

Defining the Reference Proteomes for Small Extracellular Vesicles and Non-Vesicular Components

Corresponding Author: Professor Javier Munoz

Parts of this Peer Review File have been redacted as indicated to remove third-party material.

Version 0:

Decision Letter:

*Please delete the link to your author homepage if you wish to forward this email to co-authors.

Dear Professor Munoz,

Your manuscript, "A High-Resolution Proteomic Atlas of Extracellular Vesicles and Non-Vesicular Particles", has now been seen by 2 referees, who are experts in EVs, proteomics (referee 1); and EVs, proteomics, biomarkers (referee 2). As you will see from their comments (attached below) they find this work of potential interest, but have raised substantial concerns, which in our view would need to be addressed with considerable revisions before we can consider publication in Nature Cell Biology.

Nature Cell Biology editors discuss the referee reports in detail within the editorial team, including the chief editor, to identify key referee points that should be addressed with priority, and requests that are overruled as being beyond the scope of the current study. To guide the scope of the revisions, I have listed these points below. We are committed to providing a fair and constructive peer-review process, so please feel free to contact me if you would like to discuss any of the referee comments further.

In particular, it would be essential to:

- 1) assess the physical parameters of the EVs
- 2) quantify and validate the EV markers
- 3) clarification and explanation of how proteins are grouped
- 4) All other referee concerns pertaining to strengthening existing data, providing controls, methodological details, clarifications and textual changes, should also be addressed.
- 5) Finally please pay close attention to our guidelines on statistical and methodological reporting (listed below) as failure to do so may delay the reconsideration of the revised manuscript. In particular please provide:
 - a Supplementary Figure including unprocessed images of all gels/blots in the form of a multi-page pdf file. Please ensure that blots/gels are labeled and the sections presented in the figures are clearly indicated.
 - a Supplementary Table including all numerical source data in Excel format, with data for different figures provided as different sheets within a single Excel file. The file should include source data giving rise to graphical representations and statistical descriptions in the paper and for all instances where the figures present representative experiments of multiple independent repeats, the source data of all repeats should be provided.

We would be happy to consider a revised manuscript that would satisfactorily address these points, unless a similar paper is published elsewhere, or is accepted for publication in Nature Cell Biology in the meantime.

When revising the manuscript please:

- ensure that it conforms to our format instructions and publication policies (see below and www.nature.com/nature/authors/).
- provide a point-by-point rebuttal to the full referee reports verbatim, as provided at the end of this letter.
- for any revision that includes light microscopy data, we ask our authors to please include a completed light microscopy reporting table https://www.nature.com/documents/Light_microscopy_reporting_table.xlsx to ensure the methods are described thoroughly. The table will be available to reviewers and ultimately published should the manuscript be accepted at the journal.

- provide the completed Editorial Policy Checklist (found here https://www.nature.com/authors/policies/Policy.pdf), and Reporting Summary (found here https://www.nature.com/authors/policies/ReportingSummary.pdf). This is essential for reconsideration of the manuscript and these documents will be available to editors and referees in the event of peer review. For more information see http://www.nature.com/authors/policies/availability.html or contact me.

Nature Cell Biology is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit please visit www.springernature.com/orcid.

Please submit the revised manuscript files and the point-by-point rebuttal to the referee comments using this link:

Link Redacted

*This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We would like to receive a revised submission within six months. We would be happy to consider a revision even after this timeframe, however if the resubmission deadline is missed and the paper is eventually published, the submission date will be the date when the revised manuscript was received.

We hope that you will find our referees' comments, and editorial guidance helpful. Please do not hesitate to contact me if there is anything you would like to discuss.

Best wishes,

Angela Parrish

Angela R Parrish, PhD
Locum Senior Editor
Nature Cell Biology

Reviewers' Comments:

Reviewer #1 (Remarks to the Author):

The manuscript by Morales-Sanfrutos et al. presents a comprehensive proteomic atlas of extracellular vesicles (EV) using a density gradient and protein correlation profiling (DG-PCP) approach. This research holds the potential to be a landmark study in the EV field by addressing the critical challenge of defining the "true" proteome of EV. However, I believe its impact and long-term reliability could be significantly enhanced by addressing two key areas related to experimental validation and data interpretation. The evidence presented does not sufficiently support the main claims, "A High-Resolution Proteomic Atlas of Extracellular Vesicles".

Major Points for Consideration:

1. Strengthening the Foundational Evidence for EV Separation and Reproducibility.

The manuscript relies heavily on the downstream proteomic profiles to validate the separation, with limited upstream characterization. To ensure the quality and reproducibility of the input material, I strongly recommend including more extensive validation in line with MISEV2023 guidelines (Welsh, J.A. et al., J Extracell Vesicles, 2024). The manuscript currently lacks quantitative data on the physical properties of the isolated particles.

1-1. The authors should provide particle size distribution and concentration data for the main fractions collected from the iodixanol density gradient via NTA.

1-2. There is no direct optical evidence in the manuscript to confirm the presence of vesicles. Provide representative Transmission Electron Microscopy (TEM) images of samples from the corresponding sEV-rich fractions to confirm the presence and morphology of vesicles.

Incorporating these data would provide unequivocal, independent evidence for the quality of the isolated material and strengthen the interpretation of the proteomic results.

2. Addressing the Conceptual Challenge of Proteomic Depth vs. Biological Specificity.

In total 8,831 human proteins were identified from the enriched sEV fractions of 15 human cancer cell lines. However, this impressive depth raises a crucial conceptual question for the field. As analytical sensitivity increases identifying nearly the full cellular proteome in EV samples, the risk of detecting proteins that are not selectively sorted into EVs—but rather represent low-level contaminants or stochastically included cellular proteins—also grows.

This raises a key question: At what point can a protein confidently be considered a bona fide EV cargo, as opposed to a background constituent? To preserve the biological significance of EV protein cargo profiles, I strongly encourage the authors to discuss this issue explicitly.

This manuscript has high potential to become a robust and reliable benchmark for the field. Their science is sound, and the approach is

innovative.

Therefore, I recommend this manuscript for publication after Major Revisions to address these points—particularly by strengthening the upstream validation and providing a more discussion on the identification of EV proteome and interpretation of high-depth proteomic data. I look forward to seeing the revised version.

Reviewer #2 (Remarks to the Author):

Overall comments to Author:

Due to the inherent limitations of current EV isolation methods, it remains technically challenging to obtain highly pure and well-defined subpopulations of extracellular vesicles (EVs) from complex biological samples. Conventional purification techniques frequently co-isolate non-vesicular contaminants, such as lipoproteins and abundant plasma proteins, which interfere with downstream analyses and confound the biological interpretation of EV functions. Furthermore, commonly used EV markers are insufficient to accurately assess the purity of EV preparations, adding further uncertainty to current studies. In this context, the authors present a relatively comprehensive proteomic atlas of small EVs (sEVs) derived from multiple cell lines, and importantly, extend this analysis to sEVs isolated from serum and urine. By applying a protein correlation profiling approach, the study aims to reclassify and enrich for bona fide sEV cargo proteins, distinguishing them from co-purifying background proteins. This resource represents a valuable contribution to the EV field, offering improved resolution in defining sEV composition and laying a foundation for more reliable functional and biomarker studies. While the study's methodological novelty is somewhat limited, the breadth and quality of the dataset are highly commendable. I have several suggestions that could help improve the overall clarity, presentation, and scientific rigor of the manuscript.

Summary of the Key Results:

This study presents a highly detailed proteomic characterization of small extracellular vesicles (sEVs), leveraging a combined density gradient and protein correlation profiling (DG-PCP) strategy to systematically define sEV protein cargo with high specificity. Key findings include the predominance of plasma membrane and membrane-associated proteins in sEVs, the limited presence of cytosolic proteins, and the hypothesis that most cytosolic proteins in sEV preparations are contaminants. Additionally, the authors demonstrate that urine-derived sEVs provide superior depth and tissue specificity for biomarker discovery compared to serum-derived sEVs. The presence and nature of ubiquitin within sEVs are also explored, leading to the suggestion that free, rather than conjugated, ubiquitin may dominate.

Originality and Significance:

The primary contribution of this study lies in the systematic application of established techniques to generate one of the most comprehensive small extracellular vesicle (sEV) proteomic datasets to date, which represents a valuable resource for the field. However, from a methodological standpoint, the work largely builds upon and refines existing approaches—such as density gradient separation and protein correlation profiling—rather than introducing fundamentally novel techniques. As such, while the dataset is impressive in scale and quality, the methodological innovation is relatively limited.

Data & Methodology: Validity of Approach, Quality of Data, Quality of Presentation:

The DG-PCP methodology is well-justified and appears sound for distinguishing bona fide sEV cargo from co-isolated proteins. The discussion implies a high level of technical rigor, including validation through correlation analysis with known markers. However, further clarification in the main text regarding the reproducibility across biological replicates and validation of the novel ER-derived vesicles would enhance robustness.

Appropriate Use of Statistics and Treatment of Uncertainties:

While the discussion references correlation and quantitative relationships (e.g., between intracellular and sEV protein levels), it does not clearly describe the statistical treatment or uncertainty ranges. A more explicit mention of statistical thresholds used for defining co-elution or enrichment would increase transparency. For instance, describing false discovery rates, confidence intervals, or methods to assess significance of differential protein abundance would strengthen confidence in the interpretations.

Conclusions: Robustness, Validity, Reliability:

The conclusions are generally well-supported by the data and thoughtfully contextualized. The authors convincingly argue for a reconsideration of cytosolic proteins as sEV components and propose a plausible model of stochastic cargo loading. However, the interpretation of ER-derived vesicles is speculative and would benefit from additional experimental support. Likewise, while the authors downplay the role of ubiquitinated substrates in sEVs, this conclusion is limited by the resolution of bulk proteomics and should be qualified accordingly.

References: Appropriate Credit to Previous Work?

The manuscript could do more to cite and discuss key prior contributions, including foundational studies that have examined EV heterogeneity, cargo sorting mechanisms, and EV proteome annotation. Some claims, such as the stochastic model or the role of HSPA8 in secretory autophagy, would benefit from being explicitly linked to existing literature.

Clarity and Context: Lucidity of Abstract/Summary, Appropriateness of Abstract, Introduction and Conclusions:

The discussion is scientifically rich but occasionally too dense, particularly when detailing ubiquitin chain types and sorting mechanisms. Simplifying and streamlining these sections would improve accessibility. The overall context is well-framed, though the implications for the broader field of EV-based biomarker discovery could be highlighted more clearly in the conclusions.

Suggested Improvements: Experiments, Data for Possible Revision:

1. Limited Display of EV Markers in Figure 1A

There are a wide variety of well-established EV markers reported in the literature, including but not limited to CD9, CD63, CD81, ALIX, TSG101, and syntenin. However, Figure 1A presents only three such markers. Given that more than a dozen EV-associated proteins are known to be robustly detectable in proteomic datasets, it is important to clarify whether these additional canonical markers were identified in the current study and how they were distributed between the sEV and NV groups. This information is essential for evaluating the accuracy of the proposed classification scheme.

2. Ambiguous Classification of Proteins into sEV and NV Groups

Throughout the manuscript, heatmaps classify proteins strictly into “sEV” and “NV” groups. However, such binary categorization may oversimplify the actual distribution. Proteins enriched in sEVs are not necessarily absent from the NV group, and vice versa. A more

appropriate approach might be to classify different density fractions as sEV-enriched or NV-enriched, rather than assigning proteins to exclusive categories. This would better reflect the continuous nature of protein distribution across fractions.

3. Unclear Subgrouping in Figure 1B

In Figure 1B, proteins are further subclassified into several subgroups. It is unclear whether these subgroups are mutually exclusive or hierarchical. The authors should provide a clear rationale and formal definitions for each subgroup to enhance reproducibility and interpretation.

4. Inconsistency in Supplementary Figure S1B

Supplementary Figure S1B states that three gradient formulations were used; however, only two are shown. Additionally, it is unclear whether technical replicates were performed for each formulation. Clarifying this would strengthen the methodological transparency of the study.

5. Ambiguity in Supplementary Figure S1C

The terms "top" and "bottom" in Supplementary Figure S1C are not clearly defined. Do they refer to fractions from the top and bottom of the density gradient? The authors should clarify these labels to ensure accurate interpretation of the figure.

6. Insufficient Characterization of Vesicles vs Non-vesicles

The methods for distinguishing vesicular and non-vesicular components rely heavily on a limited number of marker proteins. There is no evidence presented regarding particle size distribution, which is a critical parameter in EV characterization. Techniques such as nanoparticle tracking analysis (NTA) or transmission electron microscopy (TEM) should be included or more thoroughly discussed.

7. Improper Reagent Notation

Some reagents are not properly formatted. For example, in the Methods section, "Na₂CO₃" should be correctly written as "Na₂CO₃." Such issues, though minor, detract from the manuscript's professionalism and should be corrected throughout.

8. Missing Validation Data for a Resource Article

As a resource-type article, the manuscript should present comprehensive and rigorously validated datasets. However, key validation steps such as Western blotting (WB) for both positive and negative EV markers are missing. Although the authors mention NTA analysis, no corresponding data or figures are shown. Inclusion of such validation is essential to confirm the integrity and purity of the EV preparations.

9. Title Scope and Sample Diversity

The manuscript's title refers to a "proteomic atlas," suggesting comprehensive coverage. However, only cell line-derived EVs, along with serum and urine EVs, are included. Previous studies have highlighted that EVs from other biofluids—such as plasma, saliva, and tissue interstitial fluid—possess unique proteomic profiles. To justify the term "atlas," the authors should either broaden the sample scope or revise the title to reflect the current dataset limitations.

10. Variability in Protein Identification Across Cell Lines (Figure 2B)

Figure 2B shows substantial differences in the number of sEV proteins identified across different cell lines. The authors should comment on whether this variability is due to inherent differences in EV secretion levels between cell types. Including NTA data to demonstrate particle yield per cell line could help clarify whether secretion volume or protein diversity is the primary factor.

11. Lack of Clarity in sEV Protein Classification Methodology

The manuscript repeatedly mentions that 1,499 proteins were consistently classified as sEV proteins. However, it lacks a detailed description of the classification algorithm or criteria used for this categorization. This omission makes it difficult for readers to assess the robustness of the results or apply the same approach to their own data.

12. Unclear Interpretation of Dual and NV Proteins

The statement that "DUAL and NV proteins appeared, to some extent, interchangeable in their classification" is concerning, as it may indicate instability in the classification method. Without a clear description of the classification criteria or computational workflow, this undermines confidence in the data segmentation. If custom scripts were used for classification, the authors should deposit the code in a publicly accessible repository, as this would significantly enhance reproducibility.

13. Lack of Guidance for Filtering Non-sEV Proteins in Practical Applications

Although the authors successfully establish a reference dataset that distinguishes between sEV-enriched and non-vesicular proteins, they do not provide practical recommendations for how to exclude contaminating proteins in downstream quantitative studies. In real-world EV research, distinguishing and quantifying authentic EV cargo is crucial. Merely labeling proteins as "sEV" or "non-sEV" is insufficient unless it informs how researchers can apply these distinctions in biomarker discovery or mechanistic studies.

GUIDELINES FOR MANUSCRIPT SUBMISSION TO NATURE CELL BIOLOGY

READABILITY OF MANUSCRIPTS – Nature Cell Biology is read by cell biologists from diverse backgrounds, many of whom are not native English speakers. Authors should aim to communicate their findings clearly, explaining technical jargon that might be unfamiliar to non-specialists, and avoiding non-standard abbreviations. Titles and abstracts should concisely communicate the main findings of the study, and the background, rationale, results and conclusions should be clearly explained in the manuscript in a manner accessible to a broad cell biology audience. Nature Cell Biology uses British spelling.

MANUSCRIPT FORMAT – please follow the guidelines listed in our Guide to Authors regarding manuscript formats at Nature Cell Biology.

TITLE – should be no more than 100 characters including spaces, without punctuation and avoiding technical terms, abbreviations, and active verbs..

AUTHOR NAMES – should be given in full.

AUTHOR AFFILIATIONS – should be denoted with numerical superscripts (not symbols) preceding the names. Full addresses should be included, with US states in full and providing zip/post codes. The corresponding author is denoted by: "Correspondence should be addressed to [initials]."

ABSTRACT AND MAIN TEXT – please follow the guidelines that are specific to the format of your manuscript, as listed in our Guide to Authors (http://www.nature.com/ncb/pdf/ncb_gta.pdf) Briefly, Nature Cell Biology Articles, Resources and Technical Reports have 3500 words, including a 150 word abstract, and the main text is subdivided in Introduction, Results, and Discussion sections. Nature Cell Biology Letters have up to 2500 words, including a 180 word introductory paragraph (abstract), and the text is not subdivided in sections.

ACKNOWLEDGEMENTS – should be kept brief. Professional titles and affiliations are unnecessary. Grant numbers can be listed.

AUTHOR CONTRIBUTIONS – must be included after the Acknowledgements, detailing the contributions of each author to the paper (e.g. experimental work, project planning, data analysis etc.). Each author should be listed by his/her initials.

FINANCIAL AND NON-FINANCIAL COMPETING INTERESTS – the authors must include one of three declarations: (1) that they have no financial and non-financial competing interests; (2) that they have financial and non-financial competing interests; or (3) that they decline to respond, after the Author Contributions section. This statement will be published with the article, and in cases where financial and non-financial competing interests are declared, these will be itemized in a web supplement to the article. For further details please see <https://www.nature.com/licenceforms/nrg/competing-interests.pdf>.

REFERENCES – are limited to a total of 70 for Articles, Resources, Technical Reports; and 40 for Letters. This includes references in the main text and Methods combined. References must be numbered sequentially as they appear in the main text, tables and figure legends and Methods and must follow the precise style of Nature Cell Biology references. References only cited in the Methods should be numbered consecutively following the last reference cited in the main text. References only associated with Supplementary Information (e.g. in supplementary legends) do not count toward the total reference limit and do not need to be cited in numerical continuity with references in the main text. Only published papers can be cited, and each publication cited should be included in the numbered reference list, which should include the manuscript titles. Footnotes are not permitted.

METHODS – Nature Cell Biology publishes methods online. The methods section should be provided as a separate Word document, which will be copyedited and appended to the manuscript PDF, and incorporated within the HTML format of the paper.

Methods should be written concisely, but should contain all elements necessary to allow interpretation and replication of the results. As a guideline, Methods sections typically do not exceed 3,000 words. The Methods should be divided into subsections listing reagents and techniques. When citing previous methods, accurate references should be provided and any alterations should be noted. Information must be provided about: antibody dilutions, company names, catalogue numbers and clone numbers for monoclonal antibodies; sequences of RNAi and cDNA probes/primers or company names and catalogue numbers if reagents are commercial; cell line names, sources and information on cell line identity and authentication. Animal studies and experiments involving human subjects must be reported in detail, identifying the committees approving the protocols. For studies involving human subjects/samples, a statement must be included confirming that informed consent was obtained. Statistical analyses and information on the reproducibility of experimental results should be provided in a section titled "Statistics and Reproducibility".

All Nature Cell Biology manuscripts submitted on or after March 21 2016 must include a Data availability statement at the end of the Methods section. For Springer Nature policies on data availability see <http://www.nature.com/authors/policies/availability.html>; for more information on this particular policy see <http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf>. The Data availability statement should include:

- Accession codes for primary datasets (generated during the study under consideration and designated as "primary accessions") and secondary datasets (published datasets reanalysed during the study under consideration, designated as "referenced accessions"). For primary accessions data should be made public to coincide with publication of the manuscript. A list of data types for which submission to community-endorsed public repositories is mandated (including sequence, structure, microarray, deep sequencing data) can be found here <http://www.nature.com/authors/policies/availability.html#data>.
- Unique identifiers (accession codes, DOIs or other unique persistent identifier) and hyperlinks for datasets deposited in an approved repository, but for which data deposition is not mandated (see here for details <http://www.nature.com/sdata/data-policies/repositories>).
- At a minimum, please include a statement confirming that all relevant data are available from the authors, and/or are included with the manuscript (e.g. as source data or supplementary information), listing which data are included (e.g. by figure panels and data types) and mentioning any restrictions on availability.
- If a dataset has a Digital Object Identifier (DOI) as its unique identifier, we strongly encourage including this in the Reference list and citing the dataset in the Methods.

We recommend that you upload the step-by-step protocols used in this manuscript to protocols.io. More details can found at <https://www.protocols.io/help/publish-articles>.

DISPLAY ITEMS – main display items are limited to 6-8 main figures and/or main tables for Articles, Resources, Technical Reports; and 5 main figures and/or main tables for Letters. For Supplementary Information see below.

FIGURES – Colour figure publication costs \$600 for the first, and \$300 for each subsequent colour figure. All panels of a multi-panel figure must be logically connected and arranged as they would appear in the final version. Unnecessary figures and figure panels should be avoided (e.g. data presented in small tables could be stated briefly in the text instead).

All imaging data should be accompanied by scale bars, which should be defined in the legend.

Cropped images of gels/blots are acceptable, but need to be accompanied by size markers, and to retain visible background signal within the linear range (i.e. should not be saturated). The boundaries of panels with low background have to be demarked with black lines. Splicing of panels should only be considered if unavoidable, and must be clearly marked on the figure, and noted in the legend with a statement on whether the samples were obtained and processed simultaneously. Quantitative comparisons between samples on different gels/blots are discouraged; if this is unavoidable, it should only be performed for samples derived from the same experiment with gels/blots were processed in parallel, which needs to be stated in the legend.

Figures should be provided at approximately the size that they are to be printed at (single column is 86 mm, double column is 170 mm) and should not exceed an A4 page (8.5 x 11"). Reduction to the scale that will be used on the page is not necessary, but multi-panel figures should be sized so that the whole figure can be reduced by the same amount at the smallest size at which essential details in each panel are visible. In the interest of our colour-blind readers we ask that you avoid using red and green for contrast in figures. Replacing red with magenta and green with turquoise are two possible colour-safe alternatives. Lines with widths of less than 1 point should be avoided. Sans serif typefaces, such as Helvetica (preferred) or Arial should be used. All text that forms part of a figure should be rewritable and removable.

We accept files from the following graphics packages in either PC or Macintosh format:

- For line art, graphs, charts and schematics we prefer Adobe Illustrator (.AI), Encapsulated PostScript (.EPS) or Portable Document Format (.PDF). Files should be saved or exported as such directly from the application in which they were made, to allow us to restyle them according to our journal house style.

- We accept PowerPoint (.PPT) files if they are fully editable. However, please refrain from adding PowerPoint graphical effects to objects, as this results in them outputting poor quality raster art. Text used for PowerPoint figures should be Helvetica (preferred) or Arial.

- We do not recommend using Adobe Photoshop for designing figures, but we can accept Photoshop generated (.PSD or .TIFF) files only if each element included in the figure (text, labels, pictures, graphs, arrows and scale bars) are on separate layers. All text should be editable in 'type layers' and line-art such as graphs and other simple schematics should be preserved and embedded within 'vector smart objects' - not flattened raster/bitmap graphics.

- Some programs can generate Postscript by 'printing to file' (found in the Print dialogue). If using an application not listed above, save the file in PostScript format or email our Art Editor, Allen Beattie for advice (a.beattie@nature.com).

Regardless of format, all figures must be vector graphic compatible files, not supplied in a flattened raster/bitmap graphics format, but should be fully editable, allowing us to highlight/copy/paste all text and move individual parts of the figures (i.e. arrows, lines, x and y axes, graphs, tick marks, scale bars etc.). The only parts of the figure that should be in pixel raster/bitmap format are photographic images or 3D rendered graphics/complex technical illustrations.

All placed images (i.e. a photo incorporated into a figure) should be on a separate layer and independent from any superimposed scale bars or text. Individual photographic images must be a minimum of 300+ DPI (at actual size) or kept constant from the original picture acquisition and not decreased in resolution post image acquisition. All colour artwork should be RGB format.

FIGURE LEGENDS – must not exceed 350 words for each figure to allow fit on a single printed NCB page together with the figure. They must include a brief title for the whole figure, and short descriptions of each panel with definitions of the symbols used, but without detailing methodology.

TABLES – main tables should be provided as individual Word files, together with a brief title and legend. For supplementary tables see below.

SUPPLEMENTARY INFORMATION – Supplementary information is material directly relevant to the conclusion of a paper, but which cannot be included in the printed version in order to keep the manuscript concise and accessible to the general reader. Supplementary information is an integral part of a Nature Cell Biology publication and should be prepared and presented with as much care as the main display item, but it must not include non-essential data or text, which may be removed at the editor's discretion. All supplementary material is fully peer-reviewed and published online as part of the HTML version of the manuscript. Supplementary Figures and Supplementary Notes are appended at the end of the main PDF of the published manuscript.

Supplementary items should relate to a main text figure, wherever possible, and should be mentioned sequentially in the main manuscript, designated as Supplementary Figure, Table, Video, or Note, and numbered continuously (e.g. Supplementary Figure 1, Supplementary Figure 2, Supplementary Table 1, Supplementary Table 2 etc.).

Unprocessed scans of all key data generated through electrophoretic separation techniques need to be presented in a supplementary figure that should be labelled and numbered as the final supplementary figure, and should be mentioned in every relevant figure legend. This figure does not count towards the total number of figures and is the only figure that can be displayed over multiple pages, but should be provided as a single file, in PDF or TIFF format. Data in this figure can be displayed in a relatively informal style, but size markers and the figures panels corresponding to the presented data must be indicated.

The total number of Supplementary Figures (not including the “unprocessed scans” Supplementary Figure) should not exceed the number of main display items (figures and/or tables (see our Guide to Authors and March 2012 editorial <http://www.nature.com/ncb/authors/submit/index.html#suppinfo>; <http://www.nature.com/ncb/journal/v14/n3/index.html#ed>). No restrictions apply to Supplementary Tables or Videos, but we advise authors to be selective in including supplemental data.

Each Supplementary Figure should be provided as a single page and as an individual file in one of our accepted figure formats and should be presented according to our figure guidelines (see above). Supplementary Tables should be provided as individual Excel files. Supplementary Videos should be provided as .avi or .mov files up to 50 MB in size. Supplementary Figures, Tables and Videos must be accompanied by a separate Word document including titles and legends.

GUIDELINES FOR EXPERIMENTAL AND STATISTICAL REPORTING

REPORTING REQUIREMENTS – To improve the quality of methods and statistics reporting in our papers we have recently revised the reporting checklist we introduced in 2013. We are now asking all life sciences authors to complete two items: an Editorial Policy Checklist (found here <https://www.nature.com/authors/policies/Policy.pdf>) that verifies compliance with all required editorial policies and a reporting summary (found here <https://www.nature.com/authors/policies/ReportingSummary.pdf>) that collects information on experimental design and reagents. These documents are available to referees to aid the evaluation of the manuscript. Please note that these forms are dynamic ‘smart pdfs’ and must therefore be downloaded and completed in Adobe Reader. We will then flatten them for ease of use by the reviewers. If you would like to reference the guidance text as you complete the template, please access these flattened versions at <http://www.nature.com/authors/policies/availability.html>.

STATISTICS – Wherever statistics have been derived the legend needs to provide the n number (i.e. the sample size used to derive statistics) as a precise value (not a range), and define what this value represents. Error bars need to be defined in the legends (e.g. SD, SEM) together with a measure of centre (e.g. mean, median). Box plots need to be defined in terms of minima, maxima, centre, and percentiles. Ranges are more appropriate than standard errors for small data sets. Wherever statistical significance has been derived, precise p values need to be provided and the statistical test used needs to be stated in the legend. Statistics such as error bars must not be derived from $n < 3$. For sample sizes of $n < 5$ please plot the individual data points rather than providing bar graphs. Deriving statistics from technical replicate samples, rather than biological replicates is strongly discouraged. Wherever statistical significance has been derived, precise p values need to be provided and the statistical test stated in the legend.

Information on how many times each experiment was repeated independently with similar results needs to be provided in the legends and/or Methods for all experiments, and in particular wherever representative experiments are shown.

We strongly recommend the presentation of source data for graphical and statistical analyses as a separate Supplementary Table, and request that source data for all independent repeats are provided when representative experiments of multiple independent repeats, or averages of two independent experiments are presented. This supplementary table should be in Excel format, with data for different figures provided as different sheets within a single Excel file. It should be labelled and numbered as one of the supplementary tables, titled “Statistics Source Data”, and mentioned in all relevant figure legends.

----- Please don't hesitate to contact NCB@nature.com should you have queries about any of the above requirements -----

Version 1:

Decision Letter:

Our ref: NCB-RS57743A

26th November 2025

Dear Dr. Munoz,

Thank you for submitting your revised manuscript “Reference Proteomes of Small Extracellular Vesicles and Non-Vesicular Components” (NCB-RS57743A). It has now been seen by the original referees and their comments are below. The reviewers find that the paper has improved in revision, and therefore we'll be happy in principle to publish it in Nature Cell Biology, pending minor revisions to satisfy the referees' final requests and to comply with our editorial and formatting guidelines.

If the current version of your manuscript is in a PDF format, please email us a copy of the file in an editable format (Microsoft Word or LaTeX)-- we can not proceed with PDFs at this stage.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements in about a week. Please do not upload the final materials and make any revisions until you receive this additional information from us.

Thank you again for your interest in Nature Cell Biology Please do not hesitate to contact me if you have any questions.

Sincerely,

Angela R Parrish, PhD
Locum Senior Editor
Nature Cell Biology

Reviewer #2 (Remarks to the Author):

The revised manuscript represents a substantial improvement over the previous version. The authors have addressed most of the previous concerns with detailed explanations and additional analyses, which considerably strengthen the study. The improved interpretation of the molecular composition of sEVs and the clarification of non-vesicular components greatly enhance the manuscript's rigor and impact.

However, one issue remains unresolved. The manuscript still lacks clarification regarding the biological nature of the "unclassified (UNC)" protein group, which accounts for approximately 12% of the identified proteins. Since all analyzed proteins were derived from the P100 fraction—typically considered enriched in sEVs—it is important to specify what these UNC proteins represent. Are they low-abundance or partially associated sEV proteins, non-vesicular contaminants, or intermediate populations with atypical density distributions? Providing a brief analysis or discussion of their possible origins and biological relevance would give readers a more comprehensive understanding of the dataset.

In addition, clarification is needed for the statement that "the relative percentage of protein profiles for sEV (36%), NV-dual (30%), NV (24%), and UNC (12%) classes were remarkably similar across cell lines." It is not entirely clear how these percentages were derived—whether they reflect the proportion of total identified proteins, the number of density gradient profiles per class, or normalized averages across cell lines. A concise explanation in the Methods or figure legend would improve the transparency of the quantitative analysis.

Version 2:

Decision Letter:

Dear Dr Munoz,

I am pleased to inform you that your manuscript, "Defining the Reference Proteomes for Small Extracellular Vesicles and Non-Vesicular Components", has now been accepted for publication in *Nature Cell Biology*.

Thank you for sending us the final manuscript files to be processed for print and online production, and for returning the manuscript checklists and other forms. Your manuscript will now be passed to our production team who will be in contact with you if there are any questions with the production quality of supplied figures and text.

Over the next few weeks, your paper will be copyedited to ensure that it conforms to *Nature Cell Biology* style. Once your paper is typeset, you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required.

After the grant of rights is completed, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately.

You will not receive your proofs until the publishing agreement has been received through our system.

Due to the importance of these deadlines, we ask that you please let us know now whether you will be difficult to contact over the next month. If this is the case, we ask you provide us with the contact information (email, phone and fax) of someone who will be able to check the proofs on your behalf, and who will be available to address any last-minute problems.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

Once your paper has been scheduled for online publication, the Nature press office will be in touch to confirm the details. An online order form for reprints of your paper is available at <https://www.nature.com/reprints/author-reprints.html>. All co-authors, authors' institutions and authors' funding agencies can order reprints using the form appropriate to their geographical region.

Publication is conditional on the manuscript not being published elsewhere and on there being no announcement of this work to any media outlet until the online publication date in *Nature Cell Biology*.

You may wish to make your media relations office aware of your accepted publication, in case they consider it appropriate to organize some internal or external publicity. Once your paper has been scheduled you will receive an email confirming the publication details. This is normally 3-4 working days in advance of publication. If you need additional notice of the date and time of publication, please let the production team know when you receive the proof of your article to ensure there is sufficient time to coordinate. Further information on our embargo policies can be found here: <https://www.nature.com/authors/policies/embargo.html>

Please note that *Nature Cell Biology* is a Transformative Journal (TJ). Authors may publish their research with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. [Find out more about Transformative Journals](https://www.springernature.com/gp/open-research/transformative-journals)

Authors may need to take specific actions to achieve compliance with funder and institutional open access mandates. If your

research is supported by a funder that requires immediate open access (e.g. according to [Plan S principles](https://www.springernature.com/gp/open-science/plan-s-compliance) or the [NIH public access policy](https://www.springernature.com/gp/open-science/us-federal-agency-compliance)) then you should select the gold OA route, and we will direct you to the compliant route where possible. Because authors warrant under our subscription licensing terms that they haven't committed to licensing any version of their article under a licence inconsistent with the terms of our agreement – including the applicable embargo period – publication under the subscription model isn't suitable for authors whose funders require no embargo.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

If your paper includes color figures, please be aware that in order to help cover some of the additional cost of four-color reproduction, Nature Portfolio charges our authors a fee for the printing of their color figures. Please contact our offices for exact pricing and details.

As soon as your article is published, you will receive an automated email with your shareable link.

If you have not already done so, we strongly recommend that you upload the step-by-step protocols used in this manuscript to protocols.io (<https://protocols.io>), an open online resource that allows researchers to share their detailed experimental know-how. All uploaded protocols are made freely available and are assigned DOIs for ease of citation. Protocols and Nature Portfolio journal papers in which they are used can be linked to one another, and this link is clearly and prominently visible in the online versions of both. Authors who performed the specific experiments can act as primary authors for the Protocol as they will be best placed to share the methodology details, but the Corresponding Author of the present research paper should be included as one of the authors. By uploading your Protocols onto protocols.io, you are enabling researchers to more readily reproduce or adapt the methodology you use, as well as increasing the visibility of your protocols and papers. You can also establish a dedicated workspace to collect your lab Protocols. Further information can be found at <https://www.protocols.io/help/publish-articles>.

You can use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature Portfolio.

Nature Cell Biology encourages authors presenting evidence for cell, biological, molecular, and genetic interactions to consider communicating these findings using Biofactoid (<https://biofactoid.org/>). This tool helps users share a searchable representation of interactions (e.g. binding, gene expression, post-translational modification) between genes, gene products, or chemicals. Information added to Biofactoid, with author attribution, is shared on social media and public databases, such as Pathway Commons, where it can be discovered and analyzed in the context of a large and growing corpus of knowledge.

Please feel free to contact us if you have any questions.

With kind regards,

Melina Casadio, PhD
Senior Editor, Nature Cell Biology
ORCID ID: <https://orcid.org/0000-0003-2389-2243>

Click here if you would like to recommend Nature Cell Biology to your librarian
<http://www.nature.com/subscriptions/recommend.html#forms>

** Visit the Springer Nature Editorial and Publishing website at http://editorial-jobs.springernature.com?utm_source=ejP_NCB_email&utm_medium=ejP_NCB_email&utm_campaign=ejp_NCB for more information about our career opportunities. If you have any questions please click [here](mailto:editorial.publishing.jobs@springernature.com).

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 license, and indicate if changes were made. In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source. The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <https://creativecommons.org/licenses/by/4.0/>

Point-by-Point Response to Reviewers

We thank the reviewers for their thoughtful and constructive feedback, which has significantly improved the quality and rigor of our manuscript. We were encouraged that both reviewers recognized the importance of our work, with Reviewer #1 describing it as having the potential to be a "landmark study in the EV field" and Reviewer #2 calling it "a valuable contribution to the EV field".

Both reviewers raised valid concerns regarding the need for more extensive biophysical validation of our sEV preparations and requested further clarification of our protein classification methodology. In response to the reviewers' feedback, we have performed several new experiments and analyses, which are now presented in new and revised figures:

Biophysical Validation:

- Cryo-Electron Microscopy (cryo-EM) to visualize vesicle morphology (new **Extended Data Fig. 2d and Extended Data Fig. 6a**).
- Nanoparticle Tracking Analysis (NTA) to quantify particle size and concentration for all 15 cell lines and biofluids (new **Extended Data Fig. 2c and Extended Data Fig. 5**).
- Western Blotting for positive and negative markers to confirm sEV enrichment and purity (new **Extended Data Fig. 6b**).
- Refractometry to confirm the buoyant density of our gradient fractions (new **Extended Data Fig. 2b**).

Methodological Clarification:

- We have clarified our protein classification methodology and introduced a more precise nomenclature (NV-Dual) to avoid ambiguity.
- We have cross-validated our classifications against the Jeppesen *et al.* dataset, confirming their accuracy (new **Extended Data Fig. 8**)

The manuscript text has been substantially revised, with numerous improvements and the restructuring of key sections. Given the extensive nature of these changes, we have not highlighted every modification. Instead, to guide the review, we explicitly refer to the location of the most significant textual changes within our detailed responses below. In addition, we have corrected minor inaccuracies and improved the visual presentation of several figures throughout the manuscript to enhance clarity.

In the following point-by-point response, we outline our revisions and the rationale behind them. To facilitate the review, we have used the following colour scheme:

Black: Original reviewer comments.

Blue: Our explanation and response.

Green: We indicate where major text changes were done in the manuscript document.

Reviewer #1 (Remarks to the Author): The manuscript by Morales-Sanfrutos et al. presents a comprehensive proteomic atlas of extracellular vesicles (EV) using a density gradient and protein correlation profiling (DG-PCP) approach. This research holds the potential to be a landmark study in the EV field by addressing the critical challenge of defining the “true” proteome of EV. However, I believe its impact and long-term reliability could be significantly enhanced by addressing two key areas related to experimental validation and data interpretation. The evidence presented does not sufficiently support the main claims, “A High-Resolution Proteomic Atlas of Extracellular Vesicles”.

We thank the reviewer for their thoughtful assessment of our manuscript and for recognizing its potential as "a landmark study in the EV field." We also appreciate the constructive feedback provided. We have carefully addressed the points raised concerning experimental validation and data interpretation throughout the manuscript.

Major Points for Consideration:

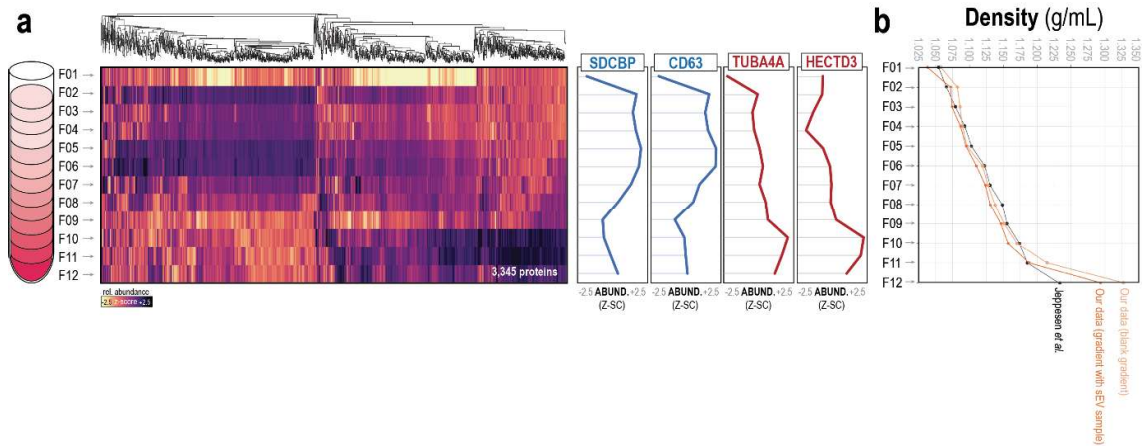
1. Strengthening the Foundational Evidence for EV Separation and Reproducibility. The manuscript relies heavily on the downstream proteomic profiles to validate the separation, with limited upstream characterization. To ensure the quality and reproducibility of the input material, I strongly recommend including more extensive validation in line with MISEV2023 guidelines (Welsh, J.A. et al., J Extracell Vesicles, 2024). The manuscript currently lacks quantitative data on the physical properties of the isolated particles.

We agree with the reviewer on the importance of incorporating upstream validation of the isolated material, in line with MISEV guidelines. To directly address this request, we have now performed several analyses of our density gradient (DG) fractions as well as crude sEV pellets (P100) using both NTA and cryoEM (see our responses to specific comments below).

In addition, to provide further validation of our methodology, we have also compared our data with the well-established reference work by Jeppesen *et al.* (Jeppesen *et al.* Cell 2019, PMID: 30951670), which utilized a highly similar iodixanol gradient approach. As shown below, our protein profiles normally aligned with the distribution of sEV markers and NV proteins reported by Jeppesen *et al.* using western blotting. In this comparison sEVs (red signal in Jeppesen *et al.*) were consistently present in these two experiments between fractions 3 and 6, while NV proteins (green signal in Jeppesen *et al.*) exhibited a higher density, predominantly appearing between fractions 7 and 10.

[REDACTED]

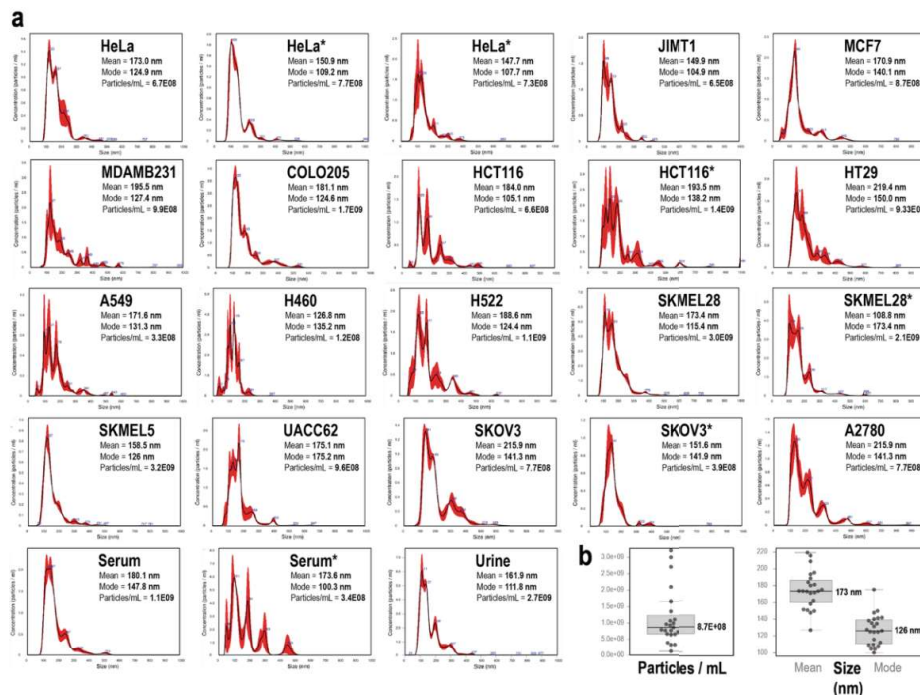
To further validate our methodology, we have now measured the buoyant density of a new iodixanol gradient using refractometry. Our proteomics data of this gradient consistently show sEVs protein enrichment in fractions 2 through 6. The refractometry analysis confirmed that these specific fractions correspond to buoyant densities ranging from 1.075 to 1.125 g/mL, which is the established range for sEVs. Remarkably, the measured density profile across the entire gradient showed excellent concordance with the reference gradient reported by Jeppesen *et al.* This analysis, now presented in the revised manuscript as new **Extended Data Fig. 2a-b**, provides robust, independent validation of our separation conditions.



(new) Extended Data Fig. 2. 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 and NV-associated proteins are shown. **b,** Refractometry analysis confirmed that sEV-enriched fractions (F2-F7) 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.

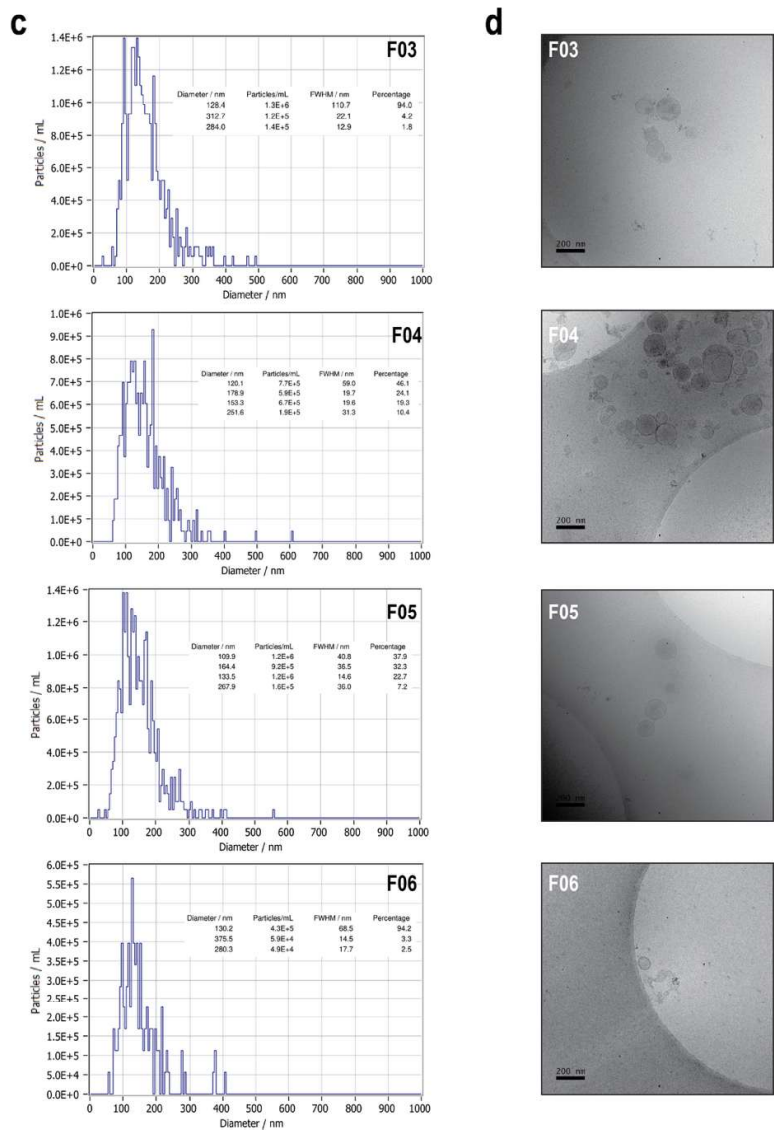
1-1. The authors should provide particle size distribution and concentration data for the main fractions collected from the iodixanol density gradient via NTA.

We have now characterized the particle size distribution and concentration of our samples by NTA. First, we have analyzed crude sEV pellets (P100) obtained by differential ultracentrifugation from all 15 cell lines (including biological replicates) used in this study as well as serum and urine. The estimated particle size was 126 nm (median of all modes) whereas the particle concentration was $8.7E+08$, consistent with values typically reported for sEVs isolated from conditioned media and biofluids. These NTA data are presented as new **Extended Data Fig. 5a-b**.



(new) **Extended Data Fig. 5. 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 of three measurements, with summary statistics for mean, mode, and particle concentration 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. Box plot elements: center lines indicate medians; box limits represent the 25th and 75th percentiles; whiskers extend to 1.5 times the interquartile range.

Additionally, we have performed NTA on new gradient-purified sEVs from one of our cell lines, UACC62. In this case, P100 sample was separated by iodixanol density gradient. Low-density fractions were collected, diluted 1:10 in PBS, re-pelleted by ultracentrifugation at $120,000 \times g$ for 3.5 hours at 4°C , and resuspended in $30 \mu\text{L}$ of PBS prior to analysis. These NTA measurements revealed a particle size of approximately 122 nm in the low-density fractions, consistent with the presence of sEVs. These NTA data are now provided as **Extended Data Fig. 2c** in the revised manuscript.



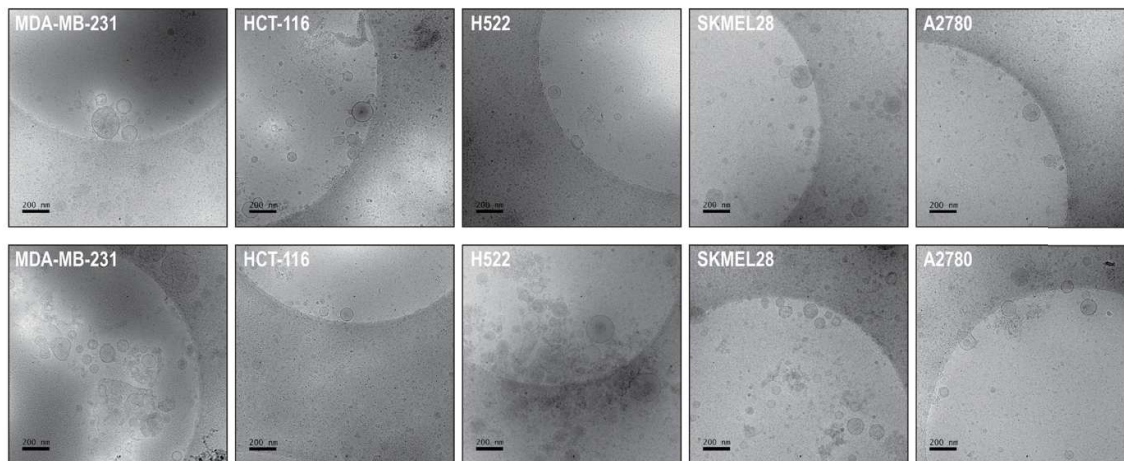
(new) Extended Data Fig. 2. Upstream validation of gradient-purified sEVs. **c**, NTA histograms of the corresponding low-density fractions, showing a primary particle population with a diameter of ~122 nm. **d**, Representative cryo-EM images of low-density fractions (F03-F06) showing intact vesicular structures consistent with sEVs.

1-2. There is no direct optical evidence in the manuscript to confirm the presence of vesicles. Provide representative Transmission Electron Microscopy (TEM) images of samples from the corresponding sEV-rich fractions to confirm the presence and morphology of vesicles. Incorporating these data would provide unequivocal, independent evidence for the quality of the isolated material and strengthen the interpretation of the proteomic results.

To provide direct visual confirmation of vesicles, we performed cryo-electron microscopy (cryo-EM) on both crude P100 and gradient-purified sEV preparations. Cryo-EM was chosen over conventional TEM due to its ability to preserve vesicles in a hydrated, near-native state, avoiding artifacts from fixation, staining, or dehydration. Unlike TEM, which can distort EV morphology (*e.g.*, inducing a cup-shaped appearance), cryo-EM allows visualization of vesicles with native ultrastructural integrity.

We first imaged crude P100 pellets from five representative cancer cell lines MDA-MB231, HCT116, H522, SKMEL-28, and A2780 using a JEM-1230 transmission electron microscope (JEOL, Japan), equipped with an UltraScan 4000 SP (4008×4008 pixels) cooled slow-scan CCD camera (GATAN, UK), and a 120-kV LaB6 thermionic gun. In all samples, we observed vesicular structures of size and morphology consistent with sEVs. These cryo-EM data are now provided as **Extended Data Fig. 6a** in the revised manuscript.

a



(new) **Extended Data Fig. 6. 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.

Subsequently, we analyzed low-density fractions from a new iodixanol gradient by cryo-EM. Vesicles with characteristic sEV morphology and size were consistently observed in the low-density fractions. Representative cryo-EM images (shown above in our response to comment 1.1) are now included as **Extended Data Fig. 2d** in the revised version of the manuscript.

2. Addressing the Conceptual Challenge of Proteomic Depth vs. Biological Specificity. In total 8,831 human proteins were identified from the enriched sEV fractions of 15 human cancer cell lines. However, this impressive depth raises a crucial conceptual question for the field. As analytical sensitivity increases identifying nearly the full cellular proteome in EV samples, the risk of detecting proteins that are not selectively sorted into EVs—but rather represent low-level contaminants or stochastically included cellular proteins—also grows. This raises a key question: At what point can a protein confidently be considered a bona fide EV cargo, as opposed to a background constituent? To preserve the biological significance of EV protein cargo profiles, I strongly encourage the authors to discuss this issue explicitly.

We thank the reviewer for raising this important point. We would first like to clarify a possible misunderstanding. The total of 8,831 proteins cited in our manuscript refers to the cumulative number of all proteins identified across the 15 cell lines analyzed, not to the total number of genuine sEV proteins. This aggregate includes both vesicular and all the other non-vesicular proteins and does not represent the size of the sEV proteome in any individual condition.

The actual number of proteins classified by DG-PCP as sEV cargo was considerably lower. As shown in **Fig. 2b**, individual cell lines typically yielded ~1,500 sEV proteins, with some (*e.g.* COLO205, SKMEL28, SKMEL5) exceeding 2,000 proteins. Interestingly, a similar number of proteins (2,398 proteins) was obtained in our analysis of urinary sEVs (**Fig. 8e**), suggesting this range might reflect a consistent estimate of the sEV proteome across diverse biological sources.

Still, even a proteome of 1,500–2,000 proteins is strikingly complex. Moreover, our quantitative proteomic data revealed that their protein abundances span over five orders of magnitude, as classical markers (*e.g.* CD9, SDCBP) were found to be >100,000 times more abundant than many other sEV-classified proteins. This wide dynamic range raises precisely the kind of questions highlighted by the reviewer: are all these low abundance proteins actively sorted into EVs, or might some represent stochastically incorporated background?

In the original manuscript, we addressed this issue explicitly (page 17), noting that a 100 nm vesicle could theoretically carry no more than ~1,000 protein molecules on its surface, assuming typical protein dimensions. The identification of hundreds or thousands of proteins per sample thus implies that most proteins must be present only in a subset of vesicles. This leads to two competing models: one in which vesicles represent discrete subpopulations with defined protein compositions, and another in which cargo incorporation is largely stochastic, generating high vesicle-to-vesicle variability. Based on the heterogeneity observed and the absence of evidence for distinct sorting pathways, we favour the latter model, though single-vesicle proteomics approaches will be essential to definitively resolve this question.

Our interpretation is supported by recent studies from Stephen J. Gould's group, which demonstrated that sEV secretion occurs predominantly at the plasma membrane via a stochastic mechanism (PMIDs: 35988652 and 38718108). These findings align with our own data (we showed that sEVs are highly enriched in plasma membrane proteins) and reinforce the idea that nonspecific incorporation of nearby proteins may be an inherent feature of the sEV biogenesis process, particularly in crowded intracellular environments.

In this context, distinguishing actively sorted from passively or stochastically included proteins becomes essential. While low-abundance proteins may reflect stochastic encapsulation, their biological relevance cannot be ruled out *a priori*, as some may exert potent effects even at low copy number. Nonetheless, several features can help prioritize likely genuine cargo:

- Abundance: highly abundant proteins, especially those consistently found across samples are strong candidates for specific loading.
- Comparative profiling: as shown in **Fig. 2d**, many genuine sEV proteins are secreted at levels that deplete their intracellular pools, without a corresponding decrease in mRNA expression. This suggests active sorting, whereas stochastically released proteins would not be expected to show this behaviour.
- Conservation: proteins that are consistently enriched in sEVs across samples, conditions and species are more likely to serve fundamental, biologically significant roles.

Although mechanistic validation of low abundance sEVs proteins, for instance by perturbing candidate proteins or sEV biogenesis pathways, is beyond the scope of this work, we agree it will be essential in future studies. Our contribution here is to provide a large-scale, quantitative, and systematically classified sEV proteome that lays the groundwork for such investigations.

We have now further expanded these points in the Discussion section of the revised manuscript (Page 17, second paragraph).

This manuscript has high potential to become a robust and reliable benchmark for the field. Their science is sound, and the approach is innovative. Therefore, I recommend this manuscript for publication after Major Revisions to address these points—particularly by strengthening the upstream validation and providing a more discussion on the identification of EV proteome and interpretation of high-depth proteomic data. I look forward to seeing the revised version.

Reviewer #2 (Remarks to the Author): Overall comments to Author: Due to the inherent limitations of current EV isolation methods, it remains technically challenging to obtain highly pure and well-defined subpopulations of extracellular vesicles (EVs) from complex biological samples. Conventional purification techniques frequently co-isolate non-vesicular contaminants, such as lipoproteins and abundant plasma proteins, which interfere with downstream analyses and confound the biological interpretation of EV functions. Furthermore, commonly used EV markers are insufficient to accurately assess the purity of EV preparations, adding further uncertainty to current studies. In this context, the authors present a relatively comprehensive proteomic atlas of small EVs (sEVs) derived from multiple cell lines, and importantly, extend this analysis to sEVs isolated from serum and urine. By applying a protein correlation profiling approach, the study aims to reclassify and enrich for bona fide sEV cargo proteins, distinguishing them from co-purifying background proteins. This resource represents a valuable contribution to the EV field, offering improved resolution in defining sEV composition and laying a foundation for more reliable functional and biomarker studies. While the study's methodological novelty is somewhat limited, the breadth and quality of the dataset are highly commendable. I have several suggestions that could help improve the overall clarity, presentation, and scientific rigor of the manuscript.

We thank the reviewer for their comprehensive assessment and for recognizing that our data represents "a valuable contribution to the EV field". We appreciate their thoughtful suggestions to enhance the manuscript's clarity and rigor.

Before addressing the specific points, we wish to clarify a key change in nomenclature that we have implemented throughout the revised manuscript and figures. In the original submission, we used the term "DUAL" to describe proteins with a "bimodal density profile". To avoid any misinterpretation that these proteins have a "dual biological origin" (*i.e.* sEV and NV), we have renamed this class to "NV-dual." We feel this new term more accurately reflects our finding that most of these bimodal-distributed proteins are non-vesicular proteins that can associate with the sEVs (and consequently also appear at low-densities). In our responses below, we will therefore refer to this class of proteins as "NV-dual" (and its sub-clusters as "NV-dual 1" and "NV-dual 2").

Summary of the Key Results: This study presents a highly detailed proteomic characterization of small extracellular vesicles (sEVs), leveraging a combined density gradient and protein correlation profiling (DG-PCP) strategy to systematically define sEV protein cargo with high specificity. Key findings include the predominance of plasma membrane and membrane-associated proteins in sEVs, the limited presence of cytosolic proteins, and the hypothesis that most cytosolic proteins in sEV preparations are contaminants. Additionally, the authors demonstrate that urine-derived sEVs provide superior depth and tissue specificity for biomarker discovery compared to serum-derived sEVs. The presence and nature of ubiquitin within sEVs are also explored, leading to the suggestion that free, rather than conjugated, ubiquitin may dominate.

We thank the reviewer for this summary, which underscores the main contributions of our work.

Originality and Significance: The primary contribution of this study lies in the systematic application of established techniques to generate one of the most comprehensive small extracellular vesicle (sEV) proteomic datasets to date, which represents a valuable resource for the field. However, from a methodological standpoint, the work largely builds upon and refines existing approaches—such as density gradient separation and protein correlation profiling—rather than introducing fundamentally novel techniques. As such, while the dataset is impressive in scale and quality, the methodological innovation is relatively limited.

We agree that our study builds upon established methodologies. However, we believe that the way these methods are integrated and applied represents a meaningful methodological advance in the field. Density gradient centrifugation is widely regarded as the gold-standard approach for purifying sEVs, yet it remains limited in its ability to fully resolve sEVs from co-isolated non-vesicular components. Protein correlation profiling (PCP), although well-established in other spatial proteomics applications, had not previously been applied to EV research. Therefore, coupling this resolving purification method with the analytical power of PCP was not only logical but essential for achieving unprecedented confidence in defining the sEV proteome.

Data & Methodology: Validity of Approach, Quality of Data, Quality of Presentation: The DG-PCP methodology is well-justified and appears sound for distinguishing bona fide sEV cargo from co-isolated proteins. The discussion implies a high level of technical rigor, including validation through correlation analysis with

known markers. However, further clarification in the main text regarding the reproducibility across biological replicates and validation of the novel ER-derived vesicles would enhance robustness.

In the original submission, we assessed biological and technical reproducibility at multiple levels. **Extended Data Fig. 3a** (formerly Suppl. Figure 1A) demonstrates the reproducibility of protein profiles across three density gradients derived from the same sEV pellet, confirming excellent technical consistency. **Extended Data Fig. 7** (formerly Suppl. Figure 3) shows additional gradients from independent P100 pellets across three distinct cell lines, supporting biological reproducibility. Most importantly, our classification system showed strong agreement across datasets: as shown in **Fig. 2c**, 82% of proteins classified as sEV in one cell line were consistently reclassified as sEV in the remaining 14 lines. This high concordance also extended to biofluids, with 85% (**Fig. 8f**) and 63% (**Fig. 8h**) of serum and urine bona fide sEV proteins, respectively, classified as genuine sEV in the cell lines dataset. In response to the reviewer's suggestion, we have now made these aspects of reproducibility more explicit in the revised manuscript (Page 3, second paragraph; Page 5, second paragraph).

Regarding the ER-derived proteins found in medium-density fractions, we agree that their presence is an intriguing observation. However, a detailed characterization of these putative vesicles lies outside the scope of this study, which is focused on delivering a high-confidence reference map of bona fide sEV proteins. We chose however to include this result as it underscores the resolving power of DG-PCP and its capacity to detect previously unappreciated populations.

Appropriate Use of Statistics and Treatment of Uncertainties: While the discussion references correlation and quantitative relationships (e.g., between intracellular and sEV protein levels), it does not clearly describe the statistical treatment or uncertainty ranges. A more explicit mention of statistical thresholds used for defining co-elution or enrichment would increase transparency. For instance, describing false discovery rates, confidence intervals, or methods to assess significance of differential protein abundance would strengthen confidence in the interpretations.

Throughout our study, we have aimed to be rigorous and to apply appropriate statistical analyses, wherever applicable, to assess the significance and robustness of our findings.

In our protein correlation profiling (PCP) approach, co-eluting proteins were identified using an unsupervised clustering strategy (the rationale for using an unsupervised strategy is explained in great detail below, in our response to comment 11). Briefly, protein intensities across density gradient fractions were log₂-transformed, and missing values, typically in Fractions 1–2 and 11–12 where protein detection is sparse, were imputed with randomized low-intensity values derived from the dataset. The data were then centered and scaled at the row level using z-scores. To ensure data quality prior to clustering, we applied two filtering steps. First, only proteins detected in at least four fractions were considered. Second, proteins with noisy profiles were excluded based on cosine fitting, using 1,000 randomizations and a false discovery rate (FDR) of 5%. Then, unsupervised hierarchical clustering (Euclidean distance, average linkage) was used to group co-migrating proteins. Clusters were manually annotated as sEV, NV, and NV-dual (formerly DUAL) based on their density profiles. In some experiments, we further refined the NV-dual category into NV-dual 1 (formerly DUAL1) and NV-dual 2 (formerly DUAL2), depending on the relative abundance at low versus high densities. While these procedures were described in detail in the Methods section, we recognize that for non-specialist readers, these steps may not be immediately intuitive. To address this, we have now incorporated a more explicit description of our PCP workflow in the Results section (Page 2, last paragraph; Page 3, last paragraph).

In the differential protein abundance analysis presented in **Fig. 5c-e**, where we assessed the effects of Na₂CO₃ treatment on sEV protein distribution, differential profiles were identified using paired one-sided t-tests. Specifically, we compared protein intensities in Fractions 3–7 between Na₂CO₃ and DMSO conditions to detect loss from low-density fractions, and Fractions 13–16 to detect enrichment at high densities. Multiple testing correction was applied using a permutation-based FDR approach. Proteins with a q-value < 0.2 and an absolute log₂ fold change > 0.3 were considered significant. To increase robustness, we only considered proteins significant in both tests (*i.e.* loss at low densities and gain at high densities), and whose profiles were well classified in the DMSO condition. This statistical approach was described in the methods section.

For most of the correlation analyses done in this work, we employed Spearman correlation, which evaluates monotonic relationships between ranked variables. When applicable, we also reported the p-values associated with Spearman's rho (ρ) (e.g. Fig. 3f). These p-values provide a measure of statistical confidence, indicating the likelihood of observing the measured correlation under the null hypothesis of no co-expression.

Conclusions: Robustness, Validity, Reliability: The conclusions are generally well-supported by the data and thoughtfully contextualized. The authors convincingly argue for a reconsideration of cytosolic proteins as sEV components and propose a plausible model of stochastic cargo loading. However, the interpretation of ER-derived vesicles is speculative and would benefit from additional experimental support. Likewise, while the authors downplay the role of ubiquitinated substrates in sEVs, this conclusion is limited by the resolution of bulk proteomics and should be qualified accordingly.

As stated above, we believe that a detailed investigation into the nature of the putative ER-derived vesicles falls beyond the scope of the present study, as it would divert from our primary objective of establishing a comprehensive proteomic atlas of sEVs and non-vesicular proteins. We agree with the reviewer that our interpretation remains speculative. Our aim was only to offer a plausible explanation for this observation, while clearly acknowledging in the text that further experimental validation will be required.

Regarding the role of ubiquitination. We agree that our bulk proteomic analysis might indeed have limitations in sensitivity for comprehensively identifying ubiquitinated cargo. Prior immune-purification of ubiquitinated peptides would certainly enhance the likelihood of detecting such modified proteins. However, given that ubiquitin itself was found among the top 5 most abundant genuine sEV proteins in our data, our initial reasoning was that highly stoichiometric ubiquitinated substrates, if present, should be detectable even without such enrichment. Because we barely detected any ubiquitinated proteins, our interpretation, rather than a definitive conclusion, was that the ubiquitin present in sEVs might largely be in its unconjugated form, potentially due to the action of deubiquitinating enzymes, which we also detected in sEVs. We have modified this part of the discussion (Page 18, first paragraph) to better reflect this interpretation.

References: Appropriate Credit to Previous Work? The manuscript could do more to cite and discuss key prior contributions, including foundational studies that have examined EV heterogeneity, cargo sorting mechanisms, and EV proteome annotation. Some claims, such as the stochastic model or the role of HSPA8 in secretory autophagy, would benefit from being explicitly linked to existing literature.

We apologize for any unintended omissions regarding the citation of prior contributions. We have carefully revised the manuscript to incorporate 22 additional pertinent references throughout the text, explicitly linking our claims regarding the stochastic model and the role of HSPA8 in secretory autophagy to the existing literature.

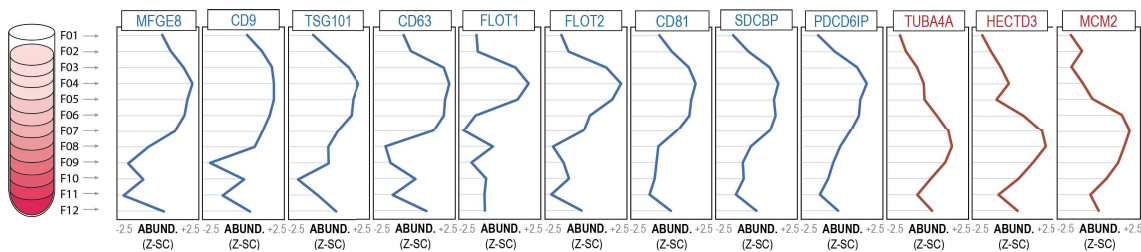
Clarity and Context: Lucidity of Abstract/Summary, Appropriateness of Abstract, Introduction and Conclusions: The discussion is scientifically rich but occasionally too dense, particularly when detailing ubiquitin chain types and sorting mechanisms. Simplifying and streamlining these sections would improve accessibility. The overall context is well-framed, though the implications for the broader field of EV-based biomarker discovery could be highlighted more clearly in the conclusions.

We thank the reviewer for these helpful suggestions. In response, we have streamlined and rephrased the relevant sections of the Discussion to improve clarity and flow, while maintaining the necessary scientific detail. We also agree that the broader implications for EV-based biomarker discovery warranted stronger emphasis. Accordingly, we have revised the Discussion section to better highlight how our classification framework and curated protein lists can enhance biomarker studies (Page 19, first paragraph).

Suggested Improvements: Experiments, Data for Possible Revision:

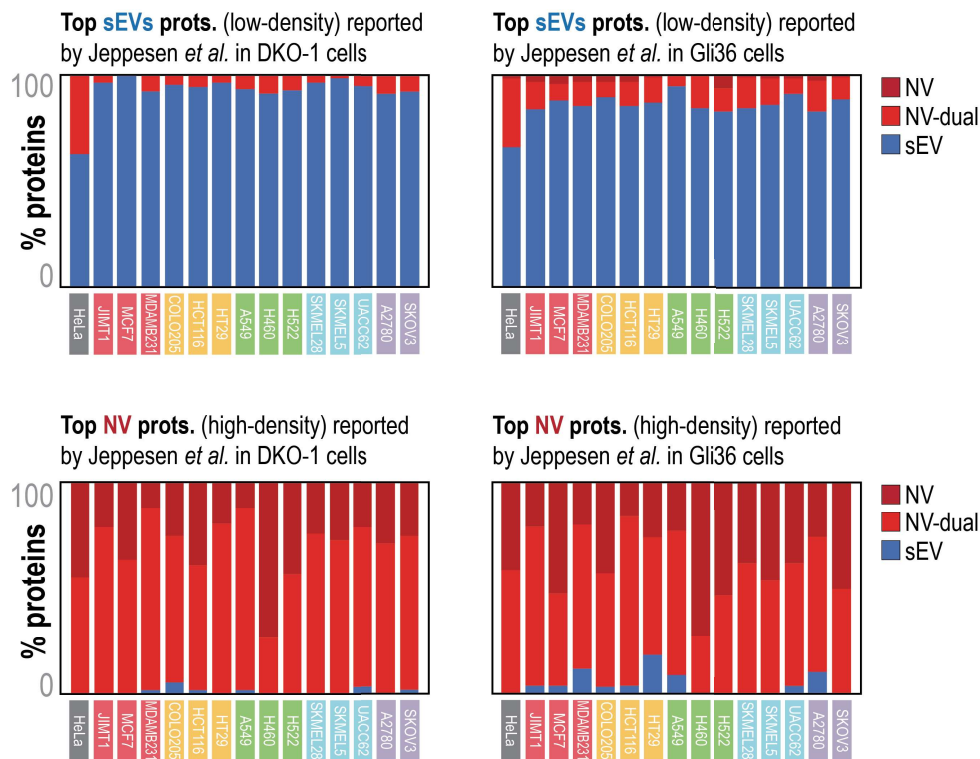
1. Limited Display of EV Markers in Figure 1A There are a wide variety of well-established EV markers reported in the literature, including but not limited to CD9, CD63, CD81, ALIX, TSG101, and syntenin. However, Figure 1A presents only three such markers. Given that more than a dozen EV-associated proteins are known to be robustly detectable in proteomic datasets, it is important to clarify whether these additional canonical markers were identified in the current study and how they were distributed between the sEV and NV groups. This information is essential for evaluating the accuracy of the proposed classification scheme.

We thank the reviewer for this important observation. All canonical sEVs markers referenced, including CD9, CD63, CD81, ALIX (PDCD6IP), TSG101, and syntenin-1 (SDCBP), were robustly identified across all 15 cell lines, as well as in serum and urine samples. Their identification and quantitative values were provided in the original submission in **Supplementary Tables 1 and 7**, respectively. Importantly, all these markers exhibited a unique low-density elution profile, which enabled their consistent classification as sEV proteins across sample types. To address the reviewer's comment, we have now included the elution profiles of all these canonical sEV markers (including also MFGE8, FLOT1 and FLOT2) in a new **Extended Data Fig. 1**, clearly illustrating their DG distribution and supporting their classification.



(new) Extended Data Fig. 1. Density gradient proteins profiles of classical sEVs markers and NV proteins. Crude sEVs were purified by dUC at 100,000 g (P100) from conditioned cell media of HeLa cells, separated by DG into 12 fractions and subsequently analyzed by LC-MS/MS. Representative protein profiles for several classical sEVs protein markers (blue) and NV-associated proteins (brown) are shown.

In addition, we have now validated our classification by cross-referencing with the top 200 sEV-enriched proteins and NV-associated proteins reported by Jeppesen *et al.* Cell, 2019 (PMID: 30951670) in DKO-1 and Gli36 cells, who used nearly identical density gradient protocols. As shown below, this analysis revealed a high concordance: 92-96% of the proteins found enriched in sEV-fractions by Jeppesen *et al.* were also confidently classified as sEV proteins in our dataset. Conversely, 84-87% of the proteins predominantly abundant in NV fractions in the Jeppesen *et al.* study were classified as NV or NV-dual (formerly DUAL) in our data. This is now included in the new **Extended Data Fig. 8**.



(new) Extended Data Fig. 8. Cross-validation of protein classifications with a reference dataset. Bar charts show how the top sEV (low-density) and NV (high-density) proteins from the Jeppesen *et al.* dataset (from DKO-1 and Gli36 cells) are classified by our DG-PCP approach into sEV, NV-dual, and NV categories in the indicated cell lines.

2. Ambiguous Classification of Proteins into sEV and NV Groups Throughout the manuscript, heatmaps classify proteins strictly into “sEV” and “NV” groups. However, such binary categorization may oversimplify the actual distribution. Proteins enriched in sEVs are not necessarily absent from the NV group, and vice versa. A more appropriate approach might be to classify different density fractions as sEV-enriched or NV-enriched, rather than assigning proteins to exclusive categories. This would better reflect the continuous nature of protein distribution across fractions.

We thank the reviewer for raising this crucial point. We share the reviewer's concern about the challenge of deconvolving a protein's biological origin from its presence across multiple fractions; in fact, our protein correlation profiling (PCP) approach is precisely designed to solve this problem.

The core issue, as the reviewer notes, is the continuous nature of protein distribution in a density gradient. Biophysical factors like diffusion and aggregation inevitably lead to overlapping populations rather than perfectly discrete bands in a density gradient. This results in "smearing," where a pure population of vesicles migrating to its buoyant low density will still leave detectable traces in subsequent high-density fractions. Given the extreme sensitivity of modern mass spectrometry, these trace amounts are easily detected. As a result, a canonical sEV marker like SDCBP was detected in our data in high-density fractions, while a highly abundant non-vesicular protein like TUBA4A was detected in low-density fractions. This creates a significant analytical challenge: an analysis of the conventionally defined "sEV-enriched" fractions would incorrectly identify TUBA4A as a potential sEV protein. We explicitly identify this pitfall in our manuscript (Fig. 1d), where we demonstrate that a staggering 80% of all proteins found in the conventionally defined "sEV-enriched" fractions did not, in fact, have a genuine sEV profile. This finding illustrates that relying on the analysis of a few "pure" fractions is a flawed strategy. The power of our DG-PCP approach is that it does not classify fractions, it classifies the entire elution profile of each protein. Instead of relying on a protein's mere presence in one or two fractions, our method uses the complete,

multi-point quantitative distribution, the shape of the profile, to make a confident assignment. While both SDCBP and TUBA4A are present in all fractions, their quantitative profiles have clearly distinct shapes. SDCBP has a sharp, unimodal peak in the low-density region, while TUBA4A's is in the high-density region. Our unsupervised clustering algorithm recognizes these distinct patterns, allowing it to deconvolve the data and confidently assign each one to its proper class, regardless of the experimental smearing.

Most importantly, this profile-based strategy was particularly crucial for interpreting proteins that exhibit a clear bimodal distribution (different from the “smearing” mentioned above) as shown in **Fig. 2b**. As pointed out by the reviewer, such bimodal profiles could suggest that these proteins are genuine components of both sEVs and NV components. However, our data indicate that most of these proteins are likely non-vesicular, with their low-density peak arising from an external association with sEVs. Several lines of evidence support this. First, many proteins that show such bimodal profile in one cell line appear as a unimodal NV profile in another (**Fig. 2c**), suggesting they belong to the same dynamic continuum of non-vesicular components. Second, GO analysis (**Extended Data Fig. 11**) reveals that such NV-dual proteins (formerly DUAL), like their unimodal NV counterparts, are mainly enriched for soluble, cytosolic proteins and lack the plasma membrane signature of genuine sEVs. If these soluble cytosolic proteins were genuine sEV cargo, they would need to be encapsulated within the vesicle lumen to be protected from the extracellular environment. However, we found that the low-density peak of bimodal proteins was sensitive to proteinase K degradation (**Fig. 6a-d**), indicating they are predominantly located outside the sEV membrane.

Finally, we do not rule out the possibility that some proteins may have a genuine dual origin. Our Na₂CO₃ treatment provides a nuanced picture of this complexity. We observed that this treatment caused some proteins with bimodal profiles to shift from low to very high density (**Fig. 5c-e**), confirming a direct vesicle association. One such protein was GAPDH. As one of the most abundant cytosolic proteins in the cell, its presence in the high-density NV fraction might be expected. However, its confirmed vesicle association in our data at low-densities is consistent with recent work showing that GAPDH can bind directly to the outer surface of sEVs via a phosphatidylserine-binding motif (Dar *et al.*, Nat. Commun., 2021; PMID: 34795295). This makes GAPDH an example of how a protein can exist as a genuine NV component (due to high abundance) and, simultaneously, as a genuine surface-associated sEV protein (due to specific interactions). Conclusively identifying all such cases will require complementary approaches beyond density gradients alone.

In the revised manuscript, we have modified pertinent sections to highlight how PCP enables confident deconvolution of protein classes (for instance in [Page 4, second paragraph](#); [Page 11, second paragraph](#)).

3. Unclear Subgrouping in Figure 1B. In Figure 1B, proteins are further subclassified into several subgroups. It is unclear whether these subgroups are mutually exclusive or hierarchical. The authors should provide a clear rationale and formal definitions for each subgroup to enhance reproducibility and interpretation.

We thank the reviewer for this question, which highlights an important detail of our classification. The subgroups are mutually exclusive, not hierarchical, as they represent distinct clusters that emerged directly from our unsupervised analysis of the protein elution profiles.

Our high-resolution DG-PCP approach revealed several distinct and reproducible protein profiles. The primary classifications are non-hierarchical and based on the overall shape of the elution profile. We identified proteins with a unimodal distribution, which were classified as either sEV (peaking at low density) or NV (peaking at high density). However, we also found that many other proteins exhibited bimodal distributions with distinct peaks in both the low- and high-density regions. We termed these profiles NV-dual (formerly DUAL), as our data strongly suggest they are non-vesicular proteins whose low-density peak arises from an association with sEVs (see our response to comment 12).

Within this broad NV-dual class, our unsupervised clustering algorithm consistently resolved two separate sub-clusters based on the quantitative partitioning of the protein's signal between the two peaks. These mutually exclusive subgroups are:

- NV-dual 1 (formerly DUAL1): Defined by a balanced bimodal distribution, where the protein's abundance in the low-density peak is comparable to its abundance in the high-density peak.

- NV-dual 2 (formerly DUAL2): Defined by an unbalanced bimodal distribution, where the protein is predominantly found in the high-density fractions, with a smaller but still distinct peak in the low-density region.

For clarity, we refer to these collectively as the "NV-dual" class for much of the manuscript. However, we do leverage this more granular distinction when it provides deeper mechanistic insight, particularly in the context of our protease protection assays (**Fig. 6a-f**). There, we show that the two subgroups exhibit different susceptibilities to proteinase K treatment, a behaviour that correlates with their differential association with the low-density sEV fractions.

Throughout the revised manuscript, we have clarified our classification system and nomenclature for the different clusters (for instance in **Page 3, last paragraph**).

4. Inconsistency in Supplementary Figure S1B. Supplementary Figure S1B states that three gradient formulations were used; however, only two are shown. Additionally, it is unclear whether technical replicates were performed for each formulation. Clarifying this would strengthen the methodological transparency of the study.

We appreciate the reviewer's attention to this detail. This was a typographical oversight in the legend. In fact, only two gradient formulations were tested in Supplementary Figure S1B: (1) the density gradient protocol described by Jeppesen *et al.* Cell 2019 (PMID: 30951670), which we used in all experiments presented in the manuscript with minor modifications, and (2) the alternative protocol from Crescitelli *et al.* Nat. Protoc. 2021 (PMID: 33495626) used here for comparison purposes only. No technical replicates were performed for this gradient comparison, as the separation profiles and protein distributions between the two protocols were highly similar. Our aim was to demonstrate the comparability of protein separation across gradient strategies rather than to generate a statistical comparison. We have corrected the figure legend accordingly (which is now renumbered to **Extended Data Fig. 3b**).

5. Ambiguity in Supplementary Figure S1C The terms "top" and "bottom" in Supplementary Figure S1C are not clearly defined. Do they refer to fractions from the top and bottom of the density gradient? The authors should clarify these labels to ensure accurate interpretation of the figure.

We thank the reviewer for pointing out this ambiguity. In our density gradient experiments, the crude sEV pellet (P100) was resuspended in 40% iodixanol and loaded at the bottom of the centrifuge tube. Sequential layers of decreasing iodixanol concentrations were then overlaid. This *bottom-loading* approach allows vesicles to float to their buoyant density, while denser non-vesicular (NV) components remain toward the bottom. An alternative is *top-loading*, where the sample is placed in the top, lowest-density layer. However, it has been suggested that *bottom-loading* is preferable, as the centrifugation conditions commonly used (*e.g.*, 15 h, 120,000 × g) may not allow NV components to equilibrate if top-loaded (Jeppesen *et al.*, Cell, 2019). To assess whether loading mode affects sEV/NV separation, we compared *bottom-loaded* and *top-loaded* gradients. Original Supplementary Figure S1C showed that the protein elution profiles and separation patterns were largely preserved, indicating that both approaches, in our hands, yield comparable results. To clarify this point, we have revised the legend and updated the figure itself (**Extended Data Fig. 3c** in the revised manuscript) to clearly define "top" and "bottom" in the context of sample loading. Although it was not explicitly requested, we have also improved **Extended Data Fig. 3d** to better illustrate the differences in the iodixanol buffer compositions investigated in this specific experiment.

6. Insufficient Characterization of Vesicles vs Non-vesicles. The methods for distinguishing vesicular and non-vesicular components rely heavily on a limited number of marker proteins. There is no evidence presented regarding particle size distribution, which is a critical parameter in EV characterization. Techniques such as nanoparticle tracking analysis (NTA) or transmission electron microscopy (TEM) should be included or more thoroughly discussed.

We thank the reviewer for this critical point, which was also raised by Reviewer #1. While our DG-PCP method provides a robust framework for distinguishing vesicular from non-vesicular proteins based on their elution

profiles, we agree that direct physical characterization is essential to complement our proteomic classifications. In response, we have now performed the requested orthogonal analyses to provide unequivocal biophysical evidence for our sEV preparations.

Validation of Gradient-Purified sEVs. We performed a biophysical characterization of our density gradient fractions to confirm the identity of the particles in the low-density, sEV-rich regions. First, refractometry measurements (**Extended Data Fig. 2b**; shown above in this document) confirmed that the fractions where sEV protein markers peaked corresponded to the established buoyant density range for sEVs (1.075-1.125 g/mL), which is consistent with published data from Jeppesen *et al.* Second, cryo-EM imaging (**Extended Data Fig. 2d**; shown above in this document) of these low-density fractions revealed intact, bilayered vesicular structures with the typical morphology and size of sEVs. Third, NTA of the same fractions (**Extended Data Fig. 2c**; shown above in this document) confirmed a particle population with the expected sEV size of ~122 nm.

Characterization of Crude sEV Starting Material (P100). In addition to characterizing the purified fractions, we performed extensive validation of the crude sEV (P100) pellets. Cryo-EM imaging of P100 pellets from five representative cancer cell lines (MDA-MB231, HCT116, H522, SKMEL-28, and A2780) confirmed the presence of vesicular structures with the correct size and morphology (**Extended Data Fig. 6a**; shown above in this document). Furthermore, we characterized the particle size and concentration of all crude P100 pellets from the 15 cell lines (including replicates), serum, and urine using NTA (**Extended Data Fig. 5a-b**; shown above in this document). The median particle size was 126 nm and concentrations were consistent with values typically reported for sEVs isolated from conditioned media and biofluids.

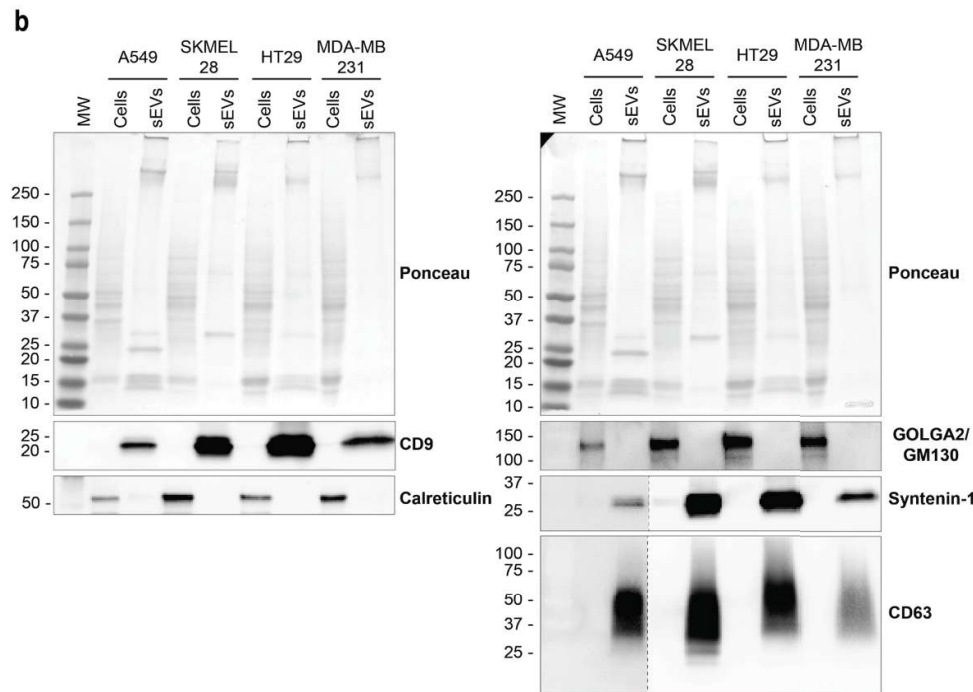
7. Improper Reagent Notation Some reagents are not properly formatted. For example, in the Methods section, “Na2CO3” should be correctly written as “Na₂CO₃.” Such issues, though minor, detract from the manuscript’s professionalism and should be corrected throughout.

We thank the reviewer for pointing this out. We have carefully reviewed the manuscript to ensure that all chemical formulas and reagent notations are correctly formatted. In the process, we also corrected several minor typographical inconsistencies elsewhere in the text to improve clarity and overall presentation.

8. Missing Validation Data for a Resource Article As a resource-type article, the manuscript should present comprehensive and rigorously validated datasets. However, key validation steps such as Western blotting (WB) for both positive and negative EV markers are missing. Although the authors mention NTA analysis, no corresponding data or figures are shown. Inclusion of such validation is essential to confirm the integrity and purity of the EV preparations.

As outlined above, we now include extensive characterization of our purified vesicle preparations. This includes measurements of buoyant density (by refractometry), particle size and concentration (by nanoparticle tracking analysis, NTA), and direct visualization by cryo-electron microscopy (cryo-EM).

Regarding the assessment of positive and negative EV markers, our manuscript already contained a systematic comparison of protein levels in our sEVs samples with their corresponding intracellular abundance in matched cell lines using published datasets (**Fig. 2e**). However, acknowledging the importance of orthogonal validation and following the reviewer’s suggestion, we have now performed Western blot analyses using established positive (CD9, CD63 and Syntenin-1) and negative (Calreticulin and GOLGA2/GM130) sEVs markers. These WB results confirm the enrichment of sEV markers and the absence of major intracellular contaminants, supporting the high purity of our preparations. The Western blot results are now included in the revised manuscript as new **Extended Data Fig. 6b**.



Extended Data Fig. 6. Characterization of P100 samples from cancer cell lines by cryo-EM and western blot. 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.

9. Title Scope and Sample Diversity The manuscript’s title refers to a “proteomic atlas,” suggesting comprehensive coverage. However, only cell line-derived EVs, along with serum and urine EVs, are included. Previous studies have highlighted that EVs from other biofluids—such as plasma, saliva, and tissue interstitial fluid—possess unique proteomic profiles. To justify the term “atlas,” the authors should either broaden the sample scope or revise the title to reflect the current dataset limitations.

We thank the reviewer for this thoughtful observation regarding the scope implied by our title. Our use of the term “*atlas*” was intended to highlight the unprecedented depth, resolution, and systematic classification of proteins in sEVs and non-vesicular fractions provided by our dataset. However, we recognize that the term may suggest broader biological coverage, in terms of sample origin, than the level accomplished here. To avoid any potential misinterpretation, we propose the following alternative:

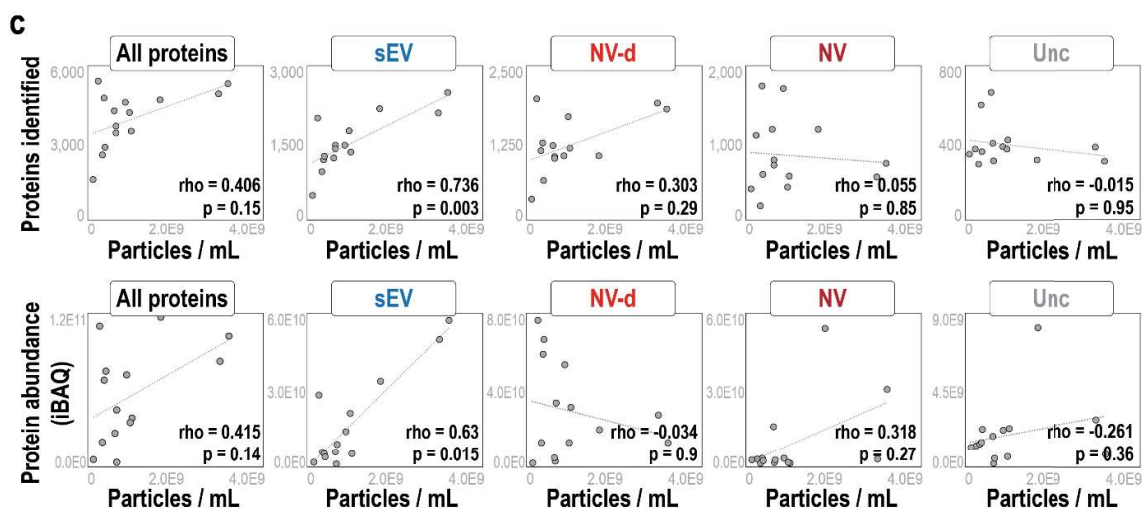
“Reference Proteomes of Small Extracellular Vesicles and Non-Vesicular Components.”

We believe this revised title better reflects the nature and scope of the work while maintaining its role as a foundational resource.

10. Variability in Protein Identification Across Cell Lines (Figure 2B) Figure 2B shows substantial differences in the number of sEV proteins identified across different cell lines. The authors should comment on whether this variability is due to inherent differences in EV secretion levels between cell types. Including NTA data to demonstrate particle yield per cell line could help clarify whether secretion volume or protein diversity is the primary factor.

It is well established that different cell types display variable sEV secretory activities. We therefore agree with the reviewer that the variability in the number of identified proteins across cell lines merits further analysis. To address this, we have compared protein identifications with NTA particle counts across our dataset.

This analysis (top panels in the figure shown below) revealed a moderate correlation between the total number of identified proteins and particle counts ($\rho = 0.406$). Strikingly, when restricted to proteins confidently classified as sEVs, the correlation was much stronger ($\rho = 0.736$). By contrast, the number of proteins assigned to other categories showed weak or negligible correlations with NTA (NV-dual, $\rho = 0.303$; NV, $\rho = 0.055$; Unclassified, $\rho = -0.015$). It is important to note that the number of proteins identified in a proteomics experiment does not necessarily scale with the total protein mass. Two samples may contain the same amount of protein, yet one can yield 500 identified proteins while another yields 1,000, depending on how the protein mass is distributed across different proteins. For this reason, we also compared particle counts with the total protein abundance estimated by iBAQ (bottom panels in the figure shown below), which better reflects protein yield than protein identifications alone. The results were consistent: particle numbers correlated closely with the cumulative abundance of sEV proteins, but not with other protein classes.



Extended Data Fig. 5. Nanoparticle Tracking Analysis (NTA) of crude sEVs preparations from human cancer cell lines and biofluids. **c**, Scatterplots showing the Spearman correlation (ρ and p -value) 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.

Taken together, these findings suggest that the variability in the number of proteins identified across cell lines is largely explained by vesicle yield: the more vesicles present in a sample, the more sEV proteins were detected. In other words, the observed differences mainly reflect technical sampling depth rather than intrinsic biological diversity in vesicle protein composition. These findings support the robustness of our classification system and further suggest that the average protein cargo per vesicle is relatively consistent across cell types. If major biological differences in protein loading existed, we would expect mismatches between particle counts and sEV protein identifications, which we did not observe. We have incorporated these analyses and their graphical representation as new **Extended Data Fig. 5c** and included a brief discussion in the main text (Page 5, first paragraph).

That said, an open question remains as to whether the observed differences in vesicle yield between cell lines are biological, reflecting genuine variation in secretion rates, or technical, arising from factors such as cell culture conditions (*e.g.*, density, passage number, conditioning time, cell cycle stage) and isolation efficiency (*e.g.*, centrifugation). Both scenarios are plausible, and several variables may contribute simultaneously. Although we minimized such effects through standardized protocols in our work, resolving their relative impact would require dedicated experiments, such as normalization to precise cell counts, systematic modulation of conditioning

periods, or orthogonal approaches like FACS-based vesicle phenotyping. These investigations lie beyond the scope of the present study but represent important avenues for future work.

11. Lack of Clarity in sEV Protein Classification Methodology The manuscript repeatedly mentions that 1,499 proteins were consistently classified as sEV proteins. However, it lacks a detailed description of the classification algorithm or criteria used for this categorization. This omission makes it difficult for readers to assess the robustness of the results or apply the same approach to their own data.

We thank the reviewer for highlighting this important point and apologize for the lack of detail in the original manuscript. We have now added a more explicit description of our PCP classification pipeline to the main text (Page 5, last paragraph) and provide a comprehensive explanation below.

In contrast to established supervised approaches in other PCP studies, in this work we deliberately chose an unsupervised framework. We think that this strategy allows us to leverage the full information contained in the co-fractionation profiles while minimizing the risk of bias that can arise when supervised methods are applied to systems with incomplete or ambiguous training sets.

Supervised approaches in PCP studies are highly effective in well-characterized systems, but they rely on extensive, high-quality prior information. For example, in PCP for protein–protein interaction mapping (Skinnider *et al.* Cell 2021; Kristensen *et al.* Nat Methods 2012), data analysis pipelines require a comprehensive reference of known protein complexes (*i.e.* the CORUM database) to train the classifier, allowing the model to learn characteristic co-elution patterns that distinguish interacting from non-interacting proteins. Similarly, in PCP for subcellular localization mapping (Christoforou *et al.* Nat Commun 2016; Geladaki *et al.* Nat Commun 2019), machine-learning models such as Support Vector Machines are trained using curated marker proteins whose subcellular locations are unequivocally established.

However, in situations where such prior knowledge is limited, supervised methods can struggle to provide reliable associations, as missing, misannotated, or incomplete markers tend to propagate errors across protein assignments, ultimately leading to misleading classifications. In the case of sEVs, canonical marker sets (including luminal, integral, or peripheral proteins) are still incomplete, and NV contaminants such as protein aggregates, lipoproteins, and other extracellular particles remain poorly characterized in terms of their biochemical and physical properties. Adding further complexity, sEVs themselves constitute a heterogeneous population that likely includes multiple subtypes, for which distinct molecular signatures are only partially understood and whose potential differences in buoyant density remain unresolved. Applying a supervised method in this scenario would risk introducing circularity (reinforcing incomplete assumptions) and systematic biases. To avoid these issues, we opted for an unsupervised approach that allows the data itself to define its structure, with cluster annotation performed only *a posteriori* using commonly used sEVs markers. This data-driven approach proved essential, as it not only reconstructed the expected sEV and NV profiles but also resolved previously unappreciated complexity, including the bimodal NV-dual profiles and a distinct class of putative ER-derived particles. These discoveries would likely have been missed by a supervised framework constrained by existing knowledge.

Technically, our PCP strategy proceeded as follows. Protein intensity profiles across gradient fractions were first compiled and log₂-transformed. Missing values, which were largely concentrated in low-signal edge fractions, were imputed with random low intensity values from the overall distribution to mimic stochastic low-abundance signals while avoiding artificial sharpening of profiles. Each protein profile was then centered and scaled (z-scored) to emphasize relative distribution rather than absolute abundance. To ensure robustness, we only retained proteins quantified in at least four fractions. In addition, we removed noisy or unstable profiles using a cosine-fit test with 1,000 randomizations and a 5% FDR threshold, which eliminated spurious profiles inconsistent with the smooth migration patterns expected from density fractionation. The resulting curated set of protein profiles was subjected to hierarchical clustering using Euclidean distance and average linkage, generating discrete groups of proteins with similar co-fractionation behaviour without any requirement for predefined labels. Clusters were annotated into biologically meaningful classes *a posteriori*. Those whose average profile was unimodal, peaked in the low-density region, and co-migrated with canonical sEV markers (*e.g.*, CD9, CD63, ALIX) were designated as sEV. The remaining clusters containing non-vesicular proteins were then annotated based on their distinct profiles: those with a unimodal high-density peak were classified as NV, while clusters exhibiting a bimodal distribution were classified as NV-dual (formerly DUAL).

Finally, to obtain a conservative and reproducible consensus sEV catalogue across cell types (Page 5, second paragraph), we applied a nominal threshold: each protein had to be annotated as sEV in at least 80% of the cell lines where it was detected and be observed in at least four different cell lines. While this threshold is somewhat arbitrary, it allowed us to generate a high-confidence consensus of 1,499 proteins. All individual protein classifications are provided in the **Supplementary Table 1**, enabling readers to apply alternative criteria if desired.

12. Unclear Interpretation of Dual and NV Proteins The statement that “DUAL and NV proteins appeared, to some extent, interchangeable in their classification” is concerning, as it may indicate instability in the classification method. Without a clear description of the classification criteria or computational workflow, this undermines confidence in the data segmentation. If custom scripts were used for classification, the authors should deposit the code in a publicly accessible repository, as this would significantly enhance reproducibility.

We appreciate the reviewer’s close reading and the opportunity to clarify our interpretation of non-vesicular protein profiles, which is indeed very relevant and perhaps was not sufficiently clear in the original manuscript. The apparent interchangeability between NV and NV-dual profiles does not reflect a methodological inconsistency but rather reveals the inherent, context-dependent nature of these proteins.

The behaviour of non-vesicular components in density gradients remains poorly characterized. Prior work by Jeppesen *et al.* established that many NV proteins exhibit a higher buoyant density than genuine sEVs. Our DG-PCP analysis confirmed this, identifying a large class of proteins with clear, unimodal high-density profiles. However, our comprehensive profiling of all gradient fractions revealed a more complex pattern not previously systematically described: many non-vesicular proteins clearly present at high densities were also detected at low densities, exhibiting a distinct bimodal distribution. These profiles, which we termed NV-dual (formerly DUAL), highlight that the co-purification of NV proteins is more intricate than a simple high-density separation.

Non-vesicular proteins, while also showing some degree of consistency (**Extended Data Fig. 9**), were more dynamic, with their classifications shifting sometimes between NV and NV-dual profiles depending on the cell line (**Fig. 2c**). While 66% (average of 15 cell lines) 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 as NV, a substantial 37% were reclassified as NV-dual in other cell lines. This non-random exchange does not suggest a methodological inconsistency but rather that NV and NV-dual profiles represent a dynamic continuum of non-vesicular proteins in which a protein’s tendency to adopt a purely high-density (NV) or a bimodal (NV-dual) profile could be a fluid, context-dependent characteristic. Several lines of evidence support this interpretation. First, the NV-dual 2 profile, in which proteins are predominantly found at high density with only a minor low-density peak, was significantly more frequent than the NV-dual 1 profile, which has a more balanced distribution. This suggests that the default state for these proteins is non-vesicular and high-density, and their co-migration with sEVs at low densities is a less common, secondary event. Furthermore, GO analysis confirmed a common origin for NV and NV-dual proteins as both classes were enriched for intracellular proteins and were distinct from the plasma membrane signature of sEVs. Finally, our protease protection assays provide crucial topological evidence. They confirm that the fraction of NV-dual proteins found at low density (co-migrating with sEVs) is located outside the vesicle, as it was sensitive to proteinase K degradation. This indicates an external interaction rather than intra-luminal encapsulation.

In the manuscript, we propose a plausible biophysical basis for this dynamic behaviour. Among the NV-dual proteins, we found that their presence in the low-density fractions correlates with smaller size and higher hydrophilicity. This suggests that these proteins might experience stronger non-specific interactions with sEVs than larger, more hydrophobic proteins, causing them to co-migrate into the low-density region. Furthermore, these interactions may be favoured by the apparent aggregated state of NV proteins. We observed that genuine soluble proteins (*e.g.* those released from vesicles by Na₂CO₃ treatment in **Fig. 5c** and the SDS-solubilized spiked-in *E. coli* lysate in **Fig. 6g**) migrated to even higher densities than typical NV proteins. This indicates that most NV proteins are not merely soluble contaminants but likely exist as aggregates, probably induced by the high centrifugation forces used for sEV purification, which may enhance their tendency to interact with sEVs.

We have revised the relevant sections of the text (for instance in [Page 5, last paragraph](#)) to provide this clearer and more detailed explanation of the NV and NV-dual protein categories, emphasizing their continuum as non-vesicular entities ([Extended Data Fig. 15c](#)) and the factors influencing their sEV association ([Extended Data Fig. 15b](#)).

13. Lack of Guidance for Filtering Non-sEV Proteins in Practical Applications Although the authors successfully establish a reference dataset that distinguishes between sEV-enriched and non-vesicular proteins, they do not provide practical recommendations for how to exclude contaminating proteins in downstream quantitative studies. In real-world EV research, distinguishing and quantifying authentic EV cargo is crucial. Merely labeling proteins as “sEV” or “non-sEV” is insufficient unless it informs how researchers can apply these distinctions in biomarker discovery or mechanistic studies.

We thank the reviewer for this important comment. We agree that distinguishing genuine sEV proteins from non-vesicular contaminants has critical implications for mechanistic studies, biomarker discovery, and quality control. We have now added a more detailed discussion of these points to the manuscript ([Page 18, third paragraph](#) and [Page 19, first paragraph](#)) and provide a comprehensive explanation below.

Quality Control

A common practice to evaluate sEV purity has been to cross-reference proteomic identifications against public repositories such as the ExoCarta “Top 100,” or to assess enrichment of GO terms like “extracellular exosome.” However, both resources are largely derived from legacy datasets that did not distinguish bona fide sEV cargo from frequently co-purifying proteins. As a result, these references carry forward systematic misannotations. In our work, this becomes evident: nearly half of the ExoCarta Top 100 proteins classify as non-vesicular contaminants ([Extended Data Fig. 10](#)), and proteins that we define as non-vesicular are paradoxically enriched for the “extracellular exosome” GO term ([Fig. 2g](#)). Together, this highlights the limited utility of such references, as their annotations reflect historical biases rather than accurate EV-specific assignments. By contrast, “plasma membrane”-related terms are much more representative of bona fide sEV proteins in our data ([Fig. 2g](#)). For this reason, we urge caution in using such databases as purity readouts and recommend instead emphasizing plasma membrane-associated annotations when characterizing sEV proteomes.

To address these challenges and provide an actionable tool for the community, we proposed in our work ([Fig. 2h](#)) two novel panels of robust positive and negative protein markers for the straightforward quality control of sEV preparations. These lists comprise the top 100 most abundant proteins classified as bona fide sEV components (positive markers) and, conversely, the top 100 most abundant proteins derived from the NV-dual, NV, and unclassified categories (negative markers). These two panels serve as a rapid and effective QC metric. For example, a high relative abundance of negative markers would strongly indicate significant contamination from NV fractions. In addition, these panels provide a standardized framework for evaluating and comparing the efficacy of novel sEVs isolation methodologies. It is important to note, however, a potential limitation: the profile of co-purified contaminants (*i.e.* negative markers) may be dependent on the specific isolation strategy employed.

Mechanistic studies

EVs are recognized mediators of cell–cell communication by transferring molecular cargo to recipient cells, thereby eliciting functional responses. This mode of action is generally thought to require docking at the plasma membrane, followed by receptor activation and signaling, vesicle internalization (endocytosis), or direct fusion with target cells. However, recent data demonstrate that EV uptake is highly inefficient, with spontaneous internalization estimated at only ~1% after one hour (Bonsergent *et al.*, Nat. Commun. 2021; PMID: 33767144). Of the EVs that are internalized, only a fraction (~30%) successfully achieve cytosolic release of luminal cargo (Bonsergent *et al.*, Nat. Commun. 2021; PMID: 33767144). Thus, only a very small proportion of the luminal contents of sEVs ever reach the cytosol of recipient cells. By contrast, other studies indicate that EVs can modulate cellular responses independently of fusion or luminal cargo delivery, instead acting primarily through modulation of cell-surface receptor activity and downstream signaling cascades (van Niel *et al.* Nat Rev Mol Cell Biol; PMID: 29339798).

Our finding that sEVs are predominantly composed of plasma membrane proteins, with comparatively few bona fide cytosolic proteins encapsulated within the lumen, supports the view that they function as active “signaling

platforms”. In this model, sEVs modulate receptor pathways in a rapid and localized manner that is less dependent of luminal cargo transfer. Notably, our proteomic portrait of sEV composition is consistent with theoretical models showing that, in vesicles below ~180–200 nm in diameter, protein mass is inherently dominated by the membrane rather than the lumen (Zendrini *et al.* Colloids Surf B Biointerfaces 2018; PMID: 35969923). In contrast, larger vesicles (*e.g.*, >800 nm) could accommodate proportionally more luminal cargo, and thus their functional impact may rely more heavily on delivery of internal contents. Consequently, small EVs may be considered more “surface-active” than “luminal-active” (Carney *et al.* Nat Nanotechnol. 2024; PMID: 39468355), with their biological effects determined largely by the abundance and accessibility of surface proteins such as integrins, tetraspanins, and adhesion molecules.

This view of sEVs as “surface-active platforms” has important implications for mechanistic studies of EV-mediated communication. Since membrane proteins are the most likely candidates for stable and functional transfer, we advise caution in interpreting the sEV-mediated delivery of soluble cytosolic or nuclear proteins unless rigorously validated by complementary approaches, such as protease protection assays or membrane disruption experiments. Additional strategies to dissect these mechanisms could include short-term receptor phosphorylation readouts, neutralizing antibodies or decoy ligands to block surface interactions, protease protection analyses to test topology, and perturbations of endosomal fusion to assess dependence on luminal cargo release.

An additional challenge in functional assays with sEVs preparations is that the responses observed in recipient cells may not be exclusively mediated by sEVs, but instead by co-isolated material (Takov *et al.* J Extracell Vesicles 2018; PMID: 30651940). Protein aggregates, lipoproteins, and other extracellular particles can interact with target cells, triggering signaling pathways or immune responses that may be mistakenly attributed to EVs (Busatto *et al.* J Extracell Vesicles 2022; PMID: 35362268). To minimize these confounding effects, we recommend additional validation steps. Functional assays should ideally be performed using density gradient-purified sEVs to reduce co-purification of non-vesicular material. Alternatively, assays may be conducted with sEVs pre-treated with proteinase K to remove non-vesicular contaminants; however, this approach must be applied with caution, as it can also compromise authentic sEV surface proteins.

Biomarker studies

In clinical biofluids, an accurate classification of sEV and NV proteins might allow researchers to prioritize proteins that are mechanistically linked to the disease or the tissue of origin while de-prioritizing those that simply reflect systemic processes such as lipoprotein or complement dynamics. However, we acknowledge the practical limitation that performing density gradient separation for every individual biological or clinical sample is not feasible, particularly in large-scale studies such as biomarker discovery across extensive cohorts. To address this, we propose a practical hybrid strategy that balances methodological rigor with high-throughput capabilities. This approach involves:

- **Initial Reference Profile:** Performing DG-PCP on pooled samples or a representative subset of the cohort (*e.g.*, a balanced mixture of disease and control samples). This critical initial step establishes a robust, study-specific classification of vesicular versus non-vesicular proteins tailored to the precise biological context under investigation.
- **High-Throughput Quantification:** Subsequently, all individual samples can be analyzed at the crude sEV (P100) level without DG separation, allowing for efficient, high-throughput quantification. It should be noted that not all the proteins detected as sEV in the DG will be detected in the P100, as the high abundance of non-vesicular proteins might compromise the sensitivity levels to detect them without this additional separation. This is particularly evident in the case of serum/plasma as we demonstrate in our study.
- **Post-Acquisition Filtering:** The classification data generated from the pooled DG-PCP analysis can then be applied as a custom filter to annotate the proteomes of the individual P100 samples. This enables the prioritization of bona fide sEV cargoes and the confident exclusion of likely contaminants during downstream statistical analysis, thereby significantly increasing the reliability of biomarker candidates identified from large cohorts.

Point-by-Point Response to Reviewer 2

We thank the reviewer for their constructive feedback. In the following point-by-point response, we outline our revisions and the rationale behind them. To facilitate the review, we have used the following colour scheme:

Black: Original reviewer comments.

Blue: Our explanation and response.

Green: Text changes

Reviewer #2 (Remarks to the Author):

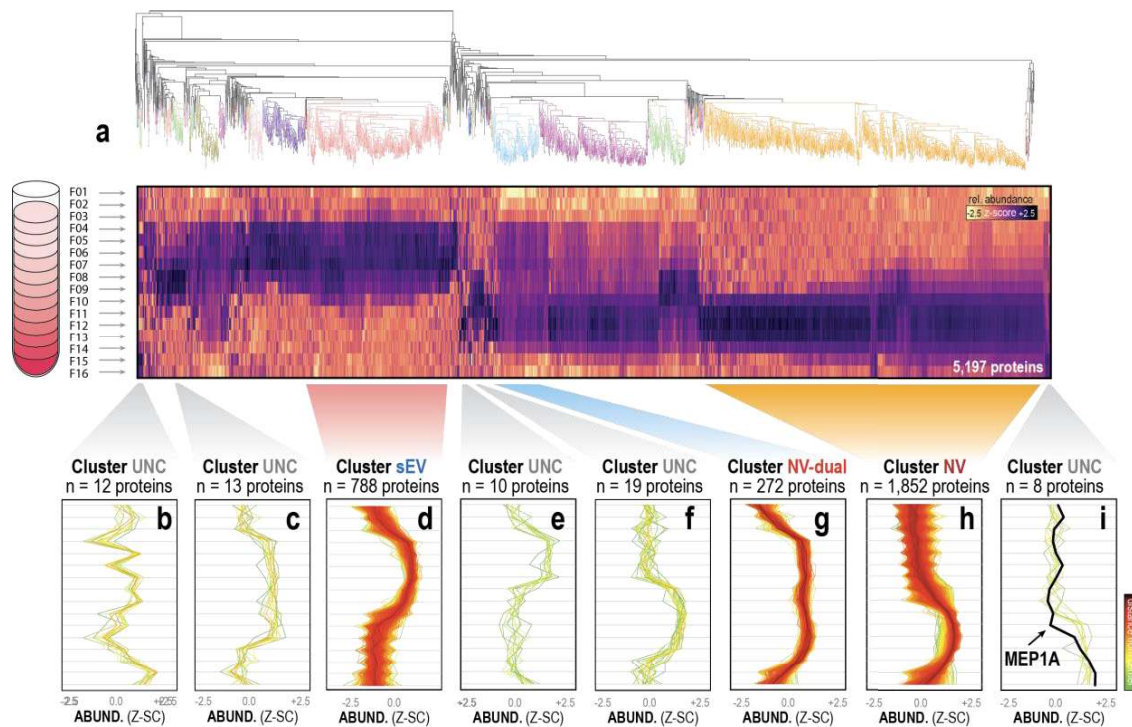
The revised manuscript represents a substantial improvement over the previous version. The authors have addressed most of the previous concerns with detailed explanations and additional analyses, which considerably strengthen the study. The improved interpretation of the molecular composition of sEVs and the clarification of non-vesicular components greatly enhance the manuscript's rigor and impact.

However, one issue remains unresolved. The manuscript still lacks clarification regarding the biological nature of the "unclassified (UNC)" protein group, which accounts for approximately 12% of the identified proteins. Since all analyzed proteins were derived from the P100 fraction—typically considered enriched in sEVs—it is important to specify what these UNC proteins represent. Are they low-abundance or partially associated sEV proteins, non-vesicular contaminants, or intermediate populations with atypical density distributions? Providing a brief analysis or discussion of their possible origins and biological relevance would give readers a more comprehensive understanding of the dataset.

We thank the reviewer for their positive assessment of our revised manuscript and for appreciating the improved rigor of our study. We also appreciate this final request for clarification regarding the "Unclassified" (UNC) protein group. We agree that defining the nature of this fraction is important for the comprehensive transparency of our dataset.

Our protein correlation profiling (PCP) strategy relies on unsupervised hierarchical clustering to group proteins with similar distribution profiles across the density gradient. Specifically, we used the "Define row clusters" function in Perseus to automatically partition the hierarchical tree into discrete clusters (typically 100–200) based on distance thresholds that maximize within-cluster profile similarity. To ensure robustness and reproducibility, we only classified clusters with a minimum size of 20 proteins, serving as a quality control filter to exclude small, low-confidence clusters, which we denoted as "unclassified" (UNC).

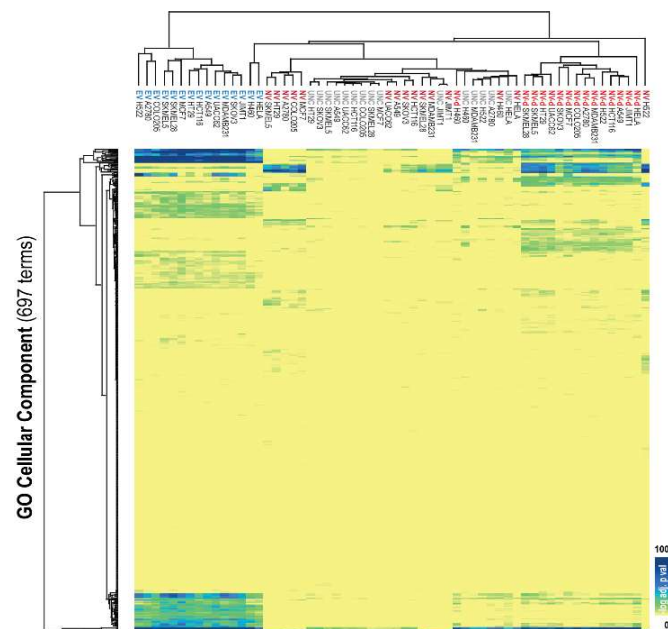
Supporting the validity of this criterion, we found that the assignment of proteins to the UNC category was highly inconsistent across cell lines, strongly suggesting a stochastic origin. Unlike the robust consensus signatures found for the sEV, NV-d and NV groups (**Fig. 2a**), no single protein was consistently annotated as UNC across all 15 cell lines. In fact, only 16 proteins in our entire data set were denoted as UNC in at least 8 cell lines. Among these few exceptions was MEP1A (labeled as UNC in 11 cell lines). Detailed inspection of its elution profile revealed a distinct high-density peak in fractions F11–F12 (**Response Fig. 1i**), consistent with its biochemical capacity to assemble into large noncovalent oligomers of up to 900 kDa (PMID: 11301339). While this specific example confirms that the unclassified group harbours valid biological entities, such as specific macromolecular complexes, our findings indicate that for most proteins, assignment to the 'Unclassified' group is a stochastic event driven by low signal intensity, rather than a stable biological feature.



Response Fig. 1. Examples of clusters classified as sEV, NV-dual and NV or remained unclassified (UNC). (a) Crude sEV preparations (HeLa is shown as a representative example) were separated by density gradient centrifugation. Unsupervised clustering of protein abundance profiles identified distinct elution patterns (labeled with different colors in the dendrogram). Only clusters containing ≥ 20 proteins were retained for manual classification while smaller clusters were annotated as unclassified (UNC). (b) Representative UNC cluster showing irregular or noisy elution profiles. (c, e and f) Examples of UNC clusters exhibiting patterns resembling the main groups. (d, g, h) Examples of major clusters classified as sEV, NV-dual, and NV. (i) Example of an UNC cluster composed of proteins migrating at very high densities, including MEP1A.

Indeed, most UNC clusters consisted of proteins with rather "noisy" profiles (**Response Fig. 1b**) while others resembled the major groups (**Response Fig. 1c, 1e, 1f**), likely bona fide members excluded by the high stringency of our classification. Mechanistically, this fragmentation is driven by low protein abundance. We found that the UNC group is significantly less abundant than the sEV or NV-dual populations, matching the abundance levels of the NV fraction (**Response Fig. 2**). In mass spectrometry, low-abundance proteins exhibit lower signal-to-noise ratios, accumulating minor stochastic deviations across the gradient fractions. Because the Euclidean distance metric used for clustering aggregates these minor deviations, it artificially inflates inter-profile dissimilarity, causing these proteins to partition into small, discrete clusters despite their visual similarity to the consensus patterns. Most importantly, Gene Ontology (GO) enrichment analysis revealed that the UNC proteins functionally cluster with the non-vesicular (NV) fraction rather than the sEV fraction (**Response Fig. 3**).

[REDACTED]



Response Fig. 3. GO enrichment analyses of sEVs, NV-dual, NV and UNC proteins. Proteins classified as sEVs, NV-dual, NV and UNC across cell lines were analyzed using g-GOSt (<https://biit.cs.ut.ee/gprofiler/gost>). Heatmap shows the unsupervised hierarchical clustering using the significance values of GO cellular components identified in the four protein classes of each cell line.

In summary, while our strict filtering criteria may inevitably exclude a minimal number of bona fide sEV proteins due to limitations of the hierarchical clustering algorithm, our data demonstrate that the vast majority of the "Unclassified" fraction represents stochastic, low-abundance non-vesicular contaminants. Therefore, we believe that this threshold serves as an effective quality control step to preserve the purity and biological relevance of the reported sEV signatures. To clarify this in the manuscript, we have:

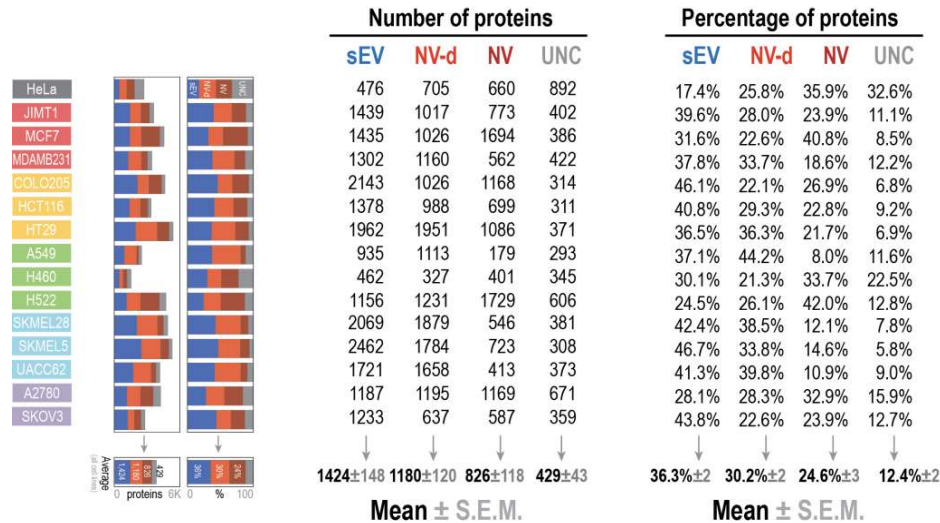
- updated **Fig. 2d** (and its legend) to show the low abundance distribution of the UNC fraction alongside the other groups and added an explanation to the main text (Page 5, first paragraph)
- expanded the **Methods** section ("Data analysis for protein correlation profiling," Page 34, first paragraph) to specify the minimum cluster-size threshold, which is also explained in the main text (Page 3, second paragraph).

Finally, this detailed re-examination of the UNC fraction prompted us to correct a labeling inconsistency in **Fig. 2h**. In this figure, we had originally labeled the grey nodes as "UNC." We wish to clarify that these proteins are not the low-abundance UNC fraction described above. Rather, they represent abundant “non-sEV proteins” that did not meet the strict consistency criteria to be part of the consensus NV or NV-dual signatures (*i.e.*, they were not consistently classified in >80% of cell lines). To avoid confusion, we have corrected the nomenclature in the revised **Fig. 2h**, legend and main text (Page 5, second paragraph) to refer to this group more accurately as "non-sEV proteins."

In addition, clarification is needed for the statement that “the relative percentage of protein profiles for sEV (36%), NV-dual (30%), NV (24%), and UNC (12%) classes were remarkably similar across cell lines.” It is not entirely clear how these percentages were derived—whether they reflect the proportion of total identified proteins, the number of density gradient profiles per class, or normalized averages across cell lines. A concise explanation in the Methods or figure legend would improve the transparency of the quantitative analysis.

We appreciate the reviewer’s careful reading and agree that our original description regarding the derivation of these percentages lacked clarity.

In **Fig. 2b**, the left stacked bar charts show the number of proteins classified as sEV, NV-d, NV, or UNC for each of the 15 cell lines. The right stacked bar charts display the relative percentage of these four classes within each cell line. In the main text, to summarize these results across the dataset, we computed the mean percentage of each class across all cell lines, yielding 36.3% sEV, 30.2% NV-d, 24.6% NV, and 12.4% UNC. The underlying numerical data used for these calculations are detailed below in **Response Fig. 4**.



Response Fig. 4. Proteins annotated as sEV, NV-d, NV and UNC across cell lines. On the left, the stacked bar charts display the number of proteins and their relative percentage for each protein class within each cell line. To summarize these results across the dataset, we computed the mean and standard error of the mean (S.E.M.) of each class across all cell lines. The table on the right lists the exact numerical values used for these plots.

These means were not explicitly shown in the initial figure, which understandably caused confusion. In the new revised **Fig. 2b**, we have now added the mean values for both absolute numbers and percentages at the bottom of the bar charts to ensure full transparency. In addition, we have updated the main text (Page 4, second paragraph) to report these means with S.E.M..