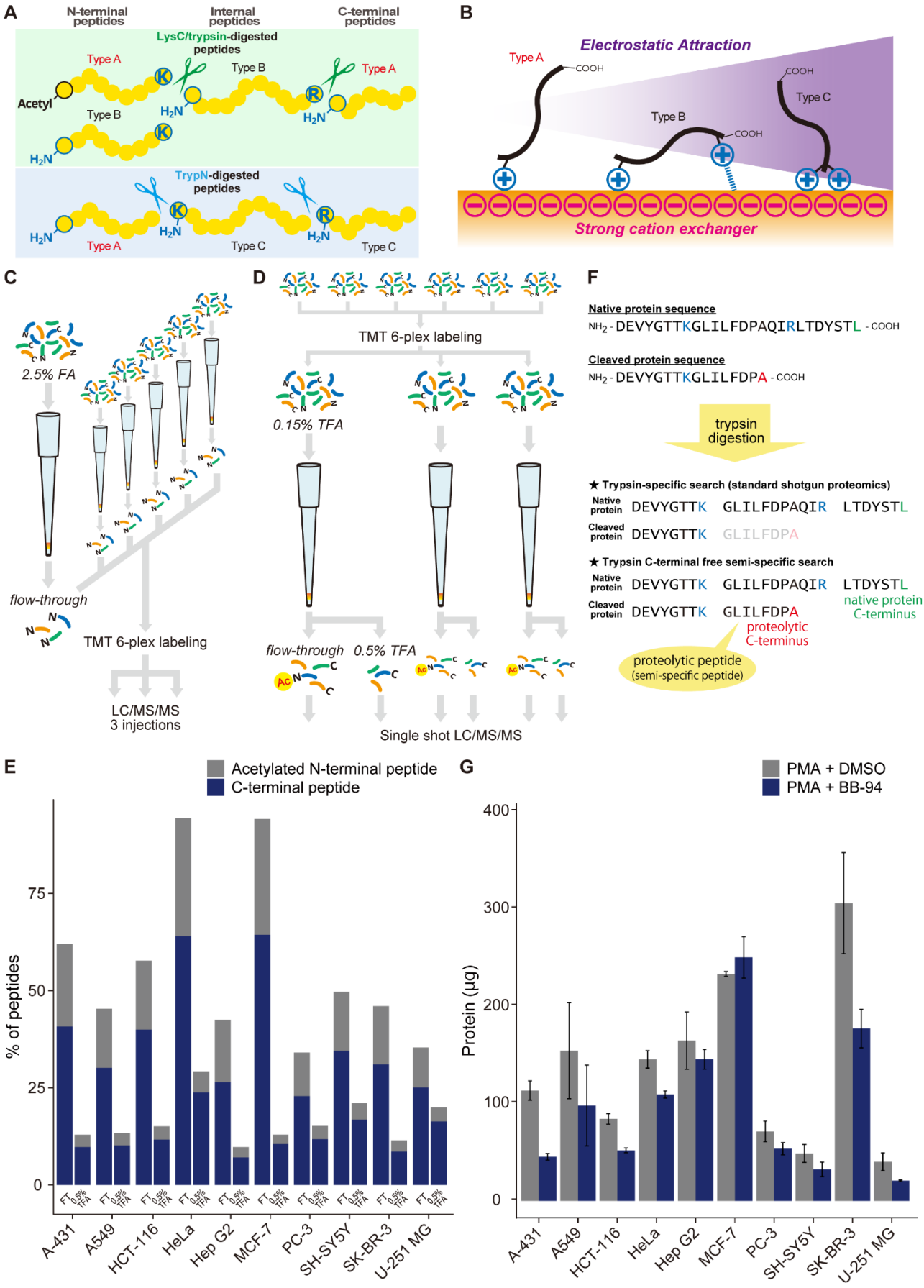

A
N-terminal peptides
Internal peptides
C-terminal peptides
LysC/trypsin-digested peptides
Type A
Type B
Type A
Acetyl
Type B
TrypN-digested peptides
Type A
Type C
Type C
B
Electrostatic Attraction
Type A
COOH
Type B
COOH
COOH
Type C
Strong cation exchanger
C
2.5% FA
flow-through
TMT 6-plex labeling
LC/MS/MS
3 injections
D
TMT 6-plex labeling
0.15% TFA
flow-through
0.5% TFA
Single shot LC/MS/MS
F
Native protein sequence
NH2 - DEVYGTTKGLILFDPAQIRLTDYSTL - COOH
Cleaved protein sequence
NH2 - DEVYGTTKGLILFDPA - COOH
trypsin digestion
★ Trypsin-specific search (standard shotgun proteomics)
Native protein DEVYGTTK GLILFDPAQIR LTDYSTL
Cleaved protein DEVYGTTK GLILFDPA
★ Trypsin C-terminal free semi-specific search
Native protein DEVYGTTK GLILFDPAQIR LTDYSTL
native protein C-terminus
Cleaved protein DEVYGTTK GLILFDPA
proteolytic C-terminus
proteolytic peptide (semi-specific peptide)
E
Acetylated N-terminal peptide
C-terminal peptide
% of peptides
0
25
50
75
FT
0.5% TFA
A-431
A549
HCT-116
HeLa
Hep G2
MCF-7
PC-3
SH-SY5Y
SK-BR-3
U-251 MG
G
PMA + DMSO
PMA + BB-94
Protein (µg)
0
100
200
300
400
A-431
A549
HCT-116
HeLa
Hep G2
MCF-7
PC-3
SH-SY5Y
SK-BR-3
U-251 MG

Figure S1. Terminal peptide enrichment by SCX, Related to Figure 1.

(A, B) Digestion with TrypN yields peptides with at least a +2 charge with Lys or Arg and an α-amino group at the N-terminus (Type C). On the other hand, peptides derived from protein N-termini have a +1 charge with neither Lys nor Arg, and only His-containing peptides with an unmodified N-terminus have a +2 charge (Type A). In addition, digestion with trypsin yields internal peptides with at least a +2 charge (Type B). In this case, C-terminal peptides without His have a +1 charge and those with His have a +2 charge (Type A). Based on these facts, we isolated N-terminal peptides and C-terminal peptides using SCX at low pH, respectively. Note that acetylated N-terminal peptides generated by trypsin/LysC digest have a 0 or +1 charge (Type A) and are isolated together with C-terminal peptides.

(C) For N-terminal peptide enrichment, TrypN-digested peptides were dissolved in 2.5% formic acid (FA) containing 30% acetonitrile (ACN) and loaded on the SCX-StageTip. The flow-through fraction was collected. Following TMT-labeling, the isolated N-terminal peptides were analyzed in triplicate.

(D) TMT-labeled peptides were divided into three parts and subjected to C-terminal peptide enrichment using three SCX-StageTips. Peptides were dissolved in 0.15% trifluoroacetic acid (TFA) containing 30% ACN and loaded on the SCX-StageTip. The flow-through and the 0.5% TFA-eluted fractions were separately collected and subjected to single shot analysis. (E) Efficiency of C-terminal peptide enrichment. As described above, acetylated N-terminal peptides were isolated together with C-terminal peptides. Remaining % peptides mainly correspond to internal peptides (Type B peptides in Figure S1A).

(F) Example of terminus and cleavage site identification by trypsin semi-specific search. In standard shotgun proteomics, only peptides whose N- and C-termini are both in accord with trypsin specificity are considered. The C-terminus of native protein C-terminal peptide does not match to the specificity of trypsin (cleavage at C-terminal side of K/R, highlighted in blue), but can be identified (highlighted in green). In C-terminal free semi-specific search, the peptides whose C-terminus is not with the specificity of trypsin (highlighted in red) can be additionally identified. We refer to the total of native protein termini and proteolytic termini as “all termini” in Figure 1E.

(G) The protein yields from a biological replicate consisting of three 15 cm dishes are shown for each condition. For each condition, three replicates were prepared, and the data are presented as mean ± standard error of the mean (SEM). The approximate number of cells per 15 cm dish at 90% confluency ranged from 10 million (HeLa) to 30 million (SH-SY5Y).

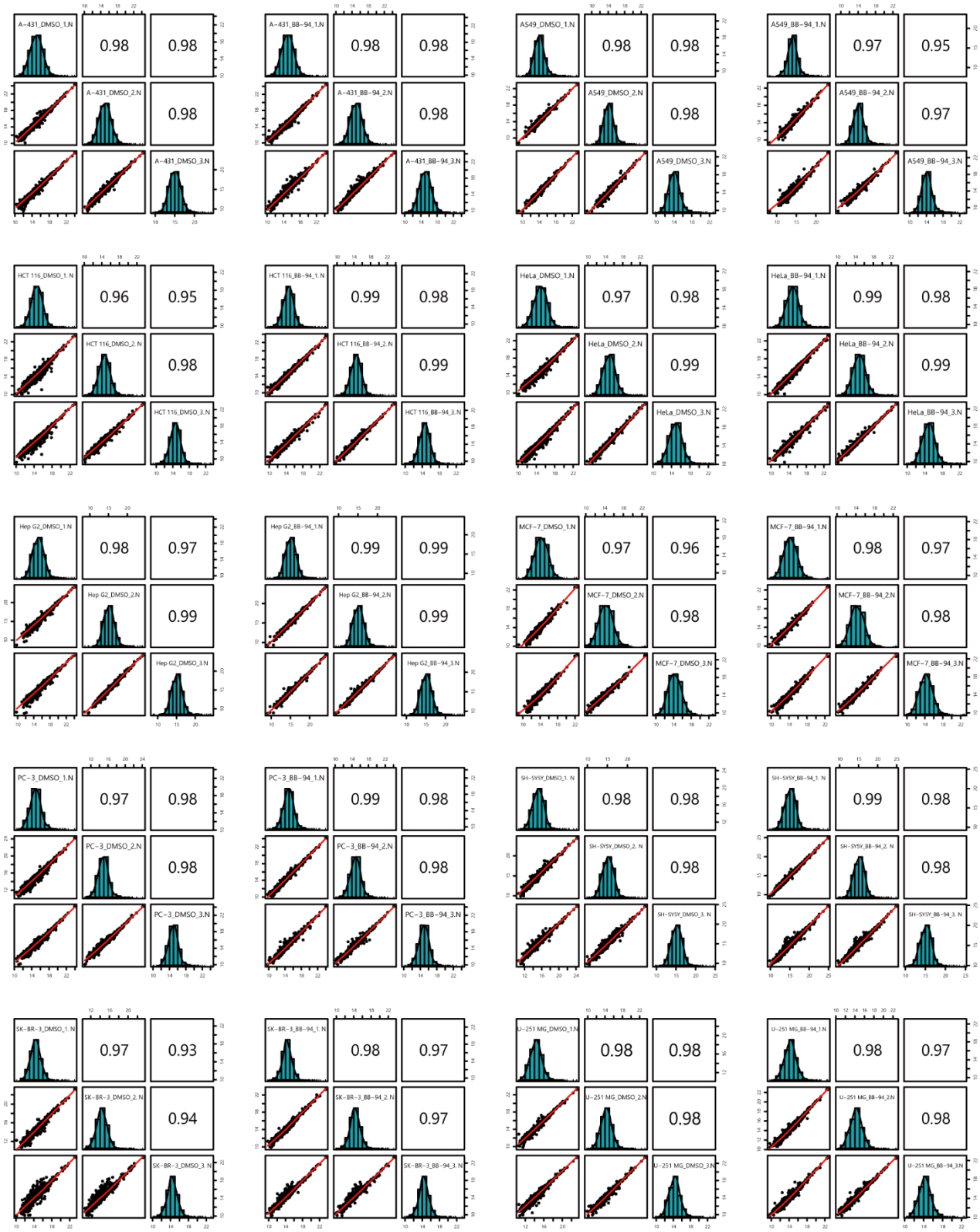


Figure S2. Reproducibility of N-terminomics, Related to Figure 1.

Logarithmized normalized TMT-reporter intensities of N-terminal peptides are plotted against each other for triplicate determinations in each condition, and the corresponding Pearson correlation coefficients are shown.

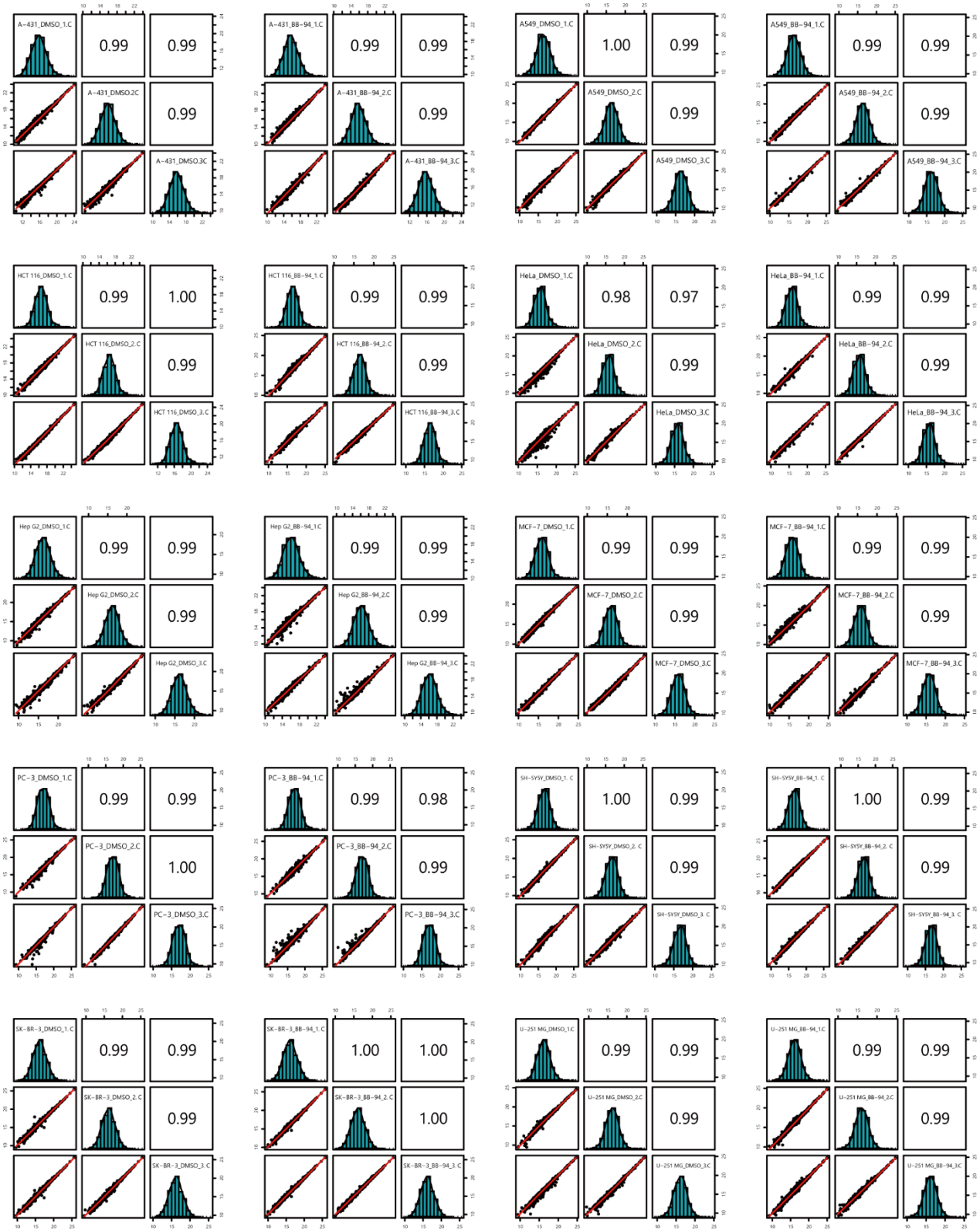


Figure S3. Reproducibility of C-terminomics, Related to Figure 1.

Logarithmized normalized TMT-reporter intensities of C-terminal peptides are plotted for against each other for triplicate determinations in each condition, and the corresponding Pearson correlation coefficients are shown.

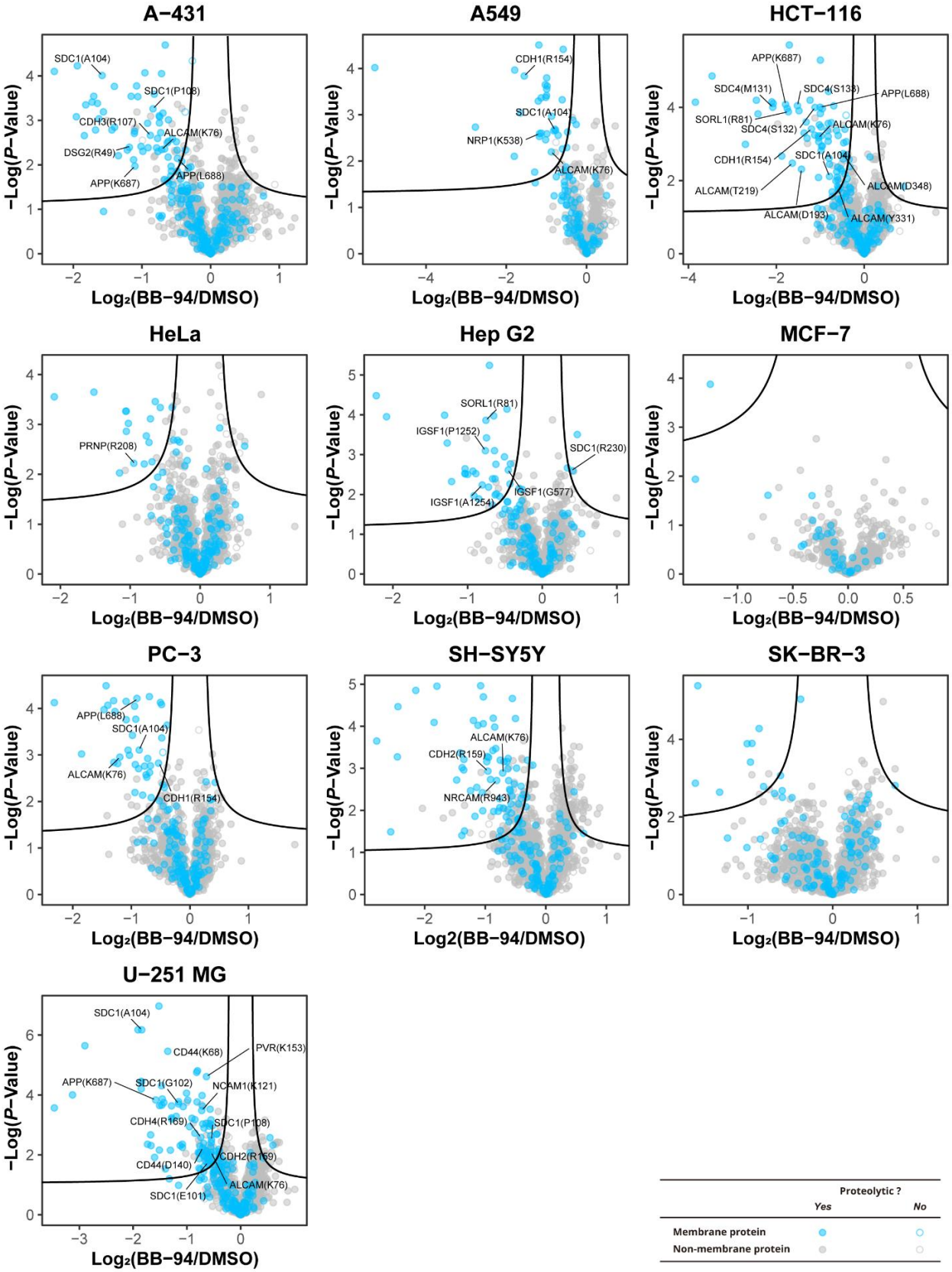

A−431
SDC1(A104)
SDC1(P108)
CDH3(R107)
ALCAM(K76)
DSG2(R49)
APP(L688)
APP(K687)
A549
CDH1(R154)
SDC1(A104)
NRP1(K538)
ALCAM(K76)
HCT−116
APP(K687)
SDC4(M131)
SDC4(S133)
APP(L688)
SORL1(R81)
SDC4(S132)
ALCAM(K76)
CDH1(R154)
SDC1(A104)
ALCAM(T219)
ALCAM(D348)
ALCAM(D193)
ALCAM(Y331)
HeLa
PRNP(R208)
Hep G2
SORL1(R81)
IGSF1(P1252)
SDC1(R230)
IGSF1(G577)
IGSF1(A1254)
MCF−7
PC−3
APP(L688)
SDC1(A104)
ALCAM(K76)
CDH1(R154)
SH−SY5Y
ALCAM(K76)
CDH2(R159)
NRCAM(R943)
SK−BR−3
U−251 MG
SDC1(A104)
CD44(K68)
PVR(K153)
APP(K687)
SDC1(G102)
NCAM1(K121)
CDH4(R169)
SDC1(P108)
CD44(D140)
CDH2(R159)
ALCAM(K76)
SDC1(E101)
−Log(P−Value)
Log₂(BB−94/DMSO)
Log2(BB−94/DMSO)
Proteolytic ?
Yes
No
Membrane protein
Non-membrane protein

Figure S4. Volcano plots for each cell line in N-terminomics, Related to Figure 1.

For quantified N-terminal peptides in the respective cell lines, volcano plots were created using Perseus with truncation at the false discovery rate of 0.05 and an artificial within groups variance (S_0) of 0.1. N-Terminal peptides of membrane proteins are highlighted in color, and proteolytic termini are highlighted with filled circles. The sites mentioned in the main text are labeled with their gene name and the P1 site.

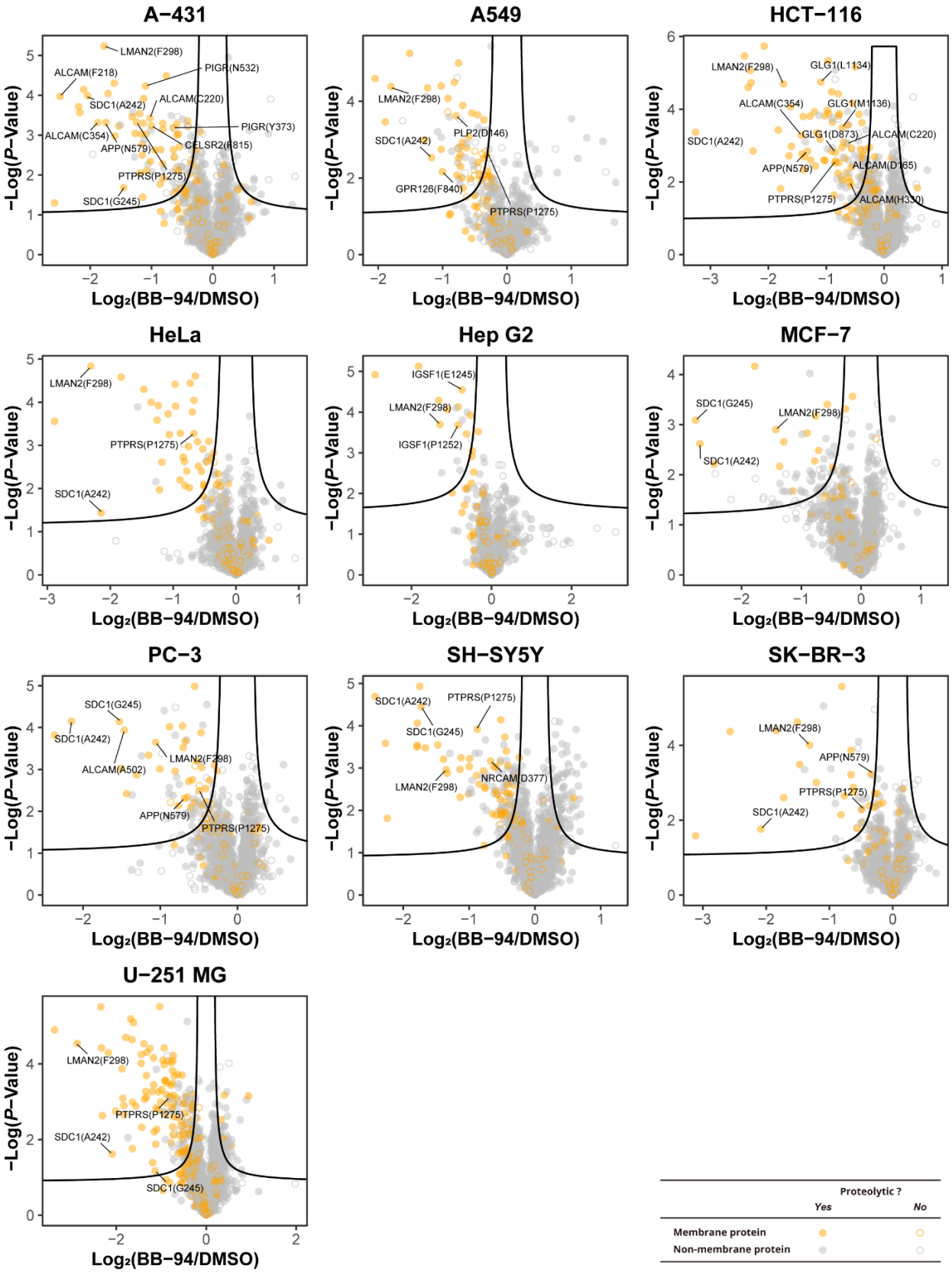

A−431
LMAN2(F298)
ALCAM(F218)
PIGR(N532)
SDC1(A242)
ALCAM(C220)
PIGR(Y373)
ALCAM(C354)
APP(N579)
CELSR2(R815)
PTPRS(P1275)
SDC1(G245)
A549
LMAN2(F298)
SDC1(A242)
PLP2(D146)
GPR126(F840)
PTPRS(P1275)
HCT−116
LMAN2(F298)
GLG1(L1134)
ALCAM(C354)
GLG1(M1136)
SDC1(A242)
GLG1(D873)
ALCAM(C220)
APP(N579)
ALCAM(D165)
PTPRS(P1275)
ALCAM(H330)
HeLa
LMAN2(F298)
PTPRS(P1275)
SDC1(A242)
Hep G2
IGSF1(E1245)
LMAN2(F298)
IGSF1(P1252)
MCF−7
SDC1(G245)
LMAN2(F298)
SDC1(A242)
PC−3
SDC1(G245)
SDC1(A242)
LMAN2(F298)
ALCAM(A502)
APP(N579)
PTPRS(P1275)
SH−SY5Y
SDC1(A242)
PTPRS(P1275)
SDC1(G245)
NRCAM(D377)
LMAN2(F298)
SK−BR−3
LMAN2(F298)
APP(N579)
PTPRS(P1275)
SDC1(A242)
U−251 MG
LMAN2(F298)
PTPRS(P1275)
SDC1(A242)
SDC1(G245)
−Log(P−Value)
Log₂(BB−94/DMSO)
Proteolytic ?
Yes
No
Membrane protein
Non-membrane protein

Figure S5. Volcano plots for each cell line in C-terminomics, Related to Figure 1.

For quantified C-terminal peptides in the respective cell lines, volcano plots were created using Perseus with truncation at the false discovery rate of 0.05 and an artificial within groups variance (S_0) of 0.1. C-Terminal peptides of membrane proteins are highlighted in color, and proteolytic termini are highlighted with filled circles. The sites mentioned in the main text are labeled with their gene name and the P1 site.

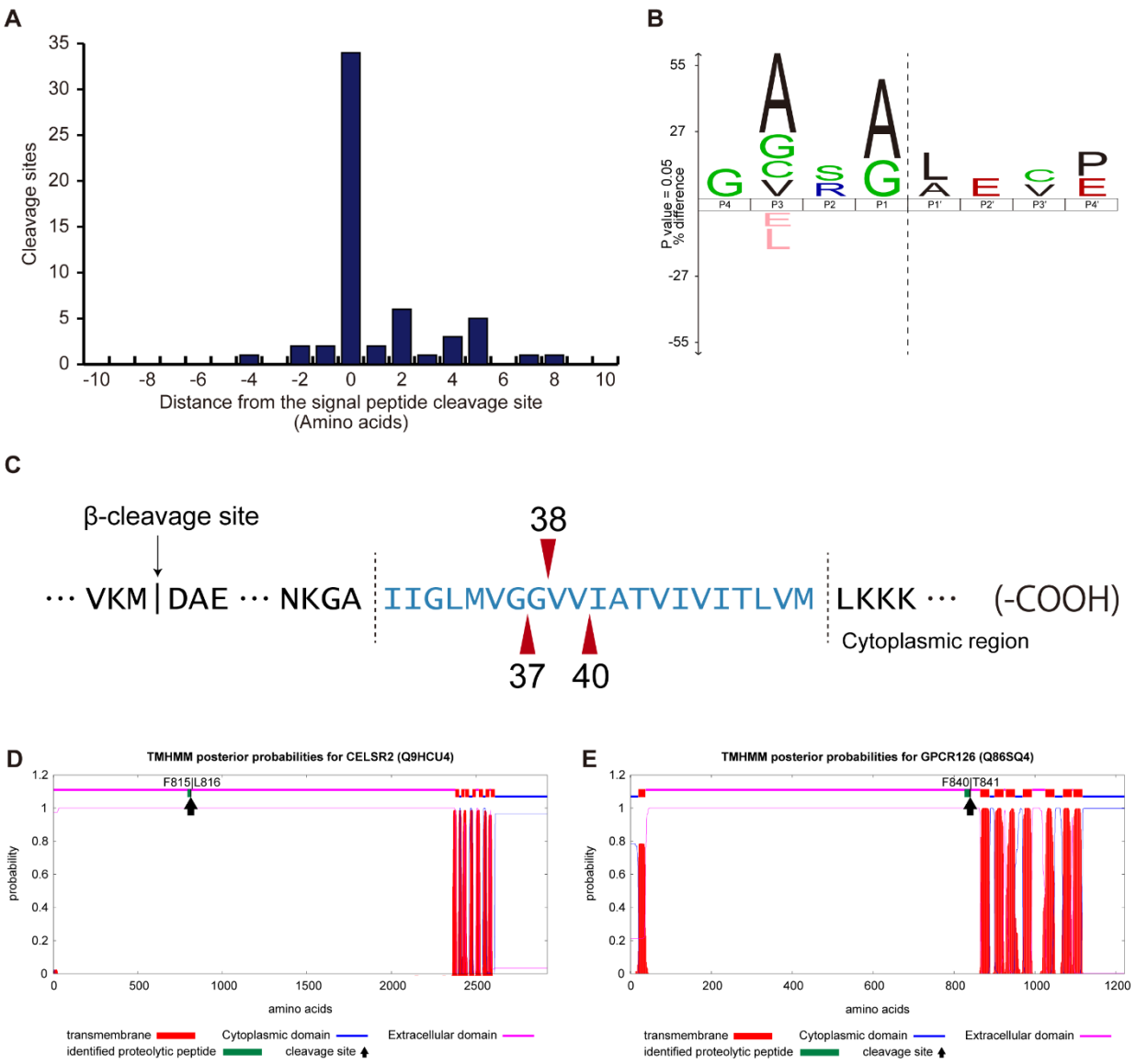


Figure S6. Signal peptide cleavage sites, APP γ-cleavage sites, and multi-pass membrane protein cleavage sites, Related to Figure 3.

(A) Distribution of the distance (number of amino acids) from the predicted signal peptide cleavage site to the identified cleavage site.

(B) Sequence logo of the cleavage sites located in the vicinity of the predicted signal peptide cleavage sites (±4 amino acids) generated by iceLogo. The dashed line shows the cleavage site.

(C) Downregulated γ-cleavage sites within the transmembrane domain of APP. The arrowheads indicate the identified cleavage sites, and the numbers indicate the length (number of amino acids) of generated amyloid β peptides.

(D, E) The extracellular domain and cytoplasmic domain were predicted for CELSR2 and GPCR126 using the TMHMM server. The identified proteolytic peptides and cleavage sites are shown together.

ADAM10
ADAM17
MMP-2
MMP-3
MMP-7
MMP-8
MMP-9
MMP-12
MMP-13
MT−MMP1
Meprin β
Furin
PCSK7
Cathepsin L
Cathepsin S
Legumain
Residue
V Y W T S P F M K L I H G E Q C D N R A
Position
P4 P3 P2 P1 P1′ P2′ P3′ P4′
Scale (log2)
−4 −3 −2 −1 0 1 2 3 4

Figure S7. PWMs for 16 selected sheddases, Related to Figure 6.

For the selected sheddases, PWMs were computed based on the sequences of ±4 amino acids flanking the substrate cleavage sites reported in MEROPS.

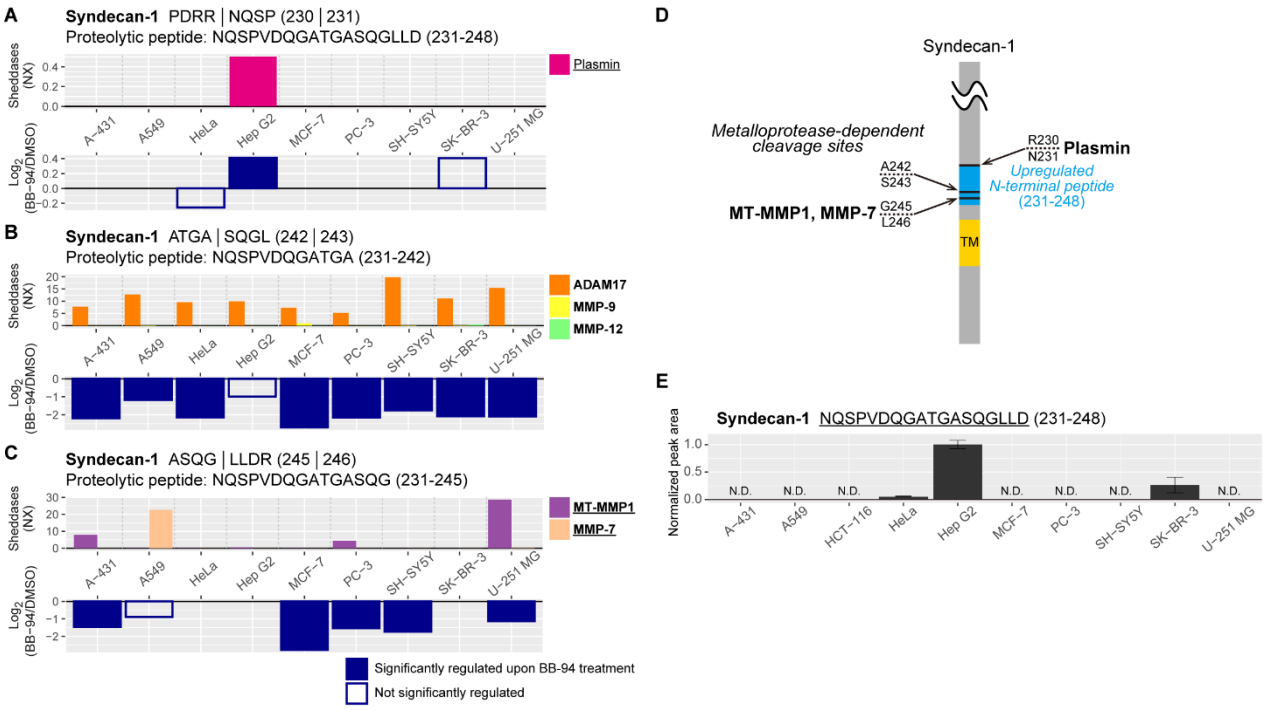


Figure S8. Sheddase expression profiles and proteolytic sites, Related to Figure 6.

(A-C) For syndecan-1 cleavage sites that have been registered with MEROPS, the quantitative results of our terminomics (BB-94/DMSO) and the mRNA expression profiles of the putative responsible sheddases (PWM score >2 in Figure 6B and/or those registered with MEROPS as responsible sheddases) obtained from the Human Protein Atlas are shown. Note that the cleavage at syndecan-1 R230↓N231 was not downregulated but upregulated by BB-94 treatment and that the expression profile in HCT-116 was not available. The sheddases with PWM score >2 are shown in bold, and the sheddases that have been registered as responsible sheddases are underlined. NX, normalized expression.

(D) Summary of syndecan-1 cleavage sites in Hep G2 cells. The identified peptide presenting the upregulated site is marked with blue. According to MEROPS, cleavage at R230↓N231 is mediated by plasmin, while cleavage at R245↓L246 is mediated by MT-MMP1 or MMP-7. The cleavage at R242↓S243 has been reported, but the responsible sheddase is unknown. TM, transmembrane domain.

(E) Peak area of the syndecan-1 proteolytic N-terminal peptide (231-NQSPVDQGATGASQGLLD-248) in each cell line (N = 3). The error bars show the standard error of the mean. N.D., not detected.

Transparent Methods

Cells

Basically, cells were cultured in Dulbecco's modified Eagle's medium (DMEM; FUJIFILM Wako) supplemented with 10% fetal bovine serum (Gibco), 100 U/mL penicillin and 100 μg/mL streptomycin (FUJIFILM Wako) on 15 cm culture dish at 37°C in a humidified incubator with 5% CO_2. PC-3 cells were cultured in Roswell Park Memorial Institute 1640 medium (RPMI1640; FUJIFILM Wako) with the same supplements as listed above. SH-SY5Y cells and U-251 MG cells were cultured in DMEM/Ham's F-12 (FUJIFILM Wako) including non-essential amino acids (FUJIFILM Wako) and the same supplements as listed above. We did not perform specific authentication of the cell lines used in this study.

Nine dishes of 90% confluent cells per condition were prepared. Supernatants of three dishes were pooled during supernatant collection to make a biological replicate. Protein yields from the supernatants for each cell line are summarized in Figure S1G.

Inhibitor treatment and supernatant preparation

Cells were washed with phosphate-buffered saline (PBS; FUJIFILM Wako) three times and incubated for 1 h with Hanks' balanced salt solution containing Ca^{2+} and Mg^{2+} (HBSS (+); FUJIFILM Wako) together with 10 μM BB-94 (Selleck) or DMSO (FUJIFILM Wako) for inhibitor pretreatment. Then, the buffer was replaced with fresh HBSS (+) containing 1 μM PMA (FUJIFILM Wako) and 20 μM BB-94 or DMSO, and the cells were incubated for 1 h. The supernatants were quickly collected on ice, and centrifuged at 7,180 *g* for 30 min to remove cell debris. Then 2 mM ethylenediaminetetraacetic acid (EDTA; DOJINDO), 2 mM ethyleneglycol bis(2-aminoethyl ether)-*N*,*N*,*N*,*N* tetraacetic acid (EGTA; DOJINDO), 1 mM phenylmethylsulfonyl fluoride (PMSF; FUJIFILM Wako) and 0.1x protease inhibitor cocktail (Sigma-Aldrich) were added to the supernatant to inhibit further proteolytic degradation, and the solution was stored at -80°C until use.

Protein purification and digestion

Culture media were concentrated up to ~50-fold using an Amicon® Ultra filter (3,000 NMWL; Merck), acidified with 2.5% trifluoroacetic acid (TFA; final concentration), and further concentrated using a SpeedVac (Thermo Fisher Scientific). Proteins were reconstituted with 100 μL water and purified by methanol-chloroform precipitation, for which the protein solution was successively mixed with 300 μL methanol, 75 μL chloroform, and 225 μL water and centrifuged at 21,000 *g* for 5 min. The resulting pellet was washed with 225 μL methanol and dissolved in phase-transfer surfactant (PTS) buffer (Masuda et al., 2009) containing 100 mM 2-amino-2-(hydroxymethyl)-1,3-propanediol hydrochloride (Tris-HCl, pH 8.5), 12 mM sodium deoxycholate (SDC; FUJIFILM Wako), 12 mM sodium *N*-lauroylsarcosinate (SLS; FUJIFILM Wako). The protein concentration was determined by means of bicinchoninic acid (BCA) assay (Thermo Fisher Scientific). For each condition, 10 μg of protein was

reduced with 10 mM dithiothreitol (DTT; FUJIFILM Wako) and alkylated with 40 mM 2-chloroacetamide (CAA; Sigma-Aldrich). Digestion was performed according to the PTS protocol (Chang et al., 2021; Masuda et al., 2009). Briefly, for C-terminomics, the protein solution was 5-fold diluted with 50 mM ammonium bicarbonate, and proteins were sequentially digested with LysC (1:50, w/w; FUJIFILM Wako) for 3 h at 37 °C and trypsin (50:1, w/w; Promega) overnight at 37 °C. In the case of N-terminomics, following 10-fold dilution with 10 mM $CaCl_2$, proteins were digested with TrypN (1:50, w/w; Protifi) overnight at 37 °C. Note that TrypN can be replaced with LysargiNase (Merck, Cat#EMS0008). Digestion was stopped with 0.5% TFA (final concentration), and the detergents were removed by liquid-liquid extraction using an equal volume of ethyl acetate. Peptides were purified with a reversed-phase SDB-XC StageTip (Rappsilber et al., 2007).

TMT labeling

Peptides were dissolved in 5 µL of 200 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES; DOJINDO)-NaOH (pH 8.5), mixed with 0.1 mg of TMT reagents dissolved in 5 µL of acetonitrile (ACN), and agitated at room temperature for 1 h. Excessive reagents were quenched with 0.33% hydroxylamine (final concentration), and the solution was acidified with 1% TFA (final concentration). Note that N-terminal peptides were TMT-labeled after N-terminal peptide enrichment, and then 6-plexed peptides were mixed. ACN was evaporated using a SpeedVac. Labeled peptides were purified using a reversed-phase SDB-XC StageTip (Rappsilber et al., 2007).

N-Terminal peptide enrichment

The TrypN digest of 10 µg protein was dissolved in 50 µL of N-loading buffer (2.5% formic acid (FA), 30% ACN) and loaded onto a StageTip with 16-gauge double-plug CationSR membranes (GL Sciences), which had been successively preconditioned with 50 µL methanol, 50 µL 80% ACN containing 0.1% TFA, 100 µL 500 mM ammonium acetate containing 30% ACN, and 300 µL N-loading buffer. Following peptide loading, the StageTip membranes were washed with 50 µL N-loading buffer, and the wash solution was collected together with the flow-through fraction as the N-terminal peptides-enriched fraction. This fraction was evaporated using a SpeedVac and TMT-labeled as described above.

C-Terminal peptide enrichment

A mixture of 6-plexed TMT-labeled peptides was dissolved in 150 µL of C-loading buffer (0.15% TFA, 30% ACN), divided into three parts, and subjected to C-terminal peptide enrichment using three StageTips. A solution of peptides in 50 µL C-loading buffer was loaded onto a StageTips with 16-gauge double-plug CationSR membranes (GL Sciences), which had been successively preconditioned with 50 µL methanol, 50 µL 80% ACN containing 0.1% TFA, 100 µL 500 mM ammonium acetate containing 30% ACN, and 300 µL C-loading buffer. Subsequently, peptides were eluted with 0.5% TFA containing 30% ACN (0.5% TFA fraction). The flow-through fraction and the

0.5% TFA fraction were separately collected and evaporated.

Liquid Chromatography/Tandem Mass Spectrometry

For N-terminal peptide analysis, one-third of the peptides was injected, and triplicate analyses were carried out (3 LC/MS/MS runs per cell line). For C-terminal peptide analysis, three sets of flow-through and 0.5% TFA fractions were subjected to single shot analysis (6 LC/MS/MS runs per cell line) (Figure S1C, D).

A nanoLC/MS/MS system comprising an UltiMate 3000RSLCnano pump (Thermo Fisher Scientific) and an Orbitrap Fusion Lumos tribrid mass spectrometer (Thermo Fisher Scientific) was employed. Peptides were injected by an HTC-PAL autosampler (CTC Analytics), loaded on a 15 cm fused-silica emitter self-packed with ReproSil-Pur C18-AQ (3 µm; Dr. Maisch), and separated by a linear gradient, that is, 5% B for 1 min, 5−15% B in 4 min, 15−40% B in 100 min, 40−99% B in 5 min, and 99% B for 10 min (Solvent A, 0.5% acetic acid; solvent B, 0.5% acetic acid in 80% ACN) at the flow rate of 500 nL/min. Peptides were ionized at 2,400 V. All MS1 spectra were acquired over the range of 375–1500 *m/z* in the Orbitrap analyzer (resolution = 120,000, maximum injection time = 50 ms, automatic gain control = 4e5). For the subsequent MS/MS analysis, precursor ions were selected and isolated in top-speed mode (cycle time = 3 sec, isolation window = 1.4 m/z), activated by higher-energy collisional dissociation (HCD; normalized collision energy = 38), and detected in the Orbitrap analyzer (resolution = 50,000, maximum injection time = 105 ms, automatic gain control = 1e5). Dynamic exclusion time was set to 30 sec.

LC/MS/MS raw data processing

LC/MS/MS raw data were processed using MaxQuant (v.1.6.7.0) with the Andromeda search engine (Cox and Mann, 2008; Cox et al., 2011). Database search was implemented against the UniprotKB/SwissProt (2019_3) human database including isoform sequences concatenated with commonly observed contaminant protein sequences set in MaxQuant. Two analysis groups were made in MaxQuant, enabling one combined analysis for TrypN with N-terminal free semi-specificity and trypsin/P with C-terminal free semi-specificity. The following parameters were applied: 10-plexed TMT quantification at the MS2 level, precursor mass tolerance of 4.5 ppm, fragment ion mass tolerance of 20 ppm, and minimal peptide length of 7 amino acids. Cysteine carbamidomethylation was set as a fixed modification, while methionine oxidation and acetylation on the protein N-terminus were allowed as variable modifications. False discovery rates were estimated by searching against a reversed decoy database and filtered for <1% at the peptide-spectrum match and protein level.

Protein inference

In terminomics, isoforms are often indistinguishable due to limited sequence coverage. Thus, we used the canonical isoform for bioinformatics analysis of membrane proteins, if the peptide can be derived from the canonical isoform. If not, we employed the leading protein given by MaxQuant.

Peptide list processing and normalization of TMT-reporter intensities

Firstly, peptides derived from contaminant proteins such as keratins and bovine serum-derived proteins were excluded. We utilized peptides with no missed cleavages for analyses. TMT-reporter intensities were log_2-converted and normalized in individual cell lines by the trimmed mean of M values (TMM) method (Robinson and Oshlack, 2010) on an R framework with the Bioconductor edgeR package (v. 3.28.1) using the default parameters. As mentioned in the main text, N-terminally acetylated N-terminal peptides are never quantified in our N-terminomics workflow, since acetylated N-terminal peptides should not have TMT-reactive sites. Nevertheless, TMT-reporter signals were sometimes observed for such peptides, e.g. due to co-fragmentation. We manually excluded 77 such N-terminally acetylated peptides, reducing the number of total identified N-terminal peptides from 6,189 to 6,181.

Statistical and bioinformatics analysis

Using Perseus (v.1.6.7.0) (Tyanova et al., 2016), we performed two-sample Welch *t*-tests to identify terminal peptides with a significant alteration upon BB-94 treatment employing a 5% permutation-based FDR filter. Proteolytic events were considered significantly altered if at least one of the N- and C-terminal peptides was significantly altered.

Membrane proteins were annotated with the UniProt Knowledgebase (UniProtKB) keywords “transmembrane” or “GPI-anchor”. Membrane protein topologies and positions of the transmembrane domain and GPI-anchor site were retrieved from UniProtKB, and further analyses were performed using Microsoft Excel.

Other statistics analyses were performed in the R framework (v.3.6.1) with the exactRankTests package (v.0.8-31), psych package (v.1.9.12.31) and pheatmap package (v.1.0.12). Bioinformatics analysis was performed using DAVID (v.6.8) (Huang et al., 2009), STRING (v.11.0) (Szklarczyk et al., 2019), and Cytoscape (v.3.8.0) (Shannon et al., 2003). Sequence logos were created using IceLogo (v.1.0) (Colaert et al., 2009).

PWM construction and scoring

Flanking sequences surrounding substrate cleavage sites for sheddases were obtained from the MEROPS database (Rawlings et al., 2018) and the previous study (Tucher et al., 2014). Sequences including non-natural amino acids were excluded. PWM construction and PWM score computation were performed essentially as described previously (Imamura et al., 2017). Briefly, the PWM score was computed by summing log_2 weights for the flanking ±4 residues surrounding each cleavage site.

In vitro sheddase assay

The substrates were designed as the flanking ±4 residues surrounding each cleavage site. Three substrate sequences were concatenated and synthesized as a 24-mer peptide. *In vitro* protease reaction was performed by incubating 10 pmol substrate peptide with 100 ng recombinant sheddase (all from R&D Systems) in 30 μL reaction buffer as instructed by the manufacturer at 37 °C for 3 h. If necessary, proteases were activated just before the reaction according to the manufacturer's instructions. The reaction was halted by acidification of the mixture with TFA. The products were desalted with SDB-XC StageTip, and analyzed using the LC/MS/MS system described above, with a short gradient: 5−40% B in 20 min, 40−99% B in 1 min, and 99% B for 4 min (Solvent A, 0.5% acetic acid; solvent B, 0.5% acetic acid in 80% ACN) at the flow rate of 500 nL/min. Peak area quantification was performed using Skyline (v20.2).

Supplemental References

Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009). Improved visualization of protein consensus sequences by iceLogo. Nat. Methods *6*, 786–787.

Masuda, T., Saito, N., Tomita, M., and Ishihama, Y. (2009). Unbiased quantitation of Escherichia coli membrane proteome using phase transfer surfactants. Mol. Cell. Proteomics *8*, 2770–2777.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A software Environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.