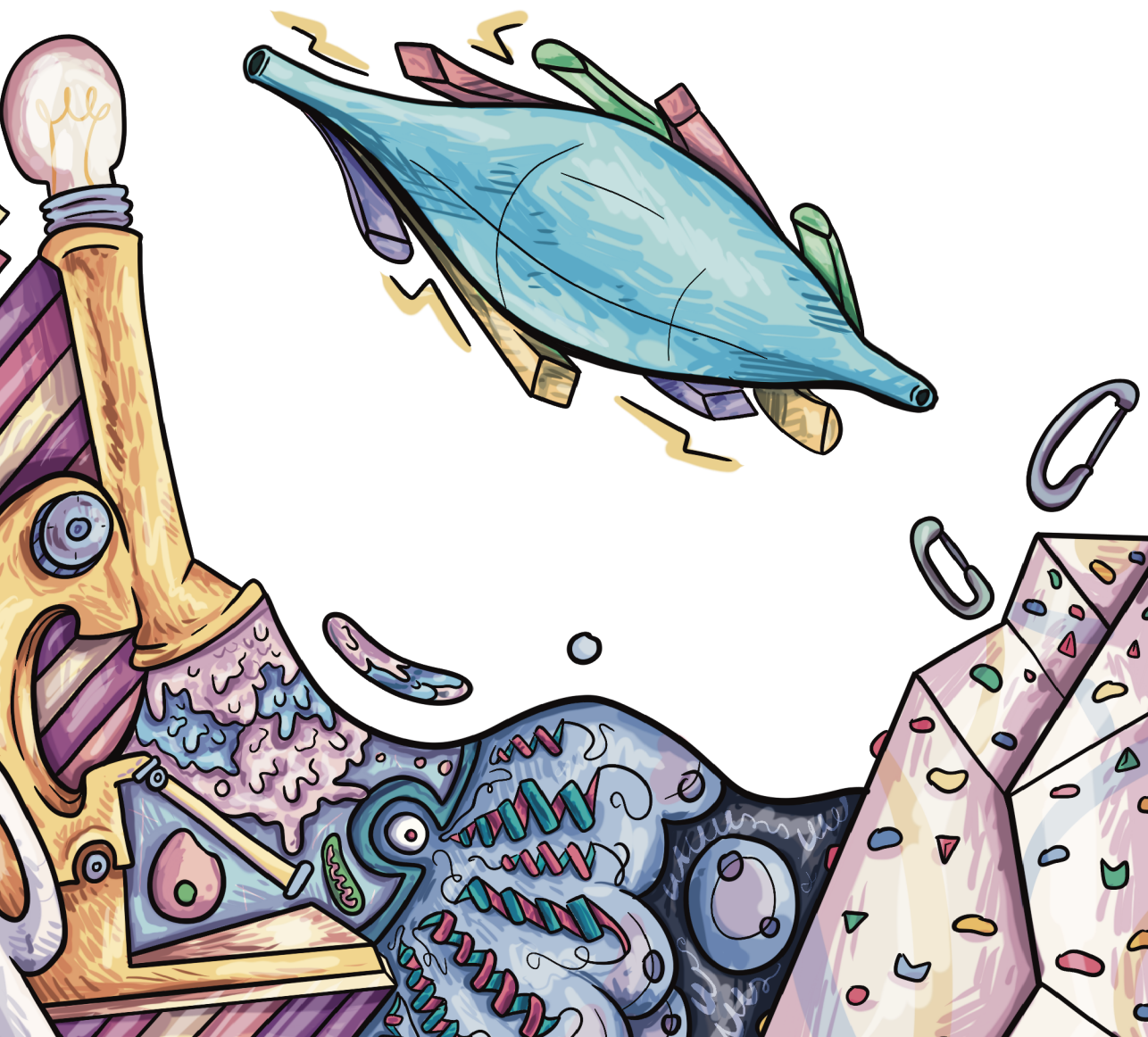# UNDERSTANDING THE NUANCES OF INTERCELLULAR COMMUNICATION

WITH SENSITIVE MASS SPECTROMETRIC APPROACHES

Julia Bauzá Martinez

UNDERSTANDING THE NUANCES OF INTERCELLULAR COMMUNICATION WITH SENSITIVE MASS SPECTROMETRIC APPROACHES

A mis padres

Julia Bauzá Martinez

UNDERSTANDING THE NUANCES OF INTERCELLULAR COMMUNICATION WITH SENSITIVE MASS SPECTROMETRIC APPROACHES

De nuances van intercellulaire communicatie begrijpen met gevoelige massa spectrometrische benaderingen
(met een samenvatting in het Nederlands)

Comprenent els matisos de la comunicació intercel·lular amb tècniques sensibles d'espectrometria de masses
(amb un resum en Català)

Comprendiendo los matices de la comunicación intercelular con técnicas sensibles de espectrometría de masas
(con un resumen en Castellano)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

maandag 21 november 2022 des middags te 2.15 uur

door

Julia Bauzá Martinez

geboren op 11 augustus 1994
te Eivissa, Spanje

"I was taught that the way of progress was neither swift nor easy."

Marie Curie *(Nobel Prize in Chemistry and Physics)*

Marie Curie - 1911
Irène Joliot-Curie - 1935
Dorothy Crowfoot Hodgkin - 1964
Ada E. Yonath - 2009
Frances H. Arnold - 2018
Jennifer A. Doudna - 2020
Emmanuelle Charpentier - 2020
Carolyn Bertozzi - 2022

•

Marie Curie - 1903
Maria Goeppert Mayer - 1963
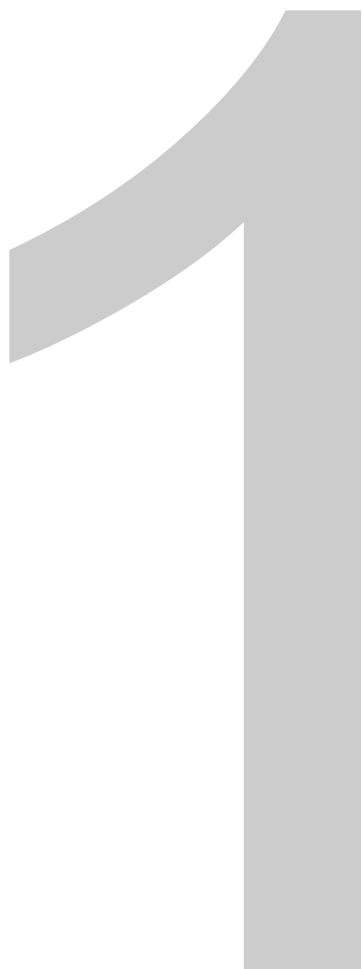Donna Strickland - 2018
Andrea Ghez - 2020

•

Gerty T. Cori - 1947
Rosalyn Yalow - 1977
Barbara McClintock - 1983
Rita Levi-Montalcini - 1986
Gertrude B. Elion - 1988
Christiane Nüsslein-Volhard - 1995
Linda B. Buck - 2004
Françoise Barré-Sinoussi - 2008
Carol W. Greider - 2009
Elizabeth H. Blackburn - 2009
May-Britt Moser - 2014
Tu Youyou - 2015

Table of contents

CHAPTER 1

Introduction

PART I – Mass spectrometry: a versatile tool for the study of peptides and proteins

The complexity of the human proteome is enormous: from the ~20,000 genes in the human genome, variations happening at the DNA and RNA levels, such as single nucleotide polymorphisms and alternative splicing, provide the first layer of complexity of the proteome. Moreover, another layer of variation is introduced at the level of post-translational modifications (PTMs) decorating the translated proteins, altogether giving rise to an 'astronomically large'[1] number of proteoforms[2]. Mass spectrometry (MS) allows for the detailed and quantitative study of peptides and proteins, and it is therefore the current gold-standard technique to study the proteome[3,4]. Advances in mass spectrometers[5–7], hand in hand with computational developments, have allowed the fast, reliable, and quantitative analysis of thousands of peptides and proteins from complex samples[3,8,9], even providing single-cell resolution[8,10]. The work presented here is based on the application of multiple MS approaches that provide diverse and insightful layers of information on various aspects of peptides and proteins.

1.1. MS fundamentals – Ionization, mass analysis and fragmentation

MS is an analytical technique that measures the mass-to-charge ratio (*m/z*) of charged analytes with high precision. The origins of MS date back to the late 19th century, when W. Wien and J. J. Thomson, working with strong electromagnetic fields and anodic or cathodic rays, measured for the first time the *m/z* of protons and electrons[11]. Further developments by Thomson, Aston, and others, led to the advent of better and more complex mass spectrometers throughout the 20th century, leading to the development of the modern mass spectrometers we have today. Nowadays, mass spectrometers can measure the *m/z* ratio of virtually any molecule, while obtaining sequence information that allows identification of the exact molecular composition. To do so, mass spectrometers rely on three main concepts that will be explained hereunder: ionization, mass analysis and fragmentation.

Ionization — MS measures the mass of charged particles (*m/z*) in the gas phase. The process of generating gas-phase charged particles (ions) for MS analysis is termed ionization, and multiple techniques to ionize molecules have been developed throughout the years. Traditionally, ionization techniques were developed for the analysis of small volatile chemicals, but when applied to fragile biomolecules they would result in partial fragmentation of the analyte, leading to wrong *m/z* readouts. This problem was circumvented with the introduction of soft ionization techniques, such as matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI). These techniques allowed volatilization of fragile biomolecules with minimal fragmentation, providing *m/z* readouts that recapitulated the original compound.

In MALDI MS, analytes are dried in a conducting metal plate in presence of a light-absorbing matrix. Then, a laser beam is directed to the crystalized sample, triggering the desorption and ionization of the analytes from the matrix. MALDI MS is characterized by generating mostly singly charged ions, therefore large biomolecules will also have large *m/z* ratios. Introduced slightly later than MALDI, the

development of electrospray ionization (ESI)[12] by M. Yamashita and J. B. Fenn was a key invention that drove the currently widespread use of MS for the analysis of biomolecules. Thanks to ESI, which allowed the ionization of large non-volatile biomolecules from a volatile solvent[13], "we learned to make elephants fly" – J. B. Fenn. Unlike MALDI, ESI easily generates multiply charged ions of the analytes, allowing the analysis of very large biomolecules at relatively low *m/z* values, and thus alleviating mass range limitations present in all mass analyzers. In ESI (**Figure 1A**), polar analytes such as peptides and proteins, are solubilized in a volatile buffer containing a conductive agent (e.g., formic acid). The analyte-containing
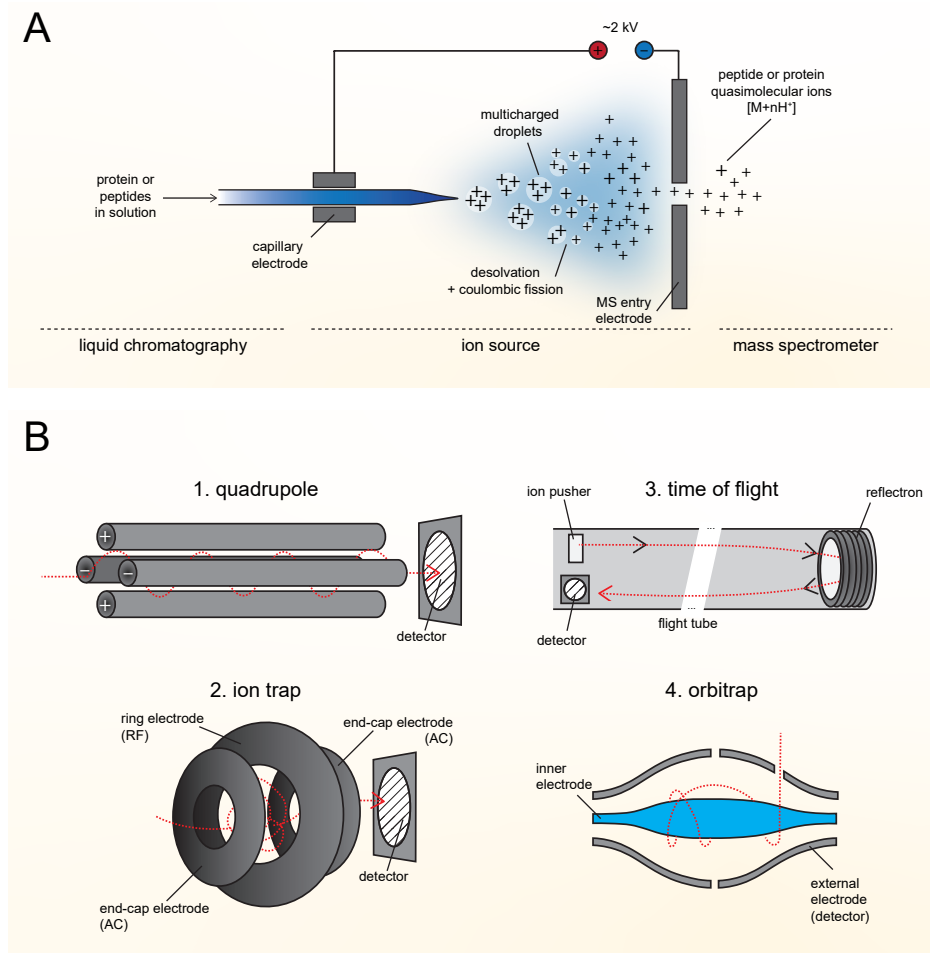
Figure 1. Schematics of electrospray ionization and main classes of mass analyzers found in commercial mass spectrometers. (**A**) Proteins or peptides are transported to a thin capillary in a solvent containing a conductive agent, a spray is formed by applying a large potential difference between the capillary and the MS entry electrode, multicharged droplets are formed and upon solvent evaporation and Coulomb fissions, quasi molecular ions [M+nH+] are formed and enter the mass spectrometer. (**B**) Four types of mass analyzer are displayed: quadrupoles and ion traps provide a lower resolution, and time of flight and orbitraps provide a higher resolution. Ion trajectory inside the mass analyzers is represented by a dashed red line.

solution is then transferred to a thin capillary where ionization takes place. This capillary is connected to an electrode, and sometimes to a heat source, to promote ionization and solvent evaporation (desolvation). Applying high voltage (kV) potential difference between the mass spectrometer entry electrode and the capillary electrode leads to spray formation, where droplets form, desolvate and undergo Coulomb fissions, leading to formation of quasi molecular ions [M+nH$^+$] of the analytes that will ultimately enter the mass spectrometer.

Mass analysis — Nowadays, mostly four types of mass analyzers are found in state-of-the-art mass spectrometers (**Figure 1B**). Generally, mass analyzers can be classified in terms of their resolving power, meaning their capacity to distinguish two ions of very similar *m/z* ratios[14]. High resolution is crucial to distinguish the naturally occurring isotopes of peptides and proteins, mostly harboring ^{13}C and ^{15}N isotopes, which ultimately translates in highly accurate mass readings that allow isobaric ion species differentiation. Moreover, isotopic resolution is also key for charge determination, which is required to calculate the mass of peptides. Since all the consecutive isotopologues of a peptidic biomolecule differ by 1 Dalton (Da), e.g., ~ 0.5 *m/z* for doubly charged ions and ~ 0.25 *m/z* for quadruply charged ions, the charge of the molecule can be extrapolated from the spacing between isotopic peaks. Finding these sequentially differing ion packages is therefore key to identifying peptide masses, as well as to accurately determining the mass of intact proteins. Thus, mass analyzers are often classified according to resolution, and can be subdivided into low-resolution analyzers, like quadrupoles and (linear) ion traps, or high-resolution analyzers, like Fourier-transform ion cyclotron resonance (FT-ICR), Orbitraps and time-of-flight (TOF), which can provide isotopic resolution for most peptides and proteins.

Quadrupoles and ion traps use direct and alternating currents to separate the ions, where different radiofrequencies (RF) allow for specific *m/z* selection[15]. In quadrupoles, ions travel through four rods to be detected (**Figure 1B-1**), while in ion traps the ions are first trapped within the electrodes and then ejected into the detector (**Figure 1B-2**). Although quadrupole and ion trap mass analyzers are very fast, they have two main limitations: low mass accuracy (50 to 100 parts per million, ppm) and, as previously stated, low resolving power. However, these type of fast mass analyzers are exceptionally good at ion isolation, focusing and transmission, and are consequently found in top-notch hybrid mass spectrometers, in combination with higher-resolution mass analyzers, such as Orbitraps and TOF.

Orbitraps, FT-ICR and TOF excel in terms of mass accuracy and resolving power [16,17], but Orbitraps and FT-ICR hold a higher resolving power. Conceptually, in TOF analyzers (**Figure 1B-3**), ions are equally accelerated to enter a field-free 'drift tube' or 'flight tube' of known distance where they travel at a certain velocity that is dependent on the ion's *m/z* value. Resolution in TOF instruments is somewhat constant over the whole *m/z* range, as it is mostly dependent on the distance travelled by the ions. Because longer travel distances will lead to higher resolution, reflectrons[18] that reverse the flight direction of ions have been incorporated to increase the flight distance, without increasing the length of the flight tube. One unique characteristic of TOF analyzers is that they can scan ions over an enormous mass

range[18] (for example, from 20 to 40,000 for timsTOF Pro 2 instrument), which allows the measurement of singly charged ions of high mass. Although recent developments in TOF-based mass spectrometers are revolutionizing ultra-high sensitivity proteomics[10], the work presented in this thesis has been acquired exclusively using Orbitrap-based mass spectrometers.Orbitraps and FT-ICRs and are ultrahigh resolution mass analyzers that belong to the ion trap category. On one hand, FT-ICR are the highest resolution mass analyzers, providing sub-ppm or even part per billion (ppb) resolution[19,20]. However, FT-ICRs are less common in many MS laboratories, mostly due to their high cost, large size and high maintenance requirements[17]. On the other hand, Orbitraps have revolutionized MS since its original development by A. Makarov in the year 2000 and are presently found in many laboratories in the world[21]. Orbitraps are modified ion traps where the ions are trapped by an electrostatic field formed between an outer, barrel-like, and an inner, spindle-like electrode[22] (**Figure 1B-4**). Upon trapping, ions enter an orbital motion around the inner electrode where they oscillate harmonically at a frequency proportional to $(m/z)^{-\frac{1}{2}}$. Fourier transformation of the recorded oscillations will provide the mass spectrum of the orbiting ions. In terms of resolution, Orbitraps provide a very high resolving power, higher than TOF-based instruments, although it is dependent on the ion's and the total oscillation time (transient time)[22,23]. Moreover, with the introduction of instrument modifications, the mass range of Orbitraps can be substantially extended to allow the analysis of megadalton assemblies[24–26].

Fragmentation and sequence deduction — Although high-resolution mass spectrometers can generate extremely precise *m/z* readouts, these would always be insufficient to distinguish molecular isomers which have an identical *m/z* value. For example, the molecules '*PEPTIDE*' and '*EEDIPPT*' have the same chemical formula and therefore, their *m/z* ratios are indistinguishable. Tandem MS (MS/MS) makes use of two events of mass analysis, in combination with ion isolation and fragmentation, to differentiate isomers based on their specific sequence composition (**Figure 2A, top**). First, a full scan (MS or MS1) is acquired at very high resolution, to resolve all the *m/z* species present in the sample. Next, a specific ion, namely precursor ion, is isolated based on its *m/z* ratio by means of quadrupoles or ion traps, and then transported to a fragmentation cell to be fragmented into smaller ions, namely product ions, that are finally transferred to the high-resolution mass analyzer to determine the MS/MS or MS2 spectrum of the original precursor ion. Therefore, product ions directly inform on the peptide backbone structure, as they contain amino acid sequence information that is key for isomer differentiation (**Figure 2A, bottom**)[27].

Ion fragmentation, also termed dissociation, is achieved by energetic destabilization of ions leading to rupture of chemical bonds. Depending on which chemical bond is ruptured, the resulting product ions are named as follows[28]: a-/x- fragments (fragmentation at the Cα—CO bonds), b-/y- fragments (fragmentation at the C—N bonds) or c-/z- fragments (fragmentation at the N—Cα bonds) (**Figure 2B**). Importantly, multiple fragmentation techniques have been developed and they preferentially generate specific ion series, leading to different applications in protein and peptide analysis. These fragmentation techniques can be classified lower-energy, such as collisional-based or infrared photon-based dissociation, or higher-energy,
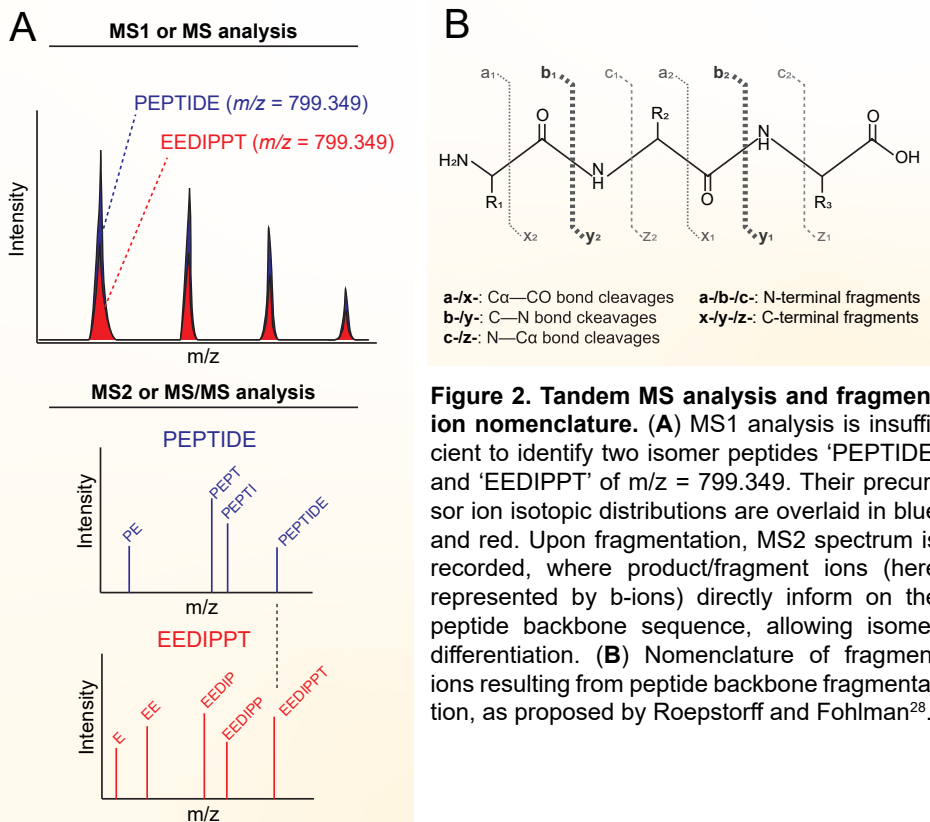
Figure 2. Tandem MS analysis and fragment ion nomenclature. (**A**) MS1 analysis is insufficient to identify two isomer peptides 'PEPTIDE' and 'EEDIPPT' of m/z = 799.349. Their precursor ion isotopic distributions are overlaid in blue and red. Upon fragmentation, MS2 spectrum is recorded, where product/fragment ions (here represented by b-ions) directly inform on the peptide backbone sequence, allowing isomer differentiation. (**B**) Nomenclature of fragment ions resulting from peptide backbone fragmentation, as proposed by Roepstorff and Fohlman[28].

such as electron-based and ultraviolet photon-based dissociation. In lower-energy dissociation techniques, enough energy must be accumulated over the ion's backbone to induce dissociation by redistribution of intramolecular vibrational energy. In higher-energy dissociation techniques, dissociation occurs rather immediately before redistribution of intramolecular vibrational energy. For instance, electron-based dissociation is a high-energetic fragmentation method, where mostly -c and -z fragment ions are generated, which retain certain PTMs[29–32] and help their accurate localization within the peptide or protein backbone. Similarly, certain photon-based dissociation methods relying on the use of high-energetic ultraviolet photons[33], produce mostly -a, -c, -x and -y fragment ions, and are frequently used to sequence the backbone of larger peptide species and intact proteins[33–36].

For the majority of the work presented in this thesis, collision-based dissociation has been the method of choice to fragment peptide precursors. This is because, compared to electron- or photon-based dissociation methods, collisional-based dissociation is readily iplemented in most commercially available mass spectrometers. Collisional-based dissociation is a type of low-energetic ergodic fragmentation method in which peptide ions are accelerated and collide with molecules of an innert gas, normally nitrogen. Upon sufficient collisions, kinetic energy accumulated in the ion's backbone exceeds the dissociation threshold, and then the ion is fragmented generating mostly b- and y- ions, as the peptide bond (C—N) is

the most labile bond under this type of dissociation. Originally, collision-induced dissociation (CID) took place inside ion traps, but the lower-energetic pathways accessed promoted loss of certain labile PTMs. Nowadays, a dedicated collision cell has been incorporated into most mass spectrometers, where higher-energy collision-induced dissociation (HCD) is performed[37] which has been shown to help retain labile PTMs[38,39] allowing better localization of the modification within the peptide backbone. Finally, the success of collision-based dissociation lies in its high charge-dependency, which makes it highly compatible with the analysis of multiply charged peptides obtained by ESI-MS.

1.2. Protein-centric and peptide-centric mass spectrometric approaches

A protein can be analyzed by MS either as a whole, or in fragments. The prior is also known as protein-centric MS. In this mode, intact proteins are ionized and injected into the mass spectrometer. This can be done in denaturing or native conditions, respectively known as top-down or native MS. On one hand, in top-down MS, a mixture of denatured intact proteins is chromatographically separated and ionized by ESI before tandem MS analysis[40]. However, intact protein fragmentation is not trivial, as lots of energy must be deposited in an efficient way over the protein backbone to achieve sufficient fragment formation as to gain sufficient sequence information. Therefore, multiple fragmentation techniques have been used, alone or in combination, to maximize the number of fragments generated and expand sequence coverage beyond the N- and C-terminal ends of the protein ion[33,36,41–43]. On the other hand, in native MS, relatively pure mixtures of proteins are dissolved in a non-denaturing buffer and directly ionized by ESI for mass spectrometric analysis[17]. In this MS mode, the tertiary (and even quaternary) structure of proteins may be retained, giving rise to more folded or compact ions, which will consequently acquire lower charges upon ionization than in denaturing top-down MS, where the full backbone of the protein is exposed. Therefore, in native MS, protein ions of lower charge will fall at higher *m/z* regions than in top-down MS, showcasing the importance of broad mass range scanning when using native approaches[17]. Overall, both top-down and native MS approaches can be extremely insightful when studying intact proteins, as they can directly inform on the protein structure, including macromolecular assemblies, and on co-existing variants of a single protein (i.e., proteoforms[2]). However, protein-centric approaches fall short in dealing with the huge complexity of the proteome. Although impressive advances have been made in protein-centric MS approaches, the field is still challenged by various technical limitations, including protein solubility, chromatographic separation, data analysis or dynamic range[44].

Considering that most crude biological samples consist of complex mixtures of different proteins that challenge protein-centric MS approaches, peptide-centric MS approaches, and more specifically bottom-up proteomics, has become the gold-standard to study complex mixtures of proteins, including various aspects of the proteome. The main driver behind the success of bottom-up proteomics is its ability to sufficiently reduce sample complexity so that multiple proteins can be simultaneously and reliably measured and/or quantified. Ironically, the first step for complexity reduction is achieved by proteolytic digestion of all proteins present in a

sample, generating an even more complex set of peptides. However, this complex peptide mixture is easily resuspended and can be chromatographically separated making it suitable for tandem MS analysis. Normally, the preferred enzyme for protein digestion is trypsin, which cleaves after lysine (K) and arginine (R) generating peptides that will carry at least two positive charges upon ionization, but alternative proteases can also be used to obtain complementary sets of peptides[45–47]. Since not all proteins contain well-spaced K/R residues over the protein backbone, obtaining complementary peptide sets can sometimes increase the coverage of the proteome.

1.3. Data acquisition strategies for peptide-centric mass spectrometric approaches

Usually, a data-dependent acquisition (DDA) strategy is used to record the MS/MS spectra of the peptides present in the sample. In DDA, a pre-scan is first recorded at high resolution, informing on the totality of precursor ions, of varying *m/z* and intensity, present in the sample at a given time. From these, typically, the top *N* most intense precursors are sequentially selected for fragmentation, and a spectral readout of the generated fragments is recorded at slightly lower resolution, to speed up the process. To avoid analyzing the same ion repeatedly, an exclusion time is set, during which the already analyzed ions will be excluded from the precursor selection algorithm. In the end, datasets containing MS1 and MS2 spectral information for tens of thousands of precursors are generated, and the protein content of the sample is reconstructed by means of database searching. In these search algorithms, the empirically measured spectral evidence is contrasted to the theoretical fragments of peptides. However, due to the multiple comparison problem, spectral evidence is prone to type I error, and therefore strict false discovery rate (FDR) corrections must be applied[48]. Normally, a decoy database is generated[49] with 'false' sequences, containing for example the reverse sequences of the proteins of interest. By contrasting the information to these 'false' sequences, as well as to the 'real' database, FDR can be estimated, and spectral matches that fall above 1 or 5 % FDR are discarded.

Another acquisition strategy named data-independent acquisition (DIA) or sequential window acquisition of all theoretical mass spectra (SWATH-MS)[50], has been more recently implemented. In this type of acquisition, narrow *m/z* isolation windows are defined so that all ions present in each pre-scan will be fragmented, instead of only the most intense ones as measured by DDA. Although DIA approaches can significantly boost sensitivity[50–53], the datasets generated are even larger than DDA approaches, and it requires reliable spectral libraries to identify multiple co-fragmenting precursors present in each isolation window. Peptide libraries were originally generated from DDA data, although novel developments using neural networks have proved key to generate reliable peptide spectral predictions for tryptic digests[51], separating the field from DDA measurements. However, the DIA field is still under development when it comes to generating spectral libraries for proteolytic digests obtaining using alternative proteases[54] or for endogenous peptides[55].

1.4. Versatility of peptide-centric approaches: from total protein content to structural characterization

Peptide-centric approaches are highly versatile, as demonstrated in various ways throughout the work presented in this thesis. Beyond proteome characterization by conventional bottom-up proteomics, peptide-centric approaches have also been applied to investigate PTMs, protein-protein interactions, proteolytic cleavage, protein function and turnover, and even protein structural determination. Peptide-centric approaches impose little restrictions on the kinds of samples; to date, shotgun methods have been instrumental in the detailed analysis of organelles, cells, or tissue extracts, and is a great toolbox for many disciplines, from systems biology to translational medicine. Here, I will briefly introduce the main peptide-centric approaches that have been used in this thesis (**Figure 3**), and how these are derived from technical modifications to the conventional bottom-up protocol.

One application of bottom-up proteomics is to study the broad range of PTMs that decorate the proteome. PTMs are known to maintain homeostasis of biological systems, as they control essential biological processes such as protein folding, degradation, signaling or secretion[56–59]. Accordingly, in diseases, dysregulations in PTMs can also be the underlying mechanisms[60–62]. In essence, PTMs involve the covalent addition of a functional group to the protein sequence, which comes with a known change in mass allowing MS detection within the peptide sequence. However, although some PTMs can be relatively abundant, other relevant PTMs occur at lower stoichiometries[63]. This highlights the need for enrichment of these low abundant post-translationally modified peptides from the non-modified fraction of the proteome.

To this end, multiple enrichment strategies have been developed to boost the coverage of specific post-translationally modified peptides. For example, phosphorylated peptides can be enriched using or titanium dioxide (TiO_2) or Fe^{3+} immobilized metal ion affinity chromatography (IMAC) which have a strong affinity for the negatively charged phosphate moieties, or glycopeptides, which can be enriched by hydrophilic interaction liquid chromatography (HILIC) due to the polar nature of the glycan moiety. Similarly, antibodies have also been used to enrich specific PTMs where a sequence motif is conserved at modification site, such as protein ubiquitination[64,65], or the protein tyrosine phosphatase (PTP) catalytic cysteine oxidation[66–68].

Another key application of bottom-up proteomics is crosslinking mass spectrometry (XLMS), capable of informing on the tertiary and quaternary structure of proteins[69]. In XLMS, a small molecule of fixed length, namely crosslinker, is used to covalently lock proteins in place before bottom-up analysis. Crosslinkers consist of a relatively short (10 to 20 Angstroms, or 0.1 to 0.2 nm) and flexible spacer arm ending with two reactive groups (typically NHS esters) that react covalently with specific amino acids (e.g., NHS ester-based crosslinkers react with amino acids containing primary amines, such as K and R). Crosslinkers can be either membrane permeable or non-permeable, and therefore can be used to crosslink from proteins in solution, to the cell periphery or the intracellular space. XLMS

1

is increasingly applied to purified protein complexes and is used to complement other structural determination techniques, such as cryogenic electron microscopy (cryo-EM) or crystallography, which usually have limited ability to resolve flexible regions of proteins. In addition, XLMS has also been applied to understand protein-protein interactions collectively at the organelle level[70], cellular fractions[71,72], or even whole cells[73,74].

Crosslinked peptides present slightly different properties than non-crosslinked tryptic peptides. Taking advantage of these different properties, multiple strategies have been developed to enrich the lower abundant crosslinked peptide fraction from the large non-crosslinked tryptic peptide fraction. Some approaches to enrich crosslinked peptides exploits the higher number of charges of these peptides (≥4⁺)

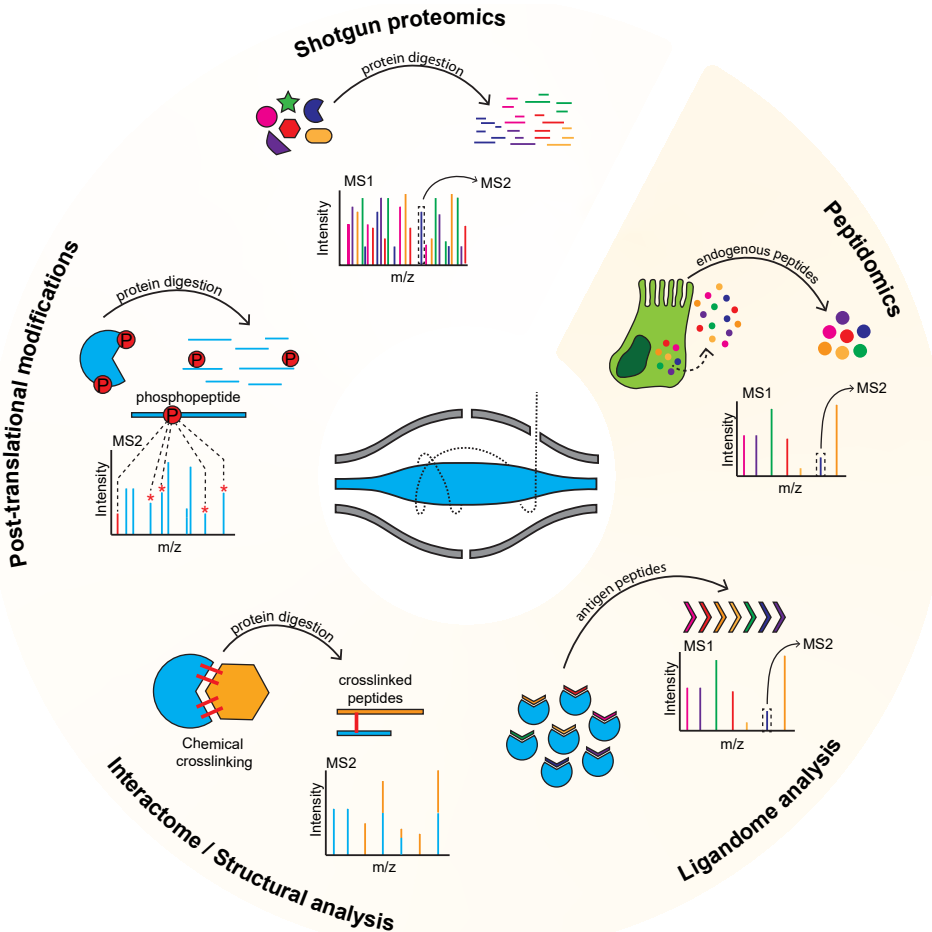**Figure 3. Peptide-centric MS approaches.** The main peptide-centric MS applications used in this thesis are presented. Various applications of peptide-centric MS are represented, including bottom-up proteomics, post-translational modification (PTMs) analysis, ligandome analysis, interactome/structural analysis and peptidomics. A brief schematic of the working principle is included for each of these applications.

when compared to non-crosslinked ones[75]. For example, cation exchange chromatography can be used to separate or enrich these crosslinked peptides[75], or alternatively, ion mobility mass spectrometry can be used to directly isolate these multiply charged and bigger peptides at the ion level[76]. Size exclusion chromatography[77] has also been used to enrich crosslinked peptides based on their larger size compared to non-crosslinked ones. Lastly, and more recently introduced, specific crosslinkers have been designed to allow direct enrichment of crosslinked peptides by affinity purification from the crosslinker itself[78,79].

Finally, peptidomics is an important branch of peptide-centric MS that studies the endogenous peptide composition of a sample, without experimental enzymatic digestion. This domain is important since the function of such endogenous peptides could differ completely from the source protein or genetically encoded polypeptide chain. These peptides may play important roles in signaling, acting as hormones, neurotransmitters, antimicrobial agents, or immune regulators[80], either as cleaved and activated translation products, or as enzymatically trimmed degradation products. Since endogenous peptides are produced by endogenous proteolytic processes in cells or biological matrices, the database search space is not restricted by enzymatic peptide ends. Consequently, enzymatic digestions are not included in peptidomics workflows, and the endogenous set of peptides is directly separated and characterized by tandem MS. Both liquid and solid phase extraction protocols have been used to extract peptides from complex samples[81–85], by separating these peptides from cellular proteins, nucleic acids and lipids. Depending on the biological origin [81,85–87], these peptides can vary greatly in their biochemical and biophysical properties. For example, peptide hormones can range from just a few amino acids to very long peptides of up to 100 amino acids[88], while class I human leukocyte antigen (HLA) bound ligands are prominently presented in lengths between 8 to 12 amino acids[89]. Therefore, analyzing such different peptidomes also requires very different analytical conditions in the mass spectrometer, as for example, the energy required to fragment and sequence short peptides is lower than for the larger ones.

It is important to mention the computational effort that certain peptide-centric MS applications suppose. The three applications described above (PTM-analysis, XLMS and peptidomics) lead to huge search spaces that result in very long computation time. For example, for a human proteomics analysis, standard databases contain already around 6 million tryptic peptides, while this number escalates to 20 million just by including phosphorylation of serine, threonine, and tyrosine. This number further escalates for crosslinked peptides[90], where each peptide in the database could be in theory crosslinked to any other peptide in the database. And for unspecific searches, where no cleavage site is defined, and windows of pre-defined lengths must be slid over the protein sequences. Specific search engines have been developed to decrease searching times[91–94], and to (partially) circumvent the search space limitations.

1.5. Improving sensitivity of peptide-centric approaches

In the work presented in this thesis, mass spectrometry has been applied to the study of specific secreted factors fundamental for intercellular communication. As detailed in the previous section, a plethora of peptide-centric MS approaches is currently available to study very intricate aspects of cell biology. However, their application to the detailed characterization of these specific secreted factors has been vastly held back by technical limitations affecting analysis sensitivity. These limitations can be narrowed down to two main factors, i.e., sample complexity and sample availability (**Figure 4A**), and should always be carefully evaluated before choosing any mass spectrometric approach. In MS, sensitivity is broadly understood as a performance metric of mass spectrometers, but it can also be understood in multiple ways[95]. For instance, sensitivity can be defined as the signal-to-noise ratio (S/N) measured for a given amount of sample, but also as the minimum signal that can be detected from the system's noise, what is better known as limit of detection. Yet another definition for sensitivity comes in terms of the measured signal for a fixed amount of analyte in different mass spectrometers. Here, newer,
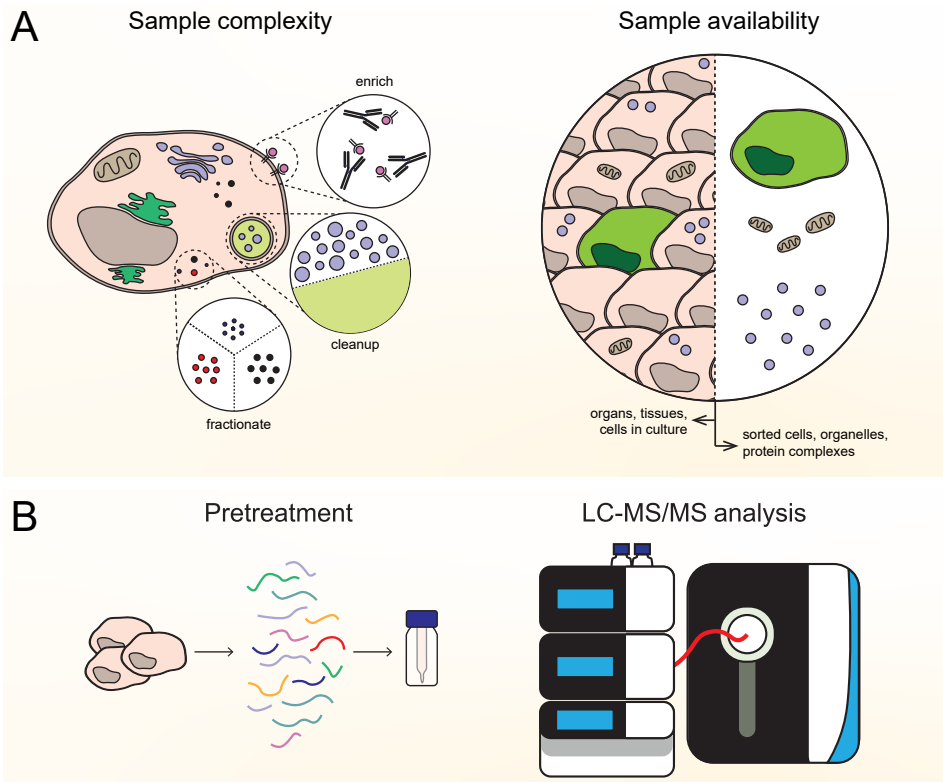

Figure 4. Sensitivity in peptide-centric mass spectrometry. (**A**) Sample complexity and sample availability are the two main factors affecting sensitivity of peptide centric approaches. (**B**) Adequately addressing these factors in a sample-specific manner in the pretreatment and LC-MS/MS steps of the experimental approach will help improve sensitivity.

more sensitive instruments provide higher readouts of intensity and therefore hold a lower limit of detection than older, less sensitive instruments. Therefore, improving any of these will consequently lead to sensitivity improvements. In this section, we will explore the concepts of sample complexity and sample availability, with a focus on strategies available to boost sensitivity, at the levels of sample pretreatment and LC-MS/MS analysis (**Figure 4B**).

On the concepts of sample complexity and availability — Many strategies to improve sensitivity intend to reduce sample complexity. Without complexity reduction, sensitivity is highly limited by co-fragmentation and the dynamic range of proteins, especially in DDA approaches. On one hand, co-fragmentation occurs when multiple peptide precursors are co-isolated in the quadrupole, and therefore fragmented together leading to mixed spectra containing fragments of heterogeneous origin. On the other hand, proteins span a wide dynamic range[96], from one to ten million copies per cell[97], translating in preferential identification of the higher abundant species. When complexity is sufficiently reduced, co-isolation and co-fragmentation are minimized, leading to highly pure spectra and therefore higher sensitivity. Similarly, reducing sample complexity increases the chance of detecting lower abundant species, alleviating the dynamic range problem. However, complexity reduction can be challenging where sample availability is limited (e.g., subcellular or single-cell proteomics), as enrichment, cleanup or fractionation strategies always gain sensitivity at the expense of sample losses.

Despite sample availability limitations, the interest in studying more specific biological processes, such as those happening in specialized cellular compartments or in individual cells, are driving the field towards lower-input proteomics. This is exemplified by the recent achievements of the single-cell proteomics community, where 1000 – 2000 proteins can be now measured from a single cell[98,99], even starting to provide insights in specific biological processes[100]. Most of the work presented here has been performed on samples of limited availability. Thus, in full consideration of the challenges that this represents, we carefully selected sensible pretreatment and LC-MS/MS strategies to minimize sample losses, reduce sample complexity, and maximize sensitivity. Some of the strategies available to boost sensitivity during sample pretreatment or LC-MS/MS analysis are described below.

Pretreatment strategies to improve sensitivity — Ideally, sample pretreatment strategies aimed to improve sensitivity should intend to reduce sample complexity, and this can be done in three main ways: enrichment, cleanup and/or fractionation (**Figure 4A**). A commonly used pretreatment strategy is *off-line* enrichment or fractionation, which makes use of physicochemical interactions of the analytes with a stationary phase to achieve separation. For example, pipette tips packed with chromatographic material (such as Octadecylsilane (C18) or strong cation exchange (SCX) stationary phases) are routinely used to capture peptides generated in a proteolytic digest, to isolate these peptides while removing salts and other contaminants, and subsequently elute these before LC-MS/MS analysis. A stepped elution of the bound analytes inside these tips can be used to generate fractions of the sample, to be individually analyzed by LC-MS/MS obtaining further

depth on the (phospho)proteome[101,102] or (immuno)peptidome[103] of interest. For this approach, the greatest depth will be achieved by using complementary chromatographic separation strategies to the one used for LC-MS/MS. Beyond its use in sample cleanup and fractionation, *off-line* chromatography can also be key to enrich for specific analytes within a complex sample, for instance in the previously explained cases of XLMS or PTM analysis. Similarly, size-based separation strategies can also help boost sensitivity for certain peptide-centric MS approaches, such as in the case of (immuno)peptidomics. To analyze endogenous peptides, these must be first separated from the intact protein fraction. This can easily be achieved by using molecular weight cut-off filter devices or protein precipitation strategies. Another way to specifically enrich analytes of interest is by means of affinity purification, in what is called affinity purification mass spectrometry (AP-MS). Multiple pull-down strategies have been developed to perform AP-MS, that employ antibodies or sequence tags to pull-down proteins or peptides of interest[67,103–106].

LC-MS/MS strategies to improve sensitivity — The most common strategy to reduce sample complexity in peptide-centric approaches is *on-line* chromatographic separation. With the introduction of ESI, liquid chromatography could be coupled to tandem mass spectrometry (LC-MS/MS), providing high-resolution separation directly at the peptide level and becoming the gold-standard technique for proteome analysis[4]. In LC-MS/MS, normally C18-based reverse phase chromatography is used to separate peptides. Briefly, peptides obtained in the pretreatment step are resuspended in a polar mobile phase and loaded into the LC system, where they bind the C18 stationary phase to be slowly eluted based on hydrophobicity using a gradient of increasing organic percentage in the mobile phase. Other types of stationary phases (C8, HILIC, SCX, etc.) can be used both alone and in combination[107–109], to further improve analyte separation based on their physicochemical properties. In terms of improving chromatographic separation sensitivity, parameters affecting peak capacity, such as column length, gradient length, and stationary phase particle size[110–112], should be considered. Generally, longer columns, longer gradient lengths and smaller particles size positively affect peak capacity and sensitivity. However, during MS analysis of low input material (i.e., sub-microgram), where the abundance of analytes is low, shorter gradients are preferred over longer gradients[113], as they maximize ion intensities, leading to more peptides triggering MS/MS events and consequently to a better proteome coverage[114].

Beyond *on-line* liquid chromatography, sensitivity can also be improved inside the mass spectrometer. For example, newer generations of mass spectrometers like Orbitrap HF-X and Exploris 480, contain improved ion transfer tubes and electromagnetic ion funnels that improve transmission efficiency[5,6] leading to improved sensitivity. In terms of ion transmission, it is important to mention the role of the S-lens. Essentially functioning as a stacked-ring ion guide that uses RF to increase the ion flux into the mass spectrometer[115], the S-lens directly improves sensitivity. In the S-lens, the RF level can be optimized to improve transmission of smaller or larger ions, depending on the peptide-centric application of choice. In general, larger RF values facilitate the transmission of larger ions, so modifications to this parameter can also aid in improving sensitivity, for example, in peptidomics. Similarly, S/N ratios can be substantially improved by filtering out the singly charged

ions[116], which are attributable to contaminants in most peptide-centric approaches. This can be done using high field asymmetric waveform ion mobility spectrometry (FAIMS)[6,116], an ion mobility technique that separates gas-phase ions before entry to the mass spectrometer.

Another strategy to improve sensitivity within the mass spectrometer involves adjustments to the automatic gain control (AGC) target and to the maximum ion injection time. The AGC controls the number of ions that enter the Orbitrap for each scan, to avoid saturation or space charge effects. Normally, AGC target values of $1\text{-}3\times10^6$ ions for MS1 and 1×10^5 for MS2 are good starting points to achieve good mass accuracy and sensitivity. The AGC target is normally calculated based on a pre-scan, where the ion current is measured for a specific mass range. Based on this pre-scan, the required time to accumulate sufficient ions as to meet the AGC target is calculated. Then, ions are accumulated in the C-trap until the AGC target value is met, followed by injection in the Orbitrap[117]. When the AGC target cannot be met, ions will be accumulated until the selected maximum injection time (*maxIT*) is reached. Therefore, when sample availability is limited, dynamic range is broad or analytes are very low in abundance, a good strategy to improve sensitivity is selecting longer *maxIT*, which will promote accumulation of as many ions as possible before Orbitrap injection. However, when the AGC target is continuously unmet, longer *maxIT* can lead to considerably fewer precursors scanned per cycle. This is because filling the C-trap accounts for most of the cycle time and this can be circumvented by slightly modifying the AGC target values.

PART 2 – From the inside-out and from the outside-in

The work presented in this thesis has been driven by the desire to better understand the elegant nuances of intercellular communication, that ironically takes place at low stoichiometry. This dissertation focuses on the study of two fundamental units of intercellular communication: peptide hormones and extracellular vesicles (EVs). Despite their longstanding relevance in intercellular communication, peptide hormones and EVs remain under-studied largely due to analytical challenges. However, taking advantage of the versatility of peptide-centric mass spectrometric approaches, and by thoughtful application of adequate experimental strategies to boost sensitivity, in this work we overcame some of these challenges and gained novel insights in peptide hormone and EV biology.

2.1. Cellular communication

Cellular communication is a universal process that allows cells to exchange information with each other and their surroundings. Cells communicate with each other by exchanging materials that, for instance, take the forms of EVs and peptide hormones. For example, in unicellular organisms, cellular communication via EVs[118,119] is key to coordinate certain responses, which promote colony growth and survival and even drive antibiotic resistance[120–122]. In multicellular organisms, even more specialized communication processes enable coordination of complex biological activities within and between tissues, ultimately allowing a huge collection of cells to interact altogether with their environment. Overall, cellular communication is an intricate process that comprises multiple specialized communication units that facilitate the exchange of information. Cell-to-cell communication can be classified as local, when it involves information transfer between cells in close proximity, or long-range, when the information is transferred over distal cells or organs (**Figure 5**). On one hand, local cell-to-cell communication can happen by either direct interaction (juxtacrine signaling), via membrane protein interactions or GAP junctions, or indirect interaction (paracrine signaling) via secreted factors like metabolites, peptides, cytokines or EVs (**Figure 5A**). On the other hand, long-range distance cell-to-cell communication happens on endocrine organs, where endocrine cells secrete signaling molecules, named hormones, capable of traveling through the bloodstream to distal parts of the body to exert their functions (**Figure 5B**).

A key organelle for cellular communication is the plasma membrane. The plasma membrane confines the cellular responses to a region in space, while integrating the cell with the extracellular world. Specifically, proteins inserted or associated with the plasma membrane act by bridging the inside and outside of cells. For example, from the *'outside-in'*, membrane receptors, such as G-coupled protein receptors or integrins, have evolved to bind a multitude of paracrine, juxtacrine or endocrine ligands. In general, ligand binding induces a conformational change in the receptor that initiates a cascade of intracellular events, such as a series of phosphorylation, culminating in a specific biological response to the initial signal, such as the transcription of a gene. Conversely, signals may also be transmitted in the reverse direction in an "inside-out" manner. A good example is juxtacrine signaling in the form of antigen presentation via HLA class I molecules. This type of

antigen presentation allows cells to communicate their health status to the immune system. Membrane-bound HLA class I molecules are loaded intracellularly with endogenous peptides, generated by proteasomal degradation of proteins in the cytosol, that act as a barcode of the health status of the cell. Therefore, unhealthy cells (e.g., tumoral cells or infected cells) display at the cell surface 'non-self' peptide antigens bound to the HLA, which upon contact with reactive T-cell receptors trigger immune synapse formation (**Figure 5A**) and cytotoxic T-cell mediated cell lysis[123]. It is even more intriguing that the same mode of juxtacrine signaling can also be packaged into EVs for long-range signaling, as demonstrated here in this thesis.
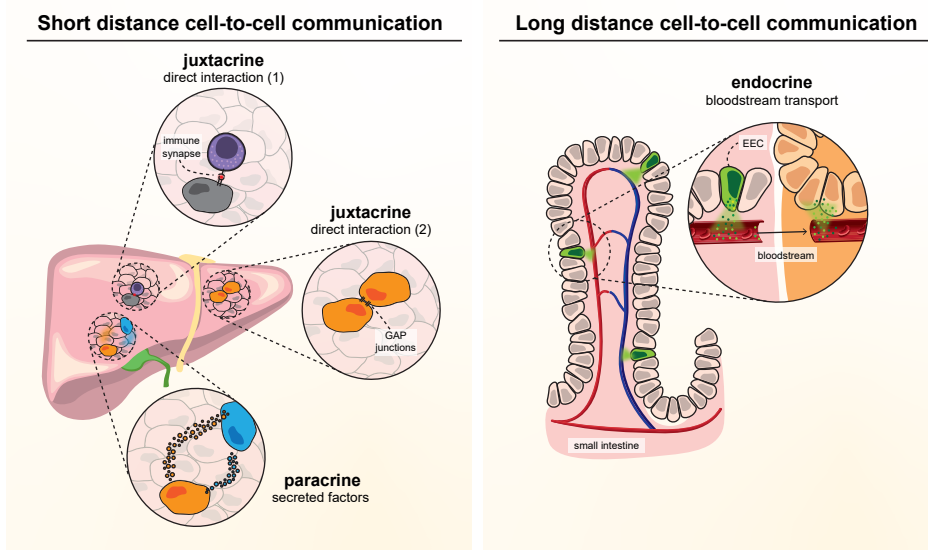

Figure 5. Modes of cell-to-cell communication. (**A**) Local cell-to-cell communication between neighboring cells can happen by direct interaction via membrane proteins, like in immune synapses, or via cytoplasmic conducts named GAP junctions (juxtacrine signaling); and by indirect interaction via secreted factors, such as peptides, cytokines or EVs (paracrine signaling). (**B**) Long-range cell-to-cell communication takes place in endocrine organs like the small intestine, where secreted factors are transported via bloodstream to act at distal body sites.

Intimately linked to the plasma membrane, endosomes are also key organelles for cellular communication. Endosomes help regulate receptor degradation and recycling, while providing an enclosed environment of highly specific conditions that promote certain signaling processes[124]. Endosomes are complex organelles, presenting multiple states of maturation and carrying out diverse functions that range from membrane receptor recycling to exosome formation. Exosomes are small EVs – 20 to 200 nanometers (nm) – that originate by endocytosis and mature in the late endosomes or multivesicular bodies. (MVB) Although MBV can fuse with lysosomes for protein degradation, they can also fuse with the plasma membrane releasing exosomes into the extracellular milieu (**Figure 6**). However, cells can also produce EVs by direct membrane budding, which tend to be bigger – 100 to

1,000 nanometers (nm) – and are named microvesicles. Although exosomes and microvesicles partly overlap in size and content, the field is slowly moving towards a definition of a more specific set of markers to distinguish these types of EVs[125].

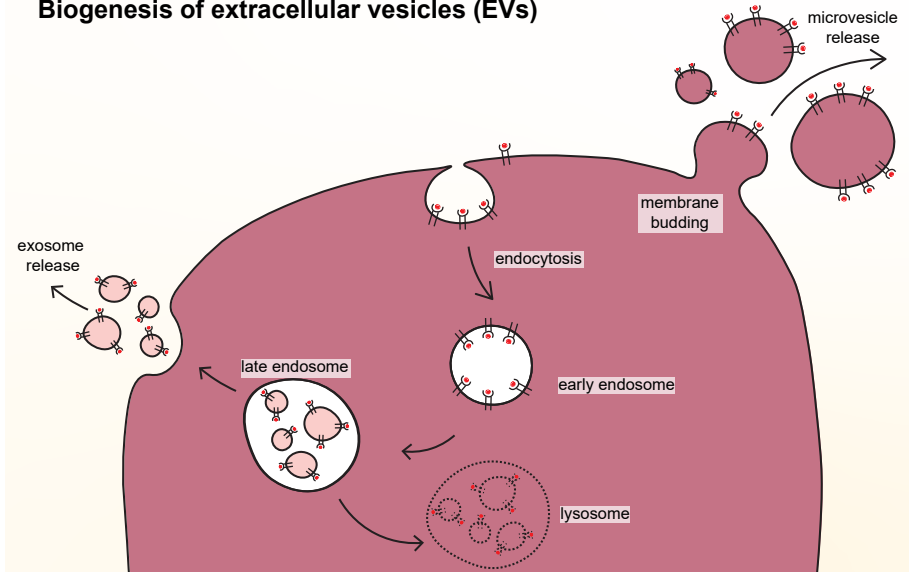**Biogenesis of extracellular vesicles (EVs)**

Figure 6. Biogenesis of extracellular vesicles (EVs). Exosomes, a type of EVs, originate by endocytosis from the plasma membrane, maturation in the endosomal pathway and release to the extracellular space by late endosome fusion to the plasma membrane. Sometimes, late endosomes can fusion with lysosomes ending up in lysis of their content. Microvesicles, another type of EVs, originate by direct buffing from the plasma membrane, and are immediately released into the extracellular space.

Independent of their origin, EVs are enclosed by a phospholipid bilayer that originates from plasma membrane invagination or evagination, and therefore always contain some plasma membrane proteins that can mediate docking with other cells by means of receptor-ligand interactions, that can also trigger signal transduction or endocytosis and consequent release of content into the recipient cells [126]. Such EV contents can vary from nucleic acids and metabolites to peptides and proteins[127]. More importantly, EV stability is great even in human plasma[128,129], and they have been found in all biological fluids[130,131]. Remarkably, proteins present in the vesicular membrane can target these EVs to specific cells or organs, and are thus responsible for EV tissue tropism. Due to these properties, EVs can be seen as hormone-like players in cell-to-cell communication[132] implying that, beyond endocrine organs, any cell can potentially perform long distance cell-to-cell communication and opens an avenue for therapeutic interventions.

2.2. Scope of this thesis

In this thesis, I have applied high-sensitivity mass spectrometry approaches to comprehensively characterize EVs and peptide hormones, two fundamental modes of cell-to-cell communication. With carefully tailored sample pretreatment

and LC-MS/MS strategies, I make bold attempts to overcome technical hurdles to obtain novel biological insights into both systems. The work presented here will lay strong foundation to better understand and perturb elements of cell-cell communication for therapy.

In **Chapter 2**, I sought to optimize an immunopeptidomics workflow to work with highly limited amounts of starting material. This pushes the limit not only in the EV field, but also in the antigen discovery field. With the strategy and considerations outlined in this chapter, I expanded the measurable EV antigen repertoire by 8-fold, to ~3,500 HLA peptides. This new depth of identification allowed me to contrast the properties of the EV ligand repertoire to that displayed in the cell surface, revealing that EVs are not only densely loaded with HLA molecules, but also over represent HLA-B and cysteinylated ligands.

In **Chapter 3**, I attempted to study the surface features of EVs. Prompted by the lack of structural information of protein complexes that reside in EV membrane, I pursued structural characterization of protein complexes in EVs, to understand the unique protein-protein interactions that may enable EV functionality. This led to an intact extracellular vesicle crosslinking (iEVXL) approach, combining disuccinimidyl suberate (DSS) crosslinking and mass spectrometry analysis. Importantly, XLMS can take a snapshot of proteins in close proximity at the surface and in the lumen of intact EVs, where proteins can interact in their native environment. With this strategy, we add a novel approach to the very limited toolbox for structural characterization of EVs.

In **Chapter 4**, I applied (phospho)proteomics to characterize the effects that short-term EV-treatment has on cardiomyocyte function, with the biological goal of understanding how EVs may support cardiac regeneration. In this chapter, the content of EVs derived from cardiac progenitor cells (CPCs) was examined by bottom-up proteomics, which shortlisted PAPPA and NID-1 as potential mediators of endothelial cell activation and migration. In parallel, changes on the phospho-proteome of endothelial cells upon treatment with CPC-EVs revealed altered Insulin/IGF pathway and was corroborated by a panel of significantly altered phosphosites downstream of the PI3K-AKT and MAPK signaling pathways. As PAPPA has been proposed to mediate extracellular release of IFG-1 by disrupting the IGF1/IGFBP4 complex, the data presented here, further supported by various validations *in vitro*, indicates that PAPPA contributes to the pro-regenerative properties of CPC-EVs.

In **Chapter 5 and Chapter 6**, I applied sensitive MS approaches to study peptide hormone secretion and processing. By combining all the low input strategies and considerations outlined in the introduction to a panel of human enteroendocrine organoids, we could follow the protein content of human EEC subpopulations. This led to an EEC atlas (Chapter 5), providing the first transcriptomic and proteomic map of Motilin-producing cells, a human EEC population that cannot be replicated and studied in mice. In this EEC atlas, we could ascribe key differences between human and murine EECs, including hormones, exopeptidases, sensory receptors, and transcription factors, as well as characterize the hormones secreted by these specialized endocrine cells. Collectively, this constitutes a rich resource to study human EEC development and function.

1

In **Chapter 6**, I further explored the endogenous processing of these gut-derived peptide hormones by measuring hormone peptides from genetically modified EECs. By knocking out or mutating key enzymes involved in the endogenous processing, the hormone peptide processing specificity was monitored. From this, we identified PCSK2-dependent glucagon production in intestinal EECs, stimulated upon BMP signaling. Integrating data from all the enzymatic mutants, we could further map the substrates and products of major EECs endo- and exopeptidases, including PCSK1, PCSK2, DPP4, CPE and CPB1. CPB1 is a novel L-cell specific carboxypeptidase we discovered in **Chapter 5**. This work provides an integral atlas of peptide hormones secreted by human EECs, identifying many bioactive peptide hormones, and describing the roles of specific proteases in their generation.

Finally, in **Chapter 7**, I will discuss the remaining challenges on the field of high-sensitivity MS analysis applied to the study of secreted mediators of cell-to-cell communication, together with the future perspectives for general high-sensitivity MS analysis of peptides and proteins.

References

1.	Aebersold, R. et al. How many human proteoforms are there? Nature Chemical Biology 14, 206–214 (2018).

2.	Smith, L. M. & Kelleher, N. L. Proteoform: a single term describing protein complexity. Nature Methods 10, 186–187 (2013).

3.	Kim, M. S. et al. A draft map of the human proteome. Nature 509, 575–581 (2014).

4.	Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. Nature 537, 347–355 (2016).

5.	Kelstrup, C. D. et al. Performance Evaluation of the Q Exactive HF-X for Shotgun Proteomics. J Proteome Res 17, 727–738 (2018).

6.	Bekker-Jensen, D. B. et al. A Compact Quadrupole-Orbitrap Mass Spectrometer with FAIMS Interface Improves Proteome Coverage in Short LC Gradients. Molecular & Cellular Proteomics 19, 716 (2020).

7.	Meier, F. et al. Online parallel accumulation–serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. Molecular & Cellular Proteomics 17, 2534–2545 (2018).

8.	Mund, A. et al. Deep Visual Proteomics defines single-cell identity and heterogeneity. Nature Biotechnology 1–10 (2022) doi:10.1038/s41587-022-01302-5.

9.	Marcu, A. et al. HLA Ligand Atlas: a benign reference of HLA-presented peptides to improve T-cell-based cancer immunotherapy. Journal for ImmunoTherapy of Cancer 9, e002071 (2021).

10.	Brunner, A.-D. et al. Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. Molecular Systems Biology 18, e10798 (2022).

11.	Budzikiewicz, H. & Grigsby, R. D. Mass spectrometry and isotopes: A century of research and discussion. Mass Spectrometry Reviews 25, 146–157 (2006).

12.	Yamashita, M. & Fenn, J. B. Electrospray ion source. Another variation on the free-jet theme. Journal of Physical Chemistry 88, 4451–4459 (1984).

13.	Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. Science (1979) 246, 64–71 (1989).

14.	Todd, J. F. J. Recommendations for nomenclature and symbolism for mass spectroscopy (including an appendix of terms used in vacuum technology). Pure and Applied Chemistry 63, 1541–1566 (1991).

15.	March, R. E. Quadrupole ion traps. Mass Spectrometry Reviews 28, 961–989 (2009).

16.	Marshall, A. G. & Hendrickson, C. L. High-Resolution Mass Spectrometers. Annual Review of Analytical Chemistry 1, 579–599 (2008).

17.	Tamara, S., den Boer, M. A. & Heck, A. J. R. High-Resolution Native Mass Spectrometry. Chemical Reviews 122, 7269–7326 (2021).

18.	Mamyrin, B. A. Time-of-flight mass spectrometry (concepts, achievements, and prospects). International Journal of Mass Spectrometry 206, 251–266 (2001).

19.	Hendrickson, C. L. et al. 21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer: A National Resource for Ultrahigh Resolution Mass Analysis. J Am Soc Mass Spectrom 26, 1626–1632 (2015).

20.	Savory, J. J. et al. Parts-per-billion fourier transform ion cyclotron resonance mass measurement accuracy with a walking calibration equation. Analytical Chemistry 83, 1732–1736 (2011).

21.	Makarov, A. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. Analytical Chemistry 72, 1156–1162 (2000).

22.	Zubarev, R. A. & Makarov, A. Orbitrap mass spectrometry. Analytical Chemistry 85, 5288–5296 (2013).

23.	Lössl, P., Snijder, J. & Heck, A. J. R. Boundaries of mass resolution in native mass spectrometry. J Am Soc Mass Spectrom 25, 906–917 (2014).

24.	Fort, K. L. et al. Expanding the structural analysis capabilities on an Orbitrap-based mass spectrometer for large macromolecular complexes. Analyst 143, 100–105 (2017).

25.	Wörner, T. P. et al. Frequency chasing of individual megadalton ions in an Orbitrap analyser improves precision of analysis in single-molecule mass spectrometry. Nature Chemistry 14, 515–522 (2022).

26.	van de Waterbeemd, M. et al. High-fidelity mass analysis unveils heterogeneity in intact ribosomal particles. Nature Methods 14, 283–286 (2017).

27.	Steen, H. & Mann, M. The abc's (and xyz's) of peptide sequencing. Nature Reviews Molecular Cell Biology 5, 699–711 (2004).

28.	Roepstorff, P. & Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomed Mass Spectrom 11, 601–601 (1984).

29.   Kleinnijenhuis, A. J., Kjeldsen, F., Kallipolitis, B., Haselmann, K. F. & Jensen, O. N. Analysis of Histidine Phosphorylation Using Tandem MS and Ion−Electron Reactions. Analytical Chemistry 79, 7450–7456 (2007).

30.   Srikanth, R. et al. Improved sequencing of oxidized cysteine and methionine containing peptides using electron transfer dissociation. J Am Soc Mass Spectrom 18, 1499–1506 (2007).

31.   Stensballe, A. et al. Electron capture dissociation of singly and multiply phosphorylated peptides. Rapid Commun. Mass Spectrom 14, 1793–1800 (2000).

32.   Mirgorodskaya, E., Roepstorff, P. & Zubarev, R. A. Localization of O-Glycosylation Sites in Peptides by Electron Capture Dissociation in a Fourier Transform Mass Spectrometer. Analytical Chemistry 71, 4431–4436 (1999).

33.   Julia, B.-M. & Oliveira, E. de. Photodissociation Mass Spectrometry of Peptides and Proteins. Encyclopedia of Analytical Chemistry 1–22 (2018) doi:10.1002/9780470027318.A9620.

34.   Fort, K. L. et al. Implementation of Ultraviolet Photodissociation on a Benchtop Q Exactive Mass Spectrometer and Its Application to Phosphoproteomics. Anal Chem 88, 2303–2310 (2016).

35.   Brodbelt, J. S., Morrison, L. J. & Santos, I. Ultraviolet Photodissociation Mass Spectrometry for Analysis of Biological Molecules. Chemical Reviews 120, 3328–3380 (2020).

36.   Greisch, J. F. et al. Expanding the mass range for UVPD-based native top-down mass spectrometry. Chemical Science 10, 7163–7171 (2019).

37.   Olsen, J. v. et al. Higher-energy C-trap dissociation for peptide modification analysis. Nature Methods 4, 709–712 (2007).

38.   Cui, L., Yapici, I., Borhan, B. & Reid, G. E. Quantification of competing H3PO4 Versus HPO 3 + H2O neutral losses from regioselective 18O-labeled phosphopeptides. J Am Soc Mass Spectrom 25, 141–148 (2014).

39.   Diedrich, J. K., Pinto, A. F. M. & Yates, J. R. Energy dependence of HCD on peptide fragmentation: stepped collisional energy finds the sweet spot. J Am Soc Mass Spectrom 24, 1690–1699 (2013).

40.   Schaffer, L. v. et al. Identification and Quantification of Proteoforms by Mass Spectrometry. Proteomics 19, 1800361 (2019).

41.   Cannon, J. R. et al. Ultraviolet photodissociation for characterization of whole proteins on a chromatographic time scale. Analytical Chemistry 86, 2185–2192 (2014).

42.   Riley, N. M. et al. Enhanced Dissociation of Intact Proteins with High Capacity Electron Transfer Dissociation. J Am Soc Mass Spectrom 27, 520–531 (2016).

43.   Riley, N. M., Westphall, M. S. & Coon, J. J. Activated Ion Electron Transfer Dissociation for Improved Fragmentation of Intact Proteins. Analytical Chemistry 87, 7109–7116 (2015).

44.   Melby, J. A. et al. Novel Strategies to Address the Challenges in Top-Down Proteomics. J Am Soc Mass Spectrom 32, 1278–1294 (2021).

45.   Tsiatsiani, L., Heck, A. J. R., Heck, A. J. R. & Bijvoet, P. Proteomics beyond trypsin. The FEBS Journal 282, 2612–2626 (2015).

46.   Giansanti, P., Tsiatsiani, L., Low, T. Y. & Heck, A. J. R. Six alternative proteases for mass spectrometry–based proteomics beyond trypsin. Nature Protocols 11, 993–1006 (2016).

47.   van der Laarse, S. A. M. et al. Targeting proline in (phospho)proteomics. The FEBS Journal 287, 2979–2997 (2020).

48.   Diz, A. P., Carvajal-Rodríguez, A. & Skibinski, D. O. F. Multiple Hypothesis Testing in Proteomics: A Strategy for Experimental Work. Molecular & Cellular Proteomics 10, (2011).

49.   Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nature Methods 4, 207–214 (2007).

50.   Gillet, L. C. et al. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. Molecular & Cellular Proteomics 11, O111.016717 (2012).

51.   Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nature Methods 17, 41–44 (2019).

52.   Bruderer, R. et al. Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. Molecular and Cellular Proteomics 14, 1400–1410 (2015).

53.   Barkovits, K. et al. Reproducibility, specificity and accuracy of relative quantification using spectral librarybased data-independent acquisition. Molecular and Cellular Proteomics 19, 181–197 (2020).

54.   Gessulat, S. et al. Prosit: proteome-wide

1

prediction of peptide tandem mass spectra by deep learning. Nature Methods 16, 509–518 (2019).

55.   Pak, H. S. et al. Sensitive immunopeptidomics by leveraging available large-scale multi-HLA spectral libraries, data-independent acquisition, and MS/MS prediction. Molecular and Cellular Proteomics 20, 100080 (2021).

56.   Hochstrasser, M. Ubiquitin-dependent protein degradation. Annu Rev Genet 30, 405–439 (1996).

57.   Chufán, E. E., De, M., Eipper, B. A., Mains, R. E. & Amzel, L. M. Amidation of Bioactive Peptides: The Structure of the Lyase Domain of the Amidating Enzyme. Structure 17, 965 (2009).

58.   Roth, J. Protein N-glycosylation along the Secretory Pathway: Relationship to organelle topography and function, protein quality control, and cell interactions. Chemical Reviews 102, 285–303 (2002).

59.   Ardito, F., Giuliani, M., Perrone, D., Troiano, G. & Muzio, L. lo. The crucial role of protein phosphorylation in cell signalingand its use as targeted therapy (Review). International Journal of Molecular Medicine 40, 271–280 (2017).

60.   Chen, L., Liu, S. & Tao, Y. Regulating tumor suppressor genes: post-translational modifications. Signal Transduction and Targeted Therapy 5, 1–25 (2020).

61.   Markiv, A., Rambaruth, N. D. S. & Dwek, M. V. Beyond the genome and proteome: targeting protein modifications in cancer. Current Opinion in Pharmacology 12, 408–413 (2012).

62.   Moumné, L., Betuing, S. & Caboche, J. Multiple Aspects of Gene Dysregulation in Huntington's Disease. Front Neurol 4, (2013).

63.   Prus, G., Hoegl, A., Weinert, B. T. & Choudhary, C. Analysis and Interpretation of Protein Post-Translational Modification Site Stoichiometry. Trends in Biochemical Sciences 44, 943–960 (2019).

64.   Udeshi, N. D., Mertins, P., Svinkina, T. & Carr, S. A. Large-Scale Identification of Ubiquitination Sites by Mass Spectrometry. Nat Protoc 8, 1950 (2013).

65.   Steger, M. et al. Time-resolved in vivo ubiquitinome profiling by DIA-MS reveals USP7 targets on a proteome-wide scale. Nature Communications 12, 1–13 (2021).

66.   Groen, A. et al. Differential Oxidation of Protein-tyrosine Phosphatases. Journal of Biological Chemistry 280, 10298–10304 (2005).

67.   Wu, W., Hale, A. J., Lemeer, S. & den Hertog, J. Differential oxidation of protein-tyrosine

phosphatases during zebrafish caudal fin regeneration. Scientific Reports 7, 1–9 (2017).

68.   Elvira, B. et al. PTPN2 Regulates the Interferon Signaling and Endoplasmic Reticulum Stress Response in Pancreatic β-Cells in Autoimmune Diabetes. Diabetes 71, 653–668 (2022).

69.   O'Reilly, F. J. & Rappsilber, J. Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. Nature Structural & Molecular Biology 25, 1000–1008 (2018).

70.   Liu, F., Lössl, P., Rabbitts, B. M., Balaban, R. S. & Heck, A. J. R. The interactome of intact mitochondria by cross-linking mass spectrometry provides evidence for coexisting respiratory supercomplexes. Molecular & Cellular Proteomics 17, 216 (2018).

71.   Armony, G., Heck, A. J. R. & Wu, W. Extracellular crosslinking mass spectrometry reveals HLA class I – HLA class II interactions on the cell surface. Molecular Immunology 136, 16–25 (2021).

72.   Fasci, D., Ingen, H. van, Scheltema, R. A. & Heck, A. J. R. Histone Interaction Landscapes Visualized by Crosslinking Mass Spectrometry in Intact Cell Nuclei. Molecular & cellular proteomics 17, 2018–2033 (2018).

73.   Lenz, S. et al. Reliable identification of protein-protein interactions by crosslinking mass spectrometry. Nature Communications 12, 1–11 (2021).

74.   Wheat, A. et al. Protein interaction landscapes revealed by advanced in vivo cross-linking-mass spectrometry. Proc Natl Acad Sci U S A 118, (2021).

75.   Fritzsche, R., Ihling, C. H., Götze, M. & Sinz, A. Optimizing the enrichment of cross-linked products for mass spectrometric protein analysis. Rapid Communications in Mass Spectrometry 26, 653–658 (2012).

76.   Ihling, C. H., Piersimoni, L., Kipping, M. & Sinz, A. Cross-Linking/Mass Spectrometry Combined with Ion Mobility on a timsTOF Pro Instrument for Structural Proteomics. Analytical Chemistry 93, 11442–11450 (2021).

77.   Leitner, A. et al. Expanding the Chemical Cross-Linking Toolbox by the Use of Multiple Proteases and Enrichment by Size Exclusion Chromatography. Molecular & Cellular Proteomics 11, (2012).

78.   Steigenberger, B., Pieters, R. J., Heck, A. J. R. & Scheltema, R. A. PhoX: An IMAC-Enrichable Cross-Linking Reagent. ACS Central Science 5, 1514–1522 (2019).

79.   Jiang, P. L. et al. A Membrane-Permeable and Immobilized Metal Affinity Chromatography (IMAC) Enrichable Cross-Linking Reagent to Advance In Vivo Cross-Linking Mass Spectrometry. Angewandte Chemie International Edition 61, e202113937 (2022).

80.   Tinoco, A. D. & Saghatelian, A. Investigating Endogenous Peptides and Peptidases using Peptidomics. Biochemistry 50, 7447 (2011).

81.   Bauzá-Martinez, J. et al. Proteolysis in septic shock patients: plasma peptidomic patterns are associated with mortality. British Journal of Anaesthesia 121, 1065–1074 (2018).

82.   Fleites, L. A. et al. Peptidomics Approaches for the Identification of Bioactive Molecules from Diaphorina citri. Journal of Proteome Research 19, 1392–1408 (2020).

83.   Miedzybrodzka, E. L. et al. Organoid Sample Preparation and Extraction for LC-MS Peptidomics. STAR Protocols 1, 100164 (2020).

84.   Keller, C. et al. Extraction optimization for combined metabolomics, peptidomics, and proteomics analysis of gut microbiota samples. Journal of mass spectrometry 56, (2021).

85.   Dingess, K. A., van den Toorn, H. W. P., Mank, M., Stahl, B. & Heck, A. J. R. Toward an efficient workflow for the analysis of the human milk peptidome. Analytical and Bioanalytical Chemistry 411, 1351–1363 (2019).

86.   Van, J. A. D. et al. Peptidomic analysis of urine from youths with early type 1 diabetes reveals novel bioactivity of uromodulin peptides in vitro. Molecular and Cellular Proteomics 19, 501–517 (2020).

87.   Hölttä, M. et al. An integrated workflow for multiplex CSF proteomics and peptidomics-identification of candidate cerebrospinal fluid biomarkers of alzheimer's disease. Journal of Proteome Research 14, 654–663 (2015).

88.   Malandrino, N. & Smith, R. J. Principles of Endocrinology and Hormone Action: Synthesis, Secretion, and Transport of Peptide Hormones. (Springer, Cham, 2018). doi:10.1007/978-3-319-44675-2_3.

89.   Rock, K. L., Reits, E. & Neefjes, J. Present Yourself! By MHC Class I and MHC Class II Molecules. Trends Immunol 37, 724 (2016).

90.   Yılmaz, Ş. et al. Cross-linked peptide identification: A computational forest of algorithms. Mass Spectrometry Reviews 37, 738–749 (2018).

91.   Yu, F. et al. Identification of modified peptides using localization-aware open search. Nature Communications 11, 1–9 (2020).

92.   Chen, Z. L. et al. A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. Nature Communications 10, 3404 (2019).

93.   Kong, A. T., Leprevost, F. v., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii, A. I. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics. Nature Methods 14, 513–520 (2017).

94.   Wenger, C. D. & Coon, J. J. A proteomics search algorithm specifically designed for high-resolution tandem mass spectra. Journal of Proteome Research 12, 1377–1386 (2013).

95.   Ingle, J. D. Sensitivity and limit of detection in quantitative spectrometric methods. Journal of Chemical Education 51, 100–105 (1974).

96.   Beck, M. et al. The quantitative proteome of a human cell line. Molecular Systems Biology 7, 549 (2011).

97.   Zubarev, R. A. The challenge of the proteome dynamic range and its implications for in-depth proteomics. Proteomics 13, 723–726 (2013).

98.   Schoof, E. M. et al. Quantitative single-cell proteomics as a tool to characterize cellular hierarchies. Nat Commun 12, 3341 (2021).

99.   Ctortecka, C. et al. Comparative Proteome Signatures of Trace Samples by Multiplexed Data-Independent Acquisition. Molecular & cellular proteomics 21, (2022).

100.   Brunner, A.-D. et al. Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. Molecular Systems Biology 18, e10798 (2022).

101.   Batth, T. S., Francavilla, C. & Olsen, J. v. Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics. Journal of Proteome Research 13, 6176–6186 (2014).

102.   Bekker-Jensen, D. B. et al. An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes Article An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. Cell Systems 4, 587–599 (2017).

103.   Demmers, L. C., Heck, A. J. R. & Wu, W. Pre-fractionation Extends but also Creates a Bias in the Detectable HLA Class I Ligandome. Journal of Proteome Research 18, 1634–1643 (2019).

104.   Keilhauer, E. C., Hein, M. Y. & Mann, M. Accurate protein complex retrieval by affinity enrichment mass spectrometry (AE-MS) rather than affinity purification mass spectrometry (AP-MS). Molecular and Cellular Proteomics 14, 120–135 (2015).

105.   Schiapparelli, L. M. et al. Direct detection of biotinylated proteins by mass spectrometry. Journal of Proteome Research 13, 3966–3978 (2014).

106.   Kleinpenning, F., Steigenberger, B., Wu, W. & Heck, A. J. R. Fishing for newly synthesized proteins with phosphonate-handles. Nat Commun 11, 3244 (2020).

107.   Gilar, M., Olivova, P., Daly, A. E. & Gebler, J. C. Orthogonality of Separation in Two-Dimensional Liquid Chromatography. Analytical Chemistry 77, 6426–6434 (2005).

108.   Zhou, F., Sikorski, T. W., Ficarro, S. B., Webber, J. T. & Marto, J. A. Online nanoflow reversed phase-strong anion exchange-reversed phase liquid chromatography-tandem mass spectrometry platform for efficient and in-depth proteome sequence analysis of complex organisms. Analytical Chemistry 83, 6996–7005 (2011).

109.   Marino, F. et al. Characterization and usage of the EASY-spray technology as part of an online 2D SCX-RP ultra-high pressure system. Analyst 139, 6520–6528 (2014).

110.   Liu, H. et al. Effects of column length, particle size, gradient length and flow rate on peak capacity of nano-scale liquid chromatography for peptide separations. Journal of Chromatography A 1147, 30–36 (2007).

111.   Petersson, P., Frank, A., Heaton, J. & Euerby, M. R. Maximizing peak capacity and separation speed in liquid chromatography. Journal of Separation Science 31, 2346–2357 (2008).

112.   Hsieh, E. J., Bereman, M. S., Durand, S., Valaskovic, G. A. & MacCoss, M. J. Effects of column and gradient lengths on peak capacity and peptide identification in nanoflow LC-MS/MS of complex proteomic samples. J Am Soc Mass Spectrom 24, 148–153 (2013).

113.   Stejskal, K. et al. Analysis of 8000 proteins and reduced carry over significantly increase the throughput of single-shot proteomics. bioRxiv 2021.11.28.470272 (2022) doi:10.1101/2021.11.28.470272.

114.   Orsburn, B. C. Evaluation of the sensitivity of proteomics methods using the absolute copy number of proteins in a single cell as a metric. Proteomes 9, 34 (2021).

115.   Olsen, J. v. et al. A dual pressure linear ion trap orbitrap instrument with very high sequencing speed. Molecular & Cellular Proteomics 8, 2759–2769 (2009).

116.   Swearingen, K. E. & Moritz, R. L. High-field asymmetric waveform ion mobility spectrometry for mass spectrometry-based proteomics. Expert Review of Proteomics 9, 505–517 (2012).

117.   Eliuk, S. & Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. Annual Review of Analytical Chemistry 8, 61–80 (2015).

118.   Ñahui Palomino, R. A., Vanpouille, C., Costantini, P. E. & Margolis, L. Microbiota–host communications: Bacterial extracellular vesicles as a common language. PLoS Pathogens 17, (2021).

119.   Díaz-Garrido, N., Badia, J. & Baldomà, L. Microbiota-derived extracellular vesicles in interkingdom communication in the gut. Journal of Extracellular Vesicles 10, (2021).

120.   Fulsundar, S. et al. Gene transfer potential of outer membrane vesicles of Acinetobacter baylyi and effects of stress on vesiculation. Appl Environ Microbiol 80, 3469–3483 (2014).

121.   Kim, S. W. et al. Outer membrane vesicles from β-lactam-resistant Escherichia coli enable the survival of β-lactam-susceptible E. coli in the presence of β-lactam antibiotics. Sci Rep 8, 5402 (2018).

122.   Kim, S. W. et al. Significant increase in the secretion of extracellular vesicles and antibiotics resistance from methicillin-resistant Staphylococcus aureus induced by ampicillin stress. Scientific Reports 10, 1–14 (2020).

123.   Rock, K. L., Reits, E. & Neefjes, J. Present Yourself! By MHC Class I and MHC Class II Molecules. Trends Immunol 37, 724 (2016).

124.   Sorkin, A. & von Zastrow, M. Endocytosis and signalling: intertwining molecular networks. Nature Reviews Molecular Cell Biology 10, 609–622 (2009).

125.   Jeppesen, D. K. et al. Reassessment of Exosome Composition. Cell 177, 428–445 (2019).

126.   Tkach, M. & Théry, C. Communication by Extracellular Vesicles: Where We Are and Where We Need to Go. Cell 164, 1226–1232 (2016).

127.   Tkach, M. & Théry, C. Communication by Extracellular Vesicles: Where We Are and Where We Need to Go. Cell 164, 1226–1232 (2016).

128.   Sokolova, V. et al. Characterisation of exosomes derived from human cells by nanoparticle tracking analysis and scanning electron microscopy. Colloids and Surfaces B: Biointerfaces 87, 146–150 (2011).

129.   Kalra, H. et al. Comparative proteomics evaluation of plasma exosome isolation techniques and assessment of the stability of exosomes in normal human blood plasma. Proteomics 13, 3354–3364 (2013).

130.   de la Torre Gomez, C., Goreham, R. v., Bech Serra, J. J., Nann, T. & Kussmann, M.

"Exosomics"-A review of biophysics, biology and biochemistry of exosomes with a focus on human breast milk. Frontiers in Genetics 27, 9–92 (2018).

131.   Raimondo, F., Morosi, L., Chinello, C., Magni, F. & Pitto, M. Advances in membranous vesicle and exosome proteomics improving biological understanding and biomarker discovery. Proteomics 11, 709–720 (2011).

132.   Salomon, C. et al. Extracellular Vesicles and Their Emerging Roles as Cellular Messengers in Endocrinology: An Endocrine Society Scientific Statement. Endocrine Reviews 43, 441–468 (2022).

1

CHAPTER 2

HLA-B and cysteinylated ligands distinguish the antigen presentation landscape of extracellular vesicles

Julia Bauzá-Martinez[1,2], Albert J.R. Heck[1,2], Wei Wu[1,2]*

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

[2] Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands

Based on:

Summary

Extracellular vesicles can modulate diverse processes ranging from proliferation and tissue repair, to chemo-resistance and cellular differentiation. With the advent of tissue and immunological targeting, extracellular vesicles are also increasingly viewed as promising vectors to deliver peptide-based cancer antigens to the human immune system. Despite the clinical relevance and therapeutic potential of such 'cell-free' approaches, the natural antigen presentation landscape exported in extracellular vesicles is still largely uncharted, due to the challenging nature of such preparations and analyses. In the context of therapeutic vesicle production, a critical evaluation of the similarity in vesicular antigen presentation is also urgently needed. In this work, we compared the HLA-I peptide ligandomes of extracellular vesicles against that of whole-cells of the same cell line. We found that extracellular vesicles not only over-represent HLA-B complexes and peptide ligands, but also cysteinylated peptides that may modulate immune responses. Collectively, these findings describe the pre-existing provision of vesicular HLA complexes that may be utilized to carry peptide vaccines, as well as the propensity for different peptide and post-translationally modified ligands to be presented, and will outline critical considerations in devising novel EV vaccination strategies.

Introduction

Cancer immunotherapy has advanced substantially in the last decade, not only with breakthroughs involving checkpoint blockade inhibitors[1], but also due to conceptual advances in anti-tumor vaccines. To date, vaccinology approaches directed at eliciting anti-tumor effects using the patient's own immune system have shown great promise in a variety of cancers such as melanoma[2–4], non-small cell lung carcinoma[5,6] and glioblastoma[7].

In practice, antigenic vaccines need to be loaded on antigen presenting cells (APCs) such as dendritic cells. This can be done exogenously before injection of these loaded APCs into the patient, or antigens may also be infused into the patient blood as carried by extracellular vesicles (EVs), to take advantage of the natural EV uptake[8] by APCs that remain in the body[9]. The latter is often referred to as 'cell-free vaccines'[10,11], which has been shown to be more immunogenic than introducing soluble antigens[12], likely due to the trained uptake of EVs by APCs. Via either approach, the ultimate goal is for these antigenic signals to reach the lymph node, the primary site of T cell priming, so as to kickstart activation and expansion of clonal T cell populations, that may in turn be reactive against the patient's own tumor[11,13–17].

Apart from RNA-based therapeutics delivered via EVs[18–21], which are already promising, peptide-based antigens may mimic the normal antigenic processes in the body more closely, as if a real tumor-specific peptide is presented. The superior stability of peptide vaccines is also desirable when considering the shelf-life and formulation of such anti-cancer therapeutics. From screening the EV contents of a large panel of cell lines[22], the conserved presence of HLA proteins in EVs was also noted. Collectively, these allude to an exciting possibility to deliver antigens by taking advantage of these EV HLAs.

With recent advances in mass spectrometry-based immunopeptidomics that we and others have contributed[23–25], greater depth has been achieved in HLA peptide profiling from cell lines, tissue specimen[26–29] and increasingly scarce and exotic material[30]. Nevertheless, to our knowledge only one publication has so far reported the identification of ~500 HLA peptide ligands eluted from EVs[31]. This limited exploration into the EV HLA peptide ligandome stems from the challenges of obtaining sufficient EV material, on top of the need for extensive purifications before immunoaffinity capture of HLA complexes. Yet, gaining more knowledge into the EV HLA peptide repertoire is critical to tap on the possibility of tailored antigen delivery. In this respect, specific questions arise, for instance, the resemblance between EV and whole-cell HLA peptide ligandomes in terms of repertoire and post-translational modifications, as well as whether there are specific characteristics of peptides that are prone to become presented via the EVs. Outlining such rules will guide the intelligent design of novel EV vaccination strategies.

Using a highly sensitive workflow adapted from our established HLA-I peptide purification method[24], we compared the EV HLA-I peptide ligandome against the whole-cell HLA-I peptide ligandome of the same cells. Such a direct comparison is much needed to establish unique features of EV HLA peptide presentation, that could be therapeutically exploitable. Our study described here led to the identification of

~3,500 HLA-I peptide ligands, many of which were uniquely detected from EVs. We also observed that HLA-B peptide binders and peptides containing cysteinylated cysteines were more prominently presented by the EV HLA complexes. Taken together, these findings describe the pre-existing provision of HLA complexes on EVs and the propensity for different peptide ligands and modified ligands to be presented on EV HLA-I complexes. Such data outlines critical considerations in devising novel EV vaccination strategies and would be a valuable resource for the tumor immunology community.

Results

Isolation and characterization of JY-derived extracellular vesicles

To compare the HLA-I content and contrast the HLA Class I peptide repertoire in EVs and whole-cells (WC), we first prepared and characterised JY whole-cells and EV material (**Figure 1A**). In order to preserve the biological validity of direct comparisons between these two sources of HLA-I proteins and HLA class I peptide ligands, JY whole-cells and JY secreted EVs from the same culture batch were always used in each paired analysis. As described in the Materials and Methods, EV-depleted culture media was replaced after 24h to harvest a second batch of EVs, to ensure the cultured cells remain well-supplied with nutrients and were not apoptotic. Both EV harvests were then pooled for subsequent use to eliminate variabilities due to cell density. From a total of 800 ml JY conditioned media, at least 150 µg of EV proteins were obtained for further characterisation by nanoparticle tracking analysis (NTA), negative stain transmission electron microscopy (NS-TEM), SDS-PAGE, western blotting and label-free quantitative proteomics, before this material was used as starting material for HLA-I protein and HLA class I peptide ligand retrieval (**Figure 1B-1G**).

The EV isolate was high in both purity and integrity, as demonstrated by a range of biophysical and proteomics approaches. Through NTA (**Figure 1B**), the JY EVs we isolated were on average about 120 nm in diameter. This is well within the expected range of 50-150 nm range to be classified as exosomes. The EV preparation was almost devoid of particles exceeding 300 nm in size, verifying that the isolates were pure and free from cell debris contaminants and apoptotic bodies. The EVs were also cup-shaped in negative stain transmission electron microscopy (NS-TEM) (**Figure 1C**), in strong agreement with documented classical morphology that is characteristic of dehydrated EVs[32]. On lysates of JY whole-cells and EVs, we also verified that dominant proteins in the EVs were distinct from prominent proteins in JY whole-cells (**Figure 1D**), and that CD81, a classical EV marker, was strongly enriched in EV lysates and below detection in the whole-cell lysate (**Figure 1E**). Together these data demonstrate that EVs were specifically isolated, with no visible contaminations of abundant cytoplasmic proteins. A deeper unbiased proteomic comparison of both lysates showed many other classical EV markers were highly enriched in the EV proteome (e.g. Rab GTPases, integrins or annexins), while intracellular compartment markers (e.g. cytochrome c, GOLGA2, HSP90) were depleted in the vesicles (**Figure 1F-G**).
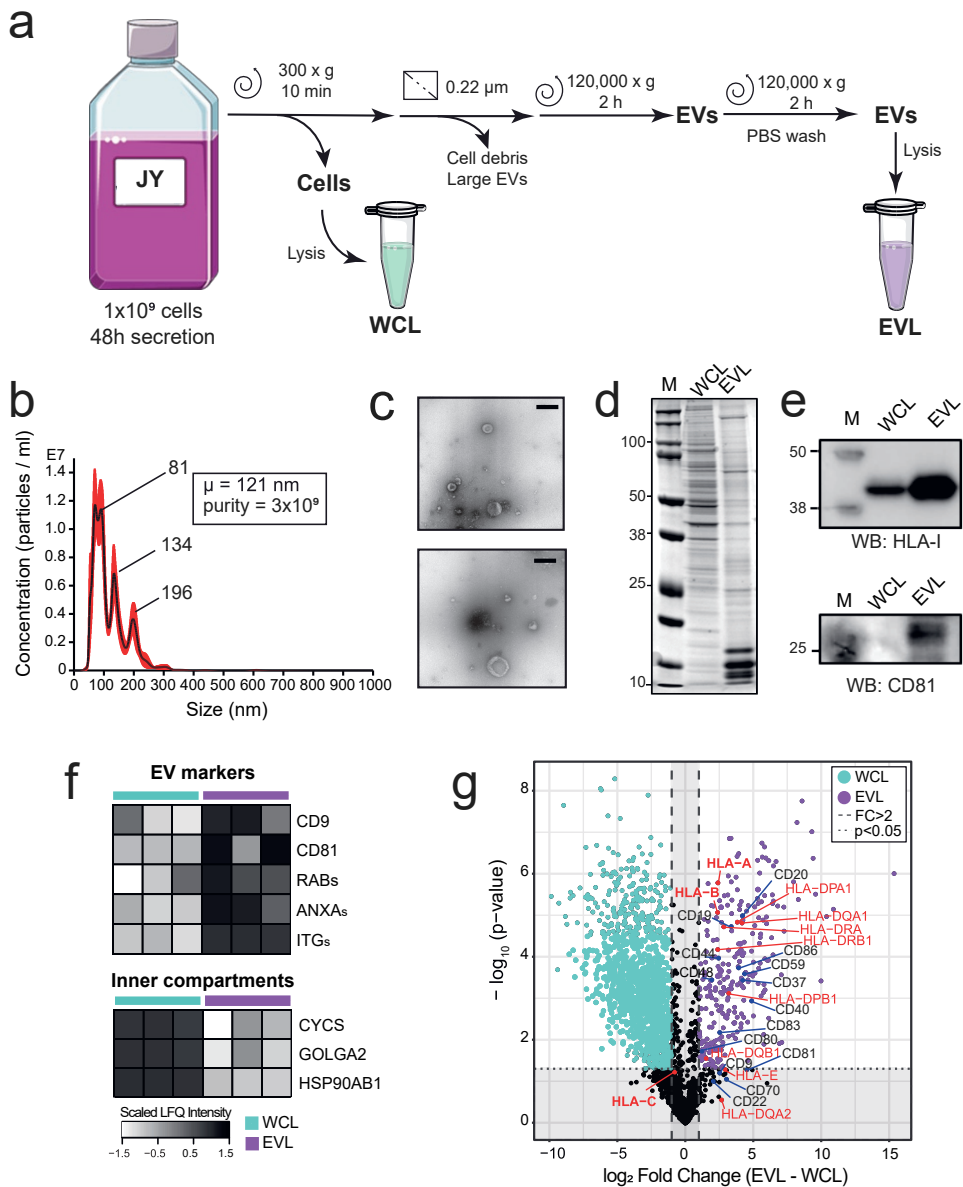
2

Figure 1. Extracellular vesicle isolation and characterization. (a) Schematic workflow of extracellular vesicle (EV) isolation. Lysates made from JY whole-cells (WCL, green) and JY-derived EVs (EVL, purple) were used subsequently as input for immunoaffinity purification of HLA complexes (Figures 2-4). (b) Nanoparticle tracking analysis (NTA). Size and concentration of isolated EVs were estimated. Purity (particle:protein ratio) of EVs was 3e9, with mean particle size of 121 nm. Representative trace. (c) Negative stain transmission electron microscopy (NS-TEM) analysis. Morphology of the isolated EVs were cup-shaped. Representative image. Scale bar: 200 nm. (Continues on next page)

Figure 1. Extracellular vesicle isolation and characterization. Full image provided in Supplementary Figure 6. (d) SDS-PAGE of protein band patterns in WCL and EVL. Dominant protein bands in WCL and EVL were mutually exclusive. Full image provided in Supplementary Figure 7. (e) HLA-I and cd81 western blot. From 5 µg load of EVL and WCL, prominent enrichment of HLA-I and CD81 were observed in EVL. Full image provided in Supplementary Figure 8. (f) Quantitative proteomics analysis of WCL (n=3) and EVL (n=3). Numerous classical EV markers were enriched in EVL (g) Quantitative proteomics comparison of EVL and WCL. HLA proteins, numerous cell surface CD proteins and tetraspanins were significantly enriched in EVL. Dashed lines indicate proteins with a 2-fold change in abundance, while dotted lines indicate proteins with p-value ≤ 0.05.

Collectively, these data provide strong and diverse evidence to show that our EV preparation from JY cells was of high quality, and that these paired JY whole-cell and EV lysates are good and valid starting material to probe differences in the cellular and extra-vesicular HLA class I peptide ligandome.

EVs and whole-cells derived HLA-I ligands display similar sequence traits

To compare the HLA-I content and contrast the HLA Class I peptide repertoire in whole-cells (WC) and EVs, we performed immunoaffinity purification of HLA-I proteins and HLA class I peptide ligands from JY whole-cell and EV lysates respectively, using the same input load and the same set of immunoaffinity capture steps (**Figure 2A**). On acid elution and molecular weight cut-off filtration, HLA class I peptide ligands were separated from HLA-I proteins. In the <10kDa filtrate (lower) fraction, HLA class I peptide ligands were recovered for LC-MS/MS analyses. From the >10kDa retentate (upper) fraction, HLA-I protein and interactors were digested prior to MS analysis. By detailed characterization of both fractions, we aim to compare and contrast both the HLA class I peptide ligand diversity, as well as the interaction network of HLA-I proteins in different spatial compartments.

Performing HLA class I peptide ligandome experiments from minute amount of EV material required a drastic down-scaling (~20x) of our established workflow[24], as it was not feasible to obtain milligrams of EVs in contrast to whole-cells. With an adapted high sensitivity workflow, we immunoprecipitated HLA-I proteins from equal input (120 µg) of JY whole-cell lysate or EV lysate. Despite using a much-reduced starting input compared to most other reports of HLA peptide profiling, we did not compromise on purification stringency. As shown in **Figure 2B**, more than 80% of all detected ligands were predicted to bind HLA-A*02:01, HLA-B*07:02, or HLA-C*07:02, the defined homozygous HLA type of our JY cell line model. This matches the high ligand purity we typically obtain from HLA class I peptide analyses of whole-cells. As shown in **Figure 2C**, more HLA-I proteins were immunoprecipitated from paired load of EV material, suggesting that EVs have much higher HLA content per microgram of protein compared to whole-cells. This was consistent also with the larger number of HLA class I peptide ligands detected from EV material (~3,500 vs ~1,400 from 120 µg of JY whole-cells, **Figure 2D**, right panel), and further corroborated by quantitative proteomics analysis of the 10 kDa filter retentate, where the total intensity of immunoaffinity-captured HLA-I proteins was also higher in EVs (**Figure 2D**, left panel). HLA class I peptide ligands from JY whole-cells and EVs distributed similarly in length (**Figure 2E**), and were also composed of peptides with similar
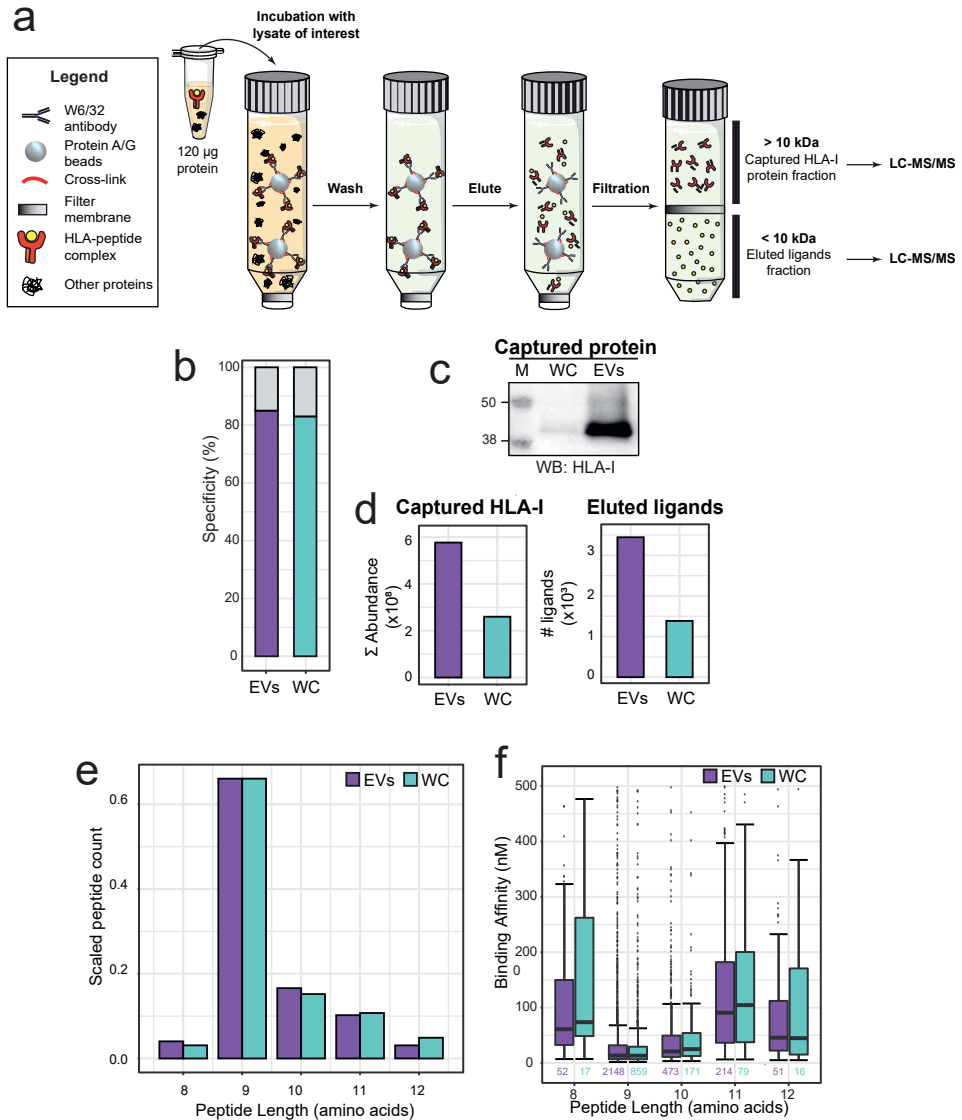
Figure 2. Global characterisation of the EV HLA-I peptide ligandome. (a) Preparative steps in the down-scaled HLA-I complex retrieval workflow. An input load of 120 µg was used for immunoaffinity purification of HLA-I complexes from both WCL and EVL. The flowthrough from 10 kDa filtration contains HLA peptide ligands that were subjected to LC-MS/MS analysis. The retentate contains HLA-I and HLA-I interacting proteins, which were digested, for the analysis of HLA-I abundance and interactome. (b) HLA-I peptide ligand specificity. From the 10 kDa flowthrough fraction, more than 80% of peptides detected were predicted to bind to the JY HLA type (HLA-A*02:01; HLA-B*07:02; HLA-C*07:02). (c) Western blot (WB) detection of HLA-I. HLA-I was detected in the eluate of WCL HLA-I IP, and more strongly in the eluate of EVL HLA-I IP. Full image provided in Supplementary Figure 9. (Continues on next page)

Figure 2. Global characterisation of the EV HLA-I peptide ligandome. (d) Mass spectrometry analysis of immuno-purified HLA-I proteins and HLA-I peptide ligands. From EVs, more HLA-I proteins (cumulative MS intensities of HLA-A, HLA-B and HLA-C) and HLA-I peptide ligand species were detected. (e) Peptide length distribution. HLA-I peptide ligands from both whole cells (WC) and EVs distributed similarly, and expectedly, between 8 to 12 amino acids. (f) Predicted binding affinity. Marginal differences in predicted binding affinity were observed in HLA-I peptides retrieved from WC and EVs. Box plots represent n peptides (where n has been annotated under each box). The 25%, 50% (median) and 75% quantiles are represented in each box, and the whiskers represent the ±1.5*IQR from the closest quantile. Bar plots represent the total pool of eluted ligands detected in three technical replicates from either WC (green) or EVs (purple).

sequence motifs (**Supplementary Figure 1**), albeit a marginally lower nM binding affinity (stronger binding) was observed in EV peptide ligands (**Figure 2F, Supplementary Figure 2**). Taken together, these data reveal that HLA class I peptides on EVs are similar in sequence properties to ligands on whole-cells, except that EVs are more densely inserted with HLA-I complexes.

EVs preferentially present HLA-B peptide ligands

Considering identification overlap in HLA class I peptide ligands, 86% of all the peptides detected from JY whole-cells were also detected in the EV ligandome. In addition, EV HLA peptide profiling revealed another 2265 ligands that were not detected from JY whole-cells (**Figure 3A**). This expanded detection may result partially from EVs being more densely loaded with HLA complexes (increased abundance leads to better MS identification), but may also reflect greater substrate peptide diversity arising from the endosomal route during EV biogenesis.

Even though the EV HLA class I peptides appeared similar in sequence properties compared to the JY whole-cell counterpart (**Figure 2**), a larger proportion of HLA-B binders was detected from EVs compared to whole-cells (**Figure 3B**). From JY whole-cells, HLA-B binders made up 40% of all the ligands detected, whereas the proportion HLA-B binders in EV ligands was substantially higher at 51% (**Figure 3C**). From JY whole-cells, the ratio of strong HLA-B binders to strong HLA-A binders is approximately 0.7:1 (**Figure 3B**, right panel, dark green), whereas this ratio was almost 1:1 in JY EVs (**Figure 3B**, left panel, dark purple). This increased presentation of HLA-B ligands on EVs could also be quantified by a higher summed MS1 intensity (52% in EVs to 36% in JY whole-cells), suggesting that HLA-B binding ligands were also more abundant in EVs (**Figure 3C**), in addition to greater sequence diversity (**Figure 3B**). This greater quantity and diversity of HLA-B ligands is likely presented on EVs by more HLA-B protein complexes that are loaded into EVs, as HLA-B protein abundance in the immunoaffinity-captured proteins (10kDa cutoff filter retentate) was also similarly increased compared to HLA-A protein complexes (**Figure 3C**, right panel). Although EVs can bud directly from the plasma membrane and contain a big portion of membrane proteins, HLA-B enrichment was not observed when measuring the proteome of plasma membranes derived from JY cells (**Figure 3D**), indicating that higher HLA-B presentation is likely an EV-specific trait-Dominant HLA-B presentation on EVs may also be visualized in an alternative man
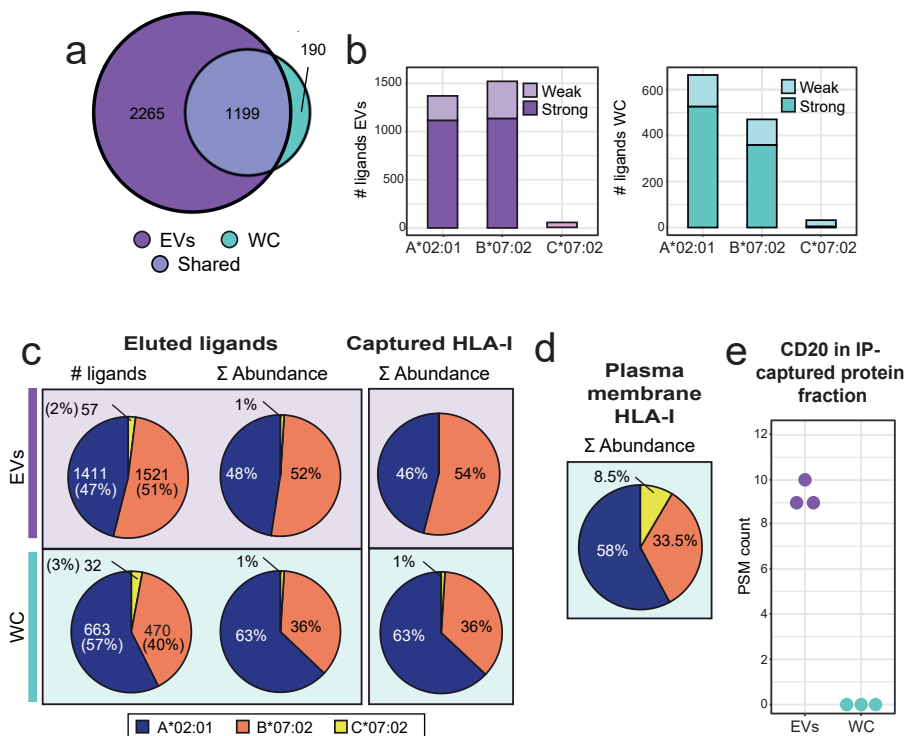
2

Figure 3. Unique characteristics of the EV HLA-I peptide ligandome. (a) HLA-I peptide species overlap. The JY whole-cell HLA-I peptide ligandome is almost completely sampled in EVs. (b) Predicted peptide binding affinity to JY HLA type. More HLA-B binding peptides were detected from EVs. Distribution of strong (≤ 50 nM affinity) and weak (≤ 500 nM affinity) peptide binders respectively. Bar plots represent the total pool of eluted ligands detected in three technical replicates from either WC (green) or EVs (purple). (c) Proportion of HLA-A (blue), HLA-B (orange), HLA-C (yellow) binders. HLA-A binders were over-represented in the JY whole-cell HLA-I peptide ligandome in both peptide species count and summed abundance, which was corroborated by the relative abundance of captured HLA-A and HLA-B proteins. A more balanced proportion of HLA-A to HLA-B was observed in EVs. (d) HLA-I complexes purified from the plasma membrane show a similar abundance distribution to what observed on whole cells. (e) Spectral count of CD20 co-immunoprecipitated with HLA-I. CD20 (MS4A1) was confidently detected exclusively in the EV HLA-I interactome.

ner. By plotting ranked intensities of all HLA peptides detected in JY EVs against the ranked intensities in whole-cells, we observed that a large number HLA-B ligands were more abundantly detected in JY EVs than whole-cells (**Supplementary Figure 3**, orange data points in grey shaded space). This observation further supports the observation that JY EVs present HLA-B complexes more dominantly than whole-cells, and imply that differential loading of HLA-I complexes into EVs may occur and be regulated by still unidentified mechanisms. In search of protein interactors that may regulate preferential EV targeting of HLA-B complexes, we compared the proteins that were co-immunoprecipitated with HLAs. Through analyses of the >10kDa retentates from both EV and whole-cell HLA pulldowns, CD20 was the only plasma membrane protein that was uniquely and prominently retrieved together with EV

HLAs (**Figure 3E**). This evidence of interaction may connect CD20 with HLA-B trafficking into the endosomal route. Indeed, a CD20 homologue has been reported to transport receptor tyrosine kinase KIT towards endocytic pathways[33], which would overlap with the route of EV biogenesis.

Cysteinylated ligands dominate the PTM-specific EV ligandome

Post-translational modifications (PTMs) on HLA class I peptide ligands have been suggested to affect HLA-I loading affinity as well as the propensity to trigger T cell activation[34,35]. Such properties are highly critical in peptide-based vaccine design and therapeutic efficacy. Particularly for peptide-loaded HLA complexes in the EV, transit through the acidic endosomal compartments, and/or exposure to complex extracellular conditions without redox protection, may already alter the prevalence of background PTMs in comparison to cell surface HLA peptide ligands. Therefore, a clear and comprehensive assessment of modification differences between whole-cell HLA ligandome and EV HLA ligandome is crucial to evaluate if HLA peptide-based vaccines delivered via EVs actually would resemble that of deactivated whole-cells. To our knowledge, such a direct comparison has not been established comprehensively in the field.

In our search of PTMs, we did not enrich for specific PTM species, as differential efficiency of the various enrichment steps could skew the relative proportion and intensity with respect to non-modified peptides. Instead, we focused on adding search modifications that have been reported to exist on HLA peptides, and allowing small modifications that are less likely to hinder HLA peptide loading and presentation. As shown in **Figure 4A**, both the whole-cell and EV ligandome contain similar proportion of peptides that have one of the six PTMs investigated, but a further breakdown in contribution of different PTMs revealed that only cysteinylation is observed much more frequently on HLA peptides of EV origin (23% vs 14% in JY whole-cells). This is unlikely to be caused by the expanded detection of HLA-B peptide binders from EVs (**Figure 3B**), since cysteinylation modifies HLA-A binders more prominently (**Figure 4A**). In contrast, the proportion of peptides with any of the other five PTMs did not change noticeably.

Intrigued by the larger proportion of cysteinylated HLA peptides in EVs, we looked further into these peptides. Compared to cysteinylated HLA peptides isolated from JY whole-cells, cysteinylated peptides from EVs appear to be slightly shorter in length (**Figure 4B**), and are mostly predicted strong binders (<50 nM predicted binding affinity). Non-cysteinylated cysteine-containing ligands, on the other hand, were rarely predicted to bind to any of HLA alleles with high affinity in both whole-cells and EVs (**Figure 4C**). Interestingly, the distribution of cysteinylation positions also differed slightly between WC and EV ligandomes. The 9-mer cores of the cysteinylated ligands were Gibbs-clustered, and the cysteinylation positions on EV ligands were in general more spread over the non-anchoring residues (**Supplementary Figure 4**). Collectively, the differences in EV HLA peptide PTMs, as well as PTM position within the peptides, seem to echo that the EV HLA peptide repertoire undergoes a slightly different processing and loading path.
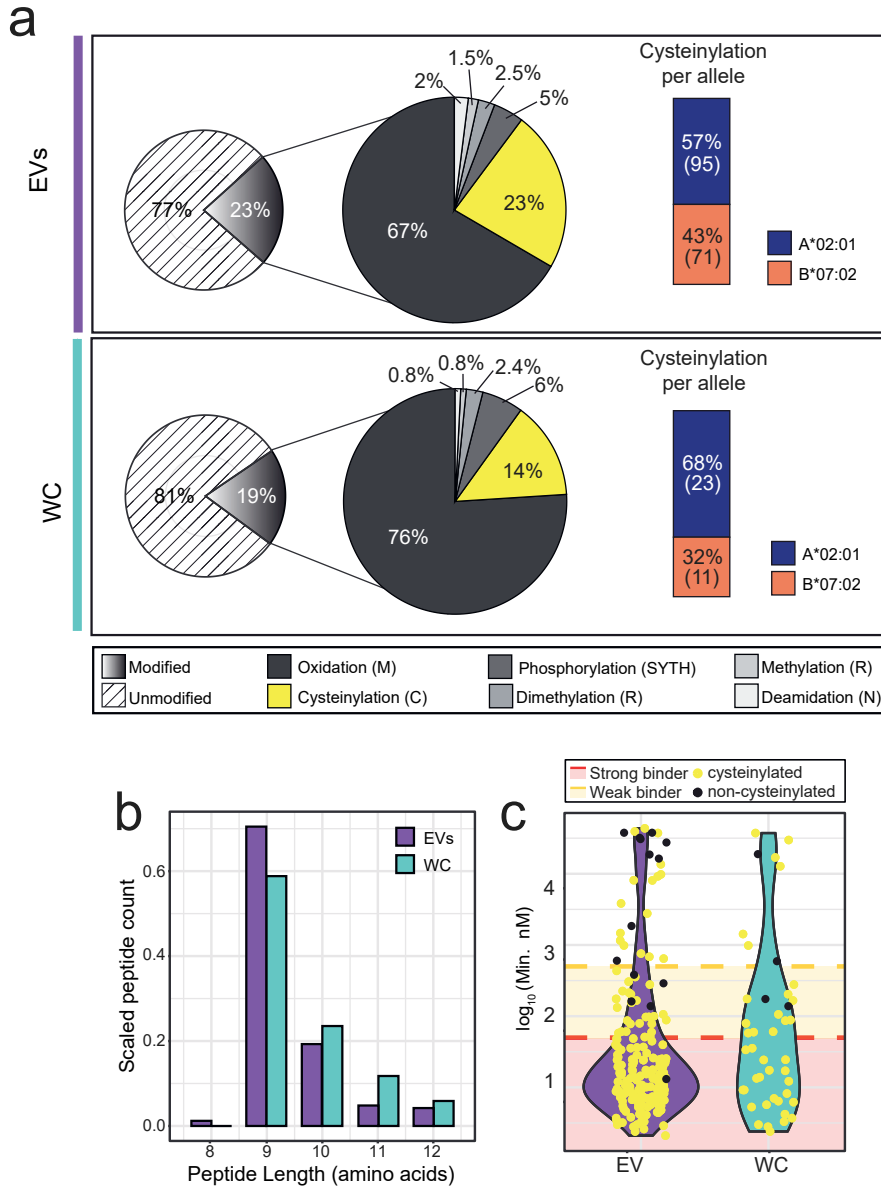
2

a

b

c

Figure 4. Post-translational modification (PTMs) analysis of EV and whole-cell HLA-I peptide ligands. (a) PTM distribution. EV HLA-I peptide ligands are more frequently cysteinylated than HLA-I peptides from whole-cells. Cysteinylation (yellow) was more abundantly detected on HLA-A binders (blue) than in HLA-B binders (orange). The number of cysteinylated ligands (n) derived from each allele is shown in parentheses. (b) Length distribution of the cysteinylated ligands. Y-axis was scaled to total number of peptides identified in three technical replicates from either WC or EVs. (c) Predicted binding affinity (nM) of all cysteine-containing peptides. Cysteinylated cysteine-containing peptides (yellow) were predicted to bind more strongly than non-cysteinylated cysteine-containing peptides.

Discussion

In recent years, EVs are emerging as promising vaccine delivery tools that may even be personalized[21,36]. To date EVs have been mostly pre-loaded artificially with therapeutic agents and RNAs[19,21,36–39], and the logical next step is to employ EVs as 'antigen delivery drones' to the human immune system. In this respect, the systemic distribution of EVs, and wide circulatory contact with other immunity components, also make EVs naturally suitable for presenting short antigen peptides. EVs are not only high-stability formulations[8,40] in the human blood, but also naturally loaded with an abundance of HLA complexes, as we demonstrate here (**Figure 2C-D**).

Compared to pre-loading EVs with a defined peptide antigen, which requires knowledge a priori, harvesting EVs from a source cell line allows the production of a complex EV repertoire with potentially broader immuno-specificities. For instance, a desirable cell line can be maintained and made to produce EVs on demand, akin to an antibody-producing hybridoma[41]. The use of cell line factories to 'manufacture' human EVs for therapy[19,42,43] has been attempted. In several phase I/II clinical trials, vaccines delivered by EVs were effective, but much milder than expected[6,44]. A possible explanation could be intrinsic peptide repertoire differences between the cellular and EV-bound ligandomes, such that giving cell-free EVs is not exactly the same as challenging the system with whole-cells. Or alternatively, post-translational modifications on exposed residues (positions 3-8) of HLA-bound peptide ligands may be modulating immunogenicity[45] or dampening immune responses[46,47]. Therefore, a comprehensive and detailed comparison of the EV-derived HLA class I ligandome, against the whole-cell counterpart, may help to rationalise the clinical differences in therapeutic efficacy, and illuminate the obstacles still impeding better application of EVs in antigen and vaccine delivery. Such an impetus motivated us to pursue this study.

Using the high-sensitivity workflow described herein, we reached an unprecedented identification of ~3,500 HLA-I peptide ligands from 120 μg EVs. This gave us ample depth to compare the HLA-I peptide ligandomes between EVs and whole-cells, both in terms of peptide sequences and post-translational modifications. In comparison to the whole-cell HLA-I peptide ligandome, we found that the EV HLA-I peptides were rather similar in length, motif and sequence properties, in agreement with by Synowsky et al.[31], albeit with a shift towards over-representation of HLA-B binders and peptides with cysteinylated-cysteines spanning positions 5 to 8 (**Supplementary Figure 3**). This higher proportion of HLA-B ligands agrees well with the higher yield of HLA-B proteins from EVs (**Figure 3C**), but the mechanism of additional HLA-B packaging into EVs remains unclear. We hypothesize this may be mediated through CD20, a specific HLA-I interactor[48], that we detected only in the EV HLA-I interactome (**Figure 3E**), but suspect that internalisation of specific HLA types along the EV biogenesis route may also be at play. Indeed, differential stability and turnover rate of different HLA allotypes have also been reported previously[49–51].

In addition to fundamental alterations in the EV HLA-I peptide repertoire due to EV HLA-I sub-composition (specifically HLA-B), the complex biogenesis pathway of EVs may also further modulate the ligand diversity. EVs are formed through multiple

routes that involve maturation in the endosomal compartments. It is thus plausible that the distinct environments, and pH, along the EV maturation path may also have an impact on the stability and exchange of HLA-I peptide ligands[52–54]. These may have contributed to the unique EV HLA-I peptide ligand species we identified here (**Figure 3A**), as well as the prevalence of modifications on redox-sensitive amino acid residues, for instance cysteine cysteinylation (**Figure 4**).

Cysteinylation results from the disulfide linkage of a free cysteine to an unpaired cysteine residue on a protein or peptide. Although free cysteines are inevitably present in the cell culture medium, cysteinylation of HLA-bound peptides has been postulated to also take place endogenously during peptide loading in the endoplasmic reticulum or in cysteine-rich organelles[46].Therefore, cysteinyl modifications on HLA peptides cannot always be easily brushed aside as artifacts. On the contrary, we still seem to observe higher cysteinylation in EV HLA-I peptide ligands when EVs are in shorter contact-time with cell culture medium in our workflow than whole-cells. In addition, we also did not detect the non-cysteinylated peptide counterpart of any cysteinylated peptides documented here. Taken together, these suggest that the modification may be favoured for loading, since 100% complete cysteinylation just from the cell culture medium seemed rather implausible. Outside the scope of this current investigation, experimental structural studies would be of great interest, to address if and how cysteinylation may affect the loading or stability of cysteine-containing peptide ligands inside the HLA-binding pocket.

With a focus on HLA peptide modifications, we quickly faced limitations in assessing or predicting the binding characteristics of post-translationally modified peptides. To date, post-translational modifications on HLA-I peptide ligands have been widely reported[45,55–59], and from this current study, about 20% of all the HLA-I peptides we detected feature modifications (**Figure 4A**). As such, the lack of computational solutions to work with these modified peptides is a perennial bottleneck still limiting our understanding of how these may impact peptide loading, half-life, stability and immunogenicity on a consensus level. On the other hand, the possibility of diverse modifications on every HLA-I peptide also further increases the complexity in the ligandome repertoire, to make systematic profiling all the more challenging.

Nonetheless, considering the major findings presented in this work, EVs indeed appear to be very promising immuno-regulatory therapeutics. EVs not only sample almost the complete whole-cell HLA-I peptide ligandome, but are also more densely loaded with antigen peptides. The source of therapeutic EVs can be renewable, and preserved in the form of a parental cell line, which also allows desired tweaks to be engineered if necessary. Indeed, specific targeting of proteins to the EVs during EV biogenesis is also already achievable by current techniques[60,61]. Here in the JY cell line, we also observed more dominant presentation of HLA-B and HLA-B peptide ligands in the EV. This may be an additional advantage, since HLA-Bs are very polymorphic[62] with diverse ligand repertoires. It is however critical to note that if the desired antigens to be delivered via EVs contain cysteines, the redox state of these

would need to be monitored closely. This is relevant, since cysteinylation has been shown to influence T-cell responses[46,47].

Methods

Cell culture and EV isolation

JY, an Epstein-Barr virus-immortalized cell line homozygous for HLA class I (HLA-A*02:01, HLA-B*07:02, and HLA-C*07:02), was obtained from ATCC, and cultured in RPMI 1640 (Lonza, Switzerland) supplemented with 10% fetal bovine serum (FBS; HyClone, USA), 10 mM L-glutamine, 50 U/mL penicillin and 50 µg/mL streptomycin (Lonza) in a humidified incubator at 37 °C with 5% CO_2. At a total count of 2e8 cells and <5% cell death, JY cells were washed thrice with PBS and plated at 5e5 cells/mL in secretion media (RPMI 1640 supplemented as previously described but containing 10% Exosome-depleted FBS (ThermoFisher, USA) instead). Conditioned media containing JY-derived EVs was collected after 24h, spun down at 300 x *g* for 10 minutes, vacuum filtered on a 0.22µm Stericup device (Millipore, USA) and kept at 4ºC. Cells were re-plated at 1e6 cells/mL and conditioned media was collected again after 24h and processed as described above. Pre-cleared media was pooled and immediately processed by ultracentrifugation at 120,000 x *g*, at 4ºC for 2h in a Sorvall T-865 rotor (ThermoFisher) to pellet EVs. The pellet was thoroughly resuspended in 10 mL PBS to wash away contaminants and purified EVs were finally pelleted again by ultracentrifugation at 120,000 x *g* at 4ºC for 2h. The EV-pellet (5%) was either resuspended in PBS for biophysical and proteome characterization (5%), or lysed separately for immunoaffinity purification of HLA-I complexes and peptide ligands (95%).

Negative stain electron microscopy (NS-TEM)

Isolated EVs were diluted 1:10 in PBS. Carbon-coated Ted Pella G400 copper grids were glow discharged, and immediately incubated with 5 µL the diluted EVs for 30s. Excess solution was blotted away, and the specimens were stained twice with 5 µL of 2% (w/v) of uranyl acetate each. NS-TEM data was collected on a Tecnai 12 transmission electron microscope (FEI, Switzerland) equipped with a SIS CCD MegaView II camera. Images were acquired with a pixel size of 6.5 Å/pixel at a defocus of -10 to -15 µm.

Nanoparticle tracking analysis (NTA)

Isolated EVs was diluted 1:1000 in PBS to a final volume of 1 ml to be analyzed in a NanoSight NS500 (Malvern Panalytica, UK), equipped with a sCMOS camera and a Blue405 laser. For the recordings, the camera level was set to 16. The sample concentrations were ~50 particles/frame, in the range of optimal operation between 30-100 particles/frame. Four videos of 1 minute were taken at 25 FPS and averaged with the built-in NanoSight Software NTA v.3.3 using a detection threshold of 5.

EV and whole-cell lysis

JY EVs and whole-cells were lysed on an end-to-end rotating platform at 4ºC in Pierce IP Lysis Buffer (ThermoFisher) supplemented with 1x cOmplete protease

inhibitors (Sigma-Aldrich,USA), 50 µg/mL DNase I (Sigma-Aldrich), and 50 µg/mL RNase A (Sigma-Aldrich). EVs were lysed in 200 µL of lysis buffer for 30 min followed by 20 cycles of sonication at 4ºC (30 seconds on, 30 seconds off) in a Bioruptor (Diagenode, Belgium). Cells were lysed in 10 mL of lysis buffer for 1h. The lysates were spun down at 20,000 g at 4ºC for 1h and the aqueous phase containing the solubilized proteins was kept. Lysate protein concentrations were determined with a microplate BCA-protein assay (ThermoFisher). Aliquots of 15 µg were taken for further characterization and all the samples were snap frozen until further processing.

SDS-page and western blot

EV and WC lysates, and 10kDa filtration retentates were resolved on 12% Bis-Tris Criterion XT precast gels (Biorad, USA) with 1x XT-MOPS buffer at fixed voltage of 150 V for about 2h. Proteins were stained in-gel using Imperial Protein Stain (ThermoFisher). For Western detection, proteins were transferred to a PVDF membrane in Towbin buffer (0.025M Tris, 0.192M glycine, 20% methanol) at 100 V and 4°C for 1h. Membranes were washed 3x with TBS buffer containing 1% Tween-20 (TBST) and then blocked for 1h in TBST supplemented with 5% Blotting-Grade Blocker (Biorad). Primary antibodies (W6/32 at 1:5,000 or α-CD81 at 1:200) were incubated at 4°C overnight in TBST supplemented with 1% milk. Secondary antibody incubations were done using HRP conjugated α-mouse IgG antibody (1:2,000 dilution) for 2h at room temperature (RT) in the same buffer. Between and after antibody incubations, membranes were washed 3x 10 min in TBST. HRP signal was visualized using SuperSignal West Dura (ThermoFisher) substrate, on an Amersham Imager 600 (GE healthcare, USA).

Immuno-affinity purification of HLA complexes and peptide ligands

To capture class I HLA complexes from EV and whole-cell lysates, 5 ug of W6/32 pan-HLA class I antibody was coupled and cross-linked to 5 µL protein A/G beads and incubated with 120 µg of lysate overnight at 4°C in an end-to-end rotating platform. The beads were washed 4 times with 100 µL of cold PBS and HLA complexes were eluted with cold 10% acetic acid. We have tested that W6/32 is specific for human HLA molecules, and does not cross-react with Bovine MHC (BoLa), which may still be present in trace amounts in EV-depleted FBS (**Supplementary Figure 5**).To maximize the retrieval of HLA complexes, lysates were re-loaded twice into the beads and the same procedure was repeated. Elutions were pooled and peptide ligands were separated from HLA complexes by ultrafiltration at 12,000 g in 10 kDa molecular weight cutoff filters (Millipore). The flow-through, containing the peptide ligands, was desalted using Pierce C18 10 µL bed Stage tips (ThermoFisher), vacuum dried and stored at -20°C before LC-MS analysis. Proteins in the retentate fraction were recovered by solubilizing in 8M Urea / 50mM ammonium bicarbonate.

Plasma membrane fractionation

JY Plasma membranes were enriched using a membrane fractionation kit (ab 65400, Abcam, Cambridge, UK) as described previously[63]. Washed cell pellets were gently resuspended in ice-cold isotonic and detergent-free homogenization buffer supple-

mented with 1 × cOmplete protease inhibitors, 50 µg/mL DNase I, and 50 µg/mL RNase A. Cells were disrupted manually at 4°C with a hand-held glass homogenizer. Any remaining intact cells were pelleted at 700 g, for 10 min and discarded. Total membranes were pelleted from the supernatant at 10,000 g, for 30 min, and further partitioned in partially miscible gradients established by mixing 200 µl each of the upper and lower phase solutions (Abcam ab 65400). The upper phase, containing plasma membrane proteins, was collected and precipitated for 30 min at 4°C by 5 times dilution with ultra-pure water. Plasma membrane proteins were then pelleted at 20,000 g for 30 min, resuspended in 8 M Urea, 50 mM Ammonium bicarbonate, 0.2% Sodium deoxycholate, and then digested as described hereafter.

Protein digestion

To perform proteomics characterization of WC and EV lysates, 5 µg of total proteins from EVs and whole-cells were digested as follows. Each sample was diluted to a final 4M Urea / 50 mM ammonium bicarbonate and proteins were reduced in 4mM dithiothreitol (DTT) for 1h, and then cysteine-alkylated in 8mM iodoacetamide for 30 min in the dark. The alkylation reaction was quenched by further addition of 4 mM DTT, and proteins were pre-digested by Lys-C (1:50 enzyme-protein ratio) at 37°C for 4h. Samples were diluted further to 1M Urea / 50 mM ammonium bicarbonate and digested overnight at 37°C by Trypsin (Promega, USA) at 1:50. Peptides were acidified to 5% Formic acid and were desalted using Pierce C18 10 µL bed Stage tips (ThermoFisher). Peptides were vacuum-dried and stored at -20°C before LC-MS analysis. The same procedure was applied for digestion of 10kDa retentates.

LC-MS/MS

All samples were analyzed in an UHPLC 1290 system (Agilent, USA) coupled to an Orbitrap HF-X mass spectrometer (Thermo Scientific, USA). Peptides were trapped (Dr. Maisch Reprosil-Pur C18-AQ, 3µm, 2.5 cm x 100 µm) for 5 minutes in solvent A (0.1% Formic acid in water) and then separated on an analytical column (Agilent Poroshell, 120 EC-C18, 2.7 µm, 50 cm x 75 µm) using a linear gradient of solvent B (0.1% Formic acid in 80% acetonitrile). For HLA peptide ligand analyses, a gradient of 7 - 40 % B in 90 min was used. For lysate proteomics, a gradient of 13 - 44 % B in 95 min. Finally, 10kDa retentate analysis, a gradient of 13 - 40% B in 65 min was used.

The mass spectrometer was operated in data-dependent mode, at a resolution of 60,000 for MS1 and 30,000 for MS2. HLA peptide data was acquired at 400-650 m/z and precursor ions were accumulated for 50 ms or until a AGC target value of 3e6 was reached. The 15 most abundant doubly and triply charged precursors were selected for fragmentation, after accumulation to the AGC target value of 1e5 within 250 ms. HCD fragmentation was performed at 27% NCE. Dynamic exclusion time was set to 16s. Proteomics data was acquired at 375-1600 m/z, with precursor accumulation of up to 3e6 AGC target value within 20 ms. The 15 most abundant ions (with charge states 2 to 5) were selected for HCD fragmentation, after accumulation to the AGC target value of 1e5 within 50ms. Dynamic exclusion times were 16s and 12s, for 95 min and 65 min gradients respectively. From all paired samples used in

this study (WC and EV ligandomes and proteomes), three technical replicates were measured by LC-MS/MS.

Database search

For HLA peptide analyses, raw data were searched in Proteome Discoverer (v_2.3, Thermo Scientific) using Sequest HT against the SwissProt human database (down-loaded on 09/2019, containing 20,431 protein sequences), curated to match JY cells HLA alleles (HLA-A*02:01, HLA-B*07:02 and HLA-C*07:02) and appended with common FBS contaminants. Unspecific searches were performed for precursors with mass between 797 and 1950 Da, corresponding to 8 – 12 amino acid peptides. Precursor ion and fragment ion tolerance were set to 10 ppm and 0.02 Da respectively, and methionine oxidation and cysteine cysteinylation were set as variable modifications for the main search. For modification-specific analyses, searches for arginine (di-)methylation, phosphorylation (S/T/Y/H) and asparagine deamidation were performed separately always including methionine oxidation. Identified peptides were filtered to 5% false discovery rate (FDR) using percolator algorithm[64].

For proteomics analysis, raw data were searched in MaxQuant (v_1.6.10.4.3)[65] against the same database using the Andromeda search engine. In this case, trypsin was set as the digestion enzyme and up to 2 missed cleavages were allowed. Carbamidomethylation of cysteines was set as a fixed modification while protein N-terminal acetylation and methionine oxidation were set as variable modifications. Label-free quantification (LFQ) was enabled using a minimum ratio count of 2 and both razor and unique peptides for quantification. Precursor ion tolerance was set to 20 ppm for the first search and 4.5 ppm after recalibration, and fragment ions tolerance was set to 20 ppm. FDR of 1% was set at both PSM and protein level by using a reverse decoy database strategy.

Data analysis

Immunopeptidomics data were analyzed using Excel and in-house built R scripts[66], and refined with Illustrator 2020 (Adobe, USA). Binding affinity (BA) predictions were performed with NetMHC 4.0[67] against the JY HLA type (HLA-A*02:01, HLA-B*07:02 and HLA-C*07:02). Predicted BAs were further curated manually to achieve a stringing binder cut-off. Peptide ligands with BA ≤ 50mM, 50 < BA ≤ 500 or BA > 500 to HLA proteins were considered strong, weak or non-binders respectively. For cysteinylation binding affinity predictions, cysteine residues of cysteinylated peptides were substituted to 'X' since algorithms cannot currently predict modified amino acids. Sequence logos were produced with Seq2Logo using the (P- weighted) Kullback-Leibler method[68] and only the NetMHC4.0 calculated 9-mer cores were used for these plots. For quantitative analysis of the ligandomes, data was processed using the Prostar software[69] and normalization of EV and WC ligandomes was achieved by mean-centering approach. Subsequently, ranks were calculated from peptides shared between EV and WC ligandomes using normalized abundance values.

All proteomics data were analyzed using Perseus software (v_1.6.14)[70]. Proteins quantified (LFQ) in 2 out of 3 replicates in EV or WC lysates or in the >10 kDa captured fractions were *log$_2$* transformed and missing values were replaced individually

2

for each sample from the normal distribution. Statistical differences were assessed by a two-sided Student's T-test and corrected p-values (q-value) were calculated using the permutation method with up to 250 iterations. Proteins were considered significant when q-value ≤ 0.05 and fold change ≥ 2. All plots were generated using R packages and refined with Illustrator 2020.

Statistics and reproducibility

All proteomics measurements were performed in technical triplicates, as indicated in the respective figure legends. For comparisons of HLA peptide proportion, identifications from three injection replicates were summed in a non-redundant manner per HLA allele. All experiments have been replicated and are reproducible.

Data Availability statement

Proteomics and immunopeptidomics raw data have been deposited to ProteomeXchange Consortium via the PRIDE repository[71] and can be accessed through the identifier PXD021177. Source data are provided with this paper. All Uncropped and unedited blot and gel images for Figures 1C-E, Figure 2C, Supplementary Figure 5 are provided in the Supplementary Figures 6-10. All data for visualisation are provided in Supplementary Data 1.

Acknowledgements

Competing interests

The authors declare no competing interests.

Author contributions

JBM and WW conceived the study and designed the approach. JBM performed all experiments, analysed and visualized the data, with supervision from WW. JBM and WW co-wrote the manuscript. AJRH and WW edited the manuscript.

References

1. Hargadon, K. M., Johnson, C. E. & Williams, C. J. Immune checkpoint blockade therapy for cancer: An overview of FDA-approved immune checkpoint inhibitors. Int. Immunopharmacol. 62, 29–39 (2018).

2. Carreno, B. M. et al. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. Science. 348, 803–808 (2015).

3. Ott, P. A. et al. An immunogenic personal neoantigen vaccine for patients with melanoma. Nature 547, 217–221 (2017).

4. Sahin, U. et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. Nature 547, 222–226 (2017).

5. Suzuki, H. et al. Multiple therapeutic peptide vaccines consisting of combined novel cancer testis antigens and anti-angiogenic peptides for patients with non-small cell lung cancer. J. Transl. Med. 11, 97 (2013).

6. Morse, M. A. et al. A phase I study of dexosome immunotherapy in patients with advanced non-small cell lung cancer. J. Transl. Med. 3, 9 (2005).

7. Keskin, D. B. et al. Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. Nature 565, 234–239 (2019).

8. Lener, T. et al. Applying extracellular vesicles based therapeutics in clinical trials-an ISEV position paper. J. Extracell. Vesicles 4, 30087 (2015).

9. Hu, Z., Ott, P. A. & Wu, C. J. Towards personalized, tumour-specific, therapeutic vaccines for cancer. Nat. Rev. Immunol. 18, 168–182 (2018).

10. Hollingsworth, R. E. & Jansen, K. Turning the corner on therapeutic cancer vaccines. npj Vaccines 4, 7 (2019).

11. Markov, O., Oshchepkova, A. & Mironova, N. Immunotherapy based on dendritic cell-targeted/-derived extracellular vesicles — A novel strategy for enhancement of the anti-tumor immune response. Front. Pharmacol. 10, 1152 (2019).

12. Zeelenberg, I. S. et al. Targeting Tumor Antigens to Secreted Membrane Vesicles In vivo Induces Efficient Antitumor Immune Responses. Cancer Res 68, 1228–1263 (2008).

13. André, F. et al. Exosomes as Potent Cell-Free Peptide-Based Vaccine. I. Dendritic Cell-Derived Exosomes Transfer Functional MHC Class I/Peptide Complexes to Dendritic Cells. J. Immunol. 172, 2126–2136 (2004).

14. Wang, L. et al. Exosomal pMHC-I complex targets T cell-based vaccine to directly stimulate CTL responses leading to antitumor immunity in transgenic FVB-neuN and HLA-A2/HER2 mice and eradicating trastuzumab-resistant tumor in athymic nude mice. Breast Cancer Res. Treat. 140, 273–284 (2013).

15. Wahlund, C. J. E. et al. Exosomes from antigen-pulsed dendritic cells induce stronger antigen-specific immune responses than microvesicles in vivo. Sci. Rep. 7, 1–9 (2017).

16. Wolfers, J. et al. Tumor-derived exosomes are a source of shared tumor rejection antigens for CTL cross-priming. Nat. Med. 7, 297–303 (2001).

17. Zeng, F. & Morelli, A. E. Extracellular vesicle-mediated MHC cross-dressing in immune homeostasis, transplantation, infectious diseases, and cancer. Semin. Immunopathol. 40, 477–490 (2018).

18. Zhang, D. et al. Exosome-Mediated Small RNA Delivery: A Novel Therapeutic Approach for Inflammatory Lung Responses. Mol. Ther. 26, 2119–2130 (2018).

19. Alvarez-Erviti, L. et al. Delivery of siRNA to the mouse brain by systemic injection of targeted exosomes. Nat. Biotechnol. 29, 341–345 (2011).

20. Jiang, L., Vader, P. & Schiffelers, R. M. Extracellular vesicles for nucleic acid delivery: progress and prospects for safe RNA-based gene therapy. Gene Ther. 24, 157–166 (2017).

21. O'Brien, K., Breyne, K., Ughetto, S., Laurent, L. C. & Breakefield, X. O. RNA delivery by extracellular vesicles in mammalian cells and its applications. Nat. Rev. Mol. Cell Biol. 21, 585–606 (2020).

22. Rontogianni, S. et al. Proteomic profiling of extracellular vesicles allows for human breast cancer subtyping. Commun. Biol. 2, 325 (2019).

23. Mommen, G. P. M. et al. Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). Proc. Natl. Acad. Sci. U. S. A. 111, 4507–4512 (2014).

24. Demmers, L. C., Heck, A. J. R. & Wu, W. Pre-fractionation Extends but also Creates a Bias in the Detectable HLA Class I Ligandome. J. Proteome Res. 18, 1634–1643 (2019).

25. Purcell, A. W., Ramarathinam, S. H. & Ternette, N. Mass spectrometry–based identification of MHC-bound peptides for immunopeptidomics. Nat. Protoc. 14, 1687–1707 (2019).

26. Ternette, N. et al. Immunopeptidomic Profiling of HLA-A2-Positive Triple Negative Breast Cancer Identifies Potential Immunotherapy Target Antigens. Proteomics 18, e1700465 (2018).

27. Chong, C. et al. High-throughput and Sensitive Immunopeptidomics Platform Reveals Profound Interferon-Mediated Remodeling of the Human Leukocyte Antigen (HLA) Ligandome. Mol. Cell. Proteomics 17, 533–548 (2018).

28. Schuster, H. et al. Data Descriptor: A tissue-based draft map of the murine MHC class I immunopeptidome. Sci. Data 5, 1–11 (2018).

29. Marcu, A. et al. HLA Ligand Atlas: a benign reference of HLA-presented peptides to improve T-cell-based cancer immunotherapy. J Immunother Cancer 9, 2071 (2021).

30. Demmers, L. C. et al. Single-cell derived tumor organoids display diversity in HLA class I peptide presentation. Nat. Commun. 11, 5338 (2020).

31. Synowsky, S. A. et al. The major histocompatibility complex class I immunopeptidome of extracellular vesicles. J. Biol. Chem. 292, 17084–17092 (2017).

32. Hong, C. S., Muller, L., Boyiadzis, M. & Whiteside, T. L. Isolation and characterization of CD34+ blast-derived exosomes in acute myeloid leukemia. PLoS One 9, e103310 (2014).

33. Cruse, G. et al. The CD20 homologue MS4A4 directs trafficking of KIT toward clathrin-independent endocytosis pathways and thus regulates receptor signaling and recycling. Mol. Biol. Cell 26, 1711–1727 (2015).

34. Engelhard, V. H., Altrich-Vanlith, M., Ostankovitch, M. & Zarling, A. L. Post-translational modifications of naturally processed MHC-binding epitopes. Curr. Opin. Immunol. 18, 92–97 (2006).

35. Petersen, J., Purcell, A. W. & Rossjohn, J. Post-translationally modified T cell epitopes: Immune recognition and immunotherapy. J. Mol. Med. 87, 1045–1051 (2009).

36. Qiao, L. et al. Tumor cell-derived exosomes home to their cells of origin and can be used as Trojan horses to deliver cancer drugs. Theranostics 10, 3474–3487 (2020).

37. Vader, P., Mol, E. A., Pasterkamp, G. & Schiffelers, R. M. Extracellular vesicles for drug delivery. Adv. Drug Deliv. Rev. 106, 148–156 (2016).

38. Elsharkasy, O. M. et al. Extracellular vesicles as drug delivery systems: Why and how? Adv. Drug Deliv. Rev. 106, 148–156 (2020).

39. B Wiklander, O. P., Á Brennan, M., Lötvall, J., Breakefield, X. O. & Andaloussi, S. EL. Advances in therapeutic applications of extracellular vesicles. Sci. Transl. Med 11, eaav8521 (2019).

40. Jeyaram, A. & Jay, S. M. Preservation and Storage Stability of Extracellular Vesicles for Therapeutic Applications. AAPS J. 20, 1–7 (2018).

41. Kojima, R. et al. Designer exosomes produced by implanted cells intracerebrally deliver therapeutic cargo for Parkinson's disease treatment. Nat. Commun. 9, 1305 (2018).

42. Chen, T. S. et al. Enabling a robust scalable manufacturing process for therapeutic exosomes through oncogenic immortalization of human ESC-derived MSCs. J. Transl. Med. 9, 47 (2011).

43. Gimona, M., Pachler, K., Laner-Plam-

berger, S., Schallmoser, K. & Rohde, E. Manufacturing of Human Extracellular Vesicle-Based Therapeutics for Clinical Use. Int. J. Mol. Sci. 18, 1190 (2017).

44. Besse, B. et al. Dendritic cell-derived exosomes as maintenance immunotherapy after first line chemotherapy in NSCLC. Oncoimmunology 5, 1071008 (2016).

45. Marino, F. et al. Arginine (Di)methylated Human Leukocyte Antigen Class i Peptides Are Favorably Presented by HLA-B□07. J. Proteome Res. 16, 34–44 (2017).

46. Meadows, L. et al. The HLA-A*0201-restricted H-Y antigen contains a posttranslationally modified cysteine that significantly affects T cell recognition. Immunity 6, 273–281 (1997).

47. Chen, W., Yewdell, J. W., Levine, R. L. & Bennink, J. R. Modification of cysteine residues in vitro and in vivo affects the immunogenicity and antigenicity of major histocompatibility complex class I- restricted viral determinants. J. Exp. Med. 189, 1757–1764 (1999).

48. Szöllósi, J., Horejsí, V., Bene, L., Angelisová, P. & Damjanovich, S. Supramolecular complexes of MHC class I, MHC class II, CD20, and tetraspan molecules (CD53, CD81, and CD82) at the surface of a B cell line JY. J. Immunol. 157, 7 (1996).

49. Neefjes, J. J. & Ploegh, H. L. Allele and locus□specific differences in cell surface expression and the association of HLA class I heavy chain with β2□microglobulin: differential effects of inhibition of glycosylation on class I subunit association. Eur. J. Immunol. 18, 801–810 (1988).

50. Yarzabek, B. et al. Variations in HLA-B cell surface expression, half-life and extracellular antigen receptivity. 7, e34961 (2018).

51. B. Sean Carey, Kay Victoria Poulton, A. P. Factors affecting HLA expression: A review. Int. J. Immunogenet. 46, 307–320 (2019).

52. Grommé, M. et al. Recycling MHC class I molecules and endosomal peptide loading. Proc. Natl. Acad. Sci. U. S. A. 96, 10326–10331 (1999).

53. Lizée, G., Basha, G. & Jefferies, W. A. Tails of wonder: Endocytic-sorting motifs key for exogenous antigen presentation. Trends Immunol. 26, 141–149 (2005).

54. Montealegre, S. & Van Endert, P. M. Endocytic recycling of MHC class I molecules in non-professional antigen presenting and dendritic cells. Front. Immunol. 10, 3098 (2019).

55. Zarling, A. L. et al. Phosphorylated peptides are naturally processed and presented by major histocompatibility complex class I molecules in vivo. J. Exp. Med. 192, 1755-1762 (2000).

56. Cobbold, M. et al. MHC class I-associated phosphopeptides are the targets of memory-like immunity in leukemia. Sci. Transl. Med. 5, 203ra125 (2013).

57. Marino, F. et al. Extended O-GlcNAc on HLA Class-I-Bound Peptides. J. Am. Chem. Soc. 137, 10922–10925 (2015).

58. Alpízar, A. et al. A molecular basis for the presentation of phosphorylated peptides by HLA-B antigens. Mol. Cell. Proteomics 16, 181–193 (2017).

59. Mei, S. et al. Immunopeptidomic analysis reveals that deamidated HLA-bound peptides arise predominantly from deglycosylated precursors. Mol. Cell. Proteomics 19, 1236-1247 (2020).

60. Anand, S., Samuel, M., Kumar, S. & Mathivanan, S. Ticket to a bubble ride: Cargo sorting into exosomes and extracellular vesicles. Biochim. Biophys. Acta - Proteins Proteomics 1867, 140203 (2019).

61. Carnino, J. M., Ni, K. & Jin, Y. Post-translational Modification Regulates Formation and Cargo-Loading of Extracellular Vesicles. Front. Immunol. 11, 948 (2020).

62. Raghavan, M. & Geng, J. HLA-B polymorphisms and intracellular assembly modes. Mol. Immunol. 68, 89–93 (2015).

63. Mezzadra, R. et al. Identification of CMTM6 and CMTM4 as PD-L1 protein regulators. Nature 549, 106–110 (2017).

64. The, M., MacCoss, M. J., Noble, W. S. & Käll, L. Fast and Accurate Protein False Discovery Rates on Large-Scale Proteomics Data Sets with Percolator 3.0. J.

Am. Soc. Mass Spectrom. 27, 1719–1727 (2016).

65. Tyanova, S., Temu, T. & Cox, J. The Max-Quant computational platform for mass spectrometry-based shotgun proteomics. Nat. Protoc. 11, 2301–2319 (2016).

66. R Development Core Team, R. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing (2011). doi:10.1007/978-3-540-74686-7.

67. Andreatta, M. & Nielsen, M. Gapped sequence alignment using artificial neural networks: Application to the MHC class i system. Bioinformatics 32, 511–517 (2016).

68. Thomsen, M. C. F. & Nielsen, M. Seq2Logo: A method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. Nucleic Acids Res. 40, W281–W287 (2012).

69. Wieczorek, S. et al. DAPAR & ProStaR: Software to perform statistical analyses in quantitative discovery proteomics. Bioinformatics 33, 135–136 (2017).

70. Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat. Methods 13, 731–740 (2016).

71. Vizcaíno, J. A. et al. 2016 update of the PRIDE database and its related tools. Nucleic Acids Res. 44, D447–D456 (2016).

2

CHAPTER 2 - SI

2

Supporting information for

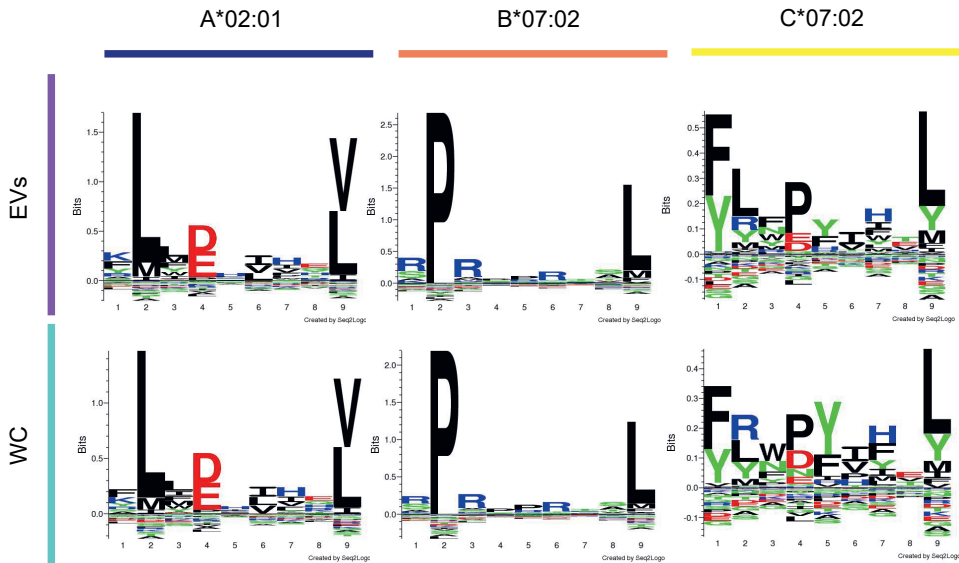# HLA-B and cysteinylated ligands distinguish the antigen presentation landscape of extracellular vesicles

Julia Bauzá-Martinez[1,2], Albert J.R. Heck[1,2], Wei Wu[1,2]*

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

[2] Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands

Based on:

Supplementary Figure 1. Allele-specific sequence motifs of EV- and WC-eluted HLA-I peptide ligands. The predicted 9-mer cores (from NetMHC 4.0) for HLA-A*02:01 (blue), B-07:02 (orange) and C*07:02 (yellow) eluted ligands are visualised by Gibbs clustering, and the empirical HLA-I peptide sequence motifs are displayed. HLA-I peptides eluted from WC (whole-cells, green) and EVs (purple) were similar in sequence motifs.

2
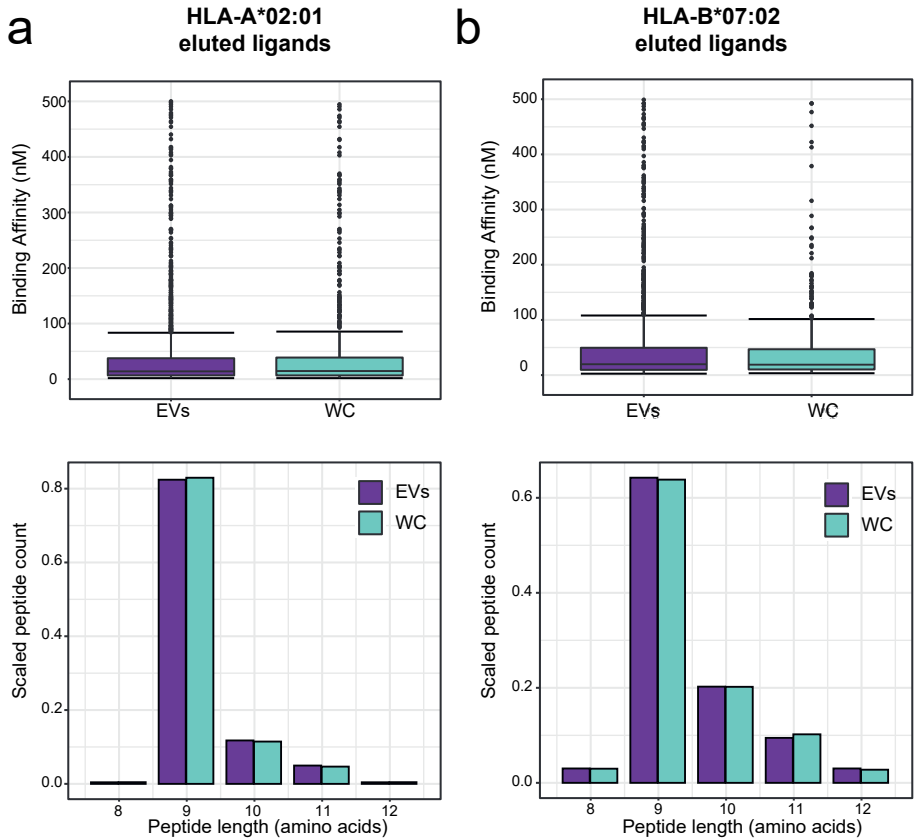
Supplementary Figure 2. Predicted binding affinity and peptide length distribution per HLA type. (a) Predicted binding affinity and length distribution of HLA-A*02:01 ligands eluted from EVs or whole-cells. (b) Predicted binding affinity and length distribution of HLA-B*07:02 ligands eluted from EVs (purple) or whole-cells (WC, green). Box plots represent peptides where the 25%, 50% (median) and 75% quantiles are represented in each box, and the wishkers represent the ±1.5*IQR from the closest quantile.
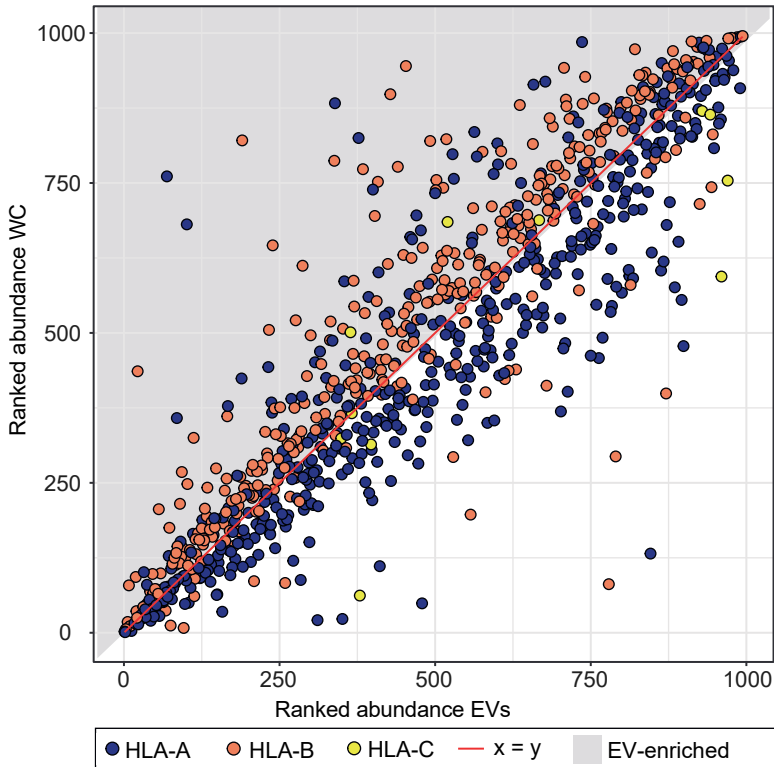
Supplementary Figure 3. Ranked abundance of HLA-I peptide ligands detected in both WC and EV ligandomes. Abundance of shared ligands between EVs and whole-cells (WC) were ranked in decreasing order (highest abundance as rank 1). Red line indicates no difference in rank. Grey shaded area marks ligands with abundance-rank higher in EVs than in WC. Over-representation of HLA-B binders (orange dots) in the grey space indicates many HLA-B binders were more abundantly detected in EVs. HLA-A derived ligands are represented by blue dots, and HLA-C ligands are represented by yellow dots.
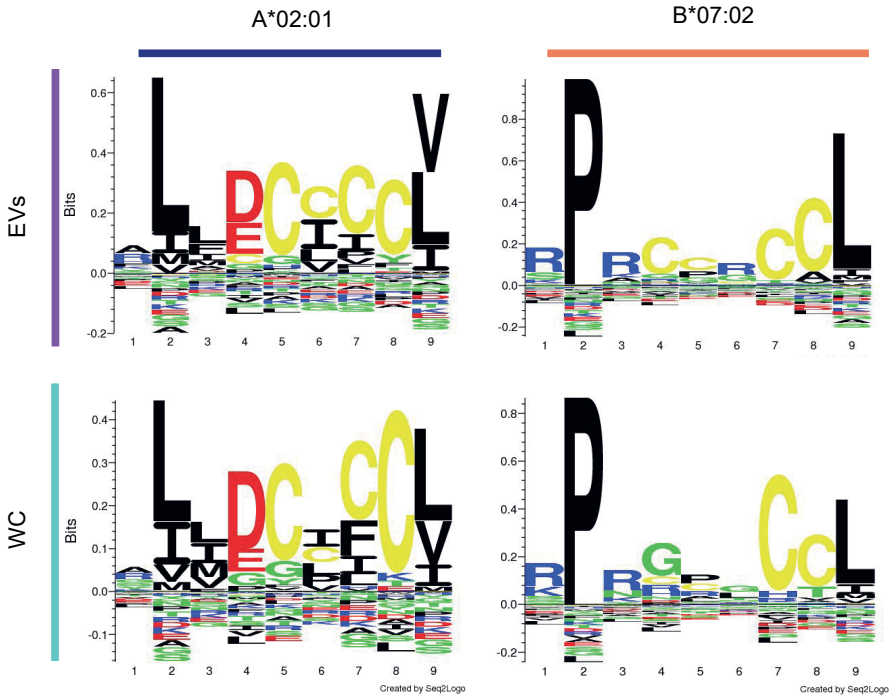
Supplementary Figure 4. Cysteinylation positions on HLA-I peptide ligands. Cysteinylation positions were localised in the predicted 9-mer cores (from NetMHC 4.0) and visualised by Gibbs clustering. All peptides detected contain only 1 cysteine residue and the plot represents the frequency of localization of the cysteinylated cysteine. Clustering shows that cysteinylation occurs most frequently at the C-terminal part of the cysteine-containing peptides in positions 5 to 8 of the ligands.

2
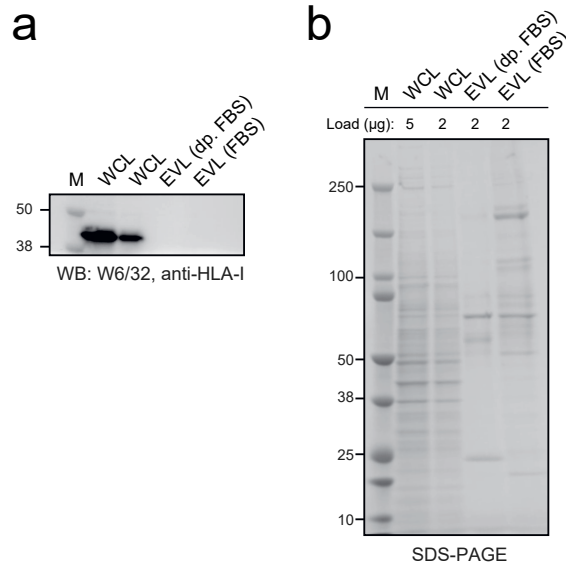
Supplementary Figure 5. W6/32 antibody does not cross-react to bovine MHC (BoLA). (a) Western blot detection performed on 5 and 2 micrograms of JY whole-cell lysates (WCL), and 2 micrograms of EV lysates (EVL), isolated from either EV-depleted fetal bovine serum (dp. FBS) or regular FBS. W6/32 antibody binds specifically to human HLA, but not bovine MHC (BoLA). To obtain bovine EVs, 20 mL of either the depleted FBS used in this study or normal FBS were diluted to 20% in raw RPMI-1640 and EVs were pelleted by ultracentrifugation as described in Materials and Methods. Full image provided in Supplementary Figure 10. **(b)** SDS-PAGE showing the protein profiles of all samples.
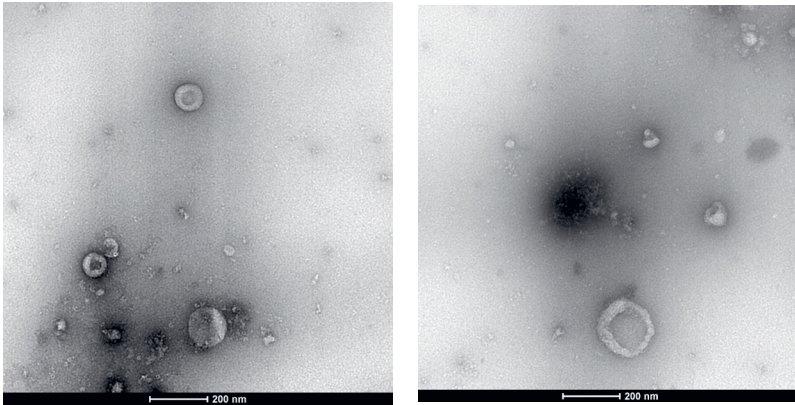
Supplementary Figure 6. Raw negative stain transmission electron microscopy (NS-TEM) scans. NS-TEM analysis of the isolated extracellular vesicles (EVs) shown in Figure 1c. Morphology of the isolated EVs were cup-shaped. Scale bars: 200 nm.
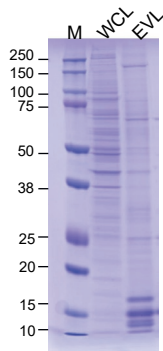
Supplementary Figure 7. Unedited SDS-PAGE image of whole-cell (WCL) and extracellular vesicle lysate (EVL) material. SDS-PAGE analysis shown Figure 1d. Dominant protein bands in WCL and EVL were mutually exclusive. M: molecular weight marker.
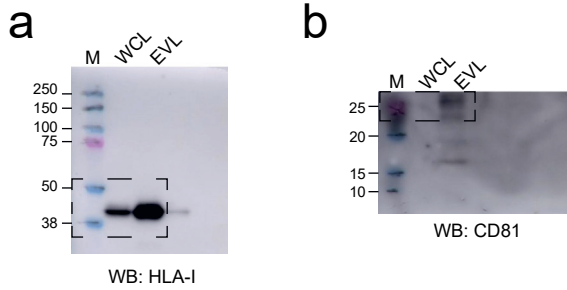
Supplementary Figure 8. Uncropped western blot images for HLA-I and CD81 detection from whole-cell (WCL) and extracellular vesicle lysate (EVL) material. (a) The upper part of the membrane was blotted against HLA-I and **(b)** the lower part of the membrane was blotted against CD81. HLA-I and CD81 western blot. From 5 µg load of EVL and WCL, prominent enrichment of HLA-I and CD81 were observed in EVL.The areas shown in Figure 1e are highlighted with a dotted line. M: molecular weight marker.
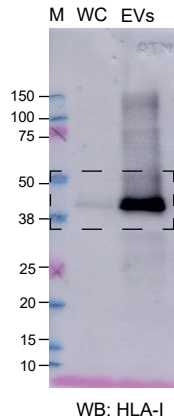
Supplementary Figure 9. Uncropped western blot images for HLA-I detection from the captured protein fraction of whole-cells (WC) or extracellular vesicles (EVs) IP eluates. The membrane was blotted against HLA-I. HLA-I was detected in the eluate of WCL HLA-I IP, and more strongly in the eluate of EVL HLA-I IP.The area shown in Figure 2c is highlighted with a dotted line. M: molecular weight marker.
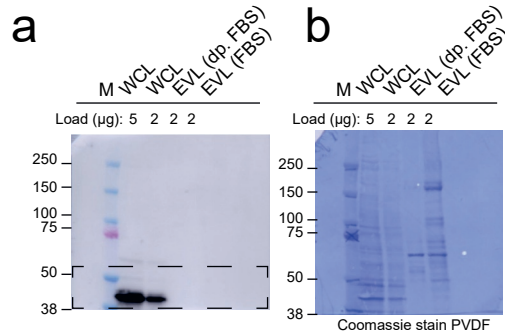
Supplementary Figure 10. Uncropped western blot images for HLA-I detection from human and bovine samples. (a) Western blot detection performed on upper membrane for 5 and 2 micrograms of JY whole-cell lysates (WCL), and 2 micrograms of EV lysates (EVL), isolated from either EV-depleted fetal bovine serum (dp. FBS) or regular FBS. W6/32 antibody binds specifically to human HLA, but not bovine MHC (BoLA). The area shown in Supplementary Figure 5a is highlighted with a dotted line. **(b)** Coomassie stain of the PVDF membrane. M: molecular weight marker.

CHAPTER 3

Characterization of protein complexes in extracellular vesicles by intact extracellular vesicle crosslinking mass spectrometry (iEVXL)

Julia Bauzá-Martinez[1,2], Gad Armony[1,2], Matti F. Pronker[1,2], Wei Wu[1,2,3,4]*

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

[2] Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands

[3] Singapore Immunology Network (SIgN), Agency for Science, Technology and Research (A*STAR), 8A Biomedical Grove, Immunos, #04-06, Singapore 138648, Singapore.

[4] Department of Pharmacy, National University of Singapore, 14 Science Drive 4, Singapore 117543, Singapore.

Based on:

Summary

Extracellular vesicles (EVs) are blood-borne messengers that coordinate signaling between different tissues and organs in the body. The specificity of such crosstalk is determined by preferential EV docking to target sites, as mediated through protein-protein interactions. As such, the need to structurally characterize the EV surface precedes further understanding of docking selectivity and recipient-cell uptake mechanisms. Here, we describe an intact extracellular vesicle crosslinking mass spectrometry (iEVXL) method that can be applied for structural characterization of protein complexes in EVs. By using a partially membrane-permeable disuccinimidyl suberate crosslinker, proteins on the EV outer-surface and inside EVs can be immobilized together with their interacting partners. This not only provides covalent stabilization of protein complexes before extraction from the membrane-enclosed environment, but also generates a set of crosslinking distance restraints that can be used for structural modeling and comparative screening of changes in EV protein assemblies. Here we demonstrate iEVXL as a powerful approach to reveal high-resolution information, about protein determinants that govern EV docking and signaling, and as a crucial aid in modeling docking interactions.

3

Introduction

Extracellular vesicles (EVs) are small heterogenous vesicles that are secreted by most cells in the body. As important mediators of intercellular communication, EVs can cross biological barriers and have consequently been detected in many bio-fluids[1,2]. Most notably, EVs in patient blood have been very informative as disease diagnostics[3,4], prognostics and means of treatment stratification[5–7]. EVs can carry specific biomolecular cargoes ranging from nucleic acids and lipids to proteins and glycoconjugates [8–11]. Quite recently, even fully loaded human leukocyte antigen molecules with presented antigen peptides[12–14] have been documented from EVs, further alluding to the immunological influence EVs might have in the body. Due to possible tissue-homing properties[15–19], EVs have also been viewed and tested as drug delivery vectors, although the mechanisms governing tissue-specific EV docking still remain poorly understood.

Through understanding the complex biogenesis of EVs[20], it appears inevitable that EVs would co-package cellular bits from their cells of origin that recapitulate their signaling state[9]. Intriguingly, the uptake of activated proteins packaged in EVs by recipient cells could also transfer and propagate oncogenic signaling in the recipient cells[21], hence bypassing the need for prior ligand activation. Such phenomena hold strong significance and implications in the treatment of malignant disease and cancer metastasis, but also fundamentally explains another mode of inter-cellular crosstalk.

The protein repertoire packaged in EVs and on the surface of EVs is a major determinant of their transit to target sites and function upon recipient-cell uptake. For instance, EVs harbor a large proportion of membrane proteins which can prime the transfer of information between cells by allowing EVs to dock to a target cell or tissue. Although one-to-one specificity between EV markers and target organs has not been extensively established, EV tissue-homing due to specific membrane proteins has been reported[22–24]. In addition, EVs are also promising vehicles for therapeutic use since they can be naturally or artificially loaded[25] with functional components such as enzymes or nucleic acids [26,27]. In view of these exciting applications, there remains a crucial knowledge gap in specific EV targeting that should be addressed. For instance, which proteins and protein assemblies can facilitate specific tissue targeting, therapeutic packaging and how these components can connect the distribution of EVs to inherited function in recipient cells.

Advances in mass spectrometry have enabled sensitive characterization of the EV protein repertoire[28–30], thereby expanding our knowledge on EV heterogeneity. Nonetheless, these studies use methods that do not inform on the structural aspects of EV protein cargo, which are critical to mechanistically explain EV docking and uptake[31]. Structural characterization of proteins in EVs requires strategies that can directly retrieve structural information without protein extraction, solubilization or detachment of proteins and protein complexes from the lipid environment. Ideally, any such procedure should also take into consideration the low sample amount and inherent heterogeneity of EVs. These pre-requisites still pose a great challenge even to state-of-the-art technologies such as super-resolution, atomic-force, and electron

microscopy, where only a few proteins can be studied simultaneously[32–35].

Leveraging on advances in crosslinking mass spectrometry[36–38], and our recent breakthrough in extracellular crosslinking of whole-cells[39], we present here an intact EV crosslinking mass spectrometry (iEVXL) approach, for the systematic characterization of protein complexes in EVs. Using a pair of metastatic breast cancer cell lines, MDA-MB-231 and the metastatic-derived counterpart LM2[40], we demonstrate here that iEVXL can provide high-resolution and comparative structural information that accurately recapitulates the native structure of previously reported complexes. Furthermore, such structural information in the form of distance restrains can aid in mapping unknown protein structures when adequately combined with computational modeling, and iEVXL can also enable hypothesis-free screening of differential protein-assemblies in EVs. This presents a significant step towards structural characterization of supramolecular complexes present in EVs, and functional elucidation of EV protein complexes responsible for homing and docking mechanisms.

Results

EV isolation and characterization

EVs were isolated from MDA-MB-231 and LM2 breast cancer cell lines using an ultracentrifugation protocol (**Fig 1A**) and the quality of the EV preparations was assessed by multiple biochemical and biophysical techniques, according to the MISEV guidelines[41] (**Fig 1B-E**). We confirmed that our EV preparations are significantly enriched for proteins from exosomes, plasma membrane and focal adhesions using a sensitive shotgun proteomics experiment (**Fig 1B**). Amongst these, EV markers such as CD9, CD63, CD81, TSG101 and PDCD6IP (Alix), were consistently enriched in EVs from both cell lines (**Fig 1C**). CD81 enrichment in EV preparations was also validated by Western Blot (**Sup Fig 1**). As previously reported, HLA proteins were also highly enriched in EVs when compared to cells[14] (**Fig 1C**). In addition, the integrity, size distribution and concentration of the EV populations were determined by biophysical methods. Imaging by negative stain transmission electron microscopy (NS-TEM) showed that EVs derived from both MDA-MB-231 and LM2 are largely intact after isolation (**Fig 1D**). Nanoparticle tracking analysis (NTA) revealed that the EV concentration and size distribution between MDA-MB-231 and LM2 are similar, with a particle:protein ratio (purity) of $2x10^9$ and particle size of 110 nm, which are expected for ultracentrifugation preparations (**Fig 1E**). Collectively, these data indicate high quality and purity of the EVs isolated from both MDA-MB-231 and LM2 cell lines.

Intact extracellular vesicle crosslinking (iEVXL)

Structural characterization of protein complexes in EVs presents unique challenges, since the phospholipid-bound entities are different in chemical accessibility and biochemical composition compared to a cell lysate. EVs are naturally densely loaded with membrane proteins which can be hard to solubilize while preserving their structure and interactions. In this work, we studied the EV interactome by iEVXL.
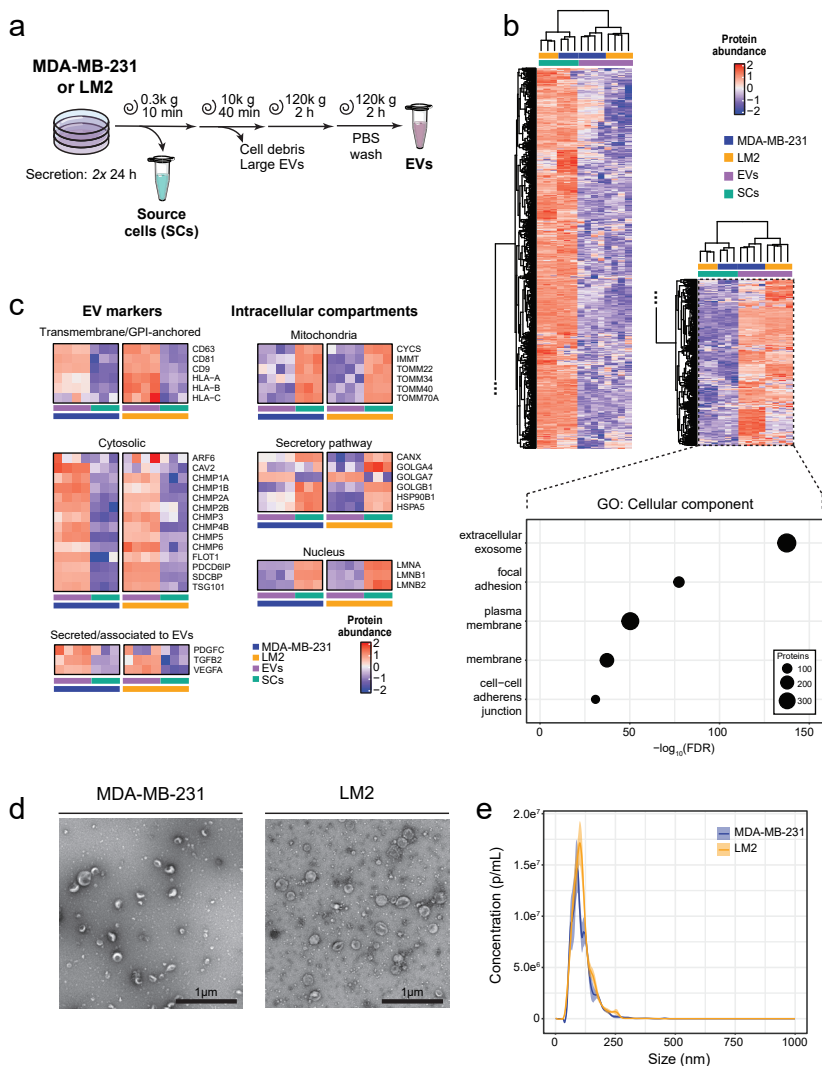
Figure 1. Extracellular vesicle isolation and characterization. (a) Schematic workflow for EV isolation. MDA-MB-231 or LM2 source cells (SCs, green) or EVs (purple) were harvested by centrifugation and ultracentrifugation respectively, for further characterization and crosslinking mass spectrometry. (b) Heatmap of proteins significantly enriched in SCs or EVs, based on Student's T-test (q-value < 0.05). EV-enriched proteins were significantly enriched for the GO:CC terms of "extracellular exosome", "plasma membrane" and "focal adhesion" (insert). Dot size represents the number of proteins mapped to each term, and the top five most significantly enriched terms based on false discovery rate (FDR) were plotted. (c) Relative abundance of EV and intracellular compartment marker proteins. EV preparations were significantly enriched with EV markers and depleted of intracellular compartment proteins. (d) Negative stain transmission electron microscopy (NS-TEM) analysis showing the morphology and integrity of the isolated EVs used for crosslinking experiments. Representative images. Scale bar: 1 μm. (e) Nanoparticle tracking analysis (NTA) showing the size and concentration of isolated EVs used for crosslinking experiments. Purity (particle:protein ratio) of EVs was ~2×10^9, and mean particle size was similar in EVs derived from either cell line.
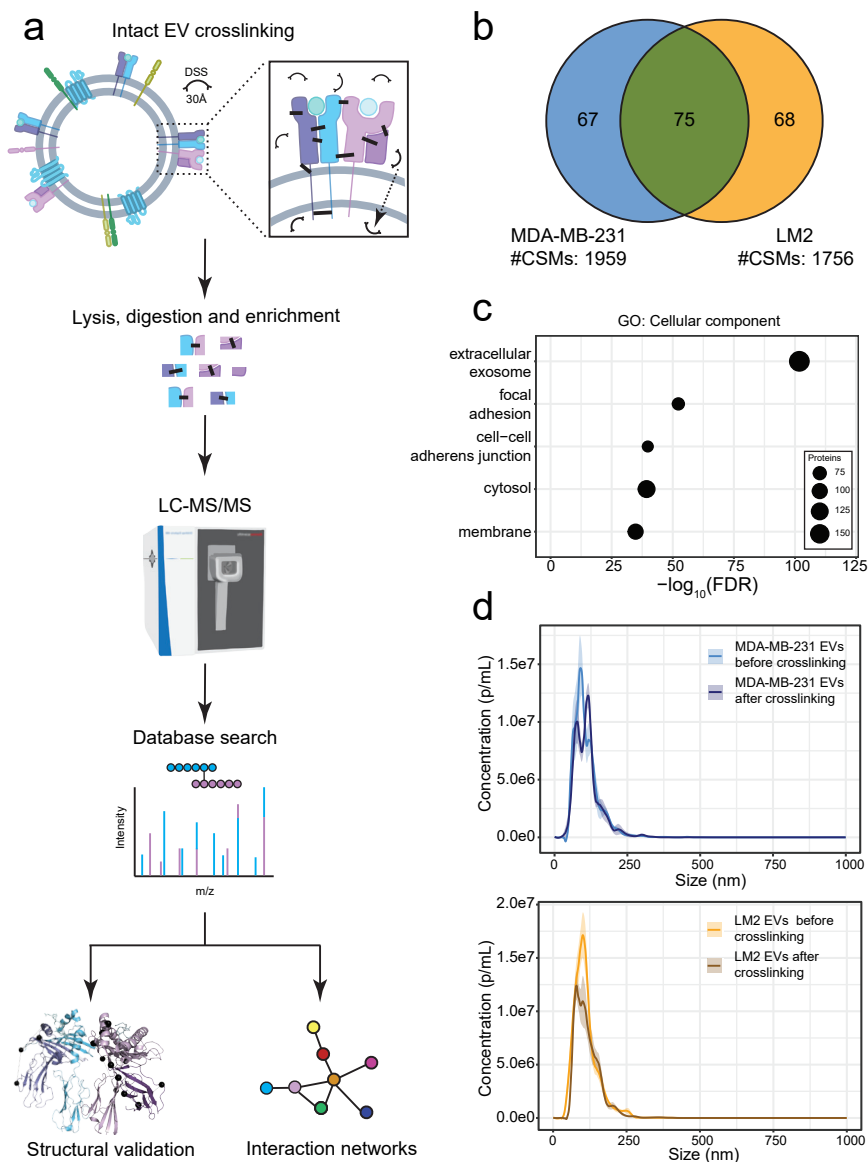
Figure 2. Intact extracellular vesicle crosslinking mass spectrometry (iEVXL). (a) Schematic workflow used for intact EV crosslinking and LC-MS/MS analysis in the iEVXL pipeline. (b) Overlap between crosslinked proteins in EVs derived from MDA-MB-231 (blue) and LM2 (yellow). The total number of crosslinked peptide spectrum matches (CSMs) is reported for each cell line. (c) GO:CC annotation of crosslinked proteins. Crosslinked proteins were significantly enriched in annotations of "extracellular exosome", "plasma membrane" and "focal adhesion". Dot size represents the number of proteins mapped to each term, and the top five most significantly enriched terms based on false discovery rate (FDR) were plotted. (d) Nanoparticle tracking analysis (NTA). EVs before and after chemical crosslinking were indistinguishable in both size or concentration.

Compared to conventional strategies of interactome profiling that require prior solubilization and extraction of interacting proteins into a non-membranous and aqueous or detergent system[42,43], crosslinking the EVs directly with a chemical crosslinker allows the preservation of labile protein-protein interactions in their native state before extensive biochemical extraction and purification. Here, we chose to crosslink with disuccinimidyl suberate (DSS), a lysine residues within a 30Å radius (C_α-C_α distance). The partial membrane permeability of DSS also allows it to penetrate the EV membrane slightly, to access peri-membrane protein complexes on the cytoplasmic side, which may otherwise be missed with membrane-impermeable crosslinkers.

Using a carefully titrated concentration of DSS (2 × 0.5 mM DSS; optimized in **Sup Fig 2**), we crosslinked EVs derived from a pair of metastatic breast cancer cells lines (**Fig 2A**). Only after chemical crosslinking, proteins were extracted from the EVs. This would ensure chemical immobilization of interacting proteins in the native orientation before harsh isolation of protein complexes from their native lipid environment. After DSS crosslinking, denaturing lysis, and trypsin digestion, crosslinked peptides were enriched by strong cation exchange (SCX) chromatography before mass spectrometry analysis of each SCX fraction. The spectral information obtained by LC-MS/MS was then compared with a protein database to retrieve linked peptide pairs (**Fig 2A**). In the analysis of protein-protein interactions, and structural mapping within protein complexes, we focused largely on the restraints imposed by these crosslinked peptide pairs.

Starting from ~100 µg of MDA-MB-231 or LM2 EV proteins, we identified 1959 crosslinked peptide spectra matches (CSMs) and 1756 CSMs respectively. These CSMs originated from about 140 unique proteins in each EV sample, where up to 52% of these proteins were found crosslinked in EVs derived from both cell lines (**Fig 2B**). Furthermore, crosslinked proteins found in EVs from both cell lines mapped to similar ontology terms (**Fig 2C**), in strong agreement with the EV proteome (**Fig 1B**). This provided confidence that crosslinks adequately represent the EV interactome, even with extensive SCX enrichment. Membrane proteins and focal adhesion terms were abundantly mapped in the set of crosslinked proteins (**Fig 2C**). In addition, the crosslinked positions found within the well-known integral membrane proteins integrin β1, Ep-CAM and CD9 were also consistent with the documented membrane topology, where most of the intralinks were found between residues in the extracellular domain (**Sup Fig 3**). After crosslinking, size-distribution and concentration of EVs did not change noticeably, as shown by NTA characterization (**Fig 2D**), indicating that EVs were also not aggregating upon crosslinking. Collectively, these readouts comprehensively ascertain that EVs remain intact and largely free from aggregation at the conditions we used for DSS crosslinking.

Structural characterization and modeling of native protein complexes by iEVXL

Experimentally, not only were EVs amenable to our chemical crosslinking workflow (**Fig 2**), but many crosslinked peptide pairs could also be identified by mass spectrometry subsequently. This provided spectral evidence of EV protein-protein interactions but may also allow structural characterization of native protein complexes in

EVs. Coupled to the use of partially membrane permeable DSS as chemical cross-linking agent, we hypothesized it could be possible to retrieve structural information for protein complexes either on the EV external surface, or encapsulated in the EV, but close to the EV membrane. Such data could potentially be very informative to study EV docking, as well as the structural basis of EV interactions with recipient cells. The untargeted nature of iEVXL allows the discovery of important EV protein complexes not known *a priori* and is also suitable for retrospective interrogation of protein-protein interactions. Most importantly, the unique membrane curvature of EVs could mean that the membrane proteins might have unique intra-membrane conformations which could only be studied with minimally disruptive techniques like iEVXL. We demonstrate this possibility here by mapping the crosslinks found in our dataset on available protein structures, focusing on high abundant EV proteins which have been previously linked to the metastatic potency of EVs.

In this direction, we first examined abundant crosslinks involving α-enolase in this dataset. A-enolase is a glycolytic enzyme commonly found in EVs derived from tumor cells[44–46], and has been shown to function as a dimer, with its structure previously resolved by X-ray crystallography[47]. We mapped the α-enolase crosslinks found in our dataset to the high-resolution crystal structure of this protein (PDB: 3b97; 2.2 Å), and all the crosslinks we detected were within the distance of 30 Å, a restraint imposed by the physical length of DSS (**Fig 3A**). This suggests that XL-MS with DSS on EVs can recapitulate well the natural dimeric structure of α-enolase. Similarly, by mapping the crosslinks involving 14-3-3 proteins detected in the same dataset, we were also able to confirm the validity of crosslinks involving 14-3-3 against the crystal structure of 14-3-3 α/β-heterodimer (PDB: 4dnk, 2.2 Å). In particular, 14-3-3 proteins have been shown to be abundant in oncogenic EVs[29,48], which is consistent with the metastatic breast cancer origin of our EVs. Since the amino acid resolution of XL-MS could distinguish between the different 14-3-3 isoforms, 'mix-n-match' assembly of 14-3-3 dimers can be distinguished. We mapped all the 14-3-3 crosslinks detected on a representative αβ-heterodimer (**Fig 3B**). Collectively, these demonstrate proof-of-principle that structural mapping by chemical crosslinking on intact EVs is feasible using a partially membrane-permeable crosslinkers such as DSS, and can detect endogenous protein complexes in documented conformations.

Taking the capabilities of structural mapping one step further, we used structural information from iEVXL to extend partially resolved protein structures. The human moesin protein is highly abundant in MDA-MB-231 and LM2 EVs and contains a poorly characterized long α-helix. To date, a full-length structure of insect (*Spodoptera frugiperda*) moesin has been determined in a closed conformation with its α-helix folded as an antiparallel coiled coil[49]. Small-angle X-ray scattering data combined with crystal structures of the moesin homologue ezrin confirmed this arrangement for human moesin and suggests these proteins can form both inactive monomers and active domain-swapped antiparallel dimers[50]. From our data, four crosslinks were within the α-helical domain of human moesin, including one between the two supposed α-helices of the coiled coil. Initially, we mapped our data into the AlphaFold[51] predicted human moesin structure, which indeed shows a coiled-coil architecture between residues 350 to 450. Although our crosslinks supported the coiled-coil section of the AlphaFold model, the coiled coil was followed by a long-disordered stretch,
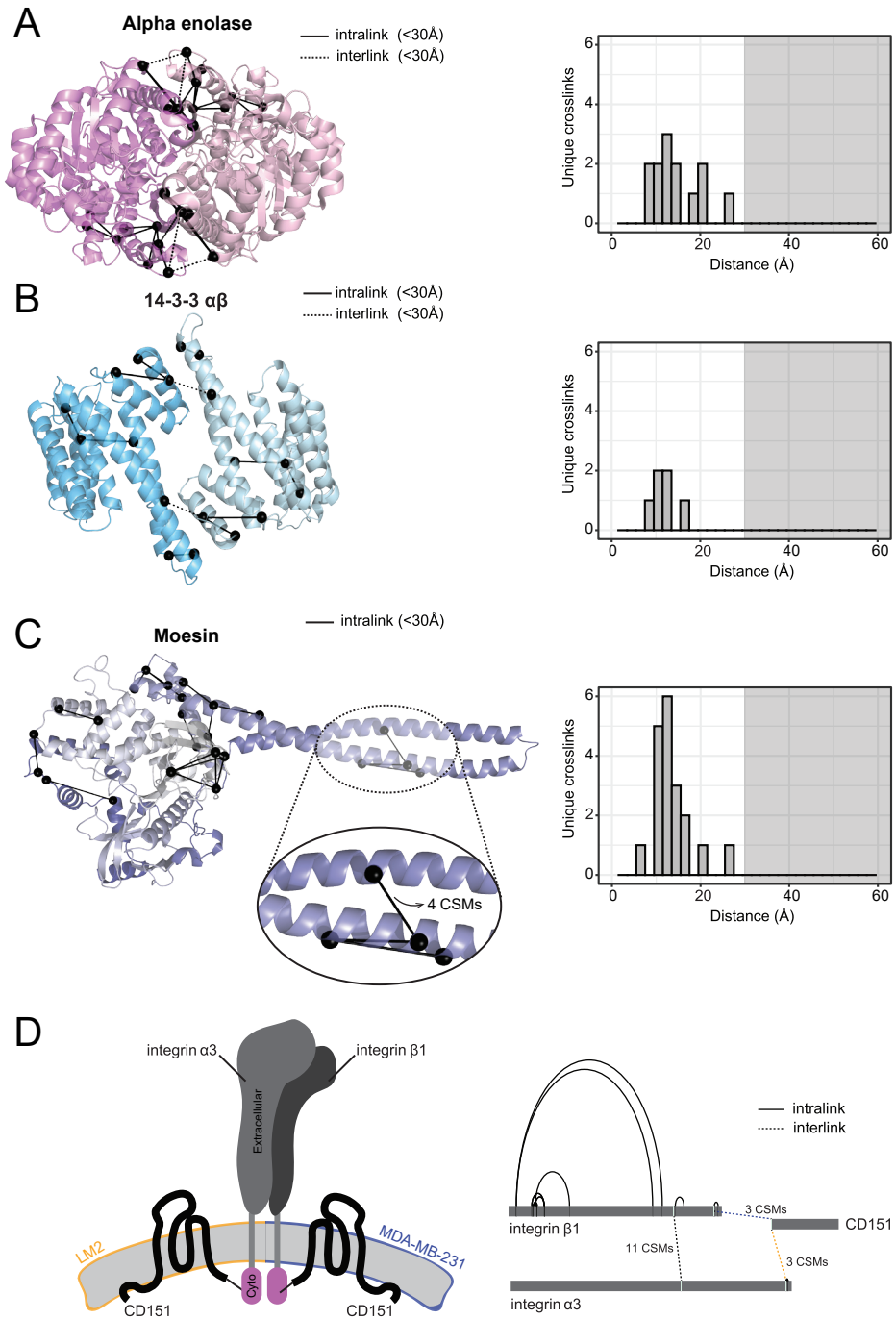
Figure 3. Visualization of detected crosslinks on protein structures. (Caption on next page)

Figure 3. Visualization of detected crosslinks on protein structures. Crosslinks of EV enolase, mapped on high-resolution crystal structure of homodimeric α-enolase (PDB: 3b97). Histogram summarizing the measured Euclidean distances (in Å) between pairs of unique-ly crosslinked lysine residues is shown on the right. (b) Crosslinks of EV 14-3-3 mapped on high-resolution crystal structure of 14-3-3 αβ complex (PDB: 4dnk). (c) Structure of ho-mology-modeled human moesin (colored in white to blue from N- to C-terminal). Crosslinks mapped on homology-modeled human moesin structure. Insert zooms into residues 375-450 of moesin, a region of the protein that has been proposed to display a coiled-coil type of ar-chitecture. Consistent with the length of DSS crosslinker, all crosslinks detected (a-c) were within the Euclidean distances of 30 Å. (d) Mode of interaction for CD151 and integrin α3-β1 dimer in MDA-MB-231 and LM2-derived EVs. Left: proposed mode of interaction via the cy-toplasmic tails; Right: protein interaction map showing unique crosslinks detected in EVs that support this interaction model. Blue crosslinks were detected in MDA-MB-231 EVs and yellow crosslinks were detected in LM2 EVs.

decreasing our confidence in the AlphaFold model. Hence, we instead generated a homology model of human moesin based on the *Spodoptera frugiperda* structure[49] (PDB: 2I1K). All the crosslinks found in our data were within 30 Å when mapped on the homology-modeled human moesin structure (**Fig 3C**, histogram). In addition, our crosslinks also supported the coiled-coil type of architecture (**Fig 3C**, insert) and thus a closed (inactive) conformation of the protein. Therefore, the EV crosslinking data we have generated may also be used to supplement and weigh-in on contradictory structural reports, to propose the more plausible structure in combination with other structural techniques.

Finally, our iEVXL approach may also allow observations about higher-order struc-tural assemblies, for instance in the interaction between integrin $\alpha_3\beta_1$ dimer and tet-raspanin CD151 (**Fig 3D**). The crosslinks revealed that CD151 interacts with the in-tegrin αβ dimer via the cytoplasmic tails (**Fig 3D**, left). Moreover, our data suggested that this interaction can occur in two ways, since the crosslink from CD151 to the α subunit was found exclusively in the LM2 EVs and the crosslink from CD151 to the β subunit exclusively in the MDA-MB-231 EVs (**Fig 3D**, yellow and blue dotted lines). This interaction has been documented previously by AP-MS[52] and strongly implicat-ed in metastasis[53–56], although crucial structural information regarding interacting do-mains could not be inferred from classical interaction studies. This highlights another key advantage of iEVXL as a complementary technique. Therefore, we demonstrate with four examples of structural modeling, that distance restraints from our iEVXL dataset are coherent with complete or partial structures, and that such data can po-tentially aid in the detailed re-construction of the EV docking interface with recipient cells.

Distinct back-to-back annexin A2 conformation in LM2-derived EVs

In the most ideal way, structural profiling should also be sensitive to changes in structural features between closely related systems. To test this, we compared the structural features in EVs derived from MDA-MB-231 and the closely related LM2 cells and found a distinct back-to-back conformation for annexin A2 dimers that was unique to LM2-derived EVs (**Fig 4**). Annexin A2 is a phospholipid-binding protein involved in the endocytic and exocytic pathways. Annexin A2 is a well-established marker of EVs[30], and was abundantly crosslinked in this current dataset. Structur-
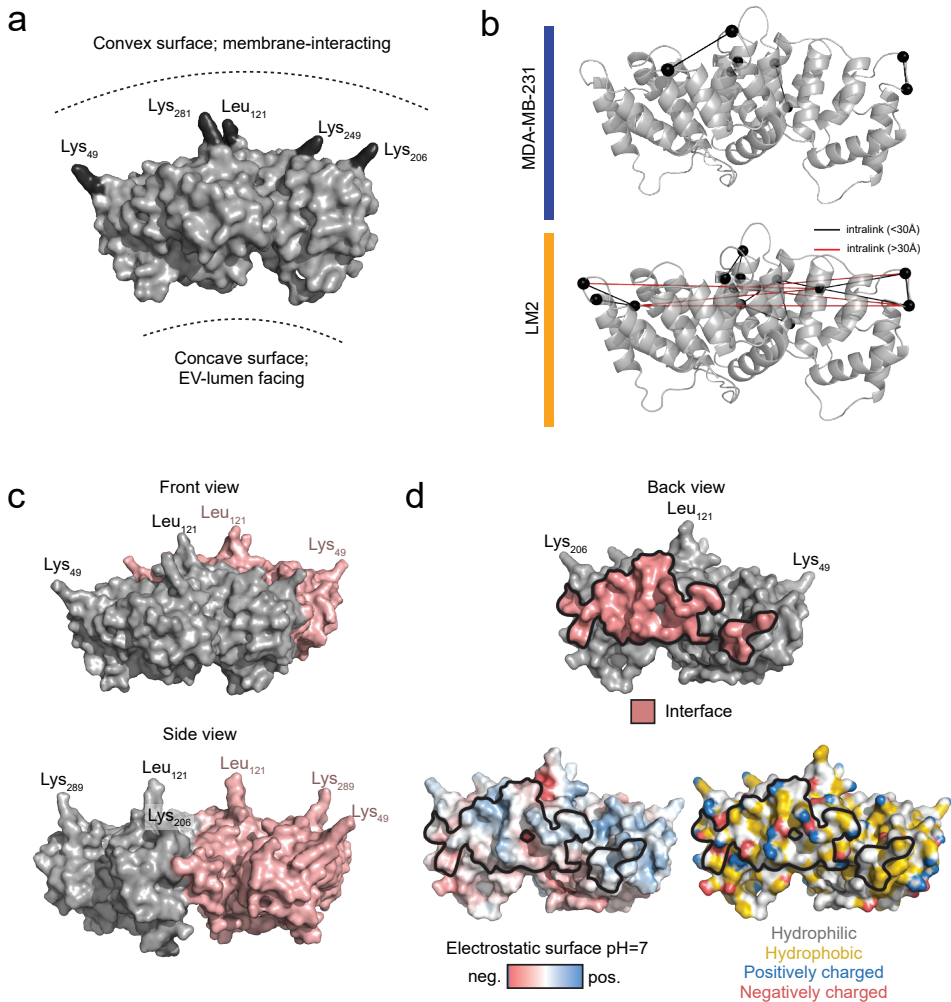
Figure 4. Crosslinking data driven structural modeling for a novel dimeric annexin A2 complex. (a) Front view of the annexin A2 monomer. Important structural features for annexin A2 functionality are annotated. Residues labelled in the convex surface are involved in plasma membrane interactions. (b) EV crosslinks mapped into the monomeric AlphaFold-predicted structure of annexin A2. Crosslinks detected in MDA-MB-231 (blue) and LM2 (yellow) EVs are plotted separately. (c) Front and side view of annexin A2 crosslinking data driven model. Both copies of annexin A2 interact through the back side, in a back-to-back dimer conformation. (d) Back view of annexin A2 monomer showing the interface (pink), electrostatic surface at pH=7.0 and hydrophobicity color-coded surface.

ally, annexin A2 has been shown to exist as a monomer, dimer or hetero-tetramer[57,58]. Monomeric annexin A2 consists of a concave surface on the bottom and a convex surface at the top, from which it is thought to attach to membranes via protruding lysine and leucine residues[59] (**Fig 4A**). Although all the crosslinks found in MDA-MB-231 derived EVs could be explained by monomeric annexin A2 (**Fig 4B**, blue), 5 crosslinks found only in LM2-derived EVs exceeded the distance restraints

of 30 Å when mapped to the same monomeric annexin A2 structure (**Fig 4B**, yellow; over-length crosslinks represented by red lines). These distance violations seem to imply that there is substantial non-monomeric annexin A2 in LM2 EVs, but not in MDA-MB-231 derived EVs. Apart from structural differences, protein abundance can sometimes explain fewer crosslinks mapping to a complex. However, Annexin A2 ranked similarly in abundance within each cell line (13th and 16th for LM2 and MDA-MB-231). Hence the non-detection of this back-to-back annexin A2 confirmation in MDA-MB-231 is unlikely to be attributed to differential protein abundance alone.

To understand further how the over-length crosslinks that were detected only in LM2-derived EVs might be biologically relevant, we used our crosslinking data for guided molecular modeling. By assuming annexin A2 forms dimers related by C2 rotational symmetry[60], we generated symmetry pairs of intra-protomer crosslinks and analyzed the accessible interaction space[61–63] on annexin A2. Complexes could be found using 4 out of 5 pairs of these crosslinks, indicating that these over-length crosslinks were in fact inter-protomer crosslinks. Therefore, we selected these as restraints to guide the subsequent docking process. Out of 200 docked complexes generated by HADDOCK, 186 modeled complexes clustered together, indicating that a single type of interaction is likely to explain the crosslink restraints. From these, we selected the best annexin A2 model based not only on the Haddock scores, but also on complex Matched and Non-accessible Crosslink (cMNXL) scores (**Sup. Fig 5A**, Complex_181w represented by a red dot). While the HADDOCK score is based on Euclidean distances, the cMNXL score is based on the solvent accessible surface distance (SASD). This model revealed a 'back-to-back' conformation, where the convex surfaces from both annexin A2 monomers face the same side (**Fig 4C and Sup Fig 5B-C**). In further support of this 'back-to-back' annexin A2 conformation, none of the other existing Annexin A2 homodimeric[58] (PDB: 1xjl) and heterotetrameric[64] (PDB: 5lpu) structures could explain the over-length crosslinks observed specifically in the LM2 dataset (**Sup Fig 4**).

The inter-protomer interaction surface of annexin A2 was quite large (1172.3 Å^2) and displayed a neutral electrostatic potential as well as quite some hydrophobic patches, indicating a hydrophobicity-based interaction (**Fig 4D**). Given that residues involved in protein interfaces tend to be more conserved than other residues, we also determined the conservation at the protein surface using all the annexin A2 orthologue sequences present in Uniprot. This analysis revealed that, as expected, the plasma-membrane interacting surface (**Sup Fig 6**, top view) was highly conserved. The interaction surface found in our 'back-to-back' model was also relatively conserved (**Sup Fig 6**, back view, interface highlighted in black) when compared to other surfaces, including the lateral surfaces (**Sup Fig 6**, lateral views 1 and 2) which were previously thought to contribute to annexin A2 oligomerization[65]. Overall, these results strongly supported a novel 'back-to-back' mode of interaction for annexin A2 in LM2-derived EVs.

Discussion

In recent years, EVs have emerged as key mediators of intercellular communication, and have been the focus of many biopharmaceutical and therapeutic developments[18,19,66,67]. The protein repertoire and structural features on EVs could critically influence the pharmaceutical utility, as these will ultimately determine tissue-targeting and uptake by recipient cells. In this respect, the advent of sensitive shotgun proteomics has made largescale and detailed EV protein characterization rather feasible, and propelled EVs into the scene of clinical diagnostics[3,29]. Nevertheless, the challenge to provide resolution on structural features or topology of EVs largely remains. Key questions of functional EV targeting require higher resolution mapping of protein complexes and protein conformations in and on the surface of EVs to be further addressed.

In recent years, the interest to understand EV structural features[31] and interactome[68] has grown significantly, but techniques amenable to structurally characterize EV protein complexes are remarkably scarce. EVs are unique and heterogenous entities that significantly challenge existing structural methods. Cryo-electron microscopy and tomography are state-of-the-art techniques that have provided a wealth of structural information on many biologically important complexes. Nonetheless, the stringent requirement for sample homogeneity and low complexity complicates the application of such techniques to study EVs. On the other hand, fluorescence-based techniques such as super resolution microscopy still hold a resolution limit of around 20 nm[69], which is much larger than the interaction space of a protein complex. In this respect, iEVXL, as we describe here, can bridge with high resolution (<30Å) structural information on interacting proteins in their native environment. Such data obtained may also complement the information obtained from tomography and fluorescence-based techniques, to collectively sketch the EV docking interface. Therefore, we envision the application of iEVXL as an important step towards understanding EV biology and selective docking mechanisms.

In this technical report, we demonstrate that iEVXL can detect differential protein interactions in EVs, when coupled to chemical crosslinking with partially membrane-permeable crosslinkers such as DSS. We show that detailed and hypothesis-free analyses of EV membrane interactomes are feasible at the scale of ~100 µg of EV proteins, and the structural information obtained are highly reliable and consistent with known experimental structures. With the demonstrated cases of α-enolase, 14-3-3α/β, moesin, CD151- integrin $α_3β_1$, and annexin A2 dimers, we showcase a range of utility for iEVXL, in protein structure interface mapping, oligomeric isoform docking, flexible region structural modeling, and determination of higher-order multimeric structural assemblies. As we overcome the generic sensitivity limitations of protein-interaction mass spectrometry, we expect the moderate sensitivity of iEVXL to improve further. Notwithstanding, we envision that the broader application of iEVXL will allow significantly better engineered EV tissue targeting.

Methods

Cell culture and EV isolation

MDA-MB-231 (obtained from ATCC) and LM2 cells (provided by The Netherlands Cancer Institute, NKI) were cultured in DMEM supplemented with 10% fetal bovine serum (FBS; HyClone, USA), 10 mM L-glutamine, 50 U/mL penicillin and 50 µg/mL streptomycin (Lonza) in a humidified incubator at 37 °C with 5% CO_2. Cells were detached using 10 mM EDTA/PBS for 5 min at 37 °C. Secretion media was prepared by depleting bovine derived EVs from the culture media. To do so, DMEM containing 20% FBS was centrifuged overnight at 100,000 × g at 4 °C in a Sorvall T-865 rotor (Thermo Fisher Scientific), filtered on a 0.22 µm Stericup device (Millipore, USA), diluted to 10% FBS, supplemented as previously described and kept at 4 ºC.

For EV secretion, cells were seeded on 10 plates (15 cm in diameter) at 50% confluence and left to attach overnight. After attachment, cells were gently washed 3× with warm PBS, before addition of secretion media. Conditioned media containing secreted EVs was collected after 24 h and 48 h, fresh secretion media was added to replace the collected conditioned media. Cell viability was measured at the start and end of secretion using the Trypan Blue method, and it remained ~95%. The conditioned media were spun down at 300 × *g* for 10 min to deplete cells, transferred to clean 50 mL Falcon tubes, spun down at 10,000 × *g* for 40 min to deplete cell debris and larger vesicles, transferred to clean 50 mL Falcon tubes and kept at 4 ºC. The two collections were pooled and EVs were immediately pelleted by ultracentrifugation at 120,000 × *g*, at 4 ºC for 2 h in a Sorvall T-865 rotor. The pellet was resuspended by gentle pipetting in 10 mL cold PBS supplemented with 50 µg/mL DNAse I (Sigma-Aldrich), to decrease nucleosome contaminations on EV preparations. Purified EVs were finally pelleted again by ultracentrifugation at 120,000 × *g* at 4 ºC for 2 h. The EV-pellet was thoroughly resuspended in 400 µL of PBS, spun down at 10,000 × *g* for 5 min and the EV-containing supernatant was kept. Aliquots were separated for further characterization of the EV populations.

EV crosslinking and protein digestion

MDA-MB-231 and LM2-derived EVs were crosslinked using disuccinimidyl suberate (DSS; Thermo Fisher Scientific) with the optimal 0.5 mM crosslinker concentration which was determined in an independent experiment (Sup Fig 1). EV samples (at an average concentration of 7.5x10⁹ p/mL, 0.5 mg/ml protein in EV lysate) were crosslinked in 0.5 mM DSS for 20 min at RT. To promote crosslinking of lower abundant species, a second round of 0.5 mM DSS was added for another 20 min at RT. The crosslinking reaction was then quenched with 100 mM TRIS pH 8.5 for 5 min. Crosslinked EVs were aliquoted for further characterization. Crosslinked EVs were then lysed by thorough vortexing in 0.5% SDC, 8 M Urea in 50 mM ammonium bicarbonate, followed by 30 min end-to-end rotation at 4 ºC and 15 cycles of sonication at 4 ºC (30 seconds on, 30 seconds off) in a Bioruptor (Diagenode, Belgium). Proteins were reduced with 4 mM dithiothreitol (DTT) at RT for 60 min, alkylated with 16 mM iodoacetamide (IAA) at RT for 30 min in the dark which was then quenched by addition of 4 mM DTT. Proteins were first digested by addition of Lys-C (at a 1:50 ratio (w/w); Wako, Japan) at 37 ºC for 2 h, followed by dilution to 2 M Urea and fur-

ther digestion with Trypsin (at a ratio 1:50 (w/w); Sigma Aldrich) at 37 ºC overnight. Protein digestions were stopped by acidification to 5% FA, and precipitated SDC was pelleted by centrifugation at 20,000 × g at 4 ºC for 30 min. Supernatants were carefully collected, desalted using Sep-Pak C18 cartridges (1cc; Waters, MA, USA), vacuum dried and stored at -20 ºC until further use.

SCX fractionation and enrichment of crosslinked peptides

The desalted peptides were dissolved in 10% formic acid, 2.5% DMSO and loaded on a Luna 100A SCX column (50 mm × 2mm, 5 µm, Phenomenex product number 00B-4398-B0) with the help of a C18 Opti-Lynx trap column (4.6mm × 5 mm, 49 µm, Optimize Technologies product number 11-02874-TA). Solvent A consisted of 0.05% formic acid, 20% acetonitrile in water and solvent B consisted of 0.05% formic acid, 20% acetonitrile, and 0.5 M NaCl. The SCX gradient was: 0-10% B in 2 min, 10-40% B in 7 min, 40-80% B in 2.5 min, 80-100% B in 4 min, 100% B for 7 min. One-minute fractions were collected and pooled into 6 or 7 approximately equal fractions by mean UV intensity. The pooled fractions were desalted with Oasis HLA 96-well µElution plate (Waters, MA, USA), vacuum-dried and stored at -80 °C.

LC-MS/MS of crosslinked SCX fractions

The data was acquired with an Ultimate 3000 system (Thermo Fisher Scientific) coupled to an Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific). Peptides were trapped (Dr Maisch Reprosil C18, 3 µM, 2 cm × 100 µM) for 2 min in 5% solvent B at a flow rate of 300 nL/min, before being separated on an analytical column (Agilent Poroshell, EC-C18, 2.7 µM, 50 cm x 75 µM). Solvent B consisted of 0.1% formic acid in 80% acetonitrile. Crosslinked peptides were then separated in the analytical column at a fixed flow rate of 300 nL/min as follows: each SCX fraction was separated using an optimal 95 min. linear gradient (ranging from 9-40% to 6-35% B) followed by a 3 min steep increase to 99% B, a 5 min wash at 99% B and a 10 min re-equilibration step at 5% A. The mass spectrometer was operated in a data-dependent mode (DDA). Peptides were ionized in a nESI source at 1.9 kV and focused at 40% amplitude of the RF lens. Full scan MS1 spectra from 350 - 2200 m/z were acquired in the Orbitrap at a resolution of 60,000 with the AGC target set to $3×10^6$ and under automated calculation of maximum injection time. Cycle time for MS2 fragmentation scans was set to 2 seconds. Only peptides with charged states 3-8 were fragmented, and dynamic exclusion was set to a duration of 16 ms. Fragmentation was done using a stepped HCD collision energy strategy (NCEs: 28, 31, 34%). Fragment ions were accumulated until a target value of $1×10^5$ ions was reached under an automated calculation of maximum injection time, with an isolation window of 1.4 m/z before injection in the Orbitrap for MS2 analysis at a resolution of 30,000. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[70] partner repository with the dataset identif'ier PXD029591.

Crosslinking database search and data analysis

For crosslinked peptide analyses, spectra were extracted from "raw" files from precursors ranging between 350 and 20,000 Da, filtered by signal-to-noise ratio of 2

3

and converted to MGF format using Proteome Discovered software (v2.4, Thermo Scientific). MGF files were searched in pLink 2[71] against a database containing the 1000 most abundant proteins of both MDA-MB-231 and LM2 EV proteomes, as determined in section 4.10., and appended with the sequences of common FBS contaminants to avoid misidentification of crosslinked peptide sequences from *Bos taurus*. The database was further curated to remove signal peptides from the protein sequences (note that positions within proteins still follow the Uniprot numbering). DSS was set as non-cleavable crosslinker and trypsin was set as the digestion enzyme, and up to 3 missed cleavages were allowed. Peptide length was set to between 6 and 60 amino-acids, precursor and fragment ion tolerance were set to 10 and 20 ppm respectively, and oxidation of methionine and acetylation of protein N-terminus were set as variable modifications while carbamidomethylation of cysteines was set as a fixed modification. FDR was set at 1% at all levels, which was calculated by using a reverse decoy database strategy. E-scores were not computed to minimize processing times.

Crosslinking and proteomics data were analyzed using Excel and in-house built R scripts[72], and plots refined with Illustrator 2020 (Adobe, USA). Spectra and site files from pLink were processed to generate the tables in the supplementary data, and the scripts used to curate pLink outputs have been deposited in GitHub (https://github.com/hecklab/pLink-results-analysis). Only crosslinks identified with sufficient spectral evidence (≥ 2 CSM per cell line) were kept while crosslinks involving histones, a debatable contaminant in EV preparations, were not analysed further. In cases of protein ID ambiguity due to shared crosslinked peptides between different proteins, intra-links were preferred over inter-links.

Molecular modeling

A homology model of human moesin (2-577) was generated using Swissmodel[73], with the (almost) full-length *Spodoptera frugiperda* moesin structure 2I1K[49] as a template (homology: 57.89% identity). The N-terminal residue missing from this model was manually added using Coot[74].

The AlphaFold predicted structure of annexin A2 (Unirpot ID: P07355) was used for the molecular docking procedures. DisVis webserver[61–63] was used to check compatibility between the crosslinks (including their symmetry mates). Molecular docking was performed with HADDOCK 2.4 web-service[63,75]. The default parameters of HADDOCK were used with unambiguous restraints based on the crosslinks. The HADDOCK structures were scored based on the crosslinks solvent accessible surface distances (SASDs) using XLM tools[76]. The best docking models were picked based on both the HADDOCK and the cMNXL scores[77]. Cocomaps[78] was used to determine which residues were in close contact. The interface was defined as the residues predicted to be in close contact for both annexin A2 copies in the modeled structure. For conservation analysis, all the annexin A2 orthologs present in Uniprot (excluding low quality proteins) were aligned using the Clustal Omega[79] algorithm and the alignment and phylogenetic tree were fed to ConSurf[80] to calculate the conservation score and visualize it on the protein structure.

Negative stain electron microscopy (NS-TEM)

Thin layer continuous formvar/carbon-coated copper mesh grids (Ted Pella 400 mesh Cu, 01754-F) were glow discharged for 10 seconds at 10 mA, and immediately incubated with 3 µL of undiluted EVs in PBS for 45s. Excess solution was blotted away, and the samples were stained first by a quick immersion in 2% (w/v) of uranyl acetate followed by re-staining for 1 min with the same reagent. After each staining step, the excess solution was blotted away. Grids were then dried at RT before electron microscopy imaging. NS-TEM data was collected on a Talos L120C transmission electron microscope (Thermo Fisher Scientific) operated at 120 kV. Images were acquired with a 4k × 4k Ceta CMOS camera (Thermo Fisher Scientific) at a magnification of 11000× corresponding to a pixel size of 13.6 Å.

Nanoparticle tracking analysis (NTA)

EVs were analyzed in a NanoSight NS500 (Malvern Panalytica, UK), equipped with a sCMOS camera and a Blue405 laser. The camera level was set to 16. The samples were diluted 1:500 in PBS to a final volume of 1 mL to be in the optimal range of operation (between 30-100 particles/frame). Four videos of 1 min were taken at 25 FPS and averaged with the built-in NanoSight Software NTA v.3.4 using a detection threshold of 5.

SDS-page and western blot

Lysed EVs or source-cell material were resolved on 12% Bis-Tris Criterion XT precast gels (Biorad, USA) with 1x XT-MOPS buffer at fixed voltage of 150 V for about 2 h. Proteins were stained in-gel using Imperial Protein Stain (Thermo Fisher Scientific). For Western detection, proteins were transferred to a PVDF membrane in Towbin buffer (0.025 M Tris, 0.192 M glycine, 20% methanol) at 100 V and 4 °C for 1 h. Membranes were washed 3x with TBS buffer containing 1% Tween-20 (TBST) and then blocked for 1 h in TBST supplemented with 5% Blotting-Grade Blocker (Biorad). Primary antibody (α-CD81 at 1:200) was incubated at 4 °C overnight in TBST supplemented with 1% milk. Secondary antibody incubation was done using HRP conjugated α-mouse IgG antibody (1:2,000 dilution) for 2 h at RT in the same buffer. Between and after antibody incubations, membranes were washed 3× 10 min in TBST. HRP signal was visualized using SuperSignal West Dura (Thermo Fisher Scientific) substrate, on an Amersham Imager 600 (GE healthcare, USA).

MDA-MB-231 and LM2 EV and source cells proteomic characterization

The proteomic characterization of the EVs and source cell proteomes has been done re-using high resolution data previously published[29]. Cell culture and EV-isolation conditions were the same and the reproducibility of the source cell and EV proteomes was assessed in independent biological replicates for both cells lines under such conditions. Briefly, the "raw" files from 4 biological replicates of EVs isolated from either MDA-MB-231 or LM2 cells, as well as 3 biological replicates of their source cells, were downloaded from Proteome Xchange (PXD012162) and re-searched using MaxQuant (v_1.6.5.0)[81] against SwissProt human database (downloaded on 09/2019, containing 20,431 protein sequences) appended with common

contaminants. Trypsin was set as the digestion enzyme and up to 2 missed cleavages were allowed. Oxidation of methionine and acetylation of protein N-terminus were set as variable modifications and carbamidomethylation of cysteine was set as a fixed modification. Label-free quantification (LFQ) was enabled using a minimum ratio count of 2 and both razor and unique peptides for quantification. Match between runs approach was enabled using default parameters. Precursor ion tolerance was set to 20 ppm for the first search and 4.5 ppm after recalibration, and fragment ions tolerance was set to 20 ppm. FDR was set at 1% for both PSM and protein level by using a reverse decoy database strategy.

Acknowledgements

Competing Interests

The authors declare no competing interests.

Author Contributions

J.B.M., G.A. and W.W. conceived this study and designed the experimental approach. JBM performed all cell culture experiments. J.B.M. and G.A. performed all crosslinking experiments and data analysis. M.F.P. performed NS-TEM analysis and assisted with structural modeling and analysis. J.B.M. and W.W. co-wrote this manuscript and G.A., M.F.P and W.W. edited this manuscript.

References

1. Raimondo, F., Morosi, L., Chinello, C., Magni, F. & Pitto, M. Advances in membranous vesicle and exosome proteomics improving biological understanding and biomarker discovery. Proteomics 11, 709–720 (2011).

2. de la Torre Gomez, C., Goreham, R. v., Bech Serra, J. J., Nann, T. & Kussmann, M. "Exosomics"-A review of biophysics, biology and biochemistry of exosomes with a focus on human breast milk. Frontiers in Genetics 9, 92 (2018).

3. Hoshino, A. et al. Extracellular Vesicle and Particle Biomarkers Define Multiple Human Cancers. Cell 182, 1044–1061 (2020).

4. Hornung, S., Dutta, S. & Bitan, G. CNS-Derived Blood Exosomes as a Promising Source of Biomarkers: Opportunities and Challenges. Frontiers in Molecular Neuroscience 13, 38 (2020).

5. Liu, J. et al. Circulating Exosomal MiR-107 Restrains Tumorigenesis in Diffuse Large B-Cell Lymphoma by Targeting 14-3-3η. Frontiers in Cell and Developmental Biology 9, 845 (2021).

6. Zhou, E. et al. Circulating extracellular vesicles are effective biomarkers for predicting response to cancer therapy. EBioMedicine 67, 103365 (2021).

7. Tian, F. et al. Protein analysis of extracellular vesicles to monitor and predict therapeutic response in metastatic breast cancer. Nature Communications 12, 1–13 (2021).

8. Thind, A. & Wilson, C. Exosomal miR-NAs as cancer biomarkers and therapeutic targets. https://doi.org/10.3402/jev.v5.31292 5, 31292 (2016).

9. Haraszti, R. A. et al. High-resolution proteomic and lipidomic analysis of exosomes and microvesicles from different cell sources. Journal of Extracellular Vesicles 5, 32570 (2016).

10. Royo, F. et al. Differences in the metabolite composition and mechanical properties of extracellular vesicles secreted by hepatic cellular models. Journal of Extracellular Vesicles 8, 1575678 (2019).

11. Williams, C. et al. Glycosylation of extracellular vesicles: current knowledge, tools and clinical perspectives. Journal of Extracellular Vesicles 7, 1442985 (2018).

12. Raposo, G. et al. B lymphocytes secrete antigen-presenting vesicles. J Exp Med 183, 1161–1172 (1996).

13. Synowsky, S. A. et al. The major histocompatibility complex class I immunopeptidome of extracellular vesicles. Journal of Biological Chemistry 292, 17084–17092 (2017).

14. Bauzá-Martinez, J., Heck, A. J. R. & Wu, W. HLA-B and cysteinylated ligands distinguish the antigen presentation landscape of extracellular vesicles. Communications Biology 4, 825 (2021).

15. Vader, P., Mol, E. A., Pasterkamp, G. & Schiffelers, R. M. Extracellular vesicles for drug delivery. Advanced Drug Delivery Reviews 106, 148–156 (2016).

16. de Jong, O. G. et al. Drug Delivery with Extracellular Vesicles: From Imagination to Innovation. Accounts of Chemical Research 52, 1761–1770 (2019).

17. Elsharkasy, O. M. et al. Extracellular vesicles as drug delivery systems: Why and how? Advanced Drug Delivery Reviews 106, 148–156 (2020).

18. Zipkin, M. Big pharma buys into exosomes for drug delivery. Nat Biotechnol 38, 1226–1228 (2020).

19. Herrmann, I. K., Wood, M. J. A. & Fuhrmann, G. Extracellular vesicles as a next-generation drug delivery platform. Nature Nanotechnology 16, 748–759 (2021).

20. Raposo, G. & Stoorvogel, W. Extracellular vesicles: Exosomes, microvesicles, and friends. Journal of Cell Biology 200, 373–383 (2013).

21. Zhang, D. X., Vu, L. T., Ismail, N. N., Le, M. T. N. & Grimson, A. Landscape of extracellular vesicles in the tumour microenvironment: Interactions with stromal cells and with non-cell components, and impacts on metabolic reprogramming, horizontal transfer of neoplastic traits, and the emergence of therapeutic resistance. Seminars in Cancer Biology 74, 24–44 (2021).

22. Joshi, B. S. & Zuhorn, I. S. Heparan sulfate proteoglycan-mediated dynamin-dependent transport of neural stem cell exosomes in an in vitro blood-brain barrier model. Eur J Neurosci 53, 706–719 (2021).

23. O'Dea, K. P. et al. Monocytes mediate homing of circulating microvesicles to the pulmonary vasculature during low-grade systemic inflammation. Journal of Extracellular Vesicles 9, 1706708 (2020).

24. Jeong Park, E. et al. Exosomal regulation of lymphocyte homing to the gut. Blood Advances 3, 1–11 (2019).

25. Pham, T. C. et al. Covalent conjugation of extracellular vesicles with peptides and nanobodies for targeted therapeutic delivery. Journal of Extracellular Vesicles 10, e12057 (2021).

26.   Jafari, D. et al. Designer Exosomes: A New Platform for Biotechnology Therapeutics. Bio-Drugs 34, 567–586 (2020).

27.   Khan, H., Pan, J. J., Li, Y., Zhang, Z. & Yang, G. Y. Native and Bioengineered Exosomes for Ischemic Stroke Therapy. Frontiers in Cell and Developmental Biology 9, 619565 (2021).

28.   Zhang, H. et al. Identification of distinct nanoparticles and subsets of extracellular vesicles by asymmetric flow field-flow fractionation. Nat Cell Biol 20, 332–343 (2018).

29.   Rontogianni, S. et al. Proteomic profiling of extracellular vesicles allows for human breast cancer subtyping. Communications Biology 2, 325 (2019).

30.   Jeppesen, D. K. et al. Reassessment of Exosome Composition. Cell 177, 428–445 (2019).

31.   Cvjetkovic, A. et al. Detailed Analysis of Protein Topology of Extracellular Vesicles-Evidence of Unconventional Membrane Protein Orientation. Sci Rep 6, 36338 (2016).

32.   Zeev-Ben-Mordehai, T., Vasishtan, D., Siebert, C. A., Whittle, C. & Grünewald, K. Extracellular Vesicles: A Platform for the Structure Determination of Membrane Proteins by Cryo-EM. Structure 22, 1687–1692 (2014).

33.   Parisse, P. et al. Atomic force microscopy analysis of extracellular vesicles. European Biophysics Journal 46, 813–820 (2017).

34.   Kim, S. Y., Khanal, D., Kalionis, B. & Chrzanowski, W. High-fidelity probing of the structure and heterogeneity of extracellular vesicles by resonance-enhanced atomic force microscopy infrared spectroscopy. Nat Protoc 14, 576–593 (2019).

35.   Lennon, K. M. et al. Single molecule characterization of individual extracellular vesicles from pancreatic cancer. Journal of Extracellular Vesicles 8, 1685634 (2019).

36.   Chen, Z. L. et al. A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. Nature Communications 10, 1–12 (2019).

37.   Chavez, J. D. et al. Chemical Crosslinking Mass Spectrometry Analysis of Protein Conformations and Supercomplexes in Heart Tissue. Cell Systems 6, 136–141 (2018).

38.   Gonzalez-Lozano, M. A. et al. Stitching the synapse: Cross-linking mass spectrometry into resolving synaptic protein interactions. Sci Adv 6, eaax5783 (2020).

39.   Armony, G., Heck, A. J. R. & Wu, W. Extracellular crosslinking mass spectrometry reveals HLA class I - HLA class II interactions on the cell surface. Mol Immunol 136, 16–25 (2021).

40.   Minn, A. J. et al. Genes that mediate breast cancer metastasis to lung. Nature 436, 518–524 (2005).

41.   Théry, C. et al. Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. J Extracell Vesicles 7, 1535750 (2018).

42.   Pankow, S., Bamberger, C., Calzolari, D., Bamberger, A. & Yates, J. R. Deep interactome profiling of membrane proteins by co-interacting protein identification technology. Nature Protocols 11, 2515–2528 (2016).

43.   Liu, F. & Heck, A. J. R. Interrogating the architecture of protein assemblies and protein interaction networks by cross-linking mass spectrometry. Current Opinion in Structural Biology 35, 100–108 (2015).

44.   Didiasova, M., Schaefer, L. & Wygrecka, M. When Place Matters: Shuttling of Enolase-1 Across Cellular Compartments. Front Cell Dev Biol 7, 61 (2019).

45.   Almaguel, F. A., Sanchez, T. W., Ortiz-Hernandez, G. L. & Casiano, C. A. Alpha-Enolase: Emerging Tumor-Associated Antigen, Cancer Biomarker, and Oncotherapeutic Target. Frontiers in Genetics 11, 28 (2020).

46.   Jiang, K. et al. Exosome-derived ENO1 regulates integrin α6β4 expression and promotes hepatocellular carcinoma growth and metastasis. Cell Death Dis 11, 972 (2020).

47.   Kang, H. J., Jung, S. K., Kim, S. J. & Chung, S. J. Structure of human alpha-enolase (hENO1), a multifunctional glycolytic enzyme. Acta Crystallogr D Biol Crystallogr 64, 651–657 (2008).

48.   Wang, X. et al. 14-3-3ζ delivered by hepatocellular carcinoma-derived exosomes impaired anti-tumor function of tumor-infiltrating T lymphocytes. Cell Death & Disease 9, 1–14 (2018).

49.   Li, Q. et al. Self-masking in an intact ERM-merlin protein: an active role for the central alpha-helical domain. J Mol Biol 365, 1446–1459 (2007).

50.   Phang, J. M. et al. Structural characterization suggests models for monomeric and dimeric forms of full-length ezrin. Biochemical Journal 473, 2763–2782 (2016).

51.   Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021).

3

52.   Huttlin, E. L. et al. Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. Cell 184, 3022–3040 (2021).

53.   Yang, X. H. et al. Disruption of laminin-integrin-CD151-focal adhesion kinase axis sensitizes breast cancer cells to ErbB2 antagonists. Cancer Res 70, 2256–2263 (2010).

54.   Brzozowski, J. S. et al. Extracellular vesicles with altered tetraspanin CD9 and CD151 levels confer increased prostate cell motility and invasion. Sci Rep 8, 8822 (2018).

55.   Zhu, J. et al. CD151 drives cancer progression depending on integrin α3β1 through EGFR signaling in non-small cell lung cancer. Journal of experimental & clinical cancer research 40, 192 (2021).

56.   Li, S. et al. Proteomic Landscape of Exosomes Reveals the Functional Contributions of CD151 in Triple-Negative Breast Cancer. Molecular & cellular proteomics 20, 100121 (2021).

57.   Waisman, D. M. & Waisman, D. M. Annexin II tetramer: structure and function. Molecular and Cellular Biochemistry 149, 301–322 (1995).

58.   Roesengarth A & Luecke H. Annexin A2: does it induce membrane aggregation by a new multimeric state of the protein. Annexins 1, 129 (2004).

59.   López-Rodríguez, J. C., Martínez-Carmona, F. J., Rodríguez-Crespo, I., Lizarbe, M. A. & Turnay, J. Molecular dissection of the membrane aggregation mechanisms induced by monomeric annexin A2. Biochimica et Biophysica Acta 1865, 863–873 (2018).

60.   Plaxco, K. W. & Gross, M. Protein Complexes: The Evolution of Symmetry. Current Biology 19, 25–26 (2009).

61.   van Zundert, G. C. P. & Bonvin, A. M. J. J. DisVis: quantifying and visualizing accessible interaction space of distance-restrained biomolecular complexes. Bioinformatics 31, 3222–3224 (2015).

62.   van Zundert, G. C. P. et al. The DisVis and PowerFit Web Servers: Explorative and Integrative Modeling of Biomolecular Complexes. J Mol Biol 429, 399–407 (2017).

63.   Honorato, R. v. et al. Structural Biology in the Clouds: The WeNMR-EOSC Ecosystem. Front Mol Biosci 8, 729513 (2021).

64.   Ecsédi, P. et al. Regulation of the Equilibrium between Closed and Open Conformations of Annexin A2 by N-Terminal Phosphorylation and S100A4-Binding. Structure 25, 1195–1207 (2017).

65.   Matos, A. L. L. et al. Membrane Binding Promotes Annexin A2 Oligomerization. Cells 9, 1169 (2020).

66.   Zhang, G. et al. Extracellular vesicles: Natural liver-accumulating drug delivery vehicles for the treatment of liver diseases. J Extracell Vesicles 10, e12030 (2020).

67.   Grossen, P. et al. Evaluation of bovine milk extracellular vesicles for the delivery of locked nucleic acid antisense oligonucleotides. European Journal of Pharmaceutics and Biopharmaceutics 158, 198–210 (2021).

68.   Rai, A., Fang, H., Claridge, B., Simpson, R. J. & Greening, D. W. Proteomic dissection of large extracellular vesicle surfaceome unravels interactive surface platform. Journal of Extracellular Vesicles 10, e12164 (2021).

69.   Galbraith, C. G. & Galbraith, J. A. Super-resolution microscopy at a glance. J Cell Sci 124, 1607–1611 (2011).

70.   Vizcaíno, J. A. et al. 2016 update of the PRIDE database and its related tools. Nucleic Acids Research 44, 447–456 (2016).

71.   Chen, Z. L. et al. A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. Nature Communications 10, 1–12 (2019).

72.   R Development Core Team, R. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing (2011).

73.   Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46, 296–303 (2018).

74.   Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60, 2126–2132 (2004).

75.   van Zundert, G. C. P. et al. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J Mol Biol 428, 720–725 (2016).

76.   Sinnott, M., Malhotra, S., Madhusudhan, M. S., Thalassinos, K. & Topf, M. Combining Information from Crosslinks and Monolinks in the Modeling of Protein Structures. Structure 28, 1061–1070 (2020).

77.   Bullock, J. M. A., Sen, N., Thalassinos, K. & Topf, M. Modeling Protein Complexes Using Restraints from Crosslinking Mass Spectrometry. Structure 26, 1015–1024 (2018).

78.   Vangone, A., Spinelli, R., Scarano, V., Cavallo, L. & Oliva, R. COCOMAPS: a web application to analyze and visualize contacts at the interface of biomolecular complexes. Bioinformatics 27, 2915–2916 (2011).

79.   Sievers, F. et al. Fast, scalable genera-
tion of high-quality protein multiple sequence align-
ments using Clustal Omega. Mol Syst Biol 7, 539
(2011).

80.   Ashkenazy, H. et al. ConSurf 2016: an
improved methodology to estimate and visualize
evolutionary conservation in macromolecules. Nu-
cleic Acids Res 44, 344–350 (2016).

81.   Tyanova, S., Temu, T. & Cox, J. The
MaxQuant computational platform for mass spec-
trometry-based shotgun proteomics. Nature Proto-
cols 11, 2301–2319 (2016).

3

CHAPTER 3 - SI

3

Supporting information for

Characterization of protein complexes in extracellular vesicles by intact extracellular vesicle crosslinking mass spectrometry (iEVXL)

Julia Bauzá-Martinez[1,2], Gad Armony[1,2], Matti F. Pronker[1,2], Wei Wu[1,2,3,4]*

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

[2] Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands

[3] Singapore Immunology Network (SIgN), Agency for Science, Technology and Research (A*STAR), 8A Biomedical Grove, Immunos, #04-06, Singapore 138648, Singapore.

[4] Department of Pharmacy, National University of Singapore, 14 Science Drive 4, Singapore 117543, Singapore.

Based on:

3

MDA-MB-231

WB: CD81

Supplementary Figure 1. Western blot validation of CD81 enrichment in extracellular vesicles when compared to source cells. This experiment was performed using the conditions described in the 'SDS-page and western blot' section of the materials and methods. EVs derived from MDA-MB-231 were chosen for this optimization, based on the data-driven observation that vesicles derived from these or LM2 cells have similar properties.

MDA-MB-231 EVs
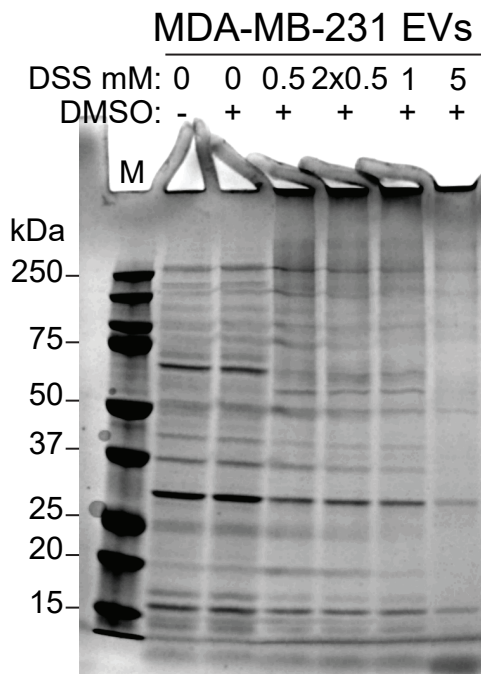
Supplementary Figure 2. Optimization of optimal DSS concentration for intact extracellular vesicle crosslinking. Vesicles were crosslinked with varying concentrations of disuccinimidyl suberate (DSS) ranging between 0.5 and 5 mM. Band shift indicating protein crosslinking, was evaluated in comparison to untreated extracellular vesicles (EVs). This experiment was performed using the conditions described in the 'EV crosslinking and protein digestion' and 'SDS-page and western blot' sections of the materials and methods. EVs derived from MDA-MB-231 were chosen for this optimization, under the assumption that vesicles derived from these or LM2 cells have similar properties.

3
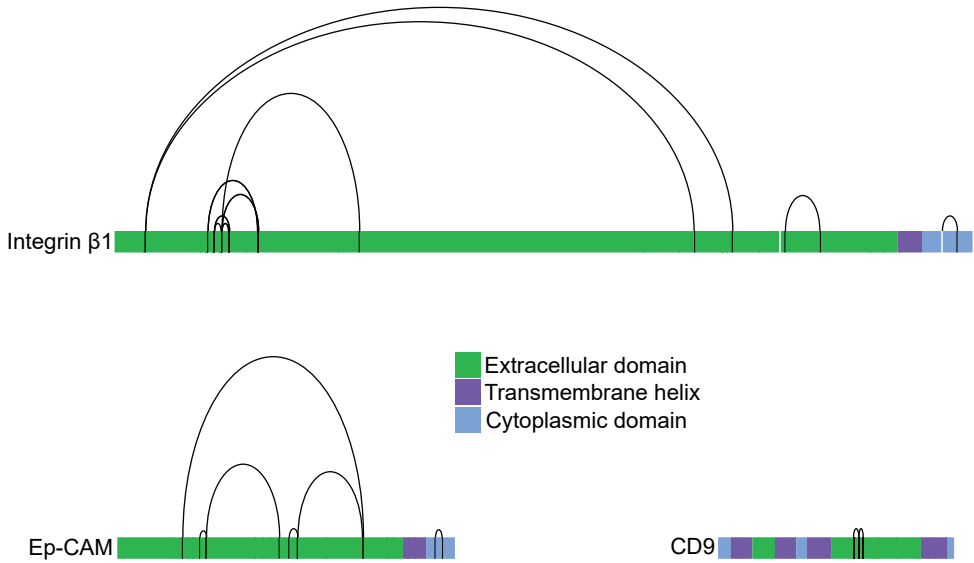
Supplementary Figure 3. Intralinks measured in our iEVXL dataset for three integral membrane proteins (integrin β1, Ep-CAM and CD9) are represented over the protein sequence, where the intracellular (blue), transmembrane (purple) and extracellular (green) domains have been annotated. Intralinks agree with the expected membrane topology, with the majority of crosslinks mapping to extracellular regions and no intralinks connecting the intracellular and extracellular domains.
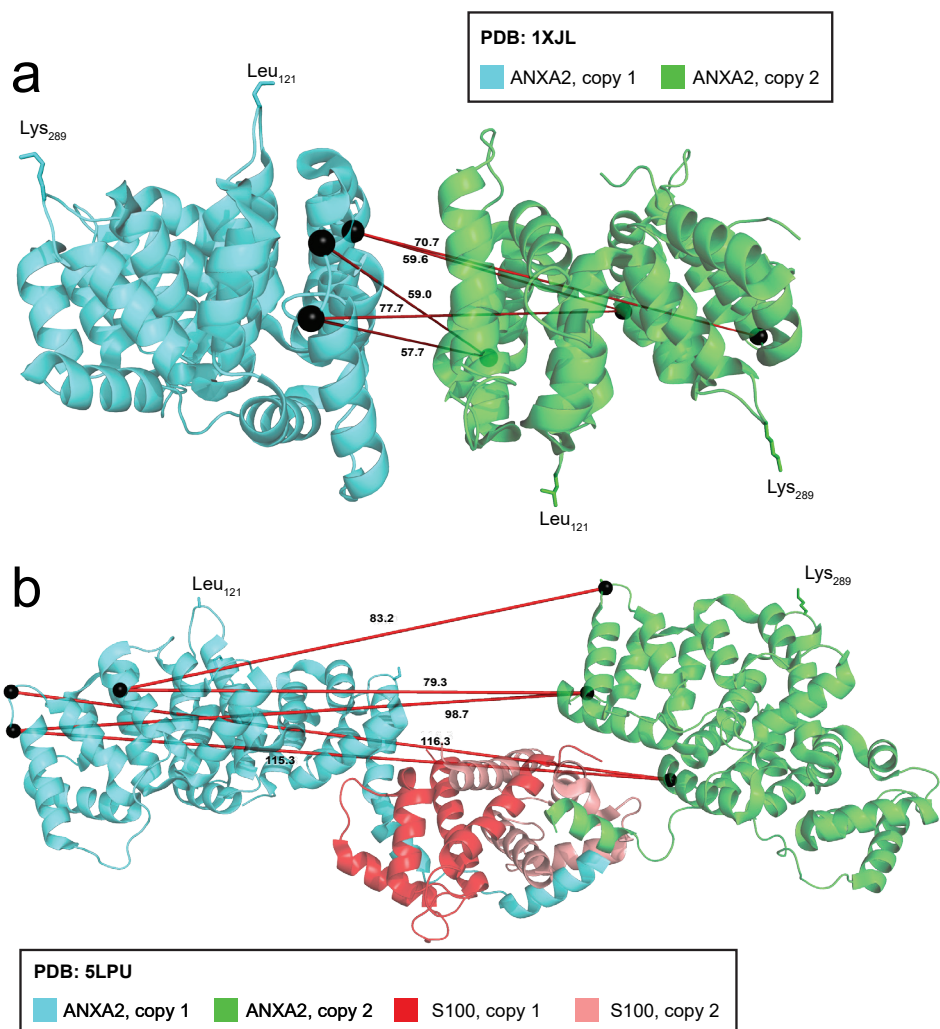
Supplementary Figure 4. Overlength crosslinks on annexin A2 (red lines in Figure 4B) found in LM2 are mapped onto (a) the crystal structure of the annexin A2 homodimer (PDB: 1xjl) and (b) the crystal structure of the annexin A2 – S100 heterotetramer (PDB: 5lpu). Euclidean distances (Å) between pairs of crosslinked lysine residues (black spheres) are annotated. All the crosslinks exceed the 30 Å Euclidean distance restraint for DSS.

a

b

c
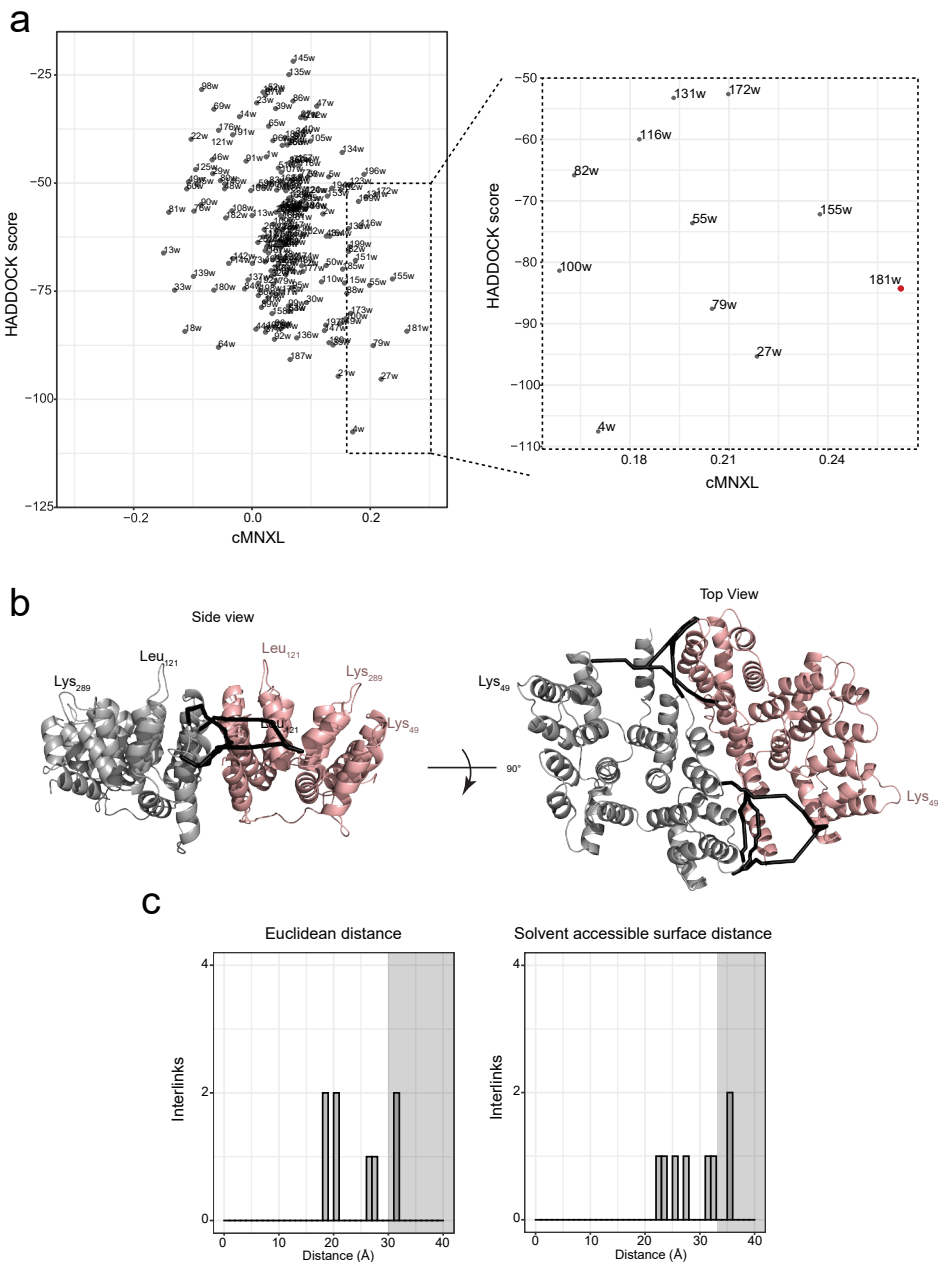
Supplementary Figure 5. (a) Correlation between Haddock and cMNXL scores for each modelled annexin A2 homodimer complex docked with HADDOCK 2.4. The region where the highest-scoring complexes were found (dashed lines) is displayed zoomed-in on the right. The selected model (complex_181w) is marked with a red dot. (b) The solvent accessible surface distances (SASDs) are represented by black lines and plotted onto the selected model (complex_181w) structure for annexin A2 homodimer. Side and top views are shown. (c) Histograms showing Euclidean distances (Å) and SASDs (Å) for the crosslinks shown in panel b.

102
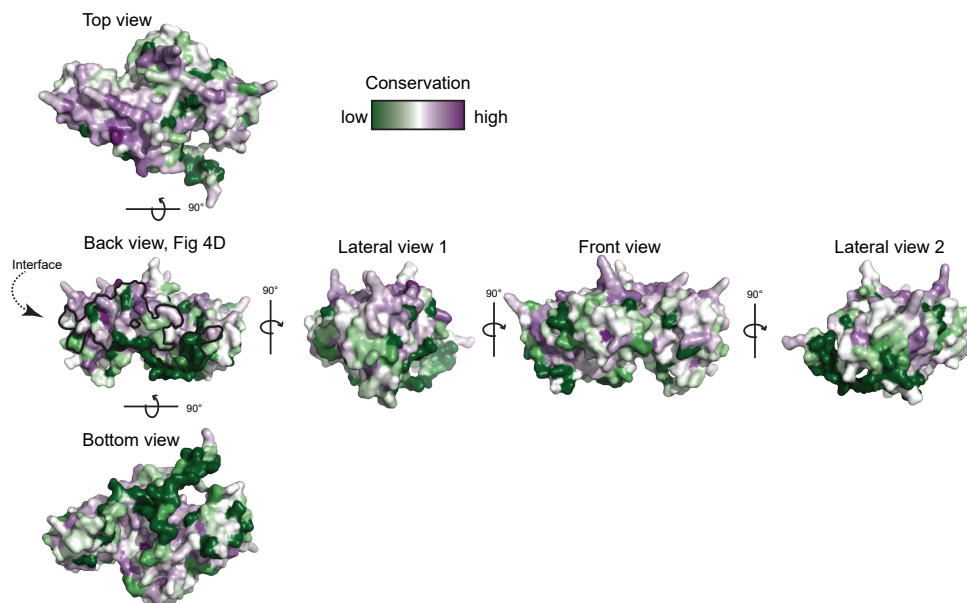
Supplementary Figure 6. Different rotational views of annexin A2 monomer conservation across all annexin A2 orthologs. The back view coincides with that of figure 4D in which the interface has been highlighted by a black line. The most conserved surface in annexin A2 is the plasma membrane interacting region (top view), followed by the predicted interface (back view) which is conserved relative to the surfaces shown in the rest of rotational views. Conserved residues are displayed in purple, while non-conserved residues are displayed in green.

CHAPTER 4

Cardiac progenitor cell-derived extracellular vesicles promote cell activation through both associated- and co-isolated proteins

Marieke Theodora Roefs[1], Julia Bauzá-Martinez[2], Jiabin Qin[1], Simonides Immanuel van de Wakker[1], Wilte Olijve[1], Robin Tuinte[1], Marjolein Rozeboom[1], Emma Alise Mol[1], Wei Wu[2,3,*], Pieter Vader[1,4,*], Joost Petrus Gerardus Sluijter[1,*]

[1] Department of Experimental Cardiology, University Medical Center Utrecht, Utrecht University, The Netherlands.

[2] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands.

[3] Singapore Immunology Network (SIgN), Agency for Science, Technology and Research (A*STAR), 8A Biomedical Grove, Immunos, #04-06, Singapore 138648, Singapore.

[4] CDL Research, University Medical Center Utrecht, The Netherlands

Summary

Extracellular vesicles (EVs) are small, cell-derived lipid bilayer enclosed particles that play a role in intercellular communication through delivery of their specific content. Cardiac progenitor cell (CPC)-derived EVs have been shown to protect the myocardium against ischemia-reperfusion injury via pro-angiogenic effects. However, underlying mechanisms remain elusive. Here, we aimed to investigate the protein-mediated effects of CPC-EVs on the endothelium and explored EV-dependent and -independent recipient cell activation. We discovered that the ability of CPC-EVs to induce endothelial cell (HMEC-1) migration and to stimulate pro-survival pathways including P13K-AKT-mTOR and (Insulin/IGF-) MAPK signaling cascades was lost upon EV donor cell exposure to calcium ionophore. This was corroborated by phosphoproteomic signatures identified from HMEC-1. MS-proteomic analysis identified proteome changes in the CPC-EVs, which composed abundantly of ECM proteins. To understand whether these proteins were truly associated to CPC-EVs, or merely co-isolated, we further purified size exclusion chromatography (SEC)-EV preparations using iodixanol gradient ultracentrifugation. These highly pure EVs lost part of their ability to activate HMEC-1 compared to crude SEC-EVs, suggesting a co-stimulatory role of co-isolated proteins in recipient cell activation. To study the contribution of individual EV-associated proteins to CPC-EV functionality, we employed the CRISPR/Cas9 technology. While knock-out of NID1 did not affect EV function, knock-out of PAPP-A resulted in a reduced potency to stimulate HMEC-1 activation and migration. The IGF-receptor inhibitor PPP abrogated CPC-EV-induced HMEC-1 activation, supporting the association of EV-associated PAPP-A with the activation of intracellular IGF-1-MAPK signaling. We identified a specific set of CPC-EV proteins, including PAPP-A, that contribute to endothelial cell activation. It is important to identify if these proteins are EV-associated or represent co-isolated factors that contribute to recipient cell activation. Our increased understanding of the mechanisms of CPC-EV-mediated cell activation will pave the way to more efficient EV-based therapeutics.

Introduction

Myocardial infarction causes massive loss of cardiomyocytes, resulting in scar for-mation and cardiac remodeling which lead to impaired cardiac function and pro-gressively into development of heart failure. Although heart failure cannot be pre-vented by currently available therapies, recent animal studies have demonstrated that cardiac function post-myocardial infarction may be improved by the therapeutic application of stem- and cardiac progenitor cell (CPC)-derived extracellular vesicles (EVs)[1,2].

EVs are cell-derived nanoparticles enclosed by a lipid bilayer that contain biological cargo including RNA, proteins and lipids and play a role in normal cellular homeo-stasis and intercellular communication[3]. EVs have the ability to activate target cells through the presence of adhesion molecules and receptors and via the delivery of bioactive molecules derived from the parent cell[2,4]. After in vivo administration, EVs released by Sca+ CPCs modulate regenerative processes in the heart by promoting angiogenesis, decreasing fibrosis and inhibiting cardiomyocyte apoptosis and there-by contributing to cardiac repair[5,6]. EVs derived from other stem cell sources have been shown to deliver distinct miRNAs and proteins to different cells in the heart to promote cardiac recovery[7–9]. Despite attempts to document CPC-EV composi-tion, there remains a lack of functional and mechanistic studies elucidating exactly which components in the complex protein repertoire are responsible for reparative function. Moreover, clinical application of stem cell-derived EVs is hampered by re-producibility issues related to differences in therapeutic activity between different EV isolations, among others[10,11]. We believe that deeper functional characterization of the CPC-EV content, and localization of this content in CPC-EV preparations is needed to get better insights into the mechanisms of action leading to a more re-producible therapeutic application of CPC-EVs. In this study, we set out to unravel the protein-mediated effects of CPC-EVs on endothelial cells. First, we identified the functional protein components of CPC-EVs involved in human microvascular endothelial cell (HMEC-1) activation by comparing the content of functional and non-functional (CPC-) EV preparations. Next, we investigated the contribution of EV-associated versus co-isolated proteins to CPC-EV function by employing iodix-anol gradient density-based purification. Lastly, we studied the contribution of in-dividual EV-associated proteins Pregnancy-associated plasma protein A (PAPP-A) and Nidogen-1 (NID1) to CPC-EV function by the generation of knock-out (KO) EVs employing CRISPR/Cas9 technology.

Results

CPC-EVs activate HMEC-1 and induce HMEC-1 migration

In the heart, pro-reparative effects evoked by the administration of CPCs and CPC-derived EVs were shown to be mediated by promoting angiogenesis[14–16]. To confirm the pro-angiogenic potential of our CPC-EVs *in vitro*, EVs were isolated from serum-free conditioned medium using size-exclusion chromatography (SEC) as described previously[13]. Successful isolation of CPC-EVs was confirmed by NTA

108
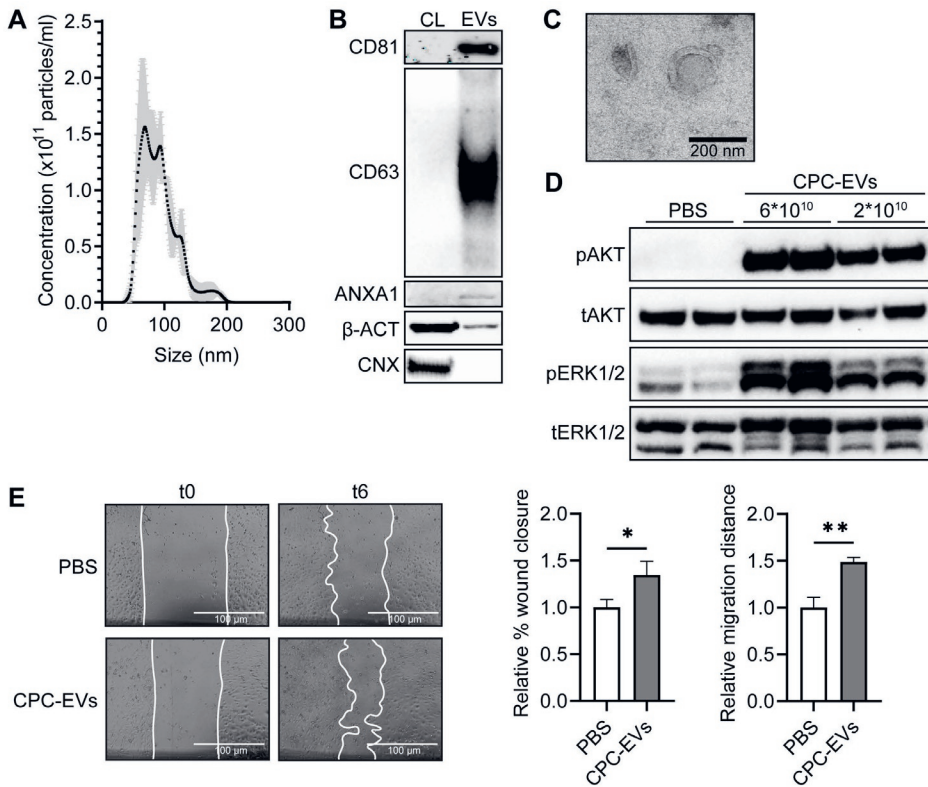
Figure 1. CPC-EVs activate intracellular signalling in HMEC-1 and induce HMEC-1 migration. (A) Representative NTA plot showing the size distribution and particle concentration of EVs isolated from the conditioned medium of CPCs (CPC-EVs) using size-exclusion chromatography. (B) Western blot analysis showing the presence of CD81, CD63, AnnexinA1 (ANXA1), β-actin (β-ACT), and absence of Calnexin (CNX) in CPC-EVs. β-ACT and CNX were present in CPC lysate (CL). (C) Representative TEM image of CPC-EVs. (D) Representative western blot analysis of phosphorylated AKT (pAKT), total AKT (tAKT), phosphorylated ERK1/2 (pERK1/2) and total ERK1/2 (tERK1/2) in HMEC-1 treated with two different CPC-EV particle doses. (E) Wound healing assay showing effects of CPC-EVs on HMEC-1 migration 6 hrs (t6) after EV addition (t0), analysed both as relative % wound closure and relative absolute migration distance compared to PBS (n=3). *= $p < 0.033$; **= $p < 0.0021$.

which demonstrated a size distribution with a peak at approximately 90 nm (**Figure 1A**). Western blot, according to MISEV2018 guidelines[17], confirmed the relative enrichment of EV marker proteins CD81, CD63, and Annexin A1 in EVs as compared to cell lysate and absence of endoplasmic reticulum protein Calnexin (**Figure 1B**). Transmission electron microscopy (TEM) of the CPC-EV preparation demonstrated the presence of vesicles containing a double leaflet membrane (**Figure 1C**).

Previous studies have demonstrated activation of the mitogen-activated protein kinase1/2 (MAPK1/2)-extracellular signal-regulated kinase1/2 (ERK1/2) and PI3 Kinase/AKT signalling pathways in HMEC-1 upon CPC-EV administration[13]. Indeed, levels of AKT and ERK1/2 phosphorylation increased within 30 min after addition of CPC-EVs in a dose-dependent manner (**Figure 1D**). The capacity of CPC-EVs to

functionally activate endothelial cells was determined in a HMEC-1 scratch wound assay. Addition of 2x10^{10} CPC-EVs to scratched HMEC-1 induced cell migration as determined by both percentage of closure and absolute migration distance compared to PBS (**Figure 1E**). This validated the pro-angiogenic potential of CPC-EVs *in vitro*.

Endothelial cell-activating capacity of CPC-EVs is lost after CPC exposure to calcium ionophore

Calcium ionophore A23187 has been hypothesized to induce EV release through complexing with calcium (Ca^{2+}), thereby inducing its transport across the cell membrane[18,19]. To stimulate EV release by CPCs, CPCs were either left untreated (0.0125% DMSO, veh-EVs) or exposed to 1 µM calcium ionophore (Ca ion-EVs) for 24 hrs. EVs were then isolated from serum-free conditioned medium using SEC. Size and total number of released EVs were not affected upon CPC exposure to calcium ionophore, as determined by NTA (**Figure 2A**). TEM confirmed the presence of round, membrane enclosed particles in both EV preparations (**Figure 2B**). Both EV isolates displayed EV marker proteins CD81, CD9 and ALIX, although their presence was reduced in Ca ion-EVs (**Figure 2C**). Ca ion-EVs had a slightly lower total protein content per 1x10^{10} particles compared to veh-EVs (**Figure 2D**). Therefore, in subsequent experiments, to prevent bias in interpreting EV functionality caused by differences in EV purity, EV supplementation was normalized between conditions both on total particle number and on total protein amount. Notably, veh-EV stimulation induced AKT and ERK1/2 phosphorylation in HMEC-1, while Ca ion-EV stimulation did not, both when EV addition was normalized based on particle number (**Figure 2E**) and on total protein levels (**Figure 2F**). EVs released from SKOV-3 cells were simultaneously isolated and characterized and served as a non-stem cell-control (**Suppl. Figure 1A,B**). SKOV-3-EVs were ineffective in inducing AKT and ERK1/2 phosphorylation as compared to veh-EVs (**Suppl. Figure 1C**). Furthermore, veh-EV stimulation, but not Ca ion-EV stimulation, induced HMEC-1 migration in a wound closure scratch assay (**Figure 2G**). These differences in EV functionality suggest that calcium ionophore treatment affects EV content. This finding was used to further explore the contributing signalling pathways involved in CPC-EV-mediated cell activation.

Phosphoproteome analysis identifies differences in HMEC-1 activation upon stimulation with veh-EV and Ca ion-EVs

To get insights into the mechanism of EV-mediated HMEC-1 activation, the signal transduction pathways triggered upon stimulation with veh-, Ca ion-EVs and negative control (PBS) were characterized by phosphoproteomics (**Suppl. Figure 2**). When comparing veh-EV- to PBS-treated HMEC-1 cells after 30 min of stimulation, significant changes were observed specifically at the phospho-proteome level, independent of proteome regulation (**Figure 3A**). The phospho-proteome analyses were highly sensitive, identifying ~6,900 class-I phosphosites (with phospho-modification localisation probability >0.75) from only 50 µg of HMEC-1 lysates. Hierarchical clustering of the significantly changing phosphosites revealed one cluster with increased
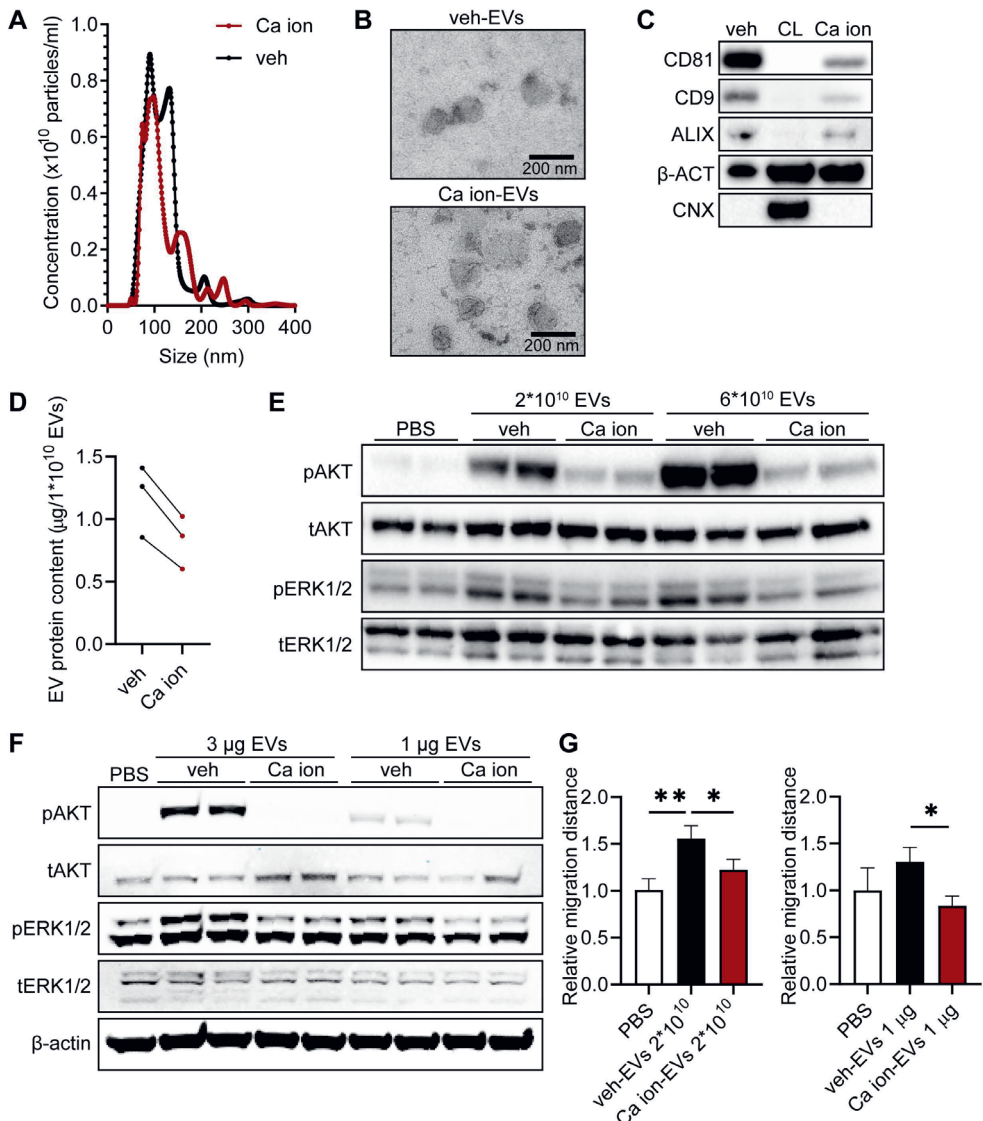
Figure 2. **EVs isolated from CPCs activate intracellular signalling in HMEC-1 and induce HMEC-1 migration, which is lost after CPC exposure to calcium ionophore.** (A) Representative NTA plot showing the size distribution and particle concentration of EVs isolated from the same volume of vehicle- (0.0125% DMSO; veh-EVs) or calcium ionophore- (Ca ion-EVs) stimulated CPCs. (B) Representative TEM images of veh- and Ca ion-EVs. (C) Western blot analysis showing the presence of CD81, CD9, ALIX, β-actin (β-ACT), and absence of Calnexin (CNX) in veh- and Ca ion-EVs. β-ACT and CNX were present in CPC lysate (CL). (D) Protein content per 1×10^{10} veh- and Ca ion-EVs of three biological triplicates. (E, F) Representative western blot analysis of phosphorylated AKT (pAKT), total AKT (tAKT), phosphorylated ERK1/2 (pERK1/2) and total ERK1/2 (tERK1/2) in HMEC-1 treated with veh- and Ca ion-EVs normalized on two doses of (E) EV particle numbers or (F) EV total protein content. β-ACT was included as housekeeping protein. (G) Wound healing assay showing effects of 2×10^{10} or 1 µg veh- and Ca ion-EVs on HMEC-1 migration (n=3). *= p < 0.033; **= p < 0.0021.

111
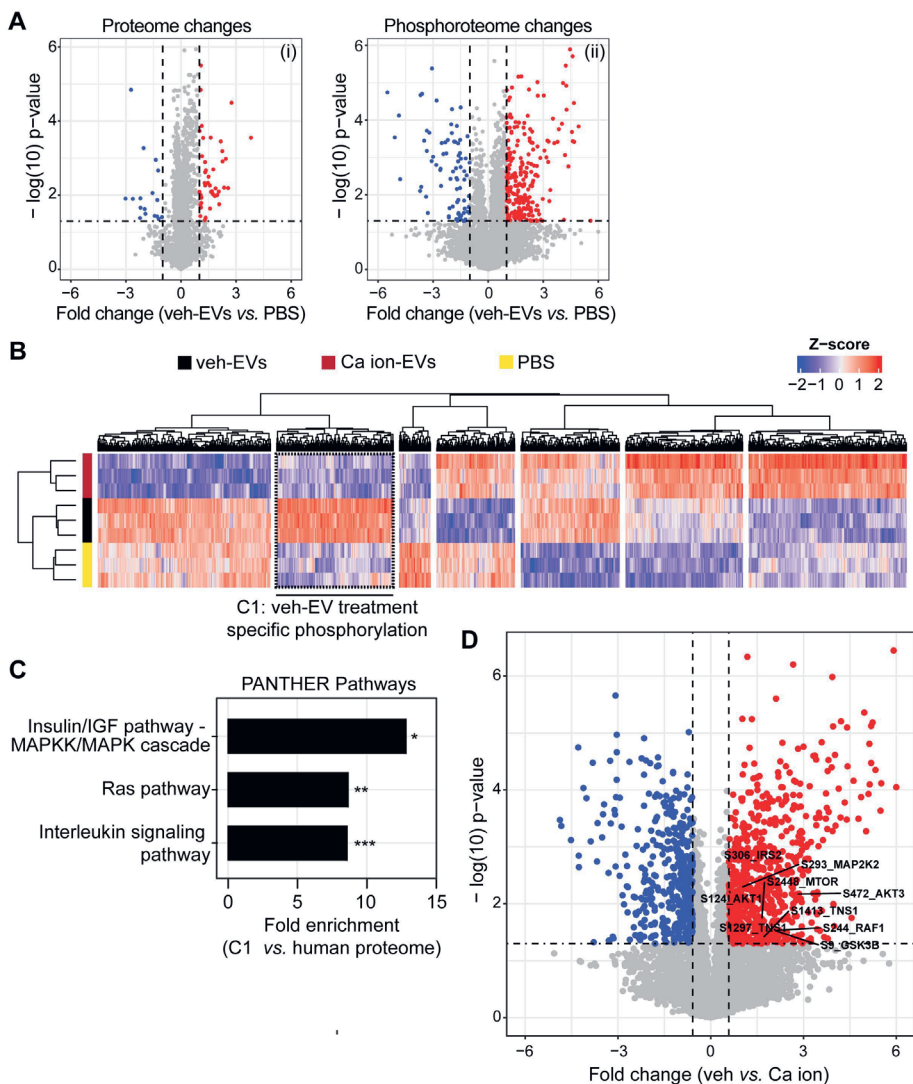
Figure 3. Phosphoproteomic analysis of HMEC-1 upon veh-EV- and Ca ion-EV stimulation. (A) Volcano plots showing changes in the (I) proteome and (ii) phosphoproteome of HMEC-1 after veh-EV stimulation compared to negative control (PBS). P-values were calculated using student's T-test, and significantly changing proteins (p-value ≤ 0.05 and fold change > 2) in veh-EV-treated HMEC-1 are highlighted in red, while significantly changing proteins in PBS are highlighted in blue. (B) Hierarchical clustering of 1,549 significantly changing phosphosites (ANOVA, q-value ≤ 0.05) found in HMEC-1 upon stimulation with veh-EVs, Ca ion-EVs and PBS. Cluster C1, including veh-EV-induced specific phosphorylation, is highlighted with dashed lines. (C) PANTHER Pathway enrichment analysis of phosphoproteins found in clusters C1, ranked on fold enrichment. *= FDR < 0.05, **= FDR < 0.01, ***= FDR < 0.005. (D) Volcano plot showing fold changes in the phosphoproteome of HMEC-1 upon veh-EV compared with Ca ion-EV stimulation. P-values were calculated using student's T-test, and significantly changing phosphosites (p-value < 0.05 and fold change > 2) after veh-EV treatment are highlighted in red, while significantly changing phosphosites after Ca ion-EV treatment are highlighted in blue.

phosphorylation in veh-EV-treated HMEC-1 compared to PBS- and Ca ion-EV-treated HMEC-1 (**Figure 3B**, dashed cluster C1). To better understand which intracellular signalling pathways were activated by veh-EVs, the phosphosites present in cluster C1 (n=195) were further annotated against PANTHER pathways. The Insulin/IGF pathway-MAPKK/MAPK cascade, Ras- and Interleukin signalling pathways were identified as the most enriched pathways in HMEC-1 (**Figure 3C**). Furthermore, significantly altered phosphosites between veh- and Ca ion-EV-stimulated HMEC-1 included members of the PI3K-AKT and MAPK signalling pathways (highlighted in **Figure 3D**, and represented in **Figure 7A**). This supports the loss in EV functionality after donor cell exposure to calcium ionophore and demonstrates the specific intracellular pathways implicated in CPC-EV-mediated HMEC-1 activation.

Proteinase K treatment decreases EV functionality on endothelial cells

CPC-EVs induce activation of intracellular signalling within 30 min after administration to HMEC-1 (**Figure 1D, 3A**), which suggests a role for direct receptor-ligand interactions of EVs with the target cell membrane. To determine the contribution of
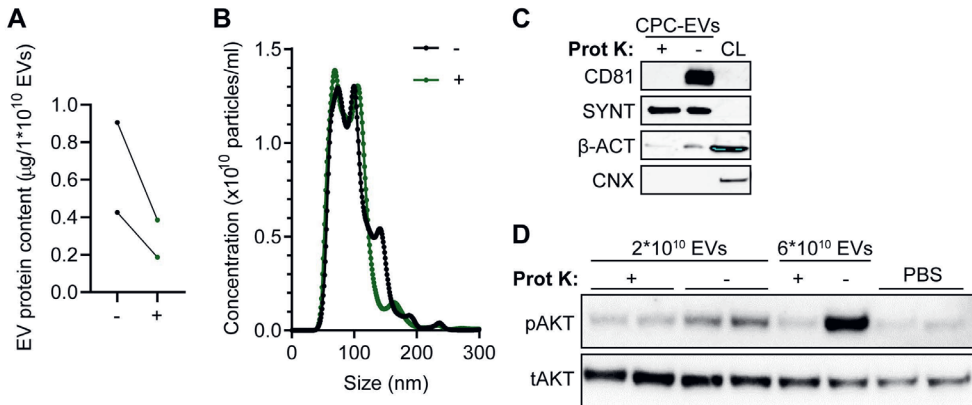
Figure 4. CPC-EVs lose HMEC-1 activating capacity after proteinase K treatment. (A) Protein content per 1×10^{10} EVs treated with (+) and without (-) Proteinase K (Prot K) of two representative experiments. (B) Representative NTA plot showing the size distribution and particle concentration of both EV populations after a second SEC isolation. (C) Western blot analysis showing the absence and presence of CD81 in Prot K-treated (+) and untreated (-) EVs, respectively. Syntenin-1 (SYNT) and β-actin (β-ACT) were present in both EV populations, while absent for Calnexin (CNX). β-ACT and CNX were present in CPC lysate (CL). (D) Representative western blot analysis of phosphorylated AKT (pAKT) and total AKT (tAKT) in HMEC-1 stimulated with Prot K- (+) and un-treated (-) CPC-EVs normalized on two doses of EV particle numbers.

EV-associated proteins to CPC-EV functionality, CPC-EV surface proteins and the extracellular receptor domains of transmembrane proteins were removed via treatment with 100 µg/mL Proteinase K. Proteinase K-treated EVs and untreated EVs subsequently underwent a second SEC separation to remove free protein fragments from EVs. A reduction in protein levels in proteinase-K treated EVs was verified with microBCA analysis (**Figure 4A**), and NTA demonstrated that EV size was not affect-

ed by proteinase K treatment (**Figure 4B**). Western blot analyses confirmed the removal of EV-surface protein CD81 in proteinase K-treated EVs, while intravesicular protein Syntenin-1 was unaffected (**Figure 4C**). Proteinase K treated-EVs were less potent to induce phosphorylation of AKT in HMEC-1 as compared to untreated EVs (**Figure 4D**), which confirmed a major contribution of CPC-EV-associated proteins to HMEC-1 activation.

CPC-EVs contain proteins involved in endothelial cell migration

Since veh-, Ca ion- and SKOV-3-EVs differed in functionality, we next sought to pinpoint EV proteins that might be responsible for the differential ability to stimulate signaling in recipient cells. Using an unbiased label-free proteomics strategy, we compared quantitatively the total EV proteome between the three sources. From this 2,077 proteins were confidently identified (1% FDR) amongst all EV populations. From these, 1,293 proteins were detected in at least 2 out of the 3 biological replicates for each EV population (**Figure 5A**), of which we further explored the 209 proteins that were significantly altered between veh-EVs and Ca ion-EVs (**Figure 5Bi**), and 146 proteins between veh-EVs and SKOV-3-EVs (**Figure 5Bii**). The enrichment of PAPP-A, Tumor necrosis factor-stimulated gene-6 (TSG-6) and Laminin subunit gamma 1 (LAMC1) in veh-EVs as compared to Ca ion-EVs was validated by western blotting (**Figure 5C**).

Further analysis of the significantly enriched proteins in veh-EVs revealed 105 proteins consistently enriched in veh-EVs compared to both the Ca ion- and SKOV-3-EV populations (top 20 in **Suppl. Table 2**), while 104 and 41 proteins were enriched exclusively when compared to Ca ion-EVs and SKOV-3-EVs, respectively (**Figure 5D**). The full list of veh-EV enriched proteins (n=105) was further annotated by Gene Ontology, which revealed an over-representation of biological processes such as "biological adhesion", "cell adhesion" and "regulation of cell migration" (**Figure 5E**). Collectively, these confirm that the effects of CPC-EVs on endothelial cells observed may be mediated through a collective subset of EV proteins that promote endothelial cell migration.

Both EV-associated factors and co-isolated proteins contribute to CPC-EV-mediated cell activation

Since ECM proteins were enriched in veh-EVs (**Suppl. Table 2**), we attempted to dissect if these proteins were packaged in EVs or co-isolated in our preparations. To this end, Optiprep™ density gradient centrifugation after SEC was employed to further separate EVs from co-isolated proteins (**Figure 6A**). Western blot analysis of the different fractions confirmed the presence of EV-markers CD81, Syntenin-1 and CD63 mainly in fractions 2-4 (**Figure 6B**). Total particle numbers in each fraction were measured using NTA and showed presence of particles in fractions 2-4 although particles were also present in the top fraction (F1) (**Figure 6B**). EV-containing fractions 1-5 (Opti-EVs) and co-isolated protein-containing fractions 7 and 8 (Opti-protein), as showed by silver stain (**Suppl. Figure 3A**), were collected and concentrated using 100- or 10 kDa cut-off spin filters. Opti-EVs and SEC-EVs showed similar size distribution (**Figure 6C**), while Opti-EVs contained less protein content

per particle compared with SEC-EVs, suggesting higher purity as expected (**Figure 6D**). Western blotting demonstrated an increased relative expression of CD81 and Syntenin-1, and absence of Calnexin in Opti-EVs compared with SEC-EVs, again confirming higher purity (**Figure 6E**). As we found that iodixanol increased EVs' potency to induce AKT- and ERK1/2 phosphorylation and wound closure (**Suppl. Figure 3B,C**), iodixanol concentrations in all samples were kept constant (2%) in all functional assays to exclude the direct functional effect of iodixanol. In an endothelial activation assay, Opti-EVs induced AKT and ERK1/2 phosphorylation, although to a lesser extent than SEC-EVs (**Figure 6F**). The Opti-protein fraction also demonstrated limited HMEC-1 stimulating capacity. In a HMEC-1 scratch assay, both Opti-EVs and the Opti-protein fraction displayed some but reduced migration-stimulating capacities compared with SEC-EVs (**Figure 6G**). These results suggest that both EVs and co-isolated proteins present in the crude SEC-EV preparations can contribute to endothelial cell activation.

Extracellular vesicle-associated PAPP-A, but not NID1, is involved in endothelial cell activation

Phosphoproteome analysis in HMEC-1 upon veh-EV stimulation revealed the specific activation of the Insulin/IGF pathway-MAPKK/MAPK cascade (**Figure 3C**) and identified the phosphorylation of members of the MAPK and AKT-mTOR signalling cascades, including IRS2, RAF1, MEK2, AKT and mTOR (**Figure 7A**). To study the contribution of the IGF receptor (IGF-R) to intracellular signalling upon EV stimulation, we employed the IGF-R inhibitor Picropodophyllin (PPP). PPP dose-dependently abrogated the increase in ERK1/2 and AKT phosphorylation in HMEC-1 upon CPC-EV stimulation (**Figure 7B**), demonstrating a contribution of the IGF-R signalling pathway in HMEC-1 activation. MS-proteomic analysis identified PAPP-A and NID1 among the 20 most strongly enriched proteins in veh-EVs compared to Ca ion- and SKOV-3-EVs (**Figure 5B, Suppl. Figure 4A, Suppl. Table 2**). A proposed mechanism of PAPP-A function is cleavage of IGFBP4 from the IGF-1/IGFBP4 complex, inducing the release of IGF-1 in the extracellular space and subsequent activation of intracellular signalling by binding to IGF-R (**Figure 7A**)[7,20]. Enriched PAPP-A expression on veh-EVs as compared to CPC lysate and Ca ion-EVs was confirmed by western blotting (**Figure 5C**). NID1 is a glycoprotein present in the basement membrane, reported to promote cell migration by being present on EVs derived from hepatocellular carcinoma cells[21]. Interestingly, western blotting confirmed NID1 expression in veh-EVs, but also showed the presence of a higher molecular weight NID1 form in Ca ion-EVs (**Figure 5C**). This could be due to differences in NID1 glycosylation status and the bias towards detection of non-glycosylated peptides by our MS-analysis.

To investigate the influence of PAPP-A and NID1 on EV-mediated endothelial cell activation, CRISPR/Cas9 technology was employed to knock-out (KO) PAPP-A and NID1 in CPCs. Three different gRNAs targeting exon 3 and 4 of *PAPPA* and targeting the 5'UTR region and exon 1 of *NID1* were selected to identify the most appropriate target site in *PAPPA* or *NID1* for gene editing, respectively, and their genomic DNA cleavage efficiency was determined using the T7 endonuclease assay (**Suppl. Figure 4B, 5A**). Based on most potent DNA cleavage efficiency, we selected gRNA 1
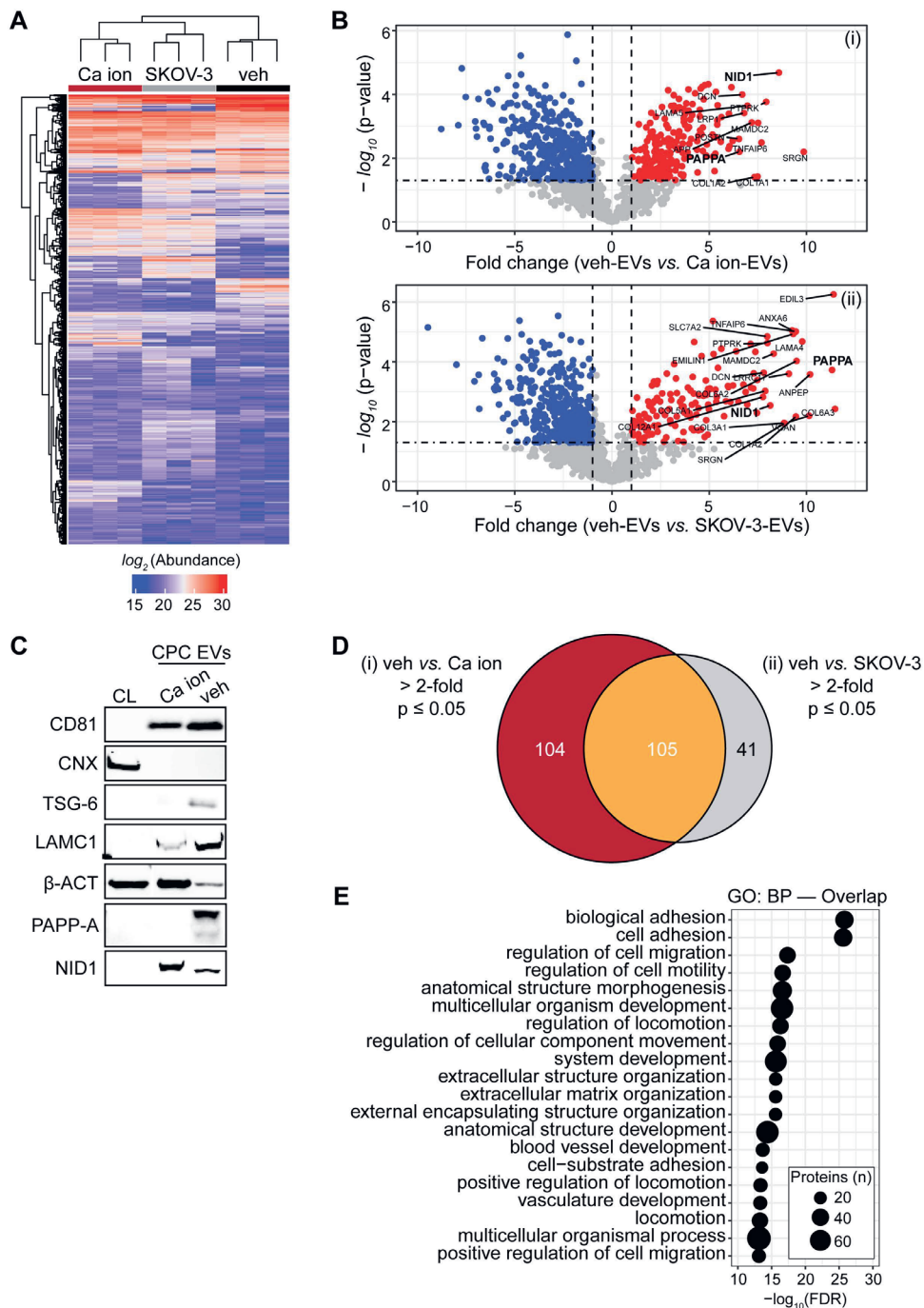
Figure 5. LC-MS/MS identified enriched proteins in veh-EVs compared with Ca ion- and SKOV-3-EVs. (Caption on next page)

Figure 5. LC-MS/MS identified enriched proteins in veh-EVs compared with Ca ion- and SKOV-3-EVs. (A) Heat map of protein abundance (log_2) of proteins identified in each biological replicate (veh-, Ca ion- and SKOV-3-EVs), as identified by LC-MS/MS. (B) Volcano plots showing average fold changes for protein abundance (log_2) of proteins identified in veh-EVs compared to (I) Ca ion-EVs and (ii) SKOV-3-EVs. P-values were calculated using student's T-test, and significantly changing proteins (p-value ≤ 0.05 and fold change > 2) in veh-EVs are highlighted in red, while significantly changing proteins in Ca ion- and SKOV-3-EVs are highlighted in blue. (C) Western blot analysis confirming the enrichment of MS-identified proteins NID1, TSG-6, LAMC1, PAPP-A, CD81 and β-actin (β-ACT) in veh-EVs compared with Ca ion-EVs. CNX was solely present in CPC lysate (CL). Complete blots of β-ACT, PAPP-A and NID1 are displayed in Suppl. Figure 8. (D) Venn diagram showing number of proteins with > 2-fold significant enrichment (p ≤ 0.05) in veh-EVs compared to Ca ion- and SKOV-3-EVs, and overlap between those two populations. (E) Gene ontology analysis using PANTHER of enriched biological processes for the 105 overlapping proteins, depicting number of identified proteins in each group, ranked on smallest corrected p-value ($-log_{10}$(FDR)).

and 2 for PAPP-A and gRNA 3 for NID1 for the generation of KO CPC lines. Sanger sequencing revealed a mixed population of mutated alleles at the gRNA target sites (**Suppl. Figure 4C, 5B**). Individual clones were generated from these CPC KO lines by clonal expansion and the presence of homozygous biallelic deletions or insertions was determined by Sanger sequencing (**Suppl. Figure 4D, 5C**). One CPC clone harbouring a homozygous insertion of 1 bp in *PAPPA* and one clone harbouring a deletion of 8 bp in *NID1* were subsequently selected for the isolation of PAPP-A-depleted (PAPPA KO) and NID1-depleted (NID KO) EVs and used for further functional studies (**Figure 7C, Suppl. Figure 5D**). A polyclonal CPC line transduced with the CRISPR/Cas9 machinery, and a non-targeting gRNA (NTgRNA) served as normal CPC-EV control. PAPP-A depletion in isolated PAPPA KO-EVs was confirmed by western blotting (**Figure 7D**). PAPP-A depletion resulted in reduced cell growth, but did not influence cell morphology (**Suppl. Figure 4E,F**) and EV production as determined by western blotting identified similar levels of CD81, Syntenin-1, Flotillin and β-actin in both the PAPPA KO- and NTgRNA-EVs (**Figure 7D**). Expression of CD63 seemed slightly reduced in the PAPPA KO-EVs. Both EV populations were similar in size, as determined by NTA (**Figure 7E**), and in purity (µg protein/$1x10^{10}$ EVs) (**Figure 7F**). Although NID1 KO-CPCs also had a slightly slower growth rate compared to NTgRNA CPCs (**Suppl. Figure 5E**) and displayed a slightly more elongated morphology (**Suppl. Figure 5F**), EV-release was not affected as determined by NTA and by the expression of EV markers CD81, Syntenin-1, and levels of β-actin (**Suppl. Figure 5G,H**). NID1 KO-EVs contained a slightly higher protein amount per particle (µg protein/$1x10^{10}$ EVs) (**Suppl. Figure 5I**).

Next, we evaluated the influence of PAPP-A and NID1 KO on EV-stimulating capacity. Compared to NTgRNA-EVs, PAPPA KO-EVs were less potent in inducing AKT- and ERK1/2 phosphorylation in HMEC-1 while NID1 KO did not affect EV-stimulating activity, when EV addition was normalized on total protein content or total particle number (**Figure 7G,H, Suppl. Figure 6A**). In a HMEC-1 scratch assay, PAPPA KO-EVs were also less potent in inducing wound closure, calculated both as percentage wound closure as absolute migration distance, when normalized for total particle number or total protein content (**Figure 7i**). In contrast, there was no difference in activity between NID1 KO- and NTgRNA-EVs in inducing wound closure (**Suppl.**
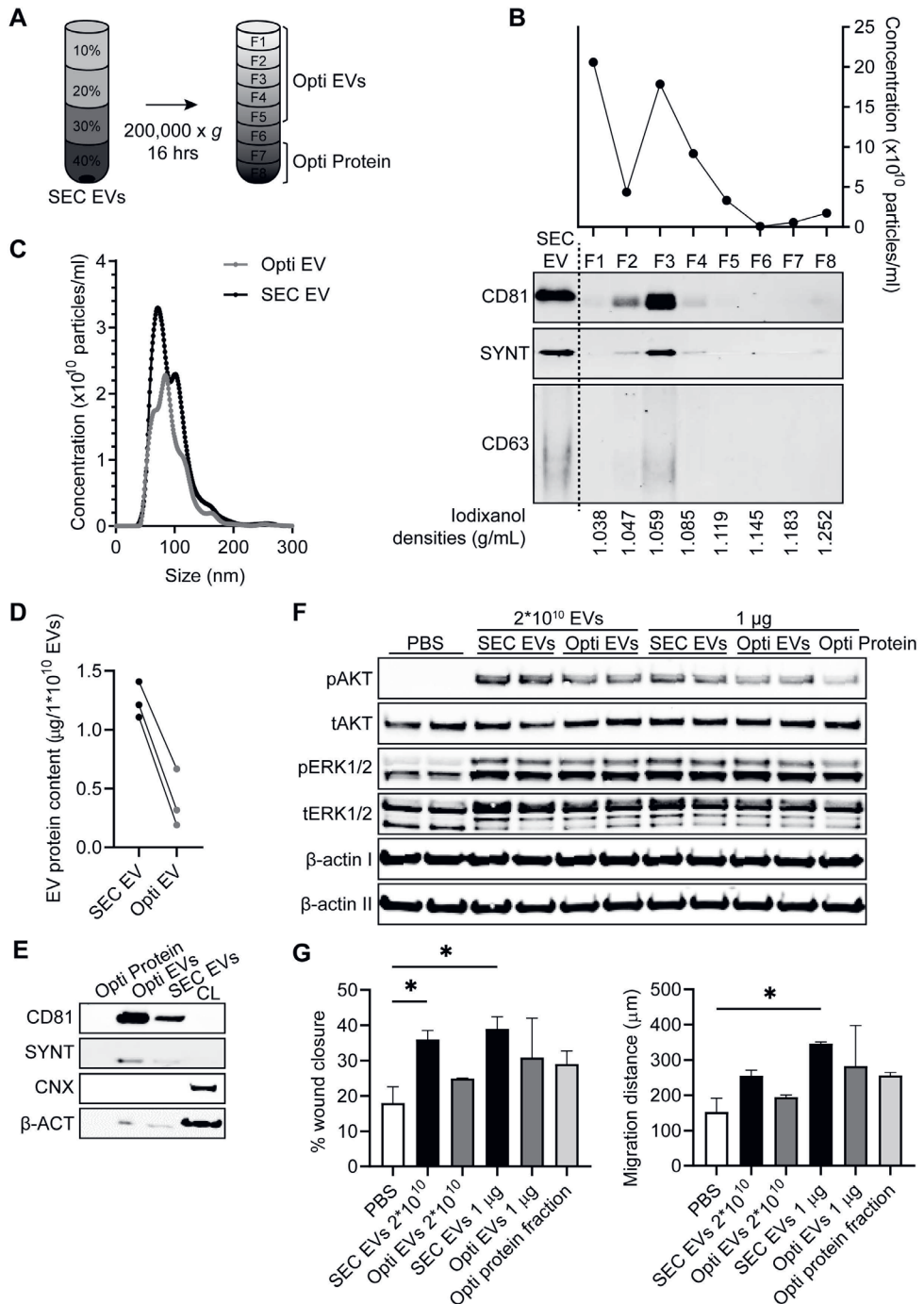
Figure 6. Both co-isolated proteins and EV-associated factors contribute to HMEC-1 activation. (Caption on next page)

Figure 6. Both co-isolated proteins and EV-associated factors contribute to HMEC-1 activation. (A) Schematic overview of the EV isolation protocol to obtain pure Opti-EVs. SEC-EVs were loaded in the bottom of a discontinuous Optiprep™ gradient and ultracentrifuged for 16 hrs. (B) Resulting fractions (F1-8) were analysed for particle number by NTA (upper panel) and the presence of EV-marker proteins CD81, CD63 and Syntenin-1 (SYNT) by western blotting. Equal volumes of each sample were analysed. Iodixanol concentration was measured in each fraction (lower panel). Fractions 1-5 were pooled and concentrated (Opti-EVs). (C) Representative NTA plot showing the size distribution and particle concentration of SEC- and purified Opti-EVs. (D) Protein content per 1×10^{10} SEC- and Opti-EVs of three representative experiments. (E) Western blot analysis showing presence of CD81, SYNT, β-actin (β-ACT) in purified Opti-EV- and protein fractions, compared with crude SEC-EVs. Calnexin (CNX) was only present in CPC lysate (CL). Complete β-ACT blot is displayed in Suppl. Figure 8. (F) Representative western blot analysis of phosphorylated AKT (pAKT), total AKT (tAKT), phosphorylated ERK1/2 (pERK1/2) and total ERK1/2 (tERK1/2) in HMEC-1 treated with SEC-EVs and Opti-EVs normalized on total particle numbers, and with SEC-EVs, Opti-EVs and Opti-protein fraction normalized on total protein content. β-ACT was included as housekeeping protein (I=phosphorylated protein blot, II=total protein blot). (G) Wound healing experiment showing effects SEC- and Opti-EVs on HMEC-1 migration normalized both on total particle number and total protein content, analysed both as % wound closure and absolute migration distance. Addition of Opti-protein fraction was normalized on volume (n=3, technical replicates. Data are representative of three independent experiments). *= p < 0.033.

Figure 6B). NID1 has been hypothesized to be involved in cardiac repair by binding to the αVβ integrin receptor and thereby activating downstream MAPK signalling[22]. However, EV-induced AKT and ERK1/2 phosphorylation and wound closure was not influenced by the pre-treatment of an IntegrinαVβ3 antibody (**Suppl. Figure 7A,B**). In addition, recombinant NID1 did not induce AKT and ERK1/2 phosphorylation nor was able to induce wound closure in the HMEC-1 scratch assay. This implies that EV-associated NID1 is not contributing to CPC-EV functionality in our *in vitro* assays. Overall, our data suggest that PAPP-A is enriched on CPC-EVs and is involved in EV-mediated endothelial cell activation via the IGF-R signalling pathway.

Discussion

CPC-EV administration has been shown to improve cardiac function by promoting different cellular processes including angiogenesis in several preclinical animal models[1]. In line with previous studies, we confirmed that CPC-EVs are potent inducers of endothelial cell activation and functionality[5,6]. Despite promising preclinical results of EV therapeutics for tissue regeneration, translation into the clinics is still hampered by *in vivo* reproducibility due to EV heterogeneity and poor understanding of functional mediators in EVs[10,11]. In order to advance CPC-EVs as therapeutics, more insights into functional EV cargo and mechanisms of EV-mediated cell activation are required. In this work, we investigated the pro-angiogenic composition of CPC-EVs.

To increase EV release by CPCs we explored the use of calcium ionophore A23187, a compound described to stimulate the release of EVs through the increase of intracellular Ca^{2+} levels[18,19,23,24]. Interestingly, CPC exposure to calcium ionophore did not increase EV release, but did result in a less functional EV population as shown in
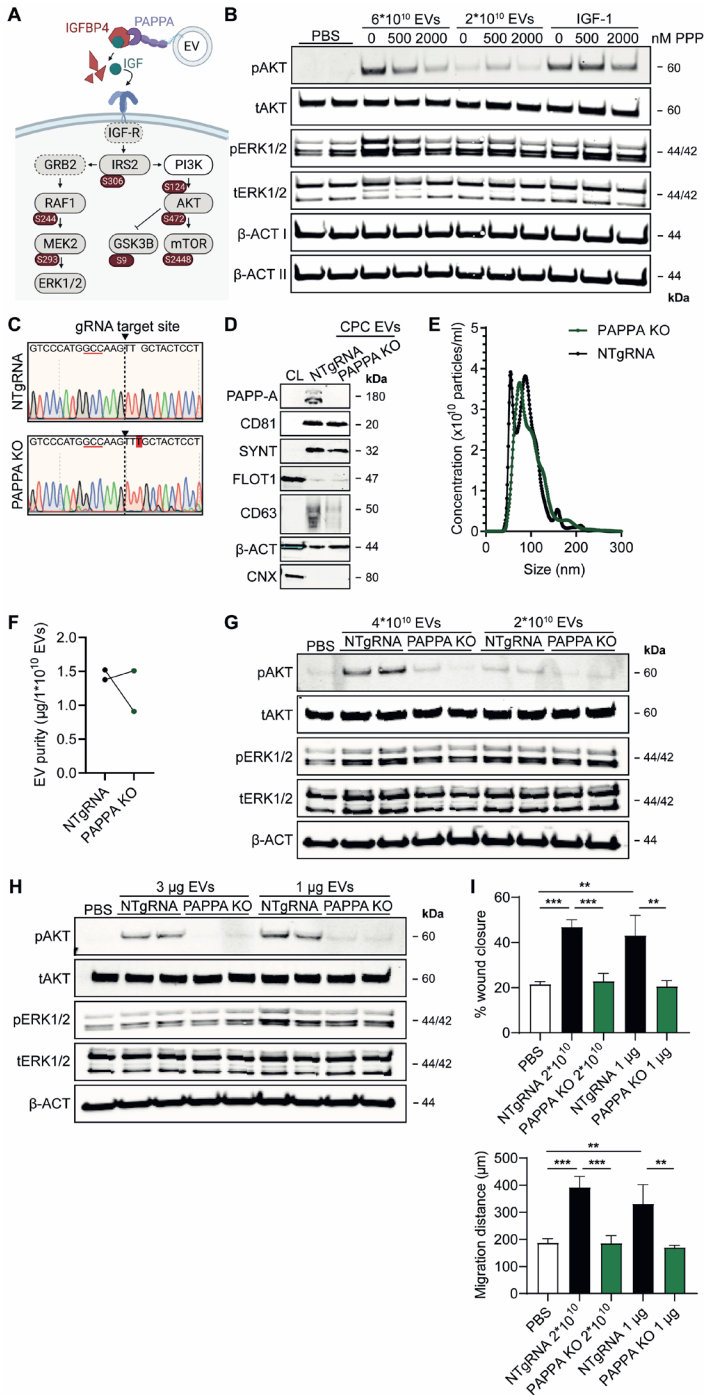
Figure 7. PAPPA KO-EVs were less potent to activate intracellular signalling in HMEC-1 and to induce HMEC-1 migration. (Caption on next page)

Figure 7. PAPPA KO-EVs were less potent to activate intracellular signalling in HMEC-1 and to induce HMEC-1 migration. (A) Schematic depicting hypothesized mechanism of (intra)cellular signalling activated by EV-associated PAPP-A, based on identified proteins and significantly altered phosphosites measured in HMEC-1 upon veh-EV stimulation by (phopho) proteomic analysis. Detected proteins in HMEC-1 are displayed in grey, while significantly changing phosphosites present in cluster C1 (see Figure 3D) are displayed in brown. (B) Representative western blot analysis of phosphorylated AKT (pAKT), total AKT (tAKT), phosphorylated ERK1/2 (pERK1/2) and total ERK1/2 (tERK1/2) in HMEC-1 treated with $6x10^{10}$ or $2x10^{10}$ CPC-EVs, or with 200 ng/mL free IGF-1 after pre-incubation with different doses of picropodophyllin (PPP). β-actin (β-ACT) was included as housekeeping protein (I=phosphorylated protein blot, II=total protein blot). (C) Sanger sequencing results confirming 1 bp insertion in exon 3 of PAPPA at the CRISPR/Cas9 target site of the PAPPA KO-CPC clone, compared with the NTgRNA polyclonal CPC line. (D) Western blot analysis showing the absence of PAPP-A in PAPPA KO-EVs, compared with NTgRNA-EVs; the presence of CD81, CD63, Syntenin-1 (SYNT), Flotillin (FLOT1), β-ACT, and absence of Calnexin (CNX) in both EV populations. FLOT1, β-ACT and CNX were present in CPC lysate (CL). (E) Representative NTA plot showing the size distribution and particle concentration of PAPPA KO- and NTgRNA-CPC-EVs. (F) Protein content per $1x10^{10}$ PAPPA KO- and NTgRNA-EVs of two representative experiments. (G,H) Representative western blot analyses of pAKT, tAKT, pERK1/2 and tERK1/2 in HMEC-1 treated with PAPPA KO- and NTgRNA-EVs normalized on two doses of (G) total particle numbers or (H) total protein content. β-ACT was included as housekeeping protein. (I) Wound healing assay showing effects of 1 µg and $2x10^{10}$ NTgRNA- and PAPPA KO-EVs on HMEC-1 migration, analysed both as % wound closure and absolute migration distance (n=3, technical replicates. Data are representative of three independent experiments). **= $p < 0.0021$, ***= $p < 0.0002$.

different HMEC-1 stimulation assays. We employed the differences in functionality between CPC veh- and Ca ion-EVs to investigate the specific activation of intracellular signalling by employing phosphoproteomic analysis of HMEC-1 upon stimulation with both EV populations. This is one of the first studies investigating the activation of intracellular signaling in recipient cells by means of phosphoproteomics upon EV stimulation, which could identify the increased phosphorylation of members of the P13K-AKT and (Insulin/IGF-) MAPK signaling pathways. We observed significant changes at the phosphoproteome level but not at the proteome level within 30 min after HMEC-1 stimulation. In addition, we demonstrated that the ability of EVs to activate intracellular signaling pathways was lost after the removal of EV-surface proteins using proteinase K, as previously shown for EVs derived from other sources[25]. This hints towards a fast signal transduction via receptor-ligand interactions instead of EV-content delivery through EV internalization. Indeed, recent studies demonstrate that EVs contain only a very limited number of specific miRNAs[26–30], corrobaorating on the importance of investigating the contribution of other EV content such as proteins to EV functionality.

Several groups have previously characterized the proteomic composition of stem- and progenitor cell-derived EVs, which is currently available via several databases[31–33]. Identification of functional components, however, was done mainly by comparisons of EVs derived from different cell types[7,34] or after differences upon cell preconditioning[35,36]. Here, we compared the proteomic composition of functionally distinct EV types, released from the same cell type and donor, thereby eliminating potential differences in content caused by differences in isolation technique and bio-

logical variability of cell type and donor. MS profiling identified a multitude of proteins differentially expressed in veh-EVs relative to Ca ion-EVs. These veh-EV enriched proteins were over-represented in PANTHER gene ontology biological process 'regulation of cell migration' (top 3), which demonstrated our approach as a method for unbiased identification of proteins potentially involved in EV-induced cell migration.

Although initial studies identified EVs as functional components of the stem cell secretome, more recent studies have hypothesized that also soluble co-isolated factors contribute to observed EV-mediated therapeutic effects, which might even be synergistic[37–41]. These observations suggest that released therapeutic stimuli are likely not to be mediated exclusively by stem cell-derived EVs. We observed that many of identified proteins enriched in veh-EVs were known components of the ECM. This raised the question whether these proteins were truly associated to EVs or just present as a result of co-isolation. Although other studies have attributed EV-mediated function to members of the ECM by being present in EV preparations[7,21,42], the influence of co-isolated proteins in functional CPC-EV preparations, isolated by SEC, has not been investigated yet. SEC has been adopted as an preferable isolation method as compared to ultracentrifugation due to a higher yield and scalability[13,43]. Here, we show that despite SEC isolation that separates CPC-EVs from proteins based on size differences, the EV preparation still contains co-isolated proteins or other factors that contribute to EV-function. Indeed, both co-isolated factors and the further purified EVs upon iodixanol density gradient centrifugation displayed some pro-angiogenic properties, as also demonstrated by others[44–48]. We hypothesize a contribution of both EV-associated factors and co-isolated proteins to CPC-EV function, which might potentially be synergistic.

MS-proteomic analysis identified candidate proteins PAPP-A and NID1 enriched in CPC-EVs. Employing CRISPR/Cas9, we set out to identify their contribution to EV function. KO of NID1 did not influence HMEC-1 stimulation capacities of CPC-EVs. In contrast, depletion of PAPP-A from CPC-EVs yielded EVs with reduced capacity to induce endothelial cell activation. The presence of PAPP-A on CPC-EVs and its involvement in cardioprotection was previously demonstrated by Barile et al[7]. PAPP-A was present on CPC-derived EVs and abrogated EV-mediated protection against staurosporine-induced cell death in HL-1 cells. As loss in protection was evoked only after addition of the IGF-1/IGFBP4 complex, where EV-bound PAPP-A is suggested to cleave the IGF-1/IGFBP4 complex and thereby leads to IGF-1 release and subsequent target cell activation[20]. Based on measured changes at the phosphoproteome level, we could identify the Insulin/IGF-MAPK pathway as most enriched pathway, providing another link between PAPP-A and the activation of IGF-1-MAPK signalling. This is further supported by our observation that the IGF-R inhibitor PPP abrogated CPC-EV-induced HMEC-1 activation.

EVs released from different stem cell types were reported to have similar pro-regenerative properties *in vitro* and *in vivo*[49], but also differences in functionality have been discovered between different organs and disease states, as reflected by differences in identified functional EV cargo[9,10]. EVs derived from different (stem) cell donors might be enriched in different functional proteomic cargo, as the presence of both PAPP-A and NID1 in EVs has been reported before[33,50], but not reproducibly[42]. The

proteomic profiles of different stem cell-derived EVs might largely differ due to EV heterogeneity caused by different methods used for EV isolation, as for example SEC is likely to better remove overabundant soluble proteins compared with other techniques[43], but also by the variability between cell donors and culture state[2]. Here, we investigated the contribution of individual CPC-EV-associated proteins to stimulate pro-angiogenic processes *in vitro*, which clearly contribute but might not explain all pro-regenerative properties of (CPC-)EVs *in vivo*. Whether PAPP-A contributes to EV-regenerative properties in other contexts remains to be investigated.

In conclusion, we demonstrate that CPC-EVs can promote endothelial cell activation and cellular migration and identified the proteomic composition of functional CPC-EVs that potentially contribute to these processes. Furthermore, we characterized for the first time downstream signaling upon CPC-EV stimulation in endothelial cells. We demonstrated that both EV-associated factors and co-isolated proteins in the CPC-EV preparation contributed to endothelial cell activation. Using CRISPR/Cas9 to knock-out individual proteins from EVs, we identified PAPP-A but not NID1 to contribute to CPC-EV function. These results support the idea that specific proteins present in SEC-EV preparations can selectively induce specific signals in recipient cells to regulate processes such as angiogenesis. This knowledge of the pro-angiogenic components and localization of these components in CPC-EV preparations can be further applied in future studies exploring the use of EVs for therapeutic application.

Methods

Cell culture and EV isolation

Cardiac progenitor cells (CPCs, donor HFH070809) were obtained from human fetal hearts as described before[12]. Human fetal heart tissue was obtained by individual permission using standard written informed consent and after approval of the ethics committee of Leiden University Medical Center, The Netherlands. This is according to the principles outlined in the Declaration of Helsinki for the use of human subjects or tissue. CPCs and human microvascular endothelial cells (HMEC-1) were cultured as reported before[13] (see Supplemental Material on Supplementary Information). EV-conditioned medium was prepared by culturing CPCs for 3 days on SP++ medium until 80% confluency was reached, followed by 24 hrs of culture on basal FBS- and supplement-free M199 medium containing 1 µM calcium ionophore A23187 (Sigma) or vehicle (0.0125% DMSO). SKOV-3 cells at 80% confluency were cultured in DMEM without any additives for 24 hrs to obtain conditioned medium. EVs were isolated from conditioned medium employing ultrafiltration and subsequent size-exclusion chromatography (SEC) (see Supplemental Material on Supplementary Information). To obtain more pure EVs, SEC-EVs were further purified employing iodixanol density gradient utracentrifugation (see Supplemental Material on Supplementary Information).

Proteinase K treatment

EVs were incubated in a final concentration of 100 µg/mL Proteinase K (Promega) for 30 min at 37 °C. Proteinase K was inactivated by diluting the EV sample to 1 mL in PBS supplemented with protease inhibitor (Roche), and EVs were subsequently separated from fractionated proteins by SEC using a Sepharose CL-4B column connected to an ÄKTA start system (GE Healthcare) containing an UV 280 nm flow cell. An EV sample without treatment with Proteinase K, but with the subsequent isolation and concentrating steps, was taken along in parallel and served as untreated control. EV-containing fractions were pooled and again concentrated using a 100-kDa MWCO Amicon Ultra-4 spin filter (Merck Millipore).

Nanoparticle tracking analysis

NTA was performed using a Nanosight NS500 system (Malvern Technologies). EVs were diluted in PBS and three videos of 30 sec were captured using a camera level of 15 and a detection threshold of 5. Size and particle concentration was determined with the Nanosight NTA 3.3 software (Malvern Technologies).

EV protein determination and western blot

Total protein concentrations of EV samples were determined using the Pierce microBCA Protein Assay Kit (Thermo Fisher Scientific) according to manufacturer's protocol after lysis in 1x RIPA buffer (Abcam). EV proteins were analyzed with western blot (see Supplementary Material on Supplementary Information).

HMEC-1 stimulation

140,000 HMEC-1 were plated in a 48-wells plate one day before stimulation. 3 hrs before stimulation, HMEC-1 were starved using basal MCDB131 medium. $6x10^{10}$, $2x10^{10}$, 3 µg or 1 µg of EVs were added in duplicate for 30 min to stimulate HMEC-1. PBS was supplemented as negative control. For inhibition experiments, different doses (100-2,000 nM) of picropodophyllin (PPP; Calbiochem) were added 3 hrs, or 10 µg/mL mouse anti-IntegrinαVβ3 antibody (Novus Biologicals) was added 1 hr before EV addition into the basal medium. Stimulation with 200 ng/mL recombinant IGF-1 (PromoCell) or 1 µg/mL Nidogen-1 (R&D systems) was included as positive control. For phophoproteomic analysis, 260,000 HMEC-1 were plated in a 24-wells plate one day before stimulation with $10x10^{10}$ EVs or PBS. Cells were lysed using cOmplete Lysis-M EDTA-free lysis buffer (Roche) supplemented with PhosSTOP phosphatase inhibitors (Roche) and centrifuged at 14,000 x*g* for 10 min at 4 °C to remove cellular debris. Supernatants were used for further western blotting experiments and (phospho)proteomic analysis as described below.

Western blot for pAKT and pERK1/2 expression

Total protein content of HMEC-1 lysates was determined using a Pierce microBCA Protein Assay Kit (Thermo Fisher Scientific) according to manufacturer's instructions, and proteins were analyzed with western blot (see Supplementary Material on Supplementary Information).

Endothelial wound closure assay

A scratch was made in a monolayer of HMEC-1 in a 48-wells plate and any floating cells were removed. Medium was replaced to basal MCDB131 with or without $2x10^{10}$ or 1 µg EVs or equal volumes of PBS. MCDB131 medium supplemented with 20% FBS was included as positive control. For inhibition experiments, 10 µg/mL mouse anti-IntegrinαVβ3 antibody (Novus Biologicals) was added simultaneously with EV addition. Stimulation with 1 µg/mL Nidogen-1 (R&D systems) was included as positive control. At 0 and 6 hrs, bright field pictures were taken using an EVOS microscope (Life Technologies). Percentage closure and absolute migration distance was determined after 6 hrs. Relative wound closure was calculated relative to the PBS control.

CRISPR/Cas9 plasmid construction and stable CPC line generation

Single guide RNAs (sgRNAs) specific for human PAPP-A and NID1 were designed using the CRISPOR design tool (http://crispor.tefor.net/) and were screened for the Homo Sapiens genome to minimize potential off-target effects. All sgRNA sequences and PCR primers are listed in Suppl. Table 1. sgRNA expressing plasmids were achieved by cloning of the sgRNA in the LentiCRISPv2 vector (Addgene Plasmid number 52961) (see Supplemental Material on Supplementary Information). To generate polyclonal stable CPC lines, CPCs were infected with CIRSPv2-expressing lentivirusses, and individual knock-out (KO) clones were obtained by serial diluting polyclonal CPC lines (see Supplemental Material on Supplementary Information).

Genomic DNA extraction and T7 endonuclease assay

Genomic DNA was extracted from stable polyclonal CPCs using the GeneJet Genomic DNA Purification kit (Thermo Scientific) following the recommended protocol. The genomic region flanking the CRISPR/Cas9 target site was amplified by PCR and subjected to a T7 endonuclease assay (see Supplemental Material on Supplementary Information).

LC-MS/MS analysis of EVs

Biological triplicates of isolated EVs were stored at -80 °C and protein composition was analysed by mass spectrometry (MS) using a using a Dionex UltiMate™ 3000 RSLCnano System coupled to a Q-Exactive mass spectrometer (Thermo Scientific) (see Supplementary Material on Supplementary Information).

LC-MS/MS analysis of EV-stimulated HMEC-1 (phospho)proteome

Total protein content of stimulated HMEC-1 lysates was determined using a microBCA Protein Assay Kit (Thermo Fisher Scientific) according to manufacturer's instructions. Two biological replicates were pooled to obtain 50 µg protein for subsequent phospho-enrichment analysis using an Ultimate 3000 system (Thermo Fischer Scientific) coupled to an Orbitrap Exploris 480 mass spectrometer (Thermo Fischer Scientific) (see Supplementary Material on Supplementary Information).

Transmission Electron Microscopy (TEM)

Concentrated EVs were adsorbed to carbon-coated formvar grids for 15 min at RT. After a PBS wash, the grids were fixed in a 1% glutaraldehyde in PBS fixing buffer for 30 min at RT, followed by counterstaining with uranyl-oxalate. Grids were embedded in a mixture of 1.8% methyl cellulose and 0.4% uranyl acetate at 4 °C and imaged on a Jeol JEM-1011 TEM microscope (Jeol).

Statistics

Statistical analyses of scratch assay results were performed using Prism 5.0 (Graph-Pad Software Inc.). The statistical difference between two groups was analyzed using an unpaired Student's T-test. Differences between more than two groups were tested with a one-way ANOVA followed by Tukey's HSD multiple comparison test as post-test. Differences with two-tailed p-values < 0.05 were considered statistically significant. All results are expressed as mean ± standard deviation.

Funding

Author contribution statement

P.V., J.S., W.W., M.T.R. and J.B. conceived and planned the experiments. M.T.R., J.B., S.W., J.Q., W.O., R.T., M.R. and E.M. carried out the experiments. M.T.R. wrote the manuscript with support from J.B., W.W., P.V. and J.S. All authors discussed the results and contributed to the final manuscript.

Data availability

The (phospho)proteomic datasets generated and analysed during the current study are available in the ProteomeXchange Consortium via the PRIDE repository, and can be accessed through the identifier PXD030779.

Acknowledgements

Conflict of interest statement

The authors declare to have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Yang L, Zhu J, Zhang C, Wang J, Yue F, Jia X, Liu H. Stem cell-derived extracellular vesicles for myocardial infarction: a meta-analysis of controlled animal studies. Aging 2019;11:1129–1150.

2. Sluijter JPG, Davidson SM, Boulanger CM, Buzás EI, Kleijn DPV de, Engel FB, Giricz Z, Hausenloy DJ, Kishore R, Lecour S, Leor J, Madonna R, Perrino C, Prunier F, Sahoo S, Schiffelers RM, Schulz R, Laake LW Van, Ytrehus K, Ferdinandy P. Extracellular vesicles in diagnostics and therapy of the ischaemic heart: Position Paper from the Working Group on Cellular Biology of the Heart of the European Society of Cardiology. Cardiovascular Research 2018;114:19–34.

3. Jeppesen DK, Fenix AM, Franklin JL, Higginbotham JN, Zhang Q, Zimmerman LJ, Liebler DC, Ping J, Liu Q, Evans R, Fissell WH, Patton JG, Rome LH, Burnette DT, Coffey RJ. Reassessment of exosome composition. Cell 2019;177:428-445.e18.

4. Niel G van, D'Angelo G, Raposo G. Shedding light on the cell biology of extracellular vesicles. Nature Reviews Molecular Cell Biology 2018;19:213–228.

5. Maring JA, Lodder K, Mol E, Verhage V, Wiesmeijer KC, Dingenouts CKE, Moerkamp AT, Deddens JC, Vader P, Smits AM, Sluijter JPG, Goumans M-J. Cardiac Progenitor Cell–Derived Extracellular Vesicles Reduce Infarct Size and Associate with Increased Cardiovascular Cell Proliferation. Journal of Cardiovascular Translational Research 2019;12:5–17.

6. Mol EA, Lei Z, Roefs MT, Bakker MH, Goumans M-J, Doevendans PA, Dankers PYW, Vader P, Sluijter JPG. Injectable Supramolecular Ureidopyrimidinone Hydrogels Provide Sustained Release of Extracellular Vesicle Therapeutics. Advanced Healthcare Materials 2019;8:1900847.

7. Barile L, Cervio E, Lionetti V, Milano G, Ciullo A, Biemmi V, Bolis S, Altomare C, Matteucci M, Silvestre D Di, Brambilla F, Fertig TE, Torre T, Demertzis S, Mauri P, Moccetti T, Vassalli G. Cardioprotection by cardiac progenitor cell-secreted exosomes: role of pregnancy-associated plasma protein-A. Cardiovascular Research 2018;114:992–1005.

8. Yuan Y, Du W, Liu J, Ma W, Zhang L, Du Z, Cai B. Stem Cell-Derived Exosome in Cardiovascular Diseases: Macro Roles of Micro Particles. Frontiers in Pharmacology 2018;9:547.

9. Roefs MT, Sluijter JPG, Vader P. Extracellular Vesicle-Associated Proteins in Tissue Repair. Trends in Cell Biology 2020;30:990–1013.

10. Varderidou-Minasian S, Lorenowicz MJ. Mesenchymal stromal/stem cell-derived extracellular vesicles in tissue repair: challenges and opportunities. Theranostics 2020;10:5979–5997.

11. Abreu RC de, Fernandes H, Costa Martins PA da, Sahoo S, Emanueli C, Ferreira L. Native and bioengineered extracellular vesicles for cardiovascular therapeutics. Nature Reviews Cardiology 2020:1–13.

12. Smits AM, Vliet P van, Metz CH, Korfage T, Sluijter JPG, Doevendans PA, Goumans MJ. Human cardiomyocyte progenitor cells differentiate into functional mature cardiomyocytes: An in vitro model for studying human cardiac physiology and pathophysiology. Nature Protocols 2009;4:232–243.

13. Mol EA, Goumans M-J, Doevendans PA, Sluijter JPG, Vader P. Higher functionality of extracellular vesicles isolated using size-exclusion chromatography compared to ultracentrifugation. Nanomedicine : nanotechnology, biology, and medicine 2017;13:2061–2065.

14. Bertero A, Murry CE. Hallmarks of cardiac regeneration. Nature Reviews Cardiology.

15. Vrijsen KR, Maring JA, Chamuleau SAJ, Verhage V, Mol EA, Deddens JC, Metz CHG, Lodder K, Eeuwijk ECM van, Dommelen SM van, Doevendans PA, Smits AM, Goumans M-J, Sluijter JPG. Exosomes from cardiomyocyte progenitor cells and mesenchymal stem cells stimulate angiogenesis via EMMPRIN. Advanced healthcare materials 2016;5:2555–2565.

16. Smits AM, Laake LW Van, Ouden K Den, Schreurs C, Szuhai K, Echteld CJ Van, Mummery CL, Doevendans PA, Goumans MJ. Human cardiomyocyte progenitor cell transplantation preserves

4

long-term function of the infarcted mouse myocardium. Cardiovascular Research 2009;83:527–535.

17. Théry C, Witwer KW, Aikawa E, Alcaraz MJ, Anderson JD, Andriantsitohaina R, Antoniou A,. Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. Journal of Extracellular Vesicles 2018;7:1535750.

18. Savina A, Furlán M, Vidal M, Colombo MI. Exosome Release Is Regulated by a Calcium-dependent Mechanism in K562 Cells. Journal of Biological Chemistry 2003;278:20083–20090.

19. Krämer-Albers E-M, Bretz N, Tenzer S, Winterstein C, Möbius W, Berger H, Nave K-A, Schild H, Trotter J. Oligodendrocytes secrete exosomes containing major myelin and stress-protective proteins: Trophic support for axons? PROTEOMICS – Clinical Applications 2007;1:1446–1461.

20. Oxvig C. The role of PAPP-A in the IGF system: location, location, location. Journal of Cell Communication and Signaling 2015;9:177–187.

21. Mao X, Tey SK, Yeung CLS, Kwong EML, Fung YME, Chung CYS, Mak L, Wong DKH, Yuen M, Ho JCM, Pang H, Wong MP, Leung CO, Lee TKW, Ma V, Cho WC, Cao P, Xu X, Gao Y, Yam JWP. Nidogen 1-enriched extracellular vesicles facilitate extrahepatic metastasis of liver cancer by activating pulmonary fibroblasts to secrete tumor necrosis factor receptor 1. Advanced Science 2020;7:2002157.

22. Zbinden A, Layland SL, Urbanczyk M, Berrio DAC, Marzi J, Zauner M, Hammerschmidt A, Brauchle EM, Sudrow K, Fink S, Templin M, Liebscher S, Klein G, Deb A, Duffy GP, Crooks GM, Eble JA, Mikkola HKA, Nsair A, Seifert M, Schenke-Layland K. Nidogen-1 Mitigates Ischemia and Promotes Tissue Survival and Regeneration. Advanced Science 2021;8:2002500.

23. Prudent M, Crettaz D, Delobel J, Seghatchian J, Tissot J-D, Lion N. Differences between calcium-stimulated and storage-induced erythrocyte-derived microvesicles. Transfusion and apheresis science : official journal of the World

Apheresis Association : official journal of the European Society for Haemapheresis 2015;53:153–158.

24. French SL, Butov KR, Allaeys I, Canas J, Morad G, Davenport P, Laroche A, Trubina NM, Italiano JE, Moses MA, Sola-Visner M, Boilard E, Panteleev MA, Machlus KR. Platelet-derived extracellular vesicles infiltrate and modify the bone marrow during inflammation. Blood Advances 2020;4:3011–3023.

25. Crain SK, Robinson SR, Thane KE, Davis AM, Meola DM, Barton BA, Yang VK, Hoffman AM. Extracellular vesicles from Wharton's Jelly mesenchymal stem cells suppress CD4 expressing T cells through transforming growth factor beta and adenosine signaling in a canine model. Stem Cells and Development 2019;28:212–226.

26. Reshke R, Taylor JA, Savard A, Guo H, Rhym LH, Kowalski PS, Trung MT, Campbell C, Little W, Anderson DG, Gibbings D. Reduction of the therapeutic dose of silencing RNA by packaging it in extracellular vesicles via a pre-microRNA backbone. Nature Biomedical Engineering 2020;4:52–68.

27. Toh WS, Lai RC, Zhang B, Lim SK. MSC exosome works through a protein-based mechanism of action. Biochemical Society Transactions 2018;46:843–853.

28. Chevillet JR, Kang Q, Ruf IK, Briggs HA, Vojtech LN, Hughes SM, Cheng HH, Arroyo JD, Meredith EK, Gallichotte EN, Pogosova-Agadjanyan EL, Morrissey C, Stirewalt DL, Hladik F, Yu EY, Higano CS, Tewari M. Quantitative and stoichiometric analysis of the microRNA content of exosomes. Proceedings of the National Academy of Sciences of the United States of America 2014;111:14888–14893.

29. Stevanato L, Thanabalasundaram L, Vysokov N, Sinden JD. Investigation of content, stoichiometry and transfer of miRNA from human neural stem cell line derived exosomes. PLoS ONE 2016;11.

30. Albanese MI, Adam Chen Y-FI, Hü ls C, Gä rtner KI, TagawaID T, Mejias-PerezID E, KepplerID OT, Gö bel C, ZeidlerID R, Shein M, Schü tzID AK, Hammerschmid-tID W. MicroRNAs are minor constituents of extracellular vesicles that are rarely

delivered to target cells. PLOS Genetics 2021;17:e1009951.

31. Keerthikumar S, Chisanga D, Ariyaratne D, Saffar H Al, Anand S, Zhao K, Samuel M, Pathan M, Jois M, Chilamkurti N, Gangoda L, Mathivanan S. ExoCarta: A Web-Based Compendium of Exosomal Cargo. Journal of Molecular Biology 2016;428:688–692.

32. Kim DK, Lee J, Kim SR, Choi DS, Yoon YJ, Kim JH, Go G, Nhung D, Hong K, Jang SC, Kim SH, Gho YS. EVpedia: A community web portal for extracellular vesicles research. Bioinformatics 2015;31:933–939.

33. Pathan M, Fonseka P, Chitti S V., Kang T, Sanwlani R, Deun J Van, Hendrix A, Mathivanan S. Vesiclepedia 2019: A compendium of RNA, proteins, lipids and metabolites in extracellular vesicles. Nucleic Acids Research 2019;47:D516–D519.

34. Haraszti RA, Didiot M-C, Sapp E, Leszyk J, Shaffer SA, Rockwell HE, Gao F, Narain NR, DiFiglia M, Kiebish MA, Aronin N, Khvorova A. High-resolution proteomic and lipidomic analysis of exosomes and microvesicles from different cell sources. Journal of Extracellular Vesicles 2016;5:32570.

35. Gorgun C, Ceresa D, Lesage R, Villa F, Reverberi D, Balbi C, Santamaria S, Cortese K, Malatesta P, Geris L, Quarto R, Tasso R. Dissecting the effects of preconditioning with inflammatory cytokines and hypoxia on the angiogenic potential of mesenchymal stromal cell (MSC)-derived soluble proteins and extracellular vesicles (EVs). Biomaterials 2021;269:120633.

36. Sicco C Lo, Reverberi D, Balbi C, Ulivi V, Principi E, Pascucci L, Becherini P, Bosco MC, Varesio L, Franzin C, Pozzobon M, Cancedda R, Tasso R. Mesenchymal Stem Cell-Derived Extracellular Vesicles as Mediators of Anti-Inflammatory Effects: Endorsement of Macrophage Polarization. STEM CELLS Translational Medicine 2017;6:1018–1028.

37. Whittaker TE, Nagelkerke A, Nele V, Kauscher U, Stevens MM. Experimental artefacts can lead to misattribution of bioactivity from soluble mesenchymal stem cell paracrine factors to extracellular vesicles. Journal of Extracellular Vesicles 2020;9:1807674.

38. Cooper TT, Sherman SE, Bell GI, Dayarathna T, McRae DM, Ma J, Lagugné-Labarthet F, Pasternak SH, Lajoie GA, Hess DA. Ultrafiltration and Injection of Islet Regenerative Stimuli Secreted by Pancreatic Mesenchymal Stromal Cells. Stem Cells and Development 2021;30:247–264.

39. Sharma S, Mishra R, Bigham GE, Wehman B, Khan MM, Xu H, Saha P, Goo YA, Datla SR, Chen L, Tulapurkar ME, Taylor BS, Yang P, Karathanasis S, Goodlett DR, Kaushal S. A Deep Proteome Analysis Identifies the Complete Secretome as the Functional Unit of Human Cardiac Progenitor Cells. Circulation Research 2017;120:816–834.

40. Forteza-Genestra MA, Antich-Rosselló M, Calvo J, Gayà A, Monjo M, Ramis JM. Purity Determines the Effect of Extracellular Vesicles Derived from Mesenchymal Stromal Cells. Cells 2020, Vol 9, Page 422 2020;9:422.

41. Gomes FG, Andrade AC, Wolf M, Hochmann S, Krisch L, Maeding N, Regl C, Poupardin R, Ebner-Peking P, Huber CG, Meisner-Kober N, Schallmoser K, Strunk D. Synergy of human platelet-derived extracellular vesicles with secretome proteins promotes regenerative functions. Biomedicines 2022;10:238.

42. Balkom BWM van, Gremmels H, Giebel B, Lim SK. Proteomic Signature of Mesenchymal Stromal Cell-Derived Small Extracellular Vesicles. PROTEOMICS 2019;19:1800163.

43. Ramirez MI, Amorim MG, Gadelha C, Milic I, Welsh JA, Freitas VM, Nawaz M, Akbar N, Couch Y, Makin L, Cooke F, Vettore AL, Batista PX, Freezor R, Pezuk JA, Rosa-Fernandes L, Carreira ACO, Devitt A, Jacobs L, Silva IT, Coakley G, Nunes DN, Carter D, Palmisano G, Dias-Neto E. Technical challenges of working with extracellular vesicles. Nanoscale 2018;10:881–906.

44. Webber J, Steadman R, Mason MD, Tabi Z, Clayton A. Cancer exosomes trigger fibroblast to myofibroblast differentiation. Cancer Research 2010;70:9621–9630.

45. Zhang H-G, Liu C, Su K, Yu S, Zhang

4

L, Zhang S, Wang J, Cao X, Grizzle W, Kimberly RP. A membrane form of TNFα presented by exosomes delays T cell activation-induced cell death. The Journal of Immunology 2006;177:2025.2-2025.

46. Plebanek MP, Angeloni NL, Vinokour E, Li J, Henkin A, Martinez-Marin D, Filleur S, Bhowmick R, Henkin J, Miller SD, Ifergan I, Lee Y, Osman I, Thaxton CS, Volpert O V. Pre-metastatic cancer exosomes induce immune surveillance by patrolling monocytes at the metastatic niche. Nature Communications 2017;8:1319.

47. Clayton A, Al-Taei S, Webber J, Mason MD, Tabi Z. Cancer exosomes express CD39 and CD73, which suppress T cells through adenosine production. The Journal of Immunology 2011;187:676–683.

48. Sung BH, Ketova T, Hoshino D, Zijlstra A, Weaver AM. Directional cell movement through tissues is controlled by exosome secretion. Nature Communications 2015;6:7164.

49. Khan M, Nickoloff E, Abramova T, Johnson J, Verma SK, Krishnamurthy P, Mackie AR, Vaughan E, Garikipati VNS, Benedict C, Ramirez V, Lambers E, Ito A, Gao E, Misener S, Luongo T, Elrod J, Qin G, Houser SR, Koch WJ, Kishore R. Embryonic stem cell-derived exosomes promote endogenous repair mechanisms and enhance cardiac function following myocardial infarction. Circulation research 2015;117:52–64.

50. Lai RC, Tan SS, Teh BJ, Sze SK, Arslan F, Kleijn DP de, Choo A, Lim SK. Proteolytic Potential of the MSC Exosome Proteome: Implications for an Exosome-Mediated Delivery of Therapeutic Proteasome. International Journal of Proteomics 2012;2012:1–14.

CHAPTER 4 - SI

Supporting information for

Cardiac progenitor cell-derived extracellular vesicles promote cell activation through both associated- and co-isolated proteins

Marieke Theodora Roefs[1], Julia Bauzá-Martinez[2], Jiabin Qin[1], Simonides Immanuel van de Wakker[1], Wilte Olijve[1], Robin Tuinte[1], Marjolein Rozeboom[1], Emma Alise Mol[1], Wei Wu[2,3,*], Pieter Vader[1,4,*], Joost Petrus Gerardus Sluijter[1,*]

[1] Department of Experimental Cardiology, University Medical Center Utrecht, Utrecht University, The Netherlands.

[2] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands.

[3] Singapore Immunology Network (SIgN), Agency for Science, Technology and Research (A*STAR), 8A Biomedical Grove, Immunos, #04-06, Singapore 138648, Singapore.

[4] CDL Research, University Medical Center Utrecht, The Netherlands

Supplemental Table 1

Sequences of all guide RNAs for each CRISPR

4

		Genomic target region
PAPP-A gRNA 1	5'-AGGAGTAGCAACTTGGCCAT-3'	exon 3
PAPP-A gRNA 2	5'-AGTTGGCAGGAGTAGCAACT-3'	exon 3
PAPPA-A gRNA 3	5'-GGCCTCTATCACGTCTTCCG-3'	exon 4
NID1 gRNA 1	5'-GGGTTTGCGGGACCGCAGTT-3'	5'UTR/exon1
NID1 gRNA 2	5'-GAACTGCGGTCCCGCAAACC-3'	5'UTR
NID1 gRNA 3	5'-TGTCGGATCTGTCGTAGAAG-3'	exon1
Non targeting	5'-ATTTCCCTACGGAGATATCC-3'	

Primer sequences of genomic DNA amplification for T7 endo-nuclease assay

PAPP-A exon 3	F: 5'-GCCTGTTATCTTTTTGGGGGGC-3'
	R: 5'-CTGTGCCAGTACAGTGTGGT-3'
PAPPA-A exon 4	F: 5'-GCTCTCAGGGCAATGGAGTC-3'
	R: 5'-AACCCAGAAAGAGAATGACCGA-3'
NID1 exon 1	F: 5'-AACAAGTTGACAGCGACCCC-3'
	R: 5'-CCGGTTACATCCCCGCCTTC-3'

Supplemental Table 2 – Top-20 proteins enriched in veh-EVs compared with Ca ion-EVs, SKOV-3 EVs or both. According to the log2 fold-change of proteins (p-value ≤ 0.05), the top 20 most significantly changing proteins are displayed.

Gene name	Mean Veh_ EVs	Mean CaIon_EVs	Mean SKOV3_EVs	q-value (Student's T-test) Veh_EV vs CaIon_EV	q-value (Student's T-test) Veh_EV vs SKOV_EV
COL1A1	31.2	23.7	19.2	0.0228	0.0056
COL6A3	30.9	24.9	19.4	0.0013	0.0045
FN1	29.9	26.6	26.5	0.0171	0.0059
LAMA4	29.5	23.3	19.7	0.0053	0.0000
LAMC1	29.4	24.8	27.1	0.0000	0.0032
COL1A2	29.3	22.0	19.8	0.0241	0.0060
APP	29.3	22.1	21.8	0.0009	0.0054
UBA52	29.1	27.4	27.7	0.0122	0.0249
VCAN	29.0	25.2	18.9	0.0050	0.0055
EDIL3	29.0	24.3	17.6	0.0029	0.0000
LAMB1	28.8	24.2	26.5	0.0000	0.0030
COL6A1	28.8	23.9	22.8	0.0049	0.0043
SRGN	28.2	18.3	19.3	0.0051	0.0066
COL6A2	28.1	22.1	18.6	0.0047	0.0000
PAPPA	28.1	21.5	16.7	0.0051	0.0000
ANPEP	28.0	24.6	17.8	0.0118	0.0000
THBS1	27.9	23.9	26.4	0.0030	0.0200
NID1	27.8	19.2	19.7	0.0000	0.0049
BASP1	27.7	25.5	23.7	0.0050	0.0057
ITGB1	27.6	25.4	26.3	0.0175	0.0068
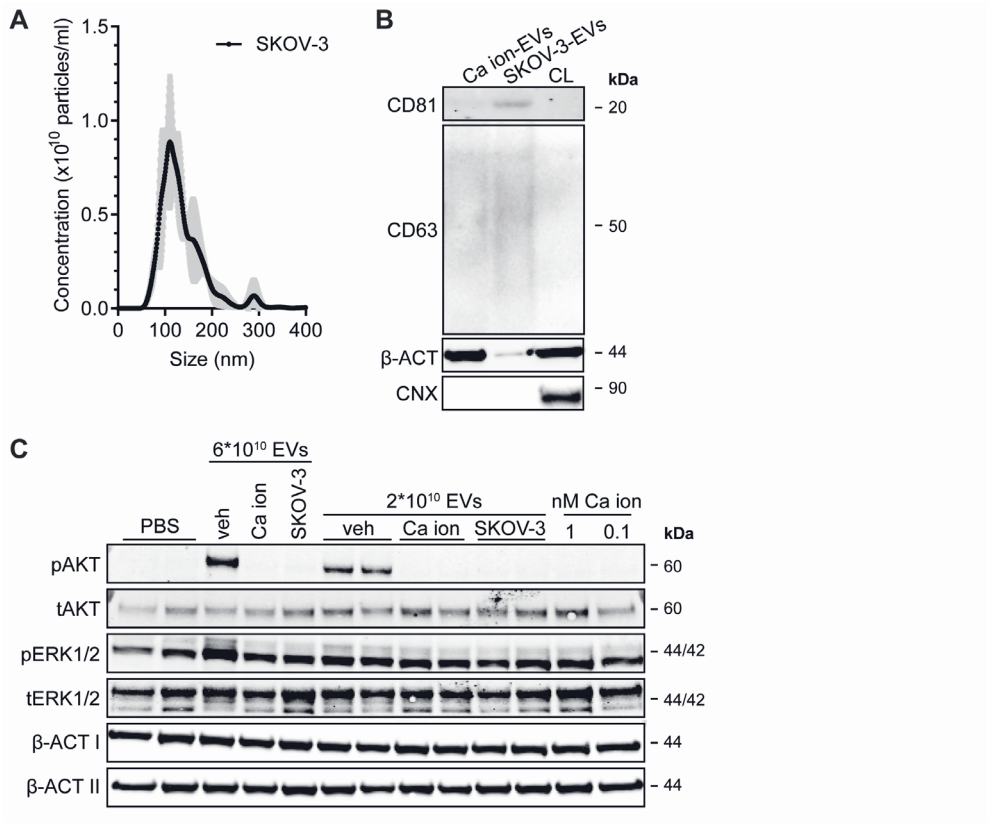
Supplementary Figure 1. (A) NTA plot showing the size distribution and particle concentration of SKOV-3 EVs. (B) Western blot analysis showing the presence of CD81, CD63, β-actin (β-ACT), and absence of Calnexin (CNX) in Ca ion-EVs and SKOV-3-EVs. β-ACT and CNX were present in Ca ion-CPC cell lysate (CL). (C) Representative western blot analysis of phosphorylated AKT (pAKT), total AKT (tAKT), phosphorylated ERK1/2 (pERK1/2) and total ERK1/2 (tERK1/2) in HMEC-1 treated with veh-, Ca ion- and SKOV-3 EVs normalized on two doses of EV particle numbers, and with 1 nM and 0.1 nM calcium ionophore A23187 in PBS. β-ACT was included as housekeeping protein.

4
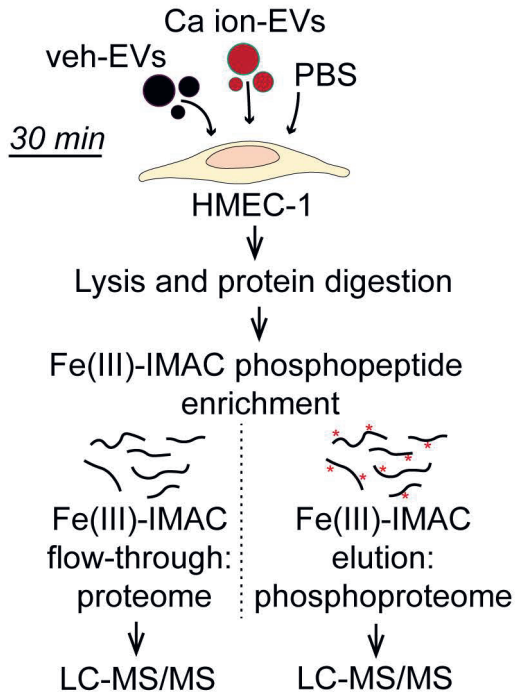
Supplementary Figure 2. Schematic of workflow used for phosphoproteomic analysis by liquid chromatography-mass spectrometry (LC-MS/MS). EV stimulated HMEC-1 were lysed and extracted proteins were digested into peptides and loaded into Fe(III)-Immobilized Metal Affinity Chromatography (Fe(III)-IMAC) cartridges for phosphopeptide enrichment. Fe(III)-IMAC flow-through containing the non-phosphorylated subset of proteome, and elutions containing the phosphorylated subset of the proteome, were analysed by (LC-MS/MS).
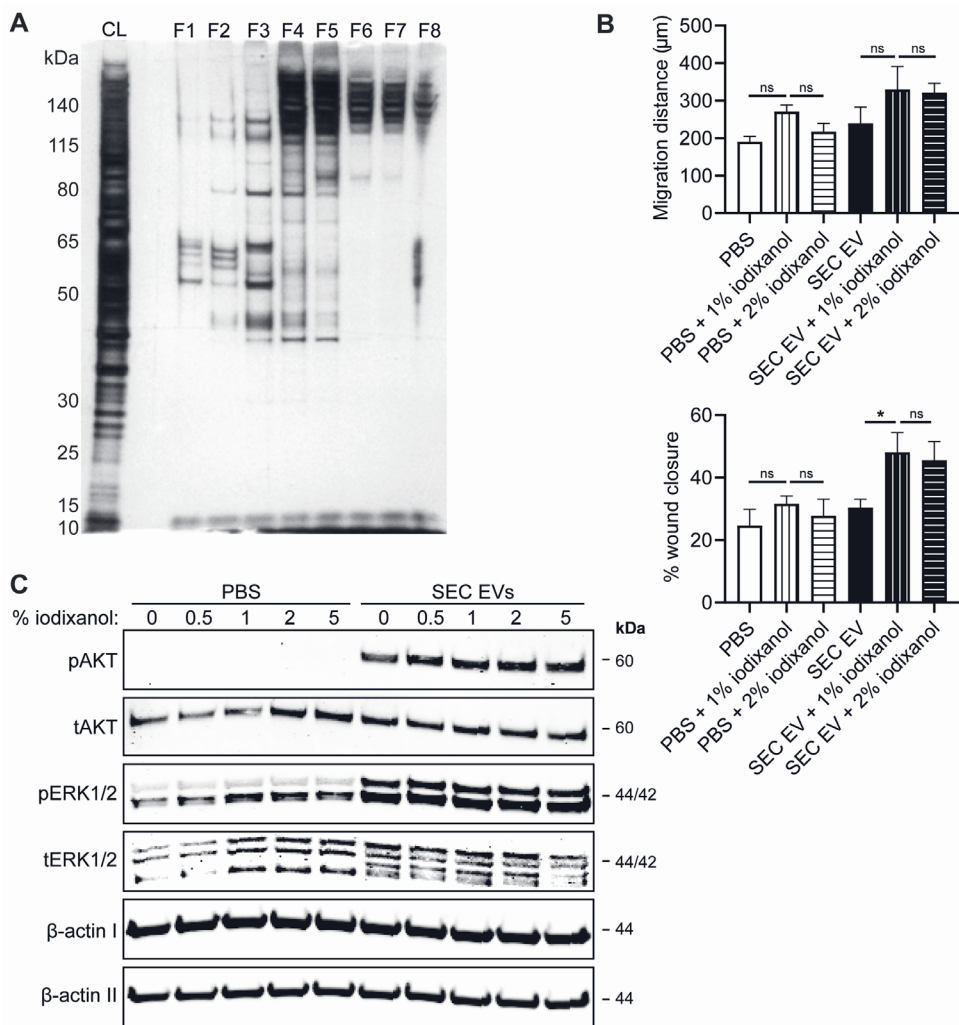
Supplementary Figure 3. (A) Silver stain of fractions (F1-8) collected after ultracentrifugation of SEC EVs loaded in the bottom of a discontinuous Optiprep™ gradient. Equal volumes of each sample were analysed, and 1 µg CPC cell lysate (CL) was included as control. (B) Representative wound healing assay showing effects of PBS and 1 µg CPC-EVs simultaneous with different concentrations of iodixanol on HMEC-1 migration, analysed both as absolute migration distance (top) and % wound closure (bottom). (C) Representative western blot analysis of phosphorylated AKT (pAKT), total AKT (tAKT), phosphorylated ERK1/2 (pERK1/2) and total ERK1/2 (tERK1/2) in HMEC-1 treated with PBS and 2×10^{10} EVs, simultaneous with different concentrations of iodixanol (0-5%). β-ACT was included as housekeeping protein.
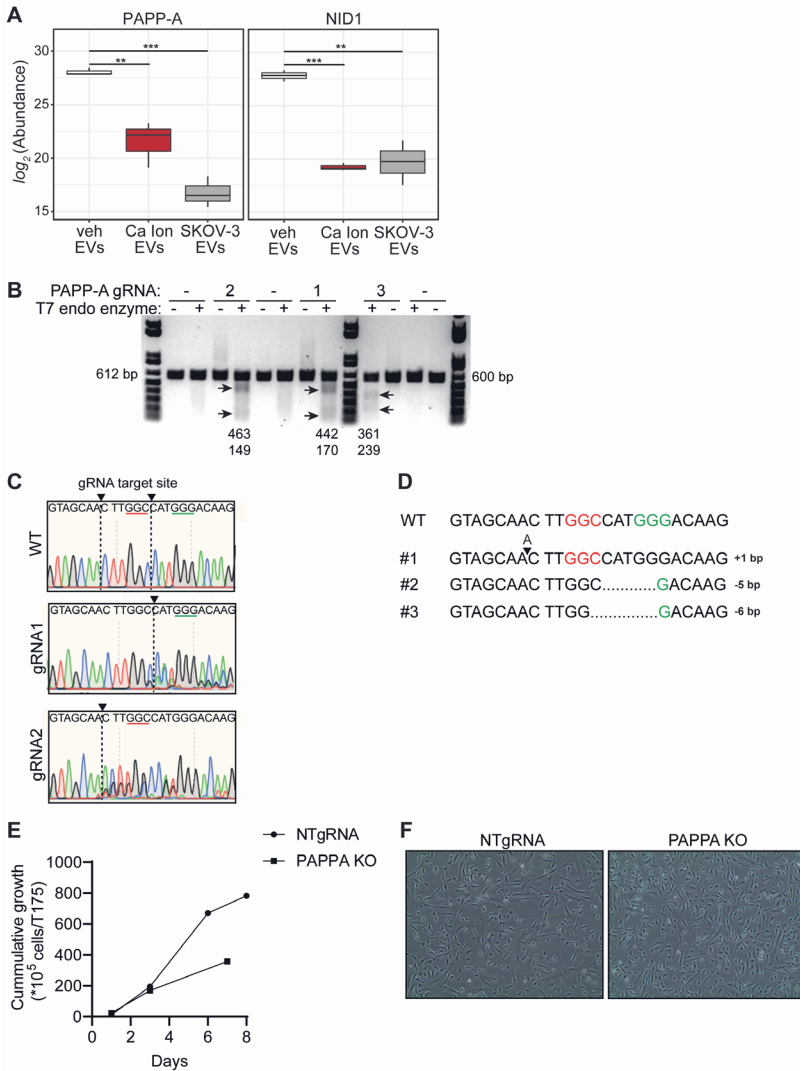
Supplementary Figure 4. (A) Boxplots representing protein abundance (log_2) of PAPP-A and NID1 in veh-, Ca ion- and SKOV-3 EVs, as measured with by LC-MS/MS. P-values were calculated using student's T-test. ** = $p<0.002101$, *** = $p<0.00102$. Boxes represent the 25%, 50% (median) and 75% quartiles and whiskers represent the ±1.5 interquartile range. (B) T7 endonuclease assay confirming a double-stranded break at the CRISPR/Cas9 target site in exon 3 of PAPPA in CPCs transduced with three different gRNAs (1-3). (C) Sanger sequencing results confirming frameshift in PAPPA exon 3 at the CRISPR/Cas9 target site of gRNA1 and gRNA2, compared with wild-type (WT) CPC genomic DNA. (D) Genomic DNA sequence surrounding two gRNA target sites in exon 3 of PAPPA of three individual PAPPA KO clones. Clones had homozygous mutations on both alleles. (E) Growth curves of NTgRNA and PAPPA KO CPC lines selected for functional studies, as visualized as cumulative cell growth. (F) Cell viability PAPPA KO and NTgRNA CPC clones.
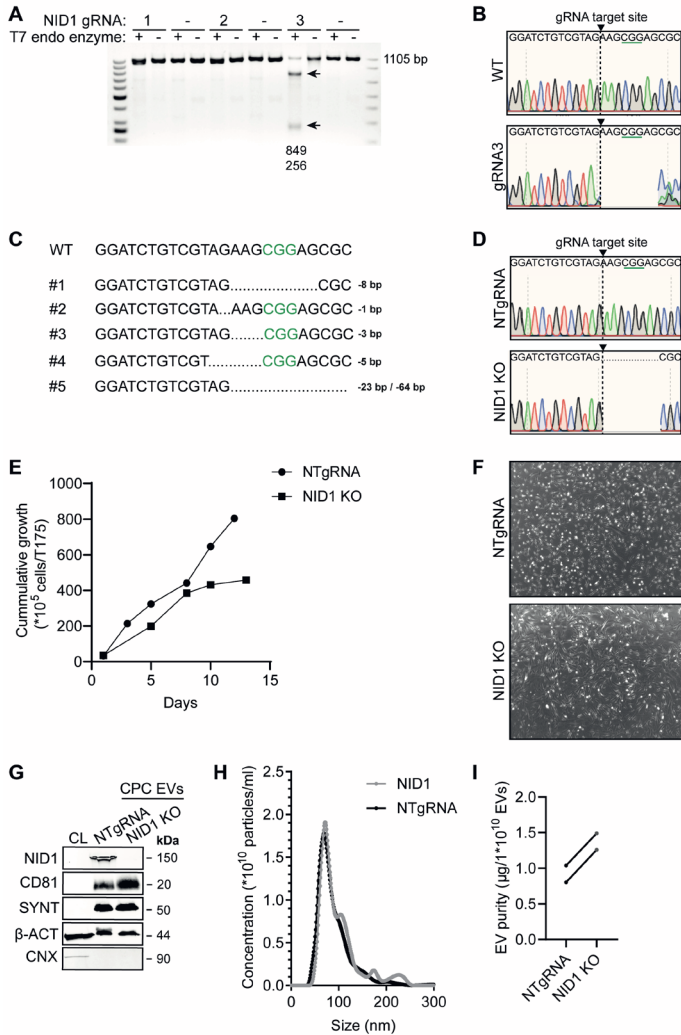
Supplementary Figure 5. (A) T7 endonuclease assay confirming a double-stranded break at the CRISPR/Cas9 target site in exon 1 of NID1 in CPCs transduced with three different gRNAs (1-3). (B) Sanger sequencing results confirming frameshift in NID1 exon 1 at the CRISPR/Cas9 target site of gRNA3, compared with wild-type (WT) CPC genomic DNA. (C) Genomic DNA sequence surrounding the gRNA3 target site in exon 1 of NID1 of five individual NID1 KO clones. Clones had homozygous mutations on both alleles. (D) Sanger sequencing results confirming 8 bp deletion in exon 1 of NID1 at the CRISPR/Cas9 target site of the NID1 KO-CPC clone, compared with the NTgRNA polyclonal CPC line. (E) Growth curves of NTgRNA and NID1 KO CPC lines selected for functional studies, as visualized as cumulative cell growth. (F) Cell viability NID1 KO and NTgRNA CPC clones. (G) Western blot analysis showing the absence of NID1 in NID1 KO-EVs compared with NTgRNA-EVs; the presence of CD81, Syntenin-1 (SYNT), β-actin (β-ACT), and absence of Calnexin (CNX) in EV populations. β-ACT and CNX were present in CPC cell lysate (CL). (H) Representative NTA plot showing the size distribution and particle concentration of NID1 KO- and NTgRNA-CPC-EVs. (I) Protein content per 1x10^10 NID1 KO- and NTgRNA-EVs of two representative experiments.

A

B
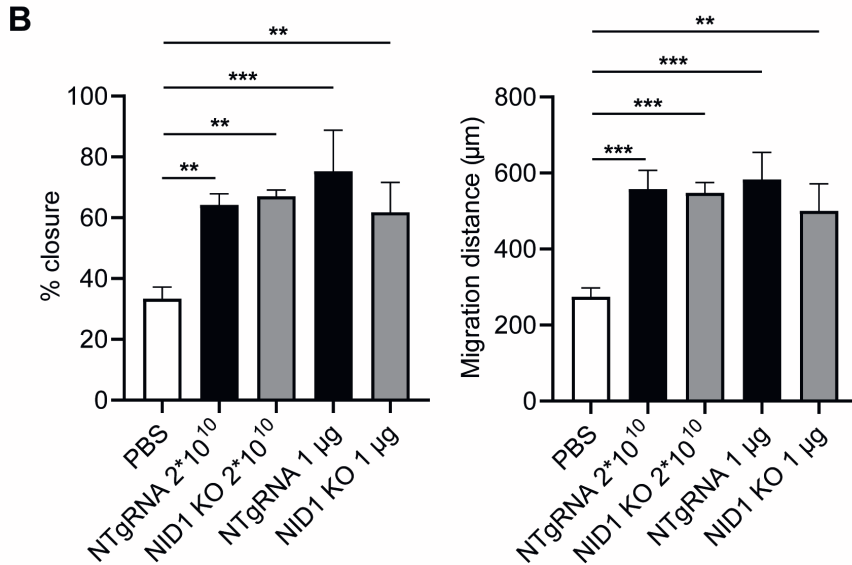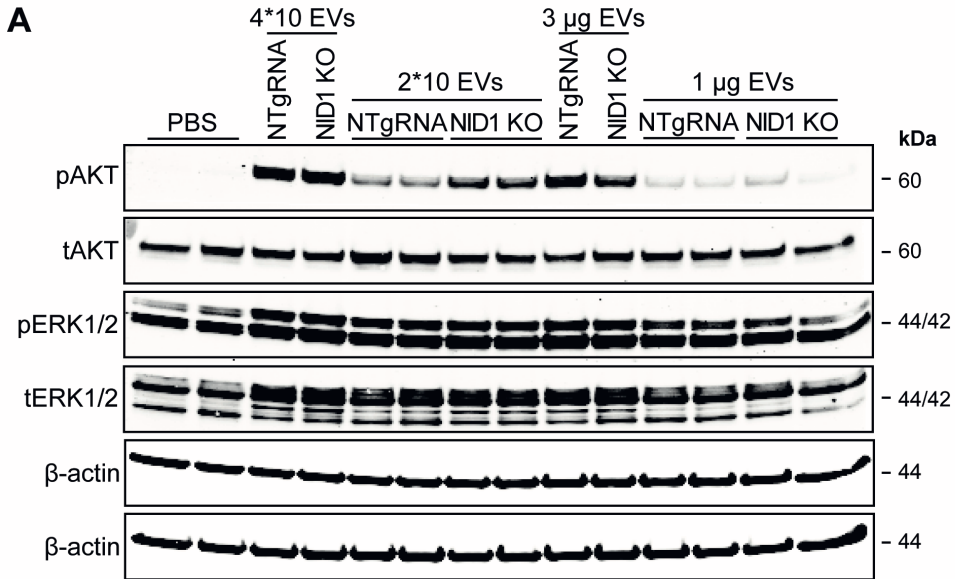
Supplementary Figure 6. (A) Representative western blot analysis of phosphorylated AKT (pAKT), total AKT (tAKT), phosphorylated ERK1/2 (pERK1/2) and total ERK1/2 (tERK1/2) in HMEC-1 treated with NID1 KO- and NTgRNA-EVs normalized on two doses of EV particle numbers or EV total protein content. β-ACT was included as housekeeping protein. (B) Representative wound healing experiment showing effects of $2x10^{10}$ and 1 µg NTgRNA- and NID1 KO-EVs on HMEC-1 migration, analysed both as % wound closure and absolute migration distance (n =3, technical replicates. Data are representative of independent experiments).

A

B
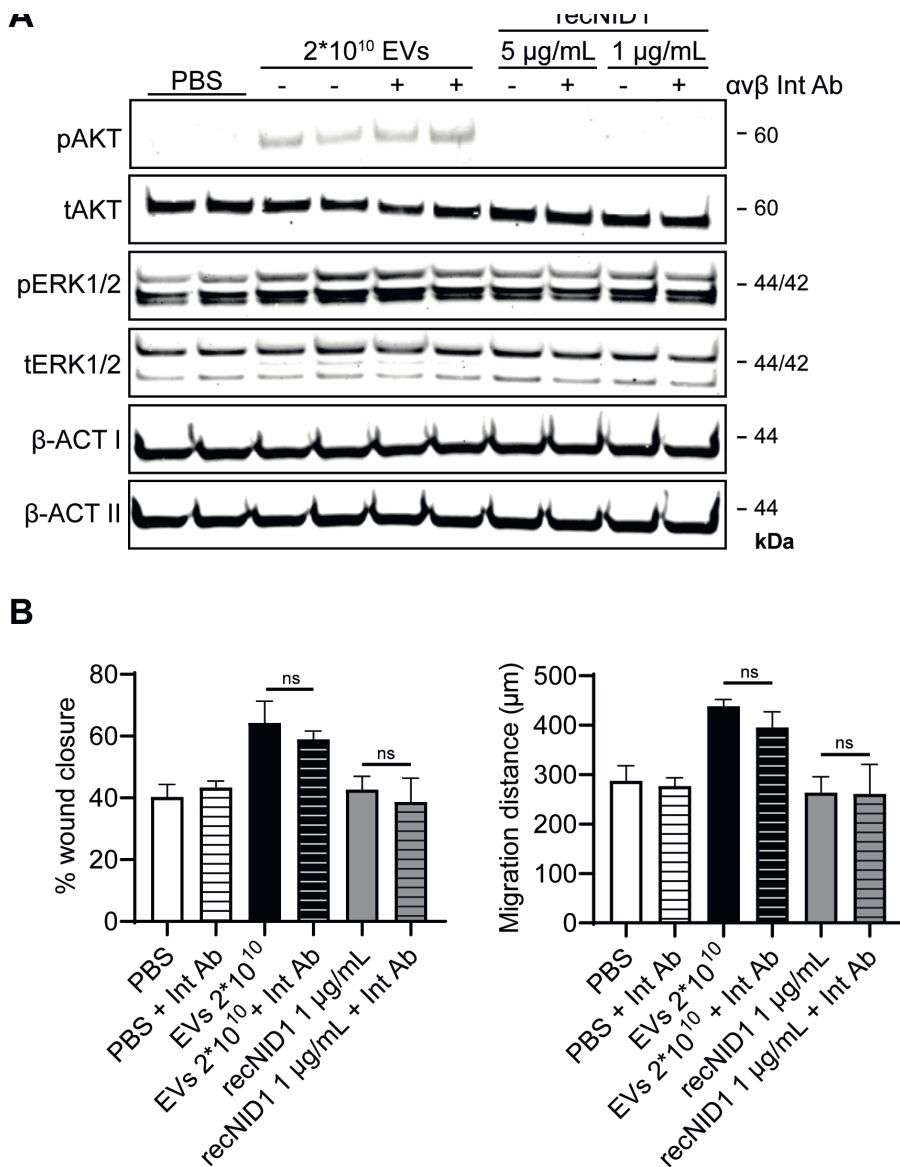
Supplementary Figure 7. (A) Representative western blot analysis of pAKT, tAKT, pERK1/2 and tERK1/2 in HMEC-1 treated with $2x10^{10}$ CPC-EVs, or with 5 µg/mL or 1 µg/mL recombinant IGF-1 after pre-incubation with (+) or without (-) 10 µg/mL IntegrinαVβ3 antibody. β-ACT was included as housekeeping protein. (B) Representative wound healing assay showing effects of 10 µg/mL IntegrinαVβ3 antibody addition to CPC-EV-induced HMEC-1 migration, analysed both as % wound closure and absolute migration distance. 5 µg/mL and 1 µg/mL recombinant NID1 (recNID1) were included as control (n =3, technical replicates. Data are representative of independent experiments). * = p<0.033, ** = p<0.0021, *** = p<0.0002.

Supplementary methods

Cell culture

CPCs were cultured in SP++ medium (66% M199 medium (Gibco), 22% EGM-2 (Lonza), 10% fetal bovine serum (FBS) (Life-Tech), 1% Penicillin/Streptomycin (P/S) (Invitrogen), and 1% MEM nonessential amino acids (Gibco). HMEC-1 were cultured in MCDB131 medium (Invitrogen) supplemented with 10% FBS, 1% P/S, 5% L-glutamine (Invitrogen), 50 nM Hydrocortisone (Sigma) and 10 ng/ml rhEGF-1 (Peprotech/Invitrogen). CPCs and HMEC-1 were cultured in flasks coated with 0.1% gelatin. SKOV-3 and HEK293fT cells were cultured in DMEM (Gibco) supplemented with 10% FBS and 1% P/S. All cells were cultured at 5% CO_2 at 37 °C and passaged at 80-90% confluency after digestion with 0.25% trypsin.

EV isolation by ultrafiltration and size-exclusion chromatography (SEC)

Conditioned medium was collected after 24 hrs, centrifuged at 2000 x*g* for 15 min and 0.45 µm filtered (0.45 µm aPES bottle top Nalgene filter) to remove cellular debris. Conditioned medium was concentrated using 100-kDA molecular weight cut-off (MWCO) Amicon Ultra-15 spin filters (Merck Millipore) or by Tangential Flow Filtration (TFF) using a Minimate TFF capsule with 100-kDa MWCO, and subsequently loaded onto a S400 high-prep column (GE Healthcare) using an ÄKTA start system (GE Healthcare) containing an UV 280 nm flow cell. Fractions containing EVs were pooled, filtered using a 0.45 µm surfactant-free cellulose acetate membrane (SCFA) syringe filter (Corning) and again concentrated using a 100-kDa MWCO Amicon Ultra-15 spin filter (Merck Millipore). EVs were stored at 4 °C for a maximum of 2 days until further use.

T7 endonuclease assay

The genomic region flanking the CRISPR/Cas9 target site was first amplified by PCR using Q5 Hot Start High-Fidelity 2x Master Mix (NEB) on 100 ng DNA template. After an initial incubation at 98 °C for 30 sec, 35 cycles of 98 °C for 10 sec, 65 °C for 30 sec, and 72 °C for 45 sec were used followed by a final extension at 72 °C for 10 min. Afterwards, amplicons were subjected to a melting and re-annealing process in NEBuffer 2 (New England Biolabs) to allow heteroduplex formation: 95 °C for 5 min, 95 °C to 85 °C ramping at 2 °C/second, 85 °C to 25 °C at 0.1 °C/second. After re-annealing, products were treated with T7 Endonuclease I (New England Biolabs) for 30 min at 37 °C following the manufacturer's recommended protocol and the digestion products were separated on 1% agarose gel containing ethidium bromide. Gels were imaged with a Gel Doc XR+ imaging system (Bio-Rad). To confirm the frameshift at the genomic DNA level at the CRISPR/Cas9 target site, amplified PCR products were analyzed by Sanger sequencing.

Western blot for EV markers

EVs were lysed in 1x RIPA buffer (Abcam). Samples were mixed with reducing sample buffer and loading buffer and heated to 90 °C for 10 min to reduce proteins. Equal protein amounts were loaded on a 4-12% Bis-Tris polyacrylamide gel

(Thermo Fisher Scientific) and subjected to electrophoresis. Proteins were blotted on a Immobilon-FL polyvinylidene difluoride (PVDF) membranes (Merck Millipore), which were subsequently blocked with 50% v/v Odyssey Blocking Buffer (LI-COR Biosciences) in Tris buffered saline (TBS). Antibodies were incubated in 50% v/v Odyssey Blocking Buffer in TBS containing 0.1% v/v Tween 20 (TBS-T). Primary antibodies used were mouse anti-CD63 (Abcam, 1:1,000), rabbit anti-CD9 (Abcam, 1:1,000), mouse anti-ALIX (Thermo Scientific, MA1-83977), rabbit anti-Calnexin (GeneTex, GTX 101676, 1:1,000), mouse anti-β-actin (Sigma, 1:5,000), rabbit anti-TSG101 (Abcam, 1:1,000), mouse anti-Syntenin-1 (Origene, TA504796, 1:1,000), rabbit anti-Annexin A1 (Abcam, ab214486, 1:1,000) and mouse anti-CD81 (clone B-11, Santa Cruz, 1:1,000). Secondary antibodies included Alexa680-conjugated goat anti-mouse (Thermo Fisher Scientific, 1:7,500) and IRG800-conjugated goat anti-rabbit (LI-COR Biosciences, 1:7,500). Imaging was performed on an Odyssey Infrared Imager (LI-COR Biosciences) at 700 and 800 nm. CD63 was detected in non-reduced protein samples.

Western blot for pAKT and pERK1/2 expression

Samples were mixed with NuPAGE sample reducing agent (Thermo Fisher Scientific) and NuPAGE sample buffer (Thermo Fisher Scientific), heated to 90 °C for 10 min, and subjected to electrophoresis over 4-12% Bis-Tris polyacrylamide gels (Thermo Fisher Scientific). Proteins were blotted on Invitrolon PVDF membranes (Thermo Fisher Scientific) using the iBlot 2 Dry blotting system (Thermo Fisher Scientific). Membranes were subsequently blocked with 5% Bovine serum albumin (BSA) or with 50% v/v Intercept Blocking Buffer (LI-COR Biosciences) in Tris buffered saline (TBS). Primary and secondary antibody incubations were performed in 0.5% BSA in TBS containing 0.1% Tween 20 (TBS-T) or in 50% v/v Intercept Blocking Buffer in TBS-T. Primary antibodies included rabbit anti-phospho-ERK (Phospho-p44/p42 MAPK (Erk1/2) (Thr202/Tyr204, Cell signaling Technology, 1:1,000), rabbit anti-ERK (p44/p42 (Erk1/2), Cell Signaling Technology, 1:1,000), rabbit anti-phospho-AKT (Ser473, Clone D9E, Cell Signaling Technology, 1:1,000), rabbit anti-AKT (Cell Signaling Technology, 1:1,000) and mouse anti-β-actin (Sigma, 1:5,000). Secondary antibodies included HRP-conjugated goat anti-rabbit antibody (DAKO) or Alexa680-conjugated goat anti-mouse (Thermo Fisher Scientific, 1:7,500) and IRG800-conjugated goat anti-rabbit (LI-COR Biosciences, 1:7,500). Proteins were detected with chemiluminescent peroxidase substrate (Sigma) using a Chemi DocTM XRS+ system (Bio-rad) and Image LabTM software, or imaging was performed on an Odyssey Infrared Imager (LI-COR Biosciences) at 700 nm and 800 nm.

LentiCRISPv2 plasmid construction

sgRNA expressing plasmids were achieved by cloning of the sgRNA in the LentiCRISPv2 vector (Addgene Plasmid number 52961) by oligo annealing of phosphorylated sgRNA oligo duplexes into Esp3I sites. Phosphorylated sgRNA oligo duplexes were formed by ligating top and bottom strand oligos (HPLC-purified, IDT) with T4 PNK ligase at 37 °C for 30 min, followed by 5 min incubation at 95 °C and a ramp-down of 5 °C per min to 25 °C.

Stable CPC line generation

HEK293fT cells were plated in 6-wells plates and transfections of plasmid DNA were performed using Lipofectamine 3000 reagent (Life Technologies) according to the manufacturer's instructions when cells were at 50-60% confluency. 1 µg LentiCRISPv2 plasmid (Addgene #52961) was mixed with 1 µg pCMV delta R8.2 (Addgene#12263) and 0.5 µg pCMV-VSV-G (Addgene #8454) helper plasmids and complexed in a 1 µg to 1 µl ratio with Lipofectamine 3000 reagent and in a 1 µg to 2 µl ratio with P3000 reagent. The complexes were added to the cells and incubated for 48 hrs before harvesting of the virus containing medium. Cell debris was removed by centrifugation at 350 x*g* for 15 min and filtration using a 0.45 µm SCFA syringe filter (Corning). CPCs were cultured in a 24-wells plate until 50% confluency before supplementation with 1 mL of virus-containing medium containing 8 ng/mL polybrene. After 24 hrs, medium was changed to SP++ medium and stable polyclonal cell lines were generated by puromycin selection starting 48 hrs after transduction. Individual knock-out (KO) clones were obtained by serial diluting polyclonal KO lines to single cell suspensions and expanding single cells in a 96-wells plate. Once confluent, genomic DNA was extracted using the GeneJet Genomic DNA Purification kit (Thermo Scientific) following the recommended protocol and investigated for a homozygous biallelic mutation using Sanger sequencing.

EV lysis and digestion for proteomic analysis

EVs were lysed by 2% SDS lysis buffer (2% SDS, 50 mM HEPES pH 7.6, 1 mM DTT) and prepared for MS analysis using a modified version of the SP3 protein clean up and digestion protocol (Moggridge et al., 2018). All extracted protein from each sample was alkylated with 4 mM Chloroacetamide. Sera-Mag SP3 bead mix (20 µl, Cytiva) was transferred into the protein sample together with 100% Acetonitrile to a final concentration of 70%. The mix was incubated under rotation at room temperature (RT) for 18 min. The mix was placed on the magnetic rack and the supernatant was discarded, followed by two washes with 70% ethanol and one with 100% acetonitrile. The beads-protein mixture was reconstituted in 100 µl LysC buffer (0.5 M Urea, 50 mM HEPES pH: 7.6 and 1:50 enzyme (LysC) to protein ratio) and incubated overnight. Finally, trypsin was added in 1:50 enzyme to protein ratio in 100 µl 50 mM HEPES pH 7.6 and incubated overnight followed by SP3 peptide clean up. Briefly, 20 µl Sera-Mag SP3 bead mix (10 µg/µl) was added to the sample. Next, 100% acetonitrile was added to achieve a final concentration of 95%. Samples were pipette-mixed and incubated for 20 min at RT and then placed on a magnetic rack. The supernatant was aspirated, discarded and the beads were washed in 180 µl of acetonitrile. Samples were removed from the magnetic rack and beads were reconstituted in 20 µl of (3% Acetonitrile, 0.1% formic acid) solution, followed by 1 min of sonication. Then the beads were placed on a magnetic rack again and the supernatant was recovered and transferred to a MS-vial. The peptides were dissolved in LC mobile phase A (3% acetonitrile (ACN), 0.01% FA) and were injected in the liquid chromatography (LC)-MS/MS system.

LC-MS/MS analysis of EVs

Data were acquired using a Dionex UltiMate™ 3000 RSLCnano System coupled to a Q-Exactive mass spectrometer (Thermo Scientific). Samples were trapped on a C18 guard desalting column (Acclaim PepMap 100, 75µm x 2 cm, nanoViper, C18, 5 µm, 100 Å), and separated on a 50 cm long C18 column (Easy spray PepMap RSLC, C18, 2 µm, 100Å, 75 µmx50cm). The nano capillary solvent A was 95% water, 5% DMSO, 0.1% formic acid; and solvent B was 5% water, 5% DMSO, 95% acetonitrile, 0.1% formic acid. At a constant flow of 0.25 µl/min, the curved gradient went from 6%B up to 43%B in 180 min, followed by a steep increase to 100% B in 5 min.

The mass spectrometer was operated in data-dependent mode. Peptides were ionized in a nESI source at 1.9 kV and focused at 60% amplitude of the RF lens. Full scan MS1 spectra from 400 - 1200 m/z were acquired in the Orbitrap at a resolution of 60,000 with the AGC target set to 1×10^6 or for a maximum injection time of 250 ms. For each cycle, the top 10 most abundant precursor ions were isolated (with an isolation window of 2 m/z) for fragmentation while precursor ions with charge states 1 or unassigned were excluded for fragmentation. Dynamic exclusion was set to a duration of 20 s. Fragmentation was done using fixed HCD normalized collision energy of 30%. Fragment ions were accumulated until a target value of 2×10^5 ions was reached or for a maximum injection time of 140 ms before injection in the Orbitrap for MS2 analysis at a resolution of 30,000.

HMEC-1 protein digestion for (phospho)proteome analysis

Proteins were precipitated by the methanol/chloroform method: 1 volume of sample was sequentially mixed with 4 volumes of methanol (Sigma-Aldrich), 1 volume of chloroform (Sigma-Aldrich) and 3 volumes of water. The mixture was centrifuged at 5,000 rpm for 10 min at RT and the upper layer was removed. Then, 3 volumes of methanol were incorporated and centrifuged at 5,000 rpm for 10 min at RT, and the liquid phase was discarded while the pellet (proteins) was allowed to air dry. The protein pellets were reconstituted in 50 µL of digestion buffer (100 mMTris-HCl pH 8.5, 1% SDC (Sigma-Aldrich), 5 mM TCEP and 30 mM CAA) to reduce and alkylate the proteins. Protein digestion was then performed by adding LysC at a 1:100 (w/w) ratio for 1 h at 37 ºC, followed by overnight incubation at 37 ºC with Trypsin at a 1:25 ratio (w/w). Digestions were quenched 0.5% FA, and the SDC precipitate was pelleted by centrifugation at 20,000 *xg* for 5 min at 4 ºC. Supernatants were loaded twice into Pierce C18 10 µL bed Stage tips (Thermo Fisher) for desalting, washed with 100 µL of 0.1% FA, and peptides were eluted twice on 80% ACN, 0.1% FA, vacuum dried and stored at −80 °C before phosphopeptide enrichment.

Automated Fe(III)-IMAC phosphopeptide enrichment

Phosphorylated peptides were enriched on Fe(III)-NTA 5 µL (Agilent technologies) cartridges using the AssayMAP Bravo Platform (Agilent Technologies). Fe(III)-NTA cartridges were primed with 250 µL of 0.1% TFA in ACN at a 100 µL/min flow rate and equilibrated with 250 µL of loading buffer (80% ACN/0.1% TFA) at a 50 µL/min flow rate. Samples were dissolved in 200 µL of loading buffer and loaded onto the cartridge at a 3 µL/min flow rate, followed by a 250 µL wash in loading buffer at 20

µL/min. The load flowthrough, containing the non-phosphorylated subset of the pro-teome, was kept for further proteomic characterization. The phosphopeptides were then eluted with 35 µL of 1% ammonia at 5 µL/min directly into 35 µL of 10% formic acid. Samples were vacuum dried and stored in −80 °C until LC–MS/MS analysis.

LC-MS/MS analysis of EV-stimulated HMEC-1

Data were acquired with an Ultimate 3000 system (Thermo Fischer Scientific) cou-pled to an Orbitrap Exploris 480 mass spectrometer (Thermo Fischer Scientific). Peptides were trapped (Dr Maisch Reprosil C18, 3µM, 2cm x 100µM) for 5min in solvent A (0.1% formic acid in water) before being separated on an analytical column (Agilent Poroshell, EC-C18, 2.7µM, 50cm x 75µM). Solvent B consisted of 0.1% formic acid in 80% acetonitrile. Trapping of peptides was performed for 2 min in 9% B at a flow rate of 300 nL/min. For full proteome analysis, peptides were separated in a gradient of 13–44% B in 95 min, while for phosphoproteomic analysis, peptides were separated in a gradient of 9–36% B in 36 min After (phospho)peptide separa-tion, gradients were followed by a steep increase to 99% B in 3 min, a 5 min wash in 99% B and a 10 min re-equilibration at 9% B. Flow rate was kept at 300 nL/min. The mass spectrometer was operated in data-dependent mode. Peptides were ionized in a nESI source at 1.9 kV and focused at 40% amplitude of the RF lens. Full scan MS1 spectra from 375 - 1600 m/z were acquired in the Orbitrap at a resolution of 60,000 with the AGC target set to 1×10^6 and under automated calculation of maxi-mum injection time. Cycle time for MS2 fragmentation scans was set to 1 second. Only peptides with charged states 2-6 were fragmented, and dynamic exclusion was set to a duration of 10 ms for 36 min gradients and to 16 ms for 95 min gradients. Fragmentation was done using fixed HCD normalized collision energy of 28%. Frag-ment ions were accumulated until a target value of $1x10^5$ ions was reached under an automated calculation of maximum injection time, with an isolation window of 1.4 m/z before injection in the Orbitrap for MS2 analysis at a resolution of 30,000. Proteomics raw data have been deposited to ProteomeXchange Consortium via the PRIDE repository[1] and can be accessed through the identifier PXD030779.

Database search and (phospho)proteomics data analysis

All (phospho)proteomics raw data were searched in MaxQuant (v_1.6.10.43)[2] against the SwissProt human reference proteome database (containing 20,381 pro-teins and downloaded from Uniprot on March 2021). Spectra were searched using MaxQuant's built-in Andromeda search engine. Trypsin was set as the digestion enzyme and up to two missed cleavages were allowed. Carbamidomethylation of cysteines was set as a fixed modification, while protein N-terminal acetylation and methionine oxidation were set as variable modifications. For phosphoproteomic data analysis, phosphorylation of serine, tyrosine and threonine were also included as variable modifications. Label-free quantification (LFQ) was enabled using a mini-mum ratio count of two and both razor and unique peptides for quantification. Match between runs was enabled, the matching time window was always set to 0.7 min while the alignment time window was set to 20 min for proteomic analysis and 10 min for phosphoproteomic analysis. Precursor ion tolerance was set to 20 ppm for the first search and 4.5 ppm after recalibration, and fragment ions tolerance was

set to 20 ppm. False discovery rate (FDR) of 1% was set at PSM, site and protein level by using a reverse decoy database strategy. Data was analyzed using Perseus software (v_1.6.14)[3]. In each analysis, proteins quantified (LFQ) in two out of three replicates were log2 transformed and missing values were replaced individually for each sample from the normal distribution. For phosphoproteomic analysis, where intensity-based MS injection balancing was not possible, phosphosite intensities were normalized to the maximum cumulative intensity (found in sample veh-EV, replicate 2). Statistical differences were always assessed by two-sided Student's T-test or One-way ANOVA and corrected p-values (q-value) were calculated using the permutation method with up to 250 iterations. Proteins were considered significant when q-value ≤ 0.05. Gene ontology analysis were done using PANTHER[4] with the human proteome as background gene set. All plots were generated using R packages[5].

References

1. Vizcaíno JA, Csordas A, Del-Toro N, Dianes JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y, Reisinger F, Ternent T, Xu QW, Wang R, Hermjakob H. 2016 update of the PRIDE database and its related tools. Nucleic Acids Research 2016;44:D447.

2. Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nature Protocols 2016 11:12 2016;11:2301–2319.

3. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nature Methods 2016 13:9 2016;13:731–740.

4. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Research 2019;47:D419.

5. Peschel I, Eisler V. Computational Many-Particle Physics. Fehske H, Schneider R, Weiße A, eds. R Foundation for Statistical Computing 2008;739:581–596.

CHAPTER 5

High Resolution mRNA and Secretome Atlas of Human Enteroendocrine Cells

Joep Beumer[1,2,14], Jens Puschhof[1,2,14], Julia Bauzá-Martinez[3,4,14], Adriana Martínez-Silgado[1,2], Rasa Elmentaite[5], Kylie R. James[5], Alexander Ross[6,7], Delilah Hendriks[1,2], Benedetta Artegiani[1,2], Georg Busslinger[1,2], Bas Ponsioen[8], Amanda Andersson-Rolf[1,2], Kai Kretzschmar[1,2], Maarten H. Geurts[1,2], Yotam E. Bar-Ephraim[1,2], Cayetano Pleguezuelos Manzano[1,2], Yorick Post[1,2], Franka van der Linden[9], Carmen Lopez Iglesias[10], Willine J. van de Wetering[1,10], Reinier van der Linden[1,2], Peter J. Peters[10], Albert .J.R. Heck[3,4], Joachim Goedhart[9], Hugo Snippert[8], Matthias Zilbauer[7], Sarah A. Teichmann[5,11,12], Wei Wu[3,4,#], Hans Clevers[1,2,13,15,*]

[1] Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences (KNAW) and UMC Utrecht, 3584 CT Utrecht, The Netherlands.
[2] Oncode Institute, Hubrecht Institute, 3584 CT Utrecht, The Netherlands.
[3] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.
[4] Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands.
[5] Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, United Kingdom, CB10 1SA.
[6] Department of Surgery, University of Cambridge, Cambridge CB2 0QQ, UK
[7] Department of Paediatrics, University of Cambridge, Cambridge CB2 0QQ, UK
[8] Oncode Institute, Center for Molecular Medicine, University Medical Centre Utrecht, Utrecht, The Netherlands.
[9] Swammerdam Institute for Life Sciences, Section of Molecular Cytology, van Leeuwenhoek Centre for Advanced Microscopy, University of Amsterdam, Amsterdam, the Netherlands.
[10] The Maastricht Multimodal Molecular Imaging institute, Maastricht University, 6229 ER Maastricht, The Netherlands.
[11] Theory of Condensed Matter, Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge, United Kingdom, CB3 0HE
[12] European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, United Kingdom, CB10 1SA
[13] The Princess Maxima Center for Pediatric Oncology, 3584 CS Utrecht, The Netherlands.
[14] Co-first author and [15] Lead Contact

Summary

Enteroendocrine cells (EECs) sense intestinal content and release hormones to regulate gastrointestinal activity, systemic metabolism and food intake. Little is known about the molecular make-up of human EEC subtypes and the regulated secretion of individual hormones. Here we describe an organoid-based platform for functional studies of human EECs. EEC formation is induced in vitro by transient expression of NEUROG3. A set of gut organoids was engineered in which the major hormones are fluorescently tagged. A single-cell mRNA atlas was generated for the different EEC subtypes, and their secreted products were recorded by mass-spectrometry. We note key differences to murine EECs, including hormones, sensory receptors and transcription factors. Notably, several novel hormone-like molecules were identified. Inter-EEC communication is exemplified by secretin-induced GLP-1 secretion. Indeed, individual EEC subtypes carry receptors for various EEC hormones. This study provides a rich resource to study human EEC development and function.

5

Introduction

The principal function of the intestine is to digest food and absorb nutrients, but it is also the largest hormone-producing organ, through its enteroendocrine cells (EECs)[1]. EECs are rare secretory cells, comprising <1% of the epithelial cells. Apical EEC receptors sense chemicals in the intestinal lumen derived from food and microbiota[2]. Hormones secreted by EECs signal to the local enteric nervous system and to distant organs including the pancreas and the brain, thus controlling food intake, insulin release, secretion of digestive enzymes and bowel movement. EECs are therapeutic targets for metabolic diseases (i.e. obesity and diabetes), illustrated by recently introduced type 2 diabetes drugs that stabilize the hormone Glucagon-like peptide 1 (GLP-1) or activate its receptor, leading to release of insulin from pancreatic β-cells[3].

EECs produce ~20 different hormones. GLP-1 and Glucose-dependent insulinotropic peptide (GIP) are the incretin hormones that stimulate insulin secretion. The Enterochromaffin (EC) cells produce 90% of body serotonin and regulate bowel movement[4]. Motilin (MLN) is a human EEC hormone, which controls gut contractions in the inter-digestive state[4]. Multiple hormones control appetite, including the appetite-inducing Ghrelin (GHRL) and the appetite-reducing Peptide YY (PYY) and Cholecystokinin (CCK). Gastrin (GAST) is secreted in the duodenum to control luminal acid by regulating proton secretion of stomach parietal cells. Somatostatin (SST) is an inhibitory peptide for most other intestinal hormones[4].

Lgr5+ cells generate all differentiated intestinal cell types[5]. The murine EEC subtypes are historically defined by their principle hormone product: L-cells (Glp-1, Pyy), I-cells (Cck), K-cells (Gastric inhibitory protein, Gip), N-cells (Neurotensin, Nts), S-cells (Secretin, Sct), enterochromaffin or EC-cells (Serotonin/5-HT), X-cells (Ghrl), G-cells (Gast) and D-cells (Sst)[6][7]. While this suggests that EEC phenotypes are hard-wired, we have recently found that the crypt-villus BMP-signaling gradient induces hormone switching within individual murine EEC lineages[8]. The relative abundance of EEC subtypes greatly differs along the proximal-distal gastrointestinal axis. Studies on EECs have largely focused on murine models, exploiting a variety of reporter mice for subsets of EECs to monitor their responses to nutritional or genetic challenges[9]. We have recently described the developmental hierarchy of murine subtypes EECs using a mouse model in which endogenous Neurogenin-3 expression, the main determinant of EEC fate, was coupled to the production of two separate fluorescent proteins with different half-lives[7]. Single cell RNA sequencing of sorted EEC progenitors allowed for construction of a time-resolved development roadmap of the mouse EEC lineage.

As the human diet and microbiome and that of rodents differ greatly[10]; secretory hormone responses may also differ between these species. The study of human EECs is challenging due to their rarity and the lack of physiologically relevant *in vitro* models. Few human EEC immortalized cell lines exist, and these differ substantially from their wildtype counterparts[11]. There is currently no atlas of human EEC subtypes. Although some inducers of hormone secretion have been described in mice, there has been no experimental model to systematically assess such secretagogues for

human EECs. Here, we describe an organoid-based platform to provide a detailed molecular and functional description of human EECs.

Results

Production of region-specific human EECs

Previous attempts to create human EECs *in vitro* have relied on growth-factor based differentiation (Beumer et al., 2018) or overexpression of *NEUROG3*, the key transcription factor to instruct EEC fate[12,13]. Both iPSC-[14] and adult stem cell (AS-C)-based[15] approaches allow studying of human EEC biology, such as modeling of hereditary NEUROG3 mutations and virally induced serotonin release[16]. However, imperfect differentiation and regional restriction of the donor material have limited these studies to a subset of human EECs.

To generate the full spectrum of human EECs, we established organoids from healthy adult proximal small intestine (duodenum), distal small intestine (ileum) and the ascending colon[17]. These organoids were transduced with a doxycycline-inducible *NEUROG3* construct (**Fig. 1A**). dTomato was inserted 3' to the *NEUROG3* reading frame, separated by a self-cleavable P2A sequence to avoid creating a fusion protein. A 48-hour-pulsed expression of *NEUROG3* in the basic medium 'ENR' promoted the expression of the broad EEC marker Chromogranin A (*CHGA)* (**Fig. 1B**). Proximal SI hormones such as *GAST, CCK and MLN* were enriched in duodenal organoids, whereas *NTS, PYY and GCG* were predominantly observed in distal SI organoids. Of note, *GCG* encodes the preproglucagon prehormone, a protein precursor to a set of hormones including GLP-1 (see below). *SST* was comparably expressed in proximal and distal organoids, consistent with its profile in the mouse gut. A recent single cell RNA sequencing study generated the profile of 83 colonic EECs, suggesting that the human colon only contains serotonin-producing ECs and L-cells positive for *GCG* and *PYY*[18]. Consistently, induced colon organoids only yielded serotonin-producing ECs and GCG-expressing EECs. We found that hormone expression peaked 5 days after initiation of *NEUROG3* expression (**Fig. S1A**). Shorter doxycycline inductions stimulated the production of fewer EECs, while continuous doxycycline treatment throughout the differentiation window enhanced EC-generation at the expense of L-cells (**Fig. S1B**). This suggests that Neurogenin-3 expression duration determines EEC subtype. We chose a 48 hours doxycycline treatment in ENR, the maximum duration of *NEUROG3* expression *in vivo*[7]. Under these conditions, EECs in organoids displayed a normal morphology as visualized by transmission electron microscopy. Note the typical basal concentration of hormone vesicles (**Fig. 1C**).

By immunofluorescent staining (**Fig. S1C**), we observed mutually exclusive expression for MLN and GAST, for GHRL and CHGA and for Serotonin and GLP-1, while a subset of GIP-positive cells co-expressed CCK. This closely resembled the co-expression patterns in mice (with the exception of MLN, a pseudogene in mice)[19] (**Fig. S1C**). Virtually all EECs, as identified by the broad marker CHGA, were derived from NEUROG3-overexpressing cells (**Fig. S1C**). A definitive hallmark of a mature

EEC is its ability to secrete hormones. Indeed, exposure to forskolin, a stimulator of adenylate cyclase, greatly enhanced secreted GLP-1 levels (**Fig. 1D**). We have previously reported that BMP signaling induces hormone expression changes in mature murine EECs in villi[8]. Consistent with our observations in murine EECs, we found that activation of BMP signaling enhances the expression of *NTS,* while reducing GLP-1 (**Fig. S2A,B**).
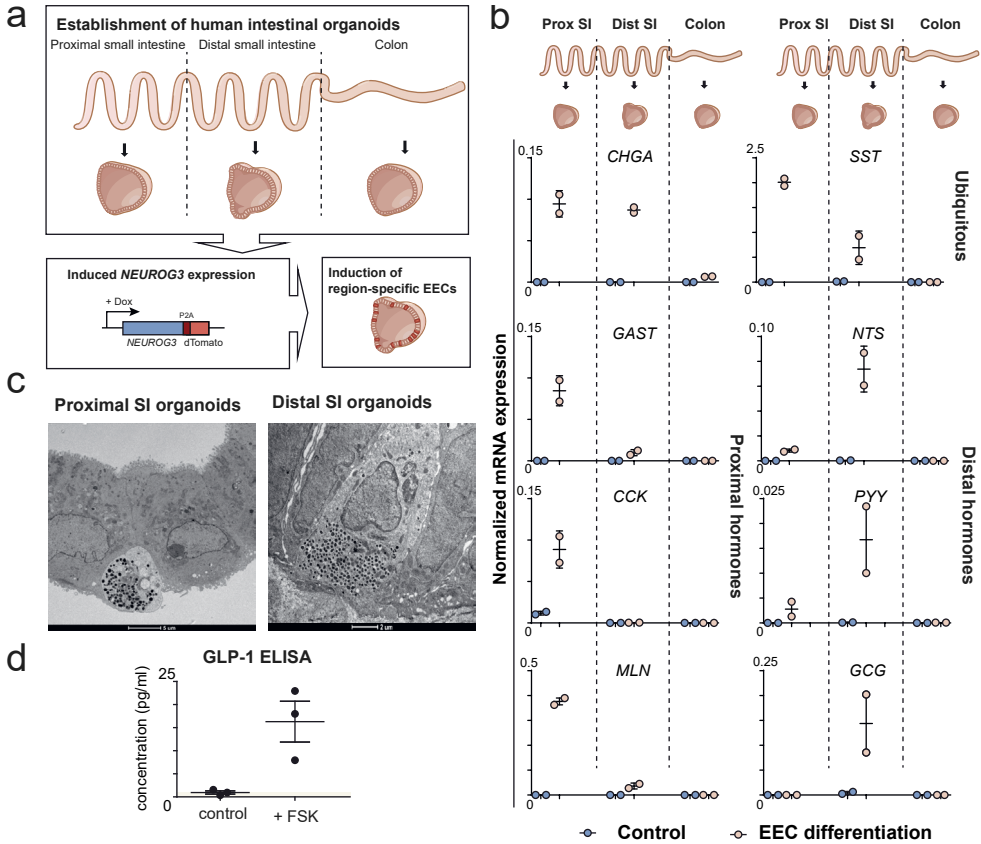
Figure 1. Production of region-specific human enteroendocrine cells in intestinal organoids. (A) Schematic representation of the generation of region-specific enteroendocrine cells (EECs). Organoids are established from different regions of the intestinal tract of different patients, after which doxycycline (dox) – induced overexpression of Neurogenin-3 (NEUROG3) can drive the production of EECs. (B) qPCR analysis showing expression of hormones, control or differentiation condition after a pulse of dox. Expression levels are normalized to GADPH. The experiment was performed in n = 2 independent experiments, and the mean expression and SEM are depicted. (C) Transmission electron microscopy (TEM) of EECs in organoids showing polarized localization of hormone vesicles. Scale bar is 5 µm (left image) and 2 µm (right image). (D) Concentration of supernatant GLP-1 determined by ELISA, in the absence (control) and presence of forskolin (FSK). The brown shaded area presents the detection threshold of GLP-1 (1 pg/ml). Forskolin induces secretion of GLP-1, confirming functionality of EECs. The experiment was performed in n = 3 independent experiments, and the mean concentration and SEM are depicted.

Since the initial expression of NEUROG3 occurs at random positions along the crypt axis in mice[7], we hypothesized that exposure to other crypt differentiation signals (i.e. Notch, Wnt) prior to this expression pulse could potentially determine EEC subtype. We modulated these signals prior to inducing NEUROG3 expression, mimicking the different initiation sites along the intestinal crypt axis (**Fig. S2C**). As a control, we modulated the same signals after NEUROG3 induction (**Fig. S2D**). Inhibition of Notch before or after expression of NEUROG3 did not affect EEC differentiation (**Fig. S2E**). Inhibition of Wnt signaling before (but not after) the NEUROG3 pulse stimulated expression of *MLN* at the expense of *GCG*, while *SST* was unaffected (**Fig. S2E**). Immunofluorescence revealed an increase in the number of MLN-producing cells rather than in the 'per cell'-expression levels (**Fig. S2F**), resulting in a strong shift in L-cell/M-cell ratio (**Fig. S2G**).

Generation of hormone reporter biobank EEC-TAG

Mouse models in which hormones are fluorescently tagged exist for several murine EEC hormones that were instrumental to study EEC subsets: Chga, Gcg, Gip, Cck, Ghrl and Pyy[20–25]. CRISPR/Cas9 targeting followed by homology-directed-repair (HDR) or Non-homologous-end-joining (NHEJ) allows the introduction of exogenous genetic material[26–28]. To mark ECs, we labeled Tryptophan hydroxylase 1 (TPH1), the rate-limiting enzyme involved in serotonin synthesis. Using HDR, we tagged TPH1 with fluorescent mClover separated by a self-cleaving P2A site. We recently optimized a strategy for site-specific introduction of DNA into organoids using NHEJ (CRISPR-HOT)[29], which allowed fluorescently labeling of multiple secreted hormones (**Fig. 2A**).

We generated a biobank of hormone reporter organoids termed EEC-TAG, consisting of duodenal, ileal and colon organoids for a total of 10 human hormones (**Fig. 2B**). All organoid lines showed complete overlap between fluorescent reporters and the corresponding hormone product (**Fig. 2C**). The fluorescently tagged hormones localized to cytoplasmic vesicles. Serial tagging into the same organoid lines for multi-hormone labeling was feasible (**Fig. 2D**).

Calcium signaling induces hormone secretion[11]. We stably introduced a Turquoise Ca2+ sensor (Tq-Ca-FLITS) into NEUROG3^{dTomato} TPH1^{mClover} reporter organoids using lentiviral transduction. The resulting genotype of the organoids is NEUROG3^{dTomato} TPH1^{mClover} CaFLITS^{Turqouise}. We stimulated the olfactory receptor OR51E2, of which the mouse homologue (*Olfr78*) is reported to be expressed in mouse EECs[30,31]. HEK cells overexpressing OR51E2 elicit a calcium response when stimulated with the selective agonist beta-ionone[32]. *OR51E2* is most strongly upregulated in distal organoids (**Fig. 2E**). When reporter organoids were stimulated with beta-ionone, we observed calcium sparking in EECs that were TPH1-negative (**Fig. 2F**), illustrating that sensors combined with hormone reporters can visualize activation of human EEC subtypes.

Single cell transcriptomics

Studies have utilized reporter mice to enrich for hormone-producing cells when performing single cell RNA sequencing. This approach cannot be used for primary hu-

157
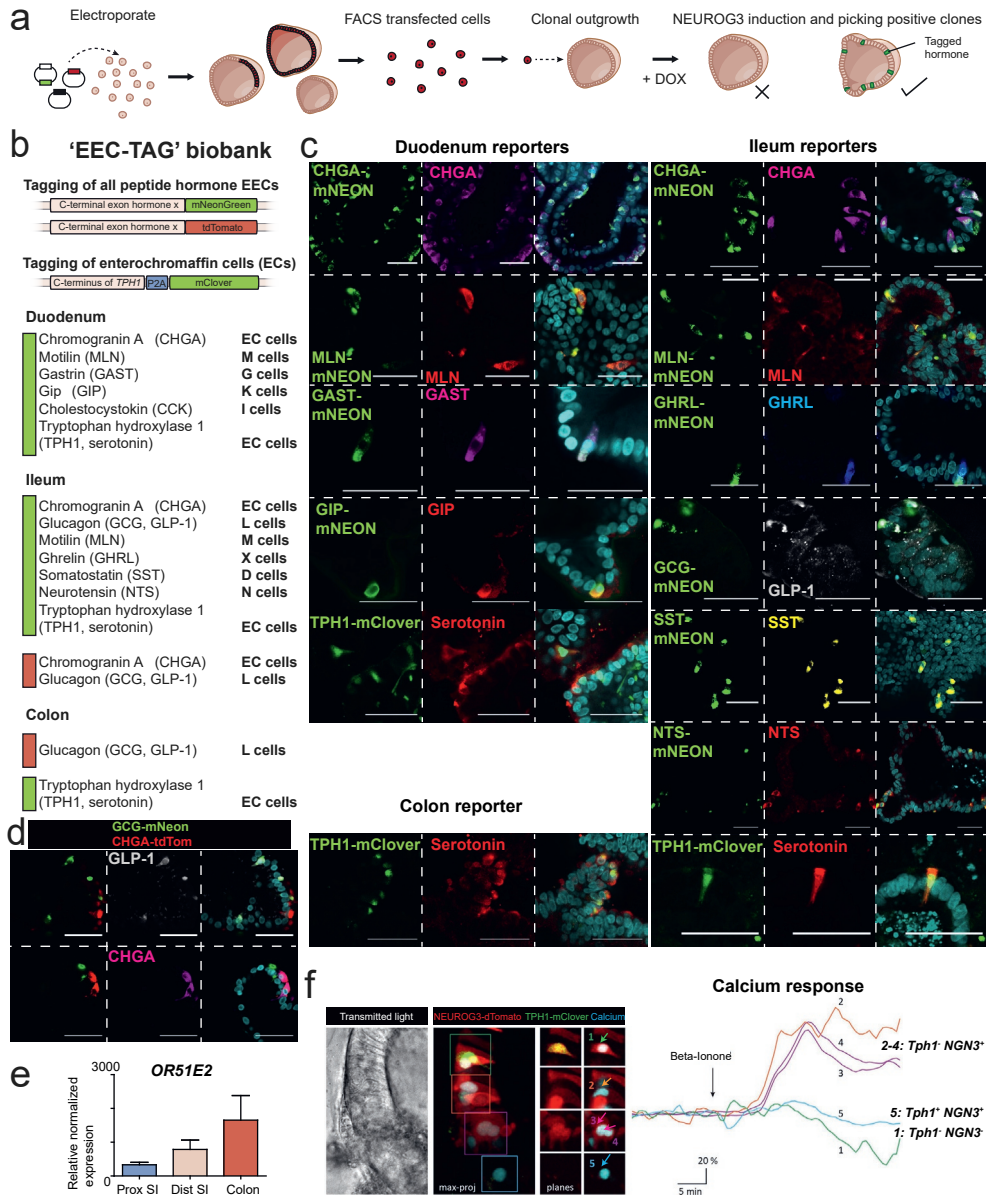
Figure 2. Generation of a human enteroendocrine cell organoid toolbox. (A) Schematic workflow of generation of reporter organoids. (B) Overview of EEC-TAG biobank comprising reporters for hormones across human duodenum, ileum and colon organoids. EECs are tagged with NHEJ (mNeon or tdTomato). TPH1-positive ECs are tagged using HDR with mClover. (C) Immunofluorescent staining confirms faithful reporter activity (knock-in left, stain middle and merge on right). Reporter expression always overlaps with the corresponding hormone. Scale bar is 50 μm. (D) Subsequent rounds of hormone tagging allows the generation of multiple-hormone reporter organoids. Immunofluorescent staining of GCG/CHGA double reporter organoid. Scale bar is 50 μm. (E) qPCR analysis showing expression of the olfactory receptor OR51E2 in different organoids enriched for EECs. (Continues on next page)

Figure 2. Generation of a human enteroendocrine cell organoid toolbox. Expression levels are normalized to GADPH and relative to control organoids without EECs. The experiment was performed in n = 2 independent experiments, and the mean expression and SEM are depicted. (F) Fluorescent image of a TPH1[mClover] NEUROG3[dTomato] organoid that is transduced with the turquoise calcium sensor Tq-Ca-FLITS, containing a nuclear localization signal. Five examples of nuclei are highlighted of which the calcium response is followed after treatment with beta-ionone, the agonist of OR51E2. EECs (marked in red) show increases in calcium flux ('2-4') with the exception of the TPH1[+] cell ('1'). The non-EEC ('5') does not show calcium increases upon beta-ionone treatment.

man EECs, making the generation of a detailed atlas from small intestinal tissue challenging. Murine EECs taken from primary tissue and from organoids are essentially identical[7,33]. We therefore exploited the human NEUROG3-induced organoids to perform single cell RNA sequencing. NEUROG3 was induced in duodenal, ileal and colon organoids in the absence or presence of BMP (to generate the crypt- and villus-'versions' of EECs, **Fig. S2**). Data from 8448 cells were generated and processed by SORT-seq[34] (**Fig. S3A,B**) and analyzed using RaceID3, a clustering method based on k-medoids[35]. After filtering, a broad intestinal cell type atlas was built from 4281 cells (**Fig. S3B,C**). This atlas contained five large clusters: CHGA-positive EECs (2255), and the following well-defined 'contaminant' lineages: FABP1-positive enterocytes (585), OLFM4-positive stem cells (113), rare MUC2-positive goblet cells (33), LYZ/MMP7-positive Paneth cells (11) and several progenitor populations (**Fig. S3D**).

Neuropeptide W (NPW) and VGF, recently observed in bulk EEC RNA sequencing[36], were broadly expressed (**Fig. S3C**). While the function of VGF remains elusive, NPW is known to increase food intake when injected in the hypothalamus[37]. Immunofluorescence confirmed protein expression of NPW by EECs in sections of human intestine (**Fig. 3A**).

All EECs and their progenitors were identified by thresholding for expression of the generic EEC marker *CHGA* and thresholding against *MUC2, FABP1, LYZ* and *OLFM4*. An EEC atlas was constructed from 2255 cells (of which 805 cells were BMP-treated) (**Fig. 3B,C, S4A**). The major clusters overlapped with their mouse equivalents, and the different EEC subtypes followed regional identity (**Fig. S4B**)[19]. The human EEC atlas significantly differed from the mouse tissue EEC atlas (**Fig. S4C**)[7]. To validate the *in vitro* EEC identities, we searched for EECs in a large single cell-dataset obtained from healthy and diseased human small intestines of various ages. Of 11302 cells represented in this dataset, we derived mRNA signatures for 39 human EECs, underscoring the challenge of deriving single EEC mRNA signatures from human ileal biopsies (**Fig. S4D**).

The largest cluster was formed by TPH1-expressing EC-cells, highly expressing CHGA (as in mouse) and representing the most frequent EEC type in vivo (**Fig. 3B,C**). ECs occurred in three 'flavors': REG4high and REG4low cells (in cluster 4), which also exist in murine intestine[19]. A third population of ECs, not observed in mice, expressed high levels of the secretogranin SCG2 and occurred mostly in proximal SI organoids (cluster 9) (**Fig. 3B,C**). All human ECs highly expressed Dopa decarboxylase (DDC) involved in serotonin biosynthesis, as well as SLC18A1, involved in sero-

5
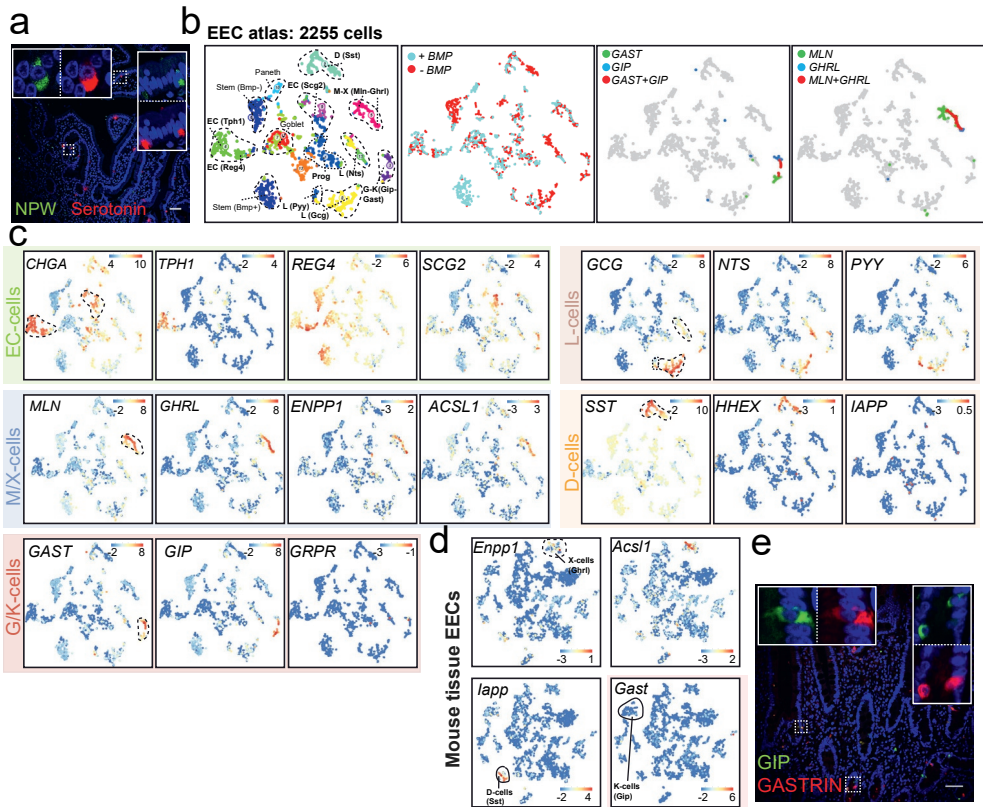
Figure 3. Single cell transcriptome atlas of human enteroendocrine cells. (A) Immuno-fluorescent staining on human intestinal section (ileum) confirms EEC-specific expression of NPW. Scale bar is 50 μm. (B) t-SNE map displaying the human EEC atlas (n= 2255 cells). Different colors represent the 13 separate clusters, and BMP treated cells are highlighted. GAST- and GIP-positive cells (defined by a minimum of 25 unique transcripts per cell) show partly overlapping expression patterns (middle t-SNE map). GHRL- and MLN -positive cells (defined by a minimum of 25 unique transcripts per cell) also overlap partly (right t-SNE map). (C) t-SNE maps displaying the expression levels of hormone and marker gene expression in the different human EEC subtypes from intestinal organoids. Bars display color-coded unique transcript expression (logarithmic scale). (D) t-SNE maps displaying the expression levels of hormone and marker gene expression in murine tissue EECs. Bars display color-coded unique transcript expression (logarithmic scale). (E) Immunofluorescent staining on duodenal sections confirms co-expression of GIP and GASTRIN. Scale bar is 50 μm.

tonin transport (**Fig. S4E**)[38]. The prototypical EC markers CHGB and GPR112 were broadly expressed by human ECs, as was the olfactory receptor OR51E1 (mouse homologue Olfr558), a marker of serotonin-producing neuroendocrine tumors in man (**Fig. S4E**)[39].

G-Cells produce Gastrin (Gast) and are largely restricted to the mouse stomach, whereas in man expression continues more distally along the GI tract in EECs[40]. Cells expressing GAST (cluster 3) co-expressed the receptor for Gastrin-releasing peptide, GRPR, a marker of G-cells in mouse stomach (**Fig. 3B,C**). GAST-expres-

sion often overlapped with high expression of the incretin GIP (same cluster 3), the main hormone product of murine K-cells. We named these cells G/K-cells (**Fig. 3B,C**). In histological sections, we confirmed overlapping expression for these two hormones (**Fig. 3E**). Some cells in cluster 3 almost exclusively produced either GIP or GAST, as seen on intestinal sections and in the atlas. The L-cell lineage clusters 8 and 13 displayed largely overlapping expression of GCG, NTS and PYY (**Fig. 3B,C**). Cluster 2 contained SST-positive D-cells, also expressing the transcription factor HHEX. HHEX has been described in murine pancreatic and intestinal Sst-producing cells[19,41]. Notably, human D-cells in tissue and organoids lacked expression of Amylin (IAPP), a peptide hormone expressed in mouse D-cells (**Fig. 3B,C, S4D**).

MLN+ cells do not exist in mice. We identified a cluster of cells producing MLN and GHRL (cluster 5). A gradient from predominantly MLN- to predominantly GHRL-expressing cells can be observed in t-SNE space (**Fig. 3C,D**). We termed these M-X or X-M cells (based on the highest expression of either MLN or GHRL, respectively) and speculate that these might represent different states of the same cell type. Indeed, BMP treatment reduced levels of GHRL, while MLN levels were slightly increased (**Fig. S6B**). M/X cells were further characterized by *ENPP1* expression*,* a known regulator of insulin responses and extracellular ATP levels[42], similarly expressed by murine X-cells (**Fig. 3C,D**). Ghrelin requires a specific acyl modification by the acyl-CoA synthetase *Acsl1* in mouse stomach X-cells[43]*.* Human M/X cells and mouse intestinal X-cells both expressed high levels of *Acsl1* (**Fig. 3C,D**).

Genes uniquely expressed by human EEC subtypes.

We next searched for EEC genes differentially expressed between human and mouse (**Fig. S4F**). The heparin-binding growth factor Midkine (MDK) was highly expressed by all human EEC types, but not by e.g. goblet cells (**Fig. 4A**). MDK is a reported biomarker of human intestinal neuroendocrine tumors[44]. Midkine has been associated with obesity and inhibits insulin signaling in adipocytes[45]. The carboxypeptidase CPB1 was produced by most EECs (highest in M/X cells), with the exception of ECs (**Fig. 4A**). Carboxypeptidases are typically involved in hormone processing[46]. Expression of CPB1 has been observed in the rat pancreas[47]. FGF14 is a human pan-EEC marker – with very limited expression in murine EECs– and belongs to a set of intracellular FGFs, that play a role in the clustering of ion channels in neurons (**Fig. 4A, S4F**)[48]. The olfactory receptor OR51E2 was sporadically expressed by different EEC subtypes, with highest levels occurring in *PYY*+ cells (**Fig. 4A**). The mouse homologue Olfr78 was lowly expressed in ECs only (**Fig. S4F**). The enzyme tryptophan 2,3-dioxygenase (*TDO2)* was found in duodenal EECs from the proximal intestine (**Fig. 4A, S4A**). *TDO2* can metabolize tryptophan through the kynurenine pathway and is one of the primary regulators of availability of this amino acid. Tryptophan is the precursor of serotonin and *TDO2* knockout mice experience increased serotonin levels[49], suggesting that *TDO2* could locally regulate serotonin production in the gut. We noted the tachykinin peptide-coding *TAC3 w*as a broadly expressed gene in human EECs, while the mouse homologue *Tac2* is not expressed in the murine intestine (**Fig. 4A**). *TAC3* codes for Neurokinin B and has been described as a regulator of secretion of gonadotropin-releasing hormone in the human hypothalamus[50]. However, the main receptor for NKB, NK$_3$ (coded by *TACR3)*, has

5
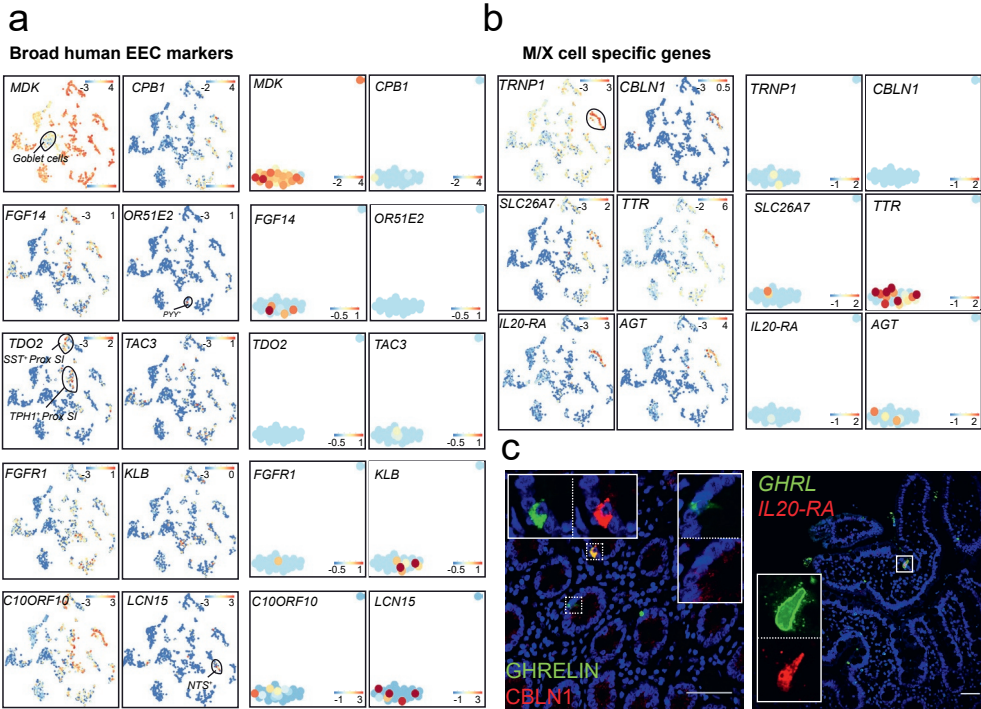
Figure 4. Novel human enteroendocrine cell markers. (A) t-SNE maps displaying the levels of hormone and marker gene expression in the different human EEC subtypes from intestinal organoids (left) and intestinal tissue (right). Bars display color-coded unique transcript expression (logarithmic scale). (B) t-SNE maps displaying the levels of hormone and marker gene in M/X cells. Bars display color-coded unique transcript expression (logarithmic scale). (C) Immunofluorescent staining on intestinal sections confirms co-expression of GHRL and the novel EEC peptide CBLN1 (left panel). Fluorescent in situ hybridization shows that GHRL+ cells express the cytokine receptor IL-20RA (right panel). All sections are from the human duodenum. Scale bar is 50 μm.

been implicated in the regulation of gastrointestinal motility[51]. The hepatokine FGF21 is a regulator of blood glucose. Several FGF21 mimetics are currently being tested for the treatment of diabetes[52]. While the receptors for FGF21 are described as a complex of FGFR1 and B-Klotho (KLB), the site of action of FGF21 is debated. We observed broad expression of FGFR1 and KLB by human EECs, suggesting that FGF21 effects could be partially mediated through the gut (**Fig. 4A**). *Fgfr1* is absent in murine EECs, while *Klb* is expressed at very low levels (**Fig. S4F**). We noted an upregulation of multiple hormones after FGF21 treatment, pointing to a potential role of signaling through FGFR1/KLB in EEC function (**Fig. S4G**). *C10ORF10* (also known as *DEPP1*) was widely expressed by human EECs. This gene is negatively regulated by insulin in liver and adipocyte tissue; its product controls the ratio between ketogenesis and gluconeogenesis[53] (**Fig. 4A**). Finally, *LCN15* was produced by *NTS+* cells (**Fig. 4A**). LCN15 is a lipocalin and one of the strongest glucose-regulated genes in Caco-2 cells[54]. Although some lipocalins have been implicated in insulin resistance, LCN15's function remains unknown.

We then focused on unique genes expressed by M/X cells, but absent in murine X-cells. *TRNP1,* involved in cortical folding in the brain[55], was the only transcription factor specific to M/X cells (**Fig. 4B**). A putative hormone, precerebellin 1 (*CBLN1),* was expressed in all M/X cells (**Fig. 4B**). *CBLN1* stimulates food intake upon intracerebroventricular injection (like Ghrelin)[56]. We confirmed CBLN1 expression in human GHRL[+] cells *in vivo* using immunofluorescence (**Fig. 4C**). We noted that M/X cells expressed the receptor for cytokines of the IL10-family (IL20-RA), an observation confirmed *in vivo* (**Fig. 4B,C**). We detected high expression of the peptide hormone Angiotensin (AGT), a regulator of blood pressure, but also of contraction of the human intestinal musculature (similar to Motilin)[57] (**Fig. 4B**). Finally, M/X cells displayed the highest expression of all EECs of the sulfate transporter *SLC26A7* and of T_4 - and Retinol-binding Transthyretin (TTR) (**Fig. 4B**). We confirmed the expression of these novel EEC genes in the (limited) number of single cell sequenced human ileal EECs (**Fig. 4A,B**).

To identify heterogeneity among the different EEC subtypes, we subclustered cells sorted from organoids carrying the individual hormone reporters. Expression of the fluorescent reporters directly correlated with the levels of the pertinent hormone transcripts within the same cell (**Fig. S5A**). A substantial number of the cells sorted for MLN-reporter expression (yet with low MLN expression) were L-cells (**Fig. S5A,B**). Surprisingly, we identified a rare subcluster of *GCG+*-reporter cells that highly expressed Pancreatic Polypeptide (*PPY*)[58], a well-described pancreas hormone involved in appetite regulation, never seen in human or mouse small intestine. We confirmed its expression and partial overlap with GLP-1 by staining on human intestinal sections (**Fig. S5B,C**).

Transcriptional networks

We analyzed expression of transcription factors known from mice to specify individual lineages (**Fig. 5A**). PAX4 specifies D/EC cells, while expression of ARX promotes all other EEC fates[59], in agreement with our expression profiles. HHEX and LMX1A defined human D and EC lineages respectively, consistent with mouse (**Fig. 5A**)[60]. The broad murine EEC transcription factors NKX2-2, PAX6, SOX4 and RFX6[7] were ubiquitously expressed in human EECs. We additionally identify ASCL1 as a broad human EEC transcription factor, absent from M/X-cells and from all mouse EECs (**Fig. 5A,B**). Ascl1 is expressed in endocrine cells in murine lung[61]. MNX1 was highly expressed by human ECs; it promotes neonatal diabetes when mutated (**Fig. 5A,B**)[62]. *MLN*[+] cells developmentally resembled murine GHRL-producing X-cells.

We chose to knockout the EC-specific *LMX1A* gene and the D-cell-specific *HHEX* gene (**Fig. 5A,C**). Organoids were transiently transfected with a Cas9-EGFP coding plasmid that included the site-specific gRNA[63]. Genotyping of clonal organoids was performed to identify homozygous loss-of-function alleles. *Lmx1a*-null mice die shortly after birth, lacking intestinal *Tph1*- and *Chga-expression,* indicative of EC loss[60]. *LMX1A*-mutant human organoids displayed a strong reduction in *TPH1* (**Fig. 5D**). We also observed a milder reduction in *SST* derived from D-cells. In contrast to mouse EECs, human *LMX1A* is also expressed in D-cells (**Fig. 5A**)

HHEX has been linked to type 2 diabetes[64]. While loss of *Hhex* in mice impairs the
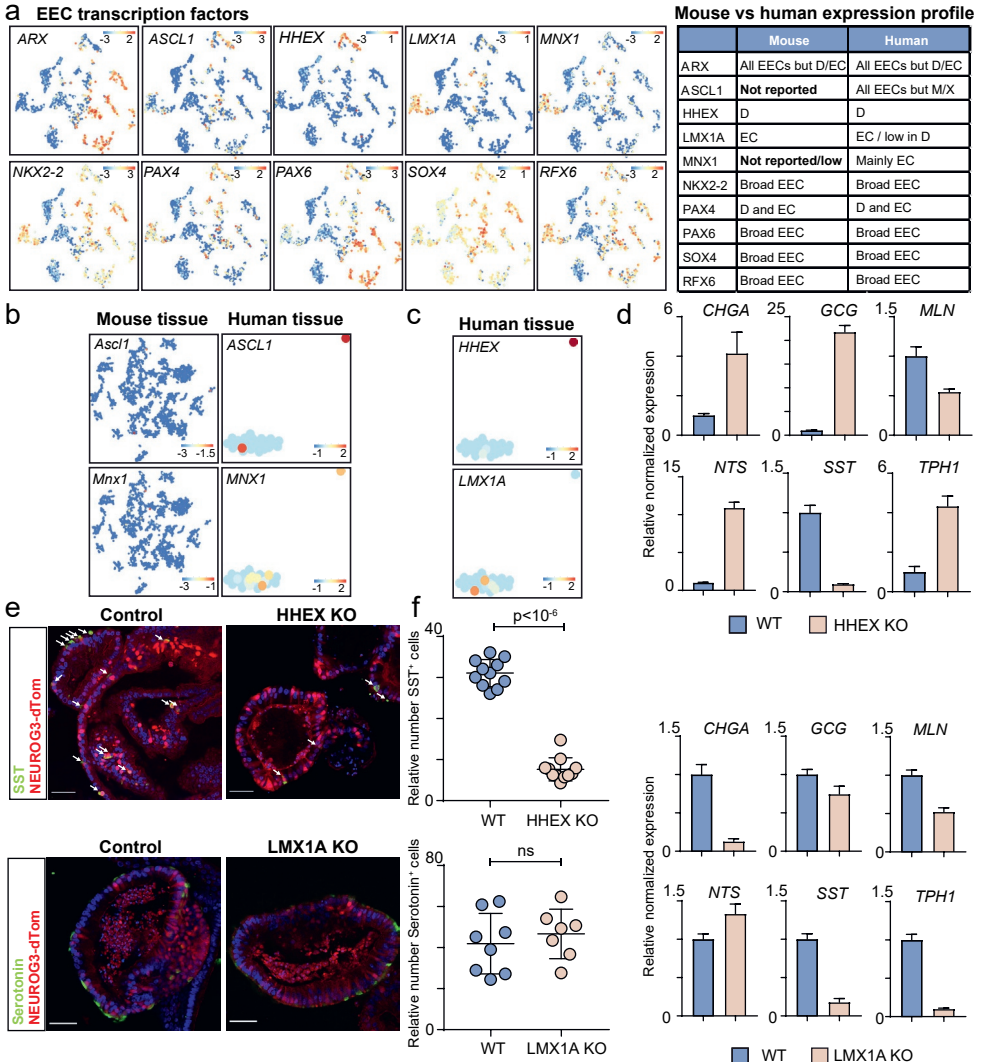
Figure 5. Transcriptional networks in human EECs. (A) t-SNE map displaying the expression level of EEC lineage transcription factors. Bars display color-coded unique transcript expression (logarithmic scale). The murine and human patterns of expression of these genes among EEC subtypes are depicted (right table). (B) t-SNE maps displaying the expression level of novel EEC transcription factor ASCL1 and MNX1 in mouse and human EECs. Bars display color-coded unique transcript expression (logarithmic scale). (C) t-SNE maps displaying the expression level of EEC transcription factors HHEX and LMX1A in human tissue EECs. Bars display color-coded unique transcript expression (logarithmic scale). (D) qPCR analysis showing expression of hormones in wildtype organoids and LMX1A and HHEX knockout (KO) organoids. Expression is normalized to GADPH, and relative to wildtype. The experiment was performed as a technical duplicate, and the mean expression and SEM are depicted. (E) Immufluorescent staining on wildtype and HHEX and LMX1A knockout organoids. D-cell are reduced after HHEX loss, while EC-cells do not decrease upon LMX1A knockout. Scale bar is 50 μm. (F) Quantification of E. Number of hormone positive cells are counted on organoid sections and shown normalized to number of tdTomato+ cells and relative to WT.

function of Sst-producing cells in pancreatic islets, effects in the murine intestinal tract were not described[65]. *HHEX* gene disruption blocked the production of *SST* (**Fig. 5D**). The most striking increase was observed in *GCG* expression (over 20-fold). In *Hhex*-knockout mice, pancreatic glucagon similarly increases[41].

Immunofluorescent staining in *HHEX* and *LMX1A* organoids revealed a 4-fold reduction in the number of SST+ D-cells upon *HHEX* loss, suggesting impaired allocation of progenitors into this lineage (**Fig. 5E,F**). Serotonin-producing cells did however not decrease in LMX1A knockout organoids, despite a decline in *TPH1* expression (**Fig. 5D-F**). We conclude that *LMX1A* is not directly important for allocation to the EC-fate*.*

BMP signaling as regulator of hormone switching in human EECs

Crypt-villus gradients were observed for human hormones such as *GCG* (**Fig. S5D**). We interrogated BMP dependency of hormone gene expression in the single cell atlas. BMP activation induced *NTS* in L-cells at the expense of *GCG* (**Fig. S5E**). Live-cell imaging of GCG-reporter organoids confirmed that BMP activation decreased reporter expression in individual L-cells (**Fig. S5F**). Additionally, we observed BMP-mediated repression of *GHRL* in M/X cells, accompanied by a mild increase in *MLN* expression (**Fig. S5E**). In murine intestine, we have found that expression of GHRL diminishes with migration of the X-cell along the crypt-villus axis[7]. Thus, human MLN/GHRL-producing EECs appeared to undergo a BMP-controlled switch in hormone expression as previously described in mouse.

High-definition transcriptomic and proteomic profiling of EECs.

Transcriptomics of pooled cells has a superior sensitivity compared to single cell RNA sequencing. We thus generated a deep transcriptomic signature of sorted and pooled EC-, L- and M-cells (**Fig. S6A**). In addition, CHGA-mNeon+ cells were sorted to generate a broad EEC signature. We identified the top 20 uniquely expressed markers from the RNA sequencing dataset for each population (**Fig. S6B**). We thus uncovered multiple EEC subtype features, unnoticed in the single cell atlas. The transcription factor IRX3, member of the Iroquois homeobox family, was one of the most defining markers of TPH1+ cells, yet has not been described in murine EECs[19] (**Fig. S6B**). IRX3 has gained attention as a neuronal regulator of energy balance, and genetic variants in IRX3 associate with obesity in humans[66].

We analyzed our bulk transcriptomic datasets for subtype-specific receptors. We noted conserved expression of receptors known from mouse EECs, including *FFAR2* (broad EEC), *GPBAR1* (L-cell), *SSTR5* (L-cell), *OR51E1* (mouse homologue Olfr558; EC), *ADGRG4* (*Gpr112;* EC) and the extracellular calcium sensor *CASR* (broad EEC)[2] (**Fig. S6C**). Human EECs expressed multiple orphan receptors, such as *GPR162* (L-cells), not found in mice (**Fig. S6C**) and reported to be expressed in brain to regulate food intake. Genetic variants in *GPR162* are linked to glucose deregulation[67]. *GPR68* is an orphan GPCR uniquely expressed by ECs (**Fig. S6C**). The orphan peptide CART (cocaine-and amphetamine-regulated protein) activates GPR68[68]; it has a role in the regulation of anxiety, reward and feeding behaviors[69]. We find broad expression of the subunit of the GABA-B receptor *GABBR2*, potentially al-

lowing a GABA response (**Fig. S7C**)[70]. We identified production of multiple hormone receptors in EECs, including the melanocortin receptor MC1R (**Fig. S6C**). MC4R in murine L-cells is a regulator of hormone secretion and can be activated by MSH-like producing bacteria[71]. ECs selectively expressed the receptor for the thyroid-stimulating-hormone, *TSHR* (**Fig. S6C**). Serotonin regulates thyroid hormone levels[72]. TSHR expression by ECs suggests that this regulation could work bidirectionally. ECs also expressed the receptor for the L-cell hormone PYY, *NPY1R* (**Fig. S6C**), reported in murine enterocytes as a regulator of electrolyte transport[11]. We did not confirm expression of *NPY1R* in the CHGA-mNeon⁻ population, which includes enterocytes. L-cells highly expressed the Secretin receptor *SCTR* as observed in our single cell atlas, but not in mice (**Fig. S6C-D**). Fluorescent in-situ-hybridization (FISH) confirmed the expression of *SCTR* in EECs *in vivo* by overlap with *CHGA* (**Fig. S6E**). Since we observed the highest expression of *SCTR* in L-cells, we measured GLP-1 secretion upon a 24-hour secretin treatment in organoids. Indeed, secretin induced GLP-1 secretion at levels comparable to forskolin as measured by ELISA, or as seen by the loss of intracellular fluorescence of GCG-neon (**Fig. S6F-H**).

We next isolated intracellular proteins for mass-spectrometry to establish subtype-specific proteomes (**Fig. S6A**). PCA-analysis revealed a clear separation of reporter populations (**Fig. S7A**). The analysis confirmed many of the novel markers for EEC populations, including the L-cell hormone PPY, EEC marker MIDKINE and the M-cell peptidase CPB1 (**Fig. S7B**). A gene identified at RNA level but not in the proteome was *CRYBA2*, a crystallin family member (**Fig. S7C**), and reported as a novel marker of human endocrine cells in pancreas and colon[34,73]. CRYBA2 protein was also absent on human intestinal sections using IHC (**Fig. S7D**). In the human genome, *CRYBA2* maps adjacent to another EEC marker gene, *FEV*[19]. *CRYBA2* and *FEV* RNAs were expressed in a virtually identical pattern in our single cell atlas (**Fig. S7C,E**).

The human EEC secretome

The human EEC organoid cultures uniquely allow proteomic analysis of hormones secreted basolaterally. We isolated supernatants of forskolin-stimulated proximal and distal small intestinal EEC organoids and control organoids and separated into >10kDa and <10kDa fractions. The latter was directly analyzed using LC-MS, whereas the former was first trypsinised (**Fig. 6A**). Proteins secreted by EEC organoids showed a large non-overlap with the bulk proteome of the different EEC populations (**Fig. 6B**). These proteins were mainly annotated to extracellular processed such as "Secretion" when compared to intracellular proteome (**Fig. 6C**). In both the >10kDa and the <10kDa (representing processed hormones) fractions, the expected EEC marker hormones were found amongst the most abundant proteins and peptides identified (**Fig. 6D**). Control organoids (not enriched for EECs) mostly secreted goblet cell products like mucins and trefoil factors. This provided strong evidence for specific hormone processing and secretion by proximal (e.g. GAST, GHRL, MLN and CCK) and distal (e.g. PYY, NTS and GCG) intestinal organoids.

Hormones known to undergo proteolytic processing were detected in the processed peptides fraction. By contrast, REG4 (biologically active as a full-length protein) was
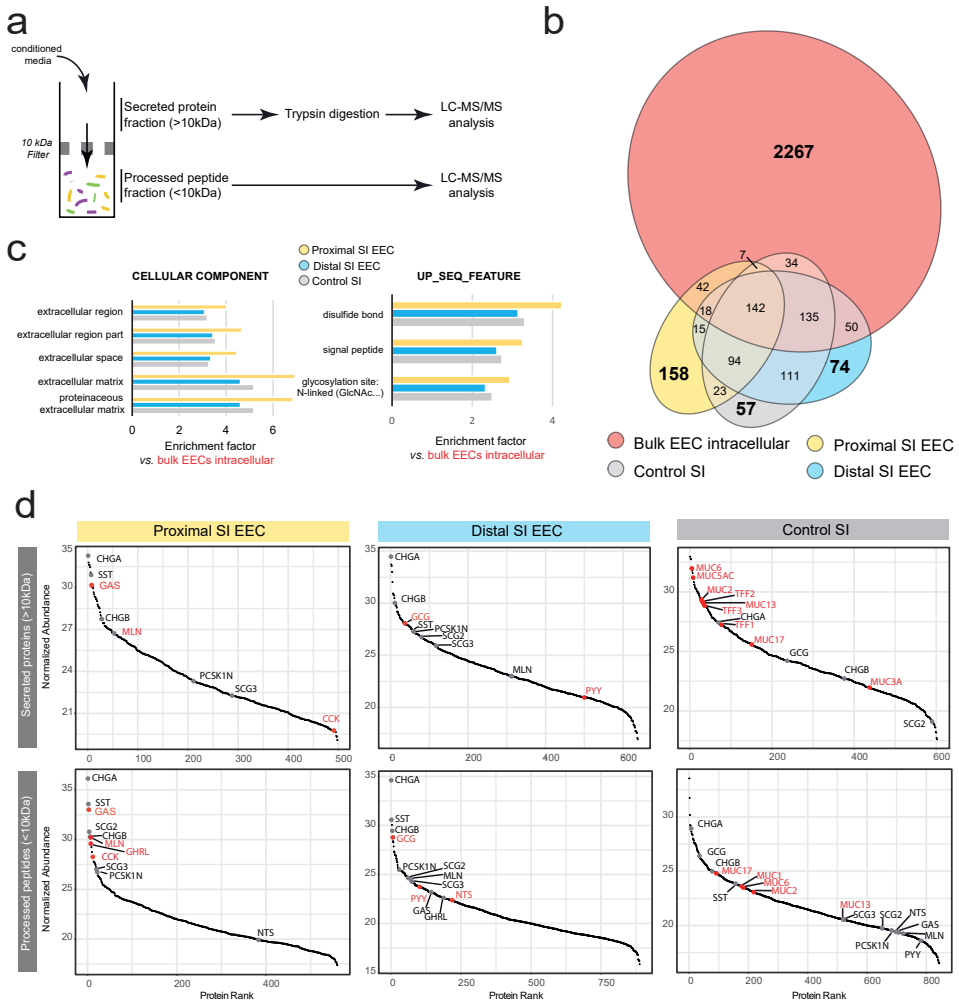
Figure 6. Human EEC secretome profiling by mass spectometry. (A) Workflow for the analysis of secreted proteins and processed peptides from conditioned media. NEUROG3 induced or control organoids were stimulated for 24 hours with Forskolin, after which conditioned media were collected. Before applying the 10 kDa cut-off filter, the conditioned media were denatured for efficient partitioning by molecular weight, and to disrupt potential protein-protein and protein-peptide interactions. (B) Venn diagram showing the overlap in protein identifications between the bulk EEC proteome (red) and the proximal EEC-enriched (orange), distal EEC-enriched (blue) and control (gray) small intestinal (SI) organoids secretomes. (C) Top enriched cellular component and sequence feature characteristics of proteins identified in the proximal EEC-enriched, distal EEC-enriched and non-induced SI organoid secretomes. The proteins detected in the bulk EEC intracellular proteome were used as reference for enrichment. (D) Ranked abundances of proteins detected in the secreted protein fraction and processed peptide fraction. EEC markers are annotated in gray. Products enriched in the different secretomes are annotated in red. Typical proximal or distal hormones are enriched in the respective organoid types, whereas control organoids uniquely secrete mucins and trefoil factors.

167
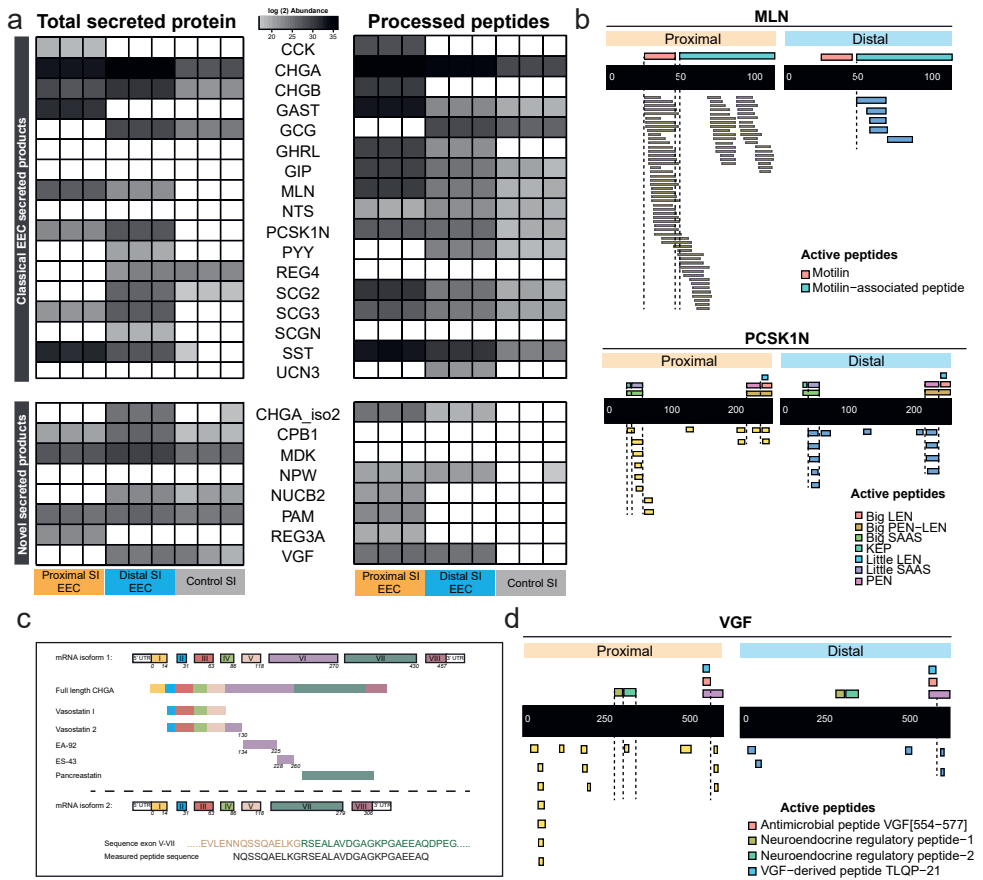
Figure 7. Processing of known and novel region-specific human EEC hormones. (A) Overview of relative abundance of known and novel secreted proteins and processed peptides. MS intensities are plotted as heatmaps for total secreted proteins (>10kDa fraction, left panel) and corresponding processed peptides (<10kDa fraction, right panel). (B) Measured peptides (<10 kDa) in the supernatant of distal SI organoids mapping to the secreted prohormone Motilin (MLN) and the PCSKN1. Data from proximal (orange) and distal SI organoid (blue) supernatants are displayed. Known bioactive peptides are shown above the black bar. (C) Alternative processing of Chromogranin A (CHGA). Compared to Isoform 1, Isoform 2 of CHGA (lower panel) lacks exon VI. Peptides spanning sequences of both exon V and VII were detected in the supernatant, indicating isoform 2 is produced in human EECs. (D) Measured peptides (<10 kDa) in the supernatant of proximal and distal SI organoids mapping to the secreted VGF protein. Other than sequences overlapping with known bioactive peptides, numerous other peptides were detected.

only found in the >10kDa fraction (**Fig. 7A**). Processed peptides generally displayed C-terminal trimming, likely due to the endogenous activity of carboxypeptidases (**Fig. 7B, S7F**). All known fragments of the proglucagon pro-hormone were observed (**Fig. S7F**). Individual fragments rarely spanned more than a single biologically active peptide. Neuronostatin, a fragment of the pro-somatostatin hormone[74], was found abundantly in the EEC secretome (**Fig. S7F**). Processed peptides were biased to-

wards bioactive fragments of multiple hormones and enzymes, including for MLN, PCSK1N and GHRL (**Fig. 7B, S7F**). The signal peptide (20-25 N-terminal amino acids of the prohormones) was consistently cleaved from all hormones (**Fig. 7B, S7F**). Apart from quantitative differences, we also detected region-specific biases in the ratio between bioactive peptides and those with no known activity. For example, duodenal bioactive peptides from proximal-enriched hormones (GHRL, MLN) were highly overrepresented.

Next, we looked for novel secreted peptides. CHGA codes for a 457-amino acid preproprotein, cleaved into many different bioactive products. A shorter second iso-form lacks exon 6 and has not been shown to be translated[75]. We now find abundant peptides spanning exon 5 and exon 7 (**Fig. 7A, C**). Consistent with the observed RNA expression, other previously unknown EEC products were also observed (**Fig. 4A, S7**). These include NPW, MIDKINE, VGF, and the peptidase CPB1 (**Fig. 7A, D**). We found expression and secretion of the antimicrobial peptide REG3A from human EECs (**Fig. 7A, S7F**). We further identified EEC-specific secretion of the enzyme PAM (Peptidyl-glycine alpha-amidating monooxygenase), which activates endocrine peptides by C-terminal amidation (**Fig. 7A, S7F**). Coding variants of PAM are asso-ciated with type 2 diabetes and can affect insulin secretion[76]. We detected the Nu-cleobindin-2 (NUC2B) precursor, processed to the neuropeptides Nesfatin-1, 2 and 3[77]. Nesfatin-1 has recently gained attention as an anorexigenic and insulinotropic peptide, produced in the hypothalamus and pancreas. Nesfatin-1 has been shown to regulate GLP-1 and GIP secretion *in vitro*[77].

Discussion

Human EECs are rare and have been largely inaccessible for *in vitro* studies. We have generated a high-resolution transcriptomic and proteomic profile of human EECs from three locations along the gastrointestinal tract, including a first assess-ment of their secreted products. This dataset yields new hormones, transcription factors and receptors, and can be mined for novel therapeutic targets. The expres-sion atlas highlights key differences with mouse. The transcriptional networks gen-erating the different EEC subtypes have been well worked out in mice[7,60,78,79]. These networks could result from a stochastically acting system that generates fixed ratios of different EECs. This would explain why organoids generate conserved ratios of EEC subtypes when compared to their tissue of origin[8]. A recent study has surveyed a broad human EEC population using antibody-based sorting approaches and bulk RNA sequencing[80] and identified novel EEC features, such as the expression of Neuropeptide W, confirmed by the current study.

We present the first transcriptomic and proteomic profiling of Motilin-producing cells. Motilin is a regulator of gut motility with intriguing evolutionary dynamics, inactivated independently in lineages leading to the mouse and rat, and guinea pigs[81]. The Mo-tilin receptor underwent a similar fate[82]. This raises questions as to how the cell type (the X-cell) that produces Motilin diverged from that point. For example, the produc-tion of a certain hormone is likely to be accompanied by the expression of dedicated sensory receptors. We found many similarities between mouse X-cells and the hu-

5

man counterparts, M-X cells. Transcription factors are conserved (**Fig. 5A**), as is expression of genes required for Ghrelin modifications such as Acsl1 (**Fig. 3C,D**). We noted important differences also, like the expression of putative hormones including CBLN1 and AGT. The latter has been proposed as motility regulator similar to Motilin (**Fig. 4**). We also identify a cytokine receptor in M-X cells, *IL-20RA,* which could link a sensory mechanism for pathogens to an expelling motility response.

The expression of receptors for some EEC hormones by EECs has been reported in mouse, particularly for Somatostatin (e.g. Sstr5 in L-cells)[83]. We now find that human EECs can sense extracellular PYY (NPY1R) and Secretin (SCTR). The PYY-receptor Npy1r has been suggested as an enterocyte marker in mouse, which we do not confirm in human[11]. Rather, we observe exclusive expression in human Serotonin-producing ECs. *SCTR* expression is low in ECs and enriched in EECs producing *GCG* and *GAST/GIP*. We show that Secretin can stimulate L-cells to secrete GLP-1. Importantly, a Secretin stimulation test is commonly used in diagnostics of Zollinger-Ellison syndrome patients that suffer from gastrin-producing tumors[84]. Secretin normally represses blood gastrin by inhibiting the secretion of gastrin from stomach G-cells (the major site of Gastrin production), likely through modulating the luminal pH. Patients suffering from small intestinal gastrinoma show sharp increases in serum gastrin upon secretin administration. Our data suggest this to occur through SCTR-expression by small intestinal gastrin-producing G-cells. More broadly, our data indicate that human EECs have an extensive capacity to cross-communicate through their hormone products.

Taken together, the EEC atlas and EEC-TAG biobank represent rich resources to identify regulators of human EEC development and function.

Acknowledgments

Author contributions

J.B., J.P. and H.C. conceptualized the project, designed the experiments, interpreted the results and wrote the manuscript. J.B.M., W.W. and A.J.R.H. performed the proteomic and secretomic experiments and analysis. A.M.S. assisted in cell culture experiments and were supervised by J.B., J.P. and H.C.; J.B. and J.P generated and analyzed the organoid-derived EEC single cell atlas. R.E., K.J., A.R., M.Z. and S.T. generated and analyzed the primary human EEC single cell dataset. G.B. generated and provided the Neurogenin-3 overexpression vector. D.H. and B.A. generated and provided the targeting vectors for CRISPR-HOT-mediated reporter organoids.

A.A.R., and M.H.G provided different constructs. A.S., C.B., Y.P. and C.P.M. contributed to functional EEC assays. F.V.D.L and J.G. provided the Tq-Ca-FLITS reporter. B.P. and H.S. performed imaging and quantification of calcium responses. Y.E.B.E. and R.V.D.L assisted with FACS experiments. K.K. provided organoid lines. C.L.I., W.J.V.D.W. and P.J.P. performed transmission electron microscopy.

Declaration of interests

H.C. is inventor on several patents related to organoid technology; his full disclosure is given at https://www.uu.nl/staff/JCClevers/.

5

Star methods

Lead contact and materials availability

Further information and requests for resources and reagents should be directed to the Lead Contact, Hans Clevers (h.clevers@hubrecht.eu).

Unique/stable reagents generated in this study are available and can be requested from the Lead Contact, a completed Materials Transfer Agreement may be required.

Experimental model and subject details

Human organoid cultures

Tissues from the human duodenum, ileum and colon were obtained from the UMC Utrecht with informed consent of each patient. The patients were diagnosed with small intestinal or colon adenocarcinoma that was resected. A sample from non-transformed, normal mucosa was taken for this study. The study was approved by the UMC Utrecht (Utrecht, The Netherlands) ethical committee and was in accordance with the Declaration of Helsinki and according to Dutch law. This study is compliant with all relevant ethical regulations regarding research involving human participants.

For immunostainings, sections of formalin-fixed, paraffin embedded human intestinal tissue were obtained from resections performed at the University Medical Center Utrecht, The Netherlands. Anonymized archival pathology material was used according to the guidelines of the UMC Utrecht's Research Ethics Committee[85].

Method details

Cell culture of human intestinal organoids

Human small intestinal cells were isolated, processed and cultured as described previously[8,17]. Instead of Wnt conditioned media, the medium was supplemented with Wnt surrogate (0,15 nM, U-Protein Express)

For differentiation towards EECs, organoids were treated with 1 μg/ml doxycycline (Sigma) in 'ENR' medium[86]. Secretin (Tocris) was used at a concentration of 10 μg/ml. Beta-ionone (Sigma) was used at 100 mg/ml. BMP activation was achieved by withdrawing Noggin from 'ENR' and addition of BMP-2 (Peprotech, 50 ng/ml) and

BMP-4 (Peprotech, 50 ng/ml). Notch signaling was inhibited by treatment with the Gamma-secretase inhibitor DAPT (Sigma, 10 µM). Wnt inhibition was performed by treatment with the Porcupine inhibitor IWP-2 (Stemcell Technologies, 5 µM). FGF21 was used at a concentration of 1 µg/ml (Peprotech).

Constructs for EEC-TAG reporter and knockout generation

The NEUROG3 was cloned in a two insert Gibson reaction into BSKS II vector. Of note, two PCR reactions were done: first, NEUROG3 was amplified from human genomic DNA, since the entire coding region lies in one exon. Second, the BSKS vector was amplified. The forward and reverse primers for Gly linker, FLAG, HA and P2A sequence were annealing to each other (**Supplementary Table 1, which can be found online at https://doi.org/10.1016/j.cell.2020.04.036**). All three DNA fragments were then combined in BSKS-NEUROG3-Flag-HA-P2A. In the next step, NEUROG3-P2A sequence was excised using EcoRI enzyme and cloned into pre-viously published pLX-NS2 vector[87]. Organoids were lentivirally transduced as de-scribed before[88].

For generation of the reporter organoid lines using CRISPR-HOT, we utilized a method described in [29]. Briefly, we used a targeting plasmid containing a fluorescent protein (mNEON or tdTomato) which can be linearized at a defined base position by a specific sgRNA and Cas9 provided from a second plasmid, which also encodes mCherry[89]. These two plasmids are co-electroporated with a plasmid encoding the sgRNA for the respective locus (**Supplementary table 1**).

The HDR donor plasmid allows C-terminal knock-in of the fluorescent reporter mClover3 in the TPH1 locus and was generated using pUC118 as a backbone. First, the endogenous SapI site in PUC118 was inactivated. Then, a selection cassette (PGK promoter driven expression of blasticidin) flanked by LoxP and two SapI sites was cloned into the SapI-inactivated pUC118 using infusion cloning (638910, Takara). Subsequently, a P2A sequence and the fluorescent protein mClover3 was PCR amplified (Phusion High fidelity DNA polymerase, M0530S, NEB) from the Ad-dgene plasmid #74252 and cloned upstream of the selection cassette using infu-sion cloning (638910, Takara) and NotI (R0189S, NEB) digestion of the pUC118 selection-cassette containing plasmid. Next, homology arms corresponding to the genomic regions, approximately 1000bp, upstream and downstream of the *TPH1* stop codon were PCR amplified (Phusion High fidelity DNA polymerase, M0530S, NEB) from genomic DNA (extracted and purified from human small intestinal organ-oid DNA). The PCR primers contained overhangs allowing subsequent Golden Gate cloning (**Supplementary Table 1**).

The PCR amplified homology arms were purified (QIAquick PCR Purification Kit, 28104, Qiagen) and finally, the targeting vector was generated by SapI (R0569S, NEB) mediated Golden Gate insertion of the homology arms into the pUC118 selec-tion-cassette containing plasmid.

The sgRNA was selected based on the WTSI website (https://www.sanger.ac.uk/htgt/wge/) and chosen as close to the TPH1 stop codon as possible. The gRNA sequence overlapped with the stop codon, so that the homology vector was not cut.

The target sequence was ordered as two complementary oligos (IDT) and cloned in the Cas9-EGFP vector (addgene plasmid #48138) following the protocol described before[90].

For the generation of *HHEX* and *LMX1A* knockout organoids, gRNAs were selected using the WTSI website and cloned in the Cas9-EGFP vector (addgene plasmid #48138) following the protocol described before[90]. gRNAs used in this story are presented in **Supplementary Table 1**.

Human intestinal organoids were transiently transfected using a NEPA21 electroporator and a previously developed protocol[91]. 3-7 days after electroporation, either mCherry (for generation of NHEJ-mediated reporter organoids) or EGFP (for generation of *HHEX* and *LMX1A* knockout lines) positive cells were sorted using a FACS-ARIA (BD Biosciences). Wnt-surrogate (0.15 nM , U-Protein Expression) and Rho kinase inhibitor (Y-27632 dihydrochloride; 10 µM, Abmole) were added to the culture medium up to 1 week after sorting to enhance single cell outgrowth. All reporter organoids were generated in organoid lines also transduced with *NEUROG3*-overexpression (with or without dTomato) vector.

For generation of CRISPR-HOT reporter organoids: Cells were transfected 1) with a gRNA targeting the hormone locus near its stop codon, 2) a vector encoding mNeon or tdTomato and 3) a vector encoding Cas9, a constitutively produced mCherry fluorescent molecule and a gRNA linearizing the vector encoding the fluorescent molecule. Five days later, transfected cells were sorted for mCherry and plated as single cells. After two weeks, NEUROG3 was induced in the resulting clonal organoids to visualize expression of the fluorescent fusion hormones. Typically, the first fluorescent organoids appeared 2-3 days later and were then clonally expanded. Organoids where fluorescent cells appeared during EEC differentiation were picked, digested using TrypLE (TryplE Express; Life Technologies) and clonally expanded to establish stable knock-in organoid lines.

Organoids grown from Cas9-EGFP transfected cells were genotyped for *HHEX* and *LMX1A* to confirm homozygous frameshift mutation (primers in Supplementary Table 1).

Calcium sensor

A red calcium probe (pTorPE-R-GECO1, addgene plasmid #32465) was used as a template to engineer a cyan genetically encoded calcium probe. The cpApple was replaced with a circular permuted mTurquoise. The resulting probe was dubbed Tq-Ca-FLITS (Turquoise Calcium Fluorescence Lifetime Indicator for Truthful Sensing). A triple nuclear localization signal (3xnls) was added to the N-terminus of the calcium probe to simplify analysis. Details of the engineering and characterization will be described elsewhere (Van der Linden et al., in preparation).

PCRs were performed on Tq-Ca-FLITS (Fw AAACAAGCGGGAGACGTGGAG-GAAAACCCTGGACCTCTCGAGatgggatcagatccaaaaaagaagag, Rev ATGG-CACTAGGCTAGTTCTAGAcCTACTTCGCTGTCATCATTTGGAC) as well as H2B-mMaroon (Fw TCGGCGCGCCACGCGT, Rev CGTCTCCCGCTTGTTTCAG-

TAGACTAAAATTCGTCGCGCCAGATCCGCTAGCattaagtttgtgcccc) and the two PCRs were cloned into a lentiviral vector using InPhusion Cloning (Takara), to produce H2B-mMaroon-P2A- Tq-Ca-FLITS, two simultaneously expressed cistrons separated by a de-optimized P2A[92].

Live cell imaging of calcium reporter organoids

H2B-mMaroon-P2A-Tq-Ca-FLITS organoids were imaged on a Leica SP8 confocal laser scanning microscope, equipped with Argon laser and White Light Laser, the latter allowing spectral flexibility for optimal visualization of all fluorophores. For cell type identification, cells were first imaged in 5 channels (Tq-Ca-FLITS-mTurquoise2, Clover, TdTomato, H2B-Maroon and transmitted light) and subsequently Tq-Ca-FLITS and H2B-mMaroon were time lapse imaged during administration of beta-ionone in XYZT-mode. Post-acquisitional analysis was done with custom-made Fiji-script.

Transmission electron microscopy

Organoids were fixed with 1.5% glutaraldehyde in 0.1M cacodylate buffer. They were kept in the fixative for 24 h at 4ºC. Then, they were washed with 0.1M cacodylate buffer and potsfixed with 1% osmium tetroxide in the same buffer containing 1.5% potassium ferricyanide for 1 hour (dark) at 4ºC. Then the samples were dehydrated in ethanol, infiltrated with Epon resin for 2 days, embedded in the same resin and polymerised at 60ºC for 48 hours. Ultrathin sections were obtained using a Leica Ultracut UCT ultramicrotome (Leica Microsystems, Vienna) and mounted on Formvar-coated copper grids. They were stained with 2% uranyl acetate in water and lead citrate. Then, sections were observed in a Tecnai T12 electron microscope equipped with an Eagle 4kx4k CCD camera (Thermo Fisher Scientific, The Netherlands).

Alternatively, organoids were chemically fixed at 4 ºC with a mixture of 2% paraformaldehyde and 0.2% glutaraldehyde in PB buffer. After washing with PB containing 50 mM glycine, cells were embedding in 12% gelatine and infused in 2.3 M sucrose. Mounted gelatine blocks were frozen in liquid nitrogen. Thin sections were prepared in an ultracryomicrotome (Leica EM Ultracut UC6/FC6, Leica Microsystems, Vienna, Austria). Ultrathin cryosections were collected with 2% methylcellulose in 2.3 M sucrose. The observations were performed in an Electron Microscope Tecnai T12 as mentioned.

Immunostaining

Organoids were stained as described before[8]. Primary antibodies used were goat anti-chromogranin A (1:500; Santa Cruz), goat anti-cholestocystokin (sc-21617,1:100; Santa Cruz), rabbit anti-neurotensin (sc-20806,1:100; Santa Cruz), goat anti-somatostatin (sc-7819, 1:100; Santa Cruz), goat anti-serotonin (ab66047, 1:1,000, Abcam), rabbit anti-gastric inhibitory polypeptide (ab22624-50, 1:500; Abcam), goat anti-GLP1 (sc-7782, 1:100; Santa Cruz), rabbit anti-GLP1 (ab22625, 1:200; Abcam), rabbit anti-MLN (HPA069392, 1:200, Atlas antibodies), mouse anti-Gastrin (60346, 1:200, Proteintech), mouse anti beta-Catenin (610154, 1:100; BD transduction laboratories), goat anti-Ghrelin (sc-10368, 1:200; Santa Cruz), rabbit anti-Neuropeptide

W (NBP2-57337, 1:100; Novus), rabbit anti-Precerebellin (ABN304, 1:100; Sigma) and rabbit anti-PPY (HPA032122, 1:200; Atlas antibodies). Organoids were incubated with the corresponding secondary antibodies Alexa488-, 568- and 647-conjugated anti-rabbit and anti-goat (1:1,000; Molecular Probes) in blocking buffer containing 4',6-diamidino-2-phenylindole (DAPI; 1;1,000, Invitrogen). Sections were embedded in Vectashield (Vector Labs) and imaged using a Sp8 confocal microscope (Leica). Image analysis was performed using ImageJ software.

Fluorescent in situ hybridization

FISH was performed using the RNAScope® Multiplex Fluorescent Reagent Kit v2 (Advanced Cell Diagnostics) according to the manufacturer's protocol[93]. In brief, paraffin embedded ileal surgical sections were deparaffinized, treated with hydrogen peroxide for 10 minutes and boiled in target retrieval buffer for 15 minutes before a 30-minute protease treatment. Probes directed against *CHGA/SCTR, CHGA/GCG and GHRL/ IL20RA* were multiplexed, respectively, amplified and detected using fluorescent probes based on opal dyes. Slides were counterstained with DAPI for 30 seconds, mounted using ProLong™ Gold Antifade Mountant (Thermo Fisher scientific) and images were obtained using a SP8 confocal fluorescent microscope (Leica).

ELISA

The supernatant from organoids either cultured in ENR for 5 days or differentiated towards EECs were collected after a 24-hour stimulation with 10 µM forskolin (Tocris). GLP-1 concentration was measured using a GLP-1 EIA kit (Rab0201 from Sigma that detects both full-length and N-terminal cleaved GLP-1) following the manufacturer's protocol.

RNA isolation and quantitative PCR

Organoid RNA was isolated using a RNAeasy kit (QIAGEN), following the manufacturer's protocol. Quantitative PCR (qPCR) analysis was performed using biological and technical duplicates as described described before[94]. Primers were designed using the NCBI primer design tool, tested using a standard curve, and are presented in **Supplementary Table 1**.

Processing human intestinal tissue for single cell RNA sequencing

Human intestinal mucosal biopsies were obtained from patients undergoing colonoscopy at Addenbrooke's Hospital, Cambridge, UK. All patients gave informed consent for extra biopsy samples to be taken for research use when undergoing elective colonoscopy (REC 17/EE/0265). Only those patients with a macroscopically normal mucosa and subsequent histological confirmation of a normal mucosa with no intestinal pathology were included in this study.

Once acquired, biopsies were immediately placed into Hanks Buffered Saline Solution (HBSS) and washed three times in fresh HBSS. Biopsies were then placed into an HBSS solution containing 1.07 Wünsch units/ml Liberase DH (Roche) and 70 U/

ml hyaluronidase (Merck) and incubated at 37 °C for 15 minutes whilst on a plate shaker at 750 rpm. The samples were then mechanically disrupted by pipetting the solution up and down using a p1000 pipette, and then incubated at 37 °C for a further 15 minutes on a plate-shaker at 750rpm. The samples were then washed three times by pelleting the cells using centrifugation at 400g for 4 minutes, removal of the supernatant and resuspending in DMEM/F12 (Thermofisher). On the third wash the cells were left in suspension, and a 10 µL aliquot was placed into a Countess ® Automated Cell Counter to estimate the cellular concentration.

3,000 cells suspended in DMEM/F12 (Thermofisher) were loaded into an individual channel of a 10x single cell chip as per the manufacturer's protocol (version 2, 3') and run in standard conditions by the chromium controller (10X genomics). cDNA libraries were prepared according to the manufacturer's protocol and sequenced on an Illumina Hi-seq 4000 (2x50bp paired-end reads).

Raw sequence reads in FASTQ format were aligned to the human transcriptome GRCh38-1.2.0 using Cellranger v2.1.1(10x Genomics) with default parameters. Scanpy v1.4[95] was used for data processing, quality control and dimensionality reduction. We excluded cells with less than 200 genes and genes expressed in fewer than 3 cells. After quality control, the dataset contained 11,302 cells. EECs were identified and subclustered based on the expression of at least 10 CHGA counts per cell.

Single cell sorting for RNA sequencing from organoids

Organoids were dissociated to single cells using a 10-minute incubation with TrypLE (TrypLE Express; Life Technologies) and repeated mechanical disruption by pipetting. Cells were sorted using a BD FACS Aria (BD Biosciences) based on fluorescence levels. For single cell RNA sequencing, individual cells were collected in 384-well plates with ERCC spike-ins (Agilent), reverse transcription primers and dNTPs (both Promega). Single cell sequencing was performed according to the Sort-seq method[34]. Sequencing libraries were generated with TruSeq small RNA primers (Illumina) and sequenced paired-end at 60 and 26 bp read length, respectively, on the Illumina NextSeq.

For bulk RNA sequencing, cells were sorted into Eppendorf tubes containing RLT buffer (RNeasy kit, QIAGEN). 5,000 – 30,000 cells were sorted per reporter in duplicates (and triplicates for tdTomato negative cells). RNA was extracted using the RNeasy mini kit (QIAGEN) following the manufacturer's instructions. Sequencing libraries were generated using a modified CELseq2 protocol[96]. 75 bp paired-end sequencing of libraries was performed on an Illumina NextSeq platform.

Single cell RNA sequencing analysis from organoids

Reads were mapped to the human GRCh37 genome assembly. Sort-seq read counts were filtered to exclude reads with identical library-, cell- and molecule barcodes. UMI counts were adjusted using Poisson counting statistics[34]. Cells with fewer than 2,000 unique transcripts were excluded from further analysis. The remaining cells in the EEC atlas were derived from the following sources: 1446 cells from duodenum, 2145 cells from ileum, 690 cells from colon.

Subsequently, RaceID3 was used for k-medoids based clustering (knn = 10) of cells and differential gene expression analysis between clusters using the standard settings described at https://github.com/dgrun/RaceID3_StemID2_package.

The dataset was then subsetted to require expression of EEC markers and exclude cells based on expression of markers of other cell types with the following transcript count cutoffs: CHGA > 5; MUC2 < 5; FABP1 < 15; LYZ < 15; OLFM4 < 10. The resulting set of EECs was again subjected to clustering (knn = 5) and differential gene expression as described above.

For reporter analyses, cells sorted by fluorescent reporter positivity were analyzed as one dataset per reporter to gain more detailed insights into single EEC subpopulations. The following deviations from standard settings were made per reporter: GCG: knn = 5; outlg = 1; probthr = 0.00001; perplexity = 10; MLN: knn = 10; probthr = 0.0000001; SST: knn = 10; perplexity = 20 .

For mouse validation, the tissue-derived single cell count tables from Gehart el al. (2019) were reanalyzed using the procedure and settings described above. No subsetting for EECs was performed.

Bulk RNA sequencing analysis

Reads were mapped to the human GRCh37 genome assembly. The counted reads were filtered to exclude reads with identical library- and molecule barcodes. Differential gene expression analysis was performed using the DESeq2 package[97]. For display in heatmaps, genes were ranked by fold change compared against tdTomato negative cells. After filtering for an adjusted p-value < 0.05, the row z-score for the top 20 genes was calculated.

Preparation of secreted peptides and proteins for LC-MS

Organoids differentiated for 5 days to EECs were washed extensively in PBS and stimulated with 10 µM forskolin (Tocris). Conditioned media was collected for 24 hours and supplemented with 1x cOmplete Protease Inhibitor Cocktail on harvest (Roche). Potential cell debris was removed by centrifugation at 10,000 x *g*, for 5 min at 4 °C. Conditioned media supernatant was denatured in final 4 M Urea, 50 mM ammonium bicarbonate and fractionated by molecular weight with a 10 kDa Vivaspin centrifugal device (Sartorius, Göttingen, Germany), at 12,000 x *g*, for 10 min at 4 °C. (i) Endogenously processed peptides recovered from the filtrate were acidified to 5% formic acid, desalted by reversed phase C18 1cc columns (Waters Corporation, Milford, USA), further purified by home-made strong cation exchange STAGE tip, and dried by vacuum centrifugation. (ii) Longer secreted proteins in the 10 kDa retentate were recovered and diluted to final 2 M Urea, 50 mM ammonium bicarbonate, for reduction with dithiothreitol, alkylation with iodoacetamide, and overnight digestion with trypsin (Promega, Madison, USA) at 37 °C. Digested peptides were similarly acidified to 5% formic acid, desalted by reversed phase C18 1 cc columns (Waters), and dried by vacuum centrifugation.

Preparation of FACS-sorted EECs for proteome analyses

FACS sorted enteroendocrine cells were lysed in 8M Urea, 50 mM Ammonium bi-carbonate, 0.5 % Sodium deoxycholate, 1x cOmplete protease inhibitor, 50 µg/mL DNAse I, and sonicated with the Biorupter (3 cycles, 20 s on, 20 s off at 4°C) (Diagenode, Liege, Belgium). Cell debris was pelleted by centrifugation at 14,000 x *g* for 1 hour at 15 °C, and supernatant containing extracted proteins were reduced, alkylated, diluted 4 times with 50 mM ammonium bicarbonate, and digested sequentially with Lys-C (Wako) and trypsin (Promega). Peptide digests were quenched to 5% formic acid, and sodium deoxycholate was precipitated and removed by centrifugation at 14,000 x g, 4 °C for 10 minutes. Peptides in the supernatant were diluted to final 20% acetonitrile and purified by SCX STAGE tips. Eluted peptides were dried by vacuum centrifugation.

LC-MS

Peptides were reconstituted in 2% FA for LC-MS injection. Data was acquired using an UHPLC 1290 system (Agilent, California, USA) coupled to an Orbitrap HF-X mass spectrometer (Thermo Scientific, Massachussets, USA). Peptides were first trapped in a 2 cm × 100 µM Reprosil C18 trap column (Dr Maisch, Ammerbuch, Germany) of 3 µM pore size for 5 min in solvent A (0.1% formic acid in water). After trapping, samples separated in an analytical column (Agilent Poroshell, EC-C18, 2.7 µm, 50 cm × 75 µm) using a gradient of 0.1% formic acid in 80% acetonitrile (solvent B). Depending on total peptide input, species complexity, and elution profiles, different LC gradient lengths were used for FACS-sorted whole proteomes (35min, 13-40% solvent B), secreted proteins (65min, 13-40% solvent B) and endogenously processed peptides (95min, 13-44% solvent B). MS acquisition was performed in data-dependent mode. Full scans (MS1) were acquired from 375 to 1600 m/z at resolution 60,000, with 20 ms injection time and $3x10^6$ AGC target value. The TOP 15 most intense precursor ions where selected for fragmentation using 1.4 m/z isolation window. Isolated precursors were fragmented using high energy C-trap dissociation (HCD) at normalized collision energy (NCE) of 27%. MS2 scans were acquired at resolution 30,000, with 50 ms injection time and an AGC target value of $1x10^5$. Exclusion times were set to 8, 12 or 16 seconds for proteomics, digested retentates and endogenously processed peptides respectively.LC-MS injection loads were adjusted to the sample of least quantity, such that all LC-MS measurements to be compared were matched in TIC intensity.

Proteomics data analysis

Collected spectral data was processed using Proteome Discoverer 2.3 (Thermo Scientific, Massachussets, USA), and searched using Sequest HT search engine, against UniProt *Homo sapiens* database (173235 entries, downloaded in August 2019, including common contaminants). Precursor and fragment mass tolerance were set to 10 ppm and 0.02 Da respectively. Protein N-terminal Acetylation and methionine oxidation were set as variable modifications. For the endogenously processed peptides, peptides of length 6 - 50 amino acids were generated from the database upon unspecific cleavage. For digested retentates and analyses of the FACS-sorted EECs proteomes, cysteine carbamidomethylation was set as fixed

modification and up to 2 missed trypsin cleavages were allowed. Identified peptides were filtered to 1% FDR using the Percolator algorithm[98].

In proteome analyses, intensities of proteins detected in 2 out of 3 replicates in at least one group were *log(2)* transformed and missing values were imputed from the normal distribution independently for each sample using Perseus software (v_1.6.2.2)[99]. Processed data was assessed for statistical significance among the groups using One-way ANOVA or Student's T-test, and resulting *p*-values were corrected for type I error using Benjamini-Hochberg approach (q-value). Tukey Honest Significant Difference test was performed to assess difference between the groups when required. Hormone processing plots were generated from peptides identified at high confidence (1 % FDR) in at least 2 out of 3 replicates. For novel secreted products, peptides identified in 3 out of 3 replicates and not observed in the secretome of control organoids were used. Peptide sequences identified based on retention time alignment (not supported by spectral evidence) were not used in hormone processing plots. Statistical analysis and plots were generated using *in-house* built R scripts (R version 3.6.0). Gene ontology enrichments were performed with Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8[100], using all the proteins identified in the bulk EEC proteome as reference list.

Quantification and statistical analysis

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to the sample allocation during experiments and outcome assessment. All data are presented as mean ± standard error of the mean (SEM), unless stated otherwise. Value of n is always displayed in the figure as individual data points, and in the legends. Statistical tests included unpaired two-tailed t-test for **Figures 5F** and **Supplementary Figure 1G**

Data and code availability

All bulk and single cell RNA sequencing data of this study have been deposited in the Gene Expression Omnibus (GEO) under accession code GSE146799.

The raw MS data is deposited in PRIDE, with accession number PXD017468.

References

1. Gribble, F. M. & Reimann, F. Signalling in the gut endocrine axis. Physiology and Behavior Preprint at https://doi.org/10.1016/j.physbeh.2017.02.039 (2017).

2. Furness, J. B., Rivera, L. R., Cho, H.-J., Bravo, D. M. & Callaghan, B. The gut as a sensory organ. Nature Reviews Gastroenterology & Hepatology 1010, 729–740 (2013).

3. Sharma, D., Verma, S., Vaidya, S., Kalia, K. & Tiwari, V. Recent updates on GLP-1 agonists: Current advancements & challenges. Biomedicine and Pharmacotherapy Preprint at https://doi.org/10.1016/j.biopha.2018.08.088 (2018).

4. Worthington, J. J., Reimann, F. & Gribble, F. M. Enteroendocrine cells-sensory sentinels of the intestinal environment and orchestrators of mucosal immunity. Mucosal Immunology Preprint at https://doi.org/10.1038/mi.2017.73 (2018).

5. Barker, N. et al. Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature 449, 1003–1007 (2007).

6. Engelstoft, M. S., Egerod, K. L., Lund, M. L. & Schwartz, T. W. Enteroendocrine cell types revisited. Current Opinion in Pharmacology Preprint at https://doi.org/10.1016/j.coph.2013.09.018 (2013).

7. Gehart, H. et al. Identification of Enteroendocrine Regulators by Real-Time Single-Cell Differentiation Mapping. Cell (2019) doi:10.1016/j.cell.2018.12.029.

8. Beumer, Joep. et al. Enteroendocrine cells switch hormone expression along the crypt-to-villus BMP signaling gradient. Nat Cell Biol 20, (2018).

9. Goldspink, D. A., Reimann, F. & Gribble, F. M. Models and Tools for Studying Enteroendocrine Cells. Endocrinology Preprint at https://doi.org/10.1210/en.2018-00672 (2018).

10. Nguyen, T. L. A., Vieira-Silva, S., Liston, A. & Raes, J. How informative is the mouse for human gut microbiota research? DMM Disease Models and Mechanisms (2015) doi:10.1242/dmm.017400.

11. Goldspink, D. A., Reimann, F. & Gribble, F. M. Models and Tools for Studying Enteroendocrine Cells. Endocrinology Preprint at https://doi.org/10.1210/en.2018-00672 (2018).

12. Sinagoga, K. L. et al. Deriving functional human enteroendocrine cells from pluripotent stem cells. Development (2018) doi:10.1242/dev.165795.

13. McCracken, K. W. et al. Modelling human development and disease in pluripotent stem-cell-derived gastric organoids. Nature (2014) doi:10.1038/nature13863.

14. Zhang, X. et al. A Comprehensive Structure-Function Study of Neurogenin3 Disease-Causing Alleles during Human Pancreas and Intestinal Organoid Development. Developmental Cell (2019) doi:10.1016/j.devcel.2019.05.017.

15. Chang-Graham, A. L. et al. Human Intestinal Enteroids With Inducible Neurogenin-3 Expression as a Novel Model of Gut Hormone Secretion. Cellular and Molecular Gastroenterology and Hepatology (2019) doi:10.1016/j.jcmgh.2019.04.010.

16. Chang-Graham, A. L. et al. Human Intestinal Enteroids With Inducible Neurogenin-3 Expression as a Novel Model of Gut Hormone Secretion. Cellular and Molecular Gastroenterology and Hepatology (2019) doi:10.1016/j.jcmgh.2019.04.010.

17. Sato, T. et al. Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. Gastroenterology 141, 1762–1772 (2011).

18. Parikh, K. et al. Colonic epithelial cell diversity in health and inflammatory bowel disease. Nature (2019) doi:10.1038/s41586-019-0992-y.

19. Haber, A. L. et al. A single-cell survey of the small intestinal epithelium. Nature 551, 333–339 (2017).

20. Reimann, F. et al. Glucose Sensing in L Cells: A Primary Cell Study. Cell Metabolism (2008) doi:10.1016/j.cmet.2008.11.002.

21. Parker, H. E., Habib, A. M., Rogers, G. J., Gribble, F. M. & Reimann, F. Nutrient-dependent secretion of glucose-dependent insulinotropic polypeptide from primary murine K cells. Diabetologia (2009) doi:10.1007/s00125-008-1202-x.

22. Sommer, C. A. & Mostoslavsky, G. Rna-seq analysis of enteroendocrine cells reveals a role for fabp5 in the control of gip secretion. Molecular Endocrinology (2014) doi:10.1210/me.2014-1194.

23. Gong, S. et al. A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. Nature (2003) doi:10.1038/nature02033.

24. Engelstoft, M. S. et al. Seven transmembrane G protein-coupled receptor repertoire of gastric ghrelin cells. Molecular Metabolism (2013) doi:10.1016/j.molmet.2013.08.006.

25. Engelstoft, M. S. et al. Research Resource: A Chromogranin A Reporter for Serotonin and Histamine Secreting Enteroendocrine Cells. Molecular Endocrinology 29, 1658–1671 (2015).

26.   Schmid-Burgk, J. L., Höning, K., Ebert, T. S. & Hornung, V. CRISPaint allows modular base-specific gene tagging using a ligase-4-dependent mechanism. Nature Communications (2016) doi:10.1038/ncomms12338.

27.   He, X. et al. Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. Nucleic Acids Research (2016) doi:10.1093/nar/gkw064.

28.   Bukhari, H. & Müller, T. Endogenous Fluorescence Tagging by CRISPR. Trends in Cell Biology (2019) doi:10.1016/j.tcb.2019.08.004.

29.   Artegiani, B. et al. Fast and efficient generation of knock-in human organoids using homology-independent CRISPR/Cas9 precision genome editing. Nature Cell Biology (2020).

30.   Fleischer, J., Bumbalo, R., Bautze, V., Strotmann, J. & Breer, H. Expression of odorant receptor Olfr78 in enteroendocrine cells of the colon. Cell and Tissue Research 361, 697–710 (2015).

31.   Jovancevic, N. et al. Odorant receptor 51E2 agonist β-ionone regulates RPE cell migration and proliferation. Frontiers in Physiology 8, (2017).

32.   Pietraszewska-Bogiel, A., van Weeren, L. & Goedhart, J. Seeing cells smell: Dynamic optical measurements of Ca²⁺ and cAMP signaling from Olfactory Receptors transiently expressed in HEK293TN cells. bioRxiv 771261 (2019) doi:10.1101/771261.

33.   Grün, D. et al. Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature 525, 251–5 (2015).

34.   Muraro, M. J. et al. A Single-Cell Transcriptome Atlas of the Human Pancreas. Cell Systems 3, 385–394 (2016).

35.   Herman, J. S., Sagar & Grün, D. FateID infers cell fate bias in multipotent progenitors from single-cell RNA-seq data. Nature Methods (2018) doi:10.1038/nmeth.4662.

36.   Roberts, G. P. et al. Comparison of human and murine enteroendocrine cells by transcriptomic and peptidomic profiling. in Diabetes (2019). doi:10.2337/db18-0883.

37.   Levine, A. S., Winsky-Sommerer, R., Huitron-Resendiz, S., Grace, M. K. & De Lecea, L. Injection of neuropeptide W into paraventricular nucleus of hypothalamus increases food intake. American Journal of Physiology - Regulatory Integrative and Comparative Physiology (2005) doi:10.1152/ajpregu.00638.2003.

38.   Lohoff, F. W. et al. Variations in the vesicular monoamine transporter 1 gene (VMAT1/SLC18A1) are associated with bipolar I disorder. Neuropsychopharmacology (2006) doi:10.1038/sj.npp.1301196.

39.   Cui, T. et al. Olfactory receptor 51E1 protein as a potential novel tissue biomarker for small intestine neuroendocrine carcinomas. European journal of endocrinology / European Federation of Endocrine Societies (2013) doi:10.1530/EJE-12-0814.

40.   Engelstoft, M. S., Egerod, K. L., Lund, M. L. & Schwartz, T. W. Enteroendocrine cell types revisited. Current Opinion in Pharmacology Preprint at https://doi.org/10.1016/j.coph.2013.09.018 (2013).

41.   Zhang, J., McKenna, L. B., Bogue, C. W. & Kaestner, K. H. The diabetes gene Hhex maintains ??-cell differentiation and islet function. Genes and Development 28, 829–834 (2014).

42.   Di Paola, R. et al. ENPP1 affects insulin action and secretion: Evidences from in vitro studies. PLoS ONE (2011) doi:10.1371/journal.pone.0019462.

43.   Bando, M. et al. High incorporation of long-chain fatty acids contributes to the efficient production of acylated ghrelin in ghrelin-producing cells. FEBS Letters 590, 992–1001 (2016).

44.   Edfeldt, K. et al. DcR3, TFF3, and Midkine Are Novel Serum Biomarkers in Small Intestinal Neuroendocrine Tumors. Neuroendocrinology (2017) doi:10.1159/000452891.

45.   Fan, N. et al. Midkine, a potential link between obesity and insulin resistance. PLoS ONE. 9, (2014).

46.   Sapio, M. R. & Fricker, L. D. Carboxypeptidases in disease: Insights from peptidomic studies. Proteomics - Clinical Applications vol. 8 327–337 Preprint at https://doi.org/10.1002/prca.201300090 (2014).

47.   Yu, S. L. et al. Phosphorylation of carboxypeptidase B1 protein regulates ß-cell proliferation. International Journal of Molecular Medicine (2017) doi:10.3892/ijmm.2017.3141.

48.   Pablo, J. L. & Pitta, G. S. FGF14 is a regulator of KCNQ2/3 channels. Proc Natl Acad Sci U S A (2017) doi:10.1073/pnas.1610158114.

49.   Too, L. K. et al. Deletion of TDO2, IDO-1 and IDO-2 differentially affects mouse behavior and cognitive function. Behavioural Brain Research 312, 102–117 (2016).

50.   Sanger, G. J. Neurokinin NK 1 and NK 3 receptors as targets for drugs to treat gastrointestinal motility disorders and pain. British Journal of Pharmacology vol. 141 1303–1312 Preprint at https://doi.org/10.1038/sj.bjp.0705742 (2004).

51. Sanger, G. J. Neurokinin NK 1 and NK 3 receptors as targets for drugs to treat gastrointestinal motility disorders and pain. British Journal of Pharmacology vol. 141 1303–1312 Preprint at https://doi.org/10.1038/sj.bjp.0705742 (2004).

52. Kuro-o, M. The Klotho proteins in health and disease. Nature Reviews Nephrology Preprint at https://doi.org/10.1038/s41581-018-0078-3 (2019).

53. Li, W. et al. DEPP/DEPP1/C10ORF10 regulates hepatic glucose andm fat metabolism partly via ROS-induced FGF21. FASEB Journal 32, 5459–5469 (2018).

54. Boztepe, T. & Gulec, S. Investigation of the influence of high glucose on molecular and genetic responses: An in vitro study using a human intestine model. Genes and Nutrition 13, (2018).

55. Stahl, R. et al. Trnp1 regulates expansion and folding of the mammalian cerebral cortex by control of radial glial fate. Cell (2013) doi:10.1016/j.cell.2013.03.027.

56. Gardiner, J. V. et al. Cerebellin1 is a novel orexigenic peptide. Diabetes, Obesity and Metabolism (2010) doi:10.1111/j.1463-1326.2010.01247.x.

57. Ewert, S. et al. Angiotensin II induced contraction of rat and human small intestinal wall musculature in vitro. Acta Physiologica (2006) doi:10.1111/j.1748-1716.2006.01600.x.

58. Cox, H. M. Neuropeptide Y receptors; antisecretory control of intestinal epithelial function. Autonomic Neuroscience: Basic and Clinical vol. 133 76–85 Preprint at https://doi.org/10.1016/j.autneu.2006.10.005 (2007).

59. Beucher, A. et al. The homeodomain-containing transcription factors Arx and Pax4 control enteroendocrine subtype specification in mice. PLoS ONE (2012) doi:10.1371/journal.pone.0036449.

60. Gross, S. et al. The novel enterochromaffin marker Lmx1a regulates serotonin biosynthesis in enteroendocrine cell lineages downstream of Nkx2.2. Development 143, 2616–2628 (2016).

61. Borges, M. et al. An achaete-scute homologue essential for neuroendocrine differentiation in the lung. Nature (1997) doi:10.1038/386852a0.

62. Pan, F. C., Brissova, M., Powers, A. C., Pfaff, S. & Wright, C. V. E. Inactivating the permanent neonatal diabetes gene Mnx1 switches insulin-producing β-cells to a δ-like fate and reveals a facultative proliferative capacity in aged β-cells. Development (Cambridge) (2015) doi:10.1242/dev.126011.

63. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. Nat Protoc. 8, (2013).

64. Scott, L. J. et al. A genome-wide association study of type 2 diabetes in finns detects multiple susceptibility variants. Science (1979) 316, 1341–1345 (2007).

65. Zhang, J., McKenna, L. B., Bogue, C. W. & Kaestner, K. H. The diabetes gene Hhex maintains ??-cell differentiation and islet function. Genes and Development 28, 829–834 (2014).

66. Schneeberger, M. Irx3, a new leader on obesity genetics. EBioMedicine vol. 39 19–20 Preprint at https://doi.org/10.1016/j.ebiom.2018.12.005 (2019).

67. Caruso, V. et al. mRNA GPR162 changes are associated with decreased food intake in rat, and its human genetic variants with impairments in glucose homeostasis in two Swedish cohorts. Gene 581, 139–145 (2016).

68. Foster, S. R. et al. Discovery of Human Signaling Systems: Pairing Peptides to G Protein-Coupled Receptors. Cell 179, 895-908.e21 (2019).

69. Shcherbina, L. et al. Intestinal CART is a regulator of GIP and GLP-1 secretion and expression. Molecular and Cellular Endocrinology 476, 8–16 (2018).

70. Hyland, N. P. & Cryan, J. F. A gut feeling about GABA: Focus on GABAB receptors. Frontiers in Pharmacology OCT, (2010).

71. Panaro, B. L. et al. The melanocortin-4 receptor is expressed in enteroendocrine I cells and regulates the release of peptide YY and glucagon-like peptide 1 in vivo. Cell Metabolism 20, 1018–1029 (2014).

72. Sullo, A., Brizzi, G. & Maffulli, N. Chronic peripheral administration of serotonin inhibits thyroid function in the rat. Muscles, Ligaments and Tendons Journal 1, 48–50 (2011).

73. Parikh, K. et al. Colonic epithelial cell diversity in health and inflammatory bowel disease. Nature (2019) doi:10.1038/s41586-019-0992-y.

74. Vainio, L. et al. Neuronostatin, a novel peptide encoded by somatostatin gene, regulates cardiac contractile function and cardiomyocyte survival. Journal of Biological Chemistry 287, 4572–4580 (2012).

75. Loh, Y. P., Cheng, Y., Mahata, S. K., Corti, A. & Tota, B. Chromogranin A and derived peptides in health and disease. in Journal of Molecular Neuroscience vol. 48 347–356 (2012).

76. Thomsen, S. K. et al. Type 2 diabetes

risk alleles in PAM impact insulin release from human pancreatic β-cells. Nature Genetics 50, 1122–1131 (2018).

77.	Ramesh, N., Mortazavi, S. & Unniappan, S. Nesfatin-1 stimulates glucagon-like peptide-1 and glucose-dependent insulinotropic polypeptide secretion from STC-1 cells in vitro. Biochemical and Biophysical Research Communications 462, 124–130 (2015).

78.	Piccand, J. et al. Rfx6 promotes the differentiation of peptide-secreting enteroendocrine cells while repressing genetic programs controlling serotonin production. Molecular Metabolism 29, 24–39 (2019).

79.	Beucher, A. et al. The homeodomain-containing transcription factors Arx and Pax4 control enteroendocrine subtype specification in mice. PLoS ONE (2012) doi:10.1371/journal.pone.0036449.

80.	Roberts, G. P. et al. Comparison of human and murine enteroendocrine cells by transcriptomic and peptidomic profiling. in Diabetes (2019). doi:10.2337/db18-0883.

81.	He, J., Irwin, D. M., Chen, R. & Zhang, Y. P. Stepwise loss of motilin and its specific receptor genes in rodents. Journal of Molecular Endocrinology 44, 37–44 (2010).

82.	He, J., Irwin, D. M., Chen, R. & Zhang, Y. P. Stepwise loss of motilin and its specific receptor genes in rodents. Journal of Molecular Endocrinology 44, 37–44 (2010).

83.	Chisholm, C. & Greenberg, G. R. Somatostatin-28 regulates GLP-1 secretion via somatostatin receptor subtype 5 in rat intestinal cultures. American Journal of Physiology - Endocrinology and Metabolism 283, (2002).

84.	Berna, M. J. et al. Serum gastrin in Zollinger-Ellison syndrome: II. Prospective study of gastrin provocative testing in 293 patients from the national institutes of health and comparison with 537 cases from the literature. Evaluation of diagnostic criteria, proposal of new . Medicine 85, 331–364 (2006).

85.	Coebergh, J. W. W., Van Veen, E. Ben, Vandenbroucke, J. P., Van Diest, P. & Oosterhuis, W. One-time general consent for research on biological samples: Opt out system for patients is optimal and endorsed in many countries [1]. British Medical Journal vol. 332 665 Preprint at https://doi.org/10.1136/bmj.332.7542.665 (2006).

86.	Sato, T. et al. Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature 459, 262–265 (2009).

87.	Sachs, N. et al. Long-term expanding human airway organoids for disease modeling.

The EMBO Journal 38, (2019).

88.	Koo, B. K. et al. Controlled gene expression in primary Lgr5 organoid cultures. Nature Methods 9, 81–83 (2012).

89.	Schmid-Burgk, J. L., Höning, K., Ebert, T. S. & Hornung, V. CRISPaint allows modular base-specific gene tagging using a ligase-4-dependent mechanism. Nature Communications (2016) doi:10.1038/ncomms12338.

90.	Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. Nat Protoc. 8, (2013).

91.	Fujii, M., Matano, M., Nanki, K. & Sato, T. Efficient genetic engineering of human intestinal organoids using electroporation. Nature Protocols 10, 1474–1485 (2015).

92.	Lo, C. A. et al. Quantification of Protein Levels in Single Living Cells. Cell Reports 13, 2634–2644 (2015).

93.	Wang, F. et al. RNAscope: A novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. Journal of Molecular Diagnostics 14, 22–29 (2012).

94.	Muñoz, J. et al. The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent "+4" cell markers. EMBO J 31, 3079–3091 (2012).

95.	Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: Large-scale single-cell gene expression data analysis. Genome Biology 19, (2018).

96.	Hashimshony, T. et al. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome Biology 17, 77 (2016).

97.	Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15, 550 (2014).

98.	The, M., MacCoss, M. J., Noble, W. S. & Käll, L. Fast and Accurate Protein False Discovery Rates on Large-Scale Proteomics Data Sets with Percolator 3.0. Journal of the American Society for Mass Spectrometry 27, 1719–1727 (2016).

99.	Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nature Methods vol. 13 731–740 Preprint at https://doi.org/10.1038/nmeth.3901 (2016).

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

CHAPTER 5 - SI

Supporting information for

High Resolution mRNA and Secretome Atlas of Human Enteroendocrine Cells

Joep Beumer[1,2,14], Jens Puschhof[1,2,14], Julia Bauzá-Martinez[3,4,14], Adriana Martínez-Silgado[1,2], Rasa Elmentaite[5], Kylie R. James[5], Alexander Ross[6,7], Delilah Hendriks[1,2], Benedetta Artegiani[1,2], Georg Busslinger[1,2], Bas Ponsioen[8], Amanda Andersson-Rolf[1,2], Kai Kretzschmar[1,2], Maarten H. Geurts[1,2], Yotam E. Bar-Ephraim[1,2], Cayetano Pleguezuelos Manzano[1,2], Yorick Post[1,2], Franka van der Linden[9], Carmen Lopez Iglesias[10], Willine J. van de Wetering[1,10], Reinier van der Linden[1,2], Peter J. Peters[10], Albert .J.R. Heck[3,4], Joachim Goedhart[9], Hugo Snippert[8], Matthias Zilbauer[7], Sarah A. Teichmann[5,11,12], Wei Wu[3,4,#], Hans Clevers[1,2,13,15,*]

[1] Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences (KNAW) and UMC Utrecht, 3584 CT Utrecht, The Netherlands.
[2] Oncode Institute, Hubrecht Institute, 3584 CT Utrecht, The Netherlands.
[3] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.
[4] Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands.
[5] Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, United Kingdom, CB10 1SA.
[6] Department of Surgery, University of Cambridge, Cambridge CB2 0QQ, UK
[7] Department of Paediatrics, University of Cambridge, Cambridge CB2 0QQ, UK
[8] Oncode Institute, Center for Molecular Medicine, University Medical Centre Utrecht, Utrecht, The Netherlands.
[9] Swammerdam Institute for Life Sciences, Section of Molecular Cytology, van Leeuwenhoek Centre for Advanced Microscopy, University of Amsterdam, Amsterdam, the Netherlands.
[10] The Maastricht Multimodal Molecular Imaging institute, Maastricht University, 6229 ER Maastricht, The Netherlands.
[11] Theory of Condensed Matter, Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge, United Kingdom, CB3 0HE
[12] European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, United Kingdom, CB10 1SA
[13] The Princess Maxima Center for Pediatric Oncology, 3584 CS Utrecht, The Netherlands.
[14] Co-first author and [15] Lead Contact

Supplementary figure 1. Enteroendocrine cells in human intestinal organoids display normal coexpression profiles. (A) qPCR analysis showing expression of hormones over the course of EEC differentiation. Expressions levels are normalized to GADPH and relative to day 0. The experiment was performed in n = 2 independent experiments, and the individual datapoints are depicted. (B) qPCR analysis showing expression of hormones after different durations of doxycycline (Dox) challenge. Organoids were differentiated for 5 days, and treated without dox, for 6 hours, 24 hours, 48 hours or 120 hours (the full differentiation time) dox. Expressions levels are normalized to GADPH and relative to 48h dox treatment. The experiment was performed in n = 2 independent experiments, and the individual datapoints are depicted. (C) Immunofluorescent staining of EEC-enriched organoids. Multiple hormones are expressed mutually exclusive. Scale bar is 50 μm.
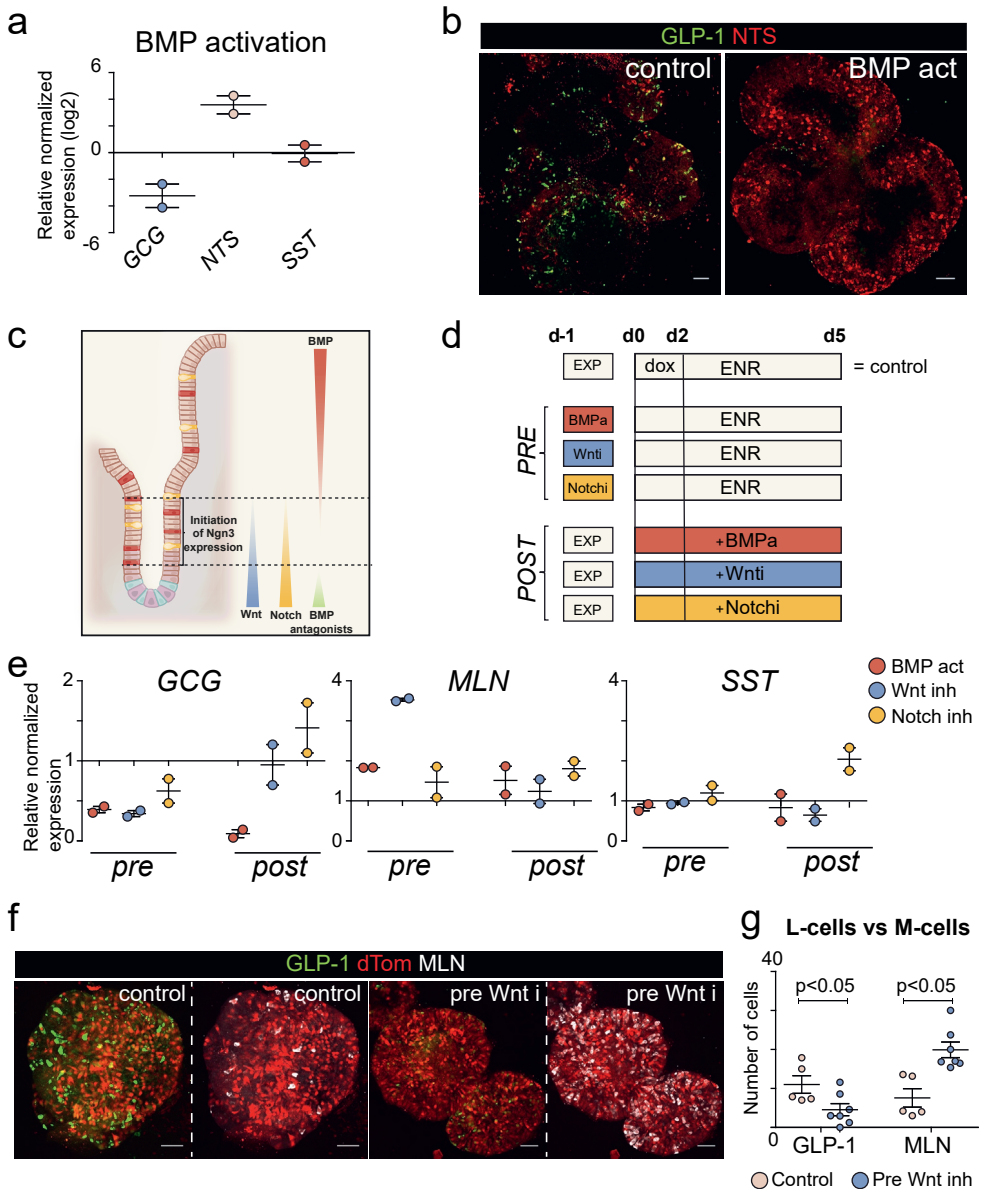
a **BMP activation**

b GLP-1 NTS / control / BMP act

c

d

e *GCG* *MLN* *SST*

f GLP-1 dTom MLN / control / control / pre Wnt i / pre Wnt i

g **L-cells vs M-cells**

Supplementary figure 2. Manipulation of Wnt and BMP signaling allows controling subspecification of EECs. (A) qPCR analysis showing expression of hormones after BMP treatment. Expression levels are normalized to GADPH, and relative to a non-treated control. The experiment was performed in n = 2 independent experiments, and the mean expression and SEM are depicted. (B) Immunofluorescent staining of BMP treated organoids. Scale bar is 50 µm. (C) Schematic representation of morphogen gradients in the intestinal crypt related to sites of initiation of Neurogenin-3 (Ngn3) expression. (D) Experimental paradigm. Different signalling pathways were modulated (BMP activation; BMPa, Wnt inhibition; Wnti, Notch in-hibition; Notchi) either 24 hours before (pre) or at the start (post) of NEUROG3 expression mediated by doxycycline (dox) treatment. (Continues on next page)

Supplementary figure 2. Manipulation of Wnt and BMP signaling allows controling subspecification of EECs. Control organoids were kept in standard expansion conditions (EXP) before dox treatment, and in standard differentiation conditions (ENR) after initiation of dox treatment. (E) qPCR analysis showing expression of hormones after different treatments shown in. Expression levels are normalized to GADPH, and relative to a non-treated control. The experiment was performed in n = 2 independent experiments, and the mean expression and SEM are depicted. (F) Immunofluorescent staining of organoids differentiated towards EECs after a 24 hour inhibition of Wnt (pre Wnt i). Organoids are shown as a maximum projection. Scale bar is 50 μm. (G) Quantification of (F). Number of positive cells were counted on n=5 organoid sections. Pre-inhibition of Wnt signalling caused a shift of L-cell to M-cell differentiation.

5

5

a   Live and single cell gating

Reporter-specific gating

b

c

d
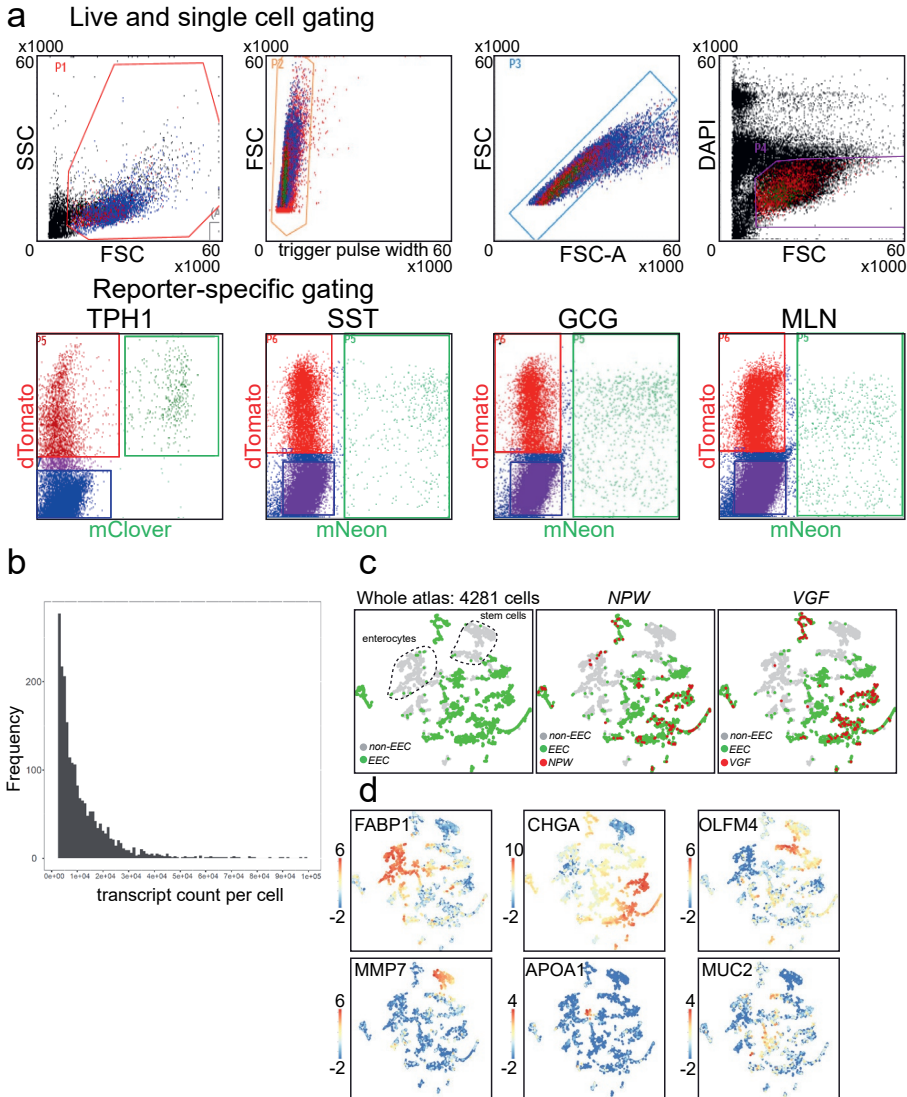
Supplementary figure 3. Generation of a human single cell RNA sequencing atlas of the intestinal tract using organoids. (A) Fluorescence-activated cell sorting (FACS) gating parameters for sorting of different EEC subtypes from reporter organoids. (B) Histogram displaying the total number of unique transcripts per cell (median number per cell is 7288 transcripts). (C) A broad human intestinal organoid atlas (n=4281 cells) generated by single cell RNA sequencing and displayed using a t-distributed stochastic neighbor-embedding (t-SNE) map. Cells defined as EECs (see methods) are shown in green. Cells expressing NPW or VGF (>1 transcripts, respectively) are highlighted in red and are found exclusively in EEC clusters. (D) t-SNE maps displaying lineage markers in the whole human intestinal organoid cell atlas (n= 8448 cells). Bars display color-coded unique transcript expression (logarithmic scale).

5

A

Prox SI
Dist SI
Colon

GCG+
MLN+
SST+
TPH1+

B

Ratio of EEC subtypes

	Prox SI	Dist SI
L (GCG)	0	13.9
L (PYY)	20.3	44.1
L (NTS)	0.8	5.6
D	18.6	13
G/K	9.6	1.3
M/X	24.1	8.1
EC (Reg4)	9.3	13.9
EC (Scg2)	17.3	0.2

C **Mouse tissue EECs**

Chga Tph1 Gcg Nts
Pyy Sst Ghrl Gip

D **Human tissue EECs**

CHGA TPH1 GCG NTS MLN
PYY SST GHRL GIP IAPP

E **Human organoid EECs**

DDC SLC18A1 OR51E1 CHGB GPR112

Mouse tissue EECs

Ddc Slc18a1 Olfr558 Chgb

F **Mouse tissue EECs**

Mdk Cpb1
Fgf14 Olfr78
Tdo2 Tac2
Fgfr1 Klb
Depp1 Lcn15
Trnp1 Cbln1
Slc26a7 Ttr
Il20-ra Agt

G +FGF21

B-ACTIN, GCG, CHGA, NTS, PYY, TPH1, MLN

Relative normalized expression

Supplementary figure 4. Single cell RNA sequencing of human EECs from organoids and tissue and mouse EECs from tissue. (A) t-SNE maps displaying the origin (left, tissue; right, reporter organoid) of cells from the human EEC atlas (n= 2255 cells). (B) The percentages of EECs corresponding to the different subtypes in proximal and distal SI organoids. (C) t-SNE maps displaying the expression levels of hormones in the different murine EEC subtypes from intestinal tissue. Bars display color-coded unique transcript expression (logarithmic scale). (D) t-SNE maps displaying the expression levels of hormones in the different human EEC subtypes from intestinal tissue. Bars display color-coded unique transcript expression (logarithmic scale). (E) t-SNE maps displaying conserved expression of different EC markers in human and mouse EECs. Bars display color-coded unique transcript expression (logarithmic scale).. (Continues on next page)

Supplementary figure 4. Single cell RNA sequencing of human EECs from organoids and tissue and mouse EECs from tissue. (F) t-SNE maps displaying the levels of hormone and marker gene expression of human M/X cells in the different murine EEC subtypes from intestinal tissue. Bars display color-coded unique transcript expression (logarithmic scale). (G) qPCR analysis showing expression of hormones after FGF-21 treatment during the 5 day EEC differentiation. Expressions levels are normalized to GADPH and relative to control organoids that are EEC differentiated without FGF-21 treatment. B-ACTIN is displayed as second housekeeping gene. The experiment was performed in n = 2 independent experiments, and the mean expression and SEM are depicted.

5

5
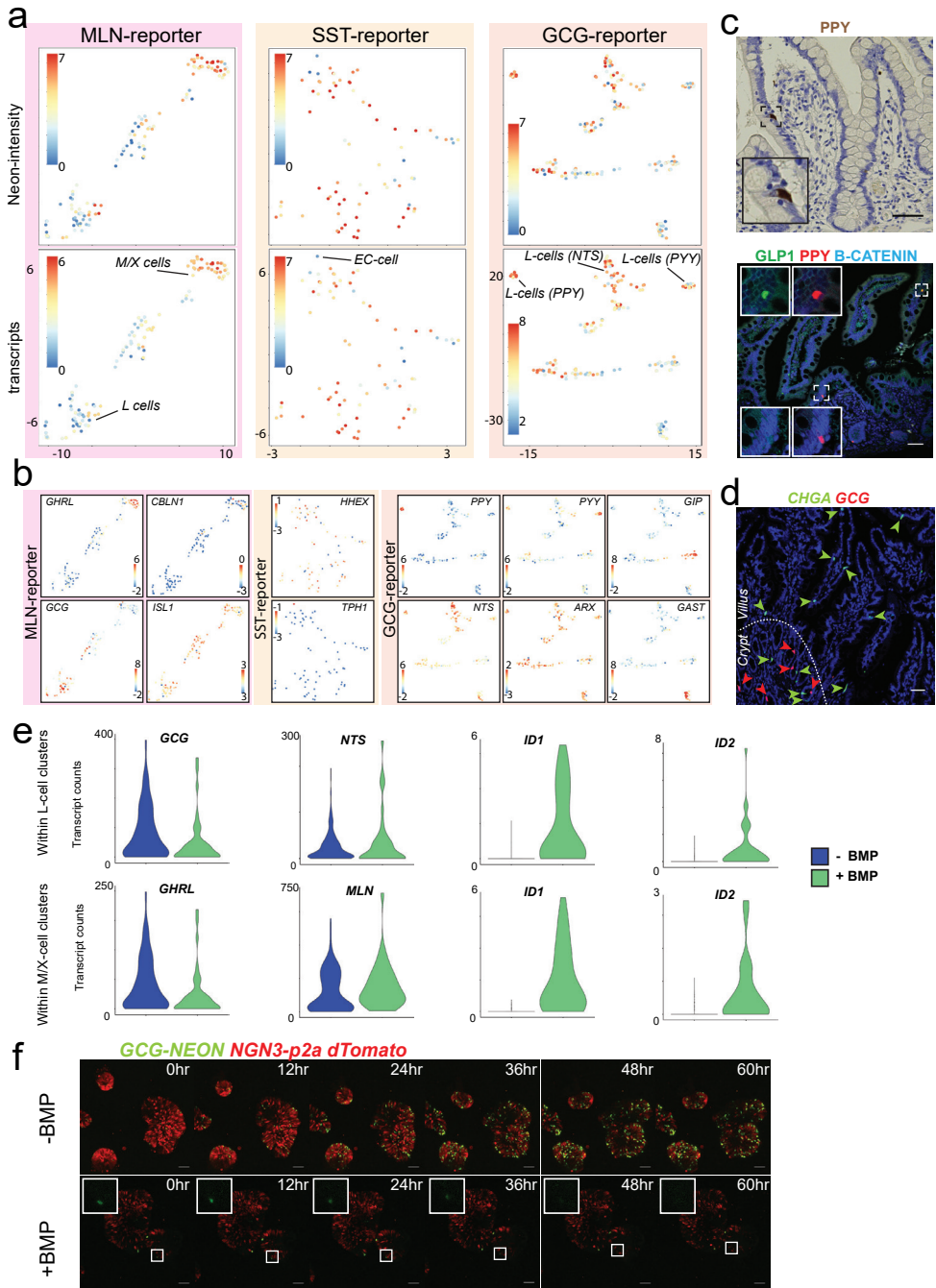
Supplementary figure 5. Subclustering of human EEC subtypes and responses to BMP activation. (A) Subclustering was performed on EECs sorted from different reporter organoids. t-SNE maps displaying the correlation between transcript levels and reporter intensity. (B) t-SNE maps displaying different hormones in EECs from the different reporter organoids. PPY-expressing cell form a distinct cluster of GCG+ cells. (Continues on next page)

Supplementary figure 5. Subclustering of human EEC subtypes and responses to BMP activation. (C) Immunohistochemistry of human ileal sections confirms PPY expression in vivo. Scale bar is 50 µm. (D) Fluorescent in situ hybridization on human ileal section shows crypt-restricted expression of GCG (red arrowheads), whereas CHGA expression (green arrowheads) is expressed also in the villus. Scale bar is 50 µm. (E) Violin plots depict the expression levels of selected hormones in single BMP-treated cells versus untreated cells in the EEC single cell RNA sequencing atlas. ID1 and ID2 are BMP target genes that confirm specific pathway activation of BMP agonist-treated cells. (F) Snapshots are shown of GCG-neon reporter organoids that were treated with BMP after 2 days of dox treatment to induce NEUROG3-dTomato expression (= 0hr timepoint). BMP treatment blocks the appearance of GCG$^+$ cells, while pre-existing L-cells downregulate GCG expression. No cell death is observed. Scale bar is 50 µm.
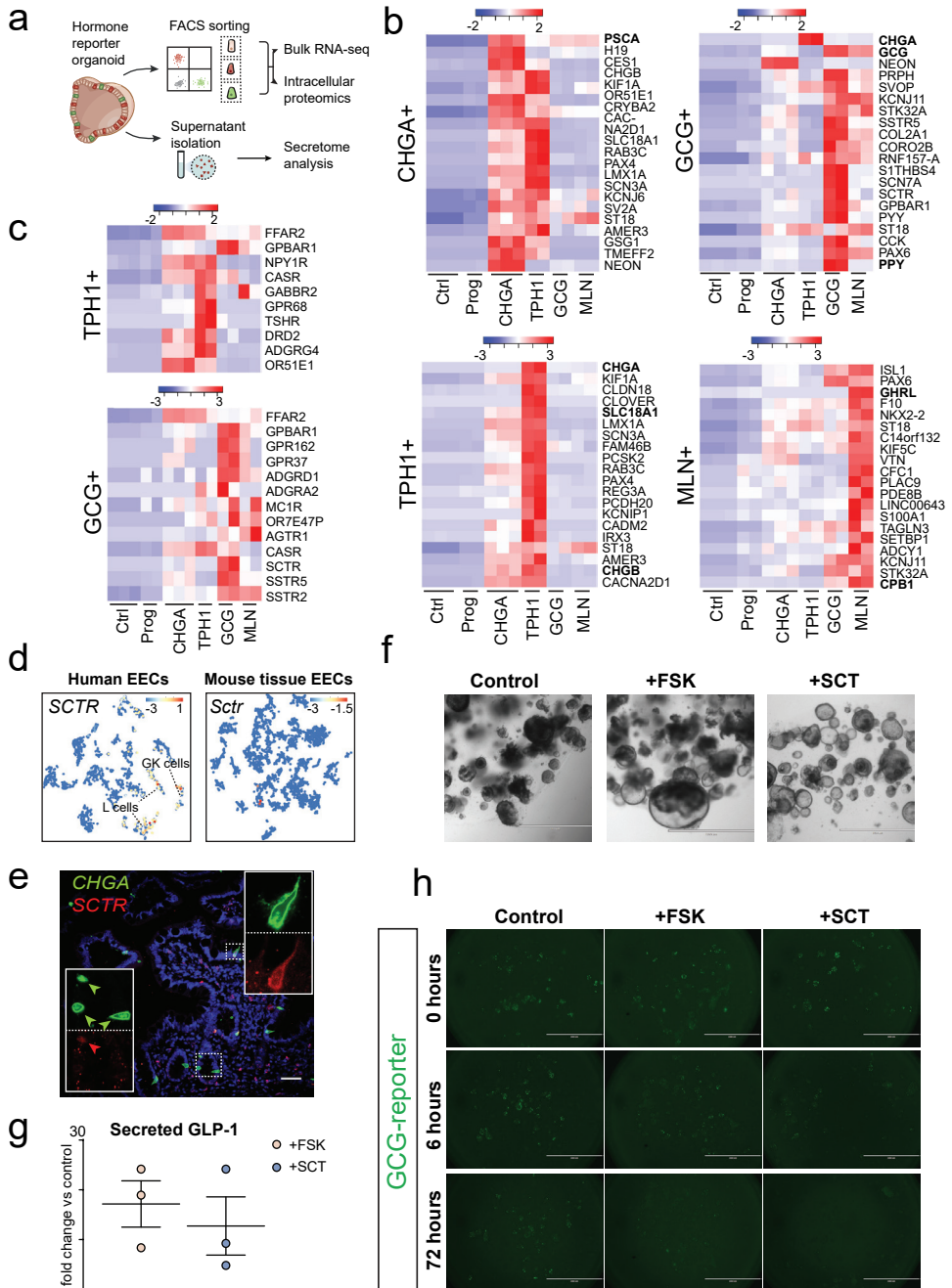
5

Supplementary figure 6. Bulk transcriptomic profiling of sorted enteroendocrine cell subtypes. (A) Experimental paradigm. Hormone reporter organoids are differentiated, after which subpopulations of EECs are sorted using FACS and processed for bulk RNA-sequencing or intracellular proteomics. In a separate experiment, the supernatant of organoids is collected after 24 hour forskolin stimulation and processed for proteomic analyses to determine the EEC secretome. (Continues on next page)

Supplementary figure 6. Bulk transcriptomic profiling of sorted enteroendocrine cell subtypes. (B) Heatmaps showing the 20 most significant RNA markers enriched in purified reporter populations. In bold genes are highlighted that are also among the 20 most significant markers on protein level. Colored bars represent Z-scores. (C) Heatmaps showing receptor expression most unique to TPH1$^+$ or GCG$^+$ cells. The receptor for EEC hormone Secretin (SCTR) is expressed highly in L-cells but not ECs, while ECs display unique expression of the PYY receptor (NPY1R). (D) t-SNE map displaying the expression level of the Secretin receptor (SCTR) in the EEC single cell sequencing atlas. Bars display color-coded unique transcript expression (logarithmic scale). (E) Fluorescent in situ hybridization on human ileal sections shows rare SCTR-expressing cells (red arrowhead) that sometimes co-express CHGA (green arrowhead). Scale bar is 50 µm. (F) Representative bright-field images of EEC-differentiated organoids after 24 hours forskolin (FSK) or Secretin (SCT) treatment. Both FSK and SCT treatment causes swelling of organoids, indicative of cAMP activation. Scale bar is 1 mm. (G) ELISA showing the fold increase in GLP-1 concentrations of EEC-enriched organoids after treatment with FSK or SCT. The experiment was performed in n = 3 independent experiments, and the mean fold change and SEM are depicted. (H) GCG-reporter organoids were differentiated towards EECs and treated with FSK or SCT. Intracellular levels of GCG-neon reduce over the course of FSK and SCT treatment. Scale bar is 2 mm.
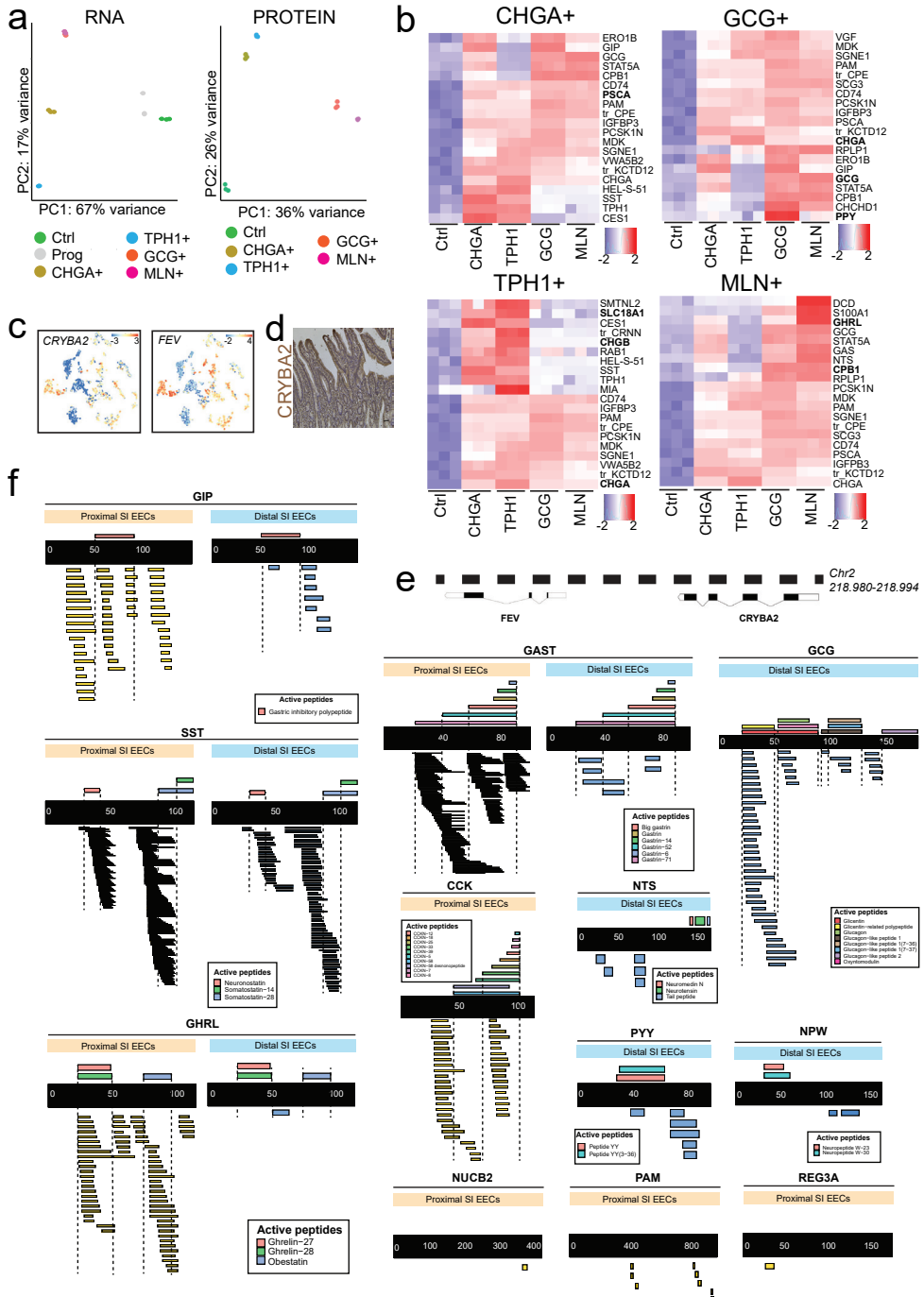
Supplementary figure 7. Proteomic and secretomic profiling of EECs. (Caption on next page)

Supplementary figure 7. Proteomic and secretomic profiling of EECs. (A) Principle component analysis (PCA) of RNA and protein data from different EEC populations. CHGA⁻ cells were used as control. CHGA⁻ cells that were positive for dTomato (induced NEUROG3-expression) were defined as EEC progenitors (Prog.). (B) Heatmaps showing the 20 most significant markers on protein level defining each EEC populations. Colored bars represent Z-scores. In bold genes are highlighted that are also among the 20 most significant markers on RNA level. (C) t-SNE maps displaying the expressions level of CRYBA2 and FEV in the EEC single cell sequencing atlas, illustrating a high degree of overlap in expression. Bars display color-coded unique transcript expression (logarithmic scale). (D) Immunohistochemistry on human duodenal sections shows a lack of CRYBA2 expression in vivo. Scale bar is 50 µm. (E) The location of the CRYBA2 and FEV genes on Chromosome 2. (F) Measured peptides (<10 kDa) in the secretome mapping to different secreted prehormones are shown below the black bar. Data from proximal (yellow background) and distal SI organoid (blue background) supernatants are displayed.

5

CHAPTER 6

Mapping prohormone processing by proteases in human enteroendocrine cells using genetically engineered organoid models

Joep Beumer[1,8], Julia Bauzá-Martinez[2,8], Tim S. Veth[2], Veerle Geurts[1], Charelle Boot[1], Filip Knop[5,6,7], Wei Wu[2,3,4,*], Hans Clevers[1,*]

1 Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences (KNAW) and UMC Utrecht, 3584 CT Utrecht, The Netherlands; Oncode Institute.
2 Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.
3 Singapore Immunology Network (SIgN), Agency for Science, Technology and Research (A*STAR), Singapore 138648, Singapore.
4 Department of Pharmacy, National University of Singapore, Singapore 117543, Singapore.
5 Center for Clinical Metabolic Research, Gentofte Hospital, University of Copenhagen, Hellerup, Denmark
6 Steno Diabetes Center Copenhagen, Herlev, Denmark
7 Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
8 These authors contributed equally.

Based on:

Summary

Enteroendocrine cells (EECs) secrete hormones in response to ingested nutrients to control physiological processes such as appetite and insulin release. EEC hormones are synthesized as large proproteins that undergo proteolytic processing to generate bioactive peptides. Mutations in EEC-enriched proteases are associated with endocrinopathies. Due to the relative rarity of EECs and paucity of in vitro models, intestinal prohormone processing remains challenging to assess. Here, human gut organoids in which EECs can efficiently be induced are subjected to CRISPR-Cas9-mediated modification of EEC-expressed endopeptidase and exopeptidase genes. We employ mass spectrometry-based analyses to monitor peptide processing and identify glucagon production in intestinal EECs, stimulated upon bone morphogenic protein (BMP) signaling. We map the substrates and products of major EECs endo- and exopeptidases. Our studies provide a comprehensive description of peptide hormones produced by human EECs and define the roles of specific proteases in their generation.

Significance Statement

Enteroendocrine cells control key physiological processes such as appetite and insulin secretion through secretion of neurotransmitters and peptide hormones. The bioactive peptides are subject to complex proteolytic processing essential for their activation or inactivation. Moreover, alternative processing allows for the generation of different peptides from the same precursor protein. We use human organoid cultures combined with CRISPR-Cas9-mediated loss-of-function and peptidomics to assay the peptide spectrum of gut proteases. We identify new substrates of these enzymes, including the production of intestinal glucagon. A more complete understanding of hormone processing could allow a rational design of therapeutic interventions targeting these proteases.

Introduction

The mammalian intestine constitute the largest hormone-producing organ through the activity of enteroendocrine cells (EECs)[1]. EECs represent only 1% of the intestinal epithelium. The >20 known bioactive products released from these cells regulate key physiological processes including appetite, insulin secretion and bowel movement. EECs secrete their products in response to luminal or absorbed nutrients[2]. The EEC products are believed to signal in a paracrine manner, locally to the enteric nervous system and musculature, and to distant endocrine organs and the CNS.

EECs synthesize prohormone proteins from which bioactive peptides are derived through complex post-translational processing by endopeptidases and exopeptidases[3]. The majority of these proteases co-localize with hormones in the secretory pathway of the EECs. Multiple EEC-enriched proteases involved in hormone processing have been described. The calcium-regulated proprotein convertase PCSK1 (also known as PC1/3), a serine endoprotease cleaves substrates C-terminally to dibasic residues and is essential for the generation of the incretin glucagon-like peptide 1 (GLP-1) from preproglucagon (encoded by *GCG*) in intestinal L cells, a subtype of EECs[4]. Pancreas islet-specific PCSK2 (PC2) provides alpha cells with the unique capacity to produce glucagon from the same prohormone. Carboxypeptidases cleave C-terminal basic residues generated after the endoprotease activity[5]. EECs express the carboxypeptidase CPE. After secretion of bioactive peptides, extracellular proteases can further process hormones to regulate their activity. Most notably, transmembrane dipeptidyl peptidase 4 (DPP4) has N-terminal exopeptidase activity and is best known for its ability to inactivate GLP-1[6]. DPP4 inhibitors, successful type 2 diabetes drugs, prevent inactivation of secreted GLP-1.

Mutations in EEC-enriched proteases are an important cause of endocrinopathies. For example, alterations in PCSK1 are linked to severe metabolic disorders[7,8]. No accurate models nor treatments have been developed for these diseases. While these proteases have expression profiles beyond the gut, mapping of their intestinal substrate spectrum may have clinical implications for these inherited diseases. Vice versa, pharmacological activation or inhibition of such enzymes could hold therapeutic potential for a range of diseases including type 2 diabetes, similar to DPP4 inhibitors. Function of EEC-enriched endopeptidases or exopeptidases have mostly been assessed using mouse models. For example, CPE mutant mice develop broad metabolic symptoms including diabetes. Yet, EEC-specific processing defects were not studied in this model[9]. Since human EECs produce several unique hormones compared to mouse EECs (e.g. motilin (MLN) and neuropeptide W (NPW)), mouse models have additional shortcomings to fully map the substrate specificity in human EECs. Moreover, we have previously identified a human EEC-enriched carboxypeptidase, CPB1, that is not produced by mice[10]. An alternative approach has involved co-expression of wildtype or mutant forms of proteases with the prohormones in cell lines[11]. Although such models allow more experimental flexibility, results might

not reflect the endogenous processing occurring in the secretory pathway of human EECs.

We have previously generated a toolset for the study of human EECs in adult stem cell-derived organoid cultures[10]. The conditional overexpression of the pro-endocrine transcription factor neurogenin 3 allows for the generation of high-purity cultures of EECs, enabling their functional study. This system allows for an in-depth study of human EECs, which includes the analysis of their secreted products Here, we employ CRISPR-Cas9-mediated gene-editing, efficient EEC differentiation and mass spectrometry-based peptidomic analysis to determine the role of endogenous gut proteases in hormone processing.

6

Results

We first assessed protease expression in different human EEC subtypes in ileal-derived organoids using our previously generated single cell RNA-sequencing dataset[10]. Highest PCSK1 expression was observed in GLP-1-producing L cells as well as in enterochromaffin cells (ECs), which secrete the neurotransmitter 5-HT (serotonin). PCSK2 was mostly restricted to ECs (**Figure 1A, S1A-B**). The *PCSK1N* gene codes for multiple bioactive peptides including proSAAS, which negatively regulates PCSK1 activity, and is broadly expressed by all EEC populations[12]. The transmembrane and soluble protease DPP4 is not restricted to EECs but produced by different intestinal lineages, most strongly by enterocytes (**Figure 1A, S1A-B**). We observed expression of the carboxypeptidase CPE in all EEC populations. A second carboxypeptidase, CPB1, was most prominently expressed in L cells and motilin/ghrelin-producing MX cells. CPB1 is not expressed in murine EECs and its function is not known (**Figure 1A, S1A-B**). To confirm its EEC-specific expression profile, we generated a knock-in reporter by targeting CRISPR-Cas9 to the CPB1 stop codon followed by NHEJ introducing fluorescent mNeon[13]. CPB1[mNeon+] cells were not observed in organoids in expansion conditions when EECs are absent. When organoids were differentiated to form EECs, we observed expression of CPB1 in a subset of EECs (**Figure 1B-C**). We generated an identical reporter construct for the known EEC protease PCSK1 that displayed a similar pattern (**Figure S1C**).

Some of the EEC subtypes alter expression of hormones while traveling from crypt, where they are 'born', to villus tips[14]. L cells express the GCG gene in crypts but switch to a cell type which enriches for neurotensin (NTS) when migrating towards the tip of the villus. This phenotype switch is controlled by a bone morphogenic protein (BMP) signaling gradient, inactive at crypt bottoms and highly active at villus tips. Bulk RNA- sequencing was performed on EEC-differentiated organoids in the presence and absence of BMP (Dataset 1-2). As previously reported[14,15], we observed strong downregulation by BMP signals of crypt hormone GCG in L cells, while villus hormones NTS and SCT increased (**Figure S1D**). Additionally, the recently discovered EEC product NPW was strongly induced by BMP activation (**Figure S1D**)[10,16]. Accordingly, we found that EECs from native human gut tissue producing the high-
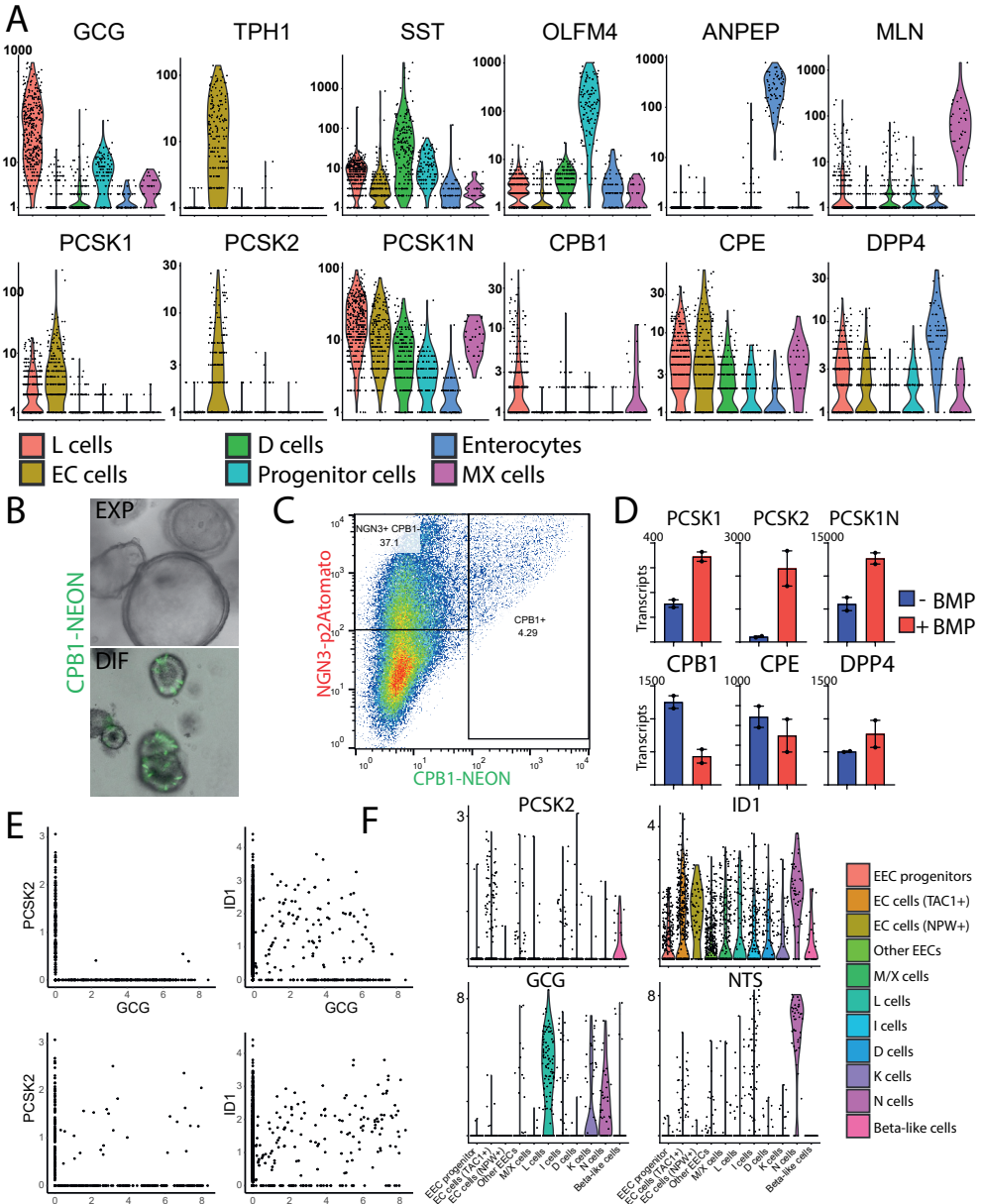
Figure 1. Expression profile of EEC-enriched proteases. (A) Violin plots depicting protease gene expression levels in different EEC populations derived from organoid cultures based on previously performed single-cell RNA-sequencing (9). (B) Overlay of brightfield and fluorescent image from CPB1mNeon organoids in culture. CPB1-neon positive cells are only observed upon EEC differentiation (DIF) and not in expanding conditions (EXP). (C) FACS analysis of differentiated NEUROG3tdTomato / CPB1mNeon organoids. Neon-positive cells can only be detected in NEUROG3-overexpressing cells, suggesting restricted expression to EECs. (D) Graphs depicting protease expression revealed by bulk RNA-sequencing of EEC-differentiated organoids in the presence and absence of BMPs. (Continues on next page)

Figure 1. Expression profile of EEC-enriched proteases. (E) Hormone and protease expression plotted from a previously published single cell RNA-sequencing atlas of the gut epithelium (17). (F) Violin plots depicting PCSK2, BMP target ID1 and the hormones GCG and NTS expression levels in different EEC populations from a previously published single cell RNA-sequencing atlas of the gut epithelium (17)

est level of NPW displayed expression of the BMP target gene ID1 (**Figure S1E**). NPW is expressed in ECs. Total numbers of ECs were not affected by BMP as based on *TPH1* expression, the rate-limiting enzyme for 5-HT production (**Figure S1D**). The majority of proteases were not significantly affected by BMP signaling, with the exception of BMP-induced *PCSK2* (**Figure 1D**). PCSK2 is known to catalyze the production of glucagon in the pancreas[17]. PCSK2 has been suggested to process GIP in murine intestinal K-cells[18]. We analyzed a recently published single cell sequencing atlas of the human gastrointestinal tract to identify *PCSK2* expression in the different EEC populations[19]. Strongest expression was observed in adult ECs, as expected, and in fetal EECs producing insulin (**Figure 1E-F**). In addition, we detected rare co-expression of *PCSK2* in L-cells producing *GCG*, and higher levels in *NTS*[+] EECs (**Figure 1E-F, S1F**). *NTS*[+] cells in part derive from *GCG*[+] cells that travel from crypt to villus while experiencing increasing BMP signals[14,20], and *NTS*[+] cells indeed expressed increased amounts of the BMP target gene *ID1* (**Figure 1E-F**). We sorted *NTS*[+] cells using a previously established reporter organoid, and found that BMP activation within these cells indeed stimulated *PCSK2* (**Figure S1G**), while decreasing *GCG*. These findings raised the possibility that crypt L cells may produce small amounts of PCSK2 and thus of glucagon (in addition to the prototypic L cell product GLP-1) in the crypt. Villus L cells downregulate *GCG* expression yet could have an increased capacity to process preexisting proglucagon to glucagon, because of increasing PCSK2 expression.

We employed CRISPR-Cas9-mediated gene editing to generate loss-of-function mutants of the various EEC proteases (**Table 1**). We included the endopeptidases PCSK1 and PCSK2, the carboxypeptidases CPE and CPB1, the aminopeptidase DPP4 and -finally- PCSK1N, regulator of PCSK1 activity and potentially of other proteases (**Figure 2A**). In addition, we used cytosine base editors to introduce previously reported point mutations in proteases as these occur in hereditary disorders. We focused on mutations reported to alter specificity rather than loss-of-function defects. Thus, we generated targeting constructs to produce the PCSK1^{S357G} mutation, suggested to be a hypermorph mutant with a 'PCSK2-like' substrate spectrum[21]. Secondly, we generated PCSK2^{R430W} mutations, previously suggested to cause increased conversion to type 2 diabetes in Amish subjects[22]. We transfected organoid cells with CRISPR-Cas9 plus a protease-specific gRNA to induce frameshifts, or with CRISPR-Cas9 base editor combinations (**Figure 2A**). Transfected cells were sorted based on GFP expression, clonally expanded and genotyped for the appropriate genotype (**Figure 2A**). Clonal knockout organoid lines were generated for all proteases and point mutations. For PCSK2^{R430W}, we only recovered a heterozygous line (**Figure 2B-C**). Nonetheless, we proceeded with this organoid line as the phenotype is predicted to be gain-of-function. We also generated PCSK1/PCSK2 double mutant organoids by sequential transfection to assess potential redundancy between
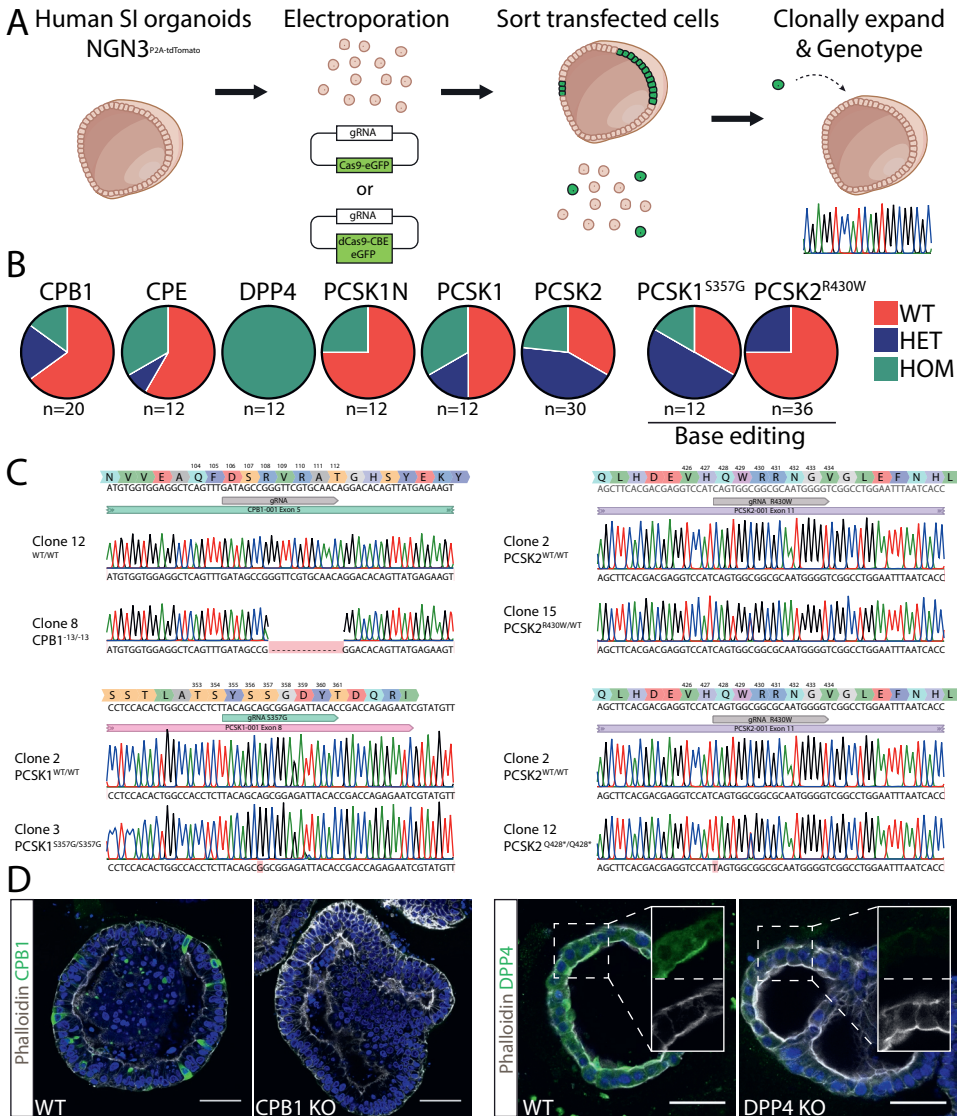
Figure 2. Genetically engineered organoids harboring protease mutations. (A) Cartoon displaying workflow to generate protease mutant organoids. (B) Targeting efficiency of the different genotypes. In-frame mutations were not counted as targeted due to the probable lack of phenotype. (C) Sanger traces of protease mutant clones. (D) Immunofluorescent staining of CPB1 and DPP4 mutant organoids. Stainings confirm genotype and absence of the targeted proteins. Scale bars are 50 μm.

these proteases. We confirmed complete loss of protein production for the CPB1 and DPP4 knockouts (**Figure 2D**).

We next employed mass spectrometry to document the proteolytic processing of gut hormones. We first focused our analyses on mutants of the endopeptidases. Wildtype and mutant organoid lines were differentiated in the presence and ab-

sence of BMPs, allowing the generation of the complete spectrum of EEC states and hormones as observed *in vivo* (**Figure S2A**). We did not note major morphological differences between wildtype and mutant organoids, with the exception of PCSK1/PCSK2 double mutant organoids: while all organoids appeared cystic under expansion conditions, PCSK1$^{-/-}$/ PCSK2$^{-/-}$ organoids displayed a dense phenotype with extensive budding (**Figure S2B**). Additionally, we always derived homozygous frameshift DPP4 mutants in this and a previous study, suggesting improved clonal outgrowth (**Figure 2B**). We did not note altered splitting ratios in these mutants. These observations suggest that (unknown) substrates of DPP4, PCSK1 and PCSK2 could play a role in clonal outgrowth, cell proliferation or morphogenetic events. Next, we collected the intracellular peptidome from differentiated organoids of all mutants. We additionally collected secreted peptides from all organoid models, induced to secrete their hormones for 90 minutes by chemical stimulation with adenylyl cyclase, inducing intracellular cAMP levels. We included both the parental and a wildtype 'sibling' organoid line for comparison. Due to the higher quality data extracted from intracellular peptidomes, which are naturally protected from unspecific proteolytic events happening extracellularly upon secretion, we focused our analyses on this fraction. The introduced mutations as well as BMP status caused major shifts in the observed number of processed bioactive peptides (**Figure 3A**).PCSK1$^{-/-}$/ PCSK2$^{-/-}$ peptidomes displayed a reduction in most bioactive peptides, stronger than single PCSK1 or PCSK2 mutants (**Figure 3A**). This indicated functional redundancy between these proteases. PCSK1 single mutants displayed a reduction in multiple peptides mapping to gastrin, promotilin - a hormone existing as a pseudogene in the mouse genome - and proglucagon. All of these were qualitatively increased in PCSK1N mutants, coding for a known negative regulator of PCSK1 activity (**Figure 3A**)[23]. We did not observe major changes in peptides extracted from PCSK1^{S357G} and PCSK2^{R430W} organoids, suggesting a minimal impact on enzymatic activity of these reported point mutations. Proglucagon is known to be heavily modified by the endopeptidases PCSK1 and PCSK2[24], which cleave prohormones after dibasic residues (lysine, K; and arginine, R). This is reflected in the strong enrichment of dibasic residues just before the N-termini of peptides mapping to proglucagon (**Figure 3C**, positions P1 and P2). Focusing on bioactive peptides derived from proglucagon, we noted a reduction of the different peptides in PCSK1 null mutants (**Figure 3B**). BMP activation negatively impacted total levels of proglucagon (**Figure 3A-B**), reflecting its effect on *GCG* transcription. Surprisingly, we detected bioactive glucagon specifically in BMP-treated conditions (**Figure 3B and Figure S3**), in line with the induction of *PCSK2* transcription by BMP (**Figure 1D**). We did not observe significant amounts of glucagon in the majority of BMP-inhibited cultures, nor in any of the BMP-activated PCSK2 mutant organoids (**Figure 3D**). Glucagon was only absent in one of the two PCSK1 mutant lines, potentially hinting to a role of this enzyme in its production. In addition to spectral evidence of glucagon obtained with our peptidomics approach, we further corroborated this evidence with MS1 and MS2 traces obtained using a parallel reaction monitoring (PRM) MS approach at optimal fragmentation and resolution (**Figure S4 and S5**). This approach yielded near complete fragmentation of the glucagon peptide (**Figure S4**) and demonstrated coeluting MS1 and MS2 traces found in the wildtype, EEC-differentiated organoids for both glucagon and active
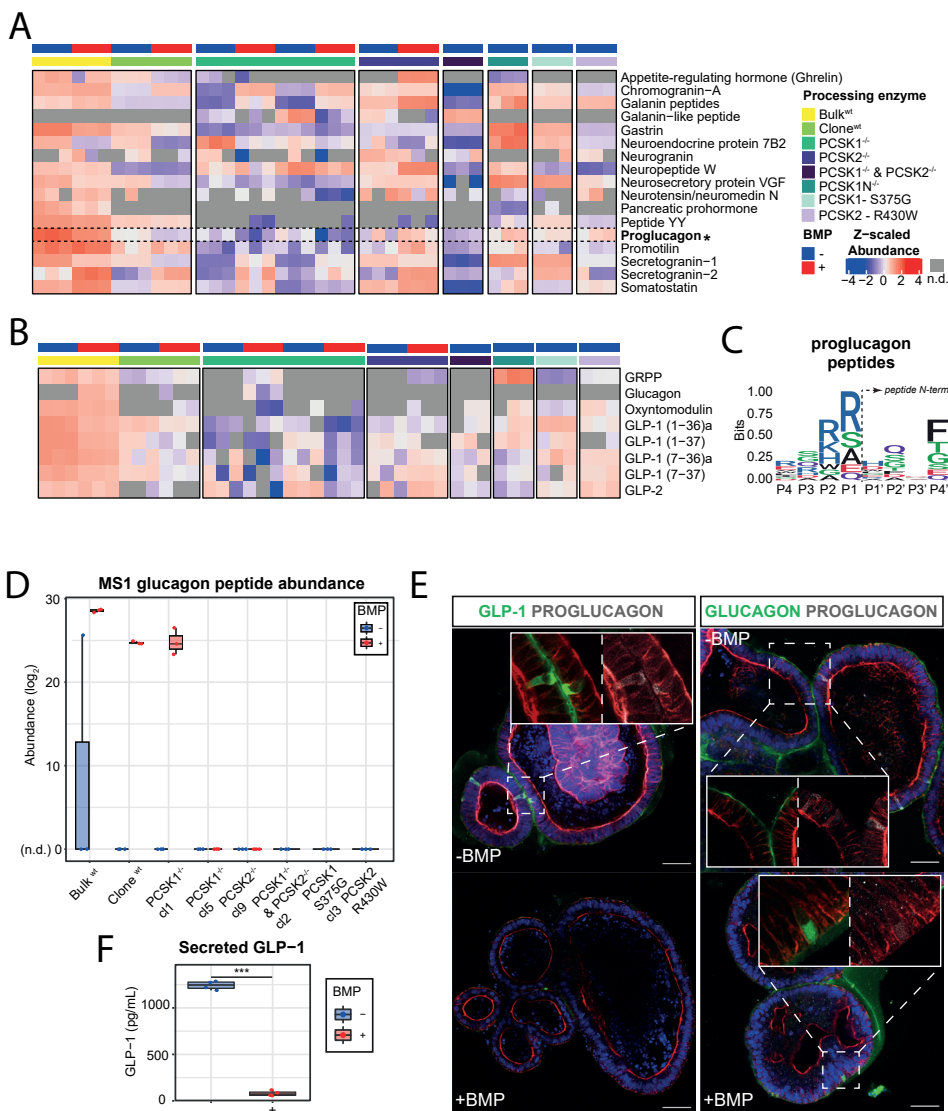
Figure 3. BMP-induced glucagon production in human EECs. (caption part 1) (A) LC-MS/MS abundance profiles of the diverse prohormones measured in the intracellular peptidome of wildtype and mutant organoids. Concentration of measured peptides was higher in the parental, non-clonal organoid ('Bulk') compared to clonal lines due to higher differentiation efficiency. Abundance is Z-scaled. Samples where the prohormone was not detected (n.d.) are displayed in gray. BMP-untreated (-) is shown in blue and BMP-treated (+) is shown in red. (B) LC-MS/MS abundance profiles of the bioactive peptides derived from proglucagon. For amidated peptides, an "a" is annotated after the peptide name. Abundance is Z-scaled. Samples where the prohormone was not detected (n.d.) are displayed in gray. BMP-untreated (-) is shown in blue and BMP-treated (+) is shown in red. (C) Sequence logo of residues found around the N-termini of all peptides derived from proglucagon. Dibasic residues are enriched at cleavage site (P1, P2) just before the N-termini (P1', P2', P3', P4') of peptides mapping to proglucagon, reflecting PCSK1 and PCSK2 cleavage specificities. (Continues on next page)

Figure 3. BMP-induced glucagon production in human EECs. Peptides measured in wild-type organoids were used, and any duplicate N-termini sequences (P-4 to P4) were excluded. In total, 11 unique N-termini (P-4 to P4) were used to generate the logo plot. (D) Total abundance for the glucagon peptide as extracted from the MS1 isotopic distributions evaluated on Skyline. Total abundance was calculated by summation of the individual intensities of all isotopes. Total abundance is displayed in log2 space. Each point in the boxplot represents individual injection replicates. BMP-untreated (-) is shown in blue and BMP-treated (+) is shown in red. When no reliable isotopic distribution was detected (n.d.), abundance was set to zero. Individual isotopic distributions for glucagon are shown in Figure S3 and Skyline result file is deposited in PRIDE Repository. (E) Immunofluorescent staining of proglucagon, GLP-1 and glucagon in BMP-untreated and -treated organoids. Scale bars are 50 μm. (F) GLP-1 ELISA on the supernatant of organoids differentiated in the presence and absence of BMPs. Organoids were stimulated to secrete for 6 hours using Forskolin and IBMX. Statistical significance was determined by two-sided Student's T-test (***, p-value < 0.001).

GLP-1 (7-37 and 7-36a) (**Figure S5A**). This confirmed the BMP-dependent increase of glucagon production and reduction in GLP-1 production (**Figure S5B**).

To validate mass spectrometry findings, we performed immunofluorescence using glucagon- and GLP-1-specific antibodies, as well as an antibody recognizing unprocessed proglucagon. We found that virtually all proglucagon positive cells in BMP-inactive conditions showed GLP-1 (but not glucagon) co-localization (**Figure 3E**). By contrast, BMP activation drastically reduced the number of proglucagon cells and intensity, but the remaining positive cells displayed glucagon reactivity. Analysis of secreted peptides by ELISA confirmed the BMP-mediated reduction of GLP-1 levels. Secreted glucagon remained below the detection threshold (**Figure 3F**). We conclude that human EECs are capable of producing bioactive glucagon, potentially mediated by BMP-stimulated expression of PCSK2.

Next, we analyzed the substrate spectrum of the carboxypeptidases CPE and CPB1. Loss of CPE resulted in the accumulation of 90 peptides mapping to multiple prohormones (**Figure 4A**). These peptides were strongly enriched in the basic residues lysine (K) and arginine (R) at the C-terminal position (**Figure 4B**). Notably, the set of peptides enriched in wildtype organoids compared to CPE mutants did not present K or R in the C-terminal position but in the contiguous residue (C-terminal +1) (**Figure 4B**). Among the peptides accumulating in the CPE$^{-/-}$ background, we found sequences derived from neuropeptide W (NPW, a hormone not observed in the murine gut), VGF, the PC2 chaperone neuroendocrine protein 72B, proglucagon, promotilin, pancreatic prohormone (PPY), somatostatin (SST) and chromogranin A (**Figure 4C**). In the secretomes of wildtype and CPE mutant organoids, we observed a slightly larger range of substrates (**Figure S6A-C**), indicating this carboxypeptidase to act both intra and extracellularly.

Deletion of the human EEC-specific CPB1 resulted in the accumulation of a smaller set of 31 hormone-derived peptides (**Figure 4D**), consistent with a narrower expression pattern as well as lower levels of this carboxypeptidase compared to CPE (Figure 1A). Accumulating peptides displayed C-terminal enrichment of the basic amino acids R and K (**Figure 4E**), although other residues also occurred in this

Figure 4. Intracellular substrate specificity of EEC exopeptidases. (A) Significantly changing hormone-derived peptides in response to CPE knockout (CPE-/-) when compared to wildtype organoids. Bulk and a clonal wildtype line and two CPE-/- clones were included. Three technical replicates were measured for all samples. (B) Sequence logo of the C-termini of peptides accumulating in response to CPE knockout (CPE-/--enriched) and peptides accumulating in wildtype (wt-enriched). Lysine (K) and arginine (R) are highlighted in red. (C) Pie chart indicating the substrate repertoire of peptides accumulating in CPE knockouts. (Continues on next page)
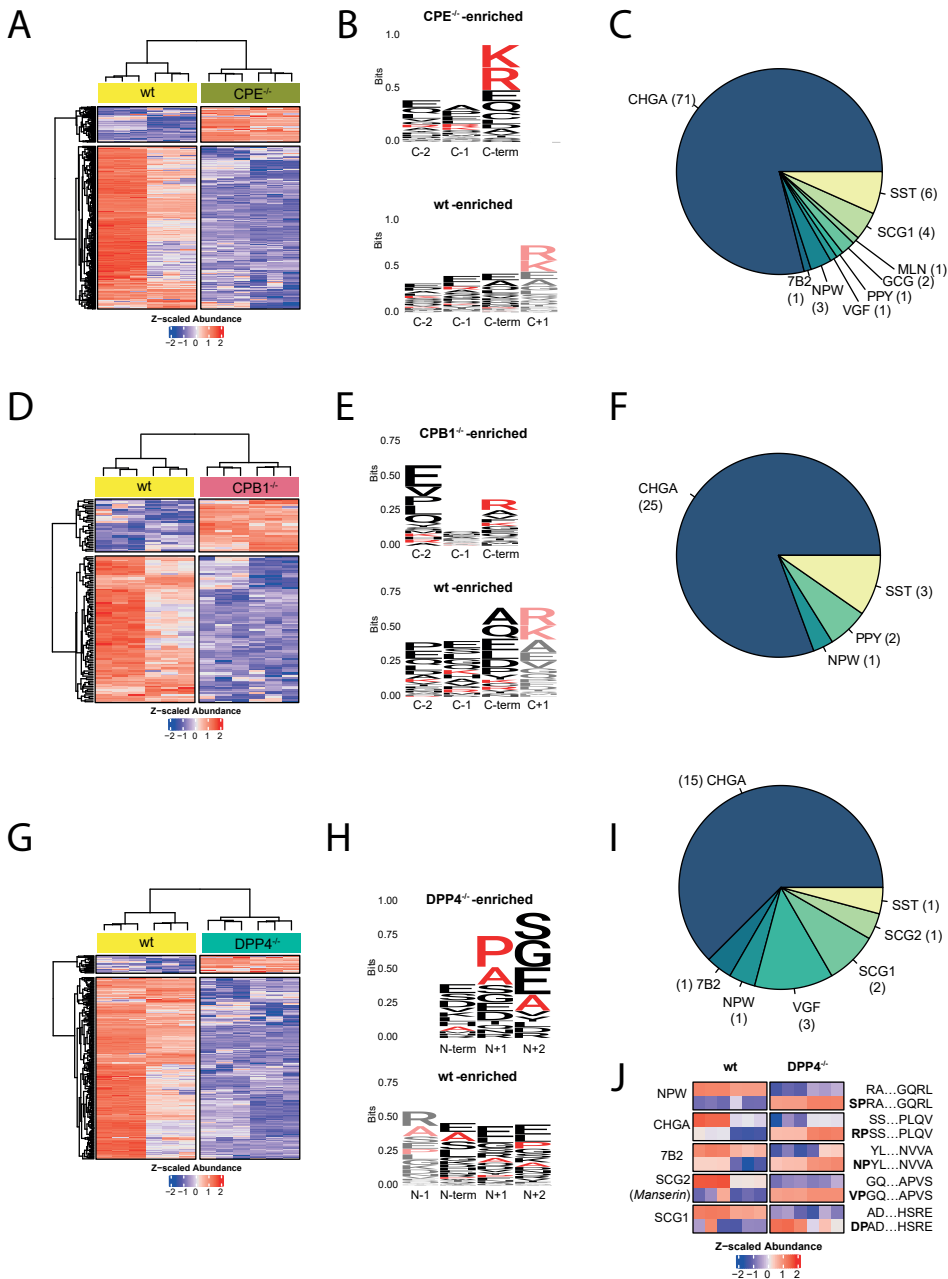
Figure 4. Intracellular substrate specificity of EEC exopeptidases. (D) Significantly changing hormone-derived peptides in response to CPB1 knockout (CPB1-/-) when compared to wildtype organoids. Bulk and a clonal wildtype line and two CPB1-/- clones were included. Three technical replicates were measured for all samples. (E) Sequence logo of the C-termini of peptides accumulating in response to CPB knockout (CPB-/--enriched) and peptides accumulating in wildtype (wt-enriched). Lysine (K) and arginine (R) are highlighted in red. (F) Pie chart indicating the substrate repertoire of peptides accumulating in CPB1 knockouts. (G) Significantly changing hormone-derived peptides in response to DPP4 knockout (DPP4-/-) when compared to wildtype organoids. Bulk and a clonal wildtype line and two DPP4-/- clones were included. Three technical replicates were measured for all samples. (H) Sequence logo of the N-termini of peptides accumulating in response to DPP4 knockout (DPP4-/--enriched) and peptides accumulating in wildtype (wt-enriched). Proline (P) and alanine (A) are highlighted in red. (I) Pie chart indicating the substrate repertoire of peptides accumulating in DPP4 knockouts. (J) Heatmap showing the Z-scaled abundance of peptide pairs differing by XP/A dipeptide cleavage when comparing DPP4-/- to wildtype clones.

6

position. Notably, we observed accumulation of SST, NPW and PPY peptides upon loss of CPB1 (**Figure 4F**). PPY-derived peptides included the C-terminal amidated versions of the full and partial sequence of pancreatic hormone bioactive region (residues 30-65 and 37-51, Dataset 3). Another bioactive region of PPY, termed pancreatic icosapeptide, was similarly enriched in one of the knockout clones of CPB1 when compared to wildtype. We measured the extracellular peptidome secreted by CPB1[-/-] organoids but could not find significantly enriched sequences. Although this is potentially due to sensitivity and the low levels of extracellular peptides, this observation may also result from low extracellular activity of this carboxypeptidase. We next investigated the peptide spectrum related to the DPP4 enzyme. Unlike the carboxypeptidases, DPP4 is a membrane-anchored and secreted aminopeptidase which cleaves dipeptides from the N-terminus of longer peptide substrates[25]. As a transmembrane protein, DPP4 is believed to mainly act on extracellular substrates, including secreted peptide hormones such as GLP-1[26]. We surprisingly find intracellular accumulation of multiple hormone-derived peptides in DPP4 mutants (**Figure 4G-J**). Validating these observations, we found a strong enrichment of proline (P) and alanine (A) residues directly adjacent to the N-terminal residue (**Figure 4H**) in accumulated peptides. The enrichment of P/A in position N+1 is in line with the reported cleavage site specificity for DPP4[27]. We found a broad range of intracellular substrates for DPP4 (**Figure 4I** and Dataset 3). Among these, we identified peptide pairs that only differed by the cleavage of the XP/A dipeptide when comparing DPP4[-/-] to wildtype clones (**Figure 4J**). These peptides indicated intracellular DPP4 activity, as the cleaved versions were enriched in wildtype while the XP/A-containing versions were enriched in the DPP4 knockouts. For example, one of the few bioactive products of secretogranin-2 (SCG2), named manserin, was found as a direct substrate of DPP4 (**Figure 4J**). We identified extracellular targets of DPP4 covering a broader range of substrates and displaying lower cleavage specificity (**Figure S6D-F**), which was reflected by lower enrichment of P/A in position N+1 of accumulating peptides (**Figure S6E**). This may result from nonspecific degradation of the secreted peptide hormones by additional extracellular (amino)peptidases.

Discussion

Processing of bioactive peptides from large prohormones is a complex process involving a wide range of peptidases. Due to the obligatory co-localization of hormone and protease, as well as the expression of co-dependent enzymes, processing is ideally studied in a physiological setting. As hormone and protease production in mice and humans differ[10,16], organoids provide the unique opportunity to assess prohormone processing in near-native human gut cells. By exploiting EEC-enrichment in human organoids with CRISPR-Cas9 engineering and applying mass spectrometry-based peptidomic analysis on endogenous peptide products, we describe the substrate spectrum of known peptidases, as well as of the newly described carboxypeptidase CPB1. We have previously shown that expression of hormone transcripts including *GCG, SCT* and *NTS* (here extended to *NPW*) is under the control of BMP signaling, and we now extend this observation to post-translational proteolytic processing. Of note, we could not detect significant production of Secretin peptides, potentially related to the relatively low expression of the prohormone in ileal organoids compared to other hormones (**Fig. S1D**). *SCT* is additionally lower expressed in human compared to mouse tissue[16]. PCSK2 is strongly induced by BMP activation, and accordingly enriched in villus EEC cells positive for NTS *in vivo*. Consequently, full-length glucagon was only observed under BMP-activated conditions in organoids. These findings support previous indications based on antibody staining[28] that the intestine -in addition to alpha-cells in the pancreas- secrete bioactive glucagon. We now append this finding with convincing spectral evidence of processed bioactive peptides.

Previous LC/MS analyses of native human intestinal mucosa could not identify pancreatic-like glucagon peptides. This could potentially be related to the relative scarcity of EECs complicating rare peptide detection, as the same study could not detect other abundant hormones in some tissue samples (such as duodenal Gastrin and stomach Ghrelin in different specimens)[16]. Moreover, measurements from native tissue could contain peptides derived beyond the intestinal epithelium. Our data imply that EECs migrating from crypt to villus alter bioactive output from prohormones through dynamic proteolytic processing, in addition to shifts in hormone gene transcription. This process could potentially be targeted by using small molecule BMP inhibitors. In addition to increasing proglucagon output per EEC, PCSK1-dependent production of GLP-1 would be enhanced in comparison to glucagon. This could favor certain metabolic diseases. In addition, our study unveil many peptides that are specifically processed by proteases without a known function. These datasets could act as a starting point to investigate potentially novel EEC-derived bioactive peptides. Taken together, our study describes the substrates of the EEC proteases and provides a framework for understanding the intestinal component of hormone processing defects in endocrinopathies as well as developing therapeutics influencing this process.

Our study has several limitations. First, dynamics in peptide processing among EECs cannot be assessed *in vivo* for non-abundant hormones due to low sensitivity of peptidomics, hampering validation. We have previously shown that organoid EECs resemble their tissue counterparts, and in this study we validate expression

profiles of EEC enzymes (**Figure 1**). Notably, *PCSK2* expression is found in tissue EECs producing high amounts of *NTS*, previously shown to be villus counterparts of L-cells[14,15,20]. We accordingly only find PCSK2 expression in sorted BMP-activated *NTS*-producing cells. Moreover, although organoid cultures can be enriched for hormone-producing EECs, we are still working on detection limits for some of the bioactive peptides (including glucagon). We detected glucagon in one PCSK1 mutant line, and not in PCSK2 mutants. We can thus not exclude a role of PCSK1 in glucagon generation in EECs. Irrespective of the protease involved, BMP was required for glucagon generation in all organoid lines assessed. Finally, we limited our studies to the distal human small intestine. We opted for this region due to a broad representation of EEC subtypes, but future work could exploit organoid models to investigate region-specific processing along the gastrointestinal tract.

6

Materials and Methods

Cell culture of human intestinal organoids

Tissue from the human ileum of a 74-year old male was obtained from the UMC Utrecht with informed consent of the patient. The patient was diagnosed with colon adenocarcinoma that was resected. A sample from non-transformed, normal mucosa was taken for this study. The study was approved by the UMC Utrecht (Utrecht, the Netherlands) ethical committee and was in accordance with the Declaration of Helsinki

Human small intestinal cells were cultured as described previousl[14,29]. Organoids harboring inducible Neurogenin-3 expression were differentiated towards EECs as described before, for 5 days in differentiation medium. For the first two days of differentiation, Neurogenin3-tdTomato expression was induced using Doxycycline (1 µg/mL). BMP activation was achieved by withdrawing Noggin from 'ENR' and addition of BMP-2 (Peprotech, 50 ng/mL) and BMP-4 (Peprotech, 50ng/mL). BMP ligands were added throughout the 5-day differentiation protocol. Medium was replaced every second day during the entire differentiation.

For the preparation of EEC secretomes, organoids were first washed for 1 hour in warm PBS containing 0.001% BSA and 10mM D-Glucose. Organoids were next stimulated to secrete for 90 minutes in PBS supplemented with 0.001% BSA, 10mM D-Glucose, 10 µM Forskolin (Tocris) and 10 µM 3-Isobutyl-1-methylxanthine (IBMX, Sigma). Medium containing secreted medium was collected and spun at 2000g for 5 minutes, after which supernatant was snap frozen.

Immunostaining

Organoid staining was performed as described previously (Beumer et al., 2018). Intact organoids were collected by gently dissolving the BME in ice-cold PBS. Material was subsequently fixed overnight at 4°C in 4% paraformaldehyde (Sigma). Permeabilization and blocking was next performed by incubating organoids in PBS containing 0.5% Triton X-100 (Sigma) and 2% normal donkey serum (Jackson ImunoResearch) for 30 min at room temperature. All stainings were performed in

blocking buffer (2% normal donkey serum in PBS). For immunofluorescence, we used the following antibodies: mouse anti-CPB1 (1:100; R&D systems, MAB2897), goat anti-DPP4 (1:200; R&D systems, AF1180), mouse anti-GLP1 (1:100; Santa Cruz, sc-73508), rabbit anti-proglucagon (1;100; Cell signaling, CST8233), mouse anti-glucagon (1:100; Mercodia, 50-5051). For immunofluorescence, organoids were incubated with the corresponding secondary antibodies Alexa 488-, 568- and 647-conjugated anti-rabbit or anti-mouse (1:1,000; Thermo Fisher Scientific), Phalloidin-Alexa-568 (A12380) (Thermo Fisher Scientific) in blocking buffer containing 4',6-diamidino-2-phenylindole (DAPI; 1; 1,000, Invitrogen). Sections were embedded in Vectashield (Vector Labs) and imaged using an Sp8 confocal microscope (Leica). Image analysis was performed using ImageJ software. Image contrast was adjusted for visualization purposes.

FISH using RNAscope®

Expression profile of *PCSK2, NTS* and *GCG* was visualized using the RNAScope® Multiplex Fluorescent Reagent Kit v2 (Advanced Cell Diagnostics) according to the manufacturer's protocols[30]. In brief, paraffin embedded sections of the human intestine were deparaffinized, treated with hydrogen peroxide for 10 minutes and boiled in target retrieval buffer for 15 minutes before a 30-minute protease treatment. Using probes for *PCSK2, NTS* and *GCG,* expression was assessed. After amplification and detection using opal dyes, slides were counterstained with DAPI for 30 seconds, mounted using ProLong™ Gold Antifade Mountant (Thermo Fisher scientific) and imaged using a SP8 confocal fluorescent microscope (Leica).

RNA-sequencing and analysis

For data in **Fig. 1A and S1**, we visualized expression of selected markers in our previously published organoid EEC single cell RNA-sequencing dataset (GSE146799). Cells with at least 500 uniquely expressed genes and fewer than 30% mitochondrial reads were included in analysis. Data was clustered and visualized using Seurat (dims=10).

For data in **Fig. 1E and 1F**, we visualized expression of selected markers from a previously published gut epithelium cell atlas, using identical cell type annotations.

For newly generated bulk RNA-sequencing, two independent samples of organoids were differentiated for 5 days towards EECs in the presence and absence of BMPs. All material was collected in Eppendorf tubes containing 350 µL RLT buffer (RNeasy kit, QIAGEN). RNA was extracted using the RNeasy mini kit (QIAGEN) following the manufacturer's instructions. Sequencing libraries were generated using a modified CELseq2 protocol[31]. 75 bp paired-end sequencing of libraries was performed on an Illumina NextSeq platform.

For this newly generated bulk RNA-sequencing dataset, reads were mapped to the human GRCh37 genome assembly. The counted reads were filtered to exclude reads with identical library- and molecule barcodes. Differential gene expression analysis was performed using the DESeq2 package[32].

Genetically engineered organoid models

For inducing frameshift mutations, we designed gRNAs targeting the different pro-teases and cloned these into a SpCas9-EGFP vector (Addgene plasmid #48138) using a protocol described before[33]. sgRNAs were selected using the ATUM gRNA design tool (https://www.atum.bio/eCommerce/cas9/input) (**Table 2**). For inducing point mutations, gRNAs were designed to generate C to T mutations within the ed-iting window of cytosine base editors (**Table 2**). These gRNAs were cloned into pSPgRNA using the same protocol (Addgene plasmid #47108). For generation of PCSK1 and CPB1 reporter organoids, gRNAs were directed close to the C-terminus and cloned into pSPgRNA (Addgene plasmid #47108) (**Table 2**).

To generate homozygous frameshift mutations in the different proteases, organoids were transfected with 15 μg SpCas9-EGFP plasmid containing the locus-specific sgRNA. To generate C to T base edits, organoids were transfected with 5 μg of the gRNA plasmid and 10 μg plasmid encoding the cytosine base editor and eGFP (Ad-dgene plasmid #112100). To generate PCSK1 and CPB1 fluorescent reporter organ-oids, we followed a protocol described before[13]. In brief, organoids were transfected with 5 μg of the locus-specific gRNA plasmid, 5 μg of a targeting plasmid encoding mNeon and 5 μg of a frameselector encoding mCherry, Cas9 and a second gRNA linearizing the targeting plasmid.

Transient transfection using a NEPA21 electroporator was performed as described before[34]. 3–7 days after transfection, organoids were dissociated using TryplE (Try-plE Express; Life Technologies) and sorted on a FACS-ARIA (BD Biosciences) for GFP or mCherry positivity. After sorting, Rho kinase inhibitor (Y-27632 dihydrochlo-ride; 10μM, Abmole) was added for 1 week to support single cell outgrowth.

To establish protease mutant organoid lines, 12-48 organoids were picked 1-2 weeks after sorting. Manually picked organoids were dissociated using TryplE and plated in ~50 μL BME. Approximately 5 μL material was left for genotyping. DNA was extracted from the cells using the Zymogen Quick-DNA microprep kit according to protocol. Regions around the sgRNA target site was amplified using Q5 high fidelity polymerase (NEB) according to manufacturer's protocol. CRISPR/Cas9-mediated indel formation was confirmed by sanger sequencing of these amplicons (Macro-gen). Subsequently, Sanger trace deconvolution was performed with the use of ICE v2 CRISPR analysis tool to call clonal organoid lines with homozygous frameshift mutations at the target site. Heterozygous or homozygous base edits could be called based on immediate manual inspection of sequence. Primers used for amplification and sanger sequencing can be found in **Table 2**.

To establish stable reporter organoid lines, organoids displaying fluorescent cells were picked 1-2 weeks after sorting. Although EECs are rare under expansion con-ditions, occasionally these are generated providing a visual way to clonally expand correctly targeted organoids.

Quantitative PCR analysis

Organoid RNA was isolated using a RNAeasy kit (QIAGEN), following the manufac-

turer's protocol. Quantitative PCR (qPCR) analysis was performed using biological and technical duplicates as described described before [35]. qPCR primers used are described in **Table 2**.

Organoid lysis for intracellular peptide hormone extraction

To extract intracellular peptide hormones from organoids, we first lysed the organoids in 8 M Urea in 50 mM ammonium bicarbonate (pH=8.5). Organoids were homogenized by vortexing, followed by sitting at 4°C for 30 min. The lysates were snap frozen and stored at -80°C until peptide extraction.

Peptide hormone extraction

For intracellular peptide hormone extraction, knock-out and wildtype organoid clones' lysates were slowly thawed at 4°C and spun down for 10 min at 20,000 x g at 4°C and the supernatants, containing proteins and peptides, were kept. Peptides were extracted by the methanol/chloroform method: 1 volume of sample was sequentially and thoroughly mixed with 4 volumes of methanol (Sigma-Aldrich), 1 volume of chloroform (Sigma-Aldrich) and 3 volumes of water. The mixture was centrifuged at 5,000 rpm for 10 min at 4°C and the upper layer, containing the water-soluble peptide fraction, was dried down in a vacuum centrifuge. Then, 3 volumes of methanol were incorporated and centrifuged at 5,000 rpm for 10 min at 4°C. The organic fraction was kept, and LC-MS/MS analysis confirmed that no peptides were present in this fraction.

The dried down peptide samples were resuspended in a reduction buffer (4 mM dithiothreitol (DTT) in 50 mM ammonium bicarbonate) and gently mixed for 1h at room temperature (RT). Following reduction, samples were alkylated in 16 mM iodoacetamide for 30 min at RT and reactions were quenched by addition of 4 mM DTT. Peptide samples were acidified to 5 % formic acid (FA) and loaded into an Oasis HLB µElution plate (Waters) by gently applying positive pressure. Wells were washed with 200 µL 0.1% FA in water followed by a low organic wash with 5% methanol in 1% FA in water. Peptides were eluted twice by addition of 30 µL of 60% methanol in 10% FA in water. Samples were dried down in a vacuum centrifuge and kept at -20°C until LC-MS/MS injection.

For extracellular peptide hormone extraction, 1 mL of knock-out and wildtype organoid clones supernatants were slowly thawed at 4°C, acidified to 5 % FA and loaded into an Oasis HLB µElution plate (Waters) to be processed as previously described. Extracted peptides were dried down in a vacuum centrifuge, and resuspended in a reduction buffer (4 mM dithiothreitol (DTT) in 50 mM ammonium bicarbonate) and gently mixed for 1h at room temperature (RT). Following reduction, samples were alkylated in 16 mM iodoacetamide for 30 min at room temperature and reactions were quenched by addition of 4 mM DTT. Samples were acidified to 5% FA and stored at -20°C until LC-MS/MS injection.

All the steps, unless indicated otherwise, were performed at 4°C to avoid peptide degradation, and all samples were collected and stored in low-binding tubes (LoBind, Sigma Aldrich), to minimize peptide losses.

LC-MS/MS analysis of intracellular and extracellular peptidomes

For mass spectrometry analysis, spectral data were acquired with an Ultimate 3000 system (Thermo Fischer Scientific) coupled to an Orbitrap Exploris 480 mass spectrometer (Thermo Fischer Scientific). Peptides were trapped (Dr Maisch Reprosil C18, 3µM, 2cm x 100µM) before being separated on an analytical column (Agilent Poroshell, EC-C18, 2.7µM, 50cm x 75µM). Solvent B consisted of 80% acetonitrile in 0.1% FA. Trapping of peptides was performed for 2 min in 9% B followed by peptide separation in the analytical column using a gradient of 13–44% B in 95 min. After peptide separation, gradients were followed by a steep increase to 99% B in 3 min, a 5 min wash in 99% B and a 10 min re-equilibration at 9% B. Flow rate was kept at 300 nL/min. The mass spectrometer was operated in data-dependent mode. Peptides were ionized in a nESI source at 2 kV and focused at 70 % amplitude of the RF lens, to improve the ionization of bigger species. Full scan MS1 spectra from 375 - 1600 m/z were acquired in the Orbitrap at a resolution of 60,000 with the AGC target set to 3×10^6 and under automated calculation of maximum injection time. Cycle time for MS2 fragmentation scans was set to 1 second. Only peptides with charged states 2-10 were fragmented, and dynamic exclusion was set to a duration of 16 s. Fragmentation was done using stepped HCD normalized collision energies of 24 and 28%. Fragment ions were accumulated until a target value of 1×10^5 ions was reached or under a maximum injection time of 300 ms, with an isolation window of 1.4 m/z before injection in the Orbitrap for MS2 analysis at a resolution of 15,000.

To validate the presence of glucagon and obtain even better spectral fragmentation, we set up a parallel reaction monitoring (PRM) assay using the same LC-MS setup, solvents and trapping of the peptides on the trap column as described above. Separation on the analytical column was performed in 65 minutes using a gradient of 20-50% B. After peptide separation, identical wash and equilibration was performed as previously described. Flow rate was kept at 300 nL/min. The mass spectrometer was operated in multiplexed mode, obtaining a MS1 scan followed by unscheduled PRM on the precursors detailed in the inclusion list (**Table 3**). An optimized RF lens amplitude was set to 70%, MS1 scans were acquired from 375 – 1600 m/z at a resolution of 120,000, with the normalized AGC target set to 100% and injection time set to 300ms. PRM scans were acquired at a resolution of 240,000, with a quadrupole isolation width of 0.4 m/z, the normalized AGC target set to 3000 % and injection time set to 512 ms. A stepped HCD approach was utilized with normalized collision energies set to 27, 30, and 33 %.

Peptidomics database search

Raw data were searched using the ultrafast search engine MSFragger (version 3.3)[36] through Proteome Discoverer software (version 2.5.0.400). Spectra were extracted and precursors with a mass between 350 and 12,000 Da were kept, and the signal-to-noise ratio was set to 1.5. To minimize search space, data were searched against a tailored database containing 369 reviewed *Homo sapiens* proteins (downloaded from Uniprot on October 2021) that contained any peptide annotation or a chain length of up to 60 amino acids, and appended with the non-hormone identified proteins retrieved from a search using SwissProt human database (840 pro-

teins). For spectral searches, precursor mass tolerances were set to ± 20 ppm and fragment mass tolerances to ± 0.02 Da. Digestion enzyme was set to unspecific, with peptides ranging from 7 to 65 amino acids in length, a maximum charge state for theoretical fragments to match of 4, and automatic clipping of N-terminal methionine. Carbamidomethylation of cysteines was set as static modification, while methionine oxidation, N-terminal acetylation and C-terminal amidation were set as variable modifications, with a maximum co-occurrence of 3 variable modifications per peptide. False discovery rate (FDR) calculations were done using Philosopher (Peptide Prophet, version 4.0)[37] and only high-confidence peptides (1 % FDR) were kept. Minora Feature Detector and Feature Mapper nodes were used for untargeted MS1-based peptide quantification, with default parameters except for the maximum retention time shift and the mass tolerance, which were set to 5 min and 5 ppm respectively in the chromatographic alignment window. Total peptide and protein abundance was normalized across all files, and normalized abundance was used for downstream analyses.

Peptidomics data analysis

For PCSK-like hormone processing analysis, protein or peptide abundance is log_2 transformed, z-scaled and plotted across replicates. For glucagon, bioactive regions annotated in Uniprot were extracted from our data and plotted separately. Amidated versions of GRPP and oxyntomodulin, where the amidation site was not followed by a glycine residue, were not included in the plots. Glucagon peptide validation was performed by manual inspection of the MS1 traces on each file using Skyline Daily (version 21.2.1.377)[38]. Isotopic distributions that highly resembled the theoretical isotope distribution of the ion (idotp ≥ 0.88), with a similar peak shape, low mass error and eluting at expected retention time range (min 89 to 91) were kept. Total abundance was calculated by summation of intensities for each isotope (up to M+4) in each sample.

PRM data was analyzed using Byonic (version 4.3.4) and Skyline Daily. Byonic searches were set to ± 10 ppm precursor tolerance, ± 20 ppm fragment mass tolerance, and limited to peptides with 0 miscleavages. A maximum of 2 variable modifications was set including methionine oxidation and C-terminal amidation, and using a fixed cysteine carbamidomethylation modification. Spectra were searched against a database of 8 glucagon peptides and decoy peptides. Only high confidence peptides were considered (1% FDR). The identified glucagon peak was further visually assessed using Skyline. Important aspects were similar peak shape of all transitions, low measured MS1 and MS2 mass errors and MS1 isotope distribution. Spectra were plotted using Interactive Peptide Spectral Annotator[39] and refined in Illustrator.

For CPB1, CPE and DPP4 specificity analysis, peptides found in all injection replicates of at least one organoid clone were included for statistical analysis. Peptide abundance was log_2 transformed and missing values were imputed according to the normal distribution of each sample, and Student's T-test was used to determine which peptides were significantly changing in response to enzyme knockouts when compared to wildtypes. Peptides were considered significantly changing if corrected *p*-values (*q*-value, calculated using the permutation method with 250 iterations) were

inferior to 0.05. Peptide abundance was scaled (z-score) and hierarchically clustered to reveal the subset of peptides accumulating in response to enzyme knockout. The N-terminal or C-terminal cleavage specificities were plotted using the R package 'ggseqlogo'[40]. Data processing was done in Perseus (version 1.6.14.0)[41] while data visualization was done using basic R packages and refined in Illustrator.

Data availability

Peptidomics data, together with curated results, have been deposited to Proteome-eXchange Consortium via the PRIDE repository[42] and can be accessed through the identifier PXD033369.

6

Author Contributions: J.B., J.B.M, W.W. and H.C conceptualized the project, designed the experiments, interpreted the results and wrote the manuscript. T.S.V. assisted in proteomic analysis and discussions. J.B. performed RNA-sequencing analysis and organoid experiments. V.G. and C.B. assisted in organoid experiments. J.B.M performed mass spectrometry experiments. F.K.K. provided the glucagon antibody and contributed to the manuscript text.

Competing Interest Statement: HC is head of Roche's Pharma Research and Early Development in Basel, Switzerland since March 2022. JB and HC are inventors on several patents related to organoid technology. HC's full disclosure is given at https://www.uu.nl/staff/JCClevers/.

References

1.   Beumer, J. & Clevers, H. Cell fate specification and differentiation in the adult mammalian intestine. Nature Reviews Molecular Cell Biology Preprint at https://doi.org/10.1038/s41580-020-0278-0 (2020).

2.   Gribble, F. M. & Reimann, F. Function and mechanisms of enteroendocrine cells and gut hormones in metabolism. Nature Reviews Endocrinology vol. 15 226–237 Preprint at https://doi.org/10.1038/s41574-019-0168-8 (2019).

3.   Kovac, S., Shulkes, A. & Baldwin, G. S. Peptide processing and biology in human disease. Current Opinion in Endocrinology, Diabetes and Obesity vol. 16 79–85 Preprint at https://doi.org/10.1097/MED.0b013e3283202555 (2009).

4.   Stijnen, P., Ramos-Molina, B., O'Rahilly, S. & Creemers, J. W. M. PCSK1 mutations and human endocrinopathies: From obesity to gastrointestinal disorders. Endocrine Reviews vol. 37 347–371 Preprint at https://doi.org/10.1210/er.2015-1117 (2016).

5.   Ji, L., Wu, H. T., Qin, X. Y. & Lan, R. Dissecting carboxypeptidase E: Properties, functions and pathophysiological roles in disease. Endocrine Connections vol. 6 R18–R38 Preprint at https://doi.org/10.1530/EC-17-0020 (2017).

6.   Mest, H. J. & Mentlein, R. Dipeptidyl peptidase inhibitors as new drugs for the treatment of type 2 diabetes. Diabetologia 48, 616–620 (2005).

7.   Ramos-Molina, B., Martin, M. G. & Lindberg, I. PCSK1 Variants and Human Obesity. in Progress in Molecular Biology and Translational Science vol. 140 47–74 (2016).

8.   Stijnen, P., Ramos-Molina, B., O'Rahilly, S. & Creemers, J. W. M. PCSK1 mutations and human endocrinopathies: From obesity to gastrointestinal disorders. Endocrine Reviews vol. 37 347–371 Preprint at https://doi.org/10.1210/er.2015-1117 (2016).

9.   Cawley, N. X. et al. The carboxypeptidase E knockout mouse exhibits endocrinological and behavioral deficits. Endocrinology 145, 5807–5819 (2004).

10.   Beumer, J. et al. High-Resolution mRNA and Secretome Atlas of Human Enteroendocrine Cells. Cell 181, 1291-1306.e19 (2020).

11.   Blanco, E. H., Peinado, J. R., Martín, M. G. & Lindberg, I. Biochemical and cell biological properties of the human prohormone convertase 1/3 Ser357Gly mutation: A PC1/3 hypermorph. Endocrinology (United States) 155, 3434–3447 (2014).

12.   Feng, Y., Reznik, S. E. & Fricker, L. D. ProSAAS and prohormone convertase 1 are broadly expressed during mouse development. Brain research. Gene expression patterns 1, 135–140 (2002).

13.   Artegiani, B. et al. Fast and efficient generation of knock-in human organoids using homology-independent CRISPR–Cas9 precision genome editing. Nature Cell Biology (2020) doi:10.1038/s41556-020-0472-5.

14.   Beumer, J. et al. Enteroendocrine cells switch hormone expression along the crypt-to-villus BMP signalling gradient. Nature Cell Biology vol. 20 909–916 Preprint at https://doi.org/10.1038/s41556-018-0143-y (2018).

15.   Grunddal, K. V. et al. Neurotensin is coexpressed, coreleased, and acts together with GLP-1 and PYY in enteroendocrine control of metabolism. Endocrinology 157, 176–194 (2016).

16.   Roberts, G. P. et al. Comparison of human and murine enteroendocrine cells by transcriptomic and peptidomic profiling. in Diabetes (2019). doi:10.2337/db18-0883.

17.   Quesada, I., Tudurí, E., Ripoll, C. & Nadal, Á. Physiology of the pancreatic α-cell and glucagon secretion: Role in glucose homeostasis and diabetes. Journal of Endocrinology Preprint at https://doi.org/10.1677/JOE-08-0290 (2008).

18.   Fujita, Y., Asadi, A., Yang, G. K., Kwok, Y. N. & Kieffer, T. J. Differential processing of pro-glucose-dependent insulinotropic polypeptide in gut. American Journal of Physiology - Gastrointestinal and Liver Physiology 298, (2010).

19.   Elmentaite, R. et al. Cells of the human intestinal tract mapped across space and time. Nature 597, 250–255 (2021).

20.   Gehart, H. et al. Identification of Enteroendocrine Regulators by Real-Time Single-Cell Differentiation Mapping. Cell 176, 1158-1173.e16 (2019).

21.   Blanco, E. H., Peinado, J. R., Martín, M. G. & Lindberg, I. Biochemical and cell biological properties of the human prohormone convertase 1/3 Ser357Gly mutation: A PC1/3 hypermorph. Endocrinology (United States) 155, 3434–3447 (2014).

22.   Winters, A. et al. Functional analysis of PCSK2 coding variants: A founder effect in the Old Order Amish population. Diabetes Research and Clinical Practice 131, 82–90 (2017).

23.   Feng, Y., Reznik, S. E. & Fricker, L. D. ProSAAS and prohormone convertase 1 are broadly expressed during mouse development. Brain research. Gene expression patterns 1, 135–140 (2002).

24.   Whalley, N. M., Pritchard, L. E., Smith, D. M. & White, A. Processing of proglucagon to GLP-1 in pancreatic α-cells: Is this a paracrine mechanism enabling GLP-1 to act on β-cells? Journal of Endocrinology 211, 99–106 (2011).

25.   Röhrborn, D., Wronkowitz, N. & Eckel, J. DPP4 in diabetes. Frontiers in Immunology vol. 6 Preprint at https://doi.org/10.3389/fimmu.2015.00386 (2015).

26.   Deacon, C. F. Physiology and Pharmacology of DPP-4 in Glucose Homeostasis and the Treatment of Type 2 Diabetes. Frontiers in Endocrinology 10, (2019).

27.   Wagner, L. Dipeptidyl Peptidase 4. in Encyclopedia of Signaling Molecules 1383–1396 (2018). doi:10.1007/978-3-319-67199-4_101580.

28.   Jorsal, T. et al. Investigating Intestinal Glucagon after Roux-en-Y Gastric Bypass Surgery. Journal of Clinical Endocrinology and Metabolism 104, 6403–6416 (2019).

29.   Sato, T. et al. Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. Gastroenterology 141, 1762–1772 (2011).

30.   Wang, F. et al. RNAscope: A novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. Journal of Molecular Diagnostics 14, 22–29 (2012).

31.   Hashimshony, T. et al. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome Biology 17, 77 (2016).

32.   Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15, 550 (2014).

33.   Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. Nat Protoc. 8, (2013).

34.   Fujii, M., Matano, M., Nanki, K. & Sato, T. Efficient genetic engineering of human intestinal organoids using electroporation. Nature Protocols 10, 1474–1485 (2015).

35.   Muñoz, J. et al. The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent "+4" cell markers. EMBO J 31, 3079–3091 (2012).

36.   Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii, A. I. MSFragger: Ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. Nature Methods 14, 513–520 (2017).

37.   da Veiga Leprevost, F. et al. Philosopher: a versatile toolkit for shotgun proteomics data analysis. Nature Methods vol. 17 869–870 Preprint at https://doi.org/10.1038/s41592-020-0912-y (2020).

38.   Schilling, B. et al. Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: Application to protein acetylation and phosphorylation. in Molecular and Cellular Proteomics vol. 11 202–214 (2012).

39.   Brademan, D. R., Riley, N. M., Kwiecien, N. W. & Coon, J. J. Interactive peptide spectral annotator: A versatile web-based tool for proteomic applications. Molecular and Cellular Proteomics 18, S193–S201 (2019).

40.   Wagih, O. Ggseqlogo: A versatile R package for drawing sequence logos. Bioinformatics 33, 3645–3647 (2017).

41.   Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nature Methods vol. 13 731–740 Preprint at https://doi.org/10.1038/nmeth.3901 (2016).

42.   Vizcaíno, J. A. et al. 2016 update of the PRIDE database and its related tools. Nucleic Acids Research 44, D447–D456 (2016).

6

CHAPTER 6 - SI

Supporting information for

Mapping prohormone processing by proteases in human enteroendocrine cells using genetically engineered organoid models

Joep Beumer[1,8], Julia Bauzá-Martinez[2,8], Tim S. Veth[2], Veerle Geurts[1], Charelle Boot[1], Filip Knop[5,6,7], Wei Wu[2,3,4,*], Hans Clevers[1,*]

1 Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences (KNAW) and UMC Utrecht, 3584 CT Utrecht, The Netherlands; Oncode Institute.
2 Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.
3 Singapore Immunology Network (SIgN), Agency for Science, Technology and Research (A*STAR), Singapore 138648, Singapore.
4 Department of Pharmacy, National University of Singapore, Singapore 117543, Singapore.
5 Center for Clinical Metabolic Research, Gentofte Hospital, University of Copenhagen, Hellerup, Denmark
6 Steno Diabetes Center Copenhagen, Herlev, Denmark
7 Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
8 These authors contributed equally.

Based on:

Table 1 Overview of protease mutant organoids. Related to Figure 2

Gene	Clone number	Allele 1	Allele 2
CPB1	8	Min13	Min13
	11	Plus 1	Plus 5
	15	Min1	Min16
CPE	4	Min17	Min17
	8	Min8	Min16
	11	Min1	Min1
	12	Plus 1	Min 1
DPP4	1	Plus 1	Plus 1
	2	Plus 1	Plus 1
	3	Plus 1	Plus 1
	4	Plus 1	Plus 1
	5	Plus 1	Min2
	6	Plus 1	Plus 1
	7	Plus 1	Plus 1
	8	Plus 1	Plus 1
	9	Plus 1	Plus 1
	10	Plus 1	Plus 1
	11	Plus 1	Plus 1
	12	Plus 1	Plus 1
PCSK1N	1	Min 51	x
	2	Plus 1	x
	5	Plus 1	
PCSK1	1	Min 2	Min 2
	3	Min 23	Min 23
	5	Min 14	Min 14
	7	Plus 13	Min 14
PCSK2	1	Plus 1	Plus 2
	9	Plus 1	Plus 1
	11	Plus 1	Min 23
	15	Plus 1	Plus 1
PCSK2 (in PCSK1 clone 1)	2	Plus 1	Min 17
	7	Min 1	Plus 1
	11	Plus 1	Min 16
PCSK1 S357G	3	S357G	S357G
	9	S357G	S357G
PCSK2 R430W	15	R430W	WT

6

Table 3 PRM inclusion list

Name	Compound	m/z	z
Oxyntomodulin	HSQGTFTSDYSKYLDSRRAQDFVQWLMNTKRNRNNIA (light)	742,201	6
GLP-1(1-36a)	HDEFERHAEGTFTSDVSSYLEGQAAKEFIAWLVK-GR[-0.984016] (light)	822,807	5
GLP-1(7-36a)	HAEGTFTSDVSSYLEGQAAKEFIAWLVKGR[-0.984016] (light)	824,922	4
GLP-1(1-37)	HDEFERHAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG (light)	834,409	5
GLP-1(7-37)	HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG (light)	839,424	4
GRPP	RSLQDTEEKSRSFSASQADPLSDPDQMNED (light)	846,628	4
Glucagon	HSQGTFTSDYSKYLDSRRAQDFVQWLMNT (light)	871,161	4
GLP-2	HADGSFSDEMNTILDNLAARDFINWLIQTKITD (light)	941,961	4

6

Table 2 Oligos used in this study

gRNA sequences	
PCSK1KO_Fw	CACCGGGCTGCTGGCATCTTCGCTC
PCSK1KO_rev	AAACGAGCGAAGATGCCAGCAGCCC
PCSK2KO_Fw	CACCGGGGCAAGCTGATGGCACTCC
PCSK2KO_rev	AAACGGAGTGCCATCAGCTTGCCCC
PCSK1NKO_Fw	CACCGGCGCGTCGGGGTCGTCGTCC
PCSK1NKO_rev	AAACGGACGACGACCCCGACGCGCC
CPB1KO_Fw	CACCGGATAGCCGGGTTCGTGCAAC
CPB1KO_rev	AAACGTTGCACGAACCCGGCTATCC
CPEKO_Fw	CACCGGCGAGGCGCTCGTGTCCGTG
CPEKO_rev	AAACCACGGACACGAGCGCCTCGCC
DPP4_Fw	CACCGATTATTCAATATCTCCTGAT
DPP4_rev	AAACATCAGGAGATATTGAATAATC
PCSK2_R430W_fw	CACCGCAGTGGCGGCGCAATGGGGT
PCSK2_R430W_rev	AAACACCCCATTGCGCCGCCACTGC
PCSK1_S357G_fw	CACCGTACAGCAGCGGAGATTACAC
PCSK1_S357G_rev	AAACGTGTAATCTCCGCTGCTGTAC

6
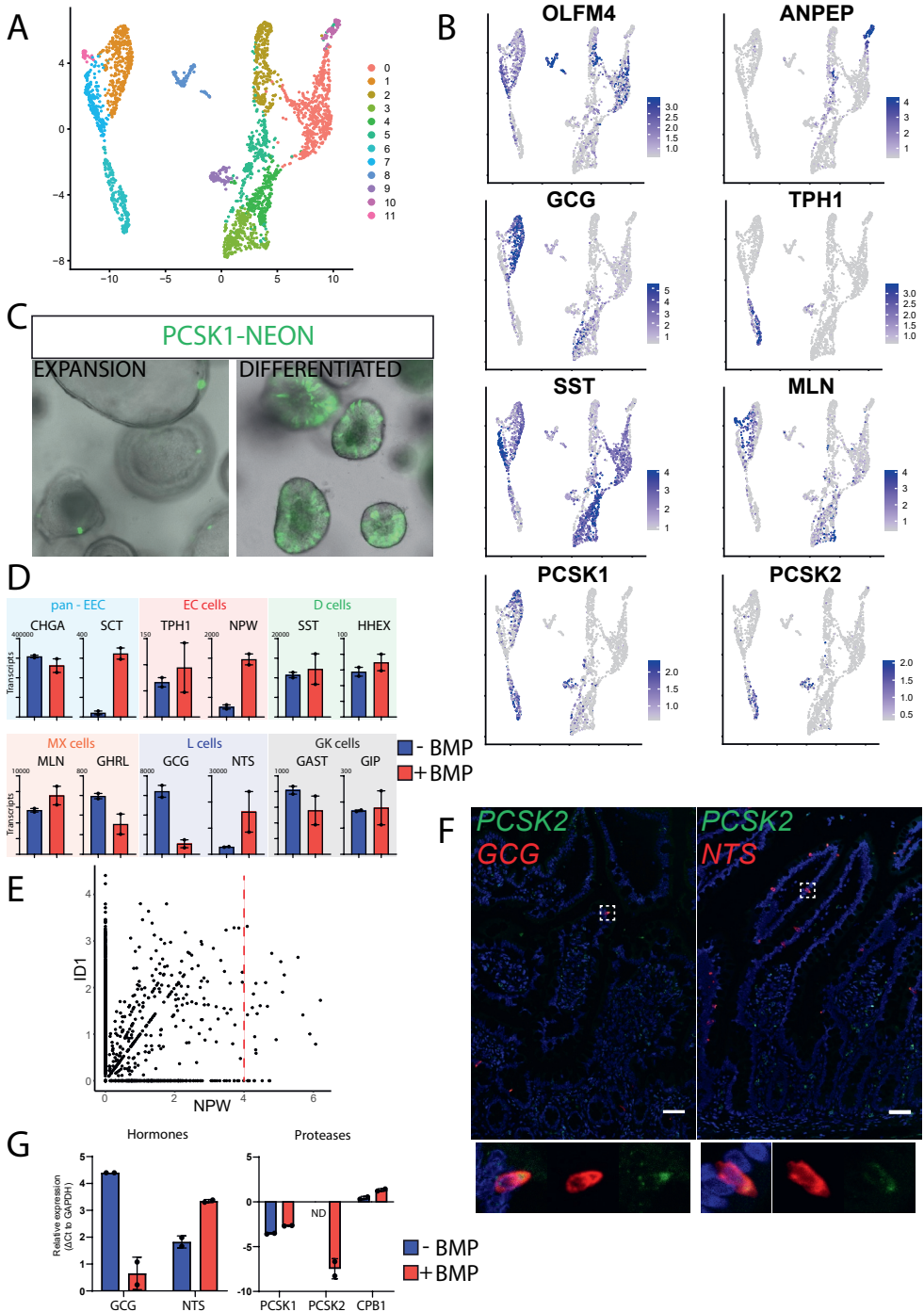
Figure S1. Expression and BMP-dependency of EEC hormones and proteases. (Caption on next page)

Figure S1. Expression and BMP-dependency of EEC hormones and proteases. (A) UMAP showing clustering of different EEC populations derived from organoid cultures based on previously performed single-cell RNA-sequencing (1). (B) Feature map showing normalized transcript levels of selected markers on dataset from (A). (C) Overlay of brightfield and fluorescent image from PCSK1mNeon organoids in culture. PCSK1-neon positive cells are strongly increased upon differentiation. (D) Graphs depicting hormone expression revealed by bulk RNA-sequencing of EEC-differentiated organoids in the presence and absence of BMPs. (E) NPW and ID1 expression plotted from a previously published single cell RNA-sequencing atlas of the gut epithelium (2). Cells expressing the highest levels of NPW (an artificial cut-off of 4 transcripts indicated by red line) are ID1-positive. (F) Fluorescent in situ hybridization (RNAscope) showing low expression of PCSK2 in GCG- and NTS-expressing cells in the human small intestinal epithelium. (G) qPCR analysis on sorted NTS-producing cells in the presence and absence of BMP. BMP stimulates the production of PCSK2. Expression was undetectable in absence of BMP (N.D: non-detectable). Experiment was performed n=2 independent experiments, and Ct values are average of technical replicate. Data is depicted as ΔCt compared to GAPDH

6
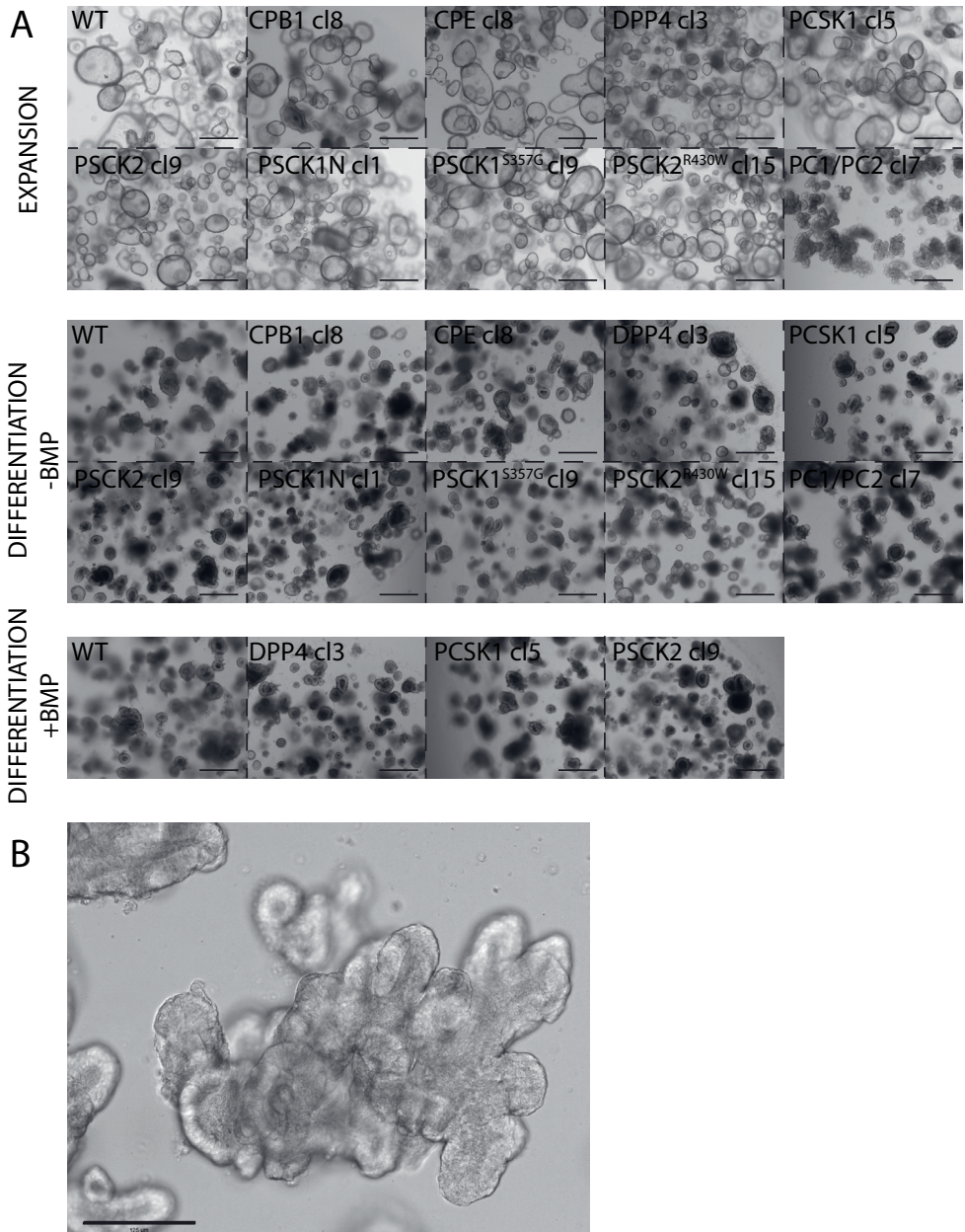
Figure S2. Biobank of protease mutant organoids. (A) Brightfield images of expanding and differentiated organoids harboring different protease mutations. Scale bars are 500 µm. (B) Zoom-in brightfield image of PCSK1/PCSK2 double mutant organoids. Scale bar is 125 µm.
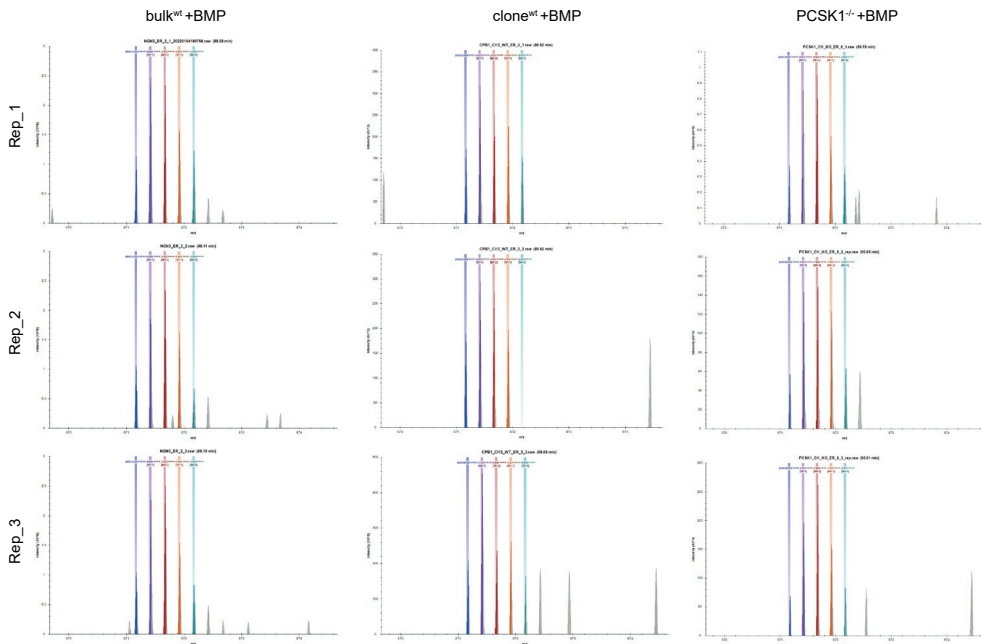
Figure S3. Extracted MS1 isotopic distributions for glucagon peptide in BMP-treated samples. MS1 isotopic distributions are extracted from raw files based on manual inspection using Skyline. The precursor ion (m/z = 871.1612, z = 4) as well as isotopes M+1 to M+4 are plotted. Retention time, similarity to the theoretical ion isotopic distribution (idotp) and low mass error are taken into consideration to determine the ion's presence or absence.

bulk^wt + BMP - PRM-MS - Glucagon MS2 spectra

H S Q G T F T S D Y S K Y L D S R R A Q D F V Q W L M N T

Precursor m/z: 871.1612   Charge: +4   q-value: 4.3e-6   Fragmented Bonds: 26/28
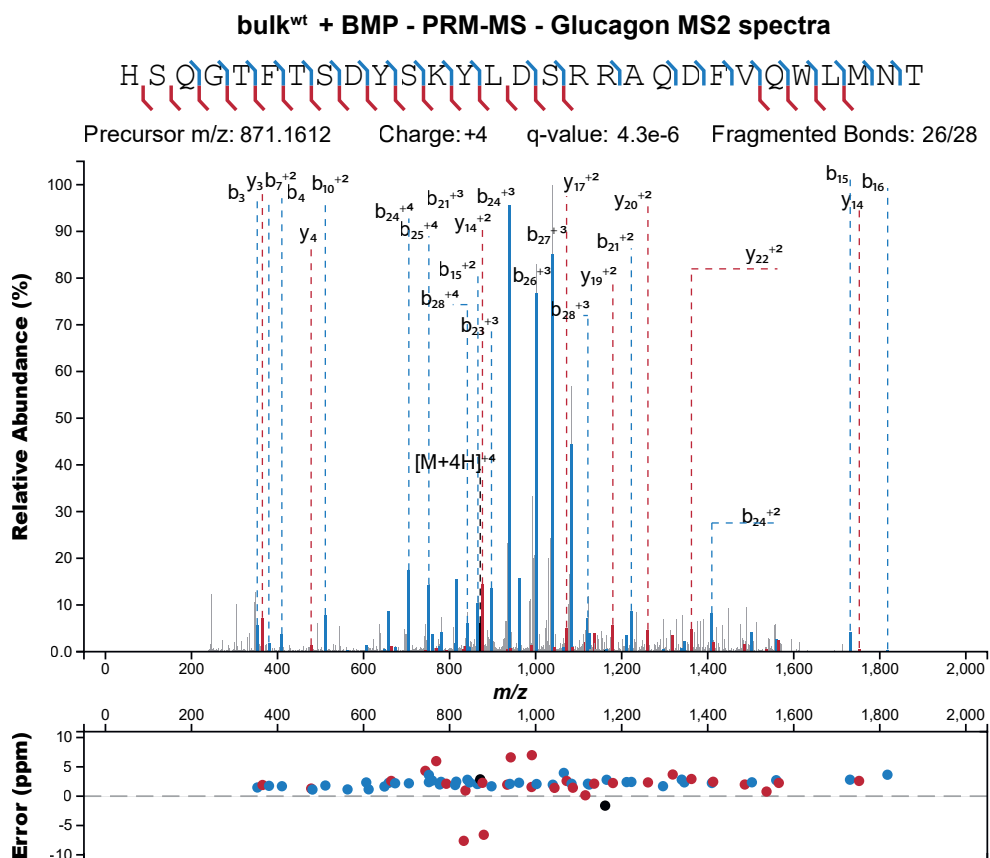
Figure S4. Annotated MS2 spectra for glucagon measured by PRM-MS. Measured MS2 spectrum of the glucagon peptide acquired at a 240k resolution. Only the precursor peak, b-ions and y-ions are used to visualize the spectrum using a 10ppm cut-off. The spectrum shows almost complete fragment annotation of the glucagon peptide with 26 out of 28 fragmented bonds found. Only the highest intensity peaks are labeled in the MS2 spectrum. Spectrum quality was visually assessed and statistically validated using a q-value calculated using a target-decoy strategy, which confirmed glucagon peptide identification.

A
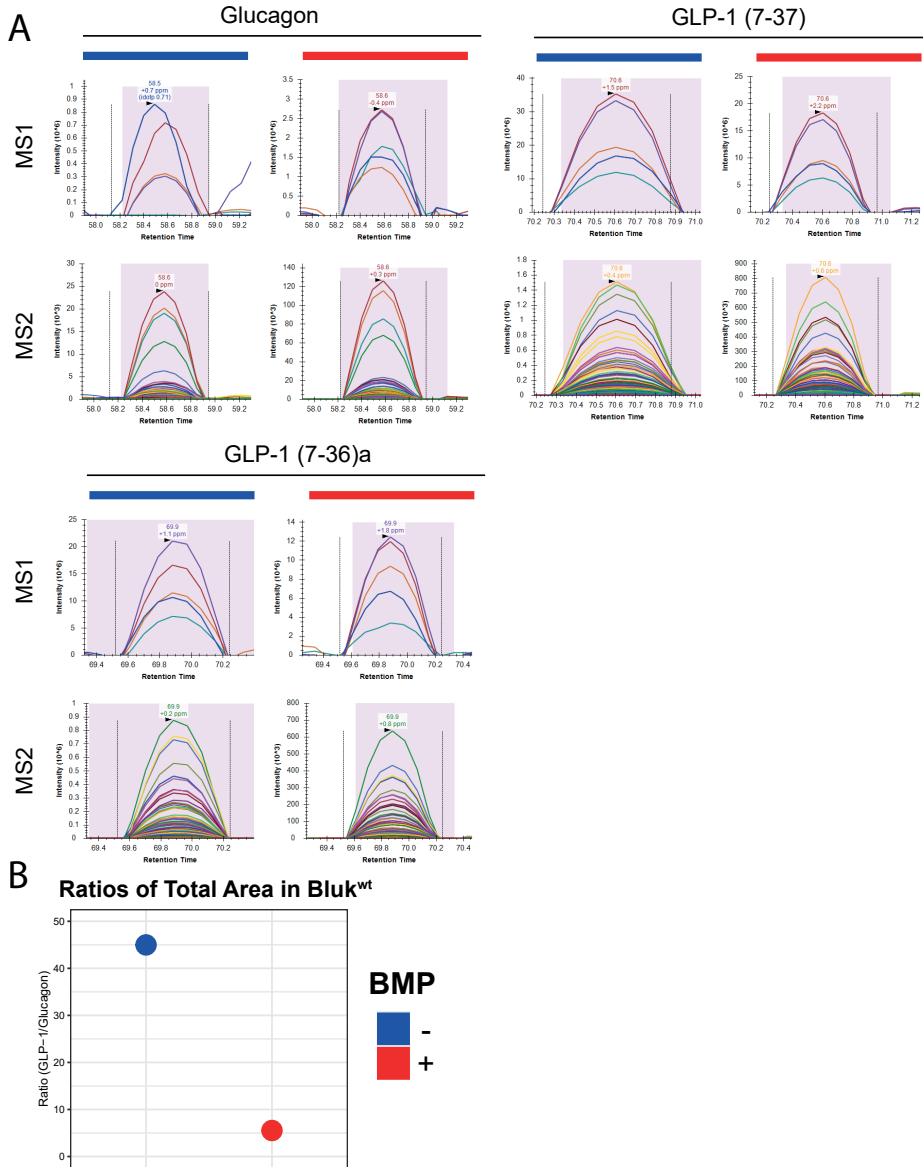
Figure S5. LC-MS/MS traces of glucagon and active GLP-1 peptides as measured by PRM-MS in BMP-treated and -untreated bulk wildtype organoids. (A) MS1 and MS2 traces for bioactive glucagon, GLP-1 (7-37) and amidated GLP-1 (7-36) as measured by PRM-MS in BMP-treated and BMP-untreated bulk wildtype organoids. The precursor ion (m/z = 871.1612, z = 4) as well as isotopes M+1 to M+4 are displayed for MS1, and all matching transitions are displayed for MS2. Mass error, in parts per million (ppm), and retention time are shown on top of the peaks. (B) The GLP-1:glucagon ratio of peak area is displayed for BMP-treated and BMP-untreated bulk wildtype organoids. The total peak area for GLP-1 is calculated by averaging GLP-1 (7-37) and GLP-1 (7-36)a areas.
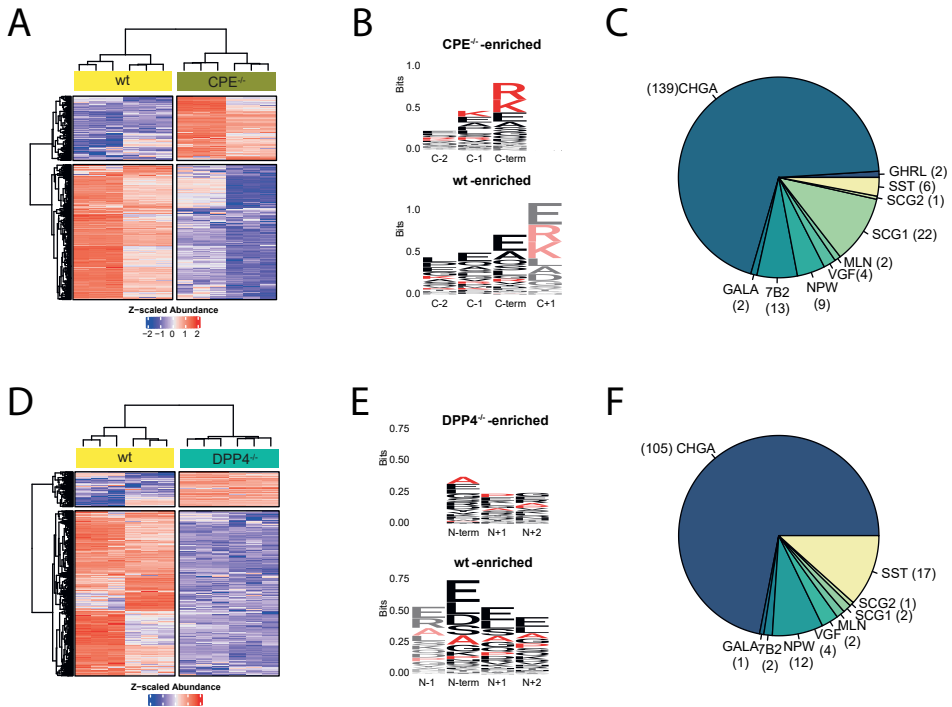
Figure S6. Extracellular substrate specificity of EEC exopeptidases. (A) Significantly changing hormone-derived peptides in response to CPE knockout (CPE-/-) when compared to wildtype organoids. Bulk and a clonal wildtype line and two CPE-/- clones were included. Three technical replicates were measured for all samples. (B) Sequence logo of the C-termini of peptides accumulating in response to CPE knockout (CPE-/--enriched) extracellularly, and peptides accumulating in wildtype (wt-enriched) extracellularly. Lysine (K) and arginine (R) are highlighted in red. (C) Pie chart indicating the substrate repertoire of peptides accumulating in CPE knockouts extracellularly. (D) Significantly changing hormone-derived peptides in response to DPP4 knockout (DPP4-/-) when compared to wildtype organoids. Bulk and a clonal wildtype line and two DPP4-/- clones were included. Three technical replicates were measured for all samples. (E) Sequence logo of the C-termini of peptides accumulating in response to DPP4 knockout (DPP4-/--enriched) extracellularly, and peptides accumulating in wildtype (wt-enriched) extracellularly. Lysine (K) and arginine (R) are highlighted in red. (F) Pie chart indicating the substrate repertoire of peptides accumulating in DPP4 knockouts extracellularly.

CHAPTER 7

Challenges and future outlook

Challenges and future outlook

Biological curiosities can only be pursued as deeply as our tools would allow us. In this respect, the analytical strategies featured in this thesis have made important advances to enable sensitive and detailed studies on specific secreted factors that mediate cell-to-cell communication, namely extracellular vesicles (EVs) and gut-derived peptide hormones. As delineated in **Chapters 2 - 6**, the versatility of peptide-centric mass spectrometry (MS), combined with thoughtful adaptations to boost sensitivity, was instrumental to reveal novel biological insights on the processing and surface architecture of these secreted moieties. The findings herein can contribute to therapeutic design, as EVs and peptide hormones are increasingly viewed as avant-garde biopharmaceutical targets towards clinical development and drug delivery. Nonetheless, in this exciting flurry, there is still room for improvement, where further developments in the field of high-sensitivity MS are still needed to add critical details that can deepen our understanding on how these secreted moieties function and may be controlled.

7

Achievements and remaining challenges of high-sensitivity mass spectrometry

On EV analysis – As discussed in **Chapters 2-4**, working with EVs is challenging, due to their small size, their inherently low sample availability, and their high heterogeneity. This may not seem problematic for traditional bottom-up proteomic approaches, where a few micrograms of protein can provide fairly detailed proteomic readouts with minimal adaptations. Nevertheless, this clashes with the application of other peptide-centric MS approaches, such as ligandomics or crosslinking MS (XLMS), that normally rely on milligrams of starting material. Moreover, such structural MS approaches can also provide new layers of information, which is key to gain the next level of insights into EV docking, signaling and functional biology.

Although EVs have been defined also as antigen presenting entities[1], and despite the exponential growth of the human immunopeptidome project (HIPP) initiative[2,3], a detailed characterization of the antigens bound to HLA class I molecules present in the surface of EVs has been lacking for at least a decade. Before our contribution, barely 500 HLA peptides have been detected, making the coverage really insufficient to address the question, of whether antigens presented in the EVs are roughly the same as antigens presented by the whole-cells. The main reason behind this is that immunoprecipitating HLA-I complexes from less than 1 mg of starting material is technically challenging, although normally more than 10 mg of material are used. In **Chapter 2**, I overcame these technical limitations by downscaling to 100 µg EVs, and obtained the first deep map of the antigens bound to HLA-I molecules presented on B-cell derived EVs. This takes the field a step further, by allowing antigen discovery work to be performed on ever-decreasing amount of input material, but also making it feasible to map the PTMs on antigen peptides, that are loaded in EVs and distributed in the systemic circulation. With these, we can now dream of asking the questions of how PTMs can influence T-cell recognition, and what is the function of these cysteinylated antigens that are being shed in EVs. The fact that we now know antigen presentation by EVs is not exactly the same as whole-cells even though EVs

241

originate from plasma membrane invaginations or evaginations, can also caution us about the intricacies in therapeutic EV production.

Similarly, protein structure is another crucial aspect of EV biology that remained largely difficult to study until recent years. The need to recharacterize protein-protein interactions and protein complex formation in the EV surface and lumen, stems from the fact that EVs have different curvature than whole cells, and have an exclusive interaction interface, given the unique biogenesis route of some of these EVs, via membrane invagination and endosome fusion. Apart from these fundamental differences, characterizing protein-protein interactions and protein complex formation in the EV surface and lumen is also key to better understand how EVs can dock and are taken up by recipient cells. The low stoichiometry and high complexity of EVs remains a perennial hindrance to this kind of work needed. In **Chapter 3**, we put together an intact extracellular vesicle crosslinking mass spectrometry (iEVXL) method that can be applied for structural characterization of protein complexes in EVs, again working with 100 µg EVs. We show that iEVXL can provide high-resolution structural information at less than 0.3 nm, which not only recapitulates the native structure of previously reported complexes but can also aid in mapping unknown protein structures when adequately combined with computational modeling. This is a big enabling step towards the structural characterization of supramolecular complexes in EVs, and functional elucidation of EV protein complexes responsible for homing and docking mechanisms.

Although these technical advancements are a step forward in EV research, obtaining 100 µg of EV proteins is still challenging. Considering the heterogeneity in EVs and EV subpopulations, the challenge remains on how to capture that diversity as each EV subset becomes ever more finely categorized. As the field takes prominent steps towards 'single-cell' proteomics, the question is whether we will ever get to 'single-EV' proteomics. In one particular instance, the development of the novel high field asymmetric waveform ion mobility spectrometry (FAIMS)[4,5] may be beneficial to give that sensitivity boost. FAIMS allows gas-phase fractionation of any sample, right before entry into the mass spectrometer. This leads to complexity reduction for in depth characterization without additional sample preparation, circumventing adsorption loss during preparation of minute starting material.

In FAIMS, different compensation voltages (CVs) are used to filter specific ions of varying charge and ratios (**Figure 1A**) right before entering the mass spectrometer. Therefore, selecting three sufficiently differing CVs (such as -40, -55 and -70) will isolate three complementary set of peptide ions from a given sample (**Figure 1B**). In this way, any given sample can be split into 3 short LC-MS/MS runs, each using one of these 3 CVs, leading to a boost in protein identifications, when compared to runs without FAIMS (**Figure 1C**). Moreover, a reference proteome, in this case 100 ng of HeLa cells digest, can be used as a reference sample to perform match between runs (MBR), further boosting identifications, especially on FAIMS runs (**Figure 1C**, right). As FAIMS helps to remove background contaminants by removing singly charged ions (z=1) at CVs greater than -30 (**Figure 1A**), the resulting FAIMS chromatograms are cleaner, thereby also improving the performance of the MBR algorithm (**Figure 1C**, right). Importantly, protein abundances are largely preserved upon FAIMS+MBR
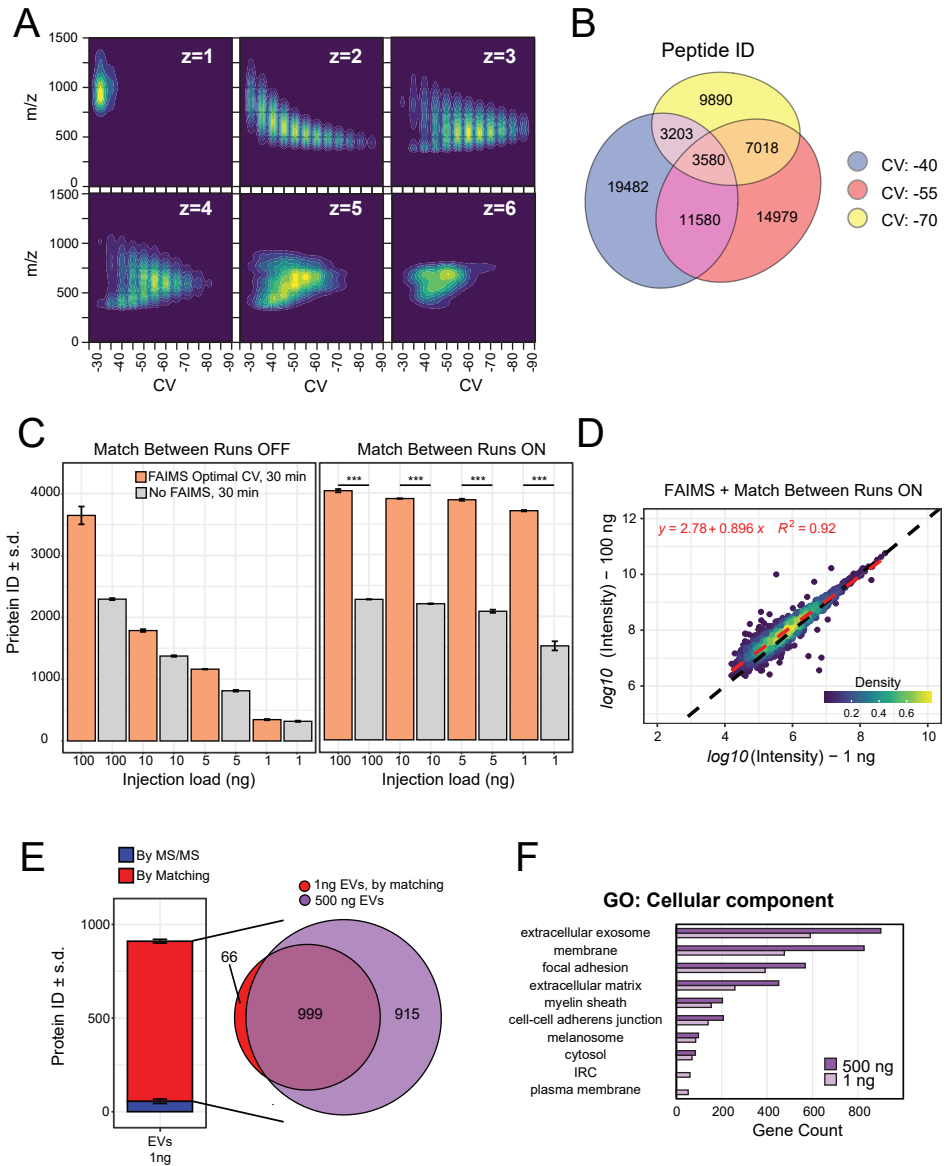
Figure 1. Principles of high field asymmetric waveform ion mobility spectrometry (FAIMS) and application to EV research. (A) Different compensation voltages (CVs) isolate specific ions based on charge (z) and m/z ratio. (B) Optimal FAIMS CV combination to obtain 3 complementary sets of peptides that (C) maximize IDs when compared to runs without FAIMS, especially combined with match between runs (MBR). (D) Abundance correlation between 1 ng and 100 ng HeLa runs using FAIMS+MBR strategy. (E) FAIMS+MBR applied to 1 ng EVs against 500 ng EVs, where a large part of the EV proteome is 'rescued' in 1 ng sample. (F) Gene ontology (GO) analysis showing the top10 enriched terms in 1ng or 500 ng EV samples using the FAIMS+MBR strategy.

strategy, as seen by the linear and very high correlation (R^2=0.92) observed on protein abundance measured in 1 ng *vs.* 100 ng of HeLa digest (**Figure 1D**).

This promising data lends hope towards deeper EV protein detection, or EV characterizations starting from fewer EVs. Indeed, in our preliminary application of FAIMS+MBR to EV analysis, we could reach high levels of EV protein identification, even with 500ng or 1ng of EV protein digest (**Figure 1**), being attributable to the MBR approach (**Figure 1E**). Importantly, gene ontology analysis revealed similar cellular component terms among the top 10 most enriched terms in both the 1ng and 500 ng EV samples (**Figure 1F**), suggesting that though MBR is used, the matched protein identifications still feature prominently EV proteins. However, with most of the identifications being attributed to the MBR algorithm, strict FDR control strategies or careful inspection and validation of individual matches should be implemented to avoid false transfer of identifications between the reference and target samples.

Although this FAIMS+MBR strategy may still need extensive verification and refinement, we believe it could eventually alleviate current challenges related to sample availability when working with EVs and ease the workload required to obtain sufficient EVs for deep proteomic characterization. Moreover, with this FAIMS+MBR strategy, a 'bulk' EV proteome could be measured from higher amounts of starting material, which could be then used as data reference to map less abundant EV subpopulations. Similarly, FAIMS alone could also provide depth of analysis on alternative aspects of EV biology, such as presented HLA peptide antigens[6], post translationally modified protein repertoire[7] or even structural components[8], on samples of even lower input material than what described on **Chapters 2-4**. However, while miniaturized sample preparation strategies for sub-microgram samples are being developed for bottom-up proteomics[9–13], this is still challenging for alternative peptide-centric MS applications and remains to be addressed.

On the field of peptide hormones – In **Chapters 5 and 6**, we applied proteomics and peptidomics to study human gut-derived enteroendocrine cells (EECs) using an EEC-enrichable organoid-based platform. While EECs naturally represent only about 1% of the cells from the intestinal epithelium, inducing EEC differentiation through overexpression of the pro-endocrine transcription factor Neurogenin 3 can boost the percentage of EECs in organoids, enabling their functional study. Using this strategy, we were able to characterize the spectrum of hormones produced by diverse regions of the intestine, as well as the proteolytic targets of a variety of hormone-processing enzymes.

In **Chapter 5**, using this organoid platform, we generated a high-resolution transcriptomic, proteomic and peptidomic atlas of human EECs from different regions of the small intestine. By proteomic characterization of sorted EECs, we were able to validate novel markers that define certain EEC populations. The very low amount of protein retrieved from sorted EECs required some adaptation for proteomic characterization. In this regard, a downscaled trypsin digestion protocol using low volumes, together with strong cation exchange (SCX) cleanup to remove any polymeric contaminants, and short chromatographic gradients to maximize signal, allowed the identification and quantification of over 3,000 proteins, validating many novel

markers for these EEC subpopulations. In this regard, the FAIMS+MBR strategy introduced in **Figure 1** would have been very helpful, but this technology was not available at the time for the instrument of choice. Then, we extracted the peptides and proteins secreted into the extracellular milieu by EEC-enriched organoids and control organoids. Proteomic and peptidomic characterization of these secretomes validated the secretion of a plethora of known and novel hormones, including bio-active peptides, in a region-specific manner. In this initial study, we opted for longer secretion times (24 h) to maximize accumulation of secreted products before sample collection. Unfortunately, the long secretion time entails longer exposure of the secreted products to extracellular peptidases, which was reflected on our data by extensive C-terminal trimming. This was improved in **Chapter 6**, where we opted for shorter secretion times (90 minutes) for secretome characterization.

In **Chapter 6**, we applied peptidomics to understand proteolytic processing of gut hormones on our EEC-enrichable organoid-based platform. Here, we generated a panel of genetically engineered organoids where key hormone-processing enzymes were either knocked-out or mutated. In this case, we opted for measuring the intracellular peptidomes, since these should be protected from unspecific proteolytic degradation while still informing on hormone processing happening inside EECs. Indeed, we detected a plethora of (bioactive) peptides, finding known and novel targets for a panel of hormone processing enzymes. Among our findings, we detected the elusive 29 amino acid peptide *glucagon*, unidentified in previous studies likely due to its very low abundance in the gut. Remarkably, *glucagon* abundance was increased upon bone morphogenic protein (BPM) treatment of the organoid cultures, which induces a villi-like phenotype[14]. To validate the *glucagon* spectrum originally obtained using an untargeted peptidomics strategy, we set up a targeted parallel reaction monitoring (PRM) strategy which allowed us to acquire a very high-resolution spectrum at optimal conditions, proving indeed the unequivocal presence of this important bioactive product.

Particularly, peptidomic analysis of low-availability samples such as sorted EECs or tissue biopsies is challenging, especially when measuring secreted peptides in the extracellular space filled with extracellular peptidases even in endogenous conditions. In this regard, PRM MS could be a great addition to the peptidomics toolbox since it represents the highest sensitivity and resolution method for peptide identification and quantification. In PRM-MS, an inclusion list containing the precursor peptides of interest is entered as 'instructions' for analysis. The mass spectrometer only screens for the precursors provided in this list, which are isolated and fragmented to obtain a very high resolution MS2 spectra in the Orbitrap. Furthermore, labeled peptide standards can be spiked-in to enabling accurate quantitation[15]. Although the analysis is biased to the peptides provided inclusion list, which must be limited to a few tens of peptides[16], this acquisition strategy would be beneficial for very low abundant or short-lived bioactive components, for example with diagnostic purposes.

Mass spectrometry-enabled development of novel endocrine therapeutic and prognostic agents

The gut is considered the largest endocrine organ in the human body[17], with at least 13 different EEC subtypes[18] secreting a huge variety of hormones that signal locally and distally. Essentially, EECs secrete hormones in response to physiological and nutritional signals to regulate key physiological processes including appetite, insulin secretion and bowel movement. Notably, these hormones are increasingly being seen as therapeutic options to target metabolic diseases. For example, the L-cell specific incretin glucagon-like peptide 1 (GLP-1) is already a therapeutic line of intervention in type 2 diabetes mellitus and obesity[19]. Similar to EECs, other cells in our body secrete EVs to communicate locally but also to mediate distal cell-to-cell communication. Just like hormones, EVs trigger specific cellular responses in recipient cells, implying that any cell is an endocrine cell to some extent. Moreover, EVs also regulate key physiological processes, and are involved in multiple diseases, including metabolic diseases, and cancer. Recently, EVs are also being explored as therapeutic and diagnostic tools, and even as targeted drug delivery vectors[20].

Hence, with EVs and peptide hormones being increasingly viewed as targets for biopharmaceutical developments, a detailed characterization of their biomolecular landscape is crucial to access the therapeutic potential and guide such diagnostic use. In a multitude of ways, MS is crucial tool for the high-throughput study of processed proteinaceous components in EVs and EECs, which escape from other 'omics' tools such as mRNA sequencing. With the advent of ultra-high sensitivity MS, proteomic analysis of incredibly small sample inputs is becoming a reality, such that the sensitivity is approaching even single cells. With all my enthusiasm and faith, I believe we will soon be able to measure the (immuno)peptidome of sorted or even single cells.

With the field of ultra-high sensitivity MS starting to become a reality, we will soon be able to gain even more exciting insights on how these secreted moieties function and may be controlled. Improvements in sensitivity could provide key information for personalized (immuno)therapies, where even a few cells swabbed almost non-invasively may suffice to understand, for example, the hormonal landscape of a metabolic disease patient, or the HLA-I ligandome of a cancer patient. Similarly, applying ultra-high sensitivity MS in combination with data-independent acquisition or targeted strategies could mean that blood from a finger-prick could suffice to reveal tumor specific antigens or prognostic protein signatures carried in circulating EVs. Optimistically, with further developments in the fields of sample preparation, instrumentation and analysis, personalized precision medicine could become a reality in the coming decade.

References

1.   Raposo, G. et al. B lymphocytes secrete antigen-presenting vesicles. J Exp Med 183, 1161–1172 (1996).

2.   Admon, A. & Bassani-Sternberg, M. The Human Immunopeptidome Project, a Suggestion for yet another Postgenome Next Big Thing. Molecular & Cellular Proteomics 10, O111.011833 (2011).

3.   Vizcaíno, J. A. et al. The Human Immunopeptidome Project: A Roadmap to Predict and Treat Immune Diseases. Mol Cell Proteomics 19, 31–49 (2020).

4.   Swearingen, K. E. & Moritz, R. L. High-field asymmetric waveform ion mobility spectrometry for mass spectrometry-based proteomics. Expert Review of Proteomics 9, 505–517 (2012).

5.   Bekker-Jensen, D. B. et al. A Compact Quadrupole-Orbitrap Mass Spectrometer with FAIMS Interface Improves Proteome Coverage in Short LC Gradients. Molecular & Cellular Proteomics : MCP 19, 716 (2020).

6.   Klaeger, S. et al. Optimized Liquid and Gas Phase Fractionation Increases HLA-Peptidome Coverage for Primary Cell and Tissue Samples. Mol Cell Proteomics 20, (2021).

7.   Adoni, K. R., Cunningham, D. L., Heath, J. K. & Leney, A. C. FAIMS Enhances the Detection of PTM Crosstalk Sites. Journal of Proteome Research 21, 930–939 (2022).

8.   Schnirch, L. et al. Expanding the Depth and Sensitivity of Cross-Link Identification by Differential Ion Mobility Using High-Field Asymmetric Waveform Ion Mobility Spectrometry. Analytical Chemistry 92, 10495–10503 (2020).

9.   Hughes, C. S. et al. Ultrasensitive proteome analysis using paramagnetic bead technology. Molecular Systems Biology 10, 757 (2014).

10.   Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. Nature Methods 2014 11:3 11, 319–324 (2014).

11.   Hughes, C. S. et al. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. Nature Protocols 2018 14:1 14, 68–85 (2018).

12.   Liang, Y. et al. Fully automated sample processing and analysis workflow for low-input proteome profiling. Analytical Chemistry 93, 1658–1666 (2021).

13.   Kostas, J. C., Greguš, M., Schejbal, J., Ray, S. & Ivanov, A. R. Simple and Efficient Micro-solid-Phase Extraction Tip-Based Sample Preparation Workflow to Enable Sensitive Proteomic Profiling of Limited Samples (200 to 10,000 Cells). Journal of Proteome Research 20, 1676–1688 (2021).

14.   Beumer, J. et al. Enteroendocrine cells switch hormone expression along the crypt-to-villus BMP signalling gradient. Nature Cell Biology 2018 20:8 20, 909–916 (2018).

15.   Carr, S. A. et al. Targeted peptide measurements in biology and medicine: Best practices for mass spectrometry-based assay development using a fit-for-purpose approach. Molecular and Cellular Proteomics 13, 907–917 (2014).

16.   Shi, T. et al. Advances in targeted proteomics and applications to biomedical research. PROTEOMICS 16, 2160–2182 (2016).

17.   Ahlman, H. & Nilsson, O. The gut as the largest endocrine organ in the body. Ann Oncol 12 Suppl 2, (2001).

18.   Beumer, J. et al. High-Resolution mRNA and Secretome Atlas of Human Enteroendocrine Cells. Cell 181, 1291-1306.e19 (2020).

19.   Sharma, D., Verma, S., Vaidya, S., Kalia, K. & Tiwari, V. Recent updates on GLP-1 agonists: Current advancements & challenges. Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie 108, 952–962 (2018).

20.   Zipkin, M. Big pharma buys into exosomes for drug delivery. Nat Biotechnol 38, 1226–1228 (2020).

7

Lay summary

Nederlandse samenvatting

Resum divulgatiu

Resumen divulgativo

List of publications

Curriculum vitae

Acknowledgements

Lay summary

Have you ever wondered how a stomach cell knows that it must roar to tell us that we are hungry? Or how do the cells of the immune system know that we have contracted an infection or cancer? Or how is it possible that can we detect breast cancer in the blood? Well, the answer to many of these questions lies in **intercellular communication,** a process mediated by factors such as **extracellular vesicles** or **peptide hormones**, two protagonists of this doctoral thesis. But what do these intercellular communication factors consist of?

Well, first, even though all our cells contain the same set of genes, passed down from our parents, only certain genes are expressed in each cell type, giving rise to sets of proteins (called **proteomes**) that orchestrate the activity and function of our tissues and organs. Proteins are made up of combinations of **20 amino acids**, and the information that determines the sequence of amino acids that gives rise to each protein is precisely encoded in our genes. Fundamentally, **proteins** are extremely important for cell function, since they allow the interaction and transmission of information between the extracellular space, which surrounds the cells, and the intracellular space and the nucleus, confined by the plasma membrane. Proteins can be in the plasma membrane (like transmembrane proteins), where they detect external signals and transmit them to the interior of the cell causing a response. Proteins can also be confined to the interior of the cell, where they carefully regulate cell activity and function. But the ones that interest me the most are the proteins that are secreted to the outside, for example, inside **extracellular vesicles**, or in the form of **peptide hormones**, to transmit information to neighboring cells. Next, I will try to explain what these intercellular communication factors are and why they are so interesting.

During the investigations carried out in this doctoral thesis, **extracellular vesicles** have been very interesting to me. These vesicles are like tiny (nanometric) bags that *all* our cells produce and that can travel through the blood and reach distant organs transmitting information. In these vesicles, our cells deposit small amounts of their proteins, along with other molecules such as RNA or metabolites. In addition, extracellular vesicles are surrounded by a plasma membrane belonging to the cell from which they originate, and therefore contain many transmembrane proteins that would allow us to identify their origin (liver, lung, breast cells, etc.). This makes extracellular vesicles exceptional candidates for the generation of new therapies, if we study in detail the proteins they contain. For example, a breast cancer cell will secrete extracellular vesicles into the bloodstream that allow us to diagnose breast cancer directly from a drop of blood. Also, the transmembrane proteins of these vesicles can be manipulated to target these vesicles to a specific location in the body. Thus, we could fill the vesicles with drugs and direct them to specific organs to minimize the toxicity of treatments.

On the other hand, in these four years I have also discovered my passion for **peptide hormones**. Many of these hormones are secreted in our intestine, by the so-called **enteroendocrine cells,** which represent only 1% of our entire intestine. To further complicate this, there are more than 10 subtypes of these cells, each specialized in the production of a few hormones. But where do these peptide hormones

251

come from? Surprise, they also come from protein! To generate **peptides**, certain proteins must be cut into smaller pieces by molecular scissors called **proteases**. Peptides can have **bioactivity**, that is, generate a specific function in neighboring cells. A great example of bioactive peptides are the famous *insulin*, responsible for the accumulation of glucose, *leptin*, responsible for the feeling of fullness after eating, and *oxytocin*, known as the hormone of love and regulator of births and contractions. Peptide **hormones**, with their small size and high solubility, have it very easy to travel through the bloodstream and reach distant cells, where they will bind to transmembrane proteins to transmit their message to these cells. In fact, **peptide hormones** also transmit their information in a localized way to the neurons that innervate our intestine, initiating a response in the nervous system. Here's the fun part: if we understand how do enteroendocrine cells, their proteases, and the peptide hormones they produce work, then we can make very interesting therapies. For example, if we discover hormones that generate satiety or a feeling of chronic hunger, we could treat obesity. Also, new therapies against diabetes are hiding among the thousands of peptide hormones produced in our intestines.

However, to make all the therapies mentioned above a reality, we must apply sufficiently sensitive techniques, such as **mass spectrometry**, the main protagonist of this thesis reflected on the cover immediately below the title. This technique allows us to determine the mass of any molecule with accuracy and high sensitivity. In our case, to study peptides and proteins, we are interested in determining the mass of the amino acids that compose them. To do this, we must first cut the huge proteins into small pieces, the peptides, using molecular scissors just like the cells do but in the laboratory. Once we have peptides, we separate them in an instrument called chromatographer, then apply a high voltage to ionize them, converting them into **peptide ions** in gaseous form (Figure 1A, Chapter 1). These peptide ions go into the instrument (the mass spectrometer), where we first measure the total mass of the ion, and then we have the ion collide with molecules of an inert gas to break it up into smaller pieces. The mass of these smaller pieces will always be less than the total mass of the ion and will vary depending on where it has been broken. The combination of all the bits that are generated from an ion is called **a mass spectrum** and is unique for each peptide, allowing us to identify its exact composition (Figure 2, Chapter 1). However, even though mass spectrometry has become a fundamental tool to study various aspects of peptides and proteins (Figure 3, Chapter 1), its application to the study of **extracellular vesicles** and **peptide hormones** can still be improved, since this type of samples are difficult to obtain and contain very small amounts of protein. Finally, the goal of the research described in this doctoral thesis has been to adjust and improve existing mass spectrometry techniques to maximize their sensitivity and obtain as much information on these intercellular communication factors as possible. In this way, we will be able to advance our biological knowledge and, in the future, generate new therapies to improve human health.

Chapter 1 represents a general introduction and is separated into two parts. In the **first part** we visit in detail the fundamental technical concepts of mass spectrometry and the alternatives that allow us to maximize the sensitivity of this technique. In the **second part** we visit the fundamental biological concepts, briefly described in this summary, which highlight the importance of the study of intercellular communication

factors, focusing on extracellular vesicles and peptide hormones. The next three chapters focus on research in the field of extracellular vesicles.

The **Chapter 2** describes the application of a variant of mass spectrometry called 'immunopeptidomics' on the elusive extracellular vesicles. Immunopeptidomics is responsible for studying peptides (called antigens) that our cells generate from *every protein* that is inside them and that are essential for developing vaccines. These antigens bind to a transmembrane protein called Human Leukocyte Antigen (HLA) that inserts into the plasma membrane and 'teaches' the immune system the health status of the cell. For example, if we had a mutated tumor cell, the antigens would contain peptides with mutations and, therefore, would be recognized by the immune system as 'dangerous', leading to the destruction of the sick cell. When we started this research, we wondered if extracellular vesicles could also show these antigens to the immune system, since if this were the case we could use them to open new avenues of diagnosis or treatment against cancer (for example, antitumor vaccines). After maximizing the sensitivity of the immunopeptidomic technique, we managed to apply it to just 100 micrograms of extracellular vesicles, which is a significant advance compared to the 10 to 20 milligrams normally used. In addition, we discovered that not only are these vesicles capable of presenting antigens, but they also present more, and a broader and different repertoire than that presented by the cells from which they originate.

Chapter 3 describes the application of a variant of mass spectrometry called 'cross-linking mass spectrometry' to the study of extracellular vesicles. This variant is very useful, since it allows us to obtain *structural information* of the proteins present in a sample. By structural information we mean how proteins fold in 3D, rather than simply studying their amino acid sequence (2D). Extracellular vesicles, as we have already mentioned, are full of transmembrane proteins, which help them find their targets cells. These proteins and others contained within the vesicles play an important role in the interaction and content delivery to a target cell and it is therefore important to know how they fold and organize. However, it is very difficult to study the structural aspect of extracellular vesicles, since it is a highly heterogeneous sample with nanometric dimensions. In this chapter, we solve these problems by developing 'iEVXL', a mass spectrometric approach to study the structure of proteins present in intact vesicles, without breaking or altering them. We demonstrate the utility and validity of the approach by studying multiple cases of known proteins, even predicting new structures never reported in cells or vesicles, such as for Annexin A2 dimers.

In **Chapter 4** we study the effect that these vesicles generate when they reach their target cell. Specifically, we put vesicles derived from cardiac stem cells in contact with endothelial cells (those that are present in the heart and are damaged when there is a heart attack). What we see is that, if we generate an artificial wound, the endothelial cells can close this wound faster when they are in the presence of the extracellular vesicles derived from cardiac stem cells. We also see that certain phosphoproteins could mediate this effect. Phosphoproteins are proteins that are phosphorylated in response to a signal, and that help transmit this signal within cells, mediating the exchange of information between the extracellular and intracellular space. We believe that the regeneration of endothelial cells could be due to the inter-

action between the vesicle and the cell, through proteins located in their respective plasma membranes that could activate phosphoproteins to regulate regeneration. This is where mass spectrometry comes in, to study the changes that occur in the proteome and phosphoproteome of these cells when in contact with the vesicles. We discovered that the cell initiates activation and migration programs, reflected in the activation of multiple phosphoproteins, in response mainly to a protein called PAPPA present in extracellular vesicles. Therefore, we conclude that vesicles with PAPPA or other proteins could favor cardiac regeneration and therefore could be explored as therapeutic options.

Next, in **Chapters 5** and **6**, we study peptide hormones produced in the human gut, combining organoid technology, gene editing, and a variant of mass spectrometry called 'peptidomics'. Thus, in **Chapter 5**, we generate for the first time an atlas of all human entero-endocrine cell populations, discovering new populations that do not exist in mice, the most widely used model organism. This is possible after genetically modifying the organoids to induce up to 50% of entero-endocrine cells which, if you remember, only represent 1% of our intestine. Then we apply the peptidomics to characterize organoid secretions generated from proximal (duodenum) and distal (ileum) areas of the intestine. Thus, we describe a multitude of peptide hormones in these regions, validating new hormones whose secretion in humans was unknown to date due to their negligible abundance. Next, in **chapter 6**, we continue with the gene editing of organoids, this time to eliminate the 'molecular scissors' (proteases) that generate peptide hormones inside entero-endocrine cells. In this way, we want to study which protease cuts which hormone and potentially discover new treatment targets. The most relevant finding of this chapter is the production of glucagon in the human gut, a hormone that was thought to be exclusively pancreatic and whose function is to increase blood glucose levels. In fact, the production of the glucagon peptide from 'proglucagon' protein seems to be dependent on the presence of BMP (another protein that controls the maturation of entero-endocrine cells along the crypt-villi axis) and on the protease that can produce this peptide (PCSK2). Finally, we study the substrates of a panel of proteases, including some proteases that are currently the target of diabetes treatments focused on stopping the degradation of a peptide hormone called GLP-1. In this case, we discovered not only an apparent intracellular activity of the protease responsible for GLP-1 degradation (DPP4), but also a wide range of substrates that goes beyond GLP-1, which should be considered in the use of these well-known therapies that inhibit this protease to prevent unwanted side effects.

Finally, in **Chapter 7**, I discuss the remaining challenges in the field of high-sensitivity mass spectrometric analysis in the realm of intercellular communication factors presented in this thesis. In addition, I raise the future perspectives in the application of high sensitivity mass spectrometry to any sample containing peptides or proteins.

Nederlandse samenvatting

Heb je je ooit afgevraagd hoe een maagcel weet dat hij ons moet vertellen dat we honger hebben? Of hoe de cellen van het immuunsysteem weten dat we een infectie hebben opgelopen of kanker hebben? Of hoe we borstkanker kunnen zien in het bloed? Het antwoord op veel van deze vragen ligt in de **intercellulaire communicatie,** een proces dat wordt gemedieerd door factoren zoals **extracellulaire blaasjes** of **peptidehormonen**, twee protagonisten van dit proefschrift. Maar waaruit bestaan deze intercellulaire communicatiefactoren?

Hoewel al onze cellen dezelfde genen bevatten, doorgegeven door onze ouders, worden in elk celtype slechts bepaalde genen tot expressie gebracht, wat zorgt voor productie van verschillende sets eiwitten (**proteomen**) die de activiteit en functie van onze weefsels en organen uitvoeren. Eiwitten zijn opgebouwd uit verschillende combinaties van **20 aminozuren**, en de informatie die de volgorde van aminozuren bepaalt staat gecodeerd in onze genen. **Eiwitten** zijn extreem belangrijk voor de celfunctie, omdat ze de interactie en overdracht van informatie mogelijk maken tussen de extracellulaire ruimte, die de cellen omringt, en de intracellulaire ruimte, waartussen zich het plasmamembraan bevindt. Eiwitten kunnen zich in het plasmamembraan bevinden (zoals transmembraaneiwitten), waar ze externe signalen opvangen en deze naar het binnenste van de cel doorgeven en een intracellulaire reactie veroorzaken. Eiwitten kunnen zich ook intracellulair bevinden, waar ze de celactiviteit en -functie zorgvuldig reguleren. Maar de eiwitten die mij het meest interesseren, zijn de eiwitten die worden uitgescheiden, bijvoorbeeld in **extracellulaire blaasjes**, of in de vorm van **peptidehormonen**, om informatie door te geven aan andere cellen. In de volgende paragraaf zal ik proberen uit te leggen wat deze intercellulaire communicatiefactoren zijn en waarom ze zo interessant zijn.

Tijdens de studies die in dit proefschrift zijn uitgevoerd, had ik altijd veel interesse in **extracellulaire blaasjes**. Deze blaasjes zijn kleine (nanometrische) zakjes die *al* onze cellen uitscheiden en vervolgens door het bloed kunnen reizen en verre organen kunnen bereiken om informatie door te geven. In deze blaasjes zetten de producerende cellen kleine hoeveelheden van hun eiwitten af, samen met andere moleculen zoals RNA of metabolieten. Bovendien zijn extracellulaire blaasjes omringd door een plasmamembraan dat hoort bij de cel waarvan ze afkomstig zijn, en daarom bevatten ze veel transmembraaneiwitten waarmee we hun oorspronkelijk weefsel kunnen identificeren (lever, long, borstcellen, enz.). Dit maakt extracellulaire blaasjes zeer interessant voor het ontwikkelen van nieuwe therapieën, als we de eiwitten die ze bevatten in detail bestuderen. Een borstkankercel zal bijvoorbeeld extracellulaire blaasjes in de bloedbaan afscheiden die wij kunnen gebruiken om borstkanker rechtstreeks met een druppel bloed te diagnosticeren. Ook kunnen de transmembraaneiwitten van deze blaasjes worden gemanipuleerd om deze blaasjes op een specifieke locatie in het lichaam te krijgen. Zo zouden we de blaasjes met medicijnen kunnen vullen en ze naar specifieke organen kunnen sturen om de toxiciteit van de therapie te minimaliseren.

Aan de andere kant heb ik in de laatste vier jaar ook mijn passie voor **peptidehormonen** ontdekt. Veel van deze hormonen worden in onze darm uitgescheiden door

255

de zogenaamde **entero-endocriene cellen,** die slechts 1% van onze hele darm vertegenwoordigen. Om het nog ingewikkelder te maken, zijn er meer dan 10 sub-types van deze cellen, elk gespecialiseerd in de productie van enkele hormonen. Maar hoe onstaan deze peptidehormonen vandaan? Verrassing, ze komen ook van eiwitten! Om **peptiden** te maken, moeten eiwitten in kleinere stukjes worden geknipt met een moleculaire schaar die **proteasen** worden genoemd**.** Peptiden kunnen **bio-activiteit** hebben, dat wil zeggen een specifieke reactie uitlokken in andere cellen. Een goed voorbeeld van bioactieve peptiden zijn *insuline*, verantwoordelijk voor de opslag van glucose, *leptine*, verantwoordelijk voor het gevoel van volheid na het eten, en *oxytocine*, bekend als het knuffelhormoon en opwekker van weeën en geboorte. **Peptidehormonen,** met hun kleine omvang en hoge oplosbaarheid in het bloed, kunnen zeer gemakkelijk door het bloed reizen en verre cellen bereiken, waar ze binden aan transmembraaneiwitten van andere cellen om hun boodschap door te geven. **peptidehormonen** kunnen hun informatie ook op lokaal niveau informatie verspreiden, bijvoorbeeld aan zenuwuiteinden die naar de darm lopen, waardoor een reactie in het zenuwstelsel wordt geïnitieerd. Het leuke hier is, als we begrijpen hoe entero-endocriene cellen, hun proteasen en peptidehormonen werken, dan kunnen we die informatie gebruiken om nieuwe therapieën te ontwikkelen. Als we bijvoorbeeld hormonen ontdekken die verzadiging of een chronisch hongergevoel opwekken, zouden we daarmee obesitas kunnen behandelen. Ook verschuilen zich nieuwe therapieën tegen diabetes tussen de duizenden peptidehormonen die in onze darmen worden geproduceerd.

Om alle bovengenoemde therapieën te realiseren, moeten we echter voldoende gevoelige technieken toepassen, zoals **massaspectrometrie**, de hoofdpersoon van dit proefschrift weerspiegeld op de omslag van dit proefschrift onder de titel. Met deze techniek kunnen we de massa van elk molecuul nauwkeurig en met hoge gevoeligheid meten. In ons geval, om peptiden en eiwitten te bestuderen, zijn we geïnteresseerd in het bepalen van de massa van de aminozuren waaruit ze zijn samengesteld. Om dit te doen, moeten we eerst de enorme eiwitten in kleine stukjes knippen, peptiden, met behulp van een proteases, net zoals cellen dat doen, maar dan in een reageerbuis. Zodra we peptiden hebben, scheiden we ze van elkaar in een instrument dat chromatograaf wordt genoemd, en passen we een hoge spanning toe om ze te ioniseren en ze als **peptide-ionen** in de gasfase te brengen (Figuur 1A, Hoofdstuk 1). Deze peptide-ionen gaan het instrument in (de massaspectrometer), waar we eerst de totale massa van het ion meten, om deze vervolgens te laten botsen met een inert gas om het in kleinere stukjes te breken. De massa van deze kleinere stukjes zal altijd kleiner zijn dan de totale massa van het ion en zal variëren afhankelijk van waar het gebroken is. De combinatie van alle stukjes die uit een ion worden gegenereerd, wordt **een massaspectrum** genoemd en is uniek voor elk peptide, waardoor we de exacte samenstelling en volgorde kunnen identificeren (Figuur 2, Hoofdstuk 1). Hoewel massaspectrometrie een fundamenteel hulpmiddel is geworden om verschillende aspecten van peptiden en eiwitten te bestuderen (Figuur 3, Hoofdstuk 1), kan de toepassing ervan op de studie van **extracellulaire blaasjes** en **peptidehormonen** echter nog steeds worden verbeterd, omdat deze monsters lastig te verkrijgen zijn en ze bevatten bovendien hele kleine hoeveelheden eiwit. Het doel van het onderzoek beschreven in dit proefschrift was dan

ook het aanpassen en verbeteren van bestaande massaspectrometrietechnieken om hun gevoeligheid te maximaliseren om zoveel mogelijk informatie te verkrijgen over deze intercellulaire communicatiefactoren. Op deze manier kunnen we onze biologische kennis vergroten en in de toekomst nieuwe therapieën ontwikkelen om de menselijke gezondheid te verbeteren.

Hoofdstuk 1 vertegenwoordigt een algemene inleiding en is opgedeeld in twee delen. In het **eerste deel** gaan we in detail in op de fundamentele technische concepten van massaspectrometrie en de alternatieven die ons in staat stellen de gevoeligheid van deze techniek te maximaliseren. In het **tweede deel** gaan we in op de fundamentele biologische concepten, kort uitgelegd in deze samenvatting, die het belang benadrukken van de studie van intercellulaire communicatiefactoren, met de nadruk op extracellulaire blaasjes en peptidehormonen. De volgende drie hoofdstukken richten zich op onderzoek op het gebied van extracellulaire blaasjes.

Hoofdstuk 2 beschrijft de toepassing van een variant van massaspectrometrie genaamd 'immunopeptidomics' op de elusieve extracellulaire blaasjes. Immunopeptidomics beschrijft het bestuderen van peptiden (antigenen genaamd) die onze cellen zelf produceren uit *elk eiwit* dat aanwezig is in de cel en die essentieel zijn voor het ontwikkelen van vaccins. Deze antigenen binden aan een transmembraan eiwit genaamd humaan leukocytenantigeen (HLA) dat in het plasmamembraan wordt ingevoegd en 'vertelt' het immuunsysteem de gezondheidstoestand van de cel. Als bijvoorbeeld een tumorcel gemuteerd is zouden de antigeen peptiden mutaties kunnen bevatten en door het immuunsysteem als 'gevaarlijk' worden herkend, wat kan leiden tot vernietiging van de tumorcel. Toen we met dit onderzoek begonnen, vroegen we ons af of extracellulaire blaasjes deze antigenen ook aan het immuunsysteem zouden kunnen tonen, want als dat het geval zou zijn, zouden we ze kunnen gebruiken om nieuwe manieren te vinden voor diagnose of behandeling van kanker (bijvoorbeeld antitumorvaccins). Na het maximaliseren van de gevoeligheid van de immunopeptidomic-techniek, zijn we erin geslaagd om het toe te passen op slechts 100 microgram extracellulaire blaasjes, wat een aanzienlijke vooruitgang is in vergelijking met de 10 tot 20 milligram die normaal wordt gebruikt. Bovendien ontdekten we dat deze blaasjes niet alleen in staat zijn om antigenen te presenteren, maar dat ze ook een breder en ander repertoire antigenen presenteren dan de cellen waaruit ze afkomstig zijn.

Hoofdstuk 3 beschrijft de toepassing van een variant van massaspectrometrie genaamd 'crosslinking mass spectrometry' op de studie van extracellulaire blaasjes. Deze variant is erg nuttig, omdat het ons vertelt over de *structurele informatie* van de eiwitten. Met structurele informatie bedoelen we hoe eiwitten vouwen in 3D, in plaats van simpelweg hun aminozuursequentie (2D) te bestuderen. Extracellulaire blaasjes, zoals we al zeiden, zitten vol met transmembraaneiwitten, die hen helpen hun doelcellen te vinden. Deze eiwitten en andere die zich in de blaasjes bevinden, spelen een belangrijke rol bij de interactie met de cel en de levering van de inhoud aan een doelcel en het is daarom belangrijk om te weten hoe ze zich vouwen en organiseren. Het is echter erg moeilijk om het structurele aspect van extracellulaire blaasjes te bestuderen, omdat het een zeer heterogeen monster is met nanometrische afmetingen. In dit hoofdstuk lossen we deze problemen op door 'iEVXL' te

257

ontwikkelen, een massaspectrometrische methode om de structuur van eiwitten in intacte blaasjes te bestuderen, zonder ze te breken of te veranderen. We demonstreren het nut en de validiteit van de aanpak door meerdere gevallen van bekende eiwitten te bestuderen, en zelfs door nieuwe structuren te voorspellen die nooit eerder zijn gerapporteerd in cellen of blaasjes, zoals voor annexin A2-dimeren.

In **Hoofdstuk 4** bestuderen we het effect dat deze blaasjes uitoefenen wanneer ze hun doelcel bereiken. Hiervoor brengen we blaasjes afkomstig van hartstamcellen in contact met endotheelcellen (die in het hart aanwezig zijn en beschadigd raken bij een hartaanval). Wat we zien is dat, als we een kunstmatige wond maken, de endotheelcellen deze wond sneller kunnen sluiten als ze in de buurt zijn van de extracellulaire blaasjes die afkomstig zijn van hartstamcellen. We zien ook dat bepaalde fosfoproteïnen dit effect kunnen mediëren. Fosfoproteïnen zijn eiwitten die worden gefosforyleerd als reactie op een signaal en die helpen dit signaal in cellen door te geven, waarbij ze de uitwisseling van informatie tussen de extracellulaire en intracellulaire ruimte faciliteren. Volgens ons wordt de regeneratie van endotheelcellen veroorzaakt door de interactie tussen het blaasje en de cel, via eiwitten die zich in de plasmamembranen bevinden en die fosfoproteïnen zouden kunnen activeren om de regeneratie te bevorderen. Dit is waar massaspectrometrie om de hoek komt kijken, om de veranderingen te bestuderen die optreden in het proteoom en fosfoproteoom van deze cellen wanneer ze in contact komen met de blaasjes. We ontdekten dat de cel activerings- en migratieprogramma's initieert, door de activering van meerdere fosfoproteïnen, voornamelijk als reactie op een eiwit genaamd PAPPA dat aanwezig is in de extracellulaire blaasjes. Daarom concluderen we dat blaasjes met PAPPA of andere eiwitten hartregeneratie kunnen bevorderen en als therapeutische opties kunnen worden onderzocht.

Vervolgens bestuderen we in **Hoofdstukken 5** en **6** peptidehormonen die in de menselijke darm worden geproduceerd, waarbij we organoïde technologie, gentherapie en 'peptidomics' combineren. Zo genereren we in **Hoofdstuk 5** voor het eerst een atlas van alle menselijke entero-endocriene celpopulaties, waarin we nieuwe populaties ontdekken die niet voorkomen in muizen, het meest gebruikte modelorganisme. Dit is mogelijk na genetische modificatie van de organoïden, zodat deze voor uit 50% entero-endocriene cellen bestaan die, als je het je herinnert, slechts voor 1% aanwezig zijn in de darmen. Daarna gebruiken we peptidomics om organoïde secreties te karakteriseren die worden geproduceerd uit de proximale (twaalfvingerige darm) en distale (kronkeldarm) delen van de darm. We vinden hier een groot aantal peptidehormonen uit deze regio's, waaronder nieuwe hormonen waarvan de uitscheiding bij mensen tot op heden onbekend was vanwege hun verwaarloosbare concentratie. Vervolgens gaan we in **hoofdstuk 6** verder met de gentherapie van organoïden, dit keer om de 'moleculaire schaar' (proteasen) te elimineren die peptidehormonen maken in entero-endocriene cellen. Op deze manier willen we onderzoeken welke protease welk hormoon produceert en mogelijk nieuwe therapeutische doelwitten ontdekken. De meest relevante ontdekking van dit hoofdstuk is de productie van glucagon in de menselijke darm, een hormoon waarvan werd gedacht dat het uitsluitend door de pancreas wordt geproduceerd en waarvan de functie het verhogen van de bloedglucosespiegels is. De productie van het glucagon-peptide uit het 'proglucagon'-eiwit blijkt afhankelijk te zijn van de aanwezigheid van BMP (een

ander eiwit dat de ontwikkeling van entero-endocriene cellen langs de crypt-villi-as reguleert) en van de protease die proglucagon knipt (PCSK2). Ten slotte bestuderen we de substraten van een paneel van proteasen, waaronder enkele proteasen die momenteel het doelwit zijn van diabetesbehandelingen die de afbraak van het peptidehormoon GLP-1 stoppen. In dit geval ontdekten we niet alleen de intracellulaire activiteit van de protease die verantwoordelijk is voor de GLP-1-afbraak (DPP4), maar ook een breed scala aan substraten naast GLP-1, waarmee rekening moet worden gehouden bij het gebruik van deze therapieën die deze protease remmen om ongewenste bijwerkingen te voorkomen.

Ten slotte bespreek ik in **Hoofdstuk 7** de resterende uitdagingen voor gevoelige massaspectrometrische analyse op het gebied van de intercellulaire communicatiefactoren die in dit proefschrift zijn beschreven. Daarnaast geef ik perspectief op de toepassing van gevoelige massaspectrometrie op elk monster dat peptiden of eiwitten bevat.

A

Resum divulgatiu

No us heu preguntat mai com sap una cèl·lula de l'estómac que s'ha de posar a rugir per indicar-nos que tenim gana?, o com saben les cèl·lules del sistema immune que hem contret una infecció o un càncer?, o com és possible que puguem detectar un càncer de mama a la sang? Bé, la resposta a moltes d'aquestes preguntes es troba en la **comunicació intercel·lular,** un procés mediat per factors com les **vesícules extracel·lulars** o les **hormones peptídiques,** dues protagonistes d'aquesta tesi doctoral. Però, en què consisteixen aquestos factors de comunicació intercel·lular?

Bé, per començar, encara que totes les nostres cèl·lules continguin el mateix conjunt de gens, transmès pels nostres progenitors, en cada tipus cel·lular només certs gens s'expressen, donant lloc a conjunts de proteïnes (anomenats **proteomes**) que orquesten l'activitat i funció dels nostres òrgans i teixits. Les proteïnes estan formades per combinacions de **20 aminoàcids**, i la informació que determina la seqüència d'aminoàcids que dóna lloc a cada proteïna està precisament codificada als nostres gens. Fonamentalment, les **proteïnes** són extremadament importants pel funcionament cel·lular, ja que permeten la interacció i la transmissió d'informació entre l'espai extracel·lular, que envolta les cèl·lules, i l'espai intracel·lular i el nucli, confinats per la membrana plasmàtica. Les proteïnes poden estar localitzades a la membrana plasmàtica (com les proteïnes transmembranals), on detecten senyals externs i els transmeten a l'interior de la cèl·lula causant una resposta. Les proteïnes també poden estar confinades a l'interior de la cèl·lula, on s'encarreguen de regular amb precisió l'activitat i la funció cel·lular. Però les que més m'interessen a mi són les proteïnes que es secreten a l'exterior, per exemple, dins de **vesícules extracel·lulars,** o en forma d'**hormones peptídiques,** per transmetre informació a les cèl·lules veïnes. A continuació, intentaré explicar-vos què són i per què són tan interessants aquestos factors de comunicació intercel·lular.

Durant les investigacions cursades en aquesta tesi doctoral, les **vesícules extracel·lulars** m'han resultat molt interessants. Aquestes vesícules són com petites bossetes (nanomètriques) que produeixen *totes* les nostres cèl·lules i que poden viatjar per la sang i arribar a **òrgans** distants transmetent informació. En aquestes vesícules, les nostres cèl·lules dipositen petites quantitats de les seves proteïnes, juntament amb altres molècules com ARN o metabòlits. A més, les vesícules extracel·lulars estan embolicades de membrana plasmàtica que prové de la cèl·lula de la que s'originen, i per tant contenen moltes proteïnes transmembranals que ens permetrien identificar-ne l'origen (cèl·lula del fetge, pulmonar, mamària, etcètera). Això fa a les vesícules extracel·lulars candidates excepcionals per a la generació de noves teràpies, sempre que estudiem detalladament les proteïnes que contenen. Per exemple, una cèl·lula cancerígena de mama secretarà vesícules extracel·lulars al torrent sanguini que ens permetran diagnosticar el càncer de mama directament en una gota de sang. També, les proteïnes transmembranals d'aquestes vesícules es poden aprofitar per dirigir aquestes vesícules a un lloc específic del cos. Així, podríem omplir les vesícules amb fàrmacs i adreçar-los a **òrgans** concrets per minimitzar la toxicitat de tractaments.

D'altra banda, en aquestos quatre anys també he descobert la meva passió per les

hormones peptídiques. Moltes d'aquestes hormones són secretades al nostre intestí, per les anomenades **cèl·lules entero-endocrines,** que representen només l'1% de tot el nostre intestí. Per complicar més la cosa, n'hi ha més de 10 subtipus d'aquestes cèl·lules, cadascun especialitzat en la producció d'unes poques hormones. Però d'on provenen aquestes hormones peptídiques? ¡Sorpresa, també provenen de les proteïnes! Per generar **pèptids,** certes proteïnes han de ser tallades per unes tisores moleculars, anomenades **proteases,** en trossets més petits. Els pèptids poden tenir **bioactivitat,** és a dir, generar una funció específica en cèl·lules veïnes. Un gran exemple de pèptids bioactius són la famosa *insulina,* responsable de l'acumulació de glucosa, *leptina,* responsable de la sensació de sacietat després de menjar, i *oxitocina,* coneguda com l'hormona de l'amor i reguladora de parts i contraccions. Les **hormones peptídiques**, amb la seva petita mida i la seva alta solubilitat, ho tenen molt fàcil per viatjar pel torrent sanguini i arribar a cèl·lules llunyanes, on s'uniran a les proteïnes transmembrana per transmetre-lis el seu missatge. De fet, les **hormones peptídiques** també transmeten la seva informació de forma localitzada a les neurones que enerven el nostre intestí, iniciant una resposta al sistema nerviós. Aquí ve el més divertit: si entenem com funcionen les cèl·lules entero-endocrines, les seves proteases i les hormones peptídiques que produeixen, aleshores podem fer teràpies molt interessants. Per exemple, si descobrim quines hormones generen sacietat o sensació de gana crònica, podríem tractar l'obesitat. També, noves teràpies contra la diabetis s'amaguen entre els milers d'hormones peptídiques produïdes al nostre intestí.

Malgrat això, per fer realitat totes les teràpies esmentades anteriorment, hem d'aplicar tècniques suficientment sensibles, com l' **espectrometria de masses,** la protagonista més gran d'aquesta tesi reflectida a la portada immediatament sota el títol. Aquesta tècnica ens permet determinar amb precisió i alta sensibilitat la massa de qualsevol molècula. En el nostre cas, per estudiar pèptids i proteïnes, ens interessa determinar la massa dels aminoàcids els componen. Per fer això, primer hem de tallar les enormes proteïnes a trossets (pèptids), usant les mateixes tisores moleculars que usen les cèl·lules, però al laboratori. Quan tenim pèptids, els separem en instrument anomenat cromatògraf, els hi apliquem un alt voltatge i els ionitzem convertint-los en **ions peptídics** en forma gasosa (Figura 1A, Capítol 1). Aquests ions peptídics entren a l'instrument (l'espectròmetre de masses), on primer mesurem la massa total de l'ió, i després fem que l'ió xoqui amb molècules d'un gas inert per fragmentar-lo a trossets més petits. La massa d'aquests trossets sempre **és** inferior a la massa total de l'ió, i varia depenent de per on s'hagi trencat. La combinació de tots els trossets que es generen d'un ió s'anomena **espectre de masses** i **és únic** per a cada pèptid, permetent identificar-ne la composició exacta (Figura 2, Capítol 1). Tot i això, malgrat que l'espectrometria de masses s'ha convertit en una eina fonamental per estudiar diversos aspectes de pèptids i proteïnes (Figura 3, Capítol 1), la seva aplicació a l'estudi de **vesícules extracel·lulars i hormones peptídiques** encara pot millorar, ja que aquest tipus de mostres són difícils d'obtenir i contenen quantitats ínfimes de proteïna. Finalment, l'objectiu de les investigacions descrites en aquesta tesi doctoral ha estat ajustar i millorar les tècniques existents d'espectrometria de masses per maximitzar-ne la sensibilitat i obtenir la màxima informació d'aquests factors de comunicació intercel·lular. D´aquesta manera, podrem avançar

261

el nostre coneixement biològic i, en el futur, generar noves teràpies per millorar la salut humana.

El **Capítol 1** representa una introducció general i està separat en dues parts. A la **primera part** visitem detalladament els conceptes tècnics fonamentals de l'espectrometria de masses i les alternatives que ens permeten maximitzar la sensibilitat d'aquesta tècnica. A la **segona part** visitem els conceptes biològics fonamentals, breument descrits en aquest resum, que denoten la importància de l'estudi de factors comunicació intercel·lular, focalitzant-nos en les vesícules extracel·lulars i les hormones peptídiques. Els tres següents capítols es centren en investigacions realitzades a l'àmbit de les vesícules extracel·lulars.

El **Capítol 2** descriu l'aplicació d'una variant de l'espectrometria de masses anomenada 'immunopeptidòmica' a les elusives vesícules extracel·lulars. La immunopeptidòmica s'encarrega d'estudiar uns pèptids (anomenats antígens) que les nostres cèl·lules generen a partir de totes les proteïnes que es trobin dins d'elles i que són fonamentals per desenvolupar vacunes. Aquests antígens s'uneixen a una proteïna transmembrana anomenada Human Leukocyte Antigen (HLA) que s'insereix a la membrana plasmàtica i 'ensenya' al sistema immune l'estat de salut de la cèl·lula. Per exemple, si tinguéssim una cèl·lula tumoral mutada, els antígens contindrien pèptids amb mutacions i, per tant, serien reconeguts pel sistema immune com a 'perillosos' comportant la destrucció de la cèl·lula malalta. Quan vam iniciar aquesta investigació, ens preguntàvem si les vesícules extracel·lulars també podien 'ensenyar' al sistema immune aquests antígens, ja que si això fos el cas podríem fer-les servir per obrir noves vies de diagnòstic o tractament contra el càncer (per exemple, vacunes antitumorals). Després de maximitzar la sensibilitat de la tècnica d'immunopeptidòmica, aconseguim aplicar-la a l'ínfima quantitat de 100 micrograms de vesícules extracel·lulars, la qual cosa és un avenç significatiu comparat amb els 10 a 20 mil·ligrams usats normalment. A més, descobrim que, no només aquestes vesícules són capaces de presentar antígens, sinó que en presenten més, i un repertori més ampli i diferent del que presenten les cèl·lules de què provenen.

El **Capítol 3** descriu l'aplicació d'una variant de l'espectrometria de masses anomenada 'crosslinking mass spectrometry' a l'estudi de les vesícules extracel·lulars. Aquesta variant és molt útil, ja que ens permet obtenir informació estructural de les proteïnes presents a una mostra. Per informació estructural ens referim a com es pleguen les proteïnes en 3D, en comptes de simplement estudiar-ne la seqüència d'aminoàcids (2D). Les vesícules extracel·lulars, com ja hem comentat, estan plenes de proteïnes transmembrana, que les ajuden a viatjar pel torrent sanguini i trobar les dianes. Aquestes proteïnes i altres contingudes dins de les vesícules juguen, per exemple, un paper important en 'l'entrega' dels seus continguts a una cèl·lula diana i per tant és important saber com es pleguen i s'organitzen. Tot i així, és molt complicat estudiar l'aspecte estructural de les vesícules extracel·lulars, ja que és una mostra altament heterogènia i de dimensions nanomètriques. En aquest capítol, proposem una sol·lució a aquests problemes posant a punt 'iEVXL', una tècnica per estudiar l'estructura de les proteïnes presents a les vesícules en el seu estat intacte, sense trencar-les ni alterar-les. Demostrem la utilitat i validesa de la tècnica estudiant múltiples casos de proteïnes conegudes, arribant fins i tot a predir

noves estructures mai reportades en cèl·lules o vesícules, com per als dímers d'Annexina A2.

El **Capítol 4** estudiem l'efecte que aquestes vesícules produeixen quan arriben a la seva cèl·lula diana. Concretament, posem en contacte vesícules derivades de cèl·lules mare cardíaques amb cèl·lules endotelials (les que són presents al cor i es fan malbé quan hi ha un infart). El que veiem és que, si generem una ferida artificial, les cèl·lules endotelials són capaces de tancar aquesta ferida més ràpid quan estan en presència de les vesícules extracel·lulars derivades de cèl·lules mare cardíaques. També veiem que certes fosfoproteïnes podrien intervindre aquest efecte. Les fosfoproteïnes són proteïnes que es fosforil·len en resposta a un senyal, i que ajuden a transmetre aquest senyal dins de les cèl·lules, promovent l'intercanvi d'informació entre l'espai extracel·lular i intracel·lular. Creiem que la regeneració de les cèl·lules endotelials podria ser deguda a la interacció entre la vesícula i la cèl·lula, a través de proteïnes situades a les seves respectives membranes plasmàtiques que podrien activar fosfoproteïnes per regular la regeneració. Aquí entra l'espectrometria de masses, per estudiar els canvis que succeeixen en el proteoma i el fosofoproteoma d'aquestes cèl·lules en estar en contacte amb les vesícules. Descobrim que la cèl·lula inicia programes d'activació i migració, reflectits en l'activació de múltiples fosfoproteïnes, en resposta principalment a una proteïna anomenada PAPPA present a les vesícules extracel·lulars. Per tant, concloem que certes vesícules amb PAPPA o altres proteïnes, podrien afavorir la regeneració cardíaca i per tant es podrien explorar com a tractaments.

Seguidament, als **capítols 5 i 6,** estudiem les hormones peptídiques produïdes a l'intestí humà, combinant la tecnologia dels organoides, l'edició genètica i una variant de l'espectrometria de masses anomenada 'peptidòmica'. Així, al **capítol 5,** generem per primer cop un atles de totes les poblacions de cèl·lules entero-endocrines humanes, descobrint noves poblacions que no existeixen en ratolí, el model d'estudi més utilitzat. Això és possible després de modificar genèticament els organoides per induir fins a un 50% de cèl·lules entero-endocrines que, si recordeu, només representen l'1% del nostre intestí. Després apliquem la petidòmica per caracteritzar per primera vegada les secrecions d'organoides generats a partir de zones proximals (duodè) i distals (ili) de l'intestí. Així, descrivim en aquestes regions una multitud d'hormones peptídiques, validant noves hormones la secreció de les quals en humans era desconeguda fins ara per la seva ínfima abundància. A continuació, al **capítol 6,** continuem amb l'edició genètica d'organoides, aquesta vegada per eliminar les tisores moleculars (proteases) que generen, a l'interior de les cèl·lules entero-endocrines, les hormones peptídiques. D'aquesta manera, volem estudiar quina proteasa talla quina hormona i potser descobrir noves dianes de tractament. El descobriment més rellevant d'aquest capítol és la producció de glucagó a l'intestí humà, una hormona que es pensava exclusivament pancreàtica que ajuda a augmentar els nivells de glucosa a la sang. En realitat, la producció de glucagó a partir del 'proglucagó' sembla ser dependent de la presència de BMP, una proteïna que controla la maduració de les cèl·lules entero-endocrines al llarg de l'eix cripta-vellositat, i de la proteasa que el talla (PCSK2). Finalment, estudiem els substrats d'un panell de proteases, incloent-hi algunes proteases que són a dia d'avui diana de tractaments contra la diabetis enfocats a frenar la degradació d'una hormona

peptídica anomenada GLP-1. En aquest cas, descobrim no només una aparent activitat intracel·lular de la proteasa responsable de la degradació de GLP-1 (DPP4), sinó també un ampli rang de substrats que va més enllà del GLP-1, la qual cosa s'hauria de tenir en compte al ús de les famoses teràpies que inhibeixen aquesta proteasa, ja que podrien resultar en efectes no desitjats.

Finalment, al **Capítol 7,** discuteixo els desafiaments restants en el camp de l'anàlisi d'espectrometria de masses d'alta sensibilitat en l'àmbit dels factors de comunicació intercel·lular presentats en aquesta tesi. A més, plantejo les perspectives futures en l'aplicació de l'espectrometria de masses d'alta sensibilitat a qualsevol mostra que contingui pèptids o proteïnes.

A

Resumen divulgativo

¿Nunca os habéis preguntado cómo sabe una célula del estómago que se tiene que poner a rugir para indicarnos que tenemos hambre?, o ¿cómo saben las células del sistema inmune que hemos contraído una infección o un cáncer?, o ¿cómo es posible que podamos detectar un cáncer de mama en la sangre? Bueno, la respuesta a muchas de estas preguntas se haya en la **comunicación intercelular**, un proceso mediado por factores como las **vesículas extracelulares** o las **hormonas peptídicas**, dos protagonistas de esta tesis doctoral. Pero ¿en qué consisten estos factores de comunicación intercelular?

Bueno, para empezar, aunque todas nuestras células contengan el mismo conjunto de genes, transmitido por nuestros progenitores, en cada tipo celular sólo ciertos genes se expresan, dando lugar a conjuntos de proteínas (llamados **proteomas**) que orquestan la actividad y función de nuestros órganos y tejidos. Las proteínas están formadas por combinaciones de **20 aminoácidos**, y la información que determina la secuencia de aminoácidos que da lugar a cada proteína está precisamente codificada en nuestros genes. Fundamentalmente, las **proteínas** son extremadamente importantes para el funcionamiento celular, ya que permiten la interacción y transmisión de información entre el espacio extracelular, que rodea las células, y el espacio intracelular y el núcleo, confinados por la membrana plasmática. Las proteínas pueden estar localizadas en la membrana plasmática (como las proteínas transmembranales), donde detectan señales externas y las transmiten al interior de la célula causando una respuesta. Las proteínas también pueden estar confinadas en el interior de la célula, donde se encargan de regular cuidadosamente la actividad y función celular. Pero, las que más me interesan a mí, son las proteínas que se secretan al exterior, por ejemplo, dentro de **vesículas extracelulares**, o en forma de **hormonas peptídicas**, para transmitir información a las células circundantes. A continuación, intentaré explicaros qué son y por qué son tan interesantes estos factores de comunicación intercelular.

Durante las investigaciones cursadas en esta tesis doctoral, las **vesículas extracelulares**, me han resultado muy interesantes. Estas vesículas son como saquitos diminutos (nanométricos) que producen *todas* nuestras **células** y que pueden viajar por la sangre y alcanzar órganos distantes transmitiendo información. En estas vesículas, nuestras células depositan pequeñas cantidades de sus proteínas, junto a otras moléculas cómo ARN o metabolitos. Además, las vesículas extracelulares están envueltas de membrana plasmática perteneciente a la célula de la que se originan, y por tanto contienen muchas proteínas transmembranales que nos permitirían identificar su origen (célula del hígado, pulmonar, mamaria, etcétera). Esto hace a las vesículas extracelulares candidatas excepcionales para la generación de nuevas terapias, siempre y cuando estudiemos en detalle las proteínas que contienen. Por ejemplo, una célula cancerígena de mama secretará vesículas extracelulares al torrente sanguíneo que nos permitirán diagnosticar el cáncer de mama directamente en una gota de sangre. También, las proteínas transmembranales de estas vesículas se pueden aprovechar para dirigir estas vesículas a un lugar específico del cuerpo. Así, podríamos llenar las vesículas con fármacos y dirigirlos a **órganos** concretos para minimizar la toxicidad de tratamientos.

Por otro lugar, en estos cuatro años también he descubierto mi pasión por las **hormonas peptídicas**. Muchas de estas hormonas son secretadas en nuestro intestino, por las llamadas **células entero-endocrinas,** que representan solamente el 1% de todo nuestro intestino. Para complicar más la cosa, existen más de 10 subtipos de estas células, cada uno especializado en la producción de unas pocas hormonas. ¿Pero, de dónde provienen estas hormonas peptídicas? ¡Sorpresa, también provienen de las proteínas! Para generar **péptidos**, ciertas proteínas deben ser cortadas por unas tijeras moleculares, llamadas **proteasas**, en trocitos más pequeños. Los péptidos pueden tener **bioactividad**, es decir, generar una función específica en células vecinas. Un gran ejemplo de péptidos bioactivos son la famosa *insulina*, responsable de la acumulación de glucosa, *leptina*, responsable de la sensación de saciedad tras comer, y *oxitocina*, conocida como la hormona del amor y reguladora de partos y contracciones. Las **hormonas peptídicas**, con su pequeño tamaño y su alta solubilidad, lo tienen muy fácil para viajar por el torrente sanguíneo y llegar a células lejanas, donde se unirán a las proteínas transmembrana para transmitir a estas células su mensaje. De hecho, las **hormonas peptídicas** también transmiten su información de forma localizada a las neuronas que enervan nuestro intestino, iniciando una respuesta en el sistema nervioso. Aquí viene lo divertido: si entendemos cómo funcionan las células entero-endocrinas, sus proteasas y las hormonas peptídicas que producen, entonces podemos hacer terapias muy interesantes. Por ejemplo, si descubrimos que hormonas generan saciedad o una sensación de apetito crónico, podríamos tratar la obesidad. También, nuevas terapias contra la diabetes se esconden entre los miles de hormonas peptídicas producidas en nuestro intestino.

No obstante, para hacer realidad todas las terapias mencionadas anteriormente, debemos aplicar técnicas suficientemente sensibles, cómo la **espectrometría de masas**, la mayor protagonista de esta tesis reflejada en la portada inmediatamente debajo del título. Esta técnica nos permite determinar con precisión y alta sensibilidad la masa de cualquier molécula. En nuestro caso, para estudiar péptidos y proteínas, nos interesa determinar la masa de los aminoácidos los componen. Para hacer esto, primero debemos cortar las enormes proteínas en trocitos, los péptidos, usando las mismas tijeras moleculares que usan las células, pero en el laboratorio. Una vez que tenemos péptidos, los separamos en instrumento llamado cromatógrafo, les aplicamos un alto voltaje y los ionizamos convirtiéndolos en **iones peptídicos** en forma gaseosa (Figura 1A, Capítulo 1). Estos iones peptídicos entran en el instrumento (el espectrómetro de masas), donde primero medimos la masa total del ion, y luego hacemos que el ion choque con moléculas de un gas inerte para fragmentarlo en trocitos más pequeños. La masa de estos trocitos siempre será inferior a la masa total del ion, y variará dependiendo de por donde se haya roto. La combinación de todos los trocitos que se generan de un ion se llama **espectro de masas** y es único para cada péptido, permitiéndonos identificar su composición exacta (Figura 2, Capítulo 1). Sin embargo, pese a que la espectrometría de masas se ha convertido en una herramienta fundamental para estudiar varios aspectos de péptidos y proteínas (Figura 3, Capítulo 1), su aplicación al estudio de **vesículas extracelulares y hormonas peptídicas** aún puede mejorarse, ya que este tipo de muestras son difíciles de obtener y contienen cantidades ínfimas de proteína. Fi-

nalmente, el objetivo de las investigaciones descritas en esta tesis doctoral ha sido ajustar y mejora las técnicas existentes de espectrometría de masas para maximizar su sensibilidad y obtener la máxima información de estos factores de comunicación intercelular. De esta manera, podremos avanzar nuestro conocimiento biológico y en el futuro generar nuevas terapias para mejorar la salud humana.

El **Capítulo 1** representa una introducción general y está separado en dos partes. En la **primera parte** visitamos en detalle los conceptos técnicos fundamentales de la espectrometría de masas y las alternativas que nos permiten maximizar la sensibilidad de esta técnica. En la **segunda parte** visitamos los conceptos biológicos fundamentales, brevemente descritos en este resumen, que denotan la importancia del estudio de factores comunicación intercelular, focalizándonos en las vesículas extracelulares y las hormonas peptídicas. Los tres próximos capítulos se centran en investigaciones realizadas en el ámbito de las vesículas extracelulares.

El **Capítulo 2** describe la aplicación de una variante de la espectrometría de masas llamada 'inmunopeptid**ó**mica' en las elusivas vesículas extracelulares. La inmunopeptid**ó**mica se encarga de estudiar unos péptidos (llamados antígenos) que nuestras células generan a partir de *todo* lo que se encuentren dentro de ellas y que son fundamentales para desarrollar vacunas. Estos antígenos se unen a una proteína transmembrana llamada Human Leukocyte Antigen (HLA) que se inserta en la membrana plasmática y 'enseña' al sistema inmune el estado de salud de la célula. Por ejemplo, si tuviéramos una célula tumoral mutada, los antígenos contendrían péptidos con mutaciones y, por tanto, serían reconocidos por el sistema inmune como 'peligrosos' conllevando la destrucción de la célula enferma. Cuando iniciamos esta investigación, nos preguntábamos si las vesículas extracelulares también podían 'enseñar' al sistema inmune estos antígenos, ya que si esto fuera el caso podríamos usarlas para abrir nuevas vías de diagnostico o tratamiento contra el cáncer (por ejemplo, vacunas antitumorales). Tras maximizar la sensibilidad de la técnica de inmunopeptidómica, conseguimos aplicarla a la ínfima cantidad de 100 microgramos de vesículas extracelulares, lo cual es un avance significativo comparado con los 10 a 20 miligramos usados normalmente. Además, descubrimos que, no solo estas vesículas son capaces de presentar antígenos, sino que presentan más, y un repertorio más amplio y diferente al que presentan las células de las que provienen.

El **Capítulo 3** describe la aplicación de una variante de la espectrometría de masas llamada 'crosslinking mass spectrometry' al estudio de las vesículas extracelulares. Esta variante es muy útil, puesto que nos permite obtener información *estructural* de las proteínas presentes en una muestra. Por información estructural nos referimos a como se pliegan las proteínas en 3D, en vez de simplemente estudiar su secuencia de aminoácidos (2D). Las vesículas extracelulares, como ya hemos comentado, están repletas de proteínas transmembrana, que les ayudan a viajar por el torrente sanguíneo y encontrar a sus dianas. Estas proteínas y otras contenidas dentro de las vesículas juegan, por ejemplo, un papel importante en 'la entrega' de sus contenidos a una célula diana y por tanto es importante saber cómo se pliegan y organizan. Sin embargo, es muy complicado estudiar el aspecto estructural en vesículas extracelulares, ya que es una muestra altamente heterogénea y de dimensiones

nanométricas. En este capítulo, solventamos estos problemas poniendo a punto 'iEVXL', una técnica para estudiar la estructura de las proteínas presentes en las vesículas en su estado intacto, sin romperlas ni alterarlas. Demostramos la utilidad y validez de la técnica estudiando múltiples casos particulares de proteínas conocidas, llegando incluso a predecir nuevas estructuras nunca reportadas en células o vesículas, como para los dímeros de Anexina A2.

El **Capítulo 4** estudiamos el efecto que estas vesículas producen cuando alcanzan a su célula diana. Concretamente, ponemos en contacto vesículas derivadas de células madre cardiacas con células endoteliales (las que están presentes en el corazón y se dañan cuando hay un infarto). Lo que vemos es que, si generamos una herida artificial, las células endoteliales son capaces de cerrar esta herida más rápido cuando están en presencia de las vesículas extracelulares derivadas de células madre cardíacas. También vemos que ciertas fosfoproteínas podrían mediar este efecto. Las fosfoproteínas son proteínas que se fosforilan en respuesta a una señal, y que ayudan a transmitir esta señal dentro de las células, mediando el intercambio de información entre el espacio extracelular e intracelular. Creemos que la regeneración de las células endoteliales podría ser debida a la interacción entre la vesícula y la célula, a través de proteínas situadas en sus respectivas membranas plasmáticas que podrían activar fosfoproteínas que regulan la regeneración. Aquí entra la espectrometría de masas, para estudiar los cambios que suceden en el proteoma y fosofoproteoma de estas células al estar en contacto con las vesículas. Descubrimos que la célula inicia programas de activación y migración, reflejados en la activación de multiples fosfoproteínas, en respuesta principalmente a una proteína llamada PAPPA presente en las vesículas extracelulares. Por lo tanto, concluimos que vesículas con PAPPA u otras proteínas, podrían favorecer la regeneración cardiaca y por tanto se podrían explorar como tratamientos.

Seguidamente, en los **capítulos 5 y 6**, estudiamos las hormonas peptídicas producidas en el intestino humano, combinando la tecnología de los organoides, la edición genética y una variante de la espectrometría de masas llamada 'peptidómica'. Así, en el **capítulo 5**, generamos por primera vez un atlas de todas las poblaciones de células entero-endocrinas humanas, descubriendo nuevas poblaciones que no existen en ratón, el modelo de estudio más usado. Esto es posible tras modificar genéticamente los organoides para inducir hasta un 50% de células entero-endocrinas que, si recordáis, sólo representan el 1% de nuestro intestino. Luego aplicamos la *peptidómica* para caracterizar por primera vez las secreciones de organoides generados a partir de zonas proximales (duodeno) y distales (íleon) del intestino. Así, describimos en estas regiones una multitud de hormonas peptídicas, validando nuevas hormonas cuya secreción en humanos era desconocida hasta la fecha por su ínfima abundancia. A continuación, en el **capítulo 6**, continuamos con la edición genética de organoides, esta vez para eliminar las 'tijeras moleculares' (proteasas) que generan, en el interior de las células entero-endocrinas, las hormonas peptídicas. De esta manera, queremos estudiar qué proteasa corta qué hormona y tal vez descubrir nuevas dianas de tratamiento. El descubrimiento más relevante de este capítulo es la producción de glucagón en el intestino humano, una hormona que se pensaba exclusivamente pancreática que ayuda a aumentar los niveles de glucosa en sangre. En realidad, la producción de glucagón a partir del 'proglucagon' parece

ser dependiente de la presencia de BMP, una proteína que controla la maduración de las células entero-endocrinas a lo largo del eje cripta-vellosidad, y de la proteasa que lo corta (PCSK2). Finalmente, estudiamos los substratos de un panel de proteasas, incluyendo algunas proteasas que son hoy en día diana de tratamientos contra la diabetes enfocados en frenar la degradación de una hormona peptídica llamada GLP-1. En este caso, descubrimos no sólo una aparente actividad intracelular de la proteasa responsable de la degradación de GLP-1 (DPP4), sino también un amplio rango de substratos que va mas allá del GLP-1, lo cual debería tenerse en cuenta en el uso de las famosas terapias que inhiben esta proteasa, ya que podrían resultar en efectos no deseados.

Finalmente, en el **Capítulo 7**, discuto los desafíos restantes en el campo del análisis de espectrometría de masas de alta sensibilidad en el ámbito de los factores comunicación intercelular presentados en esta tesis. Además, planteo las futuras perspectivas en la aplicación de la espectrometría de masas de alta sensibilidad a cualquier muestra que contenga péptidos o proteínas.

A

List of publications

Bauzá-Martinez, J., Heck, A. J. R., & Wu, W. (2021). HLA-B and cysteinylated ligands distinguish the antigen presentation landscape of extracellular vesicles. Communications Biology, 4(1). https://doi.org/10.1038/s42003-021-02364-y

Beumer, J. #, Puschhof, J. #, **Bauzá-Martinez, J.** #, Martínez-Silgado, A., Elmentaite, R., James, K. R., Ross, A., Hendriks, D., Artegiani, B., Busslinger, G. A., Ponsioen, B., Andersson-Rolf, A., Saftien, A., Boot, C., Kretzschmar, K., Geurts, M. H., Bar-Ephraim, Y. E., Pleguezuelos-Manzano, C., Post, Y., Linden, F., Lopez Iglesias, C., Wetering, W. J., Linden, R., Peters, P. J., Heck, A. J. R., Goedhart, J., Snippert, H., Zilbauer, M., Teichmann, S. A., Wu, W., Clevers, H. (2020). High-Resolution mRNA and Secretome Atlas of Human Enteroendocrine Cells. Cell, 181(6), 1291-1306.e19. https://doi.org/10.1016/j.cell.2020.04.036

Elvira, B., Vandenbempt, V., **Bauzá-Martinez, J.**, Crutzen, R., Negueruela, J., Ibrahim, H., Winder, M. L., Brahma, M. K., Vekeriotaite, B., Martens, P. J., Singh, S. P., Rossello, F., Lybaert, P., Otonkoski, T., Gysemans, C., Wu, W., & Gurzov, E. N. (2022). PTPN2 Regulates the Interferon Signaling and Endoplasmic Reticulum Stress Response in Pancreatic β-Cells in Autoimmune Diabetes. Diabetes, 71(4), 653–668. https://doi.org/10.2337/db21-0443

Bauzá-Martinez, J., Armony, G., Pronker, M. F., & Wu, W. (2022). Characterization of protein complexes in extracellular vesicles by intact extracellular vesicle crosslinking mass spectrometry (iEVXL). Journal of Extracellular Vesicles, 11(8), e12245. https://doi.org/10.1002/jev2.12245

Beumer, J. #, **Bauzá-Martinez, J.** #, Veth, T. S., Geurts, V., Boot, C., Gilliam-Vigh, H., Poulsen, S. S., Knop, F. K., Wu, W., Clevers, H. (2022). Mapping prohormone processing by proteases in human enteroendocrine cells using genetically engineered organoid models. Proc Natl Acad Sci USA (accepted).

Roefs, M.T., **Bauzá-Martinez, J.**, Qin, J., van de Wakker, S. I., Olijve, W., Tuinte, R., Rozeboom, M., Mol, E.A., Wu, W., Vader, P., Sluijter, J.P.G. (2022). Cardiac progenitor cell-derived extracellular vesicles promote endothelial cell migration through both associated- and co-isolated proteins. Communications Biology (submitted).

#: these authors contributed equally

Curriculum vitae

I was born the 11th of August of 1994 in Eivissa, Spain, where I grew up discovering my passion for science. At 18, I moved to Vic, a city near Barcelona, to pursue my studies on Biotechnology at the University of Vic. I discovered mass spectrometry during the last year of my bachelor's degree, where Dr. Eliandre de Oliveira offered me the opportunity to carry out my bachelor thesis on plasma peptidomics. After completing my Biotechnology studies, I moved to Barcelona to carry out a master's degree on Biomedical Research offered by Pompeu Fabra University. During this period, I continued to work on plasma peptidomics, joining de Olivera's lab as an assistant researcher. In 2017, excited about the recent creation of the Human Immunopeptidome Project, I contacted Albert Heck to pursue my doctoral studies in his group. In September 2018, I started my doctoral studies on the Biomolecular Mass Spectrometry and Proteomics lab of Utrecht University, under the supervision of Dr. Wei Wu, where I applied mass spectrometry to understand the immunopeptidome, the hormonal peptidome and the proteome of various biological systems, with a focus on intercellular communication. The results of the research conducted during my PhD, 2018-2022, are described in this thesis.

A

Acknowledgements

While every story has an end, every end is just a new beginning. The closure of such an important chapter of my life can only come along with a closing chapter to this book, dedicated to all the people who helped me immensely throughout these four years. And of course, I am writing these last words with teary eyes and a warm heart, but mostly with a huge feeling of gratitude. I must admit that I decided to start a PhD quite impulsively, out of my big passion for science, without really stopping to think about what it really meant. And just like that, four years ago I embarked in this PhD, which has been overall an amazing adventure. Coming to a new country to pursue my passions in an integrative and multicultural environment known as the BMSP group from Utrecht University, is beyond great. Here, not only did I grow into a *significantly* more mature person and scientist, but also discovered hobbies and people that changed my life. Therefore, I consider this chapter the most important in this entire thesis, as I can finally put together some words to thank each and every one of you.

First and foremost, I would like to thank you **Wei** for your excellent supervision and for believing in me from our first interview until now. Four years ago, you warmly welcomed into your recently formed group, putting quite a complicated project in my hands, which seemed it would never work no matter how hard we tried. Luckily, you were always willing to help in the lab with an extra pair of hands, but also to teach us all the tricks needed to perform the most meticulous experiments. Although the journey had a lot of ups and downs, you were always there, making me feel confident in the downs and calming me down when I would get too excited. Thanks as well for never cutting my wings when I proposed all sorts of crazy experiments, but also for your constant supervision even over long distances (10,000 km, such an endocrine supervisor lol). I also really appreciate the more personal side of our relationship, and the many advice we tried to give each other over the last years. I take one that you gave me at the start: "any experiment should be designed so that you can always learn something new from it". With this, I will end by saying that you are doing an outstanding job in the *rougher* path of academia, both as a professor but also as a female role model and I wish you all the best for your new adventures in Singapore.

Albert, I am much honored that you gave me the opportunity to carry out my doctoral studies in your group. Although we did not work together all the time, being Wei my direct supervisor, we still had very useful and inspiring scientific discussions that led to direct progress in my research. I am also very thankful for the times where you lifted me up when I thought my results were 'not good enough'. You are an amazing scientist, and although your ideas are sometimes crazy and hard to bring to life, you count with a group of researchers that truly believe in you, and that with your support, make amazing science possible. Therefore, thanks for being the head of the lab, for keeping the south wing filled with top-notch mass spectrometers and most importantly, for gluing all of us and our science together.

I also want to thank **Laura and Marie**, my paranymphs, friends and a 'support group' during this journey. **Laura**, you were my first contact with this group when I came to Netherlands, even becoming room- and plane-mates during my first weeks here.

Apart from transferring to me all your impressive knowledge on *how to measure polymers by MS*, you welcomed me as a friend, immersing me in the Dutch culture: carnivals, borrels, concerts and forced licorice intake in all shapes and colors. I am so thankful to have met you, and to have had each other through the ups and downs on personal matters and PhD life. I am happy to see you thrive in your new position with your own MS, "Sparky", measuring proteins as abundant as our polymer peaks, but definitely very challenging. Certainly, you are a great scientist and colleague, but most importantly you are great friend. No matter how big the distance is between both of us, you can always count with me for anything you need. I hope we can organize some trips together soon, including Ibiza and Singapore, since as you already know: I miss you a lot. Then, **Marie**, although we only met each other around my second year of PhD, we quickly grew into very good friends. I am very thankful that our paths crossed, since you became one of my best friends and I miss you enormously since you left NL. We shared so much over the last 3 years: climbing sessions (with major accidents), walks, trips, a *sometimes-worrying* plant addiction, pandemic parties (including 'balkoningsdag'), countless dinners and beers, and even a NYE celebration! You were always there to support me during rough periods, but also to share laughs and gossips. I am really excited to also have witnessed your professional growth over these 3 years, with your stunning data going into multiple high impact journals and your well-deserved new position as MS core facility manager. I'm sure we will eventually make the '*female proteomic core facility managers conference (FPCFMC)*' a reality, where I hope we can serve your excellent potato pizza for dinner. I am looking forward to more adventures together in the future.

I also want to thank my office mates from O609 over the last 4 years, who made it very fun to go to work and supported me so many times. **Kelly S.,** it was nice meeting you, and sharing nerdy conversations about biology. I am really excited for your familiar and professional growth over the last 4 years. You are an amazing scientist and no doubt you'll be a great PI, so keep it up girl! **Vojtech-ta,** I also really enjoyed meeting you, the (many) *borrels* and all the laughs around the labs. I am sure you know that your spider and I miss you a lot in the office. Best of luck with your two kids and the new work in Leiden. **Ziliang,** you came around my 3rd year, when I was the most stressed, still I hope I could teach you something useful. I have seen you progress a lot over these two years of PhD, becoming more confident and independent, so I'm sure you'll do great on these final years. I wish you all the best for the upcoming PhD and life adventures, and if you end up in Singapore, I'll see you when I come to visit. **Leticia,** you and your beautiful Spanish personality really made my last months in the lab super fun. Plus, you have a (semi) green thumb and took very good care of my plants. Special thanks for proofreading the thesis and spotting some typos! I wish you all the best on the future, I'm sure you'll shine girl. Finally, **Sem-y-sem (a.k.a. SpiderSem)** I'm glad to have met you, both as office mate and friend. I really enjoyed our many conversations, your advice on life/PhD stuff, as well as your help on coding. You are a truly clever person, and I learnt so much MS from you. Special thanks for proofreading my Introduction and giving constructive advice. Our climbing adventures are not over, I hope we can pick this up in Barcelona or anywhere else next year. All the best for the new job in Leiden and keep being the awesome, enthusiastic and wholesome person you always are.

During these 4 years, I worked very closely with some people, who taught me so many MS, EVs and organoid skills, keeping me inspired and enthusiastic. Some of the work we shaped together even made it into quite cool publications. First, from the BMSP group. **Gadi,** I really enjoyed working together, your perseverance and help were key to make my dream of crosslinking EVs a reality. Beyond and before this, we became good friends. I really appreciated the multiple dinners we shared (thanks for always cooking delicious and very elaborate things), your lovely *alfajores* that you regularly made for my birthday and the many conversations we shared over coffee or beer. I will miss you a lot and hope to see you, Tomer and a +1 soon. **Matti,** thanks for having the best hands to make grids for NS-TEM, you impressed many with your EV images! Also, you are an excellent structural scientist, and despite the bumpy road, I am sure you will become a successful PI. I wish you all the best for you and your family, which keeps getting bigger. **Tim,** the PRM master. Thanks for sharing all your knowledge with me, and for embarking in TimsTOF adventures together. I am super proud of our high-resolution *glucagon* spectrum, and I am convinced we will manage to measure this elusive peptide also on human intestinal biopsies. And if we don't, I had fun celebrating random peaks together before realizing they were not the peptide. Keep up your enthusiastic and joyful personality, and best of luck for finishing your PhD! **Franziska,** I learnt lots of DIA-MS from you, and really enjoyed our conversations, climbing sessions and beers over *borrels*. I hope you are doing great back in Switzerland and hope to see you again soon. **Riccardo Z.,** I really enjoyed working with you on our FAIMS project. Your enthusiasm, encouragement and positivism really made a difference, altogether with the delicious Italian espressos. I wish you all the best for your new UCL position. Second, **Marieke and Simon** from the UMCU. **Marieke,** it was nice collaborating on such a nice project, and learning a new MS technique to help you understand how EVs and cells interact. I think you did an outstanding amount of work for this paper and I'm sure it will get published in a great journal. All the best for the future and for your defense! **Simon,** I am really happy that you contacted me to do some proteomics on EVs, since I learnt a lot from you about EVs, their isolation, components, and the *epic* correlation between WB and MS data. I'm sure our results will make it into a good publication, even after the bitter ending. I am sure you'll make an outstanding PhD thesis. Third, **Jens and Joep** from the Hubrecht Institute. It was a pleasure to collaborate with you both, had lots of fun discussing about intestinal hormones. **Jens,** I really enjoyed our scientific conversations during the start of my PhD, I wish you all the best in your new PI position back in Germany. **Joep,** you are a brilliant collaborator to have, and it was a pleasure to work together on a second paper unraveling the proteolytic processing of gut hormones. Beyond this, I also had lots of fun discussing science, annoying reviewers, and other things with you. I am sure you will shine in Roche. Hope to see you around, or else it would be *unacceptable*!

A special mention needs to go to the amazing technicians of our group, who are responsible for keeping the labs up and running through thick and thin. **Mirjam**, you are a superhero. Not only you keep us all alive with your amazing coffee corner, but you also make sure we don't die on the labs either. I am so thankful to consider you a friend, and to have enjoyed many dinners, trips and parties together. Really appreciate the traditional Spaghetteria dinners with the mandatory Tiramisu for dessert.

I will miss you greatly and hope to have you visit Barcelona soon. **Ceri**, mistress of LCs, was really fun getting to know you and discovering a very Dutch but also very fun person. I'm very happy to see you thriving with Laura in your new position. You are an excellent technician, and we all miss you a lot in the lab. Hopefully we will manage to see MCR at some point, let's promise we'll let each other know next time they are in Europe. **Arjan**, you are really fun to be around. I enjoyed a lot the laughs, the heartbreakers and all the many things you taught me about MS. **Harm**, thanks a lot for the great conversations we shared over the years. I learnt a lot about LC-MS troubleshooting from you and was great to count with your advice when things got a bit rough. **Dominique, Nadia, Pieter,** was also a pleasure to meet you and see you all grow over these years. Best of luck in the future and stay in touch. Also taking great care of us, **Corine,** thanks for the immense help you gave me with both personal and work-related bureaucracy stuff. Would have been so tedious without your tips and tricks. Also, really enjoyed the sporadic conversations we had. **Geert,** we did not have many conversations, but it was great to always count with your help when computers decide to misbehave, specially during corona, thanks a lot! **Henk,** I really enjoyed our sporadic chats, about fun stuff but also about bioinformatics. It was always very inspiring to discuss how to visualize peptidomics data with you.

To the 'old-schoolers', PhDs, postdocs and PIs from our group – **Barbara,** I am so grateful to have met such an amazing woman, scientist, and friend. Thanks for introducing me to my favorite sport, for sharing so many dinners, trips, adventures, parties, beers and laughs together. You are one of a kind, Babsi, and that's why we all love you so much. I can't wait to see you again girl! **Jo-jo,** I am very thankful to have had you by my side during the PhD. You always knew how to lift my spirits when I was feeling insecure or upset about something. You are an outstanding PhD and scientist, and I'm sure you'll only shine brighter in Cali. I will miss greatly the fun times outside of the lab as well, but I'm looking forward to come visit you overseas to enjoy your delightful 'beer-pressure'. Keep it up, my friend, you are almost done! **Tomislav,** was fun receiving mean comments from you but also good pieces of advice, hope to cross paths some other time. **Charlotte,** thanks for teaching me lots on the MS team. I also had fun during dinners, beers, plant talk, and parties. Wish you and your family all the best, and keep up that awesome SCP work. **Bondt,** thanks for teaching me lots of MS on OR16, but also for giving me great advice over the years. Had lots of fun chatting and discussing science together and I wish you (and your family) the best in the future. **Dario,** was cool to share time together behind OR16, but also during our 'breaks' over the last years where I got to know you better. I wish you success with wrapping up the great PhD you did and hope to climb together at some point. **Baastian,** you are always fun to be around, specially at the bouldering hall. Best of luck wrapping up your PhD and success for the future, gorgeous! **Pascal,** I always enjoyed our conversations over beer, coffee, or in the MS wing. You are an exceptional scientist, as proven by your track of publications and achievements. I'm quite sure we will soon see you as a big PI in the field of structural chloroplast research. Hope to cross paths again, and best of luck in Grenoble, especially for little Luca. **Danique-y,** girlllll!!! I'm going to miss you soooo much, specially your bright and loving personality. I'm glad to have seen you become an amazing scientist and I'm sure you'll continue to shine after your PhD, since you are really clever and easy

going. Hopefully I'll hear a high-pitched "Juuuuuliaaa!" at some conference soon, meaning we can share some more laughs over a bierje. **Maurits,** it was nice getting to know you, wish you all the best for your company, as I'm sure you won't need much for finishing your excellent PhD! **Tatiana,** you impress me so much. Such a successful, strong, and independent woman. Plus, your sarcastic sense of humor is so charming. I hope we can see each other soon, to share some more laughs, conversations and nerdy MS talk over a G&T. Success in finishing your manager-PhD track, although you won't need it. **Wouter,** I'm very glad to have met you and I wish you all the best for the ending of you PhD and for your Tennis adventures. **Jeff,** I really enjoyed our very exciting and controversial conversations, both about society, life, and science. I'm sure you'd be proud to see my 100% female committee. Remember to wear a *red* t-shirt to *work*, that color looks great on you. **Nadine,** girl, I only know few people with such a nice soul. I am glad to have met you, and to have had you by my side during the ups and downs of life and PhD. Best of luck in your new position and please, be careful not to get injured! **Juanito,** you always spark kindness and positivism. I keep great memories over the last years, and wish you and your wife good luck on your new *northern* adventures. **Matina,** thanks for all your warmness and help at the start. Hope to dance some reggaeton together in the future. Good luck with "EVrything". **Kelly D., Jing, and Inge,** the milk peptidomics and proteomics experts. Thanks for very interesting discussions about peptidomics and life and wishing you all the best for the future. **Donna,** I am sorry I locked you out on the retreat. You are such an inspiration, in fashion and in life, always open-minded, understanding, and empowering. Thanks for that. I hope we can cross paths and party a bit more in the future. **Wei wei,** another fashionista of the lab. I really enjoyed our conversations, and seeing you grow as an independent and successful PhD. **Peppe and Yakub,** I learnt lots teaching you, and I'm very proud of what you both have achieved. **Joost,** et desitjo tot el millor per tu i per la teva familia, ha estat un plaer conèixer-te, riure i conversar sobre ciencia i vida. Ens veiem aviat si veniu per Barcelona! **Richard,** thanks for always willing to help, share and discuss science with a smile on. **Karli,** thanks for being an awesome PI that treats us all at equally. Best of luck for the coming years. **Niels, Fleur, Saar, Alba, Esther, Monique, Bas, Simone, Celia, and Maarten,** was great meeting you all and sharing scientific and more personal talks with you during *borrels* or in the lab. Wish you all success in your ongoing and new adventures.

To the 'newer generations' of PhDs, postdocs and PIs from our group. – For the PhDs, even if this journey is full of ups and downs, you'll eventually make it, so don't give up. And remember to celebrate all achievements and to maintain a healthy work-life balance. – **Sabrina,** you are really kind, hopefully we will cross paths again. Best of luck for the end of the PhD! **Soumya and Lai,** was awesome to share time and laughs with you. **Roos and Samiksha,** you girls are an awesome duo, glad you took over the smelly office that now smells like flowers and rainbows. Keep doing such an awesome job in you PhDs. **Maria and Cristina,** my Spanish fellows... I had lots of fun with both of you. Maria, sigue irradiando esa energía tan bonita que tienes y recuerda tomar descansos y disfrutar la vida. Cristina, gracias por empujarme a cantar y ser tan buena siempre con todos. Estoy segura de que tu PhD será excelente por la pasión que le pones a las cosas. **Anastasia, Sofia,** you showed me that Greek

people are exceptionally nice and fun to be around. I will mis you both greatly. **Yang and Fujia,** you are both really kind. I wish you great luck with your PhDs. **Kelly G.,** I must admit I still have a hard time understanding your Scottish accent, but it was a pleasure to try hard and get to know you more. I wish you, and your family, all the best back in Scotland and send you a big hug. **Jan, Andris, Joshua,** TeamsTOF, thanks for sharing all your knowledge with me, and for the good times over beers together. **Kat,** thanks for being so authentic, and sharing laughs and opinions over drinks. Also, thanks for showing me that Thermo's 'hair dryers' are not safe for human use. **Naomi,** glad to see you coming back to the group, and best of luck for the future. **Eduard, Victor and Marc,** had fun with you guys, and I wish you all the best. Lastly **Peter,** how sad I am that our paths only crossed briefly during this last year... Still, brief but awesome. I take a friend with me and I'm sure we will see each other soon. You are such a fun and clever person, and I have the feeling we would have made a great team working together. I hope we can keep doing our 'brainstorming coffee/beer sessions' remotely, or in Barcelona with Camille.

Outside of our group I also got to know a lot of amazing people that help me loads during these years. **Dounia,** thanks for all the fun nights, the mental support and for always believing in each other's science. You are an extraordinary scientist and a strong woman, with a bright future ahead of you. **Matthieu and Danielle,** thanks for all the neighbor talk and fun during corona times. Good luck with your new house! **Laura and Sjors,** was fun getting to know you and the dogs, best of luck guys. **Davide, Bruna and Ramona,** was amazing spending time with you, I take great memories with me and will miss you a lot. **Morganne and Valeria,** had lots of fun with you as well, I wish you all the best in for your future adventures. **Celine, Lisa, Ralph, Neander, Leo, Titouan, Esther, Ruben, Maxime and the rest of the climbing crew,** you guys are so kind, each of you individually. Thanks for spraying me with beta, with hugs and with your incredible personality. I hope we can see each other again in the future. **Alessandro,** te echo de menos tío, pero sé que estas muy feliz con Martina y deseo conocerla en el futuro. **Hester,** my coincidental and beloved Dutch friend. I enjoyed very much our adventures and looking forward to many more. **Maike,** thanks for being always so supportive, fun, and easy going. I hope we can see each other soon in Münster or Barcelona and wish you the best with your studies. **Astrid,** gracias por reeducarme el cuerpo después de cuatro años sentada, y por las risas, el apoyo y el 'stress release' levantando al ritmo de *la gasolina*.

Ahora a la gente que ha sido un enorme apoyo desde España y Cataluña. **Mamá y papá,** muchísimas gracias por todo. No estaría aquí sin vuestro apoyo incondicional ni sin la educación que me habéis dado. Me habéis enseñado a ser perseverante, a trabajar duro, pero también a vivir la vida y ser feliz. Y por eso, esta tesis está dedicada únicamente a vosotros, ya que, si no, siempre me faltarán las palabras para plasmar mi agradecimiento. **Chechu, Javier,** gracias por la magnífica portada que has diseñado. Sigue trabajando duro, explorando tus pasiones y siendo tan buena persona y no dudo que llegarás muy lejos en lo que decidas hacer en el futuro. Y nunca te olvides de disfrutar en el camino. **Pablito, Vir, Ricky,** gracias por todo vuestro apoyo y cariño, estoy muy orgullosa de todos vosotros y os deseo siempre lo mejor. **Titos,** gracias por siempre acordaros de mí y **Lola,** gracias por aconsejarme y enviarme detallitos estos cuatro años para que no se me olvide lo

rico que está el jamón. **Primis,** sois las mejores, que voy a decir. Gracias a todas por mucho apoyo estos años y con ganas de vernos pronto. **Natalia,** gracias por tu amistad incondicional los últimos 15 años, tú siempre sabes cómo hacerme sentir bien, y como decirme lo que no quiero escuchar. Gracias por estar a mi lado siempre, sabes que te quiero como a una hermana y que puedes contar conmigo para lo que necesites. **Canario, Dani,** sí, se me tu nombre. Gracias por la amistad incondicional desde 2012, la vida no sería tan divertida sin ti. Me impresiona verte crecer y convertirte en una persona aventurera e independiente, estoy segura de que el futuro solo te traerá cosas buenas porque no te mereces nada menos. Y con suerte, a mí me traerá un poco más de tiempo contigo. **Aitor,** mi chico, tú puedes con todo! Gracias por tus consejos y compañía estos años. **Ruben**, gràcies per tot el teu suport, pels riures, per tots els sudokus, Marius, etc. Ets un gran amic i estic tope d'orgullosa de tots els teus èxits. Moltes ganes de tenir-nos més a prop aquests anys vinents, i us desitjo el millor a tu i la Jud. **Marina,** gracias por todos tus consejos, por la amistad, las risas, y por siempre encontrar un momento estos cuatro años para vernos cuando vengo por Ibiza **Mery, Yaizi, Marta S, Marta R, Raquel, Pali,** gracias por toda la diversión y por todos los consejos que siempre me dais. **Judit,** gràcies per estar sempre al meu costat, encara que ens separin 60, 100 o 2000 km. Sé que podem contar l'una amb l'altra quan ho necessitem. Valoro molt la nostra amistat. I, sobretot, valoro que siguis la persona que ha vingut a veure'm al nord més cops, sempre pensant en portar-me records (fuet, colacao, etc.). **Cata,** que decir de ti mi chica, gracias por acompañarme siempre y por cumplir nuestra promesa de vernos una vez al año. Estoy super orgullosa de todo lo que has conseguido en la vida, y espero que sigamos apoyándonos, aunque se nos dé fatal mantener contacto frecuente. **Julia Costa,** tía lo feliz que me ha hecho reencontrarnos en Holanda. Muy agradecida por todos los momentos que hemos compartido estos últimos años, por tus consejos y tu mentalidad abierta. Os deseo lo mejor a ti y a Mathew. **Lía,** gracias por enseñarme espectrometría de masas y por siempre creer en mí. Sin ti, no estaría donde estoy hoy, y siempre que hablo de ti se me llena la boca de palabras bonitas y agradecimiento. **Carolina,** gracias por ser una gran amiga y persona y por siempre estar ahí cuando he necesitado hablar de la vida o del PhD. Me impresiona todo lo que eres capaz de hacer, sobre todo por los demás. Tengo muchas ganas de empezar esta nueva etapa juntas.

Lucas, mi pareja y compañero de vida los últimos 3 años. Gracias por tu apoyo incondicional, sin el cual habría sido aún más difícil acabar este PhD. Gracias por tus amables y acertados consejos, por creer en mí y por siempre recordarme lo que valgo. Eres una persona excepcional, buena, generosa, cariñosa y divertida, y estoy muy orgullosa de todo lo que has conseguido en la vida. Deseo seguir teniéndonos el uno al otro y continuar creciendo juntos muchos años más.

I also want to thank my **Reading and Examination Committee,** for putting the hard work of reading and evaluating this book on top of their probably busy schedules. Therefore, I am very grateful to all of you for your time, and for helping me to finish this journey.

I finally want to thank **all the women in academia** I have met so far. The path is rougher for us, according to data, but having such role models, including young fe-

male professors with impeccable tracks, makes me confident that I can also achieve this. Hopefully, in the coming decades, we will witness a shift towards a 50% of female representation in the higher rungs of the academic ladder.

A