Characterizing single extracellular vesicles by droplet barcode sequencing for protein analysis

Mahsan Banijamali¹ | Pontus Höjer¹ | Abel Nagy² | Petra Häåg³ |
Elizabeth Paz Gomero² | Christiane Stiller² | Vitaliy O. Kaminskyy³,⁴ | Simon Ekman³,⁵ |
Rolf Lewensohn³,⁵ | Amelie Eriksson Karlström² | Kristina Viktorsson³ |
Afshin Ahmadian¹

¹Royal Institute of Technology (KTH), School of Engineering Sciences in Chemistry, Biotechnology and Health, Department of Gene Technology, Science for Life Laboratory, Solna, Sweden
²Royal Institute of Technology (KTH), School of Engineering Sciences in Chemistry, Biotechnology and Health, Department of Protein Science, AlbaNova University Center, Stockholm, Sweden
³Department of Oncology-Pathology, Karolinska Institutet, Solna, Sweden
⁴Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden
⁵Theme Cancer, Medical Unit head and Neck, Lung, and Skin Tumors, Thoracic Oncology Center, Karolinska University Hospital, Solna, Sweden

Correspondence
Afshin Ahmadian, Royal Institute of Technology (KTH), School of Engineering Sciences in Chemistry, Biotechnology and Health, Department of Gene Technology, Science for Life Laboratory, SE-171 65, Solna, Sweden. Email: afshin.ahmadian@scilifelab.se

Abstract
Small extracellular vesicles (sEVs) have in recent years evolved as a source of biomarkers for disease diagnosis and therapeutic follow up. sEV samples derived from multicellular organisms exhibit a high heterogeneous repertoire of vesicles which current methods based on ensemble measurements cannot capture. In this work we present droplet barcode sequencing for protein analysis (DBS-Pro) to profile surface proteins on individual sEVs, facilitating identification of sEV-subtypes within and between samples. The method allows for analysis of multiple proteins through use of DNA barcoded affinity reagents and sequencing as readout. High throughput single vesicle profiling is enabled through compartmentalization of individual sEVs in emulsion droplets followed by droplet barcoding through PCR. In this proof-of-concept study we demonstrate that DBS-Pro allows for analysis of single sEVs, with a mixing rate below 2%. A total of over 120,000 individual sEVs obtained from a NSCLC cell line and from malignant pleural effusion (MPE) fluid of NSCLC patients have been analyzed based on their surface proteins. We also show that the method enables single vesicle surface protein profiling and by extension characterization of sEV-subtypes, which is essential to identify the cellular origin of vesicles in heterogenous samples.

KEYWORDS
droplet barcode sequencing (DBS), droplet barcode sequencing for protein analysis (DBS-Pro), protein profiling, sEV subtypes, single vesicle, small extracellular vesicles (sEVs), surface protein

1 | INTRODUCTION

Small extracellular vesicles (sEVs) are nanometre-sized vesicles secreted by cells into, for example, body fluids (György et al., 2011; Holcar et al., 2020; Pegtel & Gould, 2019). sEVs with their rich molecular cargo are involved in cell signalling and communication through transport of bioactive material and are implicated in signalling of numerous diseases, including cancer (Fang et al., 2013; Villagrasa et al., 2015; Yáñez-Mó et al., 2015). This has provided the impetus for developing technologies to characterize sEVs (Gandham et al., 2020; Shao et al., 2018). However, considering that sEVs are secreted by a variety of cells, an extremely heterogeneous repertoire of vesicles is present in biofluids. Methods for analysis based on bulk measurements of sEVs give an average quantification. Consequently, these methods do not provide information on cellular heterogeneity and cannot identify
cellular origin of sEVs. To further utilize sEVs as messengers from a specific tissue or tissue compartment there is a need for reliable and high throughput technologies that can identify sEV-populations at individual level. Several single-sEV profiling technologies have been developed in recent years (Cavallaro et al., 2021; Fraser et al., 2019; Gandham et al., 2020; Ko et al., 2020, 2021; Lee et al., 2018; Liu et al., 2018; Löf et al., 2016; Pick et al., 2018; Rissin et al., 2010; van der Vlist et al., 2012; Wang et al., 2020; Wu et al., 2019) and the common denominator for these assays is the use of immuno-reagents to target surface protein markers. In general, technologies for single-sEV characterization fall into three categories. The first uses flow cytometry as the readout (Löf et al., 2016; Pick et al., 2018; van der Vlist et al., 2012; Wang et al., 2020) while the second group relies on fluorescence microscopic imaging principles (Cavallaro et al., 2021; Fraser et al., 2019; Lee et al., 2018; Wang et al., 2020). Both these detection principles have limited potential for multiplexing due to spectral overlap between fluorophores, thus impeding their wide use. The third group of technologies are based on compartmentalization, and the assays belonging to this category may have entirely different readouts. Digital enzyme-linked immunosorbent assay (dELISA) (Pérez-Ruiz et al., 2018; Rissin et al., 2010; Rodero et al., 2017) is an alternative method to conventional ELISA with up to 1000-fold higher sensitivity. dELISA has been further developed to “droplet digital ExoELISA” for counting sEVs and quantifying surface proteins (Liu et al., 2018). While dELISA utilizes tiny reaction chambers to separate beads that carry sandwich complexes, droplet digital ExoELISA is designed to encapsulate and separate single sEVs in droplets. This method exhibits high specificity and sensitivity but because the detection is based on fluorescence signal, the multiplexing possibilities are limited. The digital PCR (dPCR) approach (Cao et al., 2017; Diehl et al., 2006) has also been further developed for analysis of single sEVs (Ko et al., 2020). This method called immuno-droplet digital PCR (iddPCR) amplifies oligonucleotide-antibody conjugates attached to sEVs in emulsion droplets. The method has excellent sensitivity but like the above-mentioned assays suffers from low multiplexing due to fluorescent readout. This limitation can be addressed by combining compartmentalization with high-throughput sequencing. In recent years, two elegant single-sEV methods with sequencing readouts have been presented with the potential to not only increase the number of targets but also allow more precise quantification of surface molecules as potential biomarkers (Ko et al., 2021; Wu et al., 2019). The proximity barcoding assay (PBA) (Wu et al., 2019) employs the concept of generating “virtual compartments” on a flat surface to incorporate rolling circle amplified barcodes to DNA-barcoded antibodies on single-sEVs. The single EV immunosequencing (seiSEQ) technology (Ko et al., 2021) on the other hand utilizes a microfluidic-based droplet-generator to encapsulate and link bead-derived DNA barcodes to individual antibody-EV complexes. The use of massively parallel sequencing (MPS) technologies in combination with molecular identifiers in PBA and seiSEQ has led to a more reliable quantification of surface proteins as compared to fluorescence sensing strategies. In the PBA article, the authors demonstrate that they could identify and distinguish the origin of different cancer cell lines based on their surface protein profile.

In this work, we present a new, alternative single-sEV assay based on droplet barcode sequencing (DBS) technology. We have previously demonstrated the use of DBS for targeted DNA phasing (Borgström et al., 2015; Redin et al., 2017), genome-wide haplotyping (Redin et al., 2019) and protein analysis (Stiller et al., 2019). Here we have further developed the DBS technology to profile surface proteins of single sEVs. The method is scalable and does not require specialized devices nor barcoded gel beads and thus may be performed in any laboratory setting. In this proof-of-concept study we report the profiling of multiple surface proteins for characterizing tens of thousands single-sEVs in the non-small cell lung cancer (NSCLC) cell line H1975 and in three malignant pleural effusion (MPE) fluid samples isolated from advanced NSCLC patients with diverse tumour genomic alterations.

2 | MATERIALS AND METHODS

2.1 | Cell culture conditions and EV isolation

In this study the NSCLC cell line H1975 (ATCC® CRL-5908™, LGC Standards, Teddington Middlesex, UK) was used as source of EVs. H1975 is a cell line of adenocarcinoma histology, with two defined mutations in Epidermal growth factor receptor (EGFR) (Cross et al., 2014). The mutations in EGFR in H1975 involve the kinase domain which is confined to the intracellular part of the protein and these mutations do not have a negative impact on the affinity reagent used in the study. The H1975 cells were maintained in RPMI-1640 medium to which 10% fetal bovine serum (FBS) was added and with 2 mM L-glutamine (Gibco, Life Technologies, Stockholm, Sweden) as additional supplement. To avoid that EVs (i.e., exosomes) within regular FBS would blur the EVs isolated from media of cultured cells, cells were put for 2 days in media with 10% of “exosome-depleted” FBS (#Gibco™ A2720801, Fisher Scientific, Gothenburg, Sweden). The EVs were isolated from about 75–100 ml cell culture media. The media was centrifuged for 5 min at 200 RCF followed by another centrifugation at 720 RCF for 20 min to clear out cell debris and other cell material. The resulting supernatant was thereafter concentrated into ~500 μl using Amicon Ultra-15 Centrifugal Filter Unit (MWCO of 3 kDa (#UFC900324, Merck Chemicals and Life Science AB, Solna, Sweden) according to manufacturer’s instruction. For isolation of EVs size-exclusion chromatography (SEC) on qEVoriginal/70 nm columns (Izon Science, Oxford, UK) were used as per manufacturer’s instruction and as described in Stiller et al. (2021). In brief, the columns were rinsed in PBS to get rid of perseverative solution prior to adding the sample on top of the column. The samples were eluted using filtered
PBS in 500 μl fractions. PBS was filtered with an Acrodisc syringe filter (0.2 μm Supor membrane, Ø 32 mm, 28143–350, Pall Corporation, distributed via VWR, Spånga, Sweden). Prior isolations of EVs from cell culture media revealed exosome sized vesicles in fractions 6–10. The content of fractions 6–10 were pooled and the resulting 2.5 ml pooled fractions were reduced to 500 μl using Amicon® Ultra-4 Centrifugal Filter Unit (MWCO of 3 kDa, #UFC800324, Merck Chemicals and Life Science AB). Particle size and amount was investigated using nanoparticle tracking analyses (NTA) (NanoSight NS300 (Malvern Panalytical, Malvern, UK). To generate appropriate particles per frame in the NTA analyses, the samples were diluted 1:50 or 1:100 in filtered PBS. The NTA settings used were: camera level 13–14, speed of syringe pump 100, threshold of analyses five and time of analyses 3 × 60 s.

The NTA confirmed that the isolated EVs were small EVs (sEVs) as the median size was about 120 nm for three individual sEVs isolates used in the study (H1975_1: 116 nm; H1975_2: 121 nm; H1975_3: 114 nm) (Figure S1A). The mean size of the extracted vesicles ranged from 161 nm (H1975_2) to 175 nm (H1975_3) likely due to the experimental conditions of isolation (Figure S1A). The amount of particles per ml differed among the three samples as different amounts of cells were used for the isolation. However, for subsequent analyses with the DBS-Pro the number of sEVs used in the assay were equal based on the NTA measurements and subsequent dilutions. We have earlier reported that the SEC based isolation of EVs from H1975 cell culture media gives vesicles that express bona fide protein markers for exosomes, for example, CD9 and TSG101 (Stiller et al., 2021), as indicated by International Society for Extracellular Vesicles (ISEV) (Théry et al., 2018). We also here confirmed with Western Blotting (WB) that sEVs isolated from another two replicates of H1975 cell culture media (during the same period as the replicates for the current study were isolated) expressed CD9, CD73, TSG101 and GAPDH with no signal for the ER-protein calnexin in sEVs (Figures S1B and S2A). All in all, sEVs characterization from samples used in this study as well as SEC-isolated sEVs from H1975 cell media reported in our previous publication support that sEVs are of exosome-size and express exosome-defining proteins.

### 2.2 Pleural effusion fluid samples and isolation of EVs

EVs were also isolated from MPE fluid obtained from NSCLC patients at Karolinska University Hospital, Stockholm, Sweden. The patient samples were gathered under ethical permit from The Ethics Review Authority in Sweden (https://etikprovningsmyndigheten.se), region Stockholm (EPN No. Dnr. 2016/2585-32/1, approval date 8th of March 2017). The patients agreed to collection of their MPE fluid for biomarker analyses via informed consent with the material collected under biobank approval. The NSCLC patients’ tumours at diagnosis were of adenocarcinoma histology with the following genomic alterations reported in the medical records: PE002: EML4-ALK, variant 3 (a/b), PE009: KRAS, exon 2, codon 12/13 and PE011: EGFR, exon 21, L858R. The total MPE fluid used in the study was centrifuged at 2000 RCF, 5 min (Rotina38, Hettich Labinstruments AB, Stockholm, Sweden) and the fluids were stored at −80°C until EV isolation. To isolate EVs from the fluids, the sample was first centrifuged (2000 RCF for 5 min) and thereafter filtered using a 0.2 μm filter (Acrodisc®, Pall Corporations, VWR, Stockholm, Sweden). The resulting fluids were concentrated on a Amicon® Ultra-4 Centrifugal Filter Unit to ~300–500 μl. The samples were thereafter added to a qEVoriginal/70 nm column in which fraction 6–10 was harvested, pooled, and concentrated prior to NTA analyses. The NTA analyses were done on the same equipment as above. The samples were diluted 1:25 in PBS resulting in at least 45 particles/frame. The syringe pump speed was 100 and the camera level was set to five with the time of analysis being 3 × 60 s. Tumour cells were also isolated from the PE-fluid using Ficoll (GE Healthcare, Uppsala, Sweden) where the tumour cells were harvested from the interphase after centrifugation at 720 RCF for 20 min. The cells were rinsed in PBS and grown ex vivo in DMEM high glucose media containing l-glutamine (SH30022.01; GE Healthcare), Pen-Strep antibiotics and 20% FBS (both from Gibco, Life Technologies). After a couple of weeks tumour cells were established, expanded and frozen. A batch of these cells from PE002, PE009 and PE011 were used in the WB analysis in the current study. The NTA profiling of these sEVs samples revealed that SEC-isolated EVs had a median size of 140 nm (PE002), 180 nm (PE009), and 160 nm (PE011), respectively (Figure S1C). The concentration of the particles per ml varied but a similar number of vesicles as determined by NTA was used for the DBS-Pro profiling. WB protein profiling of these sEVs confirmed expression of CD9 (Figures S1D and S2B).

### 2.3 Protein profiling of EVs by WB analyses

A protein extract was made from H1975 cell pellet by adding 1× RIPA buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% Triton- X100, 1 mM EDTA, 0.1% SDS) to which protease and phosphatase inhibitors (complete Mini; Phos STOP, Sigma Aldrich, Merck Chemicals and Life Science AB) were supplemented. Samples were incubated on ice for ~15 min, after which cell debris were cleared out by centrifugation (24,100 RCF (Micro 200R, Hettich Labinstrument AB). Protein concentration was analyzed by BCA assay (Pierce™ BCA Protein Assay, Fisher Scientific) and about 30 μg/lane was used for the WB analyses. Protein extract was also made from EVs isolated from H1975 cell culture media were extracted for protein by adding 5x RIPA buffer to concentrated EVs samples with a final 1x RIPA buffer. The EVs were lysed on ice for 15 min and an equal total amount of 3.77 × 10⁸ particles as...
determined by NTA were loaded onto gels from each sample. EVs from PE samples were similarly lysed in 5x RIPA and handled as above. A total amount of 6.4 × 10⁶ particles as determined by NTA were loaded for each sample. For the WB profiling of MPE fluid-isolated tumour cells and grown ex vivo a protein extract was made in 1x RIPA buffer as described above. The protein amount used for the profiling were: PE002: 2.3 µg (due to shortage of material), PE009 and PE011: 23 µg each. Prior to the WB analyses, sample buffer supplemented with reducing agent (Fisher Scientific) were added to the samples which were heated for 10 min at 70°C. The protein extracts were separated on Invitrogen NuPAGE® gel system (Fisher Scientific) using 4%-12% Bis-Tris (TB) gels using MES running buffer (NuPAGE®, Fisher Scientific) during 1 h at 180v. The proteins were transferred from gels to Odyssey® Nitrocellulose Membrane (LI-COR GmbH, BadHomburg, Germany) during 1.5 h at 30 V with a transfer buffer (10% methanol). Blocking of membranes were done during 1 h in Odyssey® blocking buffer (LI-COR GmbH)/TBS-T 1:1. Membranes were probed with primary antibodies overnight at 4°C with rotation: anti-CD73 (MAS-15537, 1:250 dilution, Invitrogen, Stockholm Sweden), anti-calnexin (#2433, 1:250 dilution, Cell Signalling Technology, BioNordica AB, Stockholm, Sweden), anti-TSG101 (ab125011, 1:250 dilution, Abcam, Cambridge, UK), anti-GAPDH: (#2275-PC-100, 1:500 dilution, Trevigen, Bio-Technne, Abingdon, UK) and anti-CD9 (#13403, 1:500 dilution, Cell Signalling Technology, BioNordica AB, Stockholm, Sweden). The membranes were after primary antibody probing washed in TBS-T for 3 × 5 min and incubated for 1 h with goat anti-Rabbit IRDye® 800CW antibody (#926-32211), or donkey anti-mouse IRDye® 680 (#926-68072) both from LI-COR. The membranes were thereafter washed again 3 × 5 min and the Odyssey® Sa Infrared Imaging System (LI-COR GmbH) used for visualizing resulting antibody-related signals.

2.4 Antibody-DNA conjugates

Antibody–DNA conjugates were prepared by employing the Z domain from protein A and using a photoaffinity-based conjugation method described earlier (Stiller et al., 2019). The purity of the antibody-DNA conjugates was improved by introducing a purification step using IgG-Sepharose 6 Fast Flow affinity resin (Cytiva I7-0969-01, Sigma-Aldrich) to remove the excess Z-DNA molecules. The IgG molecules immobilized on the resin can bind to the free Z-DNA and retain them, while the antibody-DNA conjugates are eluted and used in the DBS-Pro assay. Briefly, the photoconjugation reaction mixture was diluted to 50 µl final volume and was applied to 50 µl of resin equilibrated in PBS, followed by incubation at 4°C for 1 h. This incubation step was repeated twice with a fresh aliquot of resin. After the final incubation, the antibody-DNA conjugates were eluted using MicroSpin columns (Thermo Fisher Scientific) by centrifuging the resin-sample mixture at 1000 rpm for 2 min. The purified samples were loaded on SDS-PAGE gels and the labelling efficiency and purity were analyzed. To visualize the DNA-oligonucleotide content, the gels were stained with GelRed, and Coomassie staining was used to visualize proteins. The conjugation process results in each antibody molecule to carry on average one oligonucleotide consisting of a 6-nt long antibody-specific barcode (ABC) followed by a 6-nt long unique molecular identifier (UMI). The ABC and UMI are flanked by two universal amplification handles (H3 and H4). The antibodies, the barcodes for each antibody and the sequences for the handles flanking the ABC and the UMI are listed in Table S1 and S2.

2.5 sEV tagging and capture

sEVs were incubated with barcoded antibodies (5 nM of each) in 1x PBS buffer (Thermo Fisher Scientific; Cat. No. 18912014) in the presence of 50–500 nM of Z(WT), at 4°C overnight whilst rotating. MyOne Streptavidin Dynabeads (Thermo Fisher Scientific; Cat. No. 65306) were covered with biotinylated Cholera Toxin B subunit (CTB, Thermo Fisher Scientific; Cat. No. C34779). Approximately 20×10⁶ sEVs (as determined by NTA analyses) were incubated with circa 200 × 10⁶ CTB coated beads in PBS-CTD buffer (1x PBS with 0.05% casein (w/v), 0.1% Tween-20 (v/v), 1 mg/ml salmon sperm DNA (Thermo Fisher Scientific; Cat. No. 15632011) and 50–500 nM of Z(WT)), with rotation at 4°C for 10 minutes. The 1 sEV to 10 bead (0.1 vesicles per bead [VPB]) relationship was used to ensure that the majority of the sEV-carrying beads captured a single vesicle. Three times washing with PBS-CTD buffer was then applied and after the last wash, the beads were resuspended in 200 µl 1x PBS buffer. A fraction of the beads was used in emulsion PCR (emPCR) and another for quantitative PCR (qPCR). qPCR samples were accompanied with negative (antibody-background) controls for which the whole above-process was carried out without inclusion of sEVs in the initial step.

2.6 Emulsification and barcoding reactions

Approximately one million beads (0.1 VPB) were transferred to a 50 µl PCR mixture containing 1x Phusion Hot Start Flex (New England BioLabs; Cat. No. M0535), 2% PEG-6000 (Sigma Aldrich), 2% Tween-20 (Sigma Aldrich; Cat. No. P9416), 3% DMSO (New England BioLabs; I2611P), approximately one million molecules of H1-DBC-H2 (droplet barcode [DBC]), 0.05 µM of
the connector primers H3’-H2’ and H3 and 0.2 µM of H1 and Bio-H4’ primers (see Table S2 for sequences). All primers and oligonucleotides for emPCR were purchased from Integrated DNA Technologies (IDT). For each sample, two emulsion reactions were prepared. The protocol for generating water-in-oil emulsion droplets is described in detail by Redin et al. (2019). The emPCRs were performed in a Mastercycler Pro S (Eppendorf) machine, initiated by denaturation at 95°C for 5 min (ramp speed 100%) and then the following cycling profile: 30 cycles of 95°C for 30 s (ramp speed 50%), 60°C for 60 s (ramp speed 20%) and 72°C for 60 s (ramp speed 20%) followed by seven cycles of 95°C for 30 s (ramp speed 50%), 45°C for 60 s (ramp speed 20%) and 72°C for 60 s (ramp speed 20%). The emPCR program ended with 600 s at 72°C (ramp speed 20%) and then hold at 4°C (ramp speed 100%).

2.7 Emulsion breakage and library preparation

The emulsion breakage and library preparation steps were done as previously described by Redin et al. (2019), except for modifications in the PEG-based purification. Briefly, following aqueous phase recovery and PCR product purification by 20% PEG (Sigma Aldrich), biotinylated amplicons were enriched using MyOne Streptavidin Dynabeads (Thermo Fisher Scientific). After NaOH treatment, an indexing reaction was performed using a master mix containing Phusion Hot Start Flex (New England Biolabs; Cat. No. M0536L) and 0.2 µM of i5-H1 and i7-H4’ indexing primers (see Table S2 for sequences) using the following program: two cycles of 2 min at 95°C, 1 min at 60°C (with 20% ramp speed) and 10 min at 72°C (with 3% ramp speed). The final product was then purified three times using 16% PEG and the quality was examined by Bioanalyzer (Agilent). The libraries were quantified by Qubit fluorometer (Invitrogen) and sequenced using Reagent Kit v3 (150-cycle) for the MiSeq instrument (Illumina).

2.8 Quantitative real-time PCR

After antibody tagging and capturing of the sEVs on beads (see section 2.5), approximately one million beads (about 1/10 carrying sEVs) were used for quantitative real-time PCR in bulk. In each experiment, a negative control, in which sEVs were excluded during incubation with antibodies, was used to investigate background levels from barcoded antibodies. A second negative control containing only water was also used. The PCR mixture (50 µl) contained 1× Phusion Hot Start Flex (New England BioLabs; Cat. No. M0536L), 0.2 µM of H3 and H4’ primers (Table S2) and EvaGreen Dye (Biotium; Cat. No. 31000). Amplification was done using the following protocol: 5 min at 95°C; 40 cycles of 30 s at 94°C (ramp speed 40%), 30 s at 60°C (ramp speed 40%), 30 s at 60°C (ramp speed 60%), and 65 s at 72°C (ramp speed 30%). The real-time amplification reactions were performed on a CFX Connect Real-Time System machine (Bio-Rad).

2.9 Hashing experiments

Antibody-DNA conjugates recognizing EGFR and CD9 with different barcoding oligonucleotides were used in EV- and bead-hashing experiments. The antibody barcodes for EGFR and CD9 were ACAGTC and GTACTG in set one, and GTCGAT and AGTCTG in set two. In the EV-hashing experiments, sEVs collected from cell culture media of the H1975 cell line were incubated with set one and set two antibodies in separate reaction tubes as described above. Approximately $10 \times 10^6$ sEVs from each reaction tube were then mixed and captured on beads followed by emulsification, library preparation and sequencing, as described above. The bead-hashing experiments were carried out using the same protocol as EV-hashing except capturing of sEVs on the beads was done separately for each set and samples were pooled before emulsification.

2.10 Data analysis

FASTQs from sequencing were processed using the DBS-Pro pipeline (https://github.com/FrickTobias/DBS-Pro, v0.3). From each sequence read the DBC and the ABC+UMI combination were separately extracted using cutadapt (Martin, 2011). DBC sequences were error-corrected by clustering using starcode (Zorita et al., 2015) with an edit distance of two (parameter ‘-d 2’). The target (ABC) was identified from each ABC+UMI combination using cutadapt to compare against the known ABC sequences used in the sample. For each corrected DBC and identified ABC, UMI sequences were corrected using UMI-tools (Smith et al., 2017). Finally, the read counts were aggregated for each corrected DBC, ABC and UMI combination.

The data was filtered in several steps. First, to remove background and chimeric products, only UMIs with more than one read were kept and DBCs with only one UMI were filtered out. For same reason, for each DBC we also removed targets with only one UMI. In the hashing experiments, in droplets with mixed signals, we observed a low ratio of reads to UMIs. We therefore filtered
out DBCs based on the ratio of reads to UMIs where a lower threshold of three was used for the hashing experiment (using only two affinity reagents) while two was used for the multiplex analyses of sEVs from the HI975 cell culture media and sEVs from the MPE samples. To remove droplets carrying multiple DBCs we compared the pairwise overlap of UMI sequences using a Jaccard index. The Jaccard index was defined as the intersect divided by the union of two sets of UMIs from pair of DBCs. Any pair of DBCs with a Jaccard index greater than 0.5 was removed. For the sEVs from HI975 cell culture media and from the MPE samples we also removed DBCs with more than 25 UMIs. From the filtered data a count matrix was generated with the number of UMIs for each DBC and ABC.

Graphs were generated using python and packages matplotlib (Hunter, 2007), pandas (McKinney, 2010; pandas development team 2020), numpy (Harris et al., 2020), and seaborn (Waskom, 2021). For UMAP and PCA projections scanpy (Wolf et al., 2018) was used. First, droplets were filtered to only keep those with 3 or more targets. The count matrices were then normalized per sample using the centred log ratio (CLR) transformation over targets with the following expression:

$$CLR(x) = \left[ \log \left( \frac{x_1}{g(x_1)} + 1 \right), \ldots, \log \left( \frac{x_n}{g(x_n)} + 1 \right) \right]$$

where $x$ is the count matrix and $g(x_i)$ is the geometric mean over every non-zero entry for the target. UMAP embedding of sEVs from the MPE samples was generated using parameters n_neighbours = 100 and min_dist = 0.9. Louvain clustering was performed using a resolution value of 0.6.

3 | METHOD DESCRIPTION

In this work we have further developed the DBS technology to profile surface proteins of single sEVs. Figure 1 describes the concept of droplet barcode sequencing for protein analysis (DBS-Pro) and quantitative protein profiling of sEVs at single vesicle resolution. The assay starts with incubating the sEVs with a set of antibodies that are labelled with oligonucleotides. The oligonucleotide contains a sequence specific to the antibody-type called the antibody barcode (ABC). Each oligonucleotide also contains a UMI to enable quantitative analysis. The ABC and UMI are flanked with universal handles for amplification. The vesicles tagged with antibodies are then captured on magnetic beads covered with cholera toxin subunit B (CTB). A ratio of 0.1 VPB is used to ensure that most occupied beads carry only a single vesicle. This is followed by a series of washes to remove excess barcoded antibodies and protein contaminants in the sample. The beads are then encapsulated in emulsion droplets with a ratio of 0.1 bead per droplet. The droplets contain unique DBC sequences that are clonally amplified in the emPCR and through overlap-extension
are linked to amplified ABCs in the last cycles of the emPCR. The emPCR is followed by emulsion breakage, library indexing and sequencing using a MPS instrument. Figure S3 shows the details of construct design and library preparation. Sequencing reads are clustered based on their DBCs followed by protein identification (through ABCs) and quantification (through UMIs) to yield single sEV protein counts for further analysis. Subpopulations may be investigated through unsupervised clustering of individual vesicles based on their surface protein profiles.

4 | RESULTS

4.1 Validation of single vesicle detection

To investigate and validate detection of single vesicles using the DBS-Pro assay (Figure 1), a series of so-called hashing experiments were conducted using sEVs collected from the cell culture media of the NSCLC cell line H1975 (Figure 2A). Two different sets of barcoding oligonucleotides were used to label CD9 and EGFR antibodies. In the bead hashing experiments, vesicles were incubated with set one (CD9_1 and EFGR_1) and set two (CD9_2 and EFGR_2) antibodies in two separate reaction tubes. They were also captured on CTB-covered beads and washed separately but were mixed before emulsification. Since the two sets are physically separated until droplet generation, mixed observations would indicate presence of more than one bead in the droplets. In the EV-hashing experiments, however, the EVs are mixed before bead-capturing. In this setup, mixed observations of set one and set two antibodies would suggest presence of two or more EVs on beads or a consequence of co-encapsulation of beads. Figure 2B shows the single set and mixed rates for bead- and EV-hashing experiments. Droplets are considered single-set if 95% or more of the UMIs belong to ABCs of one set only. To demonstrate reproducibility, we performed two bead-hashing and three EV-hashing replicates. The mixed rate for bead-hashing was 1.5% and 2% while for EV-hashing we observed a mixed rate between 1.3% and 1.8%. The data shown in Figure 2B is based on a total of 46,610 single vesicles (Bead_Hashing_1: 5,270; Bead_Hashing_2: 11,928; EV_Hashing_1: 5,623; EV_Hashing_2: 14,557; and EV_Hashing_3: 9,232). Figure 2C shows the droplet counts for the two sets distributed along the axes for ascending UMI counts. Most of the droplets contained few UMIs, and predictably, a large fraction of the droplets with mixed signals had higher UMI counts. The hashing experiments clearly indicate that DBS-Pro generates data with single vesicle resolution in at least 98% of the droplets. The background noise, predominantly due to presence of excess barcoded antibodies not removed by washing, was also very low as quantitative PCR showed approximately eight to nine cycles difference between the samples and negative controls (Figure S4A,B).

4.2 Validation of DBS-Pro for multiplex protein analysis of sEVs derived from a NSCLC cell line

To investigate the potential of the technology for analysis of multiple target proteins on individual sEVs, we conducted a series of DBS-Pro experiments using a panel of eleven proteins. The experiments were performed in parallel on sEVs isolated from the media of the cultured NSCLC cell line H1975 collected at three timepoints, named H1975_1, H1975_2 and H1975_3. Low background noise levels were confirmed using quantitative PCR that showed 7 to 11 cycles difference between samples and negative controls (see Figure S4C–E). The sequencing analysis identified 8943, 10,054 and 30,350 single vesicles in H1975_1, H1975_2 and H1975_3, respectively. In this study, through WB analysis, we report on expression of CD9 in sEVs from H1975 cell culture media (Figure S1B). In addition, in a previous study (Stiller et al., 2021), WB analysis of H1975 cells and sEVs isolated from cell culture media of these cells, we identified expression of CD9 and EGFR in sEVs (see Figure S3 in Stiller et al., 2021). To confirm this, we aggregated UMI counts for each protein (Figure 3A; Table S3). As expected, a clear expression of CD9 and EGFR was observed in sEVs from all three samples. In addition, in this study we identified high expression of the epithelial marker EpCam. Looking at the fraction of droplets containing CD9, EGFR and EpCam (Figure 3B) one can see that these proteins are the three most common over all samples. Furthermore, a similar distribution of UMI counts is seen across samples for droplets containing each respective protein (Figure 3C).

To further compare the similarity of the samples we performed principal component analysis on the droplets containing three or more proteins. Figure 3D shows scatter plots of the first two principal components with relative sample density along each axis, demonstrating a very good overlap between the samples. Together with the similar results obtained for the aggregated data (Figure 3A–C), this suggests that the protein expression patterns of sEVs does not change to a large extent across the three individual sEVs isolations from cell culture media. The data presented in Figure 3 also demonstrates that the DBS-Pro technique enables multiplex protein detection of individual sEVs and captures reproducible expression patterns.
4.3 Surface protein profiling of individual vesicles from MPE fluid of NSCLC patients

After establishing that the DBS-Pro enables multiplex detection of surface proteins on individual sEVs, we applied the assay for protein profiling of single sEVs collected from MPE fluid of advanced NSCLC patients (PE002, PE009 and PE011) with different tumour genomic alterations. Quantitative PCR was applied as before to confirm low background noise in the samples (see Figure S4F). The DBS-Pro assay detected 9064, 11,920 and 4565 single vesicles in the PE002, PE009, and PE011 samples, respectively. The heatmap in Figure 4A shows the presence of each target as a proportion of the total number of DBCs. The results indicate that sEVs from all three MPE samples had relatively high expression of CD9 as it was detected in a large fraction of droplets.
Figure 3. DBS-Pro for multiplex protein analysis of single sEVs derived from cell culture media of the H1975 cell line. H1975_1, H1975_2 and H1975_3 represent sEVs from three different samples isolated from cell culture media. (A) Heatmap showing aggregated data based on the log2 transformed total number of UMI s (nUMIs) for each target. (B) Heatmap showing the fraction of the droplets containing each target. (C) Violin plots showing the distribution of UMI counts per droplet for each target and sample. Boxplot is shown inside each violin. (D) Scatter plot of the first two principal components (PCs) for droplets containing at least three targets. Relative sample density shown along each axis.
FIGURE 4  DBS-Pro for surface protein profiling of single sEVs isolated from MPE from NSCLC patients. Three different patient samples (PE002, PE009 and PE011) with different tumour genomic alterations were used. (A) Heatmap showing the presence of each target as a proportion of the total number of barcodes. (B) Heatmap showing the top 30 target combinations as a proportion of the total number of barcodes. The combinations are indicated with an upset diagram below the heatmap. (C) Uniform Manifold Approximation and Projection (UMAP) embedding of all samples' individual vesicles, showing the called Louvain clusters. (D) Stacked bar chart showing the relative proportion of sEVs from each MPE sample in each Louvain cluster. (E) UMAP embedding of PE002, PE009 and PE011 samples' individual vesicles, showing the density of each sample. (F) Heatmap showing the relative expression of each target across droplets grouped by Louvain clusters. Each column represents one droplet (i.e., one vesicle). Dendrogram on top shows the relationship between clusters.
(i.e., individual vesicles) across all samples (PE002: 53%, PE009: 65% and PE011: 75%). This was expected as CD9 is considered as a marker for sEVs of endosomal origin also by ISEV (Thérie et al., 2018), and in many studies is used to detect and enrich for sEVs because of its high abundance. In addition to CD9, higher levels of EGFR and EpCam was observed in PE009 and PE011 while high expression of PD-1 was detected in PE002. The data in Figure 4A is presented as violin plots in Figure S5, marking the distribution of UMIs for each target and sample. sEVs from PE002, PE009 and PE011 were also included in our previous work (Stiller et al., 2021). In that study, we performed WB analysis of sEVs, probing CD9, EGFR and IGF-1R (see Figure S11 in Stiller et al., 2021). WB results showed presence of CD9 in all samples but with weaker band for sEVs from PE002. Expression of EGFR was observed in sEVs from PE009 and PE011 (although weaker for PE009) while it was negative in sEVs from PE002. Immunoblotting of IGF-1R for sEVs derived from these samples turned negative. In the same study (Stiller et al., 2021), we performed immuno-PCR for detection of EGFR and IGF-1R on CD9-positive sEVs, by utilizing DNA conjugated affibodies (see Figure 4 in Stiller et al., 2021). Immuno-PCR results indicated higher expression of EGFR in sEVs from PE011 as compared to sEVs from PE002 and PE009 and very low signals (close to limit of detection) were detected for IGF-1R in all samples. All the WB results presented by Stiller et al. are confirmed in our study, when compared to Figures 4A and S5. An exception is the presence of EGFR in 19% of the droplets in PE002, a difference that may be due to the low sensitivity of WB. IGF-1R was observed in DBS-Pro but only in 1%–8% of the droplets. The only discrepancy between DBS-Pro and immuno-PCR was in detecting higher levels of EGFR in PE009 with DBS-Pro, which may be attributed to employing a different affinity reagent (affibody instead of antibody) or enrichment based on CD9-positive sEVs in the immuno-PCR assay. The results presented in Figure 4A clearly indicate that a high fraction of sEVs from these MPE of NSCLC patients do not carry CD9 on their surface illustrating that with our method further essential molecular information may be revealed. Detecting co-expressed proteins is essential for identifying subpopulations of sEVs. From the top 30 co-expressed proteins identified by DBS-Pro on single sEVs several sample specific combinations emerge (Figure 4B). Looking deeper into this data we extracted the most informative droplets (carrying three or more proteins) from the three MPE samples. Unsupervised clustering of the remaining droplets (PE002: 2,984; PE009: 427; PE011: 221) yielded six clusters here represented in a UMAP (Figure 4C). Comparing the different samples across the cluster, their relative proportion (Figure 4D) and density (Figure 4E) varies. For example, sEVs from all three MPE samples appear in cluster 1 but at different magnitude among the samples, whereas cluster 3 is almost exclusively represented by sEVs from PE002. The relative expression of each target across the droplets grouped by clusters is show in Figure 4F. Comparing clusters 1 and 3 again, one can see that the difference is in the expression of EGFR, CD8 and EpCam. Furthermore, unlike CD3, CD4, PD-L1 and IGF-1R the expression of CD9 is ubiquitous across clusters. NY-ESO1 is distinctly present in cluster 5 which mostly consists of sEVs from patient PE011. This marker has previously been associated with prognostics of NSCLC patients in sEVs from plasma characterized using protein arrays (Sandfeld-Paulsen et al., 2016). The results presented in Figure 4C–F are based on sEVs obtained from only three MPE samples, but it clearly demonstrates that DBS-Pro allows for clustering and identification of subtypes within and between samples. This may in the future be used as a diagnostic tool to uniquely identify clusters present in disease cohorts or prior and post treatments in individual patients.

5 DISCUSSION

sEVs carry molecular information of their parent cells. However, samples derived from liquid biopsies, for example plasma or PE fluid, exhibit high sEV heterogeneity due to secretion of sEVs from cells with different molecular composition. Identifying subpopulations of sEVs in these heterogeneous samples is therefore essential to further our knowledge of various disease processes, including cancer and its response to given treatments. However, much of the available data today is acquired through studies using technologies that provide ensemble measurements on sEVs, not considering cellular heterogeneity. In addition, existing single sEV assays, with the potential to detect sEV subtypes, lack multiplexing capabilities and/or throughput. In this proof-of-concept study we present DBS-Pro to characterize sEVs at single-vesicle resolution. The technology allows high throughput analysis of multiple surface proteins on individual sEVs. Compartmentalization through use of emulsion droplets followed by emPCR has enabled analysis of single vesicles with high throughput and high sensitivity. The content of single vesicles is identified by linking them to unique probes called DBC sequences that are co-encapsulated with sEVs and clonally amplified. Through hashing experiments, we demonstrated that over 98% of the data presented in this work are results of single vesicle profiling. The low rate of mixed signals was expected as we employed ratios of 0.1 VPB and 0.1 bead per droplet. Further, as we did not observe any difference in mixed sequences between the two hashing experiments, it suggests co-encapsulation of beads to be the main source of mixed signals. Although the rate of mixed signals is very low, we believe that the use of microfluidic chips for generating uniform emulsion droplets would further improve it. In this study we employed shake emulsifications, which generate droplets of different sizes, and in larger droplets we may encapsulate multiple analytes. However, advantages such as ease of use, low cost and avoiding specialized fluidic systems, make shake emulsions an attractive alternative. Furthermore, although we have analyzed a total of 121,500 single sEVs (an average of 11,046 vesicles per sample), the assay could be scaled to include a larger number of vesicles through use of the microfluidic systems, which could generate smaller droplets (i.e., larger number of droplets per sample volume) in a controlled manner. Multiplexing is facilitated through DNA barcoding of affinity reagents. In this study we targeted
eleven sEVs surface proteins being relevant for NSCLC, but the method has potential to reach considerably higher levels of multiplexity as the readout is sequencing and the length of antibody barcodes could readily be expanded with no impact on the assay procedure. In addition, the use of UMIss makes it possible to accurately determine the copy numbers of each protein present on vesicles, leading to precise characterization of each sEV in a sample. All these features allow DBS-Pro to identify subpopulations of sEVs. This was demonstrated through clustering and identification of sEV subtypes within and between three MPE samples from NSCLC patients with advanced disease and different tumor genomic make up. Identification of sEVs subtypes will play a key part towards precision medicine, treatment monitoring and detection of minimal residual disease (MRD). In this study we have pooled SEC fractions, but there is a possibility to analyze individual fractions without pooling. This may help to investigate whether there are correlations between individual subpopulations and size of EVs. However, in clinical applications of the technology such a strategy may not be feasible as the amount of the obtained material is limited. Furthermore, such a strategy would necessitate multiple DBS-Pro experiments leading to higher workload and considerably higher sequencing cost. Integrating multi-omics approaches may prove important to further improve the classification of sEV subtypes as the volume of sEVs is extremely small, leading to each vesicle carrying a limited number of biomolecules. The DBS technology is versatile and could easily be adapted to target genomic fragments, transcripts, and proteins. For high resolution sEV characterization, we envision utilizing DBS for simultaneous analysis of transcripts and proteins of single vesicles. While the number of target proteins may be increased with ease, unbiased transcript profiling could be done on the same vesicles through lysis of sEVs inside the droplets followed by reverse transcription and tagging of both types of biomarkers with DBCs.

ACKNOWLEDGEMENT
The authors acknowledge support from the National Genomics Infrastructure (NGI) in Stockholm funded by Science for Life Laboratory (SciLifeLab), the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. The data handling was enabled by resources in project (SNIC 2021/5-544 and SNIC 2021/6-314) provided by the Swedish National Infrastructure for Computing (SNIC), partially funded by the Swedish Research Council through grant agreement no. 2018–05973. The authors would also like to thank Dr. Mahya Dezfouli for discussions on chemical antibody-DNA conjugation, Dr Caroline Kamali and Dr Luigi De Petris for collection of PE samples, Vasiliki Arapi for help with sEVs isolation and Tobias Frick for helping with the development of the DBS-Pro pipeline. The supporting grants for this study were from the Erling Persson Family Foundation, Olle Engkvist Foundation (191-0475 and 193–0605), the Swedish Research Council (VR) (2018-06228), Swedish Cancer Society (CAN 2015/401 (R.L.), CAN 2018/597 (R.L.), CAN2021/1469 Pjo1 H (R.L.)), the Stockholm Cancer Society (#171123 (R.L.), #201202 (RL), #191293 (K.V.)) and Stockholm County Council (#20180404 to R.L and #962324 to S.E, R.L and K.V).

CONFLICT OF INTERESTS
The authors declare no conflict of interests.

AUTHOR CONTRIBUTION
Afshin Ahmadian conceived and conceptualized the DBS-Pro technology for single vesicle profiling. Mahsan Banijamali, Pontus Höjer and Afshin Ahmadian designed the DBS-Pro experiments. Mahsan Banijamali and Afshin Ahmadian developed and optimized DBS-Pro lab protocols. Mahsan Banijamali performed all DBS-Pro experiments and validation tests. Pontus Höjer further developed the analysis pipeline, performed all data analysis and data visualizations. Mahsan Banijamali, Pontus Höjer and Afshin Ahmadian wrote and prepared the original draft. Amelie Eriksson Karlström, Christiane Stiller and Abel Nagy developed the improved protocol for antibody-DNA conjugation. Christiane Stiller, Abel Nagy and Elizabeth Paz Gomer performed conjugations. Simon Ekman and Rolf Lewensohn provided the clinical samples and clinical expertise. Kristina Viktorsson, Petra Hååg and Vitaliy O. Kaminsky selected biomarkers, selected and cultured the cell line and prepared all sEVs used for the study. Amelie Eriksson Karlström and Christiane Stiller selected antibodies binding the biomarkers. All authors contributed to reviewing and editing of the original draft. All authors have read and agreed to submit the manuscript for publication.

DATA AVAILABILITY STATEMENT
DBS-Pro pipeline is accessible through https://github.com/FrickTobias/DBS-Pro (version 0.3). Code and data used for downstream analysis and figures is available at https://github.com/AfshinLab/dbspro_exosome_analysis.

REFERENCES
Borgström, E., Redin, D., Lundin, S., Berglund, E., Andersson, A. F., & Ahmadian, A. (2015). Phasing of single DNA molecules by massively parallel barcoding. Nature Communications, 6. https://doi.org/10.1038/ncomms8173
Cao, L., Cui, X., Hu, J., Li, Z., Choi, J. R., Yang, Q., Lin, M., Ying Hui, L., & Xu, F. (2017). Advances in digital polymerase chain reaction (dPCR) and its emerging biomedical applications. Biosensors and Bioelectronics, Elsevier Ltd; 90, 459–474.
Villagrasa, A., Álvarez, P. J., Osuna, A., Garrido, J. M., Aráñega, A., & Rodríguez-Serrano, F. (2015). Exosomes derived from breast cancer cells, small trojan horses? *Journal of Mammary Gland Biology and Neoplasia*, Springer New York LLC, 19, 303–313.

Wang, S., Khan, A., Huang, R., Ye, S., Di, K., Xiong, T. & Li, Z. (2020). Recent advances in single extracellular vesicle detection methods. *Biosensors and Bioelectronics*, Elsevier Ltd, 154, 112056.

Waskom, M. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, 6(60), 3021.

Wolf, F. A., Angerer, P., & Theis, F. J. (2018). SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biology*, 19(1), 15.

Wu, D., Yan, J., Shen, X., Sun, Y., Thulin, M., Cai, Y., Wik, L., Shen, Q., Oelrich, J., Qian, X., Dubois, K. L., Ronquist, K. G., Nilsson, M., Landegren, U., & Kamali-Moghaddam, M. (2019). Profiling surface proteins on individual exosomes using a proximity barcoding assay. *Nature Communications*, 10(1). http://doi.org/10.1038/s41467-019-11486-1

Yáñez-Mó, M., Siljander, P. R. M., Andreu, Z., Zavec, A. B., Borràs, F. E., Buzas, E. I., Buzas, K., Casal, E., Cappello, F., Carvalho, J., Colás, E., Cordeiro-da Silva, A., Fais, S., Falcon-Perez, J. M., Ghobrial, I. M., Giebel, B., Gimona, M., Graner, M., Gursel, I., … De Wever, O. (2015). Biological properties of extracellular vesicles and their physiological functions. *Journal of Extracellular Vesicles*, Co-Action Publishing, 4, 1–60.

Zorita, E., Cuscó, P., & Filion, G. J. (2015). Starcode: Sequence clustering based on all-pairs search. *Bioinformatics*, 31(12), 1913–1919.

**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Banijamali, M., Höjer, P., Nagy, A., Hååg, P., Gomero, E. P., Stiller, C., Kaminskyy, V. O., Ekman, S., Lewensohn, R., Karlström, A. E., Viktorsson, K., & Ahmadian, A. (2022). Characterizing Single Extracellular Vesicles by Droplet Barcode Sequencing for Protein Analysis. *Journal of Extracellular Vesicles*, 11, e12277. https://doi.org/10.1002/jev2.12277