Online Methods

Cell lines and cell culture.

B16-F10, B16-F1, 4T1, MDA-MB-231 series (parental, -1833, -4175, and -831, gifts from Dr. J. Massagué), LLC, SW620, HCT116 (Horizon Discovery), PANC-1, AsPC-1, Pan02 (purchased from the National Cancer Institute Tumor Repository), and NIH3T3 cells were cultured in DMEM. Human melanoma cells (SK-Mel103, A375M and A375P were obtained from MSKCC), human prostatic carcinoma cell lines PC3 and DU145, as well as BXPC-3, HPAF-II, PC-9, ET2B (gift from Dr. P. Gao and J. Bromberg), K-562 (DSMZ) and NB-4 (DSMZ) cells were cultured in RPMI, supplemented with penicillin (100 U/ml) and streptomycin (100 μg/ml) and 10% FBS. Cell lines were obtained from American Type Culture Collection, if not otherwise mentioned, and authenticated using STR profiling by commercial providers. All cells were maintained in a humidified incubator with 5% CO_2 at 37 ^{o}C and routinely tested and confirmed to be free of mycoplasma contamination. When collecting conditioned media for exosome isolation, FBS (Gibco, Thermo Fisher Scientific) was first depleted of exosomes by ultracentrifugation at 100 000 x g for 90 min. Cells were cultured for 3 days before supernatant collection.

Human specimens and processing

Fresh human tumor tissues were obtained from subjects with stage 1–3 melanoma at Memorial Sloan-Kettering Cancer Center (MSKCC) and had histologically confirmed melanoma. All individuals provided informed consent for tissue donation according to a protocol approved by the institutional review board of MSKCC (IRB# 11-033A, MSKCC; IRB#0604008488, WCM), and the study is compliant with all relevant ethical regulations regarding research involving

human participants. Tissues were cut into small pieces and cultured for 24 h in serum-free RPMI supplemented with penicillin/streptomycin. Conditioned media was processed for exosome isolation and AF4 fractionation as described below.

Exomere and exosome isolation and nanosight tracking analysis (NTA).

SEV were prepared using differential ultracentrifugation methods[41] and resuspended in phosphate buffered saline (PBS, pH7.4) for subsequent analysis and AF4 fractionation. Isolated samples were quantified using BCA assay (Pierce, Thermo Fisher Scientific). NTA analysis of exosome size and particle number was performed using the LM10 or DS500 NanoSight system (Malvern Instruments) equipped with a blue laser (405 nm) following manufacturer's instructions.

AF4 fractionation.

The detailed step-by-step AF4 fractionation protocol including sample preparation, AF4 setting parameters and running method, data collection and analysis, and fraction collection and characterization) is provided on ProtocolExchange[42].

Transmission electron microscopy (TEM) and atomic force microscopy (AFM).

For negative staining TEM analysis, 5 µl of sample solution was placed on a formvar/carbon coated grid and allowed to settle for 1 min. The sample was blotted and negative stained with 4 successive drops of 1.5% (aqu) uranyl actate, blotting between each drop. Following the last drop of stain, the grid was blotted and air-dried. Grids were imaged with a JEOL JSM 1400

(JEOL, USA, Ltd, Peabody, MA) transmission electron microscope operating at 100 Kv. Images were captured on a Veleta 2K x 2K CCD camera (Olympus-SIS, Munich, Germany).

For AFM, dilutions were made for each sample and then plated on freshly cleaved mica substrate (SPI) for ~2 min before washing with 10 mL of Molecular Biology Grade H_2O (Fisher BP2819-1) and being blown dry with nitrogen gas. Imaging was performed using an MFP-3D-BIO AFM (Asylum Research), with an Olympus AC240TS-R3 AFM probe (Asylum Research) in tapping mode at room temperature. Images were captured at 1 μm x 1 μm. Image analysis was performed using a custom-written ImageJ/FIJI (NIH) code.

Zeta potential measurement

Fractionated samples were diluted in PBS (Phosphate-buffered saline; 0.01 M phosphate buffer, 0.0027 M KCl, 0.137 M NaCl; pH 7.4 tablets, Sigma) for ζ potential analysis using Zetasizer Nano ZS90 (Malvern Instruments). Samples were freshly prepared prior to loading onto the instrument at a 90° angle (respective to the light source). All experiments were performed at a constant temperature of 25°C.

Stiffness measurement

Freshly cleaved mica coverslips were first coated with Poly-L-lysine (0.1%. w/v in H_2O) for 30 minutes and then incubated with samples on the mica surface for 45 minutes. The samples were then rinsed with 1 ml of MilliPure water, washed three times with PBS buffer, then emerged in a drop of PBS on the mica surface. A stand-alone MFP-3D atomic force microscope (Asylum Research, Santa Barbara, CA) was utilized to perform the analysis. The spring constant of cantilever was determined as 559.73 pN/nm by the thermal noise method[43]. The curvature radius

of cantilever was ~15 nm, and the resonant frequency of 325 kHz were used for the stiffness analysis (i.e., indentation of cantilever) and imaging. Force measurements were performed with an approximate force distance of 300 nm and velocity of 500 nm/s.

Western blot analysis.

Whole cell extract (WCE) and exosome fractions were lysed directly with SDS sample buffer and lysates were cleared by centrifugation at 14 000 x *g* for 10 min. 100 μg of WCE and 10 μg of input and each nanoparticle subset were separated on a Novex 4–12% Bis-Tris Plus Gel (Life Technologies), and transferred onto a PVDF membrane (Millipore). Membranes were blocked for 1 h at RT followed by primary antibody incubation overnight at 4°C. The following antibodies were used for western blot analysis: anti-Tsg101 (Santa Cruz sc-7964); anti-Alix1 (Cell Signaling 2171); anti-Hsp90 (Stressgen ADI-SPA-830-F), anti-MAT1A1 (Abcam ab174687); anti-IDH1 (Proteintech 23309-1-AP); anti-FLOT1 (BD Biosciences 610820); anti-TOLLIP (Abcam ab187198); anti-VPS4B (Santa Cruz sc-32922); anti-DNAJA1 (Abcam ab126774); anti-HSPA8/HSC70 (LifeSpan Biosciences LS-C312344-100). All primary antibodies were used at 1:1,000x dilution. IRDye 800 CW Goat-anti-mouse IgG (LI-COR Biosciences P/N 926-32210, 1:15,000x dilution), HRP-linked Sheep-anti-Mouse IgG (GE Healthcare Life Sciences NA931, 1:2,500x dilution), and HRP-linked Donkey-anti-Rabbit IgG (GE Healthcare Life Sciences NA934, 1:2,500x dilution) were used as secondary antibody. The blot was analyzed either using the Odyssey Imaging system (LI-COR Biosciences) or enhanced chemiluminescence substrates (Thermo Fisher Scientific).

Analysis of Proteomic Profiling Data.

Protein mass spectrometry analyses of fractionated exosomes were performed at the Rockefeller University Proteomics Resource Center as described previously[44, 45], and conducted on two independent biological replicates for each sample (exomere, Exo-S and Exo-L) derived from 5 different cell lines (B16-F10, 4T1, Pan02, AsPC-1 and MDA-MB-4175). Raw data were provided in Supplementary Table 2.

For proteomic data processing and Principal Component Analysis (PCA), the proteomic expression data was processed using the 'Limma' package of the R program (https://www.r-project.org, v3.2.5). Proteomic expression data was imported and was normalized using 'normalizeBetweenArrays' function (method=quantile)[46]. PCA was performed for data reduction, simplifying datasets to three dimensions for plotting purposes using 'princomp()' function with default options, and illustrated using the 'rgl' package and 'plot3d()' function.

For clustering and marker selection, Consensus clustering analysis, marker selection for each fraction, and heatmap generation were conducted using GENE-E software (http://www.broadinstitute.org/software/gene-e). Consensus clustering was conducted to assess whether proteomic expression differs between fraction[47]. To identify fraction-specific markers, the probe (based on UniProt ID) values were collapsed to protein-level using maximum probe. Only proteins detected in both replicates of a sample were included for further analysis. Proteins were sorted by signal-to-noise statistic, $(\mu_A - \mu_B)/(\alpha_A + \alpha_B)$ where μ and α represent the mean and standard deviation of proteomic expression, respectively, for each class[48]. Next, the signal to noise marker selection tool from GENE-E was used to identify fraction-specific markers with 1,000 permutations. The cutoff to select fraction-specific markers was fold change ≥ 5, false discovery rate (FDR) <0.05, and mean protein expression $\geq 10^8$ with the positivity in $\geq 80\%$ (i.e. at least 4 out 5 samples from 5 cell lines for each nanoparticle subset) of the corresponding

fraction. Heat maps for visualization of differential protein expression patterns were generated for 65 markers (39 exomere-specific markers; 5 Exo-S markers; 21 Exo-L markers) using GENE-E with relative color scheme (by subtracting each mean protein expression, divide by each standard deviation for each row).

For Gene Set Enrichment Analysis (GSEA) we used the entire proteomic expression data set[49]. Gene sets from Molecular signatures database (MSigDB, http://www.broadinstitute.org/gsea/msigdb/index.jsp) v5.1 were used for GSEA (H: 50 hallmark gene sets; C2:KEGG: 186 canonical pathways from Kyoto Encyclopedia of Genes and Genomes [KEGG] pathway database; C5: 825 gene sets based on Gene Ontology [GO] term)[50]. The default parameters were used to identify significantly enriched gene-sets (FDR q <0.25).

Glycoprotein extraction and lectin blotting

Nanoparticles were lysed with RapiGest SF (Waters) containing 1 mM sodium orthovanadate and protease inhibitor cocktail (Roche), for 30 minutes on ice and centrifuged at 16 000 x *g* for 20 min. For lectin blotting 0.5µg of total protein extracts were separated using 4-15% gradient gels (Biorad) and transferred onto nitrocellulose membranes. Samples were incubated at room temperature (RT) for 1h with the following biotinylated lectins *Aleuria aurantia* Lectin (AAL; Fucα6GlcNAc and Fucα3GlcNAc), *Sambucus nigra* Lectin (SNA; Neu5Acα6(Gal or GalNAc)), *Phaseolus vulgaris* Leucoagglutinin (L-PHA; Galβ4GlcNAcβ6(GlcNAcβ2Manα3)Manα3), and *Phaseolus vulgaris* Erythroagglutinin (E-PHA; Galβ4GlcNAcβ2Manα6(GlcNAcβ4)(GlcNAcβ4Manα3)Manβ4) (Vector Laboratories, 1:2000 dilution except 1:1000 dilution for L-PHA). Vectastain Elite ABC HRP Kit (Vector Laboratories) was used for signal detection with ECL enhanced chemiluminescence technique (GE Healthcare

Life Sciences). The total protein profile of the samples was assessed in parallel on a silver-stained gel (Supplementary figure 5a). (Abbreviations: Fuc, fucose; GlcNAc, N-acetylglucosamine; Man, mannose; Neu5Ac, neuraminic acid; Gal, galactose; GalNAc, N-acetylgalactosamine.)

Glycomics analysis

The glycoproteins extracts from the different fractions were reduced, alkylated and digested with sequencing-grade, modified trypsin (Promega) using a standard proteomics protocol[51]. The *N*-glycans were analyzed based on a modification of Jensen *et al*[52]. Briefly, *N*-Linked glycans were released with PNGase F (*Elizabethkingia* meningoseptica; Sigma), deaminated and partially purified using porous graphitized carbon solid-phase extraction cartridges (PGC-SPE, HyperSep-96-Hypercarb, 25 mg, Thermo Scientific) as described previously[53].

Glycan profiling and characterization was performed by MALDI TOF/TOF mass spectrometry (4800 Plus, SCIEX) using alpha-cyano-4-hydroxycinnamic acid (CHCA; 10 mg/mL in 50% ACN), operated in reflector negative mode (mass range of m/z 1000 to 5000) with external calibration (TOF/TOF calibration mixture, SCIEX). Three independent analytical measurements were performed. NanoHPLC- High Resolution Mass Spectrometry (HRMS) was used to validate the presence of most discriminative ions in MALDI-MS spectra using a nanoHPLC system (Dionex, 3000 Ultimate RSLCnano) coupled on-line to a LTQ-Orbitrap XL mass spectrometer (Thermo Scientific) equipped with a nano-electrospray ion source (Thermo Scientific, EASY-Spray source). *N*-Glycan chromatographic separation using Porous Graphitized Carbon (PGC) was adapted from a procedure previously described[53]. We have combined in series a nanoflow PGC column (Hypercarb, 150 mm × 75 µm ID, 3 µm particle size, Thermo

Scientific) followed by a reversed phase C18 column (EASY-Spray C18 PepMap, 100 Å, 150 mm x 75μm ID and 3 μm particle size, Thermo Scientific). This allowed a better separation of carbohydrates and remaining tryptic peptides, while minimizing salt precipitation events encountered when a nanospray emitter was utilized directly after the PGC column. The mass spectrometer was operated in negative ion mode.

The monosaccharide compositions for the glycan precursors on MALDI-MS spectra were predicted using the GlycoMod tool (http://www.expasy.ch/tools/glycomod) considering mass accuracies bellow 10 ppm. The possibility of neutral exchanges with Na^+ and K^+ was considered for sialoglycans. The glycan structures were assigned based on nanoHPLC–PGC-HRMS analysis considering: i) molecular monoisotopic mass; (ii) CID-MS/MS *de novo* sequencing; and (iii) PGC-LC relative retention times. In particular, α2,3-linked and α2,6-linked sialylated *N*-glycans were differentiated based on retention time (α2,6 < α2,3)[52]. For further validation, MS/MS fragmentation profiles were matched to glycosidic fragments calculated *in silico* on GlycoWorkBench (http://www.eurocarbdb.org/applications/ms-tools) [54]. General understanding of mammalian *N*-glycosylation was used to determine some structural aspects, yet some structural ambiguity remained in a subset of the reported *N*-glycans as indicated with brackets. A semiquantitive approach was used to compare glycan compositions based on MALDI-MS assignments, taking into account the monoisotopic peak intensity. Glycan standards and negative controls were analyzed in parallel. These results were validated based on the intensity of each specie on nanoHPLC-HRMS ion chromatograms (EIC) (m/z ± 0.01).

Lipidomics: sample preparation, mass spectrometry and data analysis

Equal amount of each sample (based on BCA quantification) was subjected to lipidomic analysis. Samples were first sonicated with a Model Q700 QSonica sonicator equipped with an Oasis 180 Chiller (4°C; Amplitude, 95; process, 5 min; pulse-on 30 sec; plus-off 55 sec), centrifuged at 14,800 rpm for 10 min at 4°C, and 50 µL of the extract supernatant was spiked with 2 µL 50 µg/mL internal standard mixture (Cer 18:1/12:0; PC 12:0/12:0; PE 14:0/14:0; PG 14:0/14:0; PS 14:0/14:0). Subsequently, the samples were analyzed by using the Thermo Q-Exactive MS system (Bremen, Germany) in the Metabolomics Laboratory of Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign. Software Xcalibur 3.0.63 was used for data acquisition and analysis. The Dionex Ultimate 3000 series HPLC system (Thermo, Germering, Germany) was used, and the LC separation was performed on a Thermo Accucore C18 column (2.1 x 150 mm, 2.6 µm) with mobile phase A (60% acetonitrile: 40% H_2O with 10 mM ammonium formate and 0.1% formic acid) and mobile phase B (90% isopropanol: 10% aceontrile with 10 mM ammonium formate and 0.1% formic acid) and a flow rate of 0.4 mL/min. The linear gradient was as follows: 0 min, 70% A; 4 min, 55% A; 12 min, 35% A; 18 min, 15% A; 20 -25 min, 0% A; 26-33 min, 70% A. The autosampler was set to 15°C and the column was kept at 45°C. The injection volume was 10 µL. Mass spectra were acquired under both positive (sheath gas flow rate, 50; aux gas flow rate: 13; sweep gas flow rate, 3; spray voltage, 3.5 kV; capillary temperature, 263 °C; Aux gas heater temperature, 425 °C) and negative electrospray ionization (sheath gas flow rate, 50; aux gas flow rate: 13; sweep gas flow rate, 3; spray voltage, -2.5 kV; capillary temperature, 263 °C; Aux gas heater temperature, 425 °C). The full scan mass spectrum resolution was set to 70,000 with the scan range of m/z 230 ~ m/z 1,600, and the AGC target was 1E6 with a maximum injection time of 200 msec. For MS/MS scan, the mass spectrum resolution was set to 17,500, and the AGC target was 5E4 with a maximum

injection time of 50 msec. Loop count was 10. Isolation window was 1.0 m/z with NCE of 25 and 30 eV. For data analysis, LipidSearch (v.4.1.30, Thermo) was used for lipid identification. The lipid signal responses were normalized to the corresponding internal standard signal response. For those lipid classes without corresponding internal standard, positive lipid ion signals were normalized with the signal of internal standard Cer 18:1/12:0 and negative ion signals were normalized with the signal of internal standard PG 14:0/14:0. The percentage of lipid classes within a sample was calculated by adding that of each of the individual molecular species quantified within a specific lipid class, and the relative abundance was represented by the mean percentage of 3 replicates for each group of samples. Differences among different subpopulations of particles derived from the same cell line were analyzed using ANOVA test ($q < 0.05$).

Nucleic acid analysis

DNA was extracted from nanoparticles using the AMPure XP beads (Agencourt) following the manufacturer's protocol. An equal volume of nanoparticles in PBS and lysis buffer AL (QIAGEN) were mixed and incubated with Proteinase K (20 µg/ml, QIAGEN) at 56 ˚C for 10 min. The mixture was mixed with one volume of each, AMPure beads, isopropanol and PEG solution (Beckman), and incubated for 5 min at RT. DNA bound to the beads was then separated from the solution/supernatant on magnet for 5 min at RT. The supernatant was removed by pipetting and bead-bound DNA was washed twice with freshly prepared 80% ethanol, then air dried for 5 min. Lastly, DNA was eluted from beads with nuclease free water and quantified using QuBit assay (Life technology). DNA extraction was performed for two independent biological replicates of each sample.

RNA was extracted using the Ambion mirVarna kit (Life techonology), following the manufacturer's protocol with one modification: one volume of nanoparticles in PBS was first lysed with 7 volumes of lysis buffer. The samples were analyzed using Agilent Total RNA Pico kits. RNA extraction was performed for two independent biological replicates of each sample.

Biodistribution assessment

Fractionated nanoparticles were first labeled with the near infrared dye CellVue NIR815 (eBioscience) following manufacturer's protocol, followed by washing with 20 ml of PBS and pelleting by ultracentrifugation at 100,000 x *g* for 70 min at 10 ^{o}C. 10 μg of labeled nanovasicles resuspended in 100 μl of PBS, or an equivalent volume of mock reaction mixture was retro-orbitally injected into naïve mice (6-week-old female C57BL/6 mice purchased from Jackson Labs). 24 h post injection, tissues were collected and analyzed using the Odyssey imaging system (LI-COR Biosciences). Two independent experiments with 3 animals per group were performed. No statistical method was used to predetermine sample size. The experiments were neither randomized, nor blinded. All animal experiments were performed in compliance with ethical regulations and in accordance with Weill Cornell Medicine institutional, IACUC and AAALAS guidelines, approved for animal protocol 0709-666A.

Code Availability

Custom-written ImageJ/FIJI (NIH) code for AFM image analysis is fully available upon request.

Statistics and Reproducibility

Error bars in graphical data represent means ± SEM. Statistical significance is determined using one way ANOVA. $P < 0.05$ was considered statistically significant. Statistical analyses were performed using GraphPad Prism software. For lipid class analysis, ANOVA test ($q < 0.05$) was performed using Qlucore Omics Explorer (Sweden). For proteomic analysis, proteins were sorted by signal-to-noise statistic, $(\mu A - \mu B)/(\alpha A + \alpha B)$ where μ and α represent the mean and standard deviation of proteomic expression, respectively. The cutoff to select fraction-specific markers was fold change ≥5, false discovery rate (FDR) <0.05, and mean protein expression $\geq 10^8$ with the positivity in ≥80% (i.e. at least 4 out 5 samples from 5 different cell lines for each subset of nanoparticles) of the corresponding fraction. For GSEA, Kolmogorov-Smirnov statistic was calculated to evaluate whether proteins from a pre-determined pathway are significantly overrepresented towards the top or bottom of the ranked gene list (FDR $q < 0.25$).

Multiple AF4 analyses were performed for each cell line studied in this work: B16-F10, > 50x (repeated times); AsPC-1, 9x; Pan02, 16x; MDA-MB-4175 (4175), 17x; and 4T1, 10x. TEM imaging analysis of fractionated particles were conducted for B16-F10, 7x; AsPC-1, 3x; Pan02, 2x; 4175, 1x; and 4T1, 4x. Four independent human melanoma specimens were analyzed using AF4 and two of them were analyzed by TEM. Proteomic profiling of exomeres, Exo-S and Exo-L was performed on two biologically independent samples of each particle derived from five different cell lines (B16-F10; AsPC-1; Pan02; 4175; and 4T1). Western blotting validation of specific signature proteins of each particle subtype was done once (noted in the legend for Fig. 1d). For *N*-glycan study, lectin blotting was repeated independently twice except for AAL and E-PHA blotting for B16-F10 and 4175 which were done once (Fig. 4a). Glycomic MS was performed on two biologically independent B16-

F10 samples and one sample of AsPC-1 and 4175 (Supplementary Fig. 5b-d). Quantification of top 6 most abundant glycans was based on 3 independent analytical measurements of one experiment (Fig. 4c, Supplementary Fig. 5c and d). Silver stained-PAGE analysis was repeated independently twice for B16-F10 and 4175 and once for AsPC-1 (Supplementary Fig. 5a). NanoHPLC-PGC-HRMS was done once (Supplementary Fig. 5 e-i). Lipidomic analysis was conducted on 3 biologically independent samples. DNA and RNA analyses of each particle subtype were repeated twice. Organ biodistribution analysis of each particle subtype was repeated 4x independently. NTA analysis was conducted using 3 biologically independent samples. TEM analysis was repeated 3 times for AF4 peaks P1 and P5 and once for HDL, LDL and VLDL (Supplementary Fig. 4d). AF4 analysis of B16-F10 sEVs collected from technical and biological replicates, and samples kept at either 4 ^{o}C or -80 ^{o}C were repeated independently 3 times, cells of different passage numbers twice, and under hypoxic *versus* normoxic conditions was repeated with 3 different cell lines independently. AF4 and TEM analysis of particles isolated from the blank media control and CM of B16-F10 and 4175 was done once (Supplementary Fig. 1 j and k).

Independent measurements of hydrodynamic diameters of exomeres, Exo-S and Exo-L derived from different cell lines in batch mode were repeated (in the order of exomere, Exo-S and Exo-L): B16-F10 (n=10, 9, and 8 independent measurements, respectively); Pan02 (*n*=11, 6, 11); AsPC-1 (*n*=5, 5, 5); 4175 (*n*=3, 5, 3); 4T1 (*n*=5, 5, 5)). For zeta potential, independent measurements were repeated: B16-F10 (*n*=8, 10, 12); Pan02 (*n*=13, 11, 13); AsPC-1 (*n*=12, 12, 12); 4175 (*n*=17, 9, 6); 4T1 (*n*=13, 3, 9). For stiffness, B16-F10 (*n*=6, 6, 6); Pan02 (*n*=6, 6, 6); AsPC-1 (*n*=21, 19, 16); 4175 (*n*=11, 10, 5); 4T1 (*n*=9, 8, 9). For AFM imaging analysis of the

height of exomeres: B16F10 (*n*=754 particles analyzed), AsPC1 (*n*=475) and 4175 (*n*=160). AFM imaging of exomeres was repeated with samples derived from 3 different cell lines.

For all experiments described above, all attempts at replication were successful with similar results.

Data Availability

The datasets for proteomic analysis of exomeres, Exo-S and Exo-L subpopulations derived from various cancer cell lines (Supplementary Table 2) have been deposited in https://figshare.com/s/302419bafecaae26b653.

Proteins that are uniquely associated with or among the top 50 most abundant proteins in exomere, Exo-S and Exo-L derived from different cancer cell lines (Supplementary Table 4) have been deposited in the figshare database: https://figshare.com/s/5081b49c6716bbc8d630. Proteomics analysis of lipoprotein particles (Supplementary Table 5) has been deposited in the figshare database: https://figshare.com/s/031571ce9dd63aca4529. Gene set enrichment analysis (GSEA) of proteins associated with exomeres, Exo-S and Exo-L derived from various cancer cell lines (Supplementary Table 6) has been deposited in the figshare database: https://figshare.com/s/633ffe2120e23acc076d. Lipid classes identified in exomeres and exosome subsets derived from different cell lines (raw data and normalized data, Supplementary Table 7) have been deposited in the Figshare: https://figshare.com/s/0573bf5335bb46ee895e. Source data for Fig. 1e, 1g, 2a-b, 2d, 4c, 5a-b, 6a-b, 7b and Supplementary Fig. 1b, 1d, 4c, 5c-d have been provided as Supplementary Table 8. All other data supporting the findings of this study are available from the corresponding authors upon reasonable request.

41. Peinado, H. *et al.* Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. *Nature medicine* **18**, 883-891 (2012).
42. Zhang, H. & Lyden, D. A protocol for Asymmetric-Flow Field-Flow Fractionation (AF4) of small extracellular vesicles. *Protocol Exchange* (2018). DOI:
43. Langlois, E.D., Shaw, G.A., Kramar, J.A., Pratt, J.R. & Hurley, D.C. Spring constant calibration of atomic force microscopy cantilevers with a piezosensor transfer standard. *The Review of scientific instruments* **78**, 093705 (2007).
44. Costa-Silva, B. *et al.* Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. *Nature cell biology* **17**, 816-826 (2015).
45. Hoshino, A. *et al.* Tumour exosome integrins determine organotropic metastasis. *Nature* **527**, 329-335 (2015).
46. Bolstad, B.M., Irizarry, R.A., Astrand, M. & Speed, T.P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185-193 (2003).
47. Monti, S., Tamayo, P., Mesirov, J. & Golub, T. Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Mach Learn* **52**, 91-118 (2003).
48. Golub, T.R. *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531-537 (1999).
49. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).
50. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739-1740 (2011).
51. Ferreira, J.A. *et al.* Synthesis and optimization of lectin functionalized nanoprobes for the selective recovery of glycoproteins from human body fluids. *Analytical chemistry* **83**, 7035-7043 (2011).
52. Kolarich, D., Windwarder, M., Alagesan, K. & Altmann, F. Isomer-Specific Analysis of Released N-Glycans by LC-ESI MS/MS with Porous Graphitized Carbon. *Methods in molecular biology* **1321**, 427-435 (2015).
53. Jensen, P.H., Karlsson, N.G., Kolarich, D. & Packer, N.H. Structural analysis of N- and O-glycans released from glycoproteins. *Nature protocols* **7**, 1299-1310 (2012).
54. Ceroni, A. *et al.* GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *Journal of proteome research* **7**, 1650-1659 (2008).