Comparative Secretomics Gives Access to High Confident Secretome Data: Evaluation of Different Methods for the Determination of Bona Fide Secreted Proteins

Gereon Poschmann, Katrin Brenig, Thomas Lenz, and Kai Stühler*

Secretome analysis is broadly applied to understand the interplay between cells and their microenvironment. In particular, the unbiased analysis by mass spectrometry-based proteomics of conditioned medium has been successfully applied. In this context, several approaches have been developed allowing to distinguish proteins actively secreted by cells from proteins derived from culture medium or proteins released from dying cells. Here, three different methods comparing conditioned medium and lysate by quantitative mass spectrometry-based proteomics to identify bona fide secreted proteins are evaluated. Evaluation in three different human cell lines reveals that all three methods give access to a similar set of bona fide secreted proteins covering a broad abundance range. In the analyzed primary cells, that is, mesenchymal stromal cells and normal human dermal fibroblasts, more than 70% of the identified proteins are linked to classical secretion pathways. Furthermore, 4–12% are predicted to be released by unconventional secretion pathways. Interestingly, evidence of release by ectodomain shedding in a large number of the remaining candidate proteins is found. In summary, it is convinced that comparative secretomics is currently the method of choice to obtain high-confident secretome data and to identify novel candidates for unconventional protein secretion which have been neglected so far.

1. Introduction

A major issue in secretome analysis is to distinguish proteins actively secreted by cells from contaminating proteins derived from the culture medium or proteins released from dying cells. This is particularly important because a multitude of studies revealed that a significant number of proteins are secreted by hitherto unknown unconventional secretory pathways.[1] While proteins destined for the classical secretory pathway are tagged by an N-terminal signal motif directing them to the ER/Golgi dependent secretion route, such specific sequence motifs have not deciphered yet for proteins released by unconventional secretory pathways.[2] Several approaches relying on experimental or in silico methods or a combination of both have been applied to reveal contaminating proteins and remove them from the list of bona fide secreted proteins.

The in silico methods rely on database annotations or are so far mainly based on protein sequence information which has been used for the development of prediction tools like SignalP,[3] Phobius,[4] SPOCTOPUS,[5] SecretomeP,[6] OutCyte,[7] SecretomeP, or database annotations.[8] Analyzing data from transcriptome and proteome studies, Uhlen and colleagues predicted that the human “classical” secretome consists of 2641 proteins. Their prediction relied on protein isoforms and UniProt KB (knowledge base) annotations.[9]

The experimental proteomics methods can be separated into three different approaches: i) using shotgun mass spectrometry-based proteomics to identify a high number of proteins,[10] ii) combining shotgun MS approaches and in silico approaches,[11] or iii) comparing quantitative data from conditioned medium (CM) and cell lysates to determine proteins significantly enriched in the CM.[12–14] The latter approach was used by Loei and colleagues who applied a 2D-LC-MS/MS-based setup in combination with iTRAQ labeling to compare the CM and cell lysate of gastric cancer cell lines. They determined 90 bona fide secreted proteins considering a ratio CM/lysate larger than 1.5.[13] The so-called localized statistics of protein abundance distribution (LSPAD) workflow compares mass spectrometric spectral counting data from CM and cell lysates to determine proteins enriched in the CM.[15] In a workflow employing LSPAD (albeit with less stringent cut-off), 382 bona fide secreted...
proteins were identified from CM of the lung cancer cell line A549.\(^{[14]}\) Furthermore, we have established an experimental approach to analyze protein secretion by comparing data from label-free quantification (LFQ) of proteins from the CM with cellular protein lysates.\(^{[12]}\) This approach has been applied to murine skeletal muscle cells, where the significance analysis of microarrays algorithm (SAM)\(^{[16]}\) was applied to determine 672 proteins as secreted due to their higher abundance in the CM compared to the cellular proteome. Brefeldin A-based blocking experiments led to a sensitivity estimation of >80% for the detection of classically secreted proteins.\(^{[12]}\)

Here, we compare three different quantitative methods making use of data from CM and cell lysates to define bona fide secreted proteins. We revealed that comparative secretomics gives access to high-confidence secretomes, and that the applied methods result in similar sets of proteins among the different cell lines. Nevertheless, the proportion of proteins predicted to be classically or unconventionally secreted varies from cell line to cell line. Furthermore, among the bona fide secreted proteins which could not be attributed to classical or unconventional secretion pathways, we detected candidate proteins released by ectodomain shedding.

2. Results

The characterization of the secretome is a current topic in biology and several approaches utilizing experimental and bioinformatic strategies alone or in combination are broadly applied. Here, we evaluated three published quantitative data analysis approaches: fold change (FC),\(^{[13]}\) SAM,\(^{[12]}\) and LSPAD\(^{[14]}\) and compared them with two qualitative methods: conditioned medium all (CMA) and conditioned medium only (CMO) (Table 1 and Data S1–S5, Supporting Information). The FC and SAM methods are based on LFQ intensities and LSPAD on peptide spectrum matches.

**Table 1.** Different data analysis methods for secretome characterization.

| Method                        | Quantification/identification method | Parameter                       | Comment                                                                 |
|-------------------------------|-------------------------------------|---------------------------------|-------------------------------------------------------------------------|
| Fold change (FC)              | LFQ intensity                       | FC > 1.5\[^{[13]}\]            | Proteins were filtered for showing at least four valid LFQ intensity values in either the conditioned medium or lysate. Proteins were considered as secreted when showing a FC (mean LFQ intensity conditioned medium)/ (mean LFQ intensity lysate) > 1.5 or which exclusively show LFQ intensities in the conditioned medium. |
| Significance analysis of microarrays (SAM) | LFQ intensity                       | FDR ≤ 5% \[^{[12]}\] | Proteins were filtered for showing at least four valid LFQ intensity values in either the conditioned medium or lysate. Proteins showing a significantly higher abundance in the conditioned medium, as determined by SAM analysis, were considered as secreted proteins. |
| Localized statistics of protein abundance distribution (LSPAD) | Peptide spectrum matches            | \(p\)-value < 0.01\[^{[12]}\] | Proteins were filtered for showing at least four valid MS/MS count values in either the conditioned medium or lysate. Based on the sums of the peptide spectrum matches, Fisher’s exact test \(p\)-values were calculated from log\(_2\) scaled and mean intensity windowed values. Proteins with a \(p\)-value derived \(p\)-value < 0.01 were considered as secreted. |
| Conditioned medium all (CMA)  | Four of five valid intensity values | This method is of low stringency toward the characterization of the secretome and based on all proteins found in the conditioned medium with at least four of five valid intensity values. |
| Conditioned medium only (CMO) | Four of five valid intensity values | This method represents the highest stringency and considers only proteins found exclusively in the conditioned medium with at least four of five valid intensity values, but with no valid intensity value in the cell lysate samples. |

### Significance Statement

Proteins released by cells to their surrounding are important in signaling and communication processes. Therefore, the detailed characterization of cellular secretomes—for example, by comparative secretome analysis methods—is important for a reliable identification of functionally relevant proteins and hitherto neglected secreted proteins. For a large number of proteins, the mode of secretion is not characterized in detail and the identification of those proteins will help to better understand the underlying unconventional secretory pathways. Furthermore, secretome analysis will help to detect novel biomarker candidates and the origin of their release. Moreover, a better understanding of secretomes and secretion processes will help to expand the druggable proteome and to access novel protein functions outside the cell.

As secretomes from different cell lines might display special characteristics, we considered three different human cell types as different experimental models: The human lung adenocarcinoma cell line A549 and the two primary cell lines normal human dermal fibroblasts (NHDF) and human mesenchymal stromal cells (MSC). For each cell line, we prepared five independent culture experiments and collected five replicates of CM and cells.

2.1. Characterization of the A549 Secretome by Different Data Analysis Approaches

First, we looked at the total number of bona fide secreted proteins by the different approaches (Figure 1A and Data S3–S5, Supporting Information). For A549 cells, we identified 1580 bona fide secreted proteins according to the CMA approach. The
Figure 1. Determination of bona fide secreted proteins by different data analysis approaches. A) Number of bona fide secreted proteins by cell line (A549, MSC, NHDF) and method (CMA, CMO, FC, LSPAD, and SAM). B) Venn diagrams showing the overlap of the bona fide secreted proteins identified by the three quantitative methods (FC, LSPAD, and SAM) for each analyzed cell line (A549, MSC, NHDF). C) Proportion of classically (SP, no TM, no KDEL) and unconventionally secreted proteins (UPS) in comparison to all bona fide secreted proteins in each analyzed cell line obtained by the different methods (CMA, CMO, FC, LSPAD, and SAM).

The comparison between the CM and cell lysate data revealed that only 93 proteins were exclusively found in the CM (CMO). The numbers of bona fide secreted proteins from the approaches relying on the comparison of quantitative data from CM and cell lysate were in a close range between 551 and 562 proteins with a high overlap of 485 proteins found with all three methods (Figure 1A, B). The quantitative data approaches determined only 3 (FC), 5 (SAM), and 64 proteins (LSPAD) exclusively for the respective method. If considering the different secretory pathways likely to contribute to the secretome, 68–69% of the bona fide secreted proteins of the quantitative data dependent methods were assigned to either classical (SP, not TM (transmembrane region), no KDEL) or unconventional protein secretion (UPS) (Figure 1C and Data S3, Supporting Information). The CMO method reached the highest proportion (77%) and the CMA method exhibited the lowest proportion 46% of annotated classically and unconventionally secreted proteins for the A549 secretome (Figure 1C and Data S3, Supporting Information).

2.2. Characterization of MSC Secretome by Different Data Analysis Approaches

The MS analysis of the MSC CM led to the identification of 411 bona fide secreted proteins according to the CMA approach. Only 19 proteins were exclusively found in the CM (CMO). For the approaches relying on the quantitative CM and lysate comparison, we determined 104–123 bona fide secreted proteins.
(Figure 1A) with a high overlap of 93 proteins (Figure 1B and Data S3, Supporting Information). Exclusively, we determined 3 (FC), 2 (SAM), and 27 (LSPAD) proteins by the three quantitative methods. Considering the different secretory pathways, the quantitative methods as well as the CMO approach resulted in a high proportion of candidate secretory proteins (FC 87%, LSPAD 84%, SAM 88%, CMO 84% classically secreted and UPS proteins, Figure 1C and Data S3, Supporting Information). The CMA method exhibited only 58% of classically and unconventionally secreted proteins among the bona fide secreted proteins.

2.3. Characterization of NHDF Secretome by Different Data Analysis Approaches

For the secretome of the primary NHDF cells, we identified 731 proteins by the CMA approach, whereas only 43 proteins were exclusively found by CMO. The numbers of bona fide secreted proteins determined by the quantitative methods were also in a comparable range between 224 and 249 proteins (Figure 1A and Data S3, Supporting Information) with a high overlap of 199 proteins (Figure 1B). Thirteen (FC), 1 (SAM), and 20 (LSPAD) proteins were exclusively predicted by the respective quantitative approach (Figure 1B). The proportion of candidate secretory proteins among the proteins detected by all three quantitative methods as well as CMO was quite high (FC 83%, LSPAD 79%, SAM 82%, CMO 84%). In contrast, only 55% of bona fide secreted proteins are predicted to be classically and unconventionally secreted if all proteins found in the CM were considered (Figure 1C and Data S3, Supporting Information).

2.4. Further Performance Comparison to the Applied Methods

Next, we analyzed the abundance range of the bona fide secreted proteins determined by the quantitative methods. In all three cell lines, protein abundances were detected over almost four orders of magnitude (Figure 2). Only in the very low abundance range no or only a few proteins were detected. Moreover, we found that the vast majority of proteins determined to be bona fide secreted is predicted to be classically or unconventionally secreted or contains a transmembrane domain (A549 84.5%, MSC 100%, NHDF 90.5%) whereas for the proteins which are not secretory candidates the proportion of this group is noticeably smaller (A549 58.8%, MSC 56.9%, NHDF 51.7%).

Next, we applied an approach for unsupervised performance evaluation of the different methods, the ROC analysis, where proteins predicted to be classically secreted (SP, no TM, no KDEL) are considered as true positives (Data S6, Supporting Information). Here, we confirmed the capability of the quantitative approaches to determine bona fide secreted proteins and revealed with an AUC at 5% FPR (false positive rate) of 0.018–0.026 a superior performance in comparison to CMA and CMO (AUC at 5% FPR of 0.004–0.014) (Data S6, Supporting Information).

2.5. Ectodomain Shedding Contributes to the Group of Novel Bona Fide Secreted Proteins

The quantitative comparison between CM and lysate revealed a high proportion (68–90%) of predictable secretory proteins (classically secreted and UPS). Nevertheless, we revealed by all three quantitative methods a group of bona fide secreted proteins which were not attributed to a secretory pathway so far (A549 148, MSC 11, and NHDF 37 proteins; Data S3, Supporting Information). One example from the A549 dataset is the known UPS protein HMGB1 which has been shown to be misclassified by OutCyte.[17] Further examples are well-known proteins with extracellular function like disintegrin and metalloproteinase domain-containing protein 10, High mobility group protein B2, Sulfhydryl oxidase 1, and Vascular cell adhesion protein 1 (Data S3, Supporting Information). Noteworthy, 38% (A549), 91% (MSC), and 49% (NHDF) of those unassigned proteins exhibit an annotated transmembrane domain, and hence are candidates for ectodomain shedding. To verify this hypothesis, we plotted the peptides identified by our mass spectrometric analysis using the lysate and secretome peptide feature plotter (LSPFP).[12] We only detected peptides of the extracellular but not intracellular domain for four transmembrane proteins (gene names: VASN, CDH2, CDH11, DAG1) in the CM of MSC and NHDF cells. That the peptides of the intracellular domain are detectable per se was confirmed by the parallel MS analysis of the cell lysate samples and therewith revealed these proteins are candidates of ectodomain shedding (Figure 3). For the A549 cells, we detected 11 proteins (gene names: DCBD2, UFO, TNR1A, LDLR, MET, CDH2, ANTR1, NRP1, DAG1, SHPS1, CRIM1) as candidates of ectodomain shedding due to the exclusive detection of peptides in the extracellular domain in the CM of A549 cells (Data S7–S9, Supporting Information). Thus, we demonstrated that comparative secretomics gave access to novel candidate proteins released by alternative pathways like, for example, ectodomain shedding.

3. Discussion

Our results revealed that comparative secretomics gave access to high-confident secretomes and to novel candidates of protein secretion, regardless of the underlying quantitative or statistical method.

All three applied quantitative methods revealed similar sets of bona fide secreted proteins. Furthermore, there was a high percentage (68–90%) of classically and unconventionally secreted proteins among the bona fide secreted proteins which agree with the stringent CMO method and other publications. For instance, using the LSPAD approach, Luo et al. showed that 85.3% of the 382 bona fide secreted proteins of the A549 secretome were annotated as secreted.[11] We also provide evidence that approaches aiming for the shot-gun analysis of CM proteins, like e.g., the CMA method, lead to a high percentage of candidate proteins (40–50%; Data S2–S4, Supporting Information) for which no secretion is predictable. Those proteins might have a high likelihood of being false-positive bona fide secreted proteins (contaminations). This agrees with previous work from our group where the CMA method in combination with bioinformatic annotation analysis has been applied to CM proteins of unrestricted somatic stem cells. Here, 661 out of 1520 identified proteins (44%) were assigned to classical and unconventional secretory pathways.[11] This clearly demonstrates that the quantitative approaches might be valuable especially in cases of cancer cell lines (like the analyzed A549 cancer cells) as those
Figure 2. Intensity distribution of bona fide secreted proteins (triangles) and proteins not predicted to be secreted (circles) determined by the overlap of the three quantitative methods (FC, LSPAD, and SAM). A) Assignment of all bona fide secreted proteins with relevant annotation (SP no TM, no KDEL; SP+TM; SP+KDEL; TM, no SP; UPS; no relevant annotation). B) Proteins not predicted to be secreted with relevant annotation (SP no TM, no KDEL; SP + TM; SP + KDEL, TM, no SP; UPS (OutCyte); no relevant annotation) in the analyzed cell lines (A549, MSC, NHDF).
are known to secrete a larger proportion of unconventionally secreted proteins\cite{17} and might help to “clean up” secretomes which are affected by a higher proportion of contaminants, for example, released from dying or broken cells. In this context, the predictive power of pure bioinformatic approaches should not be overestimated. Although the fact that the prediction of UPS has improved,\cite{7} the in silico annotation is still error prone. Moreover, we cannot be excluded that also the prediction of signal peptides and therewith the assignment to the classically secretory pathway is partly incorrect as, for example, isoform information could only partly be considered in shotgun approaches.

Nevertheless, limitations of comparative secretomics should be mentioned. The comparison of two quite different quantitative datasets, that is, CM and cell lysate, is challenging which inherently leads to stringent filter criteria and therewith to the loss of bona fide secreted proteins. The choice of parameters like cutoffs or the way of handling missing data points largely affects the accuracy and number of candidate proteins. In addition, future data based approaches should consider that a relevant amount of secreted proteins in the cell lysate might result in false negative assignments. We already estimated an amount of about 20% of false negative bona fide secreted proteins among classically secreted proteins released by C2C12 cells.\cite{12} Finally, more sensitive measurement methods\cite{18} and methods reducing the high number of missing values, for example, data independent acquisition\cite{19} will help to increase the total number of detected and bona fide secreted proteins. Moreover the secretion rate varies among cell types and amounts of proteins in the CM additionally depend on incubation length, protein degradation, and factors like cytokines and nutrients in the medium.\cite{20}

Beside the described methods, other approaches have been successfully applied to define bona fide secreted proteins and give access to unconventional secreted proteins. For example, Eichelbaum and Krijgsveld developed an approach combining SILAC labeling and click chemistry enabling the detection of bona fide secreted protein even under conditions including FCS.\cite{21} This approach has recently also applied to cardiomyocytes for which conditions including serum are beneficial for

---

Figure 3. Candidate proteins of ectodomain shedding. The sequence coverage in CM and cell lysate samples is plotted for several proteins. Extracellular regions are marked by a light-red background. A) Vasoerin; B) Dystroglycan; C) Cadherin-11; and D) Cadherin-2.
retaining important cellular characteristics.[22] An alternative approach developed by Stiess et al. made use of dual SILAC labeling strategy to detect bona fide secreted proteins by a quantitative comparison of CM and cell lysates as performed in the quantitative approaches in this work but using SILAC based quantification. Here, two channels were used to build ratios between CM and cell lysate derived signals and a third channel (light) was used to identify contaminating serum proteins.

Moreover, the secretome protein enrichment with click sugars (SPECS) method has been developed. This method detects glycosylated secreted proteins in CM after metabolomics labeling even in the presence of serum.[21] As only glycosylated proteins can be identified this method might miss unconventionally secreted proteins which are not glycosylated in the ER/Golgi. We compared our lists of bona fide secreted proteins (SAM) with a list of 1275 proteins identified using SPECS from HEK cells[21] and found overlaps of 53% (A549), 67% (MSC) and 57% (NHDF), respectively. The highest overlap is found with MSC cells which bona fide secreted proteins also contain the highest proportion of predicted classically secreted proteins which seems to be plausible as these proteins might be selectively detected by SPECS.

We demonstrated that the applied quantitative methods might give access to highly confident candidate proteins of alternative secretory pathways. Especially, they provide information about candidate proteins released by ectodomain shedding into the secretome which became in particular evident for both primary cell lines (MSC and NHDF). One example is vasorin, which was identified with exclusively extracellular located peptides in the CM of NHDF and MSC. Its extracellular domain has been shown to be released by the disintegrin and metalloproteinase domain-containing protein 17 (ADAM17).[24] The same matrix-metalloprotease (MMP) activity was reported for VCAM1 for cells are known to be unconventionally secreted.[30,31] but are misclassified by OutCyte so far. HMGB1 and its family member HMGB2 which we have also detected in the secretome of A549 cells are known to be unconventionally secreted[30,31] but are misclassified by OutCyte.[7] HMGB1, a non-histone nuclear factor, acts extracellularly as a mediator of delayed endotoxin lethality and its secretion is induced by stimuli triggering lysosome exocytosis.[30] We speculate that the physicochemical signature of some unconventionally secreted proteins like the High Mobility Group B proteins are probably underrepresented in currently available secretome data and therefore should be considered in the development of advanced prediction tools in the future.

In summary, we are convinced that comparative secretomics is currently the method of choice to obtain high-confident secretome data and to identify novel candidates of protein secretion which were neglected until now.

4. Experimental Section

Study Design: Several methods have been developed to characterize the proteins secreted or released by cultured cells. Some of those rely on a mass spectrometry proteomics-based quantitative comparison of protein intensities from the analysis of CM and cell lysates to exclude contaminants and to define bona fide secreted proteins.[12-15] In the present study, quantitative mass spectrometry-based proteomics were used to analyze the CM and lysate of three different cell lines: human A549 cells (human alveolar basal epithelial adenocarcinoma cells), MSC, and NHDF. Based on the experience with quantitative data of cell culture samples and analogously previously published approaches,[12-15] five replicates were selected per cell line which were cultured in different dishes and which were harvested and processed independently. Furthermore, the difference in protein intensity distribution between CM and cell lysate is quite pronounced so that expectedly a large number of proteins will be assigned to the CM.

Following sample preparation and mass spectrometric analysis of CM and cell lysates, five different published approaches were applied for the detection of bona fide secreted proteins. These strategies count identified proteins (CMA, CM0), others use fixed cutoff-values (LSPAD, FC), or rely on a 5% FDR (SAM).

CM and Cell Lysate Preparation: The lung adenocarcinoma cell line A549 (ATCC CCL-185) and primary NHDF (PromoCell C-12300) were cultured in a humidified incubator at 37 °C and 5% CO2 in DMEM including 4.5 g L−1 glucose, 10% FCS, and for NHDF, 100 units penicillin and 100 mg L−1 streptomycin. At cell densities of 1.1 × 105 cells cm−2 (A549) and 0.2 × 105 cells cm−2 (NHDF), the culture medium was replaced with medium without additives after washing the cells on culture dishes four times with medium without additives. CM and cells were harvested after 17 h (A549 cells) and 21.5 h (NHDF) and 48 h (MSC, please see below). This reflects on the one hand the doubling times of the different used cell types and on the other hand this represents time points where we monitored only a neglectable number of dead cells using trypan blue staining. 1.5% of dead A549 and 1% of dead NHDF cells. CM and cell lysate samples were prepared for mass spectrometry as described.[12] Briefly, proteins in the CM were precipitated with trichloroacetic acid and sodium [dodecanoyl(methyl)amino]acetate and finally dissolved in urea containing buffer (30 mM Tris, 2 μM thiourea, 7 μM, 4% CHAPS, pH 8.5). From 10 mL CM, we gained in average 20 μg protein (0.4 μg μL−1) from A549 and 12 μg protein (0.24 μg μL−1) from NHDF cells. Cell lysates were prepared in the same buffer (mean concentration 4.5 μg μL−1 A549 and 4.8 μg μL−1 NHDF). Proteins were shortly separated by SDS-PAGE (about 5 mm running distance), stained with Coomassie brilliant blue and protein-containing gel pieces were further processed by reduction with dithiothreitol, alkalyation with iodoacetamide, and overnight digestion with trypsin. Resulting peptides were extracted from the gel, dried, and finally reconstituted in 0.1% TFA.

The preparation and mass spectrometric analysis of MSC lysates and CM (48 h incubation in serum-free medium) has already been described.[12] The data of MSC from five different donors acquired in Baberg et al.[12] under analogous mass spectrometric conditions as described below was reprocessed in this study.

Liquid Chromatography and Mass Spectrometric Analysis: First, 500 ng of peptides per sample were separated by a rapid separation liquid
chromatography system (Ultimate 3000, Thermo Fisher) on a C18 column (Acclaim PepMap RSLC, 2 μm particle size, 100 Å pore size, 75 μm inner diameter, 25 cm length, Thermo Fisher) using a 2 h gradient as described.[12] Second, separated peptides were injected via a nano-source ESI interface equipped with distally coated SilicaTip emitters (New Objective) into a Q Exactive plus orbitrap containing mass spectrometer (Thermo Fisher, Software: Xcalibur 4.1.3.119, SII 1.4.0.106, Q Exactive Plus MS 2.9 SP3) operated in positive mode and analyzed with a data-dependent top ten method. After recording precursor spectra (profile mode, scan range 350–2000 m/z, resolution 70 000, target for automatic gain control 3 000 000, maximum ion time 80 ms), a maximum of ten two- or threefold charged precursor ions were selected by the quadrupole (2 m/z isolation window, 17 000 intensity threshold, minimum automatic gain control target 1000), fragmented by higher-energy collisional dissociation (normalized collision energy 30), and fragments analyzed (centroid mode, available scan range 200–2000 m/z, resolution 17 500, target for automatic gain control 100 000, maximum ion time 60 ms) and excluded from further fragmentation for the next 100 s. The above-described settings were used for the analysis of MSC and NHDF cells. For A549, the settings were identical despite the resolution for precursor ion analysis which was 140 000.

Mass Spectrometric Data Analysis, Protein Identification, and Quantification: Mass spectrometric data were processed separately for each cell line with MaxQuant version 1.6.6.0 (Max Planck Institute for Biochemistry, Planegg, Germany). If not stated otherwise, standard parameters were used for protein identification and quantification. Searches were carried out based on 74 449 Homo sapiens protein entries downloaded from the UniProtKB on November 15, 2019, using tryptic cleavage specificity (behind K and R) and a maximum of two missed cleavage sites. Methionine oxidation and N-terminal acetylation as well as a carbamidomethylation at cysteine residues were considered as variable and fixed modifications, respectively. First, an initial search was carried out using a precursor mass tolerance of 20 ppm and after recalibration, a second search was performed with 4.5 ppm precursor mass tolerance. Tolerances for fragment spectra were 20 ppm. Peptides and proteins were identified with an FDR of 1%. The “match between runs” function as well as intensity-based absolute quantification and LFQ were enabled.

Determination of Secreted Proteins: The datasets of the three cell lines were filtered by importing the protein list into Perseus 1.6.6.0 (Max Planck Institute for Biochemistry, Planegg, Germany) and removing potential contaminants, decoy hits, and proteins which were identified “by site” or identified with less than one peptide. A protein was considered as potentially classically secreted if at least one of the identifiers of the protein group carries an annotated signal peptide (SP), but no TM or KDEL ER retention signal sequence were present out based on 74 449 Homo sapiens protein entries downloaded from the UniProtKB, but no TM or KDEL ER retention signal sequence were present out based on 74 449 Homo sapiens protein entries downloaded from the UniProtKB on November 15, 2019, using tryptic cleavage specificity (behind K and R) and a maximum of two missed cleavage sites. Methionine oxidation and N-terminal acetylation as well as a carbamidomethylation at cysteine residues were considered as variable and fixed modifications, respectively. First, an initial search was carried out using a precursor mass tolerance of 20 ppm and after recalibration, a second search was performed with 4.5 ppm precursor mass tolerance. Tolerances for fragment spectra were 20 ppm. Peptides and proteins were identified with an FDR of 1%. The “match between runs” function as well as intensity-based absolute quantification and LFQ were enabled.

Comparison of Secretory Protein Candidates: The resulting datasets for the different cell lines are referred to as “rough data.” The rough data are the starting point for following filter steps and the application of the five different methods to define bona fide secreted proteins (Data S1, Supporting Information). Intensity, LFQ intensity, or peptide spectrum match (MS/MS count) values >0 are referred to as “valid values.”

Two approaches are based on the occurrence of intensity values: in CM and/or lysate samples: In the CMA approach, proteins with valid intensity values in at least four out of five CM samples were considered as bona fide secreted. The CMO approach reduces the CMA set of proteins by excluding proteins with valid intensity values in one or more cell lysate samples.

For the FC approach, at least four valid LFQ intensity values had to be present in at least one group (CM or cell lysate). Proteins were considered to be bona fide secreted if the ratio (mean LFQ intensity values CM)/(mean LFQ intensity values cell lysate) was ≥1.5 or the respective protein group did not show any valid LFQ intensity value in the lysate samples.

The SAM method[16] was based on LFQ intensity values. For this approach, a minimum of four valid values had to be present in at least one group (CM or cell lysate). LFQ intensities were processed in Perseus (1.6.6.0) by log2 transformation to reach a normal distribution like data structure and missing values were filled in with random values from down-shifted normal distribution (0.3 s.d. width, 1.8 s.d. downshift). Utilizing Student’s t-tests based SAM analysis, with a constant S0 of 0.8 and 5% FDR based cut-off, protein groups showing a significantly higher abundance in the CM were referred to as bona fide secreted proteins. As this method considers both the differences between the mean values of log2 LFQ intensities as well as the standard deviation of repeated measurements, proteins were selected as bona fide secreted proteins on the basis of abundance differences between CM and lysate as well as low p-values.

The LSPAD method was performed on peptide spectrum matches and only protein groups identified with at least four valid values in at least one group (CM or cell lysate) were considered in the starting datasets. For M–A plotting, the sums of the peptide spectrum matches X1 and X2 of the CM and lysate samples, respectively, were transformed into Y1 and Y2, respectively, using the formula $Y = \log_2 (X + 1)$. The differential protein abundance $M = Y1 − Y2$ was plotted versus the average protein abundance $A = (Y1 + Y2)/2$. The LSPAD was calculated using a local window width of 1/3 of the range of A, that is, for each particular protein, all peptide spectrum matches of neighbored proteins, whose A values were within the 33% abundance-window of the particular protein were calculated as a background to evaluate the statistical significance ($p$-value) of over- or under-representation of the particular protein by performing Fisher’s exact test on the following fourfold table:

| Peptide spectrum matches of each particular protein: | X1 (cond. media) | X2 (lysates) |
|-----------------------------------------------|----------------|-------------|
| Sum of peptide spectrum matches of all the other proteins in the window: | S1 (cond. media) | S2 (lysates) |

The p-values for each particular protein derived from the Fisher’s exact test were linearly transformed into $p$-values to obtain a continuous scale ranging from 0 (highest preference for CM) over 0.5 (no preference) to 1 (highest preference for lysate) using the formula $p = (p/2)/(X1/S1 ≥ X2/S2); 1−p/2$ for $(X1/S1 < X2/S2)$.

Comparison of Secretory Protein Candidates: Counting of annotations, visualization, venn analysis (based on the majority protein IDs, package venn, version 1.9), and ROC analysis (package ROCR, version 1.0-7) were performed within the R environment.[31]

Supporting Information
Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements
The authors greatly acknowledge Marion Gehrke and Christine Hafermann for technical assistance. This study was supported by Deutsche Forschungsgemeinschaft CRC 1208 201 (to K.S.). Open access funding enabled and organized by Projekt DEAL.

Data Availability Statement
The mass spectrometry-based proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[34] partner repository with the dataset identifier PXD018895.

Conflict of Interest
The authors declare no conflict of interest.
Keywords
ectodomain shedding, method comparison, quantitative mass spectrometry, secretome, unconventional protein secretion

Received: July 3, 2020
Revised: September 23, 2020
Published online: November 17, 2020

[1] C. Rabouille, Trends Cell Biol. 2017, 27, 230.
[2] H. Nielsen, E. I. Petsalaki, L. Zhao, K. Stuhler, Biochim. Biophys. Acta, Proteins Proteomics 2019, 1867, 140174.
[3] J. J. Almagro Armenteros, K. D. Tirigos, C. K. Sonderby, T. N. Petersen, O. Winther, S. Brunak, G. von Heijne, H. Nielsen, Nat. Biotechnol. 2019, 37, 420.
[4] L. Call, A. Krogh, E. L. Sonnhammer, Nucleic Acids Res. 2007, 35, W429.
[5] H. Viklund, A. Bensel, M. Skwark, A. Elofsson, Bioinformatics 2008, 24, 2928.
[6] J. D. Bendtsen, L. J. Jensen, N. Blom, G. Von Heijne, S. Brunak, Protein Eng., Des. Sel. 2004, 17, 349.
[7] L. Zhao, G. Poschmann, D. Waldera-Lupa, N. Rafiee, M. Kollmann, K. Stuhler, Sci. Rep. 2019, 9, 19448.
[8] D. Caccia, M. Dugo, M. Callari, I. Bongarzone, Biochim. Biophys. Acta, Proteins Proteomics 2013, 1834, 2442.
[9] M. Uhlen, M. J. Karlsson, A. Hofer, A. S. Svensson, J. Scheffel, D. Kotol, W. Zhong, A. Tebani, L. Strandberg, F. Edfors, E. Sjostedt, J. Mulder, A. Mardinoglu, A. Berling, S. Ekblad, M. Dammeyer, S. Kanje, J. Rockberg, M. Lundquist, M. Malm, A. L. Volk, P. Nilsson, A. Manberg, T. Dodig-Crnkovic, E. Pin, M. Zawhren, P. Oksvold, K., von Feilitzen, R. S. Haussler, M. G. Hong, et al., Sci. Signaling 2019, 12, eaao2074.
[10] P. Mukherjee, S. Mani, Biochim. Biophys. Acta, Proteins Proteomics 2013, 1834, 2226.
[11] J. Schira, H. Falkenberg, M. Hendricks, D. M. Waldera-Lupa, G. Kogler, H. E. Meyer, H. W. Muller, K. Stuhler, Mol. Cell. Proteomics 2015, 14, 2630.
[12] L. Grube, R. Dellen, F. Kruse, H. Schwender, K. Stuhler, G. Poschmann, J. Proteome Res. 2018, 17, 879.
[13] H. Loei, H. T. Tan, T. K. Lim, K. H. Lim, J. B. So, K. G. Yeo, M. C. Chung, J. Proteome Res. 2012, 11, 1759.
[14] X. Luo, Y. Liu, R. Wang, H. Hu, R. Zeng, H. Chen, J. Proteomics 2011, 74, 528.
[15] R. X. Li, H. B. Chen, K. Su, S. L. Zhao, H. Zhou, S. J. Li, J. Dai, Q. R. Li, S. Nie, Y. X. Li, W. P. Jia, R. Zeng, J. R. Wu, PLoS One 2008, 3, e3224.