



# Defining the reference proteomes for small extracellular vesicles and non-vesicular components

Received: 4 April 2025

Accepted: 13 January 2026

Published online: 13 February 2026

Julia Morales-Sanfrutos<sup>1,8</sup>, Joanes Etxeberria-Ugartemendia<sup>2,8</sup>,  
Orhi Barroso-Gomila<sup>1,8</sup>, Esperanza González<sup>3</sup>, María Sendino<sup>1,4</sup>,  
Pilar Ximénez-Embún<sup>1</sup>, Fernando García<sup>1</sup>, Eduardo Zarzuela<sup>1</sup>,  
Juan M. Falcón-Pérez<sup>1,3,5,6</sup>, Héctor Peinado<sup>1,7</sup> & Javier Muñoz<sup>1,2,5</sup>

Check for updates

Extracellular vesicles (EVs) are key mediators of intercellular communication and promising biomarkers. However, their molecular characterization remains challenging due to the heterogeneity of EV subtypes and co-isolated non-vesicular components. Here we leverage protein correlation profiling along density gradients to systematically analyse over 9,000 proteins in human cancer cell lines and biofluids, providing a rigorous reassessment of virtually all protein constituents associated with small EVs (sEVs) and non-vesicular entities. We show that sEVs primarily incorporate plasma membrane proteins via selective cargo-loading mechanisms, with low inclusion of intraluminal soluble proteins. By contrast, the abundant cytosolic proteins frequently detected in sEV preparations are not encapsulated within vesicles but are externally associated, probably originating from copurifying cellular debris and aggregates. Our work provides a reference resource for understanding the biogenesis, molecular determinants of cargo selection and functional roles of sEVs.

In the classical secretory pathway, signal peptide-containing proteins are translocated into the endoplasmic reticulum (ER), trafficked through the Golgi and released into the extracellular space. However, many proteins are secreted via unconventional protein secretion pathways<sup>1</sup>. Major unconventional protein secretion systems involve membranous vesicles originating from the plasma membrane and the endolysosomal system, collectively known as extracellular vesicles (EVs). Once internalized by recipient cells, EV cargo can exert important autocrine, paracrine and endocrine effects on physiology and

disease<sup>2,3</sup>. Moreover, because the cargo reflects the molecular features of the parental cell and remains stable in biofluids, EVs have emerged as promising sources of biomarkers.

However, current purification techniques yield heterogeneous samples that contain various vesicle types along with substantial amounts from non-vesicular (NV) structures<sup>4</sup>. This heterogeneity complicates the interpretation of the regulatory functions of different EV subtypes. Differential ultracentrifugation (dUC) is the most common strategy for EVs purification, separating large EVs (lEVs; 100–1,000 nm)

<sup>1</sup>Proteomics Unit, Spanish National Cancer Research Centre, Madrid, Spain. <sup>2</sup>Cell Signaling and Clinical Proteomics Group, Biobizkaia Health Research Institute, Barakaldo, Spain. <sup>3</sup>Exosomes Laboratory, Center for Cooperative Research in Biosciences, Basque Research and Technology Alliance, Dierio, Spain. <sup>4</sup>Proteomics Platform, Biobizkaia Health Research Institute, Barakaldo, Spain. <sup>5</sup>Ikerbasque, Basque Foundation for Science, Bilbao, Spain. <sup>6</sup>Biomedical Research Centre of Hepatic and Digestive Diseases, Carlos III Health Institute, Madrid, Spain. <sup>7</sup>Microenvironment and Metastasis Laboratory, Department of Immunology and Oncology, Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas, Madrid, Spain. <sup>8</sup>These authors contributed equally: Julia Morales-Sanfrutos, Joanes Etxeberria-Ugartemendia, Orhi Barroso-Gomila. e-mail: [javier.munozperalta@bio-bizkaia.eus](mailto:javier.munozperalta@bio-bizkaia.eus)

from small EVs (sEVs; <200 nm). Although many NV components copurify with sEVs during dUC, these populations can be further separated by density gradient (DG) centrifugation based on their distinct buoyant densities<sup>5</sup>. This combined dUC–DG strategy has proven essential for reassessing specific components of sEVs<sup>6</sup>. Nonetheless, the relatively low resolution of DG may introduce artifacts and misassignments, and a comprehensive, systematic and high-confidence analysis of the full spectrum of proteins in sEV and NV fractions is lacking.

To address this gap, we reconstructed complete elution profiles along the DG for over 9,000 proteins from multiple human cancer cell lines and biofluids. By applying a protein correlation profiling (PCP) strategy<sup>7,8</sup>, we achieved precise protein assignment to sEV and NV fractions. Our data confirm previous findings; challenge others; and reveal novel insights into the biogenesis, molecular determinants of cargo selection and functions of sEVs.

## Results

### DG–PCP for the analysis of sEVs

PCP relies on the principle that proteins within the same biological entity exhibit correlated quantitative profiles across biochemical fractions, such as DGs<sup>8</sup>. We applied this strategy to sEVs by collecting HeLa conditioned media, pelleting crude sEVs at 100,000g (hereafter, P100), separating them into 12 DG fractions<sup>6</sup> and analysing them by liquid chromatography–tandem mass spectrometry (LC–MS/MS) (Fig. 1a). Protein intensities were  $\log_2$  transformed, imputed and z-scored. After filtering low-quality data (see methods), DG profiles for 3,277 proteins were reconstructed. Finally, unsupervised hierarchical clustering was used to group proteins exhibiting similar profiles. As expected, classical sEVs markers such as CD9, CD63, CD81, syntenin-1 (SDCBP), ALIX (PDCD6IP) and TSG101 exhibited low-density profiles (Fig. 1a and Supplementary Fig. 1). Conversely, proteins previously associated with NV material<sup>6</sup> (for example, TUBA4A, HECTD3 and MCM2) appeared at high-densities.

Biophysical analyses confirmed the sEV identity of particles in low-density fractions. Refractometry showed buoyant densities consistent with sEVs (1.075–1.125 g ml<sup>-1</sup>) (Extended Data Fig. 1a,b). Nanoparticle tracking analysis (NTA) revealed a primary population of ~122 nm particles (Extended Data Fig. 1c) and cryo-electron microscopy (cryo-EM) visualized intact, bilayered structures with typical sEV morphology (Extended Data Fig. 1d). The DG–PCP strategy proved robust and reproducible. The separation of low- and high-density profiles was consistent across three technical replicates (Extended Data Fig. 2a) and insensitive to methodological variations, including alternative DG protocols<sup>9</sup> (Extended Data Fig. 2b), loading methods (Extended Data Fig. 2c) and buffers (Extended Data Fig. 2d). Notably, while variations in the exact fraction numbers for sEV and NV were observed between experiments, their relative separation remained evident. This highlights the fundamental strength of PCP: by analysing the entire profile shape rather than relying on specific fractions, it permits confident and reproducible classification.

### DG–PCP reveals and confidently classifies diverse protein profile

To enhance resolution, we increased DG fractions to 16 and employed a more sensitive data-independent acquisition (DIA) proteomics strategy (Fig. 1b and Supplementary Fig. 2a). To classify these high-resolution profiles, we adopted an unsupervised framework. Supervised PCP approaches are powerful but require high-quality training marker sets, which are not yet available for the full diversity of sEV and NV entities. Our unsupervised approach avoids these limitations by allowing the data itself to define its structure, with cluster annotation performed only *a posteriori*. Specifically, hierarchical trees were partitioned into 100–200 clusters based on distance thresholds. To ensure robustness, we only classified clusters with >20 proteins, serving as a filter to exclude low-confidence clusters, which we labelled as ‘unclassified’

(UNC). This strategy partitioned the data into well-defined and mutually exclusive classes. The sEV class (1,032 proteins) displayed a unimodal low-density profile containing canonical markers, while the remaining proteins were resolved into separate high-density classes. The largest group was the NV class (1,852 proteins), characterized by unimodal high-density profiles. We also resolved a distinct class of NV proteins, termed NV-dual, characterized by bimodal distributions peaking in both high- and low-densities (Fig. 1b). Among these NV-dual profiles, two subtypes could be distinguished based on the relative abundance across the low- and high-density regions: NV-dual 1 (213 proteins) showed a balanced distribution, whereas NV-dual 2 (625 proteins) was predominantly high-density. Finally, a separate class of 272 proteins with an intermediate density (termed MID) was resolved. As expected, all these profiles were less well defined when using only eight fractions (Supplementary Fig. 2b). Therefore, we opted for a 12-fraction approach for subsequent experiments (Fig. 1c), as the optimal compromise between throughput and profile reconstruction quality. This optimized DG–PCP workflow deconstructs the compositional complexity of crude sEV preparations, enabling confident classification of their constituent proteins.

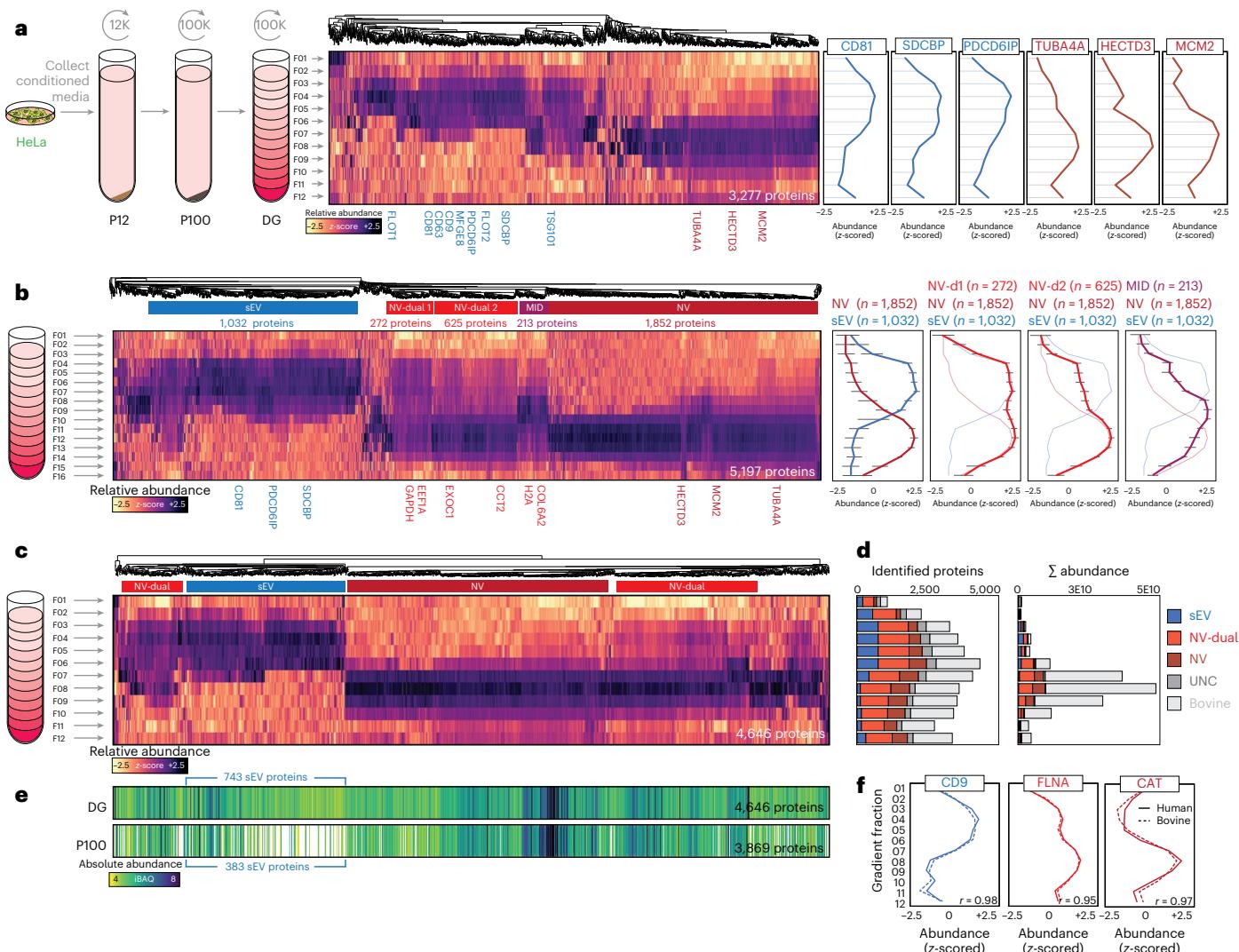
### DG–PCP resolves sEV proteins from cell culture serum contaminants

Common strategies often identify sEV-enriched DG fractions by western blot and subject only these fractions to LC–MS/MS. However, we found that 80% of proteins identified in our sEV-enriched fractions lacked genuine low-density profiles (Fig. 1c,d), indicating persistent NV contamination due to the limited resolution of DG. While a low- versus high-density fold-enrichment can be calculated to identify genuine sEVs proteins<sup>6</sup>, this requires replicates and might be confounded by bimodal distributions of NV-dual proteins, underscoring the role of PCP for determining true protein origin. Furthermore, DG separation effectively removed abundant high-density FBS contaminants (Fig. 1d) doubling the identification of human sEV proteins compared to the P100 (Fig. 1e). Despite using EV-depleted FBS, we still detected bovine sEV proteins at low-densities (Fig. 1f) and most homologous human–bovine proteins showed strong profile agreement. These results demonstrate the high sensitivity of our method to detect genuine sEV proteins, even in the presence of FBS contaminants.

### A systematic characterization of 15 cancer cell lines defines a core sEV signature

We applied DG–PCP to 15 human cancer cell lines from six cancer subtypes (Fig. 2a). NTA, cryo-EM and western blot confirmed that P100 pellets contained sEV-consistent particles enriched for classical markers and devoid of intracellular contaminants (Extended Data Figs. 3a,b and 4a,b). Subsequent DG separation and LC–MS/MS analysis identified 8,873 human proteins displaying consistent patterns across cell lines (Fig. 2a) and strong agreement between biological replicates (Supplementary Fig. 3). The 158,831 resulting profiles were categorized into sEV, NV-dual and NV classes with remaining proteins assigned as UNC (Supplementary Table 1). On average, 4,000 profiles were defined per cell line with notable variations (for example, SKMEL5 and HT29 yielded double proteins than of A549 and H460) (Fig. 2b). Protein identification counts correlated strongly with NTA particle concentration for sEV proteins ( $\rho = 0.736$ ) but not with other classes (Extended Data Fig. 3c), indicating that variability in protein counts was driven by vesicle yield and sampling depth rather than differences in protein diversity. Consistent with this, sample composition remained stable across cell lines, with mean  $\pm$  standard error of the mean (s.e.m.) proportions of 36.3%  $\pm$  2% (sEV), 30.2%  $\pm$  2% (NV-dual), 24.6%  $\pm$  3% (NV) and 12.4%  $\pm$  2% (UNC) (Fig. 2b).

Most importantly, intercell line agreement for sEV classification was exceptional, with 82% of proteins designated as sEV in one cell line retaining that same classification in all others (Fig. 2c).



**Fig. 1 | DG-PCP for the assignment of sEVs and NV proteins.** **a**, Crude sEVs were purified from conditioned cell media by dUC, separated by DG into 12 fractions and subsequently analysed by LC-MS/MS. The resulting DG profiles for 3,277 proteins (z-score) are shown as a heat map. Representative sEV markers (blue) and NV-associated proteins (brown) are displayed. **b**, Crude sEVs were separated by DG into 16 fractions. Unsupervised clustering revealed the presence of different protein profiles (sEV, NV-dual1, NV-dual2, NV and MID). Right: protein profiles for each cluster. Data ( $n$  = number of proteins in cluster) are presented as median values, with error bars representing the interquartile range. **c**, Optimized DG separation coupled with PCP strategy. sEVs were separated by DG into

12 fractions and analysed by LC-MS/MS using DIA. Unsupervised clustering was used to group protein profiles into sEV, NV-dual and NV. **d**, Number of human proteins categorized as sEV, NV-dual, NV and UNC as well as bovine contaminants detected in each fraction alongside their cumulative abundances. **e**, Absolute abundances of proteins (iBAQ) identified in both DG and P100 are visualized below the heat map, with proteins not detected in the P100 depicted in white. **f**, Elution profiles of bovine orthologs for representative sEV, NV-dual and NV human proteins are shown, along with their corresponding Pearson correlation values ( $r$ ).

Cross-validation with published data<sup>6</sup> confirmed the accuracy of our classification (Supplementary Fig. 4). We defined a core signature of 1,499 high-confidence sEV proteins, requiring classification as sEV in >80% of detected instances across at least four cell lines. Although the possibility remains that a protein could be a genuine sEV component in one cell line and NV in another, the high stability of our sEV classification and the rarity with which NV proteins were reclassified as sEVs, suggests this is uncommon.

By contrast, the classification of NV proteins was more dynamic, frequently shifting between NV and NV-dual profiles (Fig. 2c and Supplementary Fig. 5). Although 66% of NV-dual proteins consistently maintained their classification across all cell lines, 18% were reclassified as NV in at least one other instance. The inverse trend was also clear: while 55% of NV proteins consistently maintained their classification, a substantial 37% were reclassified as NV-dual in other cell lines. This suggests NV and NV-dual profiles represent a dynamic continuum

in which a protein's tendency to adopt a purely high-density (NV) or a bimodal (NV-dual) profile could be a context-dependent characteristic, rather than methodological inconsistency. Supporting this, 84–87% of the top 200 NV proteins in ref. 6 were classified as either NV or NV-dual in our analysis, confirming the NV nature of the NV-dual class (Supplementary Fig. 4).

### Intracellular protein and mRNA levels reveal sEVs loading mechanisms

We compared our extracellular data (EXO) with published intracellular protein (PRO) and mRNA (RNA) levels from matched cell lines<sup>10,11</sup>. This analysis revealed distinct patterns (Fig. 2d). (1) sEV proteins exhibited low PRO levels but average RNA levels, suggesting depletion of intracellular reservoirs via secretion. (2) NV-dual proteins, by contrast, exhibited elevated EXO, PRO and RNA levels, indicating origin from highly expressed genes. (3) NV proteins displayed low PRO and

RNA levels, indicative of low expressed genes. (4) Finally, UNC proteins also exhibited low abundance, consistent with the noisy elution profiles that prevented their classification. Correlating EXO and PRO levels for 6,188 proteins revealed a positive correlation ( $\rho = 0.467$ ) (Fig. 2e). However, sEV proteins showed significantly lower correlation ( $\rho = 0.392$ ), suggesting selective cargo loading whereas NV-dual proteins correlated strongly with the intracellular levels ( $\rho = 0.783$ ), consistent with non-selective bulk release. Notably, for sEV proteins, RNA predicted EXO abundance better than PRO levels ( $\rho = 0.437$  versus 0.392) (Fig. 2e), whereas NV-dual proteins correlated more strongly with PRO than RNA ( $\rho = 0.783$  versus 0.626), confirming that intracellular protein abundance drives their extracellular presence.

### A call for a re-evaluation of sEV proteins annotations in public repositories

The 1,499 sEVs proteins spanned five orders of magnitude in abundance. Remarkably, just six proteins (ubiquitin, CD9, SDCBP, HSPA8, CD63 and RAP1B) accounted for 25% of the sEV protein mass, with this number rising to 34 proteins for 50% of the mass (Fig. 2f). Public repositories such as Exocarta<sup>12</sup> curate lists of proteins frequently identified in sEVs. However, we found that only 51 of Exocarta's top 100 EV proteins exhibited genuine sEVs profiles (Extended Data Fig. 5) as other markers such as ACTG1, LGALS3BP and EEF1A1 were reclassified in our data as NV-dual or NV profiles. Gene Ontology (GO) analysis confirmed that sEVs proteins were enriched in 'extracellular exosome' terms ( $P_{adj} = 9.4 \times 10^{-159}$ ) (Fig. 2g and Supplementary Table 2). Surprisingly, the same terms were also enriched in NV-dual proteins ( $P_{adj} = 1.19 \times 10^{-73}$ ). These findings challenge the accuracy of current sEV protein annotations and underscore the critical need for rigorous re-evaluation. With this in mind, we propose our top 100 most abundant sEV and non-sEV proteins as refined positive and negative marker panels for assessing sEV purity (Fig. 2h).

### Subcellular localization determines composition of sEVs, NV-dual and NV proteins

Interestingly, GO 'cellular component' terms displayed far greater statistical significance than 'biological process' or 'molecular function' terms (Extended Data Fig. 6a). Unsupervised clustering showed a nearly perfect separation between 'cellular component' terms enriched in sEV, NV-dual and NV proteins in all 15 cell lines, with NV-dual and NV enriched terms showing higher agreement (Extended Data Fig. 6b). Proteins categorized as sEVs showed an extreme enrichment in 'plasma membrane' ( $P_{adj} = 4.9 \times 10^{-233}$ ) and related terms (Fig. 2g), while NV-dual and NV proteins lacked such enrichment. Instead, NV-dual and NV were enriched in 'cytosol' ( $P_{adj} = 1.1 \times 10^{-196}$ ) and 'intracellular organelle lumen' ( $P_{adj} = 3.5 \times 10^{-67}$ ) respectively. These mutually exclusive

functional enrichments strongly support our classification system and underscore the critical role of subcellular localization as a major molecular determinant leading to sEV, NV-dual and NV protein profiles.

### A subclass of ER-derived particles with distinct buoyancy

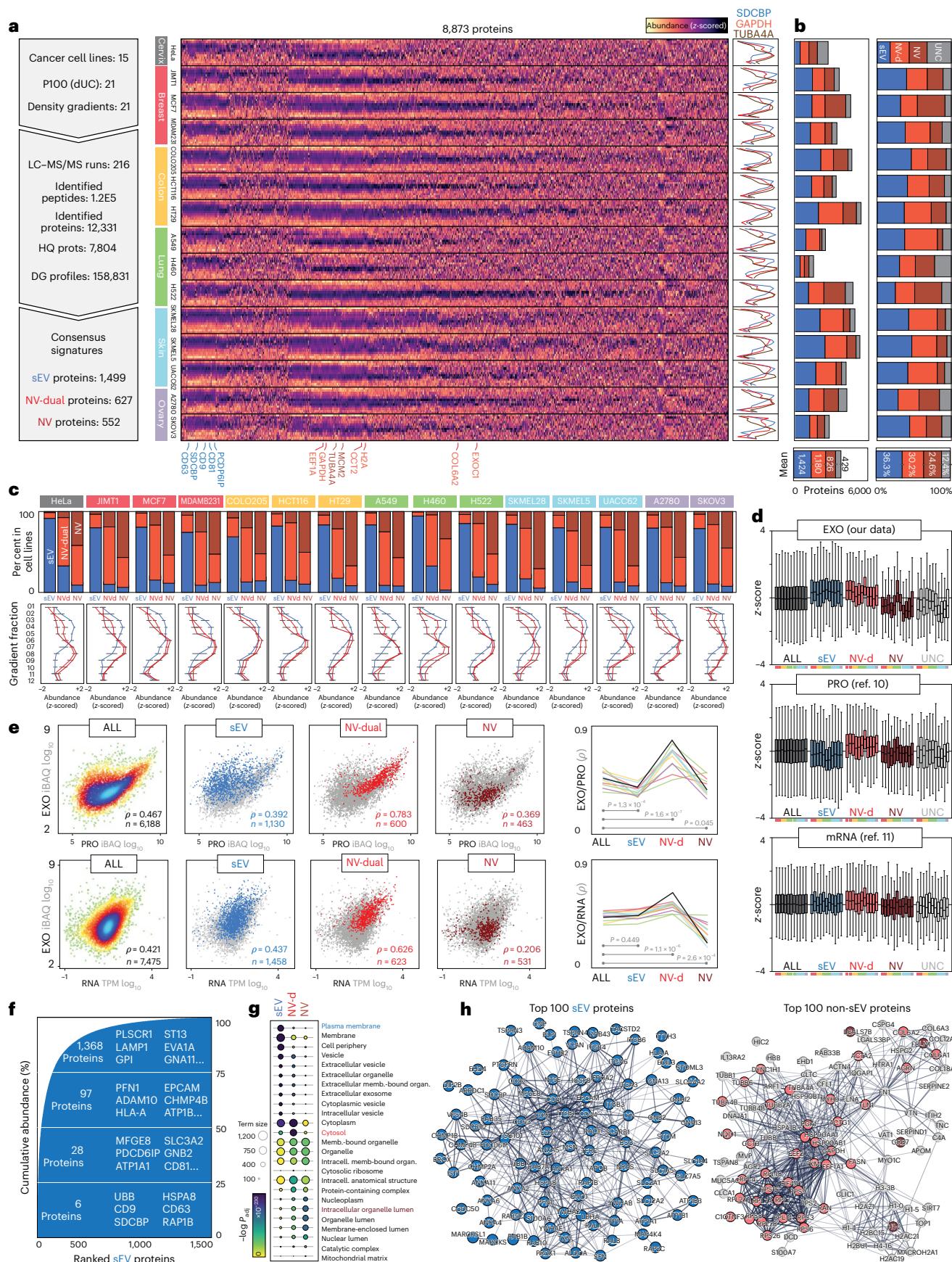
We applied rank-based GO enrichment to proteins in each DG fraction<sup>13</sup>. Visualization of GO term significance along the gradient revealed profiles mirroring protein elution patterns (Fig. 3a). While 'extracellular exosome' was ubiquitously enriched across all fractions, confirming limited utility as purity metric, 'plasma membrane' and 'intracellular organelle lumen' peaked at low and high densities, respectively, and 'cytosol' aligned NV-dual proteins. Notably, 'ER' terms were enriched at slightly higher densities (F04–05) than 'plasma membrane' (F03–04) (Fig. 3a), a pattern consistent across 15 cell lines (Extended Data Fig. 7). To investigate this in detail, we re-examined the profiles of ER-resident proteins<sup>14</sup>: DDOST and TMED10 (containing KKXX 'ER membrane' retrieval signals) and chaperones HSPA5 and HSP90B1 (containing KDEL motifs for 'ER lumen' retention). All four ER markers eluted at higher densities than sEV protein SDCBP but lower than NV protein ADH1C (Fig. 3b). Because our unsupervised clustering did not recognize their distinct densities, we retrieved from our data the best correlating proteins with these ER markers (Methods). This approach identified 32, 32, 14 and 22 proteins correlating with DDOST, TMED10, HSPA5 and HSP90B1 respectively (Fig. 3c and Supplementary Table 3). Overlap was significant between DDOST/TMED10 (20 proteins; hypergeometric test  $P = 1.7 \times 10^{-43}$ ) and HSPA5/HSP90B1 (11 proteins; hypergeometric test  $P = 1.6 \times 10^{-43}$ ), with GO analysis confirming respective 'ER membrane' and 'ER lumen' enrichments (Fig. 3c). Among the 'ER membrane' correlating proteins, we found structural ER components (RTN3 and RTN4), Translocon subunits (RPN1, RPN2 and MLEC) and ER homoeostasis regulators (CANX and HM13) (Fig. 3b). The 'ER lumen' proteins included ERP44, involved in folding, and the RNA binding protein YBX1. Notably, these ER-related proteins represented <5% of the total protein mass within sEVs (Fig. 3d). These findings highlight the capacity of DG–PCP to resolve a substoichiometric, ER-enriched protein signature with a distinct buoyancy, raising the possibility of a previously unappreciated class of ER-derived entities.

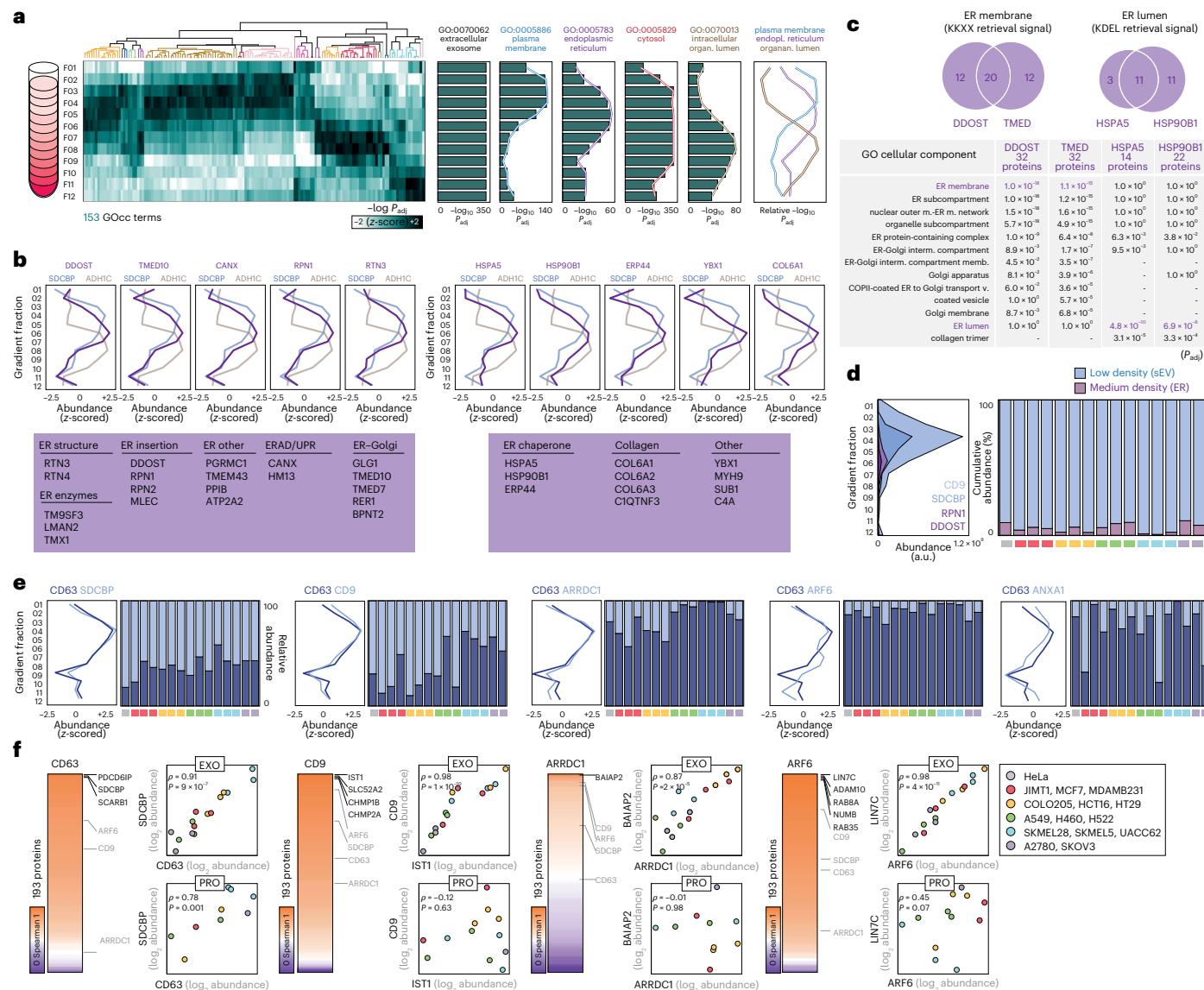
### Limitations in the characterization of sEV subtypes by DG separation

The similar physicochemical properties of sEVs subtypes pose challenges for their purification. Leveraging the resolution of our DG–PCP data, we sought to determine whether sEV subtypes could be distinguished. Exosomes, sEVs of endosomal origin, are marked by CD63 and SDCBP<sup>15</sup>. Indeed, these two exosomal proteins demonstrated

**Fig. 2 | Systematic analysis of sEV, NV-dual and NV proteins in 15 human cancer cell lines.** **a**, Crude sEVs from 15 human cancer cell lines were analysed by DG–PCP, and the resulting elution profiles for all 8,873 identified proteins are shown in the heat map. The summary metrics for all datasets (including replicates) (left) and representative elution profiles of SDCBP (sEV), GAPDH (NV-dual) and TUBA4A (NV) across all 15 cell lines (right) are shown. **b**, Protein counts and relative proportions of proteins classified as sEV, NV-dual, NV or UNC in each cell line. The mean protein counts and relative proportions (percentage of total identifications), averaged across  $n = 15$  cell lines, are indicated at the bottom with error bars representing the s.e.m. **c**, Top: for proteins classified as sEV, NV-dual or NV in one cell line, the proportion of proteins classified as sEV, NV-dual and NV in the remaining cell lines is shown as a bar chart. Bottom: the centroid profiles of sEV, NV-dual and NC proteins identified in each cell line are shown. The data ( $n$  = number of proteins in cluster) are presented as median values, with error bars representing the interquartile range. **d**, Box plots showing the abundance levels ( $z$ -score) of sEV, NV-dual, NV and UNC proteins identified in our data (EXO) compared to their intracellular protein levels (PRO)<sup>10</sup> and mRNA expression levels (RNA)<sup>11</sup> across  $n = 15$  independent cell lines. The sample size ( $n$ ) for each box plot represents the number of distinct proteins or transcripts quantified in that specific category and cell line. Box plot elements: centre lines indicate medians;

box limits represent the 25th and 75th percentiles; whiskers extend to 1.5× the interquartile range. **e**, Correlations between our data (EXO) and PRO<sup>10</sup> (top) or RNA<sup>11</sup> (bottom) data. Average values of all cell lines are used. Proteins classified as sEV, NV-dual and NV are colour-coded. Spearman correlation coefficients ( $\rho$ ) for each cell line are displayed as line charts on the right. A statistical assessment was done using a two-sided paired Student's  $t$ -test. **f**, The 1,499 proteins consistently classified as sEVs were ranked by abundance. The number of proteins constituting each quartile of fractional mass is indicated with examples of specific proteins. **g**, GO cellular component terms enriched in sEV (1,499 proteins), NV-dual (627 proteins) and NV (552 proteins) protein classes. Gene-set over-representation was assessed with g:Profiler using a one-sided cumulative hypergeometric test (Fisher's exact test) for each term.  $P$  values were adjusted for multiple testing using the g:SCS procedure (set counts and sizes) implemented in g:Profiler. **h**, Protein interaction networks constructed via the STRING database using the top 100 most abundant sEV proteins (left) and the top 100 most abundant non-sEV proteins (right). In the non-sEV network, proteins belonging to the consensus NV-dual and NV signatures are coloured red and brown, respectively. Proteins shown in grey represent abundant non-sEV components that did not meet the consensus criteria for either category. HQ, high quality; TPM, transcripts per million.

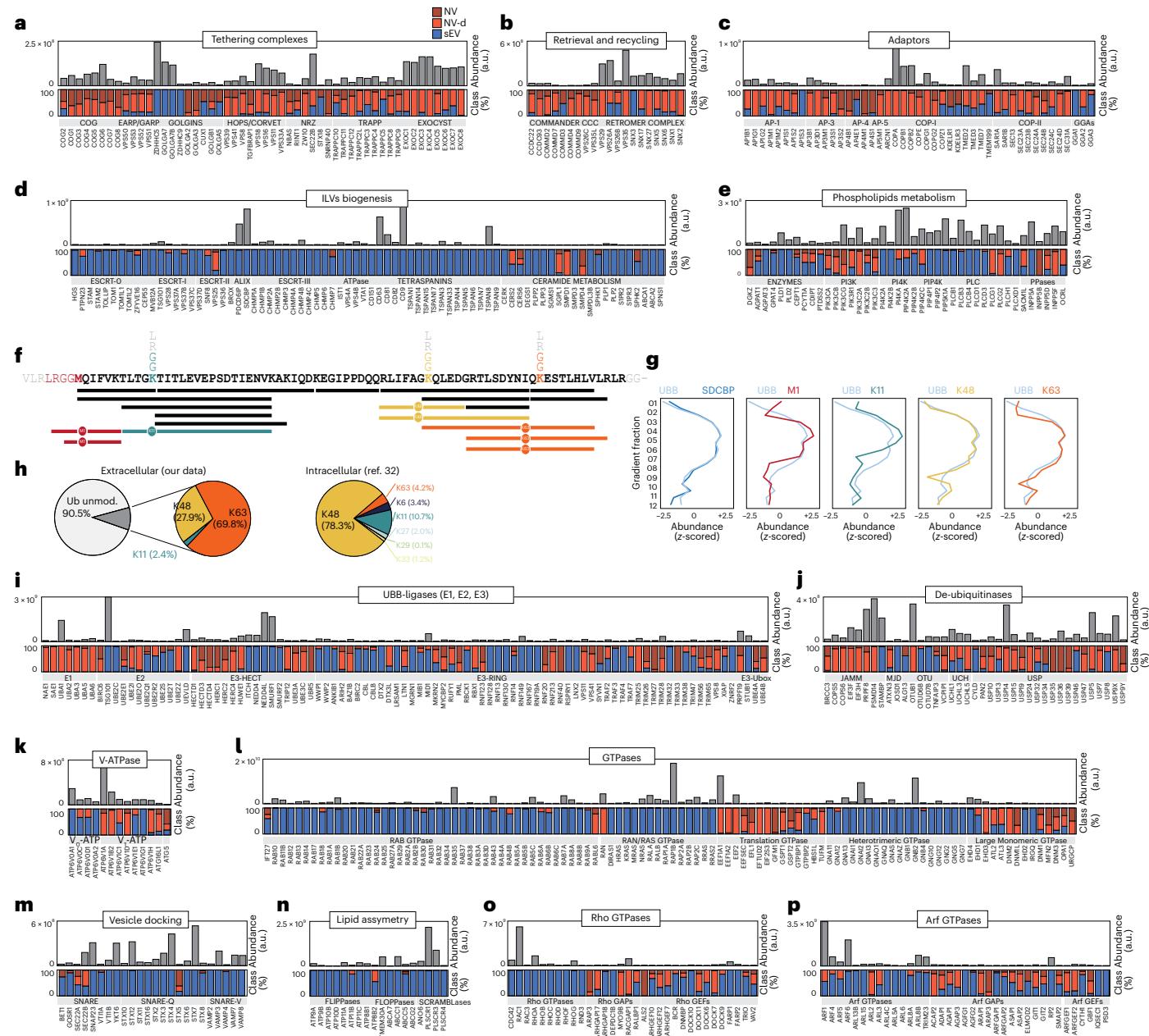




**Fig. 3 | Analysis of sEVs heterogeneity using DG-PCP.** **a**, Proteins identified in each DG fraction were ranked by descending abundance and analysed by GO enrichment. For ranked gene lists ('ordered query' in g:Profiler), pathway enrichment was assessed using a one-sided minimum hypergeometric test as implemented in g:GOST. The *P* values were adjusted for multiple testing using the g:SCS procedure (set counts and sizes). The significance ( $-\log P_{\text{adj}}$ ) of GO cellular components terms along the DG was *z*-scored and visualized as a heat map. Right: unsupervised clustering grouped GO terms with similar profiles with representative examples shown. **b**, Representative elution profiles of ER membrane proteins (left) and ER lumen proteins (right) identified at intermediate densities. For comparison, elution profiles of SDCBP (sEV) and ADH1C (NV) are provided. Additional proteins identified at intermediate densities were functionally classified and are shown below within boxes. **c**, Overlap between proteins correlating with DDOST and TMED10 (ER membrane markers) and with HSP5 and HSP90B1 (ER lumen markers). GO cellular components terms enriched in each subset are shown below along with their adjusted *P* values. Gene-set over-representation was assessed with g:Profiler

near-perfect co-elution (Fig. 3e), with their relative abundances being maintained across cell lines. Conversely, sEVs formed at the plasma membrane (that is, small ectosomes) appear to be marked by CD9<sup>15</sup>. Although CD9 displayed strong co-elution with CD63 (Fig. 3e), their relative abundances varied across cell lines, suggesting that the balance between endosomal and ectosomal secretion might be

cell-type dependent<sup>15</sup>. ARRDC1, which is present in plasma membrane ARRDC1-mediated microvesicles<sup>16</sup> paralleled the CD63 pattern but with less than 10% of its intensity, indicating substoichiometric presence. Because of their wide size range (100–1,000 nm), classical microvesicles (that is, large EVs) can also be copurified in sEV preparations. ARF6 is a critical regulator of the actin cytoskeleton contraction



**Fig. 4 | A proteome map of proteins and pathways involved in the biogenesis of sEVs. a–e**, Proteins are grouped based on relevant functional categories as discussed in the main text. Bottom: for each protein, the abundance as sEV (blue), NV-dual (red) and NV (brown) classes in all cell lines were summed and their relative percentages are displayed. Top: the corresponding absolute abundance is shown. **f**, Unmodified ubiquitin (UBB) peptides are shown in black, and UBB peptides containing the di-Gly remnant (indicative of Ub chains) are

colour-coded according to the modified K residues. **g**, Left: DG profiles of total UBB and SDCBP as a reference. The remaining plots show the DG profiles of ubiquitin chains at the indicated positions (M1, K11, K48 and K63) in comparison to total UBB. **h**, Relative levels of unmodified UBB and Ub chains identified in our data in sEVs. Right: for comparison, the relative intracellular levels of Ub chains reported by ref. 32 are shown. **i–p**, Additional functional categories are displayed as in a–e.

process of microvesicles<sup>17</sup> and also regulates CD63-positive endosomal vesicles<sup>18</sup>. Consistent with this, ARF6 and CD63 displayed similar patterns. The recently proposed microvesicle marker ANXA1<sup>6</sup> displayed, however, a slightly lower density than CD63, consistent with previous findings<sup>6</sup>. Together, these results highlight that current DG methods alone are insufficient to fully resolve different sEV subtypes due to overlapping buoyancies.

Given this limitation, we explored an approach based on the principle of 'guilt by association,' commonly used in gene co-expression analyses<sup>19</sup>. We hypothesized that proteins with conserved abundance levels across cell lines could provide compositional insights into sEV subtypes. To test this, we calculated Spearman's rank coefficients

between the above sEV subtypes markers and all other proteins in our datasets (Supplementary Table 4). Although proteins such as TSG101 exhibited similar co-expression across all four markers, supporting their involvement in various sEV biogenesis mechanisms, we did identify proteins exhibiting specific patterns. Exosomal CD63 showed best co-expression with SDCBP (syntenin-1) and PDCD6IP (ALIX) (Fig. 3f), two key components in the biogenesis of exosomes<sup>20</sup>. Conversely, the co-expression of CD63 with ectosomal markers ARF6, CD9 and, especially, ARRDC1 was significantly lower (Fig. 3f). Instead, CD9 co-expressed with ESCRT-III components (IST1, VPS4B, VTA1 and BROX). ARRDC1 co-expressed with BAIAP2, implicated in plasma membrane-derived sEVs. ARF6 showed co-expression with

GDP-binding proteins and the LIN7C–CASK complex, involved in exocytosis. Importantly, repeating this analysis with the intracellular protein levels (PRO) yielded no significant co-expressions (Fig. 3f), supporting the existence of distinct molecular signatures for sEV subtypes.

### A resource for studying sEVs biogenesis

Current mechanistic insights into sEV biogenesis are confounded by the difficulty in distinguishing authentic components from contaminants. Our high-confidence dataset facilitates a rigorous re-evaluation of established models and the identification of novel regulators.

Clathrin- and caveolae-mediated endocytosis supply proteins to endosomes using adaptors<sup>21</sup>. Although the AP-2 complex showed partial association to sEVs, other adaptors (SH3BP4), scission proteins (SH3GL1) and caveolae carriers (CAV1) appeared more enriched (Extended Data Fig. 8a). Once in endosomes, proteins can be sorted into intraluminal vesicles (ILVs) before degradation or be retrieved and recycled to the plasma membrane or Golgi<sup>22</sup>. Although retrieval complexes (that is, retromer, retriever and CCC complex), cargo adaptors (that is, AP-1, AP-3, AP-4 and GGAs) and tethering complexes (that is, COG, GARP and EARP) were not enriched in sEVs (Fig. 4a–c), proteins that are recycled back to the plasma membrane<sup>23,24</sup> or Golgi<sup>25</sup> were (Extended Data Fig. 8b,c). Hence, although recycling and degradative sub-domains are spatially segregated<sup>22</sup>, our findings show that their cargo is secreted as sEVs.

A subpopulation of late endosomes escapes degradation and fuses instead with the plasma membrane to release ILVs as exosomes. The biogenesis of these ILVs is assisted by ESCRT complexes<sup>26</sup>. Compared with the intracellular proteome, the stoichiometry of ESCRT complexes was well-conserved except MVB12A, CHMP2A, TOLLIP and IST1, which were over-represented in sEVs (Extended Data Fig. 9). Exosomal ILVs use a specific pathway with ALIX (PDCD6IP) and syntenin-1 (SDCBP)<sup>20</sup>, both of which were among the most abundant proteins in sEVs (Fig. 4d). Exosomes can also be formed using ESCRT-independent mechanisms by tetraspanins<sup>27,28</sup> and by a ceramide-dependent process<sup>29</sup>. However, neutral type II sphingomyelinase SMPD3 and phospholipid proteins PLP1/PLP2 were low abundant in sEVs (Fig. 4d). Further metabolism of ceramide into sphingosine 1-phosphate (S1P) is also involved in ILVs<sup>30</sup>, but S1P receptors were also low abundant (Fig. 4d). Nevertheless, the strong association of phospholipid metabolism proteins (Fig. 4e) and phospholipid-binding domains (Extended Data Fig. 8d) underscores the critical role of phospholipids in sEVs.

ESCRTs sort ubiquitinated cargo into ILVs<sup>31</sup>, and indeed, ubiquitin was the most abundant sEV protein (Fig. 2f). Although 90.5% of ubiquitin was unmodified, we identified di-Gly remnants within ubiquitin itself at M1, K11, K48 and K63 (Fig. 4f) which co-eluted with total ubiquitin (Fig. 4g), demonstrating the presence of branched chains in sEVs. K63-linked chains comprised 69.8% in sEVs, contrasting sharply with the ~4% found intracellularly<sup>32</sup> (Fig. 4h). This aligns with the enrichment of K63-binding proteins<sup>33</sup> (Extended Data Fig. 8e) and

HECT-domain E3 ligases SMURF1 and NEDD4L (Fig. 4i), which mediate K63-ubiquitination<sup>34</sup>. Finally, the presence in sEVs of de-ubiquitinases STAMBP (AMSH) and USP8 (Fig. 4j) supports the model that cargo is de-ubiquitinated before ILV encapsulation<sup>34</sup>.

Progressive acidification is required for the lysosomal degradation of endosomes<sup>35</sup>. However, the sequestration of V-ATPase subunits ATP6V1E1<sup>36</sup> and ATP6VOA1<sup>37</sup> into sEVs reduces late endosomes acidification and thereby enhances their exocytosis. We confirmed both subunits in sEVs, with ATP6VOA1 notably more abundant (Fig. 4k). RAB proteins, which regulate the degradative-to-secretory transition<sup>38</sup>, were nearly all incorporated into sEVs (Fig. 4l), particularly RAB7, RALA/B, RAB11B, RAB35 and its GAPs TBC1D10s (Extended Data Fig. 8f). We also identified novel sEV-associated GTPases such as RAB43, RAB5C and RAP1A. The final steps of membrane fusion are mediated by SNARE proteins (Fig. 4m). Among these, STX4, SNAP23 and VAMP7, key mediators of late endosomes–plasma membrane fusion<sup>39</sup>, were some of the most abundant docking proteins in sEVs.

ARRMs, small ectosomes and microvesicles form via plasma membrane blebbing. Although their origin from lipid rafts is debated<sup>40</sup>, we found several lipid raft-associated proteins in sEVs (Extended Data Fig. 8g). The reorganization of lipid species has thus far been uniquely described for EVs that bud at the plasma membrane<sup>2</sup>. Consistent with this, we identified scramblases, flippases and flopases in sEVs (Fig. 4n). These enzymes induce membrane bending and restructuring of the underlying actin cytoskeleton network. Key regulators of this process, including ARF6, PLD, CDC42, CFL1, RHOA and RAC1, were also found enriched in sEVs (Fig. 4o–p and Extended Data Fig. 8h). However, many of these proteins have roles in the biogenesis of both endosomal and ectosomal EVs, highlighting the overlapping mechanisms involved in their formation.

### sEVs are primarily composed of plasma membrane proteins and contain few soluble cytosolic proteins in their lumen

Among the 1,499 sEV proteins, 964 proteins were annotated as ‘plasma membrane’ ( $P_{adj} = 4.97 \times 10^{-233}$ ) and 1,229 proteins as ‘membrane’ ( $P_{adj} = 1.70 \times 10^{-220}$ ) (Fig. 2g). Notably, ‘plasma membrane’ proteins comprised 80% of the fractional mass of the sEVs proteome, a significant enrichment compared with intracellular levels (Fig. 5a). Among the ‘plasma membrane’ proteins in sEVs, single-pass type I proteins maintained a steady distribution of 25%, mirroring the intracellular proteome, whereas type II, III and IV proteins exhibited marked depletion in sEVs relative to the total cell lysate (Fig. 5b). Instead, sEVs exhibited a clear enrichment in multipass (41%) and lipid-anchored proteins (14%). Of note, NV-dual (67%) and NV (50%) were clearly enriched in peripheral membrane-associated proteins.

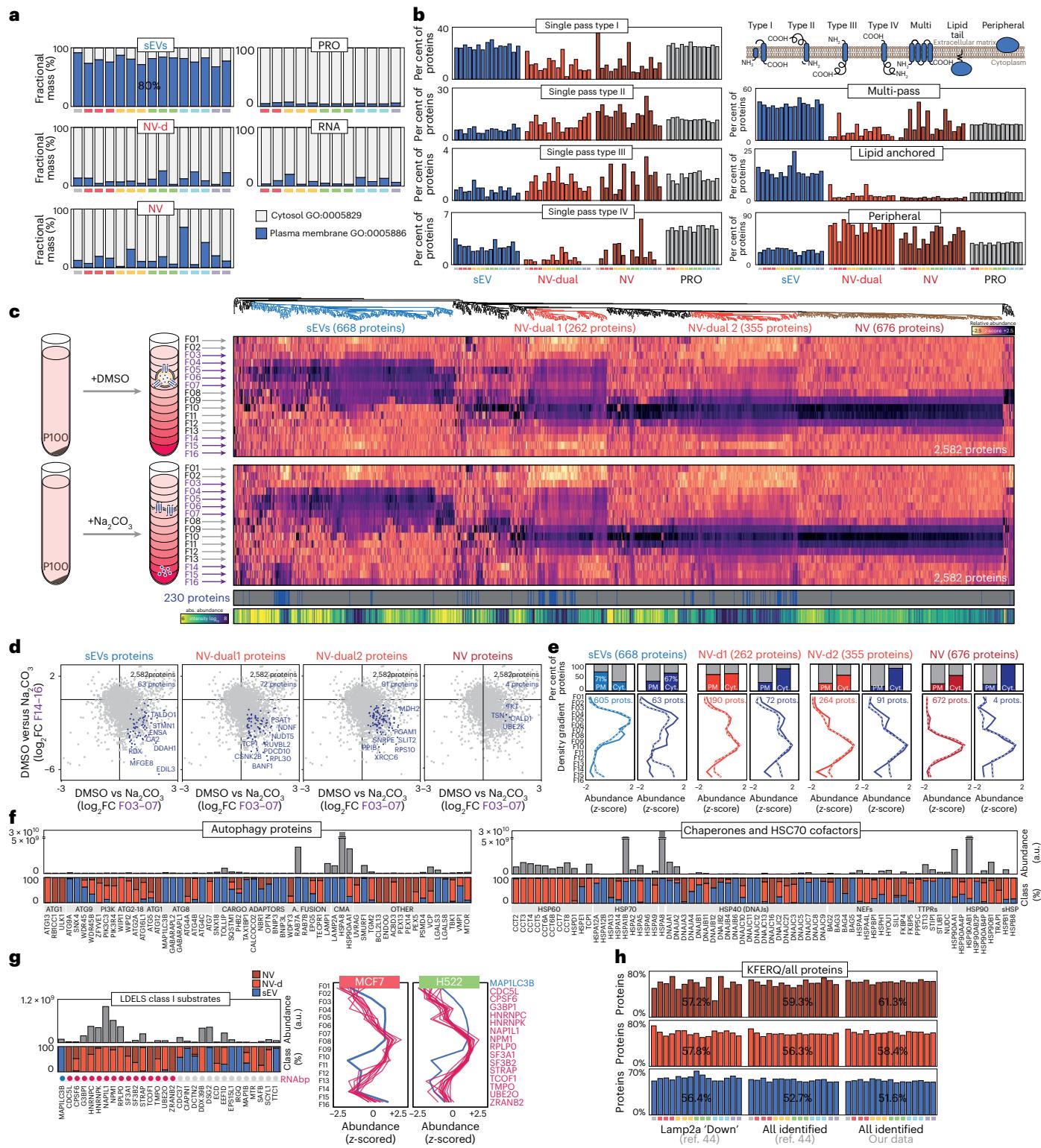
While these findings indicate that sEVs are mainly composed of plasma membrane proteins, it is well-established that soluble cytosolic proteins are also encapsulated within their lumen. To identify bona fide cytosolic proteins in sEVs, we treated crude sEVs with  $\text{Na}_2\text{CO}_3$ ,

**Fig. 5 | Plasma membrane and cytosolic protein composition in sEVs revealed by DG-PCP.** **a**, Fractional mass (sum of intensities) of proteins annotated by GO as plasma membrane and cytosol among the proteins classified as sEV, NV-dual and NV in each cell line. For comparison, fractional masses derived from intracellular protein (PRO)<sup>10</sup> mRNA (RNA)<sup>11</sup> levels data are included. **b**, Relative abundance of plasma membrane protein subclasses (as annotated in UniProt) among sEV, NV-dual and NV fractions, alongside intracellular protein (PRO) levels. **c**, Crude sEVs (P100) were treated with  $\text{Na}_2\text{CO}_3$  or DMSO and separated by DG. Elution profiles of identified proteins are visualized as a heat map. Statistical analysis identified 230 proteins increasing at high-density fractions (one-sided paired Student’s *t*-test, FDR < 0.2) while decreasing from low-density fractions (one-sided paired Student’s *t*-test, FDR < 0.2) in response to  $\text{Na}_2\text{CO}_3$ , which are highlighted in blue below the heat map. Absolute abundances (iBAQ) are displayed below. **d**, Scatterplots depict  $\log_2$ (fold changes (FC)) (DMSO versus  $\text{Na}_2\text{CO}_3$ ) in high- (F14–16) and low-density (F03–07) fractions, with the identified

significant proteins in each protein cluster highlighted in blue. **e**, Average of DG profiles for selected clusters in DMSO (solid line) and  $\text{Na}_2\text{CO}_3$  (dashed line) is shown, with proportions of plasma membrane (PM) and cytosolic (Cyt.) proteins indicated above as bar charts. **f**, The proteins involved in autophagy and chaperone networks, with their relative proportions (as sEV, NV-dual and NV classes) (bottom) and absolute abundances (top) are displayed. **g**, Proteins identified in ref. 43 as class I substrates of LDELS. Their classification as sEV, NV-dual and NV in our data is shown. Right: the DG profiles of LDELS substrates annotated as RNAbp are compared with that of LC3 (MAP1LC3B) in MCF7 and H522 cell lines as representative cell lines. **h**, Classification of proteins downregulated in LAMP2A KO cells from ref. 44 according to our DG-PCP data. The number of proteins bearing putative KFERQ motifs<sup>48</sup> was normalized with respect to all downregulated proteins. For reference, the same analysis was conducted in all the proteins identified by ref. 44 and in our data. The average percentage in each case is indicated in the bar chart.

which disrupts closed vesicles into open membrane sheets, releasing content proteins and peripheral proteins in soluble form<sup>41</sup>. Surprisingly, most sEVs protein patterns (668 proteins) remained unaffected by Na<sub>2</sub>CO<sub>3</sub>, except for a minor subset of 63 which appeared at very high-density fractions (F14–F16) (Fig. 5c,d and Supplementary Table 5). Remarkably, these proteins showed a concomitant decrease from the low-density fractions (F03–07) with their total levels remaining constant (Fig. 5d,e), indicating a genuine redistribution of protein content from low- to very high-density fractions. Among them was MFGE8,

a known peripheral sEV protein<sup>42</sup>. Importantly, these 63 proteins were enriched in 'cytosol' (67%) (Fig. 5e), including soluble proteins such as TPI, TALDO1 and LDHA, in stark contrast to the remaining 604 sEVs proteins, which were predominantly 'plasma membrane' (71%). These findings suggest that while sEVs do contain soluble cytosolic proteins, their contribution may be lower than previously assumed. Strikingly, we also identified 72 and 91 NV-dual proteins that shifted from low- to very high-density fractions, whereas their abundance at high densities (F09–11) remained unchanged (Fig. 5d,e). These proteins were



also enriched in 'cytosol' (Fig. 5e) and included GAPDH, ALDOA and ENO1. These results imply that some NV-dual proteins may represent cytosolic proteins that are associated to both vesicles (low density) and NV fractions (high density).

Three mechanisms enable cytosolic protein engulfment into ILVs: LC3-dependent EV loading and secretion (LDELS)<sup>43</sup>, exosomal LAMP2A loading of cargo (e-LLOC)<sup>44</sup> and endosomal microautophagy (eMI)<sup>45</sup>. All intersect with autophagy and, in line with this, we found several autophagic proteins present in sEVs (Fig. 5f). LDELS requires the LC3-conjugation machinery (MAP1LC3B), neutral sphingomyelinase 2 (SMPD3/nSMase2) and LC3-dependent recruitment of factor associated with nSMase2 activity (FAN or NSMaf), all of which were low abundant in sEVs (Fig. 5f). LDELS is involved in the secretion of RNA-binding proteins (RBPs) but none of the 16 RNA-binding proteins previously associated with LDELS<sup>43</sup> appeared in sEVs (Fig. 5g), casting doubt on their presence in sEVs<sup>6</sup>. Both e-LLOC and eMI rely on HSPA8 (HSC70) recognizing KFERQ motifs in protein cargo<sup>46</sup>. HSPA8 was the fourth most abundant protein in sEVs, accompanied by co-chaperones STUB1 (CHIP), DNAJB1 and STIP1 (HOP)<sup>47</sup> (Fig. 5f). In e-LLOC, HSPA8 cooperates with LAMP2A translocase<sup>44</sup>, which was also confirmed in sEVs (Supplementary Table 1). However, using a comprehensive prediction of KFERQ-containing proteins<sup>48</sup>, we did not find any enrichment in KFERQ motifs in reported e-LLOC cargoes (Fig. 5h). While 56% of the proteins reported to be downregulated in sEVs in LAMP2A KO cells<sup>44</sup> contained predicted KFERQ motifs, a similar proportion (53%) was found among all the identified proteins in the same study. Despite this, we identified 12 of 30 validated HSPA8 substrates<sup>48</sup> in our data, six of which (GAPDH, PKM, ALDOA, TPII, MDH1 and ENO1) were confirmed as luminal by Na<sub>2</sub>CO<sub>3</sub> treatment (Supplementary Table 5). Thus, while HSPA8 is likely to play a role in loading cytosolic cargo<sup>49</sup>, the number of putative substrates appears limited.

### Protease accessibility reveals the external association of NV-dual proteins with sEVs

Next, we assessed whether the pool of NV-dual proteins migrating at low densities is encapsulated within the sEVs lumen. To this end, crude sEVs were treated with proteinase K (PK) and analysed by LC-MS/MS. As expected, sEVs-classified proteins, including intraluminal marker SDCBP, remained shielded from proteolysis, confirming vesicle integrity (Fig. 6a and Supplementary Table 6). By contrast, NV

classes exhibited graded susceptibility (Fig. 6a,b): NV-dual 1 proteins showed partial degradation, NV-dual 2 were more severely affected and NV proteins showed the highest proteolysis rates. Degradation rates increased with enzyme concentration (Supplementary Fig. 6a), demonstrating PK specificity. We subsequently analysed these samples by DG-PCP (Fig. 6c). Although major degradation of NV-dual 1 and 2 proteins occurred indeed in high-densities (F10–11) (Fig. 6d), marked degradation was also evident in low-densities (F05–07), considerably exceeding that of bona fide sEV proteins (Fig. 6d,e). This indicates that NV-dual proteins comigrating with sEVs are not membrane encapsulated, suggesting instead an external association. Interestingly, we found that smaller and more hydrophilic proteins are enriched in low-density fractions (Fig. 6f and Supplementary Fig. 6b), suggesting that these biophysical properties may mediate their association with vesicles. Moreover, our data suggest that NV proteins are not merely soluble species but may exist in a centrifugation-induced aggregated state<sup>50</sup> as evidenced by their higher buoyancy compared to soluble proteins. Supporting this, when we spiked SDS-solubilized *Escherichia coli* lysate into crude sEVs before DG, the 'ground-state' soluble *E. coli* proteins exhibited lower buoyancy than the endogenous NV proteins (Fig. 6g). A similar trend was observed in our Na<sub>2</sub>CO<sub>3</sub> assay, where released soluble proteins again shifted to higher densities than NV proteins (Fig. 5c).

### Major sources of contaminating proteins in sEVs

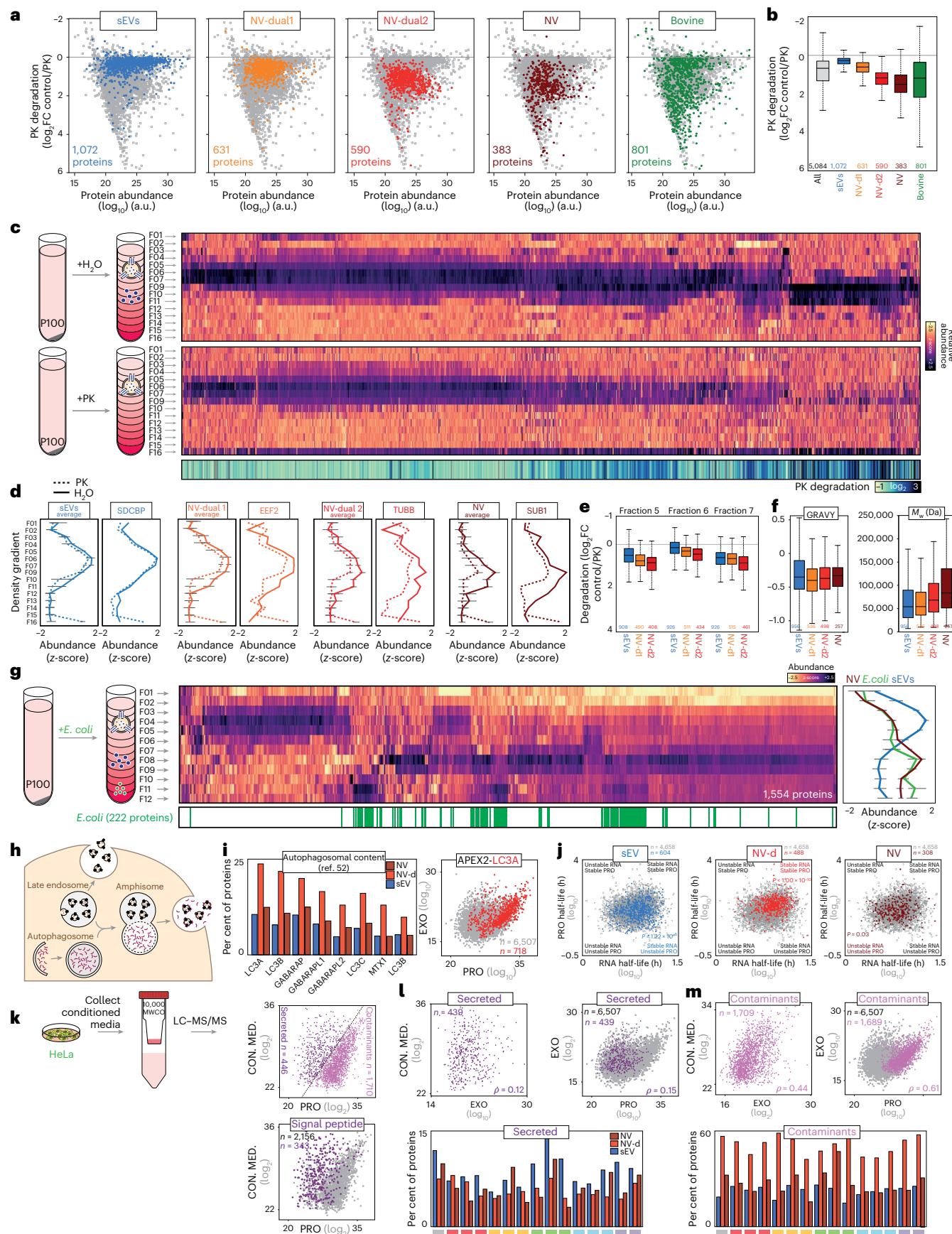
The presence of cytosolic proteins associated with sEVs may stem from two main sources. First, acidic conditions of amphisomes (that is, fusion of autophagosomes with late endosomes) can cause autophagic material to aggregate and adhere to ILVs<sup>51</sup> (Fig. 6h). Indeed, we found that known autophagosomal cargo proteins<sup>52</sup> predominantly exhibited NV-dual and NV profiles in our data (Fig. 6i) and strong extracellular-intracellular correlation (Fig. 6j), supporting their bulk capture. Further supporting an autophagic origin, NV-dual proteins were found to have significantly long half-lives<sup>53</sup> (Fig. 6j), consistent with the established role of autophagy in the clearance of old proteins<sup>54</sup>. By contrast, sEV proteins were typically short-lived despite stable mRNA levels, aligning with their active secretion into the extracellular space.

Beyond amphisomes, residual cellular debris represents another source of cytosolic contamination. Although cell viability exceeded 95% in our experiments, the large media volumes required for dUC can

### Fig. 6 | Contamination of sEVs preparations by copurifying cytosolic proteins.

**a**, Crude sEVs were either mock-treated (H<sub>2</sub>O) or digested with 45 ng  $\mu$ l<sup>-1</sup> of PK for 5 min before LC-MS/MS analysis. The extent of degradation (represented as the log<sub>2</sub>FC control/PK) is plotted against the protein abundance in control sEVs. Major protein classes (sEV, NV-dual 1, NV-dual 2 and NV) are colour-coded, with bovine proteins highlighted in green. **b**, Box plots displaying log<sub>2</sub>FC control/PK for sEV, NV-dual 1, NV-dual 2, NV and bovine proteins. For comparison, the distribution for all identified proteins is shown. Numbers above each category indicate protein counts (n). Box plot elements: centre lines indicate medians; the box limits represent the 25th and 75th percentiles; the whiskers extend to 1.5× the interquartile range. **c**, Crude sEVs treated with H<sub>2</sub>O or PK were separated by DG into 16 fractions and analysed by LC-MS/MS (F08 was excluded due to LC-MS/MS issue). PK degradation rates (that is, log<sub>2</sub>FC control/PK) are displayed below the heat map. **d**, Centroid profiles for all proteins classified as sEVs, NV-dual 1, NV-dual 2 and NV in control (solid line) and PK-digested sEVs (dashed line). Data (n = number of proteins in cluster) are presented as median values, with error bars representing the interquartile range. Examples of representative proteins from each category are shown. **e**, Box plots comparing log<sub>2</sub>FC control/PK for sEV, NV-dual 1 and NV-dual 2 proteins as measured in low-density fractions (F05–07). The numbers indicate protein counts (n). Box plot elements: centre lines indicate medians; the box limits represent the 25th and 75th percentiles; the whiskers extend to 1.5× the interquartile range. **f**, Comparison of protein hydrophobicity (GRAVY) and molecular weight (Da) distributions for sEV, NV-dual 1, NV-dual 2 and NV proteins. Numbers indicate protein counts (n). Box plot elements: the centre lines indicate medians; the box limits represent the 25th and 75th

percentiles; the whiskers extend to 1.5× the interquartile range. **g**, SDS-solubilized *E. coli* protein extract was spiked into crude sEVs and then separated by DG. Identified *E. coli* proteins are represented in green below the heat map. Centroids profiles for sEVs, NV and *E. coli* proteins are displayed on the right. Data (n = number of proteins in cluster) are presented as median values, with the error bars representing the interquartile range. **h**, A schematic illustrating how endosomal ILVs encounter autophagic material at amphisomes. **i**, Classification according to our data of proteins identified in autophagosomes via proximity labelling using different baits<sup>52</sup>. Correlation between our dataset (EXO) and intracellular protein abundance (PRO) is shown, with putative autophagosomal proteins (that is, LC3A interactors<sup>52</sup>) highlighted in red. **j**, Protein and mRNA half-lives as reported in ref. 53 for proteins classified as sEV, NV-dual and NV in our data are shown in the scatterplots. A two-dimensional enrichment analysis<sup>79</sup> was used for statistical evaluation. **k**, HeLa conditioned medium (CON. MED.) (24 h) was concentrated using a 10 kDa device filter and analysed by LC-MS/MS. Identified proteins (y axis) are plotted against their corresponding intracellular abundance (PRO) (x axis). Signal peptide-containing proteins are highlighted in purple (bottom scatterplot). Secreted proteins and potential contaminants were selected using the arbitrary displayed cut-off (top scatterplot). **l**, For proteins identified in conditioned medium as secreted, scatterplots are shown comparing protein abundance between conditioned medium versus sEVs (EXO) (left) and sEVs (EXO) versus intracellular protein levels (PRO) (right). The bar chart shows how these secreted proteins were classified in our DG-PCP datasets. **m**, Same as in **l** but for proteins identified as putative contaminants from conditioned medium. *M<sub>w</sub>*, molecular weight.



introduce trace contamination. To test this, we analysed the secretome using a 10 kDa filter (Fig. 6k). Although many signal-containing proteins appeared enriched in this secretome compared with the intracellular fraction (446 proteins), most of the proteins lacked signal peptides (1,710 proteins) and displayed a strong correlation with their intracellular abundance, implying the presence of contaminating cell debris (Fig. 6k). When mapped to our DG-PCP data, genuine secreted proteins showed no correlation with sEVs (Fig. 6l), whereas putative contaminants were strongly enriched among NV-dual and NV proteins (Fig. 6m). Collectively, these results indicate that non-vesicular proteins copurifying with sEVs during dUC may originate from both intracellular (autophagy) and extracellular (debris) sources.

### Fully assembled macromolecular complexes copurify with sEV during dUC

Our GO analysis revealed that NV and NV-dual proteins were enriched in macromolecular complexes (Supplementary Table 2). Indeed, 60% of NV-dual-classified proteins belonged to stable CORUM complexes (Fig. 7a). Among the most abundant was the ARP2/3 complex, where all seven subunits exhibited near-perfect co-elution and strong correlation with intracellular levels ( $\rho = 0.976$ ) (Fig. 7b). Similarly, chaperonin CCT subunits co-eluted and showed the expected stoichiometry of CCT6B (Fig. 7c). The ribosome was also highly abundant (72 subunits), preserving 1:1 stoichiometry for 40S/60S subcomplexes and comigrating with interactors such as PA2G4 and GCN1 (Fig. 7d–e). However, unlike ARP2/3 and CCT, ribosomal abundance in our data did not correlate with intracellular levels (Fig. 7f). Unsupervised clustering clearly distinguished sEV-associated ribosomes from intracellular pools, a finding confirmed in other datasets<sup>55,56</sup> (Supplementary Fig. 7a,b). This divergence from canonical cytosolic ribosomes persisted even when compared to monosomes and polyribosomes<sup>57</sup> (Supplementary Fig. 7c), suggesting that sEV preparations copurify a distinct ribosomal pool. Other complexes exhibited distinctive medium densities. For instance, collagen VI subunits migrated around fraction 6 (Fig. 7g), with COL6A1/2/3 probably forming the most abundant trimeric assembly and exhibiting similar profiles to other extracellular matrix proteins (AGRN1, HSPG2 and LAMC1). Similarly, abundant core and linker histones were present at medium densities, enabling identification of their specific epigenetic modifications (Fig. 7h). The proteasome was also abundant, though the 20S core and 19S regulatory particles migrated separately with differing stoichiometries (Fig. 7i–j), indicating separated subcomplexes. Interestingly, PSMD4 and ADRM1 exhibited distinct profiles and low stoichiometry (Fig. 7k), confirmed in published data<sup>56</sup> (Supplementary Fig. 7d) whereas PSMB8 co-eluted with 20S, suggesting immunoproteasome presence (Fig. 7k). Together, these findings demonstrate that numerous fully assembled macromolecular complexes with specific compositional features copurify with sEVs.

Recently, several extracellular nanoparticles have been identified<sup>58</sup>. Exosomes<sup>59</sup> can be pelleted from sEVs supernatants<sup>60</sup>, whereas supermeres are further pelleted from exosomes supernatants<sup>61</sup>. However, it is proposed that abundant cellular complexes may represent a continuum of particles that either copurify with exosomes and supermeres or constitute distinct subclasses of these entities<sup>62</sup>. Intrigued by the presence of complexes in our samples, we re-analysed a proteomic characterization of these novel nanoparticles<sup>61</sup> (Supplementary Fig. 8). Notably, while supermere pellets lacked major macromolecular assemblies (Supplementary Fig. 8c), we found that nearly 90% of the exomere protein mass consists of just four complexes: histones and the 20S proteasome account for 81%, with multimeric LGALS3BP and VCP contributing another 7% (Fig. 7l and Supplementary Fig. 8c). Whether these complexes explain the dot-shaped morphology initially attributed to exosomes—or if exosomes constitute distinct particles—remains a question for further study<sup>62</sup>.

### sEVs protein cargo serves as surrogate of their cell of origin

Despite the diversity of sEVs cargo, the extent to which it reflects parental cell molecular features remains unclear. To investigate this, we performed unsupervised clustering of all datasets and found that sEV proteins closely matched their corresponding intracellular protein and mRNA data in nearly all 15 cell lines (Fig. 8a), with similar results for NV-dual and NV proteins (Supplementary Fig. 9a). Although general cancer subtype clustering was absent, melanoma lines showed a distinct signature (Fig. 8a). Interestingly, sEVs clustered closer to RNA than PRO levels (Fig. 8a), reinforcing mRNA as a better predictor of sEV cargo. Because cancer cell lines recapitulate molecular features of their tumours of origin, we also examined tumour-specific gene signatures<sup>63</sup> (Supplementary Fig. 9b). This revealed that some proteins typically enriched in colon cancer patients (for example, GPA33 and FERMT1) were also abundant in the sEVs secreted by colon cancer cell lines. Similarly, melanoma markers MCAM, TRPV2 were enriched in melanoma sEVs. These results indicate that sEVs can serve as molecular surrogates for their parental cells.

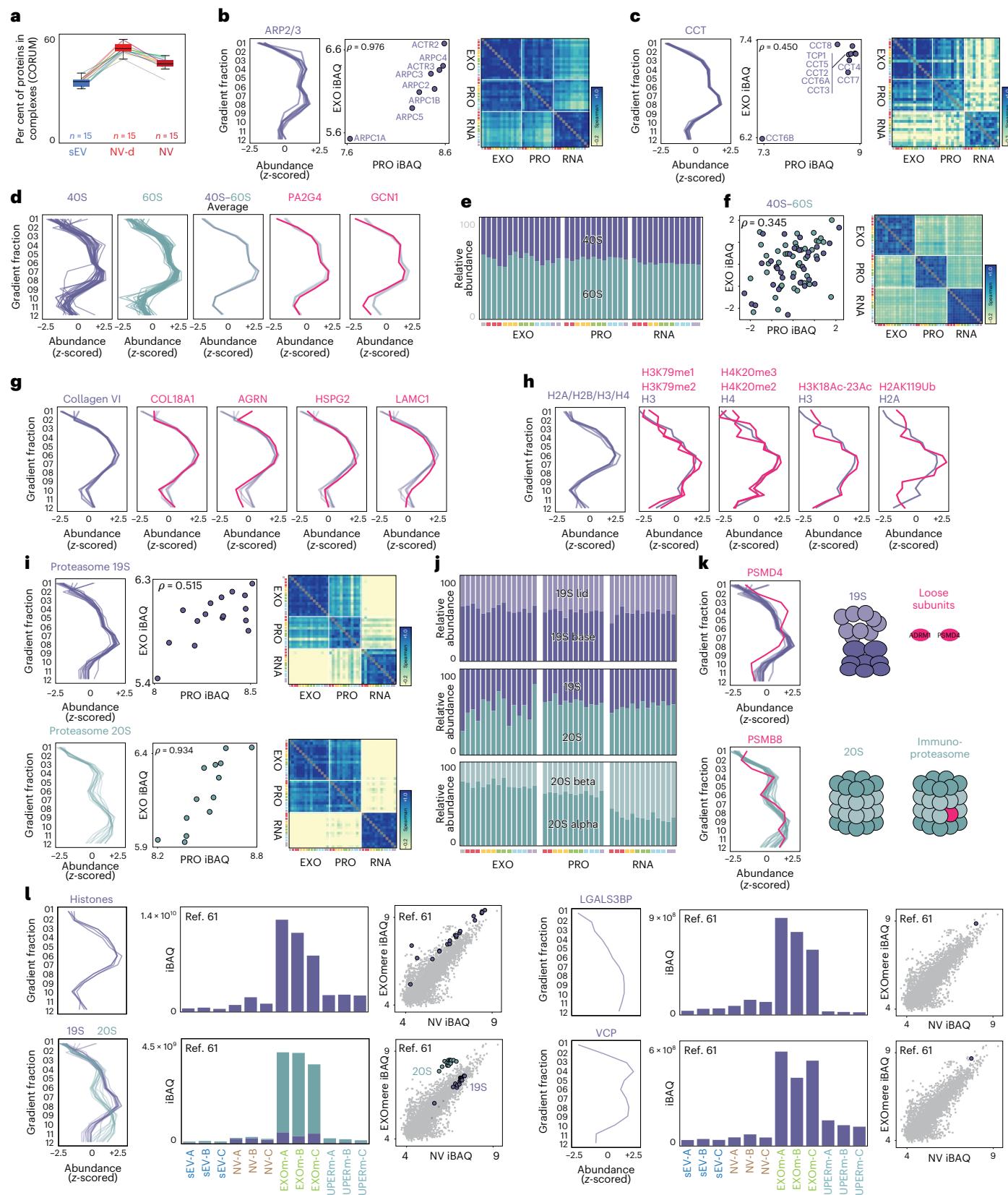
Interestingly, we found exclusive enrichment of melanosomal components in our melanoma sEVs (Fig. 8b). During melanosomal biogenesis, PMEL is sorted into the ILVs of early endosomes, which later diverge into stage I melanosomes rather than maturing into late endosomes<sup>27</sup>. However, our data showed that PMEL and other melanosomal markers co-eluted with the sEV marker CD63 (Fig. 8b). Consistently, melanosomal markers were absent in the amelanotic UACC62 cell line which expressed low levels of the melanogenesis master regulator MITF (Fig. 8b). Therefore, these findings suggest that melanoma cells, in addition to secreting melanosomes, release melanosomal proteins within sEVs, holding implications for understanding tumour-stroma interactions<sup>64</sup>.

**Fig. 7 | Characterization of macromolecular complexes copurifying with sEVs.** **a**, Percentage of proteins belonging to macromolecular complexes (as annotated in CORUM) among sEV, NV-dual and NV categories. Each colour line represents a different cell line ( $n = 15$  independent cell lines). Box plot elements: centre lines indicate medians; the box limits represent the 25th and 75th percentiles; the whiskers extend to  $1.5 \times$  the interquartile range. **b**, DG profiles of all eight ARP2/3 complex subunits in UACC62 as a representative example. The scatterplot compares their abundances in our dataset (EXO) versus intracellular levels (PRO). The heat map shows Spearman correlation values between EXO, PRO and RNA levels across all cell lines. **c**, Same as in **b**, but for CCT chaperonin complex subunits. **d**, DG profiles of 40S (purple) and 60S (green) ribosomal subunits in UACC62. Examples of top-correlating proteins with ribosomal subunits are highlighted in pink. **e**, Relative abundance of 40S and 60S ribosomal subunits in our data (EXO), intracellular protein (PRO) and mRNA (RNA) levels across all cell lines. **f**, Correlation between ribosomal protein abundance in EXO versus PRO in UACC62. The heat map shows Spearman correlation values between EXO, PRO and RNA levels across all cell lines. **g**, DG profiles of collagen VI subunits in UACC62. **h**, DG profiles of core histones in UACC62. Epigenetically modified peptides are highlighted in pink and compared with their corresponding unmodified histones. **i**, DG profiles of 19S and 20S proteasome subunits in UACC62. The scatterplot compares their abundance in EXO versus PRO. The heat map shows Spearman correlation values between EXO, PRO and RNA levels across all cell lines. **j**, Relative abundance of 19S (lid and base) and 20S (beta and alpha) proteasome subunits in EXO, PRO and RNA levels across all cell lines. **k**, DG profiles of the 19S subunit PSMD4 (top) and the 20S subunit PSMB8 (bottom), highlighted in pink, compared with all other detected 19S and 20S subunits, respectively. **l**, DG profiles of the indicated complexes and proteins in our data. The bar chart represents their abundances across sEV, NV, exomere (EXOm) and supermere (SUPERm) fractions, from the re-analysis of ref. 61 data. The scatterplot compares protein abundance between NV and exomere fractions, with the respective complex or protein highlighted.

PRO and RNA levels across all cell lines. **g**, DG profiles of collagen VI subunits in UACC62. Examples of top-correlating proteins with collagen VI are highlighted in pink. **h**, DG profiles of core histones in UACC62. Epigenetically modified peptides are highlighted in pink and compared with their corresponding unmodified histones. **i**, DG profiles of 19S and 20S proteasome subunits in UACC62. The scatterplot compares their abundance in EXO versus PRO. The heat map shows Spearman correlation values between EXO, PRO and RNA levels across all cell lines. **j**, Relative abundance of 19S (lid and base) and 20S (beta and alpha) proteasome subunits in EXO, PRO and RNA levels across all cell lines. **k**, DG profiles of the 19S subunit PSMD4 (top) and the 20S subunit PSMB8 (bottom), highlighted in pink, compared with all other detected 19S and 20S subunits, respectively. **l**, DG profiles of the indicated complexes and proteins in our data. The bar chart represents their abundances across sEV, NV, exomere (EXOm) and supermere (SUPERm) fractions, from the re-analysis of ref. 61 data. The scatterplot compares protein abundance between NV and exomere fractions, with the respective complex or protein highlighted.

Integrins are commonly secreted in sEVs and their specific patterns influence the organotropism of certain tumours<sup>65</sup>. Among the 26 known integrins, our data identified 17 strongly associated to sEVs, which exhibited pronounced differences in their composition across cell lines. For example, although sEVs from MDA-MB-231 and COLO205

were characterized by laminin-binding integrins, their specific compositions differed markedly (Fig. 8c and Supplementary Fig. 10). Moreover, while the integrin profile of sEVs (EXO) mirrored that of their cell of origin (PRO) in most cases, certain lines (for example, HCT116, H522 and UACC62) showed distinct integrin patterns in their



sEVs (Supplementary Fig. 10). Understanding these unique integrin profiles in sEVs may provide insights into the organotropism of different tumour types.

### DG-PCP enables in-depth characterization of serum and urine sEVs: implications for biomarkers

sEVs in biofluids represent a unique source of biomarkers, yet their characterization is challenged by the extreme dynamic range of protein abundance, particularly in blood. Consistent with this, we nearly identified the same proteins in crude sEVs (436 proteins) as in total serum (449 proteins) (Fig. 8d). However, upon DG separation, highly abundant soluble serum proteins (for example, albumin) migrated to high density fractions (F08-12) facilitating the detection of the less abundant, lower-density sEV proteins (F03-06) (Fig. 8d and Extended Data Fig. 10a). Consequently, we identified 204 bona fide serum sEV proteins (Supplementary Table 7), a fivefold improvement over dUC alone (39 proteins). Importantly, 85% of these 204 serum sEV proteins were also classified as sEVs in our cell lines (Fig. 8f), underscoring the high confidence of our assignments. Although lipoproteins copurify with serum sEVs during dUC<sup>66</sup>, our DG effectively separated them (Fig. 8d). In addition to lipoproteins, we noticed enrichment of other macromolecular complexes, including IgM pentamers, the APOL1-HPR complex, LGALS3BP and the 20S proteasome (Extended Data Fig. 10b). Comparing our data to blood cell proteomes<sup>67</sup> revealed that >56% of our serum sEVs proteins probably originate from platelets (for example, ITGA2B/CD41), as these proteins are nearly absent in all the other blood cell types (Fig. 8g). Similarly, 3% of sEVs proteins might originate from erythrocytes (for example, SLC4A1). Recently, studies have reported the identification of >4,000 proteins in human plasma EVs purified by strong anion exchange<sup>68</sup> and dUC<sup>69</sup>. Although these plasma EVs datasets were highly similar to each other, they differed from our serum sEV proteome (Extended Data Fig. 10c). Instead, both datasets closely resembled the platelet proteome<sup>67</sup> and were enriched in proteins released upon platelet activation<sup>70</sup> (Extended Data Fig. 10d–e). This raises the possibility of contamination by platelets and platelets-fragments, as they overlap in size and density with EVs and can be activated during plasma collection<sup>71</sup>. Collectively, our strategy enhances the characterization of genuine serum sEV proteins, clearly distinguishing them from copurifying complexes and confounding particles.

Urine is also a rich sEVs source. Our DG-PCP strategy identified 5,065 proteins in urinary sEVs (Fig. 8e and Supplementary Table 7). Among these, 2,398 proteins exhibited a low-density sEV profile, the majority of which were also classified as sEVs in our cell lines (Fig. 8h). Similarly, 57% of NV-dual proteins and 42% of NV proteins identified in urine matched their classifications in the cell lines. Urinary NV-dual fractions also contained macromolecular complexes (for example, CCT, ribosome and proteasome) and the highly abundant uromodulin

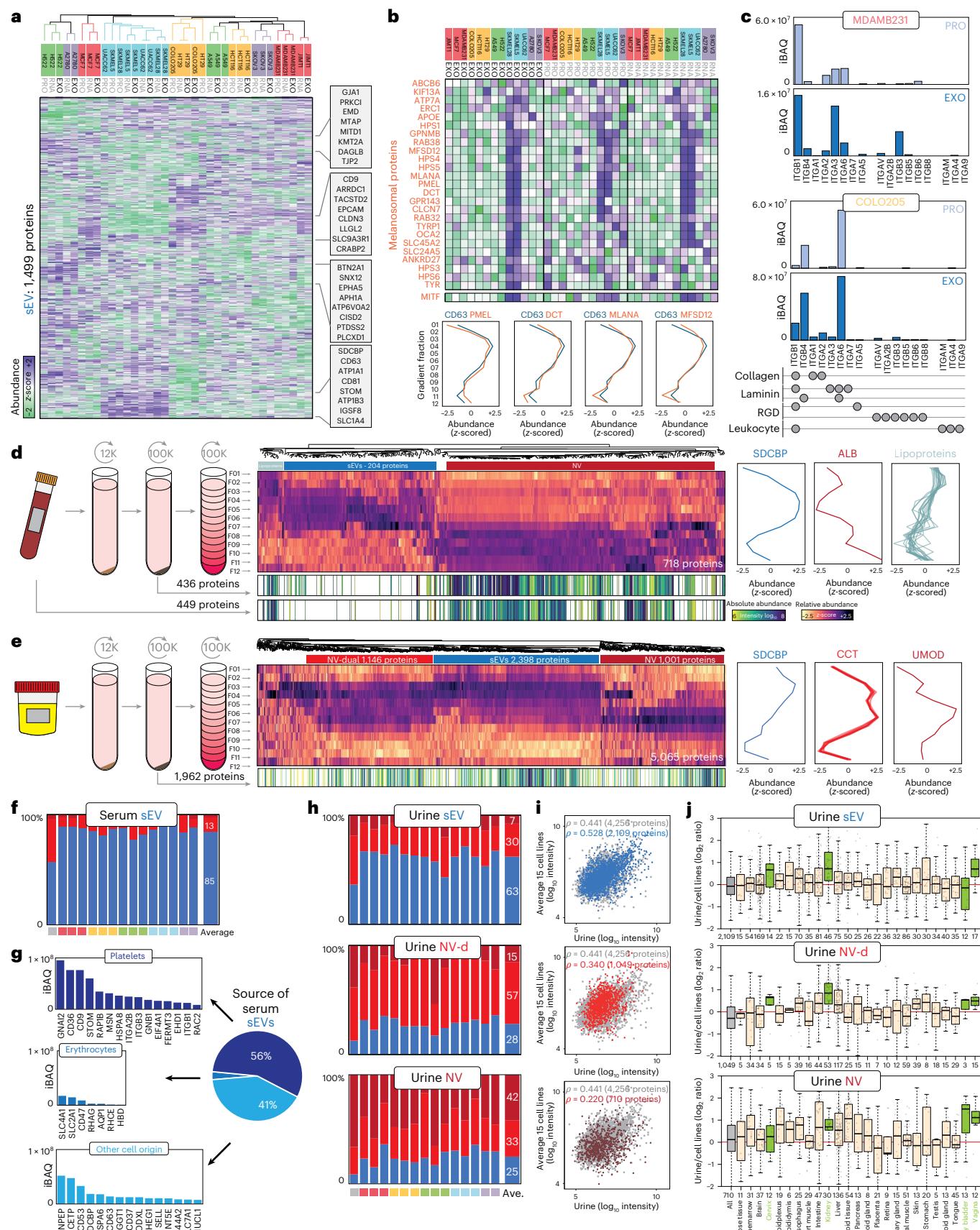
protein migrated to NV fractions (Fig. 8e). Comparing protein abundance between urine and cell lines revealed a positive correlation for sEV proteins ( $\rho = 0.528$ ), which exceeded that of NV-dual ( $\rho = 0.340$ ) and NV proteins ( $\rho = 0.220$ ) (Fig. 8i). These findings reveal the presence of a core sEV proteome with conserved abundance across cell types and biofluids. Moreover, the superior yield of sEV proteins in urine compared to serum suggests it is a more suitable biofluid for biomarker studies. Consistent with this, we found that urine sEVs contained numerous proteins (for example, MUC1, SLC12A3, DPEP1) highly specific to tissues contacting the urinary tract (Fig. 8j).

## Discussion

DG is selective yet limited in resolution, often leading to ambiguous protein assignments. We applied PCP to systematically assign sEV constituents by co-elution with established markers. This DG-PCP strategy enabled comprehensive mapping of sEVs, providing an unprecedented reference of sEVs protein cargo. We catalogued >1,500 proteins in sEVs, with some being up to 100,000× more abundant than others. Because a 100 nm vesicle could accommodate no more than ~1,000 protein molecules on its surface (assuming a 6 nm diameter for a 100 kDa protein), the identification of a number of proteins far exceeding the theoretical capacity of individual vesicles suggests that many proteins are present in only a small subset of vesicles. These findings support two models: (1) sEVs comprise distinct subpopulations, each with a defined composition, or (2) sEVs exhibit substantial variability, with each vesicle carrying a unique combination of proteins. The first determinative model would require multiple biogenesis/sorting pathways, while the second suggests a more stochastic process of cargo incorporation, in agreement with recent data<sup>72,73</sup>. Given the breadth of variability observed, we lean towards the latter hypothesis, where a single, stochastic mechanism is sufficient to generate the compositional heterogeneity of sEVs. This model of stochastic heterogeneity presents however a key conceptual challenge for the field: distinguishing bona fide, selectively sorted sEV cargo from the vast background of passively incorporated proteins. Resolving this will ultimately require single-vesicle proteomics, but our data provide a framework for prioritizing potential candidates, for instance, those showing evidence of intracellular depletion and conservation across diverse cell types. This compositional diversity is amplified by sEVs subtypes<sup>4</sup>, which DGs alone cannot resolve. However, we distinguished a distinct population of ER-enriched particles. While their intermediate density and ER membrane content suggest a vesicular nature, the absence of comigrating ribosomal proteins distinguishes them from rough microsomes, suggesting they may represent smooth microsomes or a specific ER-derived subpopulation. Regardless, this highlights the capacity of our approach to resolve copurifying nanoparticles via subtle density differences.

**Fig. 8 | Preservation of parental cell molecular signatures in sEVs: a framework for biomarker studies in serum and urine biofluids.** **a**, Unsupervised clustering of cell lines based on EXO (our data), PRO (intracellular protein levels) and RNA (mRNA expression) for the 1,499 proteins classified as sEVs. Representative proteins from specific clusters are shown in boxes. **b**, Abundance levels of melanosomal proteins across EXO, PRO and RNA datasets across cell lines. Bottom, elution profiles of representative melanosomal proteins are compared to the sEV marker CD63. **c**, Top: abundance levels of integrins identified in sEVs in our data (EXO) with corresponding intracellular protein levels (PRO). MDAMB231 and COLO205 cell lines are shown as representative examples. Major ligands for each integrin (bottom). **d**, P100 was purified from human serum via dUC and separated by DG into 12 fractions. Identified proteins are visualized in the heat map. Bottom: absolute protein levels (iBAQ) from unfractionated serum and P100 are shown. Right: elution profiles of representative proteins discussed in the text are displayed. **e**, P100 was purified from human urine via dUC and separated by DG into 12 fractions. Identified proteins are visualized in the heat map. Bottom, iBAQ from P100. Right: elution profiles of representative proteins

discussed in the text. **f**, Bar chart showing the percentage of the 204 genuine serum sEVs proteins classified as sEV, NV-dual or NV across the cell line datasets. **g**, Pie chart depicting the putative cellular origins of the 204 proteins identified as genuine serum sEV proteins. Left: examples of proteins ranked by decreasing abundance from platelets, erythrocytes and other cell types. **h**, Bar chart showing the percentage of the proteins identified in urine as sEVs, NV-dual and NV classified as sEV, NV-dual or NV across the cell line datasets. **i**, Scatterplot comparing the abundance of proteins identified in urine (x axis) and cell lines (average) (y axis), with proteins classified as sEV, NV-dual and NV in urine highlighted in different colours. **j**, Box plots of  $\log_2$  ratios of proteins abundance in urine versus cell lines of different tissue-specific proteins retrieved from The Human Protein Atlas<sup>80</sup>. Tissues in direct contact with the urinary tract are highlighted in green. The sample size ( $n$ ), representing the number of tissue-specific proteins analysed for each category, is indicated below the x axis for each box plot. Box plot elements: the centre lines indicate medians; the box limits represent the 25th and 75th percentiles; the whiskers extend to 1.5× the interquartile range.



Our finding that the sEV proteome does not simply mirror intracellular levels points to active sorting, a process where ubiquitination is involved<sup>31</sup>. Despite ubiquitin's abundance in sEVs, we barely detected ubiquitinated substrates. This suggests a predominance of unconjugated ubiquitin in sEVs, consistent with the requirement for cargo deubiquitination before ILVs packaging<sup>34</sup> but contrasts with reports suggesting that sEV ubiquitin is substrate conjugated<sup>74</sup>. As this discrepancy may stem from the limitations of our global proteomics profiling, a more definitive assessment would require targeted immuno-purification to enrich ubiquitinated cargo.

Contrary to the accepted view that sEVs encapsulate numerous cytosolic proteins, our findings show sEVs primarily contain plasma membrane proteins. This composition aligns with the two main sites of sEV formation: the plasma membrane and the endosomal system. We did confirm a limited number of bona fide cytosolic sEVs proteins. While their encapsulation in sEVs may occur passively, by incorporating proteins localized near sEV formation sites, the high levels of HSPA8 along with the presence of intraluminal proteins containing KFERQ motifs suggest an active sorting<sup>49</sup>. Given HSPA8's role in eM1<sup>45</sup>, this implies eM1-derived ILVs may be co-opted for secretory autophagy rather than degradation<sup>75</sup>.

This sparse sorted cytosolic cargo contrasts with the vast non-vesicular cytosolic proteins copurifying with sEVs. Their correlation with intracellular levels indicates they are not selectively incorporated into sEVs. While some may originate from amphisomes<sup>6</sup>, the majority probably result from residual cellular debris that might aggregate with sEVs during centrifugation. Many NV contaminants are annotated as sEVs in repositories, highlighting the urgent need for re-annotation of sEV proteomes. Similarly, numerous fully assembled macromolecular complexes sediment with sEVs. Their size and morphology make them particularly problematic, as they can interfere with the analysis of emerging extracellular nanoparticles<sup>62</sup>. The prevalence of contaminants is particularly problematic for functional studies, as biological effects may be misattributed to sEVs.

Although sEVs have traditionally been viewed as vehicles for cargo delivery, evidence shows that their uptake and the cytosolic release of luminal contents are highly inefficient<sup>76</sup>. Our data instead support a model in which sEVs function primarily as 'surface-active' signalling platforms. The enrichment of plasma membrane proteins, together with the scarcity of cytosolic cargo, is consistent with a role in modulating receptor activity at the cell surface rather than transferring luminal material. Biophysical models further show that in vesicles smaller than 200 nm, protein mass is dominated by the membrane rather than the lumen<sup>77</sup>, challenging the cargo-delivery paradigm and supporting a membrane-centred view of sEVs function.

Despite copurifying contaminants, genuine sEV proteins retain molecular signatures of their parental cells. This, together with the remarkable stability of sEVs in biofluids, highlights their potential as biomarkers. However, our findings underscore a major limitation for sEV-based biomarker discovery in blood<sup>78</sup>. Although DG separation is critical for separating the vast excess of NV proteins, the extreme dynamic range of blood proteins remains a formidable barrier to in-depth analysis, making it challenging to detect tissue-specific signals. By contrast, urine appears more promising, yielding over 2,400 genuine sEVs proteins. Since performing DG-PCP on every sample in a large cohort is unfeasible, we propose a hybrid strategy. This involves an initial DG-PCP on a representative pooled sample to create a study-specific reference map of high-confidence sEV proteins. This map can then be used as a powerful filter to analyse high-throughput data from crude sEVs of individual samples, enabling the prioritization of bona fide sEV biomarkers while excluding contaminants.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information,

acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41556-026-01878-z>.

## References

1. Rabouille, C. Pathways of unconventional protein secretion. *Trends Cell Biol.* **27**, 230–240 (2017).
2. Van Niel, G., D'Angelo, G. & Raposo, G. Shedding light on the cell biology of extracellular vesicles. *Nat. Rev. Mol. Cell Biol.* **19**, 213–228 (2018).
3. Mathieu, M., Martin-Jaular, L., Lavieu, G. & Théry, C. Specificities of secretion and uptake of exosomes and other extracellular vesicles for cell-to-cell communication. *Nat. Cell Biol.* **21**, 9–17 (2019).
4. Jeppesen, D. K., Zhang, Q., Franklin, J. L. & Coffey, R. J. Extracellular vesicles and nanoparticles: emerging complexities. *Trends Cell Biol.* **33**, 667–681 (2023).
5. Kowal, J. et al. Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes. *Proc. Natl. Acad. Sci. USA* **113**, E968–E977 (2016).
6. Jeppesen, D. K. et al. Reassessment of exosome composition. *Cell* **177**, 428–445.e18 (2019).
7. Foster, L. J. et al. A mammalian organelle map by protein correlation profiling. *Cell* **125**, 187–199 (2006).
8. Mann, M. The origins of organellar mapping by protein correlation profiling. *PROTEOMICS* **20**, 1900330 (2020).
9. Crescitelli, R., Lässer, C. & Lötvall, J. Isolation and characterization of extracellular vesicle subpopulations from tissues. *Nat. Protoc.* **16**, 1548–1580 (2021).
10. Frejno, M. et al. Proteome activity landscapes of tumor cell lines determine drug responses. *Nat. Commun.* **11**, 3639 (2020).
11. Ghandi, M. et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 (2019).
12. Keerthikumar, S. et al. ExoCarta: a web-based compendium of exosomal cargo. *J. Mol. Biol.* **428**, 688–692 (2016).
13. Kolberg, L. et al. g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res.* **51**, W207–W212 (2023).
14. Ellgaard, L. & Helenius, A. Quality control in the endoplasmic reticulum. *Nat. Rev. Mol. Cell Biol.* **4**, 181–191 (2003).
15. Mathieu, M. et al. Specificities of exosome versus small ectosome secretion revealed by live intracellular tracking of CD63 and CD9. *Nat. Commun.* **12**, 4389 (2021).
16. Nabhan, J. F., Hu, R., Oh, R. S., Cohen, S. N. & Lu, Q. Formation and release of arrestin domain-containing protein 1-mediated microvesicles (ARMMs) at plasma membrane by recruitment of TSG101 protein. *Proc. Natl. Acad. Sci. USA* **109**, 4146–4151 (2012).
17. Boshans, R. L., Szanto, S., van Aelst, L. & D'Souza-Schorey, C. ADP-ribosylation factor 6 regulates actin cytoskeleton remodeling in coordination with Rac1 and RhoA. *Mol. Cell. Biol.* **20**, 3685–3694 (2000).
18. Ghossoub, R. et al. Syntenin-ALIX exosome biogenesis and budding into multivesicular bodies are controlled by ARF6 and PLD2. *Nat. Commun.* **5**, 3477 (2014).
19. van Dam, S., Vösa, U., van der Graaf, A., Franke, L. & de Magalhães, J. P. Gene co-expression analysis for functional classification and gene–disease predictions. *Brief. Bioinform.* **19**, 575–592 (2018).
20. Baietti, M. F. et al. Syndecan-syntenin-ALIX regulates the biogenesis of exosomes. *Nat. Cell Biol.* **14**, 677–685 (2012).
21. Sigismund, S., Lanzetti, L., Scita, G. & Di Fiore, P. P. Endocytosis in the context-dependent regulation of individual and collective cell properties. *Nat. Rev. Mol. Cell Biol.* **22**, 625–643 (2021).
22. Cullen, P. J. & Steinberg, F. To degrade or not to degrade: mechanisms and significance of endocytic recycling. *Nat. Rev. Mol. Cell Biol.* **19**, 679–696 (2018).

23. Steinberg, F. et al. A global analysis of SNX27-retromer assembly and cargo specificity reveals a function in glucose and metal ion transport. *Nat. Cell Biol.* **15**, 461–471 (2013).
24. McNally, K. E. et al. Retriever is a multiprotein complex for retromer-independent endosomal cargo recycling. *Nat. Cell Biol.* **19**, 1214–1225 (2017).
25. Shin, J. J. H. et al. Spatial proteomics defines the content of trafficking vesicles captured by golgin tethers. *Nat. Commun.* **11**, 5987 (2020).
26. Vietri, M., Radulovic, M. & Stenmark, H. The many functions of ESCRTs. *Nat. Rev. Mol. Cell Biol.* **21**, 25–42 (2020).
27. van Niel, G. et al. The Tetraspanin CD63 Regulates ESCRT-Independent and -Dependent Endosomal Sorting during Melanogenesis. *Dev. Cell* **21**, 708–721 (2011).
28. Theos, A. C. et al. A luminal domain-dependent pathway for sorting to intraluminal vesicles of multivesicular endosomes involved in organelle morphogenesis. *Dev. Cell* **10**, 343–354 (2006).
29. Trajkovic, K. et al. Ceramide triggers budding of exosome vesicles into multivesicular endosomes. *Science* **319**, 1244–1247 (2008).
30. Kajimoto, T., Okada, T., Miya, S., Zhang, L. & Nakamura, S. Ongoing activation of sphingosine 1-phosphate receptors mediates maturation of exosomal multivesicular endosomes. *Nat. Commun.* **4**, 2712 (2013).
31. Raiborg, C. & Stenmark, H. The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins. *Nature* **458**, 445–452 (2009).
32. Akimov, V. et al. UbiSite approach for comprehensive mapping of lysine and N-terminal ubiquitination sites. *Nat. Struct. Mol. Biol.* **25**, 631–640 (2018).
33. Zhang, X. et al. An interaction landscape of ubiquitin signaling. *Mol. Cell* **65**, 941–955.e8 (2017).
34. Piper, R. C., Dikic, I. & Lukacs, G. L. Ubiquitin-dependent sorting in endocytosis. *Cold Spring Harb. Perspect. Biol.* **6**, a016808-a016808 (2014).
35. Huotari, J. & Helenius, A. Endosome maturation. *EMBO J.* **30**, 3481–3500 (2011).
36. Guo, H. et al. Atg5 Disassociates the V1V0-ATPase to promote exosome production and tumor metastasis independent of canonical macroautophagy. *Dev. Cell* **43**, 716–730.e7 (2017).
37. Choezom, D. & Gross, J. C. Neutral sphingomyelinase 2 controls exosome secretion by counteracting V-ATPase-mediated endosome acidification. *J. Cell Sci.* **135**, jcs259324 (2022).
38. Verweij, F. J. et al. ER membrane contact sites support endosomal small GTPase conversion for exosome secretion. *J. Cell Biol.* **221**, e202112032 (2022).
39. Liu, C. et al. Identification of the SNARE complex that mediates the fusion of multivesicular bodies with the plasma membrane in exosome secretion. *J. Extracell. Vesicles* **12**, 12356 (2023).
40. Dixson, A. C., Dawson, T. R., Di Vizio, D. & Weaver, A. M. Context-specific regulation of extracellular vesicle biogenesis and cargo selection. *Nat. Rev. Mol. Cell Biol.* <https://doi.org/10.1038/s41580-023-00576-0> (2023).
41. Fujiki, Y., Hubbard, A. L., Fowler, S. & Lazarow, P. B. Isolation of intracellular membranes by means of sodium carbonate treatment: application to endoplasmic reticulum. *J. Cell Biol.* **93**, 97–102 (1982).
42. Oshima, K., Aoki, N., Kato, T., Kitajima, K. & Matsuda, T. Secretion of a peripheral membrane protein, MFG-E8, as a complex with membrane vesicles. *Eur. J. Biochem.* **269**, 1209–1218 (2002).
43. Leidal, A. M. et al. The LC3-conjugation machinery specifies the loading of RNA-binding proteins into extracellular vesicles. *Nat. Cell Biol.* **22**, 187–199 (2020).
44. Ferreira, J. V. et al. LAMP2A regulates the loading of proteins into exosomes. *Sci. Adv.* **8**, 131–139 (2022).
45. Sahu, R. et al. Microautophagy of cytosolic proteins by late endosomes. *Dev. Cell* **20**, 131–139 (2011).
46. Fred Dice, J. Peptide sequences that target cytosolic proteins for lysosomal proteolysis. *Trends Biochem. Sci.* **15**, 305–309 (1990).
47. Kaushik, S. & Cuervo, A. M. The coming of age of chaperone-mediated autophagy. *Nat. Rev. Mol. Cell Biol.* **19**, 365–381 (2018).
48. Kirchner, P. et al. Proteome-wide analysis of chaperone-mediated autophagy targeting motifs. *PLoS Biol.* **17**, e3000301 (2019).
49. Gémard, C., Nault, F., Johnstone, R. M. & Vidal, M. Characteristics of the interaction between Hsc70 and the transferrin receptor in exosomes released during reticulocyte maturation\*. *J. Biol. Chem.* **276**, 9910–9916 (2001).
50. Linares, R., Tan, S., Gounou, C., Arraud, N. & Brisson, A. R. High-speed centrifugation induces aggregation of extracellular vesicles. *J. Extracell. Vesicles* <https://doi.org/10.3402/jev.v4.29509> (2015).
51. Buzas, E. I. Opportunities and challenges in studying the extracellular vesicle corona. *Nat. Cell Biol.* **24**, 1322–1325 (2022).
52. Guerroué, F. L. et al. Autophagosomal content profiling reveals an LC3C-dependent piecemeal mitophagy pathway. *Mol. Cell* **68**, 786–796.e6 (2017).
53. Schwahnüsser, B. et al. Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
54. Ciechanover, A. Proteolysis: from the lysosome to ubiquitin and the proteasome. *Nat. Rev. Mol. Cell Biol.* **6**, 79–87 (2005).
55. Rontogianni, S. et al. Proteomic profiling of extracellular vesicles allows for human breast cancer subtyping. *Commun. Biol.* **2**, 325 (2019).
56. Kugeratski, F. G. et al. Quantitative proteomics identifies the core proteome of exosomes with syntenin-1 as the highest abundant protein and a putative universal biomarker. *Nat. Cell Biol.* **23**, 631–641 (2021).
57. Imami, K. et al. Phosphorylation of the ribosomal protein RPL12/uL11 affects translation during mitosis. *Mol. Cell* **72**, 84–98.e9 (2018).
58. Jeppesen, D. K., Zhang, Q. & Coffey, R. J. Extracellular vesicles and nanoparticles at a glance. *J. Cell Sci.* **137**, jcs260201 (2024).
59. 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).
60. Zhang, Q. et al. Transfer of functional cargo in exosomes. *Cell Rep* **27**, 940–954.e6 (2019).
61. Zhang, Q. et al. Supermeres are functional extracellular nanoparticles replete with disease biomarkers and therapeutic targets. *Nat. Cell Biol.* **23**, 1240–1254 (2021).
62. Tosar, J. P., Cayota, A. & Witwer, K. Exosomes and supermeres: monolithic or diverse? *J. Extracell. Biol.* **1**, e45 (2022).
63. Uhlen, M. et al. A pathology atlas of the human cancer transcriptome. *Science* **357**, eaan2507 (2017).
64. Dror, S. et al. Melanoma miRNA trafficking controls tumour primary niche formation. *Nat. Cell Biol.* **18**, 1006–1017 (2016).
65. Hoshino, A. et al. Tumour exosome integrins determine organotropic metastasis. *Nature* **527**, 329–335 (2015).
66. Sódar, B. W. et al. Low-density lipoprotein mimics blood plasma-derived exosomes and microvesicles during isolation and detection. *Sci Rep.* **6**, 24316 (2016).
67. Rieckmann, J. C. et al. Social network architecture of human immune cells unveiled by quantitative proteomics. *Nat. Immunol.* **18**, 583–593 (2017).
68. Wu, C. C. et al. Enrichment of extracellular vesicles using Mag-Net for the analysis of the plasma proteome. *Nat. Commun.* **16**, 5447 (2025).
69. Kverneland, A. H., Østergaard, O., Emdal, K. B., Svane, I. M. & Olsen, J. V. Differential ultracentrifugation enables deep plasma proteomics through enrichment of extracellular vesicles. *PROTEOMICS* **23**, 2200039 (2023).

70. Houlahan, C. B. et al. Analysis of the healthy platelet proteome identifies a new form of domain-specific O-fucosylation. *Mol. Cell. Proteomics* **23**, 100717 (2024).
71. Nieuwland, R. & Siljander, P. R.-M. A beginner's guide to study extracellular vesicles in human blood plasma and serum. *J. Extracell. Vesicles* **13**, e12400 (2024).
72. Fordjour, F. K., Guo, C., Ai, Y., Daaboul, G. G. & Gould, S. J. A shared, stochastic pathway mediates exosome protein budding along plasma and endosome membranes. *J. Biol. Chem.* **298**, 102394 (2022).
73. Ai, Y. et al. Endocytosis blocks the vesicular secretion of exosome marker proteins. *Sci. Adv.* **10**, eadi9156 (2024).
74. Huebner, A. R. et al. Deubiquitylation of protein cargo is not an essential step in exosome formation\*. *Mol. Cell. Proteomics* **15**, 1556–1571 (2016).
75. Krause, G. J. et al. Reduced endosomal microautophagy activity in aging associates with enhanced exocyst-mediated protein secretion. *Aging Cell* **21**, e13713 (2022).
76. Bonsergent, E. et al. Quantitative characterization of extracellular vesicle uptake and content delivery within mammalian cells. *Nat. Commun.* **12**, 1864 (2021).
77. Zendrini, A. et al. On the surface-to-bulk partition of proteins in extracellular vesicles. *Colloids Surf. B* **218**, 112728 (2022).
78. Korff, K. et al. Pre-analytical drivers of bias in bead-enriched plasma proteomics. *EMBO Mol. Med.* <https://doi.org/10.1038/s44321-025-00309-0> (2025).
79. Cox, J. & Mann, M. 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinf.* **13**, S12 (2012).
80. Uhlen, M. et al. Tissue-based map of the human proteome. *Science* **347**, 1260419–1260419 (2015).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026

## Methods

### Cell culture

A total of 15 human cancer cell lines from different cancer types were used: cervix (HeLa), breast (JIMT1, MCF7 and MDA-MB-231), colon (Colo205, HCT116 and HT29), lung (A549, NCIH460 and NCIH522), melanoma (SKMEL28, SKMEL5 and UACC62) and ovarian (A2780 and SKOV3). Cell lines were cultured in Dulbecco's modified Eagle's medium (JIMT1, SKMEL28, SKMEL5, UACC62 and HeLa) or RPMI (MCF7, MDA-MB-231, Colo205, HCT116, HT29, A549, NCIH460, NCIH522, SKOV3 and A2780) supplemented with 10% (v/v) fetal bovine serum (FBS) and 100 units per millilitre penicillin–streptomycin at 37 °C in a 5% CO<sub>2</sub> humidified incubator. Cells were maintained by passage every 2–3 days at 80–90% confluence and tested negative for mycoplasma contamination.

### Purification of crude sEVs pellets from cell-conditioned media by dUC

Cell-conditioned medium (72 h) was collected at 90% cell confluence from 8 mm × 150 mm dishes using EV-depleted FBS (100,000g, 70 min). Cell viability was assessed as >95% in all experiments. The medium was centrifuged at 1,000g for 10 min (10 °C) to remove debris. Supernatant was centrifuged at 12,000g for 20 min (10 °C) to pellet large EVs (IEVs) (P12). The supernatant from the P12 was ultracentrifuged at 100,000g for 70 min (10 °C) to pellet sEVs (P100). The P100 was washed with 20 ml cold phosphate-buffered saline (PBS) and pelleted again in a second ultracentrifugation at 100,000g for 70 min (10 °C). The final P100 pellet was resuspended in 400 µl cold PBS and stored at –80 °C. All centrifugation steps were performed in an Optima XPN-100 ultracentrifuge (Beckman Coulter) using a Type 70 Ti Fixed-Angle rotor (Beckman Coulter). Particle content was analysed using a NanoSight NTA system (NanoSight; Malvern) equipped with a 405 nm laser.

### Purification of crude sEVs fractions (P100) from human serum and human urine by dUC

Human serum (4 ml) (H6914, Sigma-Aldrich) was centrifuged at 2,000g for 10 min (10 °C) to remove debris. The supernatant was diluted 1:1 with PBS and centrifuged at 12,000g for 20 min (10 °C) to pellet IEVs (P12). The supernatant from P12 was ultracentrifuged at 100,000g for 70 min (10 °C) to pellet sEVs (P100). The P100 pellet was washed with 8 ml cold PBS and pelleted again in a second ultracentrifugation at 100,000g for 70 min (10 °C). The final serum P100 pellet was resuspended in 100 µl cold PBS and stored at –80 °C. Human urine (170 ml) was centrifuged at 1,000g for 10 min (10 °C) to remove debris. The supernatant was centrifuged at 12,000g for 20 min (10 °C) to pellet IEVs (P12). The supernatant from P12 was ultracentrifuged at 100,000g for 70 min (10 °C) to pellet sEVs (P100). The P100 pellet was washed with 20 ml cold PBS and pelleted again in a second ultracentrifugation at 100,000g for 70 min (10 °C). The final urine P100 pellet was resuspended in 400 µl cold PBS and stored at –80 °C. All centrifugation steps were performed in an Optima XPN-100 ultracentrifuge (Beckman Coulter) using a Type 70 Ti Fixed-Angle rotor (Beckman Coulter). Particle content was analysed using a NanoSight NTA system (NanoSight; Malvern) equipped with a 405 nm laser.

### Iodixanol DG separation

Iodixanol gradient separation was done as described in Jeppesen et al.<sup>6</sup> with minor modifications. In brief, iodixanol density medium (Opti-Prep) (Stemcell, ref. 07820) was prepared in ice-cold sucrose buffer (0.25 M sucrose, 10 mM Tris pH 7.4, 1 mM EDTA) immediately before use. Crude sEVs pellets (P100) containing 30–40 µg of protein were resuspended in 900 µl of 40% iodixanol solution. The sample was loaded at the bottom of a centrifugation tube and 1.5 ml layers of decreasing iodixanol concentration (36%, 30%, 24%, 18% and 12%) were added sequentially on top. Gradients were ultracentrifuged at 100,000g for 16 h (10 °C) in an Optima XPN-100 ultracentrifuge (Beckman Coulter)

using a Type 90 Ti Fixed-Angle rotor (Beckman Coulter). Unless otherwise specified, twelve fractions of 700 µl each were sequentially collected from the top of the gradient. Particle content in the low-density fractions was analysed using a ZetaView PMX-130 (Particle Metrix, Germany) equipped with a 520 nm laser. Calibration was performed with polystyrene beads of 100 nm diameter before measurements. Samples were diluted in sterile, 0.1 µm filtered PBS to the optimal concentration range (10E7–10E9 particles per millilitre) and measured at 11 cell positions. The data were recorded and analysed using the ZetaView Software (version 8.06.01 SP1) and particle concentration was expressed as mean ± standard deviation.

### Electron microscopy

Cryo-EM was performed in the Electron Microscopy and Crystallography facility from CICbioGUNE (Derio, Spain). EV samples were adsorbed onto glow-discharged R2/1300-mesh holey carbon grids (Quantifoil), blotted at 95% humidity and rapidly vitrified in liquid ethane using a LEICA EM GP2 (Leica). Cryo-EM was performed at liquid nitrogen temperature on a JEM-1230 transmission electron microscope (JEOL), equipped with an UltraScan 4000 SP (4,008 × 4,008 pixels) cooled slow-scan CCD camera (GATAN) and a 120-kV LaB6 thermionic gun.

### Protease-protection assay

Crude sEVs (P100) from SKMEL28 and SKMEL5 were resuspended in 100 mM HEPES (pH 8.0) and 5 mM CaCl<sub>2</sub>. Samples were treated with different concentrations of PK (Promega) (45, 90, and 135 ng µl<sup>–1</sup>). Samples treated with H<sub>2</sub>O served as a control. The enzymatic reaction was carried out at 37 °C for 5 min and quenched with phenylmethylsulfonyl fluoride (5 mM final concentration). Samples were stored at –80 °C until LC–MS/MS analysis. In addition, SKMEL5 sEVs treated with 45 ng µl<sup>–1</sup> of PK and control sEVs were further separated by iodixanol gradients. In brief, P100 pellets were resuspended in a final concentration of 36% iodixanol. The sample was loaded at the bottom of a centrifugation tube and 700 µl layers of decreasing iodixanol concentration (30%, 24%, 18% and 12%) were added sequentially on top. Gradients were ultracentrifuged at 100,000g for 16 h (10 °C) in an Optima XPN-100 ultracentrifuge (Beckman Coulter) using a Type 50.4 Ti Fixed-Angle rotor (Beckman Coulter). A total of 16 fractions of 218 µl each were sequentially collected from the top of the gradient.

### Sample preparation for MS analysis

Crude sEVs pellets (P100) were lysed in 6 M urea. Protein concentration was quantified using a Qubit fluorometer. Samples (~10 µg of protein) were diluted sixfold with 100 mM Tris (pH 8.0). Samples were reduced (15 mM tris(2-carboxyethyl)phosphine, TCEP) and alkylated (30 mM chloroacetamide) at room temperature in the dark for 1 h, followed by overnight digestion with LysC/trypsin (1:50, enzyme:protein) at 37 °C. Peptides were desalted using C18 StageTips. DG fractions were processed using paramagnetic SP3 beads. In brief, fractions were solubilized in 2.5% SDS, followed by incubation at 25 °C for 1 h at 1,500 rpm. Proteins were reduced-alkylated as above, aggregated in 55% ethanol, washed with 80% ethanol and overnight digested with LysC/trypsin (1:50, enzyme:protein) at 37 °C. Peptides were desalted using a C18 desalting plate (the Nest Group, HNS S18V).

### LC–MS/MS analysis

For the analysis of the 15 cancer cell lines we used an Exploris 480 mass spectrometer (Thermo Fisher Scientific). The mass spectrometer was operated in a DIA mode using 60,000 MS1 resolution and 15,000 MS2 resolution. Ion peptides were fragmented using higher-energy collisional dissociation with a normalized collision energy of 29 and assuming a default charge state of +2. The normalized AGC target was set to 300% for MS1 (maximum injection time of 25 ms) and 1,000% for DIA MS/MS (maximum injection time of 22 ms). The 4 m/z precursor isolation windows were used in a staggered-window pattern

from 400.4 to 1004.7  $m/z$ . A precursor spectrum was interspersed every 151 DIA spectra. The scan range of the precursor spectra was 390–1,000  $m/z$ . For the analysis of HeLa triplicates as described in Fig. 1c and Extended Data Fig. 3a we used a Q Exactive HF (Thermo Fisher Scientific) and for the comparative analysis of 16 and 8 DG fractions as described in Extended Data Fig. 4b we used a Q Exactive HF-X (Thermo Fisher Scientific) using in both cases DIA. The HeLa experiment described in Fig. 1a was run in a Q Exactive Plus in DDA mode with an automatic switch between MS and MS/MS scans using a top 15 method (intensity threshold  $\geq 6.7 \times 10^4$ , dynamic exclusion of 25 s, and excluding charges +1 and +6). MS spectra were acquired from 350 to 1,400  $m/z$  with a resolution of 70,000 FWHM (200  $m/z$ ). Ion peptides were isolated using a 2.0 Th window and fragmented using higher-energy collisional dissociation with a normalized collision energy of 27. MS/MS spectra resolution was set to 17,500 (200  $m/z$ ). The normalized ion target values were  $3 \times 10^6$  for MS (maximum injection time of 25 ms) and  $1 \times 10^5$  for MS/MS (maximum injection time of 45 ms). All mass spectrometers were coupled to an UltiMate 3000 RSLC nano LC system. Peptides were loaded into a trap column (Acclaim PepMapTM 100, 100  $\mu\text{m} \times 2\text{ cm}$ , 5  $\mu\text{m}$ , Thermo Fisher Scientific) for 3 min at a flow rate of  $10 \mu\text{l min}^{-1}$  in 0.1% FA. Then, peptides were transferred to an EASY-Spray PepMap RSLC C18 column (Thermo) (2  $\mu\text{m}$ , 75  $\mu\text{m} \times 50\text{ cm}$ ) operated at 45 °C and separated using a 60 min effective gradient (buffer A: 0.1% formic acid; buffer B: 100% acetonitrile, 0.1% formic acid) at a flow rate of  $250 \text{ nl min}^{-1}$ .

#### Data analysis for PCP

DIA raw data were analysed with DIA-NN (version 1.8.1) using a library-free approach with a concatenated fasta databases containing proteins from *Homo sapiens* (UniprotKB, 20,610 sequences) and *Bos taurus* (UniprotKB, 23,847 sequences) and supplemented with frequent contaminants. Carbamidomethylation of cysteines was set as a fixed modification whereas oxidation of methionine and protein N-termini acetylation were set as variable modifications. Up to one missed cleavage was allowed, peptide length range was set to 7–30 and precursor charge range was 2–4. Match between runs was enabled whereas normalization was disabled. Protein inference was done based on the fasta protein entries. Precursor false discovery rate (FDR) was set to 1%. DDA raw data was analysed with MaxQuant (versions 2.0.1.0 and 2.4.2.0) using the same fasta database and protein modification settings used for DIA data analysis. Peptide and protein FDR were set to 1%. For PCP, protein group files from DIA-NN were loaded in Perseus (version 1.6.50). Protein intensities were  $\log_2$  transformed, missing values were imputed from a normal distribution of low intensities using the total matrix and then transformed to z-scores. Only proteins quantified in at least four DG fractions were used. Protein profiles were further filtered out by cosine fitting (FDR 5%, 1,000 randomizations). The resulting hierarchical tree was partitioned into 100 to 200 discrete clusters using the ‘Define row clusters’ function in Perseus. Clusters containing a minimum of 20 proteins were manually annotated based on their density profiles as: sEV (exclusive enrichment in low-density fractions), NV (exclusive enrichment in high-density fractions) or NV-dual (presence in both density regions). When necessary, the NV-dual category was further stratified into NV-dual1 (major peak in low-density fractions with a high-density tail) and NV-dual2 (major peak in high-density fractions with a low-density tail). Remaining clusters failing to meet this size threshold ( $N < 20$ ) were designated as UNC.

#### Statistical analysis

For the identification of proteins displaying a different DG profile in sEVs treated with  $\text{Na}_2\text{CO}_3$  we used a one-sided paired  $t$ -test comparing protein abundance between F3 and F7 of  $\text{Na}_2\text{CO}_3$  and F3 and F7 dimethyl sulfoxide (DMSO). Likewise, a one-sided paired  $t$ -test comparing F13–F16 between  $\text{Na}_2\text{CO}_3$  and DMSO was used to define proteins increasing in abundance at high density fractions in response to  $\text{Na}_2\text{CO}_3$  treatment.

Multiple testing was corrected by a permutation-based FDR. Proteins with a  $q$ -value  $< 0.2$  and a  $\log_2$  fold change  $> +0.3$  or  $< -0.3$  were defined as significant. Finally, only significant proteins in both analyses (that is, increasing at high densities and decreasing from low densities) were defined as  $\text{Na}_2\text{CO}_3$  responsive-proteins. Only proteins with well-defined sEV, NV-dual and NV clusters in the DMSO data were considered.

#### Identification of ER-associated proteins at medium density

To identify proteins associated with the selected ER displaying medium density in our DG, we performed a correlation-based analysis using Pearson correlation coefficients. Specifically, for each cell line, we selected the top 150 proteins that exhibited the highest Pearson correlation with DDOST. Proteins that showed a consistent correlation with DDOST in at least 8 out of the 15 cell lines were classified as DDOST-correlating proteins. We applied the same correlation analysis to identify proteins associated with TMED10, HSPA5 and HSP90B1, selecting the top 150 proteins most correlated with each marker in individual cell lines and retaining those that were consistently correlated in at least 8 out of the 15 cell lines.

#### Co-expression analysis

Only proteins classified as sEV in at least 14 out of the 15 cell lines used in this study were used for the co-expression analysis (ARF6 which was classified as sEVs in 12 cell lines was included). Spearman rank coefficient values ( $\rho$ ) and the corresponding  $P$  values were calculated for all protein pairs using Perseus (version 1.6.50).

#### Western blotting

Cells or sEVs (p100) were lysed in 2% SDS, 100 mM HEPES (pH 8.0), sonicated and cleared by centrifugation (20,000g, 40 min at room temperature). Then, 10  $\mu\text{g}$  of protein were loaded for SDS–polyacrylamide gel electrophoresis and transferred to nitrocellulose membranes. Blocking was performed in 5% milk in PBT (1× PBS, 0.1% Tween-20). Primary antibodies were incubated overnight at 4 °C or 2 h at room temperature. Secondary antibodies were incubated 1 h at room temperature. Primary antibodies used: Abcam, anti-GM130 (1/1,000; ab52649), anti-calreticulin (1/1,000; catalogue number ab2907), anti-syntenin (1/1,000, ab133267), anti-CD63 (1/1,000; ab193349). Millipore, anti-CD9 (1/1,000; CBL162). Secondary antibodies used, Jackson ImmunoResearch, anti-Mouse-HRP (1/10,000; 115-035-062), anti-Rabbit-HRP (1/10,000; 111-035-045). Proteins were detected using SuperSignal Pico Plus or West Femto (Thermo Fisher) in a ChemiDoc MP imaging system (Bio-Rad). All uncropped blots are provided within the Source data file.

#### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

#### Data availability

The mass spectrometry proteomics data generated in this study have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifiers [PXD062547](#) and [PXD062477](#). Intracellular protein levels were obtained from the supplementary material of Frejno et al.<sup>10</sup>. Total mRNA levels were retrieved from the Cancer Cell Line Encyclopedia at the Expression Atlas repository (<https://www.ebi.ac.uk/gxa/experiments/E-MTAB-2770/Results>). All other data supporting the findings of this study are available from the corresponding author on reasonable request. Source data are provided with this paper.

#### Acknowledgements

This work was supported by European Union Horizon 2020 programme INFRAIA project EPIC-XS (project no. 823839) and by grant PID2021-123144OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ‘ERDF/EU’ to J.M. Both J.M.F.-P. and J.M. are supported by the

Ikerbasque programme (Basque Foundation for Science). J.M.F.-P. is supported by MICIU/AEI /10.13039/501100011033 (Severo Ochoa: CEX2021-001136-S), MICIU/AEI /10.13039/501100011033) and FEDER, UE (Plan Nacional: PID2024-155622OB-I00), The European Union (Halt-Ronin: 101095679 and EVCA: 101079264). O.B.-G. is supported by the Juan de la Cierva grant JDC2022-048335-I (Ministerio de Ciencia e Innovación). We thank F. J. Sánchez for preliminary work on the project, members of the Microenvironment and Metastasis Group for technical assistance in sEV purification, and A. Rojas, D. Charro and I. Santos for their technical assistance in the cryo-EM analysis.

## Author contributions

J.M.-S. performed most DG experiments and analysed the data. O.B.-G. and J.E.-U. performed additional proteomics experiments, validations and contributed to data analysis and interpretation. E.G. and J.M.F.-P. assisted with microscopy and ultracentrifugation. M.S., P.X.-E., F.G. and E.Z. contributed to LC-MS/MS analyses. H.P. contributed to experimental design. J.M.-S. and J.M. planned all the experiments and wrote the paper. All authors contributed to and approved the final version of the paper.

## Competing interests

The authors declare no competing interests.

## Additional information

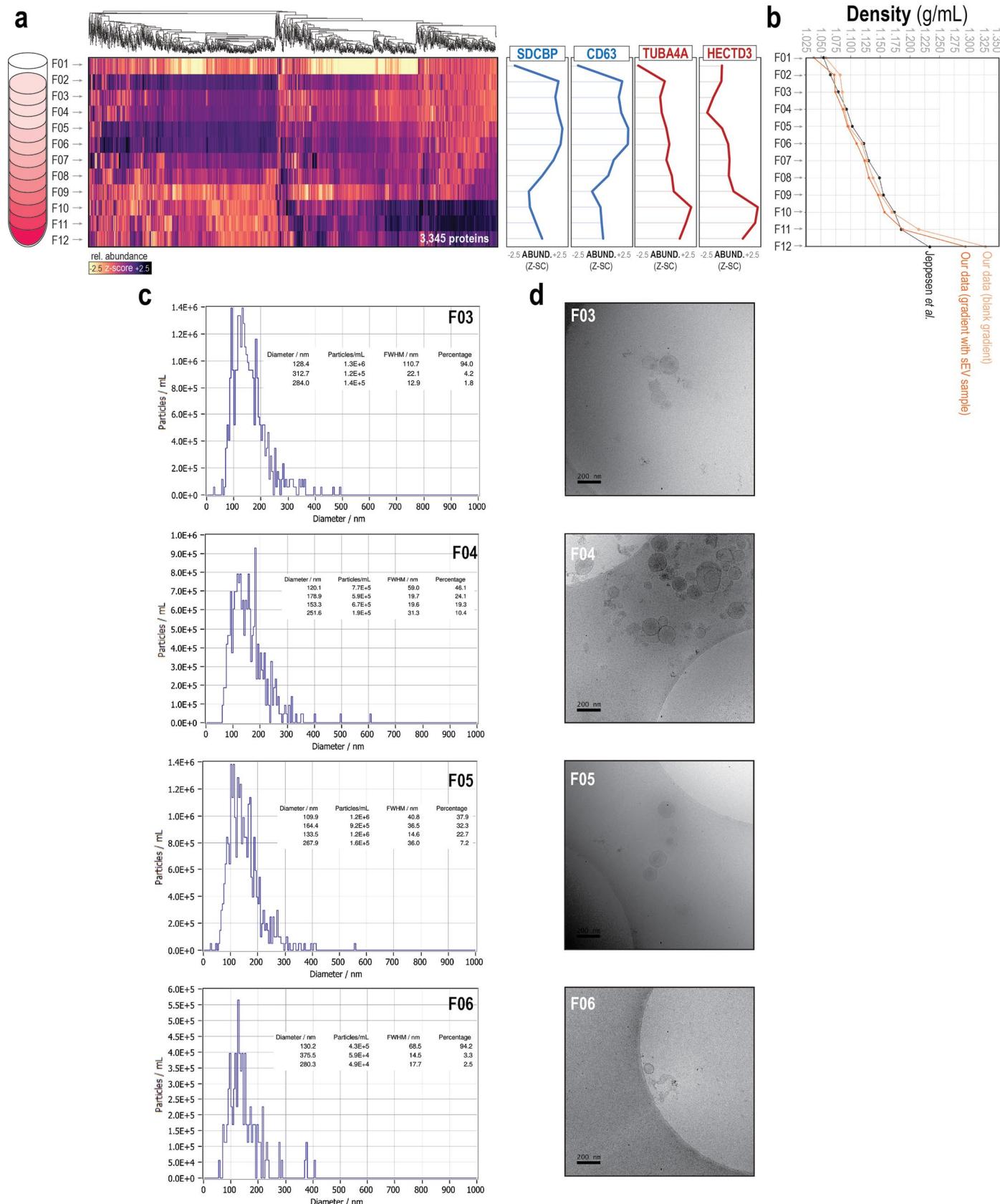
**Extended data** is available for this paper at <https://doi.org/10.1038/s41556-026-01878-z>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41556-026-01878-z>.

**Correspondence and requests for materials** should be addressed to Javier Muñoz.

**Peer review information** *Nature Cell Biology* thanks Kwang Pyo Kim and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

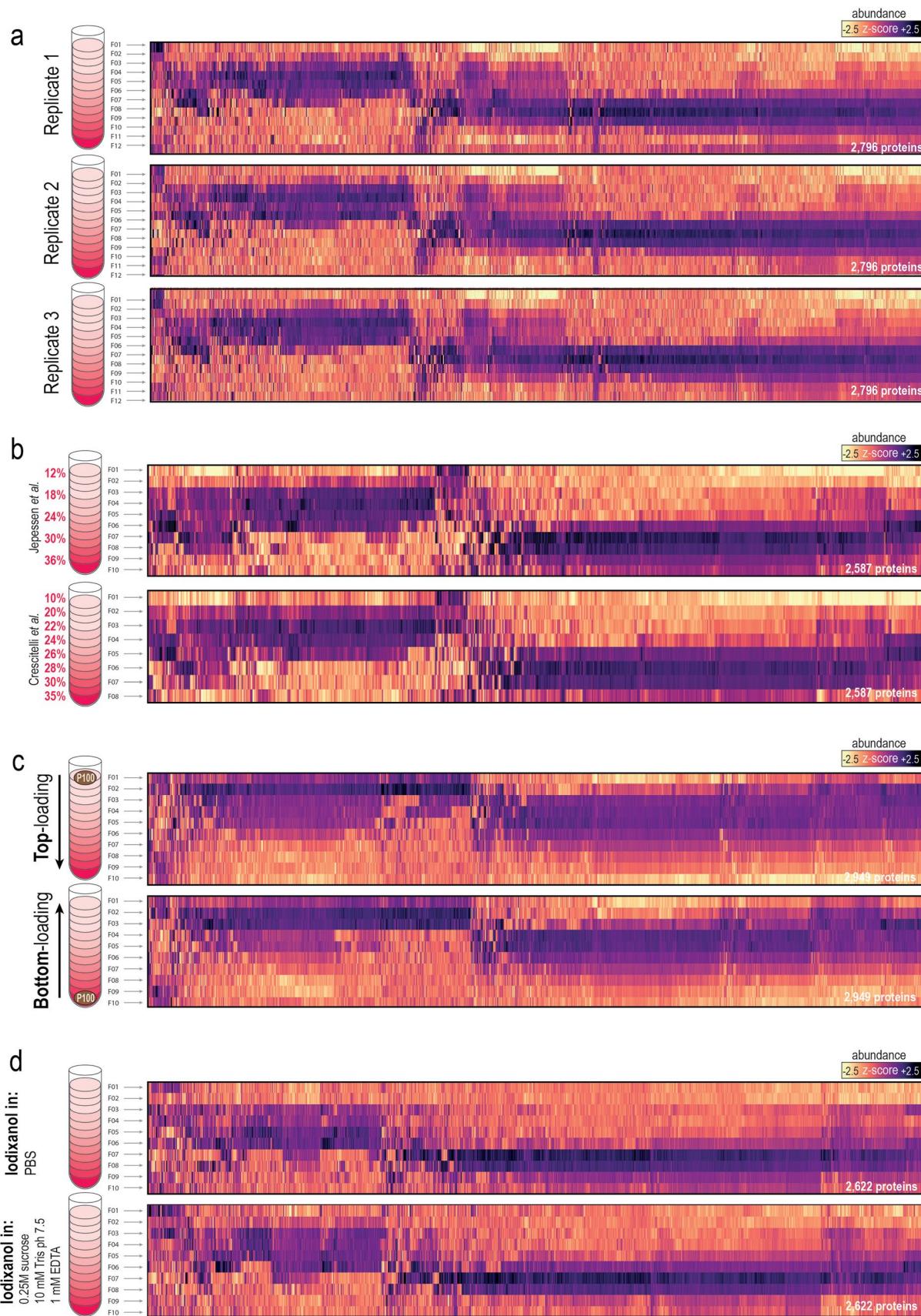
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



#### Extended Data Fig. 1 | Upstream validation of gradient-purified sEVs.

**a**, Crude sEVs (P100) from UACC62 were separated by DG into 12 fractions and analyzed by LC-MS/MS. Representative examples of sEV markers (blue) and NV-associated (brown) proteins are shown. **b**, Refractometry analysis confirmed that sEV-enriched fractions (F02-F07) showed buoyant densities ranging from 1.075 to 1.125 g/mL. As a reference, the densities measured by Jeppesen et al. using

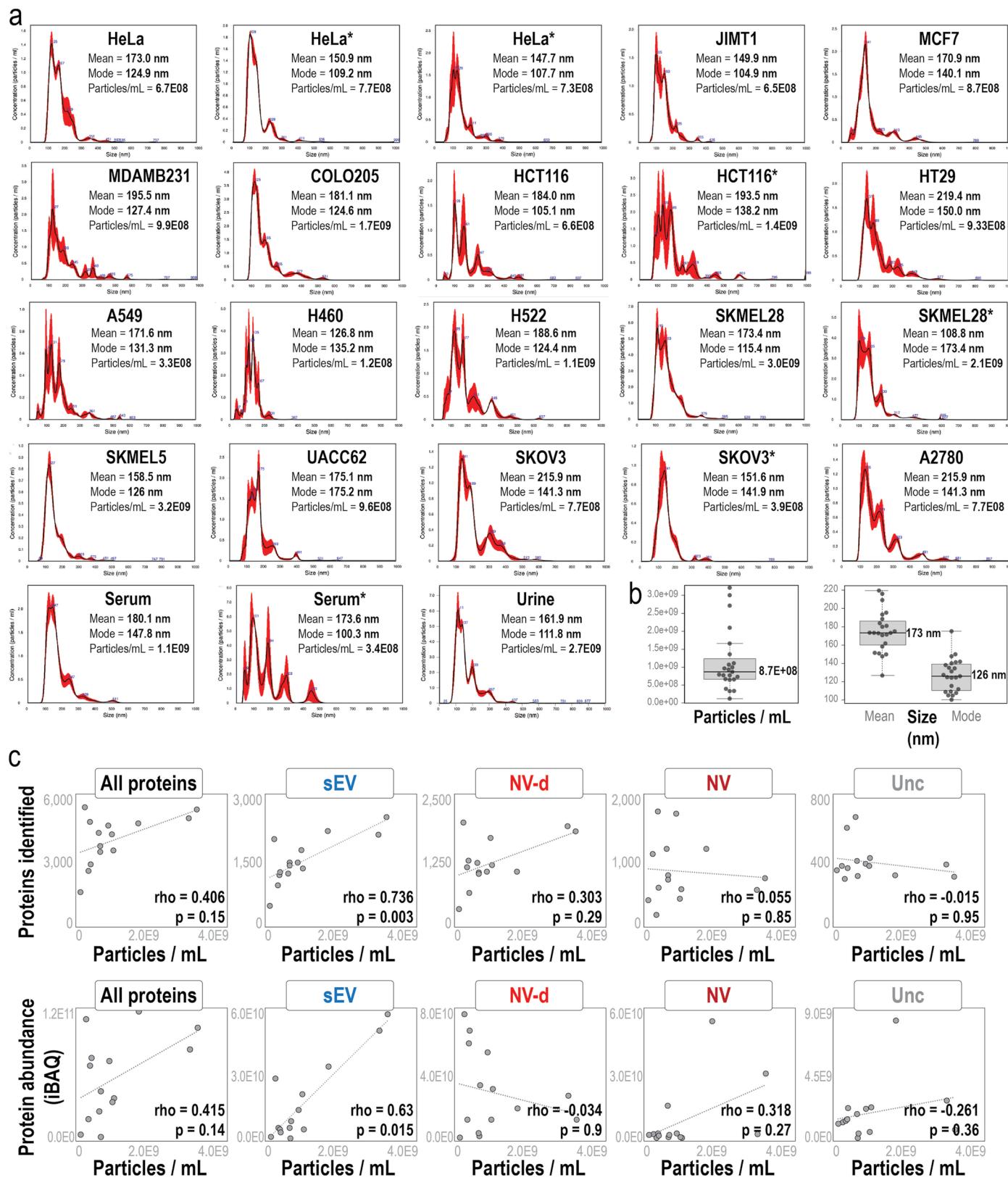
the same gradient composition are shown in black. **c**, NTA histograms of the corresponding low-density fractions, showing a primary particle population with a diameter of ~122 nm (average of F03-F06). **d**, Representative cryo-EM images of low-density fractions (F03-F06) showing intact vesicular structures consistent with sEVs. This experiment was performed once.



Extended Data Fig. 2 | See next page for caption.

**Extended Data Fig. 2 | Evaluation of the DG-PCP approach.** **a**, Crude sEVs (P100) were pelleted by dUC and bottom-loaded into three independent density gradients. **b**, Comparison of the protein profiles using the density gradient formulation by Jeppesen et al.<sup>1</sup> with those obtained by using the alternative formulation by Crescitelli et al.<sup>2</sup>. **c**, Crude sEVs pellet (P100) was loaded at the bottom of a centrifugation tube and layers of decreasing iodixanol concentration

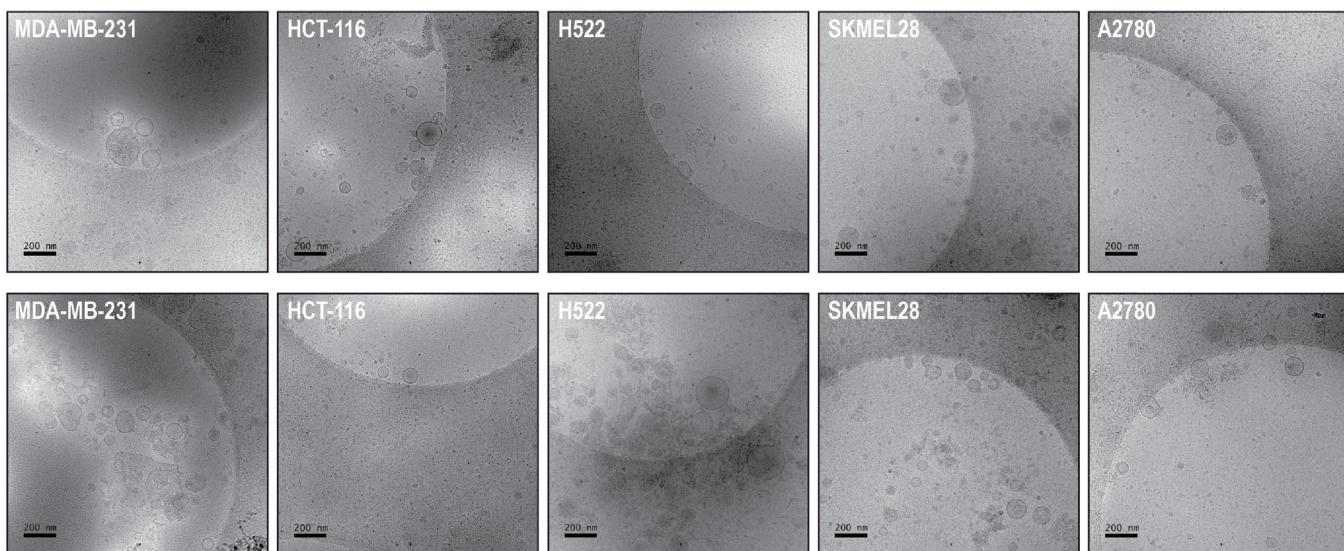
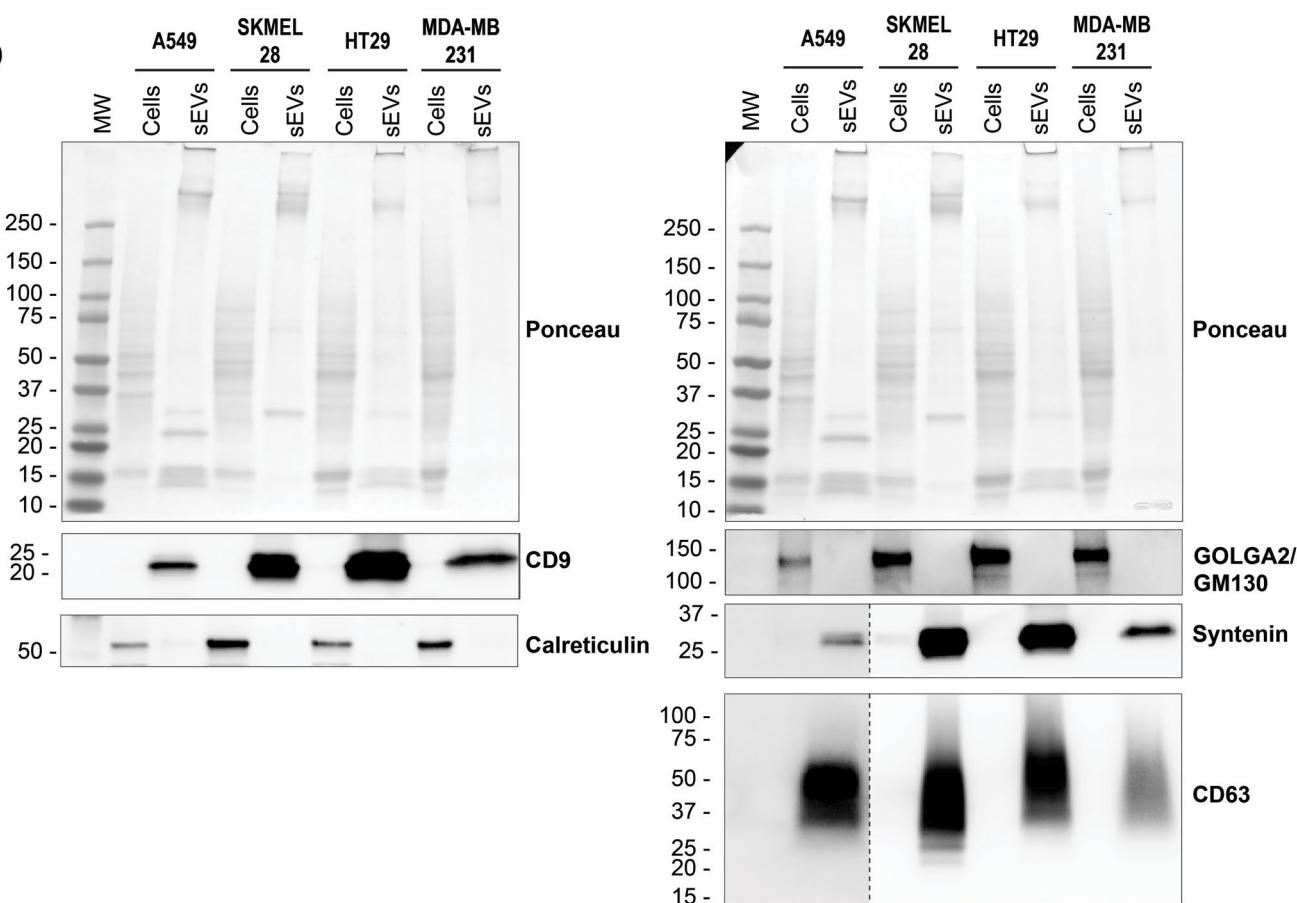
were added sequentially on top ('bottom-loading'). Alternatively, layers of decreasing iodixanol concentration were first added sequentially in the centrifugation tube and the crude sEV pellet was loaded on top of the lowest iodixanol concentration layer ('bottom-loading'). **d**, Protein elution profiles using iodixanol in PBS (as in Jeppesen et al.<sup>1</sup>) or iodixanol in 0.25 M sucrose, 10 mM Tris pH 7.5 and 1 mM EDTA (as in Crescitelli et al.<sup>2</sup>).



Extended Data Fig. 3 | See next page for caption.

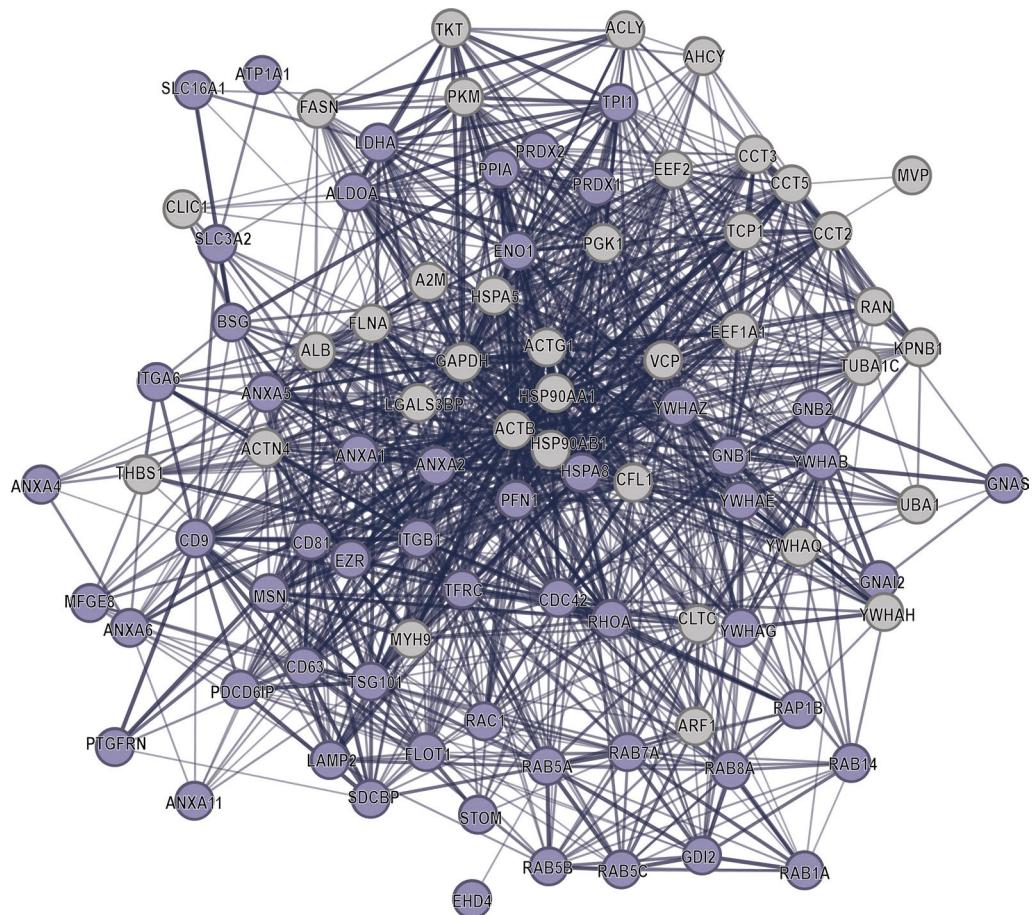
**Extended Data Fig. 3 | Nanoparticle Tracking Analysis (NTA) of crude sEVs preparations from human cancer cell lines and biofluids.** **a**, Particle size distribution and concentration in crude sEV (P100) preparations from the indicated cell lines and biofluids, analyzed by a NanoSight NTA system. Histograms display the average (black line) of three measurements ( $n = 3$  technical replicates), whereas the red shaded area around the curve represents the standard error of the mean (SEM). Summary statistics for mean, mode, and particle concentration are shown. Biological replicates are indicated with an asterisk (\*). **b**, Box plots summarizing the distribution of particle concentration

(particles/mL) and particle size (mean and mode) across all samples ( $n = 23$  biologically independent samples, comprising 20 cell line preparations, 2 serum samples, and 1 urine sample). Box plot elements: center lines indicate medians; box limits represent the 25<sup>th</sup> and 75<sup>th</sup> percentiles; whiskers extend to 1.5 times the interquartile range. **c**, Scatterplots showing the Spearman correlation ( $\rho$ ) between particle concentration and either the number of proteins identified (top) or the total protein abundance (iBAQ, bottom) for all proteins and for each designated class.  $P$  values were calculated using an unadjusted two-sided Student's *t*-test.

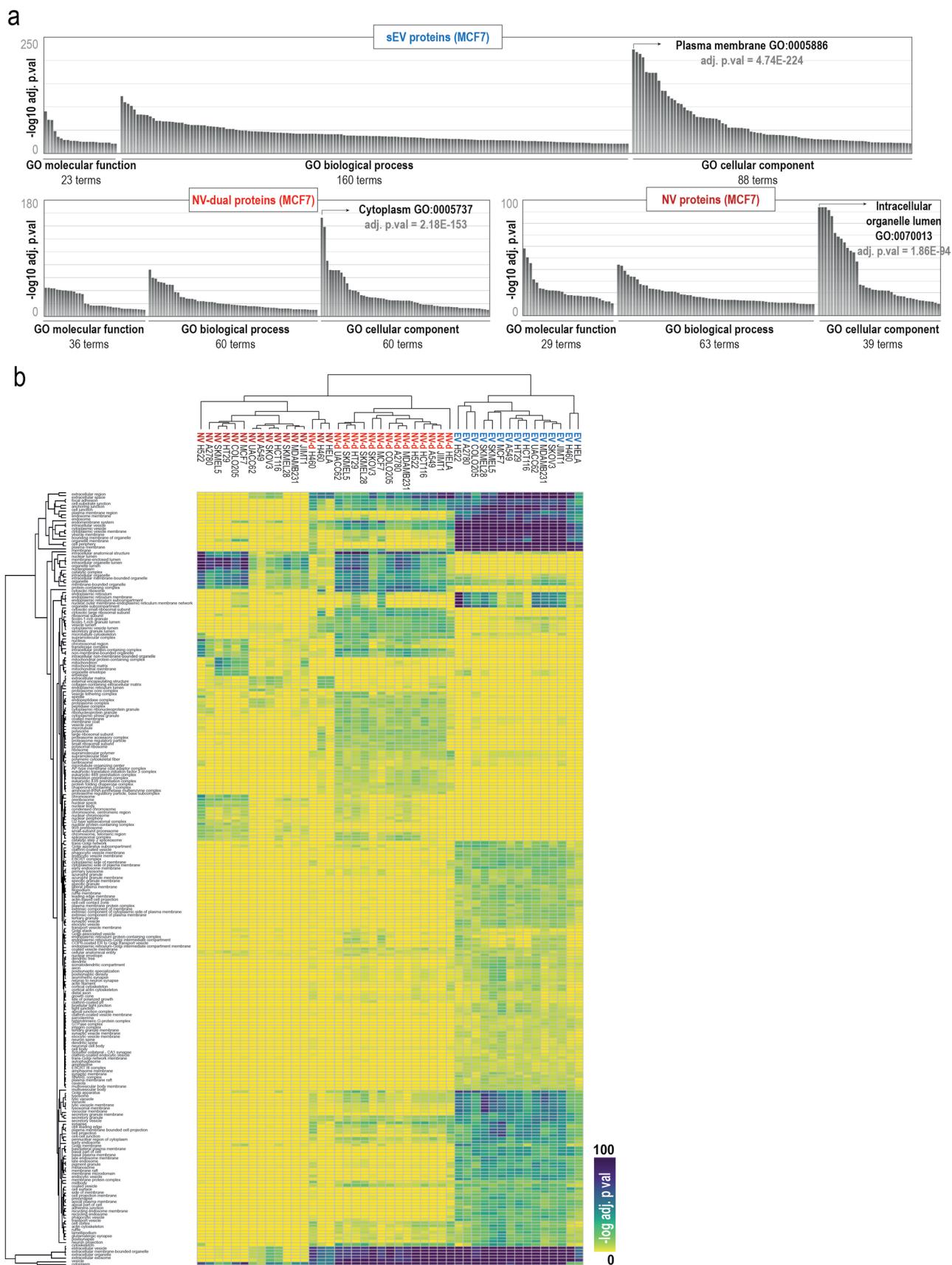
**a****b**

**Extended Data Fig. 4 | Characterization of P100 samples from cancer cell lines by cryo-EM and western blot.** **a**, Representative cryo-EM images of crude sEV (P100) preparations from the indicated cancer cell lines, showing the presence of intact, vesicle-like structures of the expected size and morphology. Cryo-EM analysis was performed once for each cell line, yielding consistent results across  $n = 5$  independent cancer cell lines. **b**, Western blot analysis of crude sEV (P100) preparations and corresponding whole-cell lysates for the indicated cell lines.

The blots demonstrate the enrichment of positive sEV markers (CD9, Syntenin-1, CD63) and the depletion of intracellular contaminant markers (Calreticulin, GOLGA2/GM130) in the sEV fractions. Ponceau S staining is shown as loading control. Dotted line represents different exposure time. The experiment was performed once for each cell line, demonstrating consistent marker enrichment profiles across  $n = 4$  independent cancer cell lines.



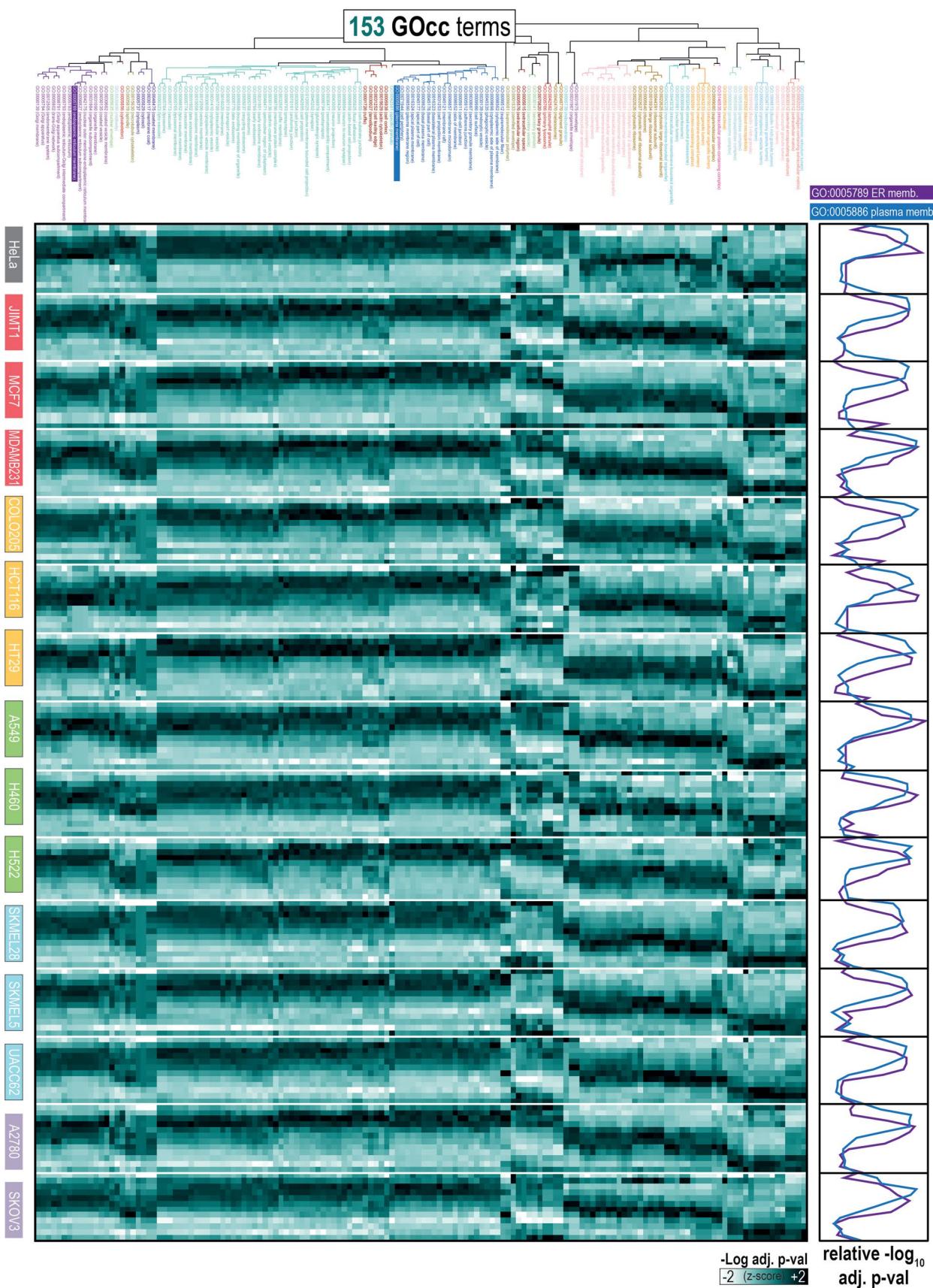
**Extended Data Fig. 5 | Re-assessment of the Exocarta top100 markers by DG-PCP.** The list of top100 most frequent proteins identified in EVs as annotated by Exocarta (<https://www.exocarta.org/>) are shown, displaying interactions as retrieved from StringDB. Proteins that were consistently classified in our data as genuine sEV proteins are displayed in blue.



Extended Data Fig. 6 | See next page for caption.

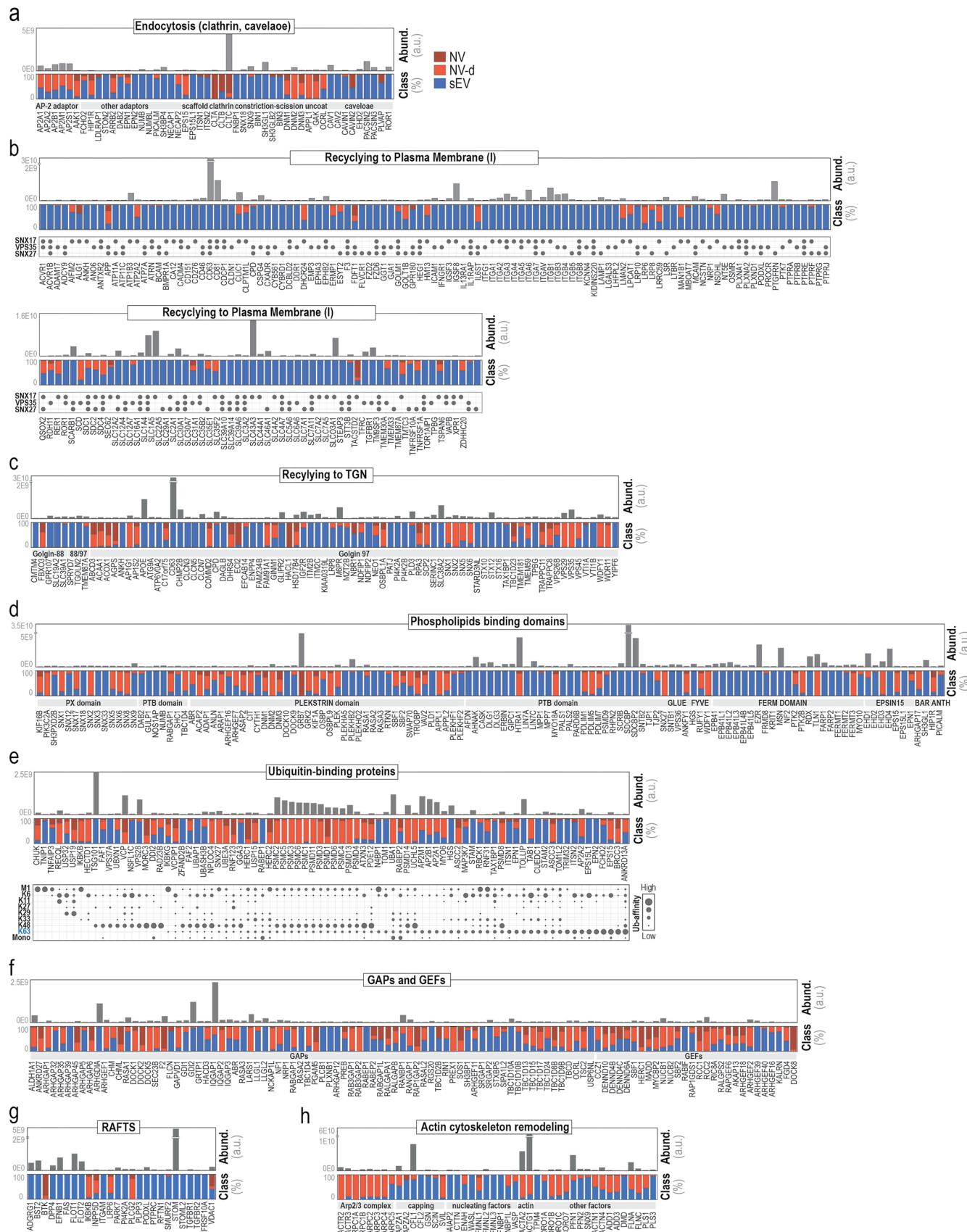
**Extended Data Fig. 6 | GO enrichment analyses of sEVs, NV-dual and NV proteins.** **a**, Proteins classified as sEVs, NV-dual and NV in the MCF7 cell line (used as a representative example) were analyzed using g:GOSt (<https://biit.cs.ut.ee/gprofiler/gost>). Gene-set over-representation was assessed with g:Profiler using a one-sided cumulative hypergeometric test (Fisher's exact test) for each term. *P* values were adjusted for multiple testing using the g:SCS

procedure (Set Counts and Sizes) implemented in g:Profiler. Terms enriched using GO biological process, GO molecular function and GO cellular component are sorted by significance. **b**, Unsupervised hierarchical clustering using the significance values of GO cellular components identified in the all protein classes of each cell line.



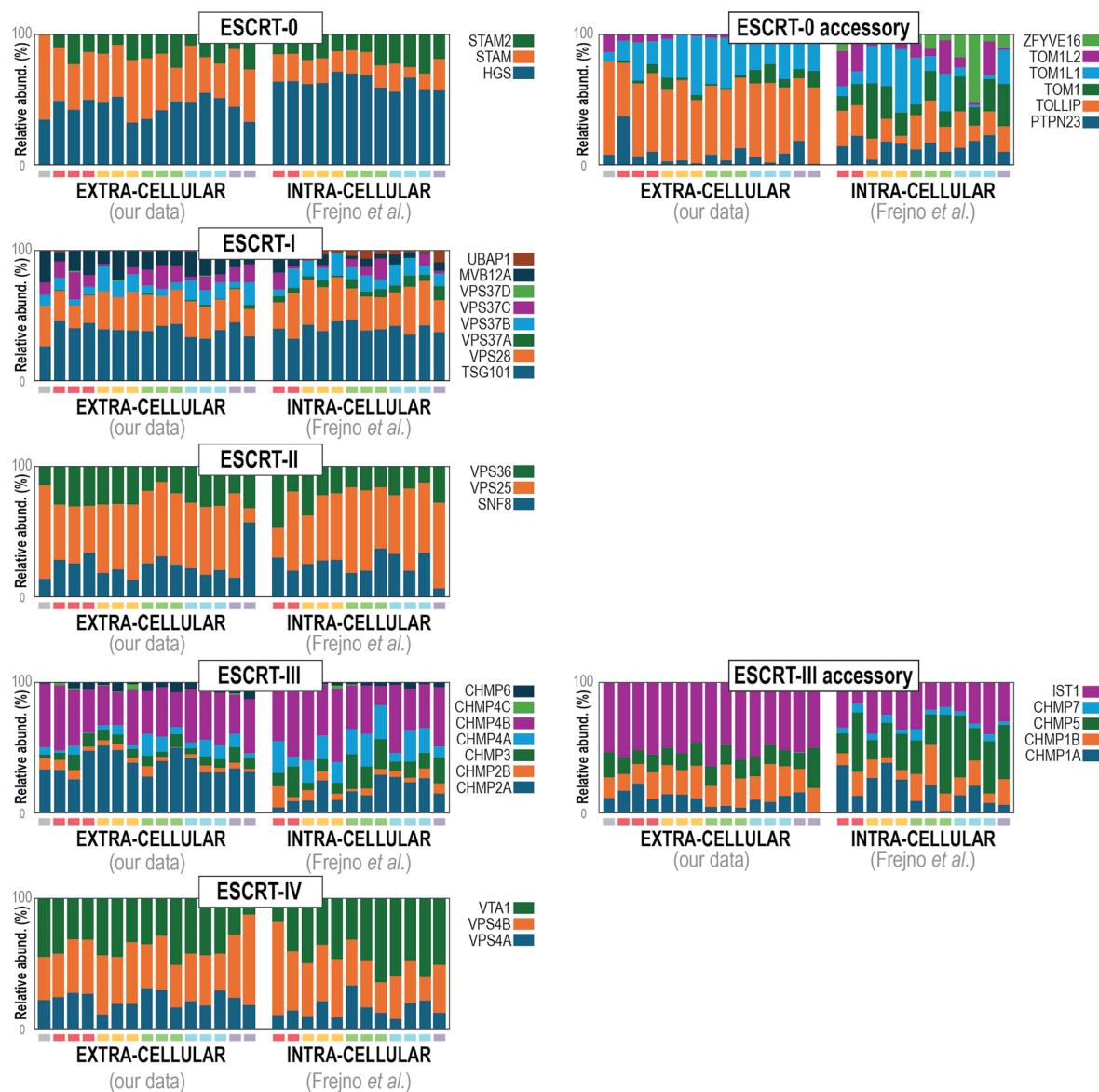
**Extended Data Fig. 7 | GO enrichment analyses across DG fractions.** Proteins identified in each DG fraction were ranked by descending abundance and analyzed by GO enrichment. For ranked gene lists ('ordered query' in g:Profiler), pathway enrichment was assessed using a one-sided minimum hypergeometric

test as implemented in g:GOSt.  $P$  values were adjusted for multiple testing using the g:SCS procedure (Set Counts and Sizes). The significance (-log adjusted  $p$ ) of GO cellular components terms along the DG was z-scored and visualized as a heatmap.

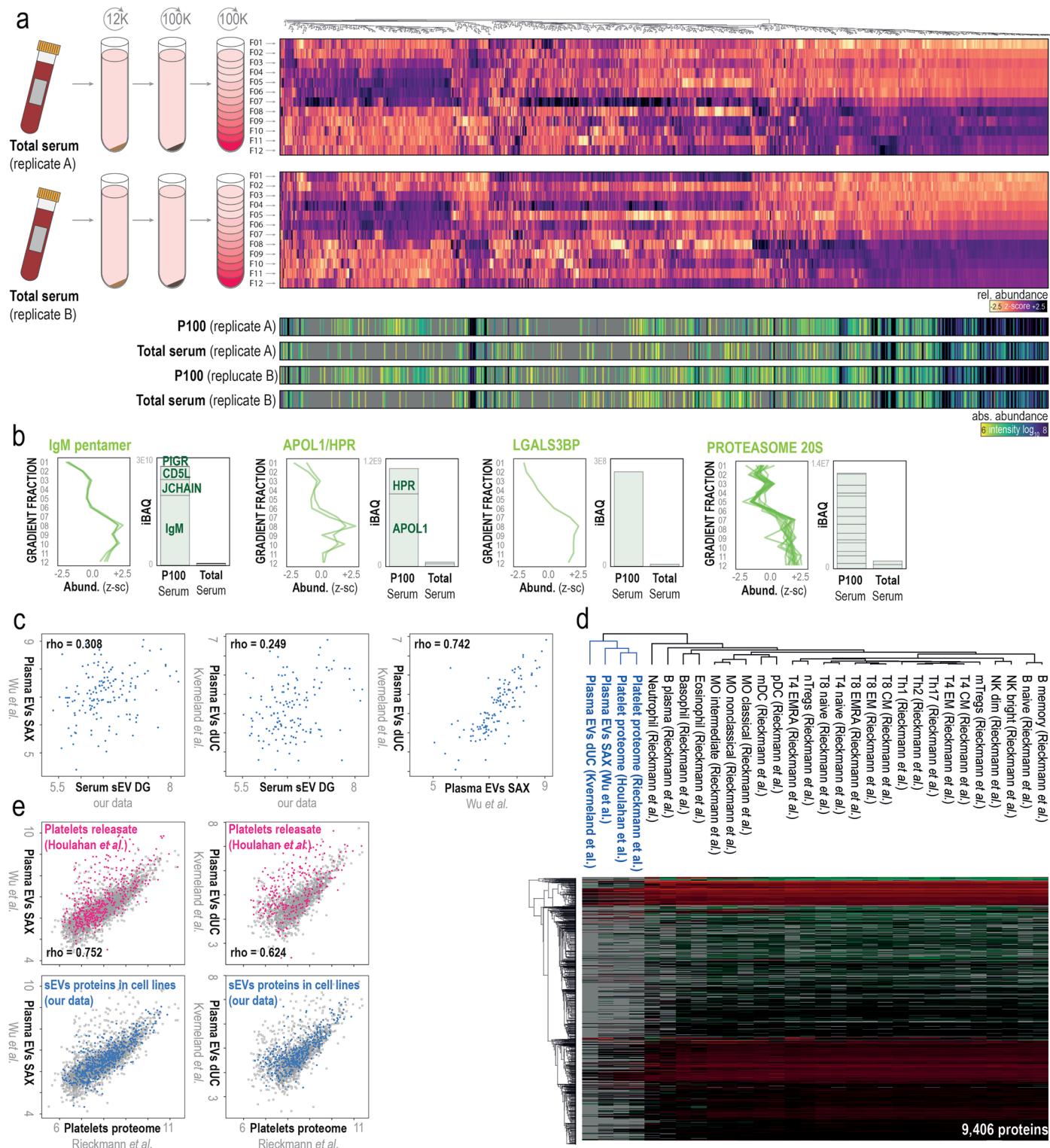


**Extended Data Fig. 8 | Classification of proteins involved in the biogenesis of sEVs.** a–h, Proteins are grouped based on relevant functional categories as discussed in the main text. For each protein, the measured abundance as sEV

(blue), NV-dual (red) and NV (brown) classes in all cell lines were summed and their relative proportions are represented. The absolute abundance is shown on top.



**Extended Data Fig. 9 | Stoichiometric analysis of ESCRT complexes.** For each ESCRT complex, their structural subunits are shown indicating the protein abundance (iBAQ) measured in each cell line. On the left, the values for extra-cellular fraction (our data) are shown. On the right, the corresponding protein levels for the intra-cellular fraction (from Frejno et al.<sup>10</sup>) are shown.



**Extended Data Fig. 10 | Analysis of serum sEVs.** **a**, P100 was purified from human serum using dUC and separated by DG into 12 fractions. Identified protein profiles are visualized in the heatmap. Below, absolute protein levels (for example iBAQ) from unfractionated serum and P100 are shown. Data from two independent biological replicates are shown. **b**, Macromolecular complexes are co-purified with sEVs from human serum using dUC. Bar charts display the abundance of the indicated protein complexes in crude sEVs (P100) and total serum. Their DG profiles are shown as a reference. **c**, The first two scatterplots compare the abundance of proteins categorized as genuine serum sEVs in our DG-PCP data to plasma EV datasets reported by Wu et al.<sup>68</sup> and Kverneland et al.<sup>69</sup> showing low correlations (Spearman's rho = 0.308 and rho = 0.249, respectively). The third scatterplot compares the Wu et al. and Kverneland et al. datasets,

revealing a strong correlation (rho = 0.742). **d**, Intra-cellular proteomes of major human hematopoietic cell populations from Rieckmann et al.<sup>67</sup> were compared with EV datasets from Wu et al. and Kverneland et al. An additional proteomic dataset of platelets from Houlahan et al.<sup>70</sup> was included in the analysis. The unsupervised HCL demonstrates the similarity between the two reported EV datasets and published platelet proteomes. **e**, The proteome of resting platelets, as reported by Rieckmann et al., is compared to the EV datasets from Wu et al. (left) and Kverneland et al. (right). Proteins released upon platelet activation (retrieved from Houlahan et al.) are highlighted in pink and show a clear enrichment in both studies compared to resting platelets. Proteins identified as sEVs in our data are highlighted in blue for reference.

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection Mass spectrometry data was acquired using an Exploris 480 (Thermo) using Xcalibur (Tune v2.11.0.3006) (Thermo)

Data analysis Raw data was processed with DIA-NN 1.8.1 and MaxQuant versions 2.0.1.0 and 2.4.2.0. Other analyses were done with Perseus v1.6.10.50

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The mass spectrometry proteomics data generated in this study have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifiers PXD062547 and PXD062477. Intracellular protein levels were obtained from the supplementary material of Frejno et al.. Total mRNA levels were retrieved from the Cancer Cell Line Encyclopedia at the Expression Atlas repository (<https://www.ebi.ac.uk/gxa/experiments/E-MTAB-2770/Results>). Source data are provided with this paper.

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	n/a
Reporting on race, ethnicity, or other socially relevant groupings	n/a
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size. The cohort of 15 distinct human cancer cell lines was selected to cover five diverse tissue origins (lung, breast, colon, melanoma, ovarian), ensuring sufficient biological variability to define a robust, universal sEV core proteome. This sample size strategy aligns with standard practices for large-scale exploratory proteomics. Additionally, human serum and urine were selected for analysis as they represent the most common and clinically relevant biofluids, demonstrating the workflow's applicability to complex biological matrices.
Data exclusions	No data was excluded throughout the manuscript except, as indicated in the manuscript, one density gradient fraction (F08 in Figure 6C) due to technical issue during LC-MS/MS
Replication	The density gradient profiles were measured in 15 independent cancer cell lines. Three of these cell lines were replicated as independent experiments. The density gradient upon Proteinase K degradation of crude sEVs was performed once (Figure 6C). The density gradient with spiked E. coli extract was performed once (Figure 6G). The density gradient from human serum was performed twice as independent biological replicates (Figure 8D). The density gradient from human urine was performed once (Figure 8E).
Randomization	Randomization was not applicable in the context of this study as each cell line could be analyzed independently. In experiments where a quantitative comparison was done (Figure 5C and 6C), density gradient fractions from control and treated conditions were alternated during LC-MS/MS to avoid technical biases.
Blinding	Blinding was not possible during sample preparation as knowledge of the experimental groups was necessary for the execution of the density gradient fractionation and enzymatic treatments. However, mass spectrometry data acquisition was performed using automated methods, and quantitative data analysis was conducted using computational algorithms (DIA-NN, Perseus).

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials &amp; experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

## Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

## Antibodies used

## Primary antibodies:

Abcam: anti-GM130 (1/1000; Cat# ab52649)  
 Abcam: anti-Calreticulin (1/1000; Cat# ab2907)  
 Abcam: anti-Syntenin (1/1000, Cat# ab133267)  
 Abcam: anti-CD63 (1/1000; Cat# ab193349)  
 Millipore: anti-CD9 (1/1000; Cat#CBL162)

## Secondary antibodies used:

Jackson ImmunoResearch: anti-Mouse-HRP (1/10000; Cat#115-035-062)  
 Jackson ImmunoResearch:anti-Rabbit-HRP (1/10000; Cat#111-035-045)

## Validation

All primary antibodies used in this study are commercially available and were validated by the respective manufacturers for use in Western Blotting (see manufacturer websites for the catalog numbers listed above).

## Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

## Cell line source(s)

Fifteen human cancer cell lines were used in this study. All cell lines were obtained as generous gifts from research groups at the Spanish National Cancer Research Centre (CNIO, Madrid, Spain). Female lines: HeLa (Cervix), JIMT1 (Breast), MCF7 (Breast), MDA-MB-231 (Breast), HT29 (Colon), SK-MEL-5 (Melanoma), UACC-62 (Melanoma), A2780 (Ovarian), SKOV-3 (Ovarian). Male lines: Colo205 (Colon), HCT116 (Colon), A549 (Lung), NCI-H460 (Lung), NCI-H522 (Lung), SK-MEL-28 (Melanoma).

## Authentication

Cells were not authenticated.

## Mycoplasma contamination

Cultures were routinely tested for mycoplasma contamination and consistently tested negative

Commonly misidentified lines  
(See [ICLAC](#) register)

None of the cell lines used in this study are listed in the ICLAC database of commonly misidentified cell lines.

## Plants

## Seed stocks

n/a

## Novel plant genotypes

n/a

## Authentication

n/a