












Article

# Is the Proteome of Bronchoalveolar Lavage Extracellular Vesicles a Marker of Advanced Lung Cancer?

Ana Sofia Carvalho <sup>1,\*</sup>, Maria Carolina Strano Moraes <sup>2</sup>, Chan Hyun Na <sup>3</sup>, Ivo Fierro-Monti <sup>1</sup>, Andreia Henriques <sup>1</sup>, Sara Zahedi <sup>1</sup>, Cristian Bodo <sup>2</sup>, Erin M Tranfield <sup>4</sup>, Ana Laura Sousa <sup>4</sup>, Ana Farinho <sup>5</sup>, Luís Vaz Rodrigues <sup>6</sup>, Paula Pinto <sup>7</sup>, Cristina Bárbara <sup>8</sup>, Leonor Mota <sup>7</sup>, Tiago Tavares de Abreu <sup>7</sup>, Júlio Semedo <sup>7</sup>, Susana Seixas <sup>9</sup>, Prashant Kumar <sup>10,11</sup>, Bruno Costa-Silva <sup>2</sup>, Akhilesh Pandey <sup>10,11,12</sup> and Rune Matthiesen <sup>1,\*</sup>

<sup>1</sup> Computational and Experimental Biology Group, Chronic Diseases Research Centre, NOVA Medical School, Faculdade de Ciências Médicas, Universidade NOVA de Lisboa, Campo dos Martíres da Patria, 130, 1169-056 Lisboa, Portugal; ivo.fierromonti@nms.unl.pt (I.F.-M.); andrea.henriques@nms.unl.pt (A.H.); sara.zahedi@nms.unl.pt (S.Z.)

<sup>2</sup> Systems Oncology Group, Champalimaud Research, Champalimaud Centre for the Unknown, Av. Brasília, Doca de Pedroucos, 1400-038 Lisbon, Portugal; carolina.stranomoraes@research.fchampalimaud.org (M.C.S.M.); cristian.bodo@research.fchampalimaud.org (C.B.); bruno.costadasilva@research.fchampalimaud.org (B.C.-S.)

<sup>3</sup> Department of Neurology, Institute for Cell Engineering, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; chanhyun@jhmi.edu

<sup>4</sup> Electron Microscopy Facility, Instituto Gulbenkian de Ciência—Rua da Quinta Grande, 6, 2780-156 Oeiras, Portugal; etranfield@igc.gulbenkian.pt (E.M.T.); alsousa@igc.gulbenkian.pt (A.L.S.)

<sup>5</sup> iNOVA4Health—Advancing Precision Medicine, CEDOC—Chronic Diseases Research Centre, NOVA Medical School/Faculdade de Ciências Médicas, Universidade NOVA de Lisboa, Campo dos Martíres da Patria, 130, 1169-056 Lisboa, Portugal; ana.farinho@nms.unl.pt

<sup>6</sup> Department of Pneumology, Unidade Local de Saúde da Guarda (USLGuarda), 6300-659 Guarda, Portugal; luis.vaz.rodrigues@gmail.com

<sup>7</sup> Unidade de Técnicas Invasivas Pneumológicas, Pneumologia II, Hospital Pulido Valente, Centro Hospitalar Lisboa Norte, 1649-028 Lisbon, Portugal; paulagpinto@gmail.com (P.P.); 37749@chln.min-saude.pt (L.M.); tavaresdeabreu@gmail.com (T.T.d.A.); jasededo@netcabo.pt (J.S.)

<sup>8</sup> Instituto de Saúde Ambiental (ISAMB), Faculdade de Medicina, Universidade de Lisboa, Centro Hospitalar Universitário Lisboa Norte, 1649-028 Lisbon, Portugal; cristina.barbara@chln.min-saude.pt

<sup>9</sup> Instituto de Investigação e Inovação em Saúde (I3S), Universidade do Porto, 4200-135 Porto, Portugal; sseixas@ipatimup.pt

<sup>10</sup> Institute of Bioinformatics, Discoverer building, ITPL, Bangalore 560066, India; prashant@ibioinformatics.org (P.K.); pandey.akhilesh@mayo.edu (A.P.)

<sup>11</sup> Manipal Academy of Higher Education (MAHE), Manipal 576104, India

<sup>12</sup> Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN 55905, USA

\* Correspondence: ana.carvalho@nms.unl.pt (A.S.C.); rune.matthiesen@nms.unl.pt (R.M.)

Received: 15 October 2020; Accepted: 19 November 2020; Published: 20 November 2020



**Simple Summary:** Bronchoalveolar lavage is routinely collected during bronchoscopy for cytology analysis in the diagnostic of lung cancer. Due to low sensitivity of this method, early-stage cancers are undetected, lowering the treatment success. In this study, we analyzed extracellular vesicles isolated from bronchoalveolar lavage of lung cancer suspects by mass spectrometry-based proteomics. The protein composition of bronchoalveolar lavage extracellular vesicles of late-stage cancer showed a higher proteome complexity associated with mortality within the two year follow-up period. We identified a potential therapeutic target DNMT3B complex which was significantly expressed in bronchoalveolar lavage extracellular vesicles as well as in tumor tissue. Bronchoalveolar lavage

extracellular vesicles proteome analysis of immune markers indicates the presence of markers of innate immune and fibroblast cells.

**Abstract:** Acellular bronchoalveolar lavage (BAL) proteomics can partially separate lung cancer from non-lung cancer patients based on principal component analysis and multivariate analysis. Furthermore, the variance in the proteomics data sets is correlated mainly with lung cancer status and, to a lesser extent, smoking status and gender. Despite these advances BAL small and large extracellular vesicles (EVs) proteomes reveal aberrant protein expression in paracrine signaling mechanisms in cancer initiation and progression. We consequently present a case-control study of 24 bronchoalveolar lavage extracellular vesicle samples which were analyzed by state-of-the-art liquid chromatography-mass spectrometry (LC-MS). We obtained evidence that BAL EVs proteome complexity correlated with lung cancer stage 4 and mortality within two years' follow-up ( $p$  value = 0.006). The potential therapeutic target DNMT3B complex is significantly up-regulated in tumor tissue and BAL EVs. The computational analysis of the immune and fibroblast cell markers in EVs suggests that patients who deceased within the follow-up period display higher marker expression indicative of innate immune and fibroblast cells (four out of five cases). This study provides insights into the proteome content of BAL EVs and their correlation to clinical outcomes.

**Keywords:** extracellular vesicles; lung cancer; proteomics; immuno oncology; bronchoalveolar lavage

---

## 1. Introduction

At present, lung cancer persisted as the most prevalent oncological disease with an estimated incidence of more than 2 million new cases worldwide, and a mortality rate of 84% with a total of 1.8 million deaths yearly [1,2]. The late detection of the disease is the main cause of such a dismal outcome as demonstrated by the five-year overall survival rates for non-small cell lung cancer (NSCLC): 14–49% for stages I to IIIA, and <5% for stage IIIB/IV [3].

The characterization of lung tumor tissue and the surrounding microenvironment at the molecular level contributed to the increased knowledge on the physiology of the disease as well as the design of novel therapies. Nevertheless, the collection of tissue biopsies of peripheral adenocarcinomas in distal airways by minimally invasive techniques is limited, and during the past years, several studies have targeted different types of biological specimens, which range from tumor tissues to different types of liquid biopsies [4]. Bronchoalveolar lavage (BAL) liquid biopsies are obtained from minimally invasive procedures, for example, fiber optic bronchoscopy. Bronchoscopy is typically performed upon suspicion of lung cancer, for example, derived from previously obtained imaging of the lungs. Due to the closer anatomical proximity of BAL compared to saliva, more accurate biomarker signatures are expected from BAL than from biofluids such as blood, saliva, or sputum. Blood and saliva, on the other hand, might serve as potential markers for early screening. In the past, several molecular technologies were applied to profile the molecular content of BAL such as miRNA [5], mRNA [6], DNA [7], DNA methylation [8], metabolites [9], microbiota [10], and proteomics. Over the years, proteomics technologies have evolved and consequently, BAL proteome profiling has been attempted multiple times. The first BAL proteomics profiling applied MALDI TOF-MS [11]. Later, 2D-PAGE and MALDI-TOF were combined [12]. Nowadays, liquid chromatography-mass spectrometry (LC-MS) approaches are typically applied for BAL proteome profiling [13–15]. Abundant proteins like albumin partially hamper the protein coverage obtainable by direct LC-MS analysis of acellular BAL samples. Therefore, Sim et al. [16] established a novel methodology based on combining antibody-based depletion of high abundant BALF proteins, high pH peptide fractionation, and label-free quantitation on a high-resolution Orbitrap Fusion Lumos instrument. However, antibody depletion of abundant

proteins risks the removal of clinically relevant target proteins. Consequently, it is imperative to explore additional sample preparation methodologies.

The enrichment of small (sEVs) and large (IEVs) EVs represents an alternative methodology for the depletion of abundant proteins in biomarker studies based on liquid biopsies [17]. Currently, no study has targeted BAL sEVs and IEVs from suspected lung cancer patients. We, therefore, performed an exploratory case-control study to evaluate the protein complexity from proteomics profiling of BAL EV fractions. The sEVs proteins constituted the most complex proteome when compared to acellular BAL, vesicle-depleted BAL, and IEVs. We identified and quantified 7484 protein isoforms from 3158 encoding genes from sEVs. The IEVs and sEVs were enriched by sequential centrifugation followed by ultracentrifugation and floatation on sucrose gradient cushion, which captures all the subpopulations of sEVs while depleting from non-vesicular protein aggregates and complexes. Besides, the potential of BAL sEV proteome as a source of biomarkers was explored. The data obtained suggested that BAL sEV proteome complexity correlates with cancer stage IV and death (Wilcoxon rank-sum test  $p$  value = 0.006). The proteomes were compared, by system biology approaches, to previously obtained proteomes from lung tumor tissues and acellular BAL proteome [13,18]. sEVs and tumor tissue displayed a common significant regulation of several enriched functional categories with potential therapeutic value. Finally, sEV protein markers for innate immune and fibroblast cells correlated with poor prognosis (mortality within two years follow-up in four out of five cases).

## 2. Results

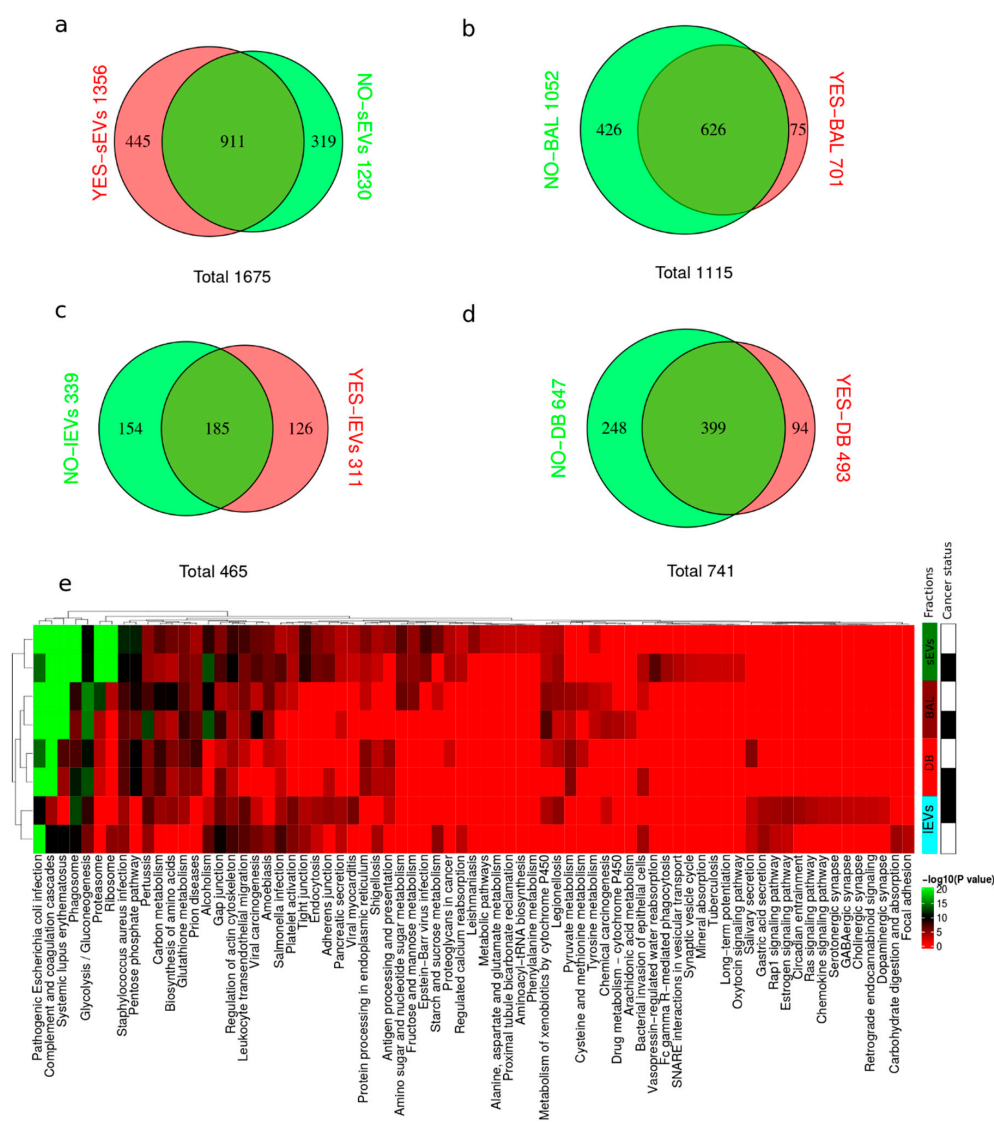
### 2.1. Outline of Study

As a proof of concept, we compared different BAL fractions by high-throughput mass spectrometry analysis such as acellular BAL, vesicle depleted BAL, IEVs and sEVs from control, and lung cancer patients. The proteome content was compared by LC-MS (Figure S1). We observed that sEVs contained the larger proteome complexity with the most significant potential for biomarker discovery (Section 2.2 and Figure S2). Therefore, sEVs were isolated from 24 acellular BAL samples and analyzed by mass spectrometry and further characterized by transmission electron microscopy (TEM), western blot, and nanoparticle tracking analysis. The quantitative MS data were compared with iBAQ values obtained from reanalyzed data of previous studies on acellular BAL (PXD004700) [13] and patient-matched tumor versus normal tissue (PXD000853) [18].

### 2.2. Proteome Content of IEVs, sEVs, Acellular BAL, and EV Depleted BAL (DB)

BAL fluids (acellular BAL, aBAL) from two different cancer status samples (NO and YES) were fractionated by sequential centrifugation and ultracentrifugation into three fractions: (1) depleted BAL (DB), (2) IEVs and (3) sEVs (Figure S1). To address which BAL fraction holds the greatest promise in terms of significant cancer status discrimination based on protein identification and quantitation, each fraction was analyzed by LC-MS in parallel with the respective aBAL fluid. Figure 1 compares the protein identification in cancer versus control for (a) sEVs, (b) BAL, (c) IEVs, and (d) DB. The sEVs fraction resulted in the highest number of protein identifications and the highest number of proteins unique to cancer (Figure 1a). KEGG functional enrichment of all the identifications from each of the fractions was compared to address the question of which fraction has the greatest potential for separating cancer status (Figure 1e). Figure S2 displays the overlap between the identified proteins in each fraction type obtained from LC-MS analysis. The highest numbers of quantified proteins were obtained in BAL sEVs from YES and NO cancer status samples. On average the number of identified proteins from sEVs was almost double compared to aBAL, while in BAL vesicle-depleted and IEVs fractions the number of identified proteins was lower, with the IEVs proteome being the simplest proteome (Figure S2). KEGG functional enrichment analysis of all identified proteins revealed functional diversity between the BAL fractions analyzed (Figure 1e). The samples clustered according to BAL fractions based on the functional KEGG annotation, strongly suggesting that each fraction contains distinct protein functional groups. The identified proteins overall

exhibited significant enrichment in pathways related to infection, inflammation, and complement and coagulation pathways. sEVs uniquely displayed significant enrichment in proteasome and ribosome factors while no enrichment in glycolysis and glycogenesis factors was observed in contrast to the other BAL fractions (Figure 1). Proteasome and ribosome factors are known as abundant factors in cancer due to increased protein turnover. Although IEV fractions resulted in the lowest amount of identifications, it presented unique functional enrichment related to, e.g., chemokine and synapse signaling in cancer. In turn, the two vesicles containing fractions displayed the maximum separation in terms of sample clustering. None of the fractions fully encapsulated all enriched functional categories. Proteomics analysis of sEV fractions showed the highest number of identified proteins and the highest diversity of functional enriched KEGG pathways. In this view, we have analyzed the proteome of sEVs for biomarkers discovery in a case-control study of 12 controls and 12 lung cancer cases.



**Figure 1.** Proteome overview in bronchoalveolar lavage (BAL) fractions. Venn diagrams comparing cancer (YES) versus control (NO) in: (a) sEVs, (b) BAL, (c) IEVs, and (d) DB; (e) Heat-map of KEGG pathway enrichment analysis. Functional enrichment analysis was performed using R. *p*-value of each KEGG pathway was <0.05. The shades of the colors reflected the  $-\log_{10}(p \text{ value})$  of the enrichment analysis of proteins identified in different subcellular BAL fractions for NO (white) and YES (black) cancer status samples. Large extracellular vesicles (IEVs), BAL vesicle-depleted (DB), acellular BAL (BAL), small extracellular vesicles (sEVs).

### 2.3. Baseline Characteristic of Clinical Samples for the Enrichment of sEVs

To explore the potential of sEVs proteome ability to classify clinical samples, we selected 12 controls (non-lung cancer cases) and 12 lung cancer samples in a case-control design (Table 1) from our previous analyzed cohort of 91 patients suspected of lung cancer [13]. The lung cancer status reflects the diagnosis after two years of follow-up. Approximately 42% of the cancer patients died within the two year follow-up period. The selection of samples for sEV characterization was optimized to prevent statistical association of age, gender, smoking status, and experimental batch effects with lung cancer status. As expected, the lung cancer status displayed an association to two-year survival and cancer staging. sEVs from the 24 clinical samples were isolated following the protocol outlined in (Figure S1).

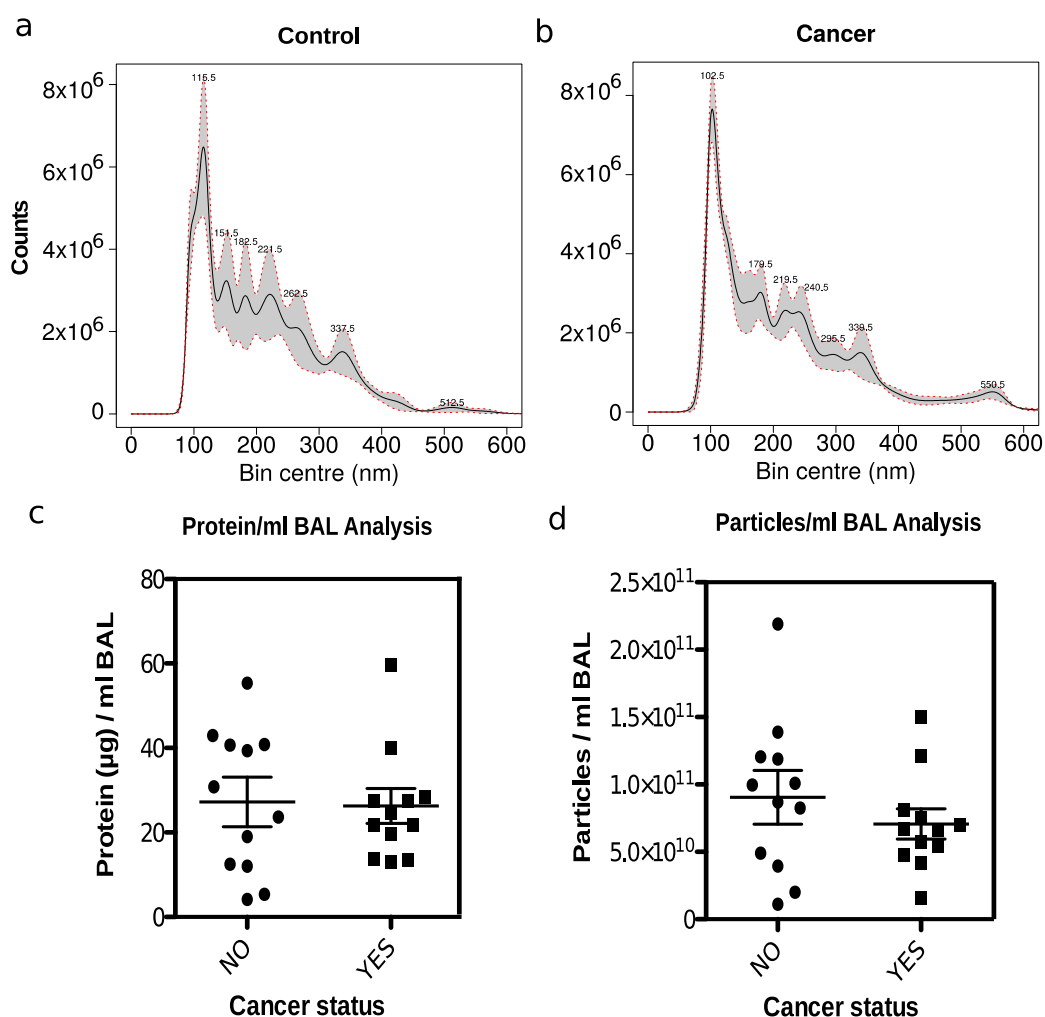
**Table 1.** Baseline characteristic of the clinical samples.

	NO (N = 12)	YES (N = 12)	Total (N = 24)	p Value
Age				0.355
<55	4 (33.3%)	1 (8.3%)	5 (12.5%)	
>55	1 (8.3%)	1 (8.3%)	2 (8.3%)	
NA	7 (58.3%)	10 (83.3%)	17 (70.8%)	
Smoking History				0.370
Current smoker	4 (33.3%)	3 (25.0%)	7 (29.2%)	
Former smoker	2 (16.7%)	6 (50.0%)	8 (33.3%)	
Nonsmoker	2 (16.7%)	1 (8.3%)	3 (12.5%)	
Unknown	4 (33.3%)	2 (16.7%)	6 (25.0%)	
Gender				0.100
NA	0 (0.0%)	1 (8.3%)	1 (4.2%)	
F	8 (66.7%)	3 (25.0%)	11 (45.8%)	
M	4 (33.3%)	8 (66.7%)	12 (50.0%)	
Batch				0.390
1	2 (16.7%)	5 (41.7%)	7 (29.2%)	
2	9 (75.0%)	6 (50.0%)	15 (62.5%)	
3	1 (8.3%)	1 (8.3%)	2 (8.3%)	
Stage				<0.001
2	0 (0.0%)	1 (8.3%)	1 (4.2%)	
3	0 (0.0%)	1 (8.3%)	1 (4.2%)	
4	0 (0.0%)	5 (41.7%)	5 (20.8%)	
NA	1 (8.3%)	5 (41.7%)	6 (25.0%)	
No	11 (91.7%)	0 (0.0%)	11 (45.8%)	
Status				0.012
Alive	12 (100.0%)	7 (58.3%)	19 (79.2%)	
Dead	0 (0.0%)	5 (41.7%)	5 (20.8%)	

### 2.4. Nanoparticle Tracking Analysis of Isolated EVs

The enriched sEVs were characterized by nanoparticle tracking analysis (Figure 2 and Figure S3). Figure 2 displays representative particle size distribution for one control (a) and one lung cancer sample (b). All the normalized particle distributions are plotted in Figure S3 and color-coded with clinical status. The distributions display high similarity across samples. Based on a t-test on the normalized counts for each particle size bin of 0.5 nm, no statistical differences between the distributions for control versus cancer were observed (Figure S3). All distributions displayed a maximum peak around 100 nm, which represents the expected size mode of sEVs. The precise position of the maximum peak differed slightly between samples. The different samples displayed slightly different peak tops at larger particle sizes likely representing different types of non-vesicle particles such as protein complexes, aggregates, and eventually virus particles. A comparison of overall EV protein content

and overall total particle counts from control versus cancer samples showed no significant differences (Figure 2c,d).

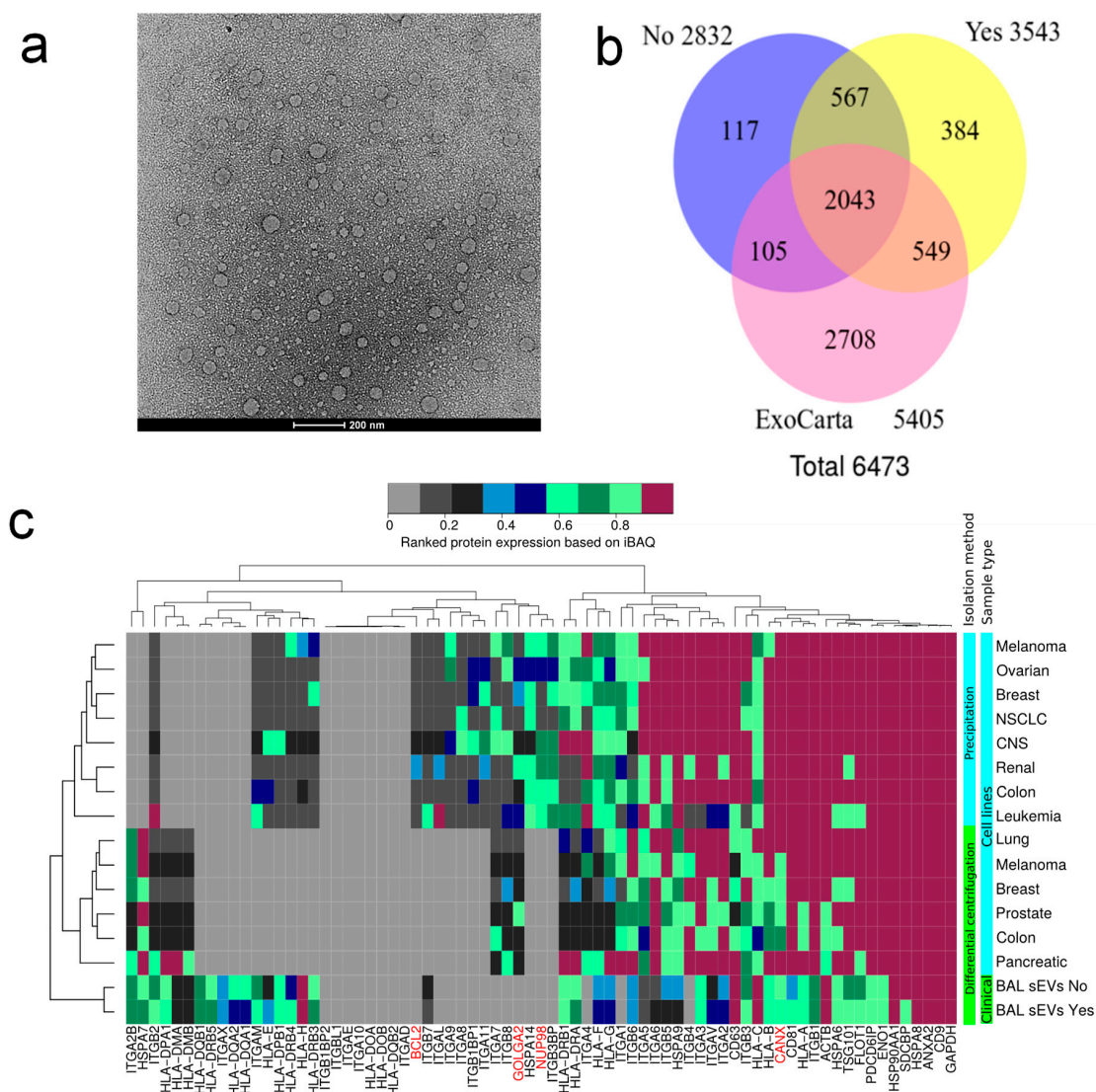


**Figure 2.** Nanoparticle tracking analysis of sEV samples: (a) particle counts versus particle size in nm for a control sample; (b) and a lung cancer sample; (c) protein concentration in sEVs versus cancer status; (d) particle concentration versus cancer status.

### 2.5. Transmission Electron Microscopy and Overall Proteome of EVs

In accordance with nanoparticle tracking analysis, TEM analysis of sEVs revealed vesicles of similar sizes around 100 nm (Figure 3a). Western blot analysis of the exosome marker CD63 comparing BAL to sEVs fraction showed a fortyfold enrichment (Figure S4). All identified proteins from controls and lung cancer in sEV were compared to proteins deposited in the exosome database ExoCarta, showing an overlap >70% for the two different cancer status samples (Figure 3b). However, such comparisons exclude protein quantitation information and consequently are not suited for assessing the exosome content in the sEV fractions. We, therefore, developed a bioinformatics methodology to quantitatively estimate exosome content in sEV fractions (Figure 3c). The heatmap displays average iBAQ values ranked from 0 to 1 for frequently reported exosome markers in the literature and the 10 most frequently reported proteins in ExoCarta. The quantitative values from cell line sEVs were obtained by re-analyzing previously published MS data and serve as reference samples. The reference data for sEVs from cell lines were obtained by two different EV enrichment methodologies: (1) PEG-based precipitation [19] and (2) differential ultra-centrifugation [20]. The two methods resulted in the identification and quantification of similar exosome markers. We compared the quantitative proteomics data on sEVs

from human clinical samples to the reference data on sEVs from cell lines. This constitutes a prudent MS-based assessment of exosome content. The clinical BAL sEVs isolated in this study exhibited similar expression levels of exosome markers when compared to the previous data from sEVs isolated from cell lines (Figure 3c) [19,20]. Additionally, the level of non EV protein is considerably lower in BAL sEVs compared to cell lines sEVs enriched using the two methods while PEG isolated EVs from cell lines displayed the highest level of non EV protein markers.

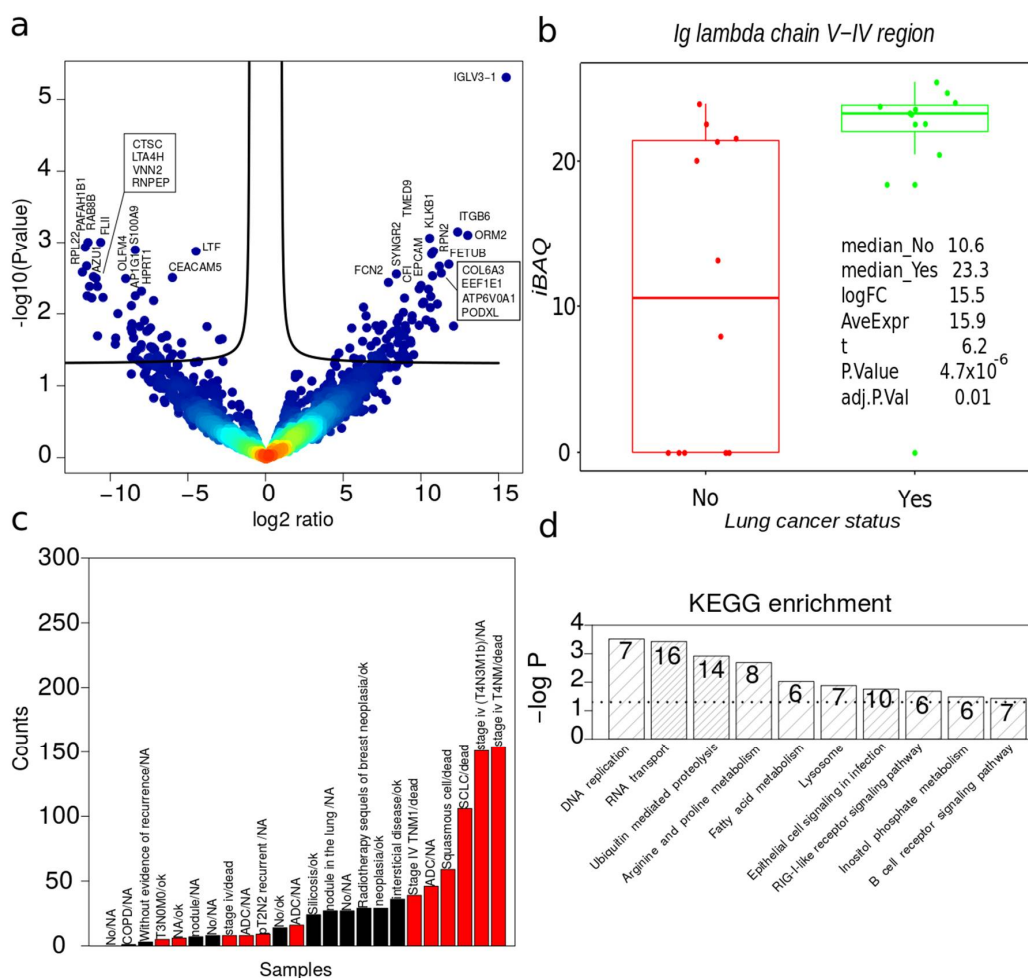


**Figure 3.** Extracellular vehicle (EV) quality assessment: (a) representative transmission electron microscopy images of enriched sEVs; (b) Venn diagram indicating overall protein identification overlap with ExoCarta database; (c) MS-based quantitative comparison between enriched BAL sEVs and cell line isolated EVs of frequently reported exosome protein markers and 10 most abundant exosome marker from ExoCarta. Red labeled proteins indicate non EV proteins.

### 2.6. Dysregulated Proteins in BAL Exosomes

Principal component analysis of all the quantitative values separated NO versus YES cancer samples based on three different principal components (Figure S5). The statistical analysis of iBAQ expression values by the R package limma applying a correction for gender and smoking resulted in 166 genes significantly regulated ( $p$ -value < 0.05) and one protein after correction of multiple testing (adjusted  $p$ -value < 0.05) between cancer positive and negative status (Table S1 and Figure 4a,b).

Previously, we reported 133 significant regulated proteins after correction of multiple testing [13], in which 91 acellular BAL samples were analyzed. We speculate that this difference primarily results from the difference in the number of cases studied (24 versus 91) rather than reflecting BAL sEVs potential as a biomarker source. In other words, due to cancer heterogeneity, a higher number of samples are required to define proteins significantly dysregulated after correction of multiple testing. To further explore protein heterogeneity of sEVs samples, uniquely identified proteins in each sample were extracted. Figure 4c depicts the number of unique proteins, finding 607 unique proteins (Table S2) in total for sEVs from lung cancer versus 176 for control. Counting the number of uniquely identified proteins per sample revealed that stage IV cancer and mortality displayed a significant tendency to contain more unique proteins identified compared to other cases ( $p$ -value = 0.006, Figure 4c). The  $p$ -value was calculated based on Wilcoxon rank-sum test without removing any outliers. This trend of increased protein complexity in cancer was confirmed in lung cancer tissue (Figure S6b) but not for acellular BAL proteome (Figure S6a). Unique cancer sEVs proteins displayed KEGG functional regulation in pathways associated with cancer (Figure 4d).

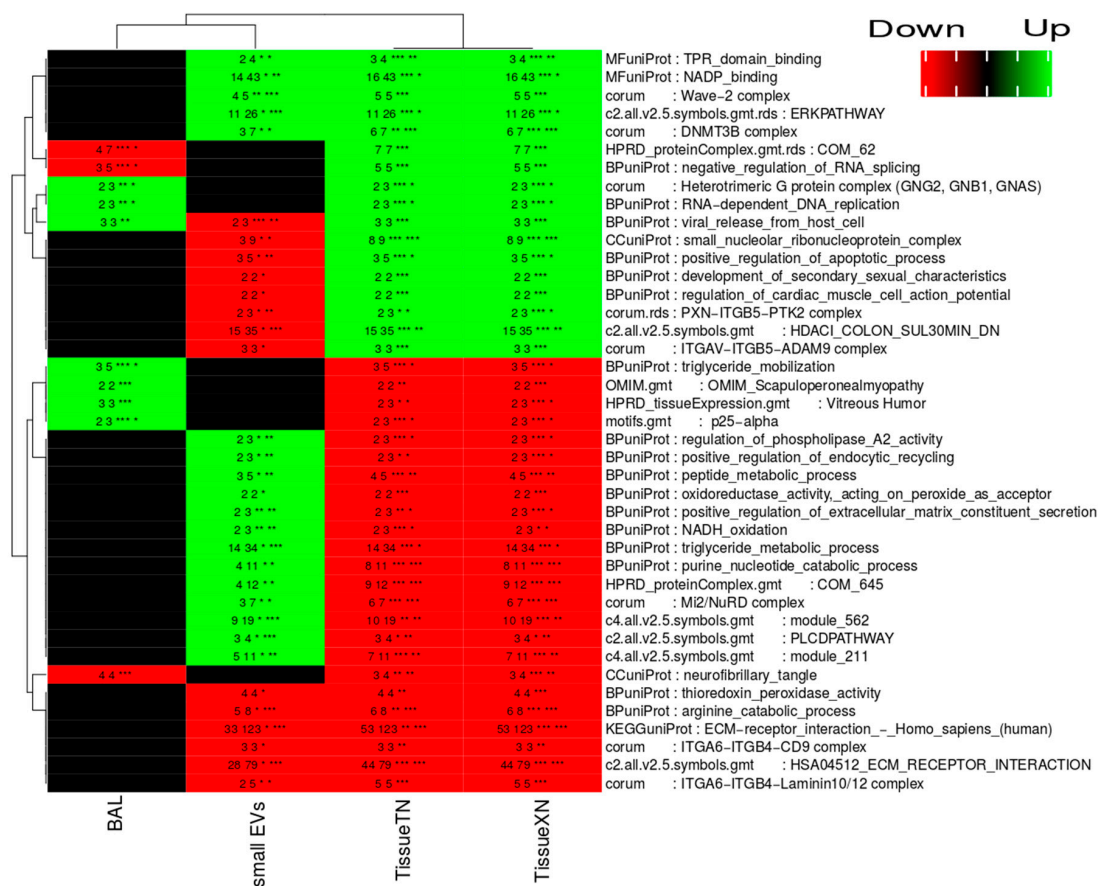


**Figure 4.** Dysregulated lung cancer proteins in BAL EVs: (a) Volcano plot summarizing the main dysregulated proteins where  $-\log_{10} p$  values were corrected for gender and smoking; (b) top regulated protein after correction for multiple testing; (c) number of unique proteins per sample (red bars indicate lung cancer and black bars represent controls); (d) KEGG functional enrichment of cancer unique proteins. The numbers in the bars represent the number of cancer unique proteins in the functional category. The dashed line indicates the  $p$ -value threshold ( $<0.05$ ).



### 2.7. Complete Functional Regulation

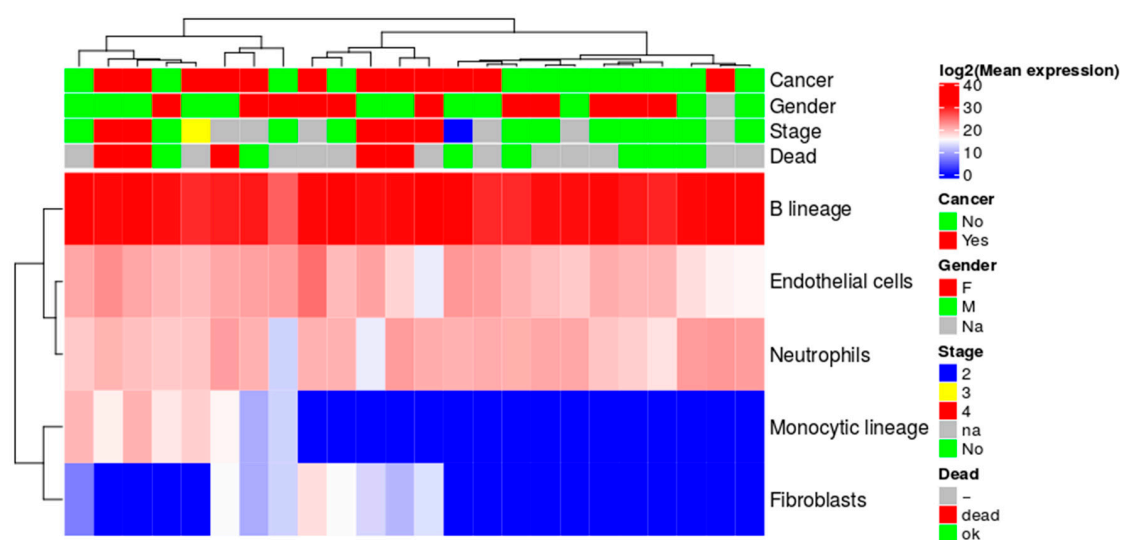
Complete functional regulation analysis tests for significant regulation of entities (genes, proteins) within a functional group together with significant enrichment in terms of the identification of a given functional annotated group [21]. The methodology combines functional regulation and enrichment analysis into a single visual summary. Figure 5 compares the results from the complete functional regulation analysis of acellular BAL, sEVs, lung tumor tissue, and mouse xenotransplant lung tumor tissue. For each functional group, the number of proteins identified in each sample type is displayed as well as the total number of proteins in the respective category. A statistical test is performed by assessing the significance of the enrichment and the significance of the overall regulation of the category. Complete functional regulation analysis revealed higher similarity between sEVs and tumor tissue when compared to acellular BAL (Figure 5). The large functional groups up-regulated in both tissue and sEVs include NADP binding and ERK pathway. ECM receptor interaction constitutes the largest down-regulated functional category common to tissue and sEVs (Figure 5). Several functional categories revealed reverse regulation when comparing sEVs with tumor tissue, for example, triglyceride metabolic processes. DNMT3B complex showed up-regulation in lung cancer for sEVs and tumor tissue. Aberrant DNA methylation caused by the DNMT3B complex in lung cancer is well established and proposed as a possible therapeutic target.



**Figure 5.** Complete functional regulation and enrichment analysis of proteins identified in acellular BAL (BAL), small EVs, lung tumor tissue (TissueTN), and xenotransplant lung tumor tissue (TissueXN). The color code in the heatmap reflects the regulation level (black unregulated, green significant up-regulated in tumor, and red significantly down-regulated in tumor). The text in the cells indicates the number of proteins identified in each functional category, the number of proteins annotated in the given functional category and the significance level of regulation, and significance level of enrichment (\*  $p < 0.05$ , \*\*  $p < 0.01$ , and \*\*\*  $p < 0.001$ ).

### 2.8. Quantitative Evaluation of Immune Cell Lineage Markers

BAL contains cells from various sources like tumor cells, epithelial cells, immune cells, and fibroblasts [22,23]. Based on the R package MCPcounter [24], we estimated the average abundance scores of markers for eight different immune and two stromal cell populations (Figure 6). MCPcounter identified six protein markers in BAL, 15 in sEVs, and 17 in tissue samples. In BAL sEVs, B lineage markers resulted in the highest average abundance scores. BAL sEVs from lung cancer tend to display higher abundance scores for monocytic lineage and fibroblasts than controls (Figure 6). This trend was not confirmed for acellular BAL proteome (Figure S7a). Markers of eight immune and two stromal cell populations could separate samples into normal, tumor, and xenotransplant tumor tissues (Figure S7b).



**Figure 6.** Average abundance scores of protein markers for eight different immune and two stromal cell populations in sEVs BAL samples.

### 3. Discussion

Deep protein coverage is paramount for identifying cancer signatures as well as contributing to the design of novel therapies [4]. Recent methodologies combining antibody-based depletion of high abundant BALF proteins, high pH peptide fractionation, and label-free quantitation on a high-resolution Orbitrap Fusion Lumos instrument demonstrated superior protein coverage [16]. In this study, we demonstrated the enrichment of sEVs by the sequential centrifugation method and quantification, resulted in a total of 7484 protein isoforms. The sEVs protocol typically yields sEVs fractions containing more than 90% vesicles according to flow cytometry experiments. Furthermore, the quality of the sEVs was accessed by nanoparticle tracking analysis, transmission electron microscopy, LC-MS, and western blotting. Webber and Clayton [25] argued that ratios  $>3 \times 10^{10}$  particles/microgram of protein corresponded to high vesicular purity whereas ratios of  $2 \times 10^9$  to  $2 \times 10^{10}$  particles/microgram of protein indicated low purity. Webber and Clayton achieved solely ratios  $>3 \times 10^{10}$  particles/microgram of protein from EVs isolated from cell culture conditioned media. On the other hand, Webber and Clayton [25] described for the body fluids urine and serum a 10-fold lower ratio similar to that obtained BAL for sEVs. In Figure 3c, we observe that the level of BCL2 (mitochondrial marker), GOLGA2 (Golgi marker), NUP98 (nuclear marker), and CANX (endoplasmic reticulum marker) is much lower in our clinical sEVs isolates than what was obtained from cell lines from previously published studies. Based on our observation, we speculate that EVs from body fluids bind more extracellular proteins than EVs from cell culture conditioned media thereby explaining the lower particles to protein ratio for clinical derived EVs. The identified proteins in aBAL and BAL sEVs displayed common functional KEGG pathways enriched in infection, inflammation, and complement

and coagulation cascades (Figure 1). Glycolysis/gluconeogenesis, together with amino acid metabolism, was enriched for BAL but completely depleted for sEVs. In contrast, “antigen processing and presentation” and proteins involved in virus infection were enriched preferentially in sEVs rather than BAL. In fact, many HLA variants were identified in BAL sEVs. HLA peptidome has recently gained interest as a source of neoantigen and biomarkers [26]. Our findings reinforce the novel methodologies for enriching for HLA for MS-based peptidome studies. The soluble HLA has been reported as a biomarker in various studies [27]. However, according to our data, neither aBAL nor depleted BAL showed enrichment for HLA suggesting that soluble HLA is mainly associated with sEVs.

Proteasome and ribosome factors were more enriched in sEVs than aBAL. These protein factors may co-purify during EVs enrichment based on densitometry. A recent study based on biotinylation of external proteins demonstrated that proteasome is located externally in EVs [28].

Here, we proposed a novel MS-based bioinformatics assessment of sEV for exosome markers. We further showed that the complete profiling of previously proposed exosome markers demonstrated that BAL sEVs contain a similar abundance of exosome markers compare to sEVs from cell lines. This is an efficient method compared to western blot (Figure 3c). We used 8 mL of acellular BAL for which we typically obtained 20 to 80 micrograms of sEV proteins. This limits the number of western blots that can be performed, as MS-based proteomics (approximately 2 micrograms) requires less protein compared to western blotting techniques (>5 micrograms).

Notably, we observed less significant regulated proteins after correction of multiple testing compared to our previous study on acellular BAL proteome [13]. We ascribe this to tumor heterogeneity, which means that larger cohorts are needed to define statically significant patterns in tumor sEVs. The unique proteins per sample supported this view (Figure 4b) and suggested that protein complexity in BAL sEVs and tissue but not acellular BAL correlates with tumor burden (Figure 4b and Figure S6). The follow-up studies on larger cohorts are needed to validate this observation. Furthermore, clinical utility must be validated in larger cohorts.

The complete functional regulation and enrichment analysis demonstrated that the functional regulation in sEV proteome revealed higher similarity to tumor tissue compared to acellular BAL. Similar results were obtained when comparing 60 NCI cell lines cellular proteome versus EV proteome using functional analysis [17]. The largest functional category in terms of identified proteins and significantly up-regulated in both BAL EVs and tissue were proteins involved in ERK pathway activation. This pathway is well described as aberrantly activated in cancer [29], including lung cancer [30,31]. Clinical targeting has been unsuccessful due to drug resistance and toxicity [32]. This highlights the emerging role of extracellular vesicle-mediated drug resistance in cancers ascribed to intercellular transfer of drug-resistant traits such as proteins or nucleic acids [33]. Efforts to elucidate the role of EVs in drug resistance implicated in downstream effectors of EGFR signaling cascade involving RAS-RAF-MEK-ERK pathways are therefore much needed. For example, it is conceivable that EGFR in EVs affects drug resistance by a similar mechanism as previously described in which CD20 targets, excessively secreted in EVs, shield the CD20 cellular target on cancer cells [34]. DNMT3B (Cytosine-5-methyltransferase 3 $\beta$ ) complex proteins were also up-regulated in sEVs and tumor tissue. DNMT3B is an epigenetic modifier and a downstream substrate of EGFR signaling cascade via AKT pathway axis. Previous studies found that cancer cells express aberrant DNMT3B transcripts encoding truncated proteins which are hypermethylated, and transcriptionally inactivate tumor suppressor genes [35]. Various solid cancers overexpress DNMT3B transcripts including lung cancer, and DNMT3B is considered a potential therapeutic target [36]. This raises the question of whether EVs can shield the target tumors from DNMT3B inhibitors. Additionally, DNMT3B has been reported to contribute in chemotherapy resistance [37] and promotes radio-resistance in nasopharyngeal carcinoma [38]. DNMT3B in BAL EVs mirrors the expression in lung tumor tissue, and this might give a novel opportunity to use it as a marker for chemotherapy and radio-resistance. A second main functional category up-regulated in sEVs showing the same direction of regulation as in tumor tissue was NADP-binding which includes proteins interacting selectively and non-covalently with the oxidized form, NADP<sup>+</sup>, or the reduced form,

NADPH. Evidence has been gathered showing that the alteration of NADPH metabolism is crucial for cell proliferation and survival. For example, the malic enzyme showed a four-fold change in tumor vs. control has been described to enhance the pentose phosphate pathway through direct binding and activating 6-phosphogluconate dehydrogenase, leading to tumor cell proliferation [39].

Integrins are bidirectional cell surface molecules are known to play a role in cancer cell survival and chemoresistance. In some tumors, higher integrin expression was associated with increased malignancy and metastasis formation. However, the down-regulation of integrins was also reported in cellular proliferation and invasion through epithelial-to-mesenchymal transition [40,41]. Herein, we demonstrate that ECM receptor interactions with its cellular constituents are downregulated in sEVs which may cause weak protein-protein interactions of adhesion proteins.

Macrophages (60%), lymphocytes (24%), and neutrophils (13%) represent the main cellular components in BAL [42]. The acellular BAL and BAL sEVs proteome mainly contained markers of B cell lineage. Several have reported the role of B cells in tumor immunity. For example, Bruno et al. observed an increased number of intratumoral B cells in freshly isolated primary tumors from NSCLC patients relative to B cells from tumor-adjacent tissues [43]. These tumor infiltrating lymphocytes B can generate anti-tumor antibodies [44], which in turn activate dendritic cells (DCs) to internalize tumor antigens and subsequently activate tumor-reactive T cells. Even though sEVs B cell markers expression was invariant between control and cancer, *IGLV3-1* was significantly upregulated, after correcting for multiple testing (>one order of magnitude), in sEVs BAL fraction from cancer patients compared to control (Table S1). It would be interesting to explore the role of immunoglobulin-bound vesicles in tumor immune evasion. The binary complex could function as a pro-tumor effect, protecting cancer cells from the attack of cytotoxic T-cells.

The abundance scores for eight immune and two stromal cell markers in acellular BAL were randomly clustered in cancer versus non-cancer samples (Figure S7a). For BAL sEVs, monocytic lineage, and fibroblasts marker scores weakly, which correlated with cancer and poor prognosis (Figure 6). The immune and two stromal cell markers' abundance scores perfectly clustered the normal tumor tissues and xenotransplanted tumor tissue into distinct clusters (Figure S7b). It is imperative to understand the tumor cell's contribution to the sEVs population in comparison to the immune and stromal cells pool.

It will be necessary to deconvolute the contribution of each entity being, vesicles, soluble proteins, and cell populations (tumor, immune system, and stromal compartment) in the BAL fluid of lung cancer patients to provide clearer evidence of the role in tumor immune evasion that can be translated into a targeted treatment. The similarity in regulated KEGG functions in sEVs and tumor tissue appears promising in terms of exploring sEVs for diagnostic and prognostic markers in lung cancer. This along with a high protein coverage compared to other proposed fractionation methodologies of BAL sEVs constitute an interesting target in lung diseases in general.

## 4. Materials and Methods

### 4.1. Patient Samples

BAL samples were collected prospectively from patients with suspected lung cancer undergoing diagnostic bronchoscopy at the Unidade de Técnicas Invasivas Pneumológicas, Pneumologia II, Hospital Pulido Valente, Centro Hospitalar Lisboa Norte. Human Ethics approval was received from Centro Hospitalar Lisboa Norte and National Health Institute Dr. Ricardo Jorge Research Ethics Committees (DIRCLN-8ABR2014-130). All methods were performed in accordance with the relevant guidelines and regulations. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Informed consent of all participants was obtained. A total number of 91 BAL samples were collected between April and July 2014, which was followed up two years later. Patient diagnosis was obtained by BAL

cytology and whenever available biopsy followed by histology. Patient disease status was assessed at two different time points in August 2014 and June 2016. Samples ( $n = 24$ ) used in the current study were selected to obtain 12 controls and 12 cancer patients. Non-lung cancer patients, referred to in this study as control, were diagnosed with diseases such as COPD, interstitial lung disease, bronchiectasis, heart failure, asthma, chronic cough, and pulmonary nodule follow-up. After the initial bronchoscopy, negative “Suspicious” patients were approached in the following way: a subset of the patients underwent CT-guided transthoracic needle biopsy, others repeated bronchoscopy and some of them were submitted to thoracic surgery. In the remaining patients presenting with small lung nodules, follow-up was performed according to Fleischner Society Guidelines [45]. The mean time length to the diagnosis of lung cancer was 100.61 days after the first bronchoscopy.

#### 4.2. BAL Processing

In most cases, BAL was targeted toward affected lung segments. The procedure was performed by wedging the bronchoscope in a subsegmental bronchus. Usually, three lavages were performed using approximately 50 mL of 0.9% saline solution per lavage. The recovered fluid was placed at 4 °C immediately. BAL was centrifuged at 320×  $g$  for 10 minutes at 4 °C to remove the cellular fraction. The resulting cell-free supernatant was immediately aliquoted and frozen at −80 °C until further analysis.

#### 4.3. Isolation of Extracellular Vesicles from Human Acellular BAL

Frozen acellular BAL specimens were thawed and centrifuged at 3000×  $g$  for 20 min at 4 °C and then at 12,000×  $g$  for 60 min at 4 °C. Clarified plasma was ultracentrifuged in a Beckman Coulter Optima™ L-80XP Ultracentrifuge at 100,000×  $g$  at 4 °C for 120 min with a Type 50.2 Ti rotor (k-factor: 157.7) to pellet EVs. The supernatant was carefully removed, and crude EV-containing pellets were resuspended in ice-cold phosphate-buffered saline (PBS), followed by floatation on sucrose cushion (30%, D<sub>2</sub>O) for 60 min at 100,000×  $g$  at 4 °C to remove non EV protein complexes. After washing by pelleting the EVs collected in the sucrose cushion for 16h at 100,000×  $g$  at 10 °C, EVs were resuspended in PBS and subjected to NanoSight NS300 analysis (Malvern Instruments, Inc., Westborough, MA, USA).

#### 4.4. Protein Measurements

Protein concentrations in isolated exosome fractions were measured using a BCA protein assay kit (Pierce Biotechnology, Rockford, IL, USA) according to the manufacturer’s instructions.

#### 4.5. Transmission Electron Microscopy

TEM of isolated EVs was performed at the Instituto Gulbenkian de Ciência—Electron Microscopy Facility. Freshly isolated EVs were put on a copper-palladium 100 mesh grid coated with 1% formvar in chloroform and carbon. The grids were glow-discharged before the adhesion of samples. Samples were fixed with 2% formaldehyde in PBS, washed ten times with distilled water, and stained with 2% uranyl acetate in ddH<sub>2</sub>O. A Tecnai G2 Spirit BioTWIN from FEI operating at 120 keV and equipped with an Olympus-SIS Veleta CCD Camera was used for imaging.

#### 4.6. Immunoblotting

BAL fluid and BAL derived-EVs proteins were quantified using Pierce™ BCA Protein Assay Kit (Thermo Scientific Pierce Biotechnology, P.O. Box 117, Rockford, USA). A constant protein amount of (10 μg) of BAL-EVs and BAL was separated on 10% SDS-PAGE under reducing conditions (4×-loading buffer: 8% SDS, 40% ( $w/v$ ) glycerol, 20% Beta-mercaptoethanol, 0.08% bromophenol blue) and transferred to a Hybond PVDF membrane (Amersham, GE, Chicago, IL USA) or stained with Coomassie brilliant blue (Merck, Kenilworth, NJ, USA). The membrane was then incubated with primary antibody directed against CD63 (AB0047-200, SICGEN at a dilution of 1:2500) and HRP-conjugated secondary antibody donkey anti-goat (Jackson ImmunoResearch, Cambridge, UK). Signals were

detected by enhanced chemiluminescence reagents ECL Western blot Detection (Amersham, GE, USA). Intensity of the bands was quantified using ImageLab Software. HeLa total cell lysate was analyzed in parallel using the same conditions as a positive control for anti-CD63. Images obtained were processed by the open-source software ImageJ ([imagej.net](http://imagej.net)) for quantitation.

#### 4.7. Peptide Sample Preparation

Protein solutions containing SDS and DTT were loaded onto filtering columns and washed exhaustively with 8M urea in HEPES buffer [46]. Proteins were reduced with DTT and alkylated with IAA. Protein digestion was performed by overnight digestion with trypsin sequencing grade (Promega).

#### 4.8. Mass Spectrometry Analysis

The peptides were analyzed on an Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Fisher Scientific, San Jose, CA, 95134, USA) coupled with an EASY nLC1200 nanoflow liquid chromatography system (Thermo Fisher Scientific). The peptides reconstituted in 0.5% FA (formic acid) were loaded on a trap column (Acclaim™ PepMap™ 100, LC C18, 5 μm, 100 μm × 2 cm, nanoViper, Thermo Fisher Scientific) at a flow rate of 4 μL/min. The peptides were resolved at 0.3 μL/min flow rate using an increasing gradient of solvent B (0.1% FA in 90% ACN) on an analytical column (Easy-Spray™ PepMap™ RSLC C18, 2 μm, 75 μm × 50 cm, Thermo Fisher Scientific), which was fitted with an EASY-Spray ion source that was operated at a voltage of about 2.1 kV. The total run time was 120 min. Mass spectrometry analysis was carried out in a DDA (data-dependent acquisition) with a full scan in the  $m/z$  (mass-to-charge ratio) range of 350 to 1800 in the “Top Speed” setting, 3 sec per cycle. MS1 and MS2 were acquired for the precursor ions and the peptide fragmentation ions, respectively. MS1 scans were measured at a resolution of 120,000 at an  $m/z$  of 200. MS2 scans were acquired by fragmenting precursor ions using the HCD (higher-energy collisional dissociation) method, which was set to 32% of collision energy with 5% of stepped collision energy, and detected at a mass resolution of 30,000 at an  $m/z$  of 200. AGC (automatic gain control) targets were set to one million ions for MS1 and 0.05 million ions for MS2. The maximum ion injection time was set to 50 ms for MS1 and 100 ms for MS2. The precursor isolation window was set to 1.6  $m/z$  with 0.4  $m/z$  of offset. Dynamic exclusion was set to 35 sec, and single charged ions were rejected. Internal calibration was carried out using the lock mass option ( $m/z$  445.12002) from ambient air.

#### 4.9. Database Search

The obtained data from the 24 LC-MS runs were searched using VEMS [47] and MaxQuant [48]. The MSMS spectra were searched against a standard human proteome database from UniProt (3AUP000005640). Permuted protein sequences, where Arg and Lys were not permuted, were included in the database for VEMS search. Trypsin cleavage allowing a maximum of 4 missed cleavages was used. Carbamidomethyl cysteine was included as a fixed modification. Methionine oxidation and N-terminal protein acetylation were included as variable modifications. 5 ppm mass accuracy was specified for precursor ions and 10  $m/z$  for fragment ions. The false discovery rate (FDR) for protein identification was set to 1% for peptide and protein identifications. No restriction was applied for minimal peptide length for VEMS search. The identified proteins were divided into evidence groups as defined by Matthiesen et al. [49].

#### 4.10. Functional Analysis of Identified Proteins in BAL Fractions

All the protein identifications by VEMS of each of the fractions BAL, depleted BAL, IEVs, and sEVs were subjected to KEGG functional analysis using the hypergeometric function as described previously [50]. The heatmap was plotted with the R package ComplexHeatmap [51].

#### 4.11. Quantitative Analysis

The quantitative comparison of acellular BAL, BAL exosomes, patient match normal versus tumor tissue, and patient-matched normal versus mouse xenotransplant tissue were based on MaxQuant [48]

analysis using the same parameter settings. Because of the difference in chromatography, each data set was analyzed separately and iBAQ values combined in R for further statistical analysis. IBAQ were preprocessed by three approaches: (1) removing common MS contaminants followed by  $\log_2(x + 1)$  transformation, (2) removing common MS contaminants followed by  $\log_2(x + 1)$  transformation and quantile normalization, and (3) removing common MS contaminants followed by  $\log_2(x + 1)$  transformation, quantile normalization and abundance filtering to optimize overall Gaussian distribution of all quantitative values. The top regulated proteins based on the different strategies were similar but with a small difference in the ranking of the proteins. The quantile normalized data were used for subsequent complete functional regulation analysis.

#### 4.12. Complete Functional Regulation Analysis

The quantitative data for acellular BAL, BAL exosomes, patient match normal versus tumor tissue, and patient-matched normal versus mouse xenotransplant tissue were subjected to completed functional regulation analysis as previously described [21]. The quantitative data was matched against the functional categories in KEGG [52], gene ontology [53], HPRD [54], CORUM [55], and MSigDB [56]. Only significantly enriched functional categories with an enrichment  $p$  value  $< 0.05$ , a regulation higher than two-fold, and  $p$  value on regulation  $< 0.05$  were maintained. The direction of the regulation of the functional categories across different sample comparisons was compared in a heatmap (Figure 5).

## 5. Conclusions

To our knowledge, we provide the first MS-based proteomic comparison of BAL fractions obtained by sequential centrifugation. We conclude that the proteome of sEVs displays the highest protein complexity of all the BAL fractions investigated. Our results prompt further studies in larger cohorts to clarify if BAL sEV proteomes can provide practical clinical information on lung cancer staging. Finally, the results highlight for the first time the role of EGFR downstream effectors in sEVs with potential implications in drug resistance.

**Supplementary Materials:** Mass spectrometry data is available in the PRIDE database (PXD021988) [57]. The following are available online at <http://www.mdpi.com/2072-6694/12/11/3450/s1>, Table S1. Statistical outcome from limma R package analysis. Figure S1. Schematic illustration of the sEVs and MV preparation. Figure S2. UpSet plot indicating the overlap between the identified proteins in each of the extracellular fractions for cancer and control samples. Figure S3. Normalized particle counts for all sEVs. Red indicates cancer patients and green controls. Figure S4. Principal component analysis using all  $\log_2$  transformed raw expression values. Figure S5. Unique proteins per sample identified in (a) acellular BAL and (b) tumor tissue. Figure S6. Western blot of CD63 enrichment in sEVs. Figure S7. Average expression of haematological markers: (a) based on acellular BAL proteomics; (b) normal (N), tumor (T), and xenotransplanted (X) tissue.

**Author Contributions:** Conceptualization, A.S.C. and R.M.; methodology, A.S.C. and R.M.; validation, A.S.C., I.F.-M., A.H. and S.Z.; formal analysis, A.S.C. and R.M.; Clinical samples, S.S., L.V.R., T.T.d.A., P.P., C.B. (Cristina Bárbara), L.M. and J.S.; EV enrichment, A.S.C., M.C.S.M., C.B. (Cristian Bodo) and B.C.-S.; MS sample preparation A.S.C.; MS analysis, C.H.N., P.K., A.P.; TEM analysis, A.F., A.L.S., E.M.T., and A.S.C.; resources, A.S.C. and R.M.; writing—original draft preparation, A.S.C. and R.M.; writing—review and editing, A.S.C. and R.M.; All authors have read and agreed to the published version of the manuscript.

**Funding:** R.M. is supported by Fundação para a Ciência e a Tecnologia (CEEC position, 2019–2025 investigator). This article is a result of the projects (iNOVA4Health—UID/Multi/04462/2013), supported by Lisboa Portugal Regional Operational Programme (Lisboa2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). This work is also funded by FEDER funds through the COMPETE 2020 Programme and National Funds through FCT—Portuguese Foundation for Science and Technology under the projects number PTDC/BTM-TEC/30087/2017 and PTDC/BTM-TEC/30088/2017. This work was supported by the Wellcome Trust/DBT India Alliance Margdarshi Fellowship (grant number IA/M/15/1/502023) awarded to A.P. B.C.-S., M.C.S.C. and C.B. are supported by the Champalimaud Foundation and the EMBO Installation Grant 3921.

**Acknowledgments:** Graphical abstract was constructed using BioRender (<https://biorender.com/>). We thank the nurses Célia Marina Cuco, Carla Lavareda, Francisco Miguel, Mafalda Ventura, and Sónia Almeida at Unidade de Técnicas Invasivas Pneumológicas, Pneumologia II, Hospital Pulido Valente, Centro Hospitalar Lisboa Norte, Lisbon, Portugal for assistance on collecting BAL samples.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Fidler, M.M.; Bray, F.; Soerjomataram, I. The global cancer burden and human development: A review. *Scand. J. Public Health* **2018**, *46*, 27–36. [[CrossRef](#)]
2. Barta, J.A.; Powell, C.A.; Wisnivesky, J.P. Global Epidemiology of Lung Cancer. *Ann. Glob. Health* **2019**, *85*, 8. [[CrossRef](#)] [[PubMed](#)]
3. Ko, E.C.; Raben, D.; Formenti, S.C. The Integration of Radiotherapy with Immunotherapy for the Treatment of Non-Small Cell Lung Cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **2018**, *24*, 5792–5806. [[CrossRef](#)] [[PubMed](#)]
4. Matthiesen, R. MS-Based Biomarker Discovery in Bronchoalveolar Lavage Fluid for Lung Cancer. *Proteom. Clin. Appl.* **2020**, *14*, e1900077. [[CrossRef](#)]
5. Lee, H.; Groot, M.; Pinilla-Vera, M.; Fredenburgh, L.E.; Jin, Y. Identification of miRNA-rich vesicles in bronchoalveolar lavage fluid: Insights into the function and heterogeneity of extracellular vesicles. *J. Control. Release Off. J. Control. Release Soc.* **2019**, *294*, 43–52. [[CrossRef](#)]
6. Venosa, A.; Katzen, J.; Tomer, Y.; Kopp, M.; Jamil, S.; Russo, S.J.; Mulugeta, S.; Beers, M.F. Epithelial Expression of an Interstitial Lung Disease-Associated Mutation in Surfactant Protein-C Modulates Recruitment and Activation of Key Myeloid Cell Populations in Mice. *J. Immunol.* **2019**, *202*, 2760–2771. [[CrossRef](#)] [[PubMed](#)]
7. Lu, R.; Zhao, X.; Li, J.; Niu, P.; Yang, B.; Wu, H.; Wang, W.; Song, H.; Huang, B.; Zhu, N.; et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *Lancet* **2020**, *395*, 565–574. [[CrossRef](#)]
8. Zhang, C.; Yu, W.; Wang, L.; Zhao, M.; Guo, Q.; Lv, S.; Hu, X.; Lou, J. DNA Methylation Analysis of the SHOX2 and RASSF1A Panel in Bronchoalveolar Lavage Fluid for Lung Cancer Diagnosis. *J. Cancer* **2017**, *8*, 3585–3591. [[CrossRef](#)]
9. Walmsley, S.; Cruickshank-Quinn, C.; Quinn, K.; Zhang, X.; Petrache, I.; Bowler, R.P.; Reisdorph, R.; Reisdorph, N. A prototypic small molecule database for bronchoalveolar lavage-based metabolomics. *Sci. Data* **2018**, *5*, 180060. [[CrossRef](#)]
10. Gomes, S.; Cavadas, B.; Ferreira, J.C.; Marques, P.I.; Monteiro, C.; Sucena, M.; Sousa, C.; Vaz Rodrigues, L.; Teixeira, G.; Pinto, P.; et al. Profiling of lung microbiota discloses differences in adenocarcinoma and squamous cell carcinoma. *Sci. Rep.* **2019**, *9*, 12838. [[CrossRef](#)]
11. Oumeraci, T.; Schmidt, B.; Wolf, T.; Zapatka, M.; Pich, A.; Brors, B.; Eils, R.; Fleischhacker, M.; Schlegelberger, B.; von Neuhoff, N. Bronchoalveolar lavage fluid of lung cancer patients: Mapping the uncharted waters using proteomics technology. *Lung Cancer* **2011**, *72*, 136–138. [[CrossRef](#)] [[PubMed](#)]
12. Pastor, M.D.; Nogal, A.; Molina-Pinelo, S.; Melendez, R.; Salinas, A.; Gonzalez De la Pena, M.; Martin-Juan, J.; Corral, J.; Garcia-Carbonero, R.; Carnero, A.; et al. Identification of proteomic signatures associated with lung cancer and COPD. *J. Proteom.* **2013**, *89*, 227–237. [[CrossRef](#)] [[PubMed](#)]
13. Carvalho, A.S.; Cuco, C.M.; Lavareda, C.; Miguel, F.; Ventura, M.; Almeida, S.; Pinto, P.; de Abreu, T.T.; Rodrigues, L.V.; Seixas, S.; et al. Bronchoalveolar Lavage Proteomics in Patients with Suspected Lung Cancer. *Sci. Rep.* **2017**, *7*, 42190. [[CrossRef](#)] [[PubMed](#)]
14. Almatroodi, S.A.; McDonald, C.F.; Collins, A.L.; Darby, I.A.; Pouniotis, D.S. Quantitative proteomics of bronchoalveolar lavage fluid in lung adenocarcinoma. *Cancer Genom. Proteom.* **2015**, *12*, 39–48.
15. Ortea, I.; Rodriguez-Ariza, A.; Chicano-Galvez, E.; Arenas Vacas, M.S.; Jurado Gamez, B. Discovery of potential protein biomarkers of lung adenocarcinoma in bronchoalveolar lavage fluid by SWATH MS data-independent acquisition and targeted data extraction. *J. Proteom.* **2016**, *138*, 106–114. [[CrossRef](#)]
16. Sim, S.Y.; Choi, Y.R.; Lee, J.H.; Lim, J.M.; Lee, S.E.; Kim, K.P.; Kim, J.Y.; Lee, S.H.; Kim, M.S. In-Depth Proteomic Analysis of Human Bronchoalveolar Lavage Fluid toward the Biomarker Discovery for Lung Cancers. *Proteom. Clin. Appl.* **2019**, *13*, e1900028. [[CrossRef](#)]
17. Carvalho, A.S.; Baeta, H.; Silva, B.C.; Moraes, M.C.S.; Bodo, C.; Beck, H.C.; Rodriguez, M.S.; Saraswat, M.; Pandey, A.; Matthiesen, R. Extra-cellular vesicles carry proteome of cancer hallmarks. *Front. Biosci.* **2020**, *25*, 398–436.
18. Li, L.; Wei, Y.; To, C.; Zhu, C.Q.; Tong, J.; Pham, N.A.; Taylor, P.; Ignatchenko, V.; Ignatchenko, A.; Zhang, W.; et al. Integrated omic analysis of lung cancer reveals metabolism proteome signatures with prognostic impact. *Nat. Commun.* **2014**, *5*, 5469. [[CrossRef](#)]



19. Hurwitz, S.N.; Rider, M.A.; Bundy, J.L.; Liu, X.; Singh, R.K.; Meckes, D.G., Jr. Proteomic profiling of NCI-60 extracellular vesicles uncovers common protein cargo and cancer type-specific biomarkers. *Oncotarget* **2016**, *7*, 86999–87015. [[CrossRef](#)]
20. Hoshino, A.; Kim, H.S.; Bojmar, L.; Gyan, K.E.; Cioffi, M.; Hernandez, J.; Zambirinis, C.P.; Rodrigues, G.; Molina, H.; Heissel, S.; et al. Extracellular Vesicle and Particle Biomarkers Define Multiple Human Cancers. *Cell* **2020**, *182*, 1044–1061. [[CrossRef](#)]
21. Carvalho, A.S.; Molina, H.; Matthiesen, R. New insights into functional regulation in MS-based drug profiling. *Sci. Rep.* **2016**, *6*, 18826. [[CrossRef](#)] [[PubMed](#)]
22. Pollock, K.; Albares, L.; Wendt, C.; Hubel, A. Isolation of fibroblasts and epithelial cells in bronchoalveolar lavage (BAL). *Exp. Lung Res.* **2013**, *39*, 146–154. [[CrossRef](#)]
23. Liao, M.; Liu, Y.; Yuan, J.; Wen, Y.; Xu, G.; Zhao, J.; Cheng, L.; Li, J.; Wang, X.; Wang, F.; et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat. Med.* **2020**, *26*, 842–844. [[CrossRef](#)] [[PubMed](#)]
24. Becht, E.; Giraldo, N.A.; Lacroix, L.; Buttard, B.; Elarouci, N.; Petitprez, F.; Selves, J.; Laurent-Puig, P.; Sautes-Fridman, C.; Fridman, W.H.; et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* **2016**, *17*, 218. [[CrossRef](#)] [[PubMed](#)]
25. Webber, J.; Clayton, A. How pure are your vesicles? *J. Extracell. Vesicles* **2013**, *2*. [[CrossRef](#)] [[PubMed](#)]
26. Purcell, A.W.; Ramarathinam, S.H.; Ternette, N. Mass spectrometry-based identification of MHC-bound peptides for immunopeptidomics. *Nat. Protoc.* **2019**, *14*, 1687–1707. [[CrossRef](#)]
27. Bassani-Sternberg, M.; Barnea, E.; Beer, I.; Avivi, I.; Katz, T.; Admon, A. Soluble plasma HLA peptidome as a potential source for cancer biomarkers. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 18769–18776. [[CrossRef](#)]
28. Cvjetkovic, A.; Jang, S.C.; Konecna, B.; Hoog, J.L.; Sihlbom, C.; Lasser, C.; Lotvall, J. Detailed Analysis of Protein Topology of Extracellular Vesicles-Evidence of Unconventional Membrane Protein Orientation. *Sci. Rep.* **2016**, *6*, 36338. [[CrossRef](#)]
29. Roberts, P.J.; Der, C.J. Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene* **2007**, *26*, 3291–3310. [[CrossRef](#)]
30. Chiu, L.Y.; Hsin, I.L.; Yang, T.Y.; Sung, W.W.; Chi, J.Y.; Chang, J.T.; Ko, J.L.; Sheu, G.T. The ERK-ZEB1 pathway mediates epithelial-mesenchymal transition in pemetrexed resistant lung cancer cells with suppression by vinca alkaloids. *Oncogene* **2017**, *36*, 242–253. [[CrossRef](#)]
31. Yang, Y.F.; Chang, Y.C.; Jan, Y.H.; Yang, C.J.; Huang, M.S.; Hsiao, M. Squalene synthase promotes the invasion of lung cancer cells via the osteopontin/ERK pathway. *Oncogenesis* **2020**, *9*, 78. [[CrossRef](#)] [[PubMed](#)]
32. Jones, G.G.; Del Rio, I.B.; Sari, S.; Sekerim, A.; Young, L.C.; Hartig, N.; Areso Zubiaur, I.; El-Bahrawy, M.A.; Hynds, R.E.; Lei, W.; et al. SHOC2 phosphatase-dependent RAF dimerization mediates resistance to MEK inhibition in RAS-mutant cancers. *Nat. Commun.* **2019**, *10*, 2532. [[CrossRef](#)]
33. Maacha, S.; Bhat, A.A.; Jimenez, L.; Raza, A.; Haris, M.; Uddin, S.; Grivel, J.C. Extracellular vesicles-mediated intercellular communication: Roles in the tumor microenvironment and anti-cancer drug resistance. *Mol. Cancer* **2019**, *18*, 55. [[CrossRef](#)] [[PubMed](#)]
34. Aung, T.; Chapuy, B.; Vogel, D.; Wenzel, D.; Oppermann, M.; Lahmann, M.; Weinlage, T.; Menck, K.; Hupfeld, T.; Koch, R.; et al. Exosomal evasion of humoral immunotherapy in aggressive B-cell lymphoma modulated by ATP-binding cassette transporter A3. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 15336–15341. [[CrossRef](#)] [[PubMed](#)]
35. Ostler, K.R.; Davis, E.M.; Payne, S.L.; Gosalia, B.B.; Exposito-Cespedes, J.; Le Beau, M.M.; Godley, L.A. Cancer cells express aberrant DNMT3B transcripts encoding truncated proteins. *Oncogene* **2007**, *26*, 5553–5563. [[CrossRef](#)] [[PubMed](#)]
36. Cheng, Y.; He, C.; Wang, M.; Ma, X.; Mo, F.; Yang, S.; Han, J.; Wei, X. Targeting epigenetic regulators for cancer therapy: Mechanisms and advances in clinical trials. *Signal Transduct. Target. Ther.* **2019**, *4*, 62. [[CrossRef](#)]
37. Sandhu, R.; Rivenbark, A.G.; Coleman, W.B. Enhancement of chemotherapeutic efficacy in hypermethylator breast cancer cells through targeted and pharmacologic inhibition of DNMT3b. *Breast Cancer Res. Treat.* **2012**, *131*, 385–399. [[CrossRef](#)]
38. Wu, C.; Guo, E.; Ming, J.; Sun, W.; Nie, X.; Sun, L.; Peng, S.; Luo, M.; Liu, D.; Zhang, L.; et al. Radiation-Induced DNMT3B Promotes Radioresistance in Nasopharyngeal Carcinoma through Methylation of p53 and p21. *Mol. Ther. Oncolytics* **2020**, *17*, 306–319. [[CrossRef](#)]
39. Yao, P.; Sun, H.; Xu, C.; Chen, T.; Zou, B.; Jiang, P.; Du, W. Evidence for a direct cross-talk between malic enzyme and the pentose phosphate pathway via structural interactions. *J. Biol. Chem.* **2017**, *292*, 17113–17120. [[CrossRef](#)]

40. Desgrosellier, J.S.; Cheresh, D.A. Integrins in cancer: Biological implications and therapeutic opportunities. *Nat. Rev. Cancer* **2010**, *10*, 9–22. [[CrossRef](#)]
41. Hamidi, H.; Pietila, M.; Ivaska, J. The complexity of integrins in cancer and new scopes for therapeutic targeting. *Br. J. Cancer* **2016**, *115*, 1017–1023. [[CrossRef](#)] [[PubMed](#)]
42. Domagala-Kulawik, J.; Guzman, J.; Costabel, U. Immune cells in bronchoalveolar lavage in peripheral lung cancer—Analysis of 140 cases. *Respir. Int. Rev. Thorac. Dis.* **2003**, *70*, 43–48. [[CrossRef](#)] [[PubMed](#)]
43. Bruno, T.C.; Ebner, P.J.; Moore, B.L.; Squalls, O.G.; Waugh, K.A.; Eruslanov, E.B.; Singhal, S.; Mitchell, J.D.; Franklin, W.A.; Merrick, D.T.; et al. Antigen-Presenting Intratumoral B Cells Affect CD4(+) TIL Phenotypes in Non-Small Cell Lung Cancer Patients. *Cancer Immunol. Res.* **2017**, *5*, 898–907. [[CrossRef](#)] [[PubMed](#)]
44. Carmi, Y.; Spitzer, M.H.; Linde, I.L.; Burt, B.M.; Prestwood, T.R.; Perlman, N.; Davidson, M.G.; Kenkel, J.A.; Segal, E.; Pusapati, G.V.; et al. Allogeneic IgG combined with dendritic cell stimuli induce antitumour T-cell immunity. *Nature* **2015**, *521*, 99–104. [[CrossRef](#)] [[PubMed](#)]
45. MacMahon, H.; Austin, J.H.; Gamsu, G.; Herold, C.J.; Jett, J.R.; Naidich, D.P.; Patz, E.F., Jr.; Swensen, S.J. Guidelines for management of small pulmonary nodules detected on CT scans: A statement from the Fleischner Society. *Radiology* **2005**, *237*, 395–400. [[CrossRef](#)] [[PubMed](#)]
46. Wisniewski, J.R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **2009**, *6*, 359–362. [[CrossRef](#)]
47. Carvalho, A.S.; Ribeiro, H.; Voabil, P.; Penque, D.; Jensen, O.N.; Molina, H.; Matthiesen, R. Global mass spectrometry and transcriptomics array based drug profiling provides novel insight into glucosamine induced endoplasmic reticulum stress. *Mol. Cell. Proteom. MCP* **2014**, *13*, 3294–3307. [[CrossRef](#)]
48. Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, *26*, 1367–1372. [[CrossRef](#)]
49. Matthiesen, R.; Prieto, G.; Amorim, A.; Aloria, K.; Fullaondo, A.; Carvalho, A.S.; Arizmendi, J.M. SIR: Deterministic protein inference from peptides assigned to MS data. *J. Proteom.* **2012**, *75*, 4176–4183. [[CrossRef](#)]
50. Hackenberg, M.; Matthiesen, R. Annotation-Modules: A tool for finding significant combinations of multisource annotations for gene lists. *Bioinformatics* **2008**, *24*, 1386–1393. [[CrossRef](#)]
51. Gu, Z.; Eils, R.; Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **2016**, *32*, 2847–2849. [[CrossRef](#)] [[PubMed](#)]
52. Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **2017**, *45*, D353–D361. [[CrossRef](#)] [[PubMed](#)]
53. Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **2019**, *47*, D330–D338. [[CrossRef](#)] [[PubMed](#)]
54. Goel, R.; Harsha, H.C.; Pandey, A.; Prasad, T.S. Human Protein Reference Database and Human Proteinpedia as resources for phosphoproteome analysis. *Mol. Biosyst.* **2012**, *8*, 453–463. [[CrossRef](#)] [[PubMed](#)]
55. Ruepp, A.; Brauner, B.; Dunger-Kaltenbach, I.; Frishman, G.; Montrone, C.; Stransky, M.; Waegle, B.; Schmidt, T.; Doudieu, O.N.; Stumpflen, V.; et al. CORUM: The comprehensive resource of mammalian protein complexes. *Nucleic Acids Res.* **2008**, *36*, D646–D650. [[CrossRef](#)]
56. Liberzon, A.; Birger, C.; Thorvaldsdottir, H.; Ghandi, M.; Mesirov, J.P.; Tamayo, P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **2015**, *1*, 417–425. [[CrossRef](#)]
57. Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D.J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; et al. The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **2018**, *47*, D442–D450. [[CrossRef](#)]

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).