

Research Article

# *In silico* approach to predict pancreatic $\beta$ -cells classically secreted proteins

Erika Pinheiro-Machado<sup>1,2</sup>, Tatiana Orli Milkewitz Sandberg<sup>3</sup>, Celina PIHL<sup>2</sup>, Per Mårten Hägglund<sup>2</sup> and  Michal Tomasz Marzec<sup>2</sup>

<sup>1</sup>Department of Pathology and Medical Biology, University Medical Center Groningen, University of Groningen, The Netherlands; <sup>2</sup>Department of Biomedical Sciences, University of Copenhagen, Denmark; <sup>3</sup>Tel Aviv University, Israel

**Correspondence:** Michal Tomasz Marzec (michal@sund.ku.dk)



Pancreatic  $\beta$ -cells, residents of the islets of Langerhans, are the unique insulin-producers in the body. Their physiology is a topic of intensive studies aiming to understand the biology of insulin production and its role in diabetes pathology. However, investigations about these cells' subset of secreted proteins, the secretome, are surprisingly scarce and a list describing islet/ $\beta$ -cell secretome upon glucose-stimulation is not yet available. *In silico* predictions of secretomes are an interesting approach that can be employed to forecast proteins likely to be secreted. In this context, using the rationale behind classical secretion of proteins through the secretory pathway, a Python tool capable of predicting classically secreted proteins was developed. This tool was applied to different available proteomic data (human and rodent islets, isolated  $\beta$ -cells,  $\beta$ -cell secretory granules, and  $\beta$ -cells supernatant), filtering them in order to selectively list only classically secreted proteins. The method presented here can retrieve, organize, search and filter proteomic lists using UniProtKB as a central database. It provides analysis by overlaying different sets of information, filtering out potential contaminants and clustering the identified proteins into functional groups. A range of 70–92% of the original proteomes analyzed was reduced generating predicted secretomes. Islet and  $\beta$ -cell signal peptide-containing proteins, and endoplasmic reticulum-resident proteins were identified and quantified. From the predicted secretomes, exemplary conservation patterns were inferred, as well as the signaling pathways enriched within them. Such a technique proves to be an effective approach to reduce the horizon of plausible targets for drug development or biomarkers identification.

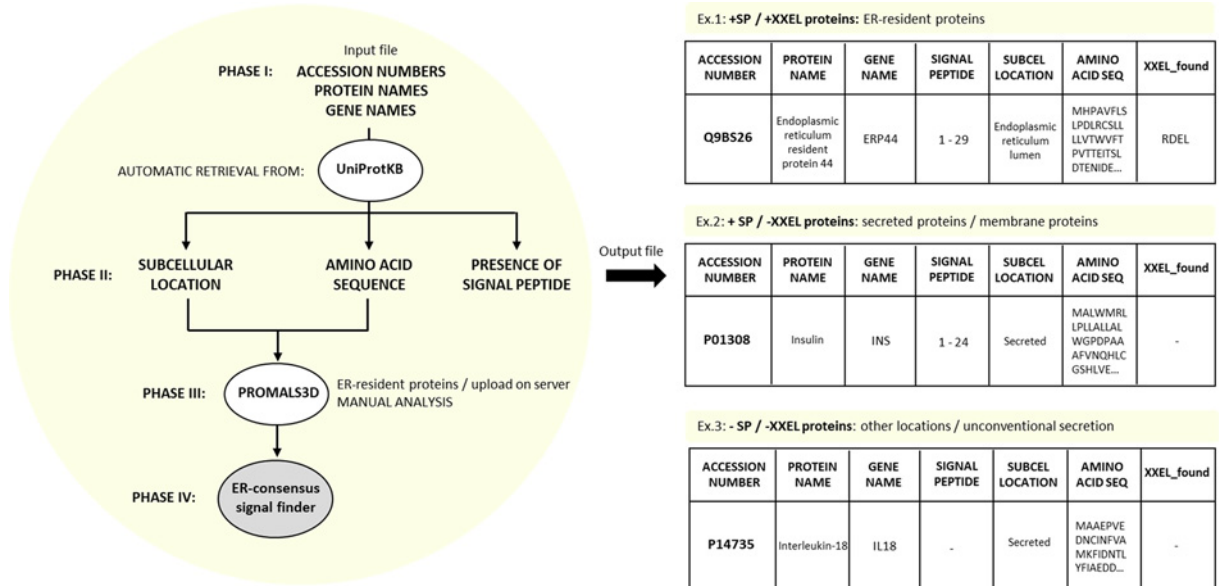
## Introduction

The term secretome was first defined as the whole subset of factors secreted by a cell [1] and later revised to 'all proteins secreted by the cell into the extracellular space' [2]. These factors can regulate a multitude of physiological processes and dictate the composition of the extracellular environment. Secretome analysis in health and disease can, therefore, bring insights about the pathophysiology of these conditions, opening new perspectives for the discovery of biomarkers and therapies [3].

Classically, proteins are secreted through the secretory pathway [4]. mRNAs are processed by ribosomes in the cytoplasm and the proteins containing a signal peptide (SP) – motif often composed by one positively charged amino acid followed by 6–12 hydrophobic amino acids [5] – are cotranslationally translocated to the endoplasmic reticulum (ER) [6]. Once there, proteins are folded by chaperones and foldases to gain proper 3D conformation [7]. The ER has its own subset of native proteins, the so-called ER-resident proteins [8]. Their continuous localization to the ER lumen is secured by an ER-retrieval signal located in the C-terminal part of the protein [8]. This signal is commonly a tetrapeptide – K (lysine/lys) D (aspartic acid/asp) E (glutamic acid/glu) L (leucine/leu) – or related sequences [8,9] and enables proteins to bind to KDEL-receptors at the early Golgi, thus preventing proteins from being secreted from cells [10].

Received: 29 October 2019  
Revised: 30 January 2020  
Accepted: 31 January 2020

Accepted Manuscript online:  
31 January 2020  
Version of Record published:  
14 February 2020



**Figure 1. The applied method workflow**

The script processes phases II and IV automatically. Phase I is done manually by the user, organizing the input file according to what is required: accession numbers, protein and gene names. Phase II (retrieval) is automatic, and the data is retrieved from UniProtKB. Phase III relies on the PROMALS3D server, in which the input data need to be manually organized and submitted. Phase IV is based on the PROMALS3D alignment results, where the ER-retrieval signal consensus is searched through all proteins in the list automatically. Examples of possible output data shown.

The non-ER-resident proteins are forwarded to the Golgi complex where different post-translational modifications occur [11]. Typically, proteins that are ready to be secreted are sorted into vesicles that: (1) can undergo secretion (exocytosis) [12]; (2) can wait for signals to undergo exocytosis [13]; or (3) can be translocated to the lysosomes for degradation [14,15].

Proteins can be secreted via conventional and unconventional secretory pathways [16,17]. The conventional secretory pathway is responsible for transport and secretion of proteins containing SP, while the mechanisms behind unconventional secretion remain elusive [18–20]. Around 39% (7649) of all human protein-coding genes (19,613) are predicted to display either SP and/or transmembrane regions [21]. The number of potentially secreted proteins (with SP and no transmembrane regions) were described to be 2623, which accounts for approximately 13% of all proteins [21,22]. Unconventionally secreted proteins often lack SP, are not localized to the secretory pathway organelles, and their secretion continues even in the presence of brefeldin A (an inhibitor of ER-Golgi transport) [23,24], imposing another layer of complexity in the secretomes' characterization.

Several approaches to define a secretome of a given cell type were employed in the past, but the introduction of proteomics greatly facilitated the overall methodology [20,25,26]. Initially, investigated with the use of 2D gel electrophoresis, radiolabeling, and enzyme activity [27,28], the secretomes were described with only a few components. Currently, at the advent of better techniques in bioinformatics and mass spectrometry (MS), the identification of hundreds to thousands of proteins is becoming routine and such approaches were successfully implemented in the description of secretomes [29–31]. However, a simple and user-friendly tool that systematically integrates experimental high-throughput proteomic datasets using proteomic data from any kind of cell type is still missing.

In the present study, a computational tool for the prediction of classically secreted proteins that only requires a list of identified proteins was developed (Figure 1). An open source Python tool that rationally integrates public databases, facilitates the analysis of protein lists, and that can be modified according to the cell type of interest, has been designed. Our method makes possible to retrieve, organize, search and filter data regarding the full amino acid sequence, protein subcellular location, presence of SP, and presence of a consensus of ER-retrieval signal (Figure 1). The approach was utilized in the analysis of available datasets from human pancreatic islets (HPI) [32], mouse pancreatic islets (MPI) [33], isolated  $\beta$ -cells [32], rat insulin secretory granules (ISG) [34],  $\beta$ -cells supernatant (SUP) [35] as well as our own supernatant of  $\beta$ -cell model cells, the insulinoma INS-1E, upon high glucose stimulation (SUP2) (Table 1). The tool predicted the classically secreted proteins by enabling clustering of proteins containing SP (+SP) and ER-resident

**Table 1** Subset of all publications that fulfill the following criteria: publication within the last 10 years; list of proteins identified available as an excel file. Publications chosen to validate our Python tool are highlighted in light yellow. N/A: non-applicable, i.e. cells were not submitted to any special experimental conditions

	Study / Author, year	Species	Condition	Technique	# Identified Proteins	Ref.
<b>Whole-cell</b>	Glucose-stimulated islet proteome / <b>Waanders et al. (2009)</b>	Mice (islets)	High glucose	LC-MS/MS (LTQ-Orbitrap)	6902	[39]
	Glucose-stimulated mice $\beta$ -cell proteome / <b>Martens et al. (2010)</b>	INS-1E	High glucose	Alternate scanning LC-MS	300	[33]
	The Human Diabetes Proteome Project / <b>Topf et al. (2013)</b>	Human (islets)	N/A	Gas-Phase Fractionation MS	5317	[32]
	The Human Diabetes Proteome Project / <b>Topf et al. (2013)</b>	INS-1E	N/A	LC-MS/MS	2625	[32]
<b><math>\beta</math>-Cell secretory granule</b>	Insulin granule / <b>Brunner et al. (2007)</b>	INS-1E	N/A	Granule purification, LC-MS/MS	130	[40]
	Insulin granule / <b>Schwartz et al. (2012)</b>	INS-1E	N/A	SILAC, 3-step gradient purification, MS/MS	140	[34]
	Insulin granule / <b>Li et al. (2018)</b>	INS-1	N/A	OptiPrep, (LC)-MS/MS, correlation profiling	81	[41]
<b><math>\beta</math>-Cell supernatant</b>	$\beta$ -Cell secretome / <b>Tattikota et al. (2013)</b>	MIN6 cells supernatant	High Glucose	Concentration (3MWKO), EASY-nLC MS/MS	1629	[42]
	$\beta$ -Cell secretome / <b>Pepaj et al. (2016)</b>	INS-1E cells supernatant	Vitamin D exposure	SILAC, LC-MS/MS	821	[35]

proteins, which were later used to filter out a range of 70–90.5% of the original data analyzed. From this predicted subsets, conservational functional patterns were inferred, as well as the signaling pathway enriched in them.

## Research design and methods

### Chemicals and materials

All reagents were purchased from Sigma, Søborg, Denmark. Additionally, trypsin from porcine pancreas (Promega, Madison, Wisconsin, U.S.A.); 1,4-dithiothreitol (DTT), iodoacetamide (IAM), urea 98%, Tris base  $\geq 99.9\%$ , trifluoroacetic acid (TFA), acetonitrile grade liquid chromatography-mass spectrometry (LC-MS) Chromasolv, and water-grade LC-MS Chromasolv (Thermo Fischer Scientific, Hvidovre, Denmark) were used. Vivacon 500 Ultra centrifugal filters with a 2-kDa molecular weight cut-off (2 MWCO) were purchased from Sartorius Stedim Lab Ltd. (Stonehouse, Gloucestershire, U.K.).

### Cell culture

The rat insulinoma INS-1E cell line was grown in RPMI-1640 GlutaMAX medium (11 mM glucose) supplemented with 10% fetal bovine serum (FBS), 100 U/ml penicillin, 100  $\mu$ g/ml streptomycin, 10 mM HEPES, 1 mM sodium pyruvate and 50  $\mu$ mol/l  $\beta$ -mercaptoethanol.

The cells were grown in 6-well plates until 85–90% confluence. Each well received fresh medium 12 h prior to the experiment. Due to reduced efficiency of MS analysis in the presence of FBS, each well was washed 3 $\times$  with RPMI-1640 glucose and FBS free, supplemented with 100 U/ml penicillin, 100  $\mu$ g/ml streptomycin, 10 mM HEPES, 1 mM sodium pyruvate and 50  $\mu$ mol/l  $\beta$ -mercaptoethanol. After washing, cells were incubated with RPMI-1640 FBS-free 20 mM glucose (2 ml) for 4 h prior to the collection of the supernatants. Cell numbers used in experiments: 3  $\times$  10<sup>6</sup> INS-1E cells (or the equivalent of 85–90% confluence of a 6-well plate).

### Sample preparation and protein digestion

After the incubation, conditioned media from INS-1E cells were collected and centrifuged (2000 g, 10 min) to remove cell debris. Samples were concentrated using Vivacon 500 2 MWCO filter units (Sartorius, Goettingen, Germany) followed by an adapted protocol from Wisniewski et al. [36]. Proteins in the filter were reduced in a denaturing buffer

(8 M urea, 0.1 M Tris-HCl pH 8.5, 500 mM DTT) for 30 min at room temperature (RT) and followed by alkylation of free sulfhydryl groups with 500 mM IAM at RT for 30 min in the dark. Reduced and alkylated samples were incubated overnight with trypsin from porcine pancreas. Next day, samples were centrifuged and the flow through, where the peptides were contained, collected.

## StageTips preparation

For micro-purification of peptides prior to MS, an adaptation of Rappsilber et al. [37] was used. The StageTips were prepared placing two Empore filter disks (3M) at the very end of a D200 200  $\mu$ l tip using a sampling tool syringe. Each sample had the pH adjusted to approximately 2 with 10% TFA. After filter activation, the samples were loaded and centrifuged (1200 g, 3 min). The StageTips were washed with 0.1% TFA and peptides were finally eluted with 60% acetonitrile, 0.1% TFA. Acetonitrile was evaporated at RT prior to MS analysis.

## Mass spectrometry

Samples were analyzed on a Bruker Impact II ESI-QTOF (Bruker Daltonics, Bremen, Germany) mass spectrometer with an on-line Dionex Ultimate 3000 chromatography system (Thermo Fisher Scientific, Waltham, Massachusetts, U.S.A.) equipped with a Bruker Nanoelute column (15 cm, 75  $\mu$ m ID). Peptides were eluted using a solvent gradient over 65 min, using acetonitrile with 0.1% formic acid at a flow rate of 300 nl min<sup>-1</sup>. The MS scan range was 150–2200 *m/z* with a cycle time of 3 min using a MS sampling rate of 2 Hz followed by intensity-based data-dependent MS/MS (4–16 Hz).

## Data analysis

Database searches were performed with MaxQuant v 1.6.1.0 [38] using the following parameters: enzyme: trypsin, with three missed cleavages; fixed modification: carbamidomethyl (cysteine); variable modification: oxidation (methionine); 1% peptide-level false discovery rate; mass tolerance: 0.07 and 0.005 Daltons (first and main searches, respectively); MS/MS mass tolerance: 40 part per million (ppm) (first and main searches). Data analysis was performed using the same software (MaxQuant v 1.6.1.0) [38] with semi-specific tryptic constraints and a 1% peptide level false discovery rate. Proteins identified by at least one unique peptide were considered only if their presence was consistent in two out of the three samples submitted for analysis. Further, the list was additionally filtered to exclude potential contaminants.

## Proteomic datasets analyzed

An overview of the publications relevant to our analysis is presented in Table 1. The publications from which the lists of proteins were acquired: (1) HPI [32], (2) MPI [39], (3) INS-1E cells (isolated  $\beta$ -cells) [32], (4) ISG from INS-1E cells [34], and (5) INS-1E cell supernatant ( $\beta$ -cells SUP) [35], were extracted and then analyzed by our tool and are highlighted in Table 1. All lists are public and available. Additional lists were found but not analyzed [33,40–42]. The sample processing and proteomic approaches vary between datasets and their descriptions can be found in respective publications. Our own list of proteins from the supernatant of INS-1E cells after 20 mM glucose stimulation, 4 h ( $\beta$ -cell supernatant – SUP2) presents 978 proteins, and it was additionally used as a proteomic dataset in the present study.

## Components of the analysis method

The method is divided into four phases (Figure 1).

### Phase I: the assembly of the data (manual)

Outputs from proteomic studies are typically presented as excel files containing a list of the identified proteins. From the lists available in the chosen publications, the identified protein names and their respective gene names and accession numbers – unique stable identifiers assigned to each protein by the UniProt Knowledge (UniProtKB, v. 2019\_03 updated in April 2019) database [43] – were extracted and copied to a new excel file. All subsets were submitted to the same handling for file organization. Consequently, a readable file for our tool is an excel file containing: (1) accession numbers, (2) protein names and (3) gene names.

### Phase II: retrieving the subcellular location, full amino acid sequence and SP (automatic)

After assembly of the input file during phase I, the data from UniProtKB regarding each protein's subcellular location, as well as their full amino acid sequence and the presence of SP are retrieved. All proteomic datasets were submitted

at least three times to check for potential inconsistencies in the data retrieval. Obsolete and deleted entries that were identified after UniProt retrieval were excluded from the respective datasets. The output file after retrieval is an excel file containing the features retrieved in separated columns. Within an excel file, the auto filter feature was employed. It facilitates the selection of +SP proteins or the ones containing an ER-retrieval consensus. The auto filter also facilitates the screening of subcellular locations (once selected only “secreted”, the excel will show only proteins annotated to be secreted according to UniProtKB). Details about the script can be found in the documentation (Supplementary File S1).

### Phase III: ER-retrieval signal consensus (manual/automatic)

Once the subcellular locations are displayed in the excel file, it is possible to filter the proteins that were experimentally proven to localize to the ER by again employing the excel auto filter. This filtering was made using the HPI list, the largest subset of proteins we had access to, and where we could find the largest number of ER-resident proteins. The list submitted to PROMALS3D is available in the Supplementary File S2. With this subset of human islet ER-proteins and their full amino acid sequence, a server for amino acid sequence alignment was used. PROMALS3D (PROfile Multiple Alignment with predicted Local Structures and 3D constraints) is a tool for multiple sequence and structure alignment [44]. The server receives input sequences in a FASTA format and it is able to return a consensus (the most common tetrapeptide variation within human islet ER-resident proteins, Supplementary File S3).

### Phase IV: ER-consensus finder (automatic)

The final step of the script executes the search of the consensus in the four last amino acids from the full amino acid sequence of each protein in the original list. All lists were submitted at least three times to perform this search. All proteins in all lists have their amino acid sequences described and appropriately recovered in phase II. If the consensus is found in the C-terminus, it will return the information in a new column in the outcome file.

The objective function of our approach can be summarized as:

Classically secreted proteins = (x2 - x3)

The objective function was implemented in our script as:

```
# Second, Grab the raw protein sequence
sequence = soup.select('sequence[length]')[0].get_text().replace('\ n; ')
excel_new['full_sequence'][i] = sequence
# If the protein has a signal peptide, parse the information
features = soup.select('feature[type = "signal peptide"]')
if len(features) > 0:

    # Grab start and end position of the signal peptide
    feature = features[0]
    start = feature.select('begin')[0].get('position') # Get start position
    end = feature.select('end')[0].get('position') # Get end position
    excel_new['signal_sequence'][i] = start+" "+end
    # Last, look for the KDEL in the final 4 amino acids
    searcher = re.compile('[ARNDCQEGHILKMFPSTWYVBZ]{2}EL')
    # Select last four characters of the full sequence
    final_aa = sequence[-4:]
    # Check if there is a match with the regular expression
    matched = searcher.match(final_aa)
    if matched:

        excel_new['kdel_found'][i] = matched.group(0)
```

## Results

The semi-automated method that integrates relevant functionalities for the prediction of classically secreted proteins (classical secretome) was constructed. The different components of the workflow were linked using a Python script. All datasets, including our own, were submitted for analysis by our tool, and the results are described and discussed below.

**Table 2** The summary of performed analysis. **SP:** signal peptide; **XXEL:** KDEL signal consensus. **+SP - XXEL:** predicted secretome (classical secretome)

	Study / Author, year	Proteins analyzed	Signal Peptide SP	ER-resident proteins +SP +XXEL	Predicted secretome +SP -XXEL
Human islet (HIP)	The Human Diabetes Proteome Project / Topf et al. (2013)	5317	725 (13.6%)	53 (1%)	672 (12.6%)
Mice islet (MIP)	Glucose-stimulated islet proteome / Waanders et al. (2009)	6745	708 (10.5%)	59 (0.9%)	649 (9.6%)
INS-1E Rat $\beta$ -cell ( $\beta$ cell)	The Human Diabetes Proteome Project / Topf et al. (2013)	2523	254 (10%)	17 (0.67%)	237 (9.4%)
INS-1E ISG (ISG)	Insulin granule / Schwartz et al. (2012)	140	43 (30.7%)	1 (0.7%)	42 (30%)
INS-1E supernatant (SUP)	$\beta$ -cell secretome / Pepaj et al. (2016)	823	104 (12.6%)	13 (1.6%)	91 (11%)
INS-1E supernatant (SUP2)	$\beta$ -cell secretome / our subset	978	92 (9.4%)	13 (1.3%)	79 (8%)

## Data selection and organization of the readable file

A literature review was conducted in order to identify data sets of the pancreatic  $\beta$ -cell secretome upon high glucose stimulation, presented as a straightforward list that could be used as a reference. To the best of our knowledge, this list is not available. Therefore, a new analysis was carried out. As a source of datasets and a proof of principle, nine specific publications containing proteomic data from HPI, MPI,  $\beta$ -cells, ISG, and SUP were selected (Table 1). By applying the criteria: (1) publication date (data reported in the last 10 years) and (2) the full list of identified proteins available as supplementary material in the original publication (excel file), five out of the nine publications were selected for further analysis (Table 1 – highlighted publications). The resulting list from our own analysis of the INS-1E cells upon 20 mM glucose stimulation (SUP2) was also submitted.

## Filtering by the presence of SP – potential structure for secreted proteins

Among the 5317 proteins identified in the HPI proteome [32], 13.6% of proteins display SP (+SP) and therefore potentially belong to the classical secretory pathway. Relatively proportional rates were found in MPI (10.5%) and  $\beta$ -cells (10%). The ISG, as a structure that undergoes exocytosis, was expected to contain only proteins that went through the secretory pathway, i.e. +SP. Indeed, it was the dataset containing the highest percentage of +SP proteins (30.7%). Filtering the list from the  $\beta$ -cell SUP as well as SUP2, both presumed to be highly enriched in secreted proteins, the percentage of +SP proteins were only 13 and 9.4% respectively. Detailed descriptions are presented in Table 2.

## ER-consensus sequence – identification of a functional tetrapeptide for pancreatic islets

A continuous retrograde transport of ER-resident proteins from the cis-Golgi is necessary to retrieve the resident soluble proteins to the ER lumen [8,10]. The selection of these proteins is dependent on a tetrapeptide present in the C-terminus, classically referred to as KDEL motif in animal cells and HDEL in *Saccharomyces cerevisiae* [8,9]. However, multiple functional variants of this motif have been reported, with the main restriction that the last two amino acids are glu-leu (EL) [45]. The process of identifying the human islets ER-resident proteins was already described in the methods session (phase III/Supplementary Files S2 and S3) and indeed the most common motif identified was EL at the last two C-terminal amino acid positions. In agreement with previous publications, high variability at two preceding positions was identified with amino acids H, S, R, P, E, F, G and D substituting K at first position of KDEL motif and T, N, E, V, G, A, I and R substituting D at the second position. Considering that protein list submitted for this analysis was biased to the experimental approaches that led in the first place to their identification as ER luminal residents, and to maintain the possibility that other ER-localized proteins with different KDEL-like motifs are not filtered out during unbiased proteome–secretome analyses, we decided to use only EL sequence for further filtering. However, as KDEL motif interaction with KDEL receptors is impacted by four C-terminal amino acids and keeping with the traditional annotation, we have further referred to the motif as XXEL, where X represents any of the 20 possible amino acids found in living organisms.

Our analysis showed that 1% of the HPI (53 proteins), 0.9% of the MPI (59 proteins) and 0.67% of  $\beta$ -cells (17 proteins) proteomes contain the motif XXEL (+XXEL). The ISG displayed only 1 protein with the XXEL motif, the endoplasmic reticulum chaperone BiP while the supernatants (SUP and SUP2) presented the largest amounts (proportionally): 1.6 and 1.3% (13 proteins both).

### The predicted subset of classically secreted proteins: classic secretomes

The final goal of our approach was to predict, *in silico*, the classically secreted proteins. Although being aware that the secretome contains 'all factors secreted by a cell to the extracellular space' [1,2], we will here name the subset of classically secreted proteins as the 'classical secretome'. All phases of the program converged to identify proteins that are concomitantly +SP and -XXEL, thus generating a list that fulfills the known structural requirements necessary for secretion through the classical secretory pathway. The auto filter feature from excel was also used within the 'subcellular location' column (phase II, previously described). The auto filter provides another layer of refinement, allowing the search for specific groups of locations. Additionally, we filtered the ISG and supernatants (SUP, SUP2) data to obtain a 'cleaner' list, excluding intracellular contaminants or proteins outside of the +SP -XXEL criteria.

Following this reasoning, the tool reduced the proteomes to the *in silico* predicted secretome – classical secretome (Table 2/Supplementary File S4). HPI, MPI,  $\beta$ -cells, ISG, SUP and SUP2 were reduced, respectively, by 87.4, 90.4, 90.5, 70, 88.8 and 92% of their full identified proteome. In other words, 12.6 (HPI), 9.6 (MPI) and 9.5% ( $\beta$ -cell) of proteins met the criteria to be included in the final list of respective, classical secretomes. Surprisingly, only 30% of the proteins identified in the ISG were +SP -XXEL proteins. Regarding the analyzed supernatants only 11% (SUP) and 8% (SUP2) fulfilled our criteria of a classical secretome (Supplementary File S5).

### Comparative analysis – conservation among the predicted secretomes

Although conservational studies among secretomes are not uncommon investigations [46,47], the analysis of how conserved the secretomes of human and rodent islets/ $\beta$ -cells are is still a loose end. Our data showed that HPI and MPI predicted secretomes are 49.4% similar. Using the  $\beta$ -cell predicted secretome as a reference, 119 proteins (out of 237) are enclosed within the HPI classical secretome, indicating that  $\beta$ -cells and HPI share 50.2% similarity in their secretomes.  $\beta$ -Cells and MPI classical secretomes share 52.3% identity. About 78% proteins reported in ISG were found in  $\beta$ -cell predicted secretome as well (32 out of 41). Finally, comparing the ISG to filtered SUP, we observed 73% similarity. Interestingly, the comparison of predicted secretomes of HPI and MPI to  $\beta$ -cell yielded 81 common proteins. However, when we compared limited with  $\beta$ -cells, predicted secretomes of SUP and ISG or SUP and SUP2, we identified only 25 and 21 common proteins. Overlapping patterns are presented in Figure 2.

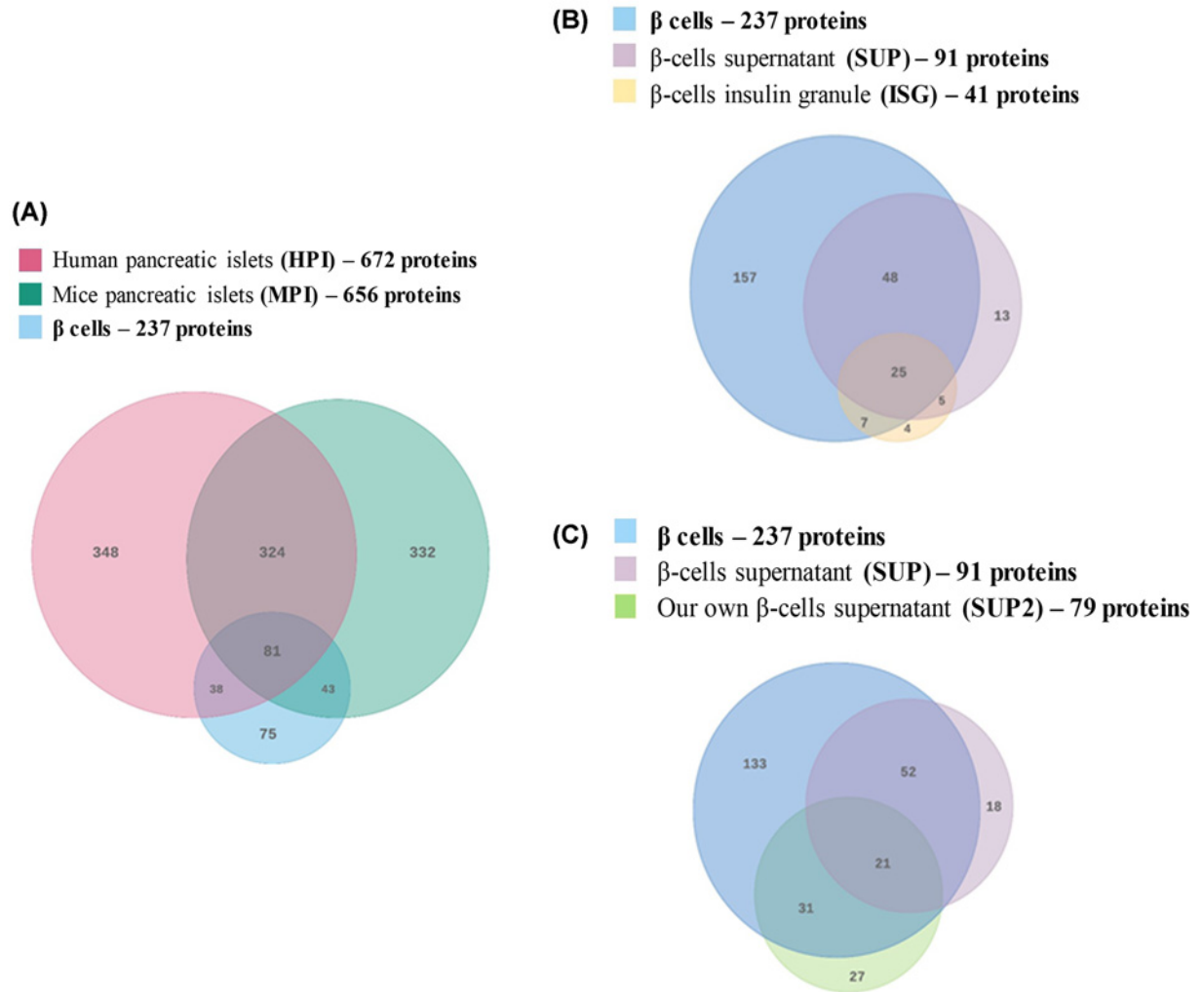
### Pathways that are enriched in the human islet predicted secretome: analysis of *in silico* obtained data set

To evaluate the biological relevance of *in silico* prediction of secretome(s), we have analyzed the biggest protein subset (*in silico* HPI secretome) for the pathway over-representation. We used Reactome, an open-source, curated and peer-reviewed pathway database [48] for analysis. Based on identification of 498 out of the 672 proteins submitted, Reactome generated a list of the 25 pathways found in the human islet predicted secretome sorted by p-value (Supplementary File S6). The list of enriched pathways clustered in three groups: (1) cell adhesion/extracellular matrix (ECM) interactions, (2) immune system and coagulation, and (3) cell proliferation (Table 3).

## Discussion

We showed here the utility of a new method to analyze proteomic data. It is a straightforward approach that can be used prior to the laboratory validation process, helping to support research questions (Figure 1). In order to accomplish its functions of filtering for classically secreted proteins, we have selected UniProtKB and PROMALDS3 as the reliable database and server, as well as Python as the programming language. All components necessary for our approach to function properly are free and publicly available at <https://github.com/tatiorli/ParseSignalPeptides>. The documentation can be found in the Supplementary File S1.

UniProtKB provides a comprehensive and freely accessible resource of protein information [49]. Manually annotated and reviewed records extracted from literature and curator-evaluated computational analysis [43,49]. From UniProtKB, we retrieved data regarding subcellular location, full amino acid sequence and the presence of SP of each protein present in the input list. The subcellular location was used to increase the analysis power of our tool although this feature has to be applied cautiously as some proteins' annotations are (1) derived experimentally while other are not thus introducing a strong bias, and (2) some proteins are shown to localize to two or more cellular compartments.



**Figure 2. Venn diagrams showing overlaps between different subsets of classically secreted proteins (secretomes)**

(A)  $\beta$ -Cell secretome and mice pancreatic islets (MPI) are approximately 50% similar to human pancreatic islets (HPI). (B) The filtered  $\beta$ -cell insulin granule (ISG) and the  $\beta$ -cell supernatant (SUP) share around 80% identity with the  $\beta$ -cell secretome. (C)  $\beta$ -Cell secretome (SUP2) shares 50% similarity with the  $\beta$ -cell supernatant (SUP). Both supernatants (SUP and SUP2) share 80.2% and 65.8%, respectively.

The proteins' full amino acid sequences, on the other hand, subsequently used for the XXEL detection were available for all tested proteins. The SP presence is dictated by UniProt if the cleavage site has been determined by direct protein sequencing or if it was predicted by at least two different predictive tools (Phobius, Predotar, SignalP, and TargetP). The credibility and wide usage of the database provides a strong argument to employ it as a central database for our method.

The ER-retention signal is crucial for the recovery of ER-resident proteins [10]. Classically, the process is mediated by receptors that recognize the KDEL motif, returning a protein that has escaped to the ER lumen. Three different receptor homologs were described in humans, forming the KDEL receptor family [50] and they exhibit varying affinities for KDEL motif variants [45]. So far, 59 KDEL variants have been identified experimentally [45]. We rationalized that other functional KDEL variants may exist and thus decided on an unbiased approach to identify them. We have aligned the amino acid sequences of experimentally proven ER-resident proteins (HPI data set) and identified only the last two C-terminal amino acids: EL as the consensus sequence. The first two amino acids within KDEL motif, varied among proteins (Supplementary Files S2 and S3) and in agreement with previous publications [45]. To avoid bias, where only experimentally proven ER-resident proteins are used for KDEL variant identification, we decided to use for protein filtering the most common identified EL consensus (for consistency referred to as XXEL) and not



**Table 3 Clusters with the enriched pathways identified in the predicted human islets secretome according to Reactome database**

CLUSTERS	PATHWAYS
<b>CELL ADHESION / ECM INTERACTIONS</b>	Laminin interactions Integrin cell surface interactions Extracellular matrix (ECM) proteoglycans Non-integrin membrane–ECM interactions Assembly of collagen fibrils and other multimeric structures Degradation of the ECM ECM organization Collagen formation Collagen degradation MET activates PTK2 signaling Collagen biosynthesis and modifying enzymes Elastic fibre formation MET promotes cell motility
<b>IMMUNE SYSTEM / COAGULATION</b>	Neutrophil degranulation Platelet degranulation Response to elevated platelet cytosolic Ca <sup>2+</sup> Formation of Fibrin Clot (Clotting cascade) Platelet activation, signaling and aggregation Innate immune system
<b>CELL PROLIFERATION</b>	Regulation of insulin-like growth factor (IGF) transport and uptake by insulin-growth factor binding proteins (IGFBPs)
<b>OTHER</b>	Syndecan interactions Post-translational protein phosphorylation Retinoid metabolism and transport Metabolism of fat-soluble vitamins Glycosphingolipid metabolism

a limited number of KDEL variants found via alignment in the tested data set (Supplementary File S3). Clearly, the future experimental verification of each possible variant functionality is necessary. This approach limits false negative selection (non-classical KDEL sequence) but maintains a 14.4% level of false positive choices. It is important to point out that the current assignment of proteins to other than ER cell compartments (false positive group) may change as new experimental data is provided, lowering the rate of false positive identifications.

Further validation of our approach derives from the decrease in the numbers of XXEL proteins proportionally to the total number of identified proteins through the models. While the HPI and MPI contained 53 and 59 +XXEL proteins respectively, the  $\beta$ -cell presented 17 and the ISG only 1 (Table 2). The only +XXEL protein found within the ISG is the endoplasmic reticulum chaperone BiP, the ER chaperone already described in the context of the extracellular compartment [51,52]. Regarding the  $\beta$ -cell supernatants (SUP and SUP2), both showed 13 +XXEL proteins (Table 2), indicating probable contamination with cellular matter and/or secretion of ER proteins by the insulinoma cell model, INS-1E. According to the Human Protein Atlas, around 2% of all human proteins have been experimentally shown to localize to the ER [53]. Another report states that 202 ER-proteins are found in the pancreas ER (corresponding to around 1% of all human proteins) [54]. Our results showed proportional trends (approximately 1%) suggesting a high level of consistency between our screening and other published approaches (Table 2). If only the cellular localization filter is used to search for ER proteins, many more than the ones in the list of +XXEL proteins will be found. The probable reason lies in the fact that 50% of all ER proteins are shown to co-localize to other compartments such as cytosol and nucleus [53].

It is reported that 39% of the human protein-coding genes are predicted to have a SP [21]. Considering that the HPI proteome is one-third of the human proteome [32] and that we showed it contains 13.6% +SP islet proteins, the results are coherent. MPI and  $\beta$ -cells exhibit similar proportions ( $\pm 10\%$ ). The ISG and supernatants (SUP, SUP2) were expected to present higher percentages since they are predicted to contain only ready-to-be secreted or already secreted proteins. Surprisingly only ISG showed substantial increase (up to 30%) while SUP and SUP2 contained only 12.6% and 9.2% of +SP proteins (Table 2). This can be explained by (1) the presence of intracellular contaminants as a result of sample processing, (2) a large number of unconventionally secreted proteins (that do not fulfill our criteria

of +SP), (3) a reflection of new biological processes by which  $\beta$ -cells can secrete proteins e.g. formation of exosomes [55] or (4) a characteristic of insulinoma INS-1E cell line. It is important to mention that the presence of SP is not a guarantee of secretion, as +SP proteins can eventually be retained in the lumen of the ER, lysosomes, any other organelle along the secretory pathway or in the membranes (e.g. membrane bound receptors).

Our ultimate goal was the creation of a list of the islet and  $\beta$ -cell classically secreted proteins that we achieved by reducing the human and  $\beta$ -cell proteomes to +SP -XXEL proteins (Supplementary File S4). With no major differences in the proportions, HPI and MPI, as well as  $\beta$ -cells contain 12.6, 9.6 and 9.4% of proteins within their respective proteomes that represent what we call the classical secretome (Table 2). This secretome includes many proteins that fit our criteria (+SP -XXEL) but that are not elsewhere annotated as secreted. It includes membrane proteins and proteins from organelles in the secretory pathway (i.e. lysosomes). To further narrow down the secretome list multiple approaches can be suggested, e.g. (1) usage of the cellular localization filter, (2) immunostaining of individual proteins or (3) exogenous expression of the tagged protein of interest.

A series of reviews and reports analyze how translatable are the discoveries and therapeutic perspectives made in rodent models to humans in the field of diabetes [56]. These discussions are especially important in order to choose between models that can best serve for the purposes of clarifying the specific mechanisms behind diabetes development and progression. The considerations at the protein expression level utilize the comparisons at the whole cell proteome or transcriptome level [57]. We decided to analyze how similar the subsets of *in silico* predicted classical secretomes of HPI and  $\beta$ -cells are (Figure 2). After filtering made by our tool, the lists of predicted classically secreted HPI and MPI proteins were overlaid. The overlap we obtained came as a surprise: 50.2% (Figure 2). The variance may represent species differences and/or reflect technical hindrances e.g. protein annotation and naming differences between animal species, databases redundancies, obsolete and deleted accession numbers, mass spectrometry related issues (machine and software used to analyze input samples) and finally, sample preparation and experimental design.

Much higher evidence of overlap was seen when 78% of the +SP -XXEL ISG proteins matched with the *in silico*  $\beta$ -cell secretome, as well as the 73% matching with the +SP -XXEL SUP (Figure 2). Moreover, the predicted  $\beta$ -cell secretome and the +SP -XXEL from SUP share 80% identity. Interestingly, our  $\beta$ -cell secretome (SUP2) is only 65.8% similar to the SUP secretome (Figure 2). This difference possibly arises from the different research designs and sample processing, where SUP is a supernatant from  $\beta$ -cells exposed to vitamin D [35], our (SUP2) comes from  $\beta$ -cells after high glucose exposure (20 mM), underscoring biases related to experimental design.

Another analysis was carried out in order to support the reliability of the filtering made by our tool. Using Reactome database, the biological processes enriched in the HPI classical secretome were investigated. The 25 most relevant pathways were identified (Supplementary File S6), and we clustered them in three groups (Table 3). The first, ECM interactions and cell adhesion group, indicate an active islet participation in creation and modeling of extracellular microenvironment where different cell types need to coexist. This subject has gained a special attention in the islet transplantation field, where ECM molecules have been explored in their clinical potential to contribute to the islet's engraftment success [58]. The second group is related to the immune system and consists of proteins related to neutrophil and platelet activation and degranulation, activation of other innate immune system components, and glycosphingolipid metabolism. These factors are tightly involved in the regulation of the inflammation levels that have been associated with the development of diabetes and islet homeostasis [59,60]. Finally, the third group was mainly represented by the insulin growth factors (IGFs) signaling and metabolism of fat-soluble vitamins. Given that the IGFs were shown to be involved in  $\beta$ -cell survival and insulin secretion, their presence in the extracellular space is of no surprise [61,62]. On the other hand, it is interesting to highlight that pathways involved in the fat-soluble vitamins (vitamins A, D, E, and K) metabolism are enriched in the HPI secretome. Recently, a report showed that a lack of vitamin A can lead to  $\beta$ -cell dysfunction and insulin production in mice [63], while vitamin D has been investigated for its potential properties to protect and repair damaged  $\beta$ -cells [64]. Although 782 pathways were identified by at least one of the secreted proteins on our list, no pathways regarding well known intracellular processes (DNA replication, chromatin organization) were significantly enriched, demonstrating that our analytical approach provides a good level of clearance of intracellular contaminants.

The pancreatic islets constitute 1–2% of the total cell mass of an adult pancreas [65]. The coding genes and proteins expressed in the pancreas are described and characterized in various publications [66]. However, to the best of our knowledge, a reference list from a free-labeled MS approach analyzing the subset of proteins secreted by the islets or  $\beta$ -cells is not yet available. Secretome proteins have a major role in central physiological processes and are considered as a rich resource for the identification of biomarkers and drug targets. Quantitative proteomics of secreted proteins is possible by MS and has been used before [67–69], but has proven difficult and highly biased due to methodological limitations. These limitations have been partially addressed with the development of techniques such as the SPECS (the secretome protein enrichment with click sugars) that allow proteome-wide identification of secreted proteins

at high-depth [70]. Further improvements can be achieved with a sophisticated *in silico* approach designed to integrate specific protein datasets already available, as presented here.

The current version of our tool does not integrate any database capable to distinguish with certainty proteins that are unconventionally secreted, which will be in an important update in the future. However, these groups of proteins could be inferred to a certain degree by first positively filtering for annotated secreted proteins and then subtracting the proteins with +SP – XXEL. Our *in silico* secretome prediction script can be adjusted to the investigation of any kind of proteomic data from any tissue or cell type.

## Conclusion

In the present study, we demonstrated how different functionalities of diverse computational tools can be used together to extract useful information in order to predict secreted proteins from proteomic data. All the programs used are open source tools freely available. Our approach reduced whole cell proteomes (HPI, MPI, and  $\beta$  cell) and filtered data from ISG and supernatants from potential contaminants and non-secreted proteins, predicting their respective classical secretomes.

As the analytical performance of proteomic platforms will continuously improve, resulting in better experimental description of secretomes, their output will identify new physiological processes and contribute to the identification of new therapeutic targets.

## Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

## Funding

This project was funded by the Danish Council for Independent Research, Augustinus Foundation, and the Department of Biomedical Sciences at the University of Copenhagen [grant number DFF-4092-00066 (to M.T.M.)].

## Author Contribution

EPM, TOMS and CP wrote the Python script and performed the analysis and wrote the manuscript draft. EPM and PMH performed experiments and analyzed data. EPM made all the figures. MTM conceived the project, analyzed the experiments, wrote and corrected the manuscript.

## Data Availability

Python tool created by us is freely available for download and usage at <https://github.com/tatiiori/ParseSignalPeptides>. The files with proteins lists (.xlsx) data used to support the findings of this study are included within the supplementary information file(s).

## Abbreviations

Asp, aspartic acid; DTT, 1,4-dithiothreitol; ECM, extracellular matrix; ER, endoplasmic reticulum; *erd2*, ER lumen protein-retaining receptor; FBS, fetal bovine serum; Glu, glutamic acid; HDEL, histidine–aspartic acid–glutamic acid–leucine; His, histidine; HPI, human pancreatic islets; IAM, iodoacetamide; ISG, insulin secretory granules; KDEL, lysine–aspartic acid–glutamic acid–leucine; LC-MS, liquid chromatography-mass spectrometry; Leu, leucine; Lys, lysine; MPI, mouse pancreatic islets; MS, mass spectrometry; MWCO, molecular weight cut-off; ppm, parts per million; PROMALS3D, PROfile Multiple Alignment with predicted Local Structures and 3D constraints; RT, room temperature; SP, signal peptide; +SP, proteins containing signal peptide; SPECS, secretome protein enrichment with click sugars; SUP,  $\beta$ -cell supernatant; SUP2, our own supernatant of  $\beta$ -cell model cells (INS-1E); TFA, trifluoroacetic acid; UniProtKB, UniProt Knowledge; XXEL, any amino acid – any amino acid – glutamic acid – leucine (the ER-retrieval signal consensus); +XXEL, proteins containing the XXEL sequence; -XXEL, proteins without the XXEL sequence.

## References

- 1 Tjalsma, H. et al. (2000) Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. *Microbiol. Mol. Biol. Rev.* **64**, 515–547, <https://doi.org/10.1128/MMBR.64.3.515-547.2000>
- 2 Agrawal, G.K. et al. (2010) Plant secretome: unlocking secrets of the secreted proteins. *Proteomics* **10**, 799–827, <https://doi.org/10.1002/pmic.200900514>
- 3 Stastna, M. and Van Eyk, J.E. (2012) Secreted proteins as a fundamental source for biomarker discovery. *Proteomics* **12**, 722–735, <https://doi.org/10.1002/pmic.201100346>
- 4 Bielenberg, D.R. and Zetter, B.R. (2015) The Contribution of Angiogenesis to the Process of Metastasis. *Cancer J.* **21**, 267–273, <https://doi.org/10.1097/PP0.000000000000138>

- 5 Shi, Z. and Gewirtz, A.T. (2018) Together Forever: Bacterial-Viral Interactions in Infection and Immunity. *Viruses* **10**, pii: E122
- 6 Lemberg, M.K. and Martoglio, B. (2002) Requirements for signal peptide peptidase-catalyzed intramembrane proteolysis. *Mol. Cell* **10**, 735–744, [https://doi.org/10.1016/S1097-2765\(02\)00655-X](https://doi.org/10.1016/S1097-2765(02)00655-X)
- 7 McCaffrey, K. and Braakman, I. (2016) Protein quality control at the endoplasmic reticulum. *Proteostasis* **60**, 227–235
- 8 Wang, J., Luo, J. and Zhang, X. (2013) From endoplasmic reticulum to Golgi apparatus: a secretory pathway controlled by signal molecules. *Zhejiang Da Xue Xue Bao Yi Xue Ban* **42**, 472–477
- 9 Stornaiuolo, M. et al. (2003) KDEL and KKXX retrieval signals appended to the same reporter protein determine different trafficking between endoplasmic reticulum, intermediate compartment, and Golgi complex. *Mol. Biol. Cell* **14**, 889–902, <https://doi.org/10.1091/mbc.e02-08-0468>
- 10 Spang, A. (2013) Retrograde Traffic from the Golgi to the Endoplasmic Reticulum. *Cold Spring Harb. Perspect. Biol.* **5**, <https://doi.org/10.1101/cshperspect.a013391>
- 11 Papanikou, E. and Glick, B.S. (2014) Golgi compartmentation and identity. *Curr. Opin. Cell Biol.* **29**, 74–81, <https://doi.org/10.1016/j.ceb.2014.04.010>
- 12 Burgoyne, R.D. and Morgan, A. (2003) Secretory granule exocytosis. *Physiol. Rev.* **83**, 581–632, <https://doi.org/10.1152/physrev.00031.2002>
- 13 Fernandez, C.J. et al. (1997) Distinct molecular events during secretory granule biogenesis revealed by sensitivities to brefeldin A. *Mol. Biol. Cell* **8**, 2171–2185, <https://doi.org/10.1091/mbc.8.11.2171>
- 14 Davis, S., Wang, J. and Ferro-Novick, S. (2017) Crosstalk between the Secretory and Autophagy Pathways Regulates Autophagosome Formation. *Dev. Cell* **41**, 23–32, <https://doi.org/10.1016/j.devcel.2017.03.015>
- 15 Powell, K. (2005) Excess secretory products fuse with lysosomes. *J. Cell Biol.* **169**, 216–216
- 16 Pickett, J. (2007) The building blocks of the secretory pathway. *Nat. Rev. Mol. Cell Biol.* **8**, 94, <https://doi.org/10.1038/nrm2103>
- 17 Rabouille, C. (2017) Pathways of Unconventional Protein Secretion. *Trends Cell Biol.* **27**, 230–240, <https://doi.org/10.1016/j.tcb.2016.11.007>
- 18 Rubartelli, A. et al. (1990) A Novel Secretory Pathway for Interleukin-1-Beta, a Protein Lacking a Signal Sequence. *EMBO J.* **9**, 1503–1510, <https://doi.org/10.1002/j.1460-2075.1990.tb08268.x>
- 19 Cooper, D.N.W. and Baronides, S.H. (1990) Evidence for Export of a Muscle Lectin from Cytosol to Extracellular-Matrix and for a Novel Secretory Mechanism. *J. Cell Biol.* **110**, 1681–1691
- 20 Hathout, Y. (2007) Approaches to the study of the cell secretome. *Exp. Rev. Proteom.* **4**, 239–248, <https://doi.org/10.1586/14789450.4.2.239>
- 21 Ponten, F. (2015) A Tissue-Based Map of the Human Proteome. *Scand. J. Immunol.* **81**, 329–329
- 22 Uhlen, M. et al. (2018) The human secretome – the proteins secreted from human cells. *bioRxiv* 465815
- 23 Misumi, Y. et al. (1986) Novel Blockade by Brefeldin-a of Intracellular-Transport of Secretory Proteins in Cultured Rat Hepatocytes. *J. Biol. Chem.* **261**, 1398–1403
- 24 Lippincottschwartz, J. et al. (1989) Rapid Redistribution of Golgi Proteins into the Er in Cells Treated with Brefeldin-a - Evidence for Membrane Cycling from Golgi to Er. *Cell* **56**, 801–813
- 25 Chevallet, M. et al. (2007) Toward a better analysis of secreted proteins: the example of the myeloid cells secretome. *Proteomics* **7**, 1757–1770, <https://doi.org/10.1002/pmic.200601024>
- 26 Brown, K.J. et al. (2012) Advances in the proteomic investigation of the cell secretome. *Exp. Rev. Proteomics* **9**, 337–345, <https://doi.org/10.1586/epr.12.21>
- 27 Zwickl, H. et al. (2005) A novel technique to specifically analyze the secretome of cells and tissues. *Electrophoresis* **26**, 2779–2785, <https://doi.org/10.1002/elps.200410387>
- 28 Della Corte, A. et al. (2008) Application of 2-dimensional difference gel electrophoresis (2D-DIGE) to the study of thrombin-activated human platelet secretome. *Platelets* **19**, 43–50, <https://doi.org/10.1080/09537100701609035>
- 29 Garg, G. and Ranganathan, S. (2011) In silico secretome analysis approach for next generation sequencing transcriptomic data. *BMC Genomics* **12**, S14, <https://doi.org/10.1186/1471-2164-12-S3-S14>
- 30 Marton, K. et al. (2018) Comprehensive analysis of *Verticillium nonalfalfae* in silico secretome uncovers putative effector proteins expressed during hop invasion. *PLoS One* **13**, e0198971, <https://doi.org/10.1371/journal.pone.0198971>
- 31 Lebre, P.H. et al. (2018) In silico characterization of the global *Geobacillus* and *Parageobacillus* secretome. *Microb. Cell Fact.* **17**, 156, <https://doi.org/10.1186/s12934-018-1005-9>
- 32 Topf, F. et al. (2013) The Human Diabetes Proteome Project (HDPP): from network biology to targets for therapies and prevention. *Transl. Proteom.* **1**, 3–11, <https://doi.org/10.1016/j.trprot.2013.03.002>
- 33 Martens, G.A. et al. (2010) Protein markers for insulin-producing beta cells with higher glucose sensitivity. *PLoS One* **5**, e14214, <https://doi.org/10.1371/journal.pone.0014214>
- 34 Schvartz, D. et al. (2012) Improved characterization of the insulin secretory granule proteomes. *J. Proteomics* **75**, 4620–4631, <https://doi.org/10.1016/j.jprot.2012.04.023>
- 35 Pepaj, M. et al. (2016) Proteomic analysis of the INS-1E secretome identify novel vitamin D-regulated proteins. *Diabetes Metab. Res. Rev.* **32**, 514–521, <https://doi.org/10.1002/dmrr.2777>
- 36 Wisniewski, J.R. et al. (2009) Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362
- 37 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, <https://doi.org/10.1038/nprot.2007.261>
- 38 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, <https://doi.org/10.1038/nbt.1511>
- 39 Waanders, L.F. et al. (2009) Quantitative proteomic analysis of single pancreatic islets. *Proc. Natl Acad. Sci. U.S.A.* **106**, 18902–18907, <https://doi.org/10.1073/pnas.0908351106>

- 40 Brunner, Y. et al. (2007) Proteomics analysis of insulin secretory granules. *Mol. Cell. Proteom.* **6**, 1007–1017, <https://doi.org/10.1074/mcp.M600443-MCP200>
- 41 Li, M. et al. (2018) Proteomic analysis of insulin secretory granules in INS-1 cells by protein correlation profiling. *Biophys. Rep.* **4**, 329–338, <https://doi.org/10.1007/s41048-018-0061-3>
- 42 Tattikota, S.G. et al. (2013) Argonaute2 regulates the pancreatic beta-cell secretome. *Mol. Cell. Proteom.* **12**, 1214–1225, <https://doi.org/10.1074/mcp.M112.024786>
- 43 UniProt Consortium, T. (2018) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **46**, 2699, <https://doi.org/10.1093/nar/gky092>
- 44 Pei, J., Kim, B.H. and Grishin, N.V. (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36**, 2295–2300, <https://doi.org/10.1093/nar/gkn072>
- 45 Raykhel, I. et al. (2007) A molecular specificity code for the three mammalian KDEL receptors. *J. Cell Biol.* **179**, 1193–1204, <https://doi.org/10.1083/jcb.200705180>
- 46 Vivek-Ananth, R.P. et al. (2018) Comparative systems analysis of the secretome of the opportunistic pathogen *Aspergillus fumigatus* and other *Aspergillus* species. *Sci. Rep.* **8**, 6617, <https://doi.org/10.1038/s41598-018-25016-4>
- 47 Morais do Amaral, A. et al. (2012) Defining the predicted protein secretome of the fungal wheat leaf pathogen *Mycosphaerella graminicola*. *PLoS One* **7**, e49904, <https://doi.org/10.1371/journal.pone.0049904>
- 48 Fabregat, A. et al. (2018) The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* **46**, D649–D655, <https://doi.org/10.1093/nar/gkx1132>
- 49 Breuza, L. et al. (2016) The UniProtKB guide to the human proteome. *Database (Oxford)* **2016**, <https://doi.org/10.1093/database/bav120>
- 50 Lewis, M.J. and Pelham, H.R. (1990) A human homologue of the yeast HDEL receptor. *Nature* **348**, 162–163, <https://doi.org/10.1038/348162a0>
- 51 Shim, S.M. et al. (2018) The endoplasmic reticulum-residing chaperone BiP is short-lived and metabolized through N-terminal arginylation. *Sci. Signal* **11**, pii: eaan0630, <https://doi.org/10.1126/scisignal.aan0630>
- 52 Delom, F. et al. (2001) Role of extracellular molecular chaperones in the folding of oxidized proteins. Refolding of colloidal thyroglobulin by protein disulfide isomerase and immunoglobulin heavy chain-binding protein. *J. Biol. Chem.* **276**, 21337–21342, <https://doi.org/10.1074/jbc.M101086200>
- 53 Friedman, J.R. and Voeltz, G.K. (2011) The ER in 3D: a multifunctional dynamic membrane network. *Trends Cell Biol.* **21**, 709–717, <https://doi.org/10.1016/j.tcb.2011.07.004>
- 54 Chen, X. et al. (2010) Molecular characterization of the endoplasmic reticulum: insights from proteomic studies. *Proteomics* **10**, 4040–4052, <https://doi.org/10.1002/pmic.201000234>
- 55 Ge, Q. et al. (2019) Exosome-Like Vesicles as New Mediators and Therapeutic Targets for Treating Insulin Resistance and  $\beta$ -Cell Mass Failure in Type 2 Diabetes Mellitus. *J. Diabetes Res.* **2019**, 7
- 56 King, A.J. (2012) The use of animal models in diabetes research. *Br. J. Pharmacol.* **166**, 877–894, <https://doi.org/10.1111/j.1476-5381.2012.01911.x>
- 57 Dolensek, J., Rupnik, M.S. and Stozar, A. (2015) Structural similarities and differences between the human and the mouse pancreas. *Islets* **7**, e1024405
- 58 Llacua, L.A., Faas, M.M. and de Vos, P. (2018) Extracellular matrix molecules and their potential contribution to the function of transplanted pancreatic islets. *Diabetologia* **61**, 1261–1272, <https://doi.org/10.1007/s00125-017-4524-8>
- 59 Tai, N., Wong, F.S. and Wen, L. (2016) The role of the innate immune system in destruction of pancreatic beta cells in NOD mice and humans with type 1 diabetes. *J. Autoimmun.* **71**, 26–34, <https://doi.org/10.1016/j.jaut.2016.03.006>
- 60 Eguchi, K. and Nagai, R. (2017) Islet inflammation in type 2 diabetes and physiology. *J. Clin. Invest.* **127**, 14–23, <https://doi.org/10.1172/JCI88877>
- 61 Lingohr, M.K. et al. (2002) Activation of IRS-2-mediated signal transduction by IGF-1, but not TGF- $\alpha$  or EGF, augments pancreatic beta-cell proliferation. *Diabetes* **51**, 966–976, <https://doi.org/10.2337/diabetes.51.4.966>
- 62 Hugl, S.R., White, M.F. and Rhodes, C.J. (1998) Insulin-like growth factor I (IGF-I)-stimulated pancreatic beta-cell growth is glucose-dependent. Synergistic activation of insulin receptor substrate-mediated signal transduction pathways by glucose and IGF-I in INS-1 cells. *J. Biol. Chem.* **273**, 17771–17779
- 63 Amisten, S. et al. (2017) Anti-diabetic action of all-trans retinoic acid and the orphan G protein coupled receptor GPRC5C in pancreatic beta-cells. *Endocr. J.* **64**, 325–338, <https://doi.org/10.1507/endocrj.EJ16-0338>
- 64 Wei, Z. et al. (2018) Vitamin D Switches BAF Complexes to Protect beta Cells. *Cell* **173**, 1135e15–1149e15
- 65 Thul, P.J. et al. (2017) A subcellular map of the human proteome. *Science* **356**, pii: eaa3321, <https://doi.org/10.1126/science.aal3321>
- 66 Zhou, J.Y. et al. (2011) Unraveling pancreatic islet biology by quantitative proteomics. *Expert Rev. Proteom.* **8**, 495–504, <https://doi.org/10.1586/epr.11.39>
- 67 Pellitteri-Hahn, M.C. et al. (2006) Improved mass spectrometric proteomic profiling of the secretome of rat vascular endothelial cells. *J. Proteome Res.* **5**, 2861–2864, <https://doi.org/10.1021/pr060287k>
- 68 Yin, X. et al. (2013) Glycoproteomic analysis of the secretome of human endothelial cells. *Mol. Cell. Proteom.* **12**, 956–978, <https://doi.org/10.1074/mcp.M112.024018>
- 69 Infante, A. and Rodriguez, C.I. (2018) Secretome analysis of in vitro aged human mesenchymal stem cells reveals IGFBP7 as a putative factor for promoting osteogenesis. *Sci. Rep.* **8**, 4632, <https://doi.org/10.1038/s41598-018-22855-z>
- 70 Serdaroglu, A. et al. (2017) An optimised version of the secretome protein enrichment with click sugars (SPECS) method leads to enhanced coverage of the secretome. *Proteomics* **17**, <https://doi.org/10.1002/pmic.201600423>