Investigating the Protein Expression Pattern of Cancer Cells as a Tool for the Detection of Primary Tumor in Lung Cancer Patients

Dissertation

With the aim of achieving a doctoral degree (Dr. rer. nat.)

At the Faculty of Mathematics and Natural Sciences

Department of Chemistry

University of Hamburg

Submitted by

Lucija Ačkar

Hamburg 2023

Reviewers of the dissertation

Prof. Dr. rer. nat Hartmut Schlüter

University Medical Center Hamburg-Eppendorf

Center for Diagnostic

Institute of Clinical chemistry and Laboratory Medicine

Campus Forschung N27

Martinistraße 52

20246 Hamburg

Prof. Dr. med. Klaus Pantel

University Medical Center Hamburg-Eppendorf

Center for Experimental Medicine

Institute of Tumor Biology

Campus Forschung N27

Martinistraße 52

20246 Hamburg

Date of disputation: 26.05.2023

Research under the dissertation title "Investigating the protein expression pattern of cancer cells as a tool for the detection of primary tumor in lung cancer patients" was performed at the Institute of Tumor Biology at the University Medical Center Hamburg-Eppendorf, in the research group of Prof. Dr. med. Klaus Pantel and under supervision of Dr. Kai Bartkowiak, in the period between September 2018 and October 2022.

I.   List of publications

1. **Ačkar L.**, Casjens S., Andreas A., Raiko I., Brüning T., Geffken M., Paine S., Kollmeier J., Johnen G., Bartkowiak K., Webber DG., Pantel K.
    Blood-based detection of lung cancer using cysteine-rich angiogenic inducer 61 (CYR61) as a circulating protein biomarker: a pilot study. Mol Oncol. 2021 Nov;15(11):2877-2890.

2. Bartkowiak K., Casjens S., Andreas A., **Ačkar L.**, Joose SA., Raiko I., Brüning T., Geffken M., Paine S., Johnen G., Weber DG., Pantel K.
    Sensitive Blood-Based Detection of Asbestos-Associated Diseases Using Cysteine-Rich Angiogenic Inducer 61 as Circulating Protein Biomarker. Clin Chem. 2021 Jan 30;67(2):363-373.

3. Heidrich I.*, **Ačkar L***, Mossahebi Mohammadi P.*, Pantel K.
    *Equal contribution to all authors
    Liquid biopsies: potential and challenges. Int J Cancer. 2021 Feb 1;148(3):528-545.

II.   Table of contents

III. Abbreviations

ACN	Acetonitrile
ADAM	Disintegrin and metalloproteinase domain-containing protein
ADC	Adenocarcinoma
AKT	Protein kinase B
ALK	Anaplastic lymphoma kinase
ANOVA	Analysis of variance
APO	Apolipoprotein
AUC	Area under the curve
BCA	Bicinchoninic acid assay
BM	Bone marrow
BMP	Bone morphogenetic protein
BRAF	B-Raf proto-oncogene
BSA	Bovine serum albumin
CA	Carbohydrate antigen
CAD	Computer-aided detection system
CAF	Cancer associated fibroblast
CAIX	Hypoxia-regulated anhydrase IX
CCN1; CYR61	Cellular communication network factor 1
CD	Cluster of differentiation
CDKN2A	Cyclin-dependent kinase inhibitor 2A
CEA	Carcinoembryonic antigen
CF	Centrifugal filtration
cfmiRNA	Cell-free microRNA
CK	Cytokeratin
CLCA	Clathrin light chain 1
CLH1	Clathrin heavy chain 1
CTC	Circulating tumor cells
ctDNA	Cell-free tumor DNA
CV	Coefficient of variance
CYFRA21-1	Serum cytokeratin 19 fragment
DAPI	4´, 6´-diamidino-2-phenylindole
DGC	Density gradient centrifugation
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DPBS	Dulbecco´s Phosphate-Buffered Saline
DTC	Disseminated tumor cells
DTT	1;4-Dithiothreitol
EC	Endothelial cells
ECM	Extracellular matrix
EDTA	Ethylenediamine tetraacetic acid
EGFR	Epidermal growth factor receptor
ELISA	Enzyme-linked immunoassay
EMT	Epithelial to mesenchymal transition
EpCAM	Epithelial cell adhesion molecule
ERK	Extracellular signal-regulated kinase

ERM	Ezrin-radixin-moesin
ESI	Electrospray ionisation
EV	Extracellular vesicle
FACS	Fluorescence-activated cell sorting
FDA	Food and Drug Administration
FDR	False discovery rate
GPM6A	Neuronal membrane glycoprotein M6-a
HCD	High energy collision dissociation
HD or HC	Healthy donor or healthy control
HDL	High density lipoproteins
HER2	Human epidermal growth factor receptor 2
HIF-1α	Hypoxia inducible factor 1α
HPLC	High-performance liquid chromatography
HPLC	High-performance liquid chromatography
HSP	Heat shock protein
IAA	Iodoacetamide
Ig	Immunoglobulin
IPG	Immobilized and non-linear pH gradient
ITB	Institute of Tumor Biology
KAEP1	Kelch like ECH associated protein 1
K_i	Inhibitor constant
K_m	Michaelis-Menten Constant
KRAS	Kirsten rat sarcoma virus
LC	Lung cancer
LCANS	Lung cancer among never smokers
LCLC	Large cell lung cancer/carcinoma
LC-MS/MS	Liquid chromatography tandem mass spectrometry
LDCT	Low-dosed computed tomography
LDL	Low density lipoproteins
LFQ	Label-free quantification
M&M	Materials and Methods
MAPK	Mitogen-activated protein kinase
MMP	Matrix metalloproteinase
mNECT-4	Membranous NECTIN-4
MS	Mass spectrometry
MSN	Moesin
MVB	Multivesicular body
NECT/PVRL	NECTIN or Poliovirus receptor-related protein
NF1	Neurofibromatosis type 1
NK	Natural killer
NOS	Not otherwise specified
NSCLC	Non-small cell lung cancer
NTA	Nanoparticle tracking assay
NTRK	Neurotrophic tyrosine kinase
P	Passage (cell culture)
PAGE	Polyacrylamide gel electrophoresis

PB	Peripheral blood
PBMC	Peripheral blood mononuclear cell
PDAP1; HASPP28	Platelet-derived growth factor associated protein 1
PDGF	Platelet-Derived Growth Factor
PEG	Polyethylene glycol
PFA	Paraformaldehyde
PRSS1	Serine protease 1
PSA	Prostate specific antigen
PVDF	Polyvinylidene fluoride
RET	Ret proto-oncogene
REV1	DNA repair protein REV1
RNA	Ribonucleic acid
ROC	Receiver operating characteristic
ROS1	Proto-oncogene tyrosine-protein kinase ROS
RT-PCR	Reverse transcription polymerase chain reaction
SAA2	Serum amyloid A2
SCC	Squamous cell carcinoma
SCLC	Small cell lung cancer
SCLC-A	Neuroendocrine SCLC subtype the most associated with ASCL1
SCLC-N	Neuroendocrine SCLC subtype the most associated with NeuroD1
SCLC-P	Non-neuroendocrine SCLC subtype the most associate with POU2F3
SCLC-Y	Non-neuroendocrine SCLC subtype the most associated with YAPI
SCLY	Selenocysteine lyase
SDC	Sodium deoxycholate
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
SEC+CF	Size exclusion chromatography coupled with centrifugal filtration
SEC+TFF	Size exclusion chromatography coupled with tangential flow filtration
SEC+UC	Size exclusion chromatography coupled with ultracentrifugation
SILAC	Stable isotope labelling of amino acids in cell culture
SILAC$_{heavy}$	Corresponding cell culture medium with SILAC heavy arginine and lysine amino acids
SILAC$_{light}$	Corresponding cell culture medium with SILAC light arginine and lysine amino acids
SLC12A6; KCC3	Solute carrier family 12 member 6; potassium-chloride cotransporter 3
STK11	Serine/threonine kinase 11
TBA	α-TUBULIN
TEAB	Triethylammonium bicarbonate
TEM	Transmission electron microscopy
TEP	Tumor educated platelets
TME	Tumor microenvironment
TP53	Tumor protein 53
UC	Ultracentrifugation
UF	Ultrafiltration
UKE	University Medical Center Hamburg-Eppendorf
VCAM	Vascular cell adhesion molecule
VEGF	Vascular endothelial growth factor
WNT-1	Wingless/Int-1

1. Zusammenfasung

Lungenkrebs war im Vergleich zu anderen Krebserkrankungen die zweithäufigste neue Krebsursache im Jahr 2020 mit 2,26 Millionen neuen Fällen, während er mit 1,80 Millionen Todesfällen die häufigste Krebstodesursache war. Lungenkrebs ist eine stille Erkrankung, die erst in der fortgeschrittenen Phase Symptome zeigt, wenn die Metastasierung bereits aufgetreten ist. Daher werden Patienten am häufigsten in späten Phasen diagnostiziert, was zu einer hohen Mortalität von Lungenkrebsfällen führt. Die einzige Screening-Methode für Hochrisikogruppen für die Entwicklung von Lungenkrebs ist „*Low-dose computed tomography*" (LDCT), die an einer hohen falsch-positiven Rate leidet, was den dringenden Bedarf an geeigneten Instrumenten für das Screening und die Früherkennung von Lungenkrebs unterstreicht. Darüber hinaus folgt auf positive LDCT-Befunde in der Regel eine konventionelle Biopsie, die aufgrund ihrer Invasivität zu verfahrensbedingten unerwünschten Ereignissen führen kann. Das minimalinvasive Verfahren „*liquid biopsy*" (Flüssigbiopsie) verwendet Analyten aus Körperflüssigkeiten, z.B. Blut, zerebrospinale Flüssigkeit, Urin oder sogar Tränen. Die am häufigsten verwendeten Analyten in der Flüssigbiopsie sind zirkulierende Tumorzellen (engl. circulating tumor cells, CTCs), zellfreie DNA (engl. cell-free DNA, cfDNA), extrazelluläre Vesikel (engl. extracellular vesicles, Evs) und sekretierte Proteine.

Die Isolierung von CTCs basiert derzeit hauptsächlich auf dem EPCAM-Protein, das ein epithelialer Biomarker ist oder auf Größenunterschieden zwischen CTCs und PBMCs. Interessanterweise ist die Erkennungsrate von CTCs, die von Lungenkrebspatienten stammen, bei Verwendung einer der oben genannten Techniken eher niedrig. Die Erkennung von Lungenkrebs-CTCs könnte durch mesenchymale Biomarker verbessert werden und möglicherweise die Erkennung von Lungenkrebs verbessern. Des Weiteren wurden extrazelluläre Vesikel, die von Tumorzellen und gesunden Zellen abgesondert werden, mit der Zell-zu-Zell Kommunikation in Verbindung gebracht. Deshalb könnte ihre Proteinzusammensetzung auch neue Biomarker für die Lungenkrebsvorsorge und die Erkennung aufdecken. Der dritte Analyt von Interesse in dieser Arbeit sind sezernierte Proteine, die, wenn sie differentiell exprimiert werden, beim Screening von Lungenkrebs-Risikogruppen verwendet werden könnten, um auch die Lungenkrebserkennung zu verbessern.

Ziel dieser Arbeit war es, neue aus der Flüssigbiopsie gewonnene Analysen zur Verbesserung der Lungenkrebsvorsorge und -erkennung bereitzustellen. Um dies zu erreichen, wurde eine SILAC-basierte Massenspektrometrie-Analyse zwischen epithelialem H1975 und mesenchymaler LC-M1-Lungenkrebszelllinie zum Nachweis potenzieller CTC-Marker und serologischer Marker zur Erkennung von Lungenkrebs verwendet. Das Zelloberflächenprotein NECT-4 und die sezernierten Proteine CCN1 und PDAP1 wurden als potenzielle Kandidaten für die Lungenkrebserkennung entdeckt. Da die extrazelluläre Domäne von NECT-4 abgespalten und freigesetzt werden kann, wurden NECT-4 und PDAP1 als potenzielle serologische Biomarker für die Lungenkrebserkennung mittels Western Blot und ELISA analysiert. NECT-4 ist in Epithelzelllinien im Vergleich zu mesenchymalen Zelllinien erhöht,

während die ROC-Analyse des ELISA-basierten Proteinnachweises in der Studienkohorte die AUC von 0,7627 lieferte, mit der Sensitivität von 43,52% und der Spezifität von 95,00% für die Unterscheidung von Lungenkrebspatienten der gesunden Kontrollen. Bei dem Adenokarzinom lieferten die Analysen eine AUC (*area under the curve*, Fläche unter der Kurve) von 0,8267 mit einer Sensitivität von 56,79% bei 95,00% Spezifität. PDAP1 wird sowohl in epithelialen als auch in mesenchymalen Zelllinien exprimiert, wobei die Spiegel in Epithelzellen im Vergleich zu den mesenchymalen Zellen erhöht sind. Die ROC-Analyse des ELISA-basierten Nachweises von PDAP1 in der Studienkohorte ergab eine AUC von 0,7020 mit einer Sensitivität von 23,32% bei einer Spezifität von 95,00%. Die Stratifizierung nach den histologischen Subtypen des Lungenkrebses ergab keine signifikante Verbesserung der Sensitivität und Spezifität der verwendeten PDAP1-Untersuchung. CCN1 wird sowohl in epithelialen als auch in mesenchymalen Zelllinien exprimiert, jedoch ist CCN1 im Gegensatz zum PDAP1 in mesenchymalen Zelllinien erhöht. Daher wurde CCN1 mittels Immunhistochemie analysiert, um die Erkennung von Lungenkrebs durch CTCs zu verbessern. CCN1 wurde in der Lungenkrebs-Zelllinie HCC-366 nachgewiesen, jedoch nicht in mononukleären Zellen des peripheren Blutes von gesunden Spendern. PBMC. Weiterhin wurden CTCs und DTCs von Lungenkrebspatienten und postmortalen Lungenkrebspatienten isoliert. Unter Verwendung von Pan-Keratin-Antikörpern wurden keine CTCs nachgewiesen, während eine Zelle in einem Patienten unter Verwendung von CCN1-Antikörpern nachgewiesen wurde. DTCs wurden bei zwei analysierten Patienten unter Verwendung von Pan-Keratin-Antikörpern nachgewiesen, und kein DTC wurde von CCN1 nachgewiesen.

Da extrazelluläre Vesikel (EV) ihr Potenzial als Quelle für Krebsbiomarker gezeigt haben, wurde die Isolierung von EVs aus dem Plasma optimiert und eine vergleichende proteomische Analyse zwischen Adenokarzinompatienten und gesunden Kontrollen mittels markierungsfreier quantifizierungsbasierter Massenspektrometrie durchgeführt. Die Analyse der proteomischen Daten zeigte unterschiedliche EV-Subproteome zwischen Adenokarzinom-Patienten und gesunden Kontrollpersonen. Hierarchisches Clustering zeigte, dass das Proteinexpressionsprofil von gesunden Rauchern teilweise mit dem von Lungenkrebspatienten überlappt, was darauf hindeutet, dass das EV-Subproteom möglicherweise einen Übergangszustand in der Lungenkrebsentwicklung aufweisen könnte. Weitere Analysen zeigten SAA2-, REV1-, SCLY-, SLC12A6-, PRSS1-, GPM6A- und MSN-Proteine als EV-Biomarker-Panel für den Nachweis von Adenokarzinom-Patienten. Dieses EV-Biomarker-Panel muss jedoch weiter validiert werden. Schließlich zeigte die auf Genontologie basierende Signalweganalyse Unterschiede in den biologischen Rollen zwischen Lungenkrebs und gesunden Kontroll-EVs. Signalwege von EVs aus gesunden Kontrollen sind am meisten mit Stoffwechsel, Lokalisation und Biogenese assoziiert, und ihre Proteine sind am stärksten mit vesikulären Komponenten und mit der enzymatischen Funktion verbunden. Proteine, die in EVs nachgewiesen wurden, die von Adenokarzinom-Lungenkrebspatienten stammen, waren am stärksten mit der Reaktion des Immunsystems assoziiert. Die Lokalisation von Adenokarzinom-Lungenkrebs-abgeleiteten-EV-Proteinen ist am stärksten mit zytoplasmatischen und extrazellulären Regionen sowie mit Lipidkomplexen assoziiert, während sie funktionell am stärksten

mit der Bindung, enzymatischen Regulation und mit der Aktivierung der Phosphatidylcholin-Sterol-O-Acyltransferase-Aktivität assoziiert sind.

2. Abstract

Lung cancer compared to other cancer entities was the second most common new cause of cancer and the leading cause of cancer deaths in 2020, accounting for 2.26 millions of new cases and 1.80 millions of deaths. Lung cancer is a silent disease which shows symptoms at advanced stage when the metastasis has already occurred. Therefore, patients are most often diagnosed at late stages resulting in the high mortality of lung cancer cases. The only screening method for high-risk group for lung cancer development is Low-dose computed tomography (LDCT). However, LDCT suffers from high false-positive rate, emphasizing the urgent need of appropriate tools for screening and for the early detection of lung cancer. Moreover, positive LDCT findings are usually followed by conventional tissue biopsy, which due to its invasiveness can lead to procedure-related adverse events. The field of *liquid biopsy* utilizes analytes from body fluids, e.g., blood, cerebrospinal fluid, urine, or even tears. Liquid biopsy is minimally invasive and allows the heterogenous analysis of the sample, contrary to the conventional tissue biopsy. The most common analytes used in liquid biopsy are circulating tumor cells (CTCs), cell-free DNA (cfDNA), extracellular vesicles (EVs), and secreted proteins. Isolation of CTCs is currently mainly based on EPCAM epithelial biomarker, or on differences in size between CTCs and PBMCs. Interestingly, detection rate of CTCs derived from lung cancer patient is rather low when using either of the above-mentioned techniques. The detection of lung cancer-CTCs might be improved using mesenchymal biomarkers and potentially improve the detection of lung cancer. Furthermore, extracellular vesicles secreted by tumor cells and normal cells have been implicated in cell-to-cell communication. Therefore, their protein composition could also reveal new biomarkers for lung cancer screening and detection. The third analyte of interest in this thesis are secreted proteins, which if being differentially expressed could be used in screening of lung cancer risk groups, improving lung cancer detection as well.

The aim of this thesis was to provide new liquid biopsy-derived analytes for the improvement of lung cancer screening and detection. To reach that, SILAC-based mass spectrometry analysis between epithelial H1975 and mesenchymal LC-M1 lung cancer cell line was used for detection of potential CTC markers and serological markers for the lung cancer detection. The cell surface protein NECT-4, and the secreted proteins CCN1 and PDAP1 have been detected as potential candidates to be utilized in lung cancer detection. Since the extracellular domain of NECT-4 can be cleaved and released, NECT-4 and PDAP1 have been analysed as potential serological biomarkers for lung cancer detection using *Western blot* and ELISA. NECT-4 signal intensities were elevated in epithelial cell lines compared to mesenchymal cell lines. The ROC analysis of ELISA-based NECT-4 detection on the study cohort provided the AUC of 0.7627, with a sensitivity of 43.52% and specificity of 95.00% for the discrimination of lung cancer patients from healthy controls. The utility of NECT-4 is the most promising for the detection of adenocarcinoma lung cancer subtype, with ROC analysis providing the AUC 0.8267, with a sensitivity of 56.79% at 95.00% specificity. PDAP1 is expressed in both, epithelial

and mesenchymal cell lines, with elevated levels in epithelial cells compared to the mesenchymal cells. The ROC analysis of ELISA-based PDAP1 detection in study cohort provided an AUC of 0.7020, with a sensitivity of 23.32% at a specificity of 95.00%. Stratification according to the lung cancer histological subtypes did not reveal significant improvement in sensitivity and specificity of the PDAP1 assay used. CCN1 is as well expressed in both, epithelial and mesenchymal cell lines, however, in contrary to PDAP1, CCN1 is elevated in mesenchymal cell lines. Therefore, CCN1 has been considered as potential biomarker for the improvement of lung cancer derived CTCs detection. CCN1 has been detected in the lung cancer cell lines by *Western blot* analysis, whereas no signals were detected in peripheral blood mononuclear cells from healthy donors. Due to the challenges in patient samples collection caused by the start of COVID-19 pandemics, the analysis of CCN1 as potential CTC biomarker has been terminated at this stage.

Since EVs have shown potential as source for cancer biomarkers, isolation of EVs from plasma has been optimized in this thesis. Size exclusion chromatography coupled with ultracentrifugation (SEC+UC) provided the highest yield and purity of EVs compared to other methods analysed. Further on, EVs from adenocarcinoma lung cancer patients and healthy controls have been isolated using SEC+UC. Comparative proteomic analysis between adenocarcinoma patients and healthy control EVs using label-free quantification-based liquid chromatography tandem mass spectrometry has been performed. The analysis of the proteomic data showed different EV subproteome between adenocarcinoma patients and healthy controls. The detected protein profile of healthy control smokers partially groups together with lung cancer patients based on hierarchical clustering results, suggesting that such an EV subproteome could potentially display a transition state in lung cancer development. Further analysis indicated SAA2, REV1, SCLY, SLC12A6, PRSS1, GPM6A, and MSN proteins as a potential EV biomarker panel for the detection of adenocarcinoma patients. However, suggested EV biomarker panel needs to be further validated. Finally, the gene ontology-based pathway analysis showed difference in the biological roles between lung cancer patient and healthy control EVs. Healthy control derived EVs are the most associated with metabolism, localization, and biogenesis, and their proteins are the most associated with vesicular components and with enzymatic function. Proteins detected on EVs deriving from adenocarcinoma lung cancer patients were the most associated with immune system response. Localization of adenocarcinoma-lung-cancer-derived-EV-proteins is the most associated with cytoplasmic and extracellular regions, as well as with complexes associated with lipids, whereas functionally they are the most associated with binding, enzymatic regulation, and with activation of phosphatidylcholine-sterol O-acyltransferase activity.

3. Introduction

3.1 Cancer

Cancer is a group of diseases characterized by uncontrolled growth and spread of cells. Cancer develops from normal cells by comprising multiple changes on a molecular level resulting in deregulation of molecular pathways essential for cellular growth, proliferation, and development (1). Currently, cancer is the leading cause of death worldwide, causing nearly 10 million deaths in 2020 (2). The most common new cases of cancer in 2020 were breast cancer and lung cancer, counting for 2.26 million and 2.21 million cases, while the most common causes of cancer deaths in 2020 were lung cancer and colorectal cancer, accounting for 1.80 million and 960,000 deaths (Figure 1) (2).
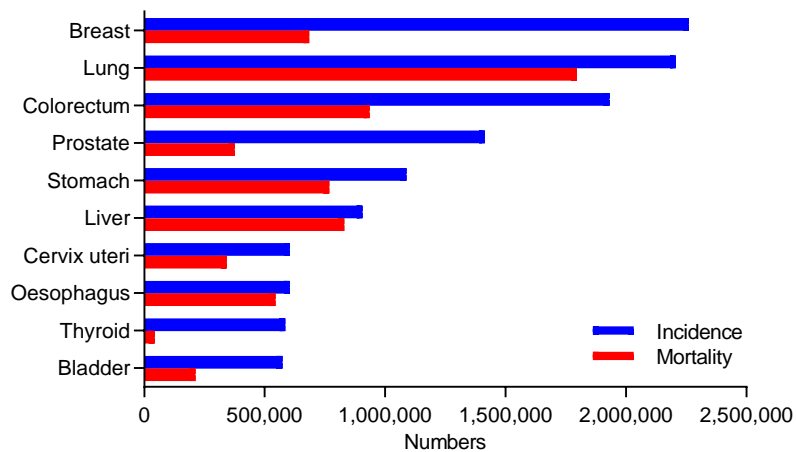
Figure 1. Estimated number of incident cases and deaths in 2020 by GLOBOCAN. Blue-incidence of cancer cases in 2020 by GLOBOCAN shows the highest incidence for breast and lung cancer. Red-estimated number of new cancer deaths in 2020 by GLOBOCAN shows the highest mortality for lung cancer, having almost double mortality rate compared to the second most mortal cancer, colorectal cancer. Data for creating graph were taken from World Health Organization, GLOBOCAN(3).

Initially, Douglas Hanahan and Robert Weinberg were the first ones who described the first 6 key traits of cancer that are acquired during tumor development, naming them hallmarks of cancer (4). In contrary to normal cells, cancer cells do not need normal growth signalling, but they are capable of mimicking growth signals and therefore promote growth on their own. Additionally, they are insensitive to antiproliferative signals such as soluble growth inhibitors or immobilized inhibitors embedded in the extracellular matrix (ECM) and on the surface of nearby cells. Upon cellular excess or cellular damages, normal cells go through precisely regulated biological steps and undergo programmed cell death-apoptosis, which cancer cells have ability to evade. Normally, cells have limited replicative potential (5), which is acquired by shortening of telomeres for 50-100 bp from every chromosome during each cell cycle (called replication). However, cancer cells developed mechanisms to prolong telomeres and to maintain immortality (6). All previously mentioned traits lead to the growth of tumors which results

in tumor cells being distant to blood vessels and nutrient supply. To progress further, cancer cells develop angiogenic ability (7). The deadliest hallmark of cancer cells is their ability to invade adjacent tissues and travel to distant sites, where they form new colonies, which are called metastases (4). Later, Hanahan and Weinberg included several new traits as cancer hallmark (8); the ability of cancer to evade immune system of a host and the ability of cancer cells to alter their energy metabolism, limiting it to glycolysis even in the presence of oxygen, named aerobic glycolysis (8). Additionally, genomic instability and tumor inflammation promoted tumorigenesis has been added as cancer hallmark (Figure 2a) (8, 9).

Recently, D. Hanahan has suggested additional cancer hallmarks (Figure 2b); phenotypic plasticity, epigenetic reprograming, polymorphic microbiomes and reversable senescence (1, 10). Cancer cells are able to regain phenotypic plasticity by dedifferentiation (11, 12) of terminally differentiated cells into pluripotent-cell-state. When being in pluripotent-cell-state, cancer cells can partially differentiate maintaining progenitor-like-state which continues to proliferate, or they can transdifferentiate into new cell type and acquire other tissue specific traits (1, 10, 13). Also, epigenetic modifications can silent or enhance gene expression, including oncogenes and tumor suppressor genes (1, 10, 14, 15). They can contribute to the metastasis by inducing epithelial-to-mesenchymal transition (EMT) (1, 16). Additionally, microbiome and its role in immunoediting and escape from the immune system, has been suggested as a hallmark of cancer (1, 10, 17-19). Further on, when exposed to the microenvironmental stress, normal cells undergo senescence as a defensive mechanism against malignancy (20, 21). On the other hand, cancer senescent cells have been found to contribute to the proliferative signalling, evade of apoptosis, induction of angiogenesis, stimulating metastasis and suppressing tumor immunity (21-25). Additionally, for normal cells senescence is an irreversible state, while cancer senescent cells can resume cell proliferation and become fully oncogenic cells again (10, 26).
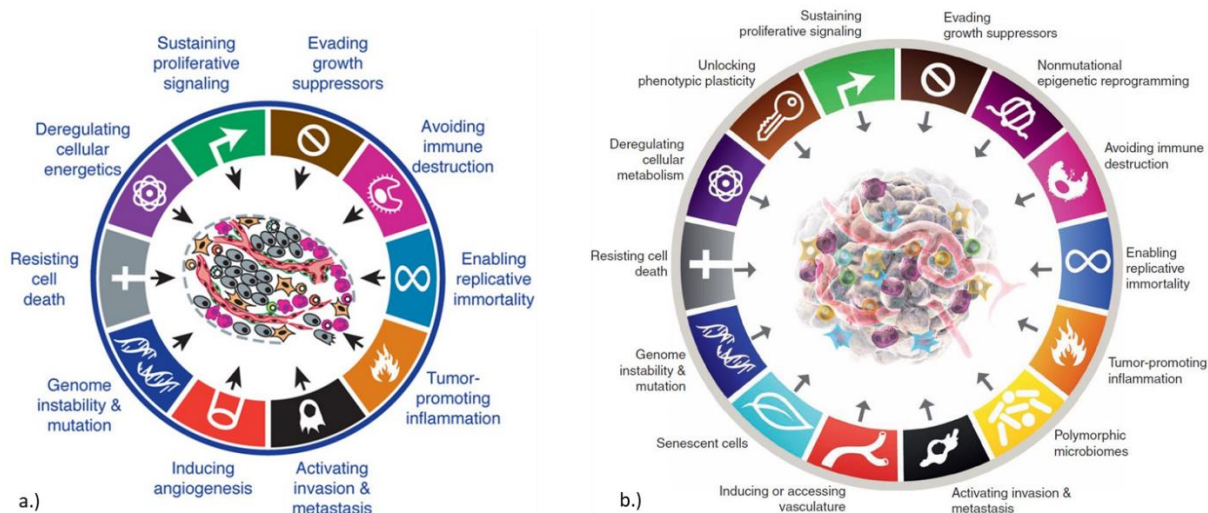
Figure 2. Hallmarks of cancer. a.) Ten hallmarks of cancer suggested by Hanahan and Weinberg in 2000 and later in 2011. b.) Five additional hallmarks of cancer suggested by Hanahan and Weinberg in 2022. Figure 2a was taken from Hanahan and Weinberg and modified (8), and figure 2b was taken from D. Hanahan (10).

Finally, the density of nerve fibres in tumor tissue is shown to be related with the tumor aggressiveness, by promoting tumorigenesis and modulating tumor environment (27-30). Therefore, Senga and Grose suggested the alteration of neuronal signalling as an additional hallmark of cancer (1).

Eventually, genomic instability together with epigenetic modifications mentioned above, yield in a variety of tumor subclones (31, 32), resulting in tumor heterogeneity, whether between patients (intertumor heterogeneity) or even within the same patient (intratumor heterogeneity). Intratumor heterogeneity is related to cancer progression, resistance to therapy and recurrence (9, 33).

3.2 Lung Cancer

Lung cancer is the second most common cancer and the leading cause of cancer deaths, often asymptomatic until late stages when it is usually diagnosed (34, 35) (Figure 1). It has resulted in more deaths than breast cancer, prostate cancer, colorectal cancer, and leukaemia combined in men ≥40 years old and women ≥60 years old (36). Diagnosis of lung cancer at early stages is strongly associated with improved survival compared with later stages. However, majority of lung cancer cases are diagnosed in symptomatic individuals, which are usually late-stage diagnosis (34, 36, 37).

Tobacco use is the greatest risk factor for developing lung cancer, including second hand smoking as well (38). In contrary of being the second most common cancer and the leading cause of cancer death, if lung cancer among never smokers (LCANS) would be considered as an individual entity, LCANS is the 11th most common cancer and 7th leading cause of deaths (39). Other risk factors for developing lung

cancer include asbestos exposure, family history of lung cancer, respiratory disease, exposure to toxic substances, heavy metals, and radon gas (34, 36).

3.2.1 Lung cancer histology

There are two major histological subtypes of lung cancer, non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC), from which SCLC is considered as more aggressive histological subtype with 5-year survival of 6%, compared to 23% for NSCLC (39). NSCLC accounts for approximately 85% of all diagnosed cancers where the most common histological subtype is adenocarcinoma (ADC, approx. 40% of diagnosis) (40), followed by squamous cell carcinoma (SCC, approx. 25% of diagnosis), large cell carcinoma (LCC, approx. 15% of diagnosis) and others and not otherwise specified (NOS, approx. 20%) (Figure 3) (39).
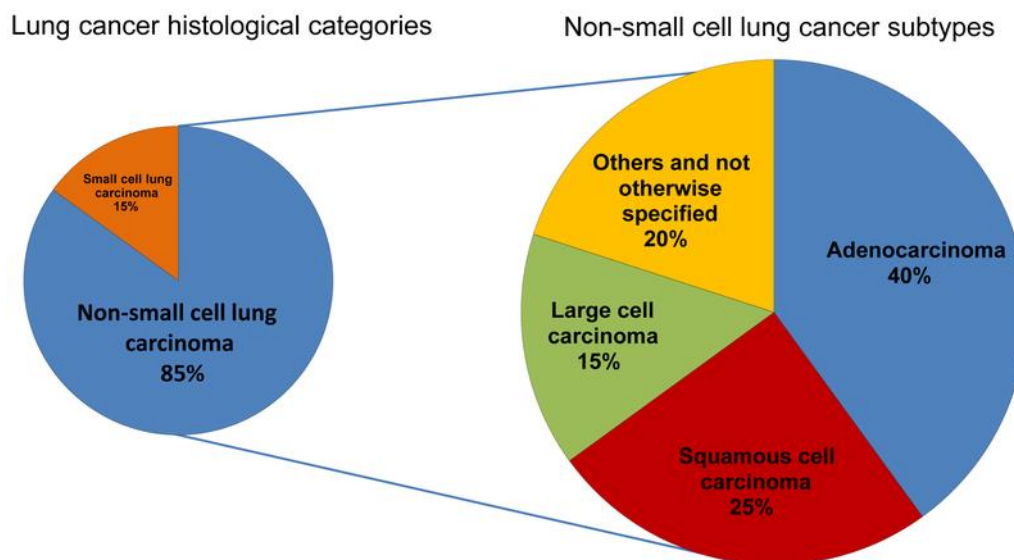
Figure 3. Histological classification of lung cancer. Lung cancer is classified into two major histological subtypes, non-small cell lung cancer and small cell lung cancer. Non-small cell lung cancer is further classified into adenocarcinoma, squamous cell carcinoma, large cell carcinoma, and others and not otherwise specified. Figure was taken from Schabath and Cote, Cancer Epidemiol Biomarker Prev. 2019 (39).

In general, driving mutations for NSCLC include the epidermal growth factor receptor (*EGFR*), Kirsten rat sarcoma virus (*KRAS*), proto-oncogene tyrosine-protein kinase ROS (*ROS1*), and anaplastic lymphoma kinase (*ALK*) (41). The most common mutations in ADCs include *EGFR* and *KRAS* genes, and tumor suppressor genes tumor protein p53 (*TP53*), Kelch like ECH associated protein 1 (*KAEP1*), serine/threonine kinase 11 (*STK11*) and neurofibromatosis type 1 (*NF1*) (41). In SCCs *TP53* is mutated in over 90% of tumors, and cyclin-dependent kinase inhibitor 2A (*CDKN2A*) is epigenetically silenced

in over 70% of tumors (41). Other mutations common in NSCLC include B-Raf proto-oncogene, serine/threonine kinase (*BRAF*) *V600E*, *MET* exon 14 skipping mutation, human epidermal growth factor receptor 2 (*HER2*) gene, rearrangements in the ret proto-oncogene (*RET*), and fusions of the neurotrophic tyrosine receptor kinase (*NTRK*) genes 1, 2, and 3 (41).

SCLC accounts for approx. 15% of all diagnosed lung cancers, among which patients are often heavy smokers (40). It is characterized by rapid growth, early metastasis, and acquired therapeutic resistance (42, 43). Tumor cells often express neuroendocrine markers, and currently SCLC is clinically considered as a single histological type. However, molecular subtypes of small cell lung cancer based on differential expression of transcription factors have been recently suggested. SCLC-A, SCLC-N have been suggested as neuroendocrine SCLC molecular subtypes, and SCLC-Y and SCLC-P as non-neuroendocrine SCLC molecular subtypes The last letter A, N, Y or P signifies the transcription regulator most strongly associated with each subtype (ASCL1, NeuroD1, YAP1, and POU2F3) among which SCLC-A has the greatest proportion, followed by SCLC-N, SCLC-Y, and SCLC-P (42, 43).

Lung cancer among never smokers is histologically most likely to be adenocarcinoma. Molecularly, the genome of lung cancer among never smokers is significantly different than the one arising in smokers – *TP53*, *KRAS* and *STK11* mutations are more common in smokers with lung cancer, while *EGFR* and *HER2* mutations, and *ALK-ELM4* fusions are more common in lung cancer among never smokers (39).

3.2.2 Lung cancer screening

Currently, the only screening method used for detection of lung cancer is low-dosed computed tomography (LDCT). The use of LDCT reduced mortality of lung cancer by 20%; however, there is still high rate of overdiagnosis and false positives (34, 36, 37), which leads to subsequent invasive procedures and procedure related adverse events (44). Many LDCT-based screening programmes have been conducted or are currently in progress and the use of appropriate multivariable risk prediction models significantly affects sensitivity and specificity of LDCT (37). Therefore, the identification of clinically significant lung cancer, while preventing overdiagnosis and false positive results is a main challenge in lung cancer diagnosis. Computer-aided detection system (CAD) could help overcoming this issue as a second reader in LDCT-based lung cancer screening (37). On the other hand, none of the validated prediction models includes biomarkers as a variable, even though extent research has been made to discover the potential of biomarkers in cancer diagnosis, including lung cancer (37). Incorporation of biomarkers into prediction models could reduce overdiagnosis and false positive rates given by LDCT. Additionally, the development and progress of cancer detection using liquid biopsy could decrease the occurrence of procedure-related adverse events, since it is minimally invasive.

3.3. Metastasis

Metastasis is the deadliest step in tumor progression, and it accounts for 90% cancer-related deaths in patient with palpable clinical traits (45). It was shown that it can occur at both early and late tumorigenesis, however with distinct pathogenesis. In early metastasis, cancer cells often harbour only truncal mutations, while in late metastasis, cancer cells harbour subclonal mutations (46). Metastasis is an organotrophic process comprised of several steps, mainly classified into invasion and intravasation, survival, extravasation, and colonization (47).

Metastasis starts with the invasion of tumor cells at tumor-stromal interface of tumor (also called invasive front), during which either single-cell migration (monoclonal metastasis) or cluster-based collective invasion (polyclonal metastasis) can occur (47, 48). Collective invasion is less common but more efficient in metastasis compared to the single-cell migration (47). In order to invade, tumor cells need to adjust their phenotype, therefore a subpopulation of cells go through the EMT (49). EMT is one of the mechanisms in promotion of metastasis, where some cells can take complete EMT, which is more common for single cells, and other cells can undergo partial (hybrid) EMT, which is more common for clusters (47). Additionally, for tumor cell migration and subsequent intravasation into blood vessels, rearrangements in the extracellular matrix (ECM) are needed. Cancer cells are being segregated from the nearby stroma by physical forces exerted by cancer associated fibroblasts (CAFs) and allow cellular invasion (50). Further on, mast cells, neutrophils, and macrophages secrete proteases (such as matrix metalloproteinase (MMP) 2 and MMP9, facilitating the remodelling of ECM (45, 51, 52). Invadopodia are tumor cell extensions rich in actin, effective for passing of tumor cell through the ECM (47). Once tumor cells are detached, they are called disseminated tumor cells (DTCs), and upon detachment these cells are under risk of apoptosis known as anoikis (programmed cell death due to the loss of integrin connections). However, tumor cells are able to avoid anoikis, for example through the expression of hypoxia inducible factor (HIF)-1α independent of hypoxia (47, 53), and intravasate blood or lymphatic vessel through the abnormal leaky vasculature instead (47). At that stage tumor cells are called circulating tumor cells (CTCs). Once tumor cells extravagate vessels at distant site, they are called DTCs again (47).

CTCs have a half-life of about 2.4 h (54), and within circulation they are exposed to multiple environmental stressors, such as oxidative stress, shear forces, and immune system (47). To avoid elimination by immune system, CTCs interact with CAFs, neutrophils and platelets. These interactions protect them from anti-tumor immune cells, such as NK cells (55). Besides protection from immune system, platelets surround CTCs in circulation to protect them from the shear stress (47), since CTC survival in circulation is affected by vascular size, flow rates and shear stress as well. Therefore, CTCs are more likely to undergo intravascular arrest and extravasate in venous vessels since the flow rate is rather moderate and shear stress is lower compared to arterial vessels (56). In order to avoid oxidative

stress due to sudden huge amount of oxygen in circulation, CTCs are able to induce reversible metabolic changes (45).

CTC extravasation is facilitated by the release of cytokines into the circulation. Platelets promote coagulation and weaken endothelial barrier (57). Additionally, tumor microenvironment (TME) in distant organs also send signals, such as vascular endothelial growth factor (VEGF), promoting CTC extravasation (58). Further on, CTCs through CAM interactions adhere to the endothelial cells (ECs) to subsequently interact with secondary site and extravagate (47, 59). Together with CTC-EC interactions, CTCs are forming invadopodia to extend in between ECs (60), which eventually leads to rupturing of the blood vessels (61).

For final cellular seeding at the secondary site and tumor colonization, DTCs interact with the ECM, and this adhesion is mediated through integrin family (62). The microenvironment at the secondary site can be either dormancy-permissive microenvironment, for example due to bone morphogenetic protein (BMP) 4, BMP7, transforming growth factor (TGF)-β2, and other factors (63, 64), or dormancy restrictive, for example if vascular cell adhesion molecule (VCAM) 1 is upregulated (65). If initially destined for dormancy, actin assembly has been found critical for promotion of a dormancy-to-proliferation switch in DTCs (47). At distant site(s), tumor cells grow and proliferate, forming secondary tumor mass.

As described above in this section, even though metastasis is generally classified in main 4 steps, each step involves plethora of interactions, and co-activity of cancer cells and host cells, showing the complexity of this process. Metastasis is an extremely demanding process for cancer cells, and at each step majority of cancer cells die, resulting in only 0.01% of cancer cells surviving and forming secondary tumors(47).

3.4 Liquid biopsy

Liquid biopsy is a minimally invasive procedure which was introduced more than a decade ago including the analysis of CTCs in the blood of cancer patients (66, 67). It is usually based on blood sampling, but urine, cerebrospinal fluid, sputum, and other body fluids can be analysed as well. Up until today, the field has been extended from the analysis of CTCs to analysis of cell-free tumor DNA (ctDNA), mRNA, cell-free microRNA (cfmiRNA), long noncoding RNA, small RNA, tumor-educated platelets (TEP), extracellular vesicles (EVs), and circulating cell-free proteins (68-70), as well as immune cells, circulating endothelial cells and cancer associated fibroblasts (Figure 4) (71).
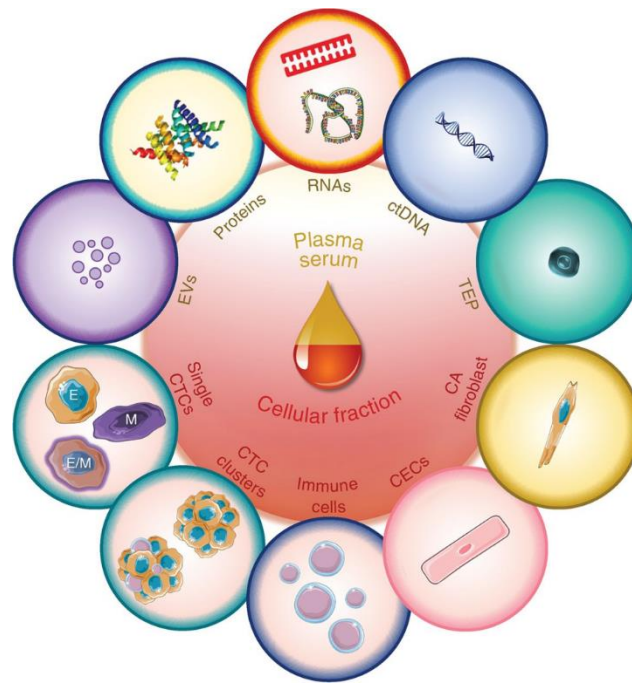
Figure 4. Circulating analytes utilized in liquid biopsy for precision medicine. The blood represents a blood sample from which plasma or serum fraction (yellow, top) and cellular fraction (red, bottom) are yielded after blood processing. EVs, proteins, RNAs, ctDNA, and TEP can be utilized as liquid biopsy biomarkers obtained from plasma or serum fraction of blood. Single CTCs, cluster CTCs, immune cells, CECs, CA fibroblasts can be utilized as liquid biopsy biomarkers and obtained from cellular fraction of blood. CA=cancer associated; CEC=circulating endothelial cells; CTC=circulating tumor cell; DNA=deoxyribonucleic acid; EV=extracellular vesicles; RNA=ribonucleic acid; TEP=tumor educated platelets. Figure was taken from Alix-Panabiere and Pantel, Cancer Discov, 2021 (71).

Tissue biopsy is still a conventional method for cancer detection. However, it is an invasive method, often related to procedure related adverse events (44). Other limitations include localized sampling of a tissue which does not accurately represent heterogeneity of the cancer, and insufficient amount or poor quality of a tissue (72). Except being minimally invasive, the advantage of liquid biopsy over tissue biopsy is laying in a possibility to overcome intratumor heterogeneity (9). Additionally, it is possible to achieve a sample even when a tumor tissue is unreachable or patient is in poor healthy state, still allowing molecular analysis of tumor. Further on, liquid biopsy allows earlier detection of minimal residual disease and relapse of the disease compared to the use of conventional methods (73). Besides detection of minimal residual disease and relapse of disease, liquid biopsy can be applied for detection and diagnosis of cancer (74, 75), including at early stages (76, 77). Further on, it can be applied for identification of therapeutic targets and mechanisms of resistance (71), as well as in therapy decision making (78) and therapy monitoring (79).

Despite big potential and progress of liquid biopsy in cancer research and in health care system, there are still many challenges remaining. For instance, CTCs, ctDNAs, cell-free tumor RNAs (ctRNAs) TEPs, and tumor derived EVs are in a very low abundance compared to the circulating analytes originating from healthy tissues. Currently, the only technology approved by Food and Drug

Administration (FDA) for CTC isolation is CellSearch, which is based on epithelial cell adhesion molecule (EpCAM) detection (77, 80). For the analysis of gene mutations in ctDNA, technologies approved by the FDA are Guardant360 (81) and FoundationOne Liquid CDx (82). For NSCLC patients for whom tissue biopsy is not available, ctDNA analysis for *EGFR* mutations has been approved by the FDA as well (83). Additionally, for better implementation of liquid biopsy in clinical setting, the standardization of analytical assay is necessary.

Therefore, there is an urgent need for development of technological platforms for more sensitive isolation of tumor specific analytes, but also for discovery of novel biomarkers of tumor specific analytes, which would facilitate the development of technological platforms. Additionally, to overcome current technological limitations, detection of circulating free proteins as potential biomarker for cancer detection or monitoring is a possibility.

3.5 Enrichment of extracellular vesicles (EVs)

Extracellular vesicles (EVs) are lipid bilayer particles composed of different lipids and proteins and they engulf DNA, RNAs, lipids and proteins as cargo (84). EVs cover wide range of sizes and are divided into two main subcategories based on their differences in biogenesis, ectosome and exosomes. Ectosomes are vesicles produced by direct outward budding of the plasma membrane in a size range from 50 nm to 1 µm in diameter (85). Exosomes are vesicles produced by the endosomal pathway through the formation of multivesicular bodies (MVBs) in a size range from 40 nm to 160 nm, with the average size of ~100 nm (85). EVs are released from cells under normal physiological or pathophysiological conditions and, among other roles, EVs serve in intercellular communication, e.g., in cancer EV-mediated communication can affect metastasis, invasion, angiogenesis, immune modulation and cell survival (86).

Isolation of EVs and especially of specific subcategories of EVs like tumor-associated or host-derived EVs remains challenging due to high heterogeneity in sizes, surface biomarkers or lipid membrane composition of EVs. Therefore, each type of isolation approach has advantages and disadvantages, where some examples are described below in this section.

3.5.1 Ultracentrifugation

Ultracentrifugation (UC) is still considered as a gold standard. Together with the commercial reagents, UC gives the highest yield of EVs compared to other isolation techniques (87). However, it is more likely that EVs isolated by UC are highly contaminated with proteins which co-precipitate with the EVs and contribute to the effect as the yield of EVs is high. Protein contamination level could be affected by the choice of rotor type, where protein contamination after the use of fixed rotor is lower compared to

the protein contamination level after the use of swing-bucket rotor type (87). When it comes to plasma samples, an additional severe contamination problem are lipoproteins due to their similarity in sizes and composition with EVs (84). Compared to polyethylene glycol (PEG) or size exclusion chromatography (SEC) based isolation methods, UC leads to a stronger reduction in lipoprotein contamination (84).

3.5.2 Density gradient centrifugation

Sucrose and iodixanol are the most commonly used media to isolate EVs by density gradient centrifugation (DGC). The isolation is based either on size and mass (top-down gradient) or on mass density (bottom-up gradient) (88). When isolating EVs according to their density gradient instead of according to the difference in size, contamination of the sample with both, proteins and lipoproteins, is greatly reduced (84). High density lipoprotein particles (HDL) have similar density as EVs and will be therefore likely isolated together with EVs (88). However, Nanoparticle Tracking Analysis (NTA) most likely will not detect HDL due to their small size (around 10 nm) and the NTA concentration measurements of EVs isolated by DGC should be more accurate than of EVs isolated by other techniques (84). On the other hand, low density lipoproteins (LDL) have different density then EVs but co-isolation with EVs still occurs (88).

3.5.3 Ultrafiltration

Ultrafiltration (UF) isolates EVs based on targeted molecular weight or on size by using membrane filters (89). The pressure is applied to pass the soluble components through the filter, or the filter is placed in a centrifuge (CF, centrifugal filtration) (88). Compared to other techniques, this method is time efficient and gives high EV recovery (88). However, problems might occur with clogging of filters (89) or the appearance of deformable particles.

3.5.4 Precipitation

Precipitation-based isolation techniques of EVs uses polymers and are commonly used in commercial products (89). Commercial kits are often PEG based - the PEG is a water-soluble molecule which binds water molecules and force less soluble components out of solution (89). The EVs isolated by precipitation have been reported as contaminated with free miRNA, albumin and HDLs, as well as the presence of morphologically-like-lipoprotein particles and the protein-aggregates-like by TEM analysis background have been reported (90). Even though this method is time and money effective as well as the low sample volume is required (such as 250 µL), it seems that it suffers from high and very heterogeneous contamination which might strongly affect downstream analysis (89, 90).

3.5.5 Size exclusion chromatography

Size exclusion chromatography (SEC) enables size-based separation on a single column (88). SEC does not induce aggregation of EVs nor alters the size and the vesicular characterization (88, 91). It provides more intact EVs and EVs of higher functionality compared to UC (92). Compared to PEG and protein organic solvent precipitation (PROSPR), SEC provides the cleanest sample, removing most of the overabundant plasma proteins (91). The major contaminants are the particles above the size cutoff, such as protein aggregates, very large proteins, LDLs, and viruses (88). EV markers, such as cluster of differentiation (CD) 9, CD63, CD81 and Alix were able to be detected after SEC isolation of EVs (91, 92). Reproducible recovery of EVs given by SEC is between 40% and 90% and the technique is fast, relatively inexpensive and has a high yield of biophysically intact EVs (88). However, the final dilution of EVs in a sample processed by SEC is often higher compared to other techniques (88).

3.5.6 Immuno-affinity capture

Affinity-based techniques are based on capturing surface proteins of EVs with specific antibodies (89). The antibodies are usually immobilized on a surface such as plate, beads, column, or chip (88, 93). This approach provides high purity of the sample and highly specific isolation of EVs (89) and it is possible to isolate subpopulations of EVs (88). Therefore, it could be used in order to investigate the specific role of different EV subtypes (87). E.g., the CSPG4 epitope is present on more than 80% of melanomas and it is carried by melanoma-derived EVs (94), which enabled the antibody-based isolation of melanoma-derived EVs. Unfortunately, surface biomarkers for various cancer derived EVs are yet to be discovered. Also, due to high specificity towards certain proteins, only subpopulation of EVs that contain the target protein will be isolated. Therefore, this method is not suitable for generalized EV studies (87). To avoid isolation of specific EV subtypes because of protein specificity, EV isolation based on capturing of lipids (phosphatidylserine) on the surface of EVs can be used (95).

To date, there is no ideal EV isolation technique but the choice depends on the downstream analysis and the final goal of the EV research. For functional studies with a need for intact EVs, the technology of choice could be SEC or immunno-affinity capturing. The same choice, together with DGC, could be made for any study where there is need for a sample of high purity. For downstream analysis that are not sensitive to contamination, precipitation, UC, or UF could be the technologies of choice, since they are fast, well established and relatively inexpensive.

3.6 Mass spectrometry (MS)–based proteomics

Proteomics is the analysis of protein mixtures or their polypeptide components in biological systems such as cells, tissues or organisms, at once (96, 97). If proteins from two different samples are being analysed, such as from healthy person and a cancer patient, proteomics is classified as classical, while if the focus is on a particular sub-proteome, it is classified as targeted proteomics (98).

Proteomics is studied through mass spectrometry (MS), where the mass spectrometer has three main components: the *ionization source*, where ions are produced, a *mass analyser*, where ions are separated based to their mass-to-charge (*m/z*) ratio, and a *detector*, where ions are detected (98). There are two types of ionization: *positive*, when the electric potential at the ion source is positive and peptides are analysed at low pH, and the Arg, Lys, and His residues are protonated. The another type of ionisation is *negative* ionization, where the electric potential at the ion source is negative and peptides are analysed at high pH and the Asp and Glu residues are de-protonated (98).

During MS analysis, ionized peptides are detected as ions in a mass spectrum in a process called *direct infusion* (in electrospray ionization-mass spectrometry, ESI-MS mode) and resulting MS displays masses of the peptides (98). For sequence identification of a peptide, the corresponding peptide, called *precursor peptide*, is isolated and fragmented in the collision cell using neutral gas resulting in MS/MS spectrum (ESI-MS/MS mode), where MS/MS spectrum represents the sum of spectra of the product ions after precursor peptide fragmentation (98). For determination of amino acid sequence, the peptides must be selected for fragmentation one at the time. Therefore, peptides can be fractionated by column chromatography coupled to a *high-performance liquid chromatography (HPLC) mass spectrometer* (98).

In general, there are two methods that can be used for MS analysis. *Top-down* approach in which intact proteins are being investigated, and *bottom-up* approach in which proteins are digested with proteolytic enzyme, usually trypsin, and the peptide mixture is analysed (96).

Except for protein identification, mass spectrometry can be used for protein quantification, which can be based on label- or on label-free techniques. Label-free methods provide relative quantification, and they are simple to use, cost effective, they have broader dynamic range and proteome coverage; however, the reproducibility is challenging (99, 100). Label-based quantification offer higher quantification precision and accuracy compared to label-free quantification due to the peptide detection in the same MS spectrum, where the difference in the isotope pattern and mass difference between native and labelled samples allow absolute quantification (99, 101). Isotopes can be incorporated metabolically, such as in stable isotope labelling of amino acids in cell culture (SILAC), or by adding tags on specific amino acids, such as in isotope-coded affinity tag, isobaric tags for relative and absolute quantification, or tandem mass tags (97, 98).

Diagnostic proteins are often hidden among the high abundant functional proteins in the cell, resulting in lack of detection if not enriched (102). With the development of state-of-the-art mass spectrometry approaches such a limitation can be overcome. Therefore, MS-based development of protein biomarkers and their clinical utility hold great potential in the future.

3.7 Clinical potential and utility of mass spectrometry

Proteins are key functional entities in the cell and executors of any cellular activity, which makes them the most prominent drug targets (103). Additionally, the genome cannot predict all disease associated proteoforms, which makes use of MS-based proteomics for protein biomarker discovery rising (102). *Proteoforms* are a group of proteins that are products of the same gene or gene family which have defined and specific amino acid sequence and localized posttranslational modification, resulting in protein variants with different functions (104, 105). Contrary to genomics or transcriptomics, proteomics can provide direct information about the structure, abundance or signalling and enzymatic activity of proteins (96).

MS-based proteomics has potential to be used in diagnostics, prognostics or in understanding the biology of disease. Protein-based diagnostic tests are already widely in use, such as pregnancy tests or rapid diagnostic tests for malaria (96). In the past decade, over a hundred protein biomarkers for clinical utility have been developed. Many of them were approved by the FDA for cancer diagnosis, monitoring therapeutic response, or disease recurrence evaluation, such as prostate-specific antigen (PSA), carbohydrate antigen (CA) 125, and CA 19-9 for prostate, ovarian, and pancreatic cancers, respectively (103). However, in the case of lung cancer, no protein biomarker for lung cancer detection has been approved by the FDA yet.

Translating MS-derived protein biomarkers into clinical application come with challenges since in most of the cases novel protein biomarkers are detected in small discovery cohort and have to be validated in large validation cohorts (102). Further on, development of clinical assays for validated biomarkers that are robust, easy-to-use, cheap and of adequate sensitivities and specificities comes with challenges as well (102).

In this thesis, based on mass spectrometry data, three proteins have been considered as proteins with a potential to be clinically used for lung cancer detection and will be further described below (3.8, 3.9 and 3.10).

3.8 NECTIN-4 (NECT-4)

NECTIN (NECT)-4 protein (Q96NY8, UniProt), also called Poliovirus receptor-related protein (PVRL)-4 was first identified in 2001 through the bioinformatic search (106). Together with NECT-1, NECT-2, and NECT-3, NECT-4 belongs to the nectin family of proteins which is part of immunoglobulin (Ig) superfamily (106, 107). NECT-4 has two described isoforms produced by alternative splicing, of which isoform 1 is a canonical isoform further on designated as cNECTIN-4 (cNECT-4) and described as follows. cNECT-4 is a type-I transmembrane protein, composed of extracellular, transmembrane, and cytoplasmic domain, where extracellular domain comprises three immunoglobulin (Ig)-like domains, one V-type and two C2-type domains (Figure 5) (108). It is a 510-amino acids long protein, with the predicted molecular weight (MW) of 55.454 kDa (109). Due to N-glycosylation site on the extracellular domain (Figure 5), cNECT-4 can migrate with a mass of 66 kDa on SDS-polyacrylamide gels (110). Extracellular domain can be shed of cNECT-4 by enzymatic activity of disintegrin and metalloproteinase domain-containing (ADAM) proteins, ADAM10/ADAM17 respectively, resulting in soluble form of NECT-4, which can be released into the blood (111, 112).
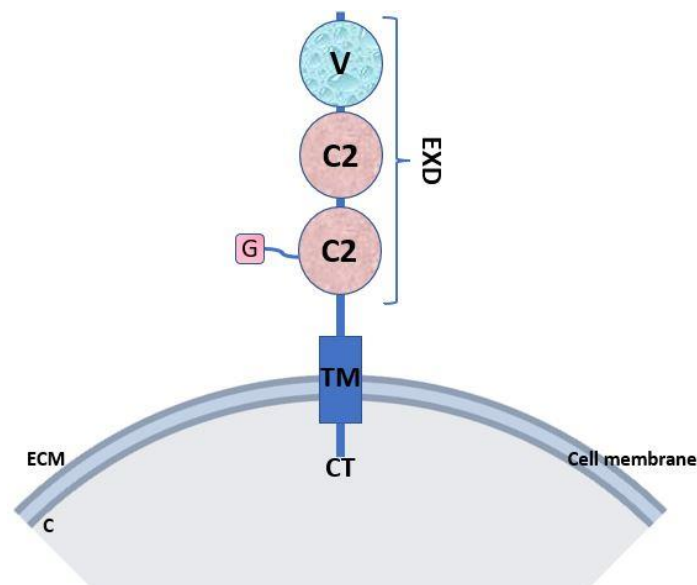
Figure 5. Scheme of cNECT-4 structure. cNECT-4 is type-I transmembrane protein with short cytoplasmic tail, transmembrane domain, and extracellular domain. Extracellular domain is composed of three immunoglobulin (Ig)-like domains, one V-type and two C2-type domains and it represents majority size of the protein. ECM=extracellular matrix; C=cytoplasm; CT=cytoplasmic tail; TM=transmembrane domain; EXD=extracellular domain; C2=C2-type domain; V=V-type domain; G=N-linked glycosylation site.

cNECT-4 can form homophylic interactions with other cNECT-4 molecules, initially in cis- followed by trans-interactions, and heterophilic trans-interactions with NECT-1 molecules (108). Contrary to the other members of nectin family which are widely expressed in adult tissue, NECT-4 is expressed mainly

in embryo and placenta (113), while expression in adult tissue is limited (113, 114). Additionally, high expression has been detected in different cancer entities, such as breast, ovarian, lung, colorectal, pancreatic and bladder cancers (108, 114).

cNECT-4 is localized on the apical surface of cells and it is involved in adherence junctions together with cadherins (107). Additionally, cytoplasmic tail of cNECT-4 interacts with actin cytoskeleton through AFADIN protein, which is also part of adhesion system (110). Therefore, cNECT-4 is involved in cell-to-cell adhesion, movement, intercellular communication, proliferation, differentiation, polarization, and the entry of viruses (108, 113).

3.9 Platelet-Derived Growth Factor-Associated protein 1 (PDAP1; HASPP28)

Platelet-Derived Growth Factor-Associated Protein 1 (PDAP1) (Q13442, UniProt), also called 28 kDa heat- and acid-stable phosphoprotein (HASPP28) was first identified in 1996 by two independent research groups (115, 116), resulting in different suggested nomenclature as well. PDAP1 has been purified associated with Platelet-Derived Growth Factor (PDGF)-A, therefore the name PDAP1 (115), and PDAP1 has been purified from heat- and acid-stable protein fraction, therefore the name HASPP28 (116). Additionally, it has phosphorylation sites for casein kinase II, protein kinase A, protein kinase C.

PDAP1 is an 181-amino acids long protein, with a predicted MW of 20.63 kDa (117). However, PDAP1 has been reported to migrate with a mass of 28 kDa on SDS-polyacrylamide gels (116). Even though it does not possess N-terminal secretory signal, PDAP1 is secreted into extracellular space (115). Further on, it can be found in cytoplasm or associated with cell membrane (117, 118), and it is broadly expressed among tissue types, as well as across different cancer cell lines (118, 119).

PDAP1 has been associated with mitosis via modulating PDGF activity (115)-PDAP1 binds to PDGF with low affinity and enhances the mitogenic effect of PDGFA, but lowers the mitogenic effect of PDGFB. Additionally, PDAP1 has been associated with cell proliferation (118), cell survival (120), membrane depolarization and apoptosis (121), diversification of Ig (120), and described as RNA-binding-protein (118). PDAP1 is overexpressed in multiple cancers, such as rectal carcinoma (122), gastric cancer (123), and glioma (121) respectively.

3.10 Cellular Communication Network Factor 1 (CCN1)

Cellular Communication Network Factor 1(CCN1) protein belongs to CCN family of proteins which has six members. CCN acronym derived from the first three discovered members of the family, cysteine-rich angiogenic inducer 61 (CCN1 or CYR61), connective tissue growth factor (CCN2 or CTGF), and

nephroblastoma overexpressed gene protein homolog (CCN3 or NOV) (124). Other members of CCN family are Wingless/Int-1 (WNT1)-inducible-signalling pathway protein 1 (CCN4 or WISP-1), WNT1-inducible-signaling pathway protein 2 (CCN5 or WISP-2), and WNT1-inducible-signaling pathway protein 3 (CCN6 or WISP-3) (75). CCN proteins are modular in their structure, consisting of an N-terminal signal sequence followed by domains with sequence similarity to insulin-like growth factor-binding protein, von Willebrand factor C, thrombospondin type 1, and a cysteine knot at the C terminus. Among all CCN proteins, WISP-2 is the only one lacking C-terminal region (125).

CCN1 is an immediate early gene, and CCN1 is secreted, extracellular protein associated with ECM and described as matricellular protein due to its dynamics (126). CCN1 is an integrin ligand (127, 128), and it is involved in multiple internal and external cellular signalling, such as in cell adhesion, migration, mitogenesis, differentiation, and survival (126). Finally, CCN1 has been associated with tumorigenesis (129) and EMT in multiple cancers (130-132).

4. Aim of the thesis

Lung cancer is the leading cause of cancer death worldwide and the detection of lung cancer is still challenging until very late stages. Therefore, the goal of this thesis was an investigation of biomarkers for an early detection of primary lung cancer. To achieve that, this thesis had two aims. One aim was an in-depth investigation of lung cancer cell line proteome and identification of novel biomarkers for the early detection of lung cancer. The other aim was an in-depth investigation of EV-proteins as biomarkers for the early detection of lung cancer.

To identify novel biomarkers for the early detection of lung cancer, a SILAC-based comparative proteomic analysis of epithelial and mesenchymal lung cancer cell line was performed. CTCs and DTCs play essential role in cancer progression and metastasis and some of these cells undergo EMT and contribute to the cancer progression. Additionally, the utilization of plasma samples and the feasibility of ELISA performance is convenient for the implementation into the clinical setting. Therefore, the aim was to use SILAC-based comparative proteomic analysis to identify epithelial and mesenchymal secreted proteins as potential serological biomarkers for the early lung cancer detection. Finally, the aim was to validate MS-based potential plasma biomarkers on the study cohort comprising lung cancer patients and healthy controls.

Additionally, extracellular vesicles emerged as a promising tool for biomarker discovery. However, there is no standardized method for the EV isolation yet. Moreover, the EVs of sufficient purity are still challenging to obtain for downstream MS analysis. Therefore, another aim of this thesis was to establish the protocol for the isolation of EVs from plasma samples suitable for downstream mass spectrometry analysis. Additionally, the aim was to perform label-free-quantification (LFQ) LC-MS/MS of plasma EVs deriving from lung cancer patients and healthy controls to identify potential EV-deriving biomarkers for the early detection of lung cancer. Finally, it was sought to provide insight into the biological role of the EVs in lung cancerogenesis using gene-ontology-based pathway analysis of genes associated with the upregulated proteins identified on EVs by LFQ-LC-MS/MS.

5. Results

The goal of this thesis was an investigation of biomarkers for the early detection of primary lung cancer. To achieve that it was aimed to investigate lung cancer cell line proteome and to identify secreted protein biomarkers for the early detection of lung cancer. Another aim was to investigate EV proteins as novel biomarkers for the early detection of lung cancer.

For an identification of potential secreted protein biomarkers for the early detection of lung cancer, a SILAC-based proteomic analysis sought to be performed. For that, the suitable pair model of cell lines for SILAC-based LC-MS/MS was identified by *Western blot* characterization of different lung cancer cell lines for the most common epithelial and mesenchymal proteins. Then, the work pursued to identification of potential epithelial and mesenchymal lung cancer biomarker proteins based on LC-MS/MS data. Proteins identified as potential biomarkers for early lung cancer detection were further validated on study cohort using ELISA. Finally, diagnostic performance of an assay was analysed using ROC analysis.

5.1 Characterisation of lung cancer cell lines for epithelial and mesenchymal biomarkers

To identify novel biomarkers for detection of lung cancer in patients, a set of lung cancer cell lines were analysed for the establishment of suitable model for mass spectrometry-based proteome analysis.

Therefore, seven different lung cancer cell lines, comprising of HCC-366, HTB-58, H1975, H1299, H1395, H1993, and LC-M1, together with two breast cancer cell lines MDA-MB-231 (MDA-231) and MCF-7 which served as controls, were characterized for multiple epithelial and mesenchymal markers, and EGFR which is the product of one of the driver genes in lung carcinogenesis (Figure 6). HCC-366, H1975, and H1993 are positive for epithelial markers, such as cytokeratins (CKs) CK18, CK19, the CKs detected by the pan-keratin antibodies A45 and AE1/AE3, and the EPCAM. Additionally, HCC-366 and H1993 were detected positive for EGFR, while HCC-366 and H1975 were positive for CK8. Together with epithelial markers, H1975 is detected positive for VIMENTIN, which is a mesenchymal marker. Therefore, HCC-366 and H1993 display an epithelial phenotype, while even though the detection of epithelial markers in H1975 is strong, this cell line displays hybrid phenotype due to the detection of VIMENTIN. H1395 is another cell line characterized by epithelial phenotype, as detected positive for A45, CK18, and EPCAM epithelial markers.

Further on, HTB-58 and H1299 display a hybrid phenotype as well. HTB-58 and H1299 were characterized by detection of CK18 and EGFR, together with N-CADHERIN and VIMENTIN which are additional mesenchymal markers. HTB-58 is also positive for CK as detected by A45 pan-antibody-cocktail of epithelial markers. LC-M1 is characterized by mesenchymal phenotype, expressing N-

CADHERIN and VIMENTIN mesenchymal markers and showing no expression of any epithelial marker. Detection of α-TUBULIN served as a loading control.
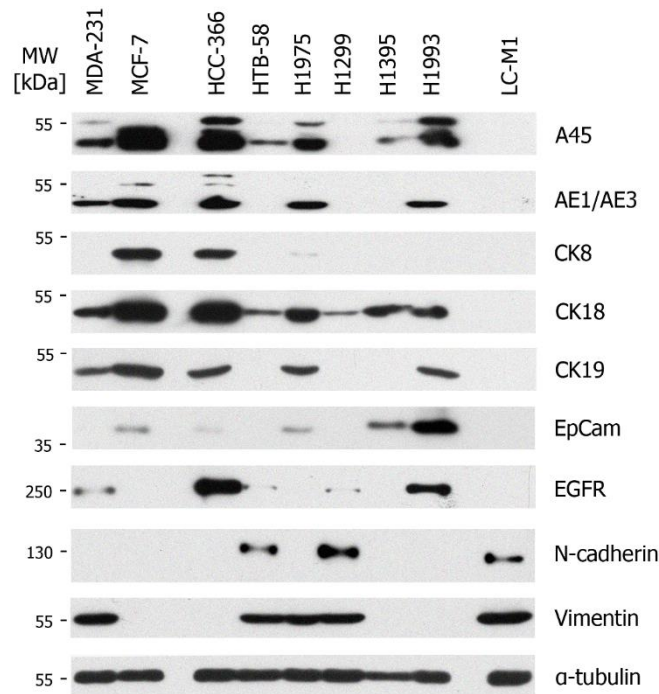
Figure 6. Characterization of different lung cancer cell lines for epithelial and mesenchymal markers using Western blot analysis. HCC-366, HTB-58, H1975, H1299, H1395, H1993 and LC-M1 lung cancer cell lines together with MDA-231 and MCF-7 breast cancer cell lines as controls, were characterized for common epithelial and mesenchymal markers, and EGFR which is product of one of the driver genes in lung carcinogenesis. α-TUBULIN served as loading control. Twenty micrograms of total proteins were loaded. n=3. A45=pan-antibody cocktail recognizing CK8/CK18 and CK9/CK19 cytokeratin heterodimers, AE1/AE3=pan-antibody cocktail where AE1 detects CK10, CK14, CK16 and AE3 detects CK1, CK2, CK3, CK4, CK5, CK6, CK7, and CK8, CK=cytokeratin, EpCAM=epithelial cell adhesion molecule, EGFR=epidermal growth factor receptor, MW=molecular weight, kDa=kilodalton, n=biological replica.

Based on results from 5.1 section, H1975 and LC-M1 cell lines were chosen as respective models for SILAC-based LC-MS/MS proteome analysis of lung cancer phenotype, including detection of potential epithelial and/or mesenchymal biomarkers for lung cancer detection.

5.2 Characterization of morphological consistency of H1975 and LC-M1 cell lines after SILAC

Morphology of H1975 and LC-M1 cells in normal medium and in SILAC heavy and light medium has been analysed by light microscopy to make sure there are no morphological changes of the cells due to the growth in SILAC medium (Figure 7). In general, H1975 cells were of an irregular, spindle-like-

shape with sharp edges. After the cultivation in heavy or light SILAC medium no morphological difference was observed between H1975 grown in SILAC medium compared to the H1975 cells grown in normal medium (Figure 7, above). Within LC-M1 cell population the mixture of elliptical cells and spindle-like-shaped cells could have been observed, and there was no morphological difference observed between LC-M1 cells grown in normal medium compared to LC-M1 cells grown in SILAC medium (Figure 7, below).
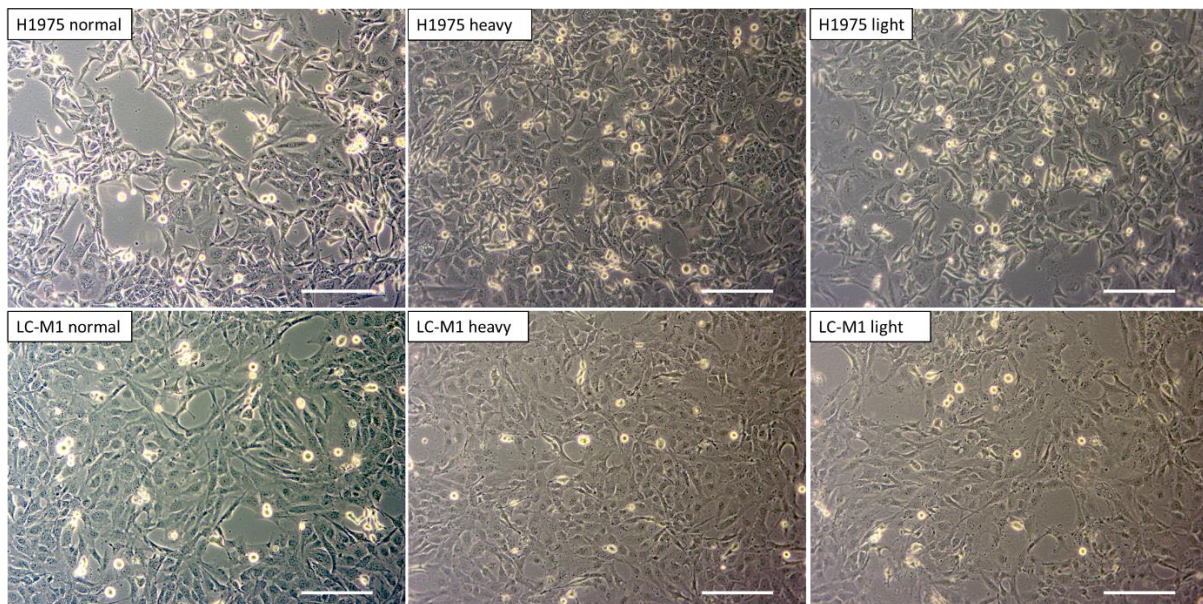
Figure 7. Morphology of H1975 and LC-M1 cells during cultivation in normal medium, SILAC heavy medium, or SILAC light medium. Above: Spindle-like morphology of H1975 cells grown in normal medium (left), heavy SILAC medium (middle), and in light SILAC medium (right). Below: Mixture of spindle-like and elliptical morphology of LC-M1 cells grown in normal medium (left), heavy SILAC medium (middle), and in light SILAC medium (right). White scale bars represent 50 µm length. Images were taken at 100× magnification. n=3. n=biological replica.

5.3 Proteome analysis of H1975 and LC-M1 cell lines

H1975 was chosen as a model for lung cancer cells with strong manifestation of epithelial characteristics and LC-M1 was selected as a model for tumor cell with mesenchymal phenotype. For the proteome analysis, the proteome of H1975 was labelled with $^{13}C_6$ arginine and $^{13}C_6$ lysine (heavy) and LC-M1 was cultured in the presence of corresponding light amino acids (light). Then, the labelling was switched in a second experiment. For the subsequent data analysis, proteins overexpressed in the heavy cell lines were predominantly analysed. Metabolic labelling was applied where cells were cultivated until almost complete labelling of cellular proteome (8.2.8 M&M section). Extracts containing heavy- and light-labelled proteins were mixed in 1:1 ratio after LC-MS/MS sample preparation was performed (8.3.7.2 M&M section), followed by tryptic digestion and prefractionation of peptides by OFF-GEL which were finally analyzed by LC-MS/MS (8.3.7, 8.3.8, 8.3.9 and 8.3.10 M&M sections). Heavy-labelled peptides

refer to the cellular protein products which metabolically integrated $^{13}C_6$ arginine or $^{13}C_6$ lysine, resulting in a mass shift of 6 Da compared to the light proteins. In this experiment, precursor mass spectra were analysed by MaxQuant software allowing identification and quantification of peptides (8.7.1 M&M section).

5.3.1 Determination of appropriate passage for complete labelling of cellular proteome

To evaluate at which passage a complete incorporation of heavy labelled isotopes occur, H1975 and LC-M1 cell lines were cultivated from passage (P) 0 until P5 in heavy SILAC medium. For each passage two flasks were generated, one for generation of the next passage and another for harvesting cells in lysis buffer (Table 18). Total protein concentration was measured for all harvested passages using BCA test (8.3.2 M&M section). For all passages (P0-P5), 20 µg of total protein extract was subjected to sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE), and the gel was stained using Coomassie blue staining (8.3.5 M&M section). Two bands were cut out for each passage and subjected to in-gel tryptic digestion (8.3.7.1 M&M section) and measured by Orbitrap Fusion mass spectrometer (8.3.10 M&M section). Complete and successful labelling was characterized by the disappearance of light peptide ion signal and the appearance of heavy peptide ion signal. Moreover, the absence of heavy peptide ion signal with mass shift of additional 5 kDa was sought indicating that conversion of stable isotope labelled arginine to proline did not occur.

For H1975 cell line, complete and successful labelling was observed in passage 3 (Figure 8). Figure 8 shows exemplary spectra of the IINEPTAAAIAYGLDR peptide, originating from Heat shock 70 kDa protein (HSPA) (HSPA1B, Swiss Prot P0DMV9; HSPA6; Swiss Prot P17066). For given peptide, the monoisotopic peak of the light peptide ion signal was predicted at m/z 844.4543 $[M + 2 H]^{2+}$. In P0, the light peptide ion signal was detected at m/z 844.4534 $[M + 2 H]^{2+}$, whereas the heavy peptide ion signal lacks as no heavy labelled amino acids were added in the medium for the given passage (Figure 8a). The signal for heavy peptide ion signal is expected at m/z 847.4543 $[M + 2 H]^{2+}$ due to the charge for this peptide of z=2 ($\frac{m}{z} heavy = \frac{m}{z} light + \frac{6kDa}{z}$). Complete and successful incorporation of heavy amino acids was observed in P3, at which the appearance of heavy peptide ion signal at m/z 847.4639 $[M + 2 H]^{2+}$ occurs, whereas the light peptide ion signal was not detected (Figure 8b). Additionally, if conversion of arginine to proline had occurred, the peptide ion signal would be expected at m/z 849.9543 $[M + 2 H]^{2+}$ ($\frac{m}{z} heavyArg \rightarrow Pro = \frac{m}{z} heavy + n\frac{Pro*5kDa}{z}$). A small peak is observed at m/z 849.9689 $[M + 2 H]^{2+}$; however, no ion signal for arginine to proline conversion is observed at expected position (Figure 8c). If above mentioned small ion signal truly corresponds to the peptide with converted arginine to proline, the intensity of the signal is anyhow very low (Figure 8b and c). Therefore, the potential rate

of conversion (if occurred) was observed as acceptable and considered as if it does not affect the final results.
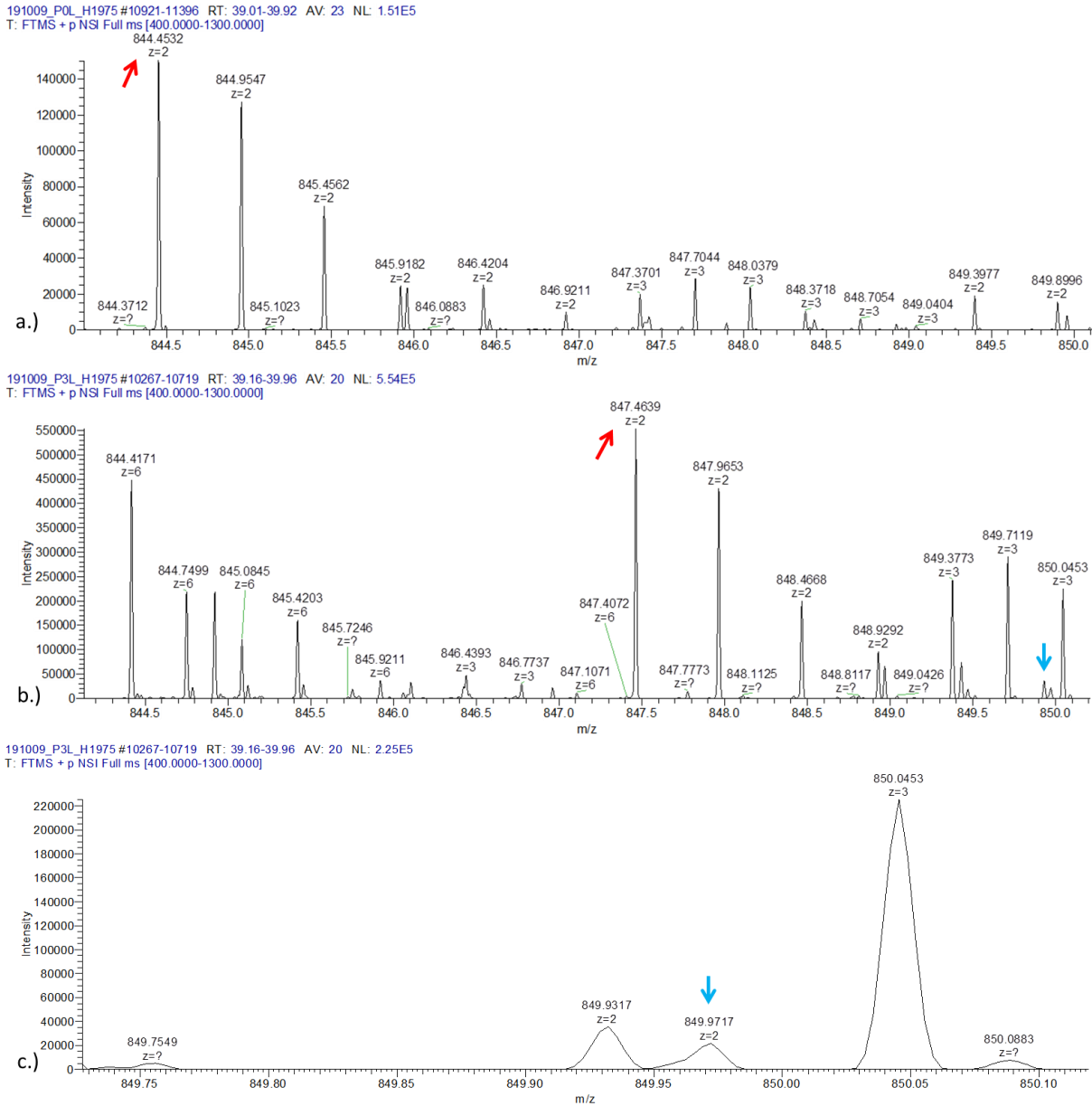
Figure 8. MS survey scan of IINEPTAAAIAYGLDR peptide in passages 0 and passage 3 of H1975 cell line. MS survey scan of the peptide IINEPTAAAIAYGLDR of the Heat shock 70 kDa protein for passages 0 and 3 show the peak intensities on the vertical axes, whereas the horizontal axes show m/z. a.) MS survey scan of the peptide IINEPTAAAIAYGLDR of the Heat shock 70 kDa protein in passage 0 shows a light peptide ion signal at the position m/z 844.4534 $[M + 2\ H]^{2+}$ (red arrow) and lack of heavy peptide ion signal. b.) MS survey scan of the peptide IINEPTAAAIAYGLDR of the Heat shock 70 kDa protein in passage 3 shows the shift of light peptide ion signal to the heavy peptide ion signal at the position m/z 847.4639 $[M + 2\ H]^{2+}$ (red arrow) and small peptide ion signal at position m/z 849.9309 $[M + 2\ H]^{2+}$, which might have occurred due to the proline conversion (blue arrow). c.) MS survey scan of the peptide IINEPTAAAIAYGLDR of the Heat shock 70 kDa protein in passage 3 with a zoomed position at which ion signal for a peptide with converted arginine to proline would have appeared. Small peptide ion signal is observed at position m/z 849.9689 $[M + 2\ H]^{2+}$ which might correspond to the ion signal of a peptide with converted arginine into proline (blue arrow).

For LC-M1 cell line, complete and successful labelling was observed in passage 4 (Figure 9). Figure 9 shows exemplary spectra of an AVFVDLEPTVIDEVR peptide, originating from α-TUBULIN (TBA) (TBA1A, Swiss Prot: Q71U36; TBA1B, Swiss Prot: P68363; TBA1C, Swiss Prot: Q9BQE3). For the given peptide, the light peptide ion signal should occur at *m/z* 851.4565 [M + 2 H]$^{2+}$. In P0, the monoisotopic peak of the light peptide ion signal at *m/z* 851.4584 [M + 2 H]$^{2+}$ can be observed, whereas the heavy peptide ion signal lacks as no heavy labelled amino acids were added in the medium for the given passage (Figure 9a). The signal for heavy peptide ion signal is expected at *m/z* 854.4565 [M + 2 H]$^{2+}$ due to the charge for this peptide of z=2 ($\frac{m}{z}\,heavy = \frac{m}{z}\,light + \frac{6kDa}{z}$). Complete and successful incorporation of heavy amino acids was observed in P4, at which the appearance of heavy peptide ion signal at *m/z*=854.4680 [M + 2 H]$^{2+}$ occurs, whereas the light peptide ion signal was not detected (Figure 9b). Additionally, if conversion of arginine to proline had occurred, the peptide ion signal would be expected at *m/z* 856.9565 [M + 2 H]$^{2+}$ ($\frac{m}{z}\,heavyArg \rightarrow Pro = \frac{m}{z}\,heavy + n\frac{Pro*5kDa}{z}$). A small peak is observed at m/z=856.9707 [M + 2 H]$^{2+}$ (Figure 9b). However, no ion signal for arginine to proline conversion is observed at expected position (Figure 9c). After more in depth analysis, an additional small peak was detected at *m/z* 856.9495 [M + 2 H]$^{2+}$, which potentially could be contributed to the ion signal for a peptide with converted arginine into proline (Figure 9c). If any of the two above mentioned small ion signals truly corresponds to the peptide with converted arginine to proline, the intensity of the signal is low (Figure 9b and c). Therefore, the potential rate of conversion (if occurred) was observed as acceptable in this work.
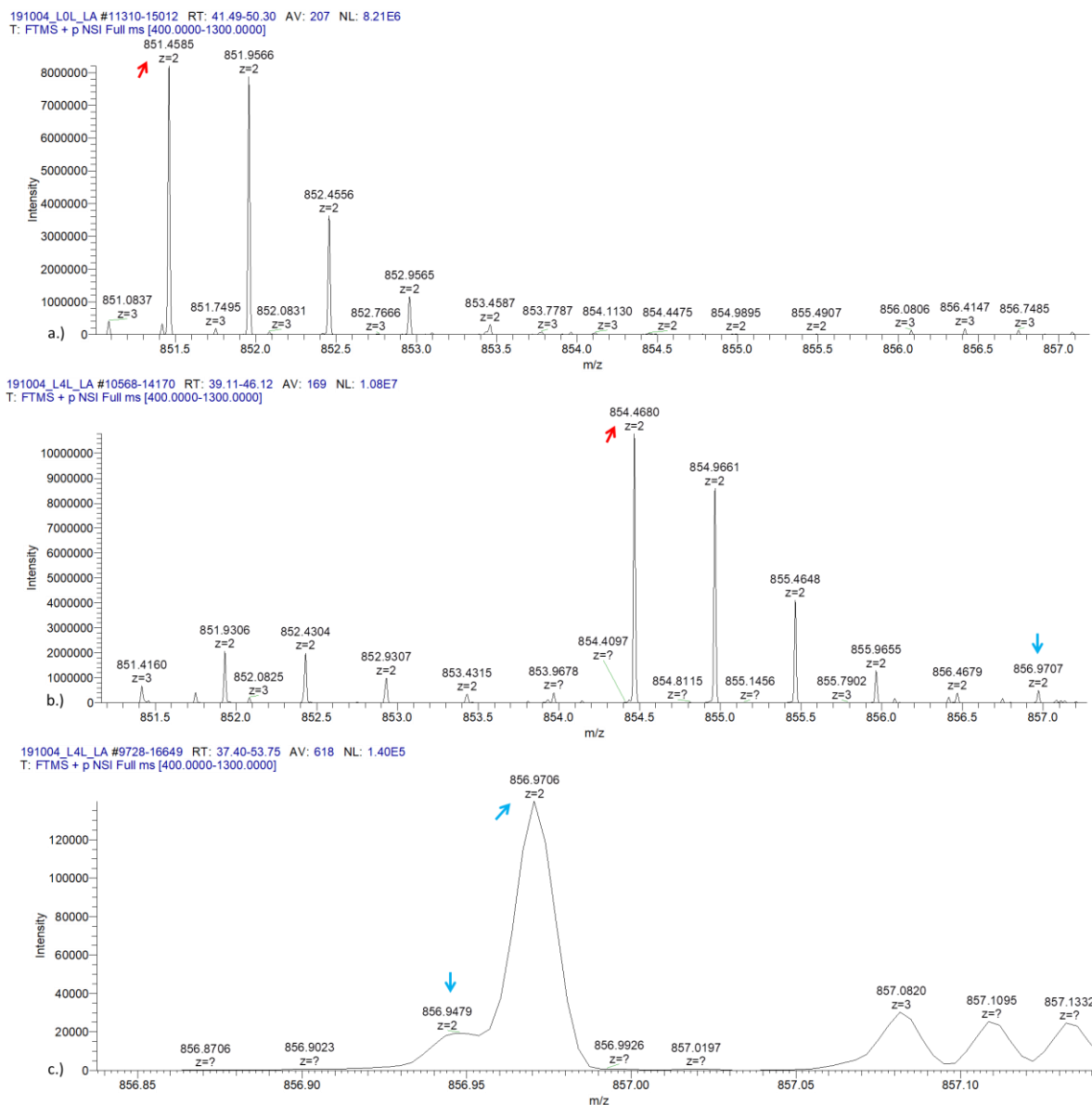
Figure 9. MS survey scan of peptide AVFVDLEPTVIDEVR in passages 0 and passage 4 of LC-M1 cell line. MS survey scan of the peptide AVFVDLEPTVIDEVR of α-TUBULIN protein for passages 0 and 4 show the peak intensities on the vertical axes, whereas the horizontal axes show *m/z*. a.) MS survey scan of the peptide AVFVDLEPTVIDEVR of α-TUBULIN in passage 0 shows a light peptide ion signal at *m/z* 851.4584 $[M + 2 H]^{2+}$ (red arrow) and lack of heavy peptide ion signal. b.) MS survey scan of the peptide AVFVDLEPTVIDEVR of α-TUBULIN in passage 4 shows conversion of the light peptide ion signal to heavy peptide ion signal at *m/z* 854.4680 $[M + 2 H]^{2+}$ (red arrow) and small peptide ion signal at *m/z* 856.9707 $[M + 2 H]^{2+}$, which might have occurred due to the proline conversion (blue arrow). c.) MS survey scan of the peptide AVFVDLEPTVIDEVR of α-TUBULIN in passage 4 with a zoomed position at which ion signal for a peptide with converted arginine to proline would have appeared. Two peptide ion signals were observed, one at *m/z* 856.9495 $[M + 2 H]^{2+}$, and another at *m/z* 856.9707 $[M + 2 H]^{2+}$, which might correspond to the ion signal of a peptide with converted arginine into proline (blue arrows).

5.3.2 Evaluation of the protein concentration ratios in SILAC H1975/LCM1 samples

For differential proteome analysis, it is essential to accurately determine protein concentration in samples, and that heavy and light proteomes are truly mixed in ratio 1:1. Therefore, the determination of protein concentration was performed as follows. After harvesting cells, the sample preparation was stopped at the point when protein pellets were dissolved in 100 µL of 9.8 M urea (8.3.7.2 M&M section). At that point, 10 µL of sample was aliquoted and diluted with 9.8 M urea in ratio 1:5. The diluted aliquot was used for determination of protein concentration and validation of obtained values. BCA test was performed (8.3.2 M&M section) and 20 µg of proteins in total from each sample, in both set of experiments, were loaded on analytical polyacrylamide gel (8.3.3 M&M section). Samples analyzed in one set of experiment were H1975$_{heavy}$ + LC-M1$_{light}$ (1:1), H1975 SILAC heavy, and LC-M1 SILAC light. Samples analyzed in another set of experiment were H1975$_{light}$ + LC-M1$_{heavy}$ (1:1), H1975 SILAC light, and LC-M1 SILAC heavy. Gels were stained with Coomassie blue, and three bands from 1:1 mixture lane were cut out (Figure 10) and protein ratio 1:1 was validated by mass spectrometry (8.3.7.2 and 8.3.10 M&M section).
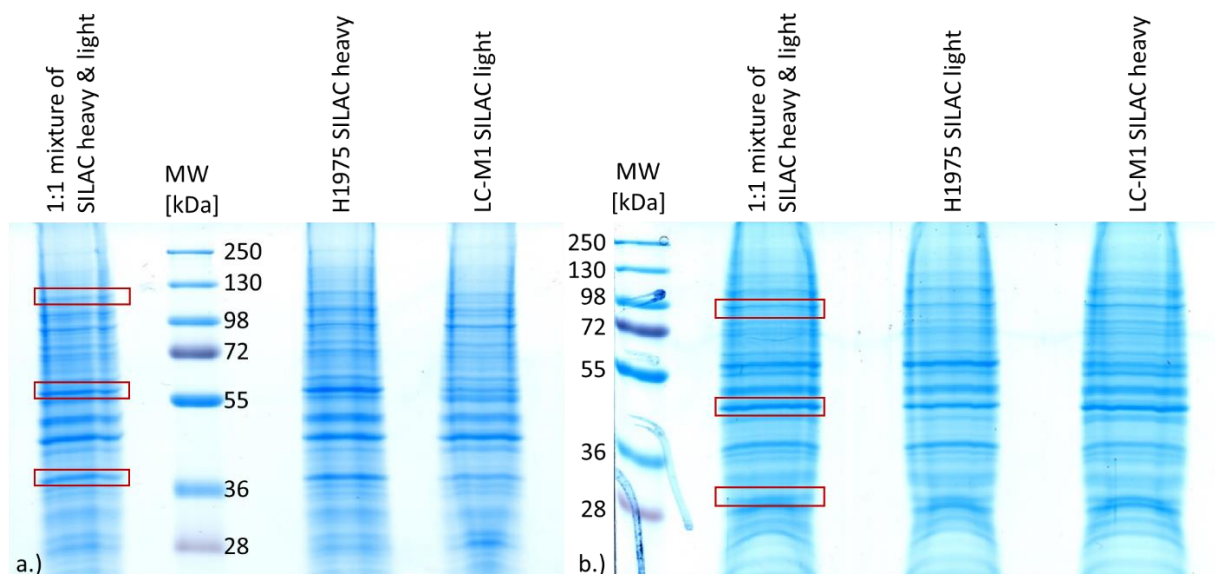
Figure 10. Analytical Coomassie blue stained polyacrylamide gel of SILAC samples. Red squares denote the bands which were analyzed by LC/MS-MS. All lines were loaded with 20 µg of total protein amount. n=3. a.) Analytical Coomassie blue stained polyacrylamide gel of H1975 SILAC heavy and LC-M1 SILAC light samples mixed in 1:1 ratio, only H1975 SILAC heavy, and only LC-M1 SILAC light (from left to the right) from replica 2. b.) Analytical Coomassie blue stained polyacrylamide gel of H1975 SILAC light and LC-M1 SILAC heavy samples mixed in 1:1 ratio, only H1975 SILAC light, and only LC-M1 SILAC heavy (from left to the right) from replica 1. n=biological replica.

Mass spectrometry data resulted in many proteins with a ratio≠1 between two analyzed samples. The ratio≠1 was expected due to the different phenotypes between H1975 and LC-M1 cell lines, hybrid and mesenchymal respectively. Therefore, they are expected to have differential expression of proteins. However, together with cell line specific expression pattern, it was assumed that some proteins, e.g., products from housekeeping genes and potentially some others, would have similar expression pattern resulting in normalized ratio H/L=1. With this assumption, on a logarithmic scale the distribution of normalized H/L ratio from all proteins have to be equal around the normalized ratio H/L=1. Taking that, the protein concentration was considered as accurate, if the saddle point was at a normalized H/L ratio between 0.9 and 1.1. In all replicates tested, normalized ratio H/L was between 0.9 and 1.1, and experiments were proceeded further on. In Figure 11a, normalized ratios H/L of all proteins detected in bands from replica 2 of a mixed H1975 SILAC heavy and LC-M1 SILAC light sample are shown. For this sample, the saddle point on the graph with logarithmic scale is at normalized ratio H/L=0.98168. In Figure 11b, normalized ratios H/L of all proteins detected in bands from replica 1 of a mixed H1975 SILAC light and LC-M1 SILAC heavy sample is shown. For this sample, the saddle point on the graph with logarithmic scale is at normalized ratio H/L=0.99743. In both graphs, equal distribution of normalized H/L for all proteins around the normalized H/L=1 can be observed, confirming the accuracy of protein concentration measurement and that the samples were mixed sufficiently close to 1:1 ratio to be considered as mixed accurately.
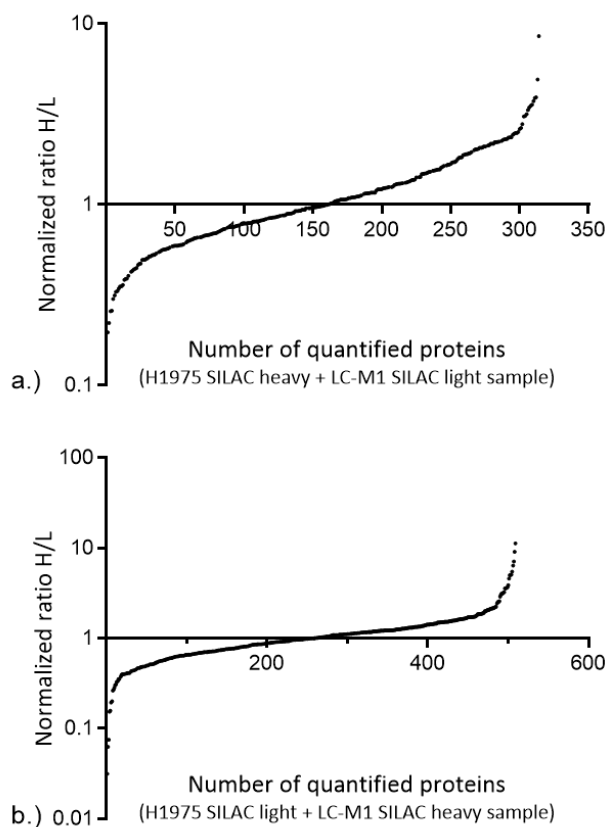
Figure 11. Distribution of normalized H/L of proteins detected in mixed SILAC heavy and light samples. n=3. a.) Normalized ratio H/L of analyzed proteins are presented on y-axis using logarithmic scale. The corresponding number of quantified proteins are presented on x-axis. Analyzed proteins correspond to the bands marked in Figure 10a of 1:1 mixed H1975 SILAC heavy & LC-M1 SILAC light sample according to the BCA test. The saddle point of the graph is at a normalized H/L=0.98168, indicating that proteins extracted from H1975 heavy and LC-M1 light samples were mixed accurately. b.) Normalized ratio H/L of analyzed proteins are presented on y-axis using logarithmic scale. The corresponding number of quantified proteins are presented on x-axis. Analyzed proteins correspond to the bands marked in Figure 10b of 1:1 mixed H1975 SILAC light & LC-M1 SILAC heavy sample according to the BCA test. The saddle point of the graph is at normalized H/L=0.99743, indicating that proteins extracted from H1975 heavy and LC-M1 light samples were mixed accurately.

5.3.3 Identification of biomarker candidates for lung cancer detection and validation of MS/MS data

The shotgun proteomics approach was used in this thesis for differential proteome analysis and the identification of potential epithelial and mesenchymal biomarkers for lung cancer detection. To increase the detection rate of analysed proteins, samples were fractionated using OFFGEL fractionator (8.3.9 M&M section). Fractionated peptides were analyzed by LC/MS-MS (8.3.10 M&M section), and identification and quantification of proteins from all three biological replicates analysed for each set of experiments was performed using MaxQuant software (8.7.1 M&M section). Trypsin was used for protein digestion, and maximum of two missed cleavages was allowed. Human proteins were identified using FASTA database and peptides matching the experimental peptide precursor were computed. Such analysis of the raw data files for experiment comprised of H1975 heavy-labelled- and LC-M1 light-labelled-derived proteins yielded in identification and quantification of 8,181 proteins, from which potential epithelial biomarkers for detection of lung cancer were identified. The analysis of raw data files for experiment conducted on H1975 light-labelled- and LC-M1 heavy-labelled-derived proteins yielded in identification and quantification of 8,147 proteins, from which potential mesenchymal biomarkers for the detection of lung cancer were identified. In both set of experiments, proteins with normalized ratio H/L$\geq$2 were considered as differentially expressed, and proteins were considered as overexpressed only if normalized ratio H/L$\geq$2 was for heavy-labelled proteins.

From the experiment where H1975 SILAC heavy & LC-M1 SILAC light proteins were analysed, out of 8,181 identified and quantified proteins, 1,929 proteins had normalized H/L$\geq$2 or normalized H/L=non-defined (n.def.). Within these 1,929 proteins, all proteins detected as contaminants were excluded, and only proteins with 4$\geq$unique peptides detected were left for further analysis, leaving 826/8,181 proteins for further analysis. 696/826 proteins had normalized H/L between 2.0002 and 97.15, whereas 130/826 proteins had normalized H/L=n.def. Among proteins with normalized H/L=n.def., only those proteins for which the intensity for light signal was 0 and the intensity for heavy signal was$\neq$0, were considered as overexpressed, leaving 79/130 proteins with normalized H/L=n.def. considered as overexpressed in H1975 compared to LC-M1. Taken together 905/8181 identified and quantified proteins were considered as overexpressed in H1975 compared to LC-M1. After literature and database search, NECT-4, SwissProt accession number Q96NY8, and PDAP1, SwissProt accession number Q13442 were chosen as potential biomarkers for detection of lung cancer. NECT-4 was detected only on H1975 cell line proteome with normalized H/L=n.def, whereas PDAP1 was detected in both cell line proteomes with normalized H/L=2.2781. Summary of LC/MS-MS results for NECT-4 and PDAP1 is listed in Table 1. MS1 and corresponding MS2 spectra of detected peptide for detection of NECT-4 and PDAP1 are exemplary presented below in Figures 12 and 13, and corresponding tables for b-and y-ion fragments masses are listed in Supplementary Tables 1 and 2 (Table S1 and Table S2).

Table 1. LC-MS/MS result summary of validation proteins detected in SILAC H1975 heavy & LC-M1 light sample.

Swiss-Prot acc no.	(UniProtKB/ Swiss-Prot) (short name)	Total number of peptides analyzed[a]	Number of unique peptides analyzed[b]	Number of biological replicates	Average of normalizied ratio H/L	Standard deviation	*p*-value[c]
Q96NY8	NECTIN-4	13	7	3	Only in H1975 heavy	n. def.	n. def.
Q13442	PDAP1	18	6	3	2.24065	0.2063	$2.7954*10^{-7}$
O00622	CCN1	50	21	3	0.19109	0.1149	$2.7732*10^{-11}$

[a] all analysed peptides identified for a protein whether the peptide is unique or not unique. Peptide can be counted once per each fraction and replicate, meaning it can be counted multiple times in total

[b] number of peptides unique for given protein. All peptides counted are different

[c] Student's *t*-test

n. def.=non-defined

Figure 12a shows the MS1 spectra of SAAVTSEFHLVPSR peptide from NECT-4 protein detected in H1975 SILAC heavy & LC-M1 SILAC light sample. The light peptide was not detected, meaning that NECT-4 was detected only in H1975 cell line, and its corresponding heavy peptide was detected at *m/z* 502.9388 [M + 3 H]$^{3+}$. Corresponding MS2 spectra of SAAVTSEFHLVPSR peptide fragmentation is shown in Figure 12b.
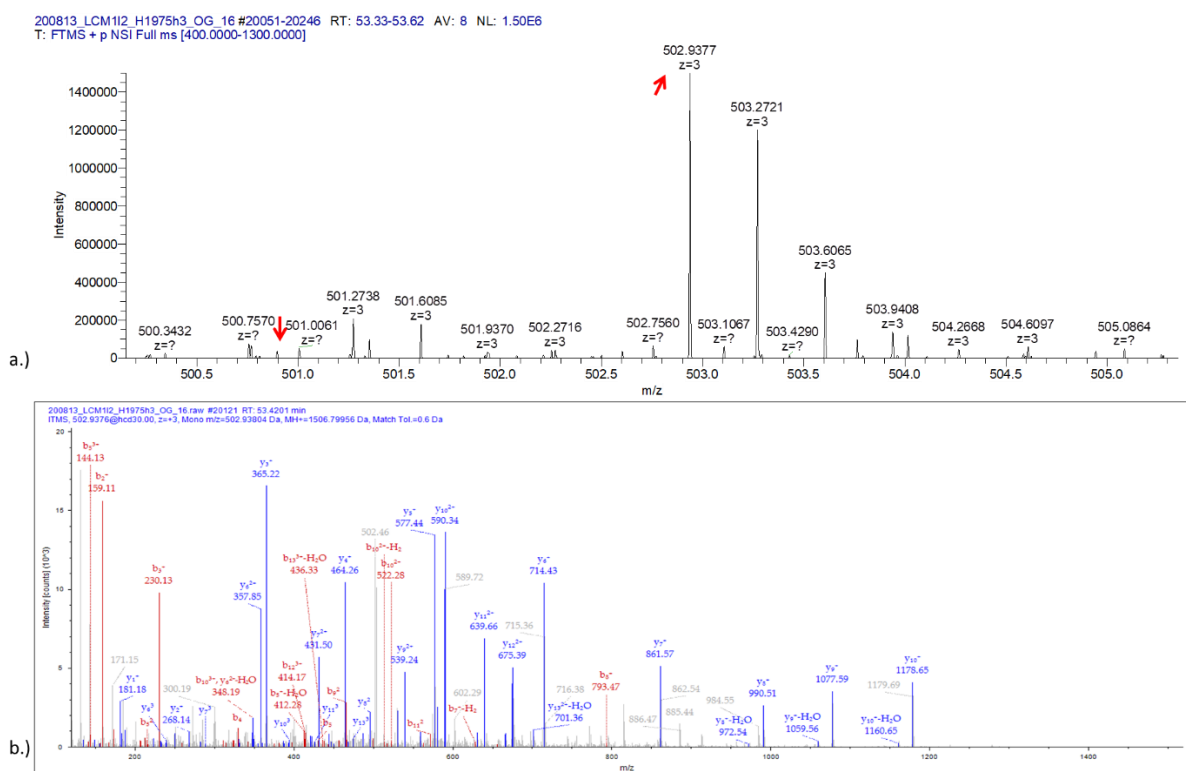
Figure 12. MS1 and MS2 spectra of SAAVTSEFHLVPSR peptide from NECT-4 protein detected in H1975 SILAC heavy & LC-M1 SILAC light sample. a.) MS1 spectra of SAAVTSEFHLVPSR peptide

from NECT-4 protein detected in H1975 SILAC heavy & LC-M1 SILAC light sample (red arrow). b.) MS2 spectra of heavy SAAVTSEFHLVPSR peptide from NECT-4 protein detected in H1975 SILAC heavy & LC-M1 SILAC light sample.

Figure 13a shows the MS1 spectra QYTSPEEIDAQLQAEK peptide from PDAP1 protein detected in H1975 SILAC heavy & LC-M1 SILAC light sample. Light peptide was detected at m/z 925.4436 [M + 2 H]$^{2+}$, and its corresponding heavy peptide at m/z 928.4539 [M + 2 H]$^{2+}$. Corresponding MS2 spectra of QYTSPEEIDAQLQAEK peptide fragmentation is shown in Figure 13b.
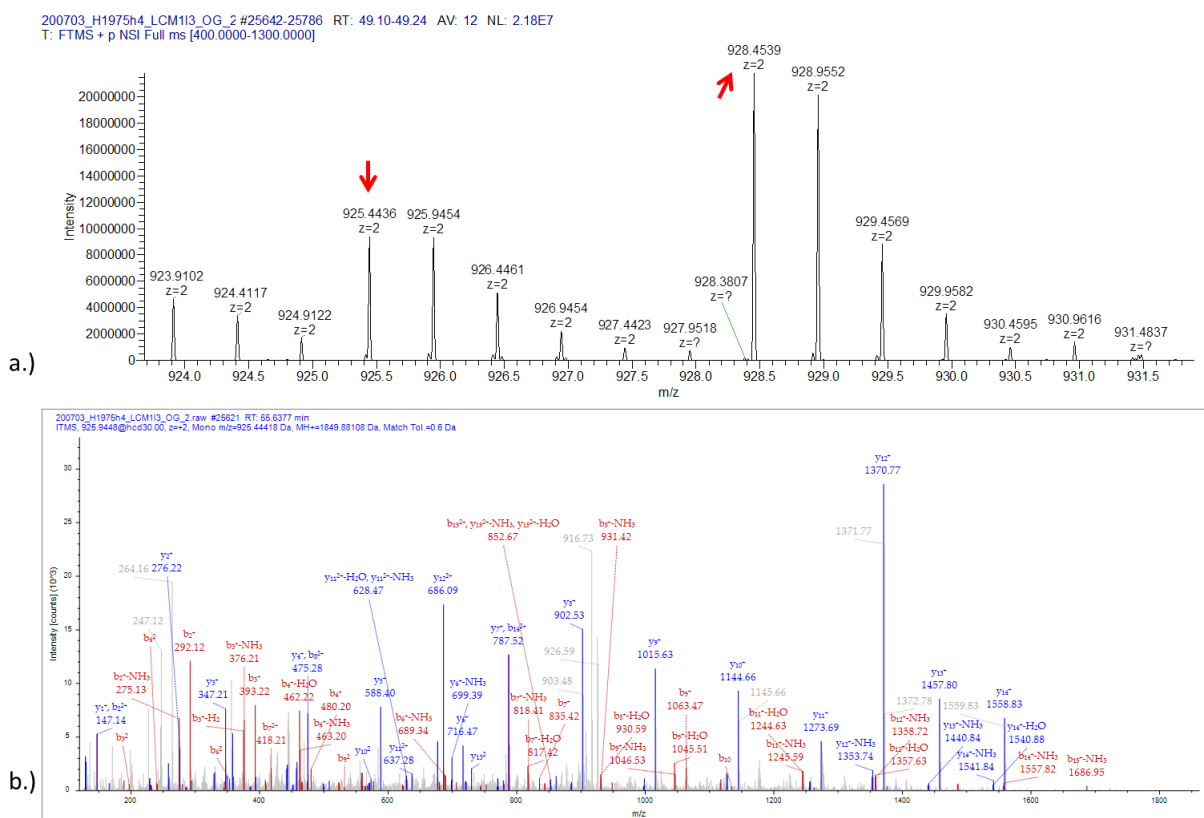
Figure 13. MS1 and MS2 spectra of QYTSPEEIDAQLQAEK peptide from PDAP1 protein detected in H1975 heavy & LC-M1light SILAC sample. a.) MS1 spectra of QYTSPEEIDAQLQAEK peptide from PDAP1 protein detected in H1975 heavy & LC-M1 light SILAC sample. b.) MS2 spectrum of light QYTSPEEIDAQLQAEK peptide from PDAP1 protein detected in H1975 heavy & LC-M1 light SILAC sample.

From the experiment where H1975 SILAC light & LC-M1 SILAC heavy proteins were analysed, out of 8,147 identified and quantified proteins, 1706 proteins had normalized H/L≥2 or normalized H/L=n.def. Within these 1,706 proteins, all proteins detected as contaminants were excluded, and only proteins with 4≥unique peptides detected were left for further analysis, leaving 677/8,147 proteins for further analysis. 443/677 proteins had normalized H/L between 2.0031 and 27.118, whereas 234/677 proteins had normalized H/L=n.def. Among proteins with normalized H/L=n.def., only those proteins for which the

intensity for light signal was 0 and the intensity for heavy signal was $\neq 0$, were considered as overexpressed, leaving 41/234 proteins with normalized H/L=n.def. considered as overexpressed in LC-M1 compared to H1975. Taken together, 484/8,147 identified and quantified proteins were considered as overexpressed in LC-M1 compared to H1975. After literature and database search, CCN family member 1 (CCN1), SwissProt accession number O00622 was chosen as potential mesenchymal biomarker for lung cancer detection. CCN1 was detected in both cell line proteomes with normalized H/L=7.18339. Summary of LC/MS-MS results for CCN1 is listed in Table 2. MS1 and corresponding MS2 spectra of detected peptide for detection of CCN1 is exemplary presented below in Figure 14, and corresponding table for b-and y-ion fragments masses is listed in Supplementary Table 3 (Table S3).

Table 2. LC-MS/MS result summary of validation proteins detected in SILAC H1975 light & LC-M1 heavy sample.

Swiss-Prot acc no.	(UniProtKB/ Swiss-Prot) (short name)	Total number of peptides analyzed[a]	Number of unique peptides analyzed[b]	Number of biological replicates	Average of normalizied ratio H/L	Standard deviation	*p*-value[c]
Q96NY8	NECTIN-4	19	10	3	Only in H1975 light	n. def.	n. def
Q13442	PDAP1	22	7	3	0.47115	0.0611	$8.0123*10^{-5}$
O00622	CCN1	42	19	3	7.18339	6.1906	$1.4414*10^{-7}$

[a] all analysed peptides identified for a protein whether the peptide is unique or not unique. Peptide can be counted once per each fraction and replicate, meaning it can be counted multiple times in total

[b] number of peptides unique for given protein. All peptides counted are different

[c] Student's *t*-test

n. def.=non-defined

Figure 14a shows MS1 spectra of FTYAGCLSVK peptide from CCN1 protein detected in H1975 SILAC light & LC-M1 SILAC heavy sample. Light peptide was detected at m/z 573.2865 $[M + 2 H]^{2+}$, and its corresponding heavy peptide at *m/z* 576.2965 $[M + 2 H]^{2+}$. Corresponding MS2 spectra of FTYAGCLSVK peptide fragmentation is shown in Figure 14b.
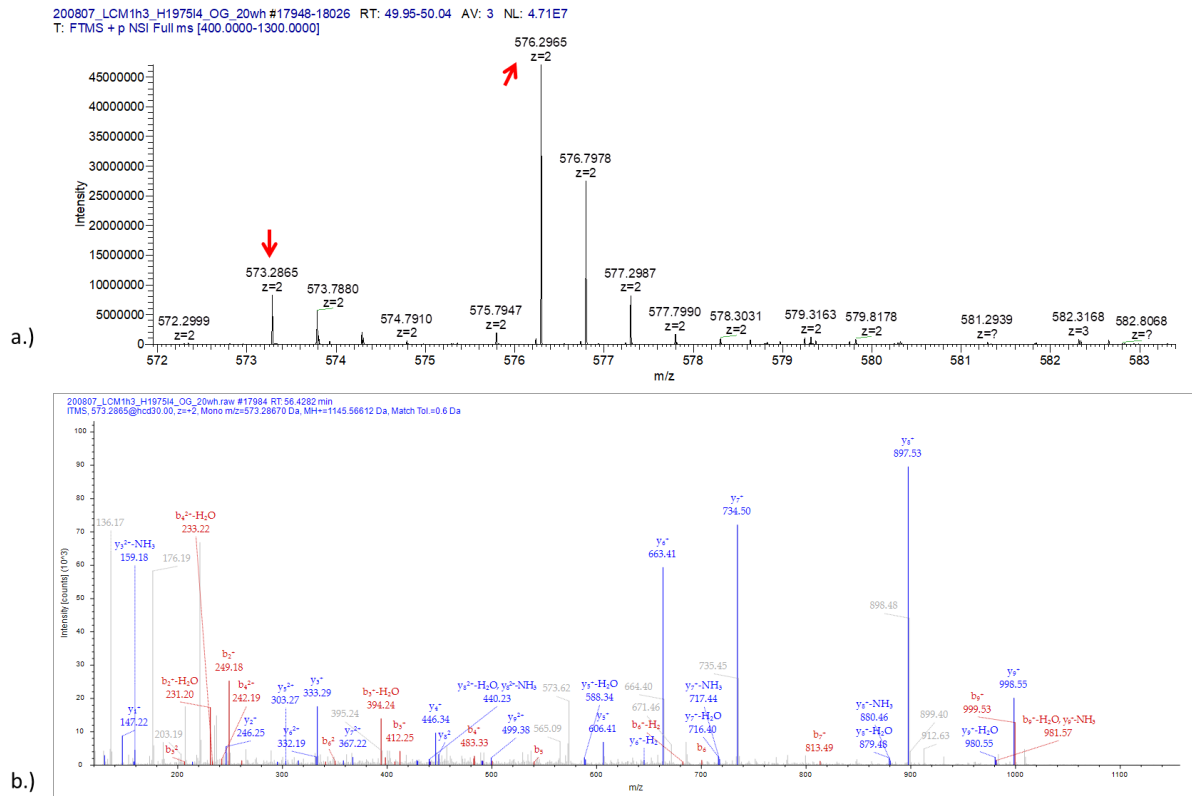
Figure 14. MS1 and MS2 spectra of FTYAGCLSVK peptide from CCN1 protein detected in H1975 light & LC-M1 heavy SILAC sample. a.) MS1 spectra of FTYAGCLSVK peptide from CCN1 protein detected in H1975 light & LC-M1 heavy SILAC sample. b.) MS2 spectra of light FTYAGCLSVK peptide from CCN1 protein detected in H1975 light & LC-M1 heavy SILAC sample.

The summary of LC/MS-MS results for CCN1 analyzed in sample H1975 SILAC heavy & LC-M1 SILAC light is listed in Table 1 (above in this section), as a confirmation of LC/MS-MS results for CCN1 from sample H1975 SILAC light & LC-M1 SILAC heavy. The summaries of LC/MS-MS results for proteins NECT-4 and PDAP1 detected in sample H1975 SILAC light & LC-M1 SILAC heavy are listed in Table 2 (above in this section), as a confirmation of LC/MS-MS results for NECT-4 and PDAP1 from sample H1975 SILAC heavy & LC-M1 SILAC light.

5.4 Validation of LC-MS/MS data for identified lung cancer biomarker candidates by *Western blot* and by ELISA

Western blot protocol was optimized for validation of lung cancer biomarker candidates identified by LC-MS/MS, NECT-4, PDAP1, and CCN1, respectively. Additional validation of NECT-4 and PDAP1 was performed using ELISA. CCN1 was previously validated as potential biomarker for lung cancer detection in men by our group (75). Therefore, CCN1 was interesting in this thesis as potential mesenchymal biomarker for detection of circulating tumor cells in lung cancer patients.

cNECT-4 is a type-I transmembrane protein, composed of extracellular domain with glycosylation site, transmembrane, and cytoplasmic domain, with the predicted molecular weight of 55.454 kDa without posttranslational modification (110). *Western blot* analysis revealed a band at 72 kDa which could be a glycosylation form of full-length-NECT-4 and the band around 60 kDa which could be non-glycosylated form of full-length-NECT-4 (Figure 15a and b). Two additional bands were detected at positions of 45 kDa and around 36 kDa which could correspond to the cleaved glycosylated extracellular domain and the cleaved non-glycosylated extracellular domain (Figure 15). Band of an unknown origin at the position slightly below 55 kDa can be observed and it was considered as non-specific binding (Figure 15). No band for NECT-4 was detected in HCC-366, HTB-58, H1299, H1395 and LC-M1 cell lines, whereas H1975 and H1993 cell lines were positive for all above mentioned forms of NECT-4. For H1975 cell line the strongest band detection was for cleaved and non-glycosylated NECT-4, while for H1993 the strongest band detection was for full length NECT-4, both glycosylated and non-glycosylated (Figure 15a). Detection of NECT-4 in H1975 and the lack of detection in LC-M1 cell lines is in accordance with the MS data (Table 1 and Table 2). Additionally, PBMCs isolated from healthy donors (n=5) were analysed for NECT-4 presence, and *Western blot* analysis revealed lack of detection of NECT-4 on PBMCs for all five healthy donors (Figure 15b). MDA-231 served as negative control and MCF-7 as positive control (Figure 15).

Figure 15. Western blot analysis of NECT-4 as a biomarker candidate for the detection of lung cancer. a.) Analysis of NECT-4 presence in MDA-231, MCF-7, HCC-366, HTB-58, H1975, H1299, H1395, H1993 and LC-M1 cell lines. α-tubulin was used as loading control. Twenty micrograms of total protein were loaded. n=3. b.) Analysis of NECT-4 presence in MDA-231, MCF-7, and PBMCs isolated from five healthy donors. α-tubulin was used as loading control. Fifty micrograms of total protein were loaded. PBMC=peripheral blood mononuclear cells, MW=molecular weight, kDa=kilodalton.

Another candidate for detection of lung cancer identified by mass spectrometry analysis was PDAP1, a secreted-28-kDa protein. *Western blot* analysis revealed a band at 28 kDa position which corresponds to the detection of PDAP1 (Figure 16). All cell lines analyzed, MDA-231, MCF-7, HCC-366, HTB-58, H1975, H1299, H1395, H1993 and LC-M1 were positive for PDAP1, with the strongest band intensity detected for H1993 cell line (Figure 16a). H1975 and LC-M1 *Western blot* results are in accordance with the mass spectrometry data (Table 1 and 2). Additionally, *Western blot* analysis on PBMCs from five healthy donors was performed and PDAP1 was not detected on healthy-donor-PBMCs, while MDA-231 and MCF-7 served as positive controls (Figure 16b).
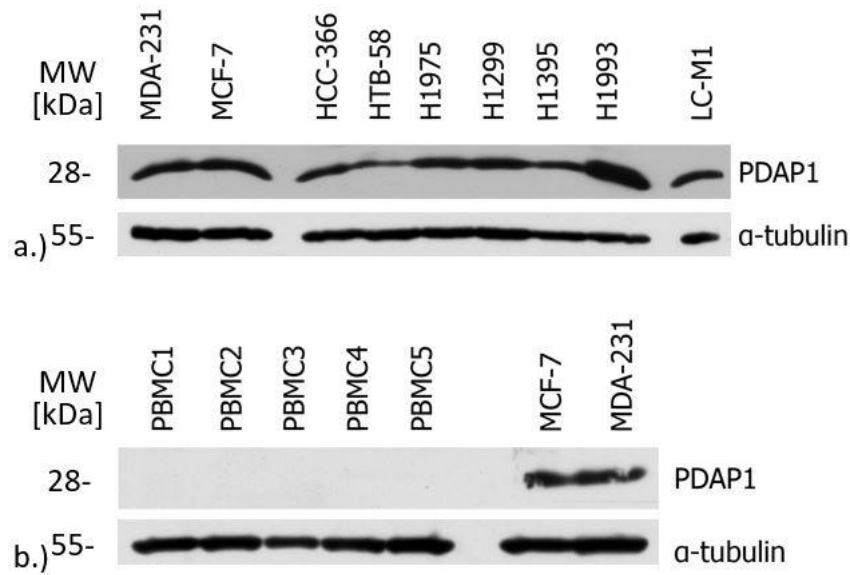
Figure 16. Western blot analysis of PDAP1 as lung cancer detection biomarker candidate. a.) Analysis of PDAP1 presence in MDA-231, MCF-7, HCC-366, HTB-58, H1975, H1299, H1395, H1993 and LC-M1 cell lines. α-tubulin was used as loading control. Twenty micrograms of total protein were loaded. n=3. b.) Analysis of PDAP1 presence in MDA-231, MCF-7, and PBMCs isolated from five healthy donors. A-tubulin was used as loading control. Fifty micrograms of total protein were loaded. PBMC=peripheral blood mononuclear cells, MW=molecular weight, kDa=kilodalton, n=biological replica.

The third candidate for detection of lung cancer identified by mass spectrometry analysis was CCN1; a secreted and 42 kDa protein. *Western blot* analysis revealed band around 40 kDa position which corresponds to the detection of CCN1 (Figure 17). HCC-366, HTB-58, H1975, H1993 and LC-M1 were positive for CCN1, with the strongest band intensity detected for HCC-366 and LC-M1 cell lines (Figure 17a). H1975 and LC-M1 *Western blot* results are in accordance with the mass spectrometry data (Table 1 and 2). Additionally, PBMCs isolated from healthy donors (n=5) were analysed for the CCN1 presence. *Western blot* analysis revealed lack of the CCN1 detection on PBMCs for all five healthy donors (Figure 17b). MDA-231 served as positive control and MCF-7 as negative control (Figure 17).
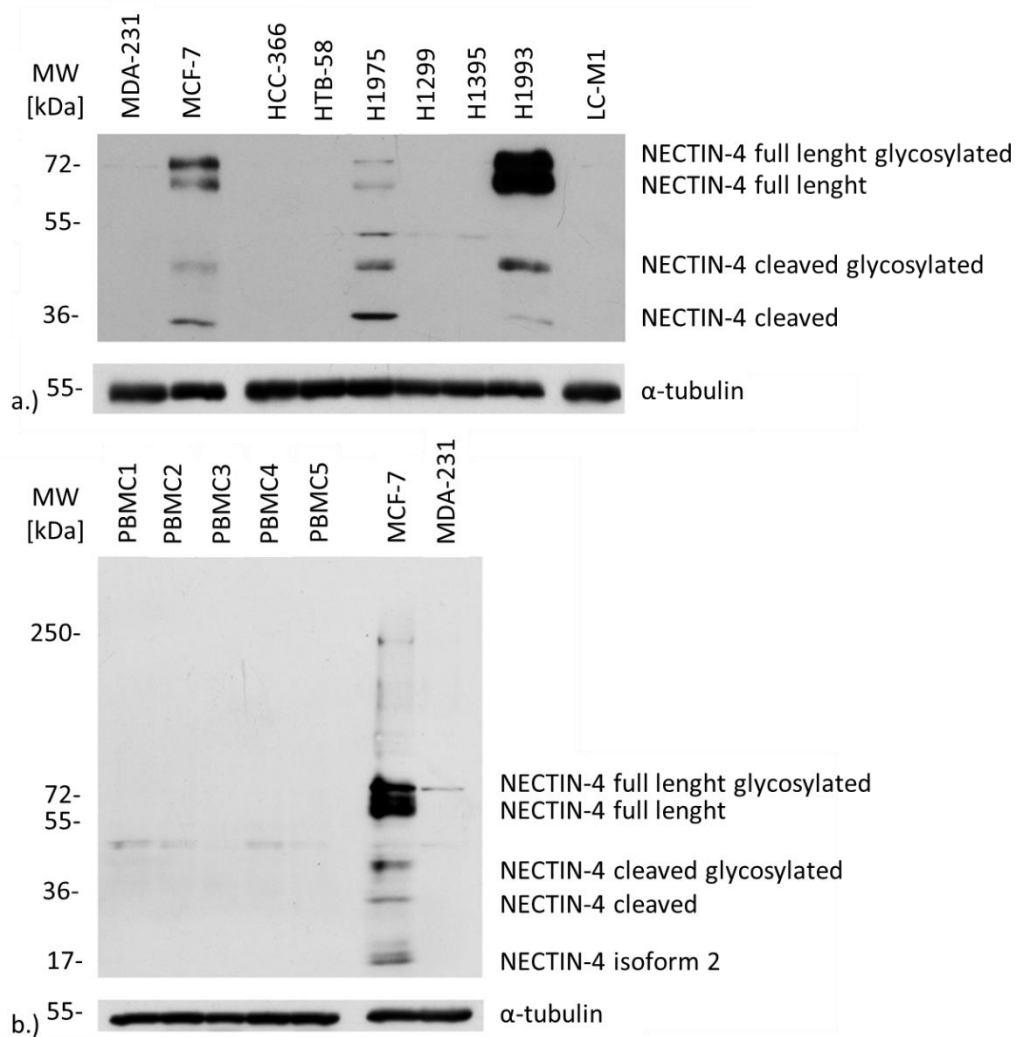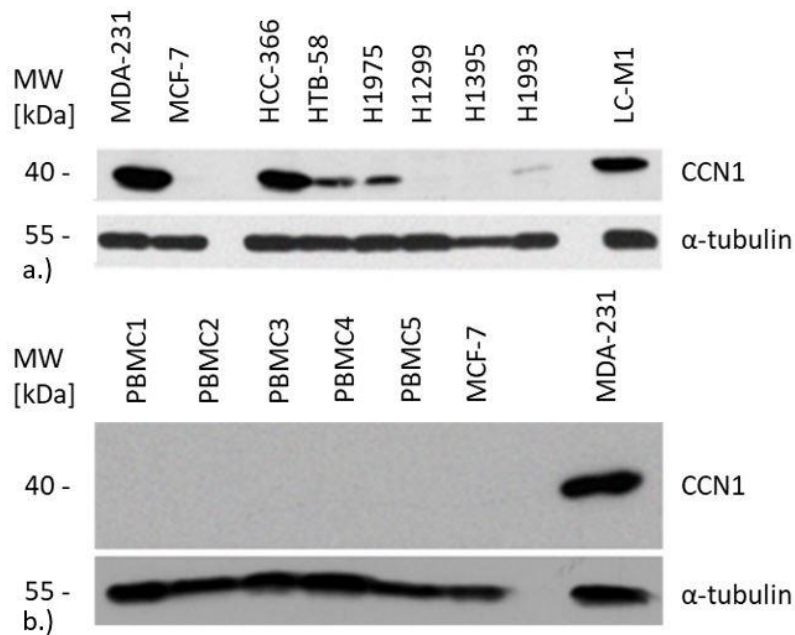
Figure 17. Western blot analysis of CCN1 as a biomarker candidate for lung cancer detection. a.) Analysis of CCN1 presence in MDA-231, MCF-7, HCC-366, HTB-58, H1975, H1299, H1395, H1993 and LC-M1 cell lines. α-tubulin was used as loading control. Twenty micrograms of total protein were loaded. n=3. b.) Analysis of CCN1 presence in MDA-231, MCF-7, and PBMCs isolated from five healthy donors. α-tubulin was used as loading control. Fifty micrograms of total protein were loaded. PBMC=peripheral blood mononuclear cells, MW=molecular weight, kDa=kilodalton, n=biological replica.

As mentioned above, the CCN1 was sought to be investigated in this thesis as a potential mesenchymal biomarker for CTC detection. However, the collection of patient samples has been challenging and limited due to the start of COVID-19 pandemic. Therefore, the investigation of CCN1 as potential biomarker for the CTC detection has not been processed further in this thesis. The further analysis of NECT-4 and PDAP1 was allowed due to the acquirement of plasma samples previously stored in the biobank and kindly provided by Dr. Georg Johnen (Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Germany).

Further on, biomarker candidates NECT-4 and PDAP1 identified by LC-MS/MS were validated by ELISA. Concentrations of NECT-4 have been measured according to the manufacturer's instruction in 50 µL of cell culture supernatants in biological triplicates (n=3). NECT-4 was detected in H1975 lung cancer cell line and not detected in LC-M1 lung cancer cell line, which is in line with mass spectrometry and *Western blot* results. In addition, concentrations of NECT-4 have been measured in MCF-7 and MDA-231 cell culture supernatants (n=3). NECT-4 was detected in MCF-7 and not detected in MDA-231, which is in line with western blot results (for cellular NECT-4) (Figure 18).
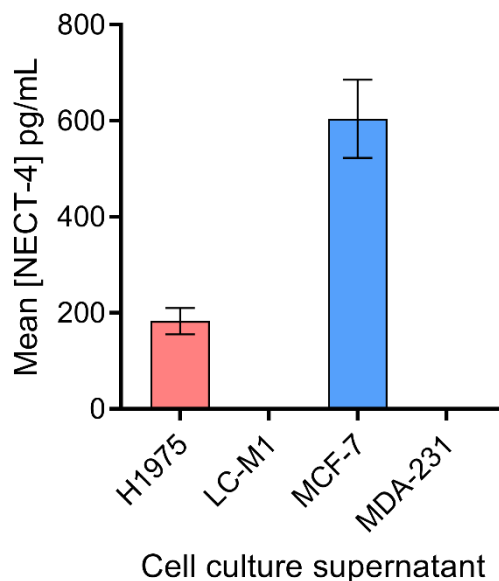
Figure 18. Detection of NECT-4 in lung cancer and breast cancer cell culture supernatants by ELISA. Concentrations of NECT-4 are presented on y-axis, whereas the cell lines from which cell culture supernatant derived are presented on x-axis. Error bars represent standard deviation. n=3. n=biological triplicates.

To cross check technical accuracy of NECT-4 ELISA kit, NECT-4 concentrations were measured in cell line lysates in technical triplicates (n=3). NECT-4 has been detected in H1975 and MCF-7 cell line lysates, whereas the NECT-4 concentrations in LC-M1 and MDA-231 were below detection limit of the assay (below 39.5 pg/mL) which is in line with previous results (Figure 19). Coefficient of variance (CV) in NECT-4 positive cell lines, H1975 and MCF-7 has been calculated as well resulting in CV=4.35% confirming the reproducibility of the assay.
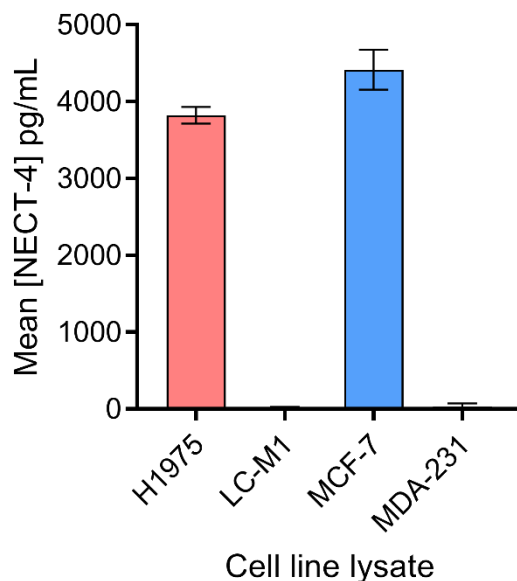
Figure 19. Detection of cellular NECT-4 in lung cancer and breast cancer cell line lysates by ELISA. Concentrations of NECT-4 are presented on y-axis, whereas the cell lines from which lysates derived are presented on x-axis. Error bars represent standard deviation. n=3. n=technical triplicates.

Further on, concentrations of PDAP1 have been measured according to the manufacturer's instruction in 100 µL of cell culture supernatants and cell line lysates in biological triplicates (n=3). This revealed the presence of PDAP1 in H1975 and in LC-M1, which is in line with mass spectrometry and *Western blot* results (Figure 20). In the experiment performed on cell line lysates, the difference in PDAP1 levels between H1975 and LC-M1 cell lines is in line with mass spectrometry and *Western blot* results as well, showing differential expression with a fold change of 1.38 (Figure 20a). However, the differential expression between H1975 and LC-M1 cell lines could not been observed from the experiment performed on cell culture supernatants (Figure 20b.). Additionally, PDAP concentration was analysed in MCF-7 and MDA-231 breast cancer cell culture supernatant and cell line lysates using ELISA. Both cell lines were positive for PDAP1 and ELISA results are in line with *Western blot* results (Figure 20).
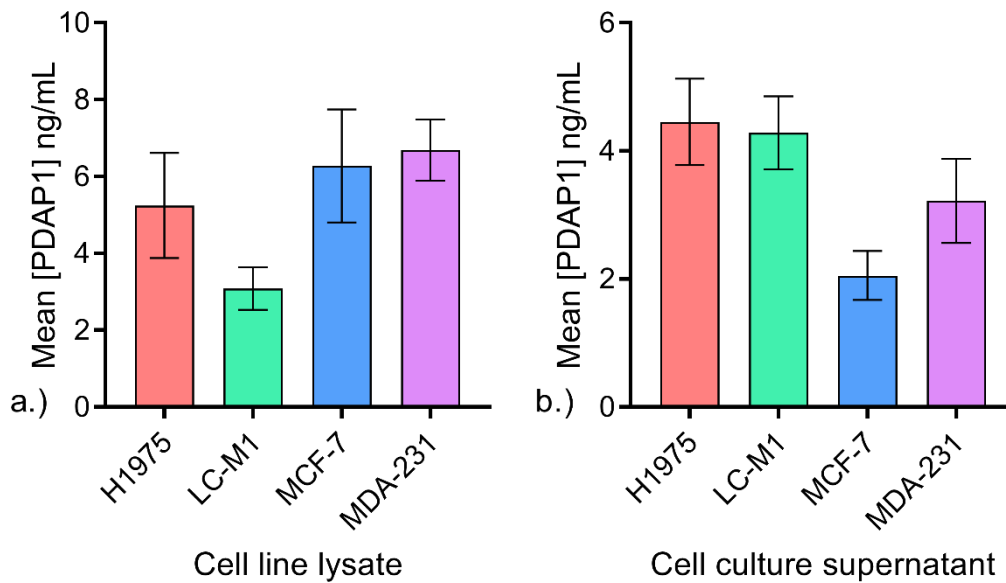
Figure 20. Detection of PDAP1 in lung cancer and breast cancer cell culture supernatants and cell line lysates by ELISA. a.) Concentrations of PDAP1 are presented on y-axis, whereas the cell lines from which cell line lysate derived are presented on x-axis b.) Concentrations of PDAP1 are presented on y-axis, whereas the cell lines from which cell line supernatants derived are presented on x-axis. In both graphs error bars represent standard deviation. n=3. n=biological triplicates.

In order to cross check technical accuracy of PDAP1 ELISA kit, the concentrations of PDAP1 were measured in cell culture supernatants in technical triplicates (n=3). H1975, LC-M1, MCF-7, and MDA-231 supernatants were positive for PDAP1, which is in line with previous results. However, the standard deviation within each sample was rather high, and the calculated coefficient of variance (CV=40.54%) was not in line with the manufacturer's instruction (Figure 21), doubting the reproducibility of the assay.
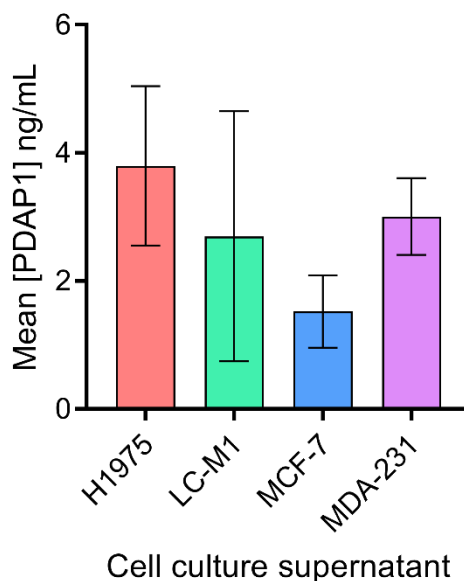
Figure 21. Detection of PDAP1 in lung cancer and breast cancer cell culture supernatants by ELISA. Concentrations of PDAP1 are presented on y-axis, whereas the cell lines from which cell culture supernatants derived are presented on x-axis. Error bars represent standard deviation. n=3. n=technical replicas.

To improve the reproducibility of the assay, manufacturer´s instruction has been modified as followed: instead of 100 µL of a sample, 50 µL of the sample were loaded into a well in which 50 µL of diluent has already been pipetted, allowing the resuspension of the sample within the pipet tip (see 8.3.11 M&M section). Optimization of the experiment was performed using healthy donor- and lung cancer-derived plasma samples. With the mentioned modification, the accuracy of pipetting, and therefore the reproducibility of the assay has been significantly improved resulting in the reduction of standard deviation as well as the achieved coefficient of variance (CV=2.53%) (Figure 22).
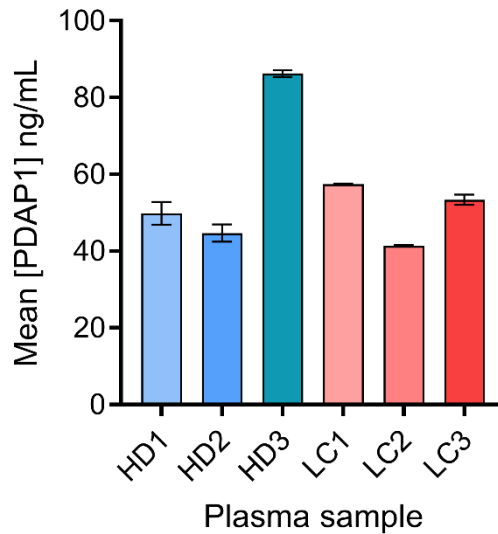
Figure 22. Detection of PDAP1 in healthy control- and lung cancer-derived plasma by ELISA. Concentrations of PDAP1 are presented on y-axis, whereas plasma samples of healthy donors and lung cancer patients that have been analysed are presented on x-axis. Error bars represent standard deviation. n=3. HD=healthy donor, LC=lung cancer patient, n=technical triplicates.

5.5. Characteristics of the clinical study and correlation of NECT-4 and PDAP1 concentrations with clinical parameters

To investigate usability of NECT-4 and PDAP1 as biomarkers for lung cancer detection and their potential for implementation into the clinical settings, optimized ELISA assays were used to measure NECT-4 and PDAP1 concentrations in plasma of lung cancer patients and healthy controls.

5.5.1 Clinical study characteristics

Investigated study group included 113 male and 80 female lung cancer patients, comprising adenocarcinoma (n=81), squamous cell carcinoma (n=48), small cell lung cancer (n=40), large cell lung cancer (n=12), non-small cell lung cancer (n=10), and other (n=2) histological subtypes, resulting in total of 193 analyzed lung cancer patients. In parallel, 200 healthy controls, comprised of 171 male and 29 female individuals, were analyzed. Distribution of NECT-4 and PDAP1 concentrations across study cohort is presented in Figure 23. One patient was described as T0-stage and for statistical purposes was excluded from statistical analysis of association of potential biomarkers NECT-4 and PDAP1 with T-stages of lung cancer. Otherwise, no other sample was excluded from any analysis.
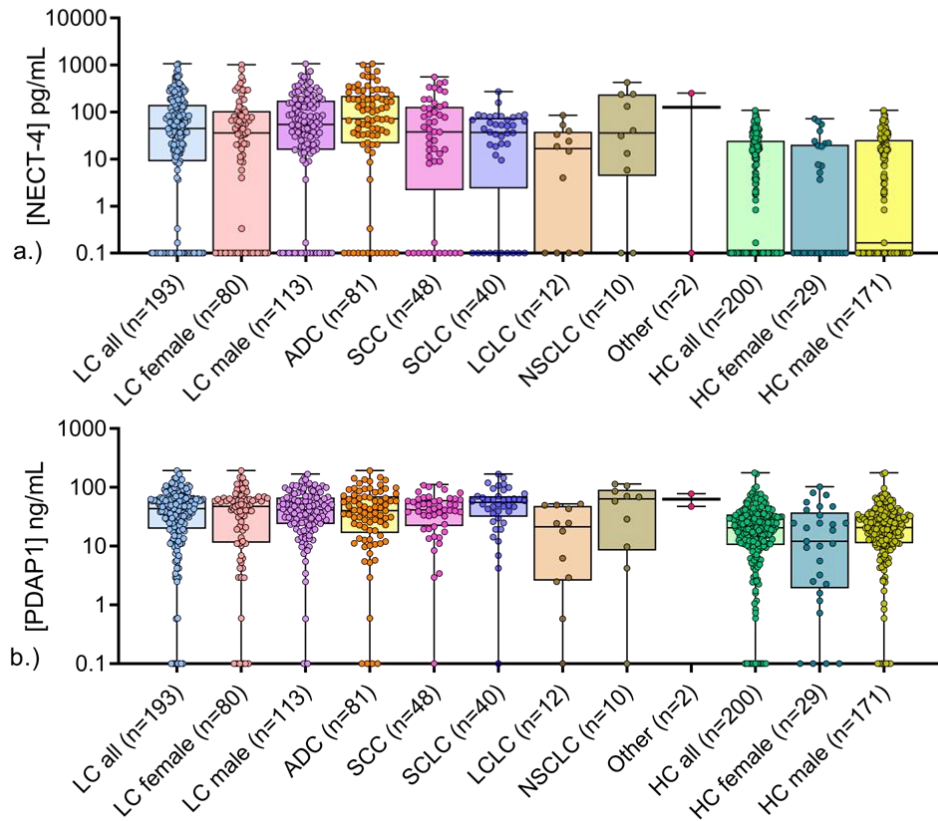
Figure 23. Distribution of NECT-4 and PDAP1 in plasma of all lung cancer patients and all healthy controls and after stratification according to histological subtype and sex. Protein concentration is presented on y-axis, whereas the analysed group is presented on x-axis. Median and the interquartile range are displayed as horizontal bars. Minimum and the maximum of protein concentration is represented by whiskers. n is represented by a dot. a.) Distribution of NECT-4 concentrations in plasma of lung cancer patients and healthy controls and after stratification according to histological subtype and sex. b.) Distribution of PDAP1 concentrations in plasma of lung cancer patients and healthy controls and after stratification according to histological subtype and sex. n (LC all)=193, n (HC all)=200, n (LC female)=80, n (HC female)=29, n (LC male)=113, n (HC male)=171, n (ADC)=81, n (SCC)=48, n (SCLC)=40, n (LCLC)=12, n (NSCLC)=10, n (Others)=2. For the values of statistical analysis refer to the Table 3 and 4 (section 5.5.1 and 5.5.3). Protein concentration of n positioned directly on the x-axis is truthfully equal to zero. However, due to the logarithmic scale it appears to be 0.1 since it is mathematically impossible to plot a zero on logarithmic scale. Therefore, some of the boxplots are also dragged to the x-axis. n=biological individual analysed, LC=lung cancer, HC=healthy control, ADC=adenocarcinoma, SCC=squamous cell carcinoma, SCLC=small cell lung cancer, LCLC=large cell lung cancer, NSCLC=non-small cell lung cancer.

5.5.2 Correlation of NECT-4 with clinical parameters

NECT-4 concentrations are elevated in lung cancer patients (113.0±171.9 ng/mL) compared to healthy controls (14.1±21.4 ng/mL) ($p<0.001$). Analysis of clinico-pathological parameters across lung cancer patients revealed association of NECT-4 with histology (Figure 24, $p=0.005$), stage (Figure 25, $p=0.016$), and metastasis (Figure 26, $p=0.006$). Other parameters analyzed, smoking status ($p=0.690$), sex ($p=0.062$), T-stage ($p=0.211$) and nodal status ($p=0.612$) respectively, did not reveal any differences between compared groups (Table 3).

41

Table 3. Concentrations of NECT-4 in lung cancer patients and healthy controls and the analysis of lung cancer associated clinico-pathological parameters.

	n	[NECT-4] pg/mL [a]	*p*-value	Minimum [NECT-4] pg/mL	Maximum [NECT-4] pg/mL
Disease status			<0.001[b]		
Lung cancer	193	113.0 ± 171.9		0	1060.6
Healthy control	200	14.1 ± 21.4		0	109.3
Total	393	62.7 ± 131.0		0	1060.6
Smoking status			0.690[c]		
Smoker (lung cancer)	51	125.4 ± 192.4		0	1060.6
Ex-smoker (lung cancer)	126	110.3 ± 169.1		0	1012.8
Never smoker (lung cancer)	13	57.6 ± 58.0		0	182.3
Unknown (lung cancer)	3	254.7 ± 222.5		0	411.7
Sex			0.062[c]		
Male (lung cancer)	113	121.5 ± 169.1		0	1060.6
Female (lung cancer)	80	101.0 ± 176.3		0	1012.8
Histology			**0.005[c]**		
ADC	81	168.9 ± 222.8		0	1060.6
SCC	48	95.6 ± 133.1		0	556.0
SCLC	40	47.2 ± 51.7		0	273.8
LCLC	12	22.7 ± 26.5		0	85.3
NSCLC	10	113.0 ± 145.5		0	428.2
Other	2	125.9 ± 178.1		0	251.8
Stage			**0.016[c]**		
I	2	18.4 ± 3.9		15.6	21.2
II	10	59.0 ± 78.1		0	189.8
III	54	65.4 ± 101.5		0	556.0
IV	127	139.0 ± 195.6		0	1060.6
T-stage			0.211[c]		
T1	13	40.2 ± 55.5		0	183.3
T2	54	135.1 ± 225.8		0	1060.6
T3	54	100.7 ± 115.8		0	556.0
T4	71	120.5 ± 173.4		0	1012.8
Nodal status			0.612[c]		
N-negative	17	108.3 ± 250.0		0	1060.6
N-positive	176	113.5 ± 163.4		0	1012.8
Metastasis			**0.006[c]**		
M0	66	63.0 ± 96.5		0	556.0
M1a	34	131.4 ± 183.4		0	738.8
M1b	29	89.8 ± 107.8		0	366.8
M1c	64	165.3 ± 227.8		0	1060.6

[a] Arithmetic mean ± standard deviation

[b] Mann Whitney *U*-test

[c] Kruskal-Wallis test

Kruskal-Wallis analysis of significant clinico-pathological parameters specified above was followed by pairwise comparison using adjusted p-values by Bonferroni correction for multiple tests. Discrimination between adenocarcinoma (168.9±222.8 ng/mL) and small cell lung cancer (47.2±51.7 ng/mL) (Figure 24; adjusted p=0.035), and adenocarcinoma (168.9±222.8 ng/mL) and large cell lung cancer (22.7±26.5 ng/mL) (Figure 24; adjusted p=0.024) has been found.

Bonferroni correction for multiple tests can be used when multiple testing is performed and false positive result is increased. By applying Bonferroni correction the probability is adjusted (p adjusted=p adj.), and the risk of false positive result decreases (133).

Further on, pairwise comparison between different lung cancer stages and the stages of metastasis, showed discrimination between stage III (65.4±101.5 ng/mL) and IV (139.0±195.6 ng/mL) (Figure 25, adjusted p=0.032), and between M0 (63.0±96.5 ng/mL) and M1c (165.3±227.8 ng/mL) (Figure 26, adjusted p=0.005) based on NECT-4 concentrations between analyzed groups.
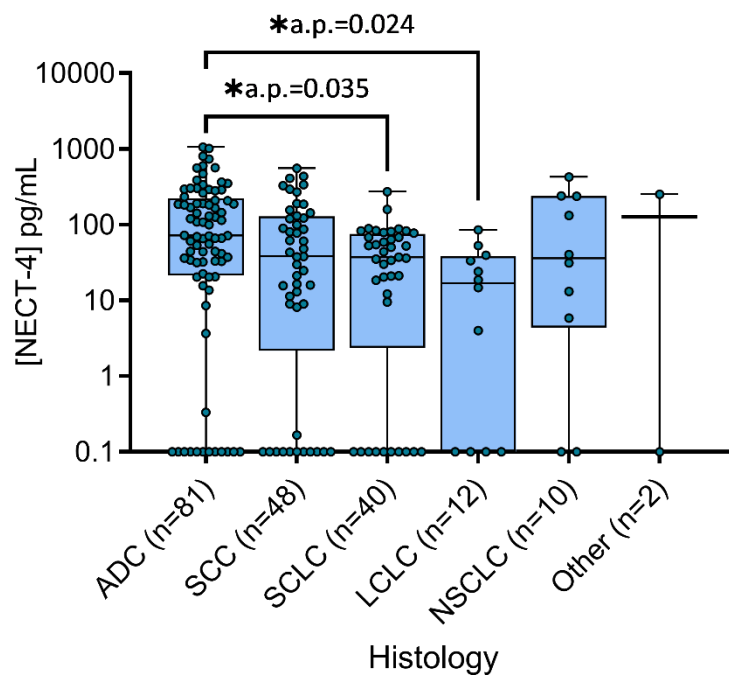
Figure 24. Association of NECT-4 with lung cancer histological subtypes. Concentrations of NECT-4 are presented on y-axis, whereas the lung cancer histological subtypes are presented on x-axis. Median and the interquartile range are displayed as horizontal bars. Minimum and the maximum of protein concentration are represented by whiskers. n is represented by a dot. Significant difference between lung cancer histological subtypes is indicated with connection lines and displayed as asterix (*). n (ADC)=81, n (SCC)=48, n (SCLC)=40, n (LCLC)=12, n (NSCLC)=10, n (Other)=2. Protein concentration of n positioned directly on the x-axis is truthfully equal to zero. However, due to the logarithmic scale it appears to be 0.1 since it is mathematically impossible to plot a zero on logarithmic scale. Therefore, some of the boxplots are also dragged to the x-axis. n=biological individual analysed, NECT-4=NECTIN-4, ADC=adenocarcinoma, SCC=squamous cell carcinoma, LCLC=large cell lung cancer,

Figure 25. Association of NECT-4 with lung cancer stages. Concentrations of NECT-4 are presented on y-axis, whereas the stages of lung cancer are presented on x-axis. Median and the interquartile range are displayed as horizontal bars. Minimum and the maximum of protein concentration are represented by whiskers. n is represented by dot. Significant difference between lung cancer stages is indicated with connection lines and displayed as asterix (*). Protein concentration of n positioned directly on the x-axis is truthfully equal to zero. However, due to the logarithmic scale it appears to be 0.1 since it is mathematically impossible to plot a zero on logarithmic scale. Therefore, some of the boxplots are also dragged to the x-axis. n (stage I)=2, n (stage II)=10, n (stage III)=54, n (stage IV)=127. n=biological individual analysed, NECT-4=NECTIN-4, I=stage I, II=stage II, III=stage III, IV=stage IV, a.p=adjusted *p*-value; *p*-values have been adjusted by the Bonferroni correction for multiple tests, pg/mL=picogram per milliliter.

Figure 26. Association of NECT-4 with lung cancer metastases. Concentrations of NECT-4 are presented on y-axis, whereas the stages of lung cancer metastasis are presented on x-axis. edian and the interquartile range are displayed as horizontal bars. Minimum and the maximum of protein concentration are represented by whiskers. n is represented by a dot. Significant difference between lung cancer metastatic stages is indicated with connection line and displayed as asterix (*). Protein concentration of n positioned directly on the x-axis is truthfully equal to zero. However, due to the logarithmic scale it appears to be 0.1 since it is mathematically impossible to plot a zero on logarithmic scale. Therefore, some of the boxplots are also dragged to the x-axis. n (stage M0)=66, n (stage M1a)=34, n (stage M1b)=29, n (stage M1c)=64. n=biological individual analysed, NECT-4=NECTIN-4, M0=stage M0, M1a=stage M1a, M1b=stage M1b, M1c=stage M1c, a. p.=adjusted *p*-value; *p*-values have been adjusted by the Bonferroni correction for multiple tests, M=metastasis, pg/mL=picogram per milliliter.

5.5.3 Correlation of PDAP1 with clinical parameters

PDAP1 concentrations are elevated in lung cancer patients (49.0±36.7 ng/mL) compared to healthy controls (26.2±24.9 ng/mL) (*p*<0.001). The analysis of clinico-pathological parameters across lung cancer patients revealed association of PDAP1 with histology (Figure 27, *p*=0.043), stage (Figure 28, *p*=0.012), T-stage (Figure 29, *p*=0.006) and metastasis (Figure 30, *p*=0.005). Across other parameters analyzed, smoking status (*p*=0.053), sex (*p*=0.796) and nodal status (*p*=0.192), respectively, no difference has been found between compared groups. However, the smoking status had marginal significance (*p*=0.053) (Table 4).

Table 4. Concentrations of PDAP1 in lung cancer patients and healthy controls and analysis of lung cancer associated clinico-pathological parameters.

	n	[PDAP1] ng/mL[a]	*p*-value	Minimum [PDAP1] ng/mL	Maximum [PDAP1] ng/mL
Disease status			<0.001[b]		
Lung cancer	193	49.0 ± 36.7		0	192.8
Healthy control	200	26.2 ± 24.9		0	177.4
Total	393	37.4 ± 33.3		0	192.8
Smoking status			0.053[c]		
Smoker (lung cancer)	51	54.6 ± 35.2		0	148.5
Ex-smoker (lung cancer	126	48.8 ± 37.8		0	192.8
Never smoker (lung cancer)	13	27.5 ± 28.3		0	94.0
Unknown (lung cancer)	3	55.2 ± 17.2		35.3	66.1
Sex			0.796[c]		
Male lung cancer	113	49.3 ± 34.7		0	169.0
Female lung cancer	80	48.7 ± 39.7		0	192.8
Histology			0.043[c]		
ADC	81	49.6 ± 41.0		0	192.8
SCC	48	45.2 ± 27.3		0	112.1
SCLC	40	58.2 ± 37.7		0	169.0
LCLC	12	22.9 ± 21.1		0	52.2
NSCLC	10	54.6 ± 42.0		0	113.9
Other	2	62.7 ± 21.7		47.4	78.0
Stage			0.012[c]		
I	2	27.9 ± 27.6		8.4	47.5
II	10	28.2 ± 20.1		0	66.5
III	54	58.1 ± 32.6		0	146.0
IV	127	47.1 ± 38.6		0	192.8
T-stage			0.006[c]		
T1	13	56.8 ± 35.6		4.2	108.7
T2	54	35.1 ± 29.4		0	118.0
T3	54	54.3 ± 35.9		0.6	169.0
T4	71	54.7 ± 40.2		0	192.8
Nodal status			0.192[c]		
N-negative	17	39.1 ± 31.8		0.6	110.9
N-positive	176	50.0 ± 37.1		0	192.8
Metastasis			0.005[c]		
M0	66	52.7 ± 32.7		0	146.0
M1a	34	51.4 ± 37.2		0	148.5
M1b	29	28.6 ± 26.0		0	106.5
M1c	64	53.2 ± 41.9		0	192.8

[a] Arithmetic mean ± standard

[b] Mann Whitney *U*-test

[c] Kruskal-Wallis test

Kruskal-Wallis analysis of significant clinico-pathological parameters specified above was followed by pairwise comparison using adjusted p-values by Bonferroni correction for multiple tests. Discrimination between large cell lung cancer (22.9±21.1 ng/mL) and small cell lung cancer (58.2±37.7 ng/mL) (Figure 27; adjusted p=0.021) has been revealed. Even though discrimination between different lung cancer stages has initially been found (see Table 4), after the application of Bonferroni correction for multiple tests, adjusted p-values were marginal, p=0.053 for comparison of stage II and stage III and p=0.058 for comparison of stage III and stage IV, respectively. However, they showed no discrimination between lung cancer stages (Figure 28). Pairwise comparison between different T-stages of lung cancer and the presence of metastasis, showed differences between T2-stage (35.1±29.4 ng/mL) and T3-stage (54.3±35.9 ng/mL) (Figure 29, adjusted p=0.018), between T2-stage (35.1±29.4 ng/mL) and T4-stage (54.7±40.2 ng/mL) (Figure 29, adjusted p=0.018), between M0 (52.7±32.7 ng/mL) and M1b (28.6±26.0 ng/mL) (Figure 30, adjusted p=0.003), and between M1b (28.6± 26.0ng/mL) and M1c (53.2±41.9 ng/mL) (Figure 30, adjusted p=0.016) based on PDAP1 concentrations between analysed groups.
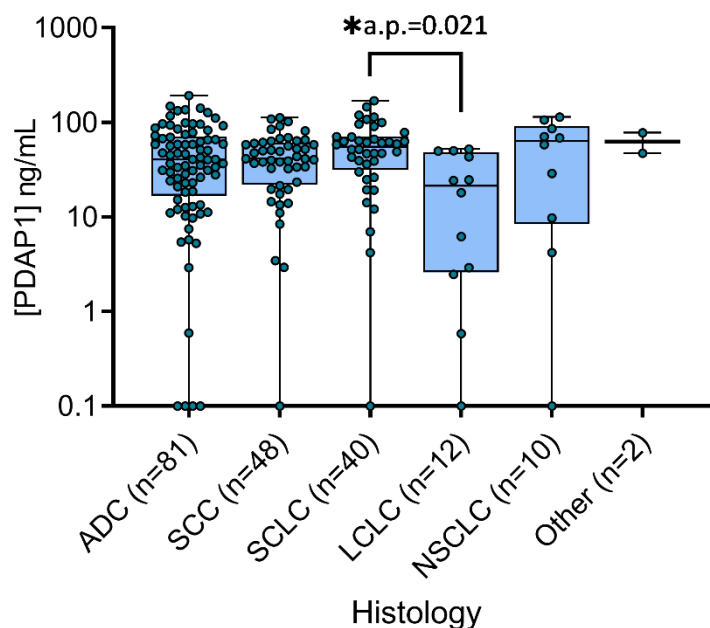
Figure 27. Association of PDAP1 with lung cancer histological subtypes. Concentrations of PDAP1 are represented on y-axis, whereas the lung cancer histological subtypes are represented on x-axis. Median and the interquartile range are displayed as horizontal bars. Minimum and the maximum of protein concentration are represented by whiskers. Significant difference between lung cancer histological subtypes is indicated with connection line and displayed as asterix (*). Protein concentration of n positioned directly on the x-axis is truthfully equal to zero. However, due to the logarithmic scale it appears to be 0.1 since it is mathematically impossible to plot a zero on logarithmic scale. Therefore, some of the boxplots are also dragged to the x-axis. n (ADC)=81, n (SCC)=48, n (SCLC)=40, n (LCLC)=12, n (NSCLC)=10, n (Other)=2. n=biological individual analysed, ADC=adenocarcinoma, SCC=squamous cell carcinoma, LCLC=large cell lung cancer, NSCLC=non-small cell lung cancer, SCLC=small cell lung cancer, a.p=adjusted p-value; p-values have been adjusted by the Bonferroni correction for multiple tests, ng/mL=nanogram per milliliter.
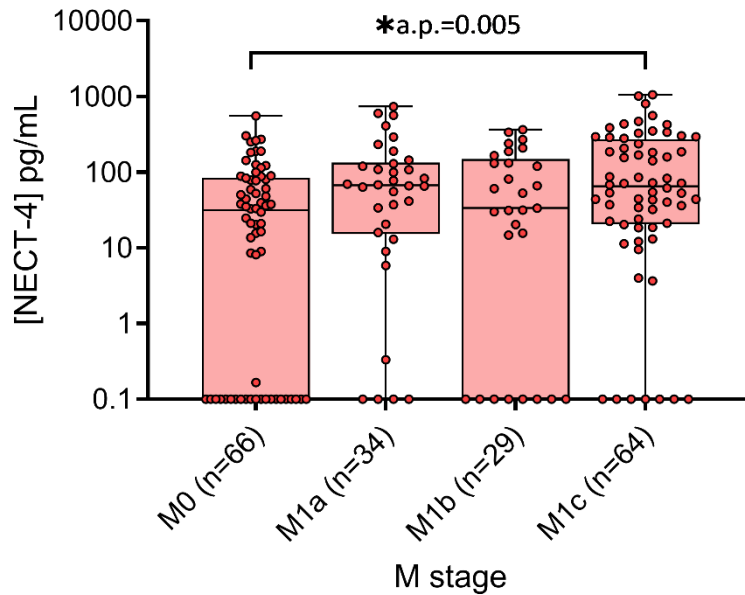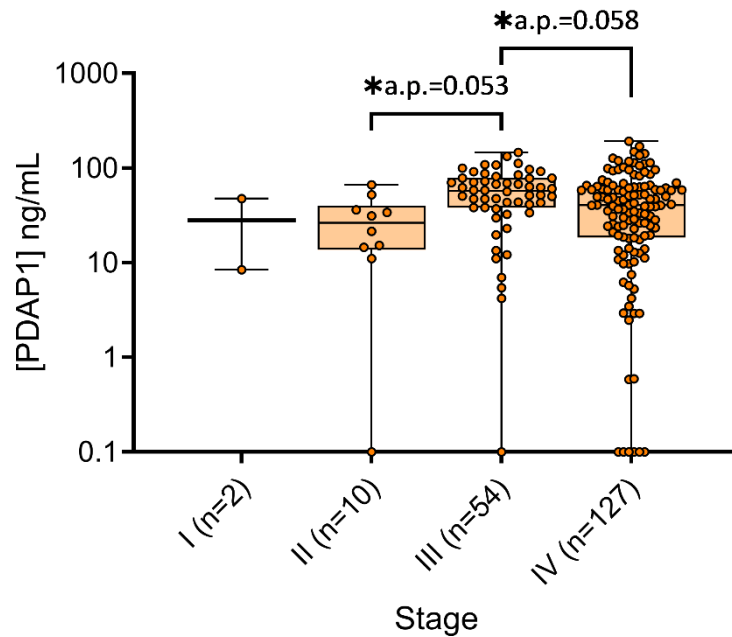
Figure 28. Association of PDAP1 with the stage of lung cancer. Concentrations of PDAP1 are presented on y-axis, whereas lung cancer stages are presented on x-axis. Median and the interquartile range are displayed as horizontal bars. Minimum and the maximum of protein concentration are represented by whiskers. Marginal *p*-values are indicated with connection lines and displayed as asterix (*). Protein concentration of n positioned directly on the x-axis is truthfully equal to zero. However, due to the logarithmic scale it appears to be 0.1 since it is mathematically impossible to plot a zero on logarithmic scale. Therefore, some of the boxplots are also dragged to the x-axis. n (stage I)=2, n (stage II)=10, n (stage III)=54, n (stage IV)=127. n=biological individual analysed, I=stage I, II=stage II, III=stage III, IV=stage IV, a.p=adjusted *p*-value; *p*-values have been adjusted by the Bonferroni correction for multiple tests, ng/mL=nanogram per milliliter.
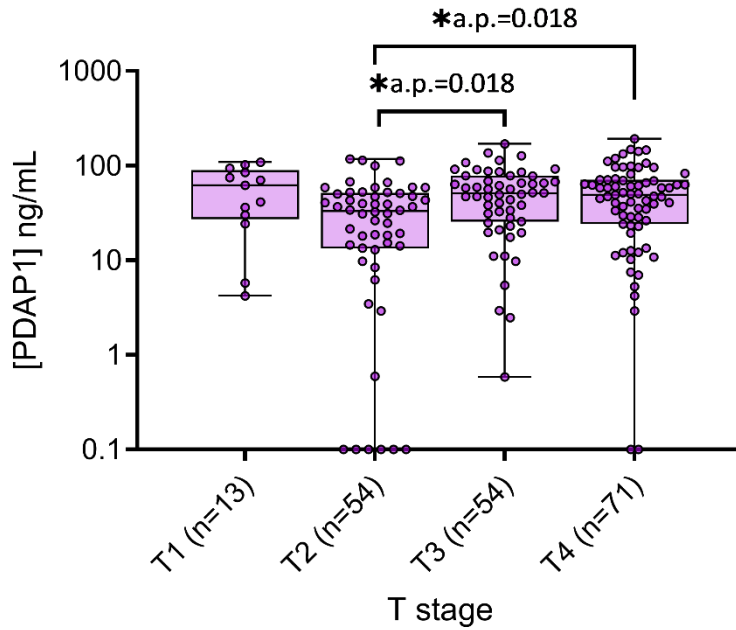
Figure 29. Association of PDAP1 with T-stage of lung cancer. Concentrations of PDAP1 are presented on y-axis, whereas T-stages of lung cancer are presented on x-axis. Median and the interquartile range are displayed as horizontal bars. Minimum and the maximum of protein concentration are represented by whiskers. Significant difference between T-stages of lung cancer is indicated with connection line and displayed as asterix (*). Protein concentration of n positioned directly on the x-axis is truthfully equal to zero. However, due to the logarithmic scale it appears to be 0.1 since it is mathematically impossible to plot a zero on logarithmic scale. Therefore, some of the boxplots are also dragged to the x-axis. n (T1-stage)=13, n (T2-stage)=54, n (T3-stage)=54, n (T4-stage)=71. n=biological individual analysed, T1=T1-stage, T2=T2-stage II, T3=T3-stage, T4=T4-stage, a. p=adjusted *p*-value; *p*-values have been adjusted by the Bonferroni correction for multiple tests, ng/mL=nanogram per milliliter.
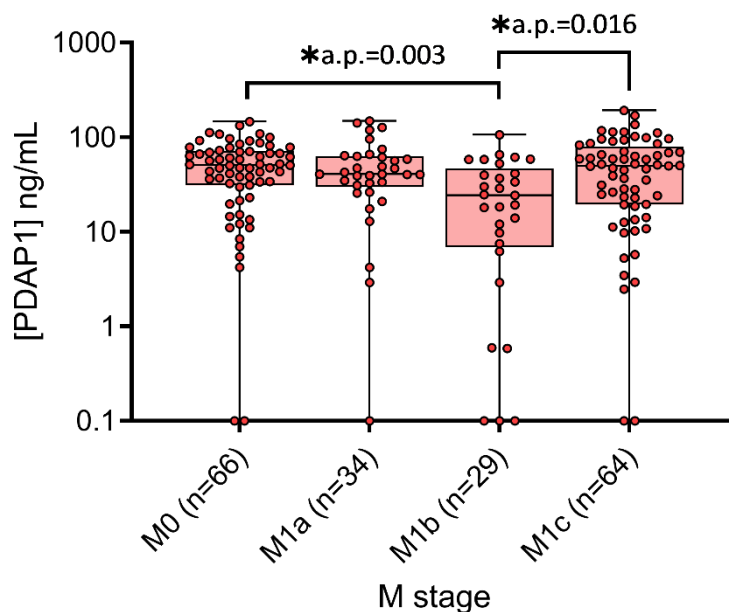
Figure 30. Association of PDAP1 with lung cancer metastases. Concentration of PDAP1 is presented on y-axis, whereas the stages of lung cancer metastasis are presented x-axis. Median and the interquartile range are displayed as horizontal bars. Minimum and the maximum of protein concentration are represented by whiskers. Significant difference between metastatic stages of lung cancer is indicated with connection line and displayed as asterix (*). Protein concentration of n positioned directly on the x-axis is truthfully equal to zero. However, due to the logarithmic scale it appears to be 0.1 since it is mathematically impossible to plot a zero on logarithmic scale. Therefore, some of the boxplots are also dragged to the x-axis. n (stage M0)=66, n (stage M1a)=34, n (stage M1b)=29, n (stage M1c)=64. n=biological individual analysed, M0=stage M0, M1a=stage M1a, M1b=stage M1b, M1c=stage M1c, a. p=adjusted *p*-value; *p*-values have been adjusted by the Bonferroni correction for multiple tests, M=metastasis, ng/mL=nanogram per milliliter.

5.6 Receiver operating characteristic (ROC) analyses of NECTIN-4 and PDAP1 in plasma

Diagnostic performance of NECT-4 and PDAP1 concentrations in plasma has been analysed using receiver operating characteristic (ROC) analysis. Summarized results are presented in Tables 5 and 6 and in Figures 31-34.

ROC analysis of NECT-4 concentrations in plasma of all lung cancer patients (n=193) against all healthy controls (n=200) showed discrimination between lung cancer patients and healthy controls ($p<0.0001$). Obtained area under the curve (AUC) was 0.7627 (95% confidence interval (CI)=0.7149-0.8105) with the sensitivity of 43.52% at the specificity of 95.00% (Table 5, Figure 31a). Stratification according to sex did not yield in significant improvement of sensitivity and specificity of NECT-4 test (Table 5, Figures 31b and c).

Table 5. Summary of ROC analysis of NECT-4 and PDAP1 in plasma and of their respective combined model in all lung cancer patients and healthy controls and stratified by sex.

Study group	*n* (LC)	*n* (HC)	Biomarker	AUC[a]	95% CI[b]	Sensitivity [%]	Specificity [%]	AUC *p*-value[c] (single biomarker)	AUC *p*-value[d] (comparison to combined)
Lung cancer all *vs*. healthy control all	193	200	Combined*	0.8269	0.7856-0.8682	50.78	95.00		
			NECTIN-4	0.7627	0.7149-0.8105	43.52	95.00	<0.0001	**0.0004**
			PDAP1	0.702	0.6495-0.7545	23.32	95.00	<0.0001	**<0.0001**
Lung cancer ♀ *vs*. healthy control ♀	80	29	Combined*	0.8071	0.7207-0.8936	42.50	96.55		
			NECTIN-4	0.7293	0.6345-0.8241	38.75	96.55	0.0003	0.0632
			PDAP1	0.7030	0.5946-0.8114	15.00	96.55	0.0012	**0.0094**
Lung cancer ♂ *vs*. healthy control ♂	113	171	Combined*	0.8526	0.8075-0.8977	56.64	95.32		
			NECTIN-4	0.7956	0.7396-0.8517	46.90	95.32	<0.0001	**0.0040**
			PDAP1	0.7177	0.6555-0.7799	26.55	95.32	<0.0001	**<0.0001**

♀=women
♂=men
Combined*=combination model of PDAP1 and NECTIN-4
[a] Area under the curve
[b] Wald confidence interval
[c] Testing the null hypothesis that AUC=0.5
[d] Chi Square test
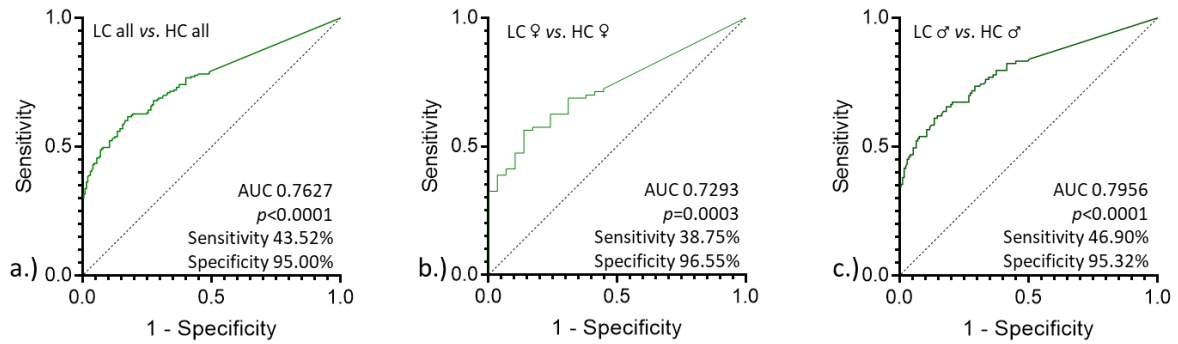
Figure 31. Receiver operating characteristic (ROC) curves of NECT-4 in all lung cancer and healthy control participants of the study cohort and stratified by sex. Sensitivities of the test are presented on y-axis, whereas the corresponding 1-specificities are presented on x-axis. a.) ROC curve of NECT-4 in all lung cancer patients (n=193) and all healthy controls (n=200). b.) ROC curve of NECT-4 in female lung cancer patients (n=80) and female healthy controls (n=29). c.) ROC curve of NECT-4 in male lung cancer patients (n=113) and male healthy controls (n=171). NECT-4=NECTIN-4, LC=lung cancer, HC=healthy control, ♀=female, ♂=male, AUC=area under the curve, p-value=testing the null hypothesis that AUC=0.5.

Further on, lung cancer patients have been stratified according to the histological subtypes of lung cancer and compared to all healthy controls using ROC analysis (Table 6, Figure 32). Since statistical analysis is not reliable for small groups, ROC analysis for large cell lung carcinoma, non-small cell lung carcinoma, and *other* was performed for the analysis completeness and stated in Table 6 but it was not taken into consideration for further observation due to the small number of patients within the corresponding groups (<20 patients).

ROC analysis of NECT-4 concentrations showed discrimination between adenocarcinoma patients and healthy controls (p<0.0001, Figure 32a), squamous cell carcinoma patients and healthy controls (p<0.0001, Figure 32b), small cell lung cancer patients and healthy controls (p<0.0001, Figure 32c), and non-small cell lung cancer patients and healthy controls (p=0.0096). ROC analysis of NECT-4 concentrations in large cell lung carcinoma (p=0.2009) and *other* lung cancer subtypes (p=0.5393) did not show discrimination between cancer patients and healthy controls. The highest AUC was obtained for adenocarcinoma patient detection (AUC=0.8267, 95% CI=0.7652-0.8882) with the sensitivity of 56.79% at the specificity of 95.00%, followed by non-small cell lung cancer (AUC=0.7428, 95% CI=0.5596-0.9259, sensitivity=40.00%, specificity=95.00%), squamous cell carcinoma (AUC=0.7303, 95% CI=0.6405-0.8200, sensitivity=39.58%, specificity=95.00%), and small cell lung cancer (AUC=0.7297, 95% CI=0.6321-0.8273, sensitivity=32.50%, specificity=95.00%) (Table 6, Figure 32).

Table 6. Summary of the ROC analysis of NECTIN-4 and PDAP1 in plasma and of their respective combined model in lung cancer patients stratified by histological subtype and healthy controls.

Study group	n (LC)	n (HC)	Biomarker	AUC[a]	95% CI[b]	Sensitivity [%]	Specificity [%]	AUC p-value[c] (single biomarker)	AUC p-value[d] (comparison to combined)
Adenocarcinoma all vs .healthy control all	81	200	Combined*	0.8951	0.8530-0.9371	60.50	90.00		
			NECTIN-4	0.8267	0.7652-0.8882	56.79	95.00	<0.0001	**0.0054**
			PDAP1	0.6799	0.6052-0.7546	25.93	95.00	<0.0001	**<0.0001**
Squamous cell carcinoma all vs . Healthy control all	48	200	Combined*	0.7891	0.7107-0.8674	52.08	95.00		
			NECTIN-4	0.7303	0.6405-0.8200	39.58	95.00	<0.0001	0.0508
			PDAP1	0.7263	0.6433-0.8093	14.58	95.00	<0.0001	0.0643
Small celll ung cancer all vs . Healthy control all	40	200	Combined*	0.8250	0.7464-0.9036	47.50	95.00		
			NECTIN-4	0.7297	0.6321-0.8273	32.50	95.00	<0.0001	**0.0076**
			PDAP1	0.7844	0.6990-0.8699	27.50	95.00	<0.0001	0.1823
Large cell lung cancer all vs . healthy control all	12	200	Combined*	0.6019	0.4359-0.7678	16.67	94.50		
			NECTIN-4	0.6100	0.4423-0.7777	8.33	95.00	0.2009	0.9014
			PDAP1	0.5406	0.3463-0.7349	8.33	95.00	0.6366	0.5086
Non-small cell lung cancer all vs . Healthy control all	10	200	Combined*	0.7308	0.5008-0.9607	40.00	95.00		
			NECTIN-4	0.7428	0.5596-0.9259	40.00	95.00	0.0096	0.8214
			PDAP1	0.6818	0.4476-0.9159	50.00	95.00	0.0526	0.5693
Other all vs . healthy control all	2	200	Combined*	0.9400	0.8203-1.0000	50.00	95.00		
			NECTIN-4	0.6263	0.0880-1.0000	50.00	95.00	0.5393	0.3177
			PDAP1	0.9250	0.8484-1.0000	50.00	95.00	0.0388	0.8922

LC=lung cancer
HC=healthy control
n=number of individuals analyzed
Combined*=combination model of PDAP1 and NECTIN-4
[a] Area under the curve
[b] Wald confidence interval
[c] Testing the null hypothesis that AUC=0.5
[d] Chi Square test

Figure 32. Receiver operating characteristic (ROC) curves of NECT-4 in lung cancer patients stratified by histological subtype (n>20) and all healthy control participants of the study cohort. Sensitivities of the test are presented on y-axis, whereas the corresponding 1-specificity is presented on x-axis. a.) ROC curve of NECT-4 in adenocarcinoma patients (n=81) and healthy controls (n=200). b.) ROC curve of NECT-4 in squamous cell carcinoma patients (n=48) and all healthy controls (n=200). c.) ROC curve of NECT-4 in small cell lung cancer patients (n=40) and all healthy controls (n=200). NECT-4=NECTIN-4, LC=lung cancer, HC=healthy control, AUC=area under the curve, p-value= testing the null hypothesis that AUC=0.5.

ROC analysis of PDAP1 concentrations in plasma of all lung cancer patients (n=193) against all healthy controls (n=200) showed discrimination between lung cancer patients and healthy controls ($p<0.0001$). Obtained AUC was 0.7020 (95% CI=0.6495-0.7545) with sensitivity of 23.32% at the specificity of 95.00% (Table 5, Figure 33a). Stratification according to sex as well did not yield in significant improvement of sensitivity and specificity of the test (Table 5, Figure 33b and c).
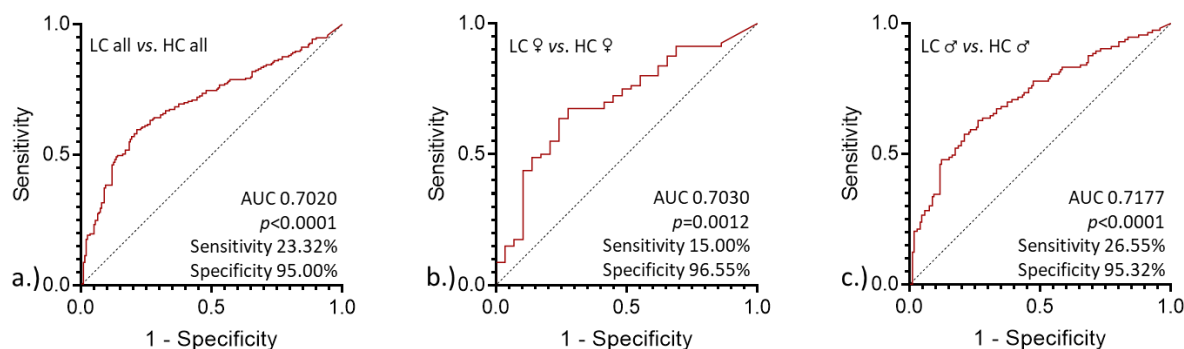
Figure 33. Receiver operating characteristic (ROC) curves of PDAP1 in all lung cancer and healthy control participants of the study cohort and stratified by sex. Sensitivities are presented on y-axis, whereas the corresponding 1-specificities are presented on x-axis. a.) ROC curve of PDAP1 in all lung cancer patients (n=193) and healthy controls (n=200). b.) ROC curve of PDAP1 in female lung cancer patients (n=80) and female healthy controls (n=29). c.) ROC curve of PDAP1 in male lung cancer patients (n=113) and male healthy controls (n=171). LC=lung cancer, HC=healthy control, ♀=female, ♂=male, AUC=area under the curve. *p*-value= testing the null hypothesis that AUC=0.5.

For further analysis of PDAP1, lung cancer patients have been stratified according to the histological subtypes of lung cancer and compared to all healthy controls using ROC analysis (Table 6, Figure 34). As stated above, ROC analysis of histological subtypes comprising <20 patients was performed for the analysis completeness (Table 6.). However, since statistical analysis is not reliable for small groups, it was not taken into consideration for further observation.

ROC analysis of PDAP1 concentrations showed discrimination between adenocarcinoma patients and healthy controls ($p<0.0001$, Figure 34a), squamous cell carcinoma patients and healthy controls ($p<0.0001$, Figure 34b), small cell lung cancer patients and healthy controls ($p<0.0001$, Figure 34c), and patients with *other* lung cancer subtypes and healthy controls ($p=0.0388$). ROC analysis of PDAP1 concentrations in large cell lung carcinoma ($p=0.6366$) and non-small cell lung cancer subtypes ($p=0.0526$) did not show discrimination between cancer patients and healthy controls (Table 6).

Comparison of PDAP1 concentrations between patients with *other* lung cancer histological subtypes and all healthy controls provided the highest AUC (AUC=0.9250, 95% CI=0.8484-1.000) compared to other histological subtypes against healthy controls, with sensitivity of 50.00% at specificity of 95.00%, followed by small cell lung cancer (AUC=0.7844, 95% CI=0.6990-0.8699, sensitivity=27.50%,

specificity=95.00%), squamous cell carcinoma (AUC=0.7263, 95% CI=0.6433-0.8093, sensitivity=14.58%, specificity=95.00%), and adenocarcinoma (AUC=0.6799, 95% CI=0.6052-0.7546, sensitivity=25.93%, specificity=95.00%) (Table 6, Figure 34).
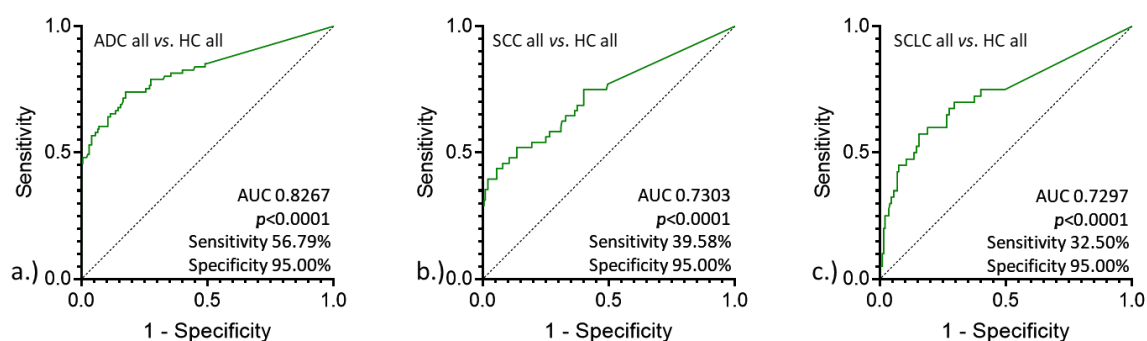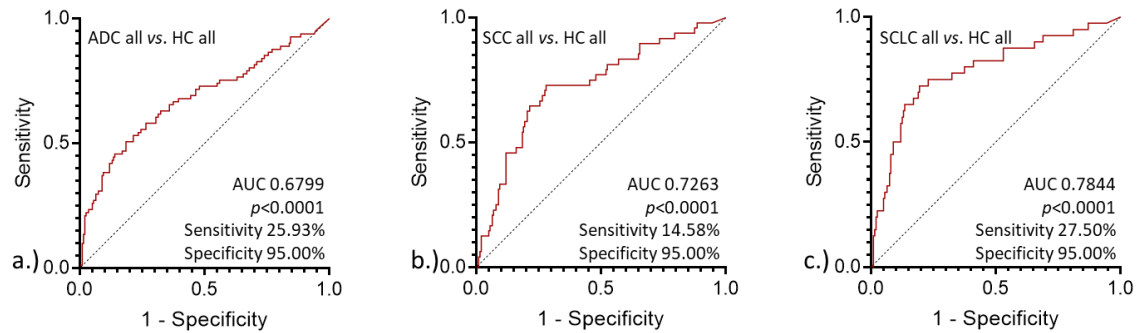

Figure 34. Receiver operating characteristic (ROC) curves of PDAP1 in lung cancer patients stratified by histological subtype and all healthy control participants of the study cohort. Sensitivities are presented on y-axis, whereas the corresponding 1-specificities are presented on x-axis. a.) ROC curve of PDAP1 in adenocarcinoma patients (n=81) and healthy controls (n=200). b.) ROC curve of PDAP1 in squamous cell carcinoma patients (n=48) and all healthy controls (n=200). c.) ROC curve of PDAP1 in small cell lung cancer patients (n=40) and all healthy controls (n=200). LC=lung cancer, HC=healthy control, AUC=area under the curve, *p*-value= testing the null hypothesis that AUC=0.5.

Since NECT-4 and PDAP1 concentrations were measured in parallel on the same samples, the combination of both proteins as combined biomarker was analyzed using ROC analysis (Tables 5 and 6, Figure 35). For discrimination whether the combined biomarker is superior compared to single biomarker for lung cancer detection, the AUC of combined biomarker was compared to the AUC of NECT-4 only and to the AUC of PDAP1 only using Chi Square test.

When comparing NECT-4 biomarker and combined biomarker (NECT-4 and PDAP1), the improvement of the test was achieved for discrimination between lung cancer patients and healthy controls (*p*=0.0004; ROC analysis parameters of NECT-4 and PDAP1 combined biomarker: AUC=0.8269, 95% CI=0.7856-0.8682, sensitivity=50.78%, specificity=95%). Moreover, the improvement of the test was achieved when using combined biomarker for discrimination between male lung cancer patients and male healthy controls (*p*=0.0040; ROC analysis parameters of NECT-4 and PDAP1 combined biomarker: AUC=0.8526, 95% CI=0.8075-0.8977, sensitivity=56.64%, specificity=95.32%) (Table 5, Figure 35).

When comparing PDAP1 biomarker and combined biomarker (NECT-4 and PDAP1), as same as for NECT-4, the improvement of the test was achieved for discrimination between lung cancer patients and healthy controls (*p*<0.0001) and for discrimination between male lung cancer patients and male healthy controls (*p*<0.0001) (ROC analysis parameters of NECT-4 and PDAP1 combined biomarker mentioned above). Additionally, combined biomarker showed improvement of the test compared to PDAP1 as

single biomarker for the discrimination between female lung cancer patients and healthy controls (p=0.0094; ROC analysis parameters of NECT-4 and PDAP1 combined biomarker: AUC=0.8071, 95% CI=0.7207-0.8936, sensitivity=42.50%, specificity=96.55%) (Table 5, Figure 35).

Details of combined model ROC analysis of lung cancer patients and healthy controls and stratified by sex and the corresponding comparison to ROC analysis of single biomarker are summarized and specified in Table 5.
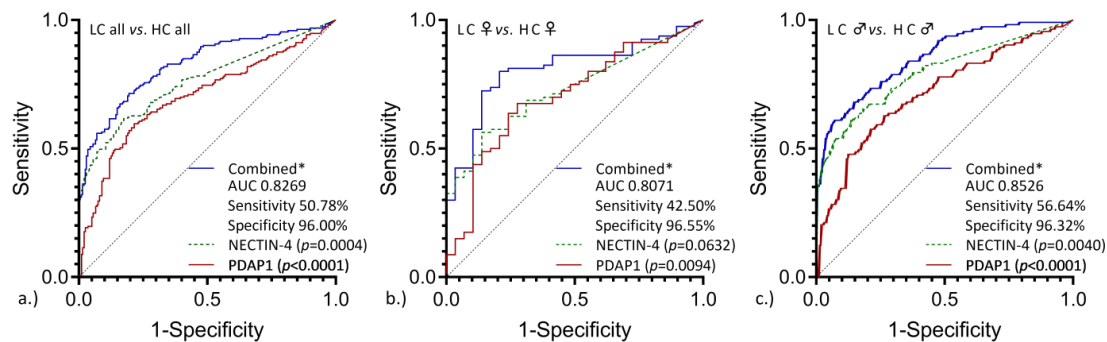
Figure 35. Combined receiver operating characteristic (ROC) curves of NECT-4 and PDP1 as single biomarkers and combined biomarker (NECT-4 and PDAP1) in all lung cancer and healthy control participants of the study cohort and stratified by sex. Sensitivities are presented on y-axis, whereas the corresponding 1-specificity is presented on x-axis. a.) ROC curves of NECT-4 and PDAP1 as single biomarkers and combined biomarker (NECT-4 and PDAP1) in all lung cancer patients (n=193) and healthy controls (n=200). b.) ROC curve s of NECT-4 and PDAP1 as single biomarkers and combined biomarker (NECT-4 and PDAP1) in female lung cancer patients (n=80) and female healthy controls (n=29). c.) ROC curves of NECT-4 and PDAP1 as single biomarkers and combined biomarker (NECT-4 and PDAP1) in male lung cancer patients (n=113) and male healthy controls (n=171). NECT-4=NECTIN-4, LC=lung cancer, HC=healthy control, ♀=female, ♂=male, AUC=area under the curve. p-value=Chi Square test.

For further analysis, lung cancer patients have been stratified according to the histological subtypes of lung cancer and compared to all healthy controls using ROC analysis (Table 6, Figure 36). To test whether the combined biomarker (NECT-4 and PDAP1) is superior for the detection of specific lung cancer histological subtype compared to single biomarker, the AUC from combined biomarker was compared to the AUC of each single biomarker, NECT-4 and PDAP1 using Chi Square test.

As it was the case for the analysis of individual biomarkers, the analysis for combined biomarker of histological subtypes comprising <20 patients was performed for the completeness (Table 6.). However, since statistical analysis is not reliable for small groups, it was not taken into consideration for further observation.

Improvement of the test was achieved for discrimination between adenocarcinoma and healthy controls (*p*=0.0054; ROC analysis parameters of NECT-4 and PDAP1 combined biomarker: AUC=0.8951, 95% CI=0.8530-0.9371, sensitivity=60%, specificity=95%) and for discrimination between small cell lung cancer patients and healthy controls (*p*=0.0076; ROC analysis parameters of NECT-4 and PDAP1 combination model: AUC=0.8250, 95% CI=0.7464-0.9036, sensitivity=47.5%, specificity=95.00%), when comparing the combined biomarker of NECT-4 and PDAP1 with NECT-4 only (Figure 36a and c). For other histological subtypes, there was no significant improvement when using combination model of NECT-4 and PDAP1 as biomarker compared to NECT-4 only (Table 6.). However, it is worth to mention that the improvement of squamous cell carcinoma detection when using combined biomarker of NECT-4 and PDAP1 compared to NECT-4 only had marginal significance (*p*=0.0508; AUC=0.7107-0.8674, sensitivity=52.08%, specificity=95.00%) (Table 6, Figure 36b).

Comparing the combined biomarker of NECT-4 and PDAP with PDAP1 only, improvement of the test was achieved for discrimination between adenocarcinoma patients and healthy controls (*p*<0.0001; ROC analysis parameters of NECT-4 and PDAP1 combination model mentioned in paragraph above) (Table 6, Figure 36a). For the other histological subtypes, there was no significant improvement when using combined biomarker of NECT-4 and PDAP1 compared to PDAP1 only (Table 6.) (Figure 36b and c).

Details of the ROC analysis of combined biomarker (NECT-4 + PDAP1) and the comparison to the single biomarker (NECT-4 or PDAP1) for the detection of each lung cancer histological subtype are specified in Table 6 (see above in this chapter).
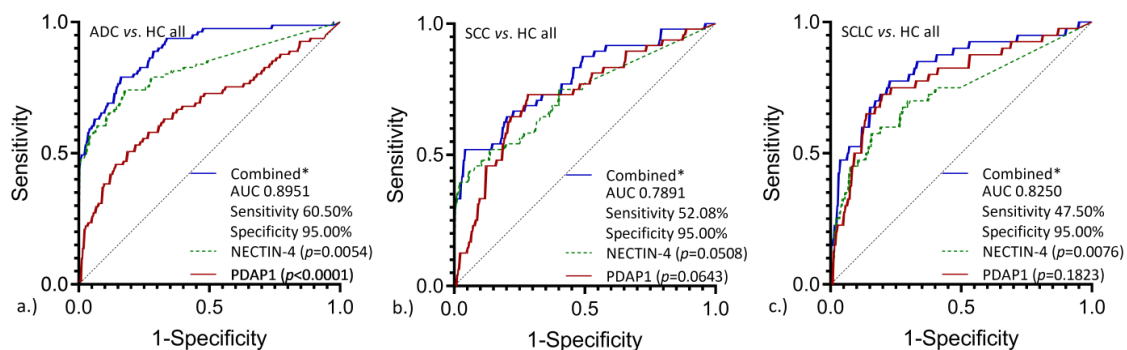
Figure 36. Combined receiver operating characteristic (ROC) curves of NECT-4 and PDAP1 as single biomarkers and combined model of both biomarkers (NECT-4 + PDAP1) in lung cancer patients stratified by histological subtype and all healthy control participants of the study cohort. Sensitivities are presented on y-axis, whereas the corresponding 1-specificities are presented on x-axis. a.) ROC curves of NECT-4 and PDAP1 as single biomarkers and of combined biomarker (NECT-4 + PDAP1) in adenocarcinoma patients (n=81) and healthy controls (n=200). b.) ROC curves of NECT-4 and PDAP1 as single biomarkers and of combined biomarker (NECT-4 + PDAP1) in squamous cell carcinoma patients (n=48) and all healthy controls (n=200). c.) ROC curves of NECT-4 and PDAP1 as single biomarkers and of combined biomarker (NECT-4 + PDAP1) in small cell lung cancer patients

(n=40) and all healthy controls (n=200). NECT-4=NECTIN-4, LC=lung cancer, HC=healthy control, AUC=area under the curve, *p*-value=Chi Square test.

Another key aim in this thesis was an identification of EV-derived protein biomarkers for an early detection of lung cancer using LFQ-MS/MS. For that, a method for EV isolation from plasma suitable for downstream MS analysis sought to be developed. Thus, three different EV-isolation protocols were compared for the yield and purity of the EVs isolated from plasma. Established protocol for the EV isolation from plasma samples was used for the isolation of EVs which were further on subjected to LFQ-MS/MS analysis. MS data was used for identification of potential EV-derived biomarkers for an early detection of lung cancer. Finally, pathway analysis of differentially expressed genes was performed to gain insight into the EV biology related to lung cancerogenesis.

5.7 Characterization of EVs isolated from healthy donor plasma using different isolation methods

Despite EV research field exponentially increasing, there is no standardized method for the isolation of EVs yet. Therefore, one of the goals in this thesis was to optimize EV isolation protocol which would yield EVs of high purity.

Three isolation methods were compared to determine which isolation method provides the highest purity and yield of plasma-derived-EVs. EVs were isolated from 2 mL of healthy donor plasma using only ultracentrifugation (UC) which is golden standard, size exclusion chromatography coupled with ultracentrifugation (SEC+UC), and size exclusion chromatography coupled with centrifugal filtration (SEC+CF). Isolated EVs were characterized using NTA, sandwich CD9 and CD63 ELISAs, and *Western blot* analysis for lipoprotein and CD9 detection (see Figure 63, section 8.4.1). Protein concentration was measured using Qubit (8.3.2 M&M section). Summary of all measured values can be found in Table 7 in the end of 5.7.2 section.

12 mL of *pre-cleared* healthy donor (HD) plasma was divided into 3 equal parts, having 4 mL of HD plasma for each isolation method to be performed in duplicates, UC only, SEC+UC, and SEC+CF (8.4.1 M&M section). For UC only, 4 mL of *pre-cleared* HD plasma was directly split into two equal parts of 2 mL and applied for UC, later referred to as UC (1) and UC (2). The rest of 8 mL of *pre-cleared* HD plasma designated for SEC (SEC+UC or SEC+CF method) was split into two equal parts of 4 mL further on referred to HD plasma (1) and HD plasma (2), and processed manually by SEC. EV-enriched fractions without or with low protein contamination were determined using CD9 and CD63 sandwich ELISAs and BCA analysis of each SEC fraction as described in details in M&M section 8.4.1.

5.7.1 Determination of EV-enriched SEC fraction and total protein concentration in SEC fractions (HD plasma)

For methods SEC+UC and SEC+CF, SEC-EV-containing fractions had to be determined and pooled in order to proceed on with downstream UC or CF as following purification step. Therefore, each SEC fraction has been analysed using CD9 and CD63 sandwich ELISAs. In parallel, total protein concentration had been measured of each SEC fraction (Figure 37 and 38). Analysis of 70 SEC fractions of HD plasma (1) showed the increase in CD9 signal at the SEC fraction 21 with OD=0.07, the highest intensity of CD9 signal in SEC fraction 33 with OD=1.59 and decrease of a CD9 signal after fraction 33 (Figure 37a). Analysis showed lower signal intensity of CD63 compared to signal intensity of CD9 across SEC fractions (Figure 37). First CD63 signal was detected in fraction 22 with OD=0.064. The highest intensity of CD63 signal was in SEC fraction 54 with OD=0.091 and decrease of a CD63 is observed after fraction 54. Disregard of low OD intensity for CD63 signal, the elution profile of CD63$^+$ EVs is wider comparing to CD9$^+$ EVs (Figure 37). Total protein concentration starts to gradually increase from fraction 34-41 after which it rapidly increases. The highest total protein concentration is observed in SEC fraction 65. Total protein concentration starts to decrease after fraction 65 (Figure 37). It is important to mention that the experiments on EV isolation from plasma were performed at ITB, UKE, Hamburg, Germany where the rotor available for the usage in this thesis allowed a maximum of 10 mL of fractions in total to be applied for UC. Thus, the corresponding equal amount was applied for CF for proper comparison of the methods. Given that, maximum of 20 mL of fractions in total could have been pooled. Taken together, SEC fractions 22-42 were considered as EV-enriched fractions without or with low protein contamination, depending on a SEC fraction. Therefore, they were pooled and split into two equal parts of 10 mL, one for ultracentrifugation (UC) and another for centrifugal filtration (CF), further on referred to as SEC+UC (1) and SEC+CF (1).

Figure 37. Analysis of 70 SEC fractions derived from HD plasma (1) for EV presence and protein contamination. CD9 or CD63 sandwich ELISA is designated in blue and represent EV-containing fractions, whereas BCA is designated in orange and increase of total protein concentration indicates increase in non-EV proteins, therefore protein contamination. a.) CD9 sandwich ELISA and BCA of

70 SEC fractions derived from HD plasma (1) b.) CD63 sandwich ELISA and BCA of 70 SEC fractions derived from HD plasma (1). a.) and b.) Total protein concentration starts to increase from fraction 34 on, and reaches its peak in fraction 50, after which it decreases. SEC fraction 22-42 (red circles) were considered as EV-enriched fractions without or with low protein contamination, pooled and split into two equal parts, and further on applied for ultracentrifugation or centrifugal filtration. n=3. n=biological replicas. ELISA=enzyme-linked immunoassay, SEC=size exclusion chromatography, BCA=bicinchoninic acid assay, HD=healthy donor, EV=extracellular vesicle.

Analysis of 70 SEC fractions of HD plasma (2) corelates with the analysis of 70 SEC fraction from HD plasma (1). CD9 signal intensity increases in SEC fraction 21 with OD=0.05 and has the highest signal intensity in SEC fraction 32 with OD=1.13. CD9 intensity decreases after fraction 32 (Figure 38a). The analysis showed lower signal intensity of CD63 compared to the signal intensity of CD9 across SEC fractions (Figure 38), as for the analysis of HD plasma (1). Increase in CD63 signal intensity starts in SEC fraction 24 with OD=0.03 and gradually increases after it. The highest signal intensity of CD63 is in fraction 55 with OD=0.08, after which it decreases (Figure 38b). As observed before in elution profile of HD plasma (1), elution profile of CD63$^+$ EVs from HD plasma (2) is wider compared to the elution profile of CD9$^+$ EVs from HD plasma (2) (Figure 38). Total protein concentration starts to gradually increase from fraction 32-48, after which it rapidly increases. The highest total protein concentration is observed in fraction 49. After SEC fraction 49, total protein concentration starts to decrease until fraction 60 (Figure 38). Signal intensity for total protein concentration was above detection limit after fraction 60. Therefore, fraction 61-70 are not included in Figure 38. As stated above, the EV isolation from plasma was performed at ITB, UKE, Hamburg, Germany where the rotor available for usage in this thesis allowed a maximum of 10 mL of fractions in total to be applied for UC. Thus, the corresponding equal amount was applied for CF for proper comparison of the methods. Given that, maximum of 20 mL of fractions in total could have been pooled. Taken together, SEC fraction 22-42 were considered as EV-enriched fractions without or with low protein contamination, depending on a SEC fraction. Therefore, they were pooled and split into two equal parts of 10 mL, one for ultracentrifugation and another for centrifugal filtration, further on referred to as SEC+UC (2) and SEC+CF (2).
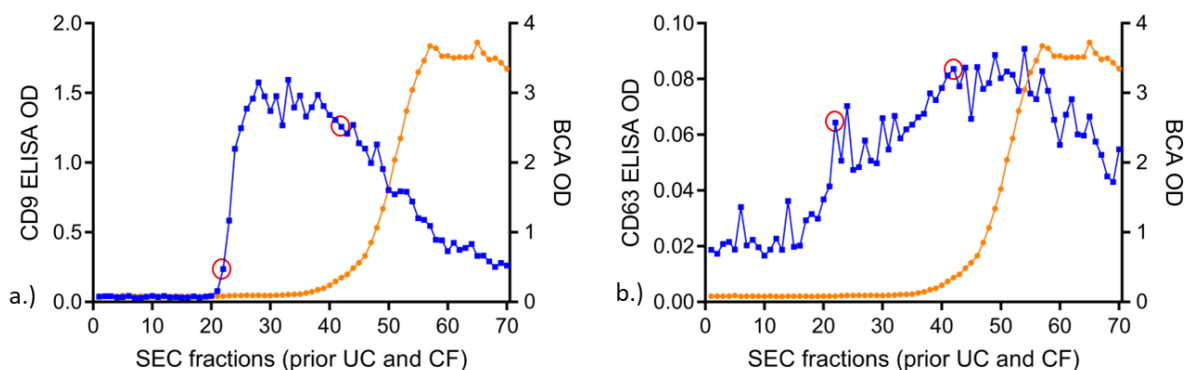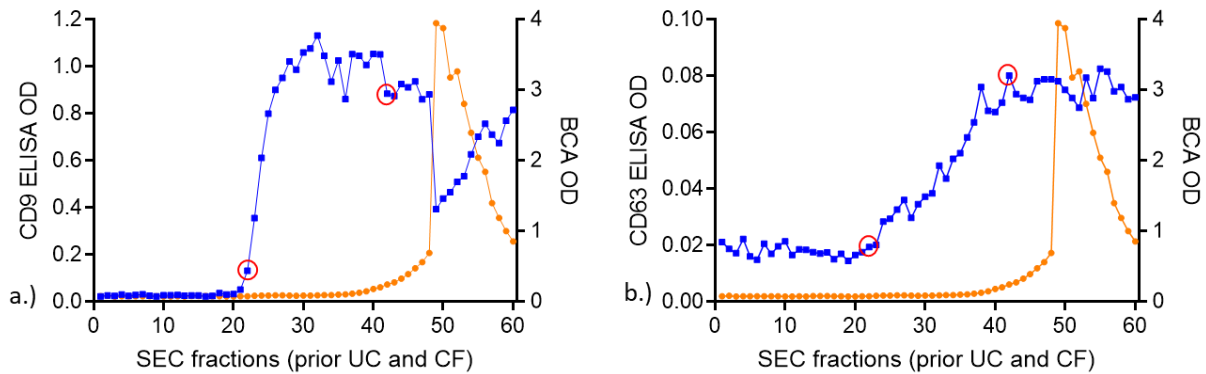
Figure 38. Representation of 60/70 analyzed SEC fractions derived from HD plasma (2) for EV presence and protein contamination. CD9 or CD63 sandwich ELISA is designated in blue and represent EV-containing fractions, whereas BCA is designated in orange and increase of total protein concentration indicates increase in non-EV proteins, therefore protein contamination. a.) CD9 sandwich ELISA and BCA of 60/70 analyzed SEC fractions derived from HD plasma (2). b.) CD63 sandwich ELISA and BCA of 60/70 analyzed SEC fractions derived from HD plasma (2). a.) and b.) Total protein concentration starts to increase from fraction 32 on, and reaches its peak in fraction 49, after which it decreases. SEC fraction 22-42 (red circles) were considered as EV-enriched fractions without or with low protein contamination, pooled and split into two equal parts, and further on applied for ultracentrifugation or centrifugal filtration. n=3. n=biological replicas. ELISA=enzyme-linked immunoassay, SEC=size exclusion chromatography, BCA=bicinchoninic acid assay, HD=healthy donor, EV=extracellular vesicle.

EVs isolated using SEC+UC and SEC+CF, together with EVs isolated using UC only, were further analysed by NTA, Qubit and Western blot analysis (see below 5.7.2 and 5.7.3).

5.7.2 Characterisation of EVs isolated from HD plasma using different isolation methods by Nanoparticle tracking assay (NTA) and Qubit

Extracellular vesicles isolated from HD plasma using UC only, SEC+UC, and SEC+CF methods were characterized for their concentration and size using NTA. Total protein concentration of each sample was measured by Qubit assay.

Mean size of EVs from UC (1) sample is 166.6±2.7 nm, whereas the mode size is 129.9±4.7 nm. Concentration of EVs from UC (1) sample is $8.24 \times 10^{10} \pm 3.39 \times 10^9$ particles/mL, and the total protein concentration of the sample is 162 µg/mL (Figure 39a and Table 7). Mean size of EVs from UC (2) sample is 165.9±2.3 nm, whereas the mode size is 117.6±5.6 nm. Concentration of EVs from UC (2) sample is $7.48 \times 10^{10} \pm 3.10 \times 10^9$ particles/mL, and the total protein concentration of the sample is 120 µg/mL (Figure 39b and Table 7).
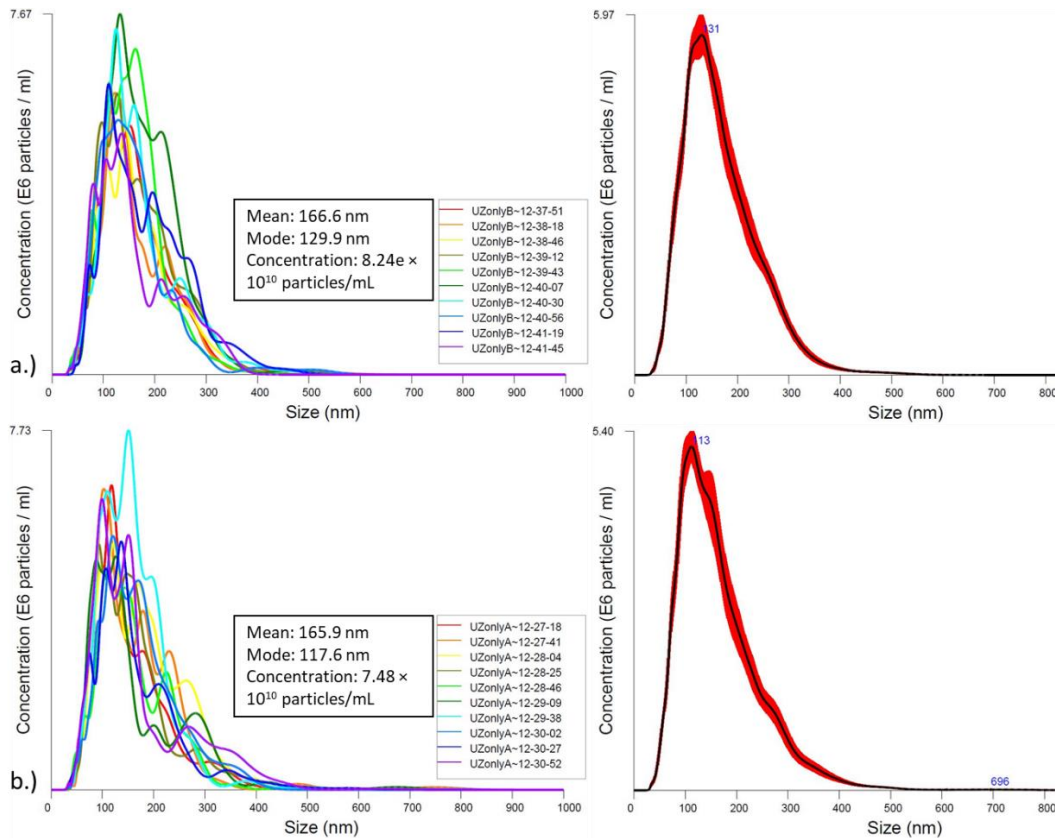
Figure 39. NTA of EVs isolated from 2 mL of HD plasma using UC only method. Graphs on the left are plotted with EV concentration on y-axis and EV size on x-axis for each of the five NTA measurements of a sample. Graphs on the right are plotted with average EV concentration on y-axes and EV size on x-axis for the experiment of each sample. ± 1 standard error of the mean is represented in red. a.) NTA of EVs from UC (1) sample, isolated from 2 mL of HD plasma using UC only. b.) NTA of EVs from UC (2) sample, isolated from 2 mL of HD plasma using UC only. n=3. NTA=nanoparticle tracking assay, HD=healthy donor, EV=extracellular vesicle, UC=ultracentrifugation. n=biological replica, nm=nanometer, ml=milliliter.

Mean size of EVs from SEC+UC (1) sample is 172.4±6.7 nm, whereas the mode size is 137.0±5.0 nm. Concentration of EVs from SEC+UC (1) sample is $7.11\times10^{10}\pm3.81\times10^{9}$ particles/mL, and the total protein concentration of the sample is 198 µg/mL (Figure 40a and Table 7). Mean size of EVs from SEC+UC (2) sample is 150.0±4.6 nm, whereas the mode size is 116.9±7.5 nm. Concentration of EVs from SEC+UC (2) sample is $4.92\times10^{10}\pm2.01\times10^{9}$ particles/mL, and the total protein concentration of the sample is 154 µg/mL (Figure 40b and Table 7).
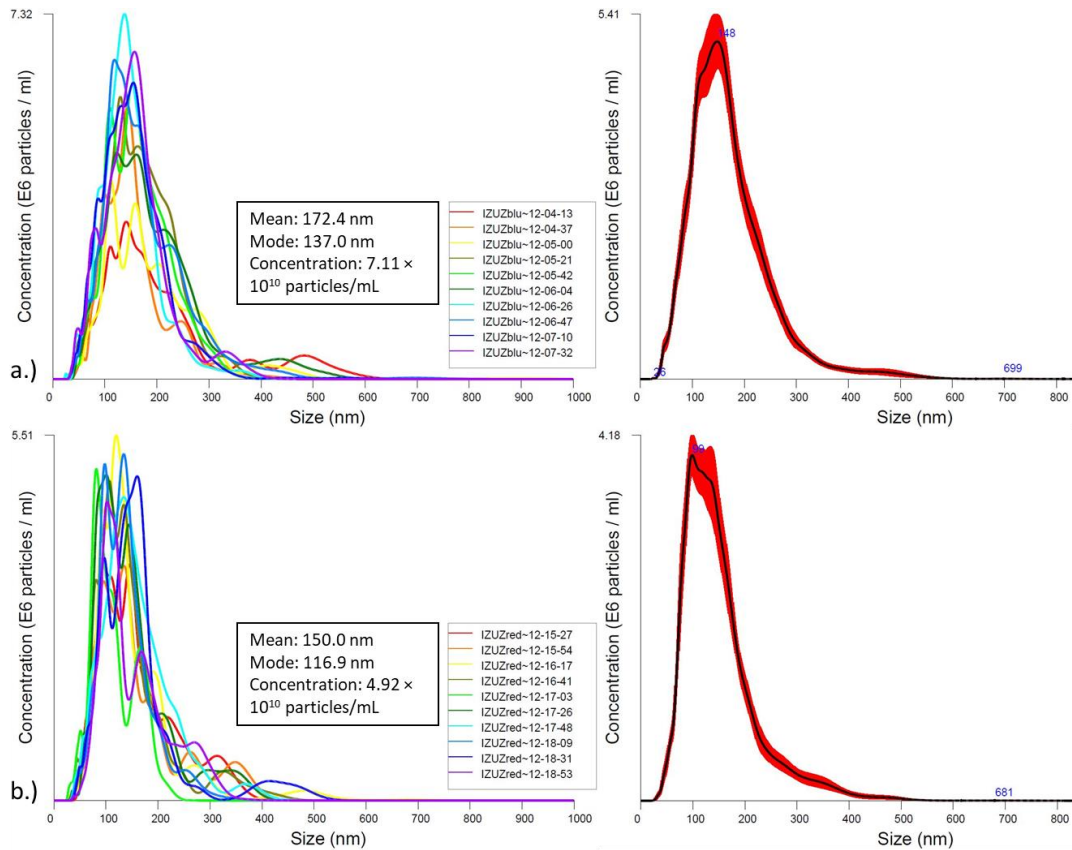
Figure 40. NTA of EVs isolated from 2 mL of HD plasma using SEC+UC method. Graphs on the left are plotted with EV concentration on y-axis and EV size on x-axis for each of the five NTA measurements of a sample. Graphs on the right are plotted with average EV concentration on y-axes and EV size on x-axis for the experiment of each sample. ± 1 standard error of the mean is represented in red. a.) NTA of EVs from SEC+UC (1) sample, isolated from 2 mL of HD plasma using SEC+UC method. b.) NTA of EVs from SEC+UC (2) sample, isolated from 2 mL of HD plasma using SEC+UC method. n=3. NTA=nanoparticle tracking assay, HD=healthy donor, EV=extracellular vesicle, SEC=size exclusion chromatography, UC=ultracentrifugation. n=biological replica, nm=nanometer, ml=milliliter.

Mean size of EVs from SEC+CF (1) sample is 104.6±2.5 nm, whereas the mode size is 90.2±4.1 nm. Concentration of EVs from SEC+CF (1) sample is $5.02 \times 10^{12} \pm 3.09 \times 10^{11}$ particles/mL, and the total protein concentration of the sample is 348 µg/mL (Figure 41a and Table 7). Mean size of EVs from SEC+CF (2) sample is 107.6±3.1 nm, whereas the mode size is 86.9±3.7 nm. Concentration of EVs from SEC+CF (2) sample is $4.22 \times 10^{12} \pm 2.70 \times 10^{11}$ particles/mL, and the total protein concentration of the sample is 284 µg/mL (Figure 41b and Table 7).
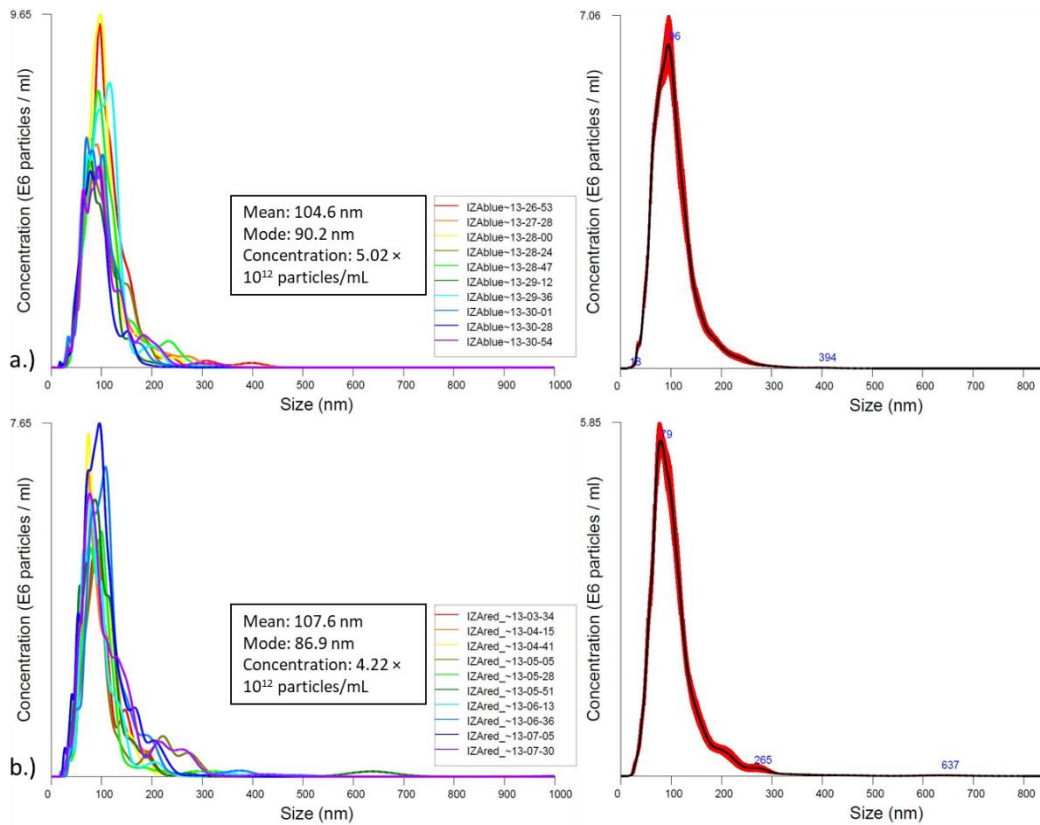
Figure 41. NTA of EVs isolated from 2 mL of HD plasma using SEC+CF method. Graphs on the left are plotted with EV concentration on y-axis and EV size on x-axis for each of the five NTA measurements of a sample. Graphs on the right are plotted with average EV concentration on y-axes and EV size on x-axis for the experiment of each sample. $\pm$ 1 standard error of the mean is represented in red. a.) NTA of EVs from SEC+CF (1) sample, isolated from 2 mL of HD plasma using SEC+CF method. b.) NTA of EVs from SEC+CF (2) sample, isolated from 2 mL of HD plasma using SEC+CF method. NTA=nanoparticle tracking assay, HD=healthy donor, EV=extracellular vesicle, SEC=size exclusion chromatography, CF=centrifugal filtration, n=biological replica, nm=nanometer, ml=milliliter.

Table 7. Summary of NTA and Qubit analysis of EVs isolated from HD plasma using different isolation methods.

EV isolation method	Technical replica no.	Mode (nm)	Mean (nm)	c (EVs) (particles / mL)	c (total proteins) (µg / mL)	Total particle no. in 2 mL of plasma
UC	1	129.9	166.6	8.24×10^{10}	162	1.65×10^{11}
UC	2	117.6	165.9	7.48×10^{10}	120	1.50×10^{11}
SEC+UC	1	137	172.4	7.11×10^{10}	198	1.42×10^{11}
SEC+UC	2	116.9	150	4.92×10^{10}	154	9.84×10^{10}
SEC+CF	1	90.2	104.6	5.02×10^{12}	348	1.00×10^{13}
SEC+CF	2	86.9	107.6	4.22×10^{12}	284	0.84×10^{13}

UC=ultracentrifugation

SEC=size exclusion chromatography

CF=centrifugal filtration

5.7.3 Characterisation of EVs isolated from HD plasma using different isolation methods by *Western blot*

Extracellular vesicles isolated from 2 mL of healthy donor plasma were further characterized for lipoprotein contamination and for the EV markers by *Western blot* analysis. Antibodies against apoliproteins (APO) A and B, anti-APOA and anti-APOB respectively, were used for detection of lipoprotein contamination. Anti-CD9 antibody was used for detection of EVs since CD9 is commonly used as EV marker.

The highest contamination of EVs with APOB has been detected in sample isolated using SEC+CF (Figure 42). The highest contamination of EVs with APOA has been detected in sample isolated using UC only (Figure 42). The strongest signal for CD9 has been detected in sample isolated by SEC+UC, even after short exposure time of the x-ray film of 1 second. After 10 seconds of exposure a double CD9 band has been detected, representing either glycosylation of CD9 or potential isoform (Figure 42).
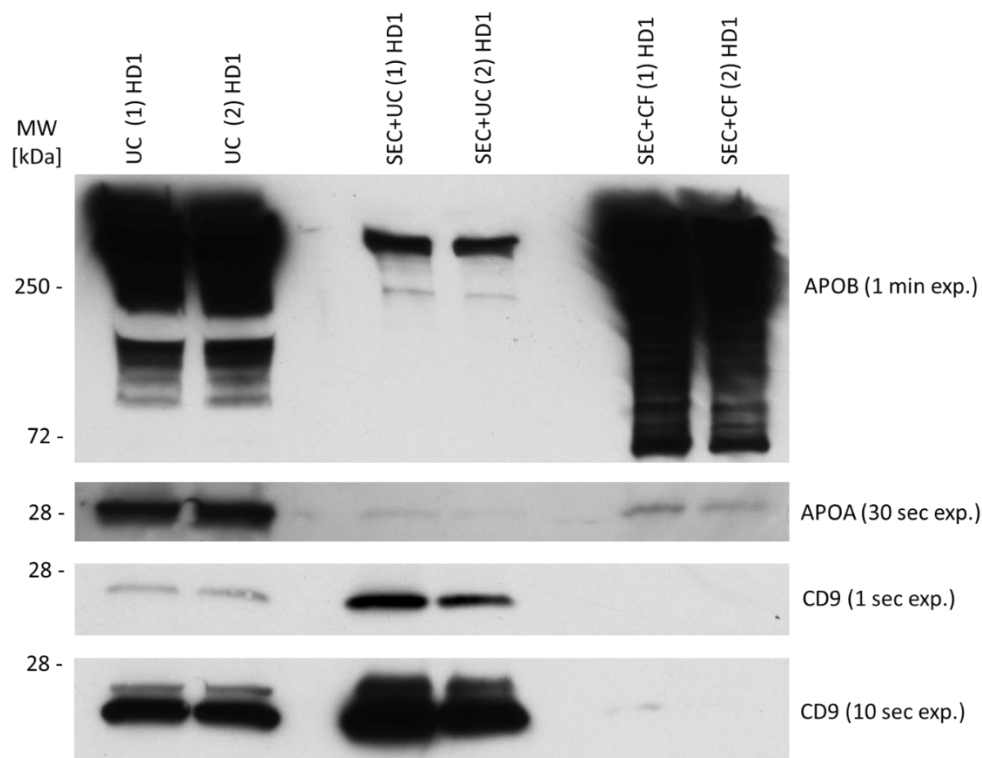
Figure 42. Lipoprotein contamination and EV presence characterization of samples isolated from 2 mL of HD plasma using different isolation methods by *Western blot*. Sample loading was normalized according to the concentration of nanoparticles obtained by NTA and maximum of 35 µL of sample was loaded. n=3. APOA=apoliprotein A, APOB=apoliproteins B, MW=molecular weight, kDa=kilodalton, EV=extracellular vesicle, HD=healthy donor. UC=ultracentrifugation, SEC=size exclusion chromatography, CF=centrifugal filtration, exp.=exposure. n=biological replica.

Taken all together, SEC+UC as EV isolation method showed the highest purity and the highest yield of EVs, compared to EVs isolated using UC only and SEC+CF. Additionally, high particle number and particle concentration derived from NTA is affected by lipoprotein contamination, as indicated by *Western blot* results of UC only and SEC+CF samples. Therefore, SEC+UC has been chosen as a method for EV isolation and further on semi-automated and validated for purity of EVs isolated from lung cancer patient plasma (see below 5.8).

5.8 Validation of SEC+UC EV isolation method on plasma derived from NSCLC patient

Final analysis of EVs includes lung cancer patient-derived plasma and the downstream analysis of isolated EVs using mass spectrometry (section 5.9). Moreover, ITB, UKE, Hamburg, Germany acquired automatic fraction collector during the experimental phase of this thesis which was used for the isolation of EVs from the study cohort comprising healthy controls and lung cancer patients (see section 8.4.3). Therefore, it was needed to test the elution profile of SEC fractions on regard to total protein concentration when using automated fraction collector. For that, plasma derived from NSCLC patient has been processed by SEC using automatic fraction collector and each fraction has been analysed by BCA. In fractions 1-4 total protein concentration is low, and from fraction 5 on the total protein concentration rapidly increases, indicating elution of plasma proteins from fraction 5 on (Figure 43).
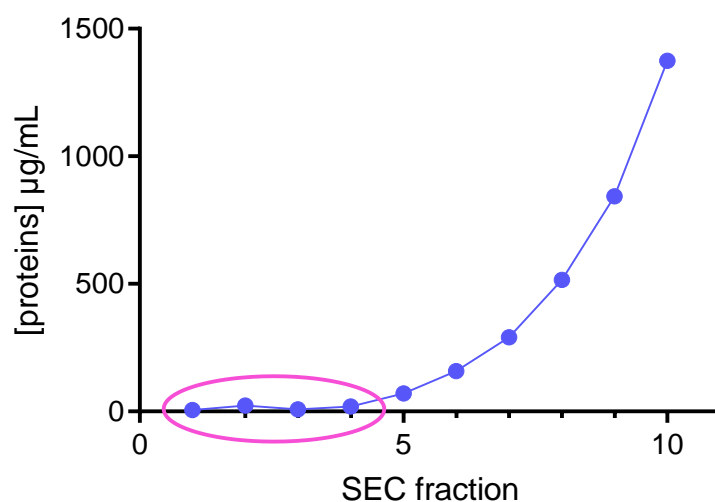
Figure 43. Total protein concentration in SEC fractions derived from NSCLC plasma sample. Fractions 1-4 show low total protein concentration, indicating enrichment of EVs and separation from plasma proteins. n=3. n=biological replica, SEC=size exclusion chromatography, NSCLC=non-small cell lung cancer.

Fractions 1-4 were pooled and further on subjected to ultracentrifugation. Total protein concentration of the sample after SEC+UC has been measured by Qubit assay and it ranged from 766.67-933.00 µg/mL (Table 8). Further on, sample has been applied for EV characterization and contamination inspection by transmission electron microscopy (TEM) (Figure 44 and Table 8 and 9). TEM sample was rich in EVs and EVs were detected with high electron density (Figure 44a and b), indicating high-degree-packed EVs in the sample. In total 449 EVs were analyzed with a mean size of 103.35 nm±20 nm and a diameter range from 30.27–299.99 nm (details specified in Table 8 and 9). EVs were characterized mainly with a round shape surrounded with a structure reminding of a halo (Figure 44c). The presence of cone-shaped EVs has been detected as well (Figure 44d). Interestingly, another membrane-closed structure has been detected inside of the EV, indicating the formation of EV inside of another EV (Figure 45). Any sign of contamination has been poorly detected, classifying the sample as a good-quality-sample.
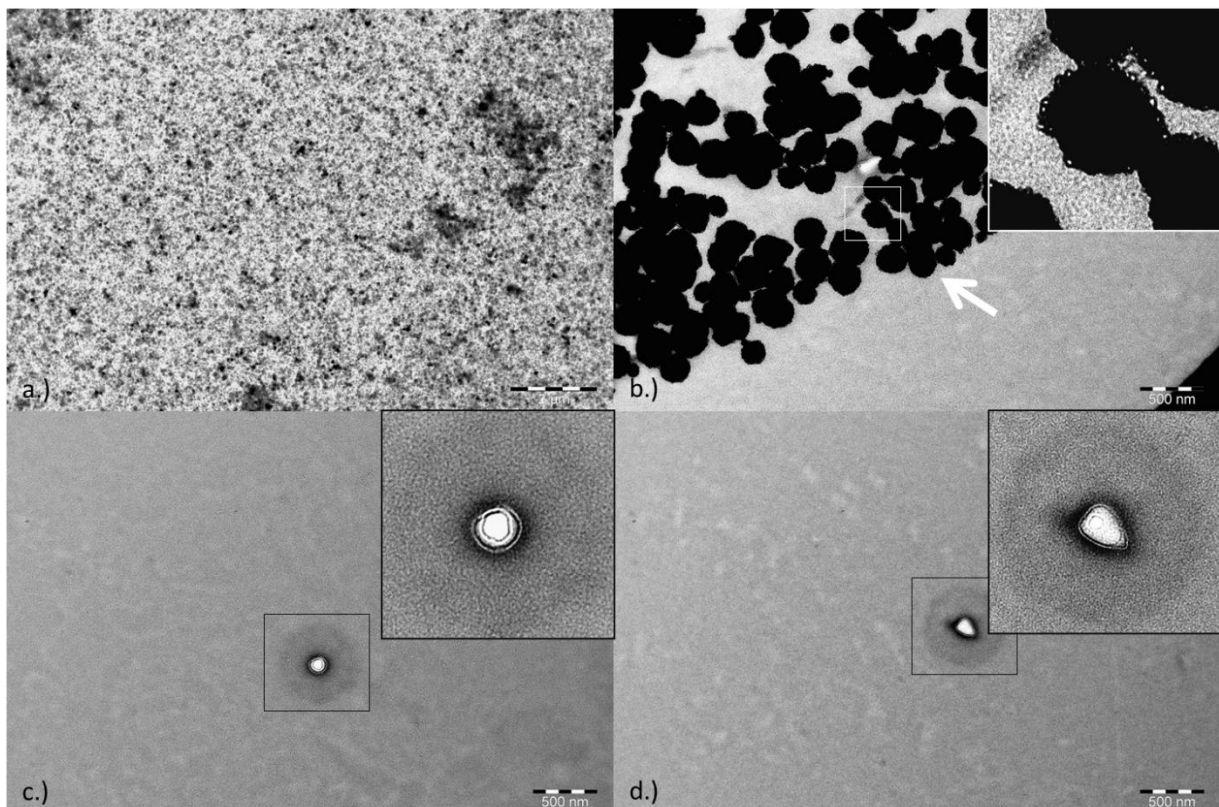
Figure 44. TEM analysis of EVs isolated from NSCLC plasma sample using SEC+UC method. a.) Sample is rich in EVs and presence of contamination has been poorly detected. b.) EVs were detected with a high electron density. c.) Single EV of a circular shape, with clear closed membrane and a halo-like structure around it. d.) Single EV of a cone-like shape, with clear closed membrane and a halo-like structure around it. n=3. n=biological replicas, NSCLC=non-small cell lung cancer, TEM=transmission electron microscopy, EV=extracellular vesicle, SEC=size exclusion chromatography, UC=ultracentrifugation.
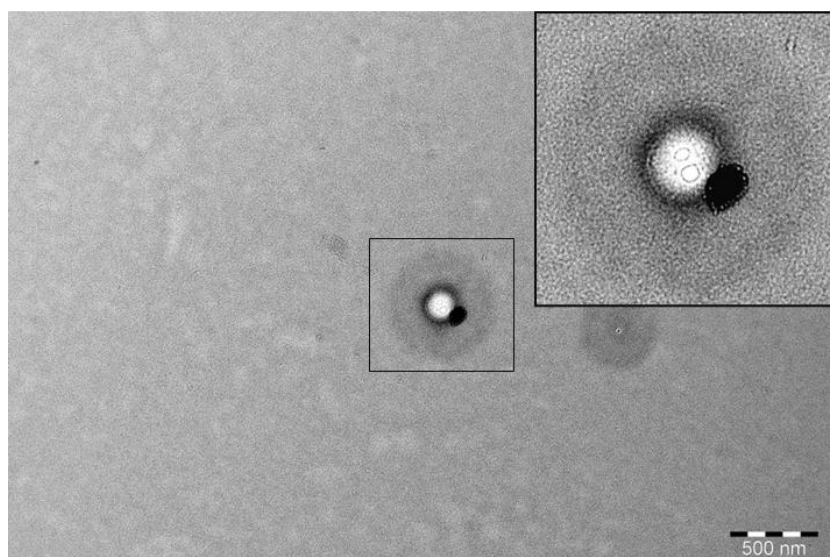
Figure 45. Indication of formation of vesicle inside of another vesicle detected by transmission electron microscopy.

Table 8. Summary of extracellular vesicle characterization by transmission electron microscopy and total protein concentration of each sample analyzed.

Sample	Mean (nm)[a]	Diameter range (nm)	n (EVs) analyzed	c (tot. prot.) (µg/mL)
LC1	112.0 ± 42.0	30.27-218.32	139	813.33
LC2	108.7 ± 50.2	36.27-299.99	202	766.67
LC3	108.7 ± 24.2	38.24-145.26	108	933

LC=lung cancer

[a] mean size in diameter ± standard deviation

Table 9. Summary of transmission electron microscopy characterization of extracellular vesicles isolated from plasma of non-small cell lung cancer patients.

n (NSCLC)	n (EVs)[a]	Mean (nm)[b]	Diameter range (nm)
3	449	103.35 ± 44.18	30.27-299.99

[a] number of total Evs analysed in three NSCLC patient samples

[b] mean size in diameter ± standard deviation

5.9 Proteome analysis of EVs isolated from study cohort using SEC+UC

To identify EV-based biomarker(s) for lung cancer detection, optimized SEC+UC protocol (see section 5.7) was used for the isolation of EVs from plasma and downstream mass spectrometry analysis of EVs has been performed. Acquired MS-data were additionally used for better understanding of the biology behind EVs as well as the potential role of EVs in lung adenocarcinoma carcinogenesis.

Study cohort comprised of 20 healthy controls with equal distribution of males and females, as well as smokers and non-smokers within each sex, was analyzed together with 16 adenocarcinoma lung cancer samples. Extracellular vesicles were isolated using SEC+UC method from 1.5 mL of plasma of each sample within study cohort and further on analyzed by mass spectrometry using Orbitrap Fusion mass spectrometer (8.3.10 M&M section). Raw mass spectrometry data was analyzed using Proteome discoverer (8.7.3 M&M section). In total, 5,435 proteins and 28,189 peptide groups were detected, whereas 4,657 out of 5,435 proteins passed a validity filtering set prior to the database search by Proteome discoverer. Therefore, further analysis was continued with 4,657 proteins which passed the validity test. The Perseus program was used for normalization and statistical analysis of data (8.7.3 and 8.7.4 M&M section). Figures 46 and 47 show log transformation of protein abundancy within each sample before and after normalization of MS data.
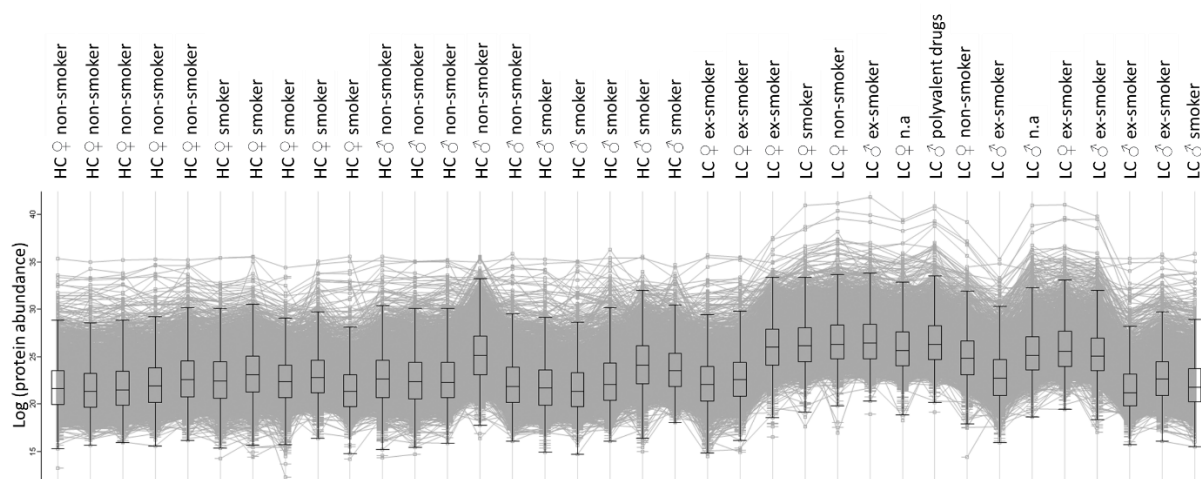
Figure 46. Box plot of log transformed protein abundances within each EV sample before data normalization. n (HC)=20, n (LC)=16. HC=healthy control, LC= adenocarcinoma lung cancer, EV=extracellular vesicle, n.a=unknown smoking status, n=number of individuals analysed.
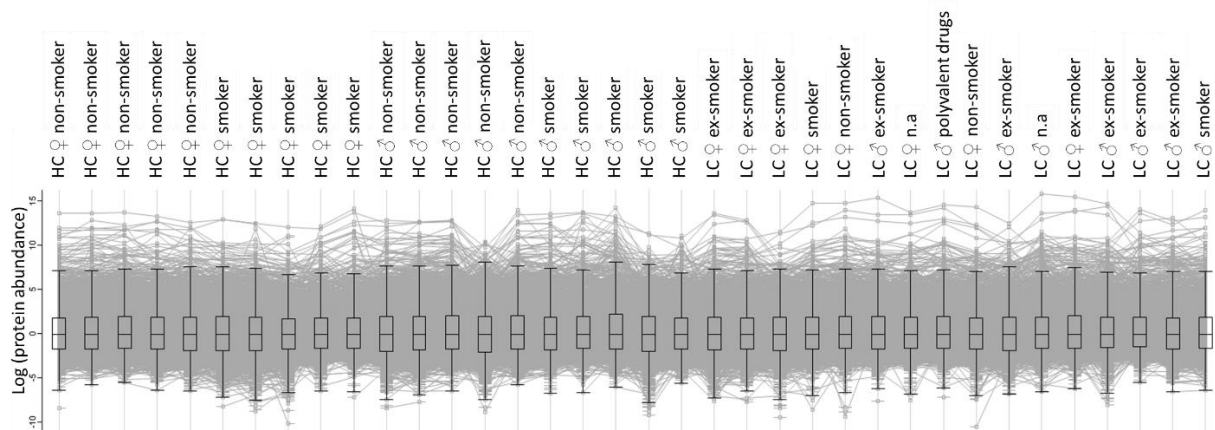
Figure 47. Box plot of log transformed protein abundances within each EV sample after data normalization. n (HC)=20, n (LC)=16. HC=healthy control, LC= adenocarcinoma lung cancer, EV=extracellular vesicle, n.a=unknown smoking status n=number of individuals analysed.

5.9.1 Validation of EV protein enrichment

Normally, APOA and APOB are highly abundant plasma proteins and most common contaminants when isolating EVs. The high abundance of APOA and APOB compared to the abundance of EV proteins within the sample could interfere with the detection of EV-specific proteins. Therefore, the sample quality has been examined through the APOA and APOB abundancies. Abundancy of APOA-I, APOA-V, and APOB were higher than median protein abundance within the sample, and of similar intensity across all samples. APOA-I and APOB were detected with higher abundancy compared to APO-V, whereas the APOA-V abundance was close to median protein abundance in most of the EV samples. (Figure 48). Taken together, the abundancy of APOA and APOB detected in EV samples in this thesis was not sufficient to consider any sample as an *outlier*, and EV isolation showed good quality.
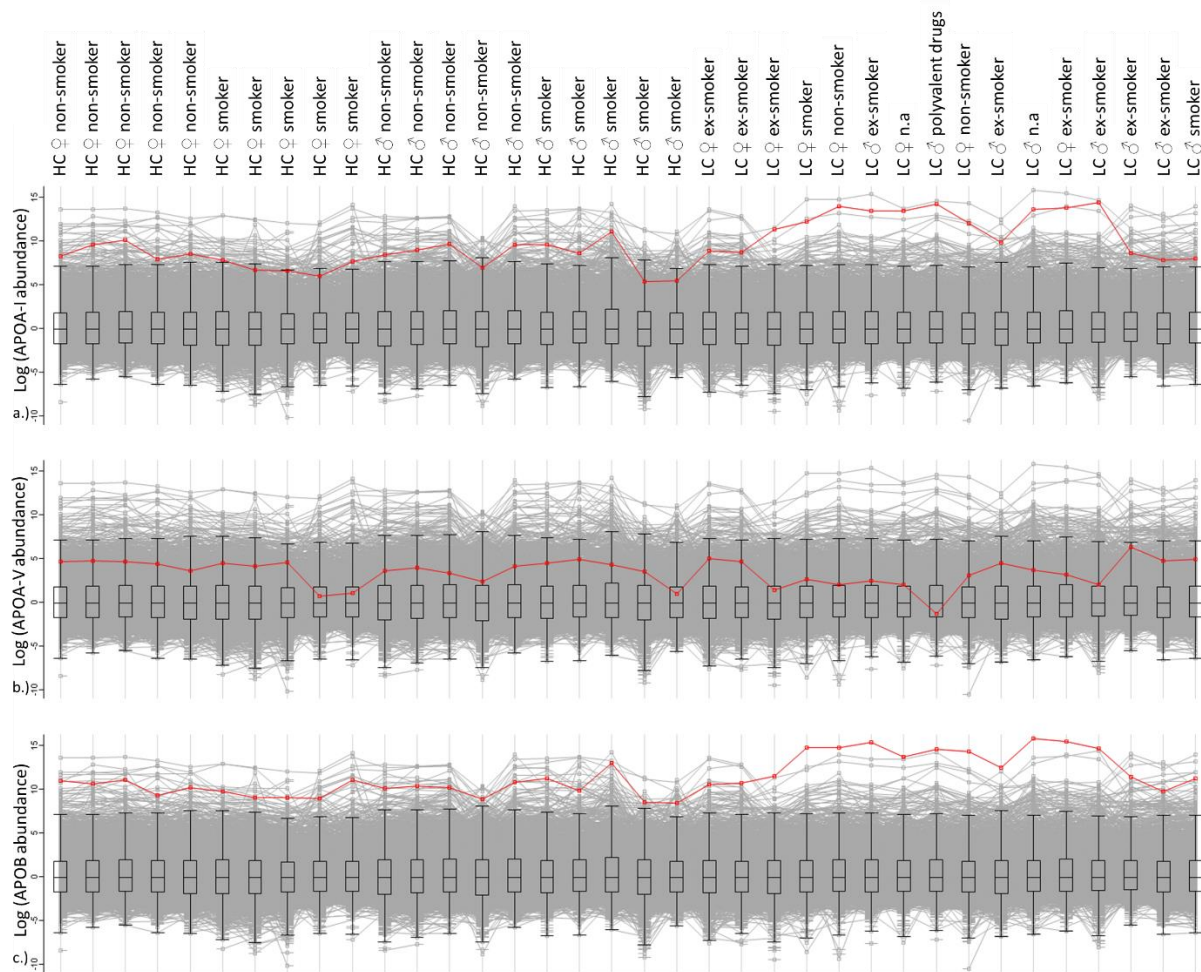
Figure 48. Box plot of log transformed protein abundances of apoliproteins within each EV sample after data normalization. a.) abundances of APOA-I across all EV samples from study cohort. b.) abundances of APOA-V across all EV samples from study cohort. c.) abundances of APOB across all EV samples from study cohort. n (HC)=20, n (LC)=16. HC=healthy control, LC= adenocarcinoma lung cancer, EV=extracellular vesicle, APOA=apoliprotein A, APOB=apoliprotein B, n=number of individuals analysed.

Additionally, as inspection of the EV protein enrichment in samples, abundance of the following EV markers has been examined: CD9, CD81, CD63 (Figure 49), Flotilin-1 (FLOT-1), Flotilin-2 (FLOT-2), Programmed cell death 6-interacting protein (ALIX), Syntenin-1 (SDCBP), and Tumor susceptibility gene 101 protein (TSG101) (Figure S2). Moreover, clathrin is a protein associated with the assembly of EVs. Therefore, abundance of clathrin has been examined as an additional inspection of EV-specific protein enrichment (Figure 50). CD9 has been detected in all samples with higher abundance compared to median protein abundance in most of the samples (30/36 samples) (Figure 49a). Comparing to the abundance of APO-V in Figure 48b, the abundance of CD9 is similar to the abundance of APO-V, indicating an enrichment of EV proteins compared to plasma contaminant proteins in general. CD81 has been detected in all samples with higher abundance compared to median protein abundance in most of

the samples (27/36 samples) (Figure 49b). CD63 has been detected in 31/36 samples with higher abundance compared to median protein abundance in 15/36 samples (Figure 49c).

In previous experiments presented in sections 5.7.1, CD9 and CD63 sandwich ELISAs showed higher concentration of CD9 in plasma samples compared to CD63 concentrations. Additionally, a wider elution profile of CD63+ EVs compared to CD9+ EVs could have been observed across SEC fractions (Figure 37 and 38). Taken that together, even though CD63 has not been detected by mass spectrometry in 5 samples, whereas CD9 and CD81 were, no sample was considered as an *outlier* and further analysis was performed including all samples.
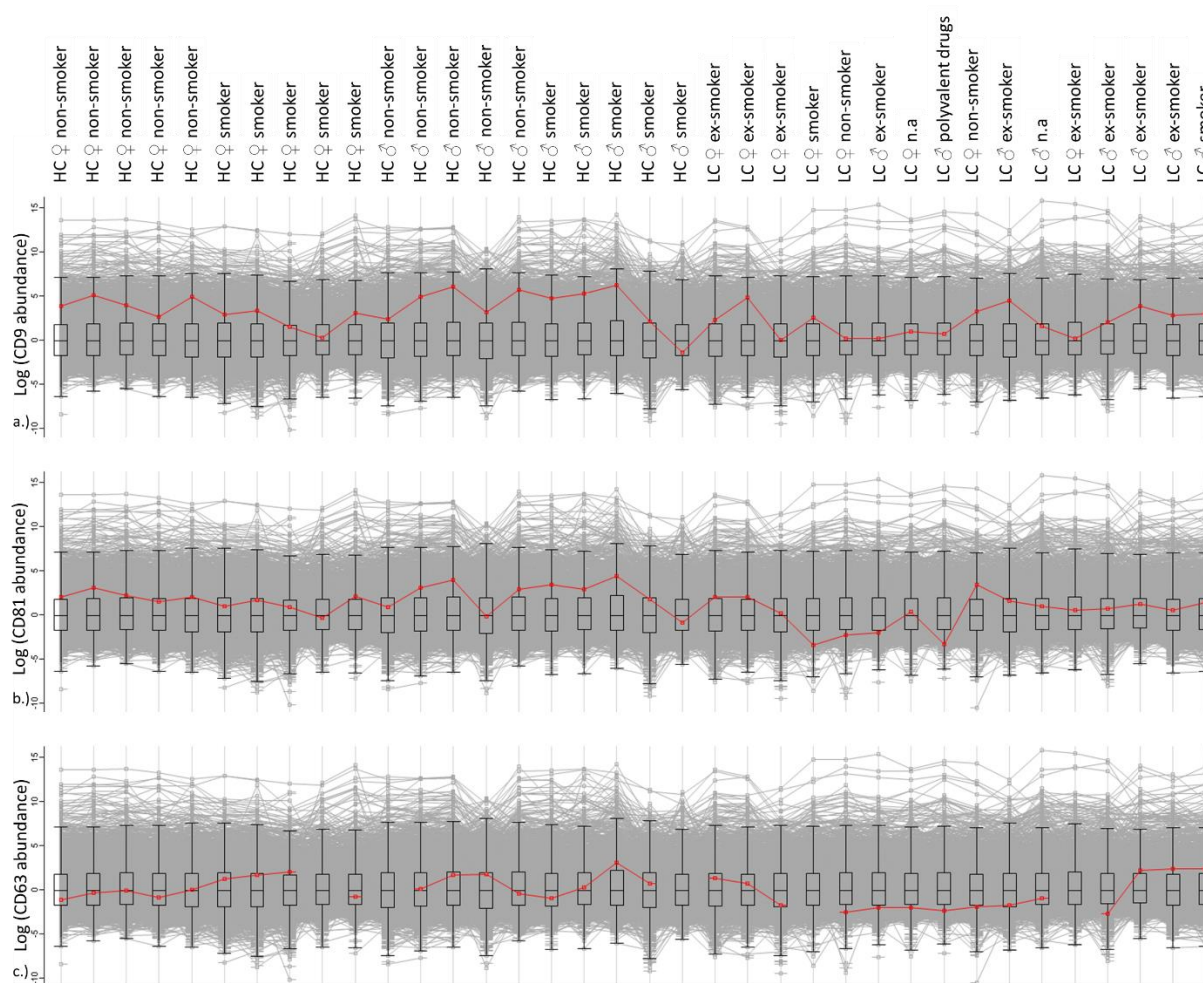
Figure 49. Box plot of log transformed protein abundances of EV markers within each EV sample after data normalization. a.) abundances of CD9 across all EV samples from study cohort. b.) abundances of CD81 across all EV samples from study cohort. c.) abundances of CD63 across all EV samples from study cohort. n (HC)=20, n (LC)=16. HC=healthy control, LC= adenocarcinoma lung cancer, EV=extracellular vesicle, n=number of individuals analysed.

As a further validation of successful EV isolation and enrichment of EV proteins compared to plasma derived contaminants, samples were analyzed for clathrin abundance. Clathrin heavy chain 1 (CLH1)

71

has been detected in all samples with similar abundance across all samples. Additionally, CLH1 abundance is higher compared to median protein abundance (Figure 50a) and APOA-V abundance in all samples, whereas the abundances of APOA-I and APOB are slightly higher compared to abundances of CLH1 (Figures 50 and 48a-c). Clathrin light chain 1 (CLCA) has been detected in 35/36 samples with higher abundance compared to median protein abundance in 28/36 samples (Figure 50b)
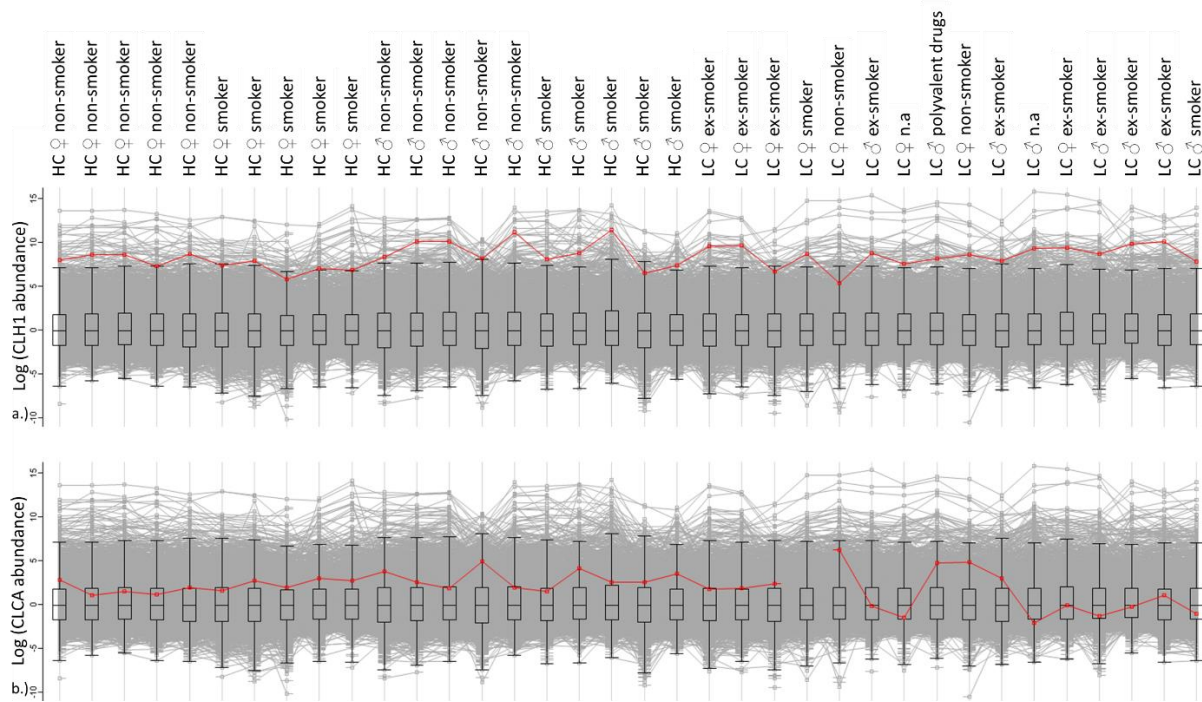
Figure 50. Box plot of log transformed protein abundances of CLH1 and CLCA1 within each EV sample after data normalization. a.) abundances of CLH1 across all EV samples from study cohort. b.) abundances of CLCA1 across all EV samples from study cohort. n (HC)=20, n (LC)=16. HC=healthy control, LC= adenocarcinoma lung cancer, EV=extracellular vesicle, CLH1=clathrin heavy chain 1, CLCA=clathrin light chain 1, n=number of individuals analysed.

5.9.2 Principal component analysis of EV derived proteins

For principal component analysis, 4,657 of total proteins detected were reduced to only those proteins present in 100% of samples. Reduction resulted in 729 proteins for the principal component analysis and no batch effect correction was required. Disease status specific signature appears to be more evident for clustering, while sex and smoking effect did not dominate clustering (Figure 51).
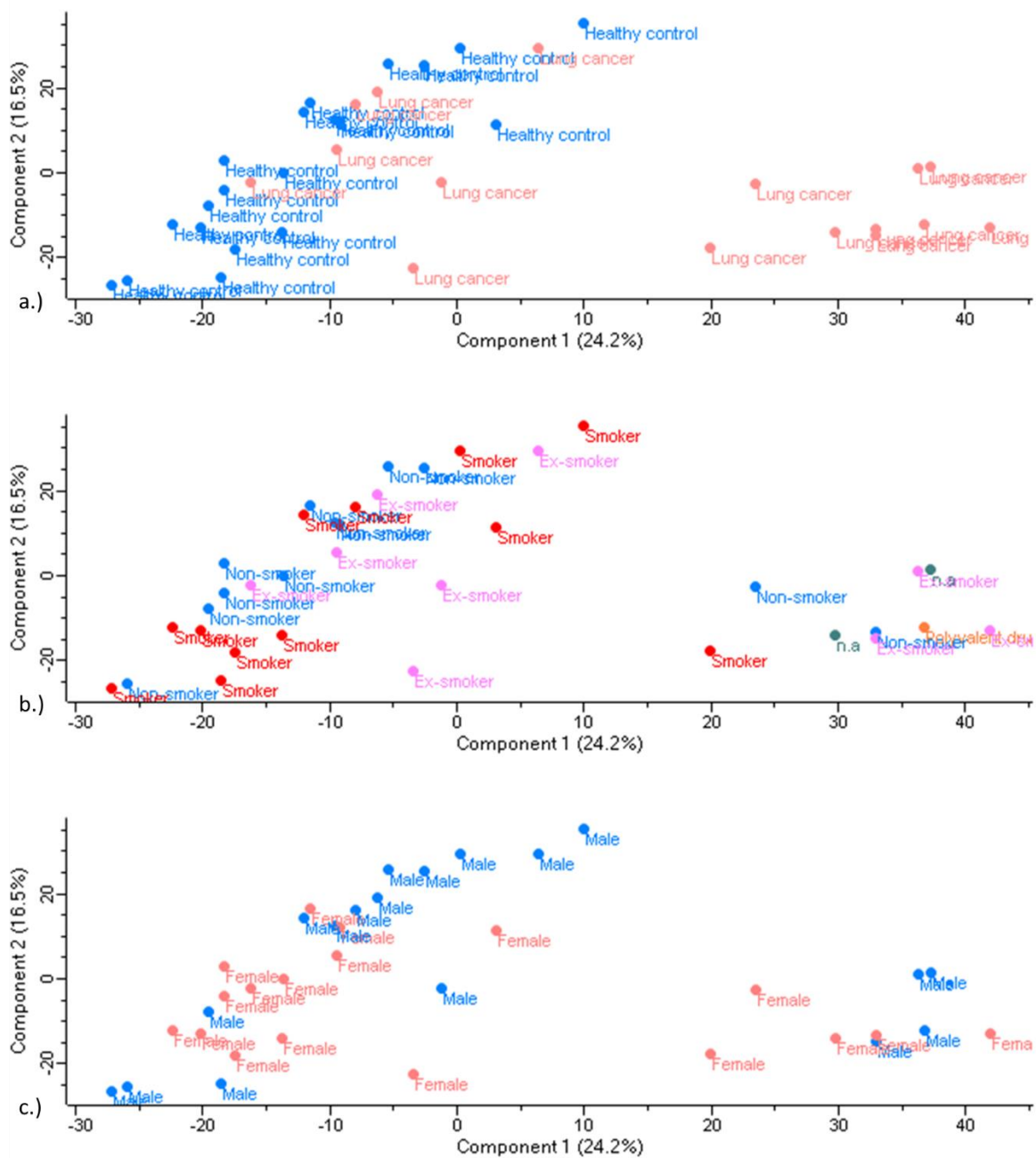
Figure 51. Principal component analysis based on 729 proteins present in 100% of samples. a.) Principal component analysis based on component 1 and component 2 showed disease status specific signature evident for clustering. b.) Principal component analysis based on component 1 and component 2 showed no domination of smoking effect for clustering. c.) Principal component analysis based on component 1 and component 2 show no domination of gender effect for clustering. n.a=polyvalent drugs.

5.9.3 Differential expression of plasma derived EV proteins

From the total EV proteins detected in 1.5 mL of healthy control- and adenocarcinoma lung cancer-derived plasma samples, t-test analysis resulted in 970 differentially expressed proteins between healthy controls and adenocarcinoma lung cancer EVs based on p-values ($p<0.05$). Adding fold change as a cut

of for significance reduced the number of differentially expressed proteins between healthy control- and adenocarcinoma lung cancer patient-EVs to 552 (1.5<FDR<-1.5) (Figure 52). Among *p*-value significant proteins, 10 proteins elevated in healthy controls compared to adenocarcinoma lung cancer, and 10 proteins elevated in adenocarcinoma lung cancer compared to healthy controls with the highest fold change are listed in Table 10.

Figure 52. Volcano plot of proteins detected in healthy control- and adenocarcinoma lung cancer-derived EVs isolated from 1.5 mL of human plasma. Proteins marked in blue are elevated in healthy control EVs compared to adenocarcinoma lung cancer patient EVs (*p*<0.05 and FDR>1.5). Proteins marked in red dots are elevated in adenocarcinoma lung cancer patient EVs compared to healthy control EVs (*p*<0.05 and FDR<-1.5). Proteins marked in grey and black dots are insignificant either based on *p*-value (grey dots; *p*≥0.05) or based on fold change (black dots, 1.5≥FDR≥-1.5). *p*-value=student's t-test. For details, see Table 10.

Table 10. Ten proteins elevated in healthy control EVs or in adenocarcinoma lung cancer patient EVs with the highest fold change between compared groups.

Gene name	*p* -value[a]	FDR[b]	Elevated in:
OSCP1	0.003	-9.13	Lung cancer EVs
CACNB3	0.034	-5.54	Lung cancer EVs
WDFY1	0.030	-5.47	Lung cancer EVs
KCNC3	0.050	-4.90	Lung cancer EVs
CEP170	0.000	-4.73	Lung cancer EVs
GUK1	0.005	-4.66	Lung cancer EVs
RUNDC3A	0.000	-4.38	Lung cancer EVs
SAMD11	0.004	-4.31	Lung cancer EVs
SLC12A6	0.000	-4.04	Lung cancer EVs
NSDHL	0.023	-3.93	Lung cancer EVs
KRTAP3-3	0.012	6.33	Healthy control EVs
NT5C3B	0.000	6.03	Healthy control EVs
MBP	0.000	4.94	Healthy control EVs
MUC19	0.001	4.82	Healthy control EVs
PRSS1	0.000	4.25	Healthy control EVs
CHDH	0.001	4.24	Healthy control EVs
IREB2	0.008	4.05	Healthy control EVs
MSN	0.000	3.88	Healthy control EVs
RAD23B	0.000	3.83	Healthy control EVs
VAT1	0.004	3.59	Healthy control EVs

[a] *p* -value=Student's *t* -test

[b] FDR=fold chancge expressed in log2 scale

Further on, differential expression of proteins between healthy controls and adenocarcinoma lung cancer patients with respect to their smoking status has been analysed by ANOVA resulting in 793 ANOVA significant proteins. Heatmap visualisation shows clustering of healthy controls against adenocarcinoma lung cancer patients. Four healthy control smokers cluster together with adenocarcinoma lung cancer patients (Figure 53).
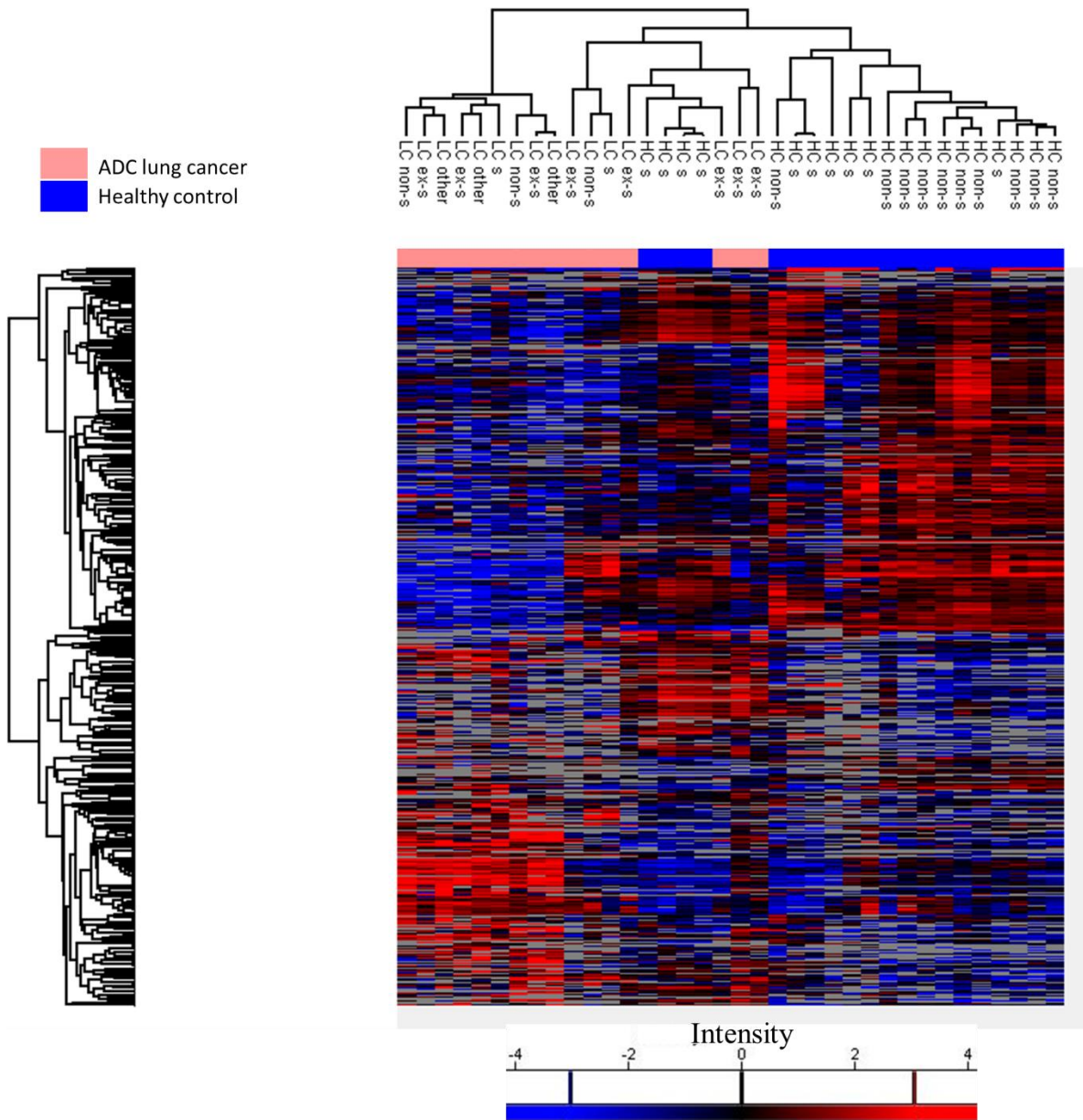
Figure 53. Heatmap visualisation of 793 ANOVA significant EV proteins with respect to disease and smoking status of healthy controls and adenocarcinoma lung cancer patients. Healthy controls cluster together (blue bars) against adenocarcinoma lung cancer patients (peach bars), whereas some healthy control smokers cluster together with adenocarcinoma lung cancer patients. LC= adenocarcinoma lung cancer, HC=healthy control, s=smoker.

Additionally, adenocarcinoma lung cancer patient EV proteome has been compared to EV proteome of each healthy control subgroup, healthy control non-smokers and healthy control smokers. Between healthy control non-smokers and adenocarcinoma lung cancer patients, 970 proteins were differentially expressed based on p-value ($p<0.05$). Adding fold change as a cut-off for significance reduced the number of differentially expressed proteins between healthy control non-smoker- and adenocarcinoma lung cancer patient-EVs to 685 ($1.5<FDR<-1.5$) (Figure 54). Among p-value significant proteins, 10 proteins elevated in healthy control non-smokers compared to lung cancer, and 10 proteins elevated in

adenocarcinoma lung cancer compared to healthy control non-smokers with the highest fold change are listed in Table 11.

Figure 54. Volcano plot of proteins detected in healthy control non-smoker- and adenocarcinoma lung cancer-derived EVs isolated from 1.5 mL of human plasma. Proteins marked in blue are elevated in healthy control non-smoker EVs compared to adenocarcinoma lung cancer patient EVs ($p<0.05$ and FDR>1.5). Proteins marked in red dots are elevated in adenocarcinoma lung cancer patient EVs compared to healthy control non-smoker EVs ($p<0.05$ and FDR<-1.5). Proteins marked in grey and black dots are insignificant either based on p-value (grey dots; $p\geq0.05$) or based on fold change (black dots, $1.5\geq$FDR≥-1.5). p-value=student's t-test. For details, see Table 11.

Table 11. Ten proteins elevated in healthy control non-smoker EVs or in adenocarcinoma lung cancer patient EVs with the highest fold change between compared groups.

Gene name	p-value[a]	FDR[b]	Elevated in:
OSCP1	0.003	-9.13	Lung cancer EVs
SAMD11	0.040	-5.81	Lung cancer EVs
RUNDC3A	0.002	-5.57	Lung cancer EVs
GUK1	0.026	-5.01	Lung cancer EVs
HOGA1	0.014	-4.75	Lung cancer EVs
SLC12A6	0.000	-4.73	Lung cancer EVs
SEPHS2	0.000	-4.60	Lung cancer EVs
NSDHL	0.022	-4.53	Lung cancer EVs
ACTR10	0.021	-4.41	Lung cancer EVs
C8A	0.001	-4.36	Lung cancer EVs
MUC19	0.000	6.51	Healthy control non-smoker EVs
NT5C3B	0.000	6.33	Healthy control non-smoker EVs
KRTAP3-3	0.049	5.88	Healthy control non-smoker EVs
VAT1	0.000	5.73	Healthy control non-smoker EVs
MBP	0.000	5.58	Healthy control non-smoker EVs
AHCYL1	0.001	4.67	Healthy control non-smoker EVs
PDCD10	0.000	4.43	Healthy control non-smoker EVs
PRSS1	0.005	4.42	Healthy control non-smoker EVs
AFP	0.000	4.27	Healthy control non-smoker EVs
GPM6A	0.000	4.17	Healthy control non-smoker EVs

[a] p-value=Student's t-test

[b] FDR=fold chancge expressed in log2 scale

Between healthy control smokers and adenocarcinoma lung cancer patients, 554 proteins were differentially expressed based on p-value ($p<0.05$). Adding fold change as a cut-off for significance reduced the number of differentially expressed proteins between healthy control smoker- and adenocarcinoma lung cancer patient-EVs to 372 (1.5<FDR<-1.5) (Figure 55). Among p-value significant proteins, 10 proteins elevated in healthy control smokers compared to adenocarcinoma lung cancer, and 10 proteins elevated in adenocarcinoma lung cancer compared to healthy control smokers with the highest fold change are listed in Table 12.
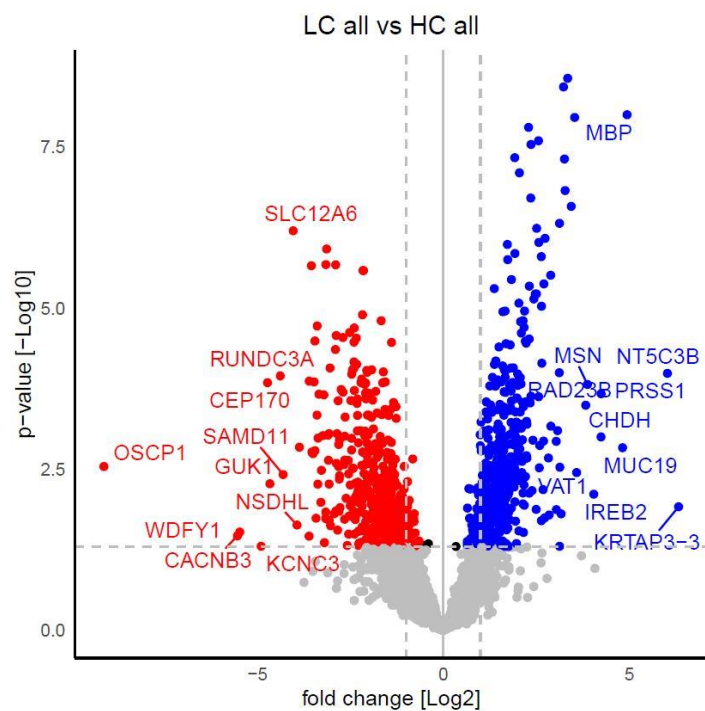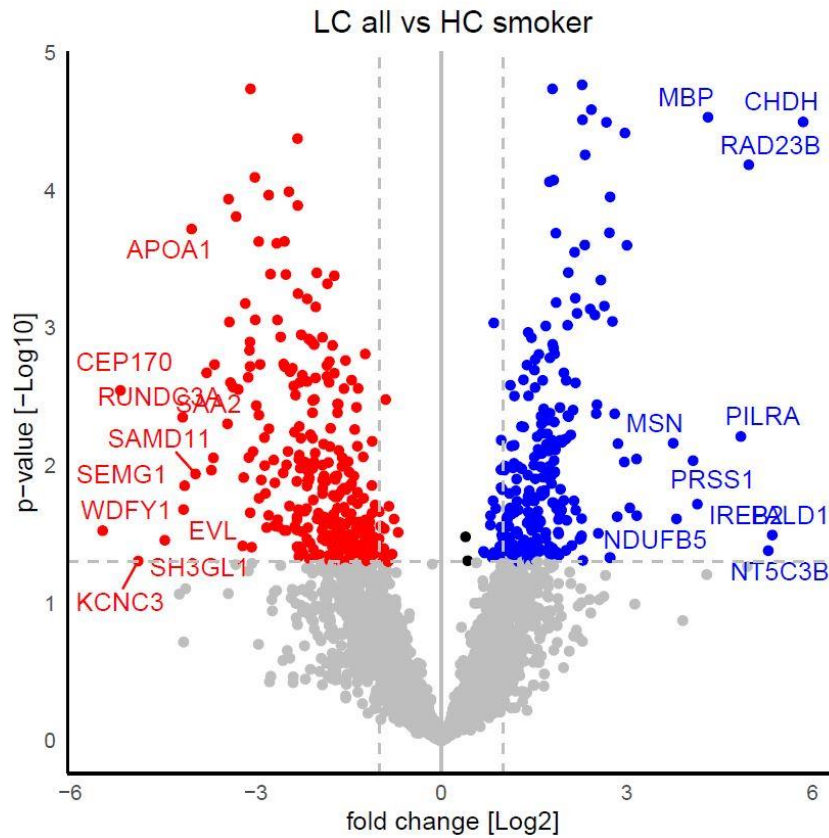
Figure 55. Volcano plot of proteins detected in healthy control smoker- and adenocarcinoma lung cancer-derived EVs isolated from 1.5 mL of human plasma. Proteins marked in blue are elevated in healthy control smoker EVs compared to adenocarcinoma lung cancer patient EVs ($p<0.05$ and FDR>1.5). Proteins marked in red dots are elevated in adenocarcinoma lung cancer patient EVs compared to healthy control smoker EVs ($p<0.05$ and FDR<-1.5). Proteins marked in grey and black dots are insignificant either based on p-value (grey dots; $p\geq0.05$) or based on fold change (black dots, $1.5\geq$ FDR$\geq$-1.5). p-value=student's t-test. For details, see Table 12.

Table 12. Ten proteins elevated in healthy control smoker EVs or in adenocarcinoma lung cancer patient EVs with the highest fold change between compared groups.

Gene name	p-value[a]	FDR[b]	Elevated in:
WDFY1	0.030	-5.47	Lung cancer EVs
CEP170	0.003	-5.18	Lung cancer EVs
KCNC3	0.050	-4.90	Lung cancer EVs
SH3GL1	0.035	-4.47	Lung cancer EVs
SAA2	0.004	-4.18	Lung cancer EVs
EVL	0.021	-4.17	Lung cancer EVs
SEMG1	0.014	-4.15	Lung cancer EVs
APOA1	0.000	-4.04	Lung cancer EVs
SAMD11	0.012	-3.97	Lung cancer EVs
RUNDC3A	0.002	-3.79	Lung cancer EVs
CHDH	0.000	5.85	Healthy control smoker EVs
PALD1	0.032	5.35	Healthy control smoker EVs
NT5C3B	0.042	5.29	Healthy control smoker EVs
RAD23B	0.000	4.97	Healthy control smoker EVs
PILRA	0.006	4.84	Healthy control smoker EVs
MBP	0.000	4.31	Healthy control smoker EVs
IREB2	0.019	4.14	Healthy control smoker EVs
PRSS1	0.009	4.07	Healthy control smoker EVs
NDUFB5	0.024	3.80	Healthy control smoker EVs
MSN	0.007	3.75	Healthy control smoker EVs

[a] p-value=Student's t-test

[b] FDR=fold chancge expressed in log2 scale

Since smoking is a major contributor for developing lung cancer, proteome difference between healthy control non-smokers and healthy control smokers was analyzed as well. Different protein pattern between healthy control-smokers and -non-smokers could provide insight into the group of proteins associated with smoking-related changes but not necessarily involved in tumorigenesis. Such information could be valuable for the reduction of false-positive cases.

Between healthy control non-smokers and healthy control smokers, 526 proteins were differentially expressed based on p-value ($p<0.05$). Adding fold change as a cut-off for significance reduced the number of differentially expressed proteins between healthy control non-smoker- and healthy control smoker-EVs to 291 (1.5<FDR<-1.5) (Figure 56). Among p-value significant proteins, 10 proteins elevated in healthy control non-smokers compared to healthy control smokers, and 10 proteins elevated in healthy control smokers compared to healthy control non-smokers with the highest fold change are listed in Table 13.
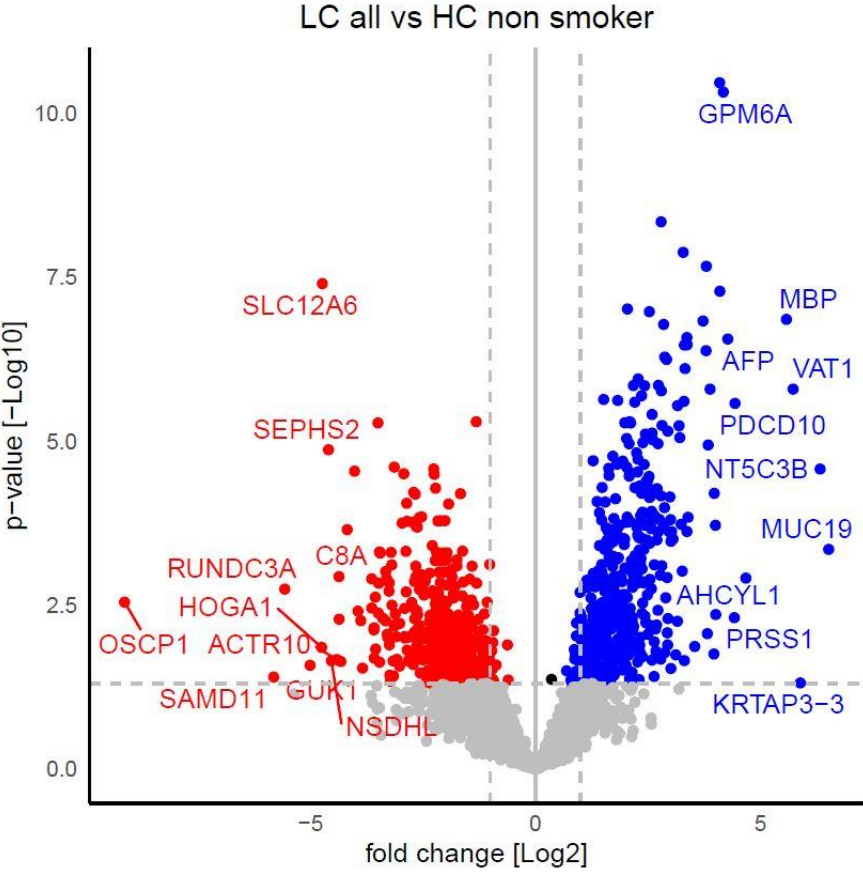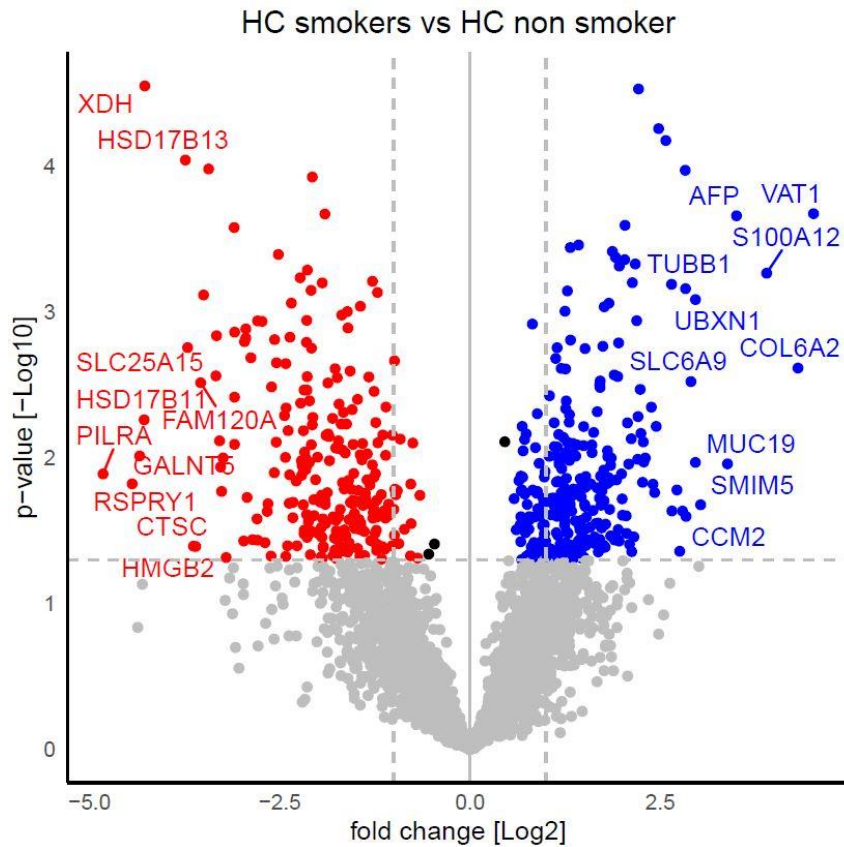
Figure 56. Volcano plot of proteins detected in healthy control non-smoker- and healthy control smoker-derived EVs isolated from 1.5 mL of human plasma. Proteins marked in blue are elevated in healthy control non-smoker EVs compared to healthy control smoker EVs ($p<0.05$ and FDR>1.5). Proteins marked in red dots are elevated in healthy control smoker EVs compared to healthy control non-smoker EVs ($p<0.05$ and FDR<-1.5). Proteins marked in grey and black dots are insignificant either based on p-value (grey dots; $p\geq0.05$) or based on fold change (black dots, 1.5$\geq$FDR$\geq$-1.5). p-value=student's t-test. For details, see Table 13.

Table 13. Ten proteins elevated in healthy control non-smoker EVs or in healthy control smoker EVs with the highest fold change between compared groups.

Gene name	p-value[a]	FDR[b]	Elevated in:
PILRA	0.013	-4.82	Healthy control smoker EVs
RSPRY1	0.015	-4.44	Healthy control smoker EVs
GALNT5	0.010	-4.34	Healthy control smoker EVs
HSD17B11	0.005	-4.28	Healthy control smoker EVs
XDH	0.000	-4.27	Healthy control smoker EVs
HSD17B13	0.000	-3.74	Healthy control smoker EVs
SLC25A15	0.002	-3.71	Healthy control smoker EVs
CTSC	0.040	-3.63	Healthy control smoker EVs
HMGB2	0.041	-3.60	Healthy control smoker EVs
FAM120A	0.003	-3.54	Healthy control smoker EVs
VAT1	0.000	4.51	Healthy control non smoker EVs
COL6A2	0.002	4.31	Healthy control non smoker EVs
S100A12	0.001	3.90	Healthy control non smoker EVs
AFP	0.000	3.50	Healthy control non smoker EVs
MUC19	0.011	3.38	Healthy control non smoker EVs
SMIM5	0.021	3.03	Healthy control non smoker EVs
UBXN1	0.001	2.96	Healthy control non smoker EVs
SLC6A9	0.003	2.90	Healthy control non smoker EVs
CCM2	0.025	2.84	Healthy control non smoker EVs
TUBB1	0.001	2.83	Healthy control non smoker EVs

[a] p-value=Student's t-test

[b] FDR=fold chancge expressed in log2 scale

5.9.4 Potential putative EV biomarker panel for lung cancer detection

Early lung cancer detection remains challenging even nowadays, and patients are most often diagnosed at late stages when disease has already advanced. In this thesis, the EV proteome was analysed for the detection of novel EV-derived lung cancer biomarkers. For that purpose, 7 differential EV proteins with the highest fold change between adenocarcinoma patients and healthy controls were chosen as potential EV-derived biomarker panel for early detection of lung cancer, adenocarcinoma, respectively. EV-proteins suggested as components of the EV-biomarker panel are independent of smoking status (Figure 57), since the smoking status showed high impact on EV proteome and differential expression between adenocarcinoma patients and healthy controls (5.9.3).

Proteins serum amyloid A2 (SAA2), DNA repair protein REV1 (REV1), selenocysteine lyase (SCLY), and solute carrier family 12 member 6 (SLC12A6) are more abundant on the EVs deriving from adenocarcinoma patients compared to EVs deriving from healthy controls, and have the highest fold change difference independent of the smoking status of the participants analyzed (Figure 57). Contrary, serine protease 1 (PRSS1), neuronal membrane glycoprotein M6-a (GPM6A), and moesin (MSN) are

more abundant on the EVs deriving from healthy controls compared to the EVs deriving from adenocarcinoma patients, independent of the smoking status of the participants (Figure 57). Proteins SAA2, REV1, SCLY, SLC12A6, PRSS1, GPM6A, and MSN could be putative signature for the adenocarcinoma lung cancer detection.
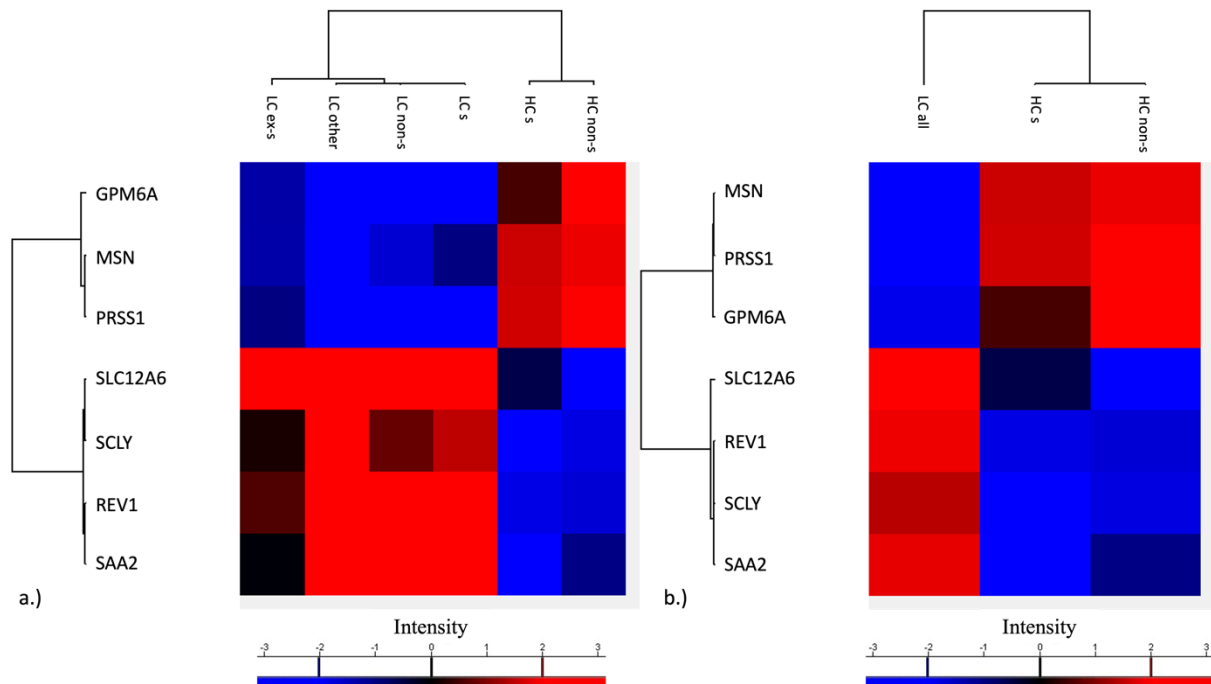
Figure 57. Potential putative EV-biomarker panel for the detection of adenocarcinoma lung cancer. a.) Seven differentially abundant proteins in adenocarcinoma lung cancer-derived EVs compared to healthy controls-derived EVs with the highest fold change independent of smoking status. b.) Seven differentially abundant proteins in adenocarcinoma lung cancer-derived EVs compared to healthy controls-derived EVs with the highest fold change independent of smoking status. Adenocarcinoma lung patient groups are joined together. es=ex-smoker; o=other; ns=non-smoker; s=smoker.

5.9.5 Pathway analysis of extracellular vesicles

To briefly gain insight into the biology behind EVs, pathway analysis of three gene otology terms has been performed. Biological process, cellular component, and molecular function gene ontology terms have been analysed for proteins enriched on EVs deriving from healthy controls, as well as for proteins enriched on EVs deriving from adenocarcinoma lung cancer patients.

Proteins enriched ($p<0.05$; FDR>1.5) in EVs deriving from healthy controls were the most associated with biological processes related to metabolism, localization, as well as with translational elongation, biogenesis of the cellular components and regulation of biological quality. On regard to the cellular component, proteins enriched in healthy control derived EVs are the most related with vesicular component, as well as cytoplasm, intracellular component, and macromolecular complex. Molecular function of healthy-control-deriving-EV-proteins is associated the most with enzymatic activity as well

as with signal transduction, protein transport, growth regulation, and polypeptide chain elongation (Figure 58).
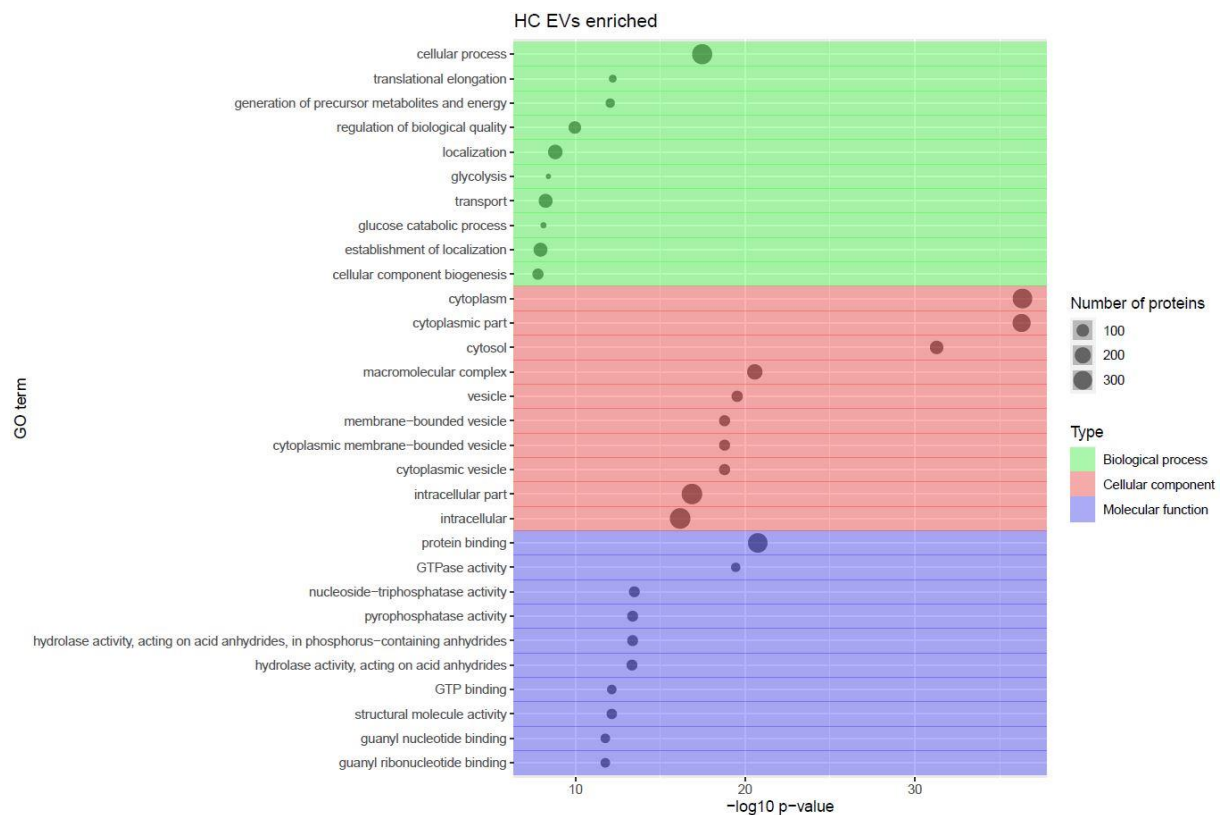
Figure 58. Enriched pathways in EVs isolated from plasma of healthy controls based on gene ontology analysis including biological process, cellular component, and molecular function gene ontology terms.

Proteins enriched ($p<0.05$; FDR<-1.5) in EVs deriving from adenocarcinoma lung cancer patients were the most associated with biological processes related to the immune system response, as well as the protein maturation and protein processing, and regulation of response to the stimulus. On regard to the cellular component, proteins enriched in adenocarcinoma lung cancer derived EVs are the most associated with cytoplasmic and extracellular regions, as well as with complexes associated with lipids. Functionally, adenocarcinoma-EV-deriving-proteins were the most associated with binding, followed by enzymatic regulation, as well as with the activation of phosphatidylcholine-sterol O-acyltransferase activity (Figure 59).
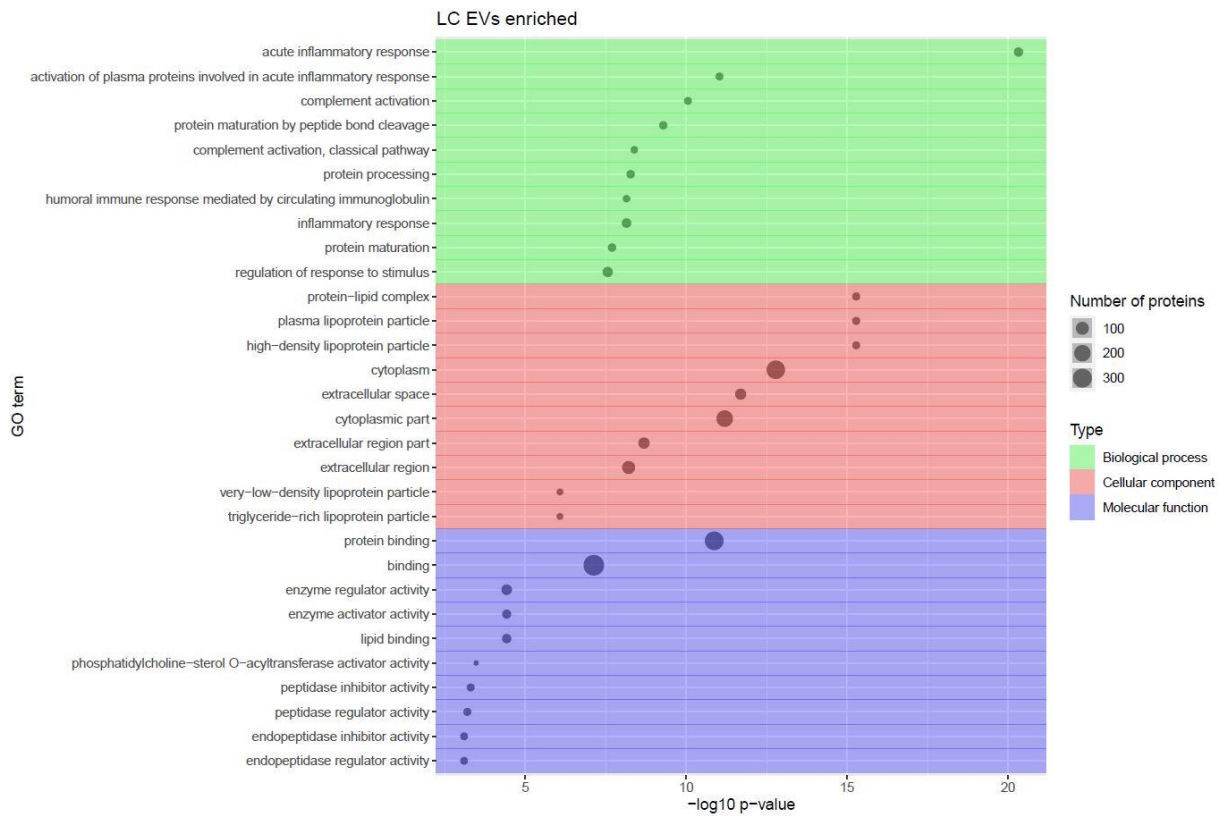
Figure 59. Enriched pathways in EVs isolated from plasma of adenocarcinoma lung cancer patients based on gene ontology analysis including biological process, cellular component, and molecular function gene ontology terms.

6. Discussion

Lung cancer has the highest mortality rate compared to other cancer entities. High mortality rate is mostly due to lung cancer being an asymptomatic disease until very late stage when patients are the most often diagnosed and the disease itself is already advanced. Therefore, patient treatment is more difficult (134). Detection of lung cancer has achieved improvement over the years, mainly with the implementation of different risk models for the assessment of high-risk patients like smokers, exposure to fumes or asbestos (135). However, up to date the only screening method for lung cancer detection is LDCT which has very high false-positive rate and it is often followed up with invasive bronchoscopy and tissue biopsy (134). For the precise diagnosis, bronchoscopy and tissue biopsy also depend on the experienced specialist (bronchoscopist or pulmonologist). Therefore, the detection of lung cancer, especially in early stages, remains challenging with an urgent need for the development of novel biomarkers for the screening of high-risk population and the detection of lung cancer.

Field of liquid biopsy has been introduced in 2010 (66), and up until today it has made a remarkable progress (9, 71, 136-140). As being minimally invasive method, the use of liquid biopsy in lung cancer detection could overcome the limitations of tissue biopsy. Field of liquid biopsy allows multiple sampling and precise follow up of the patient, as well as sampling when tissue biopsy is not possible due to the tumor location or when the patient itself is in poor condition. Many different analytes can be utilized in liquid biopsy as discussed elsewhere (141-143), including serological biomarkers and extracellular vesicles which were also the focus of this thesis.

Advantage of the utility of serological biomarkers is their cost-effectiveness compared to imaging and they are adequate for mass screening (144). With the use of highly sensitive proteomic analysis, LC-MS/MS coupled with SILAC, NECT-4 and PDAP1 proteins have been detected as potential serological biomarkers for lung cancer detection in this thesis.

6.1 Utility of NECTIN-4 and PDAP1 as biomarkers in lung cancer

6.1.1 Utility of NECTIN-4 as serological biomarker for lung cancer detection

NECT-4 is a type-I transmembrane protein composed of extracellular, transmembrane and cytoplasmic domain (108). The largest proportion of NECT-4 structure belongs to extracellular domain (ectodomain), which can be shed by enzymatic activity of ADAM10/ADAM17 and released into the blood as a soluble form after the cleavage (108, 109, 111). Unlike other members of NECT-4 family, in human tissue *NECT-4* transcripts were initially found to be expressed only in placenta and slightly in trachea, whereas the expression in adult tissue lacks (106). Recently, contradictory findings of NECT-4 presence in normal tissue have been reported. On the one hand *NECT-4* expression has been reported in

other normal tissues, including lung tissue, even though at low levels, whereas the detection of NECT-4 on protein level has not been observed in normal lung tissue (145). On the other hand, the correlation between NECT-4 transcripts and protein levels have been reported (113). This discrepancy between reports could be due to the low level of *NECT-4* detection on transcript level for most of the tissues, and the lack of detection on protein level due to the lower sensitivity of most of the protein assays compared to RT-PCR. Therefore, attributes of having the soluble form and lacking the expression in adult tissue, or potentially low expression of *NECT-4* in normal tissue, arise the potential of NECT-4 as serological biomarker. Moreover, the activation of *NECT-4* expression has been reported in breast carcinoma, whereas the expression in normal breast tissue was not detected (146, 147). In addition to breast cancer, elevated NECT-4 in tumor tissue compared to normal tissues has been previously reported in other cancers as well (113, 114, 148-152). Further on, soluble NECT-4 has been reported to have diagnostic and prognostic value in several different cancers as well (146, 153).

In this thesis, NECT-4 has been examined as potential biomarker for lung cancer detection. Concentrations of NECT-4 were measured by commercial ELISA and the information on antibodies and their epitopes is not available to the public. Therefore, it was not possible to distinguish whether the cleaved extracellular domain of NECT-4 or full-length protein has been measured by ELISA in this thesis. *Western blot* analysis would be necessary to clarify which form of NECT-4 is being detected in plasma, but the concentration of NECT-4 in plasma seems to be too low for *Western blot* analysis (Section 5.5.1, Figure 23a).

Examination of NECT-4 concentration in plasma obtained in this thesis revealed elevated levels of NECT-4 in lung cancer patients compared to healthy controls, irrespective of gender stratification, which goes in line with previous results obtained on tissue samples (113, 114). Even though the average concentration of NECT-4 detected in this thesis in healthy controls was 14.1 pg/mL, the assay range of the ELISA used in this thesis is between 39.1-2,500 pg/mL (154). Therefore, one might speculate that the healthy controls in this thesis were negative for NECT-4 expression, which is concordant with previous reports on lack of expression of *NECT-4* in normal adult tissue (113).

Analysis of clinico-pathological parameters revealed association of NECT-4 concentration with histology, stage, and metastasis. Across different histological subtypes, the highest concentration of NECT-4 in this study was in adenocarcinoma group of patients followed by *other*, non-small cell carcinoma, and squamous cell carcinoma. In this thesis, 65.43% (53/81) of adenocarcinoma patients and 47.92% (23/48) of squamous cell carcinoma patients were positive for NECT-4 (Table S4). Results from this thesis are in line with previous results which reported 54% of adenocarcinoma serum samples, 64% of adenocarcinoma tissue specimen, 54% of squamous cell carcinoma serum samples, and 62% of squamous cell carcinoma tissue specimens positive for NECT-4 (113, 114). Contrary, 47.5% (19/40) of small cell lung cancer patients were positive for NECT-4 in this thesis (Table S4), whereas only 2% with low intensity of staining of small cell lung cancer tissue specimens were positive for NECT-4 in

other study (114). The mean concentration of NECT-4 for small cell lung cancer group in this thesis was 47.2 pg/mL, which is slightly above the lower limit of an assay detection range (39.1 pg/mL), however still detectable. Higher detection rate of NECT-4 in small cell lung cancer patients in this thesis compared to the literature mentioned above could be due to higher sensitivity of ELISA assay compared to immunohistochemistry. Moreover, it could also be due to the heterogeneity of tumor tissue which often is not properly represented when performing tumor sections. Further on, to the best of my knowledge, NECT-4 concentrations in plasma were not previously associated with histological subtypes (113, 155). Additional discrepancy between results obtained in this thesis and the results obtained by the other groups is the association of NECT-4 concentrations in plasma with T-stage (113, 155), whereas such an association was not found in this thesis. Concordance of the results obtained in this thesis with the results obtained by the other groups is the positive association of NECT-4 concentrations in plasma with lung cancer stages and metastases (155).

Clinical value of NECT-4 concentration and the assay used in this thesis was analysed by ROC analysis. Cut-off level was decided by accepting the specificity level of around 95%. For the detection of lung cancer patients, obtained AUC was 0.7627, with sensitivity of 43.52% at specificity of 95.00%. For the detection of lung cancer patients, Erturk et al. obtained AUC=0.816 with sensitivity of 70.00% at 85.00% of specificity (155). For better comparison with their results, results from this thesis for the detection of lung cancer patients were extracted at 85% specificity as well-the corresponding specificity is then 56.99% according to the ROC curve shown in Figure 31. To analyse the potential improvement of the test, lung cancer patients were stratified by histological subtype and analysed by ROC analysis. The best improvement was achieved for the detection of adenocarcinoma patients, obtaining AUC of 0.8267 with sensitivity of 56.79% at 95.00% specificity. Similar results have been reported previously, with the sensitivity of 53.70% at 97.70% specificity for detection of NSCLC patients (113). If the results for ROC analysis of adenocarcinoma patients are extracted at 85.00% of specificity, the corresponding sensitivity increases to 67.90%. Comparing to other previously suggested serological markers for the detection of lung cancer, the sensitivity of NECT-4 obtained in this thesis seems to be in line or higher compared to the other serological markers when similar specificity is obtained (Table 14). Advantage of the serological biomarkers is simplicity for implementation into the clinical setting since, e.g., ELISA does not require sophisticated instruments and it is feasible to perform. Contrary, the sequencing of cfDNA as a biomarker provides higher sensitivity and specificity comparing to serological biomarkers. However, the cost-effectiveness of sequencing comparing to protein detection might be a limitation for the implementation into the clinical setting.

Table 14. Performance of different analytes for the detection of lung cancer.

Analyte group	Technology	Target	Type	Sample	Sensitivity [%]	Specificity [%]	Ref.
Antigen	ELISA	NECT-4	LC	Plasma	43.52	95	This thesis
			ADC	Plasma	56.79	95	
	ELISA	PDAP1	LC	Plasma	23.32	95	This thesis
			ADC	Plasma	25.93	95	
			SCLC	Plasma	27.50	95	
	ELISA	NECT-4 + PDAP1	LC	Plasma	50.78	95	This thesis
			ADC	Plasma	60.50	95	
	ELISA	CYFRA21-1	LC	Serum	47	95	(156)
	IRMA	CYFRA21-1	NSCLC (I-IV)	Serum	59	94	(157)
	EIA	CEA	LC	Serum	27	95	(156)
	RIA	SCC	LC	Serum	15	95	(156)
	RIA	NSE	LC	Serum	16	95	(156)
cfDNA	RT-PCR	methylation status of MGMT, p16$^{INK4\alpha}$, RASSF1A, DAPK, RAR-β	LC	Serum	49.5	85.0	(158)
	Quantitative meta-analysis	cfDNA	LC	English articles	80	77	(159)
	RT-PCR	miR-19b-3p and miR-29b-3p	NSCLC (I-IV)	Blood-PBMCs	72.62	82.61	(160)
miRNA	RT-PCR	miR-378a, miR-379, miR-139-5p, miR-200-5p	LC	Plasma-exosomes	97.5	72	(161)

In conclusion, we can speculate that NECT-4 holds potential to be used for detection of lung cancer, in particular for adenocarcinoma. Sensitivity of an assay could be improved in the future with the change of antibodies used in ELISA with different kinetic properties, such as affinity of an antibody(ies) to an epitope of antigen. Additionally, an antibody which recognizes different epitope on NECT-4 than the one used in this thesis could be used. Since antibodies used in this thesis belong to the commercial kit, the kinetic properties and the exact epitope of an antibody were unknown in this study. Therefore, the potential improvement of the assay sensitivity would require multiple comparison of different antibodies and their specifications to potentially recognize the most suitable one to be used in clinical setting and could be considered as part of the future perspective. In addition to the diagnostic potential for adenocarcinoma, NECT-4 could be of an interest as a prognostic biomarker, given its association with the stage and the status of metastasis reported in this thesis, which could be part of future studies as well.

Furthermore, the variations of NECT-4 expression have been already described as tumor-origin dependent, as well as NECT-4 has been described as histology specific biomarker for ductal type of breast carcinoma (146). Additionally, strong correlation between NECT-4 and EGFR and P53 has been described (146). Mutations in *EGFR* and *P53* are drivers for NSCLC development, and together with *KRAS* are the most common mutated genes in adenocarcinoma (41), the histological subtype that showed the most promising results in this thesis for NECT-4 implementation into clinical setting. Another study

reported increased shedding of NECT-4 ectodomain after stimulus with lysophosphatidic acid (111). Interestingly, lysophosphatidic acid is associated with increased phosphorylation of protein kinase B (AKT), EGFR, extracellular signal-regulated kinase (ERK) 1/2, c-Jun N-terminal kinases 2/3 and c-Jun indicating associated signalling pathways of NECT-4 and EGFR. Moreover, NECT-4 is shed of by proteolytic cleavage of ADAM10 and ADAM17. The same study showed that the blocking of ADAM10 and ADAM17 with different blocking antibodies or siRNA, only partially inhibits NECT-4 shedding, or does not inhibit NECT-4 shedding when lysophosphatidic acid is applied as a shedding stimulus after the treatment with ADAM10/ADAM17 blocking agents (111). Such results could be due to the biochemical differences in the blocking agents, but the shedding of NECT-4 might also be initiated through other signalling pathway, potentially associated with AKT or EGFR signalling axis which are strongly implicated in development of lung cancer as mentioned above.

6.1.2 Utility of PDAP1 as serological biomarker for lung cancer detection

Platelet-derived growth factor A-associated protein 1 (PDAP1), also called heat- and acid-stable phosphoprotein of 28 kDa (HASPP28) is a secreted phosphoprotein of 28 kDa molecular weight discovered in 1996 as casein kinase II substrate (116) and as weak interactor of platelet-derived growth factor A (PDGFA) (115). Up to date, molecular function and signalling pathway of PDAP1 is poorly understood. Recently, PDAP1 has been described as essential regulator of mature B cells physiology, maintaining their homeostasis, being required for efficient class switch recombination, and protecting B cells from stress induced cell death (120). In addition, PDAP1 has been described as an RNA-binding protein prominently contributing to the T-cell proliferation (118). Moreover, PDAP1 phosphorylation by casein kinase II, protein kinase A and protein kinase C has been observed. However, in intact cells only phosphorylation by casein kinase II has been reported (116).

In this thesis, PDAP1 has been discovered as differentially expressed between hybrid epithelial/mesenchymal lung cancer cell line and mesenchymal lung cancer cell line, as being elevated in mesenchymal cell line compared to hybrid one (section 5.3.3. and section 5.4). Such expression is of a particular interest since it gives potential to detect also the signal deriving from CTCs undergone EMT and expressing rather mesenchymal phenotype compared to epithelial. To the best of my knowledge, up to date PDAP1 has been researched on patient samples in only two studies (122, 162), and in none of these studies PDAP1 has been investigated as secreted, blood-based biomarker. Therefore, PDAP1 has been investigated as serological biomarker for lung cancer detection for the first time in this thesis, and it has been found overexpressed in lung cancer patients compared to healthy controls (section 5.5) with concentrations being almost two-fold higher in lung cancer patients (1.87) than in healthy controls. Elevated levels of PDAP1 have also been reported in human rectal carcinoma (both in cell lines and human rectal tissue) (122), grade IV glioma (121), gastric cancer (SILAC based proteomic analysis)

(123), and colorectal cancer (patient tissue samples) (162). However, in contrast to NECT-4, PDAP1 has been detected in healthy controls as well.

Analysis of clinico-pathological parameters revealed association of PDAP1 with histology, stage, T-stage, and metastasis, whereas no association has been found between PDAP1 concentrations and smoking, sex, and nodal status. To the best of my knowledge, the analysis comparing association of PDAP1 with clinical parameters of patients was not yet performed by other studies. Therefore, it is not possible to compare results from this thesis with previous researches. In order to analyse the potential of PDAP1 in clinical settings, for lung cancer diagnosis respectively, ROC analysis has been performed. For discrimination of lung cancer patients from healthy controls, the obtained AUC is 0.7020 with sensitivity of 23.32% at specificity of 95.00%. In contrast to the results obtained on NECT-4, no improvement of ROC curve was obtained after stratification of lung cancer patients according to their histological subtypes, suggesting poor use of PDAP1 assay used in this thesis in clinical setting for lung cancer diagnosis.

In colon cancer, 146 carcinoma tissues with paired remote normal tissues (163) have been reanalysed using integrated omics study of metastatic colorectal carcinoma and the higher PDAP1 expression was correlated with worse survival of colorectal carcinoma patients (162). These results, together with results obtained in this thesis on regard to association of PDAP1 with stage, T-stage and metastasis, indicate that PDAP1 could be a promising prognostic biomarker rather than a diagnostic biomarker.

6.1.3 Utility of NECTIN-4 and PDAP1 as combined serological biomarker

Up to date, carcinoembryonic antigen (CEA) and serum cytokeratin 19 fragments (CYFRA 21-1) are the only available serological biomarkers for NSCLC (164). In the range of specificity between 68-95%, the sensitivity of CEA as serological biomarker for lung cancer is in the range between 29-69% (134). Sensitivities of CYFRA are mainly in the range between 19-68%, and one study reported the sensitivity of 85% in the range of specificity between 88-95% (134). In this thesis, sensitivities obtained at 95% specificity are 43.5% and 23.3% for NECT-4 and PDAP1, respectively. When NECT-4 and PDAP1 proteins are combined as biomarkers, the sensitivity improves to 50.78% at 95% specificity. In the case of NECT-4 only, the sensitivity was 56.79% for discrimination of adenocarcinoma patients from healthy controls. When NECT-4 and PDAP1 proteins are combined as biomarkers, the sensitivity increased to 60.5% at 95% specificity for discrimination of adenocarcinoma patients against healthy controls. With respect to sensitivities and specificities, results obtained in this thesis are similar or higher compared to the results obtained for other biomarkers in other studies, as mentioned above in this paragraph and in Table 14. Additionally, CCN1 has been previously examined as serological biomarker by our group and the positive association of CCN1 with lung cancer tumorigenesis has been found (75). The advantage of the detection approach used in this thesis for the implementation into clinical setting is the simplicity

of ELISA performance, considering its cost effectiveness and no need for sophisticated instruments. Moreover, combining NECT-4 and PDAP1 with previously suggested biological biomarkers for NSCLC (75, 134) in a form of biomarker panel could further improve the sensitivity of lung cancer detection and be more clinically applicable. Such a potential remains to be investigated.

6.2 Potential of NECT-4 and PDAP1 in lung cancer therapy

6.2.1 Potential of NECTIN-4 in lung cancer therapy

Aside of lung cancer detection, NECT-4 has potential to be applied in patient therapy, and it has been suggested as potential target for antibody drug conjugate therapy in malignancies expressing NECT-4 (114). The levels of NECT-4 were associated with levels of EGFR in breast cancer (146), suggesting that NECT-4 could be used for EGFR-associated therapy as well.

Additionally, measle viruses are described as having an oncolytic activity and to use three different receptors for the infection, one of which is NECT-4. These findings lead to the development of recombinant measle viruses which used only NECT-4 as a receptor for infection and induced cell death and tumor regression (165), suggesting the potential of NECT-4 in virotherapy. Such viruses could be additionally loaded with therapeutic agents and further benefit cancer patients, potentially improving life expectancy and life quality of the patient. Additionally, the field of engineered extracellular vesicles bearing therapeutic agent is increasing (166). Such engineered extracellular vesicles targeting NECT-4 could be produced and loaded with therapeutic agent, potentially supporting the development of novel targeted therapy for lung cancer. In the same manner, the research and development could go in direction of vaccine development, targeting NECT-4 positive cancer cells. Even though some studies reported low or moderate detection of NECT-4 in normal adult tissue, given the discrepancy of NECT-4 detection in adult normal tissue it is worth of considering NECT-4 for the development of cancer vaccine (for example for adenocarcinoma and squamous cell carcinoma). However, given the discrepancy of NECT-4 detection in adult tissues between different studies, healthy candidates for receiving a potential NECT-4 based vaccination would first need to be examined for the presence of NECT-4. Additionally, the presence of NECT-4 has been repetitively detected in placenta and the role of NECT-4 in embryogenesis has been reported as essential for normal development of the embryo (107). Therefore, women who plan pregnancy would not be able to receive potential vaccination as such. However, other healthy candidates, especially who belong to the high-risk group for development of lung cancer (age, smoking status) could potentially benefit.

6.2.2 Potential of PDAP1 in lung cancer therapy

Platelet-derived growth factor (PDGF) A and B are reported as PDAP1 interacting partners (115, 121). PDGF receptors (PDGFRs) have been associated with many different tumors and described as key regulators of mesenchymal cells. Additionally, PDGFR-signalling is involved with the growth of tumor cells, as reviewed elsewhere (167). Knockdown of *PDAP1* reduced the expression of *PDGFB* and its downstream genes (*AKT1* or Protein kinase B (*PKB*)), the phosphoinositide-dependent kinase-1 (*PDK1*), as well as it reduced the proliferation of malignant glioma cells, and increased the caspase activity (121). Furthermore, more than a twofold reduction of phosphorylated AKT in PDAP1 knockdown cells has been observed (121). Phosphorylation of AKT plays important role in tumorigenesis, regulating biological processes such as proliferation, apoptosis, and growth (168, 169). Since PDAP1 is associated with regulation of PDGF/AKT/PDK1 signalling axis, and the proliferation of cells, it is worth to consider PDAP1 as therapeutic agent. Reduction in PDAP1 concentrations in high-PDAP1-concentration patients via e.g., blocking antibodies, might significantly improve patient life quality.

Besides utilization of plasma proteins, EV-derived proteins can also be utilized as serological biomarkers for an early detection of lung cancer. To analyse the potential of EV-derived proteins for implementation into clinical setting, the EV isolation method suitable for downstream MS analysis has been developed in this thesis and EV proteome has been analysed by mass spectrometry as discussed below in more details.

6.4 Purification and proteome analysis of extracellular vesicles (EVs)

The big challenge in EV isolation is due to the heterogeneity of EVs resulting in partial overlap of characteristics between different populations of EVs. Moreover, there is considerable overlap with other biomolecules, e.g., apolipoproteins, which contributes to the complexity of EV purification. urvey from 2016 comprising 30 countries from four continents, reported that ultracentrifugation, including differential centrifugation, was by far the most used primary isolation method (81% of participants), where 59% of participants used ultracentrifugation in a combination with additional method (170). Participants analysing complex biological fluids, such as plasma, serum, urine, cerebral spinal fluid, or milk, used even up to 6 methods for isolation and purification of EVs, indicating the complexity of biological fluids on regard of EV purification (170).

6.4.1 Isolation of extracellular vesicles

Isolation of EVs from plasma is in particular challenging due to the presence of lipoproteins, which significantly outnumber the EVs in blood. Separation of EVs from lipoproteins is challenging mainly because they are similar either based on their density, e.g., EVs and high-density lipoproteins (HDLs), or based on their size, e.g., EVs and (very) low-density lipoproteins ((V)LDLs) (171). The sizes and densities of different lipoparticles are overviewed in Table 15 taken from Brennan *et* al. and modified and it will be helpful for following the further discussion presented later in this chapter.

Table 15. Overview of the sizes and density ranges of different particles (172).

Particle subtype	Size (nm)	Density (g/mL)
Chylomicrons	75-1,200	<0.95
Chylomicron remnants	30-80	0-95-1.006
Very low-density lipoproteins (VLDL)	30-80	0.95-1.006
Intermediate-density lipoproteins (IDL)	23-27	1.006-1.019
Low-density lipoproteins (LDL)	18-23	1.019-1.063
High-density lipoproteins (HDL)	7-13	1.063-1.21
Extracellular vesicles (EVs)	30-1,000	1.06-1.21

Higher recovery of particles smaller than 100 nm in diameter isolated by SEC+CF compared to SEC+UC has been reported previously (173), which goes in line with the results obtained in this thesis using NTA (Table 7, section 5.7.2). Detection of particles smaller than 100 nm corresponds to the size of small EVs, exosomes in particular. However, if the EVs are detected and quantified by NTA, it is impossible to distinguish where the signal is coming from. Whether the analysed matter are truly EVs or possible contaminants or even dust cannot be distinguished using particle tracking assays. Initially in this thesis, SEC+CF also presented the highest recovery rate compared to other tested methods, and the mode size of EVs of 86.9 nm and 90.2 nm, which corresponds to small EVs in a size range of exosomes. However, the high recovery of particles and the low mode size of particles obtained by SEC+CF is highly affected by lipoprotein contamination, rather than EV population as it was presented in this thesis using *Western blot* analysis (section 5.7.3, Figure 42). Signal deriving from APOA is relatively weak in a sample isolated by SEC+CF, whereas the signal deriving from APOB is strong. Such a distribution of signals goes in line with successful separation of EVs and HDL by size since SEC separates molecules based on their relative size and the pore size of the packing (174) (Figure 60), and EVs and HDL which consist of APOA (175) differ in their size (Table 15).
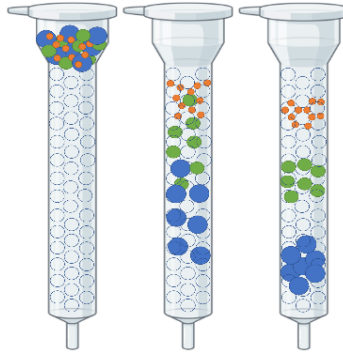
Figure 60. Size exclusion chromatography. Macromolecules are being separated based on the relative diameter of molecules and the pore size of packing. Bigger molecules do not fit the size of pores. Therefore, they are not getting slower with entering pores of the packing and they are being eluted first (blue shapes). The smallest molecules will travel the slowest since they enter inside of the pores which slows down their mobility. Therefore, smallest molecules will elute the last one (orange shapes), and intermediate molecules (green shapes) will elute in between the biggest and the smallest molecules. Created with BioRender.com.

Contrary, as mentioned above, EVs and (V)LDL are similar in their size but differ in their density which cannot be separated using SEC. Amicon filters with a cut-off of 100,000 Da were used in this thesis for centrifugal filtration, whereas the mass of only APOB itself is 515,605 Da (176). Therefore, the APOB and LDL in general would remain together with EVs and other smaller proteins would pass through the filter. This goes in line with the results in this thesis showing poor APOA contamination and high degree of APOB contamination of EVs isolated by SEC+CF. Additionally, the retention of APOB containing lipoparticles supports the high recovery of particles isolated using SEC+CF reported previously, since the same Amicon filters were used in the study (173).

Further on, (ultra)centrifuge is a method which uses centrifugal force to sediment components based on their differences in molecular weight, and therefore their density as well ($\rho\frac{m}{V}$), under high-speed and gravitation field (177). High contamination of EVs isolated by UC only was detected by *Western blot* analysis in this thesis. Separation of EVs from HDL (consisting of APOA) is limited given that the density range of EVs highly overlaps with the density of HDL (Table 15). Therefore, the strong signal deriving from APOA in a sample of EVs isolated using UC only is expected (Figure 42, section 5.7.3). Contrary, the separation of EVs from lipoparticles containing APOB when using UC only is expected due to the differences in their densities (Table 15). However, the intensity of APOB signal obtained by *Western blot* analysis in this thesis shows only a partial removal of (V)LDL when using UC only as an isolation method (Figure 42, section 5.7.3).

The third method analysed for the yield and the purity of EVs was SEC+UC. This method showed the highest purity of EVs compared to UC only and SEC+CF (Figure 42, section 5.7.3). APOA is expected to give a poor signal since SEC separates molecules based on their relative size and the pore size of the packing (174) (Figure 60), as described above. Additionally, APOB signal is expected to give a poor

signal since UC separates molecules based on their density, and EVs and (V)LDL, which consist of APOB, differ in their densities (Table 15).

Given that, the APOB signal in the sample of EVs isolated using UC only compared to the APOB signal in the sample isolated using SEC+UC should theoretically be the same, or at least similar. However, on the Figure 42 (section 5.7.3) it is notable that the ABOB signal between these two EV samples differs. It is known that APOB has glycosylation sites on N-terminus at the positions 34 and 185 (176). Glycosylation of APOB might result in an increase of the mass of (V)LDL which subsequently might lead to an increase of the lipoprotein density resulting in a co-sedimentation or partial co-sedimentation of these lipoproteins and EVs by UC. Moreover, similar to EVs, lipoproteins are heterogeneous in their composition, and lipoproteins consisted of glycosylated APOB might have different lipid composition compared to the lipoproteins consisted of non-glycosylated APOB. The composition of lipoproteins could affect the final size and the volume of lipoproteins, in both directions-lipoproteins being smaller and bigger. It is obvious that the change in the size would impact the isolation of EVs by SEC as well, potentially eliminating some of the (V)LDL already at that stage of the isolation procedure followed by UC. These events could explain the difference in the APOB pattern obtained by *Western blot* in this thesis. Additionally, APOB is a completed protein produced in the liver (APOB-100), whereas there is another splice variant, APOB-48 which is produced in the intestine. APOB-48 is the truncated version of APOB-100 caused by RNA editing producing an early stop codon and consists of N-terminal amino acids 1-2,152 of APOB-100 (178-182). Both, APOB-100 and APOB-48, are normally present in healthy individuals (183). Given that the antibody used in this thesis was anti-APOB (clone C1.4), which recognizes the N-terminal sequence between amino acids 97-526 of APOB-100, the antibody should recognize the truncated variant APOB-48 as well.

Nevertheless, the method SEC+UC showed superiority compared to UC only and SEC+CF in eliminating contamination while isolating EVs. Additionally, SEC+UC provided the strongest CD9 signal by *Western blot* analysis presenting the highest enrichment of EVs compared to other two isolation methods. Based on the results obtained by NTA, one would assume that the highest enrichment of EVs is obtained when using SEC+CF based on the EV recovery compared to two other methods (Table 7, section 5.7.2). However, poor intensity of the CD9 signal obtained by *Western blot* supports that NTA can detect other matters as nanoparticles and not only EVs (Figure 42, section 5.7.3), as stated previously in this chapter.

Prior to switching directly to the study cohort, SEC+UC isolation method for EVs was tested on plasma samples derived from lung cancer patient as well. Difference between processing plasma derived from healthy individual and cancer patients has been observed already at the pre-clearing steps of plasma prior to SEC. Usually, blood derived from healthy donor was centrifuged at 300 ×g in order to pellet the cells. Cell debris was removed by additional centrifugation at 1,800 ×g and the separated plasma would be stored at -80 degrees. After thawing the plasma derived from healthy individual no changes were

observed. Therefore, the following procedure would be centrifugation at 10,000 ×g followed by filtration, size exclusion chromatography and ultracentrifugation, as described in detail in M&M section 8.4.1. Blood derived from lung cancer patient was processed in the same manner as blood derived from healthy donor, and stored in the biobank after centrifugation step at 1,800 ×g. However, upon thawing plasma derived from cancer patient, high degree of blurriness and viscosity has been observed. Therefore, the centrifugation step at 1,800 ×g has been introduced after thawing the samples as well. Repetition of centrifugation at 1,800 ×g after sample thawing provided clearer sample and the issue of the sample handling due to the viscosity of the sample was removed. Further procedure remained unchanged.

All in all, SEC+UC showed superiority in both, contamination elimination and enrichment of EVs compared to UC, which is gold standard, and compared to SEC+CF. Therefore, size exclusion chromatography coupled with ultracentrifugation has been chosen for processing study cohort. The study cohort consisting of healthy individuals and cancer patient derived plasma was processed the same as described in M&M section 8.4.3 and analysed by LC-MS/MS.

6.4.2 Protein detection rate of EV proteome by mass spectrometry

Proteins have been reported as critical bioactive molecules of EV cargo directly involved in signalling cascade regulation, e.g., initiation of pre-metastatic niche formation in the liver (184), intercellular transfer of mutant KRAS molecules in colon cancer cells and involvement in cancer progression (185), or association with lung and liver metastasis (186, 187). Therefore, deeper insight in EV-derived proteome is needed and holds big potential to be utilized. EV proteome could be utilized for better understanding of biology behind EVs as well as for direct clinical utility of EVs in diagnostics, prognostics, therapy monitoring, or even therapy delivering. In order to analyse EV proteome derived from human plasma, and to have better insight into differences between healthy individuals and lung cancer patients, extracellular vesicles were isolated in this thesis using SEC+UC (section 8.4.3) and further analyzed by LFQ-LC-MS/MS. In total, 5,435 proteins were detected, whereas 4,657/5,435 proteins passed the validity test and were considered for further analysis (section 5.9).

Elution of EVs by SEC occurs before the elution of secreted proteins, and they are often non-detectable by common protein concentration assays, or the concentration seems very low (188), which was observed in results obtained in this thesis as well (section 5.8, Figure 43). However, the protein amount deriving from extracellular vesicles is by far enough for common proteins assays, e.g., *Western blot*, as well as mass spectrometry analysis (sections 5.7.3 and 5.9.1, Figures 42 and 49). In this thesis, the classical EV biomarkers CD9 and CD81 were detected by mass spectrometry analysis in all plasma samples analysed, whereas CD63 was detected in 31/36 samples in total. The lack of CD63 detection in

some of the samples is argued due to the extended elution profile of CD63⁺ EVs resulting in an overlap of EVs elution with plasma derived proteins (Figures 37 and 38, section 5.7.1). Since fractions in which raising protein concentration has been observed were not pooled in order to avoid the interference of plasma derived proteins with EV-proteomic analysis, potential limited detection of CD63 signal analysed by mass spectrometry was expected in this thesis, meaning also that CD63⁺ EV subpopulation might be underrepresented in such an approach. However, it still allows proteomic analysis in order to develop potential biomarkers for implementation into clinical setting.

Protein detection rate of EVs isolated using density centrifugation was in total 640 proteins (189), whereas in this thesis the identification rate was 4,657 proteins detected, suggesting better suitability of the protocol obtained in this thesis for downstream LC-MS/MS analysis. Comparison with the other studies, e.g., presented in Table 16, suggests the same, providing further confirmation of the suitability of EV isolation protocol for downstream LC-MS/MS analysis provided in this thesis.

Table 16. EV detection rate obtained by mass spectrometry analysis

Sample source	Starting sample volume (mL)	Vesicles analysed	Isolation method	Prot. Detection method	No. of identified proteins (total)	CD9, CD81, or CD63	Ref.
Plasma	1.0	Exosomes	SEC	LC-MS/MS	269	No	(190)
Plasma	-	Exosomes, MVs	Multiple MS studies of the EVs: ExoSpin™, SEC, Ultracentrifugation, density gradient (OptiPrep™), Immunoaffinity pulldown	LC-MS/MS	361	No	(190)
Plasma	0.5	Exosomes	Exo-Spin™	LC-MS/MS	152	No	(190)
CCS	35-53 (T175 flask)	Exosomes	SEC	LC-MS/MS	<200	Yes	(188)
CCS	Supernatant was collected from 2 x 10⁷ U937 cells	Necroptotic EVs in a range of nanoparticles	Ultracentrifugation	LC-MS/MS	3,337	Yes	(191)
Plasma (LACI patients)	5	Plasma membrane bound vesicles	Sequential centrifugation	LC-MS/MS	458	Yes	(192)

CCS=cell culture supernatant

LACI=lacunar infarction

MV=microvesicles

6.4.3 Selective packaging of proteins into EVs

It has been reported that the protein content differs between small and large EVs (193), which could be the result of increased EV contamination during isolation of large EVs, but it could also be the result due to the specific loading of proteins into EVs. Specific loading of proteins into EVs is further supported by the difference between the proteome of small and large EVs isolated from high metastatic breast cancer cell line compared to the proteome of small and large EVs isolated from low metastatic

breast cancer cell line (193). Further on, the differentiation of adenocarcinoma patients from healthy controls based on their EV proteome has been previously reported (189), which goes in line with the results obtained in this thesis (Figure 51 and 53, sections 5.9.2 and 5.9.3). Additionally, in this thesis healthy control smokers partially grouped with the rest of healthy controls and partially with adenocarcinoma patients (Figure 51 and 53, sections 5.9.2 and 5.9.3). To investigate that further, healthy control smoker- and healthy control non-smoker-EV proteome was compared with the EV proteome deriving from adenocarcinoma patients (section 5.9.3). Interestingly, from a total of 970 differentially expressed proteins ($p<0.05$) between healthy controls and adenocarcinoma patients, 970 proteins were differentially expressed between healthy control *non-smokers* and adenocarcinoma patients, whereas only 554 were differentially expressed between healthy control *smokers* and adenocarcinoma patients. Additionally, 526 differentially expressed proteins have been detected between healthy control smoker- and healthy control non-smoker-EV subproteome. To the best of my knowledge, this is the first report on the reflection of the EV proteome associated with the smoking effect between healthy individuals, indicating that healthy control smokers which grouped together with adenocarcinoma patients could potentially be individuals at risk of developing lung cancer. Such evidence indicates that EV subproteome could be used in screening of the high-risk healthy individuals for early detection of lung cancer. Therefore, it is important to proceed on with the analysis of current proteomic EV data obtained in this thesis together with further experiments in order to define potential biomarkers for screening of high-risk healthy individuals.

6.5 Development of the EV-biomarker panel for adenocarcinoma lung cancer detection

Investigation of EV-derived proteome to identify potential EV biomarker(s) for the detection of lung cancer was one of the aims in this thesis. Therefore, the number of significant differentially expressed proteins between each groups compared (section 5.9.3) was reduced by adding FDR difference of 1.5 or more. Seven proteins which were detected in majority of participants analysed with the highest difference between healthy controls and adenocarcinoma patients, independently of the smoking status, were chosen as potential EV-derived protein biomarkers for the detection of adenocarcinoma lung cancer. Seven proteins suggested to be part of the biomarker panel for the detection of adenocarcinoma are SAA2, REV1, SCLY, SLC12A6, PRSS1, GPM6A, MSN.

6.5.1 Serum amyloid A2 (SAA2)

SAA2 protein belongs to acute phase proteins and has been reported to increase in response to inflammation (194) and to have a role in the suppression of immune response (195). Levels of serum

SAA2 have been correlated with EVs expressing proteins associated with myeloid cells, CD14, CD16, and C-X-C motif chemokine receptor 1 in HIV-infected individuals (196), supporting the role of EVs in immune activation. Clinically, the SAA2 has been associated with different tumors. Head and neck squamous cell carcinoma patients with high expression of SAA2 have increased risk of local tumor recurrence compared to the patients with low expression of SAA2 (197). Furthermore, SAA2 is one of the most upregulated proteins detected on EVs isolated from plasma of glioma patients compared to healthy controls as analysed by LC-MS/MS (198), which goes in line with the results obtained in this thesis. According to the ExoCarta database, SAA2 is present on EVs (199). However, it is important to point out that SAA2 protein is a serum apolipoprotein and it associates with HDLs (200), arising suspicious if the detection was obtained truly from EVs or if it rather derives from contamination. It would be interesting to perform additional experiment with the use of SAA2 specific antibody for immune-based EV isolation following size exclusion chromatography or ultracentrifugation as confirmation that SAA2 is EV-derived biomarker instead of plasma-derived. The presence of SAA2-immuno based isolated EVs could be confirmed by TEM. In the case that SAA2 is enriched in lung cancer derived EVs, SAA2 could be an interesting lung cancer biomarker protein for further investigation.

6.5.2 DNA repair protein REV1 (REV1)

REV1 is a DNA polymerase involved in DNA repair mechanism and it has a central role in translesion synthesis (TLS), which is an error prone synthesis having a tolerance to the DNA damage (201). Accelerated tumorigenesis has been associated with overexpression of *REV1* (202). Inhibition of REV1 reduces the frequency of mutations in DNA after exposure to the DNA-damaging agents, e.g., UV light or cisplatin (203, 204), including in lung cancer cell lines (205), and it could overcome tumor chemoresistance during chemotherapy treatment. Except on lung cancer cell lines, REV1 was found upregulated in lung cancer patients, and the inhibition of REV1 showed anticancer effect (206). REV1 has been reported to promote cell growth and proliferation of lung cancer cells (206), therefore, promoting the lung cancer tumorigenesis. Significant differences in DNA repair capacity between lung cancer patients and healthy individuals has been previously reported (207), indicating involvement of DNA repair genes/proteins in lung cancer tumorigenesis. Additionally, poor DNA repair capacity is independent of tobacco smoking status (207). To the best of my knowledge, REV1 has not been detected in EVs yet. Therefore, REV1 is reported in this thesis as a potential novel EV protein, potentially transferred between cells in order to accelerate mutagenesis of adjacent cells, possibly in order to induce independence and resistance of the cell to gain higher metastatic potential. Additionally, REV1 is suggested in this thesis as potential EV-derived lung cancer biomarker, capable of distinguishing individuals with adenocarcinoma lung cancer from healthy individuals, independently of smoking status.

6.5.3 Selenocysteine lyase (SCLY)

SCLY is an enzyme which catalyses decomposition of L-selenocysteine down to L-alanine and H_2Se (208), and it is involved in selenium delivery (209). Selenocysteine is a cysteine analog which possess selenium atom instead of sulphur atom, whereas selenoproteins are proteins containing selenocysteine. Interestingly, SCLY has been detected on mRNA and protein level in various tissue types, including lung tissue, but the lack of detection was observed in blood (210). Detection of SCLY on EVs isolated from blood plasma in this thesis was probably facilitated due to the enrichment of EVs and elimination of blood derived secreted proteins, which would usually be interfered by more abundant blood proteins. Further on, SCLY is regulated by Jun Proto-Oncogene/Activating protein 1 axis (211). Activating protein 1 binding site on *SCLY* mRNA binds transcription factors Jun and c-Fos which have been associated with oxidative stress and inflammation. Furthermore, oxidative stress has been associated with the increase of pulmonary inflammation mediators, DNA oxidative damage, and initiation of carcinogenesis (212). Additionally, SCLY is upregulated in glomerulonephritis (211), and hypoxia induced downregulation of selenoproteins expression on posttranscriptional level has been reported (213). Further on, hypoxia has an effect on carcinogenesis and tumorigenesis, including in lung cancer as well (214). In conclusion, given the EV-derived proteomic results from this thesis (section 5.9) and literature mining, SCLY is reported as EV protein for the first time in this thesis, and suggested as potential EV-derived biomarker for discrimination between adenocarcinoma lung cancer patients and healthy individuals independently of the smoking status.

6.5.4 Solute carrier family 12 member 6 (SLC12A6)

SLC12A6 or as in literature more commonly referred as potassium-chloride cotransporter (KCC) 3 belongs to the KCC family of proteins and plays an important role in electroneutral K^+ and Cl^- cotransport through the plasma membrane (215). Therefore, it has a significant role in maintaining the homeostasis of the cells. KCC proteins have been previously associated with carcinogenesis. Upregulation of *KCC3* mRNA transcripts during cervical and ovarian carcinogenesis has been observed, as well as for *KCC1* and *KCC4* mRNA transcripts (216, 217). Further on, the activity of KCC3 has been associated with the changes in retinoblastoma protein and cell division control 2 kinase activities, which are the key regulators of the cell cycle (218). KCC3 is related with tumor proliferation and invasiveness through the insulin-like growth factor-1 and tumor necrotic factor-α regulation (218, 219), as well as activation of the nuclear factor kappa B and Ste20-related proline/alanine-rich kinase signalling cascades (220). Moreover, the overexpression of KCC3 promotes epithelial-to-mesenchymal transition through downregulation of E-cadherin/β-catenin complex formation (218, 221). KCC3 has been associated with the migration of the cancer cells as well, e.g., in high-grade glioma cells (222), breast cancer cells (223),

or in esophageal carcinoma cells (224). Clinically, KCC3 is unfavourable prognostic factor in esophageal cell carcinoma (224). According to the ExoCarta database, KCC3 has been previously detected on EVs isolated from ovarian cancer cells and from platelets (225). To the best of my knowledge, KCC3 has not been reported in lung cancer yet. Therefore, KCC3 is for the first time suggested as EV derived protein associated with lung cancer tumorigenesis in this thesis and suggested as a potential smoking independent biomarker for discrimination of adenocarcinoma lung cancer patients from healthy individuals.

6.5.5 Serine protease 1 (PRSS1)

PRSS1 is a main trypsinogen secreted by the pancreas (226). Levels of PRSS1 and *PRSS1* mutations are associated with multiple diseases, e.g., autoimmune pancreatitis (227), hereditary and chronic pancreatitis (228), as well as with cancers such as high grade serous ovarian cancer (229, 230), thyroid cancer (231), gastric cancer (232), and bladder cancer (233). Higher levels of *PRSS1* mRNA were detected in urine EVs derived from bladder cancer patients compared to urine EVs from healthy individuals (233). Further on, differentially expressed proteins were identified in gastric mucosal epithelial cancerous tissue using isobaric tags for relative and absolute quantitation labelling and 2D LC-MS/MS. Among other proteins, PRSS1 was verified as significantly overexpressed and positively correlated with differentiation, tumor size and lymph node metastasis of gastric cancer, as well as with the poor overall survival of patients with high levels of PRSS1 (232). Interestingly, in this thesis, PRSS1 has been detected at lower levels in lung cancer-derived EVs compared to healthy control-derived EVs which is in line with qPCR results reported previously (234). PRSS1 has been reported as downregulated in *early relapse* patient group compared to *no relapse* patient group of stage I radically resected NSCLC adenocarcinoma patients (234), supporting PRSS1 as potential adenocarcinoma biomarker.

Concrete mechanism of PRSS1 regulation in lung cancer is not very well known. It could be that the expression of *PRSS1* is possibly downregulated in lung cancer patients by miR-146a mechanism. miR-146a has been reported to target lung cancer relevant receptor, e.g., EGFR and Neurogenic locus notch homolog protein ½ (235-237). Moreover, knockdown of *PRSS1* expression inhibited the activation of ERK signalling pathway in gastric carcinoma cells (232), whereas ERK is known to be signalling pathway affected in lung cancer tumorigenesis as well (238). miRNA-146a-5p has a binding site for PRSS1 (232) and could be involved in silencing of *PRSS1* expression in lung cancer. According to the ExoCarta database, PRSS1 is reported for the first time on a protein level in this thesis as detected in human derived EVs, whereas PRSS1 has been previously reported on mice and rats derived EV on the protein level (239). Therefore, PRSS1 is suggested as novel EV-derived protein and as a potential EV-derived biomarker for discrimination of adenocarcinoma lung cancer patients and healthy individuals.

6.5.6 Neuronal membrane glycoprotein M6-a (GPM6A)

GPM6A is a transmembrane glycoprotein involved in differentiation and migration of neuronal and neuronal stem cells (240). Downregulation of *GPM6A* has been reported in lung cancer tissue compared to normal tissue (241), which is in line with the results obtained in this thesis on protein (GPM6A) level in EVs. Additionally, there are indications that overexpression of *GPM6A* could be the apoptosis-inducing factor (242), indicating the tumor-suppressor role of *GPM6A*. Moreover, *GPM6A* mRNA is part of tumor suppressive lncRNA-miRNA-mRNA axis in lung cancer (243). Additionally, GPM6A interacts with stemness-related transcription factors (243), implicating that dysregulation of GPM6A could be associated with epithelial-to-mesenchymal transition. Clinically, dysregulation of GPM6A has been reported in different cancers, e.g., in hepatocellular carcinoma (244), glioblastoma (245), colorectal carcinoma (246). Interestingly, low levels of GPM6A are associated with longer overall survival in colorectal cancer (246), indicating potential oncogenic function of *GPM6A* in colorectal cancer. In contrary, high expression levels of *GPM6A* mRNA are associated with better survival of lung cancer patients (243), which goes in line with potential tumor suppressing role of *GPM6A* in lung cancer mentioned above. Additionally, ROC analysis of *GPM6A* mRNA showed high diagnostic value of *GPM6A* in lung cancer by obtaining AUC of 0.993 with sensitivity of 95.88% at the specificity of 100% (243). In this thesis, GPM6A on EVs deriving from lung cancer was detected on lower levels comparing to GPM6A on EVs from healthy controls. Since patient samples for EV analysis obtained in this thesis derived from lung cancer patients with advanced lung cancer and present metastasis, given results might be due to the loss of function of *GPM6A*, further supporting the tumor suppressing role of *GPM6A* in lung cancer development. According to the ExoCarta database, GPM6A has been previously detected on EVs on protein level and together with the results in this thesis, GPM6A is suggested to be EV-derived biomarker for discrimination of adenocarcinoma lung cancer patients and healthy individuals independently of the smoking status.

6.5.7 Moesin (MSN)

Moesin was discovered in 1988 as a heparin-binding protein (247) and nowadays is known as a member of ezrin-radixin-moesin (ERM) family of proteins (248). ERM family of proteins are highly homologous and (248) essential for connecting actin cytoskeleton and the cellular membrane, as well as for cell-cell and cell-substrate adhesion or construction of microvilli (249). However, it seems that moesin individually, without ezrin and radixin members of ERM family, does not affect cell-cell nor cell-substrate adhesion but it affects the microvilli structures (249). ERM proteins have been reported to bind different transmembrane receptors involved in tumorigenesis, such as EGFR, hepatocyte growth receptor (HGFR), PDGF receptor, or co-receptors, such as CD44, CD43, ICAM1 and ICAM2 (248). Moreover, ERM proteins have been associated with MAPK signalling pathway and suggested as a scaffold proteins for the modulation of Ras superfamily activity (248).

Clinically, high moesin levels are associated with oral squamous cell carcinoma (250), high grade glioblastoma (251), breast cancer and the poor prognosis of breast cancer patients (252). Moreover, moesin is proposed as basal breast cancer biomarker (252). Downregulation of moesin in lung cancer adenocarcinoma tissue compared to normal or para-cancerous tissue on both, mRNA and protein level is reported (253, 254), which goes in line with the results obtained on EVs in this thesis. Lung adenocarcinoma patients with lower moesin levels show poor overall survival (254). Furthermore, downregulation of moesin compared to normal lung tissue has been observed in the early-stage adenocarcinoma as well (253), suggesting its potential as an early lung cancer detection biomarker. Moreover, moesin has been suggested to rather have tumor supressing role in lung cancer tumorigenesis (253). Interestingly, the correlation between *SNAIL* and *MSN* mRNA expression has been found in early-stage lung cancer patients comprising adenocarcinoma and squamous carcinoma (255), indicating the role of moesin in EMT through Snail signalling axis. It could be that moesin is associated with both, tumor suppressing and oncogenic events in lung cancer tumorigenesis due to the involvement of ERM proteins in many biological pathways and having many interacting partners, as mentioned above. In this thesis, different levels of moesin between adenocarcinoma lung cancer patients and healthy individuals have been detected on EVs. Given the association of moesin activity with tumorigenesis related processes, moesin is suggested as a potential EV-derived lung cancer biomarker independent of the smoking status.

Seven above-described proteins, SAA2, SCLY, REV1, PRSS1, SLC12A6 (KCC3), GPM6A and MSN were detected as EV-derived differential proteins between adenocarcinoma lung cancer patients and healthy individuals. SAA2, SCLY, REV1, and SCL12A6 were upregulated and PRSS1, GPM6A and MSN were downregulated on adenocarcinoma lung cancer-derived EVs compared to healthy control-derived EVs. These proteins were among the ones with the highest fold change between lung cancer derived EVs and healthy control derived EVs and differential between two compared groups independent of the smoking status. Since smoking status is one of the highest risk factors for lung cancer development (38), and the results in this thesis have shown differential EV proteome between smokers and non-smokers (section 5.9.3), smoking independent candidates were considered as the best fit for potential lung cancer (adenocarcinoma) EV-derived biomarkers. However, further evaluation and validation of the suggested EV-biomarker panel is necessary. Perhaps, immunoaffinity EV isolation could be performed targeting EV-biomarker panel proteins and subsequent TEM analysis for the confirmation of EV isolation. On the other hand, EVs could be isolated using immunoaffinity-based technique targeting EV biomarkers such as CD9 or CD63 and downstream targeted proteomics could be performed for the confirmation of differential expression of suggested EV-biomarker-panel proteins. Later on, validation of suggested biomarker panel, or each protein individually as an EV-derived biomarker for lung cancer (adenocarcinoma) detection should be validated on larger study cohort.

6.6 Potential role of extracellular vesicles in lung cancer biology

Extracellular vesicles have received high attention throughout the last two decades and the interest of the research community on the EVs keeps increasing (256). Despite intensive research so far, the concrete mechanism and biological role of EVs and the effect on tumorigenesis remains largely unknown. Even though the focus of this thesis was primarily on the identification of novel protein biomarkers for the early lung cancer detection, the chance to gain brief insight into the biology behind EVs was used. Performed pathway analysis of EV-derived proteins showed clear differences in biological function(s) between adenocarcinoma lung cancer patient-derived EV proteins and healthy control-derived EV proteins (Figures 58 and 59).

Based on the gene ontology pathway analysis, the healthy-control-derived EVs are mostly associated with vesicular components (Figure 58), whereas the highest number of lung cancer patient-derived EV proteins are associated with cytoplasm or extracellular space as cellular component. Association of lung cancer patient-derived EVs with cytoplasmic and extracellular components might be related to the biogenesis of extracellular vesicles. Whether the EVs are produced by upward membrane budding or through endosomal pathway, cytosolic and membrane proteins are assumed to be engulfed. Contrary, there are suggestions that EV packaging might be specific (257). It could be speculated that the interior of EVs is likely as the cytoplasmic part of the cell, resulting in the highest stability of cytoplasm-related proteins and the stability of the packed EV proteins during the EV transport. Additionally, cytoplasm- and ECM-related proteins might be specifically packed for cell-to-cell communication and/or preparation of metastatic niche. For example, direct modulations of cells via EVs at the predisposed secondary site (258, 259), as well as association of EV-mannered delivery of specific molecules to extracellular space (ECM), e.g., chemokines (260) have been reported. Even though by far less proteins have been associated with it (3-fold lower), the other cellular component the most associated with lung cancer derived EVs is related to protein-lipid complexes and lipoprotein particles (Figure 59, section 5.9.5).

On the one hand, it is hard to eliminate the possibility that the pathway analysis of enriched lung cancer-derived EV proteins could have been interfered by circulating free lipids. On the other hand, LDLs are involved in the catalysis of free fatty acids and cholesterol in lysosome, whereas lysosome biogenesis is associated with the biogenesis of exosomes. Furthermore, packaging of lipid nanoparticles linked to mRNAs into endocytic extracellular vesicles has been reported (261). Moreover, lipid dysregulation within EVs has been reported in multiple diseases, such as Alzheimer's disease (tissue derived EVs) (262) or metabolic diseases (263). Lipidomic profiles of EVs deriving from non-tumorigenic, tumorigenic, and metastatic prostate cell line supernatants showed differential abundance of lipids between samples. Fatty acids, glycerolipids and prenol lipids were abundant in EVs deriving from non-tumorigenic cells, while sterol lipids, sphingolipids and glycerophospholipids were abundant in EVs deriving from tumorigenic and metastatic cells (264). In general, lipid/lipoprotein contamination while

isolating EVs tend to be more associated with isolation of EVs from body fluids. Therefore, one can assume that EVs isolated from cell culture supernatant represent more "*trusty*" result of differential lipid abundances compared to the one from plasma samples, such as is the case in this thesis. However, it could be said that the results obtained on cell line supernatants (264) are in line with the results obtained in this thesis. Even though, fatty acids and glycerolipids were associated with non-tumorigenic EVs, it is not surprising that they were enriched in this thesis in EVs deriving from lung cancer patients, since all EVs were analysed from each participant, including tumorigenic and non-tumorigenic EVs. Perhaps pathways associating fatty acids and triglycerides represent more the systemic response of organism to the cancer diseases, rather than potential tumor promoting pathways induced by the tumor (or tumor cells) itself. Contrary, *phosphatidylcholine-sterol O-acyltransferase activator activity* is enriched in lung cancer patient-derived EVs based on the results obtained in this thesis. Taken it together with the detection of sterol lipids as abundant in tumorigenic and metastatic cell line supernatant-derived EVs (264), it could be speculated that these are involved in the cancer progression and in the pathways associated directly with tumor burden, since 15/16 lung cancer patient samples in this thesis were obtained from advanced/metastatic lung cancer patients (for patients' characteristics refer to the Table S5). Lastly, EVs have been reported to carry lipids and enzymes involved in their biogenesis and having chemotactic effect (265). Taken the above mentioned together with the additional steps taken in this thesis to further examine whether the obtained EV results are significantly interfered by plasma contamination (see below), it is believed that differential abundance detected on EVs deriving from lung cancer patients comparing to EVs deriving from healthy individuals in this thesis comes from EVs rather than from the interference by contamination.

Additional steps that were taken for the examination of potential plasma contamination interference with EV-MS results and observation which led to its exclusion were: (i) apoliprotein abundancies had small differences and they showed a rather uniformed pattern across healthy control and lung cancer EV samples, (ii) EV markers CD9, CD81, and CD63 abundancies had small differences and showed a rather uniformed pattern across healthy controls and lung cancer EV samples, (iii) clathrin abundancies, in particular CLH1, showed an uniformed pattern across healthy control and lung cancer EV samples, supporting that small differences observed in (i) and (ii) can be considered as rather uniformed pattern (Figures 48-50, section 5.9.1), and (iv) housekeeping proteins abundancies, e.g., RAB5C, COPZ1, had an uniformed profile plot across healthy controls and lung cancer EV samples (Figure S3).

Additionally, the differences in lipidomic profiles have been reported in NSCLC compared to healthy controls, as well as enzymes involved in lipid metabolism (266). Increase in lipid-associated content produce reactive oxygen species which in return can activate lipotoxicity-protective-proteins such as acyl-coenzyme A (CoA):cholesterol acyltransferase (ACAT) enzymes (266). ACAT proteins are acyltransferases, which is one of the detected enriched molecular functions of lung cancer-derived EV proteins (Figure 59, section 5.9.5). Interestingly, the activity of ACATs has been associated with

stimulation of cancer cell proliferation as well as tumor growth through different pathways associated with the increase of reactive oxygen species (266). These previous reports further support the results obtained by pathway analysis in this thesis associating lung cancer patient-derived EV proteins with lipid metabolism.

Functionally, healthy donor-derived-EVs are mostly associated with enzymatic activity (Figure 58), as well as lung cancer patient derived EVs (Figure 59, section 5.9.5). High degree of lung cancer patient-derived-EV molecular function involves regulation and inhibition of peptidase and endopeptidase activity, indicating active metabolism and turnover of proteins within the EVs. Perhaps it could indicate that EVs have by far bigger and rather active role than what mostly EVs are currently associated with, e.g., in "*information transfer*" during cell-cell communication. On the one hand, endopeptidase activity inside of EVs could be associated with the hydrolyses of proteins and generating free amino acids which are potentially delivered to the recipient cell. Hypothetically speaking, these free amino acids could save the energy of cancer cells for the synthesis of novel proteins. Such a mechanism would allow the cancer cell to utilize "organism energy" for the self-development and progression, especially if the localisation of tumor cell is far from the blood stream and the nutrient supply is limited. On the other hand, unmatured proteins might be packed into EVs and their propeptide processed into mature form for a delivery to the recipient cell. Such a function might be relevant for example in immunomodulating, initiating relevant signalling cascade which would contribute to the tumor-related-immunosuppression. Additionally, EVs function is associated with removal of unnecessary molecules from the cells related to cell maturation and adaptation to environmental changes and blood clotting activation (265). Therefore, high activity of peptidases in lung cancer patient-derived EVs might also be associated with cellular maturation and adaptation.

Biologically, healthy control derived EVs are mostly involved in processes associated with metabolism, localization, and biogenesis in this thesis (Figure 58). Contrary, the lung cancer patient-derived EV proteins were highly associated with the immune system as the most enriched biological processes in this thesis (Figure 59, section 5.9.5). Immune system is a very complex system comprised of different immune responses, molecular components and cells as it is in more details described elsewhere (267-272). During carcinogenesis and tumor progression the immune system is activated and protects the host organism against tumor development, whereas tumor goes through a process called *immunoediting* to eventually evade the immune system (273).

EVs have previously been associated with the immune system. For example, inflammatory cells associated with the acute immune response attract other cells by secreting EVs during their migration and creating stable chemotacting gradient in the ECM (274). Furthermore, complement factors and complement regulatory proteins can bind to the surface of EVs (275). Moreover, EVs can protect their donor cell from complement mediated cellular death through a decoy function via antibody-binding epitopes on their surface which target B cells (276). Additionally, EVs secreted by variety of cells within

tumor microenvironment, including immune cells, are able to regulate tumor cell phenotype as well (277). These findings indicate high involvement of EVs in immune modulation as well as the association of EVs with different immune systems including innate, humoral, and acute immunity associated with inflammation. Cytotoxic ability of EVs has also been reported. Fas ligand is a transmembrane protein which belongs to the tumor necrosis factor family and perforin is a pore-forming cytolytic protein. Fas ligand and perforin have been detected in higher concentration in NK-derived EVs than in the NK cells themselves, as well as these EVs had cytotoxic effect on melanoma tumor cells (278). Furthermore, tumor cell-derived EVs reduced the percentage of proliferating T cells and showed immunosuppressive activity (279). Interestingly, upon the B and T cells activation, lipid metabolism reprogramming occurs in T lymphocytes and de novo lipid synthesis and cholesterol uptake is being upregulated. Such events are observed in fast proliferating cells, whether T lymphocytes due to their increased activity, or even in cancer cells during their progression (280). Moreover, the ATP-binding cassette transporter G1 (ABCG1) is known as cholesterol efflux transporter. ABCG1 downregulation reduced tumor growth in vitro and induced the accumulation of EVs and the reduction of cell survival (281). Reduction of HIF-1α in central necrotic area of tumors is associated with the ABCG1 downregulation as well, indicating anti-angiogenic effect of targeting ABCG1 transporters (281). These findings also suggest the importance of cholesterol in tumor progression. Furthermore, HIF-1α is a known hypoxic factor which induces glycolytic metabolism and it is upregulated upon T cell activation (280). Perhaps cholesterol, or other lipids, regulate HIF-1α which would have downstream effect on the T cells and the immune response in general, and the regulation of cholesterol and other lipids delivery could be at least partially regulated through EVs.

Taken together, it is evident that the host immune system plays an essential role in both tumor suppressive and tumor promoting manner. Additionally, the EVs secreted by both tumor cells and immune cells, and probably secreted by the other cell types as well especially the one associated with tumor microenvironment, play a pivotal role in cancer progression. It could be that the EVs are the central point of the tumor regulation associated with the immunity/HIF-1α/cholesterol axis and that targeting EV-associated cholesterol might have anti-tumor effect.

6.7 Conclusion

This thesis supports the potential utility of liquid-based biomarkers in lung cancer detection. Proteins NECT-4 and PDAP1, and the proteome of EVs have been analysed for their utility in discrimination of lung cancer patients from healthy individuals.

NECT-4 is elevated in lung cancer patients compared to healthy controls. Sensitivity and the specificity of NECT-4 as a single biomarker is the highest for discrimination of adenocarcinoma patients from healthy controls. PDAP1 is elevated in lung cancer patients compared to healthy controls. However, the PDAP1 assay used in this thesis showed poorer sensitivity for discrimination between lung cancer patients and healthy individuals than NECT-4, even after histological stratification. Therefore, the PDAP1 assay used in this thesis showed less potential to be implemented into clinical setting as lung cancer diagnostic biomarker compared to the NECT-4 assay.

Moreover, this thesis was also focused on the investigation of EV subproteome and the differences between lung cancer patient EV subproteome and the EV subproteome of healthy individuals. EV isolation protocol has been optimized coupling size exclusion chromatography and ultracentrifugation, providing high yield and purity of EVs suitable for downstream mass spectrometry analysis. Based on proteomic analysis, adenocarcinoma lung cancer patients can be distinguished from healthy individuals based on the differences between their EV protein composition. Furthermore, healthy individuals who are smokers and are under risk of developing lung cancer, could potentially be recognized based on their EV protein composition, suggesting the potential utility of the EV subproteome in lung cancer screening. Moreover, EV-derived proteins SAA2, REV1, SCLY, SLC12A6, PRSS1, GPM6A, and MSN are suggested as members of a potential EV-putative biomarker panel for the detection of adenocarcinoma lung cancer patients independent of the smoking status. However, the suggested EV biomarker panel should be further evaluated and validated.

Biological role of EVs isolated from adenocarcinoma patients and healthy individuals differs as well. Healthy control derived EVs are most associated with metabolism, localization, and biogenesis, and their proteins are the most associated with vesicular components and with the enzymatic function. Proteins detected on EVs deriving from adenocarcinoma lung cancer patients are involved in the immune system response. Localization of adenocarcinoma-lung-cancer-derived-EV-proteins is the most associated with cytoplasmic and extracellular regions, as well as with complexes associated with lipids. Functionally, adenocarcinoma-lung-cancer-derived-EV-proteins are the most associated with binding, enzymatic regulation, and with the activation of phosphatidylcholine-sterol O-acyltransferase activity.

7. Outlook

NECT-4 and PDAP1 proteins have been analysed in this thesis as biomarkers for the lung cancer detection, and both of them showed a potential to be utilized with some limitations which should be overcome. Therefore, they should be further analysed on larger study cohort to obtain more precise statistical results. Furthermore, non-cancerous pulmonary disease, most commonly chronic obstructive pulmonary disease (COPD), present big challenge in lung cancer biomarker detection. In this thesis, the COPD status of healthy controls was not known since the COPD status was not initially included in the records during sample collection. Therefore, healthy controls having non-cancerous pulmonary diseases, e.g., COPD should be included in the study cohort in the future investigations.

CCN1 protein showed potential to be further analysed as CTC/DTC mesenchymal biomarker as no CCN1 detection was observed on healthy donor derived PBMCs. Such a potential continues to be investigated in the future, including the patient samples as well.

Sensitivity and the specificity of NECT-4 as a single biomarker obtained in this thesis is potentially too low to be implemented in clinical setting as a single biomarker. However, the inclusion of NECT-4 into composite biomarker panel could improve the utility and implementation of NECT-4 into clinical setting, and it should be tested in the future. Moreover, the use of NECT-4 could compensate the use of LDCT and reduce the high false positive rate obtained by LDCT. Combination of LDCT and NECT-4 in lung cancer screening holds the potential and it should be investigated in the future.

Finally, the investigation of differences in the EV proteome between adenocarcinoma patients and healthy individuals have been analysed. The analysis should be further validated on the larger cohort, as well including non-cancerous healthy controls, e.g., COPD$^+$ healthy controls. Current analysis suggested the detection of healthy individuals who are potentially developing lung cancer (at transition state). Healthy individuals recognized as at transition state should be further analysed, and potentially EV biomarkers for their detection could be developed allowing the early detection of lung cancer. Furthermore, suggested EV putative biomarker panel for the detection of adenocarcinoma patients independent of the smoking status comprising SAA2, REV1, SCLY, SLC12A6, PRSS1, GPM6A, and MSN proteins should be further investigated and validated.

Analysis of the EV proteome may also lead to new strategies for therapeutic interventions. For example, potential association of EVs in adenocarcinoma patients with the immunity/HIF-1α/cholesterol axis has been detected in this thesis and it should be investigated in the future. If such an association remains to be truthful, targeting EV-associated cholesterol might have an anti-tumor effect and to be utilized in therapy development.

8. Materials and methods

The chemicals and reagents used for experiments in this thesis are summarized in Table 17.

Table 17. Chemicals and reagents used in experiments in this thesis.

Chemical/reagent	Company
1,4-Dithiothreitol	Carl Roth, Karlsruhe, Germany
2-mercaptoethanol	Sigma-Aldrich, St. Louis, MO, USA
4',6-Diamidine-2-phenylindole (DAPI)	Carl Roth, Karlsruhe, Germany
$^{13}C_6$-Arginine-HCl	Pierce Biotechnology, Rockford, USA
$^{13}C_6$-Lysine-2HCl	Pierce Biotechnology, Rockford, USA
Acetone	Merck, Darmstadt, Germany
Acetonitrile hyper grade for LC-MS (ACN)	Merck, Darmstadt, Germany
Ammonium bicarbonate	Sigma-Aldrich, St. Louis, MO, USA
Ammonium persulfate	Carl Roth, Karlsruhe, Germany
Ammonium sulfate	Carl Roth, Karlsruhe, Germany
Bovine serum albumin (BSA)	Sigma-Aldrich, St. Louis, MO, USA
Bromphenol blue	Merck, Darmstadt, Germany
Coomassie brilliant blue G 250	Serva, Heidelberg, Germany
Dimethylsulfoxide (DMSO)	Serva, Heidelberg, Germany
DMEM for SILAC	Thermo Fisher Scientific, Waltham, MA, USA
Dulbecco's Modified Eagle's Medium (DMEM high glucose medium)	PAN Biotech, Aidenbach, Germany
Drystrip Cover Fluid	Sigma-Aldrich, St. Louis, MO, USA
Dulbecco's phosphate buffered saline (DPBS)	Thermo Fisher Scientific, Waltham, MA, USA
Epidermal growth factor (EGF)	Miltenyi Biotec, Bergisch Gladbach, Germany
Ethanol	Chemsolute/TH Geyer, Renningen, Germany
Ethylenediaminetetraacetic acid (EDTA)	Sigma-Aldrich, St. Louis, MO, USA
Fermacidal D2	IC Products SA, Gordola, Switzerland
Fetal bovine serum (FBS)	Capricorn Scientific, Ebsdorfergrung, Germany
Ficoll Paque™	GE Healthcare, Chalfont St. Giles, United Kingdom
Formic acid	Merck, Darmstadt, Germany
Glutaraldehyde	Sigma-Aldrich, St. Louis, MO, USA
Glycerol ReagentPlus®	Sigma-Aldrich, St. Louis, MO, USA
Glycine	Carl Roth, Karlsruhe, Germany
Human fibroblast growth factor (FGF-2)	Carl Roth, Karlsruhe, Germany
Insulin-Transferrin-Selen-Natrium-Pyruvat (ITS-A) (100X)	Thermo Fisher Scientific, Waltham, MA, USA
Iodoacetamide	Sigma-Aldrich, St. Louis, MO, USA
IPG buffer	GE HealthCare, Chalfont St. Giles, United Kingdom

Isopropyl alcohol	Chemsolute/ TH Geyer, Renningen, Germany
L-glutamine	Thermo Fisher Scientific, Waltham, MA, USA
L-proline for SILAC	Thermo Fisher Scientific, Waltham, MA, USA
Low fat powdered milk	Carl Roth, Karlsruhe, Germany
Methanol, BAKER ANALYZED™ A.C.S. Reagent	J.T. Baker, Deventer, Netherlands
N'-Tetramethylethylenediamine (TEMED)	Sigma-Aldrich, St. Louise, MO, USA
Nonidet P-40 (NP-40)	ICN Biomedicals, Inc., Pittsburgh, PA, USA
Ortho-Phosphoric acid 85% ROTIPURAN®	Carl Roth, Karlsruhe, Germany
PageRuler™ Plus Prestained Protein Ladder	Thermo Fisher Scentific, Waltham, MA, USA
Paraformaldehyde 16% (w/v), methanol free	Alfa Aesar, Haverhill, MA, USA
Penicillin/Streptomycin	Thermo Fisher Scientific, Waltham, MA, USA
Protease inhibitor cocktail: cOmplete ULTRA Tablets, Mini, EASYpack	Roche, Mannheim, Germany
Phosphatase inhibitor cocktail: PhosSTOP EASYpack	Roche, Mannheim, Germany
ROTIPHORESE®Gel 30	Carl Roth, Karlsruhe, Germany
RPMI 1640	PAN Biotech, Aidenbacj, Germany
Sequencing grade trypsin	Promega, Schnelldorf, Germany
Sodium deoxycholate (SDC)	Sigma-Aldrich, St. Louis, MO, USA
Sodium dodecyl sulfate (SDS)	Carl Roth, Karlsruhe, Germany
SignalFire™ Plus ECL Reagent A	Cell Signaling Technology, Danvers, MA, USA
Sodium azide	Fluka, Thermo Fisher Scientific, Waltham, MA, USA
Sodium bicarbonate (NaHCO$_3$)	Sigma-Aldrich, St. Louis, MO, USA
Sodium carbonate (Na$_2$CO$_3$)	Fluka, Thermo Fisher Scientific, Waltham, MA, USA
Sodium Chloride	Carl Roth, Karlsruhe, Germany
Sodium hydroxide	Merck, Darmstadt, Germany
Sulfuric acid	Merck, Darmstadt, Germany
Thimerosal	Sigma-Aldrich, St. Louis, MO, USA
Triethylammonium Bicarbonate (TEAB)	Thermo Fisher Scientific, Waltham, MA, USA
Tris-HCl	Carl Roth, Karlsruhe, Germany
Triton X-100	Fluka, Thermo Fisher Scientific, Waltham, MA, USA
Trizma® base	Sigma-Aldrich, St. Louis, MO, USA
Trypsin sequencing grade	Roche, Mannheim, Germany
Trypsin-EDTA solution 0.25% (w/v)	Gibco, Eggenstein, Germany
Tween20	Carl Roth, Karlsruhe, Germany
Urea	Sigma-Aldrich, St. Louis, MO, USA
Water	Honeywell Riedel-de-Haen™, Seezle, Germany

The composition of buffer solution used for experiments in this thesis are specified in Table 18.

Table 18. Composition of buffer solutions used in experiments.

Description	Composition
Lysis buffer	9.8 M Urea 30 mM Tris 15 mM EDTA deH$_2$0 (or MS-H$_2$O if used for MS)
LPIP buffer	140 mM NaCl 50 mM Tris HCl (pH 7.5) 1 mM EDTA 0.05% NP40 10% Glycerol 1:25 protease inhibitor cocktail c0mplete ULTRA Tablets, Mini, EASYpack 1:25 phosphatase inhibitor cocktail PhosSTOP EASYpack miliQ H$_2$O
Sample buffer (5 ×)	2.73 g Tris-HCl 50 g Glycerin 5 g Sodium dodecyl sulfate (SDS) 50 mg Bromophenol blue 3.86 g 1,4-Dithiothreitol (DTT) deH$_2$0 until 100 mL pH 6
Stacking gel buffer (pH 6.8)	50 mM Tris base 0.4% SDS deH$_2$O
Separation gel buffer (pH 8.8)	150 mM Tris base 1% SDS deH$_2$O
Laemmli electrophoresis running buffer	25 mM Tris base 192 mM Glycine 0.1% SDS deH$_2$O
TBST (10 ×)	48.4 g Tris base 160 g NaCl 20 mL Tween20 Until 2 L H$_2$O pH 7.6
Transfer buffer	43.6 mM Tris base 39 mM Glycine 10% Ethanol deH$_2$O
MS Extraction solution	65% ACN 5% Formic acid In MS-H$_2$O
Sep-Pak equilibration / washing buffer	5% Methanol 0.1% Formic acid In MS-H$_2$O
ELISA coating buffer	10 mM Na$_2$CO$_3$ 30 mM NaHCO$_3$ In miliQ H$_2$O pH 9.6
Stripping buffer	62.41 mM Tris-base 69.35 mM SDS 100 mM 2-mercaptoethanol In deH$_2$O pH 9.5

Coomassie fixing buffer	30% Ethanol 2% Phosphoric acid In Water Honeywell Riedel-de-Haen™
Coomassie equilibration buffer	1.14 M $(NH_4)_2SO_4$ 2% Phosphoric acid 18% Ethanol In Water Honeywell Riedel-de-Haen™
Coomassie blue staining solution (100 ×)	0.02 M Coomassie brilliant blue G 250
MS–buffer B (for LC-MS/MS)	0.1% (v/v) Formic acid In acetonitrile (ACN)
OFFGEL rehydration solution	20% Methanol 0.5% IPG buffer (pH 3-10) In Water Honeywell Riedel-de-Haen™

8.1 Patients

Analysis performed on human samples were done according to the Helsinki rules after the approval given by the ethics committee of the Medical Association Hamburg (reference number PV5392) and ethics committee of the Ruhr University Bochum (reference number 3217-08) was obtained. Written consent from all patients was obtained as well. Anonymized plasma samples from healthy controls, including both women and men, were obtained from University Medical Center Hamburg-Eppendorf, Hamburg, Germany for the analysis of NECT-4 and PDAP1 proteins using ELISA, and for the optimization of EV isolation and final analysis of EVs using mass spectrometry. Anonymized plasma samples from lung cancer patients, including women and men, were obtained from the Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, Germany for the analysis of NECT-4 and PDAP1 proteins using ELISA, and from University Medical Center Hamburg-Eppendorf, Hamburg, Germany for optimization of EV isolation and final analysis of EVs using mass spectrometry. Anonymized blood samples from healthy donors were obtained from University Medical Center Hamburg-Eppendorf, Hamburg, Germany for isolation of PBMCs using density gradient centrifugation.

8.2 Cell culture methods

To avoid contamination of cell cultures, all cell culture experiments were performed in sterile environment, including work under a vertical laminar flow cabinet using sterile disposable tips and pipettes, as well as disinfection of all material and work surfaces with Fermacidal D2 before and after the work has been performed. Upon starting cell culture work, all liquids that have been used were preheated to 37 °C in a water bath which contained 2 mL of AQUA-STABIL per 1 L of water.

8.2.1 Thawing cell lines

All cell lines stored in cryotubes were thawed on room temperature and added into $10 \times$ volume of corresponding medium to dilute DMSO. Cell suspension was centrifuged with Rotofix32 centrifuge (Hettich, Kirchlengern, Germany) at 1,200 rpm (zyto-rotor, six-fold) for three minutes, supernatant was discarded, and cells were resuspended in corresponding medium and cultured in 25 cm²-flask (Sarstedt, Nümbrecht, Germany).

8.2.2 Cryopreservation of cell lines

Cells were washed with DPBS (Thermo Fisher Scientific, Waltham, MA, USA) and detached with 2 mL of 0.25% Trypsin-EDTA (Thermo Fisher Scientific) for one minute, after which trypsin was deactivated by adding 10 mL of corresponding medium. Cell suspension was centrifuged with Rotofix32 centrifuge (Hettich) at 1,200 rpm (zyto-rotor, six-fold) for three minutes. Supernatant was discarded, and cells were resuspended in corresponding medium consisting 10% of DMSO, transferred into cryotubes and stored at -80 °C.

8.2.3 Cultivation of human cell lines

All cell lines were cultivated at 37 °C in a sterile and humified environment and used cell lines are summarized in Table 19. MDA-MB-231 and MCF-7 cell lines were cultivated in DMEM medium supplemented with 10% fetal bovine serum (Capricorn, Ebsdorfergrund, Germany), 2 mmol/L L-glutamine (Gibco/Thermo Fischer Scientific) and 100 U/mL Penicillin/Streptomycin (Gibco/Thermo Fischer Scientific) and held under 10% CO_2 and atmospheric oxygen concentration using the Heracell 150i incubator (Thermo Fischer Scientific). H2228, HCC-366, HTB-58, H1975, H1299, H1395, and H1993 cell lines were cultivated in RPMI medium (Pan-Biotech, Aidenbach, Germany) supplemented with 10% fetal bovine serum (Capricorn), 2 mmol/L L-glutamine (Gibco/Thermo Fischer Scientific) and 100 U/mL Penicillin/Streptomycin (Gibco/Thermo Fischer Scientific) and held under 5% CO_2 and atmospheric oxygen concentration. LC-M1 cell line was cultured in RPMI (Pan-Biotech, Aidenbach, Gemrnay) medium supplemented with 10% fetal bovine serum (Capricorn), 2 mmol/L L-glutamine (Gibco/Thermo Fischer Scientific), 100 U/mL Penicillin/Streptomycin (Gibco/Thermo Fischer Scientific, Waltham, USA), 1% IST (Insulin, Transferin, Selen; Gibco/Thermo Fischer Scientific), 50 ng/mL human-EGF and 10 ng/mL human-bFGF (both from Miltenyi Biotec, Bergisch Gladbach, Germany) and held under 10% O_2 and 5% CO_2 using the Heracell 15 incubator (Thermo Fischer Scientific). Oxygen level was controlled by nitrogen.

Cells were grown in 75 cm²-cell-culture-flasks (Sarstedt) until 80-90% confluency and then washed with sterile DPBS (Thermo Fischer Scientific), detached with 2 ml of 0.25% Trypsin-EDTA (Thermo Fischer

Scientific) for ca. three minutes at 37 °C after which 10 ml of corresponding medium was used to deactivate trypsin activity. Cell suspension was centrifuged with Rotofix32 centrifuge (Hettich) at 1,200 rpm (zyto-rotor, six-fold) for three minutes, supernatant was discarded, and cells were resuspended in corresponding medium and split into new cell culture flasks.

Table 19. Summary of cell lines used in experiments.

Cell line	Origin	Source	Date of authentication
MDA-231	Breast adenocarcinoma, pleural effusion	ATCC HTB-26, Manassas, VA, USA	02/2014
MCF-7	Breast adenocarcinoma, pleural effusion	ATCC HTB-22, Manassas, VA, USA	10/2017
HCC-366	Adeno-squamous carcinoma; NSCLC	DSMZ ACC 492, Braunschweig, Germany	Purchased on 09/2018
SK-MES-1 (HTB-58)	Lung, squamous cell carcinoma	ATCC HTB-58, Manassas, VA, USA, kindly provided by Dr. Harriet Wikman	2015
H1975	Lung, adenocarcinoma, NSCLC	ATCC CRL-5908, Manassas, VA, USA	05/2019
H1299	Lung; lymph node; NSCLC	ATCC CRL-5803, Manassas, VA, USA	06/2018
H1395	Lung, stage 2, adenocarcinoma	ATCC CRL-5868, Manassas, VA, USA	05/2015
H1993	Lung: lymph node; stage 3A, adenocarcinoma; NSCLC	ATCC CRL-5909, Manassas, VA, USA, kindly provided by Dr. Harriet Wikman	2015
LC-M1	Lung carcinoma, bone marrow	ITB, UKE	05/2015
H2228	Lung, adenocarcinoma; NSCLC	ATCC CRL-5935, Manassas, VA, USA	No recent authentication was performed

8.2.4 Collection of cell culture supernatant for ELISA

Cell lines were cultivated in corresponding medium in 75 cm^2-flasks (Sarstedt) until 80-90% confluence, split (as described in section 8.2.3) to new 75 cm^2-flasks (Sarstedt) and cultivated for 72 hours. Supernatant was collected, centrifuged at 2,500 ×g for 15 min at 4 °C to remove dead cells, aliquoted on ice into 1.5 mL conical polypropylene tubes (Sarstedt) and stored at -80 °.

8.2.5 Isolation of peripheral blood mononuclear cells and plasma from whole blood

Blood was collected from healthy donors (Department of Transfusion Medicine, University Medical Center Hamburg-Eppendorf) in 9.0 mL S-Monovette EDTA containing tubes (Sarstedt) and subjected to centrifugation at 300 ×g for 10 minutes at room temperature. The supernatant was transferred into a new tube, while PBMCs remained in the pellet. Supernatant was additionally centrifuged at 1 800 ×g for 10 minutes at room temperature. Supernatant consisting of plasma was transferred into a new tube and stored at -80 °C, and if needed, PBMCs were further processed as described below in this section. Otherwise, pellet was discarded.

PBMCs were isolated by density gradient centrifugation using Ficoll-Paque™ (GE Healthcare Chicago, USA). Pellet containing PBMCs and conducted as described above in this section, or 7.5 mL of blood was diluted with DPBS (Thermo Fischer Scientific) to a volume of 20 mL. Sample was gently transferred into a new 50 mL conical polypropylene screw cap tube and layered on 20 mL of Ficoll Paque™ (GE Healthcare). Suspension was centrifuged at 1,400 rpm for 30 min at room temperature without centrifuge breaks. Upper layer consisting mainly DPBS was discarded and interphase ring consisting of PBMCs was carefully collected and transferred to a new 50 mL mL conical polypropylene screw cap tube, without collecting bottom Ficoll-layer, since Ficoll is toxic for cells. Suspension consisting of PBMCs was filled up to 50 mL with DPBS (Thermo Fischer Scientific) and subjected to centrifugation at 1,400 rpm for 10 minutes at room temperature. Supernatant was discarded and pellet consisting of PBMCs was resuspended lysis buffer (Table 18) and the protein concentration was measured using BCA (section 8.3.2).

8.2.6 Authentication of cell lines

For isolation of genomic DNA (gDNA) from cell lines, NucleoSpin Tissue Kit (Macherey-Nagel, Düren, Germany) was used according to the manufacturer instruction. Nanodrop ND100 spectrometer was used for concentration and quality measurements of gDNA. Authentication was performed by the company Multiplexion GmbH (Heidelberg, Germany) and analysis was based on Single Nucleoide Polymorphism (SNP)-profiling. LC-M1 cell line cannot be commercially authenticated since the cell line was generated by Prof. Pantel in 1994 (282, 283), therefore the authentication companies do not have a reference genome. Instead, the authentication is performed at Institute of Tumor Biology, University Medical Center Hamburg-Eppendorf using immunofluorescent double staining for pan-cytokeratin/vimentin. Authentication of HCC-366 cell line was not conducted since the given cell line was freshly bought and used, therefore authentication was not needed.

Authentication dates are given in the form of a summary of cell lines in Table 19. After authentication, cells were cryopreserved and resuscitated for this work.

8.2.7 Mycoplasma test

Cell lines were tested once a month for mycoplasma contamination using the Venor GeM Classic Detection kit (Minerva Biolabs, Berlin, Germany) according to the manufacturer's instruction. Only cell lines negative for mycoplasma contamination were used in this work.

8.2.8 Stable isotope labeling by amino acids in cell culture

Stable isotope labelling by amino acids (SILAC) is used to quantify changes in expression profiles of different cell populations. To compare epithelial and mesenchymal phenotypes of lung cancer cell lines and subsequently to find potential epithelial or mesenchymal biomarkers for detection of lung cancer, H1975 cell line, which expresses an epithelial/mesenchymal hybrid phenotype, and LC-M1 cell line, which expresses mesenchymal phenotype were chosen as models to work with.

Cell lines were cultivated as described before (see 8.2.3 above) and metabolic labeling was performed using Pierce SILAC Protein Quantitation Kit (Pierce Biotechnology/Thermo Fischer Scientific) according to the manufacturer's instruction. RPMI 1640 medium for SILAC, lacking arginine and lysine was prepared as described before using dialyzed FBS. Lysine $2 \times$ HCl and arginine $\times$ HCl were added to the media in concentrations of 0.46 mM and 0.47 mM, respectively.

During protein synthesis in a cell, heavy amino acids are being incorporated resulting in a mass shift of 6 Da which is detectable by mass spectrometer and allows a proteome comparison of H1975 and LCM1 cell lines.

If there is insufficient amount of proline in medium, conversion of arginine into proline in urea cycle can occur during SILAC labeling. Conversion of arginine into proline would result in a division of the expected heavy peptide ion signal into a heavy peptide ion signal with decreased intensity and an additional heavy peptide ion signal containing heavy proline. This extra peak shows additional mass shift of 5 Da. To avoid this metabolic process and conversion of heavy arginine into heavy proline, $^{12}C_5$ proline was added into media to a final concentration of 0.48 mM.

Labeling was performed in biological triplicates, where H1975 was cultivated with RPMI 1640 medium supplemented with heavy $^{13}C_6$ arginine and $^{13}C_6$ lysine (heavy SILAC medium), and LC-M1 was cultivated in complete RPMI 1640 medium supplemented with $^{12}C_6$ arginine and $^{12}C_6$ lysine (light SILAC medium). Additionally, another labeling in triplicates was performed to avoid possible misinterpretation of results due to the contamination, where labelling was performed the other way

around. LC-M1 was cultivated with complete RPMI 1640 medium supplemented with heavy $^{13}C_6$ arginine and $^{13}C_6$ lysine (heavy SILAC medium), and H1975 was cultivated in RPMI 1640 medium supplemented with $^{12}C_6$ arginine and $^{12}C_6$ lysine (light SILAC medium).

8.3 Biochemical and immunological protein methods

8.3.1 Protein isolation from cell lines

At the confluency of 80-90% of 75 cm^2 – flasks, cells were washed three times with 37 °C pre-warmed DPBS (Thermo Fischer Scientific). After final washing, DPBS was fully removed and 300 µL of 9.8 M urea lysis buffer (Table 18) was added. Cells were scraped with a scraper and transferred to the 1.5 mL conical polypropylene tube (Sarstedt). Cell suspension was incubated for 1.5 h at room temperature with shaking on ThermoMixer® C (Eppendorf, Hamburg, Germany) and homogenized by ultrasonic treatment every 30 minutes. After incubation, cell suspension was subjected to centrifugation at 8,000 ×g for three minutes at room temperature. Supernatant was transferred into a new 1.5 mL conical polypropylene tube (Sarstedt) and stored at -20 °C.

Cellular proteins for enzyme-linked immunoassay (ELISA) were extracted as follows. Cells were detached at the confluency of 80-90% with 2 mL of 37 °C pre-warmed Trypsin and centrifuged at 1,200 rpm for three minutes at room temperature. Supernatant was removed and cell pellet was lysed on ice in 300 µL of LPIP buffer (Table 18) supplemented with protease inhibitor cocktail cOmplete ULTRA Tablets, Mini, EASYpack (Roche, Mannheim, Germany) and phosphatase inhibitor cocktail PhosSTOP EASYpack (Roche) in ratio 1:25. Cell homogenization was performed by ultrasonic treatment on ice and samples were stored at -80 °C.

8.3.2 Determination of protein concentration

Protein concentration of cells was measured using the Pierce BCA Protein Assay Kit (Thermo Fischer Scientific) according to the manufacturer's instructions. Two and a half microliter of a sample was added in a total volume of 25 µL. BSA standard series were prepared to generate standard curve in order to convert measured optical density of samples into protein concentration. Standard curve comprised BSA concentrations of 100, 200, 400, 600, and 1,000 µg/mL, each containing 2.5 µL of corresponding lysis buffer in a total volume of 25 µL. Optical densities were measured by BioPhotometer with thermal printer DPU-414 (Eppendorf). The extinction was detected at 562 nm.

To determine concentration of proteins derived from EVs, either Pierce BCA Protein Assay Kit (Thermo Fischer Scientific) or Qubit® Assay Kit (Invitrogen/Thermo Fisher Scientific) was used according to the manufacturer's instruction. Pierce BCA Protein Assay Kit was used to determine protein concentration

in SEC fractions during EV isolation optimization. Standard series were prepared for generation of standard curve with BCA concentrations of 25, 125, 250, 500, 750, 1,000, 1,500, and 2,000 µL/mL. SEC fractions and BCA standards were pipetted in 96-well microtiter plate (Sarstedt) in a volume of 25 µL and 200 µL of BCA reagents was added. Plates were incubated for 30 minutes at 37 °C and optical density was determined by microplate reader Infinite NanoQuant M200 Pro (Tecan, Männedorf, Switzerland) and the microplate reader software i-Control™ (Tecan, software version 1.11).

For EV samples of low volumes, Qubit® Assay was used instead of Pierce BCA Protein Assay, because it allows to use low amount of sample (1 µL for example instead of 25 µL). Sample dilutions and standards were prepared according to the manufacturer's instruction. Each Qubit® standard in a volume of 10 µL was added to 190 µL of Qubit® working solution. Samples were as well diluted in Qubit® working solution in total volume of 200 µL.

8.3.3 SDS-PAGE

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) with Laemmli-buffered system (284) was used for size-based separation of proteins. Prior to separation itself, 20 µg of total cell line proteins or 50 µg of total PBMC proteins were diluted and denatured in sample buffer (Table 18) under reducing conditions for 5 minutes at 95 °C and subjected to short centrifugation. Depending on follow up experiments and target protein size, separating polyacrylamide gels containing different percentages of acrylamide was used (see Table 20). Depending on the following experiment, separating gels were produced manually by casting gel into an empty novex gel cassette (Life technologies, CA, USA) or commercial Invitrogen Novex™ WedgeWell™ 4-12% Tris-Glycine Gels (Invitrogen/Thermo Fisher Scientific) or BioRad Mini-PROTEAN TGX 4-15% Gels (Bio-Rad Laboratories, Hercules, USA) were purchased (Table 20). For manually produced gels, stacking gel consisted 5% of acrylamide and novex combs (Life technologies) were used. Polyacrylamide gels used for MS analysis were manually produced as described above, but one day before the corresponding experiment to ensure that all acrylamide monomers have polymerized or oxidized before sample loading.

As a molecular weight standard, Page Ruler™ Plus pre-stain protein ladder (10-250 kDa) (Thermo Fischer Scientific) was used and electrophoresis was run in Laemmli electrophoresis buffer using Novex XCell Sure-Lock mini system (Invitrogen, Thermo Fisher Scientific, Groningen, Netherlands) for manually casted gels, or Mini Protean® Tetra Cell (Bio-Rad Laboratories) for commercially available gradient gels. For manually casted gels, voltage was set to 80 V, and it was increased to 120 V after proteins entered separating gel. For commercially available gradient gels, voltage was set to 100 V. As follow up procedure, gels were subjected either to Western blot analysis or to Colloidal Coomassie staining and to mass spectrometry sample preparation and analysis.

Table 20. Polyacrylamide gel composition and their use in experiments.

Acrylamide (%)	Composition	Experiment
5	ROTIPHORESE®Gel 30 0.65 mL H_2O 3.045 mL Stacking gel buffer (Table 18) 1.25 mL 10% ammonium persulfate 0.025 mL TEMED 0.005 mL	All experiments where manual casting gels were produced
8	ROTIPHORESE®Gel 30 2.67 mL H_2O 4.67 mL Separation gel buffer (Table 18) 2.5 mL 10% ammonium persulfate 0.1 mL 10% SDS 0.1 mL TEMED 0.01 mL	Western blot on cell lines
10	ROTIPHORESE®Gel 30 2.08 mL H_2O 2.5 mL Separation gel buffer (Table 18) 1.56 mL 10% ammonium persulfate 0.1 mL 10% SDS 0.05 mL TEMED 0.0075 mL	Western blot on cell lines; Colloidal Coomassie staining
12	ROTIPHORESE®Gel 30 4.0 mL H2O 3.4 mL Separation gel buffer (Table 18) 2.5 mL 10% ammonium persulfate 0.1 mL 10% SDS 0.1 mL TEMED 0.004 mL	Western blot on cell lines
Gradient gel 4-12	Commercial (Novex™ WedgeWell™ 4-12% Tris-Glycine Gel (Invitrogen, CA, USA))	Western blot on PBMCs
Gradient gel 4-15	Commercial (Mini-PROTEAN TGX Gels (BioRad, Illionis, USA)	Western blot on EVs

8.3.4 Western blot analysis

To detect target proteins by Western blot analysis, proteins were transferred from polyacrylamide gels onto a FluoroTrans W PVDF Transfer membrane (PALL Life Sciences, Port Washington, NY, USA) using the mini-VE vertical electrophoresis system (GE Healthcare, Chalfont St Giles, United Kingdom) containing tank blot transfer units. Prior to transfer, the gel was equilibrated in transfer buffer (Table 18) for 30 minutes, and PVDF membrane was first wet for 15 minutes in methanol and then equilibrated in transfer buffer for 15 minutes as well. These steps were performed on room temperature with mild shaking using Rotamax 120 shaker (Heidolph Instruments GmbH & Co. KG, Schwabach, Germany). In parallel, Whatman papers and thin and thick sponges were soaked in transfer buffer for minimum 30 minutes. For transferring proteins from a gel onto a membrane, a *Western blot sandwich* was assembled

starting from cathode as following: a thin sponge, a thick sponge, three Whatman papers, gel, membrane, a Whatman paper, a thick sponge, and a thin sponge. Transfer was run under following conditions: constant electric current of 160 mA per tank blot transfer unit, between 17-22 V depending on a run, and from 27-45 minutes transfer depending on a size of the protein. After the transfer, membrane was briefly washed in TBS-T (Table 18) and depending on a target protein blocked in corresponding blocking buffer (Table 21) for 1 h on a room temperature with mild shaking on a Rotamax 120 shaker (Heidolph Instruments GmbH & Co. KG). Membrane was briefly washed in TBS-T and incubated with the corresponding primary antibody (Table 21) overnight, on 4 °C, with mild shaking on a Tube roller RS-TR05 (Phoenix Instruments, Garbsen, Germany), except for APOA which was incubated for 72h. The day after, membrane was washed three times with TBS-T for 15 minutes each time, on room temperature and with mild shaking on a Rotamax 120 shaker (Heidolph Instruments GmbH & Co. KG). The incubation with corresponding horseradish peroxidase (HRP)-conjugated secondary antibody (Table 21) for 1.5 h on room temperature with mild shaking on a Tube roller RS-TR05 (Phoenix Instruments) was conducted. The SignalFire Plus ECL reagent (Cell Signaling Technology, Danvers, USA) (1:1=reagent A:reagent B) was applied as a substrate to HRP-conjugated secondary antibody to the membrane for three minutes. The release of light as a byproduct of the chemical reaction between HRP and a substrate was detected on X-ray films CEA CR5 UV (AGFA HealthCare, Mortsel, Belgium) which were placed on membranes and therefore allows the detection of target protein. X-ray films were developed with Curix 60 Processor (AGFA HealthCare, Bonn, Germany) and scanned with EPSON perfection V750 PRO (Epson, Suwa, Japan) for documentation.

To re-use membranes, membrane bound antibodies were removed by incubation with 25 mL of stripping buffer (Table 18) for 10 minutes in hot water followed by one hour incubation on a room temperature with mild shaking on a Tube roller RS-TR05 (Phoenix Instruments). Membranes were blocked in corresponding blocking buffer (Table 21) depending on a target protein and following procedure was as described above.

Table 21. Primary and secondary antibodies used for Western blot analysis.

Antibody	Clone	Pr. Ab. Company	Blocking buffer	Dilution of primary Ab	Dilution of Secondary* Ab
Anti-pan-cytokeratin antibody	A45/BB3	Micromet, Munich, Germany	5% low fat powdered milk in TBS-T	1:10,000 in 5% low fat powdered milk in TBS-T	Goat-anti-mouse 1:10,000 in 5% low fat powdered milk in TBS-T
Anti-Cytokeratin Epithelial Antibodies	AE1 and AE3 (mixed 1:1)	Merck (Chemicon), Darmstadt, Germany	5% low fat powdered milk in TBS-T	1:10,000 in 5% low fat powdered milk in TBS-T	Goat-anti-mouse 1:10,000 in 5% low fat powdered milk in TBS-T

CK8	Ks8.7	Progen Biotechnik, Heidelberg, Germany	5% low fat powdered milk in TBS-T	1:10,000 in 5% low fat powdered milk in TBS-T	Goat-anti-mouse 1:5,000 in 5% low fat powdered milk in TBS-T
CK18	Ks18.04	Progen Biotechnik, Heidelberg, Germany	5% low fat powdered milk in TBS-T	1:20,000 in 5% low fat powdered milk in TBS-T	Goat-anti-mouse 1:5,000 in 5% low fat powdered milk in TBS-T
CK19	BA-17	Antibodies-online, Atlanta, USA	5% low fat powdered milk in TBS-T	1:100,000 in 5% low fat powdered milk in TBS-T	Goat-anti-mouse-HRP 1:10,000 in 5% low fat powdered milk in TBS-T
EpCam	VU-1D9	Novocastra/Leica, Nussloch, Germany	5% low fat powdered milk in TBS-T	1:1,000 in 5% low fat powdered milk in TBS-T	Gpat-anti-mouse 1:20,000 in 5% low fat powdered milk in TBS-T
EGFR	D38B1	Cell Signaling Technology, Danvers, USA	5% low fat powdered milk in TBS-T	1:1,000 in 5% low fat powdered milk in TBS-T	Goat anti-rabbit-HRP 1:1,000 in 5% low fat powdered milk in TBS-T
N-CADHERIN	D4R1H	Cell Signaling Technology, Danvers, USA	5% low fat powdered milk in TBS-T	1:5,000 in 5% BSA in TBS-T	Goat-anti-rabbit 1:1,000 in 5% low fat powdered milk in TBS-T
VIM	RV202	BD Pharmingen, Erembodegem, Belgium	5% BSA in TBS-T or 5% low fat powdered milk in TBS-T	1:200,000 in 5% low fat powdered milk in TBS-T	Goat-anti-mouse 1:20,000 in 5% low fat powdered milk in TBS-T
CCN1	D4H5D	Cell Signaling Technology, Danvers, USA	5% BSA in TBS-T	1:10,000 in 5% BSA in TBS-T	Goat-anti-rabbit 1:5,000 in 5% low fat powdered milk in TBS-T
NECT-4	EPR15613-68	Abcam, Cambridge, United Kingdom	5% low fat powdered milk In TBS-T	1:1,000 in 5% low fat powdered milk in TBS-T	Goat-anti-rabbit 1:5,000 in 5% low fat powdered milk
PDAP1	polyclonal	Proteintech, Manchester, United Kingdom	5% low fat powdered milk in TBS-T	1:800 in 5% low fat powdered milk in TBS-T	Goat-anti-rabbit 1:4,000 in 5% low fat powdered milk
APOA	B-10	Santa Cruz Biotechnology, Dallas, Texas, USA	5% BSA in TBS-T	1:500 in 5% low fat powdered milk in TBS-T	Goat-anti-mouse 1:1,000 in 5% low fat powdered milk
APOB	C1.4	Santa Cruz Biotechnology, Dallas, Texas, USA	5% BSA in TBS-T	1:2,000 in 5% low fat powdered milk in TBS-T	Goat-anti-mouse 1:3,000 in 5% low fat powdered milk
CD9	D801A	Cell Signaling Tehcnology, Danvers, USA	5% BSA in TBS-T	1:1,000 in 5% BSA in TBS-T	Goat-anti-rabbit 1:1,000 in 5% low fat powdered milk
α-TUBULIN	11H10	Cell Signaling Technology, Danvers, MA, USA	5% BSA in TBS-T	1:20,000 in 5% BSA in TBS-T	Goat-anti-rabbit 1:2,000 in 5% low fat powdered milk

*All secondary antibodies were purchased from DAKO, Glostrup, Denmark

8.3.5 Colloidal Coomassie staining of polyacrylamide gels (Neuhoff staining)

After electrophoresis, polyacrylamide gels were stained according to Neuhoff *et al.* (285) with following modifications. All steps were performed on a room temperature and with gentle shaking. Gels were fixed in Coomassie fixing buffer (Table 18) for 30 minutes, and subsequently washed in miliQ H$_2$O for 30 minutes, followed by shrinking of a gel in 99% ethanol until it would become completely white. Shrinking of a gel was performed in order to remove SDS out of gel. Washing step was repeated 2 ×, after which gel was rehydrated in miliQ H$_2$O for 30 minutes. Afterwards, gel was incubated in Coomassie equilibration buffer (Table 18) for 30 minutes, and finally stained with Coomassie blue staining solution (Table 18) overnight. The day after, Coomassie blue staining solution was discarded, and the gel was thoroughly washed in miliQ H$_2$O until blue background in the gel was removed.

8.3.6 Precipitation of proteins

For precipitation of proteins precipitant and co-precipitant from 2-D Quant kit (GE HealthCare, Chalfont St Giles, United Kingdom) were used as described in Bartkowiak et al. (286). Half of the sample was transferred to a 1.5 mL conical polypropylene tube and 600 µL of precipitant was added. Sample was vortexed and shaken for 10 minutes on room temperature on Clay Adams Nutator shaker model no. 1106 (Marshall Scientific, Hampton, USA). Co-precipitant was added in a volume of 600 µL and sample was vortexed vigorously and shaken for additional 10 minutes on room temperature on Clay Adams Nutator shaker model no. 1106 (Marshall Scientific). After precipitation and co-precipitation, sample was subjected to centrifugation with Sigma 1-15K centrifuge (Sigma, Osterode am Harz, Germany) for 10 minutes at 16,000 ×g on room temperature and supernatant was discarded. Other half of the sample was added to the same tube and the steps of precipitation and co-precipitation were performed as described above. Supernatant was discarded and the pellet was subjected to centrifugation in Sigma 1-15K centrifuge (Sigma) for one minute at 16,000 ×g after which the remaining liquid was discarded. Pellet was washed within less than two minutes with 1 mL of 80 mM Tris in 80% acetone by resuspending with a pipette tip and vortexing, and then subjected to centrifugation in Sigma 1-15K centrifuge (Sigma) for two minutes at 16,000 ×g. Supernatant was discarded, and pellet was additionally subjected to centrifugation for one minute at 16,000 ×g after which the remaining liquid was discarded. One milliliter of 80% acetone was added to the pellet and sample was vortexed vigorously and subsequently subjected to centrifugation with a Sigma 1-15K centrifuge (Sigma) for two minutes at 16,000 ×g, after which the supernatant was discarded. Pellet was additionally subjected to centrifugation with Sigma 1-15K centrifuge (Sigma) for one minute at 16,000 ×g and remaining liquid was discarded. Pellet was solved in fresh 9.8 M urea dissolved in HPLC water (Honeywell Riedel-de-Haen, Seelze, Germany) if following analysis was mass spectrometry, or in 9.8 M urea lysis buffer (Table 18) if

following analysis was *Western Blot* and shaken on Thermomixer® C (Eppendorf) for few minutes until sample was not solved completely.

8.3.7 LC-MS/MS sample preparation

In this work bottom-up proteomics was performed, and to digest proteins into peptides a serine protease trypsin was used for proteolytic cleavage of amide bond in between amino acids. Trypsin hydrolyses amide bond on carboxyl side of arginine and lysine, except when arginine or lysine is bound to proline (Figure 61).
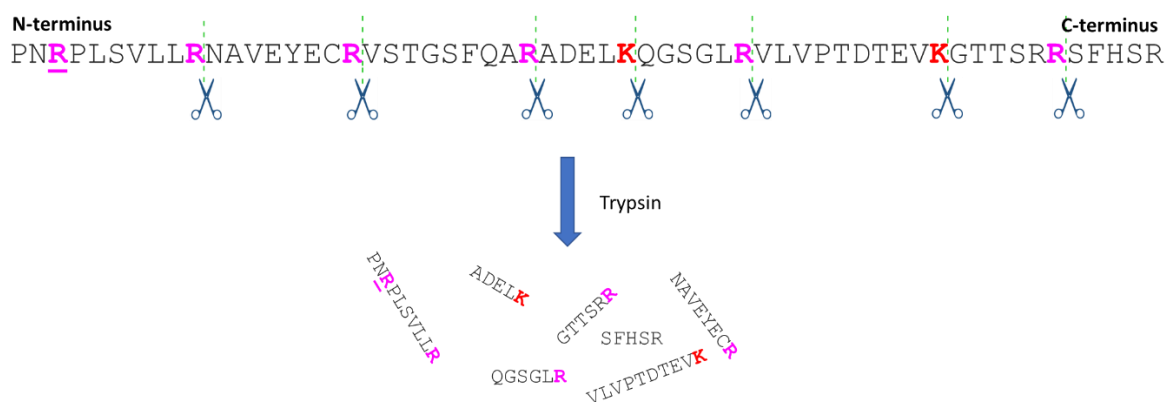

Figure 61. Tryptic digestion of proteins. Denatured proteins were treated with trypsin and proteolytically cleaved on carboxyl side of K and R, except if the next following amino acid is P. Reaction result in mixture of peptides possessing one K or R. R is before P, therefore it is not recognized by trypsin, which results in the corresponding peptide to have two Rs after tryptic digestion. R=arginine before proline; R=arginine; K=lysine. Created with BioRender.com.

Prior to the protein digestion, disulfide bonds were reduced with dithiothreitol (DTT) (Carl Roth, Karlsruhe, Germany) and free sulfhydryl groups were alkylated with iodoacetamide (IAA) as described by Suttapitugsakul et al. (287). Reaction of IAA with SH-groups and DTT creates an acidic environment in which trypsin is inactive. To increase pH to the optimal one for the tryptic digestion (7.5-8.5), 2 M Tris was added, and pH was controlled with pH-fix 0-14 stripes (Macherey-Nagel). For all dilutions used in LC-MS/MS sample preparation, HPLC water (Honeywell Riedel-de-Haen) was used.

Proteins were precipitated as described in 8.3.6 and subsequently protein concentration was measured using BCA test as described before in 8.3.2.

8.3.7.1 In-gel tryptic digestion (Labelling quality control)

For quality check of labeling, 15 µg of proteins per sample were loaded separately on SDS-gel and SDS-PAGE and Coomassie staining were performed as described in 8.3.3 and in 8.3.5. In-gel tryptic digestion was performed according to Shevchenko *et al.* (288) with modifications. Two bands were cut out per lane where one of the lanes was on the position of actin (55 kDa) to control amino acid conversion. Bands were chopped into 1 mm^2 pieces and transferred into Safe-Lock Tubes 1.5 mL (Eppendorf). For de-staining of gel pieces, 1 mL of 50% methanol was added, and samples were shaken for 30 minutes on Thermomixer® C (Eppendorf) on 50 °C. The supernatant was discarded and remaining liquid inside of gel pieces was removed by shrinking gel pieces in 250 µL of acetonitrile (ACN). Samples were shaken for 30 minutes on Thermomixer® C (Eppendorf) on room temperature, and gel pieces became shrinked and white. Supernatant was discarded, and samples were incubated with open lid on Thermomixer® C (Eppendorf) without shaking on 37 °C for 15 minutes. Samples were reduced with 250 µL of fresh 10 mM DTT (Carl Roth) in 100 mM NH_4HCO_3 for one hour on Thermomixer® C (Eppendorf) at 56 °C with shaking. Supernatant was discarded and 250 µL of fresh 55 mM IAA (AppliChem, Darmstadt, Germany) in 100 mM NH_4HCO_3 was added and incubated for 45 minutes on Thermomixer® C (Eppendorf) on room temperature in dark. Supernatant was discarded and IAA was washed away with 250 µL of 100 mM $NH4HCO3$ for 15 minutes on room temperature, after which the supernatant was discarded. Remaining liquid was removed by incubating sample with 250 µL of ACN for 15 minutes on room temperature and liquid was discarded again. Both steps were repeated twice and performed on Thermomixer® C (Eppendorf) with shaking. Samples were left with an open lid to dry completely on Thermomixer® C (Eppendorf) without shaking at 37 C for 15 minutes. 250 µL of 1 mM HCl was added to 25 µg of lyophilized trypsin (Trypsin sequencing grade, Roche) and left in the fridge for 30 minutes to allow trypsin activation. 750 µL of 50 mM NH_4HCO_3 was added to the activated trypsin and 4 µL of prepared trypsin solution was added to the gel pieces. 50 mM NH_4HCO_3 was added in a volume sufficient to cover gel pieces (approximately 70 µL) and incubated overnight on Thermomixer® C (Eppendorf) at 37 °C with shaking. The day after, tryptic peptides were extracted at 37 °C on a shaker in two repetitive cycles consisted of shrinking step using MS-extraction solution (Table 18) and swelling step using HPLC water (Honeywell Riedel-de-Haen, Seelze). Supernatants were collected in a new 1.5 mL conical polypropylene tube (Eppendorf) and dried out by evaporation until complete dryness using UNIVAPO 100K vacuum-centrifuge (Uniequip, Planegg, Germany).

8.3.7.2 In-gel tryptic digestion (BCA test quality control)

Since one of the goals of this thesis was to perform comparative proteomics and to identify biomarker proteins for detection of lung cancer cell with an epithelial or mesenchymal phenotype, it was crucial to determine protein concentration of each sample precisely. If quality check of labeling was satisfactory, quality check of BCA test of study samples was performed further on. To check the precision of BCA

test, first heavy- and light-labelled samples were treated separately. Samples were treated with 0.05 unit of sample volume with 1 M DTT (Carl Roth) in 1 M Tris for 20 minutes at 56 °C on Thermomixer® C (Eppendorf) with shaking. Two hundred millimolar IAA (AppliChem) was added to each sample, vortexed, and additionally 1.7 units of added IAA amount of 2 M Tris was added to the sample. Since reactions of IAA with SH- residues and DTT create H^+, pH of samples was tested with pH-fix 0-14 stripes (Macherey-Nagel). The pH value was adjusted to 8-9.5 using 2 M Tris. When appropriate pH was achieved, samples were incubated for 30 minutes at room temperature, in the dark with shaking. Subsequently, samples were precipitated as described above in section 8.3.6. Ten microliters of sample were diluted 1:5 and BCA test was performed as described in 8.3.2 and the rest of the sample was stored on -20 °C. After determination of protein concentration, heavy- and light-labelled samples of H1975 and LC-M1 were mixed in ratio 1:1 and total of 20 µg of protein was loaded on SDS-gel and SDS-PAGE and Coomassie staining were performed as described in 8.3.3 and in 8.3.5. The following procedure was the same as described in 8.3.7.1, except for reduction and alkylation steps with DTT and IAA, since these treatments were performed before mixing samples.

8.3.7.3 Tryptic digestion of homogenized cell lysates

If labeling quality check of each cell line, and BCA test quality check of all samples were satisfactory, samples were thawed and 400 µg of heavy and 400 µg of light labelled proteins were mixed in triplicates (1 : 1=H1975 SILAC heavy : LC-M1 SILAC light or 1 : 1=H1975 SILAC light : LC-M1 SILAC heavy). These samples were previously reduced and alkylated as described in 8.3.7.2. For preparation of trypsin solution, 400 µL of 1 mM HCl was added to 100 µg of lyophilized trypsin (Trypsin sequencing grade, Roche) and incubated for 30 minutes in fridge to allow trypsin activation. Upon activation, 40 µg of trypsin was added to the sample filled up with 100 mM NH_4HCO_3 to a volume which equates to a 10-fold of total sample volume. Reaction mixture pH was 8-9 and incubated over night at 37 °C on Thermomixer® C (Eppendorf) with shaking.

8.3.7.4 Tryptic digestion of extracellular vesicles

Extracellular vesicles (EVs) were previously isolated from plasma of healthy donors and adenocarcinoma lung cancer patient as described below in 8.4.3. Dried samples were taken out of the freezer and resuspended in 100 mM triethylammonium bicarbonate (TEAB) (Thermo Fisher Scientific, Rockford, Illinois, USA) and 1% sodium deoxycholate (SDC) (Sigma Aldrich, Saint Quentin-Fallavier, France) diluted in HPLC water (Honeywell Riedel-de-Haen, Seelze), boiled at 95 °C for 5 minutes and sonicated with a probe sonicator. Reduction of disulfide bonds was performed using 1 µL of 1 M DTT (Carl Roth) for 30 minutes, and sulfhydryl groups were alkylated using 4 µL of 0.5 M IAA (AppliChem) for 30 minutes in dark. Digestion of EV derived proteins was conducted using Sequencing grade trypsin

(Promega, Schnelldorf, Germany) over night at 37 °C. The day after, 1% of formic acid (Merck, Darmstadt, Germany) in HPLC water (Honeywell Riedel-de-Haen, Seelze) was added for precipitation of SDC and sample was subjected to centrifugation at 16,000 ×g. Supernatant was transferred into a new tube and dried using vacuum centrifuge UNIVAPO 100 H (Uniequip).

8.3.8 Desalting with Sep-Pak C18 1 cc Vac Cartridge

In this work, tryptic peptides were subjected for isoelectric focusing using OFFGEL fractionator (see below 8.3.9). Prior to isoelectric focusing, charged components which would interfere with isoelectric focusing were removed from samples using a hydrophobic, reversed-phase, silica-based bonded phase containing cartridges. The sample was filled up with 5% methanol and 0.2% formic acid to a volume of 1 mL in HPLC water (Honeywell Riedel-de-Haen, Seelze). To allow highly aqueous sample to penetrate the hydrophobic surface of a cartridge, the Sep-Pak® Vac 1cc C18 cartridge (100 mg; Waters, Manchester, United Kingdom) was conditioned with 6 mL of methanol and equilibrated with Sep-Pak equilibration buffer (Table 18). Total volume of 1 mL of a sample was loaded, and upon complete penetration of a sample into the column, 3 mL of Sep-Pak washing buffer was loaded (Table 18). Sample was eluted with 1 mL of 50% ACN and then with 0.5 mL of 75% ACN into new Safe-Lock Tubes 1.5 mL (Eppendorf). Desalted sample was completely dried out using a UNIVAPO 100 H vacuum-centrifuge (Uniequip).

8.3.9 OFFGEL-Isoelectric focusing fractionation

To decrease the sample complexity and increase the detection rate of proteins later by mass spectrometry, tryptic peptides were pre-fractionated using Agilent 3100 OFFGEL fractionator (Agilent Technologies, Santa Clara, USA). Peptides were separated and immobilized according to their isoelectric point using a 24-well tray and non-linear pH gradient (IPG) gel strip (pH 3-10, 24 cm: GE Healthcare, Munich, Germany).

Previously desalted and evaporated samples were dissolved in 3.6 mL of OFFGEL rehydration solution (Table 18), and the sample loading and device preparation was performed according to the instruction of the manufacturer using G98J program. Peptides were focused overnight, with a maximum current per strip of 50 μA until 50 kVh was achieved. Each fraction was transferred into a 1.5 mL conical polypropylene tube (Eppendorf) and dried out using UNIVAPO 100 H vacuum-centrifuge (Uniequip).

8.3.10 LC-MS/MS

Mass spectrometry analysis used in this dissertation was performed using a nano ultra-performance-liquid chromatography (nano-UPLC) system (Dionex ultimate 3,000 RSLCnano, Thermo Fisher Scientific, Bremen, Germany) coupled via electrospray-ionization (ESI) to an Orbitrap-Iontrap-Quadrupole Tribrid mass spectrometer (Orbitrap Fusion, Thermo Fisher Scientific, Waltham, USA) (Figure 62).
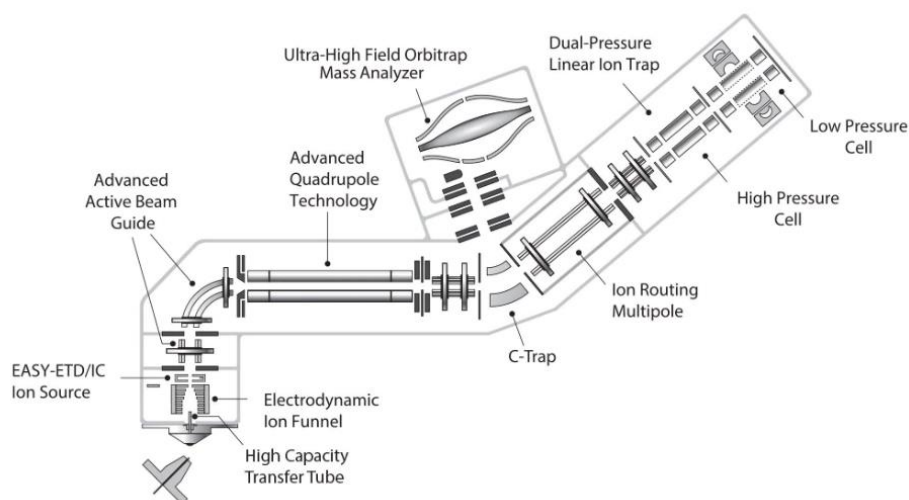

Figure 62. Scheme of an Orbitrap Tribrid Mass Spectrometer. The Orbitrap Tribrid Mass spectrometer combines a quadrupole, C – trap coupled with Orbitrap mass analyzer and Linear Ion Trap analyzer. After the separation of peptide according to their hydrophobicity by LC (not shown), peptides are directly injected into mass spectrometer and ionized using electrospray ionization, after which they enter quadrupole. In quadrupole peptides are being pre-selected following their accumulation in C-trap prior to their detection by an orbitrap. The figure was taken from Thermo Fisher Scientific (289).

Liquid chromatography is a very useful tool for purification and separation of complex samples, such as tryptic peptides from total cell protein extracts. Purification and separation are based on hydrophobic interaction between immobilized silica-based molecules on the column matrix and organic solvents from the mobile phase, such as ACN.

Afterwards, peptide ions are directly generated by nano-electrospray ionization (nano-ESI) in three following steps: (i) dispersion of a charge droplet spray, (ii) evaporation of the solvent and (iii) release of the charged droplets. The analyte is passing through the capillary due to applied high voltage, and positive ion droplets are generated in excess on the capillary tip. Given the charge attraction between positively and negatively charged molecules, the Taylor cone is being formed which results in positively charged molecules coming closer which eventually gives instability to a cone and leads to formation of positively charged droplet. Due to the solvent evaporation, droplet radius decreases and concentration of positively charged molecules within the droplet increases which results once again in droplet

instability. At the radius which meets Rayleigh limit, Coulomb explosion occurs resulting in progeny droplets, which are as well positively charged but smaller. The mechanism repeats until eventually single gas phase ions are not generated.

Gas phase ions enter quadrupole after which they are being stored in the C-trap prior to the detection by orbitrap, resulting in MS1 spectra. Peptides with high enough intensity given from MS1 spectra are being selected for fragmentation using collision-induced dissociation (CID) and N_2 for further sequencing and ion-trap for detection, resulting in MS2 spectra. A variant of CID - higher energy collision dissociation (HCD) – is commonly used in Orbitrap Mass spetrometers (290).

A half microliter of SILAC samples, one microliter of labelling quality control samples, BCA test quality control samples, or two microliters of EV samples, was applied on a trapping column (Acclaim PepMap µ-precolumn, C18, 300 µm × 5 mm, 5 µm, 100 Å, Thermo Fisher Scientific) with 2% of MS – buffer B (Table 18). Afterwards, trapping column was washed for 5 minutes with 2% MS - buffer B. Peptides were transferred onto a separation column (Acclaim PepMap 100, C18, 75 µm × 250 mm, 2 µm, 100 Å, Thermo Fisher Scientific) with a flow rate of 300 nl/min and a gradient from 3% to 28% of MS – buffer B within first 30 minutes, following by a gradient from 28% to 35% of MS – buffer B in the next 5 minutes with a total run time of a 70 minutes (for SILAC related samples) or 125 min (for EV samples) and a column temperature of 45 °C. Peptides were ionized by a fused-silica emitter (I.D. 10 µm, New Objective, Woburn, USA) using nano-electrospray ionization at a capillary voltage of 1,800 V in a positive ion mode. LC-MS/MS analysis was carried out in a data dependent acquisition (DDA) with the top speed mode. The maximum injection time was 120 ms for an automatic gain control target of 2×10^5 and an isolation width of 1.6 *m/z*. A MS scan was performed every second for peptides over a m/z range from 400-1,500 Da, with a resolution of 120,000 FWHM at *m/z* 200. Fragment ion scans were conducted for the most intense ions with an intensity greater than 1×10^4 with a HCD collision energy of 30% and a maximum injection time of 60 ms for an automatic gain control target of 1×10^4 in the ion trap.

8.3.11 Enzyme-linked Immunosorbent Assay (ELISA)

For the analysis of NECT-4 and PDAP1 concentrations in cell culture supernatants, cell line lysates or in plasma samples, and for the analysis of CD9 and CD63 in SEC fractions, quantitative sandwich-ELISAs were performed.

Concentrations of NECT-4 and PDAP1 were measured using commercial kits ELISA Human Nectin-4 Quantikine® (R&D System; # DNEC40) and Human PDAP1 (Pdgfa Associated Protein 1) (Novus Biologicals, Colorado, USA; # NBP2-68090). NECT-4 was measured according to the manufacturer's instruction, and PDAP1 was measured according to the manufacturer's instruction with the following modification: instead of loading 100 µL of sample directly onto the plate well, 50 µL of Reference

Standard and Sample Diluent was loaded on the plate and 50 µL of sample was resuspended in it. Samples analyzed for NECT-4 and PDAP1 concentration using sandwich ELISA included cell culture supernatant, cell culture lysate and plasma samples.

For the detection of EVs in SEC fractions, CD9 and CD63 ELISA were performed as follows. MiliQ H_2O and DPBS (Life Technologies) that were used for EV experiments, including ELISA, were previously filtered through 0.22 µm filters. High Binding 96 Well Strip Plates (Biomat, Trentino, Italy) were coated with 100 µL of anti-Hu CD9 Purified antibody, clone MEM-61 (ExBio, Prague, Check Republic) or anti-Hu CD63 Purified antibody, clone MEM-259 (ExBio) diluted 1:1,000 in ELISA coating buffer (Table 18) and incubated overnight at 4 °C with shaking using Rotamax 120 shaker (Heidolph Instruments GmbH & Co. KG). The day after, wells were washed with 300 µL of 0.05% Tween-20 in DPBS (Life Technologies) (0.05% PBST) for three times after which unspecific binding was blocked with 300 µL of 0.5% BSA (Sigma Aldrich, St. Louis, MO, USA) in DPBS (Life Technologies) (0.05% BSA / DPBS) for two hours at room temperature. Blocking solution was discarded and 100 µL of SEC fraction sample was loaded into well, incubated for 10 minutes at room temperature with shaking and then overnight at 4 °C with shaking as well using Rotamax 120 shaker (Heidolph Instruments GmbH & Co. KG). The following day, wells were washed three times with 0.05% PBST and 100 µL of anti-Hu CD9 Biotin antibody, clone MEM-61 (ExBio) or anti-Hu CD63 Biotin antibody, clone MEM-259 (ExBio) diluted 1:1,000 in 0.05% BSA / DPBS was applied and incubated for two hours at 4 °C with shaking using Rotamax 120 shaker (Heidolph Instruments GmbH & Co. KG). Wells were washed again for three times with 0.05% PBST, and subsequently 100 µL of Pierce™ Streptavidin Poly-HRP (Thermo Fisher Scientific) diluted 1:10,000 in 0.05% BSA / DPBS was applied and incubated for one hour at 4 °C with shaking using Rotamax 120 shaker (Heidolph Instruments GmbH & Co. KG). Unbounded antibodies were washed away with 0.05% PBST in three consecutive steps after which 100 µL of TMB 2-Component Microwell Peroxidase Substrate kit (Sera Care, Massachusetts, USA), where reagent A and reagent B were mixed in ration 1:1, was applied and incubated for 10 minutes at room temperature in the dark. To stop reaction, 100 µL of 5.56% sulfuric acid in miliQ H_2O was added and absorbance was read at 450 nm using microplate reader Infinite NanoQuant M200 Pro (Tecan, Männedorf, Switzerland) and the microplate reader software i-Control™ (Tecan, software version 1.11).

8.4. Isolation of extracellular vesicles from plasma

For all experiments performed on EVs, miliQ H_2O and fresh DPBS were used. Additionally, both liquids were previously filtered through 0.22 µm filters.

To perform LC-MS/MS analysis of EVs derived from plasma samples, the establishment of a protocol for isolation of EVs which provides high purity was needed. First, the isolation of EVs from two

milliliters of plasma samples derived from healthy donors was optimized by testing three different isolation methods. Finally, the optimized protocol was established using 1.5 mL of plasma. This final protocol was then applied to the isolation of EVs from plasma of study cohort including healthy controls and adenocarcinoma lung cancer patients (see below 8.4.3) , which were subsequently subjected to the mass spectrometry analysis.

8.4.1 Optimization of extracellular vesicle isolation from 2 mL of plasma

To gain insight into the protein profile of EVs derived from lung cancer patients compared to healthy controls, the protocol for EV isolation from plasma samples with high purity of EVs needed to be established. During all steps, 1.5 mL Ultra High Recovery Mirocentrifuge Tube Natural (StarLab GmbH) and TipOne® Tip, Natural pipette tips (StarLab GmbH) were used to avoid adhesion of EVs and therefore the loss of sample. All experiments for optimization of a protocol were performed in triplicates as described below.

Fresh plasma derived from healthy donors was isolated as described in 8.2.5. Twelve mL of plasma was subjected to centrifugation at 10,000 $\times g$ for 30 minutes at 4 °C to remove microvesicles. Resulting supernatant was filtered through 0.22 µm filters and assigned afterwards as *pre-cleared* plasma. *Pre-cleared* 12 mL of plasma was divided into three equal parts for comparison of different EV isolation protocols for yield and purity, and subsequently each part was divided into two equal parts to check technical accuracy as well (Figure 63).

First 4 mL of plasma was divided into two times of two mL of *pre-cleared* plasma, placed into Open-Top Thinwall Ultra-Clear Tube (14 mL, 14 × 96 mm, Beckman Coulter GmbH, Krefeld, Germany) ultracentrifuge tubes and applied directly for ultracentrifugation using SW40 Ti Swinging-Bucket Rotor (Beckman Coulter GmbH, Krefeld, Germany) and Optima™ L-100 WP Ultracentrifuge (Beckman Coulter GmbH, Krefeld, Germany) at 100,000 $\times g$ for two hours and 10 minutes at 4 °C. Supernatant was discarded and pellet was resuspended in 100 µL of 0.22 µm filtered DPBS, transferred into 1.5 mL conical polypropylene tube (StarLab) and stored at -80 °C.

Remaining two times 4 mL of fresh and *pre-cleared* plasma were treated separately and applied to SEC using qEV10/70 nm column (IZON, Christchurch, New Zealand) and 70 fractions of 1 mL were collected manually or automatic fraction collector device (AFC) (IZON, Christchurch, New Zealand) was used according to the manufacturer's instruction. If collected manually, resulting fractions were tested for protein concentration and the EV presence using BCA test (see 1.3.2) and CD9 and CD63 sandwich ELISAs (see 8.3.11). SEC fractions positive for CD9 and/or CD63 EV biomarkers and containing low total protein concentration were considered as EVs containing fractions with high purity. These fractions were pooled and in total were of approx. 20 mL. If the AFC device was used, BCA test and ELISA were not performed on resulting fractions, but fractions were pooled according to the

manufacturer instruction, and the further procedure was the same as if fractions were generated manually, as described next. Twenty mL were divided into two equal parts. Ten mL of pooled SEC fractions were subjected to ultracentrifugation at 100,000 ×g for two hours and 10 minutes at 4 °C using Open-Top Thinwall Ultra-Clear Tube (14 mL, 14 × 96 mm, Beckman Coulter GmbH, Krefeld, Germany) ultracentrifuge tubes placed in SW40 Ti Swinging-Bucket Rotor (Beckman Coulter GmbH, Krefeld, Germany) and Optima™ L-100 WP Ultracentrifuge (Beckman Coulter GmbH, Krefeld, Germany). Supernatant was discarded and pellet containing EVs was resuspended in 100 µL of 0.22 µm filtered DPBS, transferred into 1.5 ml conical polypropylene tubes and stored at -80 °C. Other 10 mL were filtered by centrifugal filtration using Amicon® Ultra-15 Centrifugal filter Unit MWCO 100 kDa (Merk Milipore) according to the manufacturer's instruction. Filtered EVs were transferred into 1.5 mL conical polypropylene tubes and stored at -80 °C.

Subsequently, in final samples derived from above-described isolation protocols, QUBIT assay has been used to measure protein concentration and all samples were subjected for NTA analysis. Additionally, *Western blot* analysis for apoliprotein A (APOA), apoliprotein B (APOB) and CD9 was performed to compare isolation methods for purity and yield of EVs.
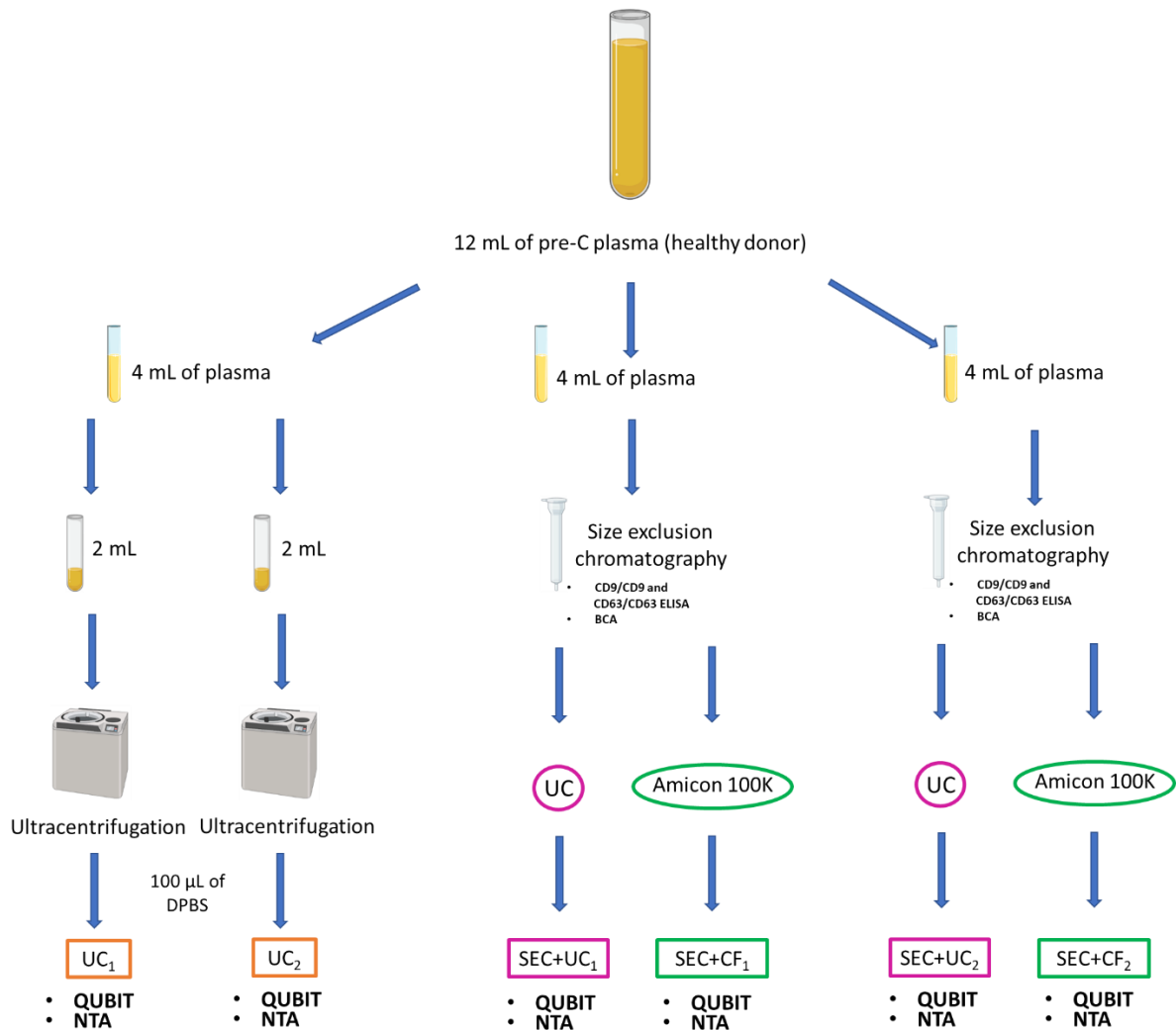
Figure 63. Scheme of compared experiments for the isolation of EVs from 2 mL of plasma. (A) 4 mL of pre-C plasma was divided into two equal parts and subjected to ultracentrifugation. Isolated EVs analyzed by NTA, and protein concentration was measured using Qubit. (B) EVs were isolated from 4 mL of pre-C plasma using SEC. SEC fractions containing EVs were pooled together and divided into two equal parts, from which one part was further purified using ultracentrifugation, while another part was additionally purified using Amicon MWCO 100 kDa. EVs=extracellular vesicles; mL=milliliter; pre-C=pre-cleared; NTA=nanoparticle tracking analysis; SEC=size exclusion chromatography; CF=centrifugal filtration; Amicon 100 kDa=amicon ultra centrifugal filters MWCO 100 kDa. Created with BioRender.com

8.4.2 Optimization of isolation of extracellular vesicles from 1.5 mL of plasma derived from lung cancer patient – purity check

Experiments described in 8.4.1 have shown that SEC+UC protocol is the most suitable for the isolation of EVs taking purity and the yield into account compared to other tested protocols. Before starting to process study cohort, it was needed to test if given protocol provides EVs of purity suitable for downstream mass spectrometry analysis. The samples acquired for the study cohort were aliquoted to 1.5 mL. Therefore, a protocol for the isolation of EVs from 1.5 mL of plasma was established.

From 3 plasma samples derived from lung cancer patients, in a volume of 1.5 mL, which were kindly provided by Dr. Harriet Wikman, EVs were isolated using SEC+UC protocol as described in 8.4.1 with modifications described below in this section and in Figure 64.

Since optimization of the isolation of EVs from plasma samples described in 8.4.1 was performed on healthy controls it could not be predicted that the additional step will be desired when isolating EVs from plasma derived from lung cancer patients. Notably, after thawing the plasma from lung cancer patient and subjecting to centrifugation at 10,000 ×g for 30 minutes at 4 °C, supernatant was still containing floating structures visible to eye. Therefore, after thawing the plasma derived from lung cancer patients, centrifugation step at 1,800 ×g for 10 minutes at 4 °C was repeated to pre-clear the plasma. Further on, SEC+UC described in 8.4.1 with the use of AFC (IZON, Christchurch, New Zealand) was used protocol for EV isolation with following modifications. SEC column used for isolation of EVs was qEV2 column (IZON, Christchurch, New Zealand) instead of qEV10/70 nm column (IZON, Christchurch, New Zealand). Also, only BCA test, but not ELISA, was performed on SEC fractions since conformation of EVs in final samples conducted after ultracentrifugation was achieved by TEM (see 8.6). SEC fractions with low total protein concentration provided by BCA test (fractions before exponential increase) were considered as non-contaminated fractions or with poor contamination, pooled together and subjected to ultracentrifugation as described in 8.4.1. Total protein concentration of a samples was measured using Qubit assay (8.3.2) and resulting samples (n=3) were dried out using UNIVAPO 100 H vacuum-centrifuge (Uniequip) and applied for the analysis by TEM for EVs characterization and sample purity (8.6). Additionally, with this test, technical accuracy of AFC was analyzed to eliminate the performance of the BCA and ELISA on SEC fractions and eventually to loss of sample.
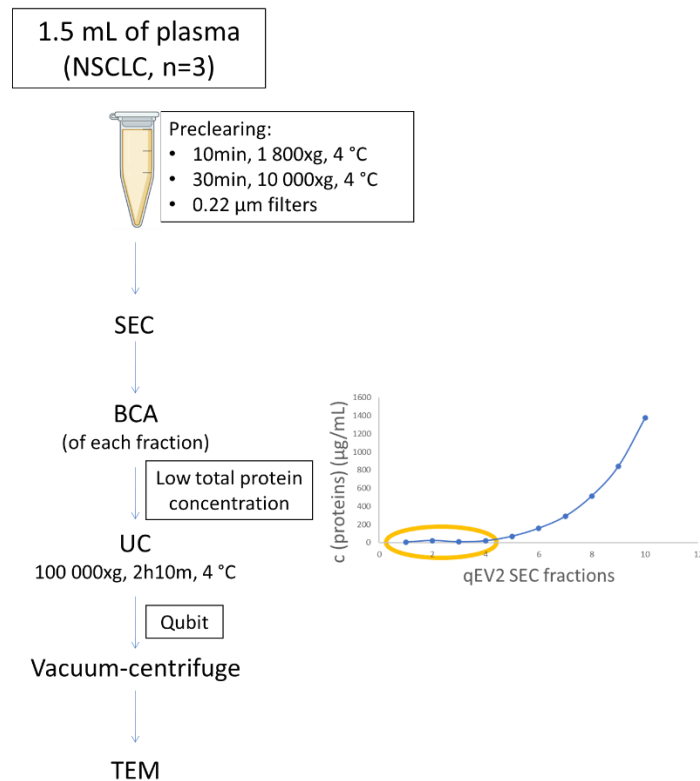
Figure 64. Scheme of experiments used for testing purity of EVs isolated from 1.5 mL of plasma samples derived from NSCLC patients. (A) 1.5 mL of plasma derived from NSCLC patients was pre-cleared using ultracentrifugation and filtration. Sample was fractionated using SEC and the presences of plasma derived contaminating proteins was measured by BCA test. Fractions containing low total protein amount were pooled together, subjected to ultracentrifugation and total protein concentration was measured using Qubit assay. Finally, samples were dried out prior to preparation and analysis by TEM for EV characterization and purity check. EVs=extracellular vesicles, mL=milliliter, NSCLC=non-small cell lung cancer, n=number of biological replicas, min=minutes, g=relative centrifugal force (RCF), °C=degree Celsius, µm=micrometer, SEC=size exclusion chromatography, BCA=bicinchoninic acid assay, UC=ultracentrifugation, TEM=transmission electron microscopy, MS=mass spectrometry. Created with BioRender.com

8.4.3 Isolation of extracellular vesicles from plasma – Study cohort

Plasma samples derived from adenocarcinoma lung cancer patients in a volume of 1.5 mL were kindly provided by Dr. Harriet Wikman and used in this study. Blood was collected from healthy donors (Department of Transfusion Medicine, University Medical Center Hamburg-Eppendorf) in 9.0 mL S-Monovette EDTA containing tubes (Sarstedt) and plasma was isolated as described in 8.2.5, and stored at -80 °C.

During all steps, 1.5 mL Ultra High Recovery Microcentrifuge Tube Natural (StarLab GmbH) were used and TipOne 200 µL Yellow UltraPoint Graduated Tips (StarLab GmbH) were used to avoid adhesion of EVs and therefore the loss of sample. The complete isolation was performed as described in 8.4.2.

To avoid degradation of proteins and EVs, plasma was slowly thawed on ice and subsequently conducted to centrifugation at 1,800 ×g for 10 minutes at 4 °C using Optima™ L-100 WP Ultracentrifuge (Beckman Coulter GmbH, Krefeld, Germany). Supernatant was transferred into a new 1.5 mL Ultra High Recovery Microcentrifuge Tube Natural (StarLab) and additionally centrifuged at 10,000 ×g for 30 minutes at 4 °C to remove microvesicles. Resulting supernatant was filtered through 0.22 µm filters and applied on previously washed qEV2 column (IZON, Christchurch, New Zealand) with 150 mL of 0.22 µm filtered DPBS. Fractions were collected using AFC (IZON, Christchurch, New Zealand) according to the manufacturer´s instruction and EVs containing fractions (SEC fractions 1-4) were pooled together and subjected to ultracentrifugation using Optima™ L-100 WP Ultracentrifuge (Beckman Coulter GmbH, Krefeld, Germany) at 100,000 ×g for two hours and 10 minutes at 4 °C. Subsequently, supernatant was discarded, and the pellet was resuspended in 100 µL of 0.22 µm pore filtered DPBS, dried out using UNIVAPO 100 H vacuum-centrifuge (Uniequip) and stored at -80 °C.

Within 4 days, the proteins were hydrolyzed by trypsin (8.3.7.4) and analyzed by mass spectrometry (8.3.10).

8.5 Nanoparticle tracking assay

During optimization of the EV isolation from plasma, EVs were analyzed and quantified by nanoparticle tracking assay using NanoSight LM14C (NanoSight NTA 2.3 Build 0033; Malvern Instruments Ltd, Malvern, United Kingdom) equipped with a Marlin F-033B IRF camera (Allied Vision Technologies, Stadtroda, Germany) and a 638 nm laser using Nanosight 3.0 software (Malwern Panalytical, Malvern, United Kingdom). For capturing of nanoparticles, camera level was set to 16 and camera screen gain was set to two for all sample analyses. Chamber was flushed first with 3 mL of 0.22 µm filtered miliQ H_2O followed by flushing with 3 mL of 0.22 µm filtered DPBS and then with 500 µL of a sample to be analyzed. Measurement was done for 5 times for 30 seconds and in between each time 100 µL of sample was flushed. For processing the analysis of the measurements, camera screen gain was set to 10 for all sample analyses measurement processing, while detection threshold was set to 5 or 6, depending on the background. Set detection threshold was kept the same within the same measurement (within all samples of the same replica). The chamber was cleaned by flushing with 0.22 µm filtered miliQ H_2O followed by flushing with 10% EtOH diluted in 0.22 µm filtered miliQ H_2O.

8.6 Analysis of extracellular vesicles by transmission electron microscopy

Previously dried out lung cancer samples (see 8.4.3) were resuspended in 100 µL of paraformaldehyde and processed for TEM analysis and subsequently analyzed by Dr. Oliver Kretz (Department of Internal

Medicine III, University Medical Center Hamburg – Eppendorf, Hamburg, Germany). Briefly, single drop of re-suspended EV pellet was placed onto parafilm, and the carbon-coated copper meshed grids (Plano, Wetzlar, Germany) were placed on it for 5 minutes to adsorb EVs on it. Samples were fixed with 1% glutaraldehyde (Sigma-Aldrich), washed 4 times for 30 seconds and negatively contrasted using 1% uranyl acetate. Grids were dried and analyzed using transmission electron microscope CM 100 (Thermo Fisher Scientific) and Olympus ITEM software (Olympus, Münster, Germany).

8.7 Data analysis

8.7.1 Protein identification and quantification of SILAC related LC/MS-MS data

Data analysis of the SILAC related LC-MS/MS raw files was performed using MaxQuant software (version 1.5.8.3). Identification of peptides and proteins was obtained using Andromeda against a reviewed human database (UniProtKB, www.uniprot.org, downloaded in October 2019; 20,430 entries), and a contaminant database (MaxQuant, 245 entries). Analysis was performed with a mass tolerance of 20 ppm on MS level and of 20 ppm or 0.5 Da on MS/MS level. Two MS1 labels were used, $^{13}C_6$ arginine and $^{13}C_6$ lysine, and maximum of two missed cleavages were allowed. As fixed modification, carbamidomethylation of cysteine was set, while oxidation on methionine, acetylation on protein N-terminus and deamidation on asparagine and glutamine were set as variable modifications. Additionally, maximum of 5 modifications per peptide were allowed, and the search was performed with false discovery rate (FDR) of 0.01 on both peptide and protein level. For quantification, only unique peptides which were unmodified and acetylated on N-terminus were considered. Student's *t*-test was used for statistical analysis of SILAC related mass spectrometry data according to the Molecular and Cellular Proteomics (MCP) guidelines as follows: I.) peptides with PEP=n. def were excluded, II.) for PDAP1 and CCN1 all peptides with H/L normalized=n. def were excluded, III.) If peptides with the same amino acid sequence were detected in several OFFGEL fractions of the same biological sample, only the peptide with the highest intensity was considered for further analysis. The represented MS survey scans were obtained from XCalibur Qual Browser, whereas corresponding positive ion mode HCD spectra were obtained from Proteome Discoverer software.

8.7.2 Selection criteria of proteins selected from SILAC proteome data

Proteins considered as lung cancer biomarker candidates from data obtained by the SILAC proteome analysis were determined based on following criteria: I.) Ratio H/L normalized $\geq$ 2 (differential expression in heavy SILAC sample), II.) Selected protein is secreted protein or membrane protein with

an extracellular domain, III.) Selected protein is not expressed on PBMCs. For determination of protein localization and PBMCs expression, UniProtKB and genecards databases were used.

8.7.3 Protein identification and quantification of EVs related LC-MS/MS data

LC-MS/MS data were searched with the Sequest algorithm integrated in the Proteome Discoverer software (v 2.41.15, Thermo Fisher Scientific) against a reviewed Homo sapiens Swissprot database, obtained in April 2020, containing 20,365 entries. Carbamidomethylation was set as fixed modification for cysteine residues and the oxidation of methionine, and pyro-glutamate formation at glutamine residues at the peptide N-terminus, as well as acetylation of the protein N-terminus were allowed as variable modifications. A maximum number of 2 missing tryptic cleavages was set. Peptides between 6 and 144 amino acids where considered. A strict cut-off (FDR <0.01) was set for peptide and protein identification. Protein quantification was carried out, using the Minora Algorithm, implemented in Proteome Discoverer. (https://pubs.acs.org/doi/pdf/10.1021/acs.jproteome.1c00143).

For statistical analysis, protein abundance values were log2 transformed and median normalized across columns to compensate for injection amount differences, and the statistical analysis was performed using Perseus software (Max Plank Institute for Biochemistry, Version 1.5.8.5).

8.7.4 Processing, analysis, and visualization of EV-related LC-MS/MS data

To visualize differences and similarities between samples, hierarchical clustering and Principal component analysis (PCA) was carried out, using the Perseus software (Max Plank Institute for Biochemistry, Version 1.5.8.5). For hierarchical clustering, Pearson correlation was used as a distance metric. Average linkage was applied. Pairwise complete correlation was used to enable the consideration of missing values. For supervised hierarchical clustering, ANOVA testing was applied. Proteins, identified with a p-value<0.05 were further considered.

Principal component analysis does not tolerate missing data points. As a result, data was filtered for complete observations prior to PCA. PCA was carried out allowing a maximum of 20 principal components. Only components, reflecting >5% of the explained variance were further considered.

For the identification of differences between groups, Student's t-testing was applied (p-value <0.05, FDR >1.5). T-test results were visualized as volcano plots, using an in-house script in the R software environment. (Version 4.2.0).

For the identification of significantly regulated signalling pathways enrichment analysis was performed using the BINGO (https://academic.oup.com/bioinformatics/article/21/16/3448/216306) in the

Cytoscape environment (Version 3.9.1) (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC403769/). Gene ontology (GO) biological processes were used as an ontology file. The whole annotation was used as a reference set. For the assessment of overrepresentations, a hypergeometric test was used. Benjamini & Hochberg FDR was used for multiple testing correction (p-value <0.05).

The top 10 enriched pathways for each condition were visualized, using an in-house script in the R software environment. (Version 4.2.0).

8.7.5 Protein quantification of ELISA data

Samples were measured as described in 8.3.11 and corresponding standard curve was obtained for protein concentration measurement. Final protein concentration of a sample was calculated using equation from the linear curve and multiplied with the corresponding sample dilution.

8.7.6 Statistical analysis of NECT-4 and PDAP1 ELISA data obtained on study cohort

Distribution of NECT-4 and PDAP1 concentrations across the study cohort were described by the use of arithmetic mean and standard deviation. Non-parametric Mann-Whitney U-test or the Kruskal-Wallis test were conducted by IBM SPSS Statistics 28.0.0.0 and used for the comparison between study groups. The results were considered as statistically significant if $p<0.05$. Kruskal-Wallis test analysis was followed by pairwise comparison and the p-values were adjusted for multiple testing using Bonferroni correction. Protein concentrations were displayed by boxplots with median, interquartile range, and whiskers, and visualized using GraphPad Prism 8.4.0.

Sensitivity and specificity of NECT-4 and PDAP1 were determined from receiver operating characteristic (ROC) curves. The area under the ROC curve (AUC) was determined by its 95% Wald confidence (CI) interval. Biomarker cut-offs were determined at predefined specificity of around 95%. The AUC, sensitivity and specificity were parameters used for determination of diagnostic accuracy and discrimination between studied groups. GraphPad Prism 8.4.0. was used for ROC curve analysis of NECT-4 and PDAP1 as single biomarkers and visualization of all ROC curves.

The ROC analysis of combined model, comprising NECT-4 and PDAP1 together, and comparison of the performance of combined model and each protein individually was obtained by Chi Square test and performed by Dr. rer. nat. Swaantje Casjens (Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Germany).

9. References

1. Senga SS, Grose RP. Hallmarks of cancer-the new testament. Open Biol. 2021;11(1):200358.
2. WHO. Cancer 2020 [updated 2nd of February 2022. Available from: https://www.who.int/news-room/fact-sheets/detail/cancer.
3. WHO. Estimated number of incident cases and deaths world, both sexes, all ages 2020 [updated 20th of January 2023. Available from: https://gco.iarc.fr/today/online-analysis-multi-bars?v=2020&mode=cancer&mode_population=countries&population=900&populations=900&key=total&sex=0&cancer=39&type=0&statistic=5&prevalence=0&population_group=0&ages_group%5B%5D=0&ages_group%5B%5D=17&nb_items=10&group_cancer=1&include_nmsc=0&include_nmsc_other=1&type_multiple=%257B%2522inc%2522%253Atrue%252C%2522mort%2522%253Atrue%252C%2522prev%2522%253Afalse%257D&orientation=horizontal&type_sort=0&type_nb_items=%257B%2522top%2522%253Atrue%252C%2522bottom%2522%253Afalse%257D.
4. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100(1):57-70.
5. Hayflick L. Mortality and immortality at the cellular level. A review. Biochemistry (Mosc). 1997;62(11):1180-90.
6. Bryan TM, Cech TR. Telomerase and the maintenance of chromosome ends. Curr Opin Cell Biol. 1999;11(3):318-24.
7. Bouck N, Stellmach V, Hsu SC. How tumors become angiogenic. Adv Cancer Res. 1996;69:135-74.
8. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74.
9. Keller L, Pantel K. Unravelling tumour heterogeneity by single-cell profiling of circulating tumour cells. Nat Rev Cancer. 2019;19(10):553-67.
10. Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov. 2022;12(1):31-46.
11. Saghafinia S, Homicsko K, Di Domenico A, Wullschleger S, Perren A, Marinoni I, et al. Cancer Cells Retrace a Stepwise Differentiation Program during Malignant Progression. Cancer Discov. 2021;11(10):2638-57.
12. Landsberg J, Kohlmeyer J, Renn M, Bald T, Rogava M, Cron M, et al. Melanomas resist T-cell therapy through inflammation-induced reversible dedifferentiation. Nature. 2012;490(7420):412-6.
13. Yuan S, Norgard RJ, Stanger BZ. Cellular Plasticity in Cancer. Cancer Discov. 2019;9(7):837-51.
14. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature. 1983;301(5895):89-92.
15. Torres CM, Biran A, Burney MJ, Patel H, Henser-Brownhill T, Cohen AS, et al. The linker histone H1.0 generates epigenetic and functional intratumor heterogeneity. Science. 2016;353(6307).
16. Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, et al. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. Cell. 2017;171(7):1611-24.e24.
17. Zitvogel L, Tesniere A, Kroemer G. Cancer despite immunosurveillance: immunoselection and immunosubversion. Nat Rev Immunol. 2006;6(10):715-27.
18. Nejman D, Livyatan I, Fuks G, Gavert N, Zwang Y, Geller LT, et al. The human tumor microbiome is composed of tumor type-specific intracellular bacteria. Science. 2020;368(6494):973-80.
19. McAllister F, Khan MAW, Helmink B, Wargo JA. The Tumor Microbiome in Pancreatic Cancer: Bacteria and Beyond. Cancer Cell. 2019;36(6):577-9.
20. Gorgoulis V, Adams PD, Alimonti A, Bennett DC, Bischof O, Bishop C, et al. Cellular Senescence: Defining a Path Forward. Cell. 2019;179(4):813-27.
21. Lee S, Schmitt CA. The dynamic nature of senescence in cancer. Nat Cell Biol. 2019;21(1):94-101.
22. Faget DV, Ren Q, Stewart SA. Unmasking senescence: context-dependent effects of SASP in cancer. Nat Rev Cancer. 2019;19(8):439-53.
23. He S, Sharpless NE. Senescence in Health and Disease. Cell. 2017;169(6):1000-11.
24. Kowald A, Passos JF, Kirkwood TBL. On the evolution of cellular senescence. Aging Cell. 2020;19(12):e13270.
25. Wang B, Kohli J, Demaria M. Senescent Cells in Cancer Therapy: Friends or Foes? Trends Cancer. 2020;6(10):838-57.

26.   De Blander H, Morel AP, Senaratne AP, Ouzounova M, Puisieux A. Cellular Plasticity: A Route to Senescence Exit and Tumorigenesis. Cancers (Basel). 2021;13(18).

27.   Huang D, Su S, Cui X, Shen X, Zeng Y, Wu W, et al. Nerve fibers in breast cancer tissues indicate aggressive tumor progression. Medicine (Baltimore). 2014;93(27):e172.

28.   Shao JX, Wang B, Yao YN, Pan ZJ, Shen Q, Zhou JY. Autonomic nervous infiltration positively correlates with pathological risk grading and poor prognosis in patients with lung adenocarcinoma. Thorac Cancer. 2016;7(5):588-98.

29.   Albo D, Akay CL, Marshall CL, Wilks JA, Verstovsek G, Liu H, et al. Neurogenesis in colorectal cancer is a marker of aggressive tumor behavior and poor outcomes. Cancer. 2011;117(21):4834-45.

30.   Magnon C, Hall SJ, Lin J, Xue X, Gerber L, Freedland SJ, Frenette PS. Autonomic nerve development contributes to prostate cancer progression. Science. 2013;341(6142):1236361.

31.   Marusyk A, Polyak K. Tumor heterogeneity: causes and consequences. Biochim Biophys Acta. 2010;1805(1):105-17.

32.   McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. Cell. 2017;168(4):613-28.

33.   Srivastava S, Koay EJ, Borowsky AD, De Marzo AM, Ghosh S, Wagner PD, Kramer BS. Cancer overdiagnosis: a biological challenge and clinical dilemma. Nat Rev Cancer. 2019;19(6):349-58.

34.   Nasim F, Sabath BF, Eapen GA. Lung Cancer. Med Clin North Am. 2019;103(3):463-73.

35.   Pantel K. Blood tests for early detection of lung cancer: challenges and promises. Lancet Respir Med. 2020;8(7):654-6.

36.   Alexander M, Kim SY, Cheng H. Update 2020: Management of Non-Small Cell Lung Cancer. Lung. 2020;198(6):897-907.

37.   Oudkerk M, Liu S, Heuvelmans MA, Walter JE, Field JK. Lung cancer LDCT screening and mortality reduction - evidence, pitfalls and future perspectives. Nat Rev Clin Oncol. 2021;18(3):135-51.

38.   Alberg AJ, Brock MV, Ford JG, Samet JM, Spivack SD. Epidemiology of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest. 2013;143(5 Suppl):e1S-e29S.

39.   Schabath MB, Cote ML. Cancer Progress and Priorities: Lung Cancer. Cancer Epidemiol Biomarkers Prev. 2019;28(10):1563-79.

40.   Heidrich I, Ačkar L, Mossahebi Mohammadi P, Pantel K. Liquid biopsies: Potential and challenges. Int J Cancer. 2021;148(3):528-45.

41.   Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. Nature. 2018;553(7689):446-54.

42.   Rudin CM, Poirier JT, Byers LA, Dive C, Dowlati A, George J, et al. Molecular subtypes of small cell lung cancer: a synthesis of human and mouse model data. Nat Rev Cancer. 2019;19(5):289-97.

43.   Rudin CM, Poirier JT, Byers LA, Dive C, Dowlati A, George J, et al. Author Correction: Molecular subtypes of small cell lung cancer: a synthesis of human and mouse model data. Nat Rev Cancer. 2019;19(7):415.

44.   Farjah F, Monsell SE, Gould MK, Smith-Bindman R, Banegas MP, Heagerty PJ, et al. Association of the Intensity of Diagnostic Evaluation With Outcomes in Incidentally Detected Lung Nodules. JAMA Intern Med. 2021;181(4):480-9.

45.   Zeeshan R, Mutahir Z. Cancer metastasis - tricks of the trade. Bosn J Basic Med Sci. 2017;17(3):172-82.

46.   Graham TA, Shibata D. Navigating the path to distant metastasis. Nat Genet. 2020;52(7):642-3.

47.   Majidpoor J, Mortezaee K. Steps in metastasis: an updated review. Med Oncol. 2021;38(1):3.

48.   Kim YH, Choi YW, Lee J, Soh EY, Kim JH, Park TJ. Senescent tumor cells lead the collective invasion in thyroid cancer. Nat Commun. 2017;8:15208.

49.   Bednarz-Knoll N, Alix-Panabières C, Pantel K. Plasticity of disseminating cancer cells in patients with epithelial malignancies. Cancer Metastasis Rev. 2012;31(3-4):673-87.

50.   Glentis A, Oertle P, Mariani P, Chikina A, El Marjou F, Attieh Y, et al. Cancer-associated fibroblasts induce metalloprotease-independent cancer cell invasion of the basement membrane. Nat Commun. 2017;8(1):924.
51.   Coussens LM, Raymond WW, Bergers G, Laig-Webster M, Behrendtsen O, Werb Z, et al. Inflammatory mast cells up-regulate angiogenesis during squamous epithelial carcinogenesis. Genes Dev. 1999;13(11):1382-97.
52.   Yan HH, Pickup M, Pang Y, Gorska AE, Li Z, Chytil A, et al. Gr-1+CD11b+ myeloid cells tip the balance of immune protection to tumor promotion in the premetastatic lung. Cancer Res. 2010;70(15):6139-49.
53.   Staller P, Sulitkova J, Lisztwan J, Moch H, Oakeley EJ, Krek W. Chemokine receptor CXCR4 downregulated by von Hippel-Lindau tumour suppressor pVHL. Nature. 2003;425(6955):307-11.
54.   Meng S, Tripathy D, Frenkel EP, Shete S, Naftalis EZ, Huth JF, et al. Circulating tumor cells in patients with breast cancer dormancy. Clin Cancer Res. 2004;10(24):8152-62.
55.   Lo HC, Xu Z, Kim IS, Pingel B, Aguirre S, Kodali S, et al. Resistance to natural killer cell immunosurveillance confers a selective advantage to polyclonal metastasis. Nat Cancer. 2020;1(7):709-22.
56.   Follain G, Herrmann D, Harlepp S, Hyenne V, Osmani N, Warren SC, et al. Fluids and their mechanics in tumour transit: shaping metastasis. Nat Rev Cancer. 2020;20(2):107-24.
57.   Gaertner F, Massberg S. Patrolling the vascular borders: platelets in immunity to infection and cancer. Nat Rev Immunol. 2019;19(12):747-60.
58.   Fukumura D, Kloepper J, Amoozgar Z, Duda DG, Jain RK. Enhancing cancer immunotherapy using antiangiogenics: opportunities and challenges. Nat Rev Clin Oncol. 2018;15(5):325-40.
59.   Bersini S, Miermont A, Pavesi A, Kamm RD, Thiery JP, Moretti M, Adriani G. A combined microfluidic-transcriptomic approach to characterize the extravasation potential of cancer cells. Oncotarget. 2018;9(90):36110-25.
60.   Leong HS, Robertson AE, Stoletov K, Leith SJ, Chin CA, Chien AE, et al. Invadopodia are required for cancer cell extravasation and are a therapeutic target for metastasis. Cell Rep. 2014;8(5):1558-70.
61.   Al-Mehdi AB, Tozawa K, Fisher AB, Shientag L, Lee A, Muschel RJ. Intravascular origin of metastasis from the proliferation of endothelium-attached tumor cells: a new model for metastasis. Nat Med. 2000;6(1):100-2.
62.   Gkretsi V, Stylianopoulos T. Cell Adhesion and Matrix Stiffness: Coordinating Cancer Cell Invasion and Metastasis. Front Oncol. 2018;8:145.
63.   Bragado P, Estrada Y, Parikh F, Krause S, Capobianco C, Farina HG, et al. TGF-β2 dictates disseminated tumour cell fate in target organs through TGF-β-RIII and p38α/β signalling. Nat Cell Biol. 2013;15(11):1351-61.
64.   Aguirre-Ghiso JA, Bragado P, Sosa MS. Metastasis awakening: targeting dormant cancer. Nat Med. 2013;19(3):276-7.
65.   Lu X, Mu E, Wei Y, Riethdorf S, Yang Q, Yuan M, et al. VCAM-1 promotes osteolytic expansion of indolent bone micrometastasis of breast cancer by engaging α4β1-positive osteoclast progenitors. Cancer Cell. 2011;20(6):701-14.
66.   Pantel K, Alix-Panabières C. Circulating tumour cells in cancer patients: challenges and perspectives. Trends Mol Med. 2010;16(9):398-406.
67.   Alix-Panabières C, Pantel K. Challenges in circulating tumour cell research. Nat Rev Cancer. 2014;14(9):623-31.
68.   Alix-Panabières C, Pantel K. Liquid biopsy: from discovery to clinical implementation. Mol Oncol. 2021;15(6):1617-21.
69.   Stadler JC, Belloum Y, Deitert B, Sementsov M, Heidrich I, Gebhardt C, et al. Current and Future Clinical Applications of ctDNA in Immuno-Oncology. Cancer Res. 2022;82(3):349-58.
70.   Lianidou E, Pantel K. Liquid biopsies. Genes Chromosomes Cancer. 2019;58(4):219-32.
71.   Alix-Panabières C, Pantel K. Liquid Biopsy: From Discovery to Clinical Application. Cancer Discov. 2021;11(4):858-73.
72.   Sholl LM, Aisner DL, Varella-Garcia M, Berry LD, Dias-Santagata D, Wistuba, II, et al. Multi-institutional Oncogenic Driver Mutation Analysis in Lung Adenocarcinoma: The Lung Cancer Mutation Consortium Experience. J Thorac Oncol. 2015;10(5):768-77.

73.   Pantel K, Alix-Panabières C. Liquid biopsy and minimal residual disease - latest advances and implications for cure. Nat Rev Clin Oncol. 2019;16(7):409-24.

74.   Lv X, Wu S, Xu X, Wu L, Chang N, Zhang J, Ti X. The combination of folate receptor-positive circulating tumor cells and serum tumor markers suggests a histological diagnosis of lung cancer. J Thorac Dis. 2022;14(1):158-68.

75.   Ačkar L, Casjens S, Andreas A, Raiko I, Brüning T, Geffken M, et al. Blood-based detection of lung cancer using cysteine-rich angiogenic inducer 61 (CYR61) as a circulating protein biomarker: a pilot study. Mol Oncol. 2021;15(11):2877-90.

76.   Ilie M, Hofman V, Long-Mira E, Selva E, Vignaud JM, Padovani B, et al. "Sentinel" circulating tumor cells allow early diagnosis of lung cancer in patients with chronic obstructive pulmonary disease. PLoS One. 2014;9(10):e111597.

77.   Vasseur A, Kiavue N, Bidard FC, Pierga JY, Cabel L. Clinical utility of circulating tumor cells: an update. Mol Oncol. 2021;15(6):1647-66.

78.   Bidard FC, Jacot W, Kiavue N, Dureau S, Kadi A, Brain E, et al. Efficacy of Circulating Tumor Cell Count-Driven vs Clinician-Driven First-line Therapy Choice in Hormone Receptor-Positive, ERBB2-Negative Metastatic Breast Cancer: The STIC CTC Randomized Clinical Trial. JAMA Oncol. 2021;7(1):34-41.

79.   Keller L, Belloum Y, Wikman H, Pantel K. Clinical relevance of blood-based ctDNA analysis: mutation detection and beyond. Br J Cancer. 2021;124(2):345-58.

80.   Riethdorf S, O'Flaherty L, Hille C, Pantel K. Clinical applications of the CellSearch platform in cancer patients. Adv Drug Deliv Rev. 2018;125:102-21.

81.   Bauml JM, Li BT, Velcheti V, Govindan R, Curioni-Fontecedro A, Dooms C, et al. Clinical validation of Guardant360 CDx as a blood-based companion diagnostic for sotorasib. Lung Cancer. 2022;166:270-8.

82.   Woodhouse R, Li M, Hughes J, Delfosse D, Skoletsky J, Ma P, et al. Clinical and analytical validation of FoundationOne Liquid CDx, a novel 324-Gene cfDNA-based comprehensive genomic profiling assay for cancers of solid tumor origin. PLoS One. 2020;15(9):e0237802.

83.   Kwapisz D. The first liquid biopsy test approved. Is it a new era of mutation testing for non-small cell lung cancer? Ann Transl Med. 2017;5(3):46.

84.   Johnsen KB, Gudbergsson JM, Andresen TL, Simonsen JB. What is the blood concentration of extracellular vesicles? Implications for the use of extracellular vesicles as blood-borne biomarkers of cancer. Biochim Biophys Acta Rev Cancer. 2019;1871(1):109-16.

85.   Kalluri R, LeBleu VS. The biology, function, and biomedical applications of exosomes. Science. 2020;367(6478).

86.   Crescitelli R, Lässer C, Jang SC, Cvjetkovic A, Malmhäll C, Karimi N, et al. Subpopulations of extracellular vesicles from human metastatic melanoma tissue identified by quantitative proteomics after optimized isolation. J Extracell Vesicles. 2020;9(1):1722433.

87.   Gudbergsson JM, Johnsen KB, Skov MN, Duroux M. Systematic review of factors influencing extracellular vesicle yield from cell cultures. Cytotechnology. 2016;68(4):579-92.

88.   Coumans FAW, Brisson AR, Buzas EI, Dignat-George F, Drees EEE, El-Andaloussi S, et al. Methodological Guidelines to Study Extracellular Vesicles. Circ Res. 2017;120(10):1632-48.

89.   Kim J, Lee H, Park K, Shin S. Rapid and Efficient Isolation of Exosomes by Clustering and Scattering. J Clin Med. 2020;9(3).

90.   Karttunen J, Heiskanen M, Navarro-Ferrandis V, Das Gupta S, Lipponen A, Puhakka N, et al. Precipitation-based extracellular vesicle isolation from rat plasma co-precipitate vesicle-free microRNAs. J Extracell Vesicles. 2019;8(1):1555410.

91.   Gámez-Valero A, Monguió-Tortajada M, Carreras-Planella L, Franquesa M, Beyer K, Borràs FE. Size-Exclusion Chromatography-based isolation minimally alters Extracellular Vesicles' characteristics compared to precipitating agents. Sci Rep. 2016;6:33641.

92.   Mol EA, Goumans MJ, Doevendans PA, Sluijter JPG, Vader P. Higher functionality of extracellular vesicles isolated using size-exclusion chromatography compared to ultracentrifugation. Nanomedicine. 2017;13(6):2061-5.

93.   Multia E, Tear CJY, Palviainen M, Siljander P, Riekkola ML. Fast isolation of highly specific population of platelet-derived extracellular vesicles from blood plasma by affinity monolithic column, immobilized with anti-human CD61 antibody. Anal Chim Acta. 2019;1091:160-8.

94.	Sharma P, Ludwig S, Muller L, Hong CS, Kirkwood JM, Ferrone S, Whiteside TL. Immunoaffinity-based isolation of melanoma cell-derived exosomes from plasma of patients with melanoma. J Extracell Vesicles. 2018;7(1):1435138.

95.	Kang YT, Purcell E, Palacios-Rolston C, Lo TW, Ramnath N, Jolly S, Nagrath S. Isolation and Profiling of Circulating Tumor-Associated Exosomes Using Extracellular Vesicular Lipid-Protein Binding Affinity Based Microfluidic Device. Small. 2019;15(47):e1903600.

96.	McArdle AJ, Menikou S. What is proteomics? Arch Dis Child Educ Pract Ed. 2021;106(3):178-81.

97.	Ma JYW, Sze YH, Bian JF, Lam TC. Critical role of mass spectrometry proteomics in tear biomarker discovery for multifactorial ocular diseases (Review). Int J Mol Med. 2021;47(5).

98.	Woods AG, Sokolowska I, Ngounou Wetie AG, Channaveerappa D, Dupree EJ, Jayathirtha M, et al. Mass Spectrometry for Proteomics-Based Investigation. Adv Exp Med Biol. 2019;1140:1-26.

99.	Xie F, Liu T, Qian WJ, Petyuk VA, Smith RD. Liquid chromatography-mass spectrometry-based quantitative proteomics. J Biol Chem. 2011;286(29):25443-9.

100.	Geiger T, Wehner A, Schaab C, Cox J, Mann M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. Mol Cell Proteomics. 2012;11(3):M111.014050.

101.	Brönstrup M. Absolute quantification strategies in proteomics based on mass spectrometry. Expert Rev Proteomics. 2004;1(4):503-12.

102.	Pauwels J, Gevaert K. Mass spectrometry-based clinical proteomics - a revival. Expert Rev Proteomics. 2021;18(6):411-4.

103.	Ding Z, Wang N, Ji N, Chen ZS. Proteomics technologies for cancer liquid biopsies. Mol Cancer. 2022;21(1):53.

104.	Lorentzian A, Uzozie A, Lange PF. Origins and clinical relevance of proteoforms in pediatric malignancies. Expert Rev Proteomics. 2019;16(3):185-200.

105.	Schaffer LV, Millikin RJ, Miller RM, Anderson LC, Fellers RT, Ge Y, et al. Identification and Quantification of Proteoforms by Mass Spectrometry. Proteomics. 2019;19(10):e1800361.

106.	Reymond N, Fabre S, Lecocq E, Adelaïde J, Dubreuil P, Lopez M. Nectin4/PRR4, a new afadin-associated member of the nectin family that trans-interacts with nectin1/PRR1 through V domain interaction. J Biol Chem. 2001;276(46):43205-15.

107.	Noyce RS, Richardson CD. Nectin 4 is the epithelial cell receptor for measles virus. Trends Microbiol. 2012;20(9):429-39.

108.	Delpeut S, Noyce RS, Richardson CD. The tumor-associated marker, PVRL4 (nectin-4), is the epithelial receptor for morbilliviruses. Viruses. 2014;6(6):2268-86.

109.	UniProt. NECT4_HUMAN 2022 [updated 22nd of June 2022. Available from: https://www.uniprot.org/uniprotkb/Q96NY8/entry#structure.

110.	Noyce RS, Bondre DG, Ha MN, Lin LT, Sisson G, Tsao MS, Richardson CD. Tumor cell marker PVRL4 (nectin 4) is an epithelial cell receptor for measles virus. PLoS Pathog. 2011;7(8):e1002240.

111.	Buchanan PC, Boylan KLM, Walcheck B, Heinze R, Geller MA, Argenta PA, Skubitz APN. Ectodomain shedding of the cell adhesion molecule Nectin-4 in ovarian cancer is mediated by ADAM10 and ADAM17. J Biol Chem. 2017;292(15):6339-51.

112.	Fabre-Lafay S, Garrido-Urbani S, Reymond N, Gonçalves A, Dubreuil P, Lopez M. Nectin-4, a new serological breast cancer marker, is a substrate for tumor necrosis factor-alpha-converting enzyme (TACE)/ADAM-17. J Biol Chem. 2005;280(20):19543-50.

113.	Takano A, Ishikawa N, Nishino R, Masuda K, Yasui W, Inai K, et al. Identification of nectin-4 oncoprotein as a diagnostic and therapeutic target for lung cancer. Cancer Res. 2009;69(16):6694-703.

114.	Challita-Eid PM, Satpayev D, Yang P, An Z, Morrison K, Shostak Y, et al. Enfortumab Vedotin Antibody-Drug Conjugate Targeting Nectin-4 Is a Highly Potent Therapeutic Agent in Multiple Preclinical Cancer Models. Cancer Res. 2016;76(10):3003-13.

115.	Fischer WH, Schubert D. Characterization of a novel platelet-derived growth factor-associated protein. J Neurochem. 1996;66(5):2213-6.

116.	Shen L, Huang KP, Chen HC, Huang FL. Molecular cloning and characterization of a novel casein kinase II substrate, HASPP28, from rat brain. Arch Biochem Biophys. 1996;327(1):131-41.

117.	UniProt. HAP28_HUMAN 2022 [updated 23rd of June 2022. Available from: https://www.uniprot.org/uniprotkb/Q13442/entry#sequences.

118. Chirichella M, Bianchi N, Džafo E, Foli E, Gualdrini F, Kenyon A, et al. RFX transcription factors control a miR-150/PDAP1 axis that restrains the proliferation of human T cells. PLoS Biol. 2022;20(2):e3001538.

119. Olsen JB, Cao XJ, Han B, Chen LH, Horvath A, Richardson TI, et al. Quantitative Profiling of the Activity of Protein Lysine Methyltransferase SMYD2 Using SILAC-Based Proteomics. Mol Cell Proteomics. 2016;15(3):892-905.

120. Delgado-Benito V, Berruezo-Llacuna M, Altwasser R, Winkler W, Sundaravinayagam D, Balasubramanian S, et al. PDGFA-associated protein 1 protects mature B lymphocytes from stress-induced cell death and promotes antibody gene diversification. J Exp Med. 2020;217(10).

121. Sharma VK, Singh A, Srivastava SK, Kumar V, Gardi NL, Nalwa A, et al. Increased expression of platelet-derived growth factor associated protein-1 is associated with PDGF-B mediated glioma progression. Int J Biochem Cell Biol. 2016;78:194-205.

122. Choi SY, Jang JH, Kim KR. Analysis of differentially expressed genes in human rectal carcinoma using suppression subtractive hybridization. Clin Exp Med. 2011;11(4):219-26.

123. Marimuthu A, Subbannayya Y, Sahasrabuddhe NA, Balakrishnan L, Syed N, Sekhar NR, et al. SILAC-based quantitative proteomic analysis of gastric cancer secretome. Proteomics Clin Appl. 2013;7(5-6):355-66.

124. Bork P. The modular architecture of a new family of growth regulators related to connective tissue growth factor. FEBS Lett. 1993;327(2):125-30.

125. Perbal B. NOV (nephroblastoma overexpressed) and the CCN family of genes: structural and functional issues. Mol Pathol. 2001;54(2):57-79.

126. Lau LF, Lam SC. The CCN family of angiogenic regulators: the integrin connection. Exp Cell Res. 1999;248(1):44-57.

127. Kireeva ML, Lam SC, Lau LF. Adhesion of human umbilical vein endothelial cells to the immediate-early gene product Cyr61 is mediated through integrin alphavbeta3. J Biol Chem. 1998;273(5):3090-6.

128. Jedsadayanmata A, Chen CC, Kireeva ML, Lau LF, Lam SC. Activation-dependent adhesion of human platelets to Cyr61 and Fisp12/mouse connective tissue growth factor is mediated through integrin alpha(IIb)beta(3). J Biol Chem. 1999;274(34):24321-7.

129. Chen PP, Li WJ, Wang Y, Zhao S, Li DY, Feng LY, et al. Expression of Cyr61, CTGF, and WISP-1 correlates with clinical features of lung cancer. PLoS One. 2007;2(6):e534.

130. Haque I, Mehta S, Majumder M, Dhar K, De A, McGregor D, et al. Cyr61/CCN1 signaling is critical for epithelial-mesenchymal transition and stemness and promotes pancreatic carcinogenesis. Mol Cancer. 2011;10:8.

131. Huang X, Xiang L, Li Y, Zhao Y, Zhu H, Xiao Y, et al. Snail/FOXK1/Cyr61 Signaling Axis Regulates the Epithelial-Mesenchymal Transition and Metastasis in Colorectal Cancer. Cell Physiol Biochem. 2018;47(2):590-603.

132. Cardoso J, Mesquita M, Dias Pereira A, Bettencourt-Dias M, Chaves P, Pereira-Leal JB. CYR61 and TAZ Upregulation and Focal Epithelial to Mesenchymal Transition May Be Early Predictors of Barrett's Esophagus Malignant Progression. PLoS One. 2016;11(9):e0161967.

133. Armstrong RA. When to use the Bonferroni correction. Ophthalmic Physiol Opt. 2014;34(5):502-8.

134. Nooreldeen R, Bach H. Current and Future Development in Lung Cancer Diagnosis. Int J Mol Sci. 2021;22(16).

135. Ye J, Liu H, Xu ZL, Zheng L, Liu RY. Identification of a multidimensional transcriptome prognostic signature for lung adenocarcinoma. J Clin Lab Anal. 2019;33(9):e22990.

136. Hofbauer LC, Bozec A, Rauner M, Jakob F, Perner S, Pantel K. Novel approaches to target the microenvironment of bone metastasis. Nat Rev Clin Oncol. 2021;18(8):488-505.

137. Wan L, Pantel K, Kang Y. Tumor metastasis: moving new biological insights into the clinic. Nat Med. 2013;19(11):1450-64.

138. Alix-Panabières C, Pantel K. Liquid Biopsy to Catch the Epigenetic Changes in Endometrial Cancer. Clin Chem. 2022;68(6):745-7.

139. Leong SP, Naxerova K, Keller L, Pantel K, Witte M. Molecular mechanisms of cancer metastasis via the lymphatic versus the blood vessels. Clin Exp Metastasis. 2022;39(1):159-79.

140. Pereira-Veiga T, Schneegans S, Pantel K, Wikman H. Circulating tumor cell-blood cell crosstalk: Biology and clinical relevance. Cell Rep. 2022;40(9):111298.

141.   Alix-Panabières C, Pantel K. Clinical Applications of Circulating Tumor Cells and Circulating Tumor DNA as Liquid Biopsy. Cancer Discov. 2016;6(5):479-91.

142.   Pantel K, Alix-Panabières C. Liquid biopsy: Potential and challenges. Mol Oncol. 2016;10(3):371-3.

143.   Pantel K. Liquid Biopsy: Blood-Based Analyses of ctDNA and CTCs. Clin Chem. 2021;67(11):1437-9.

144.   Sozzi G, Boeri M. Potential biomarkers for lung cancer screening. Transl Lung Cancer Res. 2014;3(3):139-48.

145.   Deng H, Shi H, Chen L, Zhou Y, Jiang J. Over-expression of Nectin-4 promotes progression of esophageal cancer and correlates with poor prognosis of the patients. Cancer Cell Int. 2019;19:106.

146.   Fabre-Lafay S, Monville F, Garrido-Urbani S, Berruyer-Pouyet C, Ginestier C, Reymond N, et al. Nectin-4 is a new histological and serological tumor associated marker for breast cancer. BMC Cancer. 2007;7:73.

147.   M MR, Cabaud O, Josselin E, Finetti P, Castellano R, Farina A, et al. Nectin-4: a new prognostic biomarker for efficient therapeutic targeting of primary and metastatic triple-negative breast cancer. Ann Oncol. 2017;28(4):769-76.

148.   Zhang Y, Zhang J, Shen Q, Yin W, Huang H, Liu Y, Ni Q. High expression of Nectin-4 is associated with unfavorable prognosis in gastric cancer. Oncol Lett. 2018;15(6):8789-95.

149.   Ma J, Sheng Z, Lv Y, Liu W, Yao Q, Pan T, et al. Expression and clinical significance of Nectin-4 in hepatocellular carcinoma. Onco Targets Ther. 2016;9:183-90.

150.   Derycke MS, Pambuccian SE, Gilks CB, Kalloger SE, Ghidouche A, Lopez M, et al. Nectin 4 overexpression in ovarian cancer tissues and serum: potential role as a serum biomarker. Am J Clin Pathol. 2010;134(5):835-45.

151.   Nishiwada S, Sho M, Yasuda S, Shimada K, Yamato I, Akahori T, et al. Nectin-4 expression contributes to tumor proliferation, angiogenesis and patient prognosis in human pancreatic cancer. J Exp Clin Cancer Res. 2015;34(1):30.

152.   Tomiyama E, Fujita K, Rodriguez Pena MDC, Taheri D, Banno E, Kato T, et al. Expression of Nectin-4 and PD-L1 in Upper Tract Urothelial Carcinoma. Int J Mol Sci. 2020;21(15).

153.   Nabih ES, Abdel Motaleb FI, Salama FA. The diagnostic efficacy of nectin 4 expression in ovarian cancer patients. Biomarkers. 2014;19(6):498-504.

154.   R&D. Human Nectin-4 Quantikine ELISA Kit Summary2022.

155.   Erturk K, Karaman S, Dagoglu N, Serilmez M, Duranyildiz D, Tas F. Serum nectin-2 and nectin-4 are diagnostic in lung cancer: which is superior? Wien Klin Wochenschr. 2019;131(17-18):419-26.

156.   Stieber P, Hasholzner U, Bodenmüller H, Nagel D, Sunder-Plassmann L, Dienemann H, et al. CYFRA 21-1. A new marker in lung cancer. Cancer. 1993;72(3):707-13.

157.   Wieskopf B, Demangeat C, Purohit A, Stenger R, Gries P, Kreisman H, Quoix E. Cyfra 21-1 as a biologic marker of non-small cell lung cancer. Evaluation of sensitivity, specificity, and prognostic role. Chest. 1995;108(1):163-9.

158.   Fujiwara K, Fujimoto N, Tabata M, Nishii K, Matsuo K, Hotta K, et al. Identification of epigenetic aberrant promoter methylation in serum DNA is useful for early detection of lung cancer. Clin Cancer Res. 2005;11(3):1219-25.

159.   Zhang R, Shao F, Wu X, Ying K. Value of quantitative analysis of circulating cell free DNA as a screening tool for lung cancer: a meta-analysis. Lung Cancer. 2010;69(2):225-31.

160.   Ma J, Lin Y, Zhan M, Mann DL, Stass SA, Jiang F. Differential miRNA expressions in peripheral blood mononuclear cells for diagnosis of lung cancer. Lab Invest. 2015;95(10):1197-206.

161.   Cazzoli R, Buttitta F, Di Nicola M, Malatesta S, Marchetti A, Rom WN, Pass HI. microRNAs derived from circulating exosomes as noninvasive biomarkers for screening and diagnosing lung cancer. J Thorac Oncol. 2013;8(9):1156-62.

162.   Cui HY, Wei W, Qian MR, Tian RF, Fu X, Li HW, et al. PDGFA-associated protein 1 is a novel target of c-Myc and contributes to colorectal cancer initiation and progression. Cancer Commun (Lond). 2022;42(8):750-67.

163.   Li C, Sun YD, Yu GY, Cui JR, Lou Z, Zhang H, et al. Integrated Omics of Metastatic Colorectal Cancer. Cancer Cell. 2020;38(5):734-47.e9.

164.   Cho WC. Potentially useful biomarkers for the diagnosis, treatment and prognosis of lung cancer. Biomed Pharmacother. 2007;61(9):515-9.

165. Fujiyuki T, Yoneda M, Amagai Y, Obayashi K, Ikeda F, Shoji K, et al. A measles virus selectively blind to signaling lymphocytic activation molecule shows anti-tumor activity against lung cancer cells. Oncotarget. 2015;6(28):24895-903.

166. Mendt M, Kamerkar S, Sugimoto H, McAndrews KM, Wu CC, Gagea M, et al. Generation and testing of clinical-grade exosomes for pancreatic cancer. JCI Insight. 2018;3(8).

167. Paulsson J, Ehnman M, Östman A. PDGF receptors in tumor biology: prognostic and predictive potential. Future Oncol. 2014;10(9):1695-708.

168. Gao Y, Moten A, Lin HK. Akt: a new activation mechanism. Cell Res. 2014;24(7):785-6.

169. Vivanco I, Sawyers CL. The phosphatidylinositol 3-Kinase AKT pathway in human cancer. Nat Rev Cancer. 2002;2(7):489-501.

170. Gardiner C, Di Vizio D, Sahoo S, Théry C, Witwer KW, Wauben M, Hill AF. Techniques used for the isolation and characterization of extracellular vesicles: results of a worldwide survey. J Extracell Vesicles. 2016;5:32945.

171. Woo HK, Cho YK, Lee CY, Lee H, Castro CM, Lee H. Characterization and modulation of surface charges to enhance extracellular vesicle isolation in plasma. Theranostics. 2022;12(5):1988-98.

172. Brennan K, Martin K, FitzGerald SP, O'Sullivan J, Wu Y, Blanco A, et al. A comparison of methods for the isolation and separation of extracellular vesicles from protein and lipid particles in human serum. Sci Rep. 2020;10(1):1039.

173. Lobb RJ, Becker M, Wen SW, Wong CS, Wiegmans AP, Leimgruber A, Möller A. Optimized exosome isolation protocol for cell culture supernatant and human plasma. J Extracell Vesicles. 2015;4:27031.

174. Hong P, Koza S, Bouvier ES. Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates. J Liq Chromatogr Relat Technol. 2012;35(20):2923-50.

175. von Zychlinski A, Williams M, McCormick S, Kleffmann T. Absolute quantification of apolipoproteins and associated proteins on human plasma lipoproteins. J Proteomics. 2014;106:181-90.

176. UniProt. APOB 2022 [updated 15th of October 2022. Available from: https://www.uniprot.org/uniprotkb/P04114/entry#structure

177. Pedersen KO. The development of Svedberg's ultracentrifuge. Biophys Chem. 1976;5(1-2):3-18.

178. Innerarity TL, Young SG, Poksay KS, Mahley RW, Smith RS, Milne RW, et al. Structural relationship of human apolipoprotein B48 to apolipoprotein B100. J Clin Invest. 1987;80(6):1794-8.

179. Chen SH, Habib G, Yang CY, Gu ZW, Lee BR, Weng SA, et al. Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon. Science. 1987;238(4825):363-6.

180. Hospattankar AV, Higuchi K, Law SW, Meglin N, Brewer HB, Jr. Identification of a novel in-frame translational stop codon in human intestine apoB mRNA. Biochem Biophys Res Commun. 1987;148(1):279-85.

181. Powell LM, Wallis SC, Pease RJ, Edwards YH, Knott TJ, Scott J. A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. Cell. 1987;50(6):831-40.

182. Davidson NO, Shelness GS. APOLIPOPROTEIN B: mRNA editing, lipoprotein assembly, and presecretory degradation. Annu Rev Nutr. 2000;20:169-93.

183. Behbodikhah J, Ahmed S, Elyasi A, Kasselman LJ, De Leon J, Glass AD, Reiss AB. Apolipoprotein B and Cardiovascular Disease: Biomarker and Potential Therapeutic Target. Metabolites. 2021;11(10).

184. Costa-Silva B, Aiello NM, Ocean AJ, Singh S, Zhang H, Thakur BK, et al. Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. Nat Cell Biol. 2015;17(6):816-26.

185. Demory Beckler M, Higginbotham JN, Franklin JL, Ham AJ, Halvey PJ, Imasuen IE, et al. Proteomic analysis of exosomes from mutant KRAS colon cancer cells identifies intercellular transfer of mutant KRAS. Mol Cell Proteomics. 2013;12(2):343-55.

186. Hoshino A, Costa-Silva B, Shen TL, Rodrigues G, Hashimoto A, Tesic Mark M, et al. Tumour exosome integrins determine organotropic metastasis. Nature. 2015;527(7578):329-35.

187. Greening DW, Xu R, Gopal SK, Rai A, Simpson RJ. Proteomic insights into extracellular vesicle biology - defining exosomes and shed microvesicles. Expert Rev Proteomics. 2017;14(1):69-95.

188. Abramowicz A, Marczak L, Wojakowska A, Zapotoczny S, Whiteside TL, Widlak P, Pietrowska M. Harmonization of exosome isolation from culture supernatants for optimized proteomics analysis. PLoS One. 2018;13(10):e0205496.

189. Vykoukal J, Sun N, Aguilar-Bonavides C, Katayama H, Tanaka I, Fahrmann JF, et al. Plasma-derived extracellular vesicle proteins as a source of biomarkers for lung adenocarcinoma. Oncotarget. 2017;8(56):95466-80.

190. de Menezes-Neto A, Sáez MJ, Lozano-Ramos I, Segui-Barber J, Martin-Jaular L, Ullate JM, et al. Size-exclusion chromatography as a stand-alone methodology identifies novel markers in mass spectrometry analyses of plasma-derived vesicles from healthy individuals. J Extracell Vesicles. 2015;4:27378.

191. Shlomovitz I, Erlich Z, Arad G, Edry-Botzer L, Zargarian S, Cohen H, et al. Proteomic analysis of necroptotic extracellular vesicles. Cell Death Dis. 2021;12(11):1059.

192. Datta A, Chen C, Gao YG, Sze SK. Quantitative Proteomics of Medium-Sized Extracellular Vesicle-Enriched Plasma of Lacunar Infarction for the Discovery of Prognostic Biomarkers. Int J Mol Sci. 2022;23(19).

193. Lischnig A, Bergqvist M, Ochiya T, Lässer C. Quantitative Proteomics Identifies Proteins Enriched in Large and Small Extracellular Vesicles. Mol Cell Proteomics. 2022;21(9):100273.

194. De Buck M, Gouwy M, Wang JM, Van Snick J, Opdenakker G, Struyf S, Van Damme J. Structure and Expression of Different Serum Amyloid A (SAA) Variants and their Concentration-Dependent Functions During Host Insults. Curr Med Chem. 2016;23(17):1725-55.

195. Benson MD, Aldo-Benson M. Effect of purified protein SAA on immune response in vitro: mechanisms of suppression. J Immunol. 1979;122(5):2077-82.

196. Marques de Menezes EG, Deng X, Liu J, Bowler SA, Shikuma CM, Stone M, et al. Plasma CD16(+) Extracellular Vesicles Associate with Carotid Artery Intima-Media Thickness in HIV(+) Adults on Combination Antiretroviral Therapy. mBio. 2022;13(3):e0300521.

197. Pavón MA, Parreño M, Téllez-Gabriel M, León X, Arroyo-Solera I, López M, et al. CKMT1 and NCOA1 expression as a predictor of clinical outcome in patients with advanced-stage head and neck squamous cell carcinoma. Head Neck. 2016;38 Suppl 1:E1392-403.

198. Rana R, Chauhan K, Gautam P, Kulkarni M, Banarjee R, Chugh P, et al. Plasma-Derived Extracellular Vesicles Reveal Galectin-3 Binding Protein as Potential Biomarker for Early Detection of Glioma. Front Oncol. 2021;11:778754.

199. ExoCarta. SAA2 2022 [updated 2nd of November 2022. Available from: http://exocarta.org/gene_summary?gene_id=6289

200. Benditt EP, Eriksen N. Amyloid protein SAA is associated with high density lipoprotein from human serum. Proc Natl Acad Sci U S A. 1977;74(9):4025-8.

201. Sasatani M, Zaharieva EK, Kamiya K. The in vivo role of Rev1 in mutagenesis and carcinogenesis. Genes Environ. 2020;42:9.

202. Sasatani M, Xi Y, Kajimura J, Kawamura T, Piao J, Masuda Y, et al. Overexpression of Rev1 promotes the development of carcinogen-induced intestinal adenomas via accumulation of point mutation and suppression of apoptosis proportionally to the Rev1 expression level. Carcinogenesis. 2017;38(5):570-8.

203. Okuda T, Lin X, Trang J, Howell SB. Suppression of hREV1 expression reduces the rate at which human ovarian carcinoma cells acquire resistance to cisplatin. Mol Pharmacol. 2005;67(6):1852-60.

204. Clark DR, Zacharias W, Panaitescu L, McGregor WG. Ribozyme-mediated REV1 inhibition reduces the frequency of UV-induced mutations in the human HPRT gene. Nucleic Acids Res. 2003;31(17):4981-8.

205. Pernicone N, Elias M, Onn I, Tobi D, Listovsky T. Disrupting the MAD2L2-Rev1 Complex Enhances Cell Death upon DNA Damage. Molecules. 2022;27(3).

206. Chen Y, Jie X, Xing B, Wu Z, Yang X, Rao X, et al. REV1 promotes lung tumorigenesis by activating the Rad18/SERTAD2 axis. Cell Death Dis. 2022;13(2):110.

207. Amos CI, Xu W, Spitz MR. Is there a genetic basis for lung cancer susceptibility? Recent Results Cancer Res. 1999;151:3-12.

208. Esaki N, Nakamura T, Tanaka H, Soda K. Selenocysteine lyase, a novel enzyme that specifically acts on selenocysteine. Mammalian distribution and purification and properties of pig liver enzyme. J Biol Chem. 1982;257(8):4386-91.

209. Stadtman TC. Selenocysteine Lyase. EcoSal Plus. 2004;1(1).

210. Seale LA. Selenocysteine β-Lyase: Biochemistry, Regulation and Physiological Role of the Selenocysteine Decomposition Enzyme. Antioxidants (Basel). 2019;8(9).

211. Jafari C, Panzer U, Steinmetz OM, Zahner G, Stahl RA, Harendza S. Enhanced expression of selenocysteine lyase in acute glomerulonephritis and its regulation by AP-1. Cell Mol Biol Lett. 2006;11(3):424-37.

212. Valavanidis A, Vlachogianni T, Fiotakis K, Loridas S. Pulmonary oxidative stress, inflammation and cancer: respirable particulate matter, fibrous dusts and ozone as major causes of lung carcinogenesis through reactive oxygen species mechanisms. Int J Environ Res Public Health. 2013;10(9):3886-907.

213. Becker NP, Martitz J, Renko K, Stoedter M, Hybsier S, Cramer T, Schomburg L. Hypoxia reduces and redirects selenoprotein biosynthesis. Metallomics. 2014;6(5):1079-86.

214. Sun J, Xiong Y, Jiang K, Xin B, Jiang T, Wei R, et al. Hypoxia-sensitive long noncoding RNA CASC15 promotes lung tumorigenesis by regulating the SOX4/β-catenin axis. J Exp Clin Cancer Res. 2021;40(1):12.

215. Chen YF, Chou CY, Ellory JC, Shen MR. The emerging role of KCl cotransport in tumor biology. Am J Transl Res. 2010;2(4):345-55.

216. Shen MR, Chou CY, Ellory JC. Volume-sensitive KCI cotransport associated with human cervical carcinogenesis. Pflugers Arch. 2000;440(5):751-60.

217. Chen YF, Chou CY, Wilkins RJ, Ellory JC, Mount DB, Shen MR. Motor protein-dependent membrane trafficking of KCl cotransporter-4 is important for cancer cell invasion. Cancer Res. 2009;69(22):8585-93.

218. Shen MR, Chou CY, Hsu KF, Liu HS, Dunham PB, Holtzman EJ, Ellory JC. The KCl cotransporter isoform KCC3 can play an important role in cell growth regulation. Proc Natl Acad Sci U S A. 2001;98(25):14714-9.

219. Hsu YM, Chou CY, Chen HH, Lee WY, Chen YF, Lin PW, et al. IGF-1 upregulates electroneutral K-Cl cotransporter KCC3 and KCC4 which are differentially required for breast cancer cell proliferation and invasiveness. J Cell Physiol. 2007;210(3):626-36.

220. Chiu MH, Liu HS, Wu YH, Shen MR, Chou CY. SPAK mediates KCC3-enhanced cervical cancer tumorigenesis. Febs j. 2014;281(10):2353-65.

221. Hsu YM, Chen YF, Chou CY, Tang MJ, Chen JH, Wilkins RJ, et al. KCl cotransporter-3 down-regulates E-cadherin/beta-catenin complex to promote epithelial-mesenchymal transition. Cancer Res. 2007;67(22):11064-73.

222. Gagnon KB. High-grade glioma motility reduced by genetic knockdown of KCC3. Cell Physiol Biochem. 2012;30(2):466-76.

223. Chakraborty M, Asraf H, Sekler I, Hershfinkel M. ZnR/GPR39 controls cell migration by orchestrating recruitment of KCC3 into protrusions, re-organization of actin and activation of MMP. Cell Calcium. 2021;94:102330.

224. Shiozaki A, Takemoto K, Ichikawa D, Fujiwara H, Konishi H, Kosuga T, et al. The K-Cl cotransporter KCC3 as an independent prognostic factor in human esophageal squamous cell carcinoma. Biomed Res Int. 2014;2014:936401.

225. ExoCarta. SLC12A6 2022 [updated 4th of November 2022. Available from: http://exocarta.org/gene_summary?gene_id=9990

226. Wu H, Ding F, Chen X, Chen S, Shi Z, Liu Q, et al. PRSS1 genotype is associated with prognosis in patients with pancreatic ductal adenocarcinoma. Oncol Lett. 2020;19(1):121-6.

227. Gao F, Li YM, Hong GL, Xu ZF, Liu QC, He QL, et al. PRSS1_p.Leu81Met mutation results in autoimmune pancreatitis. World J Gastroenterol. 2013;19(21):3332-8.

228. Whitcomb DC. Genetic risk factors for pancreatic disorders. Gastroenterology. 2013;144(6):1292-302.

229. Belotti Y, Lim EH, Lim CT. The Role of the Extracellular Matrix and Tumor-Infiltrating Immune Cells in the Prognostication of High-Grade Serous Ovarian Cancer. Cancers (Basel). 2022;14(2).

230. Kim BJ, Fang Y, He H, Xu B. Trypsin-Instructed Self-Assembly on Endoplasmic Reticulum for Selectively Inhibiting Cancer Cells: Dedicated to Professor George M. Whitesides on the occasion of his 80th birthday. Adv Healthc Mater. 2021;10(4):e2000416.

231. Erinjeri NJ, Nicolson NG, Deyholos C, Korah R, Carling T. Whole-Exome Sequencing Identifies Two Discrete Druggable Signaling Pathways in Follicular Thyroid Cancer. J Am Coll Surg. 2018;226(6):950-9.e5.

232.   Ye D, Li Y, Zhang H, Zhou Z, Tang Y, Wu P, et al. Silencing PRSS1 suppresses the growth and proliferation of gastric carcinoma cells via the ERK pathway. Int J Biol Sci. 2021;17(4):957-71.

233.   Huang H, Du J, Jin B, Pang L, Duan N, Huang C, et al. Combination of Urine Exosomal mRNAs and lncRNAs as Novel Diagnostic Biomarkers for Bladder Cancer. Front Oncol. 2021;11:667212.

234.   Ludovini V, Bianconi F, Siggillino A, Piobbico D, Vannucci J, Metro G, et al. Gene identification for risk of relapse in stage I lung adenocarcinoma patients: a combined methodology of gene expression profiling and computational gene network analysis. Oncotarget. 2016;7(21):30561-74.

235.   Chen G, Umelo IA, Lv S, Teugels E, Fostier K, Kronenberger P, et al. miR-146a inhibits cell growth, cell migration and induces apoptosis in non-small cell lung cancer cells. PLoS One. 2013;8(3):e60317.

236.   Wang C, Zhang W, Zhang L, Chen X, Liu F, Zhang J, et al. miR-146a-5p mediates epithelial-mesenchymal transition of oesophageal squamous cell carcinoma via targeting Notch2. Br J Cancer. 2016;115(12):1548-54.

237.   Mei J, Bachoo R, Zhang CL. MicroRNA-146a inhibits glioma development by targeting Notch1. Mol Cell Biol. 2011;31(17):3584-92.

238.   Han J, Liu Y, Yang S, Wu X, Li H, Wang Q. MEK inhibitors for the treatment of non-small cell lung cancer. J Hematol Oncol. 2021;14(1):1.

239.   ExoCarta. Detection of the serine protease family members in different organisms on a protein level 2022 [Available from: http://exocarta.org/browse_results?org_name=&cont_type=protein&tissue=&gene_symbol.

240.   UniProt. GPM6A 2022 [Available from: https://www.uniprot.org/uniprotkb/P51674/entry.

241.   Chen L, Zhuo D, Chen J, Yuan H. Screening feature genes of lung carcinoma with DNA microarray analysis. Int J Clin Exp Med. 2015;8(8):12161-71.

242.   Jiang W, Han X, Wang J, Wang L, Xu Z, Wei Q, et al. miR-22 enhances the radiosensitivity of small-cell lung cancer by targeting the WRNIP1. J Cell Biochem. 2019;120(10):17650-61.

243.   Sweef O, Yang C, Wang Z. The Oncogenic and Tumor Suppressive Long Non-Coding RNA-microRNA-Messenger RNA Regulatory Axes Identified by Analyzing Multiple Platform Omics Data from Cr(VI)-Transformed Cells and Their Implications in Lung Cancer. Biomedicines. 2022;10(10).

244.   Li ZR, Xu G, Zhu LY, Chen H, Zhu JM, Wu J. GPM6A expression is suppressed in hepatocellular carcinoma through miRNA-96 production. Lab Invest. 2022;102(11):1280-91.

245.   Lacore MG, Delmas C, Nicaise Y, Kowalski-Chauvel A, Cohen-Jonathan-Moyal E, Seva C. The Glycoprotein M6a Is Associated with Invasiveness and Radioresistance of Glioblastoma Stem Cells. Cells. 2022;11(14).

246.   Ye Z, Li Y, Xie J, Feng Z, Yang X, Wu Y, et al. Integrated bioinformatics identifies the dysregulation induced by aberrant gene methylation in colorectal carcinoma. Genes Dis. 2021;8(4):521-30.

247.   Lankes W, Griesmacher A, Grünwald J, Schwartz-Albiez R, Keller R. A heparin-binding protein involved in inhibition of smooth-muscle cell proliferation. Biochem J. 1988;251(3):831-42.

248.   Clucas J, Valderrama F. ERM proteins in cancer progression. J Cell Sci. 2014;127(Pt 2):267-75.

249.   Takeuchi K, Sato N, Kasahara H, Funayama N, Nagafuchi A, Yonemura S, et al. Perturbation of cell adhesion and microvilli formation by antisense oligonucleotides to ERM family members. J Cell Biol. 1994;125(6):1371-84.

250.   Kobayashi H, Sagara J, Kurita H, Morifuji M, Ohishi M, Kurashina K, Taniguchi S. Clinical significance of cellular distribution of moesin in patients with oral squamous cell carcinoma. Clin Cancer Res. 2004;10(2):572-80.

251.   Zhu X, Morales FC, Agarwal NK, Dogruluk T, Gagea M, Georgescu MM. Moesin is a glioma progression marker that induces proliferation and Wnt/β-catenin pathway activation via interaction with CD44. Cancer Res. 2013;73(3):1142-55.

252.   Charafe-Jauffret E, Monville F, Bertucci F, Esterni B, Ginestier C, Finetti P, et al. Moesin expression is a marker of basal breast carcinomas. Int J Cancer. 2007;121(8):1779-85.

253.   Tokunou M, Niki T, Saitoh Y, Imamura H, Sakamoto M, Hirohashi S. Altered expression of the ERM proteins in lung adenocarcinoma. Lab Invest. 2000;80(11):1643-50.

254.   Li YQ, Zheng Z, Liu QX, Lu X, Zhou D, Zhang J, et al. Moesin as a prognostic indicator of lung adenocarcinoma improves prognosis by enhancing immune lymphocyte infiltration. World J Surg Oncol. 2021;19(1):109.

255.   Zhang X, Liu W, Edaki K, Nakazawa Y, Kamioka H, Fujita A, et al. Correlations of mRNA Levels among Efflux Transporters, Transcriptional Regulators, and Scaffold Proteins in Non-Small-Cell Lung Cancer. Can J Infect Dis Med Microbiol. 2021;2021:4005327.

256.   Couch Y, Buzàs EI, Di Vizio D, Gho YS, Harrison P, Hill AF, et al. A brief history of nearly EV-erything - The rise and rise of extracellular vesicles. J Extracell Vesicles. 2021;10(14):e12144.

257.   Ozkocak DC, Phan TK, Poon IKH. Translating extracellular vesicle packaging into therapeutic applications. Front Immunol. 2022;13:946422.

258.   Dong Q, Liu X, Cheng K, Sheng J, Kong J, Liu T. Pre-metastatic Niche Formation in Different Organs Induced by Tumor Extracellular Vesicles. Front Cell Dev Biol. 2021;9:733627.

259.   Zhao H, Achreja A, Iessi E, Logozzi M, Mizzoni D, Di Raimo R, et al. The key role of extracellular vesicles in the metastatic process. Biochim Biophys Acta Rev Cancer. 2018;1869(1):64-77.

260.   Lima LG, Ham S, Shin H, Chai EPZ, Lek ESH, Lobb RJ, et al. Tumor microenvironmental cytokines bound to cancer exosomes determine uptake by cytokine receptor-expressing cells and biodistribution. Nat Commun. 2021;12(1):3543.

261.   Maugeri M, Nawaz M, Papadimitriou A, Angerfors A, Camponeschi A, Na M, et al. Linkage between endosomal escape of LNP-mRNA and loading into EVs for transport to other cells. Nat Commun. 2019;10(1):4333.

262.   Su H, Rustam YH, Masters CL, Makalic E, McLean CA, Hill AF, et al. Characterization of brain-derived extracellular vesicle lipids in Alzheimer's disease. J Extracell Vesicles. 2021;10(7):e12089.

263.   Akbar N, Azzimato V, Choudhury RP, Aouadi M. Extracellular vesicles in metabolic disease. Diabetologia. 2019;62(12):2179-87.

264.   Brzozowski JS, Jankowski H, Bond DR, McCague SB, Munro BR, Predebon MJ, et al. Lipidomic profiling of extracellular vesicles derived from prostate and prostate cancer cell lines. Lipids Health Dis. 2018;17(1):211.

265.   Buzas EI. The roles of extracellular vesicles in the immune system. Nat Rev Immunol. 2023;23(4):236-50.

266.   Zhang L, Zhu B, Zeng Y, Shen H, Zhang J, Wang X. Clinical lipidomics in understanding of lung cancer: Opportunity and challenge. Cancer Lett. 2020;470:75-83.

267.   Brodin P, Davis MM. Human immune system variation. Nat Rev Immunol. 2017;17(1):21-9.

268.   Parkin J, Cohen B. An overview of the immune system. Lancet. 2001;357(9270):1777-89.

269.   Alberts B JA, Lewis J, Raff M, Roberts K, Walter P. Innate Immunity. Molecular Biology of the Cell. 4th edition ed2002.

270.   Bernareggi D, Pouyanfard S, Kaufman DS. Development of innate immune cells from human pluripotent stem cells. Exp Hematol. 2019;71:13-23.

271.   Hossain Z, Reza A, Qasem WA, Friel JK, Omri A. Development of the immune system in the human embryo. Pediatr Res. 2022;92(4):951-5.

272.   Waldmann H. Human Monoclonal Antibodies: The Benefits of Humanization. Methods Mol Biol. 2019;1904:1-10.

273.   Abbott M, Ustoyev Y. Cancer and the Immune System: The History and Background of Immunotherapy. Semin Oncol Nurs. 2019;35(5):150923.

274.   Sung BH, Parent CA, Weaver AM. Extracellular vesicles: Critical players during cell migration. Dev Cell. 2021;56(13):1861-74.

275.   Buzás EI, Tóth E, Sódar BW, Szabó-Taylor K. Molecular interactions at the surface of extracellular vesicles. Semin Immunopathol. 2018;40(5):453-64.

276.   Capello M, Vykoukal JV, Katayama H, Bantis LE, Wang H, Kundnani DL, et al. Exosomes harbor B cell targets in pancreatic adenocarcinoma and exert decoy function against complement-mediated cytotoxicity. Nat Commun. 2019;10(1):254.

277.   Pretti MAM, Bernardes SS, da Cruz JGV, Boroni M, Possik PA. Extracellular vesicle-mediated crosstalk between melanoma and the immune system: Impact on tumor progression and therapy response. J Leukoc Biol. 2020;108(4):1101-15.

278.   Zhu L, Kalimuthu S, Gangadaran P, Oh JM, Lee HW, Baek SH, et al. Exosomes Derived From Natural Killer Cells Exert Therapeutic Effect in Melanoma. Theranostics. 2017;7(10):2732-45.

279.	Düchler M, Czernek L, Peczek L, Cypryk W, Sztiller-Sikorska M, Czyz M. Melanoma-Derived Extracellular Vesicles Bear the Potential for the Induction of Antigen-Specific Tolerance. Cells. 2019;8(7).

280.	Leone RD, Powell JD. Metabolism of immune cells in cancer. Nat Rev Cancer. 2020;20(9):516-31.

281.	Namba Y, Sogawa C, Okusha Y, Kawai H, Itagaki M, Ono K, et al. Depletion of Lipid Efflux Pump ABCG1 Triggers the Intracellular Accumulation of Extracellular Vesicles and Reduces Aggregation and Tumorigenesis of Metastatic Cancer Cells. Front Oncol. 2018;8:376.

282.	Pantel K, Dickmanns A, Zippelius A, Klein C, Shi J, Hoechtlen-Vollmar W, et al. Establishment of micrometastatic carcinoma cell lines: a novel source of tumor cell vaccines. J Natl Cancer Inst. 1995;87(15):1162-8.

283.	Putz E, Witter K, Offner S, Stosiek P, Zippelius A, Johnson J, et al. Phenotypic characteristics of cell lines derived from disseminated cancer cells in bone marrow of patients with solid epithelial tumors: establishment of working models for human micrometastases. Cancer Res. 1999;59(1):241-8.

284.	Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 1970;227(5259):680-5.

285.	Neuhoff V, Arold N, Taube D, Ehrhardt W. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. Electrophoresis. 1988;9(6):255-62.

286.	Bartkowiak K, Wieczorek M, Buck F, Harder S, Moldenhauer J, Effenberger KE, et al. Two-dimensional differential gel electrophoresis of a cell line derived from a breast cancer micrometastasis revealed a stem/ progenitor cell protein profile. J Proteome Res. 2009;8(4):2004-14.

287.	Suttapitugsakul S, Xiao H, Smeekens J, Wu R. Evaluation and optimization of reduction and alkylation methods to maximize peptide identification with MS-based proteomics. Mol Biosyst. 2017;13(12):2574-82.

288.	Shevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal Chem. 1996;68(5):850-8.

289.	Scientific T. The Orbitrap Tribrid Mass spectrometer ThermoFisher Scientific. 2022.

290.	Olsen JV, Macek B, Lange O, Makarov A, Horning S, Mann M. Higher-energy C-trap dissociation for peptide modification analysis. Nat Methods. 2007;4(9):709-12.

10. Supplement

10.1 Supplemental Figures
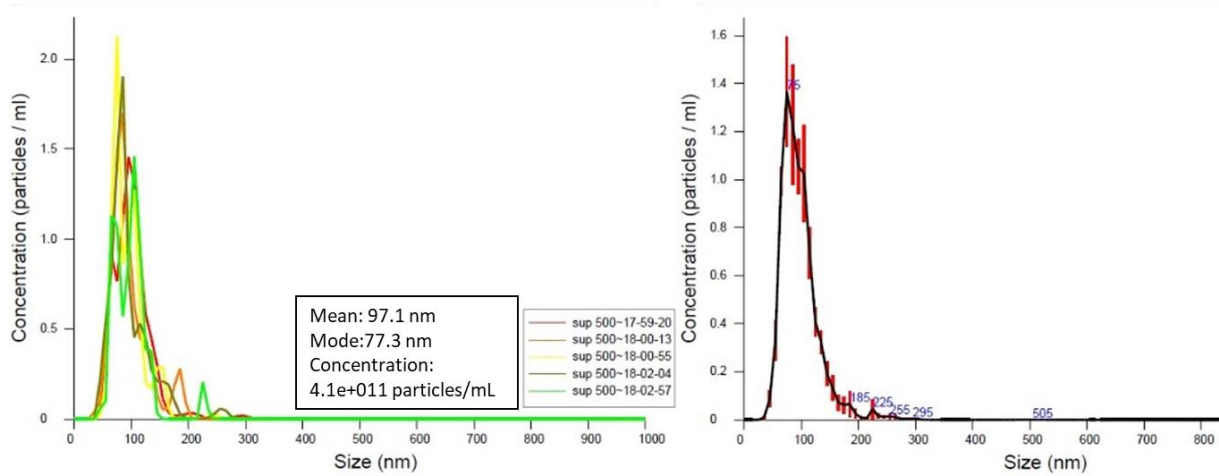
Figure S1. NTA of initial supernatant used for the isolation of EVs for comparison of three different isolation methods. Graph on the left is plotted with EV concentration on y-axis and EVs size on x-axis for each of five NTA measurements of supernatant. Graph on the right is plotted with average EV concentration on y-axes and EVs size on x-axis for the experiment of each supernatant measurement within it. NTA of initial supernatant revealed mean size of EVs 97.1 nm, mode size 77.3 nm, and EVs concentration of 4.1e+011 particles/mL.
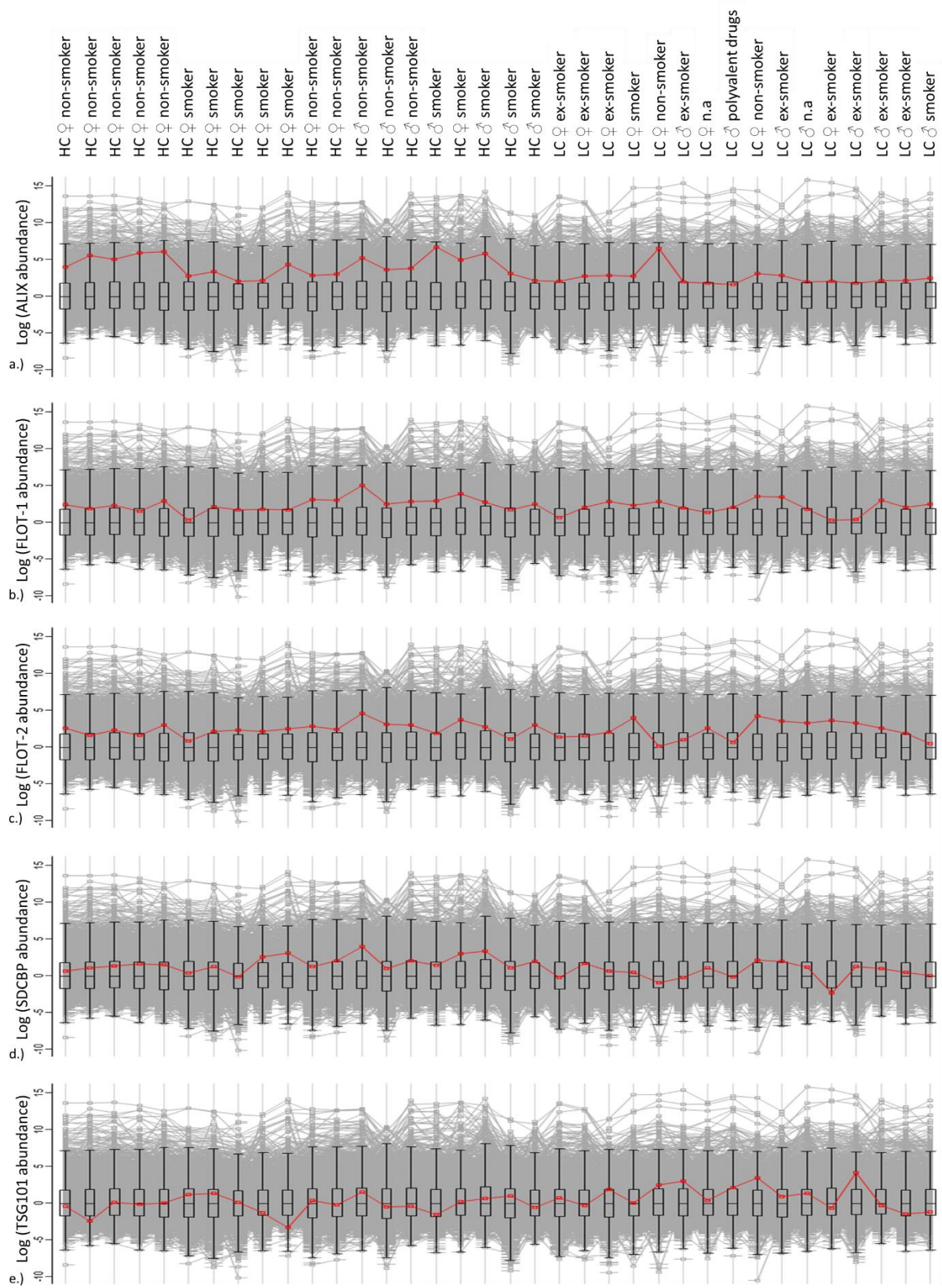
Figure S2. Profile plots of EV biomarker protein abundance across the study cohort. a.) Profile plot of ALIX. b.) Profile plot of FLOT-1. c.) Profile plot of FLOT-2. d.) Profile plot of SDCBP. e.) Profile plot TSG-101.
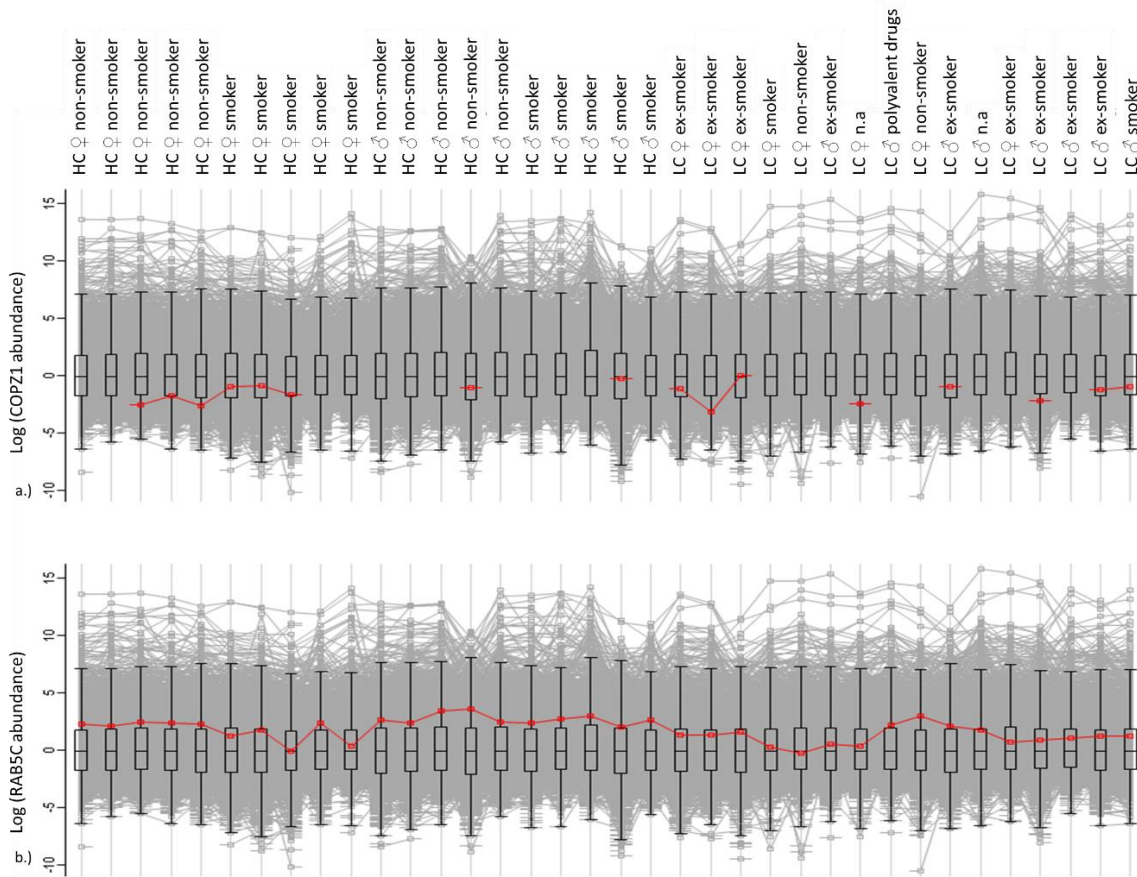
Figure S3. Profile plot of COPZ1 and RAB5C protein abundance across the study cohort.

10.2 Supplemental Tables

Table S1. b- and y-ion fragment masses of SAAVTSEFHLVPSR peptide from NECT-4.

#1	b⁺	b²⁺	b³⁺	Seq.	y⁺	y²⁺	y³⁺	#2
1	88.03930	44.52329	30.01795	S				14
2	159.07642	80.04185	53.69699	A	1419.76862	710.38795	473.92772	13
3	230.11353	115.56040	77.37603	A	1348.73150	674.86939	450.24869	12
4	329.18195	165.09461	110.39883	V	1277.69439	639.35083	426.56965	11
5	430.22962	215.61845	144.08139	T	1178.62598	589.81663	393.54684	10
6	517.26165	259.13446	173.09207	S	1077.57830	539.29279	359.86428	9
7	646.30425	323.65576	216.10627	E	990.54627	495.77677	330.85361	8
8	793.37266	397.18997	265.12907	F	861.50368	431.25548	287.83941	7
9	930.43157	465.71942	310.81538	H	714.43526	357.72127	238.81661	6
10	1043.51564	522.26146	348.51006	L	577.37635	289.19181	193.13030	5
11	1142.58405	571.79566	381.53287	V	464.29229	232.64978	155.43561	4
12	1239.63681	620.32205	413.88379	P	365.22387	183.11558	122.41281	3
13	1326.66884	663.83806	442.89447	S	268.17111	134.58919	90.06189	2
14				R-Label:13C(6)	181.13908	91.07318	61.05121	1

Table S2. b- and y-ion fragment masses of QYTSPEEIDAQLQAEK peptide from PDAP1.

#1	b⁺	b²⁺	Seq.	y⁺	y²⁺	#2
1	129.06585	65.03657	Q			16
2	292.12918	146.56823	Y	1721.82284	861.41506	15
3	393.17686	197.09207	T	1558.75951	779.88339	14
4	480.20889	240.60808	S	1457.71183	729.35955	13
5	577.26165	289.13446	P	1370.67980	685.84354	12
6	706.30425	353.65576	E	1273.62704	637.31716	11
7	835.34684	418.17706	E	1144.58444	572.79586	10
8	948.43090	474.71909	I	1015.54185	508.27456	9
9	1063.45785	532.23256	D	902.45779	451.73253	8
10	1134.49496	567.75112	A	787.43084	394.21906	7
11	1262.55354	631.78041	Q	716.39373	358.70050	6
12	1375.63760	688.32244	L	588.33515	294.67121	5
13	1503.69618	752.35173	Q	475.25109	238.12918	4
14	1574.73329	787.87028	A	347.19251	174.09989	3
15	1703.77589	852.39158	E	276.15540	138.58134	2
16			K	147.11280	74.06004	1

Table S3. b- and y-ion fragment masses of FTYAGCLSVK peptide from CCN1.

#1	b⁺	b²⁺	Seq.	y⁺	y²⁺	#2
1	148.07569	74.54148	F			10
2	249.12337	125.06532	T	998.49754	499.75241	9
3	412.18670	206.59699	Y	897.44986	449.22857	8
4	483.22381	242.11554	A	734.38654	367.69691	7
5	540.24527	270.62628	G	663.34942	332.17835	6
6	700.27592	350.64160	Carbamidometh	606.32796	303.66762	5
7	813.35999	407.18363	L	446.29731	223.65229	4
8	900.39202	450.69965	S	333.21325	167.11026	3
9	999.46043	500.23385	V	246.18122	123.59425	2
10			K	147.11280	74.06004	1

Table S4. Frequency of NECT-4⁺ lung cancer patients across different histological subtypes.

Histology	n (total)	n (NECT-4⁺)	% (NECT-4⁺)
Adenocarcinoma	81	53	65.43
Squamous cell carcinoma	48	23	47.92
Small cell carcinoma	40	19	47.5
Large cell carcinoma	12	3	25
Non-small cell carcinoma	10	5	50
Other	2	1	50

Table S5. Characteristics of patients used for the EV proteomic analysis.

Sample	UICC stage	TNM	Histology	Gender	Age	Smoking status	Pack Years
LC	IV	T4N1M1b	adenocarcinoma	female	62.68	ex-smoker	80
LC	IV	T4N2M1b	adenocarcinoma	male	77.60	n.a	n/a
LC	IV	cT3cN3cM1a	adenocarcinoma	female	62.96	ex-smoker	50
LC	IV	cT3cN0pM1b	squamous carcinoma	male	70.87	ex-smoker	50
LC	IV	cT2acN1cM1b	adenocarcinoma	female	51.51	smoker	18
LC	IV	cT3cN3cM1b	adenocarcinoma	female	27.27	non-smoker	0
LC	IV	pT3pN2cM1	adenocarcinoma	male	50.59	ex-smoker	30
LC	IV	pT1apN0pM1b	adenocarcinoma	female	45.19	n.a	n/a
LC	IV	cT3cNxcM1b	NOS	male	41.33	polyvalent drug abuse	n/a
LC	IV	cT2cN0cM1a	adenocarcinoma	female	58.02	non-smoker	0
LC	IV	pT2apN0pM1c	adenocarcinoma	female	55.04	ex-smoker	25
LC	IV	cT1bcN3cM1c	adenocarcinoma	female	55.64	ex-smoker	n/a
LC	IV	cT1bcN0cM1b	adenocarcinoma	male	74.04	ex-smoker	120
LC	IV	cTxNxM1c	adenocarcinoma	male	50.28	ex-smoker	40
LC	IV	T2aN0M1b	adenocarcinoma	female	48.64	non-smoker	0
LC	IIIa	cT4CN2M0	adenosquamous carcinoma	male	75.47	smoker	50

LC=lung cancer

NOS=not otherwise specified

n.a= status not assined

Table S6. Hazardous substances

Chemical	Hazard statement	Precautionary statement	GHS hazard pictograms
2-mercaptoethanol	H301+H331, H310, H315, H317, H318, H361d, H373, H410	P201, P262, P280, P301+P310+P330, P302+P352+P310, P305+P351+P338+P310	
$^{13}C_6$-Arginine-HCl	H319	P305+P351+P338	
$^{13}C_6$-Lysine-2HCl	N/A	N/A	N/A
Acetone	H225, H319, H336	P210, P240, P305+P351+P338, P403+P203	
Acetonitrile	H225, H302+H312+H332, H319	P210, P240, P302+P352, P305+P351+P338, P403+P233	
Acrylamide	H301, H312, H332, H315, H317, H319, H340, H350, H361f, H372	P201, P280, P302+P352, P304+P340, P305+P351+P338, P308+P310	
Ammonium bicarbonate	H302	P301+P312, P330	
Ammonium persulfate	H272, H302, H315, H317, H319, H334, H335	P220, P261, P280, P305+P351+P338, P342+P311	

158

Dithiothreitol	H302, H315, H319, H335	P261, P305+P351+P338	
Dry Strip Cover Fluid PlusOne	H304, H410	P273, P301+P310, P331, P405, P501	
Ethanol	H225, H319	P210, P240, P305+P351+P338, P403+P233	
Ethylenediaminetetraacetic acid (EDTA)	H319, H332, H373	P280, P304+P340, P312, P305+P351+P338, P337+P313	
Formic acid	H226, H302, H314, H331	P210, P280, P303+P361+P353, P304+P340+P310, P305+P351+P338, P403+P233	
Hydrochloric acid	H290, H314, H335	P260, P280, P303+P361+P353, P304+P340+P310, P305+P351+P338	
Iodoacetamide	H301, H317, H334	P261, P280, P301+P310, P342+P311	
Isopropyl alcohol	H225, H319, H336	P210, P240, P305+P351+P338, P403+P233	
Methanol	H225, H331, H311, H301, H370	P210, P233, P280, P302+P352, P304+P340, P308+P310, P403+P235	
Nonidet P-40	H302, H319, H411	P101-P103; P273, P280, P301+P312, P305+P351+P338, P337+P313, P501	
Ortho-Phosphoric acid	H290, H314	P280, P301+P330+P331, P305+P351+P338, P308+P310	
Paraformaldehyde	H228, H302+H332, H315, H317, H318, H335, H350	P201, P210, P280, P302+P352, P305+P351+P338	
Penicillin/Streptomycin	H317, H334, H361d	P201, P261, P264, P272, P280, P302+P352, P304+P340, P333+P313, P342+P311	
Protease inhibitor cocktail: cOmplete ULTRA Tablets, Mini, EASYpack	H314, H412	P260, P73, P280	
Phosphatase inhibitor cocktail: PhosSTOP EASYpack	H301	P264, P270, P301+P310+P330, P405, P501	

Sodium deoxycholate	H302, H335	P261, P264, P270, P271, P301+P312, P304+P340, P312, P330, P403+P233, P405, P501	
Sodium dodecyl sulfate	H228, H302+H332, H315, H318, H335, H412	P210, P261, P280, P301+P312+P330, P305+P351+P338+P310, P370+P378	
Sodium azide	H300+H310+H330, H373, H410	P270, P273, P280, P302+P352, P304+P340, P310	
Sulfuric acid	H290, H314, H315, H318, H319	P280, P301+P330+P331, P303+P361+P353, P305+P351+P338+P310	
Tetramethylethylenediamine	H225, H332, H302, H314	P210, P280, P305+P351+P338, P310	
Thimerosal	H300+H310+H330, H373, H410	P260, P273, P280, P301+P310+P330, P302+P352, P310, P304+P340+P310	
Triton X-100	H302, H315, H318, H410	P280, P301+P312+P330, P305+P351+P338+P310	
Trypsin sequencing grade	H315, H319, H334, H335	P261, P264, P280, P284	

11. Acknowledgements

With this, I would like to thank everyone who contributed to this doctoral thesis directly and indirectly, as well as to everyone who have encouraged me, challenged me or smoothened this doctoral journey for me.

Firstly, I would like to say big thank you to all patients who have supported science as well as contributed for the greater good for future generations by donating their blood. I have special gratitude and respect for all the patients who have made such an unselfish decision.

Particularly big thank you to Prof. Klaus Pantel for giving me an opportunity to work in his group at the Institute of Tumor Biology. It was a pleasure to learn from you and I am thankful for all exchanged ideas and information. Also, I would like to thank you for pushing me forward and supporting me in my professional growth.

I would like to thank to European Union who funded and supported this project. This project has received funding from the European Union´s Horizon 2020 research and innovation programme under the Marie Skłodovska-Curie grant agreement number No. 765492.

Additionally, I would like to thank to my supervisor Dr. Kai Bartkowiak for the support during my doctoral research. For sharing his knowledge in proteomics and mass spectrometry and allowing me to learn and grow scientifically. Moreover, I want to thank you for encouraging me in an independent work and decision making.

Additionally, I want to thank to Antje Andreas for always stepping in and helping out with her great technical knowledge and experience. Thank you for always being available and so supportive. I also want to thank you for all the laughs and jokes we shared in the lab.

Moreover, I would like to thank Dr. Laura Keller for sharing her patient samples with me allowing me to work on method optimization. Also, thank you for all the scientific talks we had and for all the useful and very helpful advices you gave to me. I was always very appreciative of your support and happy to have a colleague like you.

Special thank you to my co-supervisor Prof. Hartmut Schlüter for supporting me and advising me whenever needed. Thank you for sharing your valuable knowledge in mass spectrometry and for guiding me in right directions when needed.

Further on, I would like to thank to the WG Schlüter, Sönke Harder and Hannah Voß in particular for their technical and scientific support in mass spectrometry and bioinformatics, which allowed me to pursue with my PhD work. Thank you for having patience with me, for presenting me new skills, and for always being available.

[HRV] Veliko hvala mojoj obitelji, mojim roditeljima Iliji i Nadi Ačkar, bratu i šogorici Kristianu i Željki Ačkar, kao i našim zvrkastim mališanima, nećaku i nećakinji Karli i Franu Ačkar. Nemam riječi kojima bih opisala koliko mi znači vaša podrška. Znajući da vas imam uz sebe me motivira i potiče da slijedim sve svoje želje i snove. Bez vas ne bih bila ovdje gdje jesam. Hvala vam što vjerujete u mene kada ja ne vjerujem. Hvala vam što me kritizirate kada ja sama nisam dovoljno kritična i hvala što me nasmijavate do suza kada mi se plače. Hvala vam za sve što ste mi omogućili. Volim vas.

Finally, a massive thank you to my boyfriend Ricardo Duran Sagarnaga. You were living this PhD journey as much as I did and thank you for never giving up on me and unconditionally supporting me, even when I could not give as much of myself. Thank you for all your patience when I was stressing out and for comforting me. Thank you for every dinner you brought to me when I was staying until late. Thank you for making hard moments easier as well as for sharing the excitement with me in the good times. Above all, thank you for all the happiness you brought into my life and for every smile and laugh you caused. Having you by my side, makes everything better. I appreciate you and I am grateful for having you. I love you.

12. Declaration on Oath

I hereby declare on oath that this doctoral dissertation is written independently and solely by my own based on the original work of my PhD and has not been used other than the acknowledged resources and aids. The submitted written version corresponds to that on the electronic storage medium. I certify that this dissertation was not submitted in a previous doctoral procedure.

Hamburg, 23rd of February 2023

Lucija Ackar